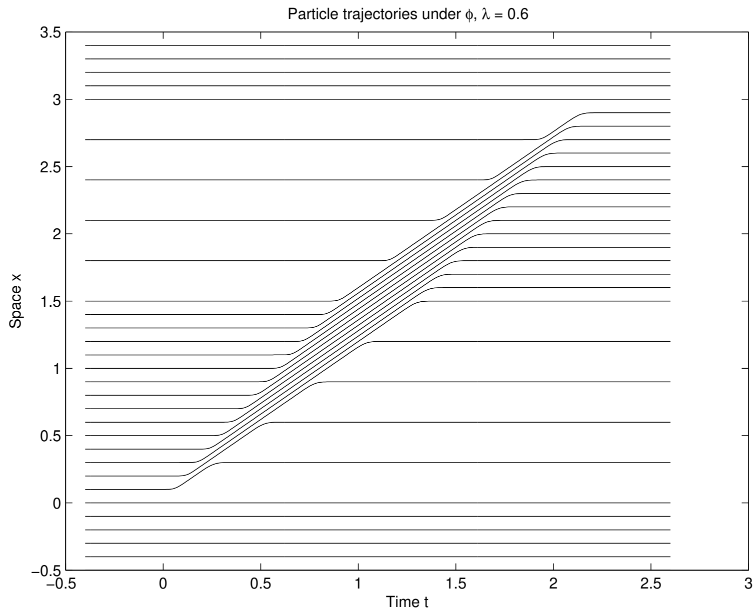
Documenta Mathematica

Journal der

Deutschen Mathematiker-Vereinigung

Gegründet 1996



Particle Trajectories, cf. page 241

Band 10 · 2005

DOCUMENTA MATHEMATICA, Journal der Deutschen Mathematiker-Vereinigung, veröffentlicht Forschungsarbeiten aus allen mathematischen Gebieten und wird in traditioneller Weise referiert. Es wird indiziert durch Mathematical Reviews, Science Citation Index Expanded, Zentralblatt für Mathematik.

Artikel können als TEX-Dateien per E-Mail bei einem der Herausgeber eingereicht werden. Hinweise für die Vorbereitung der Artikel können unter der unten angegebenen 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. It is indexed in Mathematical Reviews, Science Citation Index Expanded, Zentralblatt für Mathematik.

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

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

ISSN 1431-0635 (Print), ISSN 1431-0643 (Internet)

SPARC
LEADING EDGE

DOCUMENTA MATHEMATICA is a Leading Edge Partner of SPARC, the Scholarly Publishing and Academic Resource Coalition of the Association of Research Libraries (ARL), Washington DC, USA.

Address of Technical Managing Editor: Ulf Rehmann, Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld,

Documenta Mathematica
Band 10, 2005

Convexity, Valuations and Prüfer Extensions in Real Algebra

Manfred Knebusch and Digen Zhang

Received: February 11, 2004

Communicated by Ulf Rehmann

Abstract. We analyse the interplay between real valuations, Prüfer extensions and convexity with respect to various preorderings on a given commutative ring. We study all this first in preordered rings in general, then in f-rings. Most often Prüfer extensions and real valuations abound whenever a preordering is present. The next logical step, to focus on the more narrow class of real closed rings, is not yet taken, except in some examples.

2000 Mathematics Subject Classification: 13J30 (13A18, 13F05)
Keywords and Phrases: valuation, Prüfer extension, real holomorphy ring, convex subring

Contents

Introduction

The present paper is based on the book "Manis valuations and Prüfer extensions I" [KZ_1] by the same authors. The book provides details about all terms used here without explanation. But let us emphasize that a "ring" always means a commutative ring with 1, and a *ring extension* $A \subset R$ consists of a ring R and a subring A of R, where, of course, we always demand that the unit element of R coincides with the unit element of A.

The strength and versality of the concept of a *Prüfer extension* seems to depend a great deal on the many different ways we may look at these ring extensions and handle them. So we can say that a ring extension $A \subset R$ is Prüfer iff for every *overring* B of A in R, i.e. subring B of R containing A, the inclusion map $A \hookrightarrow B$ is an epimorphism in the category of rings, and then it follows that B is flat over A, cf. [KZ_1, Th.I.5.2, conditions (11) and (2)]. We can also say that $A \subset R$ is Prüfer iff every overring B of A in R is integrally closed in R [loc.cit., condition (4)].

On the other hand a Prüfer extension $A \subset R$ is determined by the family $S(R/A)$ of equivalence classes of all non trivial Manis valuations $v\colon R \to \Gamma \cup \infty$ on R (cf. [KZ_1, I §1]), such that $v(x) \geq 0$ for every $x \in A$, namely A is the intersection of the rings $A_v := \{x \in R \mid v(x) \geq 0\}$ with v running through $S(R/A)$. Further we can associate to each $v \in S(R/A)$ a prime ideal $\mathfrak{p} := \{x \in A \mid v(x) > 0\}$ of A , and then have

$$A_v \;=\; A_{[\mathfrak{p}]} :=\; A^R_{[\mathfrak{p}]} :=\; \{x \in R \mid \exists s \in A \setminus \mathfrak{p} \text{ with } sx \in A\}.$$

v is – up to equivalence – uniquely determined by $\mathfrak{p}$. We have a bijection $v \leftrightarrow \mathfrak{p}$ of $S(R/A)$ with the set $Y(R/A)$ of all *R-regular* prime ideals $\mathfrak{p}$ of A, i.e. prime ideals $\mathfrak{p}$ of A with $\mathfrak{p}R = R$. {Usually we do not distinguish between equivalent valuations. So we talk abusively of $S(R/A)$ as the set of non trivial Manis valuations of R over A.} Actually the $v \in S(R/A)$ are not just Manis valuations but PM (= "Prüfer-Manis") valuations. These have significantly better properties than Manis valuations in general, cf. [KZ_1, Chap.III].

We call $S(R/A)$ the *restricted PM-spectrum* of the Prüfer extension $A \subset R$ (cf.§1 below). We regard the restricted PM-spectra of Prüfer extensions as the good "complete" families of PM-valuations. In essence they are the same objects as Prüfer extensions.

The word "real algebra" in the title of the present paper is meant in a broad sense. It refers to a part of commutative algebra which is especially relevant for real algebraic geometry, real analytic geometry, and recent expansions of these topics, in particular for semialgebraic and subanalytic geometry and the now emerging *o*-minimal geometry (cf. e.g. [vd D], [vd D_1]).

Real algebra often is of non noetherian nature, but in compensation to this valuations abound. Usually these valuations are *real*, i.e. have a formally real residue class field (cf.§2 below).

A ring R has real valuations whenever R is *semireal*, i.e. -1 is not a sum of squares in R (cf.§2). We then define the *real holomorphy ring* $\mathrm{Hol}(R)$ of R as the intersection of the subrings A_v with v running through all real valuations on R. If R is a field, formally real, this is the customary definition of real holomorphy rings (e.g. [B,p.21]). In the ring case real holomorphy rings have been introduced in another way by M. Marshall, V. Powers and E. Becker ([Mar], [P], [BP]). But we will see in §3 (Cor.3.5) that their definition is equivalent to ours.

Now it can be proved under mild conditions on R, e.g. if $1+x^2$ is a unit in R for every $x \in R$, that $\mathrm{Hol}(R)$ is Prüfer in R, i.e. the extension $\mathrm{Hol}(R) \subset R$ is Prüfer (cf.§2 below). It follows that the restricted PM-spectrum $S(R/\mathrm{Hol}(R))$ is the set of all non trivial *special* (cf.[KZ_1, p.11]) real valuations on R. {Notice that every valuation v on R can be specialized to a special valuation without changing the ring A_v (loc.cit.). A Manis valuation is always special.}

Thus, under mild conditions on R, the non trivial special real valuations on R comprise one good complete family of PM-valuations on R. This fact already indicates that Prüfer extensions are bound to play a major role in real algebra.

An important albeit often difficult task in Prüfer theory is to get a hold on the complete subfamilies of $S(R/A)$ for a given Prüfer extension $A \subset R$. These are the restricted PM-spectra $S(R/B)$ with B running through the overrings of A in R. Thus there is much interest in describing and classifying these overrings of A in various ways.

Some work in this direction has been done in [KZ_1, Chapter II] by use of multiplicative ideal theory, but real algebra provides us with means which go beyond this general theory. In real algebra one very often deals with a preordering T (cf.§5 below) on a given ring R. {A case in point is that R comes as a ring of $\mathbb{R}$-valued functions on some set X, and T is the set of $f \in R$ with $f \geq 0$ everywhere on X. Here T is even a partial ordering of R, $T \cap (-T) = \{0\}$.} Then it is natural to look for T-convex subrings of R, (i.e. subrings which are convex with respect to T) and to study the T-convex hull $\mathrm{conv}_T(\Lambda)$ of a given subring Λ of R. The interplay between real valuations, Prüfer extensions and convexity for varying preorderings on R is the main theme of the present paper.

The smallest preordering in a given semireal ring R is the set $T_0 = \Sigma R^2$ of sums of squares in R. It turns out that $\mathrm{Hol}(R)$ is the smallest $\hat{T}_0$-convex subring $\mathrm{conv}_{\hat{T}_0}(\mathbb{Z})$ of R with respect to the saturation $\hat{T}_0$ (cf.§5, Def.2) of T_0 {This is essentially the *definition* of $\mathrm{Hol}(R)$ by Marshall et al. mentioned above.} Moreover, if every element of $1+T_0$ is a unit in R – an often made assumption in real algebra – then $\mathrm{Hol}(R)$ is Prüfer in R, as stated above, and every overring of $\mathrm{Hol}(R)$ in R is $\hat{T}_0$-convex in R (cf.Th.7.2 below).

Similar results can be obtained for other preorderings instead of T_0. Let (R,T) be any preordered ring. We equip every subring A of R with the preordering $T \cap A$. Convexity in A is always meant with respect to $T \cap A$. We say that A

has *bounded inversion*, if every element of $1+(T\cap A)$ is a unit in A. If R has bounded inversion, it turns out that a subring A of R is convex in R iff A itself has bounded inversion and A is Prüfer in R (cf.Th.7.2 below). Further in this case every overring on A in R again has bounded inversion and is convex in R.

Thus the relations between convexity and the Prüfer property are excellent in the presence of bounded inversion. If bounded inversion does not hold, they are still friendly, as long as $\mathrm{Hol}(R)$ is Prüfer in R. This is testified by many results in the paper.

Given a preordered ring (R,T) and a subring A of R, it is also natural to look for overrings B of A in R such that A is convex and Prüfer in B. Here we quote the following two theorems, contained in our results in §7.

Theorem 0.1 (cf.Cor.7.7 below). Assume that A has bounded inversion. There exists a unique maximal overring D of A in R such that A is convex in D and D has bounded inversion. The other overrings B of A in R with this property are just *all* overrings of A in D.

Notice that Prüfer extensions are not mentioned in this theorem. But in fact D is the Prüfer hull (cf.[KZ_1, I §5]) $P(A,R)$ of A in R. It seems to be hard to prove the theorem without employing Prüfer theory and valuations at last. We also do not know whether an analogue of the theorem holds if we omit bounded inversion.

Theorem 0.2 (cf.Cor.7.10 below). There exists a unique maximal overring E of A in R such that A is Prüfer and convex in E. The other overrings of A in R with this property are just *all* overrings of A in E.

Notice that here no bounded inversion is needed. We call E the *Prüfer convexity cover* of A in the preordered ring $R=(R,T)$ and denote it by $P_c(A,R)$.

If we start with a preordered ring $A=(A,U)$ we may ask whether for every Prüfer extension $A\subset R$ there exists a unique preordering T of R with $T\cap A=U$. In this case, taking for R the (absolute) Prüfer hull $P(A)$ (cf.[KZ_1, I §5]), we have an absolute Prüfer convexity cover $P_c(A){:=}\,P_c(A,P(A))$ at our disposal. This happens, as we will explicate in §10, if A is an *f-ring*, i.e. a lattice ordered ring which is an ℓ-subring (= subring and sublattice) of a direct product of totally ordered rings.

Another natural idea is to classify Prüfer subrings of a given preordering $R=(R,T)$ by the amount of convexity in R they admit. Assume that A is already a convex Prüfer subring of R. Does there exist a unique maximal preordering $U\supset T$ on R such that A is U-convex in R? {Without the Prüfer assumption on A this question still makes sense but seems to be very hard.}

We will see in §13 that this question has a positive answer if R is an f-ring. Let us denote this maximal preordering $U\supset T$ by T_A. Also the following holds, provided $\mathrm{Hol}(R)$ is Prüfer in R. Every overring B of A in R is convex in R

(cf.Th.9.10), and $T_B \supset T_A$. There exists a unique smallest subring H of A such that H is Prüfer and convex in A (hence in R), and $T_H = T_A$. A subring B of R is T_A-convex in R iff $B \supset H$. No bounded inversion condition is needed here.

On the contents of the paper. In §1 we develop the notion of PM-spectrum $pm(R/A)$ and restricted PM-spectrum $S(R/A)$ for any ring extension $A \subset R$. The full PM-spectrum $pm(R/A)$ is needed for functorial reasons, but nearly everything of interest happens in the subset $S(R/A)$. Actually $pm(R/A)$ carries a natural topology (not Hausdorff), but for the purposes in this paper it suffices to handle $pm(R/A)$ as a poset (= partially ordered set) under the specialization relation $\rightsquigarrow$ of that topology. For non trivial PM-valuations v and w the relation $v \rightsquigarrow w$ just means that v is a coarsening of w. {We do not discuss the topology of $pm(R/A)$.} In §1 real algebra does not play any role.

In §2 – §8 we study convexity in a preordered ring $R = (R, T)$ and its relations to real valuations, real spectra, and Prüfer extensions. We start in §2 with the smallest preordering $T_0 = \Sigma R^2$ (using the convexity concept explicitly only later), then considered prime cones in §3 and advance to arbitrary preorderings in §4.

The prime cones of R are the points of the real spectrum Sper$R$. We are eager not to assume too much knowledge about real spectra and related real algebra on the reader's side. We quote results from that area often in a detailed way but, mostly, without proofs.

We study convexity not only for subrings of R but also for ideals of a given subring A of R and more generally for A-submodules of R. Generalizing the concept of a real valuation we also study T-convex valuations on R (cf.§5). The real valuations are just the T_0-convex valuations. {Of course, these concepts exist in real algebra for long, sometimes under other names.} All this seems to be necessary to understand convex Prüfer extensions.

In the last sections, §9 – §13, we turn from preordered rings in general to f-rings. As common for f-rings (cf. e.g.[BKW]), we exploit the interplay between the lattice structure and the ring structure of an f-ring. In particular we here most often meet absolute convexity (cf.§9,Def.1) instead of just convexity. So we obtain stronger results than in the general theory, some of them described above.

Prominent examples of f-rings are the ring $C(X)$ of continuous $\mathbb{R}$-valued functions on a topological space X and the ring $CS(M, k)$ of k-valued continuous semialgebraic functions on a semialgebraic subset M of k^n ($n \in \mathbb{N}$) for k a real closed field.

These rings are fertile ground for examples illustrating our results. They are real closed (in the sense of N. Schwartz, cf.[Sch$_1$]). As Schwartz has amply demonstrated [Sch$_3$], the category of real closed rings, much smaller than the category of f-rings, is flexible enough to be a good environment for studying $C(X)$, and for studying $CS(M,k)$ anyway. Thus a logical next step beyond the study in the present paper will be to focus on real closed rings. For lack of space and time we have to leave this to another occasion.

We also give only few examples involving $C(X)$ and none involving $CS(M,k)$. It would be well possible to be more prolific here. But especially the literature on the rings $C(X)$ is so vast, that it is difficult to do justice to them without writing a much longer paper. We will be content to describe the real holomorphy ring of $C(X)$ (4.13), the minimal elements of the restricted PM-spectrum of $C(X)$ over this ring (1.3, 2.1, 4.13), and the Prüfer hull of $C(X)$ (§11) in general.

Other rings well amenable to our methods are the rings of real C^r-functions on C^r-manifolds, $r \in \mathbb{N} \cup \{\infty\}$, although they are not f-rings.

References. The present paper is an immediate continuation of the book [KZ_1], which is constantly refered to. In these references we omit the label [KZ_1]. Thus, for example, "in Chapter II" means "in [KZ_1, Chapter II]", and "by Theorem I.5.2" means "by Theorem 5.2 in [KZ_1, Chapter I §5]". All other references, which occur also in [KZ_1], are cited here by the same labels as there.

ACKNOWLEDGEMENT. We gratefully acknowledge support by the European RTNetwork RAAG. At various workshops and meetings in this network we could discuss and clarify ideas pertinent to the present paper.

§1 The PM-spectrum of a ring as a partially ordered set

Let R be any ring (as always, commutative with 1).

Definition 1. The *PM-spectrum* of R is the set of equivalence classes of PM-valuations on R. We denote this set by $pm(R)$, and we denote the subset of equivalence classes of *non-trivial* PM-valuations on R by $S(R)$. We call $S(R)$ the *restricted PM-spectrum* of the ring A.

Usually we are sloppy and think of the elements of $pm(R)$ as valuations instead of classes of valuations, replacing an equivalence class by one of its members. We introduce on $pm(R)$ a partial ordering relation "$\rightsquigarrow$" as follows.

Definition 2. Let v and w be PM-valuations of R. We decree that $v \rightsquigarrow w$ if either both v and w are nontrivial and $A_w \subset A_v$, which means that v is a coarsening of w (cf. I §1, Def. 9), or v is trivial and $\operatorname{supp} v \subset \operatorname{supp} w$. □

Remarks 1.1. a) We have a map $\operatorname{supp}: pm(R) \to \operatorname{Spec} R$ from $pm(R)$ to the Zariski spectrum $\operatorname{Spec} R$, sending a PM-valuation on R to its support. This map is compatible with the partial orderings on $pm(R)$ and $\operatorname{Spec} R$: If $v \rightsquigarrow w$ then $\operatorname{supp} v \subset \operatorname{supp} w$.

b) The restriction of the support map $\operatorname{supp}: pm(R) \to \operatorname{Spec} R$ to the subset $pm(R) \setminus S(R)$ of trivial valuations on R is an isomorphism of this poset with $\operatorname{Spec} R$. {"poset" is an abbreviation of "partially ordered set."}

c) Notice that $S(R)$ is something like a "forest". For every $v \in S(R)$ the set of all $w \in S(R)$ with $w \rightsquigarrow v$ is a chain (i.e. totally ordered). Indeed, these valuations w correspond uniquely with the R-overrings B of A_v such that $B \neq R$. Perhaps this chain does not have a minimal element. We should add on the bottom of the chain the trivial valuation v^* on R with $\operatorname{supp} v^* = \operatorname{supp} v$. The valuations v^* should be regarded as the roots of the trees of our forest. □

This last remark indicates that it is not completely silly to include the trivial valuations in the PM-spectrum, although we are interested in nontrivial valuations. Other reasons will be indicated later.

Usually we will not use the full PM-spectrum $pm(R)$ but only the part consisting of those valuations $v \in pm(R)$ such that $A_v \supset A$ for a given subring A.

Definition 3. Let $A \subset R$ be a ring extension.

a) A *valuation on R over A* is a valuation v on R with $A_v \supset A$. In this case the *center of v on A* is the prime ideal $\mathfrak{p}_v \cap A$. We denote it by $cent_A(v)$.

b) The PM-*spectrum of R over A* (or: of the extension $A \subset R$) is the partially orderd subset consisting of the PM-valuations v on R over A. We denote this poset by $pm(R/A)$. The *restricted* PM-*spectrum of R over A* is the subposet $S(R) \cap pm(R/A)$ of $pm(R/A)$. We denote it by $S(R/A)$.

c) The *maximal restricted* PM-spectrum of R over A is the set of maximal elements in the poset $S(R/A)$. We denote it by $\omega(R/A)$. It consists of all non-trivial PM-valuations of R over A which are not proper coarsenings of other such valuations.

Remark 1.2. Notice that, if v and w are elements of $pm(R/A)$ and $v \rightsquigarrow w$, then $cent_A(v) \subset cent_A(w)$. Also, if $v \in pm(R/A)$ and $\mathfrak{p} := cent_A(v)$, then $A_{[\mathfrak{p}]} \subset A_v$ and $\mathfrak{p}_v \cap A_{[\mathfrak{p}]} = \mathfrak{p}_{[\mathfrak{p}]}$. In the special case that $A \subset R$ is Prüfer the pair $(A_{[\mathfrak{p}]}, \mathfrak{p}_{[\mathfrak{p}]})$ is Manis in R. Since this pair is dominated by $(A_v, \mathfrak{p}_v)$ we have $(A_{[\mathfrak{p}]}, \mathfrak{p}_{[\mathfrak{p}]}) = (A_v, \mathfrak{p}_v)$ (cf. Th.I.2.4). It follows that, for $A \subset R$ Prüfer, the center map $cent_A \colon pm(R/A) \to \operatorname{Spec} A$ is an isomorphism from the poset $pm(R/A)$ to the poset $\operatorname{Spec} A$. {Of course, we know this for long.} It maps $S(R/A)$ onto the set $Y(R/A)$ of R-regular prime ideals of A, and $\omega(R/A)$ onto the set $\Omega(R/A)$ of maximal R-regular prime ideals of A. $\square$

Definition 4. If $A \subset R$ is Prüfer and $\mathfrak{p} \in \operatorname{Spec} A$, we denote the PM-valuation v of R over A with $cent_A(v) = \mathfrak{p}$ by $v_{\mathfrak{p}}$. If necessary, we more precisely write $v_{\mathfrak{p}}^R$ instead of $v_{\mathfrak{p}}$.

For a Prüfer extension $A \subset R$ the posets $pm(R/A)$ and $S(R/A)$ are nothing new for us. Here it is only a question of taste and comfort, whether we use the posets $\operatorname{Spec}(A)$ and $Y(R/A)$ or work directly with $pm(R/A)$ and $S(R/A)$. Recall that, if A is Prüfer in R, we have

$$A = \bigcap_{\mathfrak{p} \in Y(R/A)} A_{[\mathfrak{p}]} = \bigcap_{\mathfrak{p} \in \Omega(R/A)} A_{\mathfrak{p}},$$

hence

$$A = \bigcap_{v \in S(R/A)} A_v = \bigcap_{v \in \omega(R/A)} A_v.$$

In the same way any R-overring B of A is determined by the sets of valuations $S(R/B)$ and $\omega(R/B)$.

Example 1.3. Let X be a completely regular Hausdorff space (cf. [GJ, 3.2]). Let $R := C(X)$, the ring of continuous $\mathbb{R}$-valued functions on X, and $A := C_b(X)$, the subring of bounded functions in R.*) As proved in the book [KZ_1], and before in [G_2], the extension $A \subset R$ is Prüfer (even Bezout, cf.II.10.8). In the following we describe the set $\Omega(R/A)$ of R-regular maximal ideals of A.

Every function $f \in A$ extends uniquely to a continuous function f^β on the Stone-Čech compactification βX of X (e.g. [GJ, §6]). Thus we may identify

*) In most of the literature on $C(X)$ this ring is denoted by $C^*(X)$. We have to refrain from this notation since, for any ring R, we denote – as in [KZ_1] – the group of units of R by R^*.

$A = C(\beta X)$. As is very well known, the points $p \in \beta X$ correspond uniquely with the maximal ideals $\mathfrak{p}$ of A via

$$\mathfrak{p} = \mathfrak{m}_p := \{f \in A \mid f^\beta(p) = 0\},$$

cf. [GJ, 7.2]. In particular, $A/\mathfrak{p} = \mathbb{R}$ for every $\mathfrak{p} \in \mathrm{Max}A$. The maximal ideals of $\mathfrak{P}$ of R also correspond uniquely with the points p of βX in the following way [GJ, 7.3]: For any $f \in R$ let $Z(f)$ denote the zero set $\{x \in X \mid f(x) = 0\}$. Then the maximal ideal $\mathfrak{P}$ of R corresponding with $p \in \beta X$ is

$$\mathfrak{P} = M^p := \{f \in R \mid p \in cl_{\beta X}(Z(f))\},$$

where $cl_{\beta X}(Z(f))$ denotes the topological closure of $Z(f)$ in βX. It follows that $M^p \cap A \subset \mathfrak{m}_p$.

By definition $\Omega(R/A)$ is the set of all ideals $\mathfrak{m}_p$ with $\mathfrak{m}_p R = R$. If $\mathfrak{m}_p R = R$ then even $\mathfrak{m}_p \cap R^* \neq \emptyset$. Indeed, we have an equation $1 = \sum\limits_{i=1}^{r} f_i g_i$ with $f_i \in \mathfrak{m}_p$, $g_i \in R$. Then $h := 1 + \sum\limits_{i-1}^{r} g_i^2$ is a unit in R and the functions $\frac{g_i}{h}$ are elements of A. Thus $\frac{1}{h} = \sum\limits_{i=1}^{r} f_i \frac{g_i}{h} \in \mathfrak{m}_p$. It is known that $\mathfrak{m}_p \cap R^* = \emptyset$ iff $R/M^p = \mathbb{R}$ [GJ, 7.9.(b)]. Further the set of points $p \in \beta X$ with $R/M^p = \mathbb{R}$ is known as the real compactification υX of X [GJ, 8.4]. Thus we have

$$\Omega(R/A) = \{\mathfrak{m}_p \mid p \in \beta X \setminus \upsilon X\}.$$

By the way, every $f \in C(X)$ extends uniquely to a continuous function on υX (loc.cit.). Thus we may replace X by υX without loss of generality, i.e. assume that X is realcompact. Then

$$\Omega(R/A) = \{\mathfrak{m}_p \mid p \in \beta X \setminus X\}.$$

In Example 2.1 below we will give a description (from scratch) of the Manis pair $(A_{[\mathfrak{p}]}, \mathfrak{p}_{[\mathfrak{p}]})$ associated with $\mathfrak{p} = \mathfrak{m}_p$ for any $p \in \beta X$. □

We return to an arbitrary ring extension $A \subset R$.

Theorem 1.4. Let $A \subset R$ be a Prüfer extension and B an R-overring.

i) For every PM-valuation w of R over A the special restriction $w|_B$ of w to B is a PM-valuation of B over A.

ii) The map $w \mapsto w|_B$ from $pm(R/A)$ to $pm(B/A)$ is an isomorphism of posets.

Proof. a) Let w be a PM-valuation on R over A. Then $v := w|_B$ is a special valuation on B with $A_v = A_w \cap B$ and $\mathfrak{p}_v = \mathfrak{p}_w \cap B$. In particular, v is a valuation over A. The set $B \setminus A_v$ is closed under multiplication. Thus A_v is

PM in B (cf. Prop. I.5.1.iii). Proposition III.6.6 tells us that v is Manis, hence PM. We have $cent_A(w) = cent_A(v)$.
b) Since the center maps from $pm(R/A)$ to $\operatorname{Spec} A$ and $pm(B/A)$ to $\operatorname{Spec} A$ both are isomorphisms of posets, we have a unique isomorphism of posets $\alpha\colon pm(R/A) \xrightarrow{\sim} pm(B/A)$ such that $cent_A(w) = cent_A(\alpha(w))$ for every $w \in pm(R/A)$. From $cent_A(w) = cent_A(w|_B)$ we conclude that $\alpha(w) = w|_B$. $\square$

The theorem shows well that we sometimes should work with the full PM-spectrum $pm(R/A)$ instead of $S(R/A)$: In the situation of the proposition, whenever $R \neq B$, there exist nontrivial PM-valuations w on R over A such that $w|_B$ is trivial. (All PM-valuations w of R over B have this property.) Thus we do not have a decent map from $S(R/A)$ to $S(B/A)$.

Proposition 1.5.a. Let $B \subset R$ be a Prüfer extension. For every PM-valuation v on B there exists (up to equivalence) a unique PM-valuation w on R with $w|_B = v$.

Proof. The claim follows by applying Theorem 4 *) to the Prüfer extensions $A_v \subset B \subset R$.

Definition 5. In the situation of Proposition 5.a we denote the PM-valuation w on R with $w|_B = v$ by v^R, and we call v^R the *valuation induced by v on R*.

Proposition 1.5.b. If v_1 is a second PM-valuation on B and $v \rightsquigarrow v_1$ then $v^R \rightsquigarrow v_1^R$. Thus, if A is any subring of B, the map $v \mapsto v^R$ is an isomorphism from $pm(B/A)$ onto a sub-poset of $pm(R/A)$. It consists of all $w \in pm(R/A)$ such that $A_w \cap B$ is PM in B.

Proof. We obtain the first claim by applying again Theorem 4 to the extensions $A_{v_1} \subset B \subset R$. The second claim is obvious. $\square$

If M is a subset of $pm(B/A)$ we denote the set $\{v^R \mid v \in M\}$ by M^R.

Theorem 1.6. Assume that $A \subset B$ is a convenient extension (cf. I §6, Def.2) and $B \subset R$ a Prüfer extension. Then the map $S(B/A) \to S(B/A)^R$, $v \mapsto v^R$, is an isomorphism of posets, the inverse map being $w \mapsto w|_B$. The set $S(R/A)$ is the disjoint union of $S(B/A)^R$ and $S(R/B)$. The extension $A \subset R$ is again convenient.

Proof. a) Let $w \in S(R/A)$ be given. If $A_w \supset B$, then $w \in S(R/B)$ and $w|_B$ is trivial. Otherwise $A_w \cap B \neq B$, and the extension $A_w \cap B \subset B$ is PM, since $A \subset B$ is convenient. Now Proposition 5.b tells us that $w = v^R$ for some

*) Reference to Theorem 1.4 in this section. In later sections we will refer to this theorem as "Theorem 1.4." instead of "Theorem 4".

$v \in S(B/A)$. Of course, $v = w|_B$. The isomorphism $pm(R/A) \xrightarrow{\sim} pm(B/A)$, $w \mapsto w|_B$, stated in Theorem 4, maps $S(R/A) \setminus S(R/B)$ onto $S(B/A)$.

b) It remains to prove that R is convenient over A. Let C be an R-overring of A such that $R \setminus C$ is closed under multiplication. We have to verify that C is PM in R.

The set $B \setminus (C \cap B)$ is closed under multiplication. Thus $C \cap B$ is PM in B. It follows that $C \cap B$ is Prüfer in R, hence convenient in R. Since $C \cap B \subset C \subset R$, and $R \setminus C$ is closed under multiplication, we conclude that C is PM in R. $\square$

Various examples of convenient extensions have been given in I, §6. In the case that $A \subset B$ is Prüfer, Theorem 6 boils down to Theorem 4.

We write down a consequence of Theorem 6 for maximal restricted PM-spectra.

Corollary 1.7. Let $A \subset B$ be a convenient extension and $B \subset R$ a Prüfer extension. Then

$$\omega(B/A)^R \subset \omega(R/A) \subset \omega(B/A)^R \cup \omega(R/B).$$

Proof. a) Let $v \in \omega(B/A)^R$ be given. If $w \in S(R/A)$ and $v^R \rightsquigarrow w$ then

$$B \cap A_w \subset B \cap A_{v^R} = A_v \subsetneqq B.$$

We conclude, say by Theorem 6, that $w = u^R$ for some $u \in S(B/A)$. Then $v = v^R|_B \rightsquigarrow w|_B = u$. Since v is maximal, we have $u = v$, and $w = v^R$. Thus v^R is maximal in $S(R/A)$.

b) Let $w \in \omega(R/A)$ be given. Then either $w \in S(R/B)$ or $w = v^R$ for some $v \in S(B/A)$. In the first case certainly $w \in \omega(R/B)$ and in the second case $v \in \omega(B/A)$. {N.B. It may well happen that a given $w \in \omega(R/B)$ is not maximal in $S(R/A)$.} $\square$

§2 Real valuations and real holomorphy rings

If R is a ring and m a natural number we denote the set of sums of m-th powers $x_1^m + \cdots + x_r^m$ in R ($r \in \mathbb{N}$, all $x_i \in R$) by ΣR^m. Notice that $1 + \Sigma R^m$ is a multiplicative subset of R. If m is odd, this set contains 0, hence is of no use. But for m even the set $1 + \Sigma R^m$ will deserve interest.

Let now K be a field. Recall that K is called *formally real* if $-1 \notin \Sigma K^2$. As is very well known ([AS]) this holds iff there exists a *total ordering* on K, by which we always mean a total ordering compatible with addition and multiplication.

We will also use the less known fact, first proved by Joly, that, given a natural number d, the field K is formally real iff $-1 \notin \Sigma K^{2d}$ ([J, (6.16)], cf. also [B_4]).

In the following R is any ring (commutative, with 1, as always).

Definition 1. A prime ideal $\mathfrak{p}$ of R is called *real* if the residue class field $k(\mathfrak{p}) = \mathrm{Quot}(R/\mathfrak{p})$ is formally real.

Remark. Clearly this is equivalent to the following condition: If $a_1, \dots, a_n$ are elements of R with $\sum\limits_{i=1}^{n} a_i^2 \in \mathfrak{p}$ then $a_i \in \mathfrak{p}$ for each $i \in \{1, \dots, n\}$.

Definition 2. A valuation v on R is called *real* if the residue class field $\kappa(v)$ (cf. I, §1) is formally real.

Remark. If v is a trivial valuation on R, then clearly v is real iff the prime ideal $\operatorname{supp} v$ is real. The notion of a real valuation may be viewed as refinement of the notion of real prime ideal.

Example 2.1 (cf. [G_2, Examples 1A and 1B]). Let $R := C(X)$ be the ring of all real-valued continuous functions on a completely regular Hausdorff space X. Let further α be an ultrafilter on the lattice $\mathcal{Z}(X)$ of zero sets $Z(f) = \{x \in X \mid f(x) = 0\}$ of all $f \in R$. Given $f, g \in C(X)$ we say that $f \le g$ at α if there exists $S \in \alpha$ such that $f(x) \le g(x)$ for every $x \in S$, i.e. $\{x \in X \mid f(x) \le g(x)\} \in \alpha$. Since α is an ultrafilter we have $f \le g$ on α or $g \le f$ on α or both. We introduce the following subsets of R.

$$A_\alpha := \{f \in R \mid \exists\, n \in \mathbb{N} \text{ with } |f| \le n \text{ at } \alpha\}.$$
$$I_\alpha := \{f \in R \mid \forall\, n \in \mathbb{N}\colon |f| \le \frac{1}{n} \text{ at } \alpha\}.$$
$$\mathfrak{q}_\alpha := \{f \in R \mid \exists\, S \in \alpha \quad \text{with } f|S = 0\}.$$

We speak of the $f \in A_\alpha$ as the functions *bounded* at α, of the $f \in I_\alpha$ as the functions *infinitesimal* at α, and of the $f \in \mathfrak{q}_\alpha$ as the functions *vanishing* at α.

It is immediate that A_α is a subring of R and $\mathfrak{q}_\alpha$ is a maximal ideal of R (cf.[GJ, 2.5]). It is also clear that I_α is an ideal of A_α. We claim that this ideal is maximal.

In order to prove this, let $f \in R \setminus A_\alpha$ be given. There exists some $n \in \mathbb{N}$ such that

$$Z_1 := \{x \in X \mid \frac{1}{n} \leq |f(x)| \leq n\} \in \alpha.$$

Let $V := \{x \in X \mid \frac{1}{n+1} < |f(x)| < n+1\}$. Then $Z_0 := X \setminus V \in \mathcal{Z}(X)$ and $Z_0 \cap Z_1 = \emptyset$. Thus there exists some $h \in R$ with $h|Z_0 = 0$ and $h|Z_1 = 1$ {We do not need that X is completely regular for this, cf.[GJ, 1.15].} The function $g: X \to \mathbb{R}$ with $g = \frac{h}{f}$ on V and $g = 0$ on Z_0 is continuous, since the function $\frac{1}{f}$ on V is bounded and continuous. Thus $g \in R$. Since $fg \mid Z_1 = 1$ we conclude that $1 - fg \in \mathfrak{q}_\alpha \subset I_\alpha$.

Thus I_α is indeed a maximal ideal on R. Our binary relation "$\leq$ at α" induces a total ordering on the field A_α/I_α which clearly is archimedian. Thus $A_\alpha/I_\alpha = \mathbb{R}$.

Moreover, (A_α, I_α) is a Manis pair in R. For, if $f \in R \setminus A_\alpha$, we have $Y_n := \{x \in X \mid |f(x)| \geq n\} \in \alpha$ for every $n \in \mathbb{N}$. This implies that $\frac{1}{1+f^2} \leq \frac{1}{n}$, $\frac{f}{1+f^2} \leq \frac{1}{n}$ on Y_n, hence $\frac{1}{1+f^2} \in I_\alpha$ and $\frac{f}{1+f^2} \in I_\alpha$. We conclude that

$$f \cdot \frac{f}{1+f^2} = 1 - \frac{1}{1+f^2} \in A_\alpha \setminus I_\alpha.$$

Let $v_\alpha: R \twoheadrightarrow \Gamma_\alpha \cup \infty$ denote the associated Manis valuation on R. Then $\operatorname{supp} v_\alpha = \mathfrak{q}_\alpha$, $A_{v_\alpha} = A_\alpha$, $\mathfrak{p}_{v_\alpha} = I_\alpha$, and v_α has the residue class field $A_\alpha/I_\alpha = \mathbb{R}$ (cf. Prop.I.1.6 and Lemma 2.10 below), hence is real. v_α is trivial iff $\mathfrak{q}_\alpha = I_\alpha$ iff $R/\mathfrak{q}_\alpha = \mathbb{R}$.

The ultrafilters α on $\mathcal{Z}(X)$ can be identified with the points p of βX, cf. [GJ, 6.5]. Clearly $I_\alpha \cap A$ is the maximal ideal $\mathfrak{m}_p$ of A corresponding to the point $p = \alpha$ (cf.1.4 above). Since $A := C_b(X)$ is Prüfer in R, we conclude that (A_α, I_α) is the Manis pair $(A_{[\mathfrak{p}]}, \mathfrak{p}_{[\mathfrak{p}]})$ with $\mathfrak{p} = \mathfrak{m}_p$ in the notation of 1.4. The pair is trivial, i.e. $A_\alpha = R$, iff $p \in vX$. □

We look for a characterization of a valuation to be real in other terms. As before, R is any ring.

Proposition 2.2. Let v be a valuation on R. The following are equivalent
(1) v is real
(2) If $x_1, \dots, x_n$ are finitely many elements of R then

$$v\left(\sum_{i=1}^{n} x_i^2\right) = \min_{1 \leq i \leq n} v(x_i^2).$$

(3) There exists a natural number d such that for any finite sequence $x_1, \dots, x_n$ in R

$$v\left(\sum_{i=1}^{n} x_i^{2d}\right) \quad = \quad \min_{1\le i\le n} v(x_i^{2d}).$$

{N.B. $v(x_i^{2d}) = 2d\, v(x_i)$, of course.}

PROOF. (1) $\Rightarrow$ (2): We first study the case that R is a field. Let $x_1, \dots, x_n \in R$ be given. We assume without loss of generality that $v(x_1) \le \dots \le v(x_n)$ and $x_1 \neq 0$. We have $x_i = a_i x_1$ with $a_i \in A_v$, $a_1 = 1$. Since $A_v/\mathfrak{p}_v$ is a formally real field,

$$1 + a_1^2 + \dots + a_n^2 \notin \mathfrak{p}_v.$$

Thus $v(1 + a_1^2 + \dots + a_n^2) = 0$. This implies

$$v\left(\sum_{i=1}^{n} x_i^2\right) = v(x_1^2) = \min_{1\le i\le n} v(x_i^2).$$

Let now R be a ring and again $x_1, \dots, x_n$ a finite sequence in R. Let $\mathfrak{q} := \operatorname{supp} v$, and – as always – let $\hat{v}$ denote the valuation induced by v on $k(\mathfrak{q})$. Then with $\overline{x}_i := x_i + \mathfrak{q} \in k(\mathfrak{q})$ we have

$$v\left(\sum_{1}^{n} x_i^2\right) = \hat{v}\left(\sum_{1}^{n} \overline{x}_i^2\right) = \quad \min_{1\le i\le n} \hat{v}(\overline{x}_i^2) \quad = \quad \min_{1\le i\le n} v(x_i^2).$$

(2) $\Rightarrow$ (3): trivial.
(3) $\Rightarrow$ (1): Let $A := A_v$, $\mathfrak{p} := \mathfrak{p}_v$, $\mathfrak{q} := \operatorname{supp} v$. Property (3) for the valuation $v\colon R \to \Gamma \cup \infty$ implies the same property for $\overline{v}\colon R/\mathfrak{q} \to \Gamma \cup \infty$. Thus we may assume in advance that $\mathfrak{q} = 0$, hence R is an integral domain.

Let $K := \operatorname{Quot} R$. The valuation v extends to a valuation $\hat{v}\colon K \to \Gamma \cup \infty$. We have $\kappa(v) = \kappa(\hat{v}) = A_{\hat{v}}/\mathfrak{p}_{\hat{v}}$. Exploiting property (3) for $x_1, \dots, x_n \in A_{\hat{v}}$ we obtain

$$-1 \notin \Sigma\, \kappa(\hat{v})^{2d}.$$

Thus $\kappa(v) = \kappa(\hat{v})$ is formally real. $\square$

COROLLARY 2.3. Let $v\colon R \to \Gamma \cup \infty$ be a real valuation on R and H a convex subgroup of R. Then v/H is again a real valuation. If H contains the characteristic subgroup $c_v(\Gamma)$ (cf. I, §1, Def 3), then also $v|H$ is real.

PROOF. It is immediate that property (2) in Proposition 1 is inherited by v/H and $v|H$ from v. $\square$

COROLLARY 2.4. If v is a real valuation on R and B is a subring of R, then the valuations $v|B$ and $v|_B$ are again real.

Proof. $v|B$ inherits property (2) from v, hence is real. It follows by Corollary 3 that also $v|_B$ is real. □

Corollary 2.5. If v is a real valuation on R, then $\operatorname{supp} v$ is a real prime ideal on R.

Proof. This follows immediately from condition (2) in Proposition 2. □

We now start out to prove the remarkable fact that – under a mild condition on R – the set of all non trivial special real valuations on R coincides with the restricted PM-spectrum $S(R/A)$ over a suitable subring A of R which is Prüfer in R.

Definition 3. Let R be any ring. The *real holomorphy ring* $\operatorname{Hol}(R)$ *of* R is the intersection $\bigcap\limits_v A_v$ with v running through all real valuations on R. {If R has no real valuations, we read $\operatorname{Hol}(R) = R$.}

In this definition there is a lot of redundance. $\operatorname{Hol}(R)$ is already the intersection of the rings A_v with v running through all non trivial special real valuations on R.

We need a handy criterion for R which guarantees in sufficient generality that $\operatorname{Hol}(R)$ is Prüfer in R.

Definition 4. We say that R *has positive definite inversion*, if $\mathbb{Q} \subset R$ and if for every $x \in R$ there exists a non constant polynomial $F(t)$ in one variable t over $\mathbb{Q}$ (depending on x) which is positive definite on R_0 *) , hence on $\mathbb{R}$), such that $F(x)$ is a unit of R. {N.B. In this situation the highest coefficient of F is necessarily positive. Thus we may assume in addition that $F(t)$ is monic.}

Notice that, if R has positive definite inversion, then R is convenient over $\mathbb{Q}$ (cf. Scholium I.6.8).

Example. Assume that $\mathbb{Q} \subset R$ and for every $x \in R$ there exists some $d \in \mathbb{N}$ such that $1 + x^{2d} \in R^*$. Then R has positive definite inversion.

Theorem 2.6. If R has positive definite inversion then also $\operatorname{Hol}(R)$ has this property and $\operatorname{Hol}(R)$ is Prüfer in R.

Proof. Let $A{:=} \operatorname{Hol}(R)$. Clearly $\mathbb{Q} \subset A$. If v is any real valuation then also $\mathbb{Q} \subset \kappa(v)$. Moreover, if $F(t) \in \mathbb{Q}[t]$ is a positive definite monic polynomial, then $F(t)$ has no zero in $\kappa(v)$, since $\kappa(v)$ can be embedded into a real closed field which then contains R_0. Thus every real valuation v is an F-valuation as

*) R_0 denotes the real closure of $\mathbb{Q}$, i.e. the field of real algebraic numbers.

defined in I, § 6 (cf. Def.5 there), and we know by Theorem I.6.13 that A is Prüfer in R.

If $x \in A$, and if $F(t) \in \mathbb{Q}[t]$ is positive definite and $F(x) \in R^*$, then $A \subset A_v$ and clearly $v(F(x)) = 0$ for every real valuation v. Thus $\frac{1}{F(x)} \in A$ and $F(x) \in A^*$. $\square$

In Definition 4 we demanded that $\mathbb{Q} \subset R$. This condition, of course, is not an absolute necessity in order to guarantee that $\mathrm{Hol}(R)$ is Prüfer in R. For example, one can prove the following variant of Theorem 2.6 by the same arguments as above.

Theorem 2.6′. Assume that for every $x \in R$ there exists some $d \in \mathbb{N}$ with $1 + x^{2d} \in R^*$. Then also $\mathrm{Hol}(R)$ has this property, and $\mathrm{Hol}(R)$ is Prüfer in R. $\square$

Corollary 2.7. Under the hypothesis in Theorem 6 or 6′ every special real valuation on R is PM. Moreover, if X is any set of real valuations on R, the ring $\bigcap_{v \in X} A_v$ is Prüfer in R. $\square$

Positive definite inversion holds for many rings coming up in real algebra, namely the "strictly semireal rings", to be defined now.

Definition 5. We call a ring R *strictly semireal*, if for every maximal ideal $\mathfrak{m}$ of R the field $R/\mathfrak{m}$ is formally real.*)

Here are other characterizations of strictly semireal rings in the style of Proposition 2 above.

Proposition 2.8. For any ring R the following are equivalent.
(1) R is strictly semireal.
(2) $1 + \Sigma R^2 \subset R^*$.
(3) There exists a natural number d such that $1 + \Sigma R^{2d} \subset R^*$.

Proof. $1 + \Sigma R^2 \subset R^*$ means that $(1 + \Sigma R^2) \cap \mathfrak{m} = \emptyset$ for every maximal ideal $\mathfrak{m}$ of R, and this means that -1 is not a sum of squares in any of the fields $R/\mathfrak{m}$. In the same way we see that $1 + \Sigma R^{2d} \subset R^*$ means that -1 is not a sum of $2d$-th powers in each of these fields. $\square$

Comment. Our term "strictly semireal" alludes to property (2) in Proposition 8. Commonly a ring R is called *semireal* if $-1 \notin \Sigma R^2$ and called *real* if $a_1^2 + \cdots + a_r^2 \neq 0$ for any nonzero elements $a_1, \ldots, a_r$ of R [La_1, §2], [KS Chap III,

*) In I §6, Def.6 we coined the term "totally real" for this property. We now think it is better to reserve the label "totally real" for a ring R where the residue class fields $k(\mathfrak{p})$ of all prime ideals $\mathfrak{p}$ of R are formally real.

§2]. It may be tempting to call a ring R just "totally real" if $R/\mathfrak{m}$ is formally real for every $\mathfrak{m} \in \mathrm{Max}R$, but notice that such a ring is not necessarily real in the established terminology. Schwartz and Madden call our strictly semireal rings "rings having the weak bounded inversion property" [SchM, p.40]. This is a very suitable but lenghty term.

Corollary 2.9. If R is any ring and $d \in \mathbb{N}$, then the localisation $S_d^{-1}R$ with respect to $S_d := 1 + \Sigma R^{2d}$ is strictly semireal, and $S_d^{-1}R = S_1^{-1}R$. $\square$

In the following we need a lemma which could well have been proved in III, §1.

Lemma 2.10. If v is a PM-valuation on R then $\kappa(v) = A_v/\mathfrak{p}_v$.

Proof. We know by III, §1 that $\mathfrak{p}_v$ is a maximal ideal of A_v, hence $\overline{\mathfrak{p}}_v := \mathfrak{p}_v/\mathrm{supp}\, v$ is a maximal ideal of $\overline{A}_v := A_v/\mathrm{supp}\, v$. Proposition I.1.6 tells us that $\mathfrak{o}_v = (\overline{A}_v)_{\overline{\mathfrak{p}}_v}$. (This holds for any Manis valuation v.) Thus $\kappa(v) = \mathfrak{o}_v/\mathfrak{m}_v = A_v/\mathfrak{p}_v$ in our case. $\square$

Theorem 2.11. Assume that R is strictly semireal. Let $d \in \mathbb{N}$ be fixed and $T := \Sigma R^{2d}$. Then

$$\mathrm{Hol}(R) = \sum_{t \in T} \mathbb{Z}\frac{1}{1+t} \quad .$$

(Recall that $1 + T \subset R^*$.) $\mathrm{Hol}(R)$ is again strictly semireal.

Proof. Let $A := \sum\limits_{t \in T} \mathbb{Z}\frac{1}{1+t}$. This is a subring of A since for $t_1, t_2 \in T$

$$\frac{1}{1+t_1} \cdot \frac{1}{1+t_2} = \frac{1}{1+u}$$

with $u := t_1 + t_2 + t_1 t_2 \in T$. As in the proof of Proposition 2, (1) $\Rightarrow$ (2), we see that $v\left(\frac{1}{1+t}\right) \geq 0$ for every $t \in T$ and every real valuation v on R. Thus $A \subset \mathrm{Hol}(R)$.

From I, §6 we infer that A is Prüfer in R (I §6, Example 13). Let v be a PM-valuation on R with $A_v \supset A$. If $a_1, \dots, a_n$ are elements of A then $t := a_1^{2d} + \cdots + a_n^{2d} \in A_v$ and $\frac{1}{1+t} \in A \subset A_v$, hence $1 + t \in A_v^*$. Thus A_v is strictly semireal. Since $\mathfrak{p}_v$ is a maximal ideal of A_v, we conclude by Lemma 10 above that the field $\kappa(v)$ is formally real, i.e. v is a real valuation. It follows that $A_v \supset \mathrm{Hol}(R)$. Since A is the intersection of the rings A_v with v running through $S(R/A)$, we infer that $A \supset \mathrm{Hol}(R)$, and then that $A = \mathrm{Hol}(R)$.

If $t := a_1^{2d} + \cdots + a_r^{2d}$ with elements a_i of A then $1 + t \in A$ and $\frac{1}{1+t} \in A$, hence $1 + t \in A^*$. Thus A is strictly semireal. $\square$

Proposition 2.12. Assume that $A \subset R$ is a Prüfer extension and A is strictly semireal. Then every non trivial PM-valuation on R over A is real.

Proof. Let $\mathfrak{m}$ be an R-regular maximal ideal of A, and let v denote the associated PM-valuation on R with $A_v = A_{[\mathfrak{m}]}$, $\mathfrak{p}_v = \mathfrak{p}_{[\mathfrak{m}]}$. The natural map $A/\mathfrak{m} \to A_{[\mathfrak{m}]}/\mathfrak{p}_{[\mathfrak{m}]}$ is an isomorphism, since $A/\mathfrak{m}$ is already a field. It follows by Lemma 10 that $\kappa(v) = A/\mathfrak{m}$. By assumption this field is formally real. Thus v is real.

We now have proved that every $v \in \omega(R/A)$ is real. The other non trivial PM-valuations on R over A are coarsenings of these valuations, hence are again real, as observed in Corollary 3 above. □

We now state the first main result of this section.

Theorem 2.13. Let R be a strictly semireal ring, and let $A{:}= \mathrm{Hol}(R)$.
i) A is Prüfer in R and $S(R/A)$ is the set of all non trivial special real valuations on R.
ii) A is strictly semireal and $\mathrm{Hol}(A) = A$.
iii) The overrings of A in R are precisely all subrings of R which are strictly semireal and Prüfer in R.
iv) If B is an overring of A in R then $\mathrm{Hol}(B) = A$.

Proof. i): We know by Theorem 6 that A is Prüfer in R and by Theorem 11 that A is strictly semireal, finally by Proposition 12 that every $v \in S(R/A)$ is real. Conversely, if v is any real valuation on R, then $A_v \supset A$ by definition of $A = \mathrm{Hol}(R)$. If in addition v is special, then v is PM since A is Prüfer in R. Thus, if v is non trivial, $v \in S(R/A)$.
ii): We said already that A is strictly semireal, and now know, again by Theorems 6 and 11 (or by i)), that $\mathrm{Hol}(A)$ is strictly semireal and Prüfer in A. Since A is Prüfer in R we conclude that $\mathrm{Hol}(A)$ is Prüfer in R (cf. Th.I.5.6). Now Proposition 12 tells us that every $v \in S(R/\mathrm{Hol}(A))$ is real, hence A_v contains $A = \mathrm{Hol}(R)$. Since $\mathrm{Hol}(A)$ is the intersection of these rings A_v, we have $A \subset \mathrm{Hol}(A)$, i.e. $A = \mathrm{Hol}(A)$.
iii): Assume that B is a strictly semireal subring of R which is Prüfer in R. We see by the same arguments as in the proof of part i) that every $v \in S(R/B)$ is real. B is the intersection of the rings A_v of these valuations v. Thus $A{:}= \mathrm{Hol}(R) \subset B$.

Conversely, if B is an overring of A in R, we have $1 + t \in B$ and $\frac{1}{1+t} \in A \subset B$ for every $t \in \Sigma B^2$. Thus $1 + \Sigma B^2 \subset B^*$, and we conclude by Proposition 2 that B is strictly semireal. Of course, B is also Prüfer in R, since A is Prüfer in R.
iv): Assume that $A \subset B \subset R$. Then both A and B are strictly semireal. Applying claim iii) to the Prüfer extension $A \subset B$ we learn that $\mathrm{Hol}(B) \subset A$, and then, that $\mathrm{Hol}(B)$ is Prüfer in A. Applying the same argument to the Prüfer extension $\mathrm{Hol}(B) \subset A$ we obtain that $\mathrm{Hol}(A) \subset \mathrm{Hol}(B)$. Since $\mathrm{Hol}(A) = A$ we conclude that $\mathrm{Hol}(B) = A$. □

Scholium 2.14. Let R be a strictly semireal ring and B a subring of R which is Prüfer in R. The following are equivalent:

(1) B is strictly semireal.
(2) $S(R/B)$ consists of real valuations.
(2′) $\omega(R/B)$ consists of real valuations.
(3) $\mathrm{Hol}(R) \subset B$.

Proof. The equivalence (1) $\Leftrightarrow$ (3) has been stated in Theorem 13.iii, and the implication (3) $\Rightarrow$ (2) is clear by Theorem 13.i. (2) $\Rightarrow$ (2′) is trivial, and (2′) $\Rightarrow$ (3) is clear by definition of $\mathrm{Hol}(R)$. $\square$

Theorem 2.15. Assume that $A \subset R$ is a Prüfer extension and A is strictly semireal. Then the ring R is strictly semireal.

Proof. Let $\mathfrak{Q}$ be a maximal ideal of R. We want to verify that the field $R/\mathfrak{Q}$ is formally real. We have $\mathfrak{Q} = \mathfrak{q}R$ with $\mathfrak{q}{:}= \mathfrak{Q} \cap A$ (cf. Prop.I.4.6); and $\mathfrak{q}$ is a prime ideal of A. We choose a maximal ideal $\mathfrak{m}$ of A containing $\mathfrak{q}$. Then $\mathfrak{m}R \supset \mathfrak{Q}$.

1.Case: $\mathfrak{m}R \neq R$. This forces $\mathfrak{m}R = \mathfrak{Q}$, since $\mathfrak{Q}$ is maximal. Intersecting with A we obtain $\mathfrak{m} = \mathfrak{q}$. Since $A \subset R$ is ws we have $A_{\mathfrak{m}} = R_{\mathfrak{Q}}$ (I, §3 Def.1). This gives us $R/\mathfrak{Q} = A/\mathfrak{m}$, and $A/\mathfrak{m}$ is formally real.

2.Case: $\mathfrak{m}R = R$. Now there is a PM-valuation v on R with $A_v = A_{[\mathfrak{m}]}$, $\mathfrak{p}_v = \mathfrak{m}_{[\mathfrak{m}]}$. Proposition 12 tells us that v is real. v induces a valuation $\tilde{v}$ on $R_{\mathfrak{m}}$ with $A_{\tilde{v}} = A_{\mathfrak{m}}$, $\mathfrak{p}_{\tilde{v}} = \mathfrak{m}A_{\mathfrak{m}}$, and $\tilde{v}$ is again PM (and real, since $\kappa(\tilde{v}) = \kappa(v)$). Now we invoke Proposition I.1.3, which tells us that $R_{\mathfrak{m}}$ is a local ring with maximal ideal $\operatorname{supp} \tilde{v} = (\operatorname{supp} v)_{\mathfrak{m}}$. This implies that $\mathfrak{Q}_{\mathfrak{m}} \subset (\operatorname{supp} v)_{\mathfrak{m}}$. Taking preimages of these ideals under the localisation map $R \to R_{\mathfrak{m}}$ we obtain $\mathfrak{Q} \subset \operatorname{supp} v$, hence $\mathfrak{Q} = \operatorname{supp} v$, since $\mathfrak{Q}$ is maximal. We conclude by Corollary 5 that $\mathfrak{Q}$ is real, i.e. $R/\mathfrak{Q}$ is formally real. $\square$

Comment. Theorems 13 and 15 together tell us that for a given strictly semireal ring R we have a smallest strictly semireal subring A of R such that A is Prüfer in R, namely $A = \mathrm{Hol}(R)$, and a biggest strictly semireal ring $U \supset R$ such that R is Prüfer in U, namely $U = P(R)$, the Prüfer hull of R. Every ring B between $\mathrm{Hol}(R)$ and $P(R)$ is again strictly semireal, and $\mathrm{Hol}(B) = \mathrm{Hol}(R)$, $P(B) = P(R)$. $\square$

The following theorem may be regarded as the second main result of this section.

Theorem 2.16. Let $B \subset R$ be any Prüfer extension and let v be a real PM-valuation on B. Then the induced PM-valuation v^R on R (cf. §1, Def.5) is again real.

Proof. a) We first prove this in the special case that B is strictly semireal.

Let v be a real PM-valuation on B. Then $A_v \subset B$ is a Prüfer extension with $\omega(R/A_v) = \{v\}$. Since v is real we learn by Scholium 14 that A_v is a strictly semireal ring. The extension $A_v \subset R$ is again Prüfer and v^R is a PM-valuation on R over A_v. Proposition 12 tells us that v^R is real, provided this valuation is non trivial.

There remains the case that v^R is trivial. Then also v is trivial. The prime ideal $\mathfrak{q} := \operatorname{supp} v = \mathfrak{p}_v$ of B is real, and $\mathfrak{Q} := \operatorname{supp}(v^R)$ is a prime ideal of R with $\mathfrak{Q} \cap B = \mathfrak{q}$, hence $\mathfrak{Q} = R\mathfrak{q}$. Since $B \subset R$ is ws, we have $B_{\mathfrak{q}} = R_{\mathfrak{Q}}$. This implies $k(\mathfrak{Q}) = k(\mathfrak{q})$, which is a formally real field. Thus $\mathfrak{Q}$ is real, which means that the trivial valuation v^R is real.

b) We now prove the theorem in general. Let again v be a real valuation on B and $A := A_v$. Let $S := 1 + \Sigma A^2$. The extension $S^{-1}A \subset S^{-1}B$ is Prüfer and $S^{-1}A$ is strictly semireal. By Theorem 15 also $S^{-1}B$ is semireal (and $S^{-1}R$ as well). We have $v(s) = 0$ for every $s \in S$. Thus v extends uniquely to a valuation v' on $S^{-1}B$, and v' is PM and real, the latter since $\kappa(v') = \kappa(v)$. As proved in step a) the PM-valuation $w' := (v')^R$ on $S^{-1}R$ is again real. We have $w'(s) = 0$ for every $s \in S$, of course. Let $j_B : B \to S^{-1}B$ and $j_R : R \to S^{-1}R$ denote the localisation maps of B and R with respect to S, and let $w := w' \circ j_R$. This is a Manis valuation on R since $w(s) = w'(\frac{s}{1}) = 0$ for every $s \in S$. We have $j_R^{-1}(A_{w'}) = A_w$, $j_B^{-1}(A_{v'}) = A_v$, and $A_{w'} \cap S^{-1}B = A_{v'}$. It follows that $A_w \cap B = A_v$. In particular $A_w \supset A_v$ and thus $A_w \subset R$ is Prüfer, hence w is PM. It is now clear that $w|_B = v$, which means that $w = v^R$ (cf. §1, Def.5). We have $\kappa(w) = \kappa(w')$, and we conclude that w is real, since w' is real. $\square$

Corollary 2.17. Let $B \subset R$ be a Prüfer extension. Assume also that $\operatorname{Hol}(B)$ is Prüfer in B (e.g. B has positive definite inversion, cf. Theorem 6). Then $B \cap \operatorname{Hol}(R) = \operatorname{Hol}(B)$.

Proof. If w is a real valuation on R then the restriction $u := w|B$ is a real valuation on B and $A_u = B \cap A_w$. Thus $\operatorname{Hol}(B) \subset B \cap A_w$. Taking intersections we conclude that $\operatorname{Hol}(B) \subset B \cap \operatorname{Hol}(R)$.

On the other hand, if v is a special valuation on B we have $\operatorname{Hol}(B) \subset A_v \subset B$, and we conclude that v is PM, since $\operatorname{Hol}(B)$ is assumed to be Prüfer in B. Now Theorem 16 tells us that the valuation $w := v^R$ is again real. We have $w|_B = v$, hence $A_v = B \cap A_w \supset B \cap \operatorname{Hol}(R)$. Taking intersections we obtain $\operatorname{Hol}(B) \supset B \cap \operatorname{Hol}(R)$. $\square$

Remark 2.18. If R is any ring and B is a subring of R then $\operatorname{Hol}(B) \subset B \cap \operatorname{Hol}(R)$. This is clear by the argument at the beginning of the proof of Corollary 17. $\square$

By use of Theorem 16 we can expand a part of Theorem 13 to more general rings.

Theorem 2.19. Let R be a ring with positive definite inversion. Assume that B is an overring of $\mathrm{Hol}(R)$ in R. Then B has positive definite inversion and $\mathrm{Hol}(B) = \mathrm{Hol}(R)$.

Proof. a) Let $A := \mathrm{Hol}(R)$. We have $\mathbb{Q} \subset A \subset B$. If $x \in B$ and $F(t) \in \mathbb{Q}[t]$ then $F(x) \in B$. If in addition $F(t)$ is positive definite and $F(x) \in R^*$ then $\frac{1}{F(x)} \in A$, as has been verified in the proof of Theorem 6. Thus $\frac{1}{F(x)} \in B$ and $F(x) \in B^*$. This proves that B has positive definite inversion.
b) As observed above (Remark 18), we have $\mathrm{Hol}(B) \subset \mathrm{Hol}(R) \cap B = A$. Since A is a subring of B, we also have $\mathrm{Hol}(A) \subset \mathrm{Hol}(B) \cap A = \mathrm{Hol}(B)$. Thus $\mathrm{Hol}(A) \subset \mathrm{Hol}(B) \subset A$.
c) We finally prove that $\mathrm{Hol}(A) = A$, and then will be done. Given a real valuation v on A we have to verify that $A_v = A$. Now $u := v|_A$ is again real and $A_v = A_u$. Thus we may replace v by u and assume henceforth that v is special.

The ring A has positive definite inversion by Theorem 6 or step a) above. Thus A is convenient, hence v is PM. By Theorem 16 the induced valuation $w := v^R$ is real. This implies $A_w \supset \mathrm{Hol}(R) = A$. On the other hand $w|_A = v$ by definition of w. This implies $A_v = A_w \cap A$. It follows that $A_v = A$. $\square$

As in Theorem 6 we can replace here positive definite inverison by a slightly different condition and prove by the same arguments

Theorem 2.19′. Let R be a ring and B an overring of $\mathrm{Hol}(R)$ in R. Assume that for every $x \in R$ there exists some $d \in \mathbb{N}$ with $1 + x^{2d} \in R^*$. Then this holds for B too, and $\mathrm{Hol}(B) = \mathrm{Hol}(R)$. $\square$

We now introduce "relative" real holomorphy rings. In real algebra some of these are often more relevant objects than the "absolute" holomorphy rings $\mathrm{Hol}(R)$.

Definition 6. Let R be a ring and Λ a subring of R. The *real holomorphy ring of R over Λ* is the intersection of the rings A_v with v running through all real valuations on R over Λ (i.e. with $A_v \supset \Lambda$). We denote this ring by $\mathrm{Hol}(R/\Lambda)$.

In this terminology we have $\mathrm{Hol}(R/\mathbb{Z}) = \mathrm{Hol}(R)$ provided $\mathbb{Z} \subset R$. {If $n \cdot 1_R = 0$ for some $n \in \mathbb{N}$ we have $\mathrm{Hol}(R) = R$, since there do not exist real valuations on R.} It is also clear that $\Lambda \cdot \mathrm{Hol}(R) \subset \mathrm{Hol}(R/\Lambda)$ for any subring Λ of R.

Proposition 2.20. Assume that $\mathrm{Hol}(R)$ is Prüfer in R. {This holds for example if R has positive definite inversion, cf. Theorem 6.} Then for any subring Λ of R we have

$$\mathrm{Hol}(R/\Lambda) = \Lambda \cdot \mathrm{Hol}(R).$$

PROOF. $\mathrm{Hol}(R)$ is Prüfer in R by Theorem 6 (or Theorem 6′), hence $\mathrm{Hol}(R)\cdot\Lambda$ is Prüfer in R. It follows that $\mathrm{Hol}(R)\cdot\Lambda$ is the intersection of the rings A_v with v running through all non trivial PM-valuations on R with $\mathrm{Hol}(R)\cdot\Lambda \subset A_v$, i.e. with $\Lambda \subset A_v$ and $\mathrm{Hol}(R) \subset A_v$. These valuations are known to be real (cf. Theorem 13.i). We conclude that $\mathrm{Hol}(R)\Lambda \supset \mathrm{Hol}(R/\Lambda)$. We also have $\mathrm{Hol}(R)\Lambda \subset \mathrm{Hol}(R/\Lambda)$ as stated above. Thus both rings are equal. □

COROLLARY 2.21. Assume that $B \subset R$ is a Prüfer extension and B is strictly semireal. Then we have a factorisation (cf.II §7, Def.3)

$$\mathrm{Hol}(R/B) \;=\; \mathrm{Hol}(R) \times_{\mathrm{Hol}(B)} B.$$

PROOF. Theorem 15 tells us that R is strictly real. Then Proposition 20 says that $\mathrm{Hol}(R/B) = \mathrm{Hol}(R)\cdot B$. Finally $\mathrm{Hol}(R)\cap B = \mathrm{Hol}(B)$ by Corollary 17. □

§3 Real valuations and prime cones

As before let R be any ring (commutative, with 1, as always).

Definition 1 ([BCR, 7.1], [KS, III, §3], [La$_1$, §4]). A *prime cone* (= "*Ordnung*" in German) of R is a subset P of R with the following properties: $P + P \subset P$, $P \cdot P \subset P$, $P \cup (-P) = A$, $\mathfrak{q} := P \cap (-P)$ is a prime ideal of A. We call $\mathfrak{q}$ the *support of* P and write $\mathfrak{q} = \operatorname{supp} v$.

If R is a field and P a prime cone of R we have $P \cap (-P) = \{0\}$. Thus P is just the set of nonnegative elements of a total ordering of the field R, by which we always mean a total ordering compatible with addition and multiplication. We then call P itself an *ordering of* R.

In general, a prime cone P on R induces a total ordering $\overline{P}$ on the ring $\overline{R} := R/\mathfrak{q}$, $\mathfrak{q} = \operatorname{supp} v$, and then an ordering on $\operatorname{Quot}(\overline{R}) = k(\mathfrak{q})$ in the obvious way (loc.cit.). We denote this ordering of $k(\mathfrak{q})$ by $\hat{P}$.

Notice that P can be recovered from the pair $(\mathfrak{q}, \hat{P})$, since P is just the preimage of $\hat{P}$ under the natural homomorphism $R \to k(\mathfrak{q})$. Thus a prime cone P on the ring R is essentially the same object as pair $(\mathfrak{q}, Q)$ consisting of a prime ideal $\mathfrak{q}$ of R and an ordering Q of $k(\mathfrak{q})$.

Definition 2. The *real spectrum of* R is the set of all prime cones of R. We denote it by $\operatorname{Sper} R$.

We have a natural map

$$\operatorname{supp} \colon \operatorname{Sper} R \longrightarrow \operatorname{Spec} R$$

which sends a prime cone P on R to its support. The image of this map is the set $(\operatorname{Spec} R)_{re}$ of real prime ideals of R. Indeed, if $\mathfrak{q} \in \operatorname{Spec} R$, then $k(\mathfrak{q})$ carries at least one ordering iff $k(\mathfrak{q})$ is formally real. For any $\mathfrak{q} \in (\operatorname{Spec} R)_{re}$ the fibre $\operatorname{supp}^{-1}(\mathfrak{q})$ can be identified with $\operatorname{Sper} k(\mathfrak{q})$.

There lives a very useful topology on $\operatorname{Sper} R$, under which the support map becomes continuous. We will need this only later, cf. §4 below.

Prime cones give birth to real valuations, as we are going to explain now. We first consider the case that R is a field.

We recall some facts about convexity in an ordered field $K = (K, P)$, (cf. [La$_1$], [KS, Chap II], [BCR, 10.1]). We keep the ordering P fixed and stick to the usual notations involving the signs $<, \le$. Thus $P = \{x \in K \mid x \ge 0\}$. Also $|x| := x$ if $x \ge 0$ and $|x| := -x$ if $x \le 0$. A subset M of K is called *convex with respect to* P or P*-convex*, if for $a, b \in M$ with $a < b$ the whole interval $[a, b] := \{x \in K \mid a \le x \le b\}$ is contained in M.

Notice that an abelian subgroup M of $(K, +)$ is P-convex iff for $x \in M \cap P$ and $y, z \in P$ with $x = y + z$ we have $y \in M$ and $z \in M$.

If N is a second P-convex subgroup of $(K,+)$ then $M \subset N$ or $N \subset M$. Also K contains a smallest convex additive subgroup, which we denote by A_P. We have

$$\begin{aligned} A_P &= \{x \in K \mid \exists\, n \in \mathbb{N} \quad \text{with} \quad |x| \le n\} \\ &= \{x \in K \mid \exists\, n \in \mathbb{N} \quad \text{with} \quad n \pm x \in P\}. \end{aligned}$$

Clearly A_P is a subring of K. If x is an element of $K \setminus A_P$, then $|x| > n$ for every $n \in \mathbb{N}$, hence $|x^{-1}| < \frac{1}{n}$ for every $n \in \mathbb{N}$, and a fortiori $x^{-1} \in A_P$. This proves that A_P is a valuation domain of K (i.e. with $\mathrm{Quot}(A_P) = K$), and that

$$I_P \colon = \{x \in K \mid \forall\, n \in \mathbb{N} \colon |x| < \frac{1}{n}\} = \{x \in K \mid \forall\, n \in \mathbb{N} \colon 1 \pm nx \in P\}$$

is the maximal ideal of A_P.

If B is any P-convex subring of K then B is an overring of A_P in K and thus again a valuation domain of K. Moreover,

$$\{0\} \subset \mathfrak{m}_B \subset I_P \subset A_P \subset B \subset K,$$

and $\mathfrak{m}_B$ is a prime ideal of A_P.

Conversely we conclude easily from the fact $[0,1] \subset A_P$ that every A_P-submodule of K is P-convex in K. In particular, every overring B of A_P and every prime ideal of A_P is P-convex in K. *The overrings B of A_P in K are precisely all P-convex subrings* of K. Their maximal ideals $\mathfrak{m}_B$ are the prime ideals of A_P, and they are P-convex in A_P and in K.

More notations. Given a valuation ring B of K, let $\mathfrak{m}_B$ denote the maximal ideal of B. Let $\kappa(B)$ denote the residue class field $B/\mathfrak{m}_B$ of B and $\pi_B \colon B \twoheadrightarrow \kappa(B)$ denote the natural map from B to $\kappa(B)$. Further let v_B denote the canonical valuation associated to B with value group R^*/B^*. {In notations of I, §1 we have $\kappa(v_B) = \kappa(B)$.} For $B = A_P$ we briefly write $\kappa(P)$ instead of $\kappa(A_P)$. Thus $\kappa(P) = A_P/I_P$. In the same vein we write π_P and v_P instead of π_{A_P} and v_{A_P}.

The following facts are easily verified.

Lemma 3.1. Let B be a P-convex subring of K.
i) $Q \colon = \pi_B(P \cap B)$ is an ordering of $\kappa(B)$. In particular $\kappa(B)$ is formally real.
ii) The P-convex subrings C of K with $C \subset B$ correspond uniquely with the Q-convex subrings D of $\kappa(B)$ via $\pi_B(C) = D$ and $\pi_B^{-1}(D) = C$. We have $\pi_B(\mathfrak{m}_C) = \mathfrak{m}_D$ and $\pi_B^{-1}(\mathfrak{m}_D) = \mathfrak{m}_C$.
iii) In particular $\pi_B(A_P) = A_Q$, $\pi_B(I_P) = I_Q$, $\pi_B^{-1}(A_Q) = A_P$, $\pi_B^{-1}(I_Q) = I_P$. $\square$

We state a consequence of a famous theorem by Baer and Krull (cf. [La$_1$, Cor.3.11], [KS, II §7], [BCR, Th.10.1.10]).

Lemma 3.2. Let B be a valuation ring of K and let Q be an ordering (= prime cone) of $\kappa(B)$. Then there exists at least one ordering P of K such that B is P-convex and $\pi_B(B \cap P) = Q$. □

The theorem of Baer-Krull (loc.cit.) gives moreover a precise description of all orderings P on K with this property. We do not need this now. We refer to the literature for a proof of Lemma 2.

We return to an arbitrary ring R and a prime cone P of R. Let $\mathfrak{q} := \operatorname{supp} P$.

Definition 3. As above, $\hat{P}$ denotes the ordering on $k(\mathfrak{q}) = \operatorname{Quot}(R/\mathfrak{q})$. Let $j_\mathfrak{q}: R \to k(\mathfrak{q})$ denote the natural homomorphisms from R to $k(\mathfrak{q})$. We introduce the valuation

$$v_P := v_{\hat{P}} \circ j_\mathfrak{q}$$

on R, the ring $A_P := j_\mathfrak{q}^{-1}(A_{\hat{P}})$, and the prime ideal $I_P := j_\mathfrak{q}^{-1}(I_{\hat{P}})$ of A_P. □

For $v := v_P$ we have $\kappa(v) = \kappa(\hat{P})$, $A_v = A_P$, $\mathfrak{p}_v = I_P$, and $\operatorname{supp} v = \mathfrak{q} = \operatorname{supp} P$.

From the description of A_P and I_P above in the field case, i.e. of $A_{\hat{P}}$ and $I_{\hat{P}}$, we deduce immediately

Lemma 3.3.
$A_P = \{x \in R \mid \exists\, n \in \mathbb{N}: n \pm x \in P\}$,
$I_P = \{x \in R \mid \forall\, n \in \mathbb{N}: 1 \pm nx \in P\}$. □

Theorem 3.4. a) The real valuations on R are, up to equivalence, the coarsenings of the valuations v_P with P running through $\operatorname{Sper} R$.
b) Given a prime cone P of R, the coarsenings w of v_P correspond one-to-one with the $\hat{P}$-convex subrings B of $k(\mathfrak{q})$, $\mathfrak{q} := \operatorname{supp} P$, via $w = v_B \circ j_\mathfrak{q}$.

Proof. If P is a prime cone of R then we know by Lemma 1.i that $v_{\hat{P}}$ is real, and conclude that v_P is real. Thus every coarsening of v_P is real (cf. Cor.2.3.).

Conversely, given a real valuation w on R we have a real valuation $\hat{w}$ on $k(\mathfrak{q})$, $\mathfrak{q} := \operatorname{supp} w$, with $w = \hat{w} \circ j_\mathfrak{q}$. Applying Lemma 2 to an ordering Q on $\kappa(\hat{w}) = \kappa(w)$ we learn that there exists an ordering P' on $k(\mathfrak{q})$ such that $A_{\hat{w}} = \mathfrak{o}_w$ is P'-convex in $k(\mathfrak{q})$. This implies that $\hat{w}$ is a coarsening of $v_{P'}$.

Let $P := j_\mathfrak{q}^{-1}(P')$. This is a prime cone on R with $\operatorname{supp} P = \mathfrak{q}$, $\hat{P} = P'$. It follows that $v_P = v_{P'} \circ j_\mathfrak{q}$, and we conclude that $w = \hat{w} \circ j_\mathfrak{q}$ is a coarsening of v_P. Moreover the coarsenings w of v_P correspond uniquely with the coarsenings u of $v_{\hat{P}}$ via $u = \hat{w}$, $w = u \circ j_\mathfrak{q}$, hence with the overrings of $\mathfrak{o}_P = A_{\hat{P}}$ in $k(\mathfrak{q})$. These are the $\hat{P}$-convex subrings of $k(\mathfrak{q})$. □

Corollary 3.5. The real holomorphy ring $\mathrm{Hol}(R)$ of R is the intersection of the rings A_P with P running through $\mathrm{Sper}R$. Thus $\mathrm{Hol}(R)$ is the set of all $x \in R$, such that for every $P \in \mathrm{Sper}R$ there exists some $n \in \mathbb{N}$ with $n \pm x \in P$.

Proof. This follows from the definition of $\mathrm{Hol}(R)$ in §2 by taking into account Lemma 3 and Theorem 4.a. $\square$

We continue to work with a single prime cone P on R, and we stick to the notations from above. In particular, $\mathfrak{q} := \mathrm{supp}\, P$.

We introduce a binary relation $\leq_P$ on R by defining $x \leq_P y$ iff $y - x \in P$. This relation is reflexive and transitive, but not antisymmetric: If $x \leq_P y$ and $y \leq_P x$ then $x \equiv y \mod \mathfrak{q}$ and vice versa. For any two elements x, y of R we have $x \leq_P y$ or $y \leq_P x$. We write $x <_P y$ if $x \leq_P y$ but not $x \equiv y \mod \mathfrak{q}$.

Given elements a, b of R with $a \leq_P b$ we introduce the "intervals"

$$[a,b]_P := \{x \in R \mid a \leq_P x \leq_P b\} \quad , \quad]a,b[_P := \{x \in R \mid a <_P x <_P b\}.$$

We say that a subset M of R is *P-convex* in R if for any two elements $a, b \in R$ with $a \leq_P b$ the interval $[a,b]_P$ is contained in R.

Notice that the prime cone $\overline{P} := P/\mathfrak{q} := \{x + \mathfrak{q} \mid x \in P\}$ on $R/\mathfrak{q}$ defines a total ordering $\leq_{\overline{P}}$ on the ring $R/\mathfrak{q}$, compatible with addition and multiplication. The P-convex subsets of R are the preimages of the $\overline{P}$-convex subsets of $R/\mathfrak{q}$ under the natural map $R \twoheadrightarrow R/\mathfrak{q}$. Thus the following is evident.

Remarks 3.6. i) Let M be a subgroup of $(R, +)$. Then M is P-convex iff for any two elements x, y of P with $x + y \in M$, we have $x \in M$ and (hence) $y \in M$.
ii) The P-convex additive subgroups of R form a chain under the inclusion relation. $\square$

Lemma 3.7.
i) $\mathrm{supp}\, P$ is the smallest P-convex additive subgroup of R.
ii) A_P is the smallest P-convex additive subgroup M of R with $1 \in M$.
iii) I_P is the biggest P-convex additive subgroup M of R with $1 \notin M$.
iv) If M is any P-convex additive subgroup of R, the set $\{x \in R \mid xM \subset M\}$ is a P-convex subring of R.

Proof. i): Clear, since $\{0\}$ is the smallest $\overline{P}$-convex additive subgroup of $R/\mathfrak{q}$.
ii): An easy verification starting from the description of A_P in Lemma 3.
iii): We know by Lemma 1 that I_P is P-convex in R, and, of course, $1 \notin I_P$. Let M be any P-convex additive subgroup of R with $1 \notin M$. Suppose that $M \not\subset I_P$. We pick some $x \in M \cap P$ with $x \notin I_P$. We learn by Lemma 3 that there exists some $n \in \mathbb{N}$ with $1 - nx \notin P$, hence $nx - 1 = p \in P$. This

implies $1+p = nx \in M$. We conclude by the P-convexity of M that $1 \in M$, a contradiction. Thus $M \subset I_P$.
iv): Again an easy verification. □

As a consequence of this lemma we state

Proposition 3.8.
i) $\operatorname{supp} P$ is the smallest and I_P is the biggest P-convex prime ideal of A_P.
ii) A_P is the smallest P-convex subring of R.
iii) Every P-convex additive subgroup of R is an A_P-submodule of R. □

Definition 4. Given an additive subgroup M of R we introduce the set

$$\operatorname{conv}_P(M) := \bigcup_{z \in P \cap M} [-z, z]_P.$$

This is the smallest P-convex subset of R containing M. We call $\operatorname{conv}_P(M)$ the *P-convex hull* of M (in R). □

Lemma 3.9. $\operatorname{conv}_P(M)$ is again an additive subgroup of R, and

$$\operatorname{conv}_P(M) = \{x \in M \mid \exists\, z \in P \cap M \quad \text{with} \quad z \pm x \in P\}.$$

If M is a subring of R, then $\operatorname{conv}_P(M)$ is a subring of R.

Proof. All this is easily verified. □

Theorem 3.10. a) If w is a coarsening (cf.I §1, Def.9) of the valuation v_P on R, then A_w is a P-convex subring of R.
b) For any subring Λ of R there exists a minimal coarsening w of v_P with $A_w \supset \Lambda$, and $A_w = \operatorname{conv}_P(\Lambda)$.

Proof. a): If w is a coarsening of v_P then $\operatorname{supp}(w) = \mathfrak{q}$. The induced valuation $\hat{w}$ on $k(\mathfrak{q})$ is a coarsening of $\hat{v}_P = v_{\hat{P}}$, and $w = \hat{w} \circ j_{\mathfrak{q}}$. The ring $A_{\hat{w}}$ is $\hat{P}$-convex in $k(\mathfrak{q})$. Thus $A_w = j_{\mathfrak{q}}^{-1}(A_{\hat{w}})$ is P-convex in R.
b): Let $\overline{\Lambda} := j_{\mathfrak{q}}(\Lambda) = \Lambda + \mathfrak{q}/\mathfrak{q}$. This is a subring of $R/\mathfrak{q}$, hence of the field $k(\mathfrak{q})$. We introduce the convex hulls $B := \operatorname{conv}_P(\Lambda)$ and $\hat{B} := \operatorname{conv}_{\hat{P}}(\overline{\Lambda})$. Clearly $\hat{B}$ is the smallest $\hat{P}$-convex subring C of $k(\mathfrak{q})$ with $j_{\mathfrak{q}}^{-1}(C) = B$. There exists a unique coarsening u of $v_{\hat{P}}$ with $A_u = \hat{B}$. Then $w := u \circ j_P$ is a coarsening of v_P with $A_w = B$, and this is the minimal coarsening of v_P with valuation ring B. Since for every coarsening w' of v_P the ring $A_{w'}$ is P-convex in R, it follows that w is also the minimal coarsening of v_P with $A_w \supset \Lambda$. □

Definition 5. We call the valuation w described in Theorem 10.b the *valuation associated with P over* Λ, and denote it by $v_{P,\Lambda}$.

Corollary 3.11. Let again Λ be a subring of R. The relative holomorphy ring $\mathrm{Hol}(R/\Lambda)$ (cf.§2) is the intersection of the rings $A_{v_{P,\Lambda}} = \mathrm{conv}_P(\Lambda)$ with P running through $\mathrm{Sper}R$. It is also the set of all $x \in R$ such that for every $P \in \mathrm{Sper}R$ there exists some $\lambda \in P \cap \Lambda$ with $\lambda \pm x \in P$.

Proof. The first claim follows from Theorems 10 and 4. The second claim then follows from the description of $\mathrm{conv}_P(\Lambda)$ in Lemma 9. $\square$

We now look for P-convex prime ideals of P-convex subrings of R.

Definition 5. For any subring Λ of R we define

$$I_P(\Lambda) := \{x \in R \mid 1+\Lambda x \subset P\} \;=\; \{x \in R \mid 1 \pm \lambda x \in P \quad \text{for every} \quad \lambda \in \Lambda \cap P\}.$$

Theorem 3.12.
a) If w is a coarsening of the valuation v_P on R, then $\mathfrak{p}_w$ is a P-convex*) prime ideal of A_w.
b) Let Λ be a subring of R and $w := v_{P,\Lambda}$. Then $\mathfrak{p}_w = I_P(\Lambda)$. Moreover $I_P(\Lambda)$ is the maximal P-convex proper ideal of $A_w = \mathrm{conv}_P(\Lambda)$.

