Documenta Mathematica

Journal der Deutschen Mathematiker-Vereinigung

Band 4

1999

ISSN 1431-0635 Print   ISSN 1431-0643 Internet

Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, veröffentlicht Forschungsarbeiten aus allen mathematischen Gebieten und wird in traditioneller Weise referiert.

Documenta Mathematica erscheint am World Wide Web unter der Adresse:

`http://www.mathematik.uni-bielefeld.de/documenta`

Artikel können als TEX-Dateien per E-Mail bei einem der Herausgeber eingereicht werden. Hinweise für die Vorbereitung der Artikel können unter der obigen WWW-Adresse gefunden werden.

Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, publishes research manuscripts out of all mathematical fields and is refereed in the traditional manner.

Documenta Mathematica is published on the World Wide Web under the address:

`http://www.mathematik.uni-bielefeld.de/documenta`

Manuscripts should be submitted as TEX files by e-mail to one of the editors. Hints for manuscript preparation can be found under the above WWW-address.

ISSN 1431-0635 Documenta Mathematica (Print)
ISSN 1431-0643 Documenta Mathematica (Internet)

Anschrift des technischen geschäftsführenden Herausgebers:
Ulf Rehmann, Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld

Documenta Mathematica

Journal der Deutschen Mathematiker-Vereinigung

Band 4, 1999

Singularities, Double Points, Controlled Topology and Chain Duality

Andrew Ranicki

Received: August 7, 1998
Revised: February 10, 1999

Communicated by Joachim Cuntz

Abstract. A manifold is a Poincaré duality space without singularities. McCrory obtained a homological criterion of a global nature for deciding if a polyhedral Poincaré duality space is a homology manifold, i.e. if the singularities are homologically inessential. A homeomorphism of manifolds is a degree 1 map without double points. In this paper combinatorially controlled topology and chain complex methods are used to provide a homological criterion of a global nature for deciding if a degree 1 map of polyhedral homology manifolds has acyclic point inverses, i.e. if the double points are homologically inessential.

1991 Mathematics Subject Classification: Primary 55N45, 57R67; Secondary 55U35.
Keywords and Phrases: manifold, Poincaré space, singularity, controlled topology, chain duality.

Introduction

A *chain duality* on an additive category $\mathbb{A}$ is an involution on the derived category of finite chain complexes in $\mathbb{A}$ and chain homotopy classes of chain maps. The precise definition will be recalled in §1. Chain duality was introduced in Ranicki [29] in order to construct the algebraic surgery exact sequence of a space X

$$\cdots \to H_n(X;\mathbb{L}_\bullet) \xrightarrow{A} L_n(\mathbb{Z}[\pi_1(X)]) \to \mathbb{S}_n(X) \to H_{n-1}(X;\mathbb{L}_\bullet) \to \dots$$

with $L_*(\mathbb{Z}[\pi_1(X)])$ the surgery obstruction groups of Wall [43], and A the assembly map. Here, $\mathbb{L}_\bullet$ is the 1-connective simply-connected algebraic surgery spectrum of $\mathbb{Z}$, and the generalized homology groups are the (1-connective) L-theory of the X-controlled $\mathbb{Z}$-module category $\mathbb{A}(\mathbb{Z}, X)$ of Ranicki and Weiss [34]

$$H_*(X;\mathbb{L}_\bullet) \;=\; L_*(\mathbb{A}(\mathbb{Z},X)) \ .$$

The algebraic surgery exact sequence was used in [29, Chapter 17] to give algebraic formulations of the obstructions to the two basic questions of Browder-Novikov-Sullivan-Wall surgery theory:

A1. Is an n-dimensional Poincaré duality space X homotopy equivalent to an n-dimensional manifold?

A2. Is a homotopy equivalence $f : M \to N$ of n-dimensional manifolds homotopic to a homeomorphism?

The following are the basic questions of Chapman-Ferry-Quinn controlled topology:

B1. How close is an n-dimensional controlled Poincaré duality space X to being an n-dimensional manifold?

B2. How close is a controlled homotopy equivalence $f : M \to N$ of n-dimensional manifolds to being a homeomorphism?

Here is a very crude approximation to controlled topology. Given a topological space X define an *X-controlled space* to be a space M equipped with a map $p_M : M \to X$. A *map* of X-controlled spaces $f : M \to N$ is a map of the underlying spaces such that there is defined a commutative diagram

$$\begin{tikzcd} M \arrow[rr, "f"] \arrow[dr, "p_M"'] & & N \arrow[dl, "p_N"] \\ & X & \end{tikzcd}$$

The map f is an *X-controlled homology equivalence* if the restrictions

$$f| : p_M^{-1}(x) \to p_N^{-1}(x) \quad (x \in X)$$

induce isomorphisms

$$(f|)_* : H_*(p_M^{-1}(x)) \cong H_*(p_N^{-1}(x)) \ .$$

An *n-dimensional X-controlled Poincaré space* is an X-controlled space N with Lefschetz duality isomorphisms

$$H^{n-*}(N, N\backslash p_N^{-1}(x)) \cong H_*(p_N^{-1}(x)) \quad (x \in X) \ .$$

There are two extreme cases:

- If $X = \{\text{pt.}\}$ then:
  - an X-controlled homology equivalence $f : M \to N$ of X-controlled spaces is just a homology equivalence, with
  $$f_* : H_*(M) \cong H_*(N) \ ,$$
  - an n-dimensional X-controlled Poincaré space N is just an n-dimensional Poincaré space, with
  $$H^{n-*}(N) \cong H_*(N) \ .$$
- If $p_N = 1 : N \to N = X$ then:
  - an N-controlled homology equivalence $f : M \to N$ of N-controlled spaces is just a map with acyclic point inverses, with
  $$(f|)_* : H_*(f^{-1}(x)) \cong H_*(\{x\}) \ \ (x \in N) \ ,$$
  - an n-dimensional N-controlled Poincaré space N is just an n-dimensional homology manifold, with
  $$H^{n-*}(N, N \backslash \{x\}) \cong H_*(\{x\}) \ \ (x \in N) \ .$$

In a more sophisticated exposition of controlled topology X would be a metric space, and the condition $p_M = p_N f$ in the definition of an X-controlled map would be weakened to

$$d(p_M(x), p_N f(x)) < \epsilon \ \ (x \in M)$$

for some $\epsilon > 0$. In principle, Quinn [24] characterized ANR homology manifolds X as metrically X-controlled Poincaré duality spaces. (See Ranicki and Yamasaki [37] for a preliminary account of the metrically controlled L-theory required for the details of the characterization).

The original development of controlled topology for metric spaces involved quite complicated controlled algebra, starting with Connell and Hollingsworth [5]. However, these questions will only be considered here in the combinatorial context of compact polyhedra, homology manifolds and PL maps, for which the controlled algebra is much easier:

C1. Is a polyhedral n-dimensional Poincaré duality space X an n-dimensional homology manifold?

C2. Does a degree 1 PL map $f : M \to N$ of polyhedral n-dimensional homology manifolds have acyclic point inverses?

McCrory [17] obtained a homological obstruction for C1 (under slightly different hypotheses), which was interpreted in Ranicki [29, 8.5] in terms of the chain duality on the X-controlled $\mathbb{Z}$-module category $\mathbb{A}(\mathbb{Z}, X)$. The obstruction is the image in $H^n(X \times X \backslash \Delta_X)$ of the Poincaré dual in $H^n(X \times X)$ of the diagonal class $\Delta_*[X] \in H_n(X \times X)$. The obstruction vanishes if and only if X is an n-dimensional homology manifold, if and only if the $\mathbb{Z}$-module Poincaré duality chain equivalence

$$[X] \cap - : \Delta(X)^{n-*} \to \Delta(X')$$

is an X-controlled chain equivalence.

The main results of this paper are the following homological obstructions for C1 and C2.

Theorem A. An n-dimensional polyhedral Poincaré complex X is an n-dimensional homology manifold if and only if there is defined a Lefschetz duality isomorphism

$$H^n(X \times X, \Delta_X) \cong H_n(X \times X \backslash \Delta_X) ,$$

with

$$\Delta_X = \{(x, x) \in X \times X \mid x \in X\}$$

the diagonal of X.

Theorem B. A simplicial map $f : M \to N$ of n-dimensional polyhedral homology manifolds has acyclic point inverses if and only if it has degree 1

$$f_*[M] = [N] \in H_n(N)$$

and

$$H_n((f \times f)^{-1}\Delta_N, \Delta_M) = 0 ,$$

with

$$(f \times f)^{-1}\Delta_N = \{(x, y) \in M \times M \mid f(x) = f(y) \in N\}$$

the double point set of f.

Theorems A, B are proved in §§6,7 respectively, appearing as Theorem 6.13 and Corollary 7.5.

Here are the contents of the rest of the paper.

In §8 the obstructions of Theorems A, B are interpreted using bundles, specifically the Spivak normal bundle of a Poincaré complex and the tangent topological block bundle of a homology manifold.

In §9 the obstructions of Theorems A, B are related to the 'total surgery obstruction' $s(X) \in S_n(X)$ of Ranicki [29] for the existence of a topological manifold in the homotopy type of a Poincaré space.

In §10 chain duality is used to develop a combinatorial version of the controlled surgery theory.

In §11 some standard results on intersections and self-intersections of manifolds are interpreted in terms of the chain duality.

In §12 (resp. §13) the controlled topology point of view on Whitehead torsion (resp. fibrations) is adapted to the combinatorially controlled chain homotopy theory.

In §14 some standard results in high-dimensional knot theory are interpreted in terms of the chain duality.

In this paper only *oriented* polyhedral Poincaré complexes and homology manifolds will be considered, and orientation-preserving PL maps between them.

A preliminary version of some of the material in this paper appeared in Ranicki [32].

I am grateful to Michael Weiss for valuable comments which helped improve the exposition of the paper.

1. Chain duality

Let $\mathbb{A}$ be an additive category, and let $\mathbb{B}(\mathbb{A})$ be the additive category of finite chain complexes in $\mathbb{A}$ and chain maps. A contravariant additive functor $T : \mathbb{A} \to \mathbb{B}(\mathbb{A})$ extends to $T : \mathbb{B}(\mathbb{A}) \to \mathbb{B}(\mathbb{A})$ by defining $T(C)$ for a chain complex C to be the total of a double complex, with

$$T(C)_n = \sum_{p+q=n} T(C_{-p})_q \ .$$

Definition 1.1 (Ranicki [29, 1.1])
A *chain duality* (T, e) on $\mathbb{A}$ is a contravariant additive functor $T : \mathbb{A} \to \mathbb{B}(\mathbb{A})$, together with a natural transformation $e : T^2 \to 1$ such that for each object A in $\mathbb{A}$:

- $e(T(A))\,.\,T(e(A)) \,=\, 1 : \, T(A) \,\to\, T(A)\,,$
- $e(A)\colon T^2(A) \to A$ is a chain equivalence.

Chain duality has the following properties:

- The dual of an object A is a chain complex $T(A)$.
- The dual of a chain complex C is a chain complex $T(C)$.

Example 1.2 (i) An *involution* (T,e) on an additive category $\mathbb{A}$ is a chain duality such that $T(A)$ is a 0-dimensional chain complex (= object) for each object A in $\mathbb{A}$, with $e(A) : T^2(A) \to A$ an isomorphism.
(ii) An involution $R \to R; r \mapsto \overline{r}$ on a ring R determines the involution (T,e) on the additive category $\mathbb{A}(R)$ of f.g. free left R-modules with:

- $T(A) = \mathrm{Hom}_R(A,R)$
- $R \times T(A) \to T(A)\ ;\ (r,f) \mapsto (x \mapsto f(x)\overline{r})$
- $e(A)^{-1} : A \to T^2(A)\ ;\ x \mapsto (f \mapsto \overline{f(x)})$.

2. Simplicially controlled algebra

Let X be a simplicial complex, and let R be a commutative ring.

Definition 2.1 (Ranicki and Weiss [34])
(i) An (R,X)-*module* is a finitely generated free R-module A with direct sum decomposition

$$A\ =\ \sum_{\sigma\in X} A(\sigma)\ ,$$

such that each $A(\sigma)$ is a f.g. free R-module.
(ii) An (R,X)-module *morphism* $f : A \to B$ is an R-module morphism such that for each $\sigma \in X$

$$f(A(\sigma)) \subseteq \sum_{\tau\geq\sigma} B(\tau)\ .$$

Write the components of f as $f(\tau,\sigma) : A(\sigma) \to B(\tau)$.

Let $\mathbb{A}(R)$ be the additive category of f.g. free R-modules, and let $\mathbb{A}(R,X)$ be the additive category of (R,X)-modules. Regard the simplicial complex X as the category with objects the simplexes $\sigma \in X$, and morphisms the face inclusions $\sigma \leq \tau$. An (R,X)-module $A = \sum\limits_{\sigma\in X} A(\sigma)$ determines a contravariant functor

$$[A] : X \to \mathbb{A}(R)\ ;\ \sigma \mapsto [A][\sigma] = \sum_{\tau\geq\sigma} A(\tau)\ .$$

The (R,X)-module category $\mathbb{A}(R,X)$ is thus a full subcategory of the category of contravariant functors $X \to \mathbb{A}(R)$.

Proposition 2.2 (Ranicki and Weiss [34, 2.9])
The following conditions on a chain map $f : C \to D$ of finite chain complexes in $\mathbb{A}(R,X)$ are equivalent:
(i) *f is a chain equivalence,*
(ii) *the R-module chain maps*

$$f(\sigma,\sigma) : C(\sigma)\ \to\ D(\sigma)\quad (\sigma \in X)$$

are chain equivalences,
(iii) *the R-module chain maps*

$$[f][\sigma] : [C][\sigma] \ \to\ [D][\sigma] \quad (\sigma \in X)$$

are chain equivalences.

3. Simplicially controlled topology

The *barycentric subdivision* X' of a simplicial complex X is the simplicial complex with the same polyhedron

$$|X'| \ = \ |X|$$

and one n-simplex $\widehat{\sigma}_0\widehat{\sigma}_1 \dots \widehat{\sigma}_n$ for each sequence of simplexes in X

$$\sigma_0 < \sigma_1 < \dots < \sigma_n \ .$$

The *dual cell* of a simplex $\sigma \in X$ is the contractible subcomplex

$$D(\sigma, X) \ = \ \{\widehat{\sigma}_0\widehat{\sigma}_1 \dots \widehat{\sigma}_n \,|\, \sigma \leq \sigma_0\} \subseteq X' \ ,$$

with boundary

$$\partial D(\sigma, X) \ = \ \{\widehat{\sigma}_0\widehat{\sigma}_1 \dots \widehat{\sigma}_n \,|\, \sigma < \sigma_0\} \subseteq D(\sigma, X) \ .$$

Definition 3.1 (i) An *X-controlled simplicial complex* (M, p_M) is a finite simplicial complex M with a simplicial map $p_M : M \to X'$, the *control map.*
(ii) A *map* $f : (M, p_M) \to (N, p_N)$ of X-controlled simplicial complexes is a simplicial map $f : M \to N$ such that $p_M = p_N f : M \to X'$.

In practice, (M, p_M) will be abbreviated to M.

Definition 3.2 The *(R, X)-module chain complex* $\Delta(M; R)$ of an X-controlled simplicial complex M is the R-coefficient simplicial chain complex of M with

$$\Delta(M; R)(\sigma) \ = \ \Delta(p_M^{-1} D(\sigma, X), p_M^{-1} \partial D(\sigma, X); R) \ .$$

and

$$\begin{aligned} [\Delta(M; R)_r][\sigma] \ &= \ \sum_{\tau \geq \sigma} \Delta(M; R)(\tau)_r \\ &= \ \Delta(p_M^{-1} D(\sigma, X); R)_r \quad (r \in \mathbb{Z}, \sigma \in X) \ . \end{aligned}$$

A map of X-controlled simplicial complexes $f : M \to N$ induces an (R, X)-module chain map

$$f : \Delta(M; R) \to \Delta(N; R) \ .$$

Definition 3.3 A map of X-controlled simplicial complexes $f : M \to N$ is an *X-controlled R-homology equivalence* if the restrictions

$$f| : p_M^{-1}D(\sigma, X) \to p_N^{-1}D(\sigma, X) \quad (\sigma \in X)$$

induce isomorphisms in R-homology

$$(f|)_* : H_*(p_M^{-1}D(\sigma, X); R) \cong H_*(p_N^{-1}D(\sigma, X); R) \quad (\sigma \in X) .$$

Proposition 3.4 A map of X-controlled simplicial complexes $f : M \to N$ is an X-controlled R-homology equivalence if and only if the induced (R, X)-module chain map $f : \Delta(M; R) \to \Delta(N; R)$ is a chain equivalence.
Proof Immediate from 2.2. □

Proposition 3.5 (i) *If $X = \{\text{pt.}\}$ an X-controlled map $f : M \to N$ is an X-controlled R-homology equivalence if and only if f induces R-homology isomorphisms*

$$f_* : H_*(M; R) \cong H_*(N; R) .$$

(ii) *If $X = N$ an X-controlled map $f : M \to N$ is an X-controlled R-homology equivalence if and only if f has R-acyclic point inverses*

$$H_*(f^{-1}(x); R) \cong H_*(\{x\}; R) \quad (x \in |X|) .$$

Proof (i) Immediate from 3.4, since a chain map of finite free R-module chain complexes is a chain equivalence if and only if it induces isomorphisms in homology.
(ii) Immediate from 3.4, since every point $x \in |X|$ is in the interior $D(\sigma, X)\backslash\partial D(\sigma, X)$ of a unique dual cell $D(\sigma, X)$, and

$$H_*(\{x\}; R) \cong H_*(D(\sigma, X); R) \quad , \quad H_*(f^{-1}(x); R) \cong H_*(f^{-1}D(\sigma, X); R) .$$

□

Here is another way in which (R, X)-module chain complexes arise:

Definition 3.6 (Ranicki [29, 4.2])
Let $\Delta^{-*}(X; R)$ be the (R, X)-module chain complex defined by

$$\begin{aligned}
\Delta^{-*}(X; R) \ &= \ \mathrm{Hom}_R(\Delta(X; R), R)_{-*} \ , \\
\Delta^{-*}(X; R)_r(\sigma) \ &= \ \begin{cases} R & \text{if } r = -|\sigma| \\ 0 & \text{otherwise.} \end{cases} \quad (r \in \mathbb{Z}, \sigma \in X) \ .
\end{aligned}$$

As an R-module chain complex $\Delta^{-*}(X; R)$ is just the R-coefficient simplicial cochain complex of X regraded to be a chain complex.

4. The (R,X)-module chain duality

Proposition 4.1 (Ranicki [29, 5.1])
The additive category $\mathbb{A}(R,X)$ of (R,X)-modules has a chain duality (T,e) with the dual of an (R,X)-module A the (R,X)-module chain complex

$$T(A) \;=\; \mathrm{Hom}_R(\mathrm{Hom}_{(R,X)}(\Delta^{-*}(X;R),A),R)$$

with

- $TA(\sigma) \;=\; [A][\sigma]^{|\sigma|-*}$,
- $T(A)_r(\sigma) \;=\; \begin{cases} \sum\limits_{\tau\geq\sigma} \mathrm{Hom}_R(A(\tau),R) & \text{if } r=-|\sigma| \\ 0 & \text{if } r\neq -|\sigma| \ . \end{cases}$

The chain duality is such that

$$T(C) \simeq_R \mathrm{Hom}_{(R,X)}(C,\Delta(X';R))^{-*} \simeq_R \mathrm{Hom}_R(C,R)^{-*}$$

for any finite (R,X)-module chain complex C.

Definition 4.2 Given an X-controlled simplicial complex M let

$$\Delta(M;R)^{-*} \;=\; T(\Delta(M;R))$$

be the (R,X)-module chain complex dual to $\Delta(M;R)$.

Note that there is defined an R-module chain equivalence

$$\Delta(M;R)^{-*} \simeq_R \mathrm{Hom}_R(\Delta(M;R),R)^{-*} \ ,$$

with $\mathrm{Hom}_R(\Delta(M;R),R)^{-*}$ the simplicial R-coefficient cochain complex of M regraded to be a chain complex, and note also that

$$\Delta(M;R)^{-*}(\sigma)_r \;=\; \mathrm{Hom}_R(\Delta(p_M^{-1}D(\sigma,X);R)_{-r+|\sigma|},R) \;\; (r\in\mathbb{Z},\sigma\in X) \ .$$

A map of X-controlled simplicial complexes $f:M\to N$ induces an (R,X)-module chain map

$$f^*:\Delta(N;R)^{-*}\to\Delta(M;R)^{-*} \ .$$

The (R,X)-module chain complex $\Delta^{-*}(X;R)$ of 3.6 and the (R,X)-module chain complex $\Delta(X;R)^{-*}$ of 4.2 (with $p_M=1:M\to M=X'$) are related by the (R,X)-module chain equivalence

$$\Delta^{-*}(X;R) \simeq_{(R,X)} \Delta(X;R)^{-*}$$

induced by the projections $\Delta(D(\sigma,X);R)\to R$.

5. Products

Definition 5.1 The *product* of X-controlled simplicial complexes M, N is the pullback X-controlled simplicial complex

$$M \times_X N \;=\; \{(x,y) \in M \times N \,|\, p_M(x) = p_N(y) \in X\}$$

with control map

$$M \times_X N \to X \;;\; (x,y) \mapsto p_M(x) = p_N(y) \;.$$

(Strictly speaking, this only defines a polyhedron $M \times_X N$).

Definition 5.2 The *product* of (R,X)-modules A, B is the (R,X)-module

$$A \otimes_{(R,X)} B \;=\; \sum_{\lambda,\mu \in X, \lambda\cap\mu \neq \emptyset} A(\lambda) \otimes_R B(\mu) \subseteq A \otimes_R B$$

with

$$(A \otimes_{(R,X)} B)(\sigma) \;=\; \sum_{\lambda,\mu \in X, \lambda\cap\mu = \sigma} A(\lambda) \otimes_R B(\mu) \quad (\sigma \in X) \;.$$

Recall the following properties of the products in 5.1,5.2 from Ranicki [29, Chapter 7]. (The product $A \otimes_{(R,X)} B$ was denoted by $A \boxtimes_R B$ in [29, 7.1]).

Proposition 5.3 (i) *For any (R,X)-module chain complexes C, D*

- $C \otimes_{(R,X)} \Delta(X';R) \simeq_{(R,X)} C$,
- $TC \otimes_{(R,X)} D \simeq_R \mathrm{Hom}_{(R,X)}(C,D)$.

(ii) *For any X-controlled simplicial complexes M, N*

- $\Delta(M;R) \otimes_{(R,X)} \Delta(N;R) \simeq_{(R,X)} \Delta(M \times_X N;R)$,
- $\Delta(M;R)^{-*} \otimes_{(R,X)} \Delta(N;R)^{-*}$

$$\simeq_R \;\; Hom_R(\Delta(M \times N, M \times N \backslash M \times_X N;R),R)_{-*} \;,$$

(iii) *The Alexander-Whitney diagonal chain approximation of the barycentric subdivision X' of X is an R-module chain map*

$$\Delta : \Delta(X';R) \to \Delta(X';R) \otimes_R \Delta(X';R) \;;\; (\widehat{x}_0 \dots \widehat{x}_n) \mapsto \sum_{i=0}^{n} (\widehat{x}_0 \dots \widehat{x}_i) \otimes (\widehat{x}_i \dots \widehat{x}_n)$$

which is the composite of an (R,X)-module chain equivalence

$$\Delta(X';R) \simeq_{(R,X)} \Delta(X';R) \otimes_{(R,X)} \Delta(X';R)$$

and the inclusion

$$\Delta(X';R) \otimes_{(R,X)} \Delta(X';R) \subseteq \Delta(X';R) \otimes_R \Delta(X';R) \;.$$

(iv) *The homology classes* $[X] \in H_n(X;R)$ *are in one-one correspondence with the chain homotopy classes of* (R,X)*-module chain maps*

$$[X] \cap - : \Delta(X;R)^{n-*} \ \to\ \Delta(X';R) \ ,$$

with

$$\begin{aligned} H_0(\mathrm{Hom}_{(R,X)}(\Delta(X;R)^{n-*},\Delta(X';R))) \ &=\ H_n(\Delta(X';R)\otimes_{(R,X)}\Delta(X';R)) \\ &=\ H_n(X;R) \ . \end{aligned}$$

Remark 5.4 An X-controlled simplicial complex M is an example of a *CW* complex with a block system κ in the sense of Ranicki and Yamasaki [35]. The product $\Delta(M)\otimes_{(\mathbb{Z},X)}\Delta(M)$ is chain equivalent to the chain complex $D^{\kappa}(\Delta(M))$ of [35].

6. Homology manifolds and Poincaré complexes

Definition 6.1 An *n-dimensional R-homology manifold* is a finite simplicial complex M such that

$$H_*(M,M\backslash\widehat{\sigma};R) \ =\ \begin{cases} R & \text{if } * = n \\ 0 & \text{otherwise} \end{cases} \quad (\sigma \in M) \ .$$

Definition 6.2 An *n-dimensional R-homology Poincaré complex* is a finite simplicial complex M with a homology class $[M] \in H_n(M;R)$ such that the cap products are R-module isomorphisms

$$[M] \cap - : H^{n-*}(M;R) \cong H_*(M;R) \ .$$

Similarly for an *n-dimensional R-homology Poincaré pair* $(M,\partial M)$, with $[M] \in H_n(M,\partial M;R)$ and

$$[M] \cap - : H^{n-*}(M,\partial M;R) \cong H_*(M;R) \ .$$

Proposition 6.3 A finite simplicial complex M is an n-dimensional R-homology manifold with fundamental class $[M] \in H_n(M;R)$ if and only if each $(D(\sigma,M)$, $\partial D(\sigma,M))$ $(\sigma \in M)$ is an $(n-|\sigma|)$-dimensional R-homology Poincaré pair

$$H^{n-|\sigma|-*}(D(\sigma,M),\partial D(\sigma,M);R) \cong H_*(D(\sigma,M);R)$$

with fundamental class $[D(\sigma,M),\partial D(\sigma,M)] \in H_{n-|\sigma|}(D(\sigma,M),\partial D(\sigma,M);R)$ the image of $[M]$ under the composition of $|\sigma|$ codimension 1 boundary maps.

A $\mathbb{Z}$-homology manifold will just be called a homology manifold, and similarly for Poincaré complexes and pairs.

Definition 6.4 An *n-dimensional X-controlled R-homology Poincaré complex M* is an X-controlled simplicial complex with a homology class $[M] \in H_n(M;R)$ such that the cap product

$$[M] \cap - : \Delta(M;R)^{n-*} \to \Delta(M;R)$$

is an (R,X)-module chain equivalence.

Remark 6.5 An X-controlled simplicial complex M is an n-dimensional X-controlled R-homology Poincaré complex if and only if each

$$p_M^{-1}(D(\sigma,X), \partial D(\sigma,X)) \subseteq M \quad (\sigma \in X)$$

is an $(n-|\sigma|)$-dimensional R-homology Poincaré pair. In terms of the polyhedra $|M|$, $|X|$ this condition can be expressed as follows: for every $x \in |X|$ the inverse image $p_M^{-1}(x) \subseteq |M|$ has a closed regular neighbourhood $(U, \partial U)$ which is an n-dimensional R-homology Poincaré pair.

By analogy with 3.5:

Proposition 6.6 (i) *If $X = \{\text{pt.}\}$ an n-dimensional X-controlled R-homology Poincaré complex M is the same as an n-dimensional R-homology Poincaré complex.*
(ii) *If $X = M$ an n-dimensional X-controlled R-homology Poincaré complex M is the same as an n-dimensional R-homology manifold.*

Theorem 6.7 (Poincaré duality) *An n-dimensional R-homology manifold M is an n-dimensional X-controlled R-homology Poincaré complex, with an (R,X)-module chain equivalence*

$$\Delta(M;R)^{n-*} \simeq \Delta(M;R)$$

with respect to any control map $p_M : M \to X'$.
Proof An (R,M)-module chain equivalence

$$[M] \cap - : \Delta(M;R)^{n-*} \to \Delta(M;R)$$

can be regarded as an (R,X)-module chain equivalence, for any control map $p_M : M \to X'$. □

Corollary 6.8 (Poincaré-Lefschetz duality) *An n-dimensional R-homology manifold with boundary $(M, \partial M)$ is an n-dimensional X-controlled R-homology Poincaré pair, with an (R,X)-module chain equivalence*

$$\Delta(M;R)^{n-*} \simeq \Delta(M, \partial M;R)$$

with respect to any control map $p_M : M \to X'$.

Corollary 6.9 (Lefschetz duality) *If M is an n-dimensional R-homology manifold and $L \subseteq M$ is any subcomplex, there is defined an (R,X)-module chain equivalence*

$$\Delta(M, M\backslash L;R)^{n-*} \simeq \Delta(L;R)$$

with respect to any control map $p_M : M \to X'$. *Similarly for an* (R,X)*-module chain equivalence*

$$\Delta(M,L;R)^{n-*} \simeq \Delta(M\backslash L;R) \ .$$

Proof Let $(U,\partial U)$ be a closed regular neighbourhood of L in M, an n-dimensional R-homology manifold with boundary such that the inclusion $L \subset U$ is a homotopy equivalence. There are defined (R,X)-module chain equivalences

$$\begin{aligned}\Delta(M,M\backslash L;R)^{n-*} &\simeq \Delta(M,\mathrm{cl.}(M\backslash U);R)^{n-*} \text{ (homotopy invariance)}\\ &\simeq \Delta(U,\partial U;R)^{n-*} \text{ (excision)}\\ &\simeq \Delta(U;R) \text{ (Poincaré-Lefschetz duality)}\\ &\simeq \Delta(L;R) \text{ (homotopy invariance)} \ .\end{aligned}$$

□

Definition 6.10 Let M be an X-controlled simplicial complex, with a homology class $[M] \in H_n(M;R)$. The *X-controlled peripheral chain complex* of M is the algebraic mapping cone

$$C \ = \ \mathcal{C}([M]\cap - : \Delta(M;R)^{n-*} \to \Delta(M';R))_{*+1}$$

(with a dimension shift), a finite chain complex in $\mathbb{A}(R,X)$.

Proposition 6.11 The following conditions on an X-controlled simplicial complex M with a homology class $[M] \in H_n(M;R)$ *and peripheral chain complex C are equivalent:*
(i) *M is an n-dimensional X-controlled R-homology Poincaré complex,*
(ii) *C is chain contractible in* $\mathbb{A}(R,X)$*,*
(iii) $H_{n-1}(C \otimes_{(R,X)} C) = 0$*,*
(iv) *each* $p^{-1}(D(\sigma,X),\partial D(\sigma,X))$ $(\sigma \in X)$ *is an* $(n-|\sigma|)$*-dimensional R-homology Poincaré pair.*
Proof (i) $\Longleftrightarrow$ (ii) The chain map $[M]\cap - : \Delta(M;R)^{n-*} \to \Delta(M';R)$ is a chain equivalence in $\mathbb{A}(R,X)$ if and only if the algebraic mapping cone is chain contractible in $\mathbb{A}(R,X)$.
(ii) $\Longleftrightarrow$ (iii) The (R,X)-module chain map

$$\alpha \ = \ [M]\cap - : \Delta(M;R)^{n-*} \to \Delta(M';R)$$

is chain homotopic to its chain dual, with a chain homotopy

$$\beta : \alpha \simeq T\alpha : \Delta(M;R)^{n-*} \to \Delta(M';R) \ .$$

Define a chain equivalence in $\mathbb{A}(R,X)$

$$\phi_X : C^{n-1-*} \to C \ = \ \mathcal{C}(\alpha)_{*+1}$$

by

$$\phi_X \ = \ \begin{pmatrix}\beta & 1\\ 1 & 0\end{pmatrix} :$$
$$C^{n-1-r} = \Delta(M;R)^{n-r} \oplus \Delta(M';R)_{r+1} \to C_r = \Delta(M';R)_{r+1} \oplus \Delta(M;R)^{n-r} \ .$$

(See §9 for a more detailed discussion of the quadratic Poincaré structure on C). The abelian group

$$\begin{aligned} H_{n-1}(C \otimes_{(R,X)} C) &= H_0(\mathrm{Hom}_{(R,X)}(C^{n-1-*}, C)) \\ &= H_0(\mathrm{Hom}_{(R,X)}(C, C)) \end{aligned}$$

consists of the chain homotopy classes of chain maps $C \to C$. This group is 0 if and only if C is chain contractible.
(ii) $\Longleftrightarrow$ (iv) By 2.2 C is chain contractible if and only if each component R-module chain complexes $C(\sigma)$ ($\sigma \in X$) is chain contractible. Now

$$\begin{aligned} &C(\sigma) \simeq_R \mathcal{C}([p^{-1}D(\sigma, X)] \cap - : \\ &\qquad \Delta(p^{-1}(D(\sigma, X), \partial D(\sigma, X)); R)^{n-|\sigma|-*} \to \Delta(p^{-1}D(\sigma, X); R))_{*+1} \ , \end{aligned}$$

so that $C(\sigma) \simeq_R 0$ if and only if $p^{-1}(D(\sigma, X), \partial D(\sigma, X))$ ($\sigma \in X$) is an $(n-|\sigma|)$-dimensional R-homology Poincaré pair. □

Example 6.12 Let $X = \{\mathrm{pt.}\}$. The following conditions on a simplicial complex M with a homology class $[M] \in H_n(M; R)$ and peripheral R-module chain complex C are equivalent:
(i) M is an n-dimensional R-homology Poincaré complex with fundamental class $[M]$,
(ii) $H_*(C) = 0$,
(iii) $H_{n-1}(C \otimes_R C) = 0$.

In the following result $X = M$.

Theorem 6.13 The following conditions on an n-dimensional R-homology Poincaré complex X are equivalent:
(i) *X is an n-dimensional R-homology manifold,*
(ii) *the peripheral chain complex*

$$C \ = \ \mathcal{C}([X] \cap - : \Delta(X; R)^{n-*} \to \Delta(X'; R))_{*+1}$$

is (R, X)-module chain contractible,
(iii) $H_{n-1}(C \otimes_{(R,X)} C) = 0$,
(iv) *the cohomology class $V \in H^n(X \times X; R)$ Poincaré dual to the homology class $\Delta_*[X] \in H_n(X \times X; R)$ has image $0 \in H^n(X \times X \backslash \Delta_X; R)$,*
(v) *the fundamental class $[X] \in H_n(X; R)$ is such that*

$$[X] \in \mathrm{im}(H^n(X \times X, X \times X \backslash \Delta_X; R) \to H_n(X; R)) \ ,$$

(vi) *a particular R-module morphism*

$$H^n(X \times X \backslash \Delta_X; R) \to H_n(X \times X, \Delta_X; R)$$

(specified in the proof) is an isomorphism, namely the Lefschetz duality isomorphism.

Proof (i) $\iff$ (ii) $\iff$ (iii) This is a special case of 6.11.
(i) $\iff$ (iv) There is defined an exact sequence

$$H^n(X\times X, X\times X\backslash\Delta_X;R)\to H^n(X\times X;R)\to H^n(X\times X\backslash\Delta_X;R)\ .$$

Thus V has image $0\in H^n(X\times X\backslash\Delta_X;R)$ if and only if there exists an element

$$U\in H^n(X\times X, X\times X\backslash\Delta_X;R)$$

with image V. Now U is a chain homotopy class of (R,X)-module chain maps $\Delta(X';R)\to\Delta(X;R)^{n-*}$, since

$$\begin{aligned}H^n(X\times X, X\times X\backslash\Delta_X;R) &= H_n(\Delta(X;R)^{-*}\otimes_{(R,X)}\Delta(X;R)^{-*})\\ &= H_0(\mathrm{Hom}_{(R,X)}(\Delta(X';R),\Delta(X;R)^{n-*}))\ .\end{aligned}$$

U is a chain homotopy inverse of

$$\phi=[X]\cap-:\Delta(X;R)^{n-*}\to\Delta(X';R)$$

with

$$\begin{aligned}\phi U &= 1\in H_0(\mathrm{Hom}_{(R,X)}(\Delta(X';R),\Delta(X';R)))\ =\ H^0(X;R)\ ,\\ \phi &= T\phi\in H_0(\mathrm{Hom}_{(R,X)}(\Delta(X;R)^{n-*},\Delta(X';R)))\ ,\\ (TU)\phi &= (TU)(T\phi)\ =\ T(\phi U)\ =\ 1\\ &\quad\in H_0(\mathrm{Hom}_{(R,X)}(\Delta(X';R)^{n-*},\Delta(X;R)^{n-*}))\ .\end{aligned}$$

(iv) $\iff$ (v) $\iff$ (vi) Immediate from the commutative braid of exact sequences

$$\begin{tikzcd}[column sep=small]
H^n(X\times X,X\times X\backslash\Delta_X;R) \arrow[dr] \arrow[rr, bend left] & & H^n(X\times X;R) \arrow[dr] \arrow[rr, bend left] & & H_n(X\times X,\Delta_X;R) \\
& H_n(X;R) \arrow[ur, "\Delta_*"] \arrow[dr] & & H^n(X\times X\backslash\Delta_X;R) \arrow[ur] \arrow[dr] & \\
H_{n+1}(X\times X,\Delta_X;R) \arrow[ur, "0"] \arrow[rr, bend right, "0"'] & & H_{n-1}(C\otimes_{(R,X)}C) \arrow[ur] \arrow[rr, bend right] & & H^{n+1}(X\times X,X\times X\backslash\Delta_X;R)
\end{tikzcd}$$

on noting that $X\times X$ is a $2n$-dimensional R-homology Poincaré complex with isomorphisms

$$[X\times X]\cap-:H^n(X\times X;R)\cong H_n(X\times X;R)$$

and that the diagonal map

$$\Delta:X\to X\times X\ ;\ x\mapsto(x,x)$$

is split by the projection

$$p : X \times X \to X \ ; \ (x, y) \mapsto x \ ,$$

so that

$$H_*(X \times X; R) \ = \ H_*(X; R) \oplus H_*(X \times X, \Delta_X; R) \ .$$

The classes

$$V \in H^n(X \times X, X \times X \backslash \Delta_X; R) \ \ , \ \ \phi_X \in H_{n-1}(C \otimes_{(R,X)} C)$$

(with ϕ_X as in the proof of 6.11) are both images of the fundamental class $[X] \in H_n(X; R)$, so that they have the same image in $H^n(X \times X \backslash \Delta_X; R)$. □

Remark 6.14 The equivalence (i) $\Longleftrightarrow$ (iv) in 6.13 in the case $R = \mathbb{Z}$ is a slight generalization of the corresponding results of McCrory [17, Theorem 1] and Ranicki [29, 8.5] for n-circuits and n-dimensional pseudomanifolds respectively.

Remark 6.15 A Poincaré complex X is a homology manifold precisely when the dihomology spectral sequence of Zeeman [45] collapses. See McCrory [18] for a geometric interpretation in terms of moving cocycles in $X \times X$ off the diagonal.

There is also a version of 6.13 for Poincaré pairs with manifold boundary. Here is a special case:

Proposition 6.16 An n-dimensional R-homology Poincaré pair $(X, \partial X)$ with R-homology manifold boundary is an n-dimensional R-homology manifold with boundary if and only if the cohomology class $V \in H^n(X \times X, X \times \partial X; R)$ Poincaré-Lefschetz dual to the homology class $\Delta_*[X] \in H_n(X \times X, \partial X \times X; R)$ (with $[X] \in H_n(X, \partial X; R)$) is the image of a class

$$U \in H^n(X \times X, X \times \partial X \cup X \times X \backslash \Delta_X; R) \ .$$

Remark 6.17 In general, a singularity does not arise as a non-manifold point of a Poincaré complex, so 6.13 cannot be applied directly to obtain a homological invariant of the singularity. However, for an isolated singular point of a complex hypersurface it is possible to apply 6.16 to a related Poincaré pair with manifold boundary. Given a polynomial function $f : \mathbb{C}^{n+1} \to \mathbb{C}$ with an isolated critical point $z_0 \in V = f^{-1}(0)$ Milnor [20] relates the singularity of f at z_0 to the properties of the fibred knot

$$k : V \cap S_\epsilon \ = \ S^{2n-1} \subset S_\epsilon \ = \ S^{2n+1}$$

defined by intersecting V with

$$S_\epsilon \ = \ \{z \in \mathbb{C}^{n+1} \,|\, \|z - z_0\| = \epsilon\}$$

for a sufficiently small ϵ. (Only PL structures are considered here – the differentiable structure on $V \cap S_\epsilon$ could of course be exotic). In §14 below there will be associated to any fibred knot $k : S^{2n-1} \subset S^{2n+1}$ a $(2n+2)$-dimensional homology Poincaré pair $(X, \partial X)$ with manifold boundary, which is a homology manifold with boundary if k is unknotted; the obstruction to $(X, \partial X)$ being a homology manifold with boundary is related to homological invariants of k, and hence to the nature of the singularity.

7. Degree 1 maps and homology equivalences

This section investigates the extent to which a degree 1 map of n-dimensional homology manifolds has acyclic point inverses. It is shown that this is the case if and only if the n-dimensional homology of the double point set relative to the diagonal is zero.

Definition 7.1 The *double point set* of a map $f : M \to N$ is the pullback (5.1)

$$\begin{aligned} M \times_N M &= (f \times f)^{-1}(\Delta_N) \\ &= \{(x,y) \in M \times M \,|\, f(x) = f(y) \in N\} \ . \end{aligned}$$

If f is a simplicial map then $M \times_N M$ is an N-controlled simplicial complex.

Given a map $f : M \to N$ define the maps

$$\begin{aligned} &i : M \to M \times_N M \ ; \ x \mapsto (x,x) \ , \\ &j : M \times_N M \to N \ ; \ (x,y) \mapsto f(x) = f(y) \ , \\ &k : M \times_N M \to M \ ; \ (x,y) \mapsto x \ . \end{aligned}$$

There is defined a commutative diagram

$$\begin{tikzcd} & M \times M \arrow[r, "f\times f"] & N \times N \\ & M \times_N M \arrow[u] \arrow[r, "j"] & N \arrow[u, "\Delta_N"'] \\ M \arrow[uur, "\Delta_M"] \arrow[ur, "i"] \arrow[urr, "f"'] & & \end{tikzcd}$$

It follows from $ki = 1 : M \to M$ that

$$H_*(M \times_N M) \ = \ H_*(M) \oplus H_*(M \times_N M, \Delta_M) \ .$$

Definition 7.2 Let $f : M \to N$ be a map of X-controlled R-homology Poincaré complexes, with $\dim(M) = m$, $\dim(N) = n$.
(i) The *Umkehr* of f is the (R,X)-module chain map

$$f^! : \Delta(N;R) \simeq \Delta(N;R)^{n-*} \xrightarrow{\ f^*\ } \Delta(M;R)^{n-*} \simeq \Delta(M;R)_{*+m-n} \ .$$

(ii) f has *degree 1* if $m = n$ and

$$f_*[M] \ = \ [N] \in H_n(N;R) \ .$$

Proposition 7.3 (i) *If $f : M \to N$ is a degree 1 map of n-dimensional X-controlled R-homology Poincaré complexes the Umkehr (R,X)-module chain map $f^! : \Delta(N;R) \to \Delta(M;R)$ is such that*

$$ff^! \simeq 1 : \Delta(N;R) \to \Delta(N;R)$$

and there exists an (R,X)-module chain equivalence

$$\Delta(M;R)\simeq_{(R,X)}\Delta(N;R)\oplus\Delta(f^!)\ .$$

(ii) *If $f:M\to N$ is a degree 1 map of n-dimensional R-homology manifolds then*

$$H_n(\Delta(f^!)\otimes_{(R,N)}\Delta(f^!))\ =\ H_n(M\times_N M,\Delta_M;R)\ .$$

Proof (i) Immediate from $f_*[M]=[N]\in H_n(N;R)$ and the naturality properties of the cap product.
(ii) Apply $\Delta(M)\otimes_{(\mathbb{Z},N)}-$ to the $(\mathbb{Z},N)$-module chain equivalence given by (i)

$$\Delta(M)\simeq_{(\mathbb{Z},N)}\Delta(N)\oplus\Delta(f^!)\ ,$$

to obtain

$$\begin{aligned}&\Delta(M)\otimes_{(\mathbb{Z},N)}\Delta(M)\\ &\quad\simeq_{(\mathbb{Z},N)}(\Delta(M)\otimes_{(\mathbb{Z},N)}\Delta(N))\oplus(\Delta(M)\otimes_{(\mathbb{Z},N)}\Delta(f^!))\\ &\quad\simeq_{(\mathbb{Z},N)}(\Delta(M)\otimes_{(\mathbb{Z},N)}\Delta(N))\oplus(\Delta(N)\otimes_{(\mathbb{Z},N)}\Delta(f^!))\oplus(\Delta(f^!)\otimes_{(\mathbb{Z},N)}\Delta(f^!))\\ &\quad\simeq_{(\mathbb{Z},N)}\Delta(M)\oplus\Delta(f^!)\oplus(\Delta(f^!)\otimes_{(\mathbb{Z},N)}\Delta(f^!))\ .\end{aligned}$$

Since $H_n(f^!)=0$, it follows that

$$\begin{aligned}H_n(M\times_N M)\ &=\ H_n(\Delta(M)\otimes_{(\mathbb{Z},N)}\Delta(M))\\ &=\ H_n(M)\oplus H_n(f^!)\oplus H_n(\Delta(f^!)\otimes_{(\mathbb{Z},N)}\Delta(f^!))\\ &=\ H_n(M)\oplus H_n(\Delta(f^!)\otimes_{(\mathbb{Z},N)}\Delta(f^!))\ .\end{aligned}$$

□

Theorem 7.4 The following conditions on a degree 1 map $f:M\to N$ of n-dimensional X-controlled R-homology Poincaré complexes are equivalent:

(i) *f is an X-controlled R-homology equivalence (3.3),*

(ii) *$f:\Delta(M;R)\to\Delta(N;R)$ is an (R,X)-module chain equivalence,*

(iii) *there exists an (R,X)-module chain homotopy*

$$f^!f\simeq 1:\Delta(M;R)\to\Delta(M;R)\ ,$$

(iv) *$\Delta_*[M]=(f^!\otimes f^!)\Delta_*[N]\in H_n(M\times_X M;R)$,*

(v) *$(f^!\otimes f^!)\Delta_*[N]=0\in H_n(M\times_X M,\Delta_M;R)$,*

(vi) *$(f\times f)_*:H_n(M\times_X M;R)\cong H_n(N\times_X N;R)$.*

Proof (i) $\Longleftrightarrow$ (ii) This is a special case of 3.4.
(ii) $\Longleftrightarrow$ (iii) Immediate from 7.3.
(iii) $\Longleftrightarrow$ (iv) Immediate from the identifications

$$\begin{aligned} 1 \;=\; \Delta_*[M] \;\;,\;\; f^!f \;=\; (f^! \otimes f^!)\Delta_*[N] & \\ \in H_0(\mathrm{Hom}_{(R,X)}(\Delta(M;R),\Delta(M;R))) \;=\; H_n(M \times_X M;R)\ . & \end{aligned}$$

(iv) $\Longleftrightarrow$ (v) Immediate from the identity

$$\begin{aligned} (f^! \otimes f^!)\Delta_*[N] \;=\; ([M],(f^! \otimes f^!)\Delta_*[N] - \Delta_*[M]) & \\ \in H_n(M \times_X M;R) \;=\; H_n(M;R) \oplus H_n(M \times_X M, \Delta_M;R)\ . & \end{aligned}$$

(ii) $\Longrightarrow$ (vi) If $f : \Delta(M;R) \to \Delta(N;R)$ is an (R,X)-module chain equivalence then so is

$$f \otimes f : \Delta(M;R) \otimes_{(R,X)} \Delta(M;R) \to \Delta(N;R) \otimes_{(R,X)} \Delta(N;R)\ .$$

(vi) $\Longrightarrow$ (iv) It follows from $ff^! \simeq 1$ and

$$(f \otimes f)_*\Delta_*[M] \;=\; \Delta_*[N] \in H_n(N \times_X N;R)$$

that

$$\begin{aligned} & \Delta_*[M] - (f^! \otimes f^!)\Delta_*[N] \\ & \in \ker((f \times f)_* : H_n(M \times_X M;R) \to H_n(N \times_X N;R)) \;=\; \{0\}\ . \end{aligned}$$

□

Corollary 7.5 The following conditions on a degree 1 map $f : M \to N$ of n-dimensional homology manifolds are equivalent:
(i) *f has acyclic point inverses,*
(ii) $H_n(M \times_N M, \Delta_M) = 0$,
(iii) $H_n(\Delta(f^!) \otimes_{(\mathbb{Z},N)} \Delta(f^!)) = 0$.
Proof (i) $\Longleftrightarrow$ (ii) Apply 7.3 with $R = \mathbb{Z}$, $X = N$, so that

$$\begin{aligned} & M \times_X M \;=\; M \times_N M \;=\; (f \times f)^{-1}\Delta_N \;\;,\;\; N \times_X N \;=\; N\ , \\ & H_n(M \times_X M) \;=\; H_n(M) \oplus H_n(M \times_N M, \Delta_M)\ . \end{aligned}$$

Since $f_* : H_n(M) \cong H_n(N)$, condition 7.4 (vi)

$$(f \times f)_* : H_n(M \times_N M) \cong H_n(N \times_N N)$$

for f to be a $(\mathbb{Z},N)$-homology equivalence is equivalent to

$$H_n(M \times_N M, \Delta_M) \;=\; 0\ .$$

As in 3.5 (ii) a map f is a $(\mathbb{Z},N)$-homology equivalence if and only if it has acyclic point inverses.
(ii) $\Longleftrightarrow$ (iii) By 7.3 (ii) $H_n(\Delta(f^!) \otimes_{(\mathbb{Z},N)} \Delta(f^!)) = H_n(M \times_N M, \Delta_M)$. □

Remark 7.6 (i) A map $f : M \to N$ is injective if and only if

$$M \times_N M \;=\; \Delta_M \; .$$

The condition of 7.5 (ii) is automatically satisfied for injective f.
(ii) A degree 1 map $f : M \to N$ of n-dimensional R-homology manifolds is surjective by the following argument, which does not require M, N to be polyhedra. If $x \in N \backslash f(M)$ then

$$H_n(M, M\backslash f^{-1}(x); R) \;=\; 0 \;\;,\;\; H_n(N, N\backslash\{x\}; R) \;=\; R \;,$$

leading to a contradiction in the commutative diagram

$$\begin{array}{ccc}
H_n(M;R) = R & \xrightarrow{\;\cong\;} & H_n(N;R) = R \\
\big\downarrow & & \cong \big\downarrow \\
H_n(M, M\backslash f^{-1}(x); R) = 0 & \xrightarrow{\;f_*\;} & H_n(N, N\backslash\{x\}; R) = R
\end{array}$$

(assuming M, N are connected).

Corollary 7.7 (i) *A map $f : M \to N$ of n-dimensional R-homology Poincaré complexes is an R-homology equivalence if and only if it is degree 1 and*

$$\Delta_*[M] \;=\; (f^! \otimes f^!)\Delta_*[N] \in H_n(M \times M; R) \; .$$

(ii) *A map $f : M \to N$ of n-dimensional R-homology manifolds has R-acyclic point inverses if and only if it is degree 1 and*

$$\Delta_*[M] \;=\; (f^! \otimes f^!)\Delta_*[N] \in H_n(M \times_N M; R) \; .$$

Proof (i) Apply 7.4 with $X = \{\text{pt.}\}$.
(ii) Apply 7.4 with $X = N$. □

Definition 7.8 Given a map $f : M \to N$ of R-homology manifolds with $\dim(M) = m$, $\dim(N) = n$ let the *Umkehr* of the map

$$j : M \times_N M \to N \; ; \; (x, y) \mapsto f(x) = f(y)$$

be the (R, N)-module chain map

$$j^! : \Delta(N; R) \to \Delta(M \times_N M; R)_{*+2m-2n}$$

given by the composite

$$\begin{aligned}
j^! : \Delta(N;R) &\simeq_{(R,N)} \Delta(N \times N, N \times N\backslash \Delta_N; R)^{2n-*} \\
&\xrightarrow{(f\times f)^*} \Delta(M \times M, M \times M\backslash M \times_N M; R)^{2n-*} \\
&\simeq_{(R,N)} \Delta(M \times_N M; R)_{*+2m-2n} \; .
\end{aligned}$$

Proposition 7.9 The following conditions on a degree 1 map $f : M \to N$ of n-dimensional R-homology manifolds are equivalent:
(i) *f has R-acyclic point inverses,*
(ii) *there exists an (R,N)-module chain homotopy*

$$i_*f^! \simeq j^! : \Delta(N;R) \to \Delta(M \times_N M;R) \ ,$$

(iii) *there exists an (R,N)-module chain map $g : \Delta(N) \to \Delta(M)$ with an (R,N)-module chain homotopy*

$$i_*g \simeq j^! : \Delta(N;R) \to \Delta(M \times_N M;R) \ .$$

Proof (i) $\Longleftrightarrow$ (ii) Identify

$$\begin{aligned} i_*f^! \ &= \ \Delta_*[M] \ \ , \ \ j^! \ = \ (f^! \otimes f^!)\Delta_*[N] \\ &\in H_0(\mathrm{Hom}_{(R,N)}(\Delta(N;R), \Delta(M \times_N M;R))) \ = \ H_n(M \times_N M;R) \end{aligned}$$

and apply the equivalence (i) $\Longleftrightarrow$ (iv) of 7.4, with $X = N$.
(ii) $\Longrightarrow$ (iii) Take $g = f^!$.
(iii) $\Longrightarrow$ (i) It follows from the exact sequence

$$\begin{aligned} &H_0(\mathrm{Hom}_{(R,N)}(\Delta(N;R), \Delta(M;R))) \\ &\qquad \xrightarrow{i_*} H_0(\mathrm{Hom}_{(R,N)}(\Delta(N;R), \Delta(M \times_N M;R))) \\ &\qquad \longrightarrow H_0(\mathrm{Hom}_{(R,N)}(\Delta(N;R), \Delta(M \times_N M, \Delta_M;R))) \end{aligned}$$

that such a g exists if and only if the (R,N)-module chain homotopy class

$$j^! \in H_0(\mathrm{Hom}_{(R,N)}(\Delta(N;R), \Delta(M \times_N M;R)))$$

has 0 image in

$$H_0(\mathrm{Hom}_{(R,N)}(\Delta(N;R), \Delta(M \times_N M, \Delta_M;R))) \ = \ H_n(M \times_N M, \Delta_M;R) \ .$$

But this image is precisely the element $(f^! \otimes f^!)\Delta_*[N] \in H_n(M \times_N M, \Delta_M;R)$ of 7.4 (v) whose vanishing is (necessary and) sufficient for f to have R-acyclic point inverses. □

8. Bundles

The results of §§6,7 will now be interpreted from the bundle point of view, aftre a brief review of the various bundle theories involved.

Oriented spherical fibrations η over a space X

$$(D^k, S^{k-1}) \to (E(\eta), S(\eta)) \to X$$

are classified up to oriented fibre homotopy equivalence by the homotopy classes of maps $\eta : X \to BG(k)$ to a classifying space $BG(k)$. Every such fibration has a *Thom space*

$$T(\eta) \ = \ E(\eta)/S(\eta)$$

and a *Thom class*

$$U_\eta \in \widetilde{H}^k(T(\eta)) \ .$$

See Rourke and Sanderson [38] for the theory of (oriented) PL k-block bundles, with a classifying space $BS\widetilde{PL}(k)$. A codimension k embedding $M^n \subset N^{n+k}$ of PL manifolds has a normal PL k-block bundle $\nu_{M\subset N} : M \to BPL(k)$.

See Martin and Maunder [15] for the theory of homology cobordism bundles, with a classifying space $BSH(k)$ and forgetful maps

$$BS\widetilde{PL}(k) \to BSH(k) \ \ , \ \ BSH(k) \to BSG(k) \ .$$

A codimension k embedding $M^n \subset N^{n+k}$ of homology manifolds (i.e. a PL map which is an injection) has a normal homology cobordism S^{k-1}-bundle $\nu_{M\subset N} : M \to BSH(k)$.

See Rourke and Sanderson [39] for the theory of (oriented) topological k-block bundles, with a classifying space $BS\widetilde{TOP}(k)$ and forgetful maps

$$BS\widetilde{PL}(k) \to BS\widetilde{TOP}(k) \ \ , \ \ BS\widetilde{TOP}(k) \to BSG(k) \ .$$

Galewski and Stern [7] proved that every homology cobordism S^{k-1}-bundle has a canonical lift to a topological k-block bundle, so that there is defined a commutative diagram of classifying spaces and forgetful maps

$$\begin{tikzcd} BS\widetilde{PL}(k) \arrow[r] \arrow[d] & BS\widetilde{TOP}(k) \arrow[d] \\ BSH(k) \arrow[r] \arrow[ur] & BSG(k) \ . \end{tikzcd}$$

The diagonal embedding of an n-dimensional homology manifold M

$$\Delta : M \to M \times M \ ; \ x \mapsto (x,x)$$

has a normal homology cobordism S^{n-1}-bundle, the *tangent homology cobordism S^{n-1}-bundle* ([15, 5.3])

$$\tau_M \ = \ \nu_\Delta : M \to BSH(n) \ ,$$

and hence a *tangent topological n-block bundle* $\tau_M : M \to BS\widetilde{TOP}(n)$. The Euler class of τ_M may be identified with the Euler characteristic of M, as follows.

The *Euler characteristic* of a finite simplicial complex X is

$$\chi(X) \;=\; \sum_{r=0}^{\infty}(-)^r \dim_{\mathbb{R}} H_r(X;\mathbb{R}) \in \mathbb{Z} \;.$$

Proposition 8.1 (i) *For a connected n-dimensional Poincaré complex X*

$$\chi(X) \;=\; \Delta^*(V) \in H^n(X) \;=\; \mathbb{Z}$$

with $V \in H^n(X\times X)$ the Poincaré dual of $\Delta_*[X] \in H_n(X\times X)$.
(ii) *The obstruction to a degree 1 map $f: M \to N$ of connected n-dimensional Poincaré complexes being a homology equivalence (7.7 (i))*

$$\Delta_*[M] - (f^! \otimes f^!)\Delta_*[N] \in H_n(M\times M)$$

has image $\chi(M) - \chi(N) \in \mathbb{Z}$ under the composite

$$H_n(M\times M) \cong H^n(M\times M) \xrightarrow{\Delta^*} H^n(M) \;=\; \mathbb{Z} \;.$$

Proof (i) As for smooth manifolds (Milnor and Stasheff [21, 11.13]).
(ii) Immediate from (i). □

It is well known that $\chi(M) = \chi(\tau_M)$ for a smooth manifold M ([21, 11.13]). More generally :

Proposition 8.2 The Euler characteristic of a connected n-dimensional homology manifold M is the Euler class of the tangent n-block bundle τ_M

$$\chi(M) \;=\; \chi(\tau_M) \in H^n(M) \;=\; \mathbb{Z} \;.$$

Proof The homology tangent bundle of M (Spanier [40, p.294]) is the homology fibration

$$(M, M\backslash\{*\}) \to (M\times M, M\times M\backslash\Delta_M) \to M$$

with

$$M \to M\times M \;;\; x \mapsto (*, x) \;,$$
$$M\times M \to M \;;\; (x,y) \mapsto x \;.$$

The tangent topological n-block bundle of M

$$(D^n, S^{n-1}) \to (E(\tau_M), S(\tau_M)) \to M$$

is related to the homology tangent bundle by a homotopy pushout diagram

$$\begin{array}{ccc} S(\tau_M) & \longrightarrow & M\times M\backslash\Delta_M \\ \big\downarrow & & \big\downarrow \\ E(\tau_M)\simeq M & \xrightarrow{\;\Delta\;} & M\times M \;. \end{array}$$

The Thom space, Thom class and Euler class of τ_M are such that

$$\begin{aligned}
&T(\tau_M) \ = \ E(\tau_M)/S(\tau_M) \ = \ (M\times M)/(M\times M\backslash\Delta_M)\ ,\\
&U_M\in \widetilde{H}^n(T(\tau_M)) = H^n(M\times M, M\times M\backslash\Delta_M)\ ,\\
&e(\tau_M) \ = \ z^*(U_M)\in H^n(M)\ ,
\end{aligned}$$

with $z: M\to T(\tau_M)$ the zero section. Furthermore, there is defined a commutative diagram

$$\begin{tikzcd}
H^n(M\times M, M\times M\backslash\Delta_M) \arrow[r, "i^*"] \arrow[d, "\cong"'] & H^n(M\times M) \arrow[d, "\Delta^*"] \\
\widetilde{H}^n(T(\tau_M)) \arrow[r, "z^*"] & H^n(M)
\end{tikzcd}$$

with $i: M\times M\to (M\times M, M\times M\backslash\Delta_M)$ the natural map. As before, let $V\in H^n(M\times M)$ be the Poincaré dual of $\Delta_*[M]\in H_n(M\times M)$. The Thom class $U_M\in \widetilde{H}^n(T(\tau_M))$ has image

$$i^*(U_M) \ = \ V\in H^n(M\times M)\ ,$$

and

$$e(\tau_M) \ = \ z^*(U_M) \ = \ \Delta^*(i^*(U_M)) \ = \ \Delta^*(V) \ = \ \chi(M)\in H^n(M) \ = \ \mathbb{Z}\ .$$

$\square$

Remark 8.3 Theorem 6.13 can be regarded as a converse of 8.2 :
A connected n-dimensional Poincaré complex X is an n-dimensional homology manifold if and only if the Poincaré dual $V\in H^n(X\times X)$ of $\Delta_*[X]\in H_n(X\times X)$ is the image of a Thom class $U\in \widetilde{H}^n(T(\tau_X))$, in which case

$$\chi(X) \ = \ e(\tau_X)\in H^n(X) \ = \ \mathbb{Z}\ .$$

McCrory [17] called such U a *geometric Thom class* for X.

Proposition 8.4 A degree 1 map $f: M\to N$ of n-dimensional R-homology manifolds has acyclic point inverses if and only if the Thom classes

$$U_M\in H^n(M\times M, M\times M\backslash\Delta_M;R)\ ,\ U_N\in H^n(N\times N, N\times N\backslash\Delta_N;R)$$

have the same image in $H^n(M\times M, M\times M\backslash M\times_N M;R)$

$$c^*(U_M) \ = \ (f\times f)^*(U_N)\in H^n(M\times M, M\times M\backslash M\times_N M;R)\ ,$$

with $c: (M\times M, M\times M\backslash M\times_N M)\to (M\times M, M\times M\backslash\Delta_M)$ the inclusion of pairs.

Proof This is just the cohomology version of 7.7 (ii), after Lefschetz duality (6.8) identifications

$$\begin{aligned} U_M &= [M] \in H^n(M \times M, M \times M \backslash \Delta_M; R) = H_n(M;R) , \\ U_N &= [N] \in H^n(N \times N, N \times N \backslash \Delta_N; R) = H_n(N;R) , \\ H^n(M \times M, M \times M \backslash M \times_N M; R) &= H_n(M \times_N M; R) , \end{aligned}$$

noting that $M \times M$ and $N \times N$ are $2n$-dimensional R-homology manifolds.
□

Remark 8.5 Suppose that $f : M \to N$ is a degree 1 map of n-dimensional homology manifolds which is covered by a stable map

$$b : \tau_M \oplus \epsilon^\infty \to \tau_N \oplus \epsilon^\infty$$

of the tangent block bundles. (For example, if M, N have trivial tangent block bundles then any map $f : M \to N$ is covered by an unstable map $b : \tau_M \to \tau_N$). In general, the diagram

$$\begin{tikzcd} \widetilde{H}^n(T(\tau_N)) \arrow[rr, "T(b)^*"] \arrow[d, phantom, "\cong"] & & \widetilde{H}^n(T(\tau_M)) \arrow[d, phantom, "\cong"] \\ H^n(N \times N, N \times N \backslash \Delta_N) \arrow[dr, "(f \times f)^*"] & & H^n(M \times M, M \times M \backslash \Delta_M) \arrow[dl, "c^*"'] \\ & H^n(M \times M, M \times M \backslash M \times_N M) & \end{tikzcd}$$

is not commutative, with the obstruction in 8.4 non-zero:

$$\begin{aligned} c^* T(b)^*(U_N) - (f \times f)^*(U_N) &= c^*(U_M) - (f \times f)^*(U_N) \\ &\neq 0 \in H^n(M \times M, M \times M \backslash M \times_N M) . \end{aligned}$$

In §9 below this difference will be expressed in terms of an N-controlled refinement of the (symmetrization of the) quadratic structure used in Ranicki [27] to obtain a chain level expression for the Wall surgery obstruction.

Proposition 8.6 Let $f : M \to N$ be a degree 1 map of n-dimensional R-homology manifolds. If there exists an N-controlled map

$$a : (M \times M, M \times M \backslash \Delta_M) \to (N \times N, N \times N \backslash \Delta_N)$$

such that the diagram

$$\begin{tikzcd}[column sep=tiny] & (M\times M, M\times M\backslash M\times_N M) \arrow[dl, "c"'] \arrow[dr, "f\times f"] & \\ (M\times M, M\times M\backslash \Delta_M) \arrow[rr, "a"] & & (N\times N, N\times N\backslash \Delta_N) \end{tikzcd}$$

is N-controlled homotopy commutative, then

$$(f\times f)^*(U_N) \ = \ c^*(U_M) \in H^n(M\times M, M\times M\backslash M\times_N M;R)$$

and f has acyclic point inverses. Moreover,

$$a^*(U_N) \ = \ U_M \in H^n(M\times M, M\times M\backslash \Delta_M;R)\ .$$

Proof Define the (R,N)-module chain map

$$\begin{aligned} g : \Delta(N;R) \simeq_{(R,N)} & \Delta(N\times N, N\times N\backslash\Delta_N;R)^{2n-*} \\ & \xrightarrow{a^*} \Delta(M\times M, M\times M\backslash\Delta_M;R)^{2n-*} \simeq_{(R,N)} \Delta(M;R) \end{aligned}$$

such that

$$g[N] \ = \ a^*(U_N) \in H_n(M) \ = \ H^n(M\times M, M\times M\backslash\Delta_M)\ .$$

The N-controlled homotopy of pairs

$$ac \simeq f\times f : (M\times M, M\times M\backslash M\times_N M) \to (N\times N, N\times N\backslash\Delta_N)$$

induces an (R,N)-module chain homotopy

$$\begin{aligned} ac \simeq f\times f : \Delta(M\times M, M\times M\backslash M\times_N M;R) & \simeq_{(R,N)} \Delta(M\times_N M;R)^{2n-*} \\ \to \Delta(N\times N, N\times N\backslash\Delta_N;R) & \simeq_{(R,N)} \Delta(N;R)^{2n-*}\ . \end{aligned}$$

The chain dual is an (R,N)-module chain homotopy

$$i_*g \simeq j^! : \Delta(N;R) \to \Delta(M\times_N M;R)\ ,$$

so that

$$i_*g[N] \ = \ j^![N] \ = \ [M\times_N M] \in H_n(M\times_N M;R)\ ,$$

with dual the identity

$$c^*a^*(U_N) \ = \ (f\times f)^*(U_N) \in H^n(M\times M, M\times M\backslash M\times_N M;R)\ ,$$

so that f has R-acyclic point inverses by 8.4, and

$$\begin{aligned} g \simeq f^{-1} \simeq f^! &: \Delta(N;R) \to \Delta(M;R) \ , \\ g[N] &= [M] \in H_n(M;R) \ , \\ a^*(U_N) &= U_M \in H^n(M\times M, M\times M\backslash \Delta_M;R) \ . \end{aligned}$$

□

Remark 8.7 A degree 1 map $f: M \to N$ of n-dimensional homology manifolds which is covered by a map of the tangent n-block bundles $b: \tau_M \to \tau_N$ need not be covered by a map of homology tangent bundles a as in 8.6.

9. The total surgery obstruction

The total surgery obstruction $s(X) \in \mathbb{S}_n(X)$ of Ranicki [29] is defined for a finite simplicial complex X satisfying n-dimensional Poincaré duality with respect to all coefficients – such Poincaré complexes are considered further below. For $n \geq 5$ the total surgery obstruction is $s(X) = 0$ if and only if the polyhedron $|X|$ is homotopy equivalent to a topological manifold (which need not be triangulable). On the other hand, an n-dimensional homology Poincaré complex X is a homology manifold if and only if an obstruction in $H^n(X\times X\backslash \Delta_X)$ (6.13) is 0. The obstruction of 6.13 will now be related to the total surgery obstruction and its $\mathbb{Z}$-homology analogue.

So far, only the homology $H_*(X;R)$ and cohomology $H^*(X;R)$ of a simplicial complex X with coefficients in a commutative ring R have been considered. For non-simply-connected X the homology $H_*(X;\Lambda)$ and cohomology $H^*(X;\Lambda)$ and with coefficients in an $R[\pi_1(X)]$-module Λ will also be considered.

Given a commutative ring R and a group π let the group ring $R[\pi]$ have the involution

$$R[\pi] \to R[\pi] \ ; \ a = \sum_{g\in\pi} n_g g \mapsto \overline{a} = \sum_{g\in\pi} n_g g^{-1} \quad (n_g \in R) \ .$$

Use the involution to convert every left $R[\pi]$-module M into a right $R[\pi]$-module M^t, with the same additive group and

$$M^t \times R[\pi] \to M^t \ ; \ (x,a) \mapsto \overline{a}.x \ .$$

Define an involution (1.2) on the additive category $\mathbb{A}(R[\pi])$ of f.g. free (left) $R[\pi]$-modules

$$* : \mathbb{A}(R[\pi]) \to \mathbb{A}(R[\pi]) \ ; \ A \mapsto A^* = \mathrm{Hom}_{R[\pi]}(A, R[\pi])$$

with

$$R[\pi] \times A^* \to A^* \ ; \ (a,f) \mapsto (x \mapsto f(x).\overline{a}) \ .$$

Definition 9.1 Given a connected simplicial complex X with universal cover $\widetilde{X}$ and an $R[\pi_1(X)]$-module Λ define the Λ-*coefficient homology* and *cohomology* R-modules of X to be

$$\begin{aligned} H_*(X;\Lambda) &= H_*(\Lambda^t \otimes_{R[\pi_1(X)]} \Delta(\widetilde{X};R)) \ , \\ H^*(X;\Lambda) &= H_*(\mathrm{Hom}_{R[\pi_1(X)]}(\Delta(\widetilde{X};R),\Lambda)) \ . \end{aligned}$$

The Λ-coefficient homology and cohomology R-modules are related by a cap product pairing

$$H_n(X;R) \otimes_R H^m(X;\Lambda) \to H_{n-m}(X;\Lambda); x \otimes y \mapsto x \cap y \ .$$

For $\Lambda = R[\pi_1(X)]$ the Λ-coefficient homology and cohomology groups are $R[\pi_1(X)]$-modules

$$\begin{aligned} H_*(X;R[\pi_1(X)]) &= H_*(\Delta(\widetilde{X};R)) = H_*(\widetilde{X};R) \ , \\ H^*(X;R[\pi_1(X)]) &= H_{-*}(\mathrm{Hom}_{R[\pi_1(X)]}(\Delta(\widetilde{X};R),R[\pi_1(X)])) \end{aligned}$$

Definition 9.2 An *n-dimensional universal R-homology Poincaré complex* is a finite connected simplicial complex X with a homology class $[X] \in H_n(X;R)$ such that the cap products are $R[\pi_1(X)]$-module isomorphisms

$$[X] \cap - : H^{n-*}(X;R[\pi_1(X)]) \cong H_*(X;R[\pi_1(X)]) \ .$$

A universal $\mathbb{Z}$-homology Poincaré complex will just be called a universal homology Poincaré complex.

Remark 9.3 (i) A universal homology Poincaré complex is just a Poincaré complex in the sense of Wall [42].
(ii) If X is a universal R-homology Poincaré complex with universal cover $\widetilde{X}$ then the $R[\pi_1(X)]$-module chain map

$$[X] \cap - : \Delta(\widetilde{X};R)^{n-*} = \mathrm{Hom}_{R[\pi_1(X)]}(\Delta(\widetilde{X};R),R[\pi_1(X)])_{*-n} \to \Delta(\widetilde{X};R)$$

is a chain equivalence, and there are defined Poincaré duality isomorphisms

$$[X] \cap - : H^{n-*}(X;\Lambda) \cong H_*(X;\Lambda)$$

for any $R[\pi_1(X)]$-module Λ.
(iii) A connected finite simplicial complex X with finite fundamental group $\pi_1(X)$ is an n-dimensional universal R-homology Poincaré complex if and only if the universal cover $\widetilde{X}$ is an n-dimensional R-homology Poincaré complex in the sense of 6.2.

Proposition 9.4 A connected n-dimensional R-homology manifold X is an n-dimensional universal R-homology Poincaré complex.

Proof The assembly functor of Ranicki and Weiss [34]

$$A : \mathbb{A}(R,X) \to \mathbb{A}(R[\pi_1(X)]) \; ; \; A \;=\; \sum_{\sigma\in X} A(\sigma) \mapsto A(\widetilde{X}) \;=\; \sum_{\widetilde{\sigma}\in\widetilde{X}} A(p\widetilde{\sigma})$$

is defined for any connected simplicial complex X, with $p : \widetilde{X} \to X$ the universal covering projection. The assembly is a natural transformation of additive categories with chain duality ([29, 9.11]), so that the assembly of the n-dimensional symmetric Poincaré complex $(\Delta(X';R),\Delta[X])$ in $\mathbb{A}(R,X)$ is the n-dimensional symmetric Poincaré complex $(\Delta(\widetilde{X}';R),\Delta[X])$ in $\mathbb{A}(R[\pi_1(X)])$. (This is just a formalization of the standard dual cell proof of Poincaré duality, e.g. Wall [43, Thm. 2.1]). $\square$

In particular, a homology manifold is a universal homology Poincaré complex.

Definition 9.5 (Quinn [22])
(i) An n-dimensional *normal complex* (X,ν_X,ρ_X) is a finite simplicial complex X together with a *normal structure*

$$(\nu_X : X \to BSG(k) \;,\; \rho_X : S^{n+k} \to T(\nu_X)) \;\; (k \text{ large}) \;.$$

The homology class

$$[X] \;=\; U_{\nu_X} \cap h(\rho_X) \;=\; [X] \in H_n(X) \;\; (h = \text{Hurewicz})$$

is the *fundamental class* of X.
(ii) A *normal structure* on an n-dimensional homology Poincaré complex X is a normal structure (ν_X,ρ_X) realizing the fundamental class $[X] \in H_n(X)$.

Remark 9.6 (i) A finite simplicial complex X is an n-dimensional universal homology Poincaré complex if and only if a regular neighbourhood $(U,\partial U)$ of an embedding $X \subset S^{n+k}$ defines a fibration

$$(D^k,S^{k-1}) \to (U,\partial U) \to X$$

(Spivak [41], Wall [42], Ranicki [27]). A n-dimensional universal homology Poincaré complex X has a canonical class of Spivak normal structures $(\nu_X : X \to BSG(k), \rho_X : S^{n+k} \to T(\nu_X))$, namely those represented by such regular neighbourhoods $(U,\partial U)$ with

$$\rho_X : S^{n+k} \to S^{n+k}/\mathrm{cl.}(S^{n+k}\backslash U) \;=\; U/\partial U \;=\; T(\nu_X) \;.$$

(ii) Browder [1] used Poincaré surgery on $\pi_1(X)$ to prove that every n-dimensional homology Poincaré complex X admits normal structures $(\nu_X : X \to BSG(k), \rho_X : S^{n+k} \to T(\nu_X))$, and that for any such structure $\nu_X \oplus \epsilon : X \to BSG(k+1)$ is the normal fibration of a Poincaré embedding $X \subset S^{n+k+1}$ with complement $T(\nu_X) \cup_{\rho_X} D^{n+k+1}$.

Definition 9.7 (Ranicki [29, 17.1])
The *peripheral quadratic complex* of an n-dimensional normal complex X is the $(n-1)$-dimensional quadratic Poincaré complex (C, ψ_X) in $\mathbb{A}(\mathbb{Z}, X)$ with C the X-controlled peripheral chain complex (6.10)

$$C \;=\; \mathcal{C}([X] \cap - : \Delta(X)^{n-*} \to \Delta(X'))_{*+1}$$

and

$$\psi_X \in Q^X_{n-1}(C) \;=\; H_{n-1}(W \otimes_{\mathbb{Z}[\Sigma_2]} (C \otimes_{(\mathbb{Z},X)} C))$$

the X-controlled quadratic class obtained by the boundary construction of [29, 2.6].

Note that the normal complex X is a universal homology Poincaré complex if and only if the peripheral chain complex C is $\mathbb{A}(\mathbb{Z}[\pi_1(X)])$-contractible.

Remark 9.8 The X-controlled quadratic class $\psi_X \in Q^X_{n-1}(C)$ in 9.7 has symmetrization

$$(1+T)\psi_X \;=\; \phi_X \in H_{n-1}(C \otimes_{(\mathbb{Z},X)} C)$$

the chain homotopy class of chain equivalences $\phi_X : C^{n-1-*} \to C$ (6.11). In fact, ψ_X is an X-controlled version of the quadratic class

$$\psi \;=\; \psi_F(U_{\nu_X}) \in Q_{n-1}(C) \;=\; H_{n-1}(W \otimes_{\mathbb{Z}[\Sigma_2]} (C \otimes_{\mathbb{Z}} C))$$

obtained by evaluating the spectral quadratic construction of Ranicki [28, 7.3]

$$\psi_F : \widetilde{H}^k(T(\nu_X)) \to Q_{n-1}(C)$$

on the Thom class $U_{\nu_X} \in \widetilde{H}^k(T(\nu_X))$. Here, $F : T(\nu_X)^* \to \Sigma^\infty X_+$ is a stable map inducing the chain map $[X] \cap - : \Delta(X)^{n-*} \to \Delta(X')$, with $T(\nu_X)^*$ the spectrum S-dual of the Thom space $T(\nu_X)$. If X is homology Poincaré then $T(\nu_X)^* = \Sigma^\infty X_+$. If X is R-homology Poincaré $\psi = 0 \in Q_{n-1}(C) = 0$, but in general $\psi_X \neq 0$.

Refer to Ranicki [29, p.148] for the *algebraic surgery exact sequence* of a simplicial complex X

$$\cdots \to H_n(X; \mathbb{L}_\bullet) \xrightarrow{A} L_n(\mathbb{Z}[\pi_1(X)]) \to \mathbb{S}_n(X) \to H_{n-1}(X; \mathbb{L}_\bullet) \to \dots\ ,$$

with A the assembly map. The generalized homology group

$$H_n(X; \mathbb{L}_\bullet) \;=\; L_n(\mathbb{A}(\mathbb{Z}, X))$$

is the cobordism group of 1-connective n-dimensional quadratic Poincaré complexes (C, ψ_X) in $\mathbb{A}(\mathbb{Z}, X)$, with C an n-dimensional chain complex in $\mathbb{A}(\mathbb{Z}, X)$ and

$$\psi_X \in Q^X_n(C) \;=\; H_n(W \otimes_{\mathbb{Z}[\Sigma_2]} (C \otimes_{(\mathbb{Z},X)} C))$$

such that

$$(1+T)\psi_X \in H_n(C \otimes_{(\mathbb{Z},X)} C) \;=\; H_0(\mathrm{Hom}_{(\mathbb{Z},X)}(C^{n-*}, C))$$

is a chain homotopy class of $(\mathbb{Z}, X)$-module chain equivalences $C^{n-*} \to C$. Here, W is a free $\mathbb{Z}[\Sigma_2]$-module resolution of $\mathbb{Z}$

$$W : \cdots \to \mathbb{Z}[\Sigma_2] \xrightarrow{1-T} \mathbb{Z}[\Sigma_2] \xrightarrow{1+T} \mathbb{Z}[\Sigma_2] \xrightarrow{1-T} \mathbb{Z}[\Sigma_2]$$

and the generator $T \in \Sigma_2$ acts on $C \otimes_{(\mathbb{Z},X)} C$ by signed transposition. The quadratic L-group

$$L_n(\mathbb{Z}[\pi_1(X)]) \;=\; L_n(\mathbb{A}(\mathbb{Z}[\pi_1(X)]))$$

is the cobordism group of n-dimensional quadratic Poincaré complexes (C, ψ) over the group ring $\mathbb{Z}[\pi_1(X)]$ with

$$\psi \in Q_n(C) \;=\; H_n(W \otimes_{\mathbb{Z}[\Sigma_2]} (C \otimes_{\mathbb{Z}[\pi_1(X)]} C)) \;.$$

The *structure group* $\mathbb{S}_n(X)$ is the cobordism group of 1/2-connective $\mathbb{A}(\mathbb{Z}[\pi_1(X)])$-contractible $(n-1)$-dimensional quadratic Poincaré complexes in $\mathbb{A}(\mathbb{Z}, X)$.

Definition 9.9 (Ranicki [29, 17.4])
The *total surgery obstruction* of an n-dimensional universal homology Poincaré complex X is the cobordism class of the peripheral quadratic Poincaré complex in $\mathbb{A}(\mathbb{Z}, X)$

$$s(X) \;=\; (C, \psi_X) \in \mathbb{S}_n(X) \;.$$

Proposition 9.10 Let X be an n-dimensional universal Poincaré complex, with peripheral complex (C, ψ_X).
(i) *The following conditions are equivalent :*

(a) *X is an n-dimensional homology manifold,*

(b) *C is $\mathbb{A}(\mathbb{Z}, X)$-contractible,*

(c) *$(1+T)\psi_X = 0 \in H_{n-1}(C \otimes_{(\mathbb{Z},X)} C)$.*

(ii) *The total surgery obstruction is such that $s(X) = 0$ if (and for $n \geq 5$ only if) the polyhedron $|X|$ is homotopy equivalent to an n-dimensional topological manifold. The image of the total surgery obstruction*

$$t(X) \;=\; [s(X)] \in H_{n-1}(X; \mathbb{L}_\bullet)$$

is such that $t(X) = 0$ if and only if the Spivak normal fibration $\nu_X : X \to BSG$ admits a topological reduction $\widetilde{\nu}_X : X \to BSTOP$.

Proof (i) (a) $\iff$ (b) The peripheral quadratic complex (C, ψ_X) is $\mathbb{A}(\mathbb{Z}, X)$-contractible if and only if the peripheral chain complex C is $\mathbb{A}(\mathbb{Z}, X)$-contractible, if and only if X is a homology manifold (6.11).
(b) $\iff$ (c) The map

$$H_n(X) \to H_{n-1}(C \otimes_{(\mathbb{Z},X)} C)$$

in the braid used in the proof of Theorem 6.13 sends the fundamental class $[X] \in H_n(X)$ to the homology class

$$(1+T)\psi_X \in H_{n-1}(C \otimes_{(\mathbb{Z},X)} C) \ ,$$

and C is $\mathbb{A}(\mathbb{Z}, X)$-contractible if and only if $(1+T)\psi_X = 0$.
(ii) See [29, 17.4]. □

Remark 9.11 There is also an R-coefficient version, for any commutative ring R. The *R-coefficient peripheral complex* (C, ψ_X) of an n-dimensional universal R-homology Poincaré complex X is the $\mathbb{A}(R[\pi_1(X)])$-contractible $(n-1)$-dimensional quadratic Poincaré complex in $\mathbb{A}(R, X)$ with

$$C \ = \ \mathcal{C}([X] \cap - : \Delta(X; R)^{n-*} \to \Delta(X'; R))_{*+1}$$

The *R-coefficient total surgery obstruction* ([29, 26.1]) of X is the cobordism class

$$s(X; R) \ = \ (C, \psi_X) \in \mathbb{S}_n(X; R) \ ,$$

taking value in the R-coefficient structure group fitting into the R-coefficient algebraic surgery exact sequence

$$\dots \to H_n(X; \mathbb{L}_\bullet) \xrightarrow{A} \Gamma_n(R[\pi_1(X)] \to R) \to \mathbb{S}_n(X; R) \to H_{n-1}(X; \mathbb{L}_\bullet) \to \dots$$

with Γ_* the R-homology surgery obstruction groups of Cappell and Shaneson [3]. The R-coefficient total surgery obstruction is such that $s(X; R) = 0$ if (and for $n \geq 5$ only if) the polyhedron $|X|$ is R-homology equivalent to an n-dimensional topological manifold (Ranicki [29, 26.1]). See §14 below for the application to knot theory, with $R = \mathbb{Z}$.

10. Combinatorially controlled surgery theory

This section develops the combinatorial version of the topological controlled surgery theory proposed by Quinn [23] and Ranicki and Yamasaki [37]. In principle, it is possible to construct the topological theory using the combinatorial version and the Čech nerves of open covers (cf. Quinn [25, 1.4]), but this will not be done here.

A degree 1 map $f : M \to N$ of n-dimensional homology manifolds has acyclic point inverses if and only if

$$\Delta_*[M] - (f^! \otimes f^!)\Delta_*[N] = 0 \in H_n(M \times_N M)$$

by 7.7 (ii). For a normal map $(f,b): M \to N$ this obstruction will now be related to the chain level surgery obstruction. The Wall surgery obstruction of (f,b) was expressed in Ranicki [27],[29] as the cobordism class of a kernel n-dimensional quadratic Poincaré complex in $\mathbb{A}(\mathbb{Z}[\pi_1(N)])$

$$\sigma_*(f,b) \ = \ (\Delta(f^!),\psi_b) \in L_n(\mathbb{Z}[\pi_1(N)]) \ .$$

The quadratic class ψ_b will be refined to an N-controlled version $\psi_{b,N}$, with symmetrization

$$(1+T)\psi_{b,N} \ = \ \Delta_*[M] - (f^! \otimes f^!)\Delta_*[N] \in H_n(M \times_N M) \ .$$

Galewski and Stern [7], [8, 1.7] proved that the Spivak normal fibration $\nu_M : M \to BSG$ of a homology manifold M has a canonical topological bundle reduction $\nu_M : M \to BSTOP$, namely the canonical topological bundle reduction of the normal homology cobordism bundle $\nu_M : M \to BSH$, and that in fact for $\dim(M) \geq 5$ there exists a polyhedral topological manifold M_{TOP} with a map $M_{TOP} \to M$ with contractible point inverses.

Definition 10.1 A *normal map* $(f,b): M \to N$ from an n-dimensional homology manifold M to an n-dimensional Poincaré complex N is a degree 1 map $f : M \to N$ with a map of (stable) topological bundles $b : \nu_M \to \eta$ over f.

The surgery obstruction $\sigma_*(f,b) \in L_n(\mathbb{Z}[\pi_1(N)])$ of a normal map $(f,b): M \to N$ is defined by Maunder [16] following Wall [43]. The surgery obstruction is shown in [16] to be such that $\sigma_*(f,b) = 0$ if (and for $n \geq 5$ only if) (f,b) is normal bordant to a homotopy equivalence. The surgery obstruction can also be defined using the chain complex method of Ranicki [26], [27].

Definition 10.2 The *N-controlled quadratic structure* of a normal map $(f,b): M \to N$ of n-dimensional homology manifolds is the element

$$\psi_{b,N} \ = \ \psi_{F,N}[N] \in Q^N_n(\Delta(M)) \ = \ H_n(E\Sigma_2 \times_{\Sigma_2} (M \times_N M))$$

with $\psi_{F,N} : H_*(N) \to Q^N_*(\Delta(M))$ the N-controlled version of the quadratic construction of [27, Chapter 1]

$$\psi_F : H_*(N) \to Q_*(\Delta(M)) \ = \ H_*(E\Sigma_2 \times_{\Sigma_2} (M \times M)) \ .$$

Here, $b : \nu_M \to \eta$ is a stable bundle map over f from the stable normal bundle ν_M of M, η is a bundle over N, $E\Sigma_2$ is a contractible space with a free Σ_2-action, the generator $T \in \Sigma_2$ acts on $M \times_N M$ by transposition

$$T : M \times_N M \ \to \ M \times_N M \ ; \ (x,y) \ \mapsto \ (y,x)$$

and $F : \Sigma^\infty N_+ \to \Sigma^\infty M_+$ is a geometric Umkehr map (= the S-dual of $T(b) : \Sigma^\infty T(\nu_M) \to \Sigma^\infty T(\eta)$) inducing $f^!$ on the chain level.

As usual, write $W=\Delta(E\Sigma_2)$, so that

$$\begin{aligned} Q_n(\Delta(M)) &= H_n(W\otimes_{\mathbb{Z}[\Sigma_2]}(\Delta(M)\otimes_{\mathbb{Z}}\Delta(M)))\ ,\\ Q_n^N(\Delta(M)) &= H_n(W\otimes_{\mathbb{Z}[\Sigma_2]}(\Delta(M)\otimes_{(\mathbb{Z},N)}\Delta(M)))\ .\end{aligned}$$

Remark 10.3 As defined in [27] the quadratic construction ψ_F only gives an element $\psi_b=\psi_F[N]\in Q_n(\Delta(M))$. There are two ways of checking that there is a lift of ψ_b to an N-controlled element $\psi_{b,N}\in Q_n^N(\Delta(M))$:

- Note that the natural chain level transformation in [27, Chapter 1]

  $$\psi_F:\Delta(N)\to W\otimes_{\mathbb{Z}[\Sigma_2]}(\Delta(M)\otimes_{\mathbb{Z}}\Delta(M))$$

  factors through

  $$\psi_{F,N}:\Delta(N)\to W\otimes_{\mathbb{Z}[\Sigma_2]}(\Delta(M)\otimes_{(\mathbb{Z},N)}\Delta(M))$$

  exactly as for the Alexander-Whitney diagonal chain approximation (5.3 (iii)), so that

  $$\psi_F:H_n(N)\xrightarrow{\psi_{F,N}}Q_n^N(\Delta(M))\longrightarrow Q_n(\Delta(M))\ .$$

- Note that (f,b) determines an algebraic normal map in $\mathbb{A}(\mathbb{Z},N)$ in the sense of [29, 2.16], with a corresponding quadratic class $\psi_{b,N}$.

An n-dimensional homology manifold M determines an n-dimensional symmetric Poincaré complex in $\mathbb{A}(\mathbb{Z},N)$

$$\sigma_N^*(M) = (\Delta(M),\Delta_*[M]\in Q_N^n(\Delta(M)))$$

for any simplicial map $M\to N$. Here, the Q-group is defined by

$$Q_N^n(\Delta(M)) = H_n(\mathrm{Hom}_{\mathbb{Z}[\Sigma_2]}(W,\Delta(M)\otimes_{(\mathbb{Z},N)}\Delta(M)))\ ,$$

and $\Delta_*:H_n(M)\to Q_N^n(\Delta(M))$ is induced by the Alexander-Whitney diagonal chain approximation. (Note that Δ_* is an isomorphism for $1:M\to N=M$). The fundamental $\mathbb{L}^\bullet(\mathbb{Z})$-homology class of M (Ranicki [29, 16.16]) is the cobordism class

$$[M]_{\mathbb{L}} = \sigma_M^*(M)\in L^n(\mathbb{A}(\mathbb{Z},M)) = H_n(M;\mathbb{L}^\bullet(\mathbb{Z}))\ .$$

For a degree 1 map $f:M\to N$ the algebraic mapping cone of the Umkehr chain map $f^!:\Delta(N)\to\Delta(M)$ is a $(\mathbb{Z},N)$-module chain complex

$$\Delta(f^!) = \mathcal{C}(f^!:\Delta(N)\to\Delta(M))\ .$$

Let $e:\Delta(M)\to\Delta(f^!)$ be the inclusion. The kernel n-dimensional symmetric Poincaré complex in $\mathbb{A}(\mathbb{Z},N)$

$$\sigma_N^*(f) = (\Delta(f^!),(e\otimes e)\Delta_*[M])$$

is such that up to homotopy equivalence

$$\sigma^*_N(M) \;=\; \sigma^*_N(N) \oplus \sigma^*_N(f) \;,$$

with cobordism class the difference of the fundamental $\mathbb{L}^\bullet(\mathbb{Z})$-homology classes

$$\sigma^*_N(f) \;=\; f_*[M]_{\mathbb{L}} - [N]_{\mathbb{L}} \in H_n(N;\mathbb{L}^\bullet(\mathbb{Z})) \;.$$

Definition 10.4 (Ranicki [29, 18.3])
The *normal invariant* of a normal map $(f,b) : M \to N$ of n-dimensional homology manifolds is the cobordism class

$$\begin{aligned}[f,b]_{\mathbb{L}} \;&=\; (\Delta(f^!),(e\otimes e)\psi_{b,N})\\ &\in L_n(\mathbb{A}(\mathbb{Z},N)) \;=\; H_n(N;\mathbb{L}_\bullet) \;=\; [N,G/TOP] \;.\end{aligned}$$

The normal invariant of 10.4 is a (mild) generalization of the traditional normal invariant in surgery theory, and has the following properties:

- $[f,b]_{\mathbb{L}} \in H_n(N;\mathbb{L}_\bullet)$ is a normal bordism invariant, such that $[f,b]_{\mathbb{L}} = 0$ if f has acyclic point inverses.
- For a normal map of polyhedral topological manifolds $[f,b]_{\mathbb{L}} = 0$ if (and for $n \geq 5$ only if) (f,b) is normal bordant to a homeomorphism.
- The assembly of $[f,b]_{\mathbb{L}}$ in the Wall surgery group is the surgery obstruction of (f,b)
$$A[f,b]_{\mathbb{L}} \;=\; \sigma_*(f,b) \in L_n(\mathbb{Z}[\pi_1(N)]) \;.$$
- The image of $\sigma_*(f,b)$ in the homology surgery Γ-group of Cappell and Shaneson [3]
$$A^H[f,b]_{\mathbb{L}} \;=\; \sigma^H_*(f,b) \in \Gamma_n(\mathbb{Z}[\pi_1(N)] \to \mathbb{Z})$$
is such that $\sigma^H_*(f,b) = 0$ if (and for $n \geq 5$ only if) (f,b) is normal bordant to a homology equivalence.

For PL manifolds these are direct applications of the surgery obstruction theory of Wall [43]. In the general case, apply the extension of the theory to polyhedral homology manifolds due to Maunder [16], or else combine with the result of Galewski and Stern [7], [8, 1.7] that every polyhedral homology manifold can be resolved by a polyhedral topological manifold and the TOP version of Wall's theory.

Proposition 10.5 The N-controlled quadratic class $\psi_{b,N}$ of a normal map $(f,b) : M \to N$ of n-dimensional homology manifolds determines a kernel n-dimensional quadratic Poincaré complex in $\mathbb{A}(\mathbb{Z},N)$

$$\sigma^N_*(f,b) \;=\; (\Delta(f^!),(e\otimes e)\psi_{b,N})$$

with cobordism class the normal invariant of (f,b)

$$[f,b]_{\mathbb{L}} \;=\; \sigma^N_*(f,b) \in L_n(\mathbb{A}(\mathbb{Z},N)) \;=\; H_n(N;\mathbb{L}_\bullet)\ .$$

The Poincaré duality chain equivalence of the symmetrization

$$(1+T)\sigma^N_*(f,b) \;=\; \sigma^*_N(f)$$

is such that up to chain homotopy

$$(1+T)(e\otimes e)\psi_{b,N} \;=\; (e\otimes e)\Delta_*[M] : \Delta(f^!)^{n-*} \to \Delta(f^!)\ ,$$

which is the obstruction to f *having acyclic point inverses (7.7 (ii))*

$$\begin{aligned}(1+T)(e\otimes e)\psi_{b,N} \;&=\; \Delta_*[M] - (f^!\otimes f^!)\Delta_*[N]\\ &\in H_n(\Delta(f^!)\otimes_{(\mathbb{Z},N)}\Delta(f^!)) \;=\; H_n(M\times_N M,\Delta_M)\ \ (7.3\ \text{(ii)})\ .\end{aligned}$$

Proof The identity

$$(1+T)\psi_{b,N} \;=\; \Delta_*[M] - (f^!\otimes f^!)\Delta_*[N] \in H_n(M\times_N M)$$

is just the N-controlled analogue of the standard property of the quadratic construction ([27])

$$(1+T)\psi_b \;=\; \Delta_*[M] - (f^!\otimes f^!)\Delta_*[N] \in H_n(M\times M)\ .$$

□

Remark 10.6 The quadratic class $\psi_{b,N} \in Q^N_n(\Delta(M))$ can be defined for any degree 1 map $f : M \to N$ of n-dimensional universal Poincaré complexes with a map $b : \nu_M \to \nu_N$ of the Spivak normal fibrations, with all the properties of $\psi_{b,N}$ in 10.2 except that the n-dimensional quadratic complex $(\Delta(f^!),(e\otimes e)\psi_{b,N})$ in $\mathbb{A}(\mathbb{Z},N)$ will only be Poincaré in $\mathbb{A}(\mathbb{Z}[\pi_1(N)])$.

A homotopy equivalence $f : M \to N$ of n-dimensional homology manifolds can be regarded as a normal map $(f,b) : M \to N$ with $b : \nu_M \to (f^{-1})^*\nu_M$.

Definition 10.7 (Ranicki [29, 18.3])
The *structure invariant* of a homotopy equivalence $f : M \to N$ of n-dimensional homology manifolds is the cobordism class

$$s(f) \;=\; (\Delta(f^!),\psi_{b,N}) \in \mathbb{S}_{n+1}(N)$$

with image the normal invariant $[f,b]_{\mathbb{L}} \in H_n(N;\mathbb{L}_\bullet)$.

Proposition 10.8 (Ranicki [29, 18.3])
The structure invariant of a homotopy equivalence $f : M \to N$ *of polyhedral* n*-dimensional topological manifolds is such that* $s(f) = 0 \in \mathbb{S}_n(N)$ *if (and for* $n \geq 5$ *only if)* f *is* h*-cobordant to a homeomorphism.*

In §13 below there will be obtained controlled versions of 10.7 and 10.8.

There is also a simple version of the structure invariant, which is defined for a simple homotopy equivalence $f : M \to N$ of n-dimensional homology manifolds, taking value in the relative group $\mathbb{S}^s_n(N)$ in the exact sequence

$$\cdots \to H_n(N;\mathbb{L}_\bullet) \xrightarrow{A} L^s_n(\mathbb{Z}[\pi_1(N)]) \to \mathbb{S}^s_n(N) \to H_{n-1}(N;\mathbb{L}_\bullet) \to \dots \ .$$

Remark 10.9 The simple homotopy theory version of surgery theory allows an application of the s-cobordism theorem, to obtain:
The simple structure invariant of a simple homotopy equivalence $f : M \to N$ of polyhedral n-dimensional topological manifolds is such that $s(f) = 0 \in \mathbb{S}^s_n(N)$ if (and for $n \geq 5$ only if) f is homotopic to a homeomorphism.

Proposition 10.10 (i) *A map $f : M \to N$ of simplicial complexes with acyclic point inverses is simple, with $\tau(f) = 0 \in Wh(\pi_1(N))$.*
(ii) *A homotopy equivalence $f : M \to N$ of n-dimensional homology manifolds with acyclic point inverses is simple, with simple structure invariant $s(f) = 0 \in \mathbb{S}^s_n(N)$.*
(iii) *For $n \geq 5$ a homotopy equivalence $f : M \to N$ of n-dimensional polyhedral topological manifolds with acyclic point inverses is homotopic to a homeomorphism.*
Proof (i) As in the proof of 9.4 use the assembly functor of Ranicki and Weiss [34]

$$A : \mathbb{A}(\mathbb{Z}, N) \to \mathbb{A}(\mathbb{Z}[\pi_1(N)]) \ ; \ A \ = \ \sum_{\sigma \in N} A(\sigma) \mapsto A(\widetilde{N}) \ = \ \sum_{\widetilde{\sigma} \in \widetilde{N}} A(p\widetilde{\sigma})$$

with $p : \widetilde{N} \to N$ the universal covering projection. A choice of basis for each of the f.g. free $\mathbb{Z}$-modules $A(\sigma)$ $(\sigma \in N)$ determines a basis for the assembly f.g. free $\mathbb{Z}[\pi_1(N)]$-module $A(\widetilde{N})$, uniquely up to multiplication by the group elements $g \in \pi_1(N)$. Thus if C is a based $(\mathbb{Z}, N)$-module chain complex such that $C(\widetilde{N})$ is contractible there is a well-defined Whitehead torsion

$$\tau(C(\widetilde{N})) \in Wh(\pi_1(N)) \ .$$

For any simplicial map $f : M \to N$ there is defined a based $(\mathbb{Z}, N)$-module chain complex

$$C \ = \ \mathcal{C}(f : \Delta(M) \to \Delta(N))$$

with

$$\begin{aligned} &C(\sigma) \ = \\ &\mathcal{C}(f| : \Delta(f^{-1}D(\sigma, N), f^{-1}\partial D(\sigma, N)) \to \Delta(D(\sigma, N), \partial D(\sigma, N))) \ \ (\sigma \in N) \ . \end{aligned}$$

The assembly of C is the based f.g. free $\mathbb{Z}[\pi_1(N)]$-module chain complex

$$C(\widetilde{N}) \ = \ \mathcal{C}(\widetilde{f} : \Delta(\widetilde{M}) \to \Delta(\widetilde{N}))$$

with $\widetilde{M} = f^*\widetilde{N}$ the pullback cover of M. If $\widetilde{f}$ is a $\mathbb{Z}[\pi_1(N)]$-module chain equivalence (e.g. if f is a homotopy equivalence) the torsion of f is defined by

$$\tau(f) \ = \ \tau(C(\widetilde{N})) \in Wh(\pi_1(N)) \ .$$

If f has acyclic point inverses each $C(\sigma)$ is contractible, and $\widetilde{f}$ is a $\mathbb{Z}[\pi_1(N)]$-module chain equivalence, with the torsion of f such that

$$\tau(f) \ = \ \tau(C(\widetilde{N})) \in \mathrm{im}(Wh(\{1\}) \to Wh(\pi_1(N))) \ = \ \{0\} \ ,$$

so that $\tau(f) = 0$. (This uses $Wh(\{1\}) = 0$, exactly as in the proof of the combinatorial invariance of Whitehead torsion in Milnor [19]).
(ii) The simple structure invariant $s(f)$ is the cobordism class of the simple $\mathbb{Z}[\pi_1(N)]$-contractible n-dimensional quadratic Poincaré complex $(\Delta(f^!), \psi_{b,N})$ in $\mathbb{A}(\mathbb{Z}, N)$ with

$$f^! \ = \ f^{-1} : \Delta(N) \to \Delta(M) \ .$$

By (i) $\Delta(f^!)$ is simple $(\mathbb{Z}, N)$-contractible, and so represents 0 in the simple structure group.
(iii) By (ii) f is a simple homotopy equivalence with zero simple structure invariant, so that 10.9 applies. □

Remark 10.11 Let $n \geq 5$.
(i) A map $f : M \to N$ of n-dimensional PL manifolds with acyclic point inverses is homotopic through maps with acyclic point inverses to a PL homeomorphism if and only if the Cohen-Sato-Sullivan obstruction

$$c^H(f) \in H^3(N; \theta_3^H)$$

is 0, with θ_3^H the Kervaire-Milnor cobordism group of oriented 3-dimensional PL homology spheres (Ranicki [31, pp.26–28]). The obstruction is 0 if f has contractible point inverses. The obstruction is the homotopy class of the map

$$c^H(f) : N \to H/PL \simeq K(\theta_3^H, 3)$$

classifying the difference between the PL reductions of the normal homology cobordism bundles of M and N. The combination of the Kirby-Siebenmann result

$$TOP/PL \simeq K(\mathbb{Z}_2, 3)$$

with the work of Galewski and Stern [7] shows that the various classifying spaces are related by a commutative braid of fibration sequences

$$\begin{tikzcd}
K(\theta_3^H,3) \arrow[dr, "\alpha"] \arrow[rr, bend left] & & BPL \arrow[dr] \arrow[rr, bend left] & & BTOP \arrow[dr, "\kappa"] \arrow[rr, bend left] & & K(\ker(\alpha),5) \\
& K(\mathbb{Z}_2,3) \arrow[ur] \arrow[dr, "\delta"] & & BH \arrow[ur] \arrow[dr, "c^H"] & & K(\mathbb{Z}_2,4) \arrow[ur, "\delta"] & \\
& & K(\ker(\alpha),4) \arrow[ur] \arrow[rr, bend right] & & K(\theta_3^H,4) \arrow[ur, "\alpha"] & &
\end{tikzcd}$$

with $\alpha : \theta_3^H \to \mathbb{Z}_2$ the Rochlin map ([31, p.26]).
(ii) A homeomorphism $f : M \to N$ of n-dimensional PL manifolds is homotopic to a PL homeomorphism if and only if the Casson-Sullivan obstruction

$$\kappa(f) \;=\; \alpha(c^H(f)) \in H^3(N;\mathbb{Z}_2)$$

is 0 ([31, p.14]). The obstruction is the homotopy class of the map

$$\kappa(f) : N \to TOP/PL \simeq K(\mathbb{Z}_2, 3)$$

classifying the difference between the PL reductions of the normal topological block bundles of M and N.
(iii) A homeomorphism $f : M \to N$ of n-dimensional PL manifolds is homotopic to a simplicial map with acyclic point inverses if and only if the Galewski-Matumoto-Stern obstruction

$$\delta\kappa(f) \in H^4(N; \ker(\alpha))$$

is 0 ([31, p.28]).
(iv) Galewski and Stern [8] proved that an n-dimensional topological manifold N is polyhedral (i.e. can be triangulated by a polyhedron) if and only if the element

$$\delta\kappa(\nu_N) \in H^5(N; \ker(\alpha))$$

is 0. In particular, this obstruction is 0 for the topological manifold $N = M_{TOP}$ resolving a (polyhedral) homology manifold M given by Galewski and Stern [7], so that M_{TOP} can be taken to be polyhedral.

11. Intersections and self-intersections

The chain complex methods of this paper will now be applied to obtained a combinatorially controlled homology version of the intersection theory of homology submanifolds.

Definition 11.1 Given maps of X-controlled R-homology Poincaré complexes

$$f_1 : M_1 \to N \quad , \quad f_2 : M_2 \to N$$

with

$$\dim(M_1) \;=\; m_1 \quad , \quad \dim(M_2) \;=\; m_2 \quad , \quad \dim(N) \;=\; n$$

define the *X-controlled intersection class*

$$[M_1 \times_X M_2] \in H_{m_1+m_2-n}(M_1 \times_X M_2; R)$$

to be the chain homotopy class of the (R,X)-module chain map

$$\begin{aligned}\Delta(M_1;R)^{m_1-*} \simeq \Delta(M_1;R) &\xrightarrow{f_1} \Delta(N;R) \simeq \Delta(N;R)^{n-*} \\ &\xrightarrow{(f_2)^*} \Delta(M_2;R)^{n-*} \simeq \Delta(M_2;R)_{*+m_2-n} ,\end{aligned}$$

using the identifications

$$\begin{aligned}&H_{m_1+m_2-n}(M_1 \times_X M_2;R) \\ &\qquad = H_{m_1+m_2-n}(\Delta(M_1;R) \otimes_{(R,X)} \Delta(M_2;R)) \\ &\qquad = H_0(\mathrm{Hom}_{(R,X)}(\Delta(M_1;R)^{m_1-*}, \Delta(M_2;R)_{*+m_2-n})) .\end{aligned}$$

In terms of the Umkehr (R,X)-module chain maps (7.1)

$$\begin{aligned}f_1^! : \Delta(N;R) \simeq \Delta(N;R)^{n-*} &\xrightarrow{(f_1)^*} \Delta(M_1;R)^{n-*} \simeq \Delta(M_1;R)_{*+m_1-n} , \\ f_2^! : \Delta(N;R) \simeq \Delta(N;R)^{n-*} &\xrightarrow{(f_2)^*} \Delta(M_2;R)^{n-*} \simeq \Delta(M_2;R)_{*+m_2-n}\end{aligned}$$

the X-controlled intersection class is given by the evaluation on the fundamental class $[N] \in H_n(N)$ of the composite

$$H_n(N;R) \xrightarrow{\Delta} H_n(N \times_X N;R) \xrightarrow{f_1^! \otimes f_2^!} H_{m_1+m_2-n}(M_1 \times_X M_2;R) ,$$

that is

$$[M_1 \times_N M_2] \;=\; (f_1^! \otimes f_2^!)\Delta[N] \in H_{m_1+m_2-n}(M_1 \times_X M_2;R) .$$

For the remainder of §11 $R = \mathbb{Z}$, $X = N$, i.e. only homology manifolds will be considered.

Definition 11.2 Embeddings of homology manifolds

$$f_1 : (M_1)^{m_1} \to N^n \quad , \quad f_2 : (M_2)^{m_2} \to N^n$$

are *transverse* if

- the intersection $M_1 \cap M_2$ is an $(m_1 + m_2 - n)$-dimensional homology manifold,
- the product embedding $f_1 \times f_2 : M_1 \times M_2 \to N \times N$ has a normal homology cobordism bundle

  $$\nu_{f_1 \times f_2} \;:\; M_1 \times M_2 \to BSH(2n - m_1 - m_2)$$

  whose restriction to $M_1 \cap M_2$ (viewed as a submanifold of $M_1 \times M_2$) is a normal homology cobordism bundle for $M_1 \cap M_2 \subset N$.

(Compare with the notion of homology manifold transversality considered by Galewski and Stern [7, Chapter 3].)

Proposition 11.3 The N-controlled intersection class of transversely intersecting embeddings of homology manifolds $f_1 : (M_1)^{m_1} \to N^n$, $f_2 : (M_2)^{m_2} \to N^n$ is the fundamental class of the (m_1+m_2-n)-dimensional homology submanifold

$$M_1 \times_N M_2 \ = \ M_1 \cap M_2 \subset N \ ,$$

that is

$$[M_1 \times_N M_2] \ = \ [M_1 \cap M_2] \in H_{m_1+m_2-n}(M_1 \times_N M_2) \ .$$

Proof The normal homology cobordism bundle

$$\nu = \nu_{M_1 \cap M_2 \subset N} : M_1 \cap M_2 \to BSH(2n - m_1 - m_2)$$

is such that there are defined isomorphisms

$$\begin{aligned} H_*(N, N\backslash(M_1 \cap M_2)) \ &\cong \ H_*(N, \mathrm{cl.}(N\backslash E(\nu))) \\ &\cong \ H_*(E(\nu), S(\nu)) \\ &\cong \ H_{*+m_1+m_2-2n}(M_1 \cap M_2) \ . \end{aligned}$$

The identity $[M_1 \times_N M_2] = [M_1 \cap M_2]$ follows from the evaluation of $[N] \in H_n(N)$ in the commutative diagram

$$\begin{tikzcd} H_n(N) \arrow[r, "\cong"] \arrow[d] & H_n(N \times_N N) \arrow[d, "f_1^! \otimes f_2^!"] \\ H_n(N, N\backslash(M_1 \cap M_2)) \arrow[r, "\cong"] & H_{m_1+m_2-n}(M_1 \times_N M_2) \ . \end{tikzcd}$$

□

Given a map $f : M \to N$ define the maps

$$\begin{aligned} &i : M \to M \times_N M \ ; \ x \mapsto (x, x) \ , \\ &j : M \times_N M \to N \ ; \ (x, y) \mapsto f(x) = f(y) \ , \\ &k : M \times_N N \to M \ ; \ (x, y) \mapsto x \end{aligned}$$

(as in §7) such that

$$ji \ = \ f : M \to N \quad , \quad ki \ = \ 1 : M \to M \ .$$

The induced maps

$$i_* : H_*(M) \to H_*(M \times_N M)$$

are split injections, with

$$H_*(M \times_N M) \ = \ H_*(M) \oplus H_*(M \times_N M, \Delta_M) \ .$$

If M, N are homology manifolds with $\dim(M) = m$, $\dim(N) = n$ the Umkehr $(\mathbb{Z}, N)$-module chain maps

$$f^! : \Delta(N) \to \Delta(M)_{*+m-n} \quad , \quad j^! : \Delta(N) \to \Delta(M \times_N M)_{*+2m-2n}$$

are defined as in 7.1,7.8.

Proposition 11.4 Let $f : M^m \to N^n$ be a map of homology manifolds.
(i) *The N-controlled intersection class of f with itself*

$$[M \times_N M] \;=\; j^![N] \in H_{2m-n}(M \times_N M)$$

is such that $[M \times_N M] = 0 \in H_{2m-n}(M \times_N M, \Delta_M)$ if and only if

$$[M \times_N M] \in \mathrm{im}(i_* : H_{2m-n}(M) \to H_{2m-n}(M \times_N M)) \;.$$

(ii) *If f is an embedding then*

$$[M \times_N M] \;=\; [M] \cap e(\nu_f) \in H_{2m-n}(M \times_N M) \;=\; H_{2m-n}(M) \;,$$

with $e(\nu_f) \in H^{n-m}(M)$ the Euler class of the normal homology cobordism bundle $\nu_f : M \to BSH(n-m)$.
Proof (i) Immediate from the definition of $[M \times_N M]$, and the (split) exact sequence

$$0 \to H_{2m-n}(M) \xrightarrow{i_*} H_{2m-n}(M \times_N M) \to H_{2m-n}(M \times_N M, \Delta_M) \to 0 \;.$$

(ii) For an embedding f

$$\begin{aligned} i \;&=\; 1 : M \to M \times_N M \;=\; M \;, \\ j \;&=\; f : M \times_N M \;=\; M \to N \;. \end{aligned}$$

It follows from the commutative diagram

$$\begin{tikzcd}
H_n(N) \arrow[rrr, "\Delta_N"] \arrow[dr, "\cong"] \arrow[ddd, "j^!"'] & & & H_n(N\times N) \arrow[dl, "\cong"'] \arrow[ddd, "f^! \otimes f^!"] \\
& H^n(N\times N, N\times N\backslash\Delta_N) \arrow[r] \arrow[d, "(f\times f)^*"'] & H^n(N\times N) \arrow[d, "(f\times f)^*"] & \\
& H^n(M\times M, M\times M\backslash\Delta_M) \arrow[r] \arrow[dl, "\cong"'] & H^n(M\times M) \arrow[dr, "\cong"] & \\
H_{2m-n}(M) \arrow[rrr, "\Delta_M"] & & & H_{2m-n}(M\times M)
\end{tikzcd}$$

that

$$\Delta_M[M \times_N M] \;=\; \Delta_M j^![N] \;=\; (f^! \otimes f^!)\Delta_N[N] \in H_{2m-n}(M \times M) \;.$$

The Pontrjagin-Thom collapse map

$$F : N_+ \to N/\mathrm{cl.}(N\backslash E(\nu_f)) \;=\; E(\nu_f)/S(\nu_f) \;=\; T(\nu_f)$$

induces the Umkehr $\mathbb{Z}$-module chain map

$$F \;=\; f^! : \Delta(N) \;=\; \widetilde{\Delta}(N_+) \to \widetilde{\Delta}(T(\nu_f)) \simeq \Delta(M)_{*+m-n} \ .$$

It follows from the commutative diagram

$$\begin{tikzcd}
H_n(N) \arrow[r, "F = f^!"] \arrow[d, "\Delta_N"'] & \widetilde{H}_n(T(\nu_f)) \cong H_m(M) \arrow[d, "\Delta_{T(\nu_f)}"] \arrow[dr, "- \cap e(\nu_f)"] & \\
 & & H_{2m-n}(M) \arrow[dl, "\Delta_M"] \\
H_n(N\times N) \arrow[r, "f^! \otimes f^!"] & \widetilde{H}_n(T(\nu_f) \wedge T(\nu_f)) \cong H_{2m-n}(M \times M) &
\end{tikzcd}$$

that

$$\begin{aligned}
(f^! \otimes f^!)\Delta_N[N] \;&=\; \Delta_M(f^![N] \cap e(\nu_f)) \\
&=\; \Delta_M([M] \cap e(\nu_f)) \in H_{2m-n}(M \times M) \ .
\end{aligned}$$

Thus

$$\begin{aligned}
\Delta_M[M \times_N M] \;&=\; (f^! \otimes f^!)\Delta_N[N] \\
&=\; \Delta_M([M] \cap e(\nu_f)) \in H_{2m-n}(M \times M) \ .
\end{aligned}$$

Now $\Delta_M : H_{2m-n}(M) \to H_{2m-n}(M \times M)$ is a (split) injection, so that

$$[M \times_N M] \;=\; [M] \cap e(\nu_f) \in H_{2m-n}(M) \ .$$

□

Remark 11.5 (i) If $f : M^m \to N^n$ is a map of homology manifolds with an N-controlled map

$$a : (M \times M, M \times M\backslash\Delta_M) \to (N \times N, N \times N\backslash\Delta_N)$$

such that the diagram

$$\begin{tikzcd}
 & (M \times M, M \times M\backslash M \times_N M) \arrow[dl, "c"'] \arrow[dr, "f \times f"] & \\
(M \times M, M \times M\backslash\Delta_M) \arrow[rr, "a"] & & (N \times N, N \times N\backslash\Delta_N)
\end{tikzcd}$$

is N-controlled homotopy commutative then

$$[M \times_N M] \in \operatorname{im}(i_* : H_{2m-n}(M) \to H_{2m-n}(M \times_N M)) \ ,$$

where c is the inclusion. (For $m = n$ this is essentially the same as 8.6.) The property $ac \simeq f \times f$ is related to the necessary and sufficient condition obtained by Haefliger [10] for a map $f : M^m \to N^n$ of differentiable manifolds in the stable range $2n \geq 3(m+1)$ to be homotopic to an embedding, namely that $f \times f : M \times M \to N \times N$ be Σ_2-equivariantly homotopic to a map $h : M \times M \to N \times N$ with $h^{-1}(\Delta_N) = \Delta_M$, so that h defines a map of pairs

$$h : (M \times M, M \times M \backslash \Delta_M) \to (N \times N, N \times N \backslash \Delta_N) \ .$$

The action of Σ_2 is by transposition $(x, y) \mapsto (y, x)$. See 11.11 below for a more detailed discussion of the case $n = 2m$.
(ii) The identity of 11.4 (ii) for an embedding $f : M^m \to N^n$ can also be proved geometrically, whenever there exists an isotopic embedding $f' : M' = M \to N$ such that:

- $M, M' \subset N$ intersect transversely in a $(2m-n)$-dimensional homology submanifold $M \cap M' \subset N$,
- $[M \cap M'] \in H_{2m-n}(M)$ is Poincaré dual to $e(\nu_f) \in H^{n-m}(M)$,
- $[M \times_N M] = [M \cap M'] \in H_{2m-n}(M)$.

Applying 11.3, it follows that

$$\begin{aligned} [M \times_N M] \ &= \ [M \cap M'] \\ &= \ [M] \cap e(\nu_f) \in H_{2m-n}(M \times_N M) \ = \ H_{2m-n}(M) \ . \end{aligned}$$

(iii) Let $f : M \to X$ be a degree 1 map of n-dimensional manifolds, which is covered by a stable bundle map

$$b : \tau_M \oplus \epsilon^\infty \to \tau_X \oplus \epsilon^\infty \ .$$

The induced stable map of Thom spaces

$$T(b) : T(\tau_M \oplus \epsilon^\infty) = \Sigma^\infty T(\tau_M) \to T(\tau_X \oplus \epsilon^\infty) = \Sigma^\infty T(\tau_X)$$

sends the Thom class of τ_X to the Thom class of τ_M

$$T(b)^* : \widetilde{H}^n(T(\tau_X)) \to \widetilde{H}^n(T(\tau_M)) \ ; \ U_X \to U_M$$

The images of the Thom classes U_M, U_X under the maps

$$\begin{aligned} \text{inclusion}^* : \widetilde{H}^n(T(\tau_M)) &\cong H^n(M \times M, M \times M \backslash \Delta_M) \\ &\to H^n(M \times M, M \times M \backslash (f \times f)^{-1} \Delta_X) \ , \\ f \times f^* : H^n(X \times X, X \times X \backslash \Delta_X) &\to H^n(M \times M, M \times M \backslash (f \times f)^{-1} \Delta_X) \end{aligned}$$

are not the same (in general), since the diagram I in

$$\begin{tikzcd}
\widetilde{H}^n(T(\tau_M)) \arrow[d, phantom, "\cong" sloped] & & \widetilde{H}^n(T(\tau_X)) \arrow[ll, "T(b)^*"'] \arrow[d, phantom, "\cong" sloped] \\
H^n(M\times M, M\times M\backslash\Delta_M) \arrow[ddr, bend angle=0, to path={}] \arrow[dd] \arrow[dr] & I & H^n(X\times X, X\times X\backslash\Delta_X) \arrow[dl] \arrow[dd] \\
& H^n(M\times M, M\times M\backslash(f\times f)^{-1}\Delta_X) \arrow[dl] \arrow[dr] & \\
H^n(M\times M) & & H^n(X\times X) \arrow[ll, "(f\times f)^*"']
\end{tikzcd}$$

does not commute. However, it does commute in the unstable case $b:\tau_M\to\tau_X$, with a commutative diagram

$$\begin{tikzcd}
(M\times M, M\times M\backslash\Delta_M) \arrow[r, "b"] \arrow[d] & (X\times X, X\times X\backslash\Delta_X) \arrow[d] \\
M \arrow[r, "f"] & X
\end{tikzcd}$$

Definition 11.6 The *homotopy double point set* $P(f)$ of a map $f:M\to N$ is the homotopy pullback in the diagram

$$\begin{tikzcd}
P(f) \arrow[r] \arrow[d] & M \arrow[d, "f"] \\
M \arrow[r, "f"] & N
\end{tikzcd}$$

Thus $P(f)$ is the space of triples (x,y,ω) with $x,y\in M$ and $\omega:[0,1]\to N$ a path such that

$$\omega(0) \;=\; f(x) \quad , \quad \omega(1) \;=\; f(y)\in N\ .$$

The space $P(f)$ is a homotopy model for the actual double point set $M\times_N M$, and there is an evident inclusion

$$M\times_N M\to P(f)\ ;\ (x,y)\mapsto(x,y,\omega)$$

with $\omega(t)=f(x)=f(y)\in N$ $(0\le t\le 1)$.

Proposition 11.7 If $f:M^m\to N^n$ is a map of homology manifolds the image of $[M\times_N M]\in H_{2m-n}(M\times_N M,\Delta_M)$ in $H_{2m-n}(P(f),\Delta_M)$ is a homotopy invariant of f, which is 0 if f is homotopic to an embedding.

Proof Immediate from 11.4. □

Remark 11.8 See Hatcher and Quinn [11] for the systematic use of homotopy pullbacks to define intersection invariants of submanifolds.

Next, consider an immersion of an m-dimensional homology manifold in an n-dimensional homology manifold

$$f : M^m \to N^n$$

with $m < n$. Let $\nu_f : M \to BSH(n-m)$ classify the normal homology cobordism bundle, so that there is defined a fibration

$$(D^{n-m}, S^{n-m-1}) \to (E(\nu_f), S(\nu_f)) \to M$$

and the Thom space is given by

$$T(\nu_f) \;=\; E(\nu_f)/S(\nu_f) \;.$$

For sufficiently large k there exists a map $g : M \to \mathrm{int}(D^k)$ such that

$$f \times g : M \to N \times D^k \;;\; x \mapsto (f(x), g(x))$$

is an embedding with normal homology cobordism bundle

$$\nu_{f\times g} \;=\; \nu_f \oplus \epsilon^k : M \to BSH(n-m+k) \;.$$

The corresponding Pontrjagin-Thom collapse map

$$F : \Sigma^k N_+ \;=\; N \times D^k/N \times S^{k-1} \to T(\nu_f \oplus \epsilon^k) \;=\; \Sigma^k T(\nu_f)$$

induces the Umkehr $(\mathbb{Z}, N)$-module chain map

$$f^! : \Delta(N) \simeq \Delta(N)^{n-*} \xrightarrow{f^*} \Delta(M)^{n-*} \simeq \Delta(M)_{*+m-n} \simeq \widetilde{\Delta}(T(\nu_f)) \;.$$

Let

$$\nu_f \times_N \nu_f : M \times_N M \to BSH(2(n-m))$$

be the homology cobordism bundle defined by the restriction of the product

$$\nu_f \times \nu_f : M \times M \to BSH(2(n-m))$$

to $M \times_N M \subseteq M \times M$, with Thom space

$$\begin{aligned} T(\nu_f \times_N \nu_f) \;&=\; E(\nu_f \times_N \nu_f)/S(\nu_f \times_N \nu_f) \\ &=\; E(\nu_f) \times_N E(\nu_f)/(E(\nu_f) \times_N S(\nu_f) \cup S(\nu_f) \times_N E(\nu_f)) \;. \end{aligned}$$

Definition 11.9 The *N-controlled self-intersection class* of an immersion of homology manifolds $f : M^m \to N^n$ is the N-controlled version of the homology class of Ranicki [27, pp.279-282]

$$\begin{aligned} \mu_N(f) \;=\; & -\psi_{F,N}[N] \\ & \in \widetilde{H}_n(E\Sigma_2 \ltimes_{\Sigma_2} T(\nu_f \times_N \nu_f)) \\ & \qquad =\; H_{2m-n}(E\Sigma_2 \times_{\Sigma_2} (M \times_N M); \mathbb{Z}^{(-)^{n-m}}) \\ & \qquad =\; H_{2m-n}(W\mathbb{Z}^{(-)^{n-m}} \otimes_{\mathbb{Z}[\Sigma_2]} (\Delta(M) \otimes_{(\mathbb{Z},N)} \Delta(M))) \end{aligned}$$

with $\psi_{F,N}$ (9.2) the N-controlled version of the quadratic construction ψ_F of [27, Chapter 1] applied to a geometric Umkehr map $F : \Sigma^k N_+ \to \Sigma^k T(\nu_f)$ (k large) inducing $f^!$ on the chain level. Here, $\mathbb{Z}^{(-)^{n-m}}$ refers to $\mathbb{Z}$ twisted by the orientation character of the extended power homology cobordism bundle

$$e_2(\nu_f) : E\Sigma_2 \times_{\Sigma_2} (M \times_N M) \to BH(2(n-m))$$

with

$$\begin{aligned} E(e_2(\nu_f)) &= E\Sigma_2 \times_{\Sigma_2} (E(\nu_f) \times_N E(\nu_f)) \ , \\ S(e_2(\nu_f)) &= E\Sigma_2 \times_{\Sigma_2} (E(\nu_f) \times_N S(\nu_f) \cup S(\nu_f) \times_N E(\nu_f)) \ , \\ T(e_2(\nu_f)) &= E(e_2(\nu_f))/S(e_2(\nu_f)) \ = \ E\Sigma_2 \ltimes_{\Sigma_2} T(\nu_f \times_N \nu_f) \ , \end{aligned}$$

and $W\mathbb{Z}^{(-)^{n-m}}$ is a free $\mathbb{Z}[\Sigma_2]$-resolution of $\mathbb{Z}^{(-)^{n-m}}$.

Proposition 11.10 (i) *The N-controlled self-intersection class has symmetrization*

$$\begin{aligned} (1+T)\mu_N(f) &= [M \times_N M] - i_*(e(\nu_f) \cap [M]) \\ &\in \widetilde{H}_n(T(\nu_f \times_N \nu_f)) \ = \ H_{2m-n}(M \times_N M) \ , \end{aligned}$$

with

$$[M \times_N M] \ = \ (f^! \otimes f^!)\Delta_N[N] \in H_{2m-n}(M \times_N M) \ .$$

(ii) *The image of $\mu_N(f)$ in*

$$\begin{aligned} & H_{2m-n}(E\Sigma_2 \times_{\Sigma_2} (M \times_N M), E\Sigma_2 \times_{\Sigma_2} \Delta_M; \mathbb{Z}^{(-)^{n-m}}) \\ &\quad = \ H^{lf}_{2m-n}(E\Sigma_2 \times_{\Sigma_2} (M \times_N M \backslash \Delta_M); \mathbb{Z}^{(-)^{n-m}}) \\ &\quad = \ H^{lf}_{2m-n}((M \times_N M \backslash \Delta_M)/\Sigma_2; \mathbb{Z}^{(-)^{n-m}}) \end{aligned}$$

is a $\mathbb{Z}^{(-)^{n-m}}$-twisted fundamental class for the stratified set of unordered double points[1]

$$(M \times_N M \backslash \Delta_M)/\Sigma_2 = \{(x,y) \in M \times M \,|\, x \neq y, f(x) = f(y)\}/\{(x,y) \sim (y,x)\} \ .$$

(iii) *If $f : M \to N$ is an embedding then it is possible to chose $k = 0$, $F : N_+ \to T(\nu_f)$ and $\mu_N(f) = 0$.*
(iv) *The image of $\mu_N(f)$ in $H^{lf}_{2m-n}((M \times M \backslash \Delta_M)/\Sigma_2; \mathbb{Z}^{(-)^{n-m}})$ is a regular homotopy invariant of f, which is 0 if f is regular homotopic to an embedding.*
Proof These are the N-controlled versions of standard properties of the self-intersection form μ of Chapter 5 of Wall [43]. □

Let $f : M^m \to N^n$ be a map of connected homology manifolds with $n = 2m$, such that $f_* : \pi_1(M) \to \pi_1(N)$ is trivial. Write $\pi_1(N) = \pi$, and let $g : N \to B\pi$

[1] The unordered double point set of an immersion of manifolds $f : M^m \to N^n$ is an open $(2m-n)$-dimensional manifold in the metastable range $3m < 2n$, when there are no triple points.

be the classifying map for the universal cover $\widetilde{N} = g^*E\pi$ of N. A choice of null-homotopy $gf \simeq \{*\} : M \to B\pi$ determines a homotopy equivalence $P(gf) \simeq \pi \times M \times M$, with $P(gf)$ the homotopy double point set (11.6), as well as a lift $\widetilde{f} : M \to \widetilde{N}$ of $f : M \to N$. The N-controlled intersection class (11.1) is an element

$$[M \times_N M] \in H_0(M \times_N M)$$

with image the intersection class of Wall [43, 5.2]

$$\lambda(f,f) \in H_0(P(gf)) \;=\; \mathbb{Z}[\pi] \ ,$$

which is a homotopy invariant of f. The following result was first obtained in the differentiable category.

Proposition 11.11 (Haefliger [10])
The reduced intersection class of a map $f : M^m \to N^{2m}$

$$\widetilde{\lambda}(f,f) \;=\; [\lambda(f,f)] \in H_0(P(gf), \Delta_M) \;=\; \mathbb{Z}[\pi]/\mathbb{Z}$$

is such that $\widetilde{\lambda}(f,f) = 0$ if (and for $m \geq 3$ only if) f is homotopic to an embedding.

Now assume that $f : M^m \to N^{2m}$ is an immersion, so that the double point set $M \times_N M$ is the disjoint union of Δ_M and a finite set $M \times_N M \backslash \Delta_M$. The N-controlled self-intersection class (11.9)

$$\mu_N(f) \in H_0(E\Sigma_2 \times_{\Sigma_2} (M \times_N M); \mathbb{Z}^{(-)^m})$$

has image the self-intersection form of [43, 5.2]

$$\begin{aligned}\mu(f) \;&=\; \sum_{(x,y)\in(M\times_N M\backslash\Delta_M)/\Sigma_2} w(x,y)g(x,y) \\ &\in H_0(E\Sigma_2 \times_{\Sigma_2} P(gf); \mathbb{Z}^{(-)^m}) \;=\; \mathbb{Z}[\pi]/\{a - (-)^m\overline{a}\}\end{aligned}$$

where

- $a \mapsto \overline{a}$ is the involution on the fundamental group ring $\mathbb{Z}[\pi]$ defined (as in §9) by

$$\mathbb{Z}[\pi] \to \mathbb{Z}[\pi] \ ; \ a \;=\; \sum_{g\in\pi} n_g g \mapsto \overline{a} \;=\; \sum_{g\in\pi} n_g g^{-1} \ ,$$

- $g(x,y) \in \pi$ is the fundamental group element determined for each non-trivial ordered double point $(x,y) \in M \times_N M \backslash \Delta_N$ by

$$\widetilde{f}(x) \;=\; g(x,y)\widetilde{f}(y) \in \widetilde{N} \ ,$$

- $w(x,y) = \pm 1$ according to the matching up of the orientations of M and N at $f(x) = f(y) \in N$.

The symmetrization of $\mu(f)$ is such that

$$\mu(f) + (-)^m\overline{\mu(f)} \;=\; \lambda(f,f) - \chi(\nu_f) \in \mathbb{Z}[\pi] \ ,$$

a special case of 11.10 (i), with $\chi(\nu_f) \in \mathbb{Z} \subseteq \mathbb{Z}[\pi]$.

Proposition 11.12 (Wall [43, 5.2])
The self-intersection form of an immersion $f : M^m \to N^{2m}$

$$\mu(f) \in \mathbb{Z}[\pi]/\{a - (-)^m\overline{a}\}$$

is a regular homotopy invariant such that $\mu(f) = 0$ if (and for $m \geq 3$) if f is regular homotopic to an embedding.

In fact, the reduced self-intersection form

$$\widetilde{\mu}(f) \in \mathbb{Z}[\pi]/(\mathbb{Z} + \{a - (-)^m\overline{a}\})$$

is a homotopy invariant of f. The condition $m \geq 3$ in 11.12 is necessary for the application of the Whitney trick to remove pairs of double points, with $\mu(f) = 0$ being just the algebraic condition for the double points to appear in potentially cancelling pairs. The result of 11.12 for an immersion $f : S^m \to N^{2m}$ is of course essential for even-dimensional surgery obstruction theory.

12. Whitehead torsion

It is a commonplace of controlled topology that the Whitehead torsion of an X-controlled homotopy equivalence of X-controlled complexes has zero image in $Wh(\pi_1(X))$. See for example the controlled K-theory proof in Ranicki and Yamasaki [36] of Chapman's theorem on the topological invariance of Whitehead torsion.

Proposition 12.1 If $f : M \to N$ is a homotopy equivalence of simplicial complexes which is also an X-controlled homology equivalence then the Whitehead torsion of f is such that

$$\tau(f) \in \ker((p_N)_* : Wh(\pi_1(N)) \to Wh(\pi_1(X))) \ .$$

Proof Work as in 9.10 (i): the algebraic mapping cone of the $(\mathbb{Z}, X)$-module chain equivalence $f : \Delta(M) \to \Delta(N)$

$$C \;=\; \mathcal{C}(f : \Delta(M) \to \Delta(N))$$

is a based contractible finite chain complex in $\mathbb{A}(\mathbb{Z}, X)$, with assembly the based contractible finite chain complex in $\mathbb{A}(\mathbb{Z}[\pi_1(X)])$

$$C(\widetilde{X}) \;=\; \mathcal{C}(\widetilde{f} : \Delta(\widetilde{M}) \to \Delta(\widetilde{N}))$$

with 0 torsion in $Wh(\pi_1(X))$. The image of $\tau(f) \in Wh(\pi_1(N))$ in $Wh(\pi_1(X))$ is thus

$$(p_N)_*\tau(f) \;=\; \tau(C(\widetilde{X})) \;=\; 0 \in Wh(\pi_1(X)) \ .$$

$\square$

Definition 12.2 An X*-controlled h-cobordism* $(W; M, N)$ of homology manifolds is an h-cobordism together with a simplicial map $p_W : W \to X'$ such that the inclusions $M \to W$, $N \to W$ are X-controlled homology equivalences.

Proposition 12.3 The Whitehead torsion of an X*-controlled* h*-cobordism* $(W; M, N)$ *of homology manifolds is such that*

$$\tau(W; M, N) \in \ker((p_W)_* : Wh(\pi_1(W)) \to Wh(\pi_1(X))) \ .$$

Proof By definition

$$\tau(W; M, N) \;=\; \tau(M \to W) \in Wh(\pi_1(W)) \ .$$

Apply 12.2 to the X-controlled homotopy equivalence $M \to W$. $\square$

Corollary 12.4 If $\pi_1(W) \cong \pi_1(X)$ *an* N*-controlled* h*-cobordism* $(W; M, N)$ *of homology manifolds is an* s*-cobordism, with*

$$\tau(W; M, N) \;=\; 0 \in Wh(\pi_1(W)) \ .$$

Proof Immediate from 12.3, since in this case $p_W : W \to X = N$ is a homotopy equivalence. $\square$

In principle, it would be possible to investigate X-controlled versions of the classical h- and s-cobordism theorems of high-dimensional manifold theory, taking the controlled h-cobordism theorem of Quinn [23] as a model.

13. Homology fibrations

It is a theme of controlled topology that if $F \to E \to B$ is a fibre bundle of manifolds and $f : M \to E$ is a homotopy equivalence of manifolds then M is the total space of a fibre bundle $F \to M \to B$ if and only if f is a B-controlled homotopy equivalence. For example, see Chapman [4]. (All niceties to do with fibre bundles, block bundles, approximate fibrations etc. are being ignored here!). An analogous result will now be obtained in the combinatorial context of this paper.

Definition 13.1 A B*-controlled* R*-homology fibration* E is a B-controlled simplicial complex E such that the inclusions

$$p_E^{-1}D(\tau, B) \to p_E^{-1}D(\sigma, B) \quad (\sigma \leq \tau \in B)$$

are R-homology equivalences, i.e. induce isomorphisms

$$H_*(p_E^{-1}D(\tau, B); R) \cong H_*(p_E^{-1}D(\sigma, B); R) \ .$$

The R-module chain homotopy type of $\Delta(p_E^{-1}D(\sigma,B);R)$ is the *chain fibre* of E. (It is assumed here that B is connected, so that the chain fibre is well-defined.)

Remark 13.2 An (R,B)-module chain complex C is *homogeneous* if the inclusions define R-module chain equivalences

$$[C][\sigma] \xrightarrow{\simeq} [C][\tau] \quad (\tau \leq \sigma \in B)$$

(Ranicki and Weiss [34, 4.5], Ranicki [29, p.110]). A B-controlled simplicial complex E is a B-controlled R-homology fibration if and only if the (R,B)-module chain complex $\Delta(E;R)$ is homogeneous.

Example 13.3 Let E be a B-controlled simplicial complex.
(i) The control map $p_E : E \to B'$ has R-acyclic point inverses if and only if E is a B-controlled R-homology fibration with R-acyclic chain fibre.
(ii) The control map $p_E : E \to B'$ is a quasifibration in the sense of Dold and Thom [6] with fibre $F = p_E^{-1}(*)$ if and only if the inclusions

$$p_E^{-1}D(\tau,B) \to p_E^{-1}D(\sigma,B) \quad (\sigma \leq \tau \in B)$$

are homotopy equivalences, in which case E is a B-controlled R-homology fibration with chain fibre $\Delta(F;R)$.

Definition 13.4 A *d-dimensional B-controlled R-homology Poincaré fibration E* is a B-controlled R-homology fibration such that each $p_E^{-1}D(\sigma,B)$ $(\sigma \in B)$ is a d-dimensional R-homology Poincaré complex, with each inclusion

$$p_E^{-1}D(\tau,B) \to p_E^{-1}D(\sigma,B) \quad (\sigma \leq \tau \in B)$$

an R-homology equivalence such that the induced isomorphism

$$H_d(p_E^{-1}D(\tau,B);R) \cong H_d(p_E^{-1}D(\sigma,B);R)$$

sends $[p_E^{-1}D(\tau,B)]$ to $[p_E^{-1}D(\sigma,B)]$.

The chain fibre C of a d-dimensional B-controlled R-homology Poincaré fibration E is a d-dimensional symmetric Poincaré complex over R. (See Ranicki [26] for the theory of symmetric Poincaré complexes.)

Proposition 13.5 Let B be an n-dimensional R-homology manifold B, and let E be a d-dimensional B-controlled R-homology Poincaré fibration, with chain fibre C.
(i) *E is an $(n+d)$-dimensional B-controlled R-homology Poincaré complex.*
(ii) *$E \times_B E$ is an $(n+2d)$-dimensional B-controlled R-homology Poincaré fibration with chain fibre the $2d$-dimensional symmetric Poincaré complex $C \otimes_R C$ over R. In particular, $E \times_B E$ is an $(n+2d)$-dimensional B-controlled R-homology Poincaré complex.*

Proof (i) Use the algebraic Poincaré cycle theory of Ranicki [29], involving the symmetric L-spectrum $\mathbb{L}^\bullet(R)$ with homotopy groups the symmetric L-groups of R

$$\pi_*(\mathbb{L}^\bullet(R)) \;=\; L^*(R)\ .$$

The $\mathbb{L}^\bullet(R)$-homology group $H_m(B;\mathbb{L}^\bullet(R))$ is the cobordism group of m-dimensional symmetric Poincaré cycles in $\mathbb{A}(R,B)$, and the cap product

$$\cap : H_n(B;\mathbb{L}^\bullet(R)) \otimes H^{-d}(B;\mathbb{L}^\bullet(R)) \to H_{n+d}(B;\mathbb{L}^\bullet(R))$$

is defined using the ring spectrum structure of $\mathbb{L}^\bullet(R)$. The R-coefficient homology class

$$[E] \;=\; [B]\otimes[F] \in H_{n+d}(E;R) \;=\; H_n(B;R)\otimes_R H_d(F;R)$$

determines an $(n+d)$-dimensional symmetric cycle $[E]_{\mathbb{L}} = (\Delta(E;R),\Delta[E])$ in $\mathbb{A}(R,B)$ which is Poincaré if and only if E is an $(n{+}d)$-dimensional B-controlled R-homology Poincaré complex, in which case $[E]_{\mathbb{L}} \in H_{n+d}(E;\mathbb{L}^\bullet(R))$ is a fundamental $\mathbb{L}^\bullet(R)$-homology class. The cap product (on the algebraic Poincaré cycle level) of the fundamental $\mathbb{L}^\bullet(R)$-homology class of [29, 16.16]

$$[B]_{\mathbb{L}} \in H_n(B;\mathbb{L}^\bullet(R))$$

and the $\mathbb{L}^\bullet(R)$-cohomology class

$$[C,p_E]_{\mathbb{L}} \in H^{-d}(B;\mathbb{L}^\bullet(R))$$

of Lück and Ranicki [14, Appendix] identifies

$$[E]_{\mathbb{L}} \;=\; [B]_{\mathbb{L}} \cap [C,p_E]_{\mathbb{L}} \in H_{n+d}(B;\mathbb{L}^\bullet(R))\ ,$$

so that $[E]_{\mathbb{L}}$ is a Poincaré cycle, as required.
(ii) For any B-controlled R-homology fibration E with chain fibre C the product $E\times_B E$ is a B-controlled R-homology fibration with chain fibre $C\otimes_R C$. Thus if E is a d-dimensional B-controlled R-homology Poincaré fibration then $E\times_B E$ is a $2d$-dimensional B-controlled R-homology Poincaré fibration, and (i) applies.
□

Remark 13.6 The result of 13.5 (i) is a combinatorial version of the result of Buoncristiano, Rourke and Sanderson [2, p.21] that the total space of a mock bundle is a manifold, and of the result of Gottlieb [9] (announced by Quinn [22]) that the total space of a fibration $F \to E \to B$ with base B an n-dimensional Poincaré complex and fibre F a d-dimensional Poincaré complex is an $(n+d)$-dimensional Poincaré complex E.

Remark 13.7 Let E be a $(d{+}1)$-dimensional homology manifold with a simplicial map $p_E : E \to S^1$ such that the induced infinite cyclic cover of E

$$\overline{E} \;=\; (p_E)^*\mathbb{R}$$

is finitely dominated. Let $\zeta : \overline{E} \to \overline{E}$ be a generating covering translation, with mapping torus

$$T(\zeta) \;=\; \overline{E} \times [0,1]/\{(x,0) = (\zeta(x),1) \,|\, x \in \overline{E}\}\ .$$

The fibering obstruction of E

$$\Phi(E) \;=\; \tau(T(\zeta) \to E) \in Wh(\pi_1(E))$$

is such that $\Phi(E) = 0$ if (and for $d \geq 5$ only if) $p_E : E \to S^1$ is homotopic to the projection of a d-dimensional S^1-controlled homology Poincaré fibration. For an actual manifold E this is the original fibering obstruction of Farrell and Siebenmann, and the S^1-controlled homology Poincaré fibration can be taken to be an actual fibre bundle over S^1. See Ranicki [30],[33] and Hughes and Ranicki [12] for more recent accounts of the fibering obstruction over S^1.

Theorem 13.8 Let B be an n-dimensional R-homology manifold, and let E be a d-dimensional B-controlled R-homology Poincaré fibration with chain fibre C, so that E is an $(n+d)$-dimensional B-controlled R-homology Poincaré complex (13.5 (i)). If M is an $(n+d)$-dimensional B-controlled R-homology Poincaré complex and $f : M \to E$ is a degree 1 B-controlled map, the following conditions are equivalent:

(i) *M is a B-controlled R-homology fibration with chain fibre C,*

(ii) *f is a B-controlled R-homology equivalence,*

(iii) $(f \times f)_* : H_{n+d}(M \times_B M; R) \cong H_{n+d}(E \times_B E; R)$.

Proof (i) $\Longleftrightarrow$ (ii) A map $f : M \to E$ of B-controlled simplicial complexes is a B-controlled R-homology equivalence if and only if the restrictions

$$f| : p_M^{-1} D(\sigma, B) \to p_E^{-1} D(\sigma, B) \ \ (\sigma \in B)$$

are R-homology equivalences.
(ii) $\Longleftrightarrow$ (iii) This is a special case of 7.3. $\square$

Remark 13.9 Corollary 7.5 is the special case of 13.8 with $R = \mathbb{Z}$, $B = E$, $C = R$, $d = 0$ (cf. 13.3 (i)).

14. Knot theory

The results of §§7,13 are now illustrated by showing how they apply to high-dimensional knot theory. No actual new results are obtained in knot theory, however; known results are restated in terms of the chain theory developed in this paper.

The algebraic theory of surgery was used in Ranicki [28, 7.8], [33] to obtain a chain complex treatment of the algebraic invariants of high-dimensional knot theory, using the following construction. Let $k : S^n \subset S^{n+2}$ $(n \geq 1)$ be a locally flat n-knot, with closed regular neighbourhood

$$(U, \partial U) \;=\; (S^n \times D^2, S^n \times S^1) \subset S^{n+2} \;.$$

The knot complement

$$(T, \partial T) \;=\; (\mathrm{cl.}(S^{n+2} \backslash U), \partial U)$$

is an $(n+2)$-dimensional manifold with boundary, such that the generator $1 \in H^1(T) = H_n(U) = \mathbb{Z}$ is realized by a normal map

$$(f, b) : (T, \partial T) \to (D^{n+1} \times S^1, S^n \times S^1)$$

with $f : T \to D^{n+1} \times S^1$ a $\mathbb{Z}$-homology equivalence, and $f| = 1 : \partial T \to S^n \times S^1$. Define an $(n+3)$-dimensional $\mathbb{Z}$-homology Poincaré pair $(X, \partial X)$ with X the mapping cylinder of f, and the boundary $\partial X = T \cup_\partial D^{n+1} \times S^1$ an $(n+2)$-dimensional manifold. The peripheral complex of $(X, \partial X)$ is a $\mathbb{Z}$-contractible $(n+2)$-dimensional quadratic Poincaré complex (C, ψ_X) in $\mathbb{A}(\mathbb{Z}, X)$, with

$$C \;=\; \mathcal{C}([X] \cap - : \Delta(X, \partial X)^{n+3-*} \to \Delta(X))_{*+1} \;.$$

The cobordism class

$$s_\partial(X; \mathbb{Z}) \;=\; (C, \psi_X) \in \mathbb{S}_{n+3}(X; \mathbb{Z})$$

is the rel ∂ total homology surgery obstruction (9.11), such that $s_\partial(X; \mathbb{Z}) = 0$ if (and for $n \geq 5$ only if) $(X, \partial X)$ is homology equivalent rel ∂ to an $(n+2)$-dimensional topological manifold with boundary. The projection $X \to S^1$ is a homotopy equivalence, so that

$$\begin{aligned} \mathbb{S}_{n+3}(X; \mathbb{Z}) \;&=\; \mathbb{S}_{n+3}(S^1; \mathbb{Z}) \\ &=\; \Gamma_{n+3}\left(\begin{array}{ccc} \mathbb{Z}[z, z^{-1}] & \longrightarrow & \mathbb{Z}[z, z^{-1}] \\ \big\downarrow & & \big\downarrow \\ \mathbb{Z}[z, z^{-1}] & \longrightarrow & \mathbb{Z} \end{array} \right) \end{aligned}$$

The induced functor $\mathbb{A}(\mathbb{Z}, X) \to \mathbb{A}(\mathbb{Z}, S^1)$ sends the peripheral complex (C, ψ_X) to the kernel $\mathbb{Z}$-contractible $(n+2)$-dimensional quadratic Poincaré complex of (f, b) in $\mathbb{A}(\mathbb{Z}, S^1)$

$$\sigma_*^{S^1}(f, b) \;=\; (\Delta(f^! : \Delta(D^{n+1} \times S^1) \to \Delta(T)), \psi_b) \;.$$

The assembly functor $A : \mathbb{A}(\mathbb{Z}, S^1) \to \mathbb{A}(\mathbb{Z}[z, z^{-1}])$ sends $\sigma_*^{S^1}(f, b)$ to the $\mathbb{Z}$-contractible $(n+2)$-dimensional quadratic Poincaré complex in $\mathbb{A}(\mathbb{Z}[z, z^{-1}])$

$$A\sigma_*^{S^1}(f, b) \;=\; (\Delta(f^! : \Delta(D^{n+1} \times \mathbb{R}) \to \Delta(\overline{T})), A\psi_b) \;,$$

with $\overline{T} = f^*(D^{n+1} \times \mathbb{R})$ the canonical infinite cyclic cover of T. The total homology surgery obstruction

$$s_\partial(X;\mathbb{Z}) \; = \; A\sigma_*^{S^1}(f,b) \in \mathbb{S}_{n+3}(S^1;\mathbb{Z})$$

is a cobordism invariant of k. For $n \geq 3$ it is in fact the cobordism class of k, with $\mathbb{S}_{n+3}(S^1;\mathbb{Z}) = C_n$ the n-dimensional knot cobordism group (Ranicki [28, p.836]).

The chain homotopy type of $\sigma_*^{S^1}(f,b)$ in $\mathbb{A}(\mathbb{Z},S^1)$ is not an isotopy invariant of the n-knot k, since it depends on the choice of the map $f : T \to D^{n+1} \times S^1$ within its homotopy class. Working as in the proof of 7.3 (ii) it follows from the $(\mathbb{Z},S^1)$-module chain equivalences

$$\begin{aligned} &\Delta(T) \simeq_{(\mathbb{Z},S^1)} \Delta(f^!) \oplus \Delta(S^1) \ , \\ &\Delta(f^!)^{n+2-*} \simeq_{(\mathbb{Z},S^1)} \Delta(f^!) \end{aligned}$$

that there is defined a $\mathbb{Z}$-module chain equivalence

$$\Delta(T) \otimes_{(\mathbb{Z},S^1)} \Delta(T) \simeq_{\mathbb{Z}} (\Delta(f^!) \otimes_{(\mathbb{Z},S^1)} \Delta(f^!)) \oplus \Delta(f^!) \oplus \Delta(f^!) \oplus \Delta(S^1) \ ,$$

and that

$$\begin{aligned} H_{n+2}(T \times_{S^1} T) \; &= \; H_{n+2}(\Delta(T) \otimes_{(\mathbb{Z},S^1)} \Delta(T)) \\ &= \; H_{n+2}(\Delta(f^!) \otimes_{(\mathbb{Z},S^1)} \Delta(f^!)) \\ &= \; H_0(\mathrm{Hom}_{(\mathbb{Z},S^1)}(\Delta(f^!),\Delta(f^!))) \ . \end{aligned}$$

The following conditions are equivalent :

(a) $H_{n+2}(T \times_{S^1} T) = 0$,

(b) $\sigma_*^{S^1}(f,b)$ is chain equivalent to 0 in $\mathbb{A}(\mathbb{Z},S^1)$,

(c) $f : T \to D^{n+1} \times S^1$ is an S^1-controlled homology equivalence.

In view of 13.8 it is possible to choose f to satisfy these conditions if and only if T is an S^1-controlled homology fibration – see further below for fibred knots.

The chain homotopy type of $A\sigma_*^{S^1}(f,b)$ in $\mathbb{A}(\mathbb{Z}[z,z^{-1}])$ is an isotopy invariant of k, since it only depends on the homotopy class of $f : T \to D^{n+1} \times S^1$. Let $\zeta : \overline{T} \to \overline{T}$ be a generating covering translation of the infinite cyclic cover $\overline{T}$ of T. The quotient of $\overline{T} \times \overline{T}$ by the diagonal $\mathbb{Z}$-action

$$\overline{T} \times_{\mathbb{Z}} \overline{T} \; = \; (\overline{T} \times \overline{T})/\{(x,y) \simeq (\zeta x, \zeta y)\}$$

is such that

$$\begin{aligned} H_{n+2}(\overline{T} \times_{\mathbb{Z}} \overline{T}) \; &= \; H_{n+2}(\Delta(\overline{T}) \otimes_{\mathbb{Z}[z,z^{-1}]} \Delta(\overline{T})) \\ &= \; H_{n+2}(A\Delta(f^!) \otimes_{\mathbb{Z}[z,z^{-1}]} A\Delta(f^!)) \end{aligned}$$

The following conditions are equivalent :

(d) $H_{n+2}(\overline{T} \times_{\mathbb{Z}} \overline{T}) = 0$,

(e) $A\sigma_*^{S^1}(f,b)$ is chain equivalent to 0 in $\mathbb{A}(\mathbb{Z}[z,z^{-1}])$,

(f) $f : T \to D^{n+1} \times S^1$ is homotopic to an S^1-controlled homology equivalence.

See Ranicki [28, 7.8] for the relationship between $A\sigma_*^{S^1}(f,b)$, the Seifert form, the Alexander polynomials and the Blanchfield pairing of k. If k is simple (i.e. $H_r(\overline{T}) = 0$ for $1 \leq r \leq (n-1)/2$) and $n \geq 3$ the chain homotopy type of $A\sigma_*^{S^1}(f,b)$ is the complete isotopy invariant, by the classification results of Trotter, Levine and Kearton, and the conditions (d),(e),(f) are equivalent to k being unknotted, i.e. isotopic to the trivial n-knot $k_0 : S^n \subset S^{n+2}$.

Now suppose that $k : S^n \subset S^{n+2}$ is a fibred n-knot, i.e. that the knot complement T fibres over S^1 (cf. Remark 13.7 above). For example, the link of an isolated singular point of a complex hypersurface $f^{-1}(0) \subset \mathbb{C}^m$ $(f : \mathbb{C}^m \to \mathbb{C})$ is a fibred $(2m-3)$-knot

$$S^{2m-3} \ = \ S^{2m-1} \cap f^{-1}(0) \subset S^{2m-1} \subset \mathbb{C}^m \ ,$$

by Milnor [20] (cf. Remark 6.17 above). Let $F^{n+1} \subset S^{n+2}$ be a Seifert surface for k, with $\partial F = k(S^n)$, and let $h : F \to F$ be the monodromy. The knot complement

$$(T, \partial T) \ = \ (T(h), S^n \times S^1)$$

is the total space of a fibre bundle

$$(F^{n+1}, S^n) \to (T, \partial T) \to S^1 \ ,$$

and $f : T \to D^{n+1} \times S^1$ may be chosen to be a map of fibre bundles over S^1. The infinite cyclic cover of T is such that

$$\zeta : \overline{T} \ = \ F \times \mathbb{R} \to \overline{T} \ ; \ (x,t) \to (h(x), t+1)$$

and

$$T \times_{S^1} T \ = \ T(h \times h : F \times F \to F \times F)$$

is homotopy equivalent to

$$\overline{T} \times_{\mathbb{Z}} \overline{T} \ = \ T(h \times h) \times \mathbb{R} \ .$$

Thus

$$H_*(T \times_{S^1} T) \ = \ H_*(\overline{T} \times_{\mathbb{Z}} \overline{T})$$

and in the fibred case

$$\text{(a)} \iff \text{(b)} \iff \text{(c)} \iff \text{(d)} \iff \text{(e)} \iff \text{(f)} \ .$$

15. Other categories

Weiss [44] constructed a chain duality on the additive category of X-controlled $\mathbb{Z}$-modules, for any Δ-set X. Hutt [13] constructed a chain duality on the additive category of sheaves of $\mathbb{Z}$-modules over any space X. In principle, all the results in this paper can therefore be generalized to these categories.

References

[1] W. Browder, *Poincaré spaces, their normal fibrations and surgery*, Invent. Math. 17 (1972), 191–202.

[2] S. Buoncristiano, C. P. Rourke, and B. J. Sanderson, *A geometric approach to homology theory*, London Math. Soc. Lecture Notes 18, Cambridge University Press, Cambridge, 1976.

[3] S. E. Cappell and J. L. Shaneson, *The codimension two placement problem, and homology equivalent manifolds*, Ann. of Math. 99 (1974), 277–348.

[4] T. A. Chapman, *Approximation results in topological manifolds*, Memoirs, 251, Amer. Math. Soc., Providence R. I., 1981.

[5] E. H. Connell and J. Hollingsworth, *Geometric groups and Whitehead torsion*, Trans. Amer. Math. Soc. 140 (1969), 161–181.

[6] A. Dold and R. Thom, *Quasifaserungen und unendliche symmetrische Produkte*, Ann. of Math. 67 (1958), 239–281.

[7] D. Galewski and R. Stern, *The relationship between homology and topological manifolds via homology transversality*, Invent. Math. 39 (1977), 277–292.

[8] ———, *Classification of simplicial triangulations of topological manifolds*, Ann. of Math. 111 (1980), 1–34.

[9] D. Gottlieb, *Poincaré duality and fibrations*, Proc. Amer. Math. Soc. 76 (1979), 148–150.

[10] A. Haefliger, *Plongements différentiables dans le domaine stable*, Comment. Math. Helv. 37 (1962), 155–176.

[11] A. Hatcher and F. Quinn, *Bordism invariants of intersections of submanifolds*, Trans. Amer. Math. Soc. 200 (1974), 327–344.

[12] C. B. Hughes and A. A. Ranicki, *Ends of complexes*, Cambridge Tracts in Math. 123, Cambridge University Press, Cambridge, 1996.

[13] S. Hutt, *Poincaré sheaves on topological spaces*, preprint.

[14] W. Lück and A. A. Ranicki, *Surgery obstructions of fibre bundles*, J. Pure Appl. Algebra 81 (1992), 139–189.

[15] N. Martin and C. R. F. Maunder, *Homology cobordism bundles*, Topology 10 (1971), 93–110.

[16] C. R. F. Maunder, *Surgery on homology manifolds. I. The absolute case*, Proc. London Math. Soc. (3) 32 (1976), 480–520.

[17] C. McCrory, *A characterization of homology manifolds*, J. Lond. Math. Soc. 16 (1977), 149–159.

[18] ———, *Zeeman's filtration of homology*, Trans. Amer. Math. Soc. 250 (1979), 147–166.

[19] J. Milnor, *Whitehead torsion*, Bull. Amer. Math. Soc. 72 (1966), 358–426.

[20] ———, *Singular points of complex hypersurfaces*, Annals of Mathematics Studies 61, Princeton University Press, Princeton N. J., 1968.

[21] ——— and J. Stasheff, *Characteristic classes*, Annals of Mathematics Studies 76, Princeton University Press, Princeton N. J., 1974.

[22] F. Quinn, *Surgery on Poincaré and normal spaces*, Bull. Amer. Math. Soc. 78 (1972), 262–267.

[23] ———, *Ends of maps, II*, Invent. Math. 68 (1982), 353–424.

[24] ———, *Resolutions of homology manifolds, and the topological characterization of manifolds*, Invent. Math. 72 (1987), 267–284.

[25] ———, *Assembly maps in bordism-type theories*, Novikov Conjectures, Index Theorems and Rigidity, Vol. 1, Proc. 1993 Oberwolfach Conference (S. Ferry, A. A. Ranicki and J. Rosenberg, eds.), London Math. Soc. Lecture Notes 226, Cambridge University Press, Cambridge, 1995, 201–271.

[26] A. A. Ranicki, *The algebraic theory of surgery, I. Foundations*, Proc. London Math. Soc. (3) 40 (1980), 87–192.

[27] ———, *The algebraic theory of surgery II. Applications to topology*, Proc. London Math. Soc. (3) 40 (1980), 193–287.

[28] ———, *Exact sequences in the algebraic theory of surgery*, Math. Notes 26, Princeton University Press, Princeton N. J., 1981.

[29] ———, *Algebraic L-theory and topological manifolds*, Cambridge Tracts in Math. 102, Cambridge University Press, Cambridge, 1992.

[30] ———, *Lower K- and L-theory*, London Math. Soc. Lecture Notes, 178, Cambridge University Press, Cambridge, 1992.

[31] ——— (ed.), *The Hauptvermutung Book*, K-Monographs in Mathematics 1, Kluwer, Dordrecht, 1996, Papers on the Topology of Manifolds by A. A. Ranicki, A. J. Casson, D. P. Sullivan, M. A. Armstrong, C. P. Rourke and G. E. Cooke.

[32] ———, *45 slides on chain duality*, Surgery and Geometric Topology, Proceedings of Josai Conference, September, 1996 (A. A. Ranicki and M. Yamasaki, eds.), Science Bulletin 2 (Special Issue), Josai University, Japan, 1997, pp. 105–118. Available on WWW from
http://math.josai.ac.jp/~yamasaki/conference.html and

http://www.maths.ed.ac.uk/~aar/papers, and by anonymous FTP from math.uiuc.edu/pub/papers/K-theory/0154

[33] ______, *High-dimensional knot theory*, Springer Monographs in Mathematics, Springer, 1998.

[34] ______ and M. Weiss, *Chain complexes and assembly*, Math. Z. 204 (1990), 157–186.

[35] ______ and M. Yamasaki, *Symmetric and quadratic complexes with geometric control*, Proceedings of TGRC-KOSEF 3, 1993, pp. 139–152. Available on WWW from http://www.maths.ed.ac.uk/~aar/papers and http://math.josai.ac.jp/~yamasaki

[36] ______, *Controlled K-theory*, Topology Appl. 61 (1995), 1–59.

[37] ______, *Controlled L-theory*, Surgery and Geometric Topology, Proceedings of Josai Conference, September, 1996 (A. A. Ranicki and M. Yamasaki, eds.), Science Bulletin 2 (Special Issue), Josai University, Japan, 1997. Available on WWW from
http://math.josai.ac.jp/~yamasaki/conference.html and
http://www.maths.ed.ac.uk/~aar/papers, and by anonymous FTP from math.uiuc.edu/pub/papers/K-theory/0154, pp. 119–136.

[38] C. P. Rourke and B. J. Sanderson, *Block bundles*, Ann. of Maths. 87 (1968), I. 1–28, II. 255–277, III. 431–483.

[39] ______, *On topological neighbourhoods*, Compositio Math. 22 (1970), 387–424.

[40] E. Spanier, *Algebraic topology*, McGraw-Hill Inc., New York, 1966.

[41] M. Spivak, *Spaces satisfying Poincaré duality*, Topology 6 (1967), 77–102.

[42] C. T. C. Wall, *Poincaré complexes*, Ann. of Math. 86 (1970), 213–245.

[43] ______, *Surgery on compact manifolds*, Academic Press, New York, 1970. 2nd edition (ed. A. A. Ranicki), Mathematical Surveys and Monographs, Amer. Math. Soc., Providence R. I., 1999.

[44] M. Weiss, *Visible L-theory*, Forum Math. 4 (1992), 465–498.

[45] E. C. Zeeman, *Dihomology III. A generalization of the Poincaré duality for manifolds*, Proc. London Math. Soc. (3) 13 (1963), 155–183.

Andrew Ranicki
Department of Mathematics
and Statistics
University of Edinburgh
Edinburgh EH9 3JZ
Scotland, UK
a.ranicki@edinburgh.ac.uk

On a Conjecture of Izhboldin on Similarity of Quadratic Forms

Detlev W. Hoffmann[1]

Received: January 4, 1999

Communicated by Ulf Rehmann

Abstract. In his paper *Motivic equivalence of quadratic forms*, Izhboldin modifies a conjecture of Lam and asks whether two quadratic forms, each of which isomorphic to the product of an Albert form and a k-fold Pfister form, are similar provided they are equivalent modulo I^{k+3}. We relate this conjecture to another conjecture on the dimensions of anisotropic forms in I^{k+3}. As a consequence, we obtain that Izhboldin's conjecture is true for $k \leq 1$.

1991 Mathematics Subject Classification: Primary 11E81; Secondary 11E04.
Keywords and Phrases: Quadratic form, Pfister form, Albert form, similarity of quadratic forms.

In what follows, we will adhere to the same terminology and notations used in Izhboldin's article [I] mentioned in the abstract. In particular, if two quadratic forms ϕ and ψ are similar, we will write $\phi \sim \psi$.
Let F be a field of characteristic $\neq 2$. Recall that an Albert form α over F is a 6-dimensional quadratic form over F with signed discriminant $1 \in F^*/F^{*2}$ (i.e. $\alpha \in I^2F$), and an n-fold Pfister form over F is a form of type $\langle\langle a_1, \cdots, a_n\rangle\rangle := \langle 1, -a_1\rangle \otimes \cdots \otimes \langle 1, -a_n\rangle$, $a_i \in F^*$. In his paper [I], Izhboldin states the following conjecture:

Conjecture 1 (Cf. Conjecture 5.1 in [I].) *Let q_1 and q_2 be Albert forms over F and let π_1 and π_2 be two k-fold Pfister forms over F ($k \geq 0$) such that $q_i \otimes \pi_i$, $i = 1, 2$ is anisotropic and $q_1 \otimes \pi_1 \equiv q_2 \otimes \pi_2 \bmod I^{k+3}F$. Then $q_1 \otimes \pi_1 \sim q_2 \otimes \pi_2$.*

[1] This research has been carried out in the framework of TMR Network ERB FMRX CT-97-0107 "Algebraic K-Theory, Linear Algebraic Groups and Related Structures."

In fact, this conjecture is a special case of a question asked by Lam [L, (6.6)]. Lam's original question was as follows. Suppose σ_i, $\rho_i \in P_nF$, $i=1,2$, and let $\phi_i = (\sigma_i \perp -\rho_i)_{\mathrm{an}}$ be the anisotropic part of $\sigma_i \perp -\rho_i$. If $\phi_1 \equiv \phi_2 \bmod I^{n+1}F$, does it then follow that $\phi_1 \sim \phi_2$? By a result of Elman and Lam [EL, Theorem 4.5], it is known that $\dim \phi_i \in \{2^{n+1}-2^m, 1 \leq m \leq n+1\}$, and that if $\dim \phi_i = 2^{n+1}-2^m$, then ρ_i and σ_i are $(m-1)$-linked, i.e. there exists an $(m-1)$-fold Pfister form which divides both ρ_i and σ_i. It is an easy exercise to show that Lam's question has a positive answer if $\dim \phi_1$ (or $\dim \phi_2$) equals 0 of 2^n (i.e. $m=n+1$ or $m=n$). In [I, Section 4], Izhboldin constructs counterexamples with $\dim \phi_1$ (or $\dim \phi_2$) equal to $2^{n+1}-2^m$ with $1 \leq m \leq n-2$. The only remaining case $m=n-1$ boils down to Conjecture 1 above.
It turns out that this conjecture would have a positive answer if another well-known conjecture on quadratic forms were true, this other conjecture being

Conjecture 2 *Let $n \geq 2$ and let q be an anisotropic form in I^nF. If $\dim q > 2^n$ then $\dim q \geq 2^n + 2^{n-1}$.*

Proposition 1 *Conjecture 2 for $n = k+3$ implies Conjecture 1 for k.*

It was shown in [H 2] that Conjecture 2 holds for $n \leq 4$. As a consequence, we have

Corollary *Conjecture 1 holds for $k \leq 1$.*

Note that for $k=0$ this is essentially Jacobson's theorem saying that two Albert forms are similar if and only if their associated biquaternion algebras are isomorphic (see [MS] for a quadratic form-theoretic proof of Jacobson's theorem).

Proof of Proposition 1. Suppose that Conjecture 2 holds for $k+3$. Let q_1 and q_2 be Albert forms over F and let π_1 and π_2 be two k-fold Pfister forms over F ($k \geq 0$) such that $q_1 \otimes \pi_1 \equiv q_2 \otimes \pi_2 \bmod I^{k+3}F$ and such that $q_i \otimes \pi_i$ is anisotropic for $i=1,2$.
First, we note that we may assume $\pi_1 = \pi_2$ (cf. the remarks following Conjecture 5.1 in [I]). We denote this k-fold Pfister form by π. Since $q_i \otimes \pi \in I^{k+2}F$, we can scale q_i (and thus $q_i \otimes \pi$) without changing the equivalence $\bmod I^{k+3}F$, and we may thus assume that $q_i \cong \langle 1 \rangle \perp q_i'$, $\dim q_i' = 5$ for $i=1,2$. This yields $q_1' \otimes \pi \equiv q_2' \otimes \pi \bmod I^{k+3}F$.
In particular, $\pi \otimes (q_1' \perp -q_2')$ is a form of dimension $2^k(2^3+2) = 2^{k+3}+2^{k+1}$ in $I^{k+3}F$. By Conjecture 2, $\pi \otimes (q_1' \perp -q_2')$ is isotropic. In particular, there exists $x \in F^*$ such that x is represented by both $\pi \otimes q_1'$ and $\pi \otimes q_2'$. Using the multiplicativity of Pfister forms (cf. [EL, Theorem 1.4]), there exist 4-dimensional forms q_i'', $i=1,2$, such that $\pi \otimes q_i' \cong \pi \otimes (\langle x \rangle \perp q_i'')$.
From this, it follows readily that $\pi \otimes q_1'' \equiv \pi \otimes q_2'' \bmod I^{k+3}F$. Note that $\dim(\pi \otimes q_i'') = 2^{k+2}$, so that $\pi \otimes q_1''$ and $\pi \otimes q_2''$ are (anisotropic) half-neighbors. As a consequence, $\pi \otimes q_1''$ becomes isotropic over the function field of $\pi \otimes q_2''$ (see, e.g., [H 3, Corollary 2.6] or [I, Lemma 3.3]). By [H 1, Theorem 1.4], this

implies that $\pi \otimes q_1''$ and $\pi \otimes q_2''$ are similar, so that there exists some $y \in F^*$ such that $\pi \otimes q_1'' \cong y\pi \otimes q_2''$. Thus, we obtain

$$\begin{array}{rcll} \pi \otimes q_1 & \equiv & \pi \otimes \langle 1, x \rangle \perp \pi \otimes q_1'' & \mathrm{mod} I^{k+3}F \\ & \equiv & \pi \otimes q_2 & \mathrm{mod} I^{k+3}F \\ & \equiv & y\pi \otimes q_2 & \mathrm{mod} I^{k+3}F \\ & \equiv & y\pi \otimes \langle 1, x \rangle \perp y\pi \otimes q_2'' & \mathrm{mod} I^{k+3}F \\ & \equiv & y\pi \otimes \langle 1, x \rangle \perp \pi \otimes q_1'' & \mathrm{mod} I^{k+3}F \end{array}$$

and hence $\pi \otimes \langle 1, x \rangle \equiv y\pi \otimes \langle 1, x \rangle \bmod I^{k+3}F$. Now $\dim(\pi \otimes \langle 1, x \rangle) = 2^{k+1}$, and the Arason-Pfister Hauptsatz therefore implies that $\pi \otimes \langle 1, x \rangle \cong y\pi \otimes \langle 1, x \rangle$. We conclude that

$$\begin{array}{rcl} \pi \otimes q_1 & \cong & \pi \otimes \langle 1, x \rangle \perp \pi \otimes q_1'' \\ & \cong & y\pi \otimes \langle 1, x \rangle \perp y\pi \otimes q_2'' \\ & \cong & y\pi \otimes q_2 \ . \end{array}$$

□

Note that we didn't really make use of the fact that q_1 and q_2 are Albert forms. However, it is not difficult to show that if π is a k-fold Pfister form and $q = q' \perp \langle a \rangle \in IF$ such that $\pi \otimes q \in I^{k+2}F$, then if one chooses $b \in F^*$ such that $\tilde{q} = q' \perp \langle b \rangle \in I^2F$, one has $\pi \otimes q \cong \pi \otimes \tilde{q}$. So what is essential is the fact that $\pi \otimes q_i$ is in $I^{k+2}F$, in which case we may as well assume by what we just mentioned that q_i is an Albert form.
In the proof of Conjecture 2 for $n = 4$ in [H 2], one makes use of a certain property PD_2. It turns out that this property can be used to establish Conjecture 1 for $k = 1$ without invoking Conjecture 2 for $n = 4$. Let us recall the general definition of property PD_n.

Definition Let n be an integer ≥ 1. The field F is said to have the Pfister decomposition property for Pfister forms of fold $\leq n$, PD_n for short, if for each m $(1 \leq m \leq n)$, for each anisotropic $\pi \in P_{m-1}F$, for each $r \in \dot{F}$, and each anisotropic $\varphi \in \pi WF$, there exist forms σ and τ over F such that for $\rho := \pi \otimes \langle\langle r \rangle\rangle$ one has $\varphi \cong \pi \otimes \sigma \perp \rho \otimes \tau$ and $(\varphi_{F(\rho)})_{\mathrm{an}} \cong (\pi \otimes \sigma)_{F(\rho)}$.

Proposition 2 *Suppose that F has PD_n for some $n \geq 1$. Then Conjecture 1 holds for $k = n - 1$.*

Proof. Suppose that F has PD_n for $n = k + 1$. As in the previous proof, we may assume that we are in the situation where $\pi \otimes q_1 \equiv \pi \otimes q_2 \bmod I^{k+3}F$ with Albert forms q_i, $i = 1, 2$, a k-fold Pfister form π and with $\pi \otimes q_i$ being anisotropic for $i = 1, 2$. After scaling, we may assume that $q_1 \cong \langle 1, -r \rangle \perp q_1'$ for some $r \in F^*$. It follows that $\pi \otimes q_1$ contains the subform $\rho = \pi \otimes \langle\langle r \rangle\rangle$.
In particular, $\pi \otimes q_1$ becomes isotropic over the function field $F(\rho)$, and thus $\pi \otimes q_2$ also becomes isotropic over $F(\rho)$ (cf. [I, Theorem 4.3]). Property PD_{k+1} then implies that $\pi \otimes q_2$ contains a subform similar to ρ, and since we may scale

$\pi \otimes q_2 \in I^{k+2}F$ without changing the equivalence $\bmod I^{k+3}F$, we may assume that $\pi \otimes q_2 \cong \pi \otimes (\langle 1, -r\rangle \perp q_2')$ for some 4-dimensional form q_2'.
It follows that $\pi \otimes q_1' \equiv \pi \otimes q_2' \bmod I^{k+3}F$. As in the proof of Proposition 1, this implies that $\pi \otimes q_1'$ and $\pi \otimes q_2'$ are similar, and thus that $\pi \otimes q_1$ and $\pi \otimes q_2$ are also similar. □

It was proved by Rost that each field has property PD_2 (see [H 2, Lemma 2.6]). Again, we can conclude that Conjecture 1 holds for $k \leq 1$, this time by invoking PD_2.
In the case $n \geq 3$, we do not know whether PD_n holds for all fields nor whether PD_n for a field F implies that Conjecture 2 holds for F for $n+2$ (or vice versa).

References

[EL] Elman, R.; Lam, T.Y.: *Pfister forms and K-theory of fields.* J. Algebra 23 (1972), 181–213.

[H 1] Hoffmann, D.W.: *On quadratic forms of height* 2 *and a theorem of Wadsworth.* Trans. Amer. Math. Soc. 348 (1996), 3267–3281.

[H 2] Hoffmann, D.W.: *On the dimensions of anisotropic quadratic forms in I^4.* Invent. Math. 131 (1998), 185–198.

[H 3] Hoffmann, D.W.: *Similarity of quadratic forms and half-neighbors.* J. Algebra 204 (1998), 255–280.

[I] Izhboldin, O.T.: *Motivic equivalence of quadratic forms.* Doc. Math. J. DMV 3 (1998), 341–351.

[L] Lam, T.Y.: *Fields of u-invariant* 6 *after A. Merkurjev.* Israel Math. Conf. Proc. Vol. 1: Ring Theory 1989 (in honor of S.A. Amitsur) (ed. L. Rowen). pp.12–31. Jerusalem: Weizmann Science Press 1989.

[MS] Mammone, P.; Shapiro, D.B.: *The Albert quadratic form for an algebra of degree four.* Proc. Amer. Math. Soc. 105 (1989), 525–530.

Detlev W. Hoffmann
Équipe de Mathématiques
UMR 6623 du CNRS
Université de Franche-Comté
16, route de Gray
25030 Besançon Cedex, France
detlev@math.univ-fcomte.fr

The Local Monodromy as a Generalized Algebraic Correspondence

Caterina Consani [1]

with an Appendix by
Spencer Bloch

Received: May 25, 1998
Revised: February 24, 1999

Communicated by Peter Schneider

Abstract. For an algebraic, normal-crossings degeneration over a local field the local monodromy operator and its powers naturally define Galois equivariant classes in the ℓ-adic (middle dimensional) cohomology groups of some precise strata of the special fiber of a normal-crossings model associated to the fiber product degeneration. The paper addresses the question whether these classes are algebraic. It is shown that the answer is positive for any degeneration whose special fiber has (locally) at worst triple points singularities. These algebraic cycles are responsible for and they explain geometrically the presence of poles of local Euler L-factors at integers on the left of the left-central point.

1991 Mathematics Subject Classification: 11G25, 11S40, 14C25, 14E10

Introduction

Let X be a proper and smooth variety over a local field K and let $\mathcal{X}$ be a regular model of X defined over the ring of integers $\mathcal{O}_K$ of K. When $\mathcal{X}$ is smooth over $\mathcal{O}_K$, the Tate conjecture equates the ℓ–adic Chow groups of algebraic cycles on the geometric special fiber $X_{\bar{k}}$ of $\mathcal{X} \to \mathrm{Spec}(\mathcal{O}_K)$ with the Galois invariants in $H^{2*}(X_{\bar{K}}, \mathbf{Q}_\ell(*))$. One of the results proved in [2] (*cf.* Corollary 3.6) shows that the Tate conjecture for smooth and proper varieties over finite fields together with the monodromy–weight conjecture imply a generalization of the above result in the case of semistable reduction. Namely, let $\wp \in \mathrm{Spec}(\mathcal{O}_K)$ be a

[1] Partially supported by the NSF grant DMS-9701302

prime over which the special fiber $\mathcal{X}\times\mathrm{Spec}(k(\wp)) = Y$ is a reduced divisor with normal crossings in $\mathcal{X}$ (*i.e.* semistable fiber). Then, assuming the above two conjectures, the ℓ–adic groups of algebraic cycles modulo rational equivalence on the r–fold intersections of components of Y ($r \geq 1$) are related with Galois invariant classes on the Tate twists $H^{2*-(r-1)}(X_{\bar{K}}, \mathbf{Q}_\ell(* - (r-1)))$.
An interesting case is when one replaces X by $X \times_K X$, so that Galois invariant cycles may be identified with Galois equivariant maps $H^*(X_{\bar{K}}, \mathbf{Q}_\ell) \to H^*(X_{\bar{K}}, \mathbf{Q}_\ell(\cdot))$. Examples of such maps are the powers N^i of the logarithm of the local monodromy around $\wp$. The operators $N^i : H^*(X_{\bar{K}}, \mathbf{Q}_\ell) \to H^*(X_{\bar{K}}, \mathbf{Q}_\ell(-i))$ determine classes $[N^i] \in H^{2d}((X \times X)_{\bar{K}}, \mathbf{Q}_\ell(d-i))$ ($d = \dim\ X_{\bar{K}}$) invariant under the decomposition group. In this paper we study in detail the structure of $[N^i]$ when the special fiber Y of $\mathcal{X}$ has at worst triple points as singularities. That is, we exhibit the corresponding algebraic cycles on the (normal crossings) special fiber $T = \cup_i T_i$ of a resolution $\mathcal{Z}$ of $\mathcal{X} \times_{\mathcal{O}_K} \mathcal{X}$. Denote by $\tilde{N} = 1 \otimes N + N \otimes 1$ the monodromy on the product, and let F be the geometric Frobenius. Then the classes $[N^i]$ naturally determine elements in $\mathrm{Ker}(\tilde{N}) \cap H^{2d}((X\times X)_{\bar{K}}, \mathbf{Q}_\ell(d-i))^{F=1}$. Assuming the monodromy–weight conjecture on the product (*i.e.* the monodromy filtration $L.$ on $H^*((X \times X)_{\bar{K}}, \mathbf{Q}_\ell)$ coincides–up to a shift–with the filtration by the weights of the Frobenius *cf.* [16]) and the semisimplicity of the action of the Frobenius on the inertia invariants, the following identifications hold

$$(0.1)\quad \mathrm{Ker}(\tilde{N}) \cap H^{2d}((X \times X)_{\bar{K}}, \mathbf{Q}_\ell(d-i))^{F=1}$$
$$\simeq \left((gr^L_{2(d-i)} H^{2d}(T, \mathbf{Q}_\ell))(d-i) \right)^{F=1}$$
$$\simeq$$
$$\left[\frac{\mathrm{Ker}(\rho^{(2(i+1))} : H^{2(d-i)}(\tilde{T}^{(2i+1)}, \mathbf{Q}_\ell)(d-i) \to H^{2(d-i)}(\tilde{T}^{(2(i+1))}, \mathbf{Q}_\ell)(d-i))}{\mathrm{Image}\ \rho^{(2i+1)}}\right]^{F=1}.$$

Here $\tilde{T}^{(j)}$ denotes the normalization of the j–fold intersection on the closed fiber T. These isomorphisms show that the classes $[N^i]$ have representatives in the cohomology groups of some precise strata of T. Moreover, the Tate conjecture and the semisimplicity of the action of the Frobenius on the smooth schemes $\tilde{T}^{(j)}$ would imply that these classes are algebraic. We refer to § 1, (1.6) for the description of the restriction maps ρ in (0.1).
To better understand the geometry related to the desingulatization process $\mathcal{Z} \to \mathcal{X} \times_{\mathcal{O}_K} \mathcal{X}$, and to avoid at first, some technical complications connected to the theory of the nearby cycles in mixed characteristic, we start by investigating this problem in equal characteristic zero (*i.e.* for semistable degenerations over a disk). There, one can take full advantage of many geometric results based on the theory of the mixed Hodge structures. Under the assumption of the monodromy–weight conjecture and using some techniques of [16], our results generalize to mixed characteristic. The cycles we exhibit on $\tilde{T}^{(2i+1)}$ explain geometrically the presence of poles on specific local factors of the L–function

related to the fiber product $X \times X$. In fact, theorem 6.2 equates, under the assumption of the semisimplicity of the action of the Frobenius F on the inertia invariants $H^*((X \times X)_{\bar{K}}, \mathbf{Q}_\ell)^I$, the rank of any of the groups in (0.1) with $\mathrm{ord}_{s=d-i} \det(Id - FN(\wp)^{-s}|H^{2d}((X \times X)_{\bar{K}}, \mathbf{Q}_\ell)^I)$. Here, $N(\wp)$ denotes the number of elements of the finite field $k(\wp)$.

A study of the local geometry of the normal–crossings special fiber T shows that $[N^i]$ are represented by certain natural "diagonal cycles" on $\tilde{T}^{(2i+1)}$ together with a cycle supported on the exceptional part of the stratum that arises because the classes $[N^i]$ must belong to the kernel of the restriction map $\rho^{(2(i+1))}$ (*cf.* (0.1)). This result is obtained via the introduction of a generalized correspondence diagram for the map

$$N^i : \mathbf{H}^*(Y, gr^L_{r+i}\mathbf{R}\Psi(\mathbf{Q}_{\mathcal{X}})) \to \mathbf{H}^*(Y, (gr^L_{r-i}\mathbf{R}\Psi(\mathbf{Q}_{\mathcal{X}}))(-i)). \tag{0.2}$$

This morphism describes the monodromy action on the E_1–term of the spectral sequence of weights for the filtered complex of the nearby cycles $(\mathbf{R}\Psi(\mathbf{Q}_{\mathcal{X}}), L.)$ (*cf.* § 2, (2.1)). For $i > 0$, the classes $[N^i]$ do not describe an algebraic correspondence in the classical sense. In fact, the algebraic cycles representing them are only supported on higher strata of the special fiber T (*i.e.* on $\tilde{T}^{(2i+1)}$) and they do not naturally determine classes in the cohomology of T. This is a consequence of the fact that for $i > 0$, the cocycle $[N^i]$ does not have weight zero in the ℓ–adic cohomology of the fiber product $(X \times X)_{\bar{K}}$, as one can easily check from (0.1). Nonetheless, we expect that each of these classes supplies a refined information on the degeneration. Namely, we conjecture that the geometric description that we obtain up to triple points can be generalized to any kind of semistable singularity via a thorough combinatoric study of the toric singularities of the special fiber of the fiber product resolution $\mathcal{Z}$.

The correspondence diagram related to the map (0.2) is built up from the hypercohomology of the Steenbrink filtered resolution $(A^\bullet_{\mathcal{X}}, L.)$ of $\mathbf{R}\Psi(\mathbf{Q}_{\mathcal{X}})$. In § 3 we establish the necessary functoriality properties of the Steenbrink complex and its $L.$–filtration. A difficult point in the description of the correspondence diagram is related to the definition of a product structure on the E_1–terms of the spectral sequence of weights. Example 3.1 points out a problem related to a canonical definition of a product structure for $(A^\bullet_{\mathcal{X}}, L.)$ in the filtered category. It comes out that the monodromy filtration $L.$ is not multiplicative on the level of the filtered complexes. A partial product, canonical only on the $E_2 = E_\infty$-terms is provided in the Appendix. This suffices for purposes of our paper.

Acknowledgments. I would like to thank Kazuya Kato for suggesting the study of the local monodromy as a Galois invariant class, eventually algebraic. I also would like to acknowledge an interesting conversation with Alexander Beilinson on some of the arguments presented here and few interesting remarks received from Ofer Gabber. I gratefully thank the Institut des Hautes Études Scientifiques for the kind hospitality received during my stay there in January 1997 and the University of Cambridge where a revision of this work was carried out in Winter 1998. Finally, it is my pleasure to thank Spencer Bloch for his constant support and for many fruitful suggestions I received from him during the preparation of this paper.

1. Notations and techniques from mixed Hodge theory

In this paragraph we introduce the main notations and recall some results on the mixed Hodge theory of a degeneration.
We denote by X a connected, smooth, complex analytic manifold and we let S be the unit disk. We write $f : X \to S$ for a proper, surjective morphism and we let $Y = f^{-1}(0)$ be its special fiber. We assume that f is *smooth* at every point of $X^* = X \setminus Y$ and that the special fiber Y is an algebraic divisor with *normal–crossings*. The local description of f near a closed point $y \in Y$ is given by:

$$f(z_1, \ldots, z_m) = z_1^{e_1} \cdots z_k^{e_k}$$

for $k \leq m = \dim X$ and $\{z_1, \ldots, z_m\}$ a local coordinate system on a neighborhood of y in X centered at y and $e_i \in \mathbf{Z},\ e_i \geq 1$. The fibers of f have then dimension $d = m - 1$.
A normal–crossings divisor as above is said to have *semistable* reduction (strict normal–crossings) if one has: $e_i = 1\ \forall i$, in the local description of f.
We *fix a parameter* $t \in S$. For $t \neq 0$, let $f^{-1}(t) = X_t$ be the fiber at t. Because the restriction of f at $S^* = S \setminus \{0\}$ is a C^∞, locally trivial fiber bundle, the positive generator of $\pi_1(S^*, t) \simeq \mathbf{Z}$ induces an automorphism T_t of $H^*(X_t, \mathbf{Z})$, called *the local monodromy*. We will always suppose throughout the paper that T_t is *unipotent*. This assumption, together with the local monodromy theorem (*cf.* [7], Theorem 2.1.2), implies that $(T_t - 1)^{i+1} = 0$, on $H^i(X_t, \mathbf{Z})$. The unipotency condition of the local monodromy is for example verified when g.c.d.$(e_i,\ i \in [1, k]) = 1$, $\forall y \in Y$ (*cf. op.cit.*). Under these conditions, the logarithm of the local monodromy is defined to be the finite sum:

$$N_t := \log T_t = (T_t - 1) - \frac{1}{2}(T_t - 1)^2 + \frac{1}{3}(T_t - 1)^3 - \cdots$$

It is known (*cf.* [5]) that the automorphisms T_t of $H^i(X_t, \mathbf{C})$ $(t \in S^*)$, are the fibers of an automorphism T of the fiber bundle $\mathbf{R}^i f_*(\Omega^\bullet_{X/S}(\log Y))$ over S, whose fiber at 0 is described as $T_0 = \exp(-2\pi i N_0)$. By definition, the endomorphism N_0 is the residue at 0 of the Gauss-Manin connection ∇ on the "canonical prolongation" $\mathbf{R}^i f_*(\Omega^\bullet_{X/S}(\log Y))$ of the locally free sheaf $\mathbf{R}^i f_*(\Omega^\bullet_{X^*/S^*})$. Because of the definition of T_0, it makes sense to think of a nilpotent map $N := -\frac{1}{2\pi i} \log T$ as the monodromy operator on the degeneration $f : X \to S$. Via the canonical isomorphism (*cf.* [11], Thm. 2.18)$(t \in S)$:

$$\mathbf{R}^i f_*(\Omega^\bullet_{X/S}(\log Y)) \otimes_{\mathcal{O}_S} k(t) \xrightarrow{\simeq} \mathbf{H}^i(X_t, \Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_{X_t})$$

where $k(t)$ is the residue field of $\mathcal{O}_S$ at t, we can see the map N_0 as an endomorphism of the hypercohomology of the relative de Rham complex $\Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_Y$. This complex represents in the derived category $D^+(Y, \mathbf{C})$ of the abelian category of sheaves of $\mathbf{C}$–vector spaces on Y, the complex of the nearby cycles $\mathbf{R}\Psi(\mathbf{C})$. Namely, there exists a non-canonical quasi-isomorphism (*i.e.* depending on the choice of the parameter t on S) $\Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_Y \simeq \mathbf{R}\Psi(\mathbf{C}_{\tilde{X}^*}) := i^{-1}\mathbf{R}k_*\mathbf{C}_{\tilde{X}^*}$ (*cf.* [11], § 2). This isomorphism, composed with the canonical map $\Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_Y \to \Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_{Y^{\mathrm{red}}}$ (Y^{red} = reduced, induced structure scheme on Y), induces a quasi-isomorphism $(i^{-1}\mathbf{R}k_*\mathbf{C}_{\tilde{X}^*})_{\mathrm{un}} \simeq \Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_{Y^{\mathrm{red}}}$ (*cf. op.cit.* § 4). Here, we denote by $(i^{-1}\mathbf{R}k_*\mathbf{C}_{\tilde{X}^*})_{\mathrm{un}}$ the maximal subobject of $i^{-1}\mathbf{R}k_*\mathbf{C}_{\tilde{X}^*}$ on which $\pi_1(S^*)$ acts with unipotent automorphisms. We refer to the following commutative diagram for the description of the maps:

$$\begin{array}{ccccc}
\tilde{X}^* & \xrightarrow{k} & X & \xleftarrow{i} & Y \\
\downarrow & & \downarrow f & & \downarrow \\
\tilde{S}^* & \xrightarrow{p} & S & \longleftarrow & \{0\}.
\end{array}$$

The space $\tilde{S}^* = \{u \in \mathbf{C} \mid \mathrm{Im}\ u > 0\}$ is the upper half plane, the map $p : \tilde{S}^* \to S$ $p(u) = \exp(2\pi i u) = t$, makes $\tilde{S}^*$ in a universal covering of S^* and $\tilde{X}^*$ is the pullback $X \times_S \tilde{S}^*$ of X along p. The morphism k is the natural projection. It factorizes through X^* by means of the injection $j : X^* \to X$. Finally, i is the closed embedding.

Steenbrink, Guillen and Navarro Aznar and Masaiko Saito (*cf.* [11], [6], [12]) defined a mixed Hodge structure on the hypercohomology of the unipotent factor of the complex of the nearby cycles $\mathbf{H}^*(X, \Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_{Y^{\mathrm{red}}})$. This is frequently referred as the limiting mixed Hodge structure.

We will assume from now on that f is *projective.* Then, the weight filtration on the limiting mixed Hodge structure is the one induced by the nilpotent endomorphism N, namely by the logarithm of the unipotent Picard-Lefschetz transformation T that is already defined at the $\mathbf{Q}$-level. This filtration, which one usually refers to as the monodromy–weight filtration $L.$, is defined inductively. On the limiting cohomology $H^i(\tilde{X}^*, \mathbf{Q})$, it is increasing and has lenght at most $2i$. By the local monodromy theorem $N^{i+1} = 0$, hence one sets $L_0 = \mathrm{Im}\ N^i$ and $L_{2i-1} = \mathrm{Ker}\ N^i$. The monodromy filtration $L.$ becomes a convolution product of the kernel and the image filtration relative to the endomorphism N. These filtrations are defined as

$$K_l\ H^i(\tilde{X}^*, \mathbf{Q}) := \mathrm{Ker}\ N^{l+1}, \qquad I^j\ H^i(\tilde{X}^*, \mathbf{Q}) := \mathrm{Im}\ N^j$$

and their convolution is

$$L = K * I, \qquad L_k := \sum_{l-j=k} K_l \cap I^j. \tag{1.1}$$

It is a very interesting fact that there is no explicit construction of the monodromy-weight filtration $L.$ on $\Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_Y$ itself. The filtration $L.$ is defined on a complex $A^\bullet_{\mathbf{C}}$ which is a resolution of $\Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_{Y^{\mathrm{red}}}$. More precisely, the complex $\Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_{Y^{\mathrm{red}}}$ is isomorphic, in the derived category $D^+(Y, \mathbf{C})$, to the complex $A^\bullet_{\mathbf{C}}$ of $\mathcal{O}_X$–modules supported on Y. The complex $A^\bullet_{\mathbf{C}}$ is the simple complex associated to the double complex ($p,\ q \geq 0$):

$$A^{p,q}_{\mathbf{C}} := \Omega^{p+q+1}_X(\log Y)/W_q\Omega^{p+q+1}_X(\log Y)$$

where $W_*\Omega^\bullet_X(\log Y)$ is the weight filtration by the order of log-poles (*cf.* [3], § 3). The differentials on it are defined as follows

$$d' : A^{p,q}_{\mathbf{C}} \to A^{p+1,q}_{\mathbf{C}}, \qquad d'(\omega) = d\omega$$

is induced by the differentiation on the complex $\Omega^\bullet_X(\log Y)$ and

$$d'' : A^{p,q}_{\mathbf{C}} \to A^{p,q+1}_{\mathbf{C}}, \qquad d''(\omega) = \theta \wedge \omega$$

where $\theta := f^*(\frac{dt}{t}) = \sum_{i=1}^k e_i \frac{dz_i}{z_i}$ is the form defining the quasi-isomorphism we mentioned before (*cf.* [11], § 4)

$$\Omega^\bullet_{X/S}(\log Y) \otimes_{\mathcal{O}_X} \mathcal{O}_{Y^{\mathrm{red}}} \stackrel{\wedge\,\theta}{\to} A^\bullet_{\mathbf{C}}.$$

The total differential on $A^\bullet_{\mathbf{C}}$ is $d = d' + d''$. The weight filtration $W_*\Omega^\bullet_X(\log Y)$ induces a corresponding filtration on $A^\bullet_{\mathbf{C}}$ ($r \in \mathbf{Z}$):

$$(1.2) \qquad W_r A^{p,q}_{X,\mathbf{C}} =: W_{r+q+1}\Omega^{p+q+1}_X(\log Y)/W_q\Omega^{p+q+1}_X(\log Y).$$

The filtration that $W_r A^\bullet_{\mathbf{C}}$ induces on $\mathbf{H}^*(Y, A^\bullet_{\mathbf{C}}) \simeq \mathbf{H}^*(\tilde{X}^*, \mathbf{C})$ is the kernel filtration K (*cf.* (1.1))

$$K_r H^*(\tilde{X}^*, \mathbf{C}) = W_r\mathbf{H}^*(Y, A^\bullet_{\mathbf{C}}) =: \mathrm{Im}\Big(\mathbf{H}^*(Y, W_r A^\bullet_{\mathbf{C}}) \to \mathbf{H}^*(Y, A^\bullet_{\mathbf{C}})\Big) = \mathrm{Ker}\ N^{r+1}.$$

The monodromy-weight filtration is then defined as

$$L_r A^{p,q} := W_{2q+r+1}\Omega^{p+q+1}_X(\log Y)/W_q\Omega^{p+q+1}_X(\log Y).$$

Via Poincaré residues, the related graded pieces have the following description

$$(1.3) \qquad gr^L_r A^\bullet_{\mathbf{C}} \simeq \bigoplus_{k \geq \max(0,-r)} (a_{2k+r+1})_*\Omega^\bullet_{\tilde{Y}^{(2k+r+1)}}[-r-2k].$$

Here, we have denoted by $\tilde{Y}^{(m)}$ the disjoint union of all intersections $Y_{i_1} \cap \ldots \cap Y_{i_m}$ for $1 \leq i_1 < \ldots < i_m \leq n$ ($Y = Y_1 \cup \ldots \cup Y_n$). We write $(a_m)_* : \tilde{Y}^{(m)} \to X$ for the natural projection.
The monodromy operator N is induced by an endomorphism $\tilde{\nu}$ of $A^\bullet_{\mathbf{C}}$ which is defined as $(-1)^{p+q+1}$ times the natural projection

$$\nu : A^{p,q}_{\mathbf{C}} \to A^{p-1,q+1}_{\mathbf{C}}.$$

The endomorphism $\tilde{\nu}$ is characterized by its behavior on the L-filtration, namely

$$\tilde{\nu}(L_r A^\bullet_{\mathbf{C}}) \subset L_{r-2} A^\bullet_{\mathbf{C}}$$

and the induced map

$$\tilde{\nu}^r : gr^L_r A^\bullet_{\mathbf{C}} \to gr^L_{-r} A^\bullet_{\mathbf{C}} \tag{1.4}$$

is an isomorphism for all $r \geq 0$. The complex $A^\bullet_{\mathbf{C}}$ contains the subcomplex $W_0 A^\bullet_{\mathbf{C}} = \mathrm{Ker}(\tilde{\nu})$ that is known to be a resolution of $\mathbf{C}_Y$. The filtration L and the Hodge filtration F on $A^\bullet_{\mathbf{C}}$ induce *resp.* the kernel and F filtration on $W_0 A^\bullet_{\mathbf{C}}$. The resulting mixed Hodge structure on $H^*(Y, \mathbf{C})$ is the canonical one. Similarly, the homology $H_*(Y, \mathbf{C})$ (*i.e.* $H^*_Y(X, \mathbf{C})$) with its mixed Hodge structure is calculated by the hypercohomology of the complex $\mathrm{Coker}(\tilde{\nu})$.
Because of the description given in (1.3), the spectral sequence of hypercohomology of the filtered complex $(A^\bullet_{\mathbf{C}}, L)$ (frequently referred as the *weight spectral sequence* of $\mathbf{R}\Psi(\mathbf{C})$) has the E_1 term given by

$$\begin{aligned} E_1^{-r,n+r} &= \bigoplus_{k \geq \max(0,-r)} H^{n-r-2k}(\tilde{Y}^{(2k+r+1)}, \mathbf{C}) \\ d_1 &= \sum_k ((-1)^{r+k} d_1' + (-1)^{k-r} d_1''). \end{aligned} \tag{1.5}$$

The explicit definition of the differentials, in the strict normal–crossings case (*i.e.* semistable degeneration), is the following:

$$\begin{aligned} d_1' &= \rho^{(r+2k+2)} = \sum_{u=1}^{r+2k+2} (-1)^{u-1} \rho_u^{(r+2k+2)} \\ d_1'' &= -\gamma^{(r+2k+1)} = \sum_{u=1}^{r+2k+1} (-1)^u \gamma_u^{(r+2k+1)} \end{aligned} \tag{1.6}$$

where

$$\rho_u^{(r+2k+2)} = (\delta_u^{(r+2k+2)})^* : H^{n-r-2k}(\tilde{Y}^{(2k+r+1)}, \mathbf{C}) \to H^{n-r-2k}(\tilde{Y}^{(2k+r+2)}, \mathbf{C})$$
$$\gamma_u^{(r+2k+1)} = (\delta_u^{(r+2k+1)})_! : H^{n-r-2k}(\tilde{Y}^{(2k+r+1)}, \mathbf{C}) \to H^{n-r-2k+2}(\tilde{Y}^{(2k+r)}, \mathbf{C})$$

are the restrictions, *resp.* the Gysin maps, induced by the inclusions ($u, t \in \mathbf{Z}$)

$$\delta_u^{(t)} : Y_{i_1} \cap \cdots \cap Y_{i_t} \to Y_{i_1} \cap \cdots \cap \widehat{(Y_{i_u})} \cap \cdots \cap Y_{i_t}.$$

In the general normal–crossings case (*i.e.* fibrations locally described by $f(z_1, \ldots, z_m) = z_1^{e_1} \cdots z_k^{e_k}$, $e_i \geq 1$), the definition of d_1' has to take into account multiplicity factors $\pm e_{i_j}$ before each map $(\delta_j^{(t)})^*$. The map d_1' is infact induced from a "wedging" operation with the form $\theta = \sum_{i=1}^k e_i \frac{dz_i}{z_i}$ (*cf.* last page). The definition of d_1'' is analogous to the one given in the strict normal–crossings case.
Notice that the weight spectral sequence (1.5) is built up from a filtered *double* complex. This property distinguishes this weight spectral sequence from others

as *e.g.* the spectral sequence of weights which defines the mixed Hodge structure on a quasi–projective smooth complex variety (*cf.* [3]).
The complex $A^\bullet_{\mathbf{C}}$ is the complex part of a cohomological mixed Hodge complex $A^\bullet_{\mathbf{Q}}$ whose definition is less explicit than $A^\bullet_{\mathbf{C}}$ and for which we refer to [7]. This rational complex induces on $H^\cdot(\tilde{X}^*, \mathbf{Q})$ a rational mixed Hodge structure. The rational representative of the above spectral sequence (1.5) is

$$E_1^{-r,n+r} = \bigoplus_{k \geq \max(0,-r)} H^{n-r-2k}(\tilde{Y}^{(2k+r+1)}, \mathbf{Q})(-r-k). \tag{1.7}$$

The index in the round brackets outside the cohomology refers to the Tate twist. Both these spectral sequences degenerate at $E_2 = E_\infty$ and they converge to $H^n(\tilde{X}^*, \mathbf{C})$ and $H^n(\tilde{X}^*, \mathbf{Q})$ respectively.
For curves (*i.e.* $d = 1$), the degeneration of the weight spectral sequence provides the exact sequences

$$0 \to E_2^{-1,2} \to H^0(\tilde{Y}^{(2)}, \mathbf{Q})(-1) \overset{d_1^{-1,2}}{\to} H^2(\tilde{Y}^{(1)}, \mathbf{Q}) \to H^2(\tilde{X}^*, \mathbf{Q}) \to 0$$

and

$$0 \to H^0(\tilde{X}^*, \mathbf{Q}) \to H^0(\tilde{Y}^{(1)}, \mathbf{Q}) \overset{d_1^{0,0}}{\to} H^0(\tilde{Y}^{(2)}, \mathbf{Q}) \overset{\alpha}{\to} H^1(\tilde{X}^*, \mathbf{Q}). \tag{1.8}$$

The differentials $d_1^{-1,2}$ and $d_1^{0,0}$ are defined as in (1.6) and the map α in (1.8) is the edge map in the spectral sequence. We also have a non canonical decomposition

$$H^1(\tilde{X}^*, \mathbf{Q}) = H^1(\tilde{Y}^{(1)}, \mathbf{Q}) \oplus E_2^{-1,2} \oplus E_2^{1,0}.$$

with $E_2^{1,0} = \mathrm{Im}(\alpha)$.
Steenbrink proved that the L–filtration induced on the abutment of the spectral sequence of the nearby cycles is the Picard-Lefschetz filtration, hence it is uniquely described by the following properties

$$N(L_{n+r}H^n(\tilde{X}^*, \mathbf{Q})) \subset (L_{n+r-2}H^n(\tilde{X}^*, \mathbf{Q}))(-1)$$

and

$$N^r : gr^L_{n+r}H^n(\tilde{X}^*, \mathbf{Q}) \overset{\simeq}{\to} (gr^L_{n-r}H^n(\tilde{X}^*, \mathbf{Q}))(-r)$$

for $r > 0$. In the rest of the paper we will refer to it as the *monodromy filtration.*

2. The monodromy operator as algebraic cocycle

We keep the notations introduced in the last paragraph. As n varies in $[0, 2d]$ ($d =$ dimension of the fiber of $f : X \to S$) and $i \geq 0$, the power maps

$$N^i : H^n(\tilde{X}^*, \mathbf{Q}) \to H^n(\tilde{X}^*, \mathbf{Q})(-i)$$

induced by the endomorphism $N : \mathbf{R}^n f_*(\Omega^\bullet_{X/S}(\log Y)) \to \mathbf{R}^n f_*(\Omega^\bullet_{X/S}(\log Y))$, define elements

$$N^i \in \mathrm{Hom}(H^\cdot(\tilde{X}^*, \mathbf{Q}), H^\cdot(\tilde{X}^*, \mathbf{Q})(-i))$$

which are invariant for the action of the local monodromy group π_1. They can be naturally identified with

$$N^i \in \bigoplus_{n\geq 0}\left[H^{2d-n}(\tilde{X}^*,\mathbf{Q})(d)\otimes H^n(\tilde{X}^*,\mathbf{Q})(-i)\right]^{\pi_1} = \left[H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q})(d-i)\right]^{\pi_1}.$$

The space $\tilde{X}^*\times_S\tilde{X}^*$ is the generic fiber of the product degeneration $X\times_S X\to S$. After a suitable sequence of blow-ups along $\mathrm{Sing}(Y\times Y)\supset\mathrm{Sing}(X\times_S X)$:

$$Z\to\cdots\to X\times_S X\to S$$

we obtain a normal–crossings degeneration $h: Z\to S$ with Z non singular and whose generic fiber is still $\tilde{X}^*\times\tilde{X}^*$. Its special fiber $T=h^{-1}(0)=T_1\cup\cdots\cup T_N$ has normal crossings singularities. The local description of h along T looks like:

$$h(w_1,\ldots,w_{2m})=w_1^{e_1}\cdots w_r^{e_r}$$

for $\{w_1,\ldots,w_{2m}\}$ a set of local parameters on Z and $e_1,\ldots,e_r$ non-negative integers.
The semistable reduction theorem (*cf.* [9]) assures that modulo extensions of the basis S and up to a suitable sequence of blow-ups and down along subvarieties of the special fiber T, we may eventually obtain from h a semistable degeneration $W\to S$ with $W_0=W_{0_1}\cup\ldots\cup W_{0_M}$ as special fiber.
Because of the assumption of the unipotency of the local monodromy on $H^*(X_t,\mathbf{C})$ (*cf.* § 1), the local monodromy σ of h will be also unipotent. We then call $\tilde{N}=\log(\sigma)$. By the Künneth decomposition it results: $\tilde{N}=1\otimes N+N\otimes 1$ and we have:

$$N^i\in\left(H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}(d-i))\right)^{\pi_1}=\mathrm{Ker}(\tilde{N})\cap H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}(d-i)).$$

Let consider the monodromy filtration $L.$ relative to the degeneration h. We denote by $Hom_{MH}(\mathbf{Q}(0),V)$ ($Hom(\mathbf{Q},V)$ shortly) the subgroup of Hodge cycles of pure weight $(0,0)$ of a bifiltered $\mathbf{Q}$–vector space V: (V,L,F), endowed with the corresponding mixed Hodge structure. Then, we have the following

Proposition 2.1. *For $i\geq 1$*

$$\begin{aligned}N^i &\in Hom_{MH}\left(\mathbf{Q}(0),Ker(\tilde{N})\cap H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}(d-i))\right)\subseteq\\ &\subseteq Hom_{MH}\left(\mathbf{Q}(0),Ker(\tilde{N})\cap(gr^L_{2(d-i)}H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}))(d-i)\right)\simeq\\ &\simeq Hom_{MH}\left(\mathbf{Q}(0),(gr^L_{2(d-i)}H^{2d}(T,\mathbf{Q}))(d-i)\right)\simeq Hom(\mathbf{Q},A),\end{aligned}$$

$$A:=\frac{Ker(\rho^{2(i+1)}:H^{2(d-i)}(\tilde{T}^{(2i+1)},\mathbf{Q})(d-i)\to H^{2(d-i)}(\tilde{T}^{(2(i+1))},\mathbf{Q})(d-i))}{Image\ \rho^{(2i+1)}}.$$

Here ρ is the restriction map on cohomology and by $\tilde{T}^{(j)}$ we mean the disjoint union of all ordered j–fold intersections of the components of T (cf. §1).

Proof. The identification of N^i with a Hodge cycle is a consequence of N being a morphism in the category of Hodge structures. The first inclusion derives from the well known facts that $\mathrm{Ker}(\tilde{N})$ has monodromic weight at most zero and that its Hodge cycles are included (Hom being a functor left exact on the second place) in the corresponding ones for the graded piece $(gr^L_{2(d-i)}H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}))(d-i)$ of $\mathrm{Ker}(\tilde{N})\cap\bigoplus_j(gr^L_j H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}))(d-i)$. The second isomorphism comes from the local invariant cycle theorem, namely from the following exact sequence of pure Hodge structures (*cf.* [2], lemma 3.3 and corollary 3.4)

$$0 \to gr^L_{2(d-i)}H^{2d}(T,\mathbf{Q}) \to gr^L_{2(d-i)}H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}) \overset{N}{\twoheadrightarrow} gr^L_{2(d-i-1)}H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q})(-1)$$

Finally, the last isomorphism is a consequence of the description of the graded piece $(gr^L_{2(d-i)}H^{2d}(T,\mathbf{Q}))(d-i)$ as sub–Hodge structure of $(gr^L_{2(d-i)}H^{2d}(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}))(d-i)$ (*cf. op.cit.* lemma 3.3). □

Proposition 2.1 shows how the operators N^i can be detected by classes $[N^i]$ in the cohomology of a fixed stratum of the special fiber T. Equivalently, we can say that N^i determine classes $[N^i]\in\mathbf{H}^{2d}(T,(gr^L_{-2i}\mathbf{R}\Psi_h(\mathbf{Q}))(d-i))$ in the $(E_1^{2i,2(d-i)})(d-i)$-term of the spectral sequence of weights for the degeneration h. Here we write $gr^L_{-2i}\mathbf{R}\Psi_h(\mathbf{Q})$ for $gr^L_{-2i}A^\bullet_{W,\mathbf{Q}}$.
The goal of this paper is to identify the class $[N^i]$ with an algebraic cocycle related to the degeneration $f: X\to S$. In all those cases that we will consider in the paper, this identification is obtained via a "correspondence-type" map $(i\geq 0)$

$$N^i:\mathbf{H}^*(Y,gr^L_rA^\bullet_{X,\mathbf{Q}})\to\mathbf{H}^*(Y,(gr^L_{r-2i}A^\bullet_{X,\mathbf{Q}})(-i))=\mathbf{H}^*(Y,gr^L_r(A^\bullet_{X,\mathbf{Q}}(-i)))$$

which makes the following diagram commute

$$\begin{tikzcd}
 & E_1^{-r+2i,*+r+2(d-i)} \arrow[d, equal] \\
\mathbf{H}^*(T,gr^L_rA^\bullet_{Z,\mathbf{Q}}) \arrow[r, "{[N^i]\cdot}"] & \mathbf{H}^{2d+*}(T,(gr^L_{r-2i}A^\bullet_{Z,\mathbf{Q}})(d-i)) \arrow[d, "(p_2)_*"] \\
\mathbf{H}^*(Y,gr^L_rA^\bullet_{X,\mathbf{Q}}) \arrow[u, "(p_1)^*"] \arrow[r, "N^i"] \arrow[d, equal] & \mathbf{H}^*(Y,gr^L_r(A^\bullet_{X,\mathbf{Q}}(-i))) \arrow[d, equal] \\
E_1^{-r,*+r} & E_1^{-r+2i,*+r-2i}
\end{tikzcd} \tag{2.1}$$

The projections $p_1,\ p_2 : \tilde{X}^* \times \tilde{X}^* \to \tilde{X}^*$ on the first and second factor, determine pullbacks and pushforwards on the hypercohomology as we shall describe in § 3.
From the theory we will explain in the next paragraphs and in the Appendix it will follow that N^i has the expected shape. Namely, it is zero when $N^i = 0$ and it is the identity when N^i induces an isomorphism on $E_2^{-r,*+r}$. Also, it will result that p_1^*, $(p_2)_*$ and $[N^i]\,\cdot$ all commute with the differential on E_1. That will imply an induced commutative diagram on E_2.
For $i = 0$, *i.e.* when the correspondence map is the identity, proposition 2.1 can be slightly generalized, using the theory developed in [2] (*cf.* lemma 3.3 and corollary 3.4) and in [1] so that the identity operator is seen as an element in

$$Hom_{MH}\left(\mathbf{Q}, \frac{\mathrm{Ker}(\rho^{(2)} : H^{2d}(\tilde{T}^{(1)}, \mathbf{Q})(d) \to H^{2d}(\tilde{T}^{(2)}, \mathbf{Q})(d))}{\mathrm{Im}\ (-i^* \cdot i_* : H_{2(d-1)}(T^{(1)}, \mathbf{Q})(d-1) \to H^{2d}(\tilde{T}^{(1)}, \mathbf{Q})(d)}\right) \simeq$$

$$\simeq Hom_{MH}\left(\mathbf{Q}, \frac{\mathrm{Im}(i^* : H^{2d}(T, \mathbf{Q})(d) \to H^{2d}(\tilde{T}^{(1)}, \mathbf{Q})(d))}{\mathrm{Im}\ (-i^* \cdot i_* : H_{2(d-1)}(T^{(1)}, \mathbf{Q})(d-1) \to H^{2d}(\tilde{T}^{(1)}, \mathbf{Q})(d)}\right).$$

Here the map i^* (*resp.* i_*) represents the pullback (*resp.* pushforward) relative to the embedding $T^{(1)} \to T$. Proposition 2.1 shows this class as a Hodge cocycle in $H^{2d}(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q}(d))$. That agrees with the classical theory of algebraic correspondences describing the identity map via an algebraic correspondence with the cycle diagonal. Namely, the identity is determined by the diagonal $\Delta_{\tilde{X}^*} \subset \tilde{X}^* \times \tilde{X}^*$ seen as specialization of the cycle diagonal on $\mathcal{X} \times \mathcal{X}$ on the fiber product $\tilde{X}^* \times \tilde{X}^*$. (*cf.* [8]).
The cases described in the next paragraphs will also supply some evidence for our expectation that $[N^i]$ can be always described by an algebraic (motivic) cocycle. Finally, notice that the calculation on the E_1 involves the cohomology of individual components of the strata and it is therefore in some sense local, whereas E_2 introduces relations among components of strata, so that any calculation on it becomes of global nature. That is the reason why the description of the monodromy cycle is carried out mainly at a local level in this paper.

3. Functoriality of the Steenbrink complex and remarks on products

Let $g : Z \to X$ be a morphism between two connected, complex analytic manifolds over a disk S. Let $f : X \to S$ and $h : Z \to S$ be the degeneration maps. Let assume that both Z and X are smooth over $\mathbf{C}$ and they have algebraic special fibers $f^{-1}(0) = Y$ and $h^{-1}(0) = T$ with normal crossings. We have the following commutative diagram

$$\begin{tikzcd} T \arrow[rr] \arrow[d, "i'"] & & Y \arrow[d, "i"] \\ Z \arrow[rr, "g"] \arrow[dr, "h"'] & & X \arrow[dl, "f"] \\ & S & \end{tikzcd}.$$

Locally on the special fibers, f and h have the following description

$$f(z_1,\dots,z_m)=z_1^{e_1}\cdots z_k^{e_k};\qquad h(w_1,\dots,w_M)=w_1^{e'_1}\cdots w_K^{e'_K}$$

for $\{z_1,\dots,z_m\}$ and $\{w_1,\dots,w_M\}$ local parameters *resp.* on X and Z, $1\leq k\leq m$, $1\leq K\leq M$ and $e_1,\dots,e_k$; $e'_1,\dots,e'_K$ integers.
Because $g^{-1}(Y)=T$, at any point $y\in g(T)\subset Y$ ($y=g(t)$, for some $t\in T$) where the local description of Y is $z_1^{e_1}\cdots z_k^{e_k}=0$, the pullback sections $g^*(z_{i_j})$ ($\forall\ 1\leq i_j\leq k$) define divisors on Z supported on T (not necessarily reduced or irreducible).
Let order the components of Y as $Y=Y_1\cup\ldots\cup Y_k$ and let denote by $\tilde{Y}^{(r)}$ the disjoint union of all intersections $Y_{i_1}\cap\ldots\cap Y_{i_r}$ for $1\leq i_1<\cdots<i_r\leq k$. There is a local system ϵ of rank one on $\tilde{Y}^{(r)}$ of standard orientations of r elements (*cf.* [3]). The canonical morphism

$$g^*\Omega^\bullet_X(\log Y)\to\Omega^\bullet_Z(\log T)$$

is a map of bifiltered complexes with respect to the weight and the Hodge filtrations on X and Z *resp.* (*cf. op.cit.*). In particular it induces the following map of bicomplexes of sheaves supported on the special fibers ($r\geq 0$)

$$g^*(W_rA^\bullet_{X,\mathbf{C}})\to W_rA^\bullet_{Z,\mathbf{C}}$$

where $A^\bullet_{\mathbf{C}}$ is the Steenbrink complex which represents in the derived category the maximal subobject of the complex of nearby cycles where the action of the monodromy is unipotent (*cf.* § 1). $W_rA^\bullet_{\mathbf{C}}$ is the induced weight filtration on $A^\bullet_{\mathbf{C}}$ (*cf.* (1.2)). Because the weight filtration on the complex $A^\bullet_{\mathbf{C}}$ is induced by the weight filtration on the de Rham complex with log-poles, g induces a map in the derived category

$$g^*(W_r\mathbf{R}\Psi_f(\mathbf{Q}_X))\to W_r\mathbf{R}\Psi_h(\mathbf{Q}_Z).$$

Notice that $g^*(\frac{dz_{i_j}}{z_{i_j}})\in W_1\Omega^1_Z(\log T)$, *i.e.* pullbacks preserve poles. Hence, we deduce the functoriality of the monodromy filtration

$$g^*(L_rA^\bullet_{X,\mathbf{C}})\to L_rA^\bullet_{Z,\mathbf{C}}.$$

Because g^{-1} is an exact functor, g determines on the graded pieces a pullback map

$$g^*:gr^L_rA^\bullet_{X,\mathbf{C}}\to gr^L_rA^\bullet_{Z,\mathbf{C}}$$

where

$$gr^L_rA^\bullet_{Z,\mathbf{C}}\simeq\bigoplus_{k\geq\max(0,-r)}(a_{2k+r+1})_*\Omega^\bullet_{\tilde{T}^{(2k+r+1)}}(\epsilon^{2k+r+1})[-r-2k].$$

The functor g^{-1} is also compatible with both differentials d' and d'' on $A^\bullet_{\mathbf{C}}$. Hence, g^* induces a morphism of bifiltered mixed Hodge complexes ($F^\cdot=$ Hodge filtration *cf.* [3])

$$g^*:(A^\bullet_{X,\mathbf{C}},L,F)\to(A^\bullet_{Z,\mathbf{C}},L,F)$$

which in turn induces a map between the spectral sequences of weights

$$g^*: E_1^{-r,q+r}(X) = \mathbf{H}^q(Y, gr_r^L A_X^\bullet) \to \mathbf{H}^q(T, gr_r^L A_Z^\bullet) = E_1^{-r,q+r}(Z).$$

On the rational level this morphism between spectral sequences is described by a direct sum of maps as

(3.1)

$$g^*: H^{q-r-2k}(\tilde{Y}^{(2k+r+1)}, \mathbf{Q})(-r-k) \to H^{q-r-2k}(\tilde{T}^{(2k+r+1)}, \mathbf{Q})(-r-k).$$

Both spectral sequences degenerate at $E_2 = E_\infty$. Keeping track of the multiplicities and the signs for these pullbacks can be rather hard. Let suppose that locally the defining equations for Y and T are $t = \prod_i z_i^{e_i}$ and $t = \prod_j w_j^{e'_j}$ respectively, and we are given strata $Y_I = Y_{i_1} \cap \ldots \cap Y_{i_p}$ $(i_1 < \ldots < i_p)$ and $Y_J = Y_{j_1} \cap \ldots \cap Y_{j_p}$. Then the computation of the multiplicities involved in $g^*: H^*(Y_I, \mathbf{Q}) \to H^*(T_J, \mathbf{Q})$ essentially amounts to determine the coefficients of $\frac{dw_{j_1}}{w_{j_1}} \wedge \ldots \wedge \frac{dw_{j_p}}{w_{j_p}}$ in $g^*(\frac{dz_{j_1}}{z_{j_1}} \wedge \ldots \wedge \frac{dz_{j_p}}{z_{j_p}})$. This technique will be frequently used in the paper.

As an example, we describe the map (3.1) when $f: X \to S$ is a degeneration of curves with normal crossings singularities on its special fiber Y and Z is the blow-up of X at a closed point $P \in Y$. Let $g: Z \to X$ be the blowing up map. If P is a regular point in the special fiber, the number of components of the special fiber T of Z will simply increase by one (the exceptional divisor E) and the remaining components are the same as for Y. Hence $g^*: H^0(\tilde{Y}^{(1)}, \mathbf{C}) \to H^0(\tilde{T}^{(1)}, \mathbf{C})$ is simply the map $g^*(1_{Y_i}) = 1_{T_i} + 1_E$ on the components.

Let suppose instead that P is singular. Since the description of g^* is local around each closed point, we may assume that the degeneration f is given, in a neighborhood of P, by the equation $z_1^{e_1} z_2^{e_2} = t$, being t a chosen parameter on the disk S and e_1, e_2 positive integers. Let assume that $e_1 \leq e_2$. Then, locally around P: $\tilde{Y}^{(1)} = Y_1 \coprod Y_2$. Set-theoretically one has $Y_i = \{z_i = 0\}$ $(i = 1, 2)$ and $\tilde{Y}^{(2)} = Y_1 \cap Y_2 = \{P\}$. Then, $\tilde{T}^{(1)} = T_1 \coprod T_2 \coprod T_3$ where T_1 and T_2 are the strict transforms of the two components Y_i, while T_3 represents the exceptional divisor. We implicitly have fixed the standard orientation on $\tilde{Y}^{(r)}$ (*e.g.* $\tilde{Y}^{(2)} = Y_1 \cap Y_2 = Y_{12}$). On $\tilde{T}^{(r)}$, we choose the orientation for which the exceptional component T_3 is always considered as the last one.

There are only three graded complexes $gr_*^L A_{\mathbf{C}}^\bullet$ non zero both on X and Z. On X they have the following description

$$gr_{-1}^L A_{X,\mathbf{C}}^\bullet \simeq (a_2)_* \Omega_{\tilde{Y}^{(2)}}^\bullet[-1]$$

$$gr_0^L A_{X,\mathbf{C}}^\bullet \simeq (a_1)_* \Omega_{\tilde{Y}^{(1)}}^\bullet$$

and via the isomorphism (1.4) one has:

$$\tilde{\nu}: gr_1^L A_{X,\mathbf{C}}^\bullet \xrightarrow{\simeq} gr_{-1}^L A_{X,\mathbf{C}}^\bullet.$$

Hence $E_1^{1,q-1} = \mathbf{H}^q(Y, gr_{-1}^L A_{X,\mathbf{C}}^\bullet) = 0$ unless $q = 1$, in which case we get

$$g^*: H^0(\tilde{Y}^{(2)}, \mathbf{C}) \to H^0(\tilde{T}^{(2)}, \mathbf{C}).$$

To understand the description of this map, one has to look at the local geometry of the blow-up at P. It is quite easy to check that Z is covered by two open sets, say $Z = U \cup V$. To make the notations easier, let call $t_1 = \frac{z_1}{z_2}$ and $t_2 = \frac{z_2}{z_1}$. On U, described by $t_2^{e_2} = \frac{t}{z_1^{e_1+e_2}}$, one has coordinates $\{t_2, z_1\}$, $T_2^{\mathrm{red}} = \{t_2 = 0\}$ and $T_3^{\mathrm{red}} = \{z_1 = 0\}$. On V, described by $t_1^{e_1} = \frac{t}{z_2^{e_1+e_2}}$, one has coordinates $\{t_1, z_2\}$, $T_1^{\mathrm{red}} = \{t_1 = 0\}$ and $T_3^{\mathrm{red}} = \{z_2 = 0\}$. Then $\tilde{T}^{(2)} = T_{13} \coprod T_{23}$, here we denote $T_{ij} = T_i \cap T_j$.
On U we have $g^*(\frac{dz_1}{z_1} \wedge \frac{dz_2}{dz_2}) = \frac{dz_1}{dz_1} \wedge \frac{dt_2}{t_2}$, whereas on V one gets $g^*(\frac{dz_1}{z_1} \wedge \frac{dz_2}{dz_2}) = \frac{dt_1}{dt_1} \wedge \frac{dz_2}{z_2}$. Hence, keeping in account the fixed orientation among the components of T, the description of the pullback $g^*(1_{\tilde{Y}^{(2)}}) = g^*(1_{Y_{12}})$ is given by

$$g^*(1_{Y_{12}}) = 1_{T_{13}} - 1_{T_{23}}.$$

The presence of a negative sign is due to the change of orientation. This description defines the above map g^* on H^0. Similarly, we find that

$$g^* : H^0(\tilde{Y}^{(1)}, \mathbf{C}) \to H^0(\tilde{T}^{(1)}, \mathbf{C})$$

is given by $g^*(1_{Y_1}) = 1_{T_1} + 1_{T_3}$ and $g^*(1_{Y_2}) = 1_{T_2} + 1_{T_3}$. The description of g^* on the terms H^1 goes in parallel.
Let now consider the proper map that g induces on the closed fibers. For simplicity of notations we call it $g : T \to Y$. Let $d = (\dim\ T - \dim\ Y)$. The above arguments have shown that g induces a pullback map g^* between the cohomologies of the strata: *cf.* (3.1). Since each stratum is a smooth projective complex variety (not connected), we can use the Poicaré duality to associate to each pullback in (3.1) that contributes to the definition of the map g^* its dual so that we naturally obtain a dual pushforward on the E_1–terms of the spectral sequence of weights that is described by a direct sum of maps as

$$(3.2)\quad g_! : H^{q-r-2(k-d)}(\tilde{T}^{(2k+r+1)}, \mathbf{Q})(-r-k+d) \\ \to H^{q-r-2k}(\tilde{Y}^{(2k+r+1)}, \mathbf{Q})(-r-k).$$

On each stratum $g_!$ is defined by the following formula

$$(\frac{1}{2\pi\sqrt{-1}})^{d-2k-r} \int_{\tilde{Y}^{(2k+r+1)}} g_!(\alpha) \cup \beta = (\frac{1}{2\pi\sqrt{-1}})^{2d-2k-r} \int_{\tilde{T}^{(2k+r+1)}} \alpha \cup g^*(\beta)$$

where $\int$ denotes the morphism trace described by the cap–product with the fundamental class of each component of the stratum, for any chosen couple of elements $\alpha \in H^{q+2(2d-2k-r)}(\tilde{T}^{(2k+r+1)}, \mathbf{Q}(2d-2k-r))$ and $\beta \in H^{-q}(\tilde{Y}^{(2k+r+1)}, \mathbf{Q})$, $q \in Z$, $q \geq 0$.
Notice that although we have a notion of bifiltered pullback

$$g^* : (A_X^\bullet, L, F) \to (A_Z^\bullet, L, F)$$

this does not imply a canonical definition of a product structure on $A_{\mathbf{C}}^\bullet$ obtained via pullback along the diagonal map $\Delta : X \to X \times_S X$. In fact, the property of $f : X \to S$ to have normal crossings reduction is not preserved by the product map $f \times f : X \times_S X \to S$. The space $X \times_S X$ is in general not even smooth over $\mathbf{C}$! Finally, we remark that although the monodromy filtration is not multiplicative on the level of the filtered complexes $(A^\bullet_{\mathbf{C}}, L)$ (the simple example shown below will motivate this claim), it becomes multiplicative on the limiting cohomology with its mixed Hodge structure.

Example 3.1.

Let $f : \mathbf{P}^1_S \to S$ be a $\mathbf{P}^1$-fibration over a disk S. We blow a closed point $P \in \mathbf{P}^1_0 = Y$ in the fiber $\mathbf{P}^1_0$ over the origin $\{0\}$. The resulting map $h : Z \to S$ has a normal crossings special fiber $h^{-1}(0) = T = T_1 \cup T_2$, where T_1 is the strict transform of Y and T_2 is the exceptional component (*i.e.* $\mathbf{P}^1$). The intersection $Q = T_1 \cap T_2 = T_{12}$ is transverse. Locally around Q, h has the following description

$$h(z_1, z_2) = z_1 z_2.$$

Consider the subcomplex $W_0(A^\bullet_{Z,\mathbf{C}})$ of $A^\bullet_{Z,\mathbf{C}}$ filtered by the monodromy filtration L induced on it by the one on $A^\bullet_{Z,\mathbf{C}}$ (*cf.* § 1, (1.2)). Its hypercohomology computes $H^*(T, \mathbf{C})$ and it can be determined in terms of the homology of the complex

$$\begin{aligned}\{\mathcal{C}^\bullet :\ H^\cdot(\tilde{T}^{(1)}, \mathbf{C}) \xrightarrow{d} H^\cdot(\tilde{T}^{(2)}, \mathbf{C})\} \quad &= \\ &\{\mathcal{C}^\bullet :\ H^\cdot(T_1, \mathbf{C}) \oplus H^\cdot(T_2, \mathbf{C}) \xrightarrow{d} H^\cdot(T_{12}, \mathbf{C})\}\end{aligned}$$

where $\mathcal{C}^\bullet$ sits in degrees zero and one. The differential d on $\mathcal{C}^\bullet$ is of “Čech type” *i.e.* it is an alternate sum of pullback maps as defined in (1.6). A product in the filtered derived category $(A^\bullet_{Z,\mathbf{C}}, L)$ if any exists, should induce a product on $\mathcal{C}^\bullet$. The tensor product $\mathcal{C}^\bullet \otimes \mathcal{C}^\bullet$ is a complex sitting in degrees zero, one and two and it has the following description

$$\begin{aligned}\{\mathcal{C}^\bullet \otimes \mathcal{C}^\bullet :\ \bigoplus_{i,j\in[1,2]} (H^\cdot(T_i, \mathbf{C}) \otimes H^\cdot(T_j, \mathbf{C})) \overset{d\otimes d}{\to} \bigoplus_{i=1}^{2} \{(H^\cdot(T_i, \mathbf{C}) \otimes H^\cdot(T_{12}, \mathbf{C})) \oplus \\ \oplus (H^\cdot(T_{12}, \mathbf{C}) \otimes H^\cdot(T_i, \mathbf{C}))\} \overset{d\otimes d}{\to} (H^\cdot(T_{12}, \mathbf{C}) \otimes H^\cdot(T_{12}, \mathbf{C}))\}.\end{aligned}$$

However, there is no way to define canonically the product

$$\mu : \mathcal{C}^\bullet \otimes \mathcal{C}^\bullet \to \mathcal{C}^\bullet.$$

In fact, let's look for a possible description of it in each degree. In degree zero a product should satisfy

$$\begin{aligned} H^\cdot(T_1, \mathbf{C}) \otimes H^\cdot(T_1, \mathbf{C}) &\mapsto H^\cdot(T_1, \mathbf{C}), \\ H^\cdot(T_2, \mathbf{C}) \otimes H^\cdot(T_2, \mathbf{C}) &\mapsto H^\cdot(T_2, \mathbf{C}), \\ H^\cdot(T_i, \mathbf{C}) \otimes H^\cdot(T_j, \mathbf{C}) &\mapsto 0, \quad i, j = 1, 2. \end{aligned}$$

In degree one, one could start by setting

$$\begin{aligned} &H^{\cdot}(T_1,\mathbf{C})\otimes H^{\cdot}(T_{12},\mathbf{C})\mapsto H^{\cdot}(T_{12},\mathbf{C}),\\ &H^{\cdot}(T_{12},\mathbf{C})\otimes H^{\cdot}(T_2,\mathbf{C})\mapsto H^{\cdot}(T_{12},\mathbf{C}),\\ &H^{\cdot}(T_2,\mathbf{C})\otimes H^{\cdot}(T_{12},\mathbf{C})\mapsto 0,\\ &H^{\cdot}(T_{12},\mathbf{C})\otimes H^{\cdot}(T_1,\mathbf{C})\mapsto 0. \end{aligned}$$

Notice however, that this definition is not at all canonical, as one could alternatively set

$$\begin{aligned} &H^{\cdot}(T_2,\mathbf{C})\otimes H^{\cdot}(T_{12},\mathbf{C})\mapsto H^{\cdot}(T_{12},\mathbf{C}),\\ &H^{\cdot}(T_{12},\mathbf{C})\otimes H^{\cdot}(T_1,\mathbf{C})\mapsto H^{\cdot}(T_{12},\mathbf{C}),\\ &H^{\cdot}(T_1,\mathbf{C})\otimes H^{\cdot}(T_{12},\mathbf{C})\mapsto 0,\\ &H^{\cdot}(T_{12},\mathbf{C})\otimes H^{\cdot}(T_2,\mathbf{C})\mapsto 0. \end{aligned}$$

Finally, in degree two one would have

$$H^{\cdot}(T_{12},\mathbf{C})\otimes H^{\cdot}(T_{12},\mathbf{C})\mapsto H^{\cdot}(T_{12},\mathbf{C})$$

4. Semistable degenerations with double points

This section is mainly devoted to the determination of $[N]$ for one–dimensional semistable fibrations with at worst double points as singularities. The description of $[N]$ is obtained via the introduction of the algebraic correspondence-type square on the cohomology groups of the special fiber as described in (2.1). A one–dimensional double point degeneration is the simplest example of a normal crossings fibration. The generalization of these results to double points semistable degenerations of arbitrary dimension is done at the end of this paragraph where we also report as an example of application of these results the case of a Lefschetz pencil.

We keep the same notations as in § 3, in particular we denote by $f: X\to S$ a *semistable* fibration of fiber dimension one. Its special fiber is denoted by Y. By definition, locally around a double point $P\in Y$ the description of f looks like

$$f(z_1,z_2)=z_1z_2$$

for $\{z_1,z_2\}$ local parameters on X at P. For one dimensional fiberings, the only group where the local monodromy may act non trivially is $gr_2^L H^1(\tilde{X}^*,\mathbf{Q})$, in which case the identity map on the E_1-terms of the weight spectral sequence (1.5)

$$E_1^{-1,2}=H^0(\tilde{Y}^{(2)},\mathbf{Q})(-1)\stackrel{\mathrm{Id}}{\to}H^0(\tilde{Y}^{(2)},\mathbf{Q})(-1)=E_1^{1,0}(-1)$$

determines an isomorphism of rational Hodge structures of weight two on the related graded groups $E_2=E_\infty$. This isomorphism is induced by the action of the local monodromy N around the origin:

$$N : gr_2^L H^1(\tilde{X}^*, \mathbf{Q}) \overset{\simeq}{\to} (gr_0^L H^1(\tilde{X}^*, \mathbf{Q}))(-1)$$

It is a well known consequence of the Clemens-Schmid exact sequence (considered as a sequence of mixed Hodge structures) that (*cf.* [9])

$$\begin{aligned} &gr_2^L H^1(\tilde{X}^*, \mathbf{Q}) \neq 0 \;\Leftrightarrow \\ &\qquad \Leftrightarrow\; \mathrm{Ker}(\rho^{(2)} : H^1(\tilde{Y}^{(1)}, \mathbf{Q}) \to H^1(\tilde{Y}^{(2)}, \mathbf{Q})) \neq 0 \;\Leftrightarrow\; h^1(|\Gamma|) \neq 0 \end{aligned}$$

where $h^1(|\Gamma|)$ is the dimension of the first rational cohomology group of the geometric realization of the dual graph of Y. It follows from proposition 2.1 that $[N] \in \mathbf{H}^2(T, (gr_{-2}^L A^\bullet_{Z,\mathbf{Q}})) = H^0(\tilde{T}^{(3)}, \mathbf{Q})$ determines a Hodge class

$$\begin{aligned} (4.1) \quad & [N] \in Hom_{MH}(\mathbf{Q}(0), gr_0^L H^2(T, \mathbf{Q})) \simeq \\ & \simeq Hom_{MH}\left(\mathbf{Q}(0), \frac{H^0(\tilde{T}^{(3)}, \mathbf{Q})}{\mathrm{Image}(\rho^{(3)} : H^0(\tilde{T}^{(2)}, \mathbf{Q}) \to H^0(\tilde{T}^{(3)}, \mathbf{Q}))}\right). \end{aligned}$$

Here T is the special fiber of a normal–crossings degeneration $h : Z \to S$. The variety Z is a smooth threefold over $\mathbf{C}$ obtained via resolution of the singularities of $X \times_S X$. Notice that no more than three components of T intersect at the same closed point since $\dim Z = 3$.
We shall determine the Hodge cycle $[N] \in E_1^{2,0}(Z) = H^0(\tilde{T}^{(3)}, \mathbf{Q})$ by means of a "correspondence type" map

$$N : \mathbf{H}^*(Y, gr_r^L A^\bullet_{X,\mathbf{Q}}) \to \mathbf{H}^*(Y, (gr_{r-2}^L A^\bullet_{X,\mathbf{Q}})(-1)) = \mathbf{H}^*(Y, gr_r^L(A^\bullet_{X,\mathbf{Q}}(-1)))$$

as we explained in (2.1). From the proof it will easily follow that the map N is zero for $* \neq 1$ and is the identity for $* = 1 = r$. On the E_2-level it will induce (for $* = 1 = r$) a commutative diagram

$$\begin{array}{ccc} gr_2^L H^1(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q}) & \xrightarrow{[N]\,\cdot} & gr_2^L H^3(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q}) = E_2^{1,2} \\ {\scriptstyle (p_1)^*}\uparrow & & \downarrow{\scriptstyle (p_2)_*} \\ E_2^{-1,2} = gr_2^L H^1(\tilde{X}^*, \mathbf{Q}) & \xrightarrow{N} & (gr_0^L H^1(\tilde{X}^*, \mathbf{Q}))(-1) = (E_2^{1,0})(-1) \end{array}$$

The pullback p_1^* and pushforward $(p_2)_*$ are defined as in § 3. The above diagram will determine uniquely both $[N] \in Hom_{MH}(\mathbf{Q}(0), gr_0^L H^2(T, \mathbf{Q}))$ and the product $[N]\,\cdot$.
The following result defines the geometry of the model Z and the special fiber T after resolving the singularities of $X \times_S X$ and $Y \times Y$.

Lemma 4.1. *Let $z_1 z_2 = w_1 w_2$ be a local description of $X \times_S X$ around the point (P, P), with $P \in Y = Y_1 \cup Y_2$ a double point of f and $\{w_1,\ w_2\}$ a second set of regular parameters on X at P. After a blow-up of $X \times_S X$ with center at the origin (z_1, z_2, w_1, w_2), the resulting degeneration $h : Z \to S$ is normal–crossings. Its special fiber T is the union of five irreducible components: $T = \cup_{i=1}^5 T_i$. We number them so that the first four are the strict transforms of the irreducible components $Y_i \times Y_j$ of $Y \times Y$, namely $T_1 = (Y_1 \times Y_1)\tilde{}$, $T_2 = (Y_1 \times Y_2)\tilde{}$,*

$T_3 = (Y_2 \times Y_1)\tilde{}$, $T_4 = (Y_2 \times Y_2)\tilde{}$. The last one T_5 represents the exceptional divisor of the blow-up. We have $\tilde{T}^{(1)} = \coprod_i T_i$. The scheme Z is covered by four affine charts $\mathcal{U}_j$. On each of them there are three non empty components T_k. The scheme $\tilde{T}^{(3)}$ is the disjoint union of four zero dimensional schemes (closed points): $T_{125} \in \mathcal{U}_2$, $T_{135} \in \mathcal{U}_4$, $T_{245} \in \mathcal{U}_3$ and $T_{345} \in \mathcal{U}_1$, each of whose supports projects isomorphically onto the diagonal $\Delta_{12} : Y_{12} \to Y_{12} \times Y_{12}$.

Proof. The local description of $X \times_S X$ around (P, P) is given by the equations $z_1 z_2 = w_1 w_2$ and $z_1 z_2 = t$, for $t \in S$ a fixed parameter on the disk. We choose the standard orientation of the sets $\{z_1, z_2\}$ and $\{w_1, w_2\}$ and we write $w'_{i_1} = \frac{w_i}{z_1}$, $w'_{i_2} = \frac{w_i}{z_2}$, $w_{ij} = \frac{w_i}{w_j}$, $z'_{i_1} = \frac{z_i}{w_1}$, $z'_{i_2} = \frac{z_i}{w_2}$ and $z_{ij} = \frac{z_i}{z_j}$, for $i, j = 1, 2$. After a single blow-up of $X \times_S X$ at the origin (z_1, z_2, w_1, w_2), the resulting model Z is non singular as one can see by looking at the first of the following tables which describes Z on each of the four charts $\mathcal{U}_j$ who cover it. In the second table, we have collected for each $\mathcal{U}_j$, the description of the non empty divisors $T_k \in T^{(1)}$ there. We use the pullbacks $p_1^*(\frac{dz_1}{z_1} \wedge \frac{dz_2}{z_2})$ and $p_2^*(\frac{dw_1}{w_1} \wedge \frac{dw_2}{w_2})$ to define in the third table the pullbacks $p_i^*(1_{Y_{12}}) \in H^0(\tilde{T}^{(2)}, \mathbf{Q})$.

Open sets	Loc. coordinates and relations
$\mathcal{U}_1$	$\{w'_{1_1}, w'_{2_1}, z_1\},\ w'_{1_1} w'_{2_1} = z_{21}$
$\mathcal{U}_2$	$\{w'_{1_2}, w'_{2_2}, z_2\},\ w'_{1_2} w'_{2_2} = z_{12}$
$\mathcal{U}_3$	$\{z'_{1_1}, z'_{2_1}, w_1\},\ z'_{1_1} z'_{2_1} = w_{21}$
$\mathcal{U}_4$	$\{z'_{1_2}, z'_{2_2}, w_2\},\ z'_{1_2} z'_{2_2} = w_{12}$

Open sets	Divisors
$\mathcal{U}_1$	$T_3 = \{w'_{1_1} = 0\},\ T_4 = \{w'_{2_1} = 0\},\ T_5 = \{z_1 = 0\}$
$\mathcal{U}_2$	$T_1 = \{w'_{1_2} = 0\},\ T_2 = \{w'_{2_2} = 0\},\ T_5 = \{z_2 = 0\}$
$\mathcal{U}_3$	$T_2 = \{z'_{1_1} = 0\},\ T_4 = \{z'_{2_1} = 0\},\ T_5 = \{w_1 = 0\}$
$\mathcal{U}_4$	$T_1 = \{z'_{1_2} = 0\},\ T_3 = \{z'_{2_2} = 0\},\ T_5 = \{w_2 = 0\}$

Open sets	$p_1^*(1_{Y_{12}})$	$p_2^*(1_{Y_{12}})$
$\mathcal{U}_1$	$-1_{T_{35}} - 1_{T_{45}}$	$-1_{T_{45}} + 1_{T_{35}} + 1_{T_{34}}$
$\mathcal{U}_2$	$1_{T_{15}} + 1_{T_{25}}$	$-1_{T_{25}} + 1_{T_{15}} + 1_{T_{12}}$
$\mathcal{U}_3$	$-1_{T_{45}} + 1_{T_{25}} + 1_{T_{24}}$	$-1_{T_{25}} - 1_{T_{45}}$
$\mathcal{U}_4$	$-1_{T_{35}} + 1_{T_{15}} + 1_{T_{13}}$	$1_{T_{15}} + 1_{T_{35}}$

The global description of the pullbacks $p_1^*(1_{Y_{12}})$ and $p_2^*(1_{Y_{12}})$ is

$$p_1^*(1_{Y_{12}}) = (1_{T_{15}} + 1_{T_{25}} - 1_{T_{35}} - 1_{T_{45}}) + 1_{T_{13}} + 1_{T_{24}}$$

$$p_2^*(1_{Y_{12}}) = (1_{T_{15}} - 1_{T_{25}} + 1_{T_{35}} - 1_{T_{45}}) + 1_{T_{12}} + 1_{T_{34}}.$$

Finally, notice that each $\mathcal{U}_j$ is isomorphic to $\mathbf{A}^3$ and in each of them one has three non empty components T_k. $\square$

The following result holds

Theorem 4.2. *Let $f : X \to S$ be the semistable degeneration of curves as described above. Then, the following description of $[N] \in H^0(\tilde{T}^{(3)}, \mathbf{Q})$ (cf. (4.1)) holds:*

$$[N] = a_{125}1_{T_{125}} + a_{135}1_{T_{135}} + a_{245}1_{T_{245}} + a_{345}1_{T_{345}}$$

where the (rational) numbers a's are subject to the following requirement:

$$-2a_{125} + 2a_{135} - 2a_{245} + 2a_{345} = 1.$$

The induced class $[N]$ in $gr_0^L H^2(T, \mathbf{Q})$ (i.e. modulo boundary relations via the restriction map $\rho^{(3)}$ cf. (1.6)) determines a unique zero–cycle.

Proof. We determine $[N]$ as a cocycle making the following square commute

$$\begin{array}{ccc} gr_2^L H^1(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q}) & \xrightarrow{[N]\,\cdot} & gr_2^L H^3(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q}) = E_2^{1,2} \\ {\scriptstyle (p_1)^*}\uparrow & & \downarrow{\scriptstyle (p_2)_*} \\ E_2^{-1,2} = gr_2^L H^1(\tilde{X}^*, \mathbf{Q}) & \xrightarrow{N} & (gr_0^L H^1(\tilde{X}^*, \mathbf{Q}))(-1) = (E_2^{1,0})(-1) \end{array} \tag{4.2}$$

In terms of cohomologies of strata, we have to describe explicitly a representative of $[N]$ in $E_1^{2,0}(Z)$ that satisfies the commutativity of

$$\begin{array}{ccc} H^0(\tilde{T}^{(2)}, \mathbf{Q})(-1) & \xrightarrow{[N]\,\cdot} & H^2(\tilde{T}^{(2)}, \mathbf{Q}) \\ {\scriptstyle p_1^*}\uparrow & & \downarrow{\scriptstyle (p_2)_*} \\ E_1^{-1,2} = H^0(\tilde{Y}^{(2)}, \mathbf{Q})(-1) & = & H^0(\tilde{Y}^{(2)}, \mathbf{Q})(-1) = E_1^{1,0}(-1). \end{array} \tag{4.3}$$

With the notations used in lemma 4.1 the description of $[N]$ is given by

$$[N] = a_{125}1_{T_{125}} + a_{135}1_{T_{135}} + a_{245}1_{T_{245}} + a_{345}1_{T_{345}}.$$

For the standard choice of the orientations of $\{z_1, z_2\}$ and $\{w_1, w_2\}$ and the numbering of the T_i's defined in lemma 4.1, the local description of the pull-backs $p_i^*(1_{Y_{12}})$ for $i = 1, 2$ is given in the third table of the above lemma. Following the definition described in the Appendix (*cf.* (7.6)), the product $[N] \cdot p_1^*(1_{Y_{12}}(-1))$ is then the following

$$
\begin{aligned}
(4.4) \qquad & [N] \cdot p_1^*(1_{Y_{12}}(-1)) = \\
&= [N] \cdot (1_{T_{15}}(-1) + 1_{T_{15}}(-1) - 1_{T_{35}}(-1) - 1_{T_{45}}(-1)) \\
&= a_{125}(g_1(1_{T_{125}} \cdot 1_{T_{15}}(-1)) - g_2(1_{T_{125}} \cdot 1_{T_{25}}(-1))) \\
&\quad + a_{135}(g_1(1_{T_{135}} \cdot 1_{T_{15}}(-1)) + g_3(1_{T_{135}} \cdot 1_{T_{35}}(-1))) \\
&\quad + a_{245}(g_2(1_{T_{245}} \cdot 1_{T_{25}}(-1)) + g_4(1_{T_{245}} \cdot 1_{T_{45}}(-1))) \\
&\quad + a_{345}(-g_3(1_{T_{345}} \cdot 1_{T_{35}}(-1)) + g_4(1_{T_{345}} \cdot 1_{T_{45}}(-1))) \\
&= a_{125}(1_{T_{25}} - 1_{T_{15}}) + a_{135}(1_{T_{35}} + 1_{T_{15}}) \\
&\quad + a_{245}(1_{T_{45}} + 1_{T_{25}}) + a_{345}(-1_{T_{45}} + 1_{T_{35}}).
\end{aligned}
$$

The maps g_1, g_2, g_3 and g_4 are the pushforwards as introduced in the Appendix. The following formula illustrates the product $1_{T_{ijk}} \cdot \sum_{l,m} 1_{T_{lm}}(-1)$ following the definition of it given in the Appendix:

$$
\begin{aligned}
1_{T_{ijk}} \cdot \sum_{l,m} 1_{T_{lm}}(-1) &= 1_{T_{ijk}} \cdot (1_{T_{ik}}(-1) + 1_{T_{jk}}(-1)) \\
&= g_i(1_{T_{ijk}} \cdot 1_{T_{ik}}(-1)) - g_j(1_{T_{ijk}} \cdot 1_{T_{jk}}(-1)) = g_i(1_{T_{ijk}}) - g_j(1_{T_{ijk}}) \\
&\qquad \in \mathrm{Image}\big(\bigoplus_t g_t : H^0(\tilde{T}^{(3)}, \mathbf{Q})(-1) \to H^2(\tilde{T}^{(2)}, \mathbf{Q})\big).
\end{aligned}
$$

In (4.4), we have denoted, for simplicity of notations, the difference $g_i(1_{T_{ijk}}) - g_j(1_{T_{ijk}})$ with $1_{T_{jk}} - 1_{T_{ik}}$. The map g_i represents the pushforward on cycles deduced from the embedding $g_i : T_{ijk} \to T_{jk}$. The definition of g_j is similar. Therefore, via the local definition of the pushforward $(p_2)_*$ along the affine charts (*cf.* §3 and third table in lemma 4.1), we obtain:

$$(p_2)_*([N] \cdot p_1^*(1_{Y_{12}}(-1))) = (-2a_{125} + 2a_{135} - 2a_{245} + 2a_{345})1_{Y_{12}}(-1).$$

The commutativity of (4.3) and hence of (4.2) is then equivalent to the requirement

$$-2a_{125} + 2a_{135} - 2a_{245} + 2a_{345} = 1.$$

Hence, the operator $[N]$ is determined as a cocycle in $H^0(\tilde{T}^{(3)}, \mathbf{Q})$ by the setting

$$
(4.5) \qquad
\begin{aligned}
&[N] = a_{125}1_{T_{125}} + a_{135}1_{T_{135}} + a_{245}1_{T_{245}} + a_{345}1_{T_{345}}; \\
&\quad - 2a_{125} + 2a_{135} - 2a_{245} + 2a_{345} = 1.
\end{aligned}
$$

Up to boundary relations by means of the restriction map $\rho^{(3)}$ which connects the elements $1_{T_{125}}$ with $1_{T_{245}}$ and $1_{T_{135}}$ with $1_{T_{345}}$, (4.5) determines a unique zero–cycle in the quotient $E_2^{2,0}(Z)$ (*cf.* (4.1)). Of course, if $N = 0$, this class may be trivial. □

Remark 4.3.

The description of $[N] \in E_1^{2,0}(Z)$ as well as the relation among the coefficients a_{ijk} in (4.5) is not unique in E_1. In fact, it depends on the choice of the desingularization process, as well as on the ordering of the components $T_k \in \tilde{T}^{(1)}$. For example, for the ordering of T_k for which T_1 represents in each chart the exceptional divisor of the blow-up ($T_2 = (Y_1 \times Y_1)\tilde{}$, $T_3 = (Y_1 \times Y_2)\tilde{}$, $T_4 = (Y_2 \times Y_1)\tilde{}$, $T_5 = (Y_2 \times Y_2)\tilde{}$), the setting (4.5) becomes

$$[N] = a_{123}1_{T_{123}} + a_{124}1_{T_{124}} + a_{135}1_{T_{135}} + a_{145}1_{T_{145}};$$
$$-a_{123} + a_{124} - a_{135} + a_{145} = 1.$$

If instead we choose to desingularize $X \times_S X$ via a blowing-up along $z_1 = w_1 = 0$ and we set the order among the T_k's so that the exceptional divisor is represented in each chart by the last component (*i.e.* $T_1 = (Y_1 \times Y_2)\tilde{}$, $T_2 = (Y_2 \times Y_1)\tilde{}$, $T_3 = (Y_2 \times Y_2)\tilde{}$, $T_4 = (Y_1 \times Y_1)\tilde{}$,), then we would get

$$[N] = a_{134}1_{T_{134}} + a_{234}1_{T_{234}};$$
$$-a_{134} + a_{234} = 1.$$

It is a consequence of the uniqueness of the product structure on the corresponding E_2–terms that all these different settings determine a unique description of $[N] \in E_2^{2,0}(Z)$.
In what it follows we support some evidence for our belief that the description of $[N]$ for a double points degeneration of higher fiber dimension (*i.e.* locally described by $f(z_1,\dots,z_n) = z_iz_j$, *cf.* below) is deducible from the case worked out for curves. As already remarked, the description of $[N]$ in the cohomology of the strata of the special fiber of the fiber product resolution is of local nature, *i.e.* it can be described locally around each double point. For a higher dimensional double points degeneration $[N]$ should be again described in terms of a "diagonal" cocycle whose support projects isomorphically onto the diagonal $\Delta_{12} \in Y_{12} \times Y_{12}$ as it was shown in theorem 4.2. In general, that "diagonal" cocycle would be formally locally a bundle over the corresponding diagonal cocycle which comes up for a degeneration of curves. This is a consequence of the local description of the degeneration map around a double point. We give now some details for these ideas.
Let $f : X \to S$ be a semistable degeneration with double points of fiber dimension d over the disk S. Then, locally in a neighborhood of a double point P on Y, f has the following description

$$f(z_1,\dots,z_n) = z_iz_j$$

for $\{z_1,\dots,z_n\}$ a set of regular parameters on X at P and suitable indices $i < j$ in $I = \{1,\dots,n\}$. Let $Y = Y_1 \cup Y_2$ be the local description of Y in a neighborhood of $P \in Y_1 \cap Y_2 = Y_{12}$. Locally around P, $\{z_1,\dots,\hat{z}_i,\dots,\hat{z}_j,\dots,z_n\}$ are free parameters for this description. Hence, the special fiber is locally around the point, formally isomorphic to $\mathbf{A}^{d-1} \times \hat{Y}$ with $\hat{Y} = \hat{Y}_1 \cup \hat{Y}_2$ of dimension 1. In a formal neighborhood of P, Y is defined by $\mathrm{Spec}\big(\mathbf{C}\{\{z_1,\dots,\hat{z}_i,\dots,\hat{z}_j,\dots z_n\}\}[z_i,z_j]/z_iz_j\big)$. The model X

is formally locally isomorphic to $\mathbf{A}^{d-1} \times \hat{X}$, with $\hat{X}$ of fiber dimension 1 and special fiber $\hat{Y}$. The formal description of $X \times_S X$ is similar, namely $X \times_S X \simeq \mathbf{A}^{d-1} \times \mathbf{A}^{d-1} \times (\hat{X} \times_S \hat{X})$. Keeping the same notations introduced before, we get a formal local description of the stratum $\tilde{T}^{(3)}$ (containing the cocycle $[N]$) as $\mathbf{A}^{d-1} \times \mathbf{A}^{d-1} \times \hat{\tilde{T}}^{(3)}$, with $\hat{\tilde{T}}^{(3)}$ collection of points. $[N]$ is a cycle (of dimension $d-1$) in $\tilde{T}^{(3)}$ formally, locally described by $\Delta_{\mathbf{A}^{d-1}} \times \hat{\tilde{T}}^{(3)}$. This scheme is isomorphic to the formal completion of the diagonal $\Delta_{Y_{12}} \subset Y_{12} \times Y_{12}$, *i.e.* $\hat{\Delta}_{Y_{12}} \simeq \mathbf{A}^{d-1} \times \hat{Y}_{12}$, $(\dim\ \hat{Y}_{12} = 0)$.
In this way, the description of $[N]$ would be deduced from a formal local description of the Lefschetz pencil of fiber dimension one $f : \hat{X} \to \mathbf{C}\{\{t\}\}$. Hence, one would get a formal local class representative of N as a bundle over the diagonal cocycle which describes $[N]$ in theorem 4.2. What said so far supports evidence for the following

Conjecture 4.4. *Let $f : X \to S$ be a semistable double points degeneration of fiber dimension d. Then, the local monodromy operator is described by a unique algebraic cocycle of codimension d-1 in the stratum $\tilde{T}^{(3)}$ ($\dim\ \tilde{T}^{(3)} = 2(d-1)$) i.e.*

$$[N] \in \frac{CH^{d-1}(\tilde{T}^{(3)})}{Image\big(\rho^{(3)} : CH^{d-1}(\tilde{T}^{(2)}) \to CH^{d-1}(\tilde{T}^{(3)})\big)}.$$

The formal local description of $[N]$ is given by the algebraic cycle $\Delta_{\mathbf{A}^{d-1}} \times \hat{\tilde{T}}^{(3)}$.

Notice that for a double point degeneration of fiber dimension $d > 1$, $[N]$ may represent the monodromy map acting non trivially on different graded pieces of the limiting cohomology. However, they are all of type $gr^L_{q+1} H^q(\tilde{X}^*, \mathbf{Q}) = E_2^{-1,q+1}(X)$ for $q \in [0, d]$. In fact, for double point degenerations we have always $N = 0$ on $gr^L_q H^q(\tilde{X}^*, \mathbf{Q})$, and $\mathbf{H}^*(Y, gr^L_i A^\bullet_{X,\mathbf{C}}) = 0$ for $i \neq -1,\ 0,\ 1$ because no more than two components of Y intersect simultaneusly at the same closed point.
As an example of application of these results we consider the case of a Lefschetz pencil of fiber dimension at least three. The description of $[N]$ is the same to the one just described for a degeneration with double points. We will only show how to reduce in this case the study of $[N]$ to the previous one. A Lefschetz pencil of fiber dimension greater than one is not even normal-crossings because the special fiber is irreducible and singular. We will only consider the case of odd fiber dimension since Lefschetz pencils of even fiber dimension have trivial monodromy always.
Let $f' : \mathcal{X} \to S$ be such a pencil and let $n = 2m + 1$ be the dimension of its fiber. Locally, in a neighborhood of the singular point of the special fiber $\mathcal{Y}$, the pencil f' is described by

$$f'(z_0, \ldots, z_n) = \sum_{\nu=0}^{m} z_\nu z_{\nu+1+m}$$

where as usual $\{z_0, \dots, z_n\}$ represents a set of regular parameters on $\mathcal{X}$. It is clear from the definition that the special fiber $\mathcal{Y}$ is irreducible and singular at the origin $(z_0, \dots, z_n)$. However, after a single blow-up at that point we get a normal-crossings degeneration $f : X \to S$ with special fiber locally described by $Y = Y_1 \cup Y_2$. The component Y_1 is the exceptional divisor of the blow-up, a projective space of dimension n which intersects the strict transform Y_2 of Y along a quadric hypersurface Y_{12} of dimension $2m$. The component Y_1 appears with multiplicity $e_1 = 2$ whereas Y_2 is reduced (*i.e.* $e_2 = 1$). Let $h : X \to \mathcal{X}$ be the blow-up map. It is a (proper) map of S-schemes, therefore it induces a morphism

$$g^* \mathbf{R}\Psi_{f'}(\mathbf{Q}_{\mathcal{X}}) \to \mathbf{R}\Psi_f(\mathbf{Q}_X)$$

of complexes of nearby cycles. This morphism induces in turn a homomorphism between the corresponding hypercohomologies

$$g^* : \mathbf{H}^i(\mathcal{Y}, \mathbf{R}\Psi_{f'}(\mathbf{Q})) \to \mathbf{H}^i(Y, \mathbf{R}\Psi_f(\mathbf{Q}))$$

In order to work with the resolution $A^\bullet_{\mathbf{Q}}$ of $\mathbf{R}\Psi_f(\mathbf{Q})$ which carries the monodromy filtration, we have to consider Y with its reduced structure (the exceptional divisor has multiplicity $e_1 = 2$ as algebraic cycle on X). Because $g.c.d.(e_1, e_2) = 1 \ \forall y \in Y$ the action of the local monodromy on the complex of sheaves $\mathbf{R}\Psi_f(Y, \mathbf{Q})$ is unipotent (*cf.* § 1). That implies that the monodromy operator acts unipotently on cohomology.
Because f' is a Lefschetz pencil of fiber dimension n, the only group where N acts non trivially is $H^n(\tilde{X}^*, \mathbf{Q})$. Also, $[N]$ determines an element in $(H^{2n}(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q}(n-1)))^{\pi_1}$ and because the generic fibers of f' and f are the same, we may as well consider $[N] \in \mathbf{H}^{2n}(Y \times Y, \mathbf{R}\Psi_f(\mathbf{Q}))^{\pi_1}$.
The map f is locally described by $z_i^2 q(z_0, \dots, \hat{z}_i, \dots, z_n) = t$ for some $i \in [0, n]$, t being a local parameter on S and $q(z_0, \dots, \hat{z}_i, \dots, z_n)$ an irreducible quadratic polynomial. Via the extension of the basis $S' \to S$ $\tau \mapsto \sqrt{t}$, the degeneration f is deformed to $w_i z_i = \tau$, with $w_i = \frac{\tau}{z_i}$ and $w_i^2 = h$. It is clear that this procedure does not affect the special fibers (*i.e.* the reduced closed fibers are the same). Hence, after a possible normalization of the resulting model, we obtain a double point semistable degeneration $h : Z \to S$. Let $T = T_1 \cup T_2$ be its special fiber. Then $[N]$ can be seen as a Hodge cycle in $H^{2n}(T \times T, \mathbf{R}\Psi_h(\mathbf{Q}))^{\pi_1} = \mathrm{Ker}(\tilde{N}) \cap H^{2n}(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q})$, for $\tilde{N} = 1 \otimes N + N \otimes 1$. The geometric description of $[N]$ is then the same as the one we have shown before. The class $[N]$ represents the monodromy operator acting non trivially only on $gr^L_{n+1} H^n(\tilde{X}^*, \mathbf{Q})$.

5. Semistable degenerations with triple points

A semistable degeneration with triple points is the first case where both the operators N and N^2 may be non trivial. In this paragraph we will mainly consider a triple point degeneration of surfaces. The description of $[N]$ and $[N^2]$ for higher dimensional triple points degenerations can be deduced from

the one for surfaces using the same kind of arguments described in the last paragraph for double points degenerations of higher fiber dimension.
Let $f : X \to S$ be a surfaces degeneration with reduced normal crossings and with triple points on its special fiber Y. We keep the basic notations as in the previous sections. Then, locally around a triple point $P \in Y$ we may assume that f has the following description:

$$f(z_1, z_2, z_3) = z_1 z_2 z_3.$$

As usual, $\{z_1, z_2, z_3\}$ is a regular set of parameters on X at P. Globally on X, the special fiber can be the union of more than three components *i.e.* $Y = Y_1 \cup \ldots \cup Y_N$, but at most three of them intersect at the same closed point. The Clemens–Schmid exact sequence of mixed Hodge structures describes the behavior of the operators N and N^2 in terms of some invariants on the special fiber. Namely

Lemma 5.1. *(Monodromy criteria) Let $f : X \to S$ be a semistable degeneration of surfaces, then*

$$N = 0 \text{ on } H^1(\tilde{X}^*, \mathbf{Q}) \Leftrightarrow h^1(|\Gamma|) = 0$$

$$N = 0 \text{ on } H^2(\tilde{X}^*, \mathbf{Q}) \Leftrightarrow h^2(|\Gamma|) = 0 \text{ and } \rho^{(2)} : H^1(\tilde{Y}^{(1)}, \mathbf{Q}) \twoheadrightarrow H^1(\tilde{Y}^{(2)}, \mathbf{Q})$$

$$N^2 = 0 \text{ on } H^2(\tilde{X}^*, \mathbf{Q}) \Leftrightarrow h^2(|\Gamma|) = 0.$$

Here $h^i(|\Gamma|)$ denotes the dimension of the ith-cohomology group of the geometric realization of the dual graph of Y.

Proof. *cf.* [9]. □

A degeneration of K-3 surfaces with special fiber made by rational surfaces intersecting along a cycle of rational curves, is an example for which both N and N^2 are non zero (*cf.* [9]).
Let us suppose that at least one of the groups $gr_2^L H^1(\tilde{X}^*, Q)$ and $gr_3^L H^2(\tilde{X}^*, \mathbf{Q})$ is non zero (for the above example it is well known that $gr_2^L H^1(\tilde{X}^*, \mathbf{Q}) = 0$, as $H^1(\tilde{X}^*, \mathbf{Q}) = 0$). The map N acts on them as an isomorphism of pure Hodge structures

$$N : gr_2^L H^1(\tilde{X}^*, Q) \xrightarrow{\simeq} (gr_0^L H^1(\tilde{X}^*, \mathbf{Q}))(-1)$$

$$N : gr_3^L H^2(\tilde{X}^*, \mathbf{Q}) \xrightarrow{\simeq} (gr_1^L H^2(\tilde{X}^*, \mathbf{Q}))(-1).$$

The only group where N^2 behaves as an isomorphism is $gr_4^L H^2(\tilde{X}^*, \mathbf{Q})$. The map N^2 is defined by the composition

$$gr_4^L H^2(\tilde{X}^*, \mathbf{Q}) \xrightarrow{N} (gr_2^L H^2(\tilde{X}^*, \mathbf{Q}))(-1) \xrightarrow{N} (gr_0^L H^2(\tilde{X}^*, \mathbf{Q}))(-2).$$

The sequence is not exact in the middle. The map N on the left is injective and the one on the right surjects $(gr_2^L H^2(\tilde{X}^*, \mathbf{Q}))(-1)$ onto $(gr_0^L H^2(\tilde{X}^*, \mathbf{Q}))(-2)$. Its kernel, in term of the spectral sequence of weights is

$$\big(\mathrm{Im}(\mathbf{H}^2(Y, gr_1^W \Omega_X^{\bullet+1}(\log Y)) \otimes \mathbf{Q} \to \mathbf{H}^2(Y, A^\bullet_{X,\mathbf{Q}}))\big)(-1) \simeq$$

$$\simeq \frac{\operatorname{Ker}(\rho^{(2)}: H^2(\tilde{Y}^{(1)}, \mathbf{Q})(-1) \to H^2(\tilde{Y}^{(2)}, \mathbf{Q})(-1))}{\operatorname{Im}(\gamma^{(2)}: H^0(\tilde{Y}^{(2)}, \mathbf{Q})(-2) \to H^2(\tilde{Y}^{(1)}, \mathbf{Q})(-1))}.$$

We first consider N and its related class $[N]$. Both $gr_2^L H^1(\tilde{X}^*, \mathbf{Q})$ and $gr_3^L H^2(\tilde{X}^*, \mathbf{Q})$ are described in terms of cohomology classes on $\tilde{Y}^{(2)}$ (*cf.* (1.7)). The study of the correspondence-diagram (2.1) is similar for them. Namely, once one has found an algebraic cycle representing $[N]$, it certainly makes both the correspondence diagrams commute. For degenerations of surfaces it follows from proposition 2.1 that

(5.1)

$$[N] \in (gr_2^L H^4(T, \mathbf{Q}))(1) \simeq \frac{\operatorname{Ker}(\rho^{(4)}: H^2(\tilde{T}^{(3)}, \mathbf{Q})(1) \to H^2(\tilde{T}^{(4)}, \mathbf{Q})(1))}{\operatorname{Im}(\rho^{(3)}: H^2(\tilde{T}^{(2)}, \mathbf{Q})(1) \to H^2(\tilde{T}^{(3)}, \mathbf{Q})(1))}$$

where $h: Z \to S$ is a normal crossings degeneration with special fiber T and generic fiber $\tilde{X}^* \times \tilde{X}^*$ obtained via resolution of the singularities of $X \times_S X$. Similarly, one has

$$[N^2] \in gr_0^L H^4(T, \mathbf{Q}) \simeq \frac{H^0(\tilde{T}^{(5)}, \mathbf{Q})}{\operatorname{Im}(\rho^{(5)}: H^0(\tilde{T}^{(4)}, \mathbf{Q}) \to H^0(\tilde{T}^{(5)}, \mathbf{Q}))}. \tag{5.2}$$

Both $[N]$ and $[N^2]$ have the further property to be Hodge cycles in the cohomologies of the corresponding strata. The following lemma determines the geometry of the model Z and the special fiber T after resolving the singularities of $X \times_S X$ and $Y \times Y$.

LEMMA 5.2. *Let $z_1z_2z_3 = w_1w_2w_3$ be a local description of $X \times_S X$ around the point (P, P), being $P \in Y = \cup_{i=1}^3 Y_i$ a triple point of f and $\{w_1, w_2, w_3\}$ a second set of regular parameters on X at P. After three blows-up of $X \times_S X$ with centers at $z_i = 0 = w_i$ ($i = 1, 2, 3$) the resulting degeneration $h: Z \to S$ is normal–crossings. Its special fiber T is the union of nine irreducible components: $T = \cup_{i=1}^9 T_i$. We number them so that the first six are the strict transforms of the irreducible components $Y_i \times Y_j$ of $Y \times Y$: $T_1 = (Y_1 \times Y_2)\tilde{}$, $T_2 = (Y_1 \times Y_3)\tilde{}$, $T_3 = (Y_2 \times Y_1)\tilde{}$, $T_4 = (Y_2 \times Y_3)\tilde{}$, $T_5 = (Y_3 \times Y_1)\tilde{}$, $T_6 = (Y_3 \times Y_2)\tilde{}$. The last three components are the exceptional divisors of the three blows-up: $T_7 = (Y_1 \times Y_1)\tilde{}$, $T_8 = (Y_2 \times Y_2)\tilde{}$, $T_9 = (Y_3 \times Y_3)\tilde{}$. We have $\tilde{T}^{(1)} = \coprod_i T_i$. The scheme Z is covered by eight affine charts, on each of them there are at most five non empty components T_i. Among the components T_{ijk} whose disjoint union defines the scheme $\tilde{T}^{(3)}$, T_{178} and T_{378} contain resp. the curves "diagonal" $\tilde{\delta}_{12}$ and δ_{12} whose supports project isomorphically onto the diagonal $\Delta_{12}: Y_{12} \to Y_{12} \times Y_{12}$. Similarly, T_{279} and T_{579} contain resp. $\tilde{\delta}_{13}$ and δ_{13} whose support projects isomorphically onto $\Delta_{13}: Y_{13} \to Y_{13} \times Y_{13}$. Finally, T_{489} and T_{689} contain $\tilde{\delta}_{23}$ and δ_{23} whose support is isomorphic to Δ_{23}. The exceptional surface T_{789}–intersection of the three exceptional divisors of h–is isomorphic to the blow-up Bl of $\mathbf{P}^1 \times \mathbf{P}^1$ at the points $\{(0,1) \times (1,0)\}$ and $\{(1,0) \times (0,1)\}$. Finally, the scheme $\tilde{T}^{(5)}$ is the disjoint union of six irreducible components (points). They are: T_{12789}, T_{16789}, T_{24789}, T_{34789}, T_{35789}, T_{56789}. Their support maps isomorphically onto the (point) diagonal $\Delta_{123}: Y_{123} \to Y_{123} \times Y_{123}$.*

Proof. The local description of $X \times_S X$ at (P, P) is given by the equations $z_1z_2z_3 = w_1w_2w_3$ and $z_1z_2z_3 = t$, for $t \in S$ a fixed parameter on the disk. We choose the standard orientation of the sets $\{z_1, z_2, z_3\}$ and $\{w_1, w_2, w_3\}$ and we write $w_i' = \frac{w_i}{z_i}$, $z_i' = \frac{z_i}{w_i}$ for $i = 1, 2, 3$. After three blows-up of $X \times_S X$ along the subvarieties $z_i = 0 = w_i$, the resulting model Z is non singular as one can see by looking at the first of the following tables which describes Z on each of the eight charts $\mathcal{U}_j$ who cover it. In the second table, we have collected for each $\mathcal{U}_j$, the description of the non empty divisors $T_k \in T^{(1)}$ there and the third table shows the "diagonal" curves δ and $\tilde{\delta}$ defined in each chart. The remaining charts describe the pullbacks $p_1^*(\frac{dz_i}{z_i} \wedge \frac{dz_j}{z_j})$, $p_2^*(\frac{dw_i}{w_i} \wedge \frac{dw_j}{w_j})$, $p_1^*(\frac{dz_1}{z_1} \wedge \frac{dz_2}{z_2} \wedge \frac{dz_3}{z_3})$ and $p_2^*(\frac{dw_1}{w_1} \wedge \frac{dw_2}{w_2} \wedge \frac{dw_3}{w_3})$ in terms of the related descriptions by cocycles classes in the corresponding cohomologies.

Open sets	Loc. coordinates and relations
$\mathcal{U}_1$	$\{w_1', w_2', w_3', z_1, z_2, z_3\},\ w_1'w_2'w_3' = 1$
$\mathcal{U}_2$	$\{w_1', w_2', z_1, z_2, w_3\},\ w_1'w_2' = z_3'$
$\mathcal{U}_3$	$\{w_1', w_3', z_1, z_3, w_2\},\ w_1'w_3' = z_2'$
$\mathcal{U}_4$	$\{z_2', z_3', z_1, w_2, w_3\},\ z_2'z_3' = w_1'$
$\mathcal{U}_5$	$\{w_2', w_3', z_2, z_3, w_1\},\ w_2'w_3' = z_1'$
$\mathcal{U}_6$	$\{z_1', z_3', z_2, w_1, w_3\},\ z_1'z_3' = w_2'$
$\mathcal{U}_7$	$\{z_1', z_2', z_3, w_1, w_2\},\ z_1'z_2' = w_3'$
$\mathcal{U}_8$	$\{z_1', z_2', z_3', w_1, w_2, w_3\},\ z_1'z_2'z_3' = 1$

Open sets	Divisors
$\mathcal{U}_1$	$T_7 = \{z_1 = 0\},\ T_8 = \{z_2 = 0\},\ T_9 = \{z_3 = 0\}$
$\mathcal{U}_2$	$T_5 = \{w_1' = 0\},\ T_6 = \{w_2' = 0\},\ T_7 = \{z_1 = 0\},$ $T_8 = \{z_2 = 0\},\ T_9 = \{w_3 = 0\}$
$\mathcal{U}_3$	$T_3 = \{w_1' = 0\},\ T_4 = \{w_3' = 0\},\ T_7 = \{z_1 = 0\},$ $T_8 = \{w_2 = 0\},\ T_9 = \{z_3 = 0\}$
$\mathcal{U}_4$	$T_3 = \{z_2' = 0\},\ T_5 = \{z_3' = 0\},\ T_7 = \{z_1 = 0\},$ $T_8 = \{w_2 = 0\},\ T_9 = \{w_3 = 0\}$
$\mathcal{U}_5$	$T_1 = \{w_2' = 0\},\ T_2 = \{w_3' = 0\},\ T_7 = \{w_1 = 0\},$ $T_8 = \{z_2 = 0\},\ T_9 = \{z_3 = 0\}$
$\mathcal{U}_6$	$T_1 = \{z_1' = 0\},\ T_6 = \{z_3' = 0\},\ T_7 = \{w_1 = 0\},$ $T_8 = \{z_2 = 0\},\ T_9 = \{w_3 = 0\}$
$\mathcal{U}_7$	$T_2 = \{z_1' = 0\},\ T_4 = \{z_2' = 0\},\ T_7 = \{w_1 = 0\},$ $T_8 = \{w_2 = 0\},\ T_9 = \{z_3 = 0\}$
$\mathcal{U}_8$	$T_7 = \{w_1 = 0\},\ T_8 = \{w_2 = 0\},\ T_9 = \{w_3 = 0\}$

Open sets	"Diagonal" curves
$\mathcal{U}_1$	none
$\mathcal{U}_2$	$\delta_{13}=\{w_1'=z_1=w_3=0,\ w_2'=1\}\subset T_{579},\ \delta_{13}\cap T_8\neq\emptyset$ $\delta_{23}=\{w_2'=z_2=w_3=0,\ w_1'=1\}\subset T_{689},\ \delta_{23}\cap T_7\neq\emptyset$
$\mathcal{U}_3$	$\delta_{12}=\{w_1'=z_1=w_2=0,\ w_3'=1\}\subset T_{378},\ \delta_{12}\cap T_9\neq\emptyset$ $\tilde\delta_{23}=\{w_3'=z_3=w_2=0,\ w_1'=1\}\subset T_{489},\ \tilde\delta_{23}\cap T_7\neq\emptyset$
$\mathcal{U}_4$	$\delta_{12}=\{z_2'=z_1=w_2=0,\ z_3'=1\}\subset T_{378},\ \delta_{12}\cap T_9\neq\emptyset$ $\delta_{13}=\{z_3'=z_1=w_3=0,\ z_2'=1\}\subset T_{579},\ \delta_{13}\cap T_8\neq\emptyset$
$\mathcal{U}_5$	$\tilde\delta_{12}=\{w_2'=z_2=w_1=0,\ w_3'=1\}\subset T_{178},\ \tilde\delta_{12}\cap T_9\neq\emptyset$ $\tilde\delta_{13}=\{w_3'=z_3=w_1=0,\ w_2'=1\}\subset T_{279},\ \tilde\delta_{13}\cap T_8\neq\emptyset$
$\mathcal{U}_6$	$\tilde\delta_{12}=\{z_1'=z_2=w_1=0,\ z_3'=1\}\subset T_{178},\ \tilde\delta_{12}\cap T_9\neq\emptyset$ $\delta_{23}=\{z_3'=z_2=w_3=0,\ z_1'=1\}\subset T_{689},\ \tilde\delta_{23}\cap T_7\neq\emptyset$
$\mathcal{U}_7$	$\tilde\delta_{13}=\{z_1'=z_3=w_1=0,\ z_2'=1\}\subset T_{279},\ \tilde\delta_{13}\cap T_8\neq\emptyset$ $\tilde\delta_{23}=\{z_2'=z_3=w_2=0,\ z_1'=1\}\subset T_{489},\ \tilde\delta_{23}\cap T_7\neq\emptyset$
$\mathcal{U}_8$	none

Denote by $v_{Y_{ij}}$ a class in $H^*(Y_{ij},\mathbf{C})$ and by $v_{T_{lk}}$ a class in $H^*(\tilde T^{(2)},\mathbf{C})$. Then we have

Open sets	$p_1^*(v_{Y_{12}})$	$p_2^*(v_{Y_{12}})$
$\mathcal{U}_1$	$v_{T_{78}}$	$v_{T_{78}}$
$\mathcal{U}_2$	$v_{T_{78}}$	$v_{T_{56}}+v_{T_{58}}-v_{T_{67}}+v_{T_{78}}$
$\mathcal{U}_3$	$-v_{T_{37}}-v_{T_{47}}+v_{T_{78}}$	$v_{T_{38}}+v_{T_{78}}$
$\mathcal{U}_4$	$-v_{T_{37}}+v_{T_{78}}$	$v_{T_{78}}+v_{T_{38}}+v_{T_{58}}$
$\mathcal{U}_5$	$v_{T_{18}}+v_{T_{28}}+v_{T_{78}}$	$v_{T_{78}}-v_{T_{17}}$
$\mathcal{U}_6$	$v_{T_{18}}+v_{T_{78}}$	$v_{T_{78}}-v_{T_{17}}-v_{T_{67}}$
$\mathcal{U}_7$	$v_{T_{24}}+v_{T_{28}}-v_{T_{47}}+v_{T_{78}}$	$v_{T_{78}}$
$\mathcal{U}_8$	$v_{T_{78}}$	$v_{T_{78}}$

Hence, the global description of the pullbacks $p_1^*(v_{Y_{12}})$ and $p_2^*(v_{Y_{12}})$ are

$$p_1^*(v_{Y_{12}}) = (v_{T_{18}} + v_{T_{28}} - v_{T_{37}} - v_{T_{47}} + v_{T_{78}}) + v_{T_{24}}$$
$$p_2^*(v_{Y_{12}}) = (-v_{T_{17}} + v_{T_{38}} + v_{T_{58}} - v_{T_{67}} + v_{T_{78}}) + v_{T_{56}}.$$

Open sets	$p_1^*(v_{Y_{13}})$	$p_2^*(v_{Y_{13}})$
$\mathcal{U}_1$	$v_{T_{79}}$	$v_{T_{79}}$
$\mathcal{U}_2$	$-v_{T_{57}} - v_{T_{67}} + v_{T_{79}}$	$v_{T_{79}} + v_{T_{59}}$
$\mathcal{U}_3$	$v_{T_{79}}$	$v_{T_{79}} - v_{T_{47}} + v_{T_{39}} + v_{T_{34}}$
$\mathcal{U}_4$	$-v_{T_{57}} + v_{T_{79}}$	$v_{T_{79}} + v_{T_{39}} + v_{T_{59}}$
$\mathcal{U}_5$	$v_{T_{19}} + v_{T_{29}} + v_{T_{79}}$	$v_{T_{79}} - v_{T_{27}}$
$\mathcal{U}_6$	$v_{T_{16}} + v_{T_{19}} - v_{T_{67}} + v_{T_{79}}$	$v_{T_{79}}$
$\mathcal{U}_7$	$v_{T_{29}} + v_{T_{79}}$	$v_{T_{79}} - v_{T_{27}} - v_{T_{47}}$
$\mathcal{U}_8$	$v_{T_{79}}$	$v_{T_{79}}$

Hence we have the global descriptions

$$p_1^*(v_{Y_{13}}) = (v_{T_{19}} + v_{T_{29}} - v_{T_{57}} - v_{T_{67}} + v_{T_{79}}) + v_{T_{16}}$$
$$p_2^*(v_{Y_{13}}) = (-v_{T_{27}} + v_{T_{39}} - v_{T_{47}} + v_{T_{59}} + v_{T_{79}}) + v_{T_{34}}.$$

Open sets	$p_1^*(v_{Y_{23}})$	$p_2^*(v_{Y_{23}})$
$\mathcal{U}_1$	$v_{T_{89}}$	$v_{T_{89}}$
$\mathcal{U}_2$	$-v_{T_{58}} - v_{T_{68}} + v_{T_{89}}$	$v_{T_{89}} + v_{T_{69}}$
$\mathcal{U}_3$	$v_{T_{39}} + v_{T_{49}} + v_{T_{89}}$	$v_{T_{89}} - v_{T_{48}}$
$\mathcal{U}_4$	$v_{T_{35}} + v_{T_{39}} - v_{T_{58}} + v_{T_{89}}$	$v_{T_{89}}$
$\mathcal{U}_5$	$v_{T_{89}}$	$v_{T_{89}} - v_{T_{28}} + v_{T_{19}} + v_{T_{12}}$
$\mathcal{U}_6$	$-v_{T_{68}} + v_{T_{89}}$	$v_{T_{89}} + v_{T_{19}} + v_{T_{69}}$
$\mathcal{U}_7$	$v_{T_{49}} + v_{T_{89}}$	$v_{T_{89}} - v_{T_{28}} - v_{T_{48}}$
$\mathcal{U}_8$	$v_{T_{89}}$	$v_{T_{89}}$

Finally we have

$$p_1^*(v_{Y_{23}}) = (v_{T_{39}} + v_{T_{49}} - v_{T_{58}} - v_{T_{68}} + v_{T_{89}}) + v_{T_{35}}$$
$$p_2^*(v_{Y_{23}}) = (v_{T_{19}} - v_{T_{28}} - v_{T_{48}} + v_{T_{69}} + v_{T_{89}}) + v_{T_{12}}.$$

Using the above tables we deduce the following

Open sets	$p_1^*(1_{Y_{123}})$	$p_2^*(1_{Y_{123}})$
$\mathcal{U}_1$	$1_{T_{789}}$	$1_{T_{789}}$
$\mathcal{U}_1$	$1_{T_{578}} + 1_{T_{678}} + 1_{T_{789}}$	$1_{T_{569}} + 1_{T_{589}} - 1_{T_{679}} + 1_{T_{789}}$
$\mathcal{U}_3$	$-1_{T_{379}} - 1_{T_{479}} + 1_{T_{789}}$	$-1_{T_{348}} + 1_{T_{389}} + 1_{T_{478}} + 1_{T_{789}}$
$\mathcal{U}_4$	$1_{T_{357}} - 1_{T_{379}} + 1_{T_{578}} + v_{T_{789}}$	$1_{T_{389}} + 1_{T_{589}} + 1_{T_{789}}$
$\mathcal{U}_5$	$1_{T_{189}} + 1_{T_{289}} + 1_{T_{789}}$	$1_{T_{127}} - 1_{T_{179}} + 1_{T_{278}} + v_{T_{789}}$
$\mathcal{U}_6$	$-1_{T_{168}} + 1_{T_{189}} + 1_{T_{678}} + 1_{T_{789}}$	$-1_{T_{179}} - 1_{T_{679}} + v_{T_{789}}$
$\mathcal{U}_7$	$1_{T_{249}} + 1_{T_{289}} - 1_{T_{479}} + 1_{T_{789}}$	$1_{T_{278}} + 1_{T_{478}} - v_{T_{789}}$
$\mathcal{U}_8$	$1_{T_{789}}$	$1_{T_{789}}$

We then obtain

$$\begin{aligned} p_1^*(1_{Y_{123}}) &= (1_{T_{189}} + 1_{T_{249}} + 1_{T_{289}} - 1_{T_{379}} - 1_{T_{479}} + 1_{T_{789}}) \\ &\quad - 1_{T_{168}} + 1_{T_{357}} + 1_{T_{578}} + 1_{T_{678}} \\ p_2^*(1_{Y_{123}}) &= (-1_{T_{179}} + 1_{T_{389}} + 1_{T_{569}} + 1_{T_{589}} - 1_{T_{679}} + 1_{T_{789}}) \\ &\quad + 1_{T_{127}} + 1_{T_{278}} - 1_{T_{348}} + 1_{T_{478}}. \end{aligned}$$

Notice that with the exception of $\mathcal{U}_1$ and $\mathcal{U}_8$ that are open sets in $\mathbf{A}^5$ on which only the exceptional components T_7, T_8 and T_9 are non empty, all the remaining charts $\mathcal{U}_j$ are isomorphic to $\mathbf{A}^5$ and in each of them one has five components T_k non empty.

On $\mathcal{U}_3 \cap \mathcal{U}_4$ the surface T_{378} contains the curve δ_{12}, and on $\mathcal{U}_5 \cap \mathcal{U}_6$, T_{178} contains the curve $\tilde{\delta}_{12}$. The curves δ_{12} and $\tilde{\delta}_{12}$ are different: *i.e.* $T_1 = \emptyset$ on $\mathcal{U}_3$ and $\mathcal{U}_4$, but their supports map isomorphically onto the same diagonal $\Delta_{12} : Y_{12} \to Y_{12} \times Y_{12}$.

Similarly, $\mathcal{U}_2 \cap \mathcal{U}_4$ contains δ_{13} whose support maps isomorphically onto Δ_{13}, whereas $\mathcal{U}_5 \cap \mathcal{U}_7$ contains $\tilde{\delta}_{13}$, whose support maps still isomorphically onto Δ_{13}: $\delta_{13} \cap \tilde{\delta}_{13} = \emptyset$.

Finally, $\delta_{23} \subset \mathcal{U}_2 \cap \mathcal{U}_6$, $\delta_{23} \simeq \Delta_{23}$, while $\tilde{\delta}_{23} \subset \mathcal{U}_3 \cap \mathcal{U}_7$, $\tilde{\delta}_{23} \simeq \Delta_{23}$ and $\delta_{23} \cap \tilde{\delta}_{23} = \emptyset$.

The blow-up Z_1 of $X \times_S X$ at $z_1 = 0 = w_1$ is the strict transform of $X \times_S X$ in the blow-up of $\mathbf{A}^6$ along the corresponding linear subvariety. Let $(\tilde{z}_1, \tilde{w}_1)$ be a

couple of homogeneus coordinates. The exceptional divisor, say $E_1^{(1)}$, is locally a $\mathbf{P}^1_{(\tilde{z}_1,\tilde{w}_1)}$–bundle over $\{z_1 = 0 = w_1\}$. Then, the intersection $E_1^{(1)} \cap Z_1$ is locally defined on $E_1^{(1)}$ by $z_2 z_3 \tilde{z}_1 - w_2 w_3 \tilde{w}_1 = 0$. The blow $E_1^{(2)}$ of $\mathbf{P}^1 \times \{z_1 = 0 = w_1\}$ on $\mathbf{P}^1 \times \{z_1 = z_2 = w_1 = w_2 = 0\}$ defines the strict transform of $E_1^{(1)}$ after the second blow-up along $\{z_2 = w_2\}$. Said $E_2^{(2)}$ the exceptional divisor of the second blow-up and $(\tilde{z}_2, \tilde{w}_2)$ another couple of homogeneus coordinates, one has $E_1^{(2)} \cap E_2^{(2)} = \mathbf{P}^1_{(\tilde{z}_1,\tilde{w}_1)} \times \mathbf{P}^1_{(\tilde{z}_2,\tilde{w}_2)} \times \{z_1 = z_2 = w_1 = w_2 = 0\}$. Finally, after the third blowing at $\{z_3 = 0 = w_3\}$ the three exceptional divisors $E_1^{(3)}$, $E_2^{(3)}$ and $E_3^{(3)}$ will intersect the strict transform Z of $X \times_S X$ along the exceptional surface T_{789}. This surface is described by the equation $\tilde{z}_1 \tilde{z}_2 \tilde{z}_3 - \tilde{w}_1 \tilde{w}_2 \tilde{w}_3 = 0$ in $E_1^{(3)} \cap E_2^{(3)} \cap E_3^{(3)} = \mathbf{P}^1_{(\tilde{z}_1,\tilde{w}_1)} \times \mathbf{P}^1_{(\tilde{z}_2,\tilde{w}_2)} \times \mathbf{P}^1_{(\tilde{z}_3,\tilde{w}_3)} \times \{z_1 = z_2 = z_3 = w_1 = w_2 = w_3 = 0\} = (\mathbf{P}^1)^3$, $(\tilde{z}_3, \tilde{w}_3)$ being a third couple of homogeneus coordinates. Let consider the projection $T_{789} \to \mathbf{P}^1_{(\tilde{z}_2,\tilde{w}_2)} \times \mathbf{P}^1_{(\tilde{z}_3,\tilde{w}_3)}$. The fiber of this map over a given point in the base $(\mathbf{P}^1)^2$ is defined by a linear equation as $\alpha z_1 - \beta w_1 = 0$. If either α or β (or both) is not zero, then this fiber is reduced to a single point, so the projection map is locally an isomorphism. On the other hand, $\alpha = 0 = \beta$ happens over the two points $(1,0) \times (0,1)$ and $(0,1) \times (1,0)$, where the fiber is a $\mathbf{P}^1$. Since T_{789} is non singular, these two copies of $\mathbf{P}^1$ are Cartier divisors, so by the universal property of blow-ups the map factors through the blow-up Bl of $(\mathbf{P}^1)^2$ at the two points (*i.e.* $T_{789} \to Bl \to (\mathbf{P}^1)^2$). It is easy to see from this description that $T_{789} \simeq Bl$.
It is straighforward to verify from the second table the description of $\tilde{T}^{(5)}$ on each chart $\mathcal{U}_j$ and the statement concerning its support. $\square$

The following result generalizes the description of $[N]$ given in theorem 4.2 for double points degenerations.

Theorem 5.3. *Let $f : X \to S$ be a semistable degeneration of surfaces as we have considered above. With the same notations as in lemma 5.2, let $\pi : Bl \to \mathbf{P}^1 \times \mathbf{P}^1$ be the morphism definying the blow-up of $\mathbf{P}^1 \times \mathbf{P}^1$ at the points $\{(0,1) \times (1,0)\}$ and $\{(1,0) \times (0,1)\}$, being $Bl \simeq T_{789}$. Let $F_1 = \pi^*(\{pt\} \times \mathbf{P}^1)$ and $F_2 = \pi^*(\mathbf{P}^1 \times \{pt\})$ be the two fundamental fibers and let E_1 and E_2 be the two exceptional divisors of π. The following description of $[N] \in Ker\, \rho^{(4)}$ (cf.* (5.1)*) holds:*

$$[N] = a_{178}\tilde{\delta}_{12} + a_{279}\tilde{\delta}_{13} + a_{378}\delta_{12} + a_{489}\tilde{\delta}_{23} + a_{579}\delta_{13} + a_{689}\delta_{23} + \Gamma.$$

The 1-cycle $\Gamma \subset Bl$ and the (rational) numbers a's are subject to the following requirements:

$$\Gamma = xF_1 + yF_2 + zE_1 + wE_2, \qquad \text{with} \quad w = z - 1, \quad x, y, z, w \in \mathbf{Q}$$

$$a_{178} - a_{378} = a_{279} - a_{579} = a_{489} - a_{689} = 1$$

and the relations among them are given by the following set of equalities

$$a_{178} = -w, \qquad a_{279} = -(y+w), \qquad a_{378} = -z,$$
$$a_{489} = x+z, \qquad a_{579} = -(y+z), \qquad a_{689} = x+w.$$

Furthermore, for those degenerations with $N^2 \neq 0$, the class $[N^2] \in E_1^{0,4}(Z) = H^0(\tilde{T}^{(5)}, \mathbf{Q})$ (cf. (5.2)) can be exhibited as:

$$\begin{aligned}[N^2] = {} & b_{12789}T_{12789} + b_{16789}T_{16789} + b_{24789}T_{24789} \\ & + b_{34789}T_{34789} + b_{35789}T_{35789} + b_{56789}T_{56789}.\end{aligned}$$

The (rational) numbers b's must satisfy the following equation:

$$-b_{12789} + b_{16789} - b_{24789} + b_{34789} - b_{35789} - b_{56789} = 1.$$

Hence, the induced classes of $[N]$ in $gr_2^L H^4(T, \mathbf{Q})(1)$ and of $[N^2]$ in $gr_0^L H^4(T, \mathbf{Q})$ (i.e. modulo boundary relations via the restriction maps $\rho^{(3)}$ and $\rho^{(5)}$ cf. (1.6)) determine algebraic cocycles of dimension one and zero respectively.

Proof. We will determine $[N]$ as a cocycle making the following square commute (*i.e.* this is the one one has to study for a degeneration of K-3 surfaces of the type mentioned above)

$$\begin{array}{ccc} gr_3^L H^2(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q}) & \xrightarrow{[N]\,\cdot} & gr_5^L H^6(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q})(1) = (E_2^{1,5})(1) \\ {\scriptstyle (p_1)^*}\uparrow & & \downarrow{\scriptstyle (p_2)_*} \\ E_2^{-1,3} = gr_3^L H^2(\tilde{X}^*, \mathbf{Q}) & \xrightarrow{N} & (gr_1^L H^2(\tilde{X}^*, \mathbf{Q}))(-1) = (E_2^{1,1})(-1) \end{array}$$

Note that besides the commutativity of the square, one has to impose another condition on $[N]$ in order for it to represent the operator N. That arises from (5.1). Namely, the representative of N in $(E_1^{2,2})(1) = H^2(\tilde{T}^{(3)}, \mathbf{Q})(1)$ must belong to the kernel of the related restriction map $\rho^{(4)}$. This condition was automatically satisfied for double point degenerations since $T^{(4)} = \emptyset$ always in that case. We will explicitly describe a representative $[N]$ of N in $(E_1^{2,2})(1)$ that satisfies the commutativity of the following square

$$(5.3) \qquad \begin{array}{ccc} H^1(\tilde{T}^{(2)}, \mathbf{Q})(-1) & \xrightarrow{[N]\,\cdot} & H^5(\tilde{T}^{(2)}, \mathbf{Q})(1) \\ {\scriptstyle p_1^*}\uparrow & & \downarrow{\scriptstyle (p_2)_*} \\ H^1(\tilde{Y}^{(2)}, \mathbf{Q})(-1) & = & H^1(\tilde{Y}^{(2)}, \mathbf{Q})(-1). \end{array}$$

With the notations introduced in lemma 5.2 we first remark that the cocycles $[\delta_{ij}] = (\Delta_{ij})_*(1_{Y_{ij}})$ $(i, j = 1, 2, 3,\ i \neq j)$, $\Delta_{ij} : Y_{ij} \to Y_{ij} \times Y_{ij}$ being the diagonal embedding, evidently satisfy the cohomological equality

$$(p_2)_*(\Delta_*(1_{Y_{ij}}) \cdot (p_1)^*(v)) = (p_2)_*(\Delta_* \Delta^* p_1^*(v)) = (p_2)_*(\Delta_*(v)) = v$$

for $1_{Y_{ij}} \in H^0(Y_{ij}, \mathbf{Q})$ and any element $v \in H^1(\tilde{Y}^{(2)}, \mathbf{Q})(-1)$. However, since a simple linear combination as $a_{178}\tilde{\delta}_{12} + a_{279}\tilde{\delta}_{13} + a_{378}\delta_{12} + a_{489}\tilde{\delta}_{23} + a_{579}\delta_{13} + a_{689}\delta_{23}$ (the coefficients a's are integers) does not satisfy the requirement of being in the kernel of the restriction map $\rho^{(4)}$ (*cf.* (5.1) and (1.6)), we have to add to the above "diagonal" definition a 1-cocycle $\Gamma \subset T_{789}$, so that the completed linear combination defines an element in $(E_2^{2,2})(1)$ representing N. Notice that since the exceptional surface T_{789} projects down via p_2, onto the triple point P, this modification by Γ does not spoil the commutativity of (5.3), once we have checked it for the partial representative of $[N]$ given in terms of the above diagonals.

The 1-cycle Γ will be described as a combination of the generators $F_1,\ F_2,\ E_1,\ E_2$ of the Neron-Severi group $NS(T_{789})$. First of all, let consider the six curves T_{k789} for $k = 1, \dots, 6$. They are elements of $\tilde{T}^{(4)}$. We describe them using the generators of $NS(T_{789})$. Because $\pi(T_{1789}) = \{(0,1) \times (1,0)\}$, $T_{1789} = E_2$. Similarly, we have $T_{3789} = E_1$, as $\pi(T_{3789}) = \{(1,0) \times (0,1)\}$. The remaining four curves are described using the projection formula. For example, we know that $\pi(T_{2789}) = (0,1) \times \mathbf{P}^1$ and that $\pi^*((0,1) \times \mathbf{P}^1) = F_1 = E_2 + T_{2789}$. Hence we have $T_{2789} = F_1 - E_2$. With a similar procedure we obtain $T_{4789} = F_2 - E_1$, $T_{5789} = F_1 - E_1$ and $T_{6789} = F_2 - E_2$. The geometry of the intersections among the generators of $NS(T_{789})$ is well known, namely $E_1 \cdot E_2 = E_1 \cdot F_2 = E_1 \cdot F_1 = E_2 \cdot F_1 = E_2 \cdot F_2 = F_1 \cdot F_1 = F_2 \cdot F_2 = 0$, $E_1 \cdot E_1 = -1 = E_2 \cdot E_2$ and $F_1 \cdot F_2 = 1$.

Let $\Gamma = xF_1 + yF_2 + zE_1 + wE_2$ be an element of $NS(T_{789})$, with $x, y, z, w \in \mathbf{Q}$. Then, we must solve

$$[N] = a_{178}\tilde{\delta}_{12} + a_{279}\tilde{\delta}_{13} + a_{378}\delta_{12} + a_{489}\tilde{\delta}_{23} + a_{579}\delta_{13} + a_{689}\delta_{23} + \Gamma$$

for Γ subject to the condition that $[N]$ is in $\ker \rho^{(4)}$, for $\rho^{(4)} = \sum_{u=1}^{4} (-1)^{u-1} \rho_u^{(4)}$ (*cf.* (1.6)). For example we have $\rho^{(4)}(a_{178}\tilde{\delta}_{12}) = -a_{178}(\tilde{\delta}_{12} \cdot T_9)$, while $\rho^{(4)}(a_{279}\tilde{\delta}_{13}) = a_{279}(\tilde{\delta}_{13} \cdot T_8)$. Following these rules we obtain the system

$$(5.4) \qquad \begin{aligned} &a_{178} = \Gamma \cdot T_{1789} = -w, \qquad a_{279} = -\Gamma \cdot T_{2789} = -(y+w) \\ &a_{378} = \Gamma \cdot T_{3789} = -z, \qquad a_{489} = \Gamma \cdot T_{4789} = x + z \\ &a_{579} = -\Gamma \cdot T_{5789} = -(y+z), \qquad a_{689} = \Gamma \cdot T_{6789} = x + w. \end{aligned}$$

For the standard choice of the orientations of $\{z_1, z_2, z_3\}$ and $\{w_1, w_2, w_3\}$ and the numbering of the T_i's setted in lemma 5.2, the local description of the pullbacks $\frac{dz_i}{z_i} \wedge \frac{dz_j}{z_j}$ and $\frac{dw_i}{w_i} \wedge \frac{dw_j}{w_j}$ ($i \neq j,\ i,j = 1,2,3$) in terms of cohomology classes $v_{T_{ij}}$ and $v_{T_{ijk}}$, is given following the tables shown in the proof of lemma 5.2.

Let $v_{ij} \in H^1(\tilde{Y}^{(2)}, \mathbf{Q})(-1)$, then via the multiplicative rule described in the Appendix (*cf.* the similar calculation done in the proof of theorem 4.2) we obtain

$$[N] \cdot p_1^*(v_{12} + v_{13} + v_{23}) =$$
$$= [N] \cdot (v_{T_{18}} + v_{T_{78}} + v_{T_{29}} + v_{T_{79}} + v_{T_{49}} + v_{T_{89}}) =$$
$$= a_{178}g_1(\tilde{\delta}_{12} \cdot v_{T_{18}}) - a_{378}g_7(\delta_{12} \cdot v_{T_{78}}) + a_{279}g_2(\tilde{\delta}_{13} \cdot v_{T_{29}}) - a_{579}g_7(\delta_{13} \cdot v_{T_{79}}) +$$
$$+ a_{489}(g_4(\tilde{\delta}_{23} \cdot v_{T_{49}}) - a_{689}(g_8(\delta_{23} \cdot v_{T_{89}}) =$$
$$= a_{178}v_{78}(1) - a_{378}v_{38}(1) + a_{279}v_{79}(1) - a_{579}v_{59}(1) + a_{489}v_{89}(1) - a_{689}v_{69}(1)$$

where g_j are the pushforward maps defined in the Appendix. Applying the map $(p_2)_*$ we have

$$(p_2)_*([N]\cdot p_1^*(v_{12}+v_{13}+v_{23})) \\ = (a_{178}-a_{378})v_{12}+(a_{279}-a_{579})v_{13}+(a_{489}-a_{689})v_{23}.$$

The commutativity of the diagram (5.3) is then equivalent to the requirement

$$a_{178}-a_{378}=a_{279}-a_{579}=a_{489}-a_{689}=1 \tag{5.5}$$

The linear system (5.4) may be then read as $z-w=1$. Therefore, any curve $\Gamma = xF_1+yF_2+zE_1+wE_2$ satisfying the condition $z-w=1$ can be used in the description of $[N]\in(E_1^{2,2})(1)$.
The description of $[N^2]$ is similar. For instance, from proposition 2.1 we have

$$[N^2]\in gr_0^L H^4(T,\mathbf{Q})\simeq \frac{H^0(\tilde{T}^{(5)},\mathbf{Q})}{\mathrm{Im}(\rho^{(5)}:H^0(\tilde{T}^{(4)},\mathbf{Q})\to H^0(\tilde{T}^{(5)},\mathbf{Q}))}$$

Via the procedure described in (2.1), $[N^2]$ is then determined in terms of the commutativity of the following square

$$\begin{array}{ccc} gr_4^L H^2(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}) & \xrightarrow{[N^2]\,\cdot} & gr_4^L H^6(\tilde{X}^*\times\tilde{X}^*,\mathbf{Q}) = E_2^{2,4} \\ {\scriptstyle (p_1)^*}\uparrow & & \downarrow{\scriptstyle (p_2)_*} \\ E_2^{-2,4}=gr_4^L H^2(\tilde{X}^*,\mathbf{Q}) & \xrightarrow{N^2} & (gr_0^L H^2(\tilde{X}^*,\mathbf{Q}))(-2)=(E_2^{2,0})(-2). \end{array}$$

The related E_1 description is

$$\begin{array}{ccc} H^0(\tilde{T}^{(3)},\mathbf{Q})(-2) & \xrightarrow{[N^2]\,\cdot} & H^4(\tilde{T}^{(3)},\mathbf{Q}) \\ {\scriptstyle p_1^*}\uparrow & & \downarrow{\scriptstyle (p_2)_*} \\ H^0(\tilde{Y}^{(3)},\mathbf{Q})(-2) & = & H^0(\tilde{Y}^{(3)},\mathbf{Q})(-2). \end{array}$$

The scheme $\tilde{T}^{(5)}$ is the disjoint union of the zero dimensional schemes T_{12789}, T_{16789}, T_{35789} and T_{56789}. Their support map all isomorphically onto the diagonal $\Delta_{123}: Y_{123}\to Y_{123}\times Y_{123}$. With a similar procedure as the one used above to describe $[N]$, we write

$$[N^2]=b_{12789}T_{12789}+b_{16789}T_{16789}+b_{24789}T_{24789} \\ +b_{34789}T_{34789}+b_{35789}T_{35789}+b_{56789}T_{56789}$$

for some integers b's. Imposing the commutativity of the above diagram, by means of the description of the pullbacks $p_1^*(1_{Y_{123}})$ and $p_2^*(1_{Y_{123}})$ as shown in the last table appearing in the proof of lemma 5.2, we finally get the condition

$$-b_{12789}+b_{16789}-b_{24789}+b_{34789}-b_{35789}-b_{56789}=1.$$

□

It is straightforward to verify that both $[N]$ and $[N^2]$ make diagrams like (2.1) commute, for any choice of the indices $*$ and r.

Remark 5.4.

It is easy to verify that the description of $[N]$ and $[N^2]$ given in theorem 5.3 holds also for a normal–crossings degeneration (not semistable) like $f(z_1,\dots,z_n) = z_i^2 z_j$, $i,j \in [1,n]$, $i \neq j$. This applies in particular to the case of normal–crossings degenerations of curves with triple points as described above. The desingularization process of the threefold $X \times_S X$ is obtained via two blow-ups along $z_i = 0 = w_i$ and $z_j = 0 = w_j$ by analogy to what we have done in Remark 4.3. For the description of $[N]$ we also refer to the same Remark.

6. An arithmetic interpretation of the monodromy operator in mixed characteristic

The calculations on the geometric description of $[N^i]$ that we have done in the previous sections only involve the (local) geometry of the special fiber of a degeneration. Hence they equally hold in mixed characteristic also, *i.e.* for a degeneration $f : \mathcal{X} \to \mathrm{Spec}(\Lambda) = S$, where Λ is a Henselian discrete valuation ring with η and v as its generic and closed points respectively. In analogy with the classical case, the model $\mathcal{X}$ is assumed to be proper and the map f is supposed to be flat, smooth over the generic point η and with a normal–crossings special fiber Y defined over the *finite* field $k(v)$ of characteristic $p > 0$. Locally, for the étale topology $\mathcal{X}$ is S-isomorphic to $S[x_1,\dots,x_n]/(x_1^{e_1}\cdots x_k^{e_k} - \pi)$, where π is a uniformizing parameter in Λ and $e_i \in \mathbf{Z}$, $\forall i = 1,\dots,k$. For simplicity, we also assume that Λ is a *finite extension* of $\mathbf{Z}_\ell$ or $\mathbf{Q}_\ell$, where $l \neq p$ is a prime number.

The complex of nearby cycles is then defined as $\mathbf{R}\Psi(\Lambda) := \bar{i}^{-1}\mathbf{R}\bar{j}_*\Lambda$. Here $i : Y \to \mathcal{X}$ (*resp.* $j : \mathcal{X}_\eta \to \mathcal{X}$) is the natural closed (*resp.* open) embedding that one "extends" to the algebraic closure $k(\bar{v})$ of $k(v)$ (*resp.* a separable closure $k(\bar{\eta})$ of $k(\eta)$). Assume that the multiplicities e_i are prime to ℓ and $g.c.d.(e_i, p) = 1$. Then, the wild inertia acts trivially on $\mathbf{R}\Psi(\Lambda)$ and the theory exposed in [16] shows that the nearby cycle complex has an abstract description in the derived category $D^+(Y, \Lambda[\mathbf{Z}_\ell(1)])$ of the abelian category of complexes of sheaves of $\Lambda[\mathbf{Z}_\ell(1)]$–modules on Y, by a complex $A^\bullet_{\mathcal{X},\Lambda}$, supported on Y. $A^\bullet_{\mathcal{X},\Lambda}$ can be interpreted as the analogue of the Steenbrink resolution in the classical case. Therefore, the related study of it goes in parallel with the classical one in equal characteristic zero. We refer to *op.cit.* and [7] (*e.g.* Théorème 3.2) for further detail.

The power maps ($n \in [0,2d]$, $i \geq 0$, $d = \dim\ \mathcal{X}_\eta$) $N^i : H^n(\mathcal{X}_{\bar{\eta}}, \Lambda) \to H^n(\mathcal{X}_{\bar{\eta}}, \Lambda)(-i)$ define elements

$$N^i \in \bigoplus_{n\geq 0}\left[H^{2d-n}(\mathcal{X}_{\bar{\eta}}, \Lambda)(d) \otimes H^n(\mathcal{X}_{\bar{\eta}}, \Lambda)(-i)\right]^G = \left[H^{2d}(\mathcal{X}_{\bar{\eta}} \times \mathcal{X}_{\bar{\eta}}, \Lambda)(d-i)\right]^G$$

invariant for the action of the Galois group $G = \mathrm{Gal}(\bar{\eta}/\eta)$ on the cohomology of the product $\mathcal{X}_{\bar{\eta}} \times \mathcal{X}_{\bar{\eta}}$. Assume that $f : \mathcal{X} \to S$ has at worst triple points. Then, the singularities of both $\mathcal{X} \times_S \mathcal{X}$ and $Y \times Y$ can be resolved locally around each singular point by a sequence of at most three blows-up, as we described in details in §§ 2,4,5. The resulting degeneration $h : \mathcal{Z} \to S$ is normal–crossings with special fiber $T = T_1 \cup \ldots \cup T_N$. Let $\mathcal{X}_{\bar\eta} \times \mathcal{X}_{\bar\eta} = \mathcal{Z}_{\bar\eta}$ be its geometric generic fiber. Denote by $\tilde{N} = 1 \otimes N + N \otimes 1$ the logarithm of the local monodromy on the product degeneration h. Then, the analogue of proposition 2.1 is the following

Proposition 6.1. *Assume the monodromy-weight conjecture on $H^*(\mathcal{Z}_{\bar\eta}, \Lambda)$ and the semisimplicity of the Frobenius on the inertia invariants. Then*

$$N^i \in \left[Ker(\tilde{N}) \cap H^{2d}(\mathcal{Z}_{\bar\eta}, \Lambda(d-i))\right]^{F=1}$$
$$\simeq \left[Ker(\tilde{N}) \cap (gr^L_{2(d-i)} H^{2d}(\mathcal{Z}_{\bar\eta}, \Lambda))(d-i)\right]^{F=1}$$
$$\simeq \left((gr^L_{2(d-i)} H^{2d}(T, \Lambda))(d-i)\right)^{F=1}$$
$$\simeq$$
$$\left[\frac{Ker(\rho^{(2(i+1)} : H^{2(d-i)}(\tilde{T}^{(2i+1)}, \Lambda)(d-i) \to H^{2(d-i)}(\tilde{T}^{(2(i+1))}, \Lambda)(d-i))}{Image\ \rho}\right]^{F=1}$$

where F is the geometric Frobenius.

The following result shows the relation of proposition 6.1 with the arithmetic of the degeneration h

Theorem 6.2. *Assume the monodromy-weight conjecture on $\mathcal{Z}_{\bar\eta}$ and the semisimplicity of the action of the frobenius F on $H^*(\mathcal{Z}_{\bar\eta}, \Lambda)^I$. Then, for $i > 0$ and $d = \dim\ \mathcal{X}_{\bar\eta}$*

$$\underset{s=d-i}{ord} \det(Id - FN(v)^{-s} | H^{2d}(\mathcal{Z}_{\bar\eta}, \Lambda)^I) =$$

$$rk\left[\frac{Ker(\rho^{2(i+1)} : H^{2(d-i)}(\tilde{T}^{(2i+1)}, \Lambda)(d-i) \to H^{2(d-i)}(\tilde{T}^{(2(i+1))}, \Lambda)(d-i))}{Image\ \rho}\right]^{F=1}.$$

$N(v)$ is the number of elements of the finite residue field $k(v)$.

Proof. cf. [2], theorem 3.5. □

This result explains geometrically the pole of the local factor at v of the L-function $L(H^{2d}(\mathcal{Z}_{\bar\eta}, \mathbf{Q}_\ell), s)$ at the points $s = d-1$ and $s = d-2$, with the presence of the "diagonal" cycles representing the monodromy powers on the strata of T as we previously described.