Proof. a): $\mathfrak{p}_{\hat{w}}$ is a $\hat{P}$-convex prime ideal of $A_{\hat{w}}$. Taking preimages under $j_{\mathfrak{q}}$ we see that the same holds for $\mathfrak{p}_w$ with respect to P and A_w.
b): Let $B := \mathrm{conv}_P(\Lambda)$ and $\hat{B} := \mathrm{conv}_{\hat{P}}(\overline{\Lambda})$ with $\overline{\Lambda} := j_{\mathfrak{q}}(\Lambda)$. For any $x \in R$ we denote the image $j_{\mathfrak{q}}(x)$ by $\overline{x}$. As observed in the proof of Theorem 10, we have $B = A_w$ and $\hat{B} = A_{\hat{w}}$. From valuation theory over fields we know for $x \in (R \setminus \mathfrak{q}) \cap P$ that $\overline{x} \in \mathfrak{p}_{\hat{w}}$ iff $\overline{x}^{-1} \notin \hat{B}$. This means $\overline{x}^{-1} >_{\hat{P}} \overline{\lambda}$ for every $\lambda \in P \cap \Lambda$, i.e. $1 - \overline{\lambda}\overline{x} >_{\hat{P}} 0$. Since $\overline{x} >_{\hat{P}} 0$, this is equivalent to $1 - \overline{\lambda}\overline{x} \in \hat{P}$ for every $\lambda \in P \cap \Lambda$, hence to $1 - \lambda x \in P$ for every $\lambda \in P \cap \Lambda$. It follows easily that indeed

$$\mathfrak{p}_w = j_{\mathfrak{q}}^{-1}(\hat{\mathfrak{p}}_w) = I_P(\Lambda).$$

In particular we now know that $I_P(\Lambda)$ is a P-convex proper ideal of B. If $\mathfrak{a}$ is any such ideal, then for every $x \in \mathfrak{a}$ and $b \in B$ we have $bx \in]-1,1[_P$, hence $1 \pm bx \in P$. In particular $1 \pm \lambda x \in P$ for every $\lambda \in \Lambda$. Thus $x \in I_P(\Lambda)$. This proves that $\mathfrak{a} \subset I_P(\Lambda)$. $\square$

In the case $\Lambda = R$ the theorem tells us the following.

Scholium 3.13. $I_P(R)$ is the maximal P-convex proper ideal of R. It is a prime ideal of R. More precisely, $I_P(R) = \mathfrak{p}_w$ for w the minimal coarsening of v_P with $A_w = R$, i.e. $w = v_{P,R}$. Thus

$$I_P(R) = \{x \in R \mid Rx \subset I_P\}.$$

*) Perhaps it would be more correct to call $\mathfrak{p}_w$ a $(P \cap A_w)$-convex ideal of A_w. But this is not really necessary, since A_w is P-convex in R.

The latter fact is also obvious from the definition $I_P(R) := \{x \in R \mid 1 + Rx \subset P\}$ and the description of I_P in Lemma 3. $\square$

If Λ is a subring of R with $B := \mathrm{conv}_P(\Lambda) \neq R$ the following lemma exhibits two more P-convex ideals of B which both may be different from $I_P(\Lambda) = I_P(B)$.

Lemma 3.14. Let B be a P-convex subring of R with $B \neq R$. Then $R \setminus B$ is closed under multiplication, and the prime ideals $\mathfrak{p}_B$ and $\mathfrak{q}_B$ (cf.I §2, Def.2) of B are again P-convex.

Proof. a) We know by Theorem 10.b that $B = A_w$ for some valuation w on R. This implies that $R \setminus B$ is closed under multiplication.
b) Let $x \in R$, $z \in \mathfrak{p}_B$, and $0 \leq_P x \leq_P z$. There exists some $s \in R \setminus B$ with $sz \in B$. Eventually replacing s by $-s$ we may assume in addition that $s \in P$. Now $0 \leq_P sx \leq_P sz$. We conclude by the P-convexity of B that $sx \in B$, hence $x \in \mathfrak{p}_B$. This proves that $\mathfrak{p}_B$ is P-convex in R.
c) Let $x \in R$, $z \in \mathfrak{q}_B$, and $0 \leq_P x \leq_P z$. For any $s \in P$ we have $0 \leq_P sx \leq_P sz$ and $sz \in B$. This implies that $sx \in B$. It is now clear that $Rx \subset B$, hence $x \in \mathfrak{q}_B$. $\square$

We look for cases where *every* R-overring of A_P is P-convex. We will verify this if R is convenient over $\mathrm{Hol}(R)$. Notice that, according to §2, this happens to be true if R has positive definite inversion, and also, if for every $x \in R$ there exists some $d \in \mathbb{N}$ with $1 + x^{2d} \in R^*$. Indeed, in these cases $\mathrm{Hol}(R)$ is even Prüfer in R (cf. Theorems 2.6 and 2.6′).

We need one more lemma of general nature.

Lemma 3.15. Assume that B is a P-convex subring of R and S a multiplicative subset of R. Then $B_{[S]}$ is again P-convex in R.

Proof. Let $0 \leq_P x \leq_P z$ and $z \in B_{[S]}$. We choose some $s \in S$ with $sz \in B$. Then $s^2 z \in B$ and $0 \leq_P s^2 x \leq_P s^2 z$. Since B is P-convex in R this implies that $s^2 x \in B$. Thus $x \in B_{[S]}$. $\square$

Theorem 3.16. Assume that R is convenient over $\mathrm{Hol}(R)$. Then every R-overring B of A_P is P-convex and PM in R, and $\mathfrak{p}_B = I_P(B)$, provided $B \neq R$.

Proof. We may assume that $B \neq R$. Let $A := A_P$. The set $R \setminus A$ is closed under multiplication. A contains $\mathrm{Hol}(R)$, and R is convenient over $\mathrm{Hol}(R)$. Thus A is PM in R, hence B is PM in R. Let $\mathfrak{P}$ denote the unique R-regular maximal ideal of B (cf. III, §1), and $\mathfrak{p} := \mathfrak{P} \cap A$. Then $B = B_{[\mathfrak{P}]} = A_{[\mathfrak{p}]}$, since A is ws in B. We conclude by Lemma 15 that B is P-convex in R.

We now know by Lemma 14, that $\mathfrak{p}_B$ is P-convex in R, and then by Theorem 12, that $\mathfrak{p}_B \subset I_P(B)$. But $\mathfrak{p}_B$ is a maximal ideal of B, since B is PM in R (cf. Cor.III.1.4). This forces $\mathfrak{p}_B = I_P(B)$. $\square$

A remarkable fact here is that, given a subring B of R, there may exist various prime cones P of R such that B is P-convex. But the prime ideals $I_P(B)$ are all the same, at least if R is convenient over $\mathrm{Hol}(R)$.

Assuming again that R is convenient over $\mathrm{Hol}(R)$ we know that the special restriction $v_P^* := v_P|_R$ of v_P is a PM valuation. There remains the problem to find criteria on P which guarantee that the valuation v_P itself is PM. More generally we may ask for any given ring R and prime cone P of R whether the valuation v_P is special. We defer these questions to the next section, §4.

§4 A brief look at real spectra

Let R be any ring (commutative with 1, as always). In §3 we defined the real spectrum $\mathrm{Sper}R$ as the set of prime cones of R. We now will introduce a topology on $\mathrm{Sper}R$. For this we need some more notations in addition to the ones established in §3.

The proofs of all facts on real spectra stated below can be found in most texts on "abstract" semialgebraic geometry and related real algebra, in particular in [BCR], [KS], [La$_1$]. We will give some of these proofs for the convenience of the reader.

Notations. Given a prime cone P on R let $k(P)$ denote a fixed real closure of the residue class field $k(\mathfrak{q})$ of $\mathfrak{q} := \operatorname{supp} P$ with respect to the ordering $\hat{P}$ induced by P on $k(\mathfrak{q})$. Further let r_P denote the natural homomorphism $R \to R/\mathfrak{q} \hookrightarrow k(\mathfrak{q}) \hookrightarrow k(P)$ from R to $k(P)$. Finally, for any $f \in R$, we define the "value" $f(P)$ of f at P by $f(P) := r_P(f)$. Thus $f(P) = f + \mathfrak{q}$, regarded as an element of $k(P)$.

Given $f \in R$ and $P \in \mathrm{Sper}R$ we either have $f(P) > 0$ or $f(P) = 0$ or $f(P) < 0$. Here we refer to the unique ordering of $k(P)$ (which we do not give a name). Notice that $f(P) = 0$ means $f \in \operatorname{supp} P$, and that $f(P) \geq 0$ iff there is some $\xi \in k(P)$ with $f(P) = \xi^2$.

Remark 4.1. In these notations we can rewrite the definition of $\mathrm{conv}_P(\Lambda)$ and of $I_P(\Lambda)$ for any subring Λ of R (cf. §3) as follows.

$$\begin{aligned}
\mathrm{conv}_P(\Lambda) &= \{f \in R | \exists\, \lambda \in \Lambda \colon |f(P)| \leq |\lambda(P)|\} \\
&= \{f \in R | \exists\, \mu \in \Lambda \colon |f(P)| < |\mu(P)|\}, \\
I_P(\Lambda) &= \{f \in R | \forall\, \lambda \in \Lambda \colon |f(P)\lambda(P)| \leq 1\} \\
&= \{f \in R | \forall\, \mu \in \Lambda \colon |f(P)\mu(P)| < 1\}.
\end{aligned}$$

Here, of course, absolute values are meant with respect to the unique ordering of $k(P)$. □

If T is any subset of R, we define

$$\begin{aligned}
\mathring{H}_R(T) &:= \{P \in \mathrm{Sper}R \mid f(P) > 0 \quad \text{for every} \quad f \in T\}, \\
\overline{H}_R(T) &:= \{P \in \mathrm{Sper}R \mid f(P) \geq 0 \quad \text{for every} \quad f \in T\} \\
&= \{P \in \mathrm{Sper}R \mid P \supset T\}, \\
Z_R(T) &:= \{P \in \mathrm{Sper}R \mid f(P) = 0 \quad \text{for every} \quad f \in T\}.
\end{aligned}$$

If $T = \{f_1, \dots, f_r\}$ is finite, we more briefly write $\mathring{H}_R(f_1, \dots, f_r)$ etc. instead of $\mathring{H}_R(\{f_1, \dots, f_r\})$ etc. We usually suppress the subscript "R" if this does not lead to confusion. Notice that $Z(f) = \overline{H}(-f^2)$ and $Z(f_1, \dots, f_r) = Z(f_1^2 + \cdots + f_r^2) = \overline{H}(-f_1^2 - \cdots - f_r^2)$.

In fact we introduce two topologies on Sper$R$.

Definition 1. a) The *Harrison topology* $\mathcal{T}_{\text{Har}}$ on Sper$R$ is the topology generated by $\mathfrak{H}_R := \{\mathring{H}_R(f) \mid f \in R\}$ as a subbasis of open sets.
b) A subset X of Sper$R$ is called *constructible* if X is a boolean combination in Sper$R$ of finitely many sets $\mathring{H}_R(f), f \in R$. We denote the set of all constructible subsets of Sper$R$ by $\mathcal{K}_R$. This is the boolean lattice of subsets of Sper$R$ generated by $\mathfrak{H}_R$.
c) The *constructible topology* $\mathcal{T}_{\text{con}}$ on Sper$R$ is the topology generated by $\mathcal{K}_R$ as a basis of open sets. In this topology every $X \in \mathcal{K}_R$ is clopen, i.e. closed and open.

If nothing else is said we regard Sper$R$ as a topological space with respect to the Harrison topology $\mathcal{T}_{\text{Har}}$, while $\mathcal{T}_{\text{con}}$ will play only an auxiliary role. Of course, $\mathcal{T}_{\text{con}}$ is a much finer topology than $\mathcal{T}_{\text{Har}}$. We denote the topological space (Sper$R, \mathcal{T}_{\text{Har}}$) simply by Sper$R$ and the space (Sper$R, \mathcal{T}_{\text{con}}$) by $(\text{Sper}R)_{\text{con}}$.

$(\text{Sper}R)_{\text{con}}$ turns out to be a compact Hausdorff space. Thus Sper$R$ itself is quasicompact. Also, a constructible subset U of Sper$R$ is open iff U is the union of finitely many sets $\mathring{H}(f_1, \dots, f_r)$. We denote the family of open constructible subsets of Sper$R$ by $\mathring{\mathcal{K}}_R$ and the family of closed constructible subsets of Sper$R$ by $\overline{\mathcal{K}}_R$.

If R is a field then $\mathcal{T}_{\text{con}}$ and $\mathcal{T}_{\text{Har}}$ coincide, hence Sper$R$ is compact (= quasicompact and Hausdorff) in this case, but for R a ring Sper$R$ most often is not Hausdorff.

The support map supp : Sper$R \to \operatorname{Spec} R$ is easily seen to be continuous. Indeed, given $f \in R$, the basic open set $D(f) := \{\mathfrak{p} \in \operatorname{Spec} R \mid f \notin \mathfrak{p}\}$ of $\operatorname{Spec} R$ has the preimage $\{P \in \text{Sper}R \mid f(P) \neq 0\} = \mathring{H}(f^2)$ under this map.

Every ring homomorphism $\varphi: R \to R'$ gives us a map

$$\text{Sper}(\varphi) = \varphi^*: \quad \text{Sper}R' \longrightarrow \text{Sper}R,$$

defined by $\varphi^*(P') = \varphi^{-1}(P')$ for P' a prime cone of R'. It is easily seen (loc.cit.) that Sper(φ) is continuous with respect to the Harrison topology and also with respect to the constructible topology on both sets. In other terms, if $X \in \mathcal{K}_R$ (resp. $\mathring{\mathcal{K}}_R$, resp. $\overline{\mathcal{K}}_R$) then $(\varphi^*)^{-1}(X) \in \mathcal{K}_{R'}$ (resp. $\mathring{\mathcal{K}}_{R'}$, resp. $\overline{\mathcal{K}}_{R'}$).

Notice also that $\operatorname{supp}(\varphi^{-1}(P')) = \varphi^{-1}(\operatorname{supp}\varphi)$. Thus we have a commutative square of continuous maps

$$\begin{array}{ccc} \mathrm{Sper}R' & \overset{\mathrm{Sper}(\varphi)}{\longrightarrow} & \mathrm{Sper}R \\ \downarrow \mathrm{supp} & & \downarrow \mathrm{supp} \\ \mathrm{Spec}\,R' & \underset{\mathrm{Spec}\,(\varphi)}{\longrightarrow} & \mathrm{Spec}\,R. \end{array}$$

Before continuing our discussion of properties of real spectra, we give an application of the compactness of $(\mathrm{Sper}R)_{\mathrm{con}}$ to the theory of relative real holomorphy rings, displayed in §2 and §3, by improving Corollary 3.11.

Theorem 4.2. Let Λ be any subring of the ring R. Given an element f of R, the following are equivalent.
(i) $f \in \mathrm{Hol}(R/\Lambda)$.
(ii) There exists some $\lambda \in \Lambda$ with $|f(P)| \le |\lambda(P)|$ for every $P \in \mathrm{Sper}R$.
(iii) There exists some $\mu \in \Lambda$ with $1 + \mu^2 \pm f \in P$ for every $P \in \mathrm{Sper}R$.

Proof. The implication (iii) $\Rightarrow$ (ii) is trivial, and (ii) $\Rightarrow$ (i) is obvious by Corollary 3.11.
(i) $\Rightarrow$ (iii): For every $P \in \mathrm{Sper}R$ we choose an element $\lambda_P \in P$ with $\lambda_P \pm f \in P$. This is possible by Corollary 3.11. Then also $1 + \lambda_P^2 \pm f \in P$. In other terms, $P \in \overline{H}(1 + \lambda_P^2 + f,\, 1 + \lambda_P^2 - f)$. Thus $\mathrm{Sper}R$ is covered by the sets $X_P\colon = \overline{H}(1 + \lambda_P^2 + f,\, 1 + \lambda_P^2 - f)$ with P running through $\mathrm{Sper}R$. Since $(\mathrm{Sper}R)_{\mathrm{con}}$ is compact, there exist finitely many points $P_1, \dots, P_r$ in $\mathrm{Sper}R$ such that

$$\mathrm{Sper}R = X_{P_1} \cup \cdots \cup X_{P_r}.$$

Let $\gamma\colon = \lambda_{P_1}^2 + \cdots + \lambda_{P_r}^2 \in \Lambda$. Clearly $1 + (1+\gamma)^2 \pm f \in P$ for every $P \in \mathrm{Sper}R$. $\square$

Applying the theorem to $\Lambda = \mathbb{Z}$ we obtain

Corollary 4.3. $\mathrm{Hol}(R)$ is the set of all $f \in R$ such that there exists some $n \in \mathbb{N}$ with $n \pm f \in P$, i.e. $|f(P)| \le n$, for every $P \in \mathrm{Sper}R$. $\square$

We return to the study of the space $\mathrm{Sper}R$ for R any ring. As in any topological space we say that a point $Q \in \mathrm{Sper}R$ is a *specialization* of a point $P \in \mathrm{Sper}R$ if Q lies in the closure $\overline{\{P\}}$ of the one-point set $\{P\}$.

Proposition 4.4. If P and Q are prime cones of R, then Q is a specialization of P (in $\mathrm{Sper}R$) iff $P \subset Q$.

Proof. $Q \in \overline{\{P\}}$ iff for every open subset U of $\mathrm{Sper}R$ with $Q \in U$ also $P \in U$. It suffices to know this for the $U \in \mathfrak{H}_R$. Thus $Q \in \overline{\{P\}}$ iff for every $f \in R$ with

$f(Q) > 0$ also $f(P) > 0$; in other terms, iff for every $g \in R$ with $g(P) \geq 0$ we have $g(Q) \geq 0$. {Take $g = -f$.} This means that $P \subset Q$. □

In the following P is a fixed prime cone of R. How do we obtain the prime cones $Q \supset P$? As in §3, let $\mathfrak{q}$ denote the support of P, $\mathfrak{q} = P \cap (-P)$. Recall from §3 that $\mathfrak{q}$ is the smallest P-convex additive subgroup of R.

Lemma 4.5. Let $\mathfrak{a}$ be a P-convex additive subgroup of R and $T{:}= P + \mathfrak{a}$. Then $T = P \cup \mathfrak{a}$ and $T \cap (-T) = \mathfrak{a}$.

Proof. i) Let $p \in P$ and $a \in \mathfrak{a}$. If $p + a \notin P$ then $-(p + a) \in P$ and $-a = p - (p + a) \in \mathfrak{a}$. Since $\mathfrak{a}$ is P-convex, it follows that $-(p + a) \in \mathfrak{a}$, hence $p + a \in \mathfrak{a}$. This proves that $T = P \cup \mathfrak{a}$.
ii) Of course, $\mathfrak{a} \subset T \cap (-T)$. Let $x \in T$ be given, and assume that $x \notin \mathfrak{a}$. Then, as just proved, $x \in P$. But $x \notin -P$ since $P \cap (-P) \subset \mathfrak{a}$. Thus $x \notin -T$. This proves that $T \cap (-T) = \mathfrak{a}$. □

Theorem 4.6. The prime cones $Q \supset P$ correspond uniquely with the P-convex prime ideals $\mathfrak{r}$ of R via

$$Q = P + \mathfrak{r} = P \cup \mathfrak{r}\,, \quad \mathfrak{r} = \operatorname{supp} Q.$$

Proof. a) If $\mathfrak{r}$ is a P-convex prime ideal of R then $Q{:}= P + \mathfrak{r}$ is closed under addition and multiplication and $Q \cup (-Q) = R$. By Lemma 5 we know that $Q \cap (-Q) = \mathfrak{r}$. Thus Q is a prime cone with support $\mathfrak{r}$. Also $Q = P \cup \mathfrak{r}$ by Lemma 5.
b) Let Q be a prime cone of R containing P. Then $\mathfrak{r}{:}= \operatorname{supp} Q$ is a Q-convex prime ideal of R. Since $P \subset Q$, it follows that $\mathfrak{r}$ is P-convex. We have $P + \mathfrak{r} \subset Q$. Let $f \in Q$ be given, and assume that $f \notin P$. Then $-f \in P \subset Q$, hence $f \in \mathfrak{r}$. We conclude that $Q \subset P \cup \mathfrak{r}$. Thus $Q = P + \mathfrak{r} = P \cup \mathfrak{r}$. □

As observed in §3, the P-convex prime ideals of R form a chain under the inclusion relation. We know by §3 that $I_P(R)$ is the maximal element of this chain (cf. Scholium 3.13). Thus we infer from Proposition 4 and Theorem 6 the following

Corollary 4.7. The specialisations of $P \in \operatorname{Sper} R$ form a chain under the specialisation relation. In other terms, if Q_1 and Q_2 are prime cones with $P \subset Q_1$ and $P \subset Q_2$, then $Q_1 \subset Q_2$ or $Q_2 \subset Q_1$. The maximal specialisation of P is

$$P^*{:}=\ P \cup I_P(R)\ =\ P + I_P(R).$$ □

Thus P^* is the unique closed point of $\operatorname{Sper} R$ in the set $\overline{\{P\}}$ of specialisations of P. We now analyze the situation that P itself is a closed point of $\operatorname{Sper} R$. This will give an answer to the question posed at the end of §3.

Definition 2. a) Let Λ be a subring of R. We say that R is *archimedean over* Λ *with respect to* P if $\mathrm{conv}_P(\Lambda) = R$, i.e. for every $f \in R$ there exists some $\lambda \in \Lambda$ with $|f(P)| \le |\lambda(P)|$.
b) If K is a real closed field and Λ a subring of K, we say that K *is archimedean over* Λ if this holds with respect to the unique ordering of K.

Theorem 4.8. Let P be a prime cone of R, and $\mathfrak{q} := \mathrm{supp}\, P$. The following are equivalent.
(i) P is a closed point of $\mathrm{Sper}R$.
(ii) $\mathfrak{q} = I_P(R)$.
(ii$'$) $\mathfrak{q}$ is the only proper P-convex ideal of R.
(iii) The field $k(\mathfrak{q})$ is archimedean over $R/\mathfrak{q}$ with respect to $\hat{P}$.
(iv) $k(P)$ is archimedean over $R/\mathfrak{q}$.
(v) The valuation v_P is special.

Proof. The equivalence (i) $\Leftrightarrow$ (ii) is evident from Corollary 7, and (ii) $\Leftrightarrow$ (ii$'$) follows from the general observation (cf. §3) that $I_P(R)$ is the biggest proper P-convex ideal of R while $\mathfrak{q}$ is the smallest one. The equivalence (iii) $\Leftrightarrow$ (iv) follows from the well known fact that $k(P)$ is archimedean over $k(\mathfrak{q})$ since $k(P)$ is algebraic over $k(\mathfrak{q})$.
(ii$'$) $\Leftrightarrow$ (iii): Recall that for every $f \in R$ the image of $f + \mathfrak{q}$ of f in $\overline{R} := R/\mathfrak{q}$ has been denoted by $f(P)$. Recall also that the ordering $\hat{P}$ induced by P on $k(\mathfrak{q})$ is just the restriction of the unique ordering of $k(P)$ to $k(\mathfrak{q})$. A general element of $k(\mathfrak{q})$ has the form $\frac{f(P)}{g(P)}$ with $f, g \in R$ and $g \notin \mathfrak{q}$. The field $k(\mathfrak{q})$ is archimedean over $\overline{R}$ with respect to $\hat{P}$ iff for every such elements f, g there exists some $h \in R$ with $\left|\frac{f(P)}{g(P)}\right| \le |h(P)|$. This property can also be stated as follows: $\mathrm{conv}_{\overline{P}}(\overline{g}\overline{R}) = \overline{R}$ for every $g \in R \setminus \mathfrak{q}$ where $\overline{g} := g + \mathfrak{q}$. Translating back to R we see that (iii) means that $\mathrm{conv}_P(gR) = R$ for every $g \in R \setminus \mathfrak{q}$. Clearly this holds iff $\mathfrak{q}$ is the only proper P-convex ideal of R.
(ii) $\Leftrightarrow$ (v): Let $v := v_P$ and $A := A_P = A_v$. We have $\mathfrak{p}_v = I_P$ and $\mathrm{supp}\, v = \mathrm{supp}\, P = \mathfrak{q}$. We first study the case that $A = R$. Now $I_P = I_P(R)$, and v is special iff v is trivial. This means that $\mathrm{supp}\, v = \mathfrak{p}_v$, i.e. $\mathfrak{q} = I_P(R)$ in our case.

From now on we may assume that $A \neq R$. By Scholium 3.13 we have

$$I_P(R) = \{x \in R \mid Rx \subset I_P\} \;=\; \{x \in R \mid \forall\, y \in R{:}\, v(xy) > 0\}.$$

Since there exists some $z \in R$ with $v(z) < 0$, it follows that

$$I_P(R) = \{x \in R \mid \forall\, y \in R{:}\, v(xy) \ge 0\} \;=\; \{x \in R \mid Rx \subset A\}.$$

Thus $I_P(R)$ is the conductor $\mathfrak{q}_A$ of R in A. Proposition I.2.2 tells us that v is special iff $\mathrm{supp}\, v = \mathfrak{q}_A$. This means $\mathfrak{q} = I_P(R)$ in our case. $\square$

Taking into account the study of real valuations in §3 we obtain

Corollary 4.9. Assume that R is convenient over $\mathrm{Hol}(R)$. Then the nontrivial real PM-valuations on R are precisely the coarsenings of the valuations v_P with P running through the closed points of $\mathrm{Sper}R$. $\square$

Lemma 4.10. Assume that P and Q are prime cones of R with $P \subset Q$.
a) For every subring Λ of R, we have $\mathrm{conv}_P(\Lambda) = \mathrm{conv}_Q(\Lambda)$ and $I_P(\Lambda) = I_Q(\Lambda)$. In particular, choosing $\Lambda = \mathbb{Z}$, we have $A_P = A_Q$ and $I_P = I_Q$.
b) If M is any additive subgroup of R then $\mathrm{conv}_P(M) = \mathrm{conv}_Q(M)$.

Proof. a): First notice that for any elements $f \in R$, $g \in R$ we have $|f(P)| < |g(P)|$ iff $(g^2 - f^2)(P) > 0$ and $|f(P)| \le |g(P)|$ iff $(g^2 - f^2)(P) \ge 0$. Thus $|f(Q)| < |g(Q)|$ implies $|f(P)| < |g(P)|$, and $|f(P)| \le |g(P)|$ implies $|f(Q)| \le |g(Q)|$. The assertions now follow from the various ways to characterize the elements of $\mathrm{conv}_P(\Lambda)$, $I_P(\Lambda)$, $\dots$ either by weak inequalities ($\le$) or by strong inequalities ($<$), cf. Remark 4.1 above.
b): This can be proved in the same way. $\square$

Definition 3. a) If $v\colon R \to \Gamma \cup \infty$ is any valuation on R we denote the valuation $v|c_v(\Gamma)\colon R \to c_v(\Gamma)$ (cf. notations in I, §2) by v^*, and we call v^* the *special valuation associated to* v. {N.B. We have $v^* = v|_R$.}
b) If P is any prime cone on R we denote the maximal specialisation of P in $\mathrm{Sper}R$ (i.e. the unique closed point of $\overline{\{P\}}$) by P^*, as we did already above (Corollary 7).

Proposition 4.11. Assume that R is convenient over $\mathrm{Hol}(R)$. Given a prime cone P of R, the valuations $(v_P)^*$ and v_{P^*} are equivalent.

Proof. Let $v := v_P$, $u := v_{P^*}$. By Theorem 8 we know that u is special. By Lemma 10.a we have

$$A_v = A_P = A_{P^*} = A_u \quad , \quad \mathfrak{p}_v = I_P = I_{P^*} = \mathfrak{p}_u .$$

Both u and v^* are special valuations on R over $\mathrm{Hol}(R)$, hence are PM-valuations. We have $A_{v^*} = A_v = A_u$, $\mathfrak{p}_{v^*} = \mathfrak{p}_v = \mathfrak{p}_u$. We conclude (by I, §2) that u and v^* are equivalent. $\square$

Open problem. Does $(v_P)^* \sim v_{P^*}$ hold for any ring R and prime cone P of R?

Example 4.12 (The real spectra of $C(X)$ and $C_b(X)$). Let X be a completely regular Hausdorff space. Then the ring $R := C(X)$ is real closed in the sense of Schwartz (cf. [Sch], [Sch$_1$]). This implies that the support map $\mathrm{supp}\colon \mathrm{Sper}R \to \mathrm{Spec}\,R$ is a homeomorphism (loc.cit.). By restriction we obtain a bijection from the set $(\mathrm{Sper}R)^{\max}$ of closed points of $\mathrm{Sper}R$ to the set of closed points $(\mathrm{Spec}\,R)^{\max} = \mathrm{Max}R$ of $\mathrm{Spec}\,R$. On the other hand we have a bijection $\beta X \xrightarrow{\sim} \mathrm{Max}R$, $p \mapsto M^p$ (cf.1.4 above).

Let us regard βX as the set of ultrafilters α on the lattice $\mathcal{Z}(X)$. By what has been said there corresponds to each ultrafilter $\alpha \in \beta X$ a unique prime cone P_α of R with $\operatorname{supp} P_\alpha = M^\alpha$. We now describe this prime cone P_α. If $f \in R$ is given then both the sets $\{f \geq 0\} := \{x \in X \mid f(x) \geq 0\}$ and $\{-f \geq 0\}$ are elements of $\mathcal{Z}(X)$, and their union is X. Thus at least one of these sets is an element of α. Let

$$P := \{f \in R \mid \{f \geq 0\} \in \alpha\}.$$

Then we know already that $P \cup (-P) = R$. Clearly $P + P \subset P$ and $P \cdot P \subset P$. Also

$$P \cap (-P) = \{f \in R \mid Z(f) \in \alpha\} = M^\alpha$$

(cf.[GJ,§6]). Thus P is a prime cone of R with support M^α. We conclude that $P = P_\alpha$.

If α is not an ultrafilter but just a prime filter on the lattice $\mathcal{Z}(X)$ then we still see as above that

$$P_\alpha := \{f \in R \mid \{f \geq 0\} \in \alpha\}$$

is a prime cone on R. But not every prime cone of R is one of these P_α. The map $\alpha \mapsto P_\alpha$ is a bijection from the set of prime filters on $\mathcal{Z}(X)$ to a proconstructible subset of $\operatorname{Sper} R$, the so called *real z-Spectrum* z-$\operatorname{Sper} R$, cf.[Sch$_3$]. Under the support map we have a homeomorphism from z-$\operatorname{Sper} R$ to the space z-$\operatorname{Spec} R$ constisting of the z-prime ideals of R, which have already much been studied in [GJ].

The ring $A := C_b(X)$ of bounded continuous real functions on X is again real closed. But now the situation is simpler. We have a bijection $\beta X \xrightarrow{\sim} \operatorname{Max} A$, $\alpha \mapsto \mathfrak{m}_\alpha$ (cf.1.4) and a bijection $(\operatorname{Sper} A)^{\max} \xrightarrow{\sim} \operatorname{Max} A$ by the support map. Thus to every $\alpha \in \beta X$ there corresponds a unique prime cone $P'_\alpha \in (\operatorname{Sper} A)^{\max}$ with $\operatorname{supp} P'_\alpha = \mathfrak{m}_\alpha$. We have

$$\mathfrak{m}_\alpha = \{f \in A \mid f^\beta(\alpha) = 0\}$$

and guess easily that

$$P'_\alpha = \{f \in A \mid f^\beta(\alpha) \geq 0\}.$$

Also $A/\mathfrak{m}_\alpha = \mathbb{R}$, hence $k(P_\alpha) = \mathbb{R}$. Clearly $A \cap P_\alpha \subset P'_\alpha$. Thus P'_α is the maximal specialization of $A \cap P_\alpha$ in the real spectrum $\operatorname{Sper} A$, i.e. $P'_\alpha = (A \cap P_\alpha)^*$.

Example 4.13 (The special real valuations and the real holomorphy ring of $C(X)$). Let again X be a complete regular Hausdorff space, $R := C(X)$, $A := C_b(X)$. We retain the notations from 4.12. For every $\alpha \in \beta X$ we denote the valuation v_{P_α} more briefly by v_α. Since P_α is a closed point of $\operatorname{Sper} R$, this valuation is special. Now $1 + R^2 \subset R^*$. Thus we know, say by §2, that $\operatorname{Hol}(R)$ is Prüfer in R. This implies that every v_α is a PM-valuation, hence

$v_\alpha \in pm(R/\mathrm{Hol}(R))$. Corollary 9 tells us that $pm(R/\mathrm{Hol}(R))$ is the set of coarsenings of the valuations v_α with α running through βX.

Let $A_\alpha := A_{v_\alpha}$ and $I_\alpha := \mathfrak{p}_{v_\alpha}$. We know by Lemma 3.3 that

$$A_\alpha = \{f \in R \mid \exists\, n \in \mathbb{N}: \quad n \pm f \in P_\alpha\},$$

$$I_\alpha = \{f \in R \mid \forall\, n \in \mathbb{N}: \quad \frac{1}{n} \pm f \in P_\alpha\}.$$

For every $f \in R$ and $n \in \mathbb{N}$ we introduce the set

$$\begin{aligned} Z_n(f) &:= \{x \in X \mid n + f(x) \geq 0\} \cap \{x \in X \mid n - f(x) \geq 0\} \\ &= \{x \in X \mid |f(x)| \leq n\}. \end{aligned}$$

From the description of P_α above we read off that $f \in A_\alpha$ iff $Z_n(f)$ is an element of the ultrafilter α for some $n \in \mathbb{N}$. Thus A_α coincides with the subring A_α of R as defined in 2.1. In the same way we see that I_α is the ideal of A_α considered there and that $\mathrm{supp}\,(v_\alpha)$ is the ideal $\mathfrak{q}_\alpha$ of R considered there.

Using 2.1 we conclude that v_α is the PM-valuation of R over A corresponding to the prime ideal $\mathfrak{m}_\alpha$ of A. Thus $pm(R/\mathrm{Hol}R) = pm(R/A)$. This forces $\mathrm{Hol}(R) = A$. Using also 1.4 we conclude that

$$\omega(R/A) = \{v_\alpha \mid \alpha \in \beta X \setminus \upsilon X\}.$$

The result $\mathrm{Hol}(R) = A$ can also be verified as follows, using less information about the real valuations on R: We know by Corollary 3 above that a given element f of R is in $\mathrm{Hol}(R)$ iff there exists some $n \in \mathbb{N}$ such that $n \pm f \in P$ for every $P \in \mathrm{Sper}R$. Here we may replace $\mathrm{Sper}R$ by $(\mathrm{Sper}R)^{\max}$. Thus we see that $f \in \mathrm{Hol}(R)$ iff there exists some $n \in \mathbb{N}$ with $Z_n(f) \in \alpha$ for every ultrafilter α of the lattice $\mathcal{Z}(X)$. This means that $Z_n(f) = X$ for some $n \in \mathbb{N}$, i.e. f is bounded. $\square$

§5 Convexity of subrings and of valuations

Let R be any ring. A subset T of R is called a *preordering* of R (or: a *cone* of R [BCR, p.86]), if T is closed under addition and multiplication and contains the set $R^2 = \{x^2 \mid x \in R\}$. We call a preordering T *proper* if $-1 \notin T$.

We associate with a preordering T of R a binary relation $\leq_T$ on R, defined by

$$f \leq_T g \iff g - f \in T.$$

This relation is transitive and reflexive but in general not antisymmetric. We define the *support* of T as the set

$$\operatorname{supp} T \ = \ T \cap (-T).$$

This is an additive subgroup of R. Clearly $f \leq_T g$ and $g \leq_T f$ iff $f - g \in \operatorname{supp} T$.

Of course, the prime cones $P \in \operatorname{Sper} R$ are preorderings, but there are many more. The intersection of any family of preorderings is again a preordering. In particular R has a smallest preordering, which we denote by T_0. Clearly $T_0 = \Sigma R^2$.

In the following T is a fixed preordering of R.

Definition 1. a) A subset M of R is called T-*convex* (in R) if for any three elements x, y, z of R with $x \leq_T y \leq_T z$ and $x \in M$, $z \in M$, also $y \in M$.
b) If U is any subset of R there clearly exists a smallest T-convex subset M of R containing U. We call M the T-*convex hull* of U, and we write $M = \operatorname{conv}_T(U)$. $\square$

Remark. An additive subgroup M of R is T-convex iff for all $s \in T$, $t \in T$ with $s + t \in M$ we have $s \in M$ and (hence) $t \in M$. $\square$

It is obvious that $\operatorname{supp} T = T \cap (-T)$ is the smallest T-convex additive subgroup of R. Notice also that the set $T - T$, consisting of the differences $t_1 - t_2$ of elements t_1, t_2 of T, is a T-convex subring of R, and that $\operatorname{supp} T$ is an ideal of the ring $T - T$.

If 2 is a unit in R we have $T - T = R$, as follows from the identity

$$x = 2\left[\left(\frac{1+x}{2}\right)^2 - \left(\frac{x}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right].$$

Later only rings with 2 a unit will really matter, but we can avoid this assumption here by enlarging T slightly.

LEMMA 5.1. $T' := \{x \in R \mid \exists\, n \in \mathbb{N} : 2^n x \in T'\}$ is again a preordering of R. It is proper iff T is proper. $\square$

We omit the easy proof. We call T' the *2-saturation* of T, and we call T *2-saturated* if $T' = T$.

If T is 2-saturated, then $\operatorname{supp} T$ is an ideal of R due to the identity

$$2xy = (1+x)^2 y - x^2 y - y.$$

Of course, if $2 \in R^*$, then every preordering of R is 2-saturated. Notice also that every prime cone is a 2-saturated preordering.

Given a subring Λ and a preordering T of R we strive for an understanding and a handy description of the convex hull $\operatorname{conv}_T(\Lambda)$ of Λ in R. We introduce a new notation for this,

$$C(T, R/\Lambda) :=\ \operatorname{conv}_T(\Lambda),$$

which reflects that $\operatorname{conv}_T(\Lambda)$ also depends on the ambient ring R. It is easily seen that $C(T, R/\Lambda)$ is the set of all $x \in R$ with $\lambda_1 \underset{T}{\leq} x \underset{T}{\leq} \lambda_2$ for some elements λ_1, λ_2 of Λ. From this it is immediate that $C(T, R/\Lambda)$ is an additive subgroup of R. We also introduce the set

$$\begin{aligned} A(T, R/\Lambda) :&= \{x \in R | \exists\, \lambda \in T \cap \Lambda : \lambda \pm x \in T\} \\ &= \{x \in R | \exists\, \lambda \in T \cap \Lambda : -\lambda \leq_T x \leq_T \lambda\}. \end{aligned}$$

We use the abbreviations $C(T, R) := C(T, R/\mathbb{Z}1_R)$ and
$A(T, R) := A(T, R/\mathbb{Z}1_R)$.

Given an additive subgroup M of R let M' denote the 2-saturation of M in R, i.e. the additive group consisting of all $x \in R$ such that $2^n x \in M$ for some $n \in \mathbb{N}_0$. If M is a subring of R then also M' is a subring of R.

PROPOSITION 5.2. a) $A(T, R/\Lambda)$ is a T-convex subring of R contained in $C(T, R/\Lambda)$.
b) $C(T, R/\Lambda) = \Lambda + A(T, R/\Lambda)$.
c) $C(T, R/\Lambda) = A(T, R/\Lambda)$ iff Λ is generated by $\Lambda \cap T$ as an additive group, i.e., $\Lambda = (\Lambda \cap T) - (\Lambda \cap T)$.
d) $C(T, R) = A(T, R)$, and this is the smallest T-convex subring of R.
e) If T contains the 2-saturated hull T_0' of $T_0 = \Sigma R^2$ (e.g. T itself is 2-saturated), then $C(T, R/\Lambda) = A(T, R/\Lambda)$.
f) Without any extra assumption on T and Λ we have $A(T, R/\Lambda)' = C(T, R/\Lambda)'$ $= A(T', R/\Lambda) = C(T', R/\Lambda)$.

Proof. a) We first prove that $A(T, R/\Lambda)$ is a subring of R. Given elements x and y of $A(T, R/\Lambda)$ we choose elements λ and μ in $\Lambda \cap T$ such that $\lambda \pm x \in T$ and $\mu \pm y \in T$. Then we have

$$(\lambda + \mu) \pm (x - y) \in T,$$

which proves that $x - y \in A(T, R/\Lambda)$.

Moreover we have

$$(\lambda + x)(\mu + y) = \lambda\mu + \lambda y + \mu x + xy \in T$$

and

$$\lambda(\mu - y) \in T \quad , \quad \mu(\lambda - x) \in T.$$

By adding we obtain

$$3\lambda\mu + xy \in T.$$

Replacing x by $-x$ we obtain $3\lambda\mu - xy \in T$. This proves that $xy \in A(T, R/\Lambda)$.

Thus $A(T, R/\Lambda)$ is a subring of R. It is clear from the definition of $A(T, R/\Lambda)$ that this ring is contained in the T-convex hull $C(T, R/\Lambda)$ of Λ in R. Given elements x_1, x_2 of $A(T, R/\Lambda)$ and $y \in R$ with $x_1 \leq_T y \leq_T x_2$, we have elements λ_1, λ_2 of $\Lambda \cap T$ such that $-\lambda_1 \leq_T x_1 \leq_T \lambda_1$ and $-\lambda_2 \leq_T x_1 \leq_T \lambda_2$. These inequalities imply

$$-(\lambda_1 + \lambda_2) \;\leq_T\; x_1 \;\leq_T\; y \;\leq_T\; x_2 \;\leq_T\; (\lambda_1 + \lambda_2).$$

Thus $y \in A(T, R/\Lambda)$. This proves that $A(T, R/\Lambda)$ is T-convex in R.

b): It is evident that the additive group $M := \Lambda + A(T, R/\Lambda)$ is contained in $C(T, R/\Lambda)$. We are done if we verify that M is T-convex in R.

Let $s, t \in T$ be given with $s + t \in M$, hence $s + t = \lambda + x$ with $\lambda \in \Lambda$, $x \in A(T, R/\Lambda)$. We have $0 \leq_T s \leq_T \lambda + x$. There exists some $\mu \in \Lambda$ with $x \leq_T \mu$. Then $0 \leq_T s \leq_T \lambda + \mu$, and thus $\lambda + \mu \in \Lambda \cap T$. This proves that $s \in A(T, R/\Lambda) \subset M$.

c): $A(T, R/\Lambda) = C(T, R/\Lambda)$ means that $\Lambda \subset A(T, R/\Lambda)$. This is certainly true if $\Lambda = (\Lambda \cap T) - (\Lambda \cap T)$, since $\Lambda \cap T \subset A(T, R/\Lambda)$ by definition of $A(T, R/\Lambda)$.

It remains to verify that the inclusion $\Lambda \subset A(T, R/\Lambda)$ implies $\Lambda = (\Lambda \cap T) - (\Lambda \cap T)$. Let $\lambda \in \Lambda$ be given. There exists some $\mu \in \Lambda \cap T$ such that $\mu \pm \lambda \in T$. Then $\lambda = \mu - (\mu - \lambda)$, and both $\mu, \mu - \lambda \in \Lambda \cap T$.

d): Applying c) to $\Lambda = \mathbb{Z} \cdot 1_R$ we see that $C(T, R) = A(T, R)$. By definition $C(T, R)$ is the smallest T-convex additive subgroup of R containing 1_R, hence also the smallest T-convex subring of R.

e): For every $\lambda \in \Lambda$ we have

$$2(\lambda^2 + 1 \pm \lambda) = \lambda^2 + 1 + (\lambda \pm 1)^2 \in T_0,$$

hence $\lambda^2 + 1 \pm \lambda \in T_0' \subset T$. This implies $\Lambda \subset A(T, R/\Lambda)$, hence $C(T, R/\Lambda) = A(T, R/\Lambda)$.

f): We first verify that $A(T, R/\Lambda)' = A(T', R/\Lambda)$. Given $x \in A(T, R/\Lambda)'$, we have some $n \in \mathbb{N}$ with $2^n x \in A(T, R/\Lambda)$, hence $\lambda \pm 2^n x \in T$ for some $\lambda \in T \cap \Lambda$. It follows that $2^n(\lambda \pm x) \in T$, hence $\lambda \pm x \in T'$, hence $x \in A(T', R/\Lambda)$.

Conversely, if $x \in A(T', R/\Lambda)$ we have some $\lambda \in T' \cap \Lambda$ with $\lambda \pm x \in T'$ and then some $n \in \mathbb{N}$ with $2^n \lambda \in T \cap \Lambda$ and $2^n \lambda \pm 2^n x \in T$. Thus $2^n x \in A(T, R/\Lambda)$, and $x \in A(T, R/\Lambda)'$.

This completes the proof that $A(T, R/\Lambda)' = A(T', R/\Lambda)$. Now observe that $A(T, R/\Lambda) \subset C(T, R/\Lambda) \subset C(T', R/\Lambda)$. As proved above, $C(T', R/\Lambda) = A(T', R/\Lambda) = A(T, R/\Lambda)'$. In particular we know that $C(T', R/\Lambda)$ is 2-saturated. It follows that

$$A(T, R/\Lambda)' \subset C(T, R/\Lambda)' \subset C(T', R/\Lambda) = A(T, R/\Lambda)'.$$

Thus the groups $A(T, R/\Lambda)'$, $A(T', R/\Lambda)$, $C(T, R/\Lambda)'$, $C(T', R/\Lambda)$ are all the same. $\square$

We aim at a description of the rings between $\mathrm{Hol}(R)$ and R by T-convexity for varying preorderings T in the case that $\mathrm{Hol}(R)$ is Prüfer in R. Here preorderings will play a dominant role which are "saturated" in the sense of the following definition.

Definition 2. The *saturation* $\hat{T}$ of a preordering T of R is the intersection of all prime cones $P \supset T$ of R. In other terms,

$$\hat{T} = \{f \in R \mid \forall\, P \in \overline{H}_R(T) \colon f(P) \geq 0\}.$$

T is called *saturated* if $\hat{T} = T$. $\square$

Of course, $\hat{T}$ is always 2-saturated. More generally $\hat{T}$ is saturated with respect to the multiplicative subset $1 + T$ of R, i.e. for any $x \in R$, $t \in T$:

$$(1 + t)x \in \hat{T} \Longrightarrow x \in \hat{T}.$$

Notice that the saturation $\hat{T}_0$ of $T_0 = \Sigma R^2$ is the set of all $f \in R$ which are nonnegative on $\mathrm{Sper} R$. Thus, taking into account Proposition 2, the description of $\mathrm{Hol}(R/\Lambda)$ in Theorem 4.2 can be read as follows.

Scholium 5.3. For any ring extension $\Lambda \subset R$

$$\mathrm{Hol}(R/\Lambda) = A(\hat{T}_0, R/\Lambda) = C(\hat{T}_0, R/\Lambda). \qquad \square$$

Every proper preordering of a field is saturated, as is very well known ([BCR, p.9], [KS, p.2]). In the field case we have $T \cap (-T) = \{0\}$. Then a proper preordering is a partial ordering of the field in the usual sense.

We recall without proof the famous *abstract Positivstellensatz* about an algebraic description of $\hat{T}$ in terms of T for R an arbitrary ring.

Theorem 5.4. (cf.[BCR, p.92], [KS, p.143]). If T is any preordering of R and $a \in R$, the following are equivalent.
(1) $a \in \hat{T}$.
(2) $-a^{2n} \in T - aT$ for some $n \in \mathbb{N}_0$.
(3) There exist $t, t' \in T$ and $n \in \mathbb{N}_0$ with $a(a^{2n} + t) = t'$. □

The theorem tells us in particular (take $a = -1$) that for T proper, i.e. $-1 \notin T$, also $\hat{T}$ is proper. It follows that for a proper preordering T there always exists some prime cone $P \supset T$.

In order to get a somewhat "geometric" understanding of saturated preorderings we introduce more terminology.

Definitions 3. a) Given any subset X of Sper$R$, let $P(X)$ denote the intersection of the prime cones $P \in X$. In other terms,

$$P(X) := \{f \in R \mid \forall x \in X\colon f(x) \geq 0\}.$$

In particular, for every $x \in X$, $P(\{x\})$ is the point x itself, viewed as a prime cone, $P(\{x\}) = P_x$.
b) We call a subset X of Sper$R$ *basic closed*, if

$$X = \overline{H}_R(\Phi) = \{x \in \mathrm{Sper}R \mid f(x) \geq 0 \quad \text{for every} \quad f \in \Phi\}$$

for some subset Φ of R, i.e. X is the intersection of a family of "principal closed" sets $\overline{H}_R(f) = \{x \in \mathrm{Sper}R \mid f(x) \geq 0\}$. □
c) If X is any subset of Sper$R$, let $\hat{X}$ denote the smallest basic closed subset of Sper$R$ containing X, i.e. the intersection of all principal closed sets $\overline{H}_R(f)$ containing X. We call $\hat{X}$ the *basic closed hull* of X.
d) If Φ is any subset of R, there exists a smallest preordering T containing Φ. This is the semiring generated by $\Phi \cup R^2$ in R. We call T the *preordering generated by* Φ, and write $T = T(\Phi)$.

Remarks 5.5. i) For every $X \subset \mathrm{Sper}R$ the set $P(X)$ is a saturated preordering of R and $\overline{H}_R(P(X)) = \hat{X}$. It follows that $P(\hat{X}) = P(X)$. Moreover $\hat{X}$ is the unique maximal subset Y of Sper$R$ with $P(Y) = P(X)$.
ii) If Φ is any subset of R then $\overline{H}_R(\widehat{T(\Phi)}) = \overline{H}_R(\Phi)$. Moreover $\widehat{T(\Phi)}$ is the unique maximal subset U of R with $\overline{H}_R(U) = \overline{H}_R(\Phi)$.
iii) The basic closed subsets Z of Sper$R$ correspond uniquely with the saturated preorderings T of R via $T = P(Z)$ and $Z = \overline{H}_R(T)$.

All this can be verified easily in a straightforward way. □

If X is any subset of Sper$R$ we call a $P(X)$-*convex subset* M of R also X-*convex*. In the case that term X is a one-point set $\{x\}$, we use the term "x-convex". {Thus x-convexity is the same as P-convexity for $P = x$, regarded as prime cone.}

Instead of $A(P(X), R/\Lambda)$ we write $A_X(R/\Lambda)$. Thus

$$A_X(R/\Lambda) = \{f \in R \mid \exists\, \lambda \in \Lambda \quad \text{such that} \quad |f(x)| \leq \lambda(x) \quad \text{for every } x \in X\}.$$

{Read $A_X(R/\Lambda) = R$ if X is empty.} By Proposition 2 we have $C(P(X), R/\Lambda) = A_X(R/\Lambda)$.

Let again T be any preordering of a ring R. There exists a by now well known and well developed theory of T-convex prime ideals of R which we will need below (cf. [Br], [Br$_1$], [KS, Chap.III, §10]). The main result can be subsumed in the following theorem.

Theorem 5.6. a) Let T be a proper preordering of R and $\mathfrak{p}$ a prime ideal of R. Then $\mathfrak{p}$ is T-convex iff $\mathfrak{p}$ is $\hat{T}$-convex. In this case there exists a prime cone $P \supset T$ such that $\mathfrak{p}$ is P-convex.
b) Let X be a closed subset of Sper$R$. The X-convex prime ideals of R are precisely the supports supp$\,(P)$ of the prime cones $P \in X$. □

We do not give the proof here,*) refering the reader to [KS, Chap.III, §10] for this, but we state two key observations leading to the theorem.

Proposition 5.7 ([KS, p.148]). Let T be any preordering of R. The maximal proper T-convex ideals of R are the ideals $\mathfrak{a}$ of R which are maximal with the property $\mathfrak{a} \cap (1 + T) = \emptyset$. They are prime. □

{N.B. This holds also in the case that $-1 \in T$. Then R itself is the only T-convex ideal of R.}

Proposition 5.8 (A. Klapper, cf. [Br, p.63], [KS, p.149]). Let T_1 and T_2 be preorderings of R and $\mathfrak{p}$ a prime ideal of R. Assume that $\mathfrak{p}$ is $(T_1 \cap T_2)$-convex. Then $\mathfrak{p}$ is T_1-convex or T_2-convex. □

For later use we also mention

Lemma 5.9. Let T be a proper preordering of R and $\mathfrak{a}$ a T-convex proper ideal of R. Then $T_1 := T + \mathfrak{a}$ is again a proper preordering of R and $T_1 \cap (-T_1) = \mathfrak{a}$. The image $\overline{T} = T_1/\mathfrak{a}$ of T in $R/\mathfrak{a}$ is a proper preordering of $R/\mathfrak{a}$, and $\overline{T} \cap (-\overline{T}) = \{0\}$.

*) In fact part a) will be proved below as a special case of Theorem 16.

We leave the easy proof to the reader. □

As before let T be a fixed preordering of R.

Definition 4. We say that a *valuation* $v: R \to \Gamma \cup \infty$ is *T-convex* if the prime ideal $\operatorname{supp} v$ is T-convex in R and, for every $\gamma \in \Gamma$, the additive group $I_{\gamma,v} = \{x \in R \mid v(x) \geq \gamma\}$ is T-convex in R. In other terms, v is T-convex iff for any elements x, y of R with $0 \leq_T y \leq_T x$ we have $v(y) \geq v(x)$. If $T = P(X)$ for some set $X \subset \mathrm{Sper} R$, we also use the term "$X$-convex" instead of T-convex.

Comment. In the – not very extended – literature these valuations are usually called "compatible with T". The term "T-convex" looks more imaginative, in particular if one follows the philosophy (as we do) that valuations are refinements of prime ideals.

Several observations on real valuations stated in §2 extend readily to T-convex valuations.

Remarks 5.10. Let $v: R \to \Gamma \cup \infty$ be a valuation.
i) The following are clearly equivalent.
  (1) v is T-convex.
  (2) If $x \in T$ and $y \in T$ then $v(x) \geq v(x+y)$.
  (3) If $x \in T$ and $y \in T$ then $v(x+y) = \min(v(x), v(y))$.
In particular, v is T_0-convex iff v is real (cf. Prop.2.2.). Every T-convex valuation is real.
ii) If T is improper, i.e. $-1 \in T$, there do not exist T-convex valuations.
iii) If v is trivial then v is T-convex iff $\operatorname{supp} v$ is T-convex in R. The T_0-convex prime ideals are just the real prime ideals.
iv) If v is T-convex, both A_v and $\mathfrak{p}_v$ are T-convex in R.
v) Assume that v is T-convex. For every convex subgroup H of Γ the coarsening v/H is again T-convex. If H contains the characteristic subgroups $c_v(\Gamma)$ then also $v|H$ is T-convex.
vi) If B is a subring of R and v is T-convex, then both the valuations $v|B$ and $v|_B$ are $(T \cap B)$-convex. □

In the case of Manis valuations we have very handy criteria for T-convexity.

Theorem 5.11. Let v be a Manis valuation on R.
i) The following are equivalent.
  (1) v is T-convex.
  (2) $\mathfrak{p}_v$ is T-convex in R.
  (3) $\mathfrak{p}_v$ is $(T \cap A_v)$-convex in A_v.
ii) If v is non trivial, then (1) – (3) are also equivalent to
  (4) A_v is T-convex in R.

Proof. If v is trivial the equivalence of (1), (2), (3) is evident. Henceforth we assume that v is not trivial. The implications (1) $\Rightarrow$ (2) and (1) $\Rightarrow$ (4) are evident from the definition of T-convexity of valuations (cf. Def.4 above). The implication (2) $\Rightarrow$ (3) is trivial.

(4) $\Rightarrow$ (1): Assume that A_v is T-convex in R. Let $x, y \in R$ be given with $0 \leq_T y \leq_T x$ and $v(x) \neq \infty$. We choose some $z \in R$ with $v(xz) = 0$. This is possible since v is Manis. We have $0 \leq_T (yz)^2 \leq_T (xz)^2$. {Notice that $x^2 - y^2 = (x-y)(x+y) \in T$.} Since $(xz)^2 \in A_v$ and A_v is T-convex, it follows that $(yz)^2 \in A_v$, hence $2v(yz) \geq 0$, hence $v(y) \geq -v(z) = v(x)$. This proves that $I_{v,\gamma}$ is T-convex in R for every $\gamma \in \Gamma_v$. The support of v is the intersection of all these $I_{v,\gamma}$, since v is not trivial. Thus $\operatorname{supp} v$ is T-convex in R. This finishes the proof that v is T-convex.

(2) $\Rightarrow$ (4): Assume that $\mathfrak{p}_v$ is T-convex in R. Since v is Manis we have $A_v = \{x \in R \mid x\mathfrak{p}_v \subset \mathfrak{p}_v\}$. Let $0 \leq_T y \leq_T x$ and $x \in A_v$. For every $z \in \mathfrak{p}_v$ this implies $0 \leq_T (yz)^2 \leq_T (xz)^2 \in \mathfrak{p}_v$. Since $\mathfrak{p}_v$ is T-convex in R, we conclude that $(yz)^2 \in \mathfrak{p}_v$, and then that $yz \in \mathfrak{p}_v$. This proves that $y\mathfrak{p}_v \subset \mathfrak{p}_v$, hence $y \in A_v$.

(3) $\Rightarrow$ (2): Assume that $\mathfrak{p}_v$ is $(T \cap A_v)$-convex in A_v. We verify that $\mathfrak{p}_v$ is T-convex in R. Let $x \in \mathfrak{p}_v$ and $y \in R$ be given with $0 \leq_T y \leq_T x$. Suppose that $y \notin \mathfrak{p}_v$, i.e. $v(y) \leq 0$. We choose some $z \in R$ with $v(yz) = 0$. Then $0 \leq_T (yz)^2 \leq_T (xz)^2$. Now $z \in A_v$, hence $(xz)^2 \in \mathfrak{p}_v$, and $(yz)^2 \in A_v$. It follows that $(yz)^2 \in \mathfrak{p}_v$, hence $yz \in \mathfrak{p}_v$. This contradicts $v(yz) = 0$. Thus $\mathfrak{p}_v$ is indeed T-convex in R. $\square$

Another proof of Theorem 11 can be found in [Z_1, §2].

Corollary 5.12. Let U be a preordering (= partial ordering) of a field K. A valuation v on K is U-convex iff the valuation domain A_v is U-convex in K.

Proof. v is Manis. If v is nontrivial the claim is covered by Theorem 11.ii. If v is trivial, $\mathfrak{p}_v = \operatorname{supp} v = \{0\}$, which is U-convex. Now the claim is covered by Theorem 11.i. $\square$

Corollary 5.13. Assume that T and U are preorderings on R and that v is a Manis valuation on R which is $(T \cap U)$-convex. Then v is T-convex or U-convex.

Proof. We work with condition (3) in Theorem 11. We know that $\mathfrak{p}_v$ is $(T \cap U \cap A_v)$-convex in A_v, and we conclude that $\mathfrak{p}_v$ is $T \cap A_v$-convex or $U \cap A_v$-convex in A_v by Proposition 8 above. $\square$

Returning to valuations which are not necessarily Manis we now prove a lemma by which the study of T-convex valuations on R can be reduced to the study of U-convex valuations for preorderings U on suitable residue class fields of R.

Lemma 5.14. Let T be a proper preordering of R and v a valuation on R. We assume that $\mathfrak{q} := \operatorname{supp} v$ is T-convex.
i) $T_1 := T + \mathfrak{q}$ is a proper preordering of R and $T_1 \cap (-T_1) = \mathfrak{q}$.
ii) Let $\overline{T} := T_1/\mathfrak{q}$ denote the image of T and of T_1 in $\overline{R} := R/\mathfrak{q}$. Then the subset

$$U := \left\{ \frac{\overline{x}}{\overline{s}^2} \mid x \in T,\ s \in R \setminus \mathfrak{q} \right\}$$

of the field $k(\mathfrak{q})$ is a proper preordering (= partial ordering) of $k(\mathfrak{q})$, and $T_2 := j_{\mathfrak{q}}^{-1}(U)$ is a proper preordering of R. {Here, of course, $\overline{x} := j_{\mathfrak{q}}(x)$, $\overline{s} := j_{\mathfrak{q}}(s)$, the images of x and s in $k(\mathfrak{q})$.} We have $T \subset T_1 \subset T_2$ and $T_2 \cap (-T_2) = \mathfrak{q}$.
iii): $T_2 = \{x \in R \mid \exists\, s \in R \setminus \mathfrak{q}\colon\ s^2 x \in T_1\}$.
iv) As always (cf.I, §1) we denote the valuations induced by v on $\overline{R}$ and $k(\mathfrak{q})$ by $\overline{v}$ and $\hat{v}$ respectively. The following are equivalent:
(1) v is T-convex.
(2) $\overline{v}$ is $\overline{T}$-convex.
(3) $\hat{v}$ is U-convex.
(4) v is T_2-convex.
(5) v is T_1-convex.

Proof. i): This is covered by Lemma 9 above.
ii): We know by Lemma 9 that $\overline{T}$ is a preordering of $\overline{R}$ with $\overline{T} \cap (-\overline{T}) = \{0\}$. It then is a straightforward verification that U is a proper preordering of $k(\mathfrak{q})$. We have $\overline{T} \subset U \cap \overline{R}$, hence $T_1 = j_{\mathfrak{q}}^{-1}(\overline{T}) \subset j_{\mathfrak{q}}^{-1}(U) =: T_2$. Also $T_2 \cap (-T_2) = j_{\mathfrak{q}}^{-1}(U \cap (-U)) = \mathfrak{q}$.
iii): An easy verification.
iv): (1) $\Leftrightarrow$ (2) is completely obvious by using, say, condition (3) in Remark 10.i characterizing convexity of valuations. The implications (4) $\Rightarrow$ (5) $\Rightarrow$ (1) are trivial since $T \subset T_1 \subset T_2$, and (3) $\Rightarrow$ (4) is immediate, due to the fact that $v = \hat{v} \circ j_{\mathfrak{q}}$ and $T_2 = j_{\mathfrak{q}}^{-1}(U)$.
(1) $\Rightarrow$ (3): Let $\xi_1, \xi_2 \in U$ be given. We verify condition (3) in Remark 10.i. We write

$$\xi_1 = \frac{\overline{t_1}}{\overline{s}^2} \quad , \quad \xi_2 = \frac{\overline{t_2}}{\overline{s}^2}$$

with $t_1, t_2 \in T$, $s \in R \setminus \mathfrak{q}$. Then

$$\xi_1 + \xi_2 = \frac{\overline{t_1 + t_2}}{\overline{s}^2},$$

and $v(s) \neq \infty$, $\hat{v}(\xi_1 + \xi_2) = v(t_1 + t_2) - 2v(s) = \min(v(t_1), v(t_2)) - 2v(s)$
$= \min(v(t_1) - 2v(s),\ v(t_2) - 2v(s)) = \min(\hat{v}(\xi_1),\ \hat{v}(\xi_2))$. □

As a modest first application of Lemma 14 we analyse T-convexity for valuations in the case that T is a prime cone.

Definition 5. Given valuations v and w on R, we write $v \le w$ if w is a coarser than v (cf.I §1, Def.9).

Notice that $v \sim w$ iff $v \le w$ and $w \le v$.

Theorem 5.15. Let P be a prime cone of R and v a valuation on R.
i) v is P-convex and $\operatorname{supp} v = \operatorname{supp} P$ iff $v_P \le v$.
ii) v is P-convex iff there exists some prime cone $\tilde{P} \supset P$ such that $v_{\tilde{P}} \le v$.

Proof. i): Let $\mathfrak{q} := \operatorname{supp} P$. This is a P-convex prime ideal of R, in fact the smallest one. If $v_P \le v$ then $\operatorname{supp} v = \operatorname{supp} v_P = \mathfrak{q}$. Thus we may assume from start that $\operatorname{supp} v = \mathfrak{q}$. Lemma 14 tells us that v is P-convex iff the valuation $\hat{v}$ on $k(\mathfrak{q})$ is $\hat{P}$-convex. {Here $\hat{P}$ denotes the ordering induced by P on $k(\mathfrak{q})$, as has been decreed in §3.} By Corollary 12 $\hat{v}$ is $\hat{P}$-convex iff the valuation ring $A_{\hat{v}} = \mathfrak{o}_v$ is $\hat{P}$-convex in $k(\mathfrak{q})$. This happens to be true iff $v_{\hat{P}} \le \hat{v}$. Since $v_{\hat{P}} \circ j_{\mathfrak{q}} = v_P$ and $\hat{v} \circ j_{\mathfrak{q}} = v$, we have $v_{\hat{P}} \le \hat{v}$ iff $v_P \le v$.
ii): If there exists some prime cone $\tilde{P} \supset P$ with $v_{\tilde{P}} \le v$ then v is $\tilde{P}$-convex, as we have proved, hence v is P-convex. Conversely, assume that v is P-convex. Then $\mathfrak{p} := \operatorname{supp} v$ is P-convex (hence $\mathfrak{q} \subset \mathfrak{p}$). $\tilde{P} := P \cup \mathfrak{p} = P + \mathfrak{p}$ is a prime cone of R containing P, and $\operatorname{supp} \tilde{P} = \mathfrak{p} = \operatorname{supp} v$ (cf.Th.4.6). We claim that v is $\tilde{P}$-convex, and then will know by i) that $v_{\tilde{P}} \le v$.

This is pretty obvious. If $\tilde{x}, \tilde{y} \in \tilde{P}$, we have $\tilde{x} = x + a$, $\tilde{y} = y + b$ with $x, y \in P$ and $a, b \in \mathfrak{p}$. Then $v(\tilde{x}) = v(x)$, $v(\tilde{y}) = v(y)$, $v(\tilde{x} + \tilde{y}) = v(x+y)$, since also $a + b \in \mathfrak{p}$. We conclude that $v(\tilde{x} + \tilde{y}) = v(x + y) = \min(v(x), v(y)) = \min(v(\tilde{x}), v(\tilde{y}))$, which proves that v is $\tilde{P}$-convex. □

As before, let T be a preordering of R.

Theorem 5.16. Assume that v is a T-convex valuation on R. Then there exists a prime cone $P \supset T$ of R such that v is P-convex and $\operatorname{supp} P = \operatorname{supp} v$ {hence $v_P \le v$ by Th.15}.

Proof. a) We first prove this in the case that $R = K$ is a field. Let $B := A_v$, $\mathfrak{m} := \mathfrak{p}_v$, and $U := T \cap B$. Then B is a T-convex Krull valuation ring of K with maximal ideal $\mathfrak{m}$, and $\mathfrak{m}$ is U-convex in B. By Lemma 9 we know that $U_1 := U + \mathfrak{m}$ is a proper preordering of B and that its image $U_1/\mathfrak{m} = \overline{U}$ in the residue class field $\kappa(B) = B/\mathfrak{m}$ is a proper preordering (= partial ordering) of $\kappa(B)$. We choose a prime cone (= total ordering) $\overline{Q}$ of $\kappa(B)$ containing $\overline{U}$. {Usually this can be done in several ways.} Let $\pi : B \twoheadrightarrow \kappa(B)$ denote the residue class homomorphism from B to $\kappa(B)$. $Q := \pi^{-1}(\overline{Q})$ is a prime cone of B with $T_1 \subset Q$, $\operatorname{supp} Q = \mathfrak{m}$ and $U \subset Q$.

We now invoke the Baer-Krull theorem connecting ordering of K and $\kappa(B)$ in full strength (cf. [La, Cor.3.11], [KS, II §7], [BCR, Th.10.1.10])*). The theorem

*) We stated a rough version of this theorem already above, cf. Lemma 3.2.

can be quoted as follows. Given a group homomorphism $\chi\colon K^* \to \{\pm 1\}$ with $\chi(Q \cap B^*) = \{1\}$ and $\chi(-1) = -1$, there exists a unique prime cone (= ordering) P of K such that B is P-convex and $\mathrm{sign}_P(a) = \chi(a)$ for every $a \in K^*$.

We choose $\chi\colon K^* \to \{\pm 1\}$ in such a way that also $\chi(T \cap K^*) = 1$. By elementary character theory on the group K^*/K^{*2} this is possible, since we have $T \cap B^* \subset Q \cap B^*$ and $-1 \notin (Q \cap B^*) \cdot (T \cap K^*)$. The resulting ordering (= prime cone) P of K contains T, and B is P-convex in K, hence v is P-convex. This completes the proof for $R = K$ a field.

b) We prove the theorem in general. We are given a preordering T and a T-convex valuation v on R. The prime ideal $\mathfrak{q} := \operatorname{supp} v$ is T-convex. Thus Lemma 14 applies. We have a proper preordering U on $k(\mathfrak{q})$ as described there in part ii), and we know by part iii) of the lemma that the valuation $\hat{v}$ on $k(\mathfrak{q})$ is U-convex. As proved above in part a), there exists a prime cone (= ordering) Q on $k(\mathfrak{q})$ containing U such that $\hat{v}$ is Q-convex. It follows that $P := j_{\mathfrak{q}}^{-1}(Q)$ is a prime cone on R with $P \supset T_2 := j_{\mathfrak{q}}^{-1}(U)$, and that $v = \hat{v} \circ j_{\mathfrak{q}}$ is P-convex. As stated in the lemma, $T \subset T_2$, hence $T \subset P$. □

Notice that for v a trivial valuation the theorem boils down to part a) of Theorem 6.

Corollary 5.17. Every T-convex valuation v on R is $\hat{T}$-convex. □

This follows immediately from Theorem 16. It may be of interest – or at least amusing – to see a second proof of Corollary 17, which is based on the Positivstellensatz Theorem 4.

Second proof of Corollary 5.17 (cf.[Z_1, §2]). Suppose that v is T-convex but not $\hat{T}$-convex. We have elements a, b in $\hat{T}$ with

$$v(a+b) > \min(v(a), v(b)). \tag{1}$$

In particular $v(a) \neq \infty$, $v(b) \neq \infty$. By Theorem 4 we have natural numbers m, n and elements u, u', w, w' in T such that

$$au = a^{2m} + u' \quad , \quad bw = b^{2n} + w'.$$

Then $au \in T$, $bw \in T$ and

$$v(au) = \min(v(a^{2m}), v(u')) < \infty \ , \quad v(bw) = \min(v(b^{2n}), v(w')) < \infty.$$

Let $c := a(aubw)$, $d := b(aubw)$. We have $c \in T$, $d \in T$ and

$$v(c+d) \ = \ \min(v(c), v(d)) \ = \ \min(v(a), v(b)) + v(aubw). \tag{2}$$

On the other hand, $c + d = (a+b)aubw$, hence

$$v(c+d) \ = v(a+b) + v(aubw). \tag{3}$$

Since $v(aubw) \neq \infty$, we conclude from (2) and (3) that

$$v(a+b) \ = \ \min(v(a), v(b)), \tag{4}$$

in contradiction to (1). Thus v is $\hat{T}$-convex. □

§6 Convexity of overrings of real holomorphy rings

In this section Λ is a subring of a ring R and T a preordering of R. In §2 we defined the real holomorphy ring $\mathrm{Hol}(R/\Lambda)$ of R over Λ (§2, Def.6). We now generalise this definition.

Definition 1. a) The *T-holomorphy ring* $\mathrm{Hol}_T(R/\Lambda)$ of R over Λ is the intersection of the rings A_v with v running through all T-convex valuations of R over Λ (i.e. with $\Lambda \subset A_v$).
b) If $T = P(X)$ for some set $X \subset \mathrm{Sper} R$ we denote this ring also by $\mathrm{Hol}_X(R/\Lambda)$ and call it the holomorphy ring of the extension $\Lambda \subset R$ over X.
c) In the case $\Lambda = \mathbb{Z}1_R$ we write $\mathrm{Hol}_T(R)$ and $\mathrm{Hol}_X(R)$ instead of $\mathrm{Hol}_T(R/\Lambda)$, $\mathrm{Hol}_X(R/\Lambda)$. We call $\mathrm{Hol}_T(R)$ the *T-holomorphy ring of* R and $\mathrm{Hol}_X(R)$ the *X-holomorphy ring of* R.

Remarks 6.1. i) We know by Corollary 5.17 that

$$\mathrm{Hol}_T(R/\Lambda) = \mathrm{Hol}_{\hat{T}}(R/\Lambda).$$

ii) For the smallest preordering $T_0 = \Sigma R^2$ we have $\mathrm{Hol}_{T_0}(R/\Lambda) = \mathrm{Hol}(R/\Lambda) = \mathrm{Hol}_{\mathrm{Sper} R}(R/\Lambda)$.
iii) If $\mathrm{Hol}_T(R)$ is Prüfer in R then

$$\mathrm{Hol}_T(R/\Lambda) = \Lambda \cdot \mathrm{Hol}_T(R).$$

This can be verified by a straightforward modification of the proof of Proposition 2.20 (which settles the case $T = \Sigma R^2$). $\square$

Given a prime cone P of R we introduced in §3 (cf.Def.5 there) the P-convex valuation $v_{P,\Lambda}$. It has the valuation ring

$$A_{v_{P,\Lambda}} = \mathrm{conv}_P(\Lambda) = C(P, R/\Lambda)$$

and the center $\mathfrak{p}_{v_{P,\Lambda}} = I_P(\Lambda)$. Using these valuations we now obtain a simple description of $\mathrm{Hol}_P(R/\Lambda)$, starting from Theorem 5.15.

Theorem 6.2. Let P be any prime cone of R.
a) A valuation v of R is P-convex and $\Lambda \subset A_v$ iff there exists some prime cone $\tilde{P} \supset P$ with $v_{\tilde{P},\Lambda} \le v$.
b) For every such valuation v we have $A_v \supset \mathrm{Hol}_P(R/\Lambda)$, and

$$\mathrm{Hol}_P(R/\Lambda) = C(P, R/\Lambda) = A(P, R/\Lambda).$$

Also $\mathrm{Hol}_Q(R/\Lambda) = \mathrm{Hol}_P(R/\Lambda)$ for every prime cone $Q \supset P$.

Proof. Claim a) follows immediately Theorem 5.15 which settles the case $\Lambda = \mathbb{Z} \cdot 1_R$.

b): If $v_{Q,\Lambda} \leq v$ then $A_v \supset A_{v_{Q,\Lambda}} = C(Q, R/\Lambda)$. As observed in §5, the Q-convex hull $C(Q, R/\Lambda)$ of Λ with respect to Q does not change if we replace Q by P, and also coincides with the ring $A(Q, R/\Lambda) = A(P, R/\Lambda)$. $\square$

Theorem 6.3. As before, let T be any preordering of R.
a) $\mathrm{Hol}_T(R/\Lambda)$ is the intersection of the rings $\mathrm{Hol}_P(R/\Lambda)$ with P running through the set $\overline{H}_R(T)$ of prime cones $P \supset T$.
b) Given $f \in R$, the following are equivalent.
(1) $f \in \mathrm{Hol}_T(R/\Lambda)$.
(2) $\exists\, \lambda \in \Lambda\colon |f(P)| \leq |\lambda(P)|$ for every $P \in \overline{H}_R(T)$.
(3) $\exists\, \mu \in \Lambda\colon \quad 1 + \mu^2 \pm f \in \hat{T}$.
c) $\mathrm{Hol}_T(R/\Lambda) = C(\hat{T}, R/\Lambda) = A(\hat{T}, R/\Lambda)$.

Proof. a): This follows from the fact that every T-convex valuation v on R is P-convex for some prime cone $P \supset T$, cf. Theorem 5.16.
b): The proof runs in the same way as the proof of Theorem 4.2, which settled the case $T = T_0$.
c): We know by Proposition 5.2 that $C(\hat{T}, R/\Lambda) = A(\hat{T}, R/\Lambda)$. If $f \in \mathrm{Hol}_T(R/\Lambda)$ then condition (3) in b) is fulfilled, hence $f \in A(\hat{T}, R/\Lambda)$. Conversely, if $f \in A(\hat{T}, R/\Lambda)$ we have $-\lambda \leq_{\hat{T}} f \leq_{\hat{T}} \lambda$ for some $\lambda \in \Lambda$. This implies condition (2) in b), hence $f \in \mathrm{Hol}_T(R/\Lambda)$. $\square$

Corollary 6.4. Every $\hat{T}$-convex subring B of R is integrally closed in R.

Proof. We know by Theorem 3 that $B = \mathrm{Hol}_T(R/B)$. Thus B is an intersection of rings A_w with w running through a set of valuations on R. Each A_w is integrally closed in R (cf.Th.I.2.1). Thus B is integrally closed in R. $\square$

Remark. This corollary can be proved in a more direct way, cf.[KS, III §11, Satz 1] or §8 below.

We now turn to a study of T-convexity for subrings of R which are Prüfer in R. This will be a lot easier than studying T-convex subrings in general. We start with a general lemma on localizations.

Lemma 6.5. Let A be a subring of R, M an additive subgroup of A, and S a multiplicative subset of A with $sM \subset M$ for every $s \in S$. We define

$$M_{[S]} \colon = \{x \in R \mid \exists\, s \in S\colon sx \in M\}$$

and, as always,

$$A_{[S]} \colon = \{x \in R \mid \exists\, s \in S\colon sx \in A\}.$$

i) $M_{[S]}$ is an additive subgroup of $A_{[S]}$. If M is an ideal of A then $M_{[S]}$ is an ideal of $A_{[S]}$.
ii) If M is $(T \cap A)$-convex in A then $M_{[S]}$ is $(T \cap A_{[S]})$-convex in $A_{[S]}$.

iii) If $M_{[S]}$ is $(T \cap A_{[S]})$-convex in $A_{[S]}$ and $M_{[S]} \cap A = M$ then M is $(T \cap A)$-convex in A.

Proof. i): evident.
ii): Let $x \in M_{[S]}$ and $y \in A_{[S]}$ be given with $0 \leq_T y \leq_T x$. We choose some $s \in S$ with $sy \in A$ and $sx \in M$. Then $0 \leq_T s^2 y \leq_T s^2 x \in M$ and $s^2 y \in A$. Since M is assumed to be $(T \cap A)$-convex in A, we conclude that $s^2 y \in M$, hence $y \in M_{[S]}$. Thus $M_{[S]}$ is $(T \cap A_{[S]})$-convex in $A_{[S]}$.
iii): Let $x \in M$, $y \in A$ and $0 \leq_T y \leq_T x$. Since $M_{[S]}$ is assumed to be $(T \cap A_{[S]})$-convex in $A_{[S]}$, we conclude that $y \in M_{[S]} \cap A = M$. Thus M is $(T \cap A)$-convex in A. $\square$

We will use two special cases of this lemma, stated as follows.

Lemma 6.6. Let A be a subring of R and $\mathfrak{p}$ a prime ideal of A.
i) If A is T-convex in R then $A_{[\mathfrak{p}]}$ is T-convex in R.
ii) $\mathfrak{p}_{[\mathfrak{p}]}$ is $(T \cap A_{[\mathfrak{p}]})$-convex in $A_{[\mathfrak{p}]}$ iff $\mathfrak{p}$ is $(T \cap A)$-convex in A.

Proof. i): Apply Lemma 5 choosing $A, R, A \setminus \mathfrak{p}$ for M, A, S.
ii): Apply the lemma choosing $\mathfrak{p}, A, A \setminus \mathfrak{p}$ for M, A, S. $\square$

Theorem 6.7. Assume that A is a Prüfer subring of R. The following are equivalent.
(1) A is T-convex in R.
(2) For every R-regular maximal (or: prime) ideal $\mathfrak{p}$ of A the ring $A_{[\mathfrak{p}]}$ is T-convex in R.
(3) For every R-regular maximal (or: prime) ideal $\mathfrak{p}$ of A the ideal $\mathfrak{p}_{[\mathfrak{p}]}$ of $A_{[\mathfrak{p}]}$ is $(T \cap A_{[\mathfrak{p}]})$-convex in $A_{[\mathfrak{p}]}$.
(4) Every non trivial PM-valuation v of R over A is T-convex.
(5) Each R-regular maximal (or: prime) ideal of A is $(T \cap A)$-convex in A.
(6) Each R-regular maximal (or: prime) ideal of A is T-convex in R.
(7) A is $\hat{T}$-convex in R.

Proof. We may assume that $A \neq R$.
(1) $\Rightarrow$ (2): Evident by Lemma 6.6.i.
(2) $\Rightarrow$ (1): Clear, since A is the intersection of the rings $A_{[\mathfrak{p}]}$ with $\mathfrak{p}$ running through $\Omega(R/A)$.
(2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4): This holds by Theorem 5.11.
(3) $\Leftrightarrow$ (5): Evident by Lemma 6.6.ii.
We now have verified the equivalence of (1), (2), (3), (4), (5).
(1) $\Rightarrow$ (6): If $\mathfrak{p}$ is an R-regular prime ideal of A then $\mathfrak{p}$ is $(T \cap A)$-convex in A by (5) and A is T-convex in R. Thus $\mathfrak{p}$ is T-convex in R.
(6) $\Rightarrow$ (5): trivial.
(7) $\Rightarrow$ (1): trivial.

(4) $\Rightarrow$ (7): We know by Corollary 5.17 that $v_{\mathfrak{p}}$ is $\hat{T}$-convex for every $\mathfrak{p} \in \Omega(R/A)$. Using the implication (4) $\Rightarrow$ (1) for $\hat{T}$ instead of T we see that A is $\hat{T}$-convex in R. $\square$

Corollary 6.8. Let A be a Prüfer subring of R, and let C denote the T-convex hull of A in R, $C = C(T, R/\Lambda)$. Assume that C is a subring of R. {N.B. This is known to be true under very mild additional assumptions, cf. Prop.5.2.}
a) Then $S(R/C)$ *) is the set of all T-convex valuations $v \in S(R/A)$.
b) $C = \mathrm{Hol}_T(R/A)$, and $C = \bigcap\limits_{\mathfrak{p}} A^R_{[\mathfrak{p}]}$ with $\mathfrak{p}$ running through the set of R-regular prime ideals $\mathfrak{p}$ of A which are T-convex (i.e. $(T \cap A)$-convex)) in A. $\square$

Proof. Claim a) follows immediately from the equivalence (1) $\Leftrightarrow$ (4) in Theorem 7. We then have $C = \mathrm{Hol}_T(R/A)$ by the very definition of the relative real holomorphy ring $\mathrm{Hol}_T(R/A)$. The last statement in the corollary is evident due to the 1-1-correspondence of PM-valuations v of R over A with the R-regular prime ideals $\mathfrak{p}$ of A. $\square$

We arrive at a theorem which demonstrates well the friendly relation between T-convexity and the Prüfer condition.

Theorem 6.9. Let A be a T-convex subring of R. Then A is Prüfer in R iff every R-overring of A is $\hat{T}$-convex in R.

Proof. a) Assume that A is Prüfer in R. Let B be an R-overring of R. The ring B inherits property (4) in Theorem 7 from A, hence is $\hat{T}$-convex in R by that theorem.
b) If every R-overring of A is $\hat{T}$-convex in R then each such ring is integrally closed in R, as stated above (Corollary 4). Thus A is Prüfer in R (cf. Theorem I.5.2).

Corollary 6.10. Let Λ be a subring of R. Assume that $\mathrm{Hol}_T(R/\Lambda)$ is Prüfer in R. Then the $\hat{T}$-convex subrings of R containing Λ are precisely the overrings of $\mathrm{Hol}_T(R/\Lambda)$ in R.

Proof. We know by Theorem 3 that $\mathrm{Hol}_T(R/\Lambda)$ is the $\hat{T}$-convex hull $C(\hat{T}, R/\Lambda)$ of Λ in R. Now apply Theorem 9. $\square$

Remark. If R has positive definite inversion, or, if for every $x \in R$ there exists some $d \in \mathbb{N}$ with $1 + x^{2d} \in R^*$, we know by §2 that $\mathrm{Hol}(R)$ is Prüfer in R, hence $\mathrm{Hol}_T(R)$ is Prüfer in R, and Corollary 10 applies. Thus we have a good

*) Recall that $S(R/C)$ denotes the restricted PM-spectrum of R over C (§1).

hold on $\hat{T}$-convexity under conditions which, regarded from the view-point of real algebra, are mild.

Our proof of Theorem 7 (and hence Theorem 9) is based a great deal on Lemma 6 above. The lemma also leads us to a supplement to the theory of convex valuations developed in §5.

Proposition 6.11. Let B be a Prüfer subring of R which is T-convex in R, and let v be a $(T \cap B)$-convex PM-valuation on B. Then the induced valuation v^R on R (cf. §1, Def.5) is T-convex.

Proof: Let $A := A_v$, $\mathfrak{p} := \mathfrak{p}_v$, $w := v^R$. Since v is the special restriction $w|_B$ of w to B, we have $A_w \cap B = A$, $\mathfrak{p}_w \cap B = \mathfrak{p}$. Now A is Prüfer in R, and $A \subset A_w \subset R$. Thus $A_w = A^R_{[\mathfrak{p}]}$, $\mathfrak{p}_w = \mathfrak{p}^R_{[\mathfrak{p}]}$. The ring A is T-convex in B, hence in R. Further $\mathfrak{p}$ is T-convex in A, hence in R. By Lemma 6 it follows that A_w is T-convex in R and $\mathfrak{p}_w$ is T-convex in A_w. We conclude by Theorem 5.11 that the Manis valuation w is T-convex. $\square$

§7 The case of bounded inversion; convexity covers

Definition 1. Let (R,T) be a preordered ring, i.e. a ring R equipped with a preordering T. We say that (R,T) has *bounded inversion*, if $1+t$ is a unit of R for every $t \in T$, in short, $1+T \subset R^*$. If A is a subring of R, we say that A has *bounded inversion with respect to* T, if $(A, T\cap A)$ has bounded inversion, i.e. $1+(T\cap A) \subset A^*$.