7. Appendix (by Spencer Bloch)

Our objective in this appendix is to define a multiplication between the total complex of E_1-terms of the Steenbrink spectral sequence and the graded complex

$$H^*(Y^{(\bullet)}),\ \rho = \text{restriction} \tag{7.1}$$

which is the E_1 complex converging to the cohomology of the special fiber Y. We order the components $Y = Y_1 \cup \ldots \cup Y_N$ and write $a_{i_0,\ldots,i_m} \in H^*(Y_{i_0,\ldots,i_m}, \mathbf{Q})$. The E_1-terms of the Steenbrink spectral sequence can be arrayed in a triangular diagram (compare [7], (2.3.8.1)) where each $\bullet$ denotes some $H^*(Y^{(m)}, \mathbf{Q}(n))$.

$$
\begin{array}{ccccccc}
\bullet & & & & & & \\
\uparrow & \nwarrow & & & & & \\
\bullet & \leftarrow & \bullet & & & & \\
\uparrow & \nwarrow & \uparrow & \nwarrow & & & \\
\bullet & \leftarrow & \bullet & \leftarrow & \bullet & & \\
\vdots & \diagup_{\mathrm{wt}0} & \vdots & \diagup_{\mathrm{wt}1} & \vdots & \ddots & \\
\bullet & \leftarrow & \bullet & \leftarrow & \bullet & \leftarrow & \bullet
\end{array}
\tag{7.2}
$$

Here the horizontal arrows are Gysin maps and the vertical arrows are restriction maps. The diagonal arrows are (upto twist) the maps N which, on the level of E_1 are either the identity or 0. The Steenbrink E_1-terms, i.e. the $H^*(Y, \mathrm{gr}^L_r \mathbf{R}\Psi(\mathbf{Q}))$, are direct sums of terms on a NE-SW diagonal, with weight r meeting the "x-axis" at $x = r$. The complex $H^*(Y^{(\bullet)})$ is embedded as the left hand column, and the resulting multiplication on it is the usual (associative) product

$$
a_{i_0,\ldots,i_m} \otimes b_{j_0,\ldots,j_n} \mapsto \begin{cases} 0 & i_m \neq j_0 \\ (a \cdot b)_{i_0,\ldots,i_m,j_1,\ldots,n_n} & i_m = j_0 \end{cases}
\tag{7.3}
$$

The bottom row is a quotient complex calculating the homology of the closed fiber $H_*(Y)$ (with appropriate twist). Our multiplication induces an action of the left hand column on the bottom row, which we will show induces the cap product ([14], p. 254)

$$
H^q(Y) \otimes H_n(Y) \to H_{n-q}(Y).
\tag{7.4}
$$

This module structure, unifying and extending the classical cocycle calculations for cup and cap product, is of independent interest. Quite possibly it can be extended to a product on the whole E_1-complex, but the daunting sign calculations involved have prevented us from working it out.
We will apply this construction to calculate the product

$$
[N^i] \cdot : \mathbf{H}^*(T, gr^L_r A^\bullet_{Z,\mathbf{Q}}) \to \mathbf{H}^{*+2d}(T, gr^L_{r-2i} A^\bullet_{Z,\mathbf{Q}}(d-i))
\tag{7.5}
$$

from (2.1).
We return to the situation in section 2. In particular, $Z \to X \times_S X$ is a resolution, and $T \subset Z$ is the special fiber, which we assume is a normal crossings divisor. We write $E_1(Z)$ for the Steenbrink spectral sequence associated to the degeneration Z/S.

LEMMA 7.1. *There exists a class $[N^i]$ in $E_1(Z)$ satisfying*

1. $d_1[N^i] = 0$, *and the induced class in* E_2 *is the* i*-th power of the monodromy operator*

$$N^i \in gr^L_{2(d-i)} H^{2d}(\tilde{X}^* \times \tilde{X}^*, \mathbf{Q}(d-i))$$

2. $N([N^i]) = 0$, *i.e. in the diagram* (7.2), $[N^i]$ *lies in the left hand vertical column.*

Proof. We see from proposition (2.1) that the class of N^i is killed by N in $E_2(Z)$. Let M denote the map on E_1 which is inverse to N insofar as possible, i.e. M maps down and to the right in diagram (7.2). M is zero on the bottom line. Let $x \in E_1$ represent N^i in E_2. Then $Nx = d_1 y$. (Here $d_1 = d' + d''$ is the total differential.) Since N commutes with d' and d'', and Nx has no term on the bottom row, it follows that $[N^i] := x - d_1 My$ is supported on the left hand column, i.e. killed by N. □

Here is some notation. The special fiber will be $Y = \bigcup Y_i$, with $0 \le i \le N$. Write $H^*(Y)$ for cohomology in some fixed constant ring like $\mathbf{Z}$ or $\mathbf{C}$.

$$I = \{i_0, \dots, i_m\}; \quad J = \{j_0, \dots, j_n\} \quad \text{(strictly ordered);} \quad Y_I = \bigcap_{i_k \in I} Y_{i_k}$$

We will say the pair I, J is *admissible* if

$$\exists p \text{ such that } i_m = \max(I) = j_p \text{ and } \{j_0, \dots, j_p\} \subset I.$$

In this case, write $j_0 = i_{b_0}, \dots, j_{p-1} = i_{b_{p-1}}$. Define

$$a(I, J) := b_0 + \ldots + b_{p-1} + mp.$$

With I, J admissible as above, write

$$J' = \{j_0, \dots, j_p\};\ J'' = \{j_p, \dots, j_n\};\ J = J' \cup J'';\ J' \cap J'' = \{j_p\} = \{i_m\}$$

Write

$$I' = J';\ I'' = (I - J') \cup \{i_m\};\ I = I' \cup I'';\ \{i_m\} = I' \cap I''$$

Let $K = I'' \cup J''$, and define

$$\theta(I, J) : H^\alpha(Y_I) \otimes H^\beta(Y_J) \to H^{\alpha+\beta+2p}(Y_K) \tag{7.6}$$

$$\theta(I, J)(x \otimes y) := (-1)^{a(I,J)} g_{j_0} \circ \cdots g_{j_{p-1}}(x \cdot y). \tag{7.7}$$

Here $x \cdot y \in H^{\alpha+\beta}(Y_{I \cup J})$, the g_j are Gysin maps, and

$$g_{j_0} \circ \cdots g_{j_{p-1}} : H^*(Y_{I \cup J}) \to H^{*+2p}(Y_{I'' \cup J''}).$$

If the pair I, J is not admissible, define $\theta(I, J) = 0$. Define for I as above and $0 \le k \le N$

$$\sigma(I, k) := \#\{i \in I \mid i < k\}$$

For $k \notin I$ we have the restriction $\mathrm{rest}_k : H^*(Y_I) \to H^*(Y_{I \cup \{k\}})$. Define

$$d' := \sum_{k \notin I} (-1)^{\sigma(I,k)} \mathrm{rest}_k : H^*(Y_I) \to \bigoplus_{k \notin I} H^*(Y_{I \cup \{k\}})$$

Similarly, for $k \in I$ we have the Gysin $g_k : H^*(Y_I) \to H^{*+2}(Y_{I-\{k\}})$. We define

$$d'' = \sum_{k\in I}(-1)^{\sigma(I,k)}g_k : H^*(Y_I) \to \bigoplus_{k\in I} H^{*+2}(Y_{I-\{k\}}).$$

Theorem 7.2. *With notation as above (I, J not necessarily admissible) the following diagram is commutative:*

$$\begin{array}{ccc} H^*(Y_I)\otimes H^*(Y_J) & \xrightarrow{\theta(I,J)} & H^*(Y_K) \\ \Big\downarrow{\scriptstyle d'\otimes 1+(-1)^m 1\otimes(d'+d'')} & & \Big\downarrow{\scriptstyle d'+d''} \\ \bigoplus_{\tilde I,\tilde J} H^*(Y_{\tilde I})\otimes H^*(Y_{\tilde J}) & \xrightarrow{\theta(\tilde I,\tilde J)} & \bigoplus_{\tilde K=\tilde I''\cup\tilde J''} H^*(Y_{\tilde K}) \end{array}$$

Remark 7.3.

A priori the theorem does not suffice to determine the desired mapping

$$H^*(Y^\bullet)\otimes E_1 \to E_1 \quad a\otimes b \mapsto a*b$$

because a given $H^*(Y_K)$ occurs many times in the diagram (2.1) (at every point along a NW pointing diagonal). However, if we add the condition that the weights (SW-NE diagonals in (7.2)) should be added, the mapping is defined. It has the property that

$$a * Nb = N(a*b)$$

In particular, there is an induced action on E_1/NE_1 which we identify with the bottom row in (7.2). This simple complex calculates $H_*(Y)$, and the product coincides with the cap product. To see this, one notes that the product is correct for two elements in weight 0, and that if each $H^*(Y_I)$ is replaced by $\mathbf{Z}$, the acyclic model theorem ([14], p. 165) can be applied.

proof of theorem. The proof consists of many separate cases. In each case we will check the sign carefully (this is the delicate part) and omit checking that the maps coincide set-theoretically (which is straightforward).
case: $i_m \notin J$.
In this case, the pair I, J is not admissible, so $\theta(I,J) = 0$. We must show

$$\bigoplus_{\tilde I,\tilde J}\theta(\tilde I,\tilde J)\circ(d'\otimes 1 + (-1)^m 1\otimes(d'+d'')) = 0. \tag{7.8}$$

We may ignore non-admissible $\tilde I, \tilde J$. The only way admissible $\tilde I, \tilde J$ can occur in this situation is if for some $p \geq 0$ we have $j_{p-1} < i_m < j_p$ and $\{j_0, \ldots, j_{p-1}\} \subset I$. (If a subscript for j doesn't fall in $\{0, \ldots, n\}$, ignore it, i.e. take $j_{-1} = -\infty$, $j_{n+1} = +\infty$.) Assume these conditions hold. Then the pair $I\cup\{j_p\}$, J is admissible and occurs in the image of $d'\otimes 1$. Also the pair I, $J\cup\{i_m\}$ is admissible and occurs in the image of $(-1)^m(1\otimes d')$. We must show these two contributions cancel. Suppose $j_0 = i_{b_0}, \ldots, j_{p-1} = i_{b_{p-1}}$. Then the sign condition we need to verify is

$$\sigma(I,j_p)+b_0+\cdots+b_{p-1}+p(m+1)\equiv \\ 1+m+\sigma(J,i_m)+b_0+\cdots+b_{p-1}+pm \mod (2)$$

This is correct because $\sigma(I,j_p)=m+1$ and $\sigma(J,i_m)=p$.
case: $i_m=j_p\in J,\ \{j_0,\dots,j_{p-1}\}\not\subset J$.
This is the other case where I,J is not admissible, so $\theta(I,J)=0$. To get admissible $\tilde I,\tilde J$ we must have

$$\exists k,\ 0\le k\le p-1 \text{ such that } j_k\notin I,\ \{j_0,\dots,\hat{j}_k,\dots,j_{p-1}\}\subset I.$$

Assume this. Then the pairs $(I\cup\{j_k\},J)$ and $(I,J-\{k_k\})$ are admissible. The first occurs in $\theta(I\cup\{j_k\},J)\circ(d'\otimes 1)$ and the second in $(-1)^m\theta(I,J-\{j_k\})\circ 1\otimes d''$. The necessary sign condition for cancellation is

$$\sigma(I,j_k)+a(I\cup\{j_k\},J)\overset{?}{\equiv} m+1+k+a(I,J-\{j_k\}) \mod (2).$$

To check this sign condition write $j_r=i_{b_r}$ for $0\le r\le p-1,\ r\neq k$. Then

$$a(I,J-\{j_k\})=b_0+\cdots+b_{k-1}+b_{k+1}+\cdots+b_{p-1}+(p-1)m$$
$$a(I\cup\{j_k\},J)=b_0+\cdots+b_{k-1}+\sigma(I,j_k)+(b_{k+1}+1)+ \\ +\cdots+(b_{p-1}+1)+p(m+1).$$

This yields the necessary congruence.
For the rest of the proof we assume I,J is admissible. We examine the various terms in (7.8) and show they occur with the same signs in $(d'+d'')\circ\theta(I,J)$. We first consider terms coming from $d'\otimes 1$, so the target is labelled by $\tilde I = I\cup\{k\},\ \tilde J=J$.
case: $k<i_m=j_p$. In this case, since $j_p=\min J''$ and $k\notin I\supset J'$, we have $k\notin J$. The pair $\tilde I=I\cup\{k\},\ J$ is admissible with $\tilde I''=I''\cup\{k\}$ and the same decomposition $J=J'\cup J''$. Let $\tilde K=\tilde I''\cup J''=K\cup\{k\}$. Since $k<j_p=\min J''$, we have

$$\sigma(K,k)=\sigma(I'',k)=\sigma(I,k)-\sigma(J',k)$$

What we must show, therefore, is that

$$a(I,J)-a(\tilde I,J)\equiv\sigma(J',k) \mod (2)$$

Write

$$\tilde I=\{\tilde i_0,\dots,\tilde i_{m+1}\};\ j_0=\tilde i_{\tilde b_0},\dots,j_{p-1}=\tilde i_{\tilde b_{p-1}};$$
$$a(\tilde I,J)=\tilde b_0+\cdots+\tilde b_{p-1}+(m+1)p$$
$$I=\{i_0,\dots,i_m\};\ j_r=i_{b_r}, 0\le r\le p-1$$
$$a(I,J)=b_0+\cdots+b_{p-1}+mp$$

where

$$\tilde b_\ell=\begin{cases} b_\ell & i_{b_\ell}<k\\ b_\ell+1 & i_{b_\ell}>k\end{cases}$$

Thus

$$a(\tilde{I},J)-a(I,J)=p-\#\{j\in J'-\{j_p\}\mid j>k\}=\\ \#\{j\in J'\mid j<k\}=\sigma(J',k).$$

This is the desired congruence.
We continue to consider the contribution of $d'\otimes 1$ with I,J admissible.
case: $k>i_m,\ k\neq j_{p+1}$.
In this case $I\cup\{k\},J$ is not admissible so $\theta(\tilde{I},J)=0$.
case: $k=j_{p+1}$.
Here $\tilde{I}:=I\cup\{k\},\tilde{J}:=J$ is admissible with

$$\tilde{J}'=\{j_0,\dots,j_{p+1}\}=J'\cup\{k\}=J'\cup\{j_{p+1}\}$$
$$\tilde{J}''=\{j_{p+1},\dots,j_n\}=J''-\{j_p\};\ \tilde{K}=\tilde{I}''\cup\tilde{J}''=K-\{j_p\}$$

Note in this case $k>i_m$ so $\sigma(I,k)=m+1$. The claim is here that the diagram

$$\begin{array}{ccc} H^*(Y_I)\otimes H^*(Y_J) & \xrightarrow{\theta(I,J)} & H^*(Y_K)\\ \Big\downarrow{\scriptstyle (-1)^{m+1}\mathrm{rest.}\otimes 1} & & \Big\downarrow{\scriptstyle (-1)^{\sigma(K,j_p)}\mathrm{Gysin}_{j_p}}\\ H^*(Y_{\tilde{I}})\otimes H^*(Y_J) & \xrightarrow{\theta(\tilde{I},J)} & H^*(Y_{\tilde{K}}) \end{array}$$

commutes. Note that the right hand vertical arrow (with the sign) is part of $1\otimes d''$. To verify the signs we need

$$a(I,J)+\sigma(K,j_p)\equiv m+1+a(\tilde{I},J).$$

Since $K=I''\cup J''$ and $k=\max(I'')=\min(J'')$ it is clear that

$$\sigma(K,j_p)=\#I''-1=m-p.$$

Also $j_p=i_m$ so with the usual notation $j_r=i_{b_r}$ we get

$$a(\tilde{I},J)=b_0+\cdots+b_{p-1}+m+(m+1)(p+1).$$

Now the desired congruence becomes

$$b_0+\cdots+b_{p-1}+pm+m-p\equiv b_0+\cdots+b_{p-1}+m+(m+1)(p+1)+m+1$$

This is correct.
We now consider terms occurring in $(-1)^m(1\otimes d')$ on the left of the diagram in the statement of the theorem. We assume given $k\notin J$.
case: $k>j_p$.
Note in this case $k\notin I$. Taking $\tilde{J}=J\cup\{k\},\ \tilde{K}=K\cup\{k\}$, I claim the diagram below is commutative:

$$\begin{array}{ccc} H^*(Y_I)\otimes H^*(Y_J) & \xrightarrow{\theta(I,J)} & H^*(Y_K)\\ \Big\downarrow{\scriptstyle (-1)^{m+\sigma(J,k)}1\otimes\mathrm{rest}} & & \Big\downarrow{\scriptstyle (-1)^{\sigma(K,k)}\mathrm{rest}}\\ H^*(Y_I)\otimes H^*(Y_{\tilde{J}}) & \xrightarrow{\theta(I,\tilde{J})} & H^*(K_{\tilde{K}}) \end{array}$$

(In other words, the contribution in this case is to d' on the right.) Set

$$\tilde{J} = J' \cup \tilde{J}''; \ \tilde{J}'' = J'' \cup \{k\}; \ K = I'' \cup J''.$$

We have

$$\begin{gathered} a(I,J) = a(I,\tilde{J}) \\ \sigma(J,k) = \sigma(J'',k) + p + 1 \\ \sigma(K,k) = \sigma(J'',k) + \#I'' = \sigma(J'',k) + m + 1 - p \end{gathered}$$

It follows that

$$m + \sigma(J,k) + a(I,\tilde{J}) \equiv \sigma(K,k) + a(I,J) \mod (2)$$

which is the desired sign relation in this case.
case: $k < j_p, \ k \notin I$.
In this case, the pair $I, J \cup \{k\}$ is not admissible, so the contribution is zero.
case: $k < j_p, \ k \in I$.

In this case the pair $I, \tilde{J}$ is admissible with

$$\begin{gathered} \tilde{J} := J \cup \{k\} = \tilde{J}' \cup J''; \ \tilde{J}' = J' \cup \{k\} \\ I = \tilde{I} = \tilde{J}' \cup \tilde{I}''; \ \tilde{I}'' = I'' - \{k\}; \ \tilde{K} = K - \{k\} = \tilde{I}'' \cup J'' \end{gathered}$$

The term in question contributes to d'' on the right, and the diagram which commutes is:

$$\begin{array}{ccc} H^*(Y_I) \otimes H^*(Y_J) & \xrightarrow{\theta(I,J)} & H^*(Y_K) \\ \Big\downarrow {\scriptstyle (-1)^{m+\sigma(J,k)}\mathrm{rest}} & & \Big\downarrow {\scriptstyle (-1)^{\sigma(K,k)}\mathrm{Gysin}_k} \\ H^*(Y_I) \otimes H^*(Y_{\tilde{J}}) & \xrightarrow{\theta(I,\tilde{J})} & H^*(Y_{\tilde{K}}) \end{array}$$

The signs will be correct if

$$a(I,J) + \sigma(K,k) \equiv m + \sigma(J,k) + \theta(I,\tilde{J}) \mod (2)$$

Write $\tilde{J} = \{\tilde{j}_0, \dots, \tilde{j}_{m+1}\}$ and $\tilde{j}_r = i_{\tilde{b}_r}, \ r \le p$. The desired congruence reads

$$b_0 + \cdots + b_{p-1} + mp + \sigma(K,k) \overset{?}{\equiv} m + \sigma(J,k) + \tilde{b}_0 + \cdots + \tilde{b}_p + (p+1)m$$

We have

$$\tilde{b}_\ell = \begin{cases} b_\ell & \ell < \sigma(J',k) \\ \sigma(I,k) & \ell = \sigma(J',k) \\ b_{\ell-1} & \ell > \sigma(J',k) \end{cases}$$

The condition becomes

$$\sigma(K,k) \overset{?}{\equiv} \sigma(J,k) + \sigma(I,k) = \sigma(J',k) + \sigma(J',k) + \sigma(I'',k),$$

which is true.
Finally we consider terms coming from $(-1)^m(1 \otimes d'')$ in the lefthand vertical

arrow in the diagram of the theorem. In what follows $j \in J$.
case: $j \in J'',\ j \neq j_p$. Define

$$\tilde{J} = J - \{j\};\ K = I'' \cup J'';\ \tilde{K} = K - \{j\} = I'' \cup \tilde{J}''.$$

The diagram which commutes is:

$$\begin{array}{ccc} H^*(Y_I) \otimes H^*(Y_J) & \xrightarrow{\theta(I,J)} & H^*(Y_K) \\ \Big\downarrow{\scriptstyle 1\otimes(-1)^{m+\sigma(J,j)}\mathrm{Gysin}_j} & & \Big\downarrow{\scriptstyle (-1)^{\sigma(K,j)}\mathrm{Gysin}_j} \\ H^*(Y_I) \otimes H^*(Y_{\tilde{J}}) & \xrightarrow{\theta(I,\tilde{J})} & H^*(Y_{\tilde{K}}) \end{array}$$

The sign condition to be checked is

$$m + \sigma(J,j) + a(I,\tilde{J}) \overset{?}{\equiv} a(I,J) + \sigma(K,j) \mod (2).$$

Our conditions imply $j > j_p$ so $a(I,J) = a(I,\tilde{J})$. Also,

$$\#I'' + \#J' = m + 2 \equiv m \mod (2),$$

so

$$\begin{aligned} \sigma(K,j) &= \#I'' + \sigma(J'',j) - 1 \equiv m + \#J' + \sigma(J'',j) - 1 \\ \sigma(J,j) &= \sigma(J',j) + \sigma(J'',j) - 1 = \#J' - 1 + \sigma(J'',j). \end{aligned}$$

This is the desired condition.
case: $j = j_p$.
In this case, $I, J - \{j\}$ is not admissible, so we get no contribution.
case: $j \in J,\ j < j_p$.
In this case, $j \in J',\ j \neq j_p$. Set

$$\begin{gathered} \tilde{J} = J - \{j\};\ \tilde{J}' = J' - \{j\}; \tilde{J}'' = J'' \\ \tilde{I} = I;\ \tilde{I}'' = I'' \cup \{j\};\ I = \tilde{I} = \tilde{J}' \cup \tilde{I}'' \\ K = I'' \cup J'';\ \tilde{K} = \tilde{I}'' \cup \tilde{J}'' = K - \{j\}. \end{gathered}$$

The sign condition to show we gat a contribution to d'' on the right is

$$a(I,J) + \sigma(K,j) \overset{?}{\equiv} m + \sigma(J,j) + a(\tilde{I},\tilde{J}) \mod (2).$$

Writing $j = j_\ell = i_{b_\ell}$ the condition becomes

$$\begin{aligned} b_0 + \cdots + b_{p-1} + mp + \sigma(K,j) &\overset{?}{\equiv} \\ & b_0 + \cdots + \hat{b}_\ell + b_{\ell+1} + \cdots + b_{p-1} + m(p-1) + m + \sigma(J,j) \end{aligned}$$

This is true because

$$\begin{gathered} b_\ell = \sigma(I,j) = \sigma(I'',j) + \sigma(J',j) \\ \sigma(K,j) = \sigma(I'',j); \quad \sigma(J,j) = \sigma(J',j). \end{gathered}$$

The proof is completed by checking that all the terms on the right in the theorem (i.e. in $d' + d''$) are accounted for precisely once in the above enumeration of cases. □

References

[1] S. Bloch, H. Gillet, C. Soulé *Non–Archimedean Arakelov Theory*, Journal of Algebraic Geometry 4 no. 3, (1995), 427–485.

[2] C. Consani, *Double Complexes and Euler L–Factors*, Compositio Math. 111 no. 3, (1998), 323–358.

[3] P. Deligne, *Théorie de Hodge II*, Publ. Math. IHES Vol.40, (1972) 5–57.

[4] ———, *La Conjecture de Weil II*, Publ. Math. IHES Vol.52, (1980) 137–252.

[5] ———, *Equations Différentielles à Points Singuliers Réguliers*, LNM Vol.163 (1970).

[6] F. Guillén, V. Navarro Aznar *Sur le théorème local des cycles invariants*, Duke Mathematical Journal 61 no. 1, (1990), 133–155.

[7] L. Illusie, *Autour du Théorème de Monodromie Locale*, in Périodes p–Adiques, Asterisque 223 (1994) 9–57.

[8] S. Kleiman, *Algebraic Cycles and the Weil Conjectures*, in Dix Exposés sur la Cohomologie des Schémas, North–Holland (1968) 359–386.

[9] D. Morrison, *The Clemens–Schmid Exact Sequence and Applications*, in Topics in Trasc. alg. Geometry, Princeton Univ. Press (1984) 101–119.

[10] U. Persson, H. Pinkham *Degenerations of Surfaces with Trivial Canonical Bundle*, Ann. of Math. Vol.113, (1981) 45–66.

[11] J. H. Steenbrink, *Limits of Hodge Structures*, Inventiones Math. 31 (1976) 229–257.

[12] M. Saito, *Modules de Hodge polarisables*, Publ. RIMS 24 (1988), 849–995.

[13] J. P. Serre, *Corps Locaux*, Hermann, Paris, (1968).

[14] E. Spanier, *Algebraic Topology*, McGraw-Hill, New-York, (1966).

[15] M. Saito, S. Zucker, *The Kernel Spectral Sequence of Vanishing Cycles*, Duke Math. J. 61 (2) (1990) 329–339.

[16] M. Rapoport, Th. Zink, *On the Local Zeta Function of Shimura Varieties. Monodromy Filtration and Vanishing Cycles in Unequal Characteristic*, Inventiones Math. 68 (1982) 21–101.

C. Consani
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Ma. 02139
USA
katia@math.mit.edu

S. Bloch
Department of Mathematics
The University of Chicago
Chicago, Il. 606139
USA
bloch@math.uchicago.edu

On Rational and Periodic Solutions of Stationary KdV Equations

R. Weikard[1]

Received: February 11, 1999

Communicated by Thomas Peternell

Abstract. Stationary solutions of higher order KdV equations play an important role for the study of the KdV equation itself. They give rise to the coefficients of the associated Lax pair (P, L) for which P and L have an algebraic relationship (and are therefore called algebro-geometric). This paper gives a sufficient condition for rational and simply periodic functions which are bounded at infinity to be algebro-geometric as those potentials of L for which $Ly = zy$ has only meromorphic solutions. It also gives a new elementary proof that this is a necessary condition for any meromorphic function to be algebro-geometric.

1991 Mathematics Subject Classification: 35Q53, 34A20, 58F07
Keywords and Phrases: KdV equation, algebro-geometric solutions of integrable systems, meromorphic solutions of linear differential equations

1 Introduction

The collection of equations of the form

$$q_t = [P, L]$$

where $L = \partial^2/\partial x^2 + q$ and (P, L) is a Lax pair[2] is called the KdV hierarchy. Stationary solutions of equations in the KdV hierarchy are given as $[P, L] = 0$

[1]Based upon work supported by the US National Science Foundation under Grant No. DMS-9401816

[2]That is, P is a monic odd-order differential expression whose coefficients are polynomials in q and its x-derivatives in such a way that the commutator $[P, L]$ is a multiplication operator, see Lax [13].

and are, according to a theorem of Burchnall and Chaundy [2], [3], related to a hyperelliptic curve. For this reason they are often called algebro-geometric potentials of L. In the case of continuous, real-valued, periodic potentials q Novikov [15] and Dubrovin [4] established the fact that q is algebro-geometric if and only if the spectrum of the associated $L^2(\mathbb{R})$-operator has a finite-band structure. Recently F. Gesztesy and myself [7] discovered that an elliptic potential is algebro-geometric if and only if, for every $z \in \mathbb{C}$, every solution of the equation $Ly = y'' + qy = zy$ is a meromorphic function of the independent variable. Our proof relied on a classical theorem of Picard [16], [17], [18] which states that a linear ordinary homogeneous differential equation with elliptic coefficients has always a solution which is elliptic of the second kind provided every solution of the equation is meromorphic. Note that in Picard's theorem the independent variable is considered to be a complex variable.
By extending this result to the AKNS hierarchy (cf. [8]) we proved that the connection between the algebro-geometric property and the existence of only meromorphic solutions is not restricted to the KdV case. For a review of these and related matters see [9].
The goal of this paper is to show with the aid of theorems of Halphen [10] and Floquet [6] that this characterization of elliptic algebro-geometric potentials may be carried over to the case of rational and simply periodic potentials. This covers the case of the famous N-soliton solutions of the KdV equation, which, when viewed as depending on a complex variable, are exponentially decaying along the real axis but are periodic with a purely imaginary period. Specifically, after giving a formal definition for the term "algebro-geometric" in Definition 1, necessary and sufficient conditions for a potential to be algebro-geometric will be provided in Theorems 1 and 2, respectively.

Definition 1. Let L be the differential expression $L = d^2/dx^2 + q$. A meromorphic function $q : \mathbb{C} \to \mathbb{C}_\infty$ will be called algebro-geometric (or an algebro-geometric potential of L) if there exists an ordinary differential expression P of odd order which commutes with L.

Note that by Theorem 6.10 of Segal and Wilson [19] any algebro-geometric potential which is smooth in some real interval may be extended to a meromorphic function on $\mathbb{C}$. The restriction to meromorphic functions in Definition 1 is made to provide a concise statement.

Theorem 1. *If q is an algebro-geometric potential then the following two statements hold:*

1. *Any pole of q is a regular singular point of the differential equation $y'' + qy = zy$. The principal part of the Laurent expansion of q near x_0 is given by $-k(k+1)/(x-x_0)^2$ for a suitable positive integer k. In particular, the residue of q at x_0 is equal to zero.*

2. *For all $z \in \mathbb{C}$ all solutions of $y'' + qy = zy$ are meromorphic functions of the independent variable.*

We prove this theorem in Section 4.
At this point it should be noted that, in the case when the curve associated with q is nondegenerate, the above result follows also from a theorem of Its and Matveev [12] published in 1975. In fact, Its and Matveev showed that, under the given circumstances, the potential q and a fundamental system of solutions of $y'' + qy = zy$ may be expressed in terms of Riemann's theta-function. From these expressions one can read off immediately the conclusions of Theorem 1.
In 1985 Segal and Wilson [19] looked at this type of questions from a very different perspective. They study the Gelfand-Dickey hierarchy (which contains the KdV hierarchy as a special case) employing loop group techniques. Instead of Riemann's theta-function they use an object called τ-function which is also an entire function of its arguments and this implies the validity of Theorem 1.
In justification of offering yet another proof of Theorem 1 let me remark that it will be completely elementary using only the well-known recursion formalism of the KdV-hierarchy.
We turn now our attention from necessary conditions for the algebro-geometric property to sufficient conditions. In Section 5 the following theorem will be proven.

Theorem 2. *Suppose that the function q satisfies one of the following three conditions:*

- *q is rational and bounded near infinity,*
- *q is simply periodic with period p and there exists a positive number R such that q is bounded in $\{x : |\operatorname{Im}(x/p)| \geq R\}$, or*
- *q is elliptic.*

Furthermore assume that, for infinitely many values of $z \in \mathbb{C}$, every solution of the differential equation $Ly = y'' + qy = zy$ is meromorphic. Then q is an algebro-geometric potential of L.

Note that, when q is elliptic, this result was proven in [7]. However, the proof given below will be new and much shorter than the one in [7].
In Section 2 the KdV hierarchy is formally introduced and some of its most important properties are collected. In Section 3 Frobenius' method of establishing series solutions of linear differential equations is used to prove two crucial lemmas. Section 4 is devoted to the proof of Theorem 1 while Section 5 furnishes the proof of Theorem 2.

2 The KdV hierarchy

Suppose q is a solution of some equation in the KdV hierarchy, i.e., there exists a positive integer g and a monic differential expression $\tilde{P}$ of order $2g+1$ such that $q_t = [\tilde{P}, L]$ where $L = \partial^2/\partial x^2 + q$. Since L commutes with its own powers we may add a polynomial $K(L)$ whose degree is at most g to $\tilde{P}$ and still have a monic differential expression of order $2g+1$ whose commutator with L equals q_t. It is well known that among all these expressions one can be written as

$$P = \sum_{j=0}^{g} \left[-\frac{1}{2} f'_{g-j}(x) + f_{g-j}(x) \frac{d}{dx} \right] L^j$$

where $f_0 = 1$ and, for $n \geq 1$, the f_n can be expressed as polynomials in q and its x-derivatives which obey the recursion relation

$$f'_{n+1}(x) = \frac{1}{4} f'''_n(x) + q(x) f'_n(x) + \frac{1}{2} q'(x) f_n(x). \tag{1}$$

In fact, since $[P, L] = f'_{g+1}$, the equation satisfied by q is $q_t = f'_{g+1}$. The condition that q be a stationary solution of some equation in the KdV hierarchy is therefore equivalent to the existence of an integer g such that

$$f'_{g+1}(x) = \frac{1}{4} f'''_g(x) + q(x) f'_g(x) + \frac{1}{2} q'(x) f_g(x) = 0. \tag{2}$$

Defining

$$F_g(z, x) = \sum_{j=0}^{g} f_{g-j}(x) z^j$$

and

$$R_{2g+1}(z) = (z - q(x)) F_g(z, x)^2 - \frac{1}{2} F''_g(z, x) F_g(z, x) + \frac{1}{4} F'_g(z, x)^2$$

one can show that in this case R_{2g+1} does not depend on x and that

$$P^2 = R_{2g+1}(L)$$

which defines the hyperelliptic curve mentioned in the introduction. Since it is also true that $[P, L] = 0$ if P and L satisfy the relationship $P^2 = R_{2g+1}(L)$ one has the following result which is a special case of a theorem of Burchnall and Chaundy [2], [3].

Theorem 3. *Let $L = d^2/dx^2 + q$ and suppose P is a monic differential expression of order $2g+1$. Then L and P are commutative if and only if there exist polynomials R and K of degree $2g+1$ and $k \leq g$, respectively, such that $(P + K(L))^2 = R(L)$.*

We need subsequently the following theorem which establishes a sufficient condition for the potential q to be algebro-geometric.

Theorem 4. *Let $y_1(z, \cdot)$ and $y_2(z, \cdot)$ be two solutions of $Ly = y'' + qy = zy$ which are linearly independent for all but at most countable many values of z. Define*

$$g(z, x) = y_1(z, x) y_2(z, x).$$

If

$$g(z,x)=\frac{F(z,x)}{\gamma(z)},$$

where γ is independent of x and $F(z,x)$ is a polynomial as a function of z and meromorphic as a function of x, then q is algebro-geometric.

Proof. A straightforward calculation[3] shows that the function $g(z,\cdot)$ satisfies the differential equation

$$4(z-q(x))g^2-2gg''+g'^2=W(y_1,y_2)(z)^2 \tag{3}$$

where $W(y_1,y_2)$ is the Wronskian determinant of y_1 and y_2 and where primes denote derivatives with respect to x. Hence

$$(z-q(x))F(z,x)^2-\frac{1}{2}F(z,x)F''(z,x)+\frac{1}{4}F'(z,x)^2=\gamma(z)W(y_1,y_2)(z)^2. \tag{4}$$

As a function of z the left hand side is a polynomial of degree $2g+1$ with leading coefficient $4f_0(x)^2$ when $F(\cdot,x)$ is of degree g and has leading coefficient $f_0(x)$. Since the right hand side does not depend on x we conclude that $f_0(x)$ is constant and we may assume without loss of generality that $f_0(x)=1$. Equation (4) implies also that q is meromorphic. Therefore we may differentiate (4) with respect to x. Assuming that

$$F(z,x)=\sum_{n=0}^{g}f_n(x)z^{g-n}$$

and dropping a common factor $-2F(z,x)$ we obtain

$$\sum_{n=0}^{g-1}f'_{n+1}(x)z^{g-n}=\sum_{n=0}^{g}\left(\frac{1}{4}f'''_n(x)+\frac{1}{2}q'(x)f_n(x)+qf'_n(x)\right)z^{g-n}$$

since $f'_0=0$. This shows that the coefficients f_n satisfy the recursion relation (1) and that f_g satisfies (2). Hence, by the preceding considerations, q is algebro-geometric. $\square$

3 Frobenius' Method

In this section we prove two results concerning the structure of solutions of the differential equation $y''+qy=zy$. The first of these results is obtained from applying Frobenius' method of solving an ordinary linear differential equation by a power series to our particular case. A more general account can be found, for instance, in Ince [11], Chapter XVI. The proof of this standard result is only provided to facilitate references to it. The second result draws some further

[3] Apparently this observation was first made by Appell [1] in 1880.

conclusions in the presence of a spectral parameter and for the case when all solutions are meromorphic for infinitely many values of this spectral parameter. Suppose x_0 is a regular singular point of the equation $y''+qy=0$. Then q is meromorphic in a vicinity of x_0 and has, at worst, a second order pole there. Suppose

$$q(x)=\sum_{j=0}^{\infty} q_j(x-x_0)^{j-2}.$$

Then the indicial equation of the singularity is $r(r-1)+q_0=0$. The roots of this equation are called indices and since their sum must be equal to one we may denote them by $-k$ and $k+1$ where without loss of generality $\operatorname{Re}(-k)\le \operatorname{Re}(k+1)$. Note that $q_0=-k(k+1)$. Now introduce the series

$$w(\sigma,x)=\sum_{j=0}^{\infty} c_j(\sigma)(x-x_0)^{\sigma+j}.$$

Then

$$w''+qw=\sum_{j=0}^{\infty}\left\{(j+\sigma)(j+\sigma-1)c_j+q_0c_j+\sum_{m=0}^{j-1} q_{j-m}c_m\right\}(x-x_0)^{j+\sigma-2}.$$

Define

$$f_0(\ell)=\ell(\ell-1)+q_0=(\ell+k)(\ell-k-1)$$

and, recursively for $j\ge 1$,

$$c_j(\sigma)=\frac{-\sum_{m=0}^{j-1} q_{j-m}c_m(\sigma)}{f_0(\sigma+j)} \tag{5}$$

assuming that $f_0(\sigma+j)\neq 0$ for $j\ge 1$. Then

$$w''+qw=c_0(\sigma)f_0(\sigma)(x-x_0)^{\sigma-2}.$$

Suppose first that $-k$ and $k+1$ do not differ by an integer. Then $f_0(\sigma)=0$ but $f_0(\sigma+j)\neq 0$ for $j\ge 1$ and either choice of σ among the values $-k$ and $k+1$. Hence, choosing $c_0=1$, we find that $w(-k,\cdot)$ and $w(k+1,\cdot)$ are linearly independent solutions of $y''+qy=0$.

Next suppose that $2k+1$ is a nonnegative integer. Then $w(k+1,\cdot)$ is again a solution of $y''+qy=0$ but $w(-k,\cdot)$ becomes undefined since the requirement that $f_0(\sigma+j)\neq 0$ is not satisfied for $\sigma=-k$ and $j=2k+1$. To obtain a second solution we choose $c_0=\prod_{j=1}^{2k+1} f_0(\sigma+j)$. One shows then by induction that c_0, ..., c_{2k} are polynomials with simple zeros at $\sigma=-k$ while c_{2k+1} is a polynomial which may or may not have a zero at $-k$. Finally, c_{2k+2}, c_{2k+3}, ... are rational functions in σ which are analytic at $-k$. Now consider

$$v(\sigma,x)=\frac{\partial w}{\partial\sigma}(\sigma,x)=\sum_{j=0}^{\infty}\left(\frac{\partial c_j}{\partial\sigma}+c_j\log(x-x_0)\right)(x-x_0)^{\sigma+j}.$$

Since differentiation with respect to σ commutes with $d^2/dx^2 + q(x)$ we obtain that

$$v'' + qv = \left(\frac{\partial (c_0 f_0)}{\partial \sigma} + c_0 f_0 \log(x - x_0) \right) (x - x_0)^{\sigma - 2}.$$

Since $c_0 f_0 = \prod_{j=0}^{2k+1} f_0(\sigma + j)$ has a double zero at $\sigma = -k$ we obtain that $v(-k, \cdot)$ is a solution of $y'' + qy = 0$ which is easily seen to be independent from $w(k+1, \cdot)$. We may write

$$v(-k, x) = h_1(x) \log(x - x_0) + h_2(x)$$

where

$$h_1(x) = \sum_{j=2k+1}^{\infty} c_j(-k)(x - x_0)^{j-k} \quad \text{and} \quad h_2(z, x) = \sum_{j=0}^{\infty} \frac{\partial c_j}{\partial \sigma}(-k)(x - x_0)^{j-k}. \tag{6}$$

We collect these results for the particular case, when k is a positive integer in the following

Lemma 5. *Suppose q is meromorphic near x_0 with principal part*

$$-k(k+1)/(x - x_0)^2 + q_1/(x - x_0)$$

where k is a positive integer. Then the differential equation $y'' + qy = zy$ has a solution w which is analytic at x_0 and a solution v defined by $v(x) = h_1(x) \log(x - x_0) + h_2(x)$ where h_1 is analytic at x_0 and h_2 is meromorphic at x_0.

This lemma and its proof are the main ingredients of the following one.

Lemma 6. *Let Z be the set of all values of $z \in \mathbb{C}$ such that $y'' + qy = zy$ has only meromorphic solutions. The following statements hold:*

1. *If Z is not empty then q is meromorphic and any pole of q is of the second order at most.*
2. *Z is either a finite set or equal to $\mathbb{C}$.*
3. *If $Z = \mathbb{C}$ and if x_0 is a pole of q then the principal part of the Laurent expansion of q about x_0 is given by $-k(k+1)(x - x_0)^{-2}$ for some $k \in \mathbb{N}$, in particular, $\operatorname{res}_{x_0} q = 0$.*

Proof. The fact that $q = (y'' - zy)/y$ shows that q is meromorphic and has at most a double pole at any of its singular points even if $y'' + qy = zy$ has only one meromorphic solution for one value of z. This proves the first claim.
Hence, if $Z \neq \emptyset$, a pole x_0 of q is a regular singular point of $y'' + qy = zy$ and

$$q(x) = \sum_{j=0}^{\infty} q_j (x - x_0)^{j-2}$$

in a vicinity of x_0. The indices associated with x_0, which are given as the roots of $r(r-1)+q_0=0$ and hence are independent of z, must be distinct integers whose sum equals one. We denote them by $-k$ and $k+1$ where $k>0$ and note that $q_0=-k(k+1)$.

Note that replacing q by $q-z$ amounts to replacing q_2 by q_2-z in the Laurent expansion of q turning the recursion relation (5) into

$$c_j(\sigma,z)=\frac{-\sum_{m=0}^{j-1}(q_{j-m}-z\delta_{j-m,2})c_m(\sigma,z)}{f_0(\sigma+j)} \tag{7}$$

where $c_0=\prod_{j=1}^{2k+1} f_0(j+\sigma)$. The equation $y''+(q-z)y=0$ has a solution $v(z,x)=h_1(z,x)\log(x-x_0)+h_2(z,x)$ which is meromorphic at x_0 if and only if $h_1(z,\cdot)=0$. Recall that $c_1(-k,z)=...=c_{2k}(-k,z)=0$. Using this fact in the recursion relation (7) shows that the coefficients $c_j(-k,z)$ are zero for all j if and only if $c_{2k+1}(-k,z)=0$. Hence, because of (6), we have $h_1(z,\cdot)=0$ if and only if $c_{2k+1}(-k,z)=0$. The recursion relation (7) also implies immediately that the coefficients c_j are polynomials in their second variable. Hence $c_{2k+1}(-k,\cdot)$ has either finitely many zeros or else it is identically equal to zero. Therefore, if $c_{2k+1}(-k,\cdot)\neq 0$ for any singular point of the equation then Z is finite. However, if $c_{2k+1}(-k,\cdot)=0$ for all singular points of the equation then $Z=\mathbb{C}$. This proves the second claim.

To prove the third claim we need more detailed information about the leading coefficient of the polynomial $c_{2k+1}(-k,\cdot)$. We will show below that, if $q_1=\operatorname{res}_{x_0} q\neq 0$, then $c_{2k+1}(-k,\cdot)$ is a polynomial of degree k thus forcing Z to be a finite set and proving the last claim.

Suppose now that $q_1\neq 0$. Since $c_{2k+1}(\cdot,z)$ has a removable singularity at $-k$ we may determine $c_{2k+1}(-k,z)$ by computing $\lim_{\sigma\to -k} c_{2k+1}(\sigma,z)$ for $\sigma<-k$. Note that $c_0(\sigma,z)=\gamma_0(\sigma)$ and $c_1(\sigma,z)=-q_1\gamma_1(\sigma)$ where

$$-\gamma_0(\sigma)=-\prod_{j=1}^{2k+1} f_0(\sigma+j) \quad\text{and}\quad \gamma_1(\sigma)=\prod_{j=2}^{2k+1} f_0(\sigma+j).$$

The functions $-\gamma_0$ and γ_1 are positive in $(-k-1,-k)$ and have simple zeros at $-k$. Assume that $j\leq k$ and that $c_{2j-2}(\sigma,z)$ and $c_{2j-1}(\sigma,z)$ are polynomials in z of degree $j-1$ and that

$$c_{2j-2}(\sigma,z)=\gamma_{2j-2}(\sigma)z^{j-1}+O(z^{j-2}),$$
$$c_{2j-1}(\sigma,z)=-q_1\gamma_{2j-1}(\sigma)z^{j-1}+O(z^{j-2})$$

where $(-1)^j\gamma_{2j-2}$ and $(-1)^{j-1}\gamma_{2j-1}$ are positive in $(-k-1,-k)$ and have simple zeros at $-k$. Then, using the recursion relation (7), we obtain that $c_{2j}(\sigma,z)$ and $c_{2j+1}(\sigma,z)$ are polynomials in z of degree j and that, in particular,

$$c_{2j}(\sigma,z)=\frac{zc_{2j-2}}{f_0(\sigma+2j)}+O(z^{j-1})=\frac{\gamma_{2j-2}}{f_0(\sigma+2j)}z^j+O(z^{j-1}),$$
$$c_{2j+1}(\sigma,z)=\frac{zc_{2j-1}-q_1c_{2j}}{f_0(\sigma+2j+1)}+O(z^{j-1})=-q_1\frac{\gamma_{2j-1}+\gamma_{2j}}{f_0(\sigma+2j+1)}z^j+O(z^{j-1}).$$

Letting $\gamma_{2j} = \gamma_{2j-2}/f_0(\sigma + 2j)$ and $\gamma_{2j+1} = (\gamma_{2j-1} + \gamma_{2j})/f_0(\sigma + 2j + 1)$ we find that $(-1)^{j+1}\gamma_{2j}$ and $(-1)^j\gamma_{2j+1}$ are positive in $(-k-1,-k)$ and that γ_{2j} has a simple zero at $-k$. If $j < k$ then γ_{2j+1} has a simple zero at $-k$, too, since γ_{2j-1} and γ_{2j} have the same sign in $(-k-1,-k)$. However, if $j = k$ then both the numerator and the denominator in $(\gamma_{2k-1}+\gamma_{2k})/f_0(\sigma+2k+1)$ have a simple zero at $-k$ proving that $\gamma_{2k+1}(-k)$ is different from zero. This, however, shows that $c_{2k+1}(-k,\cdot)$ is a polynomial of degree k which has at most k distinct zeros. However, $c_{2k+1}(-k,\cdot)$ must be zero for any value of z since $Z = \mathbb{C}$. This contradiction proves our assumption $q_1 \neq 0$ wrong. $\square$

4 Necessary Conditions

In this section we will prove Theorem 1 which gives conditions which must be satisfied for any algebro-geometric potential. We start with

Theorem 7. *If q is algebro-geometric then any of its poles is a regular singular point of the differential equation $Ly = y'' + qy = zy$. Moreover, when x_0 is a pole of q then the coefficient of $(x-x_0)^{-2}$ in the Laurent expansion of q about x_0 is equal to $-k(k+1)$ for some positive integer k.*

Proof. We show first that any pole of q is a regular singular point of $y''+qy = zy$, i.e., that its order is at most equal to two. Hence assume this were not the case. That is, suppose that x_0 which, without loss of generality, may be assumed to be equal to zero is a pole of q of order $k \geq 3$. Then q has a Laurent expansion $q = \alpha x^{-k} + ...$ where $\alpha \neq 0$. Consider the recursion relation (1). One shows by induction that the order of the pole $x_0 = 0$ of f_n''' is smaller than that of $qf_n' + q'f_n/2$ and that therefore

$$f_n'(x) = -nk\alpha^n x^{-nk-1} \prod_{j=1}^{n} \frac{2j-1}{2j} + O(x^{-nk}).$$

If q were algebro-geometric there would have to be an n such that $f_n' = 0$. This contradiction shows that the order of the pole x_0 is at most two and that x_0 is a regular singular point of $y'' + qy = zy$.

Next assume that $x_0 = 0$ is a pole of order one, i.e., $q = \alpha x^{-1} + O(1)$ with $\alpha \neq 0$. We prove, again by induction, that

$$f_n'(x) = -\frac{1}{2}\alpha x^{-2n} \prod_{j=1}^{n-1} j(j+1/2) + O(x^{-2n+1})$$

using that the order of the pole $x_0 = 0$ of f_n''' is larger than that of $qf_n'+q'f_n/2$. Hence, for no n is f_n' ever zero showing that x_0 must not be a first order pole if q is algebro-geometric.

Finally, suppose that $q = \alpha x^{-2} + ...$ for some α different from any number in $\{-k(k+1) : k \in \mathbb{N}\}$. Then another induction shows that

$$f_n'(x) = -2n \left[\prod_{j=1}^{n} \frac{2j-1}{2j} (\alpha + j(j-1)) \right] x^{-2n-1} + O(x^{-2n}).$$

Again $f_n' \neq 0$ for all $n \in \mathbb{N}$ contrary to the hypothesis. □

Theorem 8. *If q is algebro-geometric then every solution of $Ly = y'' + qy = zy$ is meromorphic for every $z \in \mathbb{C}$.*

Proof. Since q is algebro-geometric there exists a differential expression P of the form

$$P = \sum_{j=0}^{g} \left[-\frac{1}{2} f_{g-j}'(x) + f_{g-j}(x) \frac{d}{dx} \right] L^j$$

for which $[P, L] = 0$ and $P^2 = R_{2g+1}(L)$. In contradiction to what we want to prove assume that there exists a point z_0 such that $y'' + qy = z_0 y$ has a solution which is not meromorphic.

Let Z be the set of all values of $z \in \mathbb{C}$ such that $y'' + qy = zy$ has only meromorphic solutions. By Lemma 6 the set Z is closed and hence its complement is open. Therefore and because the zeros of R_{2g+1} are isolated there is no harm in assuming that $R_{2g+1}(z_0) \neq 0$.

Next denote the two-dimensional space of solutions of $Ly = z_0 y$ by $W(z_0)$. The restriction of P to the space $W(z_0)$ maps back into $W(z_0)$ since P and L commute. Note that

$$P|_{W(z_0)} = F_g(z_0, x) \frac{d}{dx} - \frac{1}{2} F_g'(z_0, x).$$

Introduce the basis $\{y_1, y_2\}$ of $W(z_0)$ which is defined by $y_j^{(\ell-1)}(x_0) = \delta_{j,\ell}$. In this basis the restriction of P to $W(z_0)$ is represented by the matrix

$$M = \frac{1}{2} \begin{pmatrix} -F_g'(z_0, x_0) & 2F_g(z_0, x_0) \\ 2(z_0 - q(x_0))F_g(z_0, x_0) - F_g''(z_0, x_0) & F_g'(z_0, x_0) \end{pmatrix}.$$

Note that $\operatorname{tr} M = 0$ and $\det M = -R_{2g+1}(z_0)$ regardless of x_0. Therefore M has distinct eigenvalues $\pm w_0 = \pm\sqrt{R_{2g+1}(z_0)}$. The associated eigenfunctions $\psi_\pm$ satisfy $P\psi_\pm = \pm w_0 \psi$ and $L\psi_\pm = z_0 \psi_\pm$. Define $\varphi_\pm = \psi_\pm'/\psi_\pm$ and note that

$$\pm w_0 = \frac{P\psi_\pm}{\psi_\pm} = F_g \varphi_\pm - \frac{1}{2} F_g'.$$

Hence

$$\varphi_\pm = \frac{\pm 2w_0 + F_g'}{2F_g}$$

are meromorphic functions on $\mathbb{C}$. Not both of the solutions $\psi_\pm$ can be meromorphic since they are linearly independent. Suppose ψ_+ is not meromorphic. Then, by Lemma 5, there is a constant γ such that ψ_+ (or an appropriate multiple) is given as

$$\psi_+(x) = h_1(x)\log(x-x_0) + h_2(x) + \gamma w(x) \tag{8}$$

where h_1, h_2, and w are functions which are meromorphic at x_0. Hence

$$\begin{aligned}&(x-x_0)(\varphi_+ h_1 - h_1')\log(x-x_0)\\ &= h_1 + (x-x_0)(h_2' + \gamma w') - (x-x_0)(h_2+\gamma w)\varphi_+\end{aligned}$$

is meromorphic at x_0 and we conclude that $\varphi_+ h_1 - h_1' = 0$. This implies that $h_1 = c\psi_+$ for some constant c. If $c \neq 0$ we obtain from (8)

$$\log(x-x_0) = \left(\frac{1}{c}h_1(x) - h_2(x) - \gamma w(x)\right) h_1(x)^{-1}$$

which is impossible since the right hand side is meromorphic at x_0. Therefore $c = 0$, i.e., h_1 vanishes identically and ψ_+ is meromorphic at x_0 contrary to our assumption. □

We are now ready for the

Proof of Theorem 1. Theorem 7 proves that a pole of q is a regular singular point with principal part $-k(k+1)/(x-x_0)^2 + q_1/(x-x_0)$ for a suitable positive integer k and complex number q_1. Theorem 8 proves not only that all solutions of $y'' + qy = zy$ are meromorphic for all $z \in \mathbb{C}$ but also that the hypotheses of Lemma 6 are satisfied. This in turn shows then that $q_1 = 0$. □

5 Sufficient Conditions

In this section we will prove Theorem 2. As mentioned in the introduction the proofs rely on classical theorems by Halphen, Floquet, and Picard concerning the linear differential equation

$$q_0 y^{(n)} + q_1 y^{(n-1)} + \dots + q_n y = 0. \tag{9}$$

While Floquet's theorem is well known (see e.g. Eastham [5] or Magnus and Winkler [14]) it is appropriate to repeat the theorems of Halphen and Picard. Halphen's theorem is concerned with the rational case. A proof is given by Ince [11] and this proof can be used to state the following version which is different from Ince's version.

Theorem 9. *Let the coefficients $q_0, \dots, q_n$ in (9) be polynomials such that $\deg q_j \le \deg q_0 = s$ for $j = 1, \dots, n$. For $j = 0, \dots, n$ let A_j be the coefficient of x^s in q_j and let λ be a zero of $A_0\lambda^n + A_1\lambda^{n-1} + \dots + A_n$. If the differential equation (9) has only meromorphic solutions then it has a solution $R(x)\exp(\lambda x)$ where R is a rational function.*

Picard's theorem is concerned with the elliptic case. It may also be found in [11].

Theorem 10. *Assume that the coefficients $q_0, ..., q_n$ in* (9) *are elliptic with common fundamental periods $2\omega_1$ and $2\omega_2$ and let ρ_1 be a Floquet multiplier with respect to the period $2\omega_1$. If the differential equation* (9) *has only meromorphic solutions then it has a solution which is elliptic of the second kind and satisfies $y(x+2\omega_1) = \rho_1 y(x)$.*

5.1 Rational potentials

Suppose that q is rational and bounded at infinity. Let $z_0 = \lim_{x\to\infty} q(x)$. From Lemma 6 we know that $y'' + qy = zy$ has only meromorphic solutions for any value of z and from Halphen's theorem (Theorem 9) we obtain, for $z \neq z_0$, that there are linearly independent solutions

$$y_\pm(z,x) = R_\pm(z,x)\exp(\pm\sqrt{z-z_0}x)$$

where $R_\pm(z,\cdot)$ are rational functions. Also from Lemma 6 we obtain that

$$q = z_0 - \sum_{j=1}^m \frac{s_j(s_j+1)}{(x-b_j)^2}$$

where $b_1, ..., b_m$ are distinct complex numbers and $s_1, ..., s_m$ are positive integers. The singular point b_j of $y'' + qy = zy$ has indices $-s_j$ and s_j+1 and hence any pole of $y_\pm$ is located at one of the points b_j and has order s_j. Now define the function $g(z,x) = y_+(z,x)y_-(z,x)$. Letting $v(x) = \prod_{j=1}^m (x-b_j)^{s_j}$ we see that the functions $y_\pm v$ are entire as functions of x and hence $v^2 g(z,\cdot)$ is an entire rational function, i.e., a polynomial. Letting

$$v(x)^2 g(z,x) = \sum_{j=0}^d c_j x^j$$

the functions $v^2 g(z,x)$, $v^3 g'(z,x)$, $v^4 g''(z,x)$, and $v^5 g'''(z,x)$ are polynomials in x whose coefficients are homogeneous polynomials of degree one in $c_0, ..., c_d$. Since $v^2 q$ and $v^3 q'$ are polynomials we find that $v^5(g''' + 4(q-z)g' + 2q'g)$ is also a polynomial in x whose coefficients are homogeneous polynomials of degree one in $c_0, ..., c_d$. The coefficients of the c_ℓ in this last expression, in turn, are polynomials in z of degree at most one, i.e.,

$$v^5(g''' + 4(q-z)g' + 2q'g) = \sum_{j=0}^N \sum_{\ell=0}^d (\alpha_{j,\ell} + \beta_{j,\ell} z) c_\ell x^j \tag{10}$$

for suitable numbers N, $\alpha_{j,\ell}$, and $\beta_{j,\ell}$, which depend only on q. From Appell's equation (3) it follows upon differentiation that the expression (10) vanishes identically. This gives rise to a homogeneous system of $N+1$ linear equations for the c_ℓ of which we know that it has a nontrivial solution. Solving the system shows now that the coefficients c_ℓ are rational functions of z, i.e.,

$$c_\ell(z) = \frac{\tilde{c}_\ell(z)}{\gamma(z)}$$

where γ and $\tilde{c}_\ell$ are polynomials in z. Therefore

$$g(z,x) = \frac{\sum_{j=0}^d \tilde{c}_j(z)x^j}{\gamma(z)v(x)^2} = \frac{F(z,x)}{\gamma(z)}$$

where

$$F(z,x) = \frac{\sum_{j=0}^d \tilde{c}_j(z)x^j}{v(x)^2}$$

is a polynomial as function of z and a rational function as function of x. We have therefore proven that the hypotheses of Theorem 4 are satisfied and this shows that q is algebro-geometric.

5.2 Simply Periodic Potentials

Suppose q is meromorphic, simply periodic with period $p \in \mathbb{C}$, and bounded in $\{x : |\operatorname{Im}(x/p)| \geq R\}$ for some $R > 0$. Lemma 6 implies firstly that, for all values of z all solutions of $y'' + qy = zy$ are meromorphic. To simplify notation we assume without loss of generality that the fundamental period p of q is equal to 2π. Define $q^* : \mathbb{C} - \{0\} \to \mathbb{C}^\infty$ by $q^*(t) = q(-i \log t)$. Because of the periodicity of q the function q^* is well-defined and meromorphic. Since $q(x)$ remains bounded as $|\operatorname{Im}(x)|$ tends to infinity the points zero and infinity are removable singularities of q^* and hence q^* is a rational function which is bounded at infinity and zero. Denoting its poles by $t_1, ..., t_m$ we may write

$$q^*(t) = z_0 + \sum_{j=1}^m \sum_{k=1}^{N_j} \frac{t_j^k A_{j,k}}{(t-t_j)^k}$$

where $t_1, ..., t_m$ are distinct nonzero complex numbers. Let x_j be any complex number such that $e^{ix_j} = t_j$. Then we obtain that

$$q(x) = q^*(e^{ix}) = z_0 + \sum_{j=1}^m \sum_{k=1}^{N_j} \frac{A_{j,k}}{(e^{i(x-x_j)} - 1)^k}.$$

Since

$$e^{i(x-x_j)} - 1 = i(x-x_j)(1 + \frac{i}{2}(x-x_j) + O((x-x_j)^2))$$

we obtain from Lemma 6 that $N_j = 2$ and $A_{j,1} = A_{j,2} = s_j(s_j+1)$ with $s_j \in \mathbb{N}$. Hence

$$q^*(t) = z_0 + \sum_{j=1}^m s_j(s_j+1)\frac{tt_j}{(t-t_j)^2}.$$

In particular $q^*(0) = q^*(\infty) = z_0$.
From Floquet's theorem we know that there are solutions (called Floquet solutions) of $y'' + qy = zy$ of the form

$$\psi_\pm(z, x) = p_\pm(z, x)\mathrm{e}^{\pm i\lambda x}$$

where $p_\pm$ are periodic functions with period 2π and λ is a suitable complex number depending on z which is determined up to addition of an arbitrary integer. Unless 2λ is an integer which happens only for an isolated set of values of z the solutions $\psi_\pm$ are linearly independent.
The functions $\psi_\pm(z, \cdot)$ are meromorphic by assumption. Their poles are at the singularities of the differential equation, i.e., at the poles of q. Because the indices of the singularities $x_j = -i \log t_j$ are $-s_j$ and $s_j + 1$ the functions given by

$$\psi_\pm(z, x)\mathrm{e}^{\mp i\lambda x} \prod_{j=1}^{m} (\mathrm{e}^{ix} - \mathrm{e}^{ix_j})^{s_j}$$

are entire and periodic functions of period 2π.
Define

$$v(x) = \prod_{j=1}^{m} (\mathrm{e}^{ix} - \mathrm{e}^{ix_j})^{s_j}.$$

The substitution $y = u\mathrm{e}^{i\lambda x}/v$ transforms $y'' + qy = zy$ into

$$v^2u'' + (2i\lambda v^2 - 2vv')u' + ((-\lambda^2 - z + q)v^2 - 2i\lambda vv' + 2v'^2 - vv'')u = 0 \tag{11}$$

which has entire solutions at least one of which is periodic with period 2π.
Next define $v^*(t) = v(-i \log t) = \prod_{j=1}^{m}(t - t_j)^{s_j}$ and substitute $u(x) = u^*(t)$ where $x = -i \log t$ in (11) to obtain $u'(x) = itu^{*\prime}(t)$, $u''(x) = -t^2u^{*\prime\prime}(t) - tu^{*\prime}(t)$ and hence

$$Q_0u^{*\prime\prime} + Q_1u^{*\prime} + Q_2u^* = 0 \tag{12}$$

where

$$\begin{aligned}
Q_0 &= t^2v^{*2},\\
Q_1 &= t((1 + 2\lambda)v^{*2} - 2tv^*v^{*\prime}),\\
Q_2 &= (z - q^* + \lambda^2)v^{*2} - (2\lambda + 1)tv^*v^{*\prime} + 2t^2v^{*\prime 2} - t^2v^*v^{*\prime\prime}.
\end{aligned}$$

Because equation (11) has an entire 2π-periodic solution equation (12) has a solution which is analytic on $\mathbb{C} - \{0\}$, i.e., a solution for which zero and infinity are isolated singularities.
Since $v^*(0) \neq 0$ the point zero is a regular singular point of (12) with indicial equation

$$r^2 + 2\lambda r + z - z_0 + \lambda^2 = 0. \tag{13}$$

This equation must have at least one integer solution since otherwise no solution of (12) would be one-valued, i.e., zero would not be an isolated singularity. Thus suppose the solutions of (13) are m and $-2\lambda - m$ where $m \in \mathbb{Z}$. Then $-2\lambda m - m^2 = z - z_0 + \lambda^2$ which implies $\lambda = -m \pm i\sqrt{z - z_0}$. As we are free to change λ by adding an integer we may assume from now on that $\lambda^2 = z_0 - z$ and that the zeros of the indicial equation (13) are zero and -2λ.

Next turn to the point infinity. After introducing $1/t$ as independent variable it turns out that infinity is also a regular singular point with indicial equation

$$r^2 + (2S - 2\lambda)r + S^2 - 2\lambda S = 0 \tag{14}$$

where $S = \sum_{j=1}^m s_j = \deg v^*$. The solutions of (14) are $-S$ and $2\lambda - S$.

Now, if 2λ is not an integer then (11) has precisely one linearly independent 2π-periodic solution. Hence (12) has precisely one single-valued analytic solution in $\mathbb{C} - \{0\}$. This must therefore be the solution associated with the indices 0 and $-S$ at zero and infinity, respectively, i.e., this solution is a polynomial of degree S.

Repeating the above procedure after replacing λ by $-\lambda$ we now obtain that the differential equation $y'' + qy = zy$ has the solutions

$$y_\pm(z,x) = \frac{u_\pm^*(z, \mathrm{e}^{ix})}{v^*(\mathrm{e}^{ix})} \exp(\pm i\lambda x)$$

where $u_+^*(z,\cdot)$ and v^* are polynomials. These solutions are linearly independent except at an at most countable number of isolated points z.

Again define the function $g(z,x) = y_+(z,x)y_-(z,x)$. Then

$$v(x)^2 g(z,x) = u_+^*(z, \mathrm{e}^{ix}) u_-^*(z, \mathrm{e}^{ix}) = \sum_{j=0}^{d} c_j(z) \mathrm{e}^{ijx}.$$

The functions $v^2 g(z,x)$, $v^3 g'(z,x)$, $v^4 g''(z,x)$, and $v^5 g'''(z,x)$ are now polynomials in e^{ix} whose coefficients are homogeneous polynomials of degree one in $c_0, ..., c_d$ and so is the function $v^5(g''' + 4(q-z)g' + 2q'g)$. Specifically,

$$v^5(g''' + 4(q-z)g' + 2q'g) = \sum_{j=0}^{N} \sum_{\ell=0}^{d} (\alpha_{j,\ell} + \beta_{j,\ell} z) c_\ell \mathrm{e}^{ijx}. \tag{15}$$

As the expression (15) must vanish identically we obtain again a system of linear equations which we use to show that the coefficients c_ℓ are rational functions of z. Therefore $g(z,x) = F(z,x)/\gamma(z)$ where $F(z,x)$ is a polynomial as function of z and a rational function as function of e^{ix}. Theorem 4 gives that q is algebro-geometric.

5.3 Elliptic potentials

Finally let q be elliptic with fundamental periods $2\omega_1$ and $2\omega_3$. Assume that none of the poles of q equals zero or a half-period (modulo the fundamental period parallelogram) which may always be achieved by a slight shift of the independent variable. Then, by Lemma 6 and general properties of elliptic functions,

$$q(x) = \frac{q_1(\wp(x)) + q_2(\wp(x))\wp'(x)}{\prod_{j=1}^m (\wp(x) - p_j)^2}$$

for suitable polynomials q_1 and q_2 and suitable numbers m and $p_1, ..., p_m$. Let

$$v(x) = \prod_{j=1}^m (\wp(x) - p_j)^{s_j}$$

where $-s_j < 0$ and $s_j + 1 > 0$ are the indices of the singularity x_j for which $\wp(x_j) = p_j$. Then $v^2 q$ and $v^3 q'$ are polynomials in $\wp(x)$ and $\wp'(x)$.
Picard's theorem guarantees the existence of two linearly independent solutions $y_\pm(z, \cdot)$ of $y'' + qy = zy$ which are elliptic of the second kind for all but an at most countable number of isolated points z since we then have different Floquet multipliers with respect to $2\omega_1$. ¿From Floquet theory we know that the product of these solutions must be doubly periodic since the product of Floquet multipliers with respect to any period is equal to one in our case. As all solutions are meromorphic (by Lemma 6) we have that $g(z, \cdot)$ the product of $y_+(z, \cdot)$ and $y_-(z, \cdot)$ is elliptic. Therefore and since the only poles of $g(z, \cdot)$ are at the points where $\wp(x) = p_j$ and have order at most $2s_j$ we get

$$g(z, x) = \frac{g_1(z, \wp(x)) + g_2(z, \wp(x))\wp'(x)}{v(x)^2}$$

where $g_1(z, \cdot)$ and $g_2(z, \cdot)$ are polynomials. Introduce the coefficients $c_0, ..., c_d$ by

$$g_1(z, t) = \sum_{j=0}^{\delta} c_j(z) t^j, \quad g_2(z, t) = \sum_{j=\delta+1}^{d} c_j(z) t^{j-\delta-1}.$$

Each of the functions $v^2 g$, $v^3 g'$, $v^4 g''$, and $v^5 g'''$ are now of the form $\phi_1(\wp) + \phi_2(\wp)\wp'$ where ϕ_1 and ϕ_2 represent various polynomials. The coefficients of these are homogeneous polynomials of degree one in $c_0, ..., c_d$. Therefore $v^5(g''' + 4(q - z)g' + 2q'g) = h_1(\wp(x)) + h_2(\wp(x))\wp'(x)$ where h_1 and h_2 are polynomials whose coefficients are polynomials in the variables $z, c_0, ..., c_d$ homogeneous of degree one with respect to $c_0, ..., c_d$ and of at most first order with respect to z. This implies just as before that the c_j are rational functions of z and proves that $g(z, x) = F(z, x)/\gamma(z)$ where $F(z, x)$ is a polynomial as function of z and a rational function as function of $\wp(x)$ and $\wp'(x)$. Theorem 4 gives that q is algebro-geometric.

References

[1] P. É. Appell, Sur la transformation des équations différentielles linéaires, Comptes Rendus 91, 211–214 (1880).

[2] J. L. Burchnall and T. W. Chaundy, Commutative ordinary differential operators, Proc. London Math. Soc. Ser. 2 21, 420–440 (1923).

[3] J. L. Burchnall and T. W. Chaundy, Commutative ordinary differential operators, Proc. Roy. Soc. London A 118, 557–583 (1928).

[4] B. A. Dubrovin, Periodic problems for the Korteweg-de Vries equation in the class of finite band potentials, Funct. Anal. Appl. 9, 215–223 (1975).

[5] M. S. P. Eastham, *The Spectral Theory of Periodic Differential Equations*, Scottish Academic Press, Edinburgh and London, 1973.

[6] G. Floquet, Sur les équations différentielles linéaires à coefficients périodiques, Ann. Sci. École Norm. Sup. 12, 47–88 (1883).

[7] F. Gesztesy and R. Weikard, Picard potentials and Hill's equation on a torus, Acta Math. 176, 73-107 (1996).

[8] F. Gesztesy and R. Weikard, A characterization of all elliptic solutions of the AKNS hierarchy, Acta Math. 181, 63-108 (1998).

[9] F. Gesztesy and R. Weikard, Elliptic algebro-geometric solutions of the KdV and AKNS hierarchies - an analytic approach. Bull. AMS 35, 271-317 (1998).

[10] G. H. Halphen, Sur une nouvelle classe d'équations différentielles linéaires intégrables, C. R. Acad. Sci. Paris 101, 1238–1240 (1885).

[11] E. L. Ince, *Ordinary Differential Equations*, Dover, New York, 1956.

[12] A. R. Its and V. B. Matveev, Schrödinger operators with finite-gap spectrum and N-soliton solutions of the Korteweg-de Vries equation, Theoret. Math. Phys. 23, 343–355 (1975).

[13] P. Lax, Integrals of Nonlinear Equations of Evolution and Solitary Waves, Comm. Pure Appl. Math. 21, 467–490 (1968).

[14] W. Magnus and S. Winkler, *Hill's Equation*, Interscience Publishers, New York, London, Sydney, 1966.

[15] S. P. Novikov, The periodic problem for the Korteweg-de Vries equation, Funct. Anal. Appl. 8, 236–246 (1974).

[16] E. Picard, Sur une généralisation des fonctions périodiques et sur certaines équations différentielles linéaires, C. R. Acad. Sci. Paris 89, 140–144 (1879).

[17] E. Picard, Sur une classe d'équations différentielles linéaires, C. R. Acad. Sci. Paris 90, 128–131 (1880).

[18] E. Picard, Sur les équations différentielles linéaires à coefficients doublement périodiques, J. reine angew. Math. 90, 281–302 (1881).

[19] G. Segal and G. Wilson, Loop groups and equations of KdV type, Publ. Math. IHES 61, 5–65 (1985).

R. Weikard
Department of Mathematics
University of Alabama at Birmingham
Birmingham, AL 35294–1170
USA
rudi@math.uab.edu

Twistor Spaces With a Pencil of Fundamental Divisors

B. Kreussler

Received: March 24, 1999

Communicated by Thomas Peternell

Abstract. In this paper simply connected twistor spaces Z containing a pencil of fundamental divisors are studied. The Riemannian base for such spaces is diffeomorphic to the connected sum $n\mathbb{CP}^2$. We obtain for $n \geq 5$ a complete description of the set of curves intersecting the fundamental line bundle $K^{-\frac{1}{2}}$ negatively. For this purpose we introduce a combinatorial structure, called *blow-up graph*. We show that for generic $S \in \mid -\frac{1}{2}K \mid$ the algebraic dimension can be computed by the formula $a(Z) = 1 + \kappa^{-1}(S)$. A detailed study of the anti Kodaira dimension $\kappa^{-1}(S)$ of rational surfaces permits to read off the algebraic dimension from the blow-up graphs. This gives a characterisation of Moishezon twistor spaces by the structure of the corresponding blow-up graphs. We study the behaviour of these graphs under small deformations. The results are applied to prove the main existence result, which states that every blow-up graph belongs to a fundamental divisor of a twistor space. We show, furthermore, that a twistor space with $\dim \mid -\frac{1}{2}K \mid = 3$ is a LeBrun space [LeB2]. We characterise such spaces also by the property to contain a smooth rational non-real curve C with $C.(-\frac{1}{2}K) = 2 - n$.

1991 Mathematics Subject Classification: 32L25, 32J17, 32J20, 14M20
Keywords and Phrases: Moishezon manifold, algebraic dimension, self–dual, twistor space

1 Introduction

For a complex manifold with non-positive Kodaira dimension and zero dimensional Albanese torus, the algebraic dimension is the most basic birational invariant. By definition it is the transcendence degree over $\mathbb{C}$ of the field of meromorphic functions on the manifold. Because it is often a difficult task to compute this invariant in explicit examples, it is interesting to study the algebraic dimension in special classes of manifolds. A class where we can find interesting phenomena is the class of twistor spaces. From our point of view, a *twistor space* is a compact complex three-manifold Z equipped with

- a proper differentiable submersion $\pi : Z \longrightarrow M$ onto a real differentiable four-manifold M (called the base), whose fibres are holomorphic curves in Z which are isomorphic to the complex projective line and have normal bundle in Z isomorphic to $\mathcal{O}(1) \oplus \mathcal{O}(1)$ and

- an anti-holomorphic fixed point free involution $\sigma : Z \longrightarrow Z$ with $\pi\sigma = \pi$.

Usually, such spaces arise in 4-dimensional conformal geometry. The points of Z correspond to complex structures on the tangent spaces at M, compatible with the conformal structure. The idea for such a construction traces back to F. Hirzebruch, H. Hopf [HH] and R. Penrose [Pe]. The twistor construction in the context of Riemannian geometry was first developed by M. Atiyah, N. Hitchin, I. Singer [AHS]. It plays an important role as a bridge between conformal Riemannian geometry and complex geometry. Twistor spaces have always negative Kodaira dimension and trivial Albanese torus [H2]. If a twistor space has the maximal possible algebraic dimension $a(Z) = 3$, then it must be simply connected with base homeomorphic to either S^4 or a connected sum of $\mathbb{CP}^2$'s [C2]. Compare with Proposition 2.4 below.
The involution σ is called a real structure and we designate any σ-invariant geometric object as being "real". For example, the fibres of π are called "real twistor fibres", a line bundle $\mathcal{L} \in \operatorname{Pic} Z$ is called real if $\sigma^*\overline{\mathcal{L}} \cong \mathcal{L}$ and a subvariety $D \subset Z$ is called real if $\overline{D} := \sigma(D) = D$. The degree $\deg(\mathcal{L})$ of a line bundle $\mathcal{L} \in \operatorname{Pic} Z$ is by definition the degree of the restriction $\mathcal{L} \otimes \mathcal{O}_F$ to a real twistor fibre $F \subset Z$. The "type" of a twistor space is by definition the sign of the scalar curvature of a metric with constant scalar curvature in the conformal class of M. On every twistor space there exists a distinguished square root $K^{-\frac{1}{2}}$ of the anti-canonical line bundle of Z. This bundle is called the *fundamental* line bundle. The divisors in $| -\frac{1}{2}K |$ are called fundamental divisors. The study of the structure of these divisors and of their linear system played a fundamental role in the study of twistor spaces.
In this paper we study simply connected twistor spaces containing irreducible fundamental divisors. Some authors start with the assumption that the twistor space is of positive type, but we don't here. We show in Section 2 that a simply connected twistor space containing an irreducible fundamental divisor must necessarily have positive type. For a collection of the basic properties of such twistor spaces and appropriate references, the reader is referred to [K1, Sections 2 and 3]. In the final three sections of the paper [K1], the case $c_1^3 = 0$ is studied, whereas the case $c_1^3 > 0$ is fairly well understood (see [H2], [FK], [Po1], [KK], [Po4]). Here we focus on the general case: $c_1^3 < 0$.
The goal of this paper is an understanding of the relationship between the algebraic dimension $a(Z)$, the structure of fundamental divisors and the base locus

and dimension of the fundamental linear system on Z. The results show that a finite set of curves with certain numerical properties contains very important information on the structure of the twistor space. We study the interplay between curves and surfaces, not merely divisors inside our three-manifolds. The basic assumption for our study will be $\dim \mid -\frac{1}{2}K \mid \geq 1$. Under this assumption, we develop in Section 3 a clear picture of the possibilities for the base locus and dimension of the fundamental linear system. We also give a new characterisation of LeBrun twistor spaces (Theorem 3.6). For LeBrun twistor spaces and the twistor spaces studied in [CK2] the place among all twistor spaces becomes quite clear by Theorems 3.6 and 3.7. Curves with certain numerical properties play an important role for these results.
To compute the algebraic dimension of a simply connected twistor space one relies on the observation of Y.S. Poon [Po2] that one can compute $a(Z)$ by the Iitaka dimension of the anti-canonical line bundle $\kappa(Z, K^{-1})$. This can be deduced from the fact that K^{-1} generates the unique one-dimensional subspace in $\mathrm{Pic}\, Z \otimes \mathbb{R}$ which is invariant under the involution, induced by the real structure on Z. To compute $a(Z)$ one can use the inequality $a(Z) \leq 1 + \kappa(S, K_S^{-1})$. But in many cases this is not enough for computing the algebraic dimension. In Section 4 we improve it (under our assumption $\dim \mid -\frac{1}{2}K \mid \geq 1$) to the equality

$$a(Z) = 1 + \kappa(S, K_S^{-1})$$

for generic fundamental divisors S.
This motivates the study of the anti Kodaira dimension $\kappa^{-1}(S) := \kappa(S, K_S^{-1})$ of rational surfaces in Section 5.
To handle the structure of the base locus of the fundamental system (which is also related to the number of divisors of degree one, see [K1, Proposition 3.7]) we define the notion of a *blow-up graph* (Section 6). This is a combinatorial structure which reflects numerical properties of the components of anti-canonical divisors on rational surfaces. These graphs contain also information on the anti Kodaira dimension.
The existence of new twistor spaces can be shown with the aid of deformation theory [DonF], [C1], [LeBP], [PP2], [C3]. To be able to state interesting results on the structure of twistor spaces constructed in such an indirect manner, we study the behaviour of the blow-up graphs under small deformations in Section 7. These results will then be used in Section 8, where the relationship between $a(Z)$, $\dim \mid -\frac{1}{2}K \mid$ and the structure of anti-canonical divisors on fundamental divisors is studied. As a result we see that basic information on the structure of twistor spaces is already contained in a finite set of curves in such a space. We prove in this section a vanishing theorem for the second cohomology of the tangent sheaf:

$$H^2(Z, \Theta_Z) = 0,$$

which is necessary to show the existence of twistor spaces related to arbitrarily prescribed blow-up graphs. Our main existence result (Theorem 8.8) states that every blow-up graph appears as a blow-up graph of a fundamental divisor contained in a twistor space. To prove this we rely on recent results of N.Honda [Ho], who studies the twistor spaces constructed in [PP3].

2 Consequences of the existence of fundamental divisors

In this section we show that the existence of an irreducible fundamental divisor in a simply connected twistor space has strong consequences. We see, for example, that there is no need to assume the twistor space to be of positive type, because we obtain this from our assumption. As a consequence, we have Hitchin's vanishing theorem at our disposal. This states for simply connected twistor spaces of positive type the vanishing of $H^1(Z,\mathcal{L})$ for any line bundle $\mathcal{L}$ with $\deg(\mathcal{L}) \leq -2$ [H1].

In fact, the topology of simply connected twistor spaces containing an effective divisor can be restricted to a few cases by results of P. Gauduchon [Gau] and C. LeBrun [LeB1].

First of all, we cite the following lemma from [PP1, Lemma 2.1], which will be useful in the following.

Lemma 2.1. *Let Z be a compact twistor space and $S \subset Z$ an effective divisor of degree 2 which is irreducible and real, then S is smooth.* □

This implies, in particular, that each real irreducible fundamental divisor $S \in |-\frac{1}{2}K|$ is smooth.

From here on, we are only concerned with *simply connected* twistor spaces. Without assuming Hitchin's vanishing theorem or the twistor space to be of positive type, we can study the structure of irreducible fundamental divisors.

Lemma 2.2. *Let Z be a compact simply–connected twistor space and $S \in |-\frac{1}{2}K|$ be real and irreducible. Then there exists a real twistor fibre $F \subset S$ and $\dim|F| = 1$. The surface S is smooth and rational.*

Proof: From Lemma 2.1 we know smoothness of S. If S would not contain a real twistor fibre, the twistor fibration would give an unramified covering $S \to M$ of degree two, since Z does not contain real points. This is in contradiction with $\pi_1(M) = \pi_1(Z) = 0$. Similarly, if $\dim|F| = 0$, we obtain an unramified covering $S \setminus F \longrightarrow M \setminus \{pt\}$ of degree two. Again, we obtain a contradiction to $\pi_1(M \setminus \{pt\}) = \pi_1(M) = 0$ because $S \setminus F$ is irreducible (being open in the irreducible surface S). This implies $h^0(\mathcal{O}_S(F)) \geq 2$. The adjunction formula on S yields $(F^2)_S = (F.(-K_S))_S - 2 = F.(-\frac{1}{2}K) - 2 = 0$. Hence, we have an exact sequence $0 \longrightarrow \mathcal{O}_S \longrightarrow \mathcal{O}_S(F) \longrightarrow \mathcal{O}_F \longrightarrow 0$, implying $h^0(\mathcal{O}_S(F)) \leq h^0(\mathcal{O}_S) + h^0(\mathcal{O}_F) = 2$. Thus $|F|$ is a pencil. On the other hand, $((mF - K_S)^2)_S = 2m(F.(-K_S))_S + ((-K_S)^2)_S = 4m + ((-K_S)^2)_S > 0$ for large positive m. Therefore, S is a projective algebraic surface ([BPV, IV (5.2)]). By Noether's lemma ([GH, IV§3]) the existence of the pencil $|F|$ implies the rationality of S. □

Lemma 2.3. *If Z is a compact, simply connected twistor space containing an irreducible fundamental divisor, then $h^i(K^{\frac{1}{2}}) = 0$ for all i and $h^i(\mathcal{O}_Z) = 0$ for $i > 0$.*

Proof: By assumption $|-\frac{1}{2}K|$ contains an irreducible member, hence, the generic member of this linear system is irreducible. Therefore, we can choose an irreducible real $S \in |-\frac{1}{2}K|$, which is smooth and rational by Lemma 2.2. In particular, we have $h^1(\mathcal{O}_S) = h^2(\mathcal{O}_S) = 0$. Because the restriction defines an isomorphism $H^0(\mathcal{O}_Z) \xrightarrow{\sim} H^0(\mathcal{O}_S)$, the exact sequence $0 \longrightarrow K^{\frac{1}{2}} \longrightarrow \mathcal{O}_Z \longrightarrow \mathcal{O}_S \longrightarrow 0$ implies $h^0(K^{\frac{1}{2}}) = 0$ and $h^i(K^{\frac{1}{2}}) = h^i(\mathcal{O}_Z)$, if $i > 0$. Using the Serre duality, this gives the desired vanishing for $i \in \{0, 3\}$ and $h^1(\mathcal{O}_Z) = h^1(K^{\frac{1}{2}}) = h^2(K^{\frac{1}{2}}) = h^2(\mathcal{O}_Z)$. The simply connectedness of Z implies $0 = b_1(M) = h^1(\mathcal{O}_Z)$ (see [ES]) finishing the proof. □

Proposition 2.4. *If $Z \longrightarrow M$ is a compact, simply connected twistor space containing an irreducible fundamental divisor, then M is diffeomorphic to the connected sum $n\mathbb{CP}^2$ and M is of positive type.*

Proof: Because Z is compact and M self–dual, we obtain (see e.g. [ES, Cor. 3.2]) $b_-(M) = h^2(\mathcal{O}_Z)$ which vanishes under our assumptions by Lemma 2.3. Therefore, the intersection form on $H^2(M, \mathbb{R})$ is positive definite. To see that the type of M is positive, we recall a theorem of Gauduchon [Gau] stating that a twistor space of negative type does not contain effective divisors. Hence, in our case, the type is non–negative. If the type would be zero, we would obtain (using $\pi_1(M) = 0$) from [Pon, Cor. 4.3], that $\bar{M}$ is a Kähler surface. But, what we have seen above, implies then that the intersection form on $H^2(\bar{M}, \mathbb{R})$ would be negative definite. But for a simply connected complex surface this is impossible by the signature theorem [BPV, IV (2.13.)]. Therefore, M has positive type. In this situation a theorem of Pedersen and Poon [PP1] states that M is diffeomorphic to $n\mathbb{CP}^2$. □

From [Gau] and [LeB1] we obtain that a simply connected twistor space, which contains an effective divisor, can only be built over a self-dual four manifold M having one of the following properties ($\overline{M}$ denotes the anti-self-dual manifold obtained by reversing the orientation of M):

(a) $\overline{M}$ is a blow–up of $\mathbb{P}^2$ at $m > 9$ points or

(b) $\overline{M}$ is a K3–surface with a Ricci–flat metric or

(c) M is homeomorphic to $n\mathbb{CP}^2$ with $n \geq 0$.

From [Po3] we obtain that in case (a) $a(Z) = 0$ and in case (b) $a(Z) = 1$. The goal of the following sections is to gain more knowledge on the algebraic dimension and their relation to the geometry of Z in case (c).

3 The fundamental linear system of a twistor space

We consider a simply connected twistor space Z. In this section we study the fundamental linear system $|-\frac{1}{2}K|$. Under the assumption that it is at least a pencil, we obtain information on its dimension and the base locus. In Section 8 we study the algebraic dimension $a(Z)$ in more detail.

In the case where an irreducible fundamental divisor exists, Proposition 2.4 shows that the Riemannian base of such a twistor space is diffeomorphic to the connected sum $n\mathbb{CP}^2$ and the conformal class contains a metric with positive scalar curvature. If $n \leq 3$ it is well-known (and follows easily from the Riemann-Roch formula and Hitchin's vanishing theorem) that we have $a(Z) = 3$. The case $n = 4$ was studied in [K1]. Since a twistor space of positive type over $n\mathbb{CP}^2$ with $n \leq 4$ contains always a pencil of fundamental divisors, the picture is in this case fairly satisfactory. If, however, $n > 4$ (which is equivalent to $c_1(Z)^3 < 0$) the situation is much more rich and less understood.

In the rest of this section we denote by Z a twistor space fulfilling:

(3.0) It is simply connected, contains an irreducible fundamental divisor and satisfies $h^0(K^{-\frac{1}{2}}) \geq 2$ and $c_1(Z)^3 < 0$.

We have seen in Section 2 that such a twistor space is of positive type and is built over $n\mathbb{CP}^2$ with $n > 4$. Furthermore, $\mathrm{Pic}(Z)$ is a free abelian group of rank $n+1$ and $(-\frac{1}{2}K)^3 = 2(4-n)$ (see [K1, Section 2]).

Lemma 3.1. *Let $D \subset Z$ be an effective divisor of degree one.*
(a) If $D \cap \overline{D} \neq \emptyset$ then $D.\overline{D} = F$ is a real twistor fibre.
(b) If $h^0(D) \geq 2$, then $\dim \mid D \mid = 1$, $\dim \mid -\frac{1}{2}K \mid = 3$, *the base locus of the pencil $\mid D \mid$ is a smooth rational curve B which is disjoint to its conjugate $\overline{B}$. The surface D is rational and intersects the conjugate surface $\overline{D}$. The base locus of the fundamental linear system $\mid -\frac{1}{2}K \mid$ is the curve $B \cup \overline{B}$ and we have $B.(-\frac{1}{2}K) = 2 - n$.*

Proof: This lemma can be deduced from [Ku] and [Po4] but we prefer to give a direct proof here.

(a) Assume $D \cap \overline{D} \neq \emptyset$. Consider a point $z \in D \cap \overline{D}$ and denote by F the real twistor fibre containing z. But F and $D \cap \overline{D}$ are real, hence $\overline{z} \in F \cap D$. Using that D is of degree one we conclude $F \subset D \cap \overline{D}$. As D is smooth and irreducible we have for every real twistor fibre $F \subset D$ an exact normal bundle sequence:

$$0 \longrightarrow N_{F|D} \longrightarrow N_{F|Z} \longrightarrow \mathcal{O}_Z(D) \otimes \mathcal{O}_F \longrightarrow 0.$$

Using $N_{F|Z} \cong \mathcal{O}_F(1)^{\oplus 2}$ and $\mathcal{O}_Z(D) \otimes \mathcal{O}_F \cong \mathcal{O}_F(1)$, we obtain $N_{F|D} \cong \mathcal{O}_F(1)$, which means $(F^2)_D = 1$. As in the proof of Lemma 2.2 this implies that D is algebraic and rational. Since any two real twistor lines are disjoint we infer from the Hodge index theorem that F is the unique real twistor fibre contained in D. This implies $D \cap \overline{D} = F$ because the intersection $D \cap \overline{D}$ would contain a second twistor fibre if it contains a point outside F. We even have $D.\overline{D} = F$, since $D.\overline{D} = rF$ implies $r = ((D.\overline{D}).F)_D = \overline{D}.F = 1$.

(b) If we have $\mathcal{O}_Z(D - \overline{D}) \cong \mathcal{O}_Z$, then $c_1(\mathcal{O}_Z(D))$ would be invariant under the involution on $H^2(Z, \mathbb{Z})$. This would imply $\mathcal{O}_Z(4D) \cong K_Z^{-1}$, which is only possible if $n = 0$. In this case $Z = \mathbb{P}^3$ by [H2] and [FK]. But we assumed $n > 4$.

Hence, $\mathcal{O}_Z(D-\overline{D})$ is not the trivial line bundle. This implies $H^0(\mathcal{O}_Z(D-\overline{D})) = 0$ because there is no effective divisor of degree zero. If D and $\overline{D}$ would be disjoint, we would have $\mathcal{O}_Z(D) \otimes \mathcal{O}_{\overline{D}} \cong \mathcal{O}_{\overline{D}}$. Considering the exact sequence

$$0 \longrightarrow \mathcal{O}_Z(D-\overline{D}) \longrightarrow \mathcal{O}_Z(D) \longrightarrow \mathcal{O}_Z(D) \otimes \mathcal{O}_{\overline{D}} \longrightarrow 0$$

we would obtain a contradiction to the assumption $h^0(\mathcal{O}_Z(D)) \geq 2$. Therefore, we must have $D \cap \overline{D} \neq \emptyset$, hence $D.\overline{D} = F$.

Let now $D' \in \mid D \mid \setminus \{D\}$ and define $B := D.D'$. We obtain $B.\overline{D} = D.D'.\overline{D} = (D.\overline{D}).D' = F.D' = 1$. Since $\mid D \mid$ is at least a pencil, this implies B is smooth. This computation shows furthermore $(B.F)_D = B.\overline{D} = 1$. In particular $B \cap \overline{B} = \emptyset$, because $\overline{B} \subset \overline{D}$ and Z does not contain a real point.

Let $\mathcal{I}_B \subset \mathcal{O}_Z$ be the ideal of $B \subset Z$ and denote by $s, s' \in H^0(\mathcal{O}_Z(D))$ sections defining the divisors D, D'. By $V \subset H^0(\mathcal{O}_Z(D))$ we denote the vector space generated by s, s'. This pair of sections defines the exact Koszul complex

$$0 \longrightarrow \mathcal{O}_Z(-2D) \longrightarrow V \otimes \mathcal{O}_Z(-D) \longrightarrow \mathcal{I}_B \longrightarrow 0. \tag{1}$$

Since $\deg(-2D) = -2$ we obtain from Hitchin's vanishing $h^1(\mathcal{O}_Z(-2D)) = 0$ and $h^2(\mathcal{O}_Z(-2D)) = 0$. Hence, $H^1(\mathcal{O}_Z(-D)) \otimes V \cong H^1(\mathcal{I}_B)$. But the exact sequence $0 \longrightarrow \mathcal{O}_Z(-D) \longrightarrow \mathcal{O}_Z \longrightarrow \mathcal{O}_D \longrightarrow 0$ and $H^1(\mathcal{O}_Z) = 0$ imply $H^1(\mathcal{O}_Z(-D)) = 0$. Hence, $H^1(\mathcal{I}_B) = 0$, showing that the restriction $H^0(\mathcal{O}_Z) \longrightarrow H^0(\mathcal{O}_B)$ is surjective. Thus, B is connected, hence irreducible.

The linear system $\mid F \mid$ on D is of dimension two and does not have base points which can be seen from the exact sequence $0 \longrightarrow \mathcal{O}_D \longrightarrow \mathcal{O}_D(F) \longrightarrow \mathcal{O}_F(F) \longrightarrow 0$. Since $(B.F)_D = 1$ we see that B is the strict transform of a line in $\mathbb{CP}^2$ under the blow up $D \longrightarrow \mathbb{CP}^2$, defined by $\mid F \mid$. In particular, B is smooth and rational.

Now we can compute $B.(-\frac{1}{2}K) = B.(D+\overline{D}) = D.D'.D+B.\overline{D} = D^3+1 = 2-n$, and this is negative since we assumed $n > 4$. In particular B and $\overline{B}$ are contained in the base locus of $\mid -\frac{1}{2}K \mid$. By a lemma of Poon [Po4, Lemma 1.4], we can conclude $h^0(K^{-\frac{1}{2}}) \leq 4$.

To determine the base locus of $\mid -\frac{1}{2}K \mid$ consider a base point z of this linear system. This point is contained in every divisor of the form $D+\overline{D}$ of which there exist an infinite number. Thus there exists a pair of effective linearly equivalent divisors of degree one D, D' such that $z \in D \cap D' = B$ or $z \in \overline{D} \cap \overline{D}' = \overline{B}$. This shows that the base locus of $\mid -\frac{1}{2}K \mid$ is contained in $B \cup \overline{B}$, hence $B \cup \overline{B}$ is the base locus.

Finally, we have to compute the dimension of the fundamental linear system. For this purpose we tensor the exact sequence (1) with $\mathcal{O}_Z(D+\overline{D}) \cong K^{-\frac{1}{2}}$ and obtain an exact sequence

$$0 \longrightarrow \mathcal{O}_Z(\overline{D}-D) \longrightarrow V \otimes \mathcal{O}_Z(\overline{D}) \longrightarrow \mathcal{I}_B \otimes \mathcal{O}_Z(D+\overline{D}) \longrightarrow 0.$$

If we use $H^0(\mathcal{O}_Z(\overline{D}-D)) = 0$ we obtain an injection $H^0(V \otimes \mathcal{O}_Z(\overline{D})) = V \otimes H^0(\mathcal{O}_Z(\overline{D})) \subset H^0(\mathcal{I}_B \otimes K^{-\frac{1}{2}}) \subset H^0(K^{-\frac{1}{2}})$. Since $V \otimes \overline{V} \subset V \otimes H^0(\mathcal{O}_Z(\overline{D}))$

we obtain $4 \leq h^0(K^{-\frac{1}{2}})$ which implies by the above inequality $\dim \mid K^{-\frac{1}{2}} \mid = 3$ and hence $\dim \mid D \mid = 1$. □

Lemma 3.2. *Assume Z contains only finitely many divisors of degree one. If $A \subset Z$ is an irreducible and reduced curve with $A.(-\frac{1}{2}K) < 0$, then there exists a smooth real fundamental divisor $S \in \mid -\frac{1}{2}K \mid$ containing a real twistor fibre F with $2 \geq F.A \geq 1$. We have $A.F = 2$ if and only if A is real.*

Proof: Let $x \in A$ be a point and $x \in F \subset Z$ the real twistor fibre containing this point. Since $\mid -\frac{1}{2}K \mid$ is at least a pencil, there exists a divisor $S \in \mid -\frac{1}{2}K \mid$ containing a given point $y \in F \setminus \{x, \bar{x}\}$. Because $F.S = 2$ and $S \cap F \supset \{y, x, \bar{x}\}$ the twistor fibre F is contained in S. So the real subsystem $\mid -\frac{1}{2}K \mid_F \subset \mid -\frac{1}{2}K \mid$ of divisors containing F is not empty. Because S contains at most a real one-parameter family of real twistor fibres, the intersection points of A with real twistor fibres contained in S form at most a real one-dimensional subset of points z on A. Therefore, we obtain at least a one-parameter family of surfaces S containing a real twistor fibre F with $F.A \geq 1$. Since we assumed that there are only finitely many divisors of degree one, we can choose an irreducible real fundamental divisor among them, which is smooth by Lemma 2.1. Since $(A.(-K_S))_S = A.(-\frac{1}{2}K) < 0$ each real anti-canonical divisor $C \in \mid -K_S \mid$ contains A and $\overline{A}$. Since F is nef this implies $(F.A)_S \leq (F.(-K_S))_S = F.(-\frac{1}{2}K) = 2$. If $A \neq \overline{A}$ there even holds $(F.(A + \overline{A}))_S \leq (F.(-K_S))_S = 2$ implying $(F.A)_S = (F.\overline{A})_S = 1$. If $A = \overline{A}$, we must have $(F.A)_S \neq 1$ since S does not contain real points, hence $(F.A)_S = 2$ in this case. □

Lemma 3.3. *If Z is a twistor space satisfying condition (3.0), then:*

(a) There exists a reduced irreducible curve $A \subset Z$ with $A.(-\frac{1}{2}K) < 0$.

(b) If we have $A.(-\frac{1}{2}K) > 2 - n$ for every reduced irreducible curve $A \subset Z$, then $\dim \mid -\frac{1}{2}K \mid = 1$.

Proof: Let S be a smooth real fundamental divisor. We have $K_S^{-1} \cong K^{-\frac{1}{2}} \otimes \mathcal{O}_S$ and $\dim \mid -\frac{1}{2}K \mid = \dim \mid -K_S \mid +1$, hence $\mid -K_S \mid \neq \emptyset$. Since $((-K_S)^2)_S = (-\frac{1}{2}K)^3 = 2(4-n) < 0$ we obtain (a). To show (b) we recall from [K1, Prop. 3.6] that there exists a succession of blow-ups $\sigma : S \to \mathbb{P}^1 \times \mathbb{P}^1$ such that the anticanonical system $\mid -K_S \mid$ contains a real member C mapped onto a curve C' on $\mathbb{P}^1 \times \mathbb{P}^1$ having one of the following four types :
(0) $C' \in |\mathcal{O}(2,2)|$ is a smooth elliptic curve,
(1) C' has four components $C' = F' + \overline{F'} + G' + \overline{G'}$ where $F' \in |\mathcal{O}(0,1)|$ and $G' \in |\mathcal{O}(1,0)|$ are not real,
(2) C' has two components $C' = F' + C'_0$ where $F' \in |\mathcal{O}(0,1)|$ is real and $C'_0 \in |\mathcal{O}(2,1)|$ is real, smooth and rational,
(3) C' has two distinct components $C' = A' + \overline{A'}$ where $A', \overline{A'} \in |\mathcal{O}(1,1)|$.
At each step of blow-up a conjugate pair of points, lying on the image of C, is blown up. The pencil $\mid F \mid$ generated by a real twistor fibre $F \subset S$ is mapped to the pencil $|\mathcal{O}(0,1)|$ on $\mathbb{P}^1 \times \mathbb{P}^1$. It was shown in [K1, Prop. 3.3] that none of

the blown up points lies over a real member of $|\mathcal{O}(0,1)|$. This implies in case (2) that there is a component C_0 of C with $(C_0.(-K_S))_S = 6-2n < 2-n$. In case (0) we have $(C^2)_S = (C.(-K_S))_S = 8-2n < 0$ and C is irreducible, hence $|-K_S| = \{C\}$. If in case (3) all the blown up points lie over smooth points of $C' = A' + \overline{A'}$, then $C = A + \overline{A}$ with $(A.(-K_S))_S = (\overline{A}.(-K_S))_S = 4-n < 0$. Hence, $|-K_S| = \{C\}$. If, however, in case (3) the conjugate pair of singular points $A' \cap \overline{A'}$ is blown up, then we can choose the succession of blow-ups σ such that C is mapped to a curve of type (1) in $\mathbb{P}^1 \times \mathbb{P}^1$. This is done by an elementary transformation (see [K1, Cor. 4.3]).
To deal with case (1) we choose an irreducible reduced curve $G \subset Z$ with $G.(-\frac{1}{2}K) < 0$. By Lemma 3.1 the assumption of (b) implies that there does not exist a pencil of divisors of degree one. Hence we can apply Lemma 3.2 and can find a smooth real fundamental divisor S with $(F.(G+\overline{G}))_S = 2$ for twistor fibres $F \subset S$. Take $C \subset S$ as above in the description of type (1), then G is a component of C, hence smooth rational and not real. This implies $(F.G)_S = (F.\overline{G})_S = 1$. Thus, the curves G and $\overline{G}$ are mapped to the components G' and $\overline{G'}$ of C'. Since $G.(-\frac{1}{2}K) = \overline{G}(-\frac{1}{2}K) < 0$, at least three of the blown up points are lying over G' and their conjugates over $\overline{G'}$. Since we assumed $G.(-\frac{1}{2}K) > 2-n$, at most $n-1$ blown-up points lie over G'. Hence, a nonempty set of blown-up points lies over a conjugate pair F', $\overline{F'}$ of members of $|\mathcal{O}(0,1)|$. This implies that these curves are not movable, hence $|-K_S| = \{C\}$. □

Proposition 3.4. *Let Z be a twistor space satisfying condition (3.0) and let $A \subset Z$ be an irreducible reduced curve.*

(a) If A is not real, then $A.(-\frac{1}{2}K) \geq 2-n$.

(b) If $A.(-\frac{1}{2}K) < 2-n$, then A is real (i.e. $A = \overline{A}$) and it is the unique irreducible reduced curve having negative intersection number with $K^{-\frac{1}{2}}$. Only the following two cases are possible:

(i) $A.(-\frac{1}{2}K) = 8-2n$, $n > 6$ and A is smooth elliptic. In this case $\dim|-\frac{1}{2}K| = 1$.

(ii) $A.(-\frac{1}{2}K) = 6-2n$ and A is smooth rational. In this case $\dim|-\frac{1}{2}K| = 2$.

(c) If $A.(-\frac{1}{2}K) = 2-n$, then $\{A, \overline{A}\}$ is the set of all irreducible reduced curves in Z with negative intersection number with $K^{-\frac{1}{2}}$ and either

(i) $n = 6$ and $A = \overline{A}$ smooth and elliptic and $\dim|-\frac{1}{2}K| = 1$, *or*

(ii) A is smooth rational and not real. In this case $\dim|-\frac{1}{2}K| = 3$.

Proof: If a pencil of divisors of degree one exists, all the statements are clear by Lemma 3.1. Therefore, we assume that there exists only a finite number of divisors of degree one. This allows us to apply Lemma 3.2.

If $A.(-\frac{1}{2}K) \geq 0$ nothing is to prove. Assume $A.(-\frac{1}{2}K) < 0$ and choose a smooth fundamental divisor S as in Lemma 3.2. Let, furthermore, be $\sigma : S \to \mathbb{P}^1 \times \mathbb{P}^1$ a succession of blow-ups as in the proof of Lemma 3.3. Using the notation of that proof, we obtain that A must be a component of C. Since every curve on $\mathbb{P}^1 \times \mathbb{P}^1$ has nonnegative self-intersection number, (a) is clear from the description of the types (0) – (3).
The assumption of (b) implies that we are in types (0) or (2) which correspond to the cases (i) and (ii) respectively. In type (0), the irreducible curve C has negative intersection number with $-K_S$, hence $| -K_S |= \{C\}$, implying $\dim | -\frac{1}{2}K |= 1$ in case (i) of (b). In type (2) we obtain $| -K_S |= C_0 + |F|$ yielding $\dim | -\frac{1}{2}K |= 2$ in case (ii) of (b). Finally, if $A.(-\frac{1}{2}K) = 2 - n$ we can have type (0) only if $n = 6$, giving the case (i) of (c). Otherwise, we are in type (1) and n blown-up points are over G'. The conjugate set of blown-up points lies over $\overline{G'}$, hence the components F' and $\overline{F'}$ of C' are movable. This yields $\dim | -\frac{1}{2}K |= 3$ and $A + \overline{A}$ is mapped to $G' + \overline{G'}$ giving the statement (ii) of (c). □

Now we are ready to give new characterisations of the Moishezon twistor spaces introduced by LeBrun [LeB2] and studied by Kurke [Ku]. Recall the following result of Kurke [Ku] and Poon [Po4]:

Theorem 3.5. *If Z contains a pencil of divisors of degree one, then it is one of the Moishezon twistor spaces introduced by LeBrun [LeB2] and studied by Kurke [Ku].*

The following theorem provides new characterisations for these twistor spaces.

Theorem 3.6. *If Z is a twistor space satisfying condition (3.0) then the following properties are equivalent:*

(a) *Z contains a pencil of divisors of degree one.*

(b) $\dim | -\frac{1}{2}K |= 3$.

(c) $\dim | -\frac{1}{2}K |\geq 3$.

(d) *There exist exactly two reduced irreducible curves in Z having negative intersection number with $K^{-\frac{1}{2}}$. These two curves are smooth rational, form a conjugate pair $\{A, \overline{A}\}$ and $A.(-\frac{1}{2}K) = 2 - n$.*

(e) *There exists a smooth rational curve $A \subset Z$ with $A.(-\frac{1}{2}K) = 2 - n$.*

Proof: The implication (a) ⇒ (b) was shown in Lemma 3.1. (c) ⇒ (d) follows from Lemma 3.3(a) and Proposition 3.4. (d) ⇒ (e) is obvious. (e) ⇒ (a) can be shown by the same proof as in the case $n = 4$ given in [K1, Prop. 5.3.], therefore we omit it here. □

Theorem 3.7. *If Z is a twistor space satisfying condition (3.0) then the following properties are equivalent:*

(a) $\dim \mid -\frac{1}{2}K \mid = 2$.

(b) *There exists a smooth irreducible real rational curve $C_0 \subset Z$ with the property $C_0.(-\frac{1}{2}K) = 2(3-n)$. This is the unique irreducible reduced curve in Z having negative intersection number with $K^{-\frac{1}{2}}$.*

(c) *There exists a smooth real rational curve $C_0 \subset Z$ with $C_0.(-\frac{1}{2}K) < 0$.*

Proof: This follows from Lemma 3.3(a) and Proposition 3.4. For (c) $\Rightarrow$ (a) see also [K2, Thm. 2.1]. □

4 Computation of the algebraic dimension

The computation of the algebraic dimension of a specific compact complex manifold Z is often a very difficult task. It is known that in general there exists a line bundle $\mathcal{A} \in \operatorname{Pic} Z$ whose Iitaka dimension $\kappa(Z, \mathcal{A})$ is equal to $a(Z)$. It is an observation of Y.S. Poon [Po2], [Po3] that we can choose $\mathcal{A} = K^{-\frac{1}{2}}$ if Z is a simply connected twistor space and $\kappa(Z, K^{-\frac{1}{2}}) \neq -\infty$. If $S \in \mid -\frac{1}{2}K \mid$ is an irreducible smooth fundamental divisor on a twistor space, then the inequality $a(Z) \leq 1 + \kappa(S, K_S^{-1})$ is easy to see. But this will in general not suffice to compute $a(Z)$. The following theorem improves this situation a lot.

Theorem 4.1. *Let Z be a compact complex manifold, $\mathcal{F}$ and $\mathcal{A}$ line bundles on Z, $\Lambda \subseteq \mid \mathcal{F} \mid$ a one-dimensional linear system. Assume $a(Z) = \kappa(Z, \mathcal{A})$ and that the general member of Λ is irreducible and reduced. Then, for general $S \in \Lambda$, the following formula holds:*

$$a(Z) = 1 + \kappa(S, \mathcal{A} \otimes \mathcal{O}_S).$$

Proof: The linear system Λ does not have a fixed component since it contains an irreducible reduced member. Let $\varphi : Z \dashrightarrow \mathbb{P}^1$ be the meromorphic map defined by the pencil $\Lambda \subset \mid \mathcal{F} \mid$. By $B \subset Z$ we denote the set of indeterminacy of φ, that is the base locus of Λ. Using Hironaka's theorem on resolutions of singularities in the complex analytic case (see [AHV]), we can resolve the singularities of the graph space of φ to obtain a proper modification $\sigma : \widetilde{Z} \longrightarrow Z$ and a holomorphic map $\widetilde{\varphi} : \widetilde{Z} \longrightarrow \mathbb{P}^1$ such that: $\widetilde{Z}$ is a smooth compact complex manifold, $\widetilde{\varphi}$ is proper and surjective, σ induces an isomorphism $\widetilde{Z} \setminus \sigma^{-1}(B) \longrightarrow Z \setminus B$ and $\widetilde{\varphi} = \varphi \circ \sigma$ on $\widetilde{Z} \setminus \sigma^{-1}(B)$.
In particular, the generic fibre of $\widetilde{\varphi}$ is smooth (see [U, Cor. 1.8]). Since $\widetilde{Z}$ is irreducible and reduced and $\widetilde{\varphi}$ maps $\widetilde{Z}$ onto a smooth curve, the map $\widetilde{\varphi}$ is flat. But $\sigma^{-1}(B)$ has at least codimension one in Z and the general member S of Λ is, by assumption, an irreducible smooth divisor in Z, hence the generic fibre of $\widetilde{\varphi}$ is connected. This implies, the general fibre $\widetilde{S}$ of $\widetilde{\varphi}$ is smooth and irreducible and σ induces a proper modification $\sigma : \widetilde{S} \longrightarrow S = \sigma(\widetilde{S}) \in \Lambda$. By [U, 5.13] we obtain:

$$\kappa(\widetilde{S}, \sigma^*(\mathcal{A} \otimes \mathcal{O}_S)) = \kappa(S, \mathcal{A} \otimes \mathcal{O}_S)$$

and $\kappa(\widetilde{Z}, \sigma^*\mathcal{A}) = \kappa(Z, \mathcal{A})$.
Let m be a positive integer and consider the projective fibre space $\mathbb{P}(\widetilde{\varphi}_*\sigma^*\mathcal{A}^{\otimes m})$ over $\mathbb{P}^1$. We have meromorphic maps $\Phi_m : \widetilde{Z} \longrightarrow \mathbb{P}(\widetilde{\varphi}_*\sigma^*\mathcal{A}^{\otimes m})$ compatible with the maps to $\mathbb{P}^1$. The restriction of Φ_m to a generic fibre $\widetilde{S} \subset \widetilde{Z}$ of $\widetilde{\varphi}$ is the map given by the line bundle $(\sigma^*\mathcal{A}^{\otimes m}) \otimes \mathcal{O}_{\widetilde{S}}$ (see [U, (2.8)–(2.10)]). This implies for $m \gg 0$:

$$\dim \Phi_m(\widetilde{Z}) = 1 + \kappa(\widetilde{S}, \sigma^*\mathcal{A} \otimes \mathcal{O}_{\widetilde{S}}).$$

Since $\mathbb{P}^1$ and hence $\mathbb{P}(\widetilde{\varphi}_*\sigma^*\mathcal{A}^{\otimes m})$ are projective, we have $a(\Phi_m(\widetilde{Z})) = \dim \Phi_m(\widetilde{Z})$ and obtain

$$a(Z) = a(\widetilde{Z}) \geq a(\Phi_m(\widetilde{Z})) = 1 + \kappa(\widetilde{S}, \sigma^*\mathcal{A} \otimes \mathcal{O}_{\widetilde{S}}).$$

Finally, since we assumed $\kappa(Z, \mathcal{A}) = a(Z) \geq 0$, implying $h^0(\mathcal{A}^{\otimes m}) > 0$ for $m \gg 0$, we can apply [U, Thm. 5.11] to the proper holomorphic map $\widetilde{\varphi} : \widetilde{Z} \longrightarrow \mathbb{P}^1$ to obtain

$$\kappa(\widetilde{Z}, \sigma^*\mathcal{A}) \leq \kappa(\widetilde{S}, \sigma^*\mathcal{A} \otimes \mathcal{O}_{\widetilde{S}}) + 1.$$

All the inequalities together yield

$$a(Z) \geq 1 + \kappa(\widetilde{S}, \sigma^*\mathcal{A} \otimes \mathcal{O}_{\widetilde{S}}) \geq \kappa(\widetilde{Z}, \sigma^*\mathcal{A}) = \kappa(Z, \mathcal{A}) = a(Z)$$

which gives the claim. □

Definition 4.2. The *anti Kodaira dimension* of a compact complex variety X is the number $\kappa^{-1}(X) := \kappa(X, K_X^{-1})$.

Corollary 4.3. *Let Z be a compact, simply connected twistor space containing an irreducible fundamental divisor. If $h^0(K^{-\frac{1}{2}}) \geq 2$ and $S \in \mid -\frac{1}{2}K \mid$ is generic, then:*

$$a(Z) = 1 + \kappa^{-1}(S).$$

Proof: Our assumptions imply that there exists a pencil $\Lambda \subseteq \mid -\frac{1}{2}K \mid$ whose general member is irreducible and reduced. The general fundamental divisor of Z is contained in such a pencil. By Poon's theorem we have $a(Z) = \kappa(Z, K^{-\frac{1}{2}})$ and by the adjunction formula we obtain $K_S^{-1} \cong K^{-\frac{1}{2}} \otimes \mathcal{O}_S$. Application of Theorem 4.1 gives the result. □

5 Anti Kodaira Dimension of Rational Surfaces

The results of the previous section motivate the study of the anti Kodaira dimension of rational surfaces. Such studies were made by Sakai [Sa] but we are interested in a more detailed knowledge on the relationship between the anti Kodaira dimension and the numerical properties of the components of anti-canonical divisors. The desired results can also not be found in the papers

of E. Looijenga [Lo] and B. Harbourne [Hb] who studied surfaces containing effective anti-canonical divisors.
In contrast to the Kodaira dimension, the anti Kodaira dimension is not a birational invariant. Its behaviour under blow-ups becomes more transparent by the following results.

Lemma 5.1. *Let S' be a smooth surface, $P' \in S'$ a point and $C' \in |-K_{S'}|$ an anti-canonical divisor. By $\sigma : S \longrightarrow S'$ we denote the blow–up with centre P'. Then we have:*

(a) $\kappa^{-1}(S) \leq \kappa^{-1}(S')$ *and*

(b) if $\mathrm{mult}_{P'}(C') \geq 2$*, then* $\kappa^{-1}(S) = \kappa^{-1}(S')$.

Proof: Let $E \subset S$ be the exceptional divisor of σ. Then $\sigma^* K_{S'}^{-1} \cong K_S^{-1} \otimes \mathcal{O}_S(E)$. Because E is effective and $\sigma_* \mathcal{O}_S \cong \mathcal{O}_{S'}$ we obtain with $m \geq 1$:
$h^0(S, K_S^{-m}) \leq h^0(S, K_S^{-m} \otimes \mathcal{O}_S(mE)) = h^0(S, \sigma^* K_{S'}^{-m}) = h^0(S', K_{S'}^{-m})$. This proves (a).
Assume now $\mathrm{mult}_{P'}(C') \geq 2$, then $\tilde{C} := \sigma^* C' - 2E$ is effective. This is true for non-reduced C'. Using $K_S^{-2} \cong \sigma^* K_{S'}^{-2} \otimes \mathcal{O}_S(-2E) \cong \sigma^* K_{S'}^{-1} \otimes \mathcal{O}_S(\tilde{C})$, we obtain $h^0(S', K_{S'}^{-m}) = h^0(S, \sigma^* K_{S'}^{-m}) \leq h^0(S, \sigma^* K_{S'}^{-m} \otimes \mathcal{O}_S(m\tilde{C})) = h^0(S, K_S^{-2m})$. This implies $\kappa^{-1}(S') \leq \kappa(S, K_S^{-2}) = \kappa^{-1}(S)$ and we obtain (b). □

Theorem 5.2. *Let S' be a smooth rational surface and $C' \in |-K_{S'}|$ an anti-canonical divisor. The irreducible components (with reduced structure) are denoted by $C'_1, \dots, C'_r$. Assume that among the C'_ν there is no smooth rational (-1)–curve. Then we have:*
(a) $\exists \nu : (C'_\nu.(-K_{S'}))_{S'} > 0 \Rightarrow \kappa^{-1}(S') = 2$
(b) $\forall \nu : (C'_\nu.(-K_{S'}))_{S'} = 0 \Rightarrow \kappa^{-1}(S') \in \{0, 1\}$
(c) $\forall \mu : (C'_\mu.(-K_{S'}))_{S'} \leq 0$ and $\exists \nu : (C'_\nu.(-K_{S'}))_{S'} < 0 \Rightarrow \kappa^{-1}(S') = 0$
In the case (b) we have $\kappa^{-1}(S') = 0 \iff \forall m \geq 1 : h^0(C', N^{\otimes m}) = 0$*, with the abbreviation* $N := K_{S'}^{-1} \otimes \mathcal{O}_{C'}$.

Proof: We start with the observation that the exact sequence $0 \longrightarrow K_{S'} \longrightarrow \mathcal{O}_{S'} \longrightarrow \mathcal{O}_{C'} \longrightarrow 0$ and the rationality of S' imply $h^0(\mathcal{O}_{C'}) = 1$. As a consequence we obtain that C' is connected.
Recall that for arbitrary $D \in \mathrm{Pic}(S')$ and effective $D' \in \mathrm{Pic}(S')$ one always has $\kappa(S', D) \leq \kappa(S', D + D')$. If D is nef (i.e. for each effective divisor D' one has $(D.D')_{S'} \geq 0$), then $(D^2)_{S'} > 0$ if and only if $\kappa(S', D) = 2$.
To show (a) we assume first the existence of a component C'_ν with $({C'_\nu}^2)_{S'} > 0$. Such a divisor is nef and $\kappa(S', C'_\nu) = 2$, but $-K_{S'} - C'_\nu$ is effective, hence $\kappa^{-1}(S') = 2$.
Assume now $({C'_\mu}^2)_{S'} \leq 0$ for all μ. By assumption we have one component C'_ν with $(C'_\nu.(-K_{S'}))_{S'} > 0$. We show $({C'_\nu}^2)_{S'} = 0$ as follows: The genus formula gives $2p_a(C'_\nu) - 2 < 2p_a(C_\nu) - 2 + (C'_\nu.(-K_{S'}))_{S'} = ({C'_\nu}^2)_{S'} \leq 0$, hence the arithmetic genus $p_a(C'_\nu)$ vanishes and $C'_\nu \cong \mathbb{P}^1$. In turn, this implies

$0 \geq ({C'_\nu}^2)_{S'} = (C'_\nu.(-K_{S'}))_{S'} - 2 > -2$. By assumption, we have $({C'_\nu}^2)_{S'} \neq -1$ and conclude $({C'_\nu}^2)_{S'} = 0$. In particular, $-K_{S'}$ is not a multiple of C'_ν.
Because C' is connected, we can choose a component $C'_\mu \neq C'_\nu$ with $c := (C'_\mu.C'_\nu)_{S'} > 0$. We define $D := c \cdot C'_\mu + (1 - ({C'_\mu}^2)_{S'}) \cdot C'_\nu$ which is an effective divisor. Since $(D.C'_\mu)_{S'} = c \cdot ({C'_\mu}^2)_{S'} + (1 - ({C'_\mu}^2)_{S'}) \cdot c = c > 0$ and $(D.C'_\nu)_{S'} = c^2 + (1 - ({C'_\mu}^2)_{S'}) \cdot ({C'_\nu}^2)_{S'} = c^2 > 0$, we obtain $(D^2)_{S'} > 0$ and D is nef. If we choose $m = \max\{c, 1 - ({C'_\mu}^2)_{S'}\}$, then $\kappa^{-1}(S') = \kappa(S', K_{S'}^{-m}) \geq \kappa(S', D) = 2$ and (a) is proved.
If we have $(C'_\nu.(-K_{S'}))_{S'} = 0$ for all components of C', then $-K_{S'}$ is nef and $((-K_{S'})^2)_{S'} = 0$, hence $\kappa^{-1}(S') < 2$. Because we assumed that $\mid -K_{S'} \mid$ is non–empty, we have $\kappa^{-1}(S') \geq 0$ and (b) is shown.
To show (c) we can apply [Lo, (1.3)] which proves that the matrix $((C'_i.C'_j)_{S'})_{i,j}$ is negative definite. Hence, in the Zariski decomposition $C' = P' + N'$ of the divisor C' we have $P' = 0$ (see [Sa]). This implies $\kappa^{-1}(S') = \kappa(S', C') = \kappa(S', P') = 0$, hence (c).
To distinguish, in the case (b), anti Kodaira dimensions zero and one, we consider the exact sequence ($m \geq 1$):

$$0 \longrightarrow K_{S'}^{-(m-1)} \longrightarrow K_{S'}^{-m} \longrightarrow N^{\otimes m} \longrightarrow 0 \tag{2}$$

If $h^0(C', N^{\otimes m}) = 0$ for all $m \geq 1$, we obtain $h^0(K_{S'}^{-m}) = 1$ for $m \geq 1$ and $\kappa^{-1}(S') = 0$. On the other hand, if there exists some $m \geq 1$ with $h^0(C', N^{\otimes m}) > 0$, then we let m_0 be the smallest one with this property. From the sequence (2) we obtain $h^0(S', K_{S'}^{-m}) = 1$ for $0 \leq m < m_0$. We have $h^2(S', K_{S'}^{-m}) = h^0(S', K_{S'}^{m+1}) = 0$ for $m \geq 0$ (because S' is rational) and $((-K_{S'})^2)_{S'} = 0$ (in case (b)) and obtain from the Riemann–Roch formula $h^0(K_{S'}^{-m}) - h^1(K_{S'}^{-m}) = 1$. Therefore, $h^1(K_{S'}^{-m}) = 0$ for $0 \leq m < m_0$. The exact sequence (2) with $m = m_0$ implies now $h^0(K_{S'}^{-m_0}) > 1$, thus $\kappa^{-1}(S') > 0$. □

Remark 5.3. It is a remarkable fact that the numerical information contained in an anti-canonical divisor is not sufficient for the computation of the anti Kodaira dimension, if its components are orthogonal to the canonical class. This phenomenon also appears in the paper [Sa]. It is the reason that it is difficult to construct simply connected twistor spaces of algebraic dimension two (see [CK1]).

Corollary 5.4. *Let S be a smooth rational surface, $C \in \mid -K_S \mid$ an effective anti-canonical divisor with components $C_1, \ldots, C_r$ and denote $N := K_S^{-1} \otimes \mathcal{O}_C$. Assume that among the C_ν there is no smooth rational (-1)-curve or that S cannot be blown-down to a surface with the properties of Theorem 5.2 (b). Then, the anti–Kodaira dimension $\kappa^{-1}(S)$ is determined by the pair (C, N).*

Proof: Since $(C_\nu^2)_S = 2p_a(C_\nu) - 2 + (C_\nu.(-K_S))_S$ and $(C_\nu.(-K_S))_S = \deg(N \otimes \mathcal{O}_{C_\nu})$, this follows from Theorem 5.2. □

6 Blow-up graphs

In this section we develop a method to handle the numerical information of an anti-canonical divisor on a surface obtained by a sequence of blow-ups from $\mathbb{P}^1 \times \mathbb{P}^1$ at points lying over an anti-canonical curve with four irreducible components.

In view of our application to twistor spaces, we are only interested in blow-ups of conjugate pairs of points to have real structures on all the blown-up surfaces. We equip $S = \mathbb{P}^1 \times \mathbb{P}^1$ with the real structure given by the antipodal map on the first factor and the usual real structure on the second (cf. [K1, Ch. 3]). Choose an anti-canonical curve $C = F + \overline{F} + G + \overline{G}$ with $\overline{F} \neq F \in \mid \mathcal{O}(0,1) \mid$ and $G \in \mid \mathcal{O}(1,0) \mid$.

Let $S^{(k)} \longrightarrow S^{(k-1)} \longrightarrow \dots \longrightarrow S^{(0)} = S$ be a sequence of blow-ups at each step of which we blow up a conjugate pair of points lying on the effective anti-canonical divisor $C^{(i)}$ which is mapped to $C \subset S$. Denote by $\sigma_i : S^{(k)} \longrightarrow S^{(i)}$ the partial blow-up ($0 \leq i \leq k$). The curve $C^{(k)} \subset S^{(k)}$ is a "cycle of rational curves" as defined in [K1, Def. 3.5] with an even number of irreducible components. Denote its components by $C_1, C_2, \dots, C_{2m}$ such that C_i intersects C_{i-1} and C_{i+1} (we consider indices modulo $2m$). We have $2 \leq m \leq k+2$. We associate the following graph to the given sequence of blow-ups:

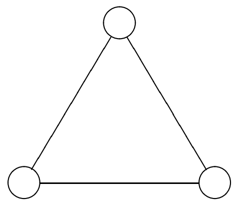
The graph contains $k+2$ vertices, some of which are possibly marked. We let m of these vertices correspond to the pairs of conjugate curves in $S^{(k)}$: (C_1, C_{m+1}), (C_2, C_{m+2}), $\dots, (C_m, C_{2m})$. We denote these vertices by $v_1, \dots, v_m$ and call them *internal* vertices of the graph. Two of these vertices v_i and v_j are joined by one edge if and only if there is an integer $0 \leq r \leq k$ such that $\sigma_r(C_i)$ and $\sigma_r(C_j)$ are curves and $\sigma_r(C_i \cup C_{m+i}) \cap \sigma_r(C_j \cup C_{m+j}) \neq \emptyset$.

The graph can also contain *external* vertices. These vertices $v_{m+1}, \dots, v_k$ correspond bijectively to conjugate pairs of irreducible smooth rational curves contracted under $\sigma_0 : S^{(k)} \longrightarrow S$. These are those strict transforms in $S^{(k)}$ of exceptional curves of the blow-ups which are not components of $C^{(k)}$. Hence, for every external vertex v there exists an integer $1 \leq r(v) \leq k-1$ such that the curves corresponding to v are the strict transforms of the exceptional curves of the blow-up $S^{(r(v)+1)} \longrightarrow S^{(r(v))}$. The number of components of $C^{(r(v))}$ and of $C^{(r(v)+1)}$ are equal and the blown-up points lie on $\sigma_{r(v)}(C_i \cup C_{m+i})$ for precisely one i. We denote this index i by $i(v)$. An external vertex is connected with an other external vertex by an edge if and only if the corresponding pairs of conjugate curves in $S^{(k)}$ have nonempty intersection. Every external vertex v is connected by an edge with precisely one internal vertex, namely with $v_{i(v)}$. Finally, we equip an external vertex v with a marking if and only if for every external vertex w connected with v by an edge we have $r(v) \leq r(w)$. Internal vertices are never marked. In our pictures we shall draw the vertices as circles and indicate the marked vertices by an asterisk inside this circle.

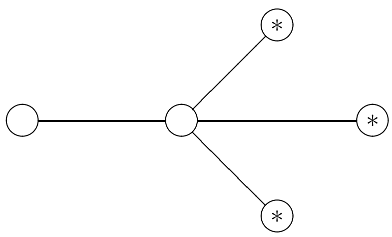
In the description of the following examples the reader should keep in mind that we consider only blow-ups of conjugate pairs of points. Thus, if there is written: "if we blow up $P \in F$, then $\dots$ ", one should read: "if we blow up

$P \in F$ and the conjugate point $\overline{P} \in \overline{F}$, then ... ".

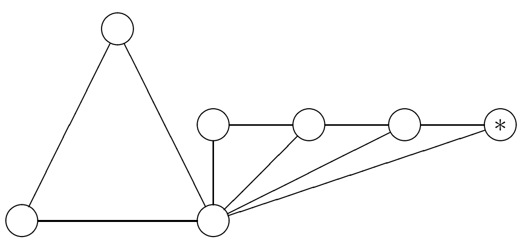
Example 6.1. If $k = 1$ and we blow up the point $F \cap G$ then the graph is the following:



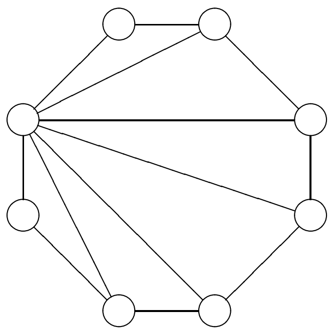
Example 6.2. If we blow up k distinct points on F, which are not contained in $G \cup \overline{G}$, then we have $m = 2$ and the graph contains k marked external vertices which are joined with one of the internal vertices. If $k = 3$, the graph looks like:



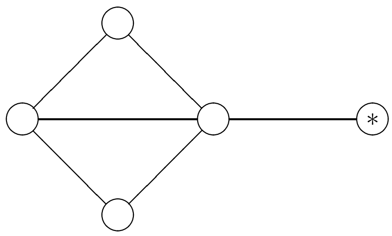
Example 6.3. If in the previous example we blow up four times the same point on the strict transforms of F, the resulting graph can be drawn as follows:



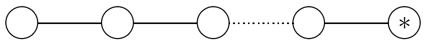
Example 6.4. If $k \geq 2$ and we always blow up the unique point over the point $F \cap G$ lying on the strict transform of F, we obtain a $(k+2)$-gon divided into triangles by the diagonals from one vertex to all other vertices. All vertices are internal in this case. If $k = 6$ the graph looks like:



Example 6.5. If $k = 3$ and we blow up the two points $F \cap (G \cup \overline{G})$ and a third point on F, the graph is the following:



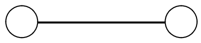
Definition 6.6. A *blow-up graph* is a graph consisting of a finite number of vertices, and edges connecting distinct vertices. Some of the vertices are marked. Between two vertices there exists at most one edge. This graph can be drawn in the real plane such that a subset of at least two non-marked vertices (called internal vertices) form a regular m-gon such that the edges connecting them are represented by mutually disjoint diagonals giving a triangulation of this m-gon. (If $m = 2$ this means that the two vertices are connected by one edge.) The remaining vertices are called external vertices. The subgraph formed by these vertices and the edges among them is the disjoint union of chains like this:



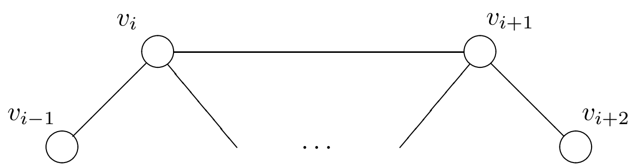
such that each chain contains precisely one marked vertex. The marked vertex of such a chain is an endpoint, i.e. is not connected with two other vertices in that chain. Finally, every external vertex is connected with precisely one internal vertex in such a way that the vertices of one chain are connected with the same internal vertex.
If v is a vertex of such a graph, we denote by $n(v)$ the number of edges adjacent to this vertex v plus the number of its markings (which is zero or one).

Proposition 6.7. *The graph associated to a blow-up in the way defined above is always a blow-up graph in the sense of Definition 6.6. Moreover, the internal vertices $v_1, \ldots, v_m$ form the vertices of the m-gon of the blow-up graph such that v_i and v_{i+1} are neighbours along the boundary of the m-gon.*

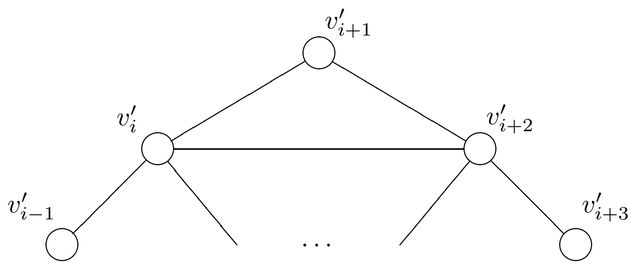
The self-intersection number of each of the curves corresponding to a vertex v is equal to $1 - n(v)$.
Every blow-up graph appears as a graph associated to a sequence of blow-ups.

Proof: We prove the proposition by induction on $k \geq 0$. If $k = 0$ we obtain $m = 2$ and the graph is a 2-gon:  consisting of 2 internal vertices. In this case the proposition is clear, because $(F^2)_S = (G^2)_S = 0$ on S.
For the inductive step let Γ be the graph associated to $S^{(k)} \longrightarrow \ldots \longrightarrow S^{(0)}$. Let $S^{(k+1)} \longrightarrow S^{(k)}$ be a further blow-up of a conjugate pair of points $\{P, \overline{P}\}$ lying on $C^{(k)}$ and denote by Γ' the graph associated to the sequence of blow-ups $S^{(k+1)} \longrightarrow S^{(k)} \longrightarrow \ldots \longrightarrow S^{(0)}$. Assume that Γ is a blow-up graph and the self-intersection numbers in $S^{(k)}$ are those given by the claim. Then there are three possibilities:

(1) P is a singular point of $C^{(k)}$, or equivalently, P is contained in two components of $C^{(k)}$. The corresponding internal vertices v_i and v_{i+1} are neighbours in Γ along the boundary of the m-gon of internal vertices. The exceptional curves of $S^{(k+1)} \longrightarrow S^{(k)}$ are components of $C^{(k+1)}$, hence correspond to a new internal vertex of Γ'. Therefore, the graph Γ' contains an $(m+1)$-gon of internal vertices $\{v'_1, \ldots, v'_{m+1}\}$ and is obtained from Γ by adding a new internal vertex, which is connected with v_i and v_{i+1}. The numbering of the vertices in Γ' can be chosen such that $v'_j = v_j$ if $1 \leq j \leq i$, v'_{i+1} is the new vertex and $v'_{j+1} = v_j$ if $i+1 \leq j \leq m$. If part of Γ looks like the following picture:

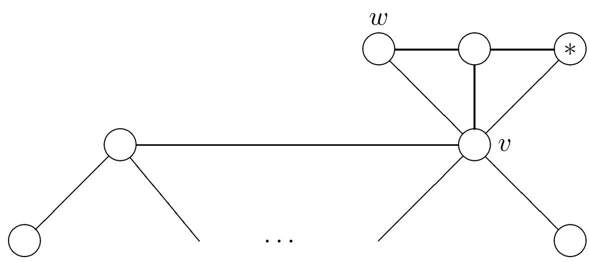


the graph Γ' is of the following kind:

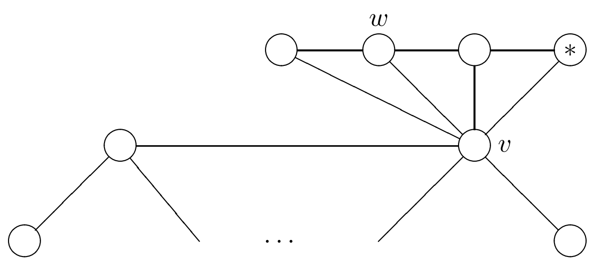


This procedure will be recalled by saying "we added an internal triangle".

(2) P is a smooth point on $C^{(k)}$. In this case, the conjugate pair of exceptional curves of the blow-up $S^{(k+1)} \longrightarrow S^{(k)}$ is not contained in $C^{(k+1)}$. It corresponds, therefore, to a new external vertex of Γ'. Assume P lies on the strict transform E of an exceptional curve of one of the previous blow-ups, which is not a component of $C^{(k)}$. Since E intersects $C^{(k)}$ we must have $E^2 = -1$. Hence, by the inductive hypothesis, the corresponding external vertex w is one end of its chain of external vertices. Moreover, if this chain consists of more that one vertex, it is the non-marked end, because this is the only vertex on this chain having $n(w) = 2$. Let v be the internal vertex being connected with w. The graph Γ' is obtained from Γ by adding a new external vertex which is connected with v and w. It, therefore, becomes the unmarked end of its chain of external vertices. For a graph Γ containing:

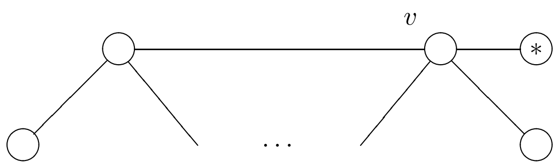


we obtain a graph Γ' like the following:

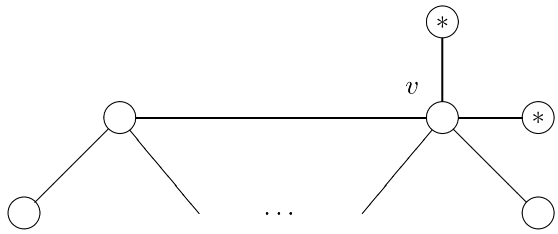


We call this procedure "adding an external triangle".

(3) P is a smooth point on $C^{(k)}$ not lying on a curve corresponding to an external vertex. Let v be the internal vertex of Γ corresponding to the components of $C^{(k)}$ containing P. Then we obtain Γ' by adding a marked external vertex to Γ and connect it with v. For example, from a graph Γ containing:



we obtain Γ' with:



We shall say, we "added a marked (external) vertex".

In each of these three cases it is clear that Γ' is again a blow-up graph and the self-intersection numbers of the curves corresponding to the vertices decrease by the number of additional edges at such a vertex. The self-intersection number of the curves corresponding to a new vertex is -1 since these are the exceptional curves of the blow-up. Hence, by the inductive hypothesis we obtain that the self-intersection numbers can be computed as $1 - n(v)$.
To show that every blow-up graph is associated to a sequence of blow-ups, we first observe that we can construct every blow-up graph Γ in the following way:

- We start with the 2-gon.
- We carry out $(m-2)$ steps of "adding an internal triangle" and obtain a triangulated m-gon.
- We add the necessary number of marked vertices.
- We add "external triangles".

As seen above, each step of this procedure corresponds to a blow-up of a conjugate pair of points, such that there exists a sequence of blow-ups determining the given graph Γ. $\square$

Remark 6.8. Observe that every vertex of a blow-up graph is connected with at least one internal vertex by an edge. For every marked vertex v we have $n(v) \in \{2, 3\}$. Every vertex v with $n(v) > 3$ is an internal vertex. The set of internal vertices is determined by the vertices, their edges and markings. Therefore, we don't need a special marking for them.

Next we give the interpretation of the results of Section 5 in terms of our graphs associated to blow-ups of surfaces.
Let Γ and Γ' be blow-up graphs such that Γ is obtained from Γ' by adding an internal triangle. This corresponds to a blow up $S \longrightarrow S'$ of a singular point (more precisely a conjugate pair of such points) on an anti-canonical divisor of a rational surface S'. By Lemma 5.1(b) these surfaces have the same anti Kodaira dimension. If for every sequence of blow-ups $S \longrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ with

associated blow-up graph Γ the anti Kodaira dimension $\kappa^{-1}(S)$ is the same, we define the anti Kodaira dimension of the graph Γ by $\kappa^{-1}(\Gamma) := \kappa^{-1}(S)$ and say that the *graph determines the anti Kodaira dimension.* This property of a graph is not changed by adding an internal triangle.

Theorem 6.9. *Let $S \longrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ be a sequence of blow-ups of conjugate pairs of points as before with anti-canonical divisor $C = \sum_{i=1}^{2m} C_i$ and Γ the associated blow-up graph. Assume that $(C_i^2)_S \neq -1$ for all i. Then:*

(a) The graph Γ cannot be obtained from an other blow-up graph by adding an internal triangle.

(b) The graph Γ determines the anti Kodaira dimension if and only if it contains an internal vertex v with $n(v) \neq 3$.

(c) If Γ contains an internal vertex v with $n(v) \leq 2$ then $\kappa^{-1}(\Gamma) = 2$.

(d) If for all internal vertices v of Γ we have $n(v) \geq 3$ and for at least one of these vertices this inequality is strict, then $\kappa^{-1}(\Gamma) = 0$.

If $m = 2$ and $n(v_1) = 2$ (that is $(C_1^2)_S = -1$), then

(b') The graph Γ determines the anti Kodaira dimension if and only if $n(v_2) \neq 5$.

(c') If $n(v_2) \leq 4$, then $\kappa^{-1}(\Gamma) = 2$.

(d') If $n(v_2) \geq 6$, then $\kappa^{-1}(\Gamma) = 0$.

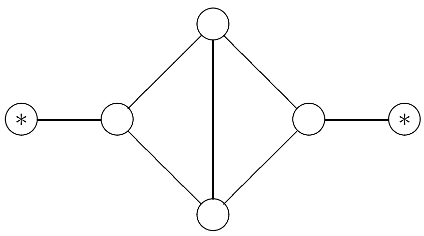
Proof: To be able to apply Theorem 5.2 we recall that the conjugate pairs of components of the anti-canonical divisor on the surface S correspond to the internal points of the associated blow-up graph. These components are irreducible smooth rational curves. By the adjunction formula for such a component C_i we obtain $(C_i.(-K_S))_S = 3 - n(v_i)$. (We keep denoting the internal vertex of Γ corresponding to C_i, $1 \leq i \leq m$ by v_i.) Therefore almost all statements are purely a translation of the statements of Theorem 5.2. We have to prove only two things.
First, the assertions (c') and (d'), if $m = 2$ and $(C_1^2)_S = (C_3^2)_S = -1$. This correspond to $n(v_1) = 2$. This case is not covered by Theorem 5.2. Let $-l = (C_2^2)_S = (C_4^2)_S$, then $n(v_2) = l + 1$. We can contract C_1 and C_3 to obtain a smooth rational surface S' with $C' = C_2' + C_4' \in\mid -K_{S'} \mid$ being the image of C. Then we have $(C_2'^2)_{S'} = (C_4'^2)_{S'} = 2 - l$ and $\kappa^{-1}(S) = \kappa^{-1}(S')$ by Lemma 5.1 (b). If $l = 3$ it is easy to see that C' is nef and big, hence $\kappa^{-1}(S') = 2$. If $l \neq 3$ we can apply Theorem 5.2 and obtain: $\kappa^{-1}(S) = \kappa^{-1}(S') = 2$ if $l < 4$, $\kappa^{-1}(S) = \kappa^{-1}(S') \in \{0, 1\}$ if $l = 4$ and $\kappa^{-1}(S) = \kappa^{-1}(S') = 0$ if $l > 4$. This shows (c') and (d').
Second, we have to prove (b) and (b'). So, we are looking for two sequences of blow-ups $S_0 \longrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ and $S_1 \longrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ with same associated graph Γ but with $\kappa^{-1}(S_j) = j$. The graph Γ is required to fulfill $m = 2, n(v_1) = 2, n(v_2) = 5$

or should be a blow-up graph whose internal vertices v all have $n(v) = 3$. Using Remark 6.11 below this follows by a similar argumentation as in [CK1, Section 4]. □

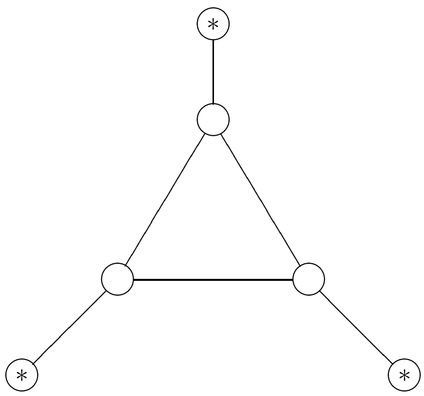
Remark 6.10. A blow-up graph Γ contains more information than necessary for computing $\kappa^{-1}(\Gamma)$. The values for all $n(v)$ at internal vertices would suffice. We shall see later (Sections 7, 8) the reason for using such graphs.

Remark 6.11. It is an easy observation that every triangulated m-gon contains at least one vertex with more than three incident edges, provided $m \geq 5$. This implies, together with Theorem 6.9 (b), that a blow-up graph determines the anti Kodaira dimension, provided it contains at least 5 internal vertices and it cannot be obtained from an other blow-up graph by adding an internal triangle. The following five blow-up graphs are the only ones with the property that each internal vertex v has $n(v) = 3$.

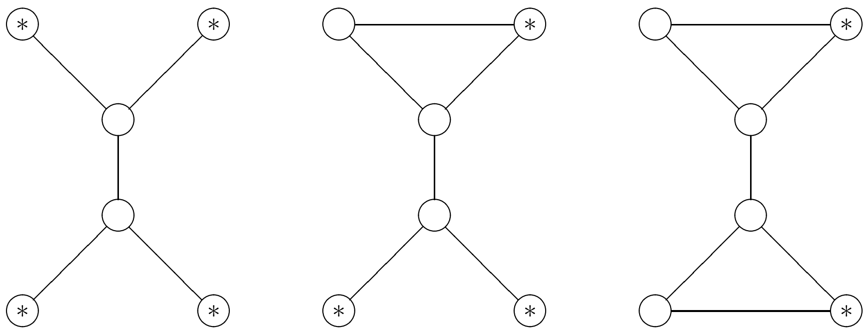
If m (the number of internal vertices) is four, there is only one possibility:



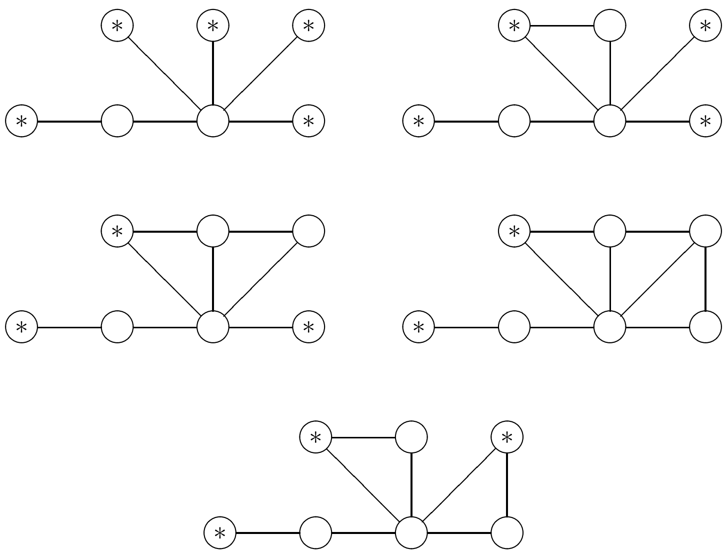
If $m = 3$ the following graph is the unique blow-up graph with precisely three edges starting at each internal vertex:



If $m = 2$ there exist three possibilities, whose differences concern only the markings and the edges between external vertices:

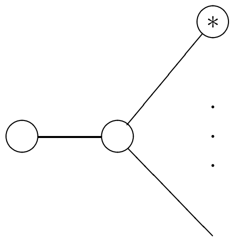


All these five graphs have six vertices. Such graphs are obtained by blowing up four pairs of conjugate points.
To obtain a complete understanding of all blow-up graphs not determining the anti Kodaira dimension, we describe below the graphs mentioned in item (b') of the above theorem. There are only five blow-up graphs with $m = 2$, $n(v_1) = 2$ and $n(v_2) = 5$. Again, they differ only in the markings and edges between the external vertices:

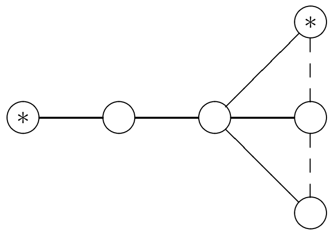


These graphs appear by blowing up five conjugate pairs of points, starting with $\mathbb{P}^1 \times \mathbb{P}^1$. But, as seen in the proof of Theorem 6.9, on such a surface we can contract a pair of (-1)-curves to arrive at a smooth rational surface, having by Lemma 5.1 (b) the same anti-Kodaira dimension as the surface we started with. The blown-down surface is obtained by blowing-up four conjugate pairs of points which are sitting on a conjugate pair of curves of type $(1, 1)$ in $\mathbb{P}^1 \times \mathbb{P}^1$. This situation appears as type (3) at the beginning of Section 8. In the paper [CK1] we studied a similar situation and showed how to construct twistor spaces of algebraic dimension one and two by moving the blown-up points a bit.

Remark 6.12. If Γ is a graph as in Theorem 6.9 (c) or (c'), then $m = 2$ or it contains an internal vertex v with $n(v) = 1$, since a vertex with $n(v) = 2$ corresponds to a (-1)-curve. Such a graph contains exactly two internal vertices (i.e. $m = 2$) and one of them is connected with at most one external vertex. If one internal vertex is not connected with an external vertex, we have no further restrictions:



If both vertices are connected with an external vertex, then the number of external vertices is at most four, one of them is connected with one internal vertex, the remaining at most three with the other internal vertex:



According to Theorems 5.2 and 6.9 the blow-up graphs associated to a sequence of blow ups resulting in a surface with anti Kodaira dimension two are precisely those which are obtained by adding a finite number of internal triangles to one of the graphs described in this remark. In particular, we find among them all blow-up graphs having no external vertex.

7 Small deformations of blow-up graphs

In this section we study small deformations of rational surfaces obtained by blowing up $\mathbb{P}^1 \times \mathbb{P}^1$. The results will be used in Section 8 to show the existence of twistor spaces containing fundamental divisors with certain properties. We study the behaviour of blow-up graphs under small deformations, so that we can apply the results of the previous sections to the deformed surfaces.

Definition 7.1. Let

$$S^{(k)} \xrightarrow{\sigma_k} S^{(k-1)} \xrightarrow{\sigma_{k-1}} \ldots \xrightarrow{\sigma_2} S^{(1)} \xrightarrow{\sigma_1} S^{(0)} = S$$

be a sequence of blow-ups of points $P^{(i)} \in S^{(i)}$ on surfaces. We call a flat family of surfaces $\mathcal{S} \longrightarrow \mathcal{T}$ together with a $\mathcal{T}$-morphism $\mathcal{S} \longrightarrow S \times \mathcal{T}$ *a family of blow-ups* of S, if we are given $\mathcal{T}$-flat families $\mathcal{S}_i \longrightarrow \mathcal{T}$ with sections $\varphi_i : \mathcal{T} \longrightarrow \mathcal{S}_i$

$(0 \le i \le k-1)$ such that $\mathcal{S}_{i+1} \longrightarrow \mathcal{T}$ is obtained by blowing up $\mathcal{S}_i \longrightarrow \mathcal{T}$ along $\varphi_i(\mathcal{T}) \subset \mathcal{S}_i$ and $\mathcal{S}_0 = S \times \mathcal{T}$, $\mathcal{S} = \mathcal{S}_k$. We say that this family is a *deformation of the given sequence of blow-ups*, if there is a point $0 \in \mathcal{T}$ such that for $1 \le i \le k$ the fibre of the blow-up morphism $\mathcal{S}_i \longrightarrow \mathcal{S}_{i-1}$ over $0 \in \mathcal{T}$ is isomorphic to the given blow-up $S^{(i)} \longrightarrow S^{(i-1)}$.

Proposition 7.2. *Let S be a smooth surface and $A, B \subset S$ smooth curves intersecting transversally at $P \in S$. Consider a sequence of morphisms*

$$S^{(k)} \xrightarrow{\sigma_k} S^{(k-1)} \xrightarrow{\sigma_{k-1}} \ldots \xrightarrow{\sigma_2} S^{(1)} \xrightarrow{\sigma_1} S^{(0)} = S$$

where σ_{i+1} is the blow-up of a point $P^{(i)} \in S^{(i)}$. Denote by $\sigma^{(i)} : S^{(i)} \longrightarrow S$ the composition $\sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_i$ and define inductively $A^{(0)} = A$, $B^{(0)} = B$ and $A^{(i)}, B^{(i)} \subseteq S^{(i)}$ to be the strict transforms of $A^{(i-1)}, B^{(i-1)} \subseteq S^{(i-1)}$.
Assume: $P^{(0)} = P \in A^{(0)} \cap B^{(0)}$ and $\sigma^{(i)}(P^{(i)}) = P$ for all $1 \le i \le k-1$. $P^{(i)} \in A^{(i)}$ if $0 \le i \le a$ and $P^{(i)} \in B^{(i)}$ if $a+1 \le i \le k-1$ for an integer $0 \le a \le k-1$.
Let $\pi : \{0, 1, \ldots, k-1\} \longrightarrow \{1, 2, \ldots, \beta\}$ be a monotone partition of $\{0, 1, \ldots, a\}$ and $\{a+1, \ldots, k-1\}$. This means β is a positive integer and π is a surjective map with the properties $i \le j \Rightarrow \pi(i) \le \pi(j)$ and $\alpha := \pi(a) < \pi(a+1)$. The fibres of π form then the usual partition sets $\pi_i := \pi^{-1}(i) \subseteq \{0, 1, \ldots, k-1\}$.
Then there exists a deformation $\mathcal{S}_\pi \longrightarrow \mathcal{T}_\pi$ of the given sequence of blow-ups, such that every neighbourhood of the special point $0 \in \mathcal{T}_\pi$ contains a point $t \in \mathcal{T}_\pi$ whose fibre $S_t := (\mathcal{S}_\pi)_t$ is isomorphic to a sequence of blow-ups

$$S_t \cong S_t^{(k)} \longrightarrow S_t^{(k-1)} \longrightarrow \ldots \longrightarrow S_t^{(1)} \longrightarrow S_t^{(0)} = S$$

at points $Q^{(i)} \in S_t^{(i)}$ with the following property (where we defined $A_t^{(i)}, B_t^{(i)}$ and $\sigma_t^{(i)}$ in the same way as $A^{(i)}, B^{(i)}$ and $\sigma^{(i)}$):
$Q^{(i)} \in A_t^{(i)}$ if $0 \le i \le a$,
$Q^{(i)} \in B_t^{(i)}$ if $a < i \le k-1$,
$\sigma_t^{(i)}(Q^{(i)}) \ne P$ for all i and
$\sigma_t^{(i)}(Q^{(i)}) = \sigma_t^{(j)}(Q^{(j)})$ if and only if $\pi(i) = \pi(j)$.
In particular, there exists a deformation $\mathcal{S} \longrightarrow \mathcal{T}$ of the given sequence of blow-ups, such that every neighbourhood of the special point $0 \in \mathcal{T}$ contains a point $t \in \mathcal{T}$ whose fibre $\mathcal{S}_t$ is isomorphic to a blow-up of S at k distinct points $Q^{(i)}$ $(0 \le i \le k-1)$ with the property $Q^{(i)} \in A \setminus \{P\}$ for $0 \le i \le a$ and $Q^{(i)} \in B \setminus \{P\}$ for $a < i \le k-1$.

The proof requires some preparation and will be given after Lemma 7.5.

Definition 7.3. We say that a quadruple (S, A, B, P) is *admissible* with parameters in T, if $S \longrightarrow T$ is a flat family of smooth projective surfaces, $A, B \subset S$ are flat sub-families of smooth curves and $P = A \cap B$ is a section of S over T.

Lemma 7.4. *Let (S, A, B, P) be admissible with parameters in T. We define $\widetilde{S} \longrightarrow S \times_T A$ to be the blow-up along the graph $\Gamma_A \subset S \times_T A$ of the embedding*

$A \subset S$, i.e. Γ_A is the intersection of $S \times_T A$ with the diagonal of $S \times_T S$. By $\widetilde{A} \subset \widetilde{S}$ and $\widetilde{B} \subset \widetilde{S}$ we denote the strict transform of $A \times_T A$ and $B \times_T A$ respectively. Let finally $\widetilde{P} := \widetilde{A} \cap \widetilde{B}$ and $\widetilde{S} \longrightarrow \widetilde{T} := A$ be the morphism induced by the projection $S \times_T A \longrightarrow A$, then $(\widetilde{S}, \widetilde{A}, \widetilde{B}, \widetilde{P})$ is admissible with parameters in $\widetilde{T}$. Furthermore, $\widetilde{A} \longrightarrow A \times_T A$ and $\widetilde{P} \longrightarrow P \times_T A$ are isomorphisms and $\widetilde{B} \longrightarrow B \times_T A$ is the blow-up of $P \times_T P$, where the morphisms are those induced by the blow-up $\widetilde{S} \longrightarrow S \times_T A$.

Proof: Since $\Gamma_A \subset S \times_T A$ is a section of the projection $S \times_T A \longrightarrow A$ we obtain flatness of $\widetilde{S} \longrightarrow \widetilde{T} = A$. Since $(A \times_T A) \cap (B \times_T A) = P \times_T A$ and $\Gamma_A \cap (P \times_T A) = P \times_T P$ is a divisor in $P \times_T A$, we obtain an isomorphism $\widetilde{P} = \widetilde{A} \cap \widetilde{B} \longrightarrow P \times_T A$, hence $\widetilde{P} \subset \widetilde{S}$ is a section of $\widetilde{S} \longrightarrow \widetilde{T}$. Because $(A \times_T A) \cap \Gamma_A$ is the diagonal in $A \times_T A$ and A has relative dimension one over T, we obtain an isomorphism $\widetilde{A} \longrightarrow A \times_T A$, which is, hence, a flat family of smooth curves.
On the other hand, $\Gamma_A \cap (B \times_T A) = P \times_T P$, hence $\widetilde{B} \longrightarrow B \times_T A$ is the blow-up of the sub-scheme of codimension two $P \times_T P \subset B \times_T A$. Since $\widetilde{B}, A$ are smooth we obtain flatness of $\widetilde{B} \longrightarrow A$ as soon as we have shown that all fibres are one-dimensional. But this is clear since the fibres of $\widetilde{S}$ over $\widetilde{T} = A$ are surfaces which are obtained by the blow-up of precisely one point of the corresponding fibre of S over T. □

In the following we denote the admissible quadruple $(\widetilde{S}, \widetilde{A}, \widetilde{B}, \widetilde{P})$ constructed in the lemma by $\mathbb{B}_A(S, A, B, P)$. Interchanging the role of A and B we obtain $\mathbb{B}_B(S, A, B, P)$ with parameters in $\widetilde{T} = B$.

We use this construction to define recursively the deformation which will be used in the proof of the proposition.

Let $\mathcal{T}^{(0)}$ be a point, $\mathcal{S}^{(0)} := S, \mathcal{A}^{(0)} := A, \mathcal{B}^{(0)} := B$ and $\mathcal{P}^{(0)} = \mathcal{A}^{(0)} \cap \mathcal{B}^{(0)} = P$. Then $(\mathcal{S}^{(0)}, \mathcal{A}^{(0)}, \mathcal{B}^{(0)}, \mathcal{P}^{(0)})$ is admissible with parameters in $\mathcal{T}^{(0)}$. We define

$$(\mathcal{S}^{(i+1)}, \mathcal{A}^{(i+1)}, \mathcal{B}^{(i+1)}, \mathcal{P}^{(i+1)}) := \begin{cases} \mathbb{B}_{\mathcal{A}^{(i)}}(\mathcal{S}^{(i)}, \mathcal{A}^{(i)}, \mathcal{B}^{(i)}, \mathcal{P}^{(i)}) & \text{if } 0 \leq i \leq a, \\ \mathbb{B}_{\mathcal{B}^{(i)}}(\mathcal{S}^{(i)}, \mathcal{A}^{(i)}, \mathcal{B}^{(i)}, \mathcal{P}^{(i)}) & \text{if } a < i < k. \end{cases}$$

The following lemma provides more information on the parameter spaces $\mathcal{T}^{(i+1)} = \mathcal{A}^{(i)}$ if $i \leq a$ and $\mathcal{T}^{(i+1)} = \mathcal{B}^{(i)}$ if $a < i \leq k-1$.

The careful reader will observe that we abuse notation a bit by using P to denote on one hand the point $P \in S$ and on the other hand the reduced closed sub-scheme $P \subset S$ supported by this point. This allows us to write $P \times A^i$ instead of $\{P\} \times A^i$ and will not cause confusion.

Lemma 7.5. *If $0 \leq i \leq a+1$ we have:*

(a) $\mathcal{A}^{(i)} \cong A \times A^i$ and the structure of a family of curves in $\mathcal{S}^{(i)}$ is given by the projection to the last i components (i.e. the first component is omitted) $\mathcal{A}^{(i)} \cong A \times A^i \longrightarrow \mathcal{A}^{(i-1)} \cong A^i = \mathcal{T}^{(i)}$.

(b) Under this isomorphism, $\mathcal{P}^{(i)} \subset \mathcal{A}^{(i)}$ corresponds to $P \times A^i$.

(c) $\mathcal{B}^{(i)} \longrightarrow B \times A^i$ *is obtained by successively blowing up first* $P \times H_i$ *and then the strict transforms of* $P \times H_{i-1}, P \times H_{i-2}, \dots, P \times H_1$*, where we denote by* $H_m \subset A^i$ *the hyper-surface being the preimage of* $P \in A$ *under the* m*-th projection* $A^i \longrightarrow A$*. Again, the projection* $B \times A^i \longrightarrow A^i$ *gives the structure map* $\mathcal{B}^{(i)} \longrightarrow A^i = \mathcal{T}^{(i)}$*.*

If $s \geq 2$ *and* $i = a + s \leq k - 1$ *we have:*

(d) $\mathcal{B}^{(i)} \longrightarrow B^s \times A^{a+1}$ *is obtained by successively blowing up the sub-varieties of codimension two* $H'_s \times H_{a+1}$*, then the strict transforms of* $H'_s \times H_a, \dots,$ $H'_s \times H_1$ *followed by the same sequence with* H'_{s-1} *replacing* H'_s*, etc. up to* $H'_1 \times H_1$*. Here we let* $H'_n \subset B^s$ *be the preimage of* $P \in B$ *under the* n*-th projection* $B^s \longrightarrow B$*. The map* $\mathcal{B}^{(i)} \longrightarrow \mathcal{B}^{(i-1)} = \mathcal{T}^{(i)}$ *is induced by the projection which forgets the first component* $B \times B^{s-1} \times A^{a+1} \longrightarrow B^{s-1} \times A^{a+1}$*.*

(e) For all $1 \leq i \leq k$ *the family* $\mathcal{S}^{(i)} \longrightarrow \mathcal{T}^{(i)}$ *is a family of blow-ups of the surface* S *(see Definition 7.1), which implies in particular that it is obtained from* $S \times \mathcal{T}^{(i)}$ *by a succession of* i *blow ups of one point in each fibre. If we consider the sequence of blow-ups of* S *corresponding to a point* $t \in \mathcal{T}^{(k)}$*, then the images in* S *of the blown-up points are precisely the components of the image of* t *under the blow-up* $\mathcal{T}^{(k)} = \mathcal{B}^{(k-1)} \longrightarrow B^{k-1-a} \times A^{a+1}$ *(if* $a = k - 1$ *one has no blow up, namely* $\mathcal{T}^{(k)} = \mathcal{A}^{(k-1)} \cong A^k$*)*

Proof: Assume $0 \leq i \leq a+1$. Since we use $\mathbb{B}_{\mathcal{A}^*}$ to construct $\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(a+1)}$ the statements (a) and (b) follow by induction from Lemma 7.4, where we use always (for different T') the natural isomorphism $(A \times_T T') \times_{T'} (A \times_T T') \cong A \times_T (A \times_T T')$ which forgets the first T'.

The statement of (c) is clear for $i = 0, 1$ from the same lemma, which also implies, that $\mathcal{B}^{(i)} \longrightarrow \mathcal{B}^{(i-1)} \times_{\mathcal{A}^{(i-2)}} \mathcal{A}^{(i-1)}$ is the blow-up at $\mathcal{P}^{(i-1)} \times_{\mathcal{A}^{(i-2)}} \mathcal{P}^{(i-1)}$. Using (a) and (b) this translates by induction to the statement that $\mathcal{B}^{(i)} \longrightarrow (B \times A^{i-1}) \times_{A^{i-1}} A^i \cong B \times A^i$ is the succession of the blow-ups of (the strict transforms of) $P \times H_i, P \times H_{i-1}, \dots, P \times H_2$ followed by the blow-up of the strict transforms of $(P \times A^{i-1}) \times_{A^{i-1}} (P \times A^{i-1}) \cong P \times P \times A^{i-1} = P \times H_1$.
We assume now $s \geq 2$ and $i = a + s \leq k - 1$. To prove (d) we first observe (cf. (a)) that the Lemma 7.4 implies that $\mathcal{B}^{(i)}$ is isomorphic to the s-fold fibre product $\mathcal{B}^{(a+1)} \times_{\mathcal{T}^{(a+1)}} \mathcal{B}^{(a+1)} \times_{\mathcal{T}^{(a+1)}} \dots \times_{\mathcal{T}^{(a+1)}} \mathcal{B}^{(a+1)}$ and the projection to $\mathcal{B}^{(i-1)}$ is by forgetting the first factor. But we know from (a) and (c) that $\mathcal{T}^{(a+1)} = A^{a+1}$ and $\mathcal{B}^{(a+1)} \longrightarrow B \times A^{a+1}$ is the blow-up of $P \times H_{a+1}, P \times H_a, \dots, P \times H_1$ and the projection to A^{a+1} gives the map $\mathcal{B}^{(a+1)} \longrightarrow \mathcal{T}^{(a+1)}$.
Induction on s implies now easily the claim of (d). The statement (e) is clear by induction since $\mathcal{S}^{(0)} = S$ and $\mathcal{S}^{(i)} \longrightarrow \mathcal{S}^{(i-1)} \times_{\mathcal{T}^{(i-1)}} \mathcal{T}^{(i)}$ is the blow-up of the section of the projection to $\mathcal{T}^{(i)}$ given by the inclusion $\mathcal{T}^{(i)} \subset \mathcal{S}^{(i-1)}$. □

Proof: (of Proposition 7.2)
Let $\mathcal{S} := \mathcal{S}^{(k)}$ and $\mathcal{T} := \mathcal{T}^{(k)}$ with the notation of Lemma 7.5. The assumptions imply that for every $0 \leq i \leq k - 1$ the surface $S^{(i)}$ is the fibre of $\mathcal{S}^{(i)} \longrightarrow \mathcal{T}^{(i)}$

over a point $P^{(i-1)} \in \mathcal{T}^{(i)} \subset \mathcal{S}^{(i-1)}$. The point $P^{(i)} \in \mathcal{S}^{(i)}$ is the intersection point of the section $\mathcal{P}^{(i)} \subset \mathcal{S}^{(i)}$ with the fibre $S^{(i)}$.
The special point $0 \in \mathcal{T} = \mathcal{T}^{(k)}$ corresponds to $P^{(k-1)} \in \mathcal{T}^{(k)} \subset \mathcal{S}^{(k-1)}$. Its image under the sequence of blow-ups $\mathcal{T} \longrightarrow B^{k-a-1} \times A^{a+1}$ is the point $P^{k-a-1} \times P^{a+1} = (P, P, \dots, P)$.
Using the partition π we can define an embedding $\delta_\pi : B^{\beta-\alpha} \times A^{\alpha} \longrightarrow B^{k-a-1} \times A^{a+1}$ by the formula $\delta_\pi(x_\beta, x_{\beta-1}, \dots, x_1) := (x_{\pi(k-1)}, x_{\pi(k-2)}, \dots, x_{\pi(0)})$. This is a kind of diagonal. The fibre product of the sequence of blow-ups $\mathcal{T} \longrightarrow B^{k-1-a} \times A^{a+1}$ with δ_π defines a variety $\mathcal{T}_\pi$ together with a morphism $\mathcal{T}_\pi \longrightarrow \mathcal{T}$ and a sequence of blow-ups $\mathcal{T}_\pi \longrightarrow B^{\beta-\alpha} \times A^{\alpha}$. $\mathcal{T}_\pi \subset \mathcal{T}$ is the strict transform of $\delta_\pi(B^{\beta-\alpha} \times A^{\alpha})$ and, therefore, the morphism $\mathcal{T}_\pi \longrightarrow B^{\beta-\alpha} \times A^{\alpha}$ is the composition of the blow-ups of $H'_{\beta-\alpha} \times H_\alpha$ followed by the blow-up of the strict transforms of $H'_{\beta-\alpha} \times H_{\alpha-1}, \dots, H'_{\beta-\alpha} \times H_1, H'_{\beta-\alpha-1} \times H_\alpha, H'_{\beta-\alpha-1} \times H_{\alpha-1}, \dots, H'_{\beta-\alpha-1} \times H_1, \dots, H'_1 \times H_\alpha, \dots, H'_1 \times H_1$. The H'_i, H_j have the same meaning as above but now as sub-varieties in $B^{\beta-\alpha}$ and A^{α} respectively. The assumption that the partition π is monotone ensures that we put the $H'_i \times H_j$ in the right order.
Since A and B are by assumption smooth irreducible curves, $\mathcal{T}_\pi$ is smooth and irreducible. The preimage in $\mathcal{T}_\pi$ of the union of all two-fold diagonals in $B^{\beta-\alpha} \times A^{\alpha}$ and the set of points with at least one component equal to P is a Zariski-closed subset of $\mathcal{T}_\pi$ containing $P^{(k-1)}$ and has codimension one in $\mathcal{T}_\pi$. On its complement the blow-up $\mathcal{T}_\pi \longrightarrow B^{\beta-\alpha} \times A^{\alpha}$ is an isomorphism (by Lemma 7.5). Hence, each (analytic) neighbourhood of $P^{(k-1)} \in \mathcal{T}_\pi$ contains a point t, whose image in $B^{\beta-\alpha} \times A^{\alpha}$ is a point whose components are distinct from each other and from P. Hence, the fibre of $\mathcal{S} \longrightarrow \mathcal{T}$ over the image of t in $\mathcal{T}$ is a sequence of blow-ups $S_t^{(i+1)} \longrightarrow S_t^{(i)}$ of points $Q^{(i)} \in S_t^{(i)}$ with the required properties.
The family $\mathcal{S}_\pi \longrightarrow \mathcal{T}_\pi$ obtained by base change via $\mathcal{T}_\pi \longrightarrow \mathcal{T}$ from $\mathcal{S} \longrightarrow \mathcal{T}$ is the family with the required properties. The particular situation with k distinct points in S corresponds to the partition $\pi : \{0, 1, \dots, k-1\} \longrightarrow \{1, 2, \dots, k\}$ given by $\pi(i) = i + 1$. In this case we have $\mathcal{T}_\pi = \mathcal{T}$. $\square$
Because our main interest is the study of sequences of blow-ups of conjugate pairs, we need an additional result to make Proposition 7.2 applicable. On the other hand, we want to patch together deformations of the kind described in Proposition 7.2 centred around different points $P \in S$. For both purposes, we can apply the following lemma.

Lemma 7.6. *Let S be a smooth surface and $\mathcal{S}' \longrightarrow S \times \mathcal{T}'$ and $\mathcal{S}'' \longrightarrow S \times \mathcal{T}''$ be two flat families of sequences of blow-ups of S. Hence, we are given sections $\varphi'_i : \mathcal{T}' \longrightarrow \mathcal{S}'_i$ $(0 \le i \le k'-1)$ of $\mathcal{T}'$-flat families $\mathcal{S}'_i \longrightarrow \mathcal{T}'$ being the blow-up of $\mathcal{S}'_{i-1} \longrightarrow \mathcal{T}'$ along $\varphi'_{i-1}(\mathcal{T}') \subset \mathcal{S}'_{i-1}$ and $\mathcal{S}'_0 = S \times \mathcal{T}'$, $\mathcal{S}' = \mathcal{S}'_{k'}$. By $\psi'_i : \mathcal{T}' \longrightarrow S$ we denote the composition of φ'_i with the projection to S. Similarly for $\mathcal{S}'' \longrightarrow \mathcal{T}''$. Assume $\bigcup_{i=0}^{k'-1} \psi'_i(\mathcal{T}') \cap \bigcup_{i=0}^{k''-1} \psi''_i(\mathcal{T}'') = \emptyset$.*
Then $\mathcal{S} := \mathcal{S}' \times_S \mathcal{S}'' \longrightarrow \mathcal{T} := \mathcal{T}' \times \mathcal{T}''$ is a flat family of blow-ups of S. The fibre over $(t', t'') \in \mathcal{T}$ is isomorphic to the blow-up of S corresponding to $t' \in \mathcal{T}'$

followed by the sequence of blow-ups corresponding to $t'' \in \mathcal{T}''$.

Proof: By the disjointness assumption we can lift φ_i'' to a section $\widetilde{\varphi}_i''$ of $\mathcal{S}_i \longrightarrow \mathcal{T}' \times \mathcal{T}''$ where $\mathcal{S}_0 := \mathcal{S}' \times \mathcal{T}'' = \mathcal{S}' \times_S (S \times \mathcal{T}'')$ and $\mathcal{S}_i \longrightarrow \mathcal{S}_{i-1}$ is the blow-up of $\widetilde{\varphi}_{i-1}''(\mathcal{T}' \times \mathcal{T}'')$, that is $\mathcal{S}_i \cong \mathcal{S}' \times_S \mathcal{S}_i''$. This gives the lemma. $\square$

Remark 7.7. We shall apply this lemma in the following situation to obtain a real structure on the family $\mathcal{S} \longrightarrow \mathcal{T}$. We assume S is a surface with a real structure (without real points) and $\mathcal{S}''$ is the conjugate family to $\mathcal{S}'$, this means $\mathcal{T}'' = \overline{\mathcal{T}'}$, $\mathcal{S}'' = \overline{\mathcal{S}'}$ and φ_i'' is the conjugate section to φ_i'. The projection $\mathcal{S}'' \longrightarrow S$ is the composition of the corresponding projection $\overline{\mathcal{S}''} \longrightarrow \overline{S}$ with the isomorphism $S \longrightarrow \overline{S}$ defining the real structure. The real structures on $\mathcal{S}$ and $\mathcal{T}$ are given by interchanging the components. This is anti-holomorphic since the identity $\mathcal{S}' \longrightarrow \overline{\mathcal{S}'}$ is.

We start now the study of small deformations with the aid of the blow-up graphs of Section 6.

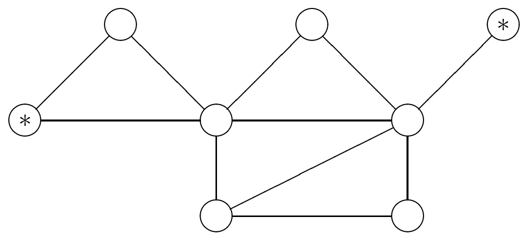
Definition 7.8. We say that a blow-up graph Γ is a *small deformation* of an other blow-up graph Γ_0 if and only if there exists a flat family of surfaces $\mathcal{S} \longrightarrow \mathcal{T}$ with real structures having special fibre S_0 over the real point $0 \in \mathcal{T}$ such that S_0 is isomorphic to a blow-up of $\mathbb{P}^1 \times \mathbb{P}^1$ with associated graph Γ_0 and every (analytic) neighbourhood of $0 \in \mathcal{T}$ contains a real point $t \in \mathcal{T}(\mathbb{R}) \setminus \{0\}$ whose fibre S_t is isomorphic to a blow-up of $\mathbb{P}^1 \times \mathbb{P}^1$ with associated graph Γ.

In the following we want to determine blow-up graphs which are small deformations of a given blow-up graph. We shall not solve the problem of determining all graphs being a small deformation of a given one, because this includes the study of different graphs belonging to isomorphic surfaces. The results obtained here are sufficient for our applications.

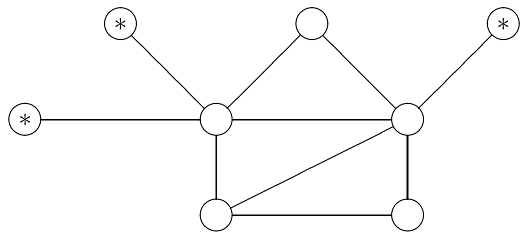
Definition 7.9. Let Γ_0, Γ be blow-up graphs with the same set of vertices. We say Γ is an *elementary deformation* of Γ_0 if we obtain Γ by removing one edge from Γ_0 which connects two internal vertices or two external vertices in Γ_0. We require that one of these vertices, call it v, is marked in Γ but not marked in Γ_0. All other markings of Γ and Γ_0 coincide.

Remark 7.10. Since the number of adjacent edges of the vertex v in Γ is one less than in the graph Γ_0 but it is marked in Γ, the number $n(v)$ must be the same for both graphs. If the removed edge connects two external vertices, the chain of external vertices in Γ_0 containing this edge splits into two chains in Γ. One of these two parts does already contain a marked point. Therefore, the vertex to be marked is in this case already determined by Γ_0. If we remove an edge connecting two internal vertices, the vertex v must fulfill $n(v) = 2$ or $n(v) = 3$, since in Γ it is an external marked vertex. In general, the vertex v which becomes marked in Γ is not determined by the graph Γ_0.

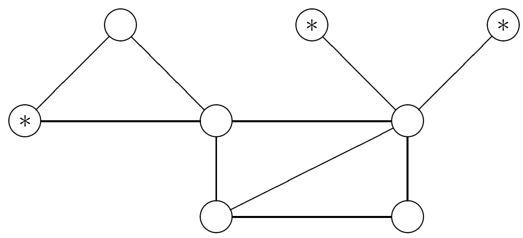
Example 7.11. If Γ_0 is the following graph:



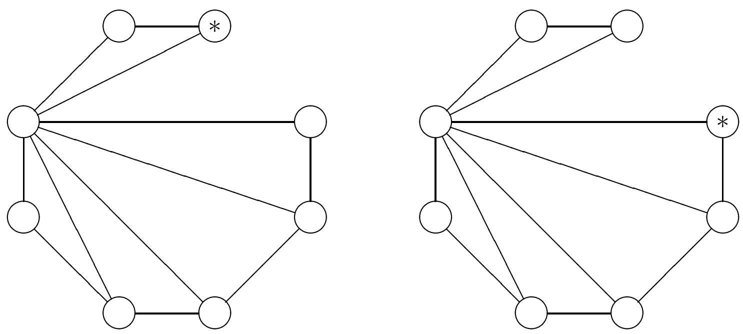
we obtain as an elementary deformation by removing a connection of external vertices the following graph:



and by removing an edge connecting two internal vertices we obtain the elementary deformation:



Example 7.12. The following two graphs are elementary deformations of the graph Γ_0 drawn in Example 6.4. Here we see that we have two possibilities for the additional marking in the graph Γ. In the first example we have $m = 6$ and in the second $m = 3$, whereas for the graph Γ_0 we have $m = 8$.



Theorem 7.13. *Let Γ_0, Γ be two blow-up graphs such that Γ can be obtained from Γ_0 by a finite number of elementary deformations. Then Γ is a small deformation of Γ_0.*

Proof: Assume Γ_0 and Γ have the same set of vertices. If $\Gamma_0 \neq \Gamma$ there is a certain set $\mathcal{V}$ of vertices which are marked in Γ and not marked in Γ_0. Every vertex $v \in \mathcal{V}$ is contained in precisely one (maximal) chain of external vertices $\mathcal{C}(v)$ in the graph Γ. By $\mathcal{C}$ we denote the union of these sets of external vertices $\mathcal{C} = \bigcup_{v \in \mathcal{V}} \mathcal{C}(v)$.

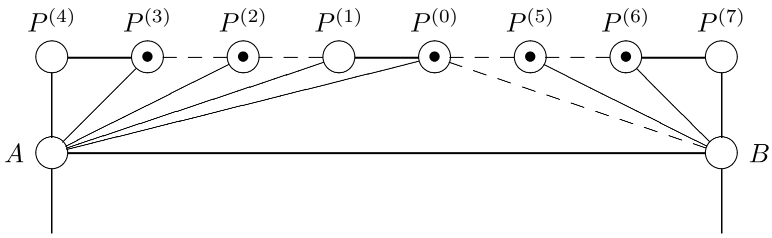
Let Γ' be the graph obtained from Γ by removing all the vertices in $\mathcal{C}$ and the edges connecting them with internal vertices of Γ. This means, Γ can be obtained from Γ' by adding marked vertices and external triangles. As seen in the proof of Proposition 6.7 there exist sequences of blow-ups $S_1 \xrightarrow{\sigma} S_1' \xrightarrow{\sigma'} S = \mathbb{P}^1 \times \mathbb{P}^1$ such that Γ (resp. Γ') is the graph associated to the sequence of blow ups $\sigma' \circ \sigma$ (resp. σ'). Furthermore, it is clear from the definitions, that we obtain Γ_0 from Γ' by adding internal triangles, external triangles and marked vertices. This implies the existence of a sequence of blow-ups $\sigma_0 : S_0 \longrightarrow S_1'$ such that Γ_0 is the graph associated to the composition $\sigma' \circ \sigma_0$.

The graph obtained from Γ_0 by removing all the internal vertices of Γ consists of certain connected components. We denote by $\mathcal{C}_i$ with $1 \leq i \leq c$ the subsets of $\mathcal{C}$ obtained by intersection with these connected components.

Every set $\mathcal{C}_i$ consists entirely of internal or of external vertices of Γ_0, because in a blow-up graph two marked vertices are not contained in the same chain of external vertices. If $\mathcal{C}_i$ contains an internal vertex of Γ_0, then it contains exactly one vertex connected with two internal vertices of Γ. All other vertices of $\mathcal{C}_i$ are connected with precisely one of these internal vertices of Γ. If $\mathcal{C}_i$ consists of external vertices of Γ_0, then all its vertices are connected with the same internal vertex in Γ.

From the relation between blow-ups and the operation of adding a marked vertex or an external triangle to a graph (described in the proof of Proposition 6.7) it is clear that the sets $\mathcal{C}_i$ are precisely the equivalence classes on $\mathcal{C}$ given by the equivalence relation: $w \sim w'$ if and only if the conjugate pairs of curves corresponding to w and w' are mapped under $\sigma_0 : S_0 \longrightarrow S_1'$ to the same conjugate pair of points. These points lie on the curves in S_1', corresponding to the internal vertices of the graph Γ' connected with $\mathcal{C}_i$.

By Lemma 7.6 it is enough to prove the theorem in the case of only one set $\mathcal{C}_i$. But in this case the result is a reformulation of Proposition 7.2 (using Lemma 7.6 to obtain a version of Proposition 7.2 with pairs of blown-up points at each step, see Remark 7.7). The partition of the set $\mathcal{C}_i$ is defined by the chains of external vertices of Γ inside $\mathcal{C}_i$. The following picture gives an example of a part of a graph Γ_0 where the edges which are not edges of Γ are drawn with broken lines. The vertices of the set $\mathcal{V}$ are indicated with bullets.



In this example we have five sets in the corresponding partition (i.e. $\beta = 5$), namely $\{P^{(0)}, P^{(1)}\}, \{P^{(2)}\}, \{P^{(3)}, P^{(4)}\}, \{P^{(5)}\}, \{P^{(6)}, P^{(7)}\}$. □

Corollary 7.14. *Every blow-up graph Γ is a small deformation of a blow-up graph Γ_0 with the same number of vertices, but having no external vertices.*

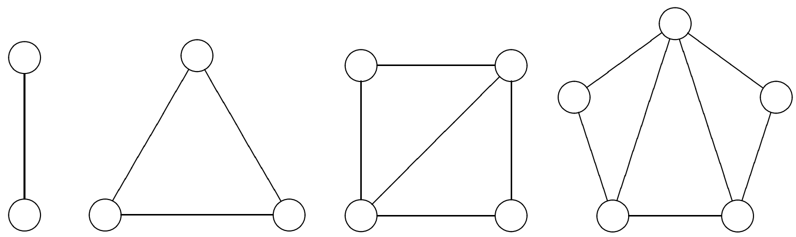
Proof: This follows easily by induction from the observation that we obtain a blow-up graph Γ_0 by the following procedure: In a blow-up graph Γ we unmark a marked external vertex v. Let w be the unique internal vertex connected with v. Then, we connect v by an edge with one of the internal vertices which are neighbours of w along the boundary of the m-gon of internal vertices of the given graph Γ. □

This motivates the following definition.

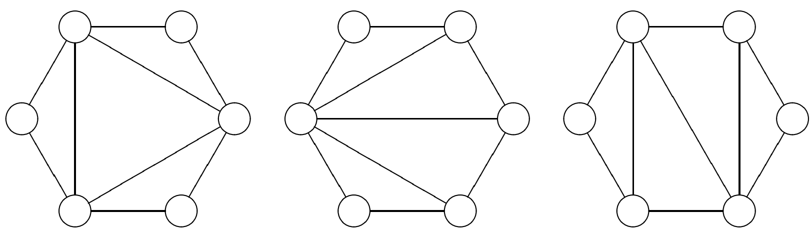
Definition 7.15. A *basic* blow-up graph is a blow-up graph which does not contain external vertices.

Remark 7.16. Let us equip the set of blow-up graphs with the partial ordering generated by the requirement: $\Gamma \geq \Gamma_0$ if Γ is an elementary deformation of Γ_0. Then the basic blow-up graphs are precisely the minimal elements in this PO-set.

Remark 7.17. For every $2 \leq m \leq 5$ there exists precisely one basic blow-up graph with m vertices. They are the following:



All other blow-up graphs with $m \leq 5$ are small deformations of them. If $m = 6$ there exist three different basic blow-up graphs:



8 Application to twistor spaces

We return to the situation of Section 3. Let Z be a compact, simply connected twistor space containing an irreducible fundamental divisor and satisfying $c_1(Z)^3 < 0$ and $h^0(K^{-\frac{1}{2}}) \geq 2$.
By Proposition 2.4 we know that the Riemannian base of such a twistor space is diffeomorphic to the connected sum $n\mathbb{CP}^2$ (with $n > 4$) and the conformal class contains a metric with positive scalar curvature.
The existence of such a pencil implies that the algebraic dimension of Z must be positive. Let S be an irreducible real fundamental divisor, then there exists a sequence of blow-ups of $n \geq 5$ conjugate pairs of points $S \longrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. We know from [K1, Prop. 3.6] that we can choose this succession of blow-ups such that the anti-canonical system $|-K_S|$ contains a real member C mapped onto a curve C' on $\mathbb{P}^1 \times \mathbb{P}^1$ having one of the following four types :

(0) $C' \in |\mathcal{O}(2,2)|$ is a smooth elliptic curve,

(1) C' has four components $C' = F' + \overline{F'} + G' + \overline{G'}$ where $F' \in |\mathcal{O}(0,1)|$ and $G' \in |\mathcal{O}(1,0)|$ are not real,

(2) C' has two components $C' = F' + C'_0$ where $F' \in |\mathcal{O}(0,1)|$ is real and $C'_0 \in |\mathcal{O}(2,1)|$ is real, smooth and rational,

(3) C' has two distinct components $C' = A' + \overline{A'}$ where $A', \overline{A'} \in |\mathcal{O}(1,1)|$.

In the case of type (0) the curve C is smooth elliptic and $(C^2)_S < 0$, hence, by Theorem 5.2 we have $\kappa^{-1}(S) = 0$. Corollary 5.4 implies that we have for generic real fundamental divisors $\kappa^{-1}(S) = 0$. Hence, by Corollary 4.3 we obtain $a(Z) = 1$.
In the type (2) case we always have $\kappa^{-1}(S) = 2$, because there is no point on F' blown up and hence the strict transform F of F' is a curve with $(F.K_S^{-1})_S = 2$. Again, by Corollaries 5.4 and 4.3 we obtain $a(Z) = 3$. This was also obtained in [K2].
The case of type (3) reduces to type (1) using elementary transformations, if the intersection points of A' and $\overline{A'}$ are blown up. Otherwise, we obtain $(A.(-K_S))_S < 0$ and $C = A + \overline{A}$. In this situation, Theorem 5.2 tells us $\kappa^{-1}(S) = 0$ and again we compute $a(Z) = 1$ using the Corollaries 5.4 and 4.3.
It remains to study the situation of type (1). This is precisely the situation where we can associate to the sequence of blow-ups a blow-up graph Γ. If Γ does not contain one of the ten graphs of Remark 6.11 as a subgraph that contains all external vertices and all edges between them, then Γ determines the anti Kodaira dimension of S by Theorem 6.9. If this is the case, the algebraic dimension $a(Z)$ is determined by Γ. This follows from Corollaries 5.4 and 4.3, because the restriction $K_S^{-1} \otimes \mathcal{O}_C \cong K^{-\frac{1}{2}} \otimes \mathcal{O}_C$ does not depend on the chosen fundamental divisor S. For example we can formulate the following theorem:

Theorem 8.1. *A simply connected twistor space Z containing at least a pencil of fundamental divisors is Moishezon if and only if it fulfills the equivalent* *conditions of Theorem 3.7 or contains a real irreducible fundamental divisor S possessing an associated blow-up graph that either*

- *contains one internal vertex which is connected with all external vertices, or*

- *contains at most four external vertices and a pair of connected internal vertices with the property that one of them is connected with precisely one of the external vertices and the other one with all the remaining external vertices.*

In particular, basic blow-up graphs appear only in Moishezon spaces.

Proof: The observations at the beginning of this section show for $n \geq 5$ that a Moishezon twistor space, not fulfilling the conditions of Theorem 3.7, contains a real fundamental divisor S possessing a blow-up graph. For $n = 4$ this follows from [K1] and in case $n \leq 3$ every twistor space contains a fundamental divisor possessing an associated blow-up graph.
Observe that a blow-up graph with at most five vertices always fulfills the conditions of the theorem. Since in the case $n \leq 3$ all twistor spaces are Moishezon, nothing is to prove then. In the case $n = 4$ (corresponding to blow-up graphs with six vertices) the result follows from previous work [K1] and the observation, that (in this case) $K^{-\frac{1}{2}}$ is not nef if and only if the corresponding blow-up graph fulfills the conditions of the theorem. By a *nef* line bundle we mean here one which has non-negative intersection number with all curves in Z. Let us, therefore, assume $n \geq 5$.
In the theorem the conditions on the graph are made to match precisely the graphs obtained by adding internal triangles to a graph fulfilling condition (c) or (c') of Theorem 6.9. If $\dim \mid -\frac{1}{2}K \mid = 1$, we can apply Corollary 5.4 and Theorem 6.9 to show that the generic $S \in \mid -\frac{1}{2}K \mid$ has $\kappa^{-1}(S) = 2$. Corollary 4.3 implies that Z is a Moishezon space. If $\dim \mid -\frac{1}{2}K \mid \geq 2$, then the result follows from Theorems 3.6, 3.7 and the observations at the beginning of this section. $\square$

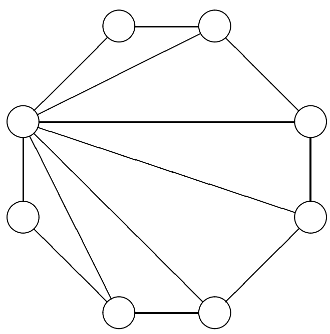
Remark 8.2. A blow-up graph fulfills the properties of the theorem precisely when it contains one of the graphs of Remark 6.12 as a subgraph that contains all external vertices of it.

Remark 8.3. In Corollary 7.14 we saw that every blow-up graph is a small deformation of a basic blow-up graph. By Theorem 8.1 basic blow-up graphs appear only in Moishezon twistor spaces. This suggests that one could hope to be able to construct twistor spaces containing a fundamental divisor associated to an arbitrarily given blow-up graph by studying small deformations of Moishezon twistor spaces. We shall see in Theorem 8.8 that this in fact works. In particular, small deformations of Moishezon twistor spaces need not to be Moishezon [C1], [LeBP].

It would be very interesting to obtain a better understanding of $a(Z)$ and $\kappa^{-1}(S)$ if Γ does not determine the anti Kodaira dimension. This would help to understand the case $a(Z) = 2$.

Definition 8.4. A blow-up graph Γ is called *twistorial* if there exists a twistor space Z containing an irreducible fundamental divisor S which is obtained from $\mathbb{P}^1 \times \mathbb{P}^1$ by a sequence of blow-ups whose associated graph is Γ.

Example 8.5. The basic blow-up graphs of Example 6.4 containing one vertex which is connected with all other vertices are twistorial.
By Proposition 6.7, Theorems 3.6 and 3.5 one should search for a corresponding fundamental divisor in a LeBrun twistor space. Such twistor spaces Z are birational to conic bundles over $\mathbb{P}^1 \times \mathbb{P}^1$ whose discriminant is the union of n irreducible divisors in the linear system $\mid \mathcal{O}(1,1) \mid$. The fundamental linear system is isomorphic to $\mid \mathcal{O}(1,1) \mid$ such that every divisor in $\mid \mathcal{O}(1,1) \mid$ corresponds to a fundamental divisor in Z. The most degenerate LeBrun spaces are those where the n components of the discriminant of the conic bundle are contained in one pencil in $\mid \mathcal{O}(1,1) \mid$. Such a pencil has two base-points on $\mathbb{P}^1 \times \mathbb{P}^1$. Every real member of this pencil, which is different from the n components of the discriminant of the conic bundle, has as its associated graph the basic blow-up graph mentioned above. (Below, the picture for the case $n = 6$ is drawn.) This degenerate case was not studied in [LeB2]. The details can be found in [Ku].



In the sequel we want to study the question which blow-up graphs are twistorial. For that purpose we have to show the existence of twistor spaces with certain properties. A very efficient tool for constructing new twistor spaces is the following theorem, which has its origin in the paper [DonF].

Theorem 8.6. *([DonF],[LeB3],[PP2],[C3]) Let Z be a Moishezon twistor space with $H^2(Z, \Theta_Z) = 0$. Then, any real member of a small deformation of Z is again a twistor space. Furthermore, any small deformation of a real irreducible fundamental divisor S with real structure is induced by a deformation of Z in the sense that the deformed surfaces are members of the fundamental system of the deformed twistor spaces.*

For LeBrun twistor spaces the vanishing of $H^2(Z, \Theta_Z)$ was shown in the papers [LeBP], [C1] and [C3]. But the authors of these papers do not take care of the degenerate case. Therefore we need the following theorem, whose proof grew out of a discussion with H. Kurke. The author is grateful to him.

Theorem 8.7. *If Z is a Moishezon twistor space containing an irreducible fundamental divisor, then $H^2(Z, \Theta_Z) = 0$.*

Proof: The space Z is simply connected [C2] and of positive type by Proposition 2.4 or [Po2]. Let $S \in \mid -\frac{1}{2}K \mid$ be an irreducible fundamental divisor. By Lemma 2.2 S is a smooth rational surface. The adjunction formula implies $K^{\frac{1}{2}} \otimes \mathcal{O}_S \cong K_S$. The exact sequence

$$0 \longrightarrow N^{\vee}_{S|Z} \longrightarrow \Omega^1_Z \otimes \mathcal{O}_S \longrightarrow \Omega^1_S \longrightarrow 0$$

implies $H^0(\Omega^1_Z \otimes \mathcal{O}_S) = 0$, because $H^0(N^{\vee}_{S|Z}) = H^0(\mathcal{O}_S(-S)) = H^0(K_S) = 0$ and $H^0(\Omega^1_S) = 0$ by the rationality of S.
On the other hand, the restriction map $\operatorname{Pic} Z \longrightarrow \operatorname{Pic} S$ is injective by [K1, Lemma 3.1]. The Fröhlicher spectral sequence (which degenerates for Moishezon varieties [U]) together with the rationality of S, the vanishing of $H^0(Z, \Omega^2_Z)$ ([H2]) and Lemma 2.3 induces natural isomorphisms $\operatorname{Pic} Z \cong H^1(Z, \Omega^1_Z)$ and $\operatorname{Pic} S \cong H^1(S, \Omega^1_S)$. The corresponding natural injective map $H^1(Z, \Omega^1_Z) \longrightarrow H^1(S, \Omega^1_S)$ is the composition of the natural maps $H^1(\Omega^1_Z) \longrightarrow H^1(\Omega^1_Z \otimes \mathcal{O}_S) \longrightarrow H^1(\Omega^1_S)$. The first morphism, which is hence injective, appears in the exact cohomology sequence of

$$0 \longrightarrow \Omega^1_Z(-S) \longrightarrow \Omega^1_Z \longrightarrow \Omega^1_Z \otimes \mathcal{O}_S \longrightarrow 0.$$

With the vanishing of $H^0(\Omega^1_Z \otimes \mathcal{O}_S)$, shown above, we obtain now: $H^1(\Omega^1_Z \otimes K^{\frac{1}{2}}) = H^1(\Omega^1_Z(-S)) = 0$.
Using the standard exact sequence $0 \longrightarrow N^{\vee}_{S|Z} \longrightarrow \Omega^1_Z \otimes \mathcal{O}_S \longrightarrow \Omega^1_S \longrightarrow 0$ we obtain, using $N^{\vee}_{S|Z} = \mathcal{O}_S(-S) = K^{\frac{1}{2}} \otimes \mathcal{O}_S$, the exact sequence

$$0 \longrightarrow K_S^{\otimes 2} \longrightarrow \Omega^1_Z \otimes K^{\frac{1}{2}} \otimes \mathcal{O}_S \longrightarrow \Omega^1_S \otimes K_S \longrightarrow 0.$$

Since S is a rational surface we have $h^0(K_S^{\otimes 2}) = 0$ and $h^0(\Omega^1_S \otimes K_S) = h^2(\Theta_S) = 0$. Hence, we obtain $h^0(\Omega^1_Z \otimes K^{\frac{1}{2}} \otimes \mathcal{O}_S) = 0$. Using the exact sequence

$$0 \longrightarrow \Omega^1_Z \otimes K \longrightarrow \Omega^1_Z \otimes K^{\frac{1}{2}} \longrightarrow \Omega^1_Z \otimes K^{\frac{1}{2}} \otimes \mathcal{O}_S \longrightarrow 0,$$

and the vanishing of $h^1(\Omega^1_Z \otimes K^{\frac{1}{2}})$, this implies $h^1(\Omega^1_Z \otimes K) = 0$. By Serre duality we obtain the desired vanishing. □
The following theorem is the main result of this section.

Theorem 8.8. *Every blow-up graph is twistorial.*

Proof: Combining Theorems 8.7, 8.6 with Theorem 7.13 and Corollary 7.14 we see that it is enough to show that every basic blow-up graph is twistorial.

The corresponding twistor spaces are provided by the equivariant version of the method of Donaldson and Friedman [DonF] to construct self-dual structures on the connected sum of two self-dual manifolds. Such a method was developed by Pedersen and Poon in the paper [PP3]. The spaces obtained in the case of the action of a two dimensional torus are investigated in detail in a recent preprint of Honda [Ho].
His main result is that the twistor spaces obtained by the equivariant version of the Donaldson-Friedman construction contain a pencil of fundamental divisors invariant under the action of the two-dimensional torus. The general member of this pencil is a smooth toric surface, which is isomorphic to a successive blow-up of $\mathbb{P}^1 \times \mathbb{P}^1$ at conjugate pairs of fixed points of the action.
Furthermore, he shows, using the results of Orlik and Raymond [OR], that every such toric surface appears as a fundamental divisor in a twistor space. But the fixed points of the torus action are precisely the singularities of the (unique) torus invariant effective anti-canonical divisor on the toric surface. This means that an arbitrary sequence of blow ups of conjugate singularities of the torus invariant effective anti-canonical divisor, starting at $\mathbb{P}^1 \times \mathbb{P}^1$, leads to a surface which appears as a fundamental divisor in a twistor space. Because every basic blow-up graph can be obtained in this way, the theorem is proven. $\square$
There is another construction of twistor spaces over $n\mathbb{CP}^2$ with the symmetry of the two-torus, introduced by D. Joyce [J]. It seems to be not clear, whether these spaces contain a pencil of fundamental divisors or not. But, observe that D. Joyce associates (in a different way) to each of his spaces one of the basic blow-up graphs [J, p. 541]. These graphs reflect the orbit structure and isotropy groups of the action of $T^2 = S^1 \times S^1$ on $n\mathbb{CP}^2$.

References

[AHV] J. Aroca, H. Hironaka, J. Vicente: Desingularisation theorems, Mem. Math. Inst. Jorge Juan No. 30, Madrid, 1977

[AHS] M.F. Atiyah, N.J. Hitchin, I.M. Singer: Self–duality in four–dimensional Riemannian geometry, *Proc. Roy. Soc. London Ser. A* 362 (1978) 425-461

[BPV] W. Barth, C. Peters, Van de Ven: Compact complex surfaces, Springer, Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Band 4, Berlin 1984

[C1] F. Campana: The class $\mathcal{C}$ is not stable by small deformations, *Math. Ann.* 290 (1991) 19-30

[C2] F. Campana: On Twistor Spaces of the Class $\mathcal{C}$, *J. Diff. Geom.* 33 (1991) 541-549

[C3] F. Campana: The class $\mathcal{C}$ is not stable by small deformations II, *Contemp. Math.* 162 (1994) 65-76

[CK1] F. Campana, B. Kreussler: Existence of twistor spaces of algebraic dimension two over the connected sum of four complex projective planes, to appear in *Proc. of the AMS*

[CK2] F. Campana, B. Kreussler: A Conic Bundle Description of Moishezon Twistor Spaces Without Effective Divisors of Degree One, *Math. Zeit.* 229 (1998) 137-162

[DonF] S. Donaldson, R. Friedman: Connected sums of self-dual manifolds and deformations of singular spaces, *Nonlinearity* 2 (1989) 197-239

[ES] M.G. Eastwood, M.A. Singer: The Fröhlicher Spectral Sequence on a Twistor Space, *J. Diff. Geom.* 38 (1993) 653-669

[FK] T. Friedrich, H. Kurke: Compact four–dimensional self–dual Einstein manifolds with positive scalar curvature, *Math. Nach.* 106 (1982) 271-299

[Gau] P. Gauduchon: Structures de Weyl et théorèmes d'annulation sur une variété conforme autoduale, *Ann. Scuola Norm. Sup. Pisa* 18 (1991) 563-629

[GH] P. Griffiths, J. Harris: Principles of algebraic geometry, New York: John Wiley, 1978

[Hb] B. Harbourne: Anticanonical Rational Surfaces, *Trans. AMS* 349 (1997) 1191-1208

[H] R. Hartshorne: Algebraic geometry, GTM 52, Springer 1977

[HH] F. Hirzebruch, H. Hopf: Felder von Flächenelementen in 4-dimensionalen Mannigfaltigkeiten, *Math. Ann.* 136 (1958) 156-172

[H1] N.J. Hitchin: Linear field equations on self–dual spaces, *Proc. Roy. Soc. London Ser. A* 370 (1980) 173-191

[H2] N.J. Hitchin: Kählerian twistor spaces, *Proc. Lond. Math. Soc., III Ser.* 43 (1981) 133-150

[Ho] N. Honda: On the structure of Pedersen-Poon twistor spaces, preprint, Hiroshima, 1998

[J] D. Joyce: Explicit construction of self-dual 4-manifolds, *Duke Math. Journal* 77 (1995) 519-552

[K1] B. Kreussler: On the algebraic dimension of twistor spaces over the connected sum of four complex projective planes, *Geometriae Dedicata* 71 (1998) 263-285

[K2] B. Kreussler: Moishezon twistor spaces without effective divisors of degree one, *J. Algebraic Geometry* 6 (1997) 379-390

[KK] B. Kreussler, H. Kurke: Twistor spaces over the connected sum of 3 projective planes, *Compositio Math.* 82 (1992) 25-55

[Ku] H. Kurke: Classification of twistor spaces with a pencil of surfaces of degree 1, Part I, *Math. Nachr.* 158 (1992) 67-85

[LeB1] C. LeBrun: On the topology of self-dual 4-manifolds, *Proc. AMS* 98 (1986) 637-640

[LeB2] C. LeBrun: Explicit self–dual metrics on $\mathbb{CP}^2\#\ldots\#\mathbb{CP}^2$, *J. Diff. Geom.* 34 (1991) 223-253

[LeB3] C. LeBrun: Twistors, Kähler manifolds, and bimeromorphic geometry. I, *J. AMS* 5 (1992) 289-316

[LeBP] C. LeBrun, Y.S. Poon: Twistors, Kähler manifolds, and bimeromorphic geometry. II, *J. AMS* 5 (1992) 317-325

[Lo] E. Looijenga: Rational surfaces with an anti-canonical cycle, *Ann. Math.* 114 (1981) 267-322

[OR] P. Orlik, F. Raymond: Action of the torus on 4-manifolds I, *Trans. AMS* 152 (1970) 531-559

[PP1] H. Pedersen, Y.S. Poon: Self–duality and differentiable structures on the connected sum of complex projective planes, *Proc. AMS* 121 (1994) 859-864

[PP2] H. Pedersen, Y.S. Poon: A relative deformation of Moishezon twistor spaces, *J. Alg. Geom.* 3 (1994) 685-701

[PP3] H. Pedersen, Y.S. Poon: Equivariant connected sums of compact self-dual manifolds, *Math. Ann.* 301 (1995) 717-749

[Pe] R. Penrose: Nonlinear gravitons and curved twistor theory, *General Relativity and Gravitation* 7 (1976) 31-52

[Pon] M. Pontecorvo: Algebraic dimension of twistor spaces and scalar curvature of anti–self–dual metrics, *Math. Ann.* 291 (1991) 113-122

[Po1] Y.S. Poon: Compact self–dual manifolds with positive scalar curvature, *J. Diff. Geom.* 24 (1986) 97-132

[Po2] Y.S. Poon: Algebraic dimension of twistor spaces, *Math. Ann.* 282 (1988) 621-627

[Po3] Y.S. Poon: Twistor spaces with Meromorphic Functions, *Proc. AMS* 111 (1991) 331-338

[Po4] Y.S. Poon: On the algebraic structure of twistor spaces, *J. Diff. Geom.* 36 (1992) 451-491

[Sa] F. Sakai: *D*-Dimension of Algebraic Surfaces and Numerically Effective Divisors, *Compositio Math.* 48 (1983) 101-118

[U] K. Ueno: Classification Theory of Algebraic Varieties and Compact Complex Spaces, *Lecture Notes in Math.* 439, Springer, Berlin 1975

[Z] O. Zariski: The Theorem of Riemann-Roch for High Multiples of an Effective Divisor on an Algebraic Surface, *Ann. Math.* 76 (1962) 560-615

B. Kreußler
FB Mathematik
Universität Kaiserslautern
PF 3049
D–67 653 Kaiserslautern
Germany
kreusler@mathematik.uni-kl.de

Les Classes de Chern Modulo p d'une Représentation Régulière

Bruno Kahn

Received: May 25, 1999

Communicated by Ulf Rehmann

Abstract. Let G be a finite group and ρ a complex linear representation of G. In 1961, Atiyah and Venkov independently defined Chern classes $c_i(\rho)$ with values in the integral or mod p cohomology of G. We consider here the mod p Chern classes of the regular representation r_G of G. Venkov claimed that $c_i(r_G) = 0$ for $i < p^n - p^{n-1}$, where p^n is the highest power of p dividing $|G|$; however his proof is only valid for G elementary abelian. In this note, we show Venkov's assertion is valid for any G. The proof also shows that the $c_i(r_G)$ are p-powers of cohomology classes invariant by $Aut(G)$ as soon as G is a non-abelian p-group.

1991 Mathematics Subject Classification: 20J06, 20C15
Keywords and Phrases: Finite groups, Chern classes, regular representation

Introduction

Soient G un groupe fini et ρ une représentation linéaire complexe de G. La définition des classes de Chern de ρ est généralement attribuée à Atiyah [1, appendice]. Toutefois, ces classes ont également été introduites par B.B. Venkov dans une note contemporaine [16].
Venkov annonce dans cette note un certain nombre de résultats[1]. En particulier, soit r_G la représentation régulière de G. Venkov [16] annonce le théorème suivant:

0.1. Théorème. *Soit p un nombre premier, et soit $\nu = v_p(|G|)$. Alors les classes de Chern $c_i(r_G) \in H^{2i}(G, \mathbb{Z}/p)$ de r_G à coefficients $\mathbb{Z}/p$ sont nulles pour $i < p^\nu$, sauf peut-être si i est de la forme $p^\nu - p^l$ pour un $l < \nu$.*

[1] Certains sont manifestement erronés, comme celui affirmant que les classes de Chern de ρ et son degré déterminent la classe de ρ dans $R(G)$.

Son esquisse de démonstration ne donne malheureusement le résultat annoncé que dans le cas où G est abélien élémentaire. Dans cet article, nous nous proposons de démontrer le théorème annoncé par Venkov. La méthode est inspirée de [7].

Je remercie Don Zagier pour son aide dans la démonstration du lemme 1.3 et le referee pour sa lecture soigneuse du manuscrit.

1. Classes de Chern modulo p d'un fibré vectoriel complexe

Soit E un fibré vectoriel complexe sur un espace topologique X, et soit p un nombre premier. Pour tout $i \geq 0$, on note $c_i(E) \in H^{2i}(X, \mathbb{Z}/p)$ la i-ème classe de Chern modulo p de E.

1.1. Théorème. *Supposons que $c_i(E) = 0$ pour $i \not\equiv 0 \pmod{p-1}$. Si les $c_i(E)$ ne sont pas tous nuls, il existe $\nu > 0$ tel que*

(i) *pour $i < p^\nu - p^{\nu-1}$, $c_i(E) = 0$; $c_{p^\nu - p^{\nu-1}}(E) \neq 0$.*

(ii) *$\mathcal{P}^k c_{p^\nu - p^{\nu-1}}(E) = \binom{p^\nu - p^{\nu-1}-1}{k} c_{p^\nu - p^{\nu-1}+(p-1)k}(E)$ pour $0 < k < p^{\nu-1}$, où $\mathcal{P}^k$ est la k-ième puissance de Steenrod ($\mathcal{P}^k = Sq^{2k}$ si $p = 2$);*

(iii) *pour $p^\nu - p^{\nu-1} \leq i < p^\nu$, $c_i(E) = 0$ si i n'est pas de la forme $p^\nu - p^r$ $(r < \nu)$.*

Démonstration. Elle procède essentiellement comme dans [7, dém. de la prop. 1.1]. Pour simplifier, notons $c_i = c_i(E)$; soit M le plus petit entier tel qu $c_M \neq 0$. D'après [11, th. 2] (voir aussi [8]), on a

$$\mathcal{P}^k c_i = \binom{i-1}{k} c_{i+(p-1)k} + \sum_{0 \leq l < (p-1)k} c_{i+l} P_l(c)$$

où $P_l(c)$ est un polynôme en les c_j isobare de poids $(p-1)k - l$, donc ne faisant intervenir c_j que pour $j \leq (p-1)k$. On a donc, pour $i < M$:

$$\binom{i-1}{k} c_{i+(p-1)k} = 0 \quad \text{pour} \quad 0 < k < \frac{M}{(p-1)}$$

et pour $i = M$:

$$\mathcal{P}^k c_M = \binom{M-1}{k} c_{M+(p-1)k} \quad \text{pour} \quad 0 < k < \frac{M}{(p-1)}.$$

Notons $C_i = c_{(p-1)i}$. Sous l'hypothèse du théorème, on a $M = (p-1)m$ pour un m convenable. Les relations ci-dessus donnent alors:

$$\binom{i(p-1)-1}{k} C_{i+k} = 0 \quad \text{pour} \quad 0 < i < m \quad \text{et} \quad 0 < k < m \tag{1}$$

et

$$\mathcal{P}^k C_m = \binom{m(p-1)-1}{k} C_{m+k} \quad \text{pour} \quad 0 < k < m. \tag{2}$$

Prenant en particulier $i = m - k$ dans (1), on obtient:

$$\binom{(m-k)(p-1)-1}{k} \equiv 0 \pmod{p} \quad \text{pour} \quad 0 < k < m. \tag{3}$$

1.2. Lemme. *Tout entier m vérifiant la condition* (3) *est une puissance de p.*

Démonstration. Écrivons $m = m_0 p^n$, avec $(m_0, p) = 1$. Supposons $m_0 > 1$. Choisissons $k = p^n$ dans (3). En écrivant dans $\mathbb{F}_p[[t]]$

$$(1+t)^{p^n(m_0-1)(p-1)-1} = (1+t^{p^n})^{(m_0-1)(p-1)}(1-t+t^2-\dots)$$

on voit que le coefficient de t^{p^n} dans $(1+t)^{p^r(m_0-1)(p-1)-1}$ est $(m_0-1)(p-1)+(-1)^{p^n}$. Par hypothèse, on a donc:

$$(m_0-1)(p-1)+(-1)^{p^n} \equiv 0 \pmod p$$

ou encore

$$m_0 \equiv 1+(-1)^{p^n} \equiv 0 \pmod p$$

ce qui contredit l'hypothèse. On a donc $m_0 = 1$ et $m = p^n$. □

Dans le théorème 1.1, (i) résulte du lemme 1.2 et (ii) résulte de (i) et de (2). Pour voir (iii), soit $j \le 2p^n - 2$ tel que $C_j \ne 0$. En vertu de (1), on a alors:

$$\binom{(j-k)(p-1)-1}{k} \equiv 0 \pmod p \quad \text{pour} \quad j-p^n < k < p^n. \tag{4}$$

il suffit donc de prouver:

1.3. Lemme. *Tout $j \in [p^n, p^n + p^{n-1} + \dots + 1]$ vérifiant* (4) *est de la forme $p^n + p^{n-1} + \dots + p^r$.*

Démonstration. (Don Zagier) Soit r le plus petit entier tel que $j \le p^n + p^{n-1} + \dots + p^r$. On a donc

$$j = p^n + p^{n-1} + \dots + p^{r+1} + x$$

avec $0 < x \le p^r$.

Supposons $x < p^r$. Choisissons $k = p^{n-1} + p^{n-2} + \dots + p^r$. Alors l'entier $N = (j-k)(p-1)-1$ vérifie l'inégalité

$$(p^n - p^r)(p-1) < N < p^n(p-1)$$

donc a un développement en base p ayant pour chiffres

chiffre	$p-2$	$p-1$	$p-1$	$\dots$	$p-1$	c	$*$	$\dots$	$*$
place	n	$n-1$	$n-2$	$\dots$	$r+1$	r	$r-1$	$\dots$	0

avec $c > 0$, tandis que k a pour développement en base p

chiffre	0	1	1	$\dots$	1	1	0	$\dots$	0
place	n	$n-1$	$n-2$	$\dots$	$r+1$	r	$r-1$	$\dots$	0

On a donc

$$\binom{(j-k)(p-1)-1}{k} \equiv \binom{p-2}{0}\binom{p-1}{1}\binom{p-1}{1}\cdots\binom{p-1}{1}\binom{c}{1}\binom{*}{0}\cdots\binom{*}{0} \not\equiv 0 \pmod p$$

ce qui contredit l'hypothèse. □

2. L'invariant ν_p

NOTATION. Soit E un fibré vérifiant la condition du théorème 1.1. L'entier ν de ce théorème est noté $\nu_p(E)$. Si tous les $c_i(E)$ sont nuls, on note $\nu_p(E)=\infty$.

2.1. PROPOSITION. *Soit $f : Y \to X$ un revêtement fini, et soit E un fibré vectoriel complexe sur Y vérifiant les conditions du théorème 1.1. Soit $\nu = \nu_p(E)$; supposons que $\operatorname{rg} E$ soit divisible par p^ν. Alors*

(i) *f_*E vérifie les conditions du théorème 1.1.*
(ii) *On a $\nu_p(f_*E) \geq \nu_p(E)$.*
(iii) *Soit $r = p^\nu - p^{\nu-1}$. Alors*

$$c_r(f_*E) = f_*c_r(E).$$

DÉMONSTRATION. Par un dévissage standard, on se ramène au cas où f est galoisien de degré p. Dans ce cas, on utilise la formule d'Evens-Kahn-Fulton-MacPherson [3], [4, th. 14.2]

$$c(f_*E) = \mathcal{N}(c(E)) + \sum_{i=0}^{n-1}\big((1-\mu^{p-1})^{n-i}-1\big)\mathcal{N}(c_i(E)) \tag{5}$$

où $n = \operatorname{rg} E$, $\mathcal{N}$ est le transfert multiplicatif d'Evens-Steiner [2], [15] et $\mu = c_1(L)$ avec $f_*\mathbf{1} = \bigoplus_{\mathbf{i=0}}^{\mathbf{p-1}} \mathbf{L}^{\otimes \mathbf{i}}$ ($\mathbf{1}$ est le fibré trivial de rang 1).

Pour voir (i), il suffit de vérifier que $\mathcal{N}(c(E))$ ne fait intervenir que des classes de degré divisible par $p-1$, ce qui résulte de [4, th. 8.1]. D'après [4, cor. 5.7], on a la formule

$$c_i(f_*E) = f_*c_i(E) + c_i(nf_*\mathbf{1}) \quad \text{pour} \quad \mathbf{i} \leq \mathbf{r}.$$

Écrivons $n = n_0p^\nu$. Alors $c(nf_*\mathbf{1}) = (\mathbf{1}-\mu^{\mathbf{p-1}})^{\mathbf{n}} = (\mathbf{1}-\mu^{(\mathbf{p-1})\mathbf{p}^\nu})^{\mathbf{n_0}}$, donc $c_i(nf_*\mathbf{1}) = \mathbf{0}$ pour $i < (p-1)p^\nu$, ce qui démontre (ii) et (iii). □

2.2. PROPOSITION. *Soient p un nombre premier, X un espace topologique et E un fibré sur X vérifiant la condition du théorème 1.1, et tel que $p^{\nu_p(E)} \mid \operatorname{rg}(E)$. Alors*

$$\nu_p(E \boxtimes B(r_{\mathbb{Z}/p})) = \nu_p(E) + 1$$

où $E \boxtimes B(r_{\mathbb{Z}/p})$ est le produit tensoriel externe de E et de $B(r_{\mathbb{Z}/p})$ sur l'espace $X \times B\mathbb{Z}/p$.

Démonstration. On peut écrire

$$r_{\mathbb{Z}/p} = \bigoplus_{\chi \in X(\mathbb{Z}/p)} \chi$$

où $X(\mathbb{Z}/p)$ est le groupe des caractères de $\mathbb{Z}/p$. Choisissons un générateur τ de ce groupe. On a alors

$$c(E \boxtimes B(r_{\mathbb{Z}/p})) = c(E \boxtimes \bigoplus_{i \in \mathbb{Z}/p} B(\tau^i)) = \prod_{i \in \mathbb{Z}/p} c(E \boxtimes B(\tau^i)).$$

Soit $n = \operatorname{rg} E$. On a

$$c(E \boxtimes B(\tau^i)) = \sum_{j=0}^{n} c_j(E) \times (1 + ic_1(\tau))^{n-j}$$

où $\times$ désigne le cross-produit. Les termes de plus bas degré sont

$$(1 + ic_1(\tau))^n + c_r(E)$$

où $r = p^{\nu_p(E)} - p^{\nu_p(E)-1}$. Par hypothèse sur n, les termes apparaissant dans $(1 + ic_1(\tau))^n - 1$ sont tous de degré $> r$. On a donc

$$c(E \boxtimes B(\tau^i)) \equiv 1 + c_r(E) \pmod{\deg r + 1}$$

et

$$c(E \boxtimes B(r_{\mathbb{Z}/p})) \equiv (1 + c_r(E))^p \equiv 1 \pmod{\deg r + 1}.$$

Il reste à voir que $c_{p^{\nu+1}-p^\nu}(E \boxtimes B(r_{\mathbb{Z}/p})) \neq 0$; pour cela, il suffit de vérifier que le coefficient numérique de $c_{(p-1)p^{\nu-1}}(E) \boxtimes c_1(\tau)^{(p-1)^2 p^{\nu-1}}$ dans sa décomposition de Künneth est $\neq 0$. Or

$$\begin{aligned}
c(E \boxtimes B(r_{\mathbb{Z}/p})) &\equiv \prod_{i=0}^{p-1} \big((1 + ic_1(\tau))^n + c_r(E) \times (1 + ic_1(\tau))^{n-r}\big) \\
&\equiv \prod_{i=0}^{p-1} (1 + ic_1(\tau))^n \prod_{i=0}^{p-1} \big(1 + c_r(E) \times (1 + ic_1(\tau))^{-r}\big) \\
&\equiv (1 - c_1(\tau)^{p-1})^n \big(1 + c_r(E) \times \sum_{i=0}^{p-1} (1 + ic_1(\tau))^{-r}\big)
\end{aligned}$$

$$\equiv (1 - c_1(\tau)^{p^{\nu+1}-p^\nu})^{n_0} \big(1 + c_r(E) \times \sum_{i=0}^{p-1} (1 + ic_1(\tau)^{p^{\nu-1}})^{-p+1}\big) \pmod{c_r(E)^2}$$

où on a posé $n = n_0 p^\nu$. Le terme cherché est donc

$$\sum_{i=0}^{p-1} \binom{-p+1}{(p-1)^2} i^{(p-1)^2} c_{(p-1)p^{\nu-1}}(E) \times c_1(\tau)^{(p-1)^2 p^{\nu-1}}$$

$$= -\binom{-p+1}{(p-1)^2} c_{(p-1)p^{\nu-1}}(E) \times c_1(\tau)^{(p-1)^2 p^{\nu-1}}$$

et on veut voir que $\binom{-p+1}{(p-1)^2} \not\equiv 0 \pmod p$. En écrivant

$$(1+t)^{-p+1} = (1+t^p)^{-1}(1+t) = \sum (-1)^i t^{pi}(1+t) \in \mathbb{F}_p[[t]]$$

on voit que

$$\binom{-p+1}{(p-1)^2} \equiv -1 \pmod p.$$

□

2.3. Remarque. Comme dans [7], définissons pour tout $n \geq 0$

$$I^{(n)}(X) = \{[E] \in K(X) \mid c_i(E) = 0 \text{ pour } i \not\equiv 0 \pmod{p-1},$$
$$\nu_p(E) \geq n \text{ et } \operatorname{rg} E \equiv 0 \pmod{p^n}\}.$$

On vérifie facilement (par le principe de scindage, par exemple) que $I^{(n)}(X)$ est un *idéal* de $K(X)$, stable par image réciproque. La proposition 2.1 montre que $I^{(n)}(X)$ se conserve également par image directe pour un revêtement fini. Comme dans [7], on peut demander:

Question. Est-il vrai que $I^{(m)}(X)I^{(n)}(X) \subset I^{(m+n)}(X)$ pour tous m, n, X?

Il est facile de voir, encore par le principe de scindage, que la réponse est oui pour $m = 1$.

3. Représentations rationnelles

Soient G un groupe et $\rho : G \to GL_n(\mathbb{C})$ une représentation de dimension finie de G. La construction de Borel associe à ρ un fibré vectoriel $B\rho$ sur l'espace classifiant BG de G. Les classes de Chern de $B\rho$ sont par définition les classes de Chern de ρ; on les note $c_i(\rho)$.

Supposons ρ définie sur $\mathbb{Q}$. Alors la condition du théorème 1.1 est vérifiée pour tout p [5]; l'invariant $\nu_p(B\rho) =: \nu_p(\rho)$ est donc défini.

Supposons G fini, et prenons $\rho = r_G$, représentation régulière de G. Alors ρ est évidemment définie sur $\mathbb{Q}$. L'entier $\nu_p(r_G)$ est noté $\nu_p(G)$.

Voici quelques propriétés de l'invariant $\nu_p(G)$:

3.1. Proposition. *Soient G un groupe fini, p un nombre premier et H un sous-groupe de G. Alors*
a) On a $\nu_p(H) \leq \nu_p(G)$ et $c_{p^{\nu_p(H)}-p^{\nu_p(H)-1}}(r_G) = \operatorname{Cor}_H^G c_{p^{\nu_p(H)}-p^{\nu_p(H)-1}}(r_H)$.
b) Supposons que H soit un p-sous-groupe de Sylow de G. Alors on a $\nu_p(H) = \nu_p(G)$.

Démonstration. a) résulte de la proposition 2.1. Pour voir b), on écrit

$$\operatorname{Res}_H^G r_G = (G : H) r_H$$

donc

$$\mathrm{Res}_H^G c(r_G) = c((G:H)r_H) = c(r_H)^{(G:H)}.$$

Comme $(G:H)$ est premier à p, on a $\nu_p((G:H)r_H) = \nu_p(r_H)$. Par ailleurs, l'homomorphisme de restriction

$$\mathrm{Res}: H^*(G, \mathbb{Z}/p) \to H^*(H, \mathbb{Z}/p)$$

est injectif. Il en résulte que $\nu_p(r_G) = \nu_p((G:H)r_H)$. □

3.2. Proposition. *Soit G un groupe fini. Alors $\nu_p(G \times \mathbb{Z}/p) = \nu_p(G) + 1$.*

Cela résulte de la proposition 2.2. □

3.3. Proposition. *Pour tout groupe fini G, on a $\nu_p(G) \leq v_p(|G|)$.*

Démonstration. Soit $r = p^{v_p(|G|)} - p^{v_p(|G|)-1}$: on va montrer que $c_r(r_G) \neq 0$. D'après la proposition 3.1, on peut supposer que G est un p-groupe. Soit E un sous-groupe d'ordre p de G. On a

$$\mathrm{Res}_E^G c(r_G) = c(\mathrm{Res}_E^G r_G) = c((G:E)r_E) = c(r_E)^{(G:E)}$$

On a $c(r_E) = 1 - c_1(\rho)^{p-1}$, où ρ est un caractère non trivial de E. On sait que $c_1(\rho)$ n'est pas nilpotent dans $H^*(E, \mathbb{Z}/p)$; par conséquent, $c_{p-1}(r_E)^{(G:E)} \neq 0$. Mais

$$c(r_E)^{(G:E)} = (1 + c_{p-1}(E) + \dots)^{(G:E)} = 1 + c_{p-1}(E)^{(G:E)} + \dots$$

puisque $(G:E)$ est une puissance de p. On en conclut que

$$\mathrm{Res}_E^G c_r(r_G) = c_{p-1}(E)^{(G:E)} \neq 0.$$

□

Vu le théorème 1.1, l'assertion de Venkov se reformule ainsi:

3.4. Assertion. (Venkov) *Pour tout groupe fini G et tout nombre premier p, on a $\nu_p(G) = v_p(|G|)$.*

La proposition 3.2 implique:

3.5. Proposition. [16] *L'assertion de Venkov est vraie pour un p-groupe abélien élémentaire.* □

4. Groupes d'ordre p^2 et p^3

4.1. Proposition. *L'assertion de Venkov est vraie pour $G = \mathbb{Z}/p^2$.*

Démonstration. On écrit $r_G = \bigoplus_{\chi \in X(G)} \chi$. Choisissant un générateur τ de $X(G)$, on obtient

$$c(r_G) = \prod_{i=0}^{p^2-1} (1 + ic_1(\tau)) = \prod_{i=0}^{p-1} (1 + ic_1(\tau))^p = (1 - c_1(\tau)^{p-1})^p = 1 - c_1(\tau)^{p^2-p}.$$

□

4.2. PROPOSITION. *Supposons p impair. L'assertion de Venkov est vraie pour le groupe non abélien G d'ordre p^3 et d'exposant p.*

DÉMONSTRATION. En effet, G est produit semi-direct de C et H pour tout sous-groupe H d'indice p de G et tout C engendré par un $g \notin H$. Par la proposition 3.5, $\nu_p(H) = 2$. On a

$$r_H = (r_{G/C})_{|H}$$

où $r_{G/C}$ est la représentation de permutation de G sur G/C, d'où

$$\operatorname{Cor}_H^G c_{p^2-p}(r_H) = \operatorname{Cor}_H^G \operatorname{Res}_H^G c_{p^2-p}(r_{G/C}) = 0.$$

D'après la proposition 3.1 a), on a donc $\nu_p(G) > 2$. D'autre part, $\nu_p(G) \leq 3$ d'après la proposition 3.3. On en conclut que $\nu_p(G) = 3$, comme souhaité. □

4.3. PROPOSITION. *Soit G un p-groupe tel que tout sous-groupe propre de G soit abélien élémentaire. Alors*

- *soit G est abélien élémentaire;*
- *soit G est cyclique d'ordre p^2;*
- *soit p est impair est G est d'ordre p^3 et d'exposant p.*

DÉMONSTRATION. Supposons que G ne soit pas abélien élémentaire. S'il est cyclique, il est nécessairement d'ordre p^2. Si G n'est pas cyclique, tout élément de G est contenu dans un sous-groupe propre de G, donc est d'ordre $\leq p$. Par conséquent, G est d'exposant p. Si $p = 2$, c'est impossible, car G serait alors abélien élémentaire.
Supposons donc $p > 2$. Comme G n'est pas cyclique, on a $(G : [G, G]) > p$; comme $[G, G]$ est distingué dans G, pour tout $g \in G$ le sous-groupe engendré par g et $[G, G]$ est propre, ce qui implique par hypothèse que $[G, G]$ est *central.* On en déduit une application bilinéaire alternée

$$[,] : G/[G, G] \times G/[G, G] \to [G, G]$$

dont l'image engendre $[G, G]$. La conclusion résulte alors du lemme suivant:

4.4. LEMME. *Soit $[,] : V \times V \to W$ une application bilinéaire alternée, où V, W sont des $\mathbb{F}_p$-espaces vectoriels. Supposons que $[,] \neq 0$, que l'image de $[,]$ engendre W et que, pour tout hyperplan H de V, $[,]_{|H} = 0$. Alors* $\dim V = 2$ *et* $\dim W = 1$.

En effet, choisissons un hyperplan H de V et soit $e \in V \setminus H$. Soit $K \subset H$ un hyperplan de H. Alors $< e, K >$ est un hyperplan de V; il en résulte que $[e, K] = 0$ et donc que $[V, K] = 0$. Mais alors on a $\dim H = 1$: sinon, tout élément de H serait contenu dans un de ses hyperplans, et on aurait $[V, H] = 0$, d'où $[V, V] = 0$ puisque $[,]$ est alternée. □

5. DÉMONSTRATION DE L'ASSERTION DE VENKOV

5.1. LEMME. *Soient G un p-groupe, H un sous-groupe de G d'indice p et $a \in H^*(H, \mathbb{Z}/p)$ une classe de cohomologie invariante sous l'action de G/H.*

Alors

$$\mathrm{Cor}_H^G\, a^p = 0.$$

Démonstration. On a

$$Res_H^G \mathcal{N}(a) = \prod_{\sigma\in G/H} \sigma a = a^p$$

d'où

$$\mathrm{Cor}_H^G\, a^p = \mathrm{Cor}_H^G \mathrm{Res}_H^G \mathcal{N}(a) = 0.$$

□

(Autre démonstration: $(\mathrm{Cor}\, a)^2 = \mathrm{Cor}(a\cdot \mathrm{Res}\,\mathrm{Cor}\, a) = 0$, d'où $(\mathrm{Cor}\, a)^p = \mathrm{Cor}(a^p) = 0$.)

5.2. Lemme clé. *Soit G un p-groupe non abélien élémentaire. Supposons que l'assertion de Venkov soit vraie pour G. Alors $c(r_G) = a^p$, où $a \in H^*(G, \mathbb{Z}/p)$ est invariant sous l'action des automorphismes de G.*

Démonstration. Soit $\Phi(G)$ le sous-groupe de Frattini de G. On peut écrire

$$r_G = r_{G/\Phi(G)} \oplus \rho$$

avec

$$\rho = \rho' \oplus p\rho''$$

où $r_{G/\Phi(G)}$ est la somme des caractères abéliens d'ordre p de G, ρ' est la somme de ses caractères abéliens d'ordre $> p$ et $p\rho''$ est la somme de ses autres caractères irréductibles, avec leur multiplicité. (Rappelons qu'un caractère irréductible intervient dans r_G avec une multiplicité égale à son degré; comme G est un p-groupe, ce degré est une puissance de p. Cf. par exemple [13, p. 30, cor. 1 et p. 68, cor. 2].)

La représentation ρ'' est évidemment invariante par automorphismes de G; il en est donc de même de $c(\rho'')$. Si χ est un caractère abélien d'ordre $p^r > p$, il en est de même de χ^i pour tout i premier à p. On a

$$c(\chi^i) = 1 + ic_1(\chi)$$

donc

$$c\Big(\bigoplus_{i\in(\mathbb{Z}/p^r)^*} \chi^i\Big) = \prod_{i\in(\mathbb{Z}/p^r)^*} (1 + ic_1(\chi)) = (1 - c_1(\chi)^{p-1})^{p^{r-1}}.$$

Soit σ un automorphisme de G. Alors σ permute les caractères abéliens d'ordre p^r ainsi que les orbites $[\chi] = \{\chi^i \mid i \in (\mathbb{Z}/p^r)^*\}$ de l'action de $(\mathbb{Z}/p^r)^*$ sur ces caractères. Par conséquent, pour tout $r > 1$, $Aut(G)$ laisse invariante la classe de cohomologie

$$\prod_{[\chi]} (1 - c_1(\chi)^{p-1})$$

où $[\chi]$ décrit les orbites ci-dessus. On en déduit que $c(\rho')$, et donc $c(\rho)$, est de la forme b^p, où b est invariant sous l'action de $Aut(G)$.

Soient $\nu = v_p(|G|)$ et $d = v_p((G : \Phi(G)))$. Par hypothèse, on a $d < \nu$. Par conséquent, $p^d - 1 < p^\nu - p^{\nu-1}$. Il en résulte que

$$c_i(r_{G/\Phi(G)}) = 0 \quad \text{pour} \quad i \geq p^\nu - p^{\nu-1}.$$

D'autre part, on a par hypothèse

$$c(r_{G/\Phi(G)}) = c(r_G)c(\rho)^{-1} = (1 + c_{p^\nu - p^{\nu-1}}(r_G) + \dots)c(\rho)^{-1}.$$

On en déduit que les $c_i(r_{G/\Phi(G)})$ sont des polynômes en les $c_i(\rho)$ (plus précisément, on a $c_i(r_{G/\Phi(G)}) = -c_i(\rho)$ puisque les seules classes de Chern éventuellement $\neq 0$ de $r_{G/\Phi(G)}$ sont celles de degré $\geq p^d - p^{d-1}$ d'après la proposition 3.5). La conclusion du lemme clé en résulte (avec une formule explicite pour a si besoin est). □

5.3. Théorème. *L'assertion de Venkov est vraie.*

Démonstration. D'après la proposition 3.1 b), il suffit de la démontrer pour tout p-groupe G. On raisonne par récurrence sur $|G|$. Si G est abélien élémentaire, le théorème résulte de la proposition 3.5. De même, si G est cyclique d'ordre p^2 ou non abélien d'ordre p^3 et d'exposant p, il résulte des propositions 4.1 et 4.2.
Supposons maintenant que G ne soit pas de l'un des types ci-dessus. D'après la proposition 4.3, G contient un sous-groupe H d'indice p qui n'est pas abélien élémentaire. Soit $\nu = v_p(|G|)$. Par hypothèse de récurrence, on a $\nu_p(H) = \nu - 1$; d'autre part, d'après le lemme-clé, on a

$$c_{p^{\nu-1} - p^{\nu-2}}(r_H) = a^p$$

où a est invariant sous l'action de $Aut(H)$. En appliquant le lemme 5.1, on en déduit que $\mathrm{Cor}_H^G\, c_{p^{\nu-1} - p^{\nu-2}}(r_H) = 0$. On conclut comme dans la démonstration de la proposition 4.2. □

5.4. Corollaire. (O. Kroll [9]) *Soit G un p-groupe qui n'est pas abélien élémentaire. Alors on a*

$$\prod_{\chi \in H^1(G, \mathbb{Z}/p) - \{0\}} \beta\chi = 0$$

où β est le Bockstein modulo p.

Démonstration. Les arguments du lemme clé montrent que $c(r_{G/\Phi(G)})$ est une puissance p-ième; en particulier, $c_{p^d-1}(r_{G/\Phi(G)}) = 0$, où $p^d = (G : \Phi(G))$, ce qui est équivalent à l'énoncé du corollaire. □

5.5. Remarques.

1. Le corollaire 5.4 a été amélioré par Serre [14]: on a $\prod \beta\chi = 0$, où χ décrit un système de représentants de l'espace projectif sur le $\mathbb{F}_p$-espace vectoriel $H^1(G, \mathbb{Z}/p)$. (Voir aussi [10].)

2. Soit $n = v_p(|G|)$, et soit $r = r_p(G)$ le maximum des rangs des p-sous-groupes abéliens élémentaires de G. En se restreignant à de tels sous-groupes, on voit facilement que les classes
$$c_{p^n-p^{n-1}}(r_G), \ldots, c_{p^n-p^{n-r}}(r_G)$$
sont non nilpotentes et algébriquement indépendantes sur $\mathbb{F}_p$. Par un théorème de Quillen [12, th. 7.1], les autres classes de Chern de r_G sont nilpotentes. D'après un autre théorème de Quillen (*loc. cit.*, cor. 2.4), l'algèbre de cohomologie $H^*(G, \mathbb{Z}/p)$ est un module de type fini sur sa sous-algèbre engendrée par les $c_i(r_G)$; c'est donc un module de type fini sur la sous algèbre engendrée par les $c_{p^n-p^k}(r_G)$ pour $k \in [n-r, n-1]$. Notons que $Aut(G)$ opère trivialement sur cette sous-algèbre. Par ailleurs, le lemme clé 5.2 montre que, si G n'est pas abélien élémentaire, les $c_{p^n-p^k}(r_G)$ sont puissances p-ièmes d'éléments également invariants par $Aut(G)$.
3. Le théorème 5.3 est à comparer avec les résultats obtenus dans [7] pour les classes de Stiefel-Whitney de r_G. Un entier $\nu(\rho)$ analogue à $\nu_p(\rho)$ y est défini pour toute représentation réelle ρ d'un groupe G. Si G est fini, on montre que
$$r_2(G) \leq \nu(r_G) \leq v_2(|G|)$$
où $r_2(G)$ est comme dans la remarque précédente. En général, on a $\nu(r_G) < v_2(|G|)$, par exemple pour $G = \mathbb{Z}/4$. Toutefois, on montre que $\nu(2r_G) = v_2(|G|) + 1$ pour tout groupe fini G.

Références

[1] M.F. Atiyah *Characters and cohomology of finite groups*, Publ. Math. I.H.É.S. 9 (1961), 23–64.

[2] L. Evens *A generalization of the transfer map in the cohomology of groups*, AMS Trans. 108 (1963), 54–65.

[3] L. Evens, D.S. Kahn *An integral Riemann-Roch formula for induced representations of finite groups*, AMS Trans. 245 (1978), 309–330.

[4] W. Fulton, R. MacPherson *Characteristic classes of direct image bundles for covering maps*, Ann. of Math. 125 (1987), 1–92.

[5] A. Grothendieck *Classes de Chern et représentations linéaires des groupes discrets*, *in* Dix exposés sur la cohomologie des schémas, North Holland, 1965.

[6] B. Kahn *Classes de Stiefel-Whitney de formes quadratiques et de représentations galoisiennes réelles*, Invent. Math. 78 (1984), 223–256.

[7] B. Kahn *The total Stiefel-Whitney class of a regular representation*, J. Alg. 144 (1991), 214–247.

[8] S. Mukohda, S. Sawaki *On the $b_p^{k,j}$ coefficient of a certain symmetric function*, J. Fac. Sci. Niigata Univ. 1 (1954), 1–6.

[9] O. Kroll *A representation-theoretical proof of a theorem of Serre*, Aarhus Univ. Preprint Series 33 (1986).

[10] T. Okuyama, H. Sasaki *Evens' norm and Serre's theorem on the cohomology algebra of a p-group*, Arch. Math. (Basel) 54 (1990), 331–339.
[11] F.P. Peterson *A* mod *p Wu formula*. Bol. Soc. Mat. Mexicana 20 (1975), 56–58.
[12] D. Quillen *The spectrum of an equivariant cohomology ring, I, II,* Ann. of Math. 94 (1971), 549–572, 573–602.
[13] J.-P. Serre Représentations linéaires des groupes finis, Hermann, Paris, 1970.
[14] J.-P. Serre *Une relation dans la cohomologie des p-groupes*, C. R. Acad. Sci., Paris, 304 (1987), 587–590.
[15] R. Steiner *Multiplicative transfers in ordinary cohomology*, Proc. Edinburgh Math. Soc. 25 (1982), 113–131.
[16] B.B. Venkov *Classes caractéristiques pour les groupes finis* (en russe), Dokl. Akad. Nauk SSSR 137 (1961), 1274–1277. Traduction anglaise: Sov. Math. Dokl. 2 (1961), 445–447.

Bruno Kahn
Institut de Mathématiques
de Jussieu
Université Paris 7
Case 7012
75251 Paris Cedex 05
France
kahn@math.jussieu.fr

Difference Scheme for the Vlasov-Manev System

R. Illner and S. Rjasanow

Received: January 22, 1999

Communicated by Alfred K. Louis

Abstract. We develop and test a finite difference scheme for the Vlasov-Manev Equation in one space and one velocity dimension. The Manev correction to the Newtonian potential produces visible qualitative differences in the behaviour of stellar systems; the most notable effect observed in this paper is a stabilisation of the separate identities of two Maxwellian concentrations at different locations.

1991 Mathematics Subject Classification: 65N06,65N12,65N22,85A05
Keywords and Phrases: Vlasov-Manev system, Finite difference methods, stability

1 Introduction

We are concerned with the numerical solution of the Vlasov equation with a Manev-type correction to the potential in $1+1$ (one space, one velocity) dimensions. In the three-dimensional case, the Newtonian potential is changed to the "Manev" potential

$$U(|x-y|) = -\frac{\gamma}{|x-y|} - \frac{\delta}{|x-y|^2}. \tag{1}$$

The correction $-\delta/|x-y|^2$ was introduced by Manev in a series of papers [7],[8],[9], [10] in the 1920s in an attempt to find a semiclassical approximation to the relativistic central force problem. For γ the universal gravitational constant and $\delta = 3\gamma^2/c^2$, where c is the speed of light, this correction gives a qualitatively accurate prediction of the precession of the perihelion of Mercury. Manev's work was the main motivation for the recent paper [2], in which the authors discuss the properties of the corresponding stellar dynamic equation. Notably, it is shown that the Cauchy problem for this equation does not, in general, admit global solutions (the corresponding result for the classical stellar dynamic equation holds only if the number of space dimensions is larger than or equal to 4, see, e.g., [4]). This means that a stellar system driven by Manev forces will typically develop features where the spatial density ρ loses smoothness such that the Manev force term, i.e., the Riesz transform of ρ

$$E_2[\rho](t,x) := -\delta \int \frac{x-y}{|x-y|^4}\rho(t,y)\,dy$$

diverges. Possible reasons for this are the local formation of singularities in ρ ("concentrations") or in $\nabla_x \rho$ (e.g., "shock waves"). As it is well known that such singularities do not occur in solutions of the classical stellar dynamic equation (see [4]), the Manev correction may be of physical relevance in modelling the evolution of large stellar systems like galaxies, globular clusters or interstellar dust clouds.

We remark that an equation with $\gamma = 0$ and $\delta > 0$ (referred to as the "pure" Manev case in [2]) possesses an interesting "projective" invariance in addition to the standard translation, scaling and Galilei invariances. Specifically, as shown in [2], if $f(t,x,v)$ is a solution of the pure Manev equation

$$\partial_t f + v \cdot \nabla_x f + E_2[\rho] \cdot \nabla_v f = 0$$

and if for some $a > 0$, $\tau = t/(1+at)$, $y = x/(1+at)$ and $w = (1+at)v - ax$, then

$$F(\tau, y, w) := f(t,x,v)$$

solves

$$\partial_\tau F + w \cdot \nabla_y F + E_2[\tilde{\rho}] \cdot \nabla_w F = 0$$

with $\tilde{\rho} = \int F\,dw$. This "projective" invariance, described in the context of the corresponding N-body problem by Bobylev and Ibragimov [1], may be of significance (and use) in regions where ρ or $\nabla \rho$ are large and the Manev correction dominates the Newtonian forces.

As also discussed in [2], Boltzmann collision terms are dimensionally of the same order as the Manev force term and should therefore be included in a proper model. However, the present study aims at the identification of effects which can be attributed to the Manev correction alone; we therefore omit Boltzmann collision terms and any other conceivable correction (such as, e.g., Smoluchowski type coagulation terms).

It is tempting to try a particle or particle-in-cell scheme for this equation, as is common for the Vlasov-Poisson (VP) system. The most advanced scheme of this type for VP is due to Greengard and Rokhlin [5]. We experimented with particle schemes for the Vlasov-Manev equation and found that they performed poorly due to the strong singularity of the Manev correction at very short range. Specifically, particles could be accelerated to extreme velocities over one time step, an effect which can also happen for VP but which is rare enough to cause no difficulties. It should also be an insignificant effect for the Vlasov-Manev equation while the densities are smooth (this follows from the existence and uniqueness proof shown in [6]), so a good numerical method should reflect this; but particle methods do not. Of course, one could mollify the Manev potential in order to avoid the difficulty, but unless the mollification parameter is chosen with great care, this may obfuscate the effects of the Manev correction relative to the Newtonian forces. It is for these reasons that we decided to avoid particle methods altogether.

The main objective of this paper is therefore the construction and testing of a difference scheme for the Vlasov-Manev (VM) equation in one-dimensional geometry; effective generalisation for multidimensional cases is planned for future work.

Our paper is organised as follows. In Section 2, we describe the VM equation and summarise its properties. In Section 3, a difference scheme for VM is derived, and its properties are formulated and proved. Section 4 contains a few informative numerical examples.

2 The Vlasov-Manev equation and its properties

The one-dimensional Vlasov-Manev equation is associated with the potential

$$U(|x-y|) = -\gamma|x-y| - \delta \ln|x-y|.$$

This potential arises from (1) by assuming homogeneity of the stellar system in the $y-$ and $z-$ directions, and the Vlasov-Manev equation can be written in the form

$$f_t + v\, f_x + E\, f_v = 0, \tag{2}$$

where $f = f(t,x,v) : \mathbb{R}_+ \times \mathbb{R}_x \times \mathbb{R}_v \to \mathbb{R}_+$ is a non-negative distribution density function, $t \geq 0$ is the time variable, x is the space variable and v is the velocity variable. The force function $E = E(t,x)$ is defined as follows:

$$\begin{aligned} E(t,x) &= -\gamma \int_{\mathbb{R}_y} \frac{x-y}{|x-y|}\rho(t,y)dy - \delta \int_{\mathbb{R}_y} \frac{x-y}{|x-y|^2}\rho(t,y)dy, \qquad (3)\\ \rho(t,x) &= \int_{\mathbb{R}_v} f(t,x,v)dv. \end{aligned}$$

or $E = -\nabla U$, with

$$U(t,x) = -\gamma \int |x-y|\rho(t,y)\,dy - \delta \int \ln|x-y|\rho(t,y)\,dy.$$

Here, ρ denotes the spatial density. The non-negative constants γ and δ are given, and typically $\gamma >> \delta$. The equation (2) is complemented with the initial condition

$$f(0,x,v) = f_0(x,v) \geq 0. \tag{4}$$

We summarise the main properties of this equation.

1. Conservation of non-negativity

$$\text{if } f_0(x,v) \geq 0, \text{ then for } t > 0 \quad f(t,x,v) \geq 0.$$

2. Conservation of mass

$$m(t) = \int_{\mathbb{R}_x} \int_{\mathbb{R}_v} f(t,x,v) dv dx = m(0) = \int_{\mathbb{R}_x} \int_{\mathbb{R}_v} f_0(x,v) dv dx.$$

3. The continuity equation reads

$$\begin{aligned} \rho_t(t,x) \quad &+ \quad j_x(t,x) = 0, \\ j(t,x) \quad &= \quad \int_{\mathbb{R}_v} v\, f(t,x,v) dv. \end{aligned} \tag{5}$$

4. Conservation of energy

$$\begin{aligned} e(t) \quad &= \quad \frac{1}{2} \int_{\mathbb{R}_v} \int_{\mathbb{R}_x} v^2 f(t,x,v) dx dv \\ &- \quad \frac{1}{2} \int_{\mathbb{R}_x} \int_{\mathbb{R}_y} (\gamma |x-y| + \delta \ln |x-y|) \rho(t,x) \rho(t,y) dx dy = e(0). \end{aligned}$$

5. Second derivative of the moment of inertia

$$\begin{aligned} \frac{d^2}{dt^2} \int_{\mathbb{R}_x} x^2 \rho(t,x) dx \quad &= \quad 2 \int_{\mathbb{R}_x} \int_{\mathbb{R}_v} v^2 f(t,x,v)\, dv\, dx \\ &- \quad \gamma \int_{\mathbb{R}_x} \int_{\mathbb{R}_y} |x-y| \rho(t,x) \rho(t,y) dy\, dx - \delta\, m^2(0). \end{aligned}$$

In the three-dimensional case and for $\gamma = 0$, the last identity is

$$\frac{d^2}{dt^2}\int x^2\rho(t,x)\,dx = 4e(0),$$

and this can be used to show non-global existence (for $\gamma = 0$) whenever the initial energy is negative. This argument is not applicable in the one-dimensional case under consideration; it may well be that solutions in this situation always exist globally.

3 A difference scheme

We begin our numerical study of the initial value problem (2),(4) with the discretisation of the physical and velocity spaces. First, we restrict the whole space $\mathbb{R}_x \times \mathbb{R}_v$ to a rectangle

$$Q_L = \{(x,v) \in \mathbb{R}_x \times \mathbb{R}_v,\ -L_x \le x \le L_x,\ -L_v \le v \le L_v\},$$

and assume that $f(t,x,v)$ has its support with respect to x and v in the box Q_L. We can then compute the force field by integrating over Q_L alone; later, we shall act as if f is extended periodically in x and v.
The next step is the discretisation of the rectangle Q_L using the nodes

$$\begin{aligned}
(x_i, v_j) &= (-L_x + i\,h_x, -L_v + j\,h_v),\ (i,j) \in Q_n \\
h_x &= \frac{2L_x}{n_x},\ n_x \in \mathbb{N},\ \text{and } n_x \text{ is even}, \\
h_v &= \frac{2L_v}{n_v},\ n_v \in \mathbb{N},\ \text{and } n_v \text{ is even}, \\
Q_n &= \{(i,j) \in \mathbb{Z}^2,\ 0 \le i \le n_x,\ 0 \le j \le n_v\}.
\end{aligned}$$

Furthermore, we introduce the index set $\tilde{Q}_n$ as a subset of all vectors in Q_n excluding those which have the form $(0,j)$ or $(i,0)$.
Let $\tau > 0$ be the time discretisation parameter, and $t_k = k\,\tau,\ k = 0,1,\ldots.$ The function $f(t_k,x,v)$ will now be represented by a vector $f^k \in \mathbb{R}^n,\ n = n_x\,n_v$ with components

$$f_l^k = f_{i,j}^k \approx f(t_k, x_i, v_j),\ (i,j) \in \tilde{Q}_n. \tag{6}$$

Here, l denotes the global index of the vector f^0, given by

$$l = (j-1)\,n_x + i,\ l = 1,\ldots,n.$$

Note that (6) defines the vector f^k for all $(i,j) \in Q_n$ by the assumed periodic extension of f.
It is also convenient to use the matrix form of the unknown function:

$$F^k \in \mathbb{R}^{n_x \times n_v}.$$

The numerical density $\rho_i^k \approx \rho(t_k, x_i)$ can be computed using the midpoint integration rule

$$\begin{aligned}\rho_i^k &= h_v \sum_{j=1}^{n_v} f_{i,j}^k,\ i = 1, \ldots, n_x, \\ \rho_0^k &= \rho_{n_x}^k,\ k = 0, 1, \ldots,\end{aligned}$$

or in the matrix form

$$\rho^k = h_v F^k e_{n_v}, \tag{7}$$

where $\rho^k \in \mathbb{R}^{n_x}$ and $e_{n_v} = (1, \ldots, 1)^T \in \mathbb{R}^{n_v}$.
The total mass of the system is

$$m^k = h_x \sum_{i=1}^{n_x} \rho_i^k$$

and can be computed as follows:

$$m^k = h_x \left(\rho^k, e_{n_x}\right) = h_x\, h_v \left(f^k, e_n\right) = h_x\, h_v \left(F^k e_{n_v}, e_{n_x}\right). \tag{8}$$

The next and most involved step is the numerical computation of the force due to (3). We take advantage of the fact that it is sufficient to integrate over one spatial period in (3), because what is really done is treat the case where the support of ρ stays inside such a period.
Using the notation

$$P(x-y) = -\gamma\, \frac{x-y}{|x-y|} - \delta\, \frac{x-y}{|x-y|^2}$$

and the piecewise representation of the density ρ we compute

$$E(t_k, x_i) \approx E_i^k = \sum_{j=1}^{n_x} \rho_j^k G_{ij}, \tag{9}$$

$$E^k = G\,\rho^k,\ E^k \in \mathbb{R}^{n_x},\ G \in \mathbb{R}^{n_x \times n_x}. \tag{10}$$

The elements of the matrix G are defined by

$$G_{ij} = \int_{x_j - h_x/2}^{x_j + h_x/2} P(x_i - y)dy. \tag{11}$$

Direct computation of the force via (9) will require $O(n_x^2)$ arithmetical operations in each time step and is therefore an "expensive" step. The combination of the special form of the matrix G and uniform discretisation leads to a special, Toeplitz form of the matrix G.

Lemma 1 *The matrix G defined in (11) is a skew-symmetric Toeplitz matrix.*

Proof:
A matrix G is Toeplitz if

$$G_{i+1,j+1} \quad = \quad G_{ij}, \ i,j = 1, \ldots, n_x - 1.$$

The analytical integration in (11) leads to

$$G_{ij} = \gamma\, h_x - \delta\, \ln \frac{j - i - 1/2}{j - i + 1/2}, \ j > i.$$

For $j < i$ we get

$$G_{ij} = -G_{ji}$$

because of (11).
The matrix G is therefore uniquely defined by its first row. The element G_{11} is a strongly singular integral which should be considered as a Cauchy integral

$$G_{11} = \lim_{\varepsilon \to 0} \left(\int_{x_1 - h_x/2}^{x_1 - \varepsilon} P(x_1 - y)dy + \int_{x_1 + \varepsilon}^{x_1 + h_x/2} P(x_1 - y)dy \right).$$

Using the substitution $y = -y' + 2x_1$ in the second integral and the obvious property

$$P(x - y) = -P(y - x)$$

we obtain

$$\begin{aligned}
\int_{x_1+\varepsilon}^{x_1+h_x/2} P(x_1-y)dy &= -\int_{x_1-\varepsilon}^{x_1-h_x/2} P(x_1+y'-2x_1)dy' \\
&= \int_{x_1-h_x/2}^{x_1-\varepsilon} P(y'-x_1)dy' \\
&= -\int_{x_1-h_x/2}^{x_1-\varepsilon} P(x_1-y')dy',
\end{aligned}$$

and therefore

$$G_{11} = 0.$$

■

Remark 1 *The multiplication of a Toeplitz matrix with a vector can be realised efficiently using the following well known trick. The matrix G can be considered as a left-upper block of the circulant matrix $\tilde{G}$ of the dimension m which is a power of two:*

$$\tilde{G} = \begin{pmatrix} G & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

The matrix $\tilde{G}$ has the following additional property

$$\tilde{G}_{i,m} = \tilde{G}_{i+1,1}, \; i = 1, \dots, m-1$$

and its first row is defined as

$$(G_{11}, \dots, G_{1,n_x}, 0, \dots, 0, -G_{n_x,1}, -G_{n_x-1,1}, \dots, -G_{2,1}) \in \mathbb{R}^{1\times m}. \tag{12}$$

The dimension m of the matrix $\tilde{G}$ is the next power of two for the number $2\,n_x - 1$ and the number of zeros in (12) is equal to $m - 2n_x + 1$. It is obviously true that $2\,n_x - 1 < m < 4\,n_x - 4$.
Each circulant matrix C of the dimension m can be represented as

$$C = m^{-1} F_m \, \Lambda \, F_m^*,$$

where F_m denotes the matrix of the Discrete Fourier Transform (DFT) of the dimension m. The diagonal matrix Λ contains the eigenvalues of C and can be computed as

$$\Lambda = diag(\lambda_1, \dots, \lambda_m) = diag(F_m C^T e_1), \tag{13}$$

e.g. as the DFT of the first row of the matrix C.

The computation of the force due to (10) therefore reduces to one DFT in a preparatory step for the computation of the eigenvalues of the matrix $\tilde{G}$ using (13) (the matrix $\tilde{G}$ remains fixed during the time steps), and then to two DFTs and one multiplication of the diagonal matrix $n_x^{-1}\Lambda$ with a vector. If m is a power of two then the computation of the force only requires $O(n_x \log_2(n_x))$ arithmetical operations using the Fast Fourier Transform (FFT) [3],[13] and is therefore much "cheaper" than the computation of the density via (7) which requires $O(n) = O(n_x\, n_v)$ arithmetical operations. The number of arithmetical operations for the multiplication

$$\tilde{G}\tilde{\rho}^k = \begin{pmatrix} G & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \begin{pmatrix} \rho_k \\ 0 \end{pmatrix} = \begin{pmatrix} G\,\rho_k \\ G_{21}\,\rho_k \end{pmatrix} \tag{14}$$

is then of the same order $O(n_x \log_2(n_x))$.

Remark 2 *In a two- $(d = 2)$ or three-dimensional case $(d = 3)$ we will obtain a circulant-block matrix [12]. Such matrices can again be efficiently multiplied with a vector using the FFT. The amount of arithmetical work would be $O(n_x^d \log_2(n_x))$ in this case.*

The next step in the numerical procedure we are describing is the discretisation of the equation (2) using a semi-implicit difference scheme. At the given time level k we compute the density ρ^k via (7) and the force E^k via (10) or (14), and then we use the following "upwind" approximation for the derivatives in (2):

$$f_t(t_k, x_i, v_j) \;\approx\; f_{t,ij}^k = \frac{f_{ij}^{k+1} - f_{ij}^k}{\tau}, \tag{15}$$

$$v\, f_x(t_k, x_i, v_j) \;\approx\; v_j\, f_{x,ij}^{k+1} = \begin{cases} v_j \dfrac{f_{ij}^{k+1} - f_{i-1,j}^{k+1}}{h_x} & , \quad v_j \geq 0 \\ v_j \dfrac{f_{i+1,j}^{k+1} - f_{ij}^{k+1}}{h_x} & , \quad v_j < 0 \end{cases}, \tag{16}$$

$$E\, f_v(t_k, x_i, v_j) \;\approx\; E_i^k\, f_{v,ij}^{k+1} = \begin{cases} E_i^k \dfrac{f_{ij}^{k+1} - f_{i,j-1}^{k+1}}{h_v} & , \quad E_i^k \geq 0 \\ E_i^k \dfrac{f_{i,j+1}^{k+1} - f_{ij}^{k+1}}{h_v} & , \quad E_i^k < 0 \end{cases}. \tag{17}$$

The resulting difference scheme can now be written in the form

$$f_{t,ij}^k + v_j\, f_{x,ij}^{k+1} + E_i^k\, f_{v,ij}^{k+1} = 0,\; i,j \in \tilde{Q}_n,\; k = 0, 1, \ldots\,. \tag{18}$$

The initial values $f_{ij}^0 = f_0(x_i, v_j)$ are given. After multiplication by τ, (18) is a system of linear equations which can be written in the matrix form

$$A_k\, f^{k+1} = f^k,\ A_k \in \mathbb{R}^{n\times n},\ f^k, f^{k+1} \in \mathbb{R}^n,\ n = n_x\, n_v.$$

The elements of the matrix A_k using the global numbering with $l = n_x(j-1)+i$ are of the form

$$\begin{aligned}
(A_k)_{ll} &= 1 + \frac{\tau}{h_x}|v_j| + \frac{\tau}{h_v}|E_i^k|, && (19)\\
(A_k)_{l,l-1} &= \begin{cases} -\frac{\tau}{h_x}|v_j| & ,\quad v_j \geq 0 \\ 0 & ,\quad v_j < 0 \end{cases}, && (20)\\
(A_k)_{l,l+1} &= \begin{cases} 0 & ,\quad v_j \geq 0 \\ -\frac{\tau}{h_x}|v_j| & ,\quad v_j < 0 \end{cases}, && (21)\\
(A_k)_{l,l-n_x} &= \begin{cases} -\frac{\tau}{h_v}|E_i^k| & ,\quad E_i^k \geq 0 \\ 0 & ,\quad E_i^k < 0 \end{cases}, && (22)\\
(A_k)_{l,l+n_x} &= \begin{cases} 0 & ,\quad E_i^k \geq 0 \\ -\frac{\tau}{h_v}|E_i^k| & ,\quad E_i^k < 0 \end{cases}, && (23)\\
l &= 1,\ldots,n\ .
\end{aligned}$$

All other elements of the matrix A_k are equal to zero, i.e. the matrix A_k is extremely sparse. Exactly three elements of each row of this matrix are unequal to zero. In a $d-$dimensional case this number would be $2d+1$.

Remark 3 *If the indices in (15)-(17) or in (20)-(24) are not from the set $\tilde{Q}_n$ then we always assume the periodic property (e.g. $f^{k+1}_{n_x+1,j} \equiv f^{k+1}_{1,j}$ etc.).*

The main properties of the difference scheme (18) correspond to the properties of the matrices $A_k,\ k = 0,\ldots$.

Lemma 2 *The matrix A_k has the following properties*

1. *A_k is a regular M-matrix,*

2. *$A_k\, e_n = e_n,\ A_k^T\, e_n = e_n,$*

3. *$\|A_k^{-1}\|_2 = 1.$*

Here $\|A_k^{-1}\|_2$ denotes the spectral norm of the matrix A_k^{-1}, i.e. its biggest singular value.

Proof:

1. The elements of the matrix A_k fulfil

$$\begin{aligned}(A_k)_{ll} &> 0,\ (A_k)_{lm} \le 0,\ l \neq m, \\ \sum_{m=1}^{n} (A_k)_{lm} &= 1,\ l = 1, \ldots, n. \end{aligned} \tag{24}$$

   By (24), the matrix A_k is strongly diagonal-dominant and therefore a regular $M-$matrix.

2. The first property is given trivially in (24). This means that the vector e_n is an eigenvector of the matrix A_k and corresponds to the eigenvalue one. The matrix A_k^T has the same eigenvector and the same eigenvalue because

$$\begin{aligned}\left(A_k^T\, e_n\right)_l &= \sum_{m=1}^{n} (A_k)_{ml} = \\ &= (A_k)_{ll} + (A_k)_{l+1,l} + (A_k)_{l-1,l} + (A_k)_{l-n_x,l} + (A_k)_{l+n_x,l}\,.\end{aligned}$$

   Using the representations $(l+1,l) = (l+1,(l+1)-1)$ and $(l-1,l) = (l-1,(l-1)+1)$, the property $l \pm 1 = (n_x-1)j + (i \pm 1)$ and (20),(21) we obtain

$$(A_k)_{l+1,l} + (A_k)_{l-1,l} = -\frac{\tau}{h_x}|v_j|.$$

   By analogy

$$(A_k)_{l-n_x,l} + (A_k)_{l+n_x,l} = -\frac{\tau}{h_v}|E_i^k|.$$

   Together with (19) we obtain the required result.

3. The matrix A_k^{-1} is element-wise non-negative, because it is the inverse of the $M-$matrix. The spectral norm of the matrix A_k^{-1} is equal to its largest singular value or to the square root of the largest eigenvalue of the matrix $A_k^{-T}A_k^{-1}$. This matrix only has non-negative elements (as a product of two element-wise non-negative matrices) and the real eigenvector e_n only has positive components. Then the corresponding eigenvalue (Perron-Frobenius theorem) is the largest Perron-eigenvalue of this matrix. In our case this eigenvalue is equal to one, because of the properties in 2. Hence the spectral norm of the matrix A_k^{-1} is equal to one.

■

More details concerning $M-$matrices can be found in [11].
The above lemma enables us to prove some important properties of the difference scheme

$$\begin{array}{ll} 1. & \text{Initial step} \\ & f^0_{ij} = f_0(x_i, v_j), \\ 2. & \text{Time step for } k = 0, 1, \ldots \\ 2.1 & \rho^k = h_v\, F^k\, e_{n_v}, \\ 2.2 & E^k = G\,\rho^k, \\ 2.3 & A_k f^{k+1} = f^k. \end{array} \tag{25}$$

Corollary 1 *The solution of the difference scheme (25) exists for all* $k = 0, 1\ldots$.

Proof: This property follows directly from the regularity of the matrix A_k for all $k = 0, 1\ldots$.

■

Corollary 2 *If the initial function* $f_0(x, v)$ *is non-negative then the vectors* f^k *remain component-wise non-negative for all* $k = 0, 1, \ldots$.

Proof:
The initial vector f^0 is component-wise non-negative because of its definition in step 1 of (25). If $f^k,\ k = 0, 1, \ldots$ is component-wise non-negative, then we obtain from Step 2.3 of (25)

$$f^{k+1} = A_k^{-1} f^k.$$

The matrix A_k^{-1} is component-wise non-negative because it is inverse of an M-matrix. The proof is then completed by induction.

■

Corollary 3 *The difference scheme (25) conserves mass.*

Proof:
The mass of the system can be computed for $k = 1, 2, \ldots$ corresponding to formula (8)

$$\begin{aligned} m^k = h_x h_v (f^k, e_n) &= h_x h_v (A_{k-1} f^{k-1}, e_n) \\ = h_x h_v (f^{k-1}, A^T_{k-1} e_n) &= h_x h_v (f^{k-1}, e_n) = m^{k-1} = \ldots = m^0. \end{aligned}$$

■

Corollary 4 *The difference scheme (25) is stable in the discrete maximum norm with respect to the initial data.*

Proof: The discrete maximum norm of $f^k,\ k = 0, 1, \ldots$ is defined as

$$\|f^k\|_\infty = \max_l |f_l^k| = \max_l f_l^k = f_{l^*}^k$$

because the components of the vector f^k are non-negative. Here we have used the global numbering $l = n_x(j-1) + i$ of the components of f^k. Using

$$(A_{k-1})_{l^*l^*} > 0,\ (A_{k-1})_{ij} \leq 0,\ i \neq j$$

in the time step $k = 1, 2, \ldots$ we obtain the following estimate for the index l^*

$$\begin{aligned}
\|f^k\|_\infty &= f_{l^*}^k = ((A_{k-1})_{l^*l^*} + (A_{k-1})_{l^*l^*-1} \\
&+ (A_{k-1})_{l^*l^*+1} + (A_{k-1})_{l^*l^*-n_x} + (A_{k-1})_{l^*l^*+n_x})\, f_{l^*}^k \\
&\leq (A_{k-1})_{l^*l^*} f_{l^*}^k + (A_{k-1})_{l^*l^*-1} f_{l^*-1}^k \\
&+ (A_{k-1})_{l^*l^*+1} f_{l^*+1}^k + (A_{k-1})_{l^*l^*-n_x} f_{l^*-n_x}^k + (A_{k-1})_{l^*l^*+n_x} f_{l^*+n_x}^k = \\
&+ (A_{k-1} f^k)_{l^*} = f_{l^*}^{k-1} \leq \|f^{k-1}\|_\infty \leq \cdots \leq \|f^0\|_\infty.
\end{aligned}$$

■

Corollary 5 *The difference scheme (25) is stable in the discrete L_2-norm with respect to the initial data.*

Proof: The discrete L_2-norm of $f^k,\ k = 0, 1, \ldots$ is defined as

$$\|f^k\|_2^2 = h_x h_v (f^k, f^k).$$

Using this definition and property 3 in Lemma 2 we obtain for $k = 1, \ldots$

$$\|f^k\|_2 = \|A_{k-1}^{-1} f^{k-1}\|_2 \leq \|A_{k-1}^{-1}\|_2 \|f^{k-1}\|_2 = \|f^{k-1}\|_2 \leq \cdots \leq \|f^0\|_2.$$

■

Corollary 6 *If the sequence $\{f^k\}$ converges then it converges to the constant*

$$\lim_{k\to\infty} f^k = \frac{m^0}{4L_x L_v}\, e_n$$

Proof: If the sequence $\{f^k\}$ converges to f^∞ then this vector fulfils

$$A_\infty f^\infty = f^\infty,$$

where A_∞ denotes the limit of the sequence of matrices $\{A_k\}$. Since the matrix A_∞ is still a regular $M-$matrix, its eigenvalue 1 is simple. It means that only a constant vector $f^\infty = \alpha e_n$ can fulfil the equation (26). The constant α can be obtained using the conservation of mass

$$m^0 = h_x h_v (f^\infty, e_n) = h_x h_v \alpha (e_n, e_n) = \alpha (h_x n_x)(h_v n_v) = \alpha (4 L_x L_v).$$

■

Next, we obtain the discrete form of the continuity equation (5). We will use the following notations

$$\begin{array}{ll}
v = (v_1, \dots, v_{n_v})^T \in \mathbb{R}^{n_v} & \text{– vector of the velocities,} \\
D_v = \mathrm{diag}(v) \in \mathbb{R}^{n_v \times n_v} & \text{– corresponding diagonal matrix,} \\
D_v^+ = \mathrm{diag}(0.5(|v| + v)) \in \mathbb{R}^{n_v \times n_v} & \text{– positive part of } D_v, \\
D_v^- = \mathrm{diag}(0.5(|v| - v)) \in \mathbb{R}^{n_v \times n_v} & \text{– negative part of } D_v, \\
w = (v_1^2, \dots, v_{n_v}^2)^T \in \mathbb{R}^{n_v} & \text{– vector of squares of the velocities,} \\
E^k = (E_1^k, \dots, E_{n_x}^k)^T \in \mathbb{R}^{n_x} & \text{– vector of the forces,} \qquad (*) \\
D_E = \mathrm{diag}(E^k) \in \mathbb{R}^{n_x \times n_x} & \text{– corresponding diagonal matrix,} \\
D_E^+ = \mathrm{diag}(0.5(|E^k| + E^k)) \in \mathbb{R}^{n_x \times n_x} & \text{– positive part of } D_x, \\
D_E^- = \mathrm{diag}(0.5(|E^k| - E^k)) \in \mathbb{R}^{n_x \times n_x} & \text{– negative part of } D_v, \\
J_m = \mathrm{circ}(0, 1, 0, \dots, 0) \in \mathbb{R}^{m \times m} & \text{– circulant matrix of the dimension } m, \\
\rho^k = h_v \, F^k \, e_{n_v} \in \mathbb{R}^{n_x} & \text{– density,} \\
j^k = h_v \, F^k \, v \in \mathbb{R}^{n_x} & \text{– numerical flux.}
\end{array}$$

Using $(*)$ we rewrite the difference scheme (18) in the matrix form

$$\begin{aligned}
\frac{F^{k+1} - F^k}{\tau} \quad &+ \quad \frac{1}{h_x} \left((I_{n_x} - J_{n_x}) F^{k+1} D_v^- + (I_{n_x} - J_{n_x}^T) F^{k+1} D_v^+ \right) \\
&+ \quad \frac{1}{h_v} \left(D_E^- F^{k+1} (I_{n_v} - J_{n_v}^T) + D_E^+ F^{k+1} (I_{n_v} - J_{n_v}) \right) = 0, \\
k \quad &= \quad 0, 1, \dots \; .
\end{aligned}$$

If we multiply this matrix with the vector $h_v e_{n_v}$ then we obtain using

$$\begin{aligned}
(I_{n_v} - J_{n_v}) e_{n_v} &= (I_{n_v} - J_{n_v}^T) e_{n_v} = 0, \\
D_v^- e_{n_v} &= v^-, \\
D_v^+ e_{n_v} &= v^+
\end{aligned}$$

the following equation

$$\frac{\rho^{k+1} - \rho^k}{\tau} + \frac{h_v}{h_x} \left((I_{n_x} - J_{n_x}) F^{k+1} v^- + (I_{n_x} - J_{n_x}^T) F^{k+1} v^+ \right),$$

$$\frac{\rho^{k+1}-\rho^k}{\tau} + \frac{h_v}{h_x}\big(0.5(I_{n_x}-J_{n_x})F^{k+1}(|v|-v)+0.5(I_{n_x}-J_{n_x}^T)F^{k+1}(|v|+v)\big),$$
$$\frac{\rho^{k+1}-\rho^k}{\tau} + \frac{1}{2h_x}(J_{n_x}-J_{n_x}^T)j^{k+1}+\frac{1}{2}h_xh_v\frac{1}{h_x^2}(2I_{n_x}-J_{n_x}-J_{n_x}^T)F^{k+1}|v|=0$$

or

$$\frac{\rho^{k+1}-\rho^k}{\tau} + \frac{1}{2h_x}(J_{n_x}-J_{n_x}^T)j^{k+1} = -\frac{1}{2}h_xh_v\frac{1}{h_x^2}(2I_{n_x}-J_{n_x}-J_{n_x}^T)F^{k+1}|v|.$$

The short form of this equation is

$$\rho_t^k + j_{\mathring{x}} = -\frac{1}{2}h_x(h_vF^{k+1}|v|)_{xx}, \tag{26}$$

where y_x denotes the central difference and y_{xx} the second difference of the grid function y. The equation (26) corresponds to the continuous equation (5). While the left hand side of (26) is a possible correct approximation of the derivatives in (5), the right hand side forms an artificial viscosity of our scheme. Because of this term which is of the order $O(h_x)$ we are not able to obtain the conservation of the energy of the scheme directly. However, our numerical tests show that the variation of the energy in one time step is small.

4 Numerical examples

In this section we calculate some examples using our difference scheme. The initial distribution $f_0(x,v)$ is given by

$$f_0(x,v)=\frac{1}{2\pi\sqrt{T_xT_v}}\left(\exp\left(-\frac{(x-x_0)^2}{2T_x}\right)+\exp\left(-\frac{(x+x_0)^2}{2T_x}\right)\right)\exp\left(-\frac{v^2}{2T_v}\right),$$

where T_x, T_v and x_0 are some positive parameters. In Figures 1,2 we present the time evolution of the density and of the force in the time interval $(0,1.4)$ for the following setting of parameters: $\gamma=4,\ \delta=0,\ T_x=2,\ T_v=0.05,\ x_0=4$ and $n_x=60,\ n_v=90$, i.e. for the pure Vlasov case. The time interval $(0,1.4)$ is sufficient to show the main numerical effects.

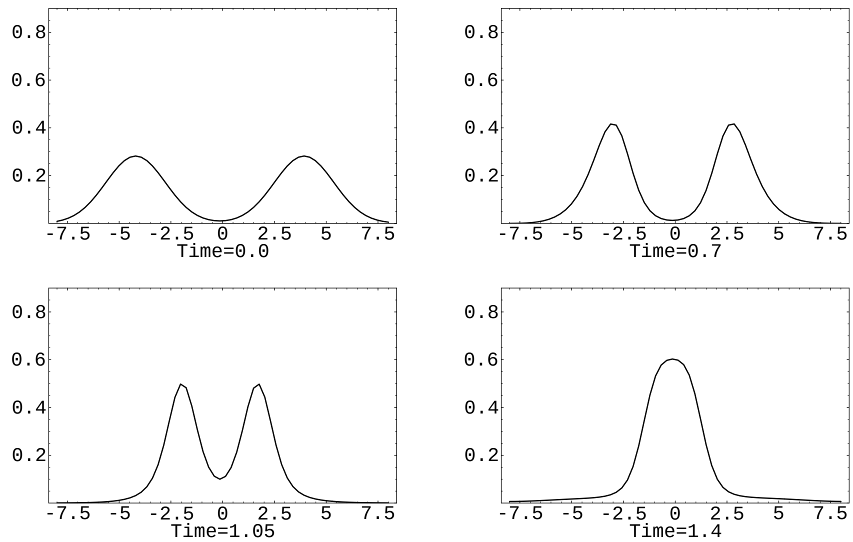


Figure 1: The density profiles for $\gamma = 4$, $\delta = 0$

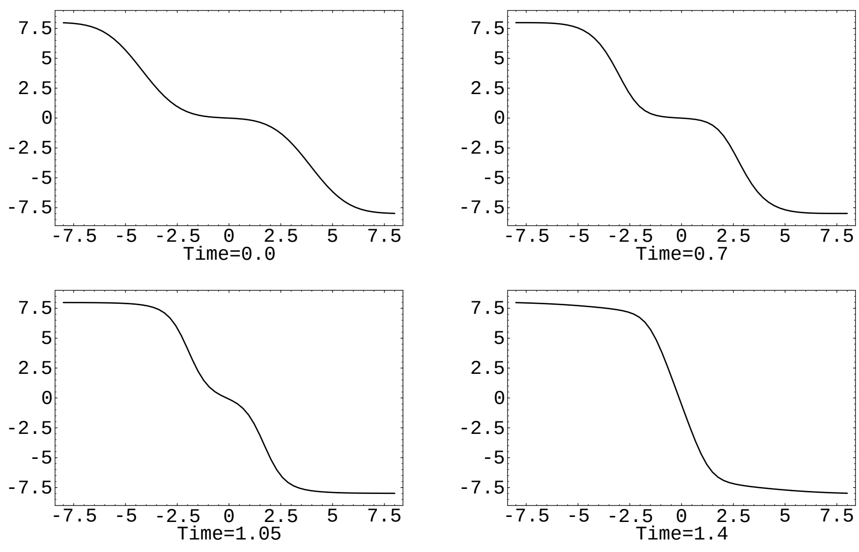


Figure 2: The force profiles for $\gamma = 4$, $\delta = 0$

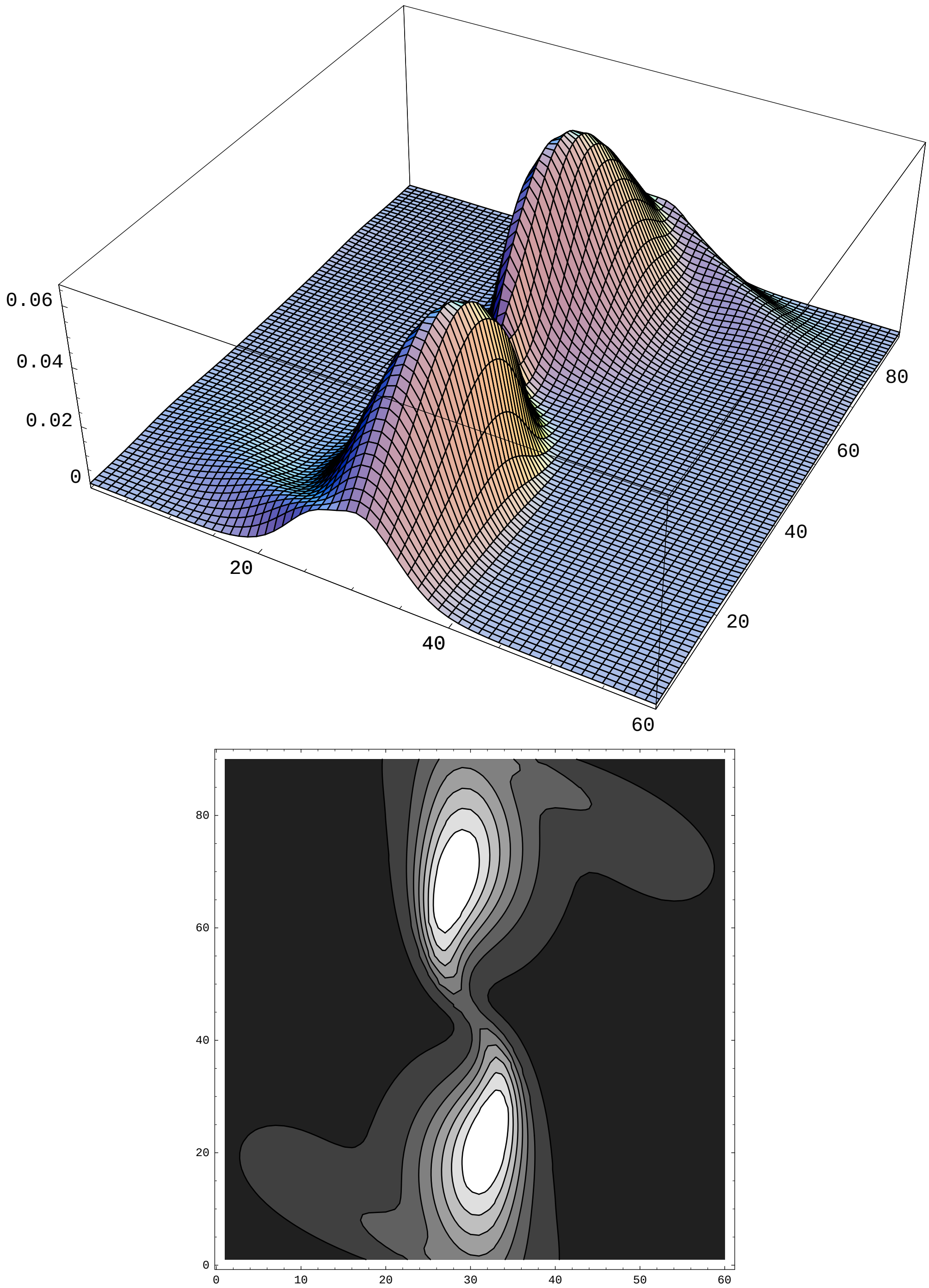


Figure 3: The solution and its iso-lines for $\gamma = 4$, $\delta = 0$

We observe a very clear unification of the two particle "clouds" in space and no remarkable concentration of mass during the time evolution. Figure 3 shows the function $f(t, x, v)$ and its iso-lines for the time $t = 1.4$.
In the second test we consider the pure Manev case with the same initial distribution and the same parameter of discretisation.

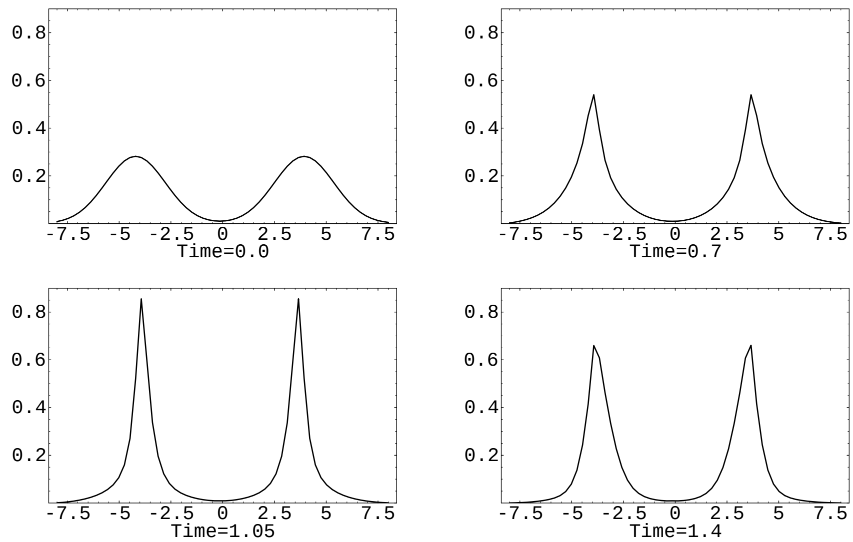


Figure 4: The density profiles for $\gamma = 0$, $\delta = 4$

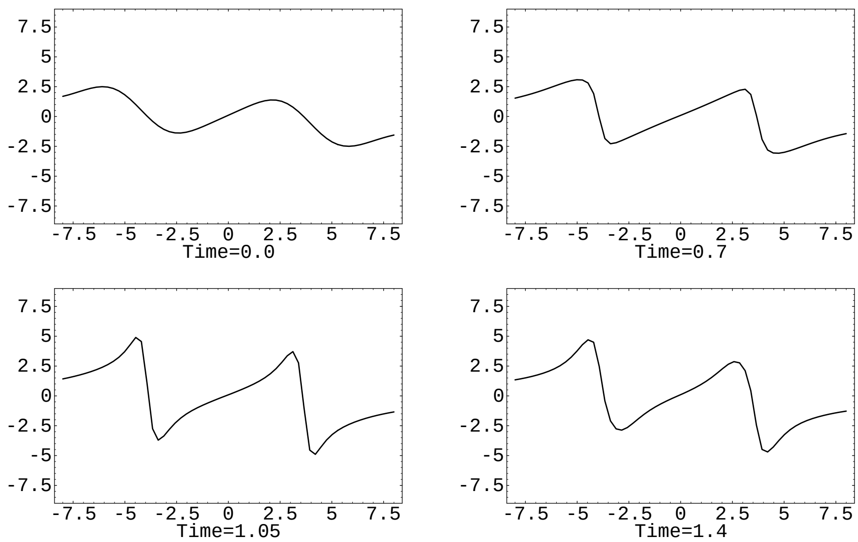


Figure 5: The force profiles for $\gamma = 0$, $\delta = 4$

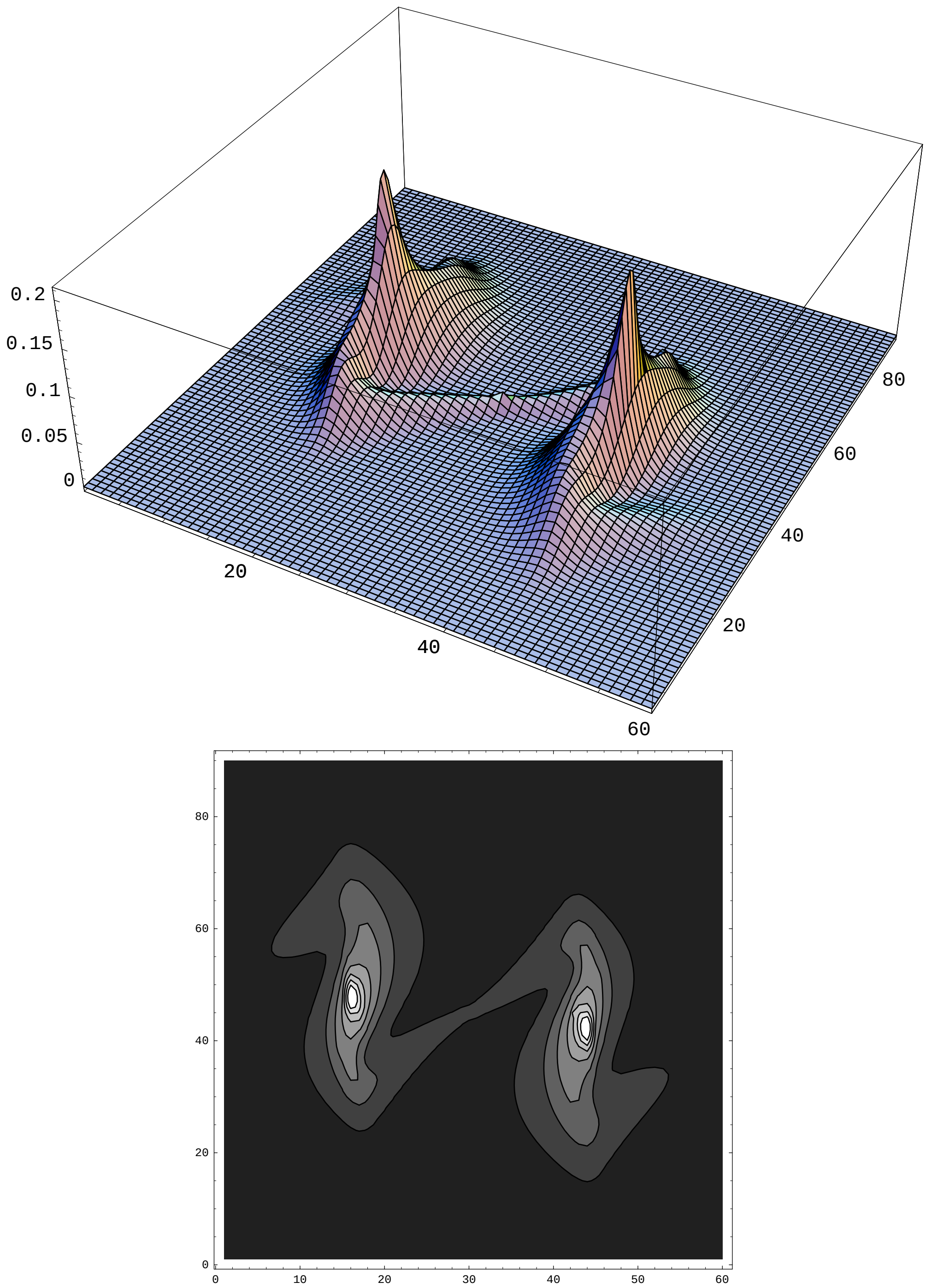


Figure 6: The solution and its iso-lines for $\gamma = 0$, $\delta = 4$

There is a very clear difference in the behaviour of the two examples. The pure Manev case leads to a significant concentration of the mass in the two "clouds", and during the evolution they remain separated. Figure 6 shows the function $f(t,x,v)$ and its iso-lines for the time $t = 1.4$.
Finally, we consider the mixed case $\gamma = 2,\ \delta = 2$ in order to illustrate the influence of the two effects: unification and concentration. The results are presented in Figures 7,8,9.

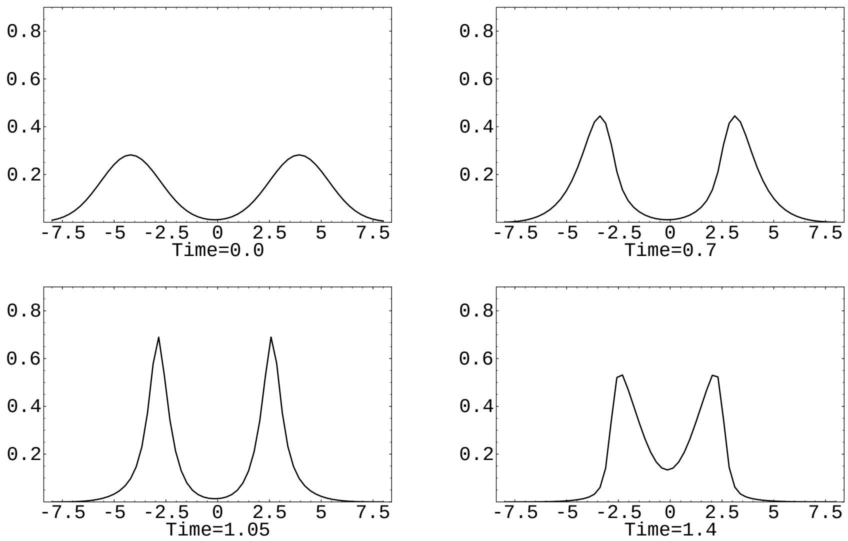


Figure 7: The density profiles for $\gamma = 2$, $\delta = 2$

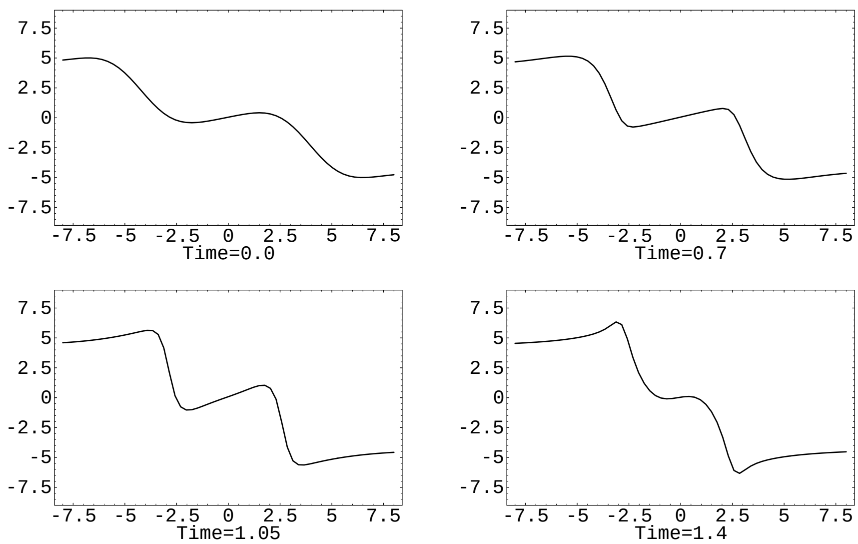


Figure 8: The force profiles for $\gamma = 2$, $\delta = 2$

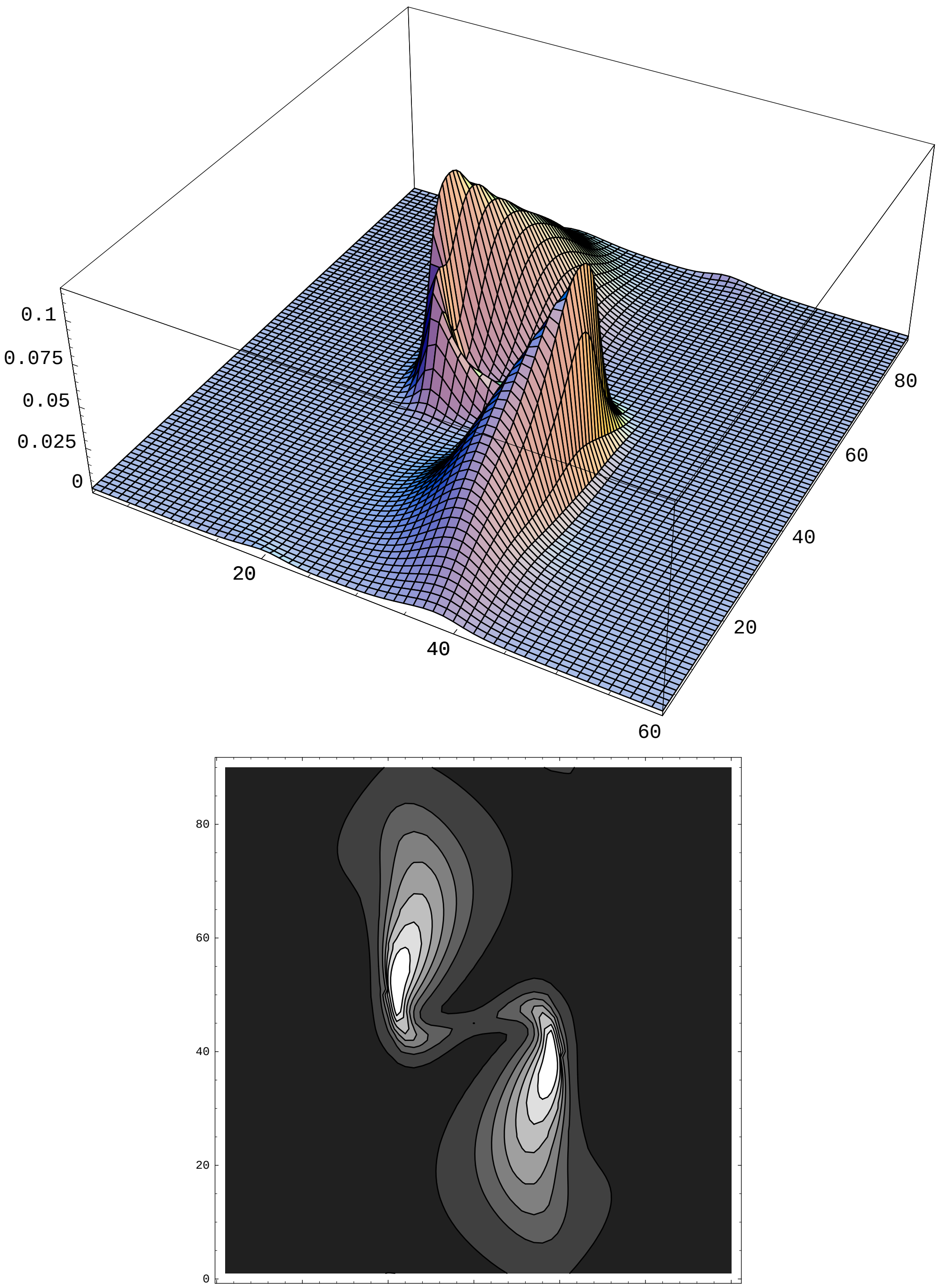


Figure 9: The solution and its iso-lines for $\gamma = 2$, $\delta = 2$

5 Conclusions

Our calculations suggest that the Manev correction will have a "stabilising" effect on isolated one-dimensional matter concentrations; this effect counteracts the tendency of the long-range Newtonian potential to accumulate all matter in one location; while this is only an isolated phenomenon which is observed here as a consequence of the Manev correction, we believe it to be evidence that truly interesting effects may occur in the more relevant three-dimensional case. Numerical experiments to this end are planned.

References

[1] A. Bobyblev and N. Ibragimov. The relationship of the group symmetry properties in dynamics, kinetic theory and hydrodynamics series. *Mathem. Modelirovanie*, 1 : 100–109, 1989.

[2] A. V. Bobylev, P. Dukes, R. Illner, and H. D. Victory. On Vlasov-Manev equations, I: Foundations, Properties and Global Nonexistence. *JSP*, 88:885–911, 1997.

[3] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, 19 : 297–301, 1965.

[4] R. Glassey. *The Cauchy Problem in Kinetic Theory*. SIAM Publ., 1996.

[5] L. Greengard and V. Rokhlin. A Fast Algorithm for Particle Simulations. *J. Comput. Phys.*, 73: 325–348, 1987.

[6] R. Illner, P. Dukes, H. D. Victory, and A. V. Bobylev. On Vlasov-Manev equations, II: Local Existence and Uniqueness. *JSP*, to appear, 1998.

[7] G. Manev. La gravitation et le principe de l'égalité de l'action et de la réaction. *Comptes Rendues*, 178: 2159–2161, 1924.

[8] G. Manev. Die Gravitation und das Prinzip von Wirkung und Gegenwirkung. *Zeitschrift für Physik*, 31: 786–802, 1925.

[9] G. Manev. La gravitation et l'eńrgie au zéro. *Comptes Rendues*, 190: 1374–1377, 1930.

[10] G. Manev. Le principe de la moindre action et la gravitation. *Comptes Rendues*, 190: 963–965, 1930.

[11] H. Minc. *Nonnegative Matrices*. J. Wiley & Sons, 1988.

[12] S. Rjasanow. Effective algorithms with circulant-block matrices. *Linear Alg. Appl.*, 202: 55–69, 1994.

[13] C. Van Loan. *Computational Frameworks for the Fast Fourier Transform*. SIAM, Philadelphia, 1992.

R. Illner
Department of Mathematics
and Statistics
University of Victoria
P.O. Box 3045
Victoria, B.C. V8W 3P4
Canada
rillner@Math.UVic.CA

S. Rjasanow
Fachbereich 9 - Mathematik
Universität des Saarlandes
Postfach 151150
66041 Saarbrücken
Germany
rjasanow@num.uni-sb.de

Sur les Formes Quadratiques de Hauteur 3 et de Degré au Plus 2

Ahmed Laghribi[1]

Received: December 9, 1998
Revised: May 12, 1999

Communicated by Ulf Rehmann

Abstract. Let F be a commutative field of characteristic not 2. In this paper, we give some results on the classification of F-quadratic forms of height 3 and degree ≤ 2.

Résumé. Soit F un corps commutatif de caractéristique différente de 2. Dans ce papier, on donne certains résultats sur la classification des F-formes quadratiques de hauteur 3 et de degré ≤ 2.

1991 Mathematics Subject Classification: 11E04, 11E81
Keywords and Phrases: Quadratic form, Generic splitting of a quadratic form, height and degree of a quadratic form

1. Introduction

A une F-forme quadratique φ de dimension ≥ 3, on associe la quadrique projective X_φ d'équation $\varphi = 0$. On désigne par $F(\varphi)$ le corps des fonctions de X_φ. Lorsque φ est anisotrope de dimension 2 (resp. φ est isotrope de dimension 2 ou $\dim\varphi \leq 1$), on pose $F(\varphi) = F(\sqrt{-\det(\varphi)})$ (resp. $F(\varphi) = F$).

D'après [19], on associe à une forme quadratique $\varphi \not\sim 0$ une suite de formes quadratiques et d'extensions de F, appelée *la tour de déploiement générique* de φ, de la manière suivante: $\varphi_0 = \varphi_{an}$ (la partie anisotrope de φ), $F_0 = F$ et pour $n \geq 1$, on définit par récurrence $F_n = F_{n-1}(\varphi_{n-1})$ et $\varphi_n = ((\varphi_{n-1})_{F_n})_{an}$. *La hauteur* de φ, noté $h = \mathrm{h}(\varphi)$, est le plus petit entier tel que $\dim\varphi_h \leq 1$. Pour $j \in \{0, \cdots, h\}$, on note $i_j(\varphi)$ l'indice de Witt de φ_{F_j}. Clairement, on a $0 \leq i_0(\varphi) < \cdots < i_h(\varphi)$. La suite $\{i_0(\varphi), \cdots, i_h(\varphi)\}$ s'appelle *la suite des indices de déploiement* de φ (splitting pattern [14]). Si $\dim\varphi$ est impaire, alors $\dim\varphi_h = 1$ et φ est dite de degré 0; sinon $\dim\varphi_h = 0$ et donc φ_{h-1} devient hyperbolique sur $F_{h-1}(\varphi_{h-1})$, ce qui implique par un résultat de Knebusch

[1] Soutenu par les fonds FDS de l'Université Catholique de Louvain, Belgique.

[19, Theorem 5.8] et Wadsworth [32], que φ_{h-1} est semblable à une forme de Pfister $\rho \in P_dF_{h-1}$, qu'on appelle *la forme dominante* (leading form) de φ. L'entier d s'appelle *le degré* de φ qu'on note $\deg(\varphi)$. La forme φ est dite *bonne* si sa forme dominante $\rho \in P_dF_{h-1}$ est définie sur F, i.e. s'il existe une F-forme quadratique τ telle que $\rho \cong \tau_{F_{h-1}}$ (dans ce cas τ est unique à isométrie près [20, Proposition 9.2]). Lorsque φ est bonne de forme dominante τ, on dit que φ est *fortement bonne* (resp. *faiblement bonne*) si $\varphi_{F(\tau)}$ est anisotrope (resp. $\varphi_{F(\tau)}$ est isotrope).

Ce procédé de déploiement générique d'une forme quadratique motive le problème suivant, dit *"problème de classification des formes quadratiques par hauteur et degré"*.

Problème: *Etant donné deux entiers positifs h et d, quelles sont les F-formes quadratiques φ telles que* $\mathrm{h}(\varphi) = h$ *et* $\deg(\varphi) = d$*?*

Jusqu'à présent, on a certaines réponses à ce problème. En effet, la caractérisation d'une forme quadratique anisotrope de hauteur 1 a été faite par Knebusch [19, Theorem 5.8] et de manière indépendante par Wadsworth [32]. Une telle forme quadratique est une voisine de codimension 0 ou 1. Dans [20, Lemma 10.1], Knebusch a caractérisé une forme quadratique anisotrope et excellente de hauteur 2 et de degré d en démontrant qu'une telle forme quadratique est de la forme $a\rho \otimes \pi'$ pour $a \in F^*$, $\rho \in P_dF$ et $\pi = \langle 1\rangle \perp \pi' \in P_nF$ avec $n \geq 2$, et il a démontré qu'une forme quadratique anisotrope de hauteur 2 et de degré 1 qui n'est pas excellente est nécessairement de dimension 4 et de discriminant $\neq 1$ [20, Theorem 10.3]. Fitzgerald [8, 1.6] et Knebusch [20, Lemma10.1, Proposition 10.8] ont obtenu certains résultats sur les formes quadratiques anisotropes et bonnes de hauteur et de degré 2. Dans [17], Kahn a caractérisé de manière complète une forme quadratique de hauteur et de degré 2 en démontrant qu'une telle forme quadratique φ est excellente, ou une forme d'Albert (i.e. $\dim\varphi = 6$ et $d_\pm\varphi = 1$), ou $\varphi \in I^2F$ de dimension 8 telle que $\operatorname{ind} c(\varphi) = 2$. Pour les formes quadratiques anisotropes de hauteur 2 et de degré ≥ 3, Hurrelbrink et Rehmann [15, 3.4] ont montré qu'une forme quadratique anisotrope de hauteur 2, de degré 3 qui est bonne mais non excellente est de dimension 16. Ce résultat a été généralisé par Hoffmann [11] qui a montré qu'une forme quadratique anisotrope de hauteur 2, de degré d qui est bonne mais non excellente est de dimension 2^{d+1}. Pour le moment, on n'a pas une caractérisation complète des formes quadratiques de hauteur 2 et de degré ≥ 3. Dans ce sens, Kahn a posé la conjecture suivante.

Conjecture 1. *(Kahn* [17, Conjecture 7]*)*
(1) Une forme quadratique φ anisotrope qui est bonne mais non excellente, est de hauteur 2 *et de degré $d \geq 1$ si et seulement si $\varphi \cong \rho \otimes \psi$ pour $\rho \in P_{d-1}F$ et* $\dim\psi = 4$.
(2) Une forme quadratique φ anisotrope qui n'est pas bonne, est de hauteur 2

et de degré $d \geq 2$ *si et seulement si* $\varphi \cong \rho \otimes \gamma$ *pour* γ *une forme d'Albert et* $\rho \in P_{d-2}F$.

Théorème 1. *(*[17, Théorème 2.12]*)*
La conjecture 1(1) est vraie en degré $d = 3$.

Dans [18], Kahn a obtenu certains résultats sur la caractérisation des formes quadratiques de hauteur 3 et de degré 1.

Théorème 2. *(Kahn* [18, Corollary 1]*)*
Soit φ *une forme quadratique anisotrope de hauteur* 3 *et de degré* 1. *Alors:*
(1) La forme φ *est de l'un des quatres types (s'excluant mutuellement):*
(i) φ *est excellente;*
(ii) φ *n'est pas excellente mais voisine de forme complémentaire une forme de dimension* 4 *et de discriminant* $\neq 1$*;*
(iii) φ *n'est ni voisine ni une forme d'Albert et* $\dim \varphi = 6$*;*
(iv) φ *n'est pas voisine et* $\dim \varphi > 6$. *Dans ce cas,* $(\varphi_{F(\varphi)})_{an}$ *est excellente.*
(2) Si φ *n'est pas voisine telle que* $\dim(\varphi_{F(\varphi)})_{an} = 6$, *alors* $\dim \varphi \leq 16$.

1.1. Les formes quadratiques de hauteur 3, de degré 1 et de dimension > 6 (resp. de hauteur 3, de degré 2 et de dimension > 16). Les formes quadratiques anisotropes de hauteur 3 et de degré 1 qui ont la plus petite dimension sont celles de type (iii) comme dans le théorème 2, i.e. de dimension 6. Dans cette section, on va s'intéresser à la caractérisation des formes quadratiques de hauteur 3, de degré 1 et de dimension > 6 (resp. de hauteur 3, de degré 2 et de dimension > 16). Les formes quadratiques de hauteur 3, de degré 2 et de dimension ≤ 16 seront traitées dans la section 1.2.

Les formes quadratiques φ anisotropes de hauteur 3 et de degré 1 telles que $\dim \varphi > 6$ (resp. de hauteur 3 et de degré 2 telles que $\dim \varphi > 16$), se partagent en quatre types qui s'excluent mutuellement:

Type I: Les formes quadratiques excellentes. Ces formes quadratiques sont décrites dans la proposition suivante.

Proposition 1. *Soit* φ *une forme quadratique anisotrope. Alors, on a équivalence entre:*
(1) φ *est excellente de hauteur* 3 *et de degré* $d \geq 1$*;*
(2) Il existe $a \in F^*$, *des formes de Pfister* $\tau \in P_dF$, $\lambda_1 = \langle 1 \rangle \perp \lambda_1'$, λ_2 *telles que* $\deg(\lambda_1) \geq 1$, $\deg(\lambda_2) \geq 2$ *et* $\varphi \cong a\tau \otimes (\lambda_1' \otimes \lambda_2 \perp \langle 1 \rangle)$.

Type II: Les formes quadratiques voisines mais non excellentes. Ces formes quadratiques sont décrites dans la proposition suivante.

Proposition 2. *Soit* φ *une forme quadratique anisotrope qui n'est pas excellente, de degré* 1 *ou* 2. *Alors, on a équivalence entre:*
(1) φ *est voisine de hauteur* 3*;*
(2) • *Si* $\deg(\varphi) = 1$*:* φ *est voisine de forme complémentaire une forme quadratique de dimension* 4 *et de discriminant* $\neq 1$. *En particulier,* $\dim \varphi = 2^n - 4$

pour un certain entier $n \geq 4$.
• *Si* $\deg(\varphi) = 2$*:* φ *est voisine de forme complémentaire une forme quadratique* ψ *telle que* ψ *est d'Albert ou bien* $\psi \in I^2F$ *de dimension* 8 *avec* $\operatorname{ind} c(\psi) = 2$. *En particulier, il existe un entier* $m \geq 5$ *qui vérifie* $\dim \varphi = 2^m - 6$ *ou* $2^m - 8$.

Type III: Les formes quadratiques φ anisotropes mais non voisines, de dimension différente de $2^{\deg(\varphi)}k$ pour tout entier k impair. Ces formes quadratiques sont décrites dans le théorème suivant.

Théorème 3. *Soit* φ *une forme quadratique anisotrope qui n'est pas voisine, de degré* $d = 1$ *ou* 2 *telle que* $\dim \varphi \neq 2^d k$ *pour tout entier* k *impair. On suppose que* $\dim \varphi > 6$ *lorsque* $d = 1$*, et que* $\dim \varphi > 16$ *lorsque* $d = 2$*. Alors, on a équivalence entre:*
(1) φ *est de hauteur* 3*;*
(2) $\varphi \in GP'_{n,d}F$ *pour un certain entier* $n \geq d + 2$ *(voir définition 3).*

Type IV: Les formes quadratiques φ anisotropes mais non voisines, de dimension $2^{\deg(\varphi)}k$ pour un certain entier k impair. Pour le moment, on n'a pas une caractérisation complète de ces formes quadratiques. D'après la proposition 3, on obtient que ces formes quadratiques sont celles qui sont faiblement bonnes, et par conséquent les formes quadratiques décrites dans le théorème 3 sont celles qui sont fortement bonnes.

Proposition 3. *Soit* φ *une forme quadratique anisotrope qui n'est pas voisine, de degré* 1 *ou* 2 *telle que* $\dim \varphi > 6$ *lorsque* $\deg(\varphi) = 1$*, et que* $\dim \varphi > 16$ *lorsque* $\deg(\varphi) = 2$*. Supposons que* φ *soit de hauteur* 3*. Alors,* φ *est bonne. De plus, on a équivalence entre:*
(1) φ *est faiblement bonne (resp. fortement bonne);*
(2) Il existe un entier k *impair tel que* $\dim \varphi = 2^{\deg(\varphi)}k$ *(resp.* $\dim \varphi \neq 2^{\deg(\varphi)}k$ *pour tout entier* k *impair);*
(3) $\varphi_{F(\tau)} \sim 0$ *(resp.* $\varphi_{F(\tau)}$ *est semblable à une forme de Pfister anisotrope) où* τ *est la forme dominante de* φ.

Le théorème suivant donne une caractérisation d'une forme quadratique φ anisotrope et faiblement bonne, de hauteur 3 et de degré 1 telle que $\dim \varphi \leq 16$. En particulier, on raffine le théorème 2(2) lorsqu'il s'agit d'une forme quadratique faiblement bonne.

Théorème 4. *Soit* φ *une forme quadratique anisotrope de discriminant à signe* $d \neq 1$*. Supposons que* φ *ne soit pas voisine et* $\dim \varphi > 6$*. Alors, on a équivalence entre:*
(1) φ *est faiblement bonne de hauteur* 3 *et de dimension* ≤ 16*;*
(2) φ *est faiblement bonne de hauteur* 3 *et* $\dim(\varphi_{F(\varphi)})_{an} = 6$*;*
(3) $\varphi \cong \langle 1, -d \rangle \otimes \xi$ *pour* ξ *une forme quadratique de dimension* 5.

1.2. Les formes quadratiques de hauteur 3, de degré 2 et de dimension au plus 16.

• Cas des formes quadratiques de dimension 8.

La proposition suivante donne une caractérisation des formes quadratiques anisotropes de dimension 8, de hauteur 3 et de degré 2.

Proposition 4. *([13], [15])*
Une forme quadratique φ anisotrope de dimension 8 est de hauteur 3 et de degré 2 si et seulement si $\varphi \in I^2F$ et $\operatorname{ind} c(\varphi) \geq 4$.

• Cas des formes quadratiques de dimension 10.

Pour les formes quadratiques de dimension 10, de hauteur 3 et de degré 2, on a la caractérisation suivante.

Proposition 5. *([13]) Soit φ une forme quadratique anisotrope de dimension 10 qui n'est pas voisine. On a équivalence entre:*
(1) φ est de hauteur 3 et de degré 2;
(2) Il existe $\pi = \langle 1 \rangle \perp \pi' \in P_3F$, $\tau = \langle 1 \rangle \perp \tau' \in P_2F$ telles que $\varphi \cong a(\pi' \perp -\tau')$ pour $a \in F^*$ convenable;
(3) $\varphi \in I^2F$ et $\operatorname{ind} c(\varphi) = 2$.

• Cas des formes quadratiques de dimension 12.

On n'a pas d'énoncé, même conjecturale, sur la caractérisation des formes quadratiques de dimension 12 fortement bonnes de hauteur 3 et de degré 2. Par contre pour celles qui sont faiblement bonnes, on pose la conjecture suivante.

Conjecture 2. *Soit φ une forme quadratique anisotrope de dimension 12 qui n'est pas voisine. Alors, on a équivalence entre:*
(1) φ est faiblement bonne, de hauteur 3 et de degré 2;
(2) Il existe $\delta \in I^2F$ de dimension 8 telle que $\operatorname{ind} c(\delta) = 2$ et $\varphi \perp \delta \in I^4F$.

La conjecture 2 est liée à la conjecture 3 sur le problème d'isotropie d'une forme quadratique sur le corps des fonctions d'une quadrique (Théorème 5).

Conjecture 3. *Soient $\pi \in P_3F$, $\tau \in P_2F$. Supposons que $\delta := (\pi \perp -\tau)_{an}$ soit de dimension 10. Si ψ est une forme quadratique telle que $\delta_{F(\psi)}$ soit isotrope, alors $\dim \psi \leq \dim \delta$.*

Théorème 5. *La conjecture 3 implique la conjecture 2.*

• Cas des formes quadratiques de dimension 14 ou 16.

On commence par un résultat général.

Proposition 6. *Soit φ une forme quadratique anisotrope de dimension 14 ou 16, de hauteur 3 et de degré 2. Alors, on a les assertions suivantes:*
(1) φ est fortement bonne.
(2) Soit $\tau \in P_2F$ la forme dominante de φ. On a:
(i) Si $\dim \varphi = 14$, alors $\varphi_{F(\tau)}$ est anisotrope de hauteur 2 et de degré 3.
(ii) Si $\dim \varphi = 16$, alors $\varphi_{F(\tau)}$ est anisotrope de hauteur 1 ou de hauteur 2 et de degré 3.

La proposition 6 permet de ramener la caractérisation des formes quadratiques de hauteur 3, de degré 2 et de dimension 14 ou 16 à celle des formes quadratiques de hauteur ≤ 2.

Jusqu'à présent, on ne connait pas un exemple d'une forme quadratique anisotrope de dimension 14, de hauteur 3 et de degré 2. Le théorème 6 précise de manière conjecturale qu'il n'existe pas une telle forme quadratique.