The theory of T-convex Prüfer subrings of R turns out to be particularly nice and good natured if (R,T) has bounded inversion, as we will explicate now.

We first observe that (R,T) has bounded inversion iff $(R,\hat{T})$ has bounded inversion, due to the following proposition.

Proposition 7.1. Given a preordering T on a ring R, the following are equivalent.
(1) $1+T\subset R^*$
(2) Every maximal ideal $\mathfrak{m}$ of R is T-convex in R.
(3) $1+\hat{T}\subset R^*$.

Proof. (1) $\Rightarrow$ (2): This follows from Proposition 5.7.*)
(2) $\Rightarrow$ (3): If $\mathfrak{m}$ is a maximal ideal of R then $\mathfrak{m}$ is T-convex in R, hence $\hat{T}$-convex in R (cf.Th.5.6). It follows that $\mathfrak{m}\cap(1+\hat{T})=\emptyset$. Since this holds for every maximal ideal of $\mathfrak{m}$, the set $1+\hat{T}$ consists of units of R.
(3) $\Rightarrow$ (1): trivial. $\square$

Thus, in the bounded inversion situation, we most often can switch from T to $\hat{T}$ and back.

Theorem 7.2. Let A be a subring of R.
i) The following are equivalent.
(1) A is Prüfer in R and $1+(T\cap A)\subset A^*$.
(2) A is Prüfer in R and $1+(\hat{T}\cap A)\subset A^*$.
(3) A is T-convex in R and $1+T\subset R^*$.
(4) A is $\hat{T}$-convex in R and $1+\hat{T}\subset R^*$.
ii) If (1) – (4) hold, every R-overring B of A is $\hat{T}$-convex in R and $B=S^{-1}A$ with $S:=T\cap A\cap B^*$.

Proof. a) We assume (1), i.e. $A\subset R$ is Prüfer and $1+(T\cap A)\subset A^*$. By Proposition 1 every maximal ideal $\mathfrak{m}$ of A is $(T\cap A)$-convex in A. Thus condition (5) in Theorem 6.7 holds and A is $\hat{T}$-convex and (hence) T-convex in R by that theorem. Applying Theorem 6.7 to $\hat{T}$ instead of T we learn that (2) holds. Since the implication (2) $\Rightarrow$ (1) is trivial we now know that (1) $\Leftrightarrow$ (2).

*) A direct proof can be found in [Z_1, p.5804 f].

b) Assuming (1) we prove that $1+T \subset R^*$. To this end let $\mathfrak{Q}$ be a maximal ideal of R. We verify that $\mathfrak{Q}$ is T-convex in R and then will be done by Proposition 1.

Let $\mathfrak{q} := \mathfrak{Q} \cap A$. Since A is ws in R, we have $A_{[\mathfrak{q}]} = R$ and $\mathfrak{q}_{[\mathfrak{q}]} = \mathfrak{Q}$ (cf.Th.I.4.8). By Lemma 6.6 it suffices to verify that $\mathfrak{q}$ is $(T \cap A)$-convex in A. We choose a maximal ideal $\mathfrak{m}$ of A containing $\mathfrak{q}$.

Case 1. $\mathfrak{m}R \neq R$. We have $\mathfrak{Q} = R\mathfrak{q} \subset R\mathfrak{m}$ (cf.Th.I.4.8). Since $\mathfrak{Q}$ is a maximal ideal of R it follows that $R\mathfrak{q} = R\mathfrak{m}$ and then, again by Th.I.4.8., that $\mathfrak{q} = \mathfrak{m}$. The ideal $\mathfrak{m}$ is $(T \cap A)$-convex in A, due to (1) and Lemma 6.6.

Case 2. $\mathfrak{m}R = R$. We have a Manis valuation v on R with $A_v = A_{[\mathfrak{m}]}$ and $\mathfrak{p}_v = \mathfrak{p}_{[\mathfrak{m}]}$. It follows by Proposition I.1.3 that $(\operatorname{supp} v)_\mathfrak{m}$ is a maximal ideal of $R_\mathfrak{m}$. Now $\mathfrak{Q}_\mathfrak{m}$ is an ideal of $R_\mathfrak{m}$ contained in the center $\mathfrak{p}_\mathfrak{m}$ of the Manis valuation $\tilde{v}$ induced by v on $R_\mathfrak{m}$. Thus $\mathfrak{Q}_\mathfrak{m} \subset \operatorname{supp}(\tilde{v}) = (\operatorname{supp} v)_\mathfrak{m}$. This implies $\mathfrak{Q} \subset R \cap (\operatorname{supp} v)_\mathfrak{m} = \operatorname{supp} v$, and then $\mathfrak{Q} = \operatorname{supp} v$, since $\mathfrak{Q}$ is a maximal ideal of R. Thus $\operatorname{supp} v = \mathfrak{q}_{[\mathfrak{q}]}$.

Since $1+(T\cap A) \subset A^*$, the ideal $\mathfrak{m}$ is $(T\cap A)$-convex in A, due to Proposition 1. Now Lemma 6.6 tells us that $\mathfrak{m}_{[\mathfrak{m}]} = \mathfrak{p}_v$ is $(T \cap A_{[\mathfrak{m}]})$-convex in $A_{[\mathfrak{m}]} = A_v$. We conclude by Theorem 5.11 that the valuation v is T-convex. It follows that $\operatorname{supp} v = \mathfrak{Q}_{[\mathfrak{q}]}$ is T-convex in R.
We have proved the implication (1) $\Rightarrow$ (3) in part i) of the theorem. Changing from T to $\hat{T}$ we also know that (2) $\Rightarrow$ (4). The implication (4) $\Rightarrow$ (3) is trivial. Altogether we have proved the implications (1) $\Leftrightarrow$ (2) $\Rightarrow$ (4) $\Rightarrow$ (3).

c) We finally prove that condition (3) implies (1) and all the assertions listed in part ii) of the theorem, and then will be done. Thus assume that that A is T-convex in R and $1 + T \subset R^*$. For every $t \in T$ we have $0 \leq_T \frac{1}{1+t} \leq_T 1$. It follows that $\frac{1}{1+t} \in A$. In particular $1 + x^2 \in R^*$ and $\frac{1}{1+x^2} \in A$ for every $x \in R$. Thus A is Prüfer in R, as is clear already by I §6, Example 13. (Take $d = 2$ there.) For $t \in A \cap T$ we have $1 + t \in A$ and $(1 + t)^{-1} \in A$, hence $1 + t \in A^*$.

Let B be an R-overring of A. If $t \in T \cap B$ then $\frac{1}{1+t} \in A \subset B$, hence $1+t \in B^*$. By the proved implication (1) $\Rightarrow$ (3) from above it follows that B is T-convex in R.

Let $b \in B$ be given. Then $s := \frac{1}{1+b^2} \in A$. Also $0 \leq_T \frac{2b}{1+b^2} \leq_T 1$, hence $a := 2bs \in A$. We have $s \in S := T \cap A \cap B^*$ and, of course, $2 \in S$. Thus $b = \frac{a}{2s} \in S^{-1}A$. We have proved all claims of the theorem. $\square$

Corollary 7.3. Let A be a Prüfer subring of R and B an overring of A in R. Then the T-convex hull $C(T, R/B)$ coincides with the saturation

$$B_{[S]} := \{x \in R \mid \exists\, s \in S \colon sx \in B\},$$

where $S := 1 + (T \cap B)$.

Proof. a) We equip the localisation $S^{-1}R$ with the preordering $S^{-1}T = \{\frac{t}{s} \mid t \in T, s \in S\}$. One easily checks that $(S^{-1}T) \cap (S^{-1}A) = S^{-1}(T \cap A)$. Applying Theorem 2 to the Prüfer extension $S^{-1}A \subset S^{-1}R$ we learn that $S^{-1}B$ is $S^{-1}T$-convex in $S^{-1}R$. Taking preimages in R we see that $B_{[S]}$ is $T_{[S]}$-convex in R, where $T_{[S]}$ denotes the preimage of $S^{-1}T$ in R. Now $T \subset T_{[S]}$. Thus $B_{[S]}$ is T-convex in R. This proves that $C(T, R/B) \subset B_{[S]}$.
b) Let $x \in B_{[S]}$ be given. There exists some $s \in S$ with $sx \in B$, $s = 1 + t$ with $t \in T \cap A$. We conclude from $0 \leq_T x^2 \leq_T s^2x^2 \in B$ that $x^2 \in C(T, R/B)$. Now B is integrally closed in R, since A is Prüfer in R. Thus $x \in C(T, R/B)$. This proves that $B_{[S]} \subset C(T, R/B)$. □

In the following we fix a preordered ring (R, T). As common in the case of ordered structures we suppress the ordering in the notation (since it is fixed), simply writing R for the pair (R, T). The subset T of R will usually be denoted by R^+. Any subring B of R is again regarded as a preordered ring, with $B^+ = T \cap B$. If we say that B has bounded inversion, we of course mean bounded inversion with respect to B^+.

Definition 2. For any subring B of R let C_B denote the smallest subring of B which is convex (= T-convex) in B. Thus, in former notation, $C_B = C(T \cap B, B) = C(T \cap B, B/\mathbb{Z})$. {Recall Prop.5.2.d.}

Proposition 7.4. Let B be a subring of R.
i) $C_B = \{x \in B \mid \exists\, n \in \mathbb{N}\colon -n \leq_T x \leq_T n\}$.
ii) C_B is contained in the real holomorphy ring $\mathrm{Hol}_{B^+}(B)$.
iii) If C_B is Prüfer in B, then $C_B = \mathrm{Hol}_{B^+}(B)$.
iv) If B has bounded inversion, then C_B is Prüfer in B and $C_B = \sum\limits_{t \in B^+} \mathbb{Z}\, \frac{1}{1+t}$.

Proof. i): Clear by Proposition 5.2.d.
ii): $\mathrm{Hol}_{B^+}(B)$ is a subring of B which is $(B^+)^\wedge$-convex in B (cf.Th.6.3.c), hence B^+-convex in B. This forces $C_B \subset \mathrm{Hol}_{B^+}(B)$.
iii): C_B is the intersection of the rings A_v with v running through the non-trivial PM-valuations of B over C_B. These are B^+-convex (cf.Th.6.7). Thus $\mathrm{Hol}_{B^+}(B) \subset C_B$. Since the reverse inclusion holds anyway, as just proved, $\mathrm{Hol}_{B^+}(B) = C_B$.
iv): The proof of Theorem 2.11 extends readily to the present situation. It gives us $\mathrm{Hol}_{B^+}(B) = \sum\limits_{t \in B^+} \mathbb{Z}\, \frac{1}{1+t}$, verifying in between that the right hand side is a Prüfer subring of B. We have $0 \leq_T \frac{1}{1+t} \leq_T 1$ for every $t \in B^+$. Thus $\mathrm{Hol}_{B^+}(B) \subset C_B$. Since $C_B \subset \mathrm{Hol}_{B^+}(B)$ anyway, both rings coincide. □

Up to now we have been rather pedantic using the term "B^+-convex" instead of just "convex". The reason was that also the saturated preordering $(B^+)^\wedge$ came into play. In the following the term "convex" will always refer to the given preordering $T = R^+$ of R.

Remark 7.5. If A and B are subrings of R with $A \subset B$, then $C_A \subset C_B$. Indeed, $A \cap C_B$ is convex in A, hence $C_A \subset A \cap C_B$.

Theorem 7.6. Let A and B be subrings of R with $A \subset B$. The following are equivalent.
(1) A has bounded inversion, and A is Prüfer in B.
(2) B has bounded inversion, and A is convex in B.
(3) Both A and B have bounded inversion, and $C_A = C_B$.

Proof. The equivalence (1) $\Longleftrightarrow$ (2) is a restatement of (1) $\Longleftrightarrow$ (3) in Theorem 2.
(1) $\wedge$ (2) $\Rightarrow$ (3): By assumption (1) and (2) both A and B have bounded inversion, and A is convex in B. Since C_A is convex in A we conclude that C_A is convex in B, and then, that $C_B \subset C_A$. Thus $C_A = C_B$.
(3) $\Rightarrow$ (1): Applying the implication (2) $\Rightarrow$ (1) to C_B and B, we see that $C_A = C_B$ is Prüfer in B. {This had already been stated in Prop.4.} Since $C_A \subset A \subset B$, it follows that A is Prüfer in B. □

Corollary 7.7. Let A be a subring of R, and let D denote the Prüfer hull of A in R, $D = P(A, R)$ (cf.I, §5, Def.2). Assume that A has bounded inversion.
a) Every overring B of A in D has bounded inversion and is convex in D, and $C_B = C_A$.
b) D is the unique maximal overring B of A in R such that B has bounded inversion and $C_B = C_A$.
c) D is the unique maximal overring B of A such that A is convex in B and B has bounded inversion.
d) C_A has bounded inversion, and D is the Prüfer hull of C_A in R. The overrings of C_A in D are precisely all subrings B of R such that $C_B = C_A$ and B has bounded inversion.

Proof. a): If B is an overring of A in D, then A is Prüfer in B. Thus, by Theorem 6, B has bounded inversion and $C_A = C_B$. In particular, D has bounded inversion and $C_A = C_D$. Applying Theorem 6 to B and D we see that B is convex in D.
b): If B is an overring of A in R with bounded inversion and $C_A = C_B$, then A is Prüfer in B by Theorem 6, hence $B \subset D$.
c): If B is an overring of A in R with bounded inversion such that A is convex in B, then again A is Prüfer in B by Theorem 6, hence $B \subset D$.
d): C_A is convex in A, hence is Prüfer in A (cf.Th.6 or Prop.4). Thus D is also the Prüfer hull of C_A in R. Now apply what has been proved about the extension $A \subset R$ to the extension $C_A \subset R$, taking into account the trivial fact that $C_A = C_B$ implies $C_A \subset B$. □

The corollary tells us in particular (part c) that A has a unique maximal overring D such that A is convex in D and D has bounded inversion. Does there hold something similar without the inverse boundedness condition? The answer is "Yes" provided A is Prüfer in R, as we are going to explain. We now denote the basic subring of R to start with Λ instead of A, since the letter A will turn up with another meaning.

Let Λ be a subring of R. We denote the subring $A(R^+, R/\Lambda)$ and the additive subgroup $C(R^+, R/\Lambda)$ (cf.§5) briefly by $A(R/\Lambda)$ and $C(R/\Lambda)$ respectively. Recall from Proposition 5.2 that $C(R/\Lambda) = \Lambda + A(R/\Lambda)$. We need the following easy

Lemma 7.8. Let B be an overring of Λ in R. Then $A(B/\Lambda) = B \cap A(R/\Lambda)$ and $C(B/\Lambda) = B \cap C(R/\Lambda)$.

Proof. The first equality is evident from the definition of $A(B/\Lambda)$ and $A(R/\Lambda)$ in §5. The second one now follows since $B \cap [\Lambda + A(R/\Lambda] = \Lambda + [B \cap A(R/\Lambda)]$. $\square$

Definition 3. Assume that Λ is Prüfer in R. The *convexity cover of* Λ *in* R is the polar $C(R/\Lambda)^\circ$ of $C(R/\Lambda)$ over Λ in R, i.e. the unique maximal R-overring E of Λ with $C(R/\Lambda) \cap E = \Lambda$ (cf.II, §7). We denote the convexity cover by $CC(R/\Lambda)$.*)

Recall that the polar I° is defined for any Λ-overmodule I of Λ in R. Thus we do not need to assume here that $C(R/\Lambda)$ itself is a subring of R.

The name "convexity cover" is justified by the following theorem.

Theorem 7.9. Assume that Λ is Prüfer in R. Let B be any R-overring of Λ. Then Λ is convex in B iff $B \subset CC(R/\Lambda)$. Thus $CC(R/\Lambda)$ is the unique maximal overring E of Λ in R such that Λ is convex in E.

Proof. Let B be any R-overring of Λ. By the lemma we have $C(B/\Lambda) = B \cap C(R/\Lambda)$. Thus Λ is convex in B iff $B \cap C(R/\Lambda) = \Lambda$. This means that $B \subset C(R/\Lambda)^\circ$. $\square$

If Λ is any subring of R then Theorem 9 still gives us the following.

Corollary 7.10. There exists a unique maximal R-overring E of Λ such that Λ is Prüfer and convex in E, namely $E = CC(P(\Lambda, R)/\Lambda)$. $\square$

Definition 4. We call this R-overring E of Λ the *Prüfer convexity cover of* Λ *in* R, and denote it by $P_c(\Lambda, R)$. $\square$

Scholium 7.11. If B_1 and B_2 are overrings of Λ in R such that Λ is Prüfer and convex in B_1 and in B_2 then Λ is also Prüfer and convex in B_1B_2. Indeed, B_1 and B_2 are both subrings of $P_c(A, R)$. Thus $B_1B_2 \subset P_c(A, R)$. $\square$

*) More precisely we write $CC(T, R/\Lambda)$, with $T = R^+$, if necessary.

We do not have such a result for “convex” alone, omitting the Prüfer condition.

In §10 we will meet a situation where a preordered (in fact partially ordered) ring A is given, such that the preordering extends to the Prüfer hull $P(A)$ in a natural way. Then we will have an *“absolute” Prüfer convexity cover* $P_c(A) := P_c(A, P(A))$ at our disposal, which is the unique maximal Prüfer extension E of A such that A is convex in E.

§8 Convexity of submodules

As before (R,T) is a preordered ring. But now we fix a subring A of R and study T-convexity for A-submodules of R instead of subrings. We will use this to develop more criteria that A is Prüfer and T-convex in R, and to find more properties of such extensions $A \subset R$. Large parts of this section may be read as a supplement to our multiplicative ideal theory in Chapter II in the presence of a preordering.

As we already did in part of the preceding section we usually simplify notation by saying "convex" instead of "T-convex", and writing $C(R/A)$ instead of $C(T,R/A)$ etc. This will cause no harm as long as we keep the preordering T fixed.

We start with an important observation by Brumfiel in his book [Br]. Brumfiel there only considers the case that T is a partial ordering of R, i.e. $T \cap (-T) = \{0\}$, but his arguments go through more generally for a preordering T.

Proposition 8.1. Let $u_1, \dots, u_{2n}, t$ be indeterminates over $\mathbb{Q}$, $u := (u_1, \dots, u_{2n})$, and $f(t) := t^{2n} + u_1 t^{2n-1} + \cdots + u_n$. Then there exists some $k \in \mathbb{N}$, polynomials $b^+(u), b^-(u) \in \mathbb{Q}[u]$, and polynomials $h_i^+(u,t), h_i^-(u,t) \in \mathbb{Q}[u,t]$, $1 \leq i \leq k$, such that

$$t - b^+(u) + \sum_{i=1}^{k} h_i^+(u,t)^2 = f(t),$$

$$b^-(u) - t + \sum_{i=1}^{k} h_i^-(u,t)^2 = f(t).$$

The proof runs by induction on n, cf. [Br, p.123 ff]. □

Inserting for the u_i elements a_i of our subring A of R we obtain the following corollary.

Corollary 8.2. Assume that $\mathbb{Q} \subset R$. If $\alpha \in R$ and $f(t) = t^{2n} + a_1 t^{2n-1} + \cdots + a_{2n}$ is a monic polynomial of even degree over A with $f(\alpha) \leq_T 0$, then

$$b^-(a_1, \dots, a_{2n}) \ \leq_T \ \alpha \ \leq_T \ b^+(a_1, \dots, a_{2n}).$$

Thus α is an element of the convex closure $C(R/A)$ of A in R.*) □

In particular we have

Corollary 8.3. If $\mathbb{Q} \subset R$, and A is convex in R, then A is integrally closed in R. □

*) Notice that $\mathbb{Q} \subset C(R/A)$.

It is possible to weaken the condition $\mathbb{Q} \subset R$ in Corollary 3 considerably.

Proposition 8.4. Assume that A is convex in R and 2-saturated in R (i.e., for every $x \in R$, $2x \in A \Rightarrow x \in A$). Then A is integrally closed in R.

Proof Let $\tilde{R} := \mathbb{Q} \otimes_{\mathbb{Z}} R$ and $\tilde{A} := \mathbb{Q} \otimes_{\mathbb{Z}} A$. As usual, we regard R as a subring of $\tilde{R}$. Then $A \subset \tilde{A}$. The preordering T extends to a preordering $\tilde{T}$ of $\tilde{R}$, and $\tilde{A}$ is $\tilde{T}$-convex in $\tilde{R}$, as is easily seen, since A is assumed to be T-convex in R.

Let $x \in R$ be integral over A. Then x is integral over $\tilde{A}$, and we know by Corollary 3 that $x \in \tilde{A}$. Thus $nx \in A$ for some $n \in \mathbb{N}$. We have $0 \underset{T}{\leq} x^2 \underset{T}{\leq} n^2x^2 \in A$. Since A is T-convex in R, it follows that $x^2 \in A$. Also $1+x$ is integral over A, and thus $(1+x)^2 \in A$. We conclude that $2x = (1+x)^2 - x^2 \in A$, and then, that $x \in A$, since A is 2-saturated in R. $\square$

Here is another observation about convexity in R. If M is any subset of R, we define

$$[A\colon M] := [A\colon_R M] := \{y \in R \mid yx \in A \quad \text{for every } x \in M\}$$

(thus $[A\colon M] = [A\colon AM]$).

Proposition 8.5. Assume again that A is convex and 2-saturated in R.

a) For every subset M of R the A-module $[A\colon M]$ is convex and 2-saturated in R.

b) Every R-invertible A-submodule of R is convex and 2-saturated in R.

Proof. a): Since $[A\colon M]$ is the intersection of the A-modules $[A\colon x]$ with x running through M, it suffices to prove the claim for $M = \{x\}$ with x a given element of R.

If $y \in R$ and $2y \in [A\colon x]$, then $2xy \in A$, hence $xy \in A$, i.e. $y \in [A\colon x]$. Thus $[A\colon x]$ is 2-saturated in R.

Let $s, t \in T$ be given with $s + t \in [A\colon x]$. Then $0 \underset{T}{\leq} s^2x^2 \underset{T}{\leq} (s+t)^2x^2 \in A$. Thus $(sx)^2 \in A$. By Proposition 4 we infer that $sx \in A$, i.e. $s \in [A\colon x]$. This proves that $[A\colon x]$ is convex in R.

b): If I is an R-invertible A-submodule of R then $I = [A\colon I^{-1}]$, and part a) applies. $\square$

Remark 8.6. Assume that A is convex in R and $2 \in R^*$. Then $2 \in A^*$, hence A is 2-saturated in R.

Proof. $0 \underset{T}{\leq} \frac{1}{2} \underset{T}{\leq} 1 \in A$, hence $\frac{1}{2} \in A$. $\square$

Thus the assumption in Propositions 4 and 5, that A is 2-saturated in R, is a very mild one.

Theorem 8.7. The following are equivalent.
(i) A is Prüfer, convex and 2-saturated in R.
(ii) Every R-regular A-submodule of R is convex and 2-saturated in R.
(iii) For every $x \in R$ the A-module $A + Ax^2$ is convex and 2-saturated in R.
(iv) Every R-overring of A is convex and 2-saturated in R.

Proof. (i) $\Rightarrow$ (ii): It suffices to study finitely generated R-regular A-modules. These are invertible in R, hence, according to Proposition 5, are convex and 2-saturated in R.
(ii) $\Rightarrow$ (iii) and (ii) $\Rightarrow$ (iv): trivial.
(iii) $\Rightarrow$ (i): By assumption $A = A + 0 \cdot A$ is convex and 2-saturated in R, and A is integrally closed in R due to Proposition 4. Let $x \in R$ be given. We have $-1 - x^2 \leq_T 2x \leq_T 1 + x^2$ and conclude by (iii) that $2x \in A + Ax^2$, then, that $x \in A + Ax^2$. Now Theorem I.5.2 tells us that A is Prüfer in R.
(iv) $\Rightarrow$ (i): Let B be an overring of A in R. By assumption B is convex and 2-saturated in R. Thus, by Proposition 4, B is integrally closed in R. We conclude by Theorem I.5.2 that A is Prüfer in R. $\square$

Remarks 8.8. i) If $2 \in R^*$ we may drop the 2-saturation assumption in all conditions (i) – (iv), since now convexity of A implies $2 \in A^*$ (cf. Remark 8.6 above). Then every A-submodule of R is 2-saturated.
ii) If $2 \in R^*$ and A is convex in R, the theorem tells us in particular that A is Prüfer in R iff every R-overring of A is convex in R. This improves Theorem 6.9 in the case $2 \in R^*$. $\square$

We now strive for criteria which start with a mild general assumption on T and the extension $A \subset R$, and then decide whether A is T-convex and Prüfer in R by looking for $(T \cap A)$-convexity in A of suitable R-regular ideals of A. One such criterion had already been given within Theorem 6.7, cf. there (1) $\Leftrightarrow$ (5).

Theorem 8.9. Assume that S is a multiplicative subset of A. Assume further that $2 \in S$, and every element of S is a nonzero divisor in A. Let $R := S^{-1}A$. The following are equivalent.
(i) A is Prüfer and convex in R.
(ii) For every $a \in A$ and $s \in S$ the ideal $As^2 + Aa^2$ is convex (i.e. $A \cap T$-convex) in A.

Proof. (i) $\Rightarrow$ (ii): Let $a \in A$ and $s \in S$ be given. Take $x := \frac{a}{s^2}$. The module $A + Ax^2$ is convex in R by Theorem 7. The map $z \mapsto s^2 z$ from R to R is an automorphism of the preordered abelian group $(R, +, T)$. Thus $As^2 + Aa^2 = s^2(A + Ax^2)$ is convex in R, hence in A.
(ii) $\Rightarrow$ (i): a) We first verify that 2 is a unit in A. Let $x := \frac{1}{2}$. Then $x \in R = S^{-1}A$ and $a := 4x \in A$. We have $0 \leq a \leq 4$, and $A \cdot 4 = A \cdot 2^2 + A \cdot 0$ is convex in A. Thus $a \in 4A$, hence $x \in A$.

b) We start out to prove that A is convex in R. {This is the main task!} Let $x \in R$ and $b \in A$ be given with $0 \underset{T}{\leq} x \underset{T}{\leq} b$. Write $x = \frac{a}{s}$ with $a \in A$, $s \in S$. We have

$$0 \underset{T}{\leq} a^2 \underset{T}{\leq} b^2 s^2 \underset{T}{\leq} s^4 + b^2 s^2.$$

Since $As^4 + Ab^2s^2$ is convex in A, this implies $a^2 \in As^4 + Ab^2s^2$, hence $x^2 \in As^2 + Ab^2 \subset A$.
Since $0 \underset{T}{\leq} x+1 \underset{T}{\leq} b+1 \in A$, also $(1+x)^2 \in A$, and thus $x = \frac{1}{2}[(1+x)^2 - x^2] \in A$. A is convex in R.

c) We finally prove for any $x \in R$ that $A + Ax^2$ is convex in R. Then we will know by Theorem 7 and Remark 8.i that A is Prüfer in R, and will be done.

Write $x = \frac{a}{s}$ with $a \in A$, $s \in S$. By assumption the A-module $Aa^2 + As^2$ is convex in A, hence convex in R. Thus also $A + Ax^2 = s^{-2}(Aa^2 + As^2)$ is convex in R. □

Lemma 8.10. Let I, J, K be A-submodules of R with $I \subset J$.
a) If I is 2-saturated in J, then $[I\colon K]$ is 2-saturated in $[J\colon K]$.
b) If the A-module K is generated by $K \cap T$ and I is convex in J, then $[I\colon K]$ is convex in $[J\colon K]$.

Proof. a): Let $x \in [J\colon K]$ and $2x \in [I\colon K]$. For any $s \in K$ we have $2sx \in I$, $sx \in J$, hence $sx \in I$. Thus $x \in [I\colon K]$.
b): Let $M := K \cap T$. Let $x \in [J\colon K]$ and $y \in [I\colon K]$ be given with $0 \underset{T}{\leq} x \underset{T}{\leq} y$. For any $s \in M$ we have $0 \underset{T}{\leq} sx \underset{T}{\leq} sy$ and $sx \in J$, $sy \in I$. It follows that $sx \in I$. Since the A-module K is generated by M, we conclude that $x \in [I\colon K]$. □

Definition 1. We say that an A-submodule I of R is *T-invertible in R*, or *(R,T)-invertible*, if I is R-invertible and both I and I^{-1} are generated by $I \cap T$ and $I^{-1} \cap T$ respectively.

Notice that the product IJ of any two (R,T)-invertible A-submodules I, J of R is again (R,T)-invertible.

Examples 8.11. i) Assume that A is Prüfer in R. Then, for every R-invertible A-module I, the module I^2 is T-invertible in R. Indeed, write $I = Aa_1 + \cdots + Aa_n$. Then $I^2 = Aa_1^2 + \cdots + Aa_n^2$ (cf. Prop.II.1.8), and $a_1^2, \dots, a_n^2 \in T$. Also I^{-2} is generated by $T \cap I^{-2}$.
ii) If $A \subset R$ is any ring extension and P is a prime cone of R then clearly every R-invertible A-submodule of R is P-invertible in R.

Lemma 8.12. Let I, J, K be A-submodules of R with $I \subset J$. Assume that K is T-invertible in R. Then I is convex in J iff IK is convex in JK, and I is 2-saturated in J iff IK is 2-saturated in JK.

Proof. This follows from Lemma 10, since, for any A-module $\mathfrak{a}$ in R, we have $\mathfrak{a}K = [\mathfrak{a}\colon K^{-1}]$ and $\mathfrak{a}K^{-1} = [\mathfrak{a}\colon K]$. □

Lemma 8.13. Let I be an A-submodule of R which is T-invertible in R. Then I is convex in A iff A is convex in R, and I is 2-saturated in A iff A is 2-saturated in R.

Proof. Apply Lemma 12 to the A-modules A, R, I. □

Definition 2. We call the ring extension $A \subset R$ T-*tight*, or say that A is T-*tight in* R, if for every $x \in R$ there exists some (R,T)-invertible ideal I of A with $Ix \subset A$.

Examples 8.14. i) If $A \subset R$ is a ring extension and $R = S^{-1}A$ with $S = A \cap R^*$, the ring A is T-tight in R for any preordering T of R. Indeed, if $x = \frac{a}{s} \in R$ is given ($a \in A, s \in S$), then $(As^2)x \subset A$, and As^2 is T-invertible in R.
ii) If A is Prüfer in R then, for every preordering T of R, A is T-tight in R. Indeed, let $x \in R$ be given. Choose an R-invertible ideal I of A with $Ix \subset A$. Then, as observed above (Example 12.ii), I^2 is T-invertible in R and $I^2x \subset A$. □

Lemma 8.15. If for any $x \in R$ there exists an (R,T)-invertible *convex* ideal I of A with $Ix \subset A$, then A is convex in R.

Proof. Let $x \in R$, $a \in A$ be given with $0 \leq_T x \leq_T a$. By the assumption, there exists an (R,T)-invertible convex ideal I of A such that $Ix \in A$, i.e. $x \in I^{-1}$. By Lemma 12, we see that I is convex in A iff A is convex in I^{-1}. Hence $x \in A$. Therefore, A is convex in R. □

Theorem 8.16. Assume that A is T-tight in R. The following are equivalent.
(i) A is Prüfer and 2-saturated in R.
(ii) Every R-regular ideal of A is 2-saturated and convex in A.
(iii) If $a \in A$ and I is an (R,T)-invertible ideal of A, then the ideal $I + Aa$ is 2-saturated and convex in A.
(iii′) Every (R,T)-invertible ideal K of A contains an (R,T)-invertible ideal I of A such that for every $a \in A$ the ideal $I + aA$ is 2-saturated and convex in A.
(iv) If I and J are finitely generated ideals of A and I^2 is (R,T)-invertible, then $I^2 + J^2$ is 2-saturated and convex in A.

Proof. (i) ⇒ (ii): Clear by Theorem 7.
(ii) ⇒ (iii) ⇒ (iii′) and (ii) ⇒ (iv): trivial.
(iii′) ⇒ (iii): We prove that any ideal J of A containing an (R,T)-invertible ideal I of A with the property listed in (iii′) is 2-saturated and convex in A.

Let $x \in A$ be given with $2x = a \in J$. Since $I + Aa$ is 2-saturated in A, we conclude that $x \in I + Aa \subset J$. Thus J is 2-saturated in A.

Let $x \in A$, $a \in J$ be given with $0 \le_T x \le_T a$. Again, since $I + Aa$ is convex in A, we conclude that $x \in I + Aa \subset J$. Thus J is convex in A.

(iii) $\Rightarrow$ (i): (a) Since by assumption every (R,T)-invertible ideal of A is convex in A, we know by Lemma 15 that A is convex in R.

(b) Let $x \in R$ be given. Since A is T-tight in R there exists some (R,T)-invertible ideal I of A having the property listed in (iii) with $Ix \subset A$. Then $I \subset I(A+Ax) \subset A$. As just proved, $I(A+Ax)$ is 2-saturated and convex in A, hence in R by (a). We conclude by Lemma 13 that $A + Ax$ is 2-saturated and convex in R. It follows by Theorem 7 (cf. there (iii) $\Rightarrow$ (ii)), that A is Prüfer in R.
(iv) $\Rightarrow$ (i): (a) We prove first that A is convex in R. Let $x \in R$ be given. We choose an (R,T)-invertible ideal I of A with $J := Ix \subset A$. By assumption, $I^2 = I^2 + A \cdot 0^2$ is 2-saturated and convex in A, and $I^2 x \subset A$. Hence A is convex in R by Lemma 15.

(b) We show that A is Prüfer in R. Let $x \in R$ be given. We again choose an (R,T)-invertible ideal I of A with $J := Ix \subset A$. By assumption, $I^2 + J^2 = I^2(A + Ax^2)$ is 2-saturated and convex in A, hence in R. Taking again into account that I^2 is (R,T)-invertible, we conclude by Lemma 13 that $A + Ax^2$ is 2-saturated and convex in R. Now Theorem 7 tells us that A is Prüfer in R. $\square$

It is the somewhat artificial looking condition (iii′) in this theorem which will turn out to be useful later (cf.Th.9.12 and Th.9.13), more than the less complicated condition (iii).

§9 Prüfer subrings and absolute convexity in f-rings

In f-rings, to be defined and discussed below, the theory of Prüfer subrings seems to be particularly well amenable to our methods. It is traditional to study f-rings within the category of lattice ordered rings. This category is slightly outside the framework we have used in §5 – §8. Thus some words of explanation are in order. Our main reference for lattice ordered rings and groups, and in particular for f-rings, is the book [BKW] by Bigard, Keimel and Wolfenstein.

We start with an abelian group G, using the additive notation. Assume that G is (partially) ordered in the usual sense, the ordering being compatible with addition. Thus $x \le y$ implies $x + z \le y + z$ and $-y \le -x$. We write $G^+ := \{x \in G \mid x \ge 0\}$, and we have $G^+ + G^+ \subset G^+$, $G^+ \cap (-G^+) = \{0\}$.

G is called *lattice-ordered* if G is a lattice with respect to its ordering. This means that the infimum and supremum

$$x \wedge y := \inf(x, y) , \quad x \vee y := \sup(x, y)$$

exist for any two elements x, y of G. As is well known, the lattice G is then automatically distributive [BKW, 1.2.14], and the group G has no torsion [BKW, 1.2.13].

We assume henceforth that G is a lattice ordered group. Clearly, for any $x, y, z \in G$ we have

$$(x + z) \wedge (y + z) = (x \wedge y) + z, \quad (x + z) \vee (y + z) = (x \vee y) + z,$$

and $(-x) \wedge (-y) = -(x \vee y)$.

For any $x \in G$ we define $x^+ := x \vee 0$, $x^- := (-x) \vee 0$. We have $x = x^+ - x^-$. Moreover, if $x = y - z$ with $y, z \in G$, then $y = x^+$, $z = x^-$ iff $y \wedge z = 0$, cf.[BKW, 1.3.4].

The *absolute value* $|x|$ of $x \in G$ is defined by $|x| := x \vee (-x)$. One proves easily that $|x| = x_+ + x_-$ [BKW, 1.3.10], more generally [BKW, 1.3.12],

$$|x - y| = (x \vee y) - (x \wedge y).$$

Of course, $|x| = 0$ iff $x = 0$, and $|x| = x$ iff $x \ge 0$.

We explicitly mention the following three facts about absolute values. Here x, y are any elements of G, and $n \in \mathbb{N}$ (The label "LO" alludes to "lattice ordered").

(LO1) $$|x| \le |y| \iff -y \le x \le y.$$

Indeed, $x \vee (-x) \leq y$ means that $x \leq y$ and $-x \leq y$, hence $x \leq y$ and $-y \leq x$.

(LO2) $$-|x| - |y| \;\leq\; x \wedge y \;\leq\; x \vee y \;\leq\; |x| + |y|$$

This follows from the trivial estimates $-|x|-|y| \leq x \leq |x|+|y|$ and $-|x|-|y| \leq y \leq |x| + |y|$.

(LO3) $$(nx)_+ \;=\; nx_+\,,\ (nx)_- \;=\; nx_-, \quad \text{hence} \quad |nx| = n|x|.$$

cf. [BKW, 1.3.7].

We now introduce a key notion for everything to follow.

Definition 1. We call a subgroup M of G *absolutely convex in* G, if $|x| \leq |a|$ implies $x \in M$ for any two elements x of G and a of M. (In [BKW] the term "solid" is used for our "absolute convex".)

On the other hand, convexity in G is defined as in §5, Definition 1. Of course, absolute convexity is a stronger property than convexity.

We will need three lemmas about absolutely convex subgroups, the first and the second being very easy.

Lemma 9.1. Every absolutely convex subgroup M of G is 2-saturated in G.

Proof. Let $x \in G$ be given with $2x \in M$. Then $2|x| = |2x|$ (cf. LO3 above), and $0 \leq |x| \leq 2|x|$. It follows that $x \in M$. $\square$

Lemma 9.2. Assume that M is a convex subgroup of the lattice ordered abelian group G. The following are equivalent.
(i) M is a sublattice of G (i.e. $x \wedge y \in M$ and $x \vee y \in M$ for any two elements x, y of M).
(ii) M is absolutely convex in G.
(iii) If $x \in M$ then $|x| \in M$.

Proof. (i) $\Rightarrow$ (ii): Let $a \in M$ and $x \in G$ be given with $|x| \leq |a|$. Then $|a| = a \vee (-a) \in M$, and we conclude from $0 \leq |x| \leq |a|$ and the convexity of M that $|x| \in M$, then from $-|x| \leq x \leq |x|$ (cf. LO1) that $x \in M$.
(ii) $\Rightarrow$ (iii): trivial.
(iii) $\Rightarrow$ (i): Let $a, b \in M$ be given. By assumption then $|a| \in M$, $|b| \in M$. As stated above (LO2), $-|a| - |b| \;\leq\; a \wedge b \;\leq\; a \vee b \;\leq\; |a| + |b|$. Since M is convex in G, this implies $a \wedge b \in M$, $a \vee b \in M$. $\square$

Lemma 9.3. Let I, J, K be absolutely convex subgroups of G. Then the subgroup $J + K$ is again absolutely convex and

$$I \cap (J + K) = (I \cap J) + (I \cap K).$$

This can be extracted from [BKW, Chap.2]. We give a direct proof of the theorem for the convenience of the reader, following arguments in [Ban, p.130 f].

Proof. i) We first verify the following: Let $a \in J^+$, $b \in K^+$, $y \in I^+$ and $y \leq a+b$. Then $y \in (I^+ \cap J^+) + (I^+ \cap K^+)$.

Starting with the triviality $y = a \wedge y + (y - a \wedge y)$, we obtain $y = a \wedge y + y + (-a) \vee (-y)$ and then

$$(*) \qquad y = a \wedge y \ + \ (y-a) \vee 0.$$

Now $0 \leq a \wedge y \leq y$ and $0 \leq a \wedge y \leq a$. Thus $a \wedge y \in I^+ \cap J^+$. We read off from $(*)$ that $(y-a) \vee 0 \in I^+$. Further $y - a \leq b$, hence $0 \leq (y-a) \vee 0 \leq b \vee 0 \in K^+$, hence $(y-a) \vee 0 \in K^+$, and we conclude that $(y-a) \vee 0 \in I^+ \cap K^+$.
ii) We use part i) with $I = G$ to verify that $J + K$ is absolutely convex in G. Let $x \in G$, $a \in J$, $b \in K$ be given with $|x| \leq |a+b|$. Then $0 \leq x^+ \leq |x| \leq |a+b| \leq |a| + |b|$. This implies, as proved, that $x^+ \in J+K$ and $|x| \in J+K$. Thus $x = 2x^+ - |x| \ \in \ J+K$.
c) Let now $a \in I \cap (J+K)$ be given. We have $a = b + c$ with $b \in J$, $c \in K$. Then we conclude from $|a| \leq |b| + |c|$ by (i) that $|a| \in (I \cap J) + (I \cap K)$. The groups $I \cap J$ and $I \cap K$ are absolutely convex in G. Thus, as proved, $(I \cap J) + (I \cap K)$ is absolutely convex in G. It follows that $a \in (I \cap J) + (I \cap K)$. This proves $I \cap (J+K) = (I \cap J) + (I \cap K)$. $\square$

We now switch to lattice ordered rings. A *ring* R (here always commutative, with 1) is called *lattice ordered*, if the set R is equipped with a partial ordering, which makes $(R,+)$ a lattice ordered abelian group, and such that $xy \geq 0$ for any two elements $x \geq 0$, $y \geq 0$ of R. Thus for $T := R^+$ the properties $T + T \subset T$, $T \cdot T \subset T$, $T \cap (-T) = \{0\}$ hold, but we do not demand that $x^2 \in T$ for $x \in R$.

We call T an *ordering* of R and sometimes speak of "the lattice ordered ring (R,T)".

A subring A of R is called an *ℓ-subring*, if A is a subring and a sublattice of R. We know by Lemma 2 that the absolutely convex subrings of R coincide with the convex ℓ-subrings of R.

A subset $\mathfrak{a}$ of R is called an *ℓ-ideal*, if $\mathfrak{a}$ is a *convex* ideal of R and a *sublattice* of R,*) equivalently (Lemma 2), if $\mathfrak{a}$ is an *absolutely convex* ideal of R.

Proposition 9.4. Let $A \subset R$ be a weakly surjective ring extension. Assume that A is lattice ordered and every R-regular ideal of A is absolutely convex in A (i.e. an ℓ-ideal). Then A is Prüfer in R.

*) The unitiated reader may object to this terminology, insisting that "ℓ" should just mean "sublattice". But observe that the ℓ-ideals, as defined here, are the kernels of the homomorphisms between lattice-ordered rings, cf.[BKW, §8.3].

PROOF. It follows from Lemma 3, applied to the lattice-ordered group $(A, +)$, that the lattice of R-regular ideals of A is distributive. Theorem II.2.8 tells us that A is Prüfer in R. $\square$

This proposition should be regarded as a preliminary result, already indicating that there are friendly relations between absolute convexity and the Prüfer property. The assumption that A is lattice ordered seems to be too weak to allow a good theory of Prüfer extensions beyond our results in Chapters I and II. But if A is an f-ring, to be defined in a minute, we will see later that the situation described in Proposition 4 is met rather often, for example for *every* Prüfer extension $A \subset R$ in case A has bounded inversion (cf.Theorems 9.15 and 10.12).

If $(C_\alpha \mid \alpha \in X)$ is a family of lattice ordered rings, the direct product $\prod\limits_{\alpha \in X} C_\alpha$ is again a lattice ordered ring in the obvious way: We equip the ring $C := \prod\limits_{\alpha \in X} C_\alpha$ with the ordering $f \le g \iff f(\alpha) \le g(\alpha)$ for every $\alpha \in X$, and we have, for $f, g \in C$, $\alpha \in X$,

$$(f \wedge g)(\alpha) = f(\alpha) \wedge g(\alpha) \quad , \quad (f \vee g)(\alpha) = f(\alpha) \vee g(\alpha).$$

{Explanation: If $h \in C$, we denote the component of h at the index α by $h(\alpha)$. Thus h is the family $(h(\alpha) \mid \alpha \in X)$.} Notice also that $f^+(\alpha) = f(\alpha)^+$, $f^-(\alpha) = f(\alpha)^-$, and $|f|(\alpha) = |f(\alpha)|$.

DEFINITION 2 [BKW, 9.11]. A lattice ordered ring R is called an f-*ring* if there exists a family $(C_\alpha \mid \alpha \in X)$ of *totally ordered* rings C_α, such that R is isomorphic (as an ordered ring) to an ℓ-subring of $\prod\limits_{\alpha \in X} C_\alpha$.

The following rules clearly hold in a totally ordered ring, hence in any f-ring R.
(F1) If $x \ge 0$ then $x(a \wedge b) = (xa) \wedge (xb)$.
(F2) If $x \ge 0$ then $x(a \vee b) = (xa) \vee (xb)$.
(F3) $|ab| = |a|\,|b|$.
(F4) $a^2 = |a|^2$.
(F5) If $a \ge 0$, $b \ge 0$, $x \ge 0$, $a \wedge b = 0$, then $a \wedge bx = 0$.
(F6) If $a \wedge b = 0$ then $ab = 0$.
(F7) $a + b = (a \wedge b) + (a \vee b)$.
(F8) $ab = (a \wedge b)(a \vee b)$.

Remarks. i) In any lattice ordered ring R the following weaker rules hold [BKW, 8.1.4]:
1) If $x \ge 0$ then $x(a \wedge b) \le xa \wedge xb$, $x(a \vee b) \ge xa \vee xb$.
2) $|ab| \le |a|\,|b|$

ii) It is known that *each* of the rules (F1), (F2), (F3), (F5) *characterizes* f-rings within the category of lattice ordered rings, thus allowing a more intrinsic definition of f-rings than Definition 2 above. {[BKW, p.173, 175 f]. Notice that, contrary to [BKW], our rings are always assumed to have a unit element. Thus [BKW, 9.1.14] applies.} □

In an f-ring R we have $x^2 \geq 0$ for every $x \in R$ (cf. F4). Thus
$R^+ = \{x \in R \mid x \geq 0\}$ is a partial ordering of R in the sense of §5, i.e. $T = R^+$ is a preordering of R with $T \cap (-T) = \{0\}$.

In the following WE ASSUME THAT R IS AN F-RING AND A IS A SUBRING OF R, if nothing else is said.

PROPOSITION 9.5. The following are equivalent.
(i) A is absolutely convex in R.
(ii) A is a convex ℓ-subring of R.
(iii) A is 2-saturated and convex in R.
(iv) A is convex and integrally closed in R.
(v) A is convex in R. If $x \in R$ and $x^2 \in A$ then $x \in A$.

PROOF. The implications (i) $\Rightarrow$ (iii) and (i) $\Leftrightarrow$ (ii) are covered by Lemmas 1 and 2, and (iii) $\Rightarrow$ (iv) is covered by Proposition 8.4. (iv) $\Rightarrow$ (v) is trivial.
(v) $\Rightarrow$ (i): If $x \in A$ then $|x|^2 = x^2 \in A$ by F4, hence $|x| \in A$. Lemma 2 tells us that A is absolutely convex in R. □

COROLLARY 9.6. If A is Prüfer and convex in R then A is absolutely convex in R. □

If M and I are subsets of R let $[I\colon M]$ or, if necessary, more precisely $[I\colon_R M]$ denote the set of all $x \in R$ with $xM \subset I$. Notice that, if I is an additive subgroup of R or an A-submodule of R, then also $[I\colon M]$ is an additive subgroup resp. an A-submodule of R.

DEFINITION 2. Let I, J be additive subgroups of R with $I \subset J$. We say that I is *absolutely convex in* J, if

$$x \in J,\ a \in I,\ |x| \leq |a| \Longrightarrow x \in I.$$

{The point here is that J is not assumed to be a sublattice of R. Thus the definition goes beyond Definition 1.}

LEMMA 9.7. Let I and J be additive subgroups of R with $I \subset J$. Assume that I is absolutely convex in J.
a) If M is any subset of R^+ then $[I\colon M]$ is absolutely convex in $[J\colon M]$.
b) If K is an additive subgroup and a sublattice of R, then $[I\colon K]$ is absolutely convex in $[J\colon K]$.

Proof. a): Let $x \in [I\colon M]$ and $y \in [J\colon M]$ be given with $0 \leq |y| \leq |x|$. For every $s \in M$ we have (using F3)

$$0 \leq s|y| = |sy| \leq s|x| = |sx|,$$

and $sx \in I$, $sy \in J$. Since I is absolutely convex in J, this implies $sy \in I$. Thus $y \in [I\colon M]$.
b): If $x \in K$, then $x = x^+ - x^-$ and $x^+ \in K$, $x^- \in K$. Thus $[I\colon K] = [I\colon K^+]$ and $[J\colon K] = [J\colon K^+]$. The claim now follows from a). □

Lemma 9.8. Assume that I is an absolutely convex additive subgroup of R.
a) $[I\colon x] = [I\colon |x|]$ for every $x \in R$.
b) For any subset K of R the additive group $[I\colon K]$ is absolutely convex in R.

Proof. a): Let $y \in [I\colon x]$ be given. We have $xy \in I$, hence (using F3)

$$|x|y^+ + |x|y^- = |x|\ |y| = |xy| \in I.$$

It follows that $|x|y_+$ and $|x|y_-$ both are elements of I. We conclude that $y = y^+ - y^- \in [I\colon |x|]$. This proves that $[I\colon x] \subset [I\colon |x|]$.

Let now $z \in [I\colon |x|]$ be given. Then $|zx| = |z \cdot |x|\ | \in I$, hence $zx \in I$, i.e. $z \in [I\colon x]$. This proves that $[I\colon |x|] \subset [I\colon x]$.
b): Let $M\colon = \{|x|\colon x \in K\}$. Using a) we obtain

$$[I\colon K] = \bigcap_{x \in K} [I\colon x] = \bigcap_{x \in K} [I\colon |x|] = [I\colon M].$$

Now apply Lemma 7.a with $J = R$. □

Lemma 9.9. Assume that A is absolutely convex in R. Then every R-invertible A-submodule of R is absolutely convex in R.

Proof. Let K be such an A-submodule. Then $K = [A\colon K^{-1}]$, and Lemma 8 applies. □

Theorem 9.10. The following are equivalent.
(1) A is Prüfer and convex in R.
(2) Every R-regular A-submodule of R is absolutely convex in R.
(3) For every $x \in R$ the A-module $A + Ax^2$ is absolutely convex in R.
(4) Every overring of A in R is absolutely convex in R.

Proof. (1) ⇒ (2): It suffices to prove that a given finitely generated R-regular A-submodule I is absolutely convex in R. Since A is Prüfer in R the A-module I is R-invertible. We know by Corollary 6 that A is absolutely convex in R. Now Lemma 9 tells us that I is absolutely convex in R.

(2) $\Rightarrow$ (3) and (2) $\Rightarrow$ (4): trivial.
(3) $\Rightarrow$ (1): It suffices to prove that A is Prüfer in R. By assumption $A = A + A \cdot 0$ is absolutely convex in R. We conlcude by Proposition 5 that A is integrally closed in R. Let $x \in R$ be given. We have $-1 - x^2 \le 2|x| \le 1 + x^2$. By (3) it follows that $2|x| \in A + Ax^2$, then that $|x| \in A + Ax^2$, finally that $x \in A + Ax^2$. Now Theorem I.5.2 tells us that A is Prüfer in R.
(4) $\Rightarrow$ (1): Let B be an R-overring of A. By assumption B is absolutely convex in R. It follows by Proposition 5 that B is integrally closed in R, then by Theorem I.5.2 that A is Prüfer in R. □

Lemma 9.11. Assume that A is absolutely convex in R.
a) Every R-invertible A-submodule I of R is R^+-invertible (cf. §8, Def.1) in R.
b) If A is tight in R, then A is R^+-tight in R (cf. §8, Def.2).

Proof. a): We know by Lemma 9 that I is absolutely convex in R. The same holds for I^{-1}. Since both I and I^{-1} are sublattices of R, they certainly are generated (as A-modules) by I^+ and $(I^{-1})^+$ respectively. Thus I is R^+-invertible in R.
b): Now obvious. □

Theorem 9.12. Assuming that A is an ℓ-subring of R, the following are equivalent.
(1) A is Prüfer and convex in R {hence absolutely convex in R by Lemma 2 or Cor.6}.
(2) A is tight in R, and every R-regular ideal of A is an ℓ-ideal of A.
(3) A is tight in R. For every R-invertible ideal I of A and every $a \in A$ the set $I + Aa$ is an ℓ-ideal of A.
(3′) A is tight in R. Every R-invertible ideal K of A contains an R-invertible ideal I of A such that $I + Aa$ is an ℓ-ideal of A for every $a \in A$.
(4) A is tight in R. For any two finitely generated ideals I, J of A with I invertible in R the set $I^2 + J^2$ is an ℓ-ideal of A.

Proof. (1) $\Rightarrow$ (2): The extension $A \subset R$ is tight since it is Prüfer. It follows by Theorem 10 that every R-regular ideal of A is absolutely convex in R, hence is absolutely convex in A.
(2) $\Rightarrow$ (3) $\Rightarrow$ (3′): trivial.
(3′) $\Rightarrow$ (1): We first prove that A is absolutely convex in R. Let $x \in R$ and $a \in A$ be given with $0 \le |x| \le |a|$. Since A is tight in R there exists an R-invertible ideal K of A such that $Kx \subset A$. By (3′) K contains an R-invertible ideal I of R having the property listed in (3′), i.e. $I + aA$ is an l-ideal of A for every $a \in A$. In particular I is absolutely convex in A, hence a sublattice of A, hence a sublattice of R. By Lemma 7.b we conclude that $A = [I{:}\,I]$ is absolutely convex in $I^{-1} = [A{:}\,I]$. We now infer from $0 \le |x| \le |a|$ and $x \in K^{-1} \subset I^{-1}$ that $x \in A$. Thus A is absolutely convex in R.

Lemma 11 tells us that A is T-tight in R, with $T = R^+$, and moreover, that all R-invertible ideals of A are (R,T)-invertible. We conclude by Theorem 8.16, using there the implication (iii$'$) $\Rightarrow$ (i), that A is Prüfer in R.

(4) $\Rightarrow$ (1): The proof runs the same way as for the implication (3$'$) $\Rightarrow$ (1). We now work with I^2 instead of I for I an R-invertible ideal such that $Ix \subset A$, and we use the implication (iv) $\Rightarrow$ (i) in Theorem 8.16. □

We also ask for criteria, in the vein of the preceding theorems 10 and 12, that A is Bezout and convex in R.

Theorem 9.13. a) The following are equivalent.
(1) A is Bezout and convex in R.
(2) For every $x \in R$ the A-module $A + Ax$ is principal and absolutely convex in R.
(3) A is an ℓ-subring of R, and $R = S^{-1}A$ with $S := A \cap R^*$. For every $a \in A$ and $s \in S$ the ideal $As + Aa$ of A is principal. For every $s \in S$ the ideal As is absolutely convex in A (i.e. an ℓ-ideal of A).
(3$'$) A is an ℓ-subring of R. There exists a multiplicative subset S of A with the following properties: $R = S^{-1}A$. For every $s \in S$ and $a \in A$ there exists some $t \in S$ such that $As + Aa = At$. For every $s \in A$ the ideal As is absolutely convex in A.

b) If $2 \in R^*$ then (1) – (3) are also equivalent to each of the following two conditions.
(4) $R = S^{-1}A$ with $S := A \cap R^*$. For every $s \in S$ and $a \in A$ the ideal $As^2 + Aa$ of A is principal and absolutely convex in A.
(4$'$) There exists a multiplicative subset S of A with $2 \in S$ and $R = S^{-1}A$, and such that, for every $a \in A$ and $s \in S$, the ideal $As^2 + Aa$ is principal and absolutely convex in A.

Comment. Given an f-ring A the somewhat artificial looking conditions (3$'$) and (4$'$) are useful for finding – theoretically – all Prüfer (hence Bezout) extensions $A \subset R$ such that R is an f-ring with $R^+ \cap A = A^+$ and A an ℓ-subring of R. Indeed, we will see in §10 (in a more general context) that, given a multiplicative subset S of A consisting of non-zero divisors of A, there exists a unique partial ordering on $R := S^{-1}A$ such that R is an f-ring, A is an ℓ-subring of R, and $R^+ \cap A = A^+$. (Actually it is not difficult, just an exercise, to give a direct proof of this fact.)

Proof of Theorem 9.13. (1) $\Rightarrow$ (2): Let $x \in R$ be given. Then $A + Ax$ is principal, since A is Bezout in R (cf.Th.II.10.2). It follows from Theorem 10 (cf. there (1) $\Rightarrow$ (2)) that $A + Ax$ is absolutely convex in R.
(2) $\Rightarrow$ (1): trivial.
(1) $\Rightarrow$ (3): Let $S := A \cap R^*$. Theorem II.10.16 tells us that $R = S^{-1}A$. We further know by Theorem 10 above (cf. there (1) $\Rightarrow$ (2)) that, for every $s \in S$

and $a \in A$, the ideal $As + Aa$ is absolutely convex in R, hence in A. Since A is Bezout in R, this ideal is also principal (cf. Th.II.10.2).
(3) $\Rightarrow$ (3′): The set $S := A \cap R^*$ has all the properties listed in (3′). This needs a verification only for the second one. Let $s \in S$ and $a \in A$ be given. By assumption (3), $As + Aa = At$ for some $t \in A$. We have $s = bt$ with some $b \in A$, and we conclude that $t \in A \cap R^* = S$.
(3′) $\Rightarrow$ (1): Let $a_1, \dots, a_r \in A$ and $s \in S$ be given. Then there exists some $t \in S$ such that $As + Aa_1 + \cdots + Aa_r = At$. Indeed, this holds for $r = 1$ by assumption (3′) and then follows for all r by an easy induction. Now Theorem 12 tells us (implication (3′) $\Rightarrow$ (1) there) that A is (Prüfer and) convex in R.
Let $x \in R$ be given. Write $x = \frac{a}{s}$ with $a \in A$, $s \in S$. Then $A + Ax = S^{-1}(As + Aa)$ and $As + Aa = At$ with $t \in S$. Thus the A-module $A + Ax$ is principal, and we conclude that A is Bezout in R (cf.Th.II.10.2).
(3) $\Rightarrow$ (4) $\Rightarrow$ (4′): trivial.
(4′) $\Rightarrow$ (1): We learn from Theorem 8.9 that A is convex in R. Let $x \in R$ be given. Write $x = \frac{a}{s^2}$ with $a \in A$, $s \in S$. The ideal $As^2 + Aa$ is principal by assumption (4′). Thus the module $A + Ax = s^{-2}(Aa + As^2)$ is principal. This proves that A is Bezout in R. □

Open Question. If A is a convex (hence absolutely convex) Prüfer subring of R, does it follow that A is Bezout in R?

We will now see that the answer is "Yes" if R or (equivalently) A has bounded inversion. Related to this, we will find more criteria, that A is Bezout in R, and results about such extensions more precise than those stated in Theorem 13.

We store our results in the following lengthy theorem 15. Here the dashed conditions (2′), (3′), (4′), (6′) are included in order to make the proof more transparent, while the undashed conditions (1) – (8) are the more interesting ones. For the proof we will need (a special case of) the following easy lemma.

Lemma 9.14. Let I be a 2-saturated additive subgroup of R. Assume that every $x \in R$ with $x^2 \in I$ is an element of I. Then I is a sublattice of R.

Proof. If $x \in I$ then $|x|^2 = x^2 \in I$, hence $|x| \in I$. It follows that $2x^+ = x + |x| \in I$ and then that $x^+ \in I$. Given elements $x, y \in I$ we conclude that

$$x \vee y = y + [(x - y) \vee 0] = y + (x - y)^+ \in I. \qquad \square$$

Theorem 9.15. The following are equivalent.
(1) A has bounded inversion and is Prüfer in R.
(2) R has bounded inversion. A is convex in R.
(2′) R has bounded inversion. A is absolutely convex in R.
(3) A is convex in R. For every $x \in R$, $A + Ax = A(1 + |x|)$.
(3′) A is absolutely convex in R. For every $x \in R$, $A + Ax = A(1 + |x|)$.

(4) For every $x \in R$, $A + Ax = A(1 + |x|)$, and this module is convex in R.
(4′) For every $x \in R$, $A+Ax = A(1+|x|)$, and this module is absolutely convex in R.
(5) R has bounded inversion. A is Bezout and convex in R.
(6) A is convex in R. For every $x \in R$, $A + Ax = A(1 \vee |x|)$.
(6′) For every $x \in R$ the module $A+Ax$ is absolutely convex in R, and $A+Ax = A(1 \vee |x|)$.
(7) $2 \in R^*$, and $R = S^{-1}A$ with $S := A \cap R^*$. For every $a \in A$, $s \in A$, the ideal $As^2 + Aa$ is an ℓ-ideal of A, and $As^2 + Aa = A(s^2 + |a|)$.
(8) There exists a multiplicative subset S of A such that $2 \in S$, $R = S^{-1}A$, and $As^2 + Aa$ is an ℓ-ideal of A for every $a \in A$ and $s \in S$.

Comment. Given an f-ring A, this time with bounded inversion, condition (8) is useful for finding – theoretically – all Prüfer (hence Bezout) extensions $A \subset R$ such that R is an f-ring with $R^+ \cap A = A^+$ and A is an ℓ-subring of R, cf. the comment following Theorem 13.

Proof of Theorem 9.15.
(1) ⇔ (2): This is covered by Theorem 7.2.
(2) ⇒ (2′): $2 \in R^*$, since R has bounded inversion. $\frac{1}{2} \in A$, since A is convex in R. Thus A is 2-saturated in R. The ring A is also convex in R. By Proposition 5 we conclude that A is absolutely convex in R.
(2′) ⇒ (2): trivial.

We now know that conditions (1), (2), (2′) are equivalent.

(1) ∧ (2) ⇒ (3′): A is Prüfer and convex in R. Let $x \in R$ be given. Theorem 10 tells us that the module $A + Ax$ is absolutely convex in R, since this module is R-regular. In particular, $|x| \in A + Ax$, hence $1 + |x| \in A + Ax$. This proves that $A(1 + |x|) \subset A + Ax$. On the other hand, $1 + |x| \in R^*$ by (2), and $(1+|x|)^{-1} \leq 1$, hence $(1+|x|)^{-1} \in A$. We also have $|x \cdot (1+|x|)^{-1}| \leq 1$, hence $x(1 + |x|)^{-1} \in A$. It follows that $1 \in A(1 + |x|)$ and $x \in A(1 + |x|)$, hence $A + Ax \subset A(1 + |x|)$. Thus $A + Ax = A(1 + |x|)$.
(3′) ⇒ (3): trivial.
(3) ⇒ (2): If $x \in R$ and $x \geq 1$ then, by (3),

$$A + Ax = A + A(x - 1) = A(1 + x - 1) = Ax.$$

Thus $1 \in Ax$, which implies $x \in R^*$. This proves that R has bounded inversion.

We now know that all conditions (1) – (3′) are equivalent.

(1) ∧ (3′) ⇒ (4′): A is Prüfer in R by (1) and absolutely convex in R by (3′). Theorem 10 tells us again that, for every $x \in R$, the module $A + Ax$ is absolutely convex in R. Also $A + Ax = A(1 + |x|)$ by (3′).
(4′) ⇒ (4) ⇒ (3): trivial.

We have proved the equivalence of all conditions (1) – (4′).

(2) $\wedge$ (4) $\Rightarrow$ (5): R has bounded inversion by (2). For every $x \in R$ the A-module $A + Ax$ is principal by (4). Thus A is Bezout in R (cf.Th.II.10.2). $A + Ax$ is also convex in R by (4). In particular ($x = 0$), A is convex in R.
(5) $\Rightarrow$ (2): trivial.
(4′) $\Rightarrow$ (6′): Let $x \in R$ be given. The module $A+Ax$ is absolutely convex in R, and $A+Ax = A(1+|x|)$. We have $1 \vee |x| \leq 1+|x|$ Thus $A(1+|x|) \supset A(1\vee|x|)$. Now $1 \vee |x| = 1 + y$ with $y \in R^+$. Thus $A(1 \vee |x|) = A + Ay$, and this module is again absolutely convex in R. Since $1 + |x| \leq 2(1 \vee |x|)$ we infer that $A(1 + |x|) \subset A(1 \vee |x|)$, and conclude that $A(1 + |x|) = A(1 \vee |x|)$.
(6′) $\Rightarrow$ (6): trivial.
(6) $\Rightarrow$ (2): For every $x \in R$ with $x \geq 1$ we have $A+Ax = Ax$, since $1 \vee |x| = x$. It follows that $1 \in Ax$, hence $x \in R^*$. Thus R has bounded inversion.

We have proved the equivalence of all conditions (1) – (6′).

(1) – (6′) $\Rightarrow$ (7): R has bounded inversion, hence $2 \in R^*$. Since A is Bezout in R, we have $R = S^{-1}A$ with $S := R^* \cap A$ (cf.Prop.II.10.16 or Th.13). Let $s \in S$ and $a \in A$ be given. By (3),

$$As^2 + Aa = s^2 \left(A + \frac{a}{s^2}\right) = s^2 A \left(A + \frac{|a|}{s^2}\right) = A(s^2 + |a|).$$

By (4) the module $A\left(1 + \frac{|a|}{s^2}\right)$ is absolutely convex in R. It follows that $A(s^2 + |a|)$ is absolutely convex in R, hence in A, i.e. $A(s^2 + |a|)$ is an ℓ-ideal of A.
(7) $\Rightarrow$ (8): trivial.
(8) $\Rightarrow$ (3): Theorem 8.9 tells us that A is Prüfer and convex in R. Let $x \in R$ be given. Write $x = \frac{a}{s^2}$ with $a \in A$, $s \in S$. Then

$$A + Ax = s^{-2}(As^2 + Aa) = s^{-2}A(s^2 + |a|) = A(1 + |x|). \qquad \square$$

§10 Rings of quotients of an f-ring

In the following A is an f-ring. We will study overrings of A in the complete ring of quotients $Q(A)$. For the general theory of $Q(A)$ we refer to Lambek's book [Lb]. (Some facts had been recapitulated in I §3.)

Recall that every element of $Q(A)$ can be represented by an A-module homomorphism $f\colon I \to A$ with I a dense ideal of A. More precisely

$$Q(A) = \varinjlim_{I \in \mathcal{D}(A)} \mathrm{Hom}_A(I, A)$$

with $\mathcal{D}(A)$ denoting the direct system of dense ideals of A, the ordering being given by reversed inclusion, $I \le J$ iff $I \supset J$. Most often we will not distinguish between such a homomorphism $f\colon I \to A$ and the corresponding element $[f]$ of $Q(A)$.

Our first goal in the present section is to prove that *there exists a unique partial ordering U on $Q(A)$ which makes $Q(A)$ an f-ring in such a way that $U \cap A = A^+$ and A is an ℓ-subring of* $Q(A)$. This is an important result due to F.W. Anderson [And]. Anderson's paper is difficult to read since he establishes such a result also for certain non commutative f-rings. For the convenience of the reader we will write down a full proof in the much easier commutative case. We then will prove the same for suitable overrings R of A in $Q(A)$ instead of $Q(A)$ itself. Among these overrings will be all Prüfer extensions of A.

Whenever it seems appropriate we will work in an arbitrary overring R of A in $Q(A)$ instead of $Q(A)$ itself. Recall that, up to isomorphism over A, these rings are all the rings of quotients of A.

Lemma 10.1. Let $a \in A^+$, $b \in A$. Then $(ab)^+ = ab^+$ and $(ab)^- = ab^-$.

Proof. $ab = ab^+ - ab^-$. Applying the property (F1) from §9 we obtain $(ab^+) \wedge (ab^-) = a(b^+ \wedge b^-) = 0$. This proves the claim. □

Corollary 10.2. If a, b, s are elements of A with $a \ge 0$, $b \ge 0$, $a = bs$, then $a = bs^+$, $0 = bs^-$.

Proof. By the lemma we have $bs^+ = (bs)^+ = a$, $bs^- = (bs)^- = 0$. □

Definitions 1 a) We call a *subset* M of A *dense in* A, if the ideal AM generated by M is dense in A. This means that for every $x \in A$ with $x \neq 0$ there exists some $m \in M$ with $xm \neq 0$.
b) If I is any ideal of A let $I^{(2)}$ denote the set $\{a^2 \mid a \in I\}$. □

Lemma 10.3. If I is a dense ideal of A the set $I^{(2)}$ is also dense in A.

Proof. Let $x \in A$ be given with $xI^{(2)} = 0$. For any two elements a, b of I we have $xa^2 = 0$, $xb^2 = 0$, $x(a+b)^2 = 0$. It follows that $2xab = 0$, and then that $xab = 0$, since the additive group of A has no torsion. Thus $xI^2 = 0$. Since I is dense in A we conclude that $xI = 0$ and then that $x = 0$. $\square$

Corollary 10.4. If I is dense ideal of A then I^+ is dense in A. $\square$

Lemma 10.5. Let M be a subset of A^+ which is dense in A. Assume that x is an element of $Q(A)$ with $xM \subset A^+$. Then $x \cdot (A\colon x)^+ \subset A^+$.

Proof. Let $a \in (A\colon x)^+$ be given. If $d \in M$, then $(ax)d = (xd)a \in A^+$ and $ax \in A$. It follows that $(ax)^- d = 0$ by Corollary 2 above. Since M is dense in A we conclude that $(ax)^- = 0$, hence $ax \in A^+$. $\square$

In the following R *is an overring of* A *in* $Q(A)$. We introduce the set

$$U := \{x \in R \mid x \cdot (A\colon x)^+ \subset A^+\}.$$

Due to Corollary 4 and Lemma 5 we can say, that U is the set of elements x of R such that there exists some dense subset M of A with $M \subset A^+$ and $Mx \subset A^+$.

Proposition 10.6.
i) U is a partial orderring of R with $x^2 \in U$ for every $x \in R$, and $U \cap A = A^+$.
ii) If T is any preordering of R with $T \cap A \subset A^+$ then $T \subset U$.

Proof. i): If $x \in U \cap (-U)$ then $x(A\colon x)^+$ is contained in $A^+ \cap (-A^+) = \{0\}$. Since $(A\colon x)^+$ is dense in A (cf.Cor.4), we conclude that $x = 0$. Thus $U \cap (-U) = \{0\}$.

Let $x, y \in U$ be given. We choose dense subsets M, N of A with $M \subset A^+$, $N \subset A^+$, $Mx \subset A^+$, $Ny \subset A^+$. The set $MN = \{uv \mid u \in M, v \in N\}$ is again dense in A and contained in A^+, and $MN(x+y) \subset A^+$, $MN(x \cdot y) \subset A^+$. Thus $U + U \subset U$ and $U \cdot U \subset U$.

Finally let $x \in U$ and $I := (A\colon x)$. We know by Lemma 3 that the subset $I^{(2)}$ of A^+ is dense in A. Since $x^2 I^{(2)} \subset A^+$, we conclude that $x^2 \in U$.

If $x \in A$ then $(A\colon x) = A$. The condition $A^+ x \subset A^+$ means that $x \in A^+$. Thus $U \cap A = A^+$.

ii): Let T be a preordering of R with $T \cap A \subset A^+$. For any $x \in T$ we have $(A\colon x)^+ \cdot x \subset T \cap A \subset A^+$, hence $x \in U$. Thus $T \subset U$. $\square$

Remark. In part ii) of the theorem we do not fully need the assumption that T is a preordering of R. It suffices to know that T is a subset of R with $T \cdot T \subset T$ and $T \cap A \subset A^+$. $\square$

Definition 2. We call U the *canonical ordering on* R induced by the ordering A^+ of A. If necessary, we write U_R instead of U. Notice that $U_R = R \cap U_{Q(A)}$. □

Lemma 10.7. Let M be a subset of A which is dense in A. Then M is dense in $Q(A)$.

Proof. Let $x \in Q(A)$ be given with $Mx = 0$. Then $M \cdot (A\colon x)x = 0$. This implies $(A\colon x)x = 0$ and then $x = 0$, since $(A\colon x)$ is dense in $Q(A)$. □

Proposition 10.8. Assume that T is a partial ordering of R with $T \cap A = A^+$. Assume further that (R, T) is an f-ring. Then $T = U_R$.

Proof. We write $U\colon = U_R$. We know by Proposition 6 that $T \subset U$. We now prove that also $U \subset T$.

In the f-ring (R, T) we use standard notation from previous sections: $T = R^+$, $x \le y$ iff $y - x \in T$, etc. Let $x \in U$ be given. We have to verify that $x \ge 0$, i.e. $x^- = 0$. Suppose that $x^- \ne 0$. The set $M\colon = (A\colon x)^+$ is dense in A by Corollary 4, hence dense in R by Lemma 7. Thus there exists some $s \in M$ with $sx^- \ne 0$. Since R is an f-ring and $s \in A^+ \subset R^+$, we conclude by Lemma 1 that $(sx)^- = sx^- \ne 0$. But $sx \in U \cap A = A^+ \subset R^+$. This is a contradiction. Thus $x^- = 0$. □

Definition 3. An *f-extension* of the f-ring A is an f-ring R which contains A as an ℓ-subring such that $R^+ \cap A = A^+$.

Theorem 10.9 (F.W. Anderson [And]). There exists a unique partial ordering T on $Q(A)$ such that $(Q(A), T)$ is an f-extension of A. This ordering T is the canonical ordering $U = U_{Q(A)}$ induced by A^+ on $Q(A)$.

Proof. We know by Proposition 8 that U is the only candidate for a partial ordering T on $Q(A)$ with these properties. We endow $Q(A)$ with the ordering U and write $U = Q(A)^+$.

Step 1. We first prove that $Q(A)$ is lattice ordered. Given $x \in Q(A)$ it suffices to verify that $x \vee 0 = \sup(x, 0)$ exists in $Q(A)$. We give an explicit construction of $x \vee 0$.

Claim. Let $a_1, \dots, a_n \in (A\colon x)^+$ and $b_1, \dots, b_n \in A$ be given with $\sum_{i=1}^{n} a_i b_i = 0$. Then $\sum_{i=1}^{n} (a_i x)^+ b_i = 0$.

Proof of the claim. Let $c \in (A\colon x)^+$. It follows by Lemma 1 from $(cx)a_i = c(a_i x)$ that $(cx)^+ a_i = (cxa_i)^+ = c(a_i x)^+$. Thus

$$c \sum_{i=1}^{n} (a_i x)^+ b_i = (cx)^+ \sum_{i=1}^{n} a_i b_i = 0.$$

Since $(A\colon x)^+$ is dense in A we obtain $\sum\limits_{i=1}^{n}(a_ix)^+b_i = 0$, as desired.

Thus there exists a well defined homomorphism $h\colon (A\colon x)^+A \to A$ of A-modules with

$$h\left(\sum_{i=1}^{n} a_ib_i\right) := \sum_{i=1}^{n}(a_ix)^+b_i$$

for all $n \in \mathbb{N}, a_i \in (A\colon x)^+$, $b_i \in A$. The map h may be viewed as an element of $Q(A)$. Notice that for every $a \in (A\colon x)^+$ we have $ah = ha = (ax)^+$.

We want to prove that $h = x \vee 0$. From $(A\colon x)^+h \subset A^+$ we conclude that $h \geq 0$. For any $a \in (A\colon x)^+$ we have $(h-x)a = h(a) - xa = (xa)^+ - xa = (xa)^- \in A^+$. Thus $h \geq x$.

Let $y \in Q(A)$ be given with $y \geq 0$ and $y \geq x$. For any $a \in (A\colon x)^+ \cap (A\colon y)^+$ the products ax, ay are in A and $ay \geq 0$, $ay \geq ax$, hence $ay \geq (ax)^+$, where, of course, $(ax)^+$ means $\sup_A(ax, 0)$. It follows that $a(y-h) \geq (ax)^+ - ah = 0$. Since $(A\colon x)^+ \cap (A\colon y)^+$ is dense in A we conclude that $y - h \geq 0$, i.e. $y \geq h$. This finishes the proof that $h = x \vee 0$.

Step 2. We prove that A is a sublattice of $Q(A)$. It suffices to verify for a given $x \in A$ that the element h constructed in Step 1 coincides with $\sup_A(x, 0) = x^+$. We have $(A\colon x)^+ = A^+$, hence by Step 1, for any $a \in A^+$, $ah = (ax)^+ = ax^+$ (cf.Lemma 1). Since A^+ is dense in $Q(A)$ it follows that indeed $h = x^+$.

Step 3. We now may use the notation x^+, x^- for any $x \in A$ unambiguously, since x^+, x^- means the same by regarding x as an element of the lattice A or of the lattice $Q(A)$. Our proof in Step 1 tells us that, for any $x \in Q(A)$, $a \in (A\colon x)^+$ we have

$$(ax)^+ = ax^+. \tag{$*$}$$

Indeed, this is just the statement that $h(a) = (ax)^+$ from Step 1. We now can prove that $Q(A)$ is an f-ring by verifying

$$s(x \vee y) = (sx) \vee (sy) \tag{$**$}$$

for given elements $x, y \in Q(A)$ and $s \in Q(A)^+$. ([BKW, 9.1.10]; we mentioned this criterion for a lattice ordered ring to be an f-ring in §9.) Subtracting sy on both sides we see that it suffices to prove $(**)$ in the case $y = 0$, i.e.

$$sx^+ = (sx)^+. \tag{$***$}$$

In order to verify this identity for given $x \in Q(A), s \in Q(A)^+$ we introduce the ideal $I\colon = ((A\colon x)\colon s) \cap (A\colon sx)$, which is dense in A. {Observe that $(A\colon x) \cdot (A\colon s) \subset ((A\colon x)\colon s)$.} For $a \in I^+$ we have, by use of $(*)$, $a(sx)^+ = (asx)^+$ since $a \in (A\colon sx)^+$, and $asx^+ = (asx)^+$ since $as \in (A\colon x)^+$. Thus $a[sx^+ - (sx)^+] = 0$

for every $a \in I^+$. Since I^+ is dense in $Q(A)$, we conclude that $sx^+ = (sx)^+$, as desired. This finishes the proof that $Q(A)$ is an f-ring. $\square$

We want to extend Theorem 9 to suitable subrings of $Q(A)$ containing A. These are the rings of type $A_{[\mathcal{F}]}$ occuring already in Theorem II.3.5 (with $R = Q(A)$ there), but now we use a more professional terminology.

Definition 4. Let A be any ring (commutative, with 1, as always). As previously let $J(A)$ denote the set of all ideals of A. We call a subset $\mathcal{F}$ of $J(A)$ a *filter on* A, if the following holds:
(1) $I \in \mathcal{F}, J \in J(A), I \subset J \;\Rightarrow\; J \in \mathcal{F}$.
(2) $I \in \mathcal{F}, J \in \mathcal{F} \;\Rightarrow\; I \cap J \in \mathcal{F}$.
(3) $A \in \mathcal{F}$.
We call a filter $\mathcal{F}$ *multiplicative* if instead of (2) the following stronger property holds:
(4) $I \in \mathcal{F}, J \in \mathcal{F} \;\Rightarrow\; IJ \in \mathcal{F}$. $\square$

We say that $\mathcal{F}$ is *of finite type* if the following holds.
(5) If $I \in \mathcal{F}$ there exists a finitely generated ideal I_0 of A with $I_0 \in \mathcal{F}$ and $I_0 \subset I$.

Notice that the subsets $\mathcal{F}$ of $J(A)$ considered in II, §3 with the properties R0-R2 (resp. R0-R3) there are just the multiplicative filters (resp. mutliplicative filters of finite type) on A.

Examples. 1) The set $\mathcal{D}(A)$ consisting of all dense ideals of A is a multiplicative filter on A.
2) If $A \subset R$ is any ring extension then the set $\mathcal{F}(R/A)$ of R-regular ideals of A is a multiplicative filter of finite type on A. $\square$

By definition we have

$$Q(A) = \varinjlim_{I \in \mathcal{D}(A)} \operatorname{Hom}_A(I, A).$$

If $\mathcal{F}$ is any filter on A contained in $\mathcal{D}(A)$ then we can form the ring

$$A_{\mathcal{F}} := \varinjlim_{I \in \mathcal{F}} \operatorname{Hom}_A(I, A).$$

in an analogous way. Since for any $I \in \mathcal{F}$ the natural map $\operatorname{Hom}_A(I, A) \to Q(A)$ is injective, we may – and will – regard $A_{\mathcal{F}}$ as a subring of $Q(A)$. For the smallest filter $\{A\}$ we obtain $A_{\{A\}} = A$. Thus $A \subset A_{\mathcal{F}} \subset Q(A)$. We have

$$A_{\mathcal{F}} \;=\; \{x \in Q(A) \mid (A\colon x) \in \mathcal{F}\} \;=\; \{x \in Q(A) \mid \exists\, I \in \mathcal{F} \text{ with } Ix \subset A\}.$$

Thus $A_{\mathcal{F}}$ is the ring $A_{[\mathcal{F}]}$ in the terminology of II, §3 (cf. Theorem II.3.5), with $R = Q(A)$ there.

Definition 4. We call a filter $\mathcal{F}$ on A *positively generated* if for any $I \in \mathcal{F}$ also $I^+A \in \mathcal{F}$.

Remark. If $\mathcal{F}$ is any filter on A then a *base* $\mathfrak{B}$ of $\mathcal{F}$ is a subset $\mathfrak{B}$ of $\mathcal{F}$ such that for every $I \in \mathcal{F}$ there exists some $K \in \mathfrak{B}$ with $K \subset I$. Of course, if $\mathcal{F}$ has a base $\mathfrak{B}$ such that $K^+A \in \mathcal{F}$ for every $K \in \mathfrak{B}$, then $\mathcal{F}$ is positively generated.

Examples 10.10. i) $\mathcal{D}(A)$ is positively generated. This is the content of Corollary 4 above.
ii) If $\mathcal{F}$ is a multiplicative filter of finite type then $\mathcal{F}$ is positively generated. Indeed, let $\mathfrak{B}$ be the set of finitely generated ideals $I \in \mathcal{F}$. It is a base of $\mathcal{F}$. If $I = Aa_1 + \cdots + Aa_n \in \mathcal{F}$, then $I^{n+1} \subset Aa_1^2 + \cdots + Aa_n^2 \subset I^+A$. Thus $I^+A \in \mathcal{F}$.
iii) Assume that $\mathcal{F}$ has a base $\mathfrak{B}$ consisting of ideals I which are sublattices of A. Then $\mathcal{F}$ is positively generated. Indeed, if $I \in \mathfrak{B}$ and $x \in I$, then $x = x^+ - x^-$ and $x^+, x^- \in I^+$. Thus $I = I^+A$.

Proposition 10.11. Assume that $\mathcal{F}$ is a positively generated multiplicative filter consisting of dense ideals.
i) $A_{\mathcal{F}}$ is an ℓ-subring of $Q(A)$. Thus, with the ordering $A_{\mathcal{F}}^+ := A : f \cap Q(A)$ on $A_{\mathcal{F}}$, both $A \subset A_{\mathcal{F}}$ and $A_{\mathcal{F}} \subset Q(A)$ are f-extensions.
ii) Let $x \in Q(A)$. Then $x \in A_{\mathcal{F}}^+$ iff there exists some $I \in \mathcal{F}$ with $I^+x \subset A^+$.

Proof. i): We verify for a given $x \in A_{\mathcal{F}}$ that $x^+ = x \vee 0 \in A_{\mathcal{F}}$. We choose some $I \in \mathcal{F}$ with $Ix \subset A$. For $a \in I^+$ we have $ax^+ = (ax)^+ \in A^+$. Thus $(I^+A)x^+ \subset A$. Since $I^+A \in \mathcal{F}$ we conclude that $x^+ \in A_{\mathcal{F}}$.
ii): Let $R := A_{\mathcal{F}}$. If $x \in Q(A)$ and $I^+x \subset A^+$ for some $I \in \mathcal{F}$ then $x \in Q(A)^+$ by definition of the ordering of $Q(A)$, since $I \in \mathcal{D}(A)$. Also $x \in A_{\mathcal{F}} = R$, since $I^+A \in \mathcal{F}$ and $(I^+A)x \subset A$. Thus $x \in R \cap Q(A)^+ = R^+$. Conversely, if $x \in R^+$, we choose some $I \in \mathcal{F}$ with $Ix \in A$. Then $I^+x \subset R^+ \cap A = A^+$. $\square$

We arrive at our main result in this section. It generalizes Theorem 9 to ws extensions of A.*) We write it down in an explicit way avoiding the technical notion of canonical ordering.

Theorem 10.12. Let A be an f-ring and $A \subset R$ a ws extension of A.

i) There exists a unique partial ordering R^+ on R such that R, equipped with this ordering, is an f-extension of A. Moreover $Q(A)$ is an f-extension of R.
ii) R^+ is the set of all $x \in R$ such that $(A : x)^+ \cdot x \subset A^+$.
iii) R^+ is the set of all $x \in R$ such that there exists some dense subset M of A with $M \subset A^+$ and $Mx \subset A^+$.

*) Recall that "ws" abbreviates "weakly surjective" (I, §3).

iv) Every overring of A in R, which is ws over A, is an ℓ-subring of R.

Proof. Defining R^+ by $R^+ := U_R = \{x \in R \mid (A:x)^+ x \subset A^+\}$ we know from above (Propositions 6 and 8), that R^+ is a partial ordering of R, and that this is the only candidate such that (R, R^+) is an f-ring and $R^+ \cap A = A^+$. We further know from above (Lemma 5) that, given a dense subset M of A with $M \subset A^+$, any $x \in R$ with $Mx \subset A^+$ is an element of R^+.

Let $\mathcal{F}$ denote the filter on A consisting of the R-regular ideals of A, $\mathcal{F} := \mathcal{F}(R/A)$. As observed above (Example 10.iii), $\mathcal{F}$ is positively generated. It follows by Propositions 11 and 8 that $A_{\mathcal{F}}$, equipped with the canonical ordering induced by A^+, is an f-ring, and both $A \subset A_{\mathcal{F}}$ and $A_{\mathcal{F}} \subset R$ are f-extensions.

Clearly $R \subset A_{\mathcal{F}}$, since $(A:x) \in \mathcal{F}$ for every $x \in R$ (Recall Th.I.3.13.) Conversely, if $x \in A_{\mathcal{F}} \subset Q(A)$, there exists some $I \in \mathcal{F}$ with $Ix \in A$. Multiplying by R we obtain $Rx = RIx \subset R$, i.e. $x \in R$. Thus $R = A_{\mathcal{F}}$. Now claims i) – iii) are evident.

Finally, if B is an overring of A in R which is ws over A, then applying what we have proved to $A \subset B$ instead of $A \subset R$, we see that B is an ℓ-subring of $Q(A)$, hence an ℓ-subring of R. $\square$

We continue to assume that A is an f-ring. We write down two corollaries of Theorem 12. Nothing new is needed to prove them.

Corollary 10.13. Let S be a multiplicative subset of A consisting of non-zero divisors. There is a unique partial ordering $(S^{-1}A)^+$ on $S^{-1}A$ such that $S^{-1}A$ becomes an f-extension of A. We have

$$(S^{-1}A)^+ = \left\{ \frac{a}{s^2} \mid a \in A^+, s \in S \right\} = \left\{ \frac{a}{s} \mid a \in A^+, s \in S^+ \right\}.$$

With this ordering $S^{-1}A$ is an ℓ-subring of $Q(A)$. $\square$

Corollary 10.14. Let $A \subset R$ be a Prüfer extension. There is a unique partial ordering R^+ on R such that R becomes an f-extension of A. An element x of R lies in R^+ iff there exists an invertible (or: R-invertible) ideal I of A with $I^+ x \subset A^+$, or alternatively, with $I^{(2)} x \subset A^+$. With this ordering $S^{-1}A$ is an ℓ-subring of $Q(A)$. $\square$

Henceforth we equip every overring R of A in $Q(A)$ with the canonical ordering R^+ induced by A^+. If $A \subset R$ is Prüfer, or more generally ws, R is an f-ring and both $A \subset R$ and $R \subset Q(A)$ are f-extensions.

It now makes sense to define an "absolute" Prüfer convexity cover of A, as announced at the end of §7.

Definition 5. Let $P_c(A)$ denote the polar $C(P(A)/A)^{\circ}$ of the convex hull $C(P(A)/A)$ of A in the f-ring $P(A)$ (over A, in $P(A)$). We call $P_c(A)$ the *Prüfer convexity cover of* A. □

From Theorem 7.9 we read off the following fact.

Theorem 10.15. $P_c(A)$ is the unique maximal overring E of A in $Q(A)$ (thus, up to isomorphy over A, the unique maximal ring of quotients of A), such that A is Prüfer and convex in E. □

Remarks 10.16. i) It follows, say, from Theorem 9.10, that every A-submodule I of $P_c(A)$, which is $P_c(A)$-regular, is *absolutely* convex in $P_c(A)$. In particular this holds for every overring of A in $P_c(A)$. Thus we may replace the word "convex" in Theorem 15 by "absolutely convex".

ii) If A has bounded inversion, it follows from Theorem 7.2 that $P_c(A) = P(A)$. Also now every overring of A in $P(A)$ has again bounded inversion (cf.Th.9.15).

iii) For R any overring of A in $Q(A)$ we obtain the Prüfer convexity cover $P_c(A, R)$ of A in R, as defined in §7, by intersecting $P_c(A)$ with R, $P_c(A, R) = R \cap P_c(A)$. Indeed, A is Prüfer and convex in $R \cap P_c(A)$, hence $R \cap P_c(A) \subset P_c(A, R)$, and A is also Prüfer and convex in $P_c(A, R)$, hence $P_c(A, R) \subset R \cap P_c(A)$.

Notice that $P_c(A, R)$ is an ℓ-subring of $Q(A)$, even if R is not. □

We want to find out which ℓ-subrings of $Q(A)$ have the same Prüfer convexity cover as A.

Definition 6. The *convex holomorphy ring* of the f-ring A is the holomorphy ring $\mathrm{Hol}_{A^+}(A)$ of A with respect to its ordering A^+ (cf.§6, Def.1). We denote this subring of A more briefly by $\mathrm{Hol}_c(A)$.

We know by Theorem 6.3 that $\mathrm{Hol}_c(A)$ is the smallest subring of A which is convex in A with respect to the saturation $(A^+)^\wedge$ (cf.§5, Def.2), i.e.

$$\mathrm{Hol}_c(A) = \{f \in A \mid \exists n \in \mathbb{N}: \quad n \pm f \in (A^+)^\wedge\}.$$

$\mathrm{Hol}_c(A)$ is an absolutely convex subring of A, in particular an ℓ-subring of A, and thus an f-ring.

Theorem 10.17. Assume that $\mathrm{Hol}(A)$ is Prüfer in A. {N.B. This is a mild condition, cf. Theorems 2.6, 2.6′.} Let B be a subring of $Q(A)$. The following are equivalent.
(1) B is an ℓ-subring of $Q(A)$ and $P_c(B) = P_c(A)$.
(2) $\mathrm{Hol}_c(A) \subset B \subset P_c(A)$.

Proof. a) Let $R := P_c(A)$ and $H := \mathrm{Hol}_c(A)$. Since $\mathrm{Hol}(A) \subset H \subset A$ and $\mathrm{Hol}(A)$ is assumed to be Prüfer in A, the ring H is Prüfer in A. It is also convex in A. We conclude that H is Prüfer and convex in R.
b) It follows by Theorem 6.7 that H is $(R^+)^\wedge$-convex in R. Thus $\mathrm{Hol}_c(R) \subset H$, and we have inclusions $\mathrm{Hol}_c(R) \subset H \subset A \subset R$. It follows that $\mathrm{Hol}_c(R)$ is Prüfer and convex in A, hence is $(A^+)^\wedge$-convex in A. This implies that $H \subset \mathrm{Hol}_c(R)$, and we conclude that $\mathrm{Hol}_c(R) = H$.
c) Since H is Prüfer and convex in R, we have $R \subset P_c(H)$, hence the inclusions $H \subset A \subset R \subset P_c(H)$. It follows by Remark 16.i that A is convex in $P_c(H)$. The ring A is also Prüfer in $P_c(H)$. This implies $P_c(H) \subset R$, and we conclude that $P_c(H) = R$.
d) If now B is any overring of H in R then we learn by Remark 16.i that B is absolutely convex in R. Thus B is an ℓ-subring of R, hence an ℓ-subring of $Q(A)$. Further we conclude from $H = \mathrm{Hol}_c(R)$ and $R = P_c(H)$ by arguments as in b) and c) that $\mathrm{Hol}_c(B) = H$ and $P_c(B) = R$.
e) Finally, if B is an ℓ-subring of $Q(A)$ with $P_c(B) = R$, then B is a subring of R which is Prüfer and convex in R, hence is $(R^+)^\wedge$-convex in R. It follows that $H \subset B \subset R$. $\square$

§11 The Prüfer hull of $C(X)$

Let X be any topological space, Hausdorff or not, and let $R := C(X)$, the ring of $\mathbb{R}$-valued continuous functions on X. We equip R with the partial ordering $R^+ := \{f \in R \mid f(x) \geq 0 \text{ for every } x \in X\}$. Obviously this makes R an f-ring. We are interested in finding the Prüfer subrings of R and the overrings of R in the complete ring of quotients $Q(R)$, in which R is Prüfer.

In this business we may assume without loss of generality that X is a *Tychonov space*, i.e. a completely regular Hausdorff space, since there exists a natural identifying continuous map $X \twoheadrightarrow X'$ onto such a space X', inducing an isomorphism of f-rings $C(X') \xrightarrow{\sim} C(X)$, cf. [GJ, §3]. But now we still refrain from the assumption that X is Tychonov. This property will become important only later in the section.

Observe that $R^+ = \{f^2 \mid f \in R\}$. Thus R^+ coincides with the smallest preordering T_0 on R. Clearly R^+ is also saturated, $R^+ = (R^+)^\wedge$. Finally $1 + R^+ \subset R^*$, i.e. R has bounded inversion. These three facts make life easier than for f-rings in general.

Since $R^+ = T_0 = \hat{T}_0$, we infer from the definitions that $\mathrm{Hol}(R) = \mathrm{Hol}_c(R)$, further from Theorem 6.3.c that $\mathrm{Hol}(R)$ coincides with the ring $C_b(X)$ of bounded continuous functions on X,

$$\mathrm{Hol}(R) = C_b(X) := \{f \in R \mid \exists\, n \in \mathbb{N}\colon\ |f| \leq n\}.$$

We had proved this by other means before (Ex.4.13).

It is clear already from Theorem 2.6 (or 2.6′) that $\mathrm{Hol}(R)$ is Prüfer in R, and it is plain that $\mathrm{Hol}(R)$ has bounded inversion.

Let $\varphi\colon S \to X$ be a continuous map from some topological space S to X. It induces a ring homomorphism $\rho := C(\varphi)$ from $C(X)$ to $C(S)$, mapping a function $f \in C(X)$ to $f \circ \varphi$. We denote the subring $\rho(C(X))$ of $C(S)$ by $C(X)|_\varphi$ and the subring $\rho(C_b(X))$ of $C_b(S)$ by $C_b(X)|_\varphi$. Since for $f, g \in C(X)$ we have $\rho(f \vee g) = \rho(f) \vee \rho(g)$ and $\rho(f \wedge g) = \rho(f) \wedge \rho(g)$, both $C(X)|_\varphi$ and $C_b(X)|_\varphi$ are ℓ-subrings of the f-ring $C(S)$.

The f-ring $A := C(X)|_\varphi$ inherits many good properties from $R = C(X)$. If $h \in A^+$, we conclude from $h = \rho(f)$ with $f \in R$, that $h = \rho(|f|) = \rho(|f|^{1/2})^2$. Thus A^+ consists of the squares of elements of A. We conclude, as above for R, that

$$\mathrm{Hol}(A) = \mathrm{Hol}_c(A) = \{h \in A \mid \exists\, n \in \mathbb{N}\colon\ |h| \leq n\}.$$

It follows that $\mathrm{Hol}(A) = C_b(X)|_\varphi$. Indeed, if $h = \rho(f)$ and $|h| \leq n$ (in A), then $h = \rho((f \wedge n) \vee (-n))$.

Since $C_b(X)$ is Prüfer in $C(X)$ and ρ maps $R = C(X)$ onto $A = C(X)|_\varphi$ and $C_b(X)$ onto $C_b(X)|_\varphi$, it follows by general principles (Prop.I.5.7) that $C_b(X)|_\varphi$ is Prüfer in $C(X)|_\varphi = A$.

Notice also that for $f \in R$ the element $1+\rho(f)^2 = \rho(1+f^2)$ is a unit of A, since $1+f^2$ is a unit of R. Thus A has bounded inversion. Theorem 2.6 (or 2.6′) tells us that $\mathrm{Hol}(A)$ is Prüfer in A. Clearly $\mathrm{Hol}(A)$ has bounded inversion. In short, A shares all the agreeable properties of R, stated above, although perhaps A is not isomorphic to a ring of continuous functions $C(Y)$.

THEOREM 11.1. *Let $\varphi: S \to X$ be a continuous map. The following are equivalent.*
(1) *$C(X)|_\varphi$ is Prüfer in $C(S)$.*
(2) *$C(X)|_\varphi$ is convex in $C(S)$.*
(3) *$C_b(X)|_\varphi = C_b(S)$.*

PROOF. This is a special case of Theorem 7.6, since both $A := C(X)|_\varphi$ and $B := C(S)$ have bounded inversion and $C_A = C_b(X)|_\varphi$, $C_B = C_b(S)$ in the notation used there. □

Assume now that S is a subspace of the topological space X and φ is the inclusion map $S \hookrightarrow X$. Then we write $C(X)|_S$ and $C_b(X)|_S$ for $C(X)|_\varphi$ and $C_b(X)|_\varphi$ respectively.

DEFINITION 1 [GJ].*) S is called *C_b-embedded* (resp. *C-embedded*) *in* X if for every $h \in C_b(S)$ (resp. $h \in C(S)$) there exists some $f \in C(X)$ with $f|_S = h$.

Notice that, if h is a bounded continuous function on S which can be extended to a continuous function on X, then h can be extended to a *bounded* continuous function on X, (as has been already observed above). Thus S is C_b-embedded in X iff $C_b(X)|_S = C_b(S)$, and, of course, S is C-embedded in X iff $C(X)|_S = C(S)$.

In this terminology Theorem 1 says the following for a subspace S of X:

COROLLARY 11.2. *$C(X)|_S$ is Prüfer in $C(S)$ iff $C(X)|_S$ is convex in $C(S)$ iff S is C_b-embedded in X.* □

We now fix an element f of $C(X)$. Associated to f we have the zero set $Z(f) := \{x \in X \mid f(x) = 0\}$ and the cozero set $coz(f) := \{x \in X \mid f(x) \neq 0\}$. We are looking for relations between the ring $C(cozf)$ and the localisation $C(X)_f = f^{-\infty}C(X)$ of $C(X)$ with respect to f.

The restriction homomorphism $\rho: C(X) \to C(cozf)$ maps f to a unit of $C(cozf)$, hence induces a ring homomorphism

$$\rho_f\colon\ C(X)_f \longrightarrow C(cozf).$$

*) Gillman and Jerison write C^* instead of C_b, as is done in most of the literature on $C(X)$. Our deviation from this labelling has been motivated in 1.3.

We claim that ρ_f is injective. Indeed, let an element $\frac{g}{f^n} \in C(X)_f$ be given $(g \in C(X),\ n \in \mathbb{N}_0)$, and assume that $\rho_f\left(\frac{g}{f^n}\right) = 0$. Then $\rho_f\left(\frac{g}{1}\right) = \rho(g) = g|_{coz f} = 0$. This implies $gf = 0$ and then $\frac{g}{f^n} = \frac{gf}{f^{n+1}} = 0$. Henceforth we regard $C(X)_f$ as a subring of $C(coz f)$ via ρ_f.

Lemma 11.3. $C(X)_f$ contains the subring $C_b(coz f)$ of $C(coz f)$.

Proof. Let $g \in C_b(coz f)$ be given. The function $h: X \to \mathbb{R}$ defined by $h(x) := f(x)g(x)$ for $x \in coz f$, $h(x) = 0$ for $x \in Z(f)$, is continuous, since g is bounded. We have $g = \rho_f\left(\frac{h}{f}\right)$. □

Theorem 11.4. For any $f \in C(X)$ the ring $C(X)_f$ is Bezout and absolutely convex in $C(coz f)$, and $C(X)_f$ has bounded inversion.

Proof. By Lemma 3 we have the inclusions $C_b(coz f) \subset C(X)_f \subset C(coz f)$. We know that $C_b(coz f)$ is Prüfer and convex in $C(coz f)$. Also both rings have bounded inversion. It follows that the extension $C_b(X) \subset C(X)_f$ is Prüfer, then by Theorem 9.15, that $C(X)_f$ has bounded inversion. Also the extension $C(X)_f \subset C(X)$ is Prüfer. We conclude by Theorem 9.15, that $C(X)_f$ is Bezout and absolutely convex in $C(X)$. □

We recall some facts about Bezout extensions from II, §10.

Definition 2 (cf.II §10, Def.6). If A is any ring, an element f of A is called a *Bezout element* of A if f is a non-zero-divisor of A and the extension $A \subset A_f$ is Bezout. The set of all Bezout elements of A is denoted by $\beta(A)$.

As has been observed in II §10, $\beta(A)$ is a saturated multiplicative subset of A. It is also clear from II §10, that for any multiplicative subset S of $\beta(A)$ the extension $A \subset S^{-1}A$ is Bezout (cf.Prop.II.10.13).Conversely any Bezout extension R of A has the shape $R = S^{-1}A$ with $S = A \cap R^*$ (cf.Prop.II.10.16).*) Thus the Bezout extensions of A in $Q(A)$ correspond uniquely with the saturated multiplicative subsets of $\beta(A)$. In particular, $\beta(A)$ itself gives us the Bezout hull $\mathrm{Bez}(A) = \beta(A)^{-1}A$ of A.

Theorem 11.5. i) Every Prüfer extension of $C(X)$ is Bezout.
ii) The Bezout elements of $C(X)$ are the non-zero-divisors f of $C(X)$ with the property that $coz(f)$ is C_b-embedded in X.

Proof. i): We know by Theorem 10.12 that every Prüfer extension $C(X) \subset R$ is an f-extension in a natural way. Since $C(X)$ has bounded inversion we read off from Theorem 9.15 that R is Bezout over $C(X)$.

---

*) Prop.II.10.16 contains a typographical error. Read "If $A \subset R$ is a Bezout extension" instead of "If A is a Bezout extension".

ii): Let f be a non-zero-divisor of $C(X)$. Then $C(X)$ embeds into $C(X)_f$. Thus we have ring extensions $C(X) \subset C(X)_f \subset C(coz f)$. We know by Theorem 4 that $C(X)_f$ is Bezout in $C(coz f)$. Thus $C(X)$ is Bezout in $C(X)_f$, i.e. f is a Bezout element, iff $C(X)$ is Bezout in $C(coz f)$ (Recall II.10.15.iii). Corollary 2 above tells us that this happens iff $coz(f)$ is C_b-embedded in X. □

Notations. We denote the set of Bezout elements $\beta(C(X))$ more briefly by $b(X)$. We further denote the set of all open subsets $coz(f)$ of X with f running through $b(X)$ by $\mathfrak{B}(X)$.

Notice that $\mathfrak{B}(X)$ is closed under finite intersections, since $coz(f_1) \cap coz(f_2) = coz(f_1 f_2)$. We have a direct system of ring extensions $(C(U) \mid U \in \mathfrak{B}(X))$ of $C(X)$. Here the index set $\mathfrak{B}(X)$ is ordered by reverse inclusion ($U \leq V$ iff $V \subset U$), and the transition maps $C(U) \to C(V)$ are the restriction homomorphisms $f \mapsto f|_V \quad (U \supset V)$. $\mathfrak{B}(X)$ has a first element $U = X = coz(1)$.

Theorems 4 and 5 lead to the following description of the Prüfer hull of $C(X)$.

COROLLARY 11.6. All transition maps in the system $(C(U) \ |U \in \mathfrak{B})$ are injective, and

$$P(C(X)) = \varinjlim_{U \in \mathfrak{B}(X)} C(U).$$

PROOF. Each ring $C(U)$ with $U \in \mathfrak{B}(X)$ is Prüfer over $C(X)$, hence embeds into the Prüfer hull $P(C(X))$ of $C(X)$ in a unique way, which (hence) is compatible with the transition maps. It follows that all transition maps are injective. Identifying the rings $C(U)$ with their images in $P(C(X))$ we may now write

$$\varinjlim_{U \in \mathfrak{B}(X)} C(U) = \bigcup_{U \in \mathfrak{B}(X)} C(U) = \bigcup_{f \in b(X)} C(coz f). \tag{1}$$

Denoting this ring by D we have $C(X) \subset D \subset P(C(X))$. It follows that D is Prüfer over $C(X)$. {We could also have invoked I.5.14.} On the other hand, every localization $C(X)_f$, with f running through $b(X)$, can be embedded in $P(C(X))$ in a unique way over $C(X)$. Since $P(C(X))$ coincides with the Bezout hull of $C(X)$, we have

$$\bigcup_{f \in b(X)} C(X)_f \ = \ P(C(X)). \tag{2}$$

We infer from (1), (2) and $C(X)_f \subset C(coz f) \subset D$ for every $f \in b(X)$, that $D = P(C(X))$. □

Starting from now *we assume that X is a Tychonov space.* Now a function $f \in C(X)$ is a non-zero-divisor in $C(X)$ iff $coz(f)$ is dense in X. {Just observe that, if a point $p \in X \setminus \overline{coz(f)}$ is given, there exists a function $g \in C(X)$ with $g|coz(f) = 0$ and $g(p) \neq 0$. Then $fg = 0$.} Thus $\mathfrak{B}(X)$ is the set of all cozero sets U in X which are dense and C_b-embedded in X.

Let $\mathcal{D}(X)$ denote the set of all dense open subsets of X, and let $\mathcal{D}_0(X)$ denote the set of all dense cozero subsets of X. Then

$$\mathfrak{B}(X) \subset \mathcal{D}_0(X) \subset \mathcal{D}(X),$$

and these three families are all closed under finite intersections. As above we have direct systems of f-rings $\{C(U) \mid U \in \mathcal{D}(X)\}$ and $\{C(U) \mid U \in \mathcal{D}_0(X)\}$ with injective transition maps.

We introduce the ring

$$Q(X) := \varinjlim_{U \in \mathcal{D}(X)} C(U),$$

which again is an f-ring in the obvious way. Every $C(U), U \in \mathcal{D}(X)$ injects into $Q(X)$ and will be regarded as a subring of $Q(X)$. We have $C(X) \subset C(U) \subset Q(X)$ for every $U \in \mathcal{D}(X)$ and

$$Q(X) = \bigcup_{U \in \mathcal{D}(X)} C(U).$$

The following has been proved by Fine, Gillman and Lambek a long time ago.

Theorem 11.7 [FGL]. $C(X)$ has the complete ring of quotients $Q(X)$ and the total ring of quotients

$$\mathrm{Quot}(C(X)) = \varinjlim_{U \in \mathcal{D}_0(X)} C(U) = \bigcup_{U \in \mathcal{D}_0(X)} C(U). \qquad \square$$

Henceforth we work in the overring $Q(C(X)) = Q(X)$ of $C(X)$. We think of the elements of $Q(X)$ as continuous functions defined on dense open subsets of X. Two such functions $g_1 : U_1 \to \mathbb{R}$, $g_2 : U_2 \to \mathbb{R}$ are identified if there exists a dense open set $V \subset U_1 \cap U_2$ with $g_1|V = g_2|V$. Of course, then g_1 and g_2 coincide on $U_1 \cap U_2$. Corollary 6 now reads as follows.

Scholium 11.8. A continuous function $g : U \to \mathbb{R}$ with U open and dense in X is an element of the Prüfer hull $P(C(X))$ iff there exists some $f \in C(X)$ such that $coz(f) \subset U$ and $coz(f)$ is dense and C_b-embedded in X. $\square$

Remark 11.9. Along the way we have proved that, if U_1, U_2 are dense cozero sets in $C(X)$, which both are C_b-embedded in X, then $U_1 \cap U_2$ is again C_b-embedded in X. In fact more generally the following holds: If U is an open subset of X, which is C_b-embedded in X, and T is a subspace of X, such that $U \cap T$ is dense in T, then $U \cap T$ is C_b-embedded in T, cf.[GJ, 9N]. □

Already from the coincidence $P(C(X)) = \mathrm{Bez}C(X)$ (Theorem 5), we know that $P(C(X))$ is contained in $\mathrm{Quot}C(X)$. Thus we have inclusions

$$C(X) \subset P(C(X)) \subset \mathrm{Quot}C(X) \subset Q(X) = Q(C(X)).$$

We now ask for cases where $P(C(X))$ is equal to one of the other three rings. Part a) of the following theorem is due to Martinez [Mart], while Part b) is due to Dashiell, Hager and Henriksen [DHH], cf. the comments below.

Theorem 11.10. i) $C(X)$ is Prüfer in its complete ring of quotients $Q(X)$ iff every dense open subset of X is C_b-embedded in X.
ii) $C(X)$ is Prüfer in $\mathrm{Quot}C(X)$ iff every dense cozero subset of X is C_b-embedded in X.

Proof. a) If $\mathfrak{B}(X) = \mathcal{D}(X)$, resp. $\mathfrak{B}(X) = \mathcal{D}_0(X)$, we know by Corollary 6 and Theorem 7 that $P(C(X)) = Q(X)$, resp. $P(C(X)) \supset \mathrm{Quot}C(X)$.
b) Assume that $C(X)$ is Prüfer in $Q(X)$. Let U be a dense open subset of X. Since $C(X) \subset C(U) \subset Q(X)$, we conclude that $C(X)$ is Prüfer in $C(U)$. Now Theorem 1, more precisely Corollary 2, tells us that U is C_b-embedded in X.
c) Assume that $C(X)$ is Prüfer in $\mathrm{Quot}C(X)$. Let f be a non-zero-divisor of $C(X)$. Since $C(X) \subset C(X)_f \subset \mathrm{Quot}C(X)$, we conclude that $C(X)$ is Prüfer, hence Bezout in $C(X)_f$, i.e. f is a Bezout element of $C(X)$. Theorem 5 tells us that $coz(f)$ is C_b-embedded in $C(X)$. □

Comments 11.11.
a) X is called *extremally disconnected* [GJ, 1H] if every open subset of X has an open closure. It is well known that this is equivalent to the property that every open subset of X is C_b-embedded in X ([GJ, 1H.6], [PW, 6.2]). Now, if all *dense* open subsets of X are C_b-embedded in X, then this is true for *all* open subsets of X. Indeed, if U is open in X and $f \in C_b(U)$, then f can be extended by zero to a bounded continuous function on the dense open set $U \cup (X \setminus \overline{U})$ of X, and this function extends to a bounded continuous function on X. Thus Theorem 11.10.a can be coined as follows: *$C(X)$ is Prüfer in $Q(C(X))$ iff X is extremally disconnected*. {[Mart, Th.2.7]; Martinez there calls a ring A which is Prüfer in $Q(A)$,*) an "I-ring" following the terminology of Eggert [Eg]}.

Extremally disconnected spaces are rare but not out of the world. For example, the Stone-Čech compactification βD of any discrete space D is extremally disconnected [PW, 6.2]. There also exist extremally disconnected spaces without isolated points, cf. [PW, 6.3].

*) more precisely, a ring A such that every overring in $Q(A)$ is integrally closed in $Q(A)$, but this means the same (Th.I.5.2).

b) A Tychonov space X is an F*-space*, if every cozero-set of X is C_b-embedded in X ([GJ, 14.25]), while X is called a *quasi-*F*-space*, if every *dense* cozero-set of X is C_b-embedded in X [DHH], which is a truly weaker condition. Thus Theorem 10.b can be coined as follows: $C(X)$ *is Prüfer in Quot* $C(X)$ *iff* X *is a quasi-*F*-space* {[DHH; A ring A which is Prüfer in Quot$A$ is traditionally called a "Prüfer ring with zero divisors" [Huc]}.

Using Theorem 9.15 we may rephrase this result as follows: $C(X)$ *is convex in Quot* $C(X)$ *iff* X *is a quasi-*F*-space.* In this way Theorem 10.b has been stated and proved by Schwartz [Sch_3, Th.6.2].

F-spaces, hence quasi-F-spaces, are not so rare. Prominent examples are the spaces $\beta Y \setminus Y$ with Y locally compact and σ-compact [GJ, 14.27]. □

Concerning the case $C(X) = P(C(X))$, i.e. Prüfer closedness of $C(X)$, we have only a partial result.

Theorem 11.12. If X is a metric space then $C(X)$ is Prüfer closed.

Proof. Suppose $C(X)$ is not Prüfer closed. Then $C(X)$ has a Bezout element f which is not a unit (cf.Theorem 5.a), and this means that the set $U := cozf$ is C_b-embedded and dense in X, but $U \neq X$ (cf.Theorem 5.b). We choose a point $p \in X \setminus U$ and then a sequence $\{x_n \mid n \in \mathbb{N}\}$ in U, consisting of pairwise different points and converging to p. The sets $Z_0 := \{x_{2n} \mid n \in \mathbb{N}\}$ and $Z_1 := \{x_{2n-1} \mid n \in \mathbb{N}\}$ are closed in U and disjoint. Let f_0 and f_1 denote the distance functions $\mathrm{dist}(-, Z_0)$ and $\mathrm{dist}(-, Z_1)$ on the metric space U. The function
$$g := \frac{|f_0|}{|f_0| + |f_1|}$$
on U is well defined, bounded and continuous. We have $g|_{Z_0} = 0$ and $g|_{Z_1} = 1$. Thus g cannot be extended continuously to $U \cup \{p\}$. This is a contradiction and proves that $C(X) = P(C(X))$. □

We mention that Schwartz has developed general criteria for $C(X)$ to be Prüfer closed, cf.[Sch_3, Th.5.3]. He also gave a description of the Prüfer hull $P(C(X))$ in general, different from our Theorem 5, by use of the real spectrum of $C(X)$, cf.[Sch_3, Th.5.5].

§12 VALUATIONS ON F-RINGS

It is somewhat remarkable that in §9 and §10 we nowhere used valuations (explicitly) for gaining results about Prüfer subrings or Prüfer extensions of a given f-ring R. But, of course, in order to complete the picture, a thorough study of valuations on R is appropriate. We will experience a relation between the convex valuations on R and the prime cones $P \supset R^+$ even closer than in the general theory in §3 and §5.

In the following R is an f-ring and $v: R \to \Gamma \cup \infty$ is a valuation on R. For any $\gamma \in \Gamma \cup \infty$ we introduce the A_v-module

$$I_{\gamma,v} := \{x \in R \mid v(x) \geq \gamma\}. \quad ^{*)}$$

PROPOSITION 12.1.
a) For every $x \in R$

$$v(x) = v(|x|) = \min(v(x^+), v(x^-)),$$

and either $v(x^+) = \infty$ or $v(x^-) = \infty$.
b) For every $\gamma \in \Gamma \cup \infty$ the set $I_{\gamma,v}$ is a sublattice of R.

PROOF. a): It follows from $x^+x^- = 0$ that either $v(x^+) = \infty$ or $v(x^-) = \infty$, and then from $x = x^+ - x^-$, $|x| = x^+ + x^-$, that $v(x) = v(|x|) = \min(v(x^+), v(x^-))$.
b) It is now clear that, for every $x \in I_{\gamma,v}$, also $x^+ \in I_{\gamma,v}$ (and $x^- \in I_{\gamma,v}$). If $x, y \in I_{\gamma,v}$ are given, we conclude that

$$x \vee y = y + [(x - y) \vee 0] = y + (x - y)^+ \in I_{\gamma,v}.$$

Also $x \wedge y = -[(-x) \vee (-y)] \in I_{\gamma,v}$. Thus $I_{\gamma,v}$ is a sublattice of R. □

In the special case that v is trivial Proposition 1 reads as follows.

COROLLARY 12.2. Every prime ideal of R is a sublattice of R. □

Here is another consequence of Proposition 1.

COROLLARY 12.3. If A is a Prüfer subring of R, every R-regular A-submodule of R is a sublattice of R.

PROOF. Let I be such a submodule of R. We may assume that I is finitely generated. I is the intersection of the R-regular $A_{[\mathfrak{p}]}$-submodules $I_{[\mathfrak{p}]}$ of R with $\mathfrak{p}$ running through the set $\Omega(R/A)$ of maximal R-regular ideals $\mathfrak{p}$ of A

*) as in Chapter III, but now allowing $\gamma \notin v(R)$ and $\gamma = \infty$. Of course, $I_{\infty,v} = \operatorname{supp} v$.

(Prop.III.1.10). To each $\mathfrak{p}$ there corresponds a non-trivial PM-valuation $v_\mathfrak{p}$ of R over A with $A_{v_\mathfrak{p}} = A_{[\mathfrak{p}]}$, and $I_{[\mathfrak{p}]}$ is a $v_\mathfrak{p}$-convex $A_{v_\mathfrak{p}}$-submodule of R (cf.Th.III.2.2). It follows from Proposition 1 that $I_{[\mathfrak{p}]}$ is a sublattice of R. Thus I is a sublattice of R. □

We return to our fixed valuation $v\colon R \to \Gamma \cup \infty$ on R.

Proposition 12.4. For any $x, y \in R$ the set of values $\{v(x \vee y), v(x \wedge y)\}$ coincides with $\{v(x), v(y)\}$.

Proof. Let $x, y \in R$ be fixed. Without loss of generality we assume that $\gamma := v(x) \leq v(y)$. Since $I_{\gamma,v}$ is a sublattice of R, we have $\gamma \leq v(x \vee y)$ and $\gamma \leq v(x \wedge y)$.

If $\gamma = \infty$ we have $v(y) = v(x \vee y) = v(x \wedge y) = \infty$, and we are done. We now assume that $\gamma \in \Gamma$. We use the identities, stated in §9,
(F7) $\quad x + y = (x \vee y) + (x \wedge y)$,
(F8) $\quad xy = (x \vee y)(x \wedge y)$.
By F8 we have

$$(*) \qquad \gamma + v(y) \;=\; v(x \vee y) + v(x \wedge y).$$

Also, as said above, $v(x \vee y) \geq \gamma$, $v(x \wedge y) \geq \gamma$. If $v(y) = \gamma$ this forces $v(x \wedge y) = v(x \vee y) = \gamma$, and we are done in this case.

There remains the case that $v(y) > \gamma$. Now $v(x + y) = \gamma$. By (F7) we have $\gamma \geq \min(v(x \vee y), v(x \wedge y))$. Since $v(x \vee y) \geq \gamma$ and $v(x \wedge y) \geq \gamma$, this forces $\gamma = \min(v(x \vee y), v(x \wedge y))$. Now $(*)$ tells us – also in the case $v(y) = \infty$ – that $v(y) = \max(v(x \vee y), v(x \wedge y))$. □

As a consequence of the proposition we have

Corollary 12.5. For any subset M of Γ the set $\{x \in R \mid v(x) \in M\}$ is either empty or a sublattice of R. In particular, A_v is an ℓ-subring of R, hence an f-ring, and both $\mathfrak{p}_v$ and $A_v \setminus \mathfrak{p}_v$ are sublattices of A_v. □

Proposition 12.6. The following are equivalent.
(1) v is convex.
(2) $v(x \vee y) = \min(v(x), v(y))$ for all $x, y \in R^+$.
(3) $v(x \wedge y) = \max(v(x), v(y))$ for all $x, y \in R^+$.

Proof. The equivalence (2) ⇔ (3) is clear from Proposition 4.
(1) ⇒ (2): Since v is convex it follows from $0 \leq x \leq x \vee y$ and $0 \leq y \leq x \vee y$ that $v(x) \geq v(x \vee y)$, $v(y) \geq v(x \vee y)$, hence $\min(v(x), v(y)) \geq v(x \vee y)$. Again invoking Proposition 4 we obtain equality here.

(2) $\Rightarrow$ (1): If $x, y \in R$ and $0 \leq y \leq x$ we have $x = x \vee y$, hence $v(x) = \min(v(x), v(y))$ by (2), i.e. $v(x) \leq v(y)$. Thus v is convex. $\square$

Remark. In the vein of Corollary 2 we obtain from Proposition 6 that, for A a convex Prüfer subring of R, every R-regular A-submodule of R is absolutely convex in R. But this we already proved in §9 in another way, cf.Theorem 9.10. $\square$

THEOREM 12.7. Let $\mathfrak{q}$ be a convex prime ideal of R.
a) Then $P := R^+ + \mathfrak{q}$ is a prime cone of R, and $P = \{x \in R \mid x^- \in \mathfrak{q}\}$.
b) P is the unique prime cone of R containing R^+ and with support $\mathfrak{q}$.

PROOF. 1) We know by Lemma 5.9 that $P := R^+ + \mathfrak{q}$ is a preordering of R and $P \cap (-P) = \mathfrak{q}$.
2) We verify that $P = \{x \in R \mid x^- \in \mathfrak{q}\}$. Let $x \in R$ be given. If $x^- \in \mathfrak{q}$, then $x = x^+ - x^- \in R^+ + \mathfrak{q} = P$. Assume now that $x \in P$. Write $x = y + z$ with $y \geq 0$ and $z \in \mathfrak{q}$. By Corollary 2 above we know that $\mathfrak{q}$ is a sublattice of R. Thus $z^- \in \mathfrak{q}$. It follows from $x = (y + z^+) - z^-$ that $0 \leq x^- \leq z^-$. Since $\mathfrak{q}$ is convex we conclude that $x^- \in \mathfrak{q}$.
3) Let $x \in R$ be given with $x \notin P$. Then $x^- \notin \mathfrak{q}$. But $x = x^+ x^- = 0 \in \mathfrak{q}$. Thus $(-x)^- = x^+ \in \mathfrak{q}$, hence $-x \in P$. This proves that $P \cup (-P) = R$. We now know that P is a prime cone of R with support $\mathfrak{q}$.
4) If P' is any prime cone of R with $P' \supset R^+$ and $\operatorname{supp} P' = \mathfrak{q}$, then $P' \supset R^+ + \mathfrak{q} = P$. Since P' and P have the same support, it follows that $P' = P$ (cf.Th.4.6). $\square$

Comment. We know for long that, if T is a proper preordering of any ring R and $\mathfrak{q}$ a T-convex prime ideal of R, there exists a prime cone $P \supset T$ with support $\mathfrak{q}$ (cf.Th.5.6 and Th.4.6). Theorem 7 states the remarkable fact that P is *unique* in the present case, where R is an f-ring and $T = R^+$. This means that we have a bijection $\mathfrak{q} \mapsto T + \mathfrak{q}$ from the set $\operatorname{Spec}_T(R)$ of all T-convex prime ideals to the set $\operatorname{Sper}_T(R)$ of prime cones $P \supset T$ of R, the inverse map being the restriction $\operatorname{Sper}_T(R) \to \operatorname{Spec}_T(R)$ of the support map $\operatorname{supp} : \operatorname{Sper}(R) \to \operatorname{Spec}(R)$.

One should view $\operatorname{Sper}_T(R)$ and $\operatorname{Spec}_T(R)$ as the real spectrum and the Zariski spectrum of the ordered ring (R, T). In the case that R is an f-ring and $T = R^+$ we leave it to the reader to verify, that our bijection $\operatorname{Sper}_T(R) \to \operatorname{Spec}_T(R)$ is a homeomorphism with respect to the subspace topologies in $\operatorname{Sper}(R)$ and $\operatorname{Spec}(R)$. $\square$

THEOREM 12.8. Let U be a preordering of R containing R^+ and v a U-convex valuation on R. Then there exists a *unique* prime cone P on R such that $U \subset P$, v is P-convex, $\operatorname{supp} P = \operatorname{supp} v$. We have $P = R^+ + \operatorname{supp} v = U + \operatorname{supp} v = \{x \in R \mid v(x^-) = \infty\}$.

PROOF. 1) Let $\mathfrak{q} := \operatorname{supp} v$. This prime ideal is U-convex, hence R^+-convex. We *define* $P := R^+ + \mathfrak{q}$. We know by Theorem 7 that P is a prime cone of R

with support $\mathfrak{q}$, and that P is the only candidate for a prime cone with the properties listed in Theorem 8.
2) We prove that v is P-convex. Given $x, y \in P$ it suffices to verify that $v(x+y) = \min(v(x), v(y))$, (cf. Remark 5.10.i). We have $x \equiv x^+ \mod \mathfrak{q}$, $y \equiv y^+ \mod \mathfrak{q}$, $x+y \equiv x^+ + y^+ \mod \mathfrak{q}$, hence $v(x) = v(x^+)$, $v(y) = v(y^+)$, $v(x+y) = v(x^+ + y^+)$. Since v is R^+-convex, we have $v(x^+ + y^+) = \min(v(x^+), v(y^+))$, and we conclude that indeed $v(x+y) = \min(v(x), v(y))$.
3) By Theorem 5.16 there exists a prime cone $P' \supset U$ such that v is P'-convex and $\operatorname{supp} P' = \mathfrak{q}$. The ideal $\mathfrak{q}$ then is P'-convex. By Theorem 7 this forces $P' = R^+ + \mathfrak{q} = P$. Since $R^+ \subset U \subset P'$, it follows that $P = U + \mathfrak{q}$. Since $P = R^+ + \mathfrak{q}$, we know by Theorem 7 that $P = \{x \in R \mid x^- \in \mathfrak{q}\}$. □

Definition 1. If v is a convex (i.e. R^+-convex) valuation on R, we denote the unique prime cone $P \supset R^+$ such that v is P-convex and $\operatorname{supp} v = \operatorname{supp} P$ by P_v, and we call P_v the *convexity prime cone of* v.

Theorem 8 tells us that P_v is the unique *maximal preordering* U of R such that $R^+ \subset U$ and v is U-convex.

Definition 2. For v is a convex valuation on R let $v^\#$ denote the valuation v_P given by the prime cone $P := P_v$.*)

Remarks 12.9. The valuation $v^\#$ is P-convex, hence convex. We have $A_{v^\#} = A_P$ (cf.§3), further $\operatorname{supp} v^\# = \operatorname{supp} P = \operatorname{supp} v$, and $P_{v^\#} = R^+ + \operatorname{supp}(v^\#) = P$. From $v^\# = v_P$ it follows that $v^\# \le v$ (cf.Th.5.15). Clearly $v^\# = (v^\#)^\#$. □

Lemma 12.10. Assume that v and w are convex valuations on R. The following are equivalent.
(1) $P_v = P_w$,
(1′) $\operatorname{supp} v = \operatorname{supp} w$,
(2) $v^\# \le w$,
(3) $v^\# = w^\#$.

Proof. (1) ⇔ (1′): Clear, since for any convex valuation u on R we have $P_u = R^+ + \operatorname{supp} u$ and $\operatorname{supp} u = \operatorname{supp} P_u$.
(1) ⇒ (3): Clear by Definition 2.
(3) ⇒ (2): Clear since $w^\# \le w$.
(2) ⇒ (1′): We have $\operatorname{supp} v^\# = \operatorname{supp} v$. From $v^\# \le w$ we conclude that $\operatorname{supp} v^\# = \operatorname{supp} w$. □

The lemma leads us to an important result about convex valuations on R.

Definition 3. Given a prime cone P of R with $P \supset R^+$ let $\mathfrak{M}_P$ denote the set of equivalence classes of convex valuations v on R with $P_v = P$. We endow $\mathfrak{M}_P$ with the partial ordering given by the coarsening relation $v \le w$.

*) v_P has been defined in §3.

As always, we do not distinguish seriously between a valuation and its equivalence class, thus speaking of the convex valuations v with $P_v = P$ as elements of $\mathfrak{M}_P$.

Theorem 12.11. Let P be a prime cone of R with $R^+ \subset P$, hence $P = R^+ + \mathfrak{q}$ with $\mathfrak{q} := \operatorname{supp} P$.
i) If v and w are convex valuations on R with $v \leq w$, and if $v \in \mathfrak{M}_P$ or if $w \in \mathfrak{M}_P$, both v and w are elements of $\mathfrak{M}_P$.
ii) $\mathfrak{M}_P$ is the set of all convex valutions v on R with $v^\# = v_P$, and also the set of all valuations v of R with $v_P \leq v$.
iii) $\mathfrak{M}_P$ is totally ordered by the coarsening relation and has a minimal and a maximal element. The minimal element is the valuation v_P. The maximal element is the trivial valuation with support $\mathfrak{q}$.

Proof. i): If $v \leq w$ then $\operatorname{supp} v = \operatorname{supp} w$, hence $P_v = P_w$ by Lemma 10.

ii): Let $u := v_P$. For every $v \in \mathfrak{M}_P$ we have $v^\# = u$ by definition of $v^\#$. Further $\operatorname{supp} u = \operatorname{supp} P$ (cf.§3, Def.3), hence $P_u = R^+ + \operatorname{supp} P = P$. Thus $u \in \mathfrak{M}_P$. If now v is a convex valuation with $v^\# = u$, then $u \leq v$ (cf.Remarks 9), hence by i), or again Remarks 9, $v \in \mathfrak{M}_P$.

Finally, if v is any valuation of R with $u \leq v$, then v is convex since u is convex (cf.Remark 5.10.v), and thus $v \in \mathfrak{M}_P$ by i).

iii): If u' is any valuation on any ring R' the coarsenings of u' correspond uniquely with the convex subgroups of the valuation group of u' (cf.I §1). Thus the coarsenings of u' form a totally ordered set. Clearly u' is the minimal element of this set, and the trivial valution with the same support as u' is the maximal one. □

Later we will also need an "relative" analogue of the valuation $v^\#$ which takes into account a given subring Λ of R. In order to define this analogue we introduce the set

$$\mathfrak{M}_{P,\Lambda} := \{v \in \mathfrak{M}_P \mid \Lambda \subset A_v\}.$$

Here – as before – P is a prime cone of R containing R^+. The set $\mathfrak{M}_{P,\Lambda}$ contains the maximal element of $\mathfrak{M}_P$, hence is certainly not empty.

Proposition 12.12. i) The valuation $w := v_{P,\Lambda}$ introduced in §3, Def.5 is the minimal element of $\mathfrak{M}_{P,\Lambda}$.
ii) $A_w = C(P, R/\Lambda) = A(P, R/\Lambda) = \operatorname{Hol}_P(R/\Lambda)$.
iii) If Λ is an ℓ-subring of R then

$$A_w \; = \; \{x \in R \mid \exists\, \lambda \in \Lambda^+ \colon \; \lambda \pm x \in P\}.$$

Proof. Claims i) and ii) are covered by Theorems 3.10 and 6.2. We have

$$A(P, R/\Lambda) \; = \; \{x \in R \mid \exists\, \lambda \in \Lambda \cap P \colon \; \lambda \pm x \in P\}.$$

If Λ is an ℓ-subring of R, then $\lambda \pm x \in P$ implies $\lambda^+ \pm x \in P$, since $\lambda^+ = \lambda + \lambda^- \in \Lambda^+$ and $\lambda^- \in R^+ \subset P$. Thus

$$A(P, R/\Lambda) \subset \{x \in R \mid \exists\, \lambda \in \Lambda^+ \colon \lambda \pm x \in P\}.$$

The reverse inclusion is trivial. $\square$

Definition 3. Let v be a convex valuation on R and $P := P_v$. Let Λ be a subring of R. We define $v_\Lambda^\# := v_{P,\Lambda}$. $\square$

The following is evident from Theorem 11 and Proposition 12.

Scholium 12.13. Let v and w be convex valuations on R. Then $v_\Lambda^\# = w_\Lambda^\#$ iff $P_v = P_w$ iff either $v \leq w$ or $w \leq v$. If $\Lambda \subset A_v$ then $v_\Lambda^\# \leq v$. If $\Lambda \not\subset A_v$ then $v \leq v_\Lambda^\#$ but $v \not\sim v_\Lambda^\#$. $\square$

§13 Convexity preorderings and holomorphy bases

The results on convex valuations in §12 will give us new insight about the interplay between convex Prüfer subrings of an f-ring R and preorderings $T \supset R^+$ of R. We make strong use of the convexity prime cone P_v of a convex valuation v on R (§12, Def.1) and also of the valuations $v^\#$ and $v^\#_\Lambda$ studied in §12.

In the whole section R *is an* f*-ring* and A *is a convex Prüfer subring of* R.

Theorem 13.1. There exists a unique maximal preordering $U \supset R^+$ of R such that A is U-convex in R. More precisely, $U \supset R^+$, A is U-convex in R, and $U \supset U'$ for every preordering $U' \supset R^+$ of R such that A is U'-convex. We have

$$U = \bigcap_{v \in \omega(R/A)} P_v \quad ,$$

where – as before (§1) – $\omega(R/A)$ denotes the maximal restricted PM-spectrum of R over A (i.e. the set of all maximal non trivial PM-valuations of R over A).

Proof. Recall that A is the intersection of the rings A_v with v running through $\omega(R/A)$. We *define* U as the intersection of prime cones P_v with v running through $\omega(R/A)$. This is a preordering of R containing R^+. Each ring A_v, $v \in \omega(R/A)$, is P_v-convex by definition of P_v, hence is U-convex in R. Thus A is U-convex.

Let now a preordering $U' \supset R^+$ of R be given such that A is U'-convex in R. Theorem 6.7 tells us that, for every $v \in \omega(R/A)$, the ring A_v ($= A_{[\mathfrak{p}]}$ with $\mathfrak{p} = A \cap \mathfrak{p}_v$) is U'-convex in R, hence the valuation v is U'-convex (cf.Th.5.11). It follows by Theorem 12.8 that $U' \subset P_v$. Since this holds for every $v \in \omega(R/A)$, we conclude that $U' \subset U$. □

Definition 1. We denote this preordering U by T^R_A, or T_A for short if R is kept fixed, and we call T_A the *convexity preordering of* A *in* R.

Remarks 13.2. i) If A is PM in R then $T_A = P_v$ with v “the” PM-valuation of R such that $A = A_v$, as is clear by Theorem 5.11.
ii) In the proof of Theorem 1 we could have worked as well with the whole restricted PM-spectrum $S(R/A)$ instead of $\omega(R/A)$. Thus also

$$T_A = \bigcap_{v \in S(R/A)} P_v \quad .$$

iii) In the case $A = R$ the set $S(R/A)$ is empty. We then should read $T_A = R$. This is the only case where the preordering T_A is improper. □

Given any proper subring A of R we denote the conductor of A in R by $\mathfrak{q}_A$, or more precisely by $\mathfrak{q}_A^R$ if necessary. By definition

$$\mathfrak{q}_A = \{x \in R \mid Rx \subset A\},$$

and $\mathfrak{q}_A$ is the largest ideal of R contained in A.

Recall from Chapter I (Prop.I.2.2) that, if v is a non trivial special valuation on R, then $\mathfrak{q}_{A_v} = \operatorname{supp} v$. In the case that v is PM this leads to pleasant relations between T_A and $\mathfrak{q}_A$ if A is Prüfer and convex in R, (which we continue to assume).

Corollary 13.3. i) $\mathfrak{q}_A = \bigcap\limits_{v\in\omega(R/A)} \operatorname{supp} v = \bigcap\limits_{v\in S(R/A)} \operatorname{supp} v$.
ii) $\mathfrak{q}_A$ is a convex ideal of R and $\mathfrak{q}_A = \sqrt{\mathfrak{q}_A}$.
iii) $\operatorname{supp} T_A = \mathfrak{q}_A$.
iv) $T_A = R^+ + \mathfrak{q}_A = \{x \in R \mid x^- \in \mathfrak{q}_A\}$.

Proof. i): This is an immediate consequence of the facts that A is the intersection of the rings A_v, with v running through $\omega(R/A)$ or $S(R/A)$, and that $\mathfrak{q}_{A_v} = \operatorname{supp} v$.
ii): Now clear, since each ideal $\operatorname{supp} v$ is prime and convex in R.
iii): $\operatorname{supp} T_A = T_A \cap (-T_A) = \bigcap\limits_{v\in\omega(R/A)} P_v \;\cap \bigcap\limits_{v\in\omega(R/A)} (-P_v) =$

$$\bigcap_{v\in\omega(R/A)} (P_v \cap -P_v) = \bigcap_{v\in\omega(R/A)} \operatorname{supp} v = \mathfrak{q}_A.$$

iv): For each $v \in \omega(R/A)$ we have $P_v = R_+ + \operatorname{supp} v = \{x \in R \mid x^- \in \operatorname{supp} v\}$. Intersecting the P_v we obtain $T_A = R_+ + \mathfrak{q}_A = \{x \in R \mid x^- \in \mathfrak{q}_A\}$. □

Example 13.4. Let X be a topological space, $R := C(X)$ and $A := C_b(X)$. Assume that X is not pseudocompact, i.e. $A \neq R$. We choose on R the partial ordering $R^+ := \{f \in R \mid f(x) \geq 0$ for every $x \in X\}$. Then R is an f-ring and A is an absolutely convex ℓ-subring of R. We know for long that A is Prüfer in R (even Bezout). By the corollary we have $T_A = R^+ + \mathfrak{q}_A$. It is clear that $\mathfrak{q}_A$ contains the ideal $C_c(X)$ of R consisting of all $f \in C(X)$ with compact support. If the space X is both locally compact and σ-compact (e.g. $X = \mathbb{R}^n$ for some n), then it is just an exercise to prove that $\mathfrak{q}_A = C_c(X)$. Thus in this case T_A is the set of all $f \in R$ such that $\{x \in X \mid f(x) < 0\}$ has a compact closure. □

We return to an arbitrary f-ring R and a convex Prüfer subring A of R.

Given an R-overring B of A in R we know that B is a sublattice of R, hence again an f-ring, since B is Prüfer in R (Cor.12.3). We state relations between T_A^R and T_A^B and, in case that B is also convex in R, between T_A^R and T_B^R.

Proposition 13.5. Let B be an overring of A in R.

i) $B \cap T_A^R \subset T_A^B$, $B \cap \mathfrak{q}_A^R \subset \mathfrak{q}_A^B$.
ii) $\mathfrak{q}_A^R \subset \mathfrak{q}_B^R$.
iii) If B is convex in R, then $T_A^R \subset T_B^R$ and $B \cap T_A^R \subset T_A^B \cap T_B^R$.

Proof. i): A is T_A^R-convex in R, hence $(B \cap T_A^R)$-convex in B. This implies $B \cap T_A^R \subset T_A^B$. Taking supports of these preorderings we obtain $B \cap \mathfrak{q}_A^R \subset \mathfrak{q}_A^B$. (By the way this trivially holds for any sequence of ring extensions $A \subset B \subset R$.)
ii): A trivial consequence of the definition of conductors.
iii): Assume now that B is convex in R. We obtain from ii) that

$$T_A^R = R^+ + \mathfrak{q}_A^R \subset R^+ + \mathfrak{q}_B^R = T_B^R.$$

It follows that $B \cap T_A^R \subset B \cap T_B^R \subset T_B^R$. By i) we have $B \cap T_A^R \subset T_A^B$. We conclude that $B \cap T_A^R \subset T_A^B \cap T_B^R$. □

Remark 13.6. If B is an overring of A in R which is convex in R, and U is a preordering of R with $U \supset R^+$, and A is U-convex, then it follows from $T_A^R \subset T_B^R$ that B is U-convex. Acutally we know more: If U is *any* preordering of R such that A is U-convex, then also B is U-convex. This holds by Theorem 8.7, cf. there (i) $\Rightarrow$ (iv). Indeed, since A is absolutely convex in R, A is 2-saturated in R, so the theorem applies. We could have used this fact in the proof of Proposition 4. □

Definition 2. We denote the holomorphy ring $\mathrm{Hol}_{T_A}(R)$ of the preordering T_A in R (cf.§6, Def.1) by H_A, more precisely by H_A^R if necessary. We call H_A the *holomorphy base of* A (*in* R). {Recall that we assume A to be Prüfer and convex in R.}

Since the preordering T_A is clearly saturated, we know by Theorem 6.3.c that H_A is the smallest T_A-convex subring of R,

$$H_A = C(T_A, R) = A(T_A, R).$$

In particular, $H_A \subset A$. By definition, H_A is the intersection of the rings A_v with v running through all T_A-convex valuations of R, hence H_A is a sublattice of R. It follows that H_A is absolutely convex in R.

We will often need the assumption that H_A is Prüfer in R. This certainly holds if the absolute holomorphy ring $\mathrm{Hol}(R)$ is Prüfer in R, since $\mathrm{Hol}(R) \subset H_A$. Thus it holds for example if R has positive definite inversion (Th.2.6) or if for every $x \in R$ there exists some $d \in \mathbb{N}$ with $1 + x^{2d} \in R^*$ (Th.2.6′).

Proposition 13.7. Assume that H_A is Prüfer in R.
i) Then T_A is also the convexity preordering of H_A.
ii) If also B is a convex Prüfer subring of R the following are equivalent.
(1) $T_A \subset T_B$,
(2) $\mathfrak{q}_A \subset \mathfrak{q}_B$,

(3) $H_B \supset H_A$,
(4) $B \supset H_A$.

Proof. i): H_A is T_A-convex in R. Thus $T_A \subset T_{H_A}$. Since $H_A \subset A$ we also have $T_{H_A} \subset T_A$ (Prop.4.iii). Thus $T_A = T_{H_A}$.
ii): (1) $\Rightarrow$ (2): Clear, since $\mathfrak{q}_A = \operatorname{supp} T_A$ and $\mathfrak{q}_B = \operatorname{supp} T_B$.
(1) $\Rightarrow$ (3): B is T_A-convex by assumption. Thus $H_B \supset H_A$.
(3) $\Rightarrow$ (4): Trivial, since $B \supset H_B$.
(4) $\Rightarrow$ (1): By Proposition 4 and i) above we have $T_B \supset T_{H_A} = T_A$. $\square$

Remark. In (ii) the implications (1) $\Leftrightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) hold under the sole assumption that both A and B are convex and Prüfer in R. {(2) $\Rightarrow$ (1) is clear, since $T_A = R^+ + \mathfrak{q}_A$ and $T_B = R^+ + \mathfrak{q}_B$.} But for (4) $\Rightarrow$ (1) we need to know that H_A is Prüfer in R. $\square$

Corollary 13.8. We assume as before that H_A is Prüfer in R. Let C be a subring of A which is convex and Prüfer in A, hence in R. Then $T_C = T_A$ iff $H_A \subset C$. In this case $H_C = H_A$.

Proof. If $T_C = T_A$ then $H_C = H_A$ by definition of H_A and H_C. Hence $H_A \subset H_C$. {For this implication we do not need that H_A is Prüfer in R.}

Assume now that $H_A \subset C$. Proposition 7 tells us that $T_A \subset T_C$. On the other hand $T_C \subset T_A$ since $C \subset A$. Thus $T_A = T_C$. $\square$

In order to understand the amount of convexity carried by subrings of R it is helpful to have also "relative holomorphy bases" at ones disposal, to be defined now. As before we assume that A is a convex Prüfer subring of R.

Definition 3. Let Λ be any subring of A. The *holomorphy base $H_{A/\Lambda}$ of A over Λ (in R)* is the holomorphy ring of R over Λ of the preordering T_A,

$$H_{A/\Lambda} \colon = H^R_{A/\Lambda} \colon = \operatorname{Hol}_{T_A}(R/\Lambda). \qquad \square$$

Remarks 13.9.
i) $\operatorname{Hol}(R) \subset H_A = H_{A/\mathbb{Z}\cdot 1_R} \subset H_{A/\Lambda} \subset A$.
ii) As in the case $\Lambda = \mathbb{Z}\cdot 1_R$ we have $H_{A/\Lambda} = C(T_A, R/\Lambda) = A(T_A, R/\Lambda)$, again by Theorem 6.3.c.
iii) Assume that H_A is Prüfer in R. Then $H_{A/\Lambda} = \Lambda \cdot H_A$, as follows from Remark 7.1.iii.
iv) If $H_{A/\Lambda}$ is Prüfer in R, all statements in Proposition 7 remain true if we replace H_A and H_B there by $H_{A/\Lambda}$, $H_{B/\Lambda}$, of course assuming that Λ is a subring of both A and B. We thus also have an obvious analogue of Corollary 8 for relative holomorphy bases. $\square$

Comment. It is already here that we can see an advantage to deal with relative instead of just "absolute" holomorphy bases. If A and B are overrings of Λ in

R then we have a result as Proposition 7 under the hypothesis that $H_{A/\Lambda}$ is Prüfer in R instead of the stronger hypothesis that H_A is Prüfer in R.

Below we will study relations between the restricted PM-spectra $S(R/A)$ and $S(R/B)$ in the case that $A \subset B$ and $T_A = T_B$. For many arguments it will again suffice to assume that $H_{A/\Lambda}$ $(= H_{B/\Lambda})$ is Prüfer in R. Without invoking relative holomorphy bases we would have to assume that H_A is Prüfer in R. $\square$

Assume – as before – that A is a convex Prüfer subring of R and $\Lambda \subset A$. Let $H := H_{A/\Lambda}$. Striving for a better understanding of holomorphy bases we look for relations between the PM-valuations of R over A and over H.

Proposition 13.10. Assume that v is a non trivial Manis valuation of R over A, i.e. $v \in S(R/A)$.
a) Then $H \subset A_{v^\#_\Lambda}$.
b) Assume in addition that H is Prüfer in A. {N.B. This holds if $\mathrm{Hol}(A)$ is Prüfer in A.} Then $v^\#_\Lambda$ is a maximal PM-valuation over H, i.e. $v^\#_\Lambda \in \omega(R/H)$.

Proof. a): Let $P := P_v$ and $v' := v^\#_\Lambda$. The valuation v is T_A-convex, since A is T_A-convex in R. Thus $T_A \subset P$. {Actually we know that $T_A = \bigcap\limits_{u \in S(R/A)} P_u$.}
The valuation v' is P-convex, hence again T_A-convex. Thus $A_{v'}$ is T_A-convex in R. This implies $H \subset A_{v'}$.
b): Let $u := v'|_R$, i.e. u is the special valuation $v'|c_{v'}(\Gamma)$ associated with $v': R \to \Gamma \cup \infty$ (cf.I,§1). We have $A_u = A_{v'} \supset H$, and we conclude that u is a PM-valuation of R over H. From $v' \le v$ we infer that $A_u \subset A_v$. Since both u and v are PM and v is not trivial, it follows that $u \le v$, and then, that $\mathrm{supp}\, u = \mathrm{supp}\, v = \mathrm{supp}\, v'$. This forces $u = v'$. The valuation u is not trivial, since $A_u \subset A_v \neq R$. Thus $v' \in S(R/H)$.

If $w \in S(R/H)$ and $w \le v'$ then it is clear that $w = v'$ since $\Lambda \subset H \subset A_w$ (cf.§12, Def.3 and Prop.12.12.i). Thus $v' \in \omega(R/H)$. $\square$

Lemma 13.11. Assume that H is Prüfer in R. For every $u \in \omega(R/H)$ we have $u = u^\#_\Lambda$.

Proof. u is T_H-convex and $T_H = T_A$. Thus $u^\#_\Lambda$ is T_A-convex. This implies $A_{u^\#_\Lambda} \supset H$. $u^\#_\Lambda$ is certainly not trivial, since $u^\#_\Lambda \le u$. Thus $u^\#_\Lambda \in S(R/H)$. Again taking into account that $u^\#_\Lambda \le u$, we conclude that $u^\#_\Lambda = u$. $\square$

Theorem 13.12. Assume that H is Prüfer in R. Let $u \in \omega(R/H)$ be given. There exists a valuation $v \in \omega(R/A)$ with $v^\#_\Lambda = u$ iff $AA_u \neq R$. In this case v is uniquely determined by u (up to equivalence). We have $A_uA = A_v$ and $v = u^\#_A$.

Proof. If $v \in S(R/A)$ and $v_\Lambda^\# = u$ then $u \le v$, hence $A_u \subset A_v$. Since also $A \subset A_v$, we conclude that $AA_u \subset A_v$. In particular, $AA_u \neq R$.

Conversely, if $AA_u \neq R$ then, since u is PM, we have $AA_u = A_v$ with v a non trivial PM-valuation on R and $u \le v$ (cf.Cor.III.3.2). Moreover $v \in S(R/A)$, since $A \subset A_v$. By Theorem 12.11 and Lemma 10 we infer that $v_\Lambda^\# = u_\Lambda^\# = u$. Clearly v is the minimal coarsening of u with valuation ring $A_v \supset A$. Thus $v = u_A^\#$ (cf.Prop.12.12).

If $w \in S(R/A)$ and $w \le v$ then w is a coarsening of u, again by Theorem 12.11, hence $v = u_A^\# \le w$, hence $v \sim w$. This proves that $v \in \omega(R/A)$.

Finally, if $w \in \omega(R/A)$ and $w_\Lambda^\# = u$ then w is again a coarsening of u. Thus $v = u_A^\# \le w$, hence $v \sim w$. □

Corollary 13.13. Assume that H is Prüfer in R. Let $v \in S(R/A)$ be given. There exists a *unique* valuation (up to equivalence) $w \in \omega(R/A)$ with $w \le v$. We have $A_w = AA_{v_\Lambda^\#}$ and $w = v_A^\#$.

Proof. There exists some $w \in \omega(R/A)$ with $w \le v$. It is clear by Theorem 12.11 that w is unique, and that $v^\# = w^\#$. Theorem 12 tells us that $A_w = AA_{w^\#} = AA_{v^\#}$, and $w = w_A^\#$. From $w \le v$ we infer that $w_A^\# = v_A^\#$ (cf.Scholium 12.13). □

The corollary generalizes readily as follows.

Proposition 13.14. Assume that H is Prüfer in R. Let C be a subring of A which is T_A-convex in A (hence in R). For every $v \in S(R/A)$ there exists a unique $w \in \omega(R/C)$ with $w \le v$. We have $A_w = CA_{v_\Lambda^\#}$ and $w = v_C^\#$.

Proof. $H_{C/\Lambda} = H$ (cf.Corollary 8 and Remark 9.iv), and $v \in S(R/C)$. The preceding corollary gives the claim. □

As before we always assume that A *is Prüfer and convex in* R *and* Λ *is a subring of* A.

Open Problem. For which subrings Λ of A is

$$\omega(R/H_{A/\Lambda}) \;=\; \{v_\Lambda^\# \mid v \in \omega(R/A)\}\ ?$$

(Do there exist subrings for which this does not hold?)

Since this problem looks rather difficult we introduce a modification of the holomorphy base $H_{A/\Lambda}$ which seems to be more tractable.

Definition 4. The *weak holomorphy base of* A *over* Λ (*in* R) is the ring

$$H'_{A/\Lambda} := \left(H^R_{A/\Lambda}\right)' := \bigcap_{v \in \omega(R/A)} A_{v_\Lambda^\#}. \qquad \square$$

It is clear from above that $H_{A/\Lambda} \subset H'_{A/\Lambda} \subset A$, and that $H'_{A/\Lambda} = H_{A/\Lambda}$ iff the question above has a positive answer for the triple (R, A, Λ).

We fix a triple (R, A, Λ) and abbreviate $H' := H'_{A/\Lambda}$, $H := H_{A/\Lambda}$. It follows from $H \subset H' \subset A$ that $T_{H'} = T_A$ (cf.Cor.8). Moreover, quite a few results stated in Proposition 10 to Proposition 14 for H take over to H' with minor modifications.

Proposition 13.15. Assume that H' is Prüfer in R.
i) $v^{\#}_{\Lambda} \in \omega(R/H')$ for every $v \in S(R/A)$.
ii) If $v \in \omega(R/A)$ and $u := v^{\#}_{\Lambda}$ then $u^{\#}_{A} = v$ and $AA_u = A_v$.

Proof. If $v \in S(R/A)$ then $H' \subset A_{v^{\#}_{\Lambda}}$ by definition of H'. Thus $v^{\#}_{\Lambda} \in S(R/H')$. Running again through the arguments in part b) of the proof of Proposition 10, with H replaced by H', we obtain all claims. $\square$

Proposition 13.16. Assume that H is Prüfer in R. Let $u \in \omega(R/H')$ be given. The following are equivalent:
(1) There exists some $v \in \omega(R/A)$ with $v^{\#}_{\Lambda} = u$.
(2) $AA_u \neq R$.
If (1), (2) hold then $u \in \omega(R/H)$.

Proof. If (1) holds then $AA_u \subset A_v$, hence $AA_u \neq R$. Assume now (2). Let $u_0 := u^{\#}_{\Lambda}$. Applying Proposition 10 and Theorem 12 to the extension $H \subset H'$, we learn that $u_0 \in \omega(R/H)$ and $H'A_{u_0} = A_u$ and $u = (u_0)^{\#}_{A}$. We have $AA_{u_0} = AH'A_{u_0} = AA_u \neq R$, and we obtain, again by Theorem 12, that there exists a unique valuation $v \in \omega(R/A)$ with $v^{\#}_{\Lambda} = u_0$. By definition of H' we have $u_0 \in S(R/H')$. We conclude from $u_0 \leq u$ that $u_0 = u$. Thus $v^{\#}_{\Lambda} = u$ and $u \in \omega(R/H)$. $\square$

We have gained a modest insight into the restricted PM-spectra of R over the holomorphy base $H_{A/\Lambda}$ and the weak holomorphy base $H'_{A/\Lambda}$ for rings $\Lambda \subset A \subset R$ with A convex and Prüfer in R. A lot remains to be done to determine $H_{A/\Lambda}$ and $H'_{A/\Lambda}$ in more concrete terms in general and in examples.

References

[And] F.W. Anderson, *Lattice ordered rings of quotients.* Canad. J. Math. 17 (1965), 434-448.

[AS] E. Artin, O. Schreier, *Algebraische Konstruktion reeller Körper.* Abh. Math. Sem. Univ. Hamburg 5 (1927), 85-97.

[B] E. Becker, *Valuations and real places in the theory of formally real fields.* Lecture Notes in Math. 959, pp. 1-40, Springer-Verlag 1982.

[B_4] E. Becker, *The real holomorphy ring and sums of 2n-th powers.* Lecture Notes in Math. 959, pp. 139–181, Springer-Verlag 1982.

[Ban] B. Banaschewski, *Pointfree topology and the spectra of f-rings*, in: "Ordered algebraic structures" (W.C. Holland, J. Martinez, eds), pp.123-148. Kluwer Acad. Publ. 1997.

[BCR] J. Bochnak, M. Coste, M.-F. Roy, *Real algebraic geometry.* Ergeb. Math. Grenzgeb. 3. Folge, Band 36, Springer-Verlag 1998.

[BKW] A. Bigard, K. Keimel, S. Wolfenstein, *Groupes et anneaux réticulés.* Lecture Notes in Math. 608, Springer-Verlag 1977.

[BP] E. Becker, V. Powers, *Sums of powers in rings and the real holomorphy ring.* J. reine angew. Math. 480 (1996), 71–103.

[Br] G.W. Brumfiel, *Partially ordered rings and semi-algebraic geometry.* London Math. Soc. Lecture Notes 37. Cambridge University Press 1979.

[Br_1] G.W. Brumfiel, *Real valuation rings and ideals.* Lecture Notes in Math. 959, pp. 55–97, Springer-Verlag 1982.

[vd D] L. van den Dries, *Tame topology and o-minimal structures.* London Math. Soc. Lecture Notes 248, Cambridge University Press 1998.

[vd D_1] L. van den Dries, *O-minimal structures on the field of real numbers.* Jber. dt. Math.-Verein. 98 (1996), 165-171.

[DHH] F. Dashiell, A. Hager, M. Henriksen, *Order-completions of rings and vector lattices of continuous functions.* Canad. J. Math. 32 (1980), 657-685.

[Eg] N. Eggert, *Rings whose overrings are integrally closed in their complete quotient ring.* J. Reine Angew. Math. 282 (1976), pp.88-95.

[FGL] N.J. Fine, L. Gillman, J. Lambek, *Rings of quotients of rings of functions.* McGill University Press, Montreal, Quebec 1966.

[G_2] M. Griffin, *Valuations and Prüfer rings.* Canad. J. Math. 26 (1974), 412-429.

[GJ] L. Gillman, J. Jerison, *Rings of continuous functions.* Van Nostrand 1960. Reprint Springer 1976.

[Huc] J.A. Huckaba, *Commutative rings with zero divisors.* Marcel Dekker, New York, 1988.

[J] J.R. Joly, *Sommes de puissances d-ièmes dans un anneau commutatif,* Acta Arithm. 17 (1970), 37-114.

[KS] M. Knebusch, C. Scheiderer, *Einführung in die reelle Algebra.* Vieweg, 1989.

[KZ_1] M. Knebusch, D. Zhang, *Manis valuations and Prüfer extensions I – A new chapter in commutative algebra.* Lecture Notes in Math. 1791, Springer-Verlag 2002.

[La_1] T.Y. Lam, *An introduction to real algebra.* Rocky Mountain J. Math. 14 (1984), 767-814.

[Lb] J. Lambek, *Lectures on rings and modules.* Blaisdell, Toronto 1966.

[Mar] M. Marshall, *Orderings and real places on commutative rings.* J. Algebra 140 (1991), 484–501.

[Mart] J. Martinez, *On commutative rings which are strongly Prüfer.* Commun. Algebra 22, N°9 (1994), 3479-3488.

[P] V. Powers, *Valuations and higher level orders in commutative rings.* J. Algebra 172 (1995), 255-272.

[PW] J.R. Porter, R.G. Woods, *Extensions and absolutes of Hausdorff spaces.* Springer-Verlag 1988.

[Sch] N. Schwartz, *The basic theory of real closed spaces.* Regensburger Math. Schriften 15, 1987.

[Sch_1] N. Schwartz, *The basic theory of real closed spaces.* Memoirs Amer. Math. Soc. 397 (1989).

[Sch_2] N. Schwartz, *Epimorphic extensions and Prüfer extensions of partially ordered rings.* manuscr. math. 102 (2000), 347-381.

[Sch_3] N. Schwartz, *Rings of continuous functions as real closed rings.* In: W.C. Holland, J. Martinez (eds), Ordered algebraic structures, Kluwer Acad. Publ. 1997, 277-313.

[SchM] N. Schwartz, J. Madden, *Semi-algebraic function rings and reflectors of partially ordered rings.* Lecture Notes in Math. 1712, Springer-Verlag 1999.

[Z] D. Zhang, *Real Prüfer extensions of commutative rings.* manuscr. math. 106 (2001), 405-409.

[Z_1] D. Zhang, *The real holomorphy rings and the Schmüdgen property.* Commun. Algebra 30, N°12 (2002), 5795-5807.

Manfred Knebusch
University of Regensburg
Department of Mathematics
Universitätsstr. 31,
D-93040 Regensburg

Digen Zhang
University of Regensburg
Department of Mathematics
Universitätsstr. 31,
D-93040 Regensburg

On the Chow Groups of Quadratic Grassmannians

A. Vishik[1]

Received: November 30, 2004
Revised: December 9, 2004

Communicated by Ulf Rehmann

Abstract. In this text we get a description of the Chow-ring (modulo 2) of the Grassmanian of the middle-dimensional planes on arbitrary projective quadric. This is only a first step in the computation of the, so-called, generic discrete invariant of quadrics. This generic invariant contains the "splitting pattern" and "motivic decomposition type" invariants as specializations. Our computation gives an important invariant J(Q) of the quadric Q. We formulate a conjecture describing the canonical dimension of Q in terms of J(Q).

2000 Mathematics Subject Classification: 11E04, 14C15, 14M15
Keywords and Phrases: Quadrics, Chow groups, Grassmannians, Steenrod operations

Contents

[1]Support of the Weyl fund is acknowledged. Partially supported by RFBR grant 22005

1 Introduction

The current article is devoted to the computation of certain invariants of smooth projective quadrics. Among the invariants of quadrics one can distinguish those which could be called *discrete*. These are invariants whose values are (roughly speaking) collections of integers. For a quadric of given dimension such an invariant takes only finitely many values. The first example is the usual dimension of anisotropic part of q. More sophisticated example is given by the *splitting pattern* of Q, or the collection of *higher Witt indices* - see [7] and [9]. The question of describing the set of possible values of this invariant is still open. Some progress in this direction was achieved by considering the interplay of the splitting pattern invariant with another discrete invariant, called, *motivic decomposition type* - see [12]. The latter invariant measures in what pieces the Chow-motive of a quadric Q could be decomposed. The splitting pattern invariant can be interpreted in terms of the existence of certain cycles on various flag varieties associated to Q, and the motivic decomposition type can be interpreted in terms of the existence of certain cycles on $Q \times Q$. So, both these invariants are faces of the following invariant $GDI(Q)$, which we will call *(quite) generic discrete invariant.* Let Q be a quadric of dimension d, and, for any $1 \leqslant m \leqslant [d/2]+1$, let $G(m,Q)$ be the Grassmanian of projective subspaces of dimension $(m-1)$ on Q. Then $GDI(Q)$ is the collection of the subalgebras

$$C^*(G(m,Q)) := image(\mathrm{CH}^*(G(m,Q))/2 \to \mathrm{CH}^*(G(m,Q)|_{\overline{k}})/2).$$

It should be noticed that this invariant has a "noncompact form", where one uses powers of quadrics $Q^{\times r}$ instead of $G(m,Q)$. The equivalence of both forms follows from the fact that the Chow-motive of $Q^{\times r}$ can be decomposed into the direct sum of the Tate-shifts of the Chow-motives of $G(m,Q)$. The varieties $G(m,Q)|_{\overline{k}}$ have natural cellular structure, so Chow-ring for them is a finite-dimensional $\mathbb{Z}$-algebra with the fixed basis parametrized by the Young diagrams of some kind. This way, $GDI(Q)$ appears as a rather combinatorial object.
The idea is to try to describe the possible values of $GDI(Q)$, rather than that of the certain faces of it. In the present article we will address the computation of $GDI(m,Q)$ for the biggest possible $m=[d/2]+1$. This case corresponds to the Grassmannian of middle-dimensional planes on Q. It should be noticed, that it is sufficient to consider the case of odd-dimensional quadrics. This follows from the fact that for the quadric P of even dimension $2n$ and arbitrary codimension 1 subquadric Q in it, $G(n+1,P) = G(n,Q) \times_{\mathrm{Spec}(k)} \mathrm{Spec}(k\sqrt{det_{\pm}(P)})$.
Below we will show that, for $m=[d/2]+1$, the $GDI(m,Q)$ can be described in a rather simple terms - see Main Theorem 5.8 and Definition 5.11. The restriction on the possible values here is given by the Steenrod operations - see Proposition 5.12. And at the moment there is no other restrictions known - see Question 5.13 (the author would expect that there is none). Finally, in the last section we show that in the case of a *generic* quadric, the Grassmannian of

middle-dimensional planes is 2-incompressible, which gives a new proof of the conjecture of G.Berhuy and Z.Reichstein (see [1, Conjecture 12.4]). Also, we formulate a conjecture describing the canonical dimension of arbitrary quadric - see Conjecture 6.6.

Most of these results were announced at the conference on "Quadratic forms" in Oberwolfach in May 2002. This text was written while I was a member at the Institute for Advanced Study at Princeton, and I would like to express my gratitude to this institution for the support, excellent working conditions and very stimulating atmosphere. The support of the the Weyl Fund is deeply appreciated. I'm very gratefull to G.Berhuy for the numerous discussions concerning canonical dimension, which made it possible for the final section of this article to appear. Finally, I want to thank a referee for suggestions and remarks which helped to improve the exposition and for pointing out a mistake.

2 The Chow ring of the last Grassmannian

Let k be a field of characteristic different from 2, and q be a nondegenerate quadratic form on a $(2n+1)$-dimensional k-vector space W_q. Denote as $G(n,Q)$ the Grassmannian of n-dimensional totally isotropic subspaces in W_q. If q is completely split, then the corresponding Grassmanian will be denoted as $G(n)$, and the underlying space of the form q will be denoted as W_n. For small n, examples are: $G(1) \cong \mathbb{P}^1$, $G(2) \cong \mathbb{P}^3$, and $G(3) \cong Q_6$ - the 6-dimensional hyperbolic quadric.

The Chow ring $\mathrm{CH}^*(G(n))$ has $\mathbf{Z}$-basis, consisting of the elements of the type z_I, where I runs over all subsets of $\{1,\ldots,n\}$ (see [2, Propositions 1,2] and [5, Proposition 4.4]). In particular, $\mathrm{rank}(\mathrm{CH}^*(G(n))) = 2^n$. The degree (codimension) of z_I is $|I| = \sum_{i\in I} i$, and this cycle can be defined as the collection of such n-dimensional totally isotropic subspaces $A \subset W_q$, that

$$\dim(A \cap \pi_{n+1-j}) \geqslant \#(i \in I, i \geqslant j), \quad \text{for all } 1 \leqslant j \leqslant n,$$

where $\pi_1 \subset \ldots \subset \pi_n$ is the fixed flag of totally isotropic subspaces in W_q. The element $z_\emptyset$ is the ring unit $1 = [G(n)]$.

Other parts of the landscape are: the tautological n-dimensional bundle V_n on $G(n)$, and the embedding $G(n-1) \stackrel{j_{n-1}}{\to} G(n)$ given by the choice of a rational point $x \in Q$.

Fixing such a point x, let $M_n \subset G(n) \times G(n)$ be the closed subvariety of pairs (A,B), satisfying the conditions:

$$x \in B, \quad \text{and} \quad \mathrm{codim}(A \cap B \subset A) \leqslant 1.$$

The projection on the first factor $(A,B) \mapsto A$ defines a birational map $g_n : M_n \to G(n)$. In particular, by the projection formula, the map $g_n^* : \mathrm{CH}^*(G(n)) \to \mathrm{CH}^*(M_n)$ is injective. On the other hand, the rule $(A,B) \mapsto (B/x)$ defines the map $\pi : M_n \to G(n-1)$. Tautological bundle V_n is naturally a subbundle in the trivial $2n+1$-dimensional bundle $pr^*(W_n)$, which we will denote still by W_n. The variety M_n can be also described as the variety of pairs $B \subset C \subset W_n$, where B is totally isotropic, $dim(B) = n$, $dim(C) = n+1$, and $x \in B$. In other words, $M_n \cong \mathbb{P}_{G(n-1)}(j_{n-1}^*(W_n/V_n))$, where the identification is given by the rule:

$$(A,B) \mapsto (A+B)/B \quad \text{(respectively,} \quad (B,B) \mapsto B^{\perp}/B \ \text{)}.$$

Clearly, $j_{n-1}^*(W_n/V_n) = (W_{n-1}/V_{n-1}) \oplus \mathcal{O}$, and $W_{n-1}/V_{n-1}^{\perp} \cong (V_{n-1}^{\vee})$. We have a 3-step filtration $V_{n-1} \subset V_{n-1}^{\perp} \subset W_{n-1}$, with first and third graded pieces mutually dual. Hence, the top exterior power of W_{n-1} is isomorphic to the middle graded piece (which is a linear bundle): $V_{n-1}^{\perp}/V_{n-1} = \Lambda^{2n-1}W_{n-1} \cong \mathcal{O}$. Thus, $M_n \cong \mathbb{P}_{G(n-1)}(Y_{n-1})$, where $[Y_{n-1}] = [V_{n-1}^{\vee}] + 2[\mathcal{O}] \in \mathrm{K}_0(G(n-1))$. We get a diagram

$$G(n) \stackrel{g_n}{\leftarrow} \mathbb{P}_{G(n-1)}(Y_{n-1}) \stackrel{\pi_{n-1}}{\rightarrow} G(n-1).$$

Using the exact sequences $0 \to A \to (A+B) \to (A+B)/A \to 0$ and $0 \to B \to (A+B) \to (A+B)/B \to 0$, and the fact that q defines a nondegenerate pairing between the spaces $(A+B)/B \cong A/(A \cap B)$ and $(A+B)/A \cong B/(A \cap B)$ for all pairs (A,B) aside from the codimension > 1 subvariety $(\Delta(G(n)) \cap M_n) \subset M_n$, we get the exact sequences:

$$0 \to g_n^*(V_n) \to X_{n-1} \to \mathcal{O}(1) \to 0, \quad \text{and}$$

$$0 \to \pi_{n-1}^*(V_{n-1}) \oplus \mathcal{O} \to X_{n-1} \to \mathcal{O}(-1) \to 0,$$

where X_{n-1} is the bundle with the fiber C. In particular,

$$[g_n^*(V_n)] = [\pi_{n-1}^*(V_{n-1})] + [\mathcal{O}] + [\mathcal{O}(-1)] - [\mathcal{O}(1)]. \quad \text{Also,}$$

$$\mathrm{CH}^*(\mathbb{P}_{G(n-1)}(Y_{n-1})) = \mathrm{CH}^*(G(n-1))[\rho]/(\rho^2 \cdot c(V_{n-1}^{\vee})(\rho)),$$

where $\rho = c_1(\mathcal{O}(1))$, and $c(E)(t) = \sum_{i=0}^{dim(E)} c_i(E)t^{dim(E)-i}$ is the total Chern polynomial of the vector bundle E.
Consider the open subvariety $\tilde{M}_n := g_n^{-1}(G(n) \backslash j_{n-1}(G(n-1))) \subset M_n$ The map $\tilde{g}_n : \tilde{M}_n \to G(n) \backslash j_{n-1}(G(n-1))$ is an isomorphism, and $\tilde{\pi}_{n-1} : \tilde{M}_n \to G(n-1)$ is an n-dimensional affine bundle over $G(n-1)$.