Théorème 6. *Supposons que toute forme quadratique anisotrope de hauteur* 2*, de degré* 3 *qui n'est pas bonne soit de dimension* 12*. Alors:*
(1) Il n'existe pas de forme quadratique anisotrope de hauteur 3*, de degré* 2 *et de dimension* 14*.*
(2) Une forme quadratique anisotrope φ *de dimension* 16 *qui n'est pas voisine est de hauteur* 3 *et de degré* 2 *si et seulement si* $\varphi \in GP_{4,2}F$*.*

Remarque. L'hypothèse qui a été faite dans le théorème 6 est motivée par la conjecture 1(2) en degré $d = 3$. ■

Remerciements. *Ce travail a été fait à l'institut de mathématiques de l'Université Catholique de Louvain, et a été révisé durant mon séjour à la Faculté Jean Perrin. Je remercie ces deux institutions pour l'aide et l'hospitalité que j'ai eues. Je tiens à remercier* B. Kahn *pour ses commentaires intéressants. Aussi, je remercie le referee pour les différentes suggestions concernant la rédaction.*

2. Définitions et rappels de résultats

Toutes les notations et définitions concernant les formes quadratiques se trouvent dans [23] et [29].

La somme orthogonale et le produit de deux formes quadratiques φ et ψ, sont notés respectivement $\varphi \perp \psi$ et $\varphi \otimes \psi$.

Si $a \in F^*$, on note $\langle a \rangle \otimes \varphi = a\varphi$. On dit que ψ est une sous-forme de φ et on note $\psi < \varphi$ s'il existe une forme quadratique ρ telle que $\varphi \cong \psi \perp \rho$ où $\cong$ désigne l'isométrie des formes quadratiques. On dit que φ et ψ sont semblables s'il existe $a \in F^*$ tel que $\varphi \cong a\psi$. Une forme quadratique φ est dite isotrope (resp. hyperbolique) si $\mathbb{H} := \langle 1, -1 \rangle < \varphi$ (resp. $\varphi \cong \mathbb{H} \perp \cdots \perp \mathbb{H}$). L'indice de Witt $i_W(\varphi)$ de φ est le plus grand entier n tel que $n \times \mathbb{H} := \underbrace{\mathbb{H} \perp \cdots \perp \mathbb{H}}_{\text{n fois}} < \varphi$.
Deux formes quadratiques φ et ψ sont dites équivalentes et on note $\varphi \sim \psi$, si $\varphi \perp -\psi$ est hyperbolique. La partie anisotrope de φ est l'unique forme quadratique anisotrope, notée φ_{an}, telle que $\varphi \sim \varphi_{an}$.

Une n-forme de Pfister est une forme de type $\langle 1, -a_1 \rangle \otimes \cdots \otimes \langle 1, -a_n \rangle$, qu'on note $\langle\langle a_1, \cdots, a_n \rangle\rangle$. On note P_nF (resp. GP_nF) l'ensemble des n-formes de Pfister à isométrie près (resp. l'ensemble des formes quadratiques qui sont semblables à des n-formes de Pfister). Une forme quadratique φ est dite voisine s'il existe une n-forme de Pfister π tel que $\dim\varphi > 2^{n-1}$ et $a\pi \cong \varphi \perp \psi$ pour un certain $a \in F^*$ et une certaine forme quadratique ψ qu'on appelle la forme complémentaire de φ; $\dim\psi$ est dite la codimension de φ. Les formes quadratiques π et ψ sont uniques à isométrie près.

On note $I^nF = (IF)^n$ où IF est l'idéal fondamental de $W(F)$ formé des formes quadratiques de dimension paire. On utilisera fréquemment le résultat, dit le Hauptsatz d'Arason-Pfister, qui affirme que si $\varphi \in I^nF$ anisotrope, alors $\dim\varphi \geq 2^n$ [29, Chapter 4, 5.6].

L'ensemble $J_n(F) = \{\varphi \in W(F) \mid \deg(\varphi) \geq n\}$ est un idéal de $W(F)$ qui vérifie $I^nF \subset J_n(F)$ pour tout $n \geq 0$ [19, Theorem 6.4, Corollary 6.6].

L'invariant de Clifford $c(\varphi)$ de φ est la classe dans le groupe de Brauer $\mathrm{Br}(F)$ de F, de $C(\varphi)$ (algèbre de Clifford de φ) ou $C_0(\varphi)$ (algèbre de Clifford paire de φ) suivant que $\dim\varphi$ est paire ou non. On désigne par $\mathrm{ind}\,c(\varphi)$ l'indice de Schur de $C(\varphi)$ ou $C_0(\varphi)$ suivant que $\dim\varphi$ est paire ou non.

Théorème 7. *(1) (Théorème de la sous-forme de Cassels-Pfister* [29, Chapter 4, 5.4(ii)]*) Soient $\psi = \langle 1\rangle \perp \psi'$, φ deux formes quadratiques telles que φ soit anisotrope et que $\varphi_{F(\psi)} \sim 0$. Alors pour tout $\alpha \in D_F(\varphi)$, on a $\alpha\psi < \varphi$. En particulier,* $\dim\varphi \geq \dim\psi$.
(2) Soit τ une forme de Pfister. Alors:
(2.1) (Pfister [29, Chapter 4, 1.5]*) On a que τ est isotrope si et seulement si $\tau \sim 0$. De plus, τ est multiplicative, i.e. $D_F(\tau) = G_F(\tau)$.*
(2.2) ([29, Chapter 4, 5.4(iv)]*)* $\mathrm{Ker}(W(F) \longrightarrow W(F(\tau))) = \tau W(F)$.

Définition 1. [20, Definition 7.7] *Toute forme quadratique de dimension ≤ 1 est dite excellente. Une forme quadratique de dimension ≥ 2 est dite excellente si elle est voisine et sa forme complémentaire est excellente.*

Définition 2. *Soit K/F une extension de corps.*
(1) On dit qu'une K-forme quadratique φ est définie sur F s'il existe une F-forme quadratique ψ telle que $\varphi \cong \psi_K$.
(2) ([20]*) On dit que K/F est excellente si pour toute F-forme quadratique φ, la K-forme quadratique $(\varphi_K)_{an}$ est définie sur F.*
(3) ([17]*) Soit $n \geq 1$ un entier. On dit que K/F satisfait à la descente pour les n-formes de Pfister si pour toute K-forme quadratique $\varphi \in P_nK - \{0\}$ qui est définie sur F, il existe $\psi \in P_nF$ telle que $\psi_K \cong \varphi$.*

Proposition 7. *(1) (*[29, Chapter 2, 5.1] *pour $d = 1$;* [2] *pour $d = 2$) Soit $\pi \in GP_dF$ avec $d \leq 2$. Alors, l'extension $F(\pi)/F$ est excellente.*
(2) ([6, 2.10]*) Si K/F est une extension excellente, alors elle satisfait à la descente pour les n-formes de Pfister quel que soit $n \geq 1$.*

Définition 3. *Soient $n > m \geq 1$ deux entiers, φ une forme quadratique anisotrope de dimension 2^n. On dit que φ appartient à $GP'_{n,m}F$ (resp.*

$GP_{n,m}F$) s'il existe $\tau \in GP_mF$ telle que $\varphi \perp \tau \in J_n(F)$ (resp. $(\varphi \perp \tau)_{an} \in GP_nF$).

L'ensemble $GP_{n,m}F$ a été introduit par Hoffmann [10]. Il est clair que $GP_{n,m}F \subset GP'_{n,m}F$. Dans [10, Conjecture 3.9], Hoffmann a conjecturé que $GP_{n,m}F = GP'_{n,m}F$ (la notation $GP'_{n,m}F$ n'a pas été introduite dans [10]).

PROPOSITION 8. *(*[10, Proposition 3.6]*) Soient $n > m \geq 1$ deux entiers. Alors:*
(1) Toute forme quadratique de $GP'_{n,m}F$ est fortement bonne.
(2) Si de plus $n \geq m+2$, alors toute forme quadratique de $GP'_{n,m}F$ est de hauteur 3 et de degré m.

Pour tout $n \geq 0$ entier, H^nF est le groupe de cohomologie galoisienne $H^n(G_s, \mathbf{Z}/2)$ où G_s est le groupe de Galois d'une clôture séparable de F. Par la théorie de Kummer, on a $H^0F \simeq \mathbf{Z}/2$, $H^1F \simeq F^*/F^{*2}$ et H^2F est isomorphe à la 2-torsion de $\mathrm{Br}(F)$.

D'après Arason [1, Satz 1.6], il existe une application $\tilde{e}^n$ de P_nF vers H^nF, définie par $\tilde{e}^n(\langle\langle a_1, \ldots, a_n\rangle\rangle) = (a_1) \cdot \ldots \cdot (a_n) \in H^nF$ où $\cdot$ est le cup produit de l'algèbre de cohomologie H^*F.

Pour $n = 0, 1, 2$, l'application $\tilde{e}^n$ se prolonge en un homomorphisme e^n de $I^nF/I^{n+1}F$ vers H^nF. Les homomorphismes e^0, e^1, e^2 correspondent à $e^0(\varphi) = \dim\varphi \pmod 2$, $e^1(\varphi) = d_\pm\varphi \pmod{F^{*2}}$ et $e^2(\varphi) = c(\varphi)$.

On a que e^0 et e^1 sont des isomorphismes. L'homomorphisme e^2 est un isomorphisme d'après Merkur'ev [24]; $\tilde{e}^3$ se prolonge en un homomorphisme e^3 de I^3F/I^4F vers H^3F d'après Arason [1], et e^3 est un isomorphisme d'après Merkur'ev-Suslin [26] et Rost [28]; $\tilde{e}^4$ se prolonge en un homomorphisme e^4 de I^4F/I^5F vers H^4F d'après Jacob-Rost [16] et Szyjewski [30], et e^4 est un isomorphisme d'après Rost (non publié). Récemment, Orlov, Vishik et Voevodsky ont montré que $\tilde{e}^n$ se prolonge en un isomorphisme e^n de $I^nF/I^{n+1}F$ vers H^nF pour tout n [27].

3. DÉMONSTRATIONS

Le long de cette section et pour une F-forme quadratique φ, on note $\varphi_1 = (\varphi_{F(\varphi)})_{an}$.

Le lemme suivant est bien connu. On en aura besoin pour la suite.

LEMME 1. *Soient φ une forme quadratique bonne, de degré $d \geq 1$ et de forme dominante τ. Si ρ est une forme quadratique telle que $\varphi \sim \rho \otimes \tau$, alors la dimension de ρ est impaire.*

DÉMONSTRATION. Si la dimension de ρ était paire, on aurait $\rho \in IF$. Puisque $\tau \in I^dF$, on aurait $\varphi \in I^{d+1}F$ et donc φ serait de degré $\geq d+1$, ceci est absurde. ■

3.1. Démonstration de la proposition 1. La démonstration de la proposition 1 se déduit de [18, Proposition 7.17] et [8, Proposition 1.2].

3.2. Démonstration de la proposition 2. La démonstration de la proposition 2 est une conséquence du lemme 2 et de la caractérisation des formes quadratiques de hauteur 2 et de degré ≤ 2.

Lemme 2. *Soient φ une forme quadratique voisine anisotrope et φ' sa forme complémentaire. Alors,* $\deg(\varphi) = \deg(\varphi')$ *et* $\mathrm{h}(\varphi) = \mathrm{h}(\varphi') + 1$.

Démonstration. D'après [15], les formes quadratiques φ' et $\varphi'_{F(\varphi)}$ ont la même suite des indices de déploiement. En particulier, $\mathrm{h}(\varphi') = \mathrm{h}(\varphi'_{F(\varphi)})$ et $\deg(\varphi') = \deg(\varphi'_{F(\varphi)})$. On a $\varphi_{F(\varphi)} \sim -\varphi'_{F(\varphi)}$. D'après [9, Theorem 1], on obtient que $\varphi'_{F(\varphi)}$ est anisotrope. Par conséquent, $\varphi_1 = ((\varphi)_{F(\varphi)})_{an} = -\varphi'_{F(\varphi)}$. Ainsi, $\mathrm{h}(\varphi) = \mathrm{h}(\varphi_1) + 1 = \mathrm{h}(\varphi'_{F(\varphi)}) + 1 = \mathrm{h}(\varphi') + 1$ et $\deg(\varphi) = \deg(\varphi_1) = \deg(\varphi'_{F(\varphi)}) = \deg(\varphi')$. ■

3.3. Un résultat sur les formes quadratiques de hauteur 3 et de degré 2. On aura besoin de la proposition suivante dans les démonstrations de la proposition 3 et du théorème 3.

Proposition 9. *Soit φ une forme quadratique anisotrope qui n'est pas voisine. Supposons que φ soit de hauteur* 3*, de degré* 2 *et de dimension* ≥ 10*. Soient $\varphi_1 = (\varphi_{F(\varphi)})_{an}$ et τ la forme dominante de φ. Alors:*
(1) La forme φ est bonne, i.e. on peut supposer que $\tau \in P_2F$. En particulier, $c(\varphi) = c(\tau)$ *et* $\operatorname{ind} c(\varphi) = 2$.
(2) La forme φ_1 satisfait l'une des deux conditions suivantes:
(2.1) φ_1 est voisine dont la forme complémentaire est semblable à $\tau_{F(\varphi)}$. En particulier, $\varphi_{F(\varphi)(\tau)} \sim 0$.
(2.2) $\dim \varphi_1 = 8$, $c(\varphi_1) = c(\tau_{F(\varphi)})$ *et* $\operatorname{ind} c(\varphi_1) = 2$.
(3) Si $\dim \varphi > 16$*, alors le cas (2.2) est impossible.*

Démonstration. (1) Cette assertion a été prouvée dans [18, Corollary 1(f)].

(2) D'après (1), la forme φ_1 est bonne de hauteur 2 et de forme dominante $\tau_{F(\varphi)} \in P_2F(\varphi)$. En utilisant la caractérisation des formes quadratiques de hauteur et de degré 2 [17], on déduit que l'une des assertions (2.1) et (2.2) est vérifiée.

(3) Supposons que $\dim \varphi > 16$ et que φ_1 vérifie la condition (2.2). On aura besoin du résultat suivant.

Théorème 8. *(*[22]*)*
Soient φ une forme quadratique de dimension > 16, $K = F(\varphi)$ *et* $\psi \in I^2K$ *de dimension* 8 *telle que* $\operatorname{ind} c(\psi) = 2$*. On suppose que* $\psi \in \operatorname{Im}(W(F) \longrightarrow W(K))$*. Alors ψ est définie sur F.*

Ce théorème permet de déduire que φ_1 est définie sur F. D'après [20, Theorem 7.13], la forme φ est voisine, ceci est absurde.

3.4. DÉMONSTRATION DE LA PROPOSITION 3. On a que φ est bonne (évident pour $\deg(\varphi) = 1$ et c'est la proposition 9 pour $\deg(\varphi) = 2$). Soit τ la forme dominante de φ.
• Si $\deg(\varphi) = 1$, alors on obtient par le théorème 2(iv) que φ_1 est excellente de hauteur 2. Ainsi, $\varphi_{F(\varphi)(\tau)} \sim 0$.
• Si $\deg(\varphi) = 2$, alors on obtient par la proposition 9(3) que $\varphi_{F(\varphi)(\tau)} \sim 0$.

(i) D'après [19, Theorem 5.8] et si $\varphi_{F(\tau)}$ est anisotrope, alors $\varphi_{F(\tau)}$ est semblable à une forme de Pfister anisotrope, en particulier φ est de dimension une puissance de 2.

(ii) Lorsque $\varphi_{F(\tau)}$ est isotrope, on déduit que $\varphi_{F(\tau)} \sim 0$, et donc $\dim \varphi = 2^{\deg(\varphi)}k$ pour un certain entier k impair (Lemme 1).

On déduit les équivalences de la proposition en combinant (i) et (ii).

3.5. DÉMONSTRATION DU THÉORÈME 3. (1) $\Longrightarrow$ (2) D'après la proposition 3, la forme φ est bonne et que $\varphi_{F(\tau)}$ est semblable à une forme de Pfister anisotrope où $\tau \in P_dF$ est la forme dominante de φ. Par la proposition 7, il existe une forme $\pi \in GP_nF$ telle que $\varphi_{F(\tau)} \cong \pi_{F(\tau)}$. Les hypothèses du théorème impliquent que $n \geq d + 2$. Puisque $\varphi \perp -\pi \in \operatorname{Ker}(W(F) \longrightarrow W(F(\tau)))$, on obtient que $\varphi \perp -\pi \sim \tau \otimes \rho$ pour une forme ρ de dimension impaire (Lemme 1). Ainsi, $\varphi \equiv \tau \otimes \rho \pmod{J_n(F)}$. Par conséquent, $\varphi \in P^w_{n,d}F$ au sens de [10, Definition 3.4(ii)]. D'après [10, Corollary 3.7], on obtient que $i_W(\varphi_{F(\varphi)}) = 2^{d-1}$. Par conséquent, $\dim \varphi_1 = 2^n - 2^d$. D'après le théorème 2(iv) et la proposition 9, on déduit que φ_1 est voisine et que sa forme complémentaire est semblable à la forme $\tau_{F(\varphi)}$ qui est de dimension 2^d. Par conséquent, il existe $a \in F(\varphi)^*$ tel que $\varphi_1 \perp a(\tau_{F(\varphi)}) \in GP_nF(\varphi) \subset I^nF(\varphi) \subset I^{d+2}F(\varphi)$. D'autre part, $\varphi \perp k\tau \sim \pi \perp \tau \otimes (\rho \perp \langle k \rangle) \in I^{d+2}F$, avec $k \in F^*$ qui vérifie $\rho \perp \langle k \rangle \in I^2F$. Par conséquent, $a(\tau_{F(\varphi)}) \equiv k(\tau_{F(\varphi)}) \pmod{I^{d+2}F(\varphi)}$. Par le Hauptsatz d'Arason-Pfister, on obtient que $a(\tau_{F(\varphi)}) \cong k(\tau_{F(\varphi)})$. Ainsi, $\varphi_1 \perp k(\tau_{F(\varphi)}) \in GP_nF(\varphi)$. Par conséquent, $\varphi_2 = (\varphi_{F(\varphi)(\varphi_1)})_{an} = (-k\tau)_{F(\varphi)(\varphi_1)}$ est définie sur F. D'après [18, Proposition 3(iii)], on obtient que $\deg(\varphi \perp k\tau) = n$. Par conséquent, $\varphi \equiv -k\tau \pmod{J_n(F)}$.

(2) $\Longrightarrow$ (1) C'est une conséquence de la proposition 8(2).

3.6. DÉMONSTRATION DU THÉORÈME 4. On aura besoin du résultat suivant.

PROPOSITION 10. *(Rost)*
Soient φ et $\tau = \langle\langle a, b \rangle\rangle$ deux formes quadratiques anisotropes. Alors, on a équivalence entre:
(1) $i_W(\varphi_{F(\tau)}) \geq 2k$ et $\varphi_{F(\sqrt{a})} \sim 0$;
(2) Il existe deux formes quadratiques λ, γ telles que $\varphi \cong \langle\langle a, b \rangle\rangle \otimes \lambda \perp \langle\langle a \rangle\rangle \otimes \gamma$ avec $\dim \lambda = k$.

Démonstration. Le résultat a été prouvé par Rost [12, Lemma 2.6] lorsque $k = 1$. Le cas général se déduit par une simple récurrence sur k. ■

Démonstration du théorème 4. Soit $d = d_{\pm}\varphi$.

(3) $\Longrightarrow$ (1) Evident.

(2) $\Longrightarrow$ (3) D'après la proposition 3, on a $\varphi_{F(\sqrt{d})} \sim 0$. Ainsi, $\varphi \cong \langle\langle d\rangle\rangle \otimes \eta$ avec $\dim \eta \geq 4$. Par le théorème 2(2), on a $\dim \eta \leq 8$. Puisque la dimension de η est impaire (Lemme 1), on a $\dim \eta = 5$ ou 7. Si $\dim \eta = 5$, alors le théorème est démontré. Supposons que $\dim \eta = 7$. Soit $b \in F^*$ tel que $\langle\langle b\rangle\rangle$ soit semblable à une sous-forme de η. Alors, $\tau \cong \langle\langle d, b\rangle\rangle$ est semblable à une sous-forme de φ. Par conséquent, $\varphi_{F(\tau)}$ est isotrope et $i_W(\varphi_{F(\tau)}) \geq i_W(\varphi_{F(\varphi)}) = 4$. Par la proposition 10, $\varphi \cong \langle\langle d, b\rangle\rangle \otimes \lambda \perp \langle\langle d\rangle\rangle \otimes \gamma$ avec $\dim \lambda = 2$. On a bien que $\dim \gamma = 3$. Ecrivons $\gamma = \langle k\rangle \perp \mu$ avec $\dim \mu = 2$. Soit $\xi = \langle k\rangle \perp \langle\langle b\rangle\rangle \otimes \lambda$. On a $\varphi = \langle\langle d\rangle\rangle \otimes (\xi \perp \mu)$. Clairement, ξ est voisine de dimension 5. Ainsi, il existe une forme quadratique ξ' de dimension 3 telle que $\xi \perp \xi' \in GP_3F$. On peut supposer que $1 \in D_F(\xi')$, et donc $\xi \perp \xi' \in P_3F$. Par conséquent, $\pi := \langle\langle d\rangle\rangle \otimes (\xi \perp \xi') \in P_4F$. On a $\varphi \perp \nu \sim \pi$, où $\nu = \langle\langle d\rangle\rangle \otimes (-\mu \perp \xi')$. Clairement, $\dim \nu = 10$. Si ν est isotrope, alors $\dim(\pi_{F(\varphi)})_{an} = \dim((\varphi \perp \nu)_{F(\varphi)})_{an} \leq 6 + 8 < 16$. Ainsi, $\pi_{F(\varphi)} \sim 0$, et donc φ est voisine, ceci est absurde. Si ν est anisotrope, alors ν admet la propriété de déploiement maximal (maximal splitting property; voir [13, Theorem 5.1]). Puisque $\varphi \perp \nu \sim \pi \in P_4F$, on obtient que $\varphi_{F(\pi)} \sim -\nu_{F(\pi)}$. Puisque $\dim \nu < \dim \varphi$, la forme $\varphi_{F(\pi)}$ est isotrope. On a $\dim(\nu_{F(\pi)})_{an} = \dim(\varphi_{F(\pi)})_{an} \leq \dim(\varphi_{F(\varphi)})_{an} = 6$. Ainsi, $\nu_{F(\pi)}$ est isotrope. D'après [9, Corollary 3], ν est voisine de π. Puisque $1 \in D_F(\nu)$, on a $\nu \subset \pi$. Ainsi, $\dim(\pi \perp -\nu)_{an} = 16 - 10 = 6$. Puisque $\varphi \sim \pi \perp -\nu$, on obtient que $\dim \varphi_{an} = 6$, ceci est absurde.

(1) $\Longrightarrow$ (2) Comme dans le début de la preuve de l'implication (2) $\Longrightarrow$ (3), on obtient que $\varphi \cong \langle\langle d\rangle\rangle \otimes \eta$ pour η de dimension 5 ou 7. Par le théorème 2(iv), la forme quadratique φ_1 est excellente de hauteur 2 et de degré 1. Ainsi, $\dim \varphi_1 = 2^n - 2$ pour un certain entier $n \geq 3$. Puisque $\dim \varphi_1 \leq \dim \varphi - 2 \leq 2 \cdot 7 - 2 = 12$, on déduit que $n = 3$ et donc $\dim \varphi_1 = 6$.

3.7. Démonstration du théorème 5. On commence par un lemme.

Lemme 3. *Soient $\varphi \in I^2F$ anisotrope, $\tau \in P_2F - \{0\}$ telles que $\dim(\varphi_{F(\varphi)})_{an} = 8$ et $c(\varphi) = c(\tau)$. Alors, on a les assertions suivantes:*
(1) Si $\varphi_{F(\tau)}$ est isotrope, alors il existe $\eta \in GP_3F$, $r \in F^*$ tels que $\varphi \perp \eta \perp r\tau \in I^4F$. Si de plus, $\dim \varphi \in \{14, 16\}$, alors $\varphi_{F(\eta)}$ est isotrope.
(2) Supposons que $\varphi_{F(\tau)}$ soit anisotrope de hauteur ≥ 2, $\dim \varphi = 16$ et qu'il existe $\delta \in GP_3F$, $s \in F^*$ tels que $\varphi \perp \delta \perp s\tau \in I^4F$. Alors, $\varphi_{F(\delta)}$ est anisotrope.

Démonstration. (1) Supposons que $\varphi_{F(\tau)}$ soit isotrope. On obtient que $i_W(\varphi_{F(\tau)}) \geq i_W(\varphi_{F(\varphi)})$. Puisque $F(\tau)/F$ est excellente et que $\dim(\varphi_{F(\varphi)})_{an} = 8$, il existe η une forme de dimension 8 telle que $\varphi_{F(\tau)} \sim -\eta_{F(\tau)}$. Puisque $c(\eta)_{F(\tau)} = 0$, on peut supposer par la proposition 7 que $\eta \in GP_3F$. Puisque $\varphi \perp \eta \in \mathrm{Ker}(W(F) \longrightarrow W(F(\tau)))$, on déduit que $\varphi \perp \eta \sim \rho \otimes \tau$ pour ρ une forme de dimension impaire (Lemme 1). Soit $r \in F^*$ tel que $\rho \perp \langle r \rangle \in I^2F$. Ainsi,

$$\varphi \perp \eta \perp r\tau \in I^4F \tag{1}$$

Supposons, de plus, que $\dim \varphi \in \{14, 16\}$. On déduit par l'équation (1) et le Hauptsatz d'Arason-Pfister que $\varphi_{F(\tau)(\eta)} \sim 0$. Si la forme $\varphi_{F(\eta)}$ est anisotrope, on obtient que $\varphi_{F(\eta)} \cong \tau \otimes \gamma$ pour γ une forme de dimension 4. Ainsi, $\dim \varphi = 16$ et $c(\varphi)_{F(\eta)} = 0$. Par le théorème de réduction d'indice ([31], [25]), on déduit que $c(\varphi) = 0$, ceci est absurde. Ainsi, $\varphi_{F(\eta)}$ est isotrope.

(2) Supposons que $\varphi_{F(\tau)}$ soit anisotrope de hauteur ≥ 2, $\dim \varphi = 16$ et qu'il existe $\delta \in GP_3F$, $s \in F^*$ tels que $\varphi \perp \delta \perp s\tau \in I^4F$. Si $\varphi_{F(\delta)}$ est isotrope, alors $i_W(\varphi_{F(\delta)}) \geq i_W(\varphi_{F(\varphi)}) = 4$. Par le Hauptsatz d'Arason-Pfister $\varphi_{F(\delta)(\tau)} \sim 0$. Puisque $\varphi_{F(\tau)}$ est anisotrope, on déduit que $\varphi_{F(\tau)} \cong \delta \otimes \rho$ pour ρ une forme de dimension 2. Ainsi, $\varphi_{F(\tau)} \in GP_4F(\tau) - \{0\}$ et donc elle est de hauteur 1, ceci est absurde. ■

Démonstration du théorème 5. Supposons que la conjecture 3 soit vraie. Soit φ une forme quadratique de dimension 12 qui n'est pas voisine.

(1) $\Longrightarrow$ (2) Supposons que φ soit faiblement bonne, de hauteur 3 et de degré 2. Soit $\tau \in P_2F$ la forme dominante de φ. On a $\varphi_{F(\tau)} \not\sim 0$ car sinon φ serait divisible par τ et donc serait une voisine. Par la proposition 9, on a $\dim \varphi_1 = 8$ et $c(\varphi) = c(\tau)$. Puisque $\varphi_{F(\tau)}$ est isotrope, on déduit par le lemme 3 l'existence de $\eta \in GP_3F$, $r \in F^*$ tels que

$$\varphi \perp \eta \perp r\tau \in I^4F \tag{2}$$

Par le Hauptsatz d'Arason-Pfister et l'équation (2), on déduit que $(\varphi_1 \perp r\tau)_{F(\varphi)(\eta)} \sim 0$. Par conséquent $\varphi_1 \perp (r\tau)_{F(\varphi)} \sim \lambda\eta$ pour $\lambda \in F(\varphi)^*$ convenable. Puisque $\dim \varphi_1 = 8$, on obtient $i_W((-r\tau)_{F(\varphi)} \perp \lambda\eta) \geq 2$. Par la multiplicativité d'une forme de Pfister, il existe $\alpha \in F(\varphi)^*$ tel que $(-r\tau)_{F(\varphi)} \perp \lambda\eta \sim \alpha(\eta' \perp -\tau')_{F(\varphi)}$ ([5, Theorem 4.5], [10, Lemma 3.2]). Puisque $i_W((-r\tau)_{F(\varphi)} \perp \lambda\eta) \geq 2$, on déduit que $(\eta' \perp -\tau')_{F(\varphi)}$ est isotrope. Par la conjecture 3, $\eta' \perp -\tau'$ est isotrope. Soit $\delta := -r(\eta' \perp -\tau')_{an} \in I^2F$. On affirme que $\dim \delta = 8$, car sinon $\dim \delta = 6$ ou 4. Or si $\dim \delta = 6$, alors δ est une forme d'Albert anisotrope et donc $\mathrm{ind}\, c(\delta) = 4$, ceci contredit l'hypothèse $c(\delta) = c(\tau)$. Si $\dim \delta = 4$, alors en remplaçant dans l'équation (2) η par $-r\eta$, on obtient $\varphi \perp \delta \in GP_4F$, et donc φ est voisine, ceci est absurde.

(2) $\Longrightarrow$ (1) D'après [20, Example 9.12], il existe $a \in F^*$ tel que $\delta_{F(\sqrt{a})} \sim 0$. Puisque $\varphi \perp \delta \in I^4F$, on obtient par le Hauptsatz d'Arason-Pfister que

$\varphi_{F(\sqrt{a})} \sim 0$. Par conséquent, pour toute extension L/F, on obtient que $(\varphi_L)_{an} \cong \langle\langle a\rangle\rangle \otimes \mu$ pour une certaine L-forme quadratique μ de dimension ≤ 6. Puisque $\varphi \in I^2F$, on déduit que la dimension de μ est paire. Ainsi, $\dim(\varphi_L)_{an} \in \{0, 4, 8, 12\}$. La classification des formes quadratiques de hauteur et de degré ≤ 2 implique que $\mathrm{h}(\varphi) \neq 1, 2$. Par conséquent, φ est de hauteur 3 et de degré 2. Soit $\tau \in P_2F$ telle que $c(\tau) = c(\delta)$. Par la proposition 9, φ est bonne de forme dominante τ. Puisque $\delta_{F(\tau)} \in GP_3F(\tau)$, on obtient que $\delta_{F(\tau)(\delta)} \sim 0$. Par le Hauptsatz d'Arason-Pfister on déduit que $\varphi_{F(\tau)(\delta)} \sim 0$. Ainsi, $\varphi_{F(\tau)}$ est isotrope, i.e. φ est faiblement bonne. ■

3.8. Démonstration de la proposition 6. D'après la proposition 9, la forme φ est bonne et φ_1 est de dimension 8 ou est excellente de dimension 12, avec $c(\varphi) = c(\tau)$ où $\tau \in P_2F$ est la forme dominante de φ.

(1) Supposons que $\varphi_{F(\tau)}$ soit isotrope.

- Supposons que φ_1 soit excellente de dimension 12. Par la proposition 9, $\varphi_{F(\tau)(\varphi)} \sim 0$. Ainsi, $\varphi \in \mathrm{Ker}(W(F) \longrightarrow W(F(\tau)))$. Si $\dim\varphi = 14$, alors φ est isotrope, ceci est absurde. Si $\dim\varphi = 16$, on déduit que φ est divisible par τ et donc $c(\varphi) = 0$, ceci est absurde. Ainsi, φ_1 ne peut être une forme excellente de dimension 12.
- Supposons $\dim\varphi_1 = 8$: Par le lemme 3, il existe $r \in F^*$, $\eta \in GP_3F$ tels que $\varphi \perp \eta \perp r\tau \in I^4F$ et $\varphi_{F(\eta)}$ isotrope. Ainsi, $i_W(\varphi_{F(\eta)}) \geq i_W(\varphi_{F(\varphi)})$. Par le Hauptsatz d'Arason-Pfister, $(\varphi \perp r\tau)_{F(\eta)} \sim 0$. Ainsi, $\varphi \perp r\tau \sim \rho \otimes \eta$ pour ρ de dimension 2. En particulier, $\varphi \perp r\tau \in I^4F$. De nouveau par le Hauptsatz d'Arason-Pfister, $\varphi_{F(\tau)} \sim 0$. Si $\dim\varphi = 14$ (resp. $\dim\varphi = 16$), on obtient que φ est isotrope (resp. φ est divisible par τ), ceci est absurde car φ est anisotrope (resp. car $c(\varphi) \neq 0$).

Ainsi, dans tous les cas φ est fortement bonne.

(2)

- Si $\dim\varphi_1 = 12$, alors $\varphi_{F(\varphi)(\tau)} \sim 0$. Puisque $\varphi_{F(\tau)}$ est anisotrope, on déduit que $\varphi_{F(\tau)}$ est de hauteur 1. Dans ce cas, $\dim\varphi = 16$.
- Si $\dim\varphi_1 = 8$, alors d'après [21, Théorème 4], $(\varphi_1)_{F(\varphi)(\tau)} \in GP_3F(\varphi)(\tau) - \{0\}$. Par conséquent, $\varphi_{F(\tau)}$ est de hauteur 2 et de degré 3.

3.9. Démonstration du théorème 6. Soient φ une forme quadratique anisotrope qui n'est pas voisine, de hauteur 3 et de degré 2 et $\tau \in P_2F$ sa forme dominante. Supposons que l'hypothèse suivante, *appelée hypothèse* (H), soit vraie:

Hypothèse (H): *Toute forme quadratique anisotrope qui n'est pas bonne, de hauteur* 2 *et de degré* 3 *est de dimension* 12.

(1) Supposons que $\dim\varphi = 14$. Par la proposition 6, on a que $\varphi_{F(\tau)}$ est anisotrope de hauteur 2 et de degré 3. D'après l'hypothèse (H), $\varphi_{F(\tau)}$ est bonne. Ceci est absurde d'après un résultat de Hurrelbrink et Rehmann sur la dimension des formes quadratiques bonnes de hauteur 2 et de degré 3 [15, 3.4]. Par conséquent, il n'existe pas de forme quadratique anisotrope de hauteur 3, de degré 2 et de dimension 14.

(2) Supposons que $\dim\varphi = 16$. Par la proposition 9, on a que $\dim\varphi_1 = 8$ ou φ_1 est excellente de dimension 12.

- Supposons que $\dim\varphi_1 = 12$. D'après la démonstration de la proposition 6, on a $\varphi_{F(\tau)} \in GP_4F(\tau) - \{0\}$. D'après la proposition 7, il existe $\eta \in GP_4F$ telle que $\varphi_{F(\tau)} \cong \eta_{F(\tau)}$. Ainsi, $\varphi \perp -\eta \sim \rho \otimes \tau$ pour ρ une forme quadratique de dimension impaire (Lemme 1). Par conséquent, $\varphi \perp b\tau \in I^4F$ avec $b \in F^*$ qui vérifie $\rho \perp \langle b\rangle \in I^2F$. D'après [12] et puisque $\dim(\varphi \perp b\tau)_{an} \leq 20$, on obtient que $\varphi \in GP_{4,2}F$.

- Supposons que $\dim\varphi_1 = 8$. Toujours par la démonstration de la proposition 6, on a que $\varphi_{F(\tau)}$ est de hauteur 2 et de degré 3. Soit $\pi \in P_3F(\tau)(\varphi)$ la forme dominante de $\varphi_{F(\tau)}$. D'après l'hypothèse (H), π est définie sur $F(\tau)$. D'après [20, Theorem 9.6], on déduit que $\varphi_{F(\tau)} \equiv \pi \pmod{J_4(F(\tau))}$. D'après [18, Proposition 4], π est définie sur F par une forme de Pfister. Par conséquent, $e^3(\varphi \perp -\pi \perp \tau) \in \mathrm{Ker}(H^3F \longrightarrow H^3F(\tau))$. D'après [1, Satz 5.5], il existe $c \in F^*$ tel que $e^3(\varphi \perp -\pi \perp \tau) = e^3(\tau \perp -c\tau)$. Par la bijectivité de e^3, on a
$$\varphi \perp -\pi \perp c\tau \in I^4F \tag{3}$$
Par le lemme 3, $\varphi_{F(\pi)}$ est anisotrope. Par l'équation (3), on a $\varphi_{F(\pi)} \equiv -c\tau_{F(\pi)} \pmod{I^4F(\pi)}$. D'après [10, Proposition 3.6], on obtient que $i_W(\varphi_{F(\pi)(\varphi)}) = 2$. Ceci contredit la condition $i_W(\varphi_{F(\varphi)}) = 4$.

Réciproquement si $\varphi \in GP_{4,2}F$, alors on déduit que φ est de hauteur 3 et de degré 2 (Proposition 8(2)).

4. Bibliographie

[1] *J. Kr. Arason*, Cohomologische Invarianten quadratischer Formen. J. Alg. 36 (1975), 448–491.

[2] *J. Kr. Arason*, Excellence of $F(\varphi)/F$ for 2-fold Pfister forms. Appendice de [6].

[3] *J. Kr. Arason,* A proof of Merkur'ev's theorem. Can. Math. Soc. Conf. Proc. 4 (1984), 121–130.

[4] *J. Kr. Arason, R. Elman et W. Jacob,* The graded Witt ring and Galois cohomology, II. Trans. Amer. Math. Soc. 314 (1989), 745–780.

[5] *R. Elman et T. Y. Lam,* Pfister forms and K-theory of fields. J. Alg. 23 (1972), 181–213.

[6] *R. Elman, T. Y. Lam et A. Wadsworth,* Amenable fields and Pfister extensions. Conf. Quadratic forms (1976) (G. Orzech, ed.). Queen's papers

on Pure and Appl. Math. Queen's Univ. Kingston, Ont., 46 (1977), 445–492.

[7] *R. W. Fitzgerald,* Functions fields of quadratic forms. Math. Z. 178 (1981), 63–73.

[8] *R. W. Fitzgerald,* Quadratic forms of height two. Trans. Amer. Math. Soc. 283 (1984), 339–351.

[9] *D. W. Hoffmann,* Isotropy of quadratic forms over the function field of a quadric. Math. Z. 220 (1995), 461–476.

[10] *D. W. Hoffmann,* Twisted Pfister forms. Doc. Math. J. DMV 1 (1996), 67–102.

[11] *D. W. Hoffmann,* Sur les dimensions des formes quadratiques de hauteur 2. C. R. Acad. Sci. Paris 324 (1997), 11–14.

[12] *D. W. Hoffmann,* On the dimensions of anisotropic forms in I^4. Invent. Math. 131 (1998), 184–198.

[13] *D. W. Hoffmann,* Splitting patterns and invariants of quadratic forms. Math. Nachr. 190 (1998), 149–168.

[14] *J. Hurrelbrink et U. Rehmann,* Splitting patterns of excellent quadratic forms. J. reine angew. Math. 444 (1993), 183–192.

[15] *J. Hurrelbrink et U. Rehmann,* Splitting patterns of quadratic forms. Math. Nachr. 176 (1995), 111–127.

[16] *W. Jacob et M. Rost,* Degree four cohomological invariants for quadratic forms. Invent. Math. 96 (1989), 551–570.

[17] *B. Kahn,* Formes quadratiques de hauteur et de degré 2. Indag. Math. 7.1 (1996), 47–66.

[18] *B. Kahn,* A descent problem for quadratic forms. Duke Math. J. 80 (1995), 139–159.

[19] *M. Knebusch,* Generic splitting of quadratic forms I. Proc. London. Math. Soc. 33 (1976), 65–93.

[20] *M. Knebusch,* Generic splitting of quadratic forms II. Proc. London. Math. Soc. 34 (1977), 1–31.

[21] *A. Laghribi,* Isotropie de certaines formes quadratiques de dimensions 7 et 8 sur le corps des fonctions d'une quadrique. Duke Math. J. 85.2 (1996), 397–410.

[22] *A. Laghribi,* Sur le problème de descente des formes quadratiques. A paraître aux Arch. Math.

[23] *T. Y. Lam,* The algebraic theory of quadratic forms. (2^e édition,) Benjamin, New York, 1980.

[24] *A. S. Merkurjev,* Le symbole de résidu normique de degré 2. En russe, Dokl. Akad. Nauk SSSR 261 (1981), 542–547. Traduction anglaise : Soviet Math. Doklady 24 (1981), 546–551.

[25] *A. S. Merkurjev,* Algèbres simples et formes quadratiques. En russe, Izv. Akad. Nauk SSSR 55 (1991), 218–224. Traduction anglaise : Math. USSR Izv. 38 (1992), 215–221.

[26] *A. S. Merkurjev et A. A. Suslin,* L'homomorphisme de résidu normique de degré 3. En russe, Izv. Akad. Nauk SSSR 54 (1990), 339–356. Traduction anglaise : Math. USSR Izv. 36 (1991), 349–367.

[27] *D. Orlov, A. Vishik et V. Voevodsky,* Motivic cohomology of Pfister quadrics and Milnor's conjecture on quadratic forms. Prépublication.

[28] *M. Rost,* Hilbert's theorem 90 for K_3^M for degree-two extensions. Prépublication, Regensburg, 1986.

[29] *W. Scharlau,* Quadratic and Hermitian forms. Springer, Berlin, 1985.

[30] *M. Szyjewski,* The fifth invariant of quadratic forms. En russe, Algebra i Analiz. 2 (1990), 213–234. Traduction anglaise: St. Petersburg Math. J. 2 (1991), 179–198.

[31] *J.-P. Tignol,* Réduction de l'indice d'une algèbre simple centrale sur le corps des fonctions d'une quadrique. Bull. Soc. Math. Belgique 42 série A (1990), 735–745.

[32] *A. R. Wadsworth,* Noetherian pairs and function fields of quadratic forms. Thèse, Université de Chicago, 1972.

Ahmed Laghribi
Faculté Jean Perrin
Rue Jean Souvraz - SP 18
F-62307 Lens Cedex
France
laghribi@euler.univ-artois.fr

Théorie d'Iwasawa et Loi Explicite de Réciprocité

Un Remake d'un Article de P. Colmez

Bernadette Perrin-Riou

Received: October 1, 1998
Revised: May 12, 1999

Communicated by Peter Schneider

Abstract. Let V be a crystalline p-adic representation of the absolute Galois group of $\mathbb{Q}_p$. The author has built the Iwasawa theory of such a representation in Invent. Math (1994) and conjectured a reciprocity law which has been proved by P. Colmez. In this text, we write the initial construction with simplification and the proof of P. Colmez in a different language. This point of view will allow us to study the universal norms in the geometric cohomology classes associated to V by Bloch and Kato in a forthcoming article.

1991 Mathematics Subject Classification: 11E95 11R23
Keywords and Phrases: p-adic representation, Iwasawa theory, exponential, reciprocity law

La loi explicite de réciprocité classique sur un corps local remonte à Artin-Hasse et Iwasawa et donne une description du symbole de Hilbert. Elle a été généralisée à des modules de Lubin-Tate, citons Wiles, Kolyvagin, Vostokov, Brückner, Coleman, Sen, de Shalit, Fesenko. On renvoie à [3] pour un historique. Le développement de ces lois s'est fait en parallèle et en liaison avec le développement de la théorie d'Iwasawa locale ; dans le cas classique, il s'agit de l'étude du comportement des unités locales sur la $\mathbb{Z}_p^\times$-extension cyclotomique K_∞ à l'aide de l'application exponentielle (Iwasawa, Coates-Wiles, Coleman). On peut envisager des généralisations de la loi de réciprocité à des représentations cristallines quelconques. Dans [4], nous avons donné une généralisation de cette étude des unités locales à des représentations cristallines V du groupe de Galois de $\mathbb{Q}_p$ générales : les unités locales sont remplacées par la limite projective $Z^1_\infty(\mathbb{Q}_p, T)$ des groupes de cohomologie galoisiennes $H^1(\mathbb{Q}_p(\mu_{p^n}), T)$ et on construit une application "exponentielle" Ω_V d'un $\mathbb{Q}_p \otimes \mathbb{Z}_p[[G_\infty]]$-module libre $\mathbb{Z}_p[[G_\infty]] \otimes_{\mathbb{Z}_p} \mathbf{D}_p(V)$

dans $\mathcal{H}(G_\infty) \otimes_{\mathbb{Z}_p[[G_\infty]]} Z^1_\infty(\mathbb{Q}_p, T)$ où $\mathbb{Z}_p[[G_\infty]]$ est l'algèbre d'Iwasawa de $G_\infty = \mathrm{Gal}(K_\infty, \mathbb{Q}_p)$, $\mathcal{H}(G_\infty)$ une algèbre de "séries formelles" avec condition de croissance, contenant $\mathbb{Z}_p[[G_\infty]]$ et $\mathbf{D}_p(V)$ le module de Fontaine associé à V. On a alors conjecturé dans ce cadre une loi explicite de réciprocité. On peut en donner deux formulations : la première (appelée $\mathrm{Réc}(V)$) dit essentiellement que pour les dualités naturelles, Ω_V et $\Omega_{V^*(1)}$ sont adjoints (ici $V^*(1)$ est le dual de Tate de V). La deuxième formulation ([6]) plus proche de la formulation traditionnelle calcule à un niveau fini (c'est-à-dire sur le corps $\mathbb{Q}_p(\mu_{p^n})$) l'application duale de l'exponentielle sur $\Omega_{V(k)}(g)$ en termes de g pour des twists à la Tate $V(k)$ convenables de V. Il n'est pas difficile de voir que les deux formulations sont équivalentes.
Cette loi vient d'être montrée par P. Colmez ([1]) et indépendemment par Kato, Kurihara, Tsuji. Plus récemment, D. Benois en a aussi donné une démonstration en utilisant la théorie des (φ, Γ)-modules de Fontaine.
Nous reprenons dans ce texte la démonstration de Colmez de la loi explicite de réciprocité pour une représentation cristalline (ou un tout petit peu plus généralement pour la partie cristalline de son module filtré). La présentation est très légèrement différente : outre que nous n'utilisons pas le langage des distributions, nous commençons par démontrer la loi explicite de réciprocité puis nous voyons la construction (un peu modifiée) de son application $\mathrm{Log}_V^{(h)}$ (DANS LE CAS OÙ V EST CRISTALLINE) comme une conséquence de cette loi, ce qui permet d'utiliser des arguments sur l'anneau B_{cris} moins subtils que les siens. Cependant, comme il a été souligné, nous ne regardons ici que la partie cristalline du module filtré associé à V, ce qui nous permet de travailler uniquement avec des fonctions analytiques. Dans [1], il est fait plus mais cette généralisation très importante est encore mal comprise (par moi en tout cas) : il semble en effet qu'il faille abandonner l'idée de raisonner avec de bonnes vieilles fonctions (ou distributions sur $\mathbb{Z}_p$). La justification de ce texte est peut-être qu'avant de sauter ce pas, nous voulions faire le point sur le cas cristallin, dans le langage "usuel". La démonstration de Colmez est alors extrêmement simple et naturelle : donnons-en les ingrédients. Si u est un générateur topologique de $1 + p\mathbb{Z}_p$, il s'agit de calculer la valeur d'une fonction analytique f sur le disque unité de $\mathbb{C}_p$ en $u^k - 1$, la fonction f étant obtenue par interpolation de nombres en h familles de points de la forme $\zeta u^j - 1$ pour ζ racine de l'unité d'ordre une puissance de p et $0 \leq j < h$; bien sûr k est différent de ces j. Il y a pour cela une formule générale qui exprime $f(u^k - 1)$ comme une limite de combinaisons linéaires des $f(\zeta u^j - 1)$. Plus précisément, on a par exemple

$$\frac{(-1)^h (h-1)!}{k(k-1)\dots(k-h+1)} \frac{f(u^k-1)}{\log u} = \lim_{n\to\infty} \sum_{i=0}^{h-1} (-1)^i \frac{p^n}{1 - u^{(k-i)p^n}} \binom{h-1}{i} R_{n,i}(f)(u^{k-i} - 1)$$

où $R_{n,i}(f)$ est le polynôme de degré $< p^n$ tel que

$$f \equiv R_{n,i}(f)(u^{-i}(1+T)-1) \mod u^{-ip^n}(1+T)^{p^n}-1 \ .$$

Maintenant, si Γ est le groupe de Galois de $\mathbb{Q}_p(\mu_{p^\infty})/\mathbb{Q}(\mu_p)$ de générateur topologique γ et si $u=\chi(\gamma)$ avec χ le caractère cyclotomique, par définition de la fonction f dont on veut calculer la valeur en u^k-1, $R_{n,i}(f)(u^{k-i}-1)$ est relié à la valeur en $\gamma_n=\gamma^{p^{n-1}}$ d'un cocycle de Γ_n à valeurs dans V : ce cocycle devient un cobord $(\gamma_n-1)c_n$ lorsqu'on étend les scalaires à $B_{\mathrm{cris}}^{G_{\mathbb{Q}(\mu_{p^\infty})}}$. On relie ainsi $\frac{R_{n,i}(f)(u^{k-i}-1)}{1-u^{(k-i)p^n}}=\chi^k(\frac{R_{n,i}(f)(u^{-i}\gamma-1)}{1-u^{-i}\gamma_n^k})$ à l'image de c_n dans $\mathbb{Q}_p(\mu_{p^n})$ par l'application $\lambda_{k,n}: B_{\mathrm{dR}}^{G_{\mathbb{Q}(\mu_{p^\infty})}} \to B_{\mathrm{dR}}/(\chi^{-k}(\gamma_n)\gamma_n-1) \to \mathbb{Q}_p(\mu_{p^n})$ (application de Tate). On peut relier ce dernier élément à l'application exponentielle duale grâce à une formule due à Kato.

Nous avons ensuite donné quelques conséquences de cette loi. Certaines sont déjà dans des articles antérieurs ([4], [7]). D'autres sont plus nouvelles.
Dans les §1 et 2, nous faisons quelques préliminaires et rappels : théorie d'Iwasawa locale, lemme fondamental d'interpolation des fonctions analytiques, résolution d'équations du type $(1-\varphi)G=g$. Dans le §3, nous reprenons complètement la construction de l'application exponentielle $\Omega_{V,h}$ faite dans [4] en tenant compte des points de $H^1_f(K_n,V)$. Nous donnons dans le §4 la démonstration due à Colmez de la loi explicite de récipocité. Dans le §5, se trouvent des conséquences de cette loi explicite (anciennement conjecture Réc(V)) et des calculs sur les valeurs spéciales de l'application logarithme que l'on peut en grande partie déjà trouvés dans [4], [5] et [7]. On espère ainsi donner un panorama complet des formules que l'on a à sa disposition. Ces formules sont très utiles dans la théorie des fonctions L p-adiques comme cela a déjà beaucoup exploité dans [5] et [4]. Dans l'appendice A, on donne quelques formules relatives au lemme de Shapiro, aux opérations de twist et de projections puis reliant différentes manières de voir les fonctions analytiques. Dans l'appendice B, on démontre la formule exprimant la valeur de $f(u^k-1)$ en termes des valeurs aux points d'interpolation pour une fonction analytique f d'ordre fini. Dans l'appendice C, on reprend la suite exacte de Coleman-Colmez en modifiant légèrement la présentation de Colmez.

Errata. Une erreur dans [4] m'a été signalée par J. Nekovář. La plupart des résultats ne sont valables que lorsque H est une extension finie de $\mathbb{Q}_p$, car on utilise à divers endroits l'accouplement local de dualité. Ainsi, il n'y a en particulier pas de résultats nouveaux sur les représentations p-adiques ordinaires sur un corps local dont le corps résiduel n'est pas fini dans [4].

Table des matières

1. Préliminaires

1.1. On pose $K = \mathbb{Q}_p$. On fixe une clôture algébrique $\overline{K}$ de K et on note $G_L = \mathrm{Gal}(\overline{K}/L)$ pour toute extension algébrique L de K (supposée contenue dans $\overline{K}$). On pose $K_n = K(\mu_{p^n})$ où μ_{p^n} est le groupe des racines p^n-ièmes de l'unité. On fixe dans tout le texte un système ϵ de racines de l'unité ζ_n d'ordre p^n vérifiant $\zeta_{n+1}^p = \zeta_n$. On note G_n le groupe de Galois de K_n/K pour $n \in \mathbb{N}\cup\{\infty\}$ et $\Lambda = \mathbb{Z}_p[[G_\infty]]$. On note χ le caractère cyclotomique $G_K \to \mathbb{Z}_p^\times$. On désigne par γ un générateur topologique de $\Gamma = \mathrm{Gal}(K_\infty/K_1)$ et on pose $\gamma_n = \gamma^{p^{n-1}}$ pour $n \geq 1$, c'est un générateur topologique de $\mathrm{Gal}(K_\infty/K_n)$. On pose $\Delta = \mathrm{Gal}(K(\mu_p)/K) = \mathrm{Gal}(K_1/K)$. Tous les groupes de cohomologie galoisienne considérés sont les groupes de cohomologie continue.

1.2. Soit $\mathcal{H}$ l'algèbre des séries formelles en une variable convergeant sur le disque unité $\{x \in \mathbb{C}_p \text{ tel que } |x| < 1\}$ où $\mathbb{C}_p$ est le complété p-adique de $\overline{\mathbb{Q}_p}$. Si ρ est un réel inférieur à 1, on note $||f||_\rho = \sup_{|x|=\rho}|f(x)| = \sup_{|x|\leq\rho}|f(x)|$.
On pose $\rho_n = p^{-\frac{1}{p^n(p-1)}}$. Il est commode d'introduire $\bar{\mathbb{R}} = \mathbb{R}\cup\{r^- \text{pour } r \in \mathbb{R}\}$ avec l'ordre total : si $r_1 < r_2$, alors $r_1 < r_2^- < r_2$. Pour $r \in \mathbb{R}^-$, on note $\lfloor r\rfloor$ le plus grand entier inférieur ou égal à r. Si r est entier, on a donc $\lfloor r^-\rfloor = r-1$. Si $r \in \bar{\mathbb{R}}$, on note $\mathcal{H}_r$ le sous-$\mathbb{Q}_p$-espace vectoriel de $\mathcal{H}$ formé des séries F telles que la suite $\frac{||F||_{\rho_n}}{p^{nr}}$ est bornée si $r \in \mathbb{R}$ et tend vers 0 si $r \in \mathbb{R}^-$. On dit que F est tempérée d'ordre $\leq r$. Si $r_1 < r_2$ dans $\bar{\mathbb{R}}$, on a $\mathcal{H}_{r_1} \subset \mathcal{H}_{r_2}$. Plus précisément, si F appartient à $\mathcal{H}_r$, la suite $\frac{||F||_{\rho_n}}{p^{nr'}}$ tend vers 0 lorsque n tend vers l'infini pour tout $r' < r$. On note $\mathcal{H}_\infty$ la réunion des $\mathcal{H}_r$. Si $g \in \mathcal{H}_\infty$, on note $\mathfrak{o}(g)$ (resp. $\mathfrak{O}(g)$) la borne inférieure des $r \in \mathbb{R}$ tel que $g \in \mathcal{H}_{r^-}$ (resp. le plus petit réel r tel que $g \in \mathcal{H}_r$ s'il existe). Pour $r \in \bar{\mathbb{R}}\cup\{\infty\}$, on note $\mathcal{H}_r(\Gamma)$ les éléments de $\mathbb{Q}_p[[\Gamma]]$ de la forme $f(\gamma-1)$ avec $f \in \mathcal{H}_r$, $\mathcal{H}_r(G_\infty) = \mathbb{Z}_p[\mathrm{Gal}(K(\mu_p)/K)] \otimes \mathcal{H}_r(\Gamma)$ et $\mathcal{H}(G_\infty)$ la réunion des $\mathcal{H}_r(G_\infty)$.
On munit $\mathcal{H}_r$ de la norme C_r définie par $C_r(F) = \sup_n \frac{||F||_{\rho_n}}{p^{nr}}$ et $\mathcal{H}_r(G_\infty)$ de la norme qui s'en déduit.

1.3. Si $g \in \mathbb{Q}_p[[T]]$, on pose $D(g) = (1+T)\frac{d}{dT}g$. On pose $\varphi(g) = g((1+T)^p-1)$ et on note ψ l'opérateur de $\mathbb{Q}_p[[T]]$ tel que $\varphi\circ\psi(g) = p^{-1}\sum_{\zeta\in\mu_p} f(\zeta(1+T)-1)$. On peut aussi voir $p\psi$ comme la trace de l'extension $\mathbb{Q}_p[[T]]/\varphi\mathbb{Q}_p[[T]]$.
Il est important de rappeler qu'on dispose d'un isomorphisme canonique d'espaces vectoriels normés entre $\mathcal{H}_r^{\Psi=0}$ et $\mathcal{H}_r(G_\infty)$. Si $\tau \in G_\infty$, on pose $\tau.(1+T) = (1+T)^{\chi(\tau)}$ et on prolonge cette action à Λ par continuité. Pour la prolonger à $\mathcal{H}_r(G_\infty)$, on montre que si $f_{n,r}$ est le polynôme d'approximation de f modulo $\prod_{i=0}^{\lfloor r\rfloor)}(\chi^{-i}(\gamma_n)\gamma_n - 1)$, la suite $f_{n,r}.(1+T)$ converge dans $\mathcal{H}_r$ et ne dépend pas des choix ; c'est par définition $f.(1+T)$. L'opérateur D sur $\mathcal{H}_r^{\Psi=0}$ correspond sur $\mathcal{H}_r(G_\infty)$ à l'opération de twist $\tau\mapsto\chi(\tau)\tau$ et est un isomorphisme topologique de $\mathcal{H}_r^{\Psi=0}$.

Soit $\mathcal{D}$ un espace vectoriel de dimension finie muni d'une norme fixée. On définit alors naturellement $||f||_\rho$ pour $f \in \mathcal{H}\otimes\mathcal{D}$.

1.4. Définition : Soit u un automorphisme de $\mathcal{D}$. On dit qu'un élément $F \in \mathcal{H}\otimes\mathcal{D}$ est u-borné (resp. u^--borné) si la suite $||(1\otimes u)^{-n}F||_{\rho_n}$ est bornée (resp. tend vers 0) lorsque n tend vers l'infini.

On note $(\mathcal{H}\otimes\mathcal{D})_{u^\epsilon}$ l'espace vectoriel des éléments u^ϵ-bornés (il est contenu dans $\mathcal{H}_\infty\otimes\mathcal{D}$) et on pose alors $C_u(F) = \sup_n(||(1\otimes u)^{-n}F||_{\rho_n})$.

Remarques : 1) Prenons $\mathcal{D} = \mathbb{Q}_p$ et pour $u = p^{-r}I$ la multiplication par p^{-r}. Alors, si $r \geq 0$, $\mathcal{H}_{p^{-r}I^\epsilon} = \mathcal{H}_{r^\epsilon}$; si $r < 0$, $\mathcal{H}_{p^{-r}I^\epsilon} = 0$. Ainsi, on a $C_{p^{-r}I} = C_r$ pour $r \geq 0$.
2) Supposons la suite d'opérateurs $p^{-nr}u^{-n}$ de $\mathcal{D}$ bornée. Alors, $\mathcal{H}_{r^\epsilon}\otimes\mathcal{D}$ est contenu dans $(\mathcal{H}\otimes\mathcal{D})_{u^\epsilon}$. On a en effet alors

$$||(1\otimes u)^{-n}F||_{\rho_n} \leq c_n\frac{||F||_{\rho_n}}{p^{nr}}$$

avec c_n bornée. Si l'on pose $||v||_r = \sup ||p^{-nr}v^n||$ pour un endomorphisme v de $\mathcal{D}$ lorsque cela existe (r peut être négatif), on obtient plus précisément

$$C_u(F) \leq ||u^{-1}||_r C_r(F) \ .$$

3) Supposons la suite $p^{ns}u^n$ bornée. Alors, $(\mathcal{H}\otimes\mathcal{D})_{u^\epsilon}$ est contenu dans $\mathcal{H}_{s^\epsilon}\otimes\mathcal{D}$. En particulier, si $s < 0$, $(\mathcal{H}\otimes\mathcal{D})_{u^\epsilon} = 0$. En effet, en écrivant $F = (1\otimes u^n)(1\otimes u^{-n})F$, on a

$$\frac{||F||_{\rho_n}}{p^{ns}} \leq c'_n||(1\otimes u)^{-n}F||_{\rho_n}$$

avec c'_n bornée et on obtient plus précisément

$$C_s(F) \leq ||u||_{-s}C_u(F) \ .$$

4) Si g est un élément de $(\mathcal{H}^{\psi=0}\otimes\mathcal{D})_{u^\epsilon}$, alors $D^k(g)$ est aussi u^ϵ-borné et on a $C_u(D^k(g)) = C_u(g)$.
5) Si $r \leq s$, on a

$$(\mathcal{H}\otimes\mathcal{D})_{p^{-r}u^\epsilon} \subset (\mathcal{H}\otimes\mathcal{D})_{p^{-s}u^\epsilon} \ .$$

6) S'il existe un réel s tel que $g \in \mathcal{H} \otimes \mathcal{D}$ soit $p^{-s}u^{-}$ bornée, on note $\mathfrak{o}_u(g)$ (resp. $\mathfrak{O}_u(g)$) la borne inférieure des $r \in \mathbb{R}$ tel que $g \in \mathcal{H}_{p^{-r}u^{-}}$ (resp. le plus petit des r tels que $g \in \mathcal{H}_{p^{-r}u}$ s'il existe). Ainsi, $\mathfrak{o}_u(g) < r$ si et seulement si la suite $||p^{rn}u^{-n}g||_{\rho_n}$ tend vers 0 lorsque $n \to \infty$. Si $\mathfrak{O}_u(g)$ existe, on a $\mathfrak{O}_u(g) = \mathfrak{o}_u(g)$.

1.5. Soit V une représentation p-adique continue de G_K de dimension finie. Si T est un réseau de V stable par G_K, on note $Z^1_\infty(K,T)$ la limite projective des $H^1(K_n,T)$ pour les applications de corestriction (encore appelées trace) et $Z^1_\infty(K,V) = \mathbb{Q}_p \otimes_{\mathbb{Z}_p} Z^1_\infty(K,T)$. On note $\tilde{Z}^1_\infty(K,T)$ le quotient du Λ-module $Z^1_\infty(K,T)$ par son sous-Λ-module de torsion et $\tilde{Z}^1_\infty(K,V) = \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \tilde{Z}^1_\infty(K,T)$. Rappelons que ce sous-$\Lambda$-module de torsion est la limite projective des $H^1(K_\infty/K_n, T^{G_{K_\infty}})$ et est isomorphe à $T^{G_{K_\infty}}$ une fois choisi un générateur de G_∞. Le lemme de Shapiro implique que $Z^1_\infty(K,T) = H^1(K, \Lambda \otimes T)$ ([1, II.1], il s'agit ici de cohohomologie continue). Grâce à la suite exacte inflation-restriction, les applications de restriction induisent l'isomorphisme canonique

$$\tilde{Z}^1_\infty(K,T) \cong H^1(K_\infty, \Lambda \otimes T)^\Gamma$$

et en tensorisant par $\mathbb{Q}_p$ l'isomorphisme canonique

$$\tilde{Z}^1_\infty(K,V) \cong H^1(K_\infty, \Lambda \otimes V)^\Gamma \ .$$

1.6. Si k est un entier, on note $V(k)$ la représentation p-adique V avec la nouvelle action de G_K donnée pour $\tau \in G_K$ par $v \mapsto \chi(\tau)^k \tau v$. On a alors un opérateur de twist $Tw^k : Z^1_\infty(K,T) \to Z^1_\infty(K,T(k))$ induit par l'identité, l'action de Galois étant modifiée : $\tau(Tw^k(x)) = \chi(\tau)^k Tw^k(\tau x)$. Pour tout entier $n \geq 0$ et pour tout entier $k \in \mathbb{Z}$, le composé des opérateurs de twists et de la projection canonique de $Z^1_\infty(K,T(k)) \to H^1(K_n,T(k))$ induisent des applications

$$\begin{aligned} \pi_{n,k} : Z^1_\infty(K,V) &\to H^1(K_n, V(k)) \\ \tilde{\pi}_{n,k} : \tilde{Z}^1_\infty(K,V) &\to H^1(K_n, V(k))/H^1(G_n, V(k)^{G_{K_\infty}}) \overset{res_\infty}{\cong} H^1(K_\infty, V(k))^{\Gamma_n} \end{aligned}$$

où res_∞ est la restriction de $H^1(K_n, V(k))$ dans $H^1(K_\infty, V(k))^{\Gamma_n}$. On démontre comme dans [4, 3.4.3] que si V est de de Rham, l'action de G_∞ sur $V^{G_{K_\infty}}$ est semi-simple et que $V^{G_{K_\infty}} = \oplus V(-j)^{G_K}(j)$. En particulier, on en déduit que $V^{G_{K_n}} = V^{G_K}$ pour tout entier n et que $V^*(1)^{G_{K_n}} = V^*(1)^{G_K}$. En utilisant la dualité locale et le fait que G_∞ est de dimension cohomologique 1, il n'est pas difficile de démontrer que l'image de $Z^1_\infty(K,T)$ dans $H^1(K_n,T(k))$ est d'indice borné par rapport à n dès que $V(k)^*(1)^{G_K}$ est nul (voir par exemple [4, 3.2.1]). Lorsque $V^*(1)^{G_K}$ est non nul, l'application

$$Z^1_\infty(K,T)_{\Gamma_n} \to H^1(K_n,T)$$

n'est pas surjective, le conoyau est isomorphe à $H^1(\Gamma_n, (V^*(1)/T^*(1))^{G_{K_\infty}})\hat{}$. L'image de $Z^1_\infty(K,T)_{\Gamma_n}$ est par contre d'indice fini borné dans l'intersection de

$H^1(K_n,T)$ et de l'image Y_n de $\mathbb{Q}_p\otimes Z^1_\infty(K,T)_{\Gamma_n}$. Un élément de $H^1(K_n,V)$ est dans Y_n si et seulement si son image dans

$$H^1(\Gamma_n,V^*(1)^{G_{K_\infty}})^* = (V^*(1)^{G_{K_\infty}}/(\gamma_n-1))^* \cong (V^*(1)^{G_{K_n}})^* \cong (V^*(1)^{G_K})^*$$

est nulle, l'application $H^1(K_n,V)\to H^1(\Gamma_n,V^*(1)^{G_{K_\infty}})^*$ provenant de la dualité locale. Nous dirons que $x\in H^1(K_n,V)$ est admissible s'il appartient à l'image de $\mathbb{Q}_p\otimes Z^1_\infty(K,T)_{\Gamma_n}$. Ce qui précède montre que x est admissible si et seulement si sa trace à K l'est.

1.7. Les applications $\pi_{n,k}$ se prolongent en des applications que l'on note de la même manière :

$$\pi_{n,k}:\mathcal{H}_r(G_\infty)\otimes_{\mathbb{Q}_p\otimes\Lambda} Z^1_\infty(K,V)\to H^1(K_n,V(k))$$

ou

$$\begin{aligned}\pi_{n,k}:\mathcal{H}_r(G_\infty)\otimes_\Lambda \tilde{Z}^1_\infty(K,V) \to & H^1(K_n,V(k))/H^1(\Gamma_n,V(k)^{G_{K_\infty}})\\ &= H^1(K_\infty,V(k))^{\Gamma_n}\end{aligned}$$

pour tout entier $n\geq 0$ et pour tout entier relatif k. Nous verrons souvent $H^1(K_\infty,V(k))^{\Gamma_n}$ comme contenu dans $H^1(K_\infty,V)$. On note $*_k$ l'action twistée : $\tau *_k m=\chi(\tau)^k\tau m$.
Nous allons donner un critère pour qu'une famille de points de $H^1(K_n,V(k))$ soit dans l'image de $\mathcal{H}_r(G_\infty)\otimes_{\mathbb{Q}_p\otimes\Lambda} Z^1_\infty(K,V)$ (interpolation de familles de points).

1.8. Proposition. *Soit $s\in\overline{\mathbb{R}}$ et h un entier $>s$. Soit $P_{n,k}$ une famille d'éléments admissibles de $H^1(K_n,V(k))$ pour $n\in\mathbb{N}$ et $k=0,\dots h-1$ telle que*

(i) $Tr_{n+1,n}(P_{n+1,k})=P_{n,k}$;
(ii) *les suites* $p^{\lfloor n(s-j)\rfloor}\sum_{k=0}^{j}(-1)^k\binom{j}{k}res_\infty P_{n,k}$ *convergent vers* 0 *dans* $H^1(K_\infty,V)$ *pour tout* $0\leq j\leq h-1$.

Alors, il existe un unique élément z de $\mathcal{H}_{s^-}(G_\infty)\otimes\tilde{Z}^1_\infty(K,T)$ tel que $\tilde{\pi}_{n,k}(z)=\tilde{P}_{n,k}$.

Démonstration. On commence par remplacer G_∞ par Γ (on a $\mathcal{H}_{s^-}(G_\infty)\otimes Z^1_\infty(K,T)=\mathcal{H}_{s^-}(\Gamma)\otimes\mathbb{Z}_p[\Delta]\otimes Z^1_\infty(K,T)$). Il existe un système libre $X_1,\cdots,X_d$ du Λ_Γ-module $Z^1_\infty(K,T)$ tels que pour tout entier k compris entre 0 et $h-1$, les éléments $\tilde{\pi}_{n,k}(X_i)$ de $res_\infty H^1(K_n,T(k))=H^1(K_\infty,T(k))^{\Gamma_n}\subset H^1(K_\infty,T)$ en forment un système libre (modulo torsion) engendrant un $\mathbb{Z}_p[G_n]$-module d'indice borné c par rapport à n dans le $\mathbb{Q}_p$-espace vectoriel des éléments admissibles de $H^1(K_\infty,T(k))^{\Gamma_n}$. On écrit alors les points $res_\infty(P_{n,k})$ dans cette base :

$$res_\infty(P_{n,k})=\sum_{i=1}^d b^{(i)}_{n,k}*_k\tilde{\pi}_{n,k}(X_i)=\sum_{i=1}^d Tw^k b^{(i)}_{n,k}\tilde{\pi}_{n,k}(X_i)$$

avec les $b^{(i)}_{n,k}$ dans $\mathbb{Q}_p[G_n]$. On peut écrire $b^{(i)}_{n,k}=R^{(i)}_{n,k}(\gamma-1)$ avec γ générateur de Γ et $R^{(i)}_{n,k}$ polynôme en T de degré $<p^n$ à coefficients dans $\mathbb{Q}_p[\Delta]$. La

première condition signifie que par l'application naturelle de $\mathbb{Q}_p[G_{n+1}]$ dans $\mathbb{Q}_p[G_n]$, l'image de $b^{(i)}_{n+1,k}$ est $b^{(i)}_{n,k}$ et donc que $R^{(i)}_{n+1,k} \equiv R^{(i)}_{n,k} \mod (1+T)^{p^n}-1$. La deuxième condition signifie que pour tout entier j avec $0 \leq j \leq h-1$ et pour $i = 1, \cdots, d$, les suites $p^{n(s-j)} \sum_{k=0}^{j}(-1)^k \binom{j}{k} R^{(i)}_{n,k}(\chi(\gamma)^k(1+T)-1)$ tendent vers 0, ce qui est la condition pour qu'il existe un unique élément de $f^{(i)} \in \mathcal{H}_{s^-}$ tel que $f^{(i)}(T) \equiv R^{(i)}_{n,k}(\chi(\gamma)^k(1+T)-1) \mod \chi(\gamma^{p^n})^k\gamma^{p^n}-1$. L'élément $\sum_{i=1}^{d} f^{(i)}(\gamma-1)X_i \in \mathcal{H}_{s^-}(G_\infty) \otimes Z^1_\infty(K,T)$ convient. L'unicité de z vient de l'unicité des $f^{(i)}$. □

Remarques : 1. Une famille d'éléments vérifiant les conditions de la proposition sera dite tempérée d'ordre $< s$. Nous parlerons de congruences pour évoquer les conditions (ii).
2. On peut mettre d'autres types de conditions : par exemple, si s est un entier, une famille de points $P_{n,k}$ pour $n \geq 1$ et $k \in \{0, \ldots, s-1\}$ telle que

(i) $Tr_{n+1,n}(P_{n+1,k}) = P_{n,k}$
(ii) les suites $p^n \sum_{k=0}^{s-1}(-a)^k \binom{s-1}{k} res_\infty P_{n,k}$ convergent vers 0 pour tout $a \in \mathbb{Z}_p$,

est tempérée d'ordre $< s$. Pour le démontrer, on écrit $(X-1)^j$ pour $j \in \{0, \ldots, s-1\}$ dans le système libre formé des $(X-a_l)^{s-1}$ pour a_l s points distincts de $\mathbb{Z}_p$ et on en déduit une relation

$$\sum_{k=0}^{j}(-1)^k \binom{j}{k} res_\infty P_{n,k} = \sum_{l=0}^{s-1} c_{k,l} \sum_{k=0}^{s-1}(-a_l)^k \binom{s-1}{k} res_\infty P_{n,k} .$$

Le résultat s'en déduit.
3. La proposition dit aussi que l'application naturelle

$$\mathcal{H}_{s^-}(G_\infty) \otimes \tilde{Z}^1_\infty(K,T) \overset{\prod \tilde{\pi}_{n,k}}{\rightarrow} \prod_{n, 0 \leq k \leq h-1} H^1(K_\infty, V(k))$$

est injective.

1.9. L'application naturelle de Λ-modules de $Z^1_\infty(K,T)$ dans $H^1(K, \Lambda \otimes T)$ se prolonge en une application δ de $\mathcal{H}_s(G_\infty) \otimes_\Lambda Z^1_\infty(K,T)$ dans $H^1(K, \mathcal{H}_s(G_\infty) \otimes T)$. Colmez démontre qu'il s'agit en fait d'un isomorphisme. Pour cela, on remarque qu'il suffit de montrer que l'homomorphisme $\mathcal{H}_s(G_\infty) \otimes_\Lambda \tilde{Z}^1_\infty(K,T) \to H^1(K_\infty, \mathcal{H}_s(G_\infty) \otimes T)^{G_\infty}$ est un isomorphisme. On a en effet un diagramme commutatif dont les lignes sont exactes :

$$\begin{array}{ccccc}
0\to & \mathcal{H}_s(G_\infty)\otimes T^{G_{K_\infty}} & \to & \mathcal{H}_s(G_\infty)\otimes Z^1_\infty(K,T) & \to \\
 & \downarrow & & \downarrow & \\
0\to & H^1(G_\infty,\mathcal{H}_s(G_\infty)\otimes T^{G_{K_\infty}}) & \to & H^1(K,\mathcal{H}_s(G_\infty)\otimes T) & \to
\end{array}$$

$$\begin{array}{cccc}
\to & \mathcal{H}_s(G_\infty)\otimes \tilde{Z}^1_\infty(K,T) & \to & 0 \\
 & \downarrow & & \\
\to & H^1(K_\infty,\mathcal{H}_s(G_\infty)\otimes T)^{G_\infty} & \to & 0
\end{array}$$

Il est facile de vérifier que la première flèche verticale est un isomorphisme. Il ne reste donc plus qu'à montrer que la troisième l'est. On a des applications naturelles de $H^1(K,\mathcal{H}_s(G_\infty)\otimes T)$ dans $H^1(K_n,V(k))$ pour tout entier $n\geq 1$ et $k\in\mathbb{Z}$ induites par l'homomorphisme $s_{n,k}$ de G_{K_n}-modules $\mathcal{H}_s(G_\infty)\otimes V\to \mathbb{Z}_p[G_n]\otimes V(k)\to V(k)$ donné par $f\mapsto R_{n,k}(f)\equiv Tw^k(f) \mod \gamma^{p^n}-1\mapsto \nu^n_{Id}(R_{n,k}(f))$ où $\nu^n_{Id}(\sum_\tau a_\tau\tau)=a_{Id}$ (cf. l'appendice A.1) et le diagramme suivant commute

$$\begin{array}{ccc}
\mathcal{H}_s(G_\infty)\otimes_\Lambda \tilde{Z}^1_\infty(K,T) & \xrightarrow{\delta} & H^1(K_\infty,\mathcal{H}_s(G_\infty)\otimes T)^{G_\infty} \\
\tilde{\pi}\downarrow & & \pi'\downarrow \\
\prod_{n,k} H^1(K_\infty,V(k))^{\Gamma_n} & = & \prod_{n,k} H^1(K_\infty,V(k))^{\Gamma_n}
\end{array}$$

L'image d'un élément de $H^1(K_\infty,\mathcal{H}_s(G_\infty)\otimes T)^{G_\infty}$ vérifie les conditions de la proposition 1.8, ce qui permet de définir un homomorphisme δ' de Λ-modules de $H^1(K_\infty,\mathcal{H}_s(G_\infty)\otimes T)^{G_\infty}$ dans $\mathcal{H}_s(G_\infty)\otimes_\Lambda\tilde{Z}^1_\infty(K,T)$ tel que $\pi'(\delta\circ\delta'(x)-x)=0$ et tel que $\tilde{\pi}(\delta'\circ\delta(x)-x)=0$. Comme $\prod_{n,0\leq k\leq h-1}\tilde{\pi}_{n,k}$ est injective, $\delta'\circ\delta(x)=x$. De même, il n'est pas difficile de voir que $\prod_{n,0\leq k\leq h-1}\pi'_{n,k}$ est injective, ce qui implique $\delta\circ\delta'(x)=x$. D'où l'isomorphisme.

2. Equations $(1-p^r\Phi)\mathcal{G}_r=D^r(g)$

On fixe un φ-module $\mathcal{D}$ de dimension finie, c'est-à-dire un $\mathbb{Q}_p$-espace vectoriel de dimension finie muni d'un automorphisme φ. Le $\mathbb{Q}_p$-espace vectoriel $\mathcal{H}_\infty\otimes\mathcal{D}$ est muni d'une action continue de G_∞, des opérateurs $\Phi=\varphi\otimes\varphi$ et $D=D\otimes 1$.

2.1. On note $\tilde{\Delta}:\mathcal{H}_\infty^{\psi=0}\otimes\mathcal{D}\to\oplus_{k\in\mathbb{Z}}\mathcal{D}/(1-p^{-k}\varphi)\mathcal{D}$ l'application définie par $g\mapsto(\frac{D^k(g)(0)}{k!} \mod (1-p^{-k}\varphi)\mathcal{D})_{k\in\mathbb{Z}}$. On note $\mathcal{D}_{\infty,e}=(\mathcal{H}_\infty^{\psi=0}\otimes\mathcal{D})^{\tilde{\Delta}=0}$ l'ensemble des éléments $g\in\mathcal{H}_\infty^{\psi=0}\otimes\mathcal{D}$ tels que $\frac{D^k(g)(0)}{k!}\in(1-p^{-k}\varphi)\mathcal{D}$ pour tout $k\in\mathbb{Z}$.

Soit $g\in\mathcal{D}_{\infty,e}$. Alors, l'équation $(1-p^r\Phi)H=D^rg$ a une solution dans $\mathcal{H}_\infty\otimes\mathcal{D}$ pour tout $r\in\mathbb{Z}$. Pour le démontrer, on fixe un entier h assez grand pour que la série $\sum_{n=0}^\infty\Phi^n(f)$ converge dans $\mathcal{H}_\infty\otimes\mathcal{D}$ dès que $f\in\mathcal{H}\cap T^h\mathbb{Q}_p[[T]]$, on remarque que $f=D^r(g)-\sum_{0\leq k<h}\frac{D^{r+k}(g)(0)}{k!}\log^k(1+T)\in T^h\mathbb{Q}_p[[T]]$; comme $\frac{D^{r+k}(g)(0)}{k!}=(1-p^{r+k}\varphi)a_k$ et que $(1-p^r\Phi)(a_k\frac{\log^k(1+T)}{k!})=\frac{D^{r+k}(g)(0)}{k!}\log^k(1+T)$, on en déduit l'existence d'une solution de l'équation $(1-p^r\Phi)H=D^r(g)$.

Définition : Soit $g \in \mathcal{D}_{\infty,e}$. On appelle solution compatible des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$ une famille $\overline{G} = (G_r)$ de solutions $G_r \in \mathcal{H}_\infty \otimes D$ de l'équation $(1-p^r\Phi)\mathcal{G}_r = D^r g$ tels que $D(G_r) = G_{r+1}$. On pose alors $D^r(\overline{G}) = G_r$ pour tout r et $D^0(\overline{G}) = G$.

Deux solutions compatibles des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$ différent d'un élément $\overline{H} = (H_r)$ avec $H_r = \sum_{i\geq 0} \frac{a_{i+r}}{i!} \log^i(1+T) = \sum_{j\geq r} \frac{a_j}{(j-r)!} \log^{j-r}(1+T)$ avec $a_j \in \mathcal{D}^{\varphi=p^{-j}}$.
D'autre part, il existe toujours une telle solution pour $g \in \mathcal{D}_{\infty,e}$: soit r vérifiant $\mathcal{D}^{\varphi=p^{-s}} = 0$ pour tout $s < r$; pour $s < r$ deux solutions H et H' de l'équation $(1-p^s\Phi)H = D^s g$ telles que $D^{r-s}(H) = D^{r-s}(H')$ sont égales. Pour exhiber une solution compatible des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$, il suffit donc de choisir $G_r \in \mathcal{H}_\infty \otimes \mathcal{D}$ tel que $(1-p^r\Phi)G_r = D^r(g)$ et de prendre $G_k = D^{k-r}(G_r)$ pour $k \geq r$ et $G_k =$ l'unique solution de $(1-p^k\Phi)H = D^k g$ telle que $D^{r-k}(H) = G_r$ pour $k < r$.
On obtient aussi que $\overline{G}$ est déterminé par G_r pour r assez petit. De plus, si $\mathcal{D}^{\varphi=p^r}$ est nul pour tout $r \in \mathbb{Z}$, la donnée de G_0 détermine tous les G_r. Il suffit donc de se donner une solution G de l'équation $(1-\Phi)G = g$.

2.2. Notons $\mathcal{D}_{\infty,f} = \mathcal{H}_\infty^{\psi=0} \otimes \mathcal{D}$. Pour tout entier r, définissons une application $L_r : \oplus_{j\in\mathbb{Z}}\mathcal{D} \to \mathcal{H}_\infty \otimes \mathcal{D}$ par $L_r(c) = \sum_{j\geq r} \frac{c_j}{(j-r)!} \log^{j-r}(1+T)$. On a $D(L_r(c)) = L_{r+1}(c)$.

Définition : Soit $g \in \mathcal{D}_{\infty,f}$. On appelle solution compatible des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$ un couple $\overline{G} = (b, G)$ où $b \in \oplus_{j\in\mathbb{Z}}\mathcal{D}$ et où $G = (G_r)_{r\in\mathbb{Z}}$ est une famille d'éléments de $\mathcal{H}_\infty \otimes \mathcal{D}$ tels que

1. $D(G_r) = G_{r+1}$
2. $(1-p^r\Phi)G_r - L_r(b) = D^r(g)$.

Il existe toujours une solution compatible des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$. Il suffit en effet de choisir des éléments a_j presque tous nuls tels que $a_j \equiv D^j(g)(0) \mod (1-p^j\varphi)\mathcal{D}$ pour tout $j \in \mathbb{Z}$, de poser $b = (a_i)_{i\in\mathbb{Z}}$ et de prendre pour r_0 assez petit une solution G_{r_0} de l'équation $(1-p^{r_0}\Phi)\mathcal{G}_{r_0} = D^{r_0}(g) - \sum_{j\geq r_0} \frac{a_j}{(j-r_0)!} \log^{j-r_0}(1+T)$. et de poser pour tout $r \geq r_0$, $G_r = D^{r-r_0}G_{r_0}$ et pour $r < r_0$ l'unique solution G_r telle que $D^{r_0-r}G_r = G_{r_0}$.
Considérons l'application $\beta : \oplus_{j\in\mathbb{Z}}\mathcal{D} \to (\oplus_{j\in\mathbb{Z}}\mathcal{D}) \oplus \oplus_{r\in\mathbb{Z}}\mathcal{H}_\infty \otimes \mathcal{D}$ donnée par

$$(\alpha_j)_{j\in\mathbb{Z}} \mapsto (((1-p^j\varphi)\alpha_j)_{j\in\mathbb{Z}}, (\sum_{j\geq r} \frac{\alpha_j}{(j-r)!}(\log(1+T))^{j-r}))_{r\in\mathbb{Z}}$$

Deux solutions compatibles des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$ différent d'un élément $(\beta_r(\alpha))$ avec $\alpha \in \oplus_{j\in\mathbb{Z}}\mathcal{D}$.
Si $\overline{G} = (b, G)$ est une solution compatible des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$, on note $D^r(\overline{G}) = (L_r(b), G_r)$ pour tout $r \in \mathbb{Z}$.

2.3. Si τ est un élément de G_∞, on pose

$$\Pi_\tau(T) = (1 - p^{-1}\varphi) \log \frac{(1+T)^{\chi(\tau)} - 1}{T} .$$

Comme $((1+T)^{\chi(\tau)} - 1)/T$ appartient à $\mathbb{Z}_p^\times + T\mathbb{Z}_p[[T]]$, $\Pi_\tau \in \mathbb{Q}_p \otimes \mathbb{Z}_p[[T]]$ par le lemme de Dwork. C'est de plus un élément de $\mathbb{Q}_p \otimes \mathbb{Z}_p[[T]]^{\psi=0}$ et l'on peut donc considérer pour tout entier $r \in \mathbb{Z}$ l'élément $D^r(\Pi_\tau) \in \mathbb{Q}_p \otimes \mathbb{Z}_p[[T]]^{\psi=0}$.

Définition : On définit $\mathcal{D}_{\infty,g}$ comme l'extension de G_∞-modules de $\oplus \mathcal{D}^{\varphi=p^i}(i+1)$ par $\mathcal{D}_{\infty,f} = \mathcal{H}_\infty^{\psi=0} \otimes \mathcal{D}$ donnée par le cocycle

$$\tau \mapsto ((a_i) \mapsto \sum_{i \in \mathbb{Z}} D^{i+1}(\Pi_\tau) a_i) ,$$

l'action de G_∞ sur $\mathcal{D}^{\varphi=p^i}(i+1) = \mathcal{D}^{\varphi=p^i}$ étant donnée par le caractère χ^{i+1}. Si $a \in \mathcal{D}^{\varphi=p^i}(i+1)$, on note $U(a)$ l'élément de $\mathcal{D}_{\infty,g}$ tel que $(\tau-1)(U(a)) = D^{i+1}(\Pi_\tau)a$.

On a donc une suite exacte de G_∞-modules

$$0 \to \mathcal{D}_{\infty,f} \to \mathcal{D}_{\infty,g} \to \oplus \mathcal{D}^{\varphi=p^i}(i+1) \to 0 .$$

Remarquons aussi que si l'on tensorise par l'anneau total des fractions $\mathcal{K}(G_\infty)$ de $\mathcal{H}(G_\infty)$, on obtient un isomorphisme $\mathcal{K}(G_\infty) \otimes \mathcal{D}_{\infty,f} \cong \mathcal{K}(G_\infty) \otimes \mathcal{D}_{\infty,g}$. On prolonge l'application U par linéarité sur $\oplus \mathcal{D}^{\varphi=p^i}(i+1)$. La formule $D \circ \tau = \chi(\tau)\tau \circ D$ et le fait que l'on a muni ici $\mathcal{D}^{\varphi=p^i}$ de l'action de G_∞ donnée par χ^{i+1} impliquent que l'expression écrite est bien un cocycle. L'équation $D \circ \varphi = p\varphi \circ D$ implique que pour $a \in \mathcal{D}^{\varphi=p^i}$,

$$(1 - \Phi)(D^{i+1} \log \frac{(1+T)^{\chi(\tau)} - 1}{T} a) = D^{i+1}(\Pi_\tau)a .$$

Ainsi, moralement pour $a \in \mathcal{D}^{\varphi=p^{-1}}$, on a $U(a) = (1-\Phi)(a \log T)$ et pour $a \in \mathcal{D}^{\varphi=p^i}$, $U(a) = (1-\Phi)((D^{i+1} \log T)a)$. On notera symboliquement $\tilde{U}(a) = (D^{i+1} \log T)a$ et $(1-\Phi)\tilde{U}(a) = U(a)$. Remarquons que si l'on veut évaluer $\log T$ en $T = \zeta_n - 1$, il est nécessaire de faire un choix du logarithme : nous prendrons lorsque cela sera nécessaire l'extension de log telle que $\log p = 0$. Si $g \in \mathcal{D}_{\infty,g}$, on note $\lambda_i(g)$ sa projection sur $\mathcal{D}^{\varphi=p^i}(i+1)$; donc

$$g - \sum_i U(\lambda_i(g)) \in \mathcal{D}_{\infty,f} .$$

Définition : Soit $g \in \mathcal{D}_{\infty,g}$. On appelle solution compatible des équations de $(1 - p^r\Phi)\mathcal{G}_r = D^r(g)$ une solution compatible des équations $(1 - p^r\Phi)\mathcal{H}_r = D^r(g - \sum_i U(\lambda_i(g))) \in \mathcal{D}_{\infty,f}$.

Ainsi, on se donne $b \in \oplus_{j \in \mathbb{Z}} \mathcal{D}$ et $H_r \in \mathcal{H}_\infty \otimes \mathcal{D}$ tels que

1. $D(H_r) = H_{r+1}$
2. $(1 - p^r\Phi)H_r - \sum_{j \geq r} \frac{b_j}{(j-r)!} \log^{j-r}(1+T) = D^r(g - \sum_i U(\lambda_i(g)))$.

Si $\overline{G}$ est une telle solution, on pose $D^r(\overline{G}) = (\lambda_r(g), L_r(b), H_r)$.

3. Application exponentielle

3.1. Nous utiliserons comme Colmez les anneaux $A_{\max}$ et $B_{\max}[t^{-1}]$ à la place de B_{cris}. La norme définie sur $B_{\max}$ par $||x||_{\max} \le 1$ si et seulement si $x \in A_{\max}$ vérifie la propriété

$$p^{-1}||x||_{\max}.||y||_{\max} \le ||xy||_{\max} \le ||x||_{\max}.||y||_{\max}$$

L'importance de la norme $|| \; ||_{\max}$ vient en particulier de son lien avec les $|| \; ||_\rho$ définies sur $\mathcal{H}$. Pour l'énoncer, introduisons $[\epsilon]$ le relèvement de Teichmüller de $\epsilon = (\zeta_n)$ dans $A_{\max}$ et $\beta_n = \varphi^{-n}([\epsilon]) = [(\zeta_{m+n})_m]$.

3.1.1. Lemme. *Si F est un élément de $\mathcal{H}$, alors $F(\beta_n - 1) = \Phi^{-n}(F([\epsilon]-1))$ vérifie*

$$||F||_{\rho_n} \le ||F(\beta_n - 1)||_{\max} \le p||F||_{\rho_n} \; .$$

Ainsi, si F est $p^{-u}\varphi^{-}$-borné, la suite $||p^{nu}(1\otimes\varphi)^{-n}F(\beta_n-1)||_{\max}$ tend vers 0 lorsque $n \to \infty$.

Démonstration. Nous ne donnons qu'une esquisse de la démonstration. Colmez ([1, corollaire V.5.5]) démontre que dans $B_{\max}^{G_{K_\infty}}$ muni de la norme $|| \; ||_{\max}$, les éléments $e_{n,k} = \frac{(\beta_n-1)^k}{p^{[\frac{k}{p^n(p-1)}]}}$ (à n fixé) forment un système libre de Banach, c'est-à-dire que la série $\sum_{k\ge 0} a_k e_{n,k}$ converge dans $B_{\max}$ pour $a_k \in K$ si et seulement si $\sup |a_k| < \infty$ et on a alors

$$||\sum_{k\ge 0} a_k e_{n,k}||_{\max} = \sup |a_k|$$

(il démontre plus que cela, mais nous ne nous servirons que de cela). Si $F \in \mathcal{H}$, on a d'autre part

$$||F||_{\rho_n} = \sup |a_k|\rho_n^k = \sup |a_k| p^{-\frac{k}{p^n(p-1)}} \; .$$

En utilisant le fait que

$$\frac{k}{p^n(p-1)} - 1 \le [\frac{k}{p^n(p-1)}] \le \frac{k}{p^n(p-1)} \; ,$$

on en déduit que $F(\beta_n - 1)$ existe dans $B_{\max}$ et que

$$||F||_{\rho_n} \le ||F(\beta_n - 1)||_{\max} \le p||F||_{\rho_n} \; ,$$

ce qui termine la démonstration. Remarquons qu'en utilisant le fait que $||FG||_\rho = ||F||_\rho ||G||_\rho$, on peut en déduire l'inégalité pour $x = F(\beta_n - 1)$ et $y = G(\beta_n - 1)$

$$p^{-2}||x||_{\max}||y||_{\max} \le ||xy||_{\max} \le p||x||_{\max}||y||_{\max} \; ,$$

ce qui est bien sûr moins fort que ce que démontre Colmez, mais suffisant. □

Soit $\mathbf{D}_p(V) = (B_{\mathrm{cris}} \otimes V)^{G_K} = (B_{\max} \otimes V)^{G_K}$ et $\mathbf{D}_{\mathrm{dR}}(V) = (B_{\mathrm{dR}} \otimes V)^{G_K}$. On note e_k la base canonique du φ-module filtré $\mathbb{Q}_p[k]$ égal à $\mathbb{Q}_p$ avec $\varphi\alpha = p^k\alpha$; on a donc un isomorphisme canonique entre $\mathbf{D}_p(V)$ et $\mathbf{D}_p(V(k))$ donné par $d \mapsto d \otimes e_{-k}$. On plonge $\mathbf{D}_p(V(k))$ dans $B_{\max} \otimes \mathbf{D}_p(V)$ par $d \mapsto t^{-k} \otimes (de_k)$, ce qui donne l'identification $\mathbf{D}_p(V(k)) = \mathbf{D}_p(V) \otimes e_{-k} \to B_{\max} \otimes \mathbf{D}_p(V)$: $d \otimes e_{-k} \mapsto t^{-k} \otimes d$. Remarquons que cette identification est compatible avec la filtration et avec l'homomorphisme de Frobenius, mais non avec l'action de Galois : ainsi, on obtient bien que $\mathrm{Fil}^0(B_{\max} \otimes \mathbf{D}_p(V(k)))^{\varphi=1}$ est $V(k)$.
Nous utiliserons la propriété suivante du φ-module filtré $\mathbf{D}_p(V)$: si $\mathrm{Fil}^{-h}\,\mathbf{D}_p(V) = \mathbf{D}_p(V)$, les applications $1 - p^s\varphi$ sont des isomorphismes de $\mathbf{D}_p(V)$ pour $s > h$.
Bloch et Kato définissent une application $\exp_V = \exp_{V,K_n} : \mathbf{D}_p(V) \oplus K_n \otimes \mathbf{D}_{\mathrm{dR}}(V) \to H^1(K_n, V)$. Plus précisément, nous noterons :

$$\begin{aligned}
\exp_{V,e} = \exp_{V,K_n,e} &: K_n \otimes \mathbf{D}_{\mathrm{dR}}(V) \to H^1_e(K_n, V),\\
\exp_{V,f} = \exp_{V,K_n,f} &: \mathbf{D}_p(V) \oplus K_n \otimes \mathbf{D}_{\mathrm{dR}}(V) \to H^1_e(K_n, V)\\
\exp_{V,g} = \exp_{V,K_n,g} &: \mathbf{D}_p(V)^{\varphi=p^{-1}} \oplus \mathbf{D}_p(V) \oplus K_n \otimes \mathbf{D}_{\mathrm{dR}}(V) \to H^1_g(K_n, V)
\end{aligned}$$

la dermière de ces exponentielles dépendant du choix d'un logarithme (nous prendrons ici $\log_p p = 0$). Rappelons les définitions des applications $\exp_*$ pour $* \in \{e, f, g\}$ sur la partie "cristalline" qui nous intéresse. Fixons un scindage continu Eul de

$$1 - \varphi : \mathrm{Fil}^0(B_{\max} \otimes \mathbf{D}_p(V)) \to B_{\max} \otimes \mathbf{D}_p(V)\ .$$

Deux tels scindages diffèrent d'un homomorphisme continu de $B_{\max} \otimes \mathbf{D}_p(V)$ dans V. Ainsi, si $b \in B_{\max} \otimes \mathbf{D}_p(V)$, $(1-\varphi)Eul(b) = b$ et $Eul(b) \in \mathrm{Fil}^0(B_{\max} \otimes \mathbf{D}_p(V))$.
Soit L une extension algébrique de K contenue dans $\bar{K}$.

3.1.2. Soit $a \in L \otimes \mathbf{D}_p(V)$. Alors $P = \exp_{V,e}(a) \in H^1(L, V)$ est la classe du cocycle

$$\tau \in G_L \mapsto (\tau - 1)(c - Eul((1 - \varphi)c))$$

où $c \in B_{\max} \otimes \mathbf{D}_p(V)$ vérifie $c - a \in \mathrm{Fil}^0(B_{\mathrm{dR}} \otimes \mathbf{D}_p(V))$. Si $C = c - Eul((1-\varphi)c)$, on a donc $(1 - \varphi)C = 0$, $NC = 0$ et $C \equiv a \mod \mathrm{Fil}^0(B_{\mathrm{dR}} \otimes \mathbf{D}_p(V))$.

3.1.3. Soit $(b, a) \in \mathbf{D}_p(V) \oplus L \otimes \mathbf{D}_p(V)$. Alors $P = \exp_{V,f}(b, a) \in H^1(L, V)$ est la classe du cocycle

$$\tau \in G_L \mapsto (\tau - 1)(c - Eul((1 - \varphi)c - b))$$

où $c \in B_{\max} \otimes \mathbf{D}_p(V)$ vérifie $c - a \in \mathrm{Fil}^0(B_{\mathrm{dR}} \otimes \mathbf{D}_p(V))$. De nouveau, si $C = c - Eul((1-\varphi)c - b)$, il vérifie $(1-\varphi)C = b$ et $C \equiv a \mod \mathrm{Fil}^0(B_{\mathrm{dR}} \otimes \mathbf{D}_p(V))$.

3.1.4. Soit $(d,b,a) \in \mathbf{D}_p(V)^{\varphi=p^{-1}} \oplus \mathbf{D}_p(V) \oplus L \otimes \mathbf{D}_p(V)$. Alors $P = \exp_{V,g}(d,b,a) \in H^1(L,V)$ est la classe du cocycle

$$\tau \in G_L \mapsto (\tau-1)(c - Eul((1-\varphi)c-b))$$

où $c \in B_{st} \otimes \mathbf{D}_p(V)$ vérifie

1. $c - a \in \mathrm{Fil}^0(B_{\mathrm{dR}} \otimes \mathbf{D}_p(V))$,
2. $Nc = d$

(remarquons que $N((1-\varphi)c-b) = (1-p^{-1}\varphi)Nc = 0$ et $Eul((1-\varphi)c-b)$ est donc définie) et $C = c - Eul((1-\varphi)c-b)$ vérifie $(1-\varphi)C = b$, $NC = d$ et $C \equiv a \mod \mathrm{Fil}^0(B_{\mathrm{dR}} \otimes \mathbf{D}_p(V))$.

Remarque : $e_B(a) = c - Eul((1-\varphi)c)$ est bien définie à valeurs dans $B_{\max}^{\varphi=1} \otimes V/V$. Lorsque L est contenue dans K_∞, on peut en fait choisir c dans $B_{\max}^{G_{K_\infty}} \otimes \mathbf{D}_p(V)$: en effet la nullité de $H^1(K_\infty, \mathrm{Fil}^0 B_{\max})$ (voir [1, IV], voir aussi le §4.1) implique la surjectivité de l'application $B_{\max}^{G_{K_\infty}} \to (B_{\mathrm{dR}}/B_{\mathrm{dR}}^+)^{G_{K_\infty}}$ et donc en tensorisant par V celle de $B_{\max}^{G_{K_\infty}} \otimes V \to (B_{\mathrm{dR}}/B_{\mathrm{dR}}^+)^{G_{K_\infty}} \otimes V$; nous allons le faire explicitement au paragraphe suivant. On en déduit que la restriction $res_\infty(P)$ de P à K_∞ est la classe du cocycle $\tau \mapsto -(\tau-1)Eul((1-\varphi)c)$, où c est un élément de $B_{\max}^{G_{K_\infty}} \otimes \mathbf{D}_p(V)$ congru à $a \mod \mathrm{Fil}^0(B_{\mathrm{dR}} \otimes \mathbf{D}_p(V))$.

3.2. On pose

$$\begin{aligned}\mathcal{D}_{\infty,e}(V) =& (\mathcal{H}_\infty^{\psi=0} \otimes \mathbf{D}_p(V))^{\tilde{\Delta}=0}\\ \mathcal{D}_{\infty,f}(V) =& \mathcal{H}_\infty^{\psi=0} \otimes \mathbf{D}_p(V)\\ \mathcal{D}_{\infty,g}(V) =& \mathcal{D}_{\infty,g}(\mathbf{D}_p(V))\end{aligned}$$

Soit h un entier ≥ 1 et k un entier tel que $h+k-1 \geq 0$.

3.2.1. Soient $g \in \mathcal{D}_{\infty,e}(V)$ et $\overline{G}$ une solution compatible des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$. Posons pour $n \geq 1$

$$\begin{aligned}\Xi_{n,k}^{(h)}(\overline{G}) =& (-1)^{h+k-1}(h+k-1)! p^{-n}(1\otimes\varphi)^{-n}(D^{-k}(\overline{G})(\zeta_n-1) \otimes e_{-k})\\ =& (-1)^{h+k-1}(h+k-1)! p^{n(k-1)}(1\otimes\varphi)^{-n} D^{-k}(\overline{G})(\zeta_n-1) \otimes e_{-k} \ ;\end{aligned}$$

c'est un élément de $K_n \otimes \mathbf{D}_p(V(k))$. Posons pour $n \geq 1$

$$P_{n,k}^{(h)}(\overline{G}) = \exp_{V(k),e}(\Xi_{n,k}^{(h)}(\overline{G})) \in H_e^1(K_n, V(k)) \ .$$

3.2.2. Soient $g \in \mathcal{D}_{\infty,f}(V)$ et $\overline{G}$ une solution compatible des équations $(1-p^r\Phi)\mathcal{G}_r = D^r(g)$: $\overline{G} = (b, (G_r)_{r\in\mathbb{Z}})$ avec $b \in \oplus_{j\in\mathbb{Z}}\mathcal{D}$ et $G_r \in \mathcal{H}_\infty \otimes \mathcal{D}$. Posons pour $n \geq 1$

$$\begin{aligned}\Xi_{n,k}^{(h)}(\overline{G}) =& (-1)^{h+k-1}(h+k-1)! p^{-n}(1\otimes\varphi)^{-n}(D^{-k}(\overline{G})(\zeta_n-1) \otimes e_{-k})\\ =& (-1)^{h+k-1}(h+k-1)! p^{n(k-1)}(1\otimes\varphi)^{-n} D^{-k}(\overline{G})(\zeta_n-1) \otimes e_{-k} \ ;\end{aligned}$$

c'est un élément de $\mathbf{D}_p(V(k)) \oplus K_n \otimes \mathbf{D}_p(V(k))$. Posons

$$P_{n,k}^{(h)}(\overline{G}) = \exp_{V(k),f}(\Xi_{n,k}^{(h)}(\overline{G})) \in H^1_f(K_n, V(k)) \ .$$

Comme $\log \zeta_n = 0$, les contributions des b_i dans $L_{-k}(b)$ pour $i > -k$ sont nulles et on a donc

$$\Xi_{n,k}^{(h)}(\overline{G}) =$$
$$(-1)^{h+k-1}(h+k-1)! p^{n(k-1)}(\varphi^{-n} b_{-k} \otimes e_{-k}, (1 \otimes \varphi)^{-n} G_{-k}(\zeta_n - 1) \otimes e_{-k}) \ .$$

3.2.3. Enfin, soit $g \in \mathcal{D}_{\infty,g}(V)$. On fixe une solution compatible $\overline{\Psi}_\tau$ des équations $(1 - p^r\varphi)\mathcal{G}_r = D^r(\Pi_\tau)$, c'est-à-dire une famille de $\Psi_\tau^{(r)} = D^r(\overline{\Psi}_\tau)$ de solutions de l'équation $(1 - p^r\varphi)\Psi_\tau^{(r)} = D^r\Pi_\tau$ pour tout entier $r \in \mathbb{Z}$ vérifiant $D(\Psi_\tau^{(r)}) = \Psi_\tau^{(r+1)}$. On vérifie alors que pour $\tau \in \mathrm{Gal}(K_\infty/K_n)$ non trivial et $i \neq 0$, l'expression $(\chi^{-i}(\tau) - 1)^{-1} D^i(\overline{\Psi}_\tau)(\zeta_n - 1)$ est un élément de K_n ne dépendant pas du choix de τ. On le note $\ell^{(1-i)}(\zeta_n - 1)$. On pose $l^{(1)}(\zeta_n - 1) = \log(\zeta_n - 1)$ (où l'on a choisi $\log_p p = 0$).
Si $a \in \mathbf{D}_p(V)^{\varphi = p^{k-1}}$, on pose

$$\Xi_{n,k}^{(h)}(\tilde{U}(a)) = (-1)^{h+k-1}(h+k-1)!(a, 0, \log(\zeta_n - 1)a) \otimes e_{-k} \ .$$

Si $a \in \mathbf{D}_p(V)^{\varphi = p^i}$ avec $k \neq i+1$, on pose

$$\begin{aligned}\Xi_{n,k}^{(h)}(\tilde{U}(a)) =& (-1)^{h+k-1}(h+k-1)!(0, 0, \ell^{(k-i)}(\zeta_n - 1)a \otimes e_{-k} \\ =& (-1)^{h+k-1}(h+k-1)!(0, 0, \frac{D^{-k+i+1}(\overline{\Psi}_\tau)(\zeta_n - 1)}{(\chi^{-k+i+1}(\tau) - 1)} a \otimes e_{-k}) \ .\end{aligned}$$

On pose $P_{n,k}^{(h)}(\tilde{U}(a)) = \exp_{V(k),g}(\Xi_{n,k}^{(h)}(\tilde{U}(a))$. On note $\tilde{P}_{n,k}^{(h)}(U(a))$ la restriction de $P_{n,k}^{(h)}(\tilde{U}(a))$ à $H^1(K_\infty, V(k))$.
Si maintenant $g = h + \sum_{i \in \mathbb{Z}} U(a_i)$ est un élément de $\mathcal{D}_{\infty,g}(V)$ et si $\overline{G} = \overline{H} + \sum_{i \in \mathbb{Z}} \tilde{U}(a_i)$ est une solution compatible des équations $(1 - p^r\Phi)\mathcal{G} = D^r(g)$, on étend les définitions de $\Xi_{n,k}^{(h)}$ et de $P_{n,k}^{(h)}$ par linéarité. On a donc $\tilde{P}_{n,k}^{(h)}(\overline{G}) = \tilde{P}_{n,k}^{(h)}(\overline{H}) + \sum_{i \in \mathbb{Z}} \tilde{P}_{n,k}^{(h)}(U(a_i))$.

3.3. Théorème. *Soit V une représentation de de Rham et soit h un entier ≥ 1 tel que $\mathrm{Fil}^{-h}\,\mathbf{D}_{\mathrm{dR}}(V) = \mathbf{D}_{\mathrm{dR}}(V)$. Soient $g \in \mathcal{D}_{\infty,f}(V)$ et $\overline{G}$ une solution compatible des équations $(1 - p^r\Phi)\mathcal{G}_r = D^r(g)$. Soit u un entier $> -h$ tel que g soit $p^{-u}\varphi^-$-bornée. Si g n'est pas dans $\mathcal{D}_{\infty,e}(V)$, on suppose de plus que $D^h(g)(0) \in (1 - p^h\varphi)\mathbf{D}_p(V)$. La famille $(P_{n,k}^{(h)}(\overline{G}))_{n \geq 1, k \geq -h+1}$ est tempérée d'ordre $\leq (u+h)^-$ et définit un élément $\Omega_{V,h}(g)$ de $\mathcal{H}_{(u+h)^-}(G_\infty) \otimes \tilde{Z}^1_\infty(K, T)$ ne dépendant que de g. Ainsi, on a*

$$\tilde{\pi}_{n,k}(\Omega_{V,h}(g)) = \tilde{P}_{n,k}^{(h)}(g)$$

pour $n \geq 1$ et pour $k \in \{1 - h, \cdots, +\infty\}$.

On a

$$\mathfrak{o}(\Omega_{V,h}(g)) \leq h + \mathfrak{o}_\varphi(g)$$

et lorsque V est cristalline,

$$\mathfrak{o}(\Omega_{V,h}(g)) = h + \mathfrak{o}_\varphi(g) \ .$$

Ainsi, si r_0 est un entier tel que la suite d'opérateurs $p^{nr_0}\varphi^{-n}$ de $\mathbf{D}_p(V)$ est bornée et si g appartient de plus à $\mathcal{H}_{s^-}^{\psi=0} \otimes \mathbf{D}_p(V)$, $\Omega_{V,h}(g)$ appartient à $\mathcal{H}_{(r_0+s+h)^-}(G_\infty) \otimes \tilde{Z}^1_\infty(T)$. En effet, g est alors $p^{-(r_0+s)}\varphi^-$-borné. Remarquons que r_0 peut être négatif, mais on a nécessairement $r_0 + h \geq 0$ et donc $r_0 + s > -h$. Par exemple, si $V = \mathbb{Q}_p(r)$, on peut prendre $h = r$ et $r_0 = -r$. La relation entre les ordres de tempérance signifie : $\Omega_{V,h}(g)$ est un $o(\log^s)$ si et seulement si g est $p^{-(s-h)}\varphi^-$-bornée et $\Omega_{V,h}(g)$ est un $O(\log^s)$ si et seulement si g est $p^{-(s-h)}\varphi$-bornée.

REMARQUES : 1) Cet homomorphisme est $(-1)^{h-1}$ fois le $\Omega_{V,h}$ de [4]. Cela permet d'éliminer certains signes : par exemple, il n'est pas difficile de déduire du théorème la relation suivante entre $\Omega_{V,h+1}$ et $\Omega_{V,h}$:

$$\Omega_{V,h+1} = \ell_h \Omega_{V,h}$$

avec

$$\ell_h = \frac{\log\gamma}{\log\chi(\gamma)} - h = \frac{\log\chi(\gamma)^{-h}\gamma}{\log\chi(\gamma)} = Tw^{-h}(\frac{\log\gamma}{\log\chi(\gamma)})$$

(attention au changement de signe par rapport à [4]). On pose pour tout entier r,

$$\Omega_{V,r} = (\prod_{r\leq j<h} \ell_j)^{-1}\Omega_{V,h} \tag{3.3.1}$$

pour $h > r$ et tel que $\mathrm{Fil}^{-h}\,\mathbf{D}_p(V) = \mathbf{D}_p(V)$. C'est un élément de $\mathcal{K}(G_\infty) \otimes \tilde{Z}^1_\infty(K,T)$ avec $\mathcal{K}(G_\infty)$ l'anneau des factions total de $\mathcal{H}(G_\infty)$ (il suffit en fait d'inverser les ℓ_j). De même, avec des identifications convenables, on a

$$\Omega_{V(j),h+j}(g) = Tw^j(\Omega_{V,h}(D^j(g))) \ .$$

2) La lettre Ω évoque pour certains une période : isomorphisme de périodes entre $\mathcal{H}_\infty \otimes \mathbf{D}_p(V)$ et $\mathcal{H}(G_\infty) \otimes \tilde{Z}^1_\infty(K,V)$. On peut le voir aussi comme une manière de mettre ensemble toutes les exponentielles de Bloch-Kato relatives à V et à ses twists cyclotomiques, d'où la notation $\mathrm{Exp}_{h,V}$ de [2]. Il est d'ailleurs amusant de remarquer que dans [2], c'est le point de vue "matrice de périodes" de cet homomorphisme qui est utilisé.
3) Ici, on n'a pas supposé V cristalline mais si V ne l'est pas, la dimension de $K_n \otimes \mathbf{D}_p(V)$ est de dimension sur K_n strictement inférieure à la dimension de V et donc le rang de $\Lambda \otimes \mathbf{D}_p(V)$ est strictement inférieure à celui de $Z^1_\infty(K,T)$.

Nous ne démontrons dans les paragraphes qui suivent que l'inégalité
$$\mathfrak{o}(\Omega_{V,h}(g)) \leq h + \mathfrak{o}_\varphi(g)\ ;$$
l'égalité dans le cas où V est cristalline sera une conséquence de la loi de réciprocité (cf. 4.2.4).
4) Si $\mathbf{D}_p(V)^{\varphi=p^{-h}} \neq 0$, V contient alors $\mathbb{Q}_p(h)$. Dans ce cas, si $D^h(g)(0) \notin (1-p^h\varphi)\mathbf{D}_p(V)$, on peut définir $\Omega_{V,h}(g)$ à valeurs dans $(\chi(\gamma)^{-h}\chi - 1)^{-1}\mathcal{H}_{(u+h)^-}(G_\infty) \otimes \tilde{Z}^1_\infty(T)$ par
$$\Omega_{V,h}(g) = (\chi(\gamma)^{-h}\gamma - 1)^{-1}\Omega_{V,h}((\chi(\gamma)^{-h}\chi - 1)g)\ .$$
Remarquons qu'alors $\Omega_{V,h+1}(g)$ est défini directement par le théorème, ce qui est cohérent avec la relation $\Omega_{V,h+1}(g) = \ell_h\Omega_{V,h}(g)$: $\Omega_{V,h+1}(g)$ n'a plus de pôles. Donnons maintenant les formules qui s'en déduisent pour $n = 0$:

3.3.1. Proposition. *Sous les hypothèses du théorème* 3.3, *on a*
$$\tilde{\pi}_{0,k}(\Omega_{V,h}(g)) = \exp_{V(k),f}(\Xi^{(h)}_{0,k}(\overline{G}))$$
avec
$$\begin{aligned}\Xi^{(h)}_{0,k}(\overline{G}) =& (-1)^{h+k-1}(h+k-1)! \times \\ & ((1-p^{k+1}\varphi^{-1})b_{-k} \otimes e_{-k}, (1-p^{k+1}\varphi^{-1})D^{-k}(G)(0) \otimes e_{-k})\ .\end{aligned}$$
En particulier, si $1 - p^{-k}\varphi$ *est un isomorphisme sur* $\mathbf{D}_p(V)$, *on a*
$$\tilde{\pi}_{0,k}(\Omega_{V,h}(g)) = \exp_{V(k),e}(\Xi^{(h)}_{0,k}(\overline{G}))$$
avec
$$\begin{aligned}\Xi^{(h)}_{0,k}(\overline{G}) =& (-1)^{h+k-1}(h+k-1)! \times \\ & (1-p^{k+1}\varphi^{-1})(1-p^{-k}\varphi)^{-1}D^{-k}(g)(0) \otimes e_{-k}\ .\end{aligned}$$

La proposition se déduit de l'équation fonctionnelle reliant G et g et de ce que $\psi(g) = 0$ (voir (4.3.2)).
Avant de commencer la démonstration du théorème, expliquons comment on peut traiter le cas où $g \in \mathcal{D}_{\infty,g}(\mathbf{D}_p(V))$. Il s'agit de définir l'image de $U(a)$ pour $a \in \mathbf{D}_p(V)^{\varphi=p^i}$. Pour cela, plutôt que de vérifier les congruences, ce que nous n'avons pas su faire, on définit directement $\Omega_{V,h}(U(a))$ puis on vérifie que $\pi_{n,k}(\Omega_{V,h}(U(a))$ est bien $\exp_{V(k),g}(\Xi^{(h)}_{n,k}(\tilde{U}(a)))$.
Pour cela, on commence par énoncer le lemme suivant :

Lemme. *Soit* τ *un élément de* G_∞ *non de torsion. Soit* $u \in \mathcal{H}(G_\infty) \otimes Z^1_\infty(K,T)$ *tel que* $\pi_{0,0}(u) = 0$. *Alors, il existe* $v_\tau \in \mathcal{H}(G_\infty) \otimes Z^1_\infty(K,T)$ *tel que* $(\tau - 1)v_\tau = u$.

On peut exprimer le lemme sous la forme suivante en le twistant : si $\pi_{0,j}(u) = 0$, il existe v_τ tel que $(\chi(\tau)^j\tau - 1)v_\tau = u$. Remarquons qu'on a alors pour tout

entier k la formule $(\chi(\tau)^{j-k}-1)\pi_{n,k}(v_\tau)=\pi_{n,k}(u)$ pour τ laissant fixe K_n, c'est-à-dire que pour tout entier $k\neq j$, $\pi_{n,k}(v_\tau)=\frac{1}{\chi(\tau)^{j-k}-1}\pi_{n,k}(u)$.
Enfin, remarquons que l'on a unicité de v_τ si l'on ne regarde que son image dans $\mathcal{H}(G_\infty)\otimes\tilde{Z}^1_\infty(K,T)$ et que si u est tempérée d'ordre $\leq r$, il en est de même de v_τ.
Appliquons ce lemme à $x_\tau=\Omega_{V,h}(D^{i+1}\Pi_\tau a)$ pour $a\in\mathbf{D}_p(V)^{\varphi=p^i}$. Par twist, on peut se ramener au cas où $i=-1$.
Comme $\pi_{n,0}(x_\tau)=\exp_{V,e}((-1)^{h-1}(h-1)!\ G_\tau(\zeta_n-1)a)$ pour $n\geq 1$ avec $G_\tau=\log\frac{(1+T)^{\chi(\tau)}-1}{T}$, on a

$$\pi_{0,0}(x_\tau)=Tr_{K_1/K}(\exp_{V,e}((-1)^{h-1}(h-1)!\log\frac{\zeta_1^{\chi(\tau)}-1}{\zeta_1-1}a))=0\ .$$

Il existe donc $y\in\mathcal{H}_\infty(G_\infty)\otimes\tilde{Z}_\infty(K,T)$ tel que $(\tau-1)y=x_\tau$ et on vérifie facilement que y ne dépend pas de τ. On pose alors $\Omega_{V,h}(U(a))=y$. On a donc

$$\Omega_{V,h}(U(a))=(\tau-1)^{-1}\Omega_{V,h}(\Pi_\tau(a)$$

et en général pour $a\in\mathbf{D}_p(V)^{\varphi=p^i}$

$$\Omega_{V,h}(U(a))=(\chi(\tau)^{i+1}\tau-1)^{-1}\Omega_{V,h}(D^{i+1}\Pi_\tau(a)\ .$$

On définit ainsi un prolongement de $\Omega_{V,h}$ à $\mathcal{D}_{\infty,g}(\mathbf{D}_p(V))$. La formule

$$\exp_{V(k),f}(\Xi^{(h)}_{n,k}(U(a)))=\pi_{n,k}(\Omega_{V,h}(U(a)))$$

est claire pour $k\neq i+1$. Pour $k=i+1$, nous la montrerons plus tard en utilisant la loi de réciprocité.

3.4. Démonstration du théorème.

3.4.1. Il s'agit de montrer que les points $P_{n,k}(\overline{G})$ vérifient les conditions de la proposition 1.8. La propriété que

$$Tr_{n+1,n}(P^{(h)}_{n+1,k}(\overline{G}))=P^{(h)}_{n,k}(\overline{G})$$

dans $H^1(K_n,V(k))$ se déduit de la condition $\psi(g)=0$. Pour montrer que les points $P_{n,k}(\overline{G})$ sont admissibles, il suffit de le faire pour $P_{0,k}(\overline{G})=Tr_{K_n/K}(P_{n,k}(\overline{G}))$. Prenons $k=0$ pour simplifier (on s'y ramène en remplaçant V par $V(k)$). Ce point est admissible si et seulement son accouplement local avec un élément v de $V^*(1)^{G_K}\cong H^1(\Gamma,V^*(1)^{G_K})\subset H^1(K,V^*(1))$ est nul (§1.6). Pour le calculer, nous utilisons les formules de Kato qui sont rappelées en 4.1.3. On en déduit que si v est vu comme élément de $\mathrm{Fil}^0\,\mathbf{D}_p(V^*(1))^{\varphi=1}\subset\mathbf{D}_p(V^*(1))$, cet accouplement est de la forme $[(1-p\varphi^{-1})(u),v]=[u,(1-\varphi)v]=0$.
Démontrons les congruences vérifiées par les $P^{(h)}_{n,k}$. Comme me l'a fait remarqué Colmez, il y a une démonstration beaucoup plus simple au niveau des calculs que celle faite dans [4]. Nous allons commencer par celle-là. Cependant, nous

avons besoin pour la démonstration de la loi de réciprocité du calcul explicite du cocycle fait dans [4]. Aussi, ferons-nous ensuite ce calcul.
Notons $[\epsilon]$ le relèvement de Teichmüller de $\epsilon = (\zeta_n)$ dans $A_{\max}$ et $\beta_n = \varphi^{-n}([\epsilon]) = [(\zeta_{m+n})_m]$. Il s'agit donc d'un relèvement de la racine de l'unité ζ_n dans $B_{\max}$ puisque l'image de β_n dans $\overline{K} \subset B_{\rm dR}/B_{\rm dR}^+$ est ζ_n.
Prenons $g \in \mathcal{D}_{\infty,e}(V)$. On se donne une solution compatible $\overline{G}$ de l'équation$(1-\Phi)\mathcal{G} = g$, c'est-à-dire des éléments G_r de $\mathcal{H}_\infty \otimes \mathbf{D}_p(V)$ vérifiant $D(G_r) = G_{r+1}$ et $(1-p^r\Phi)G_r = D^r(g)$. On pose $D^r(\overline{G}) = D^r(G) = G_r$. Le point $P_{n,k}^{(h)}(\overline{G})$ est la classe du cocycle

$$(-1)^{h+k-1}(h+k-1)!p^{n(k-1)}e_B((1\otimes\varphi)^{-n}G_{-k}(\zeta_n-1)\otimes e_{-k})\ .$$

Pour démontrer les congruences, il suffit de montrer que pour $s' \geq h+u$ et pour $\tau \in G_{K_\infty}$, la limite de la suite

$$p^{n(s'-(j+h-1))}(\tau-1)e_B(\sum_{k=1-h}^{j}(-1)^{k+h-1}\binom{j+h-1}{k+h-1}\Xi_{n,k}^{(h)}(\overline{G})t^{-k})$$

tend vers 0 lorsque $n\to\infty$. (rappelons que $\Xi_{n,k}^{(h)}(G) \in \mathbf{D}_p(V(k))$ est identifié à $\Xi_{n,k}^{(h)}(G)t^{-k}e_k \in B_{\max}\otimes\mathbf{D}_p(V)$). Notons

$$\mathcal{Y}_j^h = \sum_{k=1-h}^{j}(-1)^{k+h-1}\binom{j+h-1}{k+h-1}\Xi_{n,k}^{(h)}(\overline{G})t^{-k}$$

On a

$$\begin{aligned}\mathcal{Y}_j^h &= (j+h-1)!\sum_{k=1-h}^{j}\frac{p^{n(k-1)}}{(j-k)!}(1\otimes\varphi)^{-n}G_{-k}(\zeta_n-1)t^{-k}\\ &= (j+h-1)!p^{n(j-1)}\sum_{k=0}^{h+j-1}\frac{p^{-nk}}{k!}(1\otimes\varphi)^{-n}D^k(G_{-j})(\zeta_n-1)t^{k-j}\end{aligned}$$

en changeant k en $j-k$.
Soit $H \in \mathcal{H}_\infty\otimes\mathbf{D}_p(V)$. Posons $\tilde{H}(Z) = H(\zeta_n\exp(Z)-1)$: comme $\beta_n = \zeta_n\exp(t/p^n)$, on a $H(\beta_n-1) = \tilde{H}(t/p^n)$. Si $\mathcal{T}_{h-1}(\tilde{H})$ est le développement de Taylor d'ordre $h-1$ de $\tilde{H}$ en 0, on vérifie facilement que

$$\tilde{H}(t/p^n) - \mathcal{T}_{h-1}(\tilde{H})(t/p^n) \in t^hB_{\rm dR}^+\otimes\mathbf{D}_p(V) = \mathrm{Fil}^h B_{\rm dR}\otimes\mathbf{D}_p(V)$$

et que

$$\mathcal{T}_{h-1}(\tilde{H})(t/p^n) = \sum_{i=0}^{h-1}\frac{1}{i!}D^i(H)(\zeta_n-1)\frac{t^i}{p^{ni}}$$

On en déduit que

$$\sum_{i=0}^{h-1}\frac{1}{i!}D^i(H)(\zeta_n-1)\frac{t^i}{p^{ni}} - H(\beta_n-1) \in \mathrm{Fil}^h(B_{\rm dR})\otimes\mathbf{D}_p(V)\ ,$$

et en remplaçant h par $h+j$, que

$$\sum_{i=0}^{h+j-1}\frac{1}{i!}D^i(H)(\zeta_n-1)\frac{t^i}{p^{ni}}t^{-j}-H(\beta_n-1)t^{-j}\in \mathrm{Fil}^h(B_{\mathrm{dR}})\otimes \mathbf{D}_p(V))\ .$$

Remarquons que la condition sur h implique que $\mathrm{Fil}^h B_{\mathrm{dR}}\otimes\mathbf{D}_p(V)$ est contenu dans $\mathrm{Fil}^0(B_{\mathrm{dR}}\otimes\mathbf{D}_p(V))$. En appliquant cela à $H=(1\otimes\varphi)^{-n}G_{-j}$ et en utilisant le fait que e_B est nul sur $\mathrm{Fil}^0(B_{\mathrm{dR}}\otimes\mathcal{D})$, on obtient que

$$\begin{aligned}e_B(\mathcal{Y}_j^h)&=e_B((j+h-1)!p^{n(j-1)}(1\otimes\varphi)^{-n}G_{-j}(\beta_n-1)t^{-j})\\&=(j+h-1)!p^{n(j-1)}e_B(\Phi^{-n}(G_{-j}(\epsilon-1))t^{-j}\end{aligned}$$

Par définition de e_B et de Eul et comme $\Phi^{-n}(G_{-j}(\epsilon-1))t^{-j}$ appartient à $(B_{\max}\otimes\mathbf{D}_p(V))^{G_{K_\infty}}$, on a

$$(\tau-1)e_B(\mathcal{Y}_j^h)=-(j+h-1)!p^{n(j-1)}(\tau-1)Eul(\Phi^{-n}(D^{-j}(g)([\epsilon]-1))t^{-j})\ .$$

Il s'agit donc de montrer par continuité de Eul et stabilité de $A_{\max}$ par G_K que pour $s'-h\geq u$, la suite $p^{n(s'-h)}\Phi^{-n}(D^{-j}(g)([\epsilon]-1))$ tend vers 0 lorsque $n\to\infty$. On applique pour cela le lemme 3.1.1 à $D^{-j}(g)$ qui est $p^{-u}\varphi^{-}$-bornée :

$$||p^{n(s'-h)}\Phi^{-n}(D^{-j}(g)([\epsilon]-1))||_{\max}=p^{n(h+u-s')}||p^{nu}(1\otimes\varphi)^{-n}D^{-j}(g)||_{\rho_n}$$

tend vers 0 lorsque $n\to\infty$.
Il n'est pas difficile de voir que la même démonstration s'applique à $g\in\mathcal{D}_{\infty,f}(V)$.

3.4.2. Comme annoncé, nous allons maintenant reprendre la démonstration en calculant explicitement un cocycle représentant $P_{n,k}^{(h)}$ pour $g\in\mathcal{D}_{\infty,f}(V)$. On se donne donc une solution compatible $\overline{G}$, c'est-à-dire $b=(b_r)\in\oplus_{r\in\mathbb{Z}}\mathbf{D}_p(V)$ et des éléments G_r de $\mathcal{H}_\infty\otimes\mathbf{D}_p(V)$ vérifiant $D(G_r)=G_{r+1}$ et $(1-p^r\Phi)G_r=D^r(g)+L_r(b)$ avec $L_r(b)=\sum_{i\geq r}\frac{b_i}{(i-r)!}\log^{i-r}(1+T)$. Il est commode de noter formellement $L(b)=\sum_i\frac{b_i}{i!}\log^i(1+T)$ et $L(b)_r=D^r(L(b))=L_r(b)$. On pose $\overline{G}=(b,G)$ avec $G=(G_r)$ et $D^r(\overline{G})=(L_r(b),D^r(G))=(L_r(b),G_r)$. Posons

$$\begin{aligned}\mathcal{S}_0^{(h)}(G)=&(-1)^{h-1}(h-1)!\sum_{i=0}^{h-1}(-1)^i\frac{1}{i!}D^i(G)([\epsilon]-1)t^i\\ \mathcal{S}_k^{(h)}(G)=&(-1)^{h+k-1}(h+k-1)!\sum_{i=0}^{h+k-1}(-1)^i\frac{1}{i!}D^{i-k}(G)([\epsilon]-1)t^{i-k}\\ =&(-1)^{h-1}(h+k-1)!\sum_{i=-k}^{h-1}(-1)^i\frac{1}{(i+k)!}D^i(G)([\epsilon]-1)t^i\end{aligned}$$

et

$$\begin{aligned}\mathcal{S}^{(h)}_{n,0,cris}(G) =&(p\Phi)^{-n}(S^{(h)}_0(G))\\ \mathcal{S}^{(h)}_{n,k,cris}(G) =&(p\Phi)^{-n}(S^{(h)}_k(G))\end{aligned}$$

Ainsi, $\mathcal{S}^{(h)}_k(G) = \mathcal{S}^{(h+k)}_0(D^{-k}(G))t^{-k}$.
Lorsque $h = 1$, les formules se simplifient et deviennent

$$\begin{aligned}\mathcal{S}^{(1)}_{n,0,cris}(G) =&(p\Phi)^{-n}(G_0([\epsilon]-1))\\ \mathcal{S}^{(1)}_{n,k,cris}(G) =&(p\Phi)^{-n}(k!\sum_{i=0}^{k}(-1)^{i-k}\frac{1}{i!}D^{i-k}(G)([\epsilon]-1)t^{i-k})\\ =&(p\Phi)^{-n}(k!\sum_{u=0}^{k}(-1)^{u}\frac{1}{(k-u)!}D^{-u}(\overline{G})([\epsilon]-1)t^{-u})\ .\end{aligned}$$

Notons $b^{(r)}$ la suite b où l'on a remplacé le r-ième terme par 0. On définit par les mêmes formules $\mathcal{S}^{(h)}_{n,k,cris}(L(b))$ et $\mathcal{S}^{(h)}_{n,k,cris}(L(b^{(r)}))$.

3.4.3. Lemme. *Supposons comme dans le théorème que* $\mathrm{Fil}^{-h}\mathbf{D}_{\mathrm{dR}}(V) = \mathbf{D}_{\mathrm{dR}}(V)$ *et que* $h+k-1\geq 0$. *Alors,*
(i) $P^{(h)}_{n,k}(\overline{G})$ *est la classe du cocycle*

$$\tau\in G_{K_n}\mapsto(\tau-1)(\mathcal{S}^{(h)}_{n,k,cris}(G)-Eul(\mathcal{S}^{(h)}_{n,k,cris}(g+L(b^{(-k)}))))\ ;$$

(ii) $res_\infty(P^{(h)}_{n,k}(\overline{G}))$ *est la classe du cocycle*

$$\tau\in G_{K_\infty}\mapsto -(\tau-1)Eul(\mathcal{S}^{(h)}_{n,k,cris}(g)-p^{-n}\frac{D^h(g)(0)t^h}{h+k})$$

Si $D^h(g)(0)\in(1-p^h\varphi)\mathbf{D}_p(V)$, $res_\infty(P^{(h)}_{n,k}(\overline{G}))$ *est la classe du cocycle*

$$\tau\in G_{K_\infty}\mapsto -(\tau-1)Eul(\mathcal{S}^{(h)}_{n,k,cris}(g))\ .$$

Démonstration. Soit $H\in\mathcal{H}_\infty\otimes\mathbf{D}_p(V)$. Posons $\tilde{H}(Z)=H(\beta_n\exp(-Z)-1)$: en utilisant la formule $\zeta_n=\beta_n\exp(-t/p^n)$, on a $H(\zeta_n-1)=\tilde{H}(t/p^n)$. La même démonstration que précédemment donne que

$$H(\zeta_n-1)-\sum_{i=0}^{h-1}\frac{(-1)^i}{i!}D^i(H)(\beta_n-1)\frac{t^i}{p^{ni}}\in t^hB^+_{\mathrm{dR}}\otimes\mathbf{D}_p(V)$$

En appliquant cette formule à $D^{-k}(G)$ et à $h+k-1\geq 0$, on en déduit que

$$\Xi^h_{n,k}(G)-\mathcal{S}^{(h)}_{n,k,cris}(G)\in\mathrm{Fil}^h(B_{\mathrm{dR}})\otimes\mathbf{D}_p(V)\subset\mathrm{Fil}^0(B_{\mathrm{dR}}\otimes\mathbf{D}_p(V(k)))$$

De plus, $\mathcal{S}^{(h)}_{n,k,cris}(G)$ est invariant par G_{K_∞}. En utilisant la formule $(1-p^r\Phi)D^r(G) = D^r(g) + L_r(b)$, on a

$$(1-\Phi)\mathcal{S}^{(h)}_{n,k,cris}(G) = \mathcal{S}^{(h)}_{n,k,cris}(g) + \mathcal{S}^{(h)}_{n,k,cris}(L(b)) \ .$$

En revenant à la définition de $\exp_f$ et en remarquant que

$$\mathcal{S}^{(h)}_{n,k,cris}(L(b)) - (-1)^{h+k-1}(h+k-1)!(p\varphi)^{-n}b_{-k}t^{-k} = \mathcal{S}^{(h)}_{n,k,cris}(L(b^{(-k)})) \ ,$$

on obtient que $res_\infty(P^{(h)}_{n,k}(\overline{G}))$ est la classe du cocycle

$$\tau \in G_{K_\infty} \mapsto -(\tau-1)Eul(\mathcal{S}^{(h)}_{n,k,cris}(g + L(b^{(-k)}))) \ .$$

Calculons maintenant $\mathcal{S}^{(h)}_{n,k,cris}(L(b^{(-k)})$. On a

$$\begin{aligned}
&\mathcal{S}^{(h)}_{n,k,cris}(L(b))\\
&\qquad = (-1)^{h+k-1}(h+k-1)! \sum_{i=0}^{h+k-1} \frac{(-1)^i}{i!}\Big(\sum_{u \geq i-k} \frac{b_u}{(u-i+k)!}t^{u-i+k}\Big)t^{i-k}\\
&\qquad = (-1)^{h+k-1}(h+k-1)! \sum_{i=0}^{h+k-1} \frac{(-1)^i}{i!} \sum_{u \geq i-k} \frac{b_u}{(u-i+k)!}t^{u}\\
&\qquad = (-1)^{h+k-1}(h+k-1)! \sum_{u \geq -k} \alpha^h_{u,k} b_u t^u
\end{aligned}$$

avec

$$\alpha^h_{u,k} = \sum_{\substack{i \leq u+k \\ 0 \leq i \leq h+k-1}} \frac{(-1)^i}{i!(u-i+k)!} \ .$$

Lorque $u \leq h-1$, on a

$$(u+k)!\alpha^h_{u,k} = \sum_{0 \leq i \leq u+k} \frac{(-1)^i(u+k)!}{i!(u-i+k)!} = (1-1)^{u+k} = \begin{cases} 0 & \text{si } -k < u \leq h-1 \\ 1 & \text{si } u = -k \end{cases}$$

Ainsi,

$$\mathcal{S}^{(h)}_{n,k,cris}(L(b)) = (h+k-1)!(-1)^{h+k-1}(b_{-k}t^{-k} + \sum_{u \geq h} \alpha^h_{u,k} b_u t^u)$$

et

$$\mathcal{S}^{(h)}_{n,k,cris}(L(b^{(k)})) = (-1)^{h+k-1}(h+k-1)! \sum_{u \geq h} \alpha^h_{u,k} b_u t^u$$

et $\alpha^h_{h,k} = \frac{(-1)^{h+k-1}}{(h+k)!}$. On en déduit que

$$\mathcal{S}^{(h)}_{n,k,cris}(L(b^{(k)})) \in t^h B^{+G_{K_\infty}}_{\max} \otimes \mathbf{D}_p(V) \subset \mathrm{Fil}^0(B_{\max} \otimes \mathbf{D}_p(V)) \ .$$

Plus précisément, pour $u > h$ ou pour $b_h \in (1-p^h\varphi)\mathbf{D}_p(V)$, $b_u t^u$ appartient à $(1-\varphi)\,\mathrm{Fil}^0(B^{+G_{K_\infty}}_{\max} \otimes \mathbf{D}_p(V))$, car $1-p^u\varphi$ est alors un isomorphisme de

$\mathbf{D}_p(V)$, et on en déduit que $(\tau-1)Eul(\mathcal{S}^{(h)}_{n,k,cris}(L(b^{(k)}))=0$ pour $\tau\in G_{K_\infty}$ et $res_\infty(P^{(h)}_{n,k}(\overline{G}))$ est simplement la classe du cocycle

$$\tau\in G_{K_\infty}\mapsto -(\tau-1)Eul(\mathcal{S}^{(h)}_{n,k,cris}(g)+p^{-n(h+1)}\frac{\varphi^{-n}(b_h)t^h}{h+k})$$

et si $b_h\in(1-p^h\varphi)\mathbf{D}_p(V)$, c'est la classe du cocycle

$$\tau\in G_{K_\infty}\mapsto -(\tau-1)Eul(\mathcal{S}^{(h)}_{n,k,cris}(g))\ .$$

Sans condition sur b_h, remarquons que $b_h+D^h(g)(0)\in(1-p^h\varphi)\mathbf{D}_p(V)$ et pour les mêmes raisons,

$$\begin{aligned}&(\tau-1)Eul(\varphi^{-n}(b_h)t^h)\\&\qquad=(\tau-1)Eul(p^{nh}b_ht^h)=-p^{nh}(\tau-1)Eul(D^h(g)(0)t^h)\ .\end{aligned}$$

On en déduit le lemme. □

3.4.4. Corollaire. *La restriction de* $P^{(h)}_{n,k}(\overline{G})$ *à* $H^1(K_\infty,V(k))$ *ne dépend que de* g*, on la note* $\tilde{P}^{(h)}_{n,k}(g)$.

3.4.5. Plaçons-nous sous les hypothèses du théorème en supposant que $D^h(g)(0)\in(1-p^h\varphi)\mathbf{D}_p(V)$. Montrons que pour $s'\geq h+u$, la suite

$$\begin{aligned}&p^{n(s'-(j+h-1))}\sum_{k=1-h}^{j}(-1)^{k+h-1}\binom{j+h-1}{k+h-1}\mathcal{S}^{(h)}_{n,k,cris}(g)\\&\qquad=p^{n(s'-j-h+1)}\sum_{k=0}^{j+h-1}(-1)^k\binom{j+h-1}{k}\mathcal{S}^{(h)}_{n,k-h+1,cris}(g)\end{aligned}$$

tend vers 0. Notons

$$\begin{aligned}&\mathcal{Z}^{(h)}_j=\\&\sum_{k=0}^{j+h-1}(-1)^k\binom{j+h-1}{k}k!(-1)^k\sum_{i=0}^{k}(-1)^i\frac{1}{i!}D^{i-k+h-1}(g)([\epsilon]-1)t^{i-k+h-1})\ .\end{aligned}$$

On a donc $(p\Phi)^{-n}(\mathcal{Z}^{(h)}_j)=\sum_{k=1-h}^{j}(-1)^{k+h-1}\binom{j+h-1}{k+h-1}\mathcal{S}^{(h)}_{n,k,cris}(g)$. On a en changeant i en $k-i$

$$\begin{aligned}&\mathcal{Z}^{(h)}_j=\\&=\sum_{k=0}^{j+h-1}(-1)^k\binom{j+h-1}{k}k!\sum_{i=0}^{k}(-1)^i\frac{1}{(k-i)!}D^{-i+h-1}(g)([\epsilon]-1)t^{-i+h-1})\\&=\sum_{i=0}^{j+h-1}(-1)^iv_{i,j+h-1}D^{-i+h-1}(g)([\epsilon]-1)t^{-i+h-1}\end{aligned}$$

avec

$$\begin{aligned}
v_{i,t} =& \sum_{k=i}^{t} (-1)^k \binom{t}{k} \frac{k!}{(k-i)!} \\
=& \sum_{k=i}^{t} (-1)^k \binom{u}{k} k(k-1)\dots(k-i+1) \\
=& (\frac{d^i}{dX^i}(1-X)^t)_{|X=1}
\end{aligned}$$

Donc, seul $v_{t,t}$ est non nul, il vaut $(-1)^t t!$ et on obtient finalement

$$\begin{aligned}
p^{-nj}(p\Phi)^{-n}(\mathcal{Z}_j^{(h)}) =& p^{-nj}(p\Phi)^{-n}((j+h-1)! D^{-j}(g)([\epsilon]-1)t^{-j}) \\
=& (p\Phi)^{-n}((j+h-1)! D^{-j}(g)([\epsilon]-1))t^{-j}
\end{aligned}$$

Il s'agit donc de montrer que $p^{n(s'-1-h+1)}\Phi^{-n}(D^{-j}(g)([\epsilon]-1))$ tend vers 0 pour $s' \geq h+u$. On applique pour cela le lemme 3.1.1 à $D^{-j}(g) \in \mathcal{H}_\infty^{\psi=0} \otimes \mathbf{D}_p(V)$ qui est $p^{-u}\varphi^-$-bornée pour obtenir que

$$||p^{n(s'-h)}\Phi^{-n}(D^{-j}(g)([\epsilon]-1))||_{\max} = p^{n(h+u-s')}||p^{nu}(1\otimes\varphi)^{-n}D^{-j}(g)||_{\rho_n}$$

tend vers 0 lorsque $n \to \infty$. Cela termine la démonstration du théorème 3.3.

3.4.6. Revenons sur le cas où $D^h(g)(0) \notin (1-p^h\varphi)\mathbf{D}_p(V)$. Rappelons que le fait que $\mathbf{D}_p(V)^{\varphi=p^{-h}} \neq 0$ avec $\mathrm{Fil}^{-h}\mathbf{D}_p(V) = \mathbf{D}_p(V)$ implique que $V(-h)^{G_K}$ est non nul.

On voit alors apparaître dans le cocycle définissant le point $P_{n,k}^{(h)}(G)$ un terme de la forme $\frac{c_{k,h}}{p^n}$. Ce terme est signe de l'existence d'un pôle dans $\Omega_{V,h}(g)$. Il disparaît si l'on remplace g par $\tilde{g} = (\chi(\gamma)^{-h}\gamma - 1)g$ (on a alors $D^h(\tilde{g})(0) = 0$), d'où la définition

$$\Omega_{V,h}(g) = (\chi(\gamma)^{-h}\gamma - 1)^{-1}\Omega_{V,h}(\tilde{g}) \ .$$

Il disparaît aussi lorsqu'on remplace par h par $h+1$ et cela s'explique par la formule :

$$\Omega_{V,h+1}(g) = \ell_h \Omega_{V,h}(g)$$

et le fait que $\chi(\gamma)^{-h}\gamma - 1$ divise l_h. Que peut-on dire du résidu, c'est-à-dire de $\pi_{0,-h}((\chi(\gamma)^{-h}\gamma-1)\Omega_{V,h}(g)$? D'après la deuxième formule, il s'agit de

$$\begin{aligned}
\tilde{\pi}_{0,-h}(\Omega_{V,h+1}(g)) = \tilde{P}_{0,-h}^{(h+1)} =& Tr_{K_1/K_0}(\tilde{P}_{0,-h}^{(h+1)}) \\
=& res_\infty \exp_{V(-h),f}(b_h \otimes e_h, (1-p^{-h+1}\varphi^{-1})D^h(\overline{G})(0)\otimes e_h) \ .
\end{aligned}$$

Le cocycle associé (restreint à K_∞) est

$$\tau \mapsto (\tau - 1)Eul(D^h(g)([\epsilon]-1)t^h) \ .$$

On peut supposer que $b_h = D^h(g)(0)$ car

$$D^h(g)([\epsilon]-1)t^h \equiv \\ (1-\Phi)(D^h(G)([\epsilon]-1)t^h) + D^h(g)(0)t^h \mod (1-\Phi)(t^h \otimes \mathbf{D}_p(V)) .$$

Le premier terme disparaît lorsqu'on applique $(\tau-1)Eul$ pour $\tau \in G_{K_\infty}$, on obtient donc

$$(\tau-1)Eul(D^h(g)(0)t^h) .$$

Dans ce cas, $Eul(D^h(g)(0)t^h)$ appartient en fait à $\hat{\mathbb{Q}}_p^{nr} \otimes \mathbf{D}_p(V(-h)) = \hat{\mathbb{Q}}_p^{nr} t^h \otimes \mathbf{D}_p(V)e_h$, où $\hat{\mathbb{Q}}_p^{nr}$ est le complété de l'extension maximale non ramifiée de $\mathbb{Q}_p$, c'est-à-dire que comme il est bien connu, on n'a pas besoin de passer dans ce cas à B_{cris}. Par exemple, prenons $V = \mathbb{Q}_p(h)$ et $D^h(g)(0) = 1$, on est donc en train de construire un élément de $H^1(K_\infty, \mathbb{Q}_p)^{G_\infty}$ par la recette

$$\tau \mapsto (\tau-1)Eul(t^h) .$$

Il s'agit en fait de résoudre l'équation $(1-\varphi)\Omega = 1$, ce qui se résoud dans $\hat{\mathbb{Q}}_p^{nr}$. 0n a dans ce cas un isomorphisme

$$H^1_{f/e}(K, \mathbb{Q}_p) = H^1_{/e}(K, \mathbb{Q}_p) \cong H^1(K_\infty, \mathbb{Q}_p)^{G_\infty} = \mathrm{Hom}_{G_\infty}(G^{ab}_{K_\infty}, \mathbb{Q}_p)$$

Cette situation ne se produit pas si $V^{G_{K_\infty}} = 0$. On ne le voit non plus pas très bien si $V(-i)^{G_K} \neq 0$ pour un $i < h$ à cause des relations du type

$$\Omega_{V,h+1}(g) = \ell_h \Omega_{V,h}(g)$$

qui font disparaître le pôle. Cependant cela doit apparaître en théorie globale avec une bonne normalisation des "facteurs Γ".

4. Lois de réciprocité

On désigne toujours par γ un générateur fixé de Γ et on pose $\gamma_n = \gamma^{p^{n-1}}$.

4.1. L'application exponentielle duale. Ce paragraphe repose sur les théorèmes de Tate dont on rappelle ici l'énoncé : on note $\hat{K}_\infty$ le complété de K_∞.

4.1.1. Théorème. (Tate)

1. $H^1(K_\infty, \mathbb{C}_p) = 0$, $H^0(K_\infty, \mathbb{C}_p) = \hat{K}_\infty$;
2. *Pour* $n \geq 1$, *il existe un unique isomorphisme* $T_{K_n} : \hat{K}_\infty/(\gamma_n - 1) \to K_n$ *induisant* $\frac{1}{p^{m-n}} Tr_{K_m/K_n}$ *sur* K_m ;
3. $H^m(K_\infty/K_n, \hat{K}_\infty(i)) = 0$ *pour* $i \neq 0$; $H^0(K_\infty/K_n,, \hat{K}_\infty) = K_n$ *et* $H^1(K_\infty/K_n,, \hat{K}_\infty) \cong K_n$ *où cette dernière application est donnée par*

$$\begin{array}{ccccc} H^1(\Gamma_n, \hat{K}_\infty) & = & \hat{K}_\infty/(\gamma_n - 1) & \to & K_n \\ c & \mapsto & c_{\gamma_n} & \mapsto & \frac{1}{\log \chi(\gamma_n)} T_{K_n}(c_\gamma) \end{array}$$

Avec cette normalisation, on a $T_{K_m} = p^{n-m} Tr_{K_n/K_m} \circ T_{K_n}$ pour $n \geq m \geq 1$. On a le diagramme commutatif pour $m \leq n$

$$\begin{array}{ccccc} H^1(K_\infty/K_n, \hat{K}_\infty) & \to & \hat{K}_\infty/(\gamma^{p^n}-1) & \to & K_n \\ res \uparrow & & \uparrow & & \uparrow \\ H^1(K_\infty/K_m, \hat{K}_\infty) & \to & \hat{K}_\infty/(\gamma^{p^m}-1) & \to & K_m \end{array}$$

où la deuxième flèche verticale est donnée par $c \mapsto \sum_{i=0}^{p^{n-m}-1} \gamma^{ip^m} c$ et la troisième par l'inclusion. Remarquons aussi que si $x \in K_m$ avec $m \geq n \geq 1$, on a $T_{K_n}(x) = \frac{1}{p^{m-n}} Tr_{K_m/K_n}(x)$.

4.1.2. Construisons une famille d'applications $\lambda_{k,n} : B_{\mathrm{dR}}^{G_{K_\infty}} \to K_n$ pour $n \geq 0$ et $k \in \mathbb{Z}$.
On rappelle que $\mathrm{Fil}^i B_{\mathrm{dR}} / \mathrm{Fil}^{i+1} B_{\mathrm{dR}} = \mathbb{C}_p(i)$ en tant que G_K-modules. La nullité de $H^1(K_\infty, \mathbb{C}_p)$ implique que $\mathrm{Fil}^i B_{\mathrm{dR}}^{G_{K_\infty}} / \mathrm{Fil}^{i+1} B_{\mathrm{dR}}^{G_{K_\infty}} = \hat{K}_\infty(i)$. On déduit alors de la nullité des $H^m(G_\infty, \hat{K}_\infty(i))$ pour $i \neq 0$ que $\gamma_n - 1$ est inversible sur $\mathrm{Fil}^1 B_{\mathrm{dR}}^{G_{K_\infty}}$ de même que sur $B_{\mathrm{dR}}^{G_{K_\infty}} / \mathrm{Fil}^0 B_{\mathrm{dR}}^{G_{K_\infty}}$ et donc que

$$\begin{aligned} B_{\mathrm{dR}}^{G_{K_\infty}}/(\gamma_n - 1) B_{\mathrm{dR}}^{G_{K_\infty}} \cong & {B_{\mathrm{dR}}^+}^{G_{K_\infty}}/(\gamma_n - 1){B_{\mathrm{dR}}^+}^{G_{K_\infty}} \\ = & {B_{\mathrm{dR}}^+}^{G_{K_\infty}} / \mathrm{Fil}^1 {B_{\mathrm{dR}}}^{G_{K_\infty}}/(\gamma_n - 1) \\ = & \hat{K}_\infty/(\gamma_n - 1)\hat{K}_\infty \ . \end{aligned}$$

En composant avec T_{K_n}, on obtient une application $\lambda_{0,n} : B_{\mathrm{dR}}^{G_{K_\infty}} \to K_n$. Les applications $\lambda_{k,n} : B_{\mathrm{dR}}^{G_{K_\infty}}/(\chi(\gamma_n)^{-k}\gamma_n - 1) B_{\mathrm{dR}}^{G_{K_\infty}} \to K_n$ sont obtenues par twist :

$$B_{\mathrm{dR}}^{G_{K_\infty}} \overset{\times t^{-k}}{\to} B_{\mathrm{dR}}^{G_{K_\infty}}/(\gamma_n - 1) \overset{\lambda_{0,n}}{\to} K_n \ .$$

On a ainsi $\lambda_{k,n}(b) = \lambda_{0,n}(t^{-k} b)$.

Lemme. *Les applications $\lambda_{k,n}$ vérifient :*

1. $\lambda_{k,n}(\tau x) = \chi(\tau)^k \lambda_{k,n}(x)$;
2. $Tr_{K_m/K_n}(\lambda_{k,n}(x)) = p^{m-n} \lambda_{k,n}(x)$ *pour* $m \geq n$.
3. *Soit* $G \in \mathcal{H}_\infty$. *Alors, pour* $m \geq n$ *et* $k \geq 0$

$$\lambda_{k,n}(G(\beta_m - 1)) = \frac{1}{p^{m-n}} \frac{Tr_{K_m/K_n}(D^k(G)(\zeta_m - 1))}{p^{mk} k!}$$

Démonstration. Démontrons la troisième assertion. Remarquons d'abord qu'il est facile de calculer $\lambda_{k,n}$ sur un élément de $K_m((t))$. En effet, si $\alpha \in K_m$ et $m \geq n$, on a $\lambda_{k,n}(\alpha t^i) = 0$ si $i \neq k$ et $\frac{1}{p^{m-n}} Tr_{K_m/K_n}(\alpha)$ si $i = k$. Le premier cas vient de ce que $\chi(\gamma_n)^{i-k}\gamma_n - 1$ est un isomorphisme sur $\hat{K}_\infty$, le deuxième de ce que $\lambda_{k,n}(\alpha t^i) = \lambda_{0,n}(\alpha) = T_{K_n}(\alpha)$.

Si on utilise le fait que $\beta_m = \zeta_m e^{t/p^m}$, on obtient que

$$G(\beta_m - 1) = \sum_{j=0}^{\infty} \frac{D^j(G)(\zeta_m - 1)}{j!} \frac{t^j}{p^{mj}}$$

et donc que

$$\begin{aligned}\lambda_{k,n}(G(\beta_m - 1)) =& \lambda_{k,n}(\frac{D^k(G)(\zeta_m - 1)}{k!} \frac{t^k}{p^{mk}}) \\ =& p^{n-m} \frac{Tr_{K_m/K_n}(D^k(G)(\zeta_m - 1))}{k! p^{mk}}\end{aligned}$$

□

Remarquons (ce qui a été utilisé dans la démonstration) que si $f = \sum_j a_j(f) t^j \in K_n((t))$, on a $\lambda_{k,n}(f) = a_k(f)$ et que l'on peut définir une application T_n de $K_\infty((t))$ dans $K_n((t))$ par $T_n(f) = p^{-n} \sum_k \lambda_{k,n}(f) t^k$. Un résultat important de Colmez est que T_n se prolonge en une application continue de $B_{\max}^{G_{K_\infty}}$ dans $K_n((t))$. Nous n'en avons pas besoin pour la démonstration de la loi de réciprocité pour l'application $\Omega_{V,h}$ que nous avons construite ici. Ce prolongement semble par contre fondamental dans l'extension qu'en donne Colmez.

4.1.3. Rappelons le théorème de Kato relatif à l'application exponentielle duale. Si W est une représentation p-adique de de Rham de $G_{\mathbb{Q}_p}$ et L une extension algébrique de $\mathbb{Q}_p$, on note $\exp^*_{W^*(1),L,/u^*}$ l'application duale de l'application exponentielle $\exp_{W^*(1),L,u}$ avec $u \in \{e, f\}$ et $e^* = g$, $f^* = f$. On pose

$$\lambda_{W,L} = \exp^*_{W^*(1),L,/g} : H^1(L, W) \to L \otimes \mathrm{Fil}^0 \mathbf{D}_{\mathrm{dR}}(W) .$$

En notant par $< ., . >_{W,L}$ le cup produit :

$$H^1(L, W) \times H^1(L, W^*(1)) \to H^2(L, \mathbb{Q}_p(1)) \cong \mathbb{Q}_p$$

et $[., .]_{\mathbf{D}_{\mathrm{dR}}(W)}$ la dualité naturelle

$$L \otimes \mathbf{D}_{\mathrm{dR}}(W) \times L \otimes \mathbf{D}_{\mathrm{dR}}(W^*(1)) \to L \overset{Tr_{L/\mathbb{Q}_p}}{\to} \mathbb{Q}_p ,$$

on a donc par exemple la formule

$$< x, \exp_{W^*(1),L,e}(b) >_{W,L} = [\exp^*_{W^*(1),L,/g}(x), b]_{\mathbf{D}_p(W)} = [\lambda_{W,L}(x), b]_{\mathbf{D}_p(W)} .$$

On fera attention que

$$< \exp_{W,L,e}(a), y >_{W,L} = -[a, \exp^*_{W,L,/g}(y)]_{\mathbf{D}_p(W)} = -[a, \lambda_{W^*(1),L}(y)]_{\mathbf{D}_p(W)} .$$

Proposition. *L'application* $\lambda_{V(k),K_n} : H^1(K_n, V(k)) \to K_n \otimes \mathrm{Fil}^0 \mathbf{D}_{\mathrm{dR}}(V(k))$ *peut se calculer de la manière suivante : soit* $\tau \mapsto c_\tau$ *un* *cocycle de* G_{K_n} *à valeurs dans* $B_{\mathrm{dR}}^{G_{K_\infty}} \otimes \mathbf{D}_{\mathrm{dR}}(V)$ *ayant même image que* $x \in H^1(K_n, V(k))$ *dans* $H^1(K_n, B_{\mathrm{dR}} \otimes V)$ *; alors,*

$$\lambda_{V(k),K_n}(x) = \lambda_{-k,n}\left(\frac{c_{\gamma_n}}{\log \chi(\gamma_n)}\right) .$$

Démonstration. L'existence de c vient de ce que $H^1(K_\infty, B_{\mathrm{dR}} \otimes V) = 0$, ce qui implique que l'application inflation suivante est un isomorphisme :

$$H^1(G_\infty, (B_{\mathrm{dR}} \otimes V)^{G_{K_\infty}}) = H^1(G_\infty, B_{\mathrm{dR}}^{G_{K_\infty}} \otimes \mathbf{D}_{\mathrm{dR}}(V)) \xrightarrow{\cong} H^1(K, B_{\mathrm{dR}} \otimes V) .$$

Kato démontre que si x est représenté par un cocycle $\tau \mapsto d_\tau$, les deux cocycles de G_{K_n} à valeurs dans $B_{\mathrm{dR}} \otimes V$ donnés par $\tau \mapsto \lambda_{V,K_n}(x) \log \chi(\tau)$ et par $\tau \mapsto d_\tau$ ont même image dans $H^1(K_n, B_{\mathrm{dR}} \otimes V)$. Colmez remarque alors qu'on peut remplacer d par le cocycle $\tau \mapsto c_\tau$ ayant même image que d dans $H^1(K_n, B_{\mathrm{dR}} \otimes V)$. On en déduit que $\lambda_{V,K_n}(x) \log \chi(\gamma_n) \equiv c_{\gamma_n} \mod \gamma_n - 1$ et donc que

$$\lambda_{V,K_n}(x) \log \chi(\gamma_n) = \lambda_{0,n}(\lambda_{V,K_n}(x) \log \chi(\gamma_n)) = \lambda_{0,n}(c_{\gamma_n}) .$$

Pour passer à $V(k)$, il suffit de faire un twist convenable. □

REMARQUES : 1) L'image de $H^1(G_n, V^{G_{K_\infty}})$ dans $K_n \otimes \mathrm{Fil}^0 \mathbf{D}_{\mathrm{dR}}(V)$ par $\lambda_{V,K}$ est égale à $V^{G_K} = \mathrm{Fil}^0 \mathbf{D}_{\mathrm{dR}}(V)^{\varphi=1}$. Notons $\tilde{\lambda}_{V(k),K_n}$ le composé de $\lambda_{V(k),K_n}$ avec la projection modulo $V(k)^{G_K}$.
2) On a le diagramme commutatif

$$\begin{array}{ccccc}
H^1(K,V) & & & & \\
\downarrow & & & & \\
H^1(K, B_{\mathrm{dR}} \otimes \mathbf{D}_{\mathrm{dR}}(V)) & \xleftarrow{\cup \log \chi} & H^0(K, B_{\mathrm{dR}} \otimes \mathbf{D}_{\mathrm{dR}}(V)) & = & K \otimes \mathbf{D}_{\mathrm{dR}}(V) \\
\cong\uparrow & & \downarrow & & || \\
H^1(G_\infty, B_{\mathrm{dR}}^{G_{K_\infty}} \otimes \mathbf{D}_{\mathrm{dR}}(V)) & \xleftarrow{\cup \log \chi} & B_{\mathrm{dR}}^{G_{K_\infty}} \otimes \mathbf{D}_{\mathrm{dR}}(V)/(\gamma - 1) & \rightarrow & K \otimes \mathbf{D}_{\mathrm{dR}}(V)
\end{array}$$

où $\log \chi$ est vu comme élément de $H^1(K, \mathbb{Q}_p) = H^1(G_\infty, \mathbb{Q}_p) = \mathrm{Hom}(G_\infty, \mathbb{Q}_p)$.

4.2. Loi de réciprocité (Enoncés).

4.2.1. THÉORÈME. (Colmez) *Soit* h *un entier tel que* $\mathrm{Fil}^{-h} \mathbf{D}_{\mathrm{dR}}(V) = \mathbf{D}_{\mathrm{dR}}(V)$. *Soit* $g \in \mathcal{D}_{\infty,e}(V)$, G *une solution compatible des équations* $(1 - p^r \Phi)\mathcal{G}_r = D^r(g)$. *Alors, pour tout entier* $k \leq -h$ *et pour tout entier* $n \geq 1$, *on a*

$$\tilde{\lambda}_{V(k),K_n}(\tilde{\pi}_{n,k}(\Omega_{V,h}(g))) \equiv \frac{p^{n(k-1)}(1 \otimes \varphi)^{-n} D^{-k}(G)(\zeta_n - 1)}{(-k-h)!} \otimes e_{-k} \mod V(k)^{G_K} .$$

Remarquons que sous les hypothèses du théorème, $V(k)^{G_K}=0$ sauf peut-être pour $k=-h$.
Rappelons que l'on a par définition même de $\Omega_{V,h}(g)$ les formules pour $k\geq 1-h$,

$$\begin{aligned}\log_{V(k),n}(\pi_{n,k}\Omega_{V,h}(g)) &\equiv (-1)^{h+k-1}(h+k-1)!\times\\ &\times p^{n(k-1)}(1\otimes\varphi)^{-n}D^{-k}(G)(\zeta_n-1)\otimes e_{-k}\quad \bmod\ \mathrm{Fil}^0\,\mathbf{D}_p(V(k))\ .\end{aligned}$$

Pour uniformiser les formules, posons (cf. [5]) $\Gamma^*(k)=(k-1)!$ si $k>0$ et $(-1)^k/(-k!)$ si $k\leq 0$. On a encore l'équation fonctionnelle : $\Gamma^*(k+1)=k\Gamma^*(k)$ pour tout $k\in\mathbb{Z}$ sauf pour $k=0$. On obtient alors que

(4.2.1)

$$\frac{p^{n(k-1)}(1\otimes\varphi)^{-n}D^{-k}(G)(\zeta_n-1)}{\Gamma^*(-k-(h-1))}=\begin{cases}\log_{V(k),n}(\pi_{n,k}\Omega_{V,h}(g)) & \text{si } h+k-1>0\\ \lambda_{V(k),K_n}(\pi_{n,k}(\Omega_{V,h}(g))) & \text{si } h+k-1<0\end{cases}$$

On peut aussi remarquer que

$$\Gamma^*(-k)=\prod_{i=1}^{h-1}(-k-i).\Gamma^*(-k-(h-1))$$

à condition que $k\notin\{-h+1,\cdots-1\}$. Le produit $\prod_{i=1}^{h-1}(-k-i)$ est un produit de $h-1$ termes et c'est aussi la valeur de $\prod_{i=1}^{h-1}\ell_i$ sur le caractère χ^{-k}. On a donc aussi

$$\frac{p^{n(k-1)}(1\otimes\varphi)^{-n}D^{-k}(G)(\zeta_n-1)}{\Gamma^*(-k)}=\begin{cases}\log_{V(k),n}(\Omega_{V,1}(g)) & \text{si } k\geq 0\\ \lambda_{V(k),K_n}(\pi_{n,k}(\Omega_{V,1}(g))) & \text{si } k\leq -h\end{cases}$$

en posant $\Omega_{V,1}=(\prod_{i=1}^{h-1}\ell_i)^{-1}\Omega_{V,h}$. On a en effet alors

$$\pi_{n,k}(\Omega_{V,1}(g))=(\prod_{i=1}^{h-1}(-i-k))^{-1}\pi_{n,k}(\Omega_{V,h})$$

On a bien sûr perdu un certain nombre de termes.

4.2.2. Il est commode de transformer le théorème 4.2.1 et de le mettre sous la forme de la conjecture Réc(V) de [4]. Cela permettra ensuite d'obtenir de nouvelles formules que nous donnerons dans le §5.2. Nous supposons dans la fin du §4.2 que V est une représentation cristalline.
Rappelons que l'on a un accouplement naturel sesquilinéaire par rapport à l'involution ι induite par $\tau\mapsto\tau^{-1}$

$$Z^1_\infty(K,V)\times Z^1_\infty(K,V^*(1))\to\mathbb{Q}_p\otimes\Lambda$$

donnée par

$$<x,y>_V=\lim_{\overleftarrow{n}}\sum_{\tau\in G_n}<\tau^{-1}\pi_{n,0}(x),\pi_{n,0}(y)>_{V,K_n}\tau$$

où $< \, , \, >_{V,K_n}$ est l'accouplement de Kummer $H^1(K_n,V)\times H^1(K_n,V^*(1))\to \mathbb{Q}_p$. Il vérifie pour tout entier i

$$Tw^{-i}(< x,y >_V) = < Tw^i(x), Tw^{-i}(y) >_{V(i)} \quad .$$

On a d'autre part l'accouplement naturel

$$\mathbf{D}_p(V)\times \mathbf{D}_p(V^*(1)) \to \mathbb{Q}_p \; .$$

On prolonge respectivement ces deux accouplements par extension des scalaires à $\mathcal{K}(G_\infty)$:

$$< \, ., \, . >_V : \mathcal{K}(G_\infty)\otimes_\Lambda Z^1_\infty(K,V)\times \mathcal{K}(G_\infty)\otimes_\Lambda Z^1_\infty(K,V^*(1)) \to \mathcal{K}(G_\infty)$$

et

$$[\, ., \, .]_{\mathbf{D}_p(V)} : \mathcal{K}(G_\infty)\otimes_\Lambda \mathbf{D}_p(V)\times \mathcal{K}(G_\infty)\otimes_\Lambda \mathbf{D}_p(V^*(1)) \to \mathcal{K}(G_\infty)$$

On note de la même manière l'accouplement qui s'en déduit sur $\mathcal{H}_\infty^{\psi=0}\otimes \mathbf{D}_p(V)\times \mathcal{H}_\infty^{\psi=0}\otimes \mathbf{D}_p(V)$. Notons encore ι l'involution de $\mathcal{H}_\infty^{\psi=0}$ correspondant à l'involution ι précédemment définie sur $\mathcal{H}(G_\infty)$. Enfin, notons σ_{-1} l'élément de G_∞ agissant sur les racines de l'unité par $\zeta\mapsto\zeta^{-1}$. On a donc $\sigma_{-1}(1+T)=(1+T)^{-1}$.

4.2.3. Théorème. (Réc(V)) *Supposons que V est une représentation cristalline. On a pour tout entier h*

$$< \Omega_{V,h}(g_1), \sigma_{-1}\Omega_{V^*(1),1-h}(g_2) >_V .(1+T) = (-1)^h [g_1, \iota(g_2)]_{\mathbf{D}_p(V)} \; .$$

Autrement dit, l'inverse de $\Omega_{V,h}$ est au signe près l'adjoint de $\Omega_{V^*(1),1-h}$.

Ainsi, au lieu de commencer à construire $\Omega_{V,h}$ à partir des exponentielles de Bloch-Kato, on aurait pu construire $\Omega^*_{V^*(1),1-h}$ à partir de l'exponentielle duale, ou son inverse, ou l'adjoint de son inverse. Remarquons que l'application $\Omega_{V,h}$ dépend du choix de ϵ. Appliquer σ_{-1} revient à changer ϵ en ϵ^{-1}.
Nous donnerons au §5 les formules sur l'application inverse $\mathcal{L}_{V,h}$ de $\Omega_{V,h}$ se déduisant de ce théorème.

Démonstration. Montrons comment le théorème 4.2.3 se déduit du théorème 4.2.1. On s'appuie sur les formules données dans l'appendice A.2. On vérifie facilement que cela ne dépend pas de h, on prend alors h tel que $\mathrm{Fil}^{-h}\,\mathbf{D}_p(V)=\mathbf{D}_p(V)$. Soit h^* tel que $\mathrm{Fil}^{-h^*}\,\mathbf{D}_p(V^*(1))=\mathbf{D}_p(V^*(1))$. On a alors $\mathrm{Fil}^{h^*}\,\mathbf{D}_p(V)=0$. Prenons k tel que $-k\geq 1-h$ et $k<-h^*$. On pose $x=\Omega_{V,h}(g_1)$ et $y=\Omega_{V^*(1),h^*}(g_2)$. On note encore abusivement $\pi_{n,k}$ la projection $\mathcal{H}(G_\infty)\otimes Z^1_\infty(K,T^*(1))\to H^1(K_n,V^*(1)(k))$. On a alors

$$\begin{aligned}
&s_{n,k}(<\Omega_{V,h}(g_1),\Omega_{V^*(1),1-h}(g_2)>_V)\\
&\quad= s_{n,k}(<\Omega_{V,h}(g_1),(\prod_{-h<j<h^*}\ell_j)^{-1}\Omega_{V^*(1),h^*}(g_2)>_V)\\
&\quad= s_{n,k}(<(\prod_{-h<j<h^*}-\ell_{-j})^{-1}x,y>_V)\\
&\quad= \frac{(-k-h^*)!}{(-k+h-1)!}<\pi_{n,-k}(x),\pi_{n,k}(y)>_{V(-k),K_n}
\end{aligned}$$

car $\pi_{n,-k}(\ell_{-j}x)=(j+k)\pi_{n,-k}(x)$;

$$=-\frac{(-k-h^*)!}{(-k+h-1)!}Tr_{K_n/K}([\log_{V(-k),K_n}\pi_{n,-k}(x),\lambda_{V(k),K_n}(\pi_{n,k}(y))]_{\mathbf{D}_p(V(-k))})$$

(remarquons que $\pi_{n,-k}(x)\in H^1_e(K_n,V(-k))$ sous les hypothèses faites sur k : $\mathrm{Fil}^0\,\mathbf{D}_{\mathrm{dR}}(V(-k))=0$; le signe $-$ provient de 4.1.3). Exprimons le dernier terme à l'aide de g_1 et g_2. Avec $(1-p^k\Phi)D^k(G_1)=D^k(g_1)$ et $(1-p^{-k}\Phi)D^{-k}(G_2)=D^{-k}(g_2)$, on a

$$\begin{aligned}
&\frac{(-k-h^*)!}{(-k+h-1)!}<\pi_{n,-k}(x),\pi_{n,k}(y)>_{V(-k),K_n}\\
&\quad=(-1)^{h-k}p^{-n}Tr_{K_n/K}([D^k(G_1)(\zeta_n-1),D^{-k}(G_2)(\zeta_n-1)]_{\mathbf{D}_p(V(-k))})
\end{aligned}$$

en utilisant les formules (4.2.1) et

$$[\varphi u,\varphi v]_{\mathbf{D}_p(V(-k))}=p^{-1}[u,v]_{\mathbf{D}_p(V(-k))}\ .$$

Ce qui vaut pour $n\geq 1$,

$$(-1)^{h-k}p^{-n}\sum_{\zeta\in\mu_{p^n}-\mu_{p^{n-1}}}[D^k(G_1)(\zeta-1),D^{-k}(G_2)(\zeta-1)]_{\mathbf{D}_p(V)}\ .$$

On remarque que pour $n>1$

$$\sum_{\zeta\in\mu_{p^n}-\mu_{p^{n-1}}}[D^k(G_1)(\zeta^p-1),D^{-k}(g_2)(\zeta-1)]_{\mathbf{D}_p(V)}=0$$

car $\psi(g_2)=0$ et idem en renversant les rôles de G_2 et G_1. Donc,

$$\begin{aligned}
&\sum_{\zeta\in\mu_{p^n}-\mu_{p^{n-1}}}[D^k(G_1)(\zeta-1),D^{-k}(G_2)(\zeta-1)]_{\mathbf{D}_p(V)}\\
&\quad=\sum_{\zeta\in\mu_{p^n}-\mu_{p^{n-1}}}[D^k(g_1)(\zeta-1),D^{-k}(g_2)(\zeta-1)]_{\mathbf{D}_p(V)}\\
&\qquad+p.p^{-1}\sum_{\zeta\in\mu_{p^{n-1}}-\mu_{p^{n-2}}}[D^k(G_1)(\zeta-1),D^{-k}(G_2)(\zeta-1)]_{\mathbf{D}_p(V)}
\end{aligned}$$

En recommençant, on en déduit que

$$\sum_{\zeta\in\mu_{p^n}-\mu_{p^{n-1}}}[D^k(G_1)(\zeta-1),D^{-k}(G_2)(\zeta-1)]_{\mathbf{D}_p(V)}$$
$$=\sum_{\zeta\in\mu_{p^n}-\mu_p}[D^k(g_1)(\zeta-1),D^{-k}(g_2)(\zeta-1)]_{\mathbf{D}_p(V)}$$
$$+\sum_{\zeta\in\mu_p-\{1\}}[D^k(G_1)(\zeta-1),D^{-k}(G_2)(\zeta-1)]_{\mathbf{D}_p(V)}$$

Enfin,

$$\begin{aligned}&\sum_{\zeta\in\mu_p-\{1\}}[D^k(G_1)(\zeta-1),D^{-k}(G_2)(\zeta-1)]_{\mathbf{D}_p(V)}\\
&=\sum_{\zeta\in\mu_p}[D^k(G_1)(\zeta-1),D^{-k}(G_2)(\zeta-1)]_{\mathbf{D}_p(V)}-[D^k(G_1)(0),D^{-k}(G_2)(0)]_{\mathbf{D}_p(V)}\\
&=\sum_{\zeta\in\mu_p}[D^k(g_1)(\zeta-1),D^{-k}(g_2)(\zeta-1)]_{\mathbf{D}_p(V)}\\
&\quad+[D^k(G_1)(0),D^{-k}(G_2)(0)]_{\mathbf{D}_p(V)}-[D^k(G_1)(0),D^{-k}(G_2)(0)]_{\mathbf{D}_p(V)}\\
&=\sum_{\zeta\in\mu_p}[D^k(g_1)(\zeta-1),D^{-k}(g_2)(\zeta-1)]_{\mathbf{D}_p(V)}\end{aligned}$$

D'où l'égalité pour $n\geq 1$

$$p^{-n}\sum_{\zeta\in\mu_{p^n}-\mu_{p^{n-1}}}[D^k(G_1)(\zeta-1),D^{-k}(G_2)(\zeta-1)]_{\mathbf{D}_p(V)}$$
$$=p^{-n}\sum_{\zeta\in\mu_{p^n}}D^k(g_1)(\zeta-1),D^{-k}(g_2)(\zeta-1)]_{\mathbf{D}_p(V)}\ .$$

Comme $D^{-k}\circ\sigma_{-1}=(-1)^k\sigma_{-1}D^k$, on obtient finalement que pour $n\geq 1$,

$$\begin{aligned}&s_{n,k}(<\Omega_{V,h}(g_1),\Omega_{V^*(1),1-h}(g_2)>_V)\\
&\qquad=(-1)^hp^{-n}\sum_{\zeta\in\mu_{p^n}}[D^k(g_1)(\zeta-1),D^{-k}(\sigma_{-1}g_2)(\zeta^{-1}-1)]_{\mathbf{D}_p(V)}\\
&\qquad=(-1)^hs_{n,k}([\hat{g}_1,\sigma_{-1}\hat{g}_2^\iota]_{\mathbf{D}_p(V)})\end{aligned}$$

L'égalité ayant lieu pour tout k inférieur à h et à $-h^*$, on en déduit que

$$<\Omega_{V,h}(g_1),\Omega_{V^*(1),1-h}(g_2)>_V=(-1)^h[\hat{g}_1,\sigma_{-1}\hat{g}_2^\iota]_{\mathbf{D}_p(V)}\ ,$$

ou ce qui revient au même que

$$<\Omega_{V,h}(g_1),\Omega_{V^*(1),1-h}(g_2)>_V.(1+T)=(-1)^h[g_1,\sigma_{-1}g_2^\iota]_{\mathbf{D}_p(V)}\ .$$

□

4.2.4. Avant de passer à la démonstration de la loi de réciprocité, montrons comment elle permet de transformer l'inégalité

$$\mathfrak{o}(\Omega_{V,h}(g)) \leq h + \mathfrak{o}_\varphi(g)$$

pour $g \in \mathcal{D}_{\infty,f}(V)$ en égalité. On a de même

$$\mathfrak{o}(\Omega_{V^*(1),h^*}(g^*)) \leq h^* + \mathfrak{o}_\varphi(g^*)$$

pour $g^* \in \mathcal{D}_{\infty,f}(V^*(1))$ avec h^* comme précédemment. Choisissons g^* de manière à ce que $\mathfrak{o}_\varphi(g) + \mathfrak{o}_\varphi(g^*) = \mathfrak{o}_\varphi([g, \sigma_{-1} g^{*\iota}]_{\mathbf{D}_p(V)})$. Les inégalités déjà montrées et le théorème 4.2.3 impliquent alors que

$$\begin{aligned}\mathfrak{o}(\Omega_{V,h}(g)) + \mathfrak{o}(\Omega_{V,h^*}(g^*)) &\leq \mathfrak{o}_\varphi(g) + h + \mathfrak{o}_\varphi(g^*) + h^* \\ &\leq \mathfrak{o}(\Omega_{V,h}(g)) + \mathfrak{o}(\Omega_{V,h^*}(g^*))\end{aligned}$$

à cause du terme $(\prod_{-h<j<h^*} \ell_j)^{-1}$. On en déduit l'égalité

$$\mathfrak{o}(\Omega_{V,h}(g)) = h + \mathfrak{o}_\varphi(g) \ .$$

4.3. Loi de réciprocité (Démonstration). Soient g un élément de $\mathcal{D}_{\infty,e}(V)$ et $\overline{G}$ une solution compatible des équations $(1 - p^r\Phi)\mathcal{G}_r = D^r(g)$. On suppose que g est $p^{-u}\varphi^-$-bornée, ce qui assure que $\Omega_{V,h}(g)$ appartient à $\mathcal{H}_{(h+u)^-}(G_\infty) \otimes \tilde{Z}^1_\infty(K,T)$. Au cours de la construction de $\Omega_{V,h}(G)$, nous avons construit explicitement (lemme 3.4.3) un cocycle $Z_{n,k,\tau}$ de G_{K_n} à valeurs dans V représentant $\pi_{n,k}(\Omega_{V,h}(G))$ pour $n \geq 0$ et $k + h - 1 \geq 0$: pour $\tau \in G_{K_n}$,

$$Z_{n,k,\tau} = (\chi(\tau)^k \tau - 1)(e_{n,k})$$

avec $e_{n,k} = c_{n,k} - Eul((1-\varphi)c_{n,k})$ et $c_{n,k} = S^{(h)}_{n,k,cris}(G)$. Une remarque fondamentale de P. Colmez est qu'on peut retrouver G à partir d'un tel cocycle :

4.3.1. Proposition. *Soit g un élément $p^{-u}\varphi^-$-borné de $\mathcal{H}^{\psi=0}_\infty \otimes \mathcal{D}$ et et $\overline{G}$ une solution des équations $(1 - p^r\Phi)\mathcal{G}_r = D^r(g)$. Alors, la suite $p^m \sum_{j=0}^{u+h-1} (-1)^j \binom{u+h-1}{j} e_{m,j-h+1}$ converge dans $(B_{\max}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))^{\Phi=p^u}$ et on a pour tout entier $n \geq 0$ et tout entier k tel que $k + u \geq 0$,*

$$\lambda_{k,n}(\lim_{m\to\infty} p^m \sum_{j=0}^{u+h-1} (-1)^j \binom{u+h-1}{j} e_{m,j-h+1}) = \\ \frac{(u+h-1)!}{(k+u)!} \frac{(1\otimes\varphi)^{-n} D^k(G)(\zeta_n - 1)}{p^{nk}} \ .$$

Remarquons que pour $k + u < 0$, le membre de gauche est nul.

Démonstration. La limite de $p^m \sum_{j=0}^{u+h-1} (-1)^j \binom{u+h-1}{j} S^{(h)}_{n,j-h+1,cris}(g)$ est nulle (3.4.5, c'est d'ailleurs un argument essentiel dans l'existence de l'homomorphisme $\Omega_{V,h}$) et donc par continuité de Eul, il s'agit d'étudier la limite de

$$p^m \sum_{j=0}^{u+h-1} (-1)^j \binom{u+h-1}{j} S^{(h)}_{n,j-h+1,cris}(G) \ .$$

Nous avons calculé dans le §3.4.5 (avec g à la place de G, mais le calcul est bien sûr identique)

$$\sum_{j=0}^{u+h-1} (-1)^j \binom{u+h-1}{j} S^{(h)}_{m,j-h+1,cris}(G) =$$
$$= (u+h-1)! p^{m(u-1)} \Phi^{-m} D^{-u}(G)([\epsilon]-1) t^{-u}$$

On utilise alors le lemme suivant :

LEMME. *Si g est $p^{-u}\varphi^-$-bornée, alors, $p^{mu}\Phi^{-m}D^{-u}(G)([\epsilon]-1))$ a une limite dans $(B_{\max}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))^{\Phi=p^u}$.*

Démonstration. On s'appuie sur le fait que si $F \in \mathcal{H}_\infty$, $F(\beta_m - 1)$ existe dans $B^+_{\max}$ et que l'on a

$$\text{(4.3.1)} \qquad ||F(\beta_m - 1)||_{\max} \sim ||F||_{\rho_m} \ ,$$

(le symbole $\sim$ signifiant que $||F||_{\rho_m} \leq ||F(\beta_m - 1)||_{\max} \leq p||F||_{\rho_m}$). Rappelons que $D^{-u}(g) = (1 - p^{-u}\Phi)D^{-u}(G)$ et que $\psi(D^{-u}(g)) = 0$. Posons $u_m = p^{mu}\Phi^{-m}D^{-u}(G)([\epsilon]-1) = p^{mu}(1\otimes\varphi)^{-m}D^{-u}(G)(\beta_m - 1)$. On a

$$\begin{aligned} u_m - u_{m-1} &= p^{mu}\Phi^{-m}(1 - p^{-u}\Phi)(D^{-u}(G))([\epsilon]-1) \\ &= p^{mu}\Phi^{-m}(D^{-u}(g)([\epsilon]-1)) \ . \end{aligned}$$

On a grâce à 3.1.1

$$||u_m - u_{m+1}||_{\max} \sim ||p^{mu}(1\otimes\varphi)^{-m} f||_{\rho_m}$$

Comme g est par hypothèse $p^{-u}\varphi^-$-borné, il en est de même de $D^{-u}(g)$ (voir 1.3, 1.4) ; $u_m - u_{m+1}$ tend donc vers 0 dans $B^+_{\max} \otimes \mathbf{D}_p(V)$ et la suite (u_m) converge dans $B^+_{\max} \otimes \mathbf{D}_p(V)$. Il est clair que sa limite est fixe par G_{K_∞} puisque ce groupe de Galois laise fixe les β_m. Comme $p^{-u}\Phi(u_{m+1}) = u_m$, elle appartient à $(B_{\max}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))^{\Phi=p^u}$ □

On déduit de ce qui précède que

$$\begin{aligned} &\lambda_{k,n}(\lim_{m\to\infty} p^m \sum_{j=0}^{u+h-1} (-1)^j \binom{u+h-1}{j} \mathcal{S}_{m,j-h+1,cris}(G)) = \\ &\quad = (u+h-1)!\lambda_{k,n}(\lim_{m\to\infty} p^{mu}\Phi^{-m}D^{-u}(G)([\epsilon]-1)t^{-u}) \\ &\quad = (u+h-1)! \lim_{m\to\infty} \lambda_{k+u,n}(p^{mu}\Phi^{-m}D^{-u}(G)([\epsilon]-1)) \\ &\quad = \frac{(u+h-1)!}{(k+u)!} \lim_{m\to\infty} \frac{1}{p^{m-n}} \frac{Tr_{K_m/K_n}(p^{mu}(1\otimes\varphi)^{-m}D^k(G)(\zeta_m - 1)}{p^{m(k+u)}} \end{aligned}$$

pour $k+u \geq 0$. Comme $\Psi((1-p^k\Phi)(D^k(G)))=0$, on a

$$\sum_{\zeta\in\mu_p} D^k(G)(\zeta(1+T)-1) = p^{k+1}(1\otimes\varphi)D^k(G)((1+T)^p-1)\ . \tag{4.3.2}$$

On en déduit que

$$\begin{aligned} Tr_{K_m/K_n}((1\otimes\varphi)^{-m}(D^k(G)(\zeta_m-1)) = \\ p^{(k+1)(m-n)}(1\otimes\varphi)^{m-n}D^k(G))(\zeta_n-1)\ , \end{aligned}$$

d'où,

$$\frac{1}{p^{m-n}}\frac{Tr_{K_m/K_n}(p^{mu}(1\otimes\varphi)^{-m}D^k(G)(\zeta_m-1)}{p^{m(k+u)}} = \frac{(1\otimes\varphi)^{-n}D^k(G)(\zeta_n-1)}{p^{kn}}\ .$$

On en déduit la proposition. □

4.3.2. Posons pour simplifier $\tilde{h}=h+u$. Notons $A_{max,v}=t^{-v}A_{\max}$. On a des applications

$$\begin{aligned} \alpha : H^1(G_\infty, \mathcal{H}_{\tilde{h}^-}(G_\infty) \otimes (A_{\max,v}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))) \\ \to H^1(K_\infty, \mathcal{H}_{\tilde{h}^-}(G_\infty) \otimes (A_{\max,v}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))) \end{aligned}$$

et

$$\begin{aligned} \beta : \mathcal{H}_{\tilde{h}^-}(G_\infty) \otimes Z^1_\infty(K,T) \to H^1(K, \mathcal{H}_{\tilde{h}^-}(G_\infty) \otimes V) \\ \to H^1(K_\infty, \mathcal{H}_{\tilde{h}^-}(G_\infty) \otimes (A_{\max,v}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))) \end{aligned}$$

Lemme. *Il existe un élément z' de $H^1(G_\infty, \mathcal{H}_{\tilde{h}^-}(G_\infty) \otimes (A_{\max,v}^{G_{K_\infty}} \otimes V))$ tel que $\alpha(z)=\beta(\Omega_{V,h}(g))$.*

La démonstration utilise les résultats du type de ceux de Tate et de Sen. On renvoie à [1, chap. IV,§1-3 et lemme VI.3.2].
Choisissons un cocycle Z' représentant z'. Pour $j+h-1\geq 0$, l'image de $\Omega_{V,h}(g)$ dans $H^1(K_n,(B_{\max}\otimes V(j))^{\varphi=1})$ est nulle, il en est donc de même de celle de z' et on a donc $\pi_{n,j}(Z'_\tau)=(\tau-1)*_j d_{n,j}=(\chi(\tau)^j\tau-1)d_{n,j}$.

Lemme. *Avec les notations précédentes, la suite $p^m\sum_{j=0}^{u+h-1}(-1)^j\binom{u+h-1}{j}d_{m,j-h+1}$ a une limite dans $(B_{\max}^{G_{K_\infty}}\otimes\mathbf{D}_p(V))^{\Phi=p^u}$ et pour tout entier $n\geq 0$ et tout entier k tel que $k+u\geq 0$, on a*

$$\begin{aligned} \lambda_{k,n}(\lim_{m\to\infty} p^m \sum_{j=0}^{u+h-1}(-1)^j\binom{u+h-1}{j}d_{m,j-h+1}) = \\ \frac{(u+h-1)!}{(k+u)!}\frac{(1\otimes\varphi)^{-n}D^k(G)(\zeta_n-1)}{p^{nk}}\ . \end{aligned}$$

Démonstration. Il suffit de montrer que la limite de

$$p^m \sum_{j=0}^{u+h-1} (-1)^j \binom{u+h-1}{j} (d_{m,j-h+1} - e_{m,j-h+1})$$

est nulle. Or si Z est le cocycle représentant $\Omega_{V,h}(g)$ tel que

$$\pi_{m,k}(Z)_\tau = (\tau - 1)(e_{m,k}) \ ,$$

on a $Z_\tau - Z'_\tau = (\tau - 1)(B)$ avec $B \in \mathcal{H}_{\tilde{h}^-} \otimes (A_{\max,v}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))$ et donc $e_{m,j-h+1} - d_{m,j-h+1} = s_{m,j-h+1}(B)$. Il s'agit donc de montrer que

$$\lim_{m\to\infty} p^m \sum_{j=0}^{\tilde{h}-1} (-1)^j \binom{\tilde{h}-1}{j} \pi_{m,j-h+1}(B) = 0 \ ,$$

ce qui se déduit du fait que B appartient à $\mathcal{H}_{\tilde{h}^-} \otimes A_{\max,v} \otimes \mathbf{D}_p(V)$. □

4.3.3. Prenons donc un cocycle Z' comme dans le paragraphe précédent et posons $\tilde{Z} = Z'_\gamma$. Avec les notations précédentes, on a pour $i + h - 1 > 0$,

$$\pi_{m,i}(Z')_\tau = (\tau - 1) *_i d_{m,i} = (\chi(\tau)^i \tau - 1) d_{m,i}$$

avec $d_{m,i} \in (A_{\max,v}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))^{\Phi=1}$ et $\tau \in G_{K_m}$. D'où, pour $i + h - 1 \geq 0$,

$$\pi_{m,i}(Z')_{\gamma_m} = (\chi(\gamma_m)^i \gamma_m - 1) d_{m,i}$$

et pour $k \neq i$

$$\lambda_{-k,n}(d_{m,i}) = \frac{1}{\chi(\gamma_m)^{-k+i} - 1} \lambda_{-k,n}(\pi_{m,i}(Z')_{\gamma_m}) \ .$$

Commençons par faire le cas où $m \geq n = 1$. Dans l'appendice A, est fait le calcul explicite de $\pi_{m,i}(Z')$. Si $R_{m,i}(\tilde{Z})$ est le polynôme d'interpolation de $Tw^i \tilde{Z}$ modulo $\gamma_m - 1$ vu comme élément de $\mathbb{Z}_p[G_m] \otimes M$, on a

$$\lambda_{-k,1}(\pi_{m,i}(Z')_{\gamma_m}) = < \chi >^{k-i} (R_{m,i}(\tilde{Z}))$$

avec $M = (A_{\max,v}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))^{\Phi=1}$. D'où

$$\begin{aligned} &\lim_{m\to\infty} \lambda_{-k,1}(\sum_{j=0}^{\tilde{h}-1} (-1)^j \binom{\tilde{h}-1}{j} p^m d_{m,j-h-1}) \\ &\qquad = \lambda_{-k,1}(\lim_{m\to\infty} \sum_{j=0}^{\tilde{h}-1} (-1)^j \binom{\tilde{h}-1}{i} \frac{p^m < \chi >^{k-j+h-1} (R_{m,j-h+1}(\tilde{Z}))}{< \chi > (\gamma_m)^{j-h+1-k} - 1}) \ . \end{aligned}$$

On applique alors la proposition de l'appendice B à $Tw^{-h+1}\tilde{Z}$ (avec k remplacé par $k + h - 1$, h par $\tilde{h}$ et $< \chi > (\gamma)$ par $< \chi > (\gamma)^{-1}$) et on obtient que pour

k n'appartenant pas à $\{1-h,\cdots,u\}$

$$\lim_{m\to\infty}\lambda_{-k,1}(\sum_{j=0}^{\tilde h-1}(-1)^j\binom{\tilde h-1}{j}p^m d_{m,j-h-1})$$
$$=(-1)^{\tilde h}\frac{(\tilde h-1)!}{(k+h-1)\cdots(k-u)}\frac{\lambda_{-k,1}(<\chi>^k(\tilde Z))}{\log<\chi>(\gamma)}$$

Lorsque n est quelconque (avec toujours $m\geq n$), on décompose $\pi_{m,i}(Z')_{\gamma_m}$ sous la forme d'une somme de termes de la forme $\pi_{m,i}(Z')^{(j)}_{\gamma_m}\gamma^j\in\mathbb{Q}_p[\mathrm{Gal}(K_m/K_n)]\gamma^j$ pour j compris entre 0 et $p^{n-1}-1$, ce qui revient à remplacer le groupe Γ par le groupe Γ_n, on utilise le fait que $\lambda_{-k,n}(\tau x)=<\chi>^{-k}(\tau)\lambda_{-k,n}(x)$ pour $\tau\in\Gamma_n$ (cela n'est pas vrai pour γ^j avec $j=0,\cdots,p^{n-1}-1$) et on procède ensuite de la même manière. On obtient alors de nouveau que pour k n'appartenant pas à $\{1-h,\cdots,u\}$

$$\lim_{m\to\infty}\lambda_{-k,n}(\sum_{j=0}^{\tilde h-1}(-1)^j\binom{\tilde h-1}{j}p^m d_{m,j-h-1})$$
$$=(-1)^{\tilde h}\frac{(\tilde h-1)!}{(k+h-1)\cdots(k-u)}\frac{\lambda_{-k,n}(<\chi>^k(\tilde Z))}{\log<\chi>(\gamma)}$$

Le premier membre vaut pour $-k+u\geq 0$,

$$\frac{(\tilde h-1)!}{(-k+u)!}p^{nk}(1\otimes\varphi)^{-n}D^{-k}(G)(\zeta_n-1)\ .$$

D'où,

$$\frac{(\tilde h-1)!}{(-k+u)!}p^{nk}(1\otimes\varphi)^{-n}D^{-k}(G)(\zeta_n-1)=$$
$$(-1)^{\tilde h}\frac{(\tilde h-1)!}{(k+h-1)\cdots(k-u)}\frac{\lambda_{-k,n}(<\chi>^k(\tilde Z))}{\log\chi(\gamma)}\ .$$

Grâce au lemme de Kato

$$\frac{\lambda_{-k,n}(<\chi>^k(\tilde Z))}{p^n\log\chi(\gamma)}=\lambda_{V(k),n}(\pi_{n,k}(z))\ .$$

D'où,

$$\lambda_{V(k),K_n,/g}(\pi_{n,k}(z))=$$
$$\frac{(-1)^{\tilde h}(k+h-1)\cdots(k-u)}{(-k+u)!}p^{(k-1)n}(1\otimes\varphi)^{-n}D^{-k}(G)(\zeta_n-1)$$

pour $k-u<0$ et $k\neq 1-h,\ldots,u$, c'est-à-dire $k<1-h$. En posant $\tilde k=k-u$ (on a toujours $\tilde h=h+u$), on trouve que le coefficient dans le membre de droite est

$$C = \frac{(-1)^{\tilde{h}}(\tilde{k}+\tilde{h}-1)\cdots\tilde{k}}{(-\tilde{k})!} = \frac{\tilde{k}'\cdots(\tilde{k}'-\tilde{h}+1)}{\tilde{k}'!}$$

avec $\tilde{k}' = -\tilde{k} \geq h$

$$= \frac{1}{(\tilde{k}'-\tilde{h})!} = \frac{1}{(-k-h)!} \ .$$

D'où,

$$\lambda_{V(k),K_n}(\pi_{n,k}(z)) = \frac{p^{n(k-1)}(1\otimes\varphi)^{-n}D^{-k}(G)(\zeta_n-1)}{(-k-h)!}$$

Ce qui termine la démonstration du théorème 4.2.1.

5. Quelques conséquences

On suppose dans ce paragraphe V cristalline. Nous trouvons commode d'identifier ici $\mathbb{Z}_p[[G_\infty]]$ avec $\mathbb{Z}_p[[T]]^{\psi=0}$ et $\mathcal{H}(G_\infty)$ avec $\mathcal{H}_\infty^{\psi=0}$ par l'application induite par $\tau \mapsto (1+T)^{\chi(\tau)}$ pour $\tau \in G_\infty$. On a donc canoniquement $\mathcal{D}_{\infty,f}(V) = \mathcal{H}(G_\infty)\otimes \mathbf{D}_p(V)$.
Pour tout entier r, on note $\mathcal{L}_{V,r}$ l'inverse de $\Omega_{V,r}$. Il est donc à valeurs dans $\mathcal{K}(G_\infty)\otimes \mathbf{D}_p(V)$.

5.1. Déterminant et inverse de $\Omega_{V,h}$.

Si $\mathrm{Fil}^{-h}\,\mathbf{D}_p(V) = \mathbf{D}_p(V)$, on note $\delta_h(\Omega_V)$ l'idéal suivant de $\mathbb{Q}_p\otimes\mathcal{K}(G_\infty)$: c'est l'image par l'application déterminant $\det\Omega_{V,h}$ du $\mathbb{Q}_p\otimes\Lambda$-module $\det_{\mathbb{Q}_p\otimes\Lambda}(\Lambda\otimes\mathbf{D}_p(V))\otimes \otimes_{i\in\{1,2\}}(\det_{\mathbb{Q}_p\otimes\Lambda} Z^i_\infty(K,V))^{(-1)^i}$ où $Z^2_\infty(K,V) = \mathbb{Q}_p\otimes\varprojlim_n H^2(K_n,T) \cong (V^*(1)^{G_{K_\infty}})^*$. Ainsi, si $\mathcal{B}$ est une base du $\mathbb{Q}_p$-espace vectoriel $\mathbf{D}_p(V)$ et $\mathcal{B}'$ un système libre de $Z^1_\infty(K,T)$ engendrant un Λ-module Z de $Z^1_\infty(K,T)$, si $\det_{\mathcal{B}'}\Omega_{V,h}(\mathcal{B})$ est le déterminant de $\Omega_{V,h}$ dans les systèmes libres $\mathcal{B}$ et $\mathcal{B}'$, si $F_{Z^1_\infty(K,T)/\mathcal{B}}$ est une série caractéristique du module $Z^1_\infty(K,T)/Z$ et $F_{T^*(1)^{G_{K_\infty}}}$ une série caractéristique de $T^*(1)^{G_{K_\infty}}$, on a

$$\delta_h(\Omega_V) = F_{Z^1_\infty(K,T)/\mathcal{B}}(F^\iota_{T^*(1)^{G_{K_\infty}}})^{-1}\det_{\mathcal{B}'}\Omega_{V,h}(\mathcal{B})\Lambda \ .$$

On pose ensuite

$$\delta(\Omega_V) = \prod_{j>-h} \ell_{-j}^{-\dim_{\mathbb{Q}_p}\mathrm{Fil}^j\,\mathbf{D}_p(V)}\delta_h(V)$$

qui est indépendant de h à condition que $\mathrm{Fil}^{-h}\,\mathbf{D}_p(V) = \mathbf{D}_p(V)$. Il est démontré dans [4] que $\delta(V)$ est contenu dans $\mathbb{Q}_p\otimes\Lambda$ et que (Réc(V)) implique que $\delta(V) = \mathbb{Q}_p\otimes\Lambda$. On obtient ainsi le théorème.

5.1.1. Théorème. *($\delta(V)$) Si V est une représentation cristalline, alors*

$$\delta(\Omega_V) = \mathbb{Q}_p \otimes \Lambda \ .$$

Soit $h \geq 1$ tel que $\mathrm{Fil}^{-h} \mathbf{D}_p(V) = \mathbf{D}_p(V)$. On déduit de $\delta(V)$ que

$$\mathcal{L}_h(z) \in \prod_{j>-h} \ell_{-j}^{-\dim_{\mathbb{Q}_p} \mathrm{Fil}^j \mathbf{D}_p(V)} (F^{\iota}_{T^*(1)^{G_{K_\infty}}})^{-1} \mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V) \ .$$

En particulier, si $Frac(\Lambda)$ est l'anneau total des fractions de Λ,

$$\mathcal{L}_h(z) \in \prod_{j>-h} \ell_{-j}^{-\dim_{\mathbb{Q}_p} \mathrm{Fil}^j \mathbf{D}_p(V)} Frac(\Lambda) \mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V) \ .$$

Soit $h^* \geq 1$ tel que $\mathrm{Fil}^{h^*} \mathbf{D}_p(V) = 0$ (remarquons que cela est équivalent à dire que $\mathrm{Fil}^{-h^*} \mathbf{D}_p(V^*(1)) = \mathbf{D}_p(V^*(1))$). Dans le cas où V contient $\mathbb{Q}_p(h)$, (resp. où $V^*(1)$ contient $\mathbb{Q}_p(h^*)$), on augmente h (resp. h^*) de 1. En utilisant le fait que $V^*(1)^{G_{K_\infty}}$ est de la forme $\oplus_{j \in J} V^*(1)(-j)^{G_K}(j)$ avec J un sous-ensemble de $]-h^*, \dots h[$, on en déduit qu'il existe des entiers α_j pour $-h < j < h^*$ tels que $\mathcal{L}_{V,h}(\prod_{-h<j<h^*} \ell_{-j}^{\alpha_j} x) \in \mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V)$. On a en fait la proposition plus précise suivante.

5.1.2. Proposition. *Si $x \in \mathcal{H}(G_\infty) \otimes Z^1_\infty(K, T)$, alors*

$$\mathcal{L}_{-h^*}(x) = \prod_{-h<j<h^*} \ell_{-j} \mathcal{L}_h(x)$$

appartient à $\mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V)$.

Remarquons que $x' = \prod_{-h<j<h^*} \ell_{-j} x$ vérifie automatiquement la condition que $\pi_{n,k}(x') \in H^1_e(K_n, V(k))$ pour tout $k \geq 1 - h$.

Démonstration. Soit $g = \mathcal{L}_{V,h}(\prod_{-h<j<h^*} \ell_{-j}^{\alpha_j} x) \in \mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V)$ avec $\alpha_j \geq 1$. On désire montrer que si $\alpha_j \geq 2$, il est possible de diviser g par ℓ_{-j} dans $\mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V)$, c'est-à-dire que g s'annule sur tout caractère du type $\chi^{-j}\eta$ avec η d'ordre fini. Soit $g_2 \in \mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V^*(1))$ quelconque, on a alors en remarquant que $\Omega_{V^*(1),1-h} = (\prod_{-h<j<h^*} l_j)^{-1} \Omega_{V^*(1),h^*}$

$$\begin{aligned}(-1)^h [g, \sigma_{-1} g_2^{\iota}]_{\mathbf{D}_p(V)} &=< \Omega_{V,h}(g), (\prod_{-h<j<h^*} l_j)^{-1} \Omega_{V^*(1),h^*}(g_2) >_V \\ &=< \prod_{-h<j<h^*} \ell_{-j}^{\alpha_j - 1} x, \Omega_{V^*(1),h^*}(g_2) >_V \in \prod_{-h<j<h^*} \ell_{-j}^{\alpha_j-1} \mathcal{H}(G_\infty) \ .\end{aligned} \tag{5.1.1}$$

Ainsi, si $\alpha_j \geq 2$, le dernier terme est nul sur tout caractère $\eta\chi^{-j}$; comme g_2 est quelconque, cela implique qu'il en est de même de g qui est donc divisible par ℓ_{-j}. □

5.1.3. PROPOSITION. *Soit J un ensemble fini d'entiers contenu dans $\{-h+1,\dots,h^*-1\}$ et J^c le complémentaire de J dans $\{-h+1,\dots,h^*-1\}$. On suppose que $x\in\mathcal{H}(G_\infty)\otimes Z^1_\infty(K,T)$ vérifie $\pi_{n,k}(x)\in H^1_e(K_n,V(k))$ pour tout entier $n\geq 0$ et pour tout $k\in J$. Alors, $g_x^{h,h^*,J}=\prod_{j\in J^c}\ell_{-j}\mathcal{L}_h(x)$ appartient à $\mathcal{H}(G_\infty)\otimes\mathbf{D}_p(V)$. Autrement dit, $\prod_{j\in J^c}\ell_{-j}.x$ appartient à l'image de $\mathcal{H}(G_\infty)\otimes\mathbf{D}_p(V)$ par $\Omega_{V,h}$. De plus, $\mathfrak{o}_\varphi(g_x^{h,h^*,J})=\mathfrak{o}(x)+h^*-\sharp J-1$.*

Démonstration. On prend $g=\mathcal{L}_{V,h}(\prod_{-h<j<h^*}\ell_{-j}x)$. On a alors comme précédemment

$$(-1)^h[g,\sigma_{-1}g_2^\iota]_{\mathbf{D}_p(V)}=<x,\Omega_{V^*(1),h^*}(g_2)>_V \quad . \tag{5.1.2}$$

Soit $k\in J$. Il s'agit de montrer que ℓ_{-k} divise g. Comme $h^*-k-1\geq 0$, $\pi_{n,-k}(\Omega_{V^*(1),h^*}(g_2))$ appartient à $H^1_f(K_n,V(k)^*(1))$, on en déduit que pour tout $g_2\in\mathcal{H}(G_\infty)\otimes\mathbf{D}_p(V)$, $\chi^{-k}\eta([g,\sigma_{-1}g_2^\iota]_{\mathbf{D}_p(V)})=0$ pour tout caractère d'ordre fini (cf. Appendice A.2, rappelons que l'orthogonal de $H^1_f(K_n,V(k))$ est égal à $H^1_f(K_n,V(k)^*(1))$ pour la dualité locale). Donc g est divisible par ℓ_{-k}.

La formule sur l'ordre de tempérence se déduit de ce que

$$\mathfrak{o}_\varphi(g_x^{h,h^*,J})+h=\mathfrak{o}(x)+h+h^*-1-\sharp J\ ,$$

(cf. 3.3). □

Prenons par exemple comme dans [1] $J=\{-r+1,\cdots,0\}$ avec $r=\mathfrak{o}(x)$. On a donc alors

$$\mathfrak{o}_\varphi(g_x^{h,h^*,J})=h^*-1\ .$$

Ainsi, $g_x^{h,h^*,J}$ est $p^{-(h^*-1)}\varphi^-$-bornée. On peut alors appliquer le lemme 4.3.2 : si l'on choisit un cocycle $Z(y)$ représentant $y=\prod_{j\in J^c}\ell_{-j}x$ avec $\pi_{n,k}(Z(y)_\tau)=(\chi(\tau)^k\tau-1)d_{n,k}(y)$ pour $k>-h$, la limite de

$$p^n\sum_{j=-h+1}^{h^*-1}(-1)^{j+h-1}\binom{h^*+h-2}{j+h-1}d_{n,j}(y)$$

existe et vaut

$$\lim_{m\to\infty}p^{m(h^*-1)}\Phi^{-m}(D^{-(h^*-1)}(G)([\epsilon]-1))t^{-(h^*-1)}=$$
$$\lim_{m\to\infty}\Phi^{-m}(D^{-(h^*-1)}(G)([\epsilon]-1)t^{-(h^*-1)})\ .$$

On remarque alors que l'on peut d'abord choisir un cocycle $Z(x)$ représentant x avec $\pi_{n,k}(Z(x)_\tau)=(\chi(\tau)^k\tau-1)d_{n,k}(x)$ pour $k>-h$ et que l'on peut alors prendre

$$d_{n,k}(y)=(\prod_{j\in J^c}j-k)d_{n,k}(x)$$

pour $j \in \{-h+1, \cdots, h^*-1\}$; en particulier, $d_{n,k}(y)$ est nul pour $k \in J^c$. Un calcul élémentaire montre que

$$\frac{1}{(h+h^*-2)!}\sum_{j=-h+1}^{h^*-1}(-1)^{j+h-1}\binom{h^*+h-2}{j+h-1}d_{n,j}(y) \\ = \frac{1}{(r-1)!}\sum_{j=-r+1}^{0}(-1)^{j+r-1}\binom{r-1}{j+r-1}d_{n,j}(x)$$

Faisons-le ! Le premier terme vaut

$$\frac{1}{(h+h^*-2)!}\sum_{j=-r+1}^{0}(-1)^{j+h-1}\binom{h^*+h-2}{j+h-1}\prod_{k=-h+1}^{-r}(k-j)\prod_{k=1}^{h^*-1}(k-j)d_{n,j}(x)\ .$$

On a pour j compris entre $-r+1$ et 0

$$\begin{aligned}\frac{(-1)^{j+h-1}}{(h+h^*-2)!}&\binom{h^*+h-2}{j+h-1}\prod_{k=-h+1}^{-r}(k-j)\prod_{k=1}^{h^*-1}(k-j)\\ &= (-1)^{r+1-j}\frac{1}{(h+j-1)!(h^*-j-1)!}\frac{(h+j-1)!}{(r+j)!}\frac{(h^*-1-j)!}{(-j)!}\\ &= (-1)^{r+1-j}\frac{1}{(r-1+j)!(-j)!} = \frac{(-1)^{r-1-j}}{(r-1)!}\binom{r-1}{j+r-1}\ .\end{aligned}$$

La limite de la suite $\frac{p^n}{(r-1)!}\sum_{j=-r+1}^{0}(-1)^{j+r-1}\binom{r-1}{j+r-1}d_{n,j}(x)$ lorque $n \to \infty$ est ce que Colmez appelle $Log_V^{(r)}(x)$ modulo un isomorphisme entre $(B_{\max}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))^{\Phi=1}$ et $\mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V)$ (voir appendice C). Ainsi,

$$\frac{1}{(h+h^*-2)!}\lim_{m\to\infty}p^{m(h^*-1)}\Phi^{-m}(D^{-(h^*-1)}(G)([\epsilon]-1))t^{-(h^*-1)} = Log_V^{(r)}(x)$$

pour $\Omega_{V,h}(g) = \prod_{j\in\{-h+1,\ldots,-r\}\cup\{1,\cdots,h^*-1\}}\ell_{-j}.x$. Remarquons que l'on peut préciser dans quel cran de la filtration il est. A priori, on obtient un élément de $(\mathrm{Fil}^{-(h^*-1)} B_{\max}^{G_{K_\infty}} \otimes \mathbf{D}_p(V))^{\Phi=1}$.

5.2. Un formulaire. Nous allons essayer de donner un formulaire complet. On a les formules

$$\mathcal{L}_{V,r} = \ell_r\mathcal{L}_{V,r+1}$$
$$\mathcal{L}_{V(k),r+k}(Tw^k(z)) = D^{-k}(\mathcal{L}_{V,r})\ .$$

Nous avons ici abandonné l'idée de ne pas identifier $\mathbf{D}_p(V(k))$ avec $\mathbf{D}_p(V)$! Nous réserverons la notation h pour un entier tel que $\mathrm{Fil}^{-h}\mathbf{D}_p(V) = \mathbf{D}_p(V)$. On fait agir φ sur $K_n \otimes \mathbf{D}_p(V)$ par $1 \otimes \varphi$. Soient $\rho =< \chi >^{k_\rho}\eta_\rho$ où k_ρ est un entier et η_ρ un caractère d'ordre fini et de conducteur $p^{\mathfrak{f}(\eta_\rho)}$. On note $\log_{V(\rho)}$ et

$\exp_{V(\rho),*}$ le logarithme et les exponentielles associés à la représentation twistée $V(\rho)$ et π_ρ l'application composée

$$Z^1_\infty(K,V) \overset{\pi_{k,f(\eta_\rho)}}{\to} H^1(K_{f(\eta_\rho)},V(k)) \overset{e_{\eta_\rho^{-1}}}{\to} H^1(K_{f(\eta_\rho)},V(k))^{(\eta_\rho)} = H^1(K,V(\rho))$$

Enfin, on pose

$$\begin{aligned}
\Gamma^*(\rho) &= \Gamma^*(k-h+1)\\
P_\rho(\varphi) &= \begin{cases} p^{f_\eta k_\rho}\varphi^{-f_\eta} & \text{si } \eta \text{ est non trivial}\\ (1-p^{k_\rho+1}\varphi^{-1})(1-p^{-k_\rho}\varphi) & \text{si } \eta \text{ est le caractère trivial}\end{cases}\\
\ell_\rho &= l_{k(\rho)} = \frac{\log \rho^{-1}(\tau)\tau}{\chi(\tau)}\\
G(\rho) &= G(\eta_\rho)
\end{aligned}$$

La proposition suivante est une simple traduction de résultats déjà démontrés:

5.2.1. Proposition. *Soit* $x \in \mathcal{H}(G_\infty)\otimes Z^1_\infty(K,V)$ *appartenant à l'image de* $\mathcal{H}(G_\infty)\otimes \mathbf{D}_p(V)$. *Soit* ρ *un caractère géométrique de* G_∞.

1. *Si* $k_\rho \geq 1-h$,
$$P(\varphi_\rho)(\rho^{-1}(\mathcal{L}_{V,h}(x))) = G(\rho^{-1})\frac{\log_{V(\rho)}(\pi_\rho(x))}{\Gamma^*(\rho\chi^{h-1})}$$
2. *Si* $k_\rho < 1-h$,
$$P_\rho(\varphi)(\rho^{-1}(\mathcal{L}_{V,h}(x))) = G(\rho^{-1})\frac{\lambda_{V(\rho),}(\pi_\rho(x))}{\Gamma^*(\rho\chi^{h-1})}$$

On ne suppose maintenant plus que x est dans l'image de $\mathcal{H}(G_\infty)\otimes \mathbf{D}_p(V)$. Il ne l'est en particulier pas si $\pi_{n,k}(x)$ n'appartient pas à H^1_g pour tout $n \geq 0$ et $k \geq 1-h$. On trouvera la démonstration des formules suivantes dans [7].

5.2.2. Proposition. *Soit* $x \in \mathcal{H}(G_\infty)\otimes Z^1_\infty(K,V)$.

1. *Si* $\mathrm{Fil}^{k_\rho}\mathbf{D}_p(V) = \mathrm{Fil}^0\mathbf{D}_p(V(\rho)) = 0$ *et si* $\mathbf{D}_p(V(\rho))^{\varphi=p^{-1}} = 0$,
$$P_\rho(\varphi)\rho^{-1}(\ell^{-1}_{\rho^{-1}}\mathcal{L}_{V,h}(x)) = G(\rho^{-1})\frac{\log_{V(\rho)}(\pi_\rho(x))}{\Gamma^*(\rho\chi^{h-1})}\ .$$
2. *Si* $\mathrm{Fil}^0\mathbf{D}_p(V(\rho)) \neq 0$ *et* $\mathbf{D}_p(V(\rho))^{\varphi=p^{-1}} = 0$, *alors*
$$P_\rho(\varphi)\rho^{-1}(\mathcal{L}_{V,h}(x)) = G(\rho^{-1})\frac{\lambda_{V(\rho),}(\pi_\rho(x))}{\Gamma^*(\rho\chi^{h-1})}$$
Si de plus $\pi_\rho(x) \in H^1_f(K,V(\rho))$, *on a*
$$P_\rho(\varphi)\rho^{-1}(\ell^{-1}_{\rho^{-1}}\mathcal{L}_{V,h}(x)) \equiv G(\rho^{-1})\frac{\log_{V(\rho)}\pi_\rho(x)}{\Gamma^*(\rho\chi^{h-1})} \mod \mathrm{Fil}^0\mathbf{D}_p(V(\rho))$$

3. *Si* $\mathbf{D}_p(V(\rho))^{\varphi=p^{-1}} \neq 0$, *alors*

$$(\rho^{-1}\mathcal{L}_{V(\rho),h}(x), -(1-p^{k_\rho+1}\varphi^{-1})(\rho^{-1}\mathcal{L}_{V(\rho),h}(x))) = (1-p^{-k_\rho}\varphi)\frac{\exp^*_{V(\rho),f}(\pi_\rho(x))}{\Gamma^*(\rho\chi^{h-1})} .$$

Si de plus $\pi_\rho(x) \in H^1_f(K, V(\rho))$, *on a*

$$(1-p^{k_\rho+1}\varphi^{-1})(\rho^{-1}(\ell^{-1}_{\rho^{-1}}\mathcal{L}_{V,h}(x))) \equiv \frac{\log_{f,1}(\pi_\rho(x)) - (1-p^{-k_\rho}\varphi)\log_{f,2}(\pi_\rho(x))}{\Gamma^*(\rho\chi^{h-1})} \bmod (1-p^{-k_\rho}\varphi)\,\mathrm{Fil}^0\,\mathbf{D}_p(V(\rho))$$

où $\log_{f,1}$ *et* $\log_{f,2}$ *désignent les composantes de l'application réciproque de* $\exp_f$ *(voir 3.1).*

5.3. Conjecture de Tamagawa locale. On renvoie à [4] et à [6] pour les conséquences sur les conjectures de Tamagawa locales. La loi de réciprocité implique que ces conjectures sont invariantes par twist. En particulier, on peut pour la démontrer twister V de manière à ce que $\mathrm{Fil}^0\,\mathbf{D}_p(V) = 0$ (un des nombres de Tamagawa est alors juste un cardinal d'un groupe de torsion).

Appendice A. Formules diverses

A.1. Lemme de Shapiro.

A.1.1. Soit G un groupe profini et H un sous-groupe fermé distingué de G. Soit M un H-module. On définit $\mathrm{Ind}\ M$ comme l'ensemble des applications localement constantes f de G dans M vérifiant $f(hx) = hf(x)$ pour $h \in H$. Le groupe G opère sur $\mathrm{Ind}\ M$ par $g(f)(x) = f(xg)$. L'application $\alpha : \mathrm{Ind}\ M \to M$ donnée par $\alpha(f) = f(1)$ est un homomorphisme de H-modules. On a en effet

$$\alpha(h(f)) = (hf)(1) = f(h) = hf(1) .$$

On en déduit une application de $Z^1(G, \mathrm{Ind}\ M)$ dans $Z^1(H, M)$ puis de $H^1(G, \mathrm{Ind}\ M)$ dans $H^1(H, M)$ qui est en fait un isomorphisme.
Le cas qui nous intéresse ici est celui où M est déjà muni d'une action de G-modules et où G/H est abélien et même cyclique. On a alors un isomorphisme de G-modules

$$\mathbb{Z}[G/H] \otimes M = M[G/H] \to \mathrm{Ind}\, M \ :$$

l'image de $\sum_{\tau\in G/H} a_\tau\tau$ est l'application $f : x \mapsto xa_{x^{-1}}$ avec un abus sur $a_{x^{-1}}$: il ne dépend que de l'image de x^{-1} dans G/H ; $f \in \mathrm{Ind}\ M$ car $f(hx) = hxa_{(hx)^{-1}} = hxa_{x^{-1}} = hf(x))$; l'application réciproque est donnée par $f \mapsto \alpha = \sum_{\tau\in G/H} \tilde\tau^{-1}(f(\tilde\tau))\tau^{-1}$ (on vérifie que la définition ne dépend pas du choix des représentants $\tilde\tau$ des éléments τ de G/H, puisque pour $h \in H$, $(h\tilde\tau)^{-1}(f(h\tilde\tau)) = \tilde\tau^{-1}h^{-1}h(f(\tilde\tau)) = \tilde\tau^{-1}(f(\tilde\tau))$, le composé des deux applications est d'une part $f \mapsto g$ avec $g(x) = xa_{x^{-1}} = xx^{-1}f(x) = f(x)$,

d'autre part $\alpha \mapsto \beta$ avec $\beta = \sum_{\tau\in G/H}\tilde\tau^{-1}(\tilde\tau a_{\tilde\tau^{-1}}).\tau^{-1} = \sum_{\tau\in G/H} a_\tau.\tau = \alpha$. L'action de G sur $\mathbb{Z}[G/H]\otimes M$ est l'action diagonale (si $g\in G$, l'image de gf est $\sum_{\tau\in G/H}\tilde\tau^{-1}(g(f)(\tilde\tau))\tau^{-1} = \sum_{\tau\in G/H}\tilde\tau^{-1}(f(\tilde\tau g))\tau^{-1} = \sum_{\tau\in G/H} g\tilde\tau^{-1}(f(\tilde\tau))g\tau^{-1} = g(\sum_{\tau\in G/H}\tilde\tau^{-1}(f(\tilde\tau))\tau^{-1})$. En composant avec l'application Ind $M\to M$, on obtient un homomorphisme de $\mathbb{Z}$-modules $\nu_{Id}^{G/H} : \mathbb{Z}[G/H]\otimes M \to M$ donnée par $\sum_{\tau\in G/H} a_\tau\tau \to a_{Id}$, qui induit un isomorphisme de G/H-modules $H^1(G,\mathbb{Z}[G/H]\otimes M)\cong H^1(H,M)$, l'action sur le premier étant donnée par l'action de G/H sur $\mathbb{Z}[G/H]$ par multiplication. On a $\nu_{Id}^{G/H}(g(\sum_{\tau\in G/H} a_\tau.\tau)) = \nu_{Id}^{G/H}(\sum_{\tau\in G/H} g(a_\tau).g\tau) = \nu_{Id}^{G/H}(\sum_{\tau\in G/H} g(a_{g^{-1}\tau}).\tau) = ga_{g^{-1}}$.
Notons $\nu_g^{G/H}$ l'application $\sum_{\tau\in G/H} a_\tau.\tau \mapsto a_{g^{-1}}$. On a donc $\nu_{Id}^{G/H}(gf) = g\nu_g^{G/H}(f)$ pour $f\in\mathbb{Z}[G/H]\otimes M$ et $g\in G$.

A.1.2. Reprenons la situation du texte. Si M est un G_∞-module avec action continue de G_∞, on identifie M et $M(k)$ en tant que $\mathbb{Z}_p$-modules, on note $\tau *_k m$ l'action sur $M(k)$: $\tau *_k m = \chi(\tau)^k\tau m$. On note $\nu^{G_n} = \nu^n$ pour alléger les notations.
On considère d'abord l'isomorphisme de G_∞-modules $\iota_k : \mathcal{H}(G_\infty)\otimes M \to \mathcal{H}(G_\infty)\otimes M(k)$ induit par $\tau\otimes m\mapsto \chi(\tau)^k\tau\otimes m$ [vérifions que c'est compatible avec l'action diagonale de G_∞ : $\iota_k(g(\tau\otimes m)) = \iota(g\tau\otimes gm) = \chi(g)^k\chi(\tau)^k g\tau\otimes gm = \chi(\tau)^k g\tau\otimes g *_k m = g *_k (\chi(\tau)^k\tau\otimes m) = g *_k \iota_k(\tau\otimes m)$]. On peut aussi écrire $\iota_k = Tw^k\otimes id$. Soit R_n la projection de $\mathcal{H}(G_\infty)$ sur $\mathbb{Q}_p[G_n]$. On pose alors

$$s_{n,k} = \nu^n_{Id}\circ R_n\circ\iota_k = \nu^n_{Id}\circ R_n\circ Tw^k\ ;$$

c'est une application de $\mathcal{H}(G_\infty)\otimes M\to M(k)$. Remarquons que $R_n\circ Tw^k$ a à voir avec le "polynôme d'interpolation". Ainsi, on peut écrire avec d'autres notations $R_{n,k}(f) = R_n\circ Tw^k(f)$ et $Tw^{-k}R_{n,k}(f)\equiv f \mod \chi(\gamma)^{-kp^n}\gamma^{p^n}-1$ ou $R_{n,k}(f)\equiv Tw^k f \mod \gamma^{p^n}-1$.
Vérifions pour se rassurer que $s_{n,k}$ est bien un homomorphisme de G_{K_n}-modules : on a en effet pour $f\in\mathcal{H}(G_\infty)$ et $m\in M$

$$\begin{aligned} s_{n,k}(g(f\otimes m)) &= \nu^n_{Id}(\chi(g)^k R_n(gTw^k(f)\otimes g(m)) \\ &= \chi(g)^k\nu^n_g(R_n(Tw^k(f))\otimes g(m) \\ &= \nu^n_g(R_n(Tw^k(f))\otimes g *_k m) \end{aligned}$$

Utilisons maintenant le fait que $g\in G_{K_n}$, ce qui implique que $\nu^n_g = \nu^n_{Id}$ et donc

$$\begin{aligned} s_{n,k}(g(f\otimes m)) &= \nu^n_{Id}(R_n(Tw^k(f))\otimes g *_k m) \\ &= g *_k \nu^n_{Id}(R_n(Tw^k(f))\otimes m) \\ &= g *_k s_{n,k}(f\otimes m) \end{aligned}$$

On désire maintenant décrire l'application

$$\pi_{n,k} : H^1(G_\infty,\mathcal{H}(G_\infty)\otimes M)\to H^1(K_\infty/K_n, M(k))\ .$$

induite par $s_{n,k} : \mathcal{H}(G_\infty) \otimes M \to M(k)$. Rappelons que l'on a des isomorphismes

$$H^1(G_\infty, \mathcal{H}(G_\infty) \otimes M) \xrightarrow{\cong} \mathcal{H}(G_\infty) \otimes M/(\gamma_n - 1)$$

et

$$H^1(K_\infty/K_n, M(k)) \xrightarrow{\cong} M(k)/(\chi(\gamma_n)^k \gamma_n - 1)$$

obtenu en fixant un générateur γ de G_∞; $\gamma_n = \gamma^{p^n}$ est alors un générateur de $\mathrm{Gal}(K_\infty/K_n)$: si Z est un cocycle de G_∞ (resp. de $\mathrm{Gal}(K_\infty/K_n)$), on lui associe $Z_\gamma \in \mathcal{H}(G_\infty) \otimes M$ (resp. Z_{γ_n}).
Soit donc $\tilde{Z} \in \mathcal{H}(G_\infty) \otimes M$ et Z le cocycle déterminé par $Z_\gamma = \tilde{Z}$. On a alors $Tw^k(Z_{\gamma^r}) = Tw^k(\sum_{i=0}^{r-1} \gamma^i \tilde{Z}) = \sum_{i=0}^{r-1} \chi(\gamma)^{ik} Tw^k(\gamma^i \tilde{Z})$. Le cocycle $\pi_{n,k}(Z)$ associé dans $Z^1(K_\infty/K_n, M(k))$ est déterminé par sa valeur en γ^{p^n} qui est

$$\begin{aligned}
\pi_{n,k}(Z)_{\gamma^n} &= \nu_{Id}^n(\sum_{i=0}^{p^n-1} \chi(\gamma)^{ik} R_n(\gamma^i Tw^k \tilde{Z}) \\
&= \sum_{i=0}^{p^n-1} \chi(\gamma)^{ik} \gamma^i \nu_{\gamma^i}(R_n(Tw^k \tilde{Z})) \\
&= \sum_{i=0}^{p^n-1} \gamma^i *_k \nu_{\gamma^i}(R_n(Tw^k \tilde{Z})) \\
&= \tilde{\nu}_k^n(R_n(Tw^k \tilde{Z}))
\end{aligned}$$

avec

$$\tilde{\nu}_k^n(\sum_{i=0}^{p^n-1} a_{\gamma^i} \otimes \gamma^i) = \sum_{i=0}^{p^n-1} \gamma^i *_k a_{\gamma^{-i}} \ .$$

Si maintenant λ_s est un homomorphisme de M dans N vérifiant $\lambda_s(\tau m) = \chi^s(\tau)\lambda_s(m)$, on a pour $f \in \mathbb{Z}_p[G_n] \otimes M$, $\lambda_s \circ \tilde{\nu}_k^n(f) = \lambda_s(\chi^{-k-s}(f))$, d'où

$$\lambda_s(\pi_{n,k}(Z)_{\gamma^n}) = \lambda_s(\chi^{-k-s}(R_n(Tw^k \tilde{Z}))) = \lambda_s(\chi^{-k-s}(R_{n,k}(\tilde{Z})))$$

Si on écrit $\tilde{Z} = f(\gamma - 1)$ avec $f \in \mathcal{H} \otimes (\mathbb{Z}_p[\Delta] \otimes M)$, si $u = \chi(\gamma)$, $R_{n,k}(f)$ est le polynôme en T de degré $< p^n$ tel que $R_{n,k}(f) \equiv f(u^k(1+T) - 1)$ mod $(1+T)^{p^n} - 1$ et la formule devient

$$\lambda_s(\pi_{n,k}(Z)_{\gamma^n}) = \lambda_s(R_{n,k}(f)(u^{-k-s} - 1)) \ .$$

A.2. Formulaire d'évaluation. Rappelons que l'on a un isomorphisme canonique de G_∞-modules entre Λ et $\mathbb{Z}_p[[T]]^{\psi=0}$ qui se prolonge en un isomorphisme entre $\mathcal{H}(G_\infty)$ et $\mathcal{H}_\infty^{\psi=0}$. Il est induit par $\tau \mapsto (1+T)^{\chi(\tau)}$ pour $\tau \in G_\infty$. D'où l'isomorphisme canonique $\mathcal{D}_{\infty,f}(V) \cong \mathcal{H}(G_\infty) \otimes \mathbf{D}_p(V)$. Si $g \in \mathcal{H}_\infty^{\psi=0} \otimes \mathbf{D}_p(V)$, on posera $g = \hat{g}.(1+T)$.

A.2.1. Si η est un caractère d'ordre fini de conducteur p^n avec $n \geq 0$, on note $e_\eta = \sum_{\tau \in \mathrm{Gal}(K_n/K)} \eta(\tau)\tau$. Soit $\rho = \eta\chi^k$ un caractère continu de G_∞ à valeurs dans $\mathbb{C}_p^*$ avec η un caractère d'ordre fini de $\mathrm{Gal}(K_n/K)$ de conducteur p^n (c'est-à-dire ne se factorisant pas par K_{n-1}). On peut évaluer les éléments de $\mathcal{H}(G_\infty)$ sur un tel caractère. On a

$$e_\eta g(\zeta_n - 1) = G(\eta)\eta^{-1}(\hat{g})$$

avec

$$G(\eta) = e_\eta(\zeta_n) = \sum_{\tau \in \mathrm{Gal}(K_n/K)} \eta(\tau)\tau\zeta_n$$

la somme de Gauss associée à η. On a d'autre part

$$D^k(g) = Tw^k(\hat{g}).(1+T)\ .$$

On en déduit que

$$G(\eta)\rho^{-1}(\hat{g}) = e_\eta D^{-k}(g)(\zeta_n - 1)\ . \tag{A.2.1}$$

A.2.2. On a une application $R_{n,k} : \mathcal{H}(G_\infty) \to \mathbb{Q}_p[G_n]$, composé du twist Tw^k et de la projection sur $\mathbb{Q}_p[G_n]$. Si $x \in \mathcal{H}(G_\infty) \otimes Z^1_\infty(K,T)$ et $y \in \mathcal{H}(G_\infty) \otimes Z^1_\infty(K,T^*(1))$, l'image twistée de $< x, y >_V$ dans $\mathbb{Z}_p[G_n]$ est donnée par

$$\begin{aligned} R_{n,k}(< x, y >_V) &= R_n(< Tw^{-k}(x), Tw^k(y) >_{V(k)} \\ &= \sum_{\tau \in G_n} < \tau^{-1}.\pi_{n,-k}(x), \pi_{n,k}(y) >_{V(-k),K_n} \tau\ . \end{aligned}$$

En prenant le coefficient de $Id \in G_n$, on obtient que

$$s_{n,k}(< x, y >_V) =< \pi_{n,-k}(x), \pi_{n,k}(y) >_{V(-k),K_n}\ .$$

Enfin,

$$\begin{aligned} \rho^{-1}(< x, y >_V) &= \eta^{-1}(R_{n,-k}(< x, y >_V)) \\ &= \sum_{\tau \in G_n} < \tau^{-1}.\pi_{n,k}(x), \pi_{n,-k}(y) >_{V(k),K_n} \eta^{-1}(\tau) \\ &=< \sum_{\tau \in G_n} \eta^{-1}(\tau)\tau^{-1}.\pi_{n,k}(x), \pi_{n,-k}(y) >_{V(k),K_n} \end{aligned}$$

D'où

$$\rho^{-1}(< x, y >_V) = (\sharp G_n)^{-1} < e_\eta \pi_{n,k}(x), e_{\eta^{-1}}\pi_{n,-k}(y) >_{V(k),K_n} \tag{A.2.2}$$

A.2.3. Passons aux formules concernant le produit de convolution. On a $g_1 * g_2 = \hat{g_1}\hat{g_2}.(1+T)$. On a $D^k(g_1 * g_2) = D^k(g_1) * D^k(g_2)$. D'autre part, on note ι l'involution de $\mathcal{H}_\infty^{\psi=0}$ correspondant à l'involution ι de $\mathcal{H}(G_\infty)$ changeant τ en τ^{-1}. On a alors $D^k(g^\iota) = D^{-k}(g)^\iota$ et $D^k(g_1 * g_2^\iota) = D^k(g_1) * D^{-k}(g_2)^\iota$. Enfin, $\sigma_{-1} \circ D^k(g) = (-1)^k D^k(\sigma_{-1} g)$.

Le polynôme d'interpolation de $g_1 * g_2$ modulo $(1+T)^{p^n}-1$ est

$$R_n(g_1 * g_2) = \sum_{j=0,(j,p)=1}^{p^n-1} \frac{1}{p^n} \sum_{\zeta\in\mu_{p^n}} g_1(\zeta^{j^{-1}}-1)g_2(\zeta^{-1}-1)(1+T)^j \ .$$

D'où,

$$s_{n,k}(\hat{g_1}\hat{g}_2^{\iota}) = s_{n,0}(Tw^k(\hat{g_1}\hat{g}_2^{\iota})) = \frac{1}{p^n}\sum_{\zeta\in\mu_{p^n}} D^k(g_1)(\zeta-1)D^{-k}(g_2)^{\iota}(\zeta^{-1}-1) \ .$$

Enfin,

$$e_\eta(D^{-k}(g_1 * g_2^{\iota})) = G(\eta)\rho^{-1}(\hat{g}_1)\rho(\hat{g}_2) \ .$$

Appendice B. Interpolation

Soit u un générateur topologique de $1+p\mathbb{Z}_p$. Un élément $f \in \mathcal{H}_{h^-}$ est connu par ses polynômes d'interpolation modulo les $u^{-ip^n}(1+T)^{p^n}-1$ pour $i \in \{0,\dots h-1)$ et on peut calculer $f(u^k-1)$ pour tout entier k comme une limite de combinaisons linéaires des $R_{n,i}(f)(u^j-1)$ pour $i\in\{0,\dots h-1)$ ([4, lemme 1.3.4]. Nous allons ici démontrer la formule exacte.

Pour tout entier i, on désigne par $R_{n,i}(f)$ est le polynôme de degré $< p^n$ tel que

$$f \equiv R_{n,i}(f)(u^{-i}(1+T)-1) \mod u^{-ip^n}(1+T)^{p^n}-1 \ .$$

En particulier, on a $f(u^i-1) = R_{n,i}(f)(0)$.

Lemme. *Si $f \in \mathcal{H}_{h^-}$, alors pour tout entier $k \geq h$ (resp. pour tout élément k de $\mathbb{Z}_p - \{0,\cdots h-1\}$), on a*

$$(-1)^{h-1}\frac{(h-1)!}{k(k-1)\dots(k-h+1)}f(u^k-1) = \lim_{n\to\infty}\sum_{i=0}^{h-1}(-1)^i\binom{h-1}{i}\frac{R_{n,i}(f)(u^{k-i}-1)}{k-i}.$$

La formule peut encore s'écrire pour $f \in \mathcal{H}_{h^-}(G_\infty)$ et $k\in\mathbb{Z}_p-\{0,\cdots h-1\}$,

$$(-1)^{h-1}\frac{(h-1)!}{k(k-1)\dots(k-h+1)} <\chi>^k (f) = \lim_{n\to\infty}\sum_{i=0}^{h-1}(-1)^i\binom{h-1}{i} <\chi>^{k-i} \frac{R_n(Tw^i(f))}{k-i} \ .$$

Ici, $<\chi>$ est la projection de χ sur $1+p\mathbb{Z}_p$. La formule s'étend par continuité à tout élément de $\mathbb{Z}_p-\{0,\cdots,h-1\}$.

Ce lemme ou ses variantes est à la base de tous les calculs de valeurs de fonctions obtenues par interpolation p-adique. Lorsque $f \in \Lambda$ ou $\mathcal{H}_{1^-}$, le lemme dit simplement que

$$f(u^k-1) = \lim_{n\to\infty} R_{n,0}(f)(u^k-1)\ .$$

Démonstration. Nous avons choisi une démonstration “élémentaire”. Il suffit de démontrer la formule pour $k \geq h$ et de conclure par continuité.
Si g est une fonction sur les entiers positifs, on définit (cf [8])

$$\delta_s(g) = \sum_{r=0}^{s}(-1)^r\binom{s}{r}g(s-r)\ .$$

On a alors la formule d'inversion

$$g(n) = \sum_{s=0}^{\infty}\delta_s(g)\binom{n}{s}\ .$$

En particulier, si les $\delta_s(g)$ sont nuls pour $s \geq h$, on a

$$g(k) = \sum_{s=0}^{h-1}\delta_s(g)\binom{k}{s}\ .$$

et toute valeur de g sur un entier positif s'exprime uniquement en fonction de $g(0)$, $g(1)$, ... , $g(h-1)$. Plus précisément,

$$g(k) = \sum_{i=0}^{h-1}(-1)^i\binom{h-1}{i}c_{h,k,i}g(i)$$

avec

$$c_{h,k,i} = (-1)^i\frac{k!(h-1-i)!}{(h-1)!(k-i)!}\sum_{s=0}^{h-1-i}(-1)^s\binom{k-i}{s}\ .$$

On remarque alors que l'on a l'identité (que l'on peut montrer par récurrence)

$$\sum_{s=0}^{u}(-1)^s\binom{v}{s} = (-1)^u\binom{v-1}{u}$$

pour $0 < u < v$. On obtient alors

$$\begin{aligned} c_{h,k,i} &= (-1)^{h-1}\frac{k!(h-1-i)!}{(h-1)!(k-i)!}\binom{k-i-1}{h-1-i} \\ &= (-1)^{h-1}\frac{k!}{(k-h)!(h-1)!(k-i)}\ . \end{aligned}$$

Ainsi, si g est une fonction sur les entiers telle que $\delta_s(g) = 0$ pour tout entier $s \geq h$, on a pour tout entier $k \geq h$

$$(-1)^{h-1}\frac{(h-1)!}{k(k-1)\cdots(k-h+1)}g(k) = \sum_{i=0}^{h-1}(-1)^i\binom{h-1}{i}\frac{g(i)}{k-i}$$

Revenons au lemme à démontrer et posons $g_n(i) = R_{n,i}(f)(u^{k-i}-1)$. Le fait que $f \in \mathcal{H}_{h^-}$ implique que pour tout entier $s \geq h$, la suite $\delta_s(g_n) = \sum_{i=0}^{s}(-1)^i\binom{s}{i}g_n(s-i)$ tend vers 0 avec n. On en déduit que pour tout entier $k \geq h$, la limite de $g_n(k) - \sum_{s=0}^{h-1}\delta_s(g_n)\binom{k}{s}$ est nulle. Il ne reste plus qu'à utiliser le calcul précédent et à remarquer que $g_n(k) = f(u^k-1)$ pour tout entier $n > 0$ pour conclure. □

Lemme. *Si $f \in \mathcal{H}_{h^-}$ et P est un polynôme de degré $< t$, on a*

$$\lim_{n\to\infty} p^{n\inf(t,h)}\sum_{i=0}^{h-1}(-1)^i\binom{h-1}{i}P(k-i)R_n(f)(k-i) = 0\ .$$

Démonstration. Soit g une fonction sur les entiers vérifiant $\delta_r(g) = 0$ pour $r \geq h-1$ et P un polynôme de degré $< t$ avec $t \geq 1$; on a en $P(x) = \sum_{s=0}^{t-1}\delta_s(P)\binom{x}{s}$ (ces deux polynômes de degré $\leq t-1$ coïncident en $x = 0, \cdots, t-1$ et sont donc égaux), ainsi, $\delta_s(P) = 0$ pour $s \geq t$). On a

$$\begin{aligned}\delta_{h-1}(Pg) &= \sum_{j=0}^{h-1}\delta_j(P)\delta_{h-1-j}(g)\\ &= \sum_{j=0}^{t-1}\delta_j(P)\delta_{h-1-j}(g)\ .\end{aligned}$$

Prenons pour P le polynôme $Q = P(k-x)$ et remplaçons g par $g_n = i \mapsto R_{i,n}(f)(u^{k-i}-1)$. Rappelons que $p^n\delta_r(g_n) \to 0$ pour $r \geq h$ et que $p^{nj}\delta_{h-1-j}(g_n) \to 0$ lorsque $n \to \infty$ pour $0 \leq j \leq h-1$. On déduit alors du calcul précédent que

$$\begin{aligned}\lim_{n\to\infty} p^{nt}\sum_{i=0}^{h-1}(-1)^i\binom{h-1}{i}&P(k-i)R_{i,n}(f)(u^{k-i}-1)\\ &= \lim_{n\to\infty} p^{nt}\delta_{h-1}(Qg_n)\\ &= \lim_{n\to\infty} p^{nt}\sum_{j=0}^{t-1}\delta_j(P)\delta_{h-1-j}(g) = 0\end{aligned}$$

Si $t \leq h$, on a encore

$$\delta_{h-1}(Pg) = \sum_{j=0}^{h-1}\delta_j(P)\delta_{h-1-j}(g)$$

et

$$\lim_{n\to\infty} p^{nh} \sum_{i=0}^{h-1} (-1)^i \binom{h-1}{i} P(k-i) R_{i,n}(f)(u^{k-i}-1)$$
$$= \lim_{n\to\infty} p^{nh} \sum_{j=0}^{h-1} \delta_j(P)\delta_{h-1-j}(g)$$
$$= \lim_{n\to\infty} p^{n(h-j)} \sum_{j=0}^{h-1} \delta_j(P) p^{nj} \delta_{h-1-j}(g) = 0$$

□

PROPOSITION. *Si $f \in \mathcal{H}_{h^-}$, alors pour tout entier $k \geq h$, on a*

$$\frac{(-1)^h (h-1)!}{k(k-1)\dots(k-h+1)} \frac{f(u^k-1)}{\log u} =$$
$$\lim_{n\to\infty} \sum_{i=0}^{h-1} (-1)^i \frac{p^n}{1-u^{(k-i)p^n}} \binom{h-1}{i} R_{n,i}(f)(u^{k-i}-1) \ .$$

Nous avonc utilisé cette proposition pour des éléments de $f \in \mathcal{H}_{h^-}(G_\infty)$. Elle se traduit alors par la formule

$$\begin{aligned}&\frac{(-1)^h (h-1)!}{k(k-1)\dots(k-h+1)} \frac{\chi^k(f)}{\log \chi(\gamma)} = \\ &\lim_{n\to\infty} \sum_{i=0}^{h-1} (-1)^i \frac{p^n}{1-\chi(\gamma)^{(k-i)p^n}} \binom{h-1}{i} \chi^{k-i}(R_{n,i}(f))\end{aligned} \tag{B.0.3}$$

Démonstration. On écrit $\frac{1}{1-e^T} + \frac{1}{T} = \sum_{j=0}^{\infty} c_j T^j$ avec $p^{r_0} p^j c_j \in \mathbb{Z}_p$ pour r_0 indépendant de j. On a alors

$$\frac{p^n}{1-u^{(k-i)p^n}} + \frac{1}{(k-i)\log u} = \sum_{j=0}^{\infty} c_j p^{n(j+1)} (k-i)^j$$
$$= \sum_{j=0}^{h-1} p^{n-j} p^j c_j p^{nj} (k-i)^j + \sum_{j=h}^{\infty} p^{n(j+1-h)-j} p^j c_j p^{nh} (k-i)^j \ .$$

Comme $n(j+1-h) - j \geq 1-h$ pour $j \geq h$, on en déduit du lemme précédent que si $F_n(i) = \frac{p^n}{1-u^{(k-i)p^n}} + \frac{1}{(k-i)\log u}$ que

$$\lim_{n\to\infty} \sum_{i=0}^{h-1} (-1)^i \binom{h-1}{i} F_n(i) R_{i,n}(f)(u^{k-i}-1) = 0 \ .$$

Il ne reste plus qu'à appliquer le premier lemme. □

Appendice C. Suite exacte de Coleman-Colmez

C.1. Rappelons la définition suivante :

Définition : Soit $\mathcal{D}$ un espace vectoriel normé de dimension finie muni d'un automorphisme u. Si $\epsilon \in \{\pm\}$, on dit qu'un élément $F \in \mathcal{H} \otimes \mathcal{D}$ est u^ϵ-borné si la suite $||(1 \otimes u)^{-n} F||_{\rho_n}$ est bornée pour $\epsilon = +$ et tend vers 0 pour $\epsilon = -$.

On note $(\mathcal{H} \otimes \mathcal{D})_{u^\epsilon}$ l'ensemble des éléments u^ϵ-bornés et on pose alors $||F||_\varphi = C_u(F) = \sup_n(||(1 \otimes u)^{-n} F||_{\rho_n})$.

C.2. Fixons un $\mathbb{Q}_p$-espace vectoriel $\mathcal{D}$ de dimension finie muni d'un automorphisme φ. Colmez démontre le théorème suivant (nous avons déjà utilisé et démontré le A):

C.2.1. Théorème. (Colmez) A) *Soit F un élément de $\mathcal{H} \otimes \mathcal{D}$ tel que*

$$(1 - \Phi)F \in (\mathcal{H} \otimes \mathcal{D})_{\varphi^-} \ .$$

Alors la suite $\Phi^{-n}(F([\epsilon] - 1)) = (1 \otimes \varphi)^{-n} F(\beta_n - 1)$ converge dans $B_{\mathrm{dR}} \otimes \mathcal{D}$ vers un élément α_F de $((B^+_{\max})^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$.
B) *Réciproquement, soit α un élément de $((B^+_{\max})^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$. Il existe une série $F_\alpha \in \mathcal{H} \otimes \mathcal{D}$ telle que $(1 - \Phi)F_\alpha \in (\mathcal{H} \otimes \mathcal{D})_{\varphi^-}$ et telle que $\alpha = \alpha_{F_\alpha}$.*
C) *L'application $\alpha \mapsto F_\alpha$ est une bijection entre $((B^+_{\max})^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$ et les éléments de $\mathcal{H} \otimes \mathcal{D}$ tels que $(1 - \Phi)F \in (\mathcal{H}^{\psi=0} \otimes \mathcal{D})_{\varphi^-}$.*

On en déduit une application

$$\mathcal{C}_\mathcal{D} : ((B^+_{\max})^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1} \to (\mathcal{H}^{\psi=0} \otimes \mathcal{D})_{\varphi^-}$$

donnée par $\alpha \mapsto (1 - \Phi)F_\alpha$.

C.3. Corollaire. *On a la suite exacte de G_∞-modules*

$$\begin{aligned} 0 \to \oplus_{k \geq 0} t^k \mathcal{D}^{\varphi = p^{-k}} \to ((B^+_{\max})^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1} &\to (\mathcal{H}^{\psi=0} \otimes \mathcal{D})_{\varphi^-} \\ &\to \oplus_{k \geq 0} \mathcal{D}/(1 - p^k \varphi)(k) \to 0 \end{aligned}$$

et on a $||\mathcal{C}_\mathcal{D}(\alpha)||_\varphi = ||\alpha||_{\max}$.

C.4. Remarques. 1. Un cas particulier est celui où $\mathcal{D} = \mathbf{D}_p(\mathbb{Q}_p(1))$. On obtient alors la suite exacte de Coleman :

$$0 \to \mathbb{Z}_p t \to ((B^+_{\max})^{G_{K_\infty}})^{\varphi = p} \to \Lambda \otimes \mathbf{D}_p(\mathbb{Q}_p(1)) \to \mathbb{Z}_p(1) \to 0$$

2. L'idée fondamentale de Colmez est de montrer la convergence des éléments du type $(1 \otimes \varphi)^{-n} F(\beta_n - 1)$ vers un élément de $((B^+_{\max})^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$ et de construire l'application réciproque de $F \mapsto \alpha_F$, c'est-à-dire de construire une série tempérée à partir d'un élément de $((B^+_{\max})^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$. Pour cela, il a besoin d'opérateurs de trace sur $(B^+_{\max})^{G_{K_\infty}}$ que nous allons introduire dans le paragraphe C.6. Ces opérateurs nous permettront aussi de compléter le théorème en reliant α avec les valeurs de F_α.

C.5. QUELQUES PROPRIÉTÉS DE $(B^+_{\max})^{G_{K_\infty}}$. Enonçons sans le démontrer la proposition cruxiale suivant [1, lemme VIII.3.3] :

C.5.1. PROPOSITION. (Colmez) *Si $n \geq 1$, tout élément α de $B^+_{\max} \cap K_n[[t]]$ s'écrit de manière unique sous la forme $\alpha = F(\beta_n - 1)$ où $F \in K[[T]]$ a un rayon de convergence $\geq \rho_n$. On a de plus $||F||_{\rho_n} \leq ||F(\beta_n - 1)||_{\max} \leq p||F||_{\rho_n}$.*

Ainsi, si $F = \sum_{k=0}^{\infty} a_k T^k$, la suite $v_p(a_k) + \frac{k}{(p-1)p^{n-1}}$ en k tend vers $+\infty$. On définit un opérateur $\tilde{\delta}_k$ sur $K_\infty[[t]]$ par

$$\alpha = \sum_{k=0}^{\infty} \tilde{\delta}_k(\alpha) t^k$$

et on pose $\delta_k = \tilde{\delta}_k t^k$. L'opérateur δ_k n'est pas continu pour la topologie de B_{dR} et ne se prolonge pas à $(B^+_{\mathrm{dR}})^{G_{K_\infty}}$.

C.5.2. LEMME. *Les opérateurs $\tilde{\delta}_k$ et D sont reliés par*

$$\tilde{\delta}_k(F(\beta_n - 1)) = \frac{D^k(F)(\zeta_n - 1)}{p^{nk} k!}$$

Démonstration. Comme $\beta_n = \zeta_n \exp(t/p^n)$, et que $\tilde{\delta}$ laisse fixe K_n, on a

$$\begin{aligned}
\tilde{\delta}_k(F(\beta_n - 1)) &= \frac{1}{k!} \frac{d}{d^k T} (F(\zeta_n \exp(T/p^n) - 1))_{T=0} \\
&= \frac{1}{p^{nk} k!} \frac{d}{d^k T} (F(\zeta_n \exp(T) - 1))_{T=0} \\
&= \frac{D^k(F(\zeta_n(1+T) - 1))_{T=0}}{p^{nk} k!} \\
&= \frac{D^k(F)(\zeta_n - 1)}{p^{nk} k!}
\end{aligned}$$

□

C.6. LE PROJECTEUR T_n DE $(B^+_{\mathrm{dR}})^{G_{K_\infty}}$ SUR $K_n[[t]]$. L'inclusion de $K_n[[t]]$ dans $(B^+_{\mathrm{dR}})^{G_{K_\infty}}$ admet une section naturelle définie par Colmez et dont les propriétés sont résumées dans la proposition suivante. On note Tr_{K_m/K_n} l'application de $K_m[[t]]$ induite par la trace sur K_m et par l'identité sur t.

C.6.1. PROPOSITION. *Pour $n \geq 1$, il existe une unique application $\mathbb{Q}_p$-linéaire continue T_n de $(B^+_{\mathrm{dR}})^{G_{K_\infty}}$ dans $K_n[[t]]$ vérifiant*

$$T_n(x) = \frac{1}{p^m} Tr_{K_m/K_n}(x)$$

pour $x \in K_m[[t]]$ et $m \geq n$. Elle vérifie les propriétés suivantes :

1.
$$T_n(\beta_m) = \begin{cases} 0 & \textit{pour } m > n \\ \frac{1}{p^n}\beta_m & \textit{pour } m \leq n \end{cases}$$

2. $\lim_{n\to\infty} p^n T_n(\alpha) = \alpha$ *pour* $\alpha \in (B_{\mathrm{dR}}^+)^{G_{K_\infty}}$ *;*
3. $||p^n T_n(\alpha)||_{\max} \leq ||\alpha||_{\max}$ *pour* $\alpha \in B_{\max}^*)^{G_{K_\infty}}$*;*
4. *Si* $\alpha \in (B_{\mathrm{cont}}^+)^{G_{K_\infty}}$ *avec* $B_{\mathrm{cont}}^+ = \cap\varphi^n(B_{\max}^+)$*,* $\varphi^n p^n T_n \varphi^{-n}(\alpha)$ *est indépendant de* $n \geq 1$*.*

Ainsi, on a $T_{n-1}\circ\varphi = p\varphi\circ T_n$, $p^n T_n$ est l'identité sur $K_n[[t]]$ et fournit une section de $K_n[[t]] \to (B_{\mathrm{dR}}^+)^{G_{K_\infty}}$. Si $\alpha \in \varphi((B_{\mathrm{dR}}^+)^{G_{K_\infty}})$, par exemple si $\alpha \in (B_{\mathrm{cont}}^+)^{G_{K_\infty}}$, on pose

$$\tilde{T}_0(\alpha) = \varphi(pT_1(\varphi^{-1}(\alpha)))$$

C.6.2. Lemme. *Soit* α *un élément de* $((B_{\max}^+)^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$*. Il existe une unique série* $F_\alpha \in \mathcal{H}\otimes\mathcal{D}$ *telle que*

$$\tilde{T}_0(\alpha) = F_\alpha([\epsilon]-1)\ .$$

De plus, $(1-\Phi)F_\alpha \in (\mathcal{H}^{\psi=0}\otimes\mathcal{D})_{\varphi^-}$*.*

Démonstration. Soit $\alpha \in ((B_{\max}^+)^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$. On vérifie facilement que $((B_{\max}^+)^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1} = ((B_{\mathrm{cont}}^+)^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$. Soit $\delta = pT_1((\varphi^{-1}\otimes 1)\alpha)$. C'est un élément de $((B_{\mathrm{cont}}^+)^{G_{K_\infty}} \cap K_1[[t]])\otimes\mathcal{D})$. Grâce à la proposition C.5.1, il existe une unique série $F_\alpha \in \mathbb{Q}_p[[T]]\otimes\mathcal{D}$ de rayon de convergence $\geq p^{-\frac{1}{p-1}}$ tel que $\delta = F_\alpha(\beta_1 - 1)$. On a alors

$$\tilde{T}_0(\alpha) = (\varphi\otimes 1)\delta = (\varphi\otimes 1)F_\alpha(\beta_1-1) = F_\alpha([\epsilon]-1)\ .$$

Montrons que $F_\alpha \in \mathcal{H}\otimes\mathcal{D}$. Comme

$$p^n(\varphi^n\otimes 1)T_n((\varphi^{-n}\otimes 1)\alpha) = p(\varphi\otimes 1)T_1((\varphi^{-1}\otimes 1)\alpha) = \tilde{T}_0(\alpha)\ ,$$

on a

$$F_\alpha(\beta_n - 1) = p^n T_n((\varphi^{-n}\otimes 1)\alpha) \in (B_{\max}^{G_{K_\infty}} \cap K_n[[t]])\otimes\mathbf{D}_p(V)\ .$$

En appliquant de nouveau la proposition C.5.1, on en déduit que le rayon de convergence de F_α est $\geq \rho_n$ pour tout n et donc que $F_\alpha \in \mathcal{H}\otimes\mathcal{D}$. On vérifie que la limite de $F_\alpha(\beta_n - 1)$ est α. Utilisons maintenant l'invariance de α par Φ pour montrer que $\psi((1-\Phi)F_\alpha) = 0$. Par définition, il est équivalent de montrer que

$$\sum_{\zeta\in\mu_p} F_\alpha(\zeta(1+T)-1) = p(1\otimes\varphi)F_\alpha((1+T)^p - 1)$$

ou encore de montrer l'égalité obtenue en remplaçant T par un quelconque $\beta_{n+1}-1$ pour $n \geq 1$. On a

$$\begin{aligned}\sum_{\zeta\in\mu_p} F_\alpha(\zeta\beta_{n+1}-1) =& Tr_{K_{n+1}[[t]]/K_n[[t]]}(F_\alpha(\beta_{n+1}-1)\\ =& Tr_{K_{n+1}[[t]]/K_n[[t]]}(p^{n+1}T_{n+1}((\varphi^{-n+1}\otimes 1)\alpha))\\ =& p^{n+1}T_n((\varphi^{-n+1}\otimes 1)\alpha) = p^{n+1}T_n((\varphi^{-n}\otimes\varphi)\alpha)\end{aligned}$$

en utilisant le fait que $\varphi\otimes\varphi(\alpha)=\alpha$

$$=p(1\otimes\varphi)p^nT_n((\varphi^{-n}\otimes 1)\alpha) = p(1\otimes\varphi)F_\alpha(\beta_n-1)$$

Montrons enfin que $f=(1-\Phi)F_\alpha$ est φ^--borné. On montre facilement que

$$(1\otimes\varphi)^{-n}f(\beta_n-1) = p^nT_n(\Phi^{-n}\alpha) - p^{n-1}T_{n-1}(\Phi^{-(n-1)}\alpha)\ .$$

Comme $\Phi(\alpha)=\alpha$, cela vaut aussi

$$p^nT_n(\alpha) - p^{n-1}T_{n-1}(\alpha)\ ,$$

ce qui tend vers 0 dans $B_{\max}\otimes\mathcal{D}$. En utilisant en fait que $||(1\otimes\varphi)^{-n}f(\beta_n-1)||_{\max}\sim||(1\otimes\varphi)^{-n}f||_{\rho_n}$, on en déduit que f est φ^--borné. □

C.6.3. Démontrons le théorème C.2.1. Soit $F\in\mathcal{H}\otimes\mathcal{D}$ tel que $(1-\Phi)F\in(\mathcal{H}^{\psi=0}\otimes\mathcal{D})_{\varphi^-}$. Soit $\alpha_F=\lim_{m\to\infty}\Phi^{-m}(F([\epsilon]-1))$; calculons $T_n(\alpha_F)$. Par continuité de T_n, on a

$$T_n(\alpha_F) = \lim_{m\to\infty} T_n((1\otimes\varphi)^{-m}F(\beta_m-1))$$

En voyant $(1\otimes\varphi)^{-m}F(\beta_m-1)$ dans $K_m[[t]]$ grâce à la formule $\beta_m=\zeta_m\exp(t/p^m)$, on a pour $m\geq n\geq 1$,

$$\begin{aligned}p^nT_n((1\otimes\varphi)^{-m}F(\beta_m-1)) =& \frac{1}{p^{m-n}}Tr_{K_m[[t]]/K_n[[t]]}((1\otimes\varphi)^{-m}F(\beta_m-1))\\ =& \frac{1}{p^{m-n}}\sum_{\zeta\in\mu_{p^{m-n}}}((1\otimes\varphi)^{-m}F(\zeta\beta_m-1))\end{aligned}$$

Le fait que $\psi((1-\Phi)F)=0$ implique que $\psi(F)=(1\otimes\varphi)F$ et donc que $(1\otimes\varphi^{-n})F=\psi^{m-n}((1\otimes\varphi^{-m})F)$. On en déduit que

$$p^nT_n((1\otimes\varphi)^{-m}F(\beta_m-1)) = (1\otimes\varphi^{-n})F(\beta_n-1)\ .$$

D'où

$$p^nT_n(\alpha_F) = (1\otimes\varphi)^{-n}F(\beta_n-1)\ .$$

La formule pour $\delta_k(T_n(\alpha_F))$ se déduit du lemme C.5.2.
Si $\alpha\in((B^+_{\max})^{G_{K_\infty}}\otimes\mathcal{D})^{\Phi=1}$ et F_α construit comme dans le lemme C.6.2, on a vu dans la démonstration que

$$(1\otimes\varphi^{-n})F_\alpha(\beta_n-1) = p^nT_n(\Phi^{-n}\alpha) = p^nT_n(\alpha)\ .$$

On en déduit que $\alpha \mapsto F_\alpha \mapsto \alpha_{F_\alpha}$ est l'identité sur $((B^+_{\max})^{G_{K_\infty}} \otimes \mathcal{D})^{\Phi=1}$ et que $F \mapsto \alpha_F \mapsto F_{\alpha_F}$ est l'identité à condition de se restreindre aux F tels que $\psi((1-\Phi))F = 0$.

Références

[1] P. Colmez, *Théorie d'Iwasawa des représentations de de Rham,* Annals of Math. 148 (1998), 485-571.

[2] P. Colmez, *Représentations cristallines et représentations de hauteur finie,* prépublication LMENS 97-28, à paraître dans J. reine angew. Math.

[3] F. Destrempes, *Explicit reciprocity laws for Lubin-Tate modules,* J. reine angew. Math. 463 (1995), 27-47.

[4] B. Perrin-Riou, *Théorie d'Iwasawa des représentations cristallines,* Invent. Math. 115 (1994), 81-149.

[5] B. Perrin-Riou, Fonctions L p-adiques des représentations p-adiques, Astérisque 229 (1995), SMF (Paris).

[6] B. Perrin-Riou, *Fonctions L p-adiques,* dans Proc. Int. Congress Math., Zürich (1994), pp. 400–410, Birkhäuser Verlag (1995).

[7] B. Perrin-Riou, *Zéros triviaux des fonctions L p-adiques, un cas particulier,* Compos. Math. 114 (1998), 37-76.

[8] J.-P. Serre, *Formes modulaires et fonctions zéta p-adiques,* dans Modular functions of one vaiable 111 (1972), Lecture Notes in Math. 350.

Bernadette Perrin-Riou
Mathématiques, Bât 425
Université Paris-Sud
F-91405 Orsay Cedex
France
Bernadette.PERRIN-RIOU@math.u-psud.fr

A Minimax Principle for Eigenvalues in Spectral Gaps: Dirac Operators with Coulomb Potentials [1]

Marcel Griesemer, Roger T. Lewis, Heinz Siedentop

Received: March 29, 1999

Communicated by Bernold Fiedler

Abstract. We prove the minimax principle for eigenvalues in spectral gaps introduced in [5] based on an alternative set of hypotheses. In the case of the Dirac operator these new assumptions allow for potentials with Coulomb singularites.

1991 Mathematics Subject Classification: 47A75, 81Q10
Keywords and Phrases: Minimax principle, Dirac operator, Coulomb singularity

1 Introduction

Recently Dolbeault, Esteban, and Séré [4, 3, 2] have found a minimax principle for Dirac operators with Coulomb potentials. Independently, Griesemer and Siedentop [5] have found a minimax principle characterizing the eigenvalues of self-adjoint operators in their spectral gaps, which is flexible enough to adapt to various situations. In particular it can also be applied to Dirac operators. Such a minimax principle is of particular interest for applications, e.g., in solid state physics and relativistic quantum chemistry where differential operators having gaps in their spectra naturally arise. Apart from the computational point of view (see, e.g., Kutzelnigg [7]) it can serve as a tool to obtain non-asymptotic eigenvalue estimates, e.g., comparing the number of eigenvalues of

[1] This work has been partially supported by the European Union through the TMR network FMRX-CT 96-0001.

the Dirac operator in the gap with the number of negative eigenvalues of a corresponding Schrödinger operator (see [5]).
Comparing [3, 2] and [5] shows, that although the hypotheses for the validity of the minimax principle overlap, the methods of proof are quite different. On the other hand, with these different hypotheses different classes of operators can be treated: Dolbeault, Esteban, and Séré's result allows for Dirac operators with singular potentials of Coulomb type. Griesemer and Siedentop's result allows for a flexible formulation of the minimax principle adaptable to various situations, e.g., an earlier minimax principle for the first positive eigenvalue of the Dirac operator considered by Talman [9] and Datta and Deviah [1] can be proved.
This difference in hypotheses indicates that the optimal assumption for the abstract minimax principle is yet to be found. The present paper is a step in this direction.
In Section 2 we prove the abstract minimax principle under assumptions alternative to those in [5]. In Section 3 we show that these hypotheses allow for Dirac operators with Coulomb potentials. Applications to other self-adjoint operators with eigenvalues in spectral gaps like perturbed periodic Schrödinger operators are also conceivable.

2 The Minimax Principle

In this section we formulate and prove the abstract minimax principle. Suppose A and A_0 are self-adjoint operators in a Hilbert space $\mathfrak{H}$ and assume that their form domains are equal

$$\mathfrak{Q}(A) = \mathfrak{Q}(A_0) = \mathfrak{Q}. \tag{1}$$

Let $\mathfrak{D}(A)$ and $\mathfrak{D}(A_0)$ denote the domains of A and A_0 respectively and let $P_I(A)$ be the spectral projection of A corresponding to the interval $I \subset \mathbb{R}$. Define

$$\begin{aligned} \Lambda_+ &= P_{(0,\infty)}(A_0), \qquad & \Lambda_- &= 1 - \Lambda_+, \\ P_+ &= P_{(0,\infty)}(A), \qquad & P_- &= 1 - P_+. \end{aligned} \tag{2}$$

We set $\mathfrak{H}_\pm := \Lambda_\pm \mathfrak{H}$ and $\mathfrak{Q}_\pm := \Lambda_\pm \mathfrak{Q}$. Then $\mathfrak{H} = \mathfrak{H}_+ \oplus \mathfrak{H}_-$ and, by assumption (1), $\mathfrak{Q}_\pm \subset \mathfrak{Q}$. The minimax values in which we are interested are given by

$$\lambda_n(A) := \inf_{\substack{\mathfrak{M}_+ \subset \mathfrak{Q}_+ \\ \dim(\mathfrak{M}_+) = n}} \sup_{\substack{\psi \in \mathfrak{M}_+ \oplus \mathfrak{Q}_- \\ \|\psi\| = 1}} (\psi, A\psi), \tag{3}$$

and have been introduced in [5]. These minimax values are to be compared with the standard (Courant) minimax values

$$\mu_n(B) := \inf_{\substack{\mathfrak{M} \subset \mathfrak{Q}(B) \\ \dim(\mathfrak{M}) = n}} \sup_{\substack{\psi \in \mathfrak{M} \\ \|\psi\| = 1}} (\psi, B\psi)$$

for the eigenvalues of a self-adjoint operator B which is bounded from below. The value $\mu_n(B)$ is the n-th eigenvalue of B counting from below (see, e.g., Reed and Simon [8]).

Theorem 1. *Suppose A and A_0 are self-adjoint operators in $\mathfrak{H}$ with the same form domain $\mathfrak{Q}$ and define $\Lambda_\pm$, $P_\pm$, $\mathfrak{Q}_\pm$, $\lambda_n(A)$ and $\mu_n(\cdot)$ as above. If $(\psi, A\psi) \le 0$ for all $\psi \in \mathfrak{Q}_-$ and if*

$$\|(|A_0|+1)^{1/2}\Lambda_+ P_-(|A_0|+1)^{-1/2}\| < 1 \tag{4}$$

then $\lambda_n(A) = \mu_n(A\restriction P_+\mathfrak{H})$ for all $n \le \dim \mathfrak{H}_+$.

We remark that $|A_0|+1$ can be replaced by $|A_0|$ in (4), if we assume that 0 is in the resolvent set of A_0. This will be obvious from the proof.

Proof. We prove the theorem in two steps. Although these are partly contained in [5] we do not omit the similar parts in order to be self-contained: First, we show that it suffices to prove that $\Lambda_+ : P_+\mathfrak{Q} \to \mathfrak{Q}_+$ is a bijection. Secondly, we verify this property using assumption (4) and the negativity of $(\psi, A\psi)$ on $\mathfrak{Q}_-$.

Step 1. If $\Lambda_+ P_+ \mathfrak{Q} = \mathfrak{Q}_+$, then we have

$$\lambda_n(A) = \inf_{\substack{\mathfrak{M}_+ \subset \Lambda_+ P_+ \mathfrak{Q} \\ \dim(\mathfrak{M}_+)=n}} \sup_{\substack{\psi \in \mathfrak{M}_+ \oplus \mathfrak{Q}_- \\ \|\psi\|=1}} (\psi, A\psi) \tag{5}$$

using the defining Equation (3). Since for each $\mathfrak{M}_+ \subset \Lambda_+ P_+ \mathfrak{Q}$ with $\dim(\mathfrak{M}_+) = n$, we can find a subspace $\mathfrak{M} \subset P_+\mathfrak{Q}$ with $\dim(\mathfrak{M}) = n$ such that $\mathfrak{M}_+ = \Lambda_+\mathfrak{M}$ and since $\Lambda_+\mathfrak{M} \oplus \mathfrak{Q}_- \supset \mathfrak{M}$, we get from (5)

$$\begin{aligned}\lambda_n(A) &= \inf_{\substack{\mathfrak{M}_+ \subset \Lambda_+ P_+ \mathfrak{Q} \\ \dim(\mathfrak{M}_+)=n}} \sup_{\substack{\psi \in \mathfrak{M}_+ \oplus \mathfrak{Q}_- \\ \|\psi\|=1}} (\psi, A\psi) \\ &\ge \inf_{\substack{\mathfrak{M} \subset P_+ \mathfrak{Q} \\ \dim(\mathfrak{M})=n}} \sup_{\substack{\psi \in \mathfrak{M} \\ \|\psi\|=1}} (\psi, A\psi) = \mu_n(A\restriction P_+\mathfrak{H}).\end{aligned}$$

To prove the converse inequality we proceed as in [5]: pick $\epsilon > 0$ and let $\mathfrak{M} := P_{(0,\mu_n+\epsilon)}(A)\mathfrak{Q}$. Then $\dim(\mathfrak{M}) \ge n$ and hence $\dim(\Lambda_+\mathfrak{M}) \ge n$ by the remark above. Therefore

$$\lambda_n \le \sup_{\substack{\psi \in \Lambda_+\mathfrak{M} \oplus \mathfrak{Q}_- \\ \|\psi\|=1}} (\psi, A\psi) = \sup_{\substack{\psi \in \mathfrak{M} + \mathfrak{Q}_- \\ \|\psi\|=1}} (\psi, A\psi),$$

where $\Lambda_+\mathfrak{M} \oplus \mathfrak{Q}_- = \mathfrak{M} + \mathfrak{Q}_-$ was used. To estimate this from above we first decompose $\psi \in \mathfrak{M} + \mathfrak{Q}_-$ as $\psi = \psi_1 + \psi_2$, where $\psi_1 \in \mathfrak{M}$ and $\psi_2 \in \mathfrak{M}^\perp \cap (\mathfrak{M} + \mathfrak{Q}_-)$, and then ψ_2 as $\psi_2 = \psi_3 + \psi_-$ where $\psi_3 \in \mathfrak{M}$ and $\psi_- \in \mathfrak{Q}_-$.

Since $A\psi_3 \in \mathfrak{M}$ and $\psi_3 + \psi_- \in \mathfrak{M}^\perp$ we have $(A\psi_3, \psi_-) = -(A\psi_3, \psi_3)$. Using this, $(A\psi_3, \psi_3) \geq 0$, and $(\psi_-, A\psi_-) \leq 0$ we find

$$\begin{aligned}(\psi, A\psi) &= (\psi_1, A\psi_1) + (\psi_2, A\psi_2)\\ &= (\psi_1, A\psi_1) - (\psi_3, A\psi_3) + (\psi_-, A\psi_-) \leq (\psi_1, A\psi_1) \leq (\mu_n + \epsilon)(\psi, \psi)\end{aligned}$$

which implies $\lambda_n \leq \mu_n$.

Step 2. Surjectivity: Since $\Lambda_+ P_+ \mathfrak{Q} \subset \mathfrak{Q}_+$ it suffices that $\Lambda_+ P_+ \mathfrak{Q}_+ = \mathfrak{Q}_+$, which is equivalent to $(|A_0|+1)^{1/2}\Lambda_+ P_+ (|A_0|+1)^{-1/2}\mathfrak{H}_+ = \mathfrak{H}_+$. Now $\Lambda_+ P_+ = 1 - \Lambda_+ P_-$ on $\mathfrak{H}_+$ so that

$$(|A_0| + 1)^{1/2}\Lambda_+ P_+ (|A_0| + 1)^{-1/2} = 1 - (|A_0| + 1)^{1/2}\Lambda_+ P_- (|A_0| + 1)^{-1/2}$$

on $\mathfrak{H}_+$. By assumption (4) the latter is an isomorphism from $\mathfrak{H}_+$ to $\mathfrak{H}_+$.

Injectivity: Suppose $\Lambda_+ : P_+\mathfrak{Q} \to \mathfrak{Q}_+$ would not be one-to-one. Then there would exist a non-zero $\psi \in \mathfrak{H}_- \cap P_+\mathfrak{Q}$ such that

$$0 \geq (\psi, A\psi) = (P_+\psi, AP_+\psi) > 0.$$

$\square$

3 Application to the Dirac Operator

The hypothesis (4) of Theorem 1 contains the a priori unknown operator P_-, i.e., it is not straightforward to check. In this section we will show how to verify it for given operators nevertheless. To be specific we restrict ourselves to the Dirac operator D_γ with a screened Coulomb potential, i.e., $D_\gamma := (1/i)\nabla \cdot \boldsymbol{\alpha} + m\beta - \gamma\varphi$ in $\mathfrak{H} := L^2(\mathbb{R}^3)^4$, where $\varphi(x) = y(x)/|x|$ with measurable y and $y(\mathbb{R}^3) \subset [0, 1]$. By Hardy's inequality we have that D_γ is an operator perturbation of D_0 for $\gamma \in (-1/2, 1/2)$. We will assume this restriction on γ henceforth. In particular, perturbation theory for $|D_0| = (-\Delta + m^2)^{1/2}$ implies by Hardy's and Kato's inequality

$$\forall_{\gamma\in[0,1/2)} \quad \mathfrak{D}(D_\gamma) = H^1(\mathbb{R}^3) \otimes \mathbb{C}^4 =: \mathfrak{D}, \tag{6}$$

$$\forall_{\gamma\in[0,2/\pi)} \quad \mathfrak{Q}(D_\gamma) = H^{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^4 =: \mathfrak{Q} \tag{7}$$

for the operator and form domain of D_γ, respectively. To make connections with Section 2 we pick $A_0 := D_0$ and $A := D_\gamma$. The notation (2) is used correspondingly here.

By γ_0 we denote the real solution of $2\gamma_0^3 - 3\gamma_0^2 + 4\gamma_0 = 1$. Note that $0.305 < \gamma_0 < 0.306$ holds.

Theorem 2. *For $\gamma \in [0, \gamma_0)$*

$$\inf_{\substack{\mathfrak{M}_+\subset\mathfrak{Q}_+\\ \dim\mathfrak{M}_+=n}} \sup_{\substack{\psi\in\mathfrak{M}_+\oplus\mathfrak{Q}_-\\ \|\psi\|=1}} (\psi, D_\gamma\psi) \tag{8}$$

is equal to the n-th positive eigenvalue – counting multiplicity – of the Dirac operator D_γ or equals the mass m.

Our strategy is to roll the proof back to a verification of the hypotheses of Theorem 1. The main step is the verification of (4) which we break up into several steps:

Lemma 1. *For all* $f \in \mathfrak{H}$

$$\begin{aligned} \Lambda_+ P_- f = & -\tfrac{\gamma}{2\pi}\Lambda_+ \int_{-\infty}^{\infty}(D_0 - iz)^{-1}\varphi(D_\gamma - iz)^{-1}dzf \\ = & -\tfrac{\gamma}{\pi}\Lambda_+ \int_0^{\infty}\left[(D_0^2+z^2)^{-1}(D_0\varphi D_\gamma - z^2\varphi)(D_\gamma^2+z^2)^{-1}\right]dzf. \end{aligned} \tag{9}$$

Proof. Since for $\gamma \in [0, 2/\pi)$, zero is in the resolvent set of D_γ, we have that

$$P_\pm = \frac{1}{2} \pm \frac{1}{2\pi}\int_{-\infty}^{\infty}(D_\gamma - iz)^{-1}dz = \frac{1}{2} \pm \frac{1}{\pi}\int_0^{\infty} D_\gamma(D_\gamma^2+z^2)^{-1}dz \tag{10}$$

(Kato [6], Chapter VI.5, Lemma 5.6); $\Lambda_\pm$ is obtained from (10) by setting $\gamma = 0$. Therefore, by (10), and the second resolvent identity

$$P_- = \Lambda_- - \frac{\gamma}{2\pi}\int_{-\infty}^{\infty}(D_0 - iz)^{-1}\varphi(D_\gamma - iz)^{-1}dz$$

from which we may conclude that the first part of (9) holds.
We can simplify

$$\begin{aligned} & \int_{-\infty}^{\infty}(D_0 - iz)^{-1}\varphi(D_\gamma - iz)^{-1}dzf \\ = & \int_0^{\infty}\left[(D_0 - iz)^{-1}\varphi(D_\gamma - iz)^{-1} + (D_0 + iz)^{-1}\varphi(D_\gamma + iz)^{-1}\right]dzf \\ = & \int_0^{\infty}\left[\frac{D_0+iz}{D_0^2+z^2}\varphi\frac{D_\gamma+iz}{D_\gamma^2+z^2} + \frac{D_0-iz}{D_0^2+z^2}\varphi\frac{D_\gamma-iz}{D_\gamma^2+z^2}\right]dzf \\ = & \; 2\int_0^{\infty}\left[(D_0^2+z^2)^{-1}(D_0\varphi D_\gamma - z^2\varphi)(D_\gamma^2+z^2)^{-1}\right]dzf \end{aligned}$$

which implies that the second part of (9) holds. □

Lemma 2. *For* $\gamma \in \mathbb{R}_+$ *we have* $(1/2-\gamma)^2\varphi^2 \le |D_\gamma|^2 \le (1+2\gamma)^2|D_0|^2$.

Proof. For all $\psi \in \mathfrak{D}(D_0)$ we have $\|D_\gamma\psi\| \ge \|D_0\psi\| - \gamma\|\varphi\psi\| \ge (1/2-\gamma)\|\varphi\psi\|$, where we first use the triangle inequality and then Hardy's inequality. This implies the first stated operator inequality. The second one follows from $\|D_\gamma\psi\| \le \|D_0\psi\| + \gamma\|\varphi\psi\| \le (1+2\gamma)\|D_0\psi\|$. □

Lemma 3. *For all* $\gamma \in (0, \frac{1}{2})$ *and* $f \in \mathfrak{H}$ *we have*

$$\begin{aligned} & \||D_0|^{1/2}\int_0^{\infty}(D_0^2+z^2)^{-1}(D_0\varphi D_\gamma - z^2\varphi)(D_\gamma^2+z^2)^{-1}dz|D_0|^{-1/2}f\| \\ & \qquad \le \pi\frac{\sqrt{1+2\gamma}}{1-2\gamma}\|f\|. \end{aligned} \tag{11}$$

Proof. Using the fact that

$$\|h\| = \sup_{\|g\|=1} |(g,h)|, \quad h \in \mathfrak{H}$$

and setting $f' := |D_0|^{-1/2} f$ we see that the norm on the left hand side of (11) can be approximated by finding an upper bound for

$$|(g, |D_0|^{1/2} \int_0^\infty \left[(D_0^2+z^2)^{-1}(D_0\varphi D_\gamma - z^2\varphi)(D_\gamma^2+z^2)^{-1}\right] dz f')|, \quad \|g\| = 1. \tag{12}$$

First, consider the term

$$\begin{aligned} &|(g, |D_0|^{1/2} \int_0^\infty \left[(D_0^2+z^2)^{-1}(D_0\varphi D_\gamma)(D_\gamma^2+z^2)^{-1}\right] dz f')| \\ \leq &\left[\int_0^\infty \|D_0(D_0^2+z^2)^{-1}|D_0|^{1/2} g\|^2 dz\right]^{\frac{1}{2}} \left[\int_0^\infty \|\varphi D_\gamma (D_\gamma^2+z^2)^{-1} f'\|^2 dz\right]^{\frac{1}{2}}. \end{aligned} \tag{13}$$

Note that

$$\int_0^\infty \frac{dz}{(1+z^2)^2} = \int_0^\infty \frac{z^2 dz}{(1+z^2)^2} = \frac{\pi}{4}. \tag{14}$$

Thus, the first factor yields

$$\int_0^\infty \|D_0(D_0^2+z^2)^{-1}|D_0|^{1/2} g\|^2 dz = \int_0^\infty (g, \frac{|D_0|^3}{(D_0^2+z^2)^2} g) dz = \frac{\pi}{4}(g,g). \tag{15}$$

In a similar manner we show for $\gamma \in (0, 1/2)$

$$\int_0^\infty \|\varphi D_\gamma (D_\gamma^2+z^2)^{-1} f'\|^2 dz \tag{16}$$

$$= \int_0^\infty (f', (D_\gamma^2+z^2)^{-1} D_\gamma \varphi^2 D_\gamma (D_\gamma^2+z^2)^{-1} f') dz \tag{17}$$

$$\leq \frac{1}{(1/2-\gamma)^2} \int_0^\infty (f', (D_\gamma^2+z^2)^{-1} |D_\gamma|^4 (D_\gamma^2+z^2)^{-1} f') dz \tag{18}$$

$$= \frac{\pi}{(1-2\gamma)^2}(f', |D_\gamma| f') \leq \frac{\pi(1+2\gamma)}{(1-2\gamma)^2}(f', |D_0| f') \leq \frac{\pi(1+2\gamma)}{(1-2\gamma)^2}(f,f) \tag{19}$$

where we have used the first inequality of Lemma 2 to go from (17) to (18) and the second inequality of that Lemma in (19).
Thus we have for the product

$$|(g, |D_0|^{1/2} \int_0^\infty \left[(D_0^2+z^2)^{-1}(D_0\varphi D_\gamma)(D_\gamma^2+z^2)^{-1}\right] dz f')| \leq \frac{\pi}{2} \frac{\sqrt{1+2\gamma}}{1-2\gamma} \|f\|.$$

Likewise, we estimate the second term in (12)

$$
\begin{aligned}
&|(g, |D_0|^{1/2} \int_0^\infty (D_0^2+z^2)^{-1} z^2 \varphi (D_\gamma^2+z^2)^{-1} dz |D_0|^{-1/2} f)| \\
&\qquad = |\int_0^\infty (z(D_0^2+z^2)^{-1}|D_0|^{1/2} g, z\varphi(D_\gamma^2+z^2)^{-1} f') dz| \\
&\le \left[\int_0^\infty \|z(D_0^2+z^2)^{-1}|D_0|^{1/2} g\|^2 dz\right]^{\frac12} \left[\int_0^\infty \|z\varphi(D_\gamma^2+z^2)^{-1} f'\|^2 dz\right]^{\frac12}. \quad (20)
\end{aligned}
$$

By scaling and (14) we get for the first factor

$$\int_0^\infty \|z|D_0|^{1/2}(D_0^2+z^2)^{-1} g\|^2 dz = \frac{\pi}{4}. \tag{21}$$

The second factor yields using Lemma 2 twice

$$
\begin{aligned}
&\int_0^\infty \|z\varphi(D_\gamma^2+z^2)^{-1} f'\|^2 dz = (f', \int_0^\infty (D_\gamma^2+z^2)^{-1} \varphi^2 z^2 (D_\gamma^2+z^2)^{-1} dz f') \\
\le& \frac{1}{(1/2-\gamma)^2} (f', \int_0^\infty (D_\gamma^2+z^2)^{-1} |D_\gamma|^2 z^2 (D_\gamma^2+z^2)^{-1} dz f') \\
=& \frac{\pi}{4(1/2-\gamma)^2} (f', D_\gamma f') \le \pi \frac{1+2\gamma}{(1-2\gamma)^2} (f', D_0 f').
\end{aligned}
$$

Thus we get

$$|(g, |D_0|^{1/2} \int_0^\infty (D_0^2+z^2)^{-1} z^2 \varphi (D_\gamma^2+z^2)^{-1} dz f')| \le \frac{\pi}{2} \frac{\sqrt{1+2\gamma}}{1-2\gamma} \|f\|, \tag{22}$$

i.e., the same upper bound as for the first term. By (11), (12), and the calculations above, we have the upper bound

$$
\begin{aligned}
&\||D_0|^{1/2} \int_0^\infty \left[(D_0^2+z^2)^{-1}(D_0\varphi D_\gamma - z^2\varphi)(D_\gamma^2+z^2)^{-1}\right] dz |D_0|^{-1/2} f\| \\
&\qquad\qquad \le \pi \frac{\sqrt{1+2\gamma}}{1-2\gamma} \|f\|
\end{aligned}
$$

for $\gamma \in [0, 1/2)$ which we claimed. □

From Lemmata 1 and 3 we have the immediate

Corollary 1. *For all* $\gamma \in (0, \frac12)$

$$\||D_0|^{1/2} \Lambda_+ P_- |D_0|^{-1/2}\| \le \gamma \frac{\sqrt{1+2\gamma}}{1-2\gamma}.$$

We remark that an argument similar to the proofs of Lemmata 1 and 3 shows that $\|\Lambda_+ P_-\| = O(\gamma)$ as $\gamma \to 0$ which implies that $\Lambda_+ P_+ \mathfrak{H} = \mathfrak{H}_+$ and $\mathfrak{H}_+ \cap P_- \mathfrak{H} = \{0\}$ for small enough positive γ.
We turn now to the proof of Theorem 2.

Proof. First, we reiterate our remark (7) that for $\gamma \in [0, 2/\pi)$ the form domain of $\mathfrak{Q} := \mathfrak{Q}(D_\gamma) = H^{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^4$. In particular, it is independent of γ. This also means that $P_\pm$ and $\Lambda_\pm$ leave $\mathfrak{Q}$ invariant. Moreover, $\Lambda_- D_\gamma \Lambda_-$ is certainly non-positive. Finally, Corollary 1 implies that (4) holds true for $\gamma \in [0, \gamma_0)$ which completes the proof. □

Finally, we remark, that the construction of this Section is easily generalized to other types of potentials, as long as one can prove an analogue of Lemma 3.

Acknowledgment. This work has been partially supported by the European Union through its Training, Research, and Mobility program, grant FMRX-CT 96-0001.

REFERENCES

[1] S. N. Datta and G. Deviah. The minimax technique in relativistic Hartree-Fock calculations. *Pramana*, 30(5):387–405, May 1988.

[2] J. Dolbeault, M. J. Esteban, and E. Séré. Variational characterization for eigenvalues of Dirac operators. *Preprint, mp-arc: 98-177*, 1998.

[3] Jean Dolbeault, Maria J. Esteban, and Eric Séré. International Conference on Differential Equations and Mathematical Physics, Atlanta, Georgia, March 23–29, 1997.

[4] Maria J. Esteban and Eric Séré. Existence and multiplicity of solutions for linear and nonlinear Dirac operators. In *Paritial Differential Equations and their Applications (Toronto, ON, 1995)*, pages 107–118. Amer. Math. Soc., Providence, RI, 1997.

[5] Marcel Griesemer and Heinz Siedentop. A minimax principle for the eigenvalues in spectral gaps. *J. London Math. Soc.*, Accepted for publication. *Preprint, mp-arc 97-492*, 1997.

[6] Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1 edition, 1966.

[7] Werner Kutzelnigg. Relativistic one-electron Hamiltonians 'for electrons only' and the variational treatment of the Dirac equation. *Chemical Physics*, 1997.

[8] Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics*, volume 4: Analysis of Operators. Academic Press, New York, 1 edition, 1978.

[9] James D. Talman. Minimax principle for the Dirac equation. *Phys. Rev. Lett.*, 57(9):1091–1094, September 1986.

Marcel Griesemer
Department of Mathematics
University of Alabama
at Birmingham
Birmingham, AL 35294-1170
USA
marcel@math.uab.edu

Roger T. Lewis
Department of Mathematics
University of Alabama
at Birmingham
Birmingham, AL 35294-1170
USA
lewis@math.uab.edu

Heinz Siedentop
Mathematik I
Universität Regensburg
D-93040 Regensburg
Germany
Heinz.Siedentop@mathematik.uni-regensburg.de

Presentations of Subshifts and Their Topological Conjugacy Invariants

Kengo Matsumoto

Received: November 23, 1998
Revised: May 19, 1999

Communicated by Joachim Cuntz

Abstract. We introduce the notions of symbolic matrix system and λ-graph system that are presentations of subshifts. They are generalized notions of symbolic matrix and λ-graph for sofic subshifts to general subshifts. We then formulate strong shift equivalence and shift equivalence between symbolic matrix systems and show that two subshifts are topologically conjugate if and only if the associated canonical symbolic matrix systems are strong shift equivalent. We construct several kinds of shift equivalence invariants for symbolic matrix systems. They are the dimension groups, the Bowen-Franks groups and the nonzero spectrum that are generalizations of the corresponding notions for nonnegative matrices. The K-groups for symbolic matrix systems are introduced. They are also shift equivalence invariants and stronger than the Bowen-Franks groups but weaker than the dimension triples. These kinds of shift equivalence invariants naturally induce topological conjugacy invariants for subshifts.

1991 Mathematics Subject Classification: Primary 58F03, Secondary 54H20, 34C35, 19K14, 19K33
Keywords and Phrases: subshifts, symbolic dynamics, topological conjugacy, strong shift equivalence, shift equivalence, dimension group, Bowen-Franks group

Contents

1.Introduction

The classification of symbolic dynamical systems has been a very important and one of central problems in the theory of topological dynamical systems and the ergodic theory. The classification problem has been first examined for a class of symbolic dynamical systems called subshifts of finite type or topological Markov shifts. Each dynamical system of the class is determined by a single square matrix with entries in nonnegative integers. Hence the behavior of such a dynamical system depends on the underlying matrix. In [Wi], R. F. Williams introduced the notions of strong shift equivalence and shift equivalence between nonnegative matrices and showed that two topological Markov shifts are topologically conjugate if and only if the associated matrices are strong shift equivalent. He also showed that strong shift equivalence implies shift equivalence. Although the converse implication had been a long standing problem, Kim-Roush [KimR2] has recently solved negatively for even irreducible matrices. There is a class of subshifts called sofic subshifts that are generalized class of Markov shifts and that are determined by square matrices with entries in alphabet (see [Kit], [Kr3], [LM], [We], etc.). A square matrix with entries in alphabet is simply called a symbolic matrix. It is an equivalent object to a labeled graph called a λ-graph. M. Nasu in [N], [N2] generalized the notion of strong shift equivalence to symbolic matrices. He showed that two sofic subshifts are topologically conjugate if and only if their canonical symbolic matrices are strong shift equivalent ([N], [N2], see also [HN]). M. Boyle and W. Krieger in [BK] introduced the notion of shift equivalence for symbolic matrices and studied topologically conjugacy for sofic subshifts.

In this paper, we first introduce the notions of symbolic matrix system and λ-graph system. They are generalized notions of symbolic matrix and λ-graph for sofic subshifts. We will show that they are presentations of subshifts. A symbolic matrix system consists of two sequences of rectangular matrices $(\mathcal{M}_{l,l+1}, I_{l,l+1}), l \in \mathbb{N}$. The matrices $\mathcal{M}_{l,l+1}$ have entries in symbols and the matrices $I_{l,l+1}$ have entries in $\{0,1\}$. They satisfy the following commutation relations

$$I_{l,l+1}\mathcal{M}_{l+1,l+2} = \mathcal{M}_{l,l+1}I_{l+1,l+2}, \qquad l \in \mathbb{N}.$$

A λ-graph system is an inductive sequence of Bratteli diagrams, that come from the theory of operator algebras, with labeled edges by symbols. We will know that the symbolic matrix systems and the λ-graph systems are the same objects and give rise to subshifts. There is a canonical method to construct a symbolic matrix system from an arbitrary subshift (Theorem 3.7). The obtained symbolic matrix system is said to be canonical for the subshift. If a subshift is sofic, the canonical symbolic matrix system corresponds to the symbolic matrix of its left Krieger cover graph.

As a generalization of the notion of strong shift equivalence for nonnegative matrices and symbolic matrices, we will introduce the notion of strong shift equivalence for our symbolic matrix systems. We will prove

Theorem A (Theorem 4.2 and Theorem 4.15). *Two subshifts are topologically conjugate if and only if their canonical symbolic matrix systems are strong shift equivalent.*

Hence classification problem for subshifts are completely reduced to the classification of symbolic matrix systems up to strong shift equivalence in our sense. In the proof of the only if part of Theorem A, we provide the notion of bipartite λ-graph system. We then essentially use Nasu's factorization theorem for topological conjugacy between subshifts into bipartite codes and symbolic conjugacies.

We will next define shift equivalence between two symbolic matrix systems. That is a generalization of the corresponding notion for symbolic matrices defined by Boyle-Krieger in [BK]. We will see that strong shift equivalence implies shift equivalence even in our setting (Theorem 6.2). Similarly to the case of topological Markov shifts, we can prove that shift equivalence between two canonical symbolic matrix systems gives rise to an eventual conjugacy for the associated subshifts, that is, a topological conjugacy for their corresponding higher power shifts (Proposition 6.3). This result was motivated by a question raised by W. Krieger at a workshop at Kyushu University, Japan, March 1998.

For nonnegative matrices, there are two crucial shift equivalence invariants consisting of abelian groups. One is the dimension groups defined by W. Krieger in [Kr], [Kr2] and the other one is the Bowen-Franks groups in [BF]. They induce topological conjugacy invariants for the associated topological Markov shifts. We will generalize the two shift equivalence invariants to our symbolic matrix systems. For a symbolic matrix system $(\mathcal{M}, I)$, let $M_{l,l+1}$ be the nonnegative rectangular matrix obtained from $\mathcal{M}_{l,l+1}$ by setting all the symbols equal to 1 for each $l \in \mathbb{N}$. Then the resulting pair (M, I) still satisfies the following relations.

$$I_{l,l+1}M_{l+1,l+2} = M_{l,l+1}I_{l+1,l+2}, \qquad l \in \mathbb{N}.$$

We call (M, I) the nonnegative matrix system for $(\mathcal{M}, I)$. We say (M, I) to be canonical when $(\mathcal{M}, I)$ is canonical. More generally, for a sequence $M_{l,l+1}, l \in \mathbb{N}$ of rectangular matrices with entries in nonnegative integers and a sequence $I_{l,l+1}, l \in \mathbb{N}$ of rectangular matrices with entries in $\{0, 1\}$, the pair (M, I) is called a nonnegative matrix system if they satisfy the relations above. For a single $n \times n$ nonnegative square matrix A, if we set $M_{l,l+1} = A$ and $I_{l,l+1} = I_n$: the $n \times n$ identity matrix for all $l \in \mathbb{N}$, the pair (M, I) is a nonnegative matrix system. We will similarly formulate strong shift equivalence and shift equivalence between nonnegative matrix systems. These equivalences are generalizations of the corresponding equivalences for single nonnegative square matrices.

We will define the following three kinds of objects for a nonnegative matrix system (M, I).

(i) The dimension triple: $(\Delta_{(M,I)}, \Delta^+_{(M,I)}, \delta_{(M,I)})$.
(ii) The K-groups: $K_0(M, I)$, $K_1(M, I)$.
(iii) The Bowen-Franks groups: $BF^0(M, I)$, $BF^1(M, I)$.

The dimension triple $(\Delta_{(M,I)}, \Delta^+_{(M,I)}, \delta_{(M,I)})$ consist of an ordered abelian group $\Delta_{(M,I)}$ with positive cone $\Delta^+_{(M,I)}$ and an ordered automorphism $\delta_{(M,I)}$ on it. The K-groups $K_i(M,I), i = 0,1$ consist of a pair of abelian groups. The Bowen-Franks groups $BF^i(M,I), i = 0,1$ also consist of a pair of abelian groups. Let $m(l)$ be the row size of the matrix $I_{l,l+1}$ for each $l \in \mathbb{N}$. Let $\mathbb{Z}_{I^t}$ be the abelian group defined by the inductive limit $\mathbb{Z}_{I^t} = \varinjlim_l \{I^t_{l,l+1} : \mathbb{Z}^{m(l)} \to \mathbb{Z}^{m(l+1)}\}$. The sequence $M^t_{l,l+1}, l \in \mathbb{N}$ of the transposes of $M_{l,l+1}$ naturally yields an endomorphism on $\mathbb{Z}_{I^t}$ that is denoted by $\lambda_{(M,I)}$. The dimension group and the K-groups are defined as follows:

$$\Delta_{(M,I)} = \varinjlim \{\lambda_{(M,I)} : \mathbb{Z}_{I^t} \to \mathbb{Z}_{I^t}\}$$

and

$$K_0(M,I) = \mathbb{Z}_{I^t}/(id - \lambda_{(M,I)})\mathbb{Z}_{I^t}, \qquad K_1(M,I) = \mathrm{Ker}(id - \lambda_{(M,I)}) \text{ in } \mathbb{Z}_{I^t}.$$

The positive cone $\Delta^+_{(M,I)}$ of $\Delta_{(M,I)}$ is $\varinjlim\{\lambda_{(M,I)} : \mathbb{Z}^+_{I^t} \to \mathbb{Z}^+_{I^t}\}$ where $\mathbb{Z}^+_{I^t}$ is the natural positive cone of $\mathbb{Z}_{I^t}$ and the automorphism $\delta_{(M,I)}$ on $\Delta_{(M,I)}$ is induced one from $\lambda_{(M,I)}$. Set the projective limit of the abelian group as $\mathbb{Z}_I = \varprojlim_l \{I_{l,l+1} : \mathbb{Z}^{m(l+1)} \to \mathbb{Z}^{m(l)}\}$. The sequence $M_{l,l+1}, l \in \mathbb{N}$ acts on $\mathbb{Z}_I$ as an endomorphism that we denote by M. The identity on $\mathbb{Z}_I$ is denoted by I. The Bowen-Franks groups for (M,I) are defined by

$$BF^0(M,I) = \mathbb{Z}_I/(I-M)\mathbb{Z}_I, \qquad BF^1(M,I) = \mathrm{Ker}(I-M) \text{ in } \mathbb{Z}_I.$$

The above notions of dimension triple and Bowen-Franks group of degree zero for a nonnegative matrix system are generalizations of the corresponding notions for a single nonnegative square matrix. We will prove that the following Universal Coefficient Theorem holds (Theorem 9.6). It says that there exists a short exact sequence

$$0 \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}(K_0(M,I),\mathbb{Z}) \xrightarrow{\delta} BF^0(M,I) \xrightarrow{\gamma} \mathrm{Hom}_{\mathbb{Z}}(K_1(M,I),\mathbb{Z}) \longrightarrow 0$$

that splits unnaturally. We also see that

$$BF^1(M,I) \cong \mathrm{Hom}_{\mathbb{Z}}(K_0(M,I),\mathbb{Z}).$$

The three kinds of objects above are proved to be invariant under shift equivalence in nonnegative matrix systems. Hence they are naturally induce topological conjugacy invariants for subshifts by taking their canonical nonnegative matrix systems.

We will describe relationships among the equivalences and the invariants for the matrix systems as in the following way :

Theorem B. *For two symbolic matrix systems* $(\mathcal{M}, I), (\mathcal{M}', I')$ *and their nonnegative matrix systems* $(M, I), (M', I')$*, consider the following situations:*

(S1) $(\mathcal{M}, I) \approx (\mathcal{M}', I)$ *: strong shift equivalence,*
(N1) $(M, I) \approx (M', I)$ *: strong shift equivalence,*
(S2) $(\mathcal{M}, I) \sim (\mathcal{M}', I)$ *: shift equivalence,*
(N2) $(M, I) \sim (M', I)$ *: shift equivalence,*
(3) $(\Delta_{(M,I)}, \Delta^+_{(M,I)}, \delta_{(M,I)}) \cong (\Delta_{(M',I')}, \Delta^+_{(M',I')}, \delta_{(M',I')})$ *: isomorphic dimension triples,*
(4) $(\Delta_{(M,I)}, \delta_{(M,I)}) \cong (\Delta_{(M',I')}, \delta_{(M',I')})$ *: isomorphic dimension pairs,*
(5) $K_*(M, I) \cong K_*(M', I)$ *: isomorphic K-groups,*
(6) $BF^*(M, I) \cong BF^*(M', I)$ *: isomorphic Bowen-Franks groups.*

Then we have the following implications:

$$\begin{array}{ccccccccccc} (S1) & \Longrightarrow & (S2) & & & & & & & & \\ \Downarrow & & \Downarrow & & & & & & & & \\ (N1) & \Longrightarrow & (N2) & \Longrightarrow & (3) & \Longrightarrow & (4) & \Longrightarrow & (5) & \Longrightarrow & (6). \end{array}$$

It is well-known that the set of all nonzero eigenvalues of a nonnegative matrix A is also a shift equivalence invariant. The set for A is called the nonzero spectrum of A and plays an important rôle for studying dynamical properties of the associated topological Markov shift (cf.[LM], [Kit]). We introduce eigenvalues and eigenvectors of a nonnegative matrix system and then generalize the notion of the nonzero spectrum of a single nonnegative matrix to a nonnegative matrix system (M, I). We denote by $Sp^\times(M, I)$ the set of all nonzero eigenvalues of (M, I). A nonnegative matrix system (M, I) in general is an infinite sequence of pairs of matrices $M_{l,l+1}, I_{l,l+1}, l \in \mathbb{N}$ for which sizes of matrices are increasing. Hence it seems to be natural to deal with eigenvalues having a certain boundedness condition on the corresponding eigenvectors. We denote by $Sp^\times_b(M, I)$ the set of all nonzero eigenvalues of (M, I) with the boundedness condition on the corresponding eigenvectors. We will prove, in Section 10, that the both of the nonzero spectrums $Sp^\times(M, I)$ and $Sp^\times_b(M, I)$ are not empty and invariant under shift equivalence of (M, I). In particular, if (M, I) is the canonical nonnegative matrix system for a subshift, the set $Sp^\times_b(M, I)$ is bounded by the topological entropy of the subshift. We then define the nonzero spectrum and the nonzero bounded spectrum for subshifts by the corresponding sets for the canonical nonnegative matrix systems (Theorem 10.14).
In the final section, we present an example of the canonical symbolic matrix system for a certain nonsofic subshift, called the context free shift in [LM;Example 1.2.9]. Its K-groups and Bowen-Franks groups are calculated. We see that the types of the invariants can not appear in those of sofic shifts. The maximum of the absolute values of the bounded spectrums of the canonical nonnegative matrix system for the subshift is $1 + \sqrt{1 + \sqrt{3}}$. The value is the maximum in the bounded spectrum and coincides with the topological entropy of the subshift.

The author has recently constructed the C^*-algebra $\mathcal{O}_\Lambda$ associated with subshift Λ ([Ma]). The C^*-algebra $\mathcal{O}_\Lambda$ has a canonical action of the one dimensional torus group, called gauge action and written as α. The fixed point algebra $\mathcal{F}_\Lambda$ of $\mathcal{O}_\Lambda$ under α is an AF-algebra which is stably isomorphic to the crossed product $\mathcal{O}_\Lambda \times_\alpha \mathbb{T}$ ([Ma2]). Let (M, I) be the canonical nonnegative matrix system for the subshift Λ. The invariants mentioned above are described in terms of the K-theory for the C^*-algebras as in the following way:

$$
\begin{aligned}
(\Delta_{(M,I)}, \Delta^+_{(M,I)}, \delta_{(M,I)}) &= (K_0(\mathcal{F}_\Lambda), K_0(\mathcal{F}_\Lambda)_+, \hat{\alpha}_*),\\
K_i(M, I) &= K_i(\mathcal{O}_\Lambda), \qquad i = 0, 1,\\
BF^i(M, I) &= \mathrm{Ext}^{i+1}(\mathcal{O}_\Lambda), \qquad i = 0, 1
\end{aligned}
$$

where $\hat{\alpha}$ denotes the dual action of α and $\mathrm{Ext}^1(\mathcal{O}_\Lambda) = \mathrm{Ext}(\mathcal{O}_\Lambda), \mathrm{Ext}^0(\mathcal{O}_\Lambda) = \mathrm{Ext}(\mathcal{O}_\Lambda \otimes C_0(\mathbb{R}))$. The normalized nonnegative eigenvectors of (M, I) exactly correspond to the KMS-states for α on the C^*-algebra $\mathcal{O}_\Lambda$. Hence the set of all bounded spectrums with nonnegative eigenvectors are the set of all inverse temperatures for the admitted KMS states.

Acknowledgements: The author would like to thank Wolfgang Krieger for his valuable advices and suggestions, in particular for his question on Proposition 6.3. The author also would like to thank Toshihiro Hamachi and Yasuo Watatani for pointing out an error and an inaccuracy in an earlier version of this paper. Finally, the author wishes to express his gratitude to Joachim Cuntz and referees for advices and suggestions in the presentation of this paper.

2. SYMBOLIC MATRIX SYSTEMS AND λ-GRAPH SYSTEMS

We fix a finite set Σ and call it the alphabet. Each element of Σ is called a symbol. We always write the empty symbol $\emptyset$ in Σ as 0. We denote by $\mathfrak{S}_\Sigma$ the set of all finite formal sums of elements of Σ. A square matrix with entries in $\mathfrak{S}_\Sigma$ is called a symbolic matrix over Σ.

DEFINITION. Let $(\mathcal{M}_{l,l+1}, I_{l,l+1}), l \in \mathbb{N}$ be a pair of sequences of rectangular matrices such that the following four conditions for each $l \in \mathbb{N}$ are satisfied:

(1) $\mathcal{M}_{l,l+1}$ is an $m(l) \times m(l+1)$ rectangular matrix with entries in $\mathfrak{S}_\Sigma$.
(2) $I_{l,l+1}$ is an $m(l) \times m(l+1)$ rectangular matrix with entries in $\{0, 1\}$ satisfying the following two conditions:
(2-a) For i, there exists j such that $I_{l,l+1}(i, j) \neq 0$.
(2-b) For j, there uniquely exists i such that $I_{l,l+1}(i, j) \neq 0$.
(3) $m(l) \leq m(l+1)$.
(4) $I_{l,l+1}\mathcal{M}_{l+1,l+2} = \mathcal{M}_{l,l+1}I_{l+1,l+2}$.

The pair $(\mathcal{M}, I)$ is called a *symbolic matrix system* over Σ. For $i = 1, \dots, m(l), j = 1, \dots, m(l+1)$, we denote by $\mathcal{M}_{l,l+1}(i, j)$, $I_{l,l+1}(i, j)$ the (i, j)-components of $\mathcal{M}_{l,l+1}$, $I_{l,l+1}$ respectively. A symbolic matrix system $(\mathcal{M}, I)$ is said to be *essential* if it satisfies the following further conditions:

(5-i) For i, there exists j such that $\mathcal{M}_{l,l+1}(i, j) \neq 0$.
(5-ii) For j, there exists i such that $\mathcal{M}_{l,l+1}(i, j) \neq 0$.

We henceforth study essential symbolic matrix systems and call them symbolic matrix systems for simplicity.
The following notion of specified equivalence between symbolic matrices has been introduced by M. Nasu in [N1], [N2].
For two symbolic matrices $\mathcal{A}$ over alphabet Σ and $\mathcal{A}'$ over alphabet Σ' and bijection ϕ from a subset of Σ onto a subset of Σ', we call $\mathcal{A}$ and $\mathcal{A}'$ are specified equivalence under specification ϕ if $\mathcal{A}'$ can be obtained from $\mathcal{A}$ by replacing every symbol a appearing in $\mathcal{A}$ by $\phi(a)$. We write it as $\mathcal{A} \overset{\phi}{\simeq} \mathcal{A}'$. We call ϕ a specification from Σ to Σ'.
Two symbolic matrix systems $(\mathcal{M}, I)$ over Σ and $(\mathcal{M}', I')$ over Σ' are said to be isomorphic if there exists a specification ϕ from Σ to Σ' and an $m(l) \times m(l)$-square permutation matrix P_l for each $l \in \mathbb{N}$ such that

$$P_l \mathcal{M}_{l,l+1} \overset{\phi}{\simeq} \mathcal{M}'_{l,l+1} P_{l+1}, \qquad P_l I_{l,l+1} = I'_{l,l+1} P_{l+1} \qquad \text{for} \quad l \in \mathbb{N}.$$

The notion of symbolic matrix system is a generalized notion of symbolic matrix. We say a symbolic matrix system $(\mathcal{M}, I)$ to be *sofic* if there exists a number $L \in \mathbb{N}$ such that

$$\mathcal{M}_{l,l+1} = \mathcal{M}_{L,L+1}, \qquad I_{l,l+1} = I_{L,L+1}$$

for all $l \geq L$. Hence in this case, we see $m(L) = m(l)$ for all $l \geq L$.
A symbolic matrix corresponds to a labeled graph, called a λ-graph, that is a presentation of a sofic subshift. We will next consider a generalization of λ-graphs corresponding to symbolic matrix systems.
We first explain the notion of Bratteli diagram that appears in the theory of operator algebras (see [Bra], [Ef], [El]). A Bratteli diagram consists of a vertex set V and an edge set E satisfying the following conditions. We have a decomposition of V as a disjoint union $V = V_1 \cup V_2 \cup \cdots$ where each V_l is finite and nonempty. Similarly E decomposes as a disjoint union $E = E_{1,2} \cup E_{2,3} \cup \cdots$ where each $E_{l,l+1}$ is finite and nonempty. Moreover we have maps $s, r : E \to V$ such that $s(E_{l,l+1}) \subset V_l$, $r(E_{l,l+1}) \subset V_{l+1}$. They are called a source map and a range map respectively. A Bratteli diagram (V, E) is said to be essential if it satisfies the condition that $s^{-1}(v)$ is nonempty for all $v \in V$ and $r^{-1}(v)$ is nonempty for all $v \in V \diagdown V_1$. For $u \in V_l, v \in V_{l+1}$, put

$$E_{l,l+1}(u, v) = \{e \in E_{l,l+1} | s(e) = u, r(e) = v\}.$$

We next introduce the notion of *labeled Bratteli diagram*. A labeled Bratteli diagram over alphabet Σ consists of a Bratteli diagram (V, E) and a map λ from E to Σ.
Definition. A λ*-graph system* over alphabet Σ consists of a labeled Bratteli diagram (V, E, λ) over Σ and a surjective map ι from $V \diagdown V_1$ to V satisfying the following two conditions:

(1) $\iota(V_{l+1}) = V_l$ for $l \in \mathbb{N}$.

(2) For $u \in V_l, w \in V_{l+2}$, there exists a bijective correspondence between the edge sets

$$E_{l,l+1}(u, \iota(w)) \quad \text{and} \quad \bigcup_{v \in V_{l+1}, \iota(v)=u} E_{l+1,l+2}(v, w)$$

that is compatible with the labeling λ.

We denote by (V, E, λ, ι) the λ-graph system.

The following two conditions are implied from the above condition (2).

(2-i) For $e \in E_{l+1,l+2}$, there exists $e' \in E_{l,l+1}$ such that

$$\iota(s(e)) = s(e'), \quad \iota(r(e)) = r(e') \quad \text{and} \quad \lambda(e) = \lambda(e').$$

(2-ii) For $f \in E_{l,l+1}$, $v \in V_{l+2}$ with $\iota(v) = r(f)$, there exists $e \in E_{l+1,l+2}$ such that

$$\iota(s(e)) = s(f), \quad r(e) = v \quad \text{and} \quad \lambda(e) = \lambda(f).$$

A λ-graph system (V, E, λ, ι) is said to be *essential* if the Bratteli diagram (V, E) is essential. We always treat an essential λ-graph system and call it a λ-graph system for simplicity. We remark that by the condition (1) in Definition of λ-graph system the cardinality of the set V_{l+1} is greater than or equal to that of the set V_l.

Two λ-graph systems (V, E, λ, ι) over alphabet Σ and $(V', E', \lambda', \iota')$ over alphabet Σ' are said to be isomorphic if there exist bijections $\Phi_V : V \to V'$, $\Phi_E : E \to E'$ and a specification $\phi : \Sigma \to \Sigma'$ such that

(1) $\Phi_V(V_l) = V'_l$ and $\Phi_E(E_{l,l+1}) = E'_{l,l+1}$ for $l \in \mathbb{N}$,
(2) $\Phi_V(s(e)) = s(\Phi_E(e))$ and $\Phi_V(r(e)) = r(\Phi_E(e))$ for $e \in E$,
(3) $\iota'(\Phi_V(v)) = \Phi_V(\iota(v))$ for $v \in V$,
(4) $\lambda'(\Phi_E(e)) = \phi(\lambda(e))$ for $e \in E$.

Proposition 2.1. *There exists a bijective correspondence between the set of all isomorphism classes of symbolic matrix systems and the set of all isomorphism classes of λ-graph systems.*

Proof. 1. From symbolic matrix systems to λ-graph systems: Let $(\mathcal{M}, I)$ be a symbolic matrix system over Σ. We are always assuming that it is essential. For each $l \in \mathbb{N}$, let $V_l = \{1, \dots, m(l)\}$ be the set of all rows of the matrix $\mathcal{M}_{l,l+1}$ and $E_{l,l+1}$ the disjoint union of elements appearing in the components of $\mathcal{M}_{l,l+1}$. For each $e \in E_{l,l+1}$ we put $s(e) = i$ and $r(e) = j$ if e appears in $\mathcal{M}_{l,l+1}(i, j)$. The map $\iota : V \setminus V_1 \to V$ is defined as $\iota(j) = i$ for $j \in V_{l+1}$ if $I_{l,l+1}(i, j) = 1$. The map $\lambda : E \to \Sigma$ is defined by $\lambda(e) = e$. Then it is straightforward to see that (V, E, λ, ι) is a λ-graph system.

2. From λ-graph systems to symbolic matrix systems : Let (V, E, ι, λ) be a λ-graph system over Σ. We denote by $m(l)$ the cardinality of the vertex set V_l. We identify V_l with the set $\{1, \dots, m(l)\}$. We define $m(l) \times m(l+1)$ matrices as follows: For $i \in V_l, j \in V_{l+1}$, set $I_{l,l+1}(i, j) = 1$ if $\iota(j) = i$ otherwise $I_{l,l+1}(i, j) = 0$. For $e_k \in E_{l,l+1}, k = 1, \dots, n$ with $s(e_k) = i, r(e_k) = j$, we put $\mathcal{M}_{l,l+1}(i, j) = \lambda(e_1) + \cdots + \lambda(e_n)$. It is straightforward to see that the relations $I_{l,l+1}\mathcal{M}_{l+1,l+2} = \mathcal{M}_{l,l+1}I_{l+1,l+2}$ for $l \in \mathbb{N}$ hold.

3. Presentations of subshifts

As in the preceding section, symbolic matrix systems may be identified with λ-graph systems. We will in this section construct subshifts, a class of topological dynamical systems, from λ-graph systems. We will further show that any subshift comes from a λ-graph system. This is a generalized observation of the correspondences between the sofic subshits and the symbolic matrices. Hence studies of subshifts are completely reduced to the studies of λ-graph systems and hence symbolic matrix systems.