Proposition 2.1 *There is split exact sequence*

$$0 \to \mathrm{CH}^{*-n}(G(n-1)) \stackrel{j_{n-1*}}{\rightarrow} \mathrm{CH}^*(G(n)) \stackrel{j_{n-1}^*}{\rightarrow} \mathrm{CH}^*(G(n-1)) \to 0.$$

Proof: Consider commutative diagram:

$$\begin{array}{ccccc} G(n) & \xleftarrow{g_n} & M_n & \xrightarrow{\pi_{n-1}} & G(n-1) \\ \varphi\uparrow & & \psi\uparrow & & \| \\ G(n)\backslash j_{n-1}(G(n-1)) & \xleftarrow[\tilde{g}_n]{} & \tilde{M}_n & \xrightarrow[\tilde{\pi}_{n-1}]{} & G(n-1). \end{array}$$

Notice that the choice of a point $y \in Q\backslash T_{x,Q}$ gives a section $s : G(n-1) \to \tilde{M}_n$ of the affine bundle $\tilde{\pi}_{n-1} : \tilde{M}_n \to G(n-1)$. And the composition $\varphi \circ \tilde{g}_n \circ s$ is equal to j'_{n-1}, where j'_{n-1} is constructed from the point $y \in Q$ in the same way as j_{n-1} was constructed from the point x. Thus, the isomorphism $\mathrm{CH}^*(G(n)\backslash j_{n-1}(G(n-1))) \xrightarrow{(\tilde{\pi}^*_{n-1})^{-1}\tilde{g}^*_n} \mathrm{CH}^*(G(n-1))$ together with the localization at $G(n-1) \overset{j_{n-1}}{\to} G(n) \overset{\varphi}{\leftarrow} (G(n)\backslash G(n-1))$ gives us exact sequence

$$\mathrm{CH}^{*-n}(G(n-1)) \overset{j_{n-1*}}{\to} \mathrm{CH}^*(G(n)) \overset{{j'_{n-1}}^*}{\to} \mathrm{CH}^*(G(n-1)) \to 0.$$

Thus, $ker({j'_{n-1}}^*) = im({j_{n-1}}_*)$. Since it is true for arbitrary pair of points $x, y \in Q$ satisfying the condition that the line passing through them does not belong to a quadric, we get: $ker({j_{n-1}}^*) = im({j_{n-1}}_*)$. On the other hand, the map ${j'_{n-1}}_* : \mathrm{CH}^{*-n}(G(n-1)) \to \mathrm{CH}^*(G(n))$ is split injective, since $(\tilde{\pi}_{n-1})_* \circ \tilde{g}^*_n \circ \varphi^* \circ {j'_{n-1}}_* = id$. Then the same is true for ${j_{n-1}}_*$. And we get the desired split exact sequence. □

Lemma 2.2 *The ring* $\mathrm{CH}^*(G(n))$ *is generated by the elements of degree* $\leqslant n$.

Proof: It easily follows by induction with the help of Proposition 2.1, and projection formula. □

Proposition 2.3 *Let* q *be* $2n+1$*-dimensional split quadratic form. Then*

(1) *The group* $O(q)$ *acts trivially on* $\mathrm{CH}^*(G(n))$.

(2) *The maps* j^*_{n-1} *and* j_{n-1*} *do not depend on the choice of a point* $x \in Q$.

Proof: Use induction on n. For $n = 1$ the statement is trivial. Suppose it is true for $(n-1)$. Let $j_{n-1,x} : G(n-1)_x \to G(n)$ be the map corresponding to the point $x \in Q$. For any $\varphi \in O(q)$ such that $\varphi(x) = y$, we have the map $\varphi_{x,y} : G(n-1)_x \to G(n-1)_y$ such that $j_{n-1,y} \circ \varphi_{x,y} = \varphi \circ j_{n-1,x}$. By the inductive assumption, the maps $\varphi^*_{x,y}$ and $(\varphi_{x,y})_* = ((\varphi_{x,y})^*)^{-1}$ define canonical identification of $CH^*(G(n-1)_x)$ and $CH^*(G(n-1)_y)$ which does not depend on the choice of φ. And under this identification,

$$j^*_{n-1,x} \circ \varphi^* = j^*_{n-1,y} \text{ and } \varphi_* \circ (j_{n-1,x})_* = (j_{n-1,y})_*.$$

Let $\varphi \in O(q)$ be arbitrary element, and $x, y \in Q$ be such (rational) points that $\varphi(x) = y$. Let z be arbitrary point on Q such that neither of lines $l(x,z)$, $l(y,z)$ lives on Q. Consider reflections $\tau_{x,z}$ and $\tau_{y,z}$. They are rationally connected in $O(q)$. Consequently, for $\psi := \tau_{y,z} \circ \tau_{x,y}$, $\psi^* = id = \psi_*$. Thus, $(j_{n-1,x})^* = (j_{n-1,y})^*$ and $(j_{n-1,x})_* = (j_{n-1,y})_*$.
From Proposition 2.1 we get the commutative diagram with exact rows:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathrm{CH}^{*-n}(G(n-1)_y) & \xrightarrow{(j_{n-1,y})_*} & \mathrm{CH}^*(G(n)) & \xrightarrow{j^*_{n-1,x}} & \mathrm{CH}^*(G(n-1)_x) & \longrightarrow & 0 \\
 & & & & \Big\uparrow{\scriptstyle \varphi^*=(\varphi_*)^{-1}} & & & & \\
0 & \longrightarrow & \mathrm{CH}^{*-n}(G(n-1)_x) & \xrightarrow{(j_{n-1,x})_*} & \mathrm{CH}^*(G(n)) & \xrightarrow{j^*_{n-1,y}} & \mathrm{CH}^*(G(n-1)_y) & \longrightarrow & 0.
\end{array}$$

It implies that φ^* is identity on elements of degree $\leqslant n$. Since, by the Lemma 2.2, such elements generate $\mathrm{CH}^*(G(n))$ as a ring, $\varphi^* = id = \varphi_*$, and $j_{n-1}^*, {j_{n-1}}_*$ are well-defined. $\square$

Proposition 2.4 *There is unique set of elements $z_i \in \mathrm{CH}^i(G(n))$ defined for all $n \geqslant 1$ and satisfying the properties:*

(0) *For $G(1) \cong \mathbb{P}^1$, z_1 is the class of a point.*

(1) *As a $\mathbb{Z}$-module, $\mathrm{CH}^*(G(n)) = \oplus_{I \subset \{1,\dots,n\}} \mathbb{Z} \cdot \prod_{i \in I} z_i$.*

(2) ${j_{n-1}}_*(1) = z_n$.

(3) $j_{n-1}^* : \mathrm{CH}^*(G(n)) \to \mathrm{CH}^*(G(n-1))$ *is given by the following rule on the additive generators above:*

$$\prod_{i \in I} z_i \mapsto \begin{cases} 0, & \text{if } n \in I; \\ \prod_{i \in I} z_i, & \text{if } n \notin I. \end{cases}$$

Proof: Let us introduce the elementary cycles $z_i \in \mathrm{CH}^i(G(n))$ inductively as follows: For $n = 1$, $G(1) \cong \mathbb{P}^1$, and z_1 is just the class of a point. Let $z_i \in \mathrm{CH}^i(G(n-1))$, for $1 \leqslant i \leqslant n-1$ are defined and satisfy the condition $(1) - (3)$. Let us define similar cycles on $G(n)$.
From the Proposition 2.1 we get: ${j_{n-1}}^*$ is an isomorphism on CH^i, for $i < n$. Now, for $1 \leqslant i \leqslant n-1$, we define $z_i \in \mathrm{CH}^i(G(n))$ as unique element corresponding under this isomorphism to $z_i \in \mathrm{CH}^i(G(n-1))$. And put: $z_n := {j_{n-1}}_*(1)$. We automatically get (2) satisfied.
Let $J \subset \{1, \dots, n-1\}$. From the projection formula we get:

$${j_{n-1}}_*(\prod_{j \in J} z_j) = z_n \cdot \prod_{j \in J} z_j.$$

Applying once more Proposition 2.1, we get condition (1) and (3). $\square$

Remark: The cycle z_i we constructed is given by the set of n-dimensional totally isotropic subspaces $A \subset W_n$ satisfying the condition: $A \cap \pi_{n+1-i} \neq 0$ for fixed totally isotropic subspace π_{n+1-i} of dimension $(n+1-i)$.
Consider the commutative diagram:

$$\begin{array}{ccccc} G(n) & \xleftarrow[g_n]{} & \mathbb{P}_{G(n-1)}(Y_{n-1}) & \xrightarrow[\pi_{n-1}]{} & G(n-1) \\ {\scriptstyle j_{n-1}}\uparrow & & {\scriptstyle j}\uparrow & & \uparrow{\scriptstyle j_{n-2}} \\ G(n-1) & \xleftarrow[g_{n-1}]{} & \mathbb{P}_{G(n-2)}(Y_{n-2}) & \xrightarrow[\pi_{n-2}]{} & G(n-2). \end{array}$$

By Proposition 2.4, the ring homomorphism

$$j^* : \mathrm{CH}^*(\mathbb{P}_{G(n-1)}(Y_{n-1})) \to \mathrm{CH}^*(\mathbb{P}_{G(n-2)}(Y_{n-2}))$$

is a surjection, and it's kernel is generated as an ideal by the elements $\pi^*_{n-1}(z_{n-1})$ and $\rho^2 \cdot c(V^\vee_{n-2})(\rho)$. In particular, $(j^*)^k$ is an isomorphism for all $k < n-1$, and the kernel of $(j^*)^{n-1}$ is additively generated by $\pi^*_{n-1}(z_{n-1})$.

Theorem 2.5 *Let V_n be tautological bundle on $G(n)$, z_i be elements defined in Proposition* 2.4, *and $\rho = c_1(\mathcal{O}(1))$. Then:*

(1) $g^*_n(z_k) = \rho^k + 2\sum_{0<i<k} \rho^{k-i}\pi^*_{n-1}(z_i) + \pi^*_{n-1}(z_k)$, *for all* $0 < k < n$.

(2) $g^*_n(z_n) = \rho^n + 2\sum_{0<i<n} \rho^{n-i}\pi^*_{n-1}(z_i)$.

(3) $c(V_n)(t) = t^n + 2\sum_{1\leqslant i\leqslant n}(-1)^i z_i t^{n-i}$.

Proof: $G(1)$ is a conic and $V_1 \cong \mathcal{O}(-1)$. Hence, $c(V_1)(t) = t - 2z_1$. Let now $(1)_{m,k}$ is proven for all $m < n$ and all $0 < k < m$, and $(2)_m$ and $(3)_m$ are proven for all $m < n$.
Since $(j^*)^k : \mathrm{CH}^k(\mathbb{P}_{G(n-1)}(Y_{n-1})) \to \mathrm{CH}^k(\mathbb{P}_{G(n-2)}(Y_{n-2}))$ is an isomorphism for $k < n-1$, the condition $(1)_{n,k}$ follows from $(1)_{n-1,k}$ for all such k in view of $j^*(\rho) = \rho$ and $j^*(\pi^*_{n-1}(z_k)) = \pi^*_{n-2}(z_k)$ (Proposition 2.4(3)). Analogously, since the kernel $(j^*)^{n-1}$ is additively generated by $\pi^*_{n-1}(z_{n-1})$, the condition $(2)_{n-1}$ implies that $g^*_n(z_{n-1}) = \rho^{n-1} + 2\sum_{1\leqslant i<n-1}\rho^{n-1-i}\pi^*_{n-1}(z_i) + \lambda \cdot \pi^*_{n-1}(z_{n-1})$, where $\lambda \in \mathbf{Z}$. Since $Y_{n-1} = \mathcal{O} \oplus (V^\perp_{n-1})^\vee$, the projection $\pi_{n-1} : \mathbb{P}_{G(n-1)}(Y_{n-1}) \to G(n-1)$ has the section s (given by the rule: $(B/x) \mapsto (B,B)$). It satisfies: $g_n \circ s = j_{n-1}$. Since $s^*(\pi^*_{n-1}(z_{n-1})) = z_{n-1}$, $s^*(\rho) = 0$ and $j^*_{n-1}(z_{n-1}) = z_{n-1}$, we get $\lambda = 1$, which implies $(1)_{n,n-1}$.
Choose some rational point $y \in Q\backslash T_{x,Q}$. By Propositions 2.3(2) and 2.4(2), the cycle z_n is defined as the set of such planes A, that $y \in A$. Then the cycle $g^*_n(z_n)$ is the set of such pairs (A, B), that $y \in A$, $x \in B$ and $\dim(A+B/A) \leqslant 1$. Thus $A+B = y+B$, and $g^*_n(z_n)$ is given by the section $\mathbb{P}_{G(n-1)}(\mathcal{O}) \subset \mathbb{P}_{G(n-1)}(Y_{n-1})$. Since $c(Y_{n-1})(t) = t^2 \cdot c(\pi^*_{n-1}(V^\vee_{n-1}))(t)$, this class can be expressed as $\rho \cdot c(\pi^*_{n-1}(V^\vee_{n-1}))(\rho)$. The last expression is equal to $\rho^n + 2\rho^{n-1}\pi^*_{n-1}(z_1) + \ldots + 2\rho\pi^*_{n-1}(z_{n-1})$ because of $(3)_{n-1}$. The statement $(2)_n$ is proven.
Finally, since $[g^*_n(V_n)] = [\pi^*_{n-1}(V_{n-1})] + [\mathcal{O}] + [\mathcal{O}(-1)] - [\mathcal{O}(1)]$,
$g^*_n(c(V_n)(t)) = \pi^*_{n-1}(c(V_{n-1})(t)) \cdot \frac{t\cdot(t-\rho)}{t+\rho}$. In the light of $(3)_{n-1}$, this is equal to

$$\left(t^{n-1} + 2\sum_{1\leqslant i\leqslant n-1}(-1)^i\pi^*_{n-1}(z_i)t^{n-1-i}\right) \cdot \frac{t\cdot(t-\rho)}{t+\rho}.$$

Using the equality $\rho^2(\rho^{n-1} + 2\rho^{n-2}\pi^*_{n-1}(z_1) + \ldots + 2\pi^*_{n-1}(z_{n-1})) = 0$, as well as the conditions $(1)_{n,k}$ and $(2)_n$, we can rewrite the last expression as:

$$t^n + 2\sum_{1\leqslant i\leqslant n}(-1)^i g^*_n(z_i)t^{n-i}.$$

Since g^*_n is injective (the map g_n is birational), we get:

$$c(V_n)(t) = t^n + 2\sum_{1\leqslant i\leqslant n}(-1)^i z_i t^{n-i}.$$

The statement $(3)_n$ is proven. $\square$

3 Multiplicative structure

The multiplicative structure of $\mathrm{CH}^*(G(n))$ was studied extensively by H.Hiller, B.Boe, J.Stembridge, P.Pragacz and J.Ratajski - see [6], [11].
We can compute this ring structure from Theorem 2.5. Although, we restrict our consideration only to $(\,mod\,2)$ case, it should be pointed out that the integral case can be obtained in a similar way.
Let us denote as $\overline{u}$ the image of u under the map $\mathrm{CH}^* \to \mathrm{CH}^*/2$.

Proposition 3.1

$$\mathrm{CH}^*(G(n))/2 = \bigotimes_{1\leqslant d\leqslant n;\, d-odd} (\mathbb{Z}/2[\overline{z}_d]/(\overline{z}_d^{2^{m_d}})),$$

where $m_d = [log_2(n/d)] + 1$.

Proof: Consider the diagram:

$$G(n) \xleftarrow{g_n} \mathbb{P}_{G(n-1)}(Y_{n-1}) \xrightarrow{\pi_{n-1}} G(n-1).$$

From Theorem 2.5, $g_n^*(\overline{z}_k) = \overline{\rho}^k + \pi_{n-1}^*(\overline{z}_k)$, for $k < n$, and $g_n^*(\overline{z}_n) = \overline{\rho}^n$. Then it easily follows by the induction on n, that $\overline{z}_k^2 = \overline{z}_{2k}$ (where we assume $\overline{z}_r = 0$ if $r > n$).
Thus, we have surjective ring homomorphism

$$\bigotimes_{1\leqslant d\leqslant n;\, d-\mathrm{odd}} (\mathbb{Z}/2[\overline{z}_d]/(\overline{z}_d^{2^{m_d}})) \to \mathrm{CH}^*(G(n))/2.$$

Since the dimensions of both rings are equal to 2^n, it is an isomorphism. $\square$

Let J be a set. Let us call a *multisubset* the collection $\Lambda = \coprod_{\beta\in B}\Lambda_\beta$ of disjoint subsets of J. For a subset I of J, we will denote by the same symbol I the multisubset $\coprod_{i\in I}\{i\}$. Let $B = \coprod_{\gamma\in C} B_\gamma$, and $\Lambda'_\gamma = \coprod_{\beta\in B_\gamma}\Lambda_\beta$. Then the multisubset $\Lambda' := \coprod_{\gamma\in C}\Lambda'_\gamma$ is called the *specialization* of Λ. We call the specialization *simple* if $\#(B_\gamma) \leqslant 2$, for all $\gamma \in C$.
Let J now be some set of natural numbers (it may contain multiple entries). Then to any finite multisubset $\Lambda = \coprod_{\beta\in B}\Lambda_\beta$ of J we can assign the set of natural numbers $\overline{\Lambda} := \{\sum_{i\in\Lambda_\beta} i\}_{\beta\in B}$. We call the specialization Λ *good* if $\overline{\Lambda} \subset \{1, \ldots, n\}$.
Suppose I be some finite set of natural numbers. Let us define the element $\overline{z}_I \in \mathrm{CH}^*(G(n))/2$ by the formula:

$$\overline{z}_I = \sum_{\Lambda}\prod_{j\in\overline{\Lambda}} \overline{z}_j,$$

where we assume $\overline{z}_r = 0$, if $r > n$, and the sum is taken over all simple specializations Λ of the multisubset $I = \coprod_{i\in I}\{i\}$. Actually, $\overline{z}_I$ is just reduction modulo 2 of the Schubert cell class z_I. This follows from the Pieri formula of H.Hiller and B.Boe (see [6]) and our Proposition 3.3. We do not use this fact, but instead prove directly that $\overline{z}_I$ form basis (Proposition 3.4(1)).

Lemma 3.2 *If $I \not\subset \{1,\ldots,n\}$, then $\overline{z}_I = 0$.*

Proof: If I contains an element $r > n$, then $\overline{z}_I$ is clearly zero. Suppose now that I contains some element i twice, say as i_1 and i_2. Consider the subgroup $Z_2 \subset S_{\#(I)}$ interchanging i_1 and i_2 and keeping all other elements in place. We get $\mathbb{Z}/2$-action on our specializations. The terms which are not stable under this action will appear with multiplicity 2, so, we can restrict our attention to the stable terms. But such specializations have the property that $\{i_1, i_2\}$ is disjoint from the rest of i's, and the corresponding sum looks as: $\sum_M \prod_{j\in\overline{M}} \overline{z}_j \cdot (\overline{z}_i^2 + \overline{z}_{2i})$, where the sum is taken over all simple specializations of the multisubset $I\backslash\{i_1, i_2\}$. Since $\overline{z}_i^2 = \overline{z}_{2i}$, this expression is zero. □

We immediately get the (modulo 2) version of the Pieri formula proved by H.Hiller and B.Boe:

Proposition 3.3 ([6])

$$\overline{z}_I \cdot \overline{z}_j = \overline{z}_{I\cup j} + \sum_{i\in I} \overline{z}_{(I\backslash i)\cup(i+j)},$$

where we omit terms $\overline{z}_J$ with $J \not\subset \{1,\ldots,n\}$ (in particular, if J contains some element with multiplicity > 1).

Proof: $\overline{z}_{I\cup j} = \sum_\Lambda \prod_{l\in\overline{\Lambda}} \overline{z}_l$, where the sum is taken over all simple specializations of the multisubset $I \cup j$. We can distinguish two types of specializations: 1) j is separated from I; 2) j is not separated from I, that is, there is β such that $\Lambda_\beta = \{i, j\}$, for some $i \in I$. Let us call the latter specializations to be of type $(2, i)$. Clearly, the sum over specializations of the first kind is equal to $\overline{z}_I \cdot \overline{z}_j$, and the sum over the specializations of the type $(2, i)$ is equal to $\overline{z}_{(I\backslash i)\cup(i+j)}$. Finally, the terms with $J \not\subset \{1,\ldots,n\}$ could be omitted by Lemma 3.2. □

We also get the expression of monomials on z_i's in terms of z_I's.

Proposition 3.4 (1) *The set $\{\overline{z}_I\}_{I\subset\{1,\ldots,n\}}$ is a basis of $\mathrm{CH}^*(G(n))/2$.*

(2) $\prod_{i\in I} \overline{z}_i = \sum_\Lambda \overline{z}_{\overline{\Lambda}}$, *where sum is taken over all good specializations of I.*

Proof:
(1) On the $\mathbb{Z}/2$-vector space $\mathrm{CH}^*(G(n))/2 = \oplus_{I\subset\{1,\ldots,n\}}\mathbb{Z}/2 \cdot \prod_{i\in I} \overline{z}_i$ we have lexicographical filtration. Consider the linear map $\varepsilon : \mathrm{CH}^*(G(n))/2 \to$

$\mathrm{CH}^*(G(n))/2$ sending $\prod_{i\in I}\overline{z}_i$ to $\overline{z}_I$. Then the associated graded map: $gr(\varepsilon)$ is the identity. Thus, ε is invertible, and the set $\{\overline{z}_I\}_{I\subset\{1,\dots,n\}}$ form a basis.
(2) Consider the $\mathbb{Z}/2$-vector spaces $W_1 := \oplus_\Lambda \mathbb{Z}/2\cdot x_\Lambda$, and $W_2 := \oplus_\Lambda \mathbb{Z}/2\cdot y_\Lambda$, where Λ runs over all finite multisubsets of $\mathbb{N}$.
Consider the linear maps $\psi : W_2 \to W_1$ which sends y_Λ to the $\sum_{\Lambda'} x_{\Lambda'}$, where the sum is taken over all specializations of Λ, and $\varphi : W_1 \to W_2$ which sends x_Λ to the $\sum_{\Lambda'} y_{\Lambda'}$, where the sum is taken over all simple specializations Λ' of Λ. It is an easy exercise to show that φ and ψ are mutually inverse.
Consider the linear surjective maps: $w_1 : W_1 \to \mathrm{CH}^*(G(n))/2$ and $w_2 : W_2 \to \mathrm{CH}^*(G(n))/2$ given by the rule: $w_1(x_\Lambda) := \overline{z}_{\overline{\Lambda}}$, and $w_2(y_\Lambda) := \prod_{j\in\overline{\Lambda}}\overline{z}_j$.
Then, by the definition of $\overline{z}_I$, $w_1 = w_2\circ\varphi$. Then $w_2 = w_1\circ\psi$, which implies that $\prod_{i\in I}\overline{z}_i = \sum_\Lambda \overline{z}_{\overline{\Lambda}}$, where the sum is taken over all specializations of I. It remains to notice, that nongood specializations do not contribute to the sum (by Lemma 3.2). □

Examples: 1) $\overline{z}_i\cdot\overline{z}_j = \overline{z}_{i,j} + \overline{z}_{i+j}$, where the first term is omitted if $i=j$ and the second if $i+j>n$. 2) $\overline{z}_{i,j,k} = \overline{z}_i\cdot\overline{z}_j\cdot\overline{z}_k + \overline{z}_{i+j}\cdot\overline{z}_k + \overline{z}_{j+k}\cdot\overline{z}_i + \overline{z}_{i+k}\cdot\overline{z}_j$.

4 Action of the Steenrod algebra

On the Chow-groups modulo prime l there is the action of the Steenrod algebra. Such action was constructed by V.Voevodsky in the context of arbitrary motivic cohomology - see [13], and then a simpler construction was given by P.Brosnan for the case of usual Chow groups - see [3]. For quadratic Grassmannians we will be interested only in the case $l=2$.
We can compute the action of the Steenrod squares $S^r : \mathrm{CH}^*/2 \to \mathrm{CH}^{*+r}/2$ on the cycles $\overline{z}_i$. For convenience, let us put $z_j \in \mathrm{CH}^j(G(m))$ to be zero for $j>m$.

Theorem 4.1

$$\mathrm{S}^r(\overline{z}_i) = \binom{i}{r}\cdot\overline{z}_{i+r}$$

Proof: Use induction on n. The base is trivial. Suppose the statement is true for $(n-1)$. Since $\overline{c}(V_n)(t) = t^n$, we have: $\overline{\rho}^{n+1} = 0$. Then, by Theorem 2.5 and the assumption above, $g_n^*(\overline{z}_j) = \overline{\rho}^j + \pi_{n-1}^*(\overline{z}_j)$, for all j. Using the fact that S^r commutes with the pull-back morphisms (see [3]), and the inductive assumption, we get:

$$g_n^*(\mathrm{S}^r(\overline{z}_i)) = \mathrm{S}^r(g_n^*(\overline{z}_i)) = \mathrm{S}^r(\overline{\rho}^i + \pi_{n-1}^*(\overline{z}_i)) =$$
$$\binom{i}{r}\overline{\rho}^{i+r} + \pi_{n-1}^*(\binom{i}{r}\overline{z}_{i+r}) = \binom{i}{r}\cdot g_n^*(\overline{z}_{i+r}).$$

Now, the statement follows from the injectivity of g_n^*. □

5 Main theorem

Let X be some variety over the field k. We will denote:

$$C^*(X) := image(\mathrm{CH}^*(X)/2 \to \mathrm{CH}^*(X|_{\overline{k}})/2).$$

Let now Q be a smooth projective quadric of dimension $2n-1$, and $X = G(n,Q)$ be the Grassmanian of middle-dimensional projective planes on it. Then $X|_{\overline{k}} = G(n)$. In this section we will show that, as an algebra, $C^*(G(n,Q))$ is generated by the elementary cycles $\overline{z}_i$ contained in it.
Let $F(n,Q)$ be the variety of complete flags $(l_0 \subset l_1 \subset \ldots \subset l_{n-1})$ of projective subspaces on Q. Then $F(n,Q)$ is naturally isomorphic to the complete flag variety $F_{G(n,Q)}(V_n)$ of the tautological n-dimensional bundle on $G(n,Q)$. On the variety $F(n,Q)$ there are natural (subquotient) line bundles $\mathcal{L}_1,\ldots,\mathcal{L}_n$. The first Chern classes $c_1(\mathcal{L}_i)$, $1 \leqslant i \leqslant n$ generate the ring $\mathrm{CH}^*(F(n,Q))$ as an algebra over $\mathrm{CH}^*(G(n,Q))$, and the relations among them are: $\sigma_j(c_1(\mathcal{L}_1),\ldots,c_1(\mathcal{L}_n)) = c_j(V_n)$, $1 \leqslant j \leqslant n$ - see [4, Example 3.3.5]. Let F_n be the variety of complete flags of subspaces of the n-dimensional vector space V. It also has natural line bundles $\mathcal{L}'_1,\ldots,\mathcal{L}'_n$. Again, the first Chern classes $c_1(\mathcal{L}'_i)$ generate the ring $\mathrm{CH}^*(F_n)$. By Theorem 2.5 (3), modulo 2, all Chern classes $c_j(V_n)$ are the same as the Chern classes of the trivial n-dimensional bundle $\oplus_{i=1}^n\mathcal{O}$. Thus, modulo 2, the Chow ring of $F_{G(n,Q)}(V_n)$ is isomorphic to the Chow ring of $F_{G(n,Q)}(\oplus_{i=1}^n\mathcal{O})$. We get:

Theorem 5.1 *There is a ring isomorphism*

$$\mathrm{CH}^*(F(n,Q))/2 \cong \mathrm{CH}^*(G(n,Q))/2 \otimes_{\mathbb{Z}/2} \mathrm{CH}^*(F_n)/2,$$

where the map $\mathrm{CH}^*(G(n,Q)) \to \mathrm{CH}^*(F(n,Q))$ *is induced by the natural projection* $F(n,Q) \to G(n,Q)$*, and the map* $\mathrm{CH}^*(F_n)/2 \to \mathrm{CH}^*(F(n,Q))/2$ *is given on the generators by the rule:* $c_1(\mathcal{L}'_i) \mapsto c_1(\mathcal{L}_i)$.

□

Notice, that the change of scalar map $\mathrm{CH}^*(F_n) \to \mathrm{CH}^*(F_n|_{\overline{k}})$ is an isomorphism, and, $C^*(F(n,Q)) = C^*(G(n,Q)) \otimes_{\mathbb{Z}/2} \mathrm{CH}^*(F_n|_{\overline{k}})/2$. Thus, we have:

Statement 5.2 *Let* $v_1,\ldots,v_s$ *be linearly independent elements of* $\mathrm{CH}^*(F_n|_{\overline{k}})/2$*, and* $x_i \in \mathrm{CH}^*(G(n,Q)|_{\overline{k}})/2$*, then* $x = \sum_{i=1}^s x_i \cdot v_i$ *belongs to* $C^*(F(n,Q))$ *if and only if all* $x_i \in C^*(G(n,Q))$.

□

The ring $\mathrm{CH}^*(F_n)$ can be described as follows. Let us denote $c_1(\mathcal{L}_j)$ as h_j, and the set $\{h_j,\ldots,h_n\}$ as $\underline{h}(j)$ (and $\underline{h}(1)$ as $\underline{h}$). For arbitrary set of variables $\underline{u} = \{u_1,\ldots,u_r\}$ let us define the degree i polynomials $\sigma_i(\underline{u})$ and $\sigma_{-i}(\underline{u})$ from the equation:

$$\prod_l (1+u_l) = \sum_i \sigma_i(\underline{u}) = \left(\sum_i \sigma_{-i}(\underline{u})\right)^{-1}.$$

STATEMENT 5.3 ([4, Example 3.3.5])

$$\mathrm{CH}^*(F_n) = \mathbb{Z}[\underline{h}]/(\sigma_i(\underline{h}),\ 1 \leqslant i \leqslant n) = \mathbb{Z}[\underline{h}]/(\sigma_{-i}(\underline{h}(i)),\ 1 \leqslant i \leqslant n).$$

Since $\sigma_{-i}(\underline{h}(i))$ is the $\pm$-monic polynomial in h_i with coefficients in the subring, generated by $\underline{h}(i+1)$, we get: $\mathrm{CH}^*(F_n)$ is a free module over the subring $\mathbb{Z}[\underline{h}(n)]/(\sigma_{-n}(\underline{h}(n))) = \mathbb{Z}[h_n]/(h_n^n)$.
Let $\pi : F(n,Q) \to F(n-1,Q)$ be the natural projection between full flag varieties. We will denote by the same symbol $\overline{z}_I$ the images of $\overline{z}_I$ in $\mathrm{CH}^*(F(n,Q))/2$.
The following statement is the key for the Main Theorem.

PROPOSITION 5.4

$$\pi^*\pi_*(\overline{z}_I) = \sum_{i \in I} \overline{z}_{(I \setminus i)} \cdot \pi^*\pi_*(\overline{z}_i).$$

PROOF: $F(n,Q)$ is a conic bundle over $F(n-1,Q)$ inside the projective bundle $\mathbb{P}_{F(n-1,Q)}(V)$, where, in $K_0(F(n,Q))$, $\pi^*[V] = [\mathcal{L}_n] + [\mathcal{L}_n^{-1}] + [\mathcal{O}]$. Sheaf $\mathcal{L}_n$ is nothing else but the restriction of the sheaf $\mathcal{O}(-1)$ from $\mathbb{P}_{F(n-1,Q)}(V)$ to $F(n,Q)$.

LEMMA 5.5 *Let V be a 3-dimensional bundle over some variety X equipped with the nondegenerate quadratic form p. Let $\pi : Y \to X$ be conic bundle of p-isotropic lines in V. Then there is a $CH^*(X)$-algebra automorphism $\phi : CH^*(Y) \to CH^*(Y)$ of exponent 2 such that*

(1) $\phi(c_1(\mathcal{O}(-1)|_Y)) = c_1(\mathcal{O}(1)|_Y)$.

(2) $\pi^*\pi_*(x) \cdot c_1(\mathcal{O}(1)) = x - \phi(x)$

PROOF: Consider variety $Y \times_X Y$ with the natural projections π_1 and π_2 on the first and second factor, respectively. Then divisor $\Delta(Y) \subset Y \times_X Y$ defines an invertible sheaf $\mathcal{L}$ on $Y \times_X Y$ such that $\mathcal{L}^2 \cong \pi_1^*(\mathcal{O}(1)) \otimes \pi_2^*(\mathcal{O}(1))$ and $\Delta^*(\mathcal{L}) = \mathcal{O}(1)$. Consider the map $f := \Delta \circ \pi_2 : Y \times_X Y \to Y \times_X Y$. Define $\phi : \mathrm{CH}^*(Y) \to \mathrm{CH}^*(Y)$ as $id - \Delta^* \circ f_* \circ \pi_1^*$.
The described maps fit into the diagram:

$$\begin{tikzcd}
Y & Y \times_X Y \arrow[l, "\pi_1"'] \arrow[d, "\pi_2"] & \\
X \arrow[from=u, "\pi"'] & Y \arrow[l, "\pi"] \arrow[r, "\Delta"'] & Y \times_X Y
\end{tikzcd}$$

with the transversal Cartesian square ($\pi_1^*(T_\pi) = T_{\pi_2}$). Consequently, $\pi^* \circ \pi_* = \pi_{2*} \circ {\pi_1}^*$. Since $\mathcal{O}(\Delta(Y))|_Y = \mathcal{O}(1)$, we have:

$$\pi^*\pi_*(x) \cdot c_1(\mathcal{O}(1)) = \Delta^*\Delta_*\pi_{2*}\pi_1^*(x) = x - \phi(x).$$

Consider the map $\psi := id - f_* : \mathrm{CH}^*(Y \times_X Y) \to \mathrm{CH}^*(Y \times_X Y)$. We claim that ψ is a ring homomorphism. Really, $Y \times_X Y \cong \mathbb{P}_Y(U)$, where the projection $\mathbb{P}_Y(U) \to Y$ is given by π_2 and $c(U)(t) = t(t - c_1(\mathcal{O}(1)))$. Thus $\mathrm{CH}^*(Y \times_X Y) = \mathrm{CH}^*(Y)[\rho]/(\rho(\rho - c_1(\mathcal{O}(1))))$, where the map $\mathrm{CH}^*(Y) \to \mathrm{CH}^*(Y \times_X Y)$ is π_2^*. Notice, that $f_*\pi_2^* = \Delta_*\pi_{2*}\pi_2^* = 0$, that is, $\psi|_{\mathrm{CH}^*(Y)}$ is the identity. At the same time, $\psi(\rho) = \rho - \rho = 0$. Since $\mathrm{CH}^*(Y \times_X Y)$ is free $\mathrm{CH}^*(Y)$-module of rank 2 with the basis 1, ρ, by the projection formula, we get that ψ is an endomorphism of $\mathrm{CH}^*(Y \times_X Y)$ considered as an $\mathrm{CH}^*(Y)$-algebra.

Since, $\phi = \Delta^* \circ \psi \circ \pi_1^*$, it is a homomorphism of $\mathrm{CH}^*(X)$-algebras. Also, $\phi(c_1(\mathcal{O}(1))) = -c_1(\mathcal{O}(1))$. Finally, since the composition $\pi_{2*}\pi_1^*\Delta^*\Delta_* : \mathrm{CH}^*(Y) \to \mathrm{CH}^*(Y)$ is equal $2 \cdot id$, we get

$$(\Delta^* \circ f_* \circ \pi_1^*)^{\circ 2} = 2(\Delta^* \circ f_* \circ \pi_1^*),$$

which is equivalent to: $\phi^{\circ 2} = id$. Thus, ϕ is an automorphism of exponent 2. □

Let us compute the action of ϕ on basis elements $\overline{z}_I$. Let σ_i be elementary symmetric functions in h_i's. Since $h_i \in CH^*(F(n-1, Q))$, for $i < n$, we have equality $\phi(h_i) = h_i$ for them, and $\phi(h_n) = -h_n$. We know that $\sigma_i = (-1)^i 2z_i$. We immediately conclude:

Lemma 5.6 $\phi(z_i) = z_i + \sum_{0<l<i} 2z_{i-l}h_n^l + h_n^i$.

Lemma 5.7 $\phi(\overline{z}_I) = \overline{z}_I + \sum_{i \in I} \overline{z}_{(I\setminus i)}\overline{h}_n^i$.

Proof: Let us define the *size* $s(I)$ of I as the number of it's elements. Use induction on the size of I. The case of size $= 1$ is OK by the previous lemma. Suppose the statement is known for sizes $< s(I)$.

Let i be some element of I. We know from Proposition 3.3 that $\overline{z}_I = \overline{z}_{(I\setminus i)} \cdot$

$\overline{z}_i+\sum_{j\in I, j\neq i}\overline{z}_{(I\backslash\{i,j\})\cup(i+j)}$. Since ϕ is a ring homomorphism, we get:

$$\phi(\overline{z}_I)=\phi(\overline{z}_{(I\backslash i)})\cdot\phi(\overline{z}_i)+\sum_{j\in I\backslash i}\phi(\overline{z}_{(I\backslash\{i,j\})\cup(i+j)})=$$

$$\left(\overline{z}_{(I\backslash i)}+\sum_{l\in I\backslash i}\overline{z}_{(I\backslash\{i,l\})}h_n^l\right)\cdot(\overline{z}_i+\overline{h}_n^i)+$$

$$\sum_{j\in I\backslash i}\left(\overline{z}_{(I\backslash\{i,j\})\cup(i+j)}+\overline{z}_{(I\backslash\{i,j\})}\overline{h}_n^{i+j}+\sum_{m\in I\backslash\{i,j\}}\overline{z}_{(I\backslash\{i,j,m\})\cup(i+j)}\overline{h}_n^m\right)=$$

$$\overline{z}_I+\overline{z}_{(I\backslash i)}\overline{h}_n^i+\sum_{l\in I\backslash\{i\}}(\overline{z}_{(I\backslash\{i,l\})}\cdot\overline{z}_i)\overline{h}_n^l+\sum_{m\neq j\in I\backslash\{i\}}\overline{z}_{(I\backslash\{i,j,m\})\cup(i+j)}\overline{h}_n^m=$$

$$\overline{z}_I+\overline{z}_{(I\backslash i)}\overline{h}_n^i+\sum_{j\in I\backslash i}\overline{z}_{(I\backslash j)}\overline{h}_n^j+2\cdot\sum_{m\neq j\in I\backslash\{i\}}\overline{z}_{(I\backslash\{i,j,m\})\cup(i+j)}\overline{h}_n^m=$$

$$\overline{z}_I+\sum_{j\in I}\overline{z}_{(I\backslash j)}\overline{h}_n^j$$

(as usually, one should omit $\overline{z}_J$ with $J\not\subset\{1,\ldots,n\}$). □

Let $p=q\perp\mathbb{H}$. Then Q can be identified with the quadric of projective lines on P passing through fixed rational point y. This identifies the complete flag variety $F(r,Q)$ with the subvariety of $F(r+1,P)$ consisting of flags containing our point y. We get an embedding $i_r:F(r,Q)\to F(r+1,P)$. It is easy to see that the diagram

$$\begin{array}{ccc} F(n,Q) & \xrightarrow{i_n} & F(n+1,P)\\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi'}\\ F(n-1,Q) & \xrightarrow[i_{n-1}]{} & F(n,P)\end{array}$$

is Cartesian, and since π' is smooth, we have an equality: $\pi^*\circ\pi_*\circ i_n^*=i_n^*\circ{\pi'}^*\circ\pi'_*$.
It follows from Lemmas 5.5 and 5.7 that

$${\pi'}^*\pi'_*(\overline{z}_I)\cdot\overline{h}_{n+1}=\sum_{i\in I}\overline{z}_{(I\backslash i)}\overline{h}_{n+1}^i.$$

Thus, modulo the kernel of multiplication by $\overline{h}_{n+1}$, ${\pi'}^*\pi'_*(\overline{z}_I)\equiv\sum_{i\in I}\overline{z}_{(I\backslash i)}\overline{h}_{n+1}^{i-1}$. But, by the Statement 5.3 and Theorem 5.1, such kernel is generated by $\overline{h}_{n+1}^n$.
Since $i_n^*(\overline{z}_I)=\overline{z}_I$, $i_n^*(\overline{h}_{n+1})=\overline{h}_n$ and $\overline{h}_n^n=0$ on $F(n,Q)$, we get:

$$\pi^*\pi_*(\overline{z}_I)=i_n^*{\pi'}^*\pi'_*(\overline{z}_I)=\sum_{i\in I}\overline{z}_{(I\backslash i)}\overline{h}_n^{i-1}=\sum_{i\in I}\overline{z}_{(I\backslash i)}\pi^*\pi_*(\overline{z}_i).$$

□

Notice, that the elements $\pi^*\pi_*(z_i)$ belong to $\mathrm{CH}^*(F_n)/2$, and they are linearly independent (being nonzero and having different degrees).
As a corollary, we get:

Main Theorem 5.8 *As an algebra, $C^*(G(n,Q))$ is generated by the elementary classes $\overline{z}_i$ contained in it.*

Proof: Let $\overline{z}$ be an element of $C^*(G(n,Q))$. It can be expressed as a linear combination of the basis elements $\overline{z}_I$'s. Let us define the *size* $s(\overline{z})$ of the element $\overline{z} = \sum \overline{z}_{I_a}$ as the maximum of sizes of I_a involved. Let $m(\overline{z})$ be the main term of $\overline{z}$, that is, $\sum_{a:s(I_a)=s(\overline{z})} \overline{z}_{I_a}$.

Lemma 5.9 *Let $\overline{z} = \sum_a \overline{z}_{I_a} \in C^*(G(n,Q))$. Let $s(I_a) = s(\overline{z})$, and $i \in I_a$. Then the elementary cycle $\overline{z}_i$ belongs to $C^*(G(n,Q))$.*

Proof: Let $i \in I_a$, and $I_a \backslash i = \{j_2, \dots, j_s\}$. Denote the operation $\pi^*\pi_*$ as D. Then $D(\overline{z}) = \sum_{1 \leqslant j \leqslant n} d_j(\overline{z}) \cdot D(\overline{z}_j)$, where $d_j(\overline{z}) \in \mathrm{CH}^*(G(n,Q)|_{\overline{k}})/2$, and the elements $D(\overline{z}_j) \in \mathrm{CH}^*(F_n)/2$ are linearly independent. Since D is defined over the base field, $D(\overline{z}) \in C^*(F(n,Q))$, and, by the Statement 5.2, $d_j(\overline{z}) \in C^*(G(n,Q))$. Clearly, $m(d_j(\overline{z})) = d_j(m(\overline{z}))$. It is easy to see that $d_{j_s} \dots d_{j_2}(\overline{z}) = \overline{z}_i$, since for arbitrary I_b with $s(I_b) < s = s(\overline{z})$ we have: $d_{j_s} \dots d_{j_2}(\overline{z}_{I_b}) = 0$, or 1, and for $I_c \neq I_a$ with $s(I_c) = s$, $d_{j_s} \dots d_{j_2}(\overline{z}_{I_c})$ is either 0, or has degree different from i. Thus, $\overline{z}_i \in C^*(G(n,Q))$. □

Let us prove by induction on the size of $\overline{z}$, that $\overline{z}$ belongs to the subring of $C^*(G(n,Q))$ generated by $\overline{z}_j$'s. The base of induction, $s = 1$ is trivial. Notice that $m(\prod_{i \in I} \overline{z}_i) = \overline{z}_I$. Thus, the size of $\overline{z}' = \overline{z} - \sum_{a:s(I_a)=s(\overline{z})} \prod_{i \in I_a} \overline{z}_i$ is smaller than that of $\overline{z}$. But by the Lemma 5.9, all the $\overline{z}_i$'s appearing in this expression belong to $C^*(G(n,Q))$. By the inductive assumption, $\overline{z}'$ belongs to the subring of $C^*(G(n,Q))$ generated by $\overline{z}_j$'s. Then so is $\overline{z}$. □

Remark. Actually, for the proof of the Main Theorem one just needs the statement of the Lemma 5.5(1).

Corollary 5.10 *For arbitrary smooth projective quadric Q,*

$$C^*(G(n,Q)) = \bigotimes_{1 \leqslant d \leqslant n;\, d-odd} (\mathbb{Z}/2[\overline{z}_{d2^{l_d}}]/(\overline{z}_{d2^{l_d}}^{2^{(m_d - l_d)}})),$$

for certain $0 \leqslant l_d \leqslant m_d = [log_2(n/d)] + 1$.

Proof: It immediately follows from the Main Theorem 5.8, Proposition 3.1, and the fact that $\overline{z}_s^2 = \overline{z}_{2s}$ (or 0, if $2s > n$). □

Now we can introduce:

Definition 5.11 (1) *Let Q be a quadric of dimension $2n-1$. Denote as $J(Q)$ the subset of $\{1,\ldots,n\}$ consisting of those i, for which $\overline{z}_i \in C^*(G(n,Q))$.*

(2) *Let P be a quadric of dimension $2n$. Let Q be arbitrary subquadric of codimension 1 in P. Then $J(P)$ is a subset of $\{0,1,\ldots,n\}$, where $0 \in J(P)$ iff $det_{\pm}(P)=1$ and, for $i>0$, $i \in J(P)$ iff $i \in J(Q|_{k\sqrt{det_{\pm}(P)}})$.*

Remark: The definition (2) above is motivated by the fact that $G(n+1,P)$ is isomorphic to $G(n,Q) \times_{\mathrm{Spec}(k)} \mathrm{Spec}(k\sqrt{det_{\pm}(P)})$.

It follows from the Main Theorem 5.8 that $C^*(G(n,Q))$ is exactly the subring of $\mathrm{CH}^*(G(n,Q)|_{\overline{k}})/2$ generated by $\overline{z}_i$, $i \in J(Q)$. In particular, $J(Q)$ carries all the information about $C^*(G(n,Q))$. Notice, that the same information is contained in the sequence $\{l_d\}_{d-\mathrm{odd};1\leqslant d\leqslant n}$.
What restrictions do we have on the possible values of $J(Q)$? Because of the action of the Steenrod operations, we get:

Proposition 5.12 *Let $i \in J(Q)$, and $r \in \mathbb{N}$ is such that $\binom{i}{r} \equiv 1\,(mod\,2)$, and $i+r \leqslant n$. Then $(i+r) \in J(Q)$.*

Proof: j belongs to $J(Q)$ if and only if the cycle $\overline{z}_j \in \mathrm{CH}^j(G(n,Q)|_{\overline{k}})/2$ is defined over the base field. Since $\overline{z}_i$ has such a property, and the Steenrod operation S^r is defined over the base field too, we get: $\overline{z}_{(i+r)} = \mathrm{S}^r(\overline{z}_i)$ is also defined over the base field. □

The natural question arises:

Question 5.13 *Do we have other restrictions on $J(Q)$? In other words, let $J \subset \{1,\ldots,n\}$ be a subset satisfying the conditions of Proposition 5.12. Does there exist a quadric such that $J(Q)=J$?*

It is not difficult to check that, at least, for $n \leqslant 4$, there is no other restrictions.

6 On the canonical dimension of quadrics

In this section we will show that in the case of a *generic* quadric the variety $G(n,Q)$ is 2-incompressible, and also will formulate the conjecture describing the *canonical dimension* of arbitrary quadric. I would like to point out that the current section would not appear without the numerous discussions with G.Berhuy, who brought this problem to my attention.
We start by computing the characteristic classes of the variety $G(n,Q)$.
Let W be $(2n+1)$-dimensional vector space over k equipped with the non-degenerate quadratic form q. Let $F(r)=F(r,Q)$ be the variety of complete flags $(\pi_1 \subset \ldots \subset \pi_r)$ of totally isotropic subspaces in W. Thus,

$F(0) = \mathrm{Spec}(k)$, $F(1) = Q$, etc. ... . We get natural smooth projective maps: $\varepsilon_r : F(r+1) \to F(r)$ with fibers - quadrics of dimension $2n-2r-1$. Let $\mathcal{L}_i$ be the standard subquotient linear bundles $\mathcal{L}_i := \pi_i/\pi_{i-1}$, and $h_i = c_1(\mathcal{L}_i)$. The bundle $\mathcal{L}_i$ is defined on $F(r)$, for $r \geqslant i$. These divisors h_i are the roots of the tautological vector bundle V_n studied above.

Proposition 6.1 *The Chern polynomial of the tangent bundle $T_{F(r)}$ is equal to:*

$$c(T_{F(r)}) = \frac{\prod_{1\leqslant i\leqslant r}(1-h_i)^{2n+1}}{\prod_{1\leqslant i\leqslant r}(1-2h_i)\cdot\prod_{1\leqslant i<j\leqslant r}\left((1-h_j+h_i)\cdot(1-h_j-h_i)\right)}.$$

Proof: Let V_r be a tautological vector bundle on $F(r)$. Then V_r is an isotropic subbundle of $\eta^*(W)$ ($\eta : F(r) \to \mathrm{Spec}(k)$ is the projection), and on the subquotient $W_r := V_r^{\perp}/V_r$ we have a nondegenerate quadratic form $q_{\{r\}}$. Then the variety $F(r+1)$ is defined as zeroes of this quadratic form. Thus, $F(r+1)$ is the divisor of the sheaf $\mathcal{O}(2)$ on the projective bundle $\mathbb{P}_{F(r)}(W_r)$, and we have exact sequence:

$$0 \to T_{F(r+1)} \to T_{\mathbb{P}_{F(r)}(W_r)}|_{F(r+1)} \to \mathcal{O}(2)|_{F(r+1)} \to 0.$$

On the other hand, from the projection $\mathbb{P}_{F(r)}(W_r) \xrightarrow{\theta_r} F_r$, we have sequences:

$$0 \to T_{\theta_r} \to T_{\mathbb{P}_{F(r)}(W_r)} \to \theta_r^*(T_{F(r)}) \to 0 \text{ and}$$

$$0 \to \mathcal{O} \to W_r \otimes \mathcal{L}_{r+1}^{-1} \to T_{\theta_r} \to 0.$$

(see [4, Example 3.2.11]). It remains to notice, that in K_0, $[W_r] = (2n+1)[\mathcal{O}] - \sum_{i=1}^{r}([\mathcal{L}_i] + [\mathcal{L}_i^{-1}])$, to get the equality:

$$c(T_{F(r+1)}) = \varepsilon_r^*(c(T_{F(r)}))\cdot\frac{(1-h_{r+1})^{2n+1}}{(1-2h_{r+1})\cdot\prod_{i=1}^{r}(1-h_{r+1}+h_i)(1-h_{r+1}-h_i)}$$

The statement now easily follows by induction on r. □

Now, it is easy to compute the characteristic classes of the quadratic Grassmannians.

Proposition 6.2

$$c(T_{G(r)}) = \frac{\prod_{1\leqslant i\leqslant r}(1-h_i)^{2n+1}}{\prod_{1\leqslant i\leqslant r}(1-2h_i)\cdot\prod_{1\leqslant i<j\leqslant r}\left((1-(h_j-h_i)^2)\cdot(1-h_j-h_i)\right)},$$

where h_j are the roots of the tautological vector bundle V_r on $G(r)$.

Proof: Consider the (forgetting) projection $\delta_r : F(r) \to G(r) = G(r, Q)$. We have natural identification of $F(r)$ with the variety of complete flags corresponding to the tautological bundle V_r on $G(r)$ (we will permit ourselves to use the same notation for the tautological bundles on $G(r)$ and $F(r)$ - this is justified by the fact that they are related by the map δ_r^*). Using the fact that $\delta_r : F(r) \to G(r)$ can be decomposed into a tower of projective bundles, and [4, Example 3.2.11], we get:

$$c(T_{\delta_r}) = \prod_{1 \leqslant i < j \leqslant r} (1 + h_j - h_i),$$

and the statement follows. □

Now we can prove the following Conjecture of G.Berhuy (proven by him for $n \leqslant 4$):

Theorem 6.3 $degree(c_{dim(G(n))}(-T_{G(n)})) \equiv 2^n \ (mod\, 2^{n+1})$.

Proof: By Proposition 6.2, Chern classes of $(-T_{G(n)})$ can be expressed as polynomials in the Chern classes of the tautological vector bundle V_n. From Theorem 2.5 we know that $c_j(V_n) = \sigma_j = (-1)^j 2 z_j$, where z_j are elementary cycles defined in Proposition 2.4.
Since, in K_0, $[V_n] + [V_n^\vee] = 2n[\mathcal{O}]$, we get the relations on σ_j:

Lemma 6.4 $\sigma_i^2 = 2(-1)^i(\sigma_{2i} + \sum_{1 \leqslant j < i} (-1)^j \sigma_j \cdot \sigma_{2i-j})$.

Proof: It is just the component of degree $2i$ of the relation

$$\left(1 + \sum_i \sigma_i\right) \cdot \left(1 + \sum_i (-1)^i \sigma_i\right) = 1.$$

□

Let $A := \mathbb{Z}[\tilde{\sigma}_1, \dots, \tilde{\sigma}_n]$. We have ring homomorphism $\psi : A \to \mathrm{CH}^*(G(n))$ sending $\tilde{\sigma}_i$ to σ_i. It follows from the Lemma 6.4, that for arbitrary $f \in A$ there exists some $g \in A$ such that g does not contain squares, and $\psi(f - g) \in 2^{n+1}\,\mathrm{CH}^*(G(n))$. If f has degree $= dim(G(n))$, then g got to be monomial $\lambda \cdot \prod_{1 \leqslant i \leqslant n} \sigma_i$. Moreover, if f was a monomial divisible by 2, or containing square, then λ will be divisible by 2. Consider ideal $L \subset A$ generated by 2 and squares of elements of positive degree. Let R be a quotient ring, and $\varphi : A \to R$ be the projection.
Since $\prod_{1 \leqslant i \leqslant n} \sigma_i = (-1)^{\binom{n+1}{2}} 2^n \prod_{1 \leqslant i \leqslant n} z_i$, and $\prod_{1 \leqslant i \leqslant n} z_i$ is the class of a rational point (by Proposition 2.4), we get that for arbitrary $f \in A$, the $degree(\psi(f))$ is divisible by 2^n, and for $f \in L$ the degree is divisible by 2^{n+1}. Thus, modulo 2^{n+1}, the degree of $\psi(f)$ depends only on $\varphi(f)$.

In R we have the following equalities:

$$\varphi(1+f^2)=\varphi((1+f)^2)=1,\ \text{ for any } f \text{ of positive degree.}$$

Thus,

$$\varphi(c(-T_{G(n)}))=\varphi\left(\prod_{1\leqslant i\leqslant n}(1+h_i)\cdot\prod_{1\leqslant i<j\leqslant n}(1+h_i+h_j)\right).$$

And in the light of Giambelli's formula (see [4, Example 14.5.1(b')]),

$$\varphi\left(c(-T_{G(n)})_{\binom{n+1}{2}}\right)=\varphi\left(\sigma_n\cdot\det[\sigma_{n-2i+j}]_{1\leqslant i,j\leqslant n}\right).$$

Since we mod-out the squares, this expression is equal to $\varphi(\prod_{i=1}^{n}\sigma_i)$. Consequently, $degree(c_{dim(G(n))}(-T_{G(n)}))\equiv 2^n\ (\,mod\, 2^{n+1})$.

□

We recall from [10] that a variety X is p-compressible if there is a rational map $X \dashrightarrow Y$ to some variety Y such that $dim(Y)<dim(X)$ and $v_p(n_X)\leqslant v_p(n_Y)$, where n_Z is the image of the degree map $deg:\mathrm{CH}_0(Z)\to\mathbb{Z}$.
From the Rost degree formula ([10, Theorem 6.4]) for the characteristic number $c_{dim(G(n))}$ modulo 2 (see [10, Corollary 7.3, Proposition 7.1]), we get:

Proposition 6.5 *Let Q be a smooth $2n+1$-dimensional quadric, all splitting fields of which have degree divisible by 2^n (we call such Q - generic). Then the variety $G(n,Q)$ is 2-incompressible.*

□

Call two smooth varieties X and Y equivalent if there are rational maps $X \dashrightarrow Y$ and $Y \dashrightarrow X$. Then let $d(X)$ be the minimal dimension of a variety equivalent to X. Recall from [1] that a *canonical dimension* $cd(q)$ of a quadratic form q is defined as $d(G(n,Q))$, where $n=[dim(q)/2]+1$.
Proposition 6.5 gives another proof of the fact that the canonical dimension of a generic $(2n+1)$-dimensional form is $n(n+1)/2$, which computes the canonical dimension of the groups SO_{2n+1} and SO_{2n+2} (cf. [8, Theorem 1.1, Remark 1.3]).
Our computations of the generic discrete invariant $GDI(m,Q)$ permit to conjecture the answer in the case of arbitrary smooth quadric Q:

Conjecture 6.6 *Let Q be smooth projective quadric of dimension d. Then*

$$cd(Q)=\sum_{j\in\{1,\dots,[d+1/2]\}\backslash J(Q)} j,$$

where $J(Q)$ is the invariant from the Definition 5.11.

If Q is generic, then $J(Q)$ is empty, and $cd(Q)$ is indeed equal $\sum_{1\leqslant i\leqslant n} i = n(n+1)/2$.

References

[1] G. Berhuy, Z. Reichstein, *On the notion of canonical dimension for algebraic groups*, Linear Algebraic Groups and Related Structures, 140 (June 18, 2004), 36p. (http://www.mathematik.uni-bielefeld.de/LAG/)

[2] I. N. Bernstein, I. M. Gelfand, S. I. Gelfand, *Schubert cells and cohomology of the spaces G/P*, Russian Math. Surv., 28 (1973), n.3, 1-26.

[3] P. Brosnan, *Steenrod operations in Chow theory.*, Trans. Amer. Math. Soc. 355 (2003), no 5, 1869-1903.

[4] W. Fulton, *Intersection Theory, 2-nd edition*, Springer-Verlag, Berlin, 1998.

[5] H. Hiller, *Combinatorics and the intersections of Schubert varieties*, Comm. Math. Helv., 57 (1982), 41-59.

[6] H. Hiller, B. Boe, *Pieri formula for SO_{2n+1}/U_n and Sp_n/U_n.*, Adv. in Math. 62 (1986) 49-67.

[7] J. Hurrelbrink, U. Rehmann, *Splitting patterns of excellent quadratic forms.*, J. Reine Angew. Math. 444 (1993), 183-192.

[8] N. Karpenko, *Canonical dimension of orthogonal groups.*, Linear Algebraic Groups and Related Structures, 144 (July 26 2004), 4p. (http://www.mathematik.uni-bielefeld.de/LAG/)

[9] M. Knebusch, *Generic splitting of quadratic forms, I.*, Proc. London Math. Soc. (3) 33 (1976), 65-93.

[10] A. Merkurjev, *Steenrod operations and degree formulas.*, J. Reine Angew. Math. 565 (2003), 13-26.

[11] P. Pragacz, J.Ratajski, *A Pieri-type theorem for Lagrangian and odd Orthogonal Grassmannians.*, J. reine angew. Math. 476 (1996), 143-189.

[12] A. Vishik, *Motives of quadrics with applications to the theory of quadratic forms.*, in the Proceedings of the Summer School Geometric Methods in the Algebraic Theory of Quadratic Forms, Lens 2000 (ed. J.-P. Tignol), Lect. Notes in Math., 1835 (2004), 25-101.

[13] V. Voevodsky, *Reduced power operations in motivic cohomology*, Publ. Math. Inst. Hautes Études Sci. 98 (2003) 1-57.

On the Torsion of the Mordell-Weil Group of the Jacobian of Drinfeld Modular Curves

Ambrus Pál

Received: September 27, 2004
Revised: February 2, 2005

Communicated by Peter Schneider

Abstract. Let $Y_0(\mathfrak{p})$ be the Drinfeld modular curve parameterizing Drinfeld modules of rank two over $\mathbb{F}_q[T]$ of general characteristic with Hecke level $\mathfrak{p}$-structure, where $\mathfrak{p} \triangleleft \mathbb{F}_q[T]$ is a prime ideal of degree d. Let $J_0(\mathfrak{p})$ denote the Jacobian of the unique smooth irreducible projective curve containing $Y_0(\mathfrak{p})$. Define $N(\mathfrak{p}) = \frac{q^d-1}{q-1}$, if d is odd, and define $N(\mathfrak{p}) = \frac{q^d-1}{q^2-1}$, otherwise. We prove that the torsion subgroup of the group of $\mathbb{F}_q(T)$-valued points of the abelian variety $J_0(\mathfrak{p})$ is the cuspidal divisor group and has order $N(\mathfrak{p})$. Similarly the maximal μ-type finite étale subgroup-scheme of the abelian variety $J_0(\mathfrak{p})$ is the Shimura group scheme and has order $N(\mathfrak{p})$. We reach our results through a study of the Eisenstein ideal $\mathfrak{E}(\mathfrak{p})$ of the Hecke algebra $\mathbb{T}(\mathfrak{p})$ of the curve $Y_0(\mathfrak{p})$. Along the way we prove that the completion of the Hecke algebra $\mathbb{T}(\mathfrak{p})$ at any maximal ideal in the support of $\mathfrak{E}(\mathfrak{p})$ is Gorenstein.

2000 Mathematics Subject Classification: Primary 11G18; Secondary 11G09.
Keywords and Phrases: Drinfeld modular curves, Eisenstein ideal.

1. Introduction

Notation 1.1. Let $F = \mathbb{F}_q(T)$ denote the rational function field of transcendence degree one over a finite field $\mathbb{F}_q$ of characteristic p, where T is an indeterminate, and let $A = \mathbb{F}_q[T]$. For any non-zero ideal $\mathfrak{n}$ of A a geometrically irreducible affine algebraic curve $Y_0(\mathfrak{n})$ is defined over F, the Drinfeld modular curve parameterizing Drinfeld modules of rank two over A of general characteristic with Hecke level $\mathfrak{n}$-structure. There is a unique non-singular projective curve $X_0(\mathfrak{n})$ over F which contains $Y_0(\mathfrak{n})$ as an open subvariety. Let $J_0(\mathfrak{n})$ denote the Jacobian of the curve $X_0(\mathfrak{n})$. Let $\mathfrak{p}$ be a prime ideal of A and let d denote the degree of the residue field of $\mathfrak{p}$ over $\mathbb{F}_q$. Define $N(\mathfrak{p}) = \frac{q^d-1}{q-1}$, if d is odd, and define $N(\mathfrak{p}) = \frac{q^d-1}{q^2-1}$, otherwise.

Theorem 1.2. *The torsion subgroup $\mathcal{T}(\mathfrak{p})$ of the group of F-valued points of the abelian variety $J_0(\mathfrak{p})$ is a cyclic group of order $N(\mathfrak{p})$.*

It is possible to explicitly determine the group in the theorem above.

Definition 1.3. The geometric points of the zero dimensional complement of $Y_0(\mathfrak{n})$ in $X_0(\mathfrak{n})$ are called cusps of the curve $X_0(\mathfrak{n})$. They are actually defined over F. Since we assumed that $\mathfrak{p}$ is a prime the curve $X_0(\mathfrak{p})$ has two cusps. The cyclic group generated by the divisor which is the difference of the two cusps is called the cuspidal divisor group and it is denoted by $\mathcal{C}(\mathfrak{p})$.

Theorem 1.4. *The group $\mathcal{T}(\mathfrak{p})$ is equal to $\mathcal{C}(\mathfrak{p})$.*

Notation 1.5. The theorem above has a pair which describes the largest étale subgroup scheme of $J_0(\mathfrak{p})$ whose Cartier dual is constant. Let us introduce some additional notation in order to formulate it. Let $Y_1(\mathfrak{p})$ denote the Drinfeld modular curve parameterizing Drinfeld modules of rank two over A of general characteristic with Γ_1-type level $\mathfrak{p}$-structure. The forgetful map $Y_1(\mathfrak{p}) \to Y_0(\mathfrak{p})$ is a Galois cover defined over F with Galois group $(A/\mathfrak{p})^*/\mathbb{F}_q^*$. Let $Y_2(\mathfrak{p}) \to Y_0(\mathfrak{p})$ denote the unique covering intermediate of this covering which is a Galois covering, cyclic of order $N(\mathfrak{p})$, and let $J_2(\mathfrak{p})$ denote the Jacobian of the unique geometrically irreducible non-singular projective curve $X_2(\mathfrak{p})$ containing $Y_2(\mathfrak{p})$. The kernel of the homomorphism $J_0(\mathfrak{p}) \to J_2(\mathfrak{p})$ induced by Picard functoriality is called the Shimura group scheme and it is denoted by $\mathcal{S}(\mathfrak{p})$. For every field K let $\overline{K}$ denote the separable algebraic closure of K. We say that a finite flat subgroup scheme of $J_0(\mathfrak{p})$ is a μ-type group scheme if its Cartier dual is a constant group scheme. If this group scheme is étale, then it is uniquely determined by the group of its $\overline{F}$-valued points. The latter group actually lies in $J_0(\mathfrak{p})(\overline{\mathbb{F}}_q(T))$, where $\overline{\mathbb{F}}_q(T)$ is the maximal everywhere unramified extension of F. Let $\mathcal{M}(\mathfrak{p})$ denote the unique maximal μ-type étale subgroup scheme of $J_0(\mathfrak{p})$.

Theorem 1.6. *The group schemes $\mathcal{M}(\mathfrak{p})$ and $\mathcal{S}(\mathfrak{p})$ are equal. In particular the former is a cyclic group scheme of rank equal to $N(\mathfrak{p})$.*

These results are proved via a detailed study of the Eisenstein ideal in the Hecke algebra of the Drinfeld modular curve $Y_0(\mathfrak{p})$, defined in [18] first in this context. In particular we prove that the completion of the Hecke algebra at any prime ideal in the support of Eisenstein ideal is Gorenstein (Corollary 10.3 and Theorem 11.6). The main goal to develop such a theory in its original setting was to classify the rational torsion subgroups of elliptic curves. Some of the methods and results of this paper can be used to give a similar classification of the rational torsion subgroups of Drinfeld modules of rank two in our setting as well, whose complete proof will appear in a forthcoming paper of the author.

Contents 1.7. Of course this work is strongly influenced by [14], where Mazur proved similar theorems for elliptic modular curves, conjectured originally by Ogg. Therefore the structure of the paper is similar to [14], although there are several significant differences, too. In the next two chapters we develop the tools necessary to study congruences between automorphic forms with respect to a modulus prime to the characteristic of F: Fourier expansions and the multiplicity one theorem. Almost everything we prove is a straightforward generalization of classical results in [19]. The main idea is that the additive group of adeles of F is a pro-p group, so it is possible to do Fourier analysis for locally constant functions taking values in a ring where p is invertible. In the fourth chapter we prove an analogue of the classical Kronecker limit formula, a result of independent interest. One motivation for this result in our setting is that it connects the Eisenstein series with the geometry of the modular curve directly. We compute the Fourier coefficients of Eisenstein series in the fifth chapter and give a new, more conceptual proof of a theorem of Gekeler on the Drinfeld discriminant function. As an application of our previous results we determine the largest sub-module $\mathcal{E}_0(\mathfrak{p}, R)$ of R-valued cuspidal harmonic forms annihilated by the Eisenstein ideal in the sixth chapter, for certain rings R. The first cases of Theorem 1.4 are proved in the seventh chapter, where we connect the geometry of the modular curve to our previous observations via the uniformization theorem of Gekeler-Reversat (see [11]). With the help of a theorem of Gekeler and Nonnengardt we show that the image of the n-torsion part of $\mathcal{T}(\mathfrak{p})$, n prime to p, in the group of connected components of the Néron model of $J_0(\mathfrak{p})$ at ∞ with respect to specialization injects into $\mathcal{E}_0(\mathfrak{p}, \mathbb{Z}/n\mathbb{Z})$ without any assumptions on $t(\mathfrak{p})$, the greatest common divisor of $N(\mathfrak{p})$ and $q-1$. We also show that there is no p-torsion using a result on the reduction of $Y_0(\mathfrak{p})$ over the prime $\mathfrak{p}$, again due to Gekeler (see [6]). Then we conclude the proof of Theorem 7.19 by showing that the exponent of the kernel of the specialization map into the group of connected components at ∞ in $\mathcal{T}(\mathfrak{p})$ is only divisible by primes dividing $t(\mathfrak{p})$. We prove some important properties of the Shimura group scheme in the eight chapter. In order to do so, we first include a section on a model $M_1(\mathfrak{p})$ of $Y_1(\mathfrak{p})$ with particular emphasis on the structure of its fiber at the prime $\mathfrak{p}$ in this chapter, as the current literature on the reduction of Drinfeld modular curves is somewhat incomplete. We study an important finite étale sub-group scheme of $J_0(\mathfrak{p})$ analogous to the Dihedral subgroup of Mazur in the next chapter. The latter is an object constructed to remedy the fact that the intersection of the cuspidal and Shimura subgroups could be non-empty. Here some of the calculations overlap with the results of [5], but the author could not resist the temptation to use the methods of chapters 4 and 5 in this setting, too. The goal of the last two chapters is to fully implement Mazur's Eisenstein decent at Eisenstein primes l. The key idea here is that considerations at the prime l in Mazur's original paper should be substituted by similar arguments at the place ∞. In particular the role of the connected-étale devissage of the l-division group of the Jacobian of the classical elliptic modular curve is played by the filtration of the l-adic Tate module of $J_0(\mathfrak{p})$ defined by the monodromy-weight spectral sequence at ∞. His arguments carry over with minor modifications, but it is interesting to note that the concept of $*$-type groups is only defined for subgroups of the Jacobian $J_0(\mathfrak{p})$, unlike in the classical case considered by Mazur, where the similar concept was absolute. The main Diophantine application of the results of these chapters are Theorem 1.6 and Theorem 1.4 in the cases not taken care of by Theorem 7.19. At the end of the paper an index of notations is included for the convenience of the reader.

Acknowledgment 1.8. I wish to thank the CICMA and the CRM for their warm hospitality and the pleasant environment they created for productive research, where the most of this article was written. I also wish to thank the IHÉS for its welcoming atmosphere, where this work was completed.

2. Fourier expansion

Definition 2.1. A topological group P is a pro-p group if it is a projective limit of finite p-groups. In other words P is a compact, Haussdorf topological group which has a basis of translates of finite index subgroups and every finite quotient is a p-group. In this paper all rings are assumed to be commutative with unity. If R is a ring, we will write $1/p \in R$ if we want to say that p is invertible in R. We will call a ring R a coefficient ring if $1/p \in R$ and R is the quotient of a discrete valuation ring $\tilde{R}$ which contains p-th roots of unity. For example every algebraically closed field of characteristic different from p is a coefficient ring. Note that the image of the p-th roots of unity of $\tilde{R}$ in R are exactly the set of p-th roots of unity of R. If R is a ring, then we say that a function $f : P \to R$ is continuous, if it continuous with respect to the discrete topology on R. This is equivalent to f being a locally constant function on P.

Lemma 2.2. *There is a unique $\mathbb{Z}\langle\frac{1}{p}\rangle$-valued function μ on the open and closed subsets of P such that*
(a) *for any disjoint disjoint open set U and V we have $\mu(U\cup V) = \mu(U)+\mu(V)$,*
(b) *for any open set U and $g \in P$ we have $\mu(U) = \mu(gU) = \mu(Ug)$,*
(c) *for every open subgroup U we have $\mu(U) = \frac{1}{|P:U|}$.*

Proof. Existence and uniqueness immediately follows from the fact that every open and closed subset of P is a pairwise disjoint union of finitely many translates of some open subgroup U. □

Definition 2.3. The function μ will be called the normalized Haar-measure on P. If R is a ring with $1/p \in R$, then for every continuous function $f : P \to R$ we define its integral with respect to μ as

$$\int_P f(x)\mathrm{d}\mu(x) = \sum_{r\in R} r\mu(f^{-1}(r)).$$

Since all but finitely many terms of the sum above are zero, the integral is well-defined.

Definition 2.4. Assume that P is abelian. We denote the set of continuous homomorphisms $\chi : P \to R^*$ by $\widehat{P}(R)$. We define for each continuous function $f : P \to R$ and $\chi \in \widehat{P}(R)$ a homomorphism:

$$\widehat{f}(\chi) = \int_P f(x)\chi^{-1}(x)\mathrm{d}\mu(x) \in R.$$

Lemma 2.5. *Assume that R is a coefficient ring and P is p-torsion. Then for each continuous function $f : P \to R$ the function $\widehat{f} : \widehat{P}(R) \to R$ is supported on a finite set and*

$$f(x) = \sum_{\chi \in \widehat{P}} \widehat{f}(\chi)\chi(x).$$

Proof. Let $\tilde{R}$ denote a discrete valuation ring whose quotient is R, just like in Definition 2.1. Since P is compact, f takes finitely many values, so there is a continuous function $\tilde{f} : P \longrightarrow \tilde{R}$ lifting f, which means that the composition of $\tilde{f}$ and the surjection $\tilde{R} \longrightarrow R$ is f. Since P is p-torsion, all continuous homomorphisms $\chi : P \to R^*$ have a unique lift $\tilde{\chi} \in \widehat{P}(\tilde{R})$. Hence it is sufficient to prove the statement for $\tilde{f}$. There is an open subgroup $U \leq P$ such that $\tilde{f}$ is U-invariant. Then for all but finitely $\chi \in \widehat{P}(\tilde{R})$ we have $\mathrm{Ker}(\chi) \not\subseteq U$ and hence $\widehat{\tilde{f}}(\chi) = 0$. The formula is then a consequence of the similar formula for the group P/U, which is well-known. □

Notation 2.6. Let F denote the function field of X, where the latter is a geometrically connected smooth projective curve defined over the finite field $\mathbb{F}_q$ of characteristic p. Let $|X|$, $\mathbb{A}$, $\mathcal{O}$ denote set of places of F, the ring of adeles of F and its maximal compact subring of $\mathbb{A}$, respectively. F is embedded canonically into $\mathbb{A}$. The group $F\backslash\mathbb{A}$ is compact, totally disconnected and it is p-torsion, hence it is a pro-p group.

Lemma 2.7. *Let R be a coefficient ring. If $\tau : F\backslash\mathbb{A} \to R^*$ is a non-trivial continuous homomorphism, then all other elements of $\widehat{F\backslash\mathbb{A}}(R)$ are of the form $x \mapsto \tau(\eta x)$ for some $\eta \in F$.*

Proof. Since $F\backslash\mathbb{A}$ is p-torsion, the image of any element of $\widehat{F\backslash\mathbb{A}}(R)$ lies in the p-th roots of unity of the ring R. This group can be identified with the subgroup of p-th roots of unity in the field of complex numbers, hence the claim follows from the same statement for complex-valued characters. □

Definition 2.8. For every divisor $\mathfrak{m}$ of X let $\mathfrak{m}$ also denote the $\mathcal{O}$-module in the ring $\mathbb{A}$ generated by the ideles whose divisor is $\mathfrak{m}$, by abuse of notation. Let $\mathfrak{n}$ be an effective divisor of X. By an R-valued automorphic form over F of level $\mathfrak{n}$ we mean a locally constant function $\phi : GL_2(\mathbb{A}) \to R$ satisfying $\phi(\gamma gkz) = \phi(g)$ for all $\gamma \in GL_2(F)$, $z \in Z(\mathbb{A})$, and $k \in \mathbb{K}_0(\mathfrak{n})$, where $Z(\mathbb{A})$ is the center of $GL_2(\mathbb{A})$, and

$$\mathbb{K}_0(\mathfrak{n}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathcal{O}) | c \equiv 0 \bmod \mathfrak{n}\}.$$

Moreover, if for all $g \in GL_2(\mathbb{A})$:

$$\int_{F\backslash\mathbb{A}} \phi(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} g)\mathrm{d}\mu(x) = 0,$$

where $\mathrm{d}\mu(x)$ is the normalized Haar measure on $F\backslash\mathbb{A}$, we call ϕ a cusp form. Let $\mathcal{A}(\mathfrak{n}, R)$ (respectively $\mathcal{A}_0(\mathfrak{n}, R)$) denote the R-module of R-valued automorphic forms (respectively cuspidal automorphic forms) of level $\mathfrak{n}$.

Notation 2.9. Let $\mathrm{Pic}(X)$ and $\mathrm{Div}(X)$ denote the Picard group and the divisor group of the algebraic curve X, respectively. For every $y \in \mathbb{A}^*$ we denote the corresponding divisor and its class in $\mathrm{Pic}(X)$ by the same symbol by abuse of notation. For any idele or divisor y let $|y|$ and $\deg(y)$ denote its normalized absolute value and degree, respectively, related by the formula $|y| = q^{-\deg(y)}$.

Proposition 2.10. *Let R be a coefficient ring and let $\tau : F\backslash\mathbb{A} \to R^*$ be a nontrivial continuous homomorphism. Then for every $\phi \in \mathcal{A}(\mathfrak{n}, R)$ there are functions $\phi^0 : \mathrm{Pic}(X) \to R$ and $\phi^* : \mathrm{Div}(X) \to R$, the latter vanishing on non-effective divisors such that*

$$\phi(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}) = \phi^0(y) + \sum_{\eta \in F^*} \phi^*(\eta y \mathfrak{d}^{-1})\tau(\eta x),$$

for all $y \in \mathbb{A}^*$ and $x \in \mathbb{A}$, where the idele $\mathfrak{d}$ is such that $\mathcal{D} = \mathfrak{d}\mathcal{O}$, where $\mathcal{D}$ is the $\mathcal{O}$-module defined as

$$\mathcal{D} = \{x \in \mathbb{A} | \tau(x\mathcal{O}) = 1\}.$$

The functions ϕ^0 and ϕ^* are called the Fourier coefficients of the automorphic form ϕ with respect to the character τ.

Proof. By the condition of Definition 2.8:

$$\phi(\begin{pmatrix} 1 & \eta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}) = \phi(\begin{pmatrix} y & x+\eta \\ 0 & 1 \end{pmatrix}) = \phi(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}),$$

for every $y \in \mathbb{A}^*$ and $\eta \in F$, so there is a expansion, by Lemma 2.5 and Lemma 2.7:

$$\phi(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}) = \sum_{\eta \in F} a(\eta, y)\tau(\eta x).$$

Since

$$\phi(\begin{pmatrix} \eta & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}) = \phi(\begin{pmatrix} \eta y & \eta x \\ 0 & 1 \end{pmatrix}) = \phi(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}),$$

for every $y \in \mathbb{A}^*$ and $\eta \in F^*$, we have $a(\kappa, \eta y) = a(\kappa\eta, y) = a(\kappa\eta y)$ for some function $a : \mathbb{A}^* \to R$.

For any $k \in \mathcal{O}^*$, $l \in \mathcal{O}$

$$\phi(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} k & l \\ 0 & 1 \end{pmatrix}) = \phi(\begin{pmatrix} yk & x+yl \\ 0 & 1 \end{pmatrix}) = \phi(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}),$$

again by the definition of automorphic forms, we have $a(ky)\tau(ly) = a(y)$, which implies that $a(y)$ only depends on the divisor of y and $a(y)$ is nonzero only if y is in $\mathcal{D}$. A similar argument gives the existence of ϕ^0. $\square$

It is worth noting that this notion of Fourier coefficients coincides with the classical one when both are defined. Also note that when R contains $1/p$, then the constant Fourier coefficients $\phi^0(\cdot)$ are still defined.

Notation 2.11. For any valuation v of F we will let F_v, $\mathbf{f}_v$ and $\mathcal{O}_v$ to denote the corresponding completion of F, its constant field, or its discrete valuation ring, respectively. For any idele, adele, adele-valued matrix or function defined on the above which decomposes as an infinite product of functions defined on the individual components the subscript v will denote the v-th component. Similar convention will be applied to subsets of adeles and adele-valued matrices. Let B denote the group scheme of invertible upper triangular two by two matrices. Let P denote the group scheme of invertible upper triangular two by two matrices with 1 on the lower right corner. Let U denote the group scheme of invertible upper triangular two by two matrices with ones on the diagonal.

Lemma 2.12. *Every $\phi \in \mathcal{A}(\mathfrak{n}, R)$ is uniquely determined by its restriction to $P(\mathbb{A})$.*

Proof. It is sufficient to prove that $GL_2(F)B(\mathbb{A})$ is dense in $GL_2(\mathbb{A})$, as we can determine the values of ϕ on that set from the values of ϕ on $P(\mathbb{A})$, by Definition 2.8. This property is equivalent to the fact that $GL_2(\mathbb{A}) = GL_2(F)B(\mathbb{A})\mathbb{K}$ for every compact, open subgroup $\mathbb{K} = \prod_{v\in|X|}\mathbb{K}_v$. Take any element g of $GL_2(\mathbb{A})$. There is a finite set S of places such that if $\mathbb{K}_v$ is not $GL_2(\mathcal{O}_v)$, then $s \in S$. As the natural image of $GL_2(F)$ in $\prod_{v\in S} GL_2(F_v)$ is dense, there is a $\gamma \in GL_2(F)$ such that the v-component of $\gamma^{-1}g$ is in $\mathbb{K}_v$ for all $v \in S$. But $\gamma^{-1}g$ is in $B(F_v)\mathbb{K}_v = B(F_v)GL_2(\mathcal{O}_v)$ for all other v by the Iwasawa decomposition, so the claim above follows. $\square$

Proposition 2.13. *If R is a coefficient ring, every $\phi \in \mathcal{A}_0(\mathfrak{n}, R)$ ($\phi \in \mathcal{A}(\mathfrak{n}, R)$) is uniquely determined by the function ϕ^* (by the functions ϕ^* and ϕ^0).*

Proof. By Lemma 2.12, ϕ is uniquely determined by its restriction to $P(\mathbb{A})$, hence it is uniquely determined by the functions ϕ^* and ϕ^0. If ϕ is a cusp form then ϕ^0 is identically zero, hence ϕ is uniquely determined by the function ϕ^* alone. $\square$

3. Multiplicity one

Definition 3.1. Let $\mathfrak{m}$, $\mathfrak{n}$ be effective divisors of X. Define the set:

$$H(\mathfrak{m},\mathfrak{n}) = \\ \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathbb{A})|a,b,c,d \in \mathcal{O}, (ad-cb)=\mathfrak{m}, \mathfrak{n} \supseteq (c), (d)+\mathfrak{n}=\mathcal{O}\}.$$

The set $H(\mathfrak{m},\mathfrak{n})$ is compact and it is a double $\mathbb{K}_0(\mathfrak{n})$-coset, so it is a disjoint union of finitely many right $\mathbb{K}_0(\mathfrak{n})$-cosets. Let $R(\mathfrak{m},\mathfrak{n})$ be a set of representatives of these cosets. For any $\phi \in \mathcal{A}(\mathfrak{n},R)$ (or more generally, for any right $\mathbb{K}_0(\mathfrak{n})$-invariant R-valued function) define the function $T_{\mathfrak{m}}(\phi)$ by the formula:

$$T_{\mathfrak{m}}(\phi)(g) = \sum_{h \in R(\mathfrak{m},\mathfrak{n})} \phi(gh).$$

It is easy to check that $T_{\mathfrak{m}}(\phi)$ is independent of the choice of $R(\mathfrak{m},\mathfrak{n})$ and $T_{\mathfrak{m}}(\phi) \in \mathcal{A}(\mathfrak{n},R)$ as well. So we have an R-linear operator $T_{\mathfrak{m}} : \mathcal{A}(\mathfrak{n},R) \to \mathcal{A}(\mathfrak{n},R)$.

Lemma 3.2. *Let R be a coefficient ring. Then for every $\phi \in \mathcal{A}(\mathfrak{n},R)$ and $\mathfrak{m}$ ideal*

$$T_{\mathfrak{m}}(\phi)^*(\mathfrak{r}) = \sum_{\substack{\mathfrak{c}+\mathfrak{n}=\mathcal{O} \\ \mathfrak{r}+\mathfrak{m}\subseteq\mathfrak{c}}} \frac{|\mathfrak{c}|}{|\mathfrak{m}|}\phi^*(\frac{\mathfrak{r}\mathfrak{m}}{\mathfrak{c}^2}).$$

Proof. One particular choice of the representative system is

$$R(\mathfrak{m},\mathfrak{n}) = \{\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}|(a,d) \in S,\ b \in S(a)\},$$

where S is a $\mathcal{O}^* \times \mathcal{O}^*$-representative system to all pairs $(a,d) \in \mathcal{O}\times\mathcal{O}$ such that $(ad)=\mathfrak{m}$ and $(d)+\mathfrak{n}=\mathcal{O}$, and for each $a \in \mathcal{O}$ the set $S(a)$ is a representative system of the cosets of the ideal (a) in $\mathcal{O}$. For any adele $y \in \mathcal{O}$:

$$\begin{aligned} T_{\mathfrak{m}}(\phi)^*(y) &= \int_{F\backslash\mathbb{A}} T_{\mathfrak{m}}(\phi)(\begin{pmatrix} y\mathfrak{d} & x \\ 0 & 1 \end{pmatrix})\tau(-x)\mathrm{d}\mu(x) \\ &= \sum_{\substack{(a,d)\in S \\ b\in S(a)}} \int_{F\backslash\mathbb{A}} \phi(\begin{pmatrix} y\mathfrak{d}a & y\mathfrak{d}b+dx \\ 0 & d \end{pmatrix})\tau(-x)\mathrm{d}\mu(x) \\ &= \sum_{(a,d)\in S} \sum_{b\in S(a)} \tau(y\mathfrak{d}b/d) \int_{F\backslash\mathbb{A}} \phi(\begin{pmatrix} y\mathfrak{d}a/d & x \\ 0 & 1 \end{pmatrix})\tau(-x)\mathrm{d}\mu(x) \\ &= \sum_{(a,d)\in S} \phi^*(ya/d) \sum_{b\in S(a)} \tau(y\mathfrak{d}b/d). \end{aligned}$$

If $\phi^*(ya/d) \neq 0$ then $ya/d \in \mathcal{O}$ and the map $b \mapsto \tau(y\mathfrak{d}b/d)$ is an R-valued character on $\mathcal{O}/(a)$. In this case the sum $\sum_{b\in S(a)}\tau(y\mathfrak{d}b/d) = |a|^{-1}$, if $y/d \in \mathcal{O}$, and equal to 0, otherwise. Hence if we set $\mathfrak{c} = (d)$, we get:

$$T_{\mathfrak{m}}(\phi)^*(\mathfrak{r}) = \sum_{\substack{\mathfrak{c}+\mathfrak{n}=\mathcal{O}\\ \mathfrak{r}+\mathfrak{m}\subseteq\mathfrak{c}}} \frac{|\mathfrak{c}|}{|\mathfrak{m}|}\phi^*(\frac{\mathfrak{r}\mathfrak{m}}{\mathfrak{c}^2}). \ \square$$

Corollary 3.3. *Let R be a coefficient ring and assume that for each closed point $\mathfrak{p}$ of X an element $c_{\mathfrak{p}} \in R$ is given. Then the R-module of cuspidal automorphic forms $\phi \in \mathcal{A}_0(\mathfrak{n}, R)$ such that $T_{\mathfrak{p}}(\phi) = c_{\mathfrak{p}}\phi$ for each closed point $\mathfrak{p}$ of X is isomorphic to an ideal $\mathfrak{a} \triangleleft R$ via the map $\phi \mapsto \phi^*(1)$.*

Proof. For each effective divisor $\mathfrak{r}$ we are going to show that $\phi^*(\mathfrak{r})$ is uniquely determined by the eigenvalues $c_{\mathfrak{p}}$ and $\phi^*(1)$ by induction on the maximum $d(\mathfrak{r})$ of exponents of prime divisors of $\mathfrak{r}$. By Proposition 2.13 this implies the proposition. If $d(\mathfrak{r}) = 0$ then the claim is obvious. If $d(\mathfrak{r}) = 1$, then $\mathfrak{r} = \mathfrak{p}_1 \cdots \mathfrak{p}_n$ is the product of pair-wise different prime divisors. By Lemma 3.2 we have:

$$c_{\mathfrak{p}_1}\cdots c_{\mathfrak{p}_n}\phi^*(1) = T_{\mathfrak{p}_1}\cdots T_{\mathfrak{p}_n}(\phi)^*(1) = \frac{1}{|\mathfrak{p}_1\cdots\mathfrak{p}_n|}\phi^*(\mathfrak{p}_1\cdots\mathfrak{p}_n).$$

If $d(\mathfrak{r}) > 1$, then $\mathfrak{r} = \mathfrak{m}\mathfrak{p}^2$ for some prime ideal $\mathfrak{p}$. The lemma above implies that we have the recursive relation:

$$c_{\mathfrak{p}}\phi^*(\mathfrak{m}\mathfrak{p}) = T_{\mathfrak{p}}(\phi)^*(\mathfrak{m}\mathfrak{p}) = \frac{1}{|\mathfrak{p}|}\phi^*(\mathfrak{m}\mathfrak{p}^2) + \phi^*(\mathfrak{m}),$$

if $\mathfrak{p}$ does not lie in the support of $\mathfrak{n}$, and

$$c_{\mathfrak{p}}\phi^*(\mathfrak{m}\mathfrak{p}) = T_{\mathfrak{p}}(\phi)^*(\mathfrak{m}\mathfrak{p}) = \frac{1}{|\mathfrak{p}|}\phi^*(\mathfrak{m}\mathfrak{p}^2),$$

otherwise. $\square$

Definition 3.4. Fix a valuation ∞ of F. We may assume that the support of divisor $\mathfrak{d}$ attached to the character τ in Proposition 2.10 does not contain ∞. Let $\mathcal{H}(\mathfrak{n}, R)$ denote the R-module of automorphic forms f of level $\mathfrak{n}\infty$ satisfying the following two identities:

$$\phi(g\begin{pmatrix} 0 & 1 \\ \upsilon & 0 \end{pmatrix}) = -\phi(g), (\forall g \in GL_2(\mathbb{A})),$$

and

$$\phi(g\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}) + \sum_{\epsilon\in\mathbf{f}_\infty}\phi(g\begin{pmatrix} 1 & 0 \\ \epsilon & 1 \end{pmatrix}) = 0, (\forall g \in GL_2(\mathbb{A})),$$

where υ is a uniformizer in F_∞ and we consider $GL_2(F_\infty)$ as a subgroup of $GL_2(\mathbb{A})$ and we understand the product of their elements accordingly. Such automorphic forms are called harmonic. Let $\mathcal{H}_0(\mathfrak{n}, R)$ denote the R-module of R-valued cuspidal harmonic forms of level $\mathfrak{n}\infty$.

Lemma 3.5. *Let ϕ be an element of $\mathcal{H}(\mathfrak{n}, R)$. Then $T_\infty(\phi) = \phi$.*

Proof. For any $\epsilon \in \mathbf{f}_\infty^*$ we have the matrix identity:

$$\begin{pmatrix} 1 & 0 \\ \epsilon & 1 \end{pmatrix} \begin{pmatrix} \epsilon^{-1} & 1 \\ 0 & -\epsilon \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \upsilon & 0 \end{pmatrix} = \begin{pmatrix} \upsilon & \epsilon^{-1} \\ 0 & 1 \end{pmatrix}.$$

Hence the second identity in Definition 3.4 can be rewritten as follows:

$$\begin{aligned} 0 =&\phi(g\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}) + \sum_{\epsilon \in \mathbf{f}_\infty} \phi(g\begin{pmatrix} 1 & 0 \\ \epsilon & 1 \end{pmatrix}) \\ =&-\phi(g\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ \upsilon & 0 \end{pmatrix}) + \phi(g) - \sum_{\epsilon \in \mathbf{f}_\infty^*} \phi(g\begin{pmatrix} \upsilon & \epsilon^{-1} \\ 0 & 1 \end{pmatrix}) = \phi(g) - T_\infty(\phi) \end{aligned}$$

using the left $\mathbb{K}_0(\mathfrak{n}\infty)$-invariance and the first identity. □

Proposition 3.6. *Let R be a coefficient ring and assume that for each closed point $\mathfrak{p} \neq \infty$ of X an element $c_\mathfrak{p} \in R$ is given. Then the R-module of cuspidal harmonic forms $\phi \in \mathcal{H}_0(\mathfrak{n}, R)$ such that $T_\mathfrak{p}(\phi) = c_\mathfrak{p}\phi$ for each closed point $\mathfrak{p} \neq \infty$ of X is isomorphic to an ideal $\mathfrak{a} \triangleleft R$ via the map $\phi \mapsto \phi^*(1)$.*

Proof. By the lemma above ϕ is also an eigenvector for T_∞. The claim now follows from Corollary 3.3. □

Remark 3.7. The result above is the analogue of the classical (weak) multiplicity one result for mod p modular forms. In order to be useful for some of the applications we have in mind, we will need a multiplicity one result which does not require the eigenvalue of $T_\mathfrak{p}$ to be specified for every closed point $\mathfrak{p}$. We will prove such a result only in a special case. First let us introduce the following general notation: let $\mathbb{A}_f$, $\mathcal{O}_f$ denote the restricted direct products $\prod'_{x \neq \infty} F_x$ and $\prod'_{x \neq \infty} \mathcal{O}_x$, respectively. The former is also called the ring of finite adeles of F and the latter is its maximal compact subring. For the rest of the this chapter we assume that $F = \mathbb{F}_q(T)$ is the rational function field of transcendence degree one over $\mathbb{F}_q$, where T is an indeterminate, and ∞ is the point at infinity on $X = \mathbb{P}^1(F)$. Finally let $M[n]$ denote the n-torsion submodule of every abelian group M for any natural number $n \in \mathbb{N}$.

Proposition 3.8. *The map*

$$\mathcal{H}(1, R) \to R, \quad \phi \mapsto \phi^0(1)$$

is an isomorphism onto $R[q+1]$ for every coefficient ring R.

Proof. It is well-known that there is a natural bijection:

$$\iota : GL_2(\mathbb{F}_q[T])\backslash GL_2(F_\infty)/\Gamma_\infty Z(F_\infty) \longrightarrow GL_2(F)\backslash GL_2(\mathbb{A})/\mathbb{K}_0(\infty)Z(F_\infty),$$

where $\Gamma_\infty = K_0(\infty)_\infty$ denote the Iwahori subgroup of $GL_2(F_\infty)$:

$$\Gamma_\infty = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathcal{O}_\infty) | \infty(c) > 0 \right\},$$

and ι is induced by the natural inclusion $GL_2(F_\infty) \to GL_2(\mathbb{A})$. The former double coset is the set of edges of the Bruhat-Tits tree of the local field $F_\infty = \mathbb{F}_q((\frac{1}{T}))$ factored out by $GL_2(\mathbb{F}_q[T])$. Under this bijection elements of $\mathcal{H}(1, R)$ correspond to $GL_2(\mathbb{F}_q[T])$-invariant R-valued harmonic cochains on the Bruhat-Tits tree. This correspondence is bijective, because $Z(\mathbb{A}) = Z(F)Z(\mathcal{O})Z(F_\infty)$, so every harmonic cochain is invariant with respect to this group. The reader may find the following description of the quotient graph above in Proposition 3 of 1.6 of [17], page 86-67:

Proposition 3.9. *Let Λ_n denote the vertex of the Bruhat-Tits tree represented by the matrix $\begin{pmatrix} T^n & 0 \\ 0 & 1 \end{pmatrix}$ for every natural number $n \in \mathbb{N}$.*

- (i) *the vertices Λ_n form a fundamental domain for the action of $GL_2(\mathbb{F}_q[T])$ on the set of vertices of the Bruhat-Tits tree,*
- (ii) *the stabilizer of Λ_0 in $GL_2(\mathbb{F}_q[T])$ acts transitively on the set of edges with origin Λ_0,*
- (iii) *for every n there is an edge $\Lambda_n\Lambda_{n+1}$ with origin Λ_n and terminal vertex Λ_n,*
- (iv) *for every $n \geq 1$, the stabilizer of the edge $\Lambda_n\Lambda_{n+1}$ in $GL_2(\mathbb{F}_q[T])$ acts transitively on the set of edges with origin Λ_n distinct from $\Lambda_n\Lambda_{n+1}$.*

Proof. The second half of (i) is the corollary to the proposition quoted above on page 87 of [17]. $\square$

Let us return to the proof of Proposition 3.8. Let α denote the value of the harmonic cochain Φ corresponding to ϕ on the edge $\Lambda_0\Lambda_1$. By (ii) of the proposition above the value of Φ is α on all other edges with origin Λ_0, so $\alpha \in R[q+1]$ by harmonicity. We are going to show that $\Phi(\Lambda_n\Lambda_{n+1}) = (-1)^n\alpha$ for all n by induction. By harmonicity $\Phi(\Lambda_n\Lambda_{n-1}) = -(-1)^{n-1}\alpha = (-1)^n\alpha$. Also note that the value of Φ is $(-1)^n\alpha$ on all edges with origin Λ_n distinct from $\Lambda_n\Lambda_{n+1}$ by (iv) of the proposition above. Hence we must have $\Phi(\Lambda_n\Lambda_{n+1}) = (-q)(-1)^n\alpha = (-1)^n\alpha$ by harmonicity, also using the fact $\alpha \in R[q+1]$. We conclude that Φ is uniquely determined by its value on the edge $\Lambda_0\Lambda_1$. For every $g \in GL_2(\mathbb{A})$ the residue of the degree of the divisor $\det(g)$ modulo 2 depends only on its class in $GL_2(F)\backslash GL_2(\mathbb{A})/\mathbb{K}_0(1)Z(\mathbb{A})$. In particular if g is equivalent to the vertex Λ_n, then $n \equiv \deg(\det(g)) \mod 2$. Hence our description of Φ can be reformulated by saying that $\phi(g) = (-1)^{\deg(\det(g))}\alpha$. Moreover

$$\phi^0(1) = \int_{F\backslash\mathbb{A}} \phi(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix})\mathrm{d}\mu(x) = \int_{F\backslash\mathbb{A}} \alpha\mathrm{d}\mu(x) = \alpha,$$

because every element of the set, where the integral above is taken, has determinant 1. On the other hand for every $\alpha \in R[q+1]$ the function $H(\alpha)$, whose value is $(-1)^n\alpha$ on every edge of origin Λ_n, is clearly a harmonic cochain. The claim follows. $\square$

PROPOSITION 3.10. *Let R be a coefficient ring and let $\mathfrak{p} \neq \infty$ be a closed point of X. Then every harmonic form $\phi \in \mathcal{H}(\mathfrak{p}, R)$ such that $\phi^*(\mathfrak{m}) = 0$ for each effective divisor $\mathfrak{m}$ whose support does not contain $\mathfrak{p}$ and ∞ is an element of $\mathcal{H}(1, R)$.*

PROOF. First note that $\phi^*(\mathfrak{m}) = 0$ even for those effective divisors $\mathfrak{m}$ whose support do not contain $\mathfrak{p}$, but may contain ∞, since for any effective divisor $\mathfrak{n}$ we have:

$$\phi^*(\mathfrak{n}) = T_\infty(\phi)^*(\mathfrak{n}) = \frac{1}{|\infty|}\phi^*(\mathfrak{n}\infty),$$

by Lemma 3.2 and Lemma 3.5, so this seemingly stronger statement follows from the condition in the claim by induction on the multiplicity of ∞ in $\mathfrak{m}$. For every $y \in \mathbb{A}^*$ and a, $x \in \mathbb{A}$ we have:

$$\begin{aligned}\phi(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}) =& \phi(\begin{pmatrix} y & x+ya \\ 0 & 1 \end{pmatrix}) \\ =& \phi^0(y) + \sum_{\eta \in F^*} \phi^*(\eta y \mathfrak{d}^{-1})\tau(\eta y a)\tau(\eta x).\end{aligned}$$

If $a \in \mathfrak{p}^{-1}$, then $\phi^*(\eta y \mathfrak{d}^{-1}) = 0$ unless $\tau(\eta y a) = 1$, because $\phi^*(\eta y \mathfrak{d}^{-1}) \neq 0$ implies that $\eta y \in \mathfrak{p}\mathcal{D}$, so $\eta y a \in \mathfrak{p}\mathcal{D}\mathfrak{p}^{-1} \subset \operatorname{Ker}(\tau)$. Hence the Fourier expansion above is independent of the choice of $a \in \mathfrak{p}^{-1}$, so for every $g \in P(\mathbb{A})$ and a as above we have:

$$\phi(g\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}) = \phi(g).$$

In the proof of Lemma 2.12 we showed that $GL_2(F)P(\mathbb{A})Z(\mathbb{A})$ is dense in $GL_2(\mathbb{A})$, so the identity above holds for all $g \in GL_2(\mathbb{A})$ by continuity. Let $\pi \in \mathbb{A}_f^*$ be an idele such that $\pi\mathcal{O}_f = \mathfrak{p}$. We define the function $\psi : GL_2(\mathbb{A}) \to R$ by the formula:

$$\psi(g) = \phi(g\begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix}), \quad \forall g \in GL_2(\mathbb{A}),$$

where we consider $GL_2(\mathbb{A})$ as a $GL_2(\mathbb{A}_f)$-module and we understand the product of their elements accordingly. We claim that $\psi \in \mathcal{H}(1, R)$. It is clearly left-invariant with respect to $Z(\mathbb{A})GL_2(F)$. On the other hand we have:

$$\begin{aligned}\psi(g\begin{pmatrix} a & b \\ c & d \end{pmatrix}) = \phi(g\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix}) =& \phi(g\begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix}\begin{pmatrix} d & c/\pi \\ \pi b & a \end{pmatrix}) \\ =& \phi(g\begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix}) = \psi(g)\end{aligned}$$

for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathbb{K}_0(\mathfrak{p}) \cap GL_2(\mathbb{A}_f)$, upon using the identity:

$$\begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix}\begin{pmatrix} d & c/\pi \\ \pi b & a \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix},$$

hence $\psi \in \mathcal{A}(\mathfrak{p}\infty, R)$. The same identity may be used to show that:

$$\psi(g\begin{pmatrix}1 & 0\\ a & 1\end{pmatrix}) = \phi(g\begin{pmatrix}1 & 0\\ a & 1\end{pmatrix}\begin{pmatrix}0 & 1\\ \pi & 0\end{pmatrix}) = \phi(g\begin{pmatrix}0 & 1\\ \pi & 0\end{pmatrix}\begin{pmatrix}1 & a/\pi\\ 0 & 1\end{pmatrix}) = \phi(g\begin{pmatrix}0 & 1\\ \pi & 0\end{pmatrix}) = \psi(g)$$

for all $a \in \mathcal{O}_f$, hence ψ is even in $\mathcal{A}(\infty, R)$. Obviously the matrix $\left(\begin{smallmatrix}0 & 1\\ \pi & 0\end{smallmatrix}\right)$ commutes with the matrices in Definition 3.4, so ψ is harmonic, too. Using the $Z(\mathbb{A})$-invariance of ψ we get:

$$\phi(g) = \psi(g\begin{pmatrix}0 & \pi^{-1}\\ 1 & 0\end{pmatrix}) = \psi(g\begin{pmatrix}0 & 1\\ \pi & 0\end{pmatrix}\begin{pmatrix}\pi^{-1} & 0\\ 0 & \pi^{-1}\end{pmatrix}) = \psi(g\begin{pmatrix}0 & 1\\ \pi & 0\end{pmatrix}),$$

the claim of the proposition follows by the lemma below and applying the same argument to ψ. □

Lemma 3.11. *For every harmonic form $\phi \in \mathcal{H}(1, R)$ we have $\phi^*(\mathfrak{m}) = 0$ for every effective divisor $\mathfrak{m}$.*

Proof. It will be sufficient to show that the function $x \mapsto \phi(\left(\begin{smallmatrix}y & x\\ 0 & 1\end{smallmatrix}\right))$ is constant on $\mathbb{A}$ for each $y \in \mathcal{O} \cap \mathbb{A}^*$. The latter follows from the fact that the determinant is constant, it is equal to y. □

Theorem 3.12. *Let R be a coefficient ring and let $\mathfrak{p} \neq \infty$ be a closed point of X. Assume that for each closed point $\mathfrak{q}$ of X, different from $\mathfrak{p}$ and ∞, an element $c_{\mathfrak{q}} \in R$ is given. Then the R-module of cuspidal harmonic forms $\phi \in \mathcal{H}_0(\mathfrak{p}, R)$ such that $T_{\mathfrak{q}}(\phi) = c_{\mathfrak{q}}\phi$ for each closed point $\mathfrak{q}$ of X, different from $\mathfrak{p}$ and ∞, is isomorphic to an ideal $\mathfrak{a} \triangleleft R$ via the map $\phi \mapsto \phi^*(1)$.*

Proof. It will be sufficient to prove that any such ϕ with $\phi^*(1) = 0$ is zero by taking the difference of any two elements of the module with the same first Fourier coefficient. The argument of Corollary 3.3 implies that $\phi^*(\mathfrak{m}) = 0$ for each effective divisor $\mathfrak{m}$ whose support does not contain $\mathfrak{p}$ and ∞ for every such ϕ. By Proposition 3.10 ϕ is in $\mathcal{H}(1, R)$, hence ϕ is an element of $\mathcal{H}_0(1, R)$, too. The latter R-module is trivial by Proposition 3.8. □

4. The Kronecker limit formula

Notation 4.1. We will adopt the convention which assigns 0 or 1 as value to the empty sum or product, respectively. For every $g \in GL_2(\mathbb{A})$ (or $g \in \mathbb{A}$, *etc.*) let g_f denote its finite component in $GL_2(\mathbb{A}_f)$. Let $|\cdot|$ denote the normalized absolute value with respect to ∞ if its argument is in F_∞. For each $(u, v) \in F_\infty^2$ let $\|(u, v)\|$, $\infty(u, v)$ denote $\max(|u|, |v|)$ and $\min(\infty(u), \infty(v))$, respectively.

Definition 4.2. Let $F^2_<$ denote the set: $F^2_< = \{(a,b) \in F^2_\infty \| a| < |b|\}$. Let $\mathfrak{m}$ be an effective divisor on X whose support does not contain ∞. Let the same symbol also denote the ideal $\mathfrak{m} \cap \mathcal{O}_f$ by abuse of notation. For every $g \in GL_2(\mathbb{A})$, $(\alpha, \beta) \in (\mathcal{O}_f/\mathfrak{m})^2$, and n integer let

$$W_{\mathfrak{m}}(\alpha,\beta,g,n) = \{0 \neq f \in F^2 | fg_f \in (\alpha,\beta) + \mathfrak{m}\mathcal{O}_f^2, -n = \infty(fg_\infty)\}, \text{ and}$$

$$V_{\mathfrak{m}}(\alpha,\beta,g,n) = \{f \in W_{\mathfrak{m}}(\alpha,\beta,g,n) | fg_\infty \in F^2_<\}.$$

Also let

$$W_{\mathfrak{m}}(\alpha,\beta,g_f) = \bigcup_{n\in\mathbb{Z}} W_{\mathfrak{m}}(\alpha,\beta,g,n) \text{ and } V_{\mathfrak{m}}(\alpha,\beta,g) = \bigcup_{n\in\mathbb{Z}} V_{\mathfrak{m}}(\alpha,\beta,g,n).$$

Obviously the first set is well-defined. Finally let $E_{\mathfrak{m}}(\alpha,\beta,g,s)$ denote the $\mathbb{C}$-valued function:

$$E_{\mathfrak{m}}(\alpha,\beta,g,s) = |\det(g)|^s \sum_{f \in V_{\mathfrak{m}}(\alpha,\beta,g)} \|fg_\infty\|^{-2s},$$

for each complex number s and g, (α,β) as above, if the infinite sum is absolutely convergent.

Proposition 4.3. *The sum $E_{\mathfrak{m}}(\alpha,\beta,g,s)$ converges absolutely, if* $\mathrm{Re}(s) > 1$, *for each $g \in GL_2(\mathbb{A})$.*

Proof. The reader may find the same argument in [16]. The series $E_{\mathfrak{m}}(\alpha,\beta,g,s)$ is majorated by the series:

$$E(g,s) = |\det(g)|^s \sum_{\substack{f \in F^2 - \{0\} \\ fg \in \mathcal{O}_f^2}} \|(fg)_\infty\|^{-2s},$$

so it will be sufficient to prove that $E(g,s)$ converges absolutely for each $g \in GL_2(\mathbb{A})$ if $\mathrm{Re}(s) > 1$. For every $g \in GL_2(\mathbb{A})$ let $\mathcal{E}(g)$ denote the sheaf on X whose group of sections is for every open subset $U \subseteq X$ is

$$\mathcal{E}(g)(U) = \{f \in F^2 | fg \in \mathcal{O}_v^2, \forall v \in |U|\},$$

where we denote the set of closed points of U by $|U|$. The sheaf $\mathcal{E}(g)$ is a coherent locally free sheaf of rank two. If $\mathcal{F}_n$ denote the sheaf $\mathcal{F} \otimes \mathcal{O}_X(\infty)^n$ for every coherent sheaf $\mathcal{F}$ on X and integer n, then for every $g \in GL_2(\mathbb{A})$ and $s \in \mathbb{C}$ the series above can be rewritten as

$$E(g,s) = \sum_{n\in\mathbb{Z}} |H^0(X,\mathcal{E}(g)_n) - H^0(X,\mathcal{E}(g)_{n-1})| q^{-s\deg(\mathcal{E}(g)_n)}.$$

By the Riemann-Roch theorem for curves:

$$\dim H^0(X,\mathcal{F}) - \dim H^0(X, K_X \otimes \mathcal{F}^\vee) = 2 - 2g(X) + \deg(\mathcal{F})$$

for any coherent locally free sheaf of rank two $\mathcal{F}$ on X, where K_X, $\mathcal{F}^\vee$ and $g(X)$ is the canonical bundle on X, the dual of $\mathcal{F}$, and the genus of X, respectively. Because $\dim H^0(X,\mathcal{F}_{-n}) = 0$ for n sufficiently large depending on $\mathcal{F}$, we have that

$$|H^0(X,\mathcal{E}(g)_n)| = q^{2-2g(X)+\deg(\mathcal{E}(g))+2n\deg(\infty)} \text{ and } |H^0(X,\mathcal{E}(g)_{-n})| = 1,$$

if n is a sufficiently large positive number. Hence

$$E(g,s) = p(q^{-s}) + q^{2-2g(X)+(1-s)\deg(\mathcal{E}(g))}(1-q^{-\deg(\infty)})\sum_{n=0}^{\infty} q^{2n(1-s)\deg(\infty)},$$

where p is a polynomial. The claim now follows from the convergence of the geometric series. □

Notation 4.4. Let Ω denote the rigid analytic upper half plane, or Drinfeld's upper half plane over F_∞. The set of points of Ω is $\mathbb{C}_\infty - F_\infty$, denoted also by Ω by abuse of notation, where $\mathbb{C}_\infty$ is the completion of the algebraic closure of F_∞. For the definition of its rigid analytic structure as well as the other concepts recalled below see for example [11]. For each holomorphic function $u : \Omega \longrightarrow \mathbb{C}_\infty^*$ let $r(u) : GL_2(F_\infty) \longrightarrow \mathbb{Z}$ denote the van der Put logarithmic derivative of u (see [11], page 40). If $u : GL_2(\mathbb{A}_f) \times \Omega \longrightarrow \mathbb{C}_\infty^*$ is holomorphic in the second variable for each $g \in GL_2(\mathbb{A}_f)$ then we define $r(u)$ to be the $\mathbb{Z}$-valued function on the set $GL_2(\mathbb{A}) = GL_2(\mathbb{A}_f) \times GL_2(F_\infty)$ given by the formula $r(u)(g_f, g_\infty) = r(u(g_f,\cdot))(g_\infty)$. For each $(\alpha,\beta) \in (\mathcal{O}_f/\mathfrak{m})^2$, and N positive integer let $\epsilon_{\mathfrak{m}}(\alpha,\beta,N)(g,z)$ denote the function:

$$\epsilon_{\mathfrak{m}}(\alpha,\beta,N)(g,z) = \prod_{n\leq N}\left(\prod_{(a,b)\in W_{\mathfrak{m}}(\alpha,\beta,g,n)}(az+b)\cdot\prod_{(c,d)\in W_{\mathfrak{m}}(0,0,g,n)}(cz+d)^{-1}\right).$$

on the set $GL_2(\mathbb{A}_f) \times \Omega$.

Lemma 4.5. *The limit*

$$\epsilon_{\mathfrak{m}}(\alpha,\beta)(g,z) = \lim_{N\longrightarrow\infty} \epsilon_{\mathfrak{m}}(\alpha,\beta,N)(g,z)$$

converges uniformly in z on every admissible open subdomain of Ω for every fixed g and defines a function holomorphic in the second variable.

Proof. If $(\alpha,\beta)=(0,0)$ then the claim is trivial. Otherwise let (α,β) also denote an element of $W_{\mathfrak{m}}(\alpha,\beta,g_f)$ by abuse of notation. For sufficiently large N the product $\epsilon_{\mathfrak{m}}(\alpha,\beta,N)(g,z)$ can be rewritten as:

$$\epsilon_{\mathfrak{m}}(\alpha,\beta,N)(g,z)=(\alpha z+\beta)\cdot\prod_{\substack{n\leq N\\(a,b)\in W_{\mathfrak{m}}(0,0,g,n)}}\left(1+\frac{\alpha z+\beta}{az+b}\right).$$

The system of sets $\Omega(\omega)=\{z\in\mathbb{C}_\infty | 1/\omega\leq|z|_i,|z|\leq\omega\}$, where $1<\omega$ is any rational number and $|z|_i=\inf_{x\in F_\infty}|z+x|$ is the imaginary absolute value of z, is a cover of Ω by admissible open subdomains. On the set $\Omega(\omega)$:

$$\left|\frac{\alpha z+\beta}{az+b}\right|\leq\frac{\max(\omega|\alpha|,|\beta|)}{\max(\omega^{-1}|a|,|b|)},$$

so it converges to zero as $\|(a,b)\|\to\infty$. The claim follows at once. □

Definition 4.6. For every $\rho\in GL_2(F_\infty)$ and $z\in\mathbb{P}^1(\mathbb{C}_\infty)$ let $\rho(z)$ denote the image of z under the Möbius transformation corresponding to ρ. Let moreover $D(\rho)$ denote the open disc

$$D(\rho)=\{z\in\mathbb{P}^1(\mathbb{C}_\infty)|1<|\rho^{-1}(z)|\}.$$

Set $\delta(\rho)=-1$, if the infinite point of the projective line lies in $D(\rho)$, and let $\delta(\rho)=0$, otherwise.

Proposition 4.7. *For all $g\in GL_2(\mathbb{A})$ we have:*

$$r(\epsilon_{\mathfrak{m}}(\alpha,\beta))(g)=\delta(g_\infty)+\lim_{N\to\infty}\left(\sum_{n\leq N}|V_{\mathfrak{m}}(\alpha,\beta,g,n)|-|V_{\mathfrak{m}}(0,0,g,n)|\right).$$

Proof. The van der Put logarithmic derivative is continuous with respect to the limit of the supremum topologies on the affinoid subdomains of Ω, hence

$$r(\epsilon_{\mathfrak{m}}(\alpha,\beta))(g)=\lim_{N\to\infty}r(\epsilon_{\mathfrak{m}}(\alpha,\beta,N))(g)$$

by Lemma 4.5. More or less by definition (see [11]) for every $u\in\mathcal{O}^*(\Omega)$ rational function $r(u)(\rho)$ equals to the number of zeros z of u with $z\in D(\rho)$ counted with multiplicities minus the number of poles z of u with $z\in D(\rho)$ counted with multiplicities. If we assume that $\delta(\rho)=0$ then we can conclude that $r(az+b)(\rho)$ is 1 if and only if $(a,b)\rho\in F^2_<$ and it is 0, otherwise. Hence the claim holds for g if $\delta(g_\infty)=0$ by the additivity of the van der Put derivative. In particular the limit on the right exists in this case. Let $\Pi\in GL_2(F_\infty)$ be the matrix whose diagonal entries are zero, and its lower left and upper right entry is π and 1, respectively, where π is a uniformizer of F_∞. Clearly $F^2_\infty - \{0\} = F^2_< \coprod F^2_< \Pi$, hence

$$W_{\mathfrak{m}}(\alpha,\beta,g_f) = V_{\mathfrak{m}}(\alpha,\beta,g) \coprod V_{\mathfrak{m}}(\alpha,\beta,g\Pi)$$

for any $g \in GL_2(\mathbb{A})$. Also exactly one of the sets $D(g_\infty)$ and $D(g_\infty\Pi)$ contains the infinite point. Hence it will suffice to show that for any g and sufficiently large $N \in \mathbb{N}$ the sum

$$-1 + \sum_{n \leq N} (|W_{\mathfrak{m}}(\alpha,\beta,g,n)| - |W_{\mathfrak{m}}(0,0,g,n)|)$$

vanishes to conclude that the limit in the claim above exits in all cases. This will also imply that the expression $l(g)$ on right hand side satisfies the functional equation $l(g) + l(g\Pi) = 0$. Since the left hand side also satisfies this property the claim will follow. But the sum above vanishes because of the bijection which we already used implicitly in the proof of Lemma 4.5 when we rewrote $\epsilon_{\mathfrak{m}}(\alpha,\beta,N)(g,z)$. □

Kronecker Limit Formula 4.8. *For all $g \in GL_2(\mathbb{A})$ we have:*

$$r(\epsilon_{\mathfrak{m}}(\alpha,\beta))(g) = \delta(g_\infty) + \lim_{s\to 0^+} (E_{\mathfrak{m}}(\alpha,\beta,g,s) - E_{\mathfrak{m}}(0,0,g,s)).$$

Proof. We have to show that the limit exists on the right hand side and it equals to the left hand side. For all complex s with $\mathrm{Re}(s) > 1$ we have:

$$E_{\mathfrak{m}}(\alpha,\beta,g,s) - E_{\mathfrak{m}}(0,0,g,s) = |\det(g)|^s \sum_{n=-\infty}^{\infty} (|V_{\mathfrak{m}}(\alpha,\beta,g,n)| - |V_{\mathfrak{m}}(0,0,g,n)|)\,|\pi|^{2sn}.$$

According to the proof of Proposition 4.3 the cardinalities $|V_{\mathfrak{m}}(\alpha,\beta,g,n)|$ and $|V_{\mathfrak{m}}(0,0,g,n)|$ are zero if n is sufficiently small. Let (α,β) again denote an element of $W_{\mathfrak{m}}(\alpha,\beta,g_f)$ by abuse of notation as in the proof of Lemma 4.5. The map $f \mapsto (\alpha,\beta)+f$ defines a bijection between $V_{\mathfrak{m}}(0,0,g,n)$ and $V_{\mathfrak{m}}(\alpha,\beta,g,n)$ if n is sufficiently large, so the limit exists and

$$\lim_{s\to 0^+} (E_{\mathfrak{m}}(\alpha,\beta,g,s) - E_{\mathfrak{m}}(0,0,g,s)) = \lim_{N\to\infty} \left(\sum_{n \leq N} |V_{\mathfrak{m}}(\alpha,\beta,g,n)| - |V_{\mathfrak{m}}(0,0,g,n)| \right).$$

The claim now follows from the previous proposition. □

5. Computation of Fourier expansions

Definition 5.1. For every $\alpha \in \mathcal{O}_f/\mathfrak{m}$ and $z \in \mathbb{A}_f^*$ let

$$V_{\mathfrak{m}}(\alpha, z) = \{u \in F^* | uz \in \alpha + \mathfrak{m}\}.$$

For each α and z as above let $\zeta_{\mathfrak{m}}(\alpha, z, s)$ denote the $\mathbb{C}$-valued function

$$\zeta_{\mathfrak{m}}(\alpha, z, s) = \sum_{u \in V_{\mathfrak{m}}(\alpha, z)} |u|_\infty^{-s},$$

if this infinite sum is absolutely convergent. For every $\alpha \in \mathcal{O}_f/\mathfrak{m}$ define $\rho(\alpha)$ to be 1, if $\alpha = 0$, and to be 0, otherwise. Let μ be the unique Haar measure on the locally compact abelian topological group $\mathbb{A}$ such that $\mu(\mathcal{O})$ is equal to $|\mathfrak{d}|^{-1/2}$. Since this measure is left-invariant with respect to the discrete subgroup F by definition, it induces a measure on $F\backslash\mathbb{A}$ which will be denoted by the same letter by abuse of notation. By our choice of normalization $\mu(F\backslash\mathbb{A}) = 1$, so our notation is compatible with Definitions 2.3 and 2.8. Note that the former is the direct product of a Haar measure μ_f on $\mathbb{A}_f$ and a Haar measure μ_∞ on F_∞ such that $\mu_f(\mathcal{O}_f) = |\mathfrak{d}|^{-1/2}$ and $\mu_\infty(\mathcal{O}_\infty) = 1$. Finally let q_∞ be the cardinality of $\mathbf{f}_\infty$.

Proposition 5.2. *For each complex s with $\mathrm{Re}(s) > 1$ we have:*

$$\begin{aligned} E_{\mathfrak{m}}(\alpha, \beta, \cdot, s)^0(z) =& \rho(\alpha)|z|^s \zeta_{\mathfrak{m}}(\beta, 1, 2s) \\ &+ \frac{|\mathfrak{m}|}{|\mathfrak{d}|^{1/2}} \frac{|z|^s(q_\infty - 1)}{|z|_\infty^{2s-1}(q_\infty^{2s} - q_\infty)} \zeta_{\mathfrak{m}}(\alpha, z_f, 2s-1). \end{aligned}$$

Proof. Recall that the notion of Fourier coefficients are defined for all complex-valued automorphic forms (see [19]). The claim above should be understood in this sense. By grouping the terms in the infinite sum of Definition 4.2 we get the following identity:

$$E_{\mathfrak{m}}(\alpha, \beta, \begin{pmatrix} z & x \\ 0 & 1 \end{pmatrix}, s) = |z|^s \sum_{(0,u) \in V_{\mathfrak{m}}(\alpha, \beta, \begin{pmatrix} z & x \\ 0 & 1 \end{pmatrix})} |u|_\infty^{-2s} + |z|^s \sum_{b \in F} \sum_{\substack{a \in F^* \\ (a,0) \in V_{\mathfrak{m}}(\alpha,\beta, \begin{pmatrix} z & x+b \\ 0 & 1 \end{pmatrix})}} |a(x+b)|_\infty^{-2s}.$$

According to the Fourier inversion formula the Fourier coefficient $E_{\mathfrak{m}}(\alpha, \beta, \cdot, s)^0(z)$ is given by the formula

$$E_{\mathfrak{m}}(\alpha, \beta, \cdot, s)^0(z) = \int_{F\backslash\mathbb{A}} E_{\mathfrak{m}}(\alpha, \beta, \begin{pmatrix} z & x \\ 0 & 1 \end{pmatrix}, s) \mathrm{d}\mu(x).$$

By substituting the formula above into this integral and interchanging summation and integration we get:

$$E_{\mathfrak{m}}(\alpha, \beta, \cdot, s)^0(z) = \rho(\alpha)|z|^s \zeta_{\mathfrak{m}}(\beta, 1, 2s) + |z|^s \sum_{a \in V_{\mathfrak{m}}(\alpha, z)} |a|_\infty^{-2s} \int_{\substack{a \in V_{\mathfrak{m}}(\beta, x_f) \\ |x|_\infty > |z|_\infty}} |x|_\infty^{-2s} \mathrm{d}\mu(x).$$

Note that this computation is justified by the Lebesgue convergence theorem. The measure of the set $\{x \in \mathbb{A}_f | a \in V_{\mathfrak{m}}(\beta, x)\}$ is:

$$\begin{aligned}\mu_f(\{x \in \mathbb{A}_f | a \in V_{\mathfrak{m}}(\beta, x)\}) =& \mu_f(a_f^{-1}(\beta + \mathfrak{m}))\\ =& |a_f|^{-1}|\mathfrak{m}||\mathfrak{d}|^{-1/2} = |a|_\infty |\mathfrak{m}||\mathfrak{d}|^{-1/2}.\end{aligned}$$

On the other hand:

$$\begin{aligned}\int\limits_{|x|_\infty > |z|_\infty} |x|_\infty^{-2s} \mathrm{d}\mu_\infty(x) =& \sum_{n=1}^{\infty} |z|_\infty^{-2s} q_\infty^{-2sn} \int\limits_{\infty(x)=\infty(z)-n} \mathrm{d}\mu_\infty(x)\\ =& \sum_{n=1}^{\infty} |z|_\infty^{1-2s} q_\infty^{(1-2s)n} \cdot \frac{q_\infty - 1}{q_\infty},\end{aligned}$$

so the second term in the sum above is equal to:

$$|z|^s |\mathfrak{m}||\mathfrak{d}|^{-1/2} |z|_\infty^{1-2s} \cdot \frac{q_\infty - 1}{q_\infty^{2s} - q_\infty} \cdot \sum_{a \in V_{\mathfrak{m}}(\alpha, z)} |a|_\infty^{1-2s}. \square$$

Definition 5.3. For every $\alpha \in \mathcal{O}_f/\mathfrak{m}$ and $z \in \mathbb{A}_f^*$ let

$$S_{\mathfrak{m}}(\alpha, z) = \{u \in V_{\mathfrak{m}}(\alpha, z) | u_f^{-1}\mathfrak{m}\mathcal{O}_f \subseteq \mathfrak{d}\}.$$

For each $\beta \in \mathcal{O}_f/\mathfrak{m}$ and α, z as above let $\sigma_{\mathfrak{m}}(\alpha, \beta, z, s)$ denote the finite $\mathbb{C}$-valued sum

$$\sigma_{\mathfrak{m}}(\alpha, \beta, z, s) = \sum_{u \in S_{\mathfrak{m}}(\alpha, z)} \tau(-u_f^{-1}\beta)|u|_\infty^{-s},$$

where $\beta \in \mathcal{O}_f$ also denotes a representative of the class β by abuse of notation. The expression above is well-defined because of the condition $u_f^{-1}\mathfrak{m}\mathcal{O}_f \subseteq \mathfrak{d}$.

Proposition 5.4. *For each complex s with $\mathrm{Re}(s) > 1$ we have:*

$$E_{\mathfrak{m}}(\alpha, \beta, \cdot, s)^*(z\mathfrak{d}^{-1}) =$$

$$\Big(-q_\infty^{2s} + \frac{q_\infty - 1}{q_\infty} \cdot \sum_{n=0}^{\infty(z)-1} q_\infty^{n(2s-1)}\Big)|z|^s \frac{|\mathfrak{m}|}{|\mathfrak{d}|^{1/2}} \sigma_{\mathfrak{m}}(\alpha, \beta, z_f, 2s-1),$$

if $\infty(z) \geq 0$, and it is zero, otherwise.

Proof. The first summand in the right hand side of the first equation appearing in the proof above is constant in x, so it does not contribute to the Fourier coefficient $E_{\mathfrak{m}}(\alpha, \beta, \cdot, s)^*(z\mathfrak{d}^{-1})$. Hence

$$\begin{aligned}E_{\mathfrak{m}}(\alpha, \beta, \cdot, s)^*(z\mathfrak{d}^{-1}) =& \int_{F\backslash\mathbb{A}} E_{\mathfrak{m}}(\alpha, \beta, \begin{pmatrix} z & x \\ 0 & 1 \end{pmatrix}, s)\tau(-x)\mathrm{d}\mu(x)\\ =& |z|^s \sum_{a \in V_{\mathfrak{m}}(\alpha, z)} |a|_\infty^{-2s} \int\limits_{\substack{a \in V_{\mathfrak{m}}(\beta, x_f)\\ |x|_\infty > |z|_\infty}} |x|_\infty^{-2s} \tau(-x)\mathrm{d}\mu(x).\end{aligned}$$

interchanging summation and integration. For every $a \in V_{\mathfrak{m}}(\alpha, z)$ the integral above is a product:

$$\int\limits_{\substack{x_f \in a_f^{-1}(\beta+\mathfrak{m}) \\ |x|_\infty > |z|_\infty}} |x|_\infty^{-2s} \tau(-x) \mathrm{d}\mu(x) =$$

$$\tau(-a_f^{-1}\beta) \cdot \int\limits_{a_f^{-1}\mathfrak{m}\mathcal{O}_f} \tau(-x) \mathrm{d}\mu_f(x) \cdot \int\limits_{|x|_\infty > |z|_\infty} |x|_\infty^{-2s} \tau_\infty(-x) \mathrm{d}\mu_\infty(x),$$

where τ_∞ is the restriction of the character τ to the ∞-adic component F_∞. The first integral in the product above is zero unless additive group $a_f^{-1}\mathfrak{m}\mathcal{O}_f$ lies in the kernel of τ which is equivalent to $a \in S_{\mathfrak{m}}(\alpha, z)$. In the latter case it is equal to $\mu_f(a_f^{-1}\mathfrak{m}\mathcal{O}_f) = |a|_\infty |\mathfrak{m}||\mathfrak{d}|^{-1/2}$. By assumption $\mathcal{O}_\infty$ itself is the largest $\mathcal{O}_\infty$-submodule of F_∞ such that the restriction of τ_∞ onto this submodule is trivial, hence the integral on the right above is zero if $\infty(z) < 0$, and it is equal to:

$$\int\limits_{|x|_\infty > |z|_\infty} |x|_\infty^{-2s} \tau_\infty(-x) \mathrm{d}\mu_\infty(x) = \sum_{n=-1}^{\infty(z)-1} \int\limits_{\infty(x)=n} |x|_\infty^{-2s} \tau_\infty(-x) \mathrm{d}\mu_\infty(x)$$

$$= -q_\infty^{2s} + \frac{q_\infty - 1}{q_\infty} \cdot \sum_{n=0}^{\infty(z)-1} q_\infty^{n(2s-1)}, \text{ otherwise. } \square$$

Definition 5.5. Let $A = \mathcal{O}_f \cap F$: it is a Dedekind domain. The ideals of A and the effective divisors on X with support away from ∞ are in a bijective correspondence. These two sets will be identified in all that follows. For any ideal $\mathfrak{n} \triangleleft A$ let $Y_0(\mathfrak{n})$ denote the coarse moduli for rank two Drinfeld modules of general characteristic equipped with a Hecke level-$\mathfrak{n}$ structure. It is an affine algebraic curve defined over F. The group $GL_2(F)$ acts on the product $GL_2(\mathbb{A}_f) \times \Omega$ on the left by acting on the first factor via the natural embedding and on Drinfeld's upper half plane via Möbius transformations. The group $\mathbb{K}_f(\mathfrak{n}) = \mathbb{K}_0(\mathfrak{n}) \cap GL_2(\mathcal{O}_f)$ acts on the right of this product by acting on the first factor via the regular action. Since the quotient set $GL_2(F)\backslash GL_2(\mathbb{A}_f)/\mathbb{K}_f(\mathfrak{n})$ is finite, the set

$$GL_2(F)\backslash GL_2(\mathbb{A}_f) \times \Omega / \mathbb{K}_f(\mathfrak{n})$$

is the disjoint union of finitely many sets of the form $\Gamma\backslash\Omega$, where Γ is a subgroup of $GL_2(F)$ of the form $GL_2(F) \cap g\mathbb{K}_f(\mathfrak{n})g^{-1}$ for some $g \in GL_2(\mathbb{A}_f)$. As these groups act on Ω discretely, the set above naturally has the structure of a rigid analytic curve. Let $Y_0(\mathfrak{n})$ also denote the underlying rigid analytical space of the base change of $Y_0(\mathfrak{n})$ to F_∞ by abuse of notation.

Theorem 5.6. *There is a rigid-analytical isomorphism:*

$$Y_0(\mathfrak{n}) \cong GL_2(F)\backslash GL_2(\mathbb{A}_f) \times \Omega/\mathbb{K}_f(\mathfrak{n}).$$

Proof. See [3], Theorem 6.6. □

Notation 5.7. From now on we make the same assumptions as we did in Remark 3.7. In this case $A = \mathbb{F}_q[T]$. If $\psi : A \longrightarrow \mathbb{C}_\infty\{\tau\}$ is a Drinfeld module of rank two over A, then

$$\psi(T) = T + g(\psi)\tau + \Delta(\psi)\tau^2,$$

where Δ is the Drinfeld discriminant function. It is a Drinfeld modular form of weight q^2-1. Under the identification of Theorem 5.6 the Drinfeld discriminant function Δ is a nowhere vanishing function on $GL_2(\mathbb{A}_f) \times \Omega$ holomorphic in the second variable, and it is equal to:

$$\Delta(g,z) = \prod_{(0,0)\neq(\alpha,\beta)\in\mathcal{O}_f^2/T\mathcal{O}_f^2} \epsilon_{(T)}(\alpha,\beta)(g,z)$$

which is an immediate consequence of the uniformization theory of Drinfeld modules over $\mathbb{C}_\infty$. For every ideal $\mathfrak{n} = (n) \lhd A$ let $\Delta_\mathfrak{n}$ denote the modular form of weight q^2-1 given by the formula $\Delta_\mathfrak{n}(g,z) = \Delta(g\begin{pmatrix} n^{-1} & 0 \\ 0 & 1\end{pmatrix}, z)$. As the notation indicates $\Delta_\mathfrak{n}$ is independent of the choice of the generator $n \in \mathfrak{n}$. Finally let $E_\mathfrak{n} = r(\Delta/\Delta_\mathfrak{n})$. Since $\Delta/\Delta_\mathfrak{n}$ is a modular form of weight zero, i.e. it is a modular unit, the function $E_\mathfrak{n}$ is a $\mathbb{Z}$-valued harmonic form of level $\mathfrak{n}\infty$.

Proposition 5.8. *If T does not divide n then we have:*

$$E^0_\mathfrak{n}(1) = (q-1)q(q^{\deg(\mathfrak{n})}-1) \text{ and } E^*_\mathfrak{n}(1) = \frac{(q^2-1)(q-1)}{q}.$$

Proof. Every $\alpha \in \mathcal{O}_f/T\mathcal{O}_f$ is represented by a unique element of the constant field $\mathbb{F}_q$, which will be denoted by the same symbol by abuse of notation. For all such α and $z \in \mathbb{F}_q[T] \subset \mathbb{A}_f^*$ with $T \nmid z$ we have:

$$\zeta_{(T)}(\alpha, z^{-1}, s) = \sum_{\substack{0\neq p\in\mathbb{F}_q[T] \\ p\equiv\alpha z \bmod (zT)}} q^{-s\deg(p)}.$$

Because $p \equiv \alpha z \bmod (zT)$ holds if and only if there is a $r \in \mathbb{F}_q[T]$ with $p = \alpha z + zTr$, we have $\deg(p) = \deg(z) + 1 + \deg(r)$ in this case, unless $r = 0$ and $p = \alpha z$. Therefore

$$\begin{aligned}\zeta_{(T)}(\alpha, z^{-1}, s) =& (1-\rho(\alpha))q^{-s\deg(z)} + \sum_{k=0}^{\infty}(q-1)q^k q^{-s(\deg(z)+1+k)} \\ =& (1-\rho(\alpha))q^{-s\deg(z)} + \frac{(q-1)q^{-s(\deg(z)+1)}}{1-q^{1-s}}.\end{aligned}$$

For every $z \in \mathbb{F}_q[T]$ let z denote the unique idele whose finite component is z and its infinite component is 1, by abuse of notation. An immediate consequence of this equation and Proposition 5.2 is that the function $E_{(T)}(\alpha,\beta,\cdot,s)^0(z^{-1})$, originally defined for $\mathrm{Re}(s) > 1$ only, has a meromorphic continuation to the whole complex plane and

$$\lim_{s\to 0} E_{(T)}(\alpha,\beta,\cdot,s)^0(z^{-1}) = -\rho(\alpha)\rho(\beta) - q^{\deg(z)}(\frac{1}{q+1} - \rho(\alpha)),$$

using the fact that divisor of $\mathfrak{d}$ is in the anticanonical class, hence its degree is two. On the other hand the Limit Formula 4.8 and the description in Notation 5.7 implies that:

$$\begin{aligned} E^0_{\mathfrak{n}}(1) = & \sum_{(0,0)\neq(\alpha,\beta)\in\mathbb{F}_q^2} \lim_{s\longrightarrow 0}(E_{(T)}(\alpha,\beta,\cdot,s)^0(1) - E_{(T)}(\alpha,\beta,\cdot,s)^0(n^{-1})) \\ & -(q^2-1)\lim_{s\longrightarrow 0}(E_{(T)}(0,0,\cdot,s)^0(1) - E_{(T)}(0,0,\cdot,s)^0(n^{-1})) \\ = & \sum_{(0,0)\neq(\alpha,\beta)\in\mathbb{F}_q^2}(q^{\deg(n)}-1)(1-\rho(\alpha)) = (q-1)q(q^{\deg(\mathfrak{n})}-1). \end{aligned}$$

By Proposition 5.4 the function $E_{(T)}(\alpha,\beta,\cdot,s)^*(1)$ is a meromorphic function and:

$$E_{(T)}(\alpha,\beta,\cdot,0)^*(1) = -\sigma_{(T)}(\alpha,\beta,\mathfrak{d},-1).$$

By choosing an appropriate character τ, we may assume that $\mathfrak{d}$ any divisor of degree two, as every such divisor is linearly equivalent to the anticanonical class. In particular we may assume that $\mathfrak{d} = T_f^2$, which is in accordance with our previous assumptions. In this case:

$$S_{(T)}(\alpha,\mathfrak{d}) = \{0\neq p\in\mathbb{F}_q(T) | pT^2\in\alpha+T\mathbb{F}_q[T], p^{-1}\in T\mathbb{F}_q[T]\},$$

which is the one element set $\{\alpha T^{-2}\}$, if α is non-zero, and it is $\{\gamma T^{-1}|\gamma\in\mathbb{F}_q^*\}$, otherwise. Hence:

$$\sigma_{(T)}(\alpha,\beta,\mathfrak{d},-1) = \begin{cases} \frac{1}{q^2} & \text{, if } \alpha\neq 0, \\ -\frac{1}{q} & \text{, if } \alpha = 0 \text{ and } \beta\neq 0, \\ \frac{q-1}{q} & \text{, if } \alpha = 0 \text{ and } \beta = 0, \end{cases}$$

where in the second case we used the fact that the character is non-trivial on the set of elements $\gamma^{-1}\beta T$, where $\gamma\in\mathbb{F}_q^*$. As all Fourier coefficients $E_{(T)}(\alpha,\beta,\cdot,s)^*(n^{-1})$ are zero, because the divisor n^{-1} is not effective, we get:

$$\begin{aligned} E^*_{\mathfrak{n}}(1) = & -\sum_{(0,0)\neq(\alpha,\beta)\in\mathbb{F}_q^2}\sigma_{(T)}(\alpha,\beta,\mathfrak{d},-1) + (q^2-1)\sigma_{(T)}(0,0,\mathfrak{d},-1) \\ = & \frac{(q^2-1)(q-1)}{q}. \ \square \end{aligned}$$

Remark 5.9. The modular form $\Delta_{\mathfrak{n}}$ coincides with the function defined by Gekeler (see for example [8]), which can be seen by passing from the adelic description to the usual one. The result above is also proved in [8], but the argument applied there, unlike ours, can not be easily generalized. In particular the description of the quotient of the Bruhat-Tits tree by the full modular group (Proposition 3.9) is used which has no analogue in general.

6. Cuspidal harmonic forms annihilated by the Eisenstein ideal

Definition 6.1. Let $\mathfrak{n}$ be any ideal of A and let $H \subset GL_2(\mathbb{A}_f)$ be a compact double $\mathbb{K}_f(\mathfrak{n})$-coset. It is a disjoint union of finitely many right $\mathbb{K}_f(\mathfrak{n})$-cosets. Let R be a set of representatives of these cosets. For any function $u : GL_2(\mathbb{A}_f) \times \Omega \longrightarrow \mathbb{C}_\infty^*$ holomorphic in the second variable for each $g \in GL_2(\mathbb{A}_f)$, we define the function $T_H(u)$ by the formula:

$$T_H(u)(g) = \prod_{h \in R} u(gh).$$

If we assume that u is right $\mathbb{K}_f(\mathfrak{n})$-invariant then the function $T_H(u)$ is independent of the choice of R and $T_H(u)$ is holomorphic in the second variable for each $g \in GL_2(\mathbb{A}_f)$ as well. Moreover we have the identity:

$$r(T_H(u)) = T_H(r(u)),$$

where T_H also denotes the similarly defined linear operator on the set of right $\mathbb{K}_0(\mathfrak{n}\infty)$-invariant functions on $GL_2(\mathbb{A})$, slightly extending Definition 3.1. Let the symbol $T_{\mathfrak{m}}$ denote the operator T_H, if $H = H(\mathfrak{m}, \mathfrak{n}\infty) \cap GL_2(\mathbb{A}_f)$, where $\mathfrak{m} \lhd A$. Since we may choose the representative system $R(\mathfrak{m}, \mathfrak{n}\infty)$ to be a subset of $GL_2(\mathbb{A}_f)$, our new notation is compatible with the old one introduced in 3.1. Finally let $\mathfrak{p}$ be a prime ideal of A, and let $\pi \in \mathbb{A}_f^*$ be an idele such that $\pi\mathcal{O}_f = \mathfrak{p}$. The matrix $\begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix} \in GL_2(\mathbb{A}_f)$ introduced in the proof of Proposition 3.10 normalizes the subgroup $\mathbb{K}_0(\mathfrak{p}\infty)$, hence its double $\mathbb{K}_0(\mathfrak{n}\infty)$-coset as well as its double $\mathbb{K}_f(\mathfrak{n})$-coset consist of only one right coset. Let $W_{\mathfrak{p}}$ denote the corresponding operator.

The following lemma is also proved in [8], but we believe that our proof is simpler, and in a certain sense more revealing.

Lemma 6.2. *We have:*

$$W_{\mathfrak{p}}(E_{\mathfrak{p}}) = -E_{\mathfrak{p}} \text{ and } T_{\mathfrak{q}}(E_{\mathfrak{p}}) = (1 + q^{\deg(\mathfrak{q})})E_{\mathfrak{p}}$$

for every prime ideal $\mathfrak{q} \lhd A$ different from $\mathfrak{p}$. Moreover $E_{\mathfrak{p}}$ is an eigenvector of every Hecke operator $T_{\mathfrak{m}}$, with integral eigenvalue.

Proof. By the discussion above it is sufficient to prove the same for the modular unit $\Delta/\Delta_{\mathfrak{p}}$, up to a non-zero constant, because the van der Put derivative is zero on constant functions. Under the identification of Theorem 5.6 the modular unit $\Delta/\Delta_{\mathfrak{p}}$ corresponds to a nowhere zero rational function on the affine curve $Y_0(\mathfrak{p})$. The action of the operators $W_{\mathfrak{p}}$ and $T_{\mathfrak{m}}$ is just the usual action induced by the Atkin-Lehmer involution and the Hecke correspondence $T_{\mathfrak{m}}$, respectively. (See [6] for their definition and properties in this setting). The latter extend to correspondences on $X_0(\mathfrak{p})$, the unique non-singular projective curve which contains $Y_0(\mathfrak{p})$ as an open subvariety. The complement of $Y_0(\mathfrak{p})$ in $X_0(\mathfrak{p})$ consists of two geometric points, the cusps. These correspondences leave the group of divisors supported on the cusps invariant. In particular, the Atkin-Lehmer involution interchanges these two points, while the Hecke correspondence $T_{\mathfrak{q}}$, where $q \triangleleft A$ is a prime ideal different from $\mathfrak{p}$, maps them into themselves with multiplicity $1+q^{\deg(\mathfrak{q})}$. Since every nowhere zero rational function on the affine curve $Y_0(\mathfrak{p})$ is uniquely determined, up to a non-zero constant, by its divisor, which is of degree zero and is supported on the cusps, the claim now follows at once. □

Proposition 6.3. *A harmonic form $\phi \in \mathcal{H}(\mathfrak{p}, R)$ is cuspidal if any only if the integrals:*

$$\phi^0(1) = \int_{F\backslash\mathbb{A}} \phi(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix})d\mu(x) \text{ and}$$
$$\phi^\infty(1) = \int_{F\backslash\mathbb{A}} \phi(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix})d\mu(x)$$

are both zero.

Proof. The condition is clearly necessary. Also note that $\phi^\infty(1) = W_{\mathfrak{p}}(\phi)^0(1)$ for every $\phi \in \mathcal{H}(\mathfrak{p}, R)$, so the condition does not depend on the particular choice of π. In particular we may assume that all components π_v, where $v \triangleleft A$ is different from $\mathfrak{p}$, are actually equal to one. If we want to show that it is sufficient, we need to show that the integral

$$c(g, \phi) = \int_{F\backslash\mathbb{A}} \phi(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} g)d\mu(x)$$

is zero for every $g \in GL_2(\mathbb{A})$, if ϕ satisfies the condition of the claim. In order to do so, we first prove the lemma below. Let υ be a uniformizer in F_∞, as in Definition 3.4.

Lemma 6.4. *For every $g \in GL_2(\mathbb{A})$ and $\phi \in \mathcal{H}(\mathfrak{p}, R)$ the following holds:*

(*i*) *we have $c(g, \phi) = c(\gamma g k z, \phi)$, if $\gamma \in P(F)U(\mathbb{A})$, $k \in \mathbb{K}_0(\mathfrak{p}\infty)$ and $z \in Z(\mathbb{A})$,*

(*ii*) *we have $c(g, \phi) = -c(g\left(\begin{smallmatrix} 0 & 1 \\ \upsilon & 0 \end{smallmatrix}\right), \phi)$,*

(*iii*) *we have $c(g, \phi) = |\infty|^{-1} c(g\left(\begin{smallmatrix} \upsilon & 0 \\ 0 & 1 \end{smallmatrix}\right), \phi)$, if $g_\infty \in B(F_\infty)$.*

Proof. We first show (i). If $\gamma = \left(\begin{smallmatrix}\alpha & \beta\\ 0 & 1\end{smallmatrix}\right)$, then:

$$\begin{aligned}c(\gamma gkz,\phi) &= \int_{F\backslash\mathbb{A}}\phi(\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\gamma g)\mathrm{d}\mu(x)\\ &= \int_{F\backslash\mathbb{A}}\phi(\gamma\begin{pmatrix}1 & \alpha^{-1}x\\ 0 & 1\end{pmatrix}g)\mathrm{d}\mu(x) = c(g,\phi),\end{aligned}$$

using the right $\mathbb{K}_0(\mathfrak{p}\infty)Z(\mathbb{A})$-invariance and the left $GL_2(F)$-invariance of ϕ, as well as the fact that the map $x \mapsto \alpha^{-1}x$ leaves the Haar-measure μ of the group $F\backslash\mathbb{A}$ invariant for every $\alpha \in F^*$. Claim (ii) is an immediate consequence of the first condition in Definition 3.4. Assume now that $g_\infty = \left(\begin{smallmatrix}a & b\\ 0 & c\end{smallmatrix}\right)$. The final claim follows from the computation:

$$\begin{aligned}c(g,\phi) = c(g,T_\infty(\phi)) &= \sum_{\epsilon\in\mathbf{f}_\infty}\int_{F\backslash\mathbb{A}}\phi(\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}g\begin{pmatrix}1 & \epsilon\\ 0 & 1\end{pmatrix}\begin{pmatrix}\upsilon & 0\\ 0 & 1\end{pmatrix})\mathrm{d}\mu(x)\\ &= \sum_{\epsilon\in\mathbf{f}_\infty}\int_{F\backslash\mathbb{A}}\phi(\begin{pmatrix}1 & x+\frac{a}{c}\epsilon\\ 0 & 1\end{pmatrix}g\begin{pmatrix}\upsilon & 0\\ 0 & 1\end{pmatrix})\mathrm{d}\mu(x)\\ &= \frac{1}{|\infty|}c(g\begin{pmatrix}\upsilon & 0\\ 0 & 1\end{pmatrix},\phi),\end{aligned}$$

where we used Lemma 3.5. $\square$

Let us return to the proof of Proposition 6.3. By the Iwasawa decomposition we may write g as a product bk, where $b \in B(\mathbb{A})$ and $k \in GL_2(\mathcal{O})$. We may assume that b is a diagonal matrix with 1 in the lower left corner by multiplying g by a suitable element of $U(\mathbb{A})Z(\mathbb{A})$ on the left, according to Lemma 6.4. We may also assume that k_v is the identity matrix for all $v \in |X|$, different from $\mathfrak{p}$ and ∞, by multiplying g by a suitable element of $\mathbb{K}_0(\mathfrak{p}\infty)$ on the right, using again Lemma 6.4. Since A has class number 1, the equality $F^*\mathcal{O}_f^* = \mathbb{A}_f^*$ holds, hence we may even assume that g_v is the identity matrix for all $v \in |X|$, different from $\mathfrak{p}$ and ∞, by multiplying g by a suitable diagonal element of $GL_2(F)$ on the left and of $\mathbb{K}_0(\mathfrak{p}\infty)$ on the right. Moreover $GL_2(F_\infty) = B(F_\infty)\Gamma_\infty \cup B(F_\infty)\left(\begin{smallmatrix}0 & 1\\ \upsilon & 0\end{smallmatrix}\right)\Gamma_\infty$, hence claim (ii) of the lemma above implies that we may assume that g_∞ is a diagonal matrix with some power of υ in the upper right corner and 1 in the lower left corner, also repeating some of the arguments above. In this case (iii) of Lemma 6.4 can be used to reduce to the case when g_∞ is the identity matrix, too. Using the decomposition $GL_2(F_\mathfrak{p}) = B(F_\mathfrak{p})\Gamma_\mathfrak{p} \cup B(F_\mathfrak{p})\left(\begin{smallmatrix}0 & 1\\ \pi & 0\end{smallmatrix}\right)\Gamma_\mathfrak{p}$, where $\Gamma_\mathfrak{p} = \mathbb{K}_0(\mathfrak{p}\infty)_\mathfrak{p}$ is the Iwahori subgroup in $GL_2(F_\mathfrak{p})$, the same logic implies that $g_\mathfrak{p}$ may be assumed to be either the identity matrix or $\left(\begin{smallmatrix}0 & 1\\ \pi & 0\end{smallmatrix}\right)$. The proof is now complete. $\square$

Definition 6.5. Let $\mathcal{E}_0(\mathfrak{p},R)$ be the R-submodule of $\mathcal{H}_0(\mathfrak{p},R)$ of those cuspidal harmonic forms ϕ such that $T_\mathfrak{q}(\phi) = (1+q^{\deg(\mathfrak{q})})\phi$ for each closed point

$\mathfrak{q}$ of X, different from $\mathfrak{p}$ and ∞. By Theorem 3.12 the R-module $\mathcal{E}_0(\mathfrak{p},R)$ is isomorphic to an ideal $\mathfrak{a}\lhd R$ via the map $\phi\mapsto\phi^*(1)$. Let $d=\deg(\mathfrak{p})$ denote the degree of $\mathfrak{p}$.

Theorem 6.6. *For every coefficient ring R the map*

$$\mathcal{E}_0(\mathfrak{p},R)\longrightarrow R,\quad \phi\mapsto\phi^*(1)$$

is an isomorphism onto $R[N(\mathfrak{p})]$, if d is odd, and is an isomorphism onto $R[2N(\mathfrak{p})]$, if d is even.

Proof. Define the harmonic form $e_{\mathfrak{p}}\in\mathcal{H}(\mathfrak{p},\mathbb{Z}\langle\frac{1}{q^2-1}\rangle)$ by the formula:

$$e_{\mathfrak{p}}=\begin{cases}\frac{E_{\mathfrak{p}}}{(q-1)^2} & \text{, if } d \text{ is odd,}\\ \frac{E_{\mathfrak{p}}}{(q-1)^2(q+1)} & \text{, if } d \text{ is even.}\end{cases}$$

For every $\alpha\in R[q+1]$ let $H(\alpha)$ again denote the unique R-valued harmonic form of level ∞ with $H(\alpha)^0(1)=\alpha$, just as in the proof of Proposition 3.8. First we are going to show the following

Lemma 6.7. *The harmonic form $e_{\mathfrak{p}}$ is integer-valued.*

Proof. By Proposition 5.8 we have $e^0_{\mathfrak{p}}(1)=N(\mathfrak{p})$ and $e^*_{\mathfrak{p}}(1)=\frac{q+1}{q}$, if d is odd, and $e^*_{\mathfrak{p}}(1)=\frac{1}{q}$, if d is even. By Lemma 6.2 the form $e_{\mathfrak{p}}$ is also an eigenvector for the Hecke operator $T_{\mathfrak{m}}$, where where $\mathfrak{m}$ is any prime ideal of A, with integral eigenvalue. Hence $e^*_{\mathfrak{p}}(\mathfrak{m})\in\mathbb{Z}\langle\frac{1}{q}\rangle$ for any effective divisor $\mathfrak{m}$, arguing the same way as we did in the proof of Corollary 3.3. Moreover $e^0_{\mathfrak{p}}(y)\in\mathbb{Z}\langle\frac{1}{q}\rangle$ for any $y\in\mathbb{A}^*$ using that $\mathrm{Pic}(X)=\mathbb{Z}$ via the degree map and part *(iii)* of Lemma 6.4. The Fourier expansion formula (Proposition 2.10) implies that we must have $e_{\mathfrak{p}}\in\mathcal{H}(\mathfrak{p},\mathbb{Z}\langle\frac{1}{q}\rangle)$, hence $e_{\mathfrak{p}}$ is an integer valued harmonic form. $\square$

Let $e_{\mathfrak{p}}$ denote the image of this harmonic form in $\mathcal{H}(\mathfrak{p},R)$ for any coefficient ring R with respect to the functorial homomorphism $\mathcal{H}(\mathfrak{p},\mathbb{Z})\longrightarrow\mathcal{H}(\mathfrak{p},R)$, by abuse of notation.

Lemma 6.8. *For any $\alpha\in R[q+1]$ and $\beta\in R$ the harmonic form $H(\alpha)+\beta e_{\mathfrak{p}}$ lies in $\mathcal{E}_0(\mathfrak{p},R)$ if and only if the equations $\alpha=-\beta N(\mathfrak{p})$ and $\alpha=(-1)^d\beta N(\mathfrak{p})$ hold.*

Proof. By Lemma 6.2 the form $e_{\mathfrak{p}}$ is an eigenvector for the Hecke operator $T_{\mathfrak{q}}$, where $\mathfrak{q}$ is a prime ideal different from $\mathfrak{p}$, with $q^{\deg(\mathfrak{q})}+1$ as eigenvalue. The degree of the determinant of every element of the set $R(\mathfrak{q},\mathfrak{p}\infty)$ is $\deg(\mathfrak{q})$ for every $\mathfrak{q}$ prime of A different from $\mathfrak{p}$, hence $T_{\mathfrak{q}}(H(\alpha))^0(1)=(q^{\deg(\mathfrak{q})}+1)(-1)^{\deg(\mathfrak{q})}\alpha$. If $\deg(\mathfrak{q})$ is odd, then $q+1$ divides $q^{\deg(\mathfrak{q})}+1$, hence the expression above is equal to $0=(q^{\deg(\mathfrak{q})}+1)\alpha$ in this case. In particular $H(\alpha)$ is an eigenvector for the Hecke operator $T_{\mathfrak{q}}$ with $q^{\deg(\mathfrak{q})}+1$ as eigenvalue, too. Therefore it is sufficient to prove that $H(\alpha)+\beta e_{\mathfrak{p}}\in\mathcal{H}_0(\mathfrak{p},R)$ if and only if the equations hold in the claim above. Note that $H(\alpha)^{\infty}(1) = (-1)^d\alpha$, as every matrix of the form $\begin{pmatrix} \pi x & 1 \\ \pi & 0 \end{pmatrix}$ has determinant π, which has degree d. By Lemma 6.2 we have $W_{\mathfrak{p}}(e_{\mathfrak{p}}) = -e_{\mathfrak{p}}$, hence $e^{\infty}_{\mathfrak{p}}(1) = -e^0_{\mathfrak{p}}(1) = -N(\mathfrak{p})$. The claim now follows from Proposition 6.3. $\square$

Let's start the proof proper of Theorem 6.6. First assume that d is even. In this case every $\phi \in \mathcal{E}_0(\mathfrak{p}, R)$ can be written uniquely of the form $\phi = q\phi^*(1)e_{\mathfrak{p}} + H(\alpha)$, for some $\alpha \in R[q+1]$. By Lemma 6.8 we must have $N(\mathfrak{p})\phi^*(1) = \alpha/q = -\alpha/q$, hence $2N(\mathfrak{p})\phi^*(1) = 0$. On the other hand let $\beta \in R[2N(\mathfrak{p})]$ be arbitrary. First note that $R[2] \subseteq R[q+1]$. If q is even, then 2 is invertible in R, hence $R[2] = 0$. If q is odd, then 2 divides $q+1$, hence $R[2] \subseteq R[q+1]$. Therefore $\alpha = qN(\mathfrak{p})\beta \in R[q+1]$, so $H(\alpha)$ is well-defined. By Lemma 6.8 we have $q\beta e_{\mathfrak{p}} + H(\alpha) \in \mathcal{E}_0(\mathfrak{p}, R)$, and its image under the map of the claim is β. Now assume that d is odd. Let $\tilde{R}$ be a discrete valuation ring and let $\mathfrak{a} \triangleleft \tilde{R}$ be an ideal such that $R = \tilde{R}/\mathfrak{a}$. Define the coefficient ring R' as the quotient $\tilde{R}/(q+1)\mathfrak{a}$. The map $R \longrightarrow R'$ given by the rule $x \mapsto (q+1)x$ maps bijectively onto the ideal $(q+1) \triangleleft R'$. In particular for every $\phi \in \mathcal{E}_0(\mathfrak{p}, R)$ we have $(q+1)\phi \in \mathcal{E}_0(\mathfrak{p}, R')$, and the latter can be written of the form $(q+1)\phi = q\beta e_{\mathfrak{p}} + H(\alpha)$, for some $\alpha \in R'[q+1]$ and $\beta \in R'$ which maps to $\phi^*(1)$ under the canonical surjection $R' \longrightarrow R$. Applying Lemma 6.8 the the coeffient ring R' we get that we must have $N(\mathfrak{p})\beta = -\alpha/q$, hence $(q+1)N(\mathfrak{p})\beta = 0$. The latter is equivalent to $\phi^*(1) \in R[N(\mathfrak{p})]$. On the other hand let $\beta \in R[N(\mathfrak{p})]$ be arbitrary. For any lift $\beta' \in R'$ with respect to the natural surjection we have $\beta' \in R'[(q+1)N(\mathfrak{p})]$. Therefore $\alpha = -qN(\mathfrak{p})\beta' \in R'[q+1]$, so $H(\alpha)$ is well-defined. By Lemma 6.8 we have $q\beta' e_{\mathfrak{p}} + H(\alpha) \in \mathcal{E}_0(\mathfrak{p}, R')$, and its image under the map of the claim is $(q+1)\beta$. If we show that all values of this harmonic form lie in the ideal $(q+1)$, then we have also shown the surjectivity of the map of the claim in case of the ring R. The latter would follow if we proved that all Fourier coefficients of this harmonic form lie in the ideal $(q+1)$, by Proposition 2.10. The constant terms are obviously zero. By Lemma 3.11 the $\mathfrak{m}$-th coefficient is equal to $q\beta' e^*_{\mathfrak{p}}(\mathfrak{m})$ which lies in $(q+1)$. $\square$

Corollary 6.9. *For every natural number n relatively prime to p the module $\mathcal{E}_0(\mathfrak{p}, \mathbb{Z}/n\mathbb{Z})$ is isomorphic to $\mathbb{Z}/n\mathbb{Z}[N(\mathfrak{p})]$, if d is odd, and it is isomorphic to $\mathbb{Z}/n\mathbb{Z}[2N(\mathfrak{p})]$, if d is even.*

Proof. Since $\mathcal{E}_0(\mathfrak{p}, \mathbb{Z}/n\mathbb{Z}) = \oplus\mathcal{E}_0(\mathfrak{p}, \mathbb{Z}/k\mathbb{Z})$, where k runs through the set of components of the primary factorization of n, we may immediately reduce to the case when n is the power of a prime l. In this case the ring $\mathbb{Z}/n\mathbb{Z}$ is still not a coefficient ring in general, but it is close to it. Let $\tilde{R}$ denote the unique unramified extension of $\mathbb{Z}_l$ we get by adjoining the p-th roots of unity. The ring $R = \tilde{R}/n\tilde{R}$ is a coefficient ring which is a free $\mathbb{Z}/n\mathbb{Z}$-module. It will be sufficient to show that the map of Theorem 6.6 maps $\mathcal{E}_0(\mathfrak{p}, \mathbb{Z}/n\mathbb{Z})$ surjectively onto $\mathbb{Z}/n\mathbb{Z}[N(\mathfrak{p})]$, if d is odd, and onto $\mathbb{Z}/n\mathbb{Z}[2N(\mathfrak{p})]$, if d is even. The latter follows from the following simple observation: for every $\beta \in R[N(\mathfrak{p})]$, if d is odd, and for every $\beta \in R[2N(\mathfrak{p})]$, if d is even, the unique form $\phi \in \mathcal{E}_0(\mathfrak{p}, R)$

with the property $\phi^*(1) = \beta$ takes values in the $\mathbb{Z}/n\mathbb{Z}$-module generated by β, which is an immediate consequence of the formula for ϕ it terms of $e_{\mathfrak{p}}$ and $H(\alpha)$ in the proof above. $\square$

Remark 6.10. Another interesting consequence of our analysis is the congruence:

$$\frac{E_{\mathfrak{p}}}{(q-1)^2} \equiv H(N(\mathfrak{p})) \mod (q+1),$$

which holds for every prime $\mathfrak{p}$ of odd degree. In particular the residue of the form on the left modulo $q+1$ is invariant under the full modular group.

7. The Abel-Jacobi map

Definition 7.1. Let $\Gamma_0(\mathfrak{p})$ denote $GL_2(A) \cap \mathbb{K}_f(\mathfrak{p})$. This group also acts on Ω via Möbius transformations. By Theorem 5.6 the quotient curve $\Gamma_0(\mathfrak{p})\backslash\Omega$ is $Y_0(\mathfrak{p})$. Let moreover $\Gamma_0(\mathfrak{p})_{\text{ab}} = \Gamma_0(\mathfrak{p})/[\Gamma_0(\mathfrak{p}), \Gamma_0(\mathfrak{p})]$ be the abelianization of $\Gamma_0(\mathfrak{p})$, and let $\overline{\Gamma}_0(\mathfrak{p}) = \Gamma_0(\mathfrak{p})_{\text{ab}}/(\Gamma_0(\mathfrak{p})_{\text{ab}})_{\text{tors}}$ be its maximal torsion-free quotient. For each $\gamma \in \Gamma_0(\mathfrak{p})$ let $\overline{\gamma}$ denote its image in $\overline{\Gamma}_0(\mathfrak{p})$. We say that a meromorphic function θ on Ω is a theta function for $\Gamma_0(\mathfrak{p})$ with automorphy factor $\phi \in \text{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$, if $\theta(\gamma z) = \phi(\overline{\gamma})\theta(z)$ for all $z \in \Omega$ and $\gamma \in \Gamma_0(\mathfrak{p})$. If $D = P_1 + \cdots + P_r - Q_1 + \cdots - Q_r \in \text{Div}_0(\Omega)$ is a divisor of degree zero on Ω, define the function

$$\theta(z; D) = \prod_{\gamma \in \Gamma_0(\mathfrak{p})} \frac{(z - \gamma P_1) \cdots (z - \gamma P_r)}{(z - \gamma Q_1) \cdots (z - \gamma Q_r)}.$$

This infinite product converges and defines a meromorphic function on Ω.

Proposition 7.2. (*i*) *The function $\theta(z; D)$ is a theta function for $\Gamma_0(\mathfrak{p})$.*
(*ii*) *Given $\alpha \in \Gamma_0(\mathfrak{p})$, the theta function $\theta_{\overline{\alpha}}(z) = \theta(z; (w) - (\alpha w))$ is holomorphic, does not depend on the choice of $w \in \mathbb{C}_\infty$, and depends only on the image $\overline{\alpha}$ of α in $\overline{\Gamma}_0(\mathfrak{p})$.*

Proof. See [11], pages 62-67. Part (*ii*) is (*iv*) of Theorem 5.4.1 of [11], page 65. $\square$

Notation 7.3. Let ϕ_D be the automorphy factor of $\theta(z; D)$. By the above the value $c_\alpha(\beta) = \phi_{(z)-(\alpha z)}(\beta)$ does not depend on the choice of $z \in \mathbb{C}_\infty$, and depends only on the image of α and β in $\overline{\Gamma}_0(\mathfrak{p})$. Let $j : \overline{\Gamma}_0(\mathfrak{p}) \to \mathcal{H}(\mathfrak{p}, \mathbb{Z})$ denote the map which assigns $r(\theta_{\overline{\alpha}}(z))$ to $\overline{\alpha}$. It is a homomorphism by (*v*) of Theorem 5.4.1 of the paper quoted above.

The following result will play a crucial role.

Theorem 7.4. *The homomorphism j is an isomorphism onto $\mathcal{H}_0(\mathfrak{p}, \mathbb{Z})$.*

Proof. By Corollary 5.6.4 of [11], page 69 the image of this map lies in $\mathcal{H}_0(\mathfrak{p}, \mathbb{Z})$. The map is an isomorphism by Theorem 3.3 of [10], page 702. $\square$

Proposition 7.5. *The assignment $\alpha \mapsto c_\alpha$ defines a map*

$$c : \overline{\Gamma}_0(\mathfrak{p}) \to \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), F_\infty^*) \subset \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*),$$

which is injective and has discrete image.

Proof. See [11], pages 67-70. □

Definition 7.6. Let

$$\Phi_{AJ} : \mathrm{Div}_0(\Omega) \to \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$$

be the map which associates to the degree zero divisor D the automorphy factor ϕ_D. Let $\overline{\Gamma}_0(\mathfrak{p})$ also denote its own image in $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$ with respect to c by abuse of notation. Given a divisor D of degree zero on the curve $Y_0(\mathfrak{p})$, let $\widetilde{D}$ denote an arbitrary lift to a degree zero divisor on the Drinfeld upper half plane. The automorphy factor $\phi_{\widetilde{D}}$ depends on the choice of $\widetilde{D}$, but its image in $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)/\overline{\Gamma}_0(\mathfrak{p})$ depends only on D. Thus Φ_{AJ} induces a map $\mathrm{Div}_0(Y_0(\mathfrak{p})(\mathbb{C}_\infty)) \to \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)/\overline{\Gamma}_0(\mathfrak{p})$, which we also denote by Φ_{AJ} by abuse of notation.

Theorem 7.7. *The map*

$$\Phi_{AJ} : \mathrm{Div}_0(Y_0(\mathfrak{p})(\mathbb{C}_\infty)) \to \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)/\overline{\Gamma}_0(\mathfrak{p})$$

defined above is trivial on the group of principal divisors of $X_0(\mathfrak{p})$, and induces a $\mathrm{Gal}(\mathbb{C}_\infty|F_\infty)$-equivariant identification of the $\mathbb{C}_\infty$-rational points of the Jacobian $J_0(\mathfrak{p})$ of $X_0(\mathfrak{p})$ with the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)/\overline{\Gamma}_0(\mathfrak{p})$.

Proof. See [11], pages 77-80. □

Definition 7.8. Recall the Hecke correspondence $T_\mathfrak{q}$ on the curve $X_0(\mathfrak{p})$ for every prime $\mathfrak{q}$ different from $\mathfrak{p}$, which we introduced in the proof of Lemma 6.2. It induces an endomorphism of the Jacobian $J_0(\mathfrak{p})$ by functoriality, which will be denoted by $T_\mathfrak{q}$ by the usual abuse of notation. Our next task is to describe this action in terms of the isomorphism of Theorem 7.7.

Theorem 7.9. *For every prime $\mathfrak{q} \triangleleft A$, different from $\mathfrak{p}$, there is a unique endomorphism $T_\mathfrak{q}$ of the rigid analytic torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$, which leaves the lattice $\overline{\Gamma}_0(\mathfrak{p})$ invariant, and makes the diagram:*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \overline{\Gamma}_0(\mathfrak{p}) & \xrightarrow{c} & \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*) & \xrightarrow{\Phi_{AJ}} & J_0(\mathfrak{p}) & \longrightarrow & 0 \\ & & {\scriptstyle T_\mathfrak{q}}\downarrow & & {\scriptstyle T_\mathfrak{q}}\downarrow & & {\scriptstyle T_\mathfrak{q}}\downarrow & & \\ 0 & \longrightarrow & \overline{\Gamma}_0(\mathfrak{p}) & \xrightarrow{c} & \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*) & \xrightarrow{\Phi_{AJ}} & J_0(\mathfrak{p}) & \longrightarrow & 0 \end{array}$$

commutative. Moreover the map $j : \overline{\Gamma}_0(\mathfrak{p}) \to \mathcal{H}_0(\mathfrak{p}, \mathbb{Z})$ is equivariant with respect to this action on $\overline{\Gamma}_0(\mathfrak{p})$ and the action of the Hecke operator $T_\mathfrak{q}$ on $\mathcal{H}_0(\mathfrak{p}, \mathbb{Z})$.