We will review on subshifts. Let Σ be an alphabet. Let $\Sigma^{\mathbb{Z}}$, $\Sigma^{\mathbb{N}}$ be the infinite product spaces $\prod_{i=-\infty}^{\infty}\Sigma_i$, $\prod_{i=1}^{\infty}\Sigma_i$ where $\Sigma_i = \Sigma$, endowed with the product topology respectively. The transformation σ on $\Sigma^{\mathbb{Z}}, \Sigma^{\mathbb{N}}$ given by $(\sigma(x_i)) = (x_{i+1}), i \in \mathbb{Z}, \mathbb{N}$ is called the (full) shift. Let Λ be a shift invariant closed subset of $\Sigma^{\mathbb{Z}}$ i.e. $\sigma(\Lambda) = \Lambda$. The topological dynamical system $(\Lambda, \sigma|_\Lambda)$ is called a subshift. We denote $\sigma|_\Lambda$ by σ and write the subshift as Λ for short. We denote by $X_\Lambda(\subset \prod_{i=1}^{\infty}\Sigma_i)$ the set of all right-infinite sequences that appear in Λ. The dynamical system (X_Λ, σ) is called the right one-sided subshift for Λ. We will give examples of subshifts as follows (cf.[LM], [Kit]):
Let A be an $n \times n$ matrix with entries in nonnegative integers. Put $V_A = \{1, \dots, n\}$: the vertex set. Write $A(i,j)$ edges from $i \in V_A$ to $j \in V_A$. Hence we have a directed graph from A. We denote it by G_A. Let E_A be the set of all edges of the graph G_A. Let s_A, r_A be the map from E_A to V_A that assigns the source and the range of the edge. Let Λ_A be the set of all biinfinite sequences $(e_i)_{i\in\mathbb{Z}}$ of $e_i \in E_A$ with $r_A(e_i) = s_A(e_{i+1}), i \in \mathbb{Z}$. Then Λ_A becomes a subshift, called the topological Markov shift defined by A.
Let $\mathcal{A}$ be an $n \times n$ symbolic matrix over Σ. Each entry $\mathcal{A}(i,j), i,j = 1,\dots,n$ consists of elements of $\mathfrak{S}_\Sigma$. Similarly to the construction above, we have a directed graph $G_\mathcal{A}$ from the matrix $\mathcal{A}$ with labeled edges by the symbols in Σ. We denote by $\lambda(e) = \alpha \in \Sigma$ the label α of edge e. Let $\Lambda_\mathcal{A}$ be the set of all biinfinite sequences $\lambda(e_i)_{i\in\mathbb{Z}}$ of labels of the sequence of edges $e_i \in E_\mathcal{A}$ with $r_\mathcal{A}(e_i) = s_\mathcal{A}(e_{i+1}), i \in \mathbb{Z}$. Then $\Lambda_\mathcal{A}$ becomes a subshift, called the sofic subshift defined by $\mathcal{A}$. The labeled graph $G_\mathcal{A}$ is called a λ-graph for $\mathcal{A}$.
There are many nonsofic subshifts as seen in [LM]. We will see an example of nonsofic subshift in the final section.
A finite sequence $\mu = (\mu_1, ..., \mu_k)$ of elements $\mu_j \in \Sigma$ is called a block or a word. We denote by $|\mu|$ the length k of μ. A block $\mu = (\mu_1, ..., \mu_k)$ is said to occur or appear in $x = (x_i) \in \Sigma^{\mathbb{Z}}$ if $x_m = \mu_1, ..., x_{m+k-1} = \mu_k$ for some $m \in \mathbb{Z}$.

We will first construct subshifts from symbolic matrix systems.
Let $(\mathcal{M}, I)$ be a symbolic matrix system over Σ and (V, E, λ, ι) its corresponding λ-graph system. For $k < l$, let $P_{k,l}$ be the set of all paths from V_k to V_l, that is,

$$P_{k,l} = \{(e_k, e_{k+1}, \dots, e_{l-1}) | e_i \in E_{i,i+1}, r(e_i) = s(e_{i+1}) \text{ for } i = k, k+1, \dots, l-2\}.$$

We define the maps $s : P_{k,l} \to V_k$ and $r : P_{k,l} \to V_l$ by

$$s(e_k, e_{k+1}, \dots, e_{l-1}) = s(e_k), \quad r(e_k, e_{k+1}, \dots, e_{l-1}) = r(e_{l-1}).$$

The labeling $\lambda : P_{k,l} \to \Sigma^{l-k} = \underbrace{\Sigma \times \cdots \times \Sigma}_{l-k \text{ times}}$ is defined by

$$\lambda(e_k, e_{k+1}, \dots, e_{l-1}) = \lambda(e_k)\lambda(e_{k+1}) \cdots \lambda(e_{l-1}).$$

Set

$$L_{k,l} = \{\lambda(w) \in \Sigma^{l-k} | w \in P_{k,l}\}.$$

Put $L_l = L_{1,l+1}$ and endow it with discrete topology. The map $\pi_l : L_{l+1} \to L_l$ is defined by

$$\pi_l(\alpha_1, \dots, \alpha_{l+1}) = (\alpha_1, \dots, \alpha_l).$$

We set

$$X_{(\mathcal{M},I)} = \varprojlim \{\pi_l : L_{l+1} \to L_l\}$$

the projective limit in the category of compact Hausdorff spaces. That is

$$X_{(\mathcal{M},I)} = \{(\lambda(e_1), \lambda(e_2), \dots) \in \Sigma^{\mathbb{N}} | e_i \in E_{i,i+1}, r(e_i) = s(e_{i+1}) \text{ for } i \in \mathbb{N}\}$$

the set of all right infinite sequences consisting of labels along infinite paths. The topology on $X_{(\mathcal{M},I)}$ is defined from open sets of the form

$$U_{(\mu_1,\dots,\mu_k)} = \{(\alpha_1, \alpha_2, \dots) \in X_{(\mathcal{M},I)} | \alpha_i = \mu_i \text{ for } i = 1, \dots, k\}$$

for $(\mu_1, \dots, \mu_k) \in L_k$.

Lemma 3.1. *If $(\alpha_1, \alpha_2, \dots) \in X_{(\mathcal{M},I)}$, we have $(\alpha_2, \alpha_3, \dots) \in X_{(\mathcal{M},I)}$.*

Proof. The assertion is direct from the condition (2-i) of Definition of λ-graph system.

Lemma 3.2. *For $l > k$, if $(\alpha_k, \dots, \alpha_{l-1}) \in L_{k,l}$, we have $(\alpha_k, \dots, \alpha_{l-1}) \in L_{k+1,l+1}$.*

Proof. For $(\alpha_k, \dots, \alpha_{l-1}) \in L_{k,l}$, take $f_i \in E_{i,i+1}$ such as $\alpha_i = \lambda(f_i)$ for $i = k, k+1, \dots, l-1$ and $r(f_i) = s(f_{i+1})$ for $i = k, k+1, \dots, l-2$. We find $v_{l+1} \in V_{l+1}$ with $\iota(v_{l+1}) = r(f_{l-1})$. By the condition (2-ii) of Definition of λ-graph system, there exists $e_l \in E_{l,l+1}$ such that $\iota(s(e_l)) = s(f_{l-1})$, $r(e_l) = v_{l+1}$ and $\lambda(e_l) = \lambda(f_{l-1})$. Put $v_l = s(e_l) \in V_l$. We continue theses procedures so that we get $e_i \in E_{i,i+1}$ for $i = k+1, k+2, \dots, l$ satisfying $\iota(s(e_i)) = s(f_{i-1})$, $r(e_i) = s(e_{i+1})$ and $\lambda(e_i) = \lambda(f_{i-1})$ for $i = k+1, k+2, \dots, l$. Hence $\alpha_i = \lambda(e_{i+1})$ and $(\alpha_k, \dots, \alpha_{l-1}) \in L_{k+1,l+1}$.

As in [LM; Definition 1.3.1], a set $\mathfrak{L}$ of words of alphabet Σ is called a language if it satisfies the following conditions:

(a) Every subword of a word w in $\mathfrak{L}$ belongs to $\mathfrak{L}$.
(b) For a word w in $\mathfrak{L}$, there are nonempty words u, v in $\mathfrak{L}$ such that uwv belongs to $\mathfrak{L}$.

Let $\mathfrak{L}(\mathcal{M}, I)$ be the set of all words appearing in $X_{(\mathcal{M},I)}$. That is

$$\mathfrak{L}(\mathcal{M}, I) = \cup_{k \le l} L_{k,l}.$$

Then we have

Proposition 3.3. *$\mathfrak{L}(\mathcal{M}, I)$ is a language.*

Proof. $\mathfrak{L}(\mathcal{M}, I)$ clearly satisfies the condition (a) above. For a word $w \in L_{k,l}$, we know $w \in L_{k+1,l+1}$ by Lemma 3.2. We write $w = (\lambda(e_{k+1}), \lambda(e_{k+2}), \dots, \lambda(e_l))$ for $e_i \in E_{i,i+1}$ with $r(e_i) = s(e_{i+1})$, $i = k+1, \dots, l-1$. Since both the sets $r^{-1}(s(e_{k+1}))$ and $s^{-1}(s(e_l))$ are not empty, we may find words $u, v \in \mathfrak{L}(\mathcal{M}, I)$ such that $uwv \in \mathfrak{L}(\mathcal{M}, I)$. Thus $\mathfrak{L}(\mathcal{M}, I)$ satisfies the condition (b).

By [LM;Proposition 1.3.4], we see

Theorem 3.4. *There exists a subshift Λ over alphabet Σ whose language is $\mathfrak{L}(\mathcal{M}, I)$. Namely the set of all admissible words of the subshift Λ is $\mathfrak{L}(\mathcal{M}, I)$.*

We denote by $\Lambda_{(\mathcal{M},I)}$ the subshift Λ in the theorem above and call it *the subshift associated with symbolic matrix system* $(\mathcal{M}, I)$.

It is also possible to construct the subshift $\Lambda_{(\mathcal{M},I)}$ by using projective limit method as in the folloing way.

Lemma 3.5. *For $(\alpha_1, \alpha_2, \dots) \in X_{(\mathcal{M},I)}$, there exists a symbol $\alpha_0 \in \Sigma$ such that $(\alpha_0, \alpha_1, \alpha_2, \dots) \in X_{(\mathcal{M},I)}$.*

Proof. Put $w_k = (\alpha_1, \alpha_2, \dots, \alpha_{k-1}) \in L_{1,k}$. By Lemma 3.2 and Proposition 3.3, there exists a symbol $\beta_k \in \Sigma$ such that $\beta_k w_k \in L_{1,k+1}$. Hence we may find $y_k \in X_{(\mathcal{M},I)}$ such that $\beta_k w_k y_k \in X_{(\mathcal{M},I)}$. As the alphabet Σ is a finite set, there exists a symbol $\alpha_0 \in \Sigma$ and a subsequence of $(\beta_k)_{k\in\mathbb{N}}$ such that $\beta_{k_n} = \alpha_0$ for $n = 1, 2, \dots$ and $k_1 < k_2 < \cdots$. Put $x_{k_n} = \alpha_0 w_{k_n} y_{k_n}, n \in \mathbb{N}$. They converge to an element

$$x = (\alpha_0, \alpha_1, \alpha_2, \dots) \in X_{(\mathcal{M},I)}.$$

By Lemma 3.1, the following map

$$S : (\alpha_1, \alpha_2, \alpha_3, \dots) \in X_{(\mathcal{M},I)} \to (\alpha_2, \alpha_3, \dots) \in X_{(\mathcal{M},I)}$$

is well-defined, continuous and surjective. We set

$$\Lambda = \varprojlim \{S : X_{(\mathcal{M},I)} \to X_{(\mathcal{M},I)}\}$$

the projective limit in the category of compact Hausdorff spaces. Thus Λ is identified with the set of all biinfinite sequences arising from the sequences in $X_{(\mathcal{M},I)}$. That is

$$\Lambda = \{(\dots, \alpha_2, \alpha_1, \alpha_0, \alpha_1, \alpha_2, \dots) | (\alpha_n, \alpha_{n+1}, \dots) \in X_{(\mathcal{M},I)} \text{ for all } n \in \mathbb{Z}\}.$$

The map S induces a homeomorphism on it. We denote it by σ that satisfies $\sigma((\alpha_i)_{i\in\mathbb{Z}}) = (\alpha_{i+1})_{i\in\mathbb{Z}}$. Therefore we have a subshift (Λ, σ) from symbolic matrix system $(\mathcal{M}, I)$. It is nothing but the subshift $(\Lambda_{(\mathcal{M},I)}, \sigma)$ defined in the preceding discussion.

We will next construct symbolic matrix systems from subshifts.
For a subshift (Λ, σ) over Σ and a number $k \in \mathbb{N}$, let Λ^k be the set of all words of length k in $\Sigma^{\mathbb{Z}}$ occurring in some $x \in \Lambda$. Put $\Lambda^* = \cup_{k=0}^{\infty} \Lambda^k$ where Λ^0 denotes the empty word $\emptyset$. Set

$$\Lambda^l(x) = \{\mu \in \Lambda^l | \mu x \in X_\Lambda\} \qquad \text{for} \quad x \in X_\Lambda, \quad l \in \mathbb{N}.$$

We define a nested sequence of equivalence relations in the space X_Λ. Two points $x, y \in X_\Lambda$ are said to be *l-past equivalent* if $\Lambda^l(x) = \Lambda^l(y)$. We write this equivalence as $x \sim_l y$. We denote by $\Omega_l = X_\Lambda / \sim_l$ the quotient space by l-past equivalence classes of X_Λ ([Ma3]).

Lemma 3.6. *For $x, y \in X_\Lambda$ and $\mu \in \Lambda^k$,*

(i) *if $x \sim_l y$, we have $x \sim_m y$ for $m < l$.*
(ii) *if $x \sim_l y$ and $\mu x \in X_\Lambda$, we have $\mu y \in X_\Lambda$ and $\mu x \sim_{l-k} \mu y$ for $l > k$.*

Hence we have the following sequence of surjections in a natural way:

$$\Omega_1 \leftarrow \Omega_2 \leftarrow \cdots \leftarrow \Omega_l \leftarrow \Omega_{l+1} \leftarrow \cdots .$$

We easily see that (Λ, σ) is a sofic subshift if and only if $\Omega_l = \Omega_{l+1}$ for some $l \in \mathbb{N}$.
For a fixed $l \in \mathbb{N}$, let $F_i^l, i = 1, 2, \ldots, m(l)$ be the set of all l-past equivalence classes of X_Λ. Hence X_Λ is a disjoint union of the subsets $F_i^l, i = 1, 2, \ldots, m(l)$. We define two rectangular $m(l) \times m(l+1)$ matrices $I_{l,l+1}^\Lambda, \mathcal{M}_{l,l+1}^\Lambda$ with entries in $\{0, 1\}$ and entries in $\mathfrak{S}_\Sigma$ respectively as in the following way. For $i = 1, 2, \ldots, m(l), j = 1, 2, \ldots, m(l+1)$, the (i, j)-component $I_{l,l+1}^\Lambda(i, j)$ of $I_{l,l+1}^\Lambda$ is one if F_i^l contains F_j^{l+1} otherwise zero. Let $a_1, \ldots, a_n$ be the set of all symbols in Σ for which $a_k x \in F_i^l$ for some $x \in F_j^{l+1}$. We then define the (i, j)-component of the matrix $\mathcal{M}_{l,l+1}^\Lambda(i, j)$ as $\mathcal{M}_{l,l+1}^\Lambda(i, j) = a_1 + \cdots + a_n$: the formal sum of $a_1, \ldots, a_n$. We call $I_{l,l+1}^\Lambda$ the *inclusion matrices for* Λ and $\mathcal{M}_{l,l+1}^\Lambda$ the *symbolic representation matrices for* Λ respectively.
We next construct a labeled graph from subshift Λ for each $l \in \mathbb{N}$. The vertices of the graph consist of the sets $F_i^l, i = 1, \ldots, m(l)$ and $F_j^{l+1}, j = 1, \ldots, m(l+1)$ which we denote by V_l and V_{l+1} respectively. We write an arrow with label a, denoted by $\lambda^\Lambda(a)$, from the vertex F_i^l to F_j^{l+1} if $ax \in F_i^l$ for some $x \in F_j^{l+1}$. We denote by $E_{l,l+1}$ the set of all arrows from V_l to V_{l+1}. Since for each $j = 1, \ldots, m(l+1)$ there uniquely exists $i = 1, \ldots, m(l)$ such that $I_{l,l+1}(i, j) = 1$, we have a natural map ι_l^Λ from V_{l+1} to V_l. Set $V^\Lambda = \cup_{l=1}^{\infty} V_l$ and $E^\Lambda = \cup_{l=1}^{\infty} E_{l,l+1}$. We then see

Theorem 3.7. *For a subshift (Λ, σ), the pair $(\mathcal{M}^\Lambda, I^\Lambda)$ is a symbolic matrix system for which its λ-graph is $(V^\Lambda, E^\Lambda, \lambda^\Lambda, \iota^\Lambda)$. Moreover the subshift $\Lambda_{(\mathcal{M}^\Lambda, I^\Lambda)}$ associated with $(\mathcal{M}^\Lambda, I^\Lambda)$ coincides with the original subshift Λ.*

Proof. For each $l \in \mathbb{N}$, it is straightforward to check that the relation

$$I^\Lambda_{l,l+1}\mathcal{M}^\Lambda_{l+1,l+2} = \mathcal{M}^\Lambda_{l,l+1}I^\Lambda_{l+1,l+2}$$

holds. It then follows that the pair $(\mathcal{M}^\Lambda, I^\Lambda)$ is a symbolic matrix system whose associated λ-graph system is $(V^\Lambda, E^\Lambda, \lambda^\Lambda, \iota^\Lambda)$. It is also easy to see that the subshift associated with $(\mathcal{M}^\Lambda, I^\Lambda)$ coincides with the original subshift Λ because their forbidden words coincide.

Therefore we have a symbolic matrix system $(\mathcal{M}^\Lambda, I^\Lambda)$ and a λ-graph system $(V^\Lambda, E^\Lambda, \lambda^\Lambda, \iota^\Lambda)$ from subshift (Λ, σ). We call them the *canonical* symbolic matrix system for Λ and the *canonical* λ-graph system for Λ respectively.
It is now clear that sofic symbolic matrix systems exactly correspond to sofic subshifts.
For a symbolic matrix system $(\mathcal{M}, I)$, let $\Lambda_{(\mathcal{M},I)}$ be the associated subshift constructed from $(\mathcal{M}, I)$. Then its canonical symbolic matrix system $(\mathcal{M}^\Lambda, I^\Lambda)$ does not necessarily coincide with the original symbolic matrix system $(\mathcal{M}, I)$. We indeed see the following proposition. Its proof is direct.

Proposition 3.8. *For a subshift Λ, we have*

(i) *the representation matrices $\mathcal{M}^\Lambda_{l,l+1}$ are left resolving, i.e. the incoming edges to each vertex carry different labels.*
(ii) *the labeled Bratteli diagram $(V^\Lambda, E^\Lambda, \lambda^\Lambda)$ is predecessor-separated, i.e. distinct vertices at each level have distinct predecessor sets of labels.*

For example set, for each $l \in \mathbb{N}$, $\mathcal{M}_{l,l+1} = \begin{bmatrix} a & b \\ b & 0 \end{bmatrix}$ and $I_{l,l+1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The symbolic matrix system gives rise to the even shift that is denoted by Y. Its canonical symbolic matrix system is given by the following matrices:

$$\mathcal{M}^Y_{1,2} = \begin{bmatrix} a & a+b & b \\ b & 0 & 0 \end{bmatrix}, \qquad I^Y_{1,2} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathcal{M}^Y_{l,l+1} = \begin{bmatrix} a & a & b \\ 0 & b & 0 \\ b & 0 & 0 \end{bmatrix}, \qquad I^Y_{l,l+1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{for} \quad l \geq 2.$$

We indeed have

Proposition 3.9. *If Λ is a sofic subshift, its canonical λ-graph system is eventually realized as the left Krieger cover graph for Λ. Hence the canonical symbolic matrix system for Λ is eventually realized as the symbolic representation matrix for the left Krieger cover graph.*

4. Strong shift equivalence.

In this section, we will define two kinds of strong shift equivalences between two symbolic matrix systems. One is called the properly strong shift equivalence that exactly reflects a bipartite decomposition of the associated λ-graph systems. The other one is called the strong shift equivalence that is weaker than the former strong shift equivalence. They coincide at least between canonical symbolic matrix systems and between sofic symbolic matrix systems. The latter is easier defined and treated than the former. We will see, in the next section, that the latter strong shift equivalence directly leads to the shift equivalence between symbolic matrix systems. The main result in this section is that topological conjugacy between two subshifts are completely characterized by strong shift equivalence between their canonical symbolic matrix systems. We first define properly strong shift equivalence in 1-step between two symbolic matrix systems as a generalization of strong shift equivalence in 1-step between two nonnegative matrices defined by R. Williams in [Wi] and between two symbolic matrices defined by M. Nasu in [N](see also [BK]).

For alphabets C, D, put $C \cdot D = \{cd | c \in C, d \in D\}$. For $x = \sum_j c_j \in \mathfrak{S}_C$ and $y = \sum_k d_k \in \mathfrak{S}_D$, define $xy = \sum_{j,k} c_j d_k \in \mathfrak{S}_{C\cdot D}$.

Let $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ be symbolic matrix systems over alphabets Σ, Σ' respectively, where $\mathcal{M}_{l,l+1}, I_{l,l+1}$ are $m(l) \times m(l+1)$ matrices and $\mathcal{M}'_{l,l+1}, I'_{l,l+1}$ are $m'(l) \times m'(l+1)$ matrices.

Definition. Two symbolic matrix systems $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ are said to be *properly strong shift equivalent in 1-step* if there exist alphabets C, D and specifications

$$\varphi : \Sigma \to C \cdot D, \qquad \phi : \Sigma' \to D \cdot C$$

and increasing sequences $n(l), n'(l)$ on $l \in \mathbb{N}$ such that for each $l \in \mathbb{N}$, there exist an $n(l) \times n'(l+1)$ matrix $\mathcal{P}_l$ over C, an $n'(l) \times n(l+1)$ matrix $\mathcal{Q}_l$ over D, an $n(l) \times n(l+1)$ matrix X_l over $\{0,1\}$ and an $n'(l) \times n'(l+1)$ matrix X'_l over $\{0,1\}$ satisfying the following equations:

$$\mathcal{M}_{l,l+1} \overset{\varphi}{\simeq} \mathcal{P}_{2l}\mathcal{Q}_{2l+1}, \qquad \mathcal{M'}_{l,l+1} \overset{\phi}{\simeq} \mathcal{Q}_{2l}\mathcal{P}_{2l+1}, \tag{4.1}$$

$$I_{l,l+1} = X_{2l}X_{2l+1}, \qquad I'_{l,l+1} = X'_{2l}X'_{2l+1} \tag{4.2}$$

and

$$X_l\mathcal{P}_{l+1} = \mathcal{P}_l X'_{l+1}, \qquad X'_l\mathcal{Q}_{l+1} = \mathcal{Q}_l X_{l+1}. \tag{4.3}$$

We write this situation as

$$(\mathcal{M}, I) \underset{1-pr}{\approx} (\mathcal{M}', I').$$

It follows by (4.1) that $n(2l) = m(l)$ and $n'(2l) = m(l)$ for $l \in \mathbb{N}$.

Two symbolic matrix systems $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ are said to be *properly strong shift equivalent in N-step* if there exists a sequence of symbolic matrix systems $(\mathcal{M}^{(i)}, I^{(i)}), i = 1, 2, \dots, N-1$ such that

$$
\begin{aligned}
(\mathcal{M}, I) \underset{1-pr}{\approx} (\mathcal{M}^{(1)}, I^{(1)}) \underset{1-pr}{\approx} (\mathcal{M}^{(2)}, I^{(2)}) \\
\underset{1-pr}{\approx} \cdots \underset{1-pr}{\approx} (\mathcal{M}^{(N-1)}, I^{(N-1)}) \underset{1-pr}{\approx} (\mathcal{M}', I')
\end{aligned}
$$

We denote this situation by

$$(\mathcal{M}, I) \underset{N-pr}{\approx} (\mathcal{M}', I')$$

and simply call it a *properly strong shift equivalence.*

Proposition 4.1. *Properly strong shift equivalence is an equivalence relation on symbolic matrix systems.*

Proof. It is clear that properly strong shift equivalence is symmetric and transitive. It suffices to show that $(\mathcal{M}, I) \underset{1-pr}{\approx} (\mathcal{M}, I)$. Put $C = \Sigma, D = \{0, 1\}$. Define $\varphi : a \in \Sigma \to a \cdot 1 \in C \cdot D$ and $\phi : a \in \Sigma \to 1 \cdot a \in D \cdot C$. Let E_k be the $k \times k$ identity matrix. Set

$$
\begin{aligned}
\mathcal{P}_{2l} = \mathcal{P}_{2l+1} = \mathcal{M}_{l,l+1}, \quad \mathcal{Q}_{2l} = E_{m(l)}, \quad \mathcal{Q}_{2l+1} = E_{m(l+1)}, \\
X_{2l} = E_{m(l)}, \quad X_{2l+1} = I_{l,l+1}, \quad X'_{2l} = I_{l,l+1}, \quad X'_{2l+1} = E_{m(l+1)}.
\end{aligned}
$$

It is straightforward to see that they give a properly strong shift equivalence in 1-step between $(\mathcal{M}, I)$ and $(\mathcal{M}, I)$.

We will prove the following theorem.

Theorem 4.2. *Two subshifts Λ and Λ' are topologically conjugate if and only if their canonical symbolic matrix systems $(\mathcal{M}^{\Lambda}, I^{\Lambda})$ and $(\mathcal{M}^{\Lambda'}, I^{\Lambda'})$ are properly strong shift equivalent.*

We will first show the only if part of the theorem above. In our proof, we will use Nasu's factorization theorem for topological conjugacy between subshifts into bipartite codes ([N]).

We now introduce the notion of bipartite symbolic matrix system.

Definition. A symbolic matrix system $(\mathcal{M}, I)$ over alphabet Σ is said to be *bipartite* if there exist disjoint subsets $C, D \subset \Sigma$ and increasing sequences $n(l), n'(l)$ on $l \in \mathbb{N}$ with $m(l) = n(l) + n'(l)$ such that for each $l \in \mathbb{N}$, there exist an $n(l) \times n'(l+1)$ matrix $\mathcal{P}_{l,l+1}$ over C, an $n'(l) \times n(l+1)$ matrix $\mathcal{Q}_{l,l+1}$ over D, an $n(l) \times n(l+1)$ matrix $X_{l,l+1}$ over $\{0, 1\}$ and an $n'(l) \times n'(l+1)$ matrix $X'_{l,l+1}$ over $\{0, 1\}$ satisfying the following equations:

$$\mathcal{M}_{l,l+1} = \begin{bmatrix} 0 & \mathcal{P}_{l,l+1} \\ \mathcal{Q}_{l,l+1} & 0 \end{bmatrix}, \qquad I_{l,l+1} = \begin{bmatrix} X_{l,l+1} & 0 \\ 0 & X'_{l,l+1} \end{bmatrix}.$$

We thus see

LEMMA 4.3. *For a bipartite symbolic matrix system $(\mathcal{M}, I)$ as above, set*

$$\mathcal{P}_l = \mathcal{P}_{l,l+1}, \quad \mathcal{Q}_l = \mathcal{Q}_{l,l+1}, \quad X_l = X_{l,l+1}, \quad X'_l = X'_{l,l+1}$$

and

$$\mathcal{M}^{CD}_{l,l+1} = \mathcal{P}_{2l}\mathcal{Q}_{2l+1}, \qquad \mathcal{M}^{DC}_{l,l+1} = \mathcal{Q}_{2l}\mathcal{P}_{2l+1},$$
$$I^{CD}_{l,l+1} = X_{2l}X_{2l+1}, \qquad I^{DC}_{l,l+1} = X'_{2l}X'_{2l+1}.$$

Then the both pairs $(\mathcal{M}^{CD}, I^{CD})$ and $(\mathcal{M}^{DC}, I^{DC})$ are symbolic matrix systems over alphabets $C \cdot D$ and $D \cdot C$ respectively and they are properly strong shift equivalent in 1-step.

Proof. The relations $I_{l,l+1}\mathcal{M}_{l+1,l+2} = \mathcal{M}_{l,l+1}I_{l+1,l+2}$ and

$$I_{2l,2l+1}I_{2l+1,2l+2}\mathcal{M}_{2l+2,2l+3}\mathcal{M}_{2l+3,2l+4} = \mathcal{M}_{2l,2l+1}\mathcal{M}_{2l+1,2l+2}I_{2l+2,2l+3}I_{2l+3,2l+4}$$

shows that the both pairs $(\mathcal{M}^{CD}, I^{CD})$ and $(\mathcal{M}^{DC}, I^{DC})$ are symbolic matrix systems and they are properly strong shift equivalent in 1-step because we see

$$X_{l-1,l}\mathcal{P}_{l,l+1} = \mathcal{P}_{l-1,l}X'_{l,l+1}, \qquad X'_{l-1,l}\mathcal{Q}_{l,l+1} = \mathcal{Q}_{l-1,l}X_{l,l+1}.$$

DEFINITION. A λ-graph system (V, E, λ, ι) over alphabet Σ is said to be *bipartite* if there exist disjoint subsets $C, D \subset \Sigma$ such that $\Sigma = C \cup D$ and disjoint subsets $V_l^C, V_l^D \subset V_l$ for each $l \in \mathbb{N}$ such that $V_l^C \cup V_l^D = V_l$ and

(1) for each $e \in E_{l,l+1}$

$$s(e) \in V_l^D, \quad r(e) \in V_{l+1}^C \quad \text{if and only if} \quad \lambda(e) \in C,$$
$$s(e) \in V_l^C, \quad r(e) \in V_{l+1}^D \quad \text{if and only if} \quad \lambda(e) \in D.$$

(2)
$$\iota(V_{l+1}^D) = V_l^D, \qquad \iota(V_{l+1}^C) = V_l^C.$$

PROPOSITION 4.4. *A symbolic matrix system is bipartite if and only if its corresponding λ-graph system is bipartite.*

Proof. It is clear that a bipartite symbolic matrix system gives rise to a bipartite λ-graph system. Conversely, suppose that a λ-graph system (V, E, λ, ι) is bipartite. Let $n(l)$ and $n'(l)$ be the cardinalities of the sets V_l^D and V_l^C respectively. We may identify V_l^D and V_l^C with the sets $\{1, \dots, n(l)\}$ and $\{1, \dots, n'(l)\}$ respectively. For $i \in V_l^D, j \in V_{l+1}^C$, put $\mathcal{P}_{l,l+1}(i,j) = \lambda(e_1) + \cdots + \lambda(e_p)$ where $e_k \in E_{l,l+1}, k = 1, \dots, p$ are the set of all edges in $E_{l,l+1}$ satisfying $s(e_k) = i, r(e_k) = j$. Similarly we define for $i \in V_l^C, j \in V_{l+1}^D$, put $\mathcal{Q}_{l,l+1}(i,j) = \lambda(f_1) + \cdots + \lambda(f_q)$ where $f_k \in E_{l,l+1}, k = 1, \dots, q$ are the set of all edges in $E_{l,l+1}$ satisfying $s(f_k) = i, r(f_k) = j$. For $i \in V_l^D, j \in V_{l+1}^D$, put $X_{l,l+1}(i,j) = 1$ if $\iota(j) = i$ and $X_{l,l+1}(i,j) = 0$ otherwise. Similarly for $i \in V_l^C, j \in V_{l+1}^C$, put $X'_{l,l+1}(i,j) = 1$ if $\iota(j) = i$ and $X'_{l,l+1}(i,j) = 0$ otherwise. Then by these matrices, we know that the corresponding symbolic matrix system $(\mathcal{M}, I)$ for (V, E, λ, ι) is bipartite.

M. Nasu introduced the notion of bipartite subshift in [N] and [N2]. A subshift Λ over alphabet Σ is said to be bipartite if there exist disjoint subsets $C, D \subset \Sigma$ such that any $(x_i)_{i\in\mathbb{Z}} \in \Lambda$ is either

$$x_i \in C \text{ and } x_{i+1} \in D \text{ for all } i \in \mathbb{Z} \quad \text{or} \quad x_i \in D \text{ and } x_{i+1} \in C \text{ for all } i \in \mathbb{Z}.$$

Let $\Lambda^{(2)}$ be the 2-higher power shift for Λ. Put

$$\Lambda_{CD} = \{(c_i d_i)_{i\in\mathbb{Z}} \in \Lambda^{(2)} | c_i \in C, d_i \in D\},$$
$$\Lambda_{DC} = \{(d_i c_i)_{i\in\mathbb{Z}} \in \Lambda^{(2)} | c_i \in C, d_i \in D\}.$$

They are subshifts over alphabets $C \cdot D$ and $D \cdot C$ respectively. Hence $\Lambda^{(2)}$ is partitioned into the two subshifts Λ_{CD} and Λ_{DC}.

Proposition 4.5. *A subshift Λ is bipartite if and only if its canonical symbolic matrix system $(\mathcal{M}^\Lambda, I^\Lambda)$ is bipartite.*

Proof. It is clear that a bipartite canonical symbolic matrix system gives rise to a bipartite subshift from the preceding proposition. Suppose that Λ is bipartite with respect to alphabets C, D. It suffices to show that its canonical λ-graph system (V, E, λ, ι) is bipartite. As in the construction of the canonical λ-graph system, the vertex set V_l is the set of all l-past equivalence classes $\{F_i^l\}_{i=1,\dots,m(l)}$. Put

$$V_l^C = \{F_i^l | x_1 \in D \text{ for all } (x_1, x_2, \dots,) \in F_i^l\},$$
$$V_l^D = \{F_i^l | x_1 \in C \text{ for all } (x_1, x_2, \dots,) \in F_i^l\}$$

so that we have a disjoint union $V_l^C \cup V_l^D = V_l$. It is easy to see that this decomposition of $V_l, l \in \mathbb{N}$ yields a bipartite decomposition of the λ-graph system (V, E, λ, ι).

Let Λ be a bipartite subshift over Σ with respect to alphabets C, D. As in Lemma 4.3, we have two symbolic matrix systems $(\mathcal{M}^{CD}, I^{CD})$ and $(\mathcal{M}^{DC}, I^{DC})$ over alphabets $C \cdot D$ and $D \cdot C$ from the bipartite canonical symbolic matrix system $(\mathcal{M}^\Lambda, I^\Lambda)$ for Λ respectively. They are naturally identified with the canonical symbolic matrix systems for the subshifts Λ_{CD} and Λ_{DC} respectively.
We thus see by Lemma 4.3.

Corollary 4.6. *For a bipartite subshift Λ with respect to alphabets C, D, we have*

$$(\mathcal{M}^{CD}, I^{CD}) \underset{1-pr}{\approx} (\mathcal{M}^{DC}, I^{DC})$$

a properly strong shift equivalence in 1-step.

The following notion of bipartite conjugacy has been introduced by Nasu in [N], [N2]. The conjugacy from Λ_{CD} onto Λ_{DC} that maps $(c_i d_i)_{i\in\mathbb{Z}}$ to $(d_i c_{i+1})_{i\in\mathbb{Z}}$ is called the forward bipartite conjugacy. The conjugacy from Λ_{CD} onto Λ_{DC} that maps $(c_i d_i)_{i\in\mathbb{Z}}$ to $(d_{i-1} c_i)_{i\in\mathbb{Z}}$ is called the backward bipartite conjugacy. A topological conjugacy between subshifts is called a symbolic conjugacy if it is a 1-block map given by a bijection between the underlying alphabets of the subshifts. M. Nasu proved the following factorization theorem.

LEMMA 4.7(M.NASU [N]). *Any topological conjugacy ψ between subshifts is factorized into a composition of the form*

$$\psi = \kappa_n \zeta_n \kappa_{n-1} \zeta_{n-1} \cdots \kappa_1 \zeta_1 \kappa_0$$

where $\kappa_0, \dots, \kappa_n$ are symbolic conjugacies and $\zeta_1, \dots, \zeta_n$ are either forward or backward bipartite conjugacies.

Thanks to the Nasu's result above, we reach the following theorem

THEOREM 4.8. *For two subshifts Λ, Λ', let $(\mathcal{M}, I), (\mathcal{M}', I')$ be their canonical symbolic matrix systems for Λ, Λ' respectively. If Λ and Λ' are topologically conjugate, the symbolic matrix systems $(\mathcal{M}, I), (\mathcal{M}', I')$ are properly strong shift equivalent.*

We will prove the converse implication of the theorem above. We will indeed prove the following proposition.

PROPOSITION 4.9. *If two symbolic matrix systems are properly strong shift equivalent in 1-step, their associated subshifts are topologically conjugate.*

To prove the proposition, we provide a notation and a lemma.

Set the $m(l) \times m(l+k)$ matrices:

$$\begin{aligned} I_{l,l+k} &= I_{l,l+1} \cdot I_{l+1,l+2} \cdots I_{l+k-1,l+k}, \\ \mathcal{M}_{l,l+k} &= \mathcal{M}_{l,l+1} \cdot \mathcal{M}_{l+1,l+2} \cdots \mathcal{M}_{l+k-1,l+k} \end{aligned}$$

for each $l, k \in \mathbb{N}$.

LEMMA 4.10. *Assume that two symbolic matrix systems $(\mathcal{M}, I)$ over Σ and $(\mathcal{M}', I')$ over Σ' are properly strong shift equivalent in 1-step. Let $\varphi : \Sigma \to C \cdot D$ and $\phi : \Sigma' \to D \cdot C$ be specifications that give a properly strong shift equivalence in 1-step between them. For any word $x_1 x_2 \in (\Lambda_{(\mathcal{M},I)})^2$ of length two in the associated subshift $\Lambda_{(\mathcal{M},I)}$, put $\varphi(x_i) = c_i d_i, i = 1, 2$ where $c_i \in C, d_i \in D$. Then there uniquely exists a symbol $y_0 \in \Sigma'$ such that $\phi(y_0) = d_1 c_2$.*

Proof. Note that by definition the specification ϕ is not necessarily defined on all the elements of Σ'. It suffices to show the existence of y_0. Since $x_1 x_2 \in (\Lambda_{(\mathcal{M},I)})^*$, for any fixed $l \geq 3$, we find $j = 1, 2, \dots, m(l+2)$ and $k =$

$1, 2, \dots, m(l)$ such that x_1x_2 appears in $\mathcal{M}_{l,l+2}(k,j)$. Take $i = 1, 2, \dots, m(l-2)$ with $I_{l-2,l}(i,k) = 1$. Hence x_1x_2 appears in $I_{l-2,l}\mathcal{M}_{l,l+2}(i,j)$. As we know the equality:

$$I_{l-2,l}\mathcal{M}_{l,l+2} \overset{\varphi}{\simeq} I_{l-2,l-1}X_{2l-1}\mathcal{P}_{2l-1}\mathcal{Q}_{2l}\mathcal{P}_{2l+1}\mathcal{Q}_{2l+2}X_{2l+3},$$

the word $\varphi(x_1x_2) = c_1d_1c_2d_2$ appears in a component of the right hand symbolic matrix above. Thus the word d_1c_2 appears in a component of $\mathcal{Q}_{2l}\mathcal{P}_{2l+1}$. By the equality $\mathcal{M}'_{l,l+1} \overset{\phi}{\simeq} \mathcal{Q}_{2l}\mathcal{P}_{2l+1}$, we can find a symbol y_0 in the corresponding component of the matrix $\mathcal{M}'_{l,l+1}$ so that $\phi(y_0) = d_1c_2$.

Proof of Proposition 4.9. Suppose that $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ are properly strong shift equivalent in 1-step. We use the same notation as in Definition of properly strong shift equivalence. Set $\Lambda = \Lambda_{(\mathcal{M},I)}$ and $\Lambda' = \Lambda_{(\mathcal{M}',I')}$. By the preceding lemma, we have a 2-block map Φ from Λ^2 to Σ' defined by $\Phi(x_1x_2) = y_0$ where $\phi(y_0) = d_1c_2$ and $\varphi(x_i) = c_id_i, i = 1, 2$. Let Φ_∞ be the sliding block code induced by Φ so that Φ_∞ is a map from Λ to ${\Sigma'}^{\mathbb{Z}}$. We also write as Φ the map from Λ^* to the set of all words of Σ' defined by

$$\Phi(x_1x_2\cdots x_n) = \Phi(x_1x_2)\Phi(x_2x_3)\cdots\Phi(x_{n-1}x_n).$$

We will prove that $\Phi_\infty(\Lambda) \subset \Lambda'$. To prove this, it suffices to show that for any word w in Λ, $\Phi(w)$ is an admissible word in Λ'. For $w = w_1w_2\cdots w_n \in \Lambda^n$ and any fixed $l \geq n+1$, we find $j = 1, 2, \dots, m(l+n)$ and $k = 1, 2, \dots, m(l)$ such that w appears in $\mathcal{M}_{l,l+n}(k,j)$. Take $i = 1, 2, \dots, m(l-n)$ with $I_{l-n,l}(i,k) = 1$. Hence w appears in $I_{l-n,l}\mathcal{M}_{l,l+n}(i,j)$. Put $\varphi(w_i) = c_id_i, i = 1, 2, \dots, n$. By the equality

$$I_{l-1,l}\mathcal{M}_{l,l+n} \overset{\varphi}{\simeq} X_{2l-2}\mathcal{P}_{2l-1}\mathcal{Q}_{2l}\mathcal{P}_{2l+1}\mathcal{Q}_{2l+2}\cdots\mathcal{P}_{2l+2n-3}\mathcal{Q}_{2l+2n-2}X_{2l+2n-1},$$

the word $d_1c_2d_2c_3\cdots d_{n-1}c_n$ appears in a component of $\mathcal{Q}_{2l}\mathcal{P}_{2l+1}\mathcal{Q}_{2l+2}\cdots\mathcal{P}_{2l+2n-3}$. Hence the word $\phi^{-1}(d_1c_2)\phi^{-1}(d_2c_3)\cdots\phi^{-1}(d_{n-1}c_n)$ appears in a component of $\mathcal{M}'_{l,l+1}\cdot\mathcal{M}'_{l+1,l+2}\cdots\mathcal{M}'_{l+n-2,l+n-1}$. Thus we see that $\Phi(w)$ is an admissible word in Λ' and that the sliding block code Φ_∞ maps Λ to Λ'. Similarly, we can construct a sliding block code Ψ_∞ from Λ' to Λ that is an inverse of Φ_∞. Thus two subshifts Λ' and Λ are topologically conjugate.

Therefore we conclude the following theorem

Theorem 4.11. *If two symbolic matrix systems are properly strong shift equivalent, their associated subshifts are topologically conjugate.*

By Theorem 4.8 and Theorem 4.11, we conclude Theorem 4.2.

Remark. If there exist the matrices $\mathcal{P}_l, \mathcal{Q}_l$ for all sufficiently large number l in Definition of properly strong shift equivalence in 1-step, we may show that the associated subshifts are topologically conjugate because of the proof of Proposition 4.9.

Properly strong shift equivalence exactly corresponds to a finite sequence of bipartite decompositions of symbolic matrix systems and λ-graph systems. The definition of properly strong shift equivalence for symbolic matrix systems however needs rather complicated formulations than that of strong shift equivalence for nonnegative matrices. We will next introduce the notion of strong shift equivalence between two symbolic matrix systems that is simpler and weaker condition than properly strong shift equivalence. It is also a generalization of the notion of strong shift equivalence between nonnegative matrices defined by Williams in [Wi] and between symbolic matrices defined by Nasu in [N]. Let $(\mathcal{M}, I), (\mathcal{M}', I)$ be two symbolic matrix systems over alphabet Σ, Σ' respectively. Let $m(l), m'(l)$ be the sequences for which $\mathcal{M}_{l,l+1}, I_{l,l+1}$ are $m(l) \times m(l+1)$ matrices and $\mathcal{M}'_{l,l+1}, I'_{l,l+1}$ are $m'(l) \times m'(l+1)$ matrices respectively.

Definition. Two symbolic matrix systems $(\mathcal{M}, I), (\mathcal{M}', I)$ are said to be *strong shift equivalent in 1-step* if there exist alphabets C, D and specifications

$$\varphi : \Sigma \to C \cdot D, \qquad \phi : \Sigma' \to D \cdot C$$

such that for each $l \in \mathbb{N}$, there exist an $m(l-1) \times m'(l)$ matrix $\mathcal{H}_l$ over C and an $m'(l-1) \times m(l)$ matrix $\mathcal{K}_l$ over D satisfying the following equations:

$$I_{l-1,l}\mathcal{M}_{l,l+1} \overset{\varphi}{\simeq} \mathcal{H}_l\mathcal{K}_{l+1}, \qquad I'_{l-1,l}\mathcal{M}'_{l,l+1} \overset{\phi}{\simeq} \mathcal{K}_l\mathcal{H}_{l+1}$$

and

$$\mathcal{H}_l I'_{l,l+1} = I_{l-1,l}\mathcal{H}_{l+1}, \qquad \mathcal{K}_l I_{l,l+1} = I'_{l-1,l}\mathcal{K}_{l+1}.$$

We write this situation as

$$(\mathcal{M}, I) \underset{1-st}{\approx} (\mathcal{M}', I').$$

Two symbolic matrix systems $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ are said to be *strong shift equivalent in N-step* if there exist symbolic matrix systems $(\mathcal{M}^{(i)}, I^{(i)}), i = 1, 2, \dots, N-1$ such that

$$\begin{aligned}(\mathcal{M}, I) \underset{1-st}{\approx} (\mathcal{M}^{(1)}, I^{(1)}) &\underset{1-st}{\approx} (\mathcal{M}^{(2)}, I^{(2)}) \\ &\underset{1-st}{\approx} \cdots \underset{1-st}{\approx} (\mathcal{M}^{(N-1)}, I^{(N-1)}) \underset{1-st}{\approx} (\mathcal{M}', I').\end{aligned}$$

We denote this situation by

$$(\mathcal{M}, I) \underset{N-st}{\approx} (\mathcal{M}', I')$$

and simply call it a *strong shift equivalence*.

Similarly to the case of properly strong shift equivalence, we see that strong shift equivalence on symbolic matrix systems is an equivalence relation.

Proposition 4.12. *Properly strong shift equivalence in 1-step implies strong shift equivalence in 1-step.*

Proof. Let $\mathcal{P}_l, \mathcal{Q}_l, X_l$ and X'_l be the matrices in Definition of properly strong shift equivalence in 1-step between $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$. We set

$$\mathcal{H}_l = X_{2l-1}\mathcal{P}_{2l-1}, \qquad \mathcal{K}_l = X'_{2l-1}\mathcal{Q}_{2l-1}.$$

They give rise to a strong shift equivalence in 1-step between $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$.

Conversely we have

Proposition 4.13. *Suppose that both $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ are canonical. If they are strong shift equivalent in 1-step, they are properly strong shift equivalent in 1-step. Hence strong shift equivalence on canonical symbolic matrix systems is completely the same as properly strong shift equivalence.*

Proof. Let Λ, Λ' be the associated subshifts for $(\mathcal{M}, I), (\mathcal{M}', I')$ respectively. Suppose that $(\mathcal{M}, I) \underset{1-st}{\approx} (\mathcal{M}', I')$. We use the same notation as in Definition of strong shift equivalence. Set

$$\begin{aligned}\Lambda_\varphi = \{(\dots, c_{-1}, d_{-1}, \dot{c}_0, d_0, c_1, d_1, \dots)| \\ \text{there exists } (x_i)_{i\in\mathbb{Z}} \in \Lambda; \varphi(x_i) = c_i d_i \text{ for all } i \in \mathbb{Z}\}, \\ \Lambda'_\phi = \{(\dots, d_{-1}, c_0, \dot{d}_0, c_1, d_1, c_2, \dots)| \\ \text{there exists } (y_i)_{i\in\mathbb{Z}} \in \Lambda'; \phi(y_i) = d_i c_i \text{ for all } i \in \mathbb{Z}\}\end{aligned}$$

where $\dot{c}_0, \dot{d}_0$ locate at the position of the 0-th coordinate in the sequences. Put

$$\Lambda_o = \Lambda_\varphi \cup \Lambda'_\phi$$

that becomes a subshift over $C \cup D$ because of strong shift equivalence between $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$. It is clear that Λ_o is a bipartite subshift with respect to the alphabets C, D. Hence the 2-higher power shift $\Lambda_o^{(2)}$ is decomposed as $\Lambda_o^{(2)} = \Lambda_\varphi^{(2)} \cup \Lambda_\phi^{(2)}$. As there exist symbolic conjugacies:

$$\Lambda \overset{\varphi}{\simeq} \Lambda_\varphi^{(2)}, \qquad \Lambda' \overset{\phi}{\simeq} \Lambda_\phi^{(2)},$$

the canonical symbolic matrix systems for the subshifts Λ and Λ' are properly strong shift equivalence in 1-step by the previous discussions.

By a similar argument to the proof of Proposition 4.9, we obtain

Proposition 4.14. *If two symbolic matrix systems (not necessarily canonical) are strong shift equivalent in 1-step, their associated subshifts are topologically conjugate.*

Thus we conclude

Theorem 4.15. *If two symbolic matrix systems (not necessarily canonical) are strong shift equivalent, their associated subshifts are topologically conjugate.*

5. Higher λ-graph systems

In studies of symbolic dynamics, the operation of taking higher block presentation plays important rôles (cf.[Kit], [LM]). In topological Markov shifts, the operation of taking 2-higher block presentation is a typical example of giving strong shift equivalence in 1-step. The N-higher block presentation of an edge shift corresponds to the edge shift of the N-higher edge graph. We in this section introduce higher λ-graph systems and correspondingly higher symbolic matrix systems. It follows that the subshift associated with the N-higher λ-graph system is the N-higher block presentation of the subshift associated with the original λ-graph system. We see that a symbolic matrix system is properly strong shift equivalent in N-step to its N-higher symbolic matrix system. We treat a left resolving λ-graph system, that is, the incoming edges to each vertex carry different labels. General case and also general state splitting procedure of λ-graph systems will be treated in a forthcoming paper.

For a left resolving λ-graph system (V, E, λ, ι) over alphabet Σ and a natural number $N \in \mathbb{N}$, we will define a λ-graph system $(V^{[N]}, E^{[N]}, \lambda^{[N]}, \iota^{[N]})$ over $\Sigma^{[N]} = \underbrace{\Sigma \cdots \Sigma}_{\text{N-times}}$ as follows:

$$
\begin{aligned}
V_l^{[N]} = \{(e_1, e_2, \dots, e_{N-1}) \in E_{l,l+1} \times E_{l+1,l+2} \times \cdots \times E_{l+N-2,l+N-1} |\\
r(e_i) = s(e_{i+1}) \text{ for } i = 1, 2, \dots, N-2\},\\
E_{l,l+1}^{[N]} = \{((e_1, \dots, e_{N-1}),(f_1, \dots, f_{N-1})) \in V_l^{[N]} \times V_{l+1}^{[N]} |\\
e_{i+1} = f_i \text{ for } i = 1, 2, \dots, N-2\}.
\end{aligned}
$$

The maps

$$
s^{[N]} : E_{l,l+1}^{[N]} \to V_l^{[N]}, \qquad r^{[N]} : E_{l,l+1}^{[N]} \to V_{l+1}^{[N]}
$$

are defined by

$$
\begin{aligned}
s^{[N]}((e_1, \dots, e_{N-1}), (f_1, \dots, f_{N-1})) &= (e_1, \dots, e_{N-1}),\\
r^{[N]}((e_1, \dots, e_{N-1}), (f_1, \dots, f_{N-1})) &= (f_1, \dots, f_{N-1}).
\end{aligned}
$$

Set $V^{[N]} = \cup_{l \in \mathbb{N}} V_l^{[N]}$ and $E^{[N]} = \cup_{l \in \mathbb{N}} E_{l,l+1}^{[N]}$. Hence $(V^{[N]}, E^{[N]}, s^{[N]}, r^{[N]})$ is a Bratteli diagram. A labeling $\lambda^{[N]}$ on $(V^{[N]}, E^{[N]})$ is defined by

$$
\lambda^{[N]}((e_1, \dots, e_{N-1}), (f_1, \dots, f_{N-1})) = \lambda(e_1)\lambda(e_2)\dots\lambda(e_{N-1})\lambda(f_{N-1}) \in \Sigma^{[N]}
$$

for $((e_1, \dots, e_{N-1}), (f_1, \dots, f_{N-1})) \in E^{[N]}$. A sequence of surjections $\iota^{[N]} : V_{l+1}^{[N]} \to V_l^{[N]}, l \in \mathbb{N}$ is defined as follows. For $(e_1, \dots, e_{N-1}) \in V_{l+1}^{[N]}$, since the λ-graph system (V, E, λ, ι) is left resolving, there uniquely exist $e'_i \in E_{l+i-1,l+i}$ for $i = 1, 2, \dots, N-2$ such that

$$
\iota(s(e_i)) = s(e'_i), \quad \iota(r(e_i)) = r(e'_i), \quad \lambda(e_i) = \lambda(e'_i).
$$

As we know $(e'_1,\dots,e'_{N-1}) \in V_l^{[N]}$, by setting $\iota^{[N]}(e_1,\dots,e_{N-1}) = (e'_1,\dots,e'_{N-1})$. We get a λ-graph system $(V^{[N]},E^{[N]},\lambda^{[N]},\iota^{[N]})$ over $\Sigma^{[N]}$.

Definition. We call the λ-graph system $(V^{[N]},E^{[N]},\lambda^{[N]},\iota^{[N]})$ the *N-higher λ-graph system for* (V,E,λ,ι). For a symbolic matrix system $(\mathcal{M},I)$, the *N-higher symbolic matrix system* $(\mathcal{M}^{[N]},I^{[N]})$ is defined to be the symbolic matrix system associated with the N-higher λ-graph system for the λ-graph system of $(\mathcal{M},I)$.

It is routine to show the following proposition.

Proposition 5.1. $\Lambda_{(\mathcal{M}^{[N]},I^{[N]})} = (\Lambda_{(\mathcal{M},I)})^{[N]}$.

As seen in the case of nonnegative matrices, we see

Proposition 5.2. $(\mathcal{M},I) \underset{1-pr}{\approx} (\mathcal{M}^{[2]},I^{[2]})$: *a properly strong shift equivalence in 1-step.*

Proof. Let (V,E,λ,ι) and $(V^{[2]},E^{[2]},\lambda^{[2]},\iota^{[2]})$ be the associated λ-graph systems for $(\mathcal{M},I)$, $(\mathcal{M}^{[2]},I^{[2]})$ over alphabets Σ and $\Sigma^{[2]}$ respectively. We will construct a bipartite λ-graph system $(\hat{V},\hat{E},\hat{\lambda},\hat{\iota})$ that gives rise to a properly strong shift equivalence in 1-step between the λ-graph systems. We set for $l \in \mathbb{N}$

$$\hat{V}_{2l-1} = E_{l,l+1} \cup V_l, \qquad \hat{V}_{2l} = V_{l+1} \cup E_{l,l+1}$$

and

$$\hat{E}_{2l-1,2l} = \{(f,u) \in E_{l,l+1} \times V_{l+1} | u = r(f)\} \cup \{(v,e) \in V_l \times E_{l,l+1} | v = s(e)\},$$
$$\hat{E}_{2l,2l+1} = \{(v,e) \in V_{l+1} \times E_{l+1,l+2} | v = s(e)\} \cup \{(f,u) \in E_{l,l+1} \times V_{l+1} | u = r(f)\}.$$

The source maps $\hat{s}_{2l-1,2l} : \hat{E}_{2l-1,2l} \to \hat{V}_{2l-1}$ and $\hat{s}_{2l,2l+1} : \hat{E}_{2l,2l+1} \to \hat{V}_{2l}$ are defined as follows:

$$\hat{s}_{2l-1,2l}(f,u) = f \in E_{l,l+1}, \qquad \hat{s}_{2l-1,2l}(v,e) = v \in V_l,$$
$$\hat{s}_{2l,2l+1}(v,e) = v \in V_{l+1}, \qquad \hat{s}_{2l,2l+1}(f,u) = f \in E_{l,l+1}.$$

The range maps $\hat{r}_{2l-1,2l} : \hat{E}_{2l-1,2l} \to \hat{V}_{2l}$ and $\hat{r}_{2l,2l+1} : \hat{E}_{2l,2l+1} \to \hat{V}_{2l+1}$ are defined as follows:

$$\hat{r}_{2l-1,2l}(f,u) = u \in V_{l+1}, \qquad \hat{r}_{2l-1,2l}(v,e) = e \in E_{l,l+1},$$
$$\hat{r}_{2l,2l+1}(v,e) = e \in E_{l+1,l+2}, \qquad \hat{r}_{2l,2l+1}(f,u) = u \in V_{l+1}.$$

The maps $\hat{\iota}_{2l,2l-1} : \hat{V}_{2l} \to \hat{V}_{2l-1}$ and $\hat{\iota}_{2l+1,2l} : \hat{V}_{2l+1} \to \hat{V}_{2l}$ are defined as follows:

$$\hat{\iota}_{2l,2l-1}(u) = \iota(u) \text{ for } u \in V_{l+1}, \quad \hat{\iota}_{2l,2l-1}(f) = f \text{ for } f \in E_{l,l+1},$$
$$\hat{\iota}_{2l+1,2l}(e) = \iota(e) \text{ for } e \in E_{l+1,l+2}, \quad \hat{\iota}_{2l+1,2l}(v) = v \text{ for } v \in V_{l+1}$$

where $\iota(e) \in E_{l,l+1}$ is naturally defined for $e \in E_{l+1,l+2}$. Put $D_\Sigma = \{D_\alpha | \alpha \in \Sigma\}$, $C_\Sigma = \{C_\alpha | \alpha \in \Sigma\}$ and $\hat{\Sigma} = D_\Sigma \cup C_\Sigma$. The labeling $\hat{\lambda}$ is defined as a map from $\hat{E}$ to the alphabet $\hat{\Sigma}$ as follows: For $(f,u),(v,e)$ in $\hat{E}_{2l-1,2l} = \{(f,u) \in E_{l,l+1} \times V_{l+1} | u = r(f)\} \cup \{(v,e) \in V_l \times E_{l,l+1} | v = s(e)\}$, we set

$$\hat{\lambda}(f,u) = C_{\lambda(f)}, \quad \hat{\lambda}(v,e) = D_{\lambda(e)}.$$

For $(v,e),(f,u)$ in $\hat{E}_{2l,2l+1} = \{(v,e) \in V_{l+1} \times E_{l+1,l+2} | v = s(e)\} \cup \{(f,u) \in E_{l,l+1} \times V_{l+1} | u = r(f)\}$, we set

$$\hat{\lambda}(v,e) = D_{\lambda(e)}, \quad \hat{\lambda}(f,u) = C_{\lambda(f)}.$$

Then it is routine to check that $(\hat{V}, \hat{E}, \hat{\lambda}, \hat{\iota})$ is a bipartite λ-graph system over alphabet $\hat{\Sigma}$. Through the specifications $\varphi : \Sigma \to D_\Sigma \cdot C_\Sigma$ and $\phi : \Sigma^{[2]} \to C_\Sigma \cdot D_\Sigma$ defined by

$$\varphi(\alpha) = D_\alpha \cdot C_\alpha \quad \text{ and } \quad \phi(\alpha,\beta) = D_\alpha \cdot C_\beta,$$

we know that the symbolic matrix system for $(\hat{V}, \hat{E}, \hat{\lambda}, \hat{\iota})$ gives rise to a properly strong shift equivalence in 1-step between $(\mathcal{M}, I)$ and $(\mathcal{M}^{[2]}, I^{[2]})$.

Since $(\mathcal{M}^{[N+1]}, I^{[N+1]})$ is isomorphic to $((\mathcal{M}^{[N]})^{[2]}, (I^{[N]})^{[2]})$, we have

COROLLARY 5.3. *For any symbolic matrix system $(\mathcal{M}, I)$, we have*

$$(\mathcal{M}^{[N]}, I^{[N]}) \underset{N-st}{\approx} (\mathcal{M}, I)$$

a properly strong shift equivalence in N-step.

6. SHIFT EQUIVALENCE

By the discussions of Section 4, the topological conjugacy classes of subshifts are completely characterized by the strong shift equivalence classes of the associated canonical symbolic matrix systems. However, even for topological Markov shifts, there is no general algorithm known for deciding whether two nonnegative matrices are strong shift equivalent. R. F. Williams introduced the notion of shift equivalence between two nonnegative matrices that is weaker but easier to treat than the notion of strong shift equivalence ([Wi]). The formulation of shift equivalence between nonnegative matrices is described by certain algebraic relations between the matrices that determine a crucial invariant called the dimension group ([Kr], [Kr2]). The notion of shift equivalence has been generalized to symbolic matrices by Boyle-Krieger and studied as a topological conjugacy invariant for sofic subshifts in [BK].

We in this section introduce the notion of shift equivalence between two symbolic matrix systems as a generalization of Williams's notion for nonnegative matrices and Boyle-Krieger's notion for symbolic matrices. Let $(\mathcal{M}, I), (\mathcal{M}', I')$ be two symbolic matrix systems over alphabets Σ, Σ' respectively. For $N \in \mathbb{N}$, we put $(\Sigma)^N = \Sigma \cdots \Sigma, (\Sigma')^N = \Sigma' \cdots \Sigma'$: the N-times products.

Definition. For $N \in \mathbb{N}$, two symbolic matrix systems $(\mathcal{M}, I), (\mathcal{M}', I')$ are said to be *shift equivalent of lag* N if there exist alphabets C_N, D_N and specifications

$$\varphi_1 : \Sigma \cdot C_N \to C_N \cdot \Sigma', \qquad \varphi_2 : \Sigma' \cdot D_N \to D_N \cdot \Sigma$$

and

$$\psi_1 : (\Sigma)^N \to C_N \cdot D_N, \qquad \psi_2 : (\Sigma')^N \to D_N \cdot C_N$$

such that for each $l \in \mathbb{N}$, there exist an $m(l) \times m'(l+N)$ matrix $\mathcal{H}_l$ over C_N and an $m'(l) \times m(l+N)$ matrix $\mathcal{K}_l$ over D_N satisfying the following equations:

$$\mathcal{M}_{l,l+1}\mathcal{H}_{l+1} \overset{\varphi_1}{\simeq} \mathcal{H}_l\mathcal{M}'_{l+N,l+N+1}, \qquad \mathcal{M}'_{l,l+1}\mathcal{K}_{l+1} \overset{\varphi_2}{\simeq} \mathcal{K}_l\mathcal{M}_{l+N,l+N+1},$$

$$I_{l,l+N}\mathcal{M}_{l+N,l+2N} \overset{\psi_1}{\simeq} \mathcal{H}_l\mathcal{K}_{l+N}, \qquad I'_{l,l+N}\mathcal{M}'_{l+N,l+2N} \overset{\psi_2}{\simeq} \mathcal{K}_l\mathcal{H}_{l+N}$$

and

$$I_{l,l+1}\mathcal{H}_{l+1} = \mathcal{H}_l I'_{l+N,l+N+1}, \qquad I'_{l,l+1}\mathcal{K}_{l+1} = \mathcal{K}_l I_{l+N,l+N+1}.$$

We denote this situation by

$$(\mathcal{M}, I) \underset{lagN}{\sim} (\mathcal{M}', I') \quad \text{or} \quad (\mathcal{H}, \mathcal{K}) : (\mathcal{M}, I) \underset{lagN}{\sim} (\mathcal{M}', I')$$

and simply call it a *shift equivalence.*

Similarly to the case of nonnegative matrices and symbolic matrices, we can see the following lemma.

Lemma 6.1.

(i) $(\mathcal{M}, I) \underset{lagN}{\sim} (\mathcal{M}', I')$ *implies* $(\mathcal{M}, I) \underset{lagL}{\sim} (\mathcal{M}', I')$ *for all* $L \geq N$.

(ii) $(\mathcal{M}, I) \underset{lagN}{\sim} (\mathcal{M}', I')$ *and* $(\mathcal{M}', I') \underset{lagN'}{\sim} (\mathcal{M}'', I'')$ *implies* $(\mathcal{M}, I) \underset{lagN+N'}{\sim} (\mathcal{M}'', I'')$. *Hence shift equivalence is an equivalence relation on symbolic matrix systems.*

Proof. (i) Suppose that $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ are shift equivalent of lag N. It suffices to show that they are shift equivalent of lag $N+1$. We use the same notation as above. Set the alphabets

$$C_{N+1} = C_N, \qquad D_{N+1} = D_N \cdot \Sigma.$$

Put the specification $\varphi'_1 = \varphi_1 : \Sigma \cdot C_{N+1} \to C_{N+1} \cdot \Sigma'$. Through the specification φ_2, we have a natural specification $\varphi'_2 : \Sigma' \cdot D_{N+1} \to D_{N+1} \cdot \Sigma$. Similarly, through the specifications $\psi_1, \psi_2, \varphi_1$, we have natural specifications

$$\psi'_1 : (\Sigma)^{N+1} \to C_{N+1} \cdot D_{N+1}, \qquad \psi'_2 : (\Sigma')^{N+1} \to D_{N+1} \cdot C_{N+1}.$$

Put the matrices

$$\mathcal{H}'_l = \mathcal{H}_l I'_{l+N,l+N+1}, \quad \mathcal{K}'_l = \mathcal{K}_l \mathcal{M}_{l+N,l+N+1}.$$

Then it is straightforward to see that they give a shift equivalence of lag $N+1$ between $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$.
(ii) Assume that

$$(\mathcal{H},\mathcal{K}) : (\mathcal{M}, I) \underset{lagN}{\sim} (\mathcal{M}', I'), \qquad (\mathcal{H}',\mathcal{K}') : (\mathcal{M}', I') \underset{lagN'}{\sim} (\mathcal{M}'', I'').$$

Then it is routine to check that

$$(\mathcal{H}\mathcal{H}',\mathcal{K}'\mathcal{K}) : (\mathcal{M}, I) \underset{lagN+N'}{\sim} (\mathcal{M}'', I'').$$

Similarly to the case of matrices, we have

Theorem 6.2. *Strong shift equivalence in N-step implies shift equivalence of lag N.*

Proof. Suppose that $(\mathcal{M}, I) \underset{N-st}{\approx} (\mathcal{M}', I')$ a strong shift equivalence in N-step. There exist symbolic matrix systems $(\mathcal{M}^{(i)}, I^{(i)})$ for $i = 1, \dots, N-1$ such that

$$\begin{aligned}(\mathcal{M}, I) = (\mathcal{M}^{(0)}, I^{(0)}) &\underset{1-st}{\approx} (\mathcal{M}^{(1)}, I^{(1)}) \underset{1-st}{\approx} (\mathcal{M}^{(2)}, I^{(2)}) \underset{1-st}{\approx} \\ &\cdots \underset{1-st}{\approx} (\mathcal{M}^{(N-1)}, I^{(N-1)}) \underset{1-st}{\approx} (\mathcal{M}^{(N)}, I^{(N)}) = (\mathcal{M}', I').\end{aligned}$$

Let $\mathcal{H}_l^{(i)}, \mathcal{K}_l^{(i)}$ be rectangular symbolic matrices that give a strong shift equivalence between $(\mathcal{M}^{(i-1)}, I^{(i-1)})$ and $(\mathcal{M}^{(i)}, I^{(i)})$ where $\mathcal{H}_l^{(i)}$ is an $m^{(i-1)}(l-1) \times m^{(i)}(l)$ matrix over alphabet $C(i)$ and $\mathcal{K}_l^{(i)}$ is an $m^{(i)}(l-1) \times m^{(i-1)}(l)$ matrix over alphabet $D(i)$ for each $l \in \mathbb{N}$ and $i = 1, \dots, N$. Set the alphabets

$$C_N = C(1) \cdots C(N), \qquad D_N = D(1) \cdots D(N).$$

Put the matrices

$$\mathcal{P}_l = \mathcal{H}_{l+2}^{(1)} \mathcal{H}_{l+3}^{(2)} \cdots \mathcal{H}_{l+N+1}^{(N)}, \qquad \mathcal{Q}_l = \mathcal{K}_{l+2}^{(1)} \mathcal{K}_{l+3}^{(2)} \cdots \mathcal{K}_{l+N+1}^{(N)}$$

an $m(l) \times m'(l+N)$ matrix over C_N, an $m'(l) \times m(l+N)$ matrix over D_N respectively. We then have the following natural specifications

$$\varphi_1 : \Sigma \cdot C_N \to C_N \cdot \Sigma', \qquad \varphi_2 : \Sigma' \cdot D_N \to D_N \cdot \Sigma$$

and

$$\psi_1 : (\Sigma)^N \to C_N \cdot D_N, \qquad \psi_2 : (\Sigma')^N \to D_N \cdot C_N$$

that yield a shift equivalence of lag N between $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$.

For a subshift (Λ, σ) over Σ, its n-higher power shift $(\Lambda^{(n)}, \sigma)$ is defined to be the subshift (Λ, σ^n) over $(\Sigma)^n$ (cf.[LM]). Two subshifts is called eventually conjugate if their n-higher power shifts are conjugate for all large enough n ([Wi], [KimR]). Williams and Kim-Roush showed that two square nonnegative matrices are shift equivalent if and only if the associated topological Markov shifts are eventually conjugate. Boyle-Krieger generalized their result to symbolic matrices and sofic subshifts ([BK]). W. Krieger kindly asked the author whether or not these results can be generalized to general subshifts. The author sincerely thanks him for his question.

Proposition 6.3. *If symbolic matrix systems $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ are shift equivalent, their associated subshifts $\Lambda_{(\mathcal{M},I)}$ and $\Lambda_{(\mathcal{M}',I')}$ are eventually conjugate.*

To show the proposition, we provide a lemma that is proved by a straightforward calculation.

Lemma 6.4. *For a symbolic matrix system $(\mathcal{M}, I)$, let Λ the associated subshift. We set for $n, l \in \mathbb{N}$,*

$$I^n_{l,l+1} = I_{nl,nl+1} I_{nl+1,nl+2} \cdots I_{nl+n-1,nl+n},$$
$$\mathcal{M}^n_{l,l+1} = \mathcal{M}_{nl,nl+1} \mathcal{M}_{nl+1,nl+2} \cdots \mathcal{M}_{nl+n-1,nl+n}.$$

Then $(\mathcal{M}^n, I^n)$ becomes a symbolic matrix system whose associated subshift is the n-higher power shift $\Lambda^{(n)}$ of Λ.

Proof of Proposition 6.3. Put $\Lambda = \Lambda_{(\mathcal{M},I)}$, $\Lambda' = \Lambda_{(\mathcal{M}',I')}$ over Σ, Σ' respectively. Assume that

$$(\mathcal{H}, \mathcal{K}) : (\mathcal{M}, I) \underset{lagN}{\sim} (\mathcal{M}', I').$$

For a number $K \in \mathbb{N}$, put $n = K + N$. We will see that $\Lambda^{(n)} \underset{1-st}{\approx} \Lambda'^{(n)}$. Let C_N, D_N be alphabets as in Definition of shift equivalence. Set $C = C_N, D = D_N \cdot (\Sigma)^K$. There are natural specifications

$$(\Sigma)^n \to C \cdot D, \qquad (\Sigma')^n \to D \cdot C$$

by using the specifications in the shift equivalence between Λ and Λ'. Put the matrices

$$\mathcal{P}_l = \mathcal{H}_{nl-n} I'_{nl-K,nl-K+1} I'_{nl-K+1,nl-K+2} \cdots I'_{nl-1,nl},$$
$$\mathcal{Q}_l = \mathcal{K}_{nl-n} \mathcal{M}_{nl-K,nl-K+1} \mathcal{M}_{nl-K+1,nl-K+2} \cdots \mathcal{M}_{nl-1,nl}.$$

They are an $m(nl-n) \times m'(nl)$ matrix over C and an $m'(nl-n) \times m(nl)$ matrix over D respectively. We see that they yield a strong shift equivalence in 1-step between $(\mathcal{M}^n, I^n)$ and $(\mathcal{M'}^n, I'^n)$ so that their associated subshifts are topologically conjugate by Theorem 4.15.

We will comment on the notion of properly shift equivalence between symbolic matrix systems. The following is the definition of properly shift equivalence that is a slightly stronger than shift equivalence and weaker than properly strong shift equivalence.
Let $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ be symbolic matrix systems over alphabets Σ, Σ' respectively. Hence $\mathcal{M}_{l,l+1}, I_{l,l+1}$ are $m(l) \times m(l+1)$ matrices and $\mathcal{M}'_{l,l+1}, I'_{l,l+1}$ are $m'(l) \times m'(l+1)$ matrices.
Definition. $(\mathcal{M}, I)$ and $(\mathcal{M}', I')$ are said to be *properly shift equivalent of lag* N if there exist alphabets C_N, D_N and specifications

$$\varphi_1 : \Sigma \cdot C_N \to C_N \cdot \Sigma', \qquad \varphi_2 : \Sigma' \cdot D_N \to D_N \cdot \Sigma,$$
$$\psi_1 : (\Sigma)^N \to C_N \cdot D_N, \qquad \psi_2 : (\Sigma')^N \to D_N \cdot C_N$$

and increasing sequences $n(l), n'(l)$ on $l \in \mathbb{N}$ such that for each $l \in \mathbb{N}$, there exist an $n(k) \times n'(k+2N-1)$ matrix $\mathcal{P}_k$ over C_N, an $n'(k) \times n(k+2N-1)$ matrix $\mathcal{Q}_k$ over D_N for $k = 2l, 2l+2N-1$, an $n(l) \times n(l+1)$ matrix X_l over $\{0,1\}$ and an $n'(l) \times n'(l+1)$ matrix X'_l over $\{0,1\}$ satisfying the following equations:
(6.1)

$$\mathcal{M}_{l,l+N} I_{l+N,l+2N-1} \overset{\psi_1}{\simeq} \mathcal{P}_{2l} \mathcal{Q}_{2l+2N-1}, \quad \mathcal{M}'_{l,l+N} I'_{l+N,l+2N-1} \overset{\psi_2}{\simeq} \mathcal{Q}_{2l} \mathcal{P}_{2l+2N-1},$$

$$\mathcal{M}_{l,l+1} \mathcal{P}_{2(l+1)} X'_{2l+2N+1} \overset{\varphi_1}{\simeq} \mathcal{P}_{2l} X'_{2l+2N-1} \mathcal{M}'_{l+N,l+N+1},$$
$$\mathcal{M}'_{l,l+1} \mathcal{Q}_{2(l+1)} X_{2l+2N+1} \overset{\varphi_2}{\simeq} \mathcal{Q}_{2l} X_{2l+2N-1} \mathcal{M}_{l+N,l+N+1},$$
$$I_{l,l+1} = X_{2l} X_{2l+1}, \qquad I'_{l,l+1} = X'_{2l} X'_{2l+1}$$

and

$$X_l \mathcal{P}_{l+1} = \mathcal{P}_l X'_{l+2N-1}, \qquad X'_l \mathcal{Q}_{l+1} = \mathcal{Q}_l X_{l+2N-1}.$$

We denote this situation by

$$(\mathcal{M}, I) \underset{N-pr}{\sim} (\mathcal{M}', I').$$

It follows that by (6.1), $n(2l) = m(l)$ and $n'(2l) = m'(l)$ for $l \in \mathbb{N}$.
For $N = 1$, if we understand that the matrices $I_{l+1,l+1}$ and $I'_{l+1,l+1}$ are the $m(l+1) \times m(l+1)$ identity matrix and the $m'(l+1) \times m'(l+1)$ identity matrix respectively, the properly shift equivalence of lag 1 is exactly the same as the properly strong shift equivalence in 1-step.

This definition is also a generalization of Boyle-Krieger 's shift equivalence between symbolic matrices ([BK] see also [N2]).
The following proposition is routine.

Proposition 6.5.

(i) $(\mathcal{M}, I) \underset{N-pr}{\sim} (\mathcal{M}', I')$ *implies* $(\mathcal{M}, I) \underset{lagN}{\sim} (\mathcal{M}', I')$. *That is, properly shift equivalence implies shift equivalence.*

(ii) $(\mathcal{M}, I) \underset{N-pr}{\approx} (\mathcal{M}', I')$ *implies* $(\mathcal{M}, I) \underset{N-pr}{\sim} (\mathcal{M}', I')$. *That is, properly strong shift equivalence implies properly shift equivalence.*

We thus summarize as in the following way:

$$\begin{array}{ccc} (\mathcal{M}, I) \underset{N-pr}{\approx} (\mathcal{M}', I') & \Longrightarrow & (\mathcal{M}, I) \underset{N-pr}{\sim} (\mathcal{M}', I')) \\ \Downarrow & & \Downarrow \\ (\mathcal{M}, I) \underset{N-st}{\approx} (\mathcal{M}', I') & \Longrightarrow & (\mathcal{M}, I) \underset{lagN}{\sim} (\mathcal{M}', I'). \end{array}$$

We may define strong shift equivalence and shift equivalence between subshifts as their corresponding properties for their canonical symbolic matrix systems. Hence we can say that two subshifts are topologically conjugate if and only if they are strong shift equivalence. The strong shift equivalence for subshifts imply the shift equivalence.

7. Nonnegative matrix systems

In this section, we will introduce the notion of nonnegative matrix system that is also a generalization of nonnegative matrices. We will then generalize strong shift equivalence and shift equivalence between nonnegative matrices to between nonnegative matrix systems. Let $(A_{l,l+1}, I_{l,l+1}), l \in \mathbb{N}$ be a pair of sequences of rectangular matrices such that the following four conditions for each $l \in \mathbb{N}$ are satisfied:

(1) $A_{l,l+1}$ is an $m(l) \times m(l+1)$ rectangular matrix with entries in nonnegative integers.

(2) $I_{l,l+1}$ is an $m(l) \times m(l+1)$ rectangular matrix with entries in $\{0, 1\}$ satisfying the following two conditions:

(2-a) For i, there exists j such that $I_{l,l+1}(i,j) \neq 0$.

(2-b) For j, there uniquely exists i such that $I_{l,l+1}(i,j) \neq 0$.

(3) $m(l) \leq m(l+1)$.

(4) $I_{l,l+1}A_{l+1,l+2} = A_{l,l+1}I_{l+1,l+2}$.

The pair (A, I) is called a *nonnegative matrix system*. For $i = 1, \dots, m(l), j = 1, \dots, m(l+1)$, we denote by $A_{l,l+1}(i,j)$, $I_{l,l+1}(i,j)$ the (i,j)-components of $A_{l,l+1}$, $I_{l,l+1}$ respectively. A nonnegative matrix system (A, I) is said to be *essential* if it satisfies the following further conditions

(5-i) For i, there exists j such that $A_{l,l+1}(i,j) \neq 0$.

(5-ii) For j, there exists i such that $A_{l,l+1}(i,j) \neq 0$.

We henceforth study essential nonnegative matrix systems and call them nonnegative matrix systems for simplicity.

The property “sofic ”for nonnegative matrix systems are similarly defined to the cases of symbolic matrix systems. The following is basic.

LEMMA 7.1. *For a symbolic matrix system* $(\mathcal{M}, I)$*, let* $M_{l,l+1}$ *be the* $m(l) \times m(l+1)$ *rectangular matrix obtained from* $\mathcal{M}_{l,l+1}$ *by setting all the symbols equal to* 1*. Then the resulting pair* (M, I) *becomes a nonnegative matrix system.*

We write the matrices above as $\mathrm{supp}(\mathcal{M}_{l,l+1}) = M_{l,l+1}$ and call $M_{l,l+1}$ the support of $\mathcal{M}_{l,l+1}$. The pair (M, I) is called the nonnegative matrix system associated with $(\mathcal{M}, I)$. Conversely we see

PROPOSITION 7.2. *For a nonnegative matrix system* (A, I) *and a symbolic matrix* $\mathcal{M}_{1,2}$ *over alphabet* Σ *such that* $\mathrm{supp}(\mathcal{M}_{1,2}) = A_{1,2}$*, there exists a sequence* $\mathcal{M}_{l,l+1}, l \in \mathbb{N}$ *of symbolic matrices over* Σ *such that the pair* $(\mathcal{M}, I)$ *is a symbolic matrix system and* $\mathrm{supp}(\mathcal{M}_{l,l+1}) = A_{l,l+1}$ *for all* $l \in \mathbb{N}$*.*

Proof. We will prove the assertion by induction. Assume that a symbolic matrix $\mathcal{M}_{k,k+1}$ is determined. For $j = 1, \dots, m(k+2)$, take a unique index $j' = 1, \dots, m(k+1)$ such that $I_{k+1,k+2}(j', j) = 1$. For $i = 1, \dots, m(k)$, suppose that $\mathcal{M}_{k,k+1}(i, j') = \alpha_1 + \cdots + \alpha_n$. Let $l_1, \dots, l_p$ be the set of all numbers $l = 1, \dots, m(k+1)$ satisfying $I_{k,k+1}(i, l) = 1$. Hence we have $n = \sum_{r=1}^{p} A_{k+1,k+2}(l_r, j)$. Put $\xi_r = A_{k+1,k+2}(l_r, j)$. Now we define

$$
\begin{aligned}
\mathcal{M}_{k+1,k+2}(l_1, j) &= \alpha_1 + \cdots + \alpha_{\xi_1}, \\
\mathcal{M}_{k+1,k+2}(l_2, j) &= \alpha_{\xi_1+1} + \cdots + \alpha_{\xi_1+\xi_2}, \\
\mathcal{M}_{k+1,k+2}(l_3, j) &= \alpha_{\xi_1+\xi_2+1} + \cdots + \alpha_{\xi_1+\xi_2+\xi_3}, \\
&\cdots\cdots \\
\mathcal{M}_{k+1,k+2}(l_p, j) &= \alpha_{\xi_1+\cdots+\xi_{p-1}+1} + \cdots + \alpha_n.
\end{aligned}
$$

Since for any $l = 1, \dots, m(k+1)$, there uniquely exists $i = 1, \dots, m(k)$ such that $I_{k,k+1}(i, l) = 1$ we may define $\mathcal{M}_{k+1,k+2}(l, j)$ for all $l = 1, \dots, m(k+1)$ by the above way. The matrices satisfy $I_{k,k+1}\mathcal{M}_{k+1,k+2} = \mathcal{M}_{k,k+1}I_{k+1,k+2}$ and $\mathrm{supp}(\mathcal{M}_{k+1,k+2})(l, j) = A_{k+1,k+2}(l, j)$.

For nonnegative matrix systems we will formulate strong shift equivalence as follows.

DEFINITION. Two nonnegative matrix systems $(A, I), (A', I')$ are said to be *strong shift equivalent in 1-step* if for each $l \in \mathbb{N}$, there exist an $m(l-1) \times m'(l)$ matrix H_l with entries in nonnegative integers and an $m'(l-1) \times m(l)$ matrix K_l with entries in nonnegative integers satisfying the following equations:

$$I_{l-1,l}A_{l,l+1} = H_l K_{l+1}, \qquad I'_{l-1,l}A'_{l,l+1} = K_l H_{l+1}$$

and

$$H_l I'_{l,l+1} = I_{l-1,l}H_{l+1}, \qquad K_l I_{l,l+1} = I'_{l-1,l}K_{l+1}.$$

We write this situation as

$$(A, I) \underset{1-st}{\approx} (A', I').$$

Two nonnegative matrix systems (A, I) and (A', I') are said to be *strong shift equivalent in N-step* if there exist nonnegative matrix systems $(A^{(i)}, I^{(i)}), i = 1, 2, \dots, N-1$ such that

$$(A, I) \underset{1-st}{\approx} (A^{(1)}, I^{(1)}) \underset{1-st}{\approx} (A^{(2)}, I^{(2)}) \underset{1-st}{\approx} \cdots \underset{1-st}{\approx} (A^{(N-1)}, I^{(N-1)}) \underset{1-st}{\approx} (A', I').$$

We denote this situation by

$$(A, I) \underset{N-st}{\approx} (A', I')$$

and simply call it a *strong shift equivalence*.

This formulation is also a generalization of Williams's strong shift equivalence between nonnegative matrices ([Wi]). Similarly to symbolic matrix systems, strong shift equivalence is an equivalence relation on nonnegative matrix systems.

We directly have

Proposition 7.3. *If two symbolic matrix systems are strong shift equivalence (in N-step), then the associated nonnegative matrix systems are strong shift equivalent (in N-step).*

We will describe the matrix relations appearing in the formulation of strong shift equivalence between nonnegative matrix systems in terms of certain single homomorphisms between inductive limits of abelian groups. For a nonnegative matrix system (A, I), the transpose $I^t_{l,l+1}$ of the matrix $I_{l,l+1}$ naturally induces an ordered homomorphism from $\mathbb{Z}^{m(l)}$ to $\mathbb{Z}^{m(l+1)}$, where the positive cone $\mathbb{Z}^{m(l)}_+$ of the group $\mathbb{Z}^{m(l)}$ is defined by

$$\mathbb{Z}^{m(l)}_+ = \{(n_1, n_2, \dots, n_{m(l)}) \in \mathbb{Z}^{m(l)} | n_i \geq 0, i = 1, 2 \dots m(l)\}.$$

We put the inductive limits:

$$\mathbb{Z}_{I^t} = \varinjlim \{I^t_{l,l+1} : \mathbb{Z}^{m(l)} \to \mathbb{Z}^{m(l+1)}\},$$
$$\mathbb{Z}^+_{I^t} = \varinjlim \{I^t_{l,l+1} : \mathbb{Z}^{m(l)}_+ \to \mathbb{Z}^{m(l+1)}_+\}.$$

The condition (2-a) for the matrix $I_{l,l+1}$ says the following lemma.

Lemma 7.4. *For each $l \in \mathbb{N}$, the homomorphism $I^t_{l,l+1} : \mathbb{Z}^{m(l)} \to \mathbb{Z}^{m(l+1)}$ is injective. Hence the canonical homomorphism $\iota_l : \mathbb{Z}^{m(l)} \to \mathbb{Z}_{I^t}$ is injective.*

By the relation: $I_{l,l+1}A_{l+1,l+2} = A_{l,l+1}I_{l+1,l+2}$, the sequence of the transposed matrices $A^t_{l,l+1}, l \in \mathbb{N}$ of the matrices $A_{l,l+1}, l \in \mathbb{N}$ yields an endomorphism of the ordered group $\mathbb{Z}_{I^t}$. We write it as $\lambda_{(A,I)}$.

Definition. For nonnegative matrix systems $(A, I), (A', I')$ and $L \in \mathbb{N}$, a homomorphism ξ from the group $\mathbb{Z}_{I^t}$ to the group $\mathbb{Z}_{I'^t}$ is said to be *finite homomorphism of lag L* if it satisfies the condition

$$\xi(\mathbb{Z}^{m(l)}) \subset \mathbb{Z}^{m'(l+L)} \qquad \text{for all } l \in \mathbb{N}$$

where $\mathbb{Z}^{m(l)}$ and $\mathbb{Z}^{m'(l)}$ are naturally imbedded into $\mathbb{Z}_{I^t}$ and $\mathbb{Z}_{I'^t}$ respectively. We then have

PROPOSITION 7.5. *Two nonnegative matrix systems (A, I) and (A', I') are strong shift equivalence in 1-step if and only if there exist order preserving finite homomorphisms of lag 1: $\xi : \mathbb{Z}_{I^t} \to \mathbb{Z}_{I'^t}$ and $\eta : \mathbb{Z}_{I'^t} \to \mathbb{Z}_{I^t}$ such that*

$$\eta \circ \xi = \lambda_{(A,I)}, \qquad \xi \circ \eta = \lambda_{(A',I')}.$$

Proof. Suppose that (A, I) and (A', I') are strong shift equivalent in 1-step. Let H_l, K_l be sequences of matrices that give rise to a strong shift equivalence between them. Then by the condition $I'^t_{l,l+1} H^t_l = H^t_{l+1} I^t_{l-1,l}$, the family $H^t_l, l \in \mathbb{N}$ yields a homomorphism from $\mathbb{Z}_{I^t}$ to $\mathbb{Z}_{I'^t}$ which we denote by ξ. Similarly we define a homomorphism η from $\mathbb{Z}_{I'^t}$ to $\mathbb{Z}_{I^t}$ induced by the family $K_l, l \in \mathbb{N}$. It is easy to see that the homomorphisms ξ, η are order preserving and finite homomorphisms of lag 1. By the condition $A'^t_{l,l+1} I^t_{l-1,l} = K^t_{l+1} H^t_l$, we see $\eta \circ \xi = \lambda_{(A,I)}$. Similarly, we have $\xi \circ \eta = \lambda_{(A',I')}$.
The converse implication is also easy by using Lemma 7.4. We in fact see that the matrices H_l, K_l are given by the transposed matrices of the restrictions of the homomorphisms ξ to $\mathbb{Z}^{m(l)} (\hookrightarrow \mathbb{Z}_{I^t})$ and η to $\mathbb{Z}^{m'(l)} (\hookrightarrow \mathbb{Z}_{I'^t})$ respectively. They satisfy the required conditions of strong shift equivalence between (A, I) and (A', I').

We will next formulate shift equivalence between two nonnegative matrix systems. For a nonnegative matrix system (A, I), we set the $m(l) \times m(l + k)$ matrices:

$$I_{l,l+k} = I_{l,l+1} \cdot I_{l+1,l+2} \cdots I_{l+k-1,l+k},$$
$$A_{l,l+k} = A_{l,l+1} \cdot A_{l+1,l+2} \cdots A_{l+k-1,l+k}$$

for each $l, k \in \mathbb{N}$.
DEFINITION. Two nonnegative matrix systems $(A, I), (A', I')$ are said to be *shift equivalent of lag* N if for each $l \in \mathbb{N}$, there exist an $m(l) \times m'(l + N)$ matrix H_l with entries in nonnegative integers and an $m'(l) \times m(l + N)$ matrix K_l with entries in nonnegative integers satisfying the following equations:

$$A_{l,l+1} H_{l+1} = H_l A'_{l+N,l+N+1}, \qquad A'_{l,l+1} K_{l+1} = K_l A_{l+N,l+N+1},$$
$$H_l K_{l+N} = I_{l,l+N} A_{l+N,l+2N}, \qquad K_l H_{l+N} = I'_{l,l+N} A'_{l+N,l+2N}$$

and

$$I_{l,l+1} H_{l+1} = H_l I'_{l+N,l+N+1}, \qquad I'_{l,l+1} K_{l+1} = K_l I_{l+N,l+N+1}.$$

We write this situation as

$$(A, I) \underset{lagN}{\sim} (A', I') \quad \text{or} \quad (H, K) : (A, I) \underset{lagN}{\sim} (A', I')$$

and simply call it a *shift equivalence*.

This formulation is a generalization of Williams's shift equivalence between square matrices with entries in nonnegative integers ([Wi] see also [BK]).
Similarly to the case of shift equivalence for nonnegative matrices and symbolic matrix systems, we have.

Lemma 7.6.

(i) $(A,I) \underset{lagN}{\sim} (A',I')$ *implies* $(A,I) \underset{lagL}{\sim} (A',I')$ *for all* $L \ge N$.

(ii) $(A,I) \underset{lagN}{\sim} (A',I')$ *and* $(A',I') \underset{lagN'}{\sim} (A'',I'')$ *implies* $(A,I) \underset{lagN+N'}{\sim} (A'',I'')$. *Hence shift equivalence is an equivalence relation on nonnegative matrix systems.*

Similarly to Theorem 6.2, we have

Proposition 7.7. *For nonnegative matrix systems, strong shift equivalence in* N*-step implies shift equivalence of lag* N.

As in the case of strong shift equivalence, we may describe the matrix relations appearing in the formulation of shift equivalence in terms of single homomorphisms between inductive limits of abelian groups.

Proposition 7.8. *Two nonnegative matrix systems* (A,I) *and* $(A'I')$ *are shift equivalent of lag* N *if and only if there exist order preserving finite homomorphisms of lag* N: $\xi : \mathbb{Z}_{I^t} \to \mathbb{Z}_{I'^t}$ *and* $\eta : \mathbb{Z}_{I'^t} \to \mathbb{Z}_{I^t}$ *such that*

$$\lambda_{(A',I')} \circ \xi = \xi \circ \lambda_{(A,I)}, \qquad \lambda_{(A,I)} \circ \eta = \eta \circ \lambda_{(A',I')}$$

and

$$\eta \circ \xi = \lambda^N_{(A,I)}, \qquad \xi \circ \eta = \lambda^N_{(A',I')}.$$

Let $(\mathcal{M},I),(\mathcal{M}',I')$ be symbolic matrix systems and $(M,I),(M',I')$ be their supports respectively. The following proposition is direct.

Proposition 7.9.

(i) $(\mathcal{M},I) \underset{n-st}{\approx} (\mathcal{M}',I')$ *implies* $(M,I) \underset{n-st}{\approx} (M',I')$.

(ii) $(\mathcal{M},I) \underset{lagN}{\sim} (\mathcal{M}',I')$ *implies* $(M,I) \underset{lagN}{\sim} (M',I')$.

8. Dimension groups

In this section, we will introduce the notions of dimension group and dimension triple for nonnegative matrix systems that is shown to be a shift equivalence invariant. It is a generalization of the notions of dimension group and dimension triple for nonnegative matrices defined by W. Krieger in [Kr], [Kr2]. The Krieger's idea to define dimension groups for nonnegative matrices is based on the K-theory for C^*-algebras (cf.[Ef]). The author considered the dimension groups for subshifts by using K_0-groups for certain C^*-algebras associated with subshifts as in [Ma2],[Ma3]. It is a generalization of the original idea of Krieger. We will in this section formulate the dimension groups and the dimension triples for nonnegative matrix systems.

Let (A,I) be a nonnegative matrix system. Recall that $\mathbb{Z}_{I^t}$ denotes the ordered group of the inductive limit of the sequence of the ordered abelian groups

$\mathbb{Z}^{m(l)}, l \in \mathbb{N}$ through the transposed matrices $I^t_{l,l+1}, l \in \mathbb{N}$. As seen in the previous discussion, the sequence of the transposed matrices $A^t_{l,l+1}$ naturally induces an order preserving endomorphism on the ordered group $\mathbb{Z}_{I^t}$ that is denoted by $\lambda_{(A,I)}$. We set $\mathbb{Z}_{I^t}(k) = \mathbb{Z}_{I^t}$ and $\mathbb{Z}^+_{I^t}(k) = \mathbb{Z}^+_{I^t}$ for $k \in \mathbb{N}$. We define an abelian group and its positive cone by the following inductive limits:

$$\Delta_{(A,I)} = \varinjlim_k \{\lambda_{(A,I)} : \mathbb{Z}_{I^t}(k) \to \mathbb{Z}_{I^t}(k+1)\},$$
$$\Delta^+_{(A,I)} = \varinjlim_k \{\lambda_{(A,I)} : \mathbb{Z}^+_{I^t}(k) \to \mathbb{Z}^+_{I^t}(k+1)\}.$$

We call the ordered group $(\Delta_{(A,I)}, \Delta^+_{(A,I)})$ the *dimension group for* (A,I). Since the map $\delta_{(A,I)} : \mathbb{Z}_{I^t}(k) \to \mathbb{Z}_{I^t}(k+1)$ defined by $\delta_{(A,I)}([X,k]) = ([X,k+1])$ for $X \in \mathbb{Z}_{I^t}$ yields an automorphism on $\Delta_{(A,I)}$ that preserves the positive cone $\Delta^+_{(A,I)}$. We also denote it by $\delta_{(A,I)}$ and call it the *dimension automorphism.* We call the triple $(\Delta_{(A,I)}, \Delta^+_{(A,I)}, \delta_{(A,I)})$ the *dimension triple for* (A,I) and the pair $(\Delta_{(A,I)}, \delta_{(A,I)})$ the *dimension pair for* (A,I).

PROPOSITION 8.1. *If two nonnegative matrix systems are shift equivalent, their dimension triples are isomorphic.*

Proof. Suppose that two nonnegative matrix systems (A,I) and (A',I') are shift equivalent of lag N. By Proposition 7.8, there exist order preserving finite homomorphisms $\xi : \mathbb{Z}_{I^t} \to \mathbb{Z}_{I'^t}$ and $\eta : \mathbb{Z}_{I'^t} \to \mathbb{Z}_{I^t}$ of lag N such that

$$\lambda_{(A',I')} \circ \xi = \xi \circ \lambda_{(A,I)}, \qquad \lambda_{(A,I)} \circ \eta = \eta \circ \lambda_{(A',I')}$$

and

$$\eta \circ \xi = \lambda^N_{(A,I)}, \qquad \xi \circ \eta = \lambda^N_{(A',I')}.$$

Define the maps $\Phi_\xi : \mathbb{Z}_{I^t}(k) \to \mathbb{Z}_{I'^t}(k)$ and $\Phi_\eta : \mathbb{Z}_{I'^t}(k) \to \mathbb{Z}_{I^t}(k)$ as $\Phi_\xi([X,k]) = ([\xi(X),k])$ and $\Phi_\eta([Y,k]) = ([\eta(Y),k])$ for $X \in \mathbb{Z}_{I^t}$, $Y \in \mathbb{Z}_{I'^t}$. It is easy to see that they induce homomorphisms from $\Delta_{(A,I)}$ to $\Delta_{(A',I')}$ and $\Delta_{(A',I')}$ to $\Delta_{(A,I)}$ respectively. We still denote them by Φ_ξ and Φ_η respectively. Since the homomorphisms ξ, η are order preserving, the maps Φ_ξ, Φ_η also preserve order structures of the dimension groups. It then follows that

$$\delta_{(A,I)} \circ \Phi_\eta = \Phi_\eta \circ \delta_{(A'I')}, \qquad \delta_{(A',I')} \circ \Phi_\xi = \Phi_\xi \circ \delta_{(A,I)}$$

and

$$\Phi_\eta \circ \Phi_\xi = \delta^{-N}_{(A,I)}, \qquad \Phi_\xi \circ \Phi_\eta = \delta^{-N}_{(A',I')}.$$

Therefore we see that the both maps Φ_ξ and Φ_η are isomorphisms and the corresponding dimension triples are isomorphic.

In particular we have (cf.[BK])

Proposition 8.2. *Two sofic nonnegative matrix systems are shift equivalent if and only if their dimension triples are isomorphic. Thus the dimension triple are complete invariants for shift equivalence of sofic nonnegative matrix systems.*

Proof. The only if part is from the preceding proposition. By a similar discussion to [Kr],[Kr2], we obtain the if part of the assertion.

We will define the dimension triples for symbolic matrix systems as the dimension triples for their supports. Namely let $(\mathcal{M}, I)$ be a symbolic matrix system and (M, I) its support. Then the dimension triple $(\Delta_{(\mathcal{M},I)}, \Delta^+_{(\mathcal{M},I)}, \delta_{(\mathcal{M},I)})$ is defined to be the dimension triple $(\Delta_{(M,I)}, \Delta^+_{(M,I)}, \delta_{(M,I)})$. We may also define dimension triples for subshifts as the dimension triple for their canonical symbolic matrix systems. Let Λ be a subshift and $(\mathcal{M}, I)$ its canonical symbolic matrix system for Λ. Then the *future dimension triple* $(\Delta_\Lambda, \Delta^+_\Lambda, \delta_\Lambda)$ for subshift Λ is defined to be the dimension triple $(\Delta_{(\mathcal{M},I)}, \Delta^+_{(\mathcal{M},I)}, \delta_{(\mathcal{M},I)})$. The *past dimension triple* for Λ is defined as the future dimension triple for the transposed subshift Λ^T for Λ.
Thus we have

Proposition 8.3. *The future dimension triples for subshifts are shift equivalence invariants and in particular topological conjugacy invariants.*

The notion of dimension pair $(\Delta_\Lambda, \delta_\Lambda)$ for subshifts has been also seen in [Le].

9. K-groups and Bowen-Franks groups

The Bowen-Franks groups for nonnegative matrices and hence for topological Markov shifts have been introduced by R. Bowen and J. Franks in [BF]. For an $n \times n$ nonnegative square matrix A, its Bowen-Franks group $BF(A)$ is defined by the group $\mathbb{Z}^n/(1-A)\mathbb{Z}^n$. This group has discovered in a study of suspension flows of topological Markov shifts by Bowen and Franks (cf. [PS]). They showed that the groups are not only invariants under shift equivalence but also almost complete invariants under flow equivalence between nonnegative matrices.
We will in this section introduce and study the notion of Bowen-Franks groups for nonnegative matrix systems as a generalization of the original Bowen-Franks groups for nonnegative matrices. Our Bowen-Franks groups for a nonnegative matrix system consist of a pair of abelian groups. One corresponds to a generalization of the original Bowen-Franks group, called the Bowen-Franks group of degree zero, and the other one corresponds to its suspension, called the Bowen-Franks group of degree one. For matrices, the latter group is the torsion-free part of the original Bowen-Franks group. But in general nonnegative matrix systems the group of degree one is not necessarily the torsion-free part of the group of degree zero (see Section 10).
Before going to definition of the Bowen-Franks groups for nonnegative matrix systems, we introduce two abelian groups for nonnegative matrix systems, called K-groups, that will be proved to be invariant under shift equivalence.

Let (A, I) be a nonnegative matrix system. For $l \in \mathbb{N}$, we set the abelian groups

$$K_0^l(A, I) = \mathbb{Z}^{m(l+1)}/(I_{l,l+1}^t - A_{l,l+1}^t)\mathbb{Z}^{m(l)},$$
$$K_1^l(A, I) = \mathrm{Ker}(I_{l,l+1}^t - A_{l,l+1}^t) \text{ in } \mathbb{Z}^{m(l)}.$$

LEMMA 9.1. *The map* $I_{l,l+1}^t : \mathbb{Z}^{m(l)} \to \mathbb{Z}^{m(l+1)}$ *naturally induces homomorphisms between the following groups:*

$$i_*^l : K_*^l(A, I) \to K_*^{l+1}(A, I) \qquad \textit{for} \quad * = 0, 1.$$

The proof is straightforward by using the relations

$$I_{l,l+1}A_{l+1,l+2} = A_{l,l+1}I_{l+1,l+2}.$$

We now define the K-groups for nonnegative matrix system (A, I).
DEFINITION. The *K-groups for* (A, I) are defined as the following inductive limits of the abelian groups:

$$K_0(A, I) = \varinjlim_l \{i_0^l : K_0^l(A, I) \to K_0^{l+1}(A, I)\},$$
$$K_1(A, I) = \varinjlim_l \{i_1^l : K_1^l(A, I) \to K_1^{l+1}(A, I)\}.$$

For a symbolic matrix system $(\mathcal{M}, I)$, its K-groups $K_0(\mathcal{M}, I), K_1(\mathcal{M}, I)$ are defined to be the K-groups for the associated nonnegative matrix systems. It is easy to see that the groups $K_*(A, I)$ are also represented as in the following way

PROPOSITION 9.2.

(i) $K_0(A, I) = \mathbb{Z}_{I^t}/(id - \lambda_{(A,I)})\mathbb{Z}_{I^t}$,
(ii) $K_1(A, I) = \mathrm{Ker}(id - \lambda_{(A,I)})$ *in* $\mathbb{Z}_{I^t}$.

We will see that the groups $K_*(A, I)$ are invariant under shift equivalence.

LEMMA 9.3.

(i) $K_0(A, I) = \Delta_{(A,I)}/(id - \delta_{(A,I)})\Delta_{(A,I)}$,
(ii) $K_1(A, I) = \mathrm{Ker}(id - \delta_{(A,I)})$ *in* $\Delta_{(A,I)}$.

Proof. As the automorphism $\delta_{(A,I)}$ is given by $\lambda_{(A,I)} = \{A_{l,l+1}^t\}$ on $\Delta_{(A,I)}$, the assertions are easily proved.

Since the dimension triple $(\Delta_{(A,I)}, \Delta_{(A,I)}^+, \delta_{(A,I)})$ is invariant under shift equivalence of nonnegative matrix systems, we thus have

Proposition 9.4. *The groups* $K_i(A,I), i = 0,1$ *are invariant under shift equivalence of nonnegative matrix systems.*

Set the abelian group

$$\mathbb{Z}_I = \varprojlim_l \{I_{l,l+1} : \mathbb{Z}^{m(l+1)} \to \mathbb{Z}^{m(l)}\}$$

the projective limit of the system: $I_{l,l+1} : \mathbb{Z}^{m(l+1)} \to \mathbb{Z}^{m(l)}, l \in \mathbb{N}$. The sequence $A_{l,l+1}, l \in \mathbb{N}$ naturally acts on $\mathbb{Z}_I$ as an endomorphism that we denote by A. The identity on $\mathbb{Z}_I$ is denoted by I. We now define the Bowen-Franks groups for (A,I) as follows:

Definition. For a nonnegative matrix system (A,I),

$$BF^0(A,I) = \mathbb{Z}_I/(I-A)\mathbb{Z}_I, \qquad BF^1(A,I) = \mathrm{Ker}(I-A) \text{ in } \mathbb{Z}_I.$$

We call $BF^0(A,I)$ the Bowen-Franks group for (A,I) of degree zero and $BF^1(A,I)$ the Bowen-Franks group for (A,I) of degree one. We see

Theorem 9.5. *The Bowen-Franks groups* $BF^i(A,I), i = 0,1$ *are invariant under shift equivalence of nonnegative matrix systems.*

Proof. (i) Suppose that two nonnegative matrix systems (A,I) and (A',I') are shift equivalent of lag N. Let H_l, K_l be sequences of nonnegative matrices such that $(H,K) : (A,I) \underset{lagN}{\sim} (A',I')$. For $(x_i)_{i\in\mathbb{N}} \in \mathbb{Z}_I$, put $\Phi_K((x_i)_{i\in\mathbb{N}}) = (K_i(x_{N+i})_{i\in\mathbb{N}})$. It is easy to see that the Φ_K gives rise to a homomorphism from $\mathbb{Z}_I$ to $\mathbb{Z}_{I'}$. As we see the equality: $K_i \circ (I_{N+i,N+i+1} - A_{N+i,N+i+1}) = (I'_{i,i+1} - A'_{i,i+1}) \circ K_{i+1}$, the homomorphism induces a homomorphism from $\mathbb{Z}_I/(I-A)\mathbb{Z}_I$ to $\mathbb{Z}_{I'}/(I'-A')\mathbb{Z}_{I'}$. We denote it by $\bar{\Phi}_K$. We similarly have a homomorphism $\bar{\Phi}_H$ from $\mathbb{Z}_{I'}/(I'-A')\mathbb{Z}_{I'}$ to $\mathbb{Z}_I/(I-A)\mathbb{Z}_I$. Since we have $\Phi_H \circ \Phi_K = A^N$ on $\mathbb{Z}_I$ and $\Phi_K \circ \Phi_H = {A'}^N$ on $\mathbb{Z}_{I'}$, the homomorphisms $\bar{\Phi}_H$ and $\bar{\Phi}_K$ are inverses each other.

(ii) It is direct to see that the homomorphisms Φ_H and Φ_K induce isomorphisms between $\mathrm{Ker}(I-A)$ in $\mathbb{Z}_I$ and $\mathrm{Ker}(I'-A')$ in $\mathbb{Z}_{I'}$.

We will prove the following Universal Coefficient Theorem. It says that the Bowen-Franks groups are determined by the K-groups.

Theorem 9.6.

(i) *There exists a short exact sequence*

$$0 \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}(K_0(A,I),\mathbb{Z}) \xrightarrow{\delta} BF^0(A,I) \xrightarrow{\gamma} \mathrm{Hom}_{\mathbb{Z}}(K_1(A,I),\mathbb{Z}) \longrightarrow 0$$

that splits unnaturally.

(ii)

$$BF^1(A,I) \cong \mathrm{Hom}_{\mathbb{Z}}(K_0(A,I),\mathbb{Z}).$$

In the theorem above, $\mathrm{Ext}^1_{\mathbb{Z}}$ is the derived functor of the Hom-functor in homological algebra. The formulations above come from the Universal Coefficient Theorem for K-theory of the C^*-algebra $\mathcal{O}_\Lambda$ associated with subshift Λ ([Ma4]). General framework of the Universal Coefficient Theorem for K-theory of C^*-algebras have been proved in [Bro], [RS]. If an abelian group G is finitely generated, it is well known that

$$\begin{aligned}\mathrm{Hom}_{\mathbb{Z}}(G,\mathbb{Z}) = &\ \text{The torsion-free part of } G,\\ \mathrm{Ext}^1_{\mathbb{Z}}(G,\mathbb{Z}) = &\ \text{The torsion part of } G.\end{aligned}$$

We provide some lemmas to prove Theorem 9.6.

Lemma 9.7. $\mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z}) = 0$.

Proof. It suffices to show that an extension

$$0 \longrightarrow \mathbb{Z} \longrightarrow G \xrightarrow{\rho} \mathbb{Z}_{I^t} \longrightarrow 0$$

of abelian groups splits. For each $l \in \mathbb{N}$, let ι_l be the canonical inclusion of $\mathbb{Z}^{m(l)}$ into $\mathbb{Z}_{I^t}$. We will choose homomorphisms $\varphi_l : \mathbb{Z}^{m(l)} \to G$ such that

$$\rho \circ \varphi_l = \iota_l, \qquad \varphi_{l+1} \circ I^t_{l,l+1} = \varphi_l$$

as follows: Let $e^l_i, i = 1,\dots,m(l)$ be the standard basis of $\mathbb{Z}^{m(l)}$. We first take homomorphisms $\phi_l : \mathbb{Z}^{m(l)} \to G$ such that $\rho \circ \phi_l = \iota_l$ for $l \in \mathbb{N}$. Put $\varphi_1 = \phi_1$. Since we see $\rho((\phi_2 \circ I^t_{1,2} - \varphi_1)(e^1_i)) = 0$, we may regard the element $\phi_2 \circ I^t_{1,2}(e^1_i) - \varphi_1(e^1_i)$ as an integer m^1_i. For each $i = 1,\dots,m(1)$, take $r_i = 1,\dots,m(2)$ such that $I_{1,2}(i,r_i) = 1$. We set

$$\varphi_2(e^2_j) = \begin{cases} \phi_2(e^2_j) - m^1_i & \text{if } j = r_i \\ \phi_2(e^2_j) & \text{otherwise.}\end{cases}$$

Then it is easy to see that

$$\rho \circ \varphi_2 = \iota_2, \qquad \varphi_2 \circ I^t_{1,2} = \varphi_1.$$

By continuing these procedures, we can find a sequence of homomorphisms $\varphi_l, l \in \mathbb{N}$ that have the desired property. They give rise to a homomorphism $\varphi : \mathbb{Z}_{I^t} \to G$ such that $\rho \circ \varphi = id$.

Lemma 9.8.

(i) $\mathrm{Ext}^1_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z}) = 0$.
(ii) $\mathrm{Ext}^1_{\mathbb{Z}}(\mathrm{Ker}(id - \lambda_{(A,I)})$ *in* $\mathbb{Z}_{I^t},\mathbb{Z}) = 0$.

Proof. Regard $(id - \lambda_{(A,I)})\mathbb{Z}_{I^t}$ and $\mathrm{Ker}(id - \lambda_{(A,I)})$ in $\mathbb{Z}_{I^t}$ as subgroups of $\mathbb{Z}_{I^t}$. Consider the following short exact sequences:

(9.1)
$$0 \longrightarrow (id - \lambda_{(A,I)})\mathbb{Z}_{I^t} \xrightarrow{\iota} \mathbb{Z}_{I^t} \xrightarrow{q} \mathbb{Z}_{I^t}/(id - \lambda_{(A,I)})\mathbb{Z}_{I^t} \longrightarrow 0,$$
(9.2)
$$0 \longrightarrow \mathrm{Ker}(id - \lambda_{(A,I)}) \xrightarrow{j} \mathbb{Z}_{I^t} \xrightarrow{id-\lambda_{(A,I)}} (id - \lambda_{(A,I)})\mathbb{Z}_{I^t} \longrightarrow 0$$

of abelian groups. They yield the following exact sequences respectively:

$$\begin{aligned} \cdots \longrightarrow & \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z}) \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \\ & \longrightarrow \mathrm{Ext}^2_{\mathbb{Z}}(\mathbb{Z}_{I^t}/(id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \longrightarrow \cdots \end{aligned}$$

and

$$\begin{aligned} \cdots \longrightarrow & \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z}) \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}(\mathrm{Ker}(id - \lambda_{(A,I)}), \mathbb{Z}) \\ & \longrightarrow \mathrm{Ext}^2_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \longrightarrow \cdots \end{aligned}$$

As $\mathrm{Ext}^2_{\mathbb{Z}} = 0$, we have

$$\mathrm{Ext}^1_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) = \mathrm{Ext}^1_{\mathbb{Z}}(\mathrm{Ker}(id - \lambda_{(A,I)}), \mathbb{Z}) = 0$$

by the preceding lemma.

Lemma 9.9.

(i)
$$\begin{aligned} & \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_{I^t}/(id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \\ & \cong \mathrm{Hom}_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z})/\iota^*\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z}). \end{aligned}$$

(ii)
$$\begin{aligned} & \mathrm{Hom}_{\mathbb{Z}}(\mathrm{Ker}(id - \lambda_{(A,I)}), \mathbb{Z}) \\ & \cong \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z})/(id - \lambda_{(A,I)})^*\mathrm{Hom}_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}). \end{aligned}$$

Proof. The short exact sequences (9.1) and (9.2) make the following sequences exact:

(9.3)
$$\begin{aligned} 0 \longrightarrow & \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t}/(id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \\ & \xrightarrow{q^*} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z}) \xrightarrow{\iota^*} \mathrm{Hom}_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \\ & \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_{I^t}/(id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z}) \longrightarrow \cdots, \end{aligned}$$
(9.4)
$$\begin{aligned} 0 \longrightarrow & \mathrm{Hom}_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \xrightarrow{(id-\lambda_{(A,I)})^*} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z}) \\ & \xrightarrow{j^*} \mathrm{Hom}_{\mathbb{Z}}(\mathrm{Ker}(id - \lambda_{(A,I)}), \mathbb{Z}) \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z}) \longrightarrow \cdots. \end{aligned}$$

Hence we get the desired isomorphisms.

Proof of Theorem 9.6. (i) By Proposition 9.2 and the previous lemmas, we have

$$\begin{aligned}\mathrm{Hom}_{\mathbb{Z}}(K_1(A,I),\mathbb{Z}) \cong & \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z})/(id-\lambda_{(A,I)})^*\mathrm{Hom}_{\mathbb{Z}}((id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z}),\\ \mathrm{Ext}^1_{\mathbb{Z}}(K_0(A,I),\mathbb{Z}) \cong & \mathrm{Hom}_{\mathbb{Z}}((id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z})/\iota^*\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z}).\end{aligned}$$

The exact sequence (9.4) says the map

$$(id-\lambda_{(A,I)})^* : \mathrm{Hom}_{\mathbb{Z}}((id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z})$$

is injective. Hence we know that the group

$$\mathrm{Hom}_{\mathbb{Z}}((id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z})/\iota^*\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z})$$

is isomorphic to the group

$$(id-\lambda_{(A,I)})^*\mathrm{Hom}_{\mathbb{Z}}(id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z})/(id-\lambda_{(A,I)})^*\iota^*\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z}).$$

The map

$$(id-\lambda_{(A,I)})^*\iota^* : \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z})$$

is naturally regarded as the endomorphism

$$I-A : \mathbb{Z}_I \to \mathbb{Z}_I$$

through a natural identification between $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z})$ and $\mathbb{Z}_I$. As there exists an short exact sequence

$$\begin{aligned}0 \longrightarrow & (id-\lambda_{(A,I)})^*\mathrm{Hom}_{\mathbb{Z}}(id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z})/(id-\lambda_{(A,I)})^*\iota^*\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z})\\ \longrightarrow & \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z})/(id-\lambda_{(A,I)})^*\iota^*\mathrm{Hom}_{\mathbb{Z}}((id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z})\\ \longrightarrow & \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z})/(id-\lambda_{(A,I)})^*\mathrm{Hom}_{\mathbb{Z}}((id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z})\\ \longrightarrow & 0,\end{aligned}$$

we obtain a short exact sequence:

$$0 \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}(K_0(A,I),\mathbb{Z}) \xrightarrow{\delta} \mathbb{Z}_I/(I-A)\mathbb{Z}_I \xrightarrow{\gamma} \mathrm{Hom}_{\mathbb{Z}}(K_1(A,I),\mathbb{Z}) \longrightarrow 0.$$

The short exact sequence above splits unnaturally, since the group $\mathrm{Ext}^1_{\mathbb{Z}}(G,\mathbb{Z})$ is algebraically compact and the group $\mathrm{Hom}_{\mathbb{Z}}(H,\mathbb{Z})$ is torsion-free for any abelian groups G, H (cf. [KKS]).
(ii) By the exact sequence (9.3), we see

$$\begin{aligned}\mathrm{Hom}_{\mathbb{Z}}(K_0(A,I),\mathbb{Z}) \cong & \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t}/(id-\lambda_{(A,I)})\mathbb{Z}_{I^t},),\mathbb{Z})\\ \cong & \mathrm{Ker}\ \iota^* : \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t},\mathbb{Z}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}((id-\lambda_{(A,I)})\mathbb{Z}_{I^t},\mathbb{Z})\end{aligned}$$

By a natural identification between $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z})$ and $\mathbb{Z}_I$, we obtain Ker ι^* : $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_{I^t}, \mathbb{Z}) \longrightarrow \mathrm{Hom}_{\mathbb{Z}}((id - \lambda_{(A,I)})\mathbb{Z}_{I^t}, \mathbb{Z})$ is regarded as $\mathrm{Ker}(I - A)$ in $\mathbb{Z}_I$. Thus we end the proof of the theorem.

Remark. Lemma 9.8 (ii) means $\mathrm{Ext}^1_{\mathbb{Z}}(K_1(A,I), \mathbb{Z}) = 0$. Hence the following short exact sequence clearly holds by Theorem 9.6 (ii):

$$0 \longrightarrow \mathrm{Ext}^1_{\mathbb{Z}}(K_1(A,I), \mathbb{Z}) \xrightarrow{\delta} BF^1(A,I) \xrightarrow{\gamma} \mathrm{Hom}_{\mathbb{Z}}(K_0(A,I), \mathbb{Z}) \longrightarrow 0.$$

Example. Let M be an $n \times n$ nonnegative matrix . Put for each $l \in \mathbb{N}$

$$A_{l,l+1} = M, \qquad I_{l,l+1} = \text{ the } n \times n \text{ identity matrix.}$$

Then (A, I) is a nonnegative matrix system. The K-groups are

$$K_0(A,I) = \mathbb{Z}^n/(1 - M^t)\mathbb{Z}^n, \qquad K_1((A,I) = \mathrm{Ker}(1 - M^t) \text{ in } \mathbb{Z}^n.$$

The Bowen-Franks groups are

$$BF^0(A,I) = \mathbb{Z}^n/(1 - M)\mathbb{Z}^n, \qquad BF^1((A,I) = \mathrm{Ker}(1 - M) \text{ in } \mathbb{Z}^n.$$

Hence we have

$$K_0(A,I) \cong BF^0(A,I) = BF(M) : \text{the original Bowen-Franks group for } M,$$
$$K_1(A,I) \cong BF^1(A,I) = \text{ the torsion-free part of } BF(M)$$

We will next define K-groups and Bowen-Franks groups for subshifts.
Definition. For a subshift Λ, let (A_Λ, I_Λ) be the canonical nonnegative matrix system associated with Λ. We define

$$K_i(\Lambda) = K_i(A_\Lambda, I_\Lambda), i = 0, 1 \quad : \text{the K-groups for } \Lambda$$
$$BF^i(\Lambda) = BF^i(A_\Lambda, I_\Lambda), i = 0, 1 \quad : \text{the Bowen-Franks groups for } \Lambda$$

We thus have

Theorem 9.10. *The K-groups $K_i(\Lambda)$ and the Bowen-Franks groups $BF^i(\Lambda)$ for subshift Λ are abelian groups that are invariant under shift equivalence of subshifts. In particular, they are topological conjugacy invariants of subshifts.*

Proposition 9.11. *Let Λ be a sofic subshift. We denote by $m(\Lambda)$ the cardinality of the vertices of the left Krieger cover graph for Λ and A_Λ its adjacency matrix. Then we have*

$$BF^0(\Lambda) = \mathbb{Z}^{m(\Lambda)}/(1 - A_\Lambda)\mathbb{Z}^{m(\Lambda)}, \qquad BF^1(\Lambda) = \mathrm{Ker}(1 - A_\Lambda) \textit{ in } \mathbb{Z}^{m(\Lambda)}.$$

Proof. As we see

$$K_0(\Lambda) = \mathbb{Z}^{m(\Lambda)}/(1 - A_\Lambda)\mathbb{Z}^{m(\Lambda)}, \qquad K_1(\Lambda) = \mathrm{Ker}(1 - A_\Lambda) \text{ in } \mathbb{Z}^{m(\Lambda)},$$

the assertion is clear.

In the final section, we will see an example of a nonsofic subshift Λ for which $BF^1(\Lambda)$ is no longer the torsion-free part of the group $BF^0(\Lambda)$.

REMARK. In [Ma], the author introduced the C^*-algebra $\mathcal{O}_\Lambda$ associated with subshift Λ as a generalization of the construction of the Cuntz-Krieger algebra $\mathcal{O}_A$ associated with the topological Markov shift Λ_A determined by a matrix A with entries in $\{0, 1\}$. Cuntz-Krieger proved in [CK] that the Ext-group $\mathrm{Ext}(\mathcal{O}_A)$ of the C^*-algebra $\mathcal{O}_A$ is $\mathbb{Z}^n/(1 - A)\mathbb{Z}^n$: the Bowen-Franks group of the matrix A. The author in [Ma4] generalized the notion of the Bowen-Franks group to the subshifts as:

$$BF(\Lambda) := \mathrm{Ext}(\mathcal{O}_\Lambda).$$

From the view point of the K-theory for C^*-algebras, the invariants $K_i, BF^i, i = 0, 1$ introduced in this section appear as

$$K_0(\Lambda) = K_0(\mathcal{O}_\Lambda), \qquad K_1(\Lambda) = K_1(\mathcal{O}_\Lambda)$$

and

$$BF^0(\Lambda) = \mathrm{Ext}(\mathcal{O}_\Lambda), \qquad BF^1(\Lambda) = \mathrm{Ext}(\mathcal{O}_\Lambda \otimes C_0(\mathbb{R})).$$

The formulations in Theorem 9.6 come from the Universal Coefficients Theorem for C^*-algebras ([Bro], [RS]).

As the K-groups and the Ext-groups for C^*-algebras are stably isomorphic invariant and the stable isomorphism class of the C^*-algebra $\mathcal{O}_\Lambda$ with gauge action is invariant under topological conjugacy class of subshifts ([Ma5]), we know that the dimension triple, the K-groups and the Bowen-Franks groups for subshifts are topological conjugacy invariants without using discussions of this paper under some mild conditions for subshifts.

The Bowen-Franks group for nonnegative matrix was first invented for use as an invariant of flow equivalence of the associated topological Markov shift rather than topological conjugacy ([BF],[Fr],[PS]). We can prove that the K-groups $K_*(\Lambda)$ and hence the Bowen-Franks groups $BF^*(\Lambda)$ for subshift are also invariant under flow equivalence of subshift by using a result of Parry-Sullivan [PS]. The proof, that we do not give in this paper, will appear in a forthcoming paper (cf.[Ma4],[Ma5]).

We will finally present another candidate of Bowen-Franks groups for subshifts. For a topological Markov shift Λ_A determined by an $n \times n$ matrix A with entries in $\{0, 1\}$, the group $BF(\Lambda_A)$ is isomorphic to the K_0-group for the subshift Λ_{A^t} determined by the transpose of the matrix A. The subshift Λ_{A^t} is the transpose Λ_A^T of Λ_A as a subshift. From this point of view, it seems to be one way to define the Bowen-Franks group for canonical symbolic matrix systems as the K-groups for their transpose.

Let $(\mathcal{M}, I)$ be a canonical symbolic matrix system and $\Lambda_{(\mathcal{M},I)}$ the associated subshift. We define the transpose $(\mathcal{M}^T, I^T)$ of $(\mathcal{M}, I)$ as the canonical symbolic matrix system for the transpose $\Lambda^T_{(\mathcal{M},I)}$ of the subshift $\Lambda_{(\mathcal{M},I)}$. We will define another pair of Bowen-Franks groups as in the following way.

Definition. For a canonical symbolic matrix system $(\mathcal{M}, I)$, we define

$$BF^i_K(\mathcal{M}, I) = K_i(\mathcal{M}^T, I^T), \qquad i = 0, 1$$

where $K_i(\mathcal{M}^T, I^T)$ is defined as the K_i-groups for the nonnegative matrix system associated with $(\mathcal{M}^T, I^T)$. We call them the *Bowen-Franks groups from K for* $(\mathcal{M}, I)$. For a subshift Λ, let $(\mathcal{M}, I)$ be its canonical symbolic matrix system. We will then define Bowen-Franks groups (from K) for subshift as follows:

$$BF^i_K(\Lambda) = BF^i_K(\mathcal{M}, I), \qquad i = 0, 1.$$

We thus have

Proposition 9.12. *The Bowen-Franks groups $BF^i_K(\Lambda), i = 0, 1$ from K for subshift Λ are topological conjugacy invariants of subshifts.*

Proof. Suppose that two subshifts Λ, Λ' are topologically conjugate. We denote by $(\mathcal{M}, I), (\mathcal{M}', I')$ their canonical symbolic matrix systems respectively. Hence their transposed subshifts Λ^T, Λ'^T are topologically conjugate so that their canonical symbolic matrix systems $(\mathcal{M}^T, I^T), (\mathcal{M}'^T, I'^T)$ are strong shift equivalent and hence shift equivalent. As their corresponding nonnegative matrix systems $(M^T, I^T), (M'^T, I'^T)$ are shift equivalent, we have $K_i(M^T, I^T) = K_i(M'^T, I'^T)$ for $i = 0, 1$.

Proposition 9.13. *For a topological Markov shift Λ_A determined by an $n \times n$ square matrix A with entries in $\{0, 1\}$, we have*

$$BF^0_K(\Lambda_A) = \mathbb{Z}^n/(1 - A)\mathbb{Z}^n = BF(\Lambda_A), \quad BF^1_K(\Lambda_A) = \mathrm{Ker}(1 - A) \text{ in } \mathbb{Z}^n.$$

Hence the group $BF^1_K(\Lambda)$ is the torsion-free part of the group $BF^0_K(\Lambda)$.

We will finally present the calculation formulae for the Bowen-Franks groups from K. For a subshift Λ, let X^-_Λ be the set of all left-infinite sequences appearing in Λ. That is

$$X^-_\Lambda = \{(..., z_{-2}, z_{-1}, z_0) \in \prod_{i=-\infty}^{0} \Sigma_i | (z_i)_{i\in\mathbb{Z}} \in \Lambda\}.$$

We will define l-future equivalence in the space X^-_Λ in a symmetric way to the previous l-past equivalence. Namely, for $z \in X^-_\Lambda$ and $l \in \mathbb{N}$, put

$$\Lambda^{-l}(z) = \{\mu \in \Lambda^l | z\mu \in X^-_\Lambda\}.$$

Two points $z, w \in X_\Lambda^-$ are said to be *l-future equivalent* if $\Lambda^{-l}(z) = \Lambda^{-l}(w)$. We write this equivalence as $x \sim_{-l} y$. For a fixed $l \in \mathbb{N}$, let $P_i^l, i = 1, 2, \dots, n(l)$ be the set of all l-future equivalence classes of X_Λ^-. We define two rectangular $n(l) \times n(l+1)$ matrices $J_{l,l+1}, B_{l,l+1}$ with entries in $\{0,1\}$ and entries in nonnegative integers similarly to the matrix $I_{l,l+1}, A_{l,l+1}$. Namely, we define

$$J_{l,l+1} \text{ for } \Lambda = I_{l,l+1} \text{ for } \Lambda^T, \qquad B_{l,l+1} \text{ for } \Lambda = A_{l,l+1} \text{ for } \Lambda^T.$$

By [Ma2;Theorem 4.9], we have

THEOREM 9.14.

(i) $BF_K^0(\Lambda) = \varinjlim_l \{J_{l,l+1}^t : \mathbb{Z}^{n(l)}/(J_{l,l+1}^t - B_{l,l+1}^t)\mathbb{Z}^{n(l)}\}$.

(ii) $BF_K^1(\Lambda) = \varinjlim_l \{J_{l,l+1}^t : \mathrm{Ker}(J_{l,l+1}^t - B_{l,l+1}^t) \text{ in } \mathbb{Z}^{n(l)}\}$.

We similarly obtain by Lemma 9.2,

THEOREM 9.15. *The past dimension pair* $(\Delta_{\Lambda^T}, \delta_{\Lambda^T})$ *for subshift* Λ *determines the Bowen-Franks group* $BF_K^i(\Lambda), i = 0, 1$ *from* K *for* Λ,

10. SPECTRUM

It is well-known that the set of all nonzero eigenvalues of a nonnegative matrix M is a shift equivalence invariant. The set of M is called the nonzero spectrum of M and plays an important rôle for studying dynamical properties of the associated topological Markov shift (cf.[LM],[Ki]). In this section, we introduce the notion of spectrum of nonnegative matrix system (A, I). It is an eigenvalue of (A, I) in the sense stated bellow. We denote by $Sp(A, I)$ the set of all eigenvalues of (A, I). As the sequence of the sizes of matrices $A_{l,l+1}, I_{l,l+1}, l \in \mathbb{N}$ are increasing, it seems to be natural to deal with eigenvalues of (A, I) with a certain boundedness condition defined bellow on the corresponding eigenvectors. We denote by $Sp_b(A, I)$ the set of all eigenvalues of (A, I) with the boundedness condition on the corresponding eigenvectors. We will prove that the both of the sets of nonzero spectrum of $Sp(A, I)$ and $Sp_b(A, I)$ are invariant under shift equivalence of (A, I).

We fix a nonnegative matrix system (A, I) throughout this section.

DEFINITION. A sequence $\{v^l\}_{l\in\mathbb{N}}$ of vectors $v^l = (v_1^l, \dots, v_{m(l)}^l) \in \mathbb{C}^{m(l)}, l \in \mathbb{N}$ is called an *I-compatible vector* if it satisfies the conditions:

$$v^l = I_{l,l+1} v^{l+1} \qquad \text{for all} \quad l \in \mathbb{N}. \tag{10.1}$$

An I-compatible vector $\{v^l\}_{l\in\mathbb{N}}$ is said to be nonzero if v^l is a nonzero vector for some l. If $v_i^l \geq 0$ (resp. $v_i^l > 0$) for all $i = 1, \dots, m(l)$ and $l \in \mathbb{N}$, $\{v^l\}_{l\in\mathbb{N}}$ is said to be nonnegative (resp. positive). If there exists a number M such that $\sum_{i=1}^{m(l)} |v_i^l| \leq M$ for all $l \in \mathbb{N}$, $\{v^l\}_{l\in\mathbb{N}}$ is said to be bounded. We remark that, for an I-compatible vector $\{v^l\}_{l\in\mathbb{N}}$, $v^N \neq 0$ for some N implies $v^l \neq 0$ for all $l \geq N$.

Definition. For a complex number β, a nonzero I-compatible vector $\{v^l\}$ is called an *eigenvector* of (A, I) for *eigenvalue* β if it satisfies the conditions:

$$A_{l,l+1}v^{l+1} = \beta v^l \qquad \text{for all} \quad l \in \mathbb{N}. \tag{10.2}$$

An eigenvalue β is said to be bounded if it is an eigenvalue for a bounded eigenvector.

Remark. If a sequence v^l of vectors satisfies the above conditions (10.1),(10.2) for $l = N, N+1, \dots$ for some N, we may extendedly define vectors v^l for $l = 1, \dots, N-1$ for which $\{v^l\}_{l\in\mathbb{N}}$ satisfy the conditions (10.1),(10.2) for all $l \in \mathbb{N}$ by using the condition (10.1).

Definition. Let $Sp^\times(A, I)$ be the set of all nonzero eigenvalues of (A, I) and $Sp_b^\times(A, I)$ the set of all nonzero bounded eigenvalues of (A, I). We call them the *nonzero spectrum* of (A, I) and the *nonzero bounded spectrum* of (A, I) respectively.

We will prove

Theorem 10.1. *If two nonnegative matrix systems are shift equivalent, their nonzero spectrum coincide.*

Proof. Suppose that two nonnegative matrix systems (A, I) and (A', I') are shift equivalent of lag N. Let H_l, K_l be sequences of nonnegative matrices such that $(H, K) : (A, I) \underset{lagN}{\sim} (A', I')$. We will show $Sp^\times(A, I) \subset Sp^\times(A', I')$. For $\beta \in Sp^\times(A, I)$ with nonzero eigenvector v^l, we set $u^l = K_l v^{l+N}$ for $l \in \mathbb{N}$. It is direct to see that

$$u^l = I'_{l,l+1}u^{l+1}, \qquad A'_{l,l+1}u^{l+1} = \beta u^l.$$

Now if the vectors u^l are zero for all $l \geq l_0$ for some l_0, by the equality $H_l K_{l+N} v^{l+2N} = I_{l,l+N} A_{l+N,l+2N} v^{l+2N}$, it follows that

$$0 = A_{l,l+N} I_{l+N,l+2N} v^{l+2N} = A_{l,l+N} v^{l+N} = \beta v^l.$$

Thus $v^l = 0$ for all $l \geq l_0$ and hence for all $l \in \mathbb{N}$, a contradiction. Therefore β is a nonzero eigenvalue of (A', I').

We will next show that the nonzero bounded spectrum of (A, I) is also invariant under shift equivalence. We must provide some lemmas.

Lemma 10.2. *Put* $N_A^l = \max_j \sum_{i=1}^{m(l)} A_{l,l+1}(i,j)$ *for* $l \in \mathbb{N}$. *We have* $N_A^l = N_A^{l+1}$. *That is, the value* N_A^l *does not depend on the choice of* $l \in \mathbb{N}$.

Proof. We note that $\sum_{i=1}^{m(l)} I_{l,l+1}(i,j) = 1$ for each j. It follows that

$$\begin{aligned}\sum_{j=1}^{m(l+1)} A_{l+1,l+2}(j,k) &= \sum_{i=1}^{m(l)} \sum_{j=1}^{m(l+1)} I_{l,l+1}(i,j) A_{l+1,l+2}(j,k) \\ &= \sum_{i=1}^{m(l)} \sum_{p=1}^{m(l+1)} A_{l,l+1}(i,p) I_{l+1,l+2}(p,k).\end{aligned}$$

Hence for $k = 1, \dots, m(l+2)$, there uniquely exists $p_k = 1, \dots, m(l+1)$ such that
$$\sum_{j=1}^{m(l+1)} A_{l+1,l+2}(j,k) = \sum_{i=1}^{m(l)} A_{l,l+1}(i,p_k).$$

This implies the inequality $N_A^{l+1} \le N_A^l$. For $p = 1, \dots, m(l+1)$, take $k_p = 1, \dots, m(l+2)$ with $I_{l+1,l+2}(p,k_p) = 1$. It follows that

$$\begin{aligned}\sum_{i=1}^{m(l)} A_{l,l+1}(i,p) &= \sum_{i=1}^{m(l)} A_{l,l+1}(i,p) I_{l+1,l+2}(p,k_p) \\ &= \sum_{i=1}^{m(l)} \sum_{q=1}^{m(l+1)} A_{l,l+1}(i,q) I_{l+1,l+2}(q,k_p) \\ &= \sum_{i=1}^{m(l)} \sum_{j=1}^{m(l+1)} I_{l,l+1}(i,j) A_{l+1,l+2}(j,k_p) \\ &= \sum_{j=1}^{m(l+1)} A_{l+1,l+2}(j,k_p).\end{aligned}$$

This implies the inequality $N_A^l \le N_A^{l+1}$.

Set $N_A = \max_j \sum_{i=1}^{m(l)} A_{l,l+1}(i,j)$ that is independent of the choice of $l \in \mathbb{N}$. For an I-compatible vector $\{v^l\}_{l\in\mathbb{N}}$, we put $\|v^l\| = \sum_{i=1}^{m(l)} |v_i^l|$.

Lemma 10.3. *The sequence $\{\|v^l\|\}_{l\in\mathbb{N}}$ is increasing. If $\{v^l\}_{l\in\mathbb{N}}$ is nonnegative, $\{\|v^l\|\}_{l\in\mathbb{N}}$ is constant and hence $\{v^l\}_{l\in\mathbb{N}}$ is bounded.*

Proof. We know $\sum_{i=1}^{m(l)} |I_{l,l+1}(i,j) v_j^{l+1}| = |v_j^{l+1}|$ and

$$\|v^l\| \le \sum_{i=1}^{m(l)} \sum_{j=1}^{m(l+1)} |I_{l,l+1}(i,j) v_j^{l+1}| \le \sum_{j=1}^{m(l+1)} |v_j^{l+1}| = \|v^{l+1}\|.$$

If $\{v^l\}_{l\in\mathbb{N}}$ is nonnegative, both of the inequalities above go to equalities.

For a bounded I-compatible vector $v = \{v^l\}_{l\in\mathbb{N}}$, we put

$$\|v\|_1 = \sup_{l\to\infty} \|v^l\|.$$

Proposition 10.4. $Sp_b^\times(A,I) \subset \{z \in \mathbb{C} \,|\, |z| \le N_A\}$.

Proof. For $\beta \in Sp(A,I)$ with a bounded eigenvector $\{v^l\}_{l\in\mathbb{N}}$, we have

$$\beta \sum_{i=1}^{m(l)} |v_i^l| \le \sum_{j=1}^{m(l+1)} (\max_j \sum_{i=1}^{m(l)} A_{l,l+1}(i,j)) |v_j^{l+1}|.$$

Hence we obtain the inequality

$$\beta\|v^l\| \le N_A\|v^{l+1}\|.$$

As $\{v^l\}$ is bounded, the limit $\lim_{l\to\infty}\|v^l\| = \|v\|_1$ exists so that we have a desired assertion.

We denote by $\mathfrak{B}_I$ the set of all bounded I-compatible vectors. It is a complex Banach space with norm $\|\cdot\|_1$. A nonnegative I-compatible vector $v = \{v^l\}_{l\in\mathbb{N}}$ is called a *state* for I if $\|v\|_1 = 1$. Let $\mathfrak{S}_I$ be the set of all states for I. It is a convex subset of $\mathfrak{B}_I$.

Lemma 10.5. *For $v = \{v^l\}_{l\in\mathbb{N}} \in \mathfrak{B}_I$, put*

$$|v|_i^l = \sup_{N\ge l}\sum_{j=1}^{m(N)} I_{l,N}(i,j)|v_j^N| \qquad \text{for} \quad i=1,\dots,m(l), \quad l\in\mathbb{N}.$$

We then have

(i) $|v|_i^l < \infty$.

(ii) *The vectors defined by $|v|^l = (|v|_1^l, |v|_2^l,\dots,|v|_{m(l)}^l)$ for $l\in\mathbb{N}$ give rise to a nonnegative I-compatible vector.*

Proof. (i) By the inequality $\sum_{j=1}^{m(N)} I_{l,N}(i,j)|v_j^N| \le \sum_{j=1}^{m(N)}|v_j^N| = \|v^N\|$, we get $|v|_i^l \le \|v\|_1$.

(ii) As we easily see

$$\sum_{k=1}^{m(N)} I_{l,N}(i,k)|v_k^N| \le \sum_{j=1}^{m(N+1)} I_{l,N+1}(i,j)|v_j^{N+1}|,$$

the sequence of sums $\sum_{j=1}^{m(N)} I_{l,N}(i,j)|v_j^N|$ is increasing on N so that we have

$$|v|_i^l = \lim_{N\to\infty}\sum_{j=1}^{m(N)} I_{l,N}(i,j)|v_j^N|.$$

Hence the following equalities hold

$$\begin{aligned}
\sum_{j=1}^{m(l+1)} I_{l,l+1}(i,j)|v|_j^{l+1} &= \sum_{j=1}^{m(l+1)}\lim_{N\to\infty}\Big(\sum_{k=1}^{m(N)} I_{l,l+1}(i,j)I_{l+1,N}(j,k)|v_k^N|\Big)\\
&= \lim_{N\to\infty}\sum_{k=1}^{m(N)}\sum_{j=1}^{m(l+1)} I_{l,l+1}(i,j)I_{l+1,N}(j,k)|v_k^N|\\
&= \lim_{N\to\infty}\sum_{k=1}^{m(N)} I_{l,N}(i,k)|v_k^N| = |v|_i^l
\end{aligned}$$

so that the vectors $\{|v|^l\}_{l\in\mathbb{N}}$ yield an I-compatible vector.

The I-compatible vector $|v|$ for $v\in\mathfrak{B}_I$ is called the *total variation* of v. A bounded I-compatible vector $v\in\mathfrak{B}_I$ is said to be real if all elements v_i^l of the vectors $v^l, l\in\mathbb{N}$ are real numbers. Thus we obtain

COROLLARY 10.6. *For a real bounded I-compatible vector $v \in \mathfrak{B}_I$, there exist nonnegative bounded I-compatible vectors $v^+, v^- \in \mathfrak{B}_I$ such that*

$$v = v^+ - v^-, \qquad |v| = v^+ + v^-.$$

This decomposition is called the Jordan decomposition of v.

Proof. As $|v|_i^l \geq |v_i^l|$ for each i, l, by putting

$$v^+ = \frac{1}{2}(|v| + v), \qquad v^- = \frac{1}{2}(|v| - v)$$

we get the desired assertions.

COROLLARY 10.7. *For a bounded I-compatible vector $v \in \mathfrak{B}_I$, there exist states $v_j \in \mathfrak{S}_I$ and nonnegative real numbers $c_j \in \mathbb{R}$ such that*

$$v = c_1 v_1 - c_2 v_2 + i(c_3 v_3 - c_4 v_4).$$

PROPOSITION 10.8. *For a bounded I-compatible vector $v \in \mathfrak{B}_I$, we put*

$$(L_A v)_i^l = \sum_{j=1}^{m(l+1)} A_{l,l+1}(i,j) v_j^{l+1} \qquad \text{for} \quad i = 1, \ldots, m(l), \quad l \in \mathbb{N}.$$

Then L_A gives rise to a bounded linear operator on the Banach space $\mathfrak{B}_I$ that satisfies $\|L_A\| = N_A$, where the norm of L_A is given by $\|L_A\| = \sup_{v \neq 0} \frac{\|L_A v\|_1}{\|v\|_1}$.

To prove the proposition above, we note the following lemma.

LEMMA 10.9. *For an arbitrary fixed $l \in \mathbb{N}$ and nonnegative real numbers c_i for $i = 1, \ldots, m(l)$, there exists a nonnegative I-compatible vector $v \in \mathfrak{B}_I$ such that $v_i^l = c_i$ for $i = 1, \ldots, m(l)$.*

Proof. Put $v_i^l = c_i^l$ for $i = 1, \ldots, m(l)$. For $k \leq l$, we put $v^k = I_{k,l} v^l$. For $k = l+1$, we can choose nonnegative real numbers $v_j^{l+1}, j = 1, \ldots, m(l+1)$ such that $v_i^l = \sum_{j=1}^{m(l+1)} v_j^{l+1}$ because for each j there uniquely exists i satisfying $I_{l,l+1}(i,j) = 1$ and $I_{l,l+1}(i',j) = 0$ for other i'. Hence we may get a nonnegative I-compatible vector v by induction such that $v_i^l = c_i, i = 1, \ldots, m(l)$.

Proof of Proposition 10.8. We first show that $L_A v$ is a bounded I-compatible vector. By the relation $I_{l,l+1} A_{l+1,l+2} = A_{l,l+1} I_{l+1,l+2}$, it is direct to see that $L_A v$ is an I-compatible vector. We have $\|(L_A v)^l\| \leq N_A \|v^{l+1}\|$ so that $\|L_A v\|_1 \leq N_A \|v\|_1$. Hence $L_A v$ is bounded and $\|L_A\| \leq N_A$. Fix $l \in \mathbb{N}$. Take i_0 such that $\max_i \sum_{h=1}^{m(l-1)} A_{l-1,l}(h,i) = \sum_{h=1}^{m(l-1)} A_{l-1,l}(h,i_0)$. By the previous lemma, there exists a nonnegative I-compatible vector $v \in \mathfrak{B}_I$ such that $v_{i_0}^l = 1$ and $v_i^l \neq 0$ for $i \neq i_0$. It then follows that

$$\|(L_A v)^{l-1}\| = \sum_{h=1}^{m(l-1)} A_{l-1,l}(h, i_0) = N_A.$$

Thus we get $\|(L_A v)\|_1 = N_A$. As $\|v\|_1 = \|v^l\| = 1$, we conclude $\|L_A\| \geq N_A$ so that $\|L_A\| = N_A$.

Therefore we have

COROLLARY 10.10. *For a complex number β, it belongs to $Sp_b(A, I)$ if and only if it satisfies $L_A v = \beta v$ for some $v \in \mathfrak{B}_I$. That is, the bounded spectrum of (A, I) are nothing but the eigenvalues of the bounded positive operator L_A on the Banach space $\mathfrak{B}_I$.*

Corresponding to Theorem 10.1, we have

THEOREM 10.11. *If two nonnegative matrix systems are shift equivalent, their nonzero bounded spectrum coincide.*

Proof. Suppose that two nonnegative matrix systems (A, I) and (A', I') are shift equivalent of lag N. Let H_l, K_l be sequences of nonnegative matrices such that $(H, K) : (A, I) \underset{lagN}{\sim} (A', I')$. Following the proof of Theorem 10.1, it suffices to show that for a bounded vector $v \in \mathfrak{B}_I$, the vectors defined by $u^l = K_l v^{l+N}, l \in \mathbb{N}$ give rise to a bounded vector. As the equalities $I'_{l,l+1} K_{l+1} = K_l I_{l+N,l+N+1}$ hold, the boundedness of the vector $\{u^l\}_{l \in \mathbb{N}}$ is shown by a similar manner to the proof of the boundedness of the vector $L_A v$ as in the proof of Proposition 10.8. Hence we know $Sp_b^\times(A, I) = Sp_b^\times(A', I')$.

We will next see that the set $Sp_b^\times(A, I)$ is not empty. We will consider another topology on $\mathfrak{B}_I$. The topology is defined from the subbases of open sets of the form:

$$U_l(v, i, \epsilon) = \{u \in \mathfrak{B}_I \mid |v_i^l - u_i^l| < \epsilon\} \quad \text{for} \quad v \in \mathfrak{B}_I, i = 1, \ldots, m(l), \epsilon > 0, l \in \mathbb{N}.$$

We call it the weak topology on $\mathfrak{B}_I$. It is straightforward to see that the state space $\mathfrak{S}_I$ is compact in the topology. Let $\sigma(L_A)$ be the set of all spectrum of L_A as a bounded linear operator on the Banach space $\mathfrak{B}_I$. General theory of bounded linear operators tells us that the set $\sigma(L_A)$ is not empty. Let r_A be the spectral radius of the operator L_A on $\mathfrak{B}_I$, that is, $r_A = \sup\{|r| : r \in \sigma(L_A)\}$.

PROPOSITION 10.12. *There exists a state $v \in \mathfrak{S}_I$ such that $L_A v = r_A v$. Hence we have $r_A \in Sp_b^\times(A, I)$.*

Our proof is completely similar to the proof of [MWY;Lemma 4.1]. We will give a proof for the sake of completeness.

Proof. Let $R_A(z)$ be the resolvent of L_A that is defined by $R_A(z)v = (z - L_A)^{-1} v$ for $z \in \mathbb{C}$ with $|z| > r_A$ and $v \in \mathfrak{B}_I$. For $z \in \mathbb{C}$ with $|z| > r_A$, we see $R_A(z)v = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} L_A^k(v)$ and

$$|(L_A^k(v)_i^l| \le \sum_{j=1}^{m(l+k)} A_{l,l+k}(i,j) |v_j^{l+k}|.$$

As $|v_j^{l+k}| \le |v|_j^{l+k}$, it follows that $|(R_A(z)v)_i^l| \le (R_A(|z|)|v|)_i^l$ and hence

$$\|R_A(z)v\|_1 \le \|R_A(|z|)|v|\|_1. \tag{10.3}$$

Since $\{R_A(z)\}_{|z|>r_A}$ can not be uniformly bounded in the set $\mathcal{L}(\mathfrak{B}_I)$ of all bounded linear operators on $\mathfrak{B}_I$, by the inequality (10.3) we may find $v_0 \in \mathfrak{S}_I$ so that $\|R_A(t)v_0\|_1$ is unbounded for $t \downarrow r_A$. Put

$$v_n = \frac{R_A(r_A + \frac{1}{n})v_0}{\|R_A(r_A + \frac{1}{n})v_0\|} \qquad \text{for} \qquad n = 1, 2, \dots.$$

As L_A is a positive operator on $\mathfrak{B}_I$, the operator $R_A(t)$ is also positive so that the vectors $v_n, n = 1, 2, \dots$ are states. Hence there exists a limit point v_∞ of the sequence $\{v_n\}$ in $\mathfrak{S}_I$ in the weak topology of $\mathfrak{S}_I$. The following identity

$$(r_A - L_A)v_n = -\frac{1}{n}v_n + \frac{v_0}{\|R_A(r_A + \frac{1}{n})v_0\|}$$

implies $r_A v_\infty = L_A v_\infty$. As (A, I) is essential, the vector $L_A v_\infty$ can not be zero. Hence we have $r_A > 0$ and $r_A \in Sp_b^\times(A, I)$.

The author would like to thank Yasuo Watatani for pointing out an inaccuracy of a proof of the proposition above given in an earlier version of this paper.

We finally show that the spectrum are majorizied by topological entropy of the associated subshift. It is well-known that topological entropy $h_{\mathrm{top}}(\Lambda)$ for subshift Λ is given by

$$h_{\mathrm{top}}(\Lambda) = \lim_{k\to\infty} \frac{1}{k} \log \sharp|\Lambda^k|$$

where $\sharp|\Lambda^k|$ denotes the cardinality of the set of all admissible words of length k in the subshift Λ (cf.[LM],[Ki]).
We say a symbolic matrix system $(\mathcal{M}, I)$ to be left resolving if a symbol appearing in $\mathcal{M}(i,j)$ can not appear in $\mathcal{M}(i',j)$ for other $i' \neq i$, equivalently, its λ-graph system is left resolving. As in Proposition 3.8, a canonical symbolic matrix system is left resolving.

Proposition 10.13. *Let $(\mathcal{M}, I)$ be a left resolving symbolic matrix system and (M, I) its associated nonnegative matrix system. For any $\beta \in Sp_b(M, I)$, we have the inequalities:*

$$\log|\beta| \le \log r_M \le h_{\mathrm{top}}(\Lambda_{(\mathcal{M},I)})$$

where r_M is the spectral radius of the operator L_M on $\mathfrak{B}_I$ and $\Lambda_{(\mathcal{M},I)}$ is the associated subshift with $(\mathcal{M}, I)$.

Proof. The inequality $\log|\beta| \le \log r_M$ is clear. By the previous lemma, take $v \in \mathfrak{S}_I$ such that $L_M v = r_M v$. We have for $k \in \mathbb{N}$,

$$r_M^k v_i^1 = \sum_{j=1}^{m(k+1)} M_{1,k+1}(i,j) v_j^{k+1}.$$

As $\sum_{i=1}^{m(1)} v_i^1 = 1$, it follows that

$$r_M^k \leq (\max_j \sum_{i=1}^{m(1)} M_{1,k+1}(i,j)) \sum_{j=1}^{m(k+1)} v_j^{k+1} = \|L_M^k\|.$$

We may find j_0 such that $\|L_M^k\| = \sum_{i=1}^{m(1)} M_{1,k+1}(i,j_0)$. Since $(\mathcal{M}, I)$ is left resolving, the number $\sum_{i=1}^{m(1)} M_{1,k+1}(i,j_0)$ is majorized by the cardinality $\sharp|\Lambda^k_{(\mathcal{M},I)}|$ of the set of all admissible words of length k in the subshift $\Lambda_{(\mathcal{M},I)}$. Thus we obtain the inequalities

$$r_M^k \leq \|L_M^k\| \leq \sharp|\Lambda^k_{(\mathcal{M},I)}|.$$

As $\|L_M^k\|^{\frac{1}{k}} \to r_M$ for $k \to \infty$, we have desired inequalities.

For subshift (Λ, σ), let (M, I) be its canonical nonnegative matrix system. We define the nonzero spectrum $Sp^\times(\Lambda)$ and the nonzero bounded spectrum $Sp_b^\times(\Lambda)$ of Λ by the nonzero spectrum and the nonzero bounded spectrum of (M, I) respectively. We have thus proved

Theorem 10.14. *Both the sets $Sp^\times(\Lambda)$ and $Sp_b^\times(\Lambda)$ are not empty and topological conjugacy invariants of subshifts. In particular, $Sp_b^\times(\Lambda)$ is bounded by the topological entropy of the subshift (Λ, σ).*

11. Example

We will give an example of the canonical symbolic matrix system, the K-groups and the Bowen-Franks groups for a certain nonsofic subshift, that is called the context free shift in [LM]. Let Σ be the set of symbols $\{a, b, c\}$. The nonsofic subshift is defined to be the subshift Z over Σ whose forbidden words are

$$\mathcal{F}_Z = \{ab^m c^k a \mid m \neq k\}$$

where the word $ab^m c^k a$ means $a \underbrace{b\cdots b}_{m \text{ times}} \underbrace{c\cdots c}_{k \text{ times}} a$ (cf.[LM]). In [Ma6], the C^*-algebra $\mathcal{O}_Z$ associated with the subshift Z has been studied so that its K-groups has been calculated. By using discussions of the computation of the K-groups, we may write the canonical symbolic matrix system for Z. Let X_Z be the corresponding one-sided subshift for Z. Define sequences of subsets of X_Z in the following way.

$$P_0 = \{c^k b^\infty \mid k \geq 0\} \cup \{b^k c^m by \in X_Z \mid k \geq 0, m \geq 1, y \in X_Z\}$$

and for $n, j = 0, 1, \ldots,$

$$\begin{aligned} E_j &= \{c^j ay \in X_Z \mid y \in X_Z\}, \\ Q_n &= \cup_{j>n} E_j, \\ F_j &= \{b^m c^{m+j} ay \in X_Z \mid m \geq 1, y \in X_Z\}, \\ R_n &= \{b^m c^k ay \in X_Z \mid m \geq 1, k \geq 0, m + j \neq k \text{ for } j = 0, 1, \ldots, n\}. \end{aligned}$$

LEMMA 11.1([Ma6;Lemma 4.3]). *For each $l \in \mathbb{N}$, the space X_Z is decomposed into the disjoint union:*

$$X_Z = P_0 \cup_{j=0}^{l-1} E_j \cup Q_{l-1} \cup_{j=0}^{l-1} F_j \cup R_{l-1}.$$

This decomposition of X_Z into $2l+3$-components corresponds to the l-past equivalence classes of X_Z.

The canonical symbolic matrix systems $\mathcal{M}_{l,l+1}, I_{l,l+1}$ for Z are $m(l)(=2l+3) \times m(l+1)(=2l+5)$ matrices that are written as follows:

$$\mathcal{M}_{l,l+1} =$$

$$\begin{bmatrix}
b+c & & c & & c & & c & & c & & c & & \cdots & & \cdots & & c \\
a & a & a & & & & & & & & & & & & & & \\
 & & & b & b & & & & & & & & & & & & \\
 & c & & & & & & & & & & & & & & & \\
 & & & & & b & b & & & & & & & & & & \\
 & & & c & & & & & & & & & & & & & \\
 & & & & & & & b & b & & & & & & & & \\
 & & & & & c & & & & & & & & & & & \\
 & & & & & & & & & \ddots & \ddots & & & & & & \\
 & & & & & & & \ddots & & & & & & & & & \\
 & & & & & & & & & & & \ddots & \ddots & & & & \\
 & & & & & & & & & \ddots & & & & & & & \\
 & & & & & & & & & & & & & b & b & & \\
 & & & & & & & & & & & c & & c & & c & \\
 & b & b & & & & & & & & & & & & & b & b
\end{bmatrix},$$

$$I_{l,l+1} = \begin{bmatrix}
1 & & & & & & & & & & & & & & & & \\
 & 1 & & & & & & & & & & & & & & & \\
 & & 1 & & & & & & & & & & & & & & \\
 & & & 1 & & & & & & & & & & & & & \\
 & & & & 1 & & & & & & & & & & & & \\
 & & & & & 1 & & & & & & & & & & & \\
 & & & & & & 1 & & & & & & & & & & \\
 & & & & & & & 1 & & & & & & & & & \\
 & & & & & & & & 1 & & & & & & & & \\
 & & & & & & & & & 1 & & & & & & & \\
 & & & & & & & & & & 1 & & & & & & \\
 & & & & & & & & & & & \ddots & & & & & \\
 & & & & & & & & & & & & \ddots & & & & \\
 & & & & & & & & & & & & & 1 & & 1 & \\
 & & & & & & & & & & & & & & 1 & & 1
\end{bmatrix}$$

along the following ordered basis

$$P_0, E_0, F_0, E_1, F_1, \dots, E_{l-1}, F_{l-1}, Q_{l-1}, R_{l-1}$$

where in the matrices above, blanks denote zeros. The transposed matrices of its nonnegative matrix systems are written as:

$$A^t_{l,l+1} = \begin{bmatrix}
2 & 1 & & & & & & & & & & & & & \\
 & 1 & & 1 & & & & & & & & & & & 1 \\
1 & 1 & & & & & & & & & & & & & 1 \\
 & & 1 & & & 1 & & & & & & & & & \\
1 & & 1 & & & & & & & & & & & & \\
 & & & & 1 & & & 1 & & & & & & & \\
1 & & & & 1 & & & & & & & & & & \\
 & & & & & & 1 & & & \ddots & & & & & \\
1 & & & & & & 1 & & & & & & & & \\
 & & & & & & & & 1 & & & \ddots & & & \\
1 & & & & & & & & 1 & & & & & & \\
 & & & & & & & & & & \ddots & & & 1 & \\
\vdots & & & & & & & & & & \ddots & & & & \\
 & & & & & & & & & & & & 1 & 1 & \\
\vdots & & & & & & & & & & & & 1 & & \\
 & & & & & & & & & & & & & 1 & 1 \\
1 & & & & & & & & & & & & & & 1
\end{bmatrix},$$

$$I^t_{l,l+1} = \begin{bmatrix}
1 & & & & & & & & & & & & & & \\
 & 1 & & & & & & & & & & & & & \\
 & & 1 & & & & & & & & & & & & \\
 & & & 1 & & & & & & & & & & & \\
 & & & & 1 & & & & & & & & & & \\
 & & & & & 1 & & & & & & & & & \\
 & & & & & & 1 & & & & & & & & \\
 & & & & & & & 1 & & & & & & & \\
 & & & & & & & & 1 & & & & & & \\
 & & & & & & & & & 1 & & & & & \\
 & & & & & & & & & & 1 & & & & \\
 & & & & & & & & & & & \ddots & & & \\
 & & & & & & & & & & & & \ddots & & \\
 & & & & & & & & & & & & & 1 & \\
 & & & & & & & & & & & & & & 1 \\
 & & & & & & & & & & & & & 1 & \\
 & & & & & & & & & & & & & & 1
\end{bmatrix}.$$

Hence we have

Proposition 11.2.

$$K_0(Z) = \mathbb{Z}, \qquad K_1(Z) = 0 \quad \textit{and} \quad BF^0(Z) = 0, \qquad BF^1(Z) = \mathbb{Z}.$$

Since the subshift Z is conjugate to its transpose Z^T and by the formula for the Bowen-Franks groups from K for subshifts, we obtain

Proposition 11.3.

$$BF^0_K(Z) = \mathbb{Z}, \qquad BF^1_K(Z) = 0.$$

Hence these types of the Bowen-Franks groups can not be realized in sofic subshifts because $BF^1(Z)$ (resp. $BF^1_K(Z)$) is not the torsion-free part of $BF^0(Z)$ (resp. $BF^0_*(Z)$). We finally see

Proposition 11.4([Ma6:Theorem 6.9]). *The spectral radius of the operator* L_A *is* $1+\sqrt{1+\sqrt{3}} = 2.65289\cdots$ *that is the topological entropy for the subshift* Z. *Hence the maximum value of* $Sp_b^\times(A, I)$ *is* $1+\sqrt{1+\sqrt{3}}$.

In [KMW], the K-groups and the dimension groups for β-shifts have been calculated. The K-groups and the Bowen-Franks groups for the Dyck shifts are also calculated in [Ma7].

References

[BK] M. Boyle and W. Krieger, *Almost Markov and shift equivalent sofic systems*, Proceedings of Maryland Special Year in Dynamics 1986-87, Springer -Verlag Lecture Notes in Math **1342** (1988), 33–93.

[BF] R. Bowen and J. Franks, *Homology for zero-dimensional nonwandering sets*, Ann. Math. **106** (1977), 73–92.

[Bra] O. Bratteli, *Inductive limits of finite-dimensional C^*-algebras*, Trans. Amer. Math. Soc. **171** (1972), 195–234.

[Bro] L. G. Brown, *The universal coefficient theorem for Ext and quasidiagonality*, Operator Algebras and Group Representation, Pitmann Press **17** (1983), 60–64.

[C] J. Cuntz, *Simple C^*-algebras generated by isometries*, Comm. Math. Phys. **57** (1977), 173–185.

[C2] J. Cuntz, *A class of C^*-algebras and topological Markov chains II: reducible chains and the Ext- functor for C^*-algebras*, Inventions Math. **63** (1980), 25–40.

[CK] J. Cuntz and W. Krieger, *A class of C^*-algebras and topological Markov chains*, Inventions Math. **56** (1980), 251–268.

[DGS] M. Denker, C. Grillenberger and K. Sigmund, *Ergodic theory on compact spaces*, Springer-Verlag, Berlin, Heidelberg and New York, 1976.

[Ef] E. G. Effros, *Dimensions and C^*-algebras*, AMS-CBMS Reg. Conf. vol 46, Providence, 1981.

[El] G. A. Elliott, *On the classification of inductive limits of sequences of semisimple finite-dimensional algebras*, J. Algebra **38** (1976), 29–44.

[EFW] M. Enomoto, M. Fujii and Y. Watatani, *Tensor algebras on the sub-Fock space associated with $\mathcal{O}_A$*, Math. Japon **26** (1981), 171–177.

[Ev] D. E. Evans, *The C^*-algebras of topological Markov chains*, Tokyo Metropolitan University Lecture Note, 1982.

[Fr] J. Franks, *Flow equivalence of subshifts of finite type*, Ergod. Th. & Dynam. Sys. **4** (1984), 53–66.

[HN] T. Hamachi and M. Nasu, *Topological conjugacy for* 1*-block factor maps of subshifts and sofic covers*, Proceedings of Maryland Special Year in Dynamics 1986-87, Springer -Verlag Lecture Notes in Math **1342** (1988), 251–260.

[KKS] D. S. Kahn, J. Kaminker and C. Schochet, *Generalized homology theories on compact metric spaces*, Michigan Math. J. **24** (1977), 203–224.

[KP] J. Kaminker and I. Putnam, *K-theoretic duality for shifts of finite type*, preprint..

[Ka] G. G. Kasparov, *The operator K-functor and extensions of C^*-algebras*, Math. USSR. Izvestijia **16** (1981), 513–572.

[KMW] Y. Katayama, K. Matsumoto and Y. Watatani, *Simple C^*-algebras arising from β-expansion of real numbers*, Ergod.Th. & Dynam. Sys. **18** (1998), 937–962.

[KimR] K. H. Kim and F. W. Roush, *Some results on decidability of shift equivalence*, J. combinatorics, Info. Sys.Sci. **4** (1979), 123–146.

[KimR2] K. H. Kim and F. W. Roush, *Williams conjecture is false for irreducible subshifts*, preprint..

[Ki] E. Kirchberg, *The classification of purely infinite C^*-algebras using Kasparov's theory*, preprint. 1994.

[Kit] B. P. Kitchens, *Symbolic dynamics*, Springer-Verlag, Berlin, Heidelberg and New York, 1998.

[Kr] W. Krieger, *On dimension functions and topological Markov chains*, Invent. Math. **56** (1980), 239–250.

[Kr2] W. Krieger, *On dimension for a class of homeomorphism groups*, Math. Ann **252** (1980), 87–95.

[Kr3] W. Krieger, *On sofic systems I*, Israel J. Math. **48** (1984), 305–330.

[Kr4] W. Krieger, *On syntactically defined invariant of symbolic dynamics*, to appear in Ergod. Th. & Dynam. Sys..

[KPRR] A. Kumjian, D. Pask, I. Raeburn and J. Renault, *Graphs, groupoids and Cuntz-Krieger algebras*, J. Funct. Anal. **144** (1997), 505–541.

[Le] J. Lee, *Equivalences in subshifts*, J. Korean Math. Soc. **33** (1996), 685–692.

[LM] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press., 1995.

[Ma] K. Matsumoto, *On C^*-algebras associated with subshifts*, Internat. J. Math. **8** (1997), 357–374.

[Ma2] K. Matsumoto, *K-theory for C^*-algebras associated with subshifts*, Math. Scand. **82** (1998), 237–255.

[Ma3] K. Matsumoto, *Dimension groups for subshifts and simplicity of the associated C^*-algebras*, J. Math. Soc. Japan **51** (1999), 679–698.

[Ma4] K. Matsumoto, *Bowen-Franks groups for subshifts and Ext-groups for C^*-algebras*, preprint, 1997.

[Ma5] K. Matsumoto, *Stabilized C^*-algebras constructed from symbolic dynamical systems*, to appear in Ergod. Th. and Dyn. Sys..

[Ma6] K. Matsumoto, *A simple C^*-algebra arising from certain subshift*, to appear in J. Operator Theory.

[Ma7] K. Matsumoto, *K-theoretic invariants and conformal measures on the Dyck shifts*, preprint, 1999.

[MWY] K. Matsumoto, Y. Watatani and M. Yoshida, *KMS-states for gauge actions on C^*-algebras associated with subshifts*, Math. Z. **228** (1998), 489–509.

[N] M. Nasu, *Topological conjugacy for sofic shifts*, Ergod. Th. & Dynam. Sys. **6** (1986), 265–280.

[N2] M. Nasu, *Textile systems for endomorphisms and automorphisms of the shift*, Mem. Amer. Math. Soc. **546** (1995).

[Pa] W. Parry, *On the β-expansion of real numbers*, Acta Math. Acad. Sci. Hung. **11** (1960), 401–416.

[PS] W. Parry and D. Sullivan, *A topological invariant for flows on one-dimensional spaces*, Topology **14** (1975), 297–299.

[Ph] N. C. Phillips, *A classification theorem for nuclear purely infinite simple C^*-algebras*, preprint. 1995.

[Re] A. Rényi, *Representations for real numbers and their ergodic properties*, Acta Math. Acad. Sci. Hung **8** (1957), 477–493.

[RS] J. Rosenberg and C. Schochet, *The Künneth theorem and the universal coefficient theorem for Kasparov's generalized K-functor*, Duke Math. J. **55** (1987), 431–474.

[Tu] S.Tuncel, *A dimension, dimension modules, and Markov chains*, Proc. London Math. Soc. **46** (1983), 100–116.

[We] B. Weiss, *Subshifts of finite type and sofic systems*, Monats. Math. **77** (1973), 462–474.

[Wi] R. F. Williams, *Classification of subshifts of finite type*, Ann. Math. **98** (1973), 120–153, erratum, Ann. Math. $99(1974), 380-381$.

Kengo Matsumoto
Department of Mathematics
Joetsu University of Education
Joetsu 943-8512
Japan
matsu@juen.ac.jp

Random Matrices and K-Theory for Exact C^*-Algebras

U. Haagerup[1 2] and S. Thorbjørnsen[1 2]

Received: June 18, 1999

Communicated by Joachim Cuntz

Abstract. In this paper we find asymptotic upper and lower bounds for the spectrum of random operators of the form

$$S^*S = \Big(\sum_{i=1}^{r} a_i \otimes Y_i^{(n)}\Big)^* \Big(\sum_{i=1}^{r} a_i \otimes Y_i^{(n)}\Big),$$

where $a_1, \dots, a_r$ are elements of an exact C^*-algebra and $Y_1^{(n)}, \dots, Y_r^{(n)}$ are complex Gaussian random $n \times n$ matrices, with independent entries. Our result can be considered as a generalization of results of Geman (1981) and Silverstein (1985) on the asymptotic behavior of the largest and smallest eigenvalue of a random matrix of Wishart type. The result is used to give new proofs of:

(1) Every stably finite exact unital C^*-algebra $\mathcal{A}$ has a tracial state.

(2) If $\mathcal{A}$ is an exact unital C^*-algebra, then every state on $K_0(\mathcal{A})$ is given by a tracial state on $\mathcal{A}$.

The new proofs do not rely on quasitraces or on AW^*-algebra techniques.

1991 Mathematics Subject Classification: Primary 46L05; Secondary 46L50, 46L35, 46L80, 60F15.

Introduction

Following the terminology in [HT], we let $\mathrm{GRM}(m, n, \sigma^2)$ denote the class of $m \times n$ random matrices $B = (b_{ij})_{1\le i\le m,\ 1\le j\le n}$, for which $\big(\mathrm{Re}(b_{ij}), \mathrm{Im}(b_{ij})\big)_{1\le i\le m,\ 1\le j\le n}$ form a set of $2mn$ independent Gaussian random variables, all with mean 0 and variance $\frac{1}{2}\sigma^2$. In other words, the

[1]Department of Mathematics and Computer Science, SDU, Odense University, Denmark.

[2]MaPhySto - Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation.

entries of B are mn independent complex random variables with distribution measure on $\mathbb{C}$ given by

$$\tfrac{1}{\pi\sigma^2}\exp(-\tfrac{|z|^2}{\sigma^2})\; d\mathrm{Re}(z)\; d\mathrm{Im}(z).$$

The theory of exact C^*-algebras has been developed by Kirchberg (see [Ki1], [Ki2], [Ki3], [Was] and references given there). A C^*-algebra $\mathcal{A}$ is exact, if for all pairs $(\mathcal{B},\mathcal{J})$, of a C^*-algebra $\mathcal{B}$ and a closed two-sided ideal $\mathcal{J}$ in $\mathcal{B}$, the sequence

$$0 \longrightarrow \mathcal{A} \underset{\min}{\otimes} \mathcal{J} \longrightarrow \mathcal{A} \underset{\min}{\otimes} \mathcal{B} \longrightarrow \mathcal{A} \underset{\min}{\otimes} \big(\mathcal{B}/\mathcal{J}\big) \longrightarrow 0$$

is exact. Here, for any C^*-algebras $\mathcal{C}$ and $\mathcal{D}$, $\mathcal{C} \otimes_{\min} \mathcal{D}$ means the completion of the algebraic tensor product $\mathcal{C} \odot \mathcal{D}$ in the minimal (=spatial) tensor norm. Sub-algebras and quotients of exact C^*-algebras are again exact (cf. e.g. [Was, 2.5.2 and Corollary 9.3]), and the class of exact C^*-algebras contains most of the C^*-algebras of current interest, such as all nuclear C^*-algebras, and the non-nuclear reduced group C^*-algebras $C^*_r(\mathbb{F}_n)$, associated with the free group $\mathbb{F}_n$ on n generators $(2 \le n \le \infty)$.

For any element T of a unital C^*-algebra, we let $\mathrm{sp}(T)$ denote the spectrum of T. The main result of this paper is

0.1 Main Theorem. Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$, such that $\{a_i^*a_j \mid 1 \le i,j \le r\}$ is contained in an exact C^*-subalgebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{H})$. Let $(\Omega,\mathcal{F},P)$ be a fixed probability space, and let, for each n in $\mathbb{N}$, $Y_1^{(n)},\dots,Y_r^{(n)}$ be independent Gaussian random matrices on Ω in the class $\mathrm{GRM}(n,n,\frac{1}{n})$. Put

$$S_n = \sum_{i=1}^{r} a_i \otimes Y_i^{(n)}, \qquad (n \in \mathbb{N}),$$

and let c, d be positive real numbers. We then have

(i) If $\|\sum_{i=1}^r a_i^*a_i\| \le c$ and $\|\sum_{i=1}^r a_ia_i^*\| \le d$, then for almost all ω in Ω,

$$\limsup_{n\to\infty} \max\big[\mathrm{sp}\big(S_n^*(\omega)S_n(\omega)\big)\big] \le \big(\sqrt{c}+\sqrt{d}\big)^2.$$

(ii) If $\sum_{i=1}^r a_i^*a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$, $\|\sum_{i=1}^r a_ia_i^*\| \le d$, and $d \le c$, then for almost all ω in Ω,

$$\liminf_{n\to\infty} \min\big[\mathrm{sp}\big(S_n^*(\omega)S_n(\omega)\big)\big] \ge \big(\sqrt{c}-\sqrt{d}\big)^2. \qquad \square$$

The Main Theorem can be considered as a generalization of the results of Geman (cf. [Gem]) and Silverstein (cf. [Si]), on the asymptotic behavior of the largest and smallest eigenvalues of a random matrix of Wishart type (see also [BY], [YBK] and [HT]).

The Main Theorem has the following two immediate consequences:

0.2 Corollary. *Let $a_1, \dots, a_r$ be elements of an exact C^*-algebra $\mathcal{A}$, and for each n in $\mathbb{N}$, let $Y_1^{(1)}, \dots, Y_r^{(n)}$ be independent elements of* $\mathrm{GRM}(n, n, \frac{1}{n})$. *Then*

$$\limsup_{n\to\infty} \Big\| \sum_{i=1}^{r} a_i \otimes Y_i^{(n)}(\omega) \Big\| \le \Big\| \sum_{i=1}^{r} a_i^* a_i \Big\|^{\frac{1}{2}} + \Big\| \sum_{i=1}^{r} a_i a_i^* \Big\|^{\frac{1}{2}},$$

for almost all ω in Ω. □

0.3 Corollary. *Let $a_1, \dots, a_r$ and S_n, $n \in \mathbb{N}$, be as in the Main Theorem, and assume that $\sum_{i=1}^{r} a_i^* a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$ and $\| \sum_{i=1}^{r} a_i^* a_i \| \le d$, for some positive real numbers c, d, such that $d < c$. Then for almost all ω in Ω,*

$$0 \notin \mathrm{sp}\big(S_n^*(\omega) S_n(\omega)\big), \quad \textit{eventually as } n \to \infty. \qquad \square$$

In a subsequent paper [Th] by the second named author, it is proved, that if $a_1, \dots, a_r$ and S_n, $n \in \mathbb{N}$, are as in the Main Theorem, and if furthermore $\sum_{i=1}^{r} a_i^* a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$ and $\sum_{i=1}^{r} a_i a_i^* = d\mathbf{1}_{\mathcal{B}(\mathcal{K})}$, for some positive real numbers c, d, then

$$\lim_{n\to\infty} \max\big[\mathrm{sp}(S_n^* S_n)\big] = \big(\sqrt{c} + \sqrt{d}\big)^2, \qquad \text{almost surely,}$$

and if $c \ge d$, then

$$\lim_{n\to\infty} \min\big[\mathrm{sp}(S_n^* S_n)\big] = \big(\sqrt{c} - \sqrt{d}\big)^2, \qquad \text{almost surely.}$$

Hence the asymptotic upper and lower bounds in the Main Theorem cannot, in general, be improved.

Exactness is essential both for the Main Theorem and for the corollaries. An example of violation of the upper bound in the Main Theorem is given in Section 4. The example is based on the non-exact full C^*-algebra $C^*(\mathbb{F}_r)$ associated with the free group on r generators, for $r \ge 6$.

In [Haa], the first named author proved that bounded quasitraces on exact C^*-algebras are traces. Together with results of Handelman (cf. [Han]) and Blackadar and Rørdam (cf. [BR]), this result implies

(1) Every stably-finite exact unital C^*-algebra has a tracial state.

(2) If $\mathcal{A}$ is an exact unital C^*-algebra, then every state on $K_0(\mathcal{A})$ is given by a tracial state on $\mathcal{A}$.

The proof in [Haa] of the above mentioned quasitrace result, relies heavily on ultra product techniques for AW^*-algebras, but the starting point of the proof in [Haa] is the following fairly simple observation: Let $a_1, \dots, a_r$ be r elements in a (not necessarily exact) C^*-algebra $\mathcal{A}$, such that $\sum_{i=1}^{r} a_i^* a_i = \mathbf{1}_{\mathcal{A}}$ and $\| \sum_{i=1}^{r} a_i a_i^* \| < 1$. Let further $x_1, \dots, x_r$ be a semi-circular system (in the sense of Voiculescu; cf. [Vo2]) in some C^*-probability space $(\mathcal{B}, \psi)$. Then the operator $s = \sum_{i=1}^{r} a_i \otimes x_i$ in $\mathcal{A} \otimes C^*(x_1, \dots, x_r, \mathbf{1}_{\mathcal{B}})$, satisfies $0 \notin \mathrm{sp}(s^* s)$ but

$0 \in \mathrm{sp}(ss^*)$, and this implies that $u = s(s^*s)^{-\frac{1}{2}}$ is a non-unitary isometry in the C^*-algebra $\mathcal{A} \otimes C^*(x_1, \dots, x_r, \mathbf{1}_{\mathcal{B}})$.
Corollary 0.3 can be viewed as a random matrix version of the result that $0 \notin \mathrm{sp}(s^*s)$. The corresponding random matrix version of the result that $0 \in \mathrm{sp}(ss^*)$, holds too, i.e., if $a_1, \dots, a_r$ and S_n, $n \in \mathbb{N}$, are as in Corollary 0.3, then with probability 1, $0 \in \mathrm{sp}(S_n S_n^*)$, eventually as $n \to \infty$ (cf. [Th]). In view of Voiculescu's random matrix model for a semi-circular system (cf. [Vo1, Theorem 2.2]), it would have been more natural to substitute $Y_1^{(n)}, \dots, Y_r^{(n)}$ from $\mathrm{GRM}(n, n, \frac{1}{n})$, with a set of independent, *selfadjoint* Gaussian random matrices. However, we found it more tractable to work with the non-selfadjoint random matrices $Y_1^{(n)}, \dots, Y_r^{(n)}$.
In the last section (Section 9), we use Corollary 0.3 to give a new proof of the statements (1) and (2) above. The new proof does not rely on quasitraces or AW^*-algebra techniques. The main step in the new proof of (1) and (2) is to prove, that Corollary 0.3 implies the following

0.4 Proposition. *Let p, q be projections in an exact C^*-algebra $\mathcal{A}$, and assume that there exists an ϵ in $]0,1[$, such that*

$$\tau(q) \le (1-\epsilon)\tau(p),$$

for all lower semi-continuous (possibly unbounded) traces $\tau\colon \mathcal{A}_+ \to [0,\infty]$. Then for some n in $\mathbb{N}$, there exists a partial isometry u in $M_n(\mathcal{A}) = \mathcal{A} \otimes M_n(\mathbb{C})$, such that

$$u^*u = q \otimes \mathbf{1}_{M_n(\mathbb{C})} \qquad \textit{and} \qquad uu^* \le p \otimes \mathbf{1}_{M_n(\mathbb{C})}. \qquad \square$$

In the rest of this introduction, we shall briefly discuss the main steps of the proof of the Main Theorem. Observe first, that by a simple scaling argument, it is enough to treat the case $d = 1$. This normalization will be used throughout the paper. The proof of the Main Theorem relies on the following

0.5 Key Estimates. Let $a_1, \dots, a_r$ be elements of $\mathcal{B}(\mathcal{H}, \mathcal{K})$, let c be a positive constant, and put $S_n = \sum_{i=1}^r a_i \otimes Y_i^{(n)}$, $n \in \mathbb{N}$, as in the Main Theorem. We then have

(a) If $\|\sum_{i=1}^r a_i^* a_i\| \le c$ and $\|\sum_{i=1}^r a_i a_i^*\| \le 1$, then for $0 \le t \le \min\{\frac{n}{2c}, \frac{n}{2}\}$,

$$\mathbb{E}\big[\exp(tS_n^*S_n)\big] \le \exp\big((\sqrt{c}+1)^2 t + (c+1)^2 \tfrac{t^2}{n}\big)\mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}. \tag{0.1}$$

(b) If $\sum_{i=1}^r a_i^* a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$, $\sum_{i=1}^r a_i a_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{K})}$ and $c \ge 1$, then for $0 \le t \le \frac{n}{2c}$,

$$\mathbb{E}\big[\exp(-tS_n^*S_n)\big] \le \exp\big(-(\sqrt{c}-1)^2 t + (c+1)^2 \tfrac{t^2}{n}\big)\mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}. \tag{0.2}$$

$\square$

We emphasize that the key estimates (0.1) and (0.2) hold without the exactness assumption of the Main Theorem. Once these estimates are proved, a fairly simple application of the Borel-Cantelli Lemma yields, that if $\mathcal{H}$ is *finite dimensional*, and $\lambda_{\max}$ and $\lambda_{\min}$ denote largest and smallest eigenvalues, then one has

$$\limsup_{n\to\infty} \lambda_{\max}(S_n^* S_n) \le \big(\sqrt{c}+1\big)^2, \qquad \text{almost surely,}$$

in the situation of (a) above, and

$$\liminf_{n\to\infty} \lambda_{\min}(S_n^* S_n) \ge \big(\sqrt{c}-1\big)^2, \qquad \text{almost surely,}$$

in the situation of (b) above. (This is completely parallel to the proof of the complex version of the Geman-Silverstein result, given in [HT, Section 7]). To pass from the case $\dim(\mathcal{H}) < \infty$ to the case $\dim(\mathcal{H}) = +\infty$, we need the assumption that the C^*-algebra $C^*(\{a_i^* a_j \mid 1 \le i,j \le r\})$ is exact, as well as the following characterization of exact C^*-algebras, due to Kirchberg (cf. [Ki2] and [Was, Section 7]):

A unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{H})$ is exact if and only if the inclusion map $\iota\colon \mathcal{A} \hookrightarrow \mathcal{B}(\mathcal{H})$ has an approximate factorization

$$\mathcal{A} \xrightarrow{\varphi_\lambda} M_{n_\lambda}(\mathbb{C}) \xrightarrow{\psi_\lambda} \mathcal{B}(\mathcal{H}),$$

through a net of full matrix algebras $M_{n_\lambda}(\mathbb{C})$, $\lambda \in \Lambda$. Here, $\varphi_\lambda, \psi_\lambda$ are unital completely positive maps, and

$$\lim_{\lambda} \|\psi_\lambda \circ \varphi_\lambda(x) - x\| = 0, \quad \text{for all } x \text{ in } \mathcal{A}.$$

Finally, we use a dilation argument to pass from the condition $\sum_{i=1}^r a_i a_i^* = \mathbf{1}_{\mathcal{K}}$ of (b) above, to the less restrictive one: $\|\sum_{i=1}^r a_i a_i^*\| \le 1$, which is assumed in (ii) of the Main Theorem (when $d=1$). The proof of the fact that the key estimates (0.1) and (0.2) imply the Main Theorem, is given in Section 4 for the upper bound, and in Section 8 for the lower bound. Sections 1-3 and 5-7 are used to prove the key estimates (0.1) respectively (0.2).

In Section 1, we associate to any permutation π in the symmetric group S_p, a permutation $\hat{\pi}$ in S_{2p}, for which $\hat{\pi}^2 = \hat{\pi}\circ\hat{\pi} = \mathrm{id}$ and $\hat{\pi}(j) \ne j$ for all j, namely the permutation given by

$$\begin{aligned} \hat{\pi}(2j-1) &= 2\pi^{-1}(j), && (j \in \{1,2,\dots,p\}) \\ \hat{\pi}(2j) &= 2\pi(j)-1, && (j \in \{1,2,\dots,p\}). \end{aligned}$$

Moreover, following [Vo1], we let $\sim_{\hat{\pi}}$ denote the equivalence relation on $\{1,2,\dots,2p\}$, generated by the expression:

$$j \sim_{\hat{\pi}} \hat{\pi}(j)+1, \qquad \text{(addition formed mod. } 2p),$$

and we let $d(\hat{\pi})$ denote the number of equivalence classes for $\sim_{\hat{\pi}}$. We can write $d(\hat{\pi}) = k(\hat{\pi}) + l(\hat{\pi})$, where $k(\hat{\pi})$ (resp. $l(\hat{\pi})$) denotes the number of equivalence classes for $\sim_{\hat{\pi}}$, consisting entirely of even numbers (resp. odd numbers) in $\{1, 2, \ldots, 2p\}$. With this notation we prove, that for any random matrix B from $\mathrm{GRM}(m, n, 1)$,

$$\mathbb{E} \circ \mathrm{Tr}_n\big[(B^*B)^p\big] = \sum_{\pi \in S_p} m^{k(\hat{\pi})} n^{l(\hat{\pi})}. \tag{0.3}$$

Consider next the quantity $\sigma(\hat{\pi}) = \frac{1}{2}(p + 1 - d(\hat{\pi}))$. It turns out, that $\sigma(\hat{\pi})$ is always a non-negative integer, and that $\sigma(\hat{\pi}) = 0$ if and only if $\hat{\pi}$ is non-crossing (cf. Definition 1.14). In Section 2 we show, that if $a_1, \ldots, a_r$ are elements of $\mathcal{B}(\mathcal{H}, \mathcal{K})$ and $S = \sum_{i=1}^r a_i \otimes Y_i^{(n)}$, where $Y_1^{(n)}, \ldots, Y_r^{(n)}$ are independent elements of $\mathrm{GRM}(n, n, \frac{1}{n})$, then

$$\mathbb{E}[(S^*S)^p] = \Big(\sum_{\pi \in S_p} n^{-2\sigma(\hat{\pi})} \cdot \sum_{1 \le i_1, \ldots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}\Big) \otimes \mathbf{1}_{M_n(\mathbb{C})}. \tag{0.4}$$

In [HT, Section 6], we found explicit formulas for the quantities $\mathbb{E} \circ \mathrm{Tr}_n[\exp(tB^*B)]$ and $\mathbb{E} \circ \mathrm{Tr}_n[B^*B\exp(tB^*B)]$, where B is an element of $\mathrm{GRM}(m, n, 1)$. In Section 3, a careful comparison of the terms in (0.3) and (0.4), combined with these explicit formulas, allows us to prove, that if $\| \sum_{i=1}^r a_i^* a_i \| \le c$ and $\| \sum_{i=1}^r a_i a_i^* \| \le 1$, then for $0 \le t \le \min\{\frac{n}{2c}, \frac{n}{2}\}$,

$$\|\mathbb{E}[\exp(tS^*S)]\| \le \exp((c+1)^2 \tfrac{t^2}{n}) \int_0^\infty \exp(tx)\ d\mu_c(x), \tag{0.5}$$

where μ_c is the free (analog of the) Poisson distribution with parameter c (cf. [VDN] and [HT, Section 6]). The measure μ_c is also called the Marchenko-Pastur distribution (cf. [OP]), and it is given by

$$\mu_c = \max\{1 - c, 0\}\delta_0 + \frac{\sqrt{(x-a)(b-x)}}{2\pi x} \cdot 1_{[a,b]}(x)\ dx,$$

where $a = (\sqrt{c} - 1)^2$, $b = (\sqrt{c} + 1)^2$ and δ_0 is the Dirac measure at 0. Since $\mathrm{supp}(\mu_c) \subseteq [0, b]$, the first key estimate, (0.1), follows immediately from (0.5). To prove the second key estimate, (0.2), we show in Sections 5-6, that under the condition

$$\sum_{i=1}^r a_i^* a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}, \qquad \text{and} \qquad \sum_{i=1}^r a_i a_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{K})},$$

one has, for any q in $\mathbb{N}$, the formula:

$$\mathbb{E}\big[P_q^c(S^*S)\big] = \Big[\sum_{\rho \in S_q^{\mathrm{irr}}} n^{-2\sigma(\hat{\rho})} \Big(\sum_{1 \le i_1, \ldots, i_q \le r} a_{i_1}^* a_{i_{\rho(1)}} \cdots a_{i_q}^* a_{i_{\rho(q)}}\Big)\Big] \otimes \mathbf{1}_{M_n(\mathbb{C})}. \tag{0.6}$$

Here $P_0^c(x), P_1^c(x), P_2^c(x), \dots,$ is the sequence of monic polynomials obtained from $1, x, x^2, \dots,$ by the Gram-Schmidt orthogonalization process, w.r.t. the inner product

$$\langle f, g\rangle = \int_0^\infty f\overline{g}\ d\mu_c, \qquad (f, g \in L^2(\mathbb{R}, \mu_c)).$$

Moreover, S_q^{irr} denotes the set of permutations ρ in S_q, for which

$$1 \neq \rho(1) \neq 2 \neq \rho(2) \neq \cdots \neq q \neq \rho(q).$$

For fixed t in $\mathbb{R}$, we expand in Section 7 the exponential function $x \mapsto \exp(tx)$, in terms of the polynomials $P_q^c(x)$, $q \in \mathbb{N}_0$:

$$\exp(tx) = \sum_{q=0}^\infty \psi_q^c(t) P_q^c(x), \qquad (x \in [0, \infty[). \tag{0.7}$$

We show that the coefficients $\psi_q^c(t)$ are non-negative for all t in $[0, \infty[$, and that for any q in $\mathbb{N}_0$,

$$|\psi_q^c(-t)| \leq \left(\frac{\int_0^\infty \exp(-tx)\ d\mu_c(x)}{\int_0^\infty \exp(tx)\ d\mu_c(x)}\right) \cdot \psi_q^c(t), \qquad (t \in [0, \infty[). \tag{0.8}$$

By combining (0.6), (0.7) and (0.8) with the proof of (0.5), we obtain that for $c \geq 1$ and $0 \leq t \leq \frac{n}{2c}$,

$$\|\mathbb{E}[\exp(-tS^*S)]\| \leq \exp((c+1)^2 \tfrac{t^2}{n}) \int_0^\infty \exp(-tx)\ d\mu_c(x),$$

and since $\operatorname{supp}(\mu_c) \subseteq [a, \infty[= [(\sqrt{c}-1)^2, \infty[$, when $c \geq 1$, we obtain the second key estimate (0.2).

The rest of the paper is organized in the following way:

1 A Combinatorial Expression for $\mathbb{E}\circ\mathrm{Tr}_n\big[(B^*B)^p\big]$, for a Gaussian Random Matrix B in $\mathrm{GRM}(m,n,1)$

For ξ in $\mathbb{R}$ and σ^2 in $]0,\infty[$, we let $N(\xi,\sigma^2)$ denote the Gaussian (or normal) distribution with mean ξ and variance σ^2. In [HT], we introduced the following class of Gaussian random matrices

1.1 Definition. (cf. [HT]) Let $(\Omega,\mathcal{F},P)$ be a classical probability space, let m,n be positive integers, and let

$$B=(b(i,j))_{\substack{1\le i\le m\\ 1\le j\le n}}\colon \Omega\to M_{m,n}(\mathbb{C}),$$

be a complex, random $m\times n$ matrix defined on Ω. We say then that B is a (standard) Gaussian random $m\times n$ matrix with entries of variance σ^2, if the real valued random variables $\mathrm{Re}(b(i,j))$, $\mathrm{Im}(b(i,j))$, $1\le i\le m$, $1\le j\le n$, form a family of $2mn$ independent, identically distributed random variables, with distribution $N(0,\frac{\sigma^2}{2})$. We denote by $\mathrm{GRM}(m,n,\sigma^2)$ the set of such random matrices defined on Ω. Note that σ^2 equals the second absolute moment of the entries of an element from $\mathrm{GRM}(m,n,\sigma^2)$. □

In the following we shall omit mentioning the underlying probability space $(\Omega,\mathcal{F},P)$, and it will be understood that all considered random matrices/variables are defined on this probability space. As a matter of notation, by 1_n we denote the unit of $M_n(\mathbb{C})$, and by tr_n we denote the trace on $M_n(\mathbb{C})$ satisfying that $\mathrm{tr}_n(1_n)=1$. Moreover, we put $\mathrm{Tr}_n=n\cdot\mathrm{tr}_n$.
Let B be an element of $\mathrm{GRM}(m,n,\sigma^2)$. Then for any p in $\mathbb{N}$, $(B^*B)^p$ is a positive definite $n\times n$ random matrix, and $\mathrm{Tr}_n((B^*B)^p)$ is a positive valued, integrable, random variable. The main aim of this section is to derive a combinatorial expression for the moments $\mathbb{E}\circ\mathrm{Tr}_n((B^*B)^p)$ of B^*B w.r.t. $\mathbb{E}\circ\mathrm{Tr}_n$, where $\mathbb{E}$ denotes expectation w.r.t. P.

1.2 Lemma. *Let m,n,r,p be positive integers, let $B_1,B_2,\dots,B_r$ be independent elements of $\mathrm{GRM}(m,n,\sigma^2)$, and for each s in $\{1,2,\dots,r\}$, let $b(u,v,s)$, $1\le u\le m$, $1\le v\le n$, denote the entries of B_s. Then for any $i_1,j_1,i_2,j_2,\dots,i_p,j_p$ in $\{1,2,\dots,r\}$, we have that*

$$\begin{aligned}&\mathbb{E}\circ\mathrm{Tr}_n(B_{i_1}^*B_{j_1}B_{i_2}^*B_{j_2}\cdots B_{i_p}^*B_{j_p})\\ &=\sum_{\substack{1\le u_2,u_4,\dots,u_{2p}\le m\\ 1\le u_1,u_3,\dots,u_{2p-1}\le n}}\mathbb{E}\Big(\overline{b(u_2,u_1,i_1)}b(u_2,u_3,j_1)\cdots\overline{b(u_{2p},u_{2p-1},i_p)}b(u_{2p},u_1,j_p)\Big),\end{aligned}\tag{1.1}$$

and moreover $\mathbb{E}\circ\mathrm{Tr}_n(B_{i_1}^*B_{j_1}B_{i_2}^*B_{j_2}\cdots B_{i_p}^*B_{j_p})=0$, unless there exists a permutation π in the symmetric group S_p, such that $j_h=i_{\pi(h)}$ for all h in $\{1,2,\dots,p\}$.

Proof. Let $f(u,v)$, $1 \le u \le m$, $1 \le v \le n$, denote the usual $m \times n$ matrix units, and let $g(u,v)$, $1 \le u \le n$, $1 \le v \le m$, denote the usual $n \times m$ matrix units. We have then that

$$\begin{aligned}
&\mathbb{E} \circ \mathrm{Tr}_n(B_{i_1}^* B_{j_1} B_{i_2}^* B_{j_2} \cdots B_{i_p}^* B_{j_p}) \\
&= \sum_{\substack{1 \le v_1, u_2, v_3, u_4, \dots, v_{2p-1}, u_{2p} \le m \\ 1 \le u_1, v_2, u_3, v_4, \dots, u_{2p-1}, v_{2p} \le n}} \mathbb{E}\big(b^*(u_1, v_1, i_1) b(u_2, v_2, j_1) \cdots b^*(u_{2p-1}, v_{2p-1}, i_p) b(u_{2p}, v_{2p}, j_p)\big) \\
&\qquad\qquad\qquad \cdot \mathrm{Tr}_n\big(g(u_1, v_1) f(u_2, v_2) \cdots g(u_{2p-1}, v_{2p-1}) f(u_{2p}, v_{2p})\big) \\
&= \sum_{\substack{1 \le u_2, u_4, \dots, u_{2p} \le m \\ 1 \le u_1, u_3, \dots, u_{2p-1} \le n}} \mathbb{E}\Big(\overline{b(u_2, u_1, i_1)} b(u_2, u_3, j_1) \cdots \overline{b(u_{2p}, u_{2p-1}, i_p)} b(u_{2p}, u_1, j_p)\Big).
\end{aligned}$$

Note here, that for any $u_2, u_4, \dots, u_{2p}$ in $\{1, 2, \dots, m\}$ and $u_1, u_3, \dots, u_{2p-1}$ in $\{1, 2, \dots, n\}$, we have because of the independence assumptions,

$$\begin{aligned}
&\mathbb{E}\Big(\overline{b(u_2, u_1, i_1)} b(u_2, u_3, j_1) \cdots \overline{b(u_{2p}, u_{2p-1}, i_p)} b(u_{2p}, u_1, j_p)\Big) \\
&\qquad = \prod_{l=1}^{r} \mathbb{E}\Big(\prod_{h : i_h = l} \overline{b(u_{2h}, u_{2h-1}, l)} \prod_{h : j_h = l} b(u_{2h}, u_{2h+1}, l)\Big),
\end{aligned}$$

where $2h+1$ is calculated mod. $2p$.
Note here, that for any l in $\{1, 2, \dots, r\}$, any u in $\{1, 2, \dots, m\}$ and any v in $\{1, 2, \dots, n\}$, the distribution of $b(u,v,l)$ is invariant under multiplication by complex numbers of norm 1. Hence, for any s, t in $\mathbb{N}_0$, $\mathbb{E}\big[\overline{b(u,v,l)^s} \cdot b(u,v,l)^t\big] = 0$, unless $s = t$. Using this, and the independence assumptions, it follows that for any l in $\{1, 2, \dots, r\}$, any $u_2, u_4, \dots, u_{2p}$ in $\{1, 2, \dots, m\}$ and any $u_1, u_3, \dots, u_{2p-1}$ in $\{1, 2, \dots, n\}$, a necessary condition for the mean

$$\mathbb{E}\Big(\prod_{h : i_h = l} \overline{b(u_{2h}, u_{2h-1}, l)} \cdot \prod_{h : j_h = l} b(u_{2h}, u_{2h+1}, l)\Big)$$

to be distinct from zero is that

$$\mathrm{card}\big(\{h \in \{1, 2, \dots, p\} \mid i_h = l\}\big) = \mathrm{card}\big(\{h \in \{1, 2, \dots, p\} \mid j_h = l\}\big). \quad (1.2)$$

It follows that $\mathbb{E} \circ \mathrm{Tr}_n(B_{i_1}^* B_{j_1} B_{i_2}^* B_{j_2} \cdots B_{i_p}^* B_{j_p}) = 0$, unless (1.2) holds for all l in $\{1, 2, \dots, r\}$, and in this case, it is not hard to construct a permutation π from S_p, with the property described in the lemma. □

1.3 Definition. Let p be a positive integer, and let π be an element of S_p. We associate to π a family $\Lambda(\pi, m, n)$, $m, n \in \mathbb{N}$, of complex numbers, as follows: Let $B_1, B_2, \dots, B_p$ be independent elements of $\mathrm{GRM}(m, n, 1)$, and then define

$$\Lambda(\pi, m, n) = \mathbb{E} \circ \mathrm{Tr}_n(B_1^* B_{\pi(1)} B_2^* B_{\pi(2)} \cdots B_p^* B_{\pi(p)}). \qquad □$$

1.4 Remark. Let m,n,r,p be positive integers, and let $B_1,B_2,\dots,B_r$ be arbitrary elements of $\mathrm{GRM}(m,n,\sigma^2)$. Moreover, let $i_1,j_1,\dots,i_p,j_p$ be arbitrary elements of $\{1,2,\dots,r\}$. We shall need the fact that the quantity $\mathbb{E}\circ\mathrm{Tr}_n(B_{i_1}^*B_{j_1}\cdots B_{i_p}^*B_{j_p})$ is bounded numerically by some constant $K(m,n,p,\sigma^2)$ depending only on m,n,p,σ^2 and not on r or the distributional relations between $B_1,B_2,\dots,B_r$. For this, adapt the notation from Lemma 1.2, and note then that by (1.1) from that lemma,

$$\begin{aligned}&|\mathbb{E}\circ\mathrm{Tr}_n(B_{i_1}^*B_{j_1}\cdots B_{i_p}^*B_{j_p})|\\&\quad\le\sum_{\substack{1\le u_2,u_4,\dots,u_{2p}\le m\\1\le u_1,u_3,\dots,u_{2p-1}\le n}}\Big|\,\mathbb{E}\Big(\overline{b(u_2,u_1,i_1)}b(u_2,u_3,j_1)\cdots\overline{b(u_{2p},u_{2p-1},i_p)}b(u_{2p},u_1,j_p)\Big)\,\Big|.\end{aligned}$$

Then let $M(2p,\sigma^2)$ denote the $2p$'th absolute moment of the entries of an element from $\mathrm{GRM}(m,n,\sigma^2)$. A standard computation yields that $M(2p,\sigma^2)=\sigma^{2p}\cdot p!$, but we shall not need this explicit formula. It follows now by the generalized Hölder inequality, that for any $u_2,u_4,\dots,u_{2p}$ in $\{1,2,\dots,m\}$ and $u_1,u_3,\dots,u_{2p-1}$ in $\{1,2,\dots,n\}$,

$$\begin{aligned}&\Big|\mathbb{E}\Big(\overline{b(u_2,u_1,i_1)}b(u_2,u_3,j_1)\cdots\overline{b(u_{2p},u_{2p-1},i_p)}b(u_{2p},u_1,j_p)\Big)\,\Big|\\&\quad\le\big\|\overline{b(u_2,u_1,i_1)}\big\|_{2p}\big\|b(u_2,u_3,j_1)\big\|_{2p}\cdots\big\|\overline{b(u_{2p},u_{2p-1},i_p)}\big\|_{2p}\big\|b(u_{2p},u_1,j_p)\big\|_{2p}\\&\quad=\big(M(2p,\sigma^2)^{\frac{1}{2p}}\big)^{2p}=M(2p,\sigma^2).\end{aligned}$$

Thus it follows that we may use $K(m,n,p,\sigma^2)=m^pn^pM(2p,\sigma^2)$. □

1.5 Proposition. *Let B be an element of $\mathrm{GRM}(m,n,1)$, and let p be a positive integer. We then have*

$$\mathbb{E}\circ\mathrm{Tr}_n[(B^*B)^p]=\sum_{\pi\in S_p}\Lambda(\pi,m,n).$$

Proof. Let $(B_i)_{i\in\mathbb{N}}$ be a sequence of independent elements of $\mathrm{GRM}(m,n,1)$. Note then that for any s in $\mathbb{N}$, the matrix $\frac{1}{\sqrt{s}}(B_1+\cdots+B_s)$ is again an element of $\mathrm{GRM}(m,n,1)$, and therefore

$$\begin{aligned}\mathbb{E}\circ\mathrm{Tr}_n\left[(B^*B)^p\right]&=\mathbb{E}\circ\mathrm{Tr}_n\left[\Big(\big(s^{-\frac12}(B_1+\cdots+B_s)\big)^*\big(s^{-\frac12}(B_1+\cdots+B_s)\big)\Big)^p\right]\\&=s^{-p}\sum_{1\le i_1,j_1,\dots,i_p,j_p\le s}\mathbb{E}\circ\mathrm{Tr}_n\left[B_{i_1}^*B_{j_1}\cdots B_{i_p}^*B_{j_p}\right].\end{aligned}\tag{1.3}$$

For π in S_p we define

$$M(\pi,s)=\big\{(i_1,j_1,\dots,i_p,j_p)\in\{1,2,\dots,s\}^{2p}\bigm| j_1=i_{\pi(1)},\dots,j_p=i_{\pi(p)}\big\}.$$

It follows then from Lemma 1.2, that in (1.3), we only have to sum over those $2p$-tuples $(i_1, j_1, \dots, i_p, j_p)$ that belong to $M(\pi, s)$ for some π in S_p, and consequently

$$\mathbb{E} \circ \mathrm{Tr}_n\left[(B^*B)^p\right] = s^{-p} \sum_{(i_1,j_1,\dots,i_p,j_p)\in\cup_{\pi\in S_p} M(\pi,s)} \mathbb{E} \circ \mathrm{Tr}_n\left[B_{i_1}^* B_{j_1} \cdots B_{i_p}^* B_{j_p}\right].$$

Note though, that the sets $M(\pi, s)$, $\pi \in S_p$, are not disjoint. However, if we put

$$\mathcal{D}(s) = \left\{(i_1, j_1, \dots, i_p, j_p) \in \{1, 2, \dots, s\}^{2p} \;\middle|\; i_1, i_2, \dots, i_p \text{ are distinct}\right\},$$

then the sets $M(\pi, s) \cap \mathcal{D}(s)$, $\pi \in S_p$, are disjoint. Thus we have

$$\begin{aligned}
&\mathbb{E} \circ \mathrm{Tr}_n\left[(B^*B)^p\right] \\
&\quad = s^{-p} \sum_{\pi\in S_p} \sum_{(i_1,j_1,\dots,i_p,j_p)\in M(\pi,s)\cap\mathcal{D}(s)} \mathbb{E} \circ \mathrm{Tr}_n\left[B_{i_1}^* B_{j_1} \cdots B_{i_p}^* B_{j_p}\right] \\
&\quad\quad + s^{-p} \sum_{(i_1,j_1,\dots,i_p,j_p)\in\left(\cup_{\pi\in S_p} M(\pi,s)\right)\setminus\mathcal{D}(s)} \mathbb{E} \circ \mathrm{Tr}_n\left[B_{i_1}^* B_{j_1} \cdots B_{i_p}^* B_{j_p}\right].
\end{aligned} \tag{1.4}$$

Note here, that if $(i_1, j_1, \dots, i_p, j_p) \in M(\pi, s) \cap \mathcal{D}(s)$, then $B_{i_1}, B_{i_2}, \dots, B_{i_p}$ are independent elements of $\mathrm{GRM}(m, n, 1)$, and hence

$$\mathbb{E} \circ \mathrm{Tr}_n\left[B_{i_1}^* B_{j_1} \cdots B_{i_p}^* B_{j_p}\right] = \Lambda(\pi, m, n).$$

Thus, the first term on the right hand side of (1.4) equals

$$s^{-p} \sum_{\pi\in S_p} \mathrm{card}(M(\pi, s) \cap \mathcal{D}(s)) \cdot \Lambda(\pi, m, n).$$

Here $\mathrm{card}(M(\pi, s) \cap \mathcal{D}(s)) = s(s-1)\cdots(s-p+1)$, so

$$s^{-p} \cdot \mathrm{card}(M(\pi, s) \cap \mathcal{D}(s)) \to 1 \text{ as } s \to \infty.$$

Hence, the first term on the right hand side of (1.4) tends to $\sum_{\pi\in S_p} \Lambda(\pi, m, n)$ as $s \to \infty$, and since the left hand side of (1.4) does not depend on s, it remains thus to show that the second term on the right hand side of (1.4) tends to 0 as $s \to \infty$. This follows by noting that according to Remark 1.4, for any $(i_1, j_1, \dots, i_p, j_p)$ in $\{1, 2, \dots, s\}^{2p}$, the quantity $|\mathbb{E} \circ \mathrm{Tr}_n\big[B_{i_1}^* B_{j_1} \cdots B_{i_p}^* B_{j_p}\big]|$ is bounded by some constant $K(m, n, p)$ depending only on m, n, p; not on s. And moreover,

$$\begin{aligned}
s^{-p}\mathrm{card}\big(\big(\cup_{\pi\in S_p} M(\pi, s)\big) \setminus \mathcal{D}(s)\big) &\le \sum_{\pi\in S_p} s^{-p}\mathrm{card}(M(\pi, s) \setminus \mathcal{D}(s)) \\
&= \sum_{\pi\in S_p} \left[s^{-p}\mathrm{card}(M(\pi, s)) - s^{-p}\mathrm{card}(M(\pi, s) \cap \mathcal{D}(s))\right] \\
&= \sum_{\pi\in S_p} \left[1 - s^{-p}\mathrm{card}(M(\pi, s) \cap \mathcal{D}(s))\right] \to 0,
\end{aligned}$$

as $s \to \infty$. This concludes the proof of the proposition. □

It follows from Proposition 1.5, that in order to obtain a combinatorial expression for the moments $\mathbb{E} \circ \mathrm{Tr}_n((B^*B)^p)$ for a matrix B from $\mathrm{GRM}(m, n, 1)$, we need to derive a combinatorial expression for the quantities

$$\Lambda(\pi, m, n) = \mathbb{E} \circ \mathrm{Tr}_n(B_1^* B_{\pi(1)} B_2^* B_{\pi(2)} \cdots B_p^* B_{\pi(p)}),$$

where $\pi \in S_p$ and $B_1, \dots, B_p$ are independent elements of $\mathrm{GRM}(m, n, 1)$.
As it turns out, it shall be useful to have the relations between the factors in the product $B_1^* B_{\pi(1)} B_2^* B_{\pi(2)} \cdots B_p^* B_{\pi(p)}$ determined in terms of a permutation $\hat{\pi}$ in S_{2p}, rather than in terms of the permutation π from S_p.

1.6 DEFINITION. Let p be a positive integer, and let π be a permutation in S_p. Then the permutation $\hat{\pi}$ in S_{2p} is determined by the equations:

$$\begin{aligned} \hat{\pi}(2i-1) &= 2\pi^{-1}(i), \qquad (i \in \{1, 2, \dots, p\}), \\ \hat{\pi}(2i) &= 2\pi(i) - 1, \qquad (i \in \{1, 2, \dots, p\}). \end{aligned}$$ □

1.7 REMARK. (a) Let p, π and $\hat{\pi}$ be as in Definition 1.6. Note then that $\hat{\pi}^2 = \hat{\pi} \circ \hat{\pi} = \mathrm{id}$, the identity mapping on $\{1, 2, \dots, 2p\}$, and that $\hat{\pi}$ maps odd numbers to even numbers, i.e., that $\hat{\pi}(j) - j = 1 \pmod{2}$, for all j in $\{1, 2, \dots, 2p\}$. In particular, $\hat{\pi}$ has no fixed points. It is easy to check that $\{\hat{\pi} \mid \pi \in S_p\}$ is exactly the set of permutations γ in S_{2p}, for which $\gamma^2 = \mathrm{id}$ and $\gamma(j) - j = 1 \pmod{2}$, for all j in $\{1, 2, \dots, 2p\}$. Moreover, the mapping $\pi \mapsto \hat{\pi}$ is injective.
(b) If $B_1, B_2, \dots, B_p$ are independent elements of $\mathrm{GRM}(m, n, 1)$, then we may write the product $B_1^* B_{\pi(1)} B_2^* B_{\pi(2)} \cdots B_p^* B_{\pi(p)}$ in the form $C_1^* C_2 C_3^* C_4 \cdots C_{2p-1}^* C_{2p}$, where $C_{2i-1} = B_i$ and $C_{2i} = B_{\pi(i)}$ for all i in $\{1, 2, \dots, p\}$. Then $\hat{\pi}$ is constructed exactly so that for any j, j' in $\{1, 2, \dots, 2p\}$, we have

$$C_j = C_{j'} \Leftrightarrow j = j' \text{ or } \hat{\pi}(j) = j'.$$ □

1.8 DEFINITION. We associate to $\hat{\pi}$ an equivalence relation $\sim_{\hat{\pi}}$ on $\mathbb{Z}_{2p}$. This is the equivalence relation (introduced by Voiculescu in [Vo1, Proof of Theorem 2.2]), generated by the expression:

$$j \sim_{\hat{\pi}} \hat{\pi}(j) + 1, \qquad (j \in \{1, 2, \dots, 2p\}),$$

where addition is formed mod. $2p$. □

1.9 REMARK. For a permutation π in S_p, the $\sim_{\hat{\pi}}$-equivalence classes are precisely the orbits in $\{1, 2, \dots, 2p\}$ for the cyclic subgroup of S_{2p} generated by the permutation $j \mapsto \hat{\pi}(j) + 1$ (addition formed mod. $2p$). Since this subgroup is finite, the equivalence class $[j]_{\hat{\pi}}$ of an element j in $\{1, 2, \dots, 2p\}$ has the following form:

Let q be the number of elements in $[j]_{\hat{\pi}}$. Then

$$[j]_{\hat{\pi}} = \{j_0, j_1, \dots, j_{q-1}\},$$

where $j_0 = j, j_1 = \hat{\pi}(j_0) + 1, j_2 = \hat{\pi}(j_1) + 1, \dots, j_{q-1} = \hat{\pi}(j_{q-2}) + 1, j_0 = \hat{\pi}(j_{q-1}) + 1$, (addition formed mod. $2p$). □

It follows immediately from the definition of $\hat{\pi}$ and Remark 1.9 that each $\sim_{\hat{\pi}}$-equivalence class consists entirely of even numbers or entirely of odd numbers. This is used in the following definition:

1.10 Definition. Let p be a positive integer, let π be a permutation in S_p, and consider the corresponding permutation $\hat{\pi}$ in S_{2p}. By $k(\hat{\pi})$ and $l(\hat{\pi})$, we denote then the number of $\sim_{\hat{\pi}}$-equivalence classes consisting of even numbers, respectively the number of $\sim_{\hat{\pi}}$-equivalence classes consisting of odd numbers:

$$\begin{aligned} k(\hat{\pi}) &= \operatorname{card}\big(\big\{[j]_{\hat{\pi}} \bigm| j \in \{2, 4, \dots, 2p\}\big\}\big),\\ l(\hat{\pi}) &= \operatorname{card}\big(\big\{[j]_{\hat{\pi}} \bigm| j \in \{1, 3, \dots, 2p-1\}\big\}\big). \end{aligned}$$

Moreover, we define the quantities $d(\hat{\pi})$ and $\sigma(\hat{\pi})$ by the equations:

$$\begin{aligned} d(\hat{\pi}) &= k(\hat{\pi}) + l(\hat{\pi}) = \operatorname{card}\big(\big\{[j]_{\hat{\pi}} \bigm| j \in \{1, 2, \dots, 2p\}\big\}\big),\\ \sigma(\hat{\pi}) &= \tfrac{1}{2}\big(p + 1 - d(\hat{\pi})\big). \end{aligned}$$ □

Regarding the definition of $\sigma(\hat{\pi})$, it will be shown later (cf. Theorem 1.13), that $\sigma(\hat{\pi})$ is always a non-negative integer. The quantity $d(\hat{\pi})$ was introduced by Voiculescu in [Vo1, Proof of Theorem 2.2].

1.11 Theorem. *For any positive integers m, n and any π in S_p, we have that*

$$\Lambda(\pi, m, n) = m^{k(\hat{\pi})} n^{l(\hat{\pi})}.$$

Proof. Consider independent elements $B_1, B_2, \dots, B_p$ of $\mathrm{GRM}(m, n, 1)$, and for each j in $\{1, 2, \dots, p\}$, let $b(u, v, j)$, $1 \le u \le m$, $1 \le v \le n$, denote the entries of B_j. It follows then by (1.1) in Lemma 1.2, that

$$\begin{aligned} &\Lambda(\pi, m, n)\\ &= \mathbb{E} \circ \mathrm{Tr}_n(B_1^* B_{\pi(1)} B_2^* B_{\pi(2)} \cdots B_p^* B_{\pi(p)})\\ &= \sum_{\substack{1 \le u_1, u_3, \dots, u_{2p-1} \le n\\ 1 \le u_2, u_4, \dots, u_{2p} \le m}} \mathbb{E}\Big(\overline{b(u_2, u_1, 1)} b(u_2, u_3, \pi(1)) \cdots \overline{b(u_{2p}, u_{2p-1}, p)} b(u_{2p}, u_1, \pi(p))\Big). \end{aligned} \tag{1.5}$$

Arguing as in the proof of Lemma 1.2, it follows that the term in the above sum corresponding to $u_1, u_2, \dots, u_{2p}$ is zero, unless the corresponding matrix entries are pairwise conjugate to each other, i.e., unless we have that

$$b(u_{2i}, u_{2i+1}, \pi(i)) = b(u_{2\pi(i)}, u_{2\pi(i)-1}, \pi(i)), \qquad (i \in \{1, 2, \dots, p\}). \tag{1.6}$$

Note also, that if (1.6) is satisfied, then the corresponding term in (1.5) equals 1, and consequently

$$\begin{aligned}&\Lambda(\pi, m, n)\\&= \operatorname{card}\big(\big\{(u_1, u_2, \ldots, u_{2p}) \bigm| 1 \le u_{2i-1} \le n,\ 1 \le u_{2i} \le m,\ \text{and (1.6) holds}\big\}\big).\end{aligned}$$

To calculate this cardinality, we note first that (1.6) is equivalent to the condition

$$u_{2i} = u_{2\pi(i)} \quad \text{and} \quad u_{2i+1} = u_{2\pi(i)-1}, \qquad (i \in \{1, 2, \ldots, p\}), \tag{1.7}$$

where addition and subtraction are formed mod. $2p$. Replacing now i by $\pi^{-1}(i)$ in the first equation in (1.7), we get the equivalent condition:

$$u_{2i} = u_{2\pi^{-1}(i)} \quad \text{and} \quad u_{2i+1} = u_{2\pi(i)-1}, \qquad (i \in \{1, 2, \ldots, p\}).$$

Recall then that by definition of $\hat{\pi}$, $\hat{\pi}(2i-1) = 2\pi^{-1}(i)$, and using this formula with i replaced by $\pi(i)$, we get that also $2\pi(i) - 1 = \hat{\pi}\big(\hat{\pi}(2\pi(i) - 1)\big) = \hat{\pi}(2i)$. Thus (1.6) is equivalent to the condition

$$u_{2i} = u_{\hat{\pi}(2i-1)}, \quad \text{and} \quad u_{2i+1} = u_{\hat{\pi}(2i)}, \qquad (i \in \{1, 2, \ldots, p\}),$$

i.e., the condition

$$u_j = u_{\hat{\pi}(j-1)}, \qquad (j \in \{1, 2, \ldots, 2p\}).$$

Replacing finally j by $\hat{\pi}(j) + 1$, we conclude that (1.6) is equivalent to the condition

$$u_j = u_{\hat{\pi}(j)+1}, \qquad (j \in \{1, 2, \ldots, 2p\}),$$

where $\hat{\pi}(j) + 1$ is calculated mod. $2p$. Having realized this, it follows immediately from Remark 1.9 and the definitions of $k(\hat{\pi})$ and $l(\hat{\pi})$, that the right hand side of (1) equals $m^{k(\hat{\pi})} n^{l(\hat{\pi})}$, and hence we have the desired formula. □

1.12 Corollary. *Let m, n be positive integers and let B be an element of* $\mathrm{GRM}(m, n, 1)$. *Then for any positive integer p, we have that*

$$\mathbb{E} \circ \operatorname{Tr}_n\big[(B^*B)^p\big] = \sum_{\pi \in S_p} m^{k(\hat{\pi})} n^{l(\hat{\pi})}.$$

Proof. This follows immediately by combining Proposition 1.5 and Theorem 1.11. □

1.13 Theorem. *Let p be a positive integer, and let π be a permutation in S_p. Then*

(i) $k(\hat{\pi}) \ge 1$ *and* $l(\hat{\pi}) \ge 1$.
(ii) $k(\hat{\pi}) + l(\hat{\pi}) \le p + 1$.
(iii) $\sigma(\hat{\pi}) = \frac{1}{2}\big(p + 1 - k(\hat{\pi}) - l(\hat{\pi})\big)$ *is a non-negative integer.*

Proof. (i) This is clear from Definition 1.10.
(ii) Since $d(\hat{\pi}) = k(\hat{\pi}) + l(\hat{\pi})$ is the number of equivalence classes for $\sim_{\hat{\pi}}$, (ii) follows from [Vo1, Proof of Theorem 2.2].
(iii) The proof of (iii) requires more work. For elements p of $\mathbb{N}$ and k, l of $\mathbb{N}_0$, we define

$$\delta(p,k,l) = \operatorname{card}(\{\pi \in S_p \mid k(\hat{\pi}) = k \text{ and } l(\hat{\pi}) = l\}).$$

By (i) and (ii), $\delta(p,k,l) = 0$ unless $k \geq 1$, $l \geq 1$ and $k + l \leq p + 1$. By Corollary 1.12, we have for an element B of $\mathrm{GRM}(m,n,1)$, that

$$\mathbb{E} \circ \operatorname{Tr}_n\big[(B^*B)^p\big] = \sum_{\substack{k,l \in \mathbb{N} \\ k+l \leq p+1}} \delta(p,k,l) m^k n^l.$$

On the other hand, by the recursion formula for the moments $\mathbb{E} \circ \operatorname{Tr}_n\big[(B^*B)^p\big]$, $(p \in \mathbb{N})$, found in [HT, Theorem 8.2], it follows that for p in $\mathbb{N}$, the moment $\mathbb{E} \circ \operatorname{Tr}_n\big[(B^*B)^p\big]$ can be expressed as a polynomial in m and n of the form:

$$\mathbb{E} \circ \operatorname{Tr}_n\big[(B^*B)^p\big] = \sum_{\substack{k,l \in \mathbb{N} \\ k+l \leq p+1}} \delta'(p,k,l) m^k n^l,$$

for suitable coeffecients $\delta'(p,k,l)$. By the remarks following the proof of [HT, Theorem 8.2], only terms of homogeneous degree $p + 1 - 2j$, $j \in \{0, 1, 2, \dots, [\frac{p-1}{2}]\}$, appear in this polynomial, i.e.,

$$\delta'(p,k,l) = 0, \quad \text{when} \quad k + l = p \ (\text{mod. } 2).$$

If polynomials of two variables coincide on $\mathbb{N}^2$, then they are equal. Therefore, $\delta(p,k,l) = \delta'(p,k,l)$ for all k, l, which proves that

$$\operatorname{card}(\{\pi \in S_p \mid k(\hat{\pi}) = k \text{ and } l(\hat{\pi}) = l\}) = 0, \quad \text{if} \quad k + l = p \ (\text{mod. } 2).$$

Hence, $\sigma(\hat{\pi})$ is an integer for all π in S_p, and by (ii), $\sigma(\hat{\pi}) \geq 0$. This proves (iii). $\square$

In the rest of this section, we shall introduce a method of "reductions of permutations", which will be needed to determine the asymptotic lower bound of the spectrum of $S_n^* S_n$ (cf. Sections 5-8).
Let p be a positive integer, let π be a permutation in S_p, and consider the corresponding permutation $\hat{\pi}$ in S_{2p}, introduced in Definition 1.6. Since $\hat{\pi}^2 = \mathrm{id}$ and $\hat{\pi}$ has no fixed points, the orbits under the action of $\hat{\pi}$ form a partition of $\{1, 2, \dots, 2p\}$ into p sets, each with two elements.

1.14 Definition. Let p be a positive integer, and let π be a permutation in S_p. Following the standard definition of crossings in partitions of $\{1, 2, \dots, 2p\}$ into sets of cardinality 2 (see e.g. [Sp]), we say that (a, b, c, d) is a *crossing* for $\hat{\pi}$, if $a, b, c, d \in \{1, 2, \dots, 2p\}$ such that

$$a < b < c < d, \qquad \text{and} \qquad \hat{\pi}(a) = c, \ \hat{\pi}(b) = d. \tag{1.8}$$

If $\hat{\pi}$ has no such crossings, we say that $\hat{\pi}$ is a non-crossing permutation, and we let $S_p^{\rm nc}$ denote the set of permutations π in S_p for which $\hat{\pi}$ is non-crossing. $\square$

1.15 DEFINITION. Let p be a positive integer, let π be a permutation in S_p, and let e be an element of $\{1, 2, \dots, 2p-1\}$. We say then that $(e, e+1)$ is a *pair of neighbors* for $\hat{\pi}$, if $\hat{\pi}(e) = e+1$. Note, that a pair of neighbors for $\hat{\pi}$ is either of the form

$$(2k-1, 2k), \quad \text{where} \quad k \in \{1, \dots, p\},$$

or of the form

$$(2k, 2k+1), \quad \text{where} \quad k \in \{1, \dots, p-1\}.$$

In the first case $k = \pi(k)$, and in the second case $\pi(k) = k+1$. $\square$

1.16 DEFINITION. Let p be a positive integer, let π be a permutation in S_p, and consider the permutation $\hat{\pi}$ in S_{2p} introduced in Definition 1.6. We say then that $\hat{\pi}$ is *irreducible* if $\hat{\pi}$ has no pair of neighbors (in the sense of Definition 1.15), i.e., if $\hat{\pi}(j) \neq j+1$ for all j in $\{1, 2, \dots, 2p-1\}$. We denote by $S_p^{\rm irr}$ the set of permutations π in S_p for which $\hat{\pi}$ is irreducible. Note that

$$\pi \in S_p^{\rm irr} \iff 1 \neq \pi(1) \neq 2 \neq \pi(2) \neq \cdots \neq p \neq \pi(p).$$

If $\pi \in S_p \setminus S_p^{\rm irr}$, we say that $\hat{\pi}$ is *reducible*. Note, that we do not require that $\hat{\pi}(2p) \neq 1$ in order for $\hat{\pi}$ to be irreducible. Thus, irreducibility of $\hat{\pi}$ is *not* invariant under cyclic permutations of $\{1, 2, \dots, 2p\}$. $\square$

1.17 LEMMA. *Let p be a positive integer, and let π be a permutation in $S_p^{\rm nc}$. Then $\hat{\pi}$ has a pair of neighbors, i.e., $\hat{\pi}$ is reducible in the sense of Definition 1.16. In other words, we have the inclusion $S_p^{\rm nc} \subseteq S_p \setminus S_p^{\rm irr}$ or equivalently $S_p^{\rm irr} \subseteq S_p \setminus S_p^{\rm nc}$.*

Proof. We prove the inclusion: $S_p^{\rm irr} \subseteq S_p \setminus S_p^{\rm nc}$. So let π from $S_p^{\rm irr}$ be given, and consider the set $M = \{j \in \{1, 2, \dots, 2p\} \mid \hat{\pi}(j) \geq j\}$. Note that $M \neq \emptyset$, since clearly $1 \in M$. Define now

$$\alpha = \min\{\hat{\pi}(j) - j \mid j \in M\}.$$

Since $\hat{\pi}$ has no fixed points and no pairs of neighbors (since $\pi \in S_p^{\rm irr}$), we must have $\alpha \geq 2$. Choose j in $\{1, 2, \dots, 2p\}$ such that $\hat{\pi}(j) - j = \alpha$. Since $\alpha \geq 2$, $\hat{\pi}(j) \neq j+1$, or equivalently (since $\hat{\pi}^2 = \mathrm{id}$), $\hat{\pi}(j+1) \neq j$. Combining this with the definition of α, and the fact that $\hat{\pi}$ has no fixed points, it follows that

$$\hat{\pi}(j+1) \notin \{j, j+1, \dots, j+\alpha\} = \{j, j+1, \dots, \hat{\pi}(j)\},$$

i.e., either $\hat{\pi}(j+1) < j$ or $\hat{\pi}(j+1) > \hat{\pi}(j)$. In the first case $(\hat{\pi}(j+1), j, j+1, \hat{\pi}(j))$ is a crossing for $\hat{\pi}$, and in the second case $(j, j+1, \hat{\pi}(j), \hat{\pi}(j+1))$ is a crossing for $\hat{\pi}$. In all cases, $\pi \in S_p \setminus S_p^{\rm nc}$, as desired. $\square$

1.18 Definition. Let p be a positive integer, greater than or equal to 2, let π be a permutation in S_p, and assume that the permutation $\hat{\pi}$ in S_{2p} has a pair of neighbors $(e, e+1)$. Let φ be the order preserving bijection of $\{1, 2, \dots, 2p-2\}$ onto $\{1, 2, \dots, 2p\} \setminus \{e, e+1\}$, i.e.,

$$\varphi(i) = \begin{cases} i, & \text{if } 1 \le i \le e-1, \\ i+2, & \text{if } e \le i \le 2p-2. \end{cases} \tag{1.9}$$

By π_0 we denote then the unique permutation in S_{p-1}, satisfying that

$$\hat{\pi}_0 = \varphi^{-1} \circ \hat{\pi} \circ \varphi.$$

We say that $\hat{\pi}_0$ is obtained from $\hat{\pi}$ by *cancellation of the pair* $(e, e+1)$. □

A few words are appropriate about the introduction of π_0 in the definition above. Note first of all that $\varphi^{-1} \circ \hat{\pi} \circ \varphi$ is a well-defined permutation of $\{1, 2, \dots, 2p-2\}$, since $\hat{\pi}^2 = \mathrm{id}$ and $\hat{\pi}(e) = e+1$, so that $\hat{\pi}(\{1, 2, \dots, 2p\} \setminus \{e, e+1\}) = \{1, 2, \dots, 2p\} \setminus \{e, e+1\}$. To see that this permutation is actually of the form $\hat{\pi}_0$ for some (necessarily uniquely determined) permutation π_0 in S_{p-1}, it suffices, by Remark 1.7(a), to check that $(\varphi^{-1} \circ \hat{\pi} \circ \varphi)^2 = \mathrm{id}$, and that $\varphi^{-1} \circ \hat{\pi} \circ \varphi(j) - j = 1 \pmod{2}$, for all j in $\{1, 2, \dots, 2p-2\}$. But these properties follow from the corresponding properties of $\hat{\pi}$, and the fact that $\varphi(j) = j \pmod{2}$, for all j.

1.19 Remark. Let p be a positive integer, greater than or equal to 2, let π be a permutation in S_p, and assume that the permutation $\hat{\pi}$ in S_{2p} has a pair of neighbors $(e, e+1)$. Let π_0 be the permutation in S_{p-1} obtained from π as in Definition 1.18.
(a) If $(e, e+1) = (2k-1, 2k)$ for some k in $\{1, \dots, p\}$, then $\pi_0 = \psi^{-1} \circ \pi \circ \psi$, where $\psi\colon \{1, \dots, p-1\} \to \{1, \dots, p\} \setminus \{k\}$ is the bijection given by

$$\psi(j) = \begin{cases} j, & \text{if } 1 \le j \le k-1, \\ j+1, & \text{if } k \le j \le p-1. \end{cases} \tag{1.10}$$

(b) If $(e, e+1) = (2k, 2k+1)$ for some k in $\{1, \dots, p-1\}$, then $\pi_0 = \chi^{-1} \circ \pi \circ \psi$, where $\chi\colon \{1, \dots, p-1\} \to \{1, \dots, p\} \setminus \{k+1\}$ is the bijection given by

$$\chi(j) = \begin{cases} j, & \text{if } 1 \le j \le k, \\ j+1, & \text{if } k+1 \le j \le p-1, \end{cases} \tag{1.11}$$

and where ψ is given by (1.10). □

1.20 Lemma. *Let p be a positive integer, greater than or equal to 2, and let π be a permutation in $S_p \setminus S_p^{\mathrm{irr}}$. Let $(e, e+1)$ be a pair of neighbors for $\hat{\pi}$ and let π_0 be the permutation in S_{p-1}, for which $\hat{\pi}_0$ is the permutation obtained from $\hat{\pi}$ by cancellation of $(e, e+1)$. Then $\hat{\pi}$ is non-crossing if and only if $\hat{\pi}_0$ is non-crossing.*

Proof. Let $\varphi\colon \{1,2,\dots,2p-2\} \to \{1,2,\dots,2p\} \setminus \{e,e+1\}$ be the bijection introduced in (1.9). We show that $\hat{\pi}_0$ has a crossing if and only if $\hat{\pi}$ does.
Assume first that $\hat{\pi}_0$ has a crossing (a,b,c,d). Then since φ is (strictly) monotone and since (by definition of π_0) $\hat{\pi}(\varphi(a)) = \varphi(c)$, $\hat{\pi}(\varphi(b)) = \varphi(d)$, it follows that $(\varphi(a),\varphi(b),\varphi(c),\varphi(d))$ is a crossing for $\hat{\pi}$.
Assume conversely that $\hat{\pi}$ has a crossing (a',b',c',d'). Then clearly

$$\{a',b',c',d'\} \cap \{e,e+1\} = \emptyset,$$

so that the numbers $\varphi^{-1}(a'),\varphi^{-1}(b'),\varphi^{-1}(c'),\varphi^{-1}(d')$ are well-defined. It follows then, as above, that $(\varphi^{-1}(a'),\varphi^{-1}(b'),\varphi^{-1}(c'),\varphi^{-1}(d'))$ is a crossing for $\hat{\pi}_0$. □

1.21 Lemma. *Let m,n be positive integers, and let B be an element of* $\mathrm{GRM}(m,n,1)$. *Then*

$$\mathbb{E}(B^*B) = m\mathbf{1}_n, \qquad \text{and} \qquad \mathbb{E}(BB^*) = n\mathbf{1}_m. \tag{1.12}$$

Proof. Let $(b_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}}$ be the entries of B. Then

$$\mathbb{E}(\overline{b_{ij}}b_{st}) = \begin{cases} 1, & \text{if } (i,j) = (s,t),\\ 0, & \text{otherwise.}\end{cases} \tag{1.13}$$

Since $(B^*B)_{ij} = \sum_{s=1}^m \overline{b_{si}}b_{sj}$ and $(BB^*)_{ij} = \sum_{s=1}^n b_{is}\overline{b_{js}}$, for all i,j, (1.12) follows readily from (1.13). □

1.22 Proposition. *Let p be a positive integer, greater than or equal to 2, and let π be a permutation in $S_p \setminus S_p^{\mathrm{irr}}$. Let $(e,e+1)$ be a pair of neighbors for $\hat{\pi}$ and let π_0 be the permutation in S_{p-1}, for which $\hat{\pi}_0$ is the permutation obtained from $\hat{\pi}$ by cancellation of $(e,e+1)$. Then with $k(\cdot),l(\cdot),d(\cdot)$ and $\sigma(\cdot)$ as introduced in Definition 1.10, we have that*
(i) *If e is odd, then $k(\hat{\pi}_0) = k(\hat{\pi}) - 1$ and $l(\hat{\pi}_0) = l(\hat{\pi})$.*
(ii) *If e is even, then $k(\hat{\pi}_0) = k(\hat{\pi})$ and $l(\hat{\pi}_0) = l(\hat{\pi}) - 1$.*
In both cases, $d(\hat{\pi}_0) = d(\hat{\pi}) - 1$ and $\sigma(\hat{\pi}_0) = \sigma(\hat{\pi})$.

Proof. Let m,n be positive integers, and let $B_1,\dots,B_p$ be independent random matrices from $\mathrm{GRM}(m,n,1)$. By Theorem 1.11, we have then that

$$\mathbb{E}\circ\mathrm{Tr}_n\big[B_1^*B_{\pi(1)}\cdots B_p^*B_{\pi(p)}\big] = m^{k(\hat{\pi})}n^{l(\hat{\pi})}. \tag{1.14}$$

(i) Assume that e is odd, i.e., that $(e,e+1) = (2q-1,2q)$ for some q in $\{1,2,\dots,p\}$. Then $\pi(q) = q$, and hence the set of random matrices

$$(B_1^*,B_{\pi(1)},\dots,B_{q-1}^*,B_{\pi(q-1)},B_{q+1}^*,B_{\pi(q+1)},\dots,B_p^*,B_{\pi(p)})$$

is independent from the set $(B_q^*, B_{\pi(q)})$. Therefore,

$$\begin{aligned}\mathbb{E}\circ\mathrm{Tr}_n&\big[B_1^*B_{\pi(1)}\cdots B_p^*B_{\pi(p)}\big]\\&=\mathbb{E}\circ\mathrm{Tr}_n\big[B_1^*B_{\pi(1)}\cdots B_{q-1}^*B_{\pi(q-1)}\mathbb{E}(B_q^*B_{\pi(q)})B_{q+1}^*\cdots B_p^*B_{\pi(p)}\big]\\&=m\cdot\mathbb{E}\circ\mathrm{Tr}_n\big[B_1^*B_{\pi(1)}\cdots B_{q-1}^*B_{\pi(q-1)}B_{q+1}^*\cdots B_p^*B_{\pi(p)}\big],\end{aligned} \tag{1.15}$$

where the last equality follows from Lemma 1.21. Note that only the random matrices $B_1,\dots,B_{q-1},B_{q+1},\dots,B_p$ occur in the last expression in (1.15). Define now for i in $\{1,2,\dots,p-1\}$,

$$B_i'=\begin{cases}B_i, & \text{if } 1\le i\le q-1,\\ B_{i+1}, & \text{if } q\le i\le p-1.\end{cases}$$

Then by Remark 1.19(a), it follows that the last expression in (1.15) is equal to

$$m\cdot\mathbb{E}\circ\mathrm{Tr}_n\big[(B_1')^*B'_{\pi_0(1)}\cdots(B'_{p-1})^*B'_{\pi_0(p-1)}\big],$$

which, by Theorem 1.11, is equal to $m\cdot m^{k(\hat\pi_0)}n^{l(\hat\pi_0)}$. Altogether, we have shown that

$$m^{k(\hat\pi)}n^{l(\hat\pi)}=m\cdot m^{k(\hat\pi_0)}n^{l(\hat\pi_0)},$$

and since this holds for all positive integers m,n, it follows that $k(\hat\pi)=k(\hat\pi_0)+1$ and $l(\hat\pi)=l(\hat\pi_0)$. This proves (i).

(ii) Assume that e is even, i.e., that $(e,e+1)=(2q,2q+1)$, for some q in $\{1,2,\dots,p-1\}$. Then $\pi(q)=q+1$, and arguing now as in the proof of (i), we find that

$$\begin{aligned}m^{k(\hat\pi)}n^{l(\hat\pi)}&=\mathbb{E}\circ\mathrm{Tr}_n\big[B_1^*B_{\pi(1)}\cdots B_p^*B_{\pi(p)}\big]\\&=\mathbb{E}\circ\mathrm{Tr}_n\big[B_1^*B_{\pi(1)}\cdots B_q^*\mathbb{E}(B_{\pi(q)}B_{q+1}^*)B_{\pi(q+1)}\cdots B_p^*B_{\pi(p)}\big]\\&=n\cdot\mathbb{E}\circ\mathrm{Tr}_n\big[B_1^*B_{\pi(1)}\cdots B_q^*B_{\pi(q+1)}\cdots B_p^*B_{\pi(p)}\big],\end{aligned} \tag{1.16}$$

where the last equality follows from Lemma 1.21. Defining, this time, for each i in $\{1,2,\dots,p-1\}$,

$$B_i'=\begin{cases}B_i, & \text{if } 1\le i\le q,\\ B_{i+1}, & \text{if } q+1\le i\le p-1,\end{cases}$$

we get by application of Remark 1.19(b), that the last expression in (1.16) is equal to

$$n\cdot\mathbb{E}\circ\mathrm{Tr}_n\big[(B_1')^*B'_{\pi_0(1)}\cdots(B'_{p-1})^*B'_{\pi_0(p-1)}\big],$$

which, by Theorem 1.11, equals $n\cdot m^{k(\hat\pi_0)}n^{l(\hat\pi_0)}$. Arguing then as in the proof of (i), it follows that $k(\hat\pi)=k(\hat\pi_0)$ and $l(\hat\pi)=l(\hat\pi_0)+1$. This proves (ii).
The last statements of Proposition 1.22 follow immediately from (i), (ii) and Definition 1.10. $\square$

1.23 Proposition. *Let p be a positive integer, and let π be a permutation in S_p. By finitely many (or possibly none) successive cancellations of pairs, $\hat{\pi}$ can be reduced to either*
(i) $\hat{e}_1$, *where e_1 is the trivial permutation in S_1,*
or
(ii) $\hat{\rho}$, *where ρ is a permutation in S_q^{irr} for some q in $\{2,\dots,p\}$.*
Case (i) appears if and only if $\pi \in S_p^{\mathrm{nc}}$.

Proof. It is clear, that by finitely many (or possibly none) successive cancellations of pairs, $\hat{\pi}$ can be reduced to a permutation $\hat{\rho}$, where either $\rho \in S_1$ or $\rho \in S_q^{\mathrm{irr}}$ for some q in $\{2,3,\dots,p\}$. By Lemma 1.20, $\hat{\pi}$ is non-crossing if and only if $\hat{\rho}$ is. Since $S_1 = S_1^{\mathrm{nc}} = \{e_1\}$, and $S_q^{\mathrm{irr}} \cap S_q^{\mathrm{nc}} = \emptyset$ for all q in $\{2,3,\dots,p\}$, by Lemma 1.17, it follows thus, that either case (i) or case (ii) occurs, and that case (i) occurs if and only if $\hat{\pi}$ is non-crossing. $\square$

The following corollary is a special case of [Sh, Lemma 2.3]. For the convenience of the reader, we include a proof based on Propositions 1.22 and 1.23.

1.24 Corollary. *Let p be a positive integer and let π be a permutation in S_p. Then $\hat{\pi}$ is non-crossing if and only if $k(\hat{\pi}) + l(\hat{\pi}) = p+1$, or, equivalently, if and only if $\sigma(\hat{\pi}) = 0$.*

Proof. Assume first that $\hat{\pi}$ is non-crossing. It follows then from Proposition 1.23, that by successive cancellations of pairs, $\hat{\pi}$ may be reduced to $\hat{e}_1$, where e_1 is the unique permutation in S_1. Since $\sigma(\cdot)$ is invariant under cancellations of pairs, (cf. Lemma 1.22), it follows that $\sigma(\hat{\pi}) = \sigma(\hat{e}_1)$, and it is straightforward to check that $\sigma(\hat{e}_1) = 0$.
Assume next that $\hat{\pi}$ has a crossing. Then, by Proposition 1.23, there exist q in $\{2,\dots,p\}$ and a permutation ρ in S_q^{irr}, such that $\hat{\pi}$ may be reduced to $\hat{\rho}$ by finitely many (or possibly none) successive cancellations of pairs. By Proposition 1.22, $\sigma(\hat{\pi}) = \sigma(\hat{\rho})$, and hence it suffices to show that $\sigma(\hat{\rho}) > 0$, i.e., that $d(\hat{\rho}) < q+1$. Note for this, that since $\hat{\rho}$ is irreducible, $\hat{\rho}(j) \neq j+1$, for all j in $\{1,2,\dots,2q-1\}$. Since $\hat{\rho}^2 = \mathrm{id}$, this is equivalent to the condition that $\hat{\rho}(j) \neq j-1$, for all j in $\{2,3,\dots,2q\}$, and by Remark 1.9, this implies that $\mathrm{card}([j]_{\hat{\rho}}) \geq 2$, for all j in $\{2,3,\dots,2q\}$. Letting r denote the number of $\sim_{\hat{\rho}}$-equivalence classes, that are distinct from $[1]_{\hat{\rho}}$, we have thus the inequality

$$2r + \mathrm{card}([1]_{\hat{\rho}}) \leq 2q.$$

Since $r = d(\hat{\rho}) - 1$, and since $\mathrm{card}([1]_{\hat{\rho}}) \geq 1$, this implies that $2(d(\hat{\rho})-1)+1 \leq 2q$, and hence that $d(\hat{\rho}) \leq q$, as desired. $\square$

2 A Combinatorial Expression for the Moments of S^*S

Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, let r be a positive integer, and let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$. Moreover, let n be a fixed positive integer, and let

$Y_1, \dots, Y_r$ be independent elements of $\mathrm{GRM}(n, n, \frac{1}{n})$. We then define

$$S = \sum_{i=1}^{r} a_i \otimes Y_i.$$

Note that S is a random variable taking values in $\mathcal{B}(\mathcal{H}, \mathcal{K}) \otimes M_n(\mathbb{C})$. The aim of this section is to derive combinatorial expressions for the moments

$$(\mathrm{id}_{\mathcal{B}(\mathcal{H})} \otimes (\mathbb{E} \circ \mathrm{tr}_n))[(S^*S)^p] \quad \text{and} \quad (\mathrm{id}_{\mathcal{B}(\mathcal{H})} \otimes \mathbb{E})[(S^*S)^p], \quad (p \in \mathbb{N}),$$

where $\mathrm{id}_{\mathcal{B}(\mathcal{H})}$ denotes the identity mapping on $\mathcal{B}(\mathcal{H})$. Moreover, we shall obtain another combinatorial expression, which is an upper estimate for the norm of $(\mathrm{id}_{\mathcal{B}(\mathcal{H})} \otimes \mathbb{E})[(S^*S)^p]$. For the sake of short notation, in the following we shall just write $\mathbb{E} \circ \mathrm{tr}_n$ and $\mathbb{E}$ instead of $\mathrm{id}_{\mathcal{B}(\mathcal{H})} \otimes (\mathbb{E} \circ \mathrm{tr}_n)$ and $\mathrm{id}_{\mathcal{B}(\mathcal{H})} \otimes \mathbb{E}$.
We start with the following generalization of Proposition 1.5.

2.1 Proposition. *Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces, let r be a positive integer, and let $a_1, \dots, a_r$ be elements of $\mathcal{B}(\mathcal{H}, \mathcal{K})$. Moreover, let m, n be fixed positive integers, and let $B_1, \dots, B_r$ be independent elements of* $\mathrm{GRM}(m, n, 1)$. *Then with $T = \sum_{i=1}^{r} a_i \otimes B_i$, we have for any positive integer p, that*

$$\mathbb{E} \circ \mathrm{Tr}_n[(T^*T)^p] = \sum_{\pi \in S_p} m^{k(\hat{\pi})} n^{l(\hat{\pi})} \cdot \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}.$$

Proof. Let $(B(1,h))_{h \in \mathbb{N}}, \dots, (B(r,h))_{h \in \mathbb{N}}$ be sequences of elements from $\mathrm{GRM}(m, n, 1)$, such that (the entries of) the random matrices $B(i,h)$, $1 \le i \le r$, $h \in \mathbb{N}$, are jointly independent. Then for h in $\mathbb{N}$, we define

$$T_h = \sum_{i=1}^{r} a_i \otimes B(i,h).$$

Note then, that for each s in $\mathbb{N}$,

$$s^{-\frac{1}{2}} \sum_{h=1}^{s} T_h = s^{-\frac{1}{2}} \sum_{h=1}^{s} \sum_{i=1}^{r} a_i \otimes B(i,h) = \sum_{i=1}^{r} a_i \otimes \Big(s^{-\frac{1}{2}} \sum_{h=1}^{s} B(i,h)\Big),$$

where the random matrices $s^{-\frac{1}{2}} \sum_{h=1}^{s} B(1,h), \dots, s^{-\frac{1}{2}} \sum_{h=1}^{s} B(r,h)$ are independent elements of $\mathrm{GRM}(m, n, 1)$. It follows thus, that the moments of $s^{-1}(\sum_{h=1}^{s} T_h)^* \sum_{h=1}^{s} T_h$ w.r.t. $\mathbb{E} \circ \mathrm{Tr}_n$ are equal to those of T^*T. Thus for any p, s in $\mathbb{N}$, we have

$$\begin{aligned} \mathbb{E} \circ \mathrm{Tr}_n[(T^*T)^p] &= \mathbb{E} \circ \mathrm{Tr}_n \left[s^{-p} \Big(\Big(\sum_{h=1}^{s} T_h \Big)^* \sum_{h=1}^{s} T_h \Big)^p \right] \\ &= s^{-p} \cdot \sum_{\substack{1 \le h_1, h_2, \dots, h_p \le s \\ 1 \le g_1, g_2, \dots, g_p \le s}} \mathbb{E} \circ \mathrm{Tr}_n \left[T_{h_1}^* T_{g_1} T_{h_2}^* T_{g_2} \cdots T_{h_p}^* T_{g_p} \right]. \end{aligned} \tag{2.1}$$

Consider here an arbitrary $2p$–tuple $(h_1, g_1, \dots, h_p, g_p)$ in $\{1, 2, \dots, s\}^{2p}$. Recalling then the definition of T_h, we find that

$$\begin{aligned}
&\mathbb{E} \circ \mathrm{Tr}_n \Big[T_{h_1}^* T_{g_1} T_{h_2}^* T_{g_2} \cdots T_{h_p}^* T_{g_p}\Big] \\
&= \sum_{\substack{1 \le i_1, \dots, i_p \le r \\ 1 \le j_1, \dots, j_p \le r}} (a_{i_1}^* a_{j_1} \cdots a_{i_p}^* a_{j_p}) \cdot \mathbb{E}\big[\mathrm{Tr}_n(B(i_1, h_1)^* B(j_1, g_1) \cdots B(i_p, h_p)^* B(j_p, g_p))\big].
\end{aligned}$$

Since $B(i, h)$ is independent of $B(j, g)$ unless $i = j$ and $h = g$, it follows here from Lemma 1.2 in Section 1, that

$$\begin{aligned}
&\mathbb{E} \circ \mathrm{Tr}_n\, [B(i_1, h_1)^* B(j_1, g_1) \cdots B(i_p, h_p)^* B(j_p, g_p)] \neq 0 \\
&\Longrightarrow \exists \pi \in S_p : \ (j_1, g_1) = (i_{\pi(1)}, h_{\pi(1)}), \dots, (j_p, g_p) = (i_{\pi(p)}, h_{\pi(p)}).
\end{aligned} \tag{2.2}$$

In particular it follows that in (2.1), we only have to sum over $(h_1, g_1, \dots, h_p, g_p)$ in $\cup_{\pi \in S_p} M(\pi, s)$, where, as in the proof of Proposition 1.5 in Section 1 ,

$$M(\pi, s) = \big\{(h_1, g_1, \dots, h_p, g_p) \in \{1, 2, \dots, s\}^{2p} \bigm| g_1 = h_{\pi(1)}, \dots, g_p = h_{\pi(p)}\big\},$$

for any π in S_p. Following still the proof of Proposition 1.5 in Section 1, we define,

$$\mathcal{D}(s) = \big\{(h_1, g_1, \dots, h_p, g_p) \in \{1, 2, \dots, s\}^{2p} \bigm| h_1, \dots, h_p \text{ are distinct } \big\},$$

and then the sets $M(\pi, s) \cap \mathcal{D}(s), \ \pi \in S_p$, are disjoint and

$$\begin{aligned}
&\mathbb{E} \circ \mathrm{Tr}_n[(T^*T)^p] \\
&\quad = s^{-p} \sum_{\pi \in S_p} \sum_{(h_1, g_1, \dots, h_p, g_p) \in M(\pi, s) \cap \mathcal{D}(s)} \mathbb{E} \circ \mathrm{Tr}_n[T_{h_1}^* T_{g_1} \cdots T_{h_p}^* T_{g_p}] \\
&\qquad + s^{-p} \sum_{(h_1, g_1, \dots, h_p, g_p) \in \left(\cup_{\pi \in S_p} M(\pi, s)\right) \setminus \mathcal{D}(s)} \mathbb{E} \circ \mathrm{Tr}_n[T_{h_1}^* T_{g_1} \cdots T_{h_p}^* T_{g_p}].
\end{aligned} \tag{2.3}$$

As was noted in the proof of Proposition 1.5, we have here that

$$s^{-p} \cdot \mathrm{card}(M(\pi, s) \cap \mathcal{D}(s)) \to 1, \text{ as } s \to \infty, \quad (\pi \in S_p), \tag{2.4}$$

and that

$$s^{-p} \cdot \mathrm{card}(\big(\cup_{\pi \in S_p} M(\pi, s)\big) \setminus \mathcal{D}(s)) \to 0, \text{ as } s \to \infty. \tag{2.5}$$

Moreover, for any $h_1, g_1, \dots, h_p, g_p$ in $\mathbb{N}$, we have that

$$\begin{aligned}
&\big\|\mathbb{E}\circ\mathrm{Tr}_n[T_{h_1}^*T_{g_1}\cdots T_{h_p}^*T_{g_p}]\big\| \\
&\le \sum_{\substack{1\le i_1,\dots,i_p\le r\\ 1\le j_1,\dots,j_p\le r}} \|a_{i_1}^*a_{j_1}\cdots a_{i_p}^*a_{j_p}\|\cdot\big|\mathbb{E}\big[\mathrm{Tr}_n(B(i_1,h_1)^*B(j_1,g_1)\cdots B(i_p,h_p)^*B(j_p,g_p))\big]\big| \\
&\le K(m,n,p,1)\cdot \sum_{\substack{1\le i_1,\dots,i_p\le r\\ 1\le j_1,\dots,j_p\le r}} \|a_{i_1}^*a_{j_1}\cdots a_{i_p}^*a_{j_p}\|,
\end{aligned}$$

where $K(m,n,p,1)$ is the constant introduced in Remark 1.4 in Section 1. Since this constant does not depend on s, it follows thus, by (2.5), that the second term on the right hand side of (2.3) tends to 0 as $s\to\infty$.
Regarding the first term on the right hand side of (2.3), for any π in S_p and any $2p$–tuple $(h_1, g_1, \dots, h_p, g_p)$ in $M(\pi, s)\cap\mathcal{D}(s)$, we have that

$$\begin{aligned}
\mathbb{E}\circ\mathrm{Tr}_n[T_{h_1}^*T_{g_1}\cdots T_{h_p}^*T_{g_p}] &= \mathbb{E}\circ\mathrm{Tr}_n[T_{h_1}^*T_{h_{\pi(1)}}\cdots T_{h_p}^*T_{h_{\pi(p)}}] \\
= \sum_{\substack{1\le i_1,\dots,i_p\le r\\ 1\le j_1,\dots,j_p\le r}} &(a_{i_1}^*a_{j_1}\cdots a_{i_p}^*a_{j_p}) \\
&\cdot\mathbb{E}\big[\mathrm{Tr}_n(B(i_1,h_1)^*B(j_1,h_{\pi(1)})\cdots B(i_p,h_p)^*B(j_p,h_{\pi(p)}))\big].
\end{aligned}$$

Recalling here the statement (2.2) and that $h_1, \dots, h_p$ are distinct, it follows that the term in the above sum corresponding to $(i_1, j_1, \dots, i_p, j_p)$ is 0, unless $j_1 = i_{\pi(1)}, \dots, j_p = i_{\pi(p)}$. Thus we have that

$$\begin{aligned}
&\mathbb{E}\circ\mathrm{Tr}_n[T_{h_1}^*T_{g_1}\cdots T_{h_p}^*T_{g_p}] \\
&\quad= \sum_{1\le i_1,\dots,i_p\le r}(a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_p}^*a_{i_{\pi(p)}}) \\
&\qquad\qquad\cdot\mathbb{E}\big[\mathrm{Tr}_n(B(i_1,h_1)^*B(i_{\pi(1)},h_{\pi(1)})\cdots B(i_p,h_p)^*B(i_{\pi(p)},h_{\pi(p)}))\big].
\end{aligned}$$

Note here, that since $h_1, \dots, h_p$ are distinct, $B(i_1, h_1), \dots, B(i_p, h_p)$ are independent for any choice of $i_1, \dots, i_p$ in $\{1, 2, \dots, r\}$, and consequently

$$\mathbb{E}\big[\mathrm{Tr}_n(B(i_1,h_1)^*B(i_{\pi(1)},h_{\pi(1)})\cdots B(i_p,h_p)^*B(i_{\pi(p)},h_{\pi(p)}))\big] = \Lambda(\pi,m,n),$$

for any $i_1, \dots, i_p$ in $\{1, 2, \dots, r\}$. Thus, we may conclude that

$$\mathbb{E}\circ\mathrm{Tr}_n[T_{h_1}^*T_{g_1}\cdots T_{h_p}^*T_{g_p}] = \Lambda(\pi,m,n)\cdot\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_p}^*a_{i_{\pi(p)}},$$

and this holds for any $(h_1, g_1, \dots, h_p, g_p)$ in $M(\pi, s)\cap\mathcal{D}(s)$. Therefore the first term on the right hand side of (2.3) equals

$$\sum_{\pi\in S_p} s^{-p}\cdot\mathrm{card}(M(\pi,s)\cap\mathcal{D}(s))\cdot\Lambda(\pi,m,n)\cdot\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_p}^*a_{i_{\pi(p)}},$$

and by (2.4), this tends to

$$\sum_{\pi \in S_p} \Lambda(\pi, m, n) \cdot \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}},$$

as $s \to \infty$. Since the left hand side of (2.3) does not depend on s, we get thus by letting $s \to \infty$ in (2.3), that

$$\mathbb{E} \circ \mathrm{Tr}_n[(T^*T)^p] = \sum_{\pi \in S_p} \Lambda(\pi, m, n) \cdot \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}.$$

Combining finally with Theorem 1.11, we obtain the desired formula. □

2.2 Corollary. *Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces, let r be a positive integer, and let $a_1, \dots, a_r$ be elements of $\mathcal{B}(\mathcal{H}, \mathcal{K})$. Moreover, let n be a fixed positive integer, and let $Y_1, \dots, Y_r$ be independent elements of* $\mathrm{GRM}(n, n, \frac{1}{n})$. *Then with $S = \sum_{i=1}^r a_i \otimes Y_i$, we have for any positive integer p, that*

$$\mathbb{E} \circ \mathrm{tr}_n[(S^*S)^p] = \sum_{\pi \in S_p} n^{-2\sigma(\hat{\pi})} \cdot \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}, \tag{2.6}$$

where $\sigma(\hat{\pi})$ is the quantity introduced in Definition 1.10 in Section 1.

Proof. With $B_i = \sqrt{n} \cdot Y_i$, $i \in \{1, 2, \dots, r\}$, we have that $B_1, \dots, B_r$ are independent elements of $\mathrm{GRM}(n, n, 1)$. It follows thus from Proposition 2.1, that for any p in $\mathbb{N}$,

$$n^p \cdot \mathbb{E} \circ \mathrm{Tr}_n[(S^*S)^p] = \sum_{\pi \in S_p} n^{k(\hat{\pi}) + l(\hat{\pi})} \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}},$$

and consequently

$$\mathbb{E} \circ \mathrm{tr}_n[(S^*S)^p] = \sum_{\pi \in S_p} n^{-p-1+k(\hat{\pi})+l(\hat{\pi})} \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}.$$

Formula (2.6) now follows by noting that,

$$p + 1 - k(\hat{\pi}) - l(\hat{\pi}) = p + 1 - d(\hat{\pi}) = 2\sigma(\hat{\pi}),$$

for any π in S_p. □

Our next objective is to derive a matrix version of formula (2.6). In other words, we shall obtain a combinatorial expression for $\mathbb{E}[(S^*S)^p]$.

2.3 Lemma. *Let n, r be positive integers and let $Y_1, \dots, Y_r$ be independent elements of* $\mathrm{GRM}(n, n, \sigma^2)$. *Then for any (non–random) unitary $n \times n$ matrices $u_1, \dots, u_r$, the random matrices $u_1 Y_1 u_1^*, \dots, u_r Y_r u_r^*$ are again independent elements of* $\mathrm{GRM}(n, n, \sigma^2)$.

Proof. Note first that for each i in $\{1, 2, \dots, r\}$, the entries of $u_i Y_i u_i^*$ are all measurable w.r.t. the σ–algebra generated by the entries of Y_i. It follows therefore immediately that $u_1 Y_1 u_1^*, \dots, u_r Y_r u_r^*$ are independent random matrices. We note next, that it follows easily from Definition 1.1, that the joint distribution of the entries of an element from $\mathrm{GRM}(n, n, \sigma^2)$ has the following density w.r.t. Lebesgue measure on $M_n(\mathbb{C}) \simeq \mathbb{R}^{2n^2}$:

$$y \mapsto \big(\tfrac{1}{\pi\sigma^2}\big)^{n^2} \exp\big(-\tfrac{1}{\sigma^2} \cdot \mathrm{Tr}_n(y^* y)\big), \quad (y \in M_n(\mathbb{C})). \tag{2.7}$$

(Here the identification of $M_n(\mathbb{C})$ with $\mathbb{R}^{2n^2}$ is given by

$$y \mapsto (\mathrm{Re}(y_{jk}), \mathrm{Im}(y_{jk}))_{1 \le j,k \le n}.)$$

Now let u be a unitary $n \times n$ matrix, and consider then the linear mapping

$$\mathrm{Ad}u \colon y \mapsto uyu^* \colon M_n(\mathbb{C}) \to M_n(\mathbb{C}).$$

Under the identification of $M_n(\mathbb{C})$ with $\mathbb{R}^{2n^2}$, the Euclidean structure on $\mathbb{R}^{2n^2}$ is given by the inner product:

$$\langle y, z \rangle = \mathrm{Re}(\mathrm{Tr}_n(z^* y)), \quad (y, z \in M_n(\mathbb{C})).$$

Thus $\mathrm{Ad}u \colon \mathbb{R}^{2n^2} \to \mathbb{R}^{2n^2}$ is a linear isometry, and hence the Jacobi determinant of this mapping equals 1. Combining this fact with (2.7) and the usual transformation theorem for Lebesgue measure, it follows that for any Y in $\mathrm{GRM}(n, n, \sigma^2)$, the joint distribution of the entries of uYu^* equals that of the entries of Y. □

2.4 Lemma. *Let $\mathcal{B}$ be a C^*–algebra with unit $\mathbf{1}$, let n be a positive integer, and consider the tensor product $\mathcal{B} \otimes M_n(\mathbb{C})$. If $x \in \mathcal{B} \otimes M_n(\mathbb{C})$, such that $(\mathbf{1} \otimes u)x(\mathbf{1} \otimes u)^* = x$ for any unitary $n \times n$ matrix u, then $x \in \mathcal{B} \otimes \mathbf{1}_n$.*

Proof. Assume that $x \in \mathcal{B} \otimes M_n(\mathbb{C})$, and that $(\mathbf{1} \otimes u)x(\mathbf{1} \otimes u)^* = x$ for any unitary $n \times n$ matrix u. Since $M_n(\mathbb{C})$ is the linear span of its unitaries, it follows that

$$x \in \big\{ y \in \mathcal{B} \otimes M_n(\mathbb{C}) \;\big|\; yT = Ty \quad \text{for all } T \text{ in } \mathbf{1} \otimes M_n(\mathbb{C}) \big\} = \mathcal{B} \otimes \mathbf{1}_n,$$

where the last equality follows by standard matrix considerations; thinking of $\mathcal{B} \otimes M_n(\mathbb{C})$ as the set of $n \times n$ matrices with entries from $\mathcal{B}$. □

2.5 Proposition. *Let S be as in Corollary 2.2. Then for any positive integer p, we have that:*

$$\mathbb{E}[(S^* S)^p] = \Big(\sum_{\pi \in S_p} n^{-2\sigma(\hat{\pi})} \cdot \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}} \Big) \otimes \mathbf{1}_n.$$

Proof. Let u be an arbitrary unitary $n \times n$ matrix, and define: $S_u = \sum_{i=1}^{r} a_i \otimes (uY_iu^*)$. Note then that $S_u^* S_u = (\mathbf{1}_{\mathcal{H}} \otimes u) S^* S (\mathbf{1}_{\mathcal{H}} \otimes u)^*$, where $\mathbf{1}_{\mathcal{H}}$ denotes the unit of $\mathcal{B}(\mathcal{H})$. It follows now by Lemma 2.3, that

$$\begin{aligned}\mathbb{E}[(S^*S)^p] &= \mathbb{E}[(S_u^*S_u)^p] \\ &= \mathbb{E}[(\mathbf{1}_{\mathcal{H}} \otimes u)(S^*S)^p(\mathbf{1}_{\mathcal{H}} \otimes u)^*] = (\mathbf{1}_{\mathcal{H}} \otimes u)\mathbb{E}[(S^*S)^p](\mathbf{1}_{\mathcal{H}} \otimes u)^*.\end{aligned}$$

Since this holds for any unitary u, it follows from Lemma 2.4, that $\mathbb{E}[(S^*S)^p] \in \mathcal{B}(\mathcal{H}) \otimes \mathbf{1}_n$, and consequently

$$\mathbb{E}[(S^*S)^p] = \Big(\mathrm{tr}_n(\mathbb{E}[(S^*S)^p])\Big) \otimes \mathbf{1}_n = \Big(\mathbb{E} \circ \mathrm{tr}_n[(S^*S)^p]\Big) \otimes \mathbf{1}_n.$$

The proposition now follows by application of Corollary 2.2. □

In the next section, we shall obtain combinatorial expressions that are upper estimates for the moments $\mathbb{E}[(S^*S)^p]$. It follows from Proposition 2.5, that in order to obtain such combinatorial estimates, we should concentrate on deriving combinatorial estimates for the quantities

$$\Big\| \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}} \Big\|,$$

where $\pi \in S_p$, and $a_1, \dots, a_r$ are arbitrary bounded operators from a Hilbert space $\mathcal{H}$ to a Hilbert space $\mathcal{K}$.

2.6 Definition. Let p be a positive integer, let π be a permutation in S_p and consider the permutation $\hat{\pi}$ in S_{2p}. We then put

$$\begin{aligned}\kappa(\hat{\pi}) &= \mathrm{card}(\big\{j \in \{1, 3, \dots, 2p-1\} \bigm| \hat{\pi}(j) > j\big\}), \\ \lambda(\hat{\pi}) &= \mathrm{card}(\big\{j \in \{1, 3, \dots, 2p-1\} \bigm| \hat{\pi}(j) < j\big\}) + 1.\end{aligned}$$ □

We note, that since $\hat{\pi}$ has no fixed points, it follows that

$$\kappa(\hat{\pi}) + \lambda(\hat{\pi}) = p + 1, \quad (p \in \mathbb{N},\ \pi \in S_p). \tag{2.8}$$

Recalling that by definition of $\hat{\pi}$, $\hat{\pi}(2h-1) = 2\pi^{-1}(h)$ for all h in $\{1, 2, \dots, p\}$, it follows furthermore that

$$\begin{aligned}\kappa(\hat{\pi}) &= \mathrm{card}(\big\{h \in \{1, 2, \dots, p\} \bigm| 2\pi^{-1}(h) > 2h - 1\big\}) \\ &= \mathrm{card}(\big\{h \in \{1, 2, \dots, p\} \bigm| \pi^{-1}(h) \ge h\big\}) \\ &= \mathrm{card}(\big\{h \in \{1, 2, \dots, p\} \bigm| h \ge \pi(h)\big\}),\end{aligned} \tag{2.9}$$

where the last equality follows by replacing h by $\pi^{-1}(h)$. Similarly we have that

$$\begin{aligned}\lambda(\hat{\pi}) &= p + 1 - \kappa(\hat{\pi}) \\ &= \mathrm{card}(\big\{h \in \{1, 2, \dots, p\} \bigm| \pi^{-1}(h) < h\big\}) + 1 \\ &= \mathrm{card}(\big\{h \in \{1, 2, \dots, p\} \bigm| h < \pi(h)\big\}) + 1.\end{aligned} \tag{2.10}$$

We note also, that since $\hat{\pi}(j) = j+1$ mod. 2 and $\hat{\pi}(\hat{\pi}(j)) = j$ for all j, we have that

$$\begin{aligned}\kappa(\hat{\pi}) &= \mathrm{card}(\hat{\pi}\big[\{j \in \{1,3,\dots,2p-1\} \bigm| \hat{\pi}(j) > j\}\big]) \\ &= \mathrm{card}(\{j \in \{2,4,\dots,2p\} \bigm| \hat{\pi}(j) < j\}),\end{aligned} \tag{2.11}$$

and similarly

$$\lambda(\hat{\pi}) = \mathrm{card}(\{j \in \{2,4,\dots,2p\} \bigm| \hat{\pi}(j) > j\}) + 1. \tag{2.12}$$

In connection with products of the form $a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}$, note that $\kappa(\hat{\pi})$ denotes the number of h's in $\{1,2,\dots,p\}$ for which the factor $a_{i_h}^*$ appears before the factor a_{i_h} in this product. Similarly $\lambda(\hat{\pi}) - 1$ denotes the number of h's in $\{1,2,\dots,p\}$ for which the factor a_{i_h} appears before the factor $a_{i_h}^*$.

2.7 Proposition. *Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces, let r be a positive integer, and let $a_1, \dots, a_r$ be elements of $\mathcal{B}(\mathcal{H}, \mathcal{K})$. Let further c and d be positive real numbers, such that*

$$\Big\| \sum_{i=1}^r a_i^* a_i \Big\| \le c \qquad \text{and} \qquad \Big\| \sum_{i=1}^r a_i a_i^* \Big\| \le d. \tag{2.13}$$

Then for any positive integer p and any permutation π in S_p, we have that

$$\Big\| \sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}} \Big\| \le c^{\kappa(\hat{\pi})} d^{\lambda(\hat{\pi})-1}.$$

Proof. Let $\mathcal{V}$ be an infinite dimensional Hilbert space, and choose r isometries $s_1, \dots, s_r$ in $\mathcal{B}(\mathcal{V})$, with orthogonal ranges, i.e.,

$$s_i^* s_j = \delta_{i,j} \mathbf{1}_{\mathcal{B}(\mathcal{V})}, \qquad (i,j \in \{1,2,\dots,r\}). \tag{2.14}$$

Consider then the Hilbert space $\tilde{\mathcal{V}} = \mathcal{V} \otimes \cdots \otimes \mathcal{V}$ (p factors), and for each i in $\{1,2,\dots,r\}$ and h in $\{1,2,\dots,p\}$, define the operator $s(i,h)$ in $\mathcal{B}(\tilde{\mathcal{V}})$ by the equation

$$s(i,h) = \mathbf{1}_{\mathcal{B}(\mathcal{V})} \otimes \cdots \otimes \mathbf{1}_{\mathcal{B}(\mathcal{V})} \otimes \underset{\substack{\uparrow \\ h\text{'th position}}}{S_i} \otimes \mathbf{1}_{\mathcal{B}(\mathcal{V})} \otimes \cdots \otimes \mathbf{1}_{\mathcal{B}(\mathcal{V})}. \tag{2.15}$$

Next, put

$$t(i,h) = \begin{cases} s(i,h), & \text{if } h \le \pi^{-1}(h), \\ s(i,h)^*, & \text{if } h > \pi^{-1}(h), \end{cases} \qquad (i \in \{1,2,\dots,r\},\ h \in \{1,2,\dots,p\}), \tag{2.16}$$

and

$$A_h = \sum_{i=1}^{r} a_i \otimes t(i,h), \qquad (h \in \{1,2,\dots,p\}). \tag{2.17}$$

We consider A_h as an element of $\mathcal{B}(\mathcal{H}\otimes\tilde{\mathcal{V}},\mathcal{K}\otimes\tilde{\mathcal{V}})$ in the usual way. We claim then that

$$A_1^*A_{\pi(1)}A_2^*A_{\pi(2)}\cdots A_p^*A_{\pi(p)} = \Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big)\otimes \mathbf{1}_{\mathcal{B}(\tilde{\mathcal{V}})}. \tag{2.18}$$

To prove (2.18), observe first that

$$\begin{aligned} &A_1^*A_{\pi(1)}\cdots A_p^*A_{\pi(p)} \\ &\quad= \sum_{\substack{1\le i_1,i_2,\dots,i_p\le r\\ 1\le j_1,j_2,\dots,j_p\le r}} (a_{i_1}^*a_{j_1}a_{i_2}^*a_{j_2}\cdots a_{i_p}^*a_{j_p})\otimes \Pi(i_1,j_1,i_2,j_2,\dots,i_p,j_p), \end{aligned} \tag{2.19}$$

where

$$\begin{aligned} &\Pi(i_1,j_1,\dots,i_p,j_p) \\ &\quad= t(i_1,1)^*t(j_1,\pi(1))t(i_2,2)^*t(j_2,\pi(2))\cdots t(i_p,p)^*t(j_p,\pi(p)), \end{aligned} \tag{2.20}$$

for all $i_1,j_1,\dots,i_p,j_p$ in $\{1,2,\dots,r\}$. By (2.15) and (2.16), $t(i,h)$ and $t(i,h)^*$ both commute with $t(j,k)$ and $t(j,k)^*$, as long as $h\neq k$. Hence, we can reorder the factors in the product on the right hand side of (2.20), according to the second index in $t(\cdot,\cdot)$ and $t(\cdot,\cdot)^*$, in the following way

$$\Pi(i_1,j_1,\dots,i_p,j_p) = T(1)T(2)\cdots T(p),$$

where

$$T(h) = \begin{cases} t(i_h,h)^*t(j_{\pi^{-1}(h)},h), & \text{if } h\le \pi^{-1}(h),\\ t(j_{\pi^{-1}(h)},h)t(i_h,h)^*, & \text{if } h> \pi^{-1}(h),\end{cases}$$

for each h in $\{1,2,\dots,p\}$. By (2.16), it follows that

$$T(h) = \begin{cases} s(i_h,h)^*s(j_{\pi^{-1}(h)},h), & \text{if } h\le \pi^{-1}(h),\\ s(j_{\pi^{-1}(h)},h)^*s(i_h,h), & \text{if } h> \pi^{-1}(h),\end{cases}$$

and thus by (2.14)-(2.15), we get that for all $i_1,j_1,\dots,i_p,j_p$ in $\{1,2,\dots,r\}$ and all h in $\{1,2,\dots,p\}$,

$$T(h) = \begin{cases} \mathbf{1}_{\mathcal{B}(\tilde{\mathcal{V}})}, & \text{if } i_h = j_{\pi^{-1}(h)},\\ 0, & \text{if } i_h \neq j_{\pi^{-1}(h)}.\end{cases}$$

Therefore, $\Pi(i_1,j_1,\dots,i_p,j_p)=0$, unless $i_h = j_{\pi^{-1}(h)}$, for all h in $\{1,2,\dots,p\}$, or equivalently, unless $i_{\pi(h)} = j_h$, for all h in $\{1,2,\dots,p\}$, in which case $\Pi(i_1,j_1,\dots,i_p,j_p) = \mathbf{1}_{\mathcal{B}(\tilde{\mathcal{V}})}$. Combining this with (2.19), we obtain (2.18).

Using again that $s(i,h)^*s(j,h) = \delta_{i,j}\mathbf{1}_{\mathcal{B}(\tilde{\mathcal{V}})}$, for all i,j in $\{1,2,\dots,r\}$, we get that if $h \le \pi^{-1}(h)$,

$$A_h^*A_h = \sum_{i,j=1}^{r} a_i^*a_j \otimes s(i,h)^*s(j,h) = \sum_{i=1}^{r} a_i^*a_i \otimes \mathbf{1}_{\mathcal{B}(\tilde{\mathcal{V}})},$$

and if $h > \pi^{-1}(h)$,

$$A_hA_h^* = \sum_{i=1}^{r} a_ia_i^* \otimes \mathbf{1}_{\mathcal{B}(\tilde{\mathcal{V}})}.$$

By (2.13), it follows thus, that

$$\begin{aligned} \|A_h\|^2 &= \|A_h^*A_h\| \le c, \qquad \text{if } h \le \pi^{-1}(h),\\ \|A_h\|^2 &= \|A_hA_h^*\| \le d, \qquad \text{if } h > \pi^{-1}(h), \end{aligned}$$

so by (2.9) and (2.10),

$$\|A_1^*A_{\pi(1)}\cdots A_p^*A_{\pi(p)}\| \le \prod_{h=1}^{p}\|A_h\|^2 \le c^{\kappa(\hat{\pi})}d^{\lambda(\hat{\pi})-1}.$$

Together with (2.18), this proves the proposition. □

2.8 Corollary. *Let $\mathcal{H},\mathcal{K}$ be Hilbert spaces, let r be a positive integer, and let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$. Moreover, let n be a fixed positive integer, and let $Y_1,\dots,Y_r$ be independent elements of* $\mathrm{GRM}(n,n,\frac{1}{n})$. *Then with $S=\sum_{i=1}^r a_i\otimes Y_i$, $c=\|\sum_{i=1}^r a_i^*a_i\|$ and $d=\|\sum_{i=1}^r a_ia_i^*\|$, we have for any positive integer p, that*

$$\|\mathbb{E}[(S^*S)^p]\| \le \sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}c^{\kappa(\hat{\pi})}d^{\lambda(\hat{\pi})-1}.$$

Proof. This follows immediately by combining Propositions 2.5 and 2.7. □

In Section 3 we shall estimate further the quantity $\|\mathbb{E}[(S^*S)^p]\|$. As preparation for this, we will in Proposition 2.10 below, compare the numbers $\kappa(\hat{\pi})$ and $\lambda(\hat{\pi})$ with the numbers $k(\hat{\pi})$ and $l(\hat{\pi})$, defined in Section 1.

2.9 Lemma. *Let p be a positive integer, let π be a permutation in S_p, and consider the permutation $\hat{\pi}$ in S_{2p} and the corresponding equivalence relation $\sim_{\hat{\pi}}$. Then any equivalence class for $\sim_{\hat{\pi}}$, except possibly $[1]_{\hat{\pi}}$, contains an element j with the property that $\hat{\pi}(j) < j$.*

Proof. Let j' be an element of $\{1,2,\dots,2p\}$, such that $1 \notin [j']_{\hat{\pi}}$. We show that $[j']_{\hat{\pi}}$ contains an element j such that $\hat{\pi}(j) < j$. For this, note first, that we may assume that j' is the smallest element of $[j']_{\hat{\pi}}$. Then, by assumption, $j' \ge 2$. Now write in the usual manner (cf. Remark 1.9)

$$[j']_{\hat{\pi}} = \{j_0, j_1, \dots, j_q\}.$$

In particular, $\hat{\pi}(j_q) + 1 = j_0 = j'$ (addition formed mod. $2p$). Now, since $j' \geq 2$, we have that $j' - 1 < j'$, even when the subtraction is formed mod. $2p$. Therefore, since j' is the smallest element of $[j']_{\hat{\pi}}$, $\hat{\pi}(j_q) = j' - 1 < j' \leq j_q$. Thus we may choose $j = j_q$. □

2.10 Proposition. *Let p be a positive integer, let π be a permutation in S_p, and consider the permutation $\hat{\pi}$ in S_{2p}. We then have*
(i) $\kappa(\hat{\pi}) \geq k(\hat{\pi})$ *and* $\lambda(\hat{\pi}) \geq l(\hat{\pi})$.
(ii) $(\kappa(\hat{\pi}) - k(\hat{\pi})) + (\lambda(\hat{\pi}) - l(\hat{\pi})) = 2\sigma(\hat{\pi})$.
(iii) $\kappa(\hat{\pi}) = k(\hat{\pi})$ *and* $\lambda(\hat{\pi}) = l(\hat{\pi})$ *if and only if $\hat{\pi}$ is non–crossing.*

Proof. (i) By Lemma 2.9 and the definition of $l(\hat{\pi})$, it follows that

$$l(\hat{\pi}) - 1 \leq \operatorname{card}(\{j \in \{1, 3, \dots, 2p-1\} \mid \hat{\pi}(j) < j\}) = \lambda(\hat{\pi}) - 1.$$

Similarly we find by application of (2.11), that

$$k(\hat{\pi}) \leq \operatorname{card}(\{j \in \{2, 4, \dots, 2p\} \mid \hat{\pi}(j) < j\}) = \kappa(\hat{\pi}).$$

(ii) We find by application of (2.8), that

$$(\kappa(\hat{\pi}) - k(\hat{\pi})) + (\lambda(\hat{\pi}) - l(\hat{\pi})) = (\kappa(\hat{\pi}) + \lambda(\hat{\pi})) - d(\hat{\pi}) = p + 1 - d(\hat{\pi}) = 2\sigma(\hat{\pi}).$$

(iii) This follows immediately by combining (i), (ii) and Corollary 1.24. □

3 An upper bound for $\mathbb{E}\big[\exp(tS^*S)\big]$, $t \geq 0$

In the previous section, we computed $\mathbb{E}[(S^*S)^p]$, for p in $\mathbb{N}$ and $S = \sum_{i=1}^r a_i \otimes Y_i$, where $a_1, \dots, a_r \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, for Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, and where $Y_1, \dots, Y_r$ are independent random matrices in $\mathrm{GRM}(n, n, \frac{1}{n})$. For fixed p in $\mathbb{N}$, the function $\omega \mapsto \big(S^*(\omega)S(\omega)\big)^p$ only takes values in a finite dimensional subspace of $\mathcal{B}(\mathcal{H}) \otimes M_n(\mathbb{C})$. This is not the case for the function $\omega \mapsto \exp\big(tS^*(\omega)S(\omega)\big)$, so in order to give precise meaning to the mean $\mathbb{E}\big[\exp(tS^*S)\big]$, we will need the following definition (cf. [Ru, Definition 3.26]).

3.1 Definition. Let $\mathcal{X}$ be a Banach space, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $f \colon \Omega \to \mathcal{X}$ be a mapping, that satisfies the following two conditions
(a) $\forall \varphi \in \mathcal{X}^* \colon \varphi \circ f \in L^1(\Omega, \mathcal{F}, P)$
(b) $\exists x_0 \in \mathcal{X}\ \forall \varphi \in \mathcal{X}^* \colon \int_\Omega \varphi \circ f(\omega)\, dP(\omega) = \varphi(x_0)$.
We say then that f is *integrable* in $\mathcal{X}$, and we call x_0 the integral of f, and write

$$\mathbb{E}(f) = \int_\Omega f\, dP = x_0. \qquad \square$$

Note that in the above definition, x_0 is uniquely determined by (b). Note also, that we do not require that $\int_\Omega \|f\|\, dP < \infty$, in order for f to be integrable. However, if $\mathcal{X}$ is finite dimensional, then this follows automatically from (a).

3.2 Proposition. *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, let $a_1, \dots, a_r$ be elements of $B(\mathcal{H},\mathcal{K})$, and let γ be a strictly positive number, such that*

$$\max\left\{\|\textstyle\sum_{i=1}^r a_i^* a_i\|, \|\sum_{i=1}^r a_i a_i^*\|\right\} \leq \gamma.$$

Furthermore, let n be a positive integer, let $Y_1, \dots, Y_r$ be independent random matrices in $\mathrm{GRM}(n,n,\frac{1}{n})$, and put $S = \sum_{i=1}^r a_i \otimes Y_i$.
Then for any complex number t, such that $|t| < \frac{n}{\gamma}$, the function

$$\omega \mapsto \exp\big(tS^*(\omega)S(\omega)\big), \qquad (\omega \in \Omega),$$

is integrable in $\mathcal{B}(\mathcal{H}^n)$, in the sense of Definition 3.1, and

$$\mathbb{E}\big[\exp(tS^*S)\big] = \sum_{p=0}^{\infty} \tfrac{t^p}{p!}\mathbb{E}\big[(S^*S)^p\big], \tag{3.1}$$

where the series on the right hand side is absolutely convergent in $\mathcal{B}(\mathcal{H}^n)$.

Proof. By Proposition 2.5, we have for any p in $\mathbb{N}$,

$$\mathbb{E}\big[(S^*S)^p\big] = \Big(\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})} \sum_{1\leq i_1,\dots,i_p\leq r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}\Big) \otimes \mathbf{1}_n,$$

and by Proposition 2.7 and formula (2.8), we have here for all π in S_p, that

$$\Big\| \sum_{1\leq i_1,\dots,i_p\leq r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}} \Big\| \leq \gamma^p. \tag{3.2}$$

Hence the absolute convergence of the right hand side of (3.1) will follow, if we can prove that

$$1 + \sum_{p=1}^{\infty} \tfrac{(\gamma|t|)^p}{p!}\Big(\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\Big) < \infty, \tag{3.3}$$

whenever $|t| < \frac{n}{\gamma}$. For this, consider an element B of $\mathrm{GRM}(n,n,1)$, and recall then from Corollary 1.12, that

$$\mathbb{E}\circ\mathrm{Tr}_n\big[(B^*B)^p\big] = \sum_{\pi\in S_p} n^{k(\hat{\pi})+l(\hat{\pi})} = n^{p+1}\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}.$$

Hence for positive numbers s, we have

$$\mathbb{E}\circ\mathrm{Tr}_n\big[\exp(sB^*B)\big] = n\Big(1 + \sum_{p=1}^{\infty} \tfrac{(ns)^p}{p!}\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\Big). \tag{3.4}$$

From [HT, Theorem 6.4], we know that

$$\mathbb{E}\circ\mathrm{Tr}_n\big[\exp(sB^*B)\big] < \infty, \quad \text{when} \quad 0 \leq s < 1.$$

Hence the sum in (3.4) is finite, whenever $0 \le s < 1$, and this implies that (3.3) holds whenever $|t| < \frac{n}{\gamma}$.

Consider now the state space $\mathcal{S}(\mathcal{B}(\mathcal{H}^n))$ of $\mathcal{B}(\mathcal{H}^n)$ and an element φ of $\mathcal{S}(\mathcal{B}(\mathcal{H}^n))$. For any ω in Ω, we have then that

$$\varphi\big[\exp\big(tS^*(\omega)S(\omega)\big)\big] = \sum_{p=0}^{\infty} \tfrac{t^p}{p!}\varphi\big[(S^*(\omega)S(\omega))^p\big],$$

which is clearly a positive measurable function of ω (since φ is a state). Moreover, by Lebesgue's Monotone Convergence Theorem,

$$\begin{aligned}
\mathbb{E}\big[\varphi\big(\exp(tS^*S)\big)\big] &= \sum_{p=0}^{\infty} \tfrac{t^p}{p!}\mathbb{E}\big[\varphi\big((S^*S)^p\big)\big] \\
&= \sum_{p=0}^{\infty} \tfrac{t^p}{p!}\varphi\big(\mathbb{E}\big[(S^*S)^p\big]\big) \\
&= 1 + \sum_{p=0}^{\infty} \tfrac{t^p}{p!}\varphi\Big(\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big)\otimes \mathbf{1}_n\Big) \\
&\le 1 + \sum_{p=0}^{\infty} \tfrac{t^p}{p!}\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\Big\|\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big\|,
\end{aligned} \tag{3.5}$$

and by (3.2) and (3.3), the latter sum is finite, when $|t| < \frac{n}{\gamma}$. Since $\mathcal{B}(\mathcal{H}^n)^* = \operatorname{span}\big(\mathcal{S}(\mathcal{B}(\mathcal{H}^n))\big)$, it follows that the function $\omega \mapsto \exp(tS^*(\omega)S(\omega))$, is integrable, and (by the first two equalities in (3.5)) that $\mathbb{E}\big[\exp(tS^*S)\big]$ is given by (3.1). □

The main result of this section is the following

3.3 Theorem. *Let $\mathcal{H},\mathcal{K}$ be Hilbert spaces, and let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$, satisfying that*

$$\sum_{i=1}^{r} a_i^* a_i \le c\mathbf{1}_{\mathcal{B}(\mathcal{H})} \quad \text{and} \quad \sum_{i=1}^{r} a_i a_i^* \le \mathbf{1}_{\mathcal{B}(\mathcal{K})},$$

for some constant c in $]0,\infty[$. Consider furthermore independent elements $Y_1,\dots,Y_r$ of $\mathrm{GRM}(n,n,\frac{1}{n})$, and put $S = \sum_{i=1}^r a_i\otimes Y_i$. Then for any t in $[0,\frac{n}{2c}]\cap[0,\frac{n}{2}]$, we have that

$$\mathbb{E}\big[\exp(tS^*S)\big] \le \exp\big((\sqrt{c}+1)^2 t + (c+1)^2\cdot\tfrac{t^2}{n}\big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}.$$

For the proof of Theorem 3.3, we need three lemmas. Before stating these lemmas, we introduce some notation:

For any p,k,l in $\mathbb{N}$, we put

$$\delta(p,k,l) = \mathrm{card}(\{\pi \in S_p \mid k(\hat{\pi}) = k \text{ and } l(\hat{\pi}) = l\}). \tag{3.6}$$

Note that for any p,k,l in $\mathbb{N}$, $\delta(p,k,l) = 0$, unless $k + l \leq p + 1$ (cf. Theorem 1.13).
For any complex number w and any n in $\mathbb{N}_0$, we put

$$(w)_n = \begin{cases} 1, & \text{if } n = 0, \\ w(w+1)(w+2)\cdots(w+n-1), & \text{if } n \in \mathbb{N}. \end{cases}$$

We recall then, that the hyper-geometric function F, is defined by the formula

$$F(a,b,c;x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k k!} x^k,$$

for a,b,c,x in $\mathbb{C}$, such that $c \notin \mathbb{Z} \setminus \mathbb{N}$, and $|x| < 1$.

3.4 Lemma. *For all positive real numbers α, β, we have that*

$$\begin{aligned} \sum_{p=1}^{\infty} \frac{t^{p-1}}{(p-1)!} \sum_{\substack{k,l \in \mathbb{N} \\ k+l \leq p+1}} & \delta(p,k,l) \alpha^{k-1} \beta^{l-1} \\ &= \frac{F(1-\alpha, 1-\beta, 2, t^2)}{(1-t)^{\alpha+\beta}}, \qquad (t \in \mathbb{C},\ |t| < 1). \end{aligned} \tag{3.7}$$

Proof. Assume first that $\alpha = n$ and $\beta = m$, where $m,n \in \mathbb{N}$, and consider an element B of $\mathrm{GRM}(m,n,1)$. Then by [HT, Theorem 6.4],

$$\begin{aligned} \frac{F(1-n, 1-m, 2, t^2)}{(1-t)^{m+n}} &= \frac{1}{mn} \mathbb{E} \circ \mathrm{Tr}_n \big[B^* B \exp(t B^* B)\big] \\ &= \frac{1}{mn} \sum_{p=1}^{\infty} \frac{t^{p-1}}{(p-1)!} \mathbb{E} \circ \mathrm{Tr}_n \big[(B^* B)^p\big]. \end{aligned}$$

But from Section 1 of this paper, we know that for any p in $\mathbb{N}$

$$\mathbb{E} \circ \mathrm{Tr}_n \big[(B^* B)^p\big] = \sum_{\pi \in S_p} m^{k(\hat{\pi})} n^{l(\hat{\pi})} = \sum_{\substack{k,l \in \mathbb{N} \\ k+l \leq p+1}} \delta(p,k,l) m^k n^l,$$

and thus (3.7) holds for all α, β in $\mathbb{N}$. In particular, the left hand side (3.7) is finite for all α, β in $\mathbb{N}$. Since the left hand side of (3.7) is an increasing function of both α and β, it is therefore finite for all α, β in $]0, \infty[$.
To prove (3.7) for general positive real numbers, α, β, we get first, as in [HT, Proof of Proposition 8.1], by multiplying the power series

$$F(1-\alpha, 1-\beta, 2; t^2) = \sum_{j=0}^{\infty} \frac{1}{j+1} \binom{\alpha-1}{j} \binom{\beta-1}{j} t^{2j}, \qquad (|t| < 1),$$

and

$$(1-t)^{-(\alpha+\beta)} = \sum_{k=0}^{\infty} \binom{\alpha+\beta+k-1}{k} t^k, \qquad (|t|<1),$$

that the power series expansion for $\frac{F(1-\alpha,1-\beta,2;t^2)}{(1-t)^{\alpha+\beta}}$ is given by

$$\frac{F(1-\alpha,1-\beta,2;t^2)}{(1-t)^{\alpha+\beta}} = \sum_{p=1}^{\infty} \psi(p,\alpha,\beta) t^{p-1}, \qquad (|t|<1), \tag{3.8}$$

where for all p in $\mathbb{N}$,

$$\psi(p,\alpha,\beta) = \sum_{j=0}^{[\frac{p-1}{2}]} \frac{1}{j+1} \binom{\alpha-1}{j} \binom{\beta-1}{j} \binom{\alpha+\beta+p-2j-2}{p-2j-1}. \tag{3.9}$$

Since we know that (3.7) holds for all α, β in $\mathbb{N}$, we have, on the other hand, that

$$\psi(p,\alpha,\beta) = \frac{1}{(p-1)!} \sum_{\substack{k,l\in\mathbb{N} \\ k+l\le p+1}} \delta(p,k,l) \alpha^{k-1} \beta^{l-1}, \tag{3.10}$$

for all α, β in $\mathbb{N}$. Thus, for fixed p, the right hand sides of (3.9) and (3.10) coincide whenever $\alpha, \beta \in \mathbb{N}$, and since these two right hand sides are both polynomials in α and β, they must therefore coincide for all α, β in $]0,\infty[$. In other words, (3.10) holds for all α, β in $]0,\infty[$, and inserting this in (3.8), we get the desired formula. $\square$

3.5 Lemma. *Let α, β be positive numbers, and assume that either α or β is an integer. Then*

$$F(1-\alpha,1-\beta,2;t^2) \le \sum_{j=0}^{\infty} \frac{(\alpha\beta)^j t^{2j}}{j!(j+1)!}, \qquad \textit{whenever } 0 \le t < 1. \tag{3.11}$$

Proof. We recall first, that

$$F(1-\alpha,1-\beta,2;t^2) = \sum_{j=0}^{\infty} \frac{1}{j+1} \binom{\alpha-1}{j} \binom{\beta-1}{j} t^{2j}, \qquad (t \in [0,1[).$$

If both α and β are integers, then

$$0 \le \binom{\alpha-1}{j} \le \frac{\alpha^j}{j!} \qquad \text{and} \qquad 0 \le \binom{\beta-1}{j} \le \frac{\beta^j}{j!},$$

for all j in $\mathbb{N}_0$, and (3.11) follows immediately. By symmetry of (3.11) in α and β, it is therefore sufficient to treat the case where α is an integer and β is not. In this case, we have

$$F(1-\alpha,1-\beta,2;t^2) = \sum_{j=0}^{\alpha-1} \frac{1}{j+1} \binom{\alpha-1}{j} \binom{\beta-1}{j} t^{2j}.$$

If $\beta \geq \alpha$, we have for any j in $\{0, 1, \ldots, \alpha-1\}$, that $0 < \binom{\alpha-1}{j} \leq \frac{\alpha^j}{j!}$ and $0 < \binom{\beta-1}{j} \leq \frac{\beta^j}{j!}$, and again (3.11) follows immediately.
Assume then that $\beta < \alpha$, and let n be the integer for which $n-1 < \beta < n$. Since α is an integer, and $\alpha > \beta$, we have that $\alpha \geq n$. Forming now Taylor expansion on the function $f(s) = F(1-\alpha, 1-\beta, 2; s)$, $(s > 0)$, it follows that

$$F(1-\alpha, 1-\beta, 2; s) = \sum_{j=0}^{n-1} \frac{1}{j+1} \binom{\alpha-1}{j} \binom{\beta-1}{j} s^j + r_n(s), \qquad (s > 0), \tag{3.12}$$

where $r_n(s) = \frac{f^{(n)}(\xi(s))}{n!} s^n$, for some $\xi(s)$ in $]0, s[$. It suffices thus to show that $f^{(n)}(\xi) \leq 0$, for all ξ in $[0, 1[$, since this will imply that for all s in $[0, 1[$,

$$F(1-\alpha, 1-\beta, 2; s) \leq \sum_{j=0}^{n-1} \frac{1}{j+1} \binom{\alpha-1}{j} \binom{\beta-1}{j} s^j,$$

where, as above, $0 < \binom{\alpha-1}{j} \leq \frac{\alpha^j}{j!}$ and $0 < \binom{\beta-1}{j} \leq \frac{\beta^j}{j!}$, for all j in $\{0, 1, \ldots, n-1\}$.
To show that $f^{(n)}(\xi) \leq 0$ for all ξ in $[0, 1[$, we note that by [HTF, Vol. 1, p. 58, formula (7)],

$$\begin{aligned} f^{(n)}(\xi) &= \frac{d^n}{d\xi^n} F(1-\alpha, 1-\beta, 2; \xi) \\ &= \frac{(1-\alpha)_n (1-\beta)_n}{(n+1)!} F(n+1-\alpha, n+1-\beta, n+2; \xi), \end{aligned}$$

for all ξ in $[0, 1[$. Note here that

$$(1-\alpha)_n (1-\beta)_n = (\alpha-1)(\alpha-2)\cdots(\alpha-n)(\beta-1)(\beta-2)\cdots(\beta-n) \leq 0,$$

because $\alpha \geq n$ and $n-1 < \beta < n$. Moreover, by [HTF, Vol. 1, p. 105, formula (2)], we have for all ξ in $[0, 1[$

$$\begin{aligned} F(n+1-\alpha, n+1-\beta, n+2; \xi) &= (1-\xi)^{\alpha+\beta-n} F(\alpha+1, \beta+1, n+2; \xi) \\ &= (1-\xi)^{\alpha+\beta-n} \sum_{j=0}^{\infty} \frac{(\alpha+1)_j (\beta+1)_j}{j!(n+2)_j} \xi^j, \end{aligned}$$

and therefore $F(n+1-\alpha, n+1-\beta, n+2; \xi) > 0$ for all ξ in $[0, 1[$. Taken together, it follows that $f^{(n)}(\xi) \leq 0$ for all ξ in $[0, 1[$, as desired. $\square$

For any c in $]0, \infty[$, we let μ_c denote the free Poisson distribution with parameter c, i.e., the probability measure on $\mathbb{R}$, given by

$$\mu_c = \max\{1-c, 0\}\delta_0 + \frac{\sqrt{(x-a)(b-x)}}{2\pi x} \cdot 1_{[a,b]}(x) \cdot dx, \tag{3.13}$$

where $a = (\sqrt{c}-1)^2$, $b = (\sqrt{c}+1)^2$ and δ_0 is the Dirac measure at 0 (cf. [HT, Definition 6.5]).

3.6 Lemma. *Let α, β be strictly positive real numbers, and assume that either α or β is an integer. Then for any t in $[0, \frac{1}{2}]$,*

$$1 + \sum_{p=1}^{\infty} \frac{t^p}{p!} \sum_{\substack{k,l \in \mathbb{N} \\ k+l \le p+1}} \delta(p,k,l)\alpha^k \beta^{l-1} \le \exp((\alpha+\beta)t^2) \int_0^{\infty} \exp(\beta t x) \, d\mu_{\frac{\alpha}{\beta}}(x),$$

Proof. Using that $-\log(1-t) = \sum_{n=1}^{\infty} \frac{t^n}{n} \le t + t^2$, whenever $0 \le t \le \frac{1}{2}$, we note first that

$$(1-t)^{-(\alpha+\beta)} \le \exp((\alpha+\beta)t)\exp((\alpha+\beta)t^2), \qquad (t \in [0, \tfrac{1}{2}]).$$

Hence by Lemma 3.4 and Lemma 3.5,

$$\begin{aligned}\sum_{p=1}^{\infty} \frac{t^{p-1}}{(p-1)!} \sum_{\substack{k,l \in \mathbb{N} \\ k+l \le p+1}} &\delta(p,k,l)\alpha^{k-1}\beta^{l-1} \\ &\le \exp((\alpha+\beta)t)\exp((\alpha+\beta)t^2) \sum_{j=0}^{\infty} \frac{(\alpha\beta)^j t^{2j}}{j!(j+1)!}\end{aligned} \tag{3.14}$$

Put $c = \frac{\alpha}{\beta}$ and $s = \beta t$. From [HT, Formula (6.27)], it follows then that

$$\begin{aligned}\int_0^{\infty} x \exp(sx) \, d\mu_c(x) &= c\exp((c+1)s) \sum_{j=0}^{\infty} \frac{c^j s^{2j}}{j!(j+1)!} \\ &= c\exp((\alpha+\beta)t) \sum_{j=0}^{\infty} \frac{(\alpha\beta)^j t^{2j}}{j!(j+1)!}.\end{aligned}$$

Hence (3.14) can be rewritten as

$$\begin{aligned}\sum_{p=1}^{\infty} \frac{t^{p-1}}{(p-1)!} \sum_{\substack{k,l \in \mathbb{N} \\ k+l \le p+1}} &\delta(p,k,l)\alpha^{k-1}\beta^{l-1} \\ &\le \tfrac{\beta}{\alpha} \exp((\alpha+\beta)t^2) \int_0^{\infty} x\exp(\beta t x) \, d\mu_{\frac{\alpha}{\beta}}(x).\end{aligned} \tag{3.15}$$

Using then that $\frac{t^p}{p!} = \int_0^t \frac{u^{p-1}}{(p-1)!} \, du$, for all p in $\mathbb{N}$, and that $\exp((\alpha+\beta)u^2) \le \exp((\alpha+\beta)t^2)$, whenever $0 \le u \le t$, we get by termwise integration of (3.15) (after replacing t by u), that

$$
\begin{aligned}
&\sum_{p=1}^{\infty}\frac{t^p}{p!}\sum_{\substack{k,l\in\mathbb{N}\\ k+l\le p+1}}\delta(p,k,l)\alpha^{k-1}\beta^{l-1}\\
&\qquad\le \tfrac{\beta}{\alpha}\exp((\alpha+\beta)t^2)\int_0^t\Big(\int_0^\infty x\exp(\beta ux)\;d\mu_{\frac{\alpha}{\beta}}(x)\Big)\;du\\
&\qquad= \tfrac{\beta}{\alpha}\exp((\alpha+\beta)t^2)\int_0^\infty x\;\frac{\exp(\beta tx)-1}{\beta x}\;d\mu_{\frac{\alpha}{\beta}}(x)\\
&\qquad= \tfrac{1}{\alpha}\exp((\alpha+\beta)t^2)\int_0^\infty (\exp(\beta tx)-1)\;d\mu_{\frac{\alpha}{\beta}}(x).
\end{aligned}
$$

Hence, using that $\mu_{\frac{\alpha}{\beta}}$ is a probability measure, it follows that

$$
\begin{aligned}
&1+\sum_{p=1}^{\infty}\frac{t^p}{p!}\sum_{\substack{k,l\in\mathbb{N}\\ k+l\le p+1}}\delta(p,k,l)\alpha^{k}\beta^{l-1}\\
&\qquad\le 1+\exp((\alpha+\beta)t^2)\Big(\int_0^\infty \exp(\beta tx)\;d\mu_{\frac{\alpha}{\beta}}(x)-1\Big)\\
&\qquad\le \exp((\alpha+\beta)t^2)\int_0^\infty \exp(\beta tx)\;d\mu_{\frac{\alpha}{\beta}}(x).
\end{aligned}
$$

This conclude the proof. □

Proof of Theorem 3.3. Let $a_1,\dots,a_r$, $Y_1,\dots,Y_r$ and S be as set out in Theorem 3.3. By Proposition 2.5 and Proposition 2.7, we have then that

$$
\begin{aligned}
\mathbb{E}\big[(S^*S)^p\big] &= \Big(\sum_{\pi\in S_p} n^{-2\sigma(\hat\pi)}\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big)\otimes \mathbf{1}_n\\
&\le \Big(\sum_{\pi\in S_p} n^{-2\sigma(\hat\pi)}c^{\kappa(\hat\pi)}\Big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)},
\end{aligned}
\tag{3.16}
$$

where $\kappa(\hat\pi)$ was introduced in Definition 2.6.
We assume first that $c\ge 1$. By Proposition 2.10(i) and (ii), we have that

$$\kappa(\hat\pi)\le k(\hat\pi)+2\sigma(\hat\pi),\qquad (\pi\in S_p).$$

Hence,

$$\mathbb{E}\big[(S^*S)^p\big]\le \Big(\sum_{\pi\in S_p}\big(\tfrac{n}{c}\big)^{-2\sigma(\hat\pi)}c^{k(\hat\pi)}\Big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}.$$

Using now that $2\sigma(\hat\pi)=p+1-d(\hat\pi)=p+1-k(\hat\pi)-l(\hat\pi)$, we find that

$$
\begin{aligned}
\mathbb{E}\big[(S^*S)^p\big] &\le \Big(\big(\tfrac{c}{n}\big)^{p+1}\sum_{\pi\in S_p} n^{k(\hat\pi)}\big(\tfrac{n}{c}\big)^{l(\hat\pi)}\Big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}\\
&= \Big(\big(\tfrac{c}{n}\big)^{p}\sum_{\substack{k,l\in\mathbb{N}\\ k+l\le p+1}}\delta(p,k,l)n^k\big(\tfrac{n}{c}\big)^{l-1}\Big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)},
\end{aligned}
$$

and therefore, for $0 \leq t \leq \frac{n}{\max\{c,1\}} = \frac{n}{c}$, it follows by application of Proposition 3.2, that

$$\begin{aligned}
\mathbb{E}\big[\exp(tS^*S)\big] &= \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)} + \sum_{p=1}^{\infty} \tfrac{t^p}{p!}\mathbb{E}\big[(S^*S)^p\big] \\
&\leq \Big(1 + \sum_{p=1}^{\infty} \tfrac{1}{p!}\big(\tfrac{ct}{n}\big)^p \sum_{\substack{k,l\in\mathbb{N} \\ k+l\leq p+1}} \delta(p,k,l) n^k \big(\tfrac{n}{c}\big)^{l-1}\Big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}.
\end{aligned}$$

Using now Lemma 3.6, we get for $0 \leq \frac{ct}{n} \leq \frac{1}{2}$, that

$$\begin{aligned}
\mathbb{E}\big[\exp(tS^*S)\big] &\leq \Big(\exp\big((n+\tfrac{n}{c})(\tfrac{ct}{n})^2\big)\int_0^\infty \exp\big(\tfrac{n}{c}(\tfrac{ct}{n})x\big)\ d\mu_c(x)\Big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)} \\
&= \Big(\exp\big(c(c+1)\tfrac{t^2}{n}\big)\int_0^\infty \exp(tx)\ d\mu_c(x)\Big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)} \\
&\leq \Big(\exp\big((c+1)^2 \cdot \tfrac{t^2}{n}\big)\int_0^\infty \exp(tx)\ d\mu_c(x)\Big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}.
\end{aligned}$$

Since $\operatorname{supp}(\mu_c) \subseteq \big[0, (\sqrt{c}+1)^2\big]$, it follows that

$$\mathbb{E}\big[\exp(tS^*S)\big] \leq \exp\big((c+1)^2 \cdot \tfrac{t^2}{n}\big)\exp\big((\sqrt{c}+1)^2 t\big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)},$$

and this proves the theorem in the case where $c \geq 1$.
Assume then that $c < 1$. In this case we use (3.16) together with the fact that $\kappa(\hat{\pi}) \geq k(\hat{\pi})$ for all π in S_p, (Proposition 2.10(ii)) to obtain

$$\begin{aligned}
\mathbb{E}\big[(S^*S)^p\big] &\leq \Big(\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})} c^{k(\hat{\pi})}\Big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)} \\
&\leq \Big(\tfrac{1}{n^{p+1}} \sum_{\pi\in S_p} (nc)^{k(\hat{\pi})} n^{l(\hat{\pi})}\Big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)} \\
&= \Big(\tfrac{1}{n^p} \sum_{\substack{k,l\in\mathbb{N} \\ k+l\leq p+1}} \delta(p,k,l)(nc)^k n^{l-1}\Big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}.
\end{aligned}$$

Hence for $0 \leq t < \frac{n}{\max\{c,1\}} = n$, we get by application of Proposition 3.2,

$$\begin{aligned}
\mathbb{E}\big[\exp(tS^*S)\big] &\leq \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)} + \sum_{p=1}^{\infty} \tfrac{t^p}{p!}\mathbb{E}\big[(S^*S)^p\big] \\
&\leq \Big(1 + \sum_{p=1}^{\infty} \tfrac{1}{p!}\big(\tfrac{t}{n}\big)^p \sum_{\substack{k,l\in\mathbb{N} \\ k+l\leq p+1}} \delta(p,k,l)(nc)^k n^{l-1}\Big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}.
\end{aligned}$$

Hence by Lemma 3.6, we have for $0 \leq \frac{t}{n} \leq \frac{1}{2}$,

$$\begin{aligned}\mathbb{E}\big[\exp(tS^*S)\big] &\leq \Big(\exp\big((nc+n)(\tfrac{t}{n})^2\big)\int_0^\infty \exp\big(n(\tfrac{t}{n})x\big)\ d\mu_c(x)\Big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}\\ &= \Big(\exp\big((c+1)\tfrac{t^2}{n}\big)\int_0^\infty \exp(tx)\ d\mu_c(x)\Big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}\\ &\leq \exp\big((c+1)^2\cdot\tfrac{t^2}{n})\big)\exp\big((\sqrt{c}+1)^2 t\big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)},\end{aligned}$$

and this completes the proof. □

3.7 Remark. Assume that $a_1,\dots,a_r \in \mathcal{B}(\mathcal{H},\mathcal{K})$, satisfying that $\sum_{i=1}^r a_i^*a_i \leq c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$ and $\sum_{i=1}^r a_ia_i^* \leq d\mathbf{1}_{\mathcal{B}(\mathcal{H})}$, for some positive constants c and d. Consider furthermore independent elements $Y_1,\dots,Y_r$ of $\mathrm{GRM}(n,n,\frac{1}{n})$, and put $S = \sum_{i=1}^r a_i\otimes Y_i$. Applying then Theorem 3.3 to $a_i' = \frac{1}{\sqrt{d}}a_i$ and $c' = \frac{c}{d}$, we get the following extension of Theorem 3.3:
For any t in $[0,\frac{n}{2c}]\cap[0,\frac{n}{2d}]$,

$$\mathbb{E}\big[\exp(tS^*S)\big] \leq \exp\big((\sqrt{c}+\sqrt{d})^2 t + (c+d)^2\cdot\tfrac{t^2}{n}\big)\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}.$$ □

4 Asymptotic Upper Bound on the Spectrum of $S_n^*S_n$ in the Exact Case

Throughout this section, we consider elements $a_1,\dots,a_r$ of $\mathcal{B}(\mathcal{H},\mathcal{K})$ (for Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$), satisfying that

$$\Big\|\sum_{i=1}^r a_i^*a_i\Big\| \leq c, \quad \text{and} \quad \Big\|\sum_{i=1}^r a_ia_i^*\Big\| \leq 1, \tag{4.1}$$

for some constant c in $]0,\infty[$. Let $\mathcal{A}$ denote the unital C^*-subalgebra of $\mathcal{B}(\mathcal{H})$ generated by the family $\big\{a_i^*a_j \;\big|\; i,j\in\{1,\dots,r\}\big\}\cup\{\mathbf{1}_{\mathcal{B}(\mathcal{H})}\}$. Furthermore, for each n in $\mathbb{N}$, we consider independent elements $Y_1^{(n)},\dots,Y_r^{(n)}$ of $\mathrm{GRM}(n,n,\frac{1}{n})$, and we define

$$S_n = \sum_{i=1}^r a_i\otimes Y_i^{(n)}. \tag{4.2}$$

In this section, we shall determine (almost surely) the asymptotic behavior (as $n\to\infty$) of the largest element of the spectrum of $S_n^*S_n$ (i.e., the norm of $S_n^*S_n$), under the assumption that $\mathcal{A}$ is an exact C^*-algebra. We start by studying the corresponding asymptotic behavior for the image of $S_n^*S_n$ under certain matrix valued completely positive mappings. More precisely, let d be a fixed positive integer, and let $\Phi\colon \mathcal{A}\to M_d(\mathbb{C})$ be a unital completely positive mapping. For each n in $\mathbb{N}$, let $\mathrm{id}_n\colon M_n(\mathbb{C}) \to M_n(\mathbb{C})$ denote the identity mapping on $M_n(\mathbb{C})$. We then define

$$V_n = \big(\Phi \otimes \mathrm{id}_n\big)(S_n^* S_n) = \sum_{i,j=1}^{r} \Phi(a_i^* a_j) \otimes \big(Y_i^{(n)}\big)^* Y_j^{(n)}, \qquad (n \in \mathbb{N}). \tag{4.3}$$

Note that V_n is a random variable taking values in $M_d(\mathbb{C}) \otimes M_n(\mathbb{C}) \simeq M_{dn}(\mathbb{C})$. As indicated above, our first objective is to determine the asymptotic behavior of the largest eigenvalue of V_n. We emphasize, that this step does not require that $\mathcal{A}$ be exact.

The following lemma is a version of Jensen's Inequality, which we shall need significantly in this section and in Section 8. The lemma has been proved in much more general settings by Brown and Kosaki (cf. [BK]) and by Petz (cf. [Pe]). For the reader's convenience, we include a short proof, handling the special case needed here.

4.1 Lemma. (i) *Let $\mathcal{L}$ be a Hilbert space, and let P be a finite dimensional projection in $\mathcal{B}(\mathcal{L})$. Let* tr *denote the normalized trace on $\mathcal{B}(P(\mathcal{L}))$. Then for any selfadjoint element a of $\mathcal{B}(\mathcal{L})$, and any convex function $g\colon \mathbb{R} \to \mathbb{R}$, we have that*

$$\mathrm{tr}\big[g(PaP)\big] \le \mathrm{tr}\big[Pg(a)P\big]. \tag{4.4}$$

(ii) *Let $\mathcal{B}$ be a C^*-algebra, let m be a positive integer and let $\Psi\colon \mathcal{B} \to M_m(\mathbb{C})$ be a unital completely positive mapping. Then for any selfadjoint element a of $\mathcal{B}$ and any convex function $g\colon \mathbb{R} \to \mathbb{R}$, we have that*

$$\mathrm{tr}_m\big[g(\Psi(a))\big] \le \mathrm{tr}_m\big[\Psi(g(a))\big].$$

Proof. (i) Note first that g is continuous (being convex on the whole real line). Let m denote the dimension of $P(\mathcal{L})$, and choose an orthonormal basis $(e_1, \dots, e_m)$ for $P(\mathcal{L})$ consisting of eigenvectors for PaP. Let $\lambda_1, \dots, \lambda_m$ be the corresponding eigenvalues for PaP, i.e.,

$$\lambda_i = \langle PaPe_i, e_i\rangle = \langle ae_i, e_i\rangle, \qquad (i \in \{1, 2, \dots, m\}).$$

Then $g(\lambda_1), \dots, g(\lambda_m)$ are the eigenvalues of $g(PaP)$, and hence

$$\mathrm{tr}\big[g(PaP)\big] = \sum_{i=1}^{m} g(\lambda_i) = \sum_{i=1}^{m} g(\langle ae_i, e_i\rangle). \tag{4.5}$$

Since the trace on $\mathcal{B}(P(\mathcal{L}))$ is independent of the choice of orthonormal basis for $P(\mathcal{L})$, we have at the same time, that

$$\mathrm{tr}\big[Pg(a)P\big] = \sum_{i=1}^{m} \langle Pg(a)Pe_i, e_i\rangle = \sum_{i=1}^{m} \langle g(a)e_i, e_i\rangle. \tag{4.6}$$

Comparing (4.5) and (4.6), we see that it suffices to show that $\langle g(a)e_i, e_i\rangle \geq g(\langle ae_i, e_i\rangle)$, for all i in $\{1, 2, \dots, m\}$. But for each i, this follows from the classical Jensen Inequality, applied to the distribution of a w.r.t. the state $\langle\,\cdot\, e_i, e_i\rangle$, i.e., the probability measure μ_i supported on $\mathrm{sp}(a)$, and satisfying that $\langle f(a)e_i, e_i\rangle = \int_{\mathrm{sp}(a)} f(t)\; d\mu_i(t)$, for all functions f in $C(\mathrm{sp}(a))$. This concludes the proof of (i).

(ii) By Stinespring's Theorem, we may choose a Hilbert space $\mathcal{L}$, a $*$-representation $\pi\colon \mathcal{B} \to \mathcal{B}(\mathcal{L})$ of $\mathcal{B}$ on $\mathcal{L}$, and an embedding $\iota\colon \mathbb{C}^m \to \mathcal{L}$ of $\mathbb{C}^m$ into $\mathcal{L}$, such that

$$\Psi(b) = P_K\pi(b)P_K, \qquad (b \in \mathcal{B}), \tag{4.7}$$

where $K = \iota(\mathbb{C}^m)$, and P_K is the orthogonal projection of $\mathcal{L}$ onto K. Moreover, the equality (4.7) is modulo the natural identifications associated with ι. Let tr_K denote the normalized trace on $\mathcal{B}(K)$. By application of (i), it follows then that

$$\begin{aligned}\mathrm{tr}_m\big[g(\Psi(a))\big] &= \mathrm{tr}_K\big[g(P_K\pi(a)P_K)\big] \leq \mathrm{tr}_K\big[P_K g(\pi(a))P_K\big]\\ &= \mathrm{tr}_K\big[P_K\pi(g(a))P_K\big] = \mathrm{tr}_m\big[\Psi(g(a))\big],\end{aligned}$$

and this proves (ii). □

4.2 Lemma. *Let V_n, $n \in \mathbb{N}$, be as in* (4.3), *and let $\lambda_{\max}(V_n)$ denote the largest eigenvalue of V_n (considered as an element of $M_{dn}(\mathbb{C})$). Then for any ϵ in $]0,\infty[$, we have that*

$$\sum_{n=1}^{\infty} P\big(\lambda_{\max}(V_n) \geq (\sqrt{c}+1)^2 + \epsilon\big) < \infty. \tag{4.8}$$

Proof. The proof proceeds along the same lines as the proof of [HT, Lemma 7.3]; the main difference being that in the present situation, we have to rely on the estimate obtained in Theorem 3.3. Consider first a fixed n in $\mathbb{N}$. We find then for any t in $]0,\infty[$, that

$$\begin{aligned}P\big(\lambda_{\max}(V_n) \geq (\sqrt{c}+1)^2 + \epsilon\big) &= P\big(\exp\big(t\lambda_{\max}(V_n) - t(\sqrt{c}+1)^2 - t\epsilon\big) \geq 1\big)\\ &\leq \mathbb{E}\big[\exp\big(t\lambda_{\max}(V_n) - t(\sqrt{c}+1)^2 - t\epsilon\big)\big]\\ &= \exp(-t(\sqrt{c}+1)^2 - t\epsilon)\cdot\mathbb{E}\big[\lambda_{\max}\big(\exp(tV_n)\big)\big]\\ &\leq \exp(-t(\sqrt{c}+1)^2 - t\epsilon)\cdot\mathbb{E}\big[\mathrm{Tr}_{dn}\big(\exp(tV_n)\big)\big],\end{aligned} \tag{4.9}$$

where the last inequality follows by noting, that since $\exp(tV_n)$ is a positive $dn \times dn$ matrix, $\lambda_{\max}(\exp(tV_n)) \leq \mathrm{Tr}_{dn}(\exp(tV_n))$. Note now, that since the mapping $\Phi \otimes \mathrm{id}_n$ is unital, completely positive, and since the function $x \mapsto$

$e^{tx}\colon \mathbb{R} \to \mathbb{R}$ is convex, it follows from Lemma 4.1(ii), that

$$\begin{aligned}\mathrm{tr}_{dn}\big[\exp(tV_n)\big] &= \mathrm{tr}_{dn}\big[\exp\big(t(\Phi\otimes \mathrm{id}_n)(S_n^*S_n)\big)\big]\\ &\le \mathrm{tr}_{dn}\big[\big(\Phi\otimes \mathrm{id}_n\big)(\exp(tS_n^*S_n))\big]\\ &= \mathrm{tr}_d\otimes \mathrm{tr}_n\big[\big(\Phi\otimes \mathrm{id}_n\big)(\exp(tS_n^*S_n))\big] = \phi\otimes\mathrm{tr}_n\big[\exp(tS_n^*S_n)\big],\end{aligned} \tag{4.10}$$

where ϕ is the state $\mathrm{tr}_d\circ\Phi$ on $\mathcal{A}$. Note here, that by Definition 3.1 and Theorem 3.3,

$$\begin{aligned}\mathbb{E}\big[\phi\otimes\mathrm{tr}_n\big(\exp(tS_n^*S_n)\big)\big] &= \phi\otimes\mathrm{tr}_n\big(\mathbb{E}\big[\exp(tS_n^*S_n)\big]\big)\\ &\le \exp\big(t(\sqrt{c}+1)^2+\tfrac{t^2}{n}(c+1)^2\big),\end{aligned} \tag{4.11}$$

for all t in $]0,\frac{n}{2c}]$.
Combining now (4.9)-(4.11), we get that for all t in $]0,\frac{n}{2c}]$,

$$\begin{aligned}P\big(\lambda_{\max}(V_n) \ge&(\sqrt{c}+1)^2+\epsilon\big)\\ &\le dn\cdot\exp(-t(\sqrt{c}+1)^2-t\epsilon)\cdot\exp\big(t(\sqrt{c}+1)^2+\tfrac{t^2}{n}(c+1)^2\big)\\ &= dn\cdot\exp\big(t(\tfrac{t}{n}(c+1)^2-\epsilon)\big),\end{aligned}$$

Now choose $t=t_n=\frac{n\epsilon}{2(c+1)^2}$, and note that $t_n\in\,]0,\frac{n}{2c}]$, as long as $\epsilon\le 1$. Clearly it suffices to prove the lemma for such ϵ, so we assume that $\epsilon\le1$. It follows then that

$$P\big(\lambda_{\max}(V_n)\ge(\sqrt{c}+1)^2+\epsilon\big)\le dn\cdot\exp\big(t_n(\tfrac{t_n}{n}(c+1)^2-\epsilon)\big)=dn\cdot\exp\big(\tfrac{-n\epsilon^2}{4(c+1)^2}\big).$$

Since this estimate holds for all n in $\mathbb{N}$, it follows immediately that (4.8) holds. □

4.3 Proposition. *Let V_n, $n\in\mathbb{N}$, be as in* (4.3). *We then have*

$$\limsup_{n\to\infty}\lambda_{\max}(V_n)\le\big(\sqrt{c}+1\big)^2,\quad \textit{almost surely.}$$

Proof. It suffices to show, that for any ϵ from $]0,\infty[$,

$$P\Big(\limsup_{n\to\infty}\lambda_{\max}(V_n)\le(\sqrt{c}+1)^2+\epsilon\Big)=1,$$

and this will follow, if we show that

$$P\big(\lambda_{\max}(V_n)\le(\sqrt{c}+1)^2+\epsilon,\ \text{ for all but finitely many } n\big)=1,$$

for all ϵ in $]0,\infty[$. But this follows from the Borel-Cantelli Lemma (cf. [Bre, Lemma 3.14]) together with Lemma 4.2. □

The next step is to replace V_n in Proposition 4.3 by $S_n^*S_n$ itself. This is where we need to assume that $\mathcal{A}$ is an exact C^*-algebra. The key point in this step is the important result of E. Kirchberg that exactness implies nuclear embeddability (cf. [Ki2, Theorem 4.1] and [Was, Theorem 7.3]).
Let $\mathcal{B}$ be a unital C^*-algebra. Recall then that an operator system in $\mathcal{B}$ is a subspace E of $\mathcal{B}$, such that $\mathbf{1}_{\mathcal{B}} \in E$ and $x^* \in E$ for all x in E.

4.4 Proposition. *Let $\mathcal{B}$ be a unital exact C^*-algebra, and let E be a finite dimensional operator system in $\mathcal{B}$. Then for any ϵ in $]0, \infty[$, there exist d in $\mathbb{N}$ and a unital completely positive mapping $\Phi\colon \mathcal{B} \to M_d(\mathbb{C})$, such that*

$$\big\|\big(\Phi \otimes \mathrm{id}_n\big)(x)\big\| \geq (1-\epsilon)\|x\|,$$

for all n in $\mathbb{N}$ and all x in $M_n(E)$.

Proof. Clearly we may assume that $\mathcal{B}$ is a unital C^*-subalgebra of $\mathcal{B}(\mathcal{L})$ for some Hilbert space $\mathcal{L}$. Let N denote the dimension of E. Then by Auerbach's Lemma (cf. [LT, Proposition 1.c.3]), we may choose linear bases $e_1, \dots, e_N$ of E and $e_1^*, \dots, e_N^*$ of the dual space E^*, such that

$$\|e_i\| = \|e_i^*\| = 1, \quad \text{and} \quad e_i^*(e_j) = \delta_{i,j}, \qquad (i, j \in \{1, 2, \dots, N\}). \tag{4.12}$$

Now since $\mathcal{B}$ is exact, and hence nuclear embeddable, there exist d in $\mathbb{N}$, and unital completely positive mappings $\Phi\colon \mathcal{B} \to M_d(\mathbb{C})$ and $\Psi\colon M_d(\mathbb{C}) \to \mathcal{B}(\mathcal{L})$, such that

$$\|\Psi(\Phi(e_i)) - e_i\| \leq \tfrac{\epsilon}{N}, \qquad (i \in \{1, 2, \dots, N\}), \tag{4.13}$$

(cf. [Was, p. 60]). We show that this Φ has the property set out in the proposition. For this, it suffices to show that

$$\big\|(\Psi \circ \Phi - \iota_{\mathcal{B}})_{|E}\big\|_{\mathrm{cb}} \leq \epsilon, \tag{4.14}$$

where $\iota_{\mathcal{B}}\colon \mathcal{B} \to \mathcal{B}(\mathcal{L})$ is the embedding of $\mathcal{B}$ into $\mathcal{B}(\mathcal{L})$. Indeed, knowing the validity of (4.14), we have for n in $\mathbb{N}$ and x in $M_n(E)$, that

$$\begin{aligned}\|x\| &\leq \big\|\big((\Psi \circ \Phi) \otimes \mathrm{id}_n\big)(x) - x\big\| + \big\|\big((\Psi \circ \Phi) \otimes \mathrm{id}_n\big)(x)\big\| \\ &\leq \epsilon\|x\| + \big\|\big((\Psi \circ \Phi) \otimes \mathrm{id}_n\big)(x)\big\|,\end{aligned}$$

and hence that

$$(1-\epsilon)\|x\| \leq \big\|\big((\Psi \circ \Phi) \otimes \mathrm{id}_n\big)(x)\big\| \leq \|(\Phi \otimes \mathrm{id}_n)(x)\|,$$

where the last inequality is due to the fact that Ψ, being unital completely positive, is a complete contraction.
To verify (4.14) note first, that for x in E, we have by (4.12),

$$x = \sum_{i=1}^{N} e_i^*(x) e_i,$$

and hence

$$\Psi\circ\Phi(x) - x = \sum_{i=1}^{N} e_i^*(x)\big(\Psi\circ\Phi(e_i) - e_i\big) = \sum_{i=1}^{N} e_i^*(x) f_i,$$

where $f_i = \Psi\circ\Phi(e_i) - e_i$. Note that by (4.13), $\|f_i\| \le \frac{\epsilon}{N}$, for all i in $\{1, 2, \dots, N\}$.
Consider now n in $\mathbb{N}$ and $x = (x_{rs})_{1\le r,s\le n}$ in $M_n(E)$. We then have

$$\begin{aligned}\big((\Psi\circ\Phi)\otimes \mathrm{id}_n\big)(x) - x &= \big[(\Psi\circ\Phi)(x_{rs}) - x_{rs}\big]_{1\le r,s\le n}\\ &= \Big[\textstyle\sum_{i=1}^{N} e_i^*(x_{rs}) f_i\Big]_{1\le r,s\le n}\\ &= \sum_{i=1}^{N}\Big([e_i^*(x_{rs})]_{1\le r,s\le n}\cdot \mathrm{diag}_n(f_i, \dots, f_i)\Big),\end{aligned} \tag{4.15}$$

where $\mathrm{diag}_n(f_i, \dots, f_i)$ is the $n\times n$ diagonal matrix with f_i in all the diagonal positions. Note here that by (4.12), $\|e_i^*\|_{\mathrm{cb}} = \|e_i^*\| = 1$, for all i (cf. [Pa, Proposition 3.7]). Consequently,

$$\big\|[e_i^*(x_{rs})]_{1\le r,s\le n}\big\| \le \|e_i^*\|_{\mathrm{cb}}\cdot\|x\| = \|x\|, \qquad (i\in\{1, 2, \dots, N\}),$$

and using this in (4.15), we get that

$$\big\|\big((\Psi\circ\Phi)\otimes \mathrm{id}_n\big)(x) - x\big\| \le \sum_{i=1}^{N}\|x\|\cdot\|f_i\| \le \sum_{i=1}^{N}\|x\|\tfrac{\epsilon}{N} = \epsilon\|x\|,$$

which proves (4.14). □

4.5 Theorem. *Let $a_1, \dots, a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$, such that $\|\sum_{i=1}^{r} a_i^* a_i\| \le c$, and $\|\sum_{i=1}^{r} a_i a_i^*\| \le 1$, for some constant c in $]0,\infty[$. Assume, in addition, that the C^*-subalgebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{H})$, generated by $\{a_i^* a_j \mid i,j\in\{1, 2, \dots, r\}\}\cup\{\mathbf{1}_{\mathcal{B}(\mathcal{H})}\}$, is exact. Consider furthermore, for each n in $\mathbb{N}$, independent elements $Y_1^{(n)}, \dots, Y_r^{(n)}$ of* $\mathrm{GRM}(n,n,\frac{1}{n})$, *and put $S_n = \sum_{i=1}^{r} a_i\otimes Y_i^{(n)}$. We then have*

$$\limsup_{n\to\infty}\max\big[\mathrm{sp}(S_n^* S_n)\big] \le \big(\sqrt{c}+1\big)^2, \qquad \textit{almost surely.}$$

Proof. It suffices to show, that for any ϵ from $]0,\infty[$, the set

$$\mathcal{T}_\epsilon = \Big\{\omega\in\Omega \;\Big|\; \limsup_{n\to\infty}\max\big[\mathrm{sp}(S_n(\omega)^* S_n(\omega))\big] \le \tfrac{1}{1-\epsilon}(\sqrt{c}+1)^2\Big\},$$

has probability 1. So let ϵ from $]0,\infty[$ be given, and put

$$E = \mathrm{span}\big(\{\mathbf{1}_{\mathcal{A}}\}\cup\big\{a_i^* a_j \;\big|\; i,j\in\{1, 2, \dots, r\}\big\}\big).$$

Note that $x^* \in E$ for all x in E, and that $\mathbf{1}_{\mathcal{A}} \in E$. Hence E is a finite dimensional operator system in $\mathcal{A}$. Since $\mathcal{A}$ is exact, it follows thus from Proposition 4.4, that we may choose d in $\mathbb{N}$ and a completely positive mapping $\Phi\colon \mathcal{A} \to M_d(\mathbb{C})$, such that

$$\big\|\big(\Phi \otimes \mathrm{id}_n\big)(x)\big\| \geq (1-\epsilon)\|x\|, \qquad (n \in \mathbb{N},\ x \in M_n(E)). \tag{4.16}$$

Now put

$$V_n = \big(\Phi \otimes \mathrm{id}_n\big)(S_n^* S_n), \qquad (n \in \mathbb{N}),$$

and define furthermore

$$\mathcal{V} = \Big\{\omega \in \Omega \;\Big|\; \limsup_{n\to\infty} \|V_n(\omega)\| \leq (\sqrt{c}+1)^2\Big\}.$$

By Proposition 4.3, $P(\mathcal{V}) = 1$, and hence it suffices to show that $\mathcal{T}_\epsilon \supseteq \mathcal{V}$. But if $\omega \in \mathcal{V}$, it follows from (4.16) that

$$\limsup_{n\to\infty} \|S_n(\omega)^* S_n(\omega)\| \leq (1-\epsilon)^{-1} \limsup_{n\to\infty} \|V_n(\omega)\| \leq (1-\epsilon)^{-1}(\sqrt{c}+1)^2,$$

which shows that $\omega \in \mathcal{T}_\epsilon$. This concludes the proof. $\square$

4.6 Corollary. *Let $a_1, \dots, a_r$ be elements of an exact C^*-algebra $\mathcal{A}$, and let, for each n in $\mathbb{N}$, $Y_1^{(n)}, \dots, Y_r^{(n)}$ be independent elements of* $\mathrm{GRM}(n, n, \frac{1}{n})$. *Then*

$$\limsup_{n\to\infty} \Big\|\sum_{i=1}^r a_i \otimes Y_i^{(n)}\Big\| \leq \Big\|\sum_{i=1}^r a_i^* a_i\Big\|^{\frac{1}{2}} + \Big\|\sum_{i=1}^r a_i a_i^*\Big\|^{\frac{1}{2}}, \qquad \textit{almost surely.}$$

Proof. We may assume that not all a_i are zero. Put $\gamma = \|\sum_{i=1}^r a_i^* a_i\| > 0$ and $\delta = \|\sum_{i=1}^r a_i a_i^*\| > 0$. We may assume that $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. Then the unital C^*-algebra $\tilde{\mathcal{A}} = C^*(\mathcal{A}, \mathbf{1}_{\mathcal{B}(\mathcal{H})})$ is also exact, and hence so is every C^*-subalgebra of $\tilde{\mathcal{A}}$ (cf. [Ki1] and [Was, 2.5.2]). Therefore Corollary 4.6 follows by applying Theorem 4.5 to $a_i' = \frac{1}{\sqrt{\delta}} a_i$, $i = 1, \dots, r$, and $c = \frac{\gamma}{\delta}$. $\square$

Regarding the corollary above, consider arbitrary elements $a_1, \dots, a_r$ of an arbitrary C^*-algebra $\mathcal{A}$, and let $\{y_1, \dots, y_r\}$ be a circular (or semi-circular) system in some C^*-probability space $(\mathcal{B}, \psi)$ (cf. [Vo2]), and normalized so that $\psi(y_i^* y_i) = 1$, $i = 1, 2, \dots, r$. In [HP, Proof of Proposition 4.8], G. Pisier and the first named author showed, that in this setting, the following inequality holds:

$$\Big\|\sum_{i=1}^r a_i \otimes y_i\Big\| \leq 2 \max\Big\{\Big\|\sum_{i=1}^r a_i^* a_i\Big\|^{\frac{1}{2}}, \Big\|\sum_{i=1}^r a_i a_i^*\Big\|^{\frac{1}{2}}\Big\}. \tag{4.17}$$

In [HP, Proof of Proposition 4.8], the factor 2 on the right hand side of (4.17) is missing, but this is due to a different choice of normalization of semi-circular and circular families. By application of [Haa, Section 1], it is not hard to strengthen (4.17) to the inequality

$$\Big\| \sum_{i=1}^{r} a_i \otimes y_i \Big\| \le \Big\| \sum_{i=1}^{r} a_i^* a_i \Big\|^{\frac{1}{2}} + \Big\| \sum_{i=1}^{r} a_i a_i^* \Big\|^{\frac{1}{2}}, \tag{4.18}$$

both for semi-circular and circular systems. Since independent elements $Y_1^{(n)}, \dots, Y_r^{(n)}$ of $\mathrm{GRM}(n, n, \frac{1}{n})$ can be considered as a random matrix model for the circular system $\{y_1, \dots, y_r\}$, in the sense of [Vo1, Theorem 2.2], we should thus consider Corollary 4.6 as a random matrix version of (4.18). However, the random matrix version holds only under the assumption that the C^*-algebra $\mathcal{A}$ be exact. In fact, we shall spend the remaining part of this section, showing that the assumption in Theorem 4.5 that the C^*-algebra $\mathcal{A}$ be exact, can not be omitted. We start with two lemmas, the first of which is a slightly strengthened version of [HT, Theorem 7.4] (which, in turn, is a special case of a theorem of Wachter (cf. [Wac])).

4.7 Lemma. *Let c be a positive number, and let (m_n) be a sequence of positive integers, such that $\frac{m_n}{n} \to c$ as $n \to \infty$. Let furthermore (Y_n) be a sequence of random matrices, such that for each n in $\mathbb{N}$, $Y_n \in \mathrm{GRM}(m_n, n, \frac{1}{n})$. Then for any continuous function $f\colon [0, \infty[\, \to \mathbb{C}$, we have that*

$$\lim_{n\to\infty} \mathrm{tr}_n\big[f(Y_n^* Y_n)\big] = \int_0^b f(x)\, d\mu_c(x), \qquad \textit{almost surely}, \tag{4.19}$$

where $b = (\sqrt{c} + 1)^2$ and μ is the measure introduced in (3.13).

Proof. By splitting f in its real and imaginary parts, it is clear, that we may assume that f is a real valued continuous function on $[0, \infty[$. We note next, that it follows from [HT, Theorem 7.4] and the definition of weak convergence (cf. [HT, Definition 2.2]), that (4.19) holds for all continuous *bounded* functions $f\colon [0, \infty[\, \to \mathbb{R}$. Thus, our objective is to pass from bounded to unbounded continuous functions, and the key to this, is the fact (cf. [HT, Theorem 7.1]), that

$$\lim_{n\to\infty} \|Y_n^* Y_n\| = \big(\sqrt{c} + 1\big)^2, \qquad \text{almost surely.} \tag{4.20}$$

Indeed, it follows from (4.20), that (for example)

$$P\big(\ \|Y_n^* Y_n\| \le (\sqrt{c} + 1)^2 + 1, \text{ for all but finitely many } n\big) = 1,$$

and hence, given any ϵ in $]0, \infty[$, we may choose N in $\mathbb{N}$, such that

$$P(F_N) \ge 1 - \epsilon,$$

where

$$F_N = \big\{\omega \in \Omega \;\big|\; \|Y_n(\omega)^* Y_n(\omega)\| \le (\sqrt{c} + 1)^2 + 1, \text{ whenever } n \ge N\big\}.$$

Now, given a continuous function $f\colon [0,\infty[\ \to \mathbb{R}$, let $f_1\colon [0,\infty[\ \to \mathbb{R}$ be an arbitrary continuous function, satisfying that $f_1 = f$ on $[0,(\sqrt{c}+1)^2+1]$, and that $\mathrm{supp}(f)$ is compact. Then for any ω in F_N, we have that

$$f_1(Y_n(\omega)^*Y_n(\omega)) = f(Y_n(\omega)^*Y_n(\omega)), \quad \text{whenever } n \geq N,$$

and hence, since f_1 is bounded,

$$\begin{aligned}\lim_{n\to\infty} \mathrm{tr}_n\big[f(Y_n(\omega)^*Y_n(\omega))\big] &= \lim_{n\to\infty} \mathrm{tr}_n\big[f_1(Y_n(\omega)^*Y_n(\omega))\big] = \int_a^b f_1(x)\ d\mu_c(x)\\ &= \int_a^b f(x)\ d\mu_c(x).\end{aligned}$$

It follows thus, that

$$P\Big(\lim_{n\to\infty} \mathrm{tr}_n\big[f(Y_n^*Y_n)\big] = \textstyle\int_a^b f(x)\ d\mu_c(x)\Big) \geq P(F_N) \geq 1-\epsilon,$$

and since this holds for any ϵ in $]0,\infty[$, we obtain the desired conclusion. $\square$

Next, we shall study the polar decomposition of Gaussian random matrices. Let n be a positive integer and let Y be an element of $\mathrm{GRM}(n,n,\frac{1}{n})$, defined on $(\Omega,\mathcal{F},P)$. Furthermore, let $\mathcal{U}_n$ denote the unitary group of $M_n(\mathbb{C})$.
By a measurable unitary sign for Y, we mean a random matrix $U\colon \Omega \to \mathcal{U}_n$, such that for almost all ω in Ω, the polar-decomposition of $Y(\omega)$ is given by:

$$Y(\omega) = U(\omega)|Y(\omega)|,$$

where, as usual, $|Y(\omega)| = [Y(\omega)^*Y(\omega)]^{\frac{1}{2}}$. To see that such measurable unitary signs do exist, we note first that by [HT, Theorem 5.2], $Y(\omega)$ is invertible for almost all ω. Thus, for example the random matrix $U\colon \Omega \to \mathcal{U}_n$ given by

$$U(\omega) = \begin{cases} Y(\omega)\big[Y(\omega)^*Y(\omega)\big]^{-\frac{1}{2}}, & \text{if } Y(\omega) \text{ is invertible,}\\ \mathbf{1}_n, & \text{otherwise,}\end{cases}$$

is a measurable unitary sign for Y.

4.8 Lemma. *For each n in $\mathbb{N}$, let $Y_1^{(n)},\dots,Y_r^{(n)}$ be (not necessarily independent) random matrices in $\mathrm{GRM}(n,n,\frac{1}{n})$, and let $U_1^{(n)},\dots,U_r^{(n)}$ be measurable unitary signs for $Y_1^{(n)},\dots,Y_r^{(n)}$, respectively. Furthermore, let $\overline{U}_1^{(n)},\dots,\overline{U}_r^{(n)}$, denote the complex conjugated matrices of $U_1^{(n)},\dots,U_r^{(n)}$. We then have*

$$\liminf_{n\to\infty} \Big\| \sum_{i=1}^r \overline{U}_i^{(n)} \otimes Y_i^{(n)} \Big\| \geq \tfrac{8}{3\pi}\cdot r, \qquad \textit{almost surely.}$$

Proof. Let $(e_1,\dots,e_n)$ be the usual orthonormal basis for $\mathbb{C}^n$, and consider then the unit vector $\xi=\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\otimes e_i$ in $\mathbb{C}^n\otimes\mathbb{C}^n$. Note then that for any $A=(a_{jk})$ and $B=(b_{jk})$ in $M_n(\mathbb{C})$, we have that

$$\begin{aligned}\big\langle (A\otimes B)\xi,\xi\big\rangle &= \tfrac{1}{n}\sum_{j,k=1}^n \big\langle (A\otimes B)(e_j\otimes e_j), e_k\otimes e_k\big\rangle\\ &= \tfrac{1}{n}\sum_{j,k=1}^n \langle Ae_j,e_k\rangle\cdot\langle Be_j,e_k\rangle\\ &= \tfrac{1}{n}\sum_{j,k=1}^n a_{kj}b_{kj} = \mathrm{tr}_n(AB^t)=\mathrm{tr}_n(A^tB).\end{aligned}$$

It follows thus, that

$$\begin{aligned}\Big\|\sum_{i=1}^r \overline{U}_i^{(n)}\otimes Y_i^{(n)}\Big\| \ge \Big|\Big\langle\Big(\sum_{i=1}^r \overline{U}_i^{(n)}\otimes Y_i^{(n)}\Big)\xi,\xi\Big\rangle\Big| &= \Big|\sum_{i=1}^r \mathrm{tr}_n\big[\big(U_i^{(n)}\big)^*Y_i^{(n)}\big]\Big|\\ &= \sum_{i=1}^r \mathrm{tr}_n\big(|Y_i^{(n)}|\big),\end{aligned} \tag{4.21}$$

where the last equation holds almost surely. By Lemma 4.7, we have for all i in $\{1,\dots,r\}$, that

$$\lim_{n\to\infty}\mathrm{tr}_n\big(|Y_i^{(n)}|\big)=\int_0^4\sqrt{x}\ d\mu_1(x),\qquad\text{almost surely,}$$

and combining this with (4.21), it follows that

$$\liminf_{n\to\infty}\Big\|\sum_{i=1}^r \overline{U}_i^{(n)}\otimes Y_i^{(n)}\Big\|\ge r\int_0^4\sqrt{x}\ d\mu_1(x),\qquad\text{almost surely.}$$

We note finally that

$$\int_0^4\sqrt{x}\ d\mu_1(x)=\int_0^4\sqrt{x}\cdot\tfrac{\sqrt{x(4-x)}}{2\pi x}\ dx=\tfrac{1}{2\pi}\int_0^4\sqrt{4-x}\ dx=\tfrac{8}{3\pi},$$

and this concludes the proof. □

We are now ready to give an example where the conclusion of Theorem 4.5 fails, due to lack of exactness of the C^*-algebra $\mathcal{A}$. Consider a fixed positive integer r, greater than or equal to 2, and let $\mathbb{F}_r$ denote the free group on r generators. Let $g_1,\dots,g_r$ denote the generators of $\mathbb{F}_r$, and let $C^*(\mathbb{F}_r)$ denote the *full* C^*-algebra associated to $\mathbb{F}_r$. Recall that there is a canonical unitary representation $u_{\mathbb{F}_r}\colon\mathbb{F}_r\to C^*(\mathbb{F}_r)$, and that the pair $(C^*(\mathbb{F}_r),u_{\mathbb{F}_r})$ is characterized (up to $*$-isomorphism) by the universal property, that given any unital C^*-algebra $\mathcal{B}$

and any unitary representation $u\colon \mathbb{F}_r \to \mathcal{B}$, there exists a unique unital $*$-homomorphism $\Phi_u\colon C^*(\mathbb{F}_r) \to \mathcal{B}$, such that the following diagram commutes:

$$\begin{tikzcd} \mathbb{F}_r \arrow[r, "u_{\mathbb{F}_r}"] \arrow[d, "u"'] & C^*(\mathbb{F}_r) \arrow[dl, "\Phi_u"] \\ \mathcal{B} & \end{tikzcd}$$

It is well-known (cf. [Was, Corollary 3.7]) that $C^*(\mathbb{F}_r)$ is *not* exact. We let $u_1, \dots, u_r$ be the canonical unitaries in $C^*(\mathbb{F}_r)$ associated to $g_1, \dots, g_r$ respectively, i.e., $u_i = u_{\mathbb{F}_r}(g_i)$, $i = 1, \dots, r$. We then define

$$a_i = \tfrac{1}{\sqrt{r}} u_i, \qquad (i \in \{1, \dots, r\}). \tag{4.22}$$

Then clearly,

$$\sum_{i=1}^r a_i^* a_i = \sum_{i=1}^r a_i a_i^* = \mathbf{1}_{C^*(\mathbb{F}_r)}. \tag{4.23}$$

Consider now, in addition, for each n in $\mathbb{N}$, independent elements $Y_1^{(n)}, \dots, Y_r^{(n)}$ of $\mathrm{GRM}(n, n, \frac{1}{n})$, and define

$$S_n = \sum_{i=1}^r a_i \otimes Y_i^{(n)}, \qquad (n \in \mathbb{N}). \tag{4.24}$$

We then have the following

4.9 Proposition. *With $a_1, \dots, a_r$ and S_n, $n \in \mathbb{N}$, as introduced in* (4.22) *and* (4.24), *we have that*
(i) $\liminf_{n\to\infty} \|S_n^* S_n\| \geq \left(\frac{8}{3\pi}\right)^2 \cdot r$, *almost surely.*
(ii) *The conclusion of Theorem 4.5 does not hold for these $a_1, \dots, a_r$, whenever $r \geq 6$.*
In particular, the assumption in Theorem 4.5, that $\mathcal{A}$ be exact, can not, in general, be omitted.

Proof. (i) For each positive integer n, choose measurable unitary signs $U_1^{(n)}, \dots, U_r^{(n)}$ for $Y_1^{(n)}, \dots, Y_r^{(n)}$ respectively, and let $\overline{U}_1^{(n)}, \dots, \overline{U}_r^{(n)}$ denote the complex conjugated matrices of $U_1^{(n)}, \dots, U_r^{(n)}$. Since $\mathbb{F}_r$ is the group free product of r copies of $\mathbb{Z}$, it follows that for each ω in Ω and each n in $\mathbb{N}$, there exists a unitary representation $u_\omega^{(n)}\colon \mathbb{F}_r \to M_n(\mathbb{C})$, such that

$$u_\omega^{(n)}(g_i) = \overline{U}_i^{(n)}(\omega), \qquad (i \in \{1, \dots, r\}).$$

By the universal property of $C^*(\mathbb{F}_r)$ it follows then, that for each ω in Ω and each n in $\mathbb{N}$, we may choose a $*$-homomorphism $\Phi_\omega^{(n)}\colon C^*(\mathbb{F}_r) \to M_n(\mathbb{C})$, such that

$$\Phi_\omega^{(n)}(u_i) = \overline{U}_i^{(n)}(\omega), \qquad (i \in \{1, \dots, r\}).$$

For each ω in Ω and each n in $\mathbb{N}$, note now that

$$\begin{aligned}\Big\|\sum_{i=1}^r u_i\otimes Y_i^{(n)}(\omega)\Big\| &\ge \Big\|\big(\Phi_\omega^{(n)}\otimes \mathrm{id}_n\big)\Big(\sum_{i=1}^r u_i\otimes Y_i^{(n)}(\omega)\Big)\Big\| \\ &= \Big\|\sum_{i=1}^r \overline{U}_i^{(n)}(\omega)\otimes Y_i^{(n)}(\omega)\Big\|.\end{aligned}$$

Applying then Lemma 4.8, it follows that

$$\liminf_{n\to\infty}\Big\|\sum_{i=1}^r u_i\otimes Y_i^{(n)}\Big\| \ge \tfrac{8}{3\pi}\cdot r, \qquad \text{almost surely,}$$

and hence that

$$\liminf_{n\to\infty}\Big\|\sum_{i=1}^r a_i\otimes Y_i^{(n)}\Big\| \ge \tfrac{8}{3\pi}\cdot \sqrt{r}, \qquad \text{almost surely.}$$

Since $\|S_n^*S_n\| = \|S_n\|^2$, we get the desired formula.

(ii) By (4.23), $a_1,\dots,a_r$ introduced in (4.22) satisfy condition (4.1) in the case $c=1$. Thus, if the conclusion of Theorem 4.5 were to hold for these $a_1,\dots,a_r$, it would mean that

$$\limsup_{n\to\infty}\Big\|\sum_{i=1}^r a_i\otimes Y_i^{(n)}\Big\| \le 2, \qquad \text{almost surely.}$$

However, Proposition 4.9 shows that

$$\liminf_{n\to\infty}\Big\|\sum_{i=1}^r a_i\otimes Y_i^{(n)}\Big\| \ge \big(\tfrac{8}{3\pi}\big)\cdot \sqrt{r}, \qquad \text{almost surely,}$$

and thus the conclusion of Theorem 4.5 breaks down, for $c=1$, whenever $r > (\frac{3\pi}{4})^2 \approx 5.55$, i.e., for $r\ge 6$. $\square$

5 A New Combinatorial Expression for $\mathbb{E}\big[(S^*S)^p\big]$

Throughout this section, we consider elements $a_1,\dots,a_r$ of $\mathcal{B}(\mathcal{H},\mathcal{K})$, where $\mathcal{H}$ and $\mathcal{K}$ are Hilbert spaces. In Section 2 we proved that if $Y_1,\dots,Y_r$ are independent random matrices in $\mathrm{GRM}(n,n,\frac{1}{n})$, and we put $S=\sum_{i=1}^r a_i\otimes Y_i$, then

$$\mathbb{E}\big[(S^*S)^p\big] = \Big(\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\cdot \sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_p}^*a_{i_{\pi(p)}}\Big)\otimes \mathbf{1}_n. \tag{5.1}$$

In this section, we shall assume that $a_1,\dots,a_r$ satisfy the condition

$$\sum_{i=1}^r a_i^*a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}, \quad\text{and}\quad \sum_{i=1}^r a_ia_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{K})}, \tag{5.2}$$

for some number c in $]0,\infty[$. Under this assumption, and by application of the method of "reductions of permutations", introduced in Section 1, we show that $\mathbb{E}\big[(S^*S)^p\big]$ can be expressed as a constant plus a linear combination of the sums:

$$\sum_{\rho\in S_q^{\rm irr}} n^{-2\sigma(\hat\rho)}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^*a_{i_{\rho(1)}}\cdots a_{i_q}^*a_{i_{\rho(q)}}\Big),\qquad (q=2,\dots,p),$$

where $S_q^{\rm irr}$, as in Section 1, denotes the set of permutations ρ in S_q for which $\hat\rho$ is irreducible in the sense of Definition 1.16.

5.1 Lemma. *Let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$, and assume that* (5.2) *holds. Let p be a positive integer, greater than or equal to 2, let π be a permutation in $S_p\setminus S_p^{\rm irr}$, and let π_0 be the permutation in S_{p-1} obtained by cancellation of a pair $(e,e+1)$ for $\hat\pi$ (cf. Definition 1.18). We then have*
(i) *If e is odd, then $k(\hat\pi_0)=k(\hat\pi)-1$, and*

$$\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_p}^*a_{i_{\pi(p)}} = c\cdot\Big(\sum_{1\le i_1,\dots,i_{p-1}\le r} a_{i_1}^*a_{i_{\pi_0(1)}}\cdots a_{i_{p-1}}^*a_{i_{\pi_0(p-1)}}\Big). \tag{5.3}$$

(ii) *If e is even, then $k(\hat\pi_0)=k(\hat\pi)$, and*

$$\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_p}^*a_{i_{\pi(p)}} = \sum_{1\le i_1,\dots,i_{p-1}\le r} a_{i_1}^*a_{i_{\pi_0(1)}}\cdots a_{i_{p-1}}^*a_{i_{\pi_0(p-1)}}. \tag{5.4}$$

Proof. (i) Assume that e is odd. Then $k(\hat\pi_0)=k(\hat\pi)-1$ by Proposition 1.22. Moreover, $(e,e+1)$ is of the form $(2j-1,2j)$ for some j in $\{1,2,\dots,p\}$, and therefore $\pi(j)=j$ (cf. Definition 1.15). Hence, the index i_j occur only at the $2j-1$'th and the $2j$'th factor in the product $a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_p}^*a_{i_{\pi(p)}}$, and therefore the sum on the left hand side of (5.3) is equal to

$$\sum_{1\le i_1,\dots,i_{j-1},i_{j+1},\dots,i_p\le r} a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_{\pi(j-1)}}\Big(\sum_{i_j=1}^r a_{i_j}^*a_{i_j}\Big)a_{i_{j+1}}^*\cdots a_{i_p}^*a_{i_{\pi(p)}},$$

which by (5.2) is equal to

$$c\cdot\Big(\sum_{1\le i_1,\dots,i_{j-1},i_{j+1},\dots,i_p\le r} a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_{\pi(j-1)}}a_{i_{j+1}}^*\cdots a_{i_p}^*a_{i_{\pi(p)}}\Big). \tag{5.5}$$

Note here, that if we relabel the indices $i_{j+1},\dots,i_p$ by $i_j,\dots,i_{p-1}$, then it follows from Remark 1.19(a), that (5.5) is equal to

$$c\cdot\Big(\sum_{1\le i_1,\dots,i_{p-1}\le r} a_{i_1}^*a_{i_{\pi_0(1)}}\cdots a_{i_{p-1}}^*a_{i_{\pi_0(p-1)}}\Big),$$

and this proves (5.3).
(ii) Assume that e is even. Then $k(\hat{\pi}_0) = k(\hat{\pi})$ by Proposition 1.22, and $(e, e+1) = (2j, 2j+1)$, for some j in $\{1, 2, \dots, p-1\}$, so that $\pi(j) = j+1$ (c.f. Definition 1.15). Hence, the left hand side of (5.4) is equal to

$$\sum_{1 \le i_1, \dots, i_j, i_{j+2}, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_j}^* \Big(\sum_{i_{j+1}=1}^{r} a_{i_{j+1}} a_{i_{j+1}}^* \Big) a_{i_{\pi(j+1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}. \tag{5.6}$$

Here, $\sum_{i_{j+1}=1}^{r} a_{i_{j+1}} a_{i_{j+1}}^* = \mathbf{1}_{\mathcal{B}(\mathcal{K})}$, by (5.2), and proceeding then as in the proof of (i), we obtain by Remark 1.19(b) (after relabeling $i_{j+2}, \dots, i_p$ by $i_{j+1}, \dots, i_{p-1}$), that (5.6) is equal to

$$\sum_{1 \le i_1, \dots, i_{p-1} \le r} a_{i_1}^* a_{i_{\pi_0(1)}} \cdots a_{i_{p-1}}^* a_{i_{\pi_0(p-1)}}.$$

This proves (5.4) □

Recall that for p in $\mathbb{N}$, S_p^{nc} denotes the set of permutations π in S_p, for which the permutation $\hat{\pi}$ is non-crossing in the sense of Definition 1.14.

5.2 Lemma. *Let $a_1, \dots, a_r$ be elements of $\mathcal{B}(\mathcal{H}, \mathcal{K})$, such that* (5.2) *holds, let p be a positive integer, and let π be a permutation in S_p^{nc}. Then*

$$\sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}} = c^{k(\hat{\pi})} \mathbf{1}_{\mathcal{B}(\mathcal{H})}, \tag{5.7}$$

and

$$\sum_{1 \le i_1, \dots, i_p \le r} a_{i_1} a_{i_{\pi(1)}}^* \cdots a_{i_p} a_{i_{\pi(p)}}^* = c^{l(\hat{\pi})-1} \mathbf{1}_{\mathcal{B}(\mathcal{K})}. \tag{5.8}$$

Proof. We start by proving (5.7); proceeding by induction on p. The case $p = 1$ is clear from (5.2). Assume now that $p \ge 2$, and that (5.7) holds for $p-1$ instead of p, and all permutations in S_{p-1}^{nc}. Consider then a permutation π from S_p^{nc}, and recall from Lemma 1.17 that $\hat{\pi}$ has a pair of neighbors $(e, e+1)$. Let π_0 be the permutation in S_{p-1} obtained by cancellation of this pair. Then by Lemma 1.20, $\pi_0 \in S_{p-1}^{\mathrm{nc}}$, and hence by the induction hypothesis,

$$\sum_{1 \le i_1, \dots, i_{p-1} \le r} a_{i_1}^* a_{i_{\pi_0(1)}} \cdots a_{i_{p-1}}^* a_{i_{\pi_0(p-1)}} = c^{k(\hat{\pi}_0)} \mathbf{1}_{\mathcal{B}(\mathcal{H})}. \tag{5.9}$$

But by Lemma 5.1, (5.9) implies (5.7), both when e is odd, and when e is even. This completes the proof of (5.7).
To prove (5.8), we put $b_i = \frac{1}{\sqrt{c}} a_i^*$, $i = 1, 2, \cdots, r$. Then

$$\sum_{i=1}^{r} b_i^* b_i = c^{-1} \mathbf{1}_{\mathcal{B}(\mathcal{K})}, \quad \text{and} \quad \sum_{i=1}^{r} b_i b_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{H})}.$$

Applying then (5.7), with c replaced by c^{-1}, it follows that

$$\sum_{1\le i_1,\dots,i_p\le r} b_{i_1}^* b_{i_{\pi(1)}} \cdots b_{i_p}^* b_{i_{\pi(p)}} = c^{-k(\hat{\pi})} \mathbf{1}_{\mathcal{B}(\mathcal{K})},$$

i.e., that

$$\sum_{1\le i_1,\dots,i_p\le r} a_{i_1} a_{i_{\pi(1)}}^* \cdots a_{i_p} a_{i_{\pi(p)}}^* = c^{p-k(\hat{\pi})} \mathbf{1}_{\mathcal{B}(\mathcal{K})}.$$

Recall finally, that since $\hat{\pi}$ is non-crossing, $k(\hat{\pi}) + l(\hat{\pi}) = p+1$ (cf. Corollary 1.24), and hence it follows that (5.8) holds. □

As in Section 3, for any c in $]0,\infty[$, μ_c denotes the probability measure on $[0,\infty[$, given by

$$\mu_c = \max\{1-c,0\}\delta_0 + \frac{\sqrt{(x-a)(b-x)}}{2\pi x} \cdot 1_{[a,b]}(x) \cdot dx,$$

where $a = (\sqrt{c}-1)^2$, $b = (\sqrt{c}+1)^2$ and δ_0 is the Dirac measure at 0. Recall from [OP] or [HT, Remark 6.8], that the moments of μ_c are given by

$$\int_0^\infty x^p \, d\mu_c(x) = \tfrac{1}{p}\sum_{j=1}^{p} \tbinom{p}{j}\tbinom{p}{j-1} c^j, \qquad (p \in \mathbb{N}). \tag{5.10}$$

5.3 Lemma. *For any positive integer p, we have*

$$\sum_{\pi\in S_p^{\mathrm{nc}}} c^{k(\hat{\pi})} = \tfrac{1}{p}\sum_{j=1}^{p} \tbinom{p}{j}\tbinom{p}{j-1} c^j, \tag{5.11}$$

and

$$\sum_{\pi\in S_p^{\mathrm{nc}}} c^{l(\hat{\pi})-1} = \tfrac{1}{p}\sum_{j=1}^{p} \tbinom{p}{j}\tbinom{p}{j-1} c^{j-1}. \tag{5.12}$$

Proof. To prove (5.11), recall from Corollary 1.12, that for B in $\mathrm{GRM}(m,n,1)$, we have that

$$\mathbb{E}\circ \mathrm{Tr}_n\big[(B^*B)^p\big] = \sum_{\pi\in S_p} m^{k(\hat{\pi})} n^{l(\hat{\pi})}.$$

Hence, for Y in $\mathrm{GRM}(m,n,\frac{1}{n})$,

$$\mathbb{E}\circ \mathrm{tr}_n\big[(Y^*Y)^p\big] = n^{-p-1}\sum_{\pi\in S_p} m^{k(\hat{\pi})} n^{l(\hat{\pi})} = \sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\big(\tfrac{m}{n}\big)^{k(\hat{\pi})}, \tag{5.13}$$

where we have used that $\sigma(\hat{\pi}) = \frac{1}{2}(p+1-k(\hat{\pi})-l(\hat{\pi}))$. Consider now a sequence (m_n) of positive integers, such that $\frac{m_n}{n} \to c$ as $n\to\infty$, and for each n in $\mathbb{N}$, let Y_n be an element of $\mathrm{GRM}(m_n,n,\frac{1}{n})$. It follows then from (5.13), that

$$\lim_{n\to\infty} \mathbb{E}\circ \mathrm{tr}_n\big[(Y^*Y)^p\big] = \sum_{\substack{\pi\in S_p\\ \sigma(\hat{\pi})=0}} c^{k(\hat{\pi})} = \sum_{\pi\in S_p^{\mathrm{nc}}} c^{k(\hat{\pi})}, \tag{5.14}$$

where the last equality follows from Corollary 1.24. On the other hand, it follows from [HT, Theorem 6.7(ii)] and (5.10), that

$$\lim_{n\to\infty} \mathbb{E}\circ \mathrm{tr}_n\big[(Y^*Y)^p\big] = \int_0^\infty x^p \, d\mu_c(x) = \tfrac{1}{p}\sum_{j=1}^{p}\tbinom{p}{j}\tbinom{p}{j-1}c^j. \tag{5.15}$$

Combining (5.14) and (5.15), we obtain (5.11).
To prove (5.12), we use, again, that $k(\hat{\pi}) + l(\hat{\pi}) = p+1$ for all π in S_p^{nc}. It follows thus, that

$$\sum_{\pi\in S_p^{\mathrm{nc}}} c^{l(\hat{\pi})-1} = c^p \sum_{\pi\in S_p^{\mathrm{nc}}} c^{-k(\hat{\pi})}. \tag{5.16}$$

But by (5.11) (with c replaced by c^{-1}), the right hand side of (5.16) is equal to

$$\tfrac{1}{p}\sum_{j=1}^{p}\tbinom{p}{j}\tbinom{p}{j-1}c^{p-j}. \tag{5.17}$$

Substituting finally j with $p+1-j$ in (5.17), we obtain (5.12). □

5.4 Corollary. *Let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$, such that* (5.2) *holds. Then for any p in $\mathbb{N}$, we have that*

(i) $$\sum_{\pi\in S_p^{\mathrm{nc}}}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big) = \Big[\tfrac{1}{p}\sum_{j=1}^{p}\tbinom{p}{j}\tbinom{p}{j-1}c^j\Big]\cdot \mathbf{1}_{\mathcal{B}(\mathcal{H})},$$

and

(ii) $$\sum_{\pi\in S_p^{\mathrm{nc}}}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1} a_{i_{\pi(1)}}^*\cdots a_{i_p} a_{i_{\pi(p)}}^*\Big) = \Big[\tfrac{1}{p}\sum_{j=1}^{p}\tbinom{p}{j}\tbinom{p}{j-1}c^{j-1}\Big]\cdot \mathbf{1}_{\mathcal{B}(\mathcal{K})}.$$

Proof. Combine Lemma 5.2 and Lemma 5.3. □

5.5 Definition. (a) A subset I of $\mathbb{Z}$ is called an *interval of integers*, if it is the form

$$I = \{\alpha, \alpha+1, \dots, \beta\},$$

for some α,β in $\mathbb{Z}$, such that $\alpha\le\beta$.
(b) Let p be a positive integer, let π be a permutation in S_p, and let I be an interval of integers, such that $I\subseteq\{1,2,\dots,2p\}$. We say then that the restriction $\hat{\pi}_{|I}$ of $\hat{\pi}$ to I is non-crossing, if $\hat{\pi}(I)=I$, and $\hat{\pi}$ has no crossing (a,b,c,d) where $a,b,c,d\in I$. In this case, we refer to I as a *non-crossing interval of integers* for $\hat{\pi}$. □

5.6 Remark. Let p be a positive integer, let π be a permutation in S_p and let I be an interval of integers, such that $I \subseteq \{1,2,\dots,2p\}$ and $\hat{\pi}(I) = I$. Since $\hat{\pi}^2 = \mathrm{id}$ and $\hat{\pi}$ has no fixed points, it follows then, that $\mathrm{card}(I)$ is an even number. Put $t = \frac{1}{2}\mathrm{card}(I)$, and consider the unique order preserving bijection $\varphi\colon \{1,2,\dots,2t\} \to I$ of $\{1,2,\dots,2t\}$ onto I (i.e., $\varphi(j) = \min(I) - 1 + j$, for all j in $\{1,2,\dots,2t\}$). It is clear then, that the mapping $\varphi^{-1} \circ (\hat{\pi}_{|I}) \circ \varphi$ is a permutation of $\{1,2,\dots,2t\}$, and that we may choose a (unique) permutation π_1 in S_t, such that

$$\hat{\pi}_1 = \varphi^{-1} \circ (\hat{\pi}_{|I}) \circ \varphi, \tag{5.18}$$

(cf. Remark 1.7(a)). It is clear too, that the restriction $\hat{\pi}_{|I}$ of $\hat{\pi}$ to I is non-crossing in the sense of Definition 5.5, if and only if $\hat{\pi}_1$ is a non-crossing permutation in the usual sense (cf. Definition 1.14). □

5.7 Lemma. *Let p be a positive integer, and let π be a permutation in S_p.*
(i) *If I is an interval of integers such that $I \subseteq \{1,2,\dots,2p\}$ and $\hat{\pi}_{|I}$ is non-crossing, then there exists e in I, such that $e+1 \in I$ and $\hat{\pi}(e) = e+1$.*
(ii) *If $\pi \in S_p^{\mathrm{irr}}$, then $\hat{\pi}$ has no non-crossing interval of integers.*

Proof. (i) Assume that $I \subseteq \{1,2,\dots,2p\}$ and that $\hat{\pi}_{|I}$ is non-crossing. Put $t = \frac{1}{2}\mathrm{card}(I)$, let φ be the order preserving bijection of $\{1,2,\dots,2t\}$ onto I, and let π_1 be the permutation in S_t given by (5.18). Then $\pi_1 \in S_t^{\mathrm{nc}}$, and hence $\hat{\pi}_1$ has a pair of neighbors $(e', e'+1)$ by Lemma 1.17. Putting $e = \varphi(e')$, it follows that $e+1 = \hat{\pi}(e) \in I$, and this proves (i).
(ii) This follows immediately from (i). □

5.8 Lemma. *Let p be a positive integer, and let π be a permutation in S_p, such that $\hat{\pi}$ is reducible. Consider furthermore a family $(I_\lambda)_{\lambda\in\Lambda}$ of intervals of integers, such that $I_\lambda \subseteq \{1,2,\dots,2p\}$ for all λ, and such that the union $I = \cup_{\lambda\in\Lambda} I_\lambda$ is again an interval of integers. If each I_λ is a non-crossing interval of integers for $\hat{\pi}$, then so is I.*

Proof. Assume that each I_λ is a non-crossing interval of integers for $\hat{\pi}$. Then $\hat{\pi}(I_\lambda) = I_\lambda$ for all λ, and hence also $\hat{\pi}(I) = I$. Assume then that I contains a crossing for $\hat{\pi}$, i.e., that there exist a,b,c,d in I, such that $a < b < c < d$ and $\hat{\pi}(a) = c$, $\hat{\pi}(b) = d$. Choose λ in Λ such that $a \in I_\lambda$. Then $c = \hat{\pi}(a) \in I_\lambda$, and since I_λ is an interval of integers, also $b \in I_\lambda$. But then $d = \hat{\pi}(b) \in I_\lambda$ too, and hence (a,b,c,d) is a crossing for $\hat{\pi}$ contained in I_λ; a contradiction. Therefore I too is a non-crossing interval of integers for $\hat{\pi}$. □

5.9 Definition. Let p be a positive integer and let π be a permutation in S_p. By $\mathcal{J}(\hat{\pi})$ we denote then the family of all non-crossing intervals of integers for

$\hat{\pi}$. Moreover, we put

$$\mathrm{NC}(\hat{\pi}) = \bigcup_{I \in \mathcal{J}(\hat{\pi})} I, \tag{5.19}$$

$$\mathrm{IRR}(\hat{\pi}) = \{1, 2, \dots, 2p\} \setminus \mathrm{NC}(\hat{\pi}). \tag{5.20}$$

We refer to $\mathrm{NC}(\hat{\pi})$ (respectively $\mathrm{IRR}(\hat{\pi})$) as the non-crossing set (respectively irreducible set) for $\hat{\pi}$. □

5.10 Lemma. *Let p be a positive integer and let π be a permutation in S_p. We then have*
(i) $\mathrm{NC}(\hat{\pi}) = \{1, 2, \dots, 2p\}$ *if and only if $\hat{\pi}$ is non-crossing.*
(ii) $\mathrm{NC}(\hat{\pi}) = \emptyset$ *if and only if $\hat{\pi}$ is irreducible.*

Proof. (i) If $\mathrm{NC}(\hat{\pi}) = \{1, 2, \dots, 2p\}$, then is follows from Lemma 5.8, that $\hat{\pi}$ is non-crossing. If, conversely, $\hat{\pi}$ is non-crossing, then $\{1, 2, \dots, 2p\} \in \mathcal{J}(\hat{\pi})$, and hence $\mathrm{NC}(\hat{\pi}) = \{1, 2, \dots, 2p\}$.
(ii) If $\mathrm{NC}(\hat{\pi}) = \emptyset$, then for any j in $\{1, 2, \dots, 2p-1\}$, $\{j, j+1\}$ can not be a non-crossing interval of integers for $\hat{\pi}$. Hence $\hat{\pi}(j) \neq j+1$ for all j in $\{1, 2, \dots, 2p-1\}$, which means that $\hat{\pi}$ is irreducible. If, conversely, $\hat{\pi}$ is irreducible, then $\mathcal{J}(\hat{\pi}) = \emptyset$ by Lemma 5.7(ii), and hence also $\mathrm{NC}(\hat{\pi}) = \emptyset$. □

5.11 Proposition. *Let p be a positive integer, let π be a permutation in S_p, and assume that $\hat{\pi}$ has a crossing. Then the set $\mathrm{IRR}(\hat{\pi})$ is of the form*

$$\mathrm{IRR}(\hat{\pi}) = \{s_1, s_2, \dots, s_{2q}\},$$

where $q \in \{1, \dots, p\}$, and $1 \le s_1 < s_2 < \cdots < s_{2q} \le 2p$. Moreover, $s_1, s_2, \dots, s_{2q}$ have the following properties:
(i) *The set $\{s_1, s_2, \dots, s_{2q}\}$ is $\hat{\pi}$-invariant and $\hat{\pi}(s_i) \neq s_{i+1}$, for all i in $\{1, 2, \dots, 2q-1\}$.*
(ii) *If we put $s_0 = 0$ and $s_{2q+1} = 2p+1$, then for each i in $\{0, 1, \dots, 2q\}$, the set*

$$I_i =]s_i, s_{i+1}[\cap \mathbb{Z}$$

is either the empty set or a non-crossing interval of integers for $\hat{\pi}$.

Proof. By Definition 5.5(b), each I in $\mathcal{J}(\hat{\pi})$ is $\hat{\pi}$-invariant. Therefore $\mathrm{NC}(\hat{\pi})$ is $\hat{\pi}$-invariant too, and hence so is $\mathrm{IRR}(\hat{\pi})$. Since $\hat{\pi}^2 = \mathrm{id}$ and $\hat{\pi}$ has no fixed points, it follows that $\mathrm{card}(\mathrm{IRR}(\hat{\pi})) = 2q$ for some q in $\{0, 1, \dots, p\}$, and since $\hat{\pi}$ has a crossing, Lemma 5.10(i) shows that $q \ge 1$. Thus, we may write $\mathrm{IRR}(\hat{\pi})$ in the form $\{s_1, s_2, \dots, s_{2q}\}$, where $s_1 < s_2 < \cdots < s_{2q}$, and it remains to show that these $s_1, s_2, \dots, s_{2q}$ satisfy (i) and (ii).
We start by proving (ii). For all I from $\mathcal{J}(\hat{\pi})$, $I \cap \{s_1, s_2, \dots, s_{2q}\} = \emptyset$, and hence each such I is contained in one of the sets $I_i =]s_i, s_{i+1}[\cap \mathbb{Z}$,

$i = 0, 1, \dots, 2q$. Therefore

$$\mathcal{J}(\hat{\pi}) = \bigcup_{i=0}^{2q} \mathcal{J}_i(\hat{\pi}), \tag{5.21}$$

where $\mathcal{J}_i(\hat{\pi}) = \{I \in \mathcal{J}(\hat{\pi}) \mid I \subseteq I_i\}$, for all i in $\{0, 1, \dots, 2q\}$. Note here that

$$\bigcup_{I \in \mathcal{J}_i(\hat{\pi})} I \subseteq I_i, \qquad (i \in \{0, 1, \dots, 2q\}), \tag{5.22}$$

and that

$$\bigcup_{I \in \mathcal{J}(\hat{\pi})} I = \mathrm{NC}(\hat{\pi}) = \{1, 2, \dots, 2p\} \setminus \mathrm{IRR}(\hat{\pi}) = \bigcup_{i=0}^{2q} I_i. \tag{5.23}$$

Combining (5.21)-(5.23), it follows that we actually have equality in (5.22), i.e.,

$$\bigcup_{I \in \mathcal{J}_i(\hat{\pi})} I = I_i, \qquad (i \in \{0, 1, \dots, 2q\}). \tag{5.24}$$

Since each I_i is either empty or an interval of integers, (ii) follows now by combining (5.24) with Lemma 5.8.
It remains to prove (i). We already noted (and used) that $\mathrm{IRR}(\hat{\pi})$ is $\hat{\pi}$-invariant. Assume then that $\hat{\pi}(s_i) = s_{i+1}$ for some i in $\{1, \dots, 2q-1\}$. Then, by (ii), the set

$$\tilde{I}_i = \{s_i\} \cup I_i \cup \{s_{i+1}\},$$

is a non-crossing interval of integers for $\hat{\pi}$. But this contradicts that $s_i \notin \mathrm{NC}(\hat{\pi})$, and hence we have proved (i). $\square$

We prove next the following converse of Proposition 5.11.

5.12 Proposition. *Let p be a positive integer, let π be a permutation in S_p, and assume that there exist q in $\{1, \dots, p\}$ and $s_1 < s_2 < \cdots < s_{2q}$ in $\{1, 2, \dots, 2p\}$, such that*
(i) *The set $\{s_1, s_2, \dots, s_{2q}\}$ is $\hat{\pi}$-invariant and $\hat{\pi}(s_i) \neq s_{i+1}$, for all i in $\{1, 2, \dots, 2q-1\}$.*
(ii) *If we put $s_0 = 0$ and $s_{2q+1} = 2p+1$, then for each i in $\{0, 1, \dots, 2q\}$, the set $I_i = \,]s_i, s_{i+1}[\, \cap \mathbb{Z}$ is either the empty set or a non-crossing interval of integers for $\hat{\pi}$.*
Then $\{s_1, s_2, \dots, s_{2q}\} = \mathrm{IRR}(\hat{\pi})$.

Proof. It follows from (i), that there exists a (unique) permutation γ in S_{2q}, such that

$$\hat{\pi}(s_i) = s_{\gamma(i)}, \qquad (i \in \{1, 2, \dots, 2q\}),$$

and moreover

$$\gamma(i) \neq i+1, \qquad (i \in \{1,2,\dots,2q-1\}). \tag{5.25}$$

Our first objective is to prove that γ is of the form $\hat{\rho}$ for some (unique) permutation ρ in S_q^{irr}. For this, note first that by (ii), $\mathrm{card}(I_i)$ is an even number for all i in $\{0,1,\dots,2q\}$. Hence $s_{i+1} - s_i$ is odd for all i in $\{0,1,\dots,2q\}$, and this implies that

$$\begin{aligned} s_1, s_3, \dots, s_{2q-1} &\ \text{ are odd numbers} \\ s_2, s_4, \dots, s_{2q} &\ \text{ are even numbers} \end{aligned}$$

Since $\hat{\pi}^2 = \mathrm{id}$ and $\hat{\pi}(j) - j$ is odd for all j in $\{1,2,\dots,2p\}$, it follows now that $\gamma^2 = \mathrm{id}$ and that $\gamma(i) - i$ is odd for all i in $\{1,2,\dots,2q\}$. Therefore, by Remark 1.7(a), $\gamma = \hat{\rho}$ for some (unique) ρ in S_q, and (5.25) shows that in fact $\rho \in S_q^{\mathrm{irr}}$.

Returning now to the proof of the equation $\{s_1, s_2, \dots, s_{2q}\} = \mathrm{IRR}(\hat{\pi})$, note first that $\cup_{i=0}^{2q} I_i \subseteq \mathrm{NC}(\hat{\pi})$, and therefore

$$\{s_1, s_2, \dots, s_{2q}\} = \{1,2,\dots,2p\} \setminus \cup_{i=0}^{2q} I_i \supseteq \mathrm{IRR}(\hat{\pi}).$$

Suppose then that $\mathrm{IRR}(\hat{\pi})$ is a proper subset of $\{s_1, s_2, \dots, s_{2q}\}$. Then there exists j_0 in $\{1,2,\dots,2q\}$, such that $s_{j_0} \in \mathrm{NC}(\hat{\pi})$, i.e., $s_{j_0} \in I$, for some non-crossing interval of integers for $\hat{\pi}$. For this I, define

$$J = \big\{ j \in \{1,2,\dots,2q\} \ \big|\ s_j \in I \big\}.$$

Then $J \neq \emptyset$, and since $s_1 < s_2 < \dots < s_{2q}$, J is an interval of integers. Consider now the permutation ρ in S_q^{irr}, introduced above. Then, since $\hat{\pi}(I) = I$, we have also that $\hat{\rho}(J) = J$. Moreover, J is a non-crossing interval of integers for $\hat{\rho}$. Indeed, if (a,b,c,d) were a crossing for $\hat{\rho}$ contained in J, then clearly (s_a, s_b, s_c, s_d) would be a crossing for $\hat{\pi}$ contained in I, which is impossible. Altogether, ρ is both irreducible and has a non-crossing interval of integers, and by Lemma 5.10(ii), this is impossible. Thus, we have reached a contradiction, which means that we must also have the inclusion $\{s_1, s_2, \dots, s_{2q}\} \subseteq \mathrm{IRR}(\hat{\pi})$. $\square$

5.13 Lemma. *Let p be a positive integer, and let π be a permutation in $S_p \setminus S_p^{\mathrm{nc}}$. Write then, as in Proposition 5.11,* $\mathrm{IRR}(\hat{\pi})$ *in the form*

$$\mathrm{IRR}(\hat{\pi}) = \{s_1, s_2, \dots, s_{2q}\},$$

where $q \in \{1,\dots,p\}$ and $1 \le s_1 < s_2 < \dots < s_{2q} \le 2p$. Then $s_1, s_2, \dots, s_{2q}$ satisfy, in addition, that

(i) *$s_1, s_3, \dots, s_{2q-1}$ are odd numbers.*

(ii) *$s_2, s_4, \dots, s_{2q}$ are even numbers.*

(iii) *There is one and only one permutation ρ in S_q^{irr}, such that $\hat{\pi}(s_j) = s_{\hat{\rho}(j)}$ for all j in $\{1,2,\dots,2q\}$.*

Proof. This follows immediately from Proposition 5.11 and the first part of the proof of Proposition 5.12. □

5.14 Definition. Let p be a positive integer, let π be a permutation in $S_p \backslash S_p^{\rm nc}$, and let q, $s_1, s_2, \dots, s_{2q}$ and $I_0, I_1, \dots, I_{2q}$, be as in Proposition 5.11. Then put

$$t_i = \tfrac{1}{2}\mathrm{card}(I_i), \qquad (i \in \{0, 1, \dots, 2q\}),$$

and note that since I_i is either empty or a non-crossing interval of integers for $\hat{\pi}$, $t_i \in \mathbb{N}_0$ for all i. If $t_i > 0$, then as in Remark 5.6, we consider the order-preserving bijection φ_i of $\{1, 2, \dots, 2t_i\}$ onto I_i, and we let π_i denote the (unique) permutation in S_{t_i}, satisfying that $\hat{\pi}_i = \varphi_i^{-1} \circ (\hat{\pi}_{|I_i}) \circ \varphi$. Clearly $\pi_i \in S_p^{\rm nc}$.
It is convenient to consider the permutation group S_0 of the empty set, as a group with one element $\pi_\emptyset$. Then, in the setting considered above, we put $\pi_i = \pi_\emptyset$, for all i in $\{0, 1, \dots, 2q\}$, for which $t_i = 0$. By convention, we put

$$k(\hat{\pi}_\emptyset) = 0, \qquad \text{and} \qquad l(\hat{\pi}_\emptyset) = 1. \tag{5.26}$$

□

5.15 Lemma. *Let p be a positive integer, let π be a permutation in $S_p \setminus S_p^{\rm nc}$, and let ρ be the irreducible permutation introduced in Lemma 5.13(iii). Then $\sigma(\hat{\rho}) = \sigma(\hat{\pi})$.*

Proof. Let q, $s_1, s_2, \dots, s_{2q}$ and $I_0, I_1, \dots, I_{2q}$, be as in Proposition 5.11, and for each i in $\{0, 1, \dots, 2q\}$, let t_i and π_i be as in Definition 5.14. If $t_i > 0$, then $\hat{\pi}_i$ is non-crossing, and hence, by Proposition 1.23, $\hat{\pi}_i$ may be reduced to $\hat{e}_1$ (where e_1 is the permutation in S_1), by a series of successive cancellations of pairs. Here $\hat{e}_1$ consists exactly of one pair of neighbors, so, formally speaking, $\hat{e}_1$ can be reduced $\hat{\pi}_\emptyset$, by cancellation of this pair. Thus, $\hat{\pi}_i$ can be reduced to $\hat{\pi}_\emptyset$, by a series of successive cancellations of pairs, and forming the corresponding series of cancellations of pairs to $\hat{\pi}_{|I_i}$, it follows that $\hat{\pi}$ can be reduced to a permutation, which is, loosely speaking, obtained by "cutting out" $\hat{\pi}_{|I_i}$ from $\hat{\pi}$. Forming these reductions for each i in $\{0, 1, \dots, 2q\}$, for which $t_i > 0$, it follows that $\hat{\pi}$ can be reduced to $\hat{\rho}$ by a series of successive cancellations of pairs. By Proposition 1.22, this implies that $\sigma(\hat{\pi}) = \sigma(\hat{\rho})$. □

5.16 Proposition. *Let p be a positive integer, let π be a permutation in $S_p \setminus S_p^{\rm nc}$, and let q, $s_1, s_2, \dots, s_{2q}$ be as in Proposition 5.11. Let further ρ be the permutation in $S_q^{\rm irr}$ introduced in Lemma 5.13(iii), and let $\pi_0, \pi_1, \dots, \pi_{2q}$ be as in Definition 5.14. Then for any elements $a_1, \dots, a_r$ of $\mathcal{B}(\mathcal{H}, \mathcal{K})$ for which* (5.2) *holds, we have*

$$\sum_{1 \le i_1, \dots, i_p \le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}} = c^{h(\hat{\pi})} \sum_{1 \le i_1, \dots, i_q \le r} a_{i_1}^* a_{i_{\rho(1)}} \cdots a_{i_q}^* a_{i_{\rho(q)}}, \tag{5.27}$$

where

$$h(\hat\pi) = k(\hat\pi_0) + (l(\hat\pi_1) - 1) + k(\hat\pi_2) + \cdots + (l(\hat\pi_{2q-1}) - 1) + k(\hat\pi_{2q}). \tag{5.28}$$

Proof. We start by introducing some notation. Let t be a positive integer, and let η be a permutation in S_t. We then put

$$\Gamma(\hat\eta) = \sum_{1\le i_1,\dots,i_t\le r} a_{i_1}^* a_{i_{\eta(1)}} \cdots a_{i_t}^* a_{i_{\eta(t)}}, \tag{5.29}$$

and moreover, we put

$$\Gamma(\hat\pi_\emptyset) = \mathbf{1}_{\mathcal{B}(\mathcal{H})}. \tag{5.30}$$

Note that $\Gamma(\hat\eta)$ can be expressed in terms of $\hat\eta$ only, namely as

$$\Gamma(\hat\eta) = \sum_{(i_1,i_2,i_3,i_4,\dots,i_{2t})\in N(\hat\eta)} a_{i_1}^* a_{i_2} a_{i_3}^* a_{i_4} \cdots a_{i_{2t-1}}^* a_{i_{2t}}, \tag{5.31}$$

where

$$\begin{aligned} &N(\hat\eta) \\ &= \big\{(i_1,i_2,\dots,i_{2t}) \in \{1,2,\dots,r\}^{2t} \;\big|\; i_j = i_{\hat\eta(j)}, \text{ for all } j \text{ in } \{1,2,\dots,2t\}\big\}, \end{aligned} \tag{5.32}$$

(cf. Remark 1.7(b)). Consider next an interval of integers I, such that $I \subseteq \{1,2,\dots,2t\}$ and $\hat\eta(I) = I$. Write I in the form $\{\alpha,\alpha+1,\dots,\beta\}$, and note that $\beta-\alpha+1 = \mathrm{card}(I)$ is an even number. We then put

$$N(\hat\eta, I) = \big\{(i_\alpha,\dots,i_\beta) \in \{1,2,\dots,r\}^{\beta-\alpha+1} \;\big|\; i_j = i_{\hat\eta(j)},\ j = \alpha,\alpha+1,\dots,\beta\big\} \tag{5.33}$$

and

$$\Gamma(\hat\eta, I) = \begin{cases} \displaystyle\sum_{(i_\alpha,\dots,i_\beta)\in N(\hat\eta,I)} a_{i_\alpha}^* a_{i_{\alpha+1}} \cdots a_{i_{\beta-1}}^* a_{i_\beta}, & \text{if } \alpha \text{ is odd,} \\ \displaystyle\sum_{(i_\alpha,\dots,i_\beta)\in N(\hat\eta,I)} a_{i_\alpha} a_{i_{\alpha+1}}^* \cdots a_{i_{\beta-1}} a_{i_\beta}^*, & \text{if } \alpha \text{ is even.} \end{cases} \tag{5.34}$$

Now, to prove (5.27), consider p in $\mathbb{N}$ and π in $S_p \setminus S_p^{\mathrm{nc}}$, and let q, $s_1, s_2, \dots, s_{2q}$ and $I_0, I_1, \dots, I_{2q}$, $t_0, t_1, \dots, t_{2q}$ be as in Proposition 5.11. Note then, that we may write $N(\hat\pi)$ as

$$\begin{aligned} &N(\hat\pi) = \\ &\bigcup_{(i_{s_1},\dots,i_{s_{2q}})\in N_1(\hat\pi)} N(\hat\pi, I_0) \times \{i_{s_1}\} \times N(\hat\pi, I_1) \times \{i_{s_2}\} \times \cdots \times \{i_{s_{2q}}\} \times N(\hat\pi, I_{2p}), \end{aligned} \tag{5.35}$$

with the convention that $N(\hat{\pi}, I_i)$ is omitted in the product sets when $2t_i = \mathrm{card}(I_i) = 0$, and where

$$N_1(\hat{\pi}) = \big\{(i_{s_1}, \dots, i_{s_{2q}}) \in \{1,2,\dots,r\}^{2q} \bigm| i_{s_j} = i_{\hat{\pi}(s_j)},\ j = 1,2,\dots,2q\big\}. \tag{5.36}$$

It follows thus, by (5.31), that

$$\Gamma(\hat{\pi}) = \sum_{(i_{s_1},\dots,i_{s_{2q}})\in N_1(\hat{\pi})} \Gamma(\hat{\pi}, I_0) a^*_{i_{s_1}} \Gamma(\hat{\pi}, I_1) a_{i_{s_2}} \cdots a_{i_{s_{2q}}} \Gamma(\hat{\pi}, I_{2q}), \tag{5.37}$$

with the convention that if $\mathrm{card}(I_i) = 0$,

$$\Gamma(\hat{\pi}, I_i) = \begin{cases} \mathbf{1}_{\mathcal{B}(\mathcal{H})}, & \text{if } s_i \text{ is even,} \\ \mathbf{1}_{\mathcal{B}(\mathcal{K})}, & \text{if } s_i \text{ is odd.} \end{cases} \tag{5.38}$$

To calculate $\Gamma(\hat{\pi}, I_0), \dots, \Gamma(\hat{\pi}, I_{2q})$, consider the non-crossing permutations $\pi_0, \pi_1, \dots, \pi_{2q}$ introduced in Definition 5.14. Note then, that for each v in $\{0, 1, \dots, 2q\}$, such that $t_v > 0$, we have by a suitable relabeling of indices,

$$\begin{aligned} N(\hat{\pi}, I_v) &= \big\{(i_1, i_2, \dots, i_{2t_v}) \in \{1,2,\dots,r\}^{2t_v} \bigm| i_j = i_{\hat{\pi}_v(j)},\ j = 1,2,\dots,2t_v\big\} \\ &= N(\hat{\pi}_v). \end{aligned}$$

It follows thus, that if $t_v > 0$,

$$\Gamma(\hat{\pi}, I_v) = \begin{cases} \displaystyle\sum_{1\le i_1,\dots,i_{t_v}\le r} a^*_{i_1} a_{i_{\pi_v(1)}} \cdots a^*_{i_{t_v}} a_{i_{\pi_v(t_v)}}, & \text{if } v \text{ is even,} \\ \displaystyle\sum_{1\le i_1,\dots,i_{t_v}\le r} a_{i_1} a^*_{i_{\pi_v(1)}} \cdots a_{i_{t_v}} a^*_{i_{\pi_v(t_v)}}, & \text{if } v \text{ is odd,} \end{cases}$$

and hence by Lemma 5.2 (since $\hat{\pi}_v$ is non-crossing),

$$\Gamma(\hat{\pi}, I_v) = \begin{cases} c^{k(\hat{\pi}_v)} \mathbf{1}_{\mathcal{B}(\mathcal{H})}, & \text{if } v \text{ is even,} \\ c^{l(\hat{\pi}_v)-1} \mathbf{1}_{\mathcal{B}(\mathcal{K})}, & \text{if } v \text{ is odd.} \end{cases} \tag{5.39}$$

If $t_v = 0$, then by definition,

$$\Gamma(\hat{\pi}, I_v) = \begin{cases} \mathbf{1}_{\mathcal{B}(\mathcal{H})}, & \text{if } v \text{ is even,} \\ \mathbf{1}_{\mathcal{B}(\mathcal{K})}, & \text{if } v \text{ is odd,} \end{cases} = \begin{cases} c^{k(\hat{\pi}_v)} \mathbf{1}_{\mathcal{B}(\mathcal{H})}, & \text{if } v \text{ is even,} \\ c^{l(\hat{\pi}_v)-1} \mathbf{1}_{\mathcal{B}(\mathcal{K})}, & \text{if } v \text{ is odd,} \end{cases} \tag{5.40}$$

with $k(\hat{\pi}_\emptyset), l(\hat{\pi}_\emptyset)$ as defined in (5.26). Combining (5.37),(5.39) and (5.40), it follows that with $h(\hat{\pi})$ given in (5.28), we have

$$\Gamma(\hat{\pi}) = c^{h(\hat{\pi})} \sum_{(i_{s_1},\dots,i_{s_{2q}})\in N_1(\hat{\pi})} a^*_{i_{s_1}} a_{i_{s_2}} \cdots a^*_{i_{2q-1}} a_{i_{s_{2q}}}. \tag{5.41}$$

Note finally, that with ρ the permutation introduced in Lemma 5.13(iii), we have that

$$\begin{aligned} N_1(\hat{\pi}) &= \big\{(i_1, i_2, \dots, i_{2q}) \in \{1, 2, \dots, r\}^{2q} \bigm| i_j = i_{\hat{\rho}(j)},\ j = 1, 2, \dots, 2q\big\} \\ &= N(\hat{\rho}), \end{aligned}$$

and therefore

$$\sum_{(i_{s_1}, \dots, i_{s_{2q}}) \in N_1(\hat{\pi})} a_{i_{s_1}}^* a_{i_{s_2}} \cdots a_{i_{2q-1}}^* a_{i_{s_{2q}}} = \sum_{1 \le i_1, \dots, i_q \le r} a_{i_1}^* a_{i_{\rho(1)}} \cdots a_{i_q}^* a_{i_{\rho(q)}}.$$

Inserting this in (5.41), we obtain (5.27). □

5.17 DEFINITION. Let c be a positive number. Then for any p in $\mathbb{N}_0$, we define

$$g_c(p) = \begin{cases} \frac{1}{p} \sum_{j=1}^{p} \binom{p}{j} \binom{p}{j-1} c^j, & \text{if } p \in \mathbb{N}, \\ 1, & \text{if } p = 0, \end{cases} \tag{5.42}$$

and

$$h_c(p) = \begin{cases} \frac{1}{p} \sum_{j=1}^{p} \binom{p}{j} \binom{p}{j-1} c^{j-1}, & \text{if } p \in \mathbb{N}, \\ 1, & \text{if } p = 0. \end{cases} \tag{5.43}$$

Moreover, for p, q in $\mathbb{N}_0$, such that $p \ge q$, we put

$$\nu'(c, p, q) = \sum_{\substack{r_0, r_1, \dots, r_{2q} \ge 0 \\ r_0 + r_1 + \cdots + r_{2q} = p - q}} g_c(r_0) h_c(r_1) g_c(r_2) h_c(r_3) \cdots g_c(r_{2q}). \tag{5.44}$$

□

We are now ready to prove the main result of this section.

5.18 THEOREM. *Let $a_1, \dots, a_r$ be elements of $\mathcal{B}(\mathcal{H}, \mathcal{K})$, let c be a positive number, and assume that $\sum_{i=1}^r a_i^* a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$, and $\sum_{i=1}^r a_i a_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{K})}$. Consider furthermore independent elements $Y_1, \dots, Y_r$ of $\mathrm{GRM}(n, n, \frac{1}{n})$, and put $S = \sum_{i=1}^r a_i \otimes Y_i$. Then for any positive integer p,*

$$\begin{aligned} &\mathbb{E}\big[(S^*S)^p\big] \\ &= \Big[\nu'(c, p, 0)\mathbf{1}_{\mathcal{B}(\mathcal{H})} \\ &+ \sum_{q=1}^{p} \nu'(c, p, q) \Big(\sum_{\rho \in S_q^{\text{irr}}} n^{-2\sigma(\hat{\rho})} \sum_{1 \le i_1, \dots, i_q \le r} a_{i_1}^* a_{i_{\rho(1)}} \cdots a_{i_q}^* a_{i_{\rho(q)}} \Big) \Big] \otimes \mathbf{1}_n. \end{aligned} \tag{5.45}$$

Proof. Let p from $\mathbb{N}$ be given. Then for each q in $\{1,2,\dots,p\}$, we define

$$S_{p,q} = \big\{\pi \in S_p \;\big|\; \mathrm{card}(\mathrm{IRR}(\hat{\pi})) = 2q\big\}, \tag{5.46}$$

and

$$M_q = \sum_{\pi\in S_{p,q}} n^{-2\sigma(\hat{\pi})}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big) \tag{5.47}$$

It follows then by (5.1), that

$$\begin{aligned}\mathbb{E}\big[(S^*S)^p\big] &= \Big[\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big)\Big]\otimes \mathbf{1}_n\\ &= \sum_{q=0}^{p} M_q\otimes\mathbf{1}_n.\end{aligned} \tag{5.48}$$

By Lemma 5.10, $S_{p,0} = S_p^{\mathrm{nc}}$ and $S_{p,p} = S_p^{\mathrm{irr}}$. Hence

$$M_p = \sum_{\pi\in S_p^{\mathrm{irr}}} n^{-2\sigma(\hat{\pi})}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big), \tag{5.49}$$

and by Corollary 5.4(i) and Corollary 1.24,

$$M_0 = g_c(p)\mathbf{1}_{\mathcal{B}(\mathcal{H})} = \nu'(c,p,0)\mathbf{1}_{\mathcal{B}(\mathcal{H})}. \tag{5.50}$$

To calculate $M_1, M_2\dots, M_{p-1}$, we let, for each π in S_p, $\rho(\pi)$ denote the irreducible permutation ρ associated to π in Lemma 5.13(iii). Then for any q in $\{1,2,\dots,p-1\}$ and any ρ in S_q^{irr}, we define

$$R(p,\rho) = \{\pi\in S_{p,q} \mid \rho(\pi)=\rho\}.$$

Then we have the following disjoint union

$$S_{p,q} = \dot{\bigcup_{\rho\in S_q^{\mathrm{irr}}}} R(p,\rho),$$

and therefore

$$M_q = \sum_{\rho\in S_q^{\mathrm{irr}}}\sum_{\pi\in R(p,\rho)} n^{-2\sigma(\hat{\pi})}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big). \tag{5.51}$$

Note here, that for any ρ in S_q^{irr}, we have by Proposition 5.16 and Lemma 5.15,

$$\begin{aligned}&\sum_{\pi\in R(p,\rho)} n^{-2\sigma(\hat{\pi})}\Big(\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}}\cdots a_{i_p}^* a_{i_{\pi(p)}}\Big)\\ &\qquad = \Big(\sum_{\pi\in R(p,\rho)} c^{h(\hat{\pi})}\Big) n^{-2\sigma(\hat{\rho})}\Big(\sum_{1\le i_1,\dots,i_q\le r} a_{i_1}^* a_{i_{\rho(1)}}\cdots a_{i_q}^* a_{i_{\rho(q)}}\Big),\end{aligned} \tag{5.52}$$

where for each π in $R(p,\rho)$,

$$h(\hat{\pi}) = k(\hat{\pi}_0) + (l(\hat{\pi}_1) - 1) + k(\hat{\pi}_2) + \cdots + (l(\hat{\pi}_{2q-1}) - 1) + k(\hat{\pi}_{2q}),$$

and where $\pi_0, \pi_1, \dots, \pi_{2q}$ are the permutations introduced in Definition 5.14. For any ρ in $S_q^{\rm irr}$ and any π in $R(p,\rho)$, it follows from Proposition 5.11 and Lemma 5.13, that $\hat{\pi}$ can be obtained from $\hat{\rho}$ in a unique way, by "stuffing in" the intervals (or empty sets) $I_0, I_1, \dots, I_{2q}$, and the corresponding non-crossing permutations $\hat{\pi}_0, \hat{\pi}_1, \dots, \hat{\pi}_{2q}$. Conversely, if $\pi \in S_p$ such that $\hat{\pi}$ can be obtained from $\hat{\rho}$ by "stuffing in" intervals (or empty sets) $J_0, J_1, \dots, J_{2q}$ and corresponding non-crossing permutations $\hat{\eta}_0, \hat{\eta}_1, \dots, \hat{\eta}_{2q}$, then, by Proposition 5.12, $\pi \in R(p,\rho)$ and $J_j = I_j$, $\eta_j = \pi_j$, for all j in $\{0, 1, \dots, 2q\}$. It follows thus, that the mapping

$$\pi \mapsto (\pi_0, \pi_1, \dots, \pi_{2q})$$

is a bijection of $R(p,\rho)$ onto the set of $(2q+1)$-tuples $(\pi_0, \pi_1, \dots, \pi_{2q})$ of permutations for which there exist $t_0, t_1, \dots, t_{2q}$ in $\mathbb{N}_0$, such that $\pi_i \in S_{t_i}^{\rm nc}$ for all i, and $\sum_{i=0}^{2q} t_i = p - q$ (here we have used the convention that $S_0^{\rm nc} = S_0 = \{\pi_\emptyset\}$). Using this description of $R(p,\rho)$, it follows that

$$\sum_{\pi \in R(p,\rho)} c^{h(\hat{\pi})} = \sum_{\substack{t_0, \dots, t_{2q} \geq 0 \\ t_0 + \cdots + t_{2q} = p-q}} \sum_{\pi_0 \in S_{t_0}^{\rm nc}, \dots, \pi_{2q} \in S_{t_{2q}}^{\rm nc}} c^{k(\hat{\pi}_0)} c^{(l(\hat{\pi}_1)-1)} c^{k(\hat{\pi}_2)} \cdots c^{k(\hat{\pi}_{2q})}. \tag{5.53}$$

Recall here from Definition 5.17 and Lemma 5.3, that for any t in $\mathbb{N}$,

$$\sum_{\eta \in S_t^{\rm nc}} c^{k(\hat{\eta})} = g_c(t), \qquad \text{and} \qquad \sum_{\eta \in S_t^{\rm nc}} c^{l(\hat{\eta})-1} = h_c(t),$$

and by (5.26) this formula holds for $t = 0$ too. Using this in (5.53), it follows that

$$\begin{aligned} \sum_{\pi \in R(p,\rho)} c^{h(\hat{\pi})} &= \sum_{\substack{t_0, t_1, \dots, t_{2q} \geq 0 \\ t_0 + t_1 + \cdots + t_{2q} = p-q}} g_c(t_0) h_c(t_1) g_c(t_2) h_c(t_3) \cdots g_c(t_{2q}) \\ &= \nu'(c,p,q). \end{aligned} \tag{5.54}$$

Note, in particular, that the right hand side depends only on p and q, and not on ρ itself. Combining (5.51),(5.52) and (5.54), it follows that for any q in $\{1, 2, \dots, p-1\}$,

$$M_q = \nu'(c,p,q) \sum_{\rho \in S_q^{\rm irr}} n^{-2\sigma(\hat{\rho})} \Big(\sum_{1 \leq i_1, \dots, i_q \leq r} a_{i_1}^* a_{i_{\rho(1)}} \cdots a_{i_q}^* a_{i_{\rho(q)}} \Big). \tag{5.55}$$

Since $\nu'(c,p,p) = 1$, (5.55) holds for $q = p$ too, by (5.49), and combining this with (5.48) and (5.50), we obtain, finally, (5.45). $\square$

5.19 Proposition. *Let $a_1, \dots, a_r$ in $\mathcal{B}(\mathcal{H},\mathcal{K})$, c in $]0,\infty[$ and $S = \sum_{i=1}^r a_i \otimes Y_i$, be as in Theorem 5.18. Then for any p in $\mathbb{N}$, we have that*

$$\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\Big\| \sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^* a_{i_{\pi(1)}} \cdots a_{i_p}^* a_{i_{\pi(p)}}\Big\|$$
$$= \nu'(c,p,0) + \sum_{q=1}^p \nu'(c,p,q) \sum_{\rho\in S_q^{\mathrm{irr}}} n^{-2\sigma(\hat{\rho})}\Big\| \sum_{1\le i_1,\dots,i_q\le r} a_{i_1}^* a_{i_{\rho(1)}} \cdots a_{i_q}^* a_{i_{\rho(q)}}\Big\|.$$

Proof. This follows by exactly the same proof as for Theorem 5.18. □

5.20 Example. Let $a_1, \dots, a_r$ in $\mathcal{B}(\mathcal{H},\mathcal{K})$ and c from $]0,\infty[$ be as in Theorem 5.18.
(a) For $p=1$ or $p=2$, we have $S_p = S_p^{\mathrm{nc}}$. Hence by (5.1), Corollary 1.24 and Corollary 5.4(i), we get that

$$\mathbb{E}\big[S^*S\big] = c\mathbf{1}_{\mathcal{B}(\mathcal{H})\otimes M_n(\mathbb{C})}, \qquad \text{and} \qquad \mathbb{E}\big[(S^*S)^2\big] = (c^2+c)\mathbf{1}_{\mathcal{B}(\mathcal{H})\otimes M_n(\mathbb{C})}.$$

This can also easily be obtained directly from (5.1) and (5.2).
(b) For $p=3$, $\mathrm{card}(S_3)=6$ and $\mathrm{card}(S_3^{\mathrm{nc}})=5$. The only element of $S_3 \setminus S_3^{\mathrm{nc}}$ is the irreducible permutation π given by

$$\pi(1)=3,\ \pi(2)=1,\ \pi(3)=2.$$

For this π, $\sigma(\hat{\pi})=1$, and it follows then by (5.1) and Corollary 5.4(i), that

$$\mathbb{E}\big[(S^*S)^3\big] = (c^3+3c^2+c)\mathbf{1}_{\mathcal{B}(\mathcal{H})\otimes M_n(\mathbb{C})} + \Big(n^{-2}\sum_{i,j,k=1}^r a_i^* a_k a_j^* a_i a_k^* a_j\Big)\otimes \mathbf{1}_n.$$

This follows also from Theorem 5.18, because $S_1^{\mathrm{irr}} = S_2^{\mathrm{irr}} = \emptyset$ and $S_3^{\mathrm{irr}} = \{\pi\}$. □

6 The Sequence of Orthogonal Polynomials for the Measure μ_c

Throughout this section we consider a fixed positive constant c, and elements $a_1, \dots, a_r$ of $\mathcal{B}(\mathcal{H},\mathcal{K})$, satisfying that

$$\sum_{i=1}^r a_i^* a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})} \qquad \text{and} \qquad \sum_{i=1}^r a_i a_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{H})}.$$

Moreover, we put

$$S = \sum_{i=1}^r a_i \otimes Y_i,$$

where $Y_1, \dots, Y_r$ are independent elements of $\mathrm{GRM}(n,n,\frac{1}{n})$.

As in Section 3, we let μ_c denote the probability measure on $\mathbb{R}$, given by

$$\mu_c = \max\{1-c, 0\}\delta_0 + \frac{\sqrt{(x-a)(b-x)}}{2\pi x} \cdot 1_{[a,b]}(x) \cdot dx,$$

where $a = (\sqrt{c}-1)^2$, $b = (\sqrt{c}+1)^2$.
The asymptotic upper bound for the spectrum of S^*S obtained in Section 4 (in the exact case), was obtained by making careful estimates of the moments $\mathbb{E}\big[(S^*S)^p\big]$, $p \in \mathbb{N}$. However, these estimates cannot be used to give good asymptotic lower bounds for the spectrum of S^*S in the case $c > 1$. To obtain such lower bounds, we shall instead consider the operators $\mathbb{E}\big[P_q^c(S^*S)\big]$, where $(P_q^c)_{q\in\mathbb{N}_0}$ is the sequence of *monic* polynomials, obtained by Gram-Schmidt orthogonalization of the polynomials $1, x, x^2, \dots$, w.r.t. the inner product

$$\langle f, g\rangle = \int_0^\infty f(x)\overline{g(x)}\; d\mu_c(x), \qquad (f, g \in L^2(\mathbb{R}, \mu_c)).$$

The main result of this section is the equation

$$\mathbb{E}\big[P_q^c(S^*S)\big] = \Big[\sum_{\rho\in S_q^{\mathrm{irr}}} n^{-2\sigma(\hat{\rho})}\Big(\sum_{1\le i_1,\dots,i_q\le r} a_{i_1}^* a_{i_{\rho(1)}} \cdots a_{i_q}^* a_{i_{\rho(q)}}\Big)\Big]\otimes \mathbf{1}_n,\ (q \in \mathbb{N}),$$

where S_q^{irr} is the set of permutations ρ in S_q, satisfying that

$$1 \ne \rho(1) \ne 2 \ne \rho(2) \ne \cdots \ne \rho(q)$$

(cf. Definition 1.16).

6.1 Proposition. *Let $(P_q^c)_{q\in\mathbb{N}_0}$ be the sequence of polynomials on $\mathbb{R}$, defined by the recursion formulas:*

$$P_0^c(x) = 1, \tag{6.1}$$
$$P_1^c(x) = x - c, \tag{6.2}$$
$$P_{q+1}^c(x) = (x-c-1)P_q^c(x) - cP_{q-1}^c(x), \qquad (q \ge 1). \tag{6.3}$$

We then have
(i) *For each q in $\mathbb{N}_0$, $P_q^c(x)$ is a monic polynomial of degree q, and $P_q^c(x) \in \mathbb{R}$ for all real numbers x.*

$$\text{(ii)}\quad P_q^c(c+1+2\sqrt{c}\cos\theta) = \frac{c^{\frac{q}{2}}\sin((q+1)\theta) + c^{\frac{q-1}{2}}\sin(q\theta)}{\sin\theta}, \quad (\theta \in\,]0,\pi[\,).$$

$$\text{(iii)}\qquad \int_a^b P_q^c(x)P_{q'}^c(x)\; d\mu_c(x) = \begin{cases} c^q, & \textit{if} \quad q = q', \\ 0, & \textit{if} \quad q \ne q', \end{cases} \qquad (q, q' \in \mathbb{N}_0).$$

In particular, $(P_q^c)_{q\in\mathbb{N}_0}$ is the sequence of monic orthogonal polynomials obtained by Gram-Schmidt orthogonalization of $1, x, x^2, \dots$, in the Hilbert space $L^2(\mathbb{R}, \mu_c)$.

Proof. (i) This is clear from (6.1)-(6.3).
(ii) Consider the sequences $(R_q^c)_{q\in\mathbb{N}_0}$ and $(T_q^c)_{q\in\mathbb{N}_0}$ of polynomials, given by the recursion formulas

$$\begin{aligned} R_0^c(x) &= 1, && (6.4)\\ R_1^c(x) &= x-c-1, && (6.5)\\ R_{q+1}^c(x) &= (x-c-1)R_q^c(x) - cR_{q-1}^c(x), \qquad (q\geq 1), && (6.6)\end{aligned}$$

respectively

$$\begin{aligned} T_0^c(x) &= 0, && (6.7)\\ T_1^c(x) &= 1, && (6.8)\\ T_{q+1}^c(x) &= (x-c-1)T_q^c(x) - cT_{q-1}^c(x), \qquad (q\geq 1). && (6.9)\end{aligned}$$

Note here, that the conditions (6.6) and (6.9) are the same, and therefore, the sequence $(R_q+T_q)_{q\in\mathbb{N}_0}$ satisfies this condition too. Moreover, the sequence $(R_q+T_q)_{q\in\mathbb{N}_0}$ also satisfies (6.1) and (6.2), and it follows thus, that

$$P_q^c(x) = R_q^c(x) + T_q^c(x), \qquad (q\in\mathbb{N}_0).$$

Note also, that $T_2^c(x) = x-c-1$, so that the sequence $(T_{q+1}^c)_{q\in\mathbb{N}_0}$ satisfies (6.4)-(6.6), and hence

$$T_q^c(x) = R_{q-1}^c(x), \qquad (q\in\mathbb{N}).$$

Altogether, it follows that

$$\begin{aligned} P_q^c(x) &= R_q^c(x) + R_{q-1}^c(x), \qquad (q\geq 1), && (6.10)\\ P_0^c(x) &= R_0^c(x). && (6.11)\end{aligned}$$

To prove (ii), it suffices therefore to show, that with $x = c+1+2\sqrt{c}\cos\theta$, $\theta\in\,]0,\pi[$, one has

$$R_q^c(x) = \frac{c^{\frac{q}{2}}\sin((q+1)\theta)}{\sin\theta}, \qquad (q\in\mathbb{N}_0). \tag{6.12}$$

For $q=0$, this is clear from (6.4), and for $q=1$, it follows easily from (6.5), using that $\sin 2\theta = 2\sin\theta\cos\theta$. Proceeding by induction, assume now that $p\geq 1$ and that (6.12) has been proved for all q in $\{0,1,\dots,p\}$. Then by (6.6),

$$R_{p+1}^c(x) = \frac{2\sqrt{c}\cos\theta\cdot c^{\frac{p}{2}}\sin((p+1)\theta)}{\sin\theta} - \frac{c^{\frac{p+1}{2}}\sin(p\theta)}{\sin\theta},$$

when $x = c+1+2\sqrt{c}\cos\theta$, $\theta\in\,]0,\pi[$. But $2\cos\theta\sin((p+1)\theta) = \sin((p+2)\theta) + \sin(p\theta)$, and therefore

$$R_{p+1}^c(x) = \frac{c^{\frac{p+1}{2}}\sin((p+2)\theta)}{\sin\theta},$$

which means that (6.12) holds for $q = p + 1$. Thus, by induction, (6.12) holds for all q in $\mathbb{N}_0$, and this concludes the proof of (ii).
(iii) We show first, that for any m, n in $\mathbb{N}_0$,

$$\int_0^\infty x R_m^c(x) R_n^c(x)\ d\mu_c(x) = \begin{cases} 0, & \text{if } n \neq m, \\ c^{m+1}, & \text{if } n = m, \end{cases} \tag{6.13}$$

where $R_0^c, R_1^c, R_2^c, \ldots$, are the polynomials determined by (6.4)-(6.6). Note for this, that if $c < 1$, then the atom for μ_c at 0, does not contribute to the integral on the left hand side of (6.13). Hence, for all values of c in $]0, \infty[$, we have

$$\int_0^\infty x R_m^c(x) R_n^c(x)\ d\mu_c(x) = \tfrac{1}{2\pi} \int_a^b R_m^c(x) R_n^c(x) \sqrt{(x-a)(b-x)}\ dx. \tag{6.14}$$

By the substitution $x = c + 1 + 2\sqrt{c}\cos\theta$, $\theta \in]0, \pi[$, and by (6.12), the integral on the right hand side of (6.14) can be reduced to

$$\tfrac{2c}{\pi} \int_0^\pi c^{\frac{m+n}{2}} \sin((m+1)\theta) \sin((n+1)\theta)\ d\theta,$$

which is equal to $c^{m+1}\delta_{m,n}$. This proves (6.13).
We show next that

$$x R_m^c(x) = P_{m+1}^c(x) + c P_m^c(x), \qquad (m \in \mathbb{N}_0). \tag{6.15}$$

For $m = 0$, this is clear from (6.1),(6.2) and (6.4), and for $m \geq 1$, we get from (6.6) and (6.10), that

$$x R_m^c(x) = R_{m+1}^c(x) + (c+1) R_m^c(x) + c R_{m-1}^c(x) = P_{m+1}^c(x) + c P_m^c(x).$$

This proves (6.15). Define now

$$\gamma_{m,n} = \int_0^\infty P_m^c(x) P_n^c(x)\ d\mu_c(x), \qquad (m, n \in \mathbb{N}_0).$$

It follows then from (6.15), that

$$\gamma_{m+1,n} + c\gamma_{m,n} = \int_0^\infty x R_m^c(x) P_n^c(x)\ d\mu_c(x), \qquad (m, n \in \mathbb{N}_0),$$

and applying then (6.10),(6.11) and (6.13), we get that

$$\gamma_{m+1,n} + c\gamma_{m,n} = c^{m+1}(\delta_{m,n} + \delta_{m,n-1}), \qquad (m \in \mathbb{N}_0, n \in \mathbb{N}), \tag{6.16}$$

and

$$\gamma_{m+1,0} + c\gamma_{m,0} = c^{m+1}\delta_{m,0}, \qquad (m \in \mathbb{N}_0). \tag{6.17}$$

Since μ_c is a probability measure, $\gamma_{0,0} = 1$, and using this and induction on (6.17), it follows that $\gamma_{m,0} = 0$ for all m in $\mathbb{N}$. Thus

$$\gamma_{0,n} = \gamma_{n,0} = \begin{cases} 1, & \text{if } n = 0, \\ 0, & \text{if } n \geq 1. \end{cases} \tag{6.18}$$

Consider now a fixed n in $\mathbb{N}$. By (6.16), we have then that

$$\gamma_{m+1,n} + c\gamma_{m,n} = \begin{cases} 0, & \text{if } m \in \{0, 1, \dots, n-2\}, \\ c^n, & \text{if } m = n-1. \end{cases}$$

By induction in m $(0 \leq m \leq n)$, we get then, by application of (6.18), that

$$\gamma_{m,n} = \begin{cases} 0, & \text{if } m < n, \\ c^n, & \text{if } m = n, \end{cases}$$

and this completes the proof of (iii). □

6.2 Lemma. *For any non-negative integers p, q, put*

$$\nu(c, p, q) = c^{-q} \int_a^b x^p P_q^c(x) \, d\mu_c(x). \tag{6.19}$$

We then have
(i) *For any p in $\mathbb{N}_0$, $x^p = \sum_{q=0}^p \nu(c, p, q) P_q^c(x)$.*
(ii) *For any p, q in $\mathbb{N}_0$,*

$$\begin{aligned} \nu(c, p, q) &\geq 0, \quad \text{if } q \leq p, && (6.20) \\ \nu(c, p, p) &= 1, && (6.21) \\ \nu(c, p, q) &= 0, \quad \text{if } q > p. && (6.22) \end{aligned}$$

Proof. (i) Consider a fixed p from $\mathbb{N}_0$. By Proposition 6.1, $\operatorname{span}\{P_0^c, P_1^c, \dots, P_p^c\}$ is equal to the set of all polynomials of degree less than or equal to p. In particular we have that $x^p = \sum_{q=0}^p \gamma_q P_q^c(x)$, for suitable complex numbers $\gamma_0, \dots, \gamma_p$ (depending on c and p). Applying then the orthogonality relation in Proposition 6.1(iii), it follows that $\gamma_q = \nu(c, p, q)$ for all q in $\{0, 1, \dots, p\}$, and this proves (i).
(ii) By (6.1)-(6.3), it follows that

$$\begin{aligned} xP_0^c(x) &= P_1^c(x) + cP_0^c(x), && (6.23) \\ xP_q^c(x) &= P_{q+1}^c(x) + (c+1)P_q^c(x) + cP_{q-1}^c(x), \qquad (q \geq 1), && (6.24) \end{aligned}$$

so by induction in p, we get that $x^p (= x^p P_0^c(x))$, can be expressed as a linear combination of $P_0^c(x), P_1^c(x), \dots, P_p^c(x)$, in which all coefficients are non-negative. By (i) and the linear independence of $P_0^c(x), P_1^c(x), \dots, P_p^c(x)$, these

coefficients are exactly $\nu(c,p,0), \nu(c,p,1), \dots, \nu(c,p,p)$, and hence (6.20) follows.
Note next that (6.21) follows from (i) and the facts that $P_p^c(x)$ is a monic polynomial of degree p, whereas $P_0^c(x), \dots, P_{p-1}^c(x)$ are all of degree at most $p-1$.
Finally, (6.22) follows from (i) and the orthogonality relation in Proposition 6.1(iii). □

6.3 Lemma. *Let $\nu(c,p,q)$, $p,q \in \mathbb{N}_0$, be as in Lemma 6.2. Then for any fixed q in $\mathbb{N}_0$, the power series*

$$\sum_{p=0}^{\infty} \nu(c,p,q)t^p, \tag{6.25}$$

converges for all t in the open complex ball $B(0,\frac{1}{b})$, where $b = (\sqrt{c}+1)^2$. Moreover, the function

$$J_q^c(t) = \sum_{p=0}^{\infty} \nu(c,p,q)t^p, \qquad (t \in B(0,\tfrac{1}{b})),$$

is for all t in $B(0,\frac{1}{b}) \setminus \{0\}$, given by

$$J_q^c(t) = \frac{1-(c-1)t-\sqrt{(1-at)(1-bt)}}{2t}\left(\frac{1-(c+1)t-\sqrt{(1-at)(1-bt)}}{2ct}\right)^q, \tag{6.26}$$

where $\sqrt{\cdot}$ is the principal branch of the complex square-root.

Proof. Consider the Hilbert space $L^2(\mathbb{R},\mu_c)$, and let A be the bounded operator on $L^2(\mathbb{R},\mu_c)$, given by

$$[A(f)](x) = xf(x), \qquad (f \in L^2(\mathbb{R},\mu_c),\ x \in \mathbb{R}).$$

Note that $A^* = A$ and that $\mathrm{sp}(A) = \mathrm{supp}(\mu_c) \subseteq [0,b]$. Thus, letting $\mathbf{1}$ denote the identity operator on $L^2(\mathbb{R},\mu_c)$, $\mathbf{1}-tA$ is invertible for all complex numbers t such that $|t| < \frac{1}{b}$, and moreover, for such t,

$$(\mathbf{1}-tA)^{-1} = \sum_{p=0}^{\infty} t^p A^p, \qquad \text{(norm convergence)}.$$

For any t in $B(0,\frac{1}{b})$, we have thus that

$$\begin{aligned}\sum_{p=0}^{\infty} \nu(c,p,q)t^p &= c^{-q}\sum_{p=0}^{\infty}\langle x^p, P_q^c\rangle t^p = c^{-q}\sum_{p=0}^{\infty}\langle A^p P_0^c, P_q^c\rangle t^p \\ &= c^{-q}\langle (1-tA)^{-1}P_0^c, P_q^c\rangle.\end{aligned}$$

This shows that the series in (6.25) converges for all t in $B(0, \frac{1}{b})$, and moreover, that

$$J_q^c(t) = c^{-q}\langle (1-tA)^{-1}P_0^c, P_q^c\rangle, \qquad (t \in B(0, \tfrac{1}{b})). \tag{6.27}$$

To prove (6.26), we shall calculate the right hand side of (6.27). For this, consider for each z in $B(0, \frac{1}{\sqrt{c}})$ the series $\sum_{q=0}^{\infty} z^q P_q^c$, and note that by Lemma 6.1(iii), this series converges in $\|\cdot\|_2$-norm in $L^2(\mathbb{R}, \mu_c)$. We may thus define

$$\omega_z = \sum_{q=0}^{\infty} z^q P_q^c \in L^2(\mathbb{R}, \mu_c), \qquad (z \in B(0, \tfrac{1}{\sqrt{c}})). \tag{6.28}$$

With A as above, it follows now by (6.23) and (6.24), that for any z in $B(0, \frac{1}{\sqrt{c}}) \setminus \{0\}$,

$$\begin{aligned}
A\omega_z = \sum_{n=0}^{\infty} z^n A P_n^c &= cP_0^c + P_1^c + \sum_{n=1}^{\infty} z^n (cP_{n-1}^c + (c+1)P_n^c + P_{n+1}^c) \\
&= (c+cz)P_0^c + \sum_{n=1}^{\infty} (z^{n-1} + (c+1)z^n + cz^{n+1})P_n^c \\
&= (c+cz)P_0^c + z^{-1}(1+(c+1)z+cz^2) \sum_{n=1}^{\infty} z^n P_n^c \\
&= \big(c+cz - z^{-1}(1+(c+1)z+cz^2)\big)P_0^c + z^{-1}(1+(c+1)z+cz^2)\omega_z \\
&= -z^{-1}(1+z)P_0^c + z^{-1}(1+z)(1+cz)\omega_z,
\end{aligned}$$

where the infinite sums converge in $\|\cdot\|_2$-norm. From this it follows that

$$(z^{-1}(1+z)(1+cz)\mathbf{1} - A)\omega_z = z^{-1}(1+z)P_0^c, \qquad (z \in B(0, \tfrac{1}{\sqrt{c}}) \setminus \{0\}),$$

and hence that

$$\big(\mathbf{1} - \tfrac{z}{(1+z)(1+cz)}A\big)\omega_z = \tfrac{1}{1+cz}P_0^c, \qquad (z \in B(0, \tfrac{1}{\sqrt{c}}) \setminus \{-1, -\tfrac{1}{c}\}). \tag{6.29}$$

Define now

$$\varphi(z) = \frac{z}{(1+z)(1+cz)}, \qquad (z \in \mathbb{C} \setminus \{-1, \tfrac{1}{c}\}).$$

Since $\mathrm{sp}(A) \subseteq [0,b]$, it follows that $(\mathbf{1} - \varphi(z)A)$ is invertible whenever $\varphi(z) \notin [\frac{1}{b}, \infty[$, and in particular, as long as $|\varphi(z)| < \frac{1}{b}$. Note then, that φ is analytic on $\mathbb{C} \setminus \{-1, -\frac{1}{c}\}$, and that $\varphi(0) = 0$, $\varphi'(0) = 1$. It follows thus, that we may choose neighborhoods $\mathcal{U}$ and $\mathcal{V}$ of 0 in $\mathbb{C}$, such that φ is a bijection of $\mathcal{U}$ onto $\mathcal{V}$. We may assume, in addition, that

$$\mathcal{U} \subseteq B(0, \tfrac{1}{\sqrt{c}}) \setminus \{-1, -\tfrac{1}{c}\}, \qquad \text{and} \qquad \mathcal{V} \subseteq B(0, \tfrac{1}{b}).$$

For z in $\mathcal{U}$, it follows now from (6.29), that

$$\omega_z = \tfrac{1}{1+cz}\big(1-\varphi(z)A\big)^{-1}P_0^c,$$

and hence, by (6.27) and Lemma 6.1(iii),

$$J_q^c\big(\varphi(z)\big) = (1+cz)\cdot c^{-q}\langle \omega_z, P_q^c\rangle = (1+cz)z^q, \qquad (z\in\mathcal{U}). \tag{6.30}$$

It remains to invert φ. By solving the equation

$$t = \frac{z}{(1+z)(1+cz)},$$

w.r.t. z, we find that

$$\varphi^{-1}(t) = \frac{1-(c+1)t \pm \sqrt{(1-at)(1-bt)}}{2ct}, \qquad (t\in\mathcal{V}\setminus\{0\}),$$

where, as usual, $a=(\sqrt{c}-1)^2$ and $b=(\sqrt{c}+1)^2$. Since $\varphi^{-1}(t)\to 0$ as $t\to 0$, it follows that for some neighbourhood $\mathcal{V}_0$ of 0, such that $\mathcal{V}_0\subseteq\mathcal{V}$, we must have

$$\varphi^{-1}(t) = \frac{1-(c+1)t - \sqrt{(1-at)(1-bt)}}{2ct}, \qquad (t\in\mathcal{V}_0\setminus\{0\}), \tag{6.31}$$

where $\sqrt{\cdot}$ is the principal part of the square root. Hence, we have also that

$$1+c\varphi^{-1}(t) = \frac{1-(c-1)t - \sqrt{(1-at)(1-bt)}}{2t}, \qquad (t\in\mathcal{V}_0\setminus\{0\}). \tag{6.32}$$

Inserting (6.31) and (6.32) in (6.30), we obtain that (6.26) holds for all t in $\mathcal{V}_0\setminus\{0\}$.
To show that (6.26) actually holds for all t in $B(0,\frac{1}{b})\setminus\{0\}$, note that for all such t, $\mathrm{Re}(1-at)>0$ and $\mathrm{Re}(1-bt)>0$, so that $(1-at)(1-bt)\in\mathbb{C}\setminus]-\infty,0]$. Hence, with $\sqrt{\cdot}$ the principal branch of the square root, $t\mapsto\sqrt{(1-at)(1-bt)}$ is an analytic function of $t\in B(0,\frac{1}{b})$. By uniqueness of analytic continuation, it follows thus, that (6.26) holds for all t in $B(0,\frac{1}{b})\setminus\{0\}$. □

6.4 Lemma. *Let $g_c(p)$ and $h_c(p)$, $p\in\mathbb{N}_0$, be as in Definition 5.17. Then the power series*

$$G_c(t) = \sum_{p=0}^{\infty} g_c(p)t^p, \tag{6.33}$$

and

$$H_c(t) = \sum_{p=0}^{\infty} h_c(p)t^p, \tag{6.34}$$

are convergent for all t in $B(0,\frac{1}{b})$, and

$$J_q^c(t) = t^q G_c(t)^{q+1} H_c(t)^q, \qquad (t\in B(0,\tfrac{1}{b})). \tag{6.35}$$

Proof. By (5.10), we have

$$g_c(p) = \int_0^\infty x^p \; d\mu_c(x), \qquad (p \in \mathbb{N}),$$

and since $g_c(0) = 1$, the same formula holds for $p = 0$. Hence $g_c(p) = \nu(c, p, 0)$, for all p in $\mathbb{N}_0$, so by Lemma 6.3, the series in (6.33) converges for all t in $B(0, \frac{1}{b})$, and

$$G_c(t) = J_0^c(t) = \frac{1 - (c-1)t - \sqrt{(1-at)(1-bt)}}{2t}, \qquad (t \in B(0, \tfrac{1}{b}) \setminus \{0\}). \tag{6.36}$$

Since $h_c(0) = 1$ and since $h_c(p) = \frac{1}{c} g_c(p)$, for all p in $\mathbb{N}$, the series in (6.34) is also convergent for all t in $B(0, \frac{1}{b})$, and

$$H_c(t) = 1 + \tfrac{1}{c}(G_c(t) - 1), \qquad (t \in B(0, \tfrac{1}{c})).$$

Hence by (6.34)

$$H_c(t) = \frac{1 + (c-1)t - \sqrt{(1-at)(1-bt)}}{2ct}, \qquad (t \in B(0, \tfrac{1}{b}) \setminus \{0\}). \tag{6.37}$$

By (6.36) and (6.37), we get now for all t in $B(0, \frac{1}{b}) \setminus \{0\}$,

$$\begin{aligned}
&G_c(t)H_c(t) \\
&= \frac{\big(1 - \sqrt{(1-at)(1-bt)}\big)^2 - (c-1)^2 t^2}{4ct^2} \\
&= \frac{1 + (1-at)(1-bt) - 2\sqrt{(1-at)(1-bt)} - (c-1)^2 t^2}{4ct^2} \\
&= \frac{1 + (1 - 2(c+1)t + (c-1)^2 t^2) - 2\sqrt{(1-at)(1-bt)} - (c-1)^2 t^2}{4ct^2} \\
&= \frac{1 - (c+1)t - \sqrt{(1-at)(1-bt)}}{2ct^2}.
\end{aligned}$$

Combining this with (6.36) and (6.26), it follows that

$$J_c^q(t) = G_c(t)\big(tG_c(t)H_c(t)\big)^q, \qquad (t \in B(0, \tfrac{1}{b})),$$

and the same formula holds trivially for $t = 0$, by (6.22). This proves (6.35).
□

6.5 Lemma. *For all p, q in $\mathbb{N}_0$ such that $p \geq q$, let $\nu(c, p, q)$ be as introduced in Definition 5.17. Then*

$$\nu'(c, p, q) = \nu(c, p, q), \qquad (p, q \in \mathbb{N}_0, \; q \leq p).$$

Proof. Recall from Definition 5.17, that for p, q in $\mathbb{N}_0$, such that $p \geq q$, we have

$$\nu'(c,p,q) = \sum_{\substack{r_0,r_1,\dots,r_{2q}\geq 0\\ r_0+r_1+\cdots+r_{2q}=p-q}} g_c(r_0)h_c(r_1)g_c(r_2)h_c(r_3)\cdots g_c(r_{2q}).$$

Hence $\nu'(c,p,q)$ is the coefficient to t^{p-q} in the power series for

$$G_c(t)H_c(t)G_c(t)H_c(t)\cdots G_c(t), \qquad (2q+1 \text{ factors}),$$

and therefore $\nu'(c,p,q)$ is the coefficient to t^p in the power series for $t^q G_c(t)^{q+1}H_c(t)^q$. Thus, by Lemma 6.3 and Lemma 6.4, it follows that

$$\nu'(c,p,q) = \nu(c,p,q), \quad \text{for all } p,q \text{ in } \mathbb{N}_0, \text{ such that } p \geq q. \qquad \square$$

6.6 THEOREM. *Let $\mathcal{H}$, $\mathcal{K}$ be Hilbert spaces, and let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$, satisfying that $\sum_{i=1}^r a_i^*a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$ and $\sum_{i=1}^r a_ia_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{K})}$, for some positive real number c. Furthermore, let $Y_1,\dots,Y_r$ be independent elements of* $\mathrm{GRM}(n,n,\frac{1}{n})$, *and put $S = \sum_{i=1}^r a_i \otimes Y_i$. Then for any q in $\mathbb{N}$,*

$$\mathbb{E}\big[P_q^c(S^*S)\big] = \Big[\sum_{\rho\in S_q^{\mathrm{irr}}} n^{-2\sigma(\hat\rho)}\Big(\sum_{1\leq i_1,\dots,i_q\leq r} a_{i_1}^*a_{i_{\rho(1)}}\cdots a_{i_q}^*a_{i_{\rho(q)}}\Big)\Big]\otimes \mathbf{1}_n.$$

Proof. For each q in $\mathbb{N}$, put

$$T_q = \sum_{\rho\in S_q^{\mathrm{irr}}} n^{-2\sigma(\hat\rho)}\Big(\sum_{1\leq i_1,\dots,i_q\leq r} a_{i_1}^*a_{i_{\rho(1)}}\cdots a_{i_q}^*a_{i_{\rho(q)}}\Big),$$

and put $T_0 = \mathbf{1}_{\mathcal{B}(\mathcal{H})}$. By Theorem 5.18 and Lemma 6.5, it follows then that

$$\mathbb{E}\big[(S^*S)^p\big] = \sum_{q=0}^p \nu(c,p,q)\cdot T_q\otimes\mathbf{1}_n, \qquad (p\in\mathbb{N}_0). \tag{6.38}$$

On the other hand, it follows from Lemma 6.2(i), that

$$\mathbb{E}\big[(S^*S)^p\big] = \sum_{q=0}^p \nu(c,p,q)\mathbb{E}\big[P_q^c(S^*S)\big], \qquad (p\in\mathbb{N}_0). \tag{6.39}$$

We prove that

$$\mathbb{E}\big[P_q^c(S^*S)\big] = T_q\otimes\mathbf{1}_n, \qquad (q\in\mathbb{N}_0), \tag{6.40}$$

by induction in q. Note that (6.40) is trivial for $q = 0$. Consider then p from $\mathbb{N}$, and assume that (6.40) has been proved for $q = 0,1,\dots,p-1$. Since

$\nu(c,p,p) = 1$, by Lemma 6.2(ii), it follows then from (6.39) and (6.38), that

$$\begin{aligned}\mathbb{E}\big[P_p^c(S^*S)\big] &= \mathbb{E}\big[(S^*S)^p\big] - \sum_{q=0}^{p-1}\nu(c,p,q)\mathbb{E}\big[P_q^c(S^*S)\big] \\ &= \mathbb{E}\big[(S^*S)^p\big] - \sum_{q=0}^{p-1}\nu(c,p,q)\cdot T_q\otimes \mathbf{1}_n \\ &= T_p\otimes\mathbf{1}_n.\end{aligned}$$

Thus, (6.40) holds for $q = p$, and this completes the proof. □

6.7 Example. By (6.1)-(6.3), it follows that

$$P_1^c(x) = x - c, \tag{6.41}$$
$$P_2^c(x) = x^2 - (2c+1)x + c^2, \tag{6.42}$$
$$P_3^c(x) = x^3 - (3c+2)x^2 + (3c^2+2c+1)x - c^3. \tag{6.43}$$

By Example 5.20, $S_p^{\mathrm{irr}} = \emptyset$ if $p \in \{1,2\}$, and $S_3^{\mathrm{irr}} = \{\pi\}$, where π is the permutation given by $\pi(1) = 3, \pi(2) = 1, \pi(3) = 2$, so that $\sigma(\hat{\pi}) = 1$. It follows thus by Theorem 6.6, that

$$\begin{aligned}\mathbb{E}\big[P_1^c(S^*S)\big] &= 0, \\ \mathbb{E}\big[P_2^c(S^*S)\big] &= 0, \\ \mathbb{E}\big[P_3^c(S^*S)\big] &= n^{-2}\sum_{i,j,k=1}^{r} a_i^*a_ka_j^*a_ia_k^*a_j.\end{aligned}$$

These three formulas can also easily be derived directly from Example 5.20, using the formulas (6.41)-(6.43). □

7 An Upper Bound for $\mathbb{E}\big[\exp(-tS^*S)\big],\ t \geq 0$

Throughout this section, we consider elements $a_1,\dots,a_r$ of $\mathcal{B}(\mathcal{H},\mathcal{K})$ (for given Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$), satisfying that

$$\sum_{i=1}^{r} a_i^*a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})} \quad \text{and} \quad \sum_{i=1}^{r} a_ia_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{K})},$$

for some constant c in $[1,\infty[$. Moreover, we consider independent elements $Y_1,\dots,Y_r$ of $\mathrm{GRM}(n,n,\frac{1}{n})$, and put

$$S = \sum_{i=1}^{r} a_i\otimes Y_i.$$

As in Section 3, we let μ_c denote the probability measure on $\mathbb{R}$, given by

$$\mu_c = \frac{\sqrt{(x-a)(b-x)}}{2\pi x} \cdot 1_{[a,b]}(x) \cdot dx,$$

where $a = (\sqrt{c}-1)^2$ and $b = (\sqrt{c}+1)^2$. Furthermore, we let $(P_q^c)_{q\in\mathbb{N}_0}$ be the sequence of monic orthogonal polynomials w.r.t. μ_c as defined in Section 6. In particular $P_0^c \equiv 1$.

7.1 Lemma. *Let, as above, $a = (\sqrt{c}-1)^2$ and $b = (\sqrt{c}+1)^2$. Then for any q in $\mathbb{N}_0$,*
(i) $P_q^c(x) \geq P_q^c(b) > 0$, *for all x in $]b,\infty[$.*
(ii) $|P_q^c(x)| \leq P_q^c(b)$, *for all x in $[a,b]$.*
(iii) $|P_q^c(x)| \leq P_q^c(2c+2-x)$, *for all x in $]-\infty, a[$.*

Proof. We start by proving (ii). If $x \in [a,b]$, then $x = c+1+2\sqrt{c}\cos\theta$, for some θ in $[0,\pi]$. For θ in $]0,\pi[$, we have from Proposition 6.1(ii), that

$$P_q^c(c+1+2\sqrt{c}\cos\theta) = \frac{c^{\frac{q}{2}}\sin((q+1)\theta) + c^{\frac{q-1}{2}}\sin(q\theta)}{\sin\theta}. \tag{7.1}$$

Note here that for any k in $\mathbb{N}_0$,

$$\frac{\sin((k+1)\theta)}{\sin\theta} = e^{-k\theta}\big(1+e^{2i\theta}+e^{4i\theta}+\cdots+e^{2ki\theta}\big), \tag{7.2}$$

so that $\left|\frac{\sin((k+1)\theta)}{\sin\theta}\right| \leq k+1$. It follows thus that

$$|P_q^c(x)| \leq c^{\frac{q}{2}}(q+1) + c^{\frac{q-1}{2}}q, \qquad (x \in]a,b[), \tag{7.3}$$

and by continuity, (7.3) holds also for $x=a$ and $x=b$. By (7.2), $\lim_{\theta\to 0}\frac{\sin((k+1)\theta)}{\sin\theta} = k+1$, for any k in $\mathbb{N}_0$, and hence the right hand side of (7.3) is equal to $P_q^c(b)$. This proves (ii).
To prove (i), we note first, that by uniqueness of analytic continuation, (7.1) actually holds for all θ in $\mathbb{C}\setminus\pi\mathbb{Z}$. If we put $\theta = i\rho$, $\rho > 0$, we get the equation:

$$P_q^c(c+1+2\sqrt{c}\cosh\rho) = \frac{c^{\frac{q}{2}}\sinh((q+1)\rho) + c^{\frac{q-1}{2}}\sinh(q\rho)}{\sinh\rho}, \qquad (\rho \in]0,\infty[), \tag{7.4}$$

which covers the values of $P_q(x)$ for all x in $]b,\infty[$. Note here that for any k in $\mathbb{N}_0$,

$$\frac{\sinh((k+1)\rho)}{\sinh\rho} = e^{-k\rho}\big(1+e^{2\rho}+e^{4\rho}+\cdots+e^{2k\rho}\big),$$

and hence, if k is even,

$$\frac{\sinh((k+1)\rho)}{\sinh\rho} = 1+2\cosh(2\rho)+2\cosh(4\rho)+\cdots+2\cosh(k\rho),$$

whereas, if k is odd,

$$\frac{\sinh((k+1)\rho)}{\sinh\rho} = 2\cosh(\rho) + 2\cosh(3\rho) + \cdots + 2\cosh(k\rho),$$

so in both cases $\frac{\sinh((k+1)\rho)}{\sin\rho}$ is an increasing function of $\rho > 0$. It follows thus from (7.4), that $P_q^c(x) \geq P_q^c(b)$ for all x in $]b,\infty[$. Moreover, as we saw in the proof of (ii), $P_q^c(b) > 0$. This concludes the proof of (i).
Finally, to prove (iii), we put $\theta = \pi + i\rho$ in (7.1), and get for ρ in $]0,\infty[$, that

$$\begin{aligned}\left|P_q^c(c+1-2\sqrt{c}\cosh\rho)\right| &= \left|\frac{(-1)^q c^{\frac{q}{2}}\sinh((q+1)\rho) + (-1)^{q-1}c^{\frac{q-1}{2}}\sinh(q\rho)}{\sinh\rho}\right| \\ &\leq \frac{c^{\frac{q}{2}}\sinh((q+1)\rho) + c^{\frac{q-1}{2}}\sinh(q\rho)}{\sinh\rho} \\ &= P_q^c(c+1+2\sqrt{c}\cosh\rho).\end{aligned}$$

This proves (iii). □

7.2 Definition. For each q in $\mathbb{N}_0$, we define the function $\psi_q^c\colon \mathbb{R}\to\mathbb{R}$, by the equation

$$\psi_q^c(t) = c^{-q}\int_a^b \exp(tx)P_q^c(x)\ d\mu_c(x), \qquad (t\in\mathbb{R}).$$ □

7.3 Lemma. *Consider the sequence $(\psi_q^c)_{q\in\mathbb{N}_0}$ of functions, introduced in Definition 7.2, and for each p in $\mathbb{N}_0$, let, as in Section 6,*

$$\nu(c,p,q) = c^{-q}\int_a^b x^p P_q^c(x)\ d\mu_c(x), \qquad (p,q\in\mathbb{N}_0).$$

We then have

(i) $\psi_q^c(t) = \sum_{p=q}^\infty \frac{t^p}{p!}\nu(c,p,q)$, *for all t in $\mathbb{R}$.*

(ii) $\sum_{q=0}^\infty |\psi_q^c(t)|\cdot|P_q^c(x)| \leq \exp(|t|x) + \exp(|t|(2c+2))$, *for all t in $\mathbb{R}$ and all x in $[0,\infty[$.*

(iii) $\exp(tx) = \sum_{q=0}^\infty \psi_q^c(t)\cdot P_q^c(x)$, *for all t in $\mathbb{R}$ and x in $[0,\infty[$, and for fixed t in $\mathbb{R}$, the series converges uniformly in x on compact subsets of $[0,\infty[$.*

Proof. (i) By Lemma 6.2(ii), $\nu(c,p,q) = 0$ whenever $q > p$. Hence (i) follows from the power series expansion of $\exp(tx)$.
(ii) Let $\beta\colon \mathbb{R}\to[b,\infty[$ be the *continuous* function defined by:

$$\beta(x) = \begin{cases} x, & \text{if } x > b,\\ b, & \text{if } a\leq x\leq b,\\ 2c+2-x, & \text{if } x < a,\end{cases}$$

It follows then from Lemma 7.1, that

$$|P_q^c(x)| \le P_q^c(\beta(x)), \qquad (x \in \mathbb{R},\ q \in \mathbb{N}_0). \tag{7.5}$$

Recall that $x^p = \sum_{q=0}^{p} \nu(c,p,q) P_q^c(x)$, for all p in $\mathbb{N}$ (c.f. Lemma 6.2(i)). Hence, for x, t in $\mathbb{R}$, we have that

$$\exp(tx) = \sum_{p=0}^{\infty} \tfrac{t^p}{p!} x^p = \sum_{p=0}^{\infty} \tfrac{t^p}{p!} \Big(\sum_{q=0}^{p} \nu(c,p,q) P_q^c(x) \Big). \tag{7.6}$$

Substituting x with $\beta(x)$ and t with $|t|$ in this formula, and recalling from Lemma 6.2(ii), that $\nu(c,p,q) \ge 0$, for $0 \le q \le p$, we get by application of (7.5),

$$\begin{aligned} \sum_{p=0}^{\infty} \tfrac{|t|^p}{p!} \Big(\sum_{q=0}^{p} \nu(c,p,q) |P_q^c(x)| \Big) &\le \sum_{p=0}^{\infty} \tfrac{|t|^p}{p!} \Big(\sum_{q=0}^{p} \nu(c,p,q) P_q^c(\beta(x)) \Big) \\ &= \exp(|t|\beta(x)) < \infty. \end{aligned}$$

Hence, we can apply Fubini's theorem to the double sum in (7.6), and obtain that

$$\exp(tx) = \sum_{q=0}^{\infty} \Big(\sum_{p=q}^{\infty} \tfrac{t^p}{p!} \nu(c,p,q) \Big) P_q^c(x), \qquad (x, t \in \mathbb{R}). \tag{7.7}$$

Similarly we have that

$$\exp(|t|\beta(x)) = \sum_{q=0}^{\infty} \Big(\sum_{p=q}^{\infty} \tfrac{|t|^p}{p!} \nu(c,p,q) \Big) P_q^c(\beta(x)), \qquad (x, t \in \mathbb{R}). \tag{7.8}$$

Note here that by (i) proved above, we have that,

$$|\psi_q^c(t)| \le \sum_{p=q}^{\infty} \tfrac{|t|^p}{p!} \nu(c,p,q). \tag{7.9}$$

Since $\beta(x) \le \max\{2c+2, x\}$ for all x in $[0,\infty[$, (7.5) and (7.7)-(7.9) imply that for all t in $\mathbb{R}$ and x in $[0,\infty[$,

$$\sum_{q=0}^{\infty} |\psi_q^c(t)| \cdot |P_q^c(x)| \le \exp(|t|\beta(x)) \le \exp(|t|(2c+2)) + \exp(|t|x),$$

and this proves (ii).
(iii) The summation formula in (iii) follows from (i) and (7.7). To prove that the convergence is uniform in x on compact subsets, we observe that for any

Q in $\mathbb{N}$,

$$\begin{aligned}\Big|\exp(tx) - \sum_{q=0}^{Q}\psi_q^c(t)P_q^c(x)\Big| &\le \sum_{q=Q+1}^{\infty}|\psi_q^c(t)|\cdot|P_q^c(x)| \\ &\le \sum_{q=Q+1}^{\infty}\Big(\sum_{p=q}^{\infty}\tfrac{|t|^p}{p!}\nu(c,p,q)P_q^c(\beta(x))\Big) \\ &\le \sum_{p=Q+1}^{\infty}\tfrac{|t|^p}{p!}\Big(\sum_{q=0}^{p}\nu(c,p,q)P_q^c(\beta(x))\Big) \\ &= \sum_{p=Q+1}^{\infty}\tfrac{(|t|\beta(x))^p}{p!}.\end{aligned} \tag{7.10}$$

Since β is continuous, and hence bounded on compact sets, it follows readily from (7.10) that for fixed t in $\mathbb{R}$, the series in (iii) converges uniformly in x on compact subsets of $[0,\infty[$. □

7.4 Proposition. *Consider the sequence $(\psi_q^c)_{q\in\mathbb{N}_0}$ of functions, introduced in Definition 7.2. Then for any t in $\mathbb{R}$ such that $|t| < \frac{n}{c}$, the function $\omega \mapsto \exp(tS^*(\omega)S(\omega))$ is integrable in the sense of Definition 3.1, and*

$$\mathbb{E}\big[\exp(tS^*S)\big] = \sum_{q=0}^{\infty}\psi_q^c(t)\mathbb{E}\big[P_q^c(S^*S)\big], \tag{7.11}$$

where the sum on the right hand side is absolutely convergent in $\mathcal{B}(\mathcal{H}^n)$.

Proof. We start by proving that the right hand side of (7.11) is absolutely convergent in $\mathcal{B}(\mathcal{H}^n)$. Since $|\psi_q^c(t)| \le \psi_q^c(|t|)$ by Lemma 7.3(i) and (7.9), it suffices to consider the case where $t \ge 0$.
By Lemma 7.3(i), we have for any t in $[0,\infty[$,

$$\sum_{q=0}^{\infty}\psi_q^c(t)\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| = \sum_{p=0}^{\infty}\frac{t^p}{p!}\Big(\sum_{q=0}^{p}\nu(c,p,q)\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\|\Big). \tag{7.12}$$

Note here, that by Theorem 6.6,

$$\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| \le \sum_{\rho\in S_q^{\mathrm{irr}}} n^{-2\sigma(\hat\rho)}\Big\|\sum_{1\le i_1,\dots,i_q\le r} a_{i_1}^* a_{i_{\rho(1)}}\cdots a_{i_q}^* a_{i_{\rho(q)}}\Big\|,$$

for any q in $\mathbb{N}$, whereas

$$\big\|\mathbb{E}\big[P_0^c(S^*S)\big]\big\| = \|\mathbb{E}(\mathbf{1}_{\mathcal{B}(\mathcal{H}^n)})\| = 1.$$

Hence, by Proposition 5.19, Lemma 6.5 and Proposition 2.7, we have for any p in $\mathbb{N}$,

$$\begin{aligned}\sum_{q=0}^{p}\nu(c,p,q)\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| &\le \sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}\Big\|\sum_{1\le i_1,\dots,i_p\le r} a_{i_1}^*a_{i_{\pi(1)}}\cdots a_{i_p}^*a_{i_{\pi(p)}}\Big\| \\ &\le \sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}c^{\kappa(\hat{\pi})}.\end{aligned} \tag{7.13}$$

Using now that $c\ge 1$, and that $\kappa(\hat{\pi})\le k(\hat{\pi})+2\sigma(\hat{\pi})$ (c.f. Proposition 2.10), it follows that for any p in $\mathbb{N}$,

$$\sum_{\pi\in S_p} n^{-2\sigma(\hat{\pi})}c^{\kappa(\hat{\pi})} \le \sum_{\pi\in S_p}\big(\tfrac{n}{c}\big)^{-2\sigma(\hat{\pi})}c^{k(\hat{\pi})}. \tag{7.14}$$

For $p=0$, we note that

$$\sum_{q=0}^{p}\nu(c,p,q)\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| = 1. \tag{7.15}$$

Combining now (7.12)-(7.15), we get that

$$\sum_{q=0}^{\infty}\psi_q^c(t)\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| \le 1+\sum_{p=1}^{\infty}\tfrac{t^p}{p!}\Big(\sum_{\pi\in S_p}\big(\tfrac{n}{c}\big)^{-2\sigma(\hat{\pi})}c^{k(\hat{\pi})}\Big). \tag{7.16}$$

Using then that $-2\sigma(\hat{\pi})=k(\hat{\pi})+l(\hat{\pi})-p-1$, it follows that

$$\begin{aligned}\sum_{q=0}^{\infty}\psi_q^c(t)\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| &\le 1+\sum_{p=1}^{\infty}\tfrac{1}{p!}\big(\tfrac{ct}{n}\big)^p\sum_{\pi\in S_p} n^{k(\hat{\pi})}\big(\tfrac{n}{c}\big)^{l(\hat{\pi})-1} \\ &\le 1+ct\sum_{p=1}^{\infty}\tfrac{1}{(p-1)!}\big(\tfrac{ct}{n}\big)^{p-1}\sum_{\pi\in S_p} n^{k(\hat{\pi})-1}\big(\tfrac{n}{c}\big)^{l(\hat{\pi})-1},\end{aligned} \tag{7.17}$$

where the last equality follows by noting that $\frac{1}{p!}\le\frac{1}{(p-1)!}$ for all p in $\mathbb{N}$. By Lemma 3.4, the last quantity in (7.17) is finite whenever $0\le\frac{ct}{n}<1$, and this shows that the right hand side of (7.11) is absolutely convergent for all t in $]-\frac{n}{c},\frac{n}{c}[$, as desired.
It remains now (cf. Definition 3.1) to show, that for any state φ on $\mathcal{B}(\mathcal{H}^n)$,

$$\mathbb{E}\big[\varphi(\exp(tS^*S))\big]=\sum_{q=0}^{\infty}\psi_q^c(t)\varphi\big(\mathbb{E}\big[P_q^c(S^*S)\big]\big), \qquad (t\in]-\tfrac{n}{c},\tfrac{n}{c}[). \tag{7.18}$$

So consider a fixed t from $]-\frac{n}{c},\frac{n}{c}[$ and a fixed state φ on $\mathcal{B}(\mathcal{H}^n)$. Since the spectrum of $S^*(\omega)S(\omega)$ is compact for each ω in Ω, it follows then by Lemma 7.3, that

$$\varphi\big[\exp(tS^*(\omega)S(\omega))\big] = \sum_{p=0}^{\infty}\psi_q^c(t)\varphi\big[P_q^c(S_n(\omega)^*S_n(\omega))\big], \tag{7.19}$$

so we need to show that we can integrate termwise in the sum on the right hand side. Note for this, that by Lemma 7.3(ii), and the function calculus for selfadjoint operators on Hilbert spaces,

$$\sum_{p=0}^{\infty}|\psi_q^c(t)|\cdot|P_q^c(S(\omega)^*S(\omega))| \le \exp(2(c+1)|t|)\mathbf{1}_{\mathcal{B}(\mathcal{H}^n)} + \exp(|t|S(\omega)^*S(\omega)), \tag{7.20}$$

where $|T| = (T^2)^{\frac{1}{2}}$, for any selfadjoint T in $\mathcal{B}(\mathcal{H}^n)$. For such T, we have also that $|\varphi(T)| \le \varphi(|T|)$, and hence it follows from (7.20), that

$$\sum_{p=0}^{\infty}|\psi_q^c(t)|\cdot\big|\varphi\big[P_q^c(S(\omega)^*S(\omega))\big]\big| \le \exp(2(c+1)|t|) + \varphi\big[\exp(|t|S(\omega)^*S(\omega))\big]. \tag{7.21}$$

Since $\mathbb{E}\big[\varphi(\exp(|t|S^*S))\big] < \infty$, by Proposition 3.2, it follows from (7.21) and Lebesgue's theorem on dominated convergence, that we may integrate termwise in (7.19), and hence obtain (7.18). This concludes the proof. $\square$

In order to obtain the upper bound for $\mathbb{E}\big[\exp(-tS^*S)\big]$ in Theorem 7.8 below, we need more precise information about the behavior of the function $\psi_q^c(t)$ for $t < 0$.

7.5 Proposition. *Consider the sequence $(\psi_q^c)_{q\in\mathbb{N}_0}$ of functions, defined in Definition 7.2. Then for any q in $\mathbb{N}_0$, and any t in $]0,\infty[$, we have that*
(i) $\psi_q^c(t) > 0$.
(ii) $(-1)^q\psi_q^c(-t) > 0$.
(iii) $|\psi_q^c(-t)| \le \frac{\psi_0^c(-t)}{\psi_0^c(t)}\psi_q^c(t)$.

Proof. (i) This follows from Lemma 7.3(i), but for completeness we include a different proof, which will also be needed in the proof of (ii) and (iii). For each q in $\mathbb{N}_0$, we put

$$\rho_q^c(x) = c^{-\frac{q}{2}}P_q^c(x), \qquad (x\in\mathbb{R}).$$

Then by Proposition 6.1, $(\rho_q^c)_{q\in\mathbb{N}_0}$ is an orthonormal basis for $L_2([a,b],\mu_c)$. Let A be the (bounded) operator for multiplication by x in $L_2([a,b],\mu_c)$. Then by

(6.23) and (6.24), the matrix $M(A)$ of A w.r.t. $(\rho_q^c)_{q\in\mathbb{N}_0}$, is given by

$$M(A) = \begin{pmatrix} c & \sqrt{c} & & & 0 \\ \sqrt{c} & c+1 & \sqrt{c} & & \\ & \sqrt{c} & c+1 & \sqrt{c} & \\ & & \ddots & \ddots & \ddots \\ & & & & \\ 0 & & & & \end{pmatrix} \tag{7.22}$$

From this, it follows, that for any p in $\mathbb{N}$,

$$\begin{aligned} M(A^p)_{jk} &> 0, \quad \text{when } |j-k| \le p, \\ M(A^p)_{jk} &= 0, \quad \text{when } |j-k| > p. \end{aligned}$$

Hence, for any t in $[0, \infty[$,

$$M(\exp(tA))_{jk} = \delta_{j,k} + \sum_{p=1}^{\infty} \tfrac{t^p}{p!} M(A^p)_{jk} > 0, \qquad (j,k \in \mathbb{N}_0).$$

Since $\exp(tA)$ is the operator for multiplication by $\exp(tx)$ in $L_2([a,b],\mu_c)$, and since $P_0^c(x) \equiv 1$, we get that

$$\begin{aligned} \psi_q^c(t) &= c^{-q} \int_a^b \exp(tx) P_q^c(x) P_0^c(x) \; d\mu_c(x) = c^{-\frac{q}{2}} \langle \exp(tA)\rho_q^c, \rho_0^c \rangle \\ &= c^{-\frac{q}{2}} M(\exp(tA))_{0,q} > 0, \end{aligned} \tag{7.23}$$

and this proves (i).
(ii) To prove (ii), we consider the operator

$$B = A + 2P_0,$$

where P_0 is the projection onto $\mathbb{C}\rho_0^c$ in $\mathcal{B}\big(L_2([a,b],\mu_c)\big)$. Then

$$M(B) = \begin{pmatrix} c+2 & \sqrt{c} & & & 0 \\ \sqrt{c} & c+1 & \sqrt{c} & & \\ & \sqrt{c} & c+1 & \sqrt{c} & \\ & & \ddots & \ddots & \ddots \\ & & & & \\ 0 & & & & \end{pmatrix}, \tag{7.24}$$

so as above, we get that

$$M(\exp(tB))_{jk} > 0, \quad \text{for all } j,k \text{ in } \mathbb{N}_0. \tag{7.25}$$

Let U be the unitary operator on $L_2([a,b],\mu_c)$, defined by the equation:

$$U\rho_q^c = (-1)^q \rho_q^c, \qquad (q \in \mathbb{N}_0).$$

Then

$$M(UBU^*) = \begin{pmatrix} c+2 & -\sqrt{c} & & & 0 \\ -\sqrt{c} & c+1 & -\sqrt{c} & & \\ & -\sqrt{c} & c+1 & -\sqrt{c} & \\ & & \ddots & \ddots & \ddots \\ & & & & \\ 0 & & & & \end{pmatrix} = M(2(c+1)\mathbf{1} - A).$$

Hence $A = 2(c+1)\mathbf{1} - UBU^*$, and for t in $[0, \infty[$, we have thus that

$$\exp(-tA) = \exp(-2(c+1)t)\exp(tUBU^*) = \exp(-2(c+1)t)U\exp(tB)U^*.$$

Therefore,

$$M(\exp(-tA))_{jk} = (-1)^{j+k}\exp(-2(c+1)t)M(\exp(tB))_{jk}, \qquad (j, k \in \mathbb{N}_0), \tag{7.26}$$

so in particular, by (7.25),

$$(-1)^{j+k}M(\exp(-tA))_{jk} > 0, \qquad (j, k \in \mathbb{N}_0).$$

For t in $[0, \infty[$, we note here that

$$\psi_q^c(-t) = c^{-q}\int_a^b \exp(-tx)P_q^c(x)P_0^c(x)\, d\mu_c(x) = c^{-\frac{q}{2}}M(\exp(-tA))_{q0}, \tag{7.27}$$

and hence it follows that $(-1)^q\psi_q(-t) > 0$, which proves (ii).
To prove (iii), we need the following technical lemma:

7.6 Lemma. *Let C and D be bounded positive selfadjoint operators on $\ell_2(\mathbb{N}_0)$, and assume that the corresponding matrices $(c_{jk})_{j,k\in\mathbb{N}_0}$ and $(d_{jk})_{j,k\in\mathbb{N}_0}$ satisfy the following conditions:*

(a) $c_{jk} \geq 0$ *for all* j, k *in* $\mathbb{N}_0$.
(b) $c_{jk} = 0$ *when* $|j - k| \geq 2$.
(c) $d_{jk} = c_{jk}$, *when* $(j, k) \neq (0, 0)$.
(d) $d_{00} \geq c_{00}$.

For φ, ψ in $\ell_2(\mathbb{N}_0)$, we define

$$[\varphi, \psi]_{j,k} = \varphi(j)\psi(k) - \varphi(k)\psi(j), \qquad (j, k \in \mathbb{N}_0).$$

Consider then furthermore f, g from $\ell_2(\mathbb{N}_0)$, satisfying that

(e) $f(k) \geq 0$ *and* $g(k) \geq 0$ *for all* k *in* $\mathbb{N}_0$.
(f) $[f, g]_{j,k} \geq 0$, *for all* k, j *in* $\mathbb{N}_0$ *such that* $k > j$.

Then for all j, k in $\mathbb{N}_0$, such that $k > j$, we have that

(i) $[Cf, Cg]_{j,k} \geq 0$.
(ii) $[Df, Cg]_{j,k} \geq 0$.
(iii) $[D^n f, C^n g]_{j,k} \geq 0$, *for all* n *in* $\mathbb{N}$.
(iv) $[\exp(tD)f, \exp(tC)g]_{j,k} \geq 0$, *for all* t *in* $[0, \infty[$.

7.7 Remark. If φ, ψ are strictly positive functions in $\ell_2(\mathbb{N}_0)$, then the statement

$$[\varphi, \psi]_{j,k} \geq 0, \quad \text{for all } j,k \text{ in } \mathbb{N}_0, \text{ such that } k > j,$$

is equivalent to the condition that

$$\frac{\varphi(0)}{\psi(0)} \geq \frac{\varphi(1)}{\psi(1)} \geq \frac{\varphi(2)}{\psi(2)} \geq \cdots . \qquad \square$$

Proof of Lemma 7.6. Note first that for any φ, ψ in $\ell_2(\mathbb{N}_0)$ and j,k in $\mathbb{N}_0$, we have that $[\varphi, \psi]_{j,k} = -[\varphi, \psi]_{k,j}$. In particular,

$$[\varphi, \psi]_{j,j} = 0, \qquad (\varphi, \psi \in \ell_2(\mathbb{N}_0),\ j \in \mathbb{N}_0). \tag{7.28}$$

Note also that the positivity of C implies that

$$\det \begin{pmatrix} c_{jj} & c_{jk} \\ c_{kj} & c_{kk} \end{pmatrix} \geq 0, \qquad \text{for all } j,k \text{ in } \mathbb{N}_0, \text{ such that } j \neq k. \tag{7.29}$$

To prove (i), consider k, j in $\mathbb{N}_0$, such that $k > j \geq 0$. We then have

$$\big(Cf\big)(j) = \begin{cases} c_{j,j-1}f(j-1) + c_{j,j}f(j) + c_{j,j+1}f(j+1), & \text{if } j \geq 1, \\ c_{0,0}f(0) + c_{0,1}f(1), & \text{if } j = 0, \end{cases}$$

and since $k \neq 0$,

$$\big(Cg\big)(k) = c_{k,k-1}g(k-1) + c_{k,k}g(k) + c_{k,k+1}g(k+1).$$

Thus,

$$[Cf, Cg]_{j,k} = \begin{cases} \sum_{l=j-1}^{j+1} \sum_{m=k-1}^{k+1} c_{jl}c_{km}[f,g]_{l,m}, & \text{if } j \geq 1, \\ \sum_{l=0}^{1} \sum_{m=k-1}^{k+1} c_{0l}c_{km}[f,g]_{l,m}, & \text{if } j = 0. \end{cases}$$

Assume first that $k \geq j+2$. In this case, $l \leq j+1 \leq k-1 \leq m$, for all terms in the above sums, and thus, by (f) and (7.28), $[f,g]_{l,m} \geq 0$. Since $c_{lm} \geq 0$ for all l, m in $\mathbb{N}_0$ (by (a)), it follows thus that $[Cf, Cg]_{j,k} \geq 0$.
Assume next that $k = j+1$, and consider first the case $j \geq 1$. Then

$$[Cf, Cg]_{j,k} = \sum_{l=j-1}^{j+1} \sum_{m=j}^{j+2} c_{jl}c_{j+1,m}[f,g]_{l,m}. \tag{7.30}$$

In 8 of the 9 terms in the sum above, $l \leq m$, and hence $[f,g]_{l,m} \geq 0$. Only in the case $(l,m) = (j+1, j)$, do we have $l > m$. However, the sum of the two terms corresponding to $(l,m) = (j, j+1)$ and $(l,m) = (j+1, j)$ is non-negative, since

$$\begin{aligned} c_{jj}c_{j+1,j+1}[f,g]_{j,j+1} &+ c_{j,j+1}c_{j+1,j}[f,g]_{j+1,j} \\ &= (c_{jj}c_{j+1,j+1} - c_{j,j+1}c_{j+1,j})[f,g]_{j,j+1}, \end{aligned}$$

which is non-negative by (7.29). Since the remaining 7 terms in the sum on the right hand side of (7.30) are also non-negative, it follows that $[Cf, Cg]_{j,k} \geq 0$. If $j = 0$, and $k = j + 1 = 1$, the same argument can be used to show that

$$[Cf, Cg]_{0,1} = \sum_{l=0}^{1} \sum_{m=0}^{2} c_{0l} c_{1m} [f, g]_{l,m} \geq 0.$$

This proves (i).
To prove (ii), note first that by (a) and (c), we have

$$\big(Df\big)(j) = \big(Cf\big)(j), \quad \text{if } j \geq 1,$$

and

$$\big(Df\big)(0) = \big(Cf\big)(0) + (d_{00} - c_{00}) f(0).$$

Hence, if $k > j \geq 1$, we get from (i), that

$$[Df, Cg]_{j,k} = [Cf, Cg]_{j,k} \geq 0.$$

If $k > j = 0$, then

$$\begin{aligned} [Df, Cg]_{0,k} &= \big(Df\big)(0)\big(Cg\big)(k) - \big(Df\big)(k)\big(Cg\big)(0) \\ &= [Cf, Cg]_{0,k} + (d_{00} - c_{00}) f(0)(Cg)(k). \end{aligned}$$

But $(d_{00} - c_{00}) f(0) \geq 0$ by (d) and (e), and since also $(Cg)(k) = \sum_{l=0}^{\infty} c_{kl} g(l) \geq 0$, by (a) and (e), it follows by (i), that also $[Df, Cg]_{0,k} \geq 0$. This proves (ii).
Next, (iii) follows from (ii) and induction on n, and from noting (by induction), that $(D^n f)(j), (C^n g)(j) \geq 0$ for all n in $\mathbb{N}$ and j in $\mathbb{N}_0$.
To prove (iv), we let t be a fixed number in $[0, \infty[$, and put

$$C_n = \mathbf{1} + \tfrac{t}{n} C, \quad \text{and} \quad D_n = \mathbf{1} + \tfrac{t}{n} D, \qquad (n \in \mathbb{N}_0).$$

Then, for all n, C_n and D_n are positive selfadjoint operators on $\ell_2(\mathbb{N}_0)$, which also satisfy the requirements (a)-(d). Hence, if $f, g \in \ell_2(\mathbb{N}_0)$ which satisfy (e) and (f), we conclude from (iii), that

$$\Big[\big(\mathbf{1} + \tfrac{t}{n} D\big)^n f, \big(\mathbf{1} + \tfrac{t}{n} C\big)^n g\Big]_{j,k} \geq 0, \qquad \text{when } j > k,$$

and hence, letting $n \to \infty$, we get that

$$\Big[\exp(tD) f, \exp(tC) g\Big]_{j,k} \geq 0, \qquad \text{when } j > k,$$

as desired. $\square$

End of Proof of Proposition 7.5. Only (iii) in Proposition 7.5 remains to be proved. Let A, B from $\mathcal{B}\big(L_2([a,b], \mu_c)\big)$ be as in the first part of the proof of Proposition 7.5. Since A is the multiplication operator associated to a positive function on $[a, b]$, and since $B \geq A$, both A and B are positive selfadjoint operators on $L_2([a,b],\mu_c)$. Let C and D be the operators in $\mathcal{B}\big(\ell_2(\mathbb{N}_0)\big)$ corresponding to A and B respectively, via the natural Hilbert space isomorphism between $L_2([a,b],\mu_c)$ and $\ell_2(\mathbb{N}_0)$, given by the orthonormal basis $(\rho_q^c)_{q\in\mathbb{N}_0}$ for $L_2([a,b],\mu_c)$. Then C and D are positive selfadjoint operators and by (7.22) and (7.24), they satisfy the conditions (a)-(d) of Lemma 7.6. Now, let both f and g be the first basis vector in the natural basis for $\ell_2(\mathbb{N}_0)$ (i.e., $f(k) = g(k) = \delta_{k,0}$ for all k in $\mathbb{N}_0$). Then (e),(f) of Lemma 7.6 are also satisfied, and hence we obtain from (iv) of that lemma, that for all j,k in $\mathbb{N}_0$ such that $k > j$,

$$\big(\exp(tD)f\big)(j)\big(\exp(tC)f\big)(k) - \big(\exp(tD)f\big)(k)\big(\exp(tC)f\big)(j) \geq 0,$$

i.e.,

$$\big\langle \exp(tB)\rho_0^c, \rho_j^c\big\rangle \cdot \big\langle \exp(tA)\rho_0^c, \rho_k^c\big\rangle \geq \big\langle \exp(tB)\rho_0^c, \rho_k^c\big\rangle \cdot \big\langle \exp(tA)\rho_0^c, \rho_j^c\big\rangle.$$

For $j = 0$, we get in particular,

$$\frac{M(\exp(tB))_{k,0}}{M(\exp(tA))_{k,0}} \leq \frac{M(\exp(tB))_{0,0}}{M(\exp(tA))_{0,0}}, \qquad (k \in \mathbb{N}_0). \tag{7.31}$$

Note here, that by (7.26),

$$(-1)^k M(\exp(-tA))_{k,0} = \exp(-2(c+1)t)M(\exp(tB))_{k,0} > 0, \qquad (k \in \mathbb{N}_0).$$

Inserting this in (7.31), it follows that

$$\frac{(-1)^k M(\exp(-tA))_{k,0}}{M(\exp(tA))_{k,0}} \leq \frac{M(\exp(-tA))_{0,0}}{M(\exp(tA))_{0,0}}, \qquad (k \in \mathbb{N}_0). \tag{7.32}$$

By (7.23) and (7.27),

$$M(\exp(\pm tA))_{k,0} = c^{-\frac{k}{2}} \int_a^b \exp(\pm tx) P_k^c(x)\, d\mu_c(x) = c^{\frac{k}{2}} \psi_k^c(\pm t), \qquad (k \in \mathbb{N}_0).$$

Hence, (iii) in Proposition 7.5 follows from (7.32). $\square$

7.8 Theorem. *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and let $a_1, \dots, a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$ such that $\sum_{i=1}^r a_i^* a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$, and $\sum_{i=1}^r a_i a_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{K})}$, for some constant c in $[1,\infty[$. Consider furthermore independent elements $Y_1, \dots, Y_r$ of $\mathrm{GRM}(n,n,\frac{1}{n})$, and put $S = \sum_{i=1}^r a_i \otimes Y_i$. Then for any t in $[0, \frac{n}{2c}]$,*

$$\mathbb{E}\big[\exp(-tS^*S)\big] \leq \exp\big(-(\sqrt{c}-1)^2 t + (c+1)^2 \cdot \tfrac{t^2}{n}\big) \cdot \mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}. \tag{7.33}$$

Proof. Consider a fixed t in $[0, \frac{n}{2c}]$. By Proposition 7.4 and Proposition 7.5 we then have

$$\begin{aligned}\big\|\mathbb{E}\big[\exp(-tS^*S)\big]\big\| &\leq \sum_{q=0}^\infty |\psi_q^c(-t)| \cdot \big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| \\ &\leq \frac{\psi_0^c(-t)}{\psi_0^c(t)} \sum_{q=0}^\infty \psi_q^c(t)\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\|.\end{aligned} \tag{7.34}$$

From (7.16) in the proof of Proposition 7.4, we have here that

$$\begin{aligned}\sum_{q=0}^{\infty}\psi_q^c(t)\cdot\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| &\le \sum_{p=0}^{\infty}\tfrac{1}{p!}\big(\tfrac{ct}{n}\big)^p\sum_{\pi\in S_p}n^{k(\hat\pi)}\big(\tfrac{n}{c}\big)^{l(\hat\pi)-1}\\ &\le \sum_{p=0}^{\infty}\tfrac{1}{p!}\big(\tfrac{ct}{n}\big)^p\sum_{\substack{k,l\in\mathbb{N}\\ k+l\le p+1}}\delta(p,k,l)n^k\big(\tfrac{n}{c}\big)^{l-1},\end{aligned}$$

where $\delta(p,k,l)$ was introduced in (3.6). Applying now Lemma 3.6, we get for t in $[0,\frac{n}{2c}]$, that

$$\begin{aligned}\sum_{q=0}^{\infty}\psi_q^c(t)\cdot\big\|\mathbb{E}\big[P_q^c(S^*S)\big]\big\| &\le \exp\big((n+\tfrac{n}{c})(\tfrac{ct}{n})^2\big)\int_a^b\exp\big(\tfrac{n}{c}(\tfrac{ct}{n}x)\big)\,d\mu_c(x)\\ &\le \exp\big((c+1)^2\cdot\tfrac{t^2}{n}\big)\int_a^b\exp(tx)\,d\mu_c(x).\end{aligned}$$

Note here, that $\psi_0^c(t)=\int_a^b\exp(tx)\,d\mu_c(x)$, and hence we get by (7.34), that

$$\begin{aligned}\big\|\mathbb{E}\big[\exp(-tS^*S)\big]\big\| &\le \exp\big((c+1)^2\cdot\tfrac{t^2}{n}\big)\psi_0^c(-t)\\ &= \exp\big((c+1)^2\cdot\tfrac{t^2}{n}\big)\int_a^b\exp(-tx)\,d\mu_c(x).\end{aligned}$$

But $\exp(-tx)\le\exp(-ta)=\exp(-t(\sqrt{c}+1)^2)$ for all x in $[a,b]$, and hence it follows that

$$\big\|\mathbb{E}\big[\exp(-tS^*S)\big]\big\|\le\exp\big((c+1)^2\cdot\tfrac{t^2}{n}\big)\exp(-(\sqrt{c}-1)^2t),\qquad (t\in[0,\tfrac{n}{2c}]).$$

This proves (7.33). □

7.9 Remark. By application of the method of Remark 3.7, it is easy to extend Theorem 7.8, to the case where

$$\sum_{i=1}^{r}a_i^*a_i=c\mathbf{1}_{\mathcal{B}(\mathcal{H})},\qquad\text{and}\qquad\sum_{i=1}^{r}a_ia_i^*=d\mathbf{1}_{\mathcal{B}(\mathcal{K})},$$

for constants c,d such that $c\ge d>0$. In this case, one obtains that for t in $[0,\frac{n}{2c}]$,

$$\mathbb{E}\big[\exp(-tS^*S)\big]\le\exp\big(-(\sqrt{c}-\sqrt{d})^2t+(c+d)^2\cdot\tfrac{t^2}{n}\big)\cdot\mathbf{1}_{\mathcal{B}(\mathcal{H}^n)}.\qquad\square$$

8 Asymptotic Lower Bound on the Spectrum of $S_n^*S_n$ in the Exact Case

Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and consider elements $a_1,\dots,a_r$ of $\mathcal{B}(\mathcal{H},\mathcal{K})$. Let $\mathcal{A}$ denote the C^*-subalgebra of $\mathcal{B}(\mathcal{H})$, generated by the family $\{a_i^*a_j \mid i,j\in\{1,2,\dots,r\}\}$. Consider furthermore, for each n in $\mathbb{N}$, independent elements $Y_1^{(n)},\dots,Y_r^{(n)}$ of $\mathrm{GRM}(n,n,\frac{1}{n})$, and define

$$S_n=\sum_{i=1}^r a_i\otimes Y_i^{(n)}, \qquad (n\in\mathbb{N}). \tag{8.1}$$

In this section, we shall determine (almost surely), the asymptotic behavior of the *smallest* element of the spectrum of $S_n^*S_n$, under the assumptions that $\mathcal{A}$ is an exact C^*-algebra and that $a_1,\dots,a_r$ satisfy the condition

$$\sum_{i=1}^r a_i^*a_i=c\mathbf{1}_{\mathcal{B}(\mathcal{H})} \qquad\text{and}\qquad \sum_{i=1}^r a_ia_i^*\le\mathbf{1}_{\mathcal{B}(\mathcal{K})}, \tag{8.2}$$

for some constant c in $[1,\infty[$. We start, however, by considering the simpler case, where, instead of (8.2), $a_1,\dots,a_r$, satisfy the stronger condition

$$\sum_{i=1}^r a_i^*a_i=c\mathbf{1}_{\mathcal{B}(\mathcal{H})} \qquad\text{and}\qquad \sum_{i=1}^r a_ia_i^*=\mathbf{1}_{\mathcal{B}(\mathcal{K})}, \tag{8.3}$$

for some constant c in $[1,\infty[$. Once this simpler case has been handled, we obtain the more general case by virtue of a dilation result.
As in Section 4, we determine first the asymptotic behavior of the smallest eigenvalue of V_n, where

$$V_n=\big(\Phi\otimes\mathrm{id}_n\big)(S_n^*S_n), \qquad (n\in\mathbb{N}), \tag{8.4}$$

and $\Phi\colon\mathcal{A}\to M_d(\mathbb{C})$ is a completely positive mapping, for some d in $\mathbb{N}$.

8.1 Lemma. *Let S_n, $n\in\mathbb{N}$, and V_n, $n\in\mathbb{N}$, be as in* (8.1) *and* (8.4), *and assume that $a_1,\dots,a_r$ satisfy the condition* (8.3). *Let $\lambda_{\min}(V_n)$ denote the smallest eigenvalue of V_n (considered as an element of $M_{dn}(\mathbb{C})$). Then for any ϵ in $]0,\infty[$, we have that*

$$\sum_{n=1}^\infty P\big(\lambda_{\min}(V_n)\le(\sqrt{c}-1)^2-\epsilon\big)<\infty.$$

Proof. The proof is basically the same as the proof of Lemma 4.2; the main difference being that in this proof we apply Theorem 7.8 instead of Theorem 3.3. Consequently, we shall not repeat all details in this proof.

For fixed n in $\mathbb{N}$, and arbitrary t in $]0,\infty[$, we find that

$$\begin{aligned} P\big(\lambda_{\min}(V_n) \le (\sqrt{c}-1)^2-\epsilon\big) \\ &= P\big(\exp\big(-t\lambda_{\min}(V_n)+t(\sqrt{c}-1)^2-t\epsilon\big)\ge 1\big) \\ &\le \exp(t(\sqrt{c}-1)^2-t\epsilon)\cdot\mathbb{E}\big[\exp\big(-t\lambda_{\min}(V_n)\big)\big] \\ &\le \exp(t(\sqrt{c}-1)^2-t\epsilon)\cdot\mathbb{E}\circ\mathrm{Tr}_{dn}\big[\exp(-tV_n)\big]. \end{aligned} \tag{8.5}$$

By application of Lemma 4.1(ii), we have here, that

$$\begin{aligned} \mathrm{tr}_{dn}\big[\exp(-tV_n)\big] &= \mathrm{tr}_{dn}\big[\exp\big(-t(\Phi\otimes\mathrm{id}_n)(S_n^*S_n)\big)\big] \\ &\le \mathrm{tr}_{dn}\big[\big(\Phi\otimes\mathrm{id}_n\big)(\exp(-tS_n^*S_n))\big] \\ &= \mathrm{tr}_d\otimes\mathrm{tr}_n\big[\big(\Phi\otimes\mathrm{id}_n\big)(\exp(-tS_n^*S_n))\big] \\ &= \phi\otimes\mathrm{tr}_n\big[\exp(-tS_n^*S_n)\big], \end{aligned} \tag{8.6}$$

where ϕ is the state $\mathrm{tr}_d\circ\Phi$ on $\mathcal{A}$. It follows here from Definition 3.1 and Theorem 7.8, that

$$\begin{aligned} \mathbb{E}\big[\phi\otimes\mathrm{tr}_n\big(\exp(-tS_n^*S_n)\big)\big] &= \phi\otimes\mathrm{tr}_n\big(\mathbb{E}\big[\exp(-tS_n^*S_n)\big]\big) \\ &\le \exp\big(-t(\sqrt{c}-1)^2+\tfrac{t^2}{n}(c+1)^2\big), \end{aligned} \tag{8.7}$$

for all t in $]0,\frac{n}{2c}]$. Combining now (8.5)-(8.7), it follows that for all t in $]0,\frac{n}{2c}]$,

$$\begin{aligned} P\big(\lambda_{\min}(V_n) \le & (\sqrt{c}-1)^2-\epsilon\big) \\ &\le dn\cdot\exp(t(\sqrt{c}-1)^2-t\epsilon)\cdot\exp\big(-t(\sqrt{c}-1)^2+\tfrac{t^2}{n}(c+1)^2\big) \\ &= dn\cdot\exp\big(t(\tfrac{t}{n}(c+1)^2-\epsilon)\big). \end{aligned}$$

From here, the proof is concluded exactly as the proof of Theorem 4.2. □

8.2 Proposition. *Let S_n, $n\in\mathbb{N}$, and V_n, $n\in\mathbb{N}$, be as in* (8.1) *and* (8.4), *and assume that $a_1,\dots,a_r$ satisfy the condition* (8.3). *We then have*

$$\liminf_{n\to\infty}\lambda_{\min}(V_n)\ge\big(\sqrt{c}-1\big)^2,\quad \textit{almost surely.}$$

Proof. By Lemma 4.2 and the Borel-Cantelli Lemma (cf. [Bre, Lemma 3.14]), we have for any ϵ from $]0,\infty[$, that

$$P\big(\lambda_{\min}(V_n)\ge(\sqrt{c}-1)^2-\epsilon,\ \text{for all but finitely many } n\big)=1,$$

and from this the proposition follows readily. □

The next two lemmas enable us to pass from the situation considered in Proposition 8.2 to the more general situation, where it is only assumed that $a_1,\dots,a_r$ satisfy (8.2).

8.3 Lemma. *Let c be a number in $[1,\infty[$, and put $q = 2+[c]$, where $[c]$ denotes the integer part of c. Then there exist elements $x_1,\dots,x_q$ in the Cuntz algebra O_2, such that*

$$\sum_{i=1}^{q} x_i^* x_i = c\mathbf{1}_{O_2}, \qquad \textit{and} \qquad \sum_{i=1}^{q} x_i x_i^* = \mathbf{1}_{O_2}.$$

Proof. Recall that O_2 is the unital C^*-algebra $C^*(s_1,s_2)$ generated by two operators s_1,s_2 satisfying that $s_i^*s_j = \delta_{i,j}\mathbf{1}_{O_2}$, $i,j\in\{1,2\}$, and that $s_1s_1^* + s_2s_2^* = \mathbf{1}_{O_2}$. Define then $t_1,\dots,t_{q-1}$ in O_2, by the expression

$$t_j = \begin{cases} s_2^{j-1}s_1, & \text{if } j\in\{1,2,\dots,q-2\},\\ s_2^{q-2}, & \text{if } j = q-1.\end{cases}$$

Then $t_i^*t_j = \delta_{i,j}\mathbf{1}_{O_2}$, for all i,j in $\{1,2,\dots,q-1\}$, and

$$\sum_{j=1}^{q-1} t_jt_j^* = \sum_{i=0}^{q-3} s_2^i(\mathbf{1}_{O_2} - s_2s_2^*)(s_2^i)^* + s_2^{q-2}(s_2^{q-2})^* = \mathbf{1}_{O_2}, \tag{8.8}$$

(i.e., $t_1,\dots,t_{q-1}$ generates a copy of O_{q-1} inside O_2). Define now $x_1,\dots,x_q$ in O_2, by

$$x_i = \begin{cases} \Big(\dfrac{c-1}{q-2}\Big)^{\frac12} t_i, & \text{if } i\in\{1,2,\dots,q-1\}\\ \Big(\dfrac{q-1-c}{q-2}\Big)^{\frac12}\mathbf{1}_{O_2}, & \text{if } i = q.\end{cases}$$

Then

$$\sum_{i=1}^{q} x_i^*x_i = (q-1)\cdot\frac{c-1}{q-2}\cdot\mathbf{1}_{O_2} + \frac{q-1-c}{q-2}\cdot\mathbf{1}_{O_2} = c\mathbf{1}_{O_2},$$

and by (8.8),

$$\sum_{i=1}^{q} x_ix_i^* = \frac{c-1}{q-2}\cdot\mathbf{1}_{O_2} + \frac{q-1-c}{q-2}\cdot\mathbf{1}_{O_2} = \mathbf{1}_{O_2}.$$

Thus, $x_1,\dots,x_q$ have the desired properties. $\square$

8.4 Lemma. *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$, such that $\sum_{i=1}^r a_i^*a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$, and $\sum_{i=1}^r a_ia_i^* \le \mathbf{1}_{\mathcal{B}(\mathcal{K})}$.*
Then there exist Hilbert spaces $\tilde{\mathcal{H}},\tilde{\mathcal{K}}$, s in $\{r,r+1,r+2,\dots\}$ and elements $\tilde a_1,\dots,\tilde a_s$ of $\mathcal{B}(\tilde{\mathcal{H}},\tilde{\mathcal{K}})$, such that the following conditions hold:
(i) $\tilde{\mathcal{H}}\supseteq\mathcal{H}$ *and* $\tilde{\mathcal{K}}\supseteq\mathcal{K}$.
(ii) $\tilde a_{i|\mathcal{H}} = \begin{cases} a_i, & \textit{if } 1\le i\le r,\\ 0, & \textit{if } r+1\le i\le s.\end{cases}$
(iii) $\sum_{i=1}^s \tilde a_i^*\tilde a_i = c\mathbf{1}_{\mathcal{B}(\tilde{\mathcal{H}})}$ *and* $\sum_{i=1}^s \tilde a_i\tilde a_i^* = \mathbf{1}_{\mathcal{B}(\tilde{\mathcal{K}})}$.

Proof. By Lemma 8.3, we may choose finitely many elements $x_1, \dots, x_q$ of the Cuntz algebra O_2, such that $\sum_{i=1}^q x_i^* x_i = c\mathbf{1}_{O_2}$ and $\sum_{i=1}^q x_i x_i^* = \mathbf{1}_{O_2}$. We assume that O_2 is represented on some Hilbert space $\mathcal{L}$, so that $x_1, \dots, x_r \in \mathcal{B}(\mathcal{L})$. Define then

$$\tilde{\mathcal{H}} = (\mathcal{H} \otimes \mathcal{L}) \oplus (\mathcal{K} \otimes \mathcal{L}) \quad \text{and} \quad \tilde{\mathcal{K}} = (\mathcal{K} \otimes \mathcal{L}) \oplus (\mathcal{H} \otimes \mathcal{L}).$$

For Hilbert spaces $\mathcal{V}, \mathcal{W}$, an element v of $\mathcal{B}(\mathcal{V}, \mathcal{W})$, and an element y of $\mathcal{B}(\mathcal{L})$, we consider $v \otimes y$ as an element of $\mathcal{B}(\mathcal{V} \otimes \mathcal{L}, \mathcal{W} \otimes \mathcal{L})$ in the natural manner. Moreover, given v_{11} in $\mathcal{B}(\mathcal{H} \otimes \mathcal{L}, \mathcal{K} \otimes \mathcal{L})$, v_{12} in $\mathcal{B}(\mathcal{K} \otimes \mathcal{L})$, v_{21} in $\mathcal{B}(\mathcal{H} \otimes \mathcal{L})$ and v_{22} in $\mathcal{B}(\mathcal{K} \otimes \mathcal{L}, \mathcal{H} \otimes \mathcal{L})$, we shall consider the matrix $(v_{ij})_{1 \le i,j \le 1}$ as an element of $\mathcal{B}(\tilde{\mathcal{H}}, \tilde{\mathcal{K}})$ in the usual way. With these conventions, consider now the following elements of $\mathcal{B}(\tilde{\mathcal{H}}, \tilde{\mathcal{K}})$,

$$\begin{aligned}
\tilde{a}_i &= \begin{pmatrix} a_i \otimes \mathbf{1}_{\mathcal{B}(\mathcal{L})} & 0 \\ 0 & 0 \end{pmatrix}, \qquad (i \in \{1, 2, \dots, r\}), \\
b_j &= \begin{pmatrix} 0 & (\mathbf{1}_{\mathcal{B}(\mathcal{K})} - \sum_{i=1}^r a_i a_i^*)^{\frac{1}{2}} \otimes x_j \\ 0 & 0 \end{pmatrix}, \qquad (j \in \{1, 2, \dots, q\}), \\
c_{i,j,k} &= \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\sqrt{c}} \cdot a_i^* \otimes (x_j x_k) \end{pmatrix}, \qquad (i \in \{1, 2, \dots, r\},\ j, k \in \{1, 2, \dots, q\}).
\end{aligned}$$

It follows then by direct calculation, that

$$\begin{aligned}
&\sum_{i=1}^r \tilde{a}_i^* \tilde{a}_i + \sum_{j=1}^q b_j^* b_j + \sum_{i=1}^r \sum_{j,k=1}^q c_{i,j,k}^* c_{i,j,k} \\
&\quad = \begin{pmatrix} \big[\sum_{i=1}^r a_i^* a_i\big] \otimes \mathbf{1}_{\mathcal{B}(\mathcal{L})} & 0 \\ 0 & \big[c(\mathbf{1}_{\mathcal{B}(\mathcal{K})} - \sum_{i=1}^r a_i a_i^*) + c \sum_{i=1}^r a_i a_i^*\big] \otimes \mathbf{1}_{\mathcal{B}(\mathcal{L})} \end{pmatrix} \\
&\quad = c\mathbf{1}_{\mathcal{B}(\tilde{\mathcal{H}})},
\end{aligned}$$

and that

$$\begin{aligned}
&\sum_{i=1}^r \tilde{a}_i \tilde{a}_i^* + \sum_{j=1}^q b_j b_j^* + \sum_{i=1}^r \sum_{j,k=1}^q c_{i,j,k} c_{i,j,k}^* \\
&\quad = \begin{pmatrix} \big[\sum_{i=1}^r a_i a_i^* + (\mathbf{1}_{\mathcal{B}(\mathcal{K})} - \sum_{i=1}^r a_i a_i^*)\big] \otimes \mathbf{1}_{\mathcal{B}(\mathcal{L})} & 0 \\ 0 & \big[\frac{1}{c} \sum_{i=1}^r a_i^* a_i\big] \otimes \mathbf{1}_{\mathcal{B}(\mathcal{L})} \end{pmatrix} \\
&\quad = \mathbf{1}_{\mathcal{B}(\tilde{\mathcal{K}})}.
\end{aligned}$$

Thus, if we put $s = r + q + rq^2$, and let $\tilde{a}_{r+1}, \tilde{a}_{r+2}, \dots, \tilde{a}_s$, be new names for the elements in the set $\{b_j \mid j \in \{1, \dots, q\}\} \cup \{c_{i,j,k} \mid i \in \{1, \dots, r\},\ j, k \in \{1, \dots, q\}\}$, then it follows that $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_s$ satisfy condition (iii).

Choosing a fixed unit vector ξ in $\mathcal{L}$, we have natural embeddings $\iota_{\mathcal{H}}\colon \mathcal{H} \to \tilde{\mathcal{H}}$ and $\iota_{\mathcal{K}}\colon \mathcal{K} \to \tilde{\mathcal{K}}$ given by the equations

$$\begin{aligned} \iota_{\mathcal{H}}(h) &= (h \otimes \xi) \oplus 0, \qquad (h \in \mathcal{H}),\\ \iota_{\mathcal{K}}(k) &= (k \otimes \xi) \oplus 0, \qquad (k \in \mathcal{K}). \end{aligned}$$

This justifies (i), and moreover, it is straightforward to check, that under the identifications of $\mathcal{H}$ with $\iota_{\mathcal{H}}(\mathcal{H})$ and $\mathcal{K}$ with $\iota_{\mathcal{K}}(\mathcal{K})$, condition (ii) is satisfied. This concludes the proof. $\square$

8.5 Proposition. *Let S_n, $n \in \mathbb{N}$, and V_n, $n \in \mathbb{N}$, be as in* (8.1) *and* (8.4), *and assume now that $a_1, \dots, a_r$ satisfy the condition* (8.2). *Then*

$$\liminf_{n\to\infty} \lambda_{\min}(V_n) \geq \big(\sqrt{c} - 1\big)^2, \qquad \textit{almost surely.}$$

Proof. By Lemma 8.4, we may choose Hilbert spaces $\tilde{\mathcal{H}}, \tilde{\mathcal{K}}$, s in $\{r, r+1, \dots, \}$ and elements $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_s$ of $\mathcal{B}(\mathcal{H}, \mathcal{K})$, such that conditions (i)-(iii) of Lemma 8.4 are satisfied. If $r < s$, then for each n in $\mathbb{N}$ we choose additional elements $Y_{r+1}^{(n)}, \dots, Y_s^{(n)}$ of $\mathrm{GRM}(n, n, \frac{1}{n})$, such that $Y_1^{(n)}, Y_2^{(n)}, \dots, Y_s^{(n)}$ are independent. We then define

$$\tilde{S}_n = \sum_{i=1}^{s} \tilde{a}_i \otimes Y_i^{(n)}, \qquad (n \in \mathbb{N}).$$

Recall from (8.4), that

$$V_n = \big(\Phi \otimes \mathrm{id}_n\big)(S_n^* S_n), \qquad (n \in \mathbb{N}),$$

where $\Phi\colon \mathcal{A} \to M_d(\mathbb{C})$ is a completely positive mapping from the C^*-subalgebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{H})$ generated by $\{a_i^* a_j \mid i, j \in \{1, 2, \dots, r\}\}$, into the matrix algebra $M_d(\mathbb{C})$. By [Pa, Theorem 5.2], there exists a completely positive mapping $\Phi_1\colon \mathcal{B}(\mathcal{H}) \to M_d(\mathbb{C})$ extending Φ. Note that since Φ is unital, so is Φ_1.
Consider next the orthogonal projection $P_{\mathcal{H}}$ of $\tilde{\mathcal{H}}$ onto $\mathcal{H}$. Then the mapping

$$C_{P_{\mathcal{H}}}\colon b \mapsto P_{\mathcal{H}} b P_{\mathcal{H}}\colon \mathcal{B}(\tilde{\mathcal{H}}) \to P_{\mathcal{H}} \mathcal{B}(\tilde{\mathcal{H}}) P_{\mathcal{H}} \simeq \mathcal{B}(\mathcal{H}),$$

is unital completely positive. Hence, so is the mapping $\Phi_2\colon \mathcal{B}(\tilde{\mathcal{H}}) \to M_d(\mathbb{C})$, given by

$$\Phi_2(b) = \Phi_1(P_{\mathcal{H}} b P_{\mathcal{H}}) = \Phi_1 \circ C_{P_{\mathcal{H}}}(b), \qquad (b \in \mathcal{B}(\tilde{\mathcal{H}})).$$

Thus, if we define

$$\tilde{V}_n = \big(\Phi_2 \circ \mathrm{id}_n\big)(\tilde{S}_n^* \tilde{S}_n), \qquad (n \in \mathbb{N}),$$

then it follows from Lemma 8.4(iii) and Proposition 8.2, that

$$\liminf_{n\to\infty} \lambda_{\min}(\tilde{V}_n) \geq \big(\sqrt{c} - 1\big)^2, \qquad \text{almost surely.} \tag{8.9}$$

However, by Lemma 8.4(ii), we have here that

$$\begin{aligned}\tilde{V}_n &= \big(\Phi_2 \otimes \mathrm{id}_n\big)\Big[\sum_{i,j=1}^{s} \tilde{a}_i^*\tilde{a}_j \otimes \big(Y_i^{(n)}\big)^* Y_j^{(n)}\Big] = \sum_{i,j=1}^{s} \Phi_2\big(\tilde{a}_i^*\tilde{a}_j\big) \otimes \big(Y_i^{(n)}\big)^* Y_j^{(n)} \\ &= \sum_{i,j=1}^{s} \Phi_1\big(P_{\mathcal{H}}\tilde{a}_i^*\tilde{a}_j P_{\mathcal{H}}\big) \otimes \big(Y_i^{(n)}\big)^* Y_j^{(n)} = \sum_{i,j=1}^{r} \Phi_1(a_i^* a_j) \otimes \big(Y_i^{(n)}\big)^* Y_j^{(n)} \\ &= \sum_{i,j=1}^{r} \Phi(a_i^* a_j) \otimes \big(Y_i^{(n)}\big)^* Y_j^{(n)} = V_n.\end{aligned}$$

Therefore (8.9) yields the desired conclusion. ☐

It remains now to show that we can replace V_n in Proposition 8.5 by $S_n^*S_n$ itself. Before proceeding with this task, we draw attention to the following simple observation:

8.6 Lemma. *For each n in $\mathbb{N}$, let $\mathcal{B}_n$ be a unital C^*-algebra, and let b_n be an element of $\mathcal{B}_n$. Then for any R in $[0,\infty[$, the following two conditions are equivalent:*

(i) $\limsup_{n\to\infty} \|b_n\| \le R$.

(ii) $\limsup_{n\to\infty} \max(\mathrm{sp}(b_n)) \le R$, *and* $\liminf_{n\to\infty} \min(\mathrm{sp}(b_n)) \ge -R$.

Proof. This is clear, since, for each n, $\|b_n\|$ is the largest of the two numbers $\max(\mathrm{sp}(b_n))$ and $-\min(\mathrm{sp}(b_n))$. ☐

8.7 Theorem. *Let $a_1,\dots,a_r$ be elements of $\mathcal{B}(\mathcal{H},\mathcal{K})$, such that $\sum_{i=1}^r a_i^*a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})}$ and $\sum_{i=1}^r a_ia_i^* \le \mathbf{1}_{\mathcal{B}(\mathcal{K})}$, for some constant c in $[1,\infty[$. Assume, in addition, that the unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{H})$, generated by the set $\{a_i^*a_j \mid i,j,\in\{1,2,\dots,r\}\}$, is exact. Consider furthermore, for each n in $\mathbb{N}$, independent elements $Y_1^{(n)},\dots,Y_r^{(n)}$ of $\mathrm{GRM}(n,n,\frac{1}{n})$, and put $S_n = \sum_{i=1}^r a_i\otimes Y_i^{(n)}$, $n\in\mathbb{N}$. We then have*

$$\liminf_{n\to\infty} \min\big[\mathrm{sp}(S_n^*S_n)\big] \ge (\sqrt{c}-1)^2, \qquad \textit{almost surely.} \tag{8.10}$$

Proof. Put $E = \mathrm{span}\{a_i^*a_j \mid i,j\in\{1,2,\dots,r\}\}$, and note that $x^*\in E$ for all x in E, and that $\mathbf{1}_{\mathcal{A}} = c^{-1}\sum_{i=1}^r a_i^*a_i \in E$. Thus, E is a finite dimensional operator system, and since $\mathcal{A}$ is exact, it follows thus from Proposition 4.4, that for any ϵ from $]0,\infty[$, there exist d in $\mathbb{N}$ and a unital completely positive mapping $\Phi\colon \mathcal{A}\to M_d(\mathbb{C})$, such that

$$\big\|\big(\Phi\otimes\mathrm{id}_n\big)(x)\big\| \ge (1-\epsilon)\|x\|, \qquad (n\in\mathbb{N},\ x\in M_n(E)). \tag{8.11}$$

Consider now a fixed ϵ from $]0,\infty[$, let d, Φ be as described above, and define

$$V_n = \big(\Phi\otimes\mathrm{id}_n\big)(S_n^*S_n), \qquad (n\in\mathbb{N}).$$

Recall then from Proposition 4.3 and Proposition 8.5, that

$$\limsup_{n\to\infty} \max\left[\mathrm{sp}(V_n)\right] \le c+1+2\sqrt{c}, \qquad \text{almost surely,}$$
$$\liminf_{n\to\infty} \min\left[\mathrm{sp}(V_n)\right] \ge c+1-2\sqrt{c}, \qquad \text{almost surely,}$$

and hence that

$$\limsup_{n\to\infty} \max\left[\mathrm{sp}\big(V_n-(c+1)\mathbf{1}_{dn}\big)\right] \le \ \ 2\sqrt{c}, \qquad \text{almost surely,}$$
$$\liminf_{n\to\infty} \min\left[\mathrm{sp}\big(V_n-(c+1)\mathbf{1}_{dn}\big)\right] \ge -2\sqrt{c}, \qquad \text{almost surely.}$$

By Lemma 8.6, this means that

$$\limsup_{n\to\infty} \big\|V_n-(c+1)\mathbf{1}_{dn}\big\| \le 2\sqrt{c}, \qquad \text{almost surely.} \tag{8.12}$$

Note here, that since $S_n^*S_n-(c+1)\mathbf{1}_{\mathcal{A}\otimes M_n(\mathbb{C})}\in M_n(E)$, for all n, it follows from (8.11), that

$$\begin{aligned}\big\|S_n^*S_n-(c+1)\mathbf{1}_{\mathcal{A}\otimes M_n(\mathbb{C})}\big\| &\le (1-\epsilon)^{-1}\big\|\big(\Phi\otimes\mathrm{id}_n\big)\left[S_n^*S_n-(c+1)\mathbf{1}_{\mathcal{A}\otimes M_n(\mathbb{C})}\right]\big\|\\ &= (1-\epsilon)^{-1}\big\|V_n-(c+1)\mathbf{1}_{dn}\big\|,\end{aligned}$$

for all n in $\mathbb{N}$. Hence (8.12) implies that

$$\limsup_{n\to\infty} \big\|S_n^*S_n-(c+1)\mathbf{1}_{\mathcal{A}\otimes M_n(\mathbb{C})}\big\| \le (1-\epsilon)^{-1}\cdot 2\sqrt{c}, \qquad \text{almost surely.}$$

Since this holds for arbitrary ϵ from $]0,\infty[$, it follows that actually

$$\limsup_{n\to\infty} \big\|S_n^*S_n-(c+1)\mathbf{1}_{\mathcal{A}\otimes M_n(\mathbb{C})}\big\| \le 2\sqrt{c}, \qquad \text{almost surely.}$$

By Lemma 8.6, this implies, in particular, that

$$\liminf_{n\to\infty} \min\left[\mathrm{sp}(S_n^*S_n)-(c+1)\right] \ge -2\sqrt{c}, \qquad \text{almost surely,}$$

and this proves (8.10). □

8.8 Remark. As for the upper bound (cf. Section 4), Theorem 8.7 does not, in general, hold without the condition, that the C^*-algebra generated by $\{a_i^*a_j \mid 1\le i,j\le r\}$ be exact. In fact, for any c in $]1,\infty[$, it is possible to choose a finite set of elements $a_1,\dots,a_r$ of $\mathcal{B}(\mathcal{H})$, for an infinite dimensional Hilbert space $\mathcal{H}$, such that

$$\sum_{i=1}^r a_i^*a_i = c\mathbf{1}_{\mathcal{B}(\mathcal{H})} \quad\text{and}\quad \sum_{i=1}^r a_ia_i^* = \mathbf{1}_{\mathcal{B}(\mathcal{H})},$$

but at the same time

$$P\Big(0 \in \mathrm{sp}(S_n^* S_n), \text{ for all but finitely many } n \Big) = 1,$$

where $S_n = \sum_{i=1}^r a_i \otimes Y_i^{(n)}$, as in (8.1). The proof of this is, however, much more complicated than the corresponding proof of the possible violation for the upper bound (cf. Proposition 4.9(ii)), and it will be presented elsewhere. □

9 Comparison of Projections in Exact C^*-algebras and states on the K_0-group

In [Haa], the first named author proved that quasitraces on exact, unital C^*-algebras are traces. This result implies the following two theorems

9.1 Theorem. (cf. [Han], [Haa]) *If $\mathcal{A}$ is an exact, unital, stably finite C^*-algebra, then $\mathcal{A}$ has a tracial state.*

9.2 Theorem. (cf. [BR, Corollary 3.4]) *If $\mathcal{A}$ is an exact, unital C^*-algebra, then every state on $K_0(\mathcal{A})$ comes from a tracial state on $\mathcal{A}$.*

The proof given in [Haa] of the fact that quasitraces in exact unital C^*-algebras are traces, is based on an ultra-product argument, involving ultra products of finite AW^*-algebras. The aim of this section is to show that Theorem 9.1 and Theorem 9.2 can be obtained from the random matrix results of the previous sections, without appealing to results on quasitraces and AW^*-algebras.
We start by recapturing some of the standard notions and notation in connection with comparison theory for projections in C^*-algebras (see e.g. [Bl1],[Bl2],[Cu] and [Go2]). For a C^*-algebra $\mathcal{A}$, we put

$$M_\infty(\mathcal{A}) = \bigcup_{n\in\mathbb{N}} M_n(\mathcal{A}),$$

where elements are identified via the (non-unital) embeddings $M_n(\mathcal{A}) \hookrightarrow M_{n+1}(\mathcal{A})$, given by addition of a row and a column of zeroes. Given two projections p, q in $M_\infty(\mathcal{A})$, we say, as usual, that p and q are (Murray-von Neumann) equivalent, and write $p \sim q$, if there exists a u in $M_\infty(\mathcal{A})$, such that $u^*u = p$ and $uu^* = q$. We let $V(\mathcal{A})$ denote the set of equivalence classes $\langle p\rangle$ of projections p in $M_\infty(\mathcal{A})$, w.r.t. Murray-von Neumann equivalence, and we equip $V(\mathcal{A})$ with an order structure and an addition, as follows: For projections p, q in $M_\infty(\mathcal{A})$, we write $\langle q\rangle \le \langle p\rangle$ if $q \prec p$, i.e., if q is equivalent to a sub-projection of p. Moreover, we define $\langle p\rangle + \langle q\rangle$ to be $\langle p' + q'\rangle$, where p', q' are projections in $M_\infty(\mathcal{A})$, satisfying that $p' \sim p$, $q' \sim q$ and $p' \perp q'$. Finally, for k in $\mathbb{N}$, we let $k\langle p\rangle$ denote the equivalence class $\langle p\rangle + \cdots + \langle p\rangle$ (k summands).
Recall that for a unital C^*-algebra $\mathcal{A}$, $K_0(\mathcal{A})$ is the additive group obtained from the semi group $V(\mathcal{A})$, via the Grothendieck construction (cf. [Bl1]), and

that $K_0(\mathcal{A})_+$ denotes the range of $V(\mathcal{A})$ under the natural map

$$\rho\colon V(\mathcal{A}) \to K_0(\mathcal{A}).$$

In particular, we have that $K_0(\mathcal{A}) = K_0(\mathcal{A})_+ - K_0(\mathcal{A})_+$.
For a projection p in $M_\infty(\mathcal{A})$, we put

$$[p] = \rho(\langle p\rangle).$$

Note then, that for projections p, q in $M_\infty(\mathcal{A})$, $[p] = [q]$ if and only if there exists a projection r in $M_\infty(\mathcal{A})$, such that $\langle p\rangle + \langle r\rangle = \langle q\rangle + \langle r\rangle$.
The four lemmas 9.3-9.6 below are well known and easy, but since we have not been able to find precise references in the literature, we have included proofs of these lemmas.

9.3 Lemma. *Let $\mathcal{A}$ be a C^*-algebra, and let p, q be projections in $\mathcal{A}$. Then with $I(p)$ the ideal in $\mathcal{A}$ generated by p, the following three conditions are equivalent:*
(i) $\langle q\rangle \le k\langle p\rangle$, *for some k in $\mathbb{N}$.*
(ii) $q \in I(p)$.
(iii) $q \in \overline{I(p)}$.

Proof. (i) $\Rightarrow$ (ii) : Assume that (i) holds, i.e., that there exists k in $\mathbb{N}$ and u in $M_k(\mathcal{A})$, such that

$$u^*u = \begin{pmatrix} q & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad uu^* \le \begin{pmatrix} p & & 0 \\ & \ddots & \\ 0 & & p \end{pmatrix}.$$

This implies that u is of the form

$$u = \begin{pmatrix} u_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u_{k1} & 0 & \cdots & 0 \end{pmatrix},$$

where $u_{11}, u_{21}, \dots, u_{k1} \in p\mathcal{A}q$. It follows thus, that

$$q = \sum_{j=1}^k u_{j1}^* u_{j1} = \sum_{j=1}^k u_{j1}^* p u_{j1} \in I(p),$$

as desired.
(ii) $\Rightarrow$ (iii) : This is trivial.
(iii) $\Rightarrow$ (i) : Assume that (iii) holds. Then there exist k in $\mathbb{N}$ and $a_1, \dots, a_k$, $b_1, \dots, b_k$ in $\mathcal{A}$, such that

$$\Big\| \sum_{j=1}^k a_j p b_j - q \Big\| < 1. \tag{9.1}$$

Thus, by [Go2, 10.7],

$$\begin{pmatrix} q & 0 \\ 0 & 0 \end{pmatrix} \prec \begin{pmatrix} p & & 0 \\ & \ddots & \\ 0 & & p \end{pmatrix} \quad \text{in} \quad M_k(\mathcal{A}),$$

i.e., $\langle q \rangle \le k \langle p \rangle$. □

9.4 Lemma. *Let $\mathcal{M}$ be a von Neumann algebra, and let p be a projection in $\mathcal{M}$. Then any σ-weakly lower semi-continuous trace*

$$\tau\colon \big(p\mathcal{M}p\big)_+ \to [0,\infty],$$

has an extension to a σ-weakly lower semi-continuous trace $\tilde{\tau}$ on $\mathcal{M}_+$.

Proof. We can assume that $p \neq 0$. Choose then a maximal family $(p_i)_{i\in I}$ of pairwise orthogonal projections in $\mathcal{M}$, such that $p_i \prec p$ for all i in I. Then, by standard comparison theory, it follows that

$$\sum_{i\in I} p_i = c(p),$$

where $c(p)$ denotes the central support of p in $\mathcal{M}$. Choose next, for each i in I, a partial isometry v_i in $\mathcal{M}$, such that

$$v_i^* v_i = p_i \quad \text{and} \quad v_i v_i^* \le p, \qquad (i \in I).$$

Define then $\tilde{\tau}\colon \mathcal{M}_+ \to [0,\infty]$, by the equation

$$\tilde{\tau}(a) = \sum_{i\in I} \tau(v_i a v_i^*), \qquad (a \in \mathcal{M}_+).$$

Clearly $\tilde{\tau}$ is additive, homogeneous and σ-weakly lower semi-continuous. To show that $\tilde{\tau}$ has the trace property, note first that since $pv_i = v_i$ for all i, we have also that $c(p)v_i = v_i$ for all i. Since $c(p)$ is in the center of $\mathcal{M}$, it follows thus, that for any x in $\mathcal{M}$,

$$\begin{aligned} \tilde{\tau}(xx^*) &= \sum_{i\in I} \tau\big(v_i x x^* v_i^*\big) = \sum_{i\in I} \tau\big(c(p) v_i x x^* v_i^*\big) \\ &= \sum_{i\in I} \tau\big(v_i x c(p) x^* v_i^*\big) = \sum_{i\in I}\sum_{j\in I} \tau\big((v_i x v_j^*)(v_j x^* v_i^*)\big), \end{aligned}$$

and similarly

$$\tilde{\tau}(x^*x) = \sum_{j\in I}\sum_{i\in I} \tau\big((v_j x^* v_i^*)(v_i x v_j^*)\big).$$

But by the trace property of τ on $p\mathcal{M}p$, we have that

$$\tau\big((v_i x v_j^*)(v_j x^* v_i^*)\big) = \tau\big((v_j x^* v_i^*)(v_i x v_j^*)\big),$$

for all i, j, and since all the terms in the above sums are positive, we can permute their order without changing the sums, and thus obtain

$$\tilde{\tau}(xx^*) = \tilde{\tau}(x^*x).$$

Taken together, we have verified that $\tilde{\tau}$ is a σ-weakly lower semi-continuous trace on $\mathcal{M}_+$, and it remains thus to show that $\tilde{\tau}$ coincides with τ on $\big(p\mathcal{M}p\big)_+$. Given a from $\big(p\mathcal{M}p\big)_+$, we have that $v_i a^{\frac{1}{2}} \in p\mathcal{M}p$, for all i, and therefore

$$\tilde{\tau}(a) = \sum_{i\in I} \tau\big((v_i a^{\frac{1}{2}})(a^{\frac{1}{2}} v_i^*)\big) = \sum_{i\in I} \tau\big(a^{\frac{1}{2}} v_i^* v_i a^{\frac{1}{2}}\big) = \tau\big(a^{\frac{1}{2}} c(p) a^{\frac{1}{2}}\big) = \tau(a),$$

as desired. □

9.5 Lemma. *Let $\mathcal{M}$ be a von Neumann algebra, and let $\mathbf{1}$ denote the unit of $\mathcal{M}$. Let furthermore p, q be projections in $\mathcal{M}$, that satisfy the following two conditions:*
(i) $\mathbf{1} \in I(p)$.
(ii) $\tau(q) \le \tau(p)$, for any normal, tracial state τ on $\mathcal{M}$.
Then $q \prec p$.

Proof. Let $\mathcal{M} = e\mathcal{M} \oplus (\mathbf{1}-e)\mathcal{M}$, be the decomposition of $\mathcal{M}$ into a finite part $e\mathcal{M}$ and a properly infinite part $(\mathbf{1}-e)\mathcal{M}$, by a central projection e. Since any normal, tracial state on $\mathcal{M}$ must vanish on $(\mathbf{1}-e)\mathcal{M}$, condition (ii) is equivalent to the condition

$$\tau(eq) \le \tau(ep), \quad \text{for any normal tracial state } \tau \text{ on } e\mathcal{M}.$$

By comparison theory for finite von Neumann algebras (cf. e.g. [KR, Theorem 8.4.3(vii)]), this condition implies that

$$eq \prec ep \quad \text{in} \quad e\mathcal{M}, \tag{9.2}$$

By Lemma 9.3, condition (i) implies that there exists a k in $\mathbb{N}$, such that

$$\mathbf{1} \otimes e_{11} \prec p \otimes \mathbf{1}_k \quad \text{in} \quad M_k(\mathcal{M}),$$

where $(e_{ij})_{1\le i,j\le k}$ are the usual matrix units in $M_k(\mathbb{C})$. Therefore, we have also that

$$(\mathbf{1}-e) \otimes e_{11} \prec (\mathbf{1}-e)p \otimes \mathbf{1}_k \quad \text{in} \quad M_k((\mathbf{1}-e)\mathcal{M}).$$

At the same time, since $\mathbf{1}-e$ is a properly infinite projection in $\mathcal{M}$, we have that

$$(\mathbf{1}-e) \otimes e_{11} \sim (\mathbf{1}-e) \otimes \mathbf{1}_k \quad \text{in} \quad M_k((\mathbf{1}-e)\mathcal{M}).$$

It follows thus, that

$$(\mathbf{1}-e)q \otimes \mathbf{1}_k \le (\mathbf{1}-e) \otimes \mathbf{1}_k \sim (\mathbf{1}-e) \otimes e_{11} \prec (\mathbf{1}-e)p \otimes \mathbf{1}_k \quad \text{in} \quad M_k((\mathbf{1}-e)\mathcal{M}),$$

and by [KR, Exercise 6.9.14], this implies that

$$(\mathbf{1}-e)q \prec (\mathbf{1}-e)p \quad \text{in} \quad (\mathbf{1}-e)\mathcal{M}. \tag{9.3}$$

Combining (9.2) and (9.3), it follows that $q \prec p$, as desired. □

9.6 Lemma. *Let $\mathcal{M}$ be a von Neumann algebra, and let p, q be projections in $\mathcal{M}$. Then the following two conditions are equivalent*
(i) $q \prec p$.
(ii) $q \in I(p)$, and $\tau(q) \le \tau(p)$ for every σ-weakly lower semi-continuous trace τ on $\mathcal{M}_+$.

Proof. Clearly (i) implies (ii). To show that (ii) implies (i), assume that (ii) holds. By Lemma 9.3 there exists then a k in $\mathbb{N}$, such that $\langle q\rangle \le k\langle p\rangle$, i.e., such that

$$q \otimes e_{11} \sim q' \le p \otimes \mathbf{1}_k,$$

for some projection q' in $M_k(\mathcal{M})$. Consider now the von Neumann algebra

$$\mathcal{N} = M_k(p\mathcal{M}p),$$

with unit $\mathbf{1}_{\mathcal{N}} = p \otimes \mathbf{1}_k$. Set $p' = p \otimes e_{11}$. Then p', q' are both projections in $\mathcal{N}$, and

$$\mathbf{1}_{\mathcal{N}} \in I_{\mathcal{N}}(p'), \tag{9.4}$$

where $I_{\mathcal{N}}(p')$ is the ideal in $\mathcal{N}$ generated by p'.
We show next, that

$$\tau(q') \le \tau(p'), \quad \text{for any normal, tracial state } \tau \text{ on } \mathcal{N}. \tag{9.5}$$

Indeed, if τ is a normal, tracial state on $\mathcal{N}$, then by Lemma 9.4, the restriction $\tau_{|\mathcal{N}_+}$ of τ to $\mathcal{N}_+$ can be extended to a σ-weakly lower semi-continuous trace $\tilde{\tau}$ on $M_k(\mathcal{M})_+$. Then the mapping

$$a \mapsto \tilde{\tau}(a \otimes e_{11}), \qquad (a \in \mathcal{M}_+),$$

is a σ-weakly lower semi-continuous trace on $\mathcal{M}_+$, and hence the assumption (ii) yields that

$$\tilde{\tau}(q \otimes e_{11}) \le \tilde{\tau}(p \otimes e_{11}).$$

Since $q' \sim q \otimes e_{11}$, $p' = p \otimes e_{11}$ and $p', q' \in \mathcal{N}$, it follows thus that

$$\tau(q') = \tilde{\tau}(q') = \tilde{\tau}(q \otimes e_{11}) \le \tilde{\tau}(p \otimes e_{11}) = \tilde{\tau}(p') = \tau(p'),$$

which proves (9.5).
Applying now Lemma 9.5, it follows from (9.4) and (9.5), that $q' \prec p'$ in $\mathcal{N}$, and hence that

$$q \otimes e_{11} \sim q' \prec p' = p \otimes e_{11} \quad \text{in} \quad M_k(\mathcal{M}),$$

which implies that $q \prec p$ in $\mathcal{M}$. □

9.7 Proposition. *Let $\mathcal{A}$ be a C^*-algebra, and let p, q be projections in $\mathcal{A}$. Then the following two conditions are equivalent:*
(i) $q \prec p$ in $\mathcal{A}^{**}$.
(ii) $\tau(q) \le \tau(p)$, for every (norm) lower semi-continuous trace τ on $\mathcal{A}_+$.

Proof. (i) $\Rightarrow$ (ii) : Assume that $q \prec p$ in $\mathcal{A}^{**}$, and choose u in $\mathcal{A}^{**}$, such that $u^*u = q$ and $uu^* \le p$. Then $\|u\| \le 1$, and hence by the Kaplansky Density Theorem, we may choose a net $(u_\beta)_{\beta\in B}$ from $\mathcal{A}$, such that $\|u_\beta\| \le 1$, for all β in B, and $u_\beta \to u$ in the strong (operator) topology.
Define now: $v_\beta = pu_\beta q$, $(\beta \in B)$, and note that $v_\beta \to puq = u$ in the strong (operator) topology, so that $v_\beta^* v_\beta \to u^*u = q$ in the weak (operator) topology. Since $\|v_\beta\| \le 1$ for all β, this implies that actually

$$v_\beta^* v_\beta \to q \quad \text{in the } \sigma\text{-weak topology.}$$

Note also, that since $\|u_\beta\| \le 1$ for all β,

$$v_\beta v_\beta^* \le p, \qquad (\beta \in B). \tag{9.6}$$

Recall now that the σ-weak topology on $\mathcal{A}^{**}$ is the weak* topology i.e., the $\sigma(\mathcal{A}^{**}, \mathcal{A}^*)$-topology, and hence its restriction to $\mathcal{A}$ is the weak topology, i.e., the $\sigma(\mathcal{A}, \mathcal{A}^*)$-topology. Since $v_\beta \in \mathcal{A}$ for all β, we have thus, that

$$v_\beta^* v_\beta \to q \quad \text{in the } \sigma(\mathcal{A}, \mathcal{A}^*)\text{-topology.}$$

Consider then the convex hull K of $\{v_\beta^* v_\beta \mid \beta \in B\}$. Then $q \in K^{-\sigma(\mathcal{A},\mathcal{A}^*)}$, but since convex sets in a Banach space have the same closure in weak and norm topology (cf. [KR, Theorem 1.3.4]), it follows that actually $q \in K^{-\mathrm{norm}}$. Hence, we may choose a sequence $(w_n)_{n\in\mathbb{N}}$ from K, which converges to q in norm. Then, for any (norm) lower semi-continuous trace $\tau\colon \mathcal{A}_+ \to [0,\infty]$,

$$\tau(q) \le \liminf_{n\to\infty} \tau(w_n) \le \sup_{\beta\in B} \tau(v_\beta^* v_\beta) = \sup_{\beta\in B} \tau(v_\beta v_\beta^*) \le \tau(p), \tag{9.7}$$

and this proves (i).
(ii) $\Rightarrow$ (i) : Assume (ii) holds. We set out to show that condition (ii) in Lemma 9.6 is satisfied, in the case $\mathcal{M} = \mathcal{A}^{**}$. Consider first the function $\tau_0\colon \mathcal{A}_+ \to [0,\infty]$, defined by

$$\tau_0(a) = \begin{cases} 0, & \text{if } a \in \overline{I_{\mathcal{A}}(p)_+}, \\ \infty, & \text{if } a \in \mathcal{A}_+ \setminus \overline{I_{\mathcal{A}}(p)_+}. \end{cases}$$

Then τ_0 is a (norm) lower semi-continuous trace on $\mathcal{A}_+$, and hence the assumption (ii) yields that $\tau_0(q) \le \tau_0(p) = 0$, which means that $q \in \overline{I_{\mathcal{A}}(p)_+}$. According to Lemma 9.3, this implies that actually $q \in I_{\mathcal{A}}(p) \subseteq I_{\mathcal{A}^{**}}(p)$.
Note next, that for any σ-weakly lower semi-continuous trace τ on $(\mathcal{A}^{**})_+$, the restriction $\tau_{|\mathcal{A}_+}$ is a (norm) lower semi-continuous trace on $\mathcal{A}$, and hence, by the assumption (ii), $\tau(q) \le \tau(p)$.

Taken together, we have verified that the projections p, q satisfy the condition (ii) in Lemma 9.6, in the case $\mathcal{M} = \mathcal{A}^{**}$, and hence this lemma yields that $q \prec p$ in $\mathcal{A}^{**}$, as desired. □

9.8 Corollary. *Let $\mathcal{A}$ be a C^*-algebra, and let p, q be projections in $\mathcal{A}$. Then the following two conditions are equivalent:*
(i) $\exists k \in \mathbb{N}\colon k\langle q\rangle \le (k-1)\langle p\rangle$ in $V(A^{**})$.
(ii) $\exists \epsilon > 0\colon \tau(q) \le (1-\epsilon)\tau(p)$, for any (norm) lower semi-continuous trace τ on $\mathcal{A}_+$.

Proof. (i) $\Rightarrow$ (ii) : Assume that (i) holds, and define, for the existing k, $q' = q \otimes \mathbf{1}_k$ and $p' = p \otimes \big(\sum_{i=1}^{k-1} e_{ii}\big)$. Then q', p' are projections in $M_k(\mathcal{A})$, and the assumption (i) implies that

$$q' \prec p' \quad \text{in} \quad M_k(\mathcal{A}^{**}). \tag{9.8}$$

Given now any (norm) lower semi-continuous trace τ on $\mathcal{A}_+$, note that the expression

$$\tau_k(a) = \sum_{i=1}^{k} \tau(a_{ii}), \qquad (a = (a_{ij}) \in M_k(\mathcal{A})_+),$$

defines a (norm) lower semi-continuous trace τ on $\mathcal{M}_k(\mathcal{A})_+$. Thus, by Proposition 9.7, (9.8) implies that $\tau_k(q') \le \tau_k(p')$, i.e., that $k\tau(q) \le (k-1)\tau(p)$. This shows that (ii) holds for any ϵ in $]0, \frac{1}{k}]$.
(ii) $\Rightarrow$ (i) : Assume that (ii) holds, and choose, for the existing ϵ, a k in $\mathbb{N}$ such that $\frac{1}{k} \le \epsilon$. Define then, for this k, q' and p' as above.
Now, for any (norm) lower semi-continuous trace τ on $M_k(\mathcal{A})_+$, the mapping

$$a \mapsto \tau(a \otimes e_{11}), \qquad (a \in \mathcal{A}_+),$$

is a (norm) lower semi-continuous trace on $\mathcal{A}_+$, and thus the assumption (ii) yields that

$$\tau(q \otimes e_{11}) \le (1-\epsilon)\tau(p \otimes e_{11}) \le \tfrac{k-1}{k} \cdot \tau(p \otimes e_{11}),$$

and hence that

$$\tau(q') = k \cdot \tau(q \otimes e_{11}) \le (k-1) \cdot \tau(p \otimes e_{11}) = \tau(p').$$

According to Proposition 9.7, this means that $q' \prec p'$ in $M_k(\mathcal{A}^{**})(= M_k(\mathcal{A})^{**})$, which shows that (i) holds. □

9.9 Lemma. *Let $\mathcal{A}$ be a C^*-algebra, and let p, q be projections in $\mathcal{A}$. Then the following two conditions are equivalent:*
(i) There exists an ϵ in $]0, \infty[$, such that

$$\tau(q) \le (1-\epsilon)\tau(p), \quad \text{for any (norm) lower semi-continuous trace } \tau \text{ on } \mathcal{A}_+.$$

(ii) There exist ϵ in $]0,\infty[$, r in $\mathbb{N}$ and $a_1,\dots,a_r$ in $\mathcal{A}$, such that

$$\sum_{i=1}^r a_i^* a_i = q, \qquad \text{and} \qquad \sum_{i=1}^r a_i a_i^* \le (1-\epsilon)p.$$

Proof. The proof follows the ideas of the first section of [Haa].
Note first that (ii) clearly implies (i). To show the converse implication, assume that (i) holds. Then, by Corollary 9.8, there exists a k in $\mathbb{N}$, such that

$$q \otimes \mathbf{1}_k \prec p \otimes \big(\textstyle\sum_{i=1}^{k-1} e_{ii}\big) \quad \text{in} \quad M_k(\mathcal{A}^{**}),$$

i.e., such that

$$u^*u = q \otimes \mathbf{1}_k, \qquad \text{and} \qquad uu^* \le p \otimes \big(\textstyle\sum_{i=1}^{k-1} e_{ii}\big), \tag{9.9}$$

for some $u = (u_{ij})_{1\le i,j\le k}$ in $M_k(\mathcal{A}^{**})$. For this u, we have then that

$$\sum_{j=1}^k \sum_{i=1}^k u_{ij}^* u_{ij} = \sum_{j=1}^k (u^*u)_{jj} = kq,$$

and that

$$\sum_{i=1}^k \sum_{j=1}^k u_{ij} u_{ij}^* = \sum_{i=1}^k (uu^*)_{ii} \le (k-1)p.$$

Thus, if $b_1,\dots,b_{k^2} \in \mathcal{A}^{**}$ denote the elements $\frac{1}{\sqrt{k}} u_{ij}$, $i,j \in \{1,2,\dots,k\}$, listed in any fixed order, then we have that

$$\sum_{i=1}^{k^2} b_i^* b_i = q, \qquad \text{and} \qquad \sum_{i=1}^{k^2} b_i b_i^* \le \tfrac{k-1}{k} p.$$

Note also, that (9.9) implies that $b_i \in p\mathcal{A}^{**}q$ for all i. Consider then the subset K of $\mathcal{A} \oplus \mathcal{A}$, defined by

$$K = \Big\{ \Big(\textstyle\sum_{i=1}^r c_i^* c_i,\, g + \sum_{i=1}^r c_i c_i^* \Big) \;\Big|\; r \in \mathbb{N},\ c_1,\dots,c_r \in p\mathcal{A}q,\ g \in (p\mathcal{A}p)_+ \Big\}.$$

Then K is clearly closed under addition and multiplication by a non-negative scalar, and thus K is a convex cone in $\mathcal{A} \oplus \mathcal{A}$.
Recall next, that the σ-strong* topology on a von Neumann algebra $\mathcal{M}$, is generated by the semi-norms

$$x \mapsto \varphi(x^*x + xx^*)^{\frac{1}{2}}, \qquad (\varphi \in (\mathcal{M}_*)_+).$$

Since the σ-strong* continuous functionals on $\mathcal{M}$ are also σ-weakly continuous (i.e., belong to $\mathcal{M}_*$; cf. [Ta, Lemma II.2.4]), any convex set in $\mathcal{M}$ has the same closure in σ-strong* and σ-weak topology. In particular it follows that

$p\mathcal{A}q$ is σ-strong* dense in $p\mathcal{A}^{**}q$, and $(p\mathcal{A}p)_+$ is σ-strong* dense in $(p\mathcal{A}^{**}p)_+$.

Thus, we may choose a net $(c_1^\alpha,\dots,c_{k^2}^\alpha,g^\alpha)_{\alpha\in A}$ in $\big[\oplus_{j=1}^{k^2} p\mathcal{A}q\big]\oplus(p\mathcal{A}p)$, such that

- $c_i^\alpha\to b_i$, in the σ-strong* topology, for all i in $\{1,2,\dots,k^2\}$,
- $g^\alpha\geq 0$, for all α,
- $g^\alpha\to\frac{k-1}{k}p-\sum_{i=1}^{k^2} b_ib_i^*$, in the σ-strong* topology.

It follows then that

$$\lim_\alpha\Big(\sum_{i=1}^{k^2}(c_i^\alpha)^*c_i^\alpha\Big)=q,\quad \sigma\text{-weakly},$$

and that

$$\lim_\alpha\Big(g^\alpha+\sum_{i=1}^{k^2}c_i^\alpha(c_i^\alpha)^*\Big)=\tfrac{k-1}{k}p,\quad \sigma\text{-weakly}.$$

But since the σ-weak topology on $\mathcal{A}^{**}$ is just the weak*-topology (i.e., the $\sigma(\mathcal{A}^{**},\mathcal{A}^*)$-topology), its restriction to $\mathcal{A}$ is the weak topology (i.e., the $\sigma(\mathcal{A},\mathcal{A}^*)$-topology) on $\mathcal{A}$. It follows thus that

$$\big(q,\tfrac{k-1}{k}p\big)\in K^{-\sigma(\mathcal{A}\oplus\mathcal{A},\mathcal{A}^*\oplus\mathcal{A}^*)}.$$

But convex sets in a Banach space have the same closure in weak and norm topology (cf. [KR, Theorem 1.3.4]), so it follows that in fact

$$\big(q,\tfrac{k-1}{k}p\big)\in K^{-\text{norm}}. \tag{9.10}$$

Since $(1-\delta)^{-1}\big(\frac{k-1}{k}+\delta\big)\to\frac{k-1}{k}<1$, as $\delta\to 0$, we may choose δ,ϵ in $]0,1[$, such that

$$(1-\delta)^{-1}\big(\tfrac{k-1}{k}+\delta\big)=1-\epsilon.$$

By (9.10), there exist then r in $\mathbb{N}$, $c_1,\dots,c_r$ in $p\mathcal{A}q$ and g in $(p\mathcal{A}p)_+$, such that

$$\Big\|q-\textstyle\sum_{i=1}^r c_i^*c_i\Big\|<\delta \qquad\text{and}\qquad \Big\|\tfrac{k-1}{k}p-g-\big(\textstyle\sum_{i=1}^r c_ic_i^*\big)\Big\|<\delta. \tag{9.11}$$

The first inequality in (9.11) implies that $\sum_{i=1}^r c_i^*c_i$ is invertible in the C^*-algebra $q\mathcal{A}q$. Let $h\in(q\mathcal{A}q)_+$ denote the inverse of $\sum_{i=1}^r c_i^*c_i$ in $q\mathcal{A}q$. Since

$$(1-\delta)q\leq\sum_{i=1}^r c_i^*c_i\leq(1+\delta)q,$$

it follows then that

$$(1+\delta)^{-1}q\leq h\leq(1-\delta)^{-1}q. \tag{9.12}$$

Define now: $a_i=c_ih^{\frac12}$, $i\in\{1,2,\dots,r\}$. Then $\sum_{i=1}^r a_i^*a_i=q$, and moreover, by (9.12) and the second inequality in (9.11),

$$\begin{aligned}\sum_{i=1}^r a_ia_i^*=\sum_{i=1}^r c_ihc_i^*&\leq(1-\delta)^{-1}\sum_{i=1}^r c_ic_i^*\leq(1-\delta)^{-1}\Big(g+\sum_{i=1}^r c_ic_i^*\Big)\\ &\leq(1-\delta)^{-1}\big(\tfrac{k-1}{k}+\delta\big)p=(1-\epsilon)p.\end{aligned}$$

Thus, it follows that (ii) holds. □

9.10 Theorem. *Let $\mathcal{A}$ be an exact C^*-algebra, and let p, q be projections in $\mathcal{A}$. Assume that there exists ϵ in $]0, \infty[$, such that*

$$\tau(q) \le (1-\epsilon)\tau(p)$$

for any (norm) lower semi-continuous trace $\tau\colon \mathcal{A}_+ \to [0,\infty]$.
Then there exists n in $\mathbb{N}$, such that

$$q \otimes \mathbf{1}_n \prec p \otimes \mathbf{1}_n \quad \textit{in} \quad M_n(\mathcal{A}).$$

Proof. By Lemma 9.9, we get (after multiplying the a_i's from Lemma 9.9(ii) by $(1-\epsilon)^{-\frac{1}{2}}$), that there exist c in $]1,\infty[$, r in $\mathbb{N}$ and $a_1, \dots, a_r$ in $\mathcal{A}$, such that

$$\sum_{i=1}^r a_i^* a_i = cq, \qquad \text{and} \qquad \sum_{i=1}^r a_i a_i^* \le p. \tag{9.13}$$

We may assume that $\mathcal{A}$ is a C^*-subalgebra of $\mathcal{B}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. Then (9.13) implies that we may consider $a_1, \dots, a_r$ as elements of $\mathcal{B}(q(\mathcal{H}), p(\mathcal{H}))$, and that

$$\sum_{i=1}^r a_i^* a_i = c\mathbf{1}_{q(\mathcal{H})}, \qquad \text{and} \qquad \sum_{i=1}^r a_i a_i^* \le \mathbf{1}_{p(\mathcal{H})}.$$

Moreover, the set $\{a_i^* a_j \mid i, j \in \{1, 2, \dots, r\}\}$ is contained in the exact, unital C^*-algebra $q\mathcal{A}q$. Choosing now, for each n in $\mathbb{N}$, independent elements $Y_1^{(n)}, \dots, Y_r^{(n)}$ of $\mathrm{GRM}(n, n, \frac{1}{n})$, it follows from Theorem 8.7, that with

$$S_n = \sum_{i=1}^r a_i \otimes Y_i^{(n)}, \qquad (n \in \mathbb{N}),$$

we have that

$$\liminf_{n\to\infty} \big[\min\big\{\mathrm{sp}(S_n(\omega)^* S_n(\omega))\big\}\big] \ge \big(\sqrt{c}-1\big)^2, \quad \text{for almost all } \omega \text{ in } \Omega.$$

In particular, there exists one(!) ω in Ω, and an n in $\mathbb{N}$, such that $S_n(\omega)^* S_n(\omega)$ is invertible in the C^*-algebra $M_n(q\mathcal{A}q)$. For this pair (ω, n), we define

$$u = S_n(\omega)\big[S_n(\omega)^* S_n(\omega)\big]^{-\frac{1}{2}},$$

where the inverse is formed w.r.t. $M_n(q\mathcal{A}q)$. Then $u \in M_n(p\mathcal{A}q)$, and

$$u^* u = \mathbf{1}_{q(\mathcal{H})} \otimes \mathbf{1}_n = q \otimes \mathbf{1}_n. \tag{9.14}$$

Moreover, $uu^* \in M_n\big(\mathcal{B}(p(\mathcal{H}))\big)$, and since u^*u is a projection in $M_n\big(\mathcal{B}(q(\mathcal{H}))\big)$, uu^* is a projection in $M_n\big(\mathcal{B}(p(\mathcal{H}))\big)$, so that

$$uu^* \le \mathbf{1}_{p(\mathcal{H})} \otimes \mathbf{1}_n = p \otimes \mathbf{1}_n. \tag{9.15}$$

Combining (9.14) and (9.15), we obtain the desired conclusion. □

9.11 Corollary. *If $\mathcal{A}$ is an exact, unital and simple C^*-algebra, and p, q are projections in $\mathcal{A}$, such that $p \neq 0$ and $\tau(q) < \tau(p)$ for all tracial states τ on $\mathcal{A}$, then for some n in $\mathbb{N}$*

$$q \otimes \mathbf{1}_n \prec p \otimes \mathbf{1}_n \quad in \quad M_n(\mathcal{A}). \tag{9.16}$$

Proof. By simplicity of $\mathcal{A}$, $\tau(p) > 0$ for all tracial states τ on $\mathcal{A}$, and hence by weak* compactness of the set of tracial states on $\mathcal{A}$, there exists ϵ in $]0, \infty[$, such that

$$\tau(q) \leq (1 - \epsilon)\tau(p),$$

for all tracial states τ on $\mathcal{A}$. By the assumptions on $\mathcal{A}$, $\mathcal{A}$ is algebraically simple. Hence, every non-zero trace $\tau \colon \mathcal{A}_+ \to [0, \infty]$ is either equal to $+\infty$ on all of $\mathcal{A}_+ \setminus \{0\}$, or proportional to a tracial state. Hence we can apply Theorem 9.10. □

9.12 Remark. In the “inequality” (9.16) in Corollary 9.11, the tensoring with $\mathbf{1}_n$ can in general not be avoided. This follows from Villadsen's result in [Vi] that there exist nuclear (and hence exact) unital simple C^*-algebras with weak perforation. Recall that a unital C^*-algebra $\mathcal{A}$ has weak perforation, if there exists x in $K_0(\mathcal{A})$, such that $x \notin K_0(\mathcal{A})_+$, but $nx \in K_0(\mathcal{A})_+ \setminus \{0\}$, for some n in $\mathbb{N}$. To see how Villadsen's result implies, that we cannot, in general, avoid tensoring with $\mathbf{1}_n$ in (9.16), let $\mathcal{A}$ be a unital exact simple C^*-algebra, and assume that $x \in K_0(\mathcal{A})$, such that $x \notin K_0(\mathcal{A})_+$ and $nx \in K_0(\mathcal{A})_+ \setminus \{0\}$, for some positive integer n. Write then x in the form $x = [p] - [q]$, where p, q are projections in $M_k(\mathcal{A})$ for some k in $\mathbb{N}$. By the assumption that $nx \in K_0(\mathcal{A})_+ \setminus \{0\}$, and the simplicity of $\mathcal{A}$, it is not hard to deduce that

$$(\tau \otimes \mathrm{tr}_k)(p) > (\tau \otimes \mathrm{tr}_k)(q),$$

for all tracial states τ on $\mathcal{A}$, and hence $\tilde{\tau}(p) > \tilde{\tau}(q)$ for all tracial states $\tilde{\tau}$ on $M_k(\mathcal{A})$. However, since $x \notin K_0(\mathcal{A})_+$, q cannot be equivalent to a sub-projection of p. □

9.13 Theorem. *Let $\mathcal{A}$ be a unital, exact C^*-algebra. Then the following two conditions are equivalent:*
(i) *$\mathcal{A}$ has no tracial states.*
(ii) *For some n in $\mathbb{N}$ there exist projections p, q in $M_n(\mathcal{A})$, such that*

$$p \perp q \qquad and \qquad p \sim q \sim \mathbf{1}_{\mathcal{A}} \otimes \mathbf{1}_n.$$

Proof. Clearly, (ii) implies (i). To show the converse implication, assume that (i) holds, and consider then the two projections p', q' in $M_2(\mathcal{A})$ given by

$$p' = \begin{pmatrix} \mathbf{1}_{\mathcal{A}} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \text{and} \qquad q' = \begin{pmatrix} \mathbf{1}_{\mathcal{A}} & 0 \\ 0 & \mathbf{1}_{\mathcal{A}} \end{pmatrix}.$$

Since $\mathcal{A}$ has no tracial states, $\mathcal{A}^{**}$ has no normal tracial states, and hence $\mathcal{A}^{**}$ is a properly infinite von Neumann algebra. Therefore,

$$\langle \mathbf{1}_{\mathcal{A}} \rangle = 4\langle \mathbf{1}_{\mathcal{A}} \rangle \quad \text{in} \quad V(\mathcal{A}^{**}),$$

which implies that

$$\langle p' \rangle = 2\langle q' \rangle \quad \text{in} \quad V\big(M_2(\mathcal{A}^{**})\big).$$

Hence by Corollary 9.8 and Theorem 9.10, there exists an n in $\mathbb{N}$, such that

$$q' \otimes \mathbf{1}_n \prec p' \otimes \mathbf{1}_n \quad \text{in} \quad M_{2n}(\mathcal{A}).$$

Here, $p' \otimes \mathbf{1}_n \sim \begin{pmatrix} \mathbf{1}_{\mathcal{A}} \otimes \mathbf{1}_n & 0 \\ 0 & 0 \end{pmatrix}$, and thus there exists u in $M_{2n}(\mathcal{A})$, such that

$$u^*u = \begin{pmatrix} \mathbf{1}_{\mathcal{A}} \otimes \mathbf{1}_n & 0 \\ 0 & \mathbf{1}_{\mathcal{A}} \otimes \mathbf{1}_n \end{pmatrix}, \qquad \text{and} \qquad uu^* \le \begin{pmatrix} \mathbf{1}_{\mathcal{A}} \otimes \mathbf{1}_n & 0 \\ 0 & 0 \end{pmatrix}. \tag{9.17}$$

The inequality in (9.17) implies that u has the form

$$u = \begin{pmatrix} u_{11} & u_{12} \\ 0 & 0 \end{pmatrix},$$

for suitable u_{11}, u_{12} from $M_n(\mathcal{A})$. The equality in (9.17) yields then subsequently that

$$u_{11}^*u_{11} = u_{12}^*u_{12} = \mathbf{1}_{\mathcal{A}} \otimes \mathbf{1}_n, \qquad \text{and} \qquad u_{11}^*u_{12} = 0.$$

Defining now

$$p = u_{11}u_{11}^* \qquad \text{and} \qquad q = u_{12}u_{12}^*,$$

it follows that p, q are orthogonal projections in $M_n(\mathcal{A})$, such that $p \sim q \sim \mathbf{1}_{\mathcal{A}} \otimes \mathbf{1}_n$. This shows that (ii) holds. □

In particular, Theorem 9.13 implies the validity of Theorem 9.1:

9.14 Corollary. *If $\mathcal{A}$ is an exact, unital, stably finite C^*-algebra, then $\mathcal{A}$ has a tracial state.*

Proof. This is an obvious consequence of Theorem 9.13. □

Consider next an arbitrary unital C^*-algebra $\mathcal{A}$. A function $\varphi\colon V(\mathcal{A}) \to \mathbb{R}$ is said to be a state on $V(\mathcal{A})$, if it satisfies the following three conditions:

- $\varphi(x) \ge 0$, for all x in $V(\mathcal{A})$.
- $\varphi(x+y) = \varphi(x) + \varphi(y)$, for all x, y in $V(\mathcal{A})$.
- $\varphi\big(\langle \mathbf{1}_{\mathcal{A}} \rangle\big) = 1$.

Similarly, a function $\psi\colon K_0(\mathcal{A}) \to \mathbb{R}$ is said to be a state on $K_0(\mathcal{A})$, if it satisfies the conditions:

- $\psi(z) \ge 0$, for all z in $K_0(\mathcal{A})_+$.
- $\psi(z+w) = \psi(z) + \psi(w)$, for all z, w in $K_0(\mathcal{A})$.

- $\psi([\mathbf{1}_\mathcal{A}]) = 1$.

The set of states on $V(\mathcal{A})$ (resp. $K_0(\mathcal{A})$) is denoted by $\mathcal{S}(V(\mathcal{A}))$ (resp. $\mathcal{S}(K_0(\mathcal{A}))$). Note that $\mathcal{S}(V(\mathcal{A}))$ and $\mathcal{S}(K_0(\mathcal{A}))$ are both convex compact sets in "the topology of pointwise convergence". Let $\rho\colon V(\mathcal{A}) \to K_0(\mathcal{A})$ be the natural map introduced in the beginning of this section. Then it is clear, that the map $\psi \mapsto \psi \circ \rho$, $\psi \in \mathcal{S}(K_0(\mathcal{A}))$, gives a one-to-one correspondence between the states on $K_0(\mathcal{A})$ and the states on $V(\mathcal{A})$. Moreover, this map is an affine homeomorphism of $\mathcal{S}(K_0(\mathcal{A}))$ onto $\mathcal{S}(V(\mathcal{A}))$.

9.15 Lemma. *Let $\mathcal{A}$ be a unital, exact C^*-algebra, and let p, q be projections in $\mathcal{A}$, such that*

$$\tau(q) \le \tau(p), \quad \text{for any tracial state } \tau \text{ on } \mathcal{A}. \tag{9.18}$$

Then for any k in $\mathbb{N}$, there exists n in $\mathbb{N}$, such that

$$nk\langle q\rangle \le nk\langle p\rangle + n\langle \mathbf{1}_\mathcal{A}\rangle.$$

Proof. Let k from $\mathbb{N}$ be given, and consider then the projections p', q' in $M_{k+1}(\mathcal{A})$ defined by:

$$p' = p \otimes \big(\textstyle\sum_{i=1}^k e_{ii}\big) + \mathbf{1}_\mathcal{A} \otimes e_{k+1,k+1}, \qquad \text{and} \qquad q' = q \otimes \big(\textstyle\sum_{i=1}^k e_{ii}\big).$$

Given now an arbitrary non-zero, bounded trace τ on $M_{k+1}(\mathcal{A})$, note that the mapping

$$a \mapsto \tau(a \otimes e_{11}), \qquad (a \in \mathcal{A}),$$

is proportional to a tracial state on $\mathcal{A}$. It follows thus from the assumption (9.18), that $\tau(q \otimes e_{11}) \le \tau(p \otimes e_{11})$, and hence

$$\tau(q') = k \cdot \tau(q \otimes e_{11}) \le k \cdot \tau(p \otimes e_{11}) = \tfrac{k}{k+1} \cdot \tau(p \otimes \mathbf{1}_{k+1}) \le \tfrac{k}{k+1} \cdot \tau(p').$$

Since $\mathbf{1}_\mathcal{A} \otimes e_{11} \prec p'$, any unbounded (lower semi-continuous) trace τ on $M_{k+1}(\mathcal{A})$ must take the value $+\infty$ at p', and hence we have also in this case, that

$$\tau(q') \le \tfrac{k}{k+1} \cdot \tau(p').$$

Applying now Theorem 9.10, it follows that there exists an n in $\mathbb{N}$, such that $n\langle q'\rangle \le n\langle p'\rangle$, and hence such that $nk\langle q\rangle \le nk\langle p\rangle + n\langle \mathbf{1}_\mathcal{A}\rangle$, as desired. $\square$

Next, we need the following version of the Goodearl-Handelman theorem (see [Bl2, 3.4.7], [Go1, 7.11] and [BR, Lemma 2.9]).

9.16 Lemma. *Let $\mathcal{A}$ be a unital C^*-algebra, and consider a convex subset K of $\mathcal{S}(V(\mathcal{A}))$, which is closed in "the topology of pointwise convergence". Assume furthermore that the following implication holds*

$$\forall x, y \in V(\mathcal{A})\colon \quad \big[\forall \varphi \in K\colon \varphi(x) \le \varphi(y)\big] \Longrightarrow \big[\forall \varphi \in \mathcal{S}\big(V(\mathcal{A})\big)\colon \varphi(x) \le \varphi(y)\big]. \tag{9.19}$$

Then $K = \mathcal{S}\big(V(\mathcal{A})\big)$.

Proof. By the one-to-one correspondence between states on $K_0(\mathcal{A})$ and states on $V(\mathcal{A})$, we can find a convex compact subset L of $\mathcal{S}(K_0(\mathcal{A}))$, such that

$$K = \{\psi \circ \rho \mid \psi \in L\}.$$

Since $K_0(\mathcal{A}) = \rho(V(\mathcal{A})) - \rho(V(\mathcal{A}))$, condition (9.19) is equivalent to the condition:

$$\forall z \in K_0(\mathcal{A})\colon \quad \big[\forall \psi \in L\colon \psi(z) \geq 0\big] \Longrightarrow \big[\forall \psi \in \mathcal{S}(K_0(\mathcal{A}))\colon \psi(z) \geq 0\big].$$

Thus by [Go1, Corollary 7.11], all the extreme points of $\mathcal{S}(K_0(\mathcal{A}))$ are contained in $\overline{L} = L$. Hence by Krein-Milman's theorem,

$$\mathcal{S}(K_0(\mathcal{A})) \subseteq \overline{\mathrm{conv}(L)} = L,$$

and therefore $L = \mathcal{S}(K_0(\mathcal{A}))$, which is equivalent to the equation: $K = \mathcal{S}(V(\mathcal{A}))$. □

9.17 Theorem. *Let $\mathcal{A}$ be a unital, exact C^*-algebra. Then for any state φ on $V(\mathcal{A})$, there exists a tracial state τ on $\mathcal{A}$, such that*

$$\varphi(\langle p\rangle) = (\tau \otimes \mathrm{Tr}_m)(p), \quad \textit{for all projections } p \textit{ in } M_m(\mathcal{A}), \textit{ and } m \textit{ in } \mathbb{N}. \tag{9.20}$$

Proof. Let K denote the subset of $\mathcal{S}\big(V(\mathcal{A})\big)$ consisting of those states on $V(\mathcal{A})$, that are given by (9.20) for some tracial state τ on $\mathcal{A}$. Then K is clearly a convex, compact subset of $\mathcal{S}\big(V(A)\big)$, and hence, by Lemma 9.16, it suffices to verify that K satisfies condition (9.19). So consider projections p, q in $M_\infty(\mathcal{A})$. We may assume that $p, q \in M_m(\mathcal{A})$, for some sufficiently large positive integer m. Suppose then that

$$(\tau \otimes \mathrm{Tr}_m)(q) \leq (\tau \otimes \mathrm{Tr}_m)(p), \quad \text{for all tracial states } \tau \text{ on } \mathcal{A}.$$

Since any tracial state on $M_m(\mathcal{A})$ has the form $\frac{1}{m} \cdot \tau \otimes \mathrm{Tr}_m$, for some tracial state τ on $\mathcal{A}$, it follows then from Lemma 9.15, that for any k in $\mathbb{N}$, there exists an n in $\mathbb{N}$, such that

$$nk\langle q\rangle \leq nk\langle p\rangle + n\langle \mathbf{1}_{\mathcal{A}} \otimes \mathbf{1}_m\rangle.$$

Hence for any φ in $\mathcal{S}\big(V(\mathcal{A})\big)$, and any k in $\mathbb{N}$, we have that

$$\varphi(\langle q\rangle) \leq \varphi(\langle p\rangle) + \tfrac{m}{k},$$

and this shows that K satisfies condition (9.19). □

Using the one-to-one correspondence between states on $K_0(\mathcal{A})$ and states on $V(\mathcal{A})$, Theorem 9.17 gives a new proof, not relying on quasitraces, for the following

9.18 Corollary. *Let $\mathcal{A}$ be a unital, exact C^*-algebra. Then any state on $K_0(\mathcal{A})$ comes from a tracial state on $\mathcal{A}$, i.e., for every state ψ on $K_0(\mathcal{A})$, there exists a tracial state τ on $\mathcal{A}$, such that*

$$\psi\big([p]\big) = (\tau \otimes \mathrm{Tr}_m)(p), \ \textit{for all projections } p \textit{ in } M_m(\mathcal{A}), \textit{ and all } m \textit{ in } \mathbb{N}. \quad \square$$

References

[Bl1] B. Blackadar, *K-theory for Operator Algebras*, Math. Sci. Research Inst. Publ. 5, Springer Verlag (1986).

[Bl2] B. Blackadar, *Comparison theory for simple C^*-algebras*, Operator Algebras and Applications, London Math. Society Lecture Notes, 135 (1989), 21-54.

[BK] L.G. Brown and H. Kosaki, *Jensen's Inequality in Semi-finite von Neumann algebras*, J. Operator Theory 23 (1990), 3-19.

[BR] B. Blackadar and M. Rørdam, *Extending states on preordered semigroups and the existence of quasitraces on C^*-algebras*, J. Algebra 152 (1992), 240-247.

[Bre] L. Breiman, *Probability*, Classics In Applied Mathematics 7, SIAM (1992).

[BY] Z.D. Bai and Y.Q. Yin, *Limit of the smallest eigenvalue of a large dimensional sample covariance matrix*, Ann. Probab. 21 (1993), 1275-1294.

[Cu] J. Cuntz, *The internal structure of simple C^*-algebras*, Proc. Sympos. Pure Math. 38, Amer. Math. Soc., Providence R.I. (1982), 85-115.

[Gem] S. Geman, *A limit theorem for the norm of random matrices*, Annals of Probability 8 (1980), 252-261.

[GH] K. Goodearl and D. Handelman, *Rank functions and K_0 of regular rings*, J. Pure Appl. Algebra 7 (1976), 195-216.

[Go1] K. Goodearl, *Partial ordered abelian groups with interpolation*, Mathematical Surveys and Monographs 20, Amer. Math. Soc., Providence R.I. (1986).

[Go2] K. Goodearl, *K_0 of Multiplier algebras of C^*-algebras with real rank zero*, K-theory 10 (1996), 419-489.

[Haa] U. Haagerup, *Quasitraces on exact C^*-algebras are traces*, Manuscript distributed at the Operator Algebra Conference in Istanbul, July 1991, and paper in preparation.

[Han] D. Handelman, *Homomorphisms of C^*-algebras to finite AW^*-algebras*, Michigan Math. J. 28 (1991), 229-240.

[HP] U. Haagerup and G. Pisier, *Bounded Linear Operators Between C^*-algebras*, Duke Math. J. 71 (1993), 889-925.

[HT] U. Haagerup and S. Thorbjørnsen, *Random Matrices with Complex Gaussian Entries*, Preprint, Odense University (1998).

[HTF] Higher Transcendental Functions vol. 1-3, *A. Erdélyi, W. Magnus, F. Oberhettinger and F.G. Tricomi* (editors), based in part on notes left by *H. Bateman*, McGraw-Hill Book Company Inc. (1953-55).

[Ki1] E. Kirchberg, *The Fubini Theorem For Exact C^*-algebras*, J. Operator Theory 10 (1983), 3-8.

[Ki2] E. Kirchberg, *On Subalgebras of the Car-Algebra*, J. Functional Analysis 129 (1995), 35-63.

[Ki3] E. Kirchberg, *On non-semisplit extensions, tensor products and exactness of Group C^*-algebras*, Invent. Math. 112 (1993), 449-489.

[KR] R.V. Kadison and J.R. Ringrose, *Fundamentals of the theory of operator algebras, vol. I-II*, Academic Press (1983, 1986).

[LT] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces I and II*, Classics in Mathematics, Springer Verlag (1996).

[OP] F. Oravecz and D. Petz, *On the eigenvalue distribution of some symmetric random matrices*, Acta Sci. Math. (szeged) 63 (1997), 383-395.

[Pa] V.I. Paulsen, *Completely bounded maps and dilations*, Pitman Research Notes in Mathematics Series 146, Longman Scientific & Technical (1986).

[Pe] D. Petz, *Jensen's Inequality for Positive Contractions on Operator Algebras*, Proc. American Mathematical Society 99 (1987), 273-277.

[Ru] W. Rudin, *Functional Analysis*, McGraw-Hill Book Company (1973).

[Sh] D. Shlyakhtenko, *Limit distributions of matrices with bosonic and fermionic entries*, Free Probability Theory (ed. D.V. Voiculescu), Fields Institute Communications 12 (1997), 241-252.

[Si] J.W. Silverstein, *The smallest eigenvalue of a large dimensional Wishart matrix*, Annals of Probability 13 (1985), 1364-1368.

[Sp] R. Speicher, *Multiplicative functions on the lattice of non-crossing partitions and free convolution*, Math. Ann. 298 (1994), 611-628.

[Ta] M. Takesaki, *Theory of Operator Algebras I*, Springer Verlag (1979).

[Th] S. Thorbjørnsen, *Mixed Moments of Voiculescu's Gaussian Random Matrices*, Preprint, Odense University (1999).

[VDN] D.V. Voiculescu, K.J. Dykema and A. Nica, *Free Random Variables*, CRM Monographs Series, vol. 1 (1992).

[Vi] J. Villadsen, *Simple C^*-algebras with Perforation*, J. Functional Analysis 154 (1998), 110-116.

[Vo1] D.V. Voiculescu, *Limit laws for random matrices and free products*, Invent. Math. 104 (1991), 201-220.

[Vo2] D.V. Voiculescu, *Circular and semi-circular systems and free product factors*, Operator Algebras, Unitary Representations, Enveloping Algebras and Invariant Theory, Progress in Mathematics 92, Birkhäuser, Boston (1990), 45-60.

[Wac] K.W. Wachter, *The strong limits of random matrix spectra for sample matrices of independent elements*, Annals of Probability 6, (1978), 1-18.

[Was] S. Wassermann, *Exact C^*-algebras and Related Topics*, Seoul National University Lecture Notes Series 19 (1994).

[YBK] Y.Q. Yin, Z.D. Bai and P.R. Krishnaiah, *On the limit of the largest eigenvalue of the large dimensional sample covariance matrix*, Prob. Theory and related Fields 78 (1988), 509-521.

U. Haagerup
Department of Mathematics
and Computer Science
SDU, Odense University
Campusvej 55, 5230 Odense
Denmark
haagerup@imada.ou.dk

S. Thorbjørnsen
Department of Mathematics
and Computer Science
SDU, Odense University
Campusvej 55, 5230 Odense
Denmark
steenth@imada.ou.dk

On the Automorphism Group of a Complex Sphere

Marco Brunella

Received: May 5, 1999
Revised: September 1, 1999

Communicated by Thomas Peternell

Abstract. Let X be a compact complex threefold with the integral homology of $\mathbf{S}^6$ and let $Aut(X)$ be its holomorphic automorphism group. By [HKP] and [CDP] the dimension of $Aut(X)$ is at most 2. We prove that $Aut(X)$ cannot be isomorphic to the complex affine group.

1991 Mathematics Subject Classification: 14E05, 32J17, 32M05
Keywords and Phrases: Compact complex threefolds, holomorphic automorphisms, flops

A classical problem in the theory of complex manifolds concerns the existence of complex structures on the six-dimensional sphere $\mathbf{S}^6$. Using octonions one can construct almost-complex structures on $\mathbf{S}^6$, but they are not integrable, and in fact no integrable almost-complex structure is known; it is generally believed that they do not exist, and therefore that $\mathbf{S}^6$ provides an example of almost-complex but non-complex manifold. Examples of this kind are abundant in (real) dimension 4 (as a consequence of our rather good understanding of complex surfaces) but are still lacking in higher dimension (as a manifestation of our rather poor understanding of higher dimensional complex manifolds, except, of course, algebraic or Kähler ones). The case of $\mathbf{S}^6$ is perhaps of particular interest because a complex structure on $\mathbf{S}^6$ would give, by blowing up a point, an exotic complex structure on the familiar $\mathbf{C}P^3$. Moreover, it was proved by Borel and Serre in the fifties that $\mathbf{S}^2$ and $\mathbf{S}^6$ are the only spheres which admit an almost-complex structure.

Recently, two papers add new insights into this problem. Campana, Demailly and Peternell prove in [CDP] that a complex threefold X diffeomorphic to $\mathbf{S}^6$ has no nonconstant meromorphic function. Huckleberry, Kebekus and Peternell prove in [HKP] that a complex threefold X diffeomorphic to $\mathbf{S}^6$ is not almost-homogeneous. Due to [CDP], this last result can be reformulated as: the automorphism group $Aut(X)$ of X has dimension less than or equal to 2 (recall that the automorphism group of a compact complex manifold is a finite dimensional complex Lie group [Huc]).

Our aim is to pursue the study of $Aut(X)$. Let $Aut_0(X)$ be the connected component of the identity: it is a connected complex Lie group of dimension ≤ 2, and if it is not abelian then it is isomorphic to $Aff^k(\mathbf{C})$ for some $k \in \mathbf{N}^+ \cup \{\infty\}$, where $Aff^k(\mathbf{C})$ denotes the k-fold covering of the complex affine group $Aff(\mathbf{C})$. The Lie algebra of $Aff^k(\mathbf{C})$ is generated by two vectors ξ, η satisfying $[\xi, \eta] = \eta$, and if $k \neq \infty$ then ξ is the generator of a subgroup isomorphic to $\mathbf{C}^*$ (more precisely, $\mathbf{C}/2\pi ik\mathbf{Z}$). We shall prove that $Aut_0(X)$ cannot be isomorphic to $Aff^k(\mathbf{C})$, $k \in \mathbf{N}^+$; equivalently, if $Aut_0(X)$ contains a $\mathbf{C}^*$-action then $Aut_0(X)$ is abelian.

More generally, we shall work under the hypothesis that X is a compact complex threefold with the $\mathbf{Z}$-homology of $\mathbf{S}^6$; we shall call such an X a *complex homology sphere*. The results of [CDP] and [HKP] are still valid for any complex homology sphere: this is explicit in [CDP] and can be easily checked in [HKP].

Theorem. *Let X be a complex homology sphere. Then the groups $Aff^k(\mathbf{C})$, $k \in \mathbf{N}^+$, do not act faithfully on X.*

The main step of the proof is a "reduction" of the fixed point set of a $\mathbf{C}^*$-action on a complex sphere (incidentally, this furnishes also some simplifications of sections 7-8 of [HKP]). It has been observed in [HKP] that such a fixed point set is either a pair of points or a smooth rational curve. We shall prove that, if the former case occurs, one can find a bimeromorphic transformation $\phi : X - - \rightarrow Y$, where Y is still a complex homology sphere, which maps $Aut_0(X)$ isomorphically onto $Aut_0(Y)$ and moreover maps the $\mathbf{C}^*$-action on X to a $\mathbf{C}^*$-action on Y whose fixed point set is a rational curve. The argument is the following. Using index type considerations [Bot] we find smooth rational curves joining the two fixed points, invariant by the $\mathbf{C}^*$-action, and whose normal bundle is $\mathcal{O}(-1) \oplus \mathcal{O}(-1)$. We perform a bimeromorphic transformation (a flop [Kol]) centered on one of these curves, giving a new complex homology sphere and a new $\mathbf{C}^*$-action. A combinatorial argument shows that after a finite number of steps we arrive at a $\mathbf{C}^*$-action whose fixed point set is a rational curve, as desired.

Once that reduction of fixed points has been done, the commutativity of $Aut_0(X)$ in presence of $\mathbf{C}^*$-subgroups will be proved by a somewhat algebraic argument. Assuming (by contradiction) that $Aut_0(X)$ is not commutative, we show that X contains a rational irreducible (singular) surface, invariant by the $\mathbf{C}^*$-action and containing the stable and unstable manifolds of the fixed point set. This turns out to be impossible. We notice that one can prove the existence of such a rational surface even if $Aut_0(X)$ is commutative and bidimensional; however we are not able, in that case, to produce a contradiction and therefore we do not present here that partial result.

1. Reduction of fixed points of $\mathbf{C}^*$-actions

Let $\rho : \mathbf{C}^* \times X \to X$ be a $\mathbf{C}^*$-action on a complex homology sphere. We will assume, without loss of generality, that ρ is faithful: $\rho_t \neq id$ for $t \neq 1$. We denote by $v = \frac{d}{dt}|_{t=1}\rho_t$ its infinitesimal generator. It is a holomorphic vector field, whose flow is $2\pi i$-periodic, and its zero set coincides with the fixed point set $Fix(\rho)$ of ρ. It is a classical fact [Huc] that v is linearizable near each point of $Fix(\rho)$; in particular $Fix(\rho)$ is a smooth complex submanifold of X.

Lemma 1 [HKP]. *$Fix(\rho)$ is either a pair of points or a smooth rational curve.*

Proof. The set $Fix(\rho)$ coincides with the fixed point set of the $\mathbf{S}^1$-action contained in the $\mathbf{C}^*$-action. Therefore, and because X is a $\mathbf{Z}$-homology sphere, we have that the $\mathbf{Z}$-homology of $Fix(\rho)$ is that of $\mathbf{S}^0$ or $\mathbf{S}^2$ or $\mathbf{S}^4$ [Bor,IV.5.9]. The first case gives $Fix(\rho) = \{a, b\}$, the second one $Fix(\rho) = \mathbf{C}P^1$, and the third one is excluded because no compact complex surface has the $\mathbf{Z}$-homology of $\mathbf{S}^4$ (by the signature formula, for instance). An alternative way to exclude the third case is the following [HKP]: the adjunction formula and $b_2(X) = b_4(X) = 0$ imply that the Euler characteristic of a smooth compact complex surface $S \subset X$ is zero:

$$c_2(S) = c_2(X) \cdot S - c_1(S) \cdot c_1(\mathcal{O}_X(S)|_S) = 0$$

and so S cannot have the $\mathbf{Z}$-homology of $\mathbf{S}^4$. *q.e.d.*

Let $p \in Fix(\rho)$ and let p_1, p_2, p_3 be the eigenvalues of (the linear part of) v at p. They are integers, and the faithfulness of ρ implies that

$$GCD(p_1, p_2, p_3) = 1.$$

Lemma 2. *For every $i, j \in \{1, 2, 3\}$, $i \neq j$, we have*

$$GCD(p_i, p_j) = 1.$$

Proof. Suppose by contradiction that $GCD(p_1, p_2) = n \geq 2$ and let ω be a primitive n-root of 1. Then the periodic biholomorphism ρ_ω is not the identity but its fixed point set contains a smooth compact complex surface S with $p \in S$ and $T_pS = E_{p_1} \oplus E_{p_2}$, where E_{p_j} is the eigenspace corresponding to p_j. As observed in the proof of lemma 1, the Euler characteristic of S is zero. The action ρ restricts to S to a nontrivial (and nonfaithful) action whose fixed point set is $Fix(\rho) \cap S$. This set is nonempty (it contains p) and it is either discrete or a rational curve. In both cases the Poincaré - Hopf formula gives $c_2(S) > 0$, contradiction. *q.e.d.*

In particular, if $Fix(\rho) = \mathbf{C}P^1$ (that is, one of the eigenvalues is zero) then there are only two possibilities for (p_1, p_2, p_3) (up to renumbering and up to reversing the action): $(0, 1, 1)$ and $(0, 1, -1)$. Let us exclude the first one.

Lemma 3. *If $Fix(\rho) = \mathbf{C}P^1$ then the two nonvanishing eigenvalues of v along $Fix(\rho)$ have opposite sign.*

Proof. This can be proved using the Bott formula [Bot]. However, that formula is rather complicated in the case of nonisolated fixed points, and so we prefer to give the following elementary proof. Suppose, by contradiction, that the eigenvalues of v along $Fix(\rho)$ are $(0,1,1)$. Take the quotient of X by the $\mathbf{S}^1$-action contained in the $\mathbf{C}^*$-action. It is easy to see that it is a topological compact manifold M of dimension 5, and $Fix(\rho)$ projects on M to an embedded 2-sphere N: near a point of $Fix(\rho)$ the $\mathbf{S}^1$-action is the product of the trivial action on $\mathbf{C}$ and the action on $\mathbf{C}^2$ tangent to each 3-sphere and inducing there the Hopf fibration, so that the quotient of each 3-sphere is $\mathbf{S}^2$ and the quotient of the $\mathbf{C}^2$ factor is a cone over $\mathbf{S}^2$, that is $\mathbf{R}^3$. More explicitely, near a point of $Fix(\rho)$ we can choose local holomorphic coordinates (x,y,z) so that v is expressed by $x\frac{\partial}{\partial x}+y\frac{\partial}{\partial y}$, and then the quotient map is $\mathbf{C}^3 \to \mathbf{R}^3\times\mathbf{C}$, $(x,y,z)\mapsto (\sqrt{|x|^2+|y|^2}, \frac{x}{y}, z)$, where $\mathbf{R}^3$ is coordinatized by polar coordinates $(r,\theta) \in \mathbf{R}^+ \times \mathbf{C}P^1$. The 2-sphere N is locally given by $\{r=0\}$. The $\mathbf{R}^*$-action contained in the $\mathbf{C}^*$-action projects on M to an action generated by a vector field V vanishing on N, and only there. Up to changing V to $-V$, the sphere N is an attractor: locally, in the same coordinates (r,θ,z) as before, we have $V=-r\frac{\partial}{\partial r}$. We see that the Poincaré - Hopf index of v at N is equal to 2, hence the Euler characteristic of M is also equal to 2. Since M is odd-dimensional, this is an absurd. *q.e.d.*

We shall prove the following result.

Proposition 1. *Let X be a complex homology sphere and let $\rho : \mathbf{C}^*\times X \to X$ be a $\mathbf{C}^*$-action whose fixed point set $Fix(\rho)$ is a pair of points $\{a,b\}$. Then there exists a complex homology sphere Y and a bimeromorphism $\phi : X - - \to Y$ such that:*

i) ϕ conjugates $Aut_0(X)$ to $Aut_0(Y)$;

ii) ϕ conjugates ρ to a $\mathbf{C}^*$-action τ on Y whose fixed point set is a smooth rational curve.

The bimeromorphism ϕ will be a composition of elementary bimeromorphisms that we now describe.

Suppose that X contains a smooth rational curve R whose normal bundle $N_{R,X}$ is $\mathcal{O}(-1)\oplus\mathcal{O}(-1)$. Let $\tilde{X} \xrightarrow{\pi} X$ be the blow-up of X with center R. The exceptional divisor $D\subset\tilde{X}$ is a rational ruled surface over R, more precisely $D = P(N_{R,X}) = \mathbf{C}P^1\times\mathbf{C}P^1$. Hence there are *two* rulings on D: the ruling over R, given by $\pi|_D$, and a second ruling $D \xrightarrow{\sigma} \mathbf{C}P^1$ whose fibres are transverse to the fibres of $\pi|_D$. The normal bundle of D in $\tilde{X}$ has degree -1 on the fibres of $\pi|_D$ and also on the fibres of σ. Hence [Moi] we can contract each fibre of σ to a point: the result is a smooth complex threefold Y and a morphism $\pi' : \tilde{X}\to Y$. The image of D by π' is a smooth rational curve S with normal bundle $N_{S,Y} = \mathcal{O}(-1)\oplus\mathcal{O}(-1)$, and π' is nothing else than the blow-up of Y with center S. The bimeromorphism $\pi'\circ\pi^{-1} : X - - \to Y$ will be called a *flop with center R*. It is in fact the simplest example of a flop [Kol].

Lemma 4. *Y is a complex homology sphere.*

Proof. It follows from

$$H_k(Y,\mathbf{Z}) = H_k(\tilde{X},\mathbf{Z}) = H_k(X,\mathbf{Z}) \qquad \text{if } k \neq 2,4$$

$$H_k(Y,\mathbf{Z}) \oplus \mathbf{Z} = H_k(\tilde{X},\mathbf{Z}) = H_k(X,\mathbf{Z}) \oplus \mathbf{Z} \qquad \text{if } k = 2,4.$$

q.e.d.

It should be possible to prove also that Y is diffeomorphic to X. In fact, there should exist a smooth (non holomorphic!) diffeomorphism of $\tilde{X}$, whose support is localized on a neighbourhood of D, which exchanges the two rulings $\pi|_D$ and σ, proving the diffeomorphicity of X and Y. Remark that the fundamental groups of X and Y are isomorphic. If X is diffeomorphic to $\mathbf{S}^6$ then it is easy to see that Y also is diffeomorphic to $\mathbf{S}^6$, by classical results in differential topology (Smale, Kervaire - Milnor,...).

Lemma 5. *The flop $\pi' \circ \pi^{-1} : X - - \rightarrow Y$ realizes an isomorphism between $Aut_0(X)$ and $Aut_0(Y)$.*

Proof. We simply have to check that for every holomorphic vector field on X (resp. on Y) its transform on Y (resp. on X) is still holomorphic. This follows from the negativity of $N_{R,X}$ and $N_{S,Y}$: every holomorphic vector field on X (resp. on Y) is tangent to R (resp. to S). *q.e.d.*

In order to do flops, we have to find rational curves with normal bundle $\mathcal{O}(-1)\oplus\mathcal{O}(-1)$. This will be based on the following remarks.

Let $Fix(\rho) = \{a,b\}$ and let a_1,a_2,a_3 be the eigenvalues of v at a, b_1,b_2,b_3 those at b. Suppose that $|a_j| \geq 2$, for some j: then by the same argument of the proof of lemma 2 there is a ρ-invariant smooth complex curve $R \subset X$, with $a \in R$ and $T_aR = E_{a_j}$ (observe that, by lemma 2, this eigenspace is one-dimensional). Clearly R is rational and contains a second fixed point, that is $b \in R$. Moreover, for some i we have $T_bR = E_{b_i}$, and $b_i = -a_j$. To fix ideas, suppose $j = i = 1$. The normal bundle of R will be computed by the following formula.

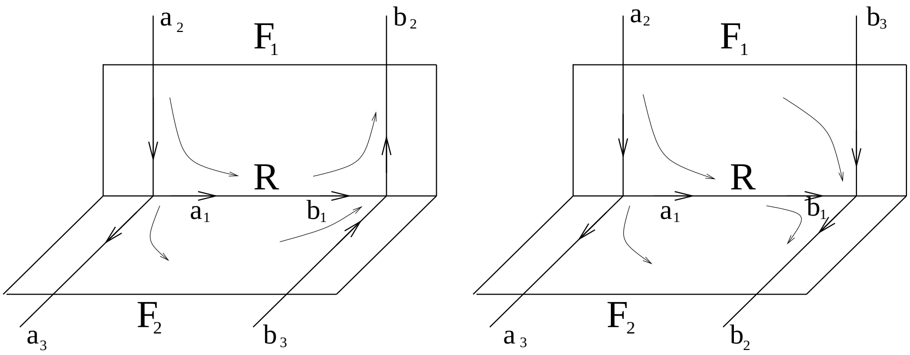
Lemma 6.

$$N_{R,X} = \mathcal{O}(n) \oplus \mathcal{O}(m)$$

where $n = \frac{a_2-b_2}{a_1}, m = \frac{a_3-b_3}{a_1}$ *or* $n = \frac{a_2-b_3}{a_1}, m = \frac{a_3-b_2}{a_1}$.

Proof. We consider the restriction of ρ to R and its natural extension to $N_{R,X}$, via the differential. We therefore are in the situation of [Bot]: a holomorphic vector field (on R) which acts on a vector bundle. Hence the characteristic numbers of that bundle are localized at zeroes of the vector field, that is at a and b.

The bundle $N_{R,X}$ has a splitting $F_1 \oplus F_2$ by line bundles which are invariant by the action: this corresponds to the fact that a $\mathbf{C}^*$-action on a rational ruled surface (in our case $P(N_{R,X})$) has always two disjoint invariant sections. The fibres $(F_i)_a$, $(F_i)_b$ are invariant and their eigenvalues are a_2,a_3,b_2,b_3. Hence there are two possibilities:



From Bott formula [Bot] we deduce in the first case

$$c_1(F_1) = \frac{a_2}{a_1} + \frac{b_2}{b_1} = \frac{a_2 - b_2}{a_1}, \qquad c_1(F_2) = \frac{a_3 - b_3}{a_1}$$

and in the second case

$$c_1(F_1) = \frac{a_2 - b_3}{a_1}, \qquad c_1(F_2) = \frac{a_3 - b_2}{a_1}.$$

q.e.d.

Observe that by adjunction formula and $c_1(X) = 0$ we have $c_1(N_{R,X}) = -c_1(R) = -2$ and consequently $n + m = -2$, i.e.

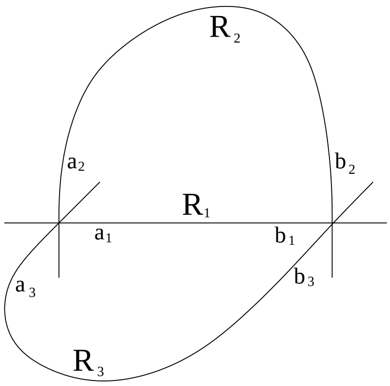
$$a_1 + a_2 + a_3 = b_1 + b_2 + b_3.$$

2. Proof of proposition 1

Let a_1, a_2, a_3 (resp. b_1, b_2, b_3) be the eigenvalues of v at a (resp. at b), with $|a_1| \leq |a_2| \leq |a_3|$.

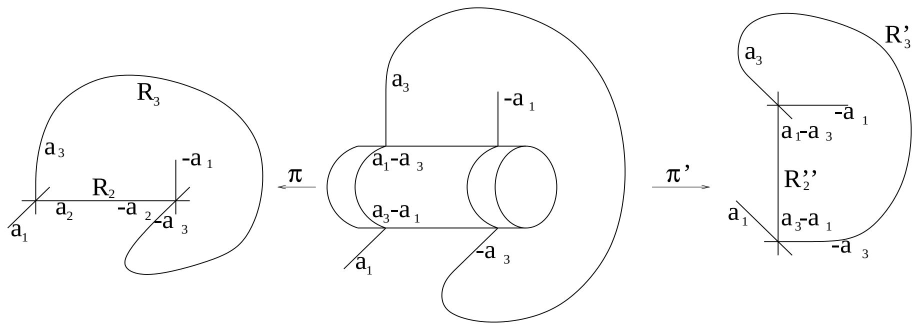
First step: from $|a_1| \geq 2$ to $|a_1| = 1$.

If $|a_1| \geq 2$ then, as explained before lemma 6, we have three smooth ρ-invariant rational curves through a and b, each one connecting a and b:



We have $b_j = -a_j$ for every $j = 1, 2, 3$ (up to renumbering the eigenvalues at b). Remark that $|a_1| \geq 2$ implies $|a_2|, |a_3| \geq 3$ (lemma 2). By lemma 6, the normal bundle $N_{R_2,X}$ is either $\mathcal{O}(\frac{2a_1}{a_2}) \oplus \mathcal{O}(\frac{2a_3}{a_2})$ or $\mathcal{O}(\frac{a_1+a_3}{a_2}) \oplus \mathcal{O}(\frac{a_1+a_3}{a_2})$. The former case is excluded because $\frac{2a_1}{a_2}$ and $\frac{2a_3}{a_2}$ are not integers. The latter case is in fact $\mathcal{O}(-1)\oplus\mathcal{O}(-1)$, because $c_1(N_{R_2,X}) = -2$. Similarly, $N_{R_3,X} = \mathcal{O}(-1)\oplus\mathcal{O}(-1)$.

Hence we can perform flops with center R_2 or R_3. Let us see how a flop with center R_2 transforms the eigenvalues of the $\mathbf{C}^*$-action. After a blow-up π with center R_2 we obtain a $\mathbf{C}^*$-action with four fixed points on the exceptional divisor, two over a and two over b. The rational curves $\pi^{-1}(a)$ and $\pi^{-1}(b)$ are invariant by the action and their eigenvalues are $\pm(a_1 - a_3)$. The blow-down π' maps these two curves onto a rational curve R_2'', with eigenvalues $\pm(a_1 - a_3)$:

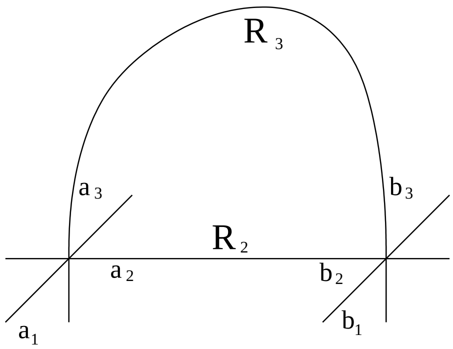


On the new complex homology sphere we therefore have a new (faithful) $\mathbf{C}^*$-action with a fixed point whose eigenvalues are $(a_1, a_3 - a_1, -a_3)$. The strict transform R_3' of R_3 has normal bundle $\mathcal{O}(-1) \oplus \mathcal{O}(-1)$, again by lemma 6. Therefore we can perform a second flop with center R_3': we obtain a new complex homology sphere and a $\mathbf{C}^*$-action with a fixed point whose eigenvalues are $(a_1, a_1 - a_3, a_3 - 2a_1) = (a_1, a_2 + 2a_1, a_3 - 2a_1)$ (recall that $\frac{a_1+a_3}{a_2} = -1$, i.e. $a_1 + a_2 + a_3 = 0$). Of course, we can reverse the order: a flop with center R_3 followed by a flop with center R_2' produces a $\mathbf{C}^*$-action with a fixed point with eigenvalues $(a_1, a_2 - 2a_1, a_3 + 2a_1)$. Remark that these new collections of eigenvalues still satisfy the GCD condition of lemma 2.

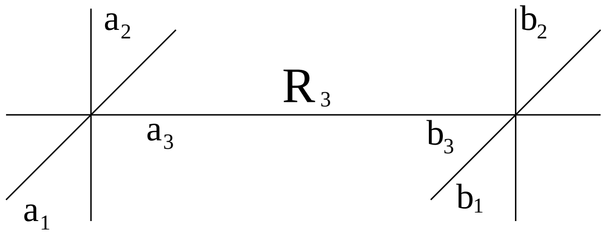
Iterating this process we arrive at a fixed point with eigenvalues (a_1, α, β) and $|\alpha| \leq |a_1|$ ($\alpha = a_2 + 2na_1$ or $a_3 + 2na_1$ for a suitable integer n). Because $GCD(a_1, \alpha) = 1$ and $|a_1| \geq 2$, we have the strict inequalities $0 < |\alpha| < |a_1|$, that is the eigenvalue with smallest modulus has modulus strictly smaller that $|a_1|$. Iterating again we finally arrive at an eigenvalue with modulus equal to 1.

Second step: from $|a_1| = 1$, $|a_2| \geq 2$ to $|a_1| = |a_2| = 1$.

Now we can guarantee only two ρ-invariant rational curves:



We have $b_j = -a_j$ for $j = 2, 3$. We also have $|b_1| = 1$: otherwise $|b_1| \geq 2$ and there would be a third ρ-invariant rational curve tangent to E_{a_1} at a and to E_{b_1} at b, giving $b_1 = -a_1$, thus $|b_1| = |a_1| = 1$, a contradiction.
From $|a_2| \geq 2$ it follows $|a_3| \geq 3$ and (lemma 6) the normal bundle $N_{R_3,X}$ is either $\mathcal{O}(\frac{a_1-b_1}{a_3}) \oplus \mathcal{O}(\frac{2a_2}{a_3})$ or $\mathcal{O}(\frac{a_1+a_2}{a_3}) \oplus \mathcal{O}(\frac{a_2-b_1}{a_3})$. As before, the first possibility is excluded because $\frac{2a_2}{a_3}$ is not an integer. Hence $N_{R_3,X} = \mathcal{O}(\frac{a_1+a_2}{a_3}) \oplus \mathcal{O}(\frac{a_2-b_1}{a_3})$. From $c_1(N_{R_3,X}) = -2$ it follows that $a_1 + 2a_2 - b_1 = -2a_3$. Because $|a_2| \neq |a_3|$, we cannot have $b_1 = a_1$ and so we have $b_1 = -a_1$. This in turn implies $a_1 + a_2 + a_3 = 0$ and $N_{R_3,X} = \mathcal{O}(-1) \oplus \mathcal{O}(-1)$.
If $|a_2| \geq 3$ the same argument applies to R_2, and we obtain $N_{R_2,X} = \mathcal{O}(-1) \oplus \mathcal{O}(-1)$. Then we proceed as in the first step: a sequence of flops produces a $\mathbf{C}^*$-action with a fixed point with eigenvalues (a_1, α, β), $|\alpha| \leq |a_1|$; that is $|\alpha| = |a_1| = 1$.
If $|a_2| = 2$, from $a_1 + a_2 + a_3 = 0$, $|a_1| = 1$ and $|a_3| \geq 3$ we find $a_2 = 2a_1, a_3 = -3a_1$. It is readily checked that a single flop along R_3 produces a $\mathbf{C}^*$-action with a fixed point with eigenvalues $(a_1, a_1, -2a_1)$.
THIRD STEP: the case $|a_1| = |a_2| = 1$, $|a_3| \geq 2$.



We have $b_3 = -a_3$. As before, $|b_1| = |b_2| = 1$. Up to exchanging b_1 and b_2 we obtain $N_{R_3,X} = \mathcal{O}(\frac{a_1-b_1}{a_3}) \oplus \mathcal{O}(\frac{a_2-b_2}{a_3})$ and $a_1 - b_1 + a_2 - b_2 = -2a_3$. From $|a_3| \geq 2$ it follows $a_1 = a_2 = -b_1 = -b_2$ and $a_3 = -2a_1$, therefore $N_{R_3,X} = \mathcal{O}(-1) \oplus \mathcal{O}(-1)$. A flop along R_3 gives a $\mathbf{C}^*$-action with a rational curve of fixed points (and eigenvalues $(0, 1, -1)$).
LAST STEP. To complete the proof of proposition 1 we need to show that the case $|a_1| = |a_2| = |a_3| = 1$ never happens. By the usual argument, if $|a_j| = 1$ for every j then also $|b_j| = 1$ for every j. We now take the Bott formula [Bot]

for $c_1^3(X)$:

$$\frac{(a_1+a_2+a_3)^3}{a_1a_2a_3} + \frac{(b_1+b_2+b_3)^3}{b_1b_2b_3} = c_1^3(X) = 0.$$

If $|a_j| = 1$ for every j then the residue $\frac{(a_1+a_2+a_3)^3}{a_1a_2a_3}$ can take only two values: 27 and -1. The same for $\frac{(b_1+b_2+b_3)^3}{b_1b_2b_3}$. Hence their sum cannot vanish. *q.e.d.*

Remark: we could use the Bott formula since the beginning of the proof and not only in the last step, but it turns out that this would give only minor simplifications (for instance, in the second step we can use the Bott formula to deduce $b_1 = -a_1$ from $|b_1| = |a_1|$). It also turns out that the analogous formula for $c_1(X) \cdot c_2(X)$ yields no further information. The formula for $c_3(X)$ is equivalent to the Poincaré - Hopf formula and was already used, more or less, in lemmata 1, 2 and 3. And, of course, all these formulae do not contradict the existence of a $\mathbf{C}^*$-action with a rational curve of fixed points, with eigenvalues $(0, 1, -1)$.

The automorphism group is abelian

From now on $\rho : \mathbf{C}^* \times X \to X$ will denote a faithful $\mathbf{C}^*$-action on a complex homology sphere with $Fix(\rho) = Z_0$ a smooth rational curve. Around each point of Z_0 we can choose local coordinates (x, y, z) such that the infinitesimal generator v is expressed by $x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}$ (and $Z_0 = \{x = y = 0\}$). If we take a sufficiently small tubular neighbourhood V of Z_0 then the sets

$$W_V^s = \{p \in V | \rho_t(p) \in V \ \forall t, |t| \geq 1, \text{ and } \lim_{t\to\infty} \rho_t(p) = \theta^+(p) \in Z_0\}$$

and

$$W_V^u = \{p \in V | \rho_t(p) \in V \ \forall t, |t| \leq 1, \text{ and } \lim_{t\to 0} \rho_t(p) = \theta^-(p) \in Z_0\}$$

are smooth complex open surfaces, containing Z_0 and intersecting transversely along Z_0. In the above local coordinates, $W_V^s = \{x = 0\}$ and $W_V^u = \{y = 0\}$. Suppose now that

$$dim\ Aut_0(X) = 2.$$

This means that there exists a second holomorphic vector field w on X, linearly independent of v. In fact, w cannot be collinear to v at a generic point of X, because X has no nonconstant meromorphic function [CDP]. The commutator $[v, w]$ is a linear combination $av + bw$, $a, b \in \mathbf{C}$, since the Lie algebra of holomorphic vector fields on X is two-dimensional, spanned by v and w. Because the flow of v is $2\pi i$-periodic, one easily sees that if $b = 0$ then also $a = 0$: when $[v, w] = av$, the flows ϕ_t $(= \rho_{\exp t})$ of v and ψ_t of w are related by $\psi_s \circ \phi_t = \phi_{t\exp(as)} \circ \psi_s$ for every $t, s \in \mathbf{C}$, in particular

$\phi_{2\pi i \exp(as)} = \psi_s \circ \phi_{2\pi i} \circ \psi_{-s} = id$ for every $s \in \mathbf{C}$, so that $\exp(as)$ is an integer for every s and therefore $a = 0$. Hence, up to replacing w by $w + \frac{a}{b}v$ (if $b \neq 0$), we have

$$[v, w] = bw$$

where $b \in \mathbf{Z}$, again for the $2\pi i$-periodicity of the flow of v. Up to changing v to $-v$, we may suppose that $b \in \mathbf{N}$.

In this section we shall prove the commutativity of $Aut_0(X)$, concluding the proof of the theorem stated in the introduction.

Proposition 2. *Let v, w be holomorphic vector fields on a complex homology sphere, where v generates a $\mathbf{C}^*$-action. Then v and w commute: $[v, w] = 0$.*

Let us consider the wedge product $v \wedge w \in H^0(X, TX \wedge TX)$, whose zero set $E \subset X$ is the analytic subset of X where v and w are collinear. Define $O(v \wedge w, W_V^s)$, resp. $O(v \wedge w, W_V^u)$, as the vanishing order of $v \wedge w$ along W_V^s, resp. W_V^u. Of course, if W_V^s is not contained in E (for instance, if W_V^s is not a piece of a compact analytic subset of X) then $O(v \wedge w, W_V^s) = 0$.

Lemma 7.

$$O(v \wedge w, W_V^s) = O(v \wedge w, W_V^u) + b.$$

Proof. We shall conclude by a local computation. Take $p \in Z_0$ and local coordinates (x, y, z) so that $v = x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}$, $w = A\frac{\partial}{\partial x} + B\frac{\partial}{\partial y} + C\frac{\partial}{\partial z}$, $W_V^s = \{x = 0\}$, $W_V^u = \{y = 0\}$. Hence

$$v \wedge w = (xB + yA)\frac{\partial}{\partial x} \wedge \frac{\partial}{\partial y} + xC\frac{\partial}{\partial x} \wedge \frac{\partial}{\partial z} - yC\frac{\partial}{\partial y} \wedge \frac{\partial}{\partial z}$$

and we see that

$$O(v \wedge w, W_V^s) = \min\{O(A, W_V^s), O(B, W_V^s) + 1, O(C, W_V^s)\}$$

$$O(v \wedge w, W_V^u) = \min\{O(A, W_V^u) + 1, O(B, W_V^u), O(C, W_V^u)\}.$$

From $[v, w] = bw$ we obtain the following system of equations:

$$\begin{cases} xA_x - yA_y = (b+1)A \\ xB_x - yB_y = (b-1)B \\ xC_x - yC_y = bC \end{cases}$$

Write $A(x, y, z) = x^h y^k a(x, y, z)$, $h = O(A, W_V^s)$, $k = O(A, W_V^u)$ (i.e., the functions $a(0, y, z)$ and $a(x, 0, z)$ are not identically zero). From the first equation we obtain:

$$xa_x - ya_y = (b + 1 - h + k)a$$

and restricting to $\{y = 0\}$:

$$xa_x(x, 0, z) = (b + 1 - h + k)a(x, 0, z)$$

that is
$$\frac{a_x(x,0,z)}{a(x,0,z)} = \frac{b+1-h+k}{x}.$$

We deduce that $b+1-h+k \geq 0$, because $a(x,0,z)$ is holomorphic and not identically zero. Restricting to $\{x=0\}$ we find the opposite inequality: $b+1-h+k \leq 0$. Hence $b+1-h+k=0$, or more explicitely

$$O(A, W_V^s) = O(A, W_V^u) + b + 1.$$

In a similar way, from the second and the third equations we find

$$O(B, W_V^s) = O(B, W_V^u) + b - 1$$

and
$$O(C, W_V^s) = O(C, W_V^u) + b$$

from which it follows that

$$O(v \wedge w, W_V^s) = O(v \wedge w, W_V^u) + b.$$

q.e.d.

In order to prove proposition 2, suppose now by contradiction that b is strictly positive. In particular $O(v \wedge w, W_V^s) > 0$, so that $v \wedge w$ does vanish on W_V^s. In other words, there exists an irreducible component $N \subset E$, *dim* $N = 2$, which contains W_V^s. Take the restriction of the $\mathbf{C}^*$-action ρ to N, and take an equivariant resolution of singularities $\tilde{N} \to N$, over which ρ can be lifted. On $\tilde{N}$ we therefore have a $\mathbf{C}^*$-action with a rational curve of fixed points (arising from Z_0). It follows from the classification of $\mathbf{C}^*$-actions on compact complex surfaces [Hau] that $\tilde{N}$ is algebraic (and even rational) and that the closure of each orbit of the $\mathbf{C}^*$-action is a (possibly singular) rational curve. Returning to X, we therefore see that for each $p \in W_V^s$ not only $\lim_{t\to\infty} \rho_t(p)$ is a single point on Z_0 (as the definition of W_V^s claims) but also $\lim_{t\to 0} \rho_t(p)$ is a single point, necessarily on Z_0, and so the ρ-orbit through p cuts W_V^u. Varying p on W_V^s we also see that the full W_V^u belongs to N. But this contradicts lemma 7: because $O(v \wedge w, W_V^s) \neq O(v \wedge w, W_V^u)$, the sets W_V^s and W_V^u cannot belong to the same irreducible component of E. This contradiction proves proposition 2.

References

[Bor] A. Borel, *Seminar on transformation groups*, Princeton Univ. Press, Ann. of Math. Studies 46 (1960)
[Bot] R. Bott, *A residue formula for holomorphic vector fields*, Jour. Diff. Geom. 1 (1967), 311-330

[CDP] F. Campana, J.-P. Demailly, T. Peternell, *The algebraic dimension of compact complex threefolds with vanishing second Betti number*, Comp. Math. 112 (1998), 77-91
[Hau] J. Hausen, *Zur Klassifikation glatter kompakter* $\mathbf{C}^*$*-flächen*, Math. Ann. 301 (1995), 763-769
[Huc] A. Huckleberry, *Actions of groups of holomorphic transformations*, Enc. Math. Sci. vol. 69 (Several complex variables VI), Springer Verlag (1990)
[HKP] A. Huckleberry, S. Kebekus, T. Peternell, *Group actions on* $\mathbf{S}^6$ *and complex structure on* $\mathbf{P}^3$, preprint math. AG/9812076 (1998)
[Kol] J. Kollár, *Flip and flop*, Proc. ICM Kyoto 1990, Springer Verlag (1991)
[Moi] B. Moishezon, *On* n*-dimensional compact varieties with* n *algebraically independent meromorphic functions*, AMS Translations 63 (1967), 51-177

Marco Brunella
CNRS UMR 5584
Laboratoire de Topologie
BP 47870, 9 Avenue Savary
21078 Dijon
France
brunella@u-bourgogne.fr

A Subshift of Finite Type in the Takens-Bogdanov Bifurcation with D_3 Symmetry

Karsten Matthies

Received: November 5, 1998
Revised: August 30, 1999

Communicated by Bernold Fiedler

Abstract. We study the versal unfolding of a vector field of codimension two, that has an algebraically double eigenvalue 0 in the linearisation of the origin and is equivariant under a representation of the symmetry group D_3. A subshift of finite type is encountered near a clover of homoclinic orbits. The subshift encodes the itinerary along the three different homoclinic orbits. In this subshift all those symbol sequences are realized for which consecutive symbols are different. In the parameter space we also locate a transcritical, three different Hopf and two global (homoclinic) bifurcations.

1991 Mathematics Subject Classification: 34C28, 58F14
Keywords and Phrases: Takens-Bogdanov, subshift of finite type, symmetry, homoclinic orbit

1 Introduction

A vector field has a Takens-Bogdanov point, if there is a Jordan block $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ in the linearisation of a steady state and if certain nondegeneracy conditions are fulfilled. This codimension two degeneracy with its unfolding is a key to understand several phenomena in dynamical systems (see [17, 3] and textbooks like [13]). Takens-Bogdanov points can also serve as a starting point for the path following in two-parameter flows of global Hopf bifurcation [6] and homoclinic orbits [7]. One parameter families of homoclinic orbits are created at Takens-Bogdanov points of two-parameter flows. Hence these bifurcation points play the same role for the creation of homoclinic orbits in two-parameter flows as Hopf bifurcation for periodic solutions in one-parameter flows. Near homoclinic orbits several bifurcations to other bounded solutions may occur. Thus Takens-Bogdanov points are important organizing centers for the bifurcation analysis of dynamical systems. Suppose the dynamical system $\dot{x} = f(x)$ has constraints given by an equivariance under a symmetry group Γ, i.e. $f(\gamma x) = \gamma f(x)$, for $\gamma \in \Gamma$. Then one often finds complicated bifurcation diagrams even at simple bifurcations, [11].

Similarly at the Takens-Bogdanov point with D_3 symmetry the dynamics are much richer than in the non-symmetric case. Applications are given to systems of three coupled oscillators in a ring. The results could be also applied to mode interactions for pattern formation in convection problems, where solutions with D_3 symmetry exist [12]. We will encounter a subshift of finite type, which is a novel dynamical feature in a bifurcation problem of dynamical systems defined by a vector field. Whereas in many bifurcations one can encounter Smale horseshoes giving rise to a full shift, the existence of a subshift of finite type is a rare phenomenon.

A subshift $\sigma(x_n)_{n\in\mathbf{Z}} = (x_{n+1})_{n\in\mathbf{Z}}$ of finite type on three symbols $\{1, 2, 3\}$ is defined on

$$X_A = \{(x_n)_{n\in\mathbf{Z}} | x_n \in \{1, 2, 3\}, a_{x_n,x_{n+1}} = 1\}$$

where $A = (a_{i,j})_{i,j\in\{1,2,3\}}$ is a 3×3 matrix with entries 0 and 1. The topology of X_A is defined as the product topology of the discrete set of symbols $\{1, 2, 3\}$. A subshift of finite type allows only those symbol sequences, for which consecutive symbols x_n, x_{n+1} are compatible with the transition matrix A. The symmetry group D_3 will act on X_A in the following manner

$$\begin{array}{llcl} \text{`flip':} & \kappa((x_n)_{n\in\mathbf{Z}}) & = & (\tilde{\kappa}x_n)_{n\in\mathbf{Z}} \text{ with } \tilde{\kappa}1 = 1, \tilde{\kappa}2 = 3, \tilde{\kappa}3 = 2 \\ \text{`rotation':} & \gamma((x_n)_{n\in\mathbf{Z}}) & = & (x_n + 1 \bmod 3)_{n\in\mathbf{Z}}. \end{array} \tag{1}$$

The bifurcation analysis will be reduced in section 2 to the discussion of a vector field on $\mathbf{R}^4 \cong \mathbf{C}^2$, where D_3 acts as

$$\begin{array}{llcl} \text{`flip':} & \kappa(v, w) & = & (\bar{v}, \bar{w}) \\ \text{`rotation':} & \gamma(v, w) & = & (\exp(i\frac{2\pi}{3})v, \exp(i\frac{2\pi}{3})w). \end{array} \tag{2}$$

A vector field in normal form can be derived. Using additional parameters (μ_1, μ_2) to unfold the singularity the normal form is generically given - up to time reversal - by

$$\begin{array}{lcl} \dot{v} & = & w \\ \dot{w} & = & \mu_1 v + \mu_2 w + \bar{v}^2 - \bar{v}\bar{w} + [A|v|^2 + B|w|^2 + C(v\bar{w} + \bar{v}w)]v + D|v|^2 w. \end{array} \tag{3}$$

A bifurcation diagram describing the complete plane of unfolding parameters is given in figure 1. The main result of this paper is formulated in theorem 1. It states the existence of a special form of a horseshoe for an open set of parameter

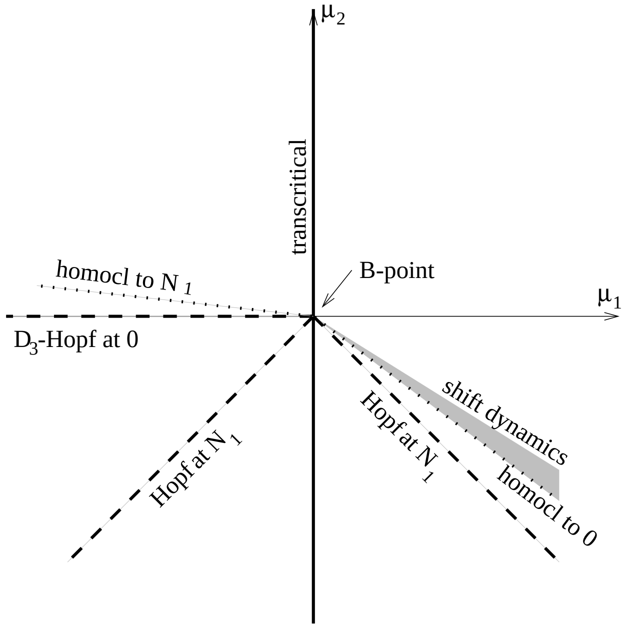


Figure 1: Bifurcation diagram in parameter space: transcritical bifurcation of steady states (bold line), three kind of Hopf bifurcation (dashed lines), two different homoclinic bifurcations (dotted lines) and shift dynamics (shaded area).

values. Later we will rigorously define three Poincaré sections $S_1^{in}, S_2^{in}, S_3^{in}$ as sections along three coexisting homoclinic orbits biasymptotic to the origin. P will be the return map on $S_1^{in} \cup S_2^{in} \cup S_3^{in}$.

Theorem 1 *For $0 < \mu_2 + \frac{6}{7}\mu_1$ small, $\mu_1 > 0$, there exists an invariant hyperbolic Cantor set $C \subset S_1^{in} \cup S_2^{in} \cup S_3^{in}$ such that the return map $P : C \to C$ induced by the flow of (3) is topological conjugate to the irreducible subshift of finite type with transition matrix $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. Here means topological conjugacy that there exists an homeomorphism $\tau : C \to X_A$ such that $\tau P = \sigma\tau$ on C.*
Furthermore P and τ can be chosen to be D_3-equivariant, when using the representation (2) on C and (1) on X_A. C is D_3-invariant.

So in fact we have a D_3-subshift of finite type as defined in [8]. Neither the incomplete bifurcation diagram in figure 1 nor theorem 1 depend on the coefficients A, B, C, D and other higher order terms. In the bifurcation diagram the lines will be bended to curves by a near-identity diffeomorphism. This is peculiar to the case of D_3 symmetry. When the system has some other symmetry

group like $O(2)$ [4] or D_4 [1] many more parameters and several different bifurcation diagrams have to be discussed. Thus our analysis can only be a first step to a general analysis of D_n equivariant Takens-Bogdanov singularities, where one might hope to encounter the $O(2)$ case as a limit.
The rest of the paper is organized as follows. In section 2 we give the Taylor expansion near the origin of a generic vector field equivariant under (2) at the Takens-Bogdanov point, derive a normal form up to third order and unfold it. We discuss the basic dynamical behavior in section 3, i.e. we analyze steady states, Hopf bifurcations and the dynamics in invariant subspaces including homoclinic orbits. The existence of the subshift will be proved in section 4, where we use a definition of a general horseshoe. In the last section 5 we will discuss some further numerical studies and applications.

2 D_3-equivariant vector fields and normal forms

Before giving a Taylor expansion near the singularity we first use some representation theory to justify the representation like in (2). There are in general two possibilities that a representation space of a compact Lie group Γ admits a non-diagonalizable Γ-equivariant linearisation A at the origin. Similar to chapter XVI of [11] there must be a Γ-invariant subspace W , that is either of the form $V \oplus V$, where V is absolutely irreducible, or that is irreducible but not absolutely irreducible. The second case is not possible for the Takens-Bogdanov singularity. The linearisation A contains the nilpotent matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Suppose W is irreducible but not absolutely irreducible, then $A(W) = 0$ or $A(W)$ is isomorphic to W [11, Lemma XII.3.4]. But if $A_{|W}$ contains the nilpotent Jordan block then $A : W \to A(W)$ cannot be an isomorphism. Hence $A_{|W} = 0$ and this is in contradiction that the Jordan block is non-zero. So we use a representation of the form $V \oplus V$. When choosing for V the standard representation of D_3 on $\mathbf{C} \cong \mathbf{R}^2$ we get the representation (2).

Proposition 2 *(i) The ring of all D_3-invariant germs acting on $\mathbf{C} \oplus \mathbf{C}$ as in (2) is generated by*

$$s_1 = v\bar{v},\ s_2 = w\bar{w},\ s_3 = v\bar{w} + \bar{v}w \text{ and}$$

$$t_j = v^j w^{3-j} + \bar{v}^j \bar{w}^{3-j}\ ,\ j \in \{0, \ldots, 3\}$$

(ii) The module of D_3-equivariant smooth mappings of $\mathbf{C} \oplus \mathbf{C} \to \mathbf{C} \oplus \mathbf{C}$ is generated by

$$g_0 = \begin{pmatrix} v \\ 0 \end{pmatrix},\ g_1 = \begin{pmatrix} 0 \\ v \end{pmatrix},\ g_2 = \begin{pmatrix} w \\ 0 \end{pmatrix},\ g_3 = \begin{pmatrix} 0 \\ w \end{pmatrix} \text{ and}$$

$$f_j = \begin{pmatrix} \bar{v}^j \bar{w}^{2-j} \\ 0 \end{pmatrix}, h_j = \begin{pmatrix} 0 \\ \bar{v}^j \bar{w}^{2-j} \end{pmatrix}\ ,\ j \in \{0, 1, 2\},$$

i.e. all D_3-equivariant smooth germs of mappings $h : \mathbf{R}^4 \to \mathbf{R}^4$ can be written in the form $h(v,\bar{v},w,\bar{w}) = p_0 g_0 + p_1 g_1 + p_2 g_2 + p_3 g_3 + q_0 f_0 + q_1 f_1 + q_2 f_2 + r_0 h_0 + r_1 h_1 + r_2 h_2$ where $p_0,\dots,p_3,q_0,q_1,q_2,r_0,r_1,r_2$ are smooth function germs of $s_1,s_2,s_3,t_0,\dots,t_3$.

Proof: For polynomials the completeness of the generators can be checked by lengthy by-hand calculations or by computer algebra, see [9]. These are then by Poénaru's theorem also the generators of the module of germs of mappings.□
Then a general D_3-Takens-Bogdanov point has the following Taylor expansion up to third order with real coefficients $a_1, b_1, \dots$:

$$\begin{array}{rclcl}
\dot{v} &=& w &+& a_1\bar{v}^2 + b_1\bar{v}\bar{w} + c_1\bar{w}^2 \\
&&&+& v(d_1 v\bar{v} + e_1 w\bar{w} + f_1(v\bar{w}+\bar{v}w)) \\
&&&+& w(g_1 v\bar{v} + h_1 w\bar{w} + i_1(v\bar{w}+\bar{v}w)) \\
\dot{w} &=& && a_2\bar{v}^2 + b_2\bar{v}\bar{w} + c_2\bar{w}^2 \\
&&&+& v(d_2 v\bar{v} + e_2 w\bar{w} + f_2(v\bar{w}+\bar{v}w)) \\
&&&+& w(g_2 v\bar{v} + h_2 w\bar{w} + i_2(v\bar{w}+\bar{v}w)).
\end{array} \tag{4}$$

First we try to remove as many second order terms as possible, therefore we choose a general near-identity D_3-equivariant coordinate change.

$$\begin{array}{rcl}
v &=& v' + \alpha_1\bar{v}^2 + \beta_1\bar{v}\bar{w} + \gamma_1\bar{w}^2 \\
w &=& w' + \alpha_2\bar{v}^2 + \beta_2\bar{v}\bar{w} + \gamma_2\bar{w}^2
\end{array}$$

We rewrite (4) in the new coordinates and this yields to

$$\begin{array}{rclcl}
\dot{v'} &=& w' &+& (a_1+\alpha_2)\bar{v}'^2 + (b_1+\beta_2-2\alpha_1)\bar{v}'\bar{w}' + (c_1+\gamma_2-\beta_1)\bar{w}'^2 \\
&&&+& v'(\tilde{d}_1 v'\bar{v}' + \tilde{e}_1 w'\bar{w}' + \tilde{f}_1(v'\bar{w}'+\bar{v}'w')) \\
&&&+& w'(\tilde{g}_1 v'\bar{v}' + \tilde{h}_1 w'\bar{w}' + \tilde{i}_1(v'\bar{w}'+\bar{v}'w')) \\
\dot{w'} &=& && a_2\bar{v}'^2 + (b_2-2\alpha_2)\bar{v}'\bar{w}' + (c_2-\beta_2)\bar{w}'^2 \\
&&&+& v'(\tilde{d}_2 v'\bar{v}' + \tilde{e}_2 w'\bar{w}' + \tilde{f}_2(v'\bar{w}'+\bar{v}'w')) \\
&&&+& w'(\tilde{g}_2 v'\bar{v}' + \tilde{h}_2 w'\bar{w}' + \tilde{i}_2(v'\bar{w}'+\bar{v}'w'))
\end{array} \tag{5}$$

where the ˜ terms depend on the original term, a_i, b_i, c_i and $\alpha_i, \beta_i, \gamma_i$ for $i = 1, 2$. By choosing

$$\alpha_1 = \frac{1}{2}(b_1+c_2), \alpha_2 = -a_1, \beta_1 = 0, \beta_2 = c_2, \gamma_1 = 0, \gamma_2 = -c_1$$

we can remove all second order terms in the first component and $\bar{w}'^2$ in the second component in (5). All the third order terms are $O(2)$-equivariant. Thus we can use exactly the same coordinate change as Dangelmayr and Knobloch [4] (after removing the second order terms) without affecting the lower order terms to get the following simplified system:

$$\begin{array}{rcl}
\dot{v} &=& w \\
\dot{w} &=& E\bar{v}^2 + F\bar{v}\bar{w} + \left[A|v|^2 + B|w|^2 + C(v\bar{w}+\bar{v}w)\right]v + D|v|^2 w,
\end{array}$$

where $E = a_2, F = b_2 + 2a_1, A = \tilde{d}_2, B = \tilde{d}_2 - \tilde{g}_1 + \tilde{f}_1 - 2\tilde{i}_2, C = \tilde{d}_1 + \tilde{f}_2$ and $D = \tilde{d}_1 + \tilde{g}_2$. We use an unfolding to describe the behavior of generic families of vector fields near the singular point. Even if there is not a general method to unfold the whole vector field, the linear part can be unfolded by $\begin{pmatrix} 0 & 0 \\ \mu_1 & \mu_2 \end{pmatrix}$ such that all nearby linear parts can be reached up to conjugation [2]. After scaling v, w, t and possibly a time-reversal we can set generically $E = 1, F = -1$, if the transformed second order terms are nonzero. Then we get the normal form as in equation (3):

$$\begin{aligned} \dot{v} &= w \\ \dot{w} &= \mu_1 v + \mu_2 w + \bar{v}^2 - \bar{v}\bar{w} + [A|v|^2 + B|w|^2 + C(v\bar{w} + \bar{v}w)]v + D|v|^2 w. \end{aligned}$$

3 Bifurcations

Standard computations show some symmetry breaking bifurcations. Here especially the behavior inside the flow-invariant fixed point space $\mathrm{Fix}(\kappa) = \{(v,w)|\kappa(v,w) = (v,w)\} = \{(v,w)|v,w \in \mathbf{R}\}$ will be considered. The same dynamics can be encountered in the rotated spaces $\gamma\mathrm{Fix}(\kappa)$ and $\gamma^2\mathrm{Fix}(\kappa)$. The bifurcation inside these planes is a Takens-Bogdanov bifurcation, in which the origin remains a singular point. This was analyzed by Hirschberg and Knobloch [14].

In general cubic and quintic terms cannot be neglected in Hopf bifurcation with D_3 symmetry. But in our situation the higher order terms are not important as long μ_1, μ_2 are small enough. To see this we have to perform the normal form calculations including these terms. The terms involving e.g. A, B, C, D are all of higher order in μ_1, μ_2 and hence can be neglected in a small neighborhood of 0 in the μ_1, μ_2 plane. For illustration we consider a Hopf bifurcation of the origin at $\mu_2 = 0, \mu_1 < 0$ inside $\mathrm{Fix}(\kappa)$. After calculating a normal form for Hopf bifurcation like in [13] the direction of branching is determined by the sign of the term $a = -\frac{1}{8|\mu_1|} + \frac{2C+D}{8}$. So the higher order terms can be neglected inside a neighborhood of $(0,0)$ in parameter space (μ_1, μ_2). Similar results hold for the other bifurcations too. We suppress therefore the dependence on these terms. They only bend some lines in the bifurcation diagram to curves by a near-identity diffeomorphism. See also figure 1.

- The only stable feature is the origin for $\mu_1, \mu_2 < 0$.
- For $\mu_1 = 0$ there is a transcritical bifurcation of secondary steady state $N_1 = (-\mu_1, 0)$ and the rotated points $N_2 = \gamma N_1$, $N_3 = \gamma^2 N_1$ each with isotropy $\mathbf{Z}_2$.
- For $\mu_2 = 0, \mu_1 < 0$ the spectrum of the origin is purely imaginary and the system undergoes a D_3-Hopf bifurcation [11], where three different types of periodic solutions appear (isotropy type $\tilde{\mathbf{Z}}_3$ for $\mu_2 < 0$; solutions of

isotropy type $\tilde{\mathbf{Z}}_2$ and inside Fix(κ) of isotropy type $\mathbf{Z}_2$ both for $\mu_2 > 0$; all these solutions are of saddle type).

- For $\mu_2 = \mu_1 < 0$ N_1, N_2, N_3 undergo Hopf bifurcations, where the imaginary eigenvalues have eigenvectors outside the invariant subspaces and the solutions have isotropy $\tilde{\mathbf{Z}}_2$.
- For $0 > \mu_2 = -\mu_1$ N_1, N_2, N_3 undergo Hopf bifurcations inside the invariant subspaces, i.e. the periodic orbit have isotropy $\mathbf{Z}_2$.
- For some curve with $\mu_2 \approx -\frac{6}{7}\mu_1, \mu_1 > 0$ there exists an orbit inside Fix(κ) homoclinic to 0, see [14].
- Similar there are orbits homoclinic to N_1, N_2, N_3 for $\mu_2 \approx -\frac{1}{7}\mu_1, \mu_1 < 0$.

For the homoclinic orbits we have even nearly explicit expressions. Scaling the equation (3)

$$\tau = \epsilon t, v = \epsilon^2 x, w = \epsilon^3 y, \mu_1 = \epsilon^2 \nu_1, \mu_2 = \epsilon^2 \nu_2.$$

and $\dot{} = \frac{d}{d\tau}$ give

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= \nu_1 x + \bar{x}^2 + \epsilon(\nu_2 y - \bar{x}\bar{y}) + O(\epsilon^2). \end{aligned} \tag{6}$$

Letting $\epsilon = 0$ the system has an explicit homoclinic orbit for $\nu_1 > 0$ inside Fix(κ):

$$q_0(t) = \begin{pmatrix} x_q(t) \\ y_q(t) \end{pmatrix} = \begin{pmatrix} -\frac{3\nu_1}{2}\left(1 - \tanh^2\left(\frac{\sqrt{\nu_1}}{2}t\right)\right) \\ \frac{3\nu_1}{2}\sqrt{\nu_1}\operatorname{sech}^2\left(\frac{\sqrt{\nu_1}}{2}t\right)\tanh\left(\frac{\sqrt{\nu_1}}{2}t\right) \end{pmatrix}.$$

Using the Melnikov method, see e.g. Guckenheimer and Holmes [13], we can then compute parameter values for which the homoclinic orbit persists for $\epsilon > 0$ to get the above results.
By symmetry there are homoclinic orbits biasymptotic to the origin inside the other two invariant fixed point spaces γFix(κ) and γ^2Fix(κ) for the same parameter values too. So there exists a 'clover' like structure of homoclinic orbits, see figure 2.

4 General Horseshoes and Proof of Theorem 1

In this section we prove the existence of the subshift of finite type near the clover of homoclinic orbits. We will compute a Poincaré map near the homoclinic orbits with varying unfolding parameters μ_1 and μ_2. For each of the three homoclinic orbits we define an 'in' and an 'out' section, called S_i^{in} and S_i^{out} (figure 3). The return map $P : S_1^{in} \cup S_2^{in} \cup S_3^{in} \to S_1^{in} \cup S_2^{in} \cup S_3^{in}$ is discussed by dividing it into local parts near the steady state, which can be described by

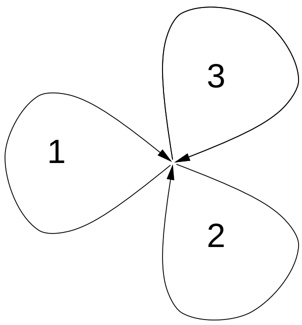


Figure 2: A sketch of the clover of homoclinic orbits. The three orbits lie all in different planes, which intersect only in the origin.

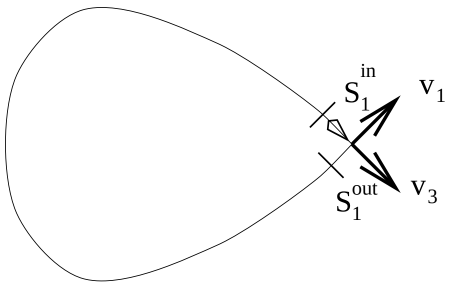


Figure 3: The sections S_1^{out} and S_1^{in} at the homoclinic orbit projected to $\mathrm{Fix}(\kappa)$.

its linearisation (lemma 4) and global parts along the homoclinic orbit. This technique can also be used to analyze several other homoclinic bifurcations, see for example the textbook [10].

Before we give the technical details of the analysis of P, we describe the geometric idea: The sections S_i^{in} and S_i^{out} are cubes in $\mathbf{R}^3$. We identify those regions in S_i^{out}, which have a preimage in some S_i^{in} under the local maps (see figure 4). Similarly we compute the regions in S_i^{in}, which are mapped by the local maps to some S_i^{out} (see figure 5). The global map P will map the cube in figure 4 to the cube in figure 5.

For appropriately chosen parameters (μ_1, μ_2) the slabs marked '2' and '3' in figure 4 will intersect the slabs '2' and '3' in figure 5. We can then show that there is a Smale horseshoe in three dimensions in the upper half of the cube. But because of the symmetry we have three copies of these cubes and the possible itineraries inside the invariant set are more complicated. In the figures 4 and 5 the sections of the homoclinic orbit marked '1' in figure 2 are shown. The trajectories of points in the regions '2' and '3' in figure 4 were in the sections S_2^{in} and S_3^{in} before. In the same way the slabs '2' and '3' in figure 5 are those regions, where the forward orbit will reach the section S_2^{out} and

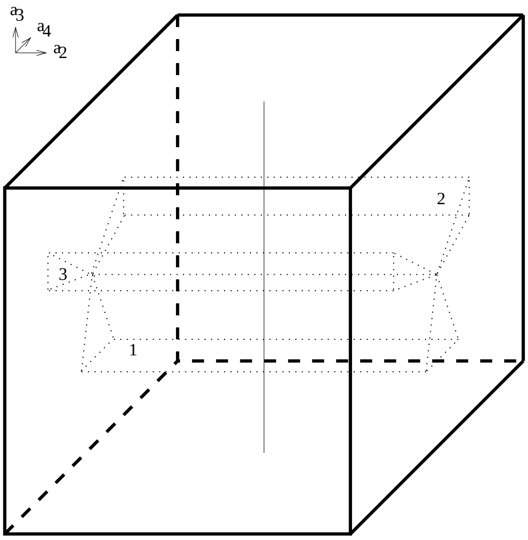


Figure 4: The section S_1^{out} with the images of S_1^{in}, S_2^{in} and S_3^{in}. The line in the middle is the section with $\mathrm{Fix}(\kappa)$.

S_3^{out} next. Hence the itineraries, described in the figures 4 and 5, have first a symbol '2' or '3' then the symbol '1', because they are now at the homoclinic orbit with symbol '1', and then proceed with '2' or '3'. At the other sections there is the same behavior after following once along the homoclinic loop: The trajectories of points inside the invariant set will lead to another section and hence to another symbol. Therefore the subshift described in theorem 1 can be realized, but no other infinite symbol sequences.
To rigorously prove the existence of the subshift, we describe briefly the notion of a general horseshoe in $\mathbf{R}^3$ following Katok and Hasselblatt [15]. First we explain the meaning of 'full intersection'. Then using cone conditions we give precise meaning to 'horizontal expansion' and 'vertical contraction'. We prove a technical lemma to justify the complete linearisation near the steady state before computing the local and global maps.
We will consider a *rectangle* $\Delta = D_1 \times D_2 \subset \mathbf{R} \oplus \mathbf{R}^2 = \mathbf{R}^3$ where D_1 and D_2 are discs. The projections on the components are denoted by π_1 ("horizontal") and π_2 ("vertical"). Let $\Delta \subset U \subset \mathbf{R}^3$ be a rectangle and $f : U \to \mathbf{R}^3$ be a diffeomorphism. Then we call a connected component $S' = fS \subset \Delta \cap f\Delta$ *full*, if

1. $\pi_2(S) = D_2$,
2. for all $z \in S, \pi_{1|f(S\cap(D_1\times\pi_2(z)))}$ is a bijection onto D_1.

The first condition implies that S reaches completely along the vertical direction and second one that the image of every horizontal fiber in S meets Δ and

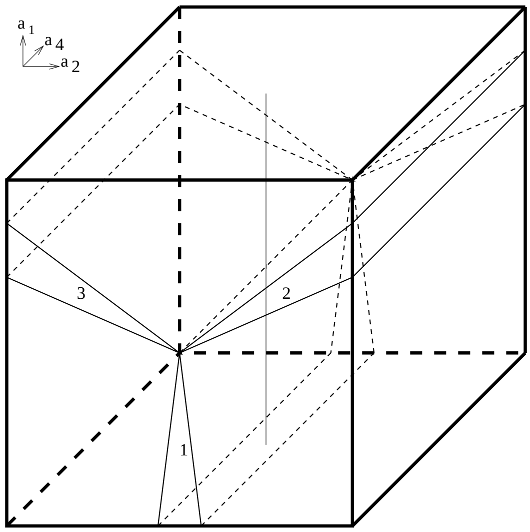


Figure 5: The section S_1^{in} where the preimages of S_1^{out}, S_2^{out} and S_3^{out} are the dotted slabs.

traverses it completely.

Next we introduce cone conditions. A horizontal s-cone H_x^s is defined by $H_x^s = \{(u,v) \in T_x\mathbf{R}^3 | \|v\| \leq s\|u\|\}$, similarly a vertical s-cone V_x^s by $V_x^s = \{(u,v) \in T_x\mathbf{R}^3 | \|u\| \leq s\|v\|\}$ at $x \in \mathbf{R}^3$ for some s. A map f preserves a family H_x of horizontal cones for $x \in U \subset \mathbf{R}^3$, if $Df_x(H_x) \subset \text{int}(H_{f(x)}) \cup \{0\}$. It is called expanding on a horizontal cone family H_x, if $\|Df_x\xi\| \geq \mu\|\xi\|$ for $\xi \in H_x$ and some fixed $\mu > 1$. We want to express a contraction property in the vertical direction, thus we consider f^{-1} on vertical cone families. It preserves the vertical cone family V_x, if $Df_x^{-1}(V_{f(x)}) \subset \text{int}(V_x) \cup \{0\}$ and f^{-1} expands them, if $\|Df_x^{-1}\xi\| \geq \lambda^{-1}\|\xi\|$ for $\xi \in V_{f(x)}$ and some uniform $\lambda < 1$. Then the appropriate generalization of a Smale horseshoe in higher space dimensions is given by

Definition 3 *[15] Let $\Delta \subset U \subset \mathbf{R}^3$ be a rectangle and $f : U \to \mathbf{R}^3$ be a diffeomorphism. $\Delta \cap f(\Delta)$ is called a horseshoe if it contains at least two full components Δ_1 and Δ_2 such that for $\Delta' = \Delta_1 \cup \Delta_2$ the following conditions hold:*

1. *$\pi_2(\Delta') \subset int(D_2)$ and $\pi_1(f^{-1}(\Delta')) \subset int(D_1)$,*
2. *$D(f_{|f^{-1}(\Delta')})$ preserves and expands a horizontal cone family on $f^{-1}(\Delta')$,*
3. *$D(f^{-1}_{|\Delta'})$ preserves and expands a vertical cone family on Δ'.*

To compute the return map P we will first prove that we can completely linearize the local maps.

Lemma 4 *Suppose that the distinct eigenvalues of the linearisation* A *at* 0 $\lambda_{1,2} = \frac{\mu_2}{2} \pm \sqrt{\frac{\mu_2^2}{4} + \mu_1}$ *are not in resonance, i.e.* $\lambda_i - (k\lambda_1 + l\lambda_2) \neq 0$ *for* $k, l \in \mathbf{N}, k + l > 1$*. Then there exists a* D_3*-equivariant smooth diffeomorphism* H *conjugating the flow* Φ_t *of (3) and* $\exp(At)$ *on some neighborhood* U *of the origin:* $H\Phi_t = \exp(At)H$*.*

Proof: We consider first the time-one-map Φ_1, again the linear part is diagonal with eigenvalues $e^{\lambda_1}, e^{\lambda_2}$. For these the non-resonance conditions for maps $\exp\lambda_i \neq \exp(k\lambda_1)\cdot\exp(l\lambda_2)$ for $k, l \in \mathbf{N}, k + l > 1$ hold. The non-resonance conditions imply that we can formally remove all terms of algebraic order by a near-identity coordinate change. This is possible even in a D_3-equivariant setting [11]. So we still have to remove flat terms and discuss convergence. To remove these flat terms we use a version of Sternberg's theorem [15, theorem 6.6.7]. The assumptions are fulfilled: The linear part is diagonal and the normal form which can be achieved by the above coordinate change is a convergent power series, since it is only linear. The theorem then gives the existence of a smooth diffeomorphism conjugating Φ_1 and its normal form. Thus there exists a smooth diffeomorphism H_1 linearizing Φ_1 in a neighborhood of the origin. Furthermore the construction in [15] can be chosen to preserve D_3-equivariance, when we use invariant cut-off functions. Then the D_3-equivariant diffeomorphism $H = \int_0^1 \exp(-At)H_1\Phi_t dt$ is the needed conjugacy for the entire flow on some neighborhood U of 0. This can be seen when using $\exp(-A)H_1\Phi_1 = H_1$

$$\begin{aligned}
&\exp(-As)H\Phi_s = \textstyle\int_0^1 \exp(-A(t+s))H_1\Phi_{t+s}dt = \int_s^{s+1}\exp(-Au)H_1\Phi_u du\\
&= H - \textstyle\int_0^s \exp(-Au)H_1\Phi_u du + \int_1^{s+1}\exp(-Au)H_1\Phi_u du\\
&= H - \textstyle\int_0^s \exp(-A(u+1))H_1\Phi_{u+1}du + \int_1^{s+1}\exp(-Au)H_1\Phi_u du\\
&= H\square
\end{aligned}$$

Now we can compute the map P. After the coordinate change of the lemma the local maps are given by a linear flow. Then the stable and unstable manifolds coincide with the stable and unstable eigenspaces. To carry out the analysis we use again the scaled coordinates $x = x_1 + ix_2, y = y_1 + iy_2 \in \mathbf{C}$ for some $\epsilon > 0$ small. We know the homoclinic orbits explicitly by section 3 up to perturbations of order $\mathrm{O}(\epsilon)$. While neglecting terms of order $\mathrm{O}(\epsilon^2)$ the system in $\mathbf{C}^2$ is given by equation (6).

Local maps: To compute the local maps we use a basis of eigenvectors of the linearized system: For the eigenvalue $\lambda_1 = \frac{\epsilon\nu_2}{2} + \sqrt{\frac{\epsilon^2\nu_2^2}{4} + \nu_1} > 0$ we choose v_1, v_2 and for the eigenvalue $\lambda_2 = \frac{\epsilon\nu_2}{2} - \sqrt{\frac{\epsilon^2\nu_2^2}{4} + \nu_1} < 0$ the vectors v_3, v_4. The original basis of $\mathbf{R}^4 \cong \mathbf{C}^2$ is given by (x_1, y_1, x_2, y_2).

$$\begin{aligned}
v_1 &= (1+\lambda_1^2)^{-\frac{1}{2}}(1, \lambda_1, 0, 0)^T, \quad v_2 = (1+\lambda_1^2)^{-\frac{1}{2}}(0, 0, 1, \lambda_1)^T,\\
v_3 &= (1+\lambda_2^2)^{-\frac{1}{2}}(1, \lambda_2, 0, 0)^T, \quad v_4 = (1+\lambda_2^2)^{-\frac{1}{2}}(0, 0, 1, \lambda_2)^T.
\end{aligned}$$

A vector $a \in \mathbf{R}^4$ is then denoted as $a = a_1v_1 + a_2v_2 + a_3v_3 + a_4v_4$. The eigenvectors v_1 and v_3 span $\mathrm{Fix}(\kappa)$. The section S_1^{out} is then defined by

$$a_1 = -c \text{ and } \max\{|a_2|, |a_3|, |a_4|\} < \delta$$

with c small and $0 < \delta \ll c$ such that the section is completely inside U, where the flow is linearized. S_1^{in} is given by

$$a_3 = -c \text{ and } \max\{|a_1|, |a_2|, |a_4|\} < \delta.$$

We will also use rotated coordinate systems with basis vectors $v_l' = \gamma v_l$ and $v_l'' = \gamma^2 v_l$ with coefficients a_l', a_l''. Thus we can define the sections of the rotated homoclinic orbits.

$$\begin{array}{ll}
S_2^{out}: & a_1' = -c, \max\{|a_2'|, |a_3'|, |a_4'|\} < \delta \\
S_3^{out}: & a_1'' = -c, \max\{|a_2''|, |a_3''|, |a_4''|\} < \delta \\
S_2^{in}: & a_3' = -c, \max\{|a_1'|, |a_2'|, |a_4'|\} < \delta \\
S_3^{in}: & a_3'' = -c, \max\{|a_1''|, |a_2''|, |a_4''|\} < \delta
\end{array}$$

First we compute $P_l^{loc}, l \in \{1,2,3\}$. The flow of the linear system is given by

$$\Phi_t(a) = a_1 v_1 e^{\lambda_1 t} + a_2 v_2 e^{\lambda_1 t} + a_3 v_3 e^{\lambda_2 t} + a_4 v_4 e^{\lambda_2 t}, \tag{7}$$

similarly in the primed versions for the rotated coordinate systems. Starting at a vector $a \in S_1^{in} \cup S_2^{in} \cup S_3^{in}$ with $P_l^{loc}(a) \in S_1^{out}$ (i.e. especially $a_1 < 0$), the time $t = (\ln \frac{c}{|a_1|})/\lambda_1$ is needed to reach the S_1^{out} section. Then $P_l^{loc}(a_1 v_1 + a_2 v_2 + a_3 v_3 + a_4 v_4)$

$$= (-c v_1 + a_2 \left|\frac{c}{a_1}\right| v_2 + a_3 \left|\frac{a_1}{c}\right|^{\frac{|\lambda_2|}{\lambda_1}} v_3 + a_4 \left|\frac{a_1}{c}\right|^{\frac{|\lambda_2|}{\lambda_1}} v_4) \tag{8}$$

with $\frac{|\lambda_2|}{\lambda_1} = 1 + \frac{18}{49}\epsilon^2\nu_1 + \frac{6}{7}\epsilon\sqrt{\frac{9}{49}\epsilon^2\nu_1^2 + \nu_1} + O(\epsilon^3)$.
To understand the geometry of the local maps we compute how the preimage of the 'out'-sections $S_l^{out}, l \in \{1,2,3\}$ intersects the 'in'-sections $S_l^{in}, l \in \{1,2,3\}$ and how the images of S_l^{in} intersect the 'out'-sections S_l^{out}. We start with the preimage of S_1^{out} intersected with S_1^{in}

$$\begin{array}{ll}
S_{1,1}^{in} & \\
= & S_1^{in} \cap {P_1^{loc}}^{-1}(S_1^{out}) \\
= & \{(a_1,a_2,a_3,a_4) | a_3 = -c, \max\{|a_1|,|a_2|,|a_4|\} < \delta\} \\
\cap & \{(a_1,a_2,a_3,a_4) | a_1 < 0, \max\{|a_2| \left|\frac{c}{a_1}\right|, |a_3| \left|\frac{a_1}{c}\right|^{\frac{|\lambda_2|}{\lambda_1}}, |a_4| \left|\frac{a_1}{c}\right|^{\frac{|\lambda_2|}{\lambda_1}}\} < \delta\} \\
= & \{(a_1,a_2,a_3,a_4) | -\delta < a_1 < 0, |a_2| < \delta \left|\frac{a_1}{c}\right|, a_3 = -c, |a_4| < \delta\}.
\end{array}$$

This is the slab with label '1' infigure 5. Then the image of S_1^{in} inside S_1^{out} is given by $S_1^{out} \cap P_1^{loc}(S_1^{in}) = P_1^{loc}(S_{1,1}^{in})$

$$= \{(a_1,a_2,a_3,a_4) | a_1 = -c, |a_2| < \delta, -\delta^{\frac{|\lambda_2|}{\lambda_1}} < a_3 < 0, |a_4| < \frac{\delta}{c}|a_3|\}.$$

This set is the slab with label '1' in figure 4. To determine the images $P_2^{loc}(S_2^{in}) \cap S_1^{out}$ and $P_3^{loc}(S_3^{in}) \cap S_1^{out}$ we have just to rotate a part of the coordinate system. Inside the stable eigenspace (v_3, v_4) is changed to (v_3', v_4') and (v_3'', v_4'') respectively. Equation (7) holds for each eigenspace independently. Thus the restrictions are essentially the same as for $S_1^{out} \cap P_1^{loc}(S_1^{in})$ just with a_3', a_4' and a_3'', a_4'' instead of a_3, a_4. Hence the slab $S_1^{out} \cap P_1^{loc}(S_1^{in})$ has just to be rotated by $2\pi/3$ and $4\pi/3$ inside the (v_3, v_4) plane to get $S_1^{out} \cap P_2^{loc}(S_2^{in})$ and $S_1^{out} \cap P_3^{loc}(S_3^{in})$. A sketch of section S_1^{out} with the images of $S_l^{in}, l \in \{1,2,3\}$ is given in figure 4.