Proof. The first claim is stated in 9.4 of [11], page 86. By definition, the action of $T_{\mathfrak{q}}$ on $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$ is the adjoint of the action of $T_{\mathfrak{q}}$ on $\overline{\Gamma}_0(\mathfrak{p})$ given by the formula 9.3.1 of the same paper on page 85. On the same page Proposition 9.3.3 states that the lattice $\overline{\Gamma}_0(\mathfrak{p})$ is invariant with respect to $T_{\mathfrak{q}}$, and its action is given by this formula. The fact that Φ_{AJ} is equivariant is an immediate consequence of its construction. The second claim is the content of Lemma 9.3.2 of [11], page 85. □

Definition 7.10. Let $\mathbb{T}(\mathfrak{p})$ denote the commutative algebra with unity generated by the endomorphisms $T_{\mathfrak{q}}$ of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$, where $\mathfrak{q} \triangleleft A$ is again any prime ideal different from $\mathfrak{p}$. Let $\mathfrak{E}(\mathfrak{p})$ denote the ideal of $\mathbb{T}(\mathfrak{p})$ generated by the elements $T_{\mathfrak{q}} - q^{\deg(\mathfrak{q})} - 1$, where $\mathfrak{q} \neq \mathfrak{p}$ is any prime. The algebra $\mathbb{T}(\mathfrak{p})$ will be called Hecke algebra and $\mathfrak{E}(\mathfrak{p})$ is its Eisenstein ideal, although these differ slightly from the usual definition, since they do not involve the Atkin-Lehmer operator. The latter will play no role in what follows. Let l be any prime ($l = p$ allowed): we define the $\mathbb{Z}_l$-algebra $\mathbb{T}_l(\mathfrak{p})$ as the tensor product $\mathbb{T}(\mathfrak{p}) \otimes \mathbb{Z}_l$. Let $\mathfrak{E}_l(\mathfrak{p})$ denote the ideal generated by the Eisenstein ideal in $\mathbb{T}_l(\mathfrak{p})$, which we will also call the Eisenstein ideal by slight abuse of terminology. We say that a prime number l is an Eisenstein prime if $l \neq p$ and the ideal $\mathfrak{E}_l(\mathfrak{p})$ is proper in $\mathbb{T}_l(\mathfrak{p})$. For any prime l different from p the l-adic Tate module of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$ will be denoted by $T_l(\mathfrak{p})$: it is a $\mathbb{T}_l(\mathfrak{p})$-module.

Proposition 7.11. *The following holds:*

(i) *the algebra $\mathbb{T}(\mathfrak{p})$ is a finitely generated, free $\mathbb{Z}$-module,*
(ii) *the $\mathbb{T}(\mathfrak{p})$-module $\overline{\Gamma}_0(\mathfrak{p})$ is faithful,*
(iii) *the $\mathbb{T}(\mathfrak{p})$-module $J_0(\mathfrak{p})$ is faithful,*
(iv) *the $\mathbb{T}_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$-module $\mathcal{H}_0(\mathfrak{p}, \mathbb{Q}_l)$ is free of rank one,*
(v) *the $\mathbb{T}_l(\mathfrak{p})$-module $T_l(\mathfrak{p})$ is locally free of rank one,*
(vi) *there is a canonical surjection $\mathbb{Z}_l/2N(\mathfrak{p})\mathbb{Z}_l \to \mathbb{T}_l(\mathfrak{p})/\mathfrak{E}_l(\mathfrak{p})$,*

where we also assume that $l \neq p$ in the last two claims.

Proof. Claim (i) is an immediate consequence of claim (ii), since the latter implies that $\mathbb{T}(\mathfrak{p})$ is a subalgebra of the endomorphism ring of a finitely generated, free $\mathbb{Z}$-module. The latter follows from the general fact that rigid analytic endomorphisms of algebraic tori are algebraic, so they act faithfully on any Zariski-dense invariant subset. Since Φ_{AJ} injects $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathcal{O}_\infty^*)$ into $J_0(\mathfrak{p})(F_\infty)$, the third claim also follows by the same token. By a classical theorem of Harder the elements of $\mathcal{H}_0(\mathfrak{p}, \mathbb{Q}_l)$ are supported on a finite set in $GL_2(F)\backslash GL_2(\mathbb{A})/\mathbb{K}_0(\mathfrak{p}\infty)Z(\mathbb{A})$, so the latter is a finite dimensional $\mathbb{Q}_l$-vectorspace, and $\mathcal{H}_0(\mathfrak{p}, \mathbb{Q}_l) = \mathcal{H}_0(\mathfrak{p}, \mathbb{Z}) \otimes \mathbb{Q}_l$. Therefore it is a faithful $\mathbb{T}_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$-module via the map j by claim (ii). As it is well known, the action of Hecke operators on $\mathcal{H}_0(\mathfrak{p}, \mathbb{Q}_l)$ is semisimple, hence the algebra $\mathbb{T}_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ itself is semisimple. By the strong multiplicity one result (Theorem 3.12) every irreducible module of $\mathbb{T}_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ has multiplicity one in $\mathcal{H}_0(\mathfrak{p}, \mathbb{Q}_l)$, so this module is free of rank one, as claim (iv) states.
As we already noted in the proof of Theorem 7.9, the action of the Hecke

algebra $\mathbb{T}(\mathfrak{p})$ on $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$ is the adjoint of the action of $\mathbb{T}(\mathfrak{p})$ on $\overline{\Gamma}_0(\mathfrak{p}) = \mathcal{H}_0(\mathfrak{p}, \mathbb{Z})$, so $\mathcal{H}_0(\mathfrak{p}, \mathbb{Z}_l) = \mathcal{H}_0(\mathfrak{p}, \mathbb{Z}) \otimes \mathbb{Z}_l$ is the $\mathbb{Z}_l$-dual of $T_l(\mathfrak{p})$. In particular $T_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ is a free $\mathbb{T}_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$-module. Since $T_l(\mathfrak{p})$ is a finitely generated, free $\mathbb{Z}_l$-module, it is a finitely generated module over $\mathbb{T}_l(\mathfrak{p})$. Hence it will be sufficient to prove that $T_l(\mathfrak{p})/\mathfrak{m}T_l(\mathfrak{p})$ is a free module of rank one over $\mathbf{k}_\mathfrak{m} = \mathbb{T}_l(\mathfrak{p})/\mathfrak{m}$ by the Nakayama lemma, where $\mathfrak{m} \triangleleft \mathbb{T}_l(\mathfrak{p})$ is any proper maximal ideal, in order to conclude claim (v). Its dimension is at least one over $\mathbf{k}_\mathfrak{m}$, since the module $T_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ is free of rank one over $\mathbb{T}_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$. For any ring R let $\mathcal{H}_{00}(\mathfrak{p}, R)$ denote the image of $\mathcal{H}_0(\mathfrak{p}, \mathbb{Z}) \otimes R$ in $\mathcal{H}_0(\mathfrak{p}, R)$ with respect to the functorial map induced by the canonical homomorphism $\mathbb{Z} \longrightarrow R$. Since l is an element of $\mathfrak{m}$, the $\mathbb{Z}_l$-duality between $\mathcal{H}_0(\mathfrak{p}, \mathbb{Z}_l)$ and $T_l(\mathfrak{p})$ induces a $\mathbb{F}_l$-duality between $T_l(\mathfrak{p})/\mathfrak{m}T_l(\mathfrak{p})$ and the submodule of $\mathcal{H}_{00}(\mathfrak{p}, \mathbb{F}_l)$ annihilated by the ideal $\mathfrak{m}$. In general, for any ring R and faithfully flat extension R' of R the natural map $\mathcal{H}_{00}(\mathfrak{p}, R) \otimes_R R' \longrightarrow \mathcal{H}_{00}(\mathfrak{p}, R')$ is an isomorphism by the theorem of Harder quoted above. This implies in particular that submodule of $\mathcal{H}_{00}(\mathfrak{p}, \mathbb{F}_l)$ annihilated by the ideal $\mathfrak{m}$ is a $\mathbf{k}_\mathfrak{m}$ sub-vectorspace of the space of elements of $\mathcal{H}_{00}(\mathfrak{p}, \mathbf{k}_\mathfrak{m})$ which are simultaneous eigenvectors for the operators $T_\mathfrak{q}$ with eigenvalue $T_\mathfrak{q}$ mod $\mathfrak{m}$. Let $\mathbf{l}_\mathfrak{m}$ be a finite extension of $\mathbf{k}_\mathfrak{m}$ which is also a coefficient ring. The eigenspace above tensored with $\mathbf{l}_\mathfrak{m}$ injects into the similar eigenspace of $\mathcal{H}_{00}(\mathfrak{p}, \mathbf{l}_\mathfrak{m})$, which is at most one dimensional over $\mathbf{l}_\mathfrak{m}$ by Theorem 3.12. Claim (v) is proved.

Finally let us concern ourselves with the proof of claim (vi). It is clear from the definition that every generator $T_\mathfrak{q}$ of $\mathbb{T}_l(\mathfrak{p})$ is congruent to an element of $\mathbb{Z}_l$ modulo the Eisenstein ideal, so the natural inclusion of $\mathbb{Z}_l$ in $\mathbb{T}_l(\mathfrak{p})$ induces a surjection $\mathbb{Z}_l \longrightarrow \mathbb{T}_l(\mathfrak{p})/\mathfrak{E}_l(\mathfrak{p})$. If this map is also injective, then the Eisenstein ideal generates a non-trivial ideal in $\mathbb{T}_l(\mathfrak{p}) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$. This implies, by claim (iv), that there is a non-zero harmonic form in $\mathcal{H}_0(\mathfrak{p}, \overline{\mathbb{Q}}_l)$ which is annihilated by the Eisenstein ideal. But this is impossible by Theorem 6.6. Therefore the map above induces an isomorphism $\mathbb{Z}_l/N\mathbb{Z}_l \longrightarrow \mathbb{T}_l(\mathfrak{p})/\mathfrak{E}_l(\mathfrak{p})$ for some non-zero $N \in \mathbb{N}$. By claim (v) the module $T_l(\mathfrak{p})/\mathfrak{E}_l(\mathfrak{p})T_l(\mathfrak{p})$ is free of rank one over $\mathbb{Z}_l/N\mathbb{Z}_l$, therefore the $\mathbb{Z}_l$-duality between $\mathcal{H}_0(\mathfrak{p}, \mathbb{Z}_l)$ and $T_l(\mathfrak{p})$ induces a $\mathbb{Z}_l/N\mathbb{Z}_l$-duality between $T_l(\mathfrak{p})/\mathfrak{E}_l(\mathfrak{p})T_l(\mathfrak{p})$ and the module $\mathcal{H}_{00}(\mathfrak{p}, \mathbb{Z}_l/N\mathbb{Z}_l) \cap \mathcal{E}_0(\mathfrak{p}, \mathbb{Z}_l/N\mathbb{Z}_l)$. The cardinality of the latter must divide $2N(\mathfrak{p})$ by Corollary 6.9, so does the cardinality of the former, because they are equal, by duality. □

An important consequence of claim (iii) above is that $\mathbb{T}(\mathfrak{p})$ may be identified with a subalgebra of the endomorphism ring of the abelian variety $J_0(\mathfrak{p})$, which we will do from now on. Also note that the factor 2 is only necessary in (vi) when $l = 2$ and d is odd.

Definition 7.12. Let $\mathbb{F}$ be local field of characteristic p and let $\mathbb{O}$, r denote its discrete valuation ring and the cardinality of its residue field, respectively. Recall that an abelian variety A defined over $\mathbb{F}$ is said to have multiplicative reduction if the connected component A_0 of the identity in the special fiber of its Néron model $\mathcal{A}$ over $\mathbb{O}$ is a torus. We also say that the abelian variety A

has totally split multiplicative reduction if it has multiplicative reduction and A_0 is a split torus.

Lemma 7.13. (*i*) *If A has multiplicative reduction then the p-primary torsion subgroup $A(\mathbb{F})[p^\infty]$ injects into the group of connected components of the special fiber of $\mathcal{A}$.*
(*ii*) *If A has totally split multiplicative reduction then the exponent of the largest torsion subgroup of $A(\mathbb{F})$ mapping into the connected component A_0 under the specialization map divides $r-1$.*

Proof. While we prove claim (*i*) we may take an unramified extension of $\mathbb{F}$, which will be denoted by the same letter, such that A_0 becomes a split torus, since it commutes with the formation of Néron models. In this case A has a rigid analytic uniformization by a torus $\mathbb{G}_m^n$. The subgroup of $A(\mathbb{F})$ mapping into A_0 under the specialization map is isomorphic to $(\mathbb{O}^*)^n$ in $(\mathbb{F}^*)^n = \mathbb{G}_m^n(\mathbb{F})$ via the uniformization map. Since $\mathbb{F}$ has characteristic p, the group $(\mathbb{O}^*)^n$ has no p-torsion. Claim (*i*) is now clear. The other half of the lemma also follows by the same reasoning as the torsion of $(\mathbb{O}^*)^n$ is $(\mathbb{F}_r^*)^n$. □

Let $M_0(\mathfrak{p})$ denote the coarse moduli of Drinfeld modules over A with Hecke $\mathfrak{p}$-level structure in the sense introduced by Katz and Mazur (see Definition 3.4 of [13], page 100). It is known that $M_0(\mathfrak{p})$ is a model of $Y_0(\mathfrak{p})$ over the spectrum of A which means that its generic fiber is canonically isomorphic to $Y_0(\mathfrak{p})$.

Proposition 7.14. *The model $M_0(\mathfrak{p})$ is contained in a scheme $\overline{M}_0(\mathfrak{p})$ which has the following properties:*

(*i*) *the scheme $\overline{M}_0(\mathfrak{p})$ is proper and flat over Spec(A),*
(*ii*) *it has good reduction over all primes $\mathfrak{q}$ different from $\mathfrak{p}$,*
(*iii*) *it has stable reduction over $\mathfrak{p}$ with two components which are rational curves over $\mathbf{f}_\mathfrak{p}$ and intersect transversally in $N(\mathfrak{p})$ points,*
(*iv*) *it is a model of $X_0(\mathfrak{p})$ over the spectrum of A,*
(*v*) *the scheme $\overline{M}_0(\mathfrak{p})$ is either regular or has a singularity of type A_q over $\mathbf{f}_\mathfrak{p}$.*

Proof. See 5.1-5.8 of [6], pages 229-233. □

Corollary 7.15. *The group $J_0(\mathfrak{p})(F)$ has no p-primary torsion.*

Proof. According to a classical theorem of Raynauld (see Proposition 1.20 of [1], page 219) the connected component of the special fiber of the Néron model over $\mathbb{O}$ of the Jacobian of any regular curve defined over $\mathbb{F}$ is isomorphic to the Picard group scheme of divisors of total degree zero of the special fiber of a regular, proper model of the curve over the spectrum of $\mathbb{O}$. If we set $\mathbb{F} = F_\mathfrak{p}$ then the curve $X_0(\mathfrak{p})$ has $\mathbb{F}$-rational points, namely the cusps. By Proposition 7.14 it has a regular, proper model over the spectrum of $\mathcal{O}_\mathfrak{p}$ such that each component in the special fiber is a rational curve and they intersect transversally. Hence $J_0(\mathfrak{p})$ has multiplicative reduction at $\mathfrak{p}$. According to Lemma 5.9 and Proposition 5.10 of [6], page 234, the order of the group of connected components of the Néron model of $J_0(\mathfrak{p})$ is $N(\mathfrak{p})$. The latter is proved the same way as the corresponding result for elliptic modular curves (see Theorem A.1 of the Appendix to [14], page 173) as it uses the description of the group of components by the intersection matrix of the special fiber, again due to Raynaud. Since $N(\mathfrak{p})$ is relatively prime to p, the claim now follows from Lemma 7.13. □

Lemma 7.16. *The torsion subgroup $\mathcal{T}(\mathfrak{p})$ of $J_0(\mathfrak{p})(F)$ is annihilated by the Eisenstein ideal $\mathfrak{E}(\mathfrak{p})$.*

Proof. For the sake of simple notation let $J_0(\mathfrak{p})$ denote the Néron model of the Jacobian over X, too. Since $J_0(\mathfrak{p})$ has good reduction over all primes $\mathfrak{q}$ different from $\mathfrak{p}$, the reduction map injects $\mathcal{T}(\mathfrak{p})$ into $J_0(\mathfrak{p})(\mathbf{f}_\mathfrak{q})$ by Corollary 7.15. Let $\mathrm{Frob}_\mathfrak{q}$ denote the Frobenius endomorphism of the abelian variety $J_0(\mathfrak{p})_{\mathbf{f}_\mathfrak{q}}$. The Hecke operator $T_\mathfrak{q}$ for each prime $\mathfrak{q}$ different from $\mathfrak{p}$ satisfies the Eichler-Shimura relation:

$$\mathrm{Frob}_\mathfrak{q}^2 - T_\mathfrak{q}\cdot\mathrm{Frob}_\mathfrak{q} + q^{\deg(\mathfrak{q})} = 0.$$

Since $\mathrm{Frob}_\mathfrak{q}$ fixes the reduction of $\mathcal{T}(\mathfrak{p})$, the endomorphism $1 - T_\mathfrak{q} + q^{\deg(\mathfrak{q})}$ annihilates this group. As the reduction map commutes with the action of the Hecke algebra, we get that $\mathfrak{E}(\mathfrak{p})$ annihilates the torsion subgroup. □

Let $t(\mathfrak{p})$ denote the greatest common divisor of $N(\mathfrak{p})$ and $q-1$.

Corollary 7.17. *If the prime l does not divide $t(\mathfrak{p})$ then the l-primary torsion subgroup of $\mathcal{T}(\mathfrak{p})$ injects into the group of connected components of the special fiber of the Néron model of $J_0(\mathfrak{p})$ at ∞ via the specialization map.*

Proof. By Corollary 7.15 we may assume that l is different from p. We may assume that l is odd, too. Otherwise $l=2$ and because it does not divide $q-1$, the number q is even, and we already covered this case. The exponent of the kernel of this map divides both $q-1$ and the cardinality of $\mathbb{T}_l(\mathfrak{p})/\mathfrak{E}_l(\mathfrak{p})$ by (ii) of Lemma 7.13 and Lemma 7.16, respectively. The former lemma could be applied as $J_0(\mathfrak{p})$ has split multiplicative reduction at ∞ by Theorem 7.7. Since the latter quantity divides $2N(\mathfrak{p})$ by (vi) of Proposition 7.11, the claim is now clear. □

Proposition 7.18. *For every natural number n the image of $\mathcal{T}(\mathfrak{p})[n]$ with respect to the specialization map into the group of connected components of the special fiber of the Néron model of $J_0(\mathfrak{p})$ at ∞ is a subgroup of $\mathcal{E}_0(\mathfrak{p},\mathbb{Z}/n\mathbb{Z})$.*

Proof. Since $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}),\mathcal{O}_\infty^*)$ is isomorphic to the subgroup of $J_0(\mathfrak{p})(F_\infty)$ mapping into the connected component under the specialization map at ∞ via the map Φ_{AJ}, the $\mathbb{T}(\mathfrak{p})$-module $n^{-1}\overline{\Gamma}_0(\mathfrak{p})/\overline{\Gamma}_0(\mathfrak{p})$ contains the n-torsion of the group of connected components at ∞ as a submodule. The former is isomorphic to $\mathcal{H}_0(\mathfrak{p},\mathbb{Z})/n\mathcal{H}_0(\mathfrak{p},\mathbb{Z})$ by Theorem 7.4, which injects into $\mathcal{H}_0(\mathfrak{p},\mathbb{Z}/n\mathbb{Z})$. Since the specialization map is $\mathbb{T}(\mathfrak{p})$-equivariant, the image of $\mathcal{T}(\mathfrak{p})[n]$ with respect to the composition of these maps must lie in the $\mathbb{T}(\mathfrak{p})$-submodule of $\mathcal{H}_0(\mathfrak{p}, \mathbb{Z}/n\mathbb{Z})$ annihilated by the Eisenstein ideal, according to Lemma 7.16. □

The following theorem is the main Diophantine result of this chapter, which implies Theorems 1.2 and 1.4 under the assumption $t(\mathfrak{p})$=1. The latter is automatic if $q = 2$, so we have a much simpler proof of this result in this case. In the general case we have to prove the Gorenstein property first.

THEOREM 7.19. *If the prime l does not divide $t(\mathfrak{p})$ then the l-primary subgroups of $\mathcal{T}(\mathfrak{p})$ and $\mathcal{C}(\mathfrak{p})$ are equal.*

PROOF. Just as in the proof of Corollary 7.17, we may assume that l is odd and different from p. This result, along with Proposition 7.18 and Corollary 6.9, also implies that the l-primary subgroup of $\mathcal{T}(\mathfrak{p})$ injects into $\mathbb{Z}_l/N(\mathfrak{p})\mathbb{Z}_l$. Since the order of $\mathcal{C}(\mathfrak{p})$ is exactly $N(\mathfrak{p})$ (see [6], Corollary 5.11 on page 235), the proof is now complete. □

8. THE GROUP SCHEME $\mathcal{S}(\mathfrak{p})$

DEFINITION 8.1. For every $\mathbb{F}_q$-algebra B let $B\{\tau\}$ denote the skew-polynomial ring over B defined by the relation $\tau b = b^q\tau$, where b is any element of B. We will also simplify our notation by using the symbol B to denote the spectrum of any ring B. For every non-zero ideal $\mathfrak{n} \lhd A$ and Drinfeld module $\phi : A \to B\{\tau\}$ let $\phi[\mathfrak{n}]$ denote the finite flat group scheme of $\mathbb{G}_a$ over B which is usually called the $\mathfrak{n}$-torsion of the Drinfeld module ϕ, where B is any A-algebra. For every scheme G over any base S and any S-scheme T let $G(T)$ denote the set of sections over T, as usual. The group of sections $\phi[\mathfrak{n}](B)$ is naturally an $A/\mathfrak{n}$-module under the action of A on $\mathbb{G}_a$ defined by ϕ.
We are going to define the concept of a Γ-level structure of a Drinfeld module ϕ of rank two over an A-algebra B, where Γ is either $\Gamma(\mathfrak{n})$ or $\Gamma_1(\mathfrak{n})$. Let $N(\Gamma)$ be the abstract A-module $(A/\mathfrak{n})^2$, if $\Gamma = \Gamma(\mathfrak{n})$, and let $N(\Gamma)$ be $A/\mathfrak{n}$, if $\Gamma = \Gamma_1(\mathfrak{n})$. A homomorphism of abstract A-modules $\iota : N(\Gamma) \to \phi[\mathfrak{n}](B)$ is said to be a Γ-level structure on ϕ over B if the effective Cartier divisor D on $\mathbb{G}_a$ over B of degree $|N(\Gamma)|$ defined by $D = \sum_{a\in N(\Gamma)}[\iota(a)]$ is a subgroup scheme of $\phi[\mathfrak{n}]$. By comparing degrees one can conclude that D is actually equal to $\phi[\mathfrak{n}]$ when $\Gamma = \Gamma(\mathfrak{n})$. Hence our concept of $\Gamma(\mathfrak{n})$-level structure is the same as what is now called a Drinfeld basis of $\phi[\mathfrak{n}]$ (see 3.1.-3.2 of chapter III in [13], page 98-99).
Let (ϕ, ι) and (ψ, κ) be ordered pairs of two Drinfeld modules ϕ and ψ of rank two over B equipped with a Γ-level structure ι and κ, respectively. We say that (ϕ, ι) and (ψ, κ) are isomorphic if there is an isomorphism $j : \mathbb{G}_a \to \mathbb{G}_a$ between ϕ and ψ such that the composition $j \circ \iota$ is equal to κ. Let $\mathcal{M}(\mathfrak{n})$ and $\mathcal{M}_1(\mathfrak{n})$ denote the functor which associates to each A-algebra B the set of isomorphism classes of pairs (ϕ, ι) as above, where ι is a $\Gamma(\mathfrak{n})$-level and $\Gamma_1(\mathfrak{n})$-level structure, respectively. If $\mathfrak{n}$ and $\mathfrak{m}$ are relatively prime non-zero ideals of A, let $\mathcal{M}(\mathfrak{n}, \mathfrak{m})$ denote the fiber product of $\mathcal{M}(\mathfrak{n})$ and $\mathcal{M}_1(\mathfrak{m})$ over $\mathcal{M}(1)$. Clearly $\mathcal{M}(\mathfrak{n}, \mathfrak{m})$ is the functor which associates to each A-algebra B the set of isomorphism classes of triples (ϕ, ι, κ), where ι, κ is a $\Gamma(\mathfrak{n})$-level and $\Gamma_1(\mathfrak{m})$-level structure of the

Drinfeld module ϕ, respectively. The following result is just the Corollary to Proposition 5.4 of [3], page 577.

Theorem 8.2. *Assume that the ideal $\mathfrak{n}$ has at least two different prime factors. Then the moduli problem $\mathcal{M}(\mathfrak{n})$ is representable by a regular fine moduli scheme $M(\mathfrak{n})$.* $\square$

Remark 8.3. The natural left action of $GL_2(A/\mathfrak{n})$ on $(A/\mathfrak{n})^2$ induces a right action of $GL_2(A/\mathfrak{n})$ on $\mathcal{M}(\mathfrak{n})$, hence a right action on $M(\mathfrak{n})$, if the latter exists. Let $\Gamma(\mathfrak{n})$ denote the kernel of the natural surjection $GL_2(A/\mathfrak{n}\mathfrak{m}) \to GL_2(A/\mathfrak{n})$ for any $\mathfrak{m} \lhd A$ non-zero ideal, by slight abuse of notation. The pull-back of the quotients $M(\mathfrak{n}\mathfrak{m}_1)/\Gamma(\mathfrak{n})$ and $M(\mathfrak{n}\mathfrak{m}_2)/\Gamma(\mathfrak{n})$ to $X - supp(\mathfrak{m}_1\mathfrak{m}_2) - \infty$ are naturally isomorphic whenever $M(\mathfrak{n}\mathfrak{m}_1)$ and $M(\mathfrak{n}\mathfrak{m}_2)$ exist and these schemes glue together to form a coarse moduli scheme for $\mathcal{M}(\mathfrak{n})$. We let $M(\mathfrak{n})$ denote this moduli scheme. Of course this notation is compatible with the previous one.

Definition 8.4. Let G be a finite flat group scheme over the base scheme S equipped with the action of a ring R. The latter implies that there is a natural R-module structure on $G(T)$ for any S-scheme T. Let N be a finite abelian group which is also an R-module. Let $\mathrm{Hom}_R(N,G)$ denote the functor which associates to each S-scheme T the set of homomorphisms of abstract R-modules $\iota : N \to G(T)$. This functor is representable by a fine moduli scheme which will be denoted by the same symbol by the usual abuse of notation.
Let $\phi : A \to B\{\tau\}$ be a Drinfeld module over the A-algebra B, let G be the kernel of a non-zero isogeny on ϕ, and let N be a finite A-module. Note that the group scheme G is naturally an A-module under the action of A on $\mathbb{G}_a$ defined by ϕ. Let $Str_A(N,G)$ denote the sub-functor of $\mathrm{Hom}_A(N,G)$ which associates to each B-algebra C the set of those homomorphisms of abstract A-modules $\iota : N \to G(C)$ such that the effective Cartier divisor D on $\mathbb{G}_a$ over C of degree $|N|$ defined by $D = \sum_{a\in N}[\iota(a)]$ is a subgroup scheme of G.

Lemma 8.5. *The functor $Str_A(N,G)$ is represented by a closed subscheme of $\mathrm{Hom}_A(N,G)$. If G is étale then $Str_A(N,G)$ is either empty or finite, étale over every connected component of B.*

Proof. In 1.5.1 of chapter in [13], page 20-21, the concept of N-level structure was defined. By Proposition 1.6.3, Corollary 1.6.3 on page 23 of the same book the functor which associates to each B-algebra C the set of N-level structures on the $\mathfrak{n}$-torsion of the pull-back of ϕ to C is represented by a closed subscheme of $\mathrm{Hom}_{\mathbb{Z}}(N,G)$. Our functor is represented by the scheme-theoretical intersection of this scheme and $\mathrm{Hom}_A(N,G)$. The second claim follows from Proposition 1.10.12 of [13], page 46-47. $\square$

Definition 8.6. We say that an A-algebra B has characteristic $\mathfrak{p}$ if the annihilator of the A-module B contains $\mathfrak{p}$. This assumption implies that B is an $\mathbf{f}_{\mathfrak{p}}$-algebra. We let $x^{\mathfrak{p}}$ denote $x^{q^{\deg(\mathfrak{p})}}$ for every $\mathbf{f}_{\mathfrak{p}}$-algebra B and element $x \in B$. We say that a Drinfeld module $\phi : A \to B\{\tau\}$ has characteristic $\mathfrak{p}$ if the

A-algebra B has characteristic $\mathfrak{p}$. For every Drinfeld module $\phi : A \to B\{\tau\}$ of characteristic $\mathfrak{p}$ we let $\phi^{(\mathfrak{p})} : A \to B\{\tau\}$ denote the Drinfeld module which as a homomorphism from A to $B\{\tau\}$ is the composition of ϕ and the unique homomorphism $F_{\mathfrak{p}} : B\{\tau\} \to B\{\tau\}$ such that $F_{\mathfrak{p}}(\tau) = \tau$ and $F_{\mathfrak{p}}(x) = x^{\mathfrak{p}}$ for every $x \in B$. Note that $\phi^{(\mathfrak{p})}$ is a Drinfeld module because the homomorphism $x \mapsto x^{\mathfrak{p}}$ fixes the field $\mathbf{f}_{\mathfrak{p}}$, so the composition of $\phi^{(\mathfrak{p})}$ and the derivation $\partial : B\{\tau\} \to B$ is the reduction map $A \to \mathbf{f}_{\mathfrak{p}}$ as required by definition. As obvious from the definition the endomorphism $x \mapsto x^{\mathfrak{p}}$ of the group scheme $\mathbb{G}_a$ defines an isogeny F from ϕ to $\phi^{(\mathfrak{p})}$ which will be called Frobenius. We let $\mathbf{k}_{\mathfrak{p}}$ denote the algebraic closure of the field $\mathbf{f}_{\mathfrak{p}}$.

Proposition 8.7. *For every Drinfeld module $\phi : A \to B\{\tau\}$ of characteristic $\mathfrak{p}$ the kernel of the isogeny F is a sub-group scheme of $\phi[\mathfrak{p}]$.*

Proof. Let $f \in A = \mathbb{F}_q[T]$ be a polynomial which generates $\mathfrak{p}$. We are going to prove the following stronger formulation of the statement which claims that $\phi(f) = \sum_n a_n\tau^n \in B\{\tau\}$ has no terms of degree less then $\deg(\mathfrak{p})$ in τ. This claim may be checked locally in the étale topology on B. Let $\mathfrak{n}$ be an ideal of A which is relatively prime to $\mathfrak{p}$ and has at least two different prime factors. By Lemma 8.5 the B-scheme $Str_A((A/\mathfrak{n})^2, \phi[\mathfrak{n}])$ is étale, since it is not empty over any component. The latter can be seen by noticing that the base change of ϕ to every geometric point of B has a $\Gamma(\mathfrak{n})$-level structure. Hence we may assume that ϕ is equipped with a $\Gamma(\mathfrak{n})$-level structure. By Theorem 8.2 the Drinfeld module ϕ is the pull-back of the universal Drinfeld module Φ on the fiber of the fine moduli scheme over $\mathbf{f}_{\mathfrak{p}}$. It will be sufficient to prove the claim for the latter. The fiber of the scheme $M(\mathfrak{n})$ over $\mathbf{f}_{\mathfrak{p}}$ is smooth, so we only have to show that the terms of $\Phi(f)$ of degree less then $\deg(\mathfrak{p})$ are vanishing at the geometric points of this fiber. The latter follows from the fact that the proposition holds for Drinfeld modules over $\mathbf{k}_{\mathfrak{p}}$. This last claim is the content of the remark following Proposition 5.1 of [4], page 178. □

Definition 8.8. By the above $\tau^{\deg(\mathfrak{p})}$ divides $\phi(f)$ on the right in the ring $B\{\tau\}$, so there is a unique isogeny V from $\phi^{(\mathfrak{p})}$ to ϕ such that the composition $V \circ F$ is $\phi(f)$. The isogeny V will be called Verschiebung. Note that V depends on the choice of f. But the latter is unique up to a non-zero element of $\mathbb{F}_q$, so $Ker(V)$ is well-defined. Let $\mathfrak{n}$ be any ideal of A relatively prime to $\mathfrak{p}$. We let $\mathcal{I}(\mathfrak{p})$ and $\mathcal{I}(\mathfrak{n}, \mathfrak{p})$ denote the functor which associates to each $\mathbf{f}_{\mathfrak{p}}$-algebra B the set of isomorphism classes of pairs (ϕ, ι) (of triples (ϕ, ι, κ), respectively), where $\phi : A \to B\{\tau\}$ is a Drinfeld module of rank two and ι is an element of $Str_A(A/\mathfrak{p}, Ker(V))$ (and κ is a $\Gamma(\mathfrak{n})$-level structure of ϕ, respectively). We say that two pairs (ϕ, ι) and (ψ, κ) as above are isomorphic if there is an isomorphism $j : \mathbb{G}_a \to \mathbb{G}_a$ between ϕ and ψ such that the composition $j^{\mathfrak{p}} \circ \iota$ is equal to κ. (Note that the definition makes sense because $j^{\mathfrak{p}}$ is an isomorphism between $\phi^{(\mathfrak{p})}$ and $\psi^{(\mathfrak{p})}$). We define the concept of isomorphism of the triples appearing in the definition of $\mathcal{I}(\mathfrak{n}, \mathfrak{p})$ similarly.

Proposition 8.9. *Let $\psi : A \to \mathbf{k}_{\mathfrak{p}}\{\tau\}$ be a Drinfeld module of rank two. The following conditions are equivalent:*

(i) *the group scheme $\psi[\mathfrak{p}]$ is connected,*
(ii) *the group scheme $Ker(V)$ is connected,*
(iii) *the group scheme $Ker(V)$ is not étale.*

Proof. The implications (i) $\Rightarrow$ (ii) and (ii) $\Rightarrow$ (iii) are obvious. If the group scheme $Ker(V)$ is not étale then all terms of $\psi(f)$ have degree greater than $\deg(\mathfrak{p})$. The latter is equivalent to (i) by Satz 5.3 of [4], page 179. □

Definition 8.10. In complete analogy with the classical theory of elliptic curves over algebraic fields of positive characteristic, such Drinfeld modules are called supersingular. Let $\mathcal{R}_{\mathfrak{p}}$ be the maximal unramified extension of $\mathcal{O}_{\mathfrak{p}}$. By definition the residue field of the latter is $\mathbf{k}_{\mathfrak{p}}$. Let $\mathcal{C}_{\mathfrak{p}}$ denote the category whose objects are artin local $\mathcal{R}_{\mathfrak{p}}$-algebras with residue field $\mathbf{k}_{\mathfrak{p}}$ and the morphisms are local $\mathcal{R}_{\mathfrak{p}}$-homomorphisms. Let $\phi : A \to \mathbf{k}_{\mathfrak{p}}\{\tau\}$ be a Drinfeld module of rank two. We say that the Drinfeld module $\Phi : A \to \mathcal{R}_{\mathfrak{p}}[[x]]\{\tau\}$ of rank two is its universal formal deformation if the latter is the universal object over $\mathcal{R}_{\mathfrak{p}}[[x]]$ pro-representing the functor which associates to each object B of $\mathcal{C}_{\mathfrak{p}}$ the set of strict isomorphism classes of Drinfeld modules over B lifting ϕ. (Recall that two Drinfeld modules over B are strictly isomorphic if there is an isomorphism between them whose pull-back to the residue field is the identity). Under our assumption $A = \mathbb{F}_q[T]$ it is very easy to see that the universal deformation exits: up to an isomorphism $\phi(T)$ is of the form $T+\tau^2$ or $T+\tau+\Delta\tau^2$ where Δ is a non-zero element of $\mathbf{k}_{\mathfrak{p}}$. Then we may choose Φ to be the unique Drinfeld module over $\mathcal{R}_{\mathfrak{p}}[[x]]$ with $\Phi(T) = T + x\tau + \tau^2$ or $\Phi(T) = T + \tau + (\Delta + x)\tau^2$.

Proposition 8.11. *Assume that the ideal $\mathfrak{n}$ has at least two different prime factors. Then the moduli problem $\mathcal{M}(\mathfrak{n}, \mathfrak{p})$ is representable by a regular fine moduli scheme $M(\mathfrak{n}, \mathfrak{p})$.*

Proof. Let (ϕ, ι) be the universal object over the fine moduli scheme $M(\mathfrak{n})$. It is clear that the moduli problem $\mathcal{M}(\mathfrak{n}, \mathfrak{p})$ is represented by $Str_A(A/\mathfrak{p}, \phi[\mathfrak{p}])$. Now we only have to show that this scheme $M(\mathfrak{n}, \mathfrak{p})$ is regular. The group scheme $\phi[\mathfrak{p}]$ is étale over the base change of $M(\mathfrak{n})$ to $X - \mathfrak{p} - \infty$. Hence the base change of $M(\mathfrak{n}, \mathfrak{p})$ to $X - \mathfrak{p} - \infty$ is étale over $M(\mathfrak{n})$, in particular it is regular. (One may see that $Str_A(A/\mathfrak{p}, \phi[\mathfrak{p}])$ is non-empty by looking at its fibers over geometric points). Therefore we only have to show that $M(\mathfrak{n}, \mathfrak{p})$ is regular at the closed points of its special fiber over $\mathfrak{p}$. By a suitable analogue of the Deligne homogeneity principle (see Theorem 5.2.1 of [13], pages 130-134), whose proof we do not include because it is completely the same as the result quoted above, we only have to check the latter at the supersingular points. This is exactly what the next proposition claims. □

Let $\psi_0 : A \to \mathbf{k}_{\mathfrak{p}}\{\tau\}$ be a supersingular Drinfeld module of rank two, and let $\Psi : A \to \mathcal{R}_{\mathfrak{p}}[[x]]\{\tau\}$ be its universal formal deformation. Fix a $\iota_0 : (A/\mathfrak{n})^2 \to \psi_0[\mathfrak{n}]$ level structure of $\Gamma(\mathfrak{n})$-type. Let $\mathcal{M}(\mathfrak{n}, \mathfrak{p}, \Psi)$ be the functor which associates to each object B of $\mathcal{C}_{\mathfrak{p}}$ the set of isomorphism classes of triples $(\Psi|_B, \iota, \kappa)$, where $\Psi|_B$ is the pull-back of the Drinfeld module Ψ to B, and ι, κ is a $\Gamma(\mathfrak{n})$-level and $\Gamma_1(\mathfrak{p})$-level structure of the Drinfeld module $\Psi|_B$, respectively, such that the base change of ι to $\mathbf{k}_{\mathfrak{p}}$ with respect to the residue map is the level structure ι_0 above.

Proposition 8.12. *The following holds:*

(i) *the set $\mathcal{M}(\mathfrak{n}, \mathfrak{p}, \Psi)(\mathbf{k}_{\mathfrak{p}})$ consists of one element,*
(ii) *the functor $\mathcal{M}(\mathfrak{n}, \mathfrak{p}, \Psi)$ is pro-represented by the spectrum of a regular local ring.*

Proof. By assumption the group scheme $\psi_0[\mathfrak{p}]$ is connected, so the Drinfeld module ψ_0 has only one $\Gamma_1(\mathfrak{p})$-level structure: the identically zero map. Hence claim (i) is clear. We may apply the argument of Proposition 5.2.2 of [13], page 135, to reduce claim (ii) to the seemingly weaker claim that the functor $\mathcal{M}(\mathfrak{n}, \mathfrak{p}, \Psi)$ is pro-represented by the spectrum of a local ring whose maximal ideal is generated by two elements. The pro-representability of $\mathcal{M}(\mathfrak{n}, \mathfrak{p}, \Psi)$ by the spectrum of a ring $\mathcal{A}$ is clear since $\mathcal{M}(\mathfrak{n}, \mathfrak{p})$ itself is representable. By claim (i) this ring $\mathcal{A}$ is local. It is also a finite $\mathcal{R}_{\mathfrak{p}}[[x]]$-algebra by Lemma 8.5, so it is complete. Let $(\Psi|_{\mathcal{A}}, \alpha, \beta)$ be the universal object over $\mathcal{A}$ with respect to the moduli problem $\mathcal{M}(\mathfrak{n}, \mathfrak{p}, \Psi)$. The section $\beta(1) \in \mathbb{G}_a(\mathcal{A})$ corresponds to an element $y \in \mathcal{A}$ which lies in the maximal ideal $\mathfrak{M}$ of $\mathcal{A}$, since the reduction of $\beta(1)$ modulo $\mathfrak{M}$ lies in the connected group scheme $\psi_0[\mathfrak{p}]$. We claim that the parameter x of $\mathcal{R}_{\mathfrak{p}}[[x]]$ and y generate the maximal ideal $\mathfrak{M}$. In light of the universal property and completeness of $\mathcal{A}$ we only need to show that for every B artin local $\mathcal{R}_{\mathfrak{p}}$-algebra and $\phi : \mathcal{A} \to B$ homomorphism of local $\mathcal{R}_{\mathfrak{p}}$-algebras with $\phi(x) = \phi(y) = 0$ the map ϕ factors through the residue map $\mathcal{A} \to \mathcal{A}/\mathfrak{M} = \mathbf{k}_{\mathfrak{p}}$, which is equivalent to the rigidity assertion below. □

Lemma 8.13. *If B is an artin local $\mathcal{R}_{\mathfrak{p}}$-algebra and if $\phi : \mathcal{A} \to B$ is a homomorphism of local $\mathcal{R}_{\mathfrak{p}}$-algebras with $\phi(x) = \phi(y) = 0$, then B is a $\mathbf{k}_{\mathfrak{p}}$-algebra and the induced triple $(\Psi|_B, \alpha|_B, \beta|_B)$ comes from the triple $(\psi_0, \iota_0, 0)$ by extension of scalars $\mathbf{k}_{\mathfrak{p}} \to B$.*

Proof. Let $f \in A = \mathbb{F}_q[T]$ be a polynomial which generates $\mathfrak{p}$. By assumption $\beta|_B(1) \in \mathbb{G}_a(B)$ is the zero section, hence the zero scheme of the polynomial $X^{q^{\deg(\mathfrak{p})}} \in B[X]$ is a subgroup scheme of $\Psi|_B[\mathfrak{p}]$. Hence it must divide the monic polynomial $\Psi|_B(f) = X^{2q^{\deg(\mathfrak{p})}} + \cdots + fX \in B[X]$. In particular f must be zero in B, so the latter is a $\mathbf{k}_{\mathfrak{p}}$-algebra. Since $\phi(x) = 0$ in B as well, the Drinfeld module $\Psi|_B$ must be constant in the sense that it is the pull-back of ψ_0 via the extension of scalars $\mathbf{k}_{\mathfrak{p}} \to B$. Since the group scheme $\Psi|_B[\mathfrak{n}]$ is étale, the Drinfeld module $\Psi|_B$ has exactly one $\Gamma(\mathfrak{n})$-level structure up to isomorphism whose base change to $\mathbf{k}_{\mathfrak{p}}$ with respect to residue map of the local ring B is isomorphic to the level structure ι_0 above, namely the pull-back of ι_0 via the extension of scalars. □

Proposition 8.14. *Assume that the ideal $\mathfrak{n}$ has at least two different prime factors. Then the moduli problem $\mathcal{I}(\mathfrak{n},\mathfrak{p})$ is representable by a smooth affine curve $I(\mathfrak{n},\mathfrak{p})$ over $\mathbf{f}_{\mathfrak{p}}$ and the natural map $I(\mathfrak{n},\mathfrak{p}) \to M(\mathfrak{n}) \times_A \mathbf{f}_{\mathfrak{p}}$ is finite and flat.*

Proof. Let $M(\mathfrak{n})_{\mathfrak{p}}$ denote the fiber of $M(\mathfrak{n})$ over $\mathbf{f}_{\mathfrak{p}}$ and let (ϕ,ι) be the universal object over the scheme $M(\mathfrak{n})_{\mathfrak{p}}$ which is a fine moduli for Drinfeld modules of characteristic $\mathfrak{p}$ equipped with a $\Gamma(\mathfrak{n})$-level structure. It is clear that the moduli problem $\mathcal{I}(\mathfrak{n},\mathfrak{p})$ is represented by $Str_A(A/\mathfrak{p}, Ker(V))$. In particular it is finite over $M(\mathfrak{n})_{\mathfrak{p}}$. By Satz 5.9 of [4], page 181, there are only finitely many $\mathbf{k}_{\mathfrak{p}}$-valued points of $M(\mathfrak{n})_{\mathfrak{p}}$ such that the corresponding Drinfeld module is supersingular. We may reformulate this claim by saying that there is a zero-dimensional closed sub-scheme $M(\mathfrak{n})^{ss}_{\mathfrak{p}}$ of the smooth affine curve $M(\mathfrak{n})_{\mathfrak{p}}$ whose base change to $\mathbf{k}_{\mathfrak{p}}$ represents supersingular Drinfeld modules equipped with a $\Gamma(\mathfrak{n})$-level structure. Here a Drinfeld module over a $\mathbf{k}_{\mathfrak{p}}$-algebra is supersingular if its $\mathfrak{p}$-torsion group scheme is connected. By Propositions 8.7 and 8.9 we may define $M(\mathfrak{n})^{ss}_{\mathfrak{p}}$ as the zero scheme of the Hasse invariant of Gekeler, i.e. the coefficient of the term of $\phi(f)$ of degree $\deg(\mathfrak{p})$, where f is a polynomial which generates the ideal $\mathfrak{p}$. The finite, flat group scheme $Ker(V)$ over the open complement $M(\mathfrak{n})^{ord}_{\mathfrak{p}}$ of $M(\mathfrak{n})^{ss}_{\mathfrak{p}}$ is étale, because its pull-back to every $\mathbf{k}_{\mathfrak{p}}$-valued point is étale by Proposition 8.7. Hence the map $I(\mathfrak{n},\mathfrak{p}) \to M(\mathfrak{n})_{\mathfrak{p}}$ is étale over the open sub-scheme $M(\mathfrak{n})^{ord}_{\mathfrak{p}}$ by Lemma 8.5. Therefore the pre-image of $M(\mathfrak{n})^{ord}_{\mathfrak{p}}$ in $I(\mathfrak{n},\mathfrak{p})$ is a smooth curve.

Hence we only have to show that $I(\mathfrak{n},\mathfrak{p})$ is smooth of dimension one at its supersingular locus, i.e. at the pre-image of $M(\mathfrak{n})^{ss}_{\mathfrak{p}}$, because every finite, almost everywhere unramified map between smooth curves is automatically flat. It is sufficient do so after base change to $\mathbf{k}_{\mathfrak{p}}$. Our argument is very similar to the proof of Proposition 8.12. Let (ψ_0,ι_0) be a pair which corresponds to a supersingular point of $M(\mathfrak{n})$, which means that $\psi_0 : A \to \mathbf{k}_{\mathfrak{p}}\{\tau\}$ is a supersingular Drinfeld module of rank two and $\iota_0 : (A/\mathfrak{n})^2 \to \psi_0[\mathfrak{n}]$ is a $\Gamma(\mathfrak{n})$-level structure. As the group scheme $Ker(V) \subseteq \psi_0^{(\mathfrak{p})}[\mathfrak{p}]$ is connected, this point has a unique lift (ψ,ι_0,κ_0) to $I(\mathfrak{n},\mathfrak{p})$. Let $\Psi : A \to \mathbf{k}_{\mathfrak{p}}[[x]]\{\tau\}$ be the universal formal deformation of ψ_0 for local artin $\mathbf{k}_{\mathfrak{p}}$-algebras. Since the group scheme $\Psi[\mathfrak{n}]$ is étale, there is a unique level structure $\iota : (A/\mathfrak{n})^2 \to \Psi[\mathfrak{n}]$ lifting ι_0 up to strict isomorphism. The pair (Ψ,ι) is the universal object over $\mathbf{k}_{\mathfrak{p}}[[x]]$ which pro-represents the deformations of the pair (ψ_0,ι_0) over local artin $\mathbf{k}_{\mathfrak{p}}$-algebras. Let $\mathcal{A}$ be the local complete $\mathbf{k}_{\mathfrak{p}}[[x]]$-algebra whose spectrum is $Str_A(A/\mathfrak{p},\Psi[\mathfrak{p}])$: this ring is the completion of the local ring of the scheme $I(\mathfrak{n},\mathfrak{p}) \times_{\mathbf{f}_{\mathfrak{p}}} \mathbf{k}_{\mathfrak{p}}$ at the closed point (ψ,ι_0,κ_0). It will be sufficient to show that $\mathcal{A}$ is a formal power series ring over $\mathbf{k}_{\mathfrak{p}}$. We only need to find a parameter in $\mathcal{A}$ because we proved already that $\mathcal{A}$ is finite over $\mathbf{k}_{\mathfrak{p}}[[x]]$ and it has dimension one. Note that $\mathcal{A}$ pro-represents the deformations of the triple $(\psi_0,\iota_0,\kappa_0)$ over local artin $\mathbf{k}_{\mathfrak{p}}$-algebras. Let (Ψ,ι,κ) be the universal object over this ring. The section $\kappa(1) \in \mathbb{G}_a(\mathcal{A})$ corresponds to an element $y \in \mathcal{A}$ which lies in the maximal ideal $\mathfrak{M}$ of $\mathcal{A}$, since the reduction of $\kappa(1)$ modulo $\mathfrak{M}$ lies in the connected group scheme $Ker(V) \subseteq \psi_0^{(\mathfrak{p})}[\mathfrak{p}]$. We claim that y generates the maximal ideal $\mathfrak{M}$. Because of the universal property of $\mathcal{A}$ it will be sufficient to show the following rigidity assertion: if B is an artin local $\mathbf{k}_\mathfrak{p}$-algebra and if $\phi : \mathcal{A} \to B$ is a homomorphism of local $\mathbf{k}_\mathfrak{p}$-algebras with $\phi(y) = 0$, then the induced triple $(\Psi|_B, \iota|_B, \kappa|_B)$ comes from the triple $(\psi_0, \iota_0, \kappa_0)$ by extension of scalars $\mathbf{k}_\mathfrak{p} \to B$. Under these assumptions $Ker(V) \subseteq \Psi^{(\mathfrak{p})}|_B[\mathfrak{p}]$ is connected, hence so does $\Psi|_B[\mathfrak{p}]$, because the latter is the extension of $Ker(F)$ by $Ker(V)$. By Lemma 5.5 of [4], page 191, the scheme $M(\mathfrak{n})^{ss}_\mathfrak{p}$ is reduced, so the pair $(\Psi|_B, \iota|_B)$ is constant. The level structure $\kappa|_B$ is constant by assumption, so does the triple $(\Psi|_B, \iota|_B, \kappa|_B)$. $\square$

Definition 8.15. The natural left action of $GL_2(A/\mathfrak{n})$ on $(A/\mathfrak{n})^2$ induces a right action of $GL_2(A/\mathfrak{n})$ on $\mathcal{M}(\mathfrak{n}, \mathfrak{p})$, hence a right action on $M(\mathfrak{n}, \mathfrak{p})$, if the latter exists. We may glue together open pieces of the quotients $M(\mathfrak{n}, \mathfrak{p})/GL_2(A/\mathfrak{n})$ for various $\mathfrak{n}$ to form a coarse moduli scheme for $\mathcal{M}_1(\mathfrak{p})$, as in Remark 8.3. We let $M_1(\mathfrak{p})$ denote this moduli scheme. Similarly we may construct a coarse moduli scheme $I(\mathfrak{p})$ representing the functor $\mathcal{I}(\mathfrak{p})$ by gluing together open pieces of the quotients $I(\mathfrak{n}, \mathfrak{p})/GL_2(A/\mathfrak{n})$. Also note that there is a morphism $I(\mathfrak{p}) \to M_1(\mathfrak{p}) \times_A \mathbf{f}_\mathfrak{p}$ induced by the natural map which assigns to every pair (ϕ, ι) of the type appearing in Definition 8.8 the pair $(\phi^{(\mathfrak{p})}, \iota)$.

Proposition 8.16. *The coarse moduli $M_1(\mathfrak{p})$ has the following properties:*

- (i) *it is a model of $Y_1(\mathfrak{p})$ over the spectrum of A,*
- (ii) *it is normal and affine over $Spec(A)$,*
- (iii) *the reduced scheme associated to its reduction over $\mathfrak{p}$ has two irreducible components which are smooth curves over $\mathbf{f}_\mathfrak{p}$ and intersect transversally in $N(\mathfrak{p})$ supersingular points.*

Proof. We start our proof by showing the following remark: if R is a normal integral domain and G is a finite group acting on R, then the subring R^G of invariants is also integrally closed. Let Q be the quotient field of R. This field is equipped with an action of G which extends the action of the latter on R. The field Q^G of invariants clearly contains the quotient field of R^G. Any element of Q^G integral over R^G must lie in $R^G = R \cap Q^G$ because R is integrally closed. Hence the remark is true.

The first claim is obvious. Zariski-locally on $Spec(A)$ the scheme $M_1(\mathfrak{p})$ is the quotient of an affine and regular scheme by a finite group, so the second claim is also clear by the remark above. Recall that the reduction of $M_0(\mathfrak{p})$ over $\mathfrak{p}$ has two irreducible components: $M_{00}(\mathfrak{p})$ and $M_{01}(\mathfrak{p})$, whose $\mathbf{k}_\mathfrak{p}$-valued points correspond to pairs $(\phi, Ker(F))$ and $(\phi^{(\mathfrak{p})}, Ker(V))$, respectively, where $\phi : A \to \mathbf{k}_\mathfrak{p}\{\tau\}$ is any Drinfeld module of rank two over $\mathbf{k}_\mathfrak{p}$. Let $M_{10}(\mathfrak{p})$ and $M_{11}(\mathfrak{p})$ denote the pre-image of $M_{00}(\mathfrak{p})$ and $M_{01}(\mathfrak{p})$ via the natural map $M_1(\mathfrak{p}) \to M_0(\mathfrak{p})$, respectively. The composition of the canonical map $M_{10}(\mathfrak{p})_{red} \to M_{10}(\mathfrak{p})$ and the restriction $M_{10}(\mathfrak{p}) \to M_{00}(\mathfrak{p})$ induces a bijection between the set of $\mathbf{k}_\mathfrak{p}$-valued points of $M_{10}(\mathfrak{p})_{red}$ and $M_{00}(\mathfrak{p})$ because the group scheme $Ker(F)$ is always connected. By Hilbert's Nullstellensatz the compo-

sition map above must be a finite map of degree 1 between irreducible curves, in particular $M_{10}(\mathfrak{p})_{red}$ is connected. But $M_{00}(\mathfrak{p})$ is normal, so this map is an isomorphism. Hence $M_{10}(\mathfrak{p})_{red}$ is smooth, too.
For every A-algebra B of characteristic $\mathfrak{p}$ the set $\mathcal{I}(\mathfrak{p})(B)$ injects into $\mathcal{M}_1(\mathfrak{p})(B)$ under the natural map which induces the map $I(\mathfrak{p}) \to M_1(\mathfrak{p}) \times_A \mathbf{f}_{\mathfrak{p}}$ of Definition 8.15, so the latter is a closed immersion. Clearly $M_{11}(\mathfrak{p})$ is the image of $I(\mathfrak{p})$, so it is smooth by Proposition 8.14. The same proposition implies that the natural map $I(\mathfrak{p}) \to M(1) \times_A \mathbf{f}_{\mathfrak{p}}$ is a branched covering which totally ramifies over the supersingular points. The latter follows from the fact every supersingular Drinfeld module of rank two over $\mathbf{k}_{\mathfrak{p}}$ has a unique $\mathcal{I}(\mathfrak{p})$-structure, because its $\mathfrak{p}$-torsion group scheme is connected. Hence $M_{11}(\mathfrak{p})$ is connected, too. For the same reason we know that every supersingular point in the reduction of $M_0(\mathfrak{p})$ over $\mathfrak{p}$ has a unique lift to $M_1(\mathfrak{p})$. Claim (iii) is now fully proved. □

Lemma 8.17. *The finite group scheme $\mathcal{S}(\mathfrak{p})$ is étale and μ-type of rank $N(\mathfrak{p})$, and as a subgroup of $J_0(\mathfrak{p})(\overline{F})$ it is cyclic.*

Proof. We will gather some facts about the cover $X_1(\mathfrak{p}) \to X_0(\mathfrak{p})$, where $X_1(\mathfrak{p})$ is the unique geometrically irreducible non-singular projective curve containing $Y_1(\mathfrak{p})$, which could be also excavated from [5], section 4 of chapter V and section 5 of chapter VII, with some effort. We call a geometric point on a Drinfeld modular curve elliptic, if the automorphism group the underlying Drinfeld module of rank two is strictly larger than $\mathbb{F}_q^*$. First note that both the cover $Y_0(\mathfrak{p}) \to Y_0(1)$ and the cover $Y_1(\mathfrak{p}) \to Y_0(1)$ could ramify only over the unique elliptic point of $Y_0(1)$. Hence the cover $X_1(\mathfrak{p}) \to X_0(\mathfrak{p})$ could ramify only at elliptic points and at the cusps. By counting the latter we get that the cover is actually unramified at them. The number of elliptic points on $Y_1(\mathfrak{p})$ is $(q^{2d}-1)/(q^2-1)$. The number of elliptic points on $Y_0(\mathfrak{p})$ is $(q^d+1)/(q+1)$, if d is odd, and it is q^d+1, if d is even. Hence the cover $X_1(\mathfrak{p}) \to X_0(\mathfrak{p})$ ramifies if and only if d is even, when the ramification index is $q+1$ at each elliptic point. We get that the cover $X_2(\mathfrak{p}) \to X_0(\mathfrak{p})$ is unramified. Since it is also Galois over F with a cyclic Galois group of order $N(\mathfrak{p})$, the lemma follows immediately by the same standard argument as in the proof of Proposition 11.6 of [14], page 100. □

Proposition 8.18. *The image of $\mathcal{S}(\mathfrak{p})$ with respect to the specialization map into the the special fiber of the Néron model of $J_0(\mathfrak{p})$*

(i) at ∞ lies in the connected component of the identity,
(ii) at $\mathfrak{p}$ does not intersect the connected component of the identity.

Proof. First note that the two claims make sense because $\mathcal{S}(\mathfrak{p})$ is étale, so it has a well-defined extension into the Néron model of $J_0(\mathfrak{p})$. In this paragraph we will use the notation and results of [11] without extra notice. (Recall that $\overline{\Gamma}_0(\mathfrak{p}) = \overline{\Gamma_0(\mathfrak{p})}$ under the notation introduced by Definition 7.1). Let K_∞ be the maximal unramified extension of F_∞ and let $\mathcal{R}_\infty$ be its discrete valuation ring. Let $J_{00}(\mathfrak{p})(K_\infty)$ and $J_{20}(\mathfrak{p})(K_\infty)$ be the pre-image of the connected component under the reduction map in the Lie groups $J_0(\mathfrak{p})(K_\infty)$ and $J_2(\mathfrak{p})(K_\infty)$, respectively. In order to prove claim (i) it will be sufficient to construct a sub-group of order $N(\mathfrak{p})$ in the kernel of the map $j : J_{00}(\mathfrak{p})(K_\infty) \to J_{20}(\mathfrak{p})(K_\infty)$ induced by Picard functoriality by the previous lemma. By the definition of the Abel-Jacobi map as an automorphy factor there is a commutative diagram of exact sequences:

$$\begin{array}{ccccccc} 0 & \longrightarrow & \mathrm{Hom}(\overline{\Gamma_0(\mathfrak{p})}, \mathcal{R}^*_\infty) & \xrightarrow{\Phi_{AJ}} & J_{00}(\mathfrak{p})(K_\infty) & \longrightarrow & 0 \\ & & {\scriptstyle i}\big\downarrow & & {\scriptstyle j}\big\downarrow & & \\ 0 & \longrightarrow & \mathrm{Hom}(\overline{\Gamma_2(\mathfrak{p})}, \mathcal{R}^*_\infty) & \xrightarrow{\Phi_{AJ}} & J_{20}(\mathfrak{p})(K_\infty) & \longrightarrow & 0 \end{array}$$

where the first vertical map is induced by the abelianization of the canonical injection $\Gamma_2(\mathfrak{p}) \to \Gamma_0(\mathfrak{p})$. Of course $\Gamma_2(\mathfrak{p})$ is the normal arithmetic subgroup of $\Gamma_0(\mathfrak{p})$ corresponding to the cover $Y_2(\mathfrak{p}) \to Y_0(\mathfrak{p})$. By the above we only need to construct a sub-group of the kernel of the map i whose order is $N(\mathfrak{p})$. Since $\mathcal{R}^*_\infty$ contains a cyclic group of order n for any natural number n relatively prime to p, it will be sufficient to construct a surjective homomorphism $h : \Gamma_0(\mathfrak{p}) \to \mathbb{Z}/N(\mathfrak{p})\mathbb{Z}$ whose kernel contains $\Gamma_2(\mathfrak{p})$. We define h as the composition of the reduction map $r : \Gamma_0(\mathfrak{p}) \to B(A/\mathfrak{p}) \subset GL_2(A/\mathfrak{p})$, the upper left corner element $a : B(A/\mathfrak{p}) \to (A/\mathfrak{p})^*$ and the unique surjection $p : (A/\mathfrak{p})^* \to \mathbb{Z}/N(\mathfrak{p})\mathbb{Z}$.

Let's start the proof of the second claim. For every projective curve C (reduced, one-dimensional, but not necessarily irreducible projective scheme over a field) let $\mathrm{Pic}^0(C)$ denote the Picard group of divisors of total degree zero. First note that there is a projective scheme $\overline{M}_1(\mathfrak{p})$ over A which contains $M_1(\mathfrak{p})$ as a Zariski-dense open sub-scheme such that the natural map $p : M_1(\mathfrak{p}) \to M_0(\mathfrak{p})$ has an extension $\overline{p} : \overline{M}_1(\mathfrak{p}) \to \overline{M}_0(\mathfrak{p})$. We may define $\overline{M}_1(\mathfrak{p})$ as the closure of the graph of p in the product of $\overline{M}_0(\mathfrak{p})$ and any projective completion of $M_1(\mathfrak{p})$ over A. Let $r : \widetilde{M}_0(\mathfrak{p}) \to \overline{M}_0(\mathfrak{p})$ be the minimal resolution of singularities of the surface $\overline{M}_0(\mathfrak{p})$ over A. Because $\overline{M}_0(\mathfrak{p})$ is either regular or has a singularity of type A_q over $\mathbf{f}_\mathfrak{p}$ at a supersingular point, the induced map $r^* : \mathrm{Pic}^0(\overline{M}_0(\mathfrak{p}) \times_A \mathbf{k}_\mathfrak{p}) \to \mathrm{Pic}^0(\widetilde{M}_0(\mathfrak{p}) \times_A \mathbf{k}_\mathfrak{p})$ is an isomorphism. Let $\widetilde{M}_1(\mathfrak{p})$ be the minimal resolution of singularities of the fiber product $\overline{M}_1(\mathfrak{p}) \times_A \widetilde{M}_0(\mathfrak{p})$ over A. By construction there is a commutative diagram:

$$\begin{array}{ccc} \widetilde{M}_1(\mathfrak{p}) & \xrightarrow{t} & \widetilde{M}_0(\mathfrak{p}) \\ {\scriptstyle s}\big\downarrow & & {\scriptstyle r}\big\downarrow \\ \overline{M}_1(\mathfrak{p}) & \xrightarrow{\overline{p}} & \overline{M}_0(\mathfrak{p}) \end{array}$$

where the vertical maps are birational. Let $\overline{S}_1$ and $\widetilde{S}_1$ be the closure of the special fiber of $M_1(\mathfrak{p})$ over $\mathbf{k}_\mathfrak{p}$ in $\overline{M}_1(\mathfrak{p}) \times_A \mathbf{k}_\mathfrak{p}$ and its pre-image with respect to

s, respectively. By Picard functoriality we have another commutative diagram:

$$\begin{array}{ccccc} \mathrm{Pic}^0((\widetilde{S}_1)_{red}) & \xleftarrow{j^*} & \mathrm{Pic}^0((\widetilde{M}_1(\mathfrak{p}) \times_A \mathbf{k}_{\mathfrak{p}})_{red}) & \xleftarrow{t^*} & \mathrm{Pic}^0(\widetilde{M}_0(\mathfrak{p}) \times_A \mathbf{k}_{\mathfrak{p}}) \\ {\scriptstyle s|^*_{\overline{S}_1}}\uparrow & & {\scriptstyle s^*}\uparrow & & {\scriptstyle r^*}\uparrow \\ \mathrm{Pic}^0((\overline{S}_1)_{red}) & \xleftarrow{i^*} & \mathrm{Pic}^0((\overline{M}_1(\mathfrak{p}) \times_A \mathbf{k}_{\mathfrak{p}})_{red}) & \xleftarrow{\overline{p}^*} & \mathrm{Pic}^0(\overline{M}_0(\mathfrak{p}) \times_A \mathbf{k}_{\mathfrak{p}}) \end{array}$$

By the classical theorem of Raynauld already quoted above it will be sufficient to show that the map t^* is injective in order to prove the second claim. Because r^* is bijective, trivial diagram chasing shows that we only have to prove that the composition $s|^*_{\overline{S}_1} \circ i^* \circ \overline{p}^*$ is injective. The scheme $(\overline{S}_1)_{red}$ has two irreducible components. Their image with respect to $\overline{p}$ intersect inside of the Zariski open set $M_0(\mathfrak{p})$ only, so they themselves intersect inside of the Zariski open set $M_1(\mathfrak{p})$ only. Hence Zariski's main theorem implies that the pre-image of every cross-point of $(\overline{S}_1)_{red}$ is connected in $(\widetilde{S}_1)_{red}$ as $M_1(\mathfrak{p})$ is normal. Therefore the restriction $s|^*_{\overline{S}_1}$ of the map s^* is injective on the toric part of the semi-abelian variety $\mathrm{Pic}^0((\overline{S}_1)_{red})$. On the other hand the composition of $\overline{p}^*$ and the map i^* induced by the closed immersion $i : (\overline{S}_1)_{red} \to (\overline{M}_1(\mathfrak{p}) \times_A \mathbf{k}_{\mathfrak{p}})_{red}$ is injective which can be seen by applying the argument in the proof of Proposition 11.9 of [14], pages 102-103. The semi-abelian variety $\mathrm{Pic}^0(\overline{M}_0(\mathfrak{p}) \times_A \mathbf{k}_{\mathfrak{p}})$ is a torus, so the composition $i^* \circ \overline{p}^*$ maps into the toric part of the semi-abelian variety $\mathrm{Pic}^0((\overline{S}_1)_{red})$. Therefore the homomorphism $s|^*_{\overline{S}_1} \circ i^* \circ \overline{p}^*$ is injective, as claimed. $\square$

In the next claim and its proof we let $J_0(\mathfrak{p})_l$ and $J_0(\mathfrak{p})$ denote the Galois module $J_0(\mathfrak{p})(\overline{F})$ and the Néron model of the Jacobian $J_0(\mathfrak{p})$, respectively.

Lemma 8.19. *Let l be an Eisenstein prime and let B be a subgroup of either $\mathcal{C}(\mathfrak{p})_l$ or $\mathcal{S}(\mathfrak{p})_l$. Then we have an exact sequence:*

$$0 \to B \to J_0(\mathfrak{p})_l^I \to (J_0(\mathfrak{p})_l/B)^I \to 0,$$

where the subscript denotes the module of elements fixed under the action of the inertia group I at $\mathfrak{p}$.

Proof. (Compare with Lemma 16.5 of [14], pages 125-126). What we need to show is that the map $J_0(\mathfrak{p})_l^I \to (J_0(\mathfrak{p})_l/B)^I$ is surjective. Any element of $J_0(\mathfrak{p})_l^I$ is fixed by the absolute Galois group of some finite, unramified extension K of $F_{\mathfrak{p}}$. Since the formation of Néron models commutes with unramified base change, the group $\mathcal{C}(\mathfrak{p})$ maps isomorphically onto the group of components of $J_0(\mathfrak{p})$ over K. Hence $J_0(\mathfrak{p})_l^I = J_{00}(\mathfrak{p})(\overline{\mathbf{f}}_{\mathfrak{p}})_l \times \mathcal{C}(\mathfrak{p})$, where $J_{00}(\mathfrak{p})$ is the connected component. Because $J_0(\mathfrak{p})$ has semi-stable reduction, the monodromy filtration on $J_0(\mathfrak{p})_l$ has two steps, in other words $(\gamma - 1)e \in J_0(\mathfrak{p})_l^I$ for any $e \in J_0(\mathfrak{p})_l$ and $\gamma \in I$. Since $J_0(\mathfrak{p})_l$ is an l-divisible group, its image under the map $\gamma - 1$ is l-divisible, too. As l divides the order of $\mathcal{C}(\mathfrak{p})$ the l-divisible part of $J_0(\mathfrak{p})_l^I$ is the factor $J_{00}(\mathfrak{p})(\overline{\mathbf{f}}_{\mathfrak{p}})_l$ of the direct product decomposition above. We may conclude that $(\gamma-1)e$ must lie in $J_{00}(\mathfrak{p})(\overline{\mathbf{f}}_{\mathfrak{p}})$. Let $\overline{e}$ be any element of $(J_0(\mathfrak{p})_l/B)^I$ and take an element e in $J_0(\mathfrak{p})_l$ which maps to $\overline{e}$. For any $\gamma\in I$ we have $(\gamma-1)e\in B$ by definition. By the above this expression also lies in $J_{00}(\mathfrak{p})(\overline{\mathbf{f}}_{\mathfrak{p}})_l$ whose intersection with B is trivial because both $\mathcal{C}(\mathfrak{p})$ and $\mathcal{S}(\mathfrak{p})$ has trivial intersection with that group. (The former is proved in 5.11 of [6], page 235, the latter is *(ii)* of Proposition 8.18). Hence $e\in J_0(\mathfrak{p})_l^I$. □

9. THE GROUP SCHEME $\mathcal{D}(\mathfrak{p})[l]$

DEFINITION 9.1. The subgroup $B(A)$ of upper triangular matrices of $GL_2(A)$ is the stabilizer of the point ∞ on the projective line in $GL_2(A)$ with respect to the Möbius action. Also note that $B(A)$ leaves the set $\Omega_c=\{z\in\Omega|c\leq|z|_i\}$ invariant for any positive $c\in\mathbb{Q}$. If $u:\Omega\to\mathbb{C}_\infty^*$ is a $B(A)$-invariant holomorphic function then its van der Put logarithmic derivative $r(u):GL_2(F_\infty)\to\mathbb{Z}$ is also invariant with respect to the left regular action of $B(A)$. In particular the integral

$$r(u)^0=\int_{A\backslash F_\infty} r(u)\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\mathrm{d}\mu_\infty(x)$$

is well-defined, where μ_∞ is the Haar measure introduced in Definition 5.1. Let $e(z):\Omega\to\mathbb{C}_\infty^*$ denote the classical Carlitz-exponential:

$$e(z)=z\prod_{0\neq\lambda\in A}\left(1-\frac{z}{\lambda}\right)$$

and define $t(z)$ as $e(z)^{q-1}$. It is well known (see for example 2.7 of [11], page 44-45) that the function t^{-1} is $B(A)$-invariant and it is a biholomorphic map between the quotient $B(A)\backslash\Omega_c$ and a small open disc around 0 punctured at 0 for a sufficiently large c. We say that the $B(A)$-invariant holomorphic function u on Ω is meromorphic at ∞ if the composition of u and the inverse of the biholomorphic map t is meromorphic at 0 for some (and hence all) such c number. In this case we can speak about its value, order of zero or order of pole at ∞. Of course our definition is just a specialization of the general definition in [5].

PROPOSITION 9.2. *Assume that the holomorphic function $u:\Omega\to\mathbb{C}_\infty^*$ is $B(A)$-invariant and it is meromorphic at ∞ in the sense defined above. Then its order of vanishing at ∞ is equal to $r(u)^0/(q-1)$.*

PROOF. It is sufficient to prove the claim in the following two cases:

(i) the function u is non-zero at ∞,
(ii) the function u is equal to $t(z)$.

In the first case we need to show that $r(u)^0=0$. Let υ be a uniformizer of F_∞ as in Definition 3.4. Since $r(u)$ is a harmonic cochain on the Bruhat-Tits tree of $GL_2(F_\infty)$, it satisfies the identity:

$$r(u)(g)=\sum_{\epsilon\in\mathbb{F}_q} r(u)(g\begin{pmatrix}\upsilon & \epsilon\\ 0 & 1\end{pmatrix})$$

for all $g \in GL_2(F_\infty)$. By an n-fold application of this identity we get the formula:

$$\begin{aligned}\int_{A\backslash F_\infty} r(u)\begin{pmatrix} \upsilon^{-n} & x \\ 0 & 1\end{pmatrix} \mathrm{d}\mu_\infty(x) =& \sum_{\epsilon\in\mathbb{F}_q}\int_{A\backslash F_\infty} r(u)\begin{pmatrix} \upsilon^{1-n} & x+\epsilon \\ 0 & 1\end{pmatrix} \mathrm{d}\mu_\infty(x) \\ =& q\int_{A\backslash F_\infty} r(u)\begin{pmatrix} \upsilon^{1-n} & x \\ 0 & 1\end{pmatrix} \mathrm{d}\mu_\infty(x) \\ =& \ldots = q^n r(u)^0.\end{aligned}$$

Because u is non-zero at ∞, its absolute value is constant on the the set Ω_c for a sufficiently large c as the latter set maps to a small neighborhood of 0 with respect to t^{-1}. Choose the natural number n large enough such that the positive number $c = |\upsilon^{-n}|$ has the property above. For every $\rho \in GL_2(F_\infty)$ let $C(\rho)$ denote the annulus

$$C(\rho) = \{z \in \mathbb{P}^1(\mathbb{C}_\infty) | 1 = |\rho^{-1}(z)|\}.$$

By our assumptions the holomorphic function u has constant absolute value on the non-empty affinoid subdomain $C(\begin{pmatrix} \upsilon^{-n} & x \\ 0 & 1\end{pmatrix}) \cap \Omega_c$ for any $x \in F_\infty$ hence either by its description in 1.7.3 of [11], page 40 as a difference of logarithms of absolute values on subdomains of this affinoid or by the results of [16], the value of $r(u)(\begin{pmatrix} \upsilon^{-n} & x \\ 0 & 1\end{pmatrix})$ is zero. Hence the integral on the left in the equation above is also zero which implies that $r(u)^0$ is zero, too.
In the second case we need to show that $r(t(z))^0 = 1 - q$. By definition:

$$r(e(z))(g) = -|\{\lambda \in A | \lambda \notin D(g)\}|$$

for every $g \in GL_2(F_\infty)$ such that $\infty \in D(g)$. As

$$\infty \in D(\begin{pmatrix} 1 & x \\ 0 & 1\end{pmatrix}) = \{z \in \mathbb{P}^1(\mathbb{C}_\infty) | 1 < |z - x|\}$$

for any $x \in F_\infty$, we get:

$$r(e(z))(\begin{pmatrix} 1 & x \\ 0 & 1\end{pmatrix}) = -|\{\lambda \in A| |\lambda - x| \leq 1\}.$$

Since for every $x \in F_\infty$ there are exactly q elements λ of A such that $|\lambda - x| \leq 1$ holds, we get that $r(t(z)) = -(q-1)q\mu_\infty(A\backslash F_\infty) = 1 - q$. $\square$

Proposition 9.3. *(i) In $\mathcal{C}(\mathfrak{p})$ the kernel of the specialization map into the group of connected components of the special fiber of the Néron model of $J_0(\mathfrak{p})$ at ∞ is its unique cyclic group of order $t(\mathfrak{p})$.*
(ii) The intersection of $\mathcal{C}(\mathfrak{p})$ and $\mathcal{S}(\mathfrak{p})$ is their unique cyclic group of order $t(\mathfrak{p})$.

Proof. Claim (i) of the proposition above is just (i) of Theorem 5.9 in [7], page 371. The intersection of $\mathcal{C}(\mathfrak{p})$ and $\mathcal{S}(\mathfrak{p})$ is a constant and a μ-type Galois module at the same time, so it is contained in the unique cyclic group of order $t(\mathfrak{p})$ in the cuspidal divisor group. Hence it is sufficient to prove that the latter lies in the kernel of the homomorphism $J_0(\mathfrak{p}) \to J_2(\mathfrak{p})$ induced by Picard functoriality. By Corollary 3.18 of [8] on page 198 the modular unit $\Delta/\Delta_{\mathfrak{p}}$ admits an $r(\mathfrak{p})$-th root in $\mathcal{O}^*(\Omega)$, where $r(\mathfrak{p}) = (q-1)^2$, if d is odd, and $r(\mathfrak{p}) = (q-1)^2(q+1)$, if d is even. (Incidentally, the latter also follows from Lemma 6.7.) Let $D_{\mathfrak{p}}$ be such a root. By Theorem 3.20 of [8], page 199 the latter transforms under $\Gamma_0(\mathfrak{p})$ through a certain character $\omega_{\mathfrak{p}} : \Gamma_0(\mathfrak{p}) \to \mathbb{C}_\infty^*$ of order $q-1$ such that $\omega_{\mathfrak{p}}^{(q-1)/t(\mathfrak{p})}$ is trivial on $\Gamma_2(\mathfrak{p})$ using the notations of the proof of Proposition 8.18. Hence $D_{\mathfrak{p}}^{(q-1)/t(\mathfrak{p})}$ defines a rational function on $X_2(\mathfrak{p})$ whose divisor generates the pull-back of the subgroup above. □

Definition 9.4. Let l be a prime dividing $t(\mathfrak{p})$. We are going to construct a group scheme $\mathcal{D}(\mathfrak{p})[l]$ which will play a role similar to $\mathcal{S}(\mathfrak{p})[l] \oplus \mathcal{C}(\mathfrak{p})[l]$ for Eisenstein primes l not dividing $t(\mathfrak{p})$. Let $l(\mathfrak{p})$ be the largest l-power dividing $t(\mathfrak{p})$. Assume first that l divides $\frac{N(\mathfrak{p})}{l(\mathfrak{p})}$. In light of the proposition above it is clear that in this case there is an $x \in \mathcal{S}(\mathfrak{p})$ and a $y \in \mathcal{C}(\mathfrak{p})$ such that

- (i) the order of x and y are both equal to $l \cdot l(\mathfrak{p})$,
- (ii) we have $lx = ly \in \mathcal{S}(\mathfrak{p}) \cap \mathcal{C}(\mathfrak{p})$,
- (iii) the natural topological generator Frob of the maximal constant field extension of F maps x to $(1+\alpha l(\mathfrak{p}))x$ for some $1 \leq \alpha < l$ integer.

Property (iii) holds because the Galois module generated by x is isomorphic to $\mu_{l \cdot l(\mathfrak{p})}$ by property (i). We define $\mathcal{D}(\mathfrak{p})[l]$ as the group generated by $u = x - y$ and $v = \alpha l(\mathfrak{p})x = \alpha l(\mathfrak{p})y$.

Lemma 9.5. *The group $\mathcal{D}(\mathfrak{p})[l]$ is l-torsion, Galois-invariant and as a Galois module everywhere unramified.*

Proof. The order of u, v is l by (i) and (ii) of the preceding paragraph above, so the first claim holds. The element v is fixed by the absolute Galois group and the latter acts on u through its maximal unramified quotient. By (iii) above $\mathrm{Frob}(u) = u + v$, so the last two claims are true as well. □

Remark 9.6. By the above $\mathcal{D}(\mathfrak{p})[l]$ contains $\mathcal{S}(\mathfrak{p})[l]$ and its quotient by this subgroup is a constant Galois module of order l which will be denoted by $\mathcal{F}(\mathfrak{p})[l]$. The simple construction above does not exist when l does not divide $\frac{N(\mathfrak{p})}{l(\mathfrak{p})}$. In this case we will give another, more involved construction which will be denoted by $\mathcal{D}(\mathfrak{p})[l]$, too. Actually this case occurs, here is a little analysis. First assume that d is odd. Since $t(\mathfrak{p})$ is the greatest common divisor of d and $q-1$ in this case, we may compute as follows:

$$N(\mathfrak{p}) = \sum_{k=0}^{d-1}(1+(q-1))^k \equiv \sum_{k=0}^{d-1} 1+k(q-1) \equiv d+(q-1)\frac{d(d-1)}{2} \equiv d \mod l \cdot l(\mathfrak{p}).$$

Hence the phenomenon occurs if and only if $l \cdot l(\mathfrak{p})$ does not divide d. Now we consider the case when d is even. In this case $t(\mathfrak{p})$ is the greatest common divisor of $d/2$ and $q-1$, so we may compute as follows:

$$\begin{aligned} N(\mathfrak{p}) &= \sum_{k=0}^{d/2-1} (1+(q-1))^{2k} \\ &\equiv \sum_{k=0}^{d/2-1} 1+2k(q-1) \equiv \frac{d}{2}+(q-1)\frac{d(d-2)}{4} \equiv \frac{d}{2} \bmod l \cdot l(\mathfrak{p}). \end{aligned}$$

Hence the phenomenon occurs if and only if $l \cdot l(\mathfrak{p})$ does not divide $d/2$. Obviously these conditions can always be satisfied by choosing an appropriate d.

Notation 9.7. We start our construction by introducing a set of new notations and definitions. Every $\alpha \in \mathbf{f}_{\mathfrak{p}}$ is represented by a unique element of $\mathbb{F}_q[T]$ whose degree is less then $\deg(\mathfrak{p})$, which will be denoted by the same symbol by abuse of notation. Let $\Gamma(\mathfrak{p}) \lhd GL_2(A)$ be the principal congruence subgroup of level $\mathfrak{p}$, that is the kernel of the reduction map $GL_2(A) \to GL_2(A/\mathfrak{p})$. For every $\underline{0} \neq (\alpha,\beta) \in \mathbf{f}_{\mathfrak{p}}^2$ let $(\alpha:\beta)$ denote the set of points $(a:b) \in \mathbb{P}^1(F)$ where a and b are in A, they are relatively prime and $(a,b) \equiv (\alpha,\beta) \mod \mathfrak{p}$. This set is an orbit of the natural left action of $\Gamma(\mathfrak{p})$ on $\mathbb{P}^1(F)$. As the quotient $\Gamma(\mathfrak{p})\backslash\mathbb{P}^1(F)$ is the set of cusps of the Drinfeld modular curve $\Gamma(\mathfrak{p})\backslash\Omega$ parameterizing Drinfeld modules of rank two equipped with a full level $\mathfrak{p}$-structure, we may identify the set $(\alpha:\beta)$ and the cusp it represents.

Definition 9.8. Let $\pi : \mathbf{f}_{\mathfrak{p}}^* \to \mathbf{f}_{\mathfrak{p}}^*/\mathbb{F}_q^*$ be the canonical surjection and let $I \subset \mathbf{f}_{\mathfrak{p}}^*$ be a complete set of representatives of the cosets of the projection π. We will specify a convenient choice of I later. Let $\phi : \mathbf{f}_{\mathfrak{p}}^*/\mathbb{F}_q^* \to \mu_l \subseteq \mathbb{F}_q^*$ be the unique surjection onto the l-th roots of unity. For every $\alpha \in \mathbf{f}_{\mathfrak{p}}^*$ let $\overline{\alpha}$ denote $\phi \circ \pi(\alpha)$ and for every $d \in A$ not in $\mathfrak{p}$ let $\overline{d}$ similarly denote the value of $\phi \circ \pi$ on the reduction of d mod $\mathfrak{p}$ by slight abuse of notation. For every $x \in \mu_l$ let $C_I(x) \subset \mathbf{f}_{\mathfrak{p}}^*$ be the set $\{\alpha \in I | \overline{\alpha} = x\}$. For any ring R let $R[\mu_l]_0$ denote the set of all R-valued functions on μ_l whose sum over the elements of μ_l is zero. For every $D \in \mathbb{Z}[\mu_l]_0$ we define the holomorphic function $\epsilon_D : \Omega \to \mathbb{C}_\infty^*$ as the product:

$$\epsilon_D(z) = \prod_{x \in \mu_l} \prod_{\alpha \in C_I(x)} \epsilon_{\mathfrak{p}}(0,\alpha)(z)^{D(x)}.$$

Definition 9.9. Let $Y_{\#l}(\mathfrak{p}) \to Y_0(\mathfrak{p})$ denote the unique covering intermediate of the covering $Y_2(\mathfrak{p}) \to Y_0(\mathfrak{p})$ which is a cyclic Galois covering of order l. Let $J_{\#l}(\mathfrak{p})$ denote the Jacobian of the unique geometrically irreducible non-singular projective curve $X_{\#l}(\mathfrak{p})$ containing $Y_{\#l}(\mathfrak{p})$. The kernel of the map $J_0(\mathfrak{p}) \to J_{\#l}(\mathfrak{p})$ induced by Picard functoriality is the unique subgroup of the Shimura group of order l. The set of geometric points of $X_{\#l}(\mathfrak{p})$ in the complement of $Y_{\#l}(\mathfrak{p})$ are the cusps of $X_{\#l}(\mathfrak{p})$. The quotients $\Gamma_1(\mathfrak{p})\backslash\Omega$ and $\Gamma_{\#l}(\mathfrak{p})\backslash\Omega$ of the arithmetic subgroups

$$\Gamma_1(\mathfrak{p}) =\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathcal{O})|c \equiv 0 \bmod \mathfrak{p}, a \equiv 1 \bmod \mathfrak{p}\} \text{ and}$$
$$\Gamma_{\#l}(\mathfrak{p}) =\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathcal{O})|c \equiv 0 \bmod \mathfrak{p}, \overline{a} = 1\}$$

of $GL_2(A)$ are the modular curves $Y_1(\mathfrak{p})$ and $Y_{\#l}(\mathfrak{p})$, respectively. Since for every subgroup $\Gamma \leq GL_2(A)$ the set of cusps of the modular curve $\Gamma\backslash\Omega$ is the quotient $\Gamma\backslash\mathbb{P}^1(F)$, the set

$$\{(\alpha : 0)|0 \neq \alpha \in \mathbf{f}_{\mathfrak{p}}\} \cup \{(0 : \beta)|0 \neq \beta \in \mathbf{f}_{\mathfrak{p}}\}$$

is a full set of representatives for the cusps of $Y_1(\mathfrak{p})$. It is also clear that set of sets above also represent the cusps of $Y_{\#l}(\mathfrak{p})$ and the sets $(\alpha : 0)$ and $(\beta : 0)$ (respectively $(0 : \alpha)$ and $(0 : \beta)$) represent the same cusp if and only if $\overline{\alpha} = \overline{\beta}$.