Next we will compute the preimage of S_2^{out} and S_3^{out} under P_1^{loc} to get the structure of S_1^{in}. When we use a rotated coordinate system (v_1', v_2') instead of (v_1, v_2) inside the unstable eigenspace, the time $t = (\ln \frac{c}{|a_1'|})/\lambda_1$ is needed to reach S_2^{out}. This yields to

$$\begin{aligned} & P_1^{loc}(a_1'v_1' + a_2'v_2' - cv_3 + a_4v_4) \\ = \quad & (-cv_1' + a_2'\left|\frac{c}{a_1'}\right| v_2' - c\left|\frac{a_1'}{c}\right|^{\frac{|\lambda_2|}{\lambda_1}} v_3 + a_4\left|\frac{a_1'}{c}\right|^{\frac{|\lambda_2|}{\lambda_1}} v_4). \end{aligned}$$

So the preimage of S_2^{out} under P_1^{loc} is just $S_{1,1}^{in}$ rotated by $2\pi/3$ inside the unstable eigenspace. And finally for the preimage of S_3^{out} the coordinate system has to be rotated by $4\pi/3$ in the unstable eigenspace. The section S_1^{in} with the preimages of $S_l^{out}, l \in \{1,2,3\}$ is drawn in figure 5.

Global maps: Next we approximate $P_l^{glo} : S_l^{out} \to S_l^{in}$ by an Taylor expansion using the linearisation along the homoclinic orbit. This approximation is valid by a general perturbation argument for hyperbolic sets, when we choose the size of the cubes δ small enough. We get a constant term of the global map when considering the splitting of the homoclinic orbit. The point $(-c, 0, 0, 0) \in S_1^{out}$ is inside $\mathrm{Fix}(\kappa)$, hence it will be mapped to $S_1^{in} \cap \mathrm{Fix}(\kappa)$. Thus the constant term is the distance of the stable and unstable manifolds inside $\mathrm{Fix}(\kappa)$. Using [13, Eq.(4.5.11)] this distance is given by $d(\nu_2, \epsilon) = \frac{\epsilon M(\nu_2)}{\|f(q)\|} + O(\epsilon^2)$, with Melnikov functional $M(\nu_2)$ and vector field f on $\mathrm{Fix}(\kappa)$. For our system this is $d(\nu_2) = \epsilon \frac{4}{5c}\sqrt{\nu_1}(\nu_2 + \frac{6}{7}\nu_1)$.

In (x_1, y_1, x_2, y_2) coordinates the linearisation along the homoclinic solution for $\epsilon > 0$ is given by $B = D_{(x,y)}f_{|(x(t),y(t))} =$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ \nu_1 + 2x_1(t) - \epsilon y_1(t) & \epsilon(\nu_2 - x_1(t)) & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \nu_1 - 2x_1(t) + \epsilon y_1(t) & \epsilon(\nu_2 + x_1(t)) \end{pmatrix}, \tag{9}$$

where $x_1(t), y_1(t)$ are the non-zero components of the homoclinic orbit. This means that we have to solve the non-autonomous linear differential equation $\dot{\xi} = B\xi$. We use the block diagonal structure of the matrix. The first block describes the behavior inside the invariant subspace $\mathrm{Fix}(\kappa)$ and the second block the orthogonal complement $\mathrm{Fix}(\kappa)^{\perp}$.

In the first block we are interested in the initial values $\xi_0 = a_3 v_3$ inside S_1^{out}. One solution of the variational equation inside $\mathrm{Fix}(\kappa)$ is given by $\dot{q}_0(t)$ for $\epsilon = 0$. Letting $q_0(0) \in S_1^{out}$ and $q_0(T) \in S_1^{in}$, then $\|q_0(0)\| = \|q_0(T)\|$ by symmetry. The vectors $\xi_0, \dot{q}_0(0)$ restricted to $\mathrm{Fix}(\kappa)$ are a fundamental system. The Wronskian W of this system is constant by Liouville's theorem: $\dot{W} = \mathrm{trace}(B_{|\mathrm{Fix}(\kappa)})W = 0$. Therefore, as $\dot{q}_0(0) = kv_1$ and $\dot{q}_0(T) = kv_3$, the projection of $\xi_0(T)$ onto v_1 is a_3. By smooth dependence on parameters this yields to $P_1^{glo}(a_3 v_3) = (1 + O(\epsilon))a_3 v_1$.
In the second block we consider initial values $\xi_1 = a_2 v_2$ and $\xi_2 = a_4 v_4$. First assume that $\epsilon = 0$. As $(\nu_1 - 2x_1(t)) > \nu_1 > 0$ and $\xi_1^{(1)}(0), \xi_1^{(2)}(0) > 0$ hold, the two components $\xi_1^{(1)}(t)$ and $\xi_1^{(2)}(t)$ are increasing. The global map also expands this vector for $\epsilon > 0$ by the smooth dependence on the parameter ϵ for finite time. Hence in linear approximation we get $P_1^{glo}(a_2 v_2) = a_2(\alpha_1 v_2 + \alpha_2 v_4)$ with $\alpha_1^2 + \alpha_2^2 \geq 1$. Furthermore

$$\alpha_1 \geq 0.9|\alpha_2| \tag{10}$$

holds because the coefficients of the solution are positive in the x_2, y_2 coordinates. Applying Liouville's theorem again for $\epsilon = 0$, the second initial vector is mapped to $P_1^{glo}(a_4 v_4) = a_4(\beta_1 v_2 + \beta_2 v_4)$ with $\alpha_1\beta_2 - \alpha_2\beta_1 = 1$. Again $\epsilon > 0$ will give perturbations of type $1 + \mathrm{O}(\epsilon)$, which we will suppress by still using the same notation.
Full Map: We now consider only those points which are mapped under the local maps from $S_1^{in} \cup S_2^{in} \cup S_3^{in}$ to S_1^{out}. When we use (v_1, v_2, v_3, v_4) as a coordinate system for all three 'in'-sections then the composed mapping is given by

$$\begin{array}{rcl} P_1^{glo} \circ P_l^{loc} : S_l^{in} & \to & S_1^{in}, l \in \{1,2,3\} \\ \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} & \mapsto & \begin{pmatrix} \frac{4\epsilon}{5c}\sqrt{\nu_1}(\frac{6}{7}\nu_1 + \nu_2) + (1+O(\epsilon))a_3 \left|\frac{a_1}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \\ \alpha_1 a_2 |\frac{c}{a_1}| + \beta_1 a_4 \left|\frac{a_1}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \\ -c \\ \alpha_2 a_2 |\frac{c}{a_1}| + \beta_2 a_4 \left|\frac{a_1}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \end{pmatrix} \end{array} \tag{11}$$

Now we can use this to determine the return map $P : S_1^{in} \cup S_2^{in} \cup S_3^{in} \to S_1^{in} \cup S_2^{in} \cup S_3^{in}$, where it is defined. Because of the symmetry the maps $P_2^{glo} \circ P_l^{loc}$ and $P_3^{glo} \circ P_l^{loc}$ are related to (11) by simple rotations of whole $\mathbf{R}^4$. When changing to the rotated coordinates, the maps $P_2^{glo} \circ P_l^{loc}$ and $P_3^{glo} \circ P_l^{loc}$ are given by equation (11) with a_i replaced by a_i' and a_i'' . Therefore it is enough to consider a reduced map $\tilde{P}$ just as a map from one section S^{in} to itself. We just have to change the original labels '1', '2' and '3' in the S_2^{in} and S_3^{in} sections. We will use a labeling relative to our position and call our position '1', the next homoclinic orbit in the direction of the rotation is called '2' and the other one '3'.
Proof of theorem 1: The existence of a horseshoe for this reduced map $\tilde{P}$ will be shown. Analyzing the implications for the full map will prove the

theorem. The conditions of definition 3 will be checked for the map:

$$\begin{array}{ccc} \tilde{P}: S^{in} & \rightarrow & S^{in} \\ \begin{pmatrix} a_1 \\ a_2 \\ a_4 \end{pmatrix} & \mapsto & \begin{pmatrix} \frac{4\epsilon}{5c}\sqrt{\nu_1}(\frac{6}{7}\nu_1+\nu_2)+(1+O(\epsilon))|a_1|\left|\frac{a_1}{c}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} \\ \alpha_1 a_2|\frac{c}{a_1}|+\beta_1 a_4\left|\frac{a_1}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \\ \alpha_2 a_2|\frac{c}{a_1}|+\beta_2 a_4\left|\frac{a_1}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \end{pmatrix} \end{array}. \quad (12)$$

The horizontal direction is $\alpha_1 v_2 + \alpha_2 v_4$ and the vertical directions are v_1 and v_4. Using these as a new basis with coefficients ζ_1, ζ_2 and ζ_3 we define Δ as the product of discs with radii 2δ in ζ_1 for D_1 and δ in (ζ_2, ζ_3) for D_2 with the further restriction $\frac{\delta}{6} < \zeta_2 < \frac{\delta}{3}$. This means we choose a coordinate system such that we can ignore any rotation of figure 4 under the global mapping to figure 5, even if $|\alpha_2|$ is not small. This can be done, because we estimated $\alpha_1 \geq 0.9|\alpha_2|$ in (10). We just have to change the labels from a_2 to ζ_1, a_1 to ζ_2 and a_4 to ζ_3. As above we denote the rotated coordinates by ζ_i' and ζ_i''. The rectangle is given in figure 6. We choose the distance of splitting $d = \frac{\delta}{6}$. The two full components Δ_1 and Δ_2, which have to be contained in $\Delta \cap \tilde{P}(\Delta)$, are the two top dotted slabs in figure 6.

We consider the preimages of these two slabs under the original return map P, i.e. we are interested in the preimages of $\Delta_1, \Delta_2 \subset S_1^{in}$ under $P_1^{glo} \circ P_{2,3}^{loc}$. Then we get $(P_1^{glo} \circ P_2^{loc})^{-1}(\Delta_1) =$

$$\tilde{\Gamma}_1 = \{(\zeta_1, \zeta_2, \zeta_3') \in \gamma \cdot \Delta | 0 < -\zeta_2 < 2\delta, |\zeta_1| \leq \delta \left|\frac{\zeta_2}{c\alpha_1}\right|\} \subset \gamma\Delta \subset S_2^{in}$$

and similarly $(P_1^{glo} \circ P_3^{loc})^{-1}(\Delta_2) =$

$$\tilde{\Gamma}_2 = \{(\zeta_1, \zeta_2, \zeta_3'') \in \gamma^2 \cdot \Delta | 0 < -\zeta_2 < 2\delta, |\zeta_1| \leq \delta \left|\frac{\zeta_2}{c\alpha_1}\right|\} \subset \gamma^2\Delta \subset S_3^{in}.$$

As we identified the three sections in this analysis of $\tilde{P}$, we deal with Γ_1 and Γ_2, which are contained in the slabs with labels 2 and 3 in figure 6. The further restrictions are due to the possible additional expanding of the global map, i.e. the slabs are defined by

$$\Gamma_1 = \gamma^{-1}\tilde{\Gamma}_1 = \{(\zeta_1'', \zeta_2'', \zeta_3) \in \Delta | 0 < -\zeta_2'' < 2\delta, |\zeta_1''| \leq \delta \left|\frac{\zeta_2''}{c\alpha_1}\right|\}. \quad (13)$$

$$\Gamma_2 = \gamma^{-2}\tilde{\Gamma}_2 = \{(\zeta_1', \zeta_2', \zeta_3) \in \Delta | 0 < -\zeta_2' < 2\delta, |\zeta_1'| \leq \delta \left|\frac{\zeta_2'}{c\alpha_1}\right|\}, \quad (14)$$

After relabeling we have $\Delta_1 = \tilde{P}(\Gamma_1)$ and $\Delta_2 = \tilde{P}(\Gamma_2)$: The slab Γ_1 is mapped by P_1^{loc} to S_3^{out} and then by P_3^{glo}, because of our relabeling it will be the slab coming from S_2^{out}, hence it is the dotted slab with label 2 and therefore $\Delta_1 = \tilde{P}(\Gamma_1)$. In the same manner we get $\Delta_2 = \tilde{P}(\Gamma_2)$.

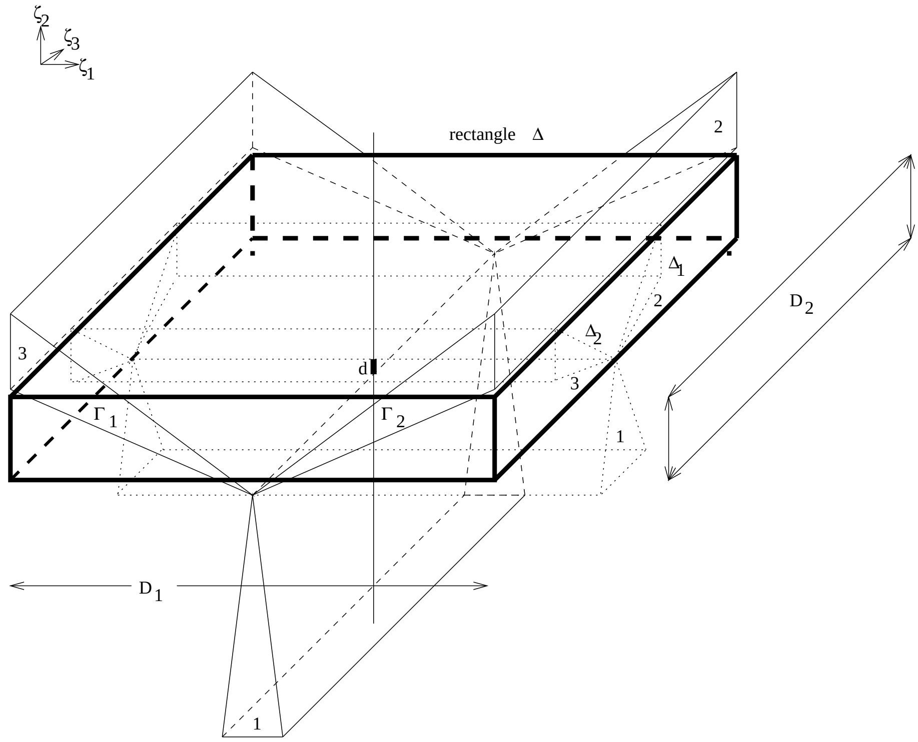


Figure 6: The section S^{in} with rectangle $\Delta = D_1 \times D_2$ including $\Delta_{1,2} = \tilde{P}(\Gamma_{1,2})$

So we can now check the conditions in the definition of the horseshoe. The two components $\Delta_1 = P(\Gamma_1)$ and $\Delta_2 = P(\Gamma_2)$ are full: $\pi_2(\Gamma_i) = D_2$ for $i = 1, 2$, because we can choose ζ_3 freely and $\zeta_2' < 0$ (respective ζ_2'') freely with $|\zeta_1'|$ (respective ζ_1'') small inside Δ, i.e. we get all wanted $\zeta_2 > 0$ in the definition of Γ_i. For all $\zeta \in \Gamma_i$ the restriction $\pi_{1|f(\Gamma_i \cap (D_1 \times \pi_2(\zeta)))}$ is a bijection onto D_1. When we vary ζ_1 for any given $z = (a_1, \alpha_1\zeta_1, \alpha_2\zeta_1 + a_4) \in \Gamma_i$ then P is affine linear (see (12)) and the projection π_1 to the ζ_1 component is injective, which is the a_2 component in P. It is also surjective onto D_1, because the restrictions on a_2 inside Γ_i ((14) and (13)) were given such that the maximal modulus of the a_2 component is δ in the image.
Next we check the first condition in definition 3. $\pi_2(\Delta') \subset \mathrm{int}(D_2)$ holds because of the contraction in the $a_1 = \zeta_2$ and ζ_3 component when choosing $\frac{6}{7}\nu_1 + \nu_2$ small enough. The ζ_3 component is given by

$$\left| (\alpha_1\zeta_3 - \alpha_2\zeta_1) \left| \frac{\zeta_2}{c} \right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \right| \ll \delta.$$

The other condition $\pi_1(P^{-1}\Delta') \subset \mathrm{int}(D_1)$ also holds, because $|\zeta_1| \leq |\zeta_1''|/2 +$

$\sqrt{3}|\zeta_2''|/2$ and ζ_1', ζ_2' and ζ_1'', ζ_2'' are small enough by the definition of Γ_1 and Γ_2 (see (14),(13)).
Finally we have to check the cone conditions, for which we need the linearisations of $\tilde{P}$ and $\tilde{P}^{-1}$. Suppressing all factors $1+O(\epsilon)$ these are given in the original (a_1, a_2, a_4) coordinates by $D\tilde{P}_x(a_1, a_2, a_4) =$

$$\begin{pmatrix} -\left|\frac{a_1}{c}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} & 0 & 0 \\ \alpha_1 a_2 \frac{c}{a_1^2} - \beta_1 a_4 \left|\frac{a_1}{c}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} & \alpha_1 \left|\frac{c}{a_1}\right| & \beta \left|\frac{a_1}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \\ \alpha_2 a_2 \frac{c}{a_1^2} - \beta_2 a_4 \left|\frac{a_1}{c}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} & \alpha_2 \left|\frac{c}{a_1}\right| & \beta_2 \left|\frac{a_1}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \end{pmatrix} \tag{15}$$

and if $(a_1, a_2, a_4) = \tilde{P}^{-1}(z)$ then $D\tilde{P}_x^{-1}(z) = (D\tilde{P}_x(a_1, a_2, a_4))^{-1}$ is

$$\begin{pmatrix} -\left|\frac{c}{a_1}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} & 0 & 0 \\ -\frac{a_2}{a_1}\left|\frac{c}{a_1}\right|^{1-\frac{6}{7}\epsilon\sqrt{\nu_1}} a_2 & \beta_2\left|\frac{a_1}{c}\right| & -\beta\left|\frac{a_1}{c}\right| \\ -a_4\left|\frac{c}{a_1}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} & -\alpha_2\left|\frac{c}{a_1}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} & \alpha\left|\frac{c}{a_1}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \end{pmatrix}. \tag{16}$$

Changing to the new ζ coordinates, we can easily check the cone conditions: The matrix $D\tilde{P}(\zeta_2, \zeta_1, \zeta_3)$ is given by

$$\begin{pmatrix} -\left|\frac{\zeta_2}{c}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} & 0 & 0 \\ \frac{\zeta_1}{\alpha_1}\frac{c}{\zeta_2^2} - (\alpha_1\zeta_3 - \alpha_2\zeta_1)\frac{\beta_1}{\alpha_1}\left|\frac{\zeta_2}{c}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} & \alpha_1\left|\frac{c}{\zeta_2}\right| + \frac{\alpha_2}{\alpha_1}\left|\frac{\zeta_2}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} & \frac{\beta_1}{\alpha_1}\left|\frac{\zeta_2}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \\ -(\alpha_1\zeta_3 - \alpha_2\zeta_1)\left|\frac{\zeta_2}{c}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} & \alpha_2\left|\frac{\zeta_2}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} & \left|\frac{\zeta_2}{c}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} \end{pmatrix} \tag{17}$$

and $D\tilde{P}^{-1}$ by

$$\begin{pmatrix} -\left|\frac{c}{\zeta_2}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} & 0 & 0 \\ -\left|\frac{c}{\zeta_2}\right|^{\frac{6}{7}\epsilon\sqrt{\nu_1}} \frac{\zeta_1}{\zeta_2\alpha_1^2} & \left|\frac{\zeta_2}{c\alpha_1}\right| & -\frac{\beta_1}{\alpha_1}\left|\frac{\zeta_2}{c}\right| \\ (\frac{\alpha_2\zeta_1}{c} - \alpha_1(\alpha_1\zeta_3 - \alpha_2\zeta_1))\left|\frac{c}{\zeta_2}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} & -\alpha_2\left|\frac{c}{\zeta_2}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} & \alpha_1^2\left|\frac{c}{\zeta_2}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}} + \beta_1\alpha_2\left|\frac{\zeta_2}{c}\right| \end{pmatrix}. \tag{18}$$

Now it is straightforward to see, that the term $\alpha_1\left|\frac{c}{\zeta_2}\right|$ is the largest entry in the matrix (17). Then it preserves horizontal cones with constant e.g. $s = 0.3$ and expands them with expansion rate $\mu = \alpha_1\frac{c}{2\delta} > 1$. Similar we see, that $\alpha_1^2\left|\frac{c}{\zeta_2}\right|^{1+\frac{6}{7}\epsilon\sqrt{\nu_1}}$ is the leading term of the last two lines in (18). Hence it preserves vertical cones with constants $s = 0.3$ and expands them with constant λ^{-1} for $\lambda = 2\frac{\delta}{c}^{\frac{6}{7}\epsilon\sqrt{\nu_1}} < 1$.
By Katok and Hasselblatt [15, p.274] we have the existence of an invariant hyperbolic Cantor set for the reduced map $\tilde{P}$, such that the dynamics are

topological conjugate to the shift on two symbols for this reduced map. Then for the complete return map P there exists the shift of finite type with the transition matrix of the theorem: if an orbit is near the loop l in the present, then as the shift is on the symbols 2 and 3 the next loop in the itinerary has to be $l+1 \bmod 3$ or $l+2 \bmod 3$. Similarly the previous one was $l+1 \bmod 3$ or $l+2 \bmod 3$. Thus possible sequences $(x_n)_{n\in\mathbf{Z}}$ have the form $x_n \neq x_{n+1}$. The realization of all these sequences are guaranteed by the existence of the full shift on two symbols for $\tilde{P}$. Proposition 6.5.3 in [15] gives then even persistence under small C^1 perturbations i.e. for an open set in parameter space. Hence we can include higher order terms. This also verifies the linear approximation of the global maps, for which all equivariant higher order terms can be neglected. It remains to check the symmetry properties of C, P and τ. The sections S_k^{in} are related by symmetry: $S_2^{in} = \gamma S_1^{in}$ and $S_3^{in} = \gamma^2 S_1^{in}$. Furthermore S_1^{in} is κ-invariant and $S_3^{in} = \kappa S_2^{in}$. Then P^{loc} is equivariant, because the linearizing diffeomorphism is D_3-equivariant. The global part is equivariant under rotation γ by construction. It is equivariant under κ by the following argument:

$$\begin{aligned} \kappa^{-1} P^{glo} \kappa x &= \kappa^{-1} \Phi_{t(\kappa x)}(\kappa x) = \Phi_{t(\kappa x)}(x) \\ P^{glo} x &= \Phi_{t(x)}(x) \end{aligned}$$

As the times $t(\kappa x)$ and $t(x)$ are both close to the time needed of the homoclinic orbits from the 'out' section to the 'in' section, we get $t(\kappa x) \approx t(x)$. As $\Phi_{t(x)}(x), \Phi_{t(\kappa x)}(x) \in S_k^{in}$ for the same k, we get $t(x) = t(\kappa x)$. Hence P^{glo} and P are equivariant. Then $C = \cap_{n=-\infty}^{\infty} P^n(\cup_{i=1,2,3} S_i^{in})$ is D_3 invariant, because P^n is equivariant and $\cup_{i=1,2,3} S_i^{in}$ is invariant. If $x \in C$ and $x = P^n(a_n)$ with $a_n \in S_{x_n}^{in}$, then $\tau(x) = (x_n)_{n\in\mathbf{Z}}$ and the equivariance of τ can be easily checked using the representations (2) and (1). □

5 Discussion

In this section we give a more complete bifurcation diagram of the Takens-Bogdanov point with D_3-symmetry, using numerical studies of the normal form equations. Then we will describe an application to coupled oscillators.

A major drawback in all further numerical studies is that there are not any stable dynamic features except the origin for some parameter values ($\mu_1, \mu_2 < 0$). Therefore all direct simulations will not give much insight. Some conjectures about the periodic solutions created at Hopf bifurcations are possible using the path-following program AUTO [5].

The dynamics are fully understood in the invariant plane $\mathrm{Fix}(\kappa)$ by [14], see also [16]. There are two branches of periodic orbits starting from the D_3-Hopf bifurcation of 0 and the Hopf of N_1 at $\mu_2 = -\mu_1, \mu_1 > 0$. These branches do not undergo any folds and end at the homoclinic orbit. The global behavior of the other branches of periodic solutions are analyzed using AUTO. These branches of periodic solutions outside $\mathrm{Fix}(\kappa)$ seem to break down at the clover

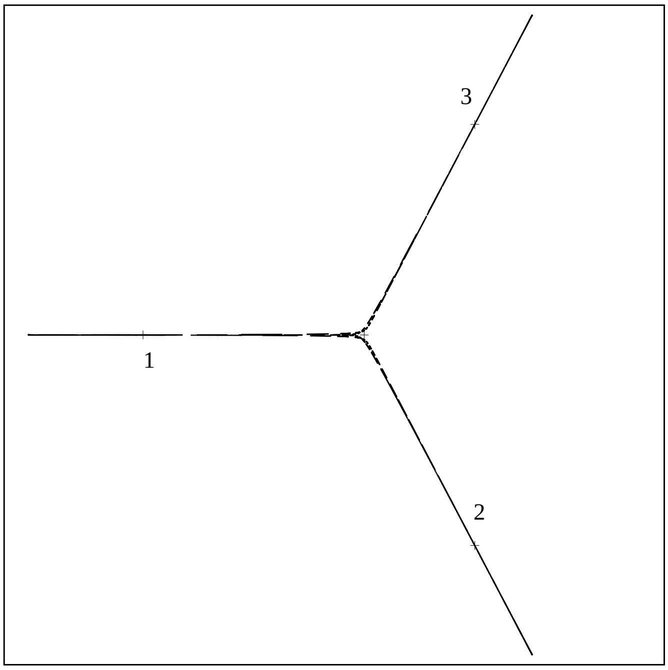


Figure 7: A periodic orbit with sequence 1213 of isotropy type $\tilde{\mathbf{Z}}_2$ of the branch coming from the D_3-Hopf. Parameter values are near the existence of the homoclinic clover. A projection on the v plane is shown, the crosses denote steady states. The trajectory of periodic orbit was approximated by integrating the differential equation starting at points, which described the periodic solution for AUTO.

structure of homoclinic orbits. Probably they are some of the periodic orbits of the subshift:

- The periodic solutions with isotropy type $\tilde{\mathbf{Z}}_2$ coming from the D_3-Hopf bifurcation have period 4 created by the sequences 1213, 2321 and 3132, see figure7.

- The solutions coming from the Hopf bifurcation of $N_{1,2,3}$ at $\mu_2 = \mu_1, \mu_1 < 0$ seem to have period 2, see figure 8.

- Even if the author could not pick up the $\tilde{\mathbf{Z}}_3$ periodic solutions starting at the D_3-Hopf bifurcation for path-following with AUTO. We might conjecture that this branch also ends at the homoclinic clover. They are probably of period 3 with sequences 123 and 132.

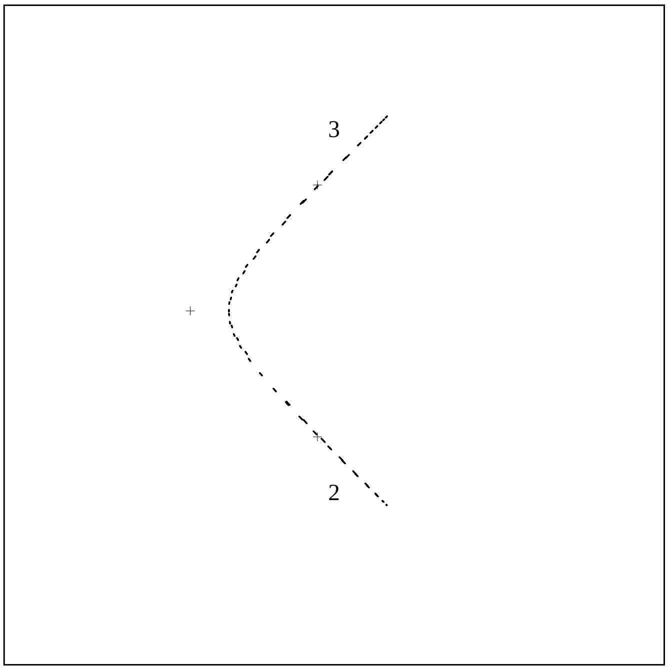


Figure 8: A periodic solution with sequence 23 with isotropy type $\tilde{\mathbf{Z}}_2$ on the branch coming from the Hopf bifurcation of N_1.

The entire horseshoe does persist for some parameter by a general perturbation argument for hyperbolic sets. It remains an open question how long for example the other periodic orbits created by the horseshoe persist. This will probably involve even more complicated bifurcations.

We will consider an application to three coupled oscillators following Fiedler [6]. The system is given by

$$\dot{x_i} = f(x_i) + D(x_{i-1} + x_{i+1} - 2x_i) \pmod 3, i = 1, 2, 3, \tag{19}$$

where $x_i \in \mathbf{R}^k$ and $D = \operatorname{diag}(d_1, \ldots, d_k)$. This system is equivariant under permutations of x_1, x_2 and x_3. The symmetry group is isomorphic to D_3. If we have a homogeneous solution, it will stay homogeneous under the evolution of time. We change to (x, y, z) coordinates where

$$x = x_1 + x_2 + x_3, y = x_1 - x_2, z = x_2 - x_3.$$

In the new coordinate system we then have:

$$
\begin{aligned}
\dot{x} &= f(\frac{x+z+2y}{3}) + f(\frac{x+z-y}{3}) + f(\frac{x-y-2z}{3}) \\
\dot{y} &= f(\frac{x+z+2y}{3}) - f(\frac{x+z-y}{3}) - 3Dy \\
\dot{z} &= f(\frac{x+z-y}{3}) - f(\frac{x-y-2z}{3}) - 3Dz
\end{aligned}
\tag{20}
$$

We consider the homogeneous equilibrium (x_0, x_0, x_0) with linearisation $f'(x_0) = A$. In the new coordinates the equilibrium is $(3x_0, 0, 0)$. Its Jacobian in the entire system is given by the block diagonal matrix $\mathrm{diag}(A, A - 3D, A - 3D)$.

We choose $k = 2$ and for f the dynamics of the Brusselator as an easy example. It gives some insight into the possible behavior of chemical oscillator. So $f = (f_1, f_2)$ is given by $f_1(\xi_1, \xi_2) = a - (b+1)\xi_1 + \xi_1^2\xi_2$, $f_2(\xi_1, \xi_2) = b\xi_1 - \xi_1^2\xi_2$ with $a, b > 0$, the equilibrium is $x_0 = (a, \frac{b}{a})$ and $A = \begin{pmatrix} b-1 & a^2 \\ -b & -a^2 \end{pmatrix}$. We choose $D = \frac{1}{3}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Then $A - 3D$ has a double eigenvalues 0 if

$$
\begin{array}{lclcl}
\det(A-3D) &=& \lambda_1\lambda_2 + \lambda_1 a^2 - \lambda_2(b-1) + a^2 &=& 0 \\
\mathrm{trace}(A-3D) &=& b - 1 - a^2 - \lambda_1 - \lambda_2 &=& 0
\end{array}
$$

The solution is given by $(\lambda_1, \lambda_2) = (b - 1 - a\sqrt{b}, -a^2 + a\sqrt{b})$, the diffusion constants λ_1, λ_2 are positive and therefore somehow realistic for $a < \frac{b-1}{\sqrt{b}}$. So this D_3-equivariant system has a Takens-Bogdanov point since there is a double zero eigenvalue and $A - 3D \neq 0$. We apply our bifurcation analysis for Takens-Bogdanov points with D_3-symmetry to this problem. It will be valid on a four-dimensional center manifold which is tangent to the subspace spanned by y and z.

As $\mathrm{trace}(A) > \mathrm{trace}(A - 3D) = 0$ holds for $\lambda_1 + \lambda_2 > 0$, the matrix A has at least one eigenvalue with positive real part. Hence all dynamical features will be unstable if we consider the entire system. We could stabilize the system when using negative diffusion rates. But still all branching solutions have unstable directions due to the Takens-Bogdanov point making them inaccessible for direct numerical simulation.

The origin will still correspond to the homogeneous solution even after the needed coordinate changes. Then an interpretation of a D_3-Hopf bifurcation in a ring of three coupled oscillators is given in [11, XVII.4]. The three different types of periodic solutions give different waveforms, phase shifts and resonances for the three cells. We furthermore expect near the bifurcation point the existence of inhomogeneous steady state solutions with two cells being in the same state. The periodic solutions coming from the Hopf bifurcations of these fixed points oscillate around these inhomogeneous steady states. In the first type two cells are in phase and in the other type two cells have a phase shift of

π. The periodic solutions collapse at the homoclinic orbits, since by moving in parameter space parts of the periodic orbits reach a state very close to the homogeneous equilibrium. For these parameter values the system is already 'chaotic' because of the existence of shift dynamics. When the solution follows one of the loops of the 'clover' structure it has nearly a $\mathbf{Z}_2$ symmetry, i.e. two cells have nearly the same state. Hence within the shift dynamics we have arbitrary changes of two out of three cells being nearly in phase. The structure of our subshift forces the system to change to another pair of cells being in phase after some time. Because of the unstable directions of the hyperbolic structure this behavior is only observable as a transient motion to infinity or to some stable solutions far away from the Takens-Bogdanov point.

Acknowledgments

I would like to thank Peter Ashwin, Bernold Fiedler, Mark Roberts, Arnd Scheel and Ian Stewart for discussions.

References

[1] Armbruster,D., Guckenheimer,J. and Kim,S. [1989]. Chaotic dynamics in systems with square symmetry, *Physics Letters A* 140 no 7/8, p 416-420.

[2] Arnold,V.I.[1983].*Geometrical methods in the theory of ordinary differential equations.* Grundlehren der mathematischen Wissenschaften 250, Springer-Verlag,New York.

[3] Bogdanov,R.I.[1975]. Versal deformations of a singular point on the plane in the case of zero eigenvalues.*Functional Anal. Appl.(2)*9, p 144-145.

[4] Dangelmayr,G. and Knobloch,E.[1987].The Takens-Bogdanov bifurcation with O(2) symmetry.*Phil. Trans. R. Soc. London* A322, p 243-279.

[5] Doedel,E.J.,Wang,X.J. and Fairgrieve,T.[1994]. *AUTO94 Software for continuation and bifurcation problems in ordinary differential equations.* California Institute of Technology.

[6] Fiedler,B. [1986]. Global Hopf bifurcation of two-parameter flows, *Arch. Rational Mech. Anal.*94, p 59-81.

[7] Fiedler,B. [1996]. Global path following of homoclinic orbits in two-parameter flows, Pitman Math. Research Notes, No 352, p 79-146.

[8] Field,M.J.[1983]. Isotopy and stability of equivariant diffeomorphisms, *Proc. London Math. Soc.* 46, p 487-516.

[9] Gatermann, K. and Guyard, F. [1999] Gröbner bases, invariant theory and equivariant dynamics, *Journal of Symbolic Calculation* 27, p 1-28; program:URL: http://www.zib.de/gatermann /symmetry.html.

[10] Glendeninning,P.[1994] *Stability, instability and Chaos.* Cambridge University Press.

[11] Golubitsky,M., Stewart,I. and Schaeffer,D.G. [1988]. *Singularities and groups in bifurcation theory, volume II.*Appl.Math.Sci. 69,Springer-Verlag, New York.

[12] Golubitsky,M., Swift,J.W. and Knobloch,E. [1984]. Symmetries and pattern selection in Rayleigh-Bénard convection,*Physica D* 10,p 249-276.

[13] Guckenheimer,J. and Holmes,P. [1983].*Nonlinear oscillation, dynamical systems, and bifurcations of vector fields.* Appl.Math.Sci.42,Springer-Verlag, New York.

[14] Hirschberg,P. and Knobloch,E.[1991]. An unfolding of the Takens-Bogdanov singularity.*Quart. Appl. Math.* 49, p 281-287.

[15] Katok,A. and Hasselblatt,B.[1995]. *Introduction to the modern theory of dynamical systems.* Encyclopedia of mathematics and its applications v.54, Cambridge University Press.

[16] Lari-Lavassani,A., Langford,W.F., Huseyin,K. and Gatermann,K.[1999]. Steady-state mode interactions for D_3 and D_4-symmetric systems. Dynamics of Continuous, Discrete and Impulsive Systems 6,p 169–209.

[17] Takens,F.[1974]. Singularities of vector fields, *Inst. Hautes Études Sci. Publ. Math.* 43, p 47-100.

Karsten Matthies
Institut für Mathematik I
Freie Universität Berlin
Arnimallee 2-6
D - 14195 Berlin
Germany
matthies@math.fu-berlin.de

Einstein Metrics and Stability for Flat Connections on Compact Hermitian Manifolds, and a Correspondence with Higgs Operators in the Surface Case

M. Lübke

Received: February 19, 1999
Revised: October 12, 1999

Communicated by Thomas Peternell

Abstract. A flat complex vector bundle (E, D) on a compact Riemannian manifold (X, g) is stable (resp. polystable) in the sense of Corlette [C] if it has no D-invariant subbundle (resp. if it is the D-invariant direct sum of stable subbundles). It has been shown in [C] that the polystability of (E, D) in this sense is equivalent to the existence of a so-called harmonic metric in E. In this paper we consider flat complex vector bundles on compact Hermitian manifolds (X, g). We propose new notions of g-(poly-)stability of such bundles, and of g-Einstein metrics in them; these notions coincide with (poly-)stability and harmonicity in the sense of Corlette if g is a Kähler metric, but are different in general. Our main result is that the g-polystability in our sense is equivalent to the existence of a g-Hermitian-Einstein metric. Our notion of a g-Einstein metric in a flat bundle is motivated by a correspondence between flat bundles and Higgs bundles over compact surfaces, analogous to the correspondence in the case of Kähler manifolds [S1], [S2], [S3].

1991 Mathematics Subject Classification: 53C07

1 Introduction.

Let X be an n-dimensional compact complex manifold. If X admits a *Kähler* metric g, then it is known by work of in particular Simpson [S1],[S2],[S3] that there exists an canonical identification of the moduli space of polystable (or semisimple) flat bundles on X with the moduli space of g-polystable Higgs-bundles with vanishing Chern classes on X. This identification has been used in showing that certain groups are not fundamental groups of compact Kähler manifolds. The construction uses the existence of *canonical* metrics, called *g-harmonic* in the case of flat bundles, and *g-Einstein* in the case of Higgs bundles.

For flat bundles, the equivalence of semisimplicity and the existence of a g-harmonic metric holds on compact *Riemannian* manifolds [C]. Furthermore, the equivalence of g-polystability and the existence of a g-Einstein metrics for Higgs bundles should generalize to the case of *Hermitian* manifolds as in the case of holomorphic vector bundles, using *Gauduchon* metrics. Nevertheless, an identification as above cannot be expected for general compact Hermitian manifolds, since it should imply restrictions on the fundamental group, but every finitely presented group is the fundamental group of a 3-dimensional compact complex manifold by a theorem of Taubes [T].

In the case of compact complex *surfaces*, however, things are different. We show that for an integrable Higgs bundle (E, d'') with vanishing real Chern numbers and of g-degree 0 with g-Einstein metric h on a compact complex surface X with Hermitian metric g, there is an canonically associated flat connection D in E, again of g-degree 0, such that h is what we call a *g-Einstein* metric for (E, D), and that the converse is also true. Furthermore, this correspondence preserves isomorphism types and hence descends to a bijection between moduli spaces.

The notion of a g-Einstein metric in a flat bundle makes sense in higher dimension, too, is equivalent to g-harmonicity in the case of a Kähler metric, but different in general, and we show that the existence of such a metric in a flat bundle (E, D) is equivalent to the g-polystability of this bundle in the sense that E is the direct sum of D-invariant g-stable flat subbundles. Here we call a flat bundle (E, D) g-stable if every D-invariant subbundle has *g-slope* larger(!) than the g-slope of (E, D). g-stability of a flat bundle is equivalent to its stability (in the sense of Corlette) in the Kähler case, but a weaker condition in general: A stable bundle is always g-stable, but the tangent bundles of certain Inoue surfaces are examples of g-stable bundles which are not stable.

We expect that for a non-Kähler surface with Hermitian metric g, there is a natural bijection between the moduli space of g-polystable Higgs bundles, with vanishing Chern numbers and g-degree, and the moduli space of g-polystable flat bundles with vanishing g-degree. In the last section we consider the special case of *line* bundles on surfaces. Here the stability is trivial, and the existence of Einstein metrics is easy to show, so we get indeed the expected natural bijection between moduli spaces of line bundles of degree 0. We further show how this can be extended (in a non-natural way) to the moduli spaces of line bundles of arbitrary degree; this extension argument works in fact for bundles of arbitrary rank once the correspondence for degree 0 has been established.

Acknowledgments. The author wishes to thank A. Teleman for the sug-

gestion to study the relation between flat connections and Higgs operators on Hermitian manifolds, and for several useful hints. Discussions with him and Ch. Okonek have been of considerable help in preparing this paper.
The author was supported in part by EC-HCM project AGE "Algebraic Geometry in Europe", contract no. ERBCHRXCT 940557.

2 Preliminaries.

Let X be a compact n-dimensional complex manifold, and $E \longrightarrow X$ a differentiable $\mathbb{C}^r$-vector bundle on X. We fix the following
Notations:
$A^p(X)$ (resp. $A^{p,q}(X)$) is the space of differentiable p-forms (forms of type (p,q)) on X.
$A^p(E)$, $A^{p,q}(E)$ are the spaces of differential forms with values in E.
$\mathcal{A}(E)$ is the space of linear connections D in E. For a connection $D \in \mathcal{A}(E)$ we write $D = D' + D''$, where D' is of type (1,0) and D'' of type (0,1).
$\mathcal{A}(E,h) \subset \mathcal{A}(E)$ is the subspace of h-unitary connections d in E, where h is a Hermitian metric in E. We write $d = \partial + \bar\partial$, where ∂ is of type (1,0) and $\bar\partial$ of type (0,1).
$\mathcal{A}_f(E) := \{\ D \in \mathcal{A}(E) \mid D^2 = 0\ \}$ is the subset of *flat* connections.
$\bar{\mathcal{A}}(E)$ is the space of semiconnections $\bar\partial$ of type (0,1) in E (i.e. $\bar\partial$ is the (0,1)-part of some $D \in \mathcal{A}(E)$).
$\mathcal{H}(E) := \{\ \bar\partial \in \bar{\mathcal{A}}(E) \mid \bar\partial^2 = 0\ \}$ is the subset of *integrable* semiconnections or *holomorphic structures* in E.
$\mathcal{A}''(E) := \bar{\mathcal{A}}(E) \oplus A^{1,0}(\mathrm{End}E) = \{\ d'' = \bar\partial + \theta \mid \bar\partial \in \bar{\mathcal{A}}(E),\ \theta \in A^{1,0}(\mathrm{End}E)\ \}$ is the space of *Higgs operators* in E.
$\mathcal{H}''(E) := \{\ d'' \in \mathcal{A}''(E) \mid (d'')^2 = 0\ \}$ is the subset of *integrable* Higgs operators.
Often the same symbol is used for a connection, semiconnection, Higgs operator etc. in E and the induced operator in End$E$.
Two connections $D_1, D_2 \in \mathcal{A}(E)$ are *isomorphic*, $D_1 \cong D_2$, if there exists a differentiable automorphism f of E such that $f \circ D_1 = D_2 \circ f$, which is equivalent to $D(f) = 0$, where D is the connection in End$E$ induced by D_1 and D_2, i.e. $D(f) = D_2 \circ f - f \circ D_1$. In the same way the isomorphy of semiconnections resp. Higgs operators is defined.
If a Hermitian metric h in E is given, then a superscript * means adjoint with respect to h.

For $D = D' + D''$ there are unique semiconnections δ'_h,δ''_h of type (1,0), (0,1) respectively such that $D' + \delta''_h$ and $\delta'_h + D''$ are h-unitary connections. Define $\delta_h := \delta'_h + \delta''_h$; then $d_h := \frac{1}{2}(D + \delta_h)$ is h-unitary, and $\Theta_h := D - d_h = \frac{1}{2}(D - \delta_h)$ is a h-selfadjoint 1-form with values in End$E$. Let $d_h = \partial_h + \bar\partial_h$ be the decomposition in the parts of type (1,0) and (0,1), and let θ_h be the $(1,0)$-part of Θ_h; then it holds

$$D = d_h + \Theta_h = \partial_h + \bar\partial_h + \theta_h + \theta_h^*.$$

The map

$$I_h : \mathcal{A}(E) \longrightarrow \mathcal{A}''(E), I_h(D) := d''_h := \bar{\partial}_h + \theta_h \in \mathcal{A}''(E)$$

is bijective; the inverse is given as follows. For $d'' = \bar{\partial} + \theta \in \mathcal{A}''(E)$ let ∂_h be the unique semiconnection of type $(1,0)$ such that the connection $d_h := \partial_h + \bar{\partial}$ is h-unitary, and define $\Theta := \theta + \theta^*$. Then

$$I_h^{-1}(d'') = D_h := d_h + \Theta \in \mathcal{A}(E).$$

Remark 2.1 *i) In general, if $D_1, D_2 \in \mathcal{A}(E)$ are isomorphic, then $I_h(D_1)$ and $I_h(D_2)$ are not isomorphic, and vice versa.*
ii) $D_h = d_h + \theta + \theta^*$ is not h-unitary unless $\theta = 0$, but the connections $d_h - \theta + \theta^*$ and $d_h + \theta - \theta^*$ are.
iii) Any metric h' in E is of the form $h' = f \cdot h$, i.e. $h'(s,t) = h(f(s),t)$, where f is a h-selfadjoint and positive definite. For a connection D it is easy to show that the operator $\delta_{h \cdot f}$ associated to D and $f \cdot h$ is given by $\delta_{h \cdot f} = f^{-1} \circ \delta_h \circ f = \delta_h + f^{-1} \circ \delta_h(f)$, so it holds

$$\begin{aligned} d''_{f \cdot h} &= d''_h + \frac{1}{2} f^{-1} \circ \delta''_h(f) - f^{-1} \circ \delta'_h(f) \\ &= d''_h + \frac{1}{2} f^{-1} \circ \bar{\partial}_h(f) - \frac{1}{2} f^{-1} \circ \theta^*_h(f) - \frac{1}{2} f^{-1} \circ \partial_h(f) + \frac{1}{2} f^{-1} \circ \theta_h(f). \end{aligned}$$

Conversely, for a given Higgs operator d'' one verifies

$$D_{f \cdot h} = D_h + f^{-1} \circ \partial_h(f) + f^{-1} \circ \theta(f).$$

In particular, if f is constant then the two maps I_h and $I_{f \cdot h}$ coincide.

Definition 2.2 *i) $G_h := (d''_h)^2$ is called the pseudocurvature of D with respect to h.*
ii) $F_h := D_h^2$ is called the curvature of d'' with respect to h.

Remark 2.3 *i) Obviously it holds: $I_h(D)$ is an integrable Higgs operator if and only if $G_h = 0$, and $I_h^{-1}(d'')$ is a flat connection if and only if $F_h = 0$.*
ii) For $i = 1, 2$, let E_i be a differentiable complex vector bundle on X with Hermitian metric h_i and connection D_i. Let h be the induced metric and D the induced connection in $\mathrm{Hom}(E_1, E_2)$. Denote by $G_{i,h}$ resp. G_h the pseudocurvature of D_i resp. D with respect to h_i resp. h. Then for $f \in A^0(\mathrm{Hom}(E_1, E_2))$ it holds $G_h(f) = G_{2,h} \circ f - f \circ G_{1,h}$.
Similarly, the curvature F_h of the Higgs operator induced in $\mathrm{Hom}(E_1, E_2)$ by Higgs operators d''_i in the E_i is given by $F_h(f) = F_{2,h} \circ f - f \circ F_{1,h}$.
iii) If D is a connection, then D^2 is the curvature of d''_h with respect to h, and if d'' is a Higgs operator, then $(d'')^2$ is the pseudocurvature of D_h with respect to h. This trivially follows from the bijectivity of I_h.

Lemma 2.4 *i) For $D \in \mathcal{A}(E)$ let $D = d_h + \Theta_h = \partial_h + \bar{\partial}_h + \theta_h + \theta_h^*$ be the decomposition induced by h as above. If D is flat, then it holds $\delta_h^2 = 0$, $d_h(\Theta_h) = 0$, i.e. $\partial_h(\theta_h) = \bar{\partial}_h(\theta_h^*) = \partial_h(\theta_h^*) + \bar{\partial}_h(\theta_h) = 0$, and furthermore $d_h^2 = -\Theta_h \wedge \Theta_h$.*
ii) For $d'' = \bar{\partial} + \theta \in \mathcal{A}''(E)$ let ∂_h, d_h and D_h be as above, and write $d'_h := \partial_h + \theta^*$.
If d'' is integrable, then it holds $(d'_h)^2 = 0$, i.e. $\partial_h^2 = \partial_h(\theta^*) = \theta^* \wedge \theta^* = 0$, $d_h^2 = [\partial_h, \bar{\partial}]$, and hence $F_h = d_h^2 + [\theta, \theta^*] + \partial_h(\theta) + \bar{\partial}(\theta^*)$.

Proof: i) For $D = D' + D'' \in \mathcal{A}_f(E)$ it holds

$$\begin{aligned} 0 &= \partial\partial h(s,t) \\ &= h((D')^2(s),t) - h(D'(s), \delta_h''(t)) + h(D'(s), \delta_h''(t)) + h(s, (\delta_h'')^2(t)) \\ &= h(s, (\delta_h'')^2(t)) \end{aligned}$$

for all $s, t \in A^0(E)$, i.e. $(\delta_h'')^2 = 0$. Similarly one sees $(\delta_h')^2 = 0 = \delta_h'\delta_h'' + \delta_h''\delta_h'$, yielding $\delta_h^2 = 0$. We conclude

$$d_h(\Theta_h) = \frac{1}{4}[D + \delta_h, D - \delta_h] = 0,$$

and

$$0 = D^2 = (d_h + \Theta_h)^2 = d_h^2 + d_h(\Theta_h) + \Theta_h \wedge \Theta_h = d_h^2 + \Theta_h \wedge \Theta_h.$$

ii) For $d'' = \bar{\partial} + \theta \in \mathcal{H}''(E)$ and $d_h = \partial_h + \bar{\partial}$ it is well known that $\partial_h^2 = 0$, and hence $d_h^2 = [\partial_h, \bar{\partial}]$. Furthermore, for all $s, t \in A^0(E)$ it holds

$$\begin{aligned} & h(\partial_h(\theta^*)(s), t) \\ &\quad = h(\partial_h \circ \theta^*(s), t) + h(\theta^* \circ \partial_h(s), t) \\ &\quad = \partial h(\theta^*(s), t) + h(\theta^*(s), \bar{\partial}(t)) - h(\partial_h(s), \theta(t)) \\ &\quad = \partial h(s, \theta(t)) + h(s, \theta \circ \bar{\partial}(t)) - h(\partial_h(s), \theta(t)) \\ &\quad = h(\partial_h(s), \theta(t)) + h(s, \bar{\partial} \circ \theta(t)) + h(s, \theta \circ \bar{\partial}(t)) - h(\partial_h(s), \theta(t)) \\ &\quad = h(s, \bar{\partial}(\theta)(t)) = 0, \end{aligned}$$

and

$$h(\theta^* \wedge \theta^*(s), t) = -h(s, \theta \wedge \theta(t)) = 0;$$

this shows $\partial_h(\theta^*) = 0 = \theta^* \wedge \theta^*$. □

Now let g be a Hermitian metric in X, and denote by ω_g the associated $(1,1)$-form on X, by Λ_g the contraction by ω_g, and by $*_g$ the associated Hodge-$*$-operator.
Recall that in the conformal class of g there exists a *Gauduchon* metric $\tilde{g}$, i.e. a metric satisfying $\bar{\partial}\partial(\omega_{\tilde{g}}^{n-1}) = 0$; $\tilde{g}$ is unique up to a constant positive factor if $n \geq 2$ ([G] p. 502, [LT] Theorem 1.2.4).

There is a natural way to define a map

$$\deg_g : \mathcal{H}(E) \longrightarrow \mathbb{R},$$

called *g-degree*, with the following properties (see [LT] sections 1.3 and 1.4):
- If g is a Gauduchon metric, and $\bar\partial \in \mathcal{H}(E)$ is a holomorphic structure, then $\deg_g(\bar\partial)$ is given as follows: Choose any Hermitian metric h in E, and let d be the *Chern connection* in (E, ∂) induced by h, i.e. the unique h-unitary connection in E with $(0,1)$-part $\bar\partial$. Then

$$\deg_g(\bar\partial) := \frac{i}{2\pi}\int\limits_X \mathrm{tr}(d^2) \wedge \omega_g^{n-1} = \frac{i}{2n\pi}\int\limits_X \mathrm{tr}\Lambda_g d^2 \cdot \omega_g^n = \frac{i}{2n\pi}\int\limits_X \mathrm{tr}\Lambda_g[\bar\partial, \partial] \cdot \omega_g^n.$$

- If g is arbitrary, then there is a unique Gauduchon metric $\tilde g$ in the conformal class of g such that $\deg_g = \deg_{\tilde g}$.
The *g-slope* of $\bar\partial$ is

$$\mu_g(\bar\partial) := \frac{\deg_g(\bar\partial)}{r},$$

where r is the rank of E.
If $D = D' + D''$ is a flat connection, then it holds $(D'')^2 = 0$, so D'' is a holomorphic structure. We define the g-degree and g-slope of D as

$$\deg_g(D) := \deg_g(D''), \mu_g(D) := \mu_g(D'').$$

Similarly, for an integrable Higgs operator $d'' = \bar\partial + \theta$ it holds $\bar\partial^2 = 0$, and we define

$$\deg_g(d'') := \deg_g(\bar\partial), \mu_g(d'') := \mu_g(\bar\partial).$$

Observe that in all three cases the g-degrees (resp. slopes) of isomorphic operators are the same.

Remark 2.5 *Suppose that g is a Kähler metric, i.e. $d(\omega_g) = 0$. Then the g-degree is a topological invariant of the bundle E, completely determined by the first real Chern class $c_1(E)_{\mathbb{R}} \in H^2(X, \mathbb{R})$. In particular, since all real Chern classes of a flat bundle vanish, it holds $\deg_g(D) = 0$ for every flat connection D in E. On the other hand, if e.g. X is a surface admitting no Kähler metric and g is Gauduchon, then every real number is the g-degree of a flat line bundle on X ([LT] Proposition 1.3.13).*

Lemma 2.6 *If g is a Gauduchon metric, then for any metric h in E it holds:*
i) If D is a flat connection, then

$$\deg_g(D) = -\frac{i}{n\pi}\int\limits_X \mathrm{tr}\Lambda_g G_h \cdot \omega_g^n,$$

where G_h is the pseudocurvature of d'' with respect to h.
ii) If d'' is an integrable Higgs operator, then

$$\deg_g(d'') = \frac{i}{2n\pi} \int_X \mathrm{tr}\Lambda_g F_h \cdot \omega_g^n,$$

where F_h is the curvature of d'' with respect to h.

Proof: i) Observe that $\Lambda_g G_h = \Lambda_g \bar\partial_h(\theta_h)$. The Chern connection in (E, D'') induced by h is $D'' + \partial_h - \theta_h = D - 2\theta_h$, and it holds

$$\mathrm{tr}\Lambda_g(D - 2\theta_h)^2 = -2\mathrm{tr}\Lambda_g((\bar\partial + \theta^*)(\theta) = -2\mathrm{tr}\Lambda_g(G_h + [\theta, \theta^*]) = -2\mathrm{tr}\Lambda_g(G_h),$$

so the claim follows by integration.
ii) Lemma 2.4 implies $\mathrm{tr}\Lambda_g F_h = \mathrm{tr}\Lambda_g d_h^2$; again the claim follows by integration. □

3 Einstein metrics and stability for flat bundles.

We fix a Hermitian metric g in X; the associated volume form is $\mathrm{vol}_g := \frac{1}{n!}\omega_g^n$, and the g-volume of X is $\mathrm{Vol}_g(X) := \int_X \mathrm{vol}_g$. We further fix a Hermitian metric h in E, and denote by $|.\,|$ the pointwise norm on forms with values in E (and associated bundles) defined by h and g.
Let $D \in \mathcal{A}_f(E)$ be a flat connection in E, and write $D = d + \Theta = \partial + \bar\partial + \theta + \theta^*$ as in section 1. Let $d''_h = I_h(D) = \bar\partial + \theta \in \mathcal{A}''(E)$ be the Higgs operator associated to D, and $G_h = (d''_h)^2$ its pseudocurvature. From $\Lambda_g G_h = \Lambda_g \bar\partial_h(\theta_h)$ and Lemma 2.4 we deduce

$$(i\Lambda_g G_h)^* = -i\Lambda_g((\bar\partial(\theta))^*) = -i\Lambda_g\partial(\theta^*) = i\Lambda_g\bar\partial(\theta) = i\Lambda_g G_h,$$

so $i\Lambda_g G_h$ is selfadjoint with respect to h.

Remark 3.1 *It also holds $i\Lambda_g G_h = \frac{i}{2}\Lambda_g(\bar\partial(\Theta) - \partial(\Theta))$, which in the case of a Kähler metric g equals $\frac{1}{2}d^*(\Theta)$, where d^* is the L^2-adjoint of $d = \partial + \bar\partial$.*

Definition 3.2 *h is called a g-Einstein metric in (E, D) if $i\Lambda_g G_h = c \cdot \mathrm{id}_E$ with a real constant c, which is called the Einstein constant.*

Lemma 3.3 *Let h be a g-Einstein metric in (E, D), and $\tilde g = \varphi \cdot g$ conformally equivalent to g. Then there exists a $\tilde g$-Einstein metric $\tilde h$ in (E, D) which is conformally equivalent to h.*

Proof: $\tilde g = \varphi \cdot g$ implies $\Lambda_{\tilde g} = \frac{1}{\varphi} \cdot \Lambda_g$. From Remark 2.1 iii) it follows that for $f \in \mathcal{C}^\infty(X, \mathbb{R})$ it holds $G_{e^f \cdot h} = G_h - \frac{1}{4}\bar\partial\partial(f) \cdot \mathrm{id}_E$. Hence the condition $i\Lambda_g G_h = c \cdot \mathrm{id}_E$ implies $i\Lambda_{\tilde g} G_{e^f \cdot h} = (\frac{c}{\varphi} - \frac{1}{4}P(f)) \cdot \mathrm{id}_E$, where $P := i\Lambda_{\tilde g}\bar\partial\partial$. Since $\mathcal{C}^\infty(X, \mathbb{R}) = \mathrm{im}P \oplus \mathbb{R}$ ([LT] Corollary 2.9), there exists an f such that $\frac{c}{\varphi} - \frac{1}{4}P(f)$ is constant. □

Lemma 3.4 *If* $i\Lambda_g G_h = c\cdot \mathrm{id}_E$ *with* $c \in \mathbb{R}$, *then it holds:*
i) $c = -\frac{\pi}{(n-1)!\cdot \mathrm{Vol}_g(X)} \cdot \mu_g(D)$ *if* g *is Gauduchon.*
ii) $\deg_g(D) = 0$ *if and only if* $c = 0$.

Proof: i) is an immediate consequence of Lemma 2.6.
ii) If g is Gauduchon, then this follows from i). If g is arbitrary, then let $\tilde{g} = \varphi \cdot g$ be the Gauduchon metric in its conformal class such that $\deg_g = \deg_{\tilde{g}}$. Now we have

$$i\Lambda_g G_h = 0 \Longleftrightarrow i\Lambda_{\tilde{g}} G_h = 0 \Longleftrightarrow \deg_{\tilde{g}}(D) = 0 \Longleftrightarrow \deg_g(D) = 0.$$

□

Remark 3.5 *i) If two flat connections* D_1,D_2 *are isomorphic via the automorphism* f *of* E, *i.e. if* $D_2 \circ f - f \circ D_1 = 0$, *and if* h *is a* g*-Einstein metric in* (E, D_1), *then* $f_* h$ *is a* g*-Einstein metric in* (E, D_2) *with the same Einstein constant.*
ii) By Remark 2.3, a necessary condition for $d''_h = I_h(D)$ *to be an integrable Higgs operator is that* h *is a* g*-Einstein metric for* D *with Einstein constant* $c = 0$, *so in particular* $\deg_g(D) = 0$. *On the other hand it holds* $d^2 = -\Theta \wedge \Theta$ *(Lemma 2.4), and, if* d''_h *is integrable,* $\theta \wedge \theta = 0$ *implying* $\theta^* \wedge \theta^* = 0$. *This gives* $\mathrm{tr}(d^2) = -\mathrm{tr}[\theta, \theta^*] = 0$, *which implies* $\deg_g(d''_h) = 0$.
iii) For complex vector bundles on compact Riemannian manifolds (X, g), *Corlette defines a* g*-harmonic metric for a flat connection by the condition* $d^*(\Theta) = 0$ *([C]). If* X *is complex and* g *is a Kähler metric, then the* g*-degree of any flat connection vanishes, so in this context* g*-harmonic is the same as* g*-Einstein (see Remarks 2.5 and 3.1), but in general the two notions are different.*

Now we prove a useful Vanishing Theorem.

Proposition 3.6 *Let* D *be a flat connection in* E, *and* h *a* g*-Einstein metric in* (E, D) *with Einstein constant* c.
If $c > 0$, *then the only section* $s \in A^0(E)$ *with* $D(s) = 0$ *is* $s = 0$.
If $c = 0$, *then for every section* $s \in A^0(E)$ *with* $D(s) = 0$ *it holds* $\bar{\partial}(s) = \theta(s) = 0$ *and* $\partial(s) = \theta^*(s) = 0$, *so in particular* $d''_h(s) = 0$.

Proof: $D(s) = 0$ is equivalent to

$$\partial(s) = -\theta(s), \bar{\partial}(s) = -\theta^*(s); \tag{1}$$

this implies

$$\bar{\partial}\partial h(s, s) = -h(\bar{\partial} \circ \theta(s), s) - h(\theta(s), \theta(s)) + h(\bar{\partial}(s), \bar{\partial}(s)) - h(s, \partial \circ \theta^*(s)). \tag{2}$$

The assumption that h is g-Einstein means $i\Lambda_g \bar{\partial}(\theta) = i\Lambda_g G_h = c\cdot \mathrm{id}_E$, which is equivalent to $i\Lambda_g \partial(\theta^*) = -c\cdot \mathrm{id}_E$ since $(i\Lambda_g \bar{\partial}(\theta))^* = -i\Lambda_g(\bar{\partial}(\theta)^*) = -i\Lambda_g \partial(\theta^*)$; these relations can be rewritten as

$$i\Lambda_g \bar{\partial} \circ \theta = -i\Lambda_g \theta \circ \bar{\partial} + c \cdot \mathrm{id}_E, \quad i\Lambda_g \partial \circ \theta^* = -i\Lambda_g \theta^* \circ \partial - c \cdot \mathrm{id}_E. \tag{3}$$

Using (1) and (3) we get

$$\begin{aligned} i\Lambda_g h(\bar\partial\circ\theta(s),s) &= -i\Lambda_g h(\theta\circ\bar\partial(s),s)+c\cdot|s|^2 = i\Lambda_g h(\bar\partial(s),\theta^*(s))+c\cdot|s|^2 \\ &= -i\Lambda_g h(\bar\partial(s),\bar\partial(s))+c\cdot|s|^2 = |\bar\partial(s)|^2+c\cdot|s|^2, \end{aligned}$$

and similarly

$$i\Lambda_g h(s,\partial\circ\theta^*(s)) = |\theta(s)|^2+c\cdot|s|^2,$$

so (2) implies

$$i\Lambda_g\bar\partial\partial h(s,s) = -2\left(|\bar\partial(s)|^2+|\theta(s)|^2+c\cdot|s|^2\right).$$

Since the image of the operator $i\Lambda_g\bar\partial\partial$ on real functions contains no non-zero functions of constant sign ([LT] Lemma 7.2.7), this gives $s=0$ in the case $c>0$, and if $c=0$ we get $\bar\partial(s)=\theta(s)=0$, implying $\partial(s)=\theta^*(s)=0$ because of (1). □

The following corollary will be used later in the context of moduli spaces.

Corollary 3.7 *For $i=1,2$ let $D_i\in\mathcal{A}_f(E)$ be a flat connection, h_i a g-Einstein metric in (E,D_i), and $d_i'':=I_{h_i}(D_i)\in\mathcal{A}''(E)$ the associated Higgs operator. If D_1 and D_2 are isomorphic via the automorphism f of E, then d_1'' and d_2'' are isomorphic via f, too.*

Proof: Let h be the metric in $\mathrm{End}E=E^*\otimes E$ induced by the dual metric of h_1 in E^* and h_2 in E, and D the connection in $\mathrm{End}E$ defined by $D(f)=D_2\circ f-f\circ D_1$ for all $f\in A^0(\mathrm{End}E)$. Then D is flat of g-degree 0 since D_1 and D_2 are flat of equal degree, and h is a g-Einstein metric in $(\mathrm{End}E,D)$ with Einstein constant $c=0$ (compare Remark 2.3). Furthermore, the Higgs operator d'' in $\mathrm{End}E$ defined by $d''(f)=d_2''\circ f-f\circ d_1''$ equals $I_h(D)$. Hence Proposition 3.6 implies that an automorphism f of E with $D(f)=0$ also satisfies $d''(f)=0$. □

If $F\subset E$ is a D-invariant subbundle of E, then it is obvious that flatness of D implies flatness of $D|_F$, and hence the following definition makes sense.

Definition 3.8 *A flat connection D in E is called g-(semi)stable iff for every proper D-invariant subbundle $0\neq F\subset E$ it holds $\mu_g(D|_F)>\mu_g(D)$ $(\mu_g(D|_F)\geq\mu_g(D))$. D is called g-polystable iff $E=E_1\oplus E_2\oplus\ldots\oplus E_k$ is a direct sum of D-invariant and g-stable subbundles E_i with $\mu_g(D|_{E_i})=\mu_g(D)$ for $i=1,2,\ldots,k$.*

Remark 3.9 *i) Let D be a flat connection in E, and $0\neq F\subset E$ a proper D-invariant subbundle. Then g-stability of D implies $\mu_g(D|_F)>\mu_g(D)$ and hence the g-instability of the holomorphic structure D'' in E (in the sense of e.g. [LT]) since F is a D''-holomorphic subbundle of E.*
ii) Suppose that g is a Kähler metric; then $\deg_g(D)=0$ for every flat connection D (Remark 2.5). Hence a flat connection D in E is

- always g-semistable,
- g-stable if and only if E has no proper non-trivial D-invariant subbundle,
- g-polystable if E is a direct sum of D-invariant g-stable subbundles.
This means that g-(poly-)stability on a Kähler manifold coincides with (poly-)stability in the sense of Corlette [C].
iii) It is obvious that stability in the sense of Corlette always implies g-stability, but at the end of this section we will give an example of a g-stable bundle which is not stable in the sense of Corlette.

Definition 3.10 *A flat connection D in E is simple if the only D-parallel endomorphisms f, i.e. those with $D_{\mathrm{End}}(f) = D \circ f - f \circ D = 0$, are the homotheties $f = a \cdot \mathrm{id}_E,\ a \in \mathbb{C}$.*

Let D be a flat connection in E, $0 \neq F \subset E$ a D-invariant subbundle, and $Q := {}^{E}/_{F}$ the quotient with natural projection $\pi : E \longrightarrow Q$. Then D induces a flat connection D_Q in Q such that $D_Q \circ \pi = \pi \circ D$. In particular, F is a holomorphic subbundle of (E, D''), and D''_Q is the induced holomorphic structure in Q. Since the g-degree of a flat connection D by definition equals the g-degree of the associated holomorphic structure D'', it follow $\deg_g(D) = \deg_g(D_1) + \deg_g(D_Q)$. Hence as in the case of holomorphic bundles one verifies (compare [K] Chapter V)

Proposition 3.11 *i) A flat connection D in E is g-(semi)stable if and only if for every D-invariant proper subbundle $0 \neq F \subset E$ with quotient $Q = {}^{E}/_{F}$ it holds $\mu_g(D_Q) < \mu_g(D)$ (resp. $\mu_g(D_Q) \leq \mu_g(D)$.)*
ii) Let (E_1, D_1) and (E_2, D_2) be g-stable flat bundles over X with $\mu_g(D_1) = \mu_g(D_2)$. If $f \in A^0(\mathrm{Hom}(E_1, E_2))$ satisfies $D_2 \circ f = f \circ D_1$, then either $f = 0$ or f is an isomorphism.
iii) A g-stable flat connection D in E is simple.

Next we prove the first half of the main result of this section.

Proposition 3.12 *Let D be a flat connection in E, and h a g-Einstein metric in (E, D) with Einstein constant c; then D is g-semistable. If D is not g-stable, then D is g-polystable; more precisely, $E = E_1 \oplus E_2 \oplus \ldots \oplus E_k$ is a h-orthogonal direct sum of D-invariant g-stable subbundles such that $\mu_g(D|_{E_i}) = \mu_g(D)$ for $i = 1, 2, \ldots, k$. Furthermore, $h|_{E_i}$ is a g-Einstein metric in $(E_i, D|_{E_i})$ with Einstein constant c for all i, and the direct sum is invariant with respect to the Higgs operator $d''_h = I_h(D)$.*

Proof: First we consider the case when g is a Gauduchon metric. Let $0 \neq F \subset E$ be a D-invariant proper subbundle of rank s; then $E = F \oplus F^{\perp}$, where $F^{\perp}$ is the h-orthogonal complement of F. With respect to this decomposition, we write operators as 2×2 matrices, so D has the form

$$D = \begin{pmatrix} D_1 & A \\ 0 & D_2 \end{pmatrix},$$

where $D_1 = D|_F$ and D_2 is a flat connection in $F^\perp$. We use notations as in section 2; it is easy to see that the operator δ associated to D by h has the form

$$\delta = \begin{pmatrix} \delta_1 & 0 \\ A^* & \delta_2 \end{pmatrix},$$

where the δ_i are the operators associated to the D_i by h. Similarly it holds

$$\bar\partial = \frac{1}{2}(D'' + \delta'') = \frac{1}{2}\begin{pmatrix} D_1'' + \delta_1'' & A'' \\ A'^* & D_2'' + \delta_2'' \end{pmatrix} = \begin{pmatrix} \bar\partial_1 & \frac{1}{2}A'' \\ \frac{1}{2}A'^* & \bar\partial_2 \end{pmatrix},$$

and

$$\theta = \frac{1}{2}(D' - \delta') = \begin{pmatrix} D_1' - \delta_1' & A' \\ -A''^* & D_2' - \delta_2' \end{pmatrix} = \begin{pmatrix} \theta_1 & \frac{1}{2}A' \\ -\frac{1}{2}A''^* & \theta_2 \end{pmatrix},$$

where A' resp. A'' is the part of A of type $(1,0)$ resp. $(0,1)$. This implies

$$\begin{aligned} \bar\partial(\theta) &= [\bar\partial, \theta] \\ &= \begin{pmatrix} \bar\partial_1(\theta_1) + \frac{1}{4}(A' \wedge A'^* - A'' \wedge A''^*) & * \\ * & \bar\partial_2(\theta_2) + \frac{1}{4}(A'^* \wedge A' - A''^* \wedge A'') \end{pmatrix}, \end{aligned}$$

hence

$$\begin{aligned} c \cdot \mathrm{id}_E &= i\Lambda_g G_h \\ &= \begin{pmatrix} i\Lambda_g G_{1,h} + \frac{i}{4}\Lambda_g(A' \wedge A'^* - A'' \wedge A''^*) & * \\ * & i\Lambda_g G_{2,h} + \frac{i}{4}\Lambda_g(A'^* \wedge A' - A''^* \wedge A'') \end{pmatrix}, \end{aligned} \tag{4}$$

and thus

$$sc = \mathrm{tr}(i\Lambda_g G_{1,h} + \frac{i}{4}\Lambda_g(A' \wedge A'^* - A'' \wedge A''^*)) = i\mathrm{tr}\Lambda_g G_{1,h} + \frac{1}{4}|A|^2 .$$

Using Lemma 2.6 and Lemma 3.4 we conclude

$$\mu_g(D_1) = -\frac{i}{sn\pi}\int_X \mathrm{tr}\Lambda_g G_{1,h} \cdot \omega_g^n \geq -\frac{c(n-1)!}{\pi}\mathrm{Vol}_g(X) = \mu_g(D); \tag{5}$$

this prove that D is g-semistable.
If D is not g-stable, then there exists a subbundle F as above such that equality holds in (5), which implies $A = 0$. This means not only that $F^\perp$ is D-invariant, too, with $D|_{F^\perp} = D_2$, but also that

$$i\Lambda_g G_{1,h} = c \cdot \mathrm{id}_F, i\Lambda_g G_{2,h} = c \cdot \mathrm{id}_{F^\perp}$$

by (4). Hence the restriction of h to F resp. $F^\perp$ is g-Einstein for D_1 resp. D_2, and it holds $\mu_g(D_1) = \mu_g(D) = \mu_g(D_2)$ by Lemma 3.4. Furthermore, the D-invariance of F means that the inclusion $i : F \hookrightarrow E$ is parallel with respect to the flat connection in $\mathrm{Hom}(F, E)$ induced by D_1 and D. Using Remark 2.3

and Proposition 3.6 as in the proof of Corollary 3.7, we conclude that i is also parallel with respect to the associated Higgs operator, i.e. that F is d''_h-invariant; the same argument works for $F^\perp$. If D_1 and D_2 are stable, then we are done; otherwise the proof is finished by induction on the rank.

Now let g be arbitrary, let $\tilde{g}$ be the Gauduchon metric in its conformal class with $\deg_g = \deg_{\tilde{g}}$, and let $\tilde{h}$ be a $\tilde{g}$-Einstein metric in the conformal class of h, which exists by Lemma 3.3; then the theorem holds for $\tilde{g}$ and $\tilde{h}$. Since g and $\tilde{g}$ define the same degree and slope, and hence stability, it follows that D is $\tilde{g}$-semistable. If D is not g-stable, then there exists a D-invariant proper subbundle F as above with $\mu_{\tilde{g}}(D_1) = \mu_g(D_1) = \mu_g(D) = \mu_{\tilde{g}}(D)$. Note that the h-orthogonal complement $F^\perp$ of F is also the $\tilde{h}$-orthogonal complement, since h and $\tilde{h}$ are conformally equivalent. Hence, using $\tilde{g}$ and $\tilde{h}$ we conclude as above that $D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$ with respect the decomposition $E = F \oplus F^\perp$; now we can proceed as in the Gauduchon case. □

Another consequence of Proposition 3.6 is

Proposition 3.13 *Let D be a simple flat connection in E. If a g-Einstein metric in (E, D) exists, then it is unique up to a positive scalar.*

Proof: Let h_1,h_2 be g-Einstein metrics in (E, D), and $c \in \mathbb{R}$ the Einstein constant. There are differentiable automorphisms f and k of E, selfadjoint with respect to both h_1 and h_2, such that $f = k^2$ and $h_2(s, t) = h_1(f(s), t) = h_1(k(s), k(t))$ for all $s, t \in A^0(E)$. Since D is simple it suffices to show $D(f) = 0$. We define a new flat connection $\tilde{D} := k \circ D \circ k^{-1}$. In what follows, operators δ, d, Θ etc. with a subscript i are associated to D by the metric h_i, without a subscript they are associated to $\tilde{D}$ by h_1. One verifies

$$\delta_2 = f^{-1} \circ \delta_1 \circ f, \delta = k^{-1} \circ \delta_1 \circ k = k \circ \delta_2 \circ k^{-1},$$

implying

$$d = \frac{1}{2}(\tilde{D} + \delta) = k \circ d_2 \circ k^{-1}, \Theta = \frac{1}{2}(\tilde{D} - \delta) = k \circ \Theta_2 \circ k^{-1}$$

and hence

$$i\Lambda_g G_{h_1} = i\Lambda_g \bar{\partial}(\theta) = ik \circ \Lambda_g \bar{\partial}_2(\theta_2) \circ k^{-1} = ik \circ \Lambda_g G_{2,h_2} \circ k^{-1} = c \cdot \mathrm{id}_E,$$

so h_1 is a g-Einstein metric in $(E, \tilde{D})$. It follows that h_1 induces a g-Einstein metric with Einstein constant 0 for the flat connection $\tilde{D}_{\mathrm{End}}(.) = . \circ D - \tilde{D} \circ .$ in End$E$. By definition it holds $\tilde{D}_{\mathrm{End}}(k) = 0$, so Proposition 3.6 implies $\tilde{d}_{\mathrm{End}}(k) = 0$. Since $\tilde{\delta}_{\mathrm{End}} = 2\tilde{d}_{\mathrm{End}} - \tilde{D}_{\mathrm{End}}$, it follows

$$0 = \tilde{\delta}_{\mathrm{End}}(k) = k \circ \delta_1 - \delta \circ k = k \circ \delta_1 - k^{-1} \circ \delta_1 \circ k^2 = k^{-1} \circ (f \circ \delta_1 - \delta_1 \circ f) \ ,$$

implying $\delta_{1,\mathrm{End}}(f) = 0$, where $\delta_{1,\mathrm{End}}$ is the operator on $\mathrm{End}E$ induced by D and h_1. But this is equivalent to $\delta'_{1,\mathrm{End}}(f) = 0$ and $\delta''_{1,\mathrm{End}}(f) = 0$, and taking adjoints with respect to h_1 we get

$$0 = (\delta'_{1,\mathrm{End}}(f))^* = D''_{\mathrm{End}}(f), 0 = (\delta''_{1,\mathrm{End}}(f))^* = D'_{\mathrm{End}}(f),$$

i.e. $D_{\mathrm{End}}(f) = 0$. □

Let (E, D), $(\tilde{E}, \tilde{D})$ be flat bundles with g-Einstein metrics h, $\tilde{h}$. Let $E = \bigoplus_{i=1}^{k} E_i$ and $\tilde{E} = \bigoplus_{i=1}^{l} \tilde{E}_i$ be the orthogonal, invariant splittings given by Proposition 3.12. We write $D_i := D|_{E_i}$, $\tilde{D}_i := \tilde{D}|_{\tilde{E}_i}$, $h_i := h|_{E_i}$, $\tilde{h}_i := \tilde{h}|_{\tilde{E}_i}$. Using Propositions 3.11 and 3.13 one verifies

Corollary 3.14 *If there exists an isomorphism $f \in A^0(\mathrm{Hom}(E, \tilde{E}))$ satisfying $f \circ D = \tilde{D} \circ f$, then it holds $k = l$, and, after renumbering of the summands if necessary, there are isomorphisms $f_i \in A^0(\mathrm{Hom}(E_i, \tilde{E}_i))$ such that $f_i \circ D_i = \tilde{D}_i \circ f$ and $f_*(h_i) = \tilde{h}_i$.*

The following result is the converse of Proposition 3.12.

Proposition 3.15 *Let (E, D) a g-stable flat bundle over X. Then there exists a g-Einstein metric for (E, D).*

Sketch of proof: The proof is very similar to the one for the existence of a g-Hermitian Einstein metric in a g-stable holomorphic vector bundle as given in Chapter 3 of [LT]. Therefore we will be brief, leaving it to the reader to fill in the necessary details.
First observe that by Lemma 3.3 we may assume that g is a Gauduchon metric. For any metric h in E it holds

$$G_h = \bar{\partial}_h(\theta_h) = \frac{1}{4}[D'' + \delta''_h, D' - \delta'_h] = -\frac{1}{4}[D'', \delta'_h] + \frac{1}{4}[D', \delta''_h]$$

since $D^2 = \delta_h^2 = 0$. Observe that $[D'', \delta'_h]$ resp. $[D', \delta''_h]$ is the curvature of the h-unitary connection $D'' + \delta'_h$ resp. $D' + \delta''_h$.
Fix a metric h_0 in E, and let $\delta = \delta' + \delta''$, $d = \partial + \bar{\partial}$, $\Theta = \theta + \theta^*$ be the operators associated to $D = D' + D''$ and h_0 as in section 2. Consider for an h_0-selfadjoint positive definite endomorphism f of E and $\varepsilon \in [0, 1]$ the differential equation

$$L_\varepsilon(f) := K^0 - \frac{i}{4}\Lambda_g D''(f^{-1} \circ \delta'(f)) + \frac{i}{4}\Lambda_g D'(f^{-1} \circ \delta''(f)) - \varepsilon \cdot \log(f) = 0, \quad (6)$$

where $K^0 := i\Lambda_g \bar{\partial}(\theta) - c \cdot \mathrm{id}_E = -\frac{i}{4}\Lambda_g([D'', \delta'] - [D', \delta'']) - c \cdot \mathrm{id}_E$, and c is the constant associated to a possible g-Einstein metric for (E, D). The metric $f \cdot h_0$, defined by $f \cdot h_0(s, t) := h_0(f(s), t)$ for sections s, t in E, is g-Einstein if and only if $L_0(f) = 0$.

The term $T_1 := i\Lambda_g D''(f^{-1} \circ \delta'(f))$ (associated to the unitary connection $d_1 := \delta' + D''$) in equation (6) is of precisely the same type as the term $T_0 := i\Lambda_g \bar{\partial}(f^{-1} \circ \partial_0(f))$ (associated to the unitary connection $d_0 = \partial_0 + \bar{\partial}$) in equation $(**)$ on page 62 in [LT], and the term $T_2 := -i\Lambda_g D'(f^{-1} \circ \delta''(f))$ (associated to the unitary connection $d_2 := D' + \delta''$) is *almost* of this type; e.g. the trace of all three terms equals $i\Lambda_g \bar{\partial}\partial(\mathrm{tr}(\log f)$, and the symbols of the differential operators $\frac{d}{df}\hat{T}_i$, where $\hat{T}_i(f) := f \circ T_i(f)$, are equal, too. Therefore most of the arguments in [LT] can easily be adapted to show first that for a simple flat connection D equation (6) has solutions f_ε for all $\varepsilon \in (0,1]$, which satisfy $\det f_\varepsilon \equiv 1$, and which converge to a solution f of $L_0(f) = 0$ if the L^2-norms of the f_ε are uniformly bounded. (There are two places where one has to argue in a slightly different way: In the proof of the analogue of [LT] Lemma 3.3.1, one uses the Laplacian $\Delta_D = D^* \circ D$ instead of $\Delta_{\bar{\partial}}$, and in the proof of the analogue of [LT] Proposition 3.3.5 the *sum* $\Delta_{d_1} + \Delta_{d_2}$ of the two Laplacians associated to d_1 and d_2 instead of just one.)
Then, under the assumptions that $\mathrm{rk}E \geq 2$ and that the L^2-norms of the f_ε are unbounded, one shows that for suitable $\varepsilon_i \longrightarrow 0$, $\rho(\varepsilon_i) \longrightarrow 0$, the limit

$$\pi := \mathrm{id}_E - \lim_{\sigma \longrightarrow 0} \left(\lim_{i \longrightarrow \infty} \rho(\varepsilon_i) \cdot f_{\varepsilon_i} \right)^\sigma$$

exists weakly in L^2_1, and satisfies in L^1 $\pi = \pi^* = \pi^2$ and

$$(\mathrm{id}_E - \pi) \circ D(\pi) = 0. \tag{7}$$

This implies $(\mathrm{id}_E - \pi) \circ D''(\pi) = 0$, so π defines a weakly holomorphic subbundle $\mathcal{F}$ of the holomorphic bundle (E, D'') by a theorem of Uhlenbeck and Yau (see [UY], [LT] Theorem 3.4.3). $\mathcal{F}$ is a coherent subsheaf of (E, D''), a holomorphic subbundle outside an analytic subset $S \subset X$ of codimension at least 2, and π is smooth on $X \setminus S$. Therefore (7) implies that $\mathcal{F}|_{X \setminus S}$ is in fact a D-invariant subbundle of $E|_{X \setminus S}$, which extends to a D-invariant subbundle F of E by the Lemma below. Again using arguments as is [LT], one finally shows that F violates the stability condition for (E, D). □

Lemma 3.16 *Let X be a differentiable manifold, E a differentiable vector bundle over X, and D a flat connection in E. Let $S \subset X$ be a subset such that $X \setminus S$ is open and dense in X, and with the following property: For every point $x \in S$ and every open neighborhood U of x in X there exists an open neighborhood $x \in U' \subset U$ such that $U' \setminus S$ is path-connected.*
Then every D-invariant subbundle $\mathcal{F}$ of $E|_{X \setminus S}$ extends to a D-invariant subbundle F of E.

Proof: For every $x \in S$ choose an open neighborhood $x \in U \subset X$ such that $U \setminus S$ is path connected and $(E|_U, D) \cong (U \times V, d)$, where V is a vector space and d the trivial flat connection. Since $\mathcal{F}$ is D-invariant and $U \setminus S$ is path connected, it holds

$$(\mathcal{F}|_{U \setminus S}, D) \cong ((U \setminus S) \times W, d),$$

where $W \subset V$ is a constant subspace. Define F over U by $F|_U :\cong U \times W$; then the topological condition on S implies that this is well defined on S, and hence gives a D-invariant extension F of $\mathcal{F}$ over X. □

The following main result of this section is a direct consequence of Propositions 3.12 and 3.15.

THEOREM 3.17 *A flat connection D in E admits a g-Einstein metric if and only if it is g-polystable.*

As for stable vector bundles and Hermitian-Einstein metrics, the gauge theoretic interpretation of our results is as follows. The group

$$\mathcal{G}^{\mathbb{C}} := A^0(GL(E))$$

of differentiable automorphisms of E acts on $\mathcal{A}(E)$ by $D \cdot f = f^{-1} \circ D \circ f$, so

$$\mathcal{A}(E)/_{\mathcal{G}^{\mathbb{C}}}$$

is the moduli space of isomorphism classes of connections in E. Observe that flatness, simplicity and g-stability are preserved under this action. Fix a metric h in E; then it holds:

COROLLARY 3.18 *The following two statements for a flat connection D are equivalent:*
i) D is g-stable.
ii) D is simple, and there is a connection D_0 in the $\mathcal{G}^{\mathbb{C}}$-orbit through D such that h is g-Einstein for D_0.

The essential uniqueness of a g-Einstein metric (Proposition 3.13) implies that the connection D_0 in ii) is unique up to the action of the subgroup

$$\mathcal{G} := A^0(U(E,h)) \subset \mathcal{G}^{\mathbb{C}}$$

of h-unitary automorphisms. This means that the moduli space

$$\mathcal{M}_f^{\mathrm{st}}(E) = {\{\ D \in \mathcal{A}_f(E) \mid D \text{ is } g-\ \text{ stable}\ \}}/_{\mathcal{G}^{\mathbb{C}}}$$

of isomorphism classes of g-stable flat connections in E coincides with the quotient

$$\{\ D \in \mathcal{A}_f(E) \mid D \text{ is simple and } h \text{ is } g-\text{Einstein for } D\ \}/_{\mathcal{G}}.$$

EXAMPLE: We now give the promised example of a flat bundle which is g-stable, but not stable in the sense of Corlette.

An Inoue surface of type $S_N^\pm$ is the quotient of $\mathbb{H}\times\mathbb{C}$ by an affine transformation group G generated by

$$\begin{aligned} g_0(w,z) &:= (\alpha w, \pm z + t), \\ g_i(w,z) &:= (w + a_i, z + b_i w + c_i),\ i = 1,2, \\ g_3(w,z) &:= (w, z + c_3), \end{aligned}$$

with certain constants $\alpha, a_i, b_i, c_3 \in \mathbb{R},\ \ c_1, c_2 \in \mathbb{C}$ (see [P] p. 160). Since the second Betti number of $S_N^\pm$ vanishes, the degree map

$$\deg_g : \mathrm{Pic}(S_N^\pm) \longrightarrow \mathbb{R}$$

associated to a Gauduchon metric g is, up to a positive factor, independent of the chosen metric g. In particular, *all* Hermitian metrics g define the same notion of g-stability ([LT] Remark 1.4.4 iii)).
The trivial flat connection d on $\mathbb{H}\times\mathbb{C}$ induces a flat connection D in the tangent bundle $E := T_{S_N^\pm}$. A D-invariant sub-line bundle of E is in particular a holomorphic subbundle, so it defines a holomorphic foliation of $S_N^\pm$. According to [B] Théorème 2, there is precisely one such foliation, namely the one induced by the G-invariant vertical foliation (i.e. with leaves $\{w\}\times\mathbb{C}$) of $\mathbb{H}\times\mathbb{C}$. The corresponding trivial line bundle L_0 on $\mathbb{H}\times\mathbb{C}$ is d-invariant, so it descends to a unique D-invariant subbundle L of E; this shows that E is not stable in the sense of Corlette. Observe that L has factors of automorphy $\chi(g_i) = \pm 1,\ i = 0,1,2,3$, so the standard flat metric in L_0 defines a metric h in L such that the associated Chern connection in $(L, D''|_L)$ is flat; this implies $\mu_g(D|_L) = \deg_g(D|_L) = 0$. On the other hand, the g-degree, and hence the g-slope, of E is negative by [P] Proposition 4.7; this implies the g-stability of E since L is the only D-invariant proper subbundle of E.

4 Einstein metrics and stability for Higgs bundles.

Again we fix Hermitian metrics g in X and h in E.
Let $d'' = \bar\partial + \theta \in \mathcal{A}_i''(E)$ be an integrable Higgs operator,

$$D_h = I_h^{-1}(d'') = d + \Theta = \partial + \bar\partial + \theta + \theta^* \in \mathcal{A}(E)$$

the connection associated to d'' as in section 2, and $F_h = D_h^2$ its curvature.

Definition 4.1 *h is called a g-Einstein metric in (E, d'') if and only if $K_h := i\Lambda_g F_h = c\cdot \mathrm{id}_E$ with a real constant c, the Einstein constant.*

Lemma 4.2 *Let h be a g-Einstein metric in (E, d''), and $\tilde g = \varphi\cdot g$ conformally equivalent to g. Then there exists a $\tilde g$-Einstein metric $\tilde h$ in (E, d'') which is conformally equivalent to h.*

Proof: From Remark 2.1 iii) it follows that for $f \in \mathcal{C}^\infty(X,\mathbb{R})$ it holds $F_{e^f\cdot h} = F_h + \bar\partial\partial(f)\cdot \mathrm{id}_E$. Using this, the proof is analogous to that of Lemma 3.3. □

Notice that since d'' is integrable it holds (compare Lemma 2.4)

$$K_h = i\Lambda_g(d^2 + [\theta,\theta^*]) = i\Lambda_g([\partial,\bar\partial] + [\theta,\theta^*])$$

where $d = \partial + \bar\partial$. An immediate consequence of Lemma 2.6 and Lemma 4.2 is (compare the proof of Lemma 3.4)

Lemma 4.3 *If $i\Lambda_g F_h = c\cdot \mathrm{id}_E$ with $c \in \mathbb{R}$, then it holds:*
i) $c = \frac{2\pi}{(n-1)!\cdot \mathrm{Vol}_g(X)}\cdot \mu_g(d'')$ if g is Gauduchon.
ii) $\deg_g(d'') = 0$ if and only if $c = 0$.

Remark 4.4 *(compare Remark 3.5)*
i) If two integrable Higgs operators d''_1,d''_2 are isomorphic via the automorphism f of E, i.e. if $d''_2\circ f - f\circ d''_1 = 0$, and if h is a g-Einstein metric in (E, d''_1), then f_*h is a g-Einstein metric in (E, d''_2), and the associated Einstein constants are equal.
ii) By Remark 2.3, a necessary condition for $D_h = I_h(d'')$ to be a flat connection is h to be Einstein with Einstein constant $c = 0$, so in particular $\deg_g(d'') = 0$. On the other hand, the Chern connection in (E, D''_h) is $\partial-\theta+\bar\partial+\theta^*$, so the g-degree of D_h is obtained by integrating $\mathrm{tr}\Lambda_g[\bar\partial+\theta^*, \partial-\theta]$ which equals $\mathrm{tr}\Lambda_g[\bar\partial, \partial]$ since d'' is integrable (Lemma 2.4 ii)). If D_h is flat, we furthermore have $d^2 = -\Theta\wedge\Theta$ (Lemma 2.4 i)), implying $\mathrm{tr}[\bar\partial,\partial] = 0$ and hence $\deg_g(D_h) = 0$.

In analogy with the case of Hermitian-Einstein metrics in holomorphic vector bundles, the following vanishing theorem holds.

Proposition 4.5 *Let h be a g-Einstein metric in (E, d'') with Einstein constant c.*
If $c < 0$, then the only section $s \in A^0(E)$ with $d''(s) = 0$ is $s = 0$.
If $c = 0$, then for every section $s \in A^0(E)$ with $d''(s) = 0$ it holds $D_h(s) = 0$.

Proof: For $s \in A^0(E)$, $d''(s) = 0$ is equivalent to $\bar\partial(s) = 0 = \theta(s)$. This implies

$$c\cdot|s|^2 = c\cdot h(s,s) = h(K_h(s),s) = i\Lambda_g\left(h(\bar\partial\partial(s),s) + h(\theta^*(s),\theta^*(s))\right). \quad (8)$$

We have

$$i\Lambda_g\bar\partial\partial h(s,s) = i\Lambda_g\left(h(\bar\partial\partial(s),s) - h(\partial(s),\partial(s))\right)$$

since $\bar\partial(s) = 0$, and using (8) we get

$$i\Lambda_g\bar\partial\partial h(s,s) = c\cdot|s|^2 - |\partial(s)|^2 - |\theta^*(s)|^2.$$

Now the claim follows as in the proof of Proposition 3.6. □

The proof of the following corollary is analogous to that of Corollary 3.7.

Corollary 4.6 *For $i = 1, 2$ let $d''_i \in \mathcal{A}''_i(E)$ be an integrable Higgs operators, h_i a g-Einstein metric in (E, d''_i), and $D_i := I^{-1}_{h_i}(d''_i) \in \mathcal{A}(E)$ the associated connection. If d''_1 and d''_2 are isomorphic via the automorphism f of E, then D_1 and D_2 are isomorphic via f, too.*

Let $d'' = \bar{\partial} + \theta$ be an integrable Higgs operator in E. A coherent subsheaf $\mathcal{F}$ of the holomorphic bundle $(E, \bar{\partial})$ is called a *Higgs-subsheaf* of (E, d'') iff it is d''-invariant. For the definition of the g-degree and g-slope of a coherent sheaf see [LT].

Definition 4.7 *An integrable Higgs operator d'' in E is called g-(semi)stable iff for every coherent Higgs-subsheaf $\mathcal{F}$ of (E, d'') with $0 < \mathrm{rk}\mathcal{F} < \mathrm{rk}E$ it holds $\mu_g(\mathcal{F}) < \mu_g(E)$ ($\mu_g(\mathcal{F}) \leq \mu_g(E)$). d'' is called g-polystable iff E is a direct sum $E = E_1 \oplus E_2 \oplus \ldots \oplus E_k$ of d''-invariant and g-stable subbundles E_i with $\mu_g(d''|_{E_i} = \mu_g(d'')$ for $i = 1, 2, \ldots, k$.*

Definition 4.8 *An integrable Higgs operator d'' in E is called simple iff for every $f \in A^0(\mathrm{End}E)$ with $d'' \circ f = f \circ d''$ it holds $f = a \cdot \mathrm{id}_E$ with $a \in \mathbb{C}$.*

As in the case of stable vector bundles or flat connections, (semi)-stability can equivalently be defined using quotients of E; again it follows

Lemma 4.9 *i) A g-stable integrable Higgs operator in E is simple.*
ii) Let d''_1, d''_2 be g-stable integrable Higgs operators in bundles E_1, E_2 on X such that $\mu_g(d''_1) = \mu_g(d''_2)$. If $f \in A^0(\mathrm{Hom}(E_1, E_2))$ satisfies $d''_2 \circ f = f \circ d''_1$, then either $f = 0$ or f is an isomorphism.

Furthermore, using arguments similar to those in the proof of Proposition 3.13, we get the following consequence of Proposition 4.5.

Proposition 4.10 *Let d'' be a simple integrable Higgs operator in E. If a g-Einstein metric in (E, d'') exists, then it is unique up to a positive scalar.*

The proof of the next result is a straightforward generalization of that in the Kähler case [S2] (just as for the proof of the corresponding statement for Hermite-Einstein metrics in vector bundles, see [LT]).

Proposition 4.11 *Let d'' be an integrable Higgs operator in E, and h a g-Einstein metric in (E, d'') with Einstein constant c; then d'' is g-semistable. If d'' is not g-stable, then d'' is g-polystable; more precisely, $E = E_1 \oplus E_2 \oplus \ldots \oplus E_k$ is an h-orthogonal direct sum of d''-invariant and g-stable subbundles such that $\mu_g(d''|_{E_i}) = \mu_g(d'')$ for $i = 1, 2, \ldots, k$. Furthermore, $h|_{E_i}$ is a g-Einstein metric in $(E_i, d''|_{E_i})$ with Einstein constant c for all i, and the direct sum is invariant with respect to the connection $D_h = I^{-1}_h(d'')$.*

Let d'', $\tilde{d}''$ be integrable Higgs operators in bundles E, $\tilde{E}$ with g-Einstein metrics h, $\tilde{h}$. Let $E = \bigoplus_{i=1}^{k} E_i$ and $\tilde{E} = \bigoplus_{i=1}^{l} \tilde{E}_i$ be the orthogonal, invariant splittings given by Proposition 4.11. We write $d_i'' := d''|_{E_i}$, $\tilde{d}_i'' := \tilde{d}''|_{\tilde{E}_i}$, $h_i := h|_{E_i}$, $\tilde{h}_i := \tilde{h}|_{\tilde{E}_i}$.

As in the previous section (but now using Lemma 4.9 and Proposition 4.10) we deduce

Corollary 4.12 *Suppose that there exists an isomorphism $f \in A^0(\mathrm{Hom}(E, \tilde{E}))$ satisfying $f \circ d'' = \tilde{d}'' \circ f$. Then it holds $k = l$, and, after renumbering of the summands if necessary, there are isomorphisms $f_i \in A^0(\mathrm{Hom}(E_i, \tilde{E}_i))$ such that $f_i \circ d_i'' = \tilde{d}_i'' \circ f$ and $f_*(h_i) = \tilde{h}_i$.*

Remark 4.13 *We expect that the existence of a g-Einstein metric for a g-stable Higgs operator d'' can be proved by solving (again using the continuity method as in [LT]) the differential equation*

$$K_h + i\Lambda_g d''(f^{-1} \circ d'(f)) = c \cdot \mathrm{id}_E$$

for a positive definite and h-selfadjoint endomorphism f of E, where h is a suitable fixed metric in E.

5 Surfaces.

In this section we consider the special case $n = 2$, i.e. where X is a compact complex surface; again we fix a Hermitian metric g in X. In this case, the real Chern numbers $c_1^2(E), c_2(E) \in H^4(X, \mathbb{R}) \cong \mathbb{R}$ can be calculated by integrating the corresponding Chern forms of any connection in E, independently of the chosen metric g. In particular, if E admits a flat connection, then these Chern numbers vanish.

Proposition 5.1 *Suppose that $D \in \mathcal{A}_f(E)$ is a flat connection of g-degree 0, and that h is a g-Einstein metric in (E, D). Then it holds $G_h = 0$. In particular, the Higgs operator d_h'' associated to D and h is integrable with $\deg_g(d_h'') = 0$, and h is a g-Einstein metric for (E, d_h'').*

Proof: (see [S2]) For $\epsilon > 0$ we define a new connection $B_\epsilon := d + \frac{1}{\epsilon}\theta + \epsilon\theta^*$, and $F_\epsilon := B_\epsilon^2$. Observe that $n = 2$ implies $F_\epsilon^2 = \frac{1}{\epsilon^2}\nabla_\epsilon^4$, where $\nabla_\epsilon = d_h'' + \epsilon d'$. The vanishing of the Chern numbers of E implies $\int_X \mathrm{tr} F_\epsilon^2 = 0$, and hence $\int_X \mathrm{tr}\nabla_\epsilon^4 = 0$ for all $\epsilon > 0$. Taking the limit $\epsilon \to 0$ it follows

$$\int_X \mathrm{tr} G_h^2 = 0. \tag{9}$$

Write $G_h = G_{1,1} + G_2$, where $G_{1,1}$ is the component of the 2-form G_h of type $(1,1)$. Then it holds

$$*_g G_{1,1} = -G_{1,1}, *_g G_2 = G_2; \tag{10}$$

the first equation is a consequence of $\Lambda_g G_h = 0$, which follows from the assumption and Lemma 3.4. On the other hand, it holds $G_h = \bar\partial^2 + \bar\partial(\theta) + \theta \wedge \theta$, so Lemma 2.4 implies

$$G_{1,1} = \bar\partial(\theta) = \partial(\theta^*)^* = -\bar\partial(\theta)^* = -G^*_{1,1}, \tag{11}$$

and

$$G_2 = \bar\partial^2 + \theta\wedge\theta = -\theta^*\wedge\theta^* - \theta\wedge\theta = (\theta\wedge\theta + \theta^*\wedge\theta^*)^* = G_2^*. \tag{12}$$

(11) and (12) combined with (10) give $*_g G_h^* = G_h$, so from (9) it follows

$$0 = \int_X \mathrm{tr} G_h^2 = \int_X \mathrm{tr}(G_h \wedge *_g G_h^*) = \int_X |G_h|^2 \mathrm{vol}_g,$$

implying $(d''_h)^2 = G_h = 0$. Hence d''_h is integrable, $\deg_g(d''_h)$ vanishes (Remark 3.5), and h is g-Einstein for $(E, d'')h)$ because the curvature of d''_h with respect to h equals $D^2 = 0$. □

Proposition 5.2 *Suppose that* $c_1^2(E) = c_2(E) = 0$, *that* d'' *is an integrable Higgs operator of g-degree 0, and that h is a g-Einstein metric in* (E, d''). *Then it holds* $F_h = 0$. *In particular, the connection* D_h *associated to* d'' *and h is flat with* $\deg_g(D_h) = 0$, *and h is a g-Einstein metric for* (E, D_h).

Proof: Define $F_{1,1} := d^2 + [\theta, \theta^*]$, $F_2 := \partial(\theta) + \bar\partial(\theta^*)$; then $F_h = F_{1,1} + F_2$. Observe that $F_{1,1}$ is of type (1,1) because d is a unitary connection in the holomorphic bundle $(E, \bar\partial)$. Since $\deg_g(d'') = 0$, Lemma 4.3 implies $0 = \Lambda_g F_h = \Lambda_g F_{1,1}$, hence it holds $*_g F_{1,1} = -F_{1,1}$ and $*_g F_2 = F_2$. On the other hand, it is easy to see that $F^*_{1,1} = -F_{1,1}$ and $F_2^* = F_2$. Combining these relations we get $*_g F_h^* = F_h$. Since $c_1^2(E)$ resp. $c_2(E)$ are obtained by integrating $-\frac{1}{4\pi^2}(\mathrm{tr} F_h)^2$ resp. $-\frac{1}{8\pi^2}((\mathrm{tr} F_h)^2 - \mathrm{tr}(F_h^2))$, we get

$$0 = \int_X \mathrm{tr}(F_h^2) = \int_X \mathrm{tr}(F_h \wedge *_g F_h^*) = \|F_h\|^2,$$

implying $D_h^2 = F_h = 0$. Hence D_h is flat, $\deg_g(D_h)$ vanishes (Remark 4.4), and h is g-Einstein for (E, D_h) because the pseudocurvature of D_h with respect to h equals $(d'')^2 = 0$. □

Remark 5.3 *The above proposition implies in particular the following: Suppose that* $c_1^2(E) = c_2(E) = 0$; *if there exists an integrable Higgs operator* d'' *in* E *with g-degree 0 admitting a g-Einstein metric, then the real Chern class* $c_1(E)_{\mathbb{R}} \in H^2(X, \mathbb{R})$ *vanishes, because there is a flat connection in E.*

We define $\mathcal{A}_f(E)^0_g$ to be the space of $D \in \mathcal{A}_f(E)$ of g-degree 0 such that there exists a g-Einstein metric in (E, D), and $\mathcal{A}''_i(E)^0_g$ to be the space of $d'' \in \mathcal{A}''_i(E)$ of g-degree 0 such that there exists a g-Einstein metric in (E, d''). By Remark 3.5 and Remark 4.4, the two moduli sets

$$\mathcal{M}_f(E)^0_g := \mathcal{A}_f(E)^0_g \big/ \text{isomorphy of connections}$$

and

$$\mathcal{M}''(E)^0_g := \mathcal{A}''_i(E)^0_g \big/ \text{isomorphy of Higgs operators}$$

are well defined. The main result of this section is

Theorem 5.4 *There is a natural bijection*

$$I : \mathcal{M}_f(E)^0_g \longrightarrow \mathcal{M}''(E)^0_g.$$

Proof: First observe that we may assume that the real Chern classes of E vanish, since otherwise both spaces are empty (see Remark 5.3).

Let D be a flat connection in E with g-degree 0, and h a g-Einstein metric in (E, D). By Proposition 5.1, the associated Higgs operator $d''_h = I_h(D)$ is integrable with g-degree 0, and h is a g-Einstein metric in (E, d''_h). We will show that the map I defined by $I([D]) := [d''_h]$ is well defined and bijective.

Suppose that $D, \tilde{D} \in \mathcal{A}_f(E)^0_g$ are isomorphic via the automorphism f of E; then f_*h is g-Einstein in $(E, \tilde{D})$ (Remark 3.5), the Higgs-operator $\tilde{d}''$ associated to $\tilde{D}$ and f_*h is isomorphic to d'' via f (Corollary 3.7), and f_*h is a g-Einstein metric in $(E, \tilde{d}'')$ (Remark 4.4). To prove that I is well defined it thus suffices to show that two different g-Einstein metrics $h, \tilde{h}$ for a fixed $D \in \mathcal{A}_f(E)^0_g$ produce isomorphic Higgs operators $d''_h, d''_{\tilde{h}}$. For this consider the D-invariant and h- resp. $\tilde{h}$-orthogonal splittings $E = \bigoplus_{i=1}^{k} E_i$ resp. $E = \bigoplus_{i=1}^{l} \tilde{E}_i$ associated to h resp. $\tilde{h}$ by Proposition 3.12. According to Corollary 3.14 (with $E = \tilde{E}$, $D = \tilde{D}$, $f = \mathrm{id}_E$) it holds $k = l$, and we may assume that there are isomorphisms $f_i : (E_i, D_i, h_i) \longrightarrow (\tilde{E}_i, \tilde{D}_i, \tilde{h}_i)$ of flat bundles of g-degree 0 with g-Einstein metrics, where $D_i := D|_{E_i}$, $\tilde{D}_i := D|_{\tilde{E}_i}$, $h_i := h|_{E_i}$, $\tilde{h}_i := \tilde{h}|_{\tilde{E}_i}$. This means in particular that the Higgs operator d''_i in E_i associated to D_i and h_i is isomorphic via f_i to the Higgs operator $\tilde{d}''_i$ in $\tilde{E}_i$ associated to $\tilde{D}_i$ and $\tilde{h}_i$. Hence $d''_h = d''_1 \oplus \dots d''_k$ is isomorphic to $d''_{\tilde{h}} = \tilde{d}''_1 \oplus \dots \oplus \tilde{d}''_k$ via the isomorphism $f := f_1 \oplus \dots \oplus f_k$.

In the same way, but using Proposition 5.2 and the results of section 4, one shows that there is a well defined map from $\mathcal{M}''(E)^0_g$ to $\mathcal{M}_f(E)^0_g$, associating to the class of an integrable Higgs operator d'' with g-Einstein metric h the class of the connection $D_h = I_h^{-1}(d'')$; this obviously is an inverse of I. □

6 Line bundles on non-Kähler surfaces.

Isomorphism classes of flat complex *line* bundles (L, D) on a manifold X are parametrized by $H^1(X, \mathbb{C}^*)$. On the other hand, an integrable Higgs operator $d'' = \bar{\partial} + \theta$ in a complex line bundle L consists of a holomorphic structure $\bar{\partial}$ in L and a holomorphic 1-form θ on X (the condition $\theta \wedge \theta = 0$ now is trivial). Furthermore, two integrable Higgs operators d''_1 and d''_2 in L are isomorphic if and only if the two holomorphic line bundles $(L, \bar{\partial}_1)$ and $(L, \bar{\partial}_2)$ are isomorphic and $\theta_1 = \theta_2$. Hence, the space parametrizing isomorphism classes of integrable Higgs operators is $H^1(X, \mathcal{O}^*) \oplus H^0(X, \Omega^1(X)) = \mathrm{Pic}(X) \oplus H^{1,0}(X)$. In particular, the moduli sets $\mathcal{M}_f(L)^0_g$ and $\mathcal{M}''(L)^0_g$ defined in the previous section are subsets of $H^1(X, \mathbb{C}^*)$ resp. $\mathrm{Pic}(X) \oplus H^{1,0}(X)$.

Lemma 6.1 *Let L be a complex line bundle on X, and g a Hermitian metric in X. Then every flat connection in L and every integrable Higgs operator in L admits a g-Einstein metric.*

Proof: Let h_0 be fixed metric in L, then every metric is of the form $h_f = e^f \cdot h_0$ with $f \in \mathcal{C}^\infty(X, \mathbb{R})$. Let D be a flat connection in L; then h_f is a g-Einstein metric for D if and only if it is a solution of the equation $i\Lambda_g G_{h_0} - \frac{i}{2}\Lambda_g \bar{\partial}\partial(f) = c$ with a real constant c. Such a solution exists by [LT] Corollary 7.2.9. A similar argument works for integrable Higgs operators. □

From now on let X be a *surface*, and g a fixed Hermitian metric in X. Then the map $\deg_g : \mathrm{Pic}(X) \longrightarrow \mathbb{R}$ is a morphism of Lie groups ([LT] Proposition 1.3.7; recall that $\deg_g = \deg_{\tilde{g}}$ for some Gauduchon metric $\tilde{g}$). We define

$$H^1(X, \mathbb{C}^*)^f := \{\ [(L, D)] \in H^1(X, \mathbb{C}^*) \mid \deg_g(D) = 0\ \},$$

$$\mathrm{Pic}(X)^T := \{\ [(L, \bar{\partial})] \in \mathrm{Pic}(X) \mid c_1(L)_{\mathbb{R}} = 0\ \},$$

and

$$\mathrm{Pic}(X)^f := \ker(\deg_g |_{\mathrm{Pic}(X)^T}).$$

Observe that $\mathrm{Pic}(X)^f$ can be identified with the set of isomorphism classes of line bundles admitting a flat *unitary* connection ([LT] Proposition 1.3.13). Theorem 5.4 and Lemma 6.1 imply

Proposition 6.2 *There is a natural bijection*

$$I_1 : H^1(X, \mathbb{C}^*)^f \longrightarrow \mathrm{Pic}(X)^f \times H^{1,0}(X).$$

If X admits a Kähler metric, i.e. if the first Betti number of $b_1(X)$ is even, then $\deg_g$ is a topological invariant for every metric g ([LT] Corollary 1.3.12 i)). Hence in this case it holds $H^1(X, \mathbb{C}^*)^f = H^1(X, \mathbb{C}^*)$ and $\mathrm{Pic}(X)^f = \mathrm{Pic}(X)^T$, and I_1 is the natural bijection from the moduli space of isomorphism classes of flat line bundles to the moduli space of integrable Higgs operators in line bundles with vanishing first real Chern class, which (e.g. by the work of Simpson) already is known to exist for a Kähler metric g.

So let us assume that $b_1(X)$ is odd. Then $\deg_g|_{\mathrm{Pic}^0(X)} : \mathrm{Pic}^0(X) \longrightarrow \mathbb{R}$ is surjective, and it holds

$$\mathrm{Pic}(X)^T \big/ \mathrm{Pic}(X)^f \cong \mathrm{Pic}^0(X) \big/ \mathrm{Pic}^0(X)^f \cong \mathbb{R}$$

([LT] Corollary 1.3.12 and Proposition 1.3.13). We will show that I_1 extends to a (non-natural) bijection from $H^1(X, \mathbb{C}^*)$ to $\mathrm{Pic}(X)^T \times H^{1,0}(X)$ in this case, too.

Lemma 6.3 *There is a bijection* $i : \mathrm{Pic}(X)^T \longrightarrow \mathrm{Pic}(X)^f \times \mathbb{R}$ *such that the diagram*

$$\begin{array}{ccc} \mathrm{Pic}(X)^T & \xrightarrow{\deg_g} & \mathbb{R} \\ i\downarrow & & \| \\ \mathrm{Pic}(X)^f \times \mathbb{R} & \xrightarrow{proj.} & \mathbb{R} \end{array}$$

commutes.

Proof: $\deg_g|_{\mathrm{Pic}^0(X)}$ is surjective, so we can choose $\mathcal{L}_1 := [(L_1, \bar\partial_1)] \in \mathrm{Pic}^0(X)$ with $\deg_g(\mathcal{L}_1) = \deg_g(\bar\partial_1) = 1$, and a class $\alpha \in H^1(X, \mathcal{O})$ such that $\mathcal{L}_1 = \pi(\alpha)$ where $\pi : H^1(X, \mathcal{O}) \longrightarrow \mathrm{Pic}^0(X)$ is the natural surjection. For $\lambda \in \mathbb{R}$ define

$$\mathcal{L}_\lambda := \pi(\lambda \cdot \alpha);$$

then $\deg_g(\mathcal{L}_\lambda) = \lambda$ since $\deg_g \circ \pi : H^1(X, \mathcal{O}) \longrightarrow \mathbb{R}$ is linear. Now define i by

$$i(\mathcal{L}) := (\mathcal{L} \otimes \mathcal{L}_{-\deg_g(\mathcal{L})}, \deg_g(\mathcal{L}));$$

then it is obvious that the inverse of i is given by $(\mathcal{L}, \lambda) \mapsto \mathcal{L} \otimes \mathcal{L}_\lambda$, and that the diagram above commutes. □

In the proof of a similar statement for $H^1(X, \mathbb{C}^*)$ we will use

Lemma 6.4 *The natural map*

$$l^1 : H^1(X, \mathbb{C}^*) \longrightarrow \mathrm{Pic}(X)^T, l^1([(L, D)]) := [(L, D'')].$$

is surjective, i.e. a holomorphic structure $\bar\partial$ *in a differentiable line bundle* L *on* X *is the (0,1)-part of a flat connection if and only if the real first Chern class* $c_1(L)_\mathbb{R}$ *vanishes.*

Proof: $\mathrm{Pic}(X)^f$ is naturally identified with $H^1(X, U(1))$, such that the inclusion $\mathrm{Pic}(X)^f \hookrightarrow \mathrm{Pic}(X)$ becomes the injection $k^1 : H^1(X, U(1)) \hookrightarrow H^1(X, \mathcal{O}^*)$ ([LT] p. 38). Observe that k^1 is the composition of the natural map $H^1(X, U(1)) \longrightarrow H^1(X, \mathbb{C}^*)$ and l^1, so it holds

$$\mathrm{Pic}(X)^f = \mathrm{im}(k^1) \subset \mathrm{im}(l^1).$$

Each component of $\mathrm{Pic}(X)^T$ contains a component of $\mathrm{Pic}(X)^f$ ([LT] Remark 1.3.10), hence for each component

$$\mathrm{Pic}^c(X) := \{\ [(L,\bar\partial)] \in \mathrm{Pic}(X) \mid c_1(L)_{\mathbb{Z}} = c\ \} \subset \mathrm{Pic}(X)^T$$

there exists a class $[(L_c, D_c)] \in H^1(X,\mathbb{C}^*)$ such that $l^1([(L_c, D_c)]) \in \mathrm{Pic}^c(X)$. Define $H^1(X,\mathbb{C}^*)^0 := \{\ [(L,D)] \in H^1(X,\mathbb{C}^*) \mid c_1(L)_{\mathbb{Z}} = 0\ \}$. The commutative diagram with exact rows

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathbb{Z} & \longrightarrow & \mathbb{C} & \xrightarrow{exp} & \mathbb{C}^* & \longrightarrow & 0 \\
 & & \| & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathbb{Z} & \longrightarrow & \mathcal{O} & \xrightarrow{exp} & \mathcal{O}^* & \longrightarrow & 0
\end{array}$$

induces the commutative diagram

$$\begin{array}{ccc}
H^1(X,\mathbb{C}) & \longrightarrow & H^1(X,\mathbb{C}^*)^0 \\
h^1 \downarrow & & \downarrow l^1 \\
H^1(X,\mathcal{O}) & \longrightarrow & \mathrm{Pic}^0(X)
\end{array}$$

with surjective horizontal arrows. Since X is a surface, the left vertical arrow h^1 is also surjective ([BPV] p. 117), hence l^1 maps $H^1(X,\mathbb{C}^*)^0$ surjectively onto $\mathrm{Pic}^0(X)$. Now it is easy to see that every element of $\mathrm{Pic}^c(X) \subset \mathrm{Pic}(X)^T$ is of the form $l^1([(L_c \otimes L, D_c \otimes D)])$ for some $[(L,D)] \in H^1(X,\mathbb{C}^*)^0$. □

Lemma 6.5 *There is a bijection $j : H^1(X,\mathbb{C}^*) \longrightarrow H^1(X,\mathbb{C}^*)^f \times \mathbb{R}$ such that the diagram*

$$\begin{array}{ccc}
H^1(X,\mathbb{C}^*) & \xrightarrow{\deg'_g} & \mathbb{R} \\
j \downarrow & & \| \\
H^1(X,\mathbb{C}^*)^f \times \mathbb{R} & \xrightarrow{proj.} & \mathbb{R}
\end{array}$$

commutes, where $\deg'_g := \deg_g \circ l^1$ *is the map associated to the g-degree of flat connections.*

Proof: Choose $\mathcal{L}_1 \in \mathrm{Pic}^0(X)$, $\alpha \in H^1(X,\mathcal{O})$ as in the proof of Lemma 6.3, and a class $\beta \in H^1(X,\mathbb{C})$ with $h^1(\beta) = \alpha$. Let $\pi' : H^1(X,\mathbb{C}) \longrightarrow H^1(X,\mathbb{C}^*)$ be the map induced by $exp : \mathbb{C} \longrightarrow \mathbb{C}^*$, and define $\mathcal{L}'_1 := \pi'(\beta) \in H^1(X,\mathbb{C}^*)$. Since the diagram

$$\begin{array}{ccc}
H^1(X,\mathbb{C}) & \xrightarrow{\pi'} & H^1(X,\mathbb{C}^*) \\
h^1 \downarrow & & \downarrow l^1 \\
H^1(X,\mathcal{O}) & \longrightarrow & \mathrm{Pic}(X)^T
\end{array}$$

commutes, it holds $\deg'_g(\mathcal{L}'_1) = 1$. The rest of the proof is as for Lemma 6.3. □
We conclude

Theorem 6.6 *The composition*

$$\bar{I} : H^1(X,\mathbb{C}^*) \xrightarrow{j} H^1(X,\mathbb{C}^*)^f \times \mathbb{R} \xrightarrow{I_1\times \mathrm{id}_{\mathbb{R}}} H^{1,0}(X) \times \mathrm{Pic}(X)^f \times \mathbb{R}$$
$$\xrightarrow{\mathrm{id}_{H^{1,0}(X)}\times i^{-1}} H^{1,0}(X) \times \mathrm{Pic}(X)^T$$

is a bijective extension of the map I_1, and preserves the g-degree.

We finish with the obvious remark that the map $l^1 : H^1(X,\mathbb{C}^*) \longrightarrow \mathrm{Pic}(X)^T$ in general does not coincide with the composition of $\bar{I}$ and projection onto $\mathrm{Pic}(X)^T$.

References

[B] M. Brunella: Feuilletages holomorphes sur les surfaces complexes compactes. *Ann. scient. Éc. Norm. Sup.*, 4^e série, t. 30 (1997), 569-594.

[BPV] W. Barth, C. Peters, A. Van de Ven: *Compact Complex Surfaces.* Springer Verlag 1984.

[C] K. Corlette: Flat G-structures with canonical metrics. *J. Diff. Geom.* 28 (1988), 361-382.

[G] P. Gauduchon: Sur la 1-forme de torsion d'une variété hermitienne compact. *Math. Ann.* 267 (1984), 495-518.

[K] S. Kobayashi: *Differential geometry of complex vector bundles.* Iwanami Shoten and Princeton Univ. Press 1987.

[LT] M. Lübke, A. Teleman: *The Kobayashi-Hitchin Correspondence.* World Scientific 1995.

[P] R. Plantiko: A rigidity property of class VII_0 surface fundamental groups. *J. reine angew. Math.* 465 (1995), 145-163.

[S1] C.T. Simpson: Constructing variations of Hodge structures using Yang-Mills theory and applications to uniformization. *J. Am. Math. Soc.* 1 (1988), 867-918.

[S2] C.T. Simpson: Higgs bundles and local systems. *Publ. Math. IHES* 75 (1992), 5-95.

[S3] C.T. Simpson: Moduli of representations of the fundamental group of a smooth projective variety I&II. Preprint, Toulouse 1992.

[T] C.H. Taubes: The existence of self-dual conformal structures. *J. Diff. Geom* 36 (1992), 163-253.

[UY] K.K. Uhlenbeck, S.-T. Yau: On the existence of Hermitian-Yang-Mills connections in stable vector bundles. *Comm. Pure Appl. Math.* 39 (1986), S257-S293.

M. Lübke
Mathematical Institute
Leiden University
PO Box 9512
NL 2300 RA Leiden
lubke@math.leidenuniv.nl

Permanence Properties of C*-exact Groups

Eberhard Kirchberg and Simon Wassermann

Received: March 31, 1999

Communicated byJoachim Cuntz

Abstract. It is shown that the class of exact groups, as defined in a previous paper, is closed under various operations, such as passing to a closed subgroup and taking extensions. Taken together, these results imply, in particular, that all almost-connected locally compact groups are exact. The proofs of the permanence properties use a result relating the exactness of sequences of maps in which corresponding algebras are strongly Morita equivalent. The statement of this result is based on a notion of reduced twisted crossed product for covariant systems which are twisted in the sense of Green. The theory of these reduced twisted crossed products and the proof of the exactness result are given in the first part of the paper.

1991 Mathematics Subject Classification: 46L05, 46L55, 22D25

1. Introduction.

Given a locally compact group G, let $\mathcal{C}^*_G$ be the category whose objects are the pairs (A,α) consisting of a C*-algebra A and a continuous action α of G on A, and whose maps are the G-equivariant *-homomorphisms between C*-algebras with continuous G-actions. Following [KW], the group G is said to be *C*-exact* (or just *exact*) if the reduced crossed product functor $A \to A \rtimes_{\alpha,r} G$, for $(A,\alpha) \in \mathcal{C}^*_G$, is short-exact. To be more precise, G is exact if and only if, whenever $(I,\alpha),(A,\beta)$ and (B,γ) are elements of $\mathcal{C}^*_G$ and there is a G-equivariant short exact sequence

$$0 \longrightarrow I \xrightarrow{\iota} A \xrightarrow{q} B \longrightarrow 0,$$

of maps, the corresponding sequence

$$0 \longrightarrow I \rtimes_{\alpha,r} G \xrightarrow{\iota_r} A \rtimes_{\beta,r} G \xrightarrow{q_r} B \rtimes_{\gamma,r} G \longrightarrow 0$$

of reduced crossed products is exact. This is equivalent to saying that for $(A,\alpha) \in \mathcal{C}^*_G$, if I is an α_G-invariant ideal of A, then the quotient $(A \rtimes_{\alpha,r} G)/(I \rtimes_{\alpha|,r} G)$ is canonically isomorphic to $(A/I) \rtimes_{\dot\alpha,r} G$, where $\alpha|$ and $\dot\alpha$ are the restriction and quotient actions of G on I and A/I, respectively.

We introduced group exactness in [KW], primarily as a criterion for the continuity of crossed products of continuous bundles of C*-algebras. Given a continuous bundle $\mathcal{A} = \{A, X, A_x\}$ over a locally compact Hausdorff space X with a continuous fibre-preserving action α of a group G on the bundle C*-algebra A, it is not in general clear that the reduced crossed product bundle $\mathcal{A} \rtimes_{\alpha,r} G = \{A \rtimes_{\alpha,r} G, X, A_x \rtimes_{\alpha_x,r} G\}$ is continuous, though we know of no instance where continuity fails. One of the main results in [KW] is that, for a given G, $\mathcal{A} \rtimes_{\alpha,r} G$ is continuous for all pairs $(\mathcal{A},\alpha)$ if and only if G is exact. It is thus of some importance to know which groups are exact, and it is this problem which is addressed in this paper.

The most basic question is whether, in fact, all locally compact groups are exact. We have so far not been able to resolve this question even in the discrete case, and to the best of our knowledge the exactness of arbitrary discrete groups remains a significant open problem. What we are able to show is that the class of exact groups is closed under various operations such as passing to closed subgroups and taking extensions. Moreover groups possessing closed exact subgroups of finite covolume or which are cocompact are themselves exact. Using these permanence results we can show that groups from a wide class, including, in particular, all connected groups, are exact.

To prove these results we use adaptations of a number of techniques from the theory of induced representations of C*-algebras. Originally formulated by Rieffel to give an interpretation of Mackey's theory of induced representations of groups in terms of C*-algebras, this theory has been developed by P. Green [Gr] and others to give powerful techniques for handling crossed products of C*-algebras. The main tools that we use to prove the permanence results are imprimitivity theorems asserting strong Morita equivalences between various C*-crossed products by a group G on the one hand and by a closed subgroup H of G on the other. These results all follow either from Green's generalisation to crossed products of Rieffel's imprimitivity theorem [Gr, §2], or from Raeburn's symmetric generalisation of Green's theorem [Rae, Theorem 1.1]. We shall use Green's notion of a twisted action of a group G on a C*-algebra [Gr] to prove that exactness is preserved on taking extensions.

Let N be a closed normal subgroup of G and suppose that $(A,\alpha) \in \mathcal{C}^*_G$. Using α to denote also the restriction $\alpha|N$, G has a canonical continuous action γ on $A \rtimes_\alpha N$, and there is a canonical homomorphic embedding $\tau : N \to \mathcal{U}(A \rtimes_\alpha N)$, where $\mathcal{U}(A \rtimes_\alpha N)$ is the unitary group of the multiplier algebra $M(A \rtimes_\alpha N)$. The map τ, which is an example of a twisting map, satisfies various compatibility conditions relative to γ (see §2). The system $\{A \rtimes_\alpha N, G, \gamma, \tau\}$ is an example of a twisted covariant system in the sense of Green.

In general a twisted covariant system $\{A, G, \alpha, \tau\}$ consists of a continuous action α of G on A and a continuous group monomorphism τ from a closed normal subgroup N of G into the unitary group $\mathcal{U}(A)$ of the multiplier algebra $M(A)$ satisfying the aforementioned compatibility conditions. There is a natural idea of a twist-preserving covariant pair of representations of a twisted covariant system and the full twisted crossed product $A \rtimes_{\alpha,\tau} G$ is defined as the unique quotient of the usual full crossed product $A \rtimes_\alpha G$ which is universal for the representations obtained as the integrated forms of the twist-preserving covariant pairs of representations of $\{A, G\}$. If (γ, τ) is the twisted action of the previous paragraph, the crossed products $(A \rtimes_{\alpha|N} N) \rtimes_{\gamma,\tau} G$ and $A \rtimes_\alpha G$ are canonically *-isomorphic, by [Gr, §1]. By a result of Echterhoff [Ech, Theorem 1], if $\{B, G, N, \alpha, \tau\}$ is a twisted covariant system, there is an associated covariant system $\{C, G/N, \beta\}$ such that the twisted crossed product $B \rtimes_{\alpha,\tau} G$ is strongly Morita equivalent to the ordinary crossed product $C \rtimes_\beta (G/N)$. Combining these results, one finds that, with A, G, α and N as above, $A \rtimes_{\alpha|N} N$ is strongly Morita equivalent to $C \rtimes_\beta (G/N)$, where $C = (C_0(G/N) \otimes A) \rtimes_{\Delta^\alpha} G$, Δ^α being the diagonal action of G on $C_0(G/N) \otimes A \cong C_0(G/N, A)$, and β is a certain action of G/N on C.

An analogous result for reduced crossed products is used in §5 to show that an extension of an exact group by an exact group is exact. This requires the definition, for a given a twisted covariant system $\{A, G, \alpha, \tau\}$, of a reduced twisted crossed product $A \rtimes_{\alpha,\tau,r} G$, which reduces to the ordinary reduced crossed product if the twisting is trivial, that is, if $N = \{1\}$. Although there are definitions of twisted reduced crossed products in the literature for twisted actions coming from cocycles, so far as we are aware none has been given hitherto for twisted actions in the sense of Green. Using the reduced twisted crossed product we show that $A \rtimes_{\alpha,r} G$ is strongly Morita equivalent to $(A \rtimes_{\alpha,r} N) \rtimes_{\gamma,\tau,r} G$ for a suitable twisted action (γ, τ) of G on $A \rtimes_{\alpha,r} N$. In fact our result is sharper, in that the Morita equivalence we establish is functorial in A in a certain sense.

The other permanence results are also proved using analogues for reduced crossed products of known imprimitivity theorems for full crossed products. In order to unify our techniques as much as possible, we give a general imprimitivity theorem in §2 which covers all the cases we need. This section also contains a brief review of twisted covariant systems, full twisted crossed products and other relevant background material. We define the reduced twisted crossed product in §3, and deduce imprimitivity results for reduced crossed products that parallel those for full crossed products in §2. These results are the basis of the proofs of the permanence properties mentioned above, which are established in §§4 and 5. The permanence properties are applied in §6 to show that groups of various types are exact.

During the writing of this paper we have benefited from discussions with a number of people. We should particularly like to thank Etienne Blanchard, Siegfried Echterhoff and George Skandalis for valuable comments and observations. We also have pleasure in thanking the following for support:

The EPSRC for a Visiting Fellowship for the first author to visit Glasgow.

The DFG for supporting a visit by the second author to Berlin.

Gert Pedersen for invitations to visit Copenhagen, and the Danish Research Council for supporting these visits.

The Volkswagen Stiftung for a Research in Pairs Fellowship at the Mathematisches Forschungsinstitut, Oberwolfach.

The second author would like to express his gratitude to Etienne Blanchard and his colleagues at the Université d'Aix-Marseille II at Luminy for inviting him to visit Luminy.

2. Imprimitivity results for full twisted crossed products.

In this section we recall the basic ideas of Green's theory of twisted group actions and state some of the imprimitivity results which will be used in later sections to prove the permanence results. Most of this material is a straightforward generalisation of [Gr,§§1,2], but we have found it necessary to make some aspects of the theory which are not immediately accessible in Green's treatment more explicit. Throughout the paper all groups will be assumed locally compact. Our notation follows that of [KW], for the most part. For each locally compact group G, m_G will denote a particular left Haar measure on G and Δ_G the modular function. For $(A, \alpha) \in \mathcal{C}^*_G$, the full and reduced crossed products of A by G are denoted by $A \rtimes_\alpha G$ and $A \rtimes_{\alpha,r} G$, respectively. If $(B, \beta) \in \mathcal{C}^*_G$, and $\theta : A \to B$ is a completely positive G-equivariant map, then θ_u and θ_r will denote the canonical extensions of the map $f \to \theta(f); C_c(G, A) \to C_c(G, B)$, where $(\theta(f))(s) = \theta(f(s))$ for $s \in G$, to completely positive maps

$$A \rtimes_\alpha G \to B \rtimes_\beta G$$

and

$$A \rtimes_{\alpha,r} G \to B \rtimes_{\beta,r} G,$$

respectively. If $\{\pi, V\}$ is a covariant pair of representations of the covariant system $\{A, G, \alpha\}$, then $\pi \rtimes V$ will denote the corresponding integrated form representation of the full crossed product $A \rtimes_\alpha G$.

Let G be a locally compact group, let H be a closed subgroup of G and let $(A, \alpha) \in \mathcal{C}^*_H$. Recall that the C*-algebra $\mathrm{Ind}(A, \alpha)$ is the *-subalgebra of the C*-algebra $C_b(G, A)$ of bounded continuous A-valued functions on G consisting of those functions f such that

$$\alpha_h(f(xh)) = f(x)$$

for $h, x \in H$, so that the function $x \to ||f(x)||$ is constant on left cosets of H, and such that the associated continuous function on G/H given by

$$xH \to ||f(x)||$$

is in $C_0(G/H)$. As is easily seen, $\mathrm{Ind}(A, \alpha)$ is closed in $C_b(G, A)$, and is, moreover, the bundle algebra of a continuous bundle of C*-algebras on G/H with constant fibre A. In general this bundle is nontrivial, though if the action α

extends to an action of G on A, then the bundle is isomorphic to the trivial bundle on G/H with fibre A. In fact when α is defined on all of G, an automorphism ν of $C_b(G, A)$ is defined by

$$(\nu(f))(s) = \alpha_s(f(s)).$$

For $f \in \mathrm{Ind}(A, \alpha)$, $x \in G$ and $h \in H$, $(\nu(f))(xh) = \alpha_x((f)(x))$, so that $\nu(f)$ is constant on left cosets of H. If we identify $\nu(f)$ with the corresponding function in $C_0(G/H, A)$, the restriction of ν to $\mathrm{Ind}(A, \alpha)$ gives an isomorphism ν_A of $\mathrm{Ind}(A, \alpha)$ onto $C_0(G/H, A)$.

A continuous action $\tilde{\alpha}$ of G on $\mathrm{Ind}(A, \alpha)$ is given by

$$(\tilde{\alpha}_g(\psi))(s) = \psi(g^{-1}s).$$

If α is defined on all of G, for $\psi \in \mathrm{Ind}(A, \alpha)$, $g, s \in G$,

$$(\nu(\tilde{\alpha}_g(\psi)))(s) = \alpha_s((\tilde{\alpha}_g(\psi))(s)) = \alpha_s(\psi(g^{-1}s)) = (\Delta^\alpha_g(\nu(\psi)))(s),$$

where Δ^α is the diagonal action of G on $C_0(G/H, A)$ given by

$$(\Delta^\alpha_g(f))(s) = \alpha_g(f(g^{-1}s))$$

for $f \in C_0(G/H, A)$. Thus ν_A is an equivariant isomorphism between the covariant systems $\{\mathrm{Ind}(A, \alpha), G, \tilde{\alpha}\}$ and $\{C_0(G/H, A), G, \Delta^\alpha\}$.

Let E_0 and B_0 be the *-algebras $C_c(G,\mathrm{Ind}(A, \alpha))$ and $B_0 = C_c(H, A)$, with the convolution products relative to the actions $\tilde{\alpha}$ and α, respectively, and let $X_0 = C_c(G, A)$. The algebras E_0 and B_0 are taken to have the C*-norms and positive cones resulting from their canonical embeddings in $\mathrm{Ind}(A, \alpha) \rtimes_{\tilde{\alpha}} G$ and $A \rtimes_\alpha H$, respectively. The linear space X_0 is given an E_0–B_0 bimodule structure and E_0- and B_0-valued inner products as follows. For $f \in E_0$, $g \in B_0$, and $x, y \in X_0$, fx, xg, $\langle x, y\rangle_{E_0}$ and $\langle x, y\rangle_{B_0}$ are defined by

$$\begin{aligned}
(fx)(r) &= \int_G f(s, r)x(s^{-1}r)dm_G(s) \\
(xg)(r) &= \int_H \delta(t)\alpha_t(x(rt)g(t^{-1}))dm_H(t) \\
\langle x, y\rangle_{E_0}(s, r) &= \int_H \Delta_G(rs^{-1}t)\alpha_t(x(rt)y(s^{-1}rt)^*)dm_H(t) \\
\langle x, y\rangle_{B_0}(t) &= \delta(t)\int_G x(s)^*\alpha_t(y(st))dm_G(s),
\end{aligned}$$

where $\delta(t) = \Delta_G(t)^{1/2}/\Delta_H(t)^{1/2}$. It is easily checked that $fx, xg \in X_0$, $\langle x, y\rangle_{E_0} \in E_0$ and $\langle x, y\rangle_{B_0} \in B_0$. The map $(f, x) \to fx$ is a left action of E_0 on X_0 and is the integrated form of the covariant pair of left actions of $\mathrm{Ind}(A, \alpha)$ and G on X_0 given by

$$\begin{aligned}
(\psi x)(r) &= \psi(r)x(r) \\
(sx)(r) &= x(s^{-1}r)
\end{aligned}$$

($\psi \in \mathrm{Ind}(A,\alpha),\ s \in G,\ x \in X_0$). The map $(x,g) \to xg$ is a right action of B_0 on X_0 and is the integrated form of the covariant pair of right actions of A and H on X_0 given by

$$\begin{array}{rcl}(xa)(r) & = & x(r)a \\ (xt)(r) & = & \Delta_G(t)^{-1/2}\Delta_H(t)^{-1/2}\alpha_{t^{-1}}(x(rt^{-1}))\end{array}$$

($a \in A,\ t \in H,\ x \in X_0$).

The following theorem generalises [Gr, Proposition 3] and [Rie1, §7]. It is straightforward, if rather tedious, to write out a proof along the lines of those of [Gr] and [Rie1], though, as Siegfried Echterhoff has pointed out to us, the result is a corollary of Raeburn's more general symmetric imprimitivity theorem [Rae, Theorem 1.1, special case 1.5]. We are grateful to Echterhoff for drawing our attention to the latter, and also for showing us how the result can, alternatively, be deduced directly from Green's original imprimitivity theorem.

Theorem 2.1 *With the structure defined above, X_0 is an E_0–B_0 equivalence (or imprimitivity) bimodule.*

Remark If the action α is actually defined on the whole of G, so that the covariant systems $\{\mathrm{Ind}(A,\alpha), G, \tilde{\alpha}\}$ and $\{C_0(G/H,A), G, \Delta^\alpha\}$ are equivariantly isomorphic, it is straightforward to verify that the E_0–B_0 equivalence bimodule X_0 is isomorphic to the $C_c(G, C_0(G/H,A))$–$C_c(H,A)$ equivalence bimodule constructed by Green in [Gr], and that Theorem 2.1 reduces to [Gr,Proposition 3].

Let $|| \ ||_u$ be the universal C*-norm on $C_c(H,A)$. If X_A is the completion of X_0 with respect to the norm $x \to ||\langle x,x\rangle_{B_0}||_u^{1/2}$, the action of B_0 extends canonically to a right action of its completion $A \rtimes_\alpha H$ on X_A. Moreover the left action of E_0 on X_0 extends to a left action on X_A by bounded operators, the operator norm on E_0 being a C*-norm, generally incomplete. If E is the completion of E_0 with respect to this norm, there is a canonical left action of E on X_A extending that of E_0, and X_A is an E–$(A \rtimes_\alpha H)$ equivalence bimodule [Rie2]. The C*-algebra E is a quotient of the full crossed product $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G$. In Corollary 2.2 we shall show that the kernel of the quotient map is trivial, so that $E \cong \mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G$ canonically. Since the proofs of the corollary and other later results use induced representations, we review the inducing process briefly.

Let A and B be strongly Morita equivalent pre-C*-algebras (i.e. normed *-algebras whose norms satisfy the C*-condition but are not necessarily complete), and let X be an A–B equivalence bimodule. If π is a contractive *-representation of B on a Hilbert space $\mathcal{H}$, then the corresponding induced representation ${}^X\pi$ acts contractively and non-degenerately on the Hilbert space ${}^X\mathcal{H}$ obtained by completing $X \otimes_B \mathcal{H}$ with respect to the semi-norm $\sum x_i \otimes \xi_i \to ||(\pi(\langle x_i, x_i\rangle_B)\xi_i|\xi_i)||^{1/2}$ and for $a \in A$,

$${}^X\pi(a)(x \otimes \xi) = ax \otimes \xi \qquad (x \in X,\ \xi \in \mathcal{H}).$$

Let X^* be the B–A equivalence bimodule dual to X. Thus X^* is the image of X by an antilinear bijection $x \to x^*$ such that

$$\bar{\lambda}x^* = (\lambda x)^*, \quad bx^* = (xb^*)^*, \quad x^*a = (a^*x)^*$$

for $\lambda \in \mathbb{C}$, $a \in A$, $b \in B$ and $x \in X$, and the B- and A-valued inner products on X^* are given by

$$\langle x^*, y^* \rangle_B = \langle y, x \rangle_B, \quad \langle x^*, y^* \rangle_A = \langle y, x \rangle_A$$

for $x, y \in X$. If σ is a contractive representation of A on a Hilbert space $\mathcal{K}$, then ${}^{X^*}\sigma$ is a contractive representation of B. The A–A equivalence bimodules $X \otimes_B X^*$ and A are isomorphic, and likewise there is an isomorphism between the B–B equivalence bimodules $X^* \otimes_A X$ and B. It follows that there are unitary equivalences ${}^{X^*}({}^X\pi) \cong \pi$ and ${}^X({}^{X^*}\sigma) \cong \sigma$ for any non-degenerate representations π and σ of B and A, respectively, so that there is a bijective correspondence between the equivalence classes of non-degenerate representations of A and B. In the rest of the paper all representations will be assumed non-degenerate.

If A and B are actually C*-algebras, by [Rie1] there are bijective correspondences between (i) ideals of A, (ii) closed A–B-invariant subspaces of X and (iii) ideals of B. If Y is a closed A–B-invariant subspace of X, the corresponding ideals of A and B are

$$\begin{aligned} A_Y &= \overline{\text{span}}\{< y, x >_A : x \in X, y \in Y\} \\ B_Y &= \overline{\text{span}}\{< x, y >_B : x \in X, y \in Y\}, \end{aligned}$$

respectively. In the opposite direction, if I and J are ideals in B and A, respectively, then the corresponding A–B-invariant subspaces of X are

$$Y_I = XI = \overline{\text{span}}\{xz : x \in X, z \in I\}$$

and

$${}_JY = JX = \overline{\text{span}}\{zx : z \in J, x \in X\}.$$

These correspondences clearly respect inclusion. When necessary we shall say that I and J correspond *via* X. It is straightforward to verify that if π is a representation of B, then the ideals $\ker^X \pi$ and $\ker \pi$ of A and B, respectively, correspond via X. In particular, ${}^X\pi$ is faithful if and only if π is faithful.

Corollary 2.2 *The operator norm on E_0 is the the universal C*-norm coming from the canonical embedding of E_0 in* $\mathrm{Ind}(A, \alpha) \rtimes_{\tilde{\alpha}} G$*, and X_A is canonically an* $(\mathrm{Ind}(A, \alpha) \rtimes_{\tilde{\alpha}} G)$–$(A \rtimes_\alpha H)$ *equivalence bimodule.*

Proof: Let $\{\pi, U\}$ be a covariant pair of representations of the system $\{A, H, \alpha\}$ on a Hilbert space $\mathcal{H}$. Writing X for X_A, let ${}^X\pi$ and ${}^XU$ denote the restrictions of the induced representation ${}^X(\pi \rtimes U)$ to $\mathrm{Ind}(A, \alpha)$ and G, respectively. We shall say that the covariant pair $\{^X\pi, ^X U\}$ is induced from the pair $\{\pi, U\}$. Similarly, given a covariant pair $\{\sigma, V\}$ of representations of $\{\mathrm{Ind}(A,\alpha), G, \tilde{\alpha}\}$, we obtain a covariant pair $\{^{X^*}\sigma, ^{X^*}U\}$ of representations of $\{A, H\}$. By the above discussion, the pairs $\{\pi, U\}$ and $\{^{X^*}(^X\pi), ^{X^*}(^X U)\}$ are unitarily equivalent, as are the pairs $\{\sigma, V\}$ and $\{^X(^{X^*}\sigma), ^X(^{X^*}V)\}$.

Now let $\{\pi, U\}$ and $\{\sigma, V\}$ be universal covariant pairs for $\{A, H, \alpha\}$ and $\{\mathrm{Ind}(A,\alpha), G, \tilde{\alpha}\}$, respectively. Replacing $\{\pi, U\}$ and $\{\sigma, V\}$ by the pairs $\{\pi \oplus ^{X^*}\sigma, U \oplus ^{X^*}V\}$ and $\{\sigma \oplus ^X\pi, V \oplus ^X U\}$, respectively, we can assume, since the inducing process respects direct sums, that

$$\sigma = ^X\pi, \quad V = ^X U, \quad \pi = ^{X^*}\sigma, \quad U = ^{X^*}V.$$

Now the representation $\sigma \rtimes V$ of $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G$ is universal, hence faithful. Also, the representation $^X(\pi \rtimes U)$ of E has restrictions σ and V to $\mathrm{Ind}(A,\alpha)$ and G, respectively. This implies that $\sigma \rtimes V$ factorises via the quotient map $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G \to E$, which implies that the quotient map is injective, so that $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G \cong E$ as required. □

REMARK If $\{\pi, U\}$ is any universal covariant pair of representations of $\{A, H, \alpha\}$, then $\pi \rtimes U$ is a faithful representation of $A \rtimes_\alpha H$, and so $^{X_A}(\pi \rtimes U)$ is a faithful representation of $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G$. Thus $\{^{X_A}\pi, ^{X_A}U\}$ is universal for $\{\mathrm{Ind}(A,\alpha), G, \tilde{\alpha}\}$.

Let G be a locally compact group with closed normal subgroup N. For $(A,\alpha) \in \mathcal{C}_G^*$ a *twisting map* for N is a strictly continuous homomorphism τ of N into the unitary group $\mathcal{U}(A)$ of $M(A)$ such that for $n \in N,\ s \in G$,

$$\tau(n)a\tau(n)^{-1} = \alpha_n(a)$$

and

$$\tau(sns^{-1}) = \alpha_s(\tau(n)).$$

The pair (α, τ) is called a *twisted action* of of G on A relative to N, and, provided A is nonzero, $\{A, G, \alpha, \tau\}$ is referred to as a twisted covariant system. A covariant pair of representations $\{\pi, V\}$ of $\{A, G\}$ on a Hilbert space $\mathcal{H}$ is *τ-covariant* or *twist-preserving* if

$$\bar{\pi}(\tau(n)) = V_n$$

for $n \in N$, where $\bar{\pi}$ denotes the canonical extension of π to the multiplier algebra $M(A)$.

Let I_τ be the closed, two-sided ideal $\bigcap_{\{\pi,V\}} \ker(\pi \rtimes V)$ of the full crossed product $A \rtimes_\alpha G$, where the supremum is over all τ-covariant pairs of representations of $\{A, G, \alpha\}$. The *full twisted crossed product* $A \rtimes_{\alpha,\tau} G$ is the C*-algebra $(A \rtimes_\alpha G)/I_\tau$. It has the universal property that if $\{\pi, V\}$ is a τ-covariant pair of representations of $\{A, G, \alpha\}$, then $I_\tau \subseteq \ker(\pi \rtimes V)$, so that $\pi \rtimes V$ is the composition of a representation $\pi \rtimes_\tau V$ of $A \rtimes_{\alpha,\tau} G$ with the quotient map $A \rtimes_\alpha G \to A \rtimes_{\alpha,\tau} G$.

Although it is not made explicit in [Gr], for a given twisted covariant system $\{A, G, \alpha, \tau\}$ it is always possible to find a twist-preserving covariant pair of representations $\{\pi, V\}$ with π faithful. In §3 we shall construct for a given faithful representation π of A on a Hilbert space $\mathcal{H}$ a τ-covariant pair of representations $\{\pi_{\alpha,\tau}, \lambda_\tau\}$ of $\{A, G, \alpha, \tau\}$ on a Hilbert space $L^2_\tau(G, \mathcal{H})$ canonically associated with π with $\pi_{\alpha,\tau}$ faithful. In the case when N is trivial, this pair reduces to the usual regular pair $\{\pi_\alpha, \lambda^G\}$. It follows that we can find a faithful representation π of $A \rtimes_{\alpha,\tau} G$ on a Hilbert space $\mathcal{H}$, with restrictions $\{\pi_A, \pi_G\}$ to $\{A, G\}$ such that π_A is injective and for $a \in A$, $g \in G$, $\pi_A(a)$ and $\pi_G(g)$ are multipliers of $\pi(A \rtimes_{\alpha,\tau} G)$. If we identify $A \rtimes_{\alpha,\tau} G$ with its image under π and the multiplier algebra $M(A \rtimes_{\alpha,\tau} G)$ with a *-subalgebra of the weak closure of this image, π_A and π_G are respectively a *-monomorphism of A and a group homomorphism of G with kernel contained in N into $M(A \rtimes_{\alpha,\tau} G)$. With these identifications, π_A and π_G are independent of π, and will be referred to as the *canonical morphisms*. It then follows that for any faithful representation of $A \rtimes_{\alpha,\tau} G$, π_A is injective. A twisted covariant pair $\{\pi, V\}$ will be called *universal* if the representation $\pi \rtimes_\tau V$ of $A \rtimes_{\alpha,\tau} G$ is faithful.

Now let G be a locally compact group with a closed normal subgroup N and let H be a closed subgroup of G containing N. If $(A, \alpha) \in \mathcal{C}^*_H$, let $\tau : N \to M(A)$ be a twisting map for N. A homomorphism $\tilde{\tau} : N \to \mathcal{U}(\mathrm{Ind}(A, \alpha))$ is defined by

$$(\tilde{\tau}(n)\psi)(s) = \tau(s^{-1}ns)\psi(s) \qquad (\psi \in \ \mathrm{Ind}(A, \alpha)).$$

It is straightforward to verify that $\{\mathrm{Ind}(A, \alpha), G, \tilde{\alpha}, \tilde{\tau}\}$ is a twisted covariant system relative to N.

Proposition 2.3 *Let $\{\pi, V\}$ be a covariant pair of representations of the covariant system $\{A, H, \alpha, \tau\}$ on a Hilbert space $\mathcal{H}$. The pair $\{{}^{X_A}\pi, {}^{X_A}V\}$ is $\tilde{\tau}$-covariant if and only if the pair $\{\pi, V\}$ is τ-covariant.*

Proof: Let $\tilde{\pi} = {}^{X_A}\pi$ and $U = {}^{X_A}V$. For $n \in N$, $f, g \in C_c(G, A) \subseteq X_A$ and $\xi, \eta \in \mathcal{H}$,

$$\begin{aligned}
&\big(U_n(f \otimes \xi)\big|g \otimes \eta\big) \\
&\quad = \big(\pi(\langle g, nf\rangle_B)\xi\big|\eta\big) \\
&\quad = \int_H \big(\pi(\langle g, nf\rangle_B(t))V_t\xi\big|\eta\big)\, dm_H(t) \\
&\quad = \int_H \int_G \delta(t)\big(\pi(g(s)^*\alpha_t(f(n^{-1}st)))V_t\xi\big|\eta\big)\, dm_G(s)\, dm_H(t) \\
&\quad = \int_G \int_H \delta(t)\big(\pi(g(s)^*\alpha_t(f(s(s^{-1}n^{-1}st))))V_t\xi\big|\eta\big)\, dm_H(t)\, dm_G(s)
\end{aligned}$$

$$\overset{t\to s^{-1}nst}{=} \int_G\int_H \delta(t)\big(\pi(g(s)^*\alpha_{s^{-1}nst}(f(st)))V_{s^{-1}ns}V_t\xi\big|\eta\big)\,dm_H(t)\,dm_G(s)$$

(Since $\Delta_G|N = \Delta_H|N = \Delta_N, N$ being a normal subgroup)

$$= \int_G\int_H \delta(t)\big(\pi(g(s)^*)V_{s^{-1}ns}\pi(\alpha_t(f(st)))V_t\xi\big|\eta\big)\,dm_H(t)\,dm_G(s)$$

and

$$\begin{aligned}
&\big(\tilde\tau(n)f\otimes\xi\big|g\otimes\eta\big)\\
&= \int_G\int_H \delta(t)\big(\pi(g(s)^*\alpha_t(\tau(t^{-1}s^{-1}nst)f(st)))V_t\xi\big|\eta\big)\,dm_H(t)\,dm_G(s)\\
&= \int_G\int_H \delta(t)\big(\pi(g(s)^*)\bar\pi(\tau(s^{-1}ns))\pi(\alpha_t(f(st)))V_t\xi\big|\eta\big)\,dm_H(t)\,dm_G(s).
\end{aligned}$$

If the pair $\{\pi, V\}$ is τ-covariant, it follows that

$$\big(U_n(f\otimes\xi)\big|g\otimes\eta\big) = \big(\tilde\tau(n)f\otimes\xi\big|g\otimes\eta\big),$$

so that $U_n = \tilde\pi(\tilde\tau(n))$, which means that the pair $\{\tilde\pi, U\}$ is $\tilde\tau$-covariant. If, conversely, $\{\tilde\pi, U\}$ is $\tilde\tau$-covariant, then

$$\big(U_n(f\otimes\xi)\big|g\otimes\eta\big) = \big(\tilde\tau(n)f\otimes\xi\big|g\otimes\eta\big),$$

and, by the above calculations,

$$\int_G\int_H\delta(t)\big(\pi(g(s)^*)(V_{s^{-1}ns} - \bar\pi(\tau(s^{-1}ns)))\pi(\alpha_t(f(st)))V_t\xi\big|\eta\big)\,dm_H(t)\,dm_G(s)=0 \qquad (*).$$

Let $a, b\in A$, let $\varepsilon > 0$ and let $\mathcal{V}$ be a symmetric compact neighbourhood of the identity in G such that for $s, t\in\mathcal{V}^2$

$$\big|\big(\pi(b^*)(V_{s^{-1}ns}-\bar\pi(\tau(s^{-1}ns)))\pi(\alpha_t(a))V_t\xi\big|\eta\big)-\big(\pi(b^*)(V_n-\bar\pi(\tau(n)))\pi(a)\xi\big|\eta\big)\big|\le\varepsilon.$$

Letting h be a continuous positive function with support in $\mathcal{V}$ such that

$$\int_G\int_H \delta(t)h(s)h(st)\,dm_H(t)\,dm_G(s) = 1$$

and taking f and g to be the functions $s\to h(s)a$ and $s\to h(s)b$, respectively, a simple calculation shows that the difference between $\pi(b^*)(V_n-\bar\pi(\tau(n)))\pi(a)$ and the integral on the left-hand side of $(*)$ has modulus less than or equal to ε. Since ε is arbitrary, this implies that

$$\pi(b^*)(V_n - \bar\pi(\tau(n)))\pi(a) = 0$$

for $a, b\in A$, so that $\bar\pi(\tau(n)) = V_n$, by the nondegeneracy assumption on π, which implies that the pair $\{\pi, V\}$ is τ-preserving. $\square$

Let I_τ be the kernel of the canonical quotient map $A \rtimes_\alpha H \to A \rtimes_{\alpha,\tau} H$. If $\tilde{I}$ is the ideal of $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G$ corresponding to I_τ via X_A, let E_τ be the quotient $(\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G)/\tilde{I}$. Then $X_{A,\tau} = X_A/X_A I_\tau$ is an E_τ–$(A \rtimes_{\alpha,\tau} H)$ equivalence bimodule. The following theorem generalises [Gr,Corollary 5].

Theorem 2.4 *The C*-algebra E_τ is canonically isomorphic to* $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G$ *and $X_{A,\tau}$ is an* $(\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G)$–$(A \rtimes_{\alpha,\tau} H)$ *equivalence bimodule.*

Proof: The proof is very similar to that of Corollary 2.2. Let $\{\pi, V\}$ and $\{\sigma, U\}$ be universal twist-covariant pairs of representations of $\{A, H, \alpha, \tau\}$ and $\{\mathrm{Ind}(A,\alpha), G, \tilde{\alpha}, \tilde{\tau}\}$, respectively. Then $\{\sigma, U\}$ is unitarily equivalent to the pair $\{{}^{X_A}({}^{X_A^*}\sigma), {}^{X_A}({}^{X_A^*}U)\}$ and, by Proposition 2.3, the pairs $\{{}^{X_A}\pi, {}^{X_A}V\}$ and $\{{}^{X_A^*}\sigma, {}^{X_A^*}U\}$ are $\tilde{\tau}$- and τ-covariant, respectively. Replacing $\{\pi, V\}$ and $\{\sigma, U\}$ by $\{\pi \oplus {}^{X_A^*}\sigma, V \oplus {}^{X_A^*}U\}$ and $\{\sigma \oplus {}^{X_A}\pi, U \oplus {}^{X_A}V\}$, respectively, we can assume that the pair of representations of $\{\mathrm{Ind}(A,\alpha), G, \tilde{\alpha}, \tilde{\tau}\}$ induced from $\{\pi, V\}$ is universal. By our earlier discussion the ideal $\tilde{I}$ is the kernel of the representation $({}^{X_A}\pi) \rtimes ({}^{X_A}V)$, which is the kernel of the canonical quotient map $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G \to \mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G$. It follows that $X_{A,\tau}$ is an $(\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G)$–$(A \rtimes_{\alpha,\tau} H)$ equivalence bimodule. □

Remarks 2.5 1. It follows, by reasoning similar to that of the remark following the proof of Corollary 2.2, that if $\{\pi, V\}$ is any universal τ-covariant pair of representations of $\{A, H, \alpha, \tau\}$, then the $\tilde{\tau}$-covariant pair $\{{}^{X_A}\pi, {}^{X_A}V\}$ is universal for $\{\mathrm{Ind}(A,\alpha), G, \tilde{\alpha}, \tilde{\tau}\}$.

2. Suppose that $H = N$, so that α is a continuous action of N on A, and that $\tau : N \to \mathcal{U}(A)$ is a twisting map. The pair of homomorphisms $\{id, \tau\}$ of $\{A, N\}$, where $id : A \to M(A)$ is the canonical embedding, is τ-covariant. If we represent $M(A)$ faithfully on a Hilbert space in such a way that the restriction of the representation to A is non-degenerate, we can regard this pair of maps as a τ-covariant pair of representations, and the integrated form of $\{id, \tau\}$ is a *-homomorphism Φ of $A \rtimes_{\alpha,\tau} N$ into $M(A)$. For $f \in C_c(N, A)$,

$$\Phi(f) = \int_N f(n)\tau(n)\, dm_N(n) \in A.$$

It follows, by taking limits, that $\Phi(A \rtimes_{\alpha,\tau} N) \subseteq A$, and, by considering f with suitably small support containing the identity of N, that the image of Φ is dense in A. Thus $\Phi(A \rtimes_{\alpha,\tau} N) = A$.

Let $\{\pi, V\}$ be a τ-covariant pair of representations of the twisted system $\{A, N, \alpha, \tau\}$ on a Hilbert space $\mathcal{H}$. Then $\bar{\pi}(\tau(n)) = V_n$ for $n \in N$ and for $f \in C_c(N, A)$,

$$\begin{aligned}(\pi \rtimes V)(f) &= \int_N \pi(f(n))\bar{\pi}(\tau(n))\, dm_N(n) \\ &= \pi\left(\int_N f(n)\tau(n)\, dm_N(n)\right) \\ &= \pi(\Phi(f)).\end{aligned}$$

Thus $\pi \rtimes V = \pi \circ \Phi$. If $\{\pi, V\}$ is a universal pair for $\{A, N, \alpha\}$, then $\pi \rtimes_\tau V$ and π are faithful. This implies that Φ is an isomorphism, i.e. $A \rtimes_{\alpha,\tau} N \cong A$. If π is a representation of A, then $\{\pi, \bar{\pi} \circ \tau\}$ is a τ-covariant pair for $\{A, N\}$ and $\pi \rtimes (\bar{\pi} \circ \tau) = \pi \circ \Phi$. If π is faithful, this implies that the pair $\{\pi, \bar{\pi} \circ \tau\}$ is universal for $\{A, N, \alpha\}$.

3. The reduced twisted crossed product.

Let G be a locally compact group and let $(A, \alpha) \in \mathcal{C}^*_G$. Let N be a closed normal subgroup of G and let $\{\pi, V\}$ be a covariant pair of representations of $\{A, N, \alpha|N\}$ on a Hilbert space $\mathcal{H}$ such that $\pi \rtimes V$ is a faithful representation of the full crossed product $A \rtimes_{\alpha|N} N$. If we identify $A \rtimes_{\alpha|N} N$ with its image under $\pi \rtimes V$, a twisted action (γ, τ) of G on $A \times_\alpha N$ relative to N is defined by

$$\begin{aligned}\gamma_s\left(\int_N \pi(f(t))V_t\, dm_N(t)\right) &= V_s\left(\int_N \pi(f(t))V_t\, dm_N(t)\right)V_s^{-1}\\ &= \int_N \frac{\Delta_G(s)}{\Delta_{G/N}(sN)}\pi(\alpha_s(f(s^{-1}ts))V_t\, dm_N(t)\end{aligned}$$

for $f \in C_c(N, A)$ and $s \in G$, and

$$\tau(n) = V_n \qquad (n \in N)$$

(cf. [Ech, proof of Theorem 1, *et seq.*]). This twisted action has the fundamental property that there is a natural isomorphism

$$A \rtimes_\alpha G \cong (A \rtimes_\alpha N) \rtimes_{\gamma,\tau} G$$

[Gr]. In §5 we shall need an analogous isomorphism with $A \rtimes_{\alpha,r} G$ on the left and $A \rtimes_{\alpha,r} N$ on the right. To formulate such a result we give a definition of reduced twisted crossed product appropriate to the present context. Although there are various definitions of reduced twisted crossed product in the literature for cocycle twistings, our definition seems to be new.

For the definition we require a twisted version of the left regular representation of a crossed product, which we construct as follows. Let π be a not necessarily faithful representation of A on a Hilbert space $\mathcal{H}$ and let $C_c(G, \mathcal{H}, \tau)$ be the set of those continuous $\mathcal{H}$-valued functions f on G whose supports have relatively compact image in G/N and which satisfy

$$\bar{\pi}(\tau(n))f(s) = f(sn^{-1}) \qquad (*)$$

(or, equivalently, $\bar{\pi}(\tau(s^{-1}ns))(f(ns)) = f(s)$) for $s \in G, n \in N$. For $f \in C_c(G, \mathcal{H}, \tau)$ the nonnegative-valued real function $s \to ||f(s)||$ is constant on each coset of N, and if we denote the common value on the coset sN by $||f(sN)||$, the function $sN \to ||f(sN)||$ is in $C_c(G/N, \mathbb{R})$. Let

$$||f||_2^\tau = \left(\int_{G/N} ||f(sN)||^2 dm_{G/N}(sN)\right)^{1/2}.$$

Then $|| \ ||_2^\tau$ is a norm on $C_c(G,\mathcal{H},\tau)$ and the completion $L^2_\tau(G,\mathcal{H})$ is a Hilbert space. It is not difficult to see that $L^2_\tau(G,\mathcal{H})$ is precisely the family of equivalence classes modulo null sets of $\mathcal{H}$-valued measurable functions f on G satisfying $(*)$ and such that

$$\left(\int_{G/N} ||f(sN)||^2 dm_{G/N}(sN)\right)^{1/2} < \infty.$$

For $a \in A$, $\xi \in L^2_\tau(G,\mathcal{H})$ and $g, s \in G$ let

$$(\pi_{\alpha,\tau}(a)\xi)(s) = \pi(\alpha_{s^{-1}}(a))\xi(s)$$

and

$$(\lambda_{\tau,g}\xi)(s) = \xi(g^{-1}s).$$

It is readily checked that $\pi_{\alpha,\tau}(a)\xi$ and $\lambda_{\tau,g}\xi$ are in $L^2_\tau(G,\mathcal{H})$, so that $\pi_{\alpha,\tau}$ and λ_τ are representations of A and G, respectively. When necessary we shall write λ_τ^G to make it clear which group is involved.

LEMMA 3.1 *1. The pair $\{\pi_{\alpha,\tau}, \lambda_\tau\}$ is τ-covariant. If π is faithful, so is $\pi_{\alpha,\tau}$.*

2. If $\dot\lambda$ denotes the representation of G on $L^2(G/N)$ obtained by composing the quotient map $G \to G/N$ with the left regular representation of G/N, then $\{\pi_{\alpha,\tau} \otimes 1_{L^2(G/N)}, \lambda_\tau \otimes \dot\lambda\}$ is a τ-covariant pair of representations of $\{A, G, \alpha, \tau\}$ which is unitarily equivalent to the pair $\{\pi_{\alpha,\tau} \otimes 1_{L^2(G/N)}, \lambda_\tau \otimes 1_{L^2(G/N)}\}$.

3. If $\{\pi, V\}$ is a τ-covariant pair of representations of $\{A, G, \alpha, \tau\}$ on a Hilbert space $\mathcal{H}$, then $\{\pi \otimes 1_{L^2(G/N)}, V \otimes \dot\lambda\}$ is a τ-covariant pair which is unitarily equivalent to the pair $\{\pi_{\alpha,\tau}, \lambda_\tau\}$.

Proof: 1. It follows readily from the definitions that, for $a \in A$ and $g \in G$,

$$\lambda_{\tau,g}\pi_{\alpha,\tau}(a)\lambda_{\tau,g}^{-1} = \pi_{\alpha,\tau}(\alpha_g(a)),$$

so that $\{\pi_{\alpha,\tau}, \lambda_\tau\}$ is a covariant pair for (A, α). Also, if $n, s \in N$, $\xi \in C_c(G,\mathcal{H},\tau)$,

$$\begin{aligned}(\lambda_{\tau,n}\xi)(s) &= \xi(n^{-1}s)\\ &= \bar\pi(\tau(s^{-1}ns))\xi(s)\\ &= \bar\pi(\alpha_{s^{-1}}(\tau(n)))\xi(s)\\ &= (\bar\pi_{\alpha,\tau}(\tau(n))\xi)(s),\end{aligned}$$

i.e. $\{\pi_{\alpha,\tau}, \lambda_\tau\}$ is τ-preserving.

For $f \in C_c(G,\mathcal{H})$ let $\bar f$ be given by

$$\bar f(s) = \int_N \bar\pi(\tau(m))f(sm)dm_N(m).$$

Then

$$\begin{aligned}\bar{f}(sn^{-1}) &= \int_N \bar{\pi}(\tau(m))f(sn^{-1}m)dm_N(m) \\ &= \int_N \bar{\pi}(\tau(n))\bar{\pi}(\tau(n^{-1}m))f(sn^{-1}m)dm_N(m) \\ &= \bar{\pi}(\tau(n))\bar{f}(s).\end{aligned}$$

Also, if $\operatorname{supp} f \subseteq C$ for some compact subset C of G, then $\operatorname{supp} \bar{f} \subseteq CN$, so that $\bar{f}$ is in $L^2_\tau(G, \mathcal{H})$.

Suppose that π is faithful. Let a be a nonzero element of A and let $\xi \in \mathcal{H}$ such that $\pi(a)\xi \neq 0$. For $\varepsilon > 0$ let C be a compact neighbourhood of the identity e in G such that $||\tau(m^{-1})\xi - \xi|| \leq \varepsilon$ for $m \in C \cap N$. Taking f a continuous nonnegative-valued real function on G with support in C^{-1} such that $\int_N f(n)dm_N(n) = 1$ and defining $F \in C_c(G, \mathcal{H})$ by $F(s) = f(s)\xi$,

$$\begin{aligned}||(\pi_{\alpha,\tau}(a)\bar{F})(e) - \pi(a)\xi|| &= ||\int_N f(m)\pi(a)(\pi(\tau(m))\xi - \xi)dm_N(m)|| \\ &\leq \sup_{n \in C \cap N} ||\pi(a)(\pi(\tau(n))\xi - \xi)|| \int_N f(m^{-1})dm_N(m) \\ &\leq \varepsilon||a||\end{aligned}$$

which implies, since ε is arbitrary and π is faithful, that $\pi_{\alpha,\tau}(a) \neq 0$, i.e. $\pi_{\alpha,\tau}$ is faithful.

2. Regarding elements of $L^2_\tau(G, \mathcal{H}) \otimes L^2(G/N)$ as equivalence classes of $\mathcal{H}$-valued functions on $G \times (G/N)$, it is straightforward to show that a unitary operator U on $L^2_\tau(G, \mathcal{H}) \otimes L^2(G/N)$ is defined by

$$(U\xi)(r, sN) = \xi(r, r^{-1}sN).$$

Then

$$U(\pi_{\alpha,\tau}(a) \otimes 1)U^* = \pi_{\alpha,\tau}(a) \otimes 1$$

and

$$U(\lambda_{\tau,g} \otimes 1)U^* = \lambda_{\tau,g} \otimes \dot{\lambda}_{gN},$$

for $a \in A$ and $g \in G$, i.e. U implements the stated equivalence.

3. For $\xi \in C_c(G/N, \mathcal{H})$ let $W\xi$ be the $\mathcal{H}$-valued function on G given by

$$(W\xi)(s) = V_{s^{-1}}\xi(sN).$$

Then for $n \in N$

$$\begin{aligned}(W\xi)(sn^{-1}) &= V_{ns^{-1}}\xi(sN) \\ &= (\bar{\pi}(\tau(n))W\xi)(s),\end{aligned}$$

so that $W\xi \in L^2_\tau(G, \mathcal{H})$. It is also immediate that $||W\xi||^\tau_2 = ||\xi||_2$, the latter norm being that on $L^2(G/N, \mathcal{H})$. Thus W extends to an isometry of $L^2(G/N,\mathcal{H})$ into $L^2_\tau(G,\mathcal{H})$. For $\xi \in L^2_\tau(G,\mathcal{H})$ the $\mathcal{H}$-valued function $s \to V_s\xi(s)$ on G is constant on each coset of N. Letting $W_1\xi$ be the $\mathcal{H}$-valued function on G/N given by

$$(W_1\xi)(sN) = V_s\xi(s),$$

W_1 is an isometry from $L^2_\tau(G,\mathcal{H})$ to $L^2(G/N,\mathcal{H})$, and $W_1 = W^{-1}$. Hence W is bijective. Moreover

$$\begin{aligned}
(\lambda_{\tau,g}W\xi)(s) &= (W\xi)(g^{-1}s)\\
&= V_{s^{-1}}V_g\xi(g^{-1}sN)\\
&= V_{s^{-1}}((V_g\otimes\dot{\lambda}_g)\xi)(sN)\\
&= (W(V_g\otimes\dot{\lambda}_g)\xi)(s)
\end{aligned}$$

and

$$\begin{aligned}
(\pi_{\alpha,\tau}(a)W\xi)(s) &= \pi(\alpha_{s^{-1}}(a))V_{s^{-1}}\xi(sN)\\
&= V_{s^{-1}}\pi(a)\xi(sN)\\
&= (W(\pi(a)\otimes 1)\xi)(s).
\end{aligned}$$

This shows that the pairs $\{\pi\otimes 1_{L^2(G/N)}, V\otimes\dot{\lambda}\}$ and $\{\pi_{\alpha,\tau},\lambda_\tau\}$ are unitarily equivalent.

□

By analogy with the untwisted case, we would like to define the reduced twisted crossed product $A\rtimes_{\alpha,\tau,r}G$ to be the image of $A\rtimes_\alpha G$ under the representation $\pi_{\alpha,\tau}\rtimes\lambda_\tau$, where π is some faithful representation of A. First, however, it is necessary to show that the resulting quotient of $A\rtimes_\alpha G$ does not depend on the choice of π. This is achieved in what follows by showing that, for a given π, the τ-covariant pair $\{\pi_{\alpha,\tau},\lambda^G_\tau\}$ is obtained by inducing the representation π of $A\cong A\rtimes_{\alpha,\tau}N$ to $M(\mathrm{Ind}(A,\alpha|N)\rtimes_{\tilde{\alpha},\tilde{\tau}}G)$ via the equivalence bimodule X_A of §2 and composing the induced representation with canonical morphisms from A and G into this multiplier algebra.

Let H be a closed subgroup of G containing N, let $\{A,\alpha\}\in\mathcal{C}^*_H$ and let $\mathrm{Ind}(A,\alpha)$ be the associated C*-algebra defined in §2. Let π be a faithful representation of A on $\mathcal{H}$ and for $\psi\in\mathrm{Ind}(A,\alpha)$ and $\xi\in L^2_\tau(G,\mathcal{H})$ let

$$(\tilde{\pi}_{\alpha,\tau}(\psi)\xi)(s) = \pi(\psi(s))\xi(s).$$

Then

$$\begin{aligned}
((\tilde{\pi}_{\alpha,\tau}(\psi)\xi)(sn^{-1}) &= \pi(\psi(sn^{-1}))\bar{\pi}(\tau(n))\xi(s)\\
&= \bar{\pi}(\tau(n))\pi(\alpha_{n^{-1}}(\psi(sn^{-1})))\xi(s)\\
&= \bar{\pi}(\tau(n))\pi(\psi(s))\xi(s)\\
&= \bar{\pi}(\tau(n))((\tilde{\pi}_{\alpha,\tau}(\psi)\xi)(s),
\end{aligned}$$

so that $\tilde{\pi}_{\alpha,\tau}(\psi)\xi \in L^2_\tau(G,\mathcal{H})$, and $\tilde{\pi}_{\alpha,\tau}$ is a representation of $\mathrm{Ind}(A,\alpha)$ on $L^2_\tau(G,\mathcal{H})$. For $g \in G$,

$$\begin{aligned}(\lambda^G_{\tau,g}\tilde{\pi}_{\alpha,\tau}(\psi)\lambda^G_{\tau,g^{-1}}\xi)(s) &= \pi(\psi(g^{-1}s))\xi(s)\\ &= (\tilde{\pi}_{\alpha,\tau}(\tilde{\alpha}_g(\psi))\xi)(s),\end{aligned}$$

so that $\{\tilde{\pi}_{\alpha,\tau},\lambda^G_\tau\}$ is a covariant pair of representations of the covariant system $\{\mathrm{Ind}(A,\alpha),G,\tilde{\alpha}\}$. The proof of Lemma 3.1 (1) shows that

$$\lambda^G_{\tau,n} = \bar{\tilde{\pi}}_{\alpha,\tau}(\tilde{\tau}(n))$$

for $n \in N$, which implies that the pair $\{\tilde{\pi}_{\alpha,\tau},\lambda^G_\tau\}$ is $\tilde{\tau}$-covariant.

Proposition 3.2 *The $\tilde{\tau}$-covariant pair of representations of $\{\mathrm{Ind}(A,\alpha),G,\tilde{\alpha},\tilde{\tau}\}$ induced from the τ-covariant pair of representations $\{\pi_{\alpha,\tau},\lambda^H_\tau\}$ of $\{A,H,\alpha,\tau\}$ via the equivalence bimodule X_A of §2 is unitarily equivalent to the pair $\{\tilde{\pi}_{\alpha,\tau},\lambda^G_\tau\}$.*

Proof: We shall also assume, as we may, that the left Haar measures m_G, m_H, m_N and $m_{G/N}$ have been chosen so that

$$\int_G f(s)\,dm_G(s) = \int_{G/N}\int_N f(sn)\,dm_N(n)dm_{G/N}(sN) \tag{1}$$

and

$$\int_H g(t)\,dm_H(t) = \int_{H/N}\int_N g(tn)\,dm_N(n)dm_{H/N}(sN) \tag{2}$$

for $f \in C_c(G)$, $g \in C_c(H)$.

For $f,g \in X_0 = C_c(G,A)$ and $\xi,\eta \in L^2_\tau(H,\mathcal{H})$, we calculate the inner product $(f\otimes\xi|g\otimes\eta)$ in ${}^{X_A}L^2_\tau(H,\mathcal{H})$. To prevent the notation becoming too cumbersome, we regard $\mathcal{H}$ as an $M(A)$-module via $\bar{\pi}$, so that, for $a \in M(A)$ and $\zeta \in \mathcal{H}$, $a\zeta$ will denote $(\bar{\pi}(a))\zeta$, and similarly regard $L^2_\tau(H,\mathcal{H})$ as an $M(A\rtimes_\alpha H)$-module via $\overline{\pi_{\alpha,\tau}\rtimes\lambda^H_\tau}$. Then

$$\begin{aligned}
&(f\otimes\xi|g\otimes\eta)\\
&= \quad (\langle g,f\rangle_{B_0}\xi|\eta)\\
&= \quad \int_{H/N}\int_H\int_G \delta(t)\Delta_G(s)^{-1}\big(\alpha_{r^{-1}}(g(s^{-1})^*)\alpha_{r^{-1}t}(f(s^{-1}t))\xi(t^{-1}r)\big|\eta(r)\big)\\
&\qquad\qquad \times dm_G(s)dm_H(t)dm_{H/N}(rN)\\
&= \quad \int_{H/N}\int_H\int_G \delta(t)\Delta_G(s)^{-1}\big(\alpha_{r^{-1}t}(f(s^{-1}t))\xi(t^{-1}r)\big|\alpha_{r^{-1}}(g(s^{-1}))\eta(r)\big)
\end{aligned}$$

$$\times dm_G(s)dm_H(t)dm_{H/N}(rN)$$

$$\overset{\substack{t\to rt\\ s\to rs}}{=} \int_{H/N}\int_H\int_G \delta(t)\delta(r)\Delta_G(r)^{-1}\Delta_G(s)^{-1}$$
$$\times\big(\alpha_t(f(s^{-1}t))\xi(t^{-1})\big|\alpha_{r^{-1}}(g(s^{-1}r^{-1}))\eta(r)\big)$$
$$\times dm_G(s)dm_H(t)dm_{H/N}(rN)$$

$$\overset{s\to s^{-1}}{=} \int_{H/N}\int_H\int_G \delta(t)\delta(r)\Delta_G(r)^{-1}\big(\alpha_t(f(st))\xi(t^{-1})\big|\alpha_{r^{-1}}(g(sr^{-1}))\eta(r)\big)$$
$$\times dm_G(s)dm_H(t)dm_{H/N}(rN)$$

$$= \int_{H/N}\int_H\int_{G/N}\int_N \delta(t)\delta(r)\Delta_G(r)^{-1}$$
$$\times\big(\alpha_t(f(smt))\xi(t^{-1})\big|\alpha_{r^{-1}}(g(smr^{-1}))\eta(r)\big)$$
$$\times dm_N(m)dm_{G/N}(sN)dm_H(t)dm_{H/N}(rN)$$

(by (1))

$$\overset{t\to m^{-1}t}{=} \int_{H/N}\int_H\int_{G/N}\int_N \delta(m)^{-1}\delta(t)\delta(r)\Delta_G(r)^{-1}$$
$$\times\big(\alpha_{m^{-1}t}(f(st))\xi(t^{-1}m)\big|\alpha_{r^{-1}}(g(smr^{-1}))\eta(r)\big)$$
$$\times dm_N(m)dm_{G/N}(sN)dm_H(t)dm_{H/N}(rN)$$

$$\overset{m\to m^{-1}}{=} \int_{H/N}\int_H\int_{G/N}\int_N \delta(t)\delta(r)\Delta_G(r)^{-1}\Delta_N(m)^{-1}$$
$$\times\big(\alpha_{mt}(f(st))\xi(t^{-1}m^{-1})\big|\alpha_{r^{-1}}(g(sm^{-1}r^{-1}))\eta(r)\big)$$
$$\times dm_N(m)dm_{G/N}(sN)dm_H(t)dm_{H/N}(rN)$$

(since $\delta(m)=1$ for $m\in N$)

$$= \int_{H/N}\int_H\int_{G/N}\int_N \delta(t)\delta(r)\Delta_G(r)^{-1}\Delta_N(m)^{-1}$$
$$\times\big(\tau(m)\alpha_t(f(st))\tau(m^{-1})\xi(t^{-1}m^{-1})\big|\alpha_{r^{-1}}(g(sm^{-1}r^{-1}))\tau(m)\eta(rm)\big)$$

$$\times dm_N(m)dm_{G/N}(sN)dm_H(t)dm_{H/N}(rN)$$

$$= \int\limits_{H/N}\int\limits_{H}\int\limits_{G/N}\int\limits_{N} \delta(t)\delta(r)\Delta_G(rm)^{-1} \times\big(\alpha_t(f(st))\xi(t^{-1})\big|\alpha_{m^{-1}r^{-1}}(g(sm^{-1}r^{-1}))\eta(rm)\big) \times dm_N(m)dm_{G/N}(sN)dm_H(t)dm_{H/N}(rN)$$

$$= \int\limits_{G/N}\int\limits_{H}\int\limits_{H} \delta(t)\delta(r)\Delta_G(r)^{-1}\big(\alpha_t(f(st))\xi(t^{-1})\big|\alpha_{r^{-1}}(g(sr^{-1}))\eta(r)\big) \times dm_H(t)dm_H(r)dm_{G/N}(sN)$$

(by (2))

$$\overset{r\to r^{-1}}{=} \int\limits_{G/N}\int\limits_{H}\int\limits_{H} \delta(t)\delta(r)\big(\alpha_t(f(st))\xi(t^{-1})\big|\alpha_r(g(sr))\eta(r^{-1})\big) \times dm_H(t)dm_H(r)dm_{G/N}(sN).$$

Let $T(f\otimes\xi)$ be the A-valued function on G given by

$$(T(f\otimes\xi))(s) = \int_H \delta(t)\alpha_t(f(st))\xi(t^{-1})\,dm_H(t).$$

Then

$$\begin{aligned} T(f\otimes\xi)(sn^{-1}) &= \int_H \delta(t)\alpha_t(f(sn^{-1}t))\xi(t^{-1})\,dm_H(t) \\ &= \int_H \delta(t)\alpha_{nt}(f(st))\xi(t^{-1}n^{-1})\,dm_H(t) \\ &= \tau(n)((T(f\otimes\xi))(s)), \end{aligned}$$

since $\delta(n)=1$ for $n\in N$, and if K_1 is the support of f in G and K_2 is the support of ξ in H, then the support of $T(f\otimes\xi)$ is contained in the set K_1K_2. The latter set has relatively compact image in G/N since the same is true of K_2 and K_1 is compact. It follows that $T(f\otimes\xi)\in C_c(G,\mathcal{H},\tau)$. By the above calculation, T is an isometric linear map from a dense subspace of ${}^{X_A}L^2_\tau(H,\mathcal{H})$ into $L^2_\tau(G,\mathcal{H})$. Standard arguments involving partitions of unity show that the image of T is dense in $L^2_\tau(G,\mathcal{H})$. Thus T has an extension to an isometry U from ${}^{X_A}L^2(H,\mathcal{H})$ onto $L^2_\tau(G,\mathcal{H})$.

For $\psi \in \mathrm{Ind}(A,\alpha)$, $f \in X_0$, $\xi \in \mathcal{H}$ and $g, s \in G$,

$$\begin{aligned}
[T(\psi f \otimes \xi)](s) &= \int_H \gamma(t)\alpha_t(\psi(st)f(st))\xi(t^{-1})\, dm_H(t) \\
&= \psi(s)(T(f \otimes \xi))(s) \\
&= [\tilde{\pi}_{\alpha,\tau}(\psi)(T(f \otimes \xi))](s)
\end{aligned}$$

and

$$\begin{aligned}
[T(gf \otimes \xi)](s) &= \int_H \gamma(t\alpha_t(f(g^{-1}st))\xi(t^{-1})\, dm_H(t) \\
&= [\lambda_{\tau,g}(T(f \otimes \xi))](s),
\end{aligned}$$

from which it follows that U implements the desired unitary equivalence. □

In the following corollary we assume that $H = N$, so that α is an action of N on A by inner automorphisms.

Corollary 3.3 *If the representation π is faithful, then the integrated form representation $\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau$ of $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G$ is faithful.*

Proof: By Proposition 3.2, $\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau$ is the representation of $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G$ induced from the representation $\pi \rtimes_\tau \lambda_\tau^N$ of $A \rtimes_{\alpha,\tau} N$ via X_A. The Hilbert space $L^2_\tau(N,\mathcal{H})$ is just the space of continuous $\mathcal{H}$-valued functions f such that

$$f(n) = \bar{\pi}(\tau(n^{-1}))f(e)$$

for $n \in N$, with norm $||f(e)||$, and the map $f \to f(e)$ is an isometry of $L^2_\tau(N,\mathcal{H})$ onto $\mathcal{H}$. This map implements a unitary equivalence between the τ-covariant pairs $\{\pi_{\alpha,\tau}, \lambda_\tau^N\}$ and $\{\pi, \bar{\pi} \circ \tau\}$. By Remark 2.5 (2), the latter pair is universal for $\{A, N, \alpha, \tau\}$, since π is faithful. Hence $\{\pi_{\alpha,\tau}, \lambda_\tau^N\}$ is universal for $\{A, N, \alpha, \tau\}$, so that $\pi_{\alpha,\tau} \rtimes_\tau \lambda_\tau^N$ is a faithful representation of $A \rtimes_\tau N$ ($\cong A$, by Remark 2.5 (2)). Since faithful representations induce faithful representations, the result follows. □

We are now ready to define the reduced twisted crossed product. Let G be a locally compact group with a closed normal subgroup N. Let $(A,\alpha) \in \mathcal{C}^*_G$ and let $\tau : N \to \mathcal{U}(A)$ be a twisting map relative to α. Let π be a faithful representation of A on a Hilbert space $\mathcal{H}$. Letting $E = \mathrm{Ind}(A, \alpha|N)$, we note in passing that by the discussion of §2 the pair $(E, \tilde{\alpha})$ is G-equivariantly isomorphic to $C_0(G/N, A)$ with G acting by left translation. By Corollary 3.3, the representation $\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau$ of $E \rtimes_{\tilde{\alpha},\tilde{\tau}} G$ on $L^2_\tau(G,\mathcal{H})$ is faithful, and hence so is its canonical extension $\overline{\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau}$ to $M(E \rtimes_{\tilde{\alpha},\tilde{\tau}} G)$. Identifying $M(E \rtimes_{\tilde{\alpha},\tilde{\tau}} G)$ with its image under $\overline{\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau}$, for $a \in A$, $\psi \in E$ and $\xi \in L^2_\tau(G,\mathcal{H})$,

$$\begin{aligned}
(\pi_{\alpha,\tau}(a)\tilde{\pi}_{\alpha,\tau}(\psi)\xi)(s) &= \pi(\alpha_{s^{-1}}(a))\pi(\psi(s))\xi(s) \\
&= (\tilde{\pi}_{\alpha,\tau}(a\psi)\xi)(s),
\end{aligned}$$

where $a\psi$ is the element of E given by

$$(a\psi)(s) = \alpha_{s^{-1}}(a)\psi(s).$$

For $f \in C_c(G,E)$ and $a \in A$, let af be the element of $C_c(G,E)$ given by $(af)(s) = af(s)$. Then

$$\begin{aligned}
\pi_{\alpha,\tau}(a)(\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau)(f) &= \pi_{\alpha,\tau}(a)\int_G \tilde{\pi}_{\alpha,\tau}(f(s))\lambda_{\tau,s}\, dm_G(s) \\
&= \int_G \tilde{\pi}_{\alpha,\tau}(af(s))\lambda_{\tau,s}\, dm_G(s) \\
&= (\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau)(af) \\
&\in \mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G,
\end{aligned}$$

and, by taking limits of sequences of such f, it follows that $\pi_{\alpha,\tau}(a)(\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau)(x) \in E \rtimes_{\tilde{\alpha},\tilde{\tau}} G$ for all $x \in E \rtimes_{\tilde{\alpha},\tilde{\tau}} G$. Similarly, if $g \in G$ and $\psi \in E$, let $g\psi$ be the element of E given by

$$(g\psi)(s) = \alpha_g(\psi(g^{-1}s)),$$

and for $f \in C_c(G,E)$ let gf be the element of $C_c(G,E)$ such that $(gf)(s) = gf(s)$. A similar calculation shows that $\lambda_{\tau,g}$ multiplies $E \rtimes_{\tilde{\alpha},\tilde{\tau}} G$. There are thus canonical homomorphisms π_0 and λ_0 from A and G into $M(E \rtimes_{\tilde{\alpha},\tilde{\tau}} G)$ given by

$$\pi_0(a)f = af, \qquad \lambda_0(g)g = gf$$

for $f \in C_c(G,E)$. Moreover $\pi_{\alpha,\tau} = (\overline{\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau}) \circ \pi_0$ and $\lambda_\tau = (\overline{\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau}) \circ \lambda_0$, from which it follows that π_0 is an isomorphism, and $\{\pi_0, \lambda_0\}$ is a τ-covariant pair.

Definition 3.4 The *reduced twisted crossed product* $A \rtimes_{\alpha,\tau,r} G$ is the image of $A \rtimes_\alpha G$ in $M(\mathrm{Ind}(A,\alpha|N) \rtimes_{\tilde{\alpha},\tilde{\tau}} G)$ under the *-homomorphism $\pi_0 \rtimes \lambda_0$.

In the next proposition we consider the natural class of mappings between twisted covariant systems with respect to given G and N. Let $\{A, G, \alpha, \tau\}$ and $\{B, G, \beta, \tau'\}$ be two such systems and let $\theta : A \to B$ be a G-equivariant *-homomorphism. We shall say that θ is *twist-equivariant* (with respect to τ and τ') if $\theta(\tau(n)a) = \tau'(n)\theta(a)$ for $n \in N$ and $a \in A$.

Proposition 3.5 *1. Let π be a representation of A on a Hilbert space $\mathcal{H}$. Then the representation $\pi_{\alpha,\tau} \rtimes_\tau \lambda_\tau$ of $A \rtimes_{\alpha,\tau} G$ is the composition of a representation $\pi_{\alpha,\tau} \rtimes_{\tau,r} \lambda_\tau$ of $A \rtimes_{\alpha,\tau,r} G$ with the canonical quotient map $A \rtimes_{\alpha,\tau} G \to A \rtimes_{\alpha,\tau,r} G$. If π is faithful, then so is $\pi_{\alpha,\tau} \rtimes_{\tau,r} \lambda_\tau$.*

*2. Let $\{A, G, \alpha, \tau\}$ and $\{B, G, \beta, \tau'\}$ be twisted covariant systems with respect to the closed normal subgroup N of G, and let $\theta : A \to B$ be a *-homomorphism which is twist-equivariant with respect to the given actions and* *twisting maps. Then there is a unique *-homomorphism* $\theta_{N,r} : A \rtimes_{\alpha,\tau,r} G \to B \rtimes_{\beta,\tau',r} G$ *such that the diagram*

$$\begin{tikzcd} A \rtimes_\alpha G \arrow[r, "\theta_u"] \arrow[d] & B \rtimes_\beta G \arrow[d] \\ A \rtimes_{\alpha,\tau,r} G \arrow[r, "\theta_{N,r}"] & B \rtimes_{\alpha,\tau',r} G \end{tikzcd}$$

*commutes, the vertical arrows denoting the canonical *-homomorphisms. The morphism* $\theta_{N,r}$ *is injective (resp. surjective) if and only if* θ *is injective (resp. surjective). If* $\operatorname{Im}\theta$ *is an ideal of* B*, then* $\operatorname{Im}\theta_{N,r}$ *is an ideal of* $B \rtimes_{\beta,\tau',r} G$.

Proof: 1. This follows immediately from the factorisations $\pi_{\alpha,\tau} = (\overline{\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau}) \circ \pi_0$ and $\lambda_\tau = (\overline{\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau}) \circ \lambda_0$, and the fact that, if π is faithful, then $\overline{\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau}} \lambda_\tau}$ is a faithful representation of $M(\operatorname{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G)$.

2. If $\{\pi, V\}$ is a τ'-covariant pair of representations of $\{B, G, \beta\}$, then $\{\pi \circ \theta, V\}$ is a covariant pair of representations of $\{A, G, \alpha\}$ on a Hilbert space $\mathcal{H}$, and for $n \in N$ and $a \in A$,

$$\begin{aligned} \overline{(\pi \circ \theta)}(\tau(n))(\pi \circ \theta)(a) &= \pi(\theta(\tau(n)a)) \\ &= \overline{\pi}(\tau'(n))(\pi \circ \theta)(a). \end{aligned}$$

If θ is surjective, this shows that $\overline{(\pi \circ \theta)}(\tau(n)) = V_n$, so that the pair $\{\pi\circ\theta, V\}$ is τ-covariant. By part 1, there is a canonical *-epimorphism $\theta_{N,r} : A \rtimes_{\alpha,\tau,r} G \to B \rtimes_{\beta,\tau',r} G$ such that

$$(\pi \rtimes_{\tau',r} V) \circ \theta_{N,r} = (\pi \circ \theta) \rtimes_{\tau,r} V, \qquad (*)$$

where $(\pi \circ \theta) \rtimes_{\tau,r} V$ and $\pi \rtimes_{\tau',r} V$ are the representations of $A \rtimes_{\alpha,\tau,r} G$ and $B \rtimes_{\beta,\tau',r} G$ associated with $\pi \circ \theta$ and π, respectively.

If, on the other hand, θ is injective, let $\mathcal{H}_1$ be the closure in $\mathcal{H}$ of $\pi(\theta(A))\mathcal{H}$ and let E be the projection onto $\mathcal{H}_1$. Then by the covariance condition, $EV_g = V_g E$ for $g \in G$. Letting $W_g = EV_g|\mathcal{H}_1$, the above identity implies that

$$\overline{(\pi \circ \theta)}(\tau(n)) = W_n E$$

for $n \in N$. Defining σ by $\sigma(a) = \pi(\theta(a))|\mathcal{H}_1$, it follows that $\{\sigma, W\}$ is a τ-covariant pair for $\{A, G, \alpha, \tau\}$. It is easily seen that the images of $A \rtimes_\alpha G$ under the representations $(\pi \circ \theta) \rtimes V$ and $\sigma \rtimes W$ are isomorphic. If π is faithful, so is σ, and the latter image is canonically isomorphic to $A \rtimes_{\alpha,\tau,r} G$ by part 1. This implies that there is a canonical *-monomorphism $\theta_{N,r} : A \rtimes_{\alpha,\tau,r} G \to B \rtimes_{\beta,\tau',r} G$ for which $(*)$ holds.

Combining these two cases, the existence of $\theta_{N,r}$ satisfying $(*)$ for arbitrary θ follows. The commutativity of the given diagram is a simple consequence of $(*)$. $\square$

We return now to the general situation where N is a closed normal subgroup of G, H is a closed subgroup of G containing N, $(A,\alpha) \in \mathcal{C}^*_H$ and $\tau : N \to \mathcal{U}(A)$ is a twisting map relative to α. Let X_A be the $(\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G)$–$(A \rtimes_\alpha H)$ equivalence bimodule constructed in §2, and let π be a faithful representation of A on a Hilbert space $\mathcal{H}$. The kernel $J_{\tau,r}$ of the canonical quotient map $A \rtimes_\alpha H \to A \rtimes_{\alpha,\tau,r} H$ is the kernel of the representation $\pi_{\alpha,\tau} \rtimes \lambda^H_\tau$. By Proposition 3.2 the representation of $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G$ induced from this representation via X_A is the integrated form of the pair $\{\tilde{\pi}_{\alpha,\tau}, \lambda^G_\tau\}$. Let I be the kernel of this representation. By Lemma 3.1, the $\tilde{\tau}$-covariant pair $\{\tilde{\pi}_{\alpha,\tau} \otimes 1_{L^2(G/N)}, \lambda^G_\tau \otimes 1_{L^2(G/N)}\}$ is unitarily equivalent to the pair $\{\tilde{\pi}_{\alpha,\tau} \otimes 1_{L^2(G/N)}, \lambda^G_\tau \otimes \dot{\lambda}\}$, which is in turn unitarily equivalent to the pair $\{(\tilde{\pi}_{\alpha,\tau})_{\tilde{\alpha},\tilde{\tau}}, \lambda_\tau\}$. Since π, and hence $\tilde{\pi}_{\alpha,\tau}$, are faithful, the kernel of the integrated form of the pair $\{(\tilde{\pi}_{\alpha,\tau})_{\tilde{\alpha},\tilde{\tau}}, \lambda_\tau\}$ is the ideal $I_{\tilde{\tau},r}$, the kernel of the canonical *-homomorphism from $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G$ to $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau},r} G$, by Proposition 3.5 (1). It follows that that I coincides with $I_{\tilde{\tau},r}$ and corresponds to $J_{\tau,r}$ via X_A. By [Rie2] $X_{A,\tau,r} = X_A / X_A J_{\tau,r}$ is an $(\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau},r} G)$–$(A \rtimes_{\alpha,\tau,r} H)$ equivalence bimodule. In fact, $X_{A,\tau,r}$ is obtained from the E_0–B_0 equivalence bimodule $X_0 = C_c(G,A)$ of §2 by completing with respect to the semi-norm $f \to ||(\pi_{\alpha,\tau} \rtimes \lambda^H_\tau)(< f, f >_B)||^{1/2}$. This proves

THEOREM 3.6 *The C*-algebras* $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau},r} G$ *and* $A \rtimes_{\alpha,\tau,r} H$ *are strongly Morita equivalent via the equivalence bimodule* $X_{A,\tau,r}$.

REMARK 3.7 When $H = N$, it follows by Remark 2.5 (2) that if π is a faithful representation of A, the pair $\{\pi_{\alpha,\tau}, \lambda^N_\tau\}$ is universal for $\{A, N\}$, and the kernels of the canonical quotient maps of $A \rtimes_\alpha N$ onto $A \rtimes_{\alpha,\tau} N$ and $A \rtimes_{\alpha,\tau,r} N$ coincide. Since these kernels correspond via X_A to the kernels of the canonical quotient maps of $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha}} G$ onto $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G$ and $\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau},r} G$, respectively, the latter kernels coincide, and there is a canonical isomorphism

$$\mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau}} G \cong \mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau},r} G.$$

Specialising to the case where $A = \mathbb{C}$, $N = \{1\}$, we recover the well known isomorphism

$$C_0(G) \rtimes_\lambda G \cong C_0(G) \rtimes_{\lambda,r} G,$$

where λ is the action of G on $C_0(G)$ by left translation. As is well-known, the reduced crossed product on the right-hand side is isomorphic to the space of compact operators on $L^2(G)$.

If $(A,\alpha), (B,\beta) \in \mathcal{C}^*_H$, and $\theta : A \to B$ is an H-equivariant *-homomorphism, an associated *-homomorphism $\tilde{\theta}$:$\mathrm{Ind}(A,\alpha) \to \mathrm{Ind}(B,\beta)$ is defined by

$$(\tilde{\theta}(f))(h) = \theta(f(h)).$$

Let $(A,\alpha) \in \mathcal{C}^*_H$ and let I be an α_H-invariant ideal of A. If $\iota : I \to A$ and $q : A \to A/I$ denote the inclusion and quotient maps, respectively, then ι and

q are equivariant when I has the restriction action $\alpha|I$ and A/I the quotient action $\dot{\alpha}$.

Lemma 3.8 *The sequence*

$$0 \longrightarrow \mathrm{Ind}\,(I,\alpha|) \xrightarrow{\tilde{\iota}} \mathrm{Ind}\,(A,\alpha) \xrightarrow{\tilde{q}} \mathrm{Ind}\,(A/I,\dot{\alpha}) \longrightarrow 0$$

is exact.

Proof: If $f \in \ker \tilde{q}$, then

$$0 = (\tilde{q}(f))(h) = q(f(h)) \Rightarrow f(h) \in I$$

for $h \in H$, i.e. $f \in \mathrm{Ind}(I,\alpha|)$. □

Let $(A,\alpha) \in \mathcal{C}^*_H$ and let $\tau : N \to \mathcal{U}(A)$ be a twisting map relative to α. If I is an α_H-invariant ideal of A, there are unital *-homomorphisms $\bar{\iota} : M(I) \to M(A)$ and $\bar{q} : M(A) \to M(A/I)$ given by $\bar{\iota}(u)x = ux$ and $\bar{q}(u)q(a) = q(ua)$ which extend ι and q, respecively. Twisting maps τ_I and $\tau_{A/I}$ relative to the restriction action $\alpha|I$ and the quotient action $\dot{\alpha}$ are given by $\tau_I = \bar{\iota} \circ \tau$ and $\tau_{A/I} = \bar{q} \circ \tau$, respectively, relative to which ι and q are twist-equivariant. To simplify the notation, we write τ for both τ_I and $\tau_{A/I}$. By the same token, there are twisting maps relative to N, which will be denoted by $\tilde{\tau}$, on $\mathrm{Ind}(I,\alpha|)$ and $\mathrm{Ind}(A/I,\dot{\alpha})$, relative to which the induced morphisms $\tilde{\iota}$ and $\tilde{q}$ are twist-equivariant.

The following theorem is the main technical result for §§4 and 5.

Theorem 3.9 *The sequence*

$$0 \longrightarrow I \rtimes_{\alpha|,\tau,r} H \xrightarrow{\iota_{N,r}} A \rtimes_{\alpha,\tau,r} H \xrightarrow{q_{N,r}} (A/I) \rtimes_{\dot{\alpha},\tau,r} H \longrightarrow 0 \qquad (*)$$

is exact if and only if the sequence

$$0 \longrightarrow \mathrm{Ind}(I,\alpha|) \rtimes_{\tilde{\alpha}|,\tilde{\tau},r} G \xrightarrow{\tilde{\iota}_{N,r}} \mathrm{Ind}(A,\alpha) \rtimes_{\tilde{\alpha},\tilde{\tau},r} G$$

$$\xrightarrow{\tilde{q}_{N,r}} \mathrm{Ind}(A/I,\dot{\alpha}) \rtimes_{\tilde{\dot{\alpha}},\tilde{\tau},r} G \longrightarrow 0 \qquad (**)$$

is exact.

Proof: Let B_I, B_A and $B_{A/I}$ denote the three C*-algebras in $(*)$ and E_I, E_A and $E_{A/I}$ the three C*-algebras in $(**)$. Let J and $\tilde{J}$ be the kernels of the homomorphisms $q_{N,\tau} : B_A \to B_{A/I}$ and $\tilde{q}_{N,\tilde{\tau}} : E_A \to E_{A/I}$. Let $J = \ker q_{N,\tau}$ and $\tilde{J} = \ker \tilde{q}_{N,\tilde{\tau}}$. Identifying B_I and E_I with their images in B_A and E_A under the embeddings $\iota_{N,\tau}$ and $\tilde{\iota}_{N,\tilde{\tau}}$, respectively, it follows by Proposition 3.5

that $B_I \subseteq J$ and $E_I \subseteq \tilde{J}$. To prove the proposition it suffices to show that the ideals B_I and J of B_A correspond to the ideals E_I and $\tilde{J}$, respectively, of E_A via the equivalence bimodule $X_{A,\tau,r}$.

The natural embedding of $C_c(G, I)$ in $C_c(G, A)$ extends to an embedding of the E_I–B_I equivalence bimodule $X_{I,\tau,r}$ as an E_I–B_I submodule X of $X_{A,\tau,r}$. It follows readily from the definitions that X is the norm closed linear span of both $X_{A,\tau,r}B_I$ and $E_I X_{A,\tau,r}$, from which it follows that E_I and B_I correspond via $X_{A,\tau,r}$.

Let σ be a faithful representation of A/I on a Hilbert space $\mathcal{H}$, and let $\pi = \sigma \circ q$. By Propositions 3.2 and 3.5, the representation of E_A induced from the representation $\pi \rtimes_{\tau,r} \lambda^H_\tau$ of B_A via $X_{A,\tau,r}$ is unitarily equivalent to $\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau},r} \lambda^G_{\tilde{\tau}}$. Since

$$\pi \rtimes_{\tau,r} \lambda^H_\tau = (\sigma \rtimes_{\tau,r} \lambda^H_\tau) \circ q_{N,r}$$

and

$$\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau},r} \lambda^G_{\tilde{\tau}} = (\tilde{\sigma}_{\dot{\alpha},\tau} \rtimes_{\tilde{\tau},r} \lambda^G_{\tilde{\tau}}) \circ \tilde{q}_{N,r}$$

the images of $\pi \rtimes_{\tau,r} \lambda^H_\tau$ and $\tilde{\pi}_{\alpha,\tau} \rtimes_{\tilde{\tau},r} \lambda^G_{\tilde{\tau}}$ are canonically isomorphic to $B_{A/I}$ and $E_{A/I}$ and their kernels are J and $\tilde{J}$, respectively. By the discussion in §2 it follows that J and $\tilde{J}$ correspond via $X_{A\tau,r}$. Since correspondence of ideals respects inclusion, this implies that $J = B_I$ if and only if $\tilde{J} = E_I$, i.e. $(*)$ is exact if and only if $(**)$ is exact. □

In §5 we shall consider a continuous action α of H on A which extends to a continuous action, also denoted by α, of G on A, and such that the ideal I is α_G-invariant, so that $\alpha|$ and $\dot{\alpha}$ also extend to continuous actions $\alpha|$ and $\dot{\alpha}$ of G on I and A/I, respectively. As noted in §2, there are then natural isomorphism

$$\theta_I : \ \mathrm{Ind}(I, \alpha|) \to C_0(G/H, I),$$

$$\theta_A : \ \mathrm{Ind}(A, \alpha) \to C_0(G/H, A),$$

and

$$\theta_{A/I} : \ \mathrm{Ind}(A/I, \dot{\alpha}) \to C_0(G/H, A/I).$$

A twisting map $\tilde{\tau} : N \to M(C_0(G/H, A))$ is defined by

$$(\tilde{\tau}(n)f)(sH) = \tau(n)f(gH) \qquad (f \in C_0(G/H, A)).$$

Relative to this twisting map and that on $\mathrm{Ind}(A, \alpha)$, the map θ_A is twist equivariant. Defining twisting maps $\tilde{\tau} : N \to M(C_0(G/H, I))$ and $\tilde{\tau} : N \to M(C_0(G/H, A/I))$ similarly, the maps θ_I and $\theta_{A/I}$ are twist-equivariant, and it follows straightforwardly that the diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Ind}(I, \alpha|) \arrow[r, "\tilde{\iota}"] \arrow[d, "\theta_I"] & \mathrm{Ind}(A, \alpha) \arrow[r, "\tilde{q}"] \arrow[d, "\theta_A"] & \mathrm{Ind}(A/I, \dot{\alpha}) \arrow[r] \arrow[d, "\theta_{A/I}"] & 0 \\
0 \arrow[r] & C_0(G/H, I) \arrow[r, "id\otimes\iota"] & C_0(G/H, A) \arrow[r, "id\otimes q"] & C_0(G/H, A/I) \arrow[r] & 0
\end{tikzcd}$$

commutes (we have identified $C_0(G/H, A)$ with $C_0(G/H) \otimes A$, etc., in the bottom row to define the horizontal maps). Taking reduced twisted crossed products by G, we obtain a commutative diagram

$$\begin{tikzcd}
0 \arrow[d] & 0 \arrow[d] \\
\mathrm{Ind}(I, \alpha|) \rtimes_{\tilde{\alpha},\tilde{\tau},r} G \arrow[r] \arrow[d] & C_0(G/H, I) \rtimes_{\Delta^\alpha,\tilde{\tau},r} G \arrow[d] \\
\mathrm{Ind}(A, \alpha) \rtimes_{\tilde{\alpha},\tilde{\tau},r} G \arrow[r] \arrow[d] & C_0(G/H, A) \rtimes_{\Delta^\alpha,\tilde{\tau},r} G \arrow[d] \\
\mathrm{Ind}(A/I, \dot{\alpha}) \rtimes_{\tilde{\dot{\alpha}},\tilde{\tau},r} G \arrow[r] \arrow[d] & C_0(G/H, A/I) \rtimes_{\Delta^{\dot{\alpha}},\tilde{\tau},r} G \arrow[d] \\
0 & 0
\end{tikzcd}$$

Since the maps denoted by the horizontal arrows are bijections, by Proposition 3.5 (2), the left-hand column is exact if and only if the same is true of the right-hand column. The following corollary is now an immediate consequence of Theorem 3.9.

Corollary 3.10 *The sequence*

$$0 \longrightarrow I \rtimes_{\alpha|,\tau,r} H \xrightarrow{\iota_{N,r}} A \rtimes_{\alpha,\tau,r} H \xrightarrow{q_{N,r}} (A/I) \rtimes_{\dot{\alpha},\tau,r} H \longrightarrow 0$$

is exact if and only if the sequence

$$0 \longrightarrow (C_0(G/H, I) \rtimes_{\Delta^{\alpha|},\tilde{\tau},r} G \xrightarrow{(id\otimes\iota)_{N,r}} (C_0(G/H, A) \rtimes_{\Delta^\alpha,\tilde{\tau},r} G$$

$$\xrightarrow{(id\otimes q)_{N,r}} (C_0(G/H, A/I)) \rtimes_{\Delta^{\dot{\alpha}},\tilde{\tau},r} G \longrightarrow 0$$

is exact.

4. Closed subgroups.

Throughout this section H will denote a closed subgroup of the locally compact group G. The following theorem is the first of the two main results of this section.

Theorem 4.1 *If G is exact, then so is H.*

Proof: Let $(A,\alpha) \in \mathcal{C}^*_H$ and let I be an α_H-invariant ideal of A. By Lemma 3.8, the associated sequence

$$0 \to \mathrm{Ind}(I,\alpha|) \to \mathrm{Ind}(A,\alpha) \to \mathrm{Ind}(A/I,\dot\alpha) \to 0$$

is exact. Since G is exact, the corresponding sequence

$$0 \to \mathrm{Ind}(I,\alpha|) \rtimes_{\tilde\alpha,r} G \to \mathrm{Ind}(A,\alpha) \rtimes_{\tilde\alpha,r} G \to \mathrm{Ind}(A/I,\dot\alpha) \rtimes_{\tilde{\dot\alpha},r} G \to 0$$

is exact. By Theorem 3.9, the sequence

$$0 \longrightarrow I \rtimes_{\alpha|,r} H \xrightarrow{\iota_r} A \rtimes_{\alpha,r} H \xrightarrow{q_r} (A/I) \rtimes_{\dot\alpha,r} H \longrightarrow 0$$

is exact. Since (A,α) and I are arbitrary, this implies the exactness of H. $\square$

For the rest of the section we assume that G is σ-compact and H is exact. By [Bo, Chap. VII, §2, Th. 2] there is a Borel measure on G/H which is quasi-invariant for the action of G on H given by left translation, though in general it is not possible to find a measure which is actually invariant for this action. If an invariant Borel probability measure on G/H exists, H is said to have *finite covolume* in G. We shall assume that this is the case for the rest of this section.

Let $(A,\alpha) \in \mathcal{G}^*_G$ and let $\bar\alpha$ denote the canonical extension of α to a continuous action of G on the multiplier algebra $M(A)$. The natural embedding of A in $M(A)$ is G-equivariant relative to this action and the corresponding crossed product map is an embedding of $A \rtimes_{\alpha,r} G$ as an ideal of $M(A) \rtimes_{\bar\alpha,r} G$. It follows that there is a canonical *-homomorphism

$$M(A) \rtimes_{\bar\alpha,r} G \to M(A \rtimes_{\alpha,r} G)$$

extending the natural embedding of $A \rtimes_{\alpha,r} G$ in $M(A \rtimes_{\alpha,r} G)$. This *-homomorphism is, in fact, an embedding. To see this, let π be a faithful representation of A on a Hilbert space $\mathcal{H}$. Then $\pi_\alpha \rtimes_r \lambda$ is a faithful representation of $A \rtimes_{\alpha,r} G$ on $\mathcal{H} \otimes L^2(G)$ which extends to a faithful representation of $M(A \rtimes_{\alpha,r} G)$ on $\mathcal{H} \otimes L^2(G)$. Let $\bar\pi$ be the canonical extension of π to $M(A)$. Then $\bar\pi$ is a faithful representation of $M(A)$ on $\mathcal{H}$ and $\bar\pi_{\bar\alpha} \rtimes_r \lambda$ is a faithful representation of $M(A) \rtimes_{\bar\alpha,r} G$ on $\mathcal{H} \otimes L^2(G)$. If we identify $M(A \rtimes_{\alpha,r} G)$ with its image on $\mathcal{H} \otimes L^2(G)$, the above *-homomorphism $M(A) \rtimes_{\bar\alpha,r} G \to M(A \rtimes_{\alpha,r} G)$ is just the *-monomorphism $\bar\pi_{\bar\alpha} \rtimes_r \lambda$.

Let $(A,\alpha) \in \mathcal{C}^*_G$ and let $E_A = (C_0(G/H) \otimes A) \rtimes_{\Delta^\alpha,r} G$. Replacing (A,α) by the pair $(C_0(G/H) \otimes A, \Delta^\alpha)$ in the argument of the previous paragraph, we get a canonical embedding

$$\kappa : M(C_0(G/H) \otimes A) \rtimes_{\bar\Delta^\alpha,r} G \to M(E_A),$$

where $\bar\Delta^\alpha$ is the canonical extension of the diagonal action Δ^α to $M(C_0(G/H) \otimes A)$. If π is the embedding $a \to 1 \otimes a$ of A in $M(C_0(G/H) \otimes A)$, π is G-equivariant and the corresponding crossed product map

$$\pi_r : A \rtimes_{\alpha,r} G \to M(C_0(G/H) \otimes A) \rtimes_{\bar\Delta^\alpha,r} G$$

is an embedding. Let Φ_A denote the embedding

$$\kappa\pi_r : A \rtimes_{\alpha,r} G \to M(E_A).$$

If I is an α_G-invariant ideal of A, and $q : A \to A/I$ is the quotient map, associated with the *-homomorphism $\tilde{q} = id \otimes q : C_0(G/H) \otimes A \to C_0(G/H) \otimes (A/I)$ is the surjective crossed product *-homomorphism

$$\tilde{q}_r : E_A \to E_{A/I}.$$

This map extends to a *-homomorphism $M(E_A) \to M(E_{A/I})$, also denoted by $\tilde{q}_r$, which need not be surjective.

Lemma 4.2 $\tilde{q}_r \Phi_A = \Phi_{A/I} q_r$.

Proof: Let π and λ be the canonical embeddings of $C_0(G/H) \otimes A$ and G in $M(E_A)$, respectively, and let $\bar{\pi}$ be the embedding of $M(C_0(G/H) \otimes A)$ in $M(E_A)$ obtained by extending π as above. Then for $a \in C_c(G, A)$,

$$\Phi_A(a) = \int_G \bar{\pi}(1 \otimes a(s))\lambda_s \, ds.$$

The linear span of the subset

$$\left\{ \int_G \pi(f \otimes b(t))\lambda_s \, ds : f \in C_0(G/H), b \in C_c(G, A) \right\}$$

of E_A is dense in E_A, and for $a, b \in C_c(G, A)$, $f \in C_0(G/H)$,

$$\begin{aligned}
\tilde{q}_r(\Phi_A(a))\bar{q}_r(\int_G \pi(f \otimes b(t)\lambda_t \; dt) &= \tilde{q}_r(\Phi_A(a) \int_G \pi(f \otimes b(t)\lambda_t \, dt) \\
&= \bar{q}_r(\int_G \int_G \pi(f \otimes a(s)\alpha_s(b(t)))\lambda_{st} \, ds \, dt) \\
&= \int_G \int_G \pi(f \otimes q(a(s))q(\alpha_s(b(t))))\lambda_s t \, ds \, dt \\
&= \Phi_{A/I}(q(a))\bar{q}_r(\int_G \pi(f \otimes b(t)\lambda_t \, dt).
\end{aligned}$$

Thus $\tilde{q}_r \Phi_A = \Phi_{A/I} q$. □

Lemma 4.3 For $f \in C_0(G/H)$ and $a \in A \rtimes_{\alpha,r} G$, $\bar{\pi}(f \otimes 1)\Phi_A(a) \in E_A$.

Proof: By the continuity of Φ_A, it is enough to check this for a in the dense subset $C_c(G, A)$ of $A \rtimes_{\alpha,r} G$. Then

$$\bar{\pi}(f \otimes 1)\Phi_A(a) = \int_G \pi(f \otimes a(s))\lambda_s \, ds \in E_A.$$

□

Fix an invariant probability measure μ on G/H. The map $P : C_0(G/H)\otimes A \to A$ given by

$$P(f\otimes a) = \left(\int_{G/H} f\,d\mu\right) a$$

is completely positive, contractive and G-equivariant. The corresponding complete contraction $P_r : E_A \to A \rtimes_{\alpha,r} G$ is given by

$$P_r(\int_G \pi(f\otimes a(s))\lambda_s\,ds) = \left(\int_G f\,d\mu\right) a$$

for $f\in C_0(G/H), a\in C_c(G,A)$.

Lemma 4.4 *(i) For $f\in C_0(G/H)$ and $a\in A\rtimes_{\alpha,r} G$,*

$$P_r(\bar\pi(f\otimes 1)\Phi_A(a)) = \left(\int_G f\,d\mu\right) a.$$

(ii) $P_r(E_I)\subseteq I\rtimes_{\alpha|,r} G$.

Proof: (i) If $a\in C_c(G,A)$,

$$\bar\pi((f\otimes 1)\Phi_A(a)) = \int_G \pi(f\otimes a(s))\lambda_s\,ds$$

and

$$P_r(\bar\pi((f\otimes 1)\Phi_A(a)) = \left(\int_G f\,d\mu\right) a,$$

by the definition of P_r. The identity for general a now follows by the continuity of Φ_A and P_r.

(ii) This is immediate from the definition of P_r.

Theorem 4.5 *Let G be a σ-compact group. If G has a closed exact subgroup H which has finite covolume in G, then G is exact.*

Proof: Let μ be an invariant probability measure on G/H. We must show that if A is a C*-algebra with an action α of G and I is an α_G-invariant ideal of A with quotient map $q : A\to A/I$, then $\ker q_r\subseteq I\rtimes_{\alpha,r} G$.

Let $x\in\ker q_r$. By Lemma 4.2, if $f\in C_0(G/H)$,

$$\bar q_r(\bar\pi(f\otimes 1)\Phi_A(x)) = \bar\pi(f\otimes 1)\bar q_r(\Phi_A(x)) = \bar\Phi_{A/I}(q_r(x)) = 0,$$

and so $\bar\pi(f\otimes 1)\Phi_A(x)\in\ker\bar q_r$. Since H is exact, the sequence

$$0\to I\rtimes_{\alpha|,r} H\to A\rtimes_{\alpha,r} H\to (A/I)\rtimes_{\dot\alpha,r} H\to 0$$

is exact. By Corollary 3.10, the sequence

$$0 \to E_I \to E_A \to E_{A/I} \to 0$$

is exact, so that $\ker \bar{q}_r = E_I$. Thus $\bar{\pi}(f \otimes 1)\Phi_A(x) \in E_I$ and

$$\left(\int_G f\,d\mu\right) x = P_r(\bar{\pi}(f \otimes 1)\Phi_A(x)) \in I \rtimes_{\alpha|,r} G,$$

by Lemma 4.4 (ii). Since G is σ-compact, there is a compact subset K of G such that $\mu(K) \neq 0$. Choosing $f \in C_0(G/H)$ such that $f(g) = 1$ for $g \in K$, it follows that $\int_G f\,d\mu \neq 0$, which implies that $x \in I \rtimes_{\alpha|,r} G$, as required. □

5. Extension of an exact group by an exact group.

The main result of this section is

Theorem 5.1 *Let G be a locally compact group and let N be a closed normal subgroup of G. If N and G/N are exact, then G is exact.*

Let N be a closed normal subgroup of G and let $(A, \alpha) \in \mathcal{C}^*_G$. As indicated at the beginning of §3, there are a twisted action (γ^α, τ) of G on $A \rtimes_\alpha N$ relative to N canonically associated with α and a natural isomorphism $A \rtimes_\alpha G \cong (A \rtimes_\alpha N) \rtimes_{\gamma^\alpha,\tau} G$. If $\{\pi, U\}$ is a universal covariant pair of representations of the system $\{A, G, \alpha\}$ on a Hilbert space $\mathcal{H}$ and $A \rtimes_\alpha N$ is identified with its image under the representation $\pi \rtimes U$ to $B(\mathcal{H})$, γ^α and τ are given by

$$\gamma^\alpha_s(x) = U_s x U_{s^{-1}}$$

and

$$\tau_n = U_n$$

for $x \in A \rtimes_\alpha N$, $s \in G$ and $n \in N$. That γ^α and τ have the stated properties follows from the proof of the reduced case, which is given in the following proposition. We define a twisted action of G on the reduced crossed product $A \rtimes_{\alpha,r} N$ with analogous properties.

By Lemma 3.1 (3), the restriction of the representation $(\pi \otimes 1) \rtimes_r (U \otimes \lambda^G)$ on $\mathcal{H} \otimes L^2(G)$ to $A \rtimes_{\alpha,r} N$ is faithful. Identifying $A \rtimes_{\alpha,r} N$ with its image under this representation, an action $\gamma^{\alpha,r}$ of G on $A \rtimes_{\alpha,r} N$ is given by

$$\gamma^{\alpha,r}_s(x) = (U_s \otimes \lambda_s) x (U_{s^{-1}} \otimes \lambda_{s^{-1}})$$

for x in $A \rtimes_{\alpha,r} N$ and $s \in G$, and a twisting map τ' is given by

$$\tau'(t) = U_t \otimes \lambda_t$$

for $t \in N$. It is immediate that the canonical quotient map $A \rtimes_\alpha N \to A \rtimes_{\alpha,r} N$ is twist-equivariant relative to this action and twisting.

Proposition 5.2 *There is an isomorphism $\Phi_A : (A \rtimes_{\alpha,r} N) \rtimes_{\gamma^{\alpha,r},\tau',r} G \to A \rtimes_{\alpha,r} G$ which is natural in the sense that if $\{B, G, \beta\}$ is another G-covariant system and $\theta : A \to B$ is a G-equivariant *-homomorphism with associated homomorphisms*

$$\theta_r : A \rtimes_{\alpha,r} G \to B \rtimes_{\beta,r} G$$

and

$$\tilde{\theta} : (A \rtimes_{\alpha,r} N) \rtimes_{\gamma^{\alpha,r},\tau',r} G \to (B \rtimes_{\beta,r} N) \rtimes_{\gamma^{\beta,r},\tau',r} G,$$

then $\Phi_B \tilde{\theta} = \theta_r \Phi_A$, i.e. the diagram

$$\begin{tikzcd}
(A \rtimes_{\alpha,r} N) \rtimes_{\gamma^{\alpha,r},\tau',r} G \arrow[r, "\tilde{\theta}"] \arrow[d, "\Phi_A"'] & (B \rtimes_{\beta,r} N) \rtimes_{\gamma^{\beta,r},\tau',r} G \arrow[d, "\Phi_B"] \\
A \rtimes_{\alpha,r} G \arrow[r, "\theta_r"] & B \rtimes_{\beta,r} G
\end{tikzcd}$$

commutes.

Proof: Let$\{\pi, U\}$ be a covariant pair of representations of $\{A, G, \alpha\}$ on the Hilbert space $\mathcal{H}$ with π faithful. The crossed product $A \rtimes_{\alpha,r} N$ can be identified with the C*-algebra generated by the operators on $\mathcal{H} \otimes L^2(G)$ of form

$$\int_G (\pi(a(n)) \otimes 1)(U_n \otimes \lambda_n) dm_N(n)$$

for $a \in C_c(N, A)$. By Lemma 3.1 (3), $(A \rtimes_{\alpha,r} N) \rtimes_{\alpha,\tau',r} G$ can be identified with the C*-algebra (in fact the closed linear span) generated by the set $\{T_f : f \in C_c(G \times N, A)\}$ of operators on $\mathcal{H} \otimes L^2(G) \otimes L^2(G/N)$, where

$$\begin{aligned}
T_f &= \int_G \int_N (\pi(f(s,n)) \otimes 1 \otimes 1)(U_n \otimes \lambda_n \otimes 1)(U_s \otimes \lambda_s \otimes \dot{\lambda}_{sN}) dm_N(n) dm_G(s) \\
&= \int_G \int_N (\pi(f(n^{-1}s, n)) \otimes 1 \otimes 1)(U_s \otimes \lambda_s \otimes \dot{\lambda}_{sN}) dm_N(n) dm_G(s) \\
&\in A \rtimes_{\alpha,r} G
\end{aligned}$$

for $f \in C_c(G \times N, A)$, $A \rtimes_{\alpha,r} G$ being identified here with its image under the integrated form representation of the pair $\{\pi \otimes 1 \otimes 1, U \otimes \lambda \otimes \dot{\lambda}\}$ (which is unitarily equivalent to the pair $\{(\pi \otimes 1) \otimes 1, (U \otimes \dot{\lambda}) \otimes \lambda\}$).

Let $g \in C_c(N)$ be a nonnegative real-valued function such that

$$\int_N g(n) dm_N(n) = 1,$$

let $a \in C_c(G, A)$ and let $f \in C_c(G \times N, A)$ be given by

$$f(s,n) = a(s)g(n).$$

Choosing g with support in a suitable neighbourhood of the identity of N, the operator of the above form corresponding to this f can be made to approximate

$$\int_G (\pi(a(s)) \otimes 1 \otimes 1)(U_s \otimes \lambda_s \otimes \dot{\lambda}_{sN}) dm_G(s)$$

arbitrarily closely in norm. Thus if we define Φ_A on $\{T_f : f \in C_c(G \times N, A)\}$ by

$$\Phi_A(T_f) = \int_G \int_N (\pi(f(n^{-1}s, n)) \otimes 1)(U_s \otimes \lambda_s) dm_N(n) dm_G(s),$$

Φ_A extends uniquely to a *-isomorphism of $(A \rtimes_{\alpha,r} N) \rtimes_{\gamma^{\alpha,r},\tau',r} G$ onto $A \rtimes_{\alpha,r} G$. The naturalness of Φ_A is a straightforward consequence of this definition. □

Combes [Co] introduced a notion of Morita equivalence for actions of a locally compact group G. Let $(A, \alpha), (B, \beta) \in \mathcal{C}^*_G$ and let X be a B–A equivalence bimodule. A continuous action u of G on X is a set of bijective linear isometries $\{u_s : s \in G\}$ of X such that the map $s \to u_s$ is strongly continuous and for each s

$$u_s(x < y, z >_A) = u_s(x) < u_s(y), u_s(z) >_A$$

for $x, y, z \in X$. The actions β and α, or more accurately the pairs (B, β) and (A, α), are *Morita equivalent* if there is a continuous action u of G on X such that for each s,

$$\alpha_s(< x, y >_A) =< u_s(x), u_s(y) >_A$$

and

$$\beta_s(< x, y >_B) =< u_s(x), u_s(y) >_B$$

for $x, y \in X$. When (A, α) and (B, β) are Morita equivalent, there are Morita equivalences between $A \rtimes_\alpha G$ and $B \rtimes_\beta G$, and between $A \rtimes_{\alpha,r} G$ and $B \rtimes_{\beta,r} G$, by [Co, §3].

Echterhoff [Ech] has extended this idea to twisted actions as follows. Let N be a closed normal subgroup of G and let (α, τ) and (β, σ) be twisted actions of G on A and B, respectively, relative to N. Then (β, σ) and (α, τ) are Morita equivalent relative to the pair (X, u) if u satisfies the above identities and also

$$u_n x = \sigma_n x \tau_n^{-1}$$

for $n \in N$ and $x \in X$. Moreover, if $\{A, G, \alpha, \tau\}$ and $\{B, G, \beta, \sigma\}$ are Morita equivalent, then $A \rtimes_{\alpha,\tau} G$ is Morita equivalent to $B \rtimes_{\beta,\sigma} G$ [Ech, p.174, Remark 2]. We shall show the corresponding result for the reduced crossed products.

Although the arguments that follow involving Morita equivalence could be expressed solely in terms of equivalence bimodules, we have found it easier to bring out some of the functorial aspects of the proof using the equivalent idea of a linking algebra, due to Brown, Green and Rieffel [BGR], which we now briefly recall. If C is a nonzero C*-algebra, a projection $p \in M(C)$ is *full* if the linear span of the set $CpC = \{xpy : x, y \in C\}$ is dense in C. Suppose that $p, q \in M(C)$ are full projections such that $p + q = 1$, and let $A = pCp$, $B = qCq$ and $X = pCq$. If we define A- and B- valued inner products on X by

$$< x, y >_A = xy^*, \ < x, y >_B = x^* y$$

and let A and B act on X by left and right multiplication, respectively, then X becomes an A–B equivalence bimodule. Conversely, if A and B are C*-algebras and X is an A–B equivalence bimodule, then we can find a C*-algebra C, known as a linking algebra for A and B, and full projections $p, q \in M(C)$ such that $p + q = 1$, $A \cong pCp$, $B \cong qCq$ and such that X and pCq are isomorphic as A–B equivalence modules. Passage from X to the corresponding C is functorial, in a sense to be made precise in what follows.

Let IMP be the category whose objects are triples $\{A, B, X\}$ consisting of a Morita equivalent pair of C*-algebras A, B and an A–B equivalence bimodule X. Given two such triples $\{A, B, X\}$ and $\{A_1, B_1, X_1\}$, a map between them is a triple $\{\phi, \psi, \omega\}$, consisting of *-homomorphisms $\phi : A \to A_1$, $\psi : B \to B_1$, and a linear map $\omega : X \to X_1$ satisfying

$$\omega(ax) = \psi(a)\omega(x), \ \omega(xb) = \omega(x)\phi(b),$$

$$\omega(x) < \omega(y), \omega(z) >_{B_1} = < \omega(x), \omega(y) >_{A_1} \omega(z),$$

$$< \omega(x), \omega(y) >_{A_1} = \phi(< x, y >_A), \ < \omega(x), \omega(y) >_{B_1} = \psi(< x, y >_B)$$

for $a \in A$, $b \in B$ and $x, y, z \in X$. Given two pairs A, A_1 and B, B_1 of C*-algebras, *-homomorphisms $\phi : A \to A_1$ and $\psi : B \to B_1$ will be said to be *Morita equivalent* if there is an $\omega : X \to X_1$ such that $\{\phi, \psi, \omega\}$ is a map in IMP. This is consistent with the definition of Morita equivalence for actions when ϕ and ψ are automorphisms of A and B, respectively.

Let A and B be Morita equivalent C*-algebras and let X be an A–B equivalence bimodule. To see that the associated linking algebra is related to X functorially, we recall its construction (cf. [BGR, proof of Theorem 1]). Let X^* be the B–A equivalence bimodule conjugate to X. There is a conjugate linear bijection $x \to x^*; X \to X^*$ such that

$$bx^* = (xb^*)^*, \ x^*a = (a^*x)^*, \ < x^*, y^* >_A = < y, x >_A, \ < x^*, y^* >_B = < y, x >_B$$

for $a \in A$, $b \in B$ and $x, y \in X$. For $x, y \in X$ let x^*y and xy^* be the elements $< x, y >_A$ and $< x, y >_B$ of A and B, respectively. The set of matrices

$$C = \{ \begin{bmatrix} a & x \\ y^* & b \end{bmatrix} : a \in A, \ b \in B, \ x, y \in X\},$$

with matrix addition and multiplication, is a *-algebra. Moreover $X \oplus B$ has left and right actions of C and B, respectively, and C- and B-valued inner products can be defined in such a way that $X \oplus B$ becomes a C–B equivalence bimodule. A norm on $X \oplus B$ is defined in terms of the B-valued inner product on X and the norm of B relative to which the associated C*-norm on C is complete. The projections

$$p = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad q = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

are in $M(C)$, are both full, and $A \cong pCp$ and $B \cong qCq$. Moreover pCq and X are isomorphic as A–B equivalence bimodules. Let $\{A, B, X\}, \{A_1, B_1, X_1\}$ be objects of IMP and let $\{\phi, \psi, \omega\} : \{A, B, X\} \to \{A_1, B_1, X_1\}$ be a morphism. If C and C_1 are the linking algebras constructed from $\{A, B, X\}$ and $\{A_1, B_1 X_1\}$, respectively, by this procedure, a *-homomorphsim $\phi = \Phi_\{\phi, \psi, \omega\} : C \to C_1$ is defined by

$$\Phi_\{\phi, \psi, \omega\}(\begin{bmatrix} a & x \\ y^* & b \end{bmatrix}) = \begin{bmatrix} \phi(a) & \omega(x) \\ (\omega(y))^* & \psi(b) \end{bmatrix}.$$

Then $\Phi(p) = p$, $\Phi(q) = q$, $\Phi|pCp = \psi$ and $\Phi|qCq = \phi$. Let LINK be the category whose objects are triples $\{C, p, q\}$ consisting of a C*-algebra C with full projections $p, q \in M(C)$ such that $p + q = 1$. If $\{C, p, q\}$ and $\{C_1, p_1, q_1\}$ are objects of LINK, a morphism θ from $\{C, p, q\} \to \{C_1, p_1, q_1\}$ is a a *-homomorphism $C \to C_1$ such that $\theta(px) = p_1\theta(x)$ for $x \in C$. Writing $C_{\{A,B,X\}}$ for the linking algebra constructed from the triple $\{A, B, X\}$, the map $\{A, B, X\} \to C_{\{A,B,X\}}$ is a functor from IMP to LINK giving an equivalence of categories.

For strongly Morita equivalent C*-algebras A and B, the order-preserving correspondence between the ideals of A and B described in §2 can be expressed elegantly in terms of C. If I is an ideal of A or B, then I_C, the ideal of C generated by I, is just the closure of the linear span of the set CIC. If I is an ideal of C, let

$$I_A = pIp, \; I_B = qIq.$$

Then $I_A = I \cap pCp = I \cap A$ and $I_B = I \cap qCq = I \cap B$, and it follows easily from the fullness of p and q that for any ideal I of C, $I = (I_A)_C = (I_B)_C$. For any ideal I of A, $I = (I_C)_A$, and similarly for B. The map $I \to (I_C)_B$ is thus an order-preserving bijection from the ideals of A to the ideals of B.

Lemma 5.3 *Let* $\{A, B, X\}, \{A_1, B_1, X_1\} \in$ IMP, *and let* $\{\phi, \psi, \omega\} : \{A, B, X\} \to \{A_1, B_1, X_1\}$ *be a morphism. Then* $\ker \psi$ *corresponds to* $\ker \phi$ *under the above bijection.*

Proof: If $\Phi : C_{\{A,B,X\}} \to C_{\{A_1,B_1,X_1\}}$ is the *-homomorphism corresponding to $\{\phi, \psi, \omega\}$, then $x \in \ker \Phi \cap A \Leftrightarrow \phi(x) = 0 \Leftrightarrow x \in \ker \phi$. Hence $\ker \phi = \ker \Phi \cap A$. Similarly $\ker \psi = \ker \Phi \cap B$. Thus $\ker \phi$, $\ker \Phi$ and $\ker \psi$ are corresponding ideals. □

Let $\{A, G, \alpha, \sigma\}$ and $\{B, G, \beta, \tau\}$ be Morita equivalent twisted covariant systems relative to a pair (X, u) consisting of an A–B equivalence bimodule X

with an action u of G, with twisting relative to the normal subgroup N of G. Then for each $s \in G$, $\{\alpha_s, \beta_s, u_s\}$ is a map in IMP, in fact an automorphism of $\{A, B, X\}$. If Γ_s denotes the corresponding automorphism of the linking algebra $C = C_{\{A,B,X\}}$, then

$$\Gamma_s(\begin{bmatrix} a & x \\ y^* & b \end{bmatrix}) = \begin{bmatrix} \alpha_s(a) & u_s x \\ (u_s y)^* & \beta_s(b) \end{bmatrix}.$$

It is immediate that $s \to \Gamma_s$ is a continuous action of G on $C_{\{A,B,X\}}$, Γ fixes p and q, $\Gamma|A = \alpha$ and $\Gamma|B = \beta$. For $n \in N$

$$\kappa_n = \begin{bmatrix} \sigma_n & 0 \\ 0 & \tau_n \end{bmatrix}$$

is in $M(C)$ and the map $\kappa : n \to \kappa_n$ is a twisting map for the action Γ.

The canonical embedding of C in $M(C \rtimes_\Gamma G)$ extends to an embedding of $M(C)$ in $M(C \rtimes_\Gamma G)$, where $x \in M(C)$ is identified with the element of $M(C \rtimes_\Gamma G)$ which sends f in $C_c(G, C)$ to xf. With this identification, p and q are in $M(C \rtimes_\Gamma G)$, are full projections for $C \rtimes_\Gamma G$, and there are canonical isomorphisms

$$p(C \rtimes_\Gamma G)p \cong A \rtimes_\alpha G, \quad q(C \rtimes_\Gamma G)q \cong B \rtimes_\beta G,$$

by [Co, §6]. In fact, if $\pi_0 : C \to M(C \rtimes_\Gamma G)$ and $U_0 : G \to M(C \rtimes_\Gamma G)$ are the canonical embeddings and A is identified with pCp, then $\{\pi_0|A, U_0\}$ is a covariant pair of representations of $\{A, G, \beta\}$, and the integrated form of this pair is a *-homomorphism $\theta_A : A \rtimes_\alpha G \to M(C \rtimes_\Gamma G)$. The image of $A \rtimes_\alpha G$ under θ_α is just $p(C \rtimes_\Gamma G)p$. To see that θ_A is injective, let $\{\pi, U\}$ be a universal covariant pair of representations of $\{A, G, \alpha\}$ on a Hilbert space $\mathcal{H}$. By the equivariant form of Stinespring's theorem, there are a Hilbert space $\mathcal{K}$ containing $\mathcal{H}$ and a covariant pair of representations $\{\pi_1, U_1\}$ of $\{C, G, \Gamma\}$ on $\mathcal{K}$ such that $\bar{\pi}_1(p)$ is the projection onto $\mathcal{H}$ and the pair $\{\bar{\pi}_1(p)(\pi_1|B)\bar{\pi}_1(p), \bar{\pi}_1(p)U_1\bar{\pi}_1(p)\}$ is a covariant pair of representations of $\{A, G, \alpha\}$ unitarily equivalent to $\pi \rtimes U$. Then $\bar{\pi}_1(p)\,((\pi_1 \rtimes U_1) \circ \theta_A)\,\bar{\pi}_1(p)$ is unitarily equivalent to $\pi \rtimes U$, which is faithful. Thus θ_A is faithful.

The canonical quotient map $q_C : C \rtimes_\Gamma G \to C \rtimes_{\Gamma,\kappa,r} G$ extends to a *-homomorphism $\bar{q}_C : M(C \rtimes_\Gamma G) \to M(C \rtimes_{\Gamma,\kappa,r} G)$, and the covariant pair $\{\bar{q}_C \circ (\pi_0|B), \bar{q}_C \circ U\}$ of homomorphisms of $\{A, G, \alpha\}$ is twist-preserving. It follows by Lemma 3.1 that $q_C \circ \theta_A$ has a factorisation $q_C \circ \theta_A = \bar{\theta}_A \circ q_A$, where $q_A : A \rtimes_\alpha G \to A \rtimes_{\alpha,\sigma,r} G$ is the quotient map and $\bar{\theta}_A : A \rtimes_{\alpha,\sigma,r} G \to C \rtimes_{\Gamma,\kappa,r} G$ is a *-isomorphism with image $\bar{p}(C \rtimes_{\Gamma,\kappa,r} G)\bar{p}$, where $\bar{p}$ is the projection $\bar{q}_C(p)$. Applying analogous considerations to the crossed products involving B, we obtain *-monomorphisms

$$\theta_B : B \rtimes_\beta G \to q(C \rtimes_\Gamma G)q$$

and

$$\bar{\theta}_B : B \rtimes_{\beta,\sigma,r} G \to \bar{q}(C \rtimes_{\Gamma,\kappa,r} G)\bar{q}$$

such that $q_C \circ \theta_B = \bar{\theta}_B \circ q_B$, where $\bar{q} = \bar{q}_C(q)$. Since the fullness of the projections p and q for $C \rtimes_\Gamma G$ implies the fullness of $\bar{p}$ and $\bar{q}$ for $C \rtimes_{\Gamma,\kappa,r} G$, we obtain

Proposition 5.4 *If $\{A, G, \alpha, \tau\}$ and $\{B, G, \beta, \sigma\}$ are Morita equivalent twisted covariant systems relative to N, then the C*-algebras $A \rtimes_{\alpha,\tau,r} G$ and $B \rtimes_{\beta,\sigma,r} G$ are strongly Morita equivalent. Moreover the diagram*

$$\begin{tikzcd}
A \rtimes_\alpha G \arrow[r, "q_A"] \arrow[d, "\theta_A"] & A \rtimes_{\alpha,\tau,r} G \arrow[d, "\bar{\theta}_A"] \\
C \rtimes_\Gamma G \arrow[r, "q_C"] & C \rtimes_{\Gamma,\kappa,r} G \\
B \rtimes_\beta G \arrow[r, "q_B"] \arrow[u, "\theta_B"] & B \rtimes_{\beta,\sigma,r} G \arrow[u, "\bar{\theta}_B"]
\end{tikzcd}$$

*commutes, so that the *-homomorphisms q_A and q_B are Morita equivalent.*

If $\{A, G, \alpha, \tau\}$ is a twisted covariant system relative to N, then there is an action β of G on $E_A = C_0(G/N, A) \rtimes_{\Delta^\alpha, \bar{\tau}} G$ such that the twisted covariant systems $\{A, G, \alpha, \tau\}$ and $\{E_A, G, \beta, 1_N\}$ are Morita equivalent, where 1_N is the trivial twisting map $n \to 1 \in M(B)$ [Ech, Theorem 1].

An equivalence bimodule Y_A giving this Morita equivalence is obtained by applying the mapping $\{\nu_A, id, \bar{\alpha}\}$ of IMP to the triple $\{\mathrm{Ind}(A, \alpha) \rtimes_{\bar{\alpha}} G, A, X_A\}$ of §2, with $H = N$, where $\nu_A : \mathrm{Ind}(A, \alpha) \to C_0(G/N, A)$ is the map isomorphism of §2 and $(\bar{\alpha}(x))(s) = \alpha_s(x(s))$ for $x \in C_c(G, A)$. Letting $X_1 = C_c(G, A)$, $E_0 = C_c(G, C_0(G/N, A)))$ and $B_0 = C_c(N, A)$, with the convolution products relative to the actions Δ^α and α, respectively, the algebras E_0 and B_0 having the C*-norms and positive cones coming from their canonical embeddings in $C_0(G/N, A) \rtimes_{\Delta^\alpha} G$ and $A \rtimes_\alpha N$, respectively, the resulting E_0–B_0 equivalence bimodule structure on X_1 is given by

$$\begin{aligned}
(fx)(r) &= \int_G f(s, rN)\alpha_s(x(s^{-1}r))dm_G(s) \\
(xg)(r) &= \int_H x(rt)\alpha_{rt}(g(t^{-1}))dm_H(t) \\
\langle x, y\rangle_{E_0}(s, rN) &= \int_H \Delta_G(s^{-1}rt)x(rt)\alpha_s(y(s^{-1}rt)^*)dm_H(t) \\
\langle x, y\rangle_{B_0}(t) &= \delta(t)\int_G \alpha_{s^{-1}}(x(s)^*y(st))dm_G(s).
\end{aligned}$$

The action β of G on E_A is given on an element f of the dense *-subalgebra $C_c(G, C_0(G/N, A))$ of E_A by

$$(\beta_s(f))(r, tN) = f(r, tsN),$$

where f is considered as a function on $G \times (G/N)$, and an action u of G on X_1 is given by

$$(u_s x)(t) = \Delta_G(s)\Delta_{G/N}(s)^{-1/2}x(ts).$$

Recalling that, by Remark 2.5 (2), there is a canonical isomorphism $A \cong A \rtimes_{\alpha,1_n} N$, the equivalence bimodule Y_A is then obtained by completing X_1 with respect to the norm $x \to || < x, x >_A ||^{1/2}$ and extending the left and right actions to E_A and A, respectively, by continuity.

Before proving Theorem 5.1 we require one further observation. Let β be a twisted action of G on a C*-algebra D relative to N with trivial twisting map 1_N, so that $\beta_n = id_D$ for $n \in N$. Then an action $\bar{\beta}$ of G/N on D is given by

$$\bar{\beta}_{sN}(x) = \beta_s(x).$$

The analogue of the following result for full twisted crossed products is implicit in [Ech].

Lemma 5.5 *There is an isomorphism $\Psi_D : D \rtimes_{\bar{\beta},r} (G/N) \to D \rtimes_{\beta,1_N,r} G$ which is natural in the sense that, if $\{D, G, \beta, 1_N\}$ and $\{D_1, G, \beta_1, 1_N\}$ are twisted covariant systems and $\theta : D_1 \to D$ is a G-equivariant twist-preserving *-homomorphism, then the diagram*

$$\begin{tikzcd}
D_1 \rtimes_{\bar{\beta}_1,r} (G/N) \arrow[r, "\theta_r"] & D \rtimes_{\bar{\beta},r} (G/N) \\
D_1 \rtimes_{\beta_1,1_N,r} G \arrow[u, "\Psi_{D_1}"'] \arrow[r, "\theta_{N,r}"] & D \rtimes_{\beta,1_N,r} G \arrow[u, "\Psi_D"']
\end{tikzcd} \qquad (*)$$

*commutes, where the horizontal maps are the *-homomorphisms corresponding canonically to θ.*

Proof: Let $\pi_0 : D \to M(D \rtimes_{\bar{\beta},r} (G/N))$ and $\lambda_0 : G/N \to M(D \rtimes_{\bar{\beta},r} (G/N))$ be the canonical monomorphisms. If $\dot{\lambda}_0$ denotes the composition of λ_0 with the quotient homomorphism $G \to G/N$, and $M(D \rtimes_{\bar{\beta},r} (G/N))$ is represented faithfully on some Hilbert space $\mathcal{H}$, the pair $\{\pi_o, \dot{\lambda}_0\}$ is a covariant pair of representations of the system $\{D, G, \beta\}$ which is twist-covariant for the twisting map 1_N. The representation $\pi_0 \rtimes_{1_n} \dot{\lambda}_0$ of $D \rtimes_{\beta,1_N} G$ in $M(D \rtimes_{\bar{\beta},r} (G/N))$ has image $D \rtimes_{\bar{\beta},r} (G/N)$, and, by Proposition 3.5 (1), has a factorisation $\pi_0 \rtimes_{1_N} \dot{\lambda}_0 = \psi_D \circ q$, where ψ_D is a *-homomorphism from $D \rtimes_{\beta,1_N,r} G$ onto $D \rtimes_{\bar{\beta},r} (G/N)$ and $q : D \rtimes_{\beta,1_N} G \to D \rtimes_{\beta,1_N,r} G$ is the canonical quotient map. Since π_0 is faithful, ψ_D is injective.

If $\theta : D_1 \to D$ is a G-covariant twist-preserving *-homomorphism, let $\bar{\theta}_{N,r} = \psi_D^{-1}\theta_r\psi_{D_1}$. By the definition of ψ_D and ψ_{D_1}, the diagram

$$\begin{tikzcd}
D_1 \rtimes_{\beta_1} G \arrow[r, "\theta_u"] \arrow[d] & D \rtimes_{\beta} G \arrow[d] \\
D_1 \times_{\beta_1,1_N,r} G \arrow[r, "\bar{\theta}_{N,r}"] & D \rtimes_{\beta,1_N,r} G
\end{tikzcd}$$

commutes, where the vertical arrows are the canonical quotient maps. By Proposition 3.5 (2), $\bar{\theta}_{N,r} = \theta_{N,r}$, from which the commutativity of $(*)$ is immediate. □

Proof of Theorem 5.1. Assume that N and G/N are exact. Let $(I, \alpha|), (A, \alpha), (B, \dot{\alpha}) \in \mathcal{C}^*_G$ and let $\iota : I \to A$ and $q : A \to B$ be G-equivariant *-homomorphisms such that $\operatorname{im} \iota = \ker q$, i.e. such that the sequence

$$0 \longrightarrow I \xrightarrow{\iota} A \xrightarrow{q} B \longrightarrow 0 \tag{1}$$

is exact. To prove the theorem we must show that the sequence

$$0 \longrightarrow I \rtimes_{\alpha|,r} G \xrightarrow{\iota_r} A \rtimes_{\alpha,r} G \xrightarrow{q_r} B \rtimes_{\dot{\alpha},r} G \longrightarrow 0 \tag{2}$$

is exact. In what follows we shall identify I with its image in A.

Since N is exact, the sequence

$$0 \longrightarrow I \rtimes_{\alpha|,r} N \xrightarrow{\iota_r} A \rtimes_{\alpha,r} N \xrightarrow{q_r} B \rtimes_{\dot{\alpha},r} N \longrightarrow 0 \tag{3}$$

is exact. Let $I_N = I \rtimes_{\alpha|,r} N$, $A_N = A \rtimes_{\alpha,r} N$ and $B_N = B \rtimes_{\dot{\alpha},r} N$. By Proposition 5.2 and the discussion preceding it, there are N-twisted actions (γ_I, τ), (γ_A, τ) and (γ_B, τ) of G on I_N, A_N and B_N, respectively, and isomorphisms

$$\Phi_I : I_N \rtimes_{\gamma_I,\tau,r} G \to I \rtimes_{\alpha|,r} G,$$

$$\Phi_A : A_N \rtimes_{\gamma_A,\tau,r} G \to A \rtimes_{\alpha,r} G$$

and

$$\Phi_B : B_N \rtimes_{\gamma_B,\tau,r} G \to B \rtimes_{\dot{\alpha},r} G$$

such that the diagram

$$\begin{tikzcd}
0 \arrow[r] & I_N \rtimes_{\gamma_I,\tau,r} G \arrow[r, "\iota_{N,r}"] \arrow[d, "\Phi_I"'] & A_N \rtimes_{\gamma_A,\tau,r} G \arrow[r, "q_{N,r}"] \arrow[d, "\Phi_A"'] & B_N \rtimes_{\gamma_B,\tau,r} G \arrow[r] \arrow[d, "\Phi_B"'] & 0 \\
0 \arrow[r] & I \rtimes_{\alpha|,r} G \arrow[r, "\iota_r"] & A \rtimes_{\alpha,r} G \arrow[r, "q_r"] & B \rtimes_{\dot{\alpha},r} G \arrow[r] & 0
\end{tikzcd}$$

commutes, where $\iota_{N,r}$ and $q_{N,r}$ are the crossed product homomorphisms corresponding to the *-homomorphisms $\iota_r|I_N$ and $q_r|A_N$, then latter G-equivariant and twist preserving relative to the twisted actions on I_N, A_N and B_N, as follows readily from the definitions. The exactness of (2) is thus equivalent to that of the sequence

$$0 \longrightarrow I_N \rtimes_{\gamma_I,\tau,r} G \xrightarrow{\iota_{\tau,r}} A_N \rtimes_{\gamma_A,\tau,r} N \xrightarrow{q_{\tau,r}} B_N \rtimes_{\gamma_B,\tau,r} N \longrightarrow 0 \tag{4}$$

By Corollary 3.10, A_N and $E_A = (C_0(G/N, A_N) \rtimes_{\Delta\gamma,\tilde{\tau}} G$ are Morita equivalent via the E_A–A_N equivalence bimodule obtained by completing $X_0 = C_c(G, A_n)$ with respect to the seminorm $|| \ ||$ given by

$$||x|| = || < x, x >_{A_N} ||^{1/2}.$$

Letting $C_A = C_{\{A,B,X\}}$, we identify E_A with pC_Ap and A_N with qC_Aq. If β_A is the action of G on E_A defined earlier such that $(\beta_A, 1_N)$ is Morita equivalent to (γ_A, τ) via (X_A, u_A), where u_A is the canonically defined action of G on X_A, let (Γ_A, κ) be the corresponding twisted action of G on C_A. We define X_I, X_B, C_I, C_B, β_I, β_B, Γ_I and Γ_B similarly. If $q_X : C_c(G, A_N) \to C_c(G, B_N)$ is the natural map given by

$$(q_X(f))(s) = q_r(f(s)),$$

for $s \in G$, then

$$||q_X(f)|| = || < q_X(f), q_X(f) >_{B_N} ||^{1/2} = ||q_r(< f, f >)_{A_N}||^{1/2} \leq ||f||.$$

If $\iota_X : C_c(G, I_N) \to C_c(G, A_N)$ is defined similarly starting from ι, since $\iota_r : I_N \to A_N$ is an isometric embedding, the same is true of ι_X. It follows that ι_X and q_X extend to an isometric embedding of X_I in X_A and a contraction of X_A onto X_B, respectively, which we will still denote by ι_X and q_X. Identifying X_I with its image in X_A we get a corresponding embedding $\iota_C : C_I \to C_A$. It is straightforward to verify that the diagram

$$\begin{tikzcd}
E_I \arrow[r, "\iota_E"] \arrow[d] & E_A \arrow[d] \\
C_I \arrow[r, "\iota_C"] & C_A \\
I_N \arrow[r, "\iota_r"] \arrow[u] & A_N \arrow[u]
\end{tikzcd}$$

commutes. Similarly the surjection q_X gives rise to a *-epimorphism $q_C : C_A \to C_B$ given by

$$q_C(\begin{bmatrix} b & x \\ y^* & a \end{bmatrix}) = \begin{bmatrix} q_E(b) & q_X(x) \\ q_X(y)^* & q_r(a) \end{bmatrix}.$$

Again it is routine to verify that the diagram

$$\begin{tikzcd}
E_A \arrow[r, "q_E"] \arrow[d] & E_B \arrow[d] \\
C_A \arrow[r, "q_C"] & C_B \\
A_N \arrow[r, "q_r"] \arrow[u] & B_N \arrow[u]
\end{tikzcd}$$

commutes. All the maps in these two diagrams are twist-preserving relative to the respective twistings for each algebra. Hence if we take the twisted crossed products of all the algebras by G, we get commuting diagrams

$$\begin{tikzcd} E_I \rtimes_{\beta_I,1_N,r} G \arrow[r, "(\iota_E)_r"] \arrow[d] & E_A \rtimes_{\beta_A,1_N,r} G \arrow[d] \\ C_I \rtimes_{\Gamma_I,\kappa,r} G \arrow[r, "(\iota_C)_r"] & C_A \rtimes_{\Gamma_A,\kappa,r} G \\ I_N \rtimes_{\gamma_I,\tau,r} G \arrow[r, "\iota_r"] \arrow[u] & A_N \rtimes_{\gamma_A,\tau,r} G \arrow[u] \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} E_A \rtimes_{\beta_A,1_N,r} G \arrow[r, "(q_E)_r"] \arrow[d] & E_B \rtimes_{\beta_B,1_N,r} G \arrow[d] \\ C_A \rtimes_{\Gamma_A,\kappa,r} G \arrow[r, "(q_C)_r"] & C_B \rtimes_{\Gamma_B,\kappa,r} G \\ A_N \rtimes_{\gamma_A,\tau,r} G \arrow[r, "q_r"] \arrow[u] & B_N \rtimes_{\gamma_B,\tau,r} G \arrow[u] \end{tikzcd}$$

In these diagrams the vertical arrows denote the embedding maps resulting from the identifications $E_I \rtimes_{\beta_I,1_N,r} G = p(C_I \rtimes_{\Gamma_I,\kappa,r} G)p$, $I_N \rtimes_{\gamma_I,\tau,r} G = q(C_I \rtimes_{\Gamma_I,\kappa,r} G)q$, etc. From the left-hand diagram it is apparent that the ideals $E_I \rtimes_{\beta_I,1_N,r} G$ of $E_A \rtimes_{\beta_A,1_N,r} G$ and $I_N \rtimes_{\gamma_I,\tau,r} G$ of $A_N \rtimes_{\gamma_A,\tau,r} G$ correspond. From the right-hand diagram the *-homomorphisms $(q_E)_r$ and q_r are seen to be Morita equivalent, so that, by Lemma 5.3, their kernels are corresponding ideals of $E_A \rtimes_{\beta_A,1_N,r} G$ and $A_N \rtimes_{\gamma_A,\tau,r} G$, respectively. Thus $\ker(q_E)_r = E_I \rtimes_{\beta_I,1_N,r} G$ if and only if $\ker q_r = I_N \rtimes_{\gamma_I,\tau,r} G$, that is, the sequence (4) is exact if and only if the sequence

$$0 \longrightarrow E_I \rtimes_{\beta_I,1_N,r} G \xrightarrow{(\iota_E)_r} E_A \rtimes_{\beta_A,1_N,r} G \xrightarrow{(q_E)_r} E_B \rtimes_{\beta_B,1_N,r} G \longrightarrow 0 \tag{5}$$

is exact.

Lemma 5.5 implies that the sequence (5) is exact if and only if the sequence

$$0 \longrightarrow E_I \rtimes_{\bar\beta_I,r} (G/N) \xrightarrow{(\iota_E)_r} E_A \rtimes_{\bar\beta_A,r} (G/N) \xrightarrow{(q_E)_r} E_B \rtimes_{\bar\beta_B,r} (G/N) \longrightarrow 0 \tag{6}$$

is exact.

Since the sequence (3) is exact by assumption, the sequence

$$0 \longrightarrow E_I \xrightarrow{\iota_E} E_A \xrightarrow{q_E} E_B \longrightarrow 0$$

is exact, by Corollary 3.10. Since G/N is exact by assumption, this implies that the sequence (6) is exact, from which the exactness of the sequences (5), (4) and (2) follow successively. Hence G is exact. □

6. Examples of exact groups.

a. Amenable Groups. The following result is essentially well-known.

Proposition 6.1 *Let G be an amenable locally compact group. Then G is exact.*

Proof: Let $(A, \alpha) \in \mathcal{C}^*_G$, and let I be an α_G-invariant ideal of A, so that the sequence

$$0 \to I \to A \to (A/I) \to 0$$

is exact. Then the diagram

$$\begin{tikzcd}
0 \arrow[r] & I \rtimes_{\alpha|} G \arrow[r] \arrow[d] & A \rtimes_{\alpha} G \arrow[r] \arrow[d] & (A/I) \rtimes_{\dot\alpha} G \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & I \rtimes_{\alpha|,r} G \arrow[r] & A \rtimes_{\alpha,r} G \arrow[r] & (A/I) \rtimes_{\dot\alpha,r} G \arrow[r] & 0
\end{tikzcd}$$

is commutative, where the vertical arrows are the canonical *-homomorphisms, and the top row is exact. Since G is amenable, the canonical *-homomorphisms $A \rtimes_\alpha G \to A \rtimes_{\alpha,r} G$, etc., are injective, from which it follows that the lower row of the diagram is exact. Hence G is exact. □

b. Discrete subgroups of semisimple Lie groups. Our goal here is to show that the discrete groups $SL_n(\mathbb{Z})$ are exact for $n = 1, 2, \ldots$, though we shall, in fact, prove a more general result. The following fact is undoubtedly known, although we lack a reference.

Proposition 6.2 *Let A and B be Morita equivalent C*-algebras. Then A is nuclear if and only if B is nuclear.*

Proof: Let (C, e, f) be a linking algebra for A and B, so that e, f are full projections in $M(C)$ such that $e + f = 1$, $A \cong eCe$ and $B \cong fCf$. Let C^{**} be represented as a von Neumann algebra on a Hilbert space $\mathcal{H}$. If we regard $M(C)$ as canonically embedded in C^{**}, the fullness conditions imply that e and f have central support 1 in C^{**}, from which it follows that $e(C^{**})'e \cong (C^{**})' \cong f(C^{**})'f$. Now $A^{**} \cong eC^{**}e$ and A is nuclear if and only if A^{**} is injective. Since a von Neumann algebra is injective if and only if its commutant is, it follows that A is nuclear if and only if $(C^{**})'$, and hence C^{**}, are injective. Since by the same reasoning B is nuclear if and only if C^{**} is injective, the result follows. □

We believe that the Morita equivalence technique used in the proof of the next proposition is essentially due to Alain Connes (unpublished), though we lack a precise attribution.

Proposition 6.3 *Let G be a locally compact group which has a closed amenable subgroup H such that G/H is compact. Then any closed discrete subgroup of G is exact.*

Proof: Assume first that H and K are just closed subgroups of G, and let $A = C_0(G/K)$. The continuous action of G on G/K by left multiplication gives rise to a continuous action α of G on A. Restricting α to

H, the algebras $C_0(G/H, A) \rtimes_{\Delta^\alpha, r} G \cong C_0((G/H) \times (G/K)) \rtimes_{\Delta, r} G$ and $A \rtimes_{\alpha,r} H \cong C_0(G/K) \rtimes_{\alpha,r} H$ are Morita equivalent, by Theorem 3.6, where Δ is the action of G on $C_0((G/H) \times (G/K))$ arising from the diagonal left action of G on $(G/H) \times (G/K)$. Interchanging H and K, it follows that $C_0(G/K) \rtimes_{\alpha,r} H$ and $C_0(G/H) \rtimes_{\beta,r} K$ are Morita equivalent, where β is the action of K coming from left multiplication on G/H.

Now assume that H is amenable, G/H is compact and K is discrete. Then $C_0(G/H) = C(G/H)$ and $C_r^*(K) \subseteq C(G/H) \rtimes_{\beta,r} K$ canonically. Also $C_0(G/K) \rtimes_{\alpha,r} H$, being a crossed product of a nuclear C*-algebra by an amenable group, is nuclear. This implies, by Proposition 6.2, that $C(G/H) \rtimes_{\beta,r} K$ is nuclear, so that $C_r^*(K)$ is exact, so that, by the equivalence of exactness for K and $C_r^*(K)$ [KW, Theorem 5.2], K is exact. □

Corollary 6.4 *Any closed discrete subgroup of a connected semisimple Lie group is exact.*

Proof: Let G be a connected semisimple Lie group. The centre Z of G is discrete, and, applying the Iwasawa decomposition [Kn], $G = KAN$, where K, A and N are connected closed subgroups of G, $Z \subseteq K$, K/Z is compact, N is nilpotent, and A is abelian. Moreover A normalises N and AN is a connected solvable Lie group. Since AN has a composition series with abelian quotients, it follows that AN, and hence ZAN, are amenable, and G/ZAN is homeomorphic to K/Z. The result now follows from Theorem 6.3. □

c. Closed linear groups. Since $SL_n(\mathbb{R})$ is semisimple [Kn] and contains $SL_n(\mathbb{Z})$ as a closed discrete subgroup, it is an immediate consequence of Corollary 6.4 that $SL_n(\mathbb{Z})$ is exact for $n = 1, 2, \ldots$. In fact the group $SL_n(\mathbb{Z})$ is a lattice in $SL_n(\mathbb{R})$, i.e. a closed discrete subgroup of finite covolume [Rag,Theorem 10.5], so that $SL_n(\mathbb{R})$ is exact by Theorem 4.5.

Proposition 6.5 *For $n \in \mathbb{N}$ any closed subgroup of $GL_n(\mathbb{R})$ is exact.*

Proof: The determinant gives a continuous homomorphisms of $GL_n(\mathbb{R})$ onto the multiplicative groups $\mathbb{R} \setminus \{0\}$ with kernel $SL_n(\mathbb{R})$. Thus $GL_n(\mathbb{R})$ is an extension of an exact group by an abelian group, hence is exact by Proposition 6.1 and Theorem 5.1. The result now follows by Theorem 4.5. □

d. Connected locally compact groups.

Proposition 6.6 *Any connected real semisimple Lie group is exact.*

Proof: Let G be a connected semisimple Lie group with centre Z and Lie algebra $\mathfrak{g}$. If Ad is the adjoint representation of G on $\mathfrak{g}$, we have a Lie group isomorphism $G/Z \cong Ad(G)$, and $Ad(G)$ coincides with $\mathrm{Aut}_0(\mathfrak{g})$, the connected component of the identity of the Lie group $\mathrm{Aut}(\mathfrak{g})$ [Kn]. Since the latter group is a closed subgroup of $GL(\mathfrak{g})$, and $\mathfrak{g}$ is finite-dimensional, it follows by Proposition 6.5 that G/Z is exact. Since Z is abelian, hence exact, the result now follows by Theorem 5.1. □

An alternative proof of this proposition follows by the technique of §4, since, by a deep theorem of Borel [Rag, Theorem 14.1], any connected noncompact semisimple Lie group contains a lattice.

Proposition 6.7 *Any connected real Lie group is exact.*

Proof: Let G be a connected Lie group with Lie algebra $\mathfrak{g}$. If rad $\mathfrak{g}$ is the radical of $\mathfrak{g}$, rad $\mathfrak{g}$ is a solvable ideal of $\mathfrak{g}$ and $\mathfrak{g}/\mathrm{rad}\,\mathfrak{g}$ is semisimple. If R is the closed normal subgroup of G with Lie algebra rad $\mathfrak{g}$, then R is solvable, hence exact, by earlier discussion, and G/R has Lie algebra $\mathfrak{g}/\mathrm{rad}\,\mathfrak{g}$. Thus G/R is semisimple, hence exact by Proposition 6.6, and the result follows using Theorem 5.2. □

Theorem 6.8 *Any connected locally compact group is exact.*

Proof: Let G be a connected locally compact group. By [MZ, Theorem 4.6] G has a closed normal compact subgroup K such that G/K is a real Lie group. Since G/K is connected, it is exact by Proposition 6.7. The result now follows by Theorem 5.2, since K, being amenable, is exact. □

Recall that a locally compact group G is *almost-connected* if the quotient group G/G_0 of G by the connected component G_0 of the identity is compact. The following corollary is an immediate consequence of Proposition 6.1, Theorem 6.8 and Theorem 5.1.

Corollary 6.9 *Any almost-connected group is exact.*

e. Exactness of certain discrete groups. By [KW, Theorem 5.2], a discrete group G is exact if and only if $C^*_r(G)$ is exact. For certain groups G, $C^*_r(G)$ can be explicitly embedded as a C*-subalgebra of a nuclear C*-algebra. For these groups, $C^*_r(G)$, being subnuclear, is exact, so that G is exact. Two classes for which $C^*_r(G)$ is known to be subnuclear are (a) the free groups and (b) the hyperbolic groups. The case of free groups was treated in [KW, Corollary 5.3], where it was shown, using a celebrated construction of Choi, that if G is a free group on at most countably many generators, then $C^*_r(G)$ can be embedded as a C*-subalgebra of the Cuntz algebra $\mathcal{O}_2$. Very recently Dykema [D] has shown that a reduced amalgamated free product of exact C*-algebras is exact. Since $C^*_r(\mathbb{Z})$ is abelian, hence nuclear, and $C^*_r(\mathbb{F}_\Lambda)$, where $\mathbb{F}_\Lambda$ is the free group on the set Λ, is the reduced free product of copies of $C^*_r(\mathbb{Z})$ indexed by Λ, Dykema's result together with our result just cited gives a new proof that free groups are exact. If G is a hyperbolic group, Adams [Ad] has shown that the natural action α of G on its Gromov boundary ∂G is amenable, which implies that the crossed product $C(\partial G) \rtimes_{\alpha,r} G$ is nuclear. Since G is discrete, $C^*_r(G)$ is a closed subalgebra of $C(\partial G) \rtimes_{\alpha,r} G$, hence exact. Germain [Ger] has recently given a concise and fairly simple proof of Adams' amenability result.

7. Concluding Remarks.

1. After we had completed most of this paper, Georges Skandalis pointed out that Corollary 6.9 can be obtained more directly by a different route. The various structure results used above together imply that if G is an almost connected group, then G contains a closed amenable subgroup H such that G/H with the quotient topology is compact. Corollary 6.9 is then an immediate consequence of Proposition 6.1 and the following theorem, which is closely related to Theorem 4.1, and has a similar, though simpler, proof.

Theorem *Let G be a locally compact group with a closed exact subgroup H. If G/H is compact, then G is exact.*

Proof: Let $(A, \alpha) \in \mathcal{C}^*_G$ and let I be an α_G-invariant ideal of A. If $\theta : A \to C_0(G/H) \otimes A \cong C_0(G/H, A)$ is the embedding $a \to 1 \otimes a$, let Φ_A denote the crossed product map

$$\Phi_A : A \rtimes_\alpha G \to C_0(G/H, A) \rtimes_{\Delta^\alpha,r} G.$$

Then Φ_A is an embedding, and if corresponding embeddings

$$\Phi_I : I \rtimes_{\alpha|,r} G \to C_0(G/H, I) \rtimes_{\Delta^\alpha,r} G$$

and

$$\Phi_{A/I} : (A/I) \rtimes_{\dot\alpha,r} G \to C_0(G/H, A/I) \rtimes_{\Delta^{\dot\alpha},r} G$$

are defined similarly, then the diagram

$$\begin{tikzcd}
0 \arrow[d] & 0 \arrow[d] \\
I \rtimes_{\alpha|,r} G \arrow[r, "\phi_I"] \arrow[d] & C_0(G/H, I) \rtimes_{\Delta^\alpha,r} G \arrow[d] \\
A \rtimes_{\alpha,r} G \arrow[r, "\Phi_A"] \arrow[d] & C_0(G/H, A) \rtimes_{\Delta^\alpha,r} G \arrow[d] \\
(A/I) \rtimes_{\dot\alpha,r} G \arrow[r, "\Phi_{A/I}"] \arrow[d] & C_0(G/H, A/I) \rtimes_{\Delta^{\dot\alpha},r} G \arrow[d] \\
0 & 0
\end{tikzcd}$$

commutes. Since H is exact, Corollary 3.10 implies, just as in the proof of Theorem 4.5, that the right-hand column is exact. If x is in the kernel of the quotient map $A \rtimes_{\alpha,r} G \to (A/I) \rtimes_{\dot{\alpha},r} G$, it follows that $\Phi_A(x)$ is in the kernel of the quotient map

$$C_0(G/H, A) \rtimes_{\Delta^\alpha,r} G \to C_0(G/H, A/I) \rtimes_{\Delta^{\dot{\alpha}},r} G,$$

which is $C_0(G/H, I) \rtimes_{\Delta^\alpha,r} G$. Thus

$$\Phi_A(x) \in (A \rtimes_{\alpha,r} G) \cap (C_0(G/H, I) \rtimes_{\Delta^\alpha,r} G).$$

Let $\{e_\mu\}$ be a bounded approximate identity for I. If we identify $1 \otimes e_\mu$ with its image in $M((C_0(G/H, I) \rtimes_{\Delta^\alpha,r} G)$ under the canonical embedding of $M(C_0(G/H, A))$ discussed in §4, it is readily checked that for each μ, $(1 \otimes e_\mu)y \in \Phi_A(I \rtimes_{\alpha|,r} G)$ for $y \in \Phi_A(A \rtimes_{\alpha,r} G)$, and $\lim_\mu (1 \otimes e_\mu)z = z$ for $z \in C_0(G/H, I) \rtimes_{\Delta^\alpha,r} G$. Then

$$\Phi_A(x) = \lim_\mu (1 \otimes e_\mu)x \in \Phi_A(I \rtimes_{\alpha|,r} G),$$

which shows that $x \in I \rtimes_{\alpha|,r} G$. Thus the left-hand column of the above diagram is exact, which implies that G is exact. □

2. If G is a locally compact group, the quotient G/G_0 by the connected component G_0 of the identity is a totally disconnected group. By Corollary 6.9, G_0 is exact. If G/G_0 is exact, it then follows, by Theorem 5.1, that G is exact. Thus to resolve the question of whether all locally compact groups are exact, it is enough to consider only totally disconnected groups. Our feeling is that if there if there are groups which are not exact, then there will probably be a discrete example.

3. In [KW, Lemma 2.5] we showed that a group which has an increasing family of exact open subgroups with union the whole group is itself exact. We have been unable to show that exactness is preserved under general inductive limits. Likewise, we do not know if exactness is preserved on passing to a quotient. If this were the case, then all discrete groups would be exact, since any discrete group is a quotient of a free group, which is exact, as noted in §5 (e).

Added note: The Referee has informed us that, in a recent preprint [Y], Guoliang Yu has studied a combinatorial property, *property A*, of discrete groups which is preserved under semi-direct products. It seems that property A is formally stronger than C*-exactness, but we only became aware of [Y] in September 1999, and have not yet studied all possible connections with our results.

References

[Ad] S. Adams, Boundary amenability for word hyperbolic groups and an application to smooth dynamics of simple groups. Topology **33** (1994), 765–783.

[BGR] L.G. Brown, P. Green, M.A. Rieffel, Stable isomorphism and strong Morita equivalence of C^*-algebras. *Pacific J. Math.* **71** (1977), 349–363

[Bo] N. Bourbaki, Élements de Mathématique, Livre VIII: Intǵration, Hermann, Paris, 1963

[Co] F. Combes, Crossed products and Morita equivalence. Proc. London Math. Soc. (3) 49 (1984), no. 2, 289–306.

[D] K.J. Dykema, Exactness of reduced amalgamated product C*-algebras, preprint.

[Ech] S. Echterhoff, Morita equivalent twisted actions and a new version of the Packer-Raeburn stabilization trick. J. London Math. Soc. (2) 50 (1994), no. 1, 170–186.

[Ger] E. Germain, Approximate invariant means for boundary actions of discrete hyperbolic groups, appendix to Amenable Groupoids by C. Anantharam-Delaroche, J. Renault (preprint)

[Gr] P. Green, The local structure of twisted covariance algebras. *Acta Math.* **140** (1978), 191–250

[Ki] E. Kirchberg, On subalgebras of the CAR-algebra. *J. Funct. Anal.* **129** (1995), 35–63.

[KW] E. Kirchberg, S. Wassermann, Exact groups and continuous bundles of C*-algebras, *Math. Ann.*, to appear.

[Kn] A. Knapp, Lie groups beyond an introduction. Progress in Mathematics, 140. Birkhäuser Boston, Inc., Boston, MA, 1996. xvi+604 pp. ISBN: 0-8176-3926-8

[MZ] D. Montgtomery, L. Zippin, Topological transformation groups. Interscience Publishers, New York-London, 1955. xi+282 pp.

[Ped] G.K. Pedersen, C*-algebras and their automorphism groups, Academic Press, London, 1979.

[Rae] I. Raeburn, Induced C^*-algebras and a symmetric imprimitivity theorem. *Math. Ann.* **280** (1988), 369–387

[Rag] M.S. Ragunatan, *Discrete subgroups of Lie groups.* Ergebnisse der Mathematik und ihrer Grenzgebiete, 68. Springer-Verlag, 1972.

[Rie1] M.A. Rieffel, Induced representations of C^*-algebras. *Advances in Math.* **13** (1974), 176–257

[Rie2] M.A. Rieffel, Unitary representations of group extensions; an algebraic approach to the theory of Mackey and Blattner. *Advances in Math. Supplementary Studies* **4** (1979), 43–82.

[Y] Guoliang Yu, The coarse Connes-Baum conjecture for spaces which admit a uniform embedding into Hilbert space. *Preprint, University of Colorado, 1998.*

Eberhard Kirchberg
Institut für Mathematik
Humboldt-Universität
Unter den Linden
6 D-10099 Berlin
Germany

Simon Wassermann
Department of Mathematics
University of Glasgow
Glasgow G12 8QW
United Kingdom

Metrics on State Spaces

This article is dedicated to Richard V. Kadison in anticipation of his completing his seventy-fifth circumnavigation of the sun.

Marc A. Rieffel[1]

Received: August 17, 1999

Communicated by Joachim Cuntz

Abstract. In contrast to the usual Lipschitz seminorms associated to ordinary metrics on compact spaces, we show by examples that Lipschitz seminorms on possibly non-commutative compact spaces are usually not determined by the restriction of the metric they define on the state space, to the extreme points of the state space. We characterize the Lipschitz norms which are determined by their metric on the whole state space as being those which are lower semicontinuous. We show that their domain of Lipschitz elements can be enlarged so as to form a dual Banach space, which generalizes the situation for ordinary Lipschitz seminorms. We give a characterization of the metrics on state spaces which come from Lipschitz seminorms. The natural (broader) setting for these results is provided by the "function spaces" of Kadison. A variety of methods for constructing Lipschitz seminorms is indicated.

1991 Mathematics Subject Classification: Primary 46L87; Secondary 58B30, 60B10

In non-commutative geometry (based on C^*-algebras), the natural way to specify a metric is by means of a suitable "Lipschitz seminorm". This idea was first suggested by Connes [C1] and developed further in [C2, C3]. Connes pointed out [C1, C2] that from a Lipschitz seminorm one obtains in a simple way an ordinary metric on the state space of the C^*-algebra. This metric generalizes

[1]The research reported here was supported in part by National Science Foundation Grant DMS–96–13833.

the Monge–Kantorovich metric on probability measures [KA, Ra, RR]. In this article we make more precise the relationship between metrics on the state space and Lipschitz seminorms.

Let ρ be an ordinary metric on a compact space X. The Lipschitz seminorm, L_ρ, determined by ρ is defined on functions f on X by

$$L_\rho(f) = \sup\{|f(x) - f(y)|/\rho(x, y) : x \neq y\}. \tag{0.1}$$

(This can take value $+\infty$.) It is known that one can recover ρ from L_ρ by the relationship

$$\rho(x, y) = \sup\{|f(x) - f(y)| : L_\rho(f) \leq 1\}.$$

But a slight extension of this relationship defines a metric, $\bar{\rho}$, on the space $S(X)$ of probability measures on X, by

$$\bar{\rho}(\mu, \nu) = \sup\{|\mu(f) - \nu(f)| : L_\rho(f) \leq 1\}. \tag{0.2}$$

This is the Monge–Kantorovich metric. The topology which it defines on $S(X)$ coincides with the weak-$*$ topology on $S(X)$ coming from viewing it as the state space of the C^*-algebra $C(X)$. The extreme points of $S(X)$ are identified with the points of X. On the extreme points, $\bar{\rho}$ coincides with ρ. Thus the relationship (0.1) can be viewed as saying that L_ρ can be recovered just from the restriction of its metric $\bar{\rho}$ on $S(X)$ to the set of extreme points of $S(X)$.

Suppose now that $\mathcal{A}$ is a unital C^*-algebra with state space $S(\mathcal{A})$, and let L be a Lipschitz seminorm on $\mathcal{A}$. (Precise definitions are given in Section 2.) Following Connes [C1, C2], we define a metric, ρ, on $S(\mathcal{A})$ by the evident analogue of (0.2). We show by simple finite dimensional examples determined by Dirac operators that L may well not be determined by the restriction of ρ to the extreme points of $S(\mathcal{A})$.

It is then natural to ask whether L is determined by ρ on all of $S(\mathcal{A})$, by a formula analogous to (0.1). One of our main theorems (Theorem 4.1) states that the Lipschitz seminorms for which this is true are exactly those which are lower semicontinuous in a suitable sense.

For ordinary compact metric spaces (X, ρ) it is known that the space of Lipschitz functions with a norm coming from the Lipschitz seminorm is the dual of a certain other Banach space. Another of our main theorems (Theorem 5.2) states that the same is true in our non-commutative setting, and we give a natural description of this predual. We also characterize the metrics on $S(\mathcal{A})$ which come from Lipschitz seminorms (Theorem 9.11).

We should make precise that we ultimately require that our Lipschitz seminorms be such that the metric on $S(\mathcal{A})$ which they determine gives the weak-$*$ topology on $S(\mathcal{A})$. An elementary characterization of exactly when this happens was given in [Rf]. (See also [P].) This property obviously holds for finite-dimensional C^*-algebras. It is known to hold in many situations for commutative C^*-algebras, as well as for C^*-algebras obtained by combining commutative ones with finite dimensional ones. But this property has not been verified for many examples beyond those. However in [Rf] this property was verified for some interesting infinite-dimensional non-commutative examples, such as the non-commutative tori, and I expect that eventually it will be found to hold in a wide variety of situations.

Actually, we will see below that the natural setting for our study is the broader one of order-unit spaces. The theory of these spaces was launched by Kadison in his memoir [K1]. For this reason it is especially appropriate to dedicate this article to him. (In [K2] Kadison uses the terminology "function systems", but we will follow [Al] in using the terminology "order-unit space" as being a bit more descriptive of these objects.)

On the other hand, most of the interesting constructions currently in view of Lipschitz seminorms on non-commutative C^*-algebras, such as those from Dirac operators, or those in [Rf], also provide in a natural way seminorms on all the matrix algebras over the algebras. Thus it is likely that "matrix Lipschitz seminorms" in analogy with the matrix norms of [Ef] will eventually be of importance. But I have not yet seen how to use them in a significant way, and so we do not deal with them here.

Let us mention here that a variety of metrics on the state spaces of full matrix algebras have been employed by the practitioners of quantum mechanics. A recent representative paper where many references can be found is [ZS]. We will later make a few comments relating some of these metrics to the considerations of the present paper.

The last three sections of this paper will be devoted to a discussion of the great variety of ways in which Lipschitz seminorms can arise, even for commutative algebras. We do discuss here some non-commutative examples, but most of our examples are commutative. I hope in a later paper to discuss and apply some other important classes of non-commutative examples. Some of the applications which I have in mind will require extending the theory developed here to quotients and sub-objects.

Finally, we should remark that while we give here considerable attention to how Dirac operators give metrics on state spaces, Connes has shown [C2] that Dirac operators encode far more than just the metric information. In particular they give extensive homological information. But we do not discuss this aspect here.

I thank Nik Weaver for suggestions for improvement of the first version of this article, which are acknowledged more specifically below.

1. Recollections on order-unit spaces

We recall [Al] that an order-unit space is a real partially-ordered vector space, $\mathcal{A}$, with a distinguished element e, the order unit, which satisfies:

1) (Order unit property) For each $a \in \mathcal{A}$ there is an $r \in \mathbb{R}$ such that $a \leq re$.
2) (Archimedean property) If $a \in \mathcal{A}$ and if $a \leq re$ for all $r \in \mathbb{R}^+$, then $a \leq 0$.

For any $a \in \mathcal{A}$ we set

$$\|a\| = \inf\{r \in \mathbb{R}^+ : -re \leq a \leq re\}.$$

We obtain in this way a norm on $\mathcal{A}$. In turn, the order can be recovered from the norm, because $0 \leq a \leq e$ iff $\|a\| \leq 1$ and $\|e - a\| \leq 1$. The primary source of examples consists of the linear spaces of all self-adjoint elements in unital C^*-algebras, with the identity element serving as order unit. But any linear space of bounded self-adjoint operators on a Hilbert space will be an order-unit space if it contains the identity operator. We expect that this broader class of examples will be important for the applications of metrics on state spaces.

We will not assume that $\mathcal{A}$ is complete for its norm. This is important for us because the domains of Lipschitz norms will be dense, but usually not closed, in the completion. (The completion is always again an order-unit space.) This also accords with the definition in [Al].

By a state of an order-unit space $(\mathcal{A}, e)$ we mean a continuous linear functional, μ, on $\mathcal{A}$ such that $\mu(e) = 1 = \|\mu\|$. States are automatically positive. We denote the collection of all the states of $\mathcal{A}$, i.e. the state space of $\mathcal{A}$, by $S(\mathcal{A})$. It is a w^*-compact convex subset of the Banach space dual, $\mathcal{A}'$, of $\mathcal{A}$.

To each $a \in \mathcal{A}$ we get a function, $\hat{a}$, on $S(\mathcal{A})$ defined by $\hat{a}(\mu) = \mu(a)$. Then $\hat{a}$ is an affine function on $S(\mathcal{A})$ which is continuous by the definition of the w^*-topology. The basic representation theorem of Kadison [K1, K2, K3] (see Theorem II.1.8 of [Al]) says that for any order-unit space the representation $a \to \hat{a}$ is an isometric order isomorphism of $\mathcal{A}$ onto a dense subspace of the space $Af(S(\mathcal{A}))$ of all continuous affine functions on $S(\mathcal{A})$, equipped with the supremem norm and the usual order on functions (and with e clearly carried to the constant function 1). In particular, if $\mathcal{A}$ is complete, then it is isomorphic to all of $Af(S(\mathcal{A}))$.

Thus we can view the order-unit spaces as exactly the dense subspaces containing 1 inside $Af(K)$, where K is any compact convex subset of a topological vector space. This provides an effective view from which to see many of the properties of order-unit spaces. Most of our theoretical discussion will be carried out in the setting of order-unit spaces and $Af(K)$, though our examples will usually involve specific C^*-algebras. We let $C(K)$ denote the real C^*-algebra of all continuous functions on K, in which $Af(K)$ sits as a closed subspace.

It will be important for us to work on the quotient vector space $\tilde{\mathcal{A}} = \mathcal{A}/(\mathbb{R}e)$. We let $\| \quad \|^\sim$ denote the quotient norm on $\tilde{\mathcal{A}}$ from $\| \quad \|$. This quotient norm is easily described. For $a \in \mathcal{A}$ set

$$\max(a) = \inf\{r : a \leq re\}$$
$$\min(a) = \sup\{r : re \leq a\},$$

so that $\|a\| = (\max(a)) \vee (-\min(a))$. Then it is easily seen that

$$\|\tilde{a}\|^\sim = (\max(a) - \min(a))/2.$$

2. The radius of the state space

Let $\mathcal{A}$ be an order-unit space. Since the term "Lipschitz seminorm" has somewhat wide but imprecise usage, we will not use this term for our main objects of precise study (which we will define in Section 5). Almost the minimal requirement for a Lipschitz seminorm is that its null-space be exactly the scalar multiples of the order unit. We will use the term "Lipschitz seminorm" in this general sense. We emphasize that a Lipschitz seminorm will usually not be continuous for $\| \quad \|$.

Let L be a Lipschitz seminorm on $\mathcal{A}$. For $\mu, \nu \in S(\mathcal{A})$ we can define a metric, ρ_L, on $S(\mathcal{A})$ by

$$\rho_L(\mu, \nu) = \sup\{|\mu(a) - \nu(a)| : L(a) \le 1\}$$

(which may be $+\infty$). Then ρ_L determines a topology on $S(\mathcal{A})$. Eventually we want to require that this topology agrees with the weak-$*$ topology. Since $S(\mathcal{A})$ is weak-$*$ compact, ρ_L must then give $S(\mathcal{A})$ finite diameter. We examine this latter aspect here, in part to establish further notation.

It is actually more convenient for us to work with "radius" (half the diameter), since this will avoid factors of 2 in various places. We would like to use the properties of order-unit spaces to express the radius in terms of L in a somewhat more precise way than was implicit in [Rf] in its more general context. The following considerations [Al] will also be used extensively later.

As in [Rf] and in the previous section, we denote the quotient vector space $\mathcal{A}/(\mathbb{R}e)$ by $\tilde{\mathcal{A}}$, with its quotient norm $\| \quad \|^{\sim}$. But in addition to this norm, the quotient seminorm $\tilde{L}$ from L is also a norm on $\tilde{\mathcal{A}}$, since L takes value 0 only on $\mathbb{R}e$.

The dual Banach space to $\tilde{\mathcal{A}}$ for $\| \quad \|^{\sim}$ is just ${\mathcal{A}'}^0$, the subspace of $\mathcal{A}'$ consisting of those $\lambda \in \mathcal{A}'$ such that $\lambda(e) = 0$. We denote the norm on $\mathcal{A}'$ dual to $\| \quad \|$ still by $\| \quad \|$. The dual norm on ${\mathcal{A}'}^0$ is just the restriction of $\| \quad \|$ to ${\mathcal{A}'}^0$. If we view $\mathcal{A}$ as a dense subspace of $Af(K) \subseteq C(K)$, then by the Hahn–Banach theorem λ extends (not uniquely) to $C(K)$ with same norm. There we can take the Jordan decomposition into disjoint non-negative measures. Note that for positive measures their norm on $C(K)$ equals their norm on $\mathcal{A}$, since $e \in \mathcal{A}$. Thus we find $\mu, \nu \ge 0$ such that $\lambda = \mu - \nu$ and $\|\lambda\| = \|\mu\| + \|\nu\|$. But $0 = \lambda(e) = \mu(e) - \nu(e) = \|\mu\| - \|\nu\|$. Consequently $\|\mu\| = \|\nu\| = \|\lambda\|/2$. Thus if $\|\lambda\| \le 2$ we have $\|\mu\| = \|\nu\| \le 1$. If $\|\lambda\| < 2$ set $t = \|\mu\| < 1$, and rescale μ and ν so that they are in $S(\mathcal{A})$. Then

$$\lambda = t\mu - t\nu = \mu - (t\nu + (1-t)\mu).$$

Now $(t\nu + (1-t)\mu)$ is no longer disjoint from μ, but we have obtained the following lemma, which will be used in a number of places.

2.1 Lemma. *The ball D_2 of radius 2 about 0 in ${\mathcal{A}'}^0$ coincides with $\{\mu - \nu : \mu, \nu \in S(\mathcal{A})\}$.*

Notice that if there is an $a \in \mathcal{A}$ such that $L(a) = 0$ but $a \notin \mathbb{R}e$, then from this lemma we can find $\mu, \nu \in S(\mathcal{A})$ such that $(\mu - \nu)(a) \neq 0$, so that $\rho_L(\mu, \nu) = +\infty$.

Thus our standing assumption that there is no such a serves to reduce the possibility of having infinite distances. But it does not eliminate this possibility, as seen by the example of the algebra of smooth (or Lipschitz) functions of compact support on the real line, with constant functions adjoined, and with the usual Lipschitz seminorm.

2.2 Proposition. *With notation as earlier, the following conditions are equivalent for an $r \in \mathbb{R}^+$:*
1) *For all $\mu, \nu \in S(\mathcal{A})$ we have $\rho_L(\mu, \nu) \leq 2r$.*
2) *For all $a \in \mathcal{A}$ we have $\|\tilde{a}\|^\sim \leq rL^\sim(\tilde{a})$.*

Proof. Suppose that condition 1 holds. Let $a \in \mathcal{A}$ and $\lambda \in D_2$. Then by the lemma $\lambda = \mu - \nu$ for some $\mu, \nu \in S(\mathcal{A})$. Thus

$$|\lambda(a)| = |(\mu - \nu)(a)| \leq L(a)\rho_L(\mu, \nu) \leq L(a)2r.$$

Since $\lambda(e) = 0$, thus inequality holds whenever a is replaced by $a + se$ for $s \in \mathbb{R}$. Thus condition 2 holds.
Conversely, suppose that condition 2 holds. Then for any $\mu, \nu \in S(\mathcal{A})$ and $a \in \mathcal{A}$ with $L(a) \leq 1$ we have

$$|\mu(a) - \nu(a)| = |(\mu - \nu)(a)| \leq 2\|\tilde{a}\|^\sim \leq 2r.$$

Thus $\rho_L(\mu, \nu) \leq 2r$ as desired. $\square$

Of course, we call the smallest r for which the conditions of this proposition hold the *radius* of $S(\mathcal{A})$.
We caution that just because a metric space has radius r, it does not follow that there is a ball of radius r which contains it, as can be seen by considering equilateral triangles in the plane. We remark that just because ρ_L gives $S(\mathcal{A})$ finite radius, it does not follow that ρ_L gives the weak-$*$ topology. Perhaps the simplest example arises when $\mathcal{A}$ is infinite dimensional and $L(a) = \|\tilde{a}\|^\sim$.

3. Lower semicontinuity for Lipschitz seminorms

Let L be any Lipschitz seminorm on an order-unit space $\mathcal{A}$. (We will not at first require that it give $S(\mathcal{A})$ finite diameter.) We would like to show that L and ρ_L contain the same information, and more specifically that we can recover L from ρ_L as being the usual Lipschitz seminorm for ρ_L. By this we mean the following. Let ρ be any metric on $S(\mathcal{A})$, possibly taking value $+\infty$. Define L_ρ on $C(S(\mathcal{A}))$ by

$$L_\rho(f) = \sup\{|f(\mu) - f(\nu)|/\rho(\mu, \nu) : \mu \neq \nu\}, \tag{3.1}$$

where this may take value $+\infty$. Let $\mathrm{Lip}_\rho = \{f : L_\rho(f) < \infty\}$. We can restrict L_ρ to $Af(S(\mathcal{A}))$. In general, few elements of $Af(S(\mathcal{A}))$ will be in Lip_ρ. However, on viewing the elements of $\mathcal{A}$ as elements of $Af(S(\mathcal{A}))$, we have:

3.2 Lemma. *Let L be a Lipschitz seminorm on $\mathcal{A}$ with corresponding metric ρ_L on $S(\mathcal{A})$. Then $\mathcal{A} \subseteq \mathrm{Lip}_{\rho_L}$, and on $\mathcal{A}$ we have $L_{\rho_L} \leq L$, in the sense that $L_{\rho_L}(a) \leq L(a)$ for all $a \in \mathcal{A}$.*

Proof. For $\mu, \nu \in S(\mathcal{A})$ and $a \in \mathcal{A}$ we have

$$|\hat{a}(\mu) - \hat{a}(\nu)| = |\mu(a) - \nu(a)| \leq L(a)\rho_L(\mu, \nu).$$

□

For later use we remark that if L and M are Lipschitz seminorms on $\mathcal{A}$ and if $M \leq L$, then $\rho_M \geq \rho_L$ in the evident sense.

We would like to recover L on $\mathcal{A}$ from ρ_L by means of formula (3.1). However, the seminorms defined by (3.1) have an important continuity property:

3.3 Definition. Let $\mathcal{A}$ be a normed vector space, and let L be a seminorm on $\mathcal{A}$, except that we permit it to take value $+\infty$. Then L is *lower semicontinuous* if for any sequence $\{a_n\}$ in $\mathcal{A}$ which converges in norm to $a \in \mathcal{A}$ we have $L(a) \leq \liminf\{L(a_n)\}$. Equivalently, for one, hence every, $t \in \mathbb{R}$ with $t > 0$, the set

$$\mathcal{L}_t = \{a \in \mathcal{A} : L(a) \leq t\}$$

is norm-closed in $\mathcal{A}$.

3.4 Proposition. *Let $\mathcal{A}$ be an order-unit space, and let ρ be any metric on $S(\mathcal{A})$, possibly taking value $+\infty$. Define L_ρ on $C(S(\mathcal{A}))$ by formula (3.1). Then L_ρ is lower semicontinuous. Consequently, the restriction of L_ρ to any subspace of $C(S(\mathcal{A}))$, such as $\mathcal{A}$ or $Af(S(\mathcal{A}))$, will be lower semicontinuous.*

Proof. When we view L_ρ as a function of f, the formula (3.1) says that L_ρ is the pointwise supremum of a collection of functions (labeled by pairs μ, ν with $\mu \neq \nu$) which are clearly continuous on $C(S(\mathcal{A}))$ for the supremum norm. But the pointwise supremum of continuous functions is lower semicontinuous. □

3.5 Example. Here is an example of a Lipschitz seminorm L whose metric can be seen to give $S(\mathcal{A})$ the weak-$*$ topology, but which is not lower semicontinuous. Let $I = [-1, 1]$, and let $\mathcal{A} = C^1(I)$, the algebra of functions which have continuous first derivatives on I. Define L on $\mathcal{A}$ by

$$L(f) = \|f'\|_\infty + |f'(0)|.$$

For each n let g_n be the function defined by $g_n(t) = n|t|$ for $|t| \leq 1/n$, and $g_n(t) = 1$ elsewhere. Let $f_n(t) = \int_{-1}^t g_n(s)ds$. Then the sequence $\{f_n\}$ converges uniformly to the function f given by $f(t) = t + 1$. But $L(f_n) = 1$ for each n, whereas $L(f) = 2$.

A substantial supply of examples of lower semicontinuous seminorms can be obtained from the W^*-derivations of Weaver [W2, W3]. These derivations will in general have large null spaces, and the seminorms from them need not give the weak-$*$ topology on the state space. But many of the specific examples of W^*-derivations which Weaver considers do in fact give the weak-$*$ topology. In terms of Weaver's terminology, which we do not review here, we have:

3.6 PROPOSITION. *Let M be a von Neumann algebra and let E be a normal dual operator M-bimodule. Let $\delta : M \to E$ be a W^*-derivation, and denote the domain of δ by $\mathcal{L}$, so that $\mathcal{L}$ is an ultra-weakly dense unital $*$-subalgebra of M. Define a seminorm, L, on $\mathcal{L}$ by $L(a) = \|\delta(a)\|_E$. Then L is lower semicontinuous, and $\mathcal{L}_1 = \{a \in \mathcal{L} : L(a) \leq 1\}$ is norm-closed in M itself.*

Proof. Let $\{a_n\}$ be a sequence in $\mathcal{L}$ which converges in norm to $b \in M$. To show that L is lower semicontinuous, it suffices to consider the case in which $\{a_n\}$ is contained in $\mathcal{L}_1$. Then the set $\{(a_n, \delta(a_n))\}$ is a bounded subset of the graph of δ for the norm $\max\{\| \quad \|_M, \| \quad \|_E\}$. Since the graph of a W^*-derivation is required to be ultra-weakly closed, and since bounded ultraweakly closed subsets are compact for the ultra-weak topology, there is a subnet which converges ultra-weakly to an element $(c, \delta(c))$ of the graph of δ. Then necessarily $c = b$, so that $b \in \mathcal{L}$, and $\delta(b)$ is in the ultra-weak closure of $\{\delta(a_n)\}$. Consequently $L(b) = \|\delta(b)\| \leq 1$. $\square$

Because of the importance of Dirac operators, it is appropriate to verify lower semicontinuity for the Lipschitz seminorms which they determine. This is close to a special case of Proposition 3.6, but does not require any kind of completeness, nor an algebra structure on $\mathcal{A}$.

3.7 PROPOSITION. *Let $\mathcal{A}$ be a linear subspace of bounded self-adjoint operators on a Hilbert space $\mathcal{H}$, containing the identity operator. Let D be an essentially self-adjoint operator on $\mathcal{H}$ whose domain, $\mathcal{D}(D)$, is carried into itself by each element of $\mathcal{A}$. Assume that $[D, a]$ is a bounded operator on $\mathcal{D}(D)$ for each $a \in \mathcal{A}$ (so that $[D, a]$ extends uniquely to a bounded operator on $\mathcal{H}$). Define L on $\mathcal{A}$ by $L(a) = \|[D, a]\|$. Then L is lower semicontinuous.*

Proof. Let $\{a_n\}$ be a sequence in $\mathcal{A}$ which converges in norm to $a \in \mathcal{A}$. Suppose that there is a constant, k, such that $L(a_n) \leq k$ for all n. For any $\xi, \eta \in \mathcal{D}(D)$ with $\|\xi\| = 1 = \|\eta\|$ we have

$$\langle [D, a]\xi, \eta\rangle = \langle a\xi, D\eta\rangle - \langle D\xi, a\eta\rangle = \lim\langle [D, a_n]\xi, \eta\rangle.$$

But $|\langle [D, a_n]\xi, \eta\rangle| \leq k$ for each n, and so $\|[D, a]\| \leq k$. $\square$

We remark that the Lipschitz seminorms constructed in [Rf] by means of actions of compact groups are easily seen to be lower semicontinuous.

4. RECOVERING L FROM ρ_L

In this section we show that a lower semicontinuous Lipschitz seminorm L can be recovered from its metric ρ_L. But before showing this we would like to emphasize the following point. Let (X, ρ) be an ordinary compact metric space, with $\mathcal{A}$ the algebra of its Lipschitz functions, with Lipschitz seminorm L. Then $S(\mathcal{A})$ consists of the probability measures on X, and the points of X correspond exactly to the extreme points of $S(\mathcal{A})$. The restriction of ρ_L to the extreme points is exactly ρ. Thus when one says that one can recover L from

the metric ρ, one is saying that one can recover L from the restriction of ρ_L on $S(\mathcal{A})$ to the extreme points of $S(\mathcal{A})$. However, for the more general situation which we are considering, it will be false in general that we can recover L from the restriction of ρ_L to the extreme points of $S(\mathcal{A})$. Simple explicit examples will be given in Section 7.
One of the main theorems of this paper is:

4.1 Theorem. *Let L be a lower semicontinuous Lipschitz seminorm on an order-unit space $\mathcal{A}$, and let ρ_L denote the corresponding metric on $S(\mathcal{A})$, possibly taking value $+\infty$. Let L_{ρ_L} be defined by formula* (3.1), *but restricted to $\mathcal{A} \subseteq Af(S(\mathcal{A}))$. Then*

$$L_{\rho_L} = L.$$

Theorem 4.1 is an immediate consequence of the following theorem, since we saw that lower semicontinuity coincides with $\mathcal{L}_1$ being norm closed.

4.2 Theorem. *Let L be any Lipschitz seminorm on an order-unit space $\mathcal{A}$, and let ρ_L denote the corresponding metric on $S(\mathcal{A})$. Let L_{ρ_L} be defined by formula* (3.1), *but restricted to $\mathcal{A} \subseteq Af(S(\mathcal{A}))$. Then $\{a \in \mathcal{A} : L_{\rho_L}(a) \le 1\}$ coincides with the norm closure, $\bar{\mathcal{L}}_1$, of $\mathcal{L}_1$ in $\mathcal{A}$. In particular, L_{ρ_L} is the largest lower semicontinuous seminorm smaller than L, and $\rho_{L_{\rho_L}} = \rho_L$.*

Proof. (An idea leading to this proof, which is simpler than my original proof, was suggested to me by Nik Weaver.) On $\mathcal{A}'$ we define the seminorm, L' , dual to L, by

$$L'(\lambda) = \sup\{|\lambda(a)| : L(a) \le 1\}.$$

Note that L' takes value $+\infty$ on any λ for which $\lambda(e) \ne 0$, and very possibly on some elements of ${\mathcal{A}'}^0$ as well. But at any rate we have the following key relationship:

4.3 Lemma. *For $\mu, \nu \in S(\mathcal{A})$ we have $\rho_L(\mu, \nu) = L'(\mu - \nu)$.*

Proof.

$$\begin{aligned} L'(\mu - \nu) &= \sup\{|(\mu - \nu)(a)| : L(a) \le 1\} \\ &= \sup\{|\mu(a) - \nu(a)| : L(a) \le 1\} = \rho_L(\mu, \nu). \end{aligned}$$

□

Because $\mathcal{L}_1$ is already convex and balanced, the bipolar theorem [Cw] says that $\bar{\mathcal{L}}_1$ is exactly the bipolar of $\mathcal{L}_1$. Thus we just need to show that $\{a \in \mathcal{A} : L_{\rho_L}(a) \le 1\}$ is the bipolar of $\mathcal{L}_1$. Now it is clear that the unit L'-ball in $\mathcal{A}'$ is exactly the polar [Cw] of $\mathcal{L}_1$. This provides the last of the following equivalences. Let $a \in \mathcal{A}$. Then:
$L_{\rho_L}(a) \le 1$ exactly if $|\mu(a) - \nu(a)| \le \rho_L(\mu, \nu)$ for all $\mu, \nu \in S(\mathcal{A})$,
exactly if $|\lambda(a)| \le L'(\lambda)$ for all $\lambda \in D_2$ (by Lemma 4.3 and Lemma 2.1),
exactly if $|\lambda(a)| \le 1$ for all $\lambda \in \mathcal{A}'$ with $L'(\lambda) \le 1$,

exactly if a is in the prepolar of $\{\lambda : L'(\lambda) \le 1\}$ (by definition [Cw]),
exactly if a is in the bipolar of $\mathcal{L}_1$.
It is clear that L_{ρ_L} is lower semicontinuous, that it is the largest such seminorm smaller than L, and that it gives the same metric. ☐

Note in particular that if L gives $S(\mathcal{A})$ finite diameter, or the weak-$*$ topology, then so does L_{ρ_L}.
We remark that a sort of dual version of Theorem 4.1 can be found later in Theorem 9.7.
We have the following related considerations. Suppose again that L is a Lipschitz seminorm on an order-unit space $\mathcal{A}$. Let $\bar{\mathcal{A}}$ denote the completion of $\mathcal{A}$ for $\| \ \|$, and let $\bar{\mathcal{L}}_1$ denote now the closure of $\mathcal{L}_1$ in $\bar{\mathcal{A}}$ rather than just in $\mathcal{A}$. Let $\bar{L}$ denote the corresponding "Minkowski functional" on $\bar{\mathcal{A}}$ obtained by setting, for $b \in \bar{\mathcal{A}}$,

$$\bar{L}(b) = \inf\{r \in \mathbb{R}^+ : b \in r\bar{\mathcal{L}}_1\}.$$

Since there may be no such r, we must allow the value $+\infty$. With this understanding, $\bar{L}$ will be a seminorm on $\bar{\mathcal{A}}$. It is easily seen that $\bar{L}(b) \le 1$ exactly if $b \in \bar{\mathcal{L}}_1$, and that $\bar{L}$ is lower semicontinuous because $\bar{\mathcal{L}}_1$ is closed.
Up to this point we did not require lower semicontinuity of L. It's import is given by:

4.4 Proposition. *Let L be a lower semicontinuous Lipschitz seminorm on an order-unit space $\mathcal{A}$. Let $\bar{L}$ on $\bar{\mathcal{A}}$ be defined as above. Then $\bar{L}$ is an extension of L, that is, for $a \in \mathcal{A}$ we have $\bar{L}(a) = L(a)$. Furthermore, $\rho_{\bar{L}} = \rho_L$.*

Proof. Suppose that $a \in \mathcal{A}$ and $L(a) = 1$. Then $a \in \mathcal{L}_1 \subseteq \bar{\mathcal{L}}_1$ and so clearly $\bar{L}(a) \le 1$. Conversely, if $\bar{L}(a) \le 1$, then $a \in \bar{\mathcal{L}}_1$. Thus there is a sequence $\{a_n\}$ in $\mathcal{L}_1$ which converges to a, with $L(a_n) \le 1$ for every n. From the lower semicontinuity of L it follows that $L(a) \le 1$. Finally, for $\mu, \nu \in S(\mathcal{A})$ we have

$$\rho_{\bar{L}}(\mu,\nu) = \sup\{|\mu(a) - \nu(a)| : a \in \bar{\mathcal{L}}_1\} = \sup\{|\mu(a) - \nu(a)| : a \in \mathcal{L}_1\} = \rho_L(\mu,\nu).$$

☐

Note in particular that if L gives $S(\mathcal{A})$ finite diameter, or the weak-$*$ topology, then so does $\bar{L}$. However, in general $\bar{L}$ need not be a Lipschitz seminorm. For example, let $\mathcal{A}$ be the algebra of real polynomials viewed as functions on the interval $[0, 2]$, and let L be the usual Lipschitz seminorm but defined using only points in $[0, 1]$.

4.5 Definition. We will call $\bar{L}$ the *closure* of L. We will say that a Lipschitz seminorm is *closed* if $L = \bar{L}$ (on the subspace where $\bar{L}$ is finite), or equivalently, if $\mathcal{L}_1$ is complete for the metric from $\| \ \|$.

Then Proposition 4.4 says that for most purposes we can assume that L is closed if convenient.

Suppose now that L is a Lipschitz seminorm on $\mathcal{A}$ which is closed. On $\mathcal{A}$ we can define a new norm, $|||\ \ |||$, by

$$|||a||| = \|a\| + L(a).$$

It is easily verified that $\mathcal{A}$ is complete for this new norm. Suppose that $\mathcal{A}$ is a $*$-algebra and $\|\ \ \|$ is a C^*-norm (this can be weakened). Suppose further that L is a closed Lipschitz seminorm on $\mathcal{A}$ which satisfies the Leibniz inequality. Then the new norm is a normed-algebra norm, and so $\mathcal{A}$ becomes a Banach algebra for the new norm. In Sections 10 and 11 we will indicate many examples of Lipschitz seminorms satisfying the Leibniz inequality. This provides a rich class of examples of Banach algebras which merit study (even in the cases when they are commutative) along the lines considered in [BCD, J, W1].

5. The pre-dual of $(\tilde{\mathcal{A}}, \tilde{L})$

It has been shown in an increasing variety of situations that the space of Lipschitz functions with a suitable Lipschitz norm is isometrically isomorphic to the dual of some Banach space. Some of the history of this phenomenon is sketched in the notes at the end of chapter 2 of [W1], or more briefly in [W2]. Within the non-commutative setting, Weaver shows in Proposition 2 of [W2] that the domains of W^*-derivations (as defined there) are dual spaces. However, his W^*-derivations can have large null spaces, and they need not give the weak-$*$ topology on $S(\mathcal{A})$. Nevertheless, Weaver's approach applies to the non-commutative tori, and gives them the same space of Lipschitz elements as the approach of the present paper (when combined with [Rf]). In fact, Weaver shows in [W3] that for the non-commutative tori one can also define Lip^α, and that Lip^α is actually the *second* dual of lip^α when $\alpha < 1$.

To show within our setting that the space of Lipschitz elements is the dual of a Banach space, we need to assume that ρ_L gives the weak-$*$ topology on $S(\mathcal{A})$. As before, let $\mathcal{L}_1 = \{a : L(a) \leq 1\}$. From theorem 1.8 of [Rf] we know that ρ_L will give the weak-$*$ topology on $S(\mathcal{A})$ exactly if the image of $\mathcal{L}_1$ in $\tilde{\mathcal{A}}$ is totally bounded for $\|\ \ \|^\sim$. Equivalently, by theorem 1.9 of [Rf], L must give $S(\mathcal{A})$ finite radius, and for one, hence all, $t \in \mathbb{R}$ with $t > 0$, the set

$$\mathcal{B}_t = \{a : L(a) \leq 1 \text{ and } \|a\| \leq t\}$$

must be totally bounded in $\mathcal{A}$ for $\|\ \ \|$. We remark that this implies that if $\{a_n\}$ is a sequence (or net) in $\mathcal{A}$ converging pointwise on $S(\mathcal{A})$ to $a \in \mathcal{A}$, and if there is a constant k such that $\|a_n\| \leq k$ and $L(a_n) \leq k$ for all n, then a_n converges to a in norm. This is because $\{a_n\}$ is contained in $k\mathcal{B}_1$ whose closure in the completion $\bar{\mathcal{A}}$ of $\mathcal{A}$ is compact. Let b be any norm limit point of $\{a_n\}$ in $\bar{\mathcal{A}}$. Then a subsequence of $\{a_n\}$ converges in norm to b. But it still converges pointwise on $S(\mathcal{A})$ to a. Consequently $b = a$, and a is the only norm limit point of $\{a_n\}$.

We now have in view all the requirements on Lipschitz seminorms which we need for our present purposes. So we now define what we expect is the correct way to specify metrics on compact non-commutative spaces:

5.1 Definition. Let $\mathcal{A}$ be an order-unit space. By a *Lip-norm* on $\mathcal{A}$ we mean a seminorm, L, on $\mathcal{A}$ (taking finite values) with the following properties:

1) For $a \in \mathcal{A}$ we have $L(a) = 0$ if and only if $a \in \mathbb{R}e$.
2) L is lower semicontinuous.
3) $\{a \in \mathcal{A} : L(a) \leq 1\}$ has image in $\tilde{\mathcal{A}}$ which is totally bounded for $\| \quad \|^{\sim}$.

We remark that it is easily checked that the closure (Definition 4.5) of a Lip-norm is again a Lip-norm.
Within the present setting the fact that the space of Lipschitz elements is a dual Banach space takes the following form (which requires the Lip-norm to be closed).

5.2 Theorem. *Let $\mathcal{A}$ be an order-unit space, and let L be a Lip-norm on $\mathcal{A}$ which is closed. Let $\mathcal{K} = \{\tilde{a} \in \tilde{\mathcal{A}} : \tilde{L}(\tilde{a}) \leq 1\}$, so that $\mathcal{K}$ is a compact (convex) set for $\| \quad \|^{\sim}$. Then $(\tilde{\mathcal{A}}, \tilde{L})$ is naturally isometrically isomorphic to the dual Banach space of $Af_0(\mathcal{K})$, the Banach space of continuous affine functions on $\mathcal{K}$ which take value 0 at $0 \in \tilde{\mathcal{A}}$, with the supremum norm.*

Proof. Let $\mathcal{L}_1$ and $\mathcal{B}_t$ be as defined as above. Because L is closed, the totally bounded sets $\mathcal{B}_t$ are complete for $\| \quad \|$, and so are compact. From the finite radius considerations of Section 2 the image of $\mathcal{L}_1$ in $\tilde{\mathcal{A}}$ will coincide with the image of $\mathcal{B}_t$ for sufficiently large t. Hence the image of $\mathcal{L}_1$ in $\tilde{\mathcal{A}}$ is compact for $\| \quad \|^{\sim}$, not just totally bounded. But the image of $\mathcal{L}_1$ is exactly $\mathcal{K}$ as defined in the statement of the theorem.
We can now argue as in the proof of proposition 1 of [W4]. We include the argument here in a form specific to our particular situation.
Let $V = Af_0(\mathcal{K})$, as defined in the statement of the theorem. Then from lemma 4.1 of [K3] each element of V extends to a linear functional (not necessarily continuous for $\| \quad \|^{\sim}$) on $\tilde{\mathcal{A}}$. But we still view V as equipped with the uniform norm $\| \quad \|_\infty$ from $C(\mathcal{K})$, for which V is complete. Then for any $f \in V$ we have

$$\|f\|_\infty = \sup\{f(\tilde{a}) : \tilde{a} \in \mathcal{K}\} = \sup\{f(\tilde{a}) : \tilde{L}(\tilde{a}) \leq 1\}.$$

Consequently $\| \quad \|_\infty$ is just the dual norm to the norm $\tilde{L}$ on $\tilde{\mathcal{A}}$. But V will usually be much smaller than the entire dual Banach space of $(\tilde{\mathcal{A}}, \tilde{L})$ because of the requirement that if $f \in V$ then f is continuous on $\mathcal{K}$.
We let V' denote the dual Banach space to V. We have the evident mapping σ from $\tilde{\mathcal{A}}$ to V' defined by $\sigma(\tilde{a})(f) = f(\tilde{a})$. Use of the Hahn–Banach theorem shows that $Af_0(\mathcal{K})$ separates the points of $\mathcal{K}$, and from this we see that σ is injective. Furthermore $|\sigma(\tilde{a})(f)| = |f(\tilde{a})| \leq \|f\|_\infty \tilde{L}(\tilde{a})$, and so $\|\sigma\| \leq 1$ for the norm $\tilde{L}$ on $\tilde{\mathcal{A}}$. In particular, $\sigma(\mathcal{K}) \subseteq (V')_1$, the unit ball of V'. From the definitions of σ and V we see immediately that σ is continuous from $\mathcal{K}$ to $(V')_1$ with its weak-$*$ topology from V. Since $\mathcal{K}$ is compact, $\sigma(\mathcal{K})$ must be compact for the weak-$*$ topology. If $\sigma(\mathcal{K})$ were not all of $(V')_1$, there would be a $\varphi_0 \in (V')_1$ and a weak-$*$ continuous linear functional separating φ_0 from

$\sigma(\mathcal{K})$. But every weak-$*$ linear functional comes from V. Thus there would be an $f \in V$ such that

$$f(\tilde{a}) \leq 1 < \varphi_0(f)$$

for every $\tilde{a} \in \mathcal{K}$. But the first inequality means that $\|f\|_\infty \leq 1$, and so the second inequality means that $\|\varphi_0\| > 1$, contradicting the assumption that $\varphi_0 \in (V')_1$. Thus $\sigma(\mathcal{K}) = (V')_1$. Consequently σ is an isometric isomorphism of $(\tilde{\mathcal{A}}, \tilde{L})$ with V'. $\square$

We remark that, if desired, we can make $\mathcal{A}$ itself into the dual of a Banach space, in a non-canonical way, as follows. Let r be the radius of $(\mathcal{A}, L)$, and let μ be any fixed state of $\mathcal{A}$. Define an actual norm, L_μ, on $\mathcal{A}$ by

$$L_\mu(a) = \max\{|\mu(a)|/r, L(a)\}.$$

Let $\tilde{L}_\mu$ be the quotient of L_μ on $\tilde{\mathcal{A}}$. It is clear that $\tilde{L}_\mu \geq \tilde{L}$. But for any given $a \in \mathcal{A}$ we can find $\alpha \in \mathbb{R}$ such that $\|a - \alpha\| \leq r\tilde{L}(\tilde{a})$, by the definition of radius. Then

$$|\mu(a - \alpha)| \leq \|a - \alpha\| \leq r\tilde{L}(\tilde{a}),$$

while $L(a - \alpha) = \tilde{L}(\tilde{a})$. Consequently $\tilde{L}_\mu(\tilde{a}) \leq \tilde{L}(\tilde{a})$, so that, in fact, $\tilde{L}_\mu = \tilde{L}$. Thus $(\mathcal{A}, L_\mu)$ has $(\tilde{\mathcal{A}}, \tilde{L})$ as quotient space. The quotient map splits by the isometric map $\tilde{a} \mapsto a - \mu(a)$. Since $(\tilde{\mathcal{A}}, \tilde{L})$ is isometrically isomorphic to a dual Banach space, it follows easily that $(\mathcal{A}, L_\mu)$ is also.

See also section 2 of [H], which gives a slightly different approach because the norm on Lip_ρ is slightly different from that implicit here.

Let $\mathcal{K}$ and $V = Af_0(\mathcal{K})$ be as in the statement of Theorem 5.2. As in Section 2, the dual of $(\tilde{\mathcal{A}}, \|\ \ \|^\sim)$ is $\mathcal{A}'^0$. By the finite diameter condition and Proposition 2.2 each $\lambda \in \mathcal{A}'^0$ defines a continuous linear functional on $(\tilde{\mathcal{A}}, \tilde{L})$. Each such functional is clearly continuous on $\mathcal{K}$ for its topology from $\|\ \ \|^\sim$. Thus each $\lambda \in \mathcal{A}'^0$ defines an element of V, and so we obtain a linear map from $\mathcal{A}'^0$ into V. From Theorem 5.2 the norm $\|\ \ \|_\infty$ on V from $C(\mathcal{K})$ coincides with the dual norm L' from $(\tilde{\mathcal{A}}, \tilde{L})$. We have the following addition to Theorem 5.2.

5.3 Proposition. *The image of $\mathcal{A}'^0$ in $Af_0(\mathcal{K})$ is dense in $Af_0(\mathcal{K})$ for its norm $\|\ \ \|_\infty = L'$.*

Proof. Let φ be any continuous linear functional on V which is 0 on the image of $\mathcal{A}'^0$. From Theorem 5.2 every continuous linear functional on V comes from an element of $\tilde{\mathcal{A}}$. If $\tilde{a}$ is the element of $\tilde{\mathcal{A}}$ corresponding to φ, we then have $\lambda(\tilde{a}) = 0$ for all $\lambda \in \mathcal{A}'^0$, which implies that $\tilde{a} = 0$ so that $\varphi \equiv 0$. It follows from the Hahn–Banach theorem that the image of $\mathcal{A}'^0$ is norm dense in V. $\square$

6. Extreme points

Let L be a Lipschitz seminorm on an order-unit space $\mathcal{A}$, and let ρ_L be the corresponding metric on $S(\mathcal{A})$. Let E denote the set of extreme points of $S(\mathcal{A})$.

Then E need not be a closed subset of $S(\mathcal{A})$, but $S(\mathcal{A})$ is the closed convex hull of E by the Krein–Milman theorem. Of course ρ_L restricts to a metric on E. We will give explicit examples in the next section to show that even when L is a Lip-norm the restriction of ρ_L to E does not determine ρ_L or L. Nevertheless, we can try to use the restriction of ρ_L to define a new Lipschitz seminorm, L^e, on $\mathcal{A}$, by

$$L^e(a) = \sup\{|\varepsilon(a) - \eta(a)|/\rho_L(\varepsilon, \eta) : \varepsilon, \eta \in E,\ \varepsilon \neq \eta\}.$$

6.1 Proposition. *With the above definition, L^e is a lower semicontinuous Lipschitz seminorm on $\mathcal{A}$, and it is the smallest such on $\mathcal{A}$ whose metric on $S(\mathcal{A})$ agrees on E with that of L. If L is a* Lip-*norm then so is L^e.*

Proof. From Theorem 4.2 it is clear that we can assume that L is lower semicontinuous. From Theorem 4.1 we know that any lower semicontinuous Lipschitz seminorm, say L_1, is recovered from its metric by a supremum as above, but ranging over all of $S(\mathcal{A})$ rather than just over E. Thus if the metric for L_1 agrees with ρ_L on E, we must have $L^e \leq L_1$. By using the argument in the proof of Proposition 3.4 it is easily seen that L^e is lower semicontinuous. Suppose that $L^e(a) = 0$ for some $a \in \mathcal{A}$. Recall that $D_2 = \{\lambda \in {\mathcal{A}'}^0 : \|\lambda\| \leq 2\}$.

6.2 Lemma. *The convex hull of $\{\varepsilon - \eta : \varepsilon, \eta \in E,\ \varepsilon \neq \eta\}$ is dense in D_2 for the weak-$*$ topology.*

Proof. From Lemma 2.1 we know that any element of D_2 can be expressed as $\mu - \nu$ for $\mu, \nu \in S(\mathcal{A})$. By the Krein–Milman theorem each of μ, ν can be approximated arbitrary closely in the weak-$*$ topology by convex combinations from E, say $\sum \alpha_j \varepsilon_j$ and $\sum \beta_k \eta_k$. But the difference of such combinations can be expressed as

$$\sum (\alpha_j \beta_k)(\varepsilon_j - \eta_k).$$

□

From this lemma it is clear that if $L^e(a) = 0$ then $L(a) = 0$, and thus $a \in \mathbb{R}e$. Also, it is easy to see that ρ_{L^e} agrees with ρ_L on E.

Finally, we must show that if L is a Lip-norm then the image of $\mathcal{K}_0 = \{a : L^e(a) \leq 1\}$ in $\tilde{\mathcal{A}}$ is totally bounded for $\|\ \ \|^\sim$. Notice that this image is larger than that for L, so we can not immediately apply the corresponding fact for L. Let $\bar{E}$ denote the closure of E in $S(\mathcal{A})$. It is clear that the supremum defining L^e could just as well be taken over $\bar{E}$, and so L^e on $\mathcal{A}$ is just the Lipschitz norm for the metric ρ_L restricted to $\bar{E}$. Thus $\mathcal{K}_0$ can be viewed as contained in $\{f \in C(\bar{E}) : L^e(f) \leq 1\}$, and the latter has totally bounded image in $C(\bar{E})/\mathbb{R}e$ since it consists of Lipschitz functions for a metric and $\bar{E}$ is compact. Thus $\mathcal{K}_0$ has totally bounded image in $C(\bar{E})/\mathbb{R}e$. But the restriction map from $Af(S(\mathcal{A}))$ to $C(\bar{E})$ is isometric for $\|\ \ \|_\infty$ since $\bar{E}$ contains the extreme points. (See Theorem II.1.8 of [Al]. We are dealing here with Kadison's smallest separating representation.) It follows easily that $\mathcal{K}_0$ has totally bounded image in $\tilde{\mathcal{A}}$ as needed. □

We remark that if F is any subset of $S(\mathcal{A})$ which contains E, then we can use F instead of E to define a Lip-norm L^F just as we defined L^e above. Then we will have

$$L^e \leq L^F \leq L$$

in the evident sense, with reverse inequalities for the corresponding metrics.

Suppose that $\mathcal{A}$ is a dense $*$-subalgebra of a C^*-algebra, $\bar{\mathcal{A}}$, and that L is a Lip-norm on $\mathcal{A}$, with corresponding metric ρ_L on $S(\mathcal{A})$. As above let E denote the set of extreme points of $S(\mathcal{A})$. Assume first that $\mathcal{A}$ is commutative. Then E is compact and $\bar{\mathcal{A}} \cong C(E)$. Assume that $L = L^e$. Then L is the usual Lipschitz norm coming from the metric on the compact set E obtained by restricting ρ_L to E. But in this case we know that L must then satisfy the Leibniz rule

$$L(ab) \leq L(a)\|b\| + \|a\|L(b).$$

It is thus natural to ask the general question:

6.3 Question. *What conditions on a* Lip*-norm L on a general unital C^*-algebra imply that L satisfies the Leibniz rule?*

In the next section we will see examples of Lip-norms which do not satisfy $L = L^e$ and yet satisfy the Leibniz rule.

7. Dirac operators and ordinary finite spaces

Connes has shown [C1, C2, C3] that for a compact Riemannian (spin) manifold all the metric information is contained in the Dirac operator. This led him to suggest that for "non-commutative spaces", metrics should be specified by some analogue of Dirac operators. We explore here some aspects of this suggestion for finite-dimensional commutative C^*-algebras, i.e. ordinary finite spaces. This will clarify some of the considerations of the previous sections. Here and throughout all the rest of this paper, when we say that an operator D is a " Dirac" operator, this is not meant to indicate any particular properties of D, but rather is meant to indicate how D is employed, namely to define a Lipschitz seminorm.

Let X be a finite set, and let $\mathcal{A} = C(X)$. In order to remain fully in the setting of the previous sections we take $C(X)$ to consist only of real-valued functions. But in the present commutative situation this is not so important because, unlike the non-commutative case, if one does not know the algebra structure, the norm for complex-valued functions is still given by a simple formula in terms of the norm for real-valued functions. (See e.g. lemma 14 of [W2].) Consequently we will be a bit careless here about this distinction.

We will suppose that $\mathcal{A}$ has been faithfully represented on a finite-dimensional complex Hilbert space $\mathcal{H}$. We suppose given on $\mathcal{H}$ an operator D (the "Dirac" operator). It is usual to take D to be self-adjoint. But we find it slightly more convenient to take D to be skew-adjoint. The two choices are related by a multiplication by i, and give the same metric results. Following Connes, we define a seminorm, L, on $\mathcal{A}$ by

$$L(a) = \|[D, a]\|,$$

where [,] denotes the usual commutator of operators, and the norm is the operator norm. We want L to be a Lip-norm. Thus we require that if $[D, a] = 0$ then $a \in \mathbb{C}I$. Because we are in a finite-dimensional setting, L is continuous for $\| \quad \|_\infty$, and indeed is a Lip-norm on $\mathcal{A}$.
From L we obtain a metric, ρ_L, on the space $S(\mathcal{A})$ of probability measures on X, as well as on its set of extreme points, which is identified with X itself. We now give a very simple example to show that ρ_L on $S(\mathcal{A})$ need not agree with the metric obtained from ρ_L on X.

7.1 Example. Consider a three-dimensional commutative C^*-algebra, $\mathcal{A}$, represented faithfully on a three-dimensional Hilbert space. Thus we can identify $\mathcal{A}$ with the algebra of diagonal matrices in the full matrix algebra $M_3 = M_3(\mathbb{C})$. We will consider Dirac operators of a special form which facilitates calculation, namely matrices D in $M_3(\mathbb{C})$ of the form

$$D = \begin{pmatrix} 0 & 0 & \alpha \\ 0 & 0 & \beta \\ -\alpha & -\beta & 0 \end{pmatrix}$$

where $\alpha > 0$ and $\beta > 0$. We will also restrict to those $f \in \mathcal{A}$ which are real, and denote the three values (or diagonal entries) of f by (f_1, f_2, f_3). Because D is skew-symmetric, $[D, f]$ is a real symmetric matrix, whose eigenvalues thus are real. In fact, we have

$$[D, f] = \begin{pmatrix} 0 & 0 & \alpha(f_3 - f_1) \\ 0 & 0 & \beta(f_3 - f_2) \\ \alpha(f_3 - f_1) & \beta(f_3 - f_2) & 0 \end{pmatrix}.$$

Because of this special form, the eigenvalues are easily calculated, and one finds that

$$L(f) = \|[D, f]\| = (\alpha^2(f_3 - f_1)^2 + \beta^2(f_3 - f_2)^2)^{1/2}.$$

It is clear from this that if $L(f) = 0$ then f is a constant function. Thus L defines a Lip-norm on $\mathcal{A}$.
We now proceed to calculate the corresponding metric on $S(\mathcal{A})$. We first calculate the dual norm, L', on ${\mathcal{A}'}^0$, the dual space of $\tilde{\mathcal{A}}$, with notation as in the previous sections. We identify ${\mathcal{A}'}^0$ with real diagonal matrices of trace 0, paired with $\mathcal{A}$ via the trace. For $\lambda \in {\mathcal{A}'}^0$ we denote its components by $\lambda = (\lambda_1, \lambda_2, \lambda_3)$. Of course

$$L'(\lambda) = \sup\{|\langle f, \lambda \rangle| : L(f) \le 1\}.$$

Now both $|\langle f, \lambda \rangle|$ and $L(f)$ are unchanged if we add a constant function to f. Thus for the supremum defining $L'(\lambda)$ we can assume that $f_3 = 0$ always.

Furthermore, we know that $\lambda_3 = -(\lambda_1 + \lambda_2)$. Thus we need only deal with the first two components of f and λ. We do this without changing notation. Then we see that

$$L'(\lambda) = \sup\{|f_1\lambda_1 + f_2\lambda_2| : \alpha^2 f_1^2 + \beta^2 f_2^2 \leq 1\}.$$

But this is just the norm of a functional on a suitable Hilbert space. Specifically, let $l^2(w)$ be the Hilbert space of functions on a 2-point space with weight function w given by (α^2, β^2). Then

$$f_1\lambda_1 + f_2\lambda_2 = f_1(\lambda_1/\alpha^2)\alpha^2 + f_2(\lambda_2/\beta^2)\beta^2,$$

and in this form the norm of the functional is the length of the vector in $l^2(w)$ defining it. This gives

$$\begin{aligned} L'(\lambda) &= ((\lambda_1/\alpha^2)^2\alpha^2 + (\lambda_2/\beta^2)^2\beta^2)^{1/2} \\ &= (\lambda_1^2/\alpha^2 + \lambda_2^2/\beta^2)^{1/2}. \end{aligned}$$

We now apply this to obtain the metric on $S(\mathcal{A})$. If $\mu, \nu \in S(\mathcal{A})$, then for the evident notation

$$\rho_L(\mu, \nu) = L'(\mu - \nu) = ((\mu_1 - \nu_1)^2/\alpha^2 + (\mu_2 - \nu_2)^2/\beta^2)^{1/2}.$$

Let X denote the maximal ideal space of $\mathcal{A}$. We identify its 3 points with the 3 extreme points of $S(\mathcal{A})$, and label them, corresponding to the coordinates in $\mathcal{A}$, by $\delta_1, \delta_2, \delta_3$. Then from the above formula for ρ_L we find that the metric on X is given by:

$$\begin{aligned} &\rho_L(\delta_1, \delta_2) = (1/\alpha^2 + 1/\beta^2)^{1/2} \\ &\rho_L(\delta_1, \delta_3) = 1/\alpha \\ &\rho_L(\delta_2, \delta_3) = 1/\beta. \end{aligned}$$

Define γ by $\rho_L(\delta_1, \delta_2) = 1/\gamma$. Let L^e denote the ordinary Lipschitz norm on $\mathcal{A}$ coming from this metric on X. Then

$$L^e(f) = \max\{|f_1 - f_2|\gamma, |f_1 - f_3|\alpha, |f_2 - f_3|\beta\}.$$

Clearly L^e is quite different from L. From Theorem 4.1 we know that the metrics on $S(\mathcal{A})$ will thus be quite different, even though they agree on the extreme points. This is, of course, also easily seen by direct calculations.

We now make some observations in preparation for the next section. It is well-known [W1, W2] that the Lipschitz seminorms $L = L_\rho$ from ordinary metrics on a metric space X have a nice relation to the lattice structure of (real-valued) $C(X)$, namely

$$L(f \vee g) \leq L(f) \vee L(g).$$

We remark that for the L of the above example this inequality fails. For instance, with notation as above, let $f = (1,0,0)$ and $g = (0,1,0)$, so that $f \vee g = (1,1,0)$. Then we see that

$$L(f) = \alpha,\ L(g) = \beta,\ \text{while } L(f \vee g) = (\alpha^2 + \beta^2)^{1/2}.$$

(This is related to the counterexample following theorem 16 of [W2].)
However, it is not difficult to check that the above L does satisfy the weaker inequality

$$L(f \vee 0) \le L(f).$$

In fact, one can prove that this holds for any choice of skew-adjoint D for the above $\mathcal{A}$. To find a counterexample for this weaker inequality one must take $\mathcal{A}$ to be 4-dimensional. I have not found a systematic way of constructing a counterexample there, but some examination of what is needed, followed by some experimentation with MATLAB yields the following (and related) example:

$$D = \begin{pmatrix} 0 & 4 & -1 & 0 \\ -4 & 0 & 2 & -2 \\ 1 & -2 & 0 & -4 \\ 0 & 2 & 4 & 0 \end{pmatrix}$$

and $f = (4,2,0,-1)$.
We remark that ordinary Lipschitz norms on compact metric spaces can all be easily obtained by means of Dirac operators. I pointed this out in a lecture in 1993, and the details are indicated after the proof of proposition 8 of [W2]. See also the discussion for graphs which we will give toward the end of Section 11.

8. A characterization of ordinary Lipschitz seminorms

Let X be a compact space, let ρ be a metric on X (giving the topology of X), and let L denote the corresponding ordinary Lip-norm on $C(X)$ (permitted to take value $+\infty$). As just mentioned in the last section, it is well-known [W1, W2] and easy to prove that L relates nicely to the lattice structure of $C(X)$ by means of the inequality

$$L(f \vee g) \le L(f) \vee L(g).$$

In Weaver's more general setting of domains of W^*-derivations he proves this inequality for W^*-derivations of Abelian structure. (See lemma 12 of [W2].)
We show here that the above inequality exactly characterizes the Lip-norms which are the ordinary Lipschitz seminorms coming from ordinary metrics on X.
We remark that we never assume here that our Lip-norms satisfy the Leibniz inequality for the algebra structure, namely

$$L(fg) \le L(f)\|g\| + \|f\|L(g).$$

But ordinary Lipschitz seminorms do satisfy this inequality. Thus one consequence of this section is that the above lattice inequality implies the Leibniz inequality. On the other hand, the Lip-norm from any "Dirac" operator will satisfy the Leibniz inequality, but can easily fail to satisfy the lattice inequality, as we saw by examples in the previous section. Thus the lattice inequality is much stronger than the Leibniz inequality.
However we should point out that for Dirac operators on compact spin Riemannian manifolds, in spite of their being defined by means of various partial derivatives and spinors, the corresponding Lip-norms do satisfy the lattice inequality. This is because Connes shows [C1, C2, C3] that the Lip-norms which those Dirac operators define coincide with the ordinary Lip-norms for the ordinary metrics on the manifolds determined by the Riemannian metrics.
Recall that for us $C(X)$ consists of real-valued functions.

8.1 Theorem. *Let X be a compact space, let $\mathcal{A}$ be a dense subspace of $C(X)$ containing the constant functions, and let L be a Lip-norm on $\mathcal{A}$. Let $\bar{L}$ denote the closure of L, viewed as defined on all of $C(X)$ as in the discussion before Proposition* 4.4*, and thus permitted to take value $+\infty$. Then the following conditions are equivalent:*

1. *The Lip-norm L is the restriction to $\mathcal{A}$ of the usual Lipschitz seminorm corresponding to a metric on X (namely the metric ρ_L).*
2. *For every $f, g \in C(X)$ we have*

$$\bar{L}(f \vee g) \leq \bar{L}(f) \vee \bar{L}(g).$$

The following lemma is somewhat parallel to lemma 13 of [W2]. For later use we state it in slightly greater generality than needed immediately.

8.2 Lemma. *Let $\mathcal{A}$ be a dense subspace of $C(X)$ containing the constant functions, and closed under the finite lattice operations (i.e. if $f, g \in \mathcal{A}$ then $f \vee g \in \mathcal{A}$). Let L be a* Lip*-norm on $\mathcal{A}$ which satisfies the inequality*

$$L(f \vee g) \leq L(f) \vee L(g)$$

for all $f, g \in \mathcal{A}$. Let $\bar{L}$ be the closure of L, defined on all of $C(X)$, permitted to take value $+\infty$. Let $\mathcal{F}$ be a bounded subset of $\mathcal{A}$ for which there is a constant, k, such that $L(f) \leq k$ for all $f \in \mathcal{F}$. Let $g = \sup\{f \in \mathcal{F}\}$. Then $g \in C(X)$ and $\bar{L}(g) \leq k$.

Proof. Let $\{g_\alpha\}$ be the net of suprema of finite subsets of $\mathcal{F}$. Then $\{g_\alpha\}$ is contained in $\mathcal{A}$, and converges up to f pointwise. By the hypothesis on L we have $L(g_\alpha) \leq k$ for all α. Thus we have

$$|g_\alpha(x) - g_\alpha(y)| \leq k\rho_L(x, y)$$

for all α and all $x, y \in X$; that is, $\{g_\alpha\}$ is equicontinuous. We can thus apply the Ascoli theorem [Ru] to conclude that the net $\{g_\alpha\}$ has a subnet which converges uniformly. But the limit of this subnet must be g, and so g must be continuous. Furthermore, from the lower semicontinuity of $\bar{L}$ we must have $\bar{L}(g) \le k$. $\square$

Proof of Theorem 8.1. As indicated above, it is basically well-known, and not hard to verify, that condition 1 implies condition 2. Suppose conversely that condition 2 holds. For any $x \in X$ let ρ_L^x be the continuous function on X defined by $\rho_L^x(y) = \rho_L(x, y)$. Set $S_x = \{f \in \mathcal{A} : f(x) = 0,\ L(f) \le 1\}$. Since $L(f)$ is unchanged when a constant function is added to f, or when f is replaced by $-f$, the definition of ρ_L can be rewritten as

$$\rho_L^x(y) = \sup\{f(y) : f \in S_x\}.$$

This means that $\rho_L^x = \sup S_x$. But S_x is a bounded set in $\mathcal{A}$ by the finite radius considerations. Thus we can apply the above lemma to conclude that $\bar{L}(\rho_L^x) \le 1$. Suppose that $\bar{L}(\rho_L^x) = c < 1$. Then $\bar{L}((1/c)\rho_L^x) = 1$, and so from the definition of ρ_L we obtain

$$(1/c)|\rho_L^x(x) - \rho_L^x(y)| \le \rho_L(x, y),$$

for all $y \in X$, that is,

$$\rho_L(x, y) \le c\rho_L(x, y)$$

for all $y \in X$, which is impossible (unless X has only one point, which we now do not permit). Thus $\bar{L}(\rho_L^x) = 1$.

Much as in Section 6, let L^e denote the ordinary Lip-norm on $C(X)$ (permitting value $+\infty$) corresponding to the restriction of ρ_L as metric on X. (Recall that X is identified with the extreme points of $S(\mathcal{A})$.) As seen in Proposition 6.1, $L^e \le \bar{L}$. We now show that $L^e = \bar{L}$ because of the inequality in the hypotheses of our theorem (and its extension in Lemma 8.2). Let $f \in C(X)$, and suppose that $L^e(f) \le 1$. Thus

$$|f(x) - f(y)| \le \rho_L(x, y)$$

for all $x, y \in X$. In particular

$$f(x) - \rho_L(x, y) \le f(y).$$

For each $x \in X$ define $h^x \in C(X)$ by

$$h^x(y) = f(x) - \rho_L(x, y).$$

Then the above inequality says that $h^x \le f$ for each x. But it is clear that $h^x(x) = f(x)$. Thus $f = \sup\{h^x : x \in X\}$. Then from the considerations of the previous paragraph we see that $\bar{L}(h^x) = 1$ for all x. Thus by Lemma 8.2 we have $\bar{L}(f) \le 1$. It follows that $\bar{L} = L^e$ as desired. $\square$

8.3 Corollary. *Let X be a compact space, and let $\mathcal{A}$ be a dense subspace of $C(X)$ which contains the constant functions and is closed under the finite lattice operations. Let L be a* Lip-*norm on $\mathcal{A}$, and suppose that*

$$L(f \vee g) \leq L(f) \vee L(g)$$

for all $f, g \in \mathcal{A}$. Then L is the restriction to $\mathcal{A}$ of the ordinary Lip-*norm on $C(X)$ corresponding to the metric ρ_L on X.*

Proof. Let $f, g \in C(X)$. Then from Lemma 8.2 we see immediately that

$$\bar{L}(f \vee g) \leq \bar{L}(f) \vee \bar{L}(g).$$

We can thus apply Theorem 8.1 to obtain the desired conclusion. □

One way of viewing Theorem 8.1 is that it characterizes the Lip-norms on commutative C^*-algebras which come from the corresponding metric on the extreme points of $S(\mathcal{A})$. It would be interesting to have a corresponding characterization for non-commutative C^*-algebras, and for general order-unit spaces.

9. Lip-norms from metrics on $S(\mathcal{A})$

It is natural to ask which metrics on $S(\mathcal{A})$ arise from Lip-norms on $\mathcal{A}$. We obtain here a characterization of such metrics. Many of the steps work for arbitrary convex sets, and so at first we will work in that setting. Thus we let V be any vector space over $\mathbb{R}$, and we let K be any convex set in V which spans V. Much as above, let $D_2 = K - K$. Note that not only is D_2 convex, but it is also balanced, in the sense that if $\lambda \in D_2$ and if $t \in [-1, 1]$, then $t\lambda \in D_2$. To see this, note that if $\lambda \in D_2$ then clearly $-\lambda \in D_2$, so we only need consider $t \geq 0$. But

$$t(\mu - \nu) = \mu - (t\nu + (1-t)\mu),$$

which is in D_2 by the convexity of K. Let $V^0 = \mathbb{R}D_2$. Then V^0 is a vector subspace of V. In the setting where $K = S(\mathcal{A})$ we know that V^0 is a proper subspace of V. Let M be a norm on V^0. Then we can define a metric, ρ, on K by $\rho(\mu, \nu) = M(\mu - \nu)$. We want to characterize the metrics which arise in this way.

The most natural property to expect is that ρ be convex (in each variable), that is:

9.1 Definition. We say that a metric ρ on K is *convex* if for every $\mu, \nu_1, \nu_2 \in K$ and $t \in [0, 1]$ we have

$$\rho(\mu, t\nu_1 + (1-t)\nu_2) \leq t\rho(\mu, \nu_1) + (1-t)\rho(\mu, \nu_2).$$

The metrics coming from norms on V^0 are convex because

$$\mu - (t\nu_1 + (1-t)\nu_2) = t(\mu - \nu_1) + (1-t)(\mu - \nu_2).$$

Given a metric ρ on K, our strategy will be to try to use ρ to define a norm, M, on V^0 by first defining it on D_2. Specifically, for $\lambda \in D_2$ we would like to set

$$M(\lambda) = \rho(\mu, \nu)$$

for $\lambda = \mu - \nu$ with $\mu, \nu \in K$. But we need to know that this is well-defined. That is, we need to know that if $\mu, \nu, \mu', \nu' \in K$ and if $\mu - \nu = \mu' - \nu'$, then $\rho(\mu, \nu) = \rho(\mu', \nu')$. This can be rewritten in terms of midpoints so as to appear a bit closer to considerations of convexity, namely, that if

$$(\mu + \nu')/2 = (\mu' + \nu)/2 \tag{9.2}$$

then $\rho(\mu, \nu) = \rho(\mu', \nu')$. This clearly holds for the metrics coming from norms. One finds an attractive geometrical interpretation when one draws a picture of this relation.

9.3 Definition. We say that a metric ρ on K is *midpoint-balanced* if whenever equation (9.2) above holds, it follows that $\rho(\mu, \nu) = \rho(\mu', \nu')$.

Let us now assume that ρ is midpoint-balanced. Then M on D_2 is well-defined as above. We wish to extend it to a norm on V^0. For this to be possible we first must have the property that if $t \in \mathbb{R}$, $|t| \leq 1$, and if $\lambda \in D_2$, then $M(t\lambda) = |t|M(\lambda)$. Now from the definition of M it is clear that $M(-\lambda) = M(\lambda)$. Thus it suffices to treat the case in which $t \geq 0$. If $\lambda = \mu - \nu$, then

$$t\lambda = t(\mu - \nu) = \mu - (t\nu + (1-t)\mu),$$

so that by the definition of M we have $M(t\lambda) = \rho(\mu, t\nu + (1-t)\mu)$. From convexity, $\rho(\mu, t\nu + (1-t)\mu) \leq t\rho(\mu, \nu)$. But also $t\lambda = (t\mu + (1-t)\nu) - \nu$, which gives a similar inequality. Then from the triangle inequality and convexity we have

$$\begin{aligned} \rho(\mu, \nu) &\leq \rho(\mu, t\nu + (1-t)\mu) + \rho(t\nu + (1-t)\mu, \nu) \\ &\leq t\rho(\mu, \nu) + (1-t)\rho(\mu, \nu) = \rho(\mu, \nu). \end{aligned}$$

Thus the inequalities must be equalities, and we obtain:

9.4 Lemma. *Let ρ be a metric on K which is convex and midpoint balanced. Define M on D_2 as above using ρ. Then for any $\mu, \nu \in S(\mathcal{A})$ and $t \in [0, 1]$ we have*

$$\rho(\mu, t\nu + (1-t)\mu) = t\rho(\mu, \nu),$$

and for any $\lambda \in D_2$ and $t \in [-1, 1]$ we have

$$M(t\lambda) = |t|M(\lambda).$$

Next, we need that M is subadditive on D_2. This means that if $\lambda, \lambda' \in D_2$ and if $\lambda + \lambda' \in D_2$, then $M(\lambda + \lambda') \leq M(\lambda) + M(\lambda)$. Let $\lambda = \mu - \nu$, $\lambda' = \mu' - \nu'$. Then

$\lambda+\lambda' = (\mu+\mu')-(\nu+\nu')$. Assuming that ρ is convex and midpoint-balanced, we obtain from Lemma 9.4 that

$$M(\lambda+\lambda') = 2M((\lambda+\lambda')/2).$$

Now $(\lambda+\lambda')/2 = (\mu+\mu')/2-(\nu+\nu')/2$, and $(\mu+\mu')/2, (\nu+\nu')/2 \in S(\mathcal{A})$. Thus

$$M((\lambda+\lambda')/2) = \rho((\mu+\mu')/2, (\nu+\nu')/2),$$

and we see that what we need is:

9.5 Definition. We say that a metric ρ on K is *midpoint concave* if for any $\mu, \nu, \mu', \nu' \in K$ we have

$$\rho((\mu+\mu')/2, (\nu+\nu')/2) \le (1/2)(\rho(\mu,\nu)+\rho(\mu',\nu')).$$

Again one finds an attractive geometrical interpretation when one draws a picture of this inequality. From the discussion above we now know that:

9.6 Lemma. *Let ρ be a metric on K which is convex, midpoint balanced, and midpoint concave. Define M on K as above. If $\lambda, \lambda' \in D_2$ and if $\lambda+\lambda' \in D_2$, then*

$$M(\lambda+\lambda') \le M(\lambda)+M(\lambda').$$

9.7 Theorem. *Let ρ be a metric on the convex subset K of V, and let $V^0 = \mathbb{R}D_2 = \mathbb{R}(K-K)$. Then there is a norm, M, on V^0 such that $\rho(\mu,\nu) = M(\mu-\nu)$ for all $\mu,\nu \in K$, if and only if ρ is convex, midpoint balanced, and midpoint concave. The norm M is unique.*

Proof. The uniqueness is clear since $V^0 = \mathbb{R}(K-K)$. We have seen above that the conditions on ρ are necessary. We now show that they are sufficient. We let M be defined on $D_2 = K-K$ as above. For any $\lambda \in V^0$ there is a $t>0$ such that $t\lambda \in D_2$. We want to extend M to V^0 by setting

$$M(\lambda) = t^{-1}M(t\lambda).$$

From Lemma 9.4 it is easily seen that M is well-defined, and furthermore that $M(s\lambda) = |s|M(\lambda)$ for all $s \in \mathbb{R}$ and $\lambda \in V^0$. The subadditivity of M then follows easily from Lemma 9.6. $\square$

We now want to apply the above ideas to $S(\mathcal{A})$ for an order-unit space $\mathcal{A}$. Note that the V^0 of just above is then the $\mathcal{A}'^0$ of earlier. We will need the following theorem, which does not involve the above ideas.

9.8 Theorem. *Let $\mathcal{A}$ be an order-unit space, and let M be a norm on ${\mathcal{A}'}^0$. Define a metric, ρ, on $S(\mathcal{A})$ by*

$$\rho(\mu,\nu) = M(\mu-\nu).$$

If the ρ-topology coincides with the weak-$*$ topology on $S(\mathcal{A})$, then

$$M = (L_\rho)'$$

on ${\mathcal{A}'}^0$.

Proof. Since Lip_ρ is a subspace of $C(S(\mathcal{A}))$, we can set $\mathcal{A}_L = (\mathrm{Lip}_\rho) \cap Af(S(\mathcal{A}))$. Note that $\mathcal{A}_L$ need not be contained in $\mathcal{A}$ unless $\mathcal{A}$ is complete. Initially it is not clear how big $\mathcal{A}_L$ is. Parallel to our earlier notation, let V denote the normed space ${\mathcal{A}'}^0$ with norm M. Note that V need not be complete. Let V' denote the Banach space dual of V, with dual norm M'. Fix any $\nu_0 \in S(\mathcal{A})$. For any $\varphi \in V'$ define a function, $\tau(\varphi)$, on $S(\mathcal{A})$ by

$$\tau(\varphi)(\mu) = \varphi(\mu - \nu_0).$$

Then for $\mu,\nu \in S(\mathcal{A})$ we have

$$|\tau(\varphi)(\mu) - \tau(\varphi)(\nu)| = |\varphi(\mu-\nu)| \le M'(\varphi)M(\mu-\nu) = M'(\varphi)\rho(\mu,\nu).$$

Thus $\tau(\varphi) \in \mathrm{Lip}_\rho$ and $L_\rho(\tau(\varphi)) \le M'(\varphi)$. In particular, $\tau(\varphi)$ is continuous on $S(\mathcal{A})$ since ρ gives the weak-$*$ topology. Furthermore it is easily seen that $\tau(\varphi)$ is affine on $S(\mathcal{A})$. Thus $\tau(\varphi) \in \mathcal{A}_L$. Consequently τ is a norm-non-increasing linear map from (V', M') to $(\mathcal{A}_L, L_\rho)$. Let $\tilde{\tau}$ denote τ composed with the map from $\mathcal{A}_L$ to $\tilde{\mathcal{A}}_L$. Then it is easily seen that $\tilde{\tau}$ does not depend on the choice of ν_0. We now need:

9.9 Lemma. *Let $\bar{\mathcal{A}} = Af(S(\mathcal{A}))$, the completion of $\mathcal{A}$ for $\|\ \ \|$, so that $\mathcal{A}_L \subseteq \bar{\mathcal{A}}$. Then $\mathcal{A}_L$ is dense in $\bar{\mathcal{A}}$.*

Proof. Since $\mathbb{R}e \subseteq \mathcal{A}_L$, it suffices to show that $\tilde{\mathcal{A}}_L$ is dense in $\bar{\mathcal{A}}^\sim$. Let $\lambda \in D_2 \subseteq {\mathcal{A}'}^0 = (\bar{\mathcal{A}}^\sim)'$. Suppose that $\lambda(\mathcal{A}_L) = 0$. Let $\lambda = \mu - \nu$ with $\mu,\nu \in S(\mathcal{A})$. For any $\varphi \in V'$ we have $\tau(\varphi) \in \mathcal{A}_L$, so

$$0 = \lambda(\tau(\varphi)) = \mu(\tau(\varphi)) - \nu(\tau(\varphi)) = \varphi(\mu-\nu_0) - \varphi(\nu-\nu_0) = \varphi(\lambda).$$

Since this is true for all $\varphi \in V'$, it follows that $\lambda = 0$. Since D_2 spans ${\mathcal{A}'}^0$, an application of the Hahn–Banach theorem now shows that $\mathcal{A}_L$ is dense on $\bar{\mathcal{A}}$. $\square$

Now let $f \in \mathcal{A}_L$. We seek to define a linear functional, $\sigma(f)$, on ${\mathcal{A}'}^0$ related to the σ in the proof of Theorem 5.2. We first try to define σ on D_2 by

$$\sigma(f)(\lambda) = f(\mu) - f(\nu),$$

where $\lambda = \mu - \nu$ for $\mu, \nu \in S(\mathcal{A})$. But we need to show that $\sigma(f)$ is well-defined. We argue much as we did before Definition 9.3. If also $\lambda = \mu_1 - \nu_1$ for $\mu_1, \nu_1 \in S(\mathcal{A})$, then $(\mu + \nu_1)/2 = (\mu_1 + \nu)/2$. But these are elements of $S(\mathcal{A})$ and so

$$f((\mu + \nu_1)/2) = f((\mu_1 + \nu)/2).$$

But from the fact that f is affine it now follows that

$$f(\mu) - f(\nu) = f(\mu_1) - f(\nu_1).$$

Thus $\sigma(f)$ is well-defined on D_2. We now need to know that $\sigma(f)$ is "linear" on D_2. The proof that $\sigma(f)(t\lambda) = t\sigma(f)(\lambda)$ for $t \in [-1, 1]$ is similar to the proof of Lemma 9.4. The proof that $\sigma(f)(\lambda + \lambda_1) = \sigma(f)(\lambda) + \sigma(f)(\lambda_1)$ if $\lambda + \lambda_1 \in D_2$ is similar to the argument just before Definition 9.5. The proof that $\sigma(f)$ then extends to a linear functional on $\mathcal{A}'^0$ is similar to the arguments in the proof of Theorem 9.7. For $\lambda = \mu - \nu$ with $\mu, \nu \in S(\mathcal{A})$ we have

$$|\sigma(f)(\lambda)| = |f(\mu) - f(\nu)| \le L_\rho(f)\rho(\mu, \nu) = L_\rho(f)M(\mu - \nu) = L_\rho(f)M(\lambda).$$

It follows that $\sigma(f) \in V'$ and $M'(\sigma(f)) \le L_\rho(f)$. Thus σ is a norm-non-increasing linear map from $(\mathcal{A}_L, L_\rho)$ to (V', M'). Note that the constant functions are in the kernel of σ, so that σ determines a norm-non-increasing linear map from $(\tilde{\mathcal{A}}_L, \tilde{L}_\rho)$ to (V', M'). But for $f \in \mathcal{A}_L$ we have

$$\tau(\sigma(f))(\mu) = \sigma(f)(\mu - \nu_0) = f(\mu) - f(\mu_0).$$

Consequently $\tilde{\tau}(\tilde{\sigma}(\tilde{f})) = \tilde{f}$. Similarly, for $\varphi \in V'$ and $\lambda = \mu - \nu$ we have

$$\tilde{\sigma}(\tilde{\tau}(\varphi))(\lambda) = \tau(\varphi)(\mu) - \tau(\varphi)(\nu) = \varphi(\mu - \nu_0) - \varphi(\nu - \nu_0) = \varphi(\lambda),$$

so that $\tilde{\sigma}(\tilde{\tau}(\varphi)) = \varphi$. Thus $\tilde{\sigma}$ and $\tilde{\tau}$ are inverses of each other. Since they are norm-non-increasing, we obtain:

9.10 Lemma. *The map $\tilde{\tau}$ is an isometric isomorphism of (V', M') onto $(\mathcal{A}_L, L_\rho)$, with inverse $\tilde{\sigma}$.*

We can now complete the proof of Theorem 9.8. Since $\mathcal{A}_L$ is dense in $\bar{\mathcal{A}}$ by Lemma 9.9, for any $\lambda \in V'$ we have

$$(L_\rho)'(\lambda) = \sup\{\lambda(\tilde{\tau}(\varphi)) : L_\rho(\tilde{\tau}(\varphi)) \le 1\} = \sup\{\varphi(\lambda) : M'(\varphi) \le 1\} = M(\lambda).$$

□

Putting together the various pieces of this section, we obtain:

9.11 Theorem. *Let $\mathcal{A}$ be an order-unit space, and let ρ be a metric on $S(\mathcal{A})$ which gives the weak-$*$ topology. Then ρ comes from a* Lip-*norm L on $\mathcal{A}$ via the relation*

$$\rho(\mu, \nu) = L'(\mu - \nu)$$

if and only if ρ is convex, midpoint balanced, and midpoint convex.

Nik Weaver has suggested to me the following alternative treatment of the material of this section. Let V, K, and V^0 be as at the beginning of this section.

9.12 Definition. We say that a metric ρ on K is *linear* if for every $\mu, \nu \in K$, every $v \in V^0$, and every $t \in \mathbb{R}^+$ such that $\mu + tv$ and $\nu + v$ are in K we have

$$\rho(\mu, \mu + tv) = t\rho(\nu, \nu + v).$$

It is easily seen that if ρ comes from a norm on V^0 then ρ is linear. Conversely, if ρ is linear, define a norm, M, on V^0 by

$$M(v) = \rho(\mu, \mu + tv)/t$$

for any $\mu \in K$ and any $t \in \mathbb{R}^+$ such that $\mu + tv \in K$. One checks that M is well-defined and is indeed a norm. Furthermore, ρ comes from M.

Weaver also points out that if V is a locally convex topological vector space and if K is compact, then for a suitable definition of ρ being compatible with the topology, one can show that when ρ is linear and compatible, then K is isometrically isomorphic to $S(Af(K))$ when the latter is given the metric coming from the Lipschitz seminorm on $Af(K)$ coming from ρ.

It is not clear that examples will come up where it is actually useful to apply the considerations of this section in order to obtain Lip-norms. Until such examples arise, it will not be clear whether my version or Weaver's will be the more useful.

10. Musings on metrics

Since the theory in the previous sections worked for order-unit spaces, which need not be algebras, the Leibniz inequality played no significant role there. Indeed, even when one has an algebra, I have not seen how to make effective use of the Leibniz inequality. Nevertheless, most constructions of Lipschitz seminorms which I have seen in the literature seem to provide ones which do satisfy the Leibniz inequality. We will briefly explore here a variety of such constructions, and the relationships between them. Our interest will be on seeing general patterns, and we will not try to deal carefully with the many technical issues which arise. Thus we will be less precise than in the previous sections.

A very natural way to look for Lipschitz seminorms, closely related to Weaver's W^*-derivations [W2], goes as follows. Let $\mathcal{A}$ be a unital algebra and let (Ω, d) be a first-order differential calculus for $\mathcal{A}$. Thus Ω (which is also often denoted Ω^1) is an $\mathcal{A}$-$\mathcal{A}$-bimodule, and d is an Ω-valued derivation on $\mathcal{A}$, that is, a linear map from $\mathcal{A}$ into Ω which satisfies the Leibniz identity

$$d(ab) = (da)b + a(db).$$

We do not require that the range of d generates Ω. Suppose now that $\mathcal{A}$ is in fact a normed algebra, and that we have a bimodule norm, N, on Ω (for the norm $\| \quad \|$ on $\mathcal{A}$), that is, a norm such that

$$N(a\omega b) \leq \|a\| N(\omega) \|b\|$$

for $a, b \in \mathcal{A}$ and $\omega \in \Omega$. Define a seminorm L on Ω by

$$L(a) = N(da).$$

It is easily seen that L satisfies the Leibniz inequality. Since $d1 = 0$, we have $L(1) = 0$. Of course, without further hypotheses the null-space of L may be much bigger. (We should mention that not all seminorms satisfying the Leibniz inequality can be constructed in this way—see the discussion in [BC].)
There is a universal first-order differential calculus for any unital algebra $\mathcal{A}$ [Ar, C2]. We approach this in a way which emphasizes more than usual those differential calculi which are *inner*, since at least conceptually that is what Dirac operators give, as we will see shortly. We form the algebraic tensor product

$$\Omega_1^u = \mathcal{A} \otimes \mathcal{A},$$

with bimodule structure defined as usual by $a(b \otimes c)d = ab \otimes cd$. We define d by

$$da = 1 \otimes a - a \otimes 1.$$

10.1 Definition. A first-order calculus (Ω, d) is *inner* if there is a $\omega_0 \in \Omega$ such that

$$da = \omega_0 a - a\omega_0.$$

Then the calculus (Ω_1^u, d) defined above is inner, with $\omega_0 = 1 \otimes 1$. Note that here ω_0 may not be in the sub-bimodule generated by the range of d. This is an indication of why we do not require this generation property. It is simple to verify:

10.2 Proposition. *The inner first-order calculus $(\Omega_1^u, d, 1 \otimes 1)$ is universal among inner first-order differential calculi over $\mathcal{A}$, in the sense that if (Ω', d', ω_0') is any other inner first-order differential calculus, then there is a bimodule homomorphism $\Phi : \Omega_1^u \to \Omega'$ such that $\Phi(da) = d'a$ and $\Phi(1 \otimes 1) = \omega_0'$. In particular,*

$$\Phi(a \otimes b) = a\omega_0' b$$

for $a, b \in \mathcal{A}$. If Ω' is generated by ω_0' as bimodule, then Φ is surjective, so that Ω' is a quotient of Ω_1^u.

10.3 Proposition. *Any first-order differential calculus is contained in an inner first-order calculus.*

Proof. Let (Ω, d) be a first-order calculus. Set $\bar{\Omega} = \Omega \oplus \mathcal{A}$ as left $\mathcal{A}$-module, set $\bar{d}a = da \oplus 0$, and set $\bar{\omega}_0 = 0 \oplus 1$. We must extend the right action of $\mathcal{A}$ on Ω to a right action on $\bar{\Omega}$ such that $\bar{d}a = \bar{\omega}_0 a - a\bar{\omega}_0$. Thus it is clear that we must set $(0 \oplus 1)a = \bar{\omega}_0 a = da \oplus 0 + a\bar{\omega}_0 = da \oplus a$, and so

$$(\omega, b)a = (\omega a + bda, ba).$$

It is simple to check that this gives the desired structure. $\square$

Now let Ω^u denote the sub-bimodule of Ω_1^u generated by the range of d, and so spanned by elements of the form

$$adb = a \otimes b - ab \otimes 1.$$

Let (Ω', d') be a first-order differential calculus which is not inner. Expand it to an inner calculus by the construction of the previous proposition, and then restrict Φ of that proposition to Ω^u. It is clear from the construction that Φ will carry Ω^u into Ω', where Ω' is viewed as a sub-bimodule of its expansion. We obtain in this way:

10.4 Proposition. *The calculus (Ω^u, d) is universal among all first-order differential calculi over $\mathcal{A}$, in the sense that if (Ω', d') is any other first-order differential calculus, then there is a bimodule homomorphism $\Phi : \Omega^u \to \Omega$ such that $\Phi(da) = d'a$. If Ω' is generated by the range of d' as bimodule, then Φ is surjective, so that Ω' is a quotient of Ω^u.*

We notice that if (Ω, d) is any first-order differential calculus and if $\mathcal{N}$ is any sub-bimodule of Ω, then we obtain a calculus $(\Omega/\mathcal{N}, d')$ where d' is the composition of d with the canonical projection of Ω onto $\Omega/\mathcal{N}$. However, unlike the universal calculus, there may now be many more elements a for which $da = 0$ beyond the scalar multiples of 1.

Let us examine briefly what the above looks like when $\mathcal{A} = C(X)$ for a compact space X. Then $\Omega_1^u (= \mathcal{A} \otimes \mathcal{A})$ is naturally viewed as a dense sub-bimodule, in fact subalgebra, of $C(X \times X)$. The bimodule actions are, of course,

$$(fF)(x,y) = f(x)F(x,y), \ (Ff)(x,y) = F(x,y)f(y),$$

and $\omega_0 = 1 \otimes 1$ is the constant function 1, so that d is given by

$$(df)(x,y) = f(y) - f(x).$$

Then Ω^u is spanned by the fdg, where

$$(fdg)(x,y) = f(x)(g(y) - g(x)).$$

Thus the elements of Ω^u take value 0 on the diagonal, Δ, of $X \times X$, and consequently $\Omega^u \subseteq C_\infty(X \times X \setminus \Delta)$. In fact it is easy to see that Ω^u is a dense subalgebra of $C_\infty(X \times X \setminus \Delta)$.

Let ρ be an ordinary metric on X (giving the topology of X). View ρ as a strictly positive function on $X \times X \setminus \Delta$, and let $\gamma = \rho^{-1}$. Then γ is a continuous function on $X \times X \setminus \Delta$, but γ is unbounded if X is not finite. Let $C(X \times X \setminus \Delta)$ denote the algebra of continuous possibly-unbounded functions on $X \times X \setminus \Delta$. Then $C(X \times X \setminus \Delta)$ can be viewed as the algebra of operators affiliated with the C^*-algebra $C_\infty(X \times X \setminus \Delta)$ in the sense studied by Baaj [Ba] and Woronowicz [Wo]. In an evident way $C(X \times X \setminus \Delta)$ is an $\mathcal{A}$-$\mathcal{A}$-bimodule, containing γ.

There are now two routes which we can take. One is to consider the inner-derivation, d_γ, defined by γ. Thus

$$(d_\gamma f)(x,y) = \gamma(x,y)f(y) - f(x)\gamma(x,y) = (f(y) - f(x))/\rho(x,y).$$

Then we can consider bimodule norms, possibly taking value $+\infty$, on $C(X \times X \smallsetminus \Delta)$, as a way to obtain Lipschitz norms on $\mathcal{A}$. The other route is to use γ (or ρ) to directly define norms on $C_\infty(X \times X \smallsetminus \Delta)$. For the first route the most obvious norm is the supremum norm, which leads to the usual definition of the Lipschitz seminorm for a metric space.

However, we choose to explore further the second route. (But most of what we find will have a fairly evident reinterpretation in terms of the first route.) There is a large variety of ways to obtain bimodule norms on $C_\infty(X \times X \smallsetminus \Delta)$. The one which gives the usual definition of the Lipschitz seminorm for a metric is clearly

$$N(F) = \|\gamma F\|_\infty,$$

permitted to take value $+\infty$. But here are some others. Let m be any positive (finite) measure on X, and assume that $m \times m$ restricted to $X \times X \smallsetminus \Delta$ has as support all of $X \times X \smallsetminus \Delta$. Then one can consider all of the L^p-norms for $m \times m$. If one wants to put γ (or ρ) explicitly into the picture, one can consider the measure $\gamma(m \times m)$, although this just represents the choice of a different measure. Note that if f is an ordinary Lipschitz function for ρ, then γdf is a bounded function on $X \times X \smallsetminus \Delta$, so that $\|\gamma df\|_{p,m\times m}$ is finite. Thus the subalgebra of elements of $\mathcal{A}$ for which this Lipschitz seminorm is finite is dense in $\mathcal{A}$.

To explore further possibilities, let us for simplicity assume that X is finite. Then $\Omega_1^u = C(X \times X)$ can be viewed as the algebra of all matrices whose entries are indexed by elements of $X \times X$. The left and right actions of $\mathcal{A}$ on Ω_1^u can be viewed as coming from embedding $\mathcal{A}$ as the diagonal matrices and using left and right matrix multiplication. Then ω_0 is the matrix with a 1 in each entry. On $\mathcal{A}$ we keep the supremum norm, but on the matrix algebra Ω_1^u we can consider any $\mathcal{A}$-$\mathcal{A}$-bimodule norm. Let $\mathcal{B}$ denote Ω_1^u viewed as matrix algebra, and equipped with the usual C^*-algebra norm. View Ω_1^u as a $\mathcal{B}$-$\mathcal{B}$-bimodule in the evident way. Then we can consider $\mathcal{B}$-$\mathcal{B}$-bimodule norms on Ω_1^u. Any such will in particular be an $\mathcal{A}$-$\mathcal{A}$-bimodule norm. But there has been extensive study of the possible $\mathcal{B}$-$\mathcal{B}$-bimodule norms on Ω_1^u. They are commonly called "symmetric norms", and among the best known are the Schatten p-norms, which include the Hilbert–Schmidt norm and the trace norm. These have, of course, also been extensively studied for operators on infinite dimensional Hilbert spaces, and play a fundamental role in Connes' theory of integration on non-commutative spaces. (See [C2] Chapter IV and its Appendix D. A nice treatment of the finite case can be found in [Bh].) From every symmetric norm we obtain a Lip-norm on $\mathcal{A}$ (since $\mathcal{A}$ is finite-dimensional). This does not exhaust the possibilities, as there is no necessity to restrict to symmetric norms in order to get $\mathcal{A}$-$\mathcal{A}$-bimodule norms.

All of the above discussion has been for the universal differential calculus. We get many more possibilities by using other differential calculi. We continue to concentrate on the case of $\mathcal{A} = C(X)$ with X compact. Now sub-$\mathcal{A}$-$\mathcal{A}$-bimodules of $C(X \times X)$, when closed in the supremum norm, will be ideals of $C(X \times X)$, and the quotient can be identified with $C(W)$ for some closed subset W of $X \times X$. We can restrict df to W. But some condition must be placed on W if we want to ensure that $df|_W = 0$ only if f is a constant function. For this purpose it is convenient to assume, to begin with, that W contains the diagonal Δ and is symmetric about Δ, that is, if $(x,y) \in W$ then $(y,x) \in W$. Given $x \in X$ we define the W-neighborhood of x to be the (closed) set of those $y \in X$ such that $(x,y) \in W$. By the W-component of x we mean the smallest closed subset of X which contains the W-neighborhood of each of its points. If $df|_W = 0$, then f is constant on the W-component of each point. Thus a sufficient condition under which $df|_W = 0$ will imply that f is constant, is that the W-component of each point is all of X. If X is a finite set, then $W \setminus \Delta$ can be viewed as consisting of the directed edges for a graph whose vertices are the points of X. Then the above condition becomes the condition that this graph is connected in the usual sense. If X is not discrete, it is usual to require that W is a neighborhood of Δ. Then each W-neighborhood of a point will be an ordinary (closed) neighborhood, and so the W-component of each point will be both closed and open. In particular, if X is connected it will be true that $df|_W = 0$ implies that f is constant.

We remark that if W is a neighborhood of Δ and is symmetric about Δ, and if we set $\Omega = C(W)$, then the first order calculus (Ω, d) obtained as above is the typical degree-one piece of the complexes (Ω^*_W, d) used in defining the Alexander–Spanier cohomology of X. The higher-degree pieces are defined similarly but in terms of X^n for various n. The Alexander–Spanier cohomology is then obtained by taking a limit of the homology of these complexes as W shrinks to Δ. Essentially this view can be seen in lemma 1.1 of [CM], where smooth functions on a manifold are used, and in Section 1 of [MW], where continuous functions are used.

Suppose now that $\Omega = C(W)$ as above, but assume now for simplicity that W and Δ are disjoint (with W no longer required closed). Let d be defined by $df = df|_W$, and assume that if $df = 0$ then f is a scalar multiple of 1. To obtain a Lipschitz seminorm on $\mathcal{A}$ we again just need to put a bimodule norm on Ω. The method which is closest to the usual Lipschitz norm is to specify a nowhere zero function γ on W and set

$$L(f) = \|\gamma df\|_\infty$$

(on W, allowing value $+\infty$). In this context however, if we set $\rho = \gamma^{-1}$, it no longer makes much sense to ask that the triangle inequality hold for ρ. About the most that is reasonable is to ask that ρ, hence γ, be positive, and that $\gamma(x,y) = \gamma(y,x)$ for $(x,y) \in W$, $x \neq y$. This is a situation which has been widely studied. Entire books [Ra, RR] have been written about the problem of finding the corresponding distance between two probability measure on X, often under the heading of "the mass transportation problem". The function ρ is then often called a "cost function". We should clarify that when ρ is not a metric we are dealing here with mass transportation "with transshipment permitted" [RR], not the original Monge–Kantorovich [KA] mass transportation problem, which does not permit transshipment, and may well not yield a metric. When transshipment is permitted and ρ is not a metric on X, the corresponding metric on $S(X)$ is called the Kantorovich–Rubenstein metric [KR1, KR2]. For a fascinating survey of some recent developments concerning the original Monge–Kantorovich problem see [Ev].

When X is a finite set and W is viewed as specifying edges for a graph which has X as set of vertices, the cost function ρ is naturally interpreted as assigning lengths to the edges (though we will see a quite different interpretation in Section 12). Then the metric on X coming from L_ρ is the usual path-length distance on the graph. There has been much study of how to compute this path-length distance efficiently for large graphs. We remark that if one prefers to have ρ defined on all of $X \times X$ one can simply set it equal to $+\infty$ on any (x, y), $x \neq y$, which is not an edge.

We remark that in the context of cost functions on compact sets there may well be no non-constant functions for which the Lipschitz seminorm is finite. As one example let X be the unit interval $[0, 1]$, and set $\rho(x, y) = |x-y|^2$. This is, in effect, because we permit transshipment — the original Monge–Kantorovich problem is quite interesting for this particular cost function, as shown in [Ev]. It is just that the minimal cost of moving one probability measure directly to another does not then give a metric on probability measures, because it may be less costly to use two or more moves.

There is a variety of other bimodule norms, such as L^p-norms, which one can use for various differential calculi, and these give a wide variety of metrics on probability measures [Ra]. A particularly deep application of such norms, for the case of graphs, and involving explicitly Connes ideas of non-commutative metrics, appears in [Da]. (I thank Nik Weaver for bringing this paper to my attention.)

Let us now discuss briefly the case in which we have $\mathcal{A} = M_n$, a full matrix algebra. As mentioned much earlier, one natural Lip-norm on $\mathcal{A}$ is just $L = \| \quad \|^\sim$. Now $\mathcal{A}'$ can be identified by means of the normalized trace, τ, with $\mathcal{A}$ itself, but equipped with the trace-norm. Then $\mathcal{A}'^0$, as in our earlier notation, consists of the matrices with trace 0. Of course, $S(\mathcal{A})$ is identified with the positive matrices of normalized trace 1. With this identification, we have

$$\rho_L(\mu, \nu) = \mathrm{trace}(|\mu - \nu|).$$

This is exactly one of the metrics listed (with references) in the introduction to [ZS]. Another one listed there uses the Hilbert–Schmidt norm instead of the trace norm. Listed also is a variety of other metrics on $S(M_n)$ which have appeared in various applications. But I have not checked whether they come from Lip-norms. There has also been much study of the differential geometry of $S(M_n)$ for a variety of Riemannian metrics, especially the "monotone metrics", which are closely related to operator monotone functions. Two very recent articles which contain many references to previous work on this topic are [Di, S]. But the emphasis of most of this work is not on the ordinary metric which a Riemannian metric induces on $S(M_n)$, but rather on the differential geometric aspects. There is also study of the volume form which is induced, and on associated probabilistic aspects. For recent related study going in the direction of non-commutative entropy see [LR].

11. Dirac operators and differential calculi

We continue our comments of the previous section, but here we focus on how Dirac operators fit into the picture. Let $\mathcal{A}$ be a unital $*$-algebra equipped with a C^*-norm (perhaps not complete), and let π be a faithful representation of $\mathcal{A}$, that is, an isometric $*$-homomorphism of $\mathcal{A}$ into the algebra $B(\mathcal{H})$ of bounded operators on a Hilbert space $\mathcal{H}$. Let D be an essentially self-adjoint, possibly unbounded, operator on $\mathcal{H}$, and assume that $\pi(a)$ carries the domain of D into itself for each $a \in \mathcal{A}$, and that on this domain $[D, \pi(a)]$ is a bounded operator, and so extends uniquely to a bounded operator on $\mathcal{H}$. Then, following Connes, we set

$$L(a) = \|[D, \pi(a)]\|.$$

As we did earlier, it is natural to require that $[D, \pi(a)] = 0$ only when a is a scalar multiple of 1. Many important examples of this situation are now known. But in general it seems difficult to ascertain whether the corresponding metric on states gives the weak-$*$ topology, though this has been shown for certain examples in [Rf]. See also [W2, W3, W5], where the sets $\mathcal{B}_t$ defined at the beginning of Section 3 are shown to be totally bounded, in fact compact, for various examples. We do not deal with this question here, but rather try to relate the bimodule picture to the Dirac picture. One direction is apparent. We view $B(\mathcal{H})$ as an $\mathcal{A}$-$\mathcal{A}$-bimodule by setting

$$aTb = \pi(a)T\pi(b).$$

Then, although D is only affiliated with $B(\mathcal{H})$, conceptually we use the inner derivation which D defines, so that

$$da = D\pi(a) - \pi(a)D = [D, \pi(a)].$$

(This, of course, is the starting point for Connes' non-commutative differential calculus [C2].) We then note that the operator norm on $B(\mathcal{H})$ is an $\mathcal{A}$-$\mathcal{A}$-bimodule norm, and so upon setting

$$L(a) = \|[D, \pi(a)]\|$$

we obtain a Lipschitz norm, which we showed to be lower semicontinuous in Proposition 3.8.

But suppose we are given instead some first order differential calculus (Ω, d) and a bimodule norm on Ω so that we obtain the corresponding Lipschitz norm L. Can we also obtain L from a Dirac operator? For this to be possible we must have $L(a^*) = L(a)$, and L must be lower semicontinuous. As mentioned earlier, L must also fit into a family of "matrix Lipschitz seminorms". These conditions are probably not enough in general, though I have not tried to find a counterexample. But the following superficial comments help to give some perspective. (In most of the considerations which follow the algebra structure on $\mathcal{A}$ is only used in order to get the Leibniz inequality. Thus much of what follows actually works for order-unit spaces.)

We saw in Proposition 10.3 that we can extend (Ω, d) to obtain an inner first-order calculus. In analogy with this idea, suppose that we can realize Ω as a subspace of $B(\mathcal{H})$ for some Hilbert space $\mathcal{H}$, in such a way that the norm on Ω is the operator norm, and the bimodule structure is given by two $*$-representations, π_1 and π_2, of $\mathcal{A}$ on $\mathcal{H}$, so that

$$a\omega b = \pi_1(a)\omega\pi_2(b)$$

for $a, b \in \mathcal{A}$ and $\omega \in \Omega$. Suppose further that there is a possibly-unbounded essentially self-adjoint operator, D_0, on $\mathcal{H}$, such that $\pi_1(a)$ and $\pi_2(a)$ carry the domain of D_0 into itself, and such that

$$da = D_0\pi_2(a) - \pi_1(a)D_0,$$

which in particular must be a bounded operator. Set $L(a) = \|da\|$. This is not exactly the Dirac operator setting, but it is not difficult to convert it into that setting. To arrange matters so that we have only one representation, we let $\pi = \pi_1 \oplus \pi_2$ on $\mathcal{H} \oplus \mathcal{H}$ and set

$$D_1 = \begin{pmatrix} 0 & D_0 \\ 0 & 0 \end{pmatrix}.$$

Then we find that

$$L(a) = \|[D_1, \pi(a)]\|.$$

But of course D_1 is not self-adjoint. We fix this in the traditional way by again doubling the Hilbert space, with representation $\pi \oplus \pi$ of $\mathcal{A}$, and setting

$$D = \begin{pmatrix} 0 & D_1^* \\ D_1 & 0 \end{pmatrix}.$$

The corresponding Lipschitz norm is $L(a) \vee L(a^*)$, but from the self-adjointness of D one can check that we actually get back L.

Anyway, we are left with

11.1 Question. For an order-unit space $\mathcal{A}$, or a $*$-algebra $\mathcal{A}$ with C^*-norm, how does one characterize those Lip-norms on $\mathcal{A}$ which come from the Dirac operator construction?

Even for finite-dimensional commutative C^*-algebras it is not clear to me what the answer is.

As mentioned earlier, a Dirac operator also gives seminorms on all of the matrix algebras over $\mathcal{A}$, so that one can speak of this family as a "matrix Lipschitz norm", in the spirit of [Ef]. Thus a related problem is to characterize these structures.

Of course a given metric on $S(\mathcal{A})$ may come from several fairly different Dirac operators. For example, suppose that we have a compact space X, and a closed neighborhood W of the diagonal Δ of $X \times X$, together with a cost function ρ on W, just as in the previous section. As discussed there, we can use ρ together with the first-order calculus determined by W to define a Lipschitz norm on $C(X)$. (Further hypotheses are needed for it to be a Lip-norm on a dense subalgebra of $C(X)$.) Then by the procedure discussed earlier in the present section we can pass to a Dirac operator. But that procedure enlarged the Hilbert space because a first-order differential calculus usually involves two representations rather than one. We will now show that there is an alternative method which does not enlarge the Hilbert space. This is a mild generalization of my lecture comments for metric spaces mentioned earlier, whose details are indicated on page 274 of [W2]. As earlier, let m be a measure on X of full support, and consider $m \times m$ on $W \setminus \Delta$. Form the Hilbert space $\mathcal{H} = L^2(W \setminus \Delta, m \times m)$. We consider only the representation π of $\mathcal{A} = C(X)$ on $\mathcal{H}$ defined by

$$(\pi_f \xi)(x, y) = f(x)\xi(x, y).$$

(This is, of course, essentially the left action on the bimodule for W.) Define an operator, F, on $\mathcal{H}$ by the flip

$$(F\xi)(x, y) = \xi(y, x).$$

Because we are using a product measure, the operator F is self-adjoint and unitary. Define an (unbounded) positive operator, P, on $\mathcal{H}$ by

$$(P\xi)(x, y) = \xi(x, y)/\rho(x, y).$$

Because we assume that $\rho(x, y) = \rho(y, x)$ for all $(x, y) \in W$, the operators F and P commute. We define the Dirac operator by

$$D = PF,$$

so that F is the phase of D and $P = |D|$. Informal calculation shows that for any $f \in C(X)$ we have

$$([D, \pi_f]\xi)(x, y) = ((f(y) - f(x))/\rho(x, y))\xi(y, x),$$

so that

$$L(f) = \|[D, \pi_f]\| = \sup\{|f(y) - f(x)|/\rho(x, y) : (x, y) \in W\}.$$

Of course, further hypotheses must be placed on ρ in order for this to give a Lip-norm. But the right-hand side of the above equality is the usual definition of a Lipschitz norm in this situation, especially in contexts such as graph theory. It will coincide with what one obtains in the corresponding bimodule approach. Notice that the resulting distance between two points $x, y \in X$ can easily be strictly smaller than $\rho(x, y)$ (if (x, y) happens to be in W).

For an interesting alternative (but closely related) method of obtaining the usual distance on a graph (including infinite graphs) from a cost function, by means of Dirac operators, see theorem 7.2 of [Da]. Furthermore, in [Da] other very interesting and quite different Dirac operators associated to cost functions on graphs are discussed in some detail, and used to obtain improved estimates for heat kernels on graphs. They can be described in terms of first-order differential calculi and Laplace operators along much the same lines as we used in Section 10. Much of this is explicit in [Da], and we will not elaborate on it here.

We should mention here that very interesting examples of Dirac operators associated with non-commutative variants of sub-Riemannian manifolds appear in the second example following axiom $4'$ of [C3], and in [W5].

12. Resistance distance

We conclude with an appealing class of examples which do not fit into the previous framework of differential calculi, and for which the Lip-norm does not satisfy the Leibniz identity. These examples come from graphs with "cost functions" on the edges, but now the graph is interpreted as an electrical circuit with resistances on the edges, whose values are given by the cost function. These examples have been extensively studied [DS, Kl, KlR, KZ], but I have not seen earlier mention of the corresponding metric on probability measures which we will define here. It is not clear to me whether this metric is more than a curiosity.

All of the discussion here can be carried out for infinite graphs, along the lines discussed extensively in [DS], but for simplicity we only discuss finite graphs here. The examples also have a fine alternative interpretation in terms of random walks [DS]. Our term "resistance distance" is taken from the title of [KlR].

The set-up, as indicated above, is a finite graph with set X of vertices, together with strictly positive real numbers $r_{xy} = r_{yx}$ assigned to each (undirected) edge. We interpret these numbers as resistances. We assume throughout that the graph is connected. Given $x, y \in X$, $x \neq y$, we can imagine putting a voltage difference across x and y, adjusted so that one unit of current flows in at x and out at y. Then Ohm's law says that the "effective resistance" is equal to the required voltage difference. We denote this effective resistance by $\rho(x, y)$. It is, in fact, a metric on X. The only reference I know for this is [KlR, K, KZ], but my friends in probability theory tell me that within the context of random walks rather than resistances this is well-known, even if no reference comes to mind.

Suppose now that μ and ν are general probability measures on X. Although it does not seem so intuitively obvious, we will see shortly that we can establish voltages on the points of X such that unit total current flows into the circuit, with the amount flowing in at each point x given by μ_x, while unit total current flows out of the circuit, with the amount at each point given by ν (with the evident interpretation when the supports of μ and ν are not disjoint). For the analysis of this situation it is useful to define a function, c, on the edges, by $c_{xy} = 1/r_{xy}$. This is commonly called the "conductance". It is convenient to extend c to all of $X \times X$ by setting $c_{xy} = 0$ if (x, y) is not an edge (or if $y = x$). Let $f \in C(X)$, interpreted as voltages applied to the points of X. We let df be defined as earlier for the universal calculus (or for the calculus corresponding to the edges). We let ∇f denote the resulting flow inside the circuit. By Ohm's law the flow (before electrons were discovered) from x to y is given by

$$(\nabla f)(x, y) = (f(x) - f(y))c_{xy} = -c(df),$$

where by $c(df)$ we mean the pointwise product of functions. Note that ∇f is a function on directed edges, with

$$(\nabla f)(x, y) = -(\nabla f)(y, x)$$

(and value 0 if (x, y) is not an edge).

Suppose now that ω is any function on directed edges such that $\omega(x, y) = -\omega(y, x)$. We interpret $\omega(x, y)$ as giving the magnitude of a current from x to y. (To be more realistic we should require 0 circulation, but we will have no need to impose this requirement.) To sustain this current, we will in general have to insert (or extract) current at various vertices. We let $\mathrm{div}(\omega)(x)$ denote the current which must be inserted at x. By Kirchhoff's laws we have

$$\mathrm{div}(\omega)(x) = \sum_y \omega(x, y).$$

Note that because $\omega(x, y) = -\omega(y, x)$, we will have

$$\sum_x \mathrm{div}(\omega)(x) = 0,$$

which accords with the fact that the total amount of current inserted must be 0.

Suppose now that $f \in C(X)$ and that we set $\omega = \nabla f$. We see from above that the currents which must be inserted to sustain the voltages given by f must be

$$\mathrm{div}(\nabla f),$$

which we denote by Δf. To accord with our earlier notation, we let ${\mathcal{A}'}^0$ denote the signed measures, λ, on X for which $\langle 1, \lambda \rangle = 0$. The discussion of the previous paragraph can be interpreted as saying that $\Delta f \in {\mathcal{A}'}^0$.

Suppose now that we are given $\lambda \in {\mathcal{A}'}^0$. Can we find f such that $\Delta f = \lambda$? Note that since $\Delta 1 = 0$, we know that f will not be unique, but rather that, as usual with potential functions, we can expect f to be unique only up to a constant function. To proceed further we must more carefully analyze the operator Δ in the traditional way [DS, K]. For $f \in C(X)$ we have

$$\begin{aligned}(\Delta f)(x) &= \sum_y (\nabla f)(x,y) \\ &= \sum_y (f(x) - f(y))c_{xy} = f(x)\sum_y c_{xy} - \sum_y f(y)c_{xy}.\end{aligned}$$

Let D denote the diagonal matrix with diagonal entries

$$D_{xx} = \sum_y c_{xy}.$$

If we view f as a column vector, we see that

$$\Delta f = (D - C)f.$$

From the Peron–Frobenius theorem and the fact that our graph is connected, it follows that the kernel of Δ consists exactly of the constant functions. If we permit ourselves to confuse vector spaces a bit, we see that Δ is self-adjoint with respect to the standard inner-product on column vectors. Thus it carries the orthogonal complement, $\mathcal{H}$, of the constant functions into itself, and it is invertible on $\mathcal{H}$. Consequently, for every $\lambda \in {\mathcal{A}'}^0$ we can find a unique $f \in \mathcal{H}$ such that $\Delta f = \lambda$. We will write this as $f = \Delta^{-1}\lambda$, where we view Δ as restricted to $\mathcal{H}$ so that it is invertible there.
Suppose now that x and y are fixed points of X, and that $\lambda = \delta_x - \delta_y$, where δ_x denotes the δ-measure at x. Thus we are inserting one unit of current at x and extracting it at y. Let $f = \Delta^{-1}\lambda$. According to our earlier comments, the effective resistance from x to y, $\rho(x,y)$, is given by $f(x) - f(y) = (\Delta^{-1}\lambda)(x) - (\Delta^{-1}\lambda)(y)$. It is now easy to see why ρ is a metric, along the lines given in [KlR]. If z is any other point of X, let

$$g = \Delta^{-1}(\delta_x - \delta_z), \qquad h = \Delta^{-1}(\delta_z - \delta_y).$$

Clearly $f = g + h$, so

$$\rho(x,y) = g(x) - g(y) + h(x) - h(y).$$

But simple considerations show that g must take its maximum and minimum values at x and z, so that

$$g(x) - g(y) \leq g(x) - g(z) = \rho(x,z).$$

Similarly $h(x) - h(y) \le \rho(x,z)$. The triangle inequality for ρ follows.
But we are interested more generally in the effective resistance between μ and ν where μ and ν are arbitrary probability measures, and it is not even clear how this should be defined. (It does not seem natural just to use the Monge–Kantorovich metric from ρ.) In view of our earlier considerations we should form $\lambda = \mu - \nu$, and so we need an appropriate norm on ${\mathcal{A}'}^0$, and this should be the dual norm of a Lip-norm, say L, on $C(X)$, probably defined by means of a norm on Ω^u. The dual norm, L', should be such that if $\lambda = \delta_x - \delta_y$, then $L'(\lambda) = (\Delta^{-1}\lambda)(x) - (\Delta^{-1}\lambda)(y)$. But as remarked above, $\Delta^{-1}\lambda$ takes its maximum and minimum values at x and y. Thus a norm which will meet this requirement is

$$L'(\lambda) = 2\|\Delta^{-1}\lambda\|_\infty^\sim,$$

where $\| \quad \|_\infty^\sim$ is as defined in Section 1. To find L on $C(X)$ we use the self-adjointness of Δ to calculate, for $g \in C(X)$ and any $\lambda \in {\mathcal{A}'}^0$,

$$\langle g, \lambda\rangle = \langle g, \Delta\Delta^{-1}\lambda\rangle = \langle \Delta g, \Delta^{-1}\lambda\rangle.$$

The supremum over λ such that $2\|\Delta^{-1}\lambda\|_\infty^\sim \le 1$ is the same as the supremum of

$$\langle (1/2)\Delta g, h\rangle$$

over h such that $\|\tilde{h}\|_\infty^\sim \le 1$. But we saw earlier that this gives just the restriction to ${\mathcal{A}'}^0$ of the dual norm for $\| \quad \|_\infty$ on $C(X)$, which is the L^1-norm. Thus we see that we must set

$$\begin{aligned} L(g) &= (1/2)\|\Delta g\|_1 = (1/2)\sum_x |(\Delta g)(x)| \\ &= (1/2)\sum_x \left|\sum_y (g(x) - g(y))c_{xy}\right| = (1/2)\sum_x \left|\sum_y dg(x,y)c_{xy}\right|. \end{aligned}$$

This is certainly rather different from the usual Lip-norms for metrics on finite sets. The above expression suggests that we define a seminorm, N, on Ω^u by

$$N(\omega) = (1/2)\sum_x \left|\sum_y \omega(x,y)c_{xy}\right|,$$

so that we have

$$L(g) = N(dg).$$

Reversal of the earlier calculation shows that the dual norm is the L' considered above, so that we obtain the desired $\rho(\mu,\nu)$. However N will not usually be a bimodule norm, so that we are not fully in the context of the previous sections, and L need not satisfy the Leibniz inequality.
I must admit that I see no particularly natural interpretation for $L(g)$, nor for $\rho(\mu,\nu)$, even if we call the latter "effective resistance". If g were interpreted as

giving voltages on X, then $L(g)$ would be half the sum of the absolute values of the currents inserted or extracted from the circuit, and thus exactly the sum of the currents inserted into the circuit (disregarding the currents extracted). But I do not see why it is natural to give g such an interpretation as voltages. If one goes back to the effective resistance between two points, then it is easily seen that this is equal to the energy dissipated by the circuit when one unit of current is inserted. This suggests using the dissipated energy in the more general case of arbitrary probability measures μ and ν. But the energy dissipated along any edge varies as the square of the current, and one can see by examples that this causes the triangle inequality to fail. One does obtain a metric if one uses the square-root of the dissipated energy, but this does not give the correct value for the effective resistance between two points. These possibilities are not far from the Lipschitz norm used right after lemma 4.1 of [Da] to define the metric denoted there by d_3. This Lipschitz norm can be interpreted as the supremum over the points x of X of the square roots of the energy dissipations in all the edges beginning at x. Perhaps the discussion of Dirichlet spaces given in section 6 of [W6], or the "twisted bimodule structure" and corresponding differential discussed beginning on page 149 of [Me] in connection with Hudson's treatment of discrete flows and stochastic differential equations, could be used to shed more light on this. Or perhaps some of the stopping rules or mixing times considered for Markov chains, as discussed in [LW], are relevant.

Finally, we remark that it would be interesting to study resistance distance in the continuous case, for example for thin plates of resistance metal of various shapes.

References

[Al] Alfsen, E. M., *Compact convex sets and boundary integrals*, Springer-Verlag, Berlin, New York, 1971.

[Ar] Arveson, W. B., *The harmonic analysis of automorphism groups*, MR 84m:46085, Proc. Sympos. Pure Math., vol. 38, Amer. Math. Soc., Providence, RI, 1982, pp. 199–269.

[Ba] Baaj, S., *Multiplicateurs non bornés*, Thèse 3eme cycle, Université Paris VI (1980).

[BCD] Bade, W. G., Curtis, P. C., and Dales, H. G., *Amenability and weak amenability for Beurling and Lipschitz algebras*, MR 88f:46098, Proc. London Math. Soc. (3) **55**, no. 2 (1987), 359–377.

[Bh] Bhatia, R., *Matrix Analysis*, MR 98i:15003, Springer-Verlag, New York, New York, 1997.

[BC] Blackadar, B. and Cuntz, J., *Differential Banach algebra norms and smooth subalgebras of C^*-algebras*, J. Operator theory **26** (1991), 255–282.

[C1] Connes, A., *Compact metric spaces, Fredholm modules and hyperfiniteness*, Ergodic Theory and Dynamical Systems **9** (1989), 207–220.

[C2] Connes, A., *Noncommutative Geometry*, Academic Press, San Diego, 1994.

[C3] Connes, A., *Gravity coupled with matter and the foundation of non commutative geometry*, hep-th/9603053., Comm. Math. Phys. **182** (1996), 155–176.

[CM] Connes, A. and Moscovici, H., *Cyclic cohomology, the Novikov conjecture and hyperbolic groups*, MR 92a:58137, Topology **29**, no. 3 (1990), 345–388.

[Cw] Conway, J. B., *A Course in Functional Analysis, 2nd edition*, Springer Verlag, New York, 1990.

[Da] Davies, E. B., *Analysis on graphs and noncommutative geometry*, MR 93m:58110, J. Funct. Anal. **111**, no. 2 (1993), 398–430.

[Di] Dittmann, J., *On the curvature of monotone metrics and a conjecture concerning the Kubo-Mori metric*, quant-ph/9906009.

[DS] Doyle, P. G. and Snell, J. L., *Random Walks and Electric Networks*, MR 89a:94023, Mathematical Association of America, Washington, DC, 1984.

[Ef] Effros, E. G., *Advances in quantized functional analysis*, MR 89e:46064, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986), Amer. Math. Soc., Providence, RI, 1987, pp. 906–916.

[Ev] Evans, L. C., *Partial differential equations and Monge–Kantorovich mass transfer*, available at: http://berkeley.math.edu/ evans/, preprint.

[H] Hanin, L. G., *Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces*, MR 92i:46026, Proc. Amer. Math. Soc. **115**, no. 2 (1992), 345–352.

[J] Johnson, B. E., *Amenability and weak amenability for Beurling and Lipschitz algebras*, MR 88f:46098, Proc. London Math. Soc. (3) **55**, no. 2 (1987), 359–377.

[K1] Kadison, R. V., *A representation theory for commutative topological algebra*, MR **13**, 360b, Mem. Amer. Math. Soc. **1951**, no. 7 (1951), 39 pp..

[K2] Kadison, R. V., *Unitary invariants for representations of operator algebras*, MR **19**, 665e, Ann. of Math. (2) **66** (1957), 304–379.

[K3] Kadison, R. V., *Transformations of states in operator theory and dynamics*, MR **29** #6328, Topology **3**, suppl. 2 (1965), 177–198.

[KA] Kantorovič, L. V. and Akilov, G. P., *Functional Analysis (Russian)*, MR 86m:46001, Third edition, "Nauka", Moscow, 1984.

[KR1] Kantorovič L. V. and Rubinšteĭn, G. Š., *On a functional space and certain extremum problems*, MR **20** #1219, Dokl. Akad. Nauk SSSR (N.S.) **115** (1957), 1058–1061.

[KR2] Kantorovič L. V. and Rubinšteĭn, G. Š., *On a space of completely additive functions*, MR **21** #808, Vestnik Leningrad. Univ. **13**, no. 7 (1958), 52–59.

[Kl] Klein, D.J., *Graph geometry, graph metrics, and Wiener*, MR 98e:05103, Match No. 35 (1997), 7–27.

[KlR] Klein, D. J. and Randić, M., *Resistance distance*, MR 94d:94041, J. Math. Chem. **12**, no. 1-4 (1993), 81–95.

[KZ] Klein, D. J. and Zhu, H.-Y., *Distances and volumina for graphs*, MR 99f:05032, J. Math. Chem. **23**, no. 1-2 (1998), 179–195.

[LR] Lesniewski, A. and Ruskai, M. B., *Monotone Riemannian metrics and relative entropy on non-commutative probability spaces*, math-ph/9808016..

[LW] Lovász L. and Winkler, P., *Mixing times*, Microsurveys in Discrete Probability (Princeton, NJ), Amer. Math. Soc., Providence, RI, 1997, pp. 85–133.

[Me] Meyer, P.-A., *Quantum probability for probabilists*, MR 94k:81152, Lecture Notes in Math., 1538, Springer-Verlag, Berlin (1993).

[MW] Moscovici, H. and Wu, F.-B., *Index theory without symbols*, MR 96g:58184, Contemp. Math. **167** (1994), Amer. Math. Soc., Providence, RI, 304–351.

[P] Pavlović, B., *Defining metric spaces via operators from unital C^*-algebras*, Pacific J. Math. **186**, no. 2 (1998), 285–313.

[Ra] Rachev, S. T., *Probability Metrics and the Stability of Stochastic Models*, John Wiley and Sons, 1991.

[RR] Rachev, S. T. and Rüschendorf, L., *Mass Transportation Problems. Vol. I, Theory*, Springer-Verlag, New York, New York, 1998.

[Rf] Rieffel, M. A., *Metrics on states from actions of compact groups*, math.OA/9807084, Doc. Math. **3** (1998), 215–229.

[Ru] Rudin, W., *Functional Analysis*, Second edition, McGraw-Hill, Inc., New York, New York, 1991.

[S] Slater, P. B., *A priori probabilities — based on volume elements of monotone metrics — of quantum disentanglements*, quant-ph/9810026, J. Phys. **A32** (1999), 5261.

[W1] Weaver, N., *Lipschitz Algebras*, World Scientific, Singapore, 1999.

[W2] Weaver, N., *Lipschitz algebras and derivations of von Neumann algebras*, MR 97f:46081, J. Funct. Anal. **139**, no. 2 (1996), 261–300.

[W3] Weaver, N., *α-Lipschitz algebras on the noncommutative torus*, J. Operator Theory **39** (1998), 123–138.

[W4] Weaver, N., *Operator spaces and noncommutative metrics*, preprint.

[W5] Weaver, N., *Sub-Riemannian metrics for quantum Heisenberg manifolds*, math.OA/9801014.

[W6] Weaver, N., *Lipschitz algebras and derivations, II: Exterior differentiation*, math.FA/9807096.

[Wo] Woronowicz, S. L., *Unbounded elements affiliated with C^*-algebras and noncompact quantum groups*, MR 92b:46117, Comm. Math. Phys. **136**, no. 2 (1991), 399–432.

[ZS] Zyczkowski, K. and W. Słomczyński, W., *The Monge distance between* *quantum states*, quant-ph/9711011, J. Phys. A **31**, no. 45 (1998), 9095–9104.

Marc A. Rieffel
Department of Mathematics
University of California
Berkeley, California 94720-3840
USA
rieffel@math.berkeley.edu

What Are Cumulants ?

Dedicated to Professor Dietrich Morgenstern,

on the occasion of his seventy-fifth birthday

Lutz Mattner

Received: July 5, 1999

Communicated by Friedrich Götze

Abstract. Let $\mathcal{P}$ be the set of all probability measures on $\mathbb{R}$ possessing moments of every order. Consider $\mathcal{P}$ as a semigroup with respect to convolution. After topologizing $\mathcal{P}$ in a natural way, we determine all continuous homomorphisms of $\mathcal{P}$ into the unit circle and, as a corollary, those into the real line. The latter are precisely the finite linear combinations of cumulants, and from these all the former are obtained via multiplication by i and exponentiation.

We obtain as corollaries similar results for the probability measures with some or no moments finite, and characterizations of constant multiples of cumulants as affinely equivariant and convolution-additive functionals. The "no moments"-case yields a theorem of Halász. Otherwise our results appear to be new even when specialized to yield characterizations of the expectation or the variance.

Our basic tool is a refinement of the convolution quotient representation theorem for signed measures of Ruzsa & Székely.

1991 Mathematics Subject Classification: 60E05, 60E10, 60-03.
Keywords and Phrases: Additive functional, characteristic function, character, convolution, equivariance, expectation, Halász, historical note, homomorphism, mean, moment, multiplicative functional, Ruzsa, semi-invariant, semiinvariant, semigroup, Székely, variance.

1 Introduction, results, and easy proofs

1.1 Aim. Cumulants are certain functionals of probability measures. This paper attempts to explain more precisely what they are by characterizing them through their most useful properties. For simplicity, only the one-dimensional case of probability measures on $\mathbb{R}$ is treated. There the most familiar examples of cumulants are the expectation and the variance. Our results yield, in particular, new descriptions of the roles played by these latter two functionals in probability theory.

1.2 Guide. The definition of cumulants is recalled in Subsection 1.4 below, as formula (4). The useful properties of cumulants, referred to above, are the homomorphism property (5) and their transformation behaviour under affine mappings, (14). The relation between cumulants and moments is recalled in Subsection 1.5.

Subsection 1.6 introduces topologies on the domains of definition of the cumulants, with the aim of formulating regularity assumptions in our theorems and corollaries. That some regularity assumptions are actually necessary, at least in the results 1.8 – 1.12, is demonstrated in 1.20.

Theorem 1.8, characterizing the continuous characters of the semigroup $\mathrm{Prob}_\infty(\mathbb{R})$, is the main result of the present paper. Its natural forerunner from the literature, namely the theorem of Halász, is recalled in 1.10 below as a special case of Corollary 1.9.

Another corollary of Theorem 1.8, and perhaps the most interesting result of this paper, is the characterization of the finite linear combinations of cumulants as the continuous, $\mathbb{R}$-valued, and convolution-additive functionals of probability measures, stated in Theorem 1.11 and Corollary 1.12. Such results were conjectured by Kemperman (1972). By restricting the functionals to be $[0,\infty[$-valued, we arrive at a characterization of the variance in 1.14. [A related result of Martin Diaz (1977) is discussed in 1.22.]

Our next results, 1.17 and 1.18, are spezializations of 1.8 and 1.11 to scale equivariant functionals, the definition of which being recalled in 1.16.

As a further corollary, we obtain in 1.19 a characterization of the expectation as the only nontrivial continuous functional homomorphic with respect to additive and multiplicative convolutions.

Historical and etymological remarks on cumulants are given in Subsection 1.21.

Subsection 1.22 discusses some further references related to the present work.

Easy proofs are given immediately after the statement of a result in Section 1. The only difficult proof of this paper, needed for the "only if" part of our main result 1.8, is the content of Section 2. Its basic technical tool, refining the convolution quotient representation theorem for signed measures of Ruzsa & Székely (1983, 1985, 1988), is supplied in Subsection 2.5.

1.3 Some notation and conventions. The positive integers are denoted by $\mathbb{N}$, the nonnegative ones by $\mathbb{N}_0$.

If $\mathcal{X}$ is a set equipped with a σ-algebra $\mathcal{A}$, we let $\mathrm{Prob}(\mathcal{X})$ denote the set of all probability measures defined on $\mathcal{A}$. The real line $\mathbb{R}$ is understood to be equipped with its Borel σ-algebra. The convolution of $P, Q \in \mathrm{Prob}(\mathbb{R})$ is denoted by $P * Q$. We write δ_a for the Dirac measure concentrated at $a \in \mathbb{R}$, and $\delta := \delta_0$ for the one concentrated at zero. For the image measure of a probability measure P under a measurable function f, we use the notation $f \Box P$. We write supp P for the support [= minimal closed set of probability one] of a $P \in \mathrm{Prob}(\mathbb{R})$.

$\mathrm{Prob}(\mathbb{R})$ will mainly be considered as a semigroup with respect to convolution. Homomorphisms of a semigroup [below always a sub-semigroup of $\mathrm{Prob}(\mathbb{R})$] into the multiplicative group $\mathbb{T}$ of complex numbers of absolute value one will be called *characters*, homomorphisms into the additive group $\mathbb{R}$ will be called *additive* functions.

1.4 Cumulants. We present below the usual introduction of cumulants and their most important properties. For $P \in \mathrm{Prob}(\mathbb{R})$, let $\widehat{P}$ denote its Fourier transform or characteristic function, defined by

$$\widehat{P}(t) \quad := \quad \int e^{itx}\, dP(x) \qquad (t \in \mathbb{R}). \tag{1}$$

The most important reason for considering Fourier transforms of probability measures is multiplicativity with respect to convolution:

$$(P * Q)\widehat{\ }(t) \quad = \quad \widehat{P}(t) \cdot \widehat{Q}(t) \qquad (P, Q \in \mathrm{Prob}(\mathbb{R}),\, t \in \mathbb{R}). \tag{2}$$

Let log denote the usual logarithm defined on, say, $\{z \in \mathbb{C} : |z-1| < 1\}$. Let $P \in \mathrm{Prob}(\mathbb{R})$. Then $\widehat{P}$ is continuous with $\widehat{P}(0) = 1$, so that $\log \circ \widehat{P}$ is defined in some P-dependent neighbourhood of zero. Now put

$$\mathrm{Prob}_r(\mathbb{R}) \quad := \quad \left\{ P \in \mathrm{Prob}(\mathbb{R}) : \int |x|^r\, dP(x) < \infty \right\} \qquad (r \in \mathbb{N}_0), \tag{3}$$

and assume that $r \in \mathbb{N}$ and $P \in \mathrm{Prob}_r(\mathbb{R})$. Then $\widehat{P}$ and thus $\log \circ \widehat{P}$ is r times continuously differentiable in the neighbourhood of zero introduced above, and the number

$$\boxed{\kappa_r(P) \;:=\; i^{-r} \left(D^r \log \circ \widehat{P} \right)(0)} \tag{4}$$

is called the rth *cumulant* of P. [Readers wondering about this strange name are referred to Subsection 1.21.] It is easy to show that the cumulants are real-valued functionals. Their most important property, which obviously follows from (2) and (4), is additivity with respect to convolution:

$$\boxed{\kappa_r(P * Q) \;=\; \kappa_r(P) + \kappa_r(Q) \qquad (r \in \mathbb{N},\, P, Q \in \mathrm{Prob}_r(\mathbb{R}))}. \tag{5}$$

In other words: For each $r \in \mathbb{N}$, $(\mathrm{Prob}_r(\mathbb{R}), *)$ is a semigroup on which κ_r is an additive function.

1.5 Examples, expression in terms of moments. The two most familiar examples of cumulants are the mean μ and the variance σ^2, since

$$\begin{aligned}\kappa_1(P) &= \mu(P) := \int x\,dP(x) \qquad (P \in \mathrm{Prob}_1(\mathbb{R})),\\ \kappa_2(P) &= \sigma^2(P) := \int (x-\mu(P))^2\,dP(x) \qquad (P \in \mathrm{Prob}_2(\mathbb{R})).\end{aligned}$$

These formulas are special cases of the relation between cumulants and the *moments*

$$\mu_r(P) := \textstyle\int x^r\,dP(x) = i^{-r}(D^n\widehat{P})(0) \qquad (r \in \mathbb{N}_0,\ P \in \mathrm{Prob}_r(\mathbb{R})).$$

One possibility of expressing this relation is to use the recursion

$$\mu_{r+1} = \sum_{l=0}^{r} \binom{r}{l} \mu_{r-l}\kappa_{l+1} \qquad (r \in \mathbb{N}_0), \tag{6}$$

which is easily proved using the Leibniz rule for the differentiation of a product and the representation of the moments as derivatives: For $P \in \mathrm{Prob}_{r+1}(\mathbb{R})$ put $\varphi := \widehat{P}$ and $\psi := \log\varphi$, in a neighbourhood of zero, and compute $D^{r+1}\varphi = D^r(\varphi \cdot D\psi) = \sum_{l=0}^r \binom{r}{l}(D^{r-l}\varphi)\cdot(D^{l+1}\psi)$, evaluate the extreme left and right hand sides at zero, and divide by i^{r+1}, to arrive at (6). Since the coefficients of μ_{r+1} and κ_{r+1} in (6) are both one, it follows by induction that

$$\kappa_r = \mu_r + \text{polynomial without constant term in } \mu_1, \ldots, \mu_{r-1} \ (r \in \mathbb{N}), \tag{7}$$

and that corresponding relations hold when μ and κ are interchanged. Various explicit fomulas derived from these relations and some examples of actual computations of cumulants may be found in Chapter 3 of Kendall, Stuart & Ord (1987). We merely note here two further examples, for convenience rewritten in terms of centered moments,

$$\begin{aligned}\kappa_3(P) &= \int (x-\mu(P))^3\,dP(x) \qquad (P \in \mathrm{Prob}_3(\mathbb{R})),\\ \kappa_4(P) &= \int (x-\mu(P))^4\,dP(x) - 3\left(\sigma^2(P)\right)^2 \qquad (P \in \mathrm{Prob}_4(\mathbb{R})).\end{aligned}$$

As one might suspect on seeing these formulas, the variance κ_2 is the only nonnegative cumulant. [This fact follows easily from 1.13 below, as can be seen from the proof of 1.14.]

1.6 Topologies on some subsets of $\mathrm{Prob}(\mathbb{R})$. One of our aims is to show that every "reasonable" homomorphism from $(\mathrm{Prob}_r(\mathbb{R}), *)$ into $(\mathbb{R}, +)$ is a linear combination of cumulants of order at most r. This is the content of Corollary 1.12, where "reasonable" is specified to mean "continuous". To this end we introduce here on each $\mathrm{Prob}_r(\mathbb{R})$ a topology. In order to make the continuity assumption in Corollary 1.12 weak, we have to choose a strong topology on $\mathrm{Prob}_r(\mathbb{R})$. We take the one induced by the weighted total variation metric d_r defined by

$$d_r(P,Q) \quad := \quad \int (1+|x|^r)\, d|P-Q|(x) \qquad (P,Q \in \mathrm{Prob}_r(\mathbb{R})). \tag{8}$$

We further consider

$$\mathrm{Prob}_\infty(\mathbb{R}) \quad := \quad \bigcap_{r\in\mathbb{N}_0} \mathrm{Prob}_r(\mathbb{R}),$$

which is the largest set of probability measures on which every cumulant is defined. We topologize $\mathrm{Prob}_\infty(\mathbb{R})$ by the family of metrics $(d_r : r \in \mathbb{N}_0)$.

1.7 Lemma. a) *Each* $\kappa_r|\mathrm{Prob}_r(\mathbb{R})$ *is continuous with respect to* d_r.

b) *Let* $r \in \mathbb{N}$ *and* $c \in]0,\infty[$. *Then there exists a sequence* (P_n) *in* $\mathrm{Prob}_\infty(\mathbb{R})$ *with*

$$\lim_{n\to\infty} d_{r-1}(P_n,\delta) \quad = \quad 0, \tag{9}$$

$$\lim_{n\to\infty} \kappa_l(P_n) \quad = \quad 0 \qquad (l=1,\dots,r-1), \tag{10}$$

$$\lim_{n\to\infty} \kappa_r(P_n) \quad = \quad c. \tag{11}$$

Proof. a) The functionals $(\mu_l : l = 1,\dots,r)$ are obviously continuous with respect to d_r, and (7) shows in particular that κ_r is a polynomial function of them.

b) We may restrict attention to those $n \in \mathbb{N}$ with $cn^{-r} \le 1$ and put $P_n := (1-cn^{-r})\delta + cn^{-r}\delta_n$. Then $P_n \in \mathrm{Prob}_\infty(\mathbb{R})$, and $d_{r-1}(P_n,\delta) = \frac{c}{n}$ yields (9). By part a), (9) implies (10). Finally, (11) follows from $\mu_l(P_n) = cn^{l-r}$ $(l=1,\dots,r)$ and (7). □

1.8 Theorem (continuous characters of $\mathrm{Prob}_\infty(\mathbb{R})$). *A function* $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ *is a continuous character iff*

$$\boxed{\chi(P) = \exp(i\sum_{l\in\mathbb{N}} c_l\kappa_l(P)) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R}))} \tag{12}$$

holds for some finitely supported sequence of real numbers $(c_l : l \in \mathbb{N})$. *The latter, if existent, is uniquely determined by* χ.

Proof. The proof of the "only if" part is the content of Section 2. The "if" part follows trivially from 1.7 a) and (5).

Finally suppose that we have (12) and an analogous representation of χ involving another finitely supported sequence $(\tilde{c}_l : l \in \mathbb{N})$. Then the sequence $(d_l) := (c_l - \tilde{c}_l)$ yields an analogous representation of the constant character 1. Suppose that not all d_l vanish. Put $r := \min\{l : d_l \neq 0\}$ and apply 1.7 b) with $c := \pi/|d_r|$. Then $1 = \exp(i \sum_{l=1}^{r} d_l \kappa_l(P_n)) \to \exp(\pm i\pi) = -1$ for $n \to \infty$. This contradiction shows that we must have $d_l = 0$ for every $l \in \mathbb{N}$, as was to be proved. □

1.9 Corollary. *Let $r \in \mathbb{N}_0$. A function $\chi|\mathrm{Prob}_r(\mathbb{R})$ is a continuous character iff* (12) *holds with $c_l = 0$ for $l > r$, and with* $\mathrm{Prob}_r(\mathbb{R})$ *in place of* $\mathrm{Prob}_\infty(\mathbb{R})$.

Proof. Again, the "if" part follows from from 1.7 a) and (5). To prove "only if": Let $\chi|\mathrm{Prob}_r(\mathbb{R})$ be a continuous character. Then, by 1.8, its restriction $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ fulfils (12) for some finitely supported sequence (c_l). Assume that $c_l \neq 0$ for some $l > r$. Put $\tilde{r} := \min\{l \in \mathbb{N} : c_l \neq 0\}$. Choose (P_n) according to 1.7 b) with $\tilde{r}$ in place of r and with $c := \pi/|c_{\tilde{r}}|$. Then, since $r < \tilde{r}$, we have $P_n \to \delta$ with respect to d_r. On the other hand, we have $\chi(P_n) \to -1 \neq 1 = \chi(\delta)$. This contradiction to the continuity of χ shows that we must have $c_l = 0$ for $l > r$. It follows that the right hand side of (12) is defined and continuous on $\mathrm{Prob}_r(\mathbb{R})$. Since $\mathrm{Prob}_\infty(\mathbb{R})$ is obviously dense in $\mathrm{Prob}_r(\mathbb{R})$, this implies that (12) also holds with $\mathrm{Prob}_r(\mathbb{R})$ in place of $\mathrm{Prob}_\infty(\mathbb{R})$. □

1.10 Theorem of Halász. The last corollary yields in particular a theorem of Halász, presented on page 132 of Ruzsa & Székely (1988), which reads:

> 1 *is the only character of* $\mathrm{Prob}(\mathbb{R})$ *continuous with respect to weak convergence.*

In fact, the special case $r = 0$ of our Corollary 1.9 is slightly stronger, since our continuity assumption refers to a stronger topology on $\mathrm{Prob}(\mathbb{R})$.

1.11 Theorem (additive functions on $\mathrm{Prob}_\infty(\mathbb{R})$). *A function $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to \mathbb{R}$ is continuous and additive iff*

$$\boxed{\kappa(P) = \sum_{l \in \mathbb{N}} c_l \kappa_l(P) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R}))} \tag{13}$$

holds for some finitely supported family of real numbers $(c_l : l \in \mathbb{N})$. The latter, if existent, is uniquely determined by κ.

Proof. The "if" part and the uniqueness of (c_l) follows via multiplication by i and subsequent exponentiation from the corresponding statements in 1.8.

To prove the "only if" part, let $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to \mathbb{R}$ be continuous and additive. Put

$$\chi(P) \quad := \quad \exp(i\kappa(P)) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R})).$$

Then χ satisfies the hypothesis of Theorem 1.8, and hence can be represented as in (12). This implies

$$\kappa(P) \quad = \quad \eta(P) + \sum_l c_l \kappa_l(P) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R})),$$

where $\eta|\mathrm{Prob}_\infty(\mathbb{R}) \to 2\pi\mathbb{Z}$. Since η must be additive, $\eta(\delta) = 0$. Since η must be continuous and $\mathrm{Prob}_\infty(\mathbb{R})$ is convex, $\eta(\mathrm{Prob}_\infty(\mathbb{R}))$ must be connected. [Here we have used the obvious fact that for $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$ the function $[0,1] \ni t \mapsto tP + (1-t)Q \in \mathrm{Prob}_\infty(\mathbb{R})$ is continuous.] Thus $\eta = 0$. □

1.12 Corollary. *Let $r \in \mathbb{N}_0$. A function $\kappa|\mathrm{Prob}_r(\mathbb{R}) \to \mathbb{R}$ is continuous and additive iff* (13) *holds with $c_l = 0$ for $l > r$ and with* $\mathrm{Prob}_r(\mathbb{R})$ *in place of* $\mathrm{Prob}_\infty(\mathbb{R})$.

Proof. Deduce 1.12 from 1.9, by arguing as in the proof of 1.11. Alternatively, deduce 1.12 from 1.11 by arguing as in the proof of 1.9. □

1.13 Lemma (cumulants of Bernoulli distributions). *For $r \in \mathbb{N}$, let $f_r|[0,1] \to \mathbb{R}$ be defined by*

$$f_r(p) \quad := \quad \kappa_r((1-p)\delta_0 + p\delta_1) \qquad (p \in [0,1]).$$

Then, for each r, f_r is a polynomial function of degree r with r simple zeros in $[0,1]$.

Proof. It is known [for example, from Kendall, Stuart & Ord (1987), exercise 5.1] that

$$f_{r+1}(p) \quad = \quad p \cdot (1-p) \cdot f_r'(p) \qquad (r \in \mathbb{N},\, p \in [0,1]),$$

where the prime denotes differentiation with respect to p. Since $f_1(p) = p$ for $p \in [0,1]$, the claim follows by an induction argument, using Rolle's theorem and the fact that f_r' has at most $r-1$ zeros, counting multiplicity. □

1.14 A characterization of the variance. *A function $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to [0,\infty[$ is continuous and additive iff $\kappa = c\kappa_2$ for some $c \in [0,\infty[$.*

Proof. The "if" claim is trivial. To prove "only if", we may by Theorem 1.11 start from the representation (13). Inserting there $P = \delta_a$ with $a \in \mathbb{R}$, we see that the assumption $\kappa \geq 0$ forces $c_1 = 0$. Thus, except for the trivial case $\kappa = 0$, we have

$$\kappa(P) = \sum_{l=2}^{r} c_l \kappa_l(P) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R}))$$

for some $r \geq 2$ with $c_r \neq 0$. Suppose now that $r \geq 3$. Then we may, by the lemma 1.13, choose a Bernoulli distribution $P_0 = (1-p)\delta_0 + p\delta_1$ with $c_r\kappa_r(P_0) < 0$. It follows that $\kappa(P) < 0$ for $P := (x \mapsto ax)\Box P_0$ with $a > 0$ sufficiently large, using (14) below. This contradiction proves our claim. $\Box$

1.15 Affine equivariance of cumulants. The second most important property of the cumulants is their behaviour under affine transformations: For $r \in \mathbb{N}$, $P \in \mathrm{Prob}_r(\mathbb{R})$ and $a, b \in \mathbb{R}$, we have

$$\kappa_r((x \mapsto ax+b)\Box P) \quad = \quad \begin{cases} a\kappa_1(P) + b & (r = 1), \\ a^r \kappa_r(P) & (r \geq 2). \end{cases} \tag{14}$$

In particular, each cumulant is affinely equivariant in the sense of the following definition and, by a trivial specialization, also scale equivariant.

1.16 Definition (equivariance). a) Let $\mathcal{X}$ be a set and $\mathcal{T}$ be a set of functions from $\mathcal{X}$ into $\mathcal{X}$. A function $\varphi|\mathcal{X}$ is called *equivariant*, with respect to $\mathcal{T}$, if we have the implication

$$x, y \in \mathcal{X},\ \varphi(x) = \varphi(y),\ T \in \mathcal{T} \quad \Longrightarrow \quad \varphi(T(x)) = \varphi(T(y)). \tag{15}$$

b) For $a, b \in \mathbb{R}$ define $T_{a,b}|\mathrm{Prob}(\mathbb{R}) \to \mathrm{Prob}(\mathbb{R})$ by

$$T_{a,b}(P) \quad := \quad (x \mapsto ax+b)\Box P \qquad (P \in \mathrm{Prob}(\mathbb{R}))$$

and put $\mathcal{T} := \{T_{a,b} : a, b \in \mathbb{R}\}$. Let $\mathcal{P} \subset \mathrm{Prob}(\mathbb{R})$ satisfy the implication $P \in \mathcal{P}$, $T \in \mathcal{T} \implies T(P) \in \mathcal{P}$. Then a function $\varphi|\mathcal{P}$ is called *affinely equivariant* if it is equivariant with respect to $\mathcal{T}$, in the sense of part a).

c) We define a function $\varphi|\mathcal{P}$ to be *scale equivariant* if it satisfies the definition given in b) above, but with $b = 0$ and $a > 0$ in the definition of $\mathcal{T}$.

1.17 Theorem (equivariant continuous characters of $\mathrm{Prob}_\infty(\mathbb{R})$). *A function $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ is a scale equivariant continuous character iff*

$$\chi(P) \quad = \quad \exp(ic\kappa_r(P)) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R})) \tag{16}$$

for some $r \in \mathbb{N}$ and some $c \in \mathbb{R}$.

Proof. The "if" part is trivial. To prove "only if": Define $S_a(P) := (x \mapsto ax)\Box P$ for $P \in \mathrm{Prob}(\mathbb{R})$ and $a \in]0, \infty[$. For $\lambda \in]0, \infty[$, let P_λ denote the Poisson distribution with expectation λ. Then

$$\kappa_l(S_a(P_\lambda)) \quad = \quad a^l \lambda \qquad (l \in \mathbb{N},\ a, \lambda \in]0, \infty[). \tag{17}$$

Now let $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ be a scale equivariant continuous character. Applying 1.8, we get (12) for some finitely supported sequence $(c_l : l \in \mathbb{N})$, and we have to show that there is at most one $l \in \mathbb{N}$ with $c_l \neq 0$. Using (17), (12) yields in particular

$$\chi(S_a(P_\lambda)) \quad = \quad \exp(i\lambda p(a)) \qquad (a, \lambda \in \,]0, \infty[) \tag{18}$$

where p is the polynomial function defined by

$$p(a) \quad := \quad \sum_{l \in \mathbb{N}} c_l a^l \qquad (a \in \mathbb{C}).$$

Now assume, to get a contradiction, that there are at least two $l \in \mathbb{N}$ with $c_l \neq 0$. Then for arbitrary $a_1, a_2 \in \,]0, \infty[$ with $a_1 \neq a_2$ and arbitrary $\lambda_1, \lambda_2 \neq 0$, there exists a number $b \in \,]0, \infty[$ with

$$\lambda_1 p(ba_1) - \lambda_2 p(ba_2) \quad \notin \quad 2\pi\mathbb{Z}. \tag{19}$$

[Proof: Assume without loss of generality that $a_1 < a_2$. If our claim is false, then the rational function $\mathbb{C} \ni z \mapsto R(z) := p(a_1 z)/p(a_2 z)$ is constant. But by our assumption on p, $\varrho := \sup\{|z| : z \in \mathbb{C}, \, p(z) = 0\} > 0$. In view of $0 < a_1 < a_2$, it is obvious that R has a zero, namely on the circle $\{|z| = \varrho/a_1\}$. Hence $R \equiv 0$ and thus $p \equiv 0$, a contradiction.]

Now choose specifically $a_1, a_2 \in \,]0, \infty[$ with $a_1 \neq a_2$ in such a way that $p(a_1) \cdot p(a_2) > 0$. Choose $\lambda_1, \lambda_2 \in \,]0, \infty[$ with

$$\lambda_1 p(a_1) \quad = \quad \lambda_2 p(a_2), \tag{20}$$

choose b as in (19), and put $Q_k := S_{a_k}(P_{\lambda_k})$ for $k = 1, 2$. Then (18) and (20) yield $\chi(Q_1) = \chi(Q_2)$, whereas (18) also yields $\chi(S_b(Q_k)) = \chi(S_{ba_k}(P_{\lambda_k})) = \exp(i\lambda_k p(ba_k))$ for $k = 1, 2$, so that (19) implies $\chi(S_b(Q_1)) \neq \chi(S_b(Q_2))$, in contradiction to the scale equivariance of χ. □

1.18 Theorem (scale equivariant additive functions on $\mathrm{Prob}_\infty(\mathbb{R})$). *A function $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to \mathbb{R}$ is continuous, additive, and scale equivariant, iff there exist $r \in \mathbb{N}$ and $c \in \mathbb{R}$ such that $\kappa = c\kappa_r$.*

Proof. Proceed as in the proof of the "only if" part of Theorem 1.11, but use equivariance of χ and 1.17 in place of 1.8. □

1.19 A characterization of the expectation. Notation: In this subsection only, we write $P \boxplus Q$ for the usual convolution $P * Q$ of $P, Q \in \mathrm{Prob}(\mathbb{R})$, and $P \boxdot Q$ for the multiplicative convolution of $P, Q \in \mathrm{Prob}(\mathbb{R})$, that is, the distribution of $X \cdot Y$ with X, Y being independent random variables with distributions P, Q.

Theorem. *Let $\kappa|\mathrm{Prob}_\infty(\mathbb{R}) \to \mathbb{R}$ be continuous. Then we have both*

$$\kappa(P \boxplus Q) = \kappa(P) + \kappa(Q), \tag{21}$$
$$\kappa(P \boxdot Q) = \kappa(P) \cdot \kappa(Q) \tag{22}$$

for $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$, iff either $\kappa = \kappa_1$ or $\kappa = 0$.

Proof. The "if" part is obvious. So assume that κ is continuous and satisfies (21) and (22). By applying (22) to $Q = \delta_a$, for every $a \in]0, \infty[$, we see that κ is scale equivariant. Hence (21) and Corollary 1.18 yield $\kappa = c\kappa_r$ for some $c \in \mathbb{R}$ and some $r \in \mathbb{N}$. Choose $P \in \mathrm{Prob}_\infty(\mathbb{R})$ with $\kappa_r(P) \neq 0$, for example P = Poisson distribution with parameter 1. Insert this P and $Q = \delta_1$ into (22), use $\kappa = c\kappa_r$, and divide by $\kappa_r(P)$. The result is $c = c^2\kappa_r(\delta_1)$. If $r \geq 2$, then $\kappa_r(\delta_1) = 0$, hence $c = 0$ and thus $\kappa = 0$. If $r = 1$, then $\kappa_r(\delta_1) = 1$, hence either again $c = 0$ and $\kappa = 0$, or $c = 1$ and thus $\kappa = \kappa_1$. □

1.20 "Counterexamples". Examples a) and b) below show that the continuity assumptions in 1.8 – 1.12 can not be omitted without substitute. Both a) and b) should be regarded as pathological. On the other hand, the examples in c) show that not only 1.8 – 1.12, but also 1.14 and, using (23), also 1.17 and 1.18 receive non-pathological counterexamples if the continuity assumption is dropped and if the domain of definition of the functionals is taken to be to small. Concerning 1.8 – 1.12, we may also refer to example d), suggested to me by I.Z. Ruzsa, where the domain of definition of κ could be thought of as being not much smaller than $\mathrm{Prob}_\infty(\mathbb{R})$.

a) By the axiom of choice, there exists a discontinuous additive function $f|\mathbb{R} \to \mathbb{R}$. Now $\kappa(P) := f(\mu(P))$ defines a discontinuous additive function $\kappa|\mathrm{Prob}_1(\mathbb{R}) \to \mathbb{R}$.

b) [Ruzsa & Székely (1988), pp. 122-123, 2.3 and 2.4] construct, using the axiom of choice, a homomorphism κ from $(\mathrm{Prob}(\mathbb{R}), *)$ into $(\mathbb{R}, +)$ which extends the expectation κ_1 defined on the subsemigroup $\mathrm{Prob}_1(\mathbb{R})$. They also show that each such κ assumes negative values for some P with support in $[0, \infty[$. It follows that the κ constructed is a discontinuous additive function from $\mathrm{Prob}(\mathbb{R})$ into $\mathbb{R}$.

c) On the semigroup

$$\mathrm{Prob}_c(\mathbb{R}) := \{P \in \mathrm{Prob}(\mathbb{R}) : \mathrm{supp}\, P \text{ compact}\} \subset \mathrm{Prob}_\infty(\mathbb{R})$$

we obtain an additive and nonnegative functional, normalized here as to satisfy additionally condition *ii)* from 1.22 below, by each of the following definitions:

$$\kappa(P) := \frac{1}{2} \cdot (\max \mathrm{supp}\, P - \min \mathrm{supp}\, P) \qquad (P \in \mathrm{Prob}_c(\mathbb{R})), \tag{23}$$
$$\kappa(P) := \frac{\log \widehat{P}(i) + \log \widehat{P}(-i)}{2 \log \cos i} \qquad (P \in \mathrm{Prob}_c(\mathbb{R})). \tag{24}$$

[In (24), we use of course the definition (1) with $\mathbb{C}$ in place of $\mathbb{R}$.]

d) Consider the semigroup

$$\mathcal{P} \quad := \quad \Big\{P \in \mathrm{Prob}_\infty(\mathbb{R}) \,:\, \widehat{P} \text{ holomorphic near zero}\Big\} \quad \subset \quad \mathrm{Prob}_\infty(\mathbb{R}).$$

Let $(a_l : l \in \mathbb{N})$ be any sequence of real numbers satisfying $a_l = O(\varepsilon^l)$, for every $\varepsilon > 0$. Then

$$\kappa(P) \quad := \quad \sum_{l=1}^{\infty} \frac{a_l}{l!} \kappa_l(P) \qquad (P \in \mathcal{P}) \tag{25}$$

defines an additive function on $\mathcal{P}$. [To see that the series in (25) always converges, observe that $\log \circ \widehat{P}$ is now holomorphic in some P-dependent neighbourhood of zero, so that its Taylor series $\sum_{l=1}^{\infty} \kappa_l(P) \cdot (iz)^l/l!$ converges for $|z|$ sufficiently small.]

1.21 Some early history and etymology. Cumulants were apparently first introduced by T.N. Thiele [1838-1910] under the name of "half-invariants". Hald (1981) describes, on pages 7-10, Thiele's contributions and their insufficient acknowledgement by K. Pearson and R.A. Fisher. According to Hald, cumulants are first defined in the book Thiele (1889). [This I did not check. Hald's formula (4.1), claimed to be Thiele's definition, is, up to an obvious misprint, the now well-known recursion (6), determining κ_{r+1} as a polynomial in the moments μ_l.] In a later and more accessible version of his book, Thiele (1903) essentially gives definition (4). Hald (1998) contains a much more comprehensive early history of cumulants.

Later authors, such as Craig (1931) and Wishart (1929), refer to the cumulants as "semi-invariants of Thiele", while Fisher (1929-30), on page 200 of his paper, simply calls them "semi-invariants", without bothering to name Thiele. But Wishart and Fisher, who obviously new about each others work before publication, prefer to use the new term "cumulative moment functions" instead. The reason for adopting this term is hinted at in Fisher's paper: On page 199, he gives an interesting although perhaps not quite precise definition of rather general "moment functions" of populations, roughly speaking by polynomial estimability, which seems at any rate to be intended to include polynomial functions of finitely many ordinary moments, and hence in particular cumulants. On page 202, Fisher then refers to a "cumulative property" of the logarithm of the Laplace transform which, expressed in terms of the cumulants, is just condition (5). Thus the the adjective "cumulative" refers, in this context, to a homomorphism condition. In particular, it is not used to distinguish a concept related to probability measures from a corresponding concept related to probability densities, as would often be the case in the older statistical literature.

Finally, "cumulative moment function" was abbreviated to "cumulant" by Fisher & Wishart (1931-32) and Fisher (1932), with Hotelling (1933) claiming to have suggested this name, which quickly became the standard one in the english language literature. The first publication having the word "cumulant" in its title seems to be the paper by Cornish & Fisher (1937), who repeat the definition, but already Haldane (1937), page 136, uses "cumulants" without definition or reference.

Readers generally interested in the history of probabilistic or statistical terms are referred to David (1995, 1998) as a useful starting point.

1.22 RELATED WORK NOT DISCUSSED ABOVE. The following papers have some relation with the present one.

Craig (1931) states on page 160 a forerunner of our Corollary 1.18. Where we assume mere continuity of κ, Craig assumes in particular that κ is a polynomial function of some finite number of moments μ_l. His treatment is not quite rigorous: For example, no domain of definition of κ is specified, his conclusion is $\kappa = \kappa_r$ for some r [instead of the correct conclusion $\kappa = c\kappa_r$ for some r and c], and a proof is offered only for the case where κ is a polynomial function of $\mu_1, \ldots, \mu_4$.

Savage (1971) characterizes moments and more general expectations of exponential polynomials as functionals κ satisfying, on the one hand, conditions like $\kappa(P * Q) = T(\kappa(P), \kappa(Q))$ with T unspecified and, on the other hand, having a representation $\kappa(P) = \int f\,dP$ with f unspecified. His first assumption is more liberal than our homomorphism assumptions, but his second assumption is rather restrictive, excluding for example every cumulant κ_r with $r \geq 2$. Thus the work of Savage is incomparable to the present one.

Martin Diaz (1977), Teorema 4, states a characterization of the variance which may be formulated as follows. We temporarily put $\mathcal{P} := \{P \in \mathrm{Prob}(\mathbb{R}) : \mathrm{supp}\, P \text{ finite}\}$.

THEOREM (MARTIN DIAZ) *Let $\kappa|\mathcal{P} \to [0, \infty[$ and assume:*

i) For every $n \in \mathbb{N}$, the map

$$\mathbb{R}^n \times \left\{ p \in]0,1]^n : \sum_{i=1}^n p_i = 1 \right\} \quad \ni \quad (x,p) \mapsto \kappa(\sum_{i=1}^n p_i \delta_{x_i})$$

is partially continuous in the two variables x and p.

ii) $\kappa(\delta_1) = 0$, $\kappa(\frac{1}{2}(\delta_{-1} + \delta_1)) = 1$.

iii) If we put $\kappa(X) := \kappa(P)$ for every random variable X with distribution $P \in \mathcal{P}$, then

$$\kappa(\sum_{i=1}^n X_i) \quad = \quad \sum_{i=1}^n \kappa(X_i)$$

> *whenever the X_i are* pairwise *independent random variables, on a common probability space, with distributions belonging to $\mathcal{P}$.*

Then $\kappa = \kappa_2$.

We observe that the word "pairwise" renders the third assumption rather confining. But without this word, a counterexample would be obtained by restricting to $\mathcal{P}$ either κ from (23) or (24). These examples may be regarded as negative solutions to the problem stated in Martin Diaz (1977) on page 96, while our result 1.14 may be regarded as a kind of positive solution.

Good (1979) speculates about the existence of a useful notion of "fractional cumulants", perhaps to be defined via fractional differentiation of $\log \circ \widehat{P}$ in analogy to (4). Such a definition, if possible, should lead to an additive function on $\mathrm{Prob}_\infty(\mathbb{R})$, and Theorem 1.11 could be taken as an indication that it will not lead to anything new and useful.

Heyer (1981) reviews, among other topics, axiomatic approaches to expectation and variances for probability measures on compact groups, referring to earlier publications of himself and of Maksimov, in particular Maksimov (1980). Although somewhat similar in spirit to the present paper, there is no overlap in the results obtained.

Characterizations of the variance not referring to the semigroup structure of $\mathrm{Prob}(\mathbb{R})$ have been provided by Bomsdorf (1974), by Gil Alvarez (1983), and by Kagan & Shepp (1998). The former two are somewhat similar to the characterization of the Shannon entropy by Fadeev's axioms, as presented in Rényi (1970), page 548.

2 The main proof

2.1 Further notation and conventions. The proof of the "only if" part of Theorem 1.8, given in 2.8 below, is prepared by the introduction of an auxiliary topological vector space $\mathcal{H}$ in 2.2 and the identification of its dual $\mathcal{H}'$ in 2.3. We will use some tools from functional and Fourier analysis as explained in Rudin (1991). In particular, we assume as known the spaces $\mathcal{C}^\infty(\mathbb{R})$, $\mathcal{D}(\mathbb{R})$, $\mathcal{D}'(\mathbb{R})$ with their usual topologies. We depart from the conventions of Rudin (1991) in that here a topological vector space is not necessarily assumed to be Hausdorff.

We let $\mathcal{U}$ denote the set of all open symmetric neighbourhoods of $0 \in \mathbb{R}$. For $U \in \mathcal{U}$, a function $h|U \to \mathbb{C}$ is called *hermitean* if

$$h(t) \quad = \quad \overline{h(-t)} \qquad (t \in U).$$

2.2 The space $\mathcal{H}$ of germs of hermitean $\mathcal{C}^\infty$ functions vanishing at zero. We consider

$$X \quad := \quad \{h \in \mathcal{C}^\infty(\mathbb{R}) \,:\, h \text{ hermitean}, \, h(0) = 0\}$$

as a topological vector space over $\mathbb{R}$, with the topology inherited from the usual topology of $\mathcal{C}^\infty(\mathbb{R})$. We further consider the vector subspace

$$N \quad := \quad \{h \in X \,:\, \exists U \in \mathcal{U} \text{ with } h|U = 0\}$$

of X, and form the quotient topological vector space

$$\mathcal{H} \quad := \quad X/N.$$

For $h \in X$, we write $[h]$ for the equivalence class $H \in \mathcal{H}$ with $h \in H$. It easy to see, though for our purposes unnecessary to check, that N is not closed, so that $\mathcal{H}$ is not Hausdorff. Since $\mathcal{C}^\infty(\mathbb{R})$ is metrizable, $\mathcal{H}$ is pseudo-metrizable, and a sequence $(H_j : j \in \mathbb{N})$ converges to $0 \in \mathcal{H}$ iff there exist $h_j \in H_j$ with $h_j \to 0 \in X$. [Proof: The discussion in Sections 1.40, 1.41 of Rudin (1991) applies with obvious changes, necessitated by the nonclosedness of our N. In particular, if d is some tranlation-invariant metric for X, the formula $\varrho([h_1],[h_2]) := \inf\{d(h_1 - h_2, g) : g \in N\}$ defines a translation-invariant pseudo-metric ϱ for $\mathcal{H}$. And if $([h_j]) : j \in \mathbb{N})$ is a sequence in $\mathcal{H}$ with $\lim \varrho([h_j],[0]) = 0$, we may choose $g_j \in N$ with $d(h_j, g_j) \le 2\varrho([h_j],[0]) + j^{-1}$, yielding $\tilde{h}_j := h_j - g_j \in [h_j]$ with $\tilde{h}_j \to 0$.]

The value at zero of the derivatives $D^l H(0)$ of a $H \in \mathcal{H}$, occuring below, is defined in the obvious way.

2.3 The dual $\mathcal{H}'$ of $\mathcal{H}$. *A function $\Lambda|\mathcal{H}$ is an $\mathbb{R}$-valued, continuous, and $\mathbb{R}$-linear functional iff there exists an $n \in \mathbb{N}_0$ and a finite sequence of real numbers $(c_l : 1 \le l \le n)$ such that*

$$\Lambda(H) \quad = \quad \sum_{l=1}^{n} c_l \cdot i^{-l}(D^l H)(0) \qquad (H \in \mathcal{H}). \tag{26}$$

Proof. The "if" claim is obviously true. To prove "only if": Let $\Lambda|\mathcal{H} \to \mathbb{R}$ be continuous and $\mathbb{R}$-linear. Define $S|\mathcal{D}(\mathbb{R}) \to \mathbb{R}$ by

$$S(\varphi) \quad := \quad \Lambda([\frac{1}{2}\left(\varphi - \varphi(0) + \overline{\varphi - \varphi(0)}^{\vee}\right)]) \qquad (\varphi \in \mathcal{D}(\mathbb{R})),$$

where $\check{\psi}(t) := \psi(-t)$. It is obvious that S is well-defined and $\mathbb{R}$-valued, as well as continuous and $\mathbb{R}$-linear. It follows that the functional $T|\mathcal{D}(\mathbb{R}) \to \mathbb{C}$ defined by

$$T(\varphi) \quad := \quad S(\varphi) - iS(i\varphi) \qquad (\varphi \in \mathcal{D}(\mathbb{R}))$$

is continuous and $\mathbb{C}$-linear, that is, a distribution $\in \mathcal{D}'(\mathbb{R})$. It is easily checked that T has support contained in $\{0\}$. Hence, by Rudin (1991), Theorem 6.24 d) and Theorem 6.25, there is an $n \in \mathbb{N}_0$ and a sequence of complex numbers $(b_l : 0 \le l \le n)$ such that

$$T(\varphi) \quad = \quad \sum_{l=0}^{n} b_l \cdot \left(D^l \varphi\right)(0) \qquad (\varphi \in \mathcal{D}(\mathbb{R})).$$

Since $S = \mathrm{Re}\, T$, we get for $H = [h] \in \mathcal{H}$, using the hermitean property of h and $h(0) = 0$,

$$\begin{aligned}\Lambda(H) &= S(h) \\ &= \mathrm{Re}\, T(h) \\ &= \sum_{l=1}^{n} \mathrm{Re}\, \left(b_l \cdot \left(D^l h\right)(0)\right) \\ &= \sum_{l=1}^{n} \mathrm{Re}\, (b_l i^l) \cdot i^{-l} \left(D^l h\right)(0),\end{aligned}$$

and thus (26) with $c_l = \mathrm{Re}\, (b_l i^l)$. □

2.4 Convergence in $\mathrm{Prob}_\infty(\mathbb{R})$. *Let P be an element of and (P_j) be a net in $\mathrm{Prob}_\infty(\mathbb{R})$. Then $\lim P_j = P$, in the topology of $\mathrm{Prob}_\infty(\mathbb{R})$, iff $\lim P_j = P$ with respect to total variation distance and*

$$\lim_j \int x^l \, dP_j(x) \quad = \quad \int x^l \, dP(x) \qquad (l \in \mathbb{N}). \tag{27}$$

Proof. Let first w be any nonnegative measurable function on a measurable space $\mathcal{X}$. Let $P, Q \in \mathrm{Prob}(\mathcal{X})$ with $\int w \, d(P+Q) < \infty$, and fix $a > 0$. Then

$$\begin{aligned}\int w \, d|P-Q| &\leq \int w \cdot (w \leq a) \, d|P-Q| + \int w \cdot (w > a) \, d(P+Q) \\ &= \int w \cdot (w \leq a) \, d|P-Q| + 2 \int w \cdot (w > a) \, dP \\ &\quad + \int w \, d(Q-P) - \int w \cdot (w \leq a) \, d(Q-P) \\ &\leq 2 \int w \cdot (w \leq a) \, d|P-Q| + 2 \int w \cdot (w > a) \, dP \\ &\quad + \int w \ \, dQ - \int w \ \, dP.\end{aligned}$$

Now let (P_j) be a net in $\mathrm{Prob}(\mathcal{X})$ with $\int w \, dP_j < \infty$ for every j. The preceding inequality shows that $\lim \int w \, d|P - P_j| = 0$ if both $\lim \int 1 \, d|P - P_j| = 0$ and $\overline{\lim} \int w \, dP_j \leq \int w \, dP$. Applied to $\mathcal{X} = \mathbb{R}$ and $w(x) = 1 + x^{2n}$, for each $n \in \mathbb{N}$, the "if" part follows. The "only if" part is trivial. □

2.5 Quotients of characteristic functions. *Let*

$$\varphi \in \Phi \quad := \quad \{\varphi \in \mathcal{D}(\mathbb{R}) \, : \, \varphi(0) = 1, \varphi \text{ hermitean}\}.$$

a) *There exist $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$ with*

$$\varphi \widehat{Q} \quad = \quad \widehat{P}. \tag{28}$$

b) *Let* (φ_j) *be a net in* Φ *with* $\lim \varphi_j = \varphi$ *in the* $\mathcal{D}(\mathbb{R})$*-topology. Then we may choose* $P_j, Q_j \in \mathrm{Prob}_\infty(\mathbb{R})$ *with* $\varphi_j \widehat{Q}_j = \widehat{P}_j$ *and*

$$\lim P_j = P, \qquad \lim Q_j = Q \quad \text{in } \mathrm{Prob}_\infty(\mathbb{R}). \tag{29}$$

Remark. As said before in 1.2, this basic tool of the present paper is a refinement of a theorem of Ruzsa & Székely. In particular, most of the following proof of part a) is as in Ruzsa & Székely (1988), pages 126-127.

Proof. We will calculate in

$$M^1(\mathbb{R}) \quad := \quad \text{set of all bounded complex measures on } \mathbb{R},$$

which is well known to be a Banach algebra, with convolution $*$ as multiplication and norm $\|\cdot\|$ defined by

$$\begin{aligned} \|\mu\| \quad &:= \quad \int 1\, d|\mu| \qquad (\mu \in M^1(\mathbb{R})), \\ |\mu| \quad &:= \quad \text{total variation measure of } \mu. \end{aligned} \tag{30}$$

For a $\mu \in M^1(\mathbb{R})$, its Fourier transform is the continuous function $\widehat{\mu}$ defined by

$$\widehat{\mu}(t) \quad := \quad \int e^{itx}\, d\mu(x) \qquad (t \in \mathbb{R}).$$

We assume as known properties of the Fourier transform as explained, for example, in Chapter 7 of Rudin (1991). All elements of $M^1(\mathbb{R})$ actually occuring below will in fact belong to

$$M^1_\infty(\mathbb{R}) \quad := \quad \left\{ \mu \in M^1(\mathbb{R}) \,:\, \int |x|^l\, d|\mu|(x) < \infty \quad (l \in \mathbb{N}_0) \right\}.$$

For $\mu \in M^1_\infty(\mathbb{R})$, we have $\widehat{\mu} \in \mathcal{C}^\infty(\mathbb{R})$.

a) We have $\varphi = \widehat{\mu}$ with $\mu \in M^1_\infty(\mathbb{R})$, μ real, $\mu(\mathbb{R}) = 1$. [Apply Theorem 7.7 of Rudin (1991).]

Choose $\alpha, \beta \in [0, \infty[$ and $R \in \mathrm{Prob}_\infty(\mathbb{R})$ with

$$\|(\mu - \delta) * R\| \;=\; \alpha \;<\; \beta \tag{31}$$

and

$$R^{*2} \quad \geq \quad \beta R. \tag{32}$$

[For example, if R is any centered normal distribution, then (32) is true with $\beta = 2^{-1/2}$, and for R sufficiently flat (31) is true as well. Alternatively, we may take $\beta = 2^{-1}$ and for R a sufficiently flat uniform distribution on an interval $[-a, a]$.]

Put

$$\begin{aligned} S &:= \beta^{-1}|(\mu-\delta)*R|, && (33)\\ Q &:= (1-\frac{\alpha}{\beta})R^{*2}*\sum_{k=0}^{\infty}S^{*k}, && (34)\\ P &:= \mu * Q. && (35) \end{aligned}$$

Since S is a sub-probability measure with $\|S\| = S(\mathbb{R}) = \alpha/\beta < 1$, the series in (34) is convergent in $M^1(\mathbb{R})$, and $Q \in \mathrm{Prob}(\mathbb{R})$. Also $P(\mathbb{R}) = 1$ and, easily verified,

$$\begin{aligned} (1-\frac{\alpha}{\beta})^{-1}P &= \mu * R^{*2} * \sum_{k=0}^{\infty} S^{*k}\\ &= R^{*2} + R^{*2} * (\mu-\delta+S) * \sum_{k=0}^{\infty} S^{*k}, \end{aligned}$$

where, using (32) and (33),

$$\begin{aligned} R^{*2} * (\mu-\delta+S) &\geq R*(R*(\mu-\delta)+\beta S)\\ &\geq 0. \end{aligned}$$

Hence $P \geq 0$ and thus $P \in \mathrm{Prob}(\mathbb{R})$.

By $0 \leq S \leq \beta^{-1}(|\mu|+\delta)*R$, $S \in M^1_\infty(\mathbb{R})$. Hence $\widehat{S} \in \mathcal{C}^\infty(\mathbb{R})$. Since (34) shows that

$$\widehat{Q}(t) = (1-\frac{\alpha}{\beta}) \cdot (\widehat{R}(t))^2 \cdot (1-\widehat{S}(t))^{-1} \qquad (t \in \mathbb{R}), \tag{36}$$

and since also $\widehat{R} \in \mathcal{C}^\infty(\mathbb{R})$, it follows that $\widehat{Q} \in \mathcal{C}^\infty(\mathbb{R})$. Since (35) implies (28), $\widehat{P}$ is $\mathcal{C}^\infty$ as well, at least in some neighbourhood of zero. Since P, Q are probability measures, it follows that $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$. [Compare, for example, Feller (1971), page 528, problem 15.]

b) We continue to use the notation of the above proof of part a). Let, additionally, μ_j denote the element of $M^1_\infty(\mathbb{R})$ with $\varphi_j = \widehat{\mu}_j$, and

$$\alpha_j := \|(\mu_j-\delta)*R\|.$$

By Theorem 7.7 of Rudin (1991), we have $\lim \mu_j = \mu$ in the Schwartz space $\mathcal{S}(\mathbb{R})$. It follows that

$$\lim \mu_j = \mu \quad \text{with respect to the norms} \quad \|\cdot\|_k \qquad (k \in \mathbb{N}_0), \tag{37}$$

where

$$\|\nu\|_k := \int (1+|x|^k)\, d|\nu|(x) \qquad (k \in \mathbb{N}_0,\ \nu \in M^1_\infty(\mathbb{R})).$$

The particular case $k = 0$ implies $\lim \mu_j = \mu$ with respect to the norm $\|\cdot\|$ from (30), hence $\lim \alpha_j = \alpha$. We may and do assume that $\alpha_j < \beta$ in what follows. Put $S_j := \beta^{-1}|(\mu_j - \delta) * R|$, $Q_j := (1-(\alpha_j/\beta))R^{*2} * \sum_{k=0}^{\infty} S_j^{*k}$, and $P_j = \mu_j * Q_j$. Then $Q_j, P_j \in \mathrm{Prob}_\infty(\mathbb{R})$ with $\varphi_j \widehat{Q}_j = \widehat{P}_j$, and what remains to be shown is (29).

By (37),

$$\lim S_j = S \quad \text{with respect to the norms} \quad \|\cdot\|_k \qquad (k \in \mathbb{N}_0). \tag{38}$$

Using (38) and the definition of Q_j, P_j, we get $\lim Q_j = Q$ and $\lim P_j = P$ with respect to $\|\cdot\|$. From (38) we also get $\lim \widehat{S}_j = \widehat{S}$ in $\mathcal{C}^\infty(\mathbb{R})$. Since we have (36) with α_j replacing α, $\widehat{Q}_j$ replacing $\widehat{Q}$, and $\widehat{S}_j$ replacing $\widehat{S}$, we may conclude that $\lim \widehat{Q}_j = \widehat{Q}$ in $\mathcal{C}^\infty(\mathbb{R})$. By $\varphi_j \widehat{P}_j = \widehat{Q}_j$, we deduce $\lim \widehat{P}_j|U = \widehat{P}|U$ in $\mathcal{C}^\infty(U)$, for some neighbourhood U of zero. Hence we have in particular (27) and the corresponding statement for (Q_j), so that we reach (29) via 2.4. □

2.6 Lemma. *Let $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ be a character, not necessarily continuous. If $P_1, P_2, Q_1, Q_2 \in \mathrm{Prob}_\infty(\mathbb{R})$, and if there exists an $U \in \mathcal{U}$ with*

$$\widehat{P}_1(t)\widehat{Q}_2(t) \quad = \quad \widehat{P}_2(t)\widehat{Q}_1(t) \qquad (t \in U),$$

then

$$\frac{\chi(P_1)}{\chi(Q_1)} \quad = \quad \frac{\chi(P_2)}{\chi(Q_2)}. \tag{39}$$

Proof. There exists an $R \in \mathrm{Prob}_\infty(\mathbb{R})$ with $\operatorname{supp} \widehat{R} \subset U$. Thus $\widehat{P}_1\widehat{Q}_2\widehat{R} = \widehat{P}_2\widehat{Q}_1\widehat{R}$ everywhere, so that we successively get

$$\begin{aligned} P_1 * Q_2 * R &= P_2 * Q_1 * R, \\ \chi(P_1)\chi(Q_2)\chi(R) &= \chi(P_2)\chi(Q_1)\chi(R), \end{aligned}$$

and hence (39). □

2.7 From χ to a linear functional Λ. *Let $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ be a continuous character. Then there exists a $\Lambda \in \mathcal{H}'$ with*

$$\chi(P) \quad = \quad \exp(i\Lambda(\log \circ [\widehat{P}])) \qquad (P \in \mathrm{Prob}_\infty(\mathbb{R})). \tag{40}$$

Here $\log \circ [\widehat{P}]$ of course denotes the element of $\mathcal{H}$ containing the functions $h \in X$ satisfying

$$h(t) \quad = \quad \log \widehat{P}(t) \qquad (t \in U)$$

for some $U \in \mathcal{U}$ with $U \subset \Big\{ t \in \mathbb{R} : |\widehat{P}(t) - 1| < 1 \Big\}$.

Proof. Follows from Steps 1-5 below. $\square$

Step 1: Construction of a function $X|\mathcal{H}$. *Let $H \in \mathcal{H}$. Then we may define $X(H) \in \mathbb{T}$ by the construction leading to (42) below, and this definition is independent of the choices of h, U, ω, P, Q made along the way.*

Proof. Choose $h \in H$. Define $\psi \in \mathcal{C}^\infty(\mathbb{R})$ by

$$\psi(t) \quad := \quad \exp(h(t)) \qquad (t \in \mathbb{R}). \tag{41}$$

Choose $U \in \mathcal{U}$ with compact closure and choose $\omega \in \mathcal{D}(\mathbb{R})$ real and symmetric with $\omega|U = 1$. Define $\varphi \in \mathcal{D}(\mathbb{R})$ by

$$\varphi(t) \quad := \quad \omega \cdot \psi.$$

Then φ is hermitean with $\varphi(0) = 1$, and hence satisfies the assumptions of 2.5. So we may choose $P, Q \in \mathrm{Prob}_\infty(\mathbb{R})$ satisfying (28), and put

$$X(H) \quad := \quad \frac{\chi(P)}{\chi(Q)}. \tag{42}$$

To show that this definition is independent of the choices made along the way, consider two choices $(h_i, U_i, \omega_i, P_i, Q_i)$, for $i \in \{1, 2\}$, yielding two values $X_i(H)$. There exists a $V \in \mathcal{U}$ with $\varphi_1|V = \varphi_2|V$. Hence (28) applied to φ_i, P_i, Q_i implies $\widehat{P}_1/\widehat{Q}_1 = \widehat{P}_2/\widehat{Q}_2$ on $U := V \cap \{t : \varphi_1(t) \neq 0\}$, so that Lemma 2.6 yields $X_1(H) = X_2(H)$. $\square$

Step 2: The relation between X and χ. *For $P \in \mathrm{Prob}_\infty(\mathbb{R})$,*

$$\chi(P) \quad = \quad X(\log \circ [\widehat{P}]).$$

Proof. Changing notation, let $P_1 \in \mathrm{Prob}_\infty(\mathbb{R})$. Put $H := \log \circ [\widehat{P}_1]$. Referring to Step 1 and its notation, let us denote one choice for the computation of $X(H)$ by (h, U, ω, P_2, Q_2), with (ψ, φ) accordingly defined. Then $\varphi = \widehat{P}_1$ in some $\tilde{U} \in \mathcal{U}$. With $Q_1 := \delta$ it follows that $\widehat{P}_1\widehat{Q}_2 = \widehat{P}_2\widehat{Q}_1$ in $\tilde{U}$. Hence (42), Lemma 2.6, and $\chi(\delta) = 1$, successively yield

$$X(H) = \frac{\chi(P_2)}{\chi(Q_2)} = \frac{\chi(P_1)}{\chi(Q_1)} = \chi(P_1).$$

$\square$

Step 3: *The function $X|\mathcal{H} \to \mathbb{T}$ defined in Step 1 is a character, with respect to addition in $\mathcal{H}$.*

Proof. We have to prove that

$$X(H_1 + H_2) \quad = \quad X(H_1) \cdot X(H_2) \qquad (H_1, H_2 \in \mathcal{H}).$$

So let $H_1, H_2 \in \mathcal{H}$. Choose $(U_i, h_i, V_i, \omega_i, P_i, Q_i)$ and define (ψ_i, φ_i) as in Step 1 to calculate $X(H_i)$ for $i \in \{1, 2\}$. Then we may use the choice

$$(h_1 + h_2,\, U_1 \cap U_2,\, \omega_1 \cdot \omega_2,\, P_1 * P_2,\, Q_1 * Q_2),$$

leading to $\psi = \psi_1 \cdot \psi_2$ and $\varphi = \varphi_1 \cdot \varphi_2$, to compute $X(H_1 + H_2)$. The result is

$$\begin{aligned} X(H_1 + H_2) &= \chi(P_1 * P_2) \cdot (\chi(Q_1 * Q_2))^{-1} \\ &= \chi(P_1) \cdot \chi(P_2) \cdot (\chi(Q_1) \cdot \chi(Q_2))^{-1} \\ &= X(H_1) \cdot X(H_2). \end{aligned}$$

□

STEP 4: CONTINUITY. *X is continuous.*

PROOF. Since $\mathcal{H}$ is pseudometrizable, it suffices to consider any given convergent sequence $(H_j : j \in \mathbb{N})$, with $\lim H_j = H$. There exist $h \in H$, $h_j \in H_j$, such that

$$\lim h_j \quad = \quad h \qquad \text{in } \mathcal{C}^\infty(\mathbb{R}).$$

Starting from the present h, choose and define, respectively, ψ, U, and ω as in Step 1 around equation (41). Analogously, define ψ_j and then φ_j, using the same U and ω as for ψ, φ. Then $\lim \varphi_j = \varphi$ in $\mathcal{D}(\mathbb{R})$. Now apply part b) of 2.5 to choose P, Q, P_j, Q_j with the properties stated there. Then, using Step 1 and the continuity of χ,

$$X(H_j) \;=\; \tfrac{\chi(P_j)}{\chi(Q_j)} \;\to\; \tfrac{\chi(P)}{\chi(Q)} \;=\; X(H).$$

□

STEP 5: *There exists a $\Lambda \in \mathcal{H}'$ with $X = \exp \circ (i\Lambda)$.*

PROOF. This is always true whenever $\mathcal{H}$ is a topological $\mathbb{R}$-vectorspace with dual $\mathcal{H}'$, and $X|\mathcal{H}$ a continuous character, with respect to the additive group of $\mathcal{H}$. See section (23.32.a) on page 370 of Hewitt & Ross (1979) for a proof assuming, and using, that $\mathcal{H}$ is additionally Hausdorff. For the general case, needed here, apply the special case to the Hausdorff quotient space of $\mathcal{H}$, obtained by identifying points $h_1, h_2 \in \mathcal{H}$ iff $h_2 - h_1$ belongs to the closure of $\{0\}$.

□

2.8 PROOF OF THE "ONLY IF" PART OF THEOREM 1.8. Let $\chi|\mathrm{Prob}_\infty(\mathbb{R})$ be a continuous character. Then there exists a linear functional Λ as in 2.7. By 2.3, Λ has a representation as in (26). Inserting this representation into (40) and applying the definition (4) yields (12). □

ACKNOWLEDGEMENT. The present work was supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft. I thank H.A. David, A. Hald, J.H.B. Kemperman, B. Roos, and I.Z. Ruzsa for helpful remarks.

References

An asterisk indicates work I have found discussed in other sources, but have not seen in the original.

Bomsdorf, E. (1974). Zur Charakterisierung von Lokations- und Dispersionsmaßen. *Metrika* 21, 223-229.

Cornish, E.A. & Fisher, R.A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* 4, 1-14.

Craig, C.C. (1931). On a property of the semi-invariants of Thiele. *Ann. Math. Statist.* 2, 154-164.

David, H.A. (1995). First (?) occurence of common terms in mathematical statistics. *The American Statistician* 49, 121-133.

David, H.A. (1998). First (?) occurence of common terms in probability and statistics-a second list, with corrections. *The American Statistician* 52, 36-40.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II, 2nd. Ed.* Wiley, N.Y.

Fisher, R.A. (1929-30). Moments and product moments of sampling distributions. *Proc. London Math. Soc.* 30, 199-238.

*Fisher, R.A. (1932). *Statistical Methods for Research Workers. 4th. Ed.* Oliver and Boyd, Edinburgh.

Fisher, R.A. & Wishart, J. (1931-32). The derivation of the pattern formulae of two-way partitions from those of simpler patterns. *Proc. London Math. Soc.* (2) 33, 195-208.

Gil Alvarez, María Angeles (1983). Caracterización axiomática para la varianza. *Trab. Estad. Invest. Oper.* 34, 40-51.

Good, I.J. (1979). Fractional moments and cumulants: some unsolved problems. *J. Statist. Comp. Simulation* 9, 314-315.

Hald, A. (1981). T. N. Thiele's contributions to statistics. *Int. Statist. Review* 49, 1-20.

Hald, A. (1998). The early history of the cumulants and the Gram-Charlier series. Preprint, Department of Theoretical Statistics, University of Copenhagen. (25 pages)

Haldane, J.B.S. (1937). The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small. *Biometrika* 29, 133-143. Correction note in *Biometrika* 31 (1939), 220.

Hewitt, E. & Ross, K.A. (1979). *Abstract Harmonic Analysis I*, 2nd ed. Springer, Berlin.

Heyer, H. (1981). Moments of probability measures on a group. *Int. J. Math. Sci.* 4, 231-249.

Hotelling, H. (1933). Review of Fisher (1932). *J. Amer. Statist. Ass.* 28, 374-375.

Kagan, A. & Shepp, L.A. (1998). Why the variance? *Statist. Probab. Letters* 38, 329-333.

Kemperman, J.H.B. (1972). Problem P 92. *Aeq. Math.* 8, 172.

Kendall, M., Stuart, A. & Ord, J.K. (1987). *Kendall's Advanced Theory of Statistics, Vol. 1, Distribution Theory.* Griffin, London.

Maksimov, V.M. (1980). Mathematical expectations for probability distributions on compact groups. *Math. Z.* 174, 49-60. [MR 82g:60020]

Martin Diaz, Miguel (1977). Caracterización de la varianza. *Trab. Estad. Invest. Oper.* 28, Nos. 2 and 3, 85-97. [MR 58, #31513]

Rényi, A. (1970). *Probability Theory.* North-Holland, Amsterdam, and Akadémiai Kiadó, Budapest.

Rudin, W. (1991). *Functional Analysis, 2nd. Ed.* McGraw-Hill, N.Y.

Ruzsa, I.Z. & Székely, G.J. (1983). Convolution quotients of nonnegative functions. *Mh. Math.* 95, 235-239.

Ruzsa, I.Z. & Székely, G.J. (1985). No distribution is prime. *Z. Wahrscheinlichkeitstheorie verw. Geb.* 70, 263-269.

Ruzsa, I.Z. & Székely, G.J. (1988). *Algebraic Probability Theory.* Wiley, Chichester.

Savage, L.J. (1971). The characteristic function characterized and the momentousness of moments. In: *Studi di probabilità, statistica e ricerca operativa in onore di Giuseppe Pompilj*, Tipografia Oderisi Editrice, Gubbio, pp. 131-141. Reprinted in: *The Writings of Leonard Jimmie Savage*, The American Statistical Association and The Institute of Mathematical Statistics, pp. 615-625 (1981).

*Thiele, T.N. (1889). *Forelæsninger over Almindelig Iagttagelseslære: Sandsynlighedsregning og mindste Kvadraters Methode.* Reitzel, København.

Thiele, T.N. (1903). *The theory of observations.* C. & E. Layton, London. Reprinted 1931 in: *Ann. Math. Statist.* 2, 165-308.

Wishart, J. (1929). A problem in combinatorial analysis giving the distribution of certain moment statistics. *Proc. London Math. Soc.* 29, 309-321.

Lutz Mattner
Universität Hamburg
Institut für Mathematische Stochastik
Bundesstr. 55
20146 Hamburg
Germany
mattner@math.uni-hamburg.de

The Automorphism Group of Linear Sections of the Grassmannians $\mathbb{G}(1,N)$

J. Piontkowski and A. Van de Ven

Received: April 4, 1999
Revised: October 27, 1999

Communicated by Thomas Peternell

Abstract. The Grassmannians of lines in projective N-space, $\mathbb{G}(1,N)$, are embedded by way of the Plücker embedding in the projective space $\mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})$. Let H^l be a general l-codimensional linear subspace in this projective space.

We examine the geometry of the linear sections $\mathbb{G}(1,N)\cap H^l$ by studying their automorphism groups and list those which are homogeneous or quasihomogeneous.

1991 Mathematics Subject Classification: 14L27, 14M15, 14J50, 14E09
Keywords and Phrases:

0 Introduction

Complete intersections in projective space have been studied extensively from many points of view. A natural generalisation is the study of complete intersections in Grassmannians. The first case that presents itself is the case of intersections with linear spaces. Indeed, there is an extensive literature on the simplest case, the Grassmannian of lines in the 3-space, where intersections are known as linear complexes and congruences of lines. L. Roth has studied the rationality of linear sections of Grassmannians of lines in general. If they are smooth and if the dimension of the intersection is greater than half the dimension of the Grassmannian, then they are rational. R. Donagi determined the cohomology and the intermediate Jacobian of some linear sections of Grassmannians of lines.

In this paper we study the linear sections from the point of automorphism groups. Let $\mathbb{G}(1,N)$ be the Grassmann variety of lines in projective N-space, canonically embedded in $\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$ and let H^l be an l-codimensional linear subspace in this space. For general H^l we determine the automorphism groups for $\mathbb{G}(1,N)\cap H$, $\mathbb{G}(1,N)\cap H^2$, $\mathbb{G}(1,4)\cap H^3$, and $\mathbb{G}(1,5)\cap H^3$. In the second case we find for example:

Theorem 3.5 *For $N = 2n-1 \geq 5$ the automorphism group of $\mathbb{G}(1,N)\cap H^2$ has $\mathrm{SL}(2,\mathbb{C})^n/\{1,-1\}$ as a normal subgroup and the quotient group is isomorphic to the permutation group $\mathrm{S}(3)$ for $n = 3$, to $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/2\mathbb{Z}$ for $n = 4$, and trivial otherwise.*

We believe that apart from trivial cases these are the only general linear sections where automorphism groups of positive dimension appear. Extensive computer checks seem to confirm this.

In particular we prove that the automorphism groups of $\mathbb{G}(1,2n)\cap H$, $\mathbb{G}(1,4)\cap H^2$, $\mathbb{G}(1,5)\cap H^2$, $\mathbb{G}(1,6)\cap H^2$, and $\mathbb{G}(1,4)\cap H^3$ are quasihomogeneous – those of $\mathbb{G}(1,2n-1)\cap H$ and $\mathbb{G}(1,3)\cap H^2$ are even homogenous – whereas all others are not.

As to our methods, in our proofs the rich geometry of the Grassmannian plays a decisive role. Otherwise, we mainly use well known tools like multilinear algebra, Lefschetz theorems, vanishing theorems etc.

We are indebted to E. Opdam and A. Pasquale for useful remarks. The first author also thanks the Stieltjes Institut of Leiden University for financial support.

1 Preliminary

The Grassmannian $\mathbb{G}(1,N)$ of lines in $\mathbb{P}_N$ is embedded by way of the Plücker embedding into $\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$

$$\begin{array}{ccc} \mathbb{G}(1,N) & \longrightarrow & \mathbb{P}(\bigwedge^2\mathbb{C}^{N+1}) \\ \mathrm{span}\,\{v,w\} & \longmapsto & \mathbb{P}(v\wedge w). \end{array}$$

We denote by H^l an l-codimensional linear subspace of $\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$. Roth [R] examined the geometry of the general linear sections of the Grassmannians and found

Theorem 1.1 *For a general H^l with $0 \leq l \leq 1/2\dim\mathbb{G}(1,N) = N-1$ the intersection with the Grassmannians, $\mathbb{G}(1,N)\cap H^l$, is rational.*

In this article we continue this study by describing the automorphism groups of these sections. As for the notation, given a subvariety Y of a variety X we define $\mathrm{Aut}(Y,X)$ to be the automorphisms of X that induce automorphisms of Y, i.e.

$$\mathrm{Aut}(Y,X) = \{\varphi\in\mathrm{Aut}(X) \mid \varphi(Y)\subseteq Y\}.$$

Recall that the automorphism group of the Grassmannian itself is computed in two steps, see e.g. [H, 10.19]. First one shows that all automorphisms are induced by automorphisms of $\mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})$, i.e.

$$\mathrm{Aut}(\mathbb{G}(1,N)) \cong \mathrm{Aut}(\mathbb{G}(1,N), \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})).$$

Then one proves that for $N \neq 3$ the right hand side group is isomorphic to $\mathbb{P}\mathrm{GL}(N+1,\mathbb{C})$ via

$$\begin{array}{ccc} \mathbb{P}\mathrm{GL}(N+1,\mathbb{C}) & \longrightarrow & \mathrm{Aut}(\mathbb{G}(1,N), \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})) \\ \mathbb{P}(T) & \longmapsto & (\mathbb{P}(\sum v_i \wedge w_i) \mapsto \mathbb{P}(\sum T v_i \wedge T w_i)) \,. \end{array}$$

For the linear sections of the Grassmannians we follow the same outline. The first step is the following theorem and its corollary; the second step will be done separately for the different cases in the next sections.

Theorem 1.2 *For $N \geq 4$ and a general linear subspace $H^l \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})$ of codimension $l \leq 2N-5$ the linear section $\mathbb{G}(1,N) \cap H^l$ spans H^l and its automorphisms are induced by automorphisms of H^l, i.e.*

$$\mathrm{Aut}(\mathbb{G}(1,N) \cap H^l) = \mathrm{Aut}(\mathbb{G}(1,N) \cap H^l, H^l)$$

Proof. We will abbreviate $\mathbb{G}(1,N)$ by $\mathbb{G}$. We want to prove that $\mathbb{G} \cap H^l$ spans H^l, i.e. for $\mathbb{G} \cap H^l \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})$

$$\mathrm{h}^0(\mathbb{G} \cap H^l, \mathcal{O}(H)) = \dim \bigwedge^2 \mathbb{C}^{N+1} - l.$$

This is known for $l = 0$. For $l \geq 1$ we take the long exact sequence associated to the restriction sequence tensored by $\mathcal{O}(H)$

$$\begin{aligned} 0 &\to \mathrm{H}^0(\mathbb{G} \cap H^{l-1}, \mathcal{O}) = \mathbb{C} \to \mathrm{H}^0(\mathbb{G} \cap \mathrm{H}^{l-1}, \mathcal{O}(H)) \to \mathrm{H}^0(\mathbb{G} \cap \mathrm{H}^l, \mathcal{O}(H)) \to \\ &\to \mathrm{H}^1(\mathbb{G} \cap \mathrm{H}^{l-1}, \mathcal{O}) = 0. \end{aligned}$$

Looking at the dimensions we get

$$\mathrm{h}^0(\mathbb{G} \cap H^l, \mathcal{O}(H)) = \mathrm{h}^0(\mathbb{G} \cap H^{l-1}, \mathcal{O}(H)) - 1,$$

and the claim follows by induction.
Now we show that all automorphisms of $\mathbb{G} \cap H^l$ are induced by automorphisms of H^l. This follows if we can show that all divisors of $\mathbb{G} \cap H^l$ are induced by divisors of H^l, i.e.

$$\mathrm{Pic}(\mathbb{G} \cap H^l) = \mathrm{Pic}(H^l) = \mathbb{Z} \cdot H,$$

because then the projective embedding of $\mathbb{G} \cap H^l$ given by the sections in the line bundle $\mathcal{O}(H)$ is equivariant for all automorphisms of $\mathbb{G} \cap H^l$
To see this, note that by the Lefschetz hyperplane section theorem

$$\mathbb{Z} \cdot H = \mathrm{H}^2(\mathbb{G}, \mathbb{Z}) = \mathrm{H}^2(\mathbb{G} \cap H, \mathbb{Z}) = \ldots = \mathrm{H}^2(\mathbb{G} \cap H^l, \mathbb{Z})$$

for $0 \le l \le 2N-5$. From the exponential sequence

$$0 \to \mathbb{Z}_{\mathbb{G}\cap H^l} \to \mathcal{O}_{\mathbb{G}\cap H^l} \to \mathcal{O}^*_{\mathbb{G}\cap H^l} \to 0$$

we get as a part of the associated long exact sequence

$$\ldots \to \mathrm{H}^1(\mathbb{G}\cap H^l, \mathcal{O}) \to \mathrm{H}^1(\mathbb{G}\cap H^l, \mathcal{O}^*) \to \mathrm{H}^2(\mathbb{G}\cap H^l, \mathbb{Z}) = \mathbb{Z}\cdot H \to 0$$

and therefore

$$\mathrm{Pic}(\mathbb{G}\cap H^l) = \mathrm{H}^1(\mathbb{G}\cap H^l, \mathcal{O}^*) = \mathbb{Z}\cdot H$$

as soon as we know that $\mathrm{H}^1(\mathbb{G}\cap H^l, \mathcal{O}) = 0$.

This is well known for $l = 0$. For $l \ge 1$ we look at the restriction sequence

$$0 \to \mathcal{O}_{\mathbb{G}\cap H^{l-1}}(-H) \to \mathcal{O}_{\mathbb{G}\cap H^{l-1}} \to \mathcal{O}_{\mathbb{G}\cap H^l} \to 0$$

and take its associated long exact sequence

$$\begin{aligned}\ldots &\to \mathrm{H}^1(\mathbb{G}\cap H^{l-1}, \mathcal{O}(-H)) \to \mathrm{H}^1(\mathbb{G}\cap H^{l-1}, \mathcal{O}) \to \mathrm{H}^1(\mathbb{G}\cap H^l, \mathcal{O}) \to \\ &\to \mathrm{H}^2(\mathbb{G}\cap H^{l-1}, \mathcal{O}(-H)) \to \ldots\end{aligned}$$

The right and left cohomology groups are trivial for $l \le 2N-4$ by Kodaira's vanishing theorem, so

$$0 = \mathrm{H}^1(\mathbb{G}, \mathcal{O}) = \mathrm{H}^1(\mathbb{G}\cap H, \mathcal{O}) = \ldots = \mathrm{H}^1(\mathbb{G}\cap H^l, \mathcal{O}). \qquad \square$$

Corollary 1.3 *For $N \ge 4$ and a general linear subspace $H^l \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})$ of codimension $l \le N-2$*

$$\mathrm{Aut}(\mathbb{G}(1,N)\cap H^l) = \mathrm{Aut}(\mathbb{G}(1,N)\cap H^l, \mathbb{P}(\textstyle\bigwedge^2\mathbb{C}^{N+1})) \cap \mathrm{Aut}(H^l, \mathbb{P}(\textstyle\bigwedge^2\mathbb{C}^{N+1})).$$

This is also true for $\mathbb{G}(1,4)\cap H^3$.

Proof. The special case of $\mathbb{G}(1,4)\cap H^3$ will be dealt with in Section 7.
In view of the theorem we need only to show that an automorphism of $\mathbb{G}(1,N)\cap H^l$ can be extended to an automorphism of $\mathbb{G}(1,N)$, which is always linear and fixes the linear subspace H^l because $\mathbb{G}(1,N)\cap H^l$ spans H^l. To do so we will study the linear subspaces of $\mathbb{G}(1,N)\cap H^l$.
The linear subspaces of the Grassmannian $\mathbb{G}(1,N)$ are the following Schubert cycles:

1. Let $p \in \mathbb{P}_N$ be a point and $K \subset \mathbb{P}_N$ a linear subspace of dimension k that contains the point p then

   $$\{L \in \mathbb{G}(1,N) \mid p \in L \subseteq K\} \subset \mathbb{G}(1,N)$$

   is a $(k-1)$-dimensional subspace and these form a

   $$\dim \mathbb{P}_N + \dim \mathbb{G}(k-1, N-1) = N + k(N-k)$$

   dimensional family F_{k-1}.

2. Let $E\subset\mathbb{P}_N$ be a plane then

$$\{L\in\mathbb{G}(1,N)\mid L\subseteq K\}\subset\mathbb{G}(1,N)$$

is a plane in $\mathbb{G}(1,N)$ and these form a

$$\dim\mathbb{G}(1,N)=3(N-2)=3N-6$$

dimensional family F_2'.

Let $d=\dim\bigwedge^2\mathbb{C}^{N+1}-1$. For the variety of 2-planes F_2 (and analogously for F_2') we consider the incidence correspondence

$$\{(E,h)\in F_2\times\mathbb{G}(d-l,d)\mid E\subset h\}\subset F_2\times\mathbb{G}(d-l,d).$$

The fibre above any point $h\in\mathbb{G}(d-l,d)$ is exactly the set of 2-planes of F_2 on $\mathbb{G}(1,N)\cap h$, which we denote by $\widetilde{F}_2$. From this correspondence we see immediately that for general H^l either $\widetilde{F}_2$ is empty or $\widetilde{F}_2$ consists of some components of the same dimension (Stein factorisation). The same argument applied to F_2' yields $\widetilde{F}_2'$. Since for $N\geq 4$ $\dim F_2\neq\dim F_2'$ we thus find that if $\widetilde{F}_2$ and $\widetilde{F}_2'$ are not both empty, then $\dim\widetilde{F}_2\neq\dim\widetilde{F}_2'$.

Now any automorphism of $\mathbb{G}(1,N)\cap H^l$ is linear, so we conclude that any automorphism of $\mathbb{G}(1,N)\cap H^l$ transforms $\widetilde{F}_2$ resp. $\widetilde{F}_2'$ onto itself. For dimension $\neq 2$ there are only linear spaces of type 1, so we can state:

Any automorphism φ of $\mathbb{G}(1,N)\cap H^l$ permutes the linear spaces of type 1.

Next, let $p\in\mathbb{P}_N$ and L_p the $(N-1)$-dimensional linear subspace of $\mathbb{G}(1,n)$ consisting of the lines through p. From $l\leq N-2$ we see $e=\dim(L_p\cap H^l)\geq 1$. Since H^l is general, for almost all points of $\mathbb{P}_N$ this dimension is exactly e. Now, and this is the crucial remark, $\varphi(L_p\cap H^l)$ being of type 1, is contained in exactly one L_q, $q\in\mathbb{P}_N$. Attaching q to p we obtain for almost all $p\in\mathbb{P}_N$ a map to $\mathbb{P}_N$. If the dimension of $L_{p_0}\cap H^l$ is bigger than the minimum, then this map can still be defined at p_0 in exactly the same way. We claim that this is continuous at p_0. This can be seen by considering a general sequence of points on $\mathbb{P}_N$, say $p_1,p_2,\ldots$, converging to p_0 with $\dim(L_{p_i}\cap H^l)$ minimal and applying the crucial remark twice. The map from $\mathbb{P}_N$ to $\mathbb{P}_N$ thus obtained has an inverse and is holomorphic, so it is a linear automorphism of $\mathbb{P}_N$. This automorphism induces an automorphism of $\mathbb{G}(1,N)$ for which it is easily verified that it coincides with φ on $\mathbb{G}(1,N)\cap H^l$. $\square$

It is tempting to assume that the groups $\operatorname{Aut}(\mathbb{G}(1,N),\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1}))$ and $\operatorname{Aut}(H^l,\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1}))$ in $\operatorname{Aut}(\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1}))$ intersect transversally. Then the dimension of $\operatorname{Aut}(\mathbb{G}(1,N)\cap H^l)$ could be computed as

$$\begin{aligned}\dim\operatorname{Aut}(\mathbb{G}(1,N)\cap H^l)&=\dim\operatorname{Aut}(\mathbb{G}(1,N))-\operatorname{codim}\operatorname{Aut}(H^l,\mathbb{P}(\textstyle\bigwedge^2\mathbb{C}^{N+1}))\\&=(N+1)^2-1-l\left(\binom{N+1}{2}-l\right).\end{aligned}$$

And we would find the following non-finite groups:

$$\begin{aligned}
\dim \operatorname{Aut}(\mathbb{G}(1,N)\cap H) &= (N^2+3N+2)/2\\
\dim \operatorname{Aut}(\mathbb{G}(1,N)\cap H^2) &= N+4\\
\dim \operatorname{Aut}(\mathbb{G}(1,4)\cap H^3) &= 3.
\end{aligned}$$

Unfortunately, the intersection is not always transversal. Our computation of the automorphism groups will show the following dimensions for $N \geq 4$:

$$\begin{aligned}
\dim \operatorname{Aut}(\mathbb{G}(1,N)\cap H) &= (N^2+3N+2)/2\\
\dim \operatorname{Aut}(\mathbb{G}(1,N)\cap H^2) &= \begin{cases} N+4 & \text{for } N \text{ even}\\ 3(N+1)/2 & \text{for } N \text{ odd}\end{cases}\\
\dim \operatorname{Aut}(\mathbb{G}(1,4)\cap H^3) &= 3\\
\dim \operatorname{Aut}(\mathbb{G}(1,5)\cap H^3) &= 1.
\end{aligned}$$

We conjecture that these are the only non-finite groups. For $N+2 \leq l$ the canonical bundle $K = \mathcal{O}(-N-1+l)$ is positive on $\mathbb{G}(1,N)\cap H^l$, and this conjecture can be proved by Serre's duality theorem and Kodaira's vanishing theorem:

$$\dim \operatorname{Aut}(\mathbb{G}\cap H^l) = h^0(\mathbb{G}\cap H^l, \Theta) = h^{2N-2-l}(\mathbb{G}\cap H^l, K\Omega^1) = 0$$

A proof for the remaining cases $3 \leq l \leq N+1$ seems difficult. For $N \leq 10$ and all l we verified the conjecture for the automorphisms induced by automorphisms of the Grassmannian by computer computations.

With this Theorem and its Corollary our task of determining the automorphisms of $\mathbb{G}(1,N)\cap H^l$ has been immensely simplified. All we need to do is to find the projective transformations of $\operatorname{Aut}(\mathbb{G}(1,N), \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})) = \mathbb{P}\mathrm{GL}(N+1,\mathbb{C})$ such that their induced action on $\mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})^*$ preserves H^l. To express this in algebraic terms we identify $(\bigwedge^2 \mathbb{C}^{N+1})^*$ with $\bigwedge^2 (\mathbb{C}^{N+1})^*$. If a particular basis of $\mathbb{C}^{N+1}$ is chosen, $\bigwedge^2 (\mathbb{C}^{N+1})^*$ as antisymmetric forms on $\mathbb{C}^{N+1}$ can also be identified with the antisymmetric matrices of size $N+1$. In concrete terms, if $(e_0,\ldots,e_N)$ is a basis of $\mathbb{C}^{N+1}$ and $E_{ij} \in \mathrm{M}(N+1,\mathbb{C})$ the matrix, which has a 1 in the position (i,j) but is otherwise zero, then

$$\begin{array}{ccc}
\left(\bigwedge^2 \mathbb{C}^{N+1}\right)^* & \longrightarrow & \operatorname{Antisym}(N+1,\mathbb{C})\\
\sum_{i,j} \lambda_{ij} (e_i \wedge e_j)^* & \longmapsto & \frac{1}{2}\sum_{i,j} \lambda_{ij}(E_{ij}-E_{ji}).
\end{array}$$

In these terms a line $l = p \wedge q \in \mathbb{G}(1,N)$ is in the hyperplane $H \in \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})$ iff for a corresponding antisymmetric matrix $A \in \operatorname{Antisym}(N+1,\mathbb{C})$ with $\mathbb{P}(A) = H$ we have ${}^t\!pAq = 0$.

Further, the action of $\mathbb{P}\mathrm{GL}(N+1,\mathbb{C})$ on $\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$, which was given for $\mathbb{P}(T)\in\mathbb{P}\mathrm{GL}(N+1,\mathbb{C})$ by

$$\begin{array}{ccc} \mathbb{P}\left(\bigwedge^2\mathbb{C}^{N+1}\right) & \longrightarrow & \mathbb{P}\left(\bigwedge^2\mathbb{C}^{N+1}\right) \\ \mathbb{P}(\sum v_i\wedge w_i) & \longmapsto & \mathbb{P}(\sum Tv_i\wedge Tw_i), \end{array}$$

induces the following action on the dual space

$$\begin{array}{ccc} \mathbb{P}(\mathrm{Antisym}(N+1,\mathbb{C})) & \longrightarrow & \mathbb{P}(\mathrm{Antisym}(N+1,\mathbb{C})) \\ \mathbb{P}(A) & \longmapsto & \mathbb{P}({}^tT^{-1}AT^{-1}). \end{array}$$

Hence an l-codimensional linear subspace $H^l\subseteq\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$ which is dually given by $\mathbb{P}(\mathrm{span}\{A_1,\dots,A_l\})$ is preserved under T iff every hyperplane containing H^l is mapped to another hyperplane containing H^l, i.e.

$$\begin{array}{ll} {}^tT^{-1}\left(\sum\lambda_iA_i\right)T^{-1}\in\mathrm{span}\{A_1,\dots,A_l\} & \text{for all } \lambda_i\in\mathbb{C} \\ \Longleftrightarrow {}^tT^{-1}A_iT^{-1}\in\mathrm{span}\{A_1,\dots,A_l\} & \text{for } i=1\dots l. \end{array}$$

We conclude

Corollary 1.4 *For $N\geq 4$, $0\leq l\leq N-2$ and a general $H^l\subset\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$ given by $\mathbb{P}(\mathrm{span}\{A_1,\dots,A_l\})\subset\mathbb{P}(\mathrm{Antisym}(N+1,\mathbb{C}))$ the automorphism group of $\mathbb{G}(1,N)\cap H^l$ is*

$$\{\mathbb{P}(T)\in\mathbb{P}\mathrm{GL}(N+1,\mathbb{C})\mid {}^tT^{-1}A_iT^{-1}\in\mathrm{span}\{A_1,\dots,A_l\}\,\forall i\}.$$

In the following sections we will compute the automorphism groups using this Corollary. In the course of the computations we will use geometric arguments for which it is essential to know if a hyperplane $H\subset\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$ is tangent to $\mathbb{G}(1,N)$ or not. We recall the basic facts together with their short proofs.

Proposition 1.5 *For any line $l_0\in\mathbb{G}(1,N)$ the Schubert cycle*

$$\sigma:=\{l\in\mathbb{G}(1,N)\mid l\cap l_0\neq\emptyset\}\subseteq\mathbb{G}(1,N)$$

lies inside the tangent space $\mathbb{T}_{l_0}\mathbb{G}(1,N)\subseteq\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$ and spans it.

Proof. Let $l\in\sigma$, $p\in l\cap l_0$, $q\in l_0\setminus\{p\}$ and $r\in l\setminus\{p\}$ then

$$\begin{array}{ccc} \mathbb{C} & \longrightarrow & \mathbb{G}(1,N) \\ \lambda & \longmapsto & p\wedge(q+\lambda r) \end{array}$$

is a line in $\sigma\subset\mathbb{G}(1,N)$ through l_0 and l. Therefore it is contained in the tangent space $\mathbb{T}_{l_0}\mathbb{G}(1,N)$, in particular $l\in\mathbb{T}_{l_0}\mathbb{G}(1,N)$.
We choose a basis $(e_0,\dots,e_N)$ of $\mathbb{C}^{N+1}$ such that $l_0=\mathbb{P}(e_0\wedge e_1)$. The $2N-1$ points $\mathbb{P}(e_0\wedge e_1)$, $\mathbb{P}(e_0\wedge e_i)$, $\mathbb{P}(e_1\wedge e_i)$ for $i=2\dots N$ lie in $\sigma\subset\mathbb{T}_{l_0}\mathbb{G}(1,N)$ and are projectively independent, hence they span $\mathbb{T}_{l_0}\mathbb{G}(1,N)$. $\square$

Corollary 1.6 *Let $H = \mathbb{P}(A) \in \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})^*$ be a hyperplane and $l_0 \in \mathbb{G}(1,N)$ a line then*

$$\mathbb{T}_{l_0}\mathbb{G}(1,N) \subseteq H \Longleftrightarrow l_0 \subseteq \ker A.$$

Proof. By the Proposition $\mathbb{T}_{l_0}\mathbb{G}(1,N) \subseteq H$ is equivalent to $\sigma \subseteq H$. If we use the same basis of $\mathbb{C}^{N+1}$ as in the proof of the proposition, this means that

$$\begin{aligned}
&\mathbb{P}((\lambda e_0 + \mu e_1) \wedge v) \in H \quad \text{for all } (\lambda : \mu) \in \mathbb{P}_1,\ v \in \mathbb{C}^{N+1} \\
&\Longleftrightarrow {}^t(\lambda e_0 + \mu e_1) A v = 0 \quad \text{for all } (\lambda : \mu) \in \mathbb{P}_1,\ v \in \mathbb{C}^{N+1} \\
&\Longleftrightarrow l_0 \subseteq \ker A.
\end{aligned}$$

□

Corollary 1.7 *The dual variety $\mathbb{G}(1,N)^* \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})^*$ of the Grassmannian variety $\mathbb{G}(1,N)$ consists of matrices of corank ≥ 2 for N odd resp. corank ≥ 3 for N even.*

For N odd it is an irreducible hypersurface of degree $(N+1)/2$; for N even it is a 3-codimensional subvariety.

Proof. By the last corollary $H = \mathbb{P}(A) \in \mathbb{P}(\bigwedge^2 \mathbb{C}^{N+1})^*$ is tangential to $\mathbb{G}(1,N)$ iff $\operatorname{corank} A \geq 2$. Recall that an antisymmetric matrix has even rank. So, for N odd the matrix $A \in \operatorname{Antisym}(N+1, \mathbb{C})$ has corank ≥ 2 iff $\det A = 0$. But again since A is antisymmetric, $\det A$ is the square of the irreducible Pfaffian polynomial $\operatorname{Pf} A$ [B, 5.2], which therefore defines $\mathbb{G}(1,N)^*$.

For N even $\operatorname{corank} A \geq 2$ is equivalent to $\operatorname{corank} A \geq 3$. We compute the dimension of $\mathbb{G}(1,N)^*$ following Mumford [M] and find

$$\begin{aligned}
\dim \begin{pmatrix} \text{space of } A \text{ with} \\ \dim\ker A = 3 \end{pmatrix} &= \dim \mathrm{G}(3, N+1) + \dim \textstyle\bigwedge^2 \mathbb{C}^{N+1}/\mathbb{C}^3 \\
&= 3(N-2) + (N-2)(N-3)/2 \\
&= (N^2 + N - 6)/2
\end{aligned}$$

$$\Longrightarrow \operatorname{codim} \mathbb{G}(1,N)^* = (N+1)N/2 - (N^2+N-6)/2 = 3.$$

□

2 $\mathbb{G}(1, 2n-1) \cap H$

Let the hyperplane $H \in \mathbb{P}(\bigwedge^2 \mathbb{C}^{2n})$ be given by an element $A \in (\bigwedge^2 \mathbb{C}^{2n})^*$, which we identify with its corresponding antisymmetric matrix. If H is general, A will be a matrix of full rank. This may be taken as the definition of a general H. We will assume from now on that H is general.

The line system $\mathbb{G}(1, 2n-1) \cap H$ in $\mathbb{P}_{2n-1}$ does not lead to obvious special points in the $\mathbb{P}_{2n-1}$. Through every point $p \in \mathbb{P}_{2n-1}$ passes a $\mathbb{P}_{2n-3}$ of lines, namely

$$p \wedge q \in \mathbb{G}(1, 2n-1) \text{ with } q \in \ker {}^t p A.$$

For $n\geq 3$ we can compute the automorphism group of $\mathbb{G}(1,2n-1)\cap H$ with the help of Theorem 1.2 and its Corollaries. It consists of elements $\mathbb{P}(T)\in \mathbb{P}\mathrm{GL}(2n,\mathbb{C})=\mathrm{Aut}(\mathbb{G}(1,2n-1))$ such that $\mathbb{P}(T)$ as an element of $\mathbb{P}\mathrm{GL}(\bigwedge^2\mathbb{C}^{2n})$ preserves H, i.e.

$$ {}^tT^{-1}AT^{-1}=\lambda A \quad \text{for suitable } \lambda\in\mathbb{C}^*. $$

We may choose coordinates on $\mathbb{P}_{2n-1}$ such that

$$ A=\begin{pmatrix} 0 & -\mathrm{E}_n \\ \mathrm{E}_n & 0 \end{pmatrix}. $$

Then by definition

$$ \mathrm{Sp}(2n,\mathbb{C})=\{T\in\mathrm{GL}(2n,\mathbb{C})\mid {}^tT^{-1}AT^{-1}=A\}, $$

and we have an isomorphism

$$ \begin{array}{ccc} \{T\in\mathrm{GL}(2n,\mathbb{C})\mid \exists\lambda_T\in\mathbb{C}^*: {}^tT^{-1}AT^{-1}=\lambda_T A\}/\mathbb{C}^* & \longrightarrow & \mathrm{Sp}(2n,\mathbb{C})/\{1,-1\} \\ \mathbb{C}^*\cdot T & \longmapsto & \pm\frac{1}{\sqrt{\lambda_T}}T. \end{array} $$

Therefore we see

Proposition 2.1 *The automorphism group of $\mathbb{G}(1,2n-1)\cap H$ for a general $H\subset\mathbb{P}(\bigwedge^2\mathbb{C}^{N+1})$ is $\mathrm{Sp}(2n,\mathbb{C})/\{1,-1\}$. Its action on $\mathbb{G}(1,2n-1)\cap H$ is homogeneous.*

Proof. The missing case of $\mathbb{G}(1,3)\cap H$ can be found in [FH, p. 278]. The transitivity of the action follows from Witt's theorem [Br, 12.31]. □

3 $\mathbb{G}(1,2n-1)\cap H^2$

A 2-codimensional linear subspace $L=H^2$ of $\mathbb{P}(\bigwedge^2\mathbb{C}^{2n})$ can be thought of as the pencil of hyperplanes containing it. So it gives a line $L^*=\mathbb{P}(\lambda A-\mu B)\subset \mathbb{P}(\bigwedge^2\mathbb{C}^{2n})^*$. We identify again $(\bigwedge^2\mathbb{C}^{2n})^*$ with the antisymmetric matrices of size $2n$. The line L^* intersects the dual Grassmannian $\mathbb{G}(1,2n-1)^*$, which consists of antisymmetric matrices of rank $\leq 2n-2$ and is a hypersurface of degree n by Corollary 1.7, in at most n points. For the moment a line L^*, and hence L, will be called general if it has n points of intersection, $H_i=\mathbb{P}(\lambda_iA-\mu_iB)\in L^*$, $i=1\dots n$, with the dual Grassmannian. These hyperplanes H_i are tangent to the Grassmannian $\mathbb{G}(1,2n-1)$ at the points $l_i:=\ker(\lambda_iA-\mu_iB)\in\mathbb{G}(1,2n-1)$ by Corollary 1.6. Therefore we get n exceptional lines $l_1,\dots,l_n$ in $\mathbb{P}_{2n-1}$.

The intersection of the Grassmannian with its tangent hyperplane H_i contains all lines that intersect l_i, because these lines are already contained in the intersection $\mathbb{G}(1,2n-1)\cap\mathbb{T}_{l_i}\mathbb{G}(1,2n-1)$ by Proposition 1.5.

So, any line through a point $p \in l_i$ will be in the subspace $L \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^{2n})$ as soon as it is contained in any other hyperplane $H \in L^* \setminus \{H_i\}$. This gives one linear restriction to lines through p, so that there is at least a $\mathbb{P}_{2n-3}$ of lines through the points of the lines l_i. In contrast, through a general point of $\mathbb{P}_{2n-1} \setminus \bigcup l_i$ there is only a $\mathbb{P}_{2n-4}$ of lines. In fact, we have

Proposition 3.1 *The points of the lines $l_1, \dots, l_n$ are characterized by the property that through each of them passes a $\mathbb{P}_{2n-3}$ of lines, i.e.*

$$\left\{ p \in \mathbb{P}_{2n-1} \,\middle|\, \begin{array}{l} \text{through } p \text{ passes a } \mathbb{P}_{2n-3} \\ \text{of lines of } \mathbb{G}(1,2n-1) \cap L \end{array} \right\} = \bigcup l_i.$$

Furthermore, the lines $l_1, \dots, l_n$ span the whole $\mathbb{P}_{2n-1}$.

This may easily be seen if we write the pencil of hyperplanes L^* in its normal form.

Proposition 3.2 (Donagi[D]) *Given a pencil of hyperplanes $L^* = \mathbb{P}(\lambda A - \mu B) \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^{2n})^*$ such that the line L^* intersects the Pfaffian hypersurface in n different points. Then there is a basis of $\mathbb{C}^{2n}$ such that*

$$A = \begin{pmatrix} J & & 0 \\ & \ddots & \\ 0 & & J \end{pmatrix} \text{ and } B = \begin{pmatrix} \lambda_1 J & & 0 \\ & \ddots & \\ 0 & & \lambda_n J \end{pmatrix} \text{ with } J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The points $(\lambda_1 : 1), \dots, (\lambda_n : 1) \in \mathbb{P}_1 \cong L^*$ are unique up to a projective transformation of $\mathbb{P}_1$.

Proof of Proposition 3.1. The hyperplane $H_i = \mathbb{P}(\lambda_i A - \mu_i B)$ has, written as an antisymmetric matrix, the kernel $l_i = \operatorname{span}\{e_{2i-1}, e_{2i}\}$ which means it is tangent to $\mathbb{G}(1, 2n-1)$ at l_i. All lines of $\mathbb{G}(1, 2n-1) \cap L$ through the point p are given by $p \wedge q$ with ${}^t\!pAq = {}^t\!pBq = 0$. In order to have a $\mathbb{P}_{2n-2}$ of lines through p, the linear forms ${}^t\!pA$ and ${}^t\!pB$ must be linear dependent, i.e. there are $\lambda, \mu \in \mathbb{C}$ with

$$0 = \lambda {}^t\!pA - \mu {}^t\!pB = {}^t\!p(\lambda A - \mu B).$$

Therefore p is in the kernel of a matrix of the pencil, but these kernels are the lines $l_1, \dots, l_n$, so p is contained in one of them. □

Knowing the exceptional lines $l_1, \dots, l_n$, one can immediately give some lines which are in the line system.

Proposition 3.3 *Any line in $\mathbb{P}_{2n-1}$ which intersects two exceptional lines is an element of the line system $\mathbb{G}(1, 2n-1) \cap L$.*

The exceptional lines themselves are not in the line system.

Proof. If a line l intersects l_i and l_j, it lies – as a point of the Grassmannian $\mathbb{G}(1,2n-1)$ – in H_i and H_j, hence in $L = H_i \cap H_j$.

Assume that the exceptional line l_1 is an element of $\mathbb{G}(1,2n-1)$. By Proposition 3.1 the lines through a point $p \in l_1$ sweep out a hyperplane. This hyperplane contains the line l_1 by assumption and the other exceptional lines $l_2,\dots,l_n$ by the first part of this proposition. But this contradicts the second statement of Proposition 3.1. □

Remark 3.4 *From Proposition 3.2 we also see that any position of the* n *points of the line* L^* *is possible. In particular, we may call a line general if the position of the points is general in the sense needed below.*

Using this geometric description we can determine the automorphisms of $\mathbb{G}(1,2n-1)\cap L$. For the moment we restrict ourselves to $n \geq 3$ in order to be able to use Theorem 1.2. By this Theorem and its Corollaries we can view an automorphism of $\mathbb{G}(1,2n-1)\cap L$ as an element $\mathbb{P}(T)$ of $\mathbb{P}\mathrm{GL}(2n,\mathbb{C})$. To make the notation simpler, we will write only T for $\mathbb{P}(T)$ if no confusion can result. Since the points of the exceptional lines are characterized by the property of Proposition 3.1, T must map the union of the lines $l_i \subset \mathbb{P}_{2n-1}$ onto itself. Permutations of the lines may occur, but – as we will presently see – not all permutations are possible.

If we view the automorphism T as an element of $\mathrm{Aut}(L,\mathbb{P}(\bigwedge^2\mathbb{C}^{2n}))$, it interchanges the hyperplanes containing L, i.e. it induces a projective transformation of the line $L^* \subset \mathbb{P}(\bigwedge^2\mathbb{C}^{2n})$. Naturally, the transformation of L^* must preserve the union of points of intersection of L^* with the dual Grassmannian, which determine the lines l_i. Now, if a transformation of $\mathbb{P}_{2n-1}$ permutes the lines l_i, then the induced transformation of L^* must permute the corresponding points of L^* in the same way.

Since not every permutation of four or more points on a line can be induced by a projective transformation, not all permutations are possible. In fact, if the points are in general position, we get the following subgroups of the permutation groups:

n	subgroup of $\mathrm{S}(n)$
3	$\mathrm{S}(3)$
4	$\{(1\,2\,3\,4),(2\,1\,4\,3),(3\,4\,1\,2),(4\,3\,2\,1)\}\cong \mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/2\mathbb{Z}$
≥ 5	$\{\mathrm{id}\}$

On the other hand, any permutation $\sigma \in \mathrm{S}(n)$ of the points on L^* that is induced by a projective transformation φ of L^* can be induced by an automorphism of $\mathbb{G}(1,2n-1)\cap L$. To see this, let us write L^* in its normal form and define $T \in \mathrm{GL}(2n,\mathbb{C})$ as

$$T(e_{2i}) := e_{2\sigma(i)} \quad \text{and} \quad T(e_{2i-1}) := e_{2\sigma(i)-1}.$$

This transformation permutes the lines in the prescribed way, and as an automorphism of $\mathbb{P}(\bigwedge^2 \mathbb{C}^{2n})$ it fixes L since the transformed line L^* is

$$
\begin{aligned}
&{}^tT^{-1}\left(\lambda\begin{pmatrix} J & & 0 \\ & \ddots & \\ 0 & & J\end{pmatrix} - \mu\begin{pmatrix} \lambda_1 J & & 0 \\ & \ddots & \\ 0 & & \lambda_n J\end{pmatrix}\right)T^{-1} \\
&= \lambda\begin{pmatrix} J & & 0 \\ & \ddots & \\ 0 & & J\end{pmatrix} - \mu\begin{pmatrix} \lambda_{\sigma^{-1}(1)} J & & 0 \\ & \ddots & \\ 0 & & \lambda_{\sigma^{-1}(n)} J\end{pmatrix}.
\end{aligned}
$$

Changing the parametrisation of the line by φ we get back the old parametrisation of the line L^* by the definition of φ. So this T is an automorphism of $\mathbb{G}(1, 2n-1) \cap L$ that induces the permutation of lines we started with.

Now we can restrict our attention to transformations that do not permute the lines since we can obtain every permutation by composing with one of the transformations from above. A transformation leaving all the lines individually fixed has the form

$$
T = \begin{pmatrix} t_1 & & 0 \\ & \ddots & \\ 0 & & t_n\end{pmatrix} \quad \text{with } t_1, \dots, t_n \in \mathrm{GL}(2, \mathbb{C}).
$$

This T will fix the line system $\mathbb{G}(1, 2n-1) \cap L$ in $\mathbb{P}_{2n-1}$ iff it preserves L^*, i.e. for all $\lambda, \mu \in \mathbb{C}$ there exists $\alpha, \beta \in \mathbb{C}$ such that

$$
{}^tT^{-1}(\lambda A - \mu B)T^{-1} = \alpha A + \beta B.
$$

It is sufficient to check this for $(\lambda, \mu) = (1, 0)$ and $(0, -1)$. Since

$$
\begin{aligned}
{}^tT^{-1}AT^{-1} &= \begin{pmatrix} \det t_1^{-1} J & & 0 \\ & \ddots & \\ 0 & & \det t_n^{-1} J\end{pmatrix} \\
{}^tT^{-1}BT^{-1} &= \begin{pmatrix} \lambda_1 \det t_1^{-1} J & & 0 \\ & \ddots & \\ 0 & & \lambda_n \det t_n^{-1} J\end{pmatrix}
\end{aligned}
$$

this is equivalent to the question if there exist $\alpha, \beta, \gamma, \delta \in \mathbb{C}$ with

$$
\begin{aligned}
(\det t_1^{-1}, \dots, \det t_n^{-1}) &= \alpha(1, \dots, 1) + \beta(\lambda_1, \dots, \lambda_n) \\
(\lambda_1 \det t_1^{-1}, \dots, \lambda_n \det t_n^{-1}) &= \gamma(1, \dots, 1) + \delta(\lambda_1, \dots, \lambda_n).
\end{aligned}
$$

It follows

$$\begin{aligned} &-\gamma(1,\ldots,1)+(\alpha-\delta)(\lambda_1,\ldots,\lambda_n)+\beta(\lambda_1^2,\ldots,\lambda_n^2)=0\\ &\Longrightarrow \alpha=\delta,\ \beta=\gamma=0\\ &\Longrightarrow \det t_1=\ldots=\det t_n. \end{aligned}$$

We normalize by $\det t_1 = 1$, i.e. $t_1,\ldots,t_n \in \mathrm{SL}(2,\mathbb{C})$. Then only T and $-T \in \mathrm{GL}(2n,\mathbb{C})$ give the same element in $\mathbb{P}\mathrm{GL}(2n,\mathbb{C})$. So that as a group the automorphisms of $\mathbb{G}(1,2n-1)\cap L$ that do not permute the exceptional lines are isomorphic to $\mathrm{SL}(2,\mathbb{C})^n/\{1,-1\}$.

Altogether we get

Theorem 3.5 *For $n \geq 3$ the automorphism group of the intersection of $\mathbb{G}(1,2n-1)$ with a general 2-codimensional linear subspace of $\mathbb{P}(\bigwedge^2\mathbb{C}^{2n})$ has $\mathrm{SL}(2,\mathbb{C})^n/\{1,-1\}$ as a normal subgroup and the quotient group is isomorphic to the permutation group $\mathrm{S}(3)$ for $n=3$, to $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/2\mathbb{Z}$ for $n=4$, and trivial otherwise.*

The automorphism group is isomorphic to the subgroup of $\mathbb{P}\mathrm{GL}(2n,\mathbb{C})$ that consists of the elements

$$P_\sigma\cdot\begin{pmatrix} t_1 & & 0\\ & \ddots & \\ 0 & & t_n\end{pmatrix}\quad \textit{with } t_1,\ldots,t_n\in\mathrm{SL}(2,\mathbb{C})$$

where P_σ is the identity for $n\geq 5$ and otherwise defined by

$$\begin{aligned} P_\sigma(e_{2i}) &= e_{2\sigma(i)}\\ P_\sigma(e_{2i-1}) &= e_{2\sigma(i)-1}\end{aligned}$$

$$\textit{for } \sigma\in\begin{cases} \mathrm{S}(n) & \textit{if } n=3\\ \{(1\,2\,3\,4),(2\,1\,4\,3),(3\,4\,1\,2),(4\,3\,2\,1)\} & \textit{if } n=4.\end{cases}$$

For the sake of completeness we recall the classical case of $\mathbb{G}(1,3)\cap H^2$.

Remark 3.6 *The automorphism group of $\mathbb{G}(1,3)\cap H^2$ is an extension of $\mathbb{Z}/2\mathbb{Z}$ by $\mathbb{P}\mathrm{GL}(2,\mathbb{C})\times\mathbb{P}\mathrm{GL}(2,\mathbb{C})$. It acts homogeneously on $\mathbb{G}(1,3)\cap H^2$.*

Proof. The Grassmannian $\mathbb{G}(1,3)$ is a smooth quadric in $\mathbb{P}(\bigwedge^2\mathbb{C}^4)\cong\mathbb{P}_5$. Therefore $\mathbb{G}(1,3)\cap H^2$ is a smooth quadric in $\mathbb{P}_3$. Hence it is isomorphic to the Segre variety $\mathbb{P}_1\times\mathbb{P}_1$ in $\mathbb{P}_3$. The automorphism group of $\mathbb{P}_1\times\mathbb{P}_1$ is generated by $\mathbb{P}\mathrm{GL}(2,\mathbb{C})\times\mathbb{P}\mathrm{GL}(2,\mathbb{C})$ together with the automorphism that exchanges the $\mathbb{P}_1$s. All the automorphisms extend to $\mathbb{P}_3$. Obviously, the group acts transitively on $\mathbb{P}_1\times\mathbb{P}_1$. □

For the rest of this section we consider the question if the action of the other automorphism groups is quasihomogeous on the corresponding line system, i.e. if there is an open orbit.

This cannot be the case for $n \geq 7$ since then the dimension of the line system $\mathbb{G}(1, 2n-1) \cap H^2$, $2(2n-2)-2 = 4n-6$, is larger than the dimension of the automorphism group, $3n$.

For $n = 3$ the action is quasihomogeneous. To see that one can adjust the $(\lambda_1, \lambda_2, \lambda_3)$ in the normal form of the line system to $(1, 0, -1)$ by a projective transformation and compute the stabiliser of the line $(1:0:1:0:1:0) \wedge (1:1:1:-2:1:1)$ by hand or computer and see that it is 3-dimensional. So the dimension of its orbit is $3 \cdot 3 - 3 = 6$, which is just the dimension of the line system.

For $n = 4, 5, 6$ the group does not act quasihomogenously anymore. For this one computes again the dimension of the stabiliser of a general line. Since the group acts transitively on $\mathbb{P}_{2n-1} \setminus \bigcup L_i$, we may restrict our attention to lines through one of those points, e.g. $(1:0:\ldots:1:0)$. Using a computer one sees that the stabilizer of a general line through this point has again dimension 3. Hence the orbit has dimension $3n-3$, which is less then the dimension of the line system, $4n-6$.

4 $\mathbb{G}(1,5) \cap H^3$

Let $L = H^3 \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^6)$ be a general 3-codimensional subspace. With our usual identification of $(\bigwedge^2 \mathbb{C}^6)^*$ with the antisymmetric matrices Antisym$(6, \mathbb{C})$ its dual plane $L^* = \mathbb{P}(\lambda A + \mu B + vC) \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^6)^*$ intersects the dual Grassmannian $\mathbb{G}(1,5)^*$, which consists of matrices of rank ≤ 4 and is a hypersurface of degree 3 by Corollary 1.7, in an irreducible cubic C^*. By Corollary 1.6 a point $(\lambda : \mu : \nu) \in C^*$ corresponds to the hyperplane $h_{(\lambda:\mu:\nu)} = \mathbb{P}(\lambda A + \mu B + \nu C)$ that is tangent to the Grassmannian at the point

$$l_{(\lambda:\mu:\nu)} := \ker(\lambda A + \mu B + \nu C) \subset \mathbb{P}_5.$$

In analogy to the former case we have

Lemma 4.1

$$\left\{ p \in \mathbb{P}_5 \;\middle|\; \begin{array}{l} \textit{through } p \textit{ passes a } \mathbb{P}_2 \\ \textit{of lines of } \mathbb{G}(1,5) \cap L \end{array} \right\} = \bigcup_{(\lambda:\mu:\nu) \in C^*} l_{(\lambda:\mu:\nu)} \subset \mathbb{P}_5$$

Proof. Since by definition the lines in $\mathbb{G}(1,5) \cap L$ that contain p are $p \wedge q$ with ${}^t\!pAq = {}^t\!pBq = {}^t\!pCq = 0$, we see that

$$\begin{aligned} &\text{through } p \text{ passes at least a } \mathbb{P}_2 \text{ of lines of } \mathbb{G}(1,5) \cap L \\ \iff &\ {}^t\!pA, {}^t\!pB, {}^t\!pC \text{ are linear dependent} \\ \iff &\ \exists (\lambda : \mu : \nu) \in \mathbb{P}_2 \text{ with } {}^t\!p(\lambda A + \mu B + \nu C) = 0 \\ \iff &\ p \in \ker(\lambda A + \mu B + \nu C) = l_{(\lambda:\mu:\nu)}. \end{aligned}$$

We also note that there cannot be a $\mathbb{P}_3$ of lines of $\mathbb{G}(1,5) \cap L$ through a point p. Because if there were one, then $\dim \operatorname{span}\{{}^t\!pA, {}^t\!pB, {}^t\!pC\} = 1$, i.e. there exist two points $(\lambda : \mu : \nu), (\lambda' : \mu' : \nu') \in \mathbb{P}_2$ with

$$ {}^t\!p(\lambda A + \mu B + \nu C) = {}^t\!p(\lambda' A + \mu' B + \nu' C) = 0. $$

It follows that all the matrices

$$ (\alpha\lambda + \beta\lambda')A + (\alpha\mu + \beta\mu')B + (\alpha\nu + \beta\nu')C \quad \text{for all } (\alpha : \beta) \in \mathbb{P}_1 $$

have a non-trivial kernel. Hence the line $(\alpha\lambda + \beta\lambda' : \alpha\mu + \beta\mu' : \alpha\nu + \beta\nu')$ must lie in $L^* \cap \mathbb{G}(1,5)^* = C^*$. But this is a contradiction since the cubic C^* is irreducible. $\square$

Proposition 4.2 *The lines $l_{(\lambda:\mu:\nu)} \subset \mathbb{P}_5$ with $(\lambda : \mu : \nu) \in C^*$ do not intersect each other.*

Proof. Assume that the line $l_{(\lambda:\mu:\nu)}$ intersects the line $l_{(\lambda':\mu':\nu')}$ in the point p, i.e.

$$ p \in \ker(\lambda A + \mu B + \nu C) \cap \ker(\lambda' A + \mu' B + \nu' C) \neq 0. $$

Then

$$ p \in \ker((\alpha\lambda + \beta\lambda')A + (\alpha\mu + \beta\mu')B + (\alpha\nu + \beta\nu')C) \neq 0 \quad \text{for all } (\alpha : \beta) \in \mathbb{P}_1, $$

and the line $(\alpha\lambda + \beta\lambda' : \alpha\mu + \beta\mu' : \alpha\nu + \beta\nu')$ must be contained in the irreducible cubic C^*, which is a contradiction. $\square$

Let us again derive a normal form:

Proposition 4.3 *For a general plane $L^* = \mathbb{P}(\lambda A + \mu B + \nu C) \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^6)^*$ there exists a choice of bases of L^* and $\mathbb{C}^6$ such that*

$$ A = \begin{pmatrix} 0 & -1 & & & & \\ 1 & 0 & & & \multicolumn{2}{c}{0} \\ & & 0 & -1 & & \\ & & 1 & 0 & & \\ \multicolumn{2}{c}{0} & & & 0 & 0 \\ & & & & 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 0 & & & & \\ 0 & 0 & & & \multicolumn{2}{c}{0} \\ & & 0 & -1 & & \\ & & 1 & 0 & & \\ \multicolumn{2}{c}{0} & & & 0 & -1 \\ & & & & 1 & 0 \end{pmatrix} $$

$$ C = \left(\begin{array}{cc|cc|cc} 0 & 0 & -\alpha & 0 & -\gamma & 0 \\ 0 & 0 & 0 & -\alpha & -\delta & -\gamma \\ \hline \alpha & 0 & 0 & -1 & -\beta & 0 \\ 0 & \alpha & 1 & 0 & 0 & -\beta \\ \hline \gamma & \delta & \beta & 0 & 0 & 0 \\ 0 & \gamma & 0 & \beta & 0 & 0 \end{array}\right). $$

Remark 4.4 *It is also possible to derive a more symmetric normal form where all three matrices look like C only with the* $\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}$ *block moved along the diagonal, but this is not more useful for our computations.*

Proof of proposition 4.3. We may assume that the line $\mathbb{P}(\lambda A + \mu B) \subset L^*$ is a general line. By Proposition 3.2 there exists a choice of coordinates (corresponding to $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = \infty$) such that A and B are of the required form. Further, if we change the coordinates of $\mathbb{C}^6$ by transformations of the type

$$T = \begin{pmatrix} t_1 & 0 & 0 \\ 0 & t_2 & 0 \\ 0 & 0 & t_3 \end{pmatrix} \quad t_1, t_2, t_3 \in \mathrm{SL}(2, \mathbb{C}),$$

then A and B will stay the same by Theorem 3.5.

We write the matrix C as

$$C = \begin{pmatrix} c_1 J & -{}^tC_{21} & -{}^tC_{31} \\ C_{21} & c_2 J & -{}^tC_{32} \\ C_{31} & C_{32} & c_3 J \end{pmatrix} \quad \text{with} \quad \begin{matrix} J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix};\ c_1, c_2, c_3 \in \mathbb{C} \\ C_{21}, C_{31}, C_{32} \in \mathrm{M}(2, \mathbb{C}). \end{matrix}$$

We may assume that $c_1 = c_3 = 0$, $c_2 = 1$, otherwise we replace C by the matrix $1/(c_2 - c_1 - c_3)(C - c_1 A - c_3 B)$. This is possible since $c_2 - c_1 - c_3 \neq 0$, because C is general. So C looks like

$$C = \begin{pmatrix} 0 & -{}^tC_{21} & -{}^tC_{31} \\ C_{21} & J & -{}^tC_{32} \\ C_{31} & C_{32} & 0 \end{pmatrix}.$$

The generality of C ensures that the matrices C_{21} and C_{32} are invertible, so

$$T = \begin{pmatrix} \frac{1}{\alpha} C_{21} & 0 & 0 \\ 0 & \mathrm{E}_2 & 0 \\ 0 & 0 & \frac{1}{\beta}{}^tC_{32} \end{pmatrix} \quad \text{with} \quad \begin{matrix} \alpha = \sqrt{\det C_{21}} \\ \beta = \sqrt{\det C_{32}} \end{matrix}$$

is of the above mentioned type and transforms C into

$$C' := {}^tT^{-1} C T^{-1} = \begin{pmatrix} 0 & -\alpha \mathrm{E}_2 & -{}^t\overline{C} \\ \alpha \mathrm{E}_2 & J & -\beta \mathrm{E}_2 \\ \overline{C} & \beta \mathrm{E}_2 & 0 \end{pmatrix} \quad \text{with } \overline{C} := \alpha\beta C_{32}^{-1} C_{31} C_{21}^{-1}.$$

This matrix will be transformed under

$$T = \begin{pmatrix} t^{-1} & 0 & 0 \\ 0 & {}^tt & 0 \\ 0 & 0 & t^{-1} \end{pmatrix} \quad \text{with } t \in \mathrm{SL}(2, \mathbb{C})$$

into

$$
{}^tT^{-1}C'T^{-1} = \begin{pmatrix} 0 & -\alpha\mathrm{E}_2 & -{}^tt{}^t\overline{C}t \\ \alpha\mathrm{E}_2 & J & -\beta\mathrm{E}_2 \\ {}^tt\overline{C}t & \beta\mathrm{E}_2 & 0 \end{pmatrix}.
$$

So, all that remains to show is: Given a general matrix $\overline{C}\in\mathrm{M}(2,\mathbb{C})$ there is a matrix $t\in\mathrm{SL}(2,\mathbb{C})$ such that

$$
{}^tt\overline{C}t = \begin{pmatrix} \gamma & \delta \\ 0 & \gamma \end{pmatrix}.
$$

If

$$
\overline{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \quad \text{and} \quad t = \begin{pmatrix} 1 & -\frac{c_{21}}{c_{11}} \\ 0 & 1 \end{pmatrix}
$$

then

$$
\overline{C}' = {}^tt\overline{C}t = \begin{pmatrix} c_{11} & c_{12}-c_{21} \\ 0 & \frac{\det\overline{C}}{c_{11}} \end{pmatrix}
$$

and an additional transformation by

$$
t = \begin{pmatrix} \frac{\sqrt[4]{\det\overline{C}}}{\sqrt{c_{11}}} & 0 \\ 0 & \frac{\sqrt{c_{11}}}{\sqrt[4]{\det\overline{C}}} \end{pmatrix}
$$

takes $\overline{C}$ into the desired form

$$
{}^tt\overline{C}'t = \begin{pmatrix} \sqrt{\det\overline{C}} & c_{12}-c_{21} \\ 0 & \sqrt{\det\overline{C}} \end{pmatrix}. \qquad \square
$$

Remark 4.5 *In terms of this coordinates the cubic $C^*\subset L^*$ is given as*

$$
\lambda^2\mu+\mu^2\lambda+\lambda\mu\nu-(\gamma^2+\beta^2)\lambda\nu^2-(\alpha^2+\gamma^2)\mu\nu^2+(\alpha\beta\delta-\gamma^2)\nu^3.
$$

One checks that the cubic is smooth for general $\alpha,\beta,\gamma,\delta$.

Now we start to determine the automorphism group of $\mathbb{G}(1,5)\cap L$. A given automorphism

$$
\varphi\in\mathrm{Aut}(\mathbb{G}(1,5)\cap L)=\mathrm{Aut}(L,\mathbb{P}(\textstyle\bigwedge^2\mathbb{C}^6))\cap\mathrm{Aut}(\mathbb{G}(1,5),\mathbb{P}(\textstyle\bigwedge^2\mathbb{C}^6))
$$

induces a dual automorphism φ^* on the dual projective space $\mathbb{P}(\bigwedge^2\mathbb{C}^6))^*$ that preserves L^* and the dual Grassmannian $\mathbb{G}(1,5)^*$, i.e.

$$
\varphi^*\in\mathrm{Aut}(L^*,\mathbb{P}(\textstyle\bigwedge^2\mathbb{C}^6)^*)\cap\mathrm{Aut}(\mathbb{G}(1,5)^*,\mathbb{P}(\textstyle\bigwedge^2\mathbb{C}^6)^*).
$$

In particular, φ^* induces a projective transformation of L^* preserving C^*. But a smooth cubic has only finitely many automorphisms that are induced by a projective linear transformation [BK, 7.3].

To find all automorphisms of $\mathbb{G}(1,5) \cap L$ that induce the identity on L^*, we look for the $T \in \mathbb{P}\mathrm{GL}(6,\mathbb{C})$ such that

$$^tT^{-1}(\lambda A + \mu B + \nu C)T^{-1} \in \mathbb{C} \cdot (\lambda A + \mu B + \nu C) \quad \text{for all } \lambda, \mu, \nu \in \mathbb{C}.$$

It suffices to check this for $(\lambda, \mu, \nu) = (1,0,0)$, $(0,1,0)$, and $(0,0,1)$. If we normalize the representation of T in $\mathrm{GL}(6,\mathbb{C})$ by $\det T = 1$, we know from the previous section that $^tT^{-1}AT^{-1} = \mathbb{C} \cdot A$ and $^tT^{-1}BT = \mathbb{C} \cdot B$ is equivalent to

$$T = \begin{pmatrix} t_1 & 0 & 0 \\ 0 & t_2 & 0 \\ 0 & 0 & t_3 \end{pmatrix} \quad \text{with } t_1, t_2, t_3 \in \mathrm{SL}(2,\mathbb{C}).$$

Furthermore, we compute

$$^tT^{-1}CT^{-1} = \begin{pmatrix} 0 & -\alpha\, {}^tt_1^{-1}t_2^{-1} & -{}^tt_1^{-1}\left(\begin{smallmatrix}\gamma & 0\\ \delta & \gamma\end{smallmatrix}\right)t_3^{-1} \\ \alpha\, {}^tt_2^{-1}t_1^{-1} & 0 & -\beta\, {}^tt_2^{-1}t_3^{-1} \\ {}^tt_3^{-1}\left(\begin{smallmatrix}\gamma & \delta\\ 0 & \gamma\end{smallmatrix}\right)t_1^{-1} & \beta\, {}^tt_3^{-1}t_2^{-1} & 0 \end{pmatrix},$$

so that $^tT^{-1}CT^{-1} = \vartheta \cdot C$ iff $t_1 = \frac{1}{\vartheta}{}^tt_2^{-1} = t_3 =: t$ and

$$^tt^{-1}\begin{pmatrix} \gamma & \delta \\ 0 & \gamma \end{pmatrix} t^{-1} = \vartheta \begin{pmatrix} \gamma & \delta \\ 0 & \gamma \end{pmatrix}.$$

Because of $\det t_1 = \det t_2 = 1$, ϑ must be either 1 or -1. Setting

$$t = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Longrightarrow t^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

the last condition together with $\det t = 1$ requires that the following polynomials vanish:

$$\begin{gathered} (d^2 + c^2 - \vartheta)\gamma - dc\delta,\ (db + ac)\gamma - (\vartheta - ad)\delta \\ (db + ac)\gamma - bc\delta,\ (b^2 + a^2 - \vartheta)\gamma - ba\delta,\ ad - bc - 1 \end{gathered}$$

The Gröbner basis of the ideal generated by these polynomials with respect to the lexicographical order $\gamma > \delta > a > b > c > d$ can be computed for $\vartheta = 1$ as

$$\gamma a + \delta c - \gamma d,\ b + c,\ ad + c^2 - 1,$$

so that

$$t = \begin{pmatrix} a & \frac{\gamma}{\delta}(a - d) \\ \frac{\gamma}{\delta}(d - a) & d \end{pmatrix} \quad \text{with } \det t = 1.$$

For $\vartheta=-1$ we get as the Gröbner basis

$$\delta, a+d, -c+b, d^2+c^2+1.$$

Since in the general case $\delta\neq 0$, this gives no further automorphisms.

The one-dimensional subgroup of $\mathbb{P}\mathrm{GL}(2,\mathbb{C})$ consisting of elements like t above acts on $\mathbb{P}_1$ with the two fixed points $(-\delta\pm\sqrt{\delta^2-4\gamma^2} : 2\gamma)$. Hence it is conjugate to the one-dimensional subgroup of $\mathbb{P}\mathrm{GL}(2,\mathbb{C})$ that acts on $\mathbb{P}_1$ with the fixed points 0 and ∞. Now this subgroup consists of the invertible diagonal matrices of $\mathbb{P}\mathrm{GL}(2,\mathbb{C})$, so it is isomorphic to $\mathbb{C}^*$. Therefore we have shown

Theorem 4.6 *The component of the automorphism group of* $\mathbb{G}(1,5)\cap H^3$ *containing the identity is isomorphic to* $\mathbb{C}^*$. *The quotient of* $\mathrm{Aut}(\mathbb{G}(1,5)\cap H^3)$ *by this component is a subgroup of the finite group of projective automorphisms of a smooth cubic in* $\mathbb{P}_2$.

5 $\mathbb{G}(1,2n)\cap H$

The hyperplane H is given by an element $A\in(\bigwedge^2\mathbb{C}^{2n+1})^*$ which can be thought of as an antisymmetric matrix of size $2n+1$. Since antisymmetric matrices have an even rank, the general H corresponds to an A of rank $2n$. The one dimensional kernel of A as a point of $\mathbb{P}_{2n}$ is called the center c of H.

The center plays a special role in the geometry of the line system $\mathbb{G}(1,2n)\cap H$ in $\mathbb{P}_{2n}$.

Proposition 5.1 *Every line through the center of the line system* $\mathbb{G}(1,2n)\cap H$ *is in the line system. The center is the only point with this property.*

Moreover, if the line $l\not\ni c$ *belongs to the line system, so does every line in the plane spanned by the line* l *and the center* c.

Proof. The line $c\wedge p$ through the center will be in the line system if ${}^t cAp=0$. But c is the kernel of A, so this is true. On the other hand, if $\overline{c}$ is a point such that every line through it belongs to the line system, then ${}^t\overline{c}Ap=0$ for all $p\in\mathbb{P}_{2n}$. Hence $\overline{c}$ must be in the kernel of A, and therefore $\overline{c}=c$.

Let the line $l=p\wedge q$ be in the line system. All the lines in the plane spanned by l and c – except the lines through c itself – can be written as

$$(\alpha p+\beta c)\wedge(\lambda q+\mu c)\quad\text{for } (\alpha:\beta),(\lambda:\mu)\in\mathbb{P}_1.$$

These will be in the line system since

$$\begin{aligned}(\alpha{}^t p+\beta{}^t c)A(\lambda q+\mu c) &= \alpha\lambda{}^t pAq+\alpha\mu{}^t pAc+\beta\lambda{}^t cAq+\beta\mu{}^t cAc\\ &= \alpha\lambda{}^t pAq=0.\end{aligned}$$

□

Let us for a moment look at the projection $\mathbb{P}(\mathbb{C}^{2n+1}/c)$ of $\mathbb{P}_{2n}$ from the center c. This projection maps all lines in a plane through the center – except the lines through c itself – to only one line. Hence we get a codimension one line system inside $\mathbb{P}_{2n-1}$. In fact, it is of the form $\mathbb{G}(1, 2n-1) \cap \overline{H}$, which is most easily seen in coordinates. We choose a basis $(e_0, \ldots, e_{2n})$ of $\mathbb{C}^{2n+1}$ such that the hyperplane H is given by the matrix

$$A = \left(\begin{array}{cc|c} 0 & -\mathrm{E}_n & 0 \\ \mathrm{E}_n & 0 & \vdots \\ & & 0 \\ \hline \multicolumn{2}{c|}{0 \cdots 0} & 0 \end{array}\right) \in \mathrm{Antisym}(2n+1, \mathbb{C}).$$

The center of H is $c = \mathbb{P}(e_{2n})$. So the projected line system is $\mathbb{G}(1, 2n-1) \cap \overline{H}$, where $\overline{H}$ is given by the matrix A with the last row and column deleted.

This description helps to determine the automorphism group of $\mathbb{G}(1, 2n) \cap H$.

First of all, any of the automorphism must – as a transformation $T \in \mathbb{P}\mathrm{GL}(2n+1, \mathbb{C})$ – preserve the center, i.e. $Tc = c$. Therefore it induces a transformation $\overline{T}$ of the projected space $\mathbb{P}(\mathbb{C}^{2n+1}/c)$. This induced transformation $\overline{T}$ has to preserve the projected line system $\mathbb{G}(1, 2n-1) \cap \overline{H}$. Since this case has been treated in Section 2, we know that if we normalize $\overline{T}$ by $\det \overline{T} = 1$, then $\overline{T} \in \mathrm{Sp}(2n, \mathbb{C})$. Therefore T must have been of the form

$$T = \left(\begin{array}{c|c} & 0 \\ \overline{T} & \vdots \\ & 0 \\ \hline a_0 \cdots a_{2n-1} & b \end{array}\right) \quad \text{with} \quad \begin{array}{l} \overline{T} \in \mathrm{Sp}(2n, \mathbb{C}) \\ a_i \in \mathbb{C} \\ b \in \mathbb{C}^*. \end{array}$$

One immediately checks that ${}^tT^{-1}AT = A$, so that the automorphism group as a subset of $\mathbb{P}\mathrm{GL}(2n+1, \mathbb{C})$ consists of all elements of the above type. Since we normalized $\overline{T}$, we have up to multiplication by -1 an unique representative in the class of $\mathbb{P}\mathrm{GL}(2n+1, \mathbb{C})$.

A small computation shows that

$$N := \left\{ \left(\begin{array}{c|c} & 0 \\ \mathrm{E}_{2n} & \vdots \\ & 0 \\ \hline a_0 \cdots a_{2n-1} & 1 \end{array}\right) \mid a_i \in \mathbb{C} \right\} \subset \mathrm{Aut}(\mathbb{G}(1, 2n) \cap H)$$

is a normal subgroup which is isomorphic to $(\mathbb{C}^{2n}, +)$.

Collecting everything together we have

Proposition 5.2 *The automorphism group of* $\mathbb{G}(1, 2n) \cap H$ *for a general hyperplane* $H \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^{2n+1})$ *is an extention of* $\mathrm{Sp}(2n, \mathbb{C}) \times \mathbb{C}^*/\{1, -1\}$ *by*

$(\mathbb{C}^{2n},+)$ *and is isomorphic to the group*

$$\left\{ \left(\begin{array}{ccc|c} & & & 0 \\ & T & & \vdots \\ & & & 0 \\ \hline a_0 & \cdots & a_{2n-1} & b \end{array} \right) \middle| \begin{array}{l} T \in \mathrm{Sp}(2n,\mathbb{C}) \\ a_i \in \mathbb{C} \\ b \in \mathbb{C}^* \end{array} \right\} / \{1,-1\}.$$

The action of the automorphism group on the line system is described by the following

Proposition 5.3 *The action of the automorphism group of* $\mathbb{G}(1,2n)\cap H$ *on the lines of* $\mathbb{G}(1,2n)\cap H$ *has two orbits:*

1. *the lines containing the center* c
2. *the lines that do not.*

Proof. Since all the automorphisms preserve the center any orbit will be contained in these two sets.

First we show that the lines containing c form one orbit. For two lines $c\wedge p$ and $c\wedge q$, we may assume $p,q\in\mathbb{P}(\mathbb{C}^{2n}\times 0)$. Take a $\overline{T}\in \mathrm{Sp}(2n,\mathbb{C})$ that maps p to q. The trivial extention of $\overline{T}$ to $T\in \mathrm{SL}(2n+1,\mathbb{C})$ will take $c\wedge p$ to $c\wedge q$.

The other lines will form the second orbit since any line not containing the center can be pushed into the hyperplane $\mathbb{P}(\mathbb{C}^{2n}\times 0)$ by a transformation with an element of the normal subgroup N. There one can use the transitive action of the $\mathrm{Aut}(\mathbb{G}(1,2n-1)\cap H)$ subgroup to show that all these lines can be mapped onto each other. □

6 $\mathbb{G}(1,2n)\cap H^2$

Let $L=H^2$ be a 2-codimensional linear subspace of $\mathbb{P}(\bigwedge^2\mathbb{C}^{2n+1})$. We want to study the linear line system $\mathbb{G}(1,2n)\cap L$. To L corresponds the line $L^*=\mathbb{P}(\lambda A-\mu B)\subset\mathbb{P}(\bigwedge^2\mathbb{C}^{2n+1})^*$ of the hyperplanes $H_{(\lambda:\mu)}=\mathbb{P}(\lambda A-\mu B)$ containing L. We identify as always $(\bigwedge^2\mathbb{C}^{2n+1})^*$ with the antisymmetric matrices $\mathrm{Antisym}(2n+1,\mathbb{C})$. The locus of antisymmetric matrices of corank 3 in $\mathrm{Antisym}(2n+1,\mathbb{C})$ is 3-codimensional by Corollary 1.7. Therefore a line L^* may be called general if it does not intersect it. Hence for the general line L^* the antisymmetric matrices $\lambda A-\mu B$ corresponding to the hyperplanes $H_{(\lambda:\mu)}$ have all corank 1. So each of the hyperplane sections $\mathbb{G}(1,2n)\cap H_{(\lambda:\mu)}$ has a unique center $c_{(\lambda:\mu)}\in\mathbb{P}_{2n}$ by Proposition 5.1. These centers play an important role in the geometry of $\mathbb{G}(1,2n)\cap L$.

Proposition 6.1 *The centers* $c_{(\lambda:\mu)}$ *are those points of* $\mathbb{P}_{2n}$ *through which there passes a* $\mathbb{P}_{2n-2}$ *of lines of the line system* $\mathbb{G}(1,2n)\cap L$*. Through all the other points of* $\mathbb{P}_{2n}$ *passes only a* $\mathbb{P}_{2n-3}$ *of lines.*

Proof. The lines of the line system through a point $p \in \mathbb{P}_{2n}$ are $p \wedge q$ with ${}^tpAq = {}^tpBq = 0$. So we need to show that ${}^tpA$ and ${}^tpB$ are linear dependent iff p is a center of a hyperplane $H_{(\lambda:\mu)}$. Now ${}^tpA$ and ${}^tpB$ are linear dependent precisely if there exists a $(\lambda:\mu) \in \mathbb{P}_1$ with $0 = \lambda{}^tpA - \mu{}^tpB = {}^tp(\lambda A - \mu B)$, i.e. p is the kernel of $\lambda A - \mu B$, which is by definition the center of $H_{(\lambda:\mu)}$. □

Remark 6.2 *Any line that contains two centers is a member of the line system* $\mathbb{G}(1, N) \cap L$.

Proof. If the line contains the centers $c_{(\alpha:\beta)}$ and $c_{(\lambda:\mu)}$, it is contained in the hyperplanes $H_{(\alpha:\beta)}$ and $H_{(\lambda:\mu)}$ by Proposition 5.1 and therefore in their intersection $L = H_{(\alpha:\beta)} \cap H_{(\lambda:\mu)}$. □

Next we want to know more about the curve $c_{(\lambda:\mu)}$.

Proposition 6.3 *Let A, B be two antisymmetric matrices of size $2n+1$ such that every non-zero linear combination of them has corank 1. Then the map*

$$\begin{array}{rccc} c: & \mathbb{P}_1 & \longrightarrow & \mathbb{P}_{2n} \\ & (\lambda:\mu) & \longmapsto & \ker(\lambda A - \mu B) \end{array}$$

is a parametrisation of a rational normal curve of degree n.

Proof. (compare [SR, X,4.3] for $n = 2$.) First we show that the map is injective. If it is not, there are two points of $\mathbb{P}_1$ with the same image. We may assume that this is the case for $(1:0)$ and $(0:1)$, i.e. A and B have the same kernel, say e_0. Writing A and B in a basis with e_0 as first element, we have

$$A = \begin{pmatrix} 0 & \cdots 0 \\ \vdots & \widetilde{A} \\ 0 & \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & \cdots 0 \\ \vdots & \widetilde{B} \\ 0 & \end{pmatrix} \quad \text{with } \widetilde{A}, \widetilde{B} \in \mathrm{Antisym}(2n, \mathbb{C}).$$

Since $\det(\lambda\widetilde{A} - \mu\widetilde{B})$ is a homogeneous polynomial of degree $2n$, there exist a $(\lambda':\mu') \in \mathbb{P}_1$ with $\det(\lambda'\widetilde{A} - \mu'\widetilde{B}) = 0$. But then $\lambda' A - \mu' B$ has corank at least two, which contradicts our assumption.

Secondly, we proof that the map is of maximal rank everywhere. If it is not, we may assume that it is not maximal at $(1:0)$. Restricting to the chart $\lambda = 1$, this means $c'(0) = 0$. Now from

$$\begin{aligned} & (A - \mu B)c(\mu) = 0 \\ \Longrightarrow\ & Ac'(\mu) - Bc(\mu) - \mu Bc'(\mu) = 0 \\ \Longrightarrow\ & Ac'(0) - Bc(0) = 0 \end{aligned}$$

$Bc(0) = 0$ follows. Therefore A and B have the same kernel $c(0)$, and we are back in the above chain of arguments.

Finally, we have to show that the embedding c is of degree n. For this we give an explicit form of the map. Recall [B, 5.2] that the determinant of an antisymmetric matrix $C = (c_{ij})$ of size $2n$ is the square of the irreducible Pfaffian polynomial $\operatorname{Pf} C$,

$$\operatorname{Pf} C := \sum_{\sigma} \operatorname{sgn}(\sigma) c_{\sigma(1)\sigma(2)} \ldots c_{\sigma(2n-1)\sigma(2n)},$$

where σ runs through all permutations $\mathrm{S}(2n)$ with $\sigma(2i-1) < \sigma(2i)$ for $i = 1 \ldots n$ and $\sigma(2i) < \sigma(2i+2)$ for $i = 1 \ldots n-1$.

Let $c_i(\lambda:\mu)$ denote $(-1)^i$-times the Pfaffian of the matrix $\lambda A - \mu B$ with the i-th row and column deleted. Then the $c_i(\lambda:\mu)$ are irreducible polynomials of degree n and by a straightforward but messy computation one can check that $(c_0(\lambda:\mu): \ldots : c_{2n}(\lambda:\mu))$ is the kernel of $\lambda A - \mu B$. Therefore $c = (c_0 : \ldots : c_{2n})$, which shows that c is a degree n embedding of $\mathbb{P}_1$. $\square$

After we have determined the special points in $\mathbb{P}_{2n}$ of the line system $\mathbb{G}(1,2n)\cap L$, we are nearly ready to compute its automorphism group. It remains to give a normal form for the line $L^* \subset \mathbb{P}(\bigwedge^2 \mathbb{C}^{2n+1})$ to make computations easier. This normal form was found by Donagi [D, 2.2]. But he did not give a proof for it since his main interest was lines in $\mathbb{P}_{2n-1}$ and not in $\mathbb{P}_{2n}$. So we give the proof here.

Proposition 6.4 *Let L^* be a line in $\mathbb{P}(\bigwedge^2 \mathbb{C}^{2n+1})^*$ such that the antisymmetric matrices corresponding to the points of L^* have all corank 1. Then there exists a basis $(e_0, \ldots, e_{2n})$ of $\mathbb{C}^{2n+1}$ such that the line can be taken as $L^* = \mathbb{P}(\lambda A - \mu B)$ with the matrices*

$$A = \left(\begin{array}{cc|c} 0 & -\mathrm{E}_n & \begin{smallmatrix}0\\ \vdots \\ 0\end{smallmatrix} \\ \mathrm{E}_n & 0 & \\ \hline 0 \cdots & 0 & 0 \end{array}\right) \quad \textit{and} \quad B = \left(\begin{array}{c|c|c} 0 & \begin{smallmatrix}0\\ \vdots\end{smallmatrix} & -\mathrm{E}_n \\ \hline 0 \cdots & 0 & \cdots 0 \\ \hline \mathrm{E}_n & \begin{smallmatrix}\vdots\\ 0\end{smallmatrix} & 0 \end{array}\right).$$

Proof. Let A and B any two matrices of L^* in an arbitrary basis. We will adjust the basis in three steps to achieve the required form for A and B.

1^{st} *Step:* We know that the map

$$\begin{array}{rccc} c: & \mathbb{P}_1 & \longrightarrow & \mathbb{P}_{2n} \\ & (\lambda:\mu) & \longmapsto & \ker(\lambda A - \mu B) \end{array}$$

is a parametrisation of a rational normal curve of degree n. Modulo projective transformations of $\mathbb{P}_1$ and $\mathbb{P}_{2n}$ such parametrisations are all the same. So we can pick a basis of $\mathbb{P}_1$ and $n+1$ linear independent vectors $e_n, \ldots, e_{2n}$ of $\mathbb{C}^{2n+1}$ such that

$$\begin{array}{rccc} c: & \mathbb{P}_1 & \longrightarrow & \mathbb{P}_{2n} \\ & (\lambda:\mu) & \longmapsto & \mathbb{P}\left(\sum\limits_{i=0}^{n} \lambda^i \mu^{n-i} e_{n+i}\right). \end{array}$$

Extending $(e_n, \ldots, e_{2n})$ to a basis $(e_0, \ldots, e_{2n})$ of $\mathbb{C}^{2n+1}$ and denoting by $a_0, \ldots, a_{2n}$ resp. $b_0, \ldots, b_{2n}$ the columns of A resp. B, the fact that $c(\lambda:\mu)$ is the kernel of $\lambda A - \mu B$ for all $(\lambda:\mu) \in \mathbb{P}_1$ has the following consequences for A and B:

$$\begin{aligned} &(\lambda A - \mu B)\left(\sum_{i=0}^{n} \lambda^i \mu^{n-i} e_{n+i}\right) = 0 \\ &\Longrightarrow \sum_{i=0}^{n} \lambda^{i+1}\mu^{n-i} a_{n+i} - \sum_{i=0}^{n} \lambda^i \mu^{n+1-i} b_{n+i} = 0 \\ &\Longrightarrow -\mu^{n+1} b_n + \sum_{i=0}^{n-1} \lambda^{i+1}\mu^{n-i}(a_{n+1} - b_{n+i+1}) + \lambda^{n+1} a_{2n} = 0 \\ &\Longrightarrow b_n = 0,\ a_{2n} = 0,\ a_{n+i} = b_{n+i+1} \ \text{ for } i = 0 \ldots n-1. \end{aligned} \tag{$*$}$$

We claim that this implies:

$$a_{n+i,n+j} = 0 \quad \text{for } i,j = 0 \ldots n.$$

Indeed, for $1 \leq i \leq n$, $0 \leq j \leq n-1$ we have using $(*)$

$$a_{n+i,n+j} = b_{n+i,n+j+1} = -b_{n+j+1,n+i} = -a_{n+j+1,n+i-1} = a_{n+i-1,n+j+1}.$$

This shows that the $a_{n+i,n+j}$ are all the same for $i+j = const$, in particular $a_{n+i,n+j} = a_{n+j,n+i}$. On the other hand, by the antisymmetricity of A we have $a_{n+i,n+j} = -a_{n+j,n+i}$, and the claim follows.

Using $(*)$ again we know that A and B look in our basis like

$$A = \left(\begin{array}{cc|c} \widetilde{A} & -{}^tM & 0 \\ M & 0 & \vdots \\ & & 0 \\ \hline \multicolumn{2}{c|}{0 \cdots 0} & 0 \end{array}\right) \quad \text{and} \quad B = \left(\begin{array}{c|c|c} \widetilde{B} & \begin{matrix}0\\ \vdots\end{matrix} & -{}^tM \\ \hline 0 \cdots & 0 & \cdots 0 \\ \hline M & \begin{matrix}\vdots\\ 0\end{matrix} & 0 \end{array}\right)$$

with $\widetilde{A}, \widetilde{B} \in \mathrm{Antisym}(n, \mathbb{C})$ and $M \in \mathrm{GL}(n, \mathbb{C})$.

2^{nd} *Step:* Here we will improve the choice of $(e_0, \ldots, e_{n-1})$ to achieve $\widetilde{A} = 0$ and $M = \mathrm{E}_n$. We claim:

Let an antisymmetric matrix $A \in \mathrm{Antisym}(2n+1, \mathbb{C})$ of rank $2n$ and linear independent vectors $e_n, \ldots, e_{2n} \in \mathbb{C}^{2n+1}$ with ${}^te_iAe_j = 0$ for $n \leq i,j \leq 2n$ and

$Ae_{2n}=0$ be given. Then $(e_n,\dots,e_{2n})$ can be extended to a basis $(e_0,\dots,e_{2n})$ of $\mathbb{C}^{2n+1}$ such that in this basis A is given as

$$A=\left(\begin{array}{cc|c} 0 & -\mathrm{E}_n & 0 \\ & & \vdots \\ \mathrm{E}_n & 0 & 0 \\ \hline \multicolumn{2}{c|}{0\cdots 0} & 0 \end{array}\right).$$

The proof is by induction. The statement is trivial for $n=0$. Assuming the claim for $n-1$, we prove it for n. Let $W:=\bigcap_{i=1}^{n}\ker {}^te_{n+i}A$, then there exists an $e_0\in W$ with ${}^te_nAe_0=1$. If not, we would have $W=W\cap\ker {}^te_nA$ and with $e_{2n}\in\ker A$

$$\dim\operatorname{span}\{{}^te_nA,\dots,{}^te_{2n}A\}=\dim\operatorname{span}\{{}^te_{n+1}A,\dots,{}^te_{2n-1}A\}\le n-1,$$

which contradicts $\operatorname{rank}A=2n$.

Set $V:=\ker {}^te_0A\cap\ker {}^te_nA$, then $\dim V=2(n-1)+1$ and $e_{n+1},\dots,e_{2n}\in V$. Therefore the induction hypothesis can be applied to $A|_V$. Together with ${}^te_0Ae_n=1$ and ${}^te_0Av={}^te_nAv=0$ for $v\in V$ this implies the stated form of the matrix.

So up to now A and B look like

$$A=\left(\begin{array}{cc|c} 0 & -\mathrm{E}_n & 0 \\ & & \vdots \\ \mathrm{E}_n & 0 & 0 \\ \hline \multicolumn{2}{c|}{0\cdots 0} & 0 \end{array}\right)\qquad B=\left(\begin{array}{c|c|c} \widetilde{B} & \begin{matrix}0\\ \vdots\end{matrix} & -\mathrm{E}_n \\ \hline 0\ \cdots & 0 & \cdots\ 0 \\ \hline \mathrm{E}_n & \begin{matrix}\vdots\\ 0\end{matrix} & 0 \end{array}\right).$$

3^{rd} *Step:* We adjust the vectors $(e_0,\dots,e_{n-1})$ so that $\widetilde{B}=0$ and A stays the same.

We note that a transformation of $\mathbb{C}^{2n+1}$ by

$$T=\left(\begin{array}{cc|c} \mathrm{E}_n & 0 & 0 \\ & & \vdots \\ t & \mathrm{E}_n & 0 \\ \hline \multicolumn{2}{c|}{0\cdots 0} & 1 \end{array}\right)^{-1}\quad\text{with } t\in\mathrm{Sym}(n,\mathbb{C})$$

does not change A since

$${}^tT^{-1}AT^{-1}={}^tT^{-1}(AT^{-1})=\left(\begin{array}{cc|c} \mathrm{E}_n & t & 0 \\ & & \vdots \\ 0 & \mathrm{E}_n & 0 \\ \hline \multicolumn{2}{c|}{0\cdots 0} & 1 \end{array}\right)\left(\begin{array}{cc|c} -t & -\mathrm{E}_n & 0 \\ & & \vdots \\ \mathrm{E}_n & 0 & 0 \\ \hline \multicolumn{2}{c|}{0\cdots 0} & 0 \end{array}\right)=A.$$

If we denote by $\overline{t}$ resp. $|t\in\mathrm{M}(n\times n,\mathbb{C})$ the matrix that we obtain by deleting the first row resp. column of t and adding a row of zeroes below resp. a column of zeroes on the right side, we can write down the transformation of B as follows:

$$
\begin{aligned}
{}^tT^{-1}BT^{-1} = {}^tT^{-1}(BT^{-1}) &= \left(\begin{array}{cc|c} \mathrm{E}_n & t & 0 \\ & & \vdots \\ 0 & \mathrm{E}_n & 0 \\ \hline 0\cdots 0 & & 1 \end{array}\right) \left(\begin{array}{c|c|c} \widetilde{B} - \overline{t} & \begin{matrix} 0 \\ \vdots \end{matrix} & -\mathrm{E}_n \\ \hline 0 \;\cdots & 0 & \cdots\; 0 \\ \hline \mathrm{E}_n & \begin{matrix} \vdots \\ 0 \end{matrix} & 0 \end{array}\right) \\
&= \left(\begin{array}{c|c|c} \widetilde{B} - \overline{t} + |t & \begin{matrix} 0 \\ \vdots \end{matrix} & -\mathrm{E}_n \\ \hline 0 \;\cdots & 0 & \cdots\; 0 \\ \hline \mathrm{E}_n & \begin{matrix} \vdots \\ 0 \end{matrix} & 0 \end{array}\right).
\end{aligned}
$$

So to finish this step, we need to show that every antisymmetric matrix $\widetilde{B} = (b_{ij}) \in \mathrm{Antisym}(n, \mathbb{C})$ can be written as $\overline{t} - |t$ for a symmetric matrix $t \in \mathrm{Sym}(n, \mathbb{C})$. The entries of $\overline{t} - |t$ are

$$(t_{i+1,j} - t_{i,j+1})_{ij}$$

where $t_{n+1,i} := t_{i,n+1} := 0$ for all $i = 1 \dots n$. Obviously, $\overline{t} - |t$ is antisymmetric. We set $t_{1i} := t_{i1} := 0$ for $i = 1 \dots n$ and define recursively for j from n down to 2

$$t_{i+1,j} := t_{i,j+1} + b_{ij} \quad \text{for } i = 1 \dots j-1.$$

Then by the symmetry of t the whole matrix t is defined and $\overline{t} - |t = B$ □

For the linear system in normal form the rational curve of centers has the parametrisation

$$
\begin{aligned}
c(\lambda : \mu) = \ker(\lambda A - \mu B) = \ker &\left(\begin{array}{c|c} \text{\Large 0} & \begin{matrix} -\lambda & \mu & & \\ & \ddots & \ddots & \\ & & -\lambda & \mu \end{matrix} \\ \hline \begin{matrix} \lambda & & \\ -\mu & \ddots & \\ & \ddots & \lambda \\ & & -\mu \end{matrix} & \text{\Large 0} \end{array}\right) \\
&= (0 : \ldots : 0 : \mu^n : \mu^{n-1}\lambda : \ldots : \lambda^n).
\end{aligned}
$$

The $\mathbb{P}_{2n-2}$ of lines through a center $c_{(\lambda:\mu)}$ is given by

$$c_{(\lambda:\mu)} \wedge q \qquad \text{where } q \in \mathbb{P}_{2n} \text{ with } {}^tc_{(\lambda:\mu)}Aq = {}^tc_{(\lambda:\mu)}Bq = 0,$$

i.e. q must be an element of the hyperplane $h_{(\lambda:\mu)} \in \mathbb{P}_{2n}^*$

$$\begin{aligned} h_{(\lambda:\mu)} &= \ker {}^t c_{(\lambda:\mu)} A \cap {}^t c_{(\lambda:\mu)} B \\ &= \ker \begin{pmatrix} \mu^n & \mu^{n-1}\lambda & \cdots & \mu\lambda^{n-1} & 0 & \cdots & 0 \\ \mu^{n-1}\lambda & \mu^{n-2}\lambda^2 & \cdots & \lambda^n & 0 & \cdots & 0 \end{pmatrix} \\ &= \ker \begin{pmatrix} \mu^{n-1} & \mu^{n-2}\lambda & \cdots & \lambda^{n-1} & 0 & \cdots & 0 \end{pmatrix}. \end{aligned}$$

So the hyperplanes $h_{(\lambda:\mu)}$, which are traced out by the $\mathbb{P}_{2n-2}$ of lines through the centers, give rise to a rational normal curve of degree $n-1$ in the space of hyperplanes containing the center curve. That the hyperplanes $h_{(\lambda:\mu)}$ contain the center curve could already be seen from the Remark 6.2, by which $h_{(\lambda:\mu)}$ must contain any line connecting $c_{(\lambda:\mu)}$ with any other point of the center curve.

Now we are ready to study the automorphism group of $\mathbb{G}(1,2n)\cap L$.

Any automorphism $T \in \mathrm{Aut}(\mathbb{G}(1,2n)\cap L) \subseteq \mathbb{P}\mathrm{GL}(2n+1,\mathbb{C})$ has to map the center curve onto itself and also the projective space $P \cong \mathbb{P}_n$ spanned by the center curve onto itself. It is known [H, 10.12] that the group of automorphisms of $\mathbb{P}_n$ fixing a rational normal curve of degree n is isomorphic to $\mathbb{P}\mathrm{GL}(2,\mathbb{C})$. If the rational normal curve is given by

$$\begin{array}{rccc} c: & \mathbb{P}_1 & \longrightarrow & \mathbb{P}_n \\ & (\lambda:\mu) & \longmapsto & (\mu^n : \mu^{n-1}\lambda : \ldots : \lambda^n), \end{array}$$

this isomorphism $\mathbb{P}\mathrm{GL}(2,\mathbb{C}) \cong \mathrm{Aut}(c,\mathbb{P}_n)$ maps

$$t = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{P}\mathrm{GL}(2,\mathbb{C})$$

to $t_{n+1} \in \mathbb{P}\mathrm{GL}(n+1,\mathbb{C})$ where t_{n+1} is the unique matrix such that

$$t_{n+1} \begin{pmatrix} \mu^n \\ \mu^{n-1}\lambda \\ \vdots \\ \lambda^n \end{pmatrix} = \begin{pmatrix} (d\mu + c\lambda)^n \\ (d\mu + c\lambda)^{n-1}(b\mu + a\lambda) \\ \vdots \\ (b\mu + a\lambda)^n \end{pmatrix};$$

for example

$$t_2 = \begin{pmatrix} d & c \\ b & a \end{pmatrix} \qquad t_3 = \begin{pmatrix} d^2 & 2cd & c^2 \\ bd & ad+bc & ac \\ b^2 & 2ab & a^2 \end{pmatrix}.$$

Applying this to the center curve restricts the form of the transformation T to

$$T = \left(\begin{array}{c|c} * & 0 \\ \hline * & t_{n+1} \end{array}\right).$$

We know further that if T maps $c_{(\lambda:\mu)}$ to $c_{(a\lambda+b\mu:c\lambda+d\mu)}$, then it must map the hyperplane $h_{(\lambda:\mu)}$ to $h_{(a\lambda+b\mu:c\lambda+d\mu)}$. Therefore it induces also an automorphism on the rational curve h of degree $n-1$ in the dual projective space $(\mathbb{P}_{2n}/P)^*$ of hyperplanes containing P. Hence T must be of the form

$$T=\left(\begin{array}{c|c}\alpha\, {}^t t_n^{-1} & 0\\ \hline * & t_{n+1}\end{array}\right)\quad \text{with } \alpha\in\mathbb{C}^*.$$

We make the following *claim*:

$$T=\left(\begin{array}{c|c}{}^t t_n^{-1} & 0\\ \hline 0 & t_{n+1}\end{array}\right)\in \mathbb{P}\mathrm{GL}(2n+1,\mathbb{C})$$

is an automorphism of the linear system $\mathbb{G}(1,2n)\cap L$.

Proof. We need to check that for every $t\in\mathbb{P}\mathrm{GL}(2,\mathbb{C})$

$${}^tT^{-1}(\lambda A-\mu B)T^{-1}\in \operatorname{span}\{A,B\}$$

for all $\lambda,\mu\in\mathbb{C}$. Since $\mathbb{P}\mathrm{GL}(2,\mathbb{C})$ is a group, this is equivalent to the statement that for every $t\in\mathbb{P}\mathrm{GL}(2,\mathbb{C})$

$${}^tT(\lambda A-\mu B)T\in \operatorname{span}\{A,B\}$$

for all $\lambda,\mu\in\mathbb{C}$. Because of the linearity it is enough to do this for $(\lambda,\mu)=(1,0)$ and $(0,-1)$. Denoting by $\overline{t_{n+1}}$ resp. $\underline{t_{n+1}}$ the matrix t_{n+1} with the first resp. last row deleted, we compute:

$${}^tT(AT)=\left(\begin{array}{c|c}t_n^{-1} & 0\\ \hline 0 & {}^t t_{n+1}\end{array}\right)\left(\begin{array}{c|c}0 & -\underline{t_{n+1}}\\ \hline \begin{matrix}{}^t t_n^{-1}\\ 0\cdots 0\end{matrix} & 0\end{array}\right)=\left(\begin{array}{c|c}0 & -t_n^{-1}\underline{t_{n+1}}\\ \hline {}^t(t_n^{-1}\underline{t_{n+1}}) & 0\end{array}\right)$$

$${}^tT(BT)=\left(\begin{array}{c|c}t_n^{-1} & 0\\ \hline 0 & {}^t t_{n+1}\end{array}\right)\left(\begin{array}{c|c}0 & -\overline{t_{n+1}}\\ \hline \begin{matrix}0\cdots 0\\ {}^t t_n^{-1}\end{matrix} & 0\end{array}\right)=\left(\begin{array}{c|c}0 & -t_n^{-1}\overline{t_{n+1}}\\ \hline {}^t(t_n^{-1}\overline{t_{n+1}}) & 0\end{array}\right).$$

So, if we show

$$\underline{t_{n+1}}=d\left(t_n{}^0_0\right)+c\left({}^0_0t_n\right)\Longrightarrow t_n^{-1}\underline{t_{n+1}}=d\left(\mathrm{E}_n{}^0_0\right)+c\left({}^0_0\mathrm{E}_n\right)$$

$$\overline{t_{n+1}}=b\left(t_n{}^0_0\right)+a\left({}^0_0t_n\right)\Longrightarrow t_n^{-1}\overline{t_{n+1}}=b\left(\mathrm{E}_n{}^0_0\right)+a\left({}^0_0\mathrm{E}_n\right),$$

where ${}^0_0$ stands for adding a column of zeroes, then

$$\begin{aligned}{}^tTAT &= dA + cB\\ {}^tTBT &= bA + aB.\end{aligned}$$

To show the equality for $\underline{t_{n+1}}$ note that on the one hand $\underline{t_{n+1}}$ is the unique matrix with

$$\underline{t_{n+1}}\begin{pmatrix}\mu^n\\ \mu^{n-1}\lambda\\ \vdots\\ \lambda^n\end{pmatrix} = \begin{pmatrix}(d\mu+c\lambda)^n\\ (d\mu+c\lambda)^{n-1}(b\mu+a\lambda)\\ \vdots\\ (d\mu+c\lambda)(b\mu+a\lambda)^{n-1}\end{pmatrix}$$

and on the other hand

$$\begin{pmatrix}(d\mu+c\lambda)^n\\ (d\mu+c\lambda)^{n-1}(b\mu+a\lambda)\\ \vdots\\ (d\mu+c\lambda)(b\mu+a\lambda)^{n-1}\end{pmatrix} = (d\mu+c\lambda)\begin{pmatrix}(d\mu+c\lambda)^{n-1}\\ (d\mu+c\lambda)^{n-2}(b\mu+a\lambda)\\ \vdots\\ (b\mu+a\lambda)^{n-1}\end{pmatrix}$$

$$= (d\mu+c\lambda)t_n\begin{pmatrix}\mu^{n-1}\\ \mu^{n-2}\lambda\\ \vdots\\ \lambda^{n-1}\end{pmatrix} = dt_n\begin{pmatrix}\mu^n\\ \mu^{n-1}\lambda\\ \vdots\\ \mu\lambda^{n-1}\end{pmatrix} + ct_n\begin{pmatrix}\mu^{n-1}\lambda\\ \mu^{n-2}\lambda^2\\ \vdots\\ \lambda^n\end{pmatrix}$$

$$= \left(d\left(t_n{}^0_0\right) + c\left({}^0_0t_n\right)\right)\begin{pmatrix}\mu^n\\ \mu^{n-1}\lambda\\ \vdots\\ \lambda^n\end{pmatrix}.$$

Of course, the proof for $\overline{t_{n+1}} = b\left(t_n{}^0_0\right) + a\left({}^0_0t_n\right)$ is analogous. □

Given any automorphism of the line system $\mathbb{G}(1,2n)\cap L$ we can compose it with one of the above automorphisms such that the composition fixes the center curve pointwise. So, we can focus our attention to automorphisms of the last type.

Lemma 6.5 *All automorphisms of $\mathbb{G}(1,2n)\cap L$ that fix the center curve pointwise are of the form*

$$T = \begin{pmatrix}\alpha \mathrm{E}_n & 0\\ S & \mathrm{E}_{n+1}\end{pmatrix} \text{ with } \alpha\in\mathbb{C}^*,\ S\in \mathrm{M}((n+1)\times n,\mathbb{C}),$$

where the matrix $S\in\mathrm{M}((n+1)\times n,\mathbb{C})$ has the same entries along the minor diagonals, i.e. $s_{ij}=s_{kl}$ for $i+j=k+l$.

As a group these matrices are isomorphic to the semi direct product $\mathbb{C}^{2n}\ltimes\mathbb{C}^*$, $(s,\alpha)\cdot(s',\alpha') = (\alpha's+s',\alpha\alpha')$.

Proof. We need only to check the property of S and the group structure. T is an automorphism iff

$$ {}^tT^{-1}AT^{-1}, {}^tT^{-1}BT^{-1} \in \operatorname{span}\{A, B\}. $$

The inverse of T is

$$ T^{-1} = \begin{pmatrix} \frac{1}{\alpha}\mathrm{E}_n & 0 \\ -\frac{1}{\alpha}S & \mathrm{E}_{n+1} \end{pmatrix}. $$

Now if $\overline{S}$ resp. $\underline{S} \in \mathrm{M}(n \times n, \mathbb{C})$ denote the matrix S with the first resp. last row deleted, then

$$ {}^tT^{-1}(AT^{-1}) = \begin{pmatrix} \frac{1}{\alpha}\mathrm{E}_n & -\frac{1}{\alpha}{}^tS \\ 0 & \mathrm{E}_{n+1} \end{pmatrix} \left(\begin{array}{cc|c} \frac{1}{\alpha}\underline{S} & -\mathrm{E}_n & 0 \\ & & \vdots \\ \frac{1}{\alpha}\mathrm{E}_n & 0 & 0 \\ \hline 0 \cdots & 0 & 0 \end{array}\right) $$

$$ = \left(\begin{array}{cc|c} \frac{1}{\alpha^2}(\underline{S} - {}^t\underline{S}) & -\frac{1}{\alpha}\mathrm{E}_n & 0 \\ & & \vdots \\ \frac{1}{\alpha}\mathrm{E}_n & 0 & 0 \\ \hline 0 \cdots & 0 & 0 \end{array}\right) $$

$$ {}^tT^{-1}(BT^{-1}) = \begin{pmatrix} \frac{1}{\alpha}\mathrm{E}_n & -\frac{1}{\alpha}{}^tS \\ 0 & \mathrm{E}_{n+1} \end{pmatrix} \left(\begin{array}{c|c|c} \frac{1}{\alpha}\overline{S} & \begin{matrix} 0 \\ \vdots \end{matrix} & -\mathrm{E}_n \\ \hline 0 \cdots & 0 & \cdots 0 \\ \hline \frac{1}{\alpha}\mathrm{E}_n & \begin{matrix} \vdots \\ 0 \end{matrix} & 0 \end{array}\right) $$

$$ = \left(\begin{array}{c|c|c} \frac{1}{\alpha^2}(\overline{S} - {}^t\overline{S}) & \begin{matrix} 0 \\ \vdots \end{matrix} & -\frac{1}{\alpha}\mathrm{E}_n \\ \hline 0 \quad \cdots & 0 & \cdots \quad 0 \\ \hline \frac{1}{\alpha}\mathrm{E}_n & \begin{matrix} \vdots \\ 0 \end{matrix} & 0 \end{array}\right). $$

Therefore T is an automorphism iff $\underline{S} = {}^t\underline{S}$ and $\overline{S} = {}^t\overline{S}$. In other words

$$ \begin{array}{rcl} s_{ij} & = & s_{ji} \\ s_{i+1,j} & = & s_{j+1,i} \end{array} \qquad \text{for } 1 \le i, j \le n, $$

so

$$ s_{ij} = s_{ji} = s_{(j-1)+1,i} = s_{i+1,j-1} $$

for $j > 1$ and $i < n$, hence $s_{ij} = s_{kl}$ for $i + j = k + l$.

The statement about the group action follows from

$$ \begin{pmatrix} \alpha\mathrm{E}_n & 0 \\ S & \mathrm{E}_{n+1} \end{pmatrix} \begin{pmatrix} \alpha'\mathrm{E}_n & 0 \\ S' & \mathrm{E}_{n+1} \end{pmatrix} = \begin{pmatrix} \alpha\alpha'\mathrm{E}_n & 0 \\ \alpha'S + S' & \mathrm{E}_{n+1} \end{pmatrix}. \qquad \square $$

Collecting the results we have

Theorem 6.6 *The automorphism group of $\mathbb{G}(1,2n)\cap L$ is an extention of $\mathbb{P}\mathrm{GL}(2,\mathbb{C})$ by the semi direct product $\mathbb{C}^{2n}\ltimes\mathbb{C}^*$.*

It is isomorphic to the matrix subgroup of $\mathbb{P}\mathrm{GL}(2n+1,\mathbb{C})$ given by

$$\begin{pmatrix} \alpha \mathrm{E}_n & 0 \\ S & \mathrm{E}_{n+1} \end{pmatrix}\begin{pmatrix} {}^t t_n^{-1} & 0 \\ 0 & t_{n+1} \end{pmatrix}$$

where $\alpha\in\mathbb{C}^*$, $S\in\mathrm{M}((n+1)\times n,\mathbb{C})$ with $s_{ij}=s_{kl}$ for $i+j=k+l$ and $t_n\in\mathrm{Aut}(h,\mathbb{P}_{n-1})$ resp. $t_{n+1}\in\mathrm{Aut}(c,\mathbb{P}_n)$ are the transformations that are induced by the $\mathbb{P}\mathrm{GL}(2,\mathbb{C})$ action on the rational normal curve $h\subset\mathbb{P}_{n-1}$ resp. $c\subset\mathbb{P}_n$.

Proof. It remains to show that the automorphism fixing the center curve point-wise form a normal subgroup, but that can be easily computed. □

Remark 6.7 *An automorphism of $\mathbb{G}(1,2n)\cap L$ is determined by its action on the lines intersecting the center curve.*

In contrast to that, the line system, i.e. the position of the line $L^*\subset\mathbb{P}(\bigwedge^2\mathbb{C}^{2n+1})^*$, is not determined by these lines, as a simple dimension count shows. Giving these lines is equivalent to giving the two rational curves $c\subset\mathbb{P}_{2n}$ and $h\subset\mathbb{P}_{2n}/P\cong\mathbb{P}_{n-1}$ and a correspondence between them, so that we have the following dimension count

$$(2(2n+1)-4)+(2n-4)+3<\dim\mathbb{G}(1,\mathbb{P}(\textstyle\bigwedge^2\mathbb{C}^{2n+1}))=2\left(\binom{2n+1}{2}-2\right).$$

Proof of the remark. We need to show that only the identity fixes these lines one by one. First a transformation T that fixes the lines must fix the center curve, hence by the Lemma 6.5 it is of the form

$$T=\begin{pmatrix} \alpha\mathrm{E}_n & 0 \\ S & \mathrm{E}_{n+1} \end{pmatrix}.$$

We compute the induced action $\widetilde{T}$ of T on $\{l\in\mathbb{G}(1,2n)\cap L\mid c_{(0:1)}\in l\}$, the $\mathbb{P}_{2n-2}$ of lines through $c_{(0:1)}=e_n$. A line $l\in\mathbb{G}(1,2n)$ through $c_{(0:1)}$ will be in the line system $\mathbb{G}(1,2n)\cap L$ iff it lies in the hyperplane $h_{(0:1)}=\ker(1:0:\ldots:0)$. Therefore the $\mathbb{P}_{2n-2}$ of lines through e_n is given by

$$e_n\wedge x \quad \text{with } x\in\mathbb{P}(\mathrm{span}\{e_1,\ldots,e_{n-1},e_{n+1},\ldots,e_{2n}\}).$$

Using $(e_1\wedge e_n,\ldots,e_{n-1}\wedge e_n,e_{n+1}\wedge e_n,\ldots,e_{2n}\wedge e_n)$ as a basis, the induced action $\widetilde{T}$ is

$$\widetilde{T}=\begin{pmatrix} \alpha\mathrm{E}_{n-1} & 0 \\ |\overline{S} & \mathrm{E}_n \end{pmatrix}.$$

Here $\lvert\overline{S}$ denotes the matrix S with the first row and column deleted. In order to have $\widetilde{T} = \mathrm{E}_{2n-1}$, we must have $\alpha = 1$ and $\lvert\overline{S} = 0$.

The same computation for the lines through $c_{(1:0)} = e_{2n}$ yields $\alpha = 1$ and $\underline{S}\rvert = 0$ from which $S = 0$ and the remark follow. □

For the rest of the section we analyze the action of the automorphism group on the line system $\mathbb{G}(1,2n) \cap L$. We start with $\mathbb{G}(1,4) \cap L$.

Proposition 6.8 *The action of* $\mathrm{Aut}(\mathbb{G}(1,4) \cap L)$ *on the lines has four orbits:*

1. *tangents of the center conic*
2. *secants of the center conic*
3. *lines through the center conic that do not lie in the plane of the center conic*
4. *lines that do not intersect the plane of the center curve.*

Proof. Since any automorphism maps the center conic onto itself, it is clear by the geometric description that all the mentioned lines lie in different orbits.

Any line in the plane P of the center conic intersects the conic twice, so by the Remark 6.2 it is a member of the line system. Since the automorphism group acts like $\mathrm{Aut}(c, P) \cong \mathbb{P}\mathrm{GL}(2, \mathbb{C})$ on the plane P, the first two orbits are obvious.

To see that the lines of 3) form one orbit, we have to exhibit an automorphism that given two lines of 3) maps one onto the other. Since the $\mathbb{P}\mathrm{GL}(2, \mathbb{C})$ part of the automorphism group acts transitively on the center conic, we may assume that both lines pass through the same point of the center conic, say $e_2 = c_{(0:1)}$. Now the induced action $\widetilde{T}$ of an automorphism T fixing the center conic pointwise on the $\mathbb{P}_2$ of lines through e_2 was computed in the proof of the Remark 6.7 as

$$\widetilde{T} = \begin{pmatrix} \alpha & 0 & 0 \\ f & 1 & 0 \\ g & 0 & 1 \end{pmatrix} \quad \text{with } \alpha \in \mathbb{C}^*;\ f, g \in \mathbb{C}.$$

These transformations act transitively on $\mathbb{P}_2 \setminus \mathbb{P}(\mathrm{span}\,\{\widetilde{e_1}, \widetilde{e_2}\})$, where the line $\mathbb{P}(\mathrm{span}\,\{\widetilde{e_1}, \widetilde{e_2}\})$ corresponds to the lines through e_2 that lie in the plane of the center conic.

The lines of 4) are all the remaining lines since there are no lines that intersect the plane P of the center conic but not the conic c itself. This is clear because the $\mathbb{P}_1$ of lines through a point $p \in P \setminus c$ is formed by the lines through p in the plane P, so there can be no other line.

Finally, we have to show that the lines of 4) form one orbit. By a small computation one checks that $\mathrm{Aut}(\mathbb{G}(1,4) \cap L) \subset \mathbb{P}\mathrm{GL}(5, \mathbb{C})$ acts transitively on

$\mathbb{P}_4\setminus P$. So, it suffices to show that the line $e_0\wedge e_1$ can be mapped to any other line through e_0 by an automorphism. Any of these lines can be written as

$$e_0\wedge(e_1+\beta e_4)\qquad \text{with } \beta\in\mathbb{C},$$

and the automorphism

$$T=\left(\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ \hline 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & \beta & 0 & 0 & 1 \end{array}\right)$$

will take $e_0\wedge e_1$ to it. □

PROPOSITION 6.9 *The automorphism group acts quasihomogeneously on* $\mathbb{G}(1,6)\cap H^2$.

Proof. For this it is enough to show that the stabiliser of the line $l=e_0\wedge e_2$ is a 2-dimensional subgroup since then

$$\begin{aligned}\dim\operatorname{Orbit}(l) &= \dim\operatorname{Aut}(\mathbb{G}(1,6)\cap H^2)-\dim\operatorname{Stab}(l)=10-2=8\\ &=\dim\mathbb{G}(1,6)\cap H^2.\end{aligned}$$

If we normalize the $t\in\mathbb{P}\mathrm{GL}(2,\mathbb{C})$ by $\det t=1$, every $T\in\operatorname{Aut}(\mathbb{G}(1,6)\cap H^2)$ can be written by the Theorem 6.6 as

$$T=\left(\begin{array}{ccc|cccc} \alpha & & 0 & & & & \\ & \alpha & & & & 0 & \\ 0 & & \alpha & & & & \\ \hline e & f & g & 1 & & & \\ f & g & h & & 1 & 0 & \\ g & h & i & & & 1 & \\ h & i & j & 0 & & & 1 \end{array}\right)\begin{pmatrix} {}^t t_3^{-1} & 0\\ 0 & t_4\end{pmatrix}$$

$$\text{with } {}^t t_3^{-1}=\begin{pmatrix} a^2 & -ab & b^2\\ -2ac & ab+cd & -2bd\\ c^2 & -cd & d^2\end{pmatrix}.$$

To compute the stabilizer we start by looking only at the first three entries of

$$\begin{aligned} Te_0 &= (\alpha a^2,-2\alpha ac,\alpha c^2,\ldots)\\ Te_2 &= (\alpha b^2,-2\alpha bd,\alpha d^2,\ldots).\end{aligned}$$

Since we must have $Te_0,Te_2\in l$, $ac=bd=0$ follows. By $\det t=ad-bc=1$ we have the two possibilities $b=c=0$, $d=a^{-1}$ and $a=d=0$, $c=-b^{-1}$. We examine only the first case, the second being similar. Now we have

$$\begin{aligned} Te_0 &= a^2(\alpha,0,0,e,f,g,h)\\ Te_2 &= a^2(0,0,\alpha,g,h,i,j).\end{aligned}$$

From $Te_0, Te_2 \in l$ we conclude $e = f = g = h = 0$ resp. $g = h = i = j = 0$. Therefore, including the case $(a = d = 0,\ c = -b^{-1})$, the stabilizer of l is

$$\operatorname{Stab}(l) = \left\{ \begin{pmatrix} \alpha a^2 & & & & & & \\ & 0 & & & & \text{\Large 0} & \\ & & \alpha a^{-2} & & & & \\ & & & a^{-3} & & & \\ & & & & a^{-1} & & \\ & \text{\Large 0} & & & & a & \\ & & & & & & a^3 \end{pmatrix}, \begin{pmatrix} 0 & \alpha b^2 & & & & & \\ & 0 & & & \text{\Large 0} & & \\ \alpha b^{-2} & 0 & & & & & \\ & & & & & & -b^{-3} \\ & & & & & b^{-1} & \\ & \text{\Large 0} & & & -b & & \\ & & & b^3 & & & \text{\Large 0} \end{pmatrix} \right\}. \qquad \square$$

Proposition 6.10 *For $n \geq 4$ the action of the automorphism group on $\mathbb{G}(1,2n) \cap H^2$ is not quasihomogeneous.*

Proof. We project the $\mathbb{P}_{2n}$ from the space P of the center curve onto $\mathbb{P}_{2n}/P \cong \mathbb{P}_{n-1}$. This projects the lines of $\mathbb{G}(1,2n) \cap H^2$ not intersecting P surjectively onto the lines $\mathbb{G}(1, \mathbb{P}_{2n}/P)$ of $\mathbb{P}_{2n}/P$. The automorphisms of $\mathbb{G}(1,2n) \cap H^2$ induce automorphisms of $\mathbb{P}_{2n}/P$. As matrices these are the upper left $n \times n$ matrices of the matrices of Theorem 6.6, i.e. they are of the form ${}^t t_n^{-1}$. So, as a group this induced automorphism group is isomorphic to $\operatorname{Aut}(h^*, \mathbb{P}_{2n}/P) \cong \mathbb{P}\mathrm{GL}(2,\mathbb{C})$. If $\operatorname{Aut}(\mathbb{G}(1,2n) \cap H^2)$ acts quasihomogeneously, then $\operatorname{Aut}(h^*, \mathbb{P}_{2n}/P)$ would have to act quasihomogeneously on $\mathbb{G}(1, \mathbb{P}_{2n}/P) \cong \mathbb{G}(1, n-1)$, but this contradicts

$$\dim \mathbb{P}\mathrm{GL}(2,\mathbb{C}) = 3 < \dim \mathbb{G}(1, n-1) = 2n-4. \qquad \square$$

7 $\mathbb{G}(1,4) \cap H^3$

Let $L = H^3$ be a general 3-codimensional subspace of $\mathbb{P}(\bigwedge^2 \mathbb{C}^5) \cong \mathbb{P}_9$. To L corresponds the plane $L^* = \mathbb{P}(\lambda A + \mu B + \nu C) \subseteq \mathbb{P}(\bigwedge^2 \mathbb{C}^5)^*$ of hyperplanes containing L. Since the locus of antisymmetric matrices of corank 3 is 3-codimensional in $\mathbb{P}(\bigwedge^2 \mathbb{C}^5)^*$ by Corollary 1.7, L^* does not contain any. Therefore to each of the hyperplanes $H_{(\lambda:\mu:\nu)} \subset L$ corresponds a unique center $c_{(\lambda:\mu:\nu)} \in \mathbb{P}_4$. In complete analogy to the last case we get

Lemma 7.1 *The centers $c_{(\lambda:\mu:\nu)}$ are those points of $\mathbb{P}_4$ through which there passes a $\mathbb{P}_1$ of lines of the line system $\mathbb{G}(1,4) \cap L$. Through all the other points passes a unique line.*

Proposition 7.2 *The map of centers*

$$\begin{array}{cccc} c: & L^* \cong \mathbb{P}_2 & \longrightarrow & \mathbb{P}_4 \\ & (\lambda : \mu : \nu) & \longmapsto & c_{(\lambda:\mu:\nu)} = \ker(\lambda A + \mu B + \nu C) \end{array}$$

is an embedding of $\mathbb{P}_2$ in $\mathbb{P}_4$ of degree 2, i.e. its image is a smooth projected Veronese surface.

Remark 7.3 *Any line that contains three centers is in the line system.*

Proof. Let $c(p_0)$, $c(p_1)$ and $c(p_2)$ with $p_0,p_1,p_2 \in L^*$ be the three centers on the line l. By the definition of the centers we have $l \in H_{p_i}$. Since c maps lines in L^* onto conics in $\mathbb{P}_4$, the three points p_0,p_1,p_2 do not lie on a line, hence they span L^*. So $l \in H_{p_0}\cap H_{p_1}\cap H_{p_2} = L$. □

From the statements we get a complete picture of the lines of $\mathbb{G}(1,4)\cap H^3$ in $\mathbb{P}_4$. We define the trisecant variety $\mathrm{Tri}(X)$ of a variety $X \subseteq \mathbb{P}_N$ by:

$$\mathrm{Tri}(X) := \overline{\{l \in \mathbb{G}(1,N) \mid \#(X\cap l) \geq 3\}}$$

Then we have

Corollary 7.4 (Castelnuovo) $\mathbb{G}(1,4)\cap L$ *is the trisecant variety of the smooth projected Veronese surface* $\operatorname{Im} c \subset \mathbb{P}_4$.

Proof.(see [C] or [SR, X, 4.4]) By the remark above the trisecant variety is contained in the irreducible variety $\mathbb{G}(1,4)\cap L$. So it is enough to show that both varieties have the same dimension. The Lemma 7.1 together with the Proposition 7.2 shows that there is an unique line of $\mathbb{G}(1,4)\cap L$ through a general point of $\mathbb{P}_4$ and that $\mathbb{G}(1,4)\cap L$ is the closure of such lines. The same statement for the trisecant variety is classical [SR, VII,3.2]. Hence both varieties have dimension three. □

The general trisecant intersects the projected Veronese surface $\operatorname{Im} c$ in three different points. Their inverse image under c are triples of points in $\mathbb{P}_2$ that have to fulfill some conditions since there is only a 3-dimensional family of these triples. To see what these conditions are, we recall some facts about the Veronese surface [H].

The Veronese surface V is the image of the embedding

$$\begin{array}{cccc} \upsilon: & \mathbb{P}_2 = \mathbb{P}(\mathbb{C}^3) & \longrightarrow & \mathbb{P}(\mathrm{Sym}^2\mathbb{C}^3) \\ & \mathbb{P}(v) & \longmapsto & \mathbb{P}(v\cdot v)\,. \end{array}$$

Its secant variety consists of the points of $\mathbb{P}(\mathrm{Sym}^2\mathbb{C}^3)$ that are the product of two vectors of $\mathbb{C}^3$,

$$\mathrm{Sec}(V) = \left\{\mathbb{P}(v\cdot w) \mid v,w \in \mathbb{C}^3\setminus\{0\}\right\}.$$

The projected Veronese surface will be smooth – like in our case – iff the center of projection P is not in the secant variety.

An intersection of the Veronese surface V with a hyperplane $H \in \mathbb{P}(\mathrm{Sym}^2\mathbb{C}^3)^*$ gives the conic $\upsilon^{-1}(V\cap H) \subset \mathbb{P}_2$ which is described by the equation H if we identify $\mathbb{P}(\mathrm{Sym}^2\mathbb{C}^3)^*$ with the polynomials of degree 2 modulo $\mathbb{C}^*$. The conics that we get as hyperplane sections of the projected Veronese surface are precisely the conics that we get as hyperplane sections of the Veronese surface by hyperplanes that contain the projection center P. So these conics fulfill one linear condition given by P. We view $P \in \mathbb{P}(\mathrm{Sym}^2\mathbb{C}^3)$ as a conic C_P^* in $\mathbb{P}_2^*$. Since P does not lie in the secant variety of the Veronese surface, P is not the product of two elements of $\mathbb{C}^3$, hence C_P^* is not the union of two lines. Therefore it is smooth. We denote the dual conic of C_P^* by $C_P \subset \mathbb{P}_2$.

Now, three different points of the projected Veronese surface $\mathrm{Im}\, c \subset \mathbb{P}_4$ lie on a line, the trisecant, iff any hyperplane that contains two of them contains all three. Under the inverse of the embedding c that means the following on the $\mathbb{P}_2$:

Three different points of $\mathbb{P}_2$ are the inverse image $c^{-1}(l)$ of a trisecant l of the projected Veronese surface $\mathrm{Im}\, c$ iff all conics that fulfill the linear condition given by P (or equivalently by C_P) and pass through two of the points pass through all three of them.

The propositions in the appendix tell us that these triples of points are the vertices of the non-degenerated polar triangles of the conic C_P. By a continuity argument the trisecants that are tangent to $\mathrm{Im}\, c$ at one point and intersect it in another correspond to the degenerated polar triangles, and the trisecants that intersect $\mathrm{Im}\, c$ in only one point "with multiplicity three" correspond to a triple point on the conic C_P. We also see that there are no 4-secants. Since if there is one, there would be four points in $\mathbb{P}_2$ such that any three of them build a different polar triangle. But this is impossible because a polar triangle is already determined by two of its vertices.

With this geometric description it is easy to compute the automorphism group of $\mathbb{G}(1,4) \cap H^3$. By Theorem 1.2 any automorphism is induced by a projective linear transformation of H^3. Hence it maps the $\mathbb{P}_1$'s of lines of Lemma 7.1 onto themselves and induces thereby an automorphism of the set of centers, the projected Veronese surface $\mathrm{Im}\, c$. Since the $\mathbb{G}(1,4) \cap H^3$ consists of the trisecants of $\mathrm{Im}\, c$, this automorphism determines the original automorphism of $\mathbb{G}(1,4) \cap H^3$.

In order to complete the proof of Corollary 1.3 we have to show that the automorphism of the projected Veronese surface $\mathrm{Im}\, c \subset \mathbb{P}_4$ is induced by one of $\mathbb{P}_4$. Since c is the composition of the Veronese map v and the projection $\pi_P : \mathbb{P}_5 \to \mathbb{P}_4$, we consider the situation in the $\mathbb{P}_5$. Since $\mathrm{Im}\, c \cong \mathbb{P}_2 \cong v(\mathbb{P}_2)$ and $\mathrm{Aut}(\mathbb{P}_2) \cong \mathrm{Aut}(v(\mathbb{P}_2), \mathbb{P}_5)$ [H, 10.9], the automorphism of $\mathrm{Im}\, c$ induces an automorphism of $\mathbb{P}_5$. We know further that the automorphism of $\mathrm{Im}\, c$ maps the trisecants of $\mathrm{Im}\, c$ onto themselves. In the $\mathbb{P}_5$ this means that it preserves the planes that contain the projection center P and three points of the Veronese surface, so it must fix their common intersection, the projection center P. Hence the automorphism descends onto the $\mathbb{P}_4$.

Now under the inverse of the embedding c, the automorphism of $\operatorname{Im} c$ preserving the trisecants corresponds to an automorphism of $\mathbb{P}_2$ preserving the polar triangles of the conic $C_P \subset \mathbb{P}_2$. Such an automorphism of $\mathbb{P}_2$ maps the degenerated polar triangles onto themselves. In particular, it maps the tangents to the conic C_P onto themselves. Therefore it has to fix the conic C_P.

So we have seen how an automorphism of $\mathbb{G}(1,4) \cap L$ induces an unique automorphism of $\mathbb{P}_2$ fixing C_P, hence an automorphism of C_P since $\operatorname{Aut}(C_P, \mathbb{P}_2) \cong \operatorname{Aut}(C_P) \cong \mathbb{P}\mathrm{GL}(2,\mathbb{C})$.

On the other hand, any projective linear transformation of $\mathbb{P}_2$ that fixes the conic C_P preserves the polar triangles of C_P. Therefore it induces via the embedding c an automorphism of the projected Veronese surface $\operatorname{Im} c$ that preserves triples of points that lie on a line. So it defines an automorphism of the trisecants of $\operatorname{Im} c$, which is the same as an automorphism of $\mathbb{G}(1,4) \cap L$.

We summarize:

THEOREM 7.5 *The automorphism group of* $\mathbb{G}(1,4) \cap H^3$ *is isomorphic to* $\mathbb{P}\mathrm{GL}(2,\mathbb{C})$.

The description of the orbits of this automorphism group follows immediately.

PROPOSITION 7.6 *The action of* $\operatorname{Aut}(\mathbb{G}(1,4) \cap H^3)$ *on the linear system* $\mathbb{G}(1,4) \cap H^3$ *has three orbits:*

1. *trisecants of the projected Veronese surface that intersect it in three points*
2. *trisecants that are tangent to the projected Veronese surface at one point and intersect it in another*
3. *trisecants that intersect the projected Veronese surface in only one point "with multiplicity three".*

Proof. By what was said above, this is equivalent to the classical statement that the action of group $\operatorname{Aut}(C_P, \mathbb{P}_2)$ on the polar triangles has three orbits: the non-degenerated triangles, the degenerated ones and the triple points on C_P. □

8 APPENDIX: POLAR TRIANGLES

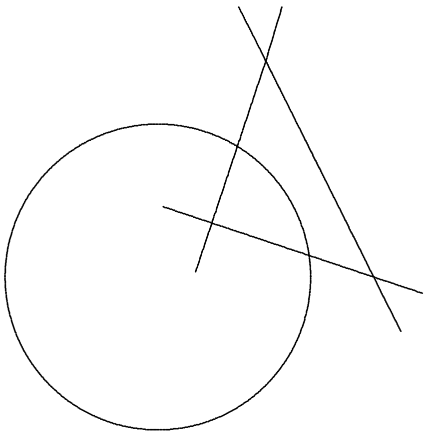
Here we prove the needed propositions about polar triangles. The whole appendix may be seen as a modern exposition of [SF, 348]. First we recall the basic definitions.

Let C_A be a smooth conic in $\mathbb{P}_2$, which is given by the quadratic equation ${}^t x A x = 0$, where $A \in \mathrm{GL}(3,\mathbb{C})$ is a symmetric, invertible matrix. Then C_A induces a polarity P by

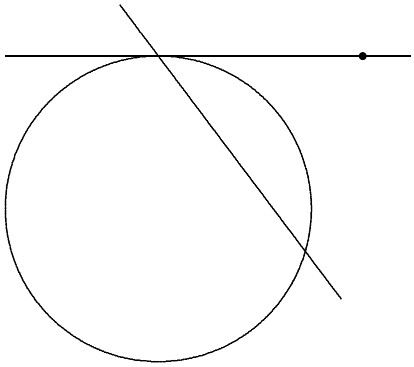
$$\begin{array}{rccc} P: & \mathbb{P}_2 & \longrightarrow & \mathbb{P}_2^* \\ & \mathbb{P}(x) & \longmapsto & \mathbb{P}({}^t x A)\,. \end{array}$$

For a point $p \in \mathbb{P}_2$ the line $P(p)$ is called the polar of p and p the pole of $P(p)$.

A polar triangle is given by three points, at least two of which are different, such that the polar of each point contains the other two points, i.e. (p, q, r) is a polar triangle if ${}^t\!pAq = {}^t\!qAr = {}^t\!rAp = 0$. The sides of the triangle are the polars of the points. In the non-degenerated case when all three points are different, the three points cannot lie on a line and therefore span the whole $\mathbb{P}_2$. In the degenerated case, (p, p, q), p lies on the conic and q on the tangent to the conic at the point p. The sides are the polar of q and twice the tangent.



Non-degenerated polar triangle



Degenerated polar triangle

Proposition 8.1 *Let $C_A = \{x \in \mathbb{P}_2 \mid {}^t\!xAx = 0\}$ be a smooth conic and $C_A^* = \{x \in \mathbb{P}_2^* \mid {}^t\!xA^{-1}x = 0\}$ its dual conic. Further, let $C_B = \{x \in \mathbb{P}_2 \mid {}^t\!xBx = 0\}$ be a conic such that*

$$\sum_{i,j=0}^{2} a^{ij} b_{ij} = 0 \qquad (*)$$

where $A^{-1} = (a^{ij}), B = (b_{ij}) \in \mathrm{Sym}(3, \mathbb{C})$. Finally, let (p, q, r) be a polar triangle of C_A then:

If two of the three points p, q, r lie on the conic C_B, then also the third.

In the case of a degenerate polar triangle, (p, p, q), the condition that C_B contains p twice means that C_B contains p and C_B is either singular in p or its tangent in b is the polar of q.

Proof. One can show that all the properties in the statement of the proposition are independent of the choice of coordinates, so we may pick nice ones. We have to distinguish between the two cases of the polar triangle being degenerated or not. We treat the case of the non-degenerated polar triangle first.

By a suitable choice of coordinates we may assume that $p=(1:0:0)$, $q=(0:1:0)$ and $r=(0:0:1)$. Then the assumption that (p,q,r) is a polar triangle of C_A translates into

$$ {}^t\!pAq = {}^t\!qAr = {}^t\!rAp = 0 \Longleftrightarrow a_{01}=a_{12}=a_{02}=0. $$

By a scaling of the coordinates we can achieve that $a_{00}=a_{11}=a_{22}=1$, so that $C_A=\{x\in\mathbb{P}_2 \mid x_0^2+x_1^2+x_2^2=0\}$. Then the condition $(*)$ reads $b_{00}+b_{11}+b_{22}=0$. If the two points p and q are on the conic C_B, we have ${}^t\!pBp=b_{00}=0$ and ${}^t\!qBq=b_{11}=0$. By $(*)$ we see $0=b_{22}={}^t\!rBr$, i.e. the third point r lies also on the conic C_B.

Now we treat the case of the degenerated polar triangle (p,p,q). We choose coordinates such that $C_A=\{x\in\mathbb{P}_2 \mid x_0^2+x_1^2+x_2^2=0\}$ and $p=(1:i:0)$. The point $q\neq p$ must lie on the tangent to C_A. So it has coordinates $q=(\lambda:i\lambda:1)$, and its polar is spanned by p and $(1:0:-\lambda)$. Now using the assumptions

$$ b_{00}+b_{11}+b_{22}=0 \qquad (*) $$

$$ p\in C_B \Longleftrightarrow {}^t\!pBp=0 \Longleftrightarrow b_{00}+2ib_{01}-b_{11}=0, \qquad (**) $$

we have to show

$$ q\in B \Longleftrightarrow C_B \text{ singular in } p \quad \text{or} \quad \mathbb{T}_pC_B = \text{ polar of } q. $$

We rewrite this as

$$ {}^t\!qBq = {}^t\!\begin{pmatrix}\lambda\\ i\lambda\\ 1\end{pmatrix} B \begin{pmatrix}\lambda\\ i\lambda\\ 1\end{pmatrix} = 0 \Longleftrightarrow {}^t\!pB\begin{pmatrix}1\\ 0\\ -\lambda\end{pmatrix} = {}^t\!\begin{pmatrix}1\\ i\\ 0\end{pmatrix} B \begin{pmatrix}1\\ 0\\ -\lambda\end{pmatrix} = 0. $$

But this is true since -2 times the left hand side plus (λ^2+1) times $(**)$ plus $(*)$ gives the right hand side. $\square$

Now we prove the converse of the last proposition.

Proposition 8.2 *Given a smooth conic $C_A=\{x\in\mathbb{P}_2 \mid {}^t\!xAx=0\}$*

$$ \mathcal{B} := \left\{ C_B=\{x\in\mathbb{P}_2 \mid {}^t\!xBx=0\} \;\middle|\; \sum_{i,j=0}^{2} a^{ij}b_{ij}=0 \right\} $$

is a four dimensional family of conics. Let $p,q,r\in\mathbb{P}_2$ be three points, at least two of which are different, with the property that if two of them lie on a conic $C_B\in\mathcal{B}$ then also the third.

Then (p,q,r) is a polar triangle of C_A.

Proof. We will show that if (p, q, r) is not a polar triangle then there exits a $C_A \in \mathcal{B}$ for which this property is violated. We have to treat several cases.

First let the three points be all different, then they cannot lie on a line. Because if they would, the conics in the at least two dimensional family

$$\mathcal{B}_{p,q} := \{C_B \in \mathcal{B} \mid p, q \in C_B\}$$

of conics of $\mathcal{B}$ passing through p and q must split off the line through the three points. If we pick coordinates such that this line is given by $\{x_2 = 0\}$, then $\mathcal{B}_{p,q}$ must be

$$\mathcal{B}_{p,q} = \{\mathrm{V}(x_2(\lambda_0 x_0 + \lambda_1 x_1 + \lambda_2 x_2)) \mid (\lambda_0 : \lambda_1 : \lambda_2) \in \mathbb{P}_2\}\,.$$

This means that the conics C_B with the matrices

$$B = \begin{pmatrix} 0 & 0 & b_{02} \\ 0 & 0 & b_{12} \\ b_{02} & b_{12} & b_{22} \end{pmatrix} \quad \text{with } b_{02}, b_{12}, b_{22} \in \mathbb{C}$$

are all in $\mathcal{B}$. Hence the matrix A^{-1} must be of the type

$$A^{-1} = \begin{pmatrix} a^{00} & a^{01} & 0 \\ a^{01} & a^{11} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{with } a^{00}, a^{01}, a^{11} \in \mathbb{C},$$

but this contradicts the invertibility of A^{-1}.

Now since p, q, r span the $\mathbb{P}_2$ we may pick coordinates such that $p = (1:0:0)$, $q = (0:1:0)$ and $r = (0:0:1)$. That (p, q, r) is not a polar triangle of C_A means that ${}^t\!pAq = a_{01} \neq 0$, ${}^t\!qAr = a_{12} \neq 0$ or ${}^t\!rAp = a_{02} \neq 0$. Assuming $\det A = 1$ we conclude that not all of the $a^{01} = a_{02}a_{12} - a_{01}a_{22}$, $a^{02} = a_{01}a_{12} - a_{02}a_{11}$ and $a^{12} = a_{01}a_{02} - a_{00}a_{12}$ can be zero. If for example $a^{02} \neq 0$, then

$$B = \begin{pmatrix} 0 & 0 & -a^{11} \\ 0 & 2a^{02} & 0 \\ -a^{11} & 0 & 0 \end{pmatrix}$$

gives a conic $C_B \in \mathcal{B}$ that contains the points p and q, but not r.

Now let us look at the case where two of the points p, q, r are the same. The points (p, p, q) will not form a polar triangle if ${}^t\!pAp \neq 0$ or ${}^t\!pAq \neq 0$.

For the case ${}^t\!pAp \neq 0$ we pick coordinates such that A is the identity matrix and $p = (1:0:0)$. Let

$$B = \begin{cases} \begin{pmatrix} 0 & -q_2 & q_1 \\ -q_2 & 0 & 0 \\ q_1 & 0 & 0 \end{pmatrix} & \text{for } q_0 \neq 0 \text{ or } q_1^2 + q_2^2 \neq 0 \\[2ex] \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & \pm i \\ 1 & \pm i & -1 \end{pmatrix} & \text{for } q = (0:1:\pm i), \end{cases}$$

then C_B is a conic of $\mathcal{B}$ that contains p and q, but is smooth in p, and its tangent in p is not the polar of q, so it does not contain p twice.

Finally, if ${}^tpAp = 0$ and ${}^tpAq \neq 0$, we pick coordinates such that A is the identity matrix and $p = (1:i:0)$. Let

$$B = \begin{cases} \begin{pmatrix} -2i & 2 & iq_0 - 2q_1 \\ 2 & 2i & -iq_1 \\ iq_0 - 2q_1 & -iq_1 & 0 \end{pmatrix} & \text{for } q = (q_0 : q_1 : 1) \\ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} & \text{for } q_2 = 0, \end{cases}$$

then we are in the same situation as above. □

References

[B] Bourbaki, N.: Eléments de Mathématique XXIV, Formes Sesquilinéaires et Formes Quadratiques. Herman, Paris 1959.

[BK] Brieskorn, E. & H. Knörrer: Plane Algebraic Curves. Birkhäuser, Basel 1986.

[Br] Brieskorn, E.: Lineare Algebra und analytische Geometrie II. Vieweg, Wiesbaden 1985.

[C] Castelnuovo, G.: Geometria della retta nello spazio a quattro dimensioni. Atti del R. Ist. Veneto (7), 2 (1891), p. 855-901.

[D] Donagi, R.: On the Geometry of Grassmannians. Duke Math. J. 44 (1977), p. 795-837.

[FH] Fulton, W. & J. Harris: Representation Theory. Springer, New York 1991.

[H] Harris, J.: Algebraic Geometry. Springer, New York 1992.

[M] Mumford, D.: Some Footnotes to the Work of C. P. Ramanujam. C. P. Ramanujam - A Tribute. Springer, Berlin 1978.

[R] Roth, L.: Some properties of Grassmannians. Rend. de Mat. e Appl. V 10 (1951), p. 96-114.

[SF] Salmon, G. & W. Fiedler: Analytische Geometrie der Kegelschnitte II. Teubner Verlag, Leipzig 1918.

[SR] Semple, J. & L. Roth: Introduction to Algebraic Geometry. Oxford University Press, Oxford 1985.

Mathematisches Institut IV
Heinrich-Heine-Universität
Universitätsstr. 1
40239 Düsseldorf
Germany
piontkow@uni-duesseldorf.de

Faculteit der Wiskunde en
Natuurwetenschappen
Afdeling Wiskunde en Informatica
Postbus 9215
2300 RA Leiden
Netherlands
VEN@rulwinw.leidenuniv.nl

Some Properties of the Symmetric Enveloping Algebra of a Subfactor, with Applications to Amenability and Property T

Sorin Popa

Received: September 27, 1999

Communicated by Joachim Cuntz

Abstract. We undertake here a more detailed study of the structure and basic properties of the symmetric enveloping algebra $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ associated to a subfactor $N \subset M$, as introduced in [Po5]. We prove a number of results relating the amenability properties of the standard invariant of $N \subset M, \mathcal{G}_{N,M}$, its graph $\Gamma_{N,M}$ and the inclusion $M \vee M^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$, notably showing that $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ is amenable relative to its subalgebra $M \vee M^{\mathrm{op}}$ iff $\Gamma_{N,M}$ (or equivalently $\mathcal{G}_{N,M}$) is amenable, i.e., $\|\Gamma_{N,M}\|^2 = [M : N]$. We then prove that the hyperfiniteness of $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ is equivalent to M being hyperfinite and $\Gamma_{N,M}$ being amenable. We derive from this a hereditarity property for the amenability of graphs of subfactors showing that if an inclusion of factors $Q \subset P$ is embedded into an inclusion of hyperfinite factors $N \subset M$ with amenable graph, then its graph $\Gamma_{Q,P}$ follows amenable as well. Finally, we use the symmetric enveloping algebra to introduce a notion of property T for inclusions $N \subset M$, by requiring $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ to have the property T relative to $M \vee M^{\mathrm{op}}$. We prove that this property doesn't in fact depend on the inclusion $N \subset M$ but only on its standard invariant $\mathcal{G}_{N,M}$, thus defining a notion of property T for abstract standard lattices $\mathcal{G}$.

1991 Mathematics Subject Classification: Primary: 46L37, secondary: 46L40

0. Introduction

Let $N \subset M$ be an inclusion of type II_1 factors with finite Jones index, $[M : N] < \infty$, and extremal. In short, its *symmetric enveloping* von Neumann algebra $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ is the unique (up to isomorphism) type II_1 factor S, generated by mutually commuting copies of M, M^{op} that satisfy $M' \cap S = M^{\mathrm{op}}$, $M^{\mathrm{op}\prime} \cap S = M$ and by a projection e_N which implements, at the same time, both the trace preserving expectation E_N of M onto N and the trace preserving expectation $E_{N^{\rm op}}$ of $M^{\rm op}$ onto $N^{\rm op}$.

One can construct this factor by first taking the C^*-algebra S_0 generated on the Hilbert space $L^2(M)$ by the operators of left and right multiplication by elements in M and by the orthogonal projection of $L^2(M)$ onto $L^2(N)$, then proving that there exists a unique trace τ on this C^*-algebra and then defining $M \underset{e_N}{\boxtimes} M^{\rm op}$ to be the type $\rm II_1$ von Neumann factor obtained via the Gelfand-Naimark-Segal representation for (S_0, τ), i.e., $M \underset{e_N}{\boxtimes} M^{\rm op} \overset{def}{=} \overline{\pi_\tau(S_0)}$. This construction doesn't in fact depend on the (binormal) representation of the triple $(N \subset M, e_N, M^{\rm op} \supset N^{\rm op})$ that one starts with: any $M - M$ bimodule with an e_N-type projection on it, instead of $L^2(M)$, will do, provided certain obvious compatibility conditions for the commutants are satisfied.

The following exemple of symmetric enveloping algebras is quite relevant: if $N \subset M$ is an inclusion associated to a finitely generated discrete group G and an outer action σ of G on a type $\rm II_1$ factor P (see e.g., 5.1.5 in [Po2]) then $(M \vee M^{\rm op} \subset M \underset{e_N}{\boxtimes} M^{\rm op})$ is isomorphic to $(P\bar{\otimes}P^{\rm op} \subset P\bar{\otimes}P^{\rm op} \rtimes_{\sigma\otimes\sigma^{\rm op}} G)$. In general, one has an interpretation of the *symmetric enveloping inclusion* $M \vee M^{\rm op} \subset M \underset{e_N}{\boxtimes} M^{\rm op}$ that is very much the same as this crossed product situation.

The symmetric enveloping algebra $M \underset{e_N}{\boxtimes} M^{\rm op}$ and the inclusions $M \vee M^{\rm op} \subset M \underset{e_N}{\boxtimes} M^{\rm op}$ were introduced in ([Po5]) in order to provide an additional tool for studying subfactors of finite index. It proved to be particularily useful for relating the analysis aspects of the theory of subfactors to its combinatorial features.

We undertake here a more detailed study of these objects and use them to get more insight into the structure of subfactors, notably proving a number of results on the amenability and the property T for subfactors $N \subset M$ and for their associated combinatorial invariants: the standard graph $\Gamma_{N,M}$ and the standard invariant $\mathcal{G}_{N,M}$.

Thus, we prove that $\mathcal{G}_{N,M}$ is amenable (by definition this means that its graph $\Gamma_{N,M}$ is amenable, i.e., it satisfies the Kesten-type condition $\|\Gamma_{N,M}\|^2 = [M : N]$) if and only if $M \underset{e_N}{\boxtimes} M^{\rm op}$ is amenable relative to $M \vee M^{\rm op}$ in the sense of [Po8]. In fact, we establish a few more additional equivalent characterizations of the amenability for $\Gamma_{N,M}$: a Følner type condition; a local Shannon-McMillan-Breiman type condition; a local bicommutation condition; a characterization in terms of the representations of $N \subset M$.

We then study the amenability in the special case of subfactors $N \subset M$ for which the algebras N, M involved are assumed amenable (or, equivalently, by Connes theorem [C1], hyperfinite) type $\rm II_1$ factors. The key result along this line shows that the algebra $M \underset{e_N}{\boxtimes} M^{\rm op}$ is itself amenable if and only if both $\mathcal{G}_{N,M}$ and the single algebras N, M are amenable.

Again, some other characterizations of this situation are proved, notably an "injectivity"-type condition requiring $N \subset M$ to be the range of a norm one projection from its standard representation, or, equivalently, to be the range of a norm one projection from any of its (smooth) representations. Also, it is proved equivalent to an Effros-Lance type condition, requiring $S_0 = C^*(M, e_N, M') \subset \mathcal{B}(L^2(M))$ to be a simple C*-algebra . We call an inclusion of factors $N \subset M$ satisfying any of these conditions an *amenable inclusion.* While proving all these results we also show that if $N \subset M$ is amenable (i.e., N, M are hyperfinite and $\Gamma_{N,M}$ is amenable) then there exists a choice of a tunnel of subfactors $M \supset N \supset P_1 \supset P_2 ...$, obtained by taking downward basic constructions for certain induced-reduced algebras in the Jones tower (the choice being dictated by information contained in the standard invariant $\mathcal{G}_{N,M}$) such that the relative commutants $P_k' \cap N \subset P_k' \cap M$ exhaust $N \subset M$. This shows in particular that amenable subfactors are completely classified by their standard invariants $\mathcal{G}_{N,M}$ (see also [Po16]).

Next we derive the result that we regard as the most significant application of the methods developed in this paper, showing that the amenability in the category of inclusions of factors with finite Jones index $N \subset M$, with "morphisms" given by commuting square embeddings between such inclusions, is a hereditarity property. In the case one takes degenerate inclusions $N = M$ we recover Connes' hereditarity result for single hyperfinite type II$_1$ factors ([C1]). In terms of graphs, the result states that if an extremal inclusion of hyperfinite factors $N \subset M$ has amenable graph then any of its sub-inclusions $Q \subset P$ (i.e., $Q \subset P$ is embedded in $N \subset M$ as a commuting square) has amenable graph. It should be noted that the embedding of $Q \subset P$ into $N \subset M$ does not require $[P : Q] = [M : N]$, nor that $[M : P] < \infty$! This hereditarity property for the amenability of graphs is somewhat surprising and there is little that could appriorically predict it. It only holds true within the class of hyperfinite subfactors, as if we drop the amenability assumption on the ambient single algebras M involved it is no longer valid, in general.

Indeed, it is proved in ([Po7]) that given any abstract standard λ- lattice $\mathcal{G}$ and any of its sublattices $\mathcal{G}_0 \subset \mathcal{G}$, there exist subfactors $N \subset M$ and $N_0 \subset M_0$ and a commuting square inclusion of $N_0 \subset M_0$ into $N \subset M$, such that $\mathcal{G}_{N,M} = \mathcal{G}$ and $\mathcal{G}_{N_0,M_0} = \mathcal{G}_0$. But any standard λ-lattice $\mathcal{G}$ contains the Temperley-Lieb-Jones standard λ-lattice with graph A_∞, which is never amenable if $\lambda^{-1} > 4$. Thus if $\mathcal{G}$ is taken to be amenable, for instance to have finite depth, then $N \subset M$ has amenable graph while $N_0 \subset M_0$, which is embedded into it, doesn't. The reason is, of course, that in the examples of subfactors $N \subset M$ constructed in ([Po7]) the algebras N, M involved are not hyperfinite.

One consequence of the hereditarity result is that, for instance, one cannot embed subfactors $Q \subset P$ of index $\alpha > 4$ that are contructed by commuting squares of finite dimensional algebras like in ([Sc], [We]) and having graph $\Gamma_{Q,P}$ equal to A_∞ (note that by [H1] $\Gamma_{Q,P} = A_\infty$ if $\alpha < (5 + \sqrt{13})/2$) into hyperfinite subfactors of finite depth and index $> \alpha$. Also, by ([H2]) there exists a subfactor of index $\alpha = (5 + \sqrt{13})/2$, constructed from commuting squares of finite dimensional algebras and having graph A_∞, which thus, by our result, cannot be embedded into Haagerup's finite depth subfactor of same index $(5+\sqrt{13})/2$ ([H1]).

Our last application to the symmetric enveloping algebra approach is the consideration of a notion of property T for standard lattices. Thus, we prove that if a standard lattice $\mathcal{G}$ is given then $M \underset{e_N}{\boxtimes} M^{\rm op}$ has the property T relative to $M \vee M^{\rm op}$, in the sense of ([A-D],[Po8]), for some $N \subset M$ for which $\mathcal{G}_{N,M} = \mathcal{G}$, if and only if it has this property for any subfactor $N \subset M$ for which $\mathcal{G}_{N,M} = \mathcal{G}$. If $\mathcal{G}$ satisfies these conditions then we say that $\mathcal{G}$ has the *property* T. Note that this definition does not require the ambient factors involved to have the property T in the sense of Connes ([C4,5]). On the other hand, if $\mathcal{G}$ is a standard lattice coming from a discrete group G as described above, then $\mathcal{G}$ has the property T if and only if the group G has the property T in the classical sense of Kazhdan. Thus, our notion generalizes this notion, from discrete groups to the larger class of group-like objects $\mathcal{G}$. Our main result in this direction shows that if a sublattice $\mathcal{G}_0$ of a standard lattice $\mathcal{G}$ has the property T then $\mathcal{G}$ has the property T. As a consequence it follows that, generically, the Temperley-Lieb-Jones standard lattices with graph A_∞ do not have the property T.

Although we only work with type II_1 factors, many of the considerations in this paper can be suitably carried over to subfactors of type III (see the remarks 1.10.3°, 2.2.2°, 2.5.2°). The corresponding symmetric enveloping type III factors may prove to be a useful tool in the analysis of the Jones-Wassermann subfactors coming from representations of loop groups ([Wa], [Xu]). In a different direction, it would be interesting to relate the symmetric enveloping algebra associated to an extremal II_1 subfactor to Jones' affine Hecke algebra associated with that subfactor ([J3,4]). An explicit description of the symmetric enveloping algebras coming from certain special classes of subfactors ([BiH], [BiJ]) would certainly be most illuminating for getting some insight on this and other related problems.

The paper is organized in 9 Sections. In the first section we introduce the C^*-analogues of the symmetric enveloping algebras, needed in order to obtain the necessary universality properties and the functoriality of the von Neumann construction. A key ingredient for these considerations is the relative Dixmier property for subfactors of finite index, that we prove in the Appendix A.1.

In Sec. 2 we define the actual symmetric enveloping type II_1 factors (2.1, 2.2) and symmetric enveloping inclusions and prove their basic properties (2.6, 2.7, 2.9, 2.10). Also, we define a more general class of enveloping inclusions, in which to two given subfactors $N \subset M$ and $Q \subset P$ having the same higher relative commutant picture one associates their *concatenation* inclusion $M \vee P^{\rm op} \subset M \boxtimes P^{\rm op}$ (2.5.1°). We end that section by introducing a notion of index $[\mathcal{G} : \mathcal{G}_0]$ for sublattices $\mathcal{G}_0$ of standard lattices $\mathcal{G}$ (2.11, 2.12).

In Sec. 3 we discuss the example of symmetric enveloping algebras associated to subfactors coming from discrete groups acting outerly on factors, case in which it becomes an actual crossed-product algebra. In Sec. 4 we prove that even for general inclusions $N \subset M$ the corresponding symmetric enveloping algebras look very much like crossed products (4.5). Also, we prove some decomposition properties for such algebras, showing for instance that when N, M are hyperfinite then, regardless of whether $M \underset{e_N}{\boxtimes} M^{\text{op}}$ is hyperfinite or not, it is a *thin* factor, i.e., $M \underset{e_N}{\boxtimes} M^{\text{op}} = \overline{\text{sp}} R_1 R_2$, for some suitable hyperfinite subfactors R_1, R_2 (4.3). Also, we prove a general ergodicity property for the higher relative commutants of a subfactor which is quite useful in applications (4.8, 4.9).

In Sec. 5 and 6 we relate the amenability properties of $\mathcal{G}_{N,M}, \Gamma_{N,M}$ and $(M \vee M^{\text{op}} \subset M \underset{e_N}{\boxtimes} M^{\text{op}})$, obtaining a number of equivalent characterizations of the amenability for standard lattices and graphs (5.3, 6.1, 6.3, 6.4).

In Sec. 7 we discuss the case when $M \underset{e_N}{\boxtimes} M^{\text{op}}$ is hyperfinite, proving this equivalent to the amenability of $N \subset M$ and to various other properties of the representation theory of $N \subset M$ (7.1). For instance, we show that for hyperfinite subfactors it is enough that the universal graph $\Gamma^{u,rf}_{N,M}$ be amenable for $\Gamma_{N,M}$ to follow amenable (7.6). We also prove here the hereditarity property for amenable inclusions (7.5). The proof uses the characterization of the amenability for $N \subset M$ by the hyperfiniteness condition on $M \underset{e_N}{\boxtimes} M^{\text{op}}$, a fact that roughly reduces the argument to Connes' hereditarity of hyperfiniteness for single type II_1 factors. Sec.8 contains the proof of the Effros-Lance type characterization of amenability (8.1).

Finally, in Sec. 9 we introduce the property T for standard lattices and prove some results about this notion.

For most notations and general technical background we refer the reader to ([Po2,4,7]). More specific references are made in the text. For the reader's convenience we included an Appendix which, besides the already mentioned relative Dixmier property for subfactors of finite index, contains a generalized version of Connes' joint distribution trick needed in the proof of the Følner condition for graphs.

The results on amenability in this paper were presented by the author in lectures and seminars, during 1991-1997. A more formal announcement of these results, with sketches of proofs, appeared in [Po5], while a couple of statements on the equivalence of the definition of amenability with representations and the Kesten condition, respectively Følner condition, were already announced in [Po2], resp. [Po4]. A rather complete discussion of the role of amenability within the overall classification of subfactors, with a presentation of most of the results in this paper (including the ones on the property T) appeared in [Po11].

1. Symmetric Enveloping C^*-Algebras

In this and the next section we discuss the definition and basic properties of the symmetric enveloping algebras (C* in this section and von Neumann in the next) associated to an extremal inclusion of factors with finite index, as introduced in [Po5]. The statements below are similar to the ones in §1 of [Po5], but the proofs, that are only briefly sketched there, are given here in details.

So let $N \subset M$ be an inclusion of type II_1 factors with finite Jones index, $[M : N] < \infty$, which we assume to be extremal, i.e., $[pMp : Np] = [M : N]\tau(p)^2$, $\forall\, p \in \mathcal{P}(N' \cap M)$.

We denote by $M \subset M_1 = \langle M, e_N\rangle$ the (abstract) basic construction for $N \subset M$, e_N being the projection implementing the trace preserving conditional expectation E_N of M onto N.

We first construct the universal C^*-algebra generated by mutually commuting copies of M, M^{op} and an e_N-like projection implementing the expectations E_N, $E_{N^{\mathrm{op}}}$ on them.

A *representation* (π, π') of $(N \subset M, e_N, M^{\mathrm{op}} \supset N^{\mathrm{op}})$ is a pair of unital $*$-representations $\pi : M_1 \to \mathcal{B}(\mathcal{H})$, $\pi' : M_1^{\mathrm{op}} \to \mathcal{B}(\mathcal{H})$ such that $[\pi(M), \pi'(M^{\mathrm{op}})] = 0$, $\pi(e_N) = \pi'(e_N^{\mathrm{op}})$. Two such representations, (π_1, π_1') on $\mathcal{H}_1$ and (π_2, π_2') on $\mathcal{H}_2$, are equivalent if there exists a unitary $U : \mathcal{H}_1 \to \mathcal{H}_2$ such that $U\pi_1(x)U^* = \pi_2(x)$, $U\pi_1'(x)U^* = \pi_2'(x)$, $\forall\, x \in M_1$. A representation (π, π') on $\mathcal{H}$ is *cyclic* if $\exists \xi \in \mathcal{H}$ such that $\overline{\mathrm{Alg}(\pi(M_1), \pi'(M_1^{\mathrm{op}}))\xi} = \mathcal{H}$.

Note that if (π, π') is a representation on $\mathcal{H}$ then there exists a representation $(\bar{\pi}, \bar{\pi}')$ on the conjugate Hilbert space $\bar{\mathcal{H}}$ defined by $\bar{\pi}(x) = \overline{\pi'(x^{*\mathrm{op}})}$, $\bar{\pi}'(x^{\mathrm{op}}) = \overline{\pi(x^*)}$, $x \in M_1$, where $T \mapsto \bar{T}$ denotes the antiisomorphism from $\mathcal{B}(\mathcal{H})$ to $\mathcal{B}(\bar{\mathcal{H}})$ implemented by the conjugation $\mathcal{H} \ni \xi \mapsto \bar{\xi} \in \bar{\mathcal{H}}$.

We denote by $\hat{\mathcal{C}}$ the set of all equivalence classes of cyclic representants of $(N \subset M, e_N, M^{\mathrm{op}} \supset N^{\mathrm{op}})$ and by $\mathcal{C}$ a set of chosen representations for $\hat{\mathcal{C}}$ such that if $(\pi, \pi') \in \mathcal{C}$ then $(\bar{\pi}, \bar{\pi}') \in \mathcal{C}$.

1.1. Proposition. *There exists a unital C^*-algebra $\mathcal{U}$ with unital embeddings $j : M_1 \hookrightarrow \mathcal{U}$, $j' : M_1^{\mathrm{op}} \hookrightarrow \mathcal{U}$ such that*
a) $[j(M), j'(M^{\mathrm{op}})] = 0$,
b) $j(e_N) = j'(e_N^{\mathrm{op}})$.
and such that given any other unital C^*-algebra $\mathcal{U}_0$ with unital embeddings $j_0 : M_1 \hookrightarrow \mathcal{U}_0$, $j_0' : M_1^{\mathrm{op}} \hookrightarrow \mathcal{U}_0$ satisfying a), b) *(with (j', j_0') instead of (j, j')), there exists a unital $*$-algebra morphism $\pi : \mathcal{U} \to \mathcal{U}_0$ such that*

$$(*) \qquad \pi(j(x)) = j_0(x), \quad \pi(j'(x^{\mathrm{op}})) = j_0'(x^{\mathrm{op}}), \qquad \forall\, x \in M_1.$$

Moreover, $\mathcal{U}$ is unique (up to an isomorphism $(*)$) with these properties. Also, $\mathcal{U}$ is generated as a C^*-algebra by $j(M)$, $j'(M^{\mathrm{op}})$, $j(e_N)$ $(= j'(e_N^{\mathrm{op}}))$ and it has a unique antiautomorphism ${}^{\mathrm{op}}$ such that $j(x)^{\mathrm{op}} = j'(x^{\mathrm{op}})$, $(j'(x^{\mathrm{op}}))^{\mathrm{op}} = j(x)$, $\forall\, x \in M_1$ (so in particular $j(e_N)^{\mathrm{op}} = j'(e_N^{\mathrm{op}})^{\mathrm{op}} = j(e_N)$).

Proof. Put

$$\mathcal{U} \stackrel{def}{=} C^*\left(\left\{\bigoplus_{(\pi,\pi')\in\mathcal{C}} \pi(x), \bigoplus_{(\pi,\pi')\in\mathcal{C}} \pi'(y^{\text{op}}) \mid x, y \in M_1\right\}\right),$$

$$j(x) \stackrel{def}{=} \bigoplus_{(\pi,\pi')\in\mathcal{C}} \pi(x), \quad j'(x^{\text{op}}) \stackrel{def}{=} \bigoplus_{(\pi,\pi')\in\mathcal{C}} \pi'(x^{\text{op}}), \qquad x \in M_1.$$

$\mathcal{U}$, j, j' this way defined clearly satisfy a), b), and $(*)$ and the uniqueness is then trivial. Then we can define $^{\text{op}}$ on $\mathcal{U}$ by:

$$\mathcal{U} \ni \bigoplus_{(\pi,\pi')\in\mathcal{C}} \pi(x) \mapsto \bigoplus_{(\pi,\pi')\in\mathcal{C}} \bar{\pi}'(x^{*\text{op}}) \in \mathcal{U}$$

$$\mathcal{U} \ni \bigoplus_{(\pi,\pi')\in\mathcal{C}} \pi'(x^{\text{op}}) \mapsto \bigoplus_{(\pi,\pi')\in\mathcal{C}} \bar{\pi}(x^*) \in \mathcal{U}$$

Q.E.D.

1.2. Definition. We denote by $C^*_{u,\max}(M, e_N, M^{\text{op}})$ the C^*-algebra $\mathcal{U}$ constructed in 1.1 and call it the *universal symmetric enveloping C^*-algebra*. Also we denote by $C^*_{u,\text{bin}}(M, e_N, M^{\text{op}}) \stackrel{def}{=} C^*_{\max}(M, e_N, M^{\text{op}})/\cap \ker\pi$, where the intersection is over all representations π of $C^*_{\max}(M, e_N, M^{\text{op}})$ for which $\pi(M)$, $\pi(M^{\text{op}})$ are von Neumann algebras and call it the *universal binormal symmetric enveloping C^*-algebra* associated with $N \subset M$ (and the trace preserving expectation). We still denote by j, j' the embeddings of M_1, M_1^{op} into $C^*_{u,\text{bin}}(M, e_N, M^{\text{op}})$ resulting from the composition of the embeddings into $C^*_{u,\max}(M, e_N, M^{\text{op}})$ with the quotient map. Note that, with the notations in the proof of 1.1, if we let $\mathcal{C}^u_{\text{bin}} = \{(\pi, \pi') \in \mathcal{C} \mid \pi(M), \pi'(M^{\text{op}})$ are von Neumann algebras$\}$, then $C^*_{u,\text{bin}}(M, e_N, M^{\text{op}})$ can alternatively be defined as

$$C^*\left(\left\{\bigoplus_{\mathcal{C}^u_{\text{bin}}} \pi(x), \bigoplus_{\mathcal{C}^u_{\text{bin}}} \pi'(y^{\text{op}}) \mid x, y \in M_1\right\}\right)$$

with

$$j(x) = \bigoplus_{\mathcal{C}^u_{\text{bin}}} \pi(x), \quad j'(x^{\text{op}}) = \bigoplus_{\mathcal{C}^u_{\text{bin}}} \pi'(x^{\text{op}}).$$

Since $(\pi, \pi') \in \mathcal{C}^u_{\text{bin}}$ implies $(\bar{\pi}, \bar{\pi}') \in \mathcal{C}^u_{\text{bin}}$, it follows that $^{\text{op}}$ implements an antiautomorphism on $C^*_{u,\text{bin}}(M, e_N, M^{\text{op}})$, still denoted $^{\text{op}}$, satisfying $j'(x^{\text{op}}) = j(x)^{\text{op}}$, $j'(x^{\text{op}})^{\text{op}} = j(x)$.

In addition, $C^*_{u,\text{bin}}(M, e_N, M^{\text{op}})$ satisfies the following universality property:

1.3. Proposition. *Given any binormal representation (π_o, π'_0) of $(N \subset M, e_N, M^{\text{op}} \supset N^{\text{op}})$ on a Hilbert space $\mathcal{H}_0$ there exists a unital $*$-representation $\pi : C^*_{u,\text{bin}}(M, e_N, M^{\text{op}}) \to \mathcal{B}(\mathcal{H}_0)$ such that $\pi(j(x)) = \pi_0(x)$, $\pi'(j'(x^{\text{op}})) = \pi'_0(x^{\text{op}})$, $\forall\ x \in M_1$. Moreover, $C^*_{u,\text{bin}}(M, e_n, M^{\text{op}})$ has a faithful representation $\tilde{\pi}$ such that $\tilde{\pi}(M)$, $\tilde{\pi}(M^{\text{op}})$ are von Neumann algebras. Also, $C^*_{u,\text{bin}}(M, e_N, M^{\text{op}})$ with the embeddings j, j' is unique (up to isomorphism) satisfying these properties.*

Proof. Trivial. Q.E.D.

1.4. Lemma. *Let $N \subset M \overset{e_1}{\subset} M_1 \overset{e_2}{\subset} M_2 \subset \cdots$ be the Jones tower for $N \subset M$, with $e_1 = e_N$, and $M \overset{e_0}{\supset} N \overset{e_{-1}}{\supset} N_1 \supset \cdots$ be a choice of a tunnel. Let $\mathcal{S}_0$ be a unital C^*-algebra with unital $*$-embeddings $j_0 : M_1 \to \mathcal{S}_0$, $j_0' : M_1^{\rm op} \to \mathcal{S}_0$, such that $[j_0(M), j_0'(M^{\rm op})] = 0$, $j_0(e_N) = j_0'(e_N)$. Then j_0, j_0' extend uniquely to $*$-embeddings of $\bigcup_{n\geq 1} M_n$, $\bigcup_{n\geq 1} M_n^{\rm op}$ into $\mathcal{S}_0$, still denoted by j_0, j_0', such that $j_0(e_{n+2}) = j_0'(e_{-n}^{\rm op}), j_0'(e_{n+2}^{\rm op}) = j_0(e_{-n})$, $n \geq 0$.*

Proof. Trivial by the abstract characterization of the basic contruction in ([PiPo2]. [Po2]). Q.E.D.

1.5. Lemma. *Let $\cdots \overset{e_{-1}}{\subset} N \overset{e_0}{\subset} M \overset{e_1}{\subset} M_1 \subset \cdots$, $\mathcal{S}_0$, j_0, j_0' be like in 1.4. Then we have*

$$\begin{aligned}\mathrm{Alg}(j_0(M_1), j_0'(M_1^{\rm op})) &= \mathrm{Alg}(j_0(M), j_0(e_N) = j_0'(e_N), j_0'(M^{\rm op}))\\ &= \bigcup_n \mathrm{sp}\, j_0'(M^{\rm op}) j_0(M_n) j_0'(M^{\rm op})\\ &= \bigcup_n \mathrm{sp}\, j_0(M) j_0'(M_n^{\rm op}) j_0(M)\\ &= \bigcup_n \mathrm{sp}\, j_0'(M^{\rm op}) j_0(M) j_0(f_{-n}^n) j_0(M) j_0'(M^{\rm op}),\end{aligned}$$

where f_{-n}^n is the Jones projection for $N_{n-1} \subset M \subset M_n$ obtained as a scalar multiple of the word of maximal length in $e_{-n+2}, \ldots, e_0, e_1, \ldots, e_n$ (cf. [PiPo2]) and it satisfies $j_0(f_{-n}^n) = j_0'((f_{-n}^n)^{\rm op})$. Similarly, for any $i \in \mathbb{Z}$ we have

$$\mathrm{Alg}(j_0(M), j_0(e_N), j_0'(M^{\rm op})) = \bigcup_n \mathrm{sp}\, j_0'(M_i^{\rm op}) j_0(M_n) j_0'(M_i^{\rm op}),$$

where $M_i = N_{-i-1}$ for $i \leq -1$, $M_0 = M$, $M_{-1} = N$.

Proof. It is sufficient to show that

$$\bigcup_n \mathrm{sp}\, j'(M^{\rm op}) j(M) j(f_{-n}^n) j(M) j(M^{\rm op})$$

is an algebra. If we denote by f_{-2n}^0 the Jones projection for $N_{2n-1} \subset N_{n-1} \subset M$ and by f_0^{2n} the one for $M \subset M_n \subset M_{2n}$, as in [PiPo2], then we have $M_n = \mathrm{sp}\, M f_{-n}^n M$, $M = \mathrm{sp}\, N_{n-1} f_{-2n}^0 N_{n-1}$ so that we get:

$$\begin{aligned}&j'(M^{\rm op}) j(M) j(f_{-n}^n) j(M) j'(M^{\rm op}) j(f_{-n}^n) j(M) j'(M^{\rm op})\\ &\quad\subset \mathrm{sp}\, j'(M^{\rm op}) j(M) j(f_{-n}^n)(j(f_{-2n}^0) j(N_{n-1}))(j'(N_{n-1}^{\rm op}) j'((f_{-2n}^0)^{\rm op}) j'(N_{n-1}^{\rm op}))\\ &\qquad\cdot j(f_{-n}^n) j(M) j'(M^{\rm op})\\ &\quad= \mathrm{sp}(j'(M^{\rm op}) j'(N_{n-1}^{\rm op}))(j(M) j(N_{n-1}))(j(f_{-n}^n) j(f_{-2n}^0) j'((f_{-2n}^0)^{\rm op}) j(f_{-n}^n))\\ &\qquad\cdot (j(N_{n-1}) j(M))(j'(N_{n-1}^{\rm op}) j'(M^{\rm op}))\\ &\quad= \mathrm{sp}\, j'(M^{\rm op}) J(M) j(f_{-2n}^{2n}) j(M) j'(M^{\rm op})\end{aligned}$$

in which we used that $[j(N_{n-1}), j(f_{-n}^n)] = 0$, $[j'(N_{n-1}^{\rm op}), j(f_{-n}^n)] = 0$ and $f_{-2n}^{2n} = \lambda^{-n} f_{-n}^n f_0^{2n} f_{-2n}^0 f_{-n}^n$, $j(f_0^{2n}) = j'((f_{-2n}^0)^{\rm op})$, $\lambda = [M : N]^{-1}$. Q.E.D.

1.6. Corollary. *Let* $\cdots \subset N_1 \overset{e_{-1}}{\subset} N \overset{e_0}{\subset} M \overset{e_1}{\subset} M_1 \subset \cdots$, f_{-n}^n *be as in* 1.5. *Then* $C^*_{u,\max}(M, e_N, M^{\mathrm{op}})$ *(respectively* $C^*_{u,\mathrm{bin}}(M, e_n, M^{\mathrm{op}})$*) is generated, as a* C^**-algebra, by* $j(M)$, $j(f_{-n}^n) = j'(f_{-n}^n)$, $j'(M^{\mathrm{op}})$ *and there exists a natural isomorphism of* $C^*_{u,\max}(M, e_N, M^{\mathrm{op}})$ *(respectively* $C^*_{u,\mathrm{bin}}(M, e_N, M^{\mathrm{op}})$*) onto* $C^*(M, e_{N_{n-1}}, M^{\mathrm{op}})$ *(respectively* $C^*_{\mathrm{bin}}(M, e_{N_{n-1}}, M^{\mathrm{op}})$*), taking the canonical images of the elements in* M, M^{op} *in one algebra into the corresponding canonical images in the other algebra and* $j(f_{-n}^n)$ *onto* $j(e_{N_{n-1}})$.

Proof. Trivial by definitions and 1.5. Q.E.D.

1.7. Lemma. *With the notations of* 1.4 *and* 1.5, *assume in addition that* $j_0(M)$, $j_0(e_N) = j_0'(e_N)$, $j_0'(M^{\mathrm{op}})$ *generate* $\mathcal{S}_0$ *as a* C^**-algebra, and that the following condition is satisfied:*

$$(*) \qquad j_0(M' \cap M_k) \subset j_0'(M^{\mathrm{op}}), \forall k \geq 1$$

Then $j_0(M_i)' \cap \mathcal{S}_0 = j_0'(M_{-i}^{\mathrm{op}})$, $(j_0'(M_i^{\mathrm{op}}))' \cap \mathcal{S}_0 = j_0(M_{-i})$, $\forall\, i \in \mathbb{Z}$, *and for all* k, i *in* $\mathbb{Z}$ *one has* $j_0(M_i' \cap M_k) = j_0(M_i)' \cap j_0(M_k) = j_0'(M_{-i}^{\mathrm{op}}) \cap (j_0'(M_{-k}^{\mathrm{op}}))' = j_0'((M_{-k}^{\mathrm{op}})' \cap M_{-i}^{\mathrm{op}})$. *Also, if* $x \in M_{-n}' \cap M_n$ *and* x' *denotes the canonical conjugate of* x $(= Jx^*J)$ *([Po2]), then* $j_0(x') = j'(x^{\mathrm{op}})$. *Moreover, for each* $i \in \mathbb{Z}$ *there exist unique conditional expectations* $\mathcal{E}_i^+ : \mathcal{S}_0 \to j_0(M_i)' \cap \mathcal{S}_0 = j_0'(M_{-i}^{\mathrm{op}})$, $\mathcal{E}_i^- : \mathcal{S}_0 \to j_0'(M_i^{\mathrm{op}})' \cap \mathcal{S}_0 = j_0(M_{-i})$ *such that* $\mathcal{E}_i^+(j_0(x)) = j_0(E_{M_i' \cap M_n}(x))$, $\mathcal{E}_i^-(j_0'(x^{\mathrm{op}})) = j_0'(E_{M_i' \cap M_n}(x)^{\mathrm{op}})$, $\forall\, x \in M_n$, $n \geq i$, *which satisfy* $\mathcal{E}_i^+ = \mathcal{E}_i^+(j_0(u) \cdot j_0(u^*))$, $\mathcal{E}_i^- = \mathcal{E}_i^-(j_0'(u^{\mathrm{op}}) \cdot j_0'(u^{\mathrm{op}})^*)$, $\forall\, u \in \mathcal{U}(M_i)$.

Proof. Since $j_0(M' \cap M_k) \subset M^{\mathrm{op}}$ and $[j_0(M_k), j_0'(M_{-k}^{\mathrm{op}}] = 0$, it follows that $j_0(M' \cap M_k) \subset j_0'(M_{-k}^{\mathrm{op}})' \cap j_0'(M^{\mathrm{op}}) = j_0'(M_{-k}^{\mathrm{op}\,\prime} \cap M^{\mathrm{op}})$. But the two finite dimensional algebras involved in this inclusion have the same dimension, so they actually follow equal. By averaging over unitaries in M_i it then follows that $j_0(M_i' \cap M_k) \subset j_0'(M^{\mathrm{op}}), \forall i \geq 1$, giving in a similar way $j_0(M_i' \cap M_k) = j_0'(M_{-k}^{\mathrm{op}\,\prime} \cap M_{-i}^{\mathrm{op}})$. Then by duality isomorphisms these equalities follow for arbitrary $i, k \in \mathbb{Z}$.

By the relative Dixmier property for subfactors of finite index (see the Appendix, A.1), if for $x \in M_n$ we denote $C_{M_i}(x) = \overline{\mathrm{co}}^n\{uxu^* \mid u \in \mathcal{U}(M_i)\} \cap M_i' \cap M_n$ then $C_{M_i}(x) = \{E_{M_i' \cap M_n}(x)\}$ and $\forall\ x_1, \ldots, x_k \in M_n$, $\forall\ \varepsilon > 0$, $\exists\ u_1, \ldots, u_m \in \mathcal{U}(M_i)$ such that

$$\left\| \frac{1}{m} \sum_{l=1}^{m} u_l x_j u_l^* - E_{M_i' \cap M_n}(x_j) \right\| < \varepsilon, \qquad j = 1, 2, \ldots, k.$$

Since, by 1.5 we have

$$\mathrm{Alg}(j_0(M_1), j_0'(M_1^{\mathrm{op}})) = \bigcup_n \mathrm{sp}\, j_0'(M_{-i}^{\mathrm{op}}) j_0(M_n) j_0'(M_{-i}^{\mathrm{op}}),$$

which is dense in $\mathcal{S}_0$, it follows that

$$\begin{aligned} j_0(M_i)' \cap \mathcal{S}_0 &= \bigcup_n j_0'(M_{-i}^{\text{op}}) j_0(M_i' \cap M_n) j_0'(M_{-i}^{\text{op}}) \\ &= \bigcup_n j_0'(M_{-i}^{\text{op}}) j_0'((M_{-n}^{\text{op}})' \cap (M_{-i}^{\text{op}})) j_0'(M_{-i}^{\text{op}}) \\ &= j_0'(M_{-i}^{\text{op}}). \end{aligned}$$

Also, it follows that if $T = \sum_l j_0'(y_{1,l}^{\text{op}}) j_0(x_l) j_0'(^{\text{op}}_{2,l})$, for some $y_{1,l}, y_{2,l} \in M_{-i}$, $x_l \in M_n$, and we denote by $C_{i,\mathcal{S}_0}(T) = \overline{\text{co}}^n\{j_0(u)Tj_0(u^*) \mid u \in \mathcal{U}(M_i)\} \cap (j_0(M_i))' \cap \mathcal{S}_0$, then

$$C_{i,\mathcal{S}_0}(T) = \left\{ \sum_l j_0'(y_{1,l}^{\text{op}}) j_0(E_{M_i' \cap M_n}(x_l)) j_0'(y_{2,l}^{\text{op}}) \right\}$$

is a single point set. Also, $C_{i,\mathcal{S}_0}(\alpha T_1 + \beta T_2) \subset \alpha C_{i,\mathcal{S}_0}(T_1) + \beta C_{i,\mathcal{S}_0}(T_2)$ and $1 \geq T \geq 0$ implies $1 \geq T' \geq 0$, $\forall\, T' \in C_{i,\mathcal{S}_0}(T)$. It follows that

$$\begin{aligned} \text{Alg}\{j_0(M_1), j_0'(M_1^{\text{op}})\} \ni T &= \sum_l j_0'(y_{1,l}^{\text{op}}) j_0(x_l) j_o'(y_{2,l}^{\text{op}}) \\ \mapsto \sum_l j_0'(y_{1,l}^{\text{op}}) j_0(E_{M_i' \cap M_n}(x_l)) & j_0'(y_{2,l}^{\text{op}}) \in (j_0(M_i))' \cap \mathcal{S}_0 = j_0'(M_{-i}^{\text{op}}) \end{aligned}$$

is a well defined positive linear norm one projection onto $j_0'(M_{-i}^{\text{op}})$ and the rest of the statement is then clear by continuity. Q.E.D.

1.8. Definition. We denote $C^*_{\max}(M, e_N, M^{\text{op}}) \stackrel{\text{def}}{=} C^*_{u,\max}(M, e_N, M^{\text{op}})/\cap \ker \pi$, where the intersection is over all *smooth* representations π of $C^*_{u,\max}(M, e_N, M^{\text{op}})$, i.e., representations satisfying the following smoothness condition (or axiom):

$$(*) \qquad \pi(j(M' \cap M_i)) \subset \pi(j'(M^{\text{op}})), \quad i \in \mathbb{N}.$$

Note that by 1.7 this condition actually implies $\pi(j(M_k' \cap M_i)) = \pi(j'(M_{-i}^{\text{op}\prime} \cap M_{-k}^{\text{op}})), \forall i, k \in \mathbb{Z}$.

We call $C^*_{\max}(M, e_N, M^{\text{op}})$ the *symmetric enveloping C^*-algebra* associated with $N \subset M$. Similarly, we put $C^*_{\text{bin}}(M, e_N, M^{\text{op}}) \stackrel{\text{def}}{=} C^*_{u,\text{bin}}(M, e_N, M^{\text{op}})/\cap \ker \pi$, where the intersection is over all representations π of $C^*_{u,\text{bin}}(M, e_N, M^{\text{op}})$ such that $\pi(j(M))$, $\pi(j'(M^{\text{op}}))$ are von Neumann algebras and such that axiom $(*)$ is satisfied. We call it the *binormal symmetric enveloping C^*-algebra* associated with $N \subset M$. Note that, since $\mathcal{B}(L^2(M))$, with the representation of M and M^{op} as operators of left and right multiplication by elements in M and $e_N = \text{proj}_{L^2(N)} \in \mathcal{B}(L^2(M))$, does satisfy the condition $(*)$, both these symmetric enveloping C*-algebras are non-degenerate.

We still denote by j, j' the canonical embeddings of M_1, $M_1^{\rm op}$ in $C^*_{\max}(M, e_N, M^{\rm op})$ and $C^*_{\rm bin}(M, e_N, M^{\rm op})$. Note that the same argument as in 1.2 shows that the antiautomorphism ${}^{\rm op}$ on $C^*_{u,\max}(M, e_N, M^{\rm op})$ (respectively ${}^{\rm op}$ on $C^*_{u,{\rm bin}}(M, e_N, M^{\rm op})$) implements an antiautomorphism, still denoted by ${}^{\rm op}$, on $C^*_{\max}(M, e_N, M^{\rm op})$ (resp. ${}^{\rm op}$ on $C^*_{\rm bin}(M, e_N, M^{\rm op})$).
Also, by universality properties of $C^*_{u,\max}(M, e_N, M^{\rm op})$ and $C^*_{u,{\rm bin}}(M, e_N, M^{\rm op})$ and the definitions, it follows that given any C^*-algebra $\mathcal{S}_0$ generated by copies of M, $M^{\rm op}$, e_N satisfying 1.1 a), b), such that the corresponding tunnel-towers $\{M_i\}_i$, $\{M_j^{\rm op}\}_j$ (cf. 1.4) satisfy the smoothness axiom 1.8 $(*)$, there exists a natural $*$-morphism of $C^*_{\max}(M, e_N, M^{\rm op})$ onto $\mathcal{S}_0$ carrying $j(M_i)$, $j'(M_j^{\rm op})$ onto the corresponding images of M_i, $M_j^{\rm op}$ $(\subset \mathcal{S}_0)$. If in addition $\mathcal{S}_0 \subset \mathcal{B}(\mathcal{H}_0)$ is so that the images of M, $M^{\rm op}$ are weakly closed, then this morphism factors to a $*$-morphism of $C^*_{\rm bin}(M, e_N, M^{\rm op})$.
The above can be regarded as the universality property satisfied by $C^*_{\max}(M, e_N, M^{\rm op})$ and $C^*_{\rm bin}(M, e_N, M^{\rm op})$. Moreover, as a consequence of the prior results and definitions, if we denote by $\mathcal{S}$ either of these two algebras, then the following properties hold true:
1.9.1. $j(M_i)' \cap \mathcal{S} = j'(M_{-i}^{\rm op})$, $j'(M_{-i}^{\rm op})' \cap \mathcal{S} = j(M_i)$, $\forall\, i \in \mathbb{Z}$.
1.9.2. If $x \in M'_{-n} \cap M_n$ and x' denotes the canonical conjugate of x $(= Jx^*J)$ then $j(x') = j'(x^{\rm op})$.
1.9.3. There exist unique conditional expectations $\mathcal{E}_i^- : \mathcal{S} \to j(M_{-i})$, $\mathcal{E}_i^+ : \mathcal{S} \to j'(M_{-i}^{\rm op})$ such that $\mathcal{E}_i^-(j'(x^{\rm op})) = j'(E_{M_i'\cap M_n}(x)^{\rm op})$, $\mathcal{E}_i^+(j'(x)) = j(E_{M_i'\cap M_n}(x))$, $\forall\ x \in M_n,\ n \geq i$. Also, these expectations satisfy $\mathcal{E}_i^- = \mathcal{E}_i^-(j'(u^{\rm op}) \cdot j'(u^{{\rm op}*}))$, $\mathcal{E}_i^+ = \mathcal{E}_i^+(j(u) \cdot j(u^*))$, $\forall\ u \in \mathcal{U}(M_i)$.
1.9.4. $C^*_{\max}(M, e_{N_{n-1}}, M^{\rm op})$ (resp. $C^*_{\rm bin}(M, e_{N_{n-1}}, M^{\rm op})$) naturally identifies with $C^*_{\max}(M, e_N, M^{\rm op})$ (resp. $C^*_{\rm bin}(M, e_N, M^{\rm op})$), as in 1.6.

1.10. *Remarks.* 1°. Note that the smoothness condition 1.8 $(*)$ is redundant if $M' \cap M_n = Alg\{1, e_1, e_2, ..., e_n\}, \forall n$, i.e., in the case the graph of $N \subset M$ is of the form $\Gamma_{N,M} = A_n$ for some $n \leq \infty$.
2°. In the case $\mathcal{S}_0 \subset \mathcal{B}(\mathcal{H})$ is so that $j_0(M)$, $j_0'(M^{\rm op})$ are von Neumann algebras (e.g., if $\mathcal{S}_0 = C^*_{\rm bin}(M, e_N, M^{\rm op})$) then one can give another proof to Lemma 1.7, which doesn't use the relative Dixmier property, as follows: if M is weakly separable (i.e., M has separable predual) then take $R \subset M$ to be a hyperfinite subfactor such that $R' \cap M_\infty = M' \cap M_\infty$ (cf. [Po2,9]), so in particular $R' \cap M_n = M' \cap M_n$, $\forall\ n$ (here $M_\infty = \overline{\cup M_n}^{\rm w}$ as usual). Then denote by Φ the conditional expectation of $\mathcal{B}(\mathcal{H})$ onto $j_0(R)' \cap \mathcal{B}(\mathcal{H})$, obtained by averaging over a suitable amenable subgroup of $\mathcal{U}(R)$. Then clearly $\Phi|_{\mathcal{S}_0} = \mathcal{E}_0^+$ and the other expectations are obtained similarly. If M is not separable one can still apply [Po2,9] to get that $\forall\ B \subset \cup j(M_n)$ countably generated, $\exists\ R \subset M$ such that $E_{R'\cap M_n}(B) = E_{M'\cap M_n}(B)$, $\forall\ n$, and the rest of the proof is then similar.
3°. The considerations in this section are easily seen to cary over to the case when instead of an extremal inclusion of type II$_1$ factors $N \subset M$ (with trace preserving expectation) we take an extremal inclusion of factors of type III,

$\mathcal{N} \subset \mathcal{M}$ ([Po3]). However, in this more general case, some adjustements of the argument in 1.7 are needed, depending on the nature of the inclusion. Then, if $\mathcal{E}$ denotes the expectation of minimal index of $\mathcal{M}$ onto $\mathcal{N}$, an argument similar to 2° above can be used to prove the existence of a unique conditional expectation $\mathcal{E}^0$ from $C^*_{\text{bin}}(\mathcal{M}, e_{\mathcal{N}}, \mathcal{M}^{\text{op}})$ onto its C*-subalgebra generated by $\mathcal{M}$ and $\mathcal{M}^{\text{op}}$.

2. Symmetric Enveloping type II$_1$ Factors

2.1. Theorem. *There exists a unique trace state* tr *on* $C^*_{\max}(M, e_N, M^{\text{op}})$ *and the corresponding ideal trace* $\mathcal{I}_{\text{tr}} = \{x \in C^*_{\max}(M, e_N, M^{\text{op}}) \mid \text{tr}(x^*x) = 0\}$ *is the unique maximal ideal in* $C^*_{\max}(M, e_N, M^{\text{op}})$*. In particular, there exists a unique state* τ_0 *on each quotient* C^**-algebra* $\mathcal{S}_0$ *of* $C^*_{\max}(M, e_N, M^{\text{op}})$ *(in particular on* $C^*_{\text{bin}}(M, e_N, M^{\text{op}})$*) and its ideal is the unique maximal ideal of* $\mathcal{S}_0$*.*

Proof. By the uniqueness properties of the expectations $\mathcal{E}_i^+$, $i \in \mathbb{Z}$, of a C^*-algebra $\mathcal{S}_0$ generated by $j_0(M_1)$, $j_0'(M_1^{\text{op}})$ onto $j_0'(M_{-i}^{\text{op}})$ like in 1.6, it follows that $\mathcal{E}_i^+ = E^{j_0'(M^{\text{op}})}_{j_0'(N_{i-1}^{\text{op}})} \circ \mathcal{E}_0^+$. Let τ be the trace on $j_0(M)$ and τ' the trace on $j_0'(M^{\text{op}})$ and define $\tau_0 = \tau' \circ \mathcal{E}_0^+$ on $\mathcal{S}_0$. Since $E^{j_0'(M^{\text{op}})}_{j_0'(N_{i-1}^{\text{op}})}$ is τ' preserving, we have for $i \geq 0$, $x \in \mathcal{S}_0$:

$$\tau_0(x) = \tau'(\mathcal{E}_0^+(x)) = \tau'\left(E^{j_0'(M^{\text{op}})}_{j_0'(N_{i-1}^{\text{op}})}(\mathcal{E}_0^+(x))\right) = \tau' \circ \mathcal{E}_i^+(x).$$

If $k \geq i$, $u \in \mathcal{U}(j_0(M_i))$, $x \in j_0(M_k)$, $y', y'' \in j_0'(N_{i-1}^{\text{op}})$ then we have:

$$\begin{aligned}\tau_0(uy'xy''u^*) &= \tau_0(y'uxu^*y'') = \tau'(\mathcal{E}_i^+(y'uxu^*y'')) = \tau'(y'\mathcal{E}_i^+(uxu^*)y'') \\ &= \tau'(y'E_{j_0(M_i' \cap M_k)}(uxu^*)y'') = \tau'(y'\mathcal{E}_i^+(x)y'') = \tau'(\mathcal{E}_i^+(y'xy'')) \\ &= \tau_0(y'xy'').\end{aligned}$$

Thus, by 1.6 it follows that $\tau_0(uTu^*) = \tau_0(T)$, $\forall\ T \in \mathcal{S}_0$, $\forall\ u \in \mathcal{U}(M_i)$. Also, if $u' \in j_0'(M^{\text{op}})$ is a unitary element and $x \in j_0(M_k)$, $y', y'' \in j_0'(M^{\text{op}})$ then we get:

$$\begin{aligned}\tau_0(u'y'xy''u'^*) &= \tau'(\mathcal{E}_0^+(u'^*y'xy''u')) = \tau'(u'y'\mathcal{E}_0^+(x)y''u'^*) \\ &= \tau'(y'\mathcal{E}_0^+(x)y'') = \tau'(\mathcal{E}_0'(y'xy'')) = \tau_0(y'xy'').\end{aligned}$$

This shows that $\tau_0(u'Tu'^*) = \tau_0(T)$, $\forall\ T \in \mathcal{S}_0$, $\forall\ u' \in \mathcal{U}(j_0'(M^{\text{op}}))$, by virtue of 1.6. Since the centralizer of τ_0 is an algebra and it contains both $\mathcal{U}(j_0(M_i))$, $\mathcal{U}(j_0'(M^{\text{op}}))$, with $i \geq 1$, τ_0 has all $\mathcal{S}_0 = C^*(j_0(M_i), j_0'(M^{\text{op}}))$ in its centralizer, thus, it is a trace.

If τ_1 is another trace on $\mathcal{S}_0$ and $(\pi_{\tau_1}, \mathcal{H}_{\tau_1}, \mathcal{E}_{\tau_1})$ is the corresponding GNS construction, then let $S_0 = \overline{\pi_{\tau_1}(\mathcal{S}_0)}^{\text{w}}$. Since the unit ball of $\pi_{\tau_1}(M_k)$ is complete in the norm given by $\|\pi_{\tau_1}(x)\xi_{\tau_1}\|$ (because the unit ball of $j_0(M_k)$ is complete in the norm $\tau_1(x^*x)^{1/2}$, by the uniqueness of the trace on the factor M_k) it follows that $\overline{\pi_{\tau_1}(M_k)} = \pi_{\tau_1}(M_k)$. Thus, for $x \in j_0(M_k)$, $y', y'' \in j_0'(M^{\mathrm{op}})$ we get:

$$\begin{aligned}
\tau_1(y'xy'') &= \langle E_{\pi_{\tau_1}(j_0(M))'\cap S_0}(\pi_{\tau_1}(y'xy''))\xi_{\tau_1}, \xi_{\tau_1}\rangle \\
&= \langle(\pi_{\tau_1}(y')E_{\pi_{\tau_1}(j_0(M))'\cap S_0}(\pi_{\tau_1}(j_0(x))\pi_{\tau_1}(y'')\xi_{\tau_1}, \xi_{\tau_1}\rangle \\
&= \langle \pi_{\tau_1}(y'E_{j_0(M'\cap M_k)}(x)y'')\xi_{\tau_1}, \xi_{\tau_1}\rangle \\
&= \tau'(y'E_{j_0(M'\cap M_k)}(x)y'') \\
&= \tau_0(y'xy'').
\end{aligned}$$

with the last part following from the uniqueness of the trace on $j_0'(M^{\mathrm{op}})$.
This shows that $\mathcal{S}_0$ has a unique trace τ_0 and also that if $\mathcal{I}_0 \subset \mathcal{S}_0$ is a two sided closed proper ideal then $\mathcal{S}_1 = \mathcal{S}_0/\mathcal{I}_0$ has a trace, which thus composed with the quotient map gives the trace on $\mathcal{S}_0$. Thus, $\mathcal{I}_0 \subset \mathcal{I}_{\tau_0}$, so $\mathcal{I}_{\tau_0}$ is the unique maximal ideal of $\mathcal{S}_0$. Q.E.D.

2.2. *Remarks.* 1°. Let

$$C^*_{\min}(M, e_N, M^{\mathrm{op}}) \stackrel{\mathrm{def}}{=} C^*_{\max}(M, e_N, M^{\mathrm{op}})/I_{\mathrm{tr}} \ (\simeq \pi_{\mathrm{tr}}(C^*_{\max}(M, e_N, M^{\mathrm{op}}))),$$

where I_{tr} is the trace ideal (= maximal ideal) of $C^*_{\max}(M, e_N, M^{\mathrm{op}})$ corresponding to the unique trace tr, as given by 2.1. From the previous theorem and its proof if follows that $C^*_{\min}(M, e_N, M^{\mathrm{op}})$ is simple, has a unique trace, still denoted tr, and has the Dixmier property, i.e., $\overline{\mathrm{co}}^n\{uxu^* \mid u \in \mathcal{U}(C^*_{\min}(M, e_N, M^{\mathrm{op}}))\} \cap \mathbb{C}1 = \{\mathrm{tr}(x)1\}$, $\forall\, x \in C^*_{\min}(M, e_N, M^{\mathrm{op}})$. In fact, by 2.1 any C^*-algebra $\mathcal{S}_0$ generated by mutually commuting copies of M, M^{op} and a projection e_N such that $N \subset M \overset{e_N}{\subset} \mathrm{Alg}(M, e_N)$ and $N^{\mathrm{op}} \subset M^{\mathrm{op}} \overset{e_N}{\subset} \mathrm{Alg}(M^{\mathrm{op}}, e_N)$ are basic constructions and such that the smoothness condition 1.8 $(*)$ is satisfied, has a unique trace tr, I_{tr} is its unique maximal ideal and $\mathcal{S}_0/I_{\mathrm{tr}} = C^*_{\min}(M, e_N, M^{\mathrm{op}})$.
Also it should be noted that if $N = M$ then $C^*_{\max}(M, e_N, M^{\mathrm{op}})$ coincides with $M \underset{\max}{\otimes} M^{\mathrm{op}}$, $C^*_{\mathrm{bin}}(M, e_N, M^{\mathrm{op}})$ with $M \underset{\mathrm{bin}}{\otimes} M^{\mathrm{op}}$ (as considered in [EL]) and $C^*_{\min}(M, e_N, M^{\mathrm{op}})$ with $M \underset{\min}{\otimes} M^{\mathrm{op}}$.
2°. Let $\mathcal{N} \subset \mathcal{M}$ be an extremal inclusion of von Neumann factors of type III, with the conditional expectation of minimal index $\mathcal{E}$, as in 1.10.3°. The construction analoguous to 2.1 is then as follows: one first considers the expectation $\mathcal{E}^0$ given by 1.10.3°; one takes a normal faithful state φ on $\mathcal{M}$ such that $\varphi \circ \mathcal{E} = \varphi$; instead of the trace tr, one defines a state ψ on $C^*_{\mathrm{bin}}(\mathcal{M}, e_{\mathcal{N}}, \mathcal{M}^{\mathrm{op}})$ by $\psi = (\varphi \otimes \varphi^{\mathrm{op}}) \circ \mathcal{E}^0$.

2.3. Corollary.

$$S = \overline{\pi_{\mathrm{tr}}(C^*_{\max}(M, e_N, M^{\mathrm{op}}))}^{\mathrm{w}} \simeq \overline{\pi_{\mathrm{tr}}(C^*_{\mathrm{bin}}(M, e_N, M^{\mathrm{op}}))}^{\mathrm{w}}$$

is a type II_1 *factor with embeddings* $\pi_{\mathrm{tr}} \circ j : M_1 \to S$, $\pi_{\mathrm{tr}} \circ j' : M_1^{\mathrm{op}} \to S$ *and an antisymmetry* ${}^{\mathrm{op}}$ *such that*
a) $[\pi_{\mathrm{tr}}(j(M)), \pi_{\mathrm{tr}}(j'(M^{\mathrm{op}}))] = 0$
b) $\pi_{\mathrm{tr}}(j(e_N)) = \pi_{\mathrm{tr}}(j'(e_N))$
c) $S = \mathrm{vN}(\pi_{\mathrm{tr}}(j(M)), \pi_{\mathrm{tr}}(j(e_N)), \pi_{\mathrm{tr}}(j'(M^{\mathrm{op}})))$
d) $\pi_{\mathrm{tr}}(j(x))^{\mathrm{op}} = \pi_{\mathrm{tr}}(j'(x^{\mathrm{op}}))$, $\forall\, x \in M$, $\pi_{\mathrm{tr}}(j(e_N))^{\mathrm{op}} = \pi_{\mathrm{tr}}(j(e_N))$.
Moreover, if S_0 *is another type* II_1 *factor with embeddings* $j_0 : M_1 \to S_0$, $j_0' : M_1^{\mathrm{op}} \to S_0$ *satisfying conditions a), b) (with* j_0 *instead of* $\pi_{\mathrm{tr}} \circ j$ *and* j_0' *instead of* $\pi_{\mathrm{tr}} \circ j'$*) and such that* $j_0(M' \cap M_n) \subset j_0'(M^{\mathrm{op}}), \forall n \geq 1$*, then there exists a unique isomorphism* σ *of* S *into* S_0 *such that* $j_0 = \sigma \circ \pi_{\mathrm{tr}} \circ j$, $j_0' = \sigma \circ \pi_{\mathrm{tr}} \circ j'$*. And if in addition* $S_0 = \mathrm{vN}(j_0(M), j_0(e_N), j_0'(M^{\mathrm{op}}))$*, then* σ *is onto.*

Proof. Trivial by 2.1. Q.E.D.

2.4. Definition. We denote by $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ the type II_1 factor $S = \pi_{\mathrm{tr}}(\overline{C^*_{\max}(M, e_N, M^{\mathrm{op}})})$ in the previous corollary and call it the *symmetric enveloping type* II_1 *factor* associated with $N \subset M$. Also, we call $M \vee M^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ the *symmetric enveloping inclusion* associated with $N \subset M$. We will identify M, e_N, M^{op} with their corresponding canonical images in $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$, more generally we will identify M_n, M_n^{op}, e_n with their canonical images via $\pi_{\mathrm{tr}} \circ j$, $\pi_{\mathrm{tr}} \circ j'$ (cf 1.4), whenever some tunnel for $N \subset M$ is chosen. We've seen in 2.3 that $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ has an antisymmetry ${}^{\mathrm{op}}$ and that it satisfies a universality and uniqueness property. Also, from now on we will use the notation τ for the unique trace on the factor $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ (as in fact for any generic factor).

2.5. *Remarks.* 1°. As one can see from 2.1-2.3, the symmetric enveloping type II_1 factor $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ associated to an inclusion $N \subset M$ can be constructed out of any C*-algebra S_0 generated by copies of M and M^{op}, satisfying $M' \cap S_0 = M^{\mathrm{op}}$, and by a projection e_N, implementing the expectations E_N on M and $E_{N^{\mathrm{op}}}$ on M^{op}: just put $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ to be the completion of the algebra S_0/I_0 in the strong topology given by its unique trace, I_0 being the maximal ideal of S_0 or, alternatively, the ideal corresponding to the unique trace on S_0. In particular, $M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = \overline{\mathrm{C}^*(M, e_N, J_M M J_M)/I_0}$. But one can also construct $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ by defining directly the Hilbert space of its standard representation. In order to show this, we will in fact consider a more general construction. Thus, let $N \subset M$ and $Q \subset P$ be extremal inclusions with the same extended higher relative commutant picture (or extended standard invariant), i.e., $\tilde{\to}\mathcal{G}_{N,M} = tilde \to \mathcal{G}_{Q,P} = \{A_{ij}\}_{i,j \in \mathbb{Z}}$. The *concatenation* algebra associated to these two inclusions is then the unique (up to isomorphism) type II_1 factor S generated by commuting copies of M, P^{op} and by a projection e, implementing both E_N and $E_{Q^{\mathrm{op}}}$, such that $M' \cap S = P^{\mathrm{op}}$. This algebra is denoted by $M \underset{e_N = e_Q}{\boxtimes} P^{\mathrm{op}}$

(or simply $M \boxtimes P^{\rm op}$, when no confusion is possible). Its uniqueness follows the same way as the uniqueness of $M \underset{e_N}{\boxtimes} M^{\rm op}$ above. To prove its existence, we consider the following construction: Take $\{m_j\}_{j\in J}$ to be be an orthonormal basis of $A_{-\infty,\infty}$ over $A_{-\infty,0} \vee A_{0,\infty}$ and identify $A_{-\infty,0} \vee A_{0,\infty}$ with its image in $M\bar{\otimes}P^{\rm op}$ (through the choice of tunnels in M and P); note that the m_j's can be chosen of bounded norm and such that the set of indices J can be written as $\cup_n J_n$, where each J_n is finite and such that $\sum_{j\in J_n} m_j M \vee M^{\rm op}$ is a $M \vee M^{\rm op}$-bimodule of finite dimension (equivalently, $\sum_{j\in J_n} m_j P \vee P^{\rm op}$ is a $P \vee P^{\rm op}$-bimodule of finite dimension) See 4.5 below for how to get these m_j's. Then let $\mathcal{H}_n \overset{\rm def}{=} \sum_{j\in J_n} m_j {\rm L}^2(M \vee P^{\rm op})$ and $\mathcal{H} \overset{\rm def}{=} \vee_n \mathcal{H}_n$, the scalar product on $\mathcal{H}$ being defined by $\langle m_j\xi, m_i\eta\rangle = \langle E_{A_{-\infty,0}\vee A_{0,\infty}}(m_i^* m_j)\xi, \eta\rangle$, $\forall \xi, \eta \in {\rm L}^2(M\vee P^{\rm op})$. Finally, we let $M, P^{\rm op}$ and $e = e_N = e_{P^{\rm op}}$ act on $\mathcal{H}$ as follows: M (and similarly $P^{\rm op}$) acts on each $\mathcal{H}_n$ by multiplication to the left, according to the relations $Mm_i \subset \sum_{j\in J_n} m_j(M \vee A_{0,\infty})$, $\forall i \in J_n$, with the latter vector space being identified with a subset of $\mathcal{H}_n$; e acts also by multiplication to the left, by regarding $\mathcal{H}$ as a left $A_{-\infty,\infty}$ module in the obvious way and letting $e = e_1$. Then $M \underset{e_N=e_Q}{\boxtimes} P^{\rm op}$ is simply the von Neumann algebra generated by $M, P^{\rm op}, e_1$ on $\mathcal{H}$.

It is easy to check that these actions of $M, P^{\rm op}, e$ on $\mathcal{H}$ are well defined, that they satisfy $M' \cap {\rm C}^*(M, e, P^{\rm op}) = P^{\rm op}$, $exe = E_N(x)e, eye = E_{P^{\rm op}}(y)e$, for $x \in M, y \in P^{\rm op}$, and that $\langle \cdot \hat{1}, \hat{1}\rangle$ defines a trace on ${\rm C}^*(M, e, P^{\rm op})$. This shows the existence of the concatenation algebra.

Note that, by using the same proofs as for $M \underset{e_N}{\boxtimes} M^{\rm op}$, it follows that the concatenation algebra has similar properties as the ones the symmetric enveloping algebras are shown to have in this section and in Sec. 4. Obviously, when $(Q \subset P) \simeq (N \subset M)$ this algebra coincides with the symmetric enveloping type ${\rm II}_1$ factor associated with $N \subset M$.

Note that any extremal hyperfinite subfactor $N \subset R$ gives rise to a canonical non-separable concatenation algebra as follows: Let ω be a free ultrafilter on $\mathbb{N}$ and denote by R^ω the corresponding ultrapower algebras associated to the hyperfinite factor R. Then $(R' \cap R^\omega)' \cap R^\omega = R$ and more generally $(N_k' \cap R^\omega)' \cap R^\omega = N_k, \forall k$, where $R \supset N \supset N_1 \supset ...$ is a tunnel for $R \supset N$ (cf. [C1]). Thus, if we denote $P^0 = R' \cap R^\omega$ and Q^0 to be the downward basic construction for $P_1^0 = N' \cap R^\omega \supset R' \cap R^\omega = P^0$ and put $(Q \subset P \subset P_1) \simeq (Q^0 \subset P^0 \subset P_1^0)^{\rm op}$ then $N \subset M$ and $Q \subset P$ have the same higher relative commutant pictute (extended standard invariant) and the von Neumann algebra S generated by R and $P_1^{\rm op} = N' \cap R^\omega$ is isomorphic to the concatenation of $(N \subset R)$ and $(Q \subset P)$ (see also Remark 2.11, 1° in [Po3], with caution to the obvious typos there...).

2°. For an extremal inclusion of type III factors $\mathcal{N} \subset \mathcal{M}$ like in 1.10.3°, 2.2.2°, one defines its symmetric enveloping von Neumann algebra as $\pi_\psi(C^*_{\rm bin}(\mathcal{M}, e_{\mathcal{N}}, \overline{\mathcal{M}^{\rm op}}))$. It is easy to see that this algebra does not in fact depend on the normal faithful state φ, with $\varphi = \varphi \circ \mathcal{E}$, taken on $\mathcal{M}$.

The next proposition summarizes the main properties of the factor $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ and its canonical subalgebras:

2.6. PROPOSITION. *$M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ with its subalgebras M_i, M_i^{op} projections e_k, $i, j, k \in \mathbb{Z}$, and antisymmetry ${}^{\mathrm{op}}$ satisfy the conditions:*
a) $[M, M^{\mathrm{op}}] = 0$;
b) $e_1^{\mathrm{op}} = e_1 = e_N$, $e_n^{\mathrm{op}} = e_{-n+2}$, $n \in \mathbb{Z}$, *and* $\cdots \subset N_1 \overset{e_{-1}}{\subset} N \overset{e_0}{\subset} M \overset{e_1}{\subset} M_1 \overset{e_2}{\subset} M_2 \subset \cdots$ *is a Jones tower-tunnel for* $N \subset M$, *where* $M_0 = M$, $M_{-1} = N$, $M_{-n} = N_{n-1}$, $n \geq 2$.
c) $\bigcup_{n\geq 1} M_i M_n^{\mathrm{op}} M_i = \bigcup_{n \geq 1} M_j^{\mathrm{op}} M_n M_j^{\mathrm{op}} = \mathrm{Alg}(M, e_N = e_1, M^{\mathrm{op}})$, $\forall\ i, j \in \mathbb{Z}$ *and it is a dense $*$-subalgebra in* $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$.
d) $M_j' \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = M_{-j}^{\mathrm{op}}$ *and* $(M_j^{\mathrm{op}})' \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = M_{-j}$, $\forall\ j \in \mathbb{Z}$.

Proof. Clear by 1.9 and the definition of $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$. Q.E.D.

The bicommutant relations in d) above can actually be taken as an abstract characterization of the symmetric enveloping algebra:

2.7. PROPOSITION. *Let $N \subset M$ be an extremal inclusion and S be a type II_1 von Neumann algebra containing M. If $(M' \cap S \subset N' \cap S) \simeq (M^{\mathrm{op}} \subset M_1^{\mathrm{op}})$, and S is generated by M and $N' \cap S$ then $M \vee M' \cap S \subset S$ is naturally isomorphic to $M \vee M^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$*

Proof. Let $e_0 \in M$ be a Jones projection for $N \subset M$ and $\{m_j\}_j$ an orthonormal basis of N over $N_1 = \{e_0\}' \cap N$ such that one of the m_j's equals 1. For $x \in S$ define $E(x) = \Sigma_j m_j e_0 x e_0 m_j^* \in S$. Note that if $x \in M' \cap S$ then m_j and e_0 commute with x so $E(x) = x$. Also, if $x \in N' \cap S$ then for each $y \in M$ we have

$$yE(x) = y \sum_j m_j e_0 x e_0 m_j^* = \lambda^{-1} \sum_{i,j} m_i e_0 E_N^M(e_0 m_i^* y m_j e_0) x e_0 m_j^*$$

$$= \lambda^{-1} \sum_{i,j} m_i e_0 x E_N^M(e_0 m_i^* y m_j e_0) e_0 m_j^* = \sum_i m_i e_0 x e_0 m_i^* y = E(x) y$$

showing that $[E(x), y] = 0$. Thus $E(x) \in M' \cap S$. This shows that E is a norm one projection of $N' \cap S$ onto $M' \cap S$ so by Tomiyama's theorem it is a conditional expectation. Also, if $x \in N' \cap S$ then we have

$$\tau(E(x)) = \tau(\Sigma_j m_j e_0 x e_0 m_j^*) = \tau(x e_0 \Sigma_j m_j^* m_j e_0)$$

$$= \Sigma_j \tau(x e_0 E_{N_1}^N(m_j^* m_j)) = \Sigma_j \tau(E_{N_1' \cap S}(x e_0 E_{N_1}^N(m_j^* m_j))) = \lambda^{-1}\tau(x e_0) = \tau(x)$$

Thus E is trace preserving as well, so it must coincide with the unique trace preserving expectation of $N' \cap S$ onto $M' \cap S$. Also, from the definition of $E(x)$, if $x \in N' \cap S$ then $e_0 x e_0 = E(x)e_0$. Thus, if $e_1 \in N' \cap S$ is a Jones projection for $(M' \cap S \subset N' \cap S) = (M^{\rm op} \subset M_1^{\rm op})$ then $e_0 e_1 e_0 = \lambda e_0$. But

$$\tau(e_0)^{-1} = [M : N] = [M^{\rm op} : N^{\rm op}] = [M_1^{\rm op} : M^{\rm op}] = \tau(e_1)^{-1},$$

so that $\tau(e_0) = \tau(e_1)$. Together with $e_0 e_1 e_0 = \lambda e_0$, this implies that $e_1 e_0 e_1 = \lambda e_1$. Thus, if $x, y \in N$ then $e_1(x e_0 y)e_1 = \lambda x y e_1 = E_N^M(x e_0 y)e_1$, showing that e_1 implements the conditional expectation E_N^M. But, by its definition, e_1 also implements the conditional expectation of $M' \cap S$ onto $\{e_1\}' \cap (M' \cap S)$.
Since we also have the isomorphism $(M' \cap S \subset N' \cap S) \simeq (M^{\rm op} \subset M_1^{\rm op})$, which in turn implements an isomorphism $(\{e_1\}' \cap M' \cap S \subset M' \cap S \subset N' \cap S) \simeq (N^{\rm op} \subset M^{\rm op} \subset M_1^{\rm op})$, 2.3 applies to yield $(M \vee M' \cap S \subset S) \simeq (M \vee M^{\rm op} \subset M \underset{e_N}{\boxtimes} M^{\rm op})$
. Q.E.D.

Note that from the above proposition and [Po2] it follows that if M is hyperfnite and the graph $\Gamma_{N,M}$ of $N \subset M$ is strongly amenable (see [Po2] for the definitions) then the inclusion $M \vee M^{\rm op} \subset M \underset{e_N}{\boxtimes} M^{\rm op}$ is isomorphic to the inclusion $M \vee M' \cap M_\infty \subset M_\infty$. The inclusions $M \vee M' \cap M_\infty \subset M_\infty$ for $N \subset M$ hyperfinite with finite depth, i.e., with finite (thus strongly amenable) graph, were considered and extensively studied by Ocneanu ([Oc], see also [EvKa]). Note that if M is an arbitrary type $\rm II_1$ factor and $N \subset M$ is a subfactor of finite depth and we denote by $Q \subset P$ the standard model $N^{\rm st} \subset M^{\rm st}$ then $\mathcal{G}_{Q,P} = \mathcal{G}_{N,M}$ and $M \vee M' \cap M_\infty \subset M_\infty$ naturally identifies with the "concatenation" inclusion considered in 2.5.1°, i.e., with $M \vee P^{\rm op} \subset M \boxtimes P^{\rm op}$.
The next lemma provides some useful localization properties relating the Jones projections, the relative commutants and the antiisomorphism $^{\rm op}$. They are reminiscent of some well known facts (see e.g., [PiPo1] page 83, [Bi1] page 205).

2.8. Lemma. *Let $N \subset M$ be an extremal inclusion, $N \subset M \overset{e_N}{\subset} M_1$ its basic construction and $^{\rm op}$ the canonical antiautomorphism of $N' \cap M$ onto $M' \cap M_1$ (so $x^{\rm op} = J_M x^* J_M$, $x \in N' \cap M$).*
a) *If $x \in N' \cap M$ then $x e_N = x^{\rm op} e_N$ and $x^{\rm op}$ is the unique element $y' \in M' \cap M_1$ such that $y' e_N = x e_N$.*
b) *$e_N x y^{\rm op} e_N = \tau(xy) e_N$ and $\tau(x y^{\rm op} e_N) = \lambda \tau(xy)$, $\forall\ x, y \in N' \cap M$, where $\lambda = [M : N]^{-1}$.*
c) *If $q \in \mathcal{P}(N' \cap M)$, $q \neq 0$, then $N q q^{\rm op} \subset q M q q^{\rm op} \subset q q^{\rm op} M_1 q q^{\rm op}$ is a basic construction with Jones projections equal to*

$$\tau(q)^{-1} q q^{\rm op} e_N q q^{\rm op} = \tau(q)^{-1} q e_N q = \tau(q)^{-1} q^{\rm op} e_N q^{\rm op}.$$

d) *$E_{M \vee M' \cap M_1}(e_N) = \lambda \sum_{i,j,k} \tau(f_{jj}^k)^{-1} f_{ij}^k f_{ji}^{k\,{\rm op}}$, where $\{f_{ij}^k\}$ is a matrix unit for $N' \cap M$.*

Proof. a) If $y \in \hat{M}$ then $e_N(\hat{y}) = \widehat{E_N(y)}$ so that

$$x e_N(\hat{y}) = x \widehat{E_N(y)} = \widehat{E_N(y)} x = x^{\rm op} e_N(\hat{y}).$$

The uniqueness is clear because $y'e_N = 0$ implies $e_N y'^* y' e_N = 0$ so that $E_{N^{\text{op}}}(y'^* y') = 0$, thus $y' = 0$.
b) By a) we have

$$e_N x y^{\text{op}} e_N = e_N x y e_N = E_N(xy) e_N = E_{N' \cap N}(xy) = \tau(xy) e_N,$$

whenever $x, y \in N' \cap M$. The second part is then trivial.
c) Since

$$q q^{\text{op}} \mathcal{B}(L^2(M)) q q^{\text{op}} = \mathcal{B}(L^2(qMq))$$

and

$$(N q q^{\text{op}}) \cap q q^{\text{op}} \mathcal{B}(L^2(M)) q q^{\text{op}} = q q^{\text{op}} M_1 q q^{\text{op}},$$

it follows that

$$N q q^{\text{op}} \subset q M q q^{\text{op}} \subset q q^{\text{op}} M_1 q q^{\text{op}}$$

is a basic construction. Also, if $e = \tau(q)^{-1} q q^{\text{op}} e_N q q^{\text{op}}$ then by a) we have

$$e = \tau(q)^{-1} q e_N q = \tau(q)^{-1} q^{\text{op}} e_N q^{\text{op}}.$$

Also, the range of $e = \tau(q)^{-1} q q^{\text{op}} e_N q q^{\text{op}}$ is clearly

$$L^2(Nq) = L^2(qNq) = q L^2(N) q = L^2(N) q$$

and so, since e is a projection we get $e = \text{proj}_{L^2(N)q} = \text{proj}_{L^2(N q q^{\text{op}})}$ as an element in $\mathcal{B}(L^2(qMq))$.
d) To prove this it is sufficient to show that $\tau(x y^{\text{op}} e_N) = \tau(x y^{\text{op}} a)$, $\forall\, x, y \in N' \cap M$, where $a = \lambda \sum_{i,j,k} \tau(f^k_{jj})^{-1} f^k_{ij} {f^k_{ji}}^{\text{op}}$. It is then enough to check this for $x = f^{k'}_{rs}$, $y = f^{k''}_{s'r'}$. For the left hand side, by b) we have:

$$\tau(f^{k'}_{rs} {f^{k''}_{s'r'}}^{\text{op}} e_N) = \lambda \delta_{k'k''} \delta_{ss'} \delta_{rr'} \tau(f^{k'}_{rr}).$$

For the right hand side we have:

$$\begin{aligned}
\tau(f^{k'}_{rs} {f^{k''}_{s'r'}}^{\text{op}} a) &= \lambda \sum_{i,j,k} \tau(f^k_{jj})^{-1} \tau(f^k_{ij} {f^k_{ji}}^{\text{op}} f^{k'}_{rs} {f^{k''}_{s'r'}}^{\text{op}}) \\
&= \lambda \sum_{i,j,k} \tau(f^{k'}_{rr})^{-1} \delta_{kk'} \delta_{kk''} \delta_{jr} \delta_{jr'} \tau(f^k_{is}) \tau(f^k_{is'}) \\
&= \lambda \tau(f^{k'}_{rr})^{-1} \delta_{k'k''} \delta_{rr'} \delta_{ss'} \tau(f^{k'}_{ss})^2 \\
&= \lambda \delta_{k'k''} \delta_{ss'} \delta_{rr'} \tau(f^{k'}_{rr}).
\end{aligned}$$

Q.E.D.

2.9. Proposition. *Let $N \subset M$ be an extremal inclusion of type* II_1 *factors. Then we have*
a) $M \underset{e_{N_{n-1}}}{\boxtimes} M^{\mathrm{op}}$ *naturally identifies with* $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$, *by letting* $e_{N_{n-1}} \mapsto f^n_{-n}$.
b) *The inclusion* $M_1 \vee N^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ *naturally identifies with the reduced by* e_1^{op} *of the symmetric enveloping inclusion of* $M \subset M_1$, $M_1 \vee M_1^{\mathrm{op}} \subset M_1 \underset{e_M}{\boxtimes} M_1^{\mathrm{op}}$. *More generally,* $M_n \vee N_{n-1}^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ *is isomorphic to the reduced by* $(f^n_{-n})^{\mathrm{op}}$ *of* $M_n \vee M_n^{\mathrm{op}} \subset M_n \underset{e_M}{\boxtimes} M_n$.
c) *If* $p \in \mathcal{P}(N' \cap M)$ *and we denote by* $L \subset K$ *the inclusion* $Np \subset pMp$ *then* $K \underset{e_L}{\boxtimes} K^{\mathrm{op}}$ *is naturally embedded in* $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ *as the weakly closed* $*$*-subalgebra generated by* $pp^{\mathrm{op}}(M \vee M^{\mathrm{op}})pp^{\mathrm{op}}$ *and by*

$$e'_L \stackrel{\mathrm{def}}{=} \sigma(p)^{-1} pp^{\mathrm{op}} e_N pp^{\mathrm{op}}.$$

Also, the ientity of this algebra is pp^{op}.
d) *If* $T \subset S$ *denotes the symmetric enveloping inclusion associated with* $N \subset M$ *and* $T_0 \subset S_0$ *the symmetric enveloping inclusion associated with some other extremal inclusion of type* II_1 *factors* $N_0 \subset M_0$, *then the symmetric enveloping inclusion associated with* $N \bar{\otimes} N_0 \subset M \bar{\otimes} M_0$ *is naturally isomorphic to* $T \bar{\otimes} T_0 \subset S \bar{\otimes} S_0$.

Proof. a) Is clear by 1.6, 1.9.4 and 2.4.
b) follows then immediately, from 2.1, 2.3 and the fact that $(f^n_{-n})^{\mathrm{op}} M_n^{\mathrm{op}} (f^n_{-n})^{\mathrm{op}} = N_{n-1}^{\mathrm{op}} (f^n_{-n})^{\mathrm{op}} \simeq N_{n-1}^{\mathrm{op}}$.
To prove c) note that if π is the canonical representation of $C^*(M, e_N, JMJ)$ into $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ then the C^*-algebra generated by $pp^{\mathrm{op}}(M \cup M^{\mathrm{op}})pp^{\mathrm{op}}$ and e'_L is the image under π of $C^*(pJpJ(M \cup JMJ)pJpJ, pJpJe_NpJpJ)$ which naturally identifies with the C*-algebra generated by K, $J_K K J_K$ and e_L in $\mathcal{B}(L^2(K))$, where $L = Np \subset pMp = K$. Since this representation of $C^*_{u,\max}(K, e_L, K^{\mathrm{op}})$ is smooth, it follows that π implements a smooth representation of $C^*_{u,\max}(K, e_L, K^{\mathrm{op}})$ into $pp^{\mathrm{op}}(M \underset{e_N}{\boxtimes} M^{\mathrm{op}})pp^{\mathrm{op}}$. Since the latter has a trace, it follows by 2.1, 2.2 that $K \underset{e_L}{\boxtimes} K^{\mathrm{op}} = \overline{(C^*(pp^{\mathrm{op}} M pp^{\mathrm{op}}, e'_L, pp^{\mathrm{op}} M^{\mathrm{op}} pp^{\mathrm{op}}))}^w \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$.
d) follows trivially from any of the characterizing universality properties of the symmetric enveloping algebras (e.g., from 2.7). Q.E.D.

2.10. Proposition. *Let $N \subset M$ be an extremal inclusion of type* II_1 *factors.*
a) *If* $Q \subset N$ *is an extremal subfactor of* N *then* $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ *is unitally embedded as a subfactor of* $M \underset{e_Q}{\boxtimes} M^{\mathrm{op}}$, *by taking* $e_N \mapsto \sum_j m_j e_Q m_j^*$ $(= \sum_j m_j^{*\mathrm{op}} e_Q m_j^{\mathrm{op}})$, *where* $\{m_j\}_j$ *is a orthonormal basis of* N *over* Q. *Moreover, if there exists a tunnel* $M \supset N \supset N_1 \supset \cdots$ *for* $N \subset M$ *such that* $N_k \subset Q$ *for some* k, *then this unital embedding is in fact an equality.*

b) *If* $Q \subset P$ *is an extremal inclusion of factors embedded in* $N \subset M$ *as a commuting square, such that* $[P : Q] = [M : N]$ *and* $P' \cap P_n \subset M' \cap M_n$, $\forall\, n$ *then* $P \underset{e_Q}{\boxtimes} P^{\mathrm{op}}$ *is unitally embedded in* $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$, *by taking* $P \hookrightarrow M$, $P^{\mathrm{op}} \hookrightarrow M^{\mathrm{op}}$ *and* $e_Q \mapsto e_N$. *Also,* $P \underset{e_Q}{\boxtimes} P^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ *has finite index iff* $P \subset M$ *has finite index, with the estimate* $[M \underset{e_N}{\boxtimes} M^{\mathrm{op}} : P \underset{e_Q}{\boxtimes} P^{\mathrm{op}}] \leq [M : P]^2$. *Moreover, if* $P' \cap P_n = M' \cap M_n$, *then this embedding implements the nondegenerate commuting square:*

$$\begin{array}{ccc} M \vee M^{\mathrm{op}} & \subset & M \underset{e_N}{\boxtimes} M^{\mathrm{op}} \\ \cup & & \cup \\ P \vee P^{\mathrm{op}} & \subset & P \underset{e_Q}{\boxtimes} P^{\mathrm{op}}. \end{array}$$

Proof. a) It is easy to check by direct computation that

$$e_N = \sum_j m_j e_Q m_j^* = \sum_j (J m_j J) e_Q (J m_j^* J)$$

in $\mathcal{B}(L^2(M))$, so a) follows from 2.2-2.6. The last part of a) then follows from 2.7 a) and the first part.

b) To prove the first part we only need to show that the representation of $C^*_{u,\mathrm{bin}}(P, e_Q, P^{\mathrm{op}})$ in $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ satisfies the faithfulness condition 1.8 $(*)$, i.e., we need to show that $P' \cap P_n = {Q^{\mathrm{op}}_{n-1}}' \cap P^{\mathrm{op}}$, $\forall\, n$. By 2.1–2.4, it is sufficient to check this equality in a representation of $C^*_{\mathrm{bin}}(M, e_N, M^{\mathrm{op}})$, and we'll choose $C^*(M, e_N, JMJ) \subset \mathcal{B}(L^2(M))$ to do this. Let $e_P^M \in \mathcal{B}(L^2(M))$ be the orthogonal projection of $L^2(M)$ onto $L^2(P)$. Note that all the elements in $C^*(P, e_N, J_M P J_M) \subset \mathcal{B}(L^2(M))$ commute with e_P^M and that if $x \in JMJ$ then $x \in JPJ$ iff $[x, e_P^M] = 0$. Now, if $x \in P' \cap P_n$ then $x \in M' \cap P_n$ by hypothesis, so $x \in JMJ \cap P_n$. Also, $[x, e_P^M] = 0$, because $P_n \subset C^*(P, e_N, JPJ)$. Thus, $x \in JPJ \cap P_n$. But $P_n \subset M_n = J N_{n-1} J' \cap \mathcal{B}(L^2(M)) \subset J Q_{n-1} J'$. Thus, $x \in JPJ \cap (J Q_{n-1} J)'$. This proves the first part of b).

Further on, assume $[M : P] < \infty$ and take $\{m_j\}_j$ to be a finite orthonormal basis of M over P. Note that $M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = \overline{\mathrm{sp}}(M \vee M^{\mathrm{op}}) \mathrm{vN}\{e_j\}_{j \in \mathbb{Z}}$ and $P \underset{e_Q}{\boxtimes} P^{\mathrm{op}} = \overline{\mathrm{sp}}(P \vee P^{\mathrm{op}}) \mathrm{vN}\{e_j\}_{j \in \mathbb{Z}}$, with $\{e_j\}_j \subset P_\infty$ being the Jones projections for a tower-tunnel for $Q \subset P$ (see 4.1, 4.2), and thus for $N \subset M$ as well. Since $M \vee M^{\mathrm{op}} = \Sigma_{i,j} m_i m_j^{*\mathrm{op}} P \vee P^{\mathrm{op}}$ it thus follows that $M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = \Sigma_{i,j} m_i m_j^{*\mathrm{op}} P \underset{e_Q}{\boxtimes} P^{\mathrm{op}}$, showing that $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ is a finitely generated left module over $P \underset{e_Q}{\boxtimes} P^{\mathrm{op}}$, with the estimate $[M \underset{e_N}{\boxtimes} M^{\mathrm{op}} : P \underset{e_Q}{\boxtimes} P^{\mathrm{op}}] \leq [M : P]^2$ as a bonus. For the last part, we have that $\bigcup_n \mathrm{sp} P^{\mathrm{op}} P_n P^{\mathrm{op}}$ is so-dense in $P \underset{e_Q}{\boxtimes} P^{\mathrm{op}}$ and writing P_n as $\mathrm{sp} P f_{-n}^n P$ we get $E_{M \vee M^{\mathrm{op}}}(P \underset{e_Q}{\boxtimes} P^{\mathrm{op}}) = \overline{\mathrm{sp}} \bigcup_n ((P \vee P^{\mathrm{op}}) E_{M \vee M^{\mathrm{op}}}(f_{-n}^n)(P \vee P^{\mathrm{op}}))$. But since $P' \cap P_n = M' \cap M_n$ and $Q'_{n-1} \cap P =$

$N'_{n-1} \cap M$, $\forall\ n$, it follows that $E_{M\vee M^{\rm op}}(f^n_{-n}) = E_{P\vee P^{\rm op}}(f^n_{-n})$, proving the desired commuting square condition. Q.E.D.

Let us end this section by considering a notion of index for sublattices of standard λ-lattices (see [Po7] for the definition of abstract standard lattices and for the notations and results used hereafter). We relate this notion with the content of this section by showing that the index of a sublattice coincides with the index of a certain canonically associated inclusion of symmetric enveloping type II_1 factors. This latter result will be used in Sections 5 and 8.

2.11. Definition. Let $\mathcal{G} = (A_{ij})_{0\leq i\leq j}$ be a standard λ-lattice and $\mathcal{G}_0 = (A^0_{ij})_{0\leq i\leq j}$ a sublattice. We define the index of $\mathcal{G}_0$ in $\mathcal{G}$ by $[\mathcal{G} : \mathcal{G}_0] \stackrel{def}{=} \lim_{n\to\infty} \mathrm{Ind}E^{A_{0n}}_{A^0_{0n}} = \mathrm{Ind}E^{A_{0,\infty}}_{A^0_{0,\infty}}$, where $\mathrm{Ind}(E)$ denotes as usual the index ([PiPo1]) of the conditional expectation E and $A_{i,\infty} = \overline{\cup_n A_{in}}$, $A^0_{i,\infty} = \overline{\cup_n A_{in}}$.

Let us make right away some comments on this definition. By (1.1.6 in [Po3]), if $\{m_j\}_j$ is an orthonormal basis of $A_{0,\infty}$ over $A^0_{0,\infty}$ (apriorically made up of square summable operators) then $\|\Sigma_j m_j m_j^*\| = \mathrm{Ind}E^{A_{0,\infty}}_{A^0_{0,\infty}}$. But both $A_{1,\infty} \subset A_{0,\infty}$ and $A^0_{1,\infty} \subset A^0_{0,\infty}$ are λ-Markov inclusions (see 1.1.5 in [Po2] for the definition), so the commuting square embedding of the latter into the former is nondegenerate (1.5,1.6 in [Po2]). Thus, by (1.6 in [Po2]) any orthonormal basis of $A_{1,\infty}$ over $A^0_{1,\infty}$ is an orthonormal basis of $A_{0,\infty}$ over $A^0_{0,\infty}$. Thus, in the above we may assume that $\{m_j\}_j$ lies in $A_{1,\infty}$. On the other hand, if bounded, $\Sigma_j m_j m_j^*$ belongs to the center of $A_{0,\infty}$ (see e.g. 1.1.5 in [Po3]), thus $\Sigma_j m_j m_j^* \in \mathcal{Z}(A_{0,\infty}) \cap A_{1,\infty} = \mathcal{Z}(A_{0,\infty}) \cap \mathcal{Z}(A_{1,\infty})$. But by (Corollary 1.4.2 in [Po2]) this latter intersection is in fact equal to the scalar multiples of the identity. Thus, $\Sigma_j m_j m_j^* \in \mathbb{C}1$. Altogether, this shows that we may as well take $[\mathcal{G} : \mathcal{G}_0] \stackrel{def}{=} \|\Sigma_j m_j m_j^*\| = \Sigma_j m_j m_j^*$, $\{m_j\}_j$ being an arbitrary orthonormal basis of $A_{i,\infty}$ over $A^0_{i,\infty}$, for some $i \geq 0$. The next proposition gives more ways to look at this index.

2.12. Proposition. *Let $\mathcal{G}$ be a standard λ-lattice with a sublattice $\mathcal{G}_0$. Let Q_0 be a non-atomic finite von Neumann algebra with a faithful trace and $N^{\mathcal{G}}(Q_0) \subset M^{\mathcal{G}}(Q_0)$, respectively $N^{\mathcal{G}_0}(Q_0) \subset M^{\mathcal{G}_0}(Q_0)$ be the associated extremal inclusions of type II_1 factors having $\mathcal{G}$, respectively $\mathcal{G}_0$ as standard invariants, given by the universal construction in ([Po7]). Let*

$$\begin{array}{ccc} N^{\mathcal{G}}(Q_0) & \subset & M^{\mathcal{G}}(Q_0) \\ \cup & & \cup \\ N^{\mathcal{G}_0}(Q_0) & \subset & M^{\mathcal{G}_0}(Q_0). \end{array}$$

be the corresponding commuting square like in ([Po7]). Then we have

$$[\mathcal{G} : \mathcal{G}_0] = [M^{\mathcal{G}}(Q_0) : M^{\mathcal{G}_0}(Q_0)] = [S : S_0],$$

where S and respectively S_0 denote the symmetric enveloping algebras of $N^{\mathcal{G}}(Q_0) \subset M^{\mathcal{G}}(Q_0)$ and respectively $N^{\mathcal{G}_0}(Q_0) \subset M^{\mathcal{G}_0}(Q_0)$.

Proof. Recall from [Po7] that $M^{\mathcal{G}}_\infty(Q_0)$ identifies with the free product with amalgamation $Q_0 \bar{\otimes} A_{1,\infty} *_{A_{1,\infty}} A_{0,\infty}$, with $M^{\mathcal{G}_0}_\infty(Q_0)$ identifying with the subalgebra generated by Q_0 and $A^0_{0,\infty}$. By the resulting commuting square relations for these inclusions (see [Po7], pages 435 and 438), it follows that any orthonormal basis of $A_{1,\infty}$ over $A^0_{1,\infty}$ is an orthonormal basis of $Q_0 \bar{\otimes} A_{1,\infty} *_{A_{1,\infty}} A_{0,\infty}$ over $Q_0 \bar{\otimes} A^0_{1,\infty} *_{A^0_{1,\infty}} A^0_{0,\infty}$, thus of $M^{\mathcal{G}}_\infty(Q_0)$ over $M^{\mathcal{G}_0}_\infty(Q_0)$. But the commuting square embedding of $M^{\mathcal{G}_0}(Q_0) \subset M^{\mathcal{G}}(Q_0)$ into $M^{\mathcal{G}_0}_\infty(Q_0) \subset M^{\mathcal{G}}_\infty(Q_0)$ is nondegenarate (cf. [Po7]), so that in the end, if $\{m_j\}$ denotes an orthonormal basis of $A_{1,\infty}$ over $A^0_{1,\infty}$, we get $[M^{\mathcal{G}}(Q_0) : M^{\mathcal{G}_0}(Q_0)] = [M^{\mathcal{G}}_\infty(Q_0) : M^{\mathcal{G}_0}_\infty(Q_0)] = \Sigma_j m_j m_j^* = [\mathcal{G} : \mathcal{G}_0]$.
Finally, from the universality properties of the symmteric enveloping algebras and the definition of $N^{\mathcal{G}}(Q_0) \subset M^{\mathcal{G}}(Q_0)$ and $N^{\mathcal{G}_0}(Q_0) \subset M^{\mathcal{G}_0}(Q_0)$, we see that, if we denote by $N \subset M$ and $N_0 \subset M_0$ these two inclusions then $S_0 \subset S$ identifies with the inclusion $Q_0 \bar{\otimes} N_0^{\mathrm{op}} *_{N_0^{\mathrm{op}}} M_0^{\mathrm{op}} \subset Q_0 \bar{\otimes} N^{\mathrm{op}} *_{N^{\mathrm{op}}} M^{\mathrm{op}}$. But from the above we have that any orthonormal basis of N^{op} over N_0^{op} will be an orthonormal basis of S over S_0. Q.E.D.

3. A Class of Examples

Let Q be a type II_1 factor and $\sigma_1, \dots, \sigma_n$ a n-tuple of automorphisms of Q. Let $N \subset M$ be the locally trivial inclusion of factors associated with $\sigma_1, \dots, \sigma_n$ (see e.g. [Po2]), i.e., $M = Q \otimes M_{n+1}(\mathbb{C})$, $N = \{\sum_{i=0}^n \sigma_i(x) \otimes e_{ii} \mid x \in Q \simeq Q \otimes \mathbb{C}1\}$, where $\sigma_0 = \mathrm{id}_Q$ and $\{e_{ij}\}_{0 \le i,j \le n}$ is a matrix unit for $M_{n+1}(\mathbb{C})$.
We still denote by σ_i the automorphism of $M = Q \otimes M_{n+1}(\mathbb{C})$ defined by $\sigma_i(x \otimes e_{kl}) = \sigma_i(x) \otimes e_{kl}$, $\forall\ x \in Q$, $0 \le k,l \le n$. Denote by G the discrete group generated by $\sigma_1, \dots, \sigma_n$ in $\mathrm{Aut}(M)/\mathrm{Int}(M)$. Also, we let $\sigma : G \to \mathrm{Aut}(M)/\mathrm{Int}(M)$ be the corresponding faithful G-kernel. Then note that the faithful G-kernel $\sigma \otimes \sigma^{\mathrm{op}}$ on $M \bar{\otimes} M^{\mathrm{op}}$ has vanishing $H^3(G, \mathbb{T})$ cohomology obstruction ([J5]), so that it can be viewed as a (properly outer) cocycle action of G on $M \bar{\otimes} M^{\mathrm{op}}$.
In this section we show that, with the above notations, we have

$$(M \vee M^{\mathrm{op}} \subset M \underset{e_Q}{\boxtimes} M^{\mathrm{op}}) \simeq (M \bar{\otimes} M^{\mathrm{op}} \subset (M \bar{\otimes} M^{\mathrm{op}}) \rtimes_{\sigma \otimes \sigma^{\mathrm{op}}} G),$$

in which the cross product is associated with the cocycle action $\sigma \otimes \sigma^{\mathrm{op}}$ as in (4.1 of [J5]). Since by the previous sections $M \vee M^{\mathrm{op}} \subset M \underset{e_Q}{\boxtimes} M^{\mathrm{op}}$ is the (weak closure of the) quotient of $C^*(M, JMJ) \subset C^*(M, e_N, JMJ)$, it will be sufficient to study this latter inclusion of algebras.
So let U_i be the unitary element acting on $L^2(M, \tau)$, defined on the dense subset $\hat{M} \subset L^2(M, \tau)$ by $U_i(\hat{x}) = \widehat{\sigma_i(x)}$, $x \in M$, $0 \le i \le n$. Note that $U_i x U_i^* = \sigma_i(x)$, $\forall\ x \in M$, $0 \le i \le n$, and $[J, U_i] = 0$. In particular, since $\sigma_i(e_{kl}) = e_{kl}$, $0 \le k,l \le n$, we also have $[U_i, e_{kl}] = 0$, $[U_i, J e_{kl} J] = 0$, $\forall\ i,k,l$.

3.1. Lemma.

a) $e_N = \dfrac{1}{n+1}\sum_{i,j=0}^{n} U_jU_i^* e_{ji}Je_{ji}J$.

b) $U_j = (n+1)\sum_{k,l=0}^{n} Je_{lj}Je_{kj}e_Ne_{0k}Je_{0l}J$.

Proof. a) If $x = \sum x_{ij}\otimes e_{ij} \in Q\otimes M_{n+1}(\mathbb{C}) = M$, then

$$\widehat{E_N(x)} = \frac{1}{n+1}\sum_{i,j}(\sigma_j\sigma_i^{-1}(x_{ii})\otimes e_{jj})\hat{\ } = \frac{1}{n+1}\sum_{i,j=0}^{n} U_jU_i^* e_{ji}Je_{ji}J(\hat{x})$$

proving the first formula.

b) By a) we have $e_{jj}e_NJe_{00}J = \dfrac{1}{n+1}U_je_{j0}Je_{j0}J$, so that $U_je_{00}Je_{00}J = (n+1)e_{0j}Je_{0j}Je_{jj}e_NJe_{00}J$. Thus we get

$$U_j = \sum_{k,l=0}^{n} e_{k0}Je_{l0}J(U_je_{00}Je_{00}J)Je_{0l}Je_{0k} = (n+1)\sum_{k,l=0}^{n} e_{kj}Je_{lj}Je_NJe_{0l}Je_{0k}.$$

Q.E.D.

3.2. Corollary. *$C^*(M, e_N, JMJ) = C^*(M, \{U_i\}_{i\le n}, JMJ)$. In fact,*

$$C^*(M_{n+1}(\mathbb{C}), e_N, JM_{n+1}(\mathbb{C})J) = C^*(M_{n+1}(\mathbb{C}), \{U_i\}_{0\le i\le n}, JM_{n+1}(\mathbb{C})J).$$

Proof. Trivial by the previous lemma. Q.E.D.

Describing $M\vee M^{\rm op} \subset M\underset{e_N}{\boxtimes} M^{\rm op}$ as a cross product is now an immediate consequence of the previous lemma and of 2.1, once we notice that $U_j(xJy^*J)U_j^* = \sigma_i(x)J\sigma_i(y^*)J$. To write the corresponding isomorphism in more specific terms, denote by u_i the image of U_i in $M\underset{e_N}{\boxtimes} M^{\rm op}$ (cf. 2.1) and by g_i the image of σ_i as an element of the group G.

3.3 Theorem. *There exists a unique isomorphism γ, of $(M\vee M^{\rm op} \subset M\underset{e_N}{\boxtimes} M^{\rm op})$ onto $(M\bar{\otimes}M^{\rm op} \subset M\bar{\otimes}M^{\rm op}\rtimes_{\sigma\otimes\sigma^{\rm op}} G)$, satisfying:*

a) $\gamma(xy^{\rm op}) = x\otimes y^{\rm op}$, $x,y\in M$.

b) $u_{g_i} \overset{\rm def}{=} \gamma(u_i)$ are unitary elements in the cross product $M\bar{\otimes}M^{\rm op}\rtimes_{\sigma\otimes\sigma^{\rm op}} G$ which implement the automorphism $\sigma\otimes\sigma^{\rm op}(g_i)$, $0\le i\le n$.

c) $\gamma(e_N) = \dfrac{1}{n+1}\sum_{i,j=0}^{n} u_{g_j}u_{g_i}^* e_{ji}\otimes e_{ij}^{\rm op}$.

d) $\gamma^{-1}(u_{g_j}) = u_j = (n+1)\sum_{k,l=0}^{n} e_{kj}e_{jl}^{\rm op}e_Ne_{0k}e_{l0}^{\rm op}$, $0\le j\le n$.

Proof. Trivial by 2.1 and 3.1. Q.E.D.

In the next section we will see that even for arbitrary extremal subfactors $N \subset M$ the resulting inclusion $M \vee M^{\text{op}} \subset M \underset{e_N}{\boxtimes} M^{\text{op}}$ can be interpreted as a 'cross-product'-type structure.

3.4 *Remarks.* 1°. As one knows (see e.g. 5.1.5 in [Po2]), the standard invariant $\mathcal{G}_{N,M}$ of the above locally trivial subfactor $N \subset M$ only depends on the cohomology obstruction in $H^3(G, \mathbb{T})$ ([J5]) of the corresponding G-kernel σ on Q. Thus, if we take another G-kernel σ', on another type II_1 factor Q' but with the same $H^3(G, \mathbb{T})$-obstruction as σ, and denote the similar locally trivial inclusion (corresponding to the same generators of G) by $N' \subset M'$, then $\mathcal{G}_{N',M'} = \mathcal{G}_{N,M}$ and we can thus consider the concatenation algebra 2.5.1° associated with these two inclusions. Then $M \vee M'^{\text{op}} \subset M \boxtimes M'^{\text{op}}$ is isomorphic to a cocycle cross product $M \bar{\otimes} M'^{\text{op}} \subset (M \bar{\otimes} M'^{\text{op}}) \rtimes_{\sigma \otimes \sigma'^{\text{op}}} G$.
2°. Let $\mathcal{G} = (A_{ij})_{0 \leq i \leq j}$ be the standard λ-lattice associated to the locally trivial subfactor $N \subset M$, constructed from the automorphisms $\sigma_1, ..., \sigma_n$ acting on the factor Q as above, with G denoting the group generated by the σ_i's in $\text{Aut}(Q)/\text{Int}(Q)$ (and with the corresponding generators denoted hereafter by $g_1, ..., g_n$). Let $\mathcal{G}_0 = (A^0_{ij})_{i,j}$ be a sublattice of $\mathcal{G}$ with the property that A^0_{01} is a maximal abelian subalgebra of A_{01}. Note that this amounts to saying that $\mathcal{G}_0$ has same "generators" but possibly lesser "relations" than $\mathcal{G}$. Now take Q_0 to be an arbitrary finite von Neumann algebra without atoms. With Q_0 as "initial data", do the universal construction [Po7] of subfactors $N^{\mathcal{G}}(Q_0) \subset M^{\mathcal{G}}(Q_0)$ and $N^{\mathcal{G}_0}(Q_0) \subset M^{\mathcal{G}_0}(Q_0)$ with higher relative commutants picture given by $\mathcal{G}$ respectively $\mathcal{G}_0$, like at the end of Sec. 2, thus obtaining the non-degenerate commuting square of inclusions:

$$\begin{array}{ccc} N^{\mathcal{G}}(Q_0) & \subset & M^{\mathcal{G}}(Q_0) \\ \cup & & \cup \\ N^{\mathcal{G}_0}(Q_0) & \subset & M^{\mathcal{G}_0}(Q_0). \end{array}$$

One can then show that the above algebras and the inclusions involved can be alternatively described in terms of the following objects:
a). A type II_1 factor Q' with a faithful G kernel σ' on it such that if $N \subset M$ denotes the locally trivial subfactor constructed out of this G-kernel and the generators $g_1, ..., g_n$, like at the beginning of this section, then $(N \subset M) \simeq (N^{\mathcal{G}}(Q_0) \subset M^{\mathcal{G}}(Q_0))$;
b). An irreducible regular (in the sense of [D1]) subfactor $Q'_0 \subset Q'$, a group G_0 with generators $g'_1, ..., g'_n$ and a G_0-kernel σ'_0 on Q'_0 such that if $N_0 \subset M_0$ denotes the associated locally trivial subfactor, constructed from this G_0-kernel and the generators $g'_1, ..., g'_n$, like at the beginning of this section, then $(N_0 \subset M_0) \simeq (N^{\mathcal{G}_0}(Q_0) \subset M^{\mathcal{G}_0}(Q_0))$;
c). A group morphism ρ of G_0 onto G such that $\rho(g'_i) = g_i$ and such that if $H = \ker(\rho)$ denotes the corresponding kernel group then H is isomorphic to $\mathcal{N}_{Q'}(Q'_0)/\mathcal{U}(Q'_0)$ (so that Q' is a cocycle cross-product of Q'_0 by H), in such a way that if we denote by $\{u_h\}_{h \in H}$ a set of unitaries in $\mathcal{N}_{Q'}(Q'_0)$ that give

a cross-section for H then, modulo perturbations by inner automorphisms, σ' and σ_0' are related as follows: $\sigma'(\rho(g'))(u_h a_0') = u_{hgh^{-1}}\sigma_0'(g')(a_0'), \forall a_0' \in Q_0', h \in H, g' \in G_0$.
Moreover, through these identifications, $N_0 \subset M_0$ is embedded in $N \subset M$ by the inclusion $M_0 = Q_0' \otimes M_{n+1}(\mathbb{C}) \subset Q' \otimes M_{n+1}(\mathbb{C}) = M$, and the corresponding commuting square is isomorphic to the above commuting square.
Thus, in this exemple the sublattice $\mathcal{G}_0$ of the lattice $\mathcal{G}$ (which was associated to the group G) corresponds to a "covering" group G_0 of the group G. Note that, with these identifications, we have that the index of $\mathcal{G}_0$ in $\mathcal{G}$ equals the order of the group H, $[\mathcal{G} : \mathcal{G}_0] = |H|$.
Finally, let us see what the symmetric enveloping algebras become in this case: if we extend the atomorphisms $\sigma'(g), \sigma_0'(g')$ to M, M_0 by putting them to act as the identity on $M_{n+1}(\mathbb{C})$, then the symmetric enveloping algebras S, S_0 of $N \subset M$ respectively $N_0 \subset M_0$, and the corresponding inclusion $S_0 \subset S$ (cf. 2.10, b)), are given by

$$S_0 = M_0 \bar{\otimes} M_0^{\mathrm{op}} \rtimes_{\sigma_0' \otimes \sigma_0'^{\mathrm{op}}} G_0 \subset M \bar{\otimes} M^{\mathrm{op}} \rtimes_{\sigma' \otimes \sigma'^{\mathrm{op}}} G = S$$

with the inclusion being described similarily to c).

4. Thinness and Quasi-Regularity Properties

We've already seen that $\mathrm{sp}\bigcup_n M M_n^{\mathrm{op}} M = \mathrm{sp}\bigcup_n M^{\mathrm{op}} M_n M^{\mathrm{op}}$ is a $*$-subalgebra which is dense in $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ in the weak (or strong) operator topology. Let $\{e_n\}_{n\in\mathbb{Z}}$ be the Jones projections for the Jones tunnel-tower $\cdots N_1 \subset N \subset M \subset M_1 \subset \cdots$, with $e_N = e_1$, as in Sections 1–2, and denote by P the von Neumann algebra they generate in $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$. Fix $n \geq 0$ and choose an orthonormal basis $\{m_j\}_j$ of M over N_{n-1} that belongs to $\mathrm{vN}\{e_k\}_{k\leq 0} \subset P$ and an orthonormal basis $\{m_k^n\}_k$ of M_n over M that belongs to $\mathrm{vN}\{e_k\}_{k\leq n} \subset P$. Thus we have

$$\begin{aligned} M M_n^{\mathrm{op}} M &\subset M\left(\sum_k M^{\mathrm{op}} m_k^{n\,\mathrm{op}}\right)\left(\sum_j N_{n-1} m_j^*\right) = \sum_{j,k} M M^{\mathrm{op}} N_{n-1} m_k^{n\,\mathrm{op}} m_j^* \\ &= \sum_{j,k} M M^{\mathrm{op}} m_k^{n\,\mathrm{op}} m_j^* \subset \mathrm{sp}\, M M^{\mathrm{op}} P. \end{aligned}$$

Thus we obtain $\mathrm{sp}\bigcup_n M M_n^{\mathrm{op}} M \subset \mathrm{sp}(M \vee M^{\mathrm{op}})P$. Similarly, since $\sum_{j,k} M m_k^{n\,\mathrm{op}} m_j^* \subset M_\infty$, we get $\mathrm{sp}\bigcup_n M M_n^{\mathrm{op}} M \subset \mathrm{sp}\, M^{\mathrm{op}} M_\infty$, giving us the following:

4.1. Proposition. *With the above notations we have:*

$$\begin{aligned} S \overset{\mathrm{def}}{=} M \underset{e_N}{\boxtimes} M^{\mathrm{op}} &= \overline{\mathrm{sp}}(M \vee M^{\mathrm{op}})P = \overline{\mathrm{sp}}\, M_\infty M^{\mathrm{op}} \\ &= \overline{\mathrm{sp}}(M \vee M^{\mathrm{op}})(\mathrm{Alg}\{f_{-n}^n\}_n)(M \vee M^{\mathrm{op}}) \end{aligned}$$

the closure being taken in either of the wo, so *or* $\| \|_2$ *topologies in* S.

Proof. Since $\mathrm{sp}(M \vee M^{\mathrm{op}})P$, $M_\infty M^{\mathrm{op}}$ and $\mathrm{sp}(M \vee M^{\mathrm{op}})(Alg\{f_{-n}^n\}_n)(M \vee M^{\mathrm{op}})$ contain $\mathrm{sp} \bigcup_n M M_n^{\mathrm{op}} M$, which is a dense $*$-subalgebra in $S = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$, we are done. Q.E.D.

Note that if M is hyperfinite then $M \vee M^{\mathrm{op}}$, M^{op}, M_∞ are all hyperfinite. Thus, in this case $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ can be written as a "product" of two hyperfinite subfactors. Recall from ([Po5]) that such situation is singled out by the following:

4.2. DEFINITION. A type II_1 factor S for which there exist two hyperfinite type II_1 subfactors $R_1, R_2 \subset S$ such that $S = \overline{\mathrm{sp}}\, R_1 R_2$, the closure being taken in $\| \|_2$, is called a *thin* type II_1 factor.

With this terminology the above observation takes the form:

4.3. COROLLARY. *If* $N \subset M$ *is a extremal inclusion of hyperfinite type* II_1 *factors then* $S = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ *is a thin type* II_1 *factor.*

From the above, the previous section and Connes' fundamental theorem ([C1]) we can already conclude:

4.4. COROLLARY. *If* $N \subset M$ *is an inclusion of factors associated to a faithful* G*-kernel* σ *on a hyperfinite type* II_1 *factor* R *like on Section* 3, *where* G *is a finitely generated discrete group, then* $M \underset{e_Q}{\boxtimes} M^{\mathrm{op}} \simeq R \otimes R^{\mathrm{op}} \rtimes_{\sigma \otimes \sigma^{\mathrm{op}}} G$ *is thin but it is hyperfinte iff* G *is amenable.*

More precise statements along these lines will be obtained in Sec.5 and 7. Let us note now that the Hilbert space $\mathcal{K}_n$ obtained as the closure of

$$(\mathrm{sp} M^{\mathrm{op}} M_n M^{\mathrm{op}})\hat{} = (\mathrm{sp} M M_n^{\mathrm{op}} M)\hat{}$$

in $L^2(M \underset{e_N}{\boxtimes} M^{\mathrm{op}}, \tau)$ is invariant to multiplication from left and right by both M and M^{op}, thus by $T = M \vee M^{\mathrm{op}}$. Thus $\mathcal{K}_n$ is a T-T bimodule.
Since $\mathrm{sp}\, M M_n^{\mathrm{op}} M = \mathrm{sp} \sum_{j,k} m_k m_j^{*\mathrm{op}} f_{-n}^n\ M M^{\mathrm{op}} = \mathrm{sp} \sum_{j,k} M M^{\mathrm{op}} f_{-n}^n m_j^{\mathrm{op}} m_k^*$, it follows that $\mathcal{K}_n$ has finite dimension both as a left and as a right T module. Thus, if p_n is the orthogonal projection of $L^2(S, \tau)$ onto $\mathcal{K}_n$ the p_n commutes with the operators of left and right multiplication by elements in T, i.e., $p_n \in T' \cap \langle S, T \rangle$. Also, since $\overline{\bigcup_n \mathcal{K}_n} = L^2(S, \tau)$, we have $p_n \nearrow 1$ and the above shows that $\mathrm{Tr}\, p_n < \infty$, $\forall\, n$, where $\mathrm{Tr} = \mathrm{Tr}_{\langle S, T \rangle}$ denotes the unique trace on $\langle S, T \rangle$ satisfying $\mathrm{Tr}(e_T) = 1$.
Thus, $T' \cap \langle S, T \rangle$ is generated by finite projections of $\langle S, T \rangle$ and the inclusion of factors $T p_n \subset p_n \langle S, T \rangle p_n$ has finite index for all n. Since $T' \cap S = \mathbb{C}1$ (cf. 2.3), by (1.8 in [PiPo1]) we can already conclude that $\mathrm{Tr}\, p \geq 1$, $\forall\, p \in T' \cap \langle S, T \rangle$ (so in particular $T' \cap \langle S, T \rangle$ is atomic) and that the multiplicity of any minimal projection p in $T' \cap \langle S, T \rangle$ is $\leq \mathrm{Tr}\, p$.
In fact we have the following more precise statement:

4.5. Theorem. *Let $N \subset M$ be an extremal inclusion of type* II_1 *factors and denote* $S = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$, $T = M \vee M^{\mathrm{op}} \subset S$.

a) *If* $\{\mathcal{H}_k\}_{k\in K}$ *denotes the set of irreducible* M-M *bimodules corresponding to the set of even vertices of the standard graph* $\Gamma_{N,M}$ *of* $N \subset M$ *then* $L^2(S,\tau)$ *is isomorphic as a* T-T *bimodule with* $\bigoplus_{k\in K} \mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}_k}^{\mathrm{op}}$ *and* $\mathcal{K}_n$ *with* $\bigoplus_{k\in K_n} \mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}_k}^{\mathrm{op}}$ *(in which* $T \simeq M \bar{\otimes} M^{\mathrm{op}}$*).*

b) *If* $L^2(S,\tau)$ *is identified with* $\bigoplus_{k\in K} \mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}_k}^{\mathrm{op}}$ *as in* a) *and* s_k *denotes the orthogonal projection of* $L^2(S,\tau)$ *onto its direct summand* $\mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}_k}^{\mathrm{op}}$ *then* s_k *is a minimal projection in* $T' \cap \langle S, T\rangle$, $p_n = \sum_{k\in K_n} s_k$ *and* $T' \cap \langle S,T\rangle = \mathrm{vN}\{s_k\}_{k\in K} \simeq \ell^\infty(K)$. *Moreover,* $(\mathrm{Tr}\, s_k)^2 = [s_k\langle S,T\rangle s_k : T s_k] = v_k^4$, *where* $\vec{v} = (v_k)_{k\in K}$ *is the standard vector giving the weights at the even vertices of* $\Gamma_{N,M}$.

c) *The antiautomorphism* ${}^{\mathrm{op}}$ *on* S *leaves* T *invariant and thus implements an antiautomorphism on* $\langle S,T\rangle$, *still denoted by* ${}^{\mathrm{op}}$. *We have* $(T' \cap \langle S,T\rangle)^{\mathrm{op}} = T' \cap \langle S,T\rangle$, *the projection* s_k^{op} *coincides with* $J_S s_k J_S$ *and the corresponding bimodule is* $(\mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}_k}^{\mathrm{op}})^- = \bar{\mathcal{H}}_k \bar{\otimes} \mathcal{H}_k^{\mathrm{op}}$.

Proof. Let $k \in K_1$ and choose $q = q_k \in N_1' \cap M$ to be a minimal projection in the direct summand labeled by k. Denote $v_q' = (\lambda\tau(q))^{1/2} q q^{\mathrm{op}} e_1 e_0 e_0^{\mathrm{op}}$ and $v_q = \lambda^{-2} E_{N'\cap M_1}(v_q')$. Note that $f = v_q' {v_q'}^*$ is the Jones projection for the irreducible inclusion $q^{\mathrm{op}} q N_1 \subset q^{\mathrm{op}} q M q \subset q^{\mathrm{op}} q M_2 q q^{\mathrm{op}}$ (cf. 2.8.b) and 2.8.c)). Note also that by applying twice the "push down lemma" (1.2 in [PiPo1]) and using the above definitions we get:

$$v_q e_0^{\mathrm{op}} e_0 = \lambda^{-2} E_{N'\cap M_1}(v_q') e_0^{\mathrm{op}} e_0$$

$$= \lambda^{-1} E_{N'\cap M_2}(\lambda^{-1} E_{N_1'\cap M_1}(v_q') e_0^{\mathrm{op}}) e_0 = \lambda^{-1} E_{N'\cap M_2}(v_q') e_0 = v_q',$$

implying that:

$$v_q e_0 e_0^{\mathrm{op}} v_q^* = v_q' e_0 e_0^{\mathrm{op}} {v_q'}^* = v_q' {v_q'}^* = f \le q^{\mathrm{op}} q.$$

Step I. We first prove that $L^2(\mathrm{sp} M v_q M) \simeq \mathcal{H}_k$ and that $L^2(\mathrm{sp} M^{\mathrm{op}} v_q M^{\mathrm{op}}) \simeq \bar{\mathcal{H}}_k^{\mathrm{op}}$. Indeed, since $e_0^{\mathrm{op}} = e_2$, by the definition of $\mathcal{H}_k$ we have $\mathcal{H}_k = L^2(\sum_j m_j M)$, where $\{m_j\}_j \subset M_1 = \langle M, e_1\rangle$ $(\subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}})$ are so that $\{m_j e_0^{\mathrm{op}} m_j^*\}_j$ are mutually orthogonal projections with $\sum_j m_j e^{\mathrm{op}} m_j^* = q^{\mathrm{op}} \in M' \cap M_2$. Since $v_q \in M_1$ and $v_q e_0^{\mathrm{op}} = q^{\mathrm{op}} v_q e_0^{\mathrm{op}}$, it follows that $v_q \in \sum m_j M$. Thus $M v_q M \subset \sum_j m_j M$, so that $L^2(\mathrm{sp} M v_q M) \subset \mathcal{H}_k$. Since $\mathcal{H}_k$ is irreducible and $L^2(\mathrm{sp} M v_q M)$ is a M-M bimodule, we actually have the equality $L^2(\mathrm{sp} M v_q M) = \mathcal{H}_k$.

To prove the second isomorphism, note that given any $T - T$ (resp. $M - M$) bimodule $\mathcal{H} \subset L^2(S,\tau)$, its conjugate $T - T$ (resp. $M - M$) bimodule $\bar{\mathcal{H}}$ can be identified with $(\mathcal{H})^* = \{\xi^* \mid \xi \in \mathcal{H}\}$ and its opposite $T^{\mathrm{op}} - T^{\mathrm{op}}$ (resp.

$M^{\rm op}-M^{\rm op}$) bimodule $\mathcal{H}^{\rm op}$ can be identified with $(\mathcal{H})^{\rm op} = \{\xi^{\rm op} \mid \xi \in \mathcal{H}\}$ (all this is trivial by the definitions). As a consequence, we also have $\bar{\mathcal{H}}^{\rm op} \simeq ((\mathcal{H})^*)^{\rm op}$. By taking into account that $(v_q^*)^{\rm op} = v_q$ and that $M^* = M$, from the isomorphism $L^2({\rm sp} M v_q M) \simeq \mathcal{H}_k$ and the above remark it thus follows that $L^2({\rm sp} M^{\rm op} v_q M^{\rm op}) \simeq \bar{\mathcal{H}}_k^{\rm op}$ as well.

Step II. We now prove that $\mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}}_k^{\rm op} \simeq L^2({\rm sp} M M^{\rm op} v_q M M^{\rm op})$. To see this, by Step I it is sufficient to prove that there exists $\alpha \in \mathbb{C}$ such that:

$$\langle x_1 x_2^{\rm op} v_q y_1 y_2^{\rm op}, x_3 x_4^{\rm op} v_q y_3 y_4^{\rm op}\rangle = \alpha \langle x_1 v_q y_1, x_3 v_q y_3\rangle \langle x_2^{\rm op} v_q y_2^{\rm op}, x_4^{\rm op} v_q y_4^{\rm op}\rangle,$$

$\forall\, x_i, y_j \in M$, $1 \le i, j \le 4$. By denoting $a = x_3^* x_1$, $b = y_1 y_3^*$, $c = x_2 x_4^*$, $d = y_4^* y_2$, it follows that it is sufficient to prove that

$$\langle a v_q b, c^{*{\rm op}} v_q d^{*{\rm op}}\rangle = \alpha \langle a v_q b, v_q\rangle \langle c^{\rm op} v_q d^{\rm op}, v_q\rangle,$$

$\forall\ a, b, c, d \in M$. Writing $b = b_1 e_0 b_2$, $d = d_1 e_0 d_2$ for $b_{1,2}, d_{1,2} \in N$ and using that

$$\begin{aligned}\langle a v_q b, c^{*{\rm op}} v_q d^{*{\rm op}}\rangle &= \langle b_2 a b_1 v_q e_0, d_2^{*{\rm op}} c^{*{\rm op}} d_1^{*{\rm op}} v_q e_0^{\rm op}\rangle,\\ \langle a v_q b, v_q\rangle &= \langle b_2 a b_1 v_q e_0, v_q\rangle,\\ \langle c^{\rm op} v_q d^{\rm op}, v_q\rangle &= \langle d_1^{\rm op} c^{\rm op} d_2^{\rm op} v_q e_0^{\rm op}, v_q\rangle,\end{aligned}$$

by putting a for $b_2 a b_1$ and c for $d_2 c d_1$, it follows that we only need to check that:

$$\langle a v_q e_0, c^{*{\rm op}} v_q e_0^{\rm op}\rangle = \alpha \langle a v_q e_0, v_q\rangle \langle c^{\rm op} v_q e_0^{\rm op}, v_q\rangle,$$

$\forall\ a, c \in M$. But

$$\langle a v_q e_0, c^{*{\rm op}} v_q e_0^{\rm op}\rangle = \tau(a c^{\rm op} v_q e_0 v_q^*) = \tau(a c^{\rm op} f)$$

and also

$$\begin{aligned}\langle a v_q e_0, v_q\rangle = \langle a v_q e_0, v_q e_0\rangle &= \tau(a v_q e_0 v_q^*) = \lambda^{-1} \tau(a v_q e_0 e_0^{\rm op} v_q^*)\\ &= \lambda^{-1}\tau(a f),\end{aligned}$$

and similarly $\langle c^{\rm op} v_q e_0^{\rm op}, v_q\rangle = \lambda^{-1}\tau(c^{\rm op} f)$, where $f = v_q e_0 e_0^{\rm op} v_q^*$ is the Jones projection for the irreducible inclusion $N_1 q q^{\rm op} \subset q M q q^{\rm op} \subset q^{\rm op} q M_2 q q^{\rm op}$. Since $E_{M \vee M^{\rm op}}(f) = v_k^{-2} q q^{\rm op}$ (where $\vec{v} = (v_k)_{k \in K}$ is the standard vector as usual), we have $\tau(a c^{\rm op} f) = \alpha_0 \tau(a q) \tau(c^{\rm op} q^{\rm op})$, $\forall\ a, c \in M$, for some constant $\alpha_0 \in \mathbb{R}_+$. Also, we have $\tau(a f) = \alpha_1 \tau(a q)$, $\tau(c^{\rm op} f) = \alpha_1 \tau(c^{\rm op} q^{\rm op})$, $\forall\ a, c \in M$ for some constant $\alpha_1 \in \mathbb{R}_+$. This ends the proof.

Step III. We next show that

$$L^2({\rm sp} M M^{\rm op} e_1 M M^{\rm op}) = \sum_{k \in K_1} L^2({\rm sp} M M^{\rm op} v_{q_k} M M^{\rm op}).$$

To see this let $q' \in N_1' \cap M$ be a minimal projection in the same simple direct summand as q and $u \in \mathcal{U}(N_1' \cap M)$ such that $u^* q u = q'$. Let $v'' = \lambda^{-1} E_{M_1}(u^{\mathrm{op}} v_q u^{*\mathrm{op}}) \in N' \cap M_1$ and note that $v'' e_0^{\mathrm{op}} v''^* = u^{\mathrm{op}}(v_q e_0^{\mathrm{op}} v_q^*) u^{*\mathrm{op}} \leq q'^{\mathrm{op}}$. By the same reasoning as in Step I, it follows that $L^2(\mathrm{sp} M v'' M) = L^2(\sum_j m_j' M)$, where $\{m_j'\}_j \subset M_1$ is an orthonormal system such that $\sum m_j' e_0^{\mathrm{op}} m_j'^* = q'^{\mathrm{op}}$. But $v'' \in \mathrm{sp} M M^{\mathrm{op}} v_q M^{\mathrm{op}} M$, because $u^{\mathrm{op}} v_q u^{*\mathrm{op}} \in \mathrm{sp} M^{\mathrm{op}} v_q M^{\mathrm{op}}$ and $E_{M_1}(u^{\mathrm{op}} v_q u^{*\mathrm{op}}) = \lambda \sum_j b_j^{\mathrm{op}} (u^{\mathrm{op}} v_q u^{*\mathrm{op}}) b_j^{*\mathrm{op}} \in \mathrm{sp} M^{\mathrm{op}} v_q M^{\mathrm{op}}$ as well, where $\{b_j^*\}_j$ is an orthonormal basis of N over N_1.
Thus we have $\mathrm{sp} M M^{\mathrm{op}} v_q M^{\mathrm{op}} M \supset \mathrm{sp} M M^{\mathrm{op}} v_{q'} M M^{\mathrm{op}}$, $\forall\ q'$ chosen this way. Thus, if $\{m_j^k\}_j \subset M_1$ is a orthonormal system such that $\sum_j m_j^k e_0^{\mathrm{op}} m_j^k$ is the central support of q^{op} in $M' \cap M_2$ then $\mathrm{sp} M^{\mathrm{op}}(\sum_j m_j^k M) M^{\mathrm{op}} = \mathrm{sp} M M^{\mathrm{op}} v_q M^{\mathrm{op}} M$. Summing up over k and using that $\sum_k \sum_j m_j^k M = M_1 = \mathrm{sp} M e_1 M$, the statement follows.

Step IV. We now derive that

$$L^2(\mathrm{sp} M M^{\mathrm{op}} f_{-n}^n M M^{\mathrm{op}}) \simeq \bigoplus_{k \in K_n} \mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}}_k^{\mathrm{op}}$$

and then

$$L^2\left(M \underset{e_N}{\boxtimes} M^{\mathrm{op}}\right) = \bigoplus_{k \in K} \mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}}_k^{\mathrm{op}}.$$

To see this, note first that $\mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}}_k^{\mathrm{op}} \simeq \mathcal{H}_{k'} \bar{\otimes} \bar{\mathcal{H}}_{k'}^{\mathrm{op}}$ if and only if $\mathcal{H}_k \simeq \mathcal{H}_k'$. This fact follows immediately by interpreting $\mathcal{H}_k$ as irreducible representation of $M \otimes M^{\mathrm{op}}$, according to Connes' alternative view on correspondences (see [C4], [Po8]).
Since by Steps II and III we have $\vee_{k \in K_1} \mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}}_k^{\mathrm{op}} = L^2(\mathrm{sp} M M^{\mathrm{op}} e_1 M M^{\mathrm{op}})$, with $\mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}}_k^{\mathrm{op}}$ mutually nonisomorphic, the first part of the statement follows for $n = 1$. By using this fact for $N_{n-1} \subset M \overset{f_{-n}^n}{\subset} M_n$, $n \geq 1$, we get it for any $n \geq 1$. The last part is now clear, since $\cup_n \mathrm{sp} M M^{\mathrm{op}} f_{-n}^n M M^{\mathrm{op}}$ is dense in $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$.

Step V. We finally show that if s_k denotes the minimal projection in $T' \cap \langle S, T\rangle$ labeled by $k \in K$ then $\mathrm{Tr}\, s_k = v_k^2$. This fact can be checked directly by using a similar strategy as in Step III. Instead, we will use the following more elegant argument: Since

$$v_k^4 = (\mathrm{Tr}_{\langle S,T\rangle} s_k)(\mathrm{Tr}_{T'} s_k) = (\mathrm{Tr}_{\langle S,T\rangle} s_k)(\mathrm{Tr}_{\langle S,T\rangle} J_S s_k J_S)$$

(cf. [J1]), we only need to show that $T' \cap \langle S, T\rangle \ni s_k \mapsto J_S s_k J_S \in T' \cap \langle S, T\rangle$ is $\mathrm{Tr}_{\langle S,T\rangle}$-preserving.
To see this note that since ${}^{\mathrm{op}}$ acts on S leaving T invariant, it implements a $\mathrm{Tr}_{\langle S,T\rangle}$-preserving anti-automorphism on $\langle S, T\rangle$, thus a $\mathrm{Tr}_{\langle S,T\rangle}$-preserving automorphism on the commutative algebra $T' \cap \langle S, T\rangle$. Moreover, if we put

$s_k L^2(S,\tau) = L^2(\text{sp}Tv_qT)$ as in Steps I and II and use that $v_q^{\text{op}} = v_q^*$, then we have

$$s_k^{\text{op}} L^2(S,\tau) = (L^2(\text{sp}Tv_qT))^{\text{op}}$$

$$= L^2(\text{sp}T^{\text{op}}v_q^{\text{op}}T^{\text{op}}) = L^2(\text{sp}Tv_q^{\text{op}}T) = L^2(\text{sp}Tv_q^*T)$$

$$= L^2(\text{sp}T^*v_q^*T^*) = L^2(\text{sp}Tv_qT)^* = J_S s_k J_S L^2(S,\tau)$$

Thus, $s_k^{\text{op}} = J_S s_k J_S$ so that $\text{Tr}(s_k) = \text{Tr}(s_k^{\text{op}}) = \text{Tr}(J_S s_k J_S)$. Q.E.D.

Note that the above theorem agrees with the exemples in Section 3. Indeed, if $\sigma \in \text{Aut}(P)$ is an automorphism of a type II_1 factor P and $\mathcal{H}_\sigma = L^2(\sigma)$ denotes the $P-P$ bimodule associated with σ as in [Po8] then an easy calculation shows that $\bar{\mathcal{H}_\sigma}^{\text{op}} = \mathcal{H}_{\sigma^{\text{op}}}$.

4.6. COROLLARY. *Let $N \subset M$ be an extremal inclusion. Then $N \subset M$ has finite depth if and only if $[M \underset{e_N}{\boxtimes} M^{\text{op}} : M \vee M^{\text{op}}] < \infty$. Moreover if these conditions are satisfied then $M \vee M^{\text{op}}$ has finite depth in $M \underset{e_N}{\boxtimes} M^{\text{op}}$.*

Proof. With the notations used in 4.5 and its proof, if we assume that $N \subset M$ has finite depth then K is finite so that by 4.5 we have $\dim(S' \cap \langle S, T\rangle) < \infty$ and each of the local indices is finite. But then, by Jones' formula ([J1]), it follows that $[S : T] < \infty$.
Conversely, if $[S : T] < \infty$ then $\dim(S' \cap \langle S, T\rangle) < \infty$, so that K follows finite, i.e., $N \subset M$ has finite depth.
Moreover, we see from 4.5 that if $[S : T] < \infty$ then the set of all $T-T$ irreducible bimodules generated by $L^2(S,\tau)$ under Connes' tensor product (fusion) are contained in the set of bimodules $\{\mathcal{H}_k \bar{\otimes} \bar{\mathcal{H}_{k'}}^{\text{op}}\}_{k,k' \in K}$ and is thus finite, i.e., $T \subset S$ has finite depth. Q.E.D.

4.7. *Remark.* As mentioned before, if M is hyperfinite and $N \subset M$ is a subfactor of finite depth then by ([Po15]) we have $M \vee M^{\text{op}} \subset M \underset{e_N}{\boxtimes} M^{\text{op}}$ is isomorphic to the inclusion $M \vee M' \cap M_\infty \subset M_\infty$ of [Oc]. This latter inclusion was already shown to have finite depth in [Oc] and in fact all its standard invariant (paragroup) has been calculated ([Oc], see also [EvKa]). In particular, for this class of symmetric enveloping inclusions, part b) of 4.5 can be recovered from ([Oc]). If $N \subset M$ is a finite depth subfactor with M not necessarily hyperfinite, then it is imediate to see that $M \vee M^{\text{op}} \subset M \underset{e_N}{\boxtimes} M^{\text{op}}$ has the same standard invariant (paragroup) as $P \vee P^{\text{op}} \subset P \underset{e_Q}{\boxtimes} P^{\text{op}}$ where $Q \subset P$ denotes the standard model for $N \subset M$, which is thus an inclusion of hyperfinite factors. Thus, for any $N \subset M$ with finite depth the standard invariant (paragroup) of $M \vee M^{\text{op}} \subset M \underset{e_N}{\boxtimes} M^{\text{op}}$ can be recovered from these results.
Theorem 4.5 shows that in the case $(T \subset S) = (M \vee M^{\text{op}} \subset M \underset{e_N}{\boxtimes} M^{\text{op}})$ then $L^2(S,\tau)$ is spanned by $T-T$ bimodules which are finitely generated both as left

and right T-modules. Equivalently, $T \subset S$ is such that $T' \cap \langle S, T\rangle$ is generated by finite projections of $\langle S, T\rangle$. Inclusions $T \subset S$ verifying this latter condition are called *discrete* in [ILP]. We'll introduce here a new terminology for such subfactors based on the former, more intrinsic characterization.

4.8. Definition. Let $T \subset S$ be an irreducible inclusion of type II_1 factors. We denote by $q\mathcal{N}_S(T) \overset{\text{def}}{=} \{x \in S \mid \exists\, x_1, x_2, \dots, x_n \in S$ such that $xT \subset \sum_{i=1}^n Tx_i$ and $Tx \subset \sum_{i=1}^n x_i T\}$. We call $q\mathcal{N}_S(T)$ the *quasi-normalizer* of T in S.

Note that the condition "$xT \subset \sum Tx_i$, $Tx \subset \sum x_i T$" is equivalent to "$TxT \subset (\sum_{i=1}^n Tx_i) \cap (\sum_{i=1}^n x_i T)$" and also to "$\mathrm{sp}TxT$ is finitely generated both as left and as a right T-module." It then follows readily that $\mathrm{sp}(q\mathcal{N}_S(T))$ is a $*$-algebra. Thus $P \overset{\text{def}}{=} \overline{\mathrm{sp}}(q\mathcal{N}_S(T)) = q\mathcal{N}_S(T)''$ is a subfactor of S containing T. Note also that $L^2(P) = \vee\{\mathcal{H} \mid \mathcal{H} \subset L^2(P),\ \mathcal{H}$ is a T-T bimodule, $\dim({}_T\mathcal{H}) < \infty$, $\dim(\mathcal{H}_T) < \infty\}$ and that the orthogonal projection e_P, of $L^2(S)$ onto $L^2(P)$, satisfies $e_P = \vee\{f \in T' \cap \langle S, T\rangle \mid \mathrm{Tr} f < \infty,\ \mathrm{Tr} J_S f J_S < \infty\}$. All these facts are just reformulations of some results in [PiPo1] and [ILP], but can also be proved as exercises.
The terminology we wanted to introduce is then as follows:

4.9. Definition. Let $T \subset S$ be an irreducible inclusion. If $q\mathcal{N}_S(T)'' = S$, we say that T is *quasi-regular* in S. From the above remarks we see that an irreducible inclusion $T \subset S$ is discrete (as defined in [ILP]) iff T is quasi-regular in S.

Thus, from 4.5 it follows that if $N \subset M$ is an extremal inclusion of type II_1 factors then $M \vee M^{\mathrm{op}}$ is quasi-regular in $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$. Note that, even more, we showed that each irreducible T-T bimodule in $L^2(S)$ (where $T = M \vee M^{\mathrm{op}}$, $S = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$) has multiplicity 1 and its (finite) dimension as a left T module coincides with its dimension as a right T-module. Thus, our symmetric enveloping inclusions have very similar properties to the inclusions given by cross-products of factors by outer actions of discrete groups.
We wanted to emphasize even more this aspect by choosing the terminology "quasi-normalizer", "quasi-regular" in analogy with Dixmier's notions of "normalizer" and "regularity" for an irreducible subfactor ([D1]). This is particularily justified by noticing that exemples of quasi-regular subfactors $T \subset S$ can be obtained by requiring S to be generated by unitary elements u such that uTu^* is included in T and has finite index in it (see the Appendix in [ILP] for a concrete exemple of such a situation).
Let us end this section with a result showing that the extended sequence of Jones projections in a tunnel-tower associated to a subfactor $N \subset M$ has a certain general ergodicity property with respect to the higher relative commutants that is very useful in applications (see e.g. 2.2 and 2.3 in [GePo]). We'll refer to this result as the *Ergodicity Theorem for Higher Relative Commutants*.

4.10. Theorem. *Let $N \subset M$ be a subfactor with finite index (but not necessarily extremal). Let $\{M_j\}_{j\in\mathbb{Z}}$ be a tunnel-tower for $N \subset M$, where*

$M_0 = M, M_{-1} = N$, and $\{e_j\}_{j\in\mathbb{Z}}$ be its corresponding Jones projections. Denote $A_{ij} = M_i' \cap M_j$ and $A_{-\infty,i} = \overline{\bigcup_{n\leq i} A_{ni}}$, $A_{-\infty,\infty} = \overline{\bigcup_i A_{-\infty,i}}$. Then we have:
a) $\{e_j\}'_{j\in\mathbb{Z}} \cap A_{-\infty,\infty} = \mathbb{C}$. *In particular, $A_{-\infty,\infty}$ is a factor.*
b) *If M has separable predual then the tunnel $\{M_j\}_{j\leq 0}$ can be chosen such that $\{e_j\}'_{j\leq k} \cap M_n \subset A_{-\infty,n}, \forall k \leq n$ in $\mathbb{Z}$.*
c) *If $N \subset M$ is extremal and its tunnel is chosen to satisfy condition* b) *then $\{e_j\}'_{j\in\mathbb{Z}} \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = \mathbb{C}$.*

Proof. a). Let θ be the trace preserving automorphism on $A_{-\infty,\infty}$ implemented by the duality isomorphism (1.5 of [PiPo1] or 1.3.3 of [Po2]), i.e., θ satisfies $\theta(A_{ij}) = A_{i+2,j+2}, \theta(e_k) = e_{k+2}, \forall i,j,k \in \mathbb{Z}$, with $\theta_{|A_{ij}}$ being defined as the restriction to $M_i' \cap M_j = M_i^{\alpha\prime} \cap M_j^\alpha$ of $\sigma'_{ij} : (M_i \subset M_{i+1} \subset ... \subset M_j)^\alpha \to (M_{i+2} \subset ... \subset M_{j+2})$, where $\sigma'_{ij}((x_{rs})_{r,s}) = \lambda^{i-j+1} \sum_{r,s} m_r e_{i+2} e_{i+3} ... e_{j+2} x_{rs} e_{j+2} ... e_{i+2} m_s^*$, in which $\{m_r\}_r$ is an orthonormal basis of $\mathrm{vN}\{e_n\}_{n\leq i+1}$ over $\mathrm{vN}\{e_n\}_{n\leq i}$ and $\lambda^{-1} = [M : N]$.
We first show that this automorphism satisfies the identity $\theta(z)e_{j+2} = ze_{j+2}$ for all $z \in \{e_k\}'_{k\leq j} \cap A_{-\infty,j}$. To this end let $\varepsilon > 0$ and $i \leq j$ be so that $\|E_{A_{ij}}(z) - z\|_2 < \varepsilon$. Put $z_0 = E_{A_{ij}}(z) \in \{e_{i+2}, ..., e_j\}' \cap A_{ij}$. From the above local formula for θ we have

$$e_{j+2}\theta(z_0)e_{j+2} = \lambda^{i-j+1} e_{j+2} (\sum_r m_r e_{i+2} ... e_{j+2} z_0 e_{j+2} ... e_{i+2} m_r^*) e_{j+2}$$

$$= \lambda^{i-j+3} \sum_r m_r e_{i+2} ... e_j (z_0 e_{j+2}) e_j ... e_{i+2} m_r^*$$

$$= \sum_r m_r (z_0 e_{i+2} e_{j+2}) m_r^* = (\sum_r m_r (z_0 e_{i+2}) m_r^*) e_{j+2}.$$

By taking into account that the orthonormal basis $\{m_r\}_r$ can be taken to be made up of no more than $[M : N] + 1$ elements, we thus get the estimates:

$$\|\theta(z)e_{j+2} - ze_{j+2}\|_2$$

$$\leq \|\theta(z) - \theta(z_0)\|_2 + \|z - z_0\|_2 + \|e_{J+2}\theta(z_0)e_{j+2} - z_0 e_{j+2}\|_2$$

$$\leq 2\varepsilon + \|\Sigma_r m_r (z_0 e_{i+2}) m_r^* e_{j+2} - z_0 e_{j+2}\|_2$$

$$\leq 2\varepsilon + \sum_r \|[m_r e_{i+2}, z_0]\|_2$$

$$\leq 2\varepsilon + ([M : N] + 1)^2 \varepsilon.$$

Letting ε tend to 0, we get the desired identity.
Now to prove part a) of the statement let $z \in \mathrm{vN}\{e_n\}'_{n\in\mathbb{Z}} \cap A_{-\infty,\infty}$ with $\tau(z) = 0$ and take $z_0 = E_{A_{-\infty,j}}(z)$ for some j. Note that $\tau(z_0) = 0$ as well. For such a z_0, and in fact for any z_0 in $\mathrm{vN}\{e_n\}'_{n\leq j} \cap A_{-\infty,j}$, we then have the estimates:

$$\|(z - z_0)e_{j+2}\|_2^2 = \tau((z - z_0)^*(z - z_0)e_{j+2})$$

$$= \tau(E_{\{e_l\}'_{l\geq j+2}\cap A_{-\infty,\infty}}((z-z_0)^*(z-z_0)e_{j+2}))$$

$$= \tau((z-z_0)^*(z-z_0)E_{\{e_l\}'_{l\geq j+2}\cap A_{-\infty,\infty}}(e_{j+2})) = \lambda\|z-z_0\|_2^2,$$

in which we used that by Jones ergodicity theorem we have $E_{\{e_l\}'_{l\geq j+2}\cap A_{-\infty,\infty}}(e_{j+2}) = \lambda 1$.
Since $z_0 e_{j+2} = \theta(z_0)e_{j+2}$, we get similarily:

$$\|(z-z_0)e_{j+2}\|_2^2 = \|(z-\theta(z_0))e_{j+2}\|_2^2$$

$$= \tau((z-\theta(z_0)^*(z-\theta(z_0))e_{j+2})$$

$$= \tau(E_{\{e_l\}'_{l\leq j+2}\cap A_{-\infty,\infty}}((z-\theta(z_0))^*(z-\theta(z_0))e_{j+2}))$$

$$= \tau((z-\theta(z_0))^*(z-\theta(z_0))E_{\{e_l\}'_{l\leq j+2}\cap A_{-\infty,\infty}}(e_{j+2}))$$

$$= \lambda\|z-\theta(z_0)\|_2^2,$$

in which we used the fact that $\theta(z_0)$ commutes with $\mathrm{vN}\{e_l\}_{l\leq j+2}$ and that by Jones ergodicity theorem we have $E_{\{e_l\}'_{l\leq j+2}\cap A_{-\infty,\infty}}(e_{j+2}) = \lambda 1$.
Altogether, the above shows that $\|z-z_0\|_2 = \|z-\theta(z_0)\|_2$ and by applying this recursively n times we get $\|z-z_0\|_2 = \|z-\theta^n(z_0)\|_2, \forall n \geq 1$.
On the other hand $\theta^n(A_{ij}) = A_{i+2n,j+2n}$ and so, if n is so that $2n > j-i$ then $\tau(z_1\theta^n(z_1)) = \tau(z_1)^2, \forall z_1 \in A_{ij}$, showing that θ is mixing on $A_{-\infty,\infty} = \overline{\bigcup_{i,j} A_{ij}}$. Thus, for $z_0 \in A_{-\infty,j}$ with $\tau(z_0) = 0$ we have

$$\lim_{n\to\infty}\|z_0-\theta^n(z_0)\|_2^2 = 2\|z_0\|_2^2.$$

Since $\|z_0-\theta^n(z_0)\|_2 \leq \|z_0-z\|_2 + \|z-\theta^n(z_0)\|_2$ and since for $z \in \mathrm{vN}\{e_n\}'_{n\in\mathbb{Z}} \cap A_{-\infty,\infty}$ we proved that $\|z-z_0\|_2 = \|z-\theta^n(z_0)\|_2, \forall n \geq 1$, in which $z_0 = E_{-\infty,j}(z)$, it follows that for each j we have the estimate:

$$2\|z_0\|_2^2 = \lim_{n\to\infty}\|z_0-\theta^n(z_0)\|_2^2 \leq 4\|z-z_0\|_2^2.$$

Now, letting j tend to infinity we get $\|z-z_0\|_2$ tend to 0 and $\|z_0\|_2$ tend to $\|z\|_2$, which from the above estimate forces $z=0$. This ends the proof of a).
b). Let $\{x_n\}_{n\geq 1} \subset M$ be a sequence of elements dense in the unit ball of M in the so-topology. We construct recursively a sequence of integers $0 < k_1 < k_2 < \ldots$ and a tunnel $M \supset N \supset N_1 \ldots \supset N_{k_1} \supset \ldots \supset N_{k_n} \supset \ldots$ for $N \subset M$ such that if $\{e_n\}_{n\leq 0}$ are the corresponding Jones projections and we denote by $B_n = \mathrm{Alg}\{e_j\}_{-k_n+1\leq j\leq -k_{n-1}-1}$ then we have:

$$\|E_{B_n'\cap M}(x_j) - E_{N'_{k_{n-1}}\cap M}(x_j)\|_2 < 2^{-n}, \forall j \leq n$$

Assume we have this up to some n. By ([Po1]) there exists a hyperfinite subfactor $R \subset N_{k_n}$ such that $E_{R'\cap M}(x) = E_{N'_{k_n}\cap M}(x), \forall x \in M$. On the other hand, by Jones ergodicity theorem, we can regard R as being generated by a sequence of Jones λ-projections e_j indexed over the integers $\leq -k_n - 1$. Thus, there will exist a sufficiently large k_{n+1} such that if we denote $B_{n+1} = \mathrm{Alg}\{e_j\}_{-k_{n+1}+1 \leq j \leq -k_n - 1}$ then

$$\|E_{B'_{n+1}\cap M}(x_j) - E_{N'_{k_n}\cap M}(x_j)\|_2 < 2^{-n-1}, \forall j \leq n+1.$$

Now choose a Jones projection e_{-k_n} for $N_{k_n} \subset N_{k_n-1}$ such that it commutes with $e_j \in B_{n+1}$ for $j \leq -k_n - 2$ and such that it satisfies the Jones-Temperley-Lieb relation for $j = -k_n - 1$ (see the proof of 4.4 on page 33 of [Po15]), i.e., $e_{-k_n} e_{-k_n-1} e_{-k_n} = \lambda e_{-k_n}$, and then simply define the corresponding tunnel $N_{k_n} \supset N_{k_n+1} \supset \ldots \supset N_{k_{n+1}}$ as given by these newly chosen Jones projections e_j with $-k_{n+1} + 1 \leq j \leq -k_n$.
Thus, if we take $A_n = \overline{\cup_{m\geq n} B_m} \subset \mathrm{vN}\{e_j\}_{j\leq -k_n-1}$ then it follows from the above that $E_{A'_n\cap M}(x) \in \overline{\cup_k N'_k \cap M}$ for all $x \in \{x_j\}_j$ and thus by density for all $x \in M$. Thus even more so $\mathrm{vN}\{e_j\}'_{j\leq -m} \cap M \subset \overline{\cup_k N'_k \cap M}$ for $m = k_n$ and thus in fact for all $m \geq 0$.
Finally, if $x \in M_n$ for some $n \geq 0$ then for any $\varepsilon > 0$ there exists $k \leq 0$ and $x' \in \mathrm{sp}((\mathrm{Alg}\{e_j\}_{k\leq j\leq n})M)$ such that $\|x - x'\|_2 < \varepsilon$. But then $x'' = E_{\{e_l\}'_{l\leq k-2}\cap M_n}(x')$ belongs to $\mathrm{sp}((\mathrm{Alg}(\{e_j\}_{k\leq j\leq n})\overline{\cup_i N'_i \cap M})$ which in turn is icluded into $\overline{\cup_i N'_i \cap M_n}$ and we have:

$$\|E_{\{e_l\}'_{l\leq k-2}\cap M_n}(x) - x''\|_2 = \|E_{\{e_l\}'_{l\leq k-2}\cap M_n}(x - x')\|_2 \leq \|x - x'\|_2 \leq \varepsilon.$$

Letting ε go to 0 and j to $-\infty$ we get

$$\lim_{j\to-\infty} E_{\{e_l\}'_{l\leq j-2}\cap M_n}(x) = E_{A_{-\infty,n}}(x), \forall x \in M_n$$

This ends the proof of b) and c) follows then imediately, by taking into account that $\bigcup_n \mathrm{sp} M M_n^{\mathrm{op}} M$ is dense in $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ and applying a) and b). Q.E.D.

4.11. Corollary. *Let $N \subset M$ be an extremal inclusion of type II_1 factors with separable preduals. There exists a choice of a tunnel $\{M_j\}_{j\leq 0}$ for $N \subset M$ such that if we denote $M_n = (M_{-n})^{\mathrm{op}\prime} \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$, $n \geq 1$, $M_\infty = \overline{\cup_n M_n}$, $M_\infty^{\mathrm{op}} = \overline{\cup_n M_n^{\mathrm{op}}}$ and $A_{-\infty,\infty} = \overline{\cup_n M'_{-n} \cap M_n}$ then $M_\infty, M_\infty^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ satisfy the conditions:*
a) $\overline{\mathrm{sp}} M_\infty M_\infty^{\mathrm{op}} = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$.
b) $M_\infty \cap M_\infty^{\mathrm{op}} = A_{-\infty,\infty}$ *and* $E_{M_\infty} E_{M_\infty^{\mathrm{op}}} = E_{A_{-\infty,\infty}}$.
c) $A'_{-\infty,\infty} \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = \mathbb{C}1$.

Proof.. Conditions a) and b) are actually valid for any choice of the tunnel while 4.9 clearly implies c). Q.E.D.

5. Relating the Amenability Properties of $\Gamma_{N,M}$, $\mathcal{G}_{N,M}$ and $(M \vee M^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}})$

In [Po8] one considers a notion of relative amenability for inclusions of finite von Neumann algebras $T \subset S$ by requiring the existence of norm one projections from $\langle S, T\rangle$ onto S, equivalently of Connes-type S-hypertraces on $\langle S, T\rangle$. In the case $S = T \rtimes G$ for some discrete group G this condition on the inclusion $T \subset S$ is equivalent to the amenability of the group G.

As we have seen in the previous section, when $T = M \vee M^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = S$, for $N \subset M$ a locally trivial subfactor associated to some faithful G-kernel σ, with G a finitely generated discrete group, then $(T \subset S) \simeq (T \subset T \rtimes_{\sigma \otimes \sigma^{\mathrm{op}}} G)$. Thus, the relative amenability of the inclusion $T \subset S$ is equivalent in this case to the amenability of G. On the other hand, one of the equivalent characterizations of the amenability of G is Kesten's condition requiring that the Cayley graph of G, Γ, corresponding to some finite, self-adjoint set of generators $g_0 = 1, g_1, \dots, g_n$, satisfies $\|\Gamma\| = n+1$.

Recalling from [Po2,5] that the *standard graph* of a subfactor $\Gamma_{N,M}$ is called *amenable* if it satisfies the Kesten-type condition $\|\Gamma_{N,M}\|^2 = [M : N]$ and that its *standard invariant* $\mathcal{G}_{N,M}$ is called *amenable* if $\Gamma_{N,M}$ is amenable, and noticing that for the locally trivial subfactor $N \subset M$ corresponding to the above $(G; g_0, \dots, g_n; \sigma)$ the Cayley graph Γ coincides with the standard graph $\Gamma_{N,M}$, while $[M : N] = (n+1)^2$, it follows that in this case the amenability of G (thus, the relative amenability of $T \subset S$) is equivalent to the amenability of $\mathcal{G}_{N,M}$.

We prove in this section that in fact even for arbitrary extremal subfactors of finite index $N \subset M$ the relative amenability condition on $T = M \vee M^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = S$ is equivalent to the amenability of the standard lattice $\mathcal{G}_{N,M}$. Along the lines, we will obtain some other related characterizations of the amenability of $\mathcal{G}_{N,M}$, thus of $\Gamma_{N,M}$.

Before stating the result, recall some terminology and notations from [Po2].

So let $(\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}) = \bigoplus((N \otimes P^{\mathrm{op}})^{**} \overset{(E \otimes \mathrm{id})^{**}}{\subset} (M \otimes P^{\mathrm{op}})^{**})$, the sum being taken over all isomorphism classes of type II_1 factors P that can be embedded with finite index in some amplification of M, i.e., factors P that are weakly stably equivalent to M in the sense of 1.4.3 in [Po8] (like for instance $P = M$). Then take first the atomic part of this inclusion, $(\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M})_{\mathrm{at}}$, and next the binormal part of the latter inclusion, $((\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M})_{\mathrm{at}})_{\mathrm{bin}}$ (i.e., the largest direct summand in which both M and P^{op} sit as von Neumann algebras), which we denote by $\mathcal{N}^u \overset{\mathcal{E}^u}{\subset} \mathcal{M}^u$, and call *the universal atomic (binormal) representation* of $N \subset M$. Also, the inclusion graph (or matrix) of $\mathcal{N}^u \subset \mathcal{M}^u$ is denoted by $\Gamma^u_{N,M}$ and called the *universal graph (or matrix)* of $N \subset M$.

Finally one defines $(\mathcal{N}^{\mathrm{st}} \overset{\mathcal{E}^{\mathrm{st}}}{\subset} \mathcal{M}^{\mathrm{st}})$ to be the minimal direct summand of $\mathcal{N}^u \subset$

$\mathcal{M}^u$ (or, equivalently, of $(\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M})_{\rm at}$) containing the standard representation of $M \otimes M^{\rm op}$, $\mathcal{B}(L^2(M))$ and call it *the standard representation* of $N \subset M$. It is easy to see (cf. e.g., [Po2]) that the commuting square embedding

$$\begin{array}{ccc} \mathcal{N}^{\rm st} & \overset{\mathcal{E}^{\rm st}}{\subset} & \mathcal{M}^{\rm st} \\ \cup & & \cup \\ N \otimes M^{\rm op} & \overset{E\otimes {\rm id}}{\subset} & M \otimes M^{\rm op} \end{array}$$

can be identified with the embedding

$$\begin{array}{ccc} \bigoplus\limits_{\ell \in L} \mathcal{B}(\mathcal{K}_\ell) & \overset{\mathcal{E}^{\rm st}}{\subset} & \bigoplus\limits_{k \in K} \mathcal{B}(\mathcal{H}_k) \\ \cup & & \cup \\ N \otimes M^{\rm op} & \overset{E\otimes {\rm id}}{\subset} & M \otimes M^{\rm op} \end{array}$$

in which $\{\mathcal{H}_k\}_{k\in K}$ (respectively, $\{\mathcal{K}_\ell\}_{\ell \in L}$) is the list of all irreducible M-M (resp. N-M) bimodules appearing as direct summands in $L^2(M_j)$, $j = 0, 1, 2, \dots$, and $M \otimes M^{\rm op}$ (resp, $N \otimes M^{\rm op}$) is represented on each $\mathcal{H}_k$ (resp. $\mathcal{K}_\ell$) by operators of left and right multiplication by elements in M (respectively, right multiplication by elements in N and right multiplication by elements in M). Moreover, the inclusion matrix (or graph) for

$$\mathcal{N}^{\rm st} \subset \bigoplus_{\ell \in L} \mathcal{B}(\mathcal{K}_\ell) \subset \bigoplus_{k \in K} \mathcal{B}(\mathcal{H}_k) = \mathcal{M}^{\rm st}$$

(which is thus a direct summand of the universal graph $\Gamma^u_{N,M}$) is given by $(\Gamma_{N,M})^{\rm t}$, while $\mathcal{E}^{\rm st}$ is the unique expectation that preserves the trace Tr on $\mathcal{M}^{\rm st} = \bigoplus\limits_{k \in K} \mathcal{B}(\mathcal{H}_k)$ given by the weight vector $\vec{v} = (v_k)_{k \in K}$, with $v_k = \dim({}_M\mathcal{H}_{k\,M})^{1/2}$.
Finally, note that $N \subset M$ is in fact embedded in the smaller inclusion

$$\mathcal{N}^{\rm st,f} \overset{\rm def}{:=} (1 \otimes M^{\rm op})' \cap \mathcal{N}^{\rm st} \overset{\mathcal{E}^{\rm st,f}}{\subset} (1 \otimes M^{\rm op})' \cap \mathcal{M}^{\rm st} \overset{\rm def}{=:} \mathcal{M}^{\rm st,f}$$

where $\mathcal{E}^{\rm st,f}$ is the restriction of $\mathcal{E}^{\rm st}$ to $\mathcal{M}^{\rm st,f}$.

5.1. Definition. The commuting square embedding:

$$\begin{array}{ccc} \mathcal{N}^{{\rm st},f} & \overset{\mathcal{E}^{\rm st,f}}{\subset} & \mathcal{M}^{\rm st,f} \\ \cup & & \cup \\ N & \overset{E_N}{\subset} & M \end{array}$$

is called the *finite (or reduced) standard representation* of $N \subset M$.

5.2. Lemma. *$\mathcal{N}^{\mathrm{st,f}}$, $\mathcal{M}^{\mathrm{st,f}}$ are finite type II_1 von Neumann algebras with atomic centers $\mathcal{Z}(\mathcal{N}^{\mathrm{st,f}}) = \mathcal{Z}(\mathcal{N}^{\mathrm{st}}) \simeq \ell^\infty(L)$, $\mathcal{Z}(\mathcal{M}^{\mathrm{st,f}}) = \mathcal{Z}(\mathcal{M}^{\mathrm{st}}) \simeq \ell^\infty(K)$. Moreover, the inclusion $\mathcal{N}^{\mathrm{st,f}} \subset \mathcal{M}^{\mathrm{st,f}}$ is a matricial inclusion having inclusion matrix (or graph) $(\Gamma_{N,M})^{\mathrm{t}}$. Also, $\mathcal{E}^{\mathrm{st,f}}$ is the unique conditional expectation of $\mathcal{M}^{\mathrm{st,f}}$ onto $\mathcal{N}^{\mathrm{st,f}}$ preserving the trace Tr on $\mathcal{M}^{\mathrm{st,f}}$ given by the weights $\{v_k^2\}_{k\in K}$ on the center $\mathcal{Z}(\mathcal{M}^{\mathrm{st,f}}) \simeq \ell^\infty(K)$.*

Proof. The first part is trivial, by the definition of $\mathcal{N}^{\mathrm{st,f}} \subset \mathcal{M}^{\mathrm{st,f}}$ and the properties of $\mathcal{N}^{\mathrm{st}} \subset \mathcal{M}^{\mathrm{st}}$. Then the last part is an immediate consequence of the first part and of 2.7 in [PiPo2]. Q.E.D.

5.3. Theorem. *Let $N \subset M$ be an extremal inclusion of type II_1 factors. The following conditions are equivalent:*
1) $\mathcal{G}_{N,M}$ is amenable.
1′) $\Gamma_{N,M}$ is amenable, i.e., $\Gamma_{N,M}$ satisfies the Kesten type condition $\|\Gamma_{N,M}\|^2 = [M : N]$.
2) $(\Gamma_{N,M}, \vec{v})$ satisfies the Følner-type condition: $\forall\ \varepsilon > 0$, $\exists F \subset K$ finite such that

$$\sum_{k\in\partial F} v_k^2 < \varepsilon \sum_{k\in F} v_k^2,$$

where

$$\partial F = \{k \in K \setminus F \mid \exists\ k_0 \in F \text{ such that } (\Gamma_{N,M}\Gamma_{N,M}^{\mathrm{t}})_{kk_0} \neq 0\}.$$

3) There exists a state ψ_0 on $\ell^\infty(K) \simeq T' \cap \langle S, T\rangle$ such that $\psi_0 \circ E$ has $S = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ in its centralizer, where E is the unique Tr-preserving conditional expectation of $\langle S, T\rangle$ onto $T' \cap \langle S, T\rangle$.
4) $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ is amenable relative to $M \vee M^{\mathrm{op}}$.
5) There exists a norm one projection from $(\mathcal{N}^{\mathrm{st},f} \overset{\mathcal{E}^{\mathrm{st,f}}}{\subset} \mathcal{M}^{\mathrm{st,f}})$ onto $(N \overset{E_N}{\subset} M)$.
5′) There exists a $(N \subset M)$-hypertrace on $(\mathcal{N}^{\mathrm{st,f}} \overset{\mathcal{E}^{\mathrm{st,f}}}{\subset} \mathcal{M}^{\mathrm{st,f}})$.

Proof. 1) $\Longleftrightarrow$ 1′) is clear by the definitions.
To prove 1′) $\Longrightarrow$ 2) let $\Phi = \lambda V^{-1}\Gamma\Gamma^{\mathrm{t}}V$, where V is the diagonal matrix over K with entries $(v_k)_{k\in K}$. Note that Φ defines a bounded positive linear operator from $P \overset{\mathrm{def}}{=} T' \cap \langle S, T\rangle \simeq \ell^\infty(K)$ into itself such that $\Phi(1) = 1$. Note also that the trace Tr on P inherited from $\langle S, T\rangle$ has weights $(v_k^2)_{k\in K}$ as a measure on K, i.e., if $b \in P \simeq \ell^\infty(K)$ then

$$\|b\|_{1,\mathrm{Tr}} = \sum_{k\in K} |b_k| v_k^2.$$

For $a, b : K \to \mathbb{C}$, at least one of which has finite support, we denote $\langle a, b\rangle = \sum_{k\in K} a_k \bar{b}_k$. For each $b \in P \simeq \ell^\infty(K)$ with finite support we then have:

$$\begin{aligned}\mathrm{Tr}(\Phi(b)) &= \langle \Phi(b), V^2(1)\rangle = \langle b, \lambda V\Gamma\Gamma^{\mathrm{t}}V^{-1}V^2(1)\rangle \\ &= \langle b, \lambda V\Gamma\Gamma^{\mathrm{t}}V(1)\rangle = \langle b, V^2(1)\rangle = \mathrm{Tr}(b).\end{aligned}$$

Thus $\mathrm{Tr} \circ \Phi = \mathrm{Tr}$. In particular, by Kadison's inequality, this implies $\|\Phi(a)\|_{2,\mathrm{Tr}} \leq \|a\|_{2,\mathrm{Tr}}$, $\forall\, a \in L^2(T, \mathrm{Tr})$.
Since $\|\lambda\Gamma\Gamma^{\mathrm{t}}\| = 1$, it follows that $\forall\ \delta > 0\ \exists\ F_0 \subset K$ finite such that $T_0 = {}_{F_0}(\lambda\Gamma\Gamma^{\mathrm{t}})_{F_0}$ satisfies $1 \geq \|T_0\| \geq 1 - \delta^2/2$. By the classical Perron-Frobenius theorem applied to T_0 (which is a finite symmetric matrix with nonnegative entries) it follows that there exists $b_0 \in \ell^\infty(K) \simeq P$, supported in the set F_0, with $b_0(k) \geq 0$, $\forall\ k$, and $\langle b_0, b_0\rangle = 1$, such that $T_0 b_0 \geq (1 - \delta^2/2)b_0$. Thus, $\lambda\Gamma\Gamma^{\mathrm{t}} b_0 \geq (1 - \delta^2/2)b_0$.
Let then $b \overset{\mathrm{def}}{=} V^{-1}(b_0) \in \ell^\infty(K) \simeq P$ and note that

$$\|b\|_{2,\mathrm{tr}}^2 = \langle V^{-1}(b_0), V^2 V^{-1}(b_0)\rangle = \langle b_0, b_0\rangle = 1.$$

Moreover, we have:

$$\begin{aligned}
\|\Phi(b) - b\|_{2,\mathrm{Tr}}^2 &\leq 2 - 2\mathrm{Tr}(\Phi(b)b) \\
&= 2 - 2\langle \lambda V^{-1}\Gamma\Gamma^{\mathrm{t}}(b_0), V(b_0)\rangle \\
&= 2 - 2\langle \lambda\Gamma\Gamma^{\mathrm{t}}(b_0), b_0\rangle \\
&\leq 2 - 2(1 - \delta^2/2) = 2\delta^2/2 = \delta^2.
\end{aligned}$$

Thus $\|b - \Phi(b)\|_{2,\mathrm{Tr}} < \delta$ and $\|\Phi(b)\|_{2,\mathrm{Tr}} \geq 1 - \delta$, while $\|b\|_{2,\mathrm{Tr}} = 1$.
By Theorem A.2 it follows that if $\delta < 10^{-4}$ then there exists a finite spectral projection e of b such that $\|\Phi(e) - e\|_{2,\mathrm{Tr}} < \delta^{1/4}\|e\|_{2,\mathrm{Tr}}$.
In particular we have:

$$\begin{aligned}
\|(1 - e)\Phi(e)\|_{2,\mathrm{Tr}}^2 &\leq \|(1 - e)\Phi(e)\|_{2,\mathrm{Tr}}^2 + \|e - e\Phi(e)\|_{2,\mathrm{Tr}}^2 \\
&= \|e - \Phi(e)\|_{2,\mathrm{Tr}}^2 < \delta^{1/4}\|e\|_{2,\mathrm{Tr}}^2.
\end{aligned}$$

Let $F \subset K$ be the support set of $e \in \ell^\infty(K) \simeq P$. By the first 3 lines of the proof of Lemma 3.2 on page 281 of [Po3], we have $v_k^{-1} v_{k_0} \geq \lambda$ for all $k_0, k \in K$ for which $(\Gamma\Gamma^{\mathrm{t}})_{kk_0} \neq 0$. Thus we have

$$(\Phi)_{kk_0} = \lambda v_k^{-1} v_{k_0} \sum_{l \in L} a_{kl} a_{k_0 l} \geq \lambda^2,$$

forall $k, k_0 \in K$ for which the entry (k, k_0) of Φ is nonzero. In particular, this shows that $\Phi(e)(1 - e) \geq \lambda^2 \chi_{\partial F}$, where $\chi_{\partial F} \in \ell^\infty(K)$ is the characteristic function of $\partial F \subset K$. Thus we have

$$\begin{aligned}
\lambda^4 \sum_{k \in \partial F} v_k^2 &= \|\lambda^2 \chi_F\|_{2,\mathrm{Tr}}^2 \\
&\leq \|(1 - e)\Phi(e)\|_{2,\mathrm{Tr}}^2 < \delta^{1/4}\|e\|_{2,\mathrm{Tr}}^2 \\
&= \delta^{1/4} \sum_{k \in F} v_k^2.
\end{aligned}$$

Thus, if $\varepsilon > 0$ was given and we take $\delta = (\lambda^4\varepsilon)^4$ then

$$\sum_{k\in\partial F} v_k^2 < \varepsilon \sum_{k\in F} v_k^2$$

thus proving $1') \Longrightarrow 2)$.
Proof of $2) \Longrightarrow 3)$. By 2), for each $\varepsilon = 2^{-n}$ there exisits a finite subset $F_n \subset K$ such that

$$\sum_{k\in\partial F_n} v_k^2 < 2^{-n} \sum_{k\in F_n} v_k^2$$

or, equivalently,

$$\left(\sum_{k\in\Gamma\Gamma^t F_n} v_k^2 - \sum_{k\in F_n} v_k^2\right) < 2^{-n} \sum_{k\in F_n} v_k^2.$$

Let $f_n \in T' \cap \langle S, T\rangle \simeq \ell^\infty(K)$ be the support projection of F_n. Let ω be a free ultrafilter on $\mathbb{N} \simeq K$ and define ψ_0 on $\ell^\infty(K) \simeq T' \cap \langle S, T\rangle$ by

$$\psi_0 = \lim_{n\to\omega} \mathrm{Tr}(\cdot f_n)/\mathrm{Tr} f_n.$$

Let $\psi \stackrel{\mathrm{def}}{=} \psi_0 \circ E$ and note that $\psi = \lim_{n\to\omega} \mathrm{Tr}(\cdot f_n)/\mathrm{Tr} f_n$ on $\langle S, T\rangle$ as well. Note that for each n we have that $\mathrm{Tr}(\cdot f_n)/\mathrm{Tr} f_n$ has T in its centralizer and it is a normal state on $\langle S, T\rangle$. Since $T' \cap S = \mathbb{C}$ this implies that $\mathrm{Tr}(\cdot f_n)/\mathrm{Tr} f_n$ coincides with the trace τ when restricted to S. Thus, $\psi|_S = \tau$ and ψ has T in its centralizer. Let us show that ψ also has e_N in its centralizer. To do this, it is sufficient to prove that

$$\lim_{n\to\infty} (\|f_n e_N - e_N f_n\|_{1,\mathrm{Tr}}/\mathrm{Tr} f_n) = 0.$$

Let $f_n' \in T' \cap \langle S, T\rangle \simeq \ell^\infty(K)$ be the support projection of $F_n \cup \partial F_n$ and note that we have

$$\lim_{n\to\infty} (\|f_n' - f_n\|_{1,\mathrm{Tr}}/\mathrm{Tr} f_n) = \lim_{n\to\infty} \left(\sum_{k\in\partial F_n} v_k^2 / \sum_{k\in F_n} v_k^2\right) = 0.$$

Also, we have

$$\|f_n e_N - e_N f_n\|_{1,\mathrm{Tr}} \le \|e_N f_n - f_n' e_N f_n\| + 2\|f_n' - f_n\|_{1,\mathrm{Tr}}.$$

So, to prove that $[e_N, \psi] = 0$, it is in fact sufficient to prove that $f_n' e_N f_n = e_N f_n$, $\forall\, n$. We will show that, more generally, we have $s_{F'} e_N s_F = e_N s_F$, $\forall\, F \subset K$, where $F' = F \cup \partial F$ and $s_F = \sum_{k\in F} s_k$, $s_{F'} = \sum_{k\in F'} s_k$. To this end, it is clearly sufficient to do it for single element sets $F = \{k_1\}$. It then amounts to show that if $k_2 \in K \setminus F'$, then $s_{k_2} e_N s_{k_1} = 0$. By the proof of 4.5

we thus need to show that if $k_1, k_2 \in K_{n-1}$ for some n, with $k_2 \notin \{k_1\} \cup \partial\{k_1\}$ and we take a minimal projection q_i in the direct summand labeled by k_i in $N'_{2n-1} \cap M$, for each $i = 1, 2$, then we have $MM^{\rm op} v_{q_2} \perp e_1 MM^{\rm op} v_{q_1} MM^{\rm op}$, where $e_1 = e_N$ and $v_{q_i} = E_{N'_{n-1} \cap M_n}(q_i q_i^{\rm op} f_{-n}^n f_{-2n}^0 f_0^{2n})$, $i = 1, 2$.

Before proving this, note that for such q_1, q_2 we have $q_2 e_N q_1 = 0$ and in fact $q_2(N'_{2n-1} \cap M_1)q_1 = 0$. Now, if we take $x_{1,2} \in M$, $y_{1,2} \in M^{\rm op}$, $x, x_0 \in N_{n-1}$, $y, y_0 \in N^{\rm op}_{n-1}$, then we get

$$\begin{aligned}
&\tau(v_{q_2}^* y_0^{\rm op} y_2^{\rm op} x x_2 e_1 x_1 y_1^{\rm op} v_{q_1} x_0 f_{-2n}^0 x y^{\rm op} f_0^{2n} y_0^{\rm op}) \\
&\quad = \tau(E_{N'_{n-1} \cap M_n}(f_0^{2n} f_{-2n}^0 f_{-n}^n q_2 q_2^{\rm op}) y_0^{\rm op} y_2^{\rm op} x x_2 e_1 x_1 x_0 y_1^{\rm op} y^{\rm op} \\
&\qquad \cdot E_{N'_{n-1} \cap M_n}(q_1 q_1^{\rm op} f_{-n}^n f_0^{2n} f_{-2n}^0) f_{-2n}^0 f_0^{2n}).
\end{aligned}$$

Taking the conditional expectation onto $N'_{2n-1} \cap (M \underset{e_N}{\boxtimes} M^{\rm op})$ and denoting $Y_1^{\rm op} = y_0^{\rm op} y_2^{\rm op} \in M^{\rm op}$, $Y_2^{\rm op} = y_1^{\rm op} y^{\rm op} \in M^{\rm op}$, $X' = E_{N'_{2n-1}}(x x_2 e_1 x_1 x_0) \in N'_{2n-1} \cap M_1$, we thus obtain that the above is equal to:

$$\begin{aligned}
&\tau(f_0^{2n} f_{-2n}^0 f_{-n}^n q_2^{\rm op} q_2 Y^{\rm op} X' Y_2^{\rm op} q_1 q_1^{\rm op} f_{-n}^n f_0^{2n} f_{-2n}^0) \\
&= \tau(f_0^{2n} f_{-2n}^0 f_{-n}^n q_2^{\rm op} Y_1^{\rm op} (q_2 X' q_1) Y_2^{\rm op} q_1^{\rm op} f_{-n}^n) \\
&= 0
\end{aligned}$$

in which we first used that $v_{q_i} f_{-2n}^0 f_0^{2n} = q_i q_i^{\rm op} f_{-n}^n f_0^{2n} f_{-2n}^0$ and then we used that $q_2 X' q_1 = 0$.

Since the elements of the form $x_0 f_{-2n}^0 x$ with $x, x_0 \in N_{n-1}$ span all M, this finishes the proof of the fact that e_N is in the centralizer of ψ. Since ψ is equal to the trace on $S = M \underset{e_N}{\boxtimes} M^{\rm op}$ and has in its centralizer the weakly dense $*$-subalgebra generated by $T = M \vee M^{\rm op}$ and e_N in S, by [C3] it follows that ψ has all S in its centralizer. This ends the proof of 2) $\Longrightarrow$ 3).

The proof of 3) $\Longrightarrow$ 4) is then trivial, since the relative amenability of $T = M \vee M^{\rm op} \subset M \underset{e_N}{\boxtimes} M^{\rm op} = S$ merely requires the existence of a state on $\langle S, T\rangle$ which has S in its centralizer, while condition 3) provides very special such states.

To prove 4) $\Longrightarrow$ 5) we need the following lemma.

5.4. Lemma. *Let $\mathcal{M}_0 = {\rm vN}(M \cup J_S M J_S)$, $\mathcal{N}_0 = {\rm vN}(N \cup J_S M J_S)$ and $\Phi_0 : \mathcal{B}(L^2(S)) \to \mathcal{B}(L^2(S))$ be defined by $\Phi_0(T) = \lambda \sum_j m_j^{\rm op} T m_j^{\rm op *}$, where $\{m_j^{\rm op}\}_j$ is an orthonormal basis of $M_1^{\rm op}$ over $M^{\rm op}$ and $\lambda = [M : N]^{-1}$ as usual. Then $\Phi_0(\mathcal{M}_0) = \mathcal{N}_0$, $\mathcal{E}_0 = \Phi_0|_{\mathcal{M}_0}$ is a conditional expectation and in fact*

$$\begin{array}{ccc}
\mathcal{N}_0 & \overset{\mathcal{E}_0}{\subset} & \mathcal{M}_0 \\
\cup & & \cup \\
N & \overset{E_N}{\subset} & M
\end{array}$$

is a commuting square embedding of $N \subset M$, which is isomorphic to the standard representation of $N \subset M$. Moreover, if $\mathcal{N}_0^{\mathrm{f}} = J_S M J_S' \cap \mathcal{N}_0 \subset J_S M J_S' \cap \mathcal{M}_0 = \mathcal{M}_0^{\mathrm{f}}$, then

$$\begin{array}{ccc} \mathcal{N}_0^{\mathrm{f}} & \overset{\mathcal{E}_0^{\mathrm{f}}}{\subset} & \mathcal{M}_0^{\mathrm{f}} \\ \cup & & \cup \\ N & \subset & M \end{array}$$

is a commuting square embedding isomorphic to the finite standard representation of $N \subset M$.

Proof. By construction, we see that $\mathcal{N}_0 \subset \mathcal{M}_0$ is a direct summand of $(N \otimes M^{\mathrm{op}})^{**} \overset{(E\otimes \mathrm{id})^{**}}{\subset} (M \otimes M^{\mathrm{op}})^{**}$. Also, since $\mathcal{N}_0 \subset (M_1^{\mathrm{op}} \cup J_S M^{\mathrm{op}} J_S)' \cap \mathcal{B}(L^2(S))$, we have

$$\begin{aligned} \mathcal{M}_0' \cap \mathcal{N}_0 &\subset (M \cup J_S M J_S)' \cap (M_1^{\mathrm{op}} \cup J_S M^{\mathrm{op}} J_S)' \\ &= \mathrm{vN}(M \cup M_1^{\mathrm{op}} \cup J_S M J_S \cup J_S M^{\mathrm{op}} J_S)' \cap \mathcal{B}(L^2(S)) \\ &= (M \underset{e_N}{\boxtimes} M^{\mathrm{op}} \cup J_S (M \vee M^{\mathrm{op}}) J_S)' \\ &= J_S((M \vee M^{\mathrm{op}})' \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}}) J_S = \mathbb{C}1. \end{aligned}$$

Thus $\mathcal{Z}(\mathcal{M}_0) \cap \mathcal{Z}(\mathcal{N}_0) = \mathbb{C}$. But if p_0 denotes the projection of $L^2(S)$ onto $L^2(M)$ then clearly $p_0 \mathcal{M}_0 p_0 = \mathcal{M}_0 p_0$ is isomorphic to $\mathcal{B}(L^2(M))$ as a $M \otimes M^{\mathrm{op}}$ representation. Thus, $\mathcal{N}_0 \subset \mathcal{M}_0$ must in fact coincide with $\mathcal{N}^{\mathrm{st}} \subset \mathcal{M}^{\mathrm{st}}$. The last part is now clear, since this isomorphism sends $1 \otimes M^{\mathrm{op}}$ onto $J_S M J_S$. Q.E.D.

Proof of $4) \Longrightarrow 5) \Longleftrightarrow 5')$. The equivalence of 5) and 5′) was proved in [Po2], the argument being identical to Connes' single algebra analogue statement. Let us then prove $4) \Longrightarrow 5')$. So let ψ be a S-hypertrace on $\langle S, T\rangle = J_S T J_S' \cap \mathcal{B}(L^2(S))$. Since $T = M \vee M^{\mathrm{op}}$ and $\mathcal{M}_0 \subset J_S M^{\mathrm{op}} J_S' \cap \mathcal{B}(L^2(S))$ (we've already noticed this in the above lemma) it follows that

$$\begin{aligned} \mathcal{M}_0^f &= (J_S M J_S)' \cap \mathcal{M}_0 \subset (J_S(\mathrm{vN}(M \cup M^{\mathrm{op}}))J_S)' \cap \mathcal{B}(L^2(S)) \\ &= (J_S T J_S)' \cap \mathcal{B}(L^2(S)) = \langle S, T\rangle. \end{aligned}$$

Thus ψ restricts to a state ϕ on $\mathcal{M}_0^{\mathrm{f}}$ which has M in its centralizer (since ψ has S in its centralizer and S contains M).
Note now that if $T \in \mathcal{M}_0^f$ then

$$\psi(e_N T) = \psi(u^{\mathrm{op}}(e_N T)u^{\mathrm{op}*}) = \psi((u^{\mathrm{op}} e_N u^{\mathrm{op}*})T), \qquad \forall\ u^{\mathrm{op}} \in \mathcal{U}(M^{\mathrm{op}}).$$

Averaging by unitaries in $\mathcal{U}(M^{\mathrm{op}})$ and using that $\overline{\mathrm{co}}^n\{u^{\mathrm{op}} e_N u^{\mathrm{op}*} \mid u^{\mathrm{op}} \in \mathcal{U}(M^{\mathrm{op}})\} \cap \mathbb{C}1 = \{\lambda 1\}$ (see the Appendix A.1), it follows that

$$\psi(e_N T) = \lambda \psi(T) = \lambda \phi(T).$$

But $e_N \in S$ is in the centralizer of ψ so

$$\psi(e_N T) = \psi(e_N T e_N) = \psi(\mathcal{E}_0(T) e_N).$$

By the same argument as above, the latter equals

$$\lambda\psi(\mathcal{E}_0(T)) = \lambda\phi(\mathcal{E}_0(T)).$$

Thus $\phi = \phi \circ \mathcal{E}_0^{\mathrm{f}}$ showing that ϕ is a $(N \subset M)$-hypertrace on $\mathcal{N}_0^f \overset{\mathcal{E}_0^{\mathrm{f}}}{\subset} \mathcal{M}_0^f$ thus on $\mathcal{N}_0^{\mathrm{st,f}} \overset{\mathcal{E}_0^{\mathrm{st,f}}}{\subset} \mathcal{M}_0^{\mathrm{st},f}$, proving $5')$.

Proof of $5') \implies 1)$. Since $\mathcal{N}_0^{\mathrm{st}f} \overset{\mathcal{E}_0^{\mathrm{st}f}}{\subset} \mathcal{M}_0^{\mathrm{st},f}$ has inclusion matrix $(\Gamma_{N,M})^{\mathrm{t}}$ and the trace Tr on $\mathcal{M}^{\mathrm{st},f}$ defined in 5.2 is preserved by $\mathcal{E}^{\mathrm{st},f}$, it follows by the general result in [Po13] that $\|\Gamma_{N,M}\|^2 = \|\Gamma_{N,M}^{\mathrm{t}}\|^2 = [M : N]$.
This ends the proof of the theorem. Q.E.D.

5.5 *Remarks.* 1°. Of all the equivalent characterisations of amenability for standard graphs, the Kesten-type amenability condition $\|\Gamma_{N,M}\|^2 = [M : N]$ seems to remain the easiest to check in practice. For instance, it immediately implies that if $[M : N] \leq 4$ then $\Gamma_{N,M}$ is amenable, and it is the condition that was used by Bisch and Haagerup to construct many examples of infinite depth subfactors with amenable graphs, by taking compositions between a fixed point algebra inclusion and a cross product inclusion, corresponding to actions of finite groups ([BiH]). Nevertheless, each of the other equivalent characterizations of amenability provided in [Po2-5] and in this paper has its own role in understanding various combinatorial and functional analytical aspects of this concept. The main interest in this notion of amenability comes from the fact that the hyperfinite subfactors having amenable graphs are precisely those that can be recovered from their standard invariants and are thus, in particular, completely classified by this invariant (see 7.1, 7.2 later in this paper, and also [Po16]).
2°. Note that in the proof of the Følner condition 5.3.2 for $\Gamma_{N,M}$, from the Kesten-type condition $\|\Gamma_{N,M}\|^2 = [M : N]$ (taken as the definition of the amenability for a graph) we do not actually use the fact that $\Gamma_{N,M}$ is standard, i.e., the fact that it comes from a subfactor. Indeed, the proof goes the same for any weighted bipartite graph (see [Po14] for more comments on this). However, by using the ergodicity property 4.8 of the standard invariant and of its subalgebra generated by the Jones projections, one can prove an interesting sharper Følner type condition for standard graphs. This will be discussed in a forthcoming paper.

5.6 Corollary. (a). *Let $\mathcal{G}$ be a standard λ-lattice and $\mathcal{G}_0$ a sublattice. If $\mathcal{G}_0$ is amenable then $\mathcal{G}$ is amenable. Conversely, if $[\mathcal{G} : \mathcal{G}_0] < \infty$ and $\mathcal{G}$ is amenable then $\mathcal{G}_0$ is amenable.*
(b). *Let $\mathcal{G}_k = \{A_{ij}^k\}_{i,j\geq 0}$ be standard λ_k-lattices with corresponding graphs $\Gamma_{\mathcal{G}_k} = \Gamma_k$, $k = 1, 2$. Let $\mathcal{G}$ denote the system of finite dimensional algebras*

$A_{ij} \overset{def}{=} A^1_{ij} \otimes A^2_{ij}$, $i, j \geq 0$, *with the tensor product trace. Then* $\mathcal{G}$ *is a standard* $\lambda_1\lambda_2$*- lattice, its graph* Γ *is naturally identified with the tensor product of the graphs* Γ_k *(regarded as matrices) and we have that* $\mathcal{G}$ *is amenable if and only if both* $\mathcal{G}_1$ *and* $\mathcal{G}_2$ *are amenable.*

Proof. (a). The first part follows trivially from the (Kesten-type) definition of amenability, since $\mathcal{G}_0 \subset \mathcal{G}$ implies $\|\Gamma_{\mathcal{G}}\| \geq \|\Gamma_{\mathcal{G}_0}\|$. The second part follows from 2.11, 5.3.4 and [Po8].
(b). The first part follows imediately from the axiomatization of standard lattices in [Po7]. The second part follows from the definition of the amenability, because we have $\|\Gamma\| = \|\Gamma_1\|\|\Gamma_2\|$, so that $(\lambda_1\lambda_2)^{-1} = \|\Gamma\|^2$ iff $\lambda_1^{-1} = \|\Gamma_1\|^2$ and $\lambda_2^{-1} = \|\Gamma_2\|^2$. Note that, by using 2.9.d) and [Po8], this part is an imediate consequence of 5.3.1 as well. Q.E.D.

6. Some More Characterizations of the Amenability for $\Gamma_{N,M}$ and $\mathcal{G}_{N,M}$

In this section we prove several more equivalent characterizations of the amenability for standard graphs and lattices, which clarify some of the results and ideas of the approach to amenability in [Po2,12]. We mention that, while related in spirit with the rest of the paper, the present section will not make explicit use of the symmetric enveloping algebras. So, in this respect, it can be regarded as a digression.
To state the first result, recall that if $B \subset A$ is an inclusion of von Neumann subalgebras of an ambient type II_1 factor then $H(A \mid B)$ denotes its Connes-Størmer relative entropy. By [PiPo1], if $N \subset M$ is an extremal inclusion of type II_1 factors then $H(M \mid N) = \ln([M : N])$. Also, if $N \subset M \subset M_1 \subset \cdots$ is the Jones tower associated to $N \subset M$ then

$$H(M' \cap M_{k+1} \mid M' \cap M_k) \leq H(M_{k+1} \mid M_k) = \ln([M_{k+1} : M_k]) = \ln([M : N]),$$

for all $k \geq 0$. More generally, if p is a projection in $M' \cap M_k$ then by [PiPo1] we have

$$\begin{aligned} &H(p(M' \cap M_{k+1})p \mid p(M' \cap M_k)p) \\ &\qquad\qquad \leq H(pM_{k+1}p \mid pM_kp) = \ln([pM_{k+1}p : pM_kp]) \\ &\qquad\qquad = \ln([M_{k+1} : M_k]) = \ln([M : N]) = H(M \mid N). \end{aligned}$$

Similarly, if $N^{\mathrm{st}} \subset M^{\mathrm{st}}$ denotes as usual the "model" inclusion generated by the higher relative commutants, as in [Po2], then the same remark as above shows that $H(pM^{\mathrm{st}}p \mid pN^{\mathrm{st}}p) \leq H(M \mid N)$, $\forall\, p \in \mathcal{P}(N^{\mathrm{st}})$.
The result that follows states that this "upper bound" for the "local relative entropies" is attained precisely when $\mathcal{G}_{N,M}$ (equivalently $\Gamma_{N,M}$) is amenable. Since $H(pM' \cap M_{k+1}p \mid pM' \cap M_kp)$ also represents the conditional entropy from step k to step $k+1$ of the restriction to the support set of p (in K or

L) of the random walk on the graph $\Gamma \circ \Gamma^{\rm t} \circ \Gamma \circ \Gamma^{\rm t} \cdots$, for $\Gamma = \Gamma_{N,M}$, with transition probabilities determined by $v = (v_k)_{k\in K}$, this maximality condition on the entropy can be interpreted as a *local Shanon-McMillan-Breimann type condition*, in the same spirit as 5.3.5 in [Po2].

6.1. Theorem. *Let $N \subset M$ be an extremal inclusion of type* II_1 *factors. The following conditions are equivalent.*
1) *$\mathcal{G}_{N,M}$ is amenable.*
2) *$\forall\ \varepsilon > 0$, $\exists\ n \geq 1$ and $p \in \mathcal{P}(M' \cap M_n)$ such that*

$$\|E_{(pM'\cap M_{n+1}p)'\cap(pM'\cap M_{n+2}p)}(e_{n+2}p) - \lambda p\|_2 < \varepsilon \|p\|_2.$$

3) *$\forall\ \varepsilon > 0$, $\exists\ p \in \mathcal{P}(N_1^{\rm st})$ such that*

$$\|E_{pN^{\rm st}p'\cap pM^{\rm st}p}(e_0 p) - \lambda p\|_2 < \varepsilon\|p\|_2.$$

4)
$$\lim_k \sup_{p\in\mathcal{P}(M'\cap M_k)} H(pM'\cap M_{k+1}p \mid pM'\cap M_k p)$$
$$= H(M \mid N) = \ln([M:N]).$$

5)
$$\sup_{p\in\mathcal{P}(N^{\rm st})} H(pM^{\rm st}p \mid pN^{\rm st}p) = H(M\mid N).$$

Proof. First of all, note that since by [Po2] we have that $\Gamma_{N,M}$ is amenable if and only if Γ_{M,M_1} is amenable, it is suficient to prove the above equivalences in the case n is even in condition 2) and the k's are taken odd in condition 4).
1) $\Longrightarrow$ 2). If $\mathcal{G}_{N,M}$ is amenable then by 5.3 its graph $\Gamma_{N,M}$ verifies the Følner condition 5.3.2). Thus, $\forall\ \varepsilon > 0$, $\exists\ F \subset K$ finite non-empty such that

$$\sum_{k\in\partial F} v_k^2 < (\varepsilon/2)\sum v_k^2.$$

Let $n_0 \geq 1$ be such that $F' \overset{\rm def}{=} F \cup \partial F$ is included in K_n, $\forall\ n \geq n_0$.
For each $n \geq n_0$ let $\{p_k^n\}_{k\in K_n}$ be the list of minimal central projections of $M' \cap M_{2n}$. Note that $\forall\ k \in K$ we have

$$\lim_{n\to\infty} \dim(M'\cap M_{2n}p_k^n) = \infty.$$

Let $\delta > 0$. Let $\{m_k\}_{k\in F'}$ be positive integers such that

$$(*) \qquad \left|\frac{m_k}{m_{k'}} - \frac{v_k}{v_{k'}}\right| < \delta \min\{v_r/v'_r \mid r, r' \in F'\}, \qquad \forall\ k, k' \in F'.$$

Fix $n \geq n_0$ large enough such that

$$\dim(M'\cap M_{2n}p_k^n) \geq m_k^2, \qquad \forall\ k \in F'.$$

Then for each $k \in F'$ choose $q_k \in \mathcal{P}(M' \cap M_{2n}p_k^n)$ such that $\dim(q_k M' \cap M_{2n} q_k) = m_k^2$. Let $p \overset{\text{def}}{=} \sum_{k \in F'} q_k$. We will show that, for $\delta > 0$ small enough, p satisfies condition 2).

To this end denote by $G = \Gamma^{\mathrm{t}} F'$ the set of simple summands of $pM' \cap M_{2n+1}p$ and by $\{\bar{q}_l\}_{l \in G}$ the corresponding minimal central projections. Let also $\{s_k\}_{k \in F'}$, $\{t_l\}_{l \in G}$ denote the traces of the minimal projections in $pM' \cap M_{2n}p$ and respectively $pM' \cap M_{2n+1}p$.

Thus, if $\Gamma = (a_{kl})_{k \in K, l \in L}$ then for each $k \in F$, $l \in G$ with $a_{kl} \neq 0$ we have $t_l = \lambda \sum_{k' \in K} a_{k'l} s_{k'} = \lambda \sum_{k' \in F'} a_{k'l} s_{k'}$. Also, if we denote by $n_l^2 = \dim(\bar{q}_l M' \cap M_{2n+1}\bar{q}_l)$ and ${m'}_k^2 = \dim(q'_k M' \cap M_{2n+2} q'_k)$, where $\{q'_k\}_{k \in F''}$, $F'' = F' \cup \partial F'$, are the minimal central projections of $pM' \cap M_{2n+2}p$, then $n_l = \sum_{k' \in F'} a_{k'l} m_{k'}$ and $m'_k = \sum_{k'' \in F'} b_{kk''} m_{k''}$, where $(b_{kk'})_{k,k' \in K} = \Gamma\Gamma^{\mathrm{t}}$.

From $(*)$ it follows that for $k \in F$ and $l \in G$ with $a_{kl} \neq 0$ we have the estimates:

$$\begin{aligned}
\left| \frac{t_l}{s_k} - \frac{n_l}{m'_k} \right| &= \left| \lambda \sum_{k' \in F'} a_{k'l} s_{k'} / s_k - \sum_{k' \in F'} a_{k'l} m_{k'} \Big/ \sum_{k'' \in F'} b_{kk''} m_{k''} \right| \\
&\leq \left| \lambda \sum_{k' \in F'} a_{k'l} s_{k'} / s_k - \sum_{k' \in F'} a_{k'l} s_{k'} \Big/ \sum_{k'' \in F'} b_{kk''} s_{k''} \right| + f(\delta) \\
&= \left| \lambda \sum_{k' \in F'} a_{k'l} s_{k'} / s_k - \sum_{k' \in F'} a_{k'l} s_{k'} / \lambda^{-1} s_k \right| + f(\delta) \\
&= f(\delta)
\end{aligned}$$

where $f(\delta) \to 0$ as $\delta \to 0$ and in which we used that for $k \in F$ we have

$$(\Gamma\Gamma^{\mathrm{t}}(s_{k''})_{k'' \in F'})_k = \lambda^{-1} s_k.$$

With these estimates in mind recall that, with the above notations, we have (see e.g. Sec.6 in [PiPo1]):

$$\begin{aligned}
E_{(pM' \cap M_{2n+1}p)' \cap (pM' \cap M_{2n+2}p)}(e_{2n+2}p) &= \sum_{k \in F'',\ l \in G} (\tau(q'_k \bar{q}_l)^2 / a_{kl}^2 \tau(q'_k) \tau(\bar{q}_l)) \bar{q}_l q'_k \\
&= \sum_{k \in F'',\ l \in G} (\lambda s_k n_l / m'_k t_l) \bar{q}_l q'_k.
\end{aligned}$$

But from the above estimates we see that for all $k \in F$ and $l \in G$ with $a_{kl} \neq 0$ we have:

$$|\lambda s_k n_l / m'_k t_l - \lambda| < f'(\delta)$$

where $f'(\delta) \to 0$ as $\delta \to 0$.

This would finish the proof if we could show that the trace of the sum of the projections $\bar{q}_l q'_k$ for $l \in G$ and $k \in F'' \setminus F$ is small with respect to the trace of p. To show this, it is sufficient to show that $\sum_{k\in F''-F}\tau(q'_k)$ is small with respect to $\tau(p)$. To this end, note first that we have

$$\begin{aligned}\sum_{k\in F}\tau(q'_k) &= \sum_{k\in F}\lambda m'_k t_k = \lambda\sum_{k\in F}\left(\sum_{k''\in F'} b_{kk''}m_{k''}\right)t_k\\ &\geq \lambda\sum_{k\in F}\sum_{k''\in F'} b_{kk''}m_k t_{k''} - \lambda\delta\sum_{k\in F,\ k''\in F'} b_{kk''}m_k t_k\\ &\geq \sum_{k\in F}\tau(q_k) - \lambda^{-2}\delta\sum_{k\in F}\tau(q_k)\end{aligned}$$

in which we first used $(*)$ and then the fact that $\sum_{k''\in K} b_{kk''}\leq\lambda^{-3}$, $\forall\ k\in K$ (see e.g., [Po3], page 281). Thus we get:

$$\begin{aligned}\sum_{k\in F''-F}\tau(q'_k) &= \sum_{k\in F''}\tau(q'_k) - \sum_{k\in F}\tau(q'_k)\\ = \tau(p) - \sum_{k\in F}\tau(q'_k) &\leq \tau(p) - \sum_{k\in F}\tau(q_k) + \lambda^{-2}\delta\sum_{k\in F}\tau(q_k)\\ &\leq \sum_{k'\in\partial F}\tau(q_{k'}) + \lambda^{-2}\delta\tau(p).\end{aligned}$$

But by applying $(*)$ again we also have

$$\begin{aligned}\sum_{k'\in\partial F}\tau(q_{k'})\Big/\tau(p) &= \sum_{k'\in\partial F}\tau(q_{k'})\Big/\sum_{k\in F'}\tau(q_k) \leq \sum_{k'\in\partial F}\tau(q_{k'})\Big/\sum_{k\in F}\tau(q_k)\\ &= \sum_{k'\in\partial F} m_{k'}s_{k'}\Big/\sum_{k\in F} m_k s_k = \sum_{k'\in\partial F}\left(s_{k'}\Big/\sum_{k\in F}\frac{m_k}{m_{k'}}s_k\right)\\ &\leq (1-\delta)^{-1}\sum_{k'\in\partial F}\left(s_{k'}\Big/\sum_{k\in F}\frac{v_k}{v_{k'}}\cdot s_k\right)\\ &= (1-\delta)^{-1}\sum_{k'\in\partial F} v_{k'}^2\Big/\sum_{k\in F} v_k^2 < (1-\delta)^{-1}\varepsilon/2.\end{aligned}$$

Altogether we get:

$$\begin{aligned}&\|E_{(pM'\cap M_{2n+1}p)'\cap(pM'\cap M_{2n+2}p)}(e_{2n+2}p) - \lambda p\|_2^2\\ &\qquad\leq f'(\delta)^2\tau(p) + \lambda^{-2}\delta\tau(p) + ((1-\delta)^{-1}\varepsilon^2/2)\tau(p)\\ &= (f''(\delta) + \varepsilon^2/2)\tau(p).\end{aligned}$$

Thus, if δ is chosen sufficiently small to make $f''(\delta) < \varepsilon^2/2$ then the above is majorized by $\varepsilon^2\tau(p)$, thus finishing the proof of 1) $\Longrightarrow$ 2).

Now, 2) $\Longleftrightarrow$ 3) is trivial by the definition of $N^{\mathrm{st}} \subset M^{\mathrm{st}}$. Also 4 $\Longleftrightarrow$ 5) is clear from the continuity properties of the relative entropy under commuting square conditions ([PiPo1]).
Then 2) $\Longrightarrow$ 4) follows from (4.2 in [PiPo1]).
Finally, to prove 4) $\Longrightarrow$ 1), recall from [PiPo2] that if $B \subset A$ is an inclusion of finite dimensional algebras with inclusion matrix T then $\mathrm{Ind}(E_A^B) \geq \|T\|^2 \geq \exp(H(B \mid A))$. Since the inclusion matrix T of $pM' \cap M_{2n+1}p \subset pM' \cap M_{2n+2}p$ is a restriction of $\Gamma_{N,M}$, we have

$$\|\Gamma_{N,M}\|^2 \geq \|T\|^2 \geq \exp(H(pM' \cap M_{2n+2}p \mid pM' \cap M_{2n+1}p)).$$

Thus, if the right hand side term can be made arbitrarily close to $\exp(H(M \mid N)) = [M : N]$ then we obtain $\|\Gamma_{N,M}\|^2 = [M : N]$, i.e., $\mathcal{G}_{N,M}$ follows amenable. Q.E.D.

6.2. Notation. We denote by $\tilde{M}$ the bicommutant of M in its enveloping algebra M_∞, i.e., $\tilde{M} = (M' \cap M_\infty)' \cap M_\infty$. Similarily we put $\tilde{N} = (N' \cap M_\infty)' \cap M_\infty$ and more generally $\tilde{M}_i = (M_i' \cap M_\infty)' \cap M_\infty$, $i \in \mathbb{Z}$, $\{M_i\}_{i\in\mathbb{Z}}$ being as usual a Jones tunnel-tower for $N \subset M$ and $M_0 = M, M_{-1} = N, M_{-n} = N_{n-1}, n \geq 2$. Note that there exists a unique conditional expectation $\tilde{E}$ from $\tilde{M}$ onto $\tilde{N}$ defined by $\tilde{E}(X) = \lambda \Sigma_j m_j X m_j^*$, for $X \in \tilde{M}$, $\{m_j\}_j$ being any orthonormal basis of $N' \cap M_\infty$ over $M' \cap M_\infty$ (e.g., an orthonormal basis of $\mathrm{vN}\{e_k\}_{k\geq 1}$ over $\mathrm{vN}\{e_k\}_{k\geq 2}$ will do, as the definition of $\tilde{E}$ is anyway easily seen to be independent of the choice of $\{m_j\}_j$) and that $\tilde{E}$ is implemented by e_1, i.e., $e_1 X e_1 = \tilde{E}(X) e_1$ (see Sec. 2.2 in [Po2] or 6.9 in [Po6]). The inclusion $\tilde{N} \overset{\tilde{E}}{\subset} \tilde{M}$ is in fact homogeneous λ-Markov in the sense of (1.2.3 and 1.2.11 of [Po3]) and we have a non-degenerate commuting square

$$\begin{array}{ccc} \tilde{N} & \overset{\tilde{E}}{\subset} & \tilde{M} \\ \cup & & \cup \\ N & \overset{E_N}{\subset} & M \end{array}$$

It should also be noted that, while $\tilde{E}(Y_1 e_0 Y_2) = \lambda Y_1 Y_2, \forall Y_{1,2} \in \tilde{N}$ (this relation can in fact be taken as the definition of $\tilde{E}$), in general $\tilde{E}$ is not trace preserving. In fact, one can easily show (see the proof of 6.4 hereafter) that it is trace preserving if and only if $\tilde{M} = M$, i.e., when the bicommutant relation holds true, $(M' \cap M_\infty)' \cap M_\infty = M$, equivalently when $\Gamma_{N,M}$ is strongly amenable (cf. 5.3.1 in [Po2]).
The Jones tower-tunnel of the above commuting square is obtained by defining the conditional expectations $\tilde{E}_i$ from $\tilde{M}_{i-1}$ onto $\tilde{M}_{i-2}$ in a similar manner with $\tilde{E}$.
Recalling from ([Po2]) that a representation $\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}$ of $N \subset M$ is *smooth* if $N' \cap M_n \subset \mathcal{N}' \cap \mathcal{M}_n, \forall n$, note that by its construction, $\tilde{N} \subset \tilde{M}$ is obviously a smooth representation of $N \subset M$.

6.3. Theorem. *Let $N \subset M$ be an extremal inclusion of type* II_1 *factors. The following conditions are equivalent:*
1) $\Gamma_{N,M}$ *is amenable.*
2) *There exists a (possibly singular) trace ψ on $\tilde{M}$ such that $\psi \circ \tilde{E} = \psi$.*
3) *There exists a norm one projection of $\tilde{N} \overset{\tilde{E}}{\subset} \tilde{M}$ onto $N \subset M$.*
4) *If*

$$\begin{array}{ccc} \mathcal{N} & \subset & \mathcal{M} \\ \cup & & \cup \\ N & \subset & M \end{array}$$

is a smooth representation of $N \subset M$ such that there exists a norm one projection of $\mathcal{M}$ onto M (equivalently, a M-hypertrace on $\mathcal{M}$), then there exists a norm one projection of $\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}$ onto $N \subset M$ (equivalently, a $N \subset M$-hypertrace on $\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}$).
5) *For any smooth representation of $N \subset M$ into an inclusion of type* II_1 *von Neumann algebras $\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}$, there exists a norm one projection of $\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}$ onto $N \subset M$ (equivalently, a $N \subset M$-hypertrace on $\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}$).*

Proof. 1) $\Longrightarrow$ 2) By Theorem 6.1 (see condition 6.1.3) applied to $M \subset M_1$ and the anti-isomorphism between $N_1^{\mathrm{st}} \subset N^{\mathrm{st}} \overset{e_0}{\subset} M^{\mathrm{st}}$ and $M' \cap M_\infty \subset N' \cap M_\infty \overset{e_0}{\subset} N_1' \cap M_\infty$, it follows that there exist projections $p_n \in M' \cap M_\infty$ such that

$$\|E_{(p_n N' \cap M_\infty p_n)' \cap (p_n N_1' \cap M_\infty p_n)}(e_0 p_n) - \lambda p_n\|_2 / \|p_n\|_2 \leq 2^{-n}, \qquad \forall\ n.$$

We then define on M_∞ the state $\varphi \overset{\text{def}}{=} \lim_{n\to\omega} \tau(p_n)^{-1}\tau(\,\cdot\, p_n)$. Note that, since $p_n \in M' \cap M_\infty$, we have $[p_n, (M' \cap M_\infty)' \cap M_\infty] = 0$, in other words $[p_n, \tilde{M}] = 0$. Thus, $[\varphi, \tilde{M}] = 0$, in particular $\varphi|_{\tilde{M}}$ is a trace. Moreover, by noting that $\tau(\,\cdot\, p_n) = \tau(E_{p_n B p_n}(\,\cdot\,)p_n)$ for any von Neumann subalgebra $B \subset M_\infty$ with $p_n \in B$, taking $B = (p_n N' \cap M_\infty p_n)' \cap p_n M_\infty p_n$ and using the above and the Cauchy-Schwartz inequality it follows that for all $x, y \in \tilde{N}$ we have:

$$\begin{aligned}
&|\tau(x e_0 y p_n)/\tau(p_n) - \lambda \tau(x y p_n)/\tau(p_n)| \\
&\quad = |\tau(p_n y x e_0 p_n)/\tau(p_n) - \tau(p_n x y \lambda p_n)/\tau(p_n)| \\
&\quad = |\tau(E_{p_n B p_n}(p_n y x e_0 p_n)/\tau(p_n) - \tau(p_n x y \lambda p_n)/\tau(p_n)| \\
&\quad = \tau(p_n y x (E_{p_n B p_n}(e_0 p_n) - \lambda p_n))/\tau(p_n)) \\
&\quad \leq \|p\|_2 \|yx\|\ \|E_{p_n B p_n}(e_0) - \lambda p_n\|_2/\tau(p_n) \\
&\quad \leq 2^{-n}\|yx\|.
\end{aligned}$$

Since $\tilde{E}(x e_0 y) = \lambda x y$, $\forall\ x, y \in \tilde{N}$, and $\mathrm{sp}\tilde{N} e_0 \tilde{N} = \tilde{M}$, it follows that

$$\lim_{n\to\infty} \|\tau(X p_n)/\tau(p_n) - \tau(\tilde{E}(X) p_n)/\tau(p_n)| = 0 \qquad \forall\ X \in \tilde{M}.$$

Thus, $\varphi(X) = \varphi(\tilde{E}(X))$, $\forall\ X \in \tilde{M}$. All this shows that $\psi \stackrel{\text{def}}{=} \varphi|_{\tilde{M}}$ is both a trace and satisfies $\psi = \psi \circ \tilde{E}$.
2) $\Longrightarrow$ 3). Since ψ is a trace on $\tilde{M}$, it is in particular a M-hypertrace and $\psi = \psi \circ \tilde{E}$ implies it is actually a $(N \subset M)$-hypertrace on $(\tilde{N} \stackrel{\tilde{E}}{\subset} \tilde{M})$; equivalently, there exists a conditional expectation of $\tilde{N} \subset \tilde{M}$ onto $N \subset M$.
3) $\Longrightarrow$ 4). If there exists a conditional expectation Φ of $\mathcal{M}$ onto M then by amplification it follows that there exist conditional expectations Φ_{2n} of $\mathcal{M}_{2n}$ onto M_{2n}, $\forall\ n \geq 0$. Let $\mathcal{F}_{2n} : \cup_k \mathcal{M}_k \to \mathcal{M}_{2n}$ be the conditional expectation implemented by $\cdots \circ \mathcal{E}_{2n+2} \circ \mathcal{E}_{2n+1}$ and denote $\Psi_{2n} : \cup_k \mathcal{M}_k \to M_\infty$ the aplications defined by $\Psi_{2n}(X) = \Phi_{2n} \circ \mathcal{F}_{2n}(X) \in M_{2n} \subset M_\infty$. Note that Ψ_{2n} is M_{2n}-M_{2n} linear. Finally, we put $\Psi(X) \stackrel{\text{def}}{=} \lim_{n\to\omega} \psi_{2n}(X)$, for $X \in \cup_k \mathcal{M}_k$, where ω is a free ultrafilter on $\mathbb{N}$. Thus, $\Psi(1) = 1$ and Ψ is M_{2n}-M_{2n} linear $\forall\ n$. Since the representation of $N \subset M$ into $\mathcal{N} \subset \mathcal{M}$ is smooth, $M' \cap M_j \subset \mathcal{M}' \cap \mathcal{M}_j$, $\forall\ j$. Thus, if $X \in \mathcal{M}$ then $[X, M' \cap M_j] = 0$ and by applying ψ we get $[\Psi(X), M' \cap M_j] = 0$. Thus, $\psi(X) \subset (\cup_j M' \cap M_j)' \cap M_\infty = \tilde{M}$. Similarly, we obtain that if $X \in \mathcal{N}$ then $\Psi(X) \in \tilde{N}$. But by 3) we have a conditional expectation of $\tilde{M}$ onto M, say Ψ_0, such that $\Psi_0(\tilde{N}) = N$.
We then define $\Psi_1 : \mathcal{M} \to M$ by $\Psi_1(X) = \Psi_0(\Psi(X))$, which is a conditional expectation and satisfies $\Psi_1(\mathcal{N}) = \Psi_0(\psi(\mathcal{N})) \subset \Psi_0(\tilde{N}) \subset N$.
4) $\Longrightarrow$ 5). Since $\mathcal{M}$ has projections $p \in \mathcal{Z}(\mathcal{M})$ such that $\mathcal{M}p$ is finite, it follows that there is a conditional expectation of $\mathcal{M}p$ onto $Mp \simeq M$, thus of $\mathcal{M}$ onto M and so 4) applies.
5) $\Longrightarrow$ 1) If 5) holds true then in particular there exists a norm one projection from the finite standard representation onto $N \subset M$, so by Theorem 5.3 we have 1). Q.E.D.

Recall from [Po2] that a standard λ-graph $(\Gamma, \vec{s})$ is called *ergodic* if $\vec{s}$ is the unique $\vec{s}$-bounded eigenvector for $\Gamma\Gamma^t$ corresponding to the eigenvalue λ^{-1}, equivalently, if $\mathcal{Z}(A_{0,\infty}) = \mathbb{C}$, where $A_{0,\infty}$ is the finite von Neumann algebra obtained as an inductive limit with the Bratteli diagram given by $\Gamma, \Gamma^t, \Gamma, ...$, starting from the even vertex $*$ of Γ, and having trace given by $\vec{s} = (s_k)_{k\in K}$. Note that if $N \subset M$ is a subfactor having standard graph $(\Gamma, \vec{s})$ then the algebra $A_{0,\infty}$ equals $M' \cap M_\infty$, where $N \subset M = M_0 \subset M_1 \subset ...$ is the Jones tower for $N \subset M$ and $M_\infty = (\cup_n M_n)^-$ as usual.
In what follows we'll call the standard λ-graph *almost ergodic* if $\dim\mathcal{Z}(A_{0,\infty}) < \infty$. This is equivalent to the fact that, up to scalar multiples, there are only finitely many $\vec{s}$-bounded eigenvectors for $\Gamma\Gamma^t$ corresponding to the eigenvalue λ^{-1} (see the proof of 1.4.2 in [Po2]). Note that Haagerup constructed extremal hyperfinite subfactors of index $\lambda^{-1} = 2 \cdot 4\cos^2\pi/5 = 3 + \sqrt{5}$ which have almost ergodic, but not ergodic, standard graph. The following consequence of 6.3 shows that this cannot happen if Γ is amenable.

6.4. Corollary. *If an amenable, extremal standard graph $(\Gamma, \vec{s})$ is almost ergodic then it is ergodic, and thus it is strongly amenable.*

Proof. Let $N \subset M$ be a subfactor having (weighted) standard graph equal to $(\Gamma, \vec{s})$. Denote like in 6.2 by $\tilde{N}_1 = (N_1' \cap M_\infty)' \cap M_\infty, \tilde{N} = (N' \cap M_\infty)' \cap M_\infty, \tilde{M} = (M' \cap M_\infty)' \cap M_\infty$ and by $\tilde{F}$ the expectation from $\tilde{M}$ onto $\tilde{N}_1$ defined as in 6.2 (so that in fact, with the notations there, we also have $\tilde{F} = \tilde{E}_{-1} \circ \tilde{E}_0$). Note that $\mathcal{Z}(\tilde{M}) = \mathcal{Z}(M' \cap M_\infty)$. Since $N_1 \subset M$ has amenable graph ($= \Gamma\Gamma^t$), by 6.3 there exists a trace τ' on $\tilde{M}$ such that $\tau' \circ \tilde{F} = \tau'$. Since $\dim \mathcal{Z}(\tilde{M}) = \dim \mathcal{Z}(M' \cap M_\infty) < \infty$, it follows that there exists $a \in \mathcal{Z}(\tilde{M})_+$ such that $\tau'(X) = \tau(Xa), \forall X \in \tilde{M}$. Since $\tilde{E}$ is τ'-preserving, this implies that $a = \tilde{F}(a) \in \tilde{F}(\mathcal{Z}(\tilde{M})) = \mathcal{Z}(\tilde{N}_1)$. Thus $a \in \mathcal{Z}(\tilde{M}) \cap \mathcal{Z}(\tilde{N}_1) = \mathbb{C}1$, so $a = 1$ and $\tau' = \tau$.

Thus $\tilde{F}$ coincides with the trace preserving expectation F of $\tilde{M}$ onto $\tilde{N}$. In particular, this implies that $E_{(N_1' \cap M_\infty)' \cap M_\infty}(f) = F(f) = \lambda^2 1$, where $f \in M$ is the Jones projection for $N_1 \subset M$. By duality it follows that $E_{(M_{2j}' \cap M_\infty)' \cap M_\infty}(f_j) = \lambda^2 1$ for any $j \in \mathbb{Z}$, where f_j is the Jones projection for the inclusion $M_{2j} \subset M_{2j+2}$. By (5.3 in [Po2]) it follows that $M = (M' \cap M_\infty)' \cap M_\infty$, so in particular $M' \cap M_\infty$ is a factor, i.e., $(\Gamma, \vec{s})$ is ergodic. Q.E.D.

We now examine the effect of amenability on the universal graph $\Gamma^u_{N,M}$. To this end, let us denote, like in [Po2], by $\mathcal{N}^{u,f} \overset{\mathcal{E}^{u,f}}{\subset} \mathcal{M}^{u,f}$ the direct summand of $\mathcal{N}^u \subset \mathcal{M}^u$ given by all the irreducible representations $\mathcal{B}(\mathcal{H})$ of $M \otimes P^{\mathrm{op}}$ which, when regarded as $M-P$ bimodules, have finite dimension, $\dim({}_M\mathcal{H}_P) < \infty$, P denoting here a generic "dummy" type II_1 factor weakly stably equivalent to M (in the sense of 1.4.3 in [Po8], i.e., P can be embedded with finite index in the amplification by some $\alpha > 0$ of M). Let $\Gamma^{u,f}_{N,M}$ denote its inclusion graph (or matrix). Recall from [Po2] that $\Gamma^{u,f}_{N,M}$ is in a natural way a weighted bipartite graph, the weights being given by the vector $((\dim_{M,P}\mathcal{H})^{1/2})$, which in fact also gives the weights of an $\mathcal{E}^{u,f}$-invariant trace on $\mathcal{M}^{u,f}$.

6.5. Theorem. *Let $N \subset M$ be an extremal inclusion of type II_1 factors. The following conditions are equivalent:*

1) *The standard graph $\Gamma_{N,M}$ is amenable, i.e., $\|\Gamma_{N,M}\|^2 = [M : N]$.*
2) *The graph $\Gamma^{u,f}_{N,M}$ is amenable, i.e., $\|\Gamma^{u,f}_{N,M}\|^2 = [M : N]$.*
3) *Each irreducible component Γ of $\Gamma^u_{N,M}$ satisfies $\|\Gamma\|^2 = [M : N]$.*
4) *For any $\varepsilon > 0$ there exists a subfactor $Q \subset N$, with $[N : Q] < \infty$, such that the inclusion matrix $T_0 = T_{Q' \cap N \subset Q' \cap M}$ satisfies $\|T_0\|^2 \geq [M : N] - \varepsilon$.*
4′) *For any $\varepsilon > 0$ there exists a factor P containing M with $[P : M] < \infty$, such that $\|T_{M' \cap P \subset N' \cap P}\|^2 \geq [M : N] - \varepsilon$.*

Proof. 3) $\Longrightarrow$ 2) is trivial.

2) $\Longrightarrow$ 1). For simplicity of notations, we let $(\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}) = (\mathcal{N}^{u,f} \overset{\mathcal{E}^{u,f}}{\subset} \mathcal{M}^{u,f})$. Let K' be the set of simple summands of $\mathcal{M}$ and TT^{t} be the inclusion matrix of $\mathcal{M} \subset \mathcal{M}_2$. It follows that $\forall\, \varepsilon > 0$, $\exists\, k_0 \in K'$ such that

$$\lim_{n \to \infty} \|(TT^{\mathrm{t}})^n \delta_{k_0}\|^{1/n} \geq \|TT^t\| - \varepsilon = [M : N] - \varepsilon.$$

But if for each $n \geq 0$ we denote by p_{2n} the minimal central projection in $\mathcal{M}_{2n}$ corresponding to $k_0 \in K'$, then

$$\|(TT^{\mathrm{t}})^n \delta_{k_0}\|^2 = \dim(\mathcal{M}p_0' \cap p_0 \mathcal{M}_{2n} p_0) = \dim(\mathcal{M}p_{2n}' \cap \mathcal{M}_{2n} p_{2n}).$$

Moreover, by using that if $R \subset Q \subset P$ are inclusions of type II_1 factors with finite index then $\dim R' \cap P \leq ([P : Q]) \dim R' \cap Q$ (because the inclusion matrix of $R' \cap Q \subset R' \cap P$ has square norm $\leq [P : Q]$), it follows that, with the notations $M^0 = P' \cap \mathcal{M}, M_k^0 = P' \cap \mathcal{M}_k, k \geq 0$, we have:

$$\begin{aligned}
\dim(\mathcal{M}p_{2n}' \cap \mathcal{M}_{2n} p_{2n}) &\leq \dim(Mp_{2n}' \cap M_{2n}^0 p_{2n}) \\
&\leq ([M_{2n}^0 p_{2n} : M_{2n} p_{2n}]) \dim(Mp_{2n}' \cap M_{2n} p_{2n}) \\
&= ([M^0 p_0 : Mp_0]) \dim(M' \cap M_{2n}) \\
&= ([M^0 p_0 : Mp_0]) \|(\Gamma_{N,M} \Gamma_{N,M}^{\mathrm{t}})^n \delta_* \|^2.
\end{aligned}$$

Thus,

$$\lim_{n\to\infty} \|(TT^{\mathrm{t}})^n \delta_{k_0}\|^{1/n} \leq \lim_{n\to\infty} ([\mathcal{M}p_0 : Mp_0])^{1/2n} \|(\Gamma\Gamma^{\mathrm{t}})^n \delta_*\|^{1/n} = \|\Gamma\Gamma^{\mathrm{t}}\|$$

showing that $\|\Gamma\|^2 \geq [M : N] - \varepsilon$. Since ε was arbitrary, $\|\Gamma_{N,M}\|^2 = [M : N]$.
1) $\Longrightarrow$ 3). Let k' be any of the labels corresponding to an even vertex of Γ and let $q \in \mathcal{M}_{2n}^u$ be a minimal central projection corresponding to that same label. Then $\dim((\mathcal{M}^u q)' \cap \mathcal{M}_{2n}^u q) = \|(\Gamma\Gamma^t)^n \delta_{k'}\|^2$. But by smoothness, $(M' \cap M_{2n})q \subset (\mathcal{M}^u q)' \cap \mathcal{M}_{2n}^u q$, thus $\|\Gamma\Gamma^t\| \geq \|\Gamma_{N,M}\Gamma_{N,M}^t\|$.
1) $\Longrightarrow$ 4). This is clear, by simply taking $Q = N_k$, a subfactor in a Jones tunnel, with k large enough.
4) $\Longleftrightarrow$ 4$'$). This follows immediately by taking into account that if $Q \subset N (\subset M)$ is a subfactor of finite index in N and we denote $Q \subset N \subset M \subset M_1 \subset Q_1$ its basic construction, then $T_{Q' \cap N \subset Q' \cap M} = T_{M_1' \cap Q_1 \subset M' \cap Q_1}$.
4$'$) $\Longrightarrow$ 2). If $P \supset M$ is as in condition 4$'$) for some ε then let $p \in \mathcal{Z}(\mathcal{M}^{u,f})$ be the central projection supporting all the $N - M$ bimodules appearing as direct summands in ${}_N L^2(P)_M$. Then clearly $\|\Gamma\| \geq \|\Gamma_p\| \geq \|T_{M' \cap P \subset N' \cap P}\|$. Q.E.D.

6.6. Corollary. *Let $Q \subset N \subset M$ be inclusions of II_1 factors with finite index (not necessarily extremal). (i). If $Q \subset M$ has amenable graph then $Q \subset N$ and $N \subset M$ have amenable graphs. (ii). If $N \subset M$ has amenable graph and $p \in N' \cap M$ is a projection, then $Np \subset pMp$ has amenable graph.*

Proof. By [L], there exist extremal inclusions $Q_0 \subset N_0 \subset M_0$ such that: a). The higher relative commutants of $Q_0 \subset N_0$, $N_0 \subset M_0$ and respectively $Q_0 \subset M_0$ are algebraically isomorphic to those of $Q \subset N$, $N \subset M$ and respectively $Q \subset M$; so, in particular, the graphs of the induced-reduced algebras in the Jones towers of the corresponding subfactors are equa. b). $[N_0 : Q_0] = \mathrm{Ind}E_{\min}^{Q,N}$, $[M_0 : N_0] = \mathrm{Ind}E_{\min}^{N,M}$ and $[M_0 : Q_0] = \mathrm{Ind}E_{\min}^{Q,M}$ and the local indices in the Jones tower for the inclusions $Q_0 \subset N_0, N_0 \subset M_0$,

respectively $Q_0 \subset M_0$ are the same as for the initial incusions $Q \subset N, N \subset M$, respectively $Q \subset M$. With these in mind, let us prove (i) and (ii).
(i). Let $\Gamma^{u,f}_{Q,N}$, $\Gamma^{u,f}_{N,M}$ be as in [Po2] the inclusion matrices describing the inclusions $(Q \otimes M^{\rm op})^{**}_{{\rm at},f} \subset (N \otimes M^{\rm op})^{**}_{{\rm at},f} \subset (M \otimes M^{\rm op})^{**}_{{\rm at},f}$. Recall from ([Po2]) that $\Gamma^{u,f}_{Q,M} = \Gamma^{u,f}_{Q,N} \circ \Gamma^{u,f}_{N,M}$. Thus, if $\|\Gamma^{u,f}_{Q,M}\|^2 = {\rm Ind}E^{Q,M}_{\min}$ then we get

$$\begin{aligned}{\rm Ind}E^{Q,N}_{\min} \cdot {\rm Ind}E^{N,M}_{\min} &= {\rm Ind}E^{Q,M}_{\min} = \|\Gamma^{u,f}_{Q,M}\|^2 \leq \|\Gamma^{u,f}_{Q,N}\|^2 \|\Gamma^{u,f}_{N,M}\|^2 \\ &\leq {\rm Ind}E^{Q,N}_{\min} \cdot {\rm Ind}E^{N,M}_{\min},\end{aligned}$$

forcing the equalities $\|\Gamma^{u,f}_{Q,N}\|^2 = {\rm Ind}E^{Q,N}_{\min}$, $\|\Gamma^{u,f}_{N,M}\|^2 = {\rm Ind}E^{N,M}_{\min}$. But by the above considerations and 6.5 this implies $\Gamma_{Q,N}$ and $\Gamma_{N,M}$ are amenable.
(ii). This can be easily deduced from 6.5, by using the universal graphs as in the proof of (i) above. Instead, we'll use the following simpler argument: By the first part of the proof, we may assume $N \subset M$ is extremal. Then by 2.9.c) it follows that the finite standard representation of $Np \subset pMp$ is given by $\mathcal{N}^{{\rm st},f}p \overset{\mathcal{E}}{\subset} p\mathcal{M}^{\rm st,f}p$, where $\mathcal{E}$ is defined by $\mathcal{E}(pXp) = \tau(p)^{-1}\mathcal{E}^{\rm st,f}(pXp)$, for $X \in \mathcal{M}^{\rm st,f}$. But then, if Φ is a conditional expectation from $\mathcal{M}^{\rm st,f}$ onto $\mathcal{N}^{\rm st,f}$ sending M onto N then clearly Φ also sends $p\mathcal{M}^{\rm st,f}p$ onto pMp and $\mathcal{N}^{\rm st,f}p$ onto Np. By 5.3, this implies that $Np \subset pMp$ has amenable graph. Q.E.D.

We mention one last hereditarity property for the amenability of the graphs of subfactors, which has a self-contained and rather elementary proof.

6.7. Proposition. *Let*

$$\begin{array}{ccc} N & \subset & M \\ \cup & & \cup \\ Q & \subset & P \end{array}$$

be a nondegenerate commuting square of inclusions of type $\rm II_1$ *factors with finite index (thus,* $[M:N]=[P:Q]<\infty$, $[M:P]=[N:Q]<\infty$*). Then we have:*
a) $\|\Gamma_{N,M}\| = \|\Gamma_{Q,P}\|$, $H(M \mid N) = H(P \mid Q)$, ${\rm Ind}E^{N,M}_{\min} = {\rm Ind}E^{Q,P}_{\min}$ *and* $E_{Q'\cap P}(e_0) = E_{N'\cap M}(e_0)$, *where* $e_0 \in P$ *is a Jones projection for* $Q \subset P$ *(and thus for* $N \subset M$ *as well).*
b) $N^{\rm st} \subset M^{\rm st}$ *has atomic centers iff* $Q^{\rm st} \subset P^{\rm st}$ *has atomic centers.*
c) $\mathcal{G}_{N,M}$ *is amenable (resp. strongly amenable, resp. has finite depth) iff* $\mathcal{G}_{Q,P}$ *is amenable (resp. strongly amenable, resp. has finite depth).*

Proof. Let

$$\begin{array}{ccccc} \dots N_1 & \subset & N & \subset & M \\ \cup & & \cup & & \cup \\ \dots Q_1 & \subset & Q & \subset & P \end{array}$$

be a tunel for the given commuting square. Then $\dim N_k' \cap M \leq \dim Q_k' \cap M \leq [M:P] \dim Q_k' \cap P$, so that

$$\|\Gamma_{N,M}\|^2 = \lim_{k\to\infty} (\dim N_k' \cap M)^{1/k} \leq \lim_{k\to\infty} (\dim Q_k' \cap P)^{1/k} = \|\Gamma_{Q,P}\|^2.$$

Taking

$$\begin{array}{ccc} \langle N,Q\rangle & \subset & \langle M,P\rangle \\ \cup & & \cup \\ N & \subset & M \end{array}$$

and using that $\langle N,Q\rangle \subset \langle M,P\rangle$ is an amplified of $Q \subset P$ (so that $\Gamma_{\langle N,Q\rangle,\langle M,P\rangle} = \Gamma_{Q,P}$), by the first part we also get $\|\Gamma_{Q,P}\| \geq \|\Gamma_{N,M}\|$, thus $\|\Gamma_{N,M}\| = \|\Gamma_{Q,P}\|$.

Now remark that $E_P(Q'_k \cap M) = Q'_k \cap P$ and $E_P(\mathcal{Z}(Q'_k \cap M)) \subset \mathcal{Z}(Q'_k \cap P)$. Also, we have

$$\mathrm{Ind}(E^{Q'_k\cap M}_{Q'_k\cap P}) \leq [M:P],$$
$$\mathrm{Ind}(E^{Q'_k\cap M}_{N'_k\cap M}) \leq [Q'_k : N'_k] = [N_k : Q_k] = [M:P].$$

It follows that if we denote $R = \overline{\cup_k Q'_k \cap M}$ then $\mathrm{Ind}(E^R_{P^{\mathrm{st}}}) \leq [M:P]$, $\mathrm{Ind}(E^R_{M^{\mathrm{st}}}) \leq [M:P]$. Thus, P^{st} has atomic center iff R has atomic center iff M^{st} has atomic center.

Also, the above shows that

$$\sup_n \dim \mathcal{Z}(N'_k \cap M) < \infty \Longleftrightarrow \sup_n \dim \mathcal{Z}(Q'_k \cap M) < \infty$$
$$\Longleftrightarrow \sup_k \dim \mathcal{Z}(Q'_k \cap P) < \infty.$$

Thus, $N \subset M$ has finite depth iff $Q \subset P$ has finite depth.

Since $Q \subset P$ is embedded as a commuting square in $N \subset M$, by the definition of relative entropy we have $H(P \mid Q) \leq H(M \mid N) \leq H(\langle M,P\rangle \mid \langle N,Q\rangle) = H(P \mid Q)$, thus, $H(M \mid N) = H(P \mid Q)$.

Next, if $e_0 \in P$ is a Jones projection then

$$E_{N'\cap M}(e_0) = E_{N'\cap M}(E_{Q'\cap M}(e_0)) = E_{N'\cap M}(E_{Q'\cap P}(e_0))$$

so that $\|E_{N'\cap M}(e_0)\|_2 \leq \|E_{Q'\cap P}(e_0)\|_2$ with equality iff $E_{N'\cap M}(e_0) = E_{Q'\cap P}(e_0)$. But $N \subset M$ is embedded as a commuting square in $\langle N,Q\rangle \subset \langle M,P\rangle$ which is an amplified of $Q \subset P$, so we get similarly $\|E_{Q'\cap P}(e_0)\|_2 \leq \|E_{N'\cap M}(e_0)\|_2$ giving $E_{N'\cap M}(e_0) = E_{Q'\cap P}(e_0)$.

To prove the statemnt about the minimal index, note from the formula of the Jones projection in ([PiPo1], page 83-84) that $E^{N,M}_{\min} = E^M_N(b^{1/2} \cdot b^{1/2})$ with $b \in \mathrm{Alg}\{E_{N'\cap M}(e_0)\} = \mathrm{Alg}\{E_{Q'\cap P}(e_0)\}$. Thus, $b \in P$ and $E^{N,M}_{\min}(P) = Q$, implying that $\mathrm{Ind}(E^{N,M}_{\min}) \geq \mathrm{Ind}(E^{Q,P}_{\min})$. Similarily, $\mathrm{Ind}(E^{N,M}_{\min}) \leq \mathrm{Ind}(E^{\langle N,Q\rangle,\langle M,P\rangle}_{\min})$ Thus, $\mathrm{Ind}E^{M,N}_{\min} = \mathrm{Ind}E^{P,Q}_{\min}$.

From the above, it follows in particular that $\mathrm{Ind}E^{N,M}_{\min} = \|\Gamma_{N,M}\|^2$ iff $\mathrm{Ind}E^{Q,P}_{\min} = \|\Gamma_{Q,P}\|^2$ so $\Gamma_{N,M}$ is amenable iff $\Gamma_{Q,P}$ is amenable (without the extremality assumtion required).

If $\Gamma_{N,M}$ is amenable and M^{st} is a factor (i.e., $\mathcal{G}_{N,M}$ is strongly amenable) then $\Gamma_{Q,P}$ is amenable and P^{st} has finite dimensional center. Thus 6.4 applies to get that $\Gamma_{Q,P}$ follows strongly amenable. Alternatively, and in order to keep the proof of this Proposition elementary and self-contained, note that the same proof as on (pages 235 and 183 of [Po2]) can be used to get the same conclusion, i.e., that P^{st} follows a factor and thus $\Gamma_{Q,P}$ strongly amenable. Q.E.D.

7. Hyperfiniteness of $M \underset{e_N}{\boxtimes} M^{\text{op}}$ and Hereditarity of the Amenability for Subfactors

Recall from [Po2] that an inclusion of factors $N \subset M$ is called *amenable* if it is the range of a norm one projection from any of its smooth representations, equivalently, if the algebras N, M are themselves amenable (i.e., hyperfinite by [C1]) and the graph $\Gamma_{N,M}$ is amenable ([Po2,3,4]), i.e., $\|\Gamma_{N,M}\|^2 = \text{Ind}E_{\min}^{N,M}$. In this section we will show that, in the case the inclusion $N \subset M$ is extremal, the amenability of $N \subset M$ is in fact equivalent to the hyperfiniteness of its symmetric enveloping algebra. We will then derive that the amenability of an inclusion is inherited by its "sub-inclusions"

7.1. Theorem. *Let $N \subset M$ be an extremal inclusion of type II_1 factors. The following conditions are equivalent:*
1) *$N \subset M$ is amenable.*
2) *$\Gamma_{N,M}$ is amenable and M is hyperfinite.*
3) *$\forall\ x_1, \dots, x_m \in M$, $\forall\ \varepsilon > 0$, $\exists\ n$, a projection f in $N' \cap M_n$, a subfactor $P \subset N$ such that $Pf \subset Nf \subset fM_nf$ is a basic construction and a finite dimensional subfactor $Q_0 \subset P$ such that*

$$x_i \in_\varepsilon Q_0 \vee (P' \cap M), \qquad i = 1, 2, \dots, m.$$

4) *$\forall\ x_1, \dots, x_m \in M$, $\forall\ \varepsilon > 0$, $\exists\ Q \subset N$ with $[N : Q] < \infty$ such that*

$$x_i \in_\varepsilon Q' \cap M, \qquad i = 1, 2, \dots, m.$$

5) *$M \underset{e_N}{\boxtimes} M^{\text{op}}$ is isomorphic to the hyperfinite type II_1 factor.*
6) *There exists a $M \underset{e_N}{\boxtimes} M^{\text{op}}$-hypertrace on $\mathcal{B}(L^2(M \underset{e_N}{\boxtimes} M^{\text{op}})$.*
7) *There exists a $(N \subset M)$-hypertrace on $\mathcal{N}^{\text{st}} \subset \mathcal{M}^{\text{st}}$ (equivalently, a norm one projection of $\mathcal{N}^{\text{st}} \subset \mathcal{M}^{\text{st}}$ onto $N \subset M$).*

Proof. We will prove 1) $\Longrightarrow$ 7) $\Longrightarrow$ 2) $\Longrightarrow$ 3) $\Longrightarrow$ 4) $\Longrightarrow$ 5) $\Longrightarrow$ 6) $\Longrightarrow$ 7) and 2) $\Longrightarrow$ 1).
The implication 1) $\Longrightarrow$ 7) is trivial, as $\mathcal{N}^{\text{st}} \subset \mathcal{M}^{\text{st}}$ is just a particular case of a smooth representation.
If 7) is satisfied then by [Po13] we have $\|\Gamma_{N,M}\|^2 = [M : N]$ and N, M follow amenable (as ranges of norm one projections from the amenable von Neumann algebras $\mathcal{N}^{\text{st}}$, $\mathcal{M}^{\text{st}}$). Thus we have 7) $\Longrightarrow$ 2).

2) $\Longrightarrow$ 3). This is essentially (4.4.1 in [Po2]), or the proof of (4.1 in [Po4], up to Step VI on page 291), with some changes and additional considerations that we explain below.

Like in the proof of 1) $\Longrightarrow$ 2) in Theorem 6.1, we let F be an ε'-Følner set for $\Gamma_{N_1,N}$ (by 2) we have that $\Gamma_{N,M}$ is amenable, equivalently $\Gamma_{N_1,N}$ is amenable), then we choose a large n and some integers $m_k \leq (\dim N' \cap M_{2n+1} p_k^{n+1})^{1/2}$ such that

$$\left|\frac{m_k}{m_{k'}} - \frac{v_k}{v_{k'}}\right| < \delta \qquad \forall\ k, k' \in F, \tag{1}$$

where $\{p_k^{n+1}\}_k$ is now the list of minimal central projections in $N' \cap M_{2n+1}$ and $\vec{v} = (v_k)_{k \in K}$ is the standard vector of local indices (at even levels) for $\Gamma_{N_1,N}$. We then take $q_k \in \mathcal{P}(N' \cap M_{2n+1} p_k^{n+1})$ such that $\dim(q_k N' \cap M_{2n+1} q_k) = m_k^2$ and define $p = \sum_{k \in F} q_k$.

Let then $P_0 \subset N$ be a downward basic construction for $Np \subset pM_{2n+1}p$. By the choice of F (i.e., satisfying $\sum_{k \in F \cup \partial F} v_k^2 \leq (1+\varepsilon') \sum_{k \in F} v_k^2$) it follows that if $\{\bar{x}_j\}_j$ is an orthonormal basis of N over $P_0 \vee P_0' \cap N$ then $\{\bar{x}_j\}_j$ is almost an orthonormal basis of M over $P_0 \vee P_0' \cap M$ as well. Moreover, by the choice of integers $\{m_k\}_{k \in F}$ it follows that

$$\sum_j \tau(E_{P \vee P' \cap N}(\bar{x}_j^* \bar{x}_j) \bar{p}_{k_0}) / \tau(\bar{p}_{k_0}) \approx \sum_{k \in F} v_k^2, \qquad \forall\ k_0 \in F, \tag{2}$$

$\{\bar{p}_k\}_{k \in F}$ being the minimal central projections in $P_0' \cap N$. Also, since P_0 is a type II_1 factor, we may assume $E_{P_0 \vee P_0' \cap N}(\bar{x}_j^* \bar{x}_j) \bar{p}_k \in P_0 \bar{p}_k$, $\forall\ k \in F$. But then, by using first the approximate innerness of $Np \subset pM_{2n+1}p$ then the central freeness of $P_0 \subset M$, like in (Steps I, II, III in the proof of 4.1 in [Po4]), we obtain a conjugate of P_0 by a unitary element in N, say P_1, such that we have the type of estimates (a)–(f) on page 285 of [Po4] with P_1 instead of N_{m_0}. Then we go through Step IV on pages 286–288 of [Po4], noting that due to the condition (2) above, we don't need to take a further tunnel and that taking P_1 for N_m will do.

Then Step V on page 289 can be taken unchanged. Altogether, after doing all this we end up obtaining the following: $\forall\ x_1, \dots, x_l \in M$, $\forall\ \varepsilon > 0$, if $F \subset \mathrm{Even}(\Gamma_{N_1,N})$ is a ε-Følner set, n is sufficiently large and $\{m_k\}_{k \in F}$ satisfy (1) with δ suficiently small, then there exists a choice of a downward basic construction $P_1 \subset N$ for $Np \subset pM_{2n+1}p$, where $p = \sum_{k \in F} q_k$ as above, and a projection $s_0 \in P_1$, such that for all $1 \leq i \leq l$ we have

$$\begin{gathered} \|[s_0, x_i]\|_2 < f(\varepsilon')\|s_0\|_2, \\ \|s_0 x_i s_0 - E_{s_0(P_1 \vee P_1' \cap M)s_0}(s_0 x_i s_0)\|_2 < f(\varepsilon')\|s_0\|_2 \end{gathered} \tag{3}$$

where $f(\varepsilon') \to 0$ as $\varepsilon' \to 0$.

Arguing like in Step VI on page 290 of [Po4] we obtain a family of such choices of downward basic constructions $(P_i)_{i \in I}$ with projections $(s_i)_{i \in I}$, $s_i \in P_i$, such that (P_i, s_i) satisfy (3) and $\sum_i s_i = 1$. But then there exists a downward basic construction $P \subset N$ for $Np \subset pM_{2n+1}p$ such that $s_i \in P$, $\forall\ i$, and $s_i P s_i = s_i P_i s_i$, $\forall\ i$. Thus P will satisfy

$$\|E_{P \vee P' \cap M}(x_i) - x_i\|_2 < f(\varepsilon'), \qquad 1 \leq i \leq n.$$

Since P is hyperfinite, by taking ε' so that $f(\varepsilon') < \varepsilon$ we get 7.1.3).
3) $\Longrightarrow$ 4) is trivial, by simply taking $Q = Q_0' \cap P$ in 3).
4) $\Longrightarrow$ 5). since $\mathrm{Alg}(M, e_N, M^{\mathrm{op}})$ is so-dense in $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ it is sufficient to prove that $\forall\ x_1, x_2, \ldots, x_n \in M$, $\forall\ \varepsilon > 0$, $\exists\ B \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ finite dimensional such that

$$\|E_B(x_i) - x_i\|_2 < \varepsilon$$
$$\|E_B(e_N) - e_N\|_2 < \varepsilon$$
$$\|E_B(x_i^{\mathrm{op}}) - x_i^{\mathrm{op}}\|_2 < \varepsilon.$$

By 4) there exists $Q \subset N$ with $[N : Q] < \infty$ such that $x_i \in_\varepsilon Q' \cap M$. But

$$[Q^{\mathrm{op}\prime} \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}} : M^{\mathrm{op}\prime} \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}}] \leq [M^{\mathrm{op}} : Q^{\mathrm{op}}] < \infty$$

and thus

$$[Q^{\mathrm{op}\prime} \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}} : Q] \leq [M : Q]^2 < \infty,$$

implying that $B \overset{\mathrm{def}}{=} Q' \cap (Q^{\mathrm{op}\prime} \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}})$ has finite dimension. Since $e_N \in B$ and $Q' \cap M$, $Q^{\mathrm{op}\prime} \cap M^{\mathrm{op}} \subset B$, we are done.
5) $\Longrightarrow$ 6) is trivial, because hyperfinite algebras are amenable, so they have hypertraces.
6) $\Longrightarrow$ 7) By 5.2 we have $\mathcal{M}^{\mathrm{st}} = \mathrm{vN}(M, J_S M J_S) \subset \mathcal{B}(L^2(S))$, $\mathcal{N}^{\mathrm{st}} = \mathrm{vN}(N, J_S M J_S)$, where $S = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$. Let then $\Phi : \mathcal{B}(L^2(S)) \to S$ be a conditional expectation. Since Φ is S-S linear and $[\mathcal{M}^{\mathrm{st}}, M^{\mathrm{op}}] = 0$ it follows that $\Phi(\mathcal{M}^{\mathrm{st}}) \subset M^{\mathrm{op}\prime} \cap S = M$. Similarly, since $[\mathcal{M}^{\mathrm{st}}, M_1^{\mathrm{op}}] = 0$, we get $\Phi(\mathcal{N}^{\mathrm{st}}) \subset M_1^{\mathrm{op}\prime} \cap S = N$.
2 $\Longrightarrow$ 1). If

$$\begin{array}{ccc} \mathcal{N} & \subset & \mathcal{M} \\ \cup & & \cup \\ N & \subset & M \end{array}$$

is an arbitrary smooth representation of $N \subset M$ then, M being hyperfinite, it follows that there exists a conditional expectation of $\mathcal{M}$ onto M. By Theorem 5.7 it then follows that there exists a conditional expectation of $\mathcal{N} \overset{\mathcal{E}}{\subset} \mathcal{M}$ onto $N \subset M$. Q.E.D.

7.2. *Remarks.* 1°. Note that by using condition 7.1.3 one can easily proceed to construct recursively a sequence of appropriate downward basic constructions

for suitable local inclusions in the Jones tower-tunnel, say $M \supset N \supset P \supset P_1 \subset ...$, such that if we let $Q = \cap_n P_n$ and $(N^{0,\text{st}} \subset M^{0,\text{st}}) = (\overline{\cup_n P_n' \cap N} \subset \overline{\cup_n P_n' \cap M})$ then $(N \subset M) = (Q \bar{\otimes} N^{0,\text{st}} \subset Q \bar{\otimes} M^{0,\text{st}})$, with the isomorphism class of $N^{0,\text{st}} \subset M^{0,\text{st}}$ only depending on $\mathcal{G}_{N,M}$.
Indeed, from the proof of 7.1.2) $\Rightarrow$ 7.1.3) we see that, up to conjugacy by a unitary in N, the choice of the subfactor $P = P_1$ (equivalently, the choice of the projection $p \in N' \cap M_{2n+1}$) is determined by a choice of the ε-Følner set F and by a choice of the integers $\{m_k\}_{k \in F}$, which in turn both depend on ε. Similarily, each time one goes from step n to step $n+1$, one uses the Følner-type amenability condition for $P_n \subset M$ and some $\varepsilon = \varepsilon_{n+1}$ to get the next subfactor P_{n+1} (up to conjugation by a unitary element in P_n), from a downward basic construction that only depends on some choice of an ε_{n+1}-Følner set F_{n+1} and of some integers $\{m_j^{n+1}\}_{j \in F_{n+1}}$. Thus, if we let for instance $\varepsilon = 2^{-n}, \forall n$, and make the choice of the F_n's and m_j^n's this way, once for all, then the isomorphism class of $\{P_n' \cap N \subset P_n' \cap M\}_n$ will only depend on $\mathcal{G}_{N,M}$, as all the above choices can be "read" from this object through its amenability properties. In particular, the isomorphism class of $N^{0,\text{st}} \subset M^{0,\text{st}}$ will only depend on $\mathcal{G}_{N,M}$.
Thus, when complemented with this remark, we see that condition 7.1.3 in the above theorem shows that hyperfinite subfactors with amenable graphs are completely classified by their standard invariants (for more on this, see [Po16]).
2°. Recently, F. Hiai and M. Izumi have further investigated our notion of amenability for standard lattices and weighted graphs coming from subfactors and obtained two more equivalent characterizations ([HiIz]): the first one requires the existence of invariant means on the (weighted) fusion algebra of all $M-M$ bimodules in the Jones tower of $N \subset M$; the second one is a ratio limit condition on the weight vector $\vec{v}$, stating that the (weighted) graph $(\Gamma_{N,M}, \vec{v})$ is amenable if and only if for every vertex $k \in K$ one has

$$\lim_{n \to \infty} \frac{\langle (\Gamma\Gamma^t)^n \delta_*, \delta_k \rangle}{\langle (\Gamma\Gamma^t)^n \delta_*, \delta_* \rangle} = v_k,$$

where $\Gamma = \Gamma_{N,M}$. This "ratio limit" result for group-like objects coming from subfactors, which generalizes in a non-trivial way a prior result of Avez for discrete groups ([Av]), shows that in fact the projections $q_k \in (N' \cap M_{2n+1}) p_k^{n+1}$ in the proof of 2) $\Longrightarrow$ 3) of Theorem 7.1 can be taken equal to p_k^{n+1}. It also shows that the standard weight vector $\vec{v}$ of an amenable standard λ-lattice $\mathcal{G}$ can be completely recovered from its graph Γ.
It should be noted however that there exist no known examples of standard graphs Γ which admit two distinct standard weights, say $\vec{v}_1, \vec{v}_2$, for the same value of the index, i.e., such that $(\Gamma, \vec{v}_1) \not\simeq (\Gamma, \vec{v}_2)$. Whether such examples exist or not seems to be an interesting problem.
In order to prove the hereditarity result in its largest generality, namely without assuming that the inclusions involved are extremal, we'll need the following:

7.3. Lemma. *Let $N \subset M$ be an inclusion of type II_1 factors, of finite index (but not necessarily extremal). Let B be a C^*-algebra containing M, such that $B = C^*(M, \bigcup_k N_k' \cap B)$ and such that B has a state ϕ with $[\varphi, M] = 0$. Let $(\pi_\varphi, \mathcal{H}_\varphi, \xi_\varphi)$ be the GNS representation for (B, φ). Then, as a M-M Hilbert bimodule, $\mathcal{H}_\varphi$ is a direct sum of irreducible bimodules $\mathcal{H}_\varphi = \bigoplus_j \mathcal{H}_j'$ with each $\mathcal{H}_j'$ isomorphic to a bimodule in the list $\{\mathcal{H}_k\}_{k \in K}$.*

Proof. Note first that if $\{\mathcal{K}_l^{\mathrm{op}}\}_{l \in L}$ denotes the list of all irreducible M-N bimodules contained in $\bigoplus_{k \in K} {}_M\mathcal{H}_{kN}$ (see the beginning of Section 5) and $\mathcal{H}_0' \simeq \bigoplus_j \mathcal{K}_{l_j}^{\mathrm{op}}$ is a M-N Hilbert bimodule contained in some $M-M$ bimodule $\mathcal{H}$, (i.e., $\mathcal{H}_0' \subset {}_M\mathcal{H}_N$) then $\overline{\mathrm{sp}} M \mathcal{H}_0' M \simeq \bigoplus_i \mathcal{H}_{k_i}$.
Then note that $\cup_k \mathrm{sp} M(N_k' \cap B)\xi_\varphi$ is dense in $\mathcal{H}_\varphi$. Indeed, we have

$$\begin{aligned}\mathrm{sp}M(N_k' \cap B)M(N_k' \cap B)M &= \mathrm{sp}M(N_k' \cap B)N_k f_{-k-1}^0 N_k \\ (N_k' \cap B)M &= \mathrm{sp}M(N_k' \cap B) f_{-k-1}^0 (N_k' \cap B)M \\ &\subset \mathrm{sp}M(N_{2k+1}' \cap B)M.\end{aligned}$$

showing that

$$\mathrm{Alg}(M, \cup_k N_k' \cap B) = \cup_k \mathrm{sp}M(N_k' \cap B)M = \cup_k \mathrm{sp}M(N_k' \cap B)R,$$

where $R = \overline{\cup_k N_k' \cap M}$, the closure being taken in the norm $\|\cdot\|_2$ in M. But $\cup_k (N_k' \cap M)\xi_\varphi$ is dense in $R\xi_\varphi$ (because φ implements τ on M), so $\cup_k \mathrm{sp}M(N_k' \cap B)\xi_\varphi$ is dense in $\cup_k \mathrm{sp}M(N_k' \cap B)R\xi_\varphi = \cup_k \mathrm{sp}M(N_k' \cap B)M\xi_\varphi$ which is dense in $\mathcal{H}_\varphi$.
Let then $\mathcal{H}_\varphi' \overset{\mathrm{def}}{=} \vee\{\mathcal{H}' \subset \mathcal{H}_\varphi \mid \exists\, k \in K \text{ such that } \mathcal{H}' \simeq \mathcal{H}_k \text{ as } M\text{-}M \text{ bimodules}\}$. Assume $\mathcal{H}_\varphi' \neq \mathcal{H}_\varphi$. Thus, there exists $\xi \in M(N_k' \cap B)\xi_\varphi$ such that $\xi \notin \mathcal{H}_\varphi'$. Let $\xi = X_0 Y_0' \xi_\varphi$, for some $X_0 \in M$, $Y_0' \in N_k' \cap B$. It follows that if $x \in M$, $y \in N_k$ then

$$\langle x\xi y, \xi\rangle = \langle xX_0 Y_0'\xi_\varphi y, X_0 Y_0' \xi_\varphi\rangle = \langle X_0^* x X_0 y Y_0' \xi_\varphi, Y_0'\xi_\varphi\rangle.$$

But the state on M defined by $\psi(X) = \langle XY_0'\xi_\varphi, Y_0'\xi_\varphi\rangle$, $X \in M$ has N_k in its centralizer so by A.1 it is automatically normal and of the form $\psi(X) = \tau(Xa)$ for some $a \in N_k' \cap M$. Thus we get

$$\langle x\xi y, \xi\rangle = \tau(X_0^* x X_0 y a) = \langle x(X_0 a^{1/2}\xi_\tau)y, (X_0 a^{1/2}\xi_\tau)\rangle$$

so if we define $\xi' = X_0 a^{1/2}\xi_\tau \in M\xi_\tau \subset L^2(M)$ then the above shows that $\mathcal{H}_0' = \overline{\mathrm{sp}} M\xi N_k$ is a M-N_k bimodule isomorphic to a sub-bimodule of ${}_M L^2(M)_{N_k}$. By the first part applied to $N = N_k$ it follows that $\overline{\mathrm{sp}} M\mathcal{H}_0' M$ is a sub-bimodule of $(\bigoplus_{k \in K} \mathcal{H}_k)^n$ for some multiplicity $n \leq \infty$, giving a contradiction. Q.E.D.

7.4. Corollary. *Let $N \subset M$ be an inclusion of type II_1 factors with finite index. Assume that for any $\varepsilon > 0$ there exists an amenable type II_1 von Neumann algebra B containing M such that*

$$\|E_{(N'\cap B)'\cap B}(e_0) - \lambda 1\|_2 < \varepsilon.$$

Then there exists a norm one projection from $\mathcal{N}^{\mathrm{st}} \subset \mathcal{M}^{\mathrm{st}}$ onto $N \subset M$.

Proof. If $B \supset M$ satisfies the condition in the hypothesis for some ε, then there exist some finite many unitary elements $u_1, \ldots, u_n \in N' \cap B$ such that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} u_i e_0 u_i^* - \lambda 1 \right\|_2 < \varepsilon.$$

Thus, by taking instead of B the von Neumann algebra generated by M and $\{u_1, \ldots, u_n\}$, it follows that we may assume B is separable in the norm $\| \ \|_2$. Let then $\mathcal{H}$ be the M-M Hilbert bimodule obtained by summing up countably many copies of each $\mathcal{H}_k$, $k \in K$. By 7.3 we have $L^2(B) \subset \mathcal{H}$, for each B as in the hypothesis, where $L^2(B)$ has the M-M bimodule structure given by left-right multiplication by elements of M.
For each $\varepsilon = 1/n$ we choose an algebra B_n satisfying the hypothesis. We let $\Phi_n : \mathcal{B}(L^2(B_n)) \to B_n$ be norm one projections and define the state φ on $\mathcal{B}(\mathcal{H})$ by

$$\varphi(T) = \lim_{n\to\omega} \tau_n \circ \Phi_n(p_n T|_{L^2(B_n)})$$

where $p_n = \mathrm{proj}_{L^2(B_n)}$, τ_n is the trace on B_n and ω is a free ultrafilter on $\mathbb{N}$. Since each $\tau_n \circ \Phi_n$ is a M-hypertrace, φ follows a M-hypertrace. Moreover, if we identify $\mathcal{M}^{\mathrm{st}}$ with the von Neumann algebra generated in $\mathcal{B}(\mathcal{H})$ by the operators of left and right multiplication by M and $\mathcal{N}^{\mathrm{st}}$ with its von Neumann subalgebra generated by the operators of left multiplication by N and right multiplication by M, then $\mathcal{M}^{\mathrm{st}} = \mathrm{sp}\mathcal{N}^{\mathrm{st}} e_0 N$. Let $Y \in \mathcal{N}^{\mathrm{st}}$, $y \in N$. We want to show that $\varphi = \varphi \circ \mathcal{E}^{\mathrm{st}}$ on $\mathcal{M}^{\mathrm{st}}$, thus we need to show that $\varphi(Ye_0y) = \lambda\varphi(Yy)$. But $[p_n, \mathcal{M}^{\mathrm{st}}] = 0$ and $[\mathcal{N}^{\mathrm{st}} p_n, N' \cap B_n] = 0$, so that $\Phi_n(\mathcal{N}^{\mathrm{st}} p_n) \subset (N' \cap B_n)' \cap B_n$. Thus $\Phi_n((Ye_0y)p_n) = \Phi_n(Yp_n)e_0y = y'e_0y$, with $y' \in (N' \cap B_n)' \cap B_n$. Thus $\tau_n(y'e_0y) = \tau_n(E_{(N'\cap B_n)'\cap B_n}(y'e_0y)) = \tau_n(y'E_{(N'\cap B_n)'\cap B_n}(e_0)y)$. It follows that

$$\begin{aligned}
|\tau_n \circ \Phi_n((Ye_0y)p_n) - &\lambda\tau_n \circ \Phi_n((Yy)p_n)| \\
&= |\tau_n(y'E_{(N'\cap B_n)'\cap B_n}(e_0)y) - \lambda\tau_n(y'y)| \\
&\le \|y'\| \, \|y\| \, \|E_{(N'\cap B_n)\cap B_n}(e_0) - \lambda 1\|_2 \\
&\le \frac{1}{n}\|Y\| \, \|y\|.
\end{aligned}$$

This proves that indeed $\varphi(Ye_0y) = \lambda\varphi(Yy)$, so $\varphi = \varphi \circ \mathcal{E}^{\mathrm{st}}$ on $\mathcal{M}^{\mathrm{st}}$. Q.E.D.

We can now prove the announced hereditarity property for amenable inclusions.

7.5. Theorem. *Let $N \subset M$ be an extremal inclusion of hyperfinite type II_1 factors with amenable graph (equivalently, with amenable standard invariant $\mathcal{G}_{N,M}$), i.e., $\|\Gamma_{N,M}\|^2 = [M : N]$. Assume $Q \subset P$ is an inclusion of factors embedded in $N \subset M$ as commuting squares (i.e., such that $E_N(P) = Q$), but without necessarily being extremal and not necessarily having the same index as $N \subset M$. Then $\Gamma_{Q,P}$ is amenable (equivalently, $\mathcal{G}_{Q,P}$ is amenable), i.e., $\|\Gamma_{Q,P}\|^2$ equals the minimal index of $Q \subset P$.*

Proof. By Theorem 7.1, $S = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ follows amenable so in particular the von Neumann algebra B generated in S by P and $Q' \cap S$ is also amenable. Let $e_0 \in P$ be a Jones projection for $Q \subset P$. Thus, $E_Q(e_0) = E_N(e_0) = \lambda_0 1 = [P : Q]^{-1}1$. Since $(N' \cap S)' \cap S = N$ it follows that $E_{(N'\cap S)'\cap S}(e_0) = \lambda_0 1$. Since $N' \cap S \subset Q' \cap S = Q' \cap B$, this implies that $E_{(Q'\cap B)'\cap B}(e_0) = \lambda_0 1$ as well.
Thus $Q \subset P$ satisfies the conditions in the hypothesis of 7.4, so there exists a norm one projection from the standard representation $\mathcal{Q}^{\mathrm{st}} \subset \mathcal{P}^{\mathrm{st}}$ onto $Q \subset P$. By [Po13] this implies $\|\Gamma_{Q,P}\|^2 = \|T_{\mathcal{Q}^{\mathrm{st}}\subset\mathcal{P}^{\mathrm{st}}}\|^2 = \mathrm{Ind}E_{\min}^{Q,P}$. Q.E.D.

In Sec. 6 we've seen that for extremal inclusions of arbitrary type II_1 factors $N \subset M$ the condition $\|\Gamma^{u,f}_{N,M}\|^2 = [M : N]$ is sufficient to insure the amenability of the standard graph $\Gamma_{N,M}$. We now show that for inclusions of hyperfinite factors the weaker condition $\|\Gamma^{u,rf}_{N,M}\|^2 = [M : N]$ is enough, where $\Gamma^{u,rf}_{N,M}$ denotes the inclusion graph of the direct summand $\mathcal{N}^{u,rf} \subset \mathcal{M}^{u,rf}$ of $\mathcal{N}^u \subset \mathcal{M}^u$, in which $\mathcal{M}^{u,rf}$ consists of all irreducible representations $\mathcal{B}(\mathcal{H})$ of $M \otimes M^{\mathrm{op}}$, with $\mathcal{H}$ having finite right dimension over M, i.e., $\dim(\mathcal{H}_M) < \infty$, but leaving the left dimensions $\dim({}_M\mathcal{H})$ arbitrary.

7.6. Theorem. *Let $N \subset M$ be an extremal inclusion of hyperfinite type* II_1 *factors. The following conditions are equivalent:*
1) *$N \subset M$ has amenable graph, i.e., $\|\Gamma_{N,M}\|^2 = [M : N]$.*
2) *$\forall\ \varepsilon > 0$, $\exists\ P$ a hyperfinite factor containing M, such that* $\dim M' \cap P < \infty$ *and $\|T_{M'\cap P\subset N'\cap P}\|^2 \geq [M : N] - \varepsilon$.*
3) *$\|\Gamma^{u,rf}_{N,M}\|^2 = [M : N]$.*

Proof. 1) $\Longrightarrow$ 3) is trivial because $\Gamma^{u,rf}_{N,M} \supset \Gamma_{N,M}$.
3) $\Longrightarrow$ 2) By 3) there exists a direct summand $\mathcal{N} \subset \mathcal{M} = \oplus_{l\in L'}\mathcal{B}(\mathcal{K}'_l) \subset \oplus_{k\in K'}\mathcal{B}(\mathcal{H}'_k)$ of $\mathcal{N}^{u,rf} \subset \mathcal{M}^{u,rf}$ such that its inclusion graph Γ is connected and $\|\Gamma\|^2 > [M : N] - \varepsilon$. Take $K'_0 \subset K'$ finite and sufficiently large so that we still have $\|\Gamma^t_{K'_0}\|^2 > [M : N] - \varepsilon$.
By the definition of the universal representation $\mathcal{N}^{u,rf} \subset \mathcal{M}^{u,rf}$, if $Q = M' \cap \mathcal{N}$ then Q is a factor of type II_1, $\mathcal{N} = N \vee Q \subset M \vee Q = \mathcal{M}$ and Q has finite coupling constant in each direct summand $\mathcal{B}(\mathcal{H}'_k)$ of $\mathcal{M}$. But then, if one takes $P = Q' \cap \mathcal{B}(\oplus_{k\in K'_0}\mathcal{H}'_k)$ then $\|T_{M'\cap P\subset N'\cap P}\|^2 \geq \|\Gamma^t_{K'_0}\|^2 \geq [M : N] - \varepsilon$.
2) $\Longrightarrow$ 3) follows by noticing that $\mathcal{M}^{u,rf}$ contains the von Neumann algebra generated by the operators of left multiplication by M and right multiplication

by P on $L^2(P)$ as a direct summand (by taking $M^{\rm op} \simeq P^{\rm op}$, both being hyperfinite factors). Thus $T_{M'\cap P\subset N'\cap P}$ will be a restriction of the graph $\Gamma^{u,rf}_{N,M}$
2) $\Longrightarrow$ 1). Let $\varepsilon > 0$ and choose P a hyperfinite II$_1$ factor satisfying 2) for ε^{32}. Denote by T the bipartite graph describing the inclusions $M' \cap P \subset N' \cap P \subset N_1' \cap P \subset N_2' \cap P \subset \cdots$. Thus $\|T\|^2 \geq [M : N] - \varepsilon^{32}$ and there is a positive vector $w = (w_j)_{j\in J}$ such that $TT^{\rm t}w = \lambda^{-1}w$, giving the traces on $\{N_k' \cap P\}_{k\geq 1}$. But then A.2 applies the same way as in the proof of 1) $\Longrightarrow$ 2) in Theorem 5.3 to get a finite set $F \subset J$ such that $\sum_{j\in\partial F} w_j^2 < \varepsilon \sum_{j\in F} w_j^2$ (see 5.5.2° and [Po14]). Arguing like in the proof of 1) $\Longrightarrow$ 2) in Theorem 6.1 it then follows that there exist $k \geq 1$ and a projection $p \in N_{2k}' \cap P$ such that

$$\|E_{(pN'_{2k+1}\cap Pp)'\cap pPp}(e_{-2k-1}p) - \lambda p\|_2 < \varepsilon\|p\|_2.$$

By taking an $\alpha = [M : N]^{k+1}$-amplification of the inclusion $N_{2k+2} \subset N_{2k+1} \hookrightarrow pPp$ and using that $(N_{2k+2} \subset N_{2k+1})^\alpha = (N \subset M)$ it follows that there exists a hyperfinite type II$_1$ factor $P_0 \simeq (pPp)^\alpha$ such that $N \subset M \subset P_0$ and

$$\|E_{(N'\cap P_0)'\cap P_0}(e_0) - \lambda 1\|_2 < \varepsilon.$$

But then 7.3 applies to get that $\Gamma_{N,M}$ is amenable. Q.E.D

8. An Effros-Lance Type Characterization of Amenability

We will prove in this section yet another equivalent characterization for the amenability of a subfactor $N \subset M$, in terms of simplicity properties of the C^*-algebra $C^*_{\rm bin}(M, e_N, M^{\rm op})$. In the case $N = M$ our result reduces to the implication "$C^*_{\rm bin}(M, M^{\rm op})$ simple $\Longrightarrow \exists$ conditional expectations of $\mathcal{B}(L^2(M))$ onto M", which is one of the well known results of Effros and Lance in [EL], relating various amenability conditions for single von Neumann algebras (semidiscreteness, injectivity, etc).

8.1. Theorem. *Let $N \subset M$ be an extremal inclusion of type* II$_1$ *factors. The following conditions are equivalent:*
1°. *$N \subset M$ is amenable.*
2°. *$C^*_{\rm bin}(M, e_N, M^{\rm op})$ is simple.*
3°. *$C^*(M, e_N, JMJ)$ is simple.*
4°. *$C^*_{\rm bin}(M, e_N, M^{\rm op}) \simeq C^*(M, e_N, JMJ) \simeq C^*_{\rm min}(M, e_N, M^{\rm op})$, with the isomorphisms being given by the natural quotient maps.*

Proof. 1° $\Longrightarrow$ 2°. Let $C^*_{\rm bin}(M, e_N, M^{\rm op}) \hookrightarrow \mathcal{B}(\mathcal{H})$ be a faithful representation of $C^*_{\rm bin}(M, e_N, M^{\rm op})$ such that M and $M^{\rm op}$ are von Neumann algebras in $\mathcal{B}(\mathcal{H})$. It is sufficient to prove that if

$$x \in \mathrm{Alg}(M, e_N, M^{\rm op}) = \bigcup_k \mathrm{sp} M^{\rm op} M_k M^{\rm op} \subset C^*_{\rm bin}(M, e_N, M^{\rm op}) \subset \mathcal{B}(\mathcal{H}),$$

then $\|x\|_{\mathcal{B}(\mathcal{H})} \leq \|x\|_{\min}$, where $\|x\|_{\min}$ is the norm of (the image of) x in $C^*_{\min}(M, e_N, M^{\text{op}})$. For such $x \in \text{Alg}(M, e_N, M^{\text{op}})$ let k be large enough such that

$$x = \sum_{i=n}^{n} y_i^{\text{op}} z_i x_i^{\text{op}} \in \text{sp} M^{\text{op}} M_k M^{\text{op}},$$

for some $x_i, y_i \in M, z_i \in M^{\text{op}}$, $1 \leq i \leq n$. We will prove that x can be approximated in the so topology on $\mathcal{B}(\mathcal{H})$ by elements $x' \in \text{Alg}(M, e_N, M^{\text{op}})$ such that $\|x'\|_{\mathcal{B}(\mathcal{H})} \leq \|x\|_{\min}$. By the inferior semicontinuity of the norm $\| \ \|_{\mathcal{B}(\mathcal{H})}$ with respect to the so-topology on $\mathcal{B}(\mathcal{H})$, this will show that $\|x\|_{\mathcal{B}(\mathcal{H})} \leq \|x\|_{\min}$ and will thus end the proof of $1^\circ \implies 2^\circ$.
To prove this approximation, let us first note that $\forall\ \xi_1, \dots, \xi_p \in \mathcal{H}$, $\forall\ \varepsilon > 0$, $\exists\ \delta > 0$ such that if $z_i' \in M_k$, with $\|z_i' - z_i\|_2 < \delta$, $\|z_i'\| \leq \|z_i\|$ then

$$x' \stackrel{\text{def}}{=} \sum_{i=1}^{n} y_i^{\text{op}} z_i' x_i^{\text{op}}$$

satisfies $\|(x - x')\xi_j\| < \varepsilon$, $\forall\ 1 \leq j \leq p$. Indeed we have:

$$\|(x - x')\xi_j\| \leq \sum_{i=1}^{n} \|y_i^{\text{op}}\| \ \|(z_i - z_i')x_i^{\text{op}}\xi_j\|$$

and since the so-topology on the ball of radius $\|z_i\|$ (in the uniform norm) in M_k coincides with the topology given by the norm $\| \ \|_2$ on this ball, it follows that there exists $\delta > 0$ such that if $\|z_i - z_i'\|_2 < \delta$ then $\|(z_i - z_i')x_i^{\text{op}}\xi_j\| < \varepsilon/n\|y_i\|$, $\forall\ i$. But then we have

$$\|(x - x')\xi_j\| < \sum_{i=1}^{n} \|y_i^{\text{op}}\| \varepsilon/n \|y_i^{\text{op}}\| = \varepsilon.$$

Now, if we assume $N \subset M$ is amenable then $M \subset M_k$ follows amenable and by [Po2] we get that $\forall\ \delta > 0$, $\exists$ finitely many tunnels $\{N_k^r\}_{1 \leq k \leq n_r}$, $r = 1, \dots, m$, and projections $p_r \in \mathcal{P}({N_{n_r}^r}' \cap M)$, $r = 1, \dots, m$, such that $\{p_r\}_r$ are mutually orthogonal, $\Sigma_r p_r = 1$ and

$$z_i' \stackrel{\text{def}}{=} \sum_{r=1}^{m} p_r E^{M_k}_{{N^r}'_{n_r} \cap M_k}(z_i) p_r$$

satisfies $\|z_i' - z_i\|_2 < \delta$. Also, by its definition, z_i' checks $\|z_i'\| \leq \|z_i\|$. Furthermore, since $p_r \in M$ commute with $x_i^{\text{op}}, y_i^{\text{op}} \in M^{\text{op}}$, $\forall\ r, i$, it follows that if we let $x' = \Sigma_i y_i^{\text{op}} z_i' x_i^{\text{op}}$ as above, then $\|x'\|_{\mathcal{B}(\mathcal{H})} = \sup_r \|x' p_r\|_{\mathcal{B}(\mathcal{H})}$. But since

$$x' p_r = \sum_i y_i^{\text{op}} p_r E_{{N_{n_r}^r}' \cap M_k}(z_i) p_r x_i^{\text{op}} = p_r \left(\sum_{i=1}^{n} y_i^{\text{op}} E_{{N_{n_r}^r}' \cap M_k}(z_i) x_i^{\text{op}} \right) p_r,$$

each $\|x'p_r\|_{\mathcal{B}(\mathcal{H})}$ is majorized by

$$\begin{aligned}\left\|\sum_{i=1}^{n} y_i^{\mathrm{op}} E_{N^r_{n_r}{}'\cap M_k}(z_i)x_i^{\mathrm{op}}\right\|_{\mathcal{B}(\mathcal{H})} &= \left\|\sum_{i=1}^{n} y_i^{\mathrm{op}} E_{N^r_{n_r}{}'\cap M_k}(z_i)x_i^{\mathrm{op}}\right\|_{M^{\mathrm{op}}_{n_r}} \\ &= \left\|\sum_{i=1}^{n} y_i^{\mathrm{op}} E_{N^r_{n_r}{}'\cap M_k}(z_i)x_i^{\mathrm{op}}\right\|_{C^*_{\min}(M,e_N,M^{\mathrm{op}})} \\ &\leq \|x\|_{\min},\end{aligned}$$

with the last inequality following from the fact that, in the algebra $C^*_{\min}(M,e_N,M^{\mathrm{op}})$, the element $\Sigma_i y_i^{\mathrm{op}} E_{N^r_{n_r}{}'\cap M_k}(z_i)x_i^{\mathrm{op}}$ is the image of x under a conditional expectation.
Thus, from the the above remarks, if we take δ sufficiently small, we are done.
$2^\circ \implies 3^\circ$ is trivial, since by the definition of $C^*_{\mathrm{bin}}(M,e_N,M^{\mathrm{op}})$, $C^*(M,e_N,JMJ)$ is its quotient.
$3^\circ \implies 1^\circ$. If $C^*(M,e_N,JMJ)$ is simple then there exists an isomorphism

$$\varphi : S^0 \stackrel{\mathrm{def}}{=} C^*_{\min}(M,e_N,M^{\mathrm{op}}) \simeq C^*(M,e_N,JMJ) \subset \mathcal{B}(L^2(M)).$$

Since $S^0 \subset S \subset \mathcal{B}(L^2(S))$, where $S = M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ as usual, by Arveson's theorem φ can be extended to a completely positive map Φ from all $\mathcal{B}(L^2(S))$ to $\mathcal{B}(L^2(M))$. (Note that in fact we only use here a particular case of Arveson's theorem stating that if $B \subset A$ are unital C^*-algebras and $\pi_0 : B \to \mathcal{B}(\mathcal{H}_0)$ is a representation of B then there exists a Hilbert space $\mathcal{H} \supset \mathcal{H}_0$ and a representation $\pi : A \to \mathcal{B}(\mathcal{H})$ such that $\pi_0(b) = \mathrm{proj}_{\mathcal{H}_0}\pi(b)|_{\mathcal{H}_0}$, $\forall\, b \in B$. See 2.10.2 in [D2]). Since Φ is a unital $*$-homomorphism when restricted to S^0, it follows that it is a S^0-S^0 bimodule map. In particular, if $x_{1,2}^{\mathrm{op}} \in M^{\mathrm{op}} \subset S^0$ ($\subset \mathcal{B}(L^2(S))$) then $\Phi(x_1^{\mathrm{op}} T x_2^{\mathrm{op}}) = \varphi(x_1^{\mathrm{op}})\Phi(T)\varphi(x_2^{\mathrm{op}})$, $\forall\, T \in \mathcal{B}(L^2(S))$. Thus, if T satisfies $Tx^{\mathrm{op}} - x^{\mathrm{op}}T = 0$, $\forall\, x^{\mathrm{op}} \in M^{\mathrm{op}}$, then $\Phi(T)\varphi(x^{\mathrm{op}}) = \varphi(x^{\mathrm{op}})\Phi(T)$, $\forall\, x^{\mathrm{op}} \in M^{\mathrm{op}}$.
Thus we have $\Phi((M^{\mathrm{op}})' \cap \mathcal{B}(L^2(S)) = \varphi(M^{\mathrm{op}})' \cap \mathcal{B}(L^2(M))$. But $\varphi(M^{\mathrm{op}}) = JMJ$ and $JMJ' \cap \mathcal{B}(L^2(M)) = M$, so that $\Phi((M^{\mathrm{op}})' \cap \mathcal{B}(L^2(S))) = M$. Similarily we get

$$\begin{aligned}\Phi((M_1^{\mathrm{op}})' \cap \mathcal{B}(L^2(S))) &= \varphi(M_1^{\mathrm{op}})' \cap \mathcal{B}(L^2(M)) \\ &= JM_1J' \cap \mathcal{B}(L^2(M)) \\ &= N.\end{aligned}$$

But from 5.3 we have that $\mathcal{M}^{\mathrm{st}} \subset (M^{\mathrm{op}})' \cap \mathcal{B}(L^2(S))$ and $\mathcal{N}^{\mathrm{st}} \subset (M_1^{\mathrm{op}})' \cap \mathcal{B}(L^2(S))$, so Φ implements a positive unital M-M bimodule map from $\mathcal{M}^{\mathrm{st}}$ onto M carrying $\mathcal{N}^{\mathrm{st}}$ onto N. This shows that there exists a conditional expectation of $(\mathcal{N}^{\mathrm{st}} \subset \mathcal{M}^{\mathrm{st}})$ onto $(N \subset M)$, so $N \subset M$ follows amenable.

All this shows that the conditions 1°–3° are equivalent. Since clearly $4^\circ \Longleftrightarrow 2^\circ$, all the conditions 1°–4° follow equivalent. Q.E.D.

8.2. *Remarks.* 1°. Note that when applied to the case $N = M$ the above proof of the implication $3^\circ \Rightarrow 1^\circ$ in Theorem 8.1 reduces to a very short and elementary proof to one of the results in ([EL]).
2°. Recall from ([Bi3]) that $C^*(M, e_N, JMJ)$ contains the ideal $\mathcal{K}$ of compact operators over the Hilbert space $L^2(M, \tau)$ if and only if N contains no non-trivial central sequences of M. Thus, 8.1 implies that amenable inclusions always have non-trivial cental sequences contained in the subfactor (because if $C^*(M, e_N, JMJ)$ is simple then it cannot contain the ideal $\mathcal{K}$). In fact, 7.1.4 shows that there even exist non-commuting such central sequences, so that amenable inclusions split off the hyperfinite type II_1 factor (this is, of course, a consequence of the classification result 7.2.1° as well).

9. Property T for Subfactors and Standard Lattices

In this section we introduce a notion of property T for standard λ-lattices $\mathcal{G}$ (or, equivalently, for paragroups). When restricted to the class of standard lattices associated with subfactors coming from finitely generated discrete groups, our notion coincides with the classical property T of Kazhdan, which it thus generalizes, from discrete groups to the larger class of group-like objects $\mathcal{G}$. In order to define this notion, we will use a strategy similar to the approach to amenability in Section 5. Thus, the property T for a standard λ-lattice $\mathcal{G}$ will be defined by requiring $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ to have the property T relative to $M \vee M^{\mathrm{op}}$ in the sense of ([A-D], [Po8]), where $N \subset M$ is an extremal subfactor with $\mathcal{G}_{N,M} = \mathcal{G}$. This definition however depends on proving that such a property does not in fact depend on the subfactor $N \subset M$ one takes. We do prove this in the next few lemmas.
First of all, let us recall the definition of the relative property T, as introduced in ([A-D], [Po8]):

$(*)$. *Let U be a type II_1 factor and $B \subset U$ a von Neumann subalgebra of U. Then we say that U has the property* T *relative to B if there exists $\varepsilon > 0$ and $x_1, x_2, ..., x_n \in U$ such that whenever $\mathcal{H}$ is a given $U-U$ bimodule with a vector $\xi \in \mathcal{H}$ satisfying $\|\xi\| = 1, [\xi, B] = 0, \|[\xi, x_i]\| < \varepsilon$, it follows that $\mathcal{H}$ must contain a non-zero vector ξ_0 satisfying $[\xi_0, U] = 0$.*

Note that in the case $B = \mathbb{C}$ the above definition reduces to Connes' definition of property T for single type II_1 factors U. In general though, the definition $(*)$ does not require the ambient algebra U to have the property T. Instead, note that by ([A-D], [Po8]), if V is a type II_1 factor and G is a discrete group acting outerly on V, then $U = V \rtimes G$ has the property T relative to V if and only if the group G has Kazhdan's property T.
With this in mind, let us proceed with some technical results.

9.1. Lemma. *Let $V \subset U$ be an inclusion of type* II_1 *factors with $V' \cap U = \mathbb{C}1$. Then U has the property* T *relative to V if and only if $\forall\ \varepsilon > 0\ \exists\ \delta > 0$*

and $x_1, \dots, x_n \in U$ such that if $\varphi : U \to U$ is completely positive, unital, trace preserving, with $\varphi(v_1 x v_2) = v_1 \varphi(x) v_2$, $\forall\, v_1, v_2 \in V$, $x \in U$, and $\|\varphi(x_i) - x_i\|_2 < \delta$ then $\|\varphi(x) - x\|_2 < \varepsilon$, $\forall\, x \in U$, $\|x\| \leq 1$.

Proof. If U has the property T relative to V then the condition on completely positive maps holds true by 4.1.4 in [Po8].

Conversely, if this latter condition holds, then let $\mathcal{H}$ be a U-U bimodule with $\xi \in \mathcal{H}$, $\|\xi\| = 1$, $v\xi = \xi v$, $\forall\, v \in V$, $\|x_i \xi - \xi x_i\| < \delta' \stackrel{\text{def}}{=} \delta / 2 \sum \|x_i\|_2$. Let $\varphi : U \to U$ be defined by $\tau(y \varphi(x)) = \langle x \xi y, \xi \rangle$, $x, y \in V$ as in [C4] (see [Po8]). Then φ is a well defined completely positive map and $\tau(\varphi(x)) = \langle x\xi, \xi \rangle$. Since $\xi = v \xi v^*$, $\forall\, v \in \mathcal{U}(V)$, one gets

$$\begin{aligned} \langle v x v^* \xi, \xi \rangle &= \langle v x \xi v^*, \xi \rangle = \langle v x \xi, \xi v \rangle = \langle v x \xi, v \xi \rangle \\ &= \langle v^* v x \xi, \xi \rangle = \langle x \xi, \xi \rangle \end{aligned}$$

for all $x \in U$. Averaging by unitaries $v \in \mathcal{U}(V)$ like in [Po1] and using that $V' \cap U = \mathbb{C}1$ and $E_{V' \cap U}(x) = \tau(x)1$, it follows that

$$\langle x \xi, \xi \rangle = \langle \tau(x) \xi, \xi \rangle = \tau(x), \qquad \forall\, x \in U.$$

Similarly, we obtain that

$$\tau(y \varphi(1)) = \langle \xi y, \xi \rangle = \tau(y), \qquad \forall\, y \in U.$$

Thus $\varphi(1) = 1$. Also, if $x, y \in U$, $v_1, v_2 \in V$ then

$$\begin{aligned} \tau(y \varphi(v_1 x v_2)) &= \langle v_1 x v_2 \xi y, \xi \rangle = \langle x \xi v_2 y, v_1^* \xi \rangle \langle x \xi v_2 y, \xi v_1^* \rangle \\ &= \langle x \xi v_2 y v_1, \xi \rangle = \tau(v_2 y v_1 \varphi(x)) = \tau(y(v_1 \varphi(x) v_2)). \end{aligned}$$

This shows that $\varphi(v_1 x v_2) = v_1 \varphi(x) v_2$.

Finally, since $\|x_i \xi - \xi x_i\|_2 < \delta'$ we have

$$\begin{aligned} \|\varphi(x_i) - x_i\|_2^2 &= \tau(\varphi(x_i^*)\varphi(x_i)) = \tau(x_i^* x_i) - 2\mathrm{Re}\, \tau(x_i^* \varphi(x_i)) \\ &\leq \tau(\varphi(x_i^* x_i)) + \tau(x_i^* x_i) - 2\mathrm{Re}\, \tau(x_i^* \varphi(x_i)) \\ &= 2\tau(x_i^* x_i) - 2\mathrm{Re}\, \tau(x_i^* \varphi(x_i)) \\ &= 2\langle \xi x_i, \xi x_i \rangle - 2\mathrm{Re} \langle x_i \xi, \xi x_i \rangle \\ &\leq 2\|x_i \xi - \xi x_i\| \, \|\xi x_i\| \leq 2\delta' \|x_i\|_2 < \delta^2. \end{aligned}$$

Thus, φ this way defined satisfies the required condition, so $\|\varphi(x) - x\|_2 < \varepsilon$, $\forall\, x \in U$, $\|x\| \leq 1$. In particular, we have

$$\|\varphi(u) - u\|_2 < \varepsilon, \qquad \forall\, u \in \mathcal{U}(U).$$

Thus,

$$\begin{aligned} \|\xi u - u \xi\|^2 &= 2 - 2\mathrm{Re}\langle \xi u, \xi u \rangle = 2 - 2\mathrm{Re}\langle u^* \xi u, \xi \rangle \\ &= 2 - 2\mathrm{Re}\, \tau(\varphi(u) u^*) = 2\mathrm{Re}(\tau((u - \varphi(u)) u^*)) \\ &\leq 2\|\varphi(u) - u\|_2 \leq 2\varepsilon. \end{aligned}$$

Thus, if $\varepsilon < 1/2$ then $\|u^* \xi u - \xi\| < 1$, $\forall\, u \in \mathcal{U}(U)$. But then $\exists\, \xi_0 \in \mathcal{H}$, $\|\xi_0 - \xi\| < 1$, such that $u \xi_0 = \xi_0 u$, $\forall\, u$ (see e.g., [Po1]). Thus $\mathcal{H}$ has a nonzero vector commuting with U, showing that U has the property T relative to V Q.E.D.

9.2. Lemma. *Let*

$$\begin{array}{ccc} V & \subset & U \\ \cup & & \cup \\ Q & \subset & P \end{array}$$

be a nondegenerate commuting square of type II_1 *von Neumann algebras with a countable set* $\mathcal{X} = \{f_n\}_n \subset P$ *such that* $\mathrm{sp}\, Q\mathcal{X}Q$ *is* $\| \ \|_2$*-dense in* P *and* $\mathrm{sp}\, V\mathcal{X}V$ *is* $\| \ \|_2$*-dense in* U*. Let* $\varphi : P \to P$ *be a unital,* τ*-preserving, completely positive map such that*

$$\varphi(q_1 x q_2) = q_1 \varphi(x) q_2, \qquad \forall\ q_1, q_2 \in Q,\ x \in P$$

and assume that $\forall\ n \geq 1$, $\exists\ \{m_j\}_j \subset L^2(V, \tau)$ *orthonormal basis of* V *over* Q *such that* $[m_j, f_n] = 0$, $[m_j, \varphi(f_n)] = 0$, $\forall\ j$*. Then there exists a unique unital,* τ*-preserving, completely positive map* $\tilde{\varphi} : U \to U$ *such that* $\tilde{\varphi}|_P = \varphi$ *and* $\tilde{\varphi}(v_1 x v_2) = v_1 \varphi(x) v_2$, $\forall\ v_1, v_2 \in V$, $x \in P$.

Proof. Let $e = e_P^U$. Let $\{m_j\}_j \subset L^2(V)$ be a fixed orthonormal basis of V over Q and note that any element in $\langle U, P\rangle$ can be written in the form $\Sigma_{i,j} m_i p_{ij} e m_j^*$, with $p_{ij} \in P$ (see Ch.1 in [Po2]). We first define an application $\tilde{\varphi} : \langle U, e\rangle \to \langle U, e\rangle$ by

$$\tilde{\varphi}\left(\sum_{i,j} m_i p_{ij} e m_j^*\right) = \sum_{i,j} m_i \varphi(p_{ij}) e m_j^*, \qquad p_{ij} \in P.$$

It is easy to see that $\tilde{\varphi}$ this way defined is completely positive and Tr-preserving and satisfies $\tilde{\varphi}(1) = 1, \tilde{\varphi}(Y_1 X Y_2) = Y_1 \tilde{\varphi}(X) Y_2, \forall\ Y_1, Y_2 \in \langle V, e\rangle,\ X \in \langle U, e\rangle$. Let us next show that $\tilde{\varphi}$ does not depend on the choice of the orthonormal basis $\{m_j\}$ of V over Q. So let $\{m'_j\}_j \subset L^2(V, \tau)$ be another such orthonormal basis. Then $m_i = \sum_k m'_k E_P^U(m_k^{'*} m_i)$ so that if $p \in P$ then $m_i p e m_j^* = \sum_{k,l} m'_k E_P^U(m_k^{'*} m_i) p E(m_j^* m'_l) e m_l^{'*}$ (note that the sums do make sense in $L^2(U, \tau)$, with convergence in $\| \ \|_2$, respectively so-topologies). By definition we thus have $\tilde{\varphi}(m_i p e m_j^*) = m_i \varphi(p) e m_j^*$ and since $E_P^U(m_k^{'*} m_i) \in Q$ and

$$\varphi(E_P^U(m_k^{\prime *} m_i) p E_P^U(m_j^* m'_l)) = E_P^U(m_k^{\prime *} m_i) \varphi(p) E_P^U(m_j^* m'_l),$$

we further get

$$m_i \varphi(p) e m_j^* = \sum_{k,l} m'_k (E_P^U(m_k^{\prime *} m_i) \varphi(p) E_P^U(m_j^* m'_l)) e m_l^{\prime *}$$

$$= \sum_{k,l} m'_k \varphi(E_P^U(m_k^{\prime *} m_i) p E_P^U(m_j^* m'_l)) e m_l^{\prime *}.$$

Taking linear combinations and limits, this shows that if

$$\sum_{i,j} m_i p_{ij} e m_j^* = \sum_{i,j} m'_i p'_{ij} e m_j^{\prime *}$$

then
$$\sum_{i,j} m_i\varphi(p_{ij})em_j^* = \sum_{i,j} m_i'\varphi(p_{ij}')em_j'^*,$$
showing that $\tilde{\varphi}$ does not depend on $\{m_j\}_j$.
We will now show that $\tilde{\varphi}(U) = U$ and that $\tilde{\varphi}|_P = \varphi$. To this end, let us first note that $\tilde{\varphi}(f_n) = \varphi(f_n)$, $\forall\ n$. Indeed, we have $f_n = f_n 1 = f_n \sum_i m_i e m_i^*$ in which we may assume $[m_j, f_n] = 0$, $\forall\ j$ (by the hypothesis and the above). Thus we get $f_n = \sum_i m_i f_n e m_i^*$.
According to the definition of $\tilde{\varphi}$ we get $\tilde{\varphi}(f_n) = \sum_i m_i\varphi(f_n)em_i^*$. But by the hypothesis we may also assume $[m_i, \varphi(f_n)] = 0$ so that we get $\sum_j m_j\varphi(f_n)em_j^* = \varphi(f_n)\sum_j m_j e m_j^* = \varphi(f_n)$.
Since $\tilde{\varphi}$ is V-bilinear (being $\langle Q, e\rangle$-bilinear) it follows that $\tilde{\varphi}(V\mathcal{X}V) = V\tilde{\varphi}(\mathcal{X})V = V\varphi(\mathcal{X})V \subset U$. In particular $\tilde{\varphi}_{|Q\mathcal{X}Q} = \varphi$.
The rest of the statement thus follows by continuity. Q.E.D.

9.3. Corollary. *Let*
$$\begin{array}{ccc} N & \subset & M \\ \cup & & \cup \\ N_0 & \subset & M_0 \end{array}$$
be a nondegenerate commuting square of type II_1 *factors with* $N \subset M$, $N_0 \subset M_0$ *extremal and* $N' \cap M_j = N_0' \cap M_{0j}$, $\forall\ j$. *Let* $T = M \vee M^{\mathrm{op}} \subset M \underset{e_N}{\boxtimes} M^{\mathrm{op}} = S$ *and* $T_0 = M_0 \vee M_0^{\mathrm{op}} \subset M_0 \underset{e_{N_0}}{\boxtimes} M_0^{\mathrm{op}} = S_0$. *If* S *has the property* T *relative to* T *then* S_0 *has the property* T *relative to* T_0.

Proof. By 2.5 we have $T_0' \cap S_0 = \mathbb{C}$, $T' \cap S = \mathbb{C}$. Also, by 2.8 S_0 is naturally included in S and we have a nondegenerate commuting square
$$\begin{array}{ccc} T & \subset & S \\ \cup & & \cup \\ T_0 & \subset & S_0. \end{array}$$

Let $\{N_{0,m}\}_m$ be some tunnel for $N_0 \subset M_0$ and N_m be the corresponding tunnel for $N \subset M$ and denote by $f_n = f_{-n}^n \in M_{0,n}$ the Jones projection for $N_{0,n-1} \subset M_0 \subset M_{0,n}$. By 4.1.4 in [Po8], since S has the property T relative to T and $\mathrm{sp}\cup_n T f_n T$ contains the dense $*$-subalgebra $\cup_n M M_n^{\mathrm{op}} M$ in S (cf. 4.1), it follows that $\forall\ \varepsilon > 0$ there exists n and δ such that if $\varphi : S \to S$ is unital, trace preserving, completely positive, T-T bimodule map with $\|\varphi(f_n) - f_n\|_2 < \delta$ then $\|\varphi(x) - x\|_2 < \varepsilon$, $\forall\ x \in S$, $\|x\| \le 1$. Since
$$f_m \in (N_{0,m} \vee N_{0,m}^{\mathrm{op}})' \cap M_0 \underset{e_{N_0}}{\boxtimes} M_0^{\mathrm{op}} = (N_m \vee N_m^{\mathrm{op}})' \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}}, \forall\ m,$$
it follows that $\forall\ k$, $\exists\ \{m_j^k\}_j \subset N_k \vee N_k^{\mathrm{op}}$ an orthonormal basis of $N_k \vee N_k^{\mathrm{op}}$ over $N_{0,k} \vee N_{0,k}^{\mathrm{op}}$ (which will therefore be an orthonormal basis of T over T_0 as well). Thus $[m_j^k, f_k] = 0, \forall j$.

Let $\varphi_0 : S_0 \to S_0$ be a unital, trace preserving, completely positive, T_0-T_0 bimodule map satisfying $\|\varphi_0(f_n) - f_n\|_2 < \delta$. Since φ_0 is $T_0 - T_0$ bilinear and since $[f_k, N_{0,k} \vee N_{0,k}^{\mathrm{op}}] = 0$, we get $[\varphi_0(f_k), N_{0,k} \vee N_{0,k}^{\mathrm{op}}] = 0$. Thus we also have:

$$\varphi_0(f_m) \in (N_{0,m} \vee N_{0,m}^{\mathrm{op}})' \cap M_0 \underset{e_{N_0}}{\boxtimes} M_0^{\mathrm{op}} = (N_m \vee N_m^{\mathrm{op}})' \cap M \underset{e_N}{\boxtimes} M^{\mathrm{op}}, \forall\, m.$$

So we may apply Lemma 9.2 to get $\varphi : S \to S$ unital, τ-preserving, completely positive T-T bimodule map with $\varphi|_{S_0} = \varphi_0$. Thus $\|\varphi(f_n) - f_n\|_2 = \|\varphi_0(f_n) - f_n\|_2 < \delta$, implying that $\|\varphi(x) - x\|_2 < \varepsilon$, $\forall\, x \in S$, $\|x\| \le 1$.
In particular, $\|\varphi_0(x) - x\|_2 < \varepsilon\ \forall\, x \in S_0$. By Lemma 9.1, this is suficient to ensure that S_0 has the property T relative to T_0. Q.E.D.

9.4. PROPOSITION. *Let*

$$\begin{array}{ccc} N & \subset & M \\ \cup & & \cup \\ N_0 & \subset & M_0 \end{array}$$

be a nondegenerate commuting square of type II_1 *factors with* $N_0 \subset M_0$*,* $N \subset M$ *extremal and* $N_0' \cap M_{0,j} \subset N' \cap M_j$*,* $\forall\, j$ *(i.e.,* $N_0 \subset M_0$ *is smoothly embedded in* $N \subset M$*, in the sense of* [Po2]*). Let* $T_0 \subset S_0$*,* $T \subset S$ *be the corresponding symmetric enveloping inclusions. If* S_0 *has the property* T *relative to* T_0*, then* S *has the property* T *relative to* T*.*

Proof. By [Po8], $\forall\, \varepsilon > 0$, $\exists\, \delta > 0$ and $x_1, \dots, x_n \in S_0$ such that if $\mathcal{H}_0$ is a S_0-S_0 bimodule with a unit vector $\xi_0 \in \mathcal{H}_0$ satisfying $[y, \xi_0] = 0$, $\forall\, y \in T_0$, and $\|[x_i, \xi_0]\| < \delta$, $\forall\, i$, then there exists $\xi_1 \in \mathcal{H}_0$ satisfying $[x, \xi_1] = 0$, $\forall\, x \in S_0$, and $\|\xi_1 - \xi_0\| < \varepsilon$.
Let then $\mathcal{H}$ be a S-S bimodule with a unit vector $\xi_0 \in \mathcal{H}$ such that $[y, \xi_0] = 0$, $\forall\, y \in T$, and $\|[x_i, \xi_0]\| < \delta$, $\forall\, i$. Regarding $\mathcal{H}$ as a S_0-S_0 bimodule it follows that there exists $\xi_0' \in \mathcal{H}$ such that $[x, \xi_0'] = 0$, $\forall\, x \in S_0$, and $\|\xi_0' - \xi_0\| < \varepsilon$.
Denote

$$\begin{aligned} \mathcal{K} &= \{\xi \in \mathcal{H} \mid x\xi = \xi x,\ \forall\, x \in S_0\}, \\ \mathcal{K}_0 &= \{\eta_0 \in \mathcal{H} \mid [y, \eta_0] = 0,\ \forall\, y \in T = M \vee M^{\mathrm{op}}\}, \\ \mathcal{K}_1 &= \{\eta_1 \in \mathcal{H} \mid [y, \eta_1] = 0,\ \forall\, y \in M_1 \vee N^{\mathrm{op}}\}. \end{aligned}$$

With these notations, it follows that $\xi_0 \in \mathcal{K}_0$ and $\xi_0' \in \mathcal{K}$. We then need to construct some positive contractions $A, B \in \mathcal{B}(\mathcal{H})$ such that $0 \le A, B \le 1$, $A\xi = \xi = B\xi$, $\forall\, \xi \in \mathcal{K}$, $A\mathcal{K}_0 \subset \mathcal{K}_1$, $B\mathcal{K}_1 \subset \mathcal{K}_0$. For if we have such A and B, then

$$\|(BA)^n\xi_0 - \xi_0'\| = \|(BA)^n\xi_0 - (BA)^n\xi_0'\| \le \|\xi_0 - \xi_0'\| < \varepsilon$$

so that if ξ_0'' is a weak limit point of $\{(1/n)\sum_{k=1}^n (BA)^k\xi_0\}_n$ then $BA\xi_0'' = \xi_0''$, $\xi_0'' \in \mathcal{K}_0$ (because all $(BA)^k\xi_0$ belong to $\mathcal{K}_0$) and $\|\xi_0'' - \xi_0'\| < \varepsilon$. But $0 \le A \le 1$, $0 \le B \le 1$, $BA\xi_0'' = \xi_0''$ implies that $A\xi_0'' = \xi_0''$, $B\xi_0'' = \xi_0''$, so that $\xi_0'' \in \mathcal{K}_0 \cap \mathcal{K}_1$. Thus $e_1\xi_0'' = \xi_0''e_1$, $y\xi_0'' = \xi_0''y$, $\forall\, y \in T$, and since T and e_1 generate S we get

$x\xi_0'' = \xi_0'' x$, $\forall\, x \in S$. This shows that $\mathcal{H}$ has a nonzero vector commuting with S.

Finally, in order to construct A, B with the required properties, let $\{p_j^n\}_{1\le j\le k_n} \subset M_{0,1} = \langle M_0, e_1\rangle$, $\{q_i^k\}_{1\le i\le m} \subset M_0^{\mathrm{op}}$ be partitions of the identity such that if p^n, respectively q^k denote the spectral projection of $|\sum_j p_j^n e_2 p_j^n - \lambda 1|$, respectively $|\sum_i q_i^k e_1 q_i^k - \lambda 1|$, corresponding to the interval $[\varepsilon, \infty]$, where $\lambda = [M_0 : N_0]^{-1} = [M : N]^{-1}$, then $\tau(p^n) < (1/n)\min_j \tau(p_j^n)$ and $\tau(q^k) < (1/k)\min_i \tau(q_i^k)$ (cf. the Appendix in [Po2], or [Po9]). We claim that if A is a weak limit of the sequence of operators $\{\sum_j p_j^n \cdot p_j^n\}_n \subset \mathcal{B}(\mathcal{H})$ and B is a weak limit of $\{\sum_i q_i^k \cdot q_i^k\}_k \subset \mathcal{B}(\mathcal{H})$ then A, B do satisfy the required conditions. Indeed, since $p_j^n, q_i^k \in S_0$ we have

$$\sum_j p_j^n \xi p_j^n = \xi, \quad \sum_i q_i^k \xi q_i^k = \xi, \qquad \forall\, \xi \in \mathcal{K},\ \forall\, k, n.$$

Thus, $A\xi = \xi = B\xi$, $\forall\, \xi \in \mathcal{K}$. Since $p_j^n \in \langle M_0, e_N\rangle \subset N^{\mathrm{op}\prime} \cap S$ it follows that if $[y, \eta_0] = 0$, $\forall\, y \in T = M \vee M^{\mathrm{op}}$, then

$$\left[x^{\mathrm{op}}, \sum_j p_j^n \eta_0 p_j^n\right] = 0, \qquad \forall\, x^{\mathrm{op}} \in N^{\mathrm{op}}.$$

Thus,

$$[x^{\mathrm{op}}, A\eta_0] = 0, \qquad \forall\, x^{\mathrm{op}} \in N^{\mathrm{op}},\ \forall\, \eta_0 \in \mathcal{K}_0.$$

Let $\eta_0 \in \mathcal{K}_0$ with $\|\eta_0\| = 1$ and note that, since η_0 commutes with T and $T' \cap S = \mathbb{C}$, η_0 follows a trace vector for S. Let also $\xi \in \mathcal{H}$, and $x \in M_1 = \langle M, e_1\rangle$ and note that

$$\begin{aligned}
&\left\|\lambda^{-1}\sum_j p_j^n E_M^{M_1}(p_j^n x p_i^n) - x p_i^n\right\|_2 \\
&\quad = \lambda^{-1/2}\left\|\lambda^{-1}\sum_j p_j^n e_2 p_j^n x p_i^n e_2 - x p_i^n e_2\right\|_2 \\
&\quad \le \lambda^{-1/2}\left\|(1-p^n)\left(\lambda^{-1}\sum_j p_j^n e_2 p_j^n - 1\right)\right\| \, \|x p_i^n e_2\|_2 + \lambda^{-1/2}\|p^n\|_2 \\
&\quad \le 2\lambda^{-1/2}(\|x\|/n)\|p_i^n\|_2.
\end{aligned}$$

Thus we get

$$\begin{aligned}
&\left\| x \sum_i p_i^n \eta_0 p_i^n - \lambda^{-1} \sum_{i,j} p_j^n E_M^{M_1}(p_j^n x p_i^n) \eta_0 p_i^n \right\|^2 \\
&\quad = \sum_i \left\| x p_i^n \eta_0 p_i^n - \lambda^{-1} \sum_j p_j^n E_M^{M_1}(p_j^n x p_i^n) \eta_0 p_i^n \right\|^2 \\
&\quad \leq \sum_i \left\| x p_i^n - \lambda^{-1} \sum_j p_j^n E_M^{M_1}(p_j^n x p_i^n) \right\|_2^2 \\
&\quad \leq 4\lambda^{-1}(\|x\|^2/n^2) \sum_j \|p_i^n\|_2^2 = 4\lambda^{-1}\|x\|^2/n^2.
\end{aligned}$$

Similarly we get

$$\left\| \sum_i p_i^n \eta_0 p_i^n x - \lambda^{-1} \sum_{i,j} p_j^n \eta_0 E_M^{M_1}(p_j^n x p_i^n) p_i^n \right\|^2 \leq 4\lambda^{-1}\|x\|^2/n^2$$

as well.

But since $y\eta_0 = \eta_0 y$, $\forall\ y \in M \subset M \vee M^{\mathrm{op}}$, we have

$$\sum_i \left(\sum_j p_j^n E_M^{M_1}(p_j^n x p_i^n) \right) \eta_0 p_i^n = \sum_j p_j^n \eta_0 \left(\sum_i E_M^{M_1}(p_j^n x p_i^n) p_i^n \right)$$

so by the above estimates we get

$$\left\| x \sum_i p_i^n \eta_0 p_i^n - \sum_j p_j^n \eta_0 p_j^n x \right\| < 8\lambda^{-1/2}\|x\|/n \to 0.$$

Since A is a weak limit of $\{\sum_i p_i^n \cdot p_i^n\}_n$ it follows that $[x, A\eta_0] = 0$, thus $A\mathcal{K}_0 \subset \mathcal{K}_1$. Similar calculations show that $B\mathcal{K}_1 \subset \mathcal{K}_0$ and A, B are thus constructed. As we have previously shown, this was sufficient to ensure that $\mathcal{H}$ has a nonzero vector commuting with S. Thus S has the property T relative to T. Q.E.D.

We can now conclude with the following:

9.5. Theorem. *Let $N_0 \subset M_0$ be an extremal inclusion of type* II_1 *factors such that $M_0 \underset{e_{N_0}}{\boxtimes} M_0^{\mathrm{op}}$ has the property* T *relative to $M_0 \vee M_0^{\mathrm{op}}$. Then $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ has* *the property* T *relative to* $M \vee M^{\mathrm{op}}$ *for any extremal inclusion* $N \subset M$ *with* $\mathcal{G}_{N,M} = \mathcal{G}_{N_0,M_0}$

Proof. Since $N_0 \subset M_0$ is embedded smoothly in $N_0^\omega \subset M_0^\omega$ and the two subfactors have the same higher relative commutants, by 9.4 it follows that $M_0^\omega \underset{e_{N_0^\omega}}{\boxtimes} M_0^{\omega\mathrm{op}}$ has the property T relative to $M_0^\omega \vee M_0^{\omega\mathrm{op}}$. But by [Po9], the inclusion of factors $N^{\mathcal{G}}(R) \subset M^{\mathcal{G}}(R)$, where $\mathcal{G} = \mathcal{G}_{N_0,M_0}$ and R is the hyperfinite II_1 factors, is also embedded as a commuting square with same higher relative commutants in $N_0^\omega \subset M_0^\omega$. Thus, by 9.2 it follows that $M^{\mathcal{G}}(R) \boxtimes (M^{\mathcal{G}}(R))^{\mathrm{op}}$ has the property T relative to $M^{\mathcal{G}}(R) \vee (M^{\mathcal{G}}(R))^{\mathrm{op}}$. But $N^{\mathcal{G}}(R) \subset M^{\mathcal{G}}(R)$ is included in $N^\omega \subset M^\omega$ as well ([Po9]), so $M^\omega \underset{e_{N^\omega}}{\boxtimes} M^{\omega\mathrm{op}}$ has the property T relative to $M^\omega \vee M^{\omega\mathrm{op}}$, by 9.4. Then, again by 9.2 it follows that $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ has the property T relative to $M \vee M^{\mathrm{op}}$. Q.E.D.

9.6. Definition. We say that a standard λ-lattice $\mathcal{G}$ *has the property* T if $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ has the property T relative to $M \vee M^{\mathrm{op}}$ for some (and thus all!) subfactor $N \subset M$ with $\mathcal{G}_{N,M} = \mathcal{G}$.

The following class of examples shows that our notion of property T agrees with Kazhdan's classical notion for groups.

9.7. Proposition. *Let* $\mathcal{G}$ *be the standard* λ*-lattice of a locally trivial subfactor associated to some faithful* G*-kernel on some type* II_1 *factor. Then* $\mathcal{G}$ *has the property T if and only if the group* G *has the property* T.

Proof. Let P be the factor on which G acts and σ be the G-kernel on P. By Section 3 and 9.6, $\mathcal{G}$ has the property T iff $P \bar{\otimes} P^{\mathrm{op}} \rtimes_{\sigma\otimes\sigma^{\mathrm{op}}} G$ has the property T relative to $P \bar{\otimes} P^{\mathrm{op}}$. By ([A-D], [Po8]) this is equivalent to G having the property T. Q.E.D.

Let us next note some simple properties of this notion.

9.8. Proposition. (i) $\mathcal{G}$ *is both amenable and has the property* T *if and only if it has finite depth.*
(ii) *If* $\mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_1$ *(see part (b) in 5.6 for the definition) then* $\mathcal{G}$ *has the property* T *if and only if both* $\mathcal{G}_1$ *and* $\mathcal{G}_2$ *have the property* T.
(iii) $\mathcal{G}$ *has the property* T *if and only if* $\mathcal{G}^{\mathrm{op}}$ *has it.*
(iv) *If* $N \subset M$ *is an extremal inclusion* $\{M_i\}_i$ *is its tower, then* $\mathcal{G}_{N,M}$ *has the property* T *iff* $\mathcal{G}_{M_i,M_j}$ *has the property* T *for some* $i < j$ *iff* G_{M_i,M_j} *has the property* T *for all* $i < j$.

Proof. To prove (i), let $N \subset M$ be an extremal inclusion such that $\mathcal{G}_{N,M} = \mathcal{G}$. Then $\mathcal{G}$ is both amenable and has the property T iff $M \underset{e_N}{\boxtimes} M^{\mathrm{op}}$ is both amenable and has the property T relative to $M \vee M^{\mathrm{op}}$. And by (4.1.4 in [Po8]) this is further equivalent to $[M \underset{e_N}{\boxtimes} M^{\mathrm{op}} : M \vee M^{\mathrm{op}}] < \infty$. But by 4.6, $[M \underset{e_N}{\boxtimes} M^{\mathrm{op}} : M \vee M^{\mathrm{op}}] < \infty$ is equivalent to $N \subset M$ having finite depth.

To prove (ii) let $N_j \subset M_j$, $j = 1, 2$, be such that $\mathcal{G}_{N_j,M_j} = \mathcal{G}_j$ and note that $\mathcal{G}_{N,M} = \mathcal{G}$ where $\mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_2$, $N = N_1 \bar{\otimes} N_2 \subset M_1 \bar{\otimes} M_2 = M$. Then $(T \subset S) = (T_1 \bar{\otimes} T_2 \subset S_1 \bar{\otimes} S_2)$, where $T \subset S$, $T_1 \subset S_1$, $T_2 \subset S_2$ are the symmetric enveloping inclusions associated with $N \subset M$, $N_1 \subset M_1$, respectively $N_2 \subset M_2$. If $T \subset S$ has the relative property T and $\{x_i\}_{1 \le i \le n} \subset S$ is its critical set for some $\varepsilon > 0$, then by [Po8] we may assume x_i are in the algebraic tensor product $S_1 \otimes S_2$, i.e., $x_i = \sum_j x^i_j \otimes y^i_j$, $x^i_j \in S_1$, $y^i_j \in S_2$. Let $\mathcal{H}_1$ be a S_1-S_1 bimodule with a unit vector $\xi_1 \in \mathcal{H}_1$ commuting with T_1 and δ_1-commuting with $\{x^i_j\}_{i,j}$. Denote by $\mathcal{H} = \mathcal{H}_1 \bar{\otimes} L^2(S_2)$, $\xi = \xi_1 \otimes \hat{1}$ and note that if δ_1 is sufficiently small then ξ ε-commutes with $\{x_i\}_i$. It follows that there exists $\xi' \in \mathcal{H}$, commuting with S at distance $K\varepsilon$ from ξ (see [Po8]), where K is a universal constant. But then, if $K\varepsilon < 1$, the projection ξ'' of ξ' onto $\mathcal{H}_1 \otimes \mathbb{C}1 \simeq \mathcal{H}_1$ is a nonzero vector commuting with S_1. This shows that S_1 has the property T relative to T_1. Similarly, S_2 has the property T relative to T_2.

Conversely, if S_i has property T relative to T_i for $i = 1, 2$ and $\mathcal{H}$ is a S-S bimodule with $\xi \in \mathcal{H}$ a unit vector commuting with $\{x_i \otimes y_j\}_{i,j}$, where $\{x_i\}_i \subset S_1$, $\{y_j\}_j \subset S_2$ are the critical sets for $T_1 \subset S_1$, respectively $T_2 \subset S_2$ then it follows that

$$\|u\xi - \xi u\| < \varepsilon, \qquad \forall\, u \in \mathcal{U}(S_1 \otimes 1) \cup \mathcal{U}(1 \otimes S_2).$$

Thus,

$$\|(u \otimes v)\xi - \xi(u \otimes v)\| < 2\varepsilon, \qquad \forall\, u \in \mathcal{U}(S_1),\ v \in \mathcal{U}(S_2).$$

A simple convexity argument in Hilbert space, or Ryll-Nardjewski's fixed point theorem then shows that there exists $\xi' \in \mathcal{H}$, $\|\xi' - \xi\| < 2\varepsilon$, commuting with all elements in the group $\mathcal{F} = \{u \otimes v \mid u \in \mathcal{U}(S_1),\ v \in \mathcal{U}(S_2)\}$. Since $\operatorname{sp} \mathcal{F} \supset S_1 \otimes S_2$ it follows that ξ' commutes with $S = S_1 \bar{\otimes} S_2$. Taking $\varepsilon < 1/2$, this shows that $\mathcal{H}$ has a nonzero vector commuting with S, so S has the property T relative to T.

To prove (iii) we only need to remark that the symmetric enveloping inclusions associated to $N \subset M$ and $N^{\rm op} \subset M^{\rm op}$ are identical, so that 9.5 applies to get that $\mathcal{G}_{N,M}$ has T iff $\mathcal{G}_{N^{\rm op},M^{\rm op}}$ $(= (\mathcal{G}_{N,M})^{\rm op})$ has this property.

Finally, to prove (iv) recall from [Po8] that if $V_0 \subset V \subset U$ are inclusions of factors and $[V : V_0] < \infty$ then U has the property T relative to V iff U has the property T relative to V_0. Thus, if $N \subset M$ is an extremal inclusion and we put $U = M \underset{e_N}{\boxtimes} M^{\rm op}$, $V = M \vee M^{\rm op}$, $V_0 = M \vee N^{\rm op}$, $V_1 = M_1 \vee N^{\rm op}$, it follows that U has the property T relative to V iff U has the property T relative V_1. But $V_1 \subset U$ is a reduced of the symmetric enveloping inclusion for $M \subset M_1$ (cf. 2.6) so, by [Po8] again, it has the relative property T iff $M_1 \vee M_1^{\rm op} \subset M_1 \underset{e_N}{\boxtimes} M_1^{\rm op}$ has relative propert T. Thus, $\mathcal{G}_{N,M}$ has T iff $\mathcal{G}_{M,M_1}$ has T. The rest follows from 2.6 a). Q.E.D.

We do not have more examples of property T standard λ-lattices other than the ones coming from groups (in 9.7) or the obvious ones that can be constructed by using jointly 9.7 and 9.8. For example, we do not know whether there exist standard lattices with the property T that come from irreducible subfactors. As for the minimal standard lattices generated by the Jones projections only, i.e., the so-called Temperley-Lieb-Jones standard lattices, we will prove below that generically they do not have the property T. This will in fact be an imediate corollary of the following more important consequence of 9.4:

9.9. Theorem. *Let $\mathcal{G}$ be a standard λ-lattice and $\mathcal{G}_0$ be a sublattice of $\mathcal{G}$. If $\mathcal{G}_0$ has the property* T *then $\mathcal{G}$ has the property* T. *Conversely, if $[\mathcal{G} : \mathcal{G}_0] < \infty$ and $\mathcal{G}$ has T then $\mathcal{G}_0$ has T.*

Proof. By [Po7] there exists a commuting square

$$\begin{array}{ccc} N & \subset & M \\ \cup & & \cup \\ N_0 & \subset & M_0 \end{array}$$

such that $\mathcal{G}_0 = \mathcal{G}_{N_0,M_0}$ and $\mathcal{G} = \mathcal{G}_{N,M}$. By 9.4 and the definition of property T for standard latices 9.6, it follows that if $\mathcal{G}_{N_0,M_0}$ has T then $\mathcal{G}_{N,M}$ has this property as well.
The last part is trivial, by [Po8], 2.7, 2.9 and 2.10. Q.E.D.

9.10. Corollary. *If a standard λ-lattice $\mathcal{G}_0$ is a sublattice of an amenable standard λ-lattice with infinite graph then $\mathcal{G}_0$ doesn't have the property* T. *In particular, if there exists an amenable subfactor of index λ^{-1} and infinite depth then the Temperley-Lieb-Jones standard lattice of graph A_∞ and index λ^{-1} does not have the property* T.

Proof. Trivial by 9.9. Q.E.D.

Let us end by mentioning a problem which at this point seems of interest:

9.11. *Problem.* Is it true that the property T for a standard lattice $\mathcal{G}$ only depends on its graph, i.e., if $\mathcal{G}, \mathcal{G}_0$ have the same (weighted) graph Γ and $\mathcal{G}$ has T, does it follow that $\mathcal{G}_0$ has T ? Note that in all the examples of property T standard lattices that we have in this paper (obtained by combining 9.7 with 9.8) this is indeed the case.
We strongly believe that this question has a positive answer. If this would be indeed the case, then one would have a notion of property T for standard graphs. We mention that in the combinatorial theory of groups there has been a steady interest towards generalizing the property T from groups to more general objects, in particular to (certain classes of) graphs. Since the standard lattices do generalize discrete groups and certain classes of Kac algebras and compact quantum groups ([Ba]), our definition of property T does provide a generalization along these lines.

Appendix

A.1. Relative Dixmier Property for Subfactors of Finite Index

We prove in this section a version for inclusions of type II_1 factors with finite Jones index of Dixmier's classical result on the norm closure of "averaging" elements by unitaries, as follows:

Theorem. *Let $N \subset M$ be an inclusion of factors of finite index. Then $N \subset M$ has the relative Dixmier property, i.e., for any $x \in M$, we have $\overline{\mathrm{co}}^n\{uxu^* \mid u$ unitary element in $N\} \cap N' \cap M = \{E_{N'\cap M}(x)\}$.*

Proof. For $x \in M$ denote $C_N(x) = \overline{\mathrm{co}}^n\{uxu^* \mid u \in \mathcal{U}(N)\}$. Since $E_{N'\cap M}(uxu^*) = E_{N'\cap M}(x), \forall u \in \mathcal{U}(N)$, it follows that $E_{N'\cap M}(y) = E_{N'\cap M}(x), \forall y \in C_N(x)$. Thus, if for some $x \in M$ we have $C_N(x) \cap N' \cap M \neq \emptyset$ then $C_N(x) \cap N' \cap M = \{E_{N'\cap M}(x)\}$.

By replacing if necessary x by $x - E_{N'\cap M}(x)$, it follows that it is sufficient to check that $0 \in C_N(x)$ for all $x \in M$ with $E_{N'\cap M}(x) = 0$. Moreover, by arguing like in the single algebra case ([D3]), it is sufficient to check this property for selfadjoint such elements x.

We will proceed by contradiction, assuming there exists an element $x = x^*$ in M, with $E_{N'\cap M}(x) = 0$, such that $0 \notin C_N(x)$. By the Hahn-Banach theorem there exists a functional $\Phi = \Phi^* \in M^*$ and $\varepsilon_0 > 0$ such that $\Phi(y) \geq \varepsilon_0,\ \forall y \in C_N(x)$. It follows that $\Psi(x) \geq \varepsilon_0,\ \forall \Psi \in \mathrm{co}\{\Phi(u \cdot u^*) \mid u \in \mathcal{U}(N)\}$ so that $\Psi(x) \geq \varepsilon_0,\ \forall \Psi \in C_N(\Phi) \overset{\mathrm{def}}{=} \overline{\mathrm{co}}^{\sigma(M^*,M)}\{\Phi(u \cdot u^*) \mid u \in \mathcal{U}(N)\}$ and in fact $\Psi(y) \geq \varepsilon_0,\ \forall y \in C_N(x)$ as well.

To get to a contradiction we first show that there exists Ψ in $C_N(\Phi)$ which can be written as $\Psi = \Psi_1 - \Psi_2$, with $\Psi_{1,2}$ positive functionals on M which are scalar multiples of the trace τ when restricted to N. To this end let $\Phi = \Phi_1 - \Phi_2$ be the polar decomposition of Φ, into its positive and negative parts.

Let $\mathcal{V} = \{F \subset (N)_1 \mid F \text{ finite}\}$. By Dixmier's classical Theorem $\forall F \in \mathcal{V}, \exists u_F = (u_1^F, ..., u_{n_F}^F) \subset \mathcal{U}(N)$ such that $\|T_{u_F}(y) - \tau(y)1\| < 1/|F|, \forall y \in F$, where for $X \in M$ we denote $T_{u_F}(X) \overset{\mathrm{def}}{=} (n_F)^{-1}\Sigma_j u_j^F X u_j^{F*}$. Then let ω be a free ultrafilter majorizing the filter $\mathcal{V}$ and for each $i = 1, 2$ define $\Psi_i(X) = \lim_{F\to\omega} \Phi_i(T_{u_F}(X))$, the limit being taken in the usual Banach sense. Then we clearly have $\Psi_{i|N} = c_i\tau_{|N}$, where $c_i = \Phi_i(1), i = 1, 2$. Also, if we let $\Psi = \Psi_1 - \Psi_2$ then $\Psi(X) = \lim_{F\to\omega} \Phi(T_{u_F}(X)), \forall X \in M$ and since $\Phi(T_{u_F}(\quad)) \in C_N(\Phi), \forall F$, it follows that Ψ belongs to $C_N(\Phi)$. Thus $\Psi = \Psi_1 - \Psi_2$ satisfies the desired conditions.

But by [PiPo1] we have $E_N(X) \geq \lambda X, \forall X \in M_+$, so by applying $\Psi_{1,2}$ to both sides we get $c_i\tau(X) = c_i\tau(E_N(X)) = \Psi_i(E_N(X)) \geq \lambda\Psi_i(X)$, implying that $\Psi_i \leq \lambda^{-1}c_i\tau,\ i = 1, 2$. Thus $\Psi_{1,2}$ actually follow normal on all M and so does Ψ. By Sakai's Radon-Nykodim type theorem there exists $a = a^* \in M$ such that $\Psi(X) = \tau(aX), \forall X \in M$. Putting this into the relation that Ψ satisfies gives $\tau(ya) \geq \varepsilon_0, \forall y \in C_N(x)$.

In particular, from the last relation and the trace property we get $\tau(xu^*au) = \tau(uxu^*a) \geq \varepsilon_0$. By taking convex combinations of elements of the form u^*au and weak limits and using that $\overline{\text{co}}^w\{u^*au \mid u \text{ unitary element in } N\} \cap N' \cap M = \{E_{N'\cap M}(a)\}$ (cf. [Po1]), we deduce that $\tau(xE_{N'\cap M}(a)) \geq \varepsilon_0$. But $\tau = \tau \circ E_{N'\cap M}$ and, since $E_{N'\cap M}(xE_{N'\cap M}(a)) = E_{N'\cap M}(x)E_{N'\cap M}(a)$ and x was assumed to satisfy $E_{N'\cap M}(x) = 0$, we obtain $0 \geq \varepsilon_0$ a contradiction which ends the proof of the theorem. Q.E.D.

A.2. A generalized version of Connes' perturbation theorem

In [C1] A. Connes proved a technical result about Hilbert norm perturbations of square integrable operators in semifinite von Neumann algebras.

We will use here a slight modification of his argument (essentially, of his "joint distribution trick") to derive the following version of his result, needed in the proof of Theorem 5.4:

A.2.1. Theorem. *Let P be a semifinite von Neumann algebra with a normal semifinite faithful trace denoted by* Tr*. Let Φ be a positive map on P satisfying the conditions:*

(1) $\Phi(1) = 1$, $\text{Tr} \circ \Phi \leq \text{Tr}$.

(2) $\sup\{\|\Phi(x)\|_{2,\text{Tr}} \mid x \in P,\ \|x\|_{2,\text{Tr}} \leq 1\} \leq 1$.

Let $\delta > 0$ be such that $\delta < (5)^{-4}$ and $b \in P_+$ satisfy the conditions:

(3) $\|b\|_{2,\text{Tr}} = 1$, $\|\Phi(b)\|_{2,\text{Tr}} \geq 1 - \delta$.

(4) $\|b - \Phi(b)\|_{2,\text{Tr}} < \delta$.

Then there exists $s > 0$ such that $\|e_s(b) - \Phi(e_s(b))\|_{2,\text{Tr}} < \delta^{1/4}\|e_s(b)\|$.

Proof. Like in [C1], let $X = \mathbb{R}_+^2 \setminus \{0\}$ and $H_0(x,y) = x$, $H_1(x,y) = y$. As on page 77 in [C1] it then follows that

$$\mu(A_0 \times A_1) \stackrel{\text{def}}{=} \text{Tr}(e_{A_0}(b)\Phi(e_{A_1}(b)))$$

for $A_i \subset \mathbb{R}_+$ Borel sets such that either $0 \neq \bar{A}_0$ or $0 \neq \bar{A}_1$, defines a Radon measure μ on X, which satisfies the properties:

a) $\|f(H_i)\|_{1,\mu} = \text{Tr}(\Phi_i(|f|(b)))$ (respectively $\|f(H_i)\|_{2,\mu}^2 = \text{Tr}(\Phi_i(|f|^2(b))) \leq \|f(b)\|_{2,\text{Tr}}^2$), for all $f : [0,\infty) \to \mathbb{C}$ Borel function with $f(0) = 0$ and $f(b) \in L^1(P,\text{Tr})$ (respectively $f(b) \in L^2(P,\text{Tr})$), $i = 0,1$, where $\Phi_0 = id, \Phi_1 = \Phi$.

b) $\displaystyle\int_X f_0(H_0)\overline{f_1(H_1)}\text{d}\mu = \text{Tr}(f_0(b))\Phi(\bar{f}_1(b)))$, for all $f_i : [0,\infty) \to \mathbb{C}$ Borel with $f_i(0) = 0$ and $f_i(b) \in L^2(P,\text{Tr})$, $i = 0,1$.

c) $\|f_0(H_0) - f_1(H_1)\|_{2,\mu} \geq \|f_0(b) - \Phi(f_1(b))\|_{2,\text{Tr}}$, $\forall\ f_i$ as in b).

d) $\|H_0 - H_1\|_{2,\mu}^2 = \text{Tr}(b^2) + \text{Tr}(\Phi(b^2)) - 2\text{Tr}(b\Phi(b)) \leq 6\delta$.

Indeed, a) and b) are clear by the proof of I.1 in [C1] and the definition of μ.

Further on, by a), b), (1), and Kadison's inequality we get:

$$\begin{aligned}\|f_0(H_0) - f_1(H_1)\|_{2,\mu}^2 &= \|f_0(H_0)\|_{2,\mu}^2 + \|f_1(H_1)\|_{2,\mu}^2 \\ &\quad - 2\mathrm{Re}\int_X f_0(H_0)\overline{f_1(H_1)}\mathrm{d}\mu \\ &= \mathrm{Tr}(f_0(b)^* f_0(b) + \mathrm{Tr}(\Phi(f_1(b)^* f_1(b))) \\ &\quad - 2\mathrm{Re}\ \mathrm{Tr}(f_0(b))\Phi(\overline{f_1}(b))) \\ &\geq \mathrm{Tr}(f_0(b)^* f_0(b_0)) \\ &\quad + \mathrm{Tr}(\Phi(f_1(b))^*\Phi(f_1(b))) \\ &\quad - 2\mathrm{Re}\ \mathrm{Tr}(f_0(b))\Phi(f_1(b))^*) \\ &= \|f_0(b_0) - \Phi(f_1(b))\|_{2,\mathrm{Tr}}^2.\end{aligned}$$

This proves c). Then d) is clear by noticing that the hypothesis and the Cauchy-Schwartz inequality imply:

$$\begin{aligned}&\mathrm{Tr}(b^2) + \mathrm{Tr}(\Phi(b^2)) - 2\mathrm{Tr}(b\Phi(b)) \\ &\quad \leq \mathrm{Tr}(b^2) + \mathrm{Tr}(b^2) - 2\mathrm{Tr}(b\Phi(b)) \\ &\quad = 2 - 2\mathrm{Tr}(b\Phi(b)) \\ &\quad \leq 2 - 2\mathrm{Tr}(b^2) + 2\delta \\ &\quad \leq 2(1-(1-\delta)^2) + 2\delta \leq 6\delta\end{aligned}$$

Remark now that we have, like in proof of 1.2.6 in [C1], the estimate:

$$\begin{aligned}&\int_{\mathbb{R}_+^*} \|e_{t^{1/2}}(H_0) - e_{t^{1/2}}(H_1)\|_{2,\mu}^2 \mathrm{d}t \\ &\quad = \|H_0^2 - H_1^2\|_{1,\mu} \leq \|H_0 - H_1\|_{2,\mu}\|H_0 + H_1\|_{2,\mu}.\end{aligned}$$

But d) implies $\|H_0 - H_1\|_{2,\mu} \leq (6\delta)^{1/2}$ and a) implies $\|H_0+H_1\|_{2,\mu} \leq \|H_0\|_{2,\mu} + \|H_1\|_{2,\mu} \leq \|b\|_{2,\mathrm{Tr}} + \|b\|_{2,\mathrm{Tr}} = 2$. Thus, by applying c) to the function $f = \chi_{[t^{1/2},\infty)}$, for each $t > 0$, we obtain

$$\begin{aligned}&\int_{\mathbb{R}_+^*} \|e_{t^{1/2}}(b) - \Phi(e_{t^{1/2}}(b))\|_{2,\mathrm{Tr}}^2 \mathrm{d}t \\ (*) \qquad &\quad \leq 2(6\delta)^{1/2} = 2(6\delta)^{1/2}\int_{\mathbb{R}_+^*} \|e_{t^{1/2}}(b_0)\|_{2,\mathrm{Tr}}^2 \mathrm{d}t.\end{aligned}$$

This implies that if we denote by D the set of all $t > 0$ for which

$$g(t) \stackrel{\mathrm{def}}{=} \|e_{t^{1/2}}(b_0) - \Phi(e_{t^{1/2}}(b))\|_{2,\mathrm{Tr}}^2 \mathrm{d}t < \delta^{1/4}\|e_{t^{1/2}}(b)\|_{2,\mathrm{Tr}}^2$$

then

$$\int_D \|e_{t^{1/}}(b)\|_{2,\mathrm{Tr}}^2 \mathrm{d}t \geq 1 - 5\delta^{1/4}.$$

Indeed, for if $\int_D \|e_{t^{1/2}}(b_0)\|_{2,\mathrm{Tr}}^2 \mathrm{d}t < 1 - 5\delta^{1/4}$, by taking into account that $g(t) \geq \delta^{1/4}\|e_{t^{1/2}}(b_0)\|_{2,\mathrm{Tr}}^2$ for $t \in \mathbb{R}_+^* \setminus D$, we would get:

$$\begin{aligned}\int_{\mathbb{R}_+^*} g(t)\mathrm{d}t &\geq \int_{\mathbb{R}_+^* \setminus D} g(t)\mathrm{d}t \\ &\geq \delta^{1/4} \int_{\mathbb{R}_+^* \setminus D} \|e_{t^{1/2}}(b_0)\|_{2,\mathrm{Tr}}^2 \mathrm{d}t \\ &\geq 5\delta^{1/2} > 2(6\delta)^{1/2}.\end{aligned}$$

which is in contradiction with $(*)$.
In particular, since $\delta < 5^{-4}$, we have $1 - 5\delta^{1/4} > 0$ so that $D \neq \emptyset$. Thus, any $s > 0$ with $s^2 \in D$ will satisfy the condition in the conclusion. Q.E.D.

References

[A-D] C. Anantharam-Delaroche: *On Connes' property T for von Neumann algebras*, Math. Japonica, 32 (1987), 337-355.

[Av] A. Avez: *Limite de quotients de marchés aléatoires sur des groupes*, C. R. Acad. Sci. Paris, 275 (1973), 317-320.

[Ba] T. Banica: *Representations of compact quantum groups and subfactors*, preprint 1997.

[Bi1] D. Bisch: *A note on intermediate subfactors*, Pac. J. Math., 163 (1994), 201-215.

[Bi2] D. Bisch: *Bimodules, higher relative commutants and the fusion algebra associated to a subfactor*, The Fields Institute for Research in Mathematical Sciences Communications Series, Vol. 13, 1997, AMS, Providence, Rhode Island, pp. 13-63.

[Bi3] D. Bisch: *On the existence of central sequences in subfactors*, Trans. Amer. Math. Soc., 321, (1990), 117-128.

[BiH] D. Bisch and U. Haagerup: *Composition of subfactors: new examples of infinite depth subfactors*, Ann. Scient. Ec. Norm. Sup., 29 (1996), 329–383.

[BiJ] D. Bisch and V. F. R. Jones: *Algebras associated to intermediate subfactors*, Invent. Math., 128 (1997), 89-157.

[C1] A. Connes: *Classification of injective factors*, Ann. of Math., 104 (1976), 73–115.

[C2] A. Connes: *On the classification of von Neumann algebras and their automorphisms*, in Symposia Mathematica, Vol. 20, Academic Press, London and New York, 1976, pp 435-478.

[C3] A. Connes: *Compact metric spaces, Fredholm modules and hypertraces*, Ergod. Th. & Dynam. Sys., 9 (1989), 207-220.

[C4] A. Connes: *Notes on correpondences*, manuscript 1980.

[C5] A. Connes: *Classification des facteurs*, in Symposia Mathematica, Vol. 38, Academic Press, London and New York, 1982, pp 43-109.

[CS] A. Connes and E. Størmer: *Entropy of automorphims of type II_1 von Neumann algebras*, Acta. Math. 134 (1975), 288-306.

[CT] A. Connes and M. Takesaki: *The flow of weights on a factor of type III*, Tohoku Math. J., 29 (1977), 473–575.

[D1] J. Dixmier: *Sous-anneaux abéliens maximaux dans les facteurs de type fini*, Ann. Math., 59 (1954), 279-286.

[D2] J. Dixmier: Les algébres d'operateurs dans l'éspace de Hilbert (Algébres de von Neumann), Gauthier-Villars, Paris, 1957.

[D3] J. Dixmier: Les C*-algébres et leurs représentations, Gauthier-Villars, Paris, 1964.

[EL] E. Effros and C. Lance *Tensor products of operator algebras*, Advances in Math. 25 (1977), 1-34.

[EvKa] D. Evans and Y. Kawahigashi: Quantum symmetries on operator algebras, book manuscript, to appear.

[GePo] L. Ge and S. Popa: *On some decomposition properties for factors of type II_1*, to appear in Duke Math. J., 1998.

[GHJ] F. Goodman, P. de la Harpe and V. F. R. Jones: Coxeter grpahs and towers of algebras, Math. Sci. Res. Inst. Publ., 14, Springer-Verlag, 1989.

[Gr] F. Greenleaf: Invariant Means on Topological Groups, Van Nostrand Math. Studies, New York 1969.

[H1] U. Haagerup: *Principal graphs of subfactors in the index range* $4 < [M : N] < 3 + \sqrt{2}$. In "Subfactors", World Scientfic, Singapore-New Jersey-London-Hong Kong, 1994, pp. 1–39.

[H2] U. Haagerup: Private communication.

[Hi] F. Hiai: *Minimizing indices of conditional expectations onto a subfactor*, Publ. RIMS, Kyoto Univ., 24 (1988), 673–678.

[HiIz] F. Hiai and M. Izumi: *Amenability and strong amenability for fusion algebras with applications to subfactor theory*, preprint 1996.

[ILP] M. Izumi, R. Longo and S. Popa: *A Galois correspondence for compact groups of automorphisms of von Neumann algebras with a generalization to Kac algebras*, J. Funct. Analysis, 155 (1998), 25-63.

[J1] V. F. R. Jones: *Index for subfactors*, Invent. Math., 72 (1983), 1–25.

[J2] V. F. R. Jones: *Subfactors of II_1 factors and related topics*, in Proceedings of the International Congress of Mathematics, Berkeley 1986, pp 939-947.

[J3] V. F. R. Jones: *Planar Algebras*, preprint 1997.

[J4] V. F. R. Jones: *An affine Hecke algebra quotient in the Brauer algebra*, L'Enseign. Mathem. 1996.

[J5] V. F. R. Jones: Actions of finite groups on the hyperfinite type II_1 factor, Mem. Amer. Math. Soc., 237, 1980.

[K] D. Kazhdan: *Connection of the dual space of a group with the structure of its closed subgroups*, Funct. Anal. and its Appl., 1 (1967), 63-65.

[L] R. Longo: *Minimal index and braided subfactors*, J. Funct. Anal., 109 (1992), 97-112.

[Lo] P. Loi: *On the theory of index and type III factors*, thesis 1988.

[Oc] A. Ocneanu: *Quantized groups, string algebras and Galois theory for von Neumann algebras.* In "Operator Algebras and Applications", London Math. Soc. Lect. Notes Series, Vol. 136, 1988, pp. 119–172.

[PiPo1] M. Pimsner and S. Popa: *Entropy and index for subfactors*, Ann. Sci. Ecole Norm. Sup., 19 (1986), 57–106.

[PiPo2] M. Pimsner and S. Popa: *Finite dimensional approximation for pairs of algebras and obstructions for the index*, J. Funct. Anal., 98 (1991), 270–291.

[PiPo3] M. Pimsner and S. Popa: *Iterating the basic construction*, Trans. Amer. Math. Soc., 310 (1988), 127-133.

[Po1] S. Popa: *On a problem of R. V. Kadison on maximal abelian *-subalgebras in factors*, Invent. Math., 65 (1981), 269-281.

[Po2] S. Popa: *Classification of amenable subfactors of type II*, Acta Math., 172 (1994), 163–255.

[Po3] S. Popa: *Classification of subfactors and their endomorphisms*, CBMS Lecture Notes, 86 (1995).

[Po4] S. Popa: *Approximate innerness and central freeness for subfactors: A classification result.* In "Subfactors", World Scientific, Singapore-New Jersey-London-Hong Kong, 1994, pp. 274–293.

[Po5] S. Popa: *Symmetric enveloping algebras, amenability and AFD properties for subfactors*, Math. Res. Lett., 1 (1994), 409–425.

[Po6] S. Popa: *Markov traces on universal Jones algebras and subfactors of finite index*, Invent. Math., 111 (1993), 375-405.

[Po7] S. Popa: *An axiomatization of the lattice of higher relative commutants of a subfactor*, Invent. Math., 120 (1995), 427–445.

[Po8] S. Popa: *Correspondences*, INCREST preprint 1986.

[Po9] S. Popa: *Free independent sequences in type II_1 factors and related problems*, Astérisque 232 (1995), 187-202.

[Po10] S. Popa: *The relative Dixmier property for inclusions of von Neumann algebras*, preprint 1997.

[Po11] S. Popa: *Amenability in the theory of subfactors*, in "Operator Algebras and Quantum Field Theory", International Press, Editors S. Doplicher et al, pp 199-211.

[Po12] S. Popa: *Classification of hyperfinite subfactors of type II and $III_\lambda, 0 < \Lambda \leq 1$*, UCLA preprint 1991, unpublished.

[Po13] S. Popa: *Biduals associated to subfactors, hypertraces and restrictions for the Jones index*, preprint 1997.

[Po14] S. Popa: *On A. Connes' joint distribution trick*, preprint 1997.

[Po15] S. Popa: *Classification of subfactors: the reduction to commuting squares*, Invent. Math., 101 (1990), 19-43.

[Po16] S. Popa: *The classification of hyperfinite subfactors with amenable graph*, preprint 1997.

[PS] R. Powers and E. Størmer: *Free states of the canonical anticommutation relations*, Comm. Math. Phys., 16 (1970), 1-33.

[Sa] S. Sakai: C*-algebras and W*-algebras, Springer-Verlag, Berlin-Heidelber-New York, 1971.

[Sc] J. Schou: *Commuting squares and index for subfactors*, Ph. D. Thesis, Odense University, 1990.

[T] M. Takesaki: *Duality for crossed products and the structure of von Neumann algebras of type III*, Acta Math., 131 (1973), 249-310.

[To] J. Tomiyama: *On the projection of norm one in W*-algebras*, Proc. Japan Acad., 33 (1957), 608-612.

[Wa] A. Wassermann: *Operator algebras and conformal field theory*, Proc. ICM Zurich 1994, Birkhauser Verlag, Basel, Switzerland 1995, pp. 967-979.

[Xu] F. Xu: *Jones-Wassermann subfactors from disconnected intervals*, preprint 1997.

Sorin Popa
Math. Dept.
University of California
Los Angeles, CA 90095-155505
USA
popa@math.ucla.edu