Proposition 9.10. *The function ϵ_D is a modular unit on $Y_{\#l}(\mathfrak{p})$ defined over F.*

Proof. For every $(\alpha, \beta) \in \mathbf{f}_{\mathfrak{p}}^2$ we have the following transformation law:

$$\epsilon_{\mathfrak{p}}(\alpha,\beta)\left(\frac{az+b}{cz+d}\right) = \frac{1}{cz+d}\epsilon_{\mathfrak{p}}(a\alpha+c\beta, b\alpha+d\beta)(z), \quad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(A).$$

From this formula it is clear the every holomorphic function which is the the product of functions of the form $\epsilon_{\mathfrak{p}}(\alpha,\beta)(z)/\epsilon_{\mathfrak{p}}(\alpha',\beta')(z)$, such as ϵ_D, is invariant under the action of $\Gamma(\mathfrak{p})$, so it defines a holomorphic function on the Drinfeld modular curve $Y(\mathfrak{p}) = \Gamma(\mathfrak{p})\backslash\Omega$ parameterizing Drinfeld modules with full level $\mathfrak{p}$-structure. Moreover every function on $Y(\mathfrak{p})$ arising from such a quotient is the base change to $\mathbb{C}_\infty$ of the universal modular object associating to every rank two Drinfeld module $\psi : A \to K\{\tau\}$ of general characteristic equipped with a level structure $\iota : \mathbf{f}_{\mathfrak{p}}^2 \to \psi[\mathfrak{p}]$ the fraction $\iota(\alpha,\beta)(z)/\iota(\alpha',\beta')$, so it is a modular unit defined over F. Hence we only have to show that the function ϵ_D is actually invariant under the action of $\Gamma_{\#l}(\mathfrak{p})$, too.
For every $\beta \in \mathbf{f}_{\mathfrak{p}}^*$ and $\alpha \in C_I(x)$ there is a unique $\alpha_\beta \in C_I(\overline{\beta}x)$ and a $t_\alpha(\beta) \in \mathbb{F}_q^*$ such that $\beta\alpha = t_\alpha(\beta)\alpha_\beta$. Clearly the map $C_I(x) \to C_I(\overline{\beta}x)$ given by the rule $\alpha \mapsto \alpha_\beta$ is bijective. Hence

$$\prod_{\alpha\in C_I(x)} t_\alpha(\beta) \cdot \prod_{\gamma\in C_I(\overline{\beta}x)} \gamma = \prod_{\alpha\in C_I(x)} t_\alpha(\beta)\alpha_\beta = \beta^{\frac{q^d-1}{l(q-1)}} \cdot \prod_{\alpha\in C_I(x)} \alpha.$$

Substituting the equation above into the third line of the equation below we get the following identity:

$$\begin{aligned}\epsilon_D\left(\frac{az+b}{cz+d}\right) &= \prod_{x\in\mu_l}\prod_{\alpha\in C_I(x)}\frac{1}{(cz+d)^{D(x)}}\epsilon_{\mathfrak{p}}(0,d\alpha)(z)^{D(x)}\\ &= \prod_{x\in\mu_l}(cz+d)^{-\frac{D(x)(q^d-1)}{l(q-1)}}\prod_{\alpha\in C_I(x)}\epsilon_{\mathfrak{p}}(0,t_\alpha(d)\alpha_d)(z)^{D(x)}\\ &= \prod_{x\in\mu_l}\prod_{\alpha\in C_I(x)}t_\alpha(d)^{D(x)}\cdot\prod_{y\in\mu_l}\prod_{\beta\in C_I(\overline{d}y)}\epsilon_{\mathfrak{p}}(0,\beta)(z)^{D(y)}\\ &= \prod_{\alpha\in I}\alpha^{D(\overline{\alpha})-D(\overline{d\alpha})}\cdot\epsilon_{D(\overline{d}\cdot)}(z),\quad\forall\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in\Gamma_0(\mathfrak{p}),\end{aligned}$$

where we also used the transformation law at the start of our proof in the first equation and the simple identity $\epsilon_{\mathfrak{p}}(\gamma\alpha,\gamma\beta)(z)=\gamma\epsilon_{\mathfrak{p}}(\alpha,\beta)(z)$ valid for all $\gamma\in\mathbb{F}_q^*$ in the third equation. From this identity the claim follows immediately. $\square$

Lemma 9.11. *For any $0\neq z\in A\subset\mathbb{A}_f^*$ and $(\alpha,\beta)\in(\mathbf{f}_{\mathfrak{p}})^2$ we have:*

$$\int_{F\backslash\mathbb{A}}r(\epsilon_{\mathfrak{p}}(\alpha,\beta))(\begin{pmatrix}z^{-1} & x\\ 0 & 1\end{pmatrix})\mathrm{d}\mu(x)=1-\rho(\alpha)\rho(\beta)-q^{1+\deg(\alpha z)-\deg(\mathfrak{p})}(1-\rho(\alpha)).$$

Proof. Recall our convention which for every $v\in\mathbb{A}_f^*$ denote the unique idele whose finite component is v and whose ∞-adic component is 1 by the symbol v as well. The equation above should be understood in this sense. By the Limit Formula 4.8 the restriction of $r(\epsilon_{\mathfrak{p}}(\alpha,\beta))$ onto $B(\mathbb{A})$ is the limit of automorphic forms, so in particular it is $B(F)$-invariant. Hence the integral on the right hand side in the equation above, which we will denote by $r(\epsilon_{\mathfrak{p}}(\alpha,\beta))^0(z^{-1})$, is well-defined. Fix an $f\in A$ generator of the ideal $\mathfrak{p}$. For all $\alpha\in\mathbf{f}_{\mathfrak{p}}$ we have:

$$\zeta_{\mathfrak{p}}(\alpha,z^{-1},s)=\sum_{\substack{0\neq u\in\mathbb{F}_q[T]\\ u\equiv\alpha z\bmod(zf)}}q^{-s\deg(u)}.$$

Applying the same argument as in the proof of Proposition 5.8, we get that:

$$\zeta_{\mathfrak{p}}(\alpha,z^{-1},s)=(1-\rho(\alpha))q^{-s\deg(\alpha z)}+\frac{(q-1)q^{-s\deg(fz)}}{1-q^{1-s}}.$$

An immediate consequence of this equation and Proposition 5.2 is that the function $E_{\mathfrak{p}}(\alpha,\beta,\cdot,s)^0(z^{-1})$, originally defined for $\mathrm{Re}(s)>1$ only, has a meromorphic continuation to the whole complex plane and

$$E_{\mathfrak{p}}(\alpha,\beta,\cdot,s)^0(z^{-1})=-\rho(\alpha)\rho(\beta)-q^{1+\deg(\alpha z)-\deg(\mathfrak{p})}(1-\rho(\alpha))+\frac{q^{1+\deg(z)}}{q+1},$$

arguing the same was as in the proof of Proposition 5.8 and using the fact that $|\mathfrak{p}|=q^{-\deg(\mathfrak{p})}$. Hence by the Limit Formula 4.8 the following equation holds:

$$\begin{aligned}r(\epsilon_{\mathfrak{p}}(\alpha,\beta))^0(z^{-1})=&E_{\mathfrak{p}}(\alpha,\beta,\cdot,0)^0(z^{-1})-E_{\mathfrak{p}}(0,0,\cdot,0)^0(z^{-1}))\\ =&1-\rho(\alpha)\rho(\beta)-q^{1+\deg(\alpha z)-\deg(\mathfrak{p})}(1-\rho(\alpha)).\ \square\end{aligned}$$

Proposition 9.12. *For any $D \in \mathbb{Z}[\mu_l]_0$ and $\beta \in \mathbf{f}_{\mathfrak{p}}^*$ the order of vanishing of the modular unit ϵ_D at the cusp $(0:\beta)$ is zero and at the cusp $(\beta:0)$ is equal to:*

$$\frac{1}{1-q}\sum_{\alpha\in I} D(\overline{\alpha}/\overline{\beta})q^{\deg(\alpha)}.$$

Proof. Let $\mathrm{ord}_{(\alpha:\beta)}(u)$ denote the order of vanishing of any modular unit u on the curve $Y_{\#l}(\mathfrak{p})$ at the cusp $(\alpha:\beta)$. Let $\begin{pmatrix} a & b \\ c & d\end{pmatrix} \in \Gamma_0(\mathfrak{p})$ be a matrix such that $d \equiv \beta \mod \mathfrak{p}$. Then:

$$\begin{aligned}\mathrm{ord}_{(0:\beta)}\epsilon_D(z) =& \mathrm{ord}_{(0:\beta)}\epsilon_{D(\overline{\beta}^{-1}\cdot)}(\frac{az+b}{cz+d}) = \mathrm{ord}_{(0:1)}\epsilon_{D(\overline{\beta}^{-1}\cdot)}(z) = \frac{r(\epsilon_{D(\overline{\beta}^{-1}\cdot)})^0}{q-1}\\ =&\frac{1}{q^2-q}\sum_{\alpha\in I} D(\overline{\alpha}/\overline{\beta})r(\epsilon_{\mathfrak{p}}(0,\alpha))^0(1) = \frac{1}{q^2-q}\sum_{\alpha\in I} D(\overline{\alpha}/\overline{\beta}) = 0.\end{aligned}$$

Let us explain why this sequence of equalities hold. The first equation is the consequence of the transformation rule we derived at the end of the proof of Proposition 9.10. The second equation follows from the fact that the image of the cusp $(0:1)$ under the automorphism of $Y_{\#l}(\mathfrak{p})$ induced by $\begin{pmatrix} a & b \\ c & d\end{pmatrix}$ is the cusp $(0:\beta)$. The group $\Gamma_{\#l}(\mathfrak{p})$ contains $B(A)$, so the third equation is just a special case of Proposition 9.2. Note that for every $g: F\backslash\mathbb{A} \to \mathbb{C}$ continuous and $\mathcal{O}_f$-translation invariant function there is a unique $g_\infty : A\backslash F_\infty \to \mathbb{C}$ continuous function such that $g(x) = g_\infty(x)$ for every $x \in F_\infty$. Moreover

$$\int_{F\backslash\mathbb{A}} g\mathrm{d}\mu(x) = \mu_f(\mathcal{O}_f)\int_{A\backslash F_\infty} g_\infty(x)\mathrm{d}\mu_\infty(x)$$

using the notation of Definition 5.1. Hence the fourth equation follows from the relation between the usual van der Put derivative and its adelic version introduced in Notation 4.4. The fifth equation is just a special case of Lemma 9.11 and the last equation holds by definition.

For any $\begin{pmatrix} h & j \\ m & n\end{pmatrix} \in \Gamma_{\#l}(\mathfrak{p})$ we have $hn \in \mathbb{F}_q^* \subset \mathbf{f}_{\mathfrak{p}}^* \mod \mathfrak{p}$ and $\overline{h} = 1$ by definition, hence the equation $\overline{n} = 1$ also holds as $\mathbb{F}_q^*$ is in the kernel of $\phi\circ\pi$. Therefore the group $\Gamma_{\#l}(\mathfrak{p})$ is normalized by the matrix $\begin{pmatrix} 0 & 1 \\ f & 0\end{pmatrix}$, where $f \in A$ is again a generator of the prime ideal $\mathfrak{p}$. Hence this matrix induces an involution of the modular curve $Y_{\#l}(\mathfrak{p})$ exchanging the cusps $(0:\beta)$ and $(\beta:0)$. For every $H \in \mathbb{Z}[\mu_l]_0$ we define the holomorphic function $\widehat{\epsilon}_H : \Omega \to \mathbb{C}_\infty^*$ as the product:

$$\widehat{\epsilon}_H(z) = \prod_{x\in\mu_l}\prod_{\alpha\in C_I(x)} \epsilon_{\mathfrak{p}}(\alpha,0)(fz)^{H(x)}.$$

Then the transformation law at the start of the proof of Proposition 9.10 imply that

$$\widehat{\epsilon}_H(\frac{1}{fz}) = \epsilon_H(z),$$

in particular $\widehat{\epsilon}_H$ is also a modular unit on the curve $Y_{\#l}(\mathfrak{p})$. Therefore

$$\begin{aligned}
\mathrm{ord}_{(\beta:0)}\epsilon_D(z) =&\mathrm{ord}_{(\beta:0)}\epsilon_{D(\overline{\beta}^{-1}\cdot)}(\frac{az+b}{cz+d}) = \mathrm{ord}_{(1:0)}\epsilon_{D(\overline{\beta}^{-1}\cdot)}(z)\\
=&\mathrm{ord}_{(1:0)}\widehat{\epsilon}_{D(\overline{\beta}^{-1}\cdot)}(\frac{1}{fz}) = \mathrm{ord}_{(0:1)}\widehat{\epsilon}_{D(\overline{\beta}^{-1}\cdot)}(z) = \frac{r(\widehat{\epsilon}_{D(\overline{\beta}^{-1}\cdot)})^0}{q-1}\\
=&\frac{1}{q^2-q}\sum_{\alpha\in I} D(\overline{\alpha}/\overline{\beta})r(\epsilon_{\mathfrak{p}}(\alpha,0)(f\cdot))^0(1)\\
=&\frac{1}{q^2-q}\sum_{\alpha\in I} D(\overline{\alpha}/\overline{\beta})r(\epsilon_{\mathfrak{p}}(\alpha,0))^0(f^{-1})\\
=&\frac{1}{q^2-q}\sum_{\alpha\in I} D(\overline{\alpha}/\overline{\beta})(1-q^{1+\deg(\alpha)})\\
=&\frac{1}{1-q}\sum_{\alpha\in I} D(\overline{\alpha}/\overline{\beta})q^{\deg(\alpha)}. \ \square
\end{aligned}$$

Corollary 9.13. *Let $D \in \mathbb{Z}[\mu_l]_0$ be a function such that $D(y)$ mod l does not depend on $y \in \mu_l$. Then the divisor of the modular unit ϵ_D is divisible by l.*

Proof. The property of D in the claim above is clearly invariant under the action of the group ring $\mathbb{Z}[\mu_l]$, hence it is sufficient to show that

$$\sum_{y\in\mu_l} D(y) \sum_{\alpha\in C_I(y)} q^{\deg(\alpha)} \equiv 0 \mod (q-1)l$$

when $D(y) = 1 + lC(y)$ for some function $C : \mu_l \to \mathbb{Z}$. We have

$$\begin{aligned}
\sum_{\alpha\in C_I(y)} q^{\deg(\alpha)} &= \sum_{\alpha\in C_I(y)} (1+(q-1))^{\deg(\alpha)}\\
&\equiv \sum_{\alpha\in C_I(y)} (1+(q-1)\deg(\alpha)) \mod (q-1)l
\end{aligned}$$

for any $y \in \mu_l$, therefore

$$\begin{aligned}
\sum_{y\in\mu_l} D(y) \sum_{\alpha\in C_I(y)} q^{\deg(\alpha)} \equiv&\frac{q^d-1}{(q-1)l}\sum_{y\in\mu_l} D(y)\\
&+ (q-1)\sum_{y\in\mu_l}(1+lC(y))\sum_{\alpha\in C_I(y)}\deg(\alpha)\\
\equiv&(q-1)\sum_{\alpha\in I}\deg(\alpha) \equiv (q-1)\sum_{j=0}^{d-1} jq^j\\
\equiv&(q-1)\sum_{j=0}^{d-1} j \equiv \frac{(q-1)d(d-1)}{2} \mod (q-1)l.
\end{aligned}$$

If l is odd then 2 is invertible modulo l and l divides d. If $l=2$ then d is even and l divides $d/2$. □

Lemma 9.14. *(i) If $M\leq\mathbb{Z}[\mu_l]_0$ is a μ_l-invariant $\mathbb{Z}$-submodule then M is either trivial or its $\mathbb{Z}$-rank is $l-1$.*
(ii) If $M_1\lneqq M_2\leq\mathbb{Z}_l[\mu_l]_0$ are non-trivial μ_l-invariant $\mathbb{Z}_l$-submodules and the natural μ_l-action on the quotient M_2/M_1 is trivial then M_2/M_1 is a cyclic group of order l.

Proof. The $\overline{\mathbb{Q}}$-span $M_{\overline{\mathbb{Q}}}$ of M in the $\overline{\mathbb{Q}}$-vectorspace $\overline{\mathbb{Q}}[\mu_l]_0$ has the same $\overline{\mathbb{Q}}$-rank as the $\mathbb{Z}$-rank of the free $\mathbb{Z}$-module M. Since $M_{\overline{\mathbb{Q}}}$ is also μ_l-invariant, it is the direct sum of some of the irreducible μ_l-invariant subspaces. On the other hand it is also fixed by the natural action of the absolute Galois group of $\mathbb{Q}$ on the tensor product $\overline{\mathbb{Q}}[\mu_l]_0=\mathbb{Q}[\mu_l]_0\otimes\overline{\mathbb{Q}}$. This action permutes the irreducible μ_l-invariant subspaces transitively, therefore $M_{\overline{\mathbb{Q}}}$ is either trivial or it is the whole $\overline{\mathbb{Q}}$-vector space.
We start the proof of the second claim by noting that the first claim also holds when the role of the ring $\mathbb{Z}$ is played by the ring $\mathbb{Z}_l$. The proof is identical. Hence for every non-trivial μ_l-invariant $\mathbb{Z}_l$-submodule $M\leq\mathbb{Z}_l[\mu_l]_0$ there is a unique natural number $n(M)\in\mathbb{N}$ such that $l^{n(M)}\mathbb{Z}_l[\mu_l]_0\leq M$ but $l^{n(M)-1}\mathbb{Z}_l[\mu_l]_0\not\leq M$. Let σ be a generator of μ_l. We are going to show that there is a natural number $m(M)\in\mathbb{N}$ such that M is the image of the endomorphism $x\mapsto(1-\sigma)^{m(M)}$ by induction on $n(M)$. The μ_l-invariant subgroups of the quotient $\mathbb{Z}_l[\mu_l]_0/\mathbb{Z}_l[\mu_l]_0=\mathbb{F}_l[\mu_l]_0$ are exactly the proper ideals of the group ring $\mathbb{F}_l[\mu_l]=\mathbb{F}_l[T]/(T-1)^l$. As the latter form a chain whose Jordan-Hölder components are all isomorphic to $\mathbb{F}_l$, the claim is now obvious when $n(M)=1$. Since the map $x\mapsto(1-\sigma)^{m(M)}$ is injective, the general case follows using induction and the same argument where the role of $\mathbb{Z}_l[\mu_l]_0$ is played by $M+l^{n(M)-1}\mathbb{Z}_l[\mu_l]_0$. Now claim (ii) follows. □

Definition 9.15. Let $\mathcal{F}_{\#l}(\mathfrak{p})\subset J_{\#l}(\mathfrak{p})(\overline{F})$ denote the Galois module generated by the linear equivalence classes of degree zero divisors supported on the cusps of $X_{\#l}(\mathfrak{p})$ mapping to the cusp 0 of the curve $X_0(\mathfrak{p})$. Moreover let $\mathcal{F}(\mathfrak{p})[l]\subseteq\mathcal{F}_{\#l}(\mathfrak{p})$ denote subgroup of elements of l-primary order fixed by the decking transformations of the cover $X_{\#l}(\mathfrak{p})\to X_0(\mathfrak{p})$. The next proposition partially justifies our choice of notation.

Proposition 9.16. *The group $\mathcal{F}(\mathfrak{p})[l]$ is cyclic of order l.*

Proof. Let $J\subset\mathrm{Ker}(\phi\circ\pi)\subset\mathbf{f}_{\mathfrak{p}}^*$ be a complete set of representatives of the cosets of the restriction of the projection π onto $\mathrm{Ker}(\phi\circ\pi)$. Moreover let ξ be a generator of the cyclic group $\mathbf{f}_{\mathfrak{p}}^*$. We define the set I as the union $\cup_{j=0}^{l-1}\xi^jJ$. Pick a $\begin{pmatrix} h & j \\ m & n\end{pmatrix}\in\Gamma_0(\mathfrak{p})$ matrix with $n=\xi$ and let $D\in\mathbb{Z}[\mu_l]_0$ be the function with $D(1)=1$, $D(\overline{\xi})=-1$ and all the other values are zero. Then

$$\prod_{j=0}^{l-1}\epsilon_D(\begin{pmatrix} h & j \\ m & n\end{pmatrix}^j z) = \prod_{\alpha\in J}\frac{\epsilon_{\mathfrak{p}}(0,\alpha)(z)}{\epsilon_{\mathfrak{p}}(0,\xi\alpha)(z)}\cdot\prod_{\alpha\in J}\frac{\epsilon_{\mathfrak{p}}(0,\xi\alpha)(z)}{\epsilon_{\mathfrak{p}}(0,\xi^2\alpha)(z)}\cdots\prod_{\alpha\in J}\frac{\epsilon_{\mathfrak{p}}(0,\xi^{l-1}\alpha)(z)}{\epsilon_{\mathfrak{p}}(0,\xi^l\alpha)(z)}$$
$$=\prod_{\alpha\in J}\frac{\epsilon_{\mathfrak{p}}(0,\alpha)(z)}{\epsilon_{\mathfrak{p}}(0,\xi^l\alpha)(z)} = \prod_{\alpha\in J} t_{\xi^l}(\alpha)^{-1} = \xi^{\frac{q^d-1}{1-q}} \notin (F^*)^l$$

using the notation of the proof of Proposition 9.10. Let $\mathcal{Y}$ and $\mathcal{P}$ denote the group of degree zero divisors supported on the cusps of $X_{\#l}(\mathfrak{p})$ mapping to the cusp 0 of the curve $X_0(\mathfrak{p})$ and its subgroup of principal divisors, respectively. Let $\mathcal{U}$ denote the group of divisors of units of the form ϵ_D introduced in Definition 9.8. Fix a non-zero element $\sigma\in\mu_l$. For every $y\in\mu_l$ let $z\mapsto z^y$ denote the decking transformation of $X_{\#l}(\mathfrak{p})$ corresponding to y. It is characterized by the property that it maps the cusp $(0:1)$ to $(0:\alpha)$ where $\overline{\alpha}=y$. For every non-zero F-rational function g on the curve $X_{\#l}(\mathfrak{p})$ whose divisor lies in $\mathcal{P}$ the divisor of the product $N(g)=\prod_{y\in\mu_l}g(z^y)$ is μ_l-invariant, hence it is trivial. Therefore $N(g)$ is constant. It is also clear that its class in $F^*/(F^*)^l$ only depends on the divisor of g modulo l. We let N denote the corresponding homomorphism $\mathcal{P}/l\mathcal{P}\to F^*/(F^*)^l$ as well. It is clear from the above that this homomorphism is non-trivial restricted to $\mathcal{U}/l\mathcal{U}$. An immediate consequence is that the μ_l-invariant modules $\mathcal{P}$ and $\mathcal{U}$ are non-trivial. Hence they have $\mathbb{Z}$-rank $l-1$ by claim (i) of Lemma 9.14. In particular the group $\mathcal{F}_{\#l}(\mathfrak{p})$ is torsion.
Note that the map N is μ_l-invariant, so it induces an embedding of $\mathcal{U}/(1-\sigma)\mathcal{U}$ into $F^*/(F^*)^l$. We claim that $\mathcal{U}\otimes\mathbb{Z}_l=\mathcal{P}\otimes\mathbb{Z}_l$. If this were false then there would be an element H of $\mathcal{P}$ such that $(1-\sigma)H$ lies in $\mathcal{U}$ but it does not lie in $(1-\sigma)\mathcal{U}$ by claim (ii) of Lemma 9.14. The latter can be applied as the module $\mathcal{Y}\otimes\mathbb{Z}_l$ is isomorphic to $\mathbb{Z}_l[\mu_l]_0$ as a μ_l-module. Since $N(g(z)/g(z^\sigma))=1$ for any F-rational function g on $X_{\#}(\mathfrak{p})$ whose divisor is in $\mathcal{P}$, we get a contradiction. On the other hand we claim that $\mathcal{P}\otimes\mathbb{Z}_l$ is strictly smaller than $\mathcal{Y}\otimes\mathbb{Z}_l$. By the above we only have to prove this for $\mathcal{U}\otimes\mathbb{Z}_l$. It will be enough to show that the unique smallest μ_l-invariant $\mathbb{Z}_l$-submodule of $\mathcal{U}\otimes\mathbb{Z}_l$ strictly larger than $l(\mathcal{U}\otimes\mathbb{Z}_l)$ is contained in $l(\mathcal{Y}\otimes\mathbb{Z}_l)$. But this is exactly the content of Corollary 9.13. Therefore the l-torsion of $\mathcal{F}_{\#l}(\mathfrak{p})$ is non-trivial, and the claim now follows from claim (ii) of Lemma 9.14. □

Definition 9.17. We define $\mathcal{D}(\mathfrak{p})[l]\subset J_0(\mathfrak{p})(\overline{F})$ to be the pre-image of $\mathcal{F}(\mathfrak{p})[l]$ under the map $J_0(\mathfrak{p})\to J_{\#l}(\mathfrak{p})$ induced by Picard functoriality. In this paragraph let $\overline{S}$ denote the base change of the F-scheme S to $\overline{F}$. Since the map $X_{\#l}(\mathfrak{p})\to X_0(\mathfrak{p})$ is a Galois covering with Galois group μ_l, there is a Hochschild-Serre spectral sequence $H^p(\mu_l,H^q(\overline{X_{\#l}(\mathfrak{p})},\mathbb{Q}_l/\mathbb{Z}_l))\Rightarrow H^{p+q}(\overline{X_0(\mathfrak{p})},\mathbb{Q}_l/\mathbb{Z}_l))$ which gives rise to an exact sequence

$$H^1(\overline{X_0(\mathfrak{p})},\mathbb{Q}_l/\mathbb{Z}_l))\to H^1(\overline{X_{\#l}(\mathfrak{p})},\mathbb{Q}_l/\mathbb{Z}_l)^{\mu_l}\to H^2(\mu_l,H^0(\overline{X_{\#l}(\mathfrak{p})},\mathbb{Q}_l/\mathbb{Z}_l))=0$$

By definition $\mathcal{D}(\mathfrak{p})[l]$ contains $\mathcal{S}(\mathfrak{p})[l]$ and its quotient by this subgroup is isomorphic to the Galois module $\mathcal{F}(\mathfrak{p})[l]$ by the above. Also note that $\mathcal{D}(\mathfrak{p})[l]$ is

l-torsion as our choice of notation indicates. We argue as follows: the composition of the morphisms $J_0(\mathfrak{p}) \to J_{\#l}(\mathfrak{p})$ and $J_{\#l}(\mathfrak{p}) \to J_0(\mathfrak{p})$ induced by Picard and Albanese functoriality respectively is multiplication by l on $J_0(\mathfrak{p})$. On the other hand the image of every element of $\mathcal{F}(\mathfrak{p})[l]$ under the Albanese map is represented by the direct image of a divisor supported on the pre-image of the cusp 0 under the map $X_{\#l}(\mathfrak{p}) \to X_0(\mathfrak{p})$, hence it must be zero.

Proposition 9.18. *The following holds:*

(*i*) *the Galois modules* $\mathcal{S}(\mathfrak{p})[l]$ *and* $\mathcal{F}(\mathfrak{p})[l]$ *are constant of order* l,

(*ii*) *the Galois module* $\mathcal{D}(\mathfrak{p})[l]$ *is everywhere unramified,*

(*iii*) *both* $\mathcal{S}(\mathfrak{p})[l]$ *and* $\mathcal{D}(\mathfrak{p})[l]$ *are* $\mathbb{T}(\mathfrak{p})$*-invariant and annihilated by the Eisenstein ideal,*

(*iv*) *the exact sequence:*

$$0 \to \mathcal{S}(\mathfrak{p})[l] \to \mathcal{D}(\mathfrak{p})[l] \to \mathcal{F}(\mathfrak{p})[l] \to 0$$

of Galois modules does not split over F,

(*v*) *the intersection of* $\mathcal{D}(\mathfrak{p})[l]$ *and* $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)[l]$ *is* $\mathcal{S}(\mathfrak{p})[l]$.

Proof. First consider the case when l divides $N(\mathfrak{p})/l(\mathfrak{p})$. Claims (*i*) and (*ii*) are immediate consequences of Lemma 9.5. In order to show claim (*iii*) it will be sufficient to show that both $\mathcal{C}(\mathfrak{p})$ and $\mathcal{S}(\mathfrak{p})$ are $\mathbb{T}(\mathfrak{p})$-invariant and annihilated by the Eisenstein ideal, since in this case every subgroup of the sum $\mathcal{C}(\mathfrak{p})+\mathcal{S}(\mathfrak{p})$ is fixed by the Eisenstein ideal as it acts on the latter by scalar multiplication. Using the same argument again we are reduced to show that $\mathcal{T}(\mathfrak{p})$ and $\mathcal{M}(\mathfrak{p})$ are $\mathbb{T}(\mathfrak{p})$-invariant and annihilated by the Eisenstein ideal. These groups are obviously Hecke-invariant, and the annihilation by the Eisenstein ideal follows from the Eichler-Shimura relation, spelled out in Lemma 7.16 and Lemma 10.4, respectively. By the proof of Lemma 9.5 the exact sequence above is not even split over F_∞, hence claim (*iv*) holds. The the intersection of $\mathcal{D}(\mathfrak{p})[l]$ and $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)[l]$ contains $\mathcal{S}(\mathfrak{p})[l]$ by Proposition 8.18. If it were larger, then the group scheme $\mathcal{D}(\mathfrak{p})[l]$ would be μ-type over F_∞ which it is not by the above, so claim (*v*) is true.

Now consider the case when l does not divide $N(\mathfrak{p})/l(\mathfrak{p})$. The cusps of $X_{\#l}(\mathfrak{p})$ mapping to the cusp 0 of the curve $X_0(\mathfrak{p})$ are actually defined over F, so the group $\mathcal{F}(\mathfrak{p})[l]$ is constant as a Galois-module. The Galois module $\mathcal{S}(\mathfrak{p})[l]$ is μ-type of order l, so it is constant, too. This proves the first claim. Lemma 8.19 and claim (*i*) implies that $\mathcal{D}$ is unramified at $\mathfrak{p}$. Note that that $\mathcal{D}(\mathfrak{p})[l]$ is a tamely ramified Galois module. It is the extension of the constant Galois module $\mathbb{F}_l$ by itself, so there is an $\mathbb{F}_l$-basis of this module where the Galois action is given by upper triangular matrices with ones on the diagonal. So the Galois action is given by a homomorphism from the absolute Galois group of F into $\mathbb{F}_l$. That is a tame abelian extension of F. As every tamely ramified Galois module which only ramifies at ∞ is in fact everywhere unramified, we get that claim (*ii*) holds.

As we already noted the group $\mathcal{S}(\mathfrak{p})[l]$ is both $\mathbb{T}(\mathfrak{p})$-invariant and Galois-invariant. Hence the quotient module $J_0(\mathfrak{p})(\mathbb{C}_\infty^*)[l]/\mathcal{S}(\mathfrak{p})[l]$ is equipped with a commuting action of $\mathbb{T}(\mathfrak{p})$ and the absolute Galois group which also satisfies the Eichler-Shimura relations. By repeating the arguments above we get that the Galois submodule $\mathcal{F}(\mathfrak{p})[l]$ of the quotient Galois module above is $\mathbb{T}(\mathfrak{p})$-invariant. Therefore its pre-image $\mathcal{D}(\mathfrak{p})[l]$ in $J_0(\mathfrak{p})(\mathbb{C}_\infty^*)[l]$ is also $\mathbb{T}(\mathfrak{p})$-invariant. Since $\mathcal{D}(\mathfrak{p})[l]$ is the extension of a constant Galois module by a μ-type Galois module, the identity $(\mathrm{Frob}_\mathfrak{q}-1)(\mathrm{Frob}_\mathfrak{q}-q^{\deg(\mathfrak{q})})$ holds on $\mathcal{D}(\mathfrak{p})[l]$ for every $\mathfrak{q}\neq\mathfrak{p}$ prime. By subtracting this identity from the Eichler-Shimura relations we get that $\mathrm{Frob}_\mathfrak{q}(T_\mathfrak{q}-1-q^{\deg(\mathfrak{q})})=0$. Since $\mathrm{Frob}_\mathfrak{q}$ is invertible we get that $\mathcal{D}(\mathfrak{p})[l]$ is annihilated by the Eisenstein ideal. This concludes the proof of claim (iii).

We will continue to use the notation introduced in the proof of Proposition 9.16. Take an element H of $\mathcal{Y}$ which represents a non-zero element of $\mathcal{F}(\mathfrak{p})[l]$. Then $(1-\sigma)H$ lies in $\mathcal{P}$ but it does not lie in $(1-\sigma)\mathcal{P}$. Since N is μ_l-invariant, it is trivial on $(1-\sigma)(\mathcal{P}/l\mathcal{P})$, therefore $(1-\sigma)H$ is the divisor of a non-zero rational function e such that $N(e)\notin(F^*)^l$. Assume that claim (iv) is false. Then there is an F-rational divisor E on $X_0(\mathfrak{p})$ whose pull-back E^* to $X_{\#l}(\mathfrak{p})$ is linearly equivalent to H, that is there is a a non-zero F-rational function g such that $H=E^*+(g)$. Since σ fixes this pull-back E^* there is a constant $u\in F^*$ such that $e(z)=ug(z)/g(z^\sigma)$. Hence $N(e)=u^l$ which is a contradiction. Because the Galois module $\mathcal{D}(\mathfrak{p})[l]$ is unramified, it does not split over F_∞ either. Hence claim (v) follows from claim (iv), as we already saw. □

Remark 9.19. The integers $\sum_{\alpha\in C_I(y)}q^{\deg(\alpha)}$ are analogues of the Bernoulli numbers. This is more or less clear from the computations of this chapter, but we will give an alternative argument here. We continue to let f denote a generator of the prime ideal $\mathfrak{p}$. Let $\mathbb{O}$ denote the ring of integers in the extension of $\mathbb{Q}_l$ we get by adjoining the l-th roots of unity. We define the $\mathbb{O}$-valued Dirichlet character χ by requiring that $\chi(f)=0$ and $\chi(g)=\overline{g}$ for every $g\in\mathbb{F}_q[T]$ relatively prime to f, where we consider μ_l as a subset of $\mathbb{O}$. We let U denote $\mathbb{P}^1_{\mathbb{F}_q}-\{\mathfrak{p}\}$ and let $\overline{U}$ denote its base change to $\overline{\mathbb{F}}_q$. By class field theory we have a corresponding Galois representation $\chi:\widehat{\pi}_1^{ab}(U)\to\mathbb{O}^*$ which is tamely ramified at $\mathfrak{p}$ and it is totally split at ∞. More precisely the Artin L-function of χ is

$$L(\chi,t)=\prod_{\mathfrak{p}\neq x\in|\mathbb{P}^1_{\mathbb{F}_q}|}(1-\chi(\mathrm{Fr}_x)t^{\deg(x)})^{-1}=\frac{1}{(1-t)(q-1)}\sum_{f\not| g\in\mathbb{F}_q[T]}\chi(g)t^{\deg(g)},$$

where $\mathrm{Fr}_x\in\widehat{\pi}_1^{ab}(U)$ is the arithmetic Frobenius at the place x. For every l-adic Galois representation ρ of $\widehat{\pi}_1(U,\infty)$ will use the same symbol to denote the lisse sheaf on U corresponding to ρ as well its base change to $\overline{U}$. By the Grothendieck-Verdier trace formula:

$$L(\chi,t)=\prod_{i=0}^{2}\det(1-Ft|H^1_c(\overline{U},\chi^{-1})^{(-1)^{i+1}},$$

where F is the Frobenius operator acting on the étale cohomology of χ^{-1}. Since χ as a representation of $\widehat{\pi}_1(\overline{U},\infty)$ is irreducible and non-trivial, the groups $H^0_c(\overline{U},\chi^{-1})$ and $H^2_c(\overline{U},\chi^{-1})$ are zero, and by the Ogg-Shafarevich formula the dimension of $H^1_c(\overline{U},\chi^{-1}))$ is $\deg(\mathfrak{p})-2$. Hence $L(\chi,t)$ is a polynomial of degree $\deg(\mathfrak{p})-2$ and

$$L(\chi,t)=\frac{1}{(1-t)(q-1)}\sum_{\substack{0\neq g\in\mathbb{F}_q[T]\\ \deg(g)<\deg(\mathfrak{p})}}\overline{g}\,t^{\deg(d)}=\sum_{y\in\mu_l}\frac{y}{1-t}\cdot\sum_{\alpha\in C_I(y)}t^{\deg(\alpha)}.$$

10. Mazur's Eisenstein decent at primes l not dividing $t(\mathfrak{p})$

Definition 10.1. For the rest of the paper, unless we say otherwise explicitly, we fix an Eisenstein prime l. Introduce the shorthand notation $\mathfrak{E}=\mathfrak{E}_l(\mathfrak{p})$ for the Eisenstein ideal in $\mathbb{T}_l(\mathfrak{p})$. Let $\mathfrak{P}\lhd\mathbb{T}_l(\mathfrak{p})$ be the unique prime ideal lying above $\mathfrak{E}$. As $\mathbb{Z}_l$ surjects onto $\mathbb{T}_l(\mathfrak{p})/\mathfrak{E}$ via its natural inclusion into $\mathbb{T}_l(\mathfrak{p})$, clearly $\mathfrak{P}=(\mathfrak{E},l)$. Hence the latter is a maximal ideal with residue field $\mathbb{F}_l$. Let $\eta_{\mathfrak{q}}$ denote the element $T_{\mathfrak{q}}-q^{\deg(\mathfrak{q})}-1\in\mathbb{T}(\mathfrak{p})$, where $\mathfrak{q}\lhd A$ is any prime ideal different from $\mathfrak{p}$. Let $\mathfrak{q}\lhd A$ be a prime ideal and let $r(T)\in A$ be the unique monic polynomial which generates $\mathfrak{q}$. We say that $\mathfrak{q}$ is a good prime if the following holds:

- (*i*) the prime ideal $\mathfrak{q}$ is not equal to $\mathfrak{p}$,
- (*ii*) the image of the reduction of the polynomial $r(T)$ modulo $\mathfrak{p}$ in the quotient $(A/\mathfrak{p})^*/\mathbb{F}_q^*$ is not an l-th power,
- (*iii*) if l does not divide $t(\mathfrak{p})$ then it also does not divide $q^{\deg(\mathfrak{q})}-1$,
- (*iv*) if l does divide $t(\mathfrak{p})$ then it does not divide $\deg(\mathfrak{q})$.

Note that every Eisenstein prime l divides $\frac{q^d-1}{q-1}$. This number is the order of the quotient group $(A/\mathfrak{p})^*/\mathbb{F}_q^*$, so the l-power map is not invertible on the latter. Hence the Chebotarev density theorem implies that there are infinitely many good primes.

For the rest of this chapter we assume that l does not divide $t(\mathfrak{p})$, unless we say otherwise explicitly. Now we can state the main result of this section:

Theorem 10.2. *The ideal $\mathfrak{P}$ is generated by l and $\eta_{\mathfrak{q}}$ for every good prime $\mathfrak{q}$.*

As explained in [14], Propositions 15.3 and 16.2, this theorem implies the following

Corollary 10.3. *The completion $\mathbb{T}_{\mathfrak{P}}$ of the Hecke algebra $\mathbb{T}_l(\mathfrak{p})$ at the prime ideal $\mathfrak{P}$ is Gorenstein.*

Before we start to prove Theorem 10.2, let us deduce its main Diophantine application from the corollary above. Let $\mathcal{E}(\mathfrak{p})$ denote the largest torsion subgroup of $J_0(\mathfrak{p})(\overline{F})$ annihilated by the Eisenstein ideal $\mathfrak{E}(\mathfrak{p})\lhd\mathbb{T}(\mathfrak{p})$. We will need the following preliminary result.

Lemma 10.4. *The group $\mathcal{E}(\mathfrak{p})$ contains $\mathcal{M}(\mathfrak{p})$.*

Proof. For the sake of simple notation let $J_0(\mathfrak{p})$ denote the Néron model of the Jacobian over X, too. The Cartier dual of a constant p-torsion group scheme is not étale in characteristic p, so the group scheme $\mathcal{M}(\mathfrak{p})$ has no p-torsion. Hence the reduction map injects $\mathcal{M}(\mathfrak{p})$ into $J_0(\mathfrak{p})(\overline{\mathbf{f}}_{\mathfrak{q}})$, for every prime $\mathfrak{q}$ different from $\mathfrak{p}$. The Frobenius endomorphism $\mathrm{Frob}_{\mathfrak{q}}$ of the abelian variety $J_0(\mathfrak{p})_{\mathbf{f}_{\mathfrak{q}}}$ acts as multiplication by $q^{\deg(\mathfrak{q})}$ on the reduction of $\mathcal{M}(\mathfrak{p})$. Therefore the Eichler-Shimura relation implies that the endomorphism $1 - T_{\mathfrak{q}} + q^{\deg(\mathfrak{q})}$ annihilates this group. □

Theorem 10.5. *The group schemes $\mathcal{M}(\mathfrak{p})_l$ and $\mathcal{S}(\mathfrak{p})_l$ are equal for any prime l not dividing $t(\mathfrak{p})$.*

Proof. Clearly the claim only needs demonstration when l is Eisenstein. The Frobenius Frob_∞ at ∞ acts non-trivially on the l-primary subgroup of $\mathcal{M}(\mathfrak{p})$, hence the latter must lie in the torsion of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$ annihilated by the ideal $\mathfrak{E}$ according to Lemma 10.4. The latter module is dual to $T_l^\vee/\mathfrak{E}T_l^\vee$, where the subscript $^\vee$ denotes the $\mathbb{T}_l(\mathfrak{p})$-dual. As $\mathbb{T}_{\mathfrak{P}}$ is Gorenstein, the completion of the locally free $\mathbb{T}_l(\mathfrak{p})$-module T_l at $\mathfrak{P}$ is isomorphic to its dual, so the module above is isomorphic to $\mathbb{T}_{\mathfrak{P}}/\mathfrak{E}\mathbb{T}_{\mathfrak{P}}$, because $\mathfrak{E}$ is supported on $\mathfrak{P}$. The latter has the same order as $\mathbb{Z}_l/N(\mathfrak{p})\mathbb{Z}_l$, hence it has the same order as the l-primary component of $\mathcal{S}(\mathfrak{p})$. □

We start our proof of Theorem 10.2 by proving a useful proposition about finite étale group schemes over the base $\mathbb{P}^1_{\mathbb{F}_q} - \{\mathfrak{p}\}$ which will function as a suitable analogue for the criteria for constancy and purity of [14] (Lemma 3.4 on page 57 and Proposition 4.5 on page 59, respectively).

Definition 10.6. In this paragraph, the next proposition and its proof l is any Eisenstein prime. We say that the group scheme G over the base S is μ-type if it is finite, flat and its Cartier dual is a constant group scheme over S. We say that the group scheme G is pure if it is the direct sum of a constant and a μ-type group scheme. Let $\mathbb{Z}/l^n\mathbb{Z}$ and μ_{l^n} denote the constant group scheme of order l^n and its Cartier dual, respectively. We say that a group scheme G is admissible if it is finite, étale and has a filtration by group schemes such that the successive quotients are pure. Clearly all these concepts make sense for the special case of finite Galois modules over fields.

Proposition 10.7. *Let G be an admissible group scheme of l-primary rank over the base $\mathbb{P}^1_{\mathbb{F}_q} - \{\mathfrak{p}\}$ and let $\mathfrak{q}$ be a good prime. Then the group scheme G is constant (resp. μ-type) if and only if it is constant (resp. μ-type) as a Galois module both over $F_{\mathfrak{q}}$ and over F_∞.*

Proof. For the sake of simple notation let U denote $\mathbb{P}^1_{\mathbb{F}_q} - \{\mathfrak{p}\}$ and let $\overline{U}$ denote its base change to $\overline{\mathbb{F}}_p$. First note that the criterion for constancy implies the other criterion by taking the Cartier-dual. In the former case clearly what we have to show is that the cardinality of the étale cohomology group $H^0_{et}(U, G)$

is the same as the rank of G. We are going to show the latter by induction on the rank of G. Since G is admissible, it contains a group scheme H isomorphic to either μ_l or $\mathbb{Z}/l\mathbb{Z}$. The group scheme H is constant as a Galois module over F_∞, hence it is isomorphic to $\mathbb{Z}/l\mathbb{Z}$. Therefore G is an extension:

$$0 \to \mathbb{Z}/l\mathbb{Z} \to G \to M \to 0.$$

The group scheme M is also admissible of l-primary rank which is constant as a Galois module both over $F_\mathfrak{q}$ and over F_∞. Hence by the induction hypothesis M is constant. Therefore it will be enough to show that the coboundary map $\delta : H^0_{et}(U, M) \to H^1_{et}(U, \mathbb{Z}/l\mathbb{Z})$ of the cohomological exact sequence of the short exact sequence above is trivial. Since G is constant as a Galois module both over $F_\mathfrak{q}$ and over F_∞, the coboundary maps $\delta : H^0(F_\mathfrak{q}, M) \to H^1(F_\mathfrak{q}, \mathbb{Z}/l\mathbb{Z})$ and $\delta : H^0(F_\infty, M) \to H^1(F_\infty, \mathbb{Z}/l\mathbb{Z})$ of the base change of the short exact sequence to the spectrum of $F_\mathfrak{q}$ and F_∞ respectively are trivial. (Of course the cohomology groups above are Galois cohomology groups.) Therefore we only have to show that the natural map $H^1_{et}(U, \mathbb{Z}/l\mathbb{Z}) \to H^1(F_\mathfrak{q}, \mathbb{Z}/l\mathbb{Z}) \oplus H^1(F_\infty, \mathbb{Z}/l\mathbb{Z})$ is injective. The cohomology group $H^1_{et}(U, \mathbb{Z}/l\mathbb{Z})$ is equal to the group cohomology $H^1(\widehat{\pi}_1^{ab}(U), \mathbb{Z}/l\mathbb{Z}) = \mathrm{Hom}(\widehat{\pi}_1^{ab}(U), \mathbb{Z}/l\mathbb{Z})$, where $\widehat{\pi}_1^{ab}(U)$ denotes the abelianization of the étale fundamental group of U. The map above is just the evaluation of the corresponding homomorphism $\widehat{\pi}_1^{ab}(U) \to \mathbb{Z}/l\mathbb{Z}$ on the Frobenius elements $\mathrm{Frob}_\mathfrak{q}$ and Frob_∞ in $\widehat{\pi}_1^{ab}(U)$. Hence we only have to prove that the image of $\mathrm{Frob}_\mathfrak{q}$ and Frob_∞ in $\widehat{\pi}_1^{ab}(U)/l\widehat{\pi}_1^{ab}(U)$ generate this group. By class field theory the latter is a consequence of (in fact it is equivalent to) the second condition in the definition of good primes. Let us give a quick proof of this fact. By class field theory the group $\widehat{\pi}_1^{ab}(U)$ is isomorphic to $F^*\backslash\mathbb{A}^*/U_\mathfrak{p}$, where $U_\mathfrak{p}$ is the direct product $\prod_{x\neq\mathfrak{p}} \mathcal{O}_x^*$. Under this identification the Frobenius elements $\mathrm{Frob}_\mathfrak{q}$ and Frob_∞ are represented by ideles $\pi_\mathfrak{q}$ and π_∞ whose divisor is $\mathfrak{q}$ and ∞, respectively, such that all components of π_v, where $v \neq \mathfrak{q}$ or ∞, which are different from $\mathfrak{q}$ or ∞, respectively, are actually equal to one. This identification also implies that there is an exact sequence

$$0 \to \mathcal{O}_\mathfrak{p}^*/(l\mathcal{O}_\mathfrak{p}^*)\mathbb{F}_q^* \to \widehat{\pi}_1^{ab}(U)/l\widehat{\pi}_1^{ab}(U) \to \mathbb{Z}/l\mathbb{Z} \to 0,$$

where the second map is the degree mod l of the divisor of any idele representing the class in $\widehat{\pi}_1^{ab}(U)/l\widehat{\pi}_1^{ab}(U)$. In particular $\widehat{\pi}_1^{ab}(U)/l\widehat{\pi}_1^{ab}(U)$ is two-dimensional as a vector space over $\mathbb{F}_l$, because l divides $\frac{q^d-1}{q-1}$. We also get that if the image of $\pi_\mathfrak{q}$ and π_∞ in $\widehat{\pi}_1^{ab}(U)/l\widehat{\pi}_1^{ab}(U)$ do not generate this group then the image of $\pi_\mathfrak{q}\pi_\infty^{-\deg(\mathfrak{q})}$ is trivial in $\widehat{\pi}_1^{ab}(U)/l\widehat{\pi}_1^{ab}(U)$. The latter can be reformulated by saying that $\pi_\mathfrak{q}\pi_\infty^{-\deg(\mathfrak{q})} = f(T)ug^l$, where $f(T) \in F^*$, $u \in U_\mathfrak{p}$ and $g \in \mathbb{A}^*$. It is clear from this equation that $f(T)$ is an l-th power in $F_\mathfrak{p}^*$. Because every degree zero divisor on $\mathbb{P}^1$ is principal, we get that $f(T) = cr(T)s(T)^l$ by comparing the divisors of the two sides of the equation above, where $c \in \mathbb{F}_q^*$ is a constant, $r(T) \in A$ is again the unique monic polynomial generating $\mathfrak{q}$ and $s(T) \in F^*$.

Looking at the $\mathfrak{p}$-adic components of the two sides of the equation above we get that $cr(T)$ is an l-th power modulo $\mathfrak{p}$. $\square$

For any smooth group scheme G let G_l denote its maximal l-primary subgroup scheme.

Proposition 10.8. *The group $\mathcal{E}(\mathfrak{p})_l$ is the direct sum of $\mathcal{C}(\mathfrak{p})_l$ and $\mathcal{M}(\mathfrak{p})_l$.*

Proof. We first prove that $\mathcal{E}(\mathfrak{p})_l$ is admissible. The latter has a filtration by the subgroups $\mathcal{E}(\mathfrak{p})[l^n]$, where $n \in \mathbb{Z}$. The quotient $\mathcal{E}(\mathfrak{p})[l^{n+1}]/\mathcal{E}(\mathfrak{p})[l^n]$ injects into $\mathcal{E}(\mathfrak{p})[l]$ via the map $x \mapsto l^n x$, hence it will be sufficient to prove that $\mathcal{E}(\mathfrak{p})[l]$ is admissible. Let $\mathcal{W}(\mathfrak{p})$ denote the direct sum of $\mathcal{E}(\mathfrak{p})[l]$ and its Cartier dual. It is a $\mathbb{T}_l(\mathfrak{p})$-module annihilated by $\mathfrak{P}$. It is also a Galois module over F which is unramified for every prime $\mathfrak{q} \neq \mathfrak{p}$ of A such that the action of the Galois group commutes with the action of the Hecke algebra. The fact that the action of the Hecke operator $T_\mathfrak{q}$ on $\mathcal{W}(\mathfrak{p})$ satisfies the Eichler-Shimura relations implies that the action of the Frobenius $\mathrm{Frob}_\mathfrak{q}$ for any prime $\mathfrak{q} \neq \mathfrak{p}$ of A satisfies the relation

$$(\mathrm{Frob}_\mathfrak{q} - 1)(\mathrm{Frob}_\mathfrak{q} - q^{\deg(\mathfrak{q})}) = 0.$$

Hence the only eigenvalues possible for the action of $\mathrm{Frob}_\mathfrak{q}$ on $\mathcal{W}(\mathfrak{p})$ are 1 and $q^{\deg(\mathfrak{p})}$. Since the latter is Cartier self-dual, the multiplicities of these eigenvalues must be the same, hence the characteristic polynomial of $\mathrm{Frob}_\mathfrak{q}$ acting on $\mathcal{W}(\mathfrak{p})$ must be $(x-1)^m(x-q^{\deg(\mathfrak{q})})^m$, where $2m$ is the dimension of $\mathcal{W}(\mathfrak{p})$ as a vector space over $\mathbb{F}_l$. By the Chebotarev theorem we get that the characteristic polynomial of any element in the absolute Galois group of F acting on $\mathcal{W}(\mathfrak{p})$ is the same as the characteristic polynomial of its action on the Galois module $(\mathbb{Z}/l\mathbb{Z})^m \oplus (\mu_l)^m$. The Brauer-Nesbitt theorem implies that the semi-simplification of these modules must be equal, so $\mathcal{W}(\mathfrak{p})$, and therefore $\mathcal{E}(\mathfrak{p})_l$, are admissible.

As l does not divide $q-1$, the intersection of $\mathcal{C}(\mathfrak{p})_l$ and $\mathcal{M}(\mathfrak{p})_l$ is trivial. Now we only have to show that their direct sum is the whole l-primary subgroup of $\mathcal{E}(\mathfrak{p})$. Since $\mathcal{C}(\mathfrak{p})$ is fixed by the absolute Galois group of F, the quotient $\mathcal{H}(\mathfrak{p}) = \mathcal{E}(\mathfrak{p})/\mathcal{C}(\mathfrak{p})$ is a Galois module. This module is unramified at all places different from ∞ and $\mathfrak{p}$, because $\mathcal{E}(\mathfrak{p})$ is. The proof of Proposition 7.18 shows that the quotient of $\mathcal{E}(\mathfrak{p})_l$ by the torsion of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$ injects into $\mathbb{Z}_l/N(\mathfrak{p})\mathbb{Z}_l$. The restriction of this map onto $\mathcal{C}(\mathfrak{p})_l$ is surjective, as we already saw in the proof of Theorem 7.19. Hence $\mathcal{H}(\mathfrak{p})_l$ as a Galois module over F_∞ is isomorphic to a submodule of the torsion of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$, in particular it is also unramified at ∞. As $\mathcal{E}(\mathfrak{p})_l$ is admissible as a Galois module over F, so does $\mathcal{H}(\mathfrak{p})_l$. Therefore the unique finite étale group scheme over $\mathbb{P}^1_{\mathbb{F}_q} - \{\mathfrak{p}\}$ prolonging $\mathcal{H}(\mathfrak{p})_l$ is also admissible. Moreover this admissible group scheme is μ-type as a Galois module over F_∞, because the l-primary torsion of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$ is. We also get that all Jordan-Hölder components of this admissible group scheme must be isomorphic to μ_l. Let $\mathfrak{q}$ be any admissible prime of A. The operator $\eta_\mathfrak{q}$ annihilates $\mathcal{E}(\mathfrak{p})$, so does the endomorphism $(\mathrm{Frob}_\mathfrak{q}-1)(\mathrm{Frob}_\mathfrak{q}-q^{\deg(\mathfrak{q})})$ by the Eichler-Shimura relations. By the above $\mathrm{Frob}_{\mathfrak{q}} - 1$ must be invertible on $\mathcal{H}(\mathfrak{p})_l$, so we get that $\mathrm{Frob}_{\mathfrak{q}} - q^{\deg(\mathfrak{q})}$ must annihilate this Galois module. Hence the module $\mathcal{H}(\mathfrak{p})_l$ must be μ-type by Proposition 10.7. In particular it is fixed under the action of the inertia group I at $\mathfrak{p}$. By Lemma 8.19 we get that the whole module $\mathcal{E}(\mathfrak{p})_l$ is fixed by I, too. As $\mathcal{E}(\mathfrak{p})_l$ is the direct sum of $\mathcal{C}(\mathfrak{p})_l$ and $\mathcal{E}(\mathfrak{p})_l \cap \mathrm{Hom}(\overline{\Gamma}(\mathfrak{p}), \mathbb{C}_\infty^*)_l$, where the latter is a Galois sub-module over F_∞ isomorphic to $\mathcal{H}(\mathfrak{p})_l$, the module $\mathcal{E}(\mathfrak{p})_l$ is unramified at ∞, too, so it is in fact everywhere unramified. Since it is pure as a Galois module over F_∞, it is pure as a Galois module over F, and the claim is now obvious. □

Fix a good prime $\mathfrak{q}$. For any natural number r let $\mathcal{G}_r$ denote the largest subgroup-scheme of $J_0(\mathfrak{p})_l$ annihilated by the ideal $(\mathfrak{E}\mathfrak{P}^r, \eta_{\mathfrak{q}})$.

Proposition 10.9. *The group scheme $\mathcal{G}_r$ is the direct sum of the group $\mathcal{C}(\mathfrak{p})_l$ and a μ-type group $\mathcal{M}_r$.*

Proof. We are going to prove the claim by induction on r. As $\mathcal{G}_0 = \mathcal{E}(\mathfrak{p})_l$, this case has already been proved. Now we assume that the claim has been proved for $\mathcal{G}_r$, and we are going to show it for $\mathcal{G}_{r+1}$. Let $a_1, a_2, \ldots, a_m$ be a set of elements of $\mathfrak{P}^r$ such that their class mod $\mathfrak{P}^{r+1}$ is a basis of the $\mathbb{F}_l$ vector space $\mathfrak{P}^r/\mathfrak{P}^{r+1}$. The map $x \mapsto a_1 x \oplus \cdots \oplus a_m x$ defines a homomorphism $\mathcal{G}_{r+1} \to \mathcal{E}(\mathfrak{p})_l^m$ with kernel $\mathcal{G}_r$, hence the quotient Galois module $\mathcal{G}_{r+1}/\mathcal{G}_r$ is pure as a submodule of a pure Galois module. Let $\mathcal{G}_{r+1}/\mathcal{G}_r = \mathcal{A}_r \oplus \mathcal{N}_r$, where $\mathcal{A}_r$, $\mathcal{N}_r$ are constant and μ-type Galois modules, respectively.
Let $\mathcal{G}$ be the pre-image of $\mathcal{A}_r$ in $\mathcal{G}_{r+1}$ and let $\overline{\mathcal{G}}$ be the quotient $\mathcal{G}/\mathcal{M}_r$ (recall that $\mathcal{M}_r$ is the μ-type component of $\mathcal{G}_r$). Clearly $\overline{\mathcal{G}}$ is a Galois module over F which is admissible, because it is the extension of the constant module $\mathcal{A}_r$ by the constant module $\mathcal{C}(p)_l$. The natural action of the Hecke algebra on the quotient Galois module $\mathcal{G}_{r+1}/\mathcal{G}_r$ commutes with the action of the Galois group, so it must preserve the eigenspace $\mathcal{A}_r$ of the latter. Therefore it leaves the Galois module $\mathcal{G}$ invariant, moreover it acts on its quotient $\overline{\mathcal{G}}$, because it leaves the module $\mathcal{M}_r$ invariant. The module $\overline{\mathcal{G}}$ injects into the quotient of $J_0(\mathfrak{p})_l$ by the l-primary torsion of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$. Therefore it is constant as a Galois module over F_∞. The operator $\eta_{\mathfrak{q}}$ annihilates $\overline{\mathcal{G}}$, so does the endomorphism $(\mathrm{Frob}_{\mathfrak{q}} - 1)(\mathrm{Frob}_{\mathfrak{q}} - q^{\deg(\mathfrak{q})})$ by the Eichler-Shimura relations. By the above $\mathrm{Frob}_{\mathfrak{q}} - q^{\deg(\mathfrak{q})}$ must be invertible on $\overline{\mathcal{G}}$, so we get that $\mathrm{Frob}_{\mathfrak{q}} - 1$ must annihilate this Galois module as well. Now we may apply Proposition 10.7 to conclude that $\overline{\mathcal{G}}$ is actually constant as a Galois module over F. As we already saw in the proof of Lemma 7.16, this fact and the Eichler-Shimura relations imply that $\overline{\mathcal{G}}$ is annihilated by the Eisenstein ideal. Hence $\overline{\mathcal{G}} = \mathcal{C}(p)_l$ according to the proof of Proposition 7.18.
We get that $\mathcal{A}_r = 0$, so $\mathcal{G}_{r+1}$ is the extension of $C(\mathfrak{p})_l$ by a group scheme which the extension of the μ-type group scheme $\mathcal{N}_r$ by the μ-type group scheme $\mathcal{M}_r$. In particular the latter is admissible, and it must lie in the l-primary torsion of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}_\infty^*)$. Therefore the argument presented above shows that this module is μ-type over F, and the claim is now proved. □

Proof of Theorem 10.2. Let $x \in J_0(\mathfrak{p})_l$ be any element annihilated by $\eta_{\mathfrak{q}}$. Then x is actually annihilated by the ideal $(\eta_{\mathfrak{q}}, l^n)$ for some n. The latter contains $\mathfrak{E}\mathfrak{P}^r$ for some r, hence x is an element of $\mathcal{G}_r$ in this case. Since $\eta_{\mathfrak{q}}$ annihilates $\mathcal{E}(\mathfrak{p})_l$, we get that the group of elements $x \in J_0(\mathfrak{p})_l$ annihilated by $\eta_{\mathfrak{q}}$ is $\mathcal{M}(\mathfrak{p})_l \oplus \mathcal{C}(\mathfrak{p})_l$, using Propositions 10.8 and 10.9. Also note that $\eta_{\mathfrak{q}}$ is actually an isogeny of $J_0(\mathfrak{p})$. If it were not, then $J_0(\mathfrak{p})$ would contain an abelian subvariety such that the action of the Frobenius at $\mathfrak{q}$ on this variety would have 1 or $q^{\deg(\mathfrak{q})}$ as an eigenvalue by the Eichler-Shimura relations. The latter is impossible by Weil's theorem. Therefore $\eta_{\mathfrak{q}}$ is injective as an endomorphism of T_l. By dualizing we get that it is surjective as an endomorphism of $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$. Let $y \in \mathcal{H}_{00}(\mathfrak{p}, \mathbb{F}_l)$ be any element annihilated by $\eta_{\mathfrak{q}}$. Pick an element $x \in J_0(\mathfrak{p})_l$ whose specialization (i.e. its class in the quotient of $J_0(\mathfrak{p})_l$ by the l-primary torsion of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)$) is y. Then $\eta_{\mathfrak{q}}(x) \in \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$. By the above there is a $z \in \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$ such that $\eta_{\mathfrak{q}}(z) = \eta_{\mathfrak{q}}(x)$. Then the element $x - z$ is annihilated by $\eta_{\mathfrak{q}}$ and its specialization is y. We get that the specialization map from $C(\mathfrak{p})[l]$ into the submodule of $\mathcal{H}_{00}(\mathfrak{p}, \mathbb{F}_l)$ annihilated by $\eta_{\mathfrak{q}}$ is an isomorphism, in particular the latter is 1-dimensional as a vector space over $\mathbb{F}_l$. The latter is also dual to $T_l/(\eta_{\mathfrak{q}}, l)$. Since T_l is locally free of rank one as a $\mathbb{T}_l(\mathfrak{p})$-module, we get that $(\eta_{\mathfrak{q}}, l)$ is a prime ideal, hence the claim holds. □

11. Mazur's Eisenstein decent for primes l dividing $t(\mathfrak{p})$

Definition 11.1. For the rest of this chapter we fix a prime l dividing $t(\mathfrak{p})$. Then l is automatically an Eisenstein prime. We also introduce the shorthand notation $\mathcal{S} = \mathcal{S}(\mathfrak{p})[l]$, $\mathcal{F} = \mathcal{F}(\mathfrak{p})[l]$ and $\mathcal{D} = \mathcal{D}(\mathfrak{p})[l]$. A Galois sub-module $G \subset J_0(\mathfrak{p})_l$ is $*$-type, if

- (i) it contains $\mathcal{D}$,
- (ii) the intersection $G_0 = G \cap \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$ is Galois-invariant,
- (iii) the Galois module G_0 is admissible,
- (iv) the quotient G/G_0 is equal to $\mathcal{F}$.

In this case let $G_{00} \subseteq G_0$ denote pre-image of the largest μ-type subgroup of $G_0/\mathcal{S}$ under the quotient map. Note that under this definition $\mathcal{D}$ itself is a $*$-type group by Proposition 9.18.

Lemma 11.2. *Let $G \subset J_0(\mathfrak{p})_l$ be a $*$-type Galois module. Then G_{00} is μ-type.*

Proof. By Lemma 8.19 the Galois module G_{00} is unramified at $\mathfrak{p}$. Since every tame Galois module which only ramifies at ∞ is in fact everywhere unramified, we get that G_{00} is everywhere unramified. It is μ-type as a Galois module over F_∞, being a sub-module of G_0, hence it is μ-type as a Galois module over F, too. □

The following proposition corresponds to Lemma 17.5 of [14], pages 131-133.

Proposition 11.3. *Let $\mathfrak{q}$ be a good prime and let $G \subset H \subset J_0(\mathfrak{p})_l$ be two $\mathbb{T}(\mathfrak{p})$-invariant Galois modules annihilated by $\eta_{\mathfrak{q}}$ and assume that*

- (i) *the Galois module G is $*$-type,* (*ii*) *the quotient H/G has order l,*

(*iii*) *the quotient $H/H \cap \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$ is l-torsion.*

Then H is $*$-type, too.

Proof. Let H_0 denote the intersection $H \cap \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$. Since the quotient groups H/G and G/G_0 both have order l, the quotient H/G_0 must have order l^2. Hence it is either isomorphic to $\mathbb{Z}/l^2\mathbb{Z}$ or to $\mathbb{F}_l^2$ as a group. In the first case H_0 must be equal to G, since H_0/G_0 must be a proper subgroup of H/G_0 by condition (*iii*), but $\mathbb{Z}/l^2\mathbb{Z}$ has only one proper subgroup. Since G does not lie in $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$, this is a contradiction. Hence H/G_0 is l-torsion.
Because H is annihilated by the operator $\eta_\mathfrak{q}$, the Eichler-Shimura relation has the shape $(\mathrm{Frob}_\mathfrak{q} - 1)(\mathrm{Frob}_\mathfrak{q} - q^{\deg(\mathfrak{q})}) = 0$ in H for the prime $\mathfrak{q}$. The Galois module H/G is also equipped with an action of the Hecke algebra $\mathbb{T}(\mathfrak{p})$ which satisfies the Eichler-Shimura relations. Since $q \equiv 1 \mod l$ by assumption, we get that $(\mathrm{Frob}_\mathfrak{q} - 1)^2 = 0$ on H/G. Since the latter is a one-dimensional vector space over $\mathbb{F}_l$, we get that $\mathrm{Frob}_\mathfrak{q} - 1$ annihilates H/G, in other words $(\mathrm{Frob}_\mathfrak{q} - 1)(H)$ lies in G. Using the Eichler-Shimura relation for the prime $\mathfrak{q}$ in H again we get that the image of $\mathrm{Frob}_\mathfrak{q} - 1$ actually lies in the kernel M of $\mathrm{Frob}_\mathfrak{q} - q^{\deg(\mathfrak{q})}$ in G.
Note that $G/\mathcal{D}$ is μ-type as a Galois module over F_∞. Hence the image of M in this group under the quotient map is μ-type by Proposition 10.7. As the natural map $G_0/\mathcal{S} \to G/\mathcal{D}$ is an isomorphism, the image of M in $G/\mathcal{D}$ must lie in the image of G_{00} by the above. Hence M lies in the group generated by G_{00} and $\mathcal{D}$. Assume that M does not lie in G_{00}. By Lemma 11.2 the module G_{00} is μ-type, hence it is annihilated by $\mathrm{Frob}_\mathfrak{q} - q^{\deg(\mathfrak{q})}$, or in other words it is in M. This implies that M must contain $\mathcal{D}$, too. The latter is everywhere unramified, but does not split by (*ii*) and (*iv*) of Proposition 9.18. Therefore the action of $\mathrm{Frob}_\mathfrak{q}$ could not be trivial as $\mathrm{Frob}_\mathfrak{q}$ generates the maximal everywhere unramified l-torsion abelian Galois extension of F because of the condition that l does not divide $\deg(\mathfrak{q})$. This is a contradiction, so M lies in $G_{00} \subseteq G_0$. Hence we get that $\mathrm{Frob}_\mathfrak{q} - 1$ annihilates H/G_0.
Now assume that $H_0 = G_0$. In this case H/G_0 injects naturally into the quotient $J_0(\mathfrak{p})_l/\mathrm{Hom}(\overline{\Gamma}(\mathfrak{p}), \mathbb{C}^*_\infty)_l$. Hence it is trivial as a Galois module over F_∞, so it is even trivial as a Galois module over F by Proposition 10.7. The Galois module G_0 is also $\mathbb{T}(\mathfrak{p})$-invariant, so there is an induced action of the Hecke algebra $\mathbb{T}(\mathfrak{p})$ on H/G_0. The latter satisfies the Eichler-Shimura relations, so the Eisenstein ideal annihilates H/G_0 applying again the argument in the proof of Lemma 7.16. Since the inclusion of H/G_0 in $J_0(\mathfrak{p})_l/\mathrm{Hom}(\overline{\Gamma}(\mathfrak{p}), \mathbb{C}^*_\infty)_l$ is $\mathbb{T}(\mathfrak{p})$-equivariant, we get that the former must be one-dimensional as a vector space over $\mathbb{F}_l$ by the strong multiplicity one theorem.
This is a contradiction, so H_0 is strictly larger than G_0. As we already see in the first paragraph, the group H_0/G_0 can not be equal to G/G_0, so H/G_0 has two proper subgroups invariant under the action of the absolute Galois group of F_∞. Hence H/G_0 must be trivial as a Galois module over F_∞. By repeating the argument above we get that H/G_0 is trivial as a Galois module over F. In

particular H_0/G_0 is Galois-invariant, hence H_0 is a Galois-invariant subgroup of H. Since it is the extension of $\mathbb{Z}/l\mathbb{Z}$ by the admissible Galois module G_0, it must be admissible, too. The quotient H/H_0 has order l, so it must be equal to $\mathcal{F}$. Since condition (i) of Definition 11.1 is automatic for H, the claim is now proved. $\square$

Definition 11.4. Fix a good prime $\mathfrak{q}$. Let $\mathfrak{P} = (\mathfrak{E}, l)$ be the Eisenstein prime ideal above l. For any natural number r let $\mathcal{H}(r)$ denote the largest subgroup of $J_0(\mathfrak{p})_l$ annihilated by the ideal $(l^r, \eta_{\mathfrak{q}})$. Let $\mathcal{G}(0)$ be $\mathcal{D}$, and for every positive integer r let $\mathcal{G}(r)$ be the pre-image of the largest submodule of $\mathcal{H}_{00}(\mathfrak{p}, \mathbb{F}_l)$ annihilated by $\eta_{\mathfrak{q}}$ in $\mathcal{H}(r)$ under the specialization map and let $\mathcal{G}_0(r)$ denote the intersection $\mathcal{G}(r) \cap \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$. Both groups are invariant under the action of $\mathbb{T}_l$ and the absolute Galois group of F_∞. What is not clear that these groups are Galois modules over F.

The following proposition corresponds to Lemma 17.7 of [14], pages 133-134.

Proposition 11.5. *The group $\mathcal{G}(r)$ is Galois-invariant, and as a Galois module it is $*$-type.*

Proof. We are going to prove the claim by induction on r. As $\mathcal{G}(0) = \mathcal{D}$, the claim is clear for $r = 0$. Now we assume that the claim is true for r and then we are going to prove it for $r+1$. If $x \in \mathcal{H}(r+1)$, then by the defining property of $\mathcal{G}(r+1)$ we have $x \in \mathcal{G}(r+1)$ if and only if $lx \in \mathcal{G}_0(r)$. (This is true even when $r = 0$ as $\mathcal{G}(1) = \mathcal{H}(1)$ is l-torsion.) We first need to show that $\sigma(x) \in \mathcal{G}(r+1)$ for any $\sigma \in \mathrm{Gal}(\overline{F}|F)$. Equivalently we have to show that $l\sigma(x) \in \mathcal{G}_0(r)$, but this is true because σ leaves $\mathcal{G}_0(r)$ stable by the induction hypothesis and $\sigma(lx) = l\sigma(x)$.
The Galois module $\mathcal{G}(r+1)$ is admissible because it is a Galois sub-module of $\mathcal{H}(r+1)$, which is admissible. The latter can be seen by noting that $\mathcal{H}(r+1)$ has a filtration by $\mathbb{T}_l$-invariant Galois submodules whose components are annihilated by the ideal $(l, \eta_{\mathfrak{q}})$, hence by some power of the Eisenstein ideal. Therefore the arguments at the start of the proof of Propositions 10.8 and 10.9 can be applied to these components to show that they are admissible.
The Galois modules $\mathcal{G}(r)$ and $\mathcal{G}(r+1)$ are both $\mathbb{T}_l$-invariant, so there is a filtration:

$$\mathcal{G}(r) = F_0 \subset F_1 \subset \ldots \subset F_j \subset \ldots \subset F_m = \mathcal{G}(r+1)$$

by $\mathbb{T}_{\mathfrak{P}}[\mathrm{Gal}(\overline{F}|F)]$-modules such that the successive quotients are irreducible modules over the group algebra $\mathbb{T}_{\mathfrak{P}}[\mathrm{Gal}(\overline{F}|F)]$, where $\mathbb{T}_{\mathfrak{P}}$ is the completion of the Hecke algebra $\mathbb{T}_l(\mathfrak{p})$ at the prime ideal $\mathfrak{P}$. These modules must be annihilated by $\mathfrak{P}$, because they are irreducible. But $\mathbb{T}_{\mathfrak{P}}/\mathfrak{P} = \mathbb{Z}/l\mathbb{Z}$, so these components are actually irreducible $\mathrm{Gal}(\overline{F}|F)$-modules. Since they are admissible, too, their order is l. Therefore it follows that F_j is $*$-type using Proposition 11.3 by induction on j: the modules F_j are $\mathbb{T}_l$-invariant by their construction, condition (i) is the induction hypothesis, condition (ii) has just been proved, and condition (iii) holds because $\mathcal{G}(r+1)/\mathcal{G}_0(r+1)$ is l-torsion by definition. $\square$

Theorem 11.6. *The ideal $\mathfrak{P}$ is generated by l and $\eta_{\mathfrak{q}}$ for every good prime $\mathfrak{q}$. In particular $\mathbb{T}_{\mathfrak{P}}$ is Gorenstein.*

Proof. Let $y \in \mathcal{H}_{00}(\mathfrak{p}, \mathbb{F}_l)$ be any element annihilated by $\eta_{\mathfrak{q}}$. Pick an element $x \in J_0(\mathfrak{p})_l$ whose specialization (i.e. its class in the quotient of $J_0(\mathfrak{p})_l$ by the l-primary torsion of the torus $\mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)$) is y. Then $\eta_{\mathfrak{q}}(x) \in \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$. Since $\eta_{\mathfrak{q}}$ is an isogeny of $J_0(\mathfrak{p})$, there is a $z \in \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$ such that $\eta_{\mathfrak{q}}(z) = \eta_{\mathfrak{q}}(x)$. Then the element $u = x - z$ is annihilated by $\eta_{\mathfrak{q}}$ and its specialization is y. As u must be an element of $\mathcal{G}(r)$ for some natural number r, we get that the submodule of $\mathcal{H}_{00}(\mathfrak{p}, \mathbb{F}_l)$ annihilated by $\eta_{\mathfrak{q}}$ is 1-dimensional as a vector space over $\mathbb{F}_l$ by Proposition 11.5. The latter is also dual to $T_l/(\eta_{\mathfrak{q}}, l)$. Since T_l is locally free of rank one as a $\mathbb{T}_l(\mathfrak{p})$-module, we get that $(\eta_{\mathfrak{q}}, l)$ is a prime ideal, hence the claim holds. □

Corollary 11.7. *The groups $\mathcal{E}(\mathfrak{p})[l]$ and $\mathcal{D}$ are equal.*

Proof. As we already noted, $\mathcal{E}(\mathfrak{p})[l]$ contains $\mathcal{D}$. By the strong multiplicity one theorem the image of the specialization of $\mathcal{E}(\mathfrak{p})[l]$ is equal to the image of the specialization of $\mathcal{D}$ (see the proof of Proposition 7.18). Because $\mathbb{T}_{\mathfrak{P}}$ is Gorenstein by Theorem 11.6, the intersection $\mathcal{E}(\mathfrak{p}) \cap \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)[l]$ is a free $\mathbb{T}_{\mathfrak{P}}/\mathfrak{E}\mathbb{T}_{\mathfrak{P}}$ module of rank one. Hence it has the same order as $\mathcal{S}$, so they are equal, too. Since this module is the kernel of the specialization map, the claim is now obvious. □

Corollary 11.8. *The l-primary subgroups of $\mathcal{M}(\mathfrak{p})$ (resp. $\mathcal{T}(\mathfrak{p})$) and $\mathcal{S}(\mathfrak{p})$ (resp. $\mathcal{C}(\mathfrak{p})$) are equal.*

Proof. First note that the intersection $\mathcal{E}(\mathfrak{p})_l \cap \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$ is $\mathcal{S}(\mathfrak{p})_l$. This can be seen very easily by repeating the proof of Theorem 10.5 if either d is odd or $l \neq 2$. This condition is necessary to guarantee that the order of $\mathcal{E}(\mathfrak{p})_l \cap \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_l$ is the same as the order of $\mathcal{S}(\mathfrak{p})_l$ while using claim (vi) of Proposition 7.11, which rests on Theorem 6.6. If $d = \deg(\mathfrak{p})$ is even and $l = 2$ then the same argument (and the claim quoted above) only shows that $\mathcal{S}(\mathfrak{p})_2$ is a subgroup of index at most two in the group $\mathcal{E}_2 = \mathcal{E}(\mathfrak{p}) \cap \mathrm{Hom}(\overline{\Gamma}_0(\mathfrak{p}), \mathbb{C}^*_\infty)_2$ as the order of the latter is the same as the index of the Eisenstein ideal $\mathfrak{E}_2(\mathfrak{p}) \triangleleft \mathbb{T}_2(\mathfrak{p})$. Note that $\mathcal{E}_2$ is the intersection of $\mathcal{E}(\mathfrak{p})$ and the union $\cup_{r \in \mathbb{N}} \mathcal{G}_0(r)$, so it is a Galois module. The quotient group $\mathcal{E}_2/\mathcal{S}(\mathfrak{p})$ is admissible of order at most two, so it must be μ-type. Hence Lemma 8.19 can be applied to show that $\mathcal{E}_2$ is unramified at $\mathfrak{p}$. By the Néron property this group has a specialization map into the group of components of $J_0(\mathfrak{p})$ at $\mathfrak{p}$. The restriction of this map to $\mathcal{S}(\mathfrak{p})$ is injective by (ii) of Proposition 8.18, so it is injective as $\mathcal{E}_2$ is a cyclic group by the strong multiplicity one theorem. The order of the maximal 2-primary subgroup of the group of components of $J_0(\mathfrak{p})$ at $\mathfrak{p}$ is the same as the order of $\mathcal{S}(\mathfrak{p})_2$, so the latter is the whole group $\mathcal{E}_2$. By reversing the logic of the argument at the start of this paragraph we get that $\mathbb{T}_l/\mathfrak{E}_l(\mathfrak{p}) = \mathbb{Z}_l/N(\mathfrak{p})\mathbb{Z}_l$ even when d is even and $l = 2$.

Let $l(\mathfrak{p})$ denote the largest power of l dividing $N(\mathfrak{p})$. If the claim above is false then there is an element x in $\mathcal{M}(\mathfrak{p})_l - \mathcal{S}(\mathfrak{p})_l$ (resp. in $\mathcal{T}(\mathfrak{p})_l - \mathcal{C}(\mathfrak{p})_l$) such that lx is in $\mathcal{S}(\mathfrak{p})_l$ (resp. in $\mathcal{C}(\mathfrak{p})_l$). The element x is annihilated by $l(\mathfrak{p})$, since it is annihilated by the Eisenstein ideal. Therefore lx is annihilated by $\frac{l(\mathfrak{p})}{l}$. Since both $\mathcal{S}(\mathfrak{p})_l$ and $\mathcal{C}(\mathfrak{p})_l$ are cyclic of order $l(\mathfrak{p})$, the element lx must have an l-root u in $\mathcal{S}(\mathfrak{p})_l$ (resp. in $\mathcal{C}(\mathfrak{p})_l$) by the above. Subtracting u from x we get that we may assume that x is l-torsion. By Corollary 11.7 we must have $x \in \mathcal{D}$. Since the Galois module $\mathcal{D}$ is not pure, we conclude that x is actually in $\mathcal{S}$. The intersection of $\mathcal{S}(\mathfrak{p})_l$ and $\mathcal{C}(\mathfrak{p})_l$ is exactly the largest constant Galois submodule of the former by Proposition 9.3, so the claim is now clear. $\square$

Remark 11.9. An interesting corollary of the proof above that the inclusion $\mathcal{H}_{00}(\mathfrak{p},\mathbb{Z}_2/2N(\mathfrak{p})\mathbb{Z}_2) \rightarrow \mathcal{H}_0(\mathfrak{p},\mathbb{Z}_2/2N(\mathfrak{p})\mathbb{Z}_2)$ is not surjective if $d = \deg(\mathfrak{p})$ is even, i.e. there is a cuspidal harmonic form with values in $\mathbb{Z}_2/2N(\mathfrak{p})\mathbb{Z}_2$ which cannot be lifted to an integer-valued cuspidal harmonic form. Our proof of this fact is quite involved and geometric, and wanders out of the natural algebraic universe where this question lives. It would be nice to see a more conceptual and general proof.

References

1. M. Artin, *Néron models*, Arithmetic geometry (G. Cornell, J. H. Silverman, eds.), Springer, Berlin-Heidelberg-New York, 1986, pp. 213–230.
2. P. Deligne and M. Rapoport, *Schémas de modules des courbes elliptiques*, L.N.M. 349, Springer, Berlin-Heidelberg-New York, 1973, pp. 143-316.
3. V. G. Drinfeld, *Elliptic modules*, [translation], Mat. Sbornik **23** (1974), 561–592.
4. E.-U. Gekeler, *Zur Arithmetic von Drinfeld-Monduln*, Math. Ann. **262** (1983), 167-182.
5. E.-U. Gekeler, *Drinfeld modular curves*, L.N.M. 1231, Springer, Berlin-Heidelberg-New York, 1986.
6. E.-U. Gekeler, *Über Drinfeld'sche Modulkurven von Hecke-Typ*, Comp. Math. **57** (1986), 219-236.
7. E.-U. Gekeler, *On the cuspidal divisor class group of a Drinfeld modular curve*, Doc. Math. **2** (1997), 351-374.
8. E.-U. Gekeler, *On the Drinfeld discriminant function*, Comp. Math. **106** (1997), 181-202.
9. E.-U. Gekeler, *Cuspidal divisor class groups of modular curves*, Algebraic number theory and Diophantine analysis (Graz, 1998), de Gruyter, Berlin, 2000, pp. 163–189.
10. E.-U. Gekeler and U. Nonnengardt, *Fundamental domains of some arithmetic groups over function fields*, Int. J. Math. **6** (1995), 689–708.
11. E.-U. Gekeler and M. Reversat, *Jacobians of Drinfeld modular curves*, J. reine angew. Math. **476** (1996), 27–93.
12. H. Jacquet and R.P. Langlands, *Automorphic forms on $GL(2)$*, Springer, Berlin-Heidelberg-New York, 1970.
13. N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Princeton University Press, Princeton, 1985.

14. B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. **47** (1977), 33–186.
15. J. S. Milne, *Étale cohomology*, Princeton University Press, Princeton, 1980.
16. A. Pál, *The rigid analytical regulator and K_2 of Drinfeld modular curves*, submitted.
17. J.-P. Serre, *Trees*, Springer, Berlin-Heidelberg-New York, 1980.
18. A. Tamagawa, *The Eisenstein quotient of the Jacobian variety of a Drinfeld modular curve*, Publ. Res. Inst. Math. Sci. **31** (1995), 203–246.
19. A. Weil, *Dirichlet series and automorphic forms*, Springer, Berlin-Heidelberg-New York, 1971.

Index of notations

Ambrus Pál
IHÉS
Le Bois-Marie
35, Route de Chartres
F-91440
Bures-sur-Yvette, France
pal@ihes.fr

Cohomology of Arithmetic Groups with Infinite Dimensional Coefficient Spaces

Anton Deitmar and Joachim Hilgert

Received: December 7, 2004

Communicated by Patrick Delorme

Abstract. The cuspidal cohomology groups of arithmetic groups in certain infinite dimensional modules are computed. As a result we get a simultaneous generalization of the Patterson-Conjecture and the Lewis-Correspondence.

2000 Mathematics Subject Classification: 11F75

Introduction

Let G be a semisimple Lie group and $\Gamma \subset G$ an arithmetic subgroup. For a finite dimensional representation (ρ, E) of G the cohomology groups $H^\bullet(\Gamma, E)$ are related to automorphic forms and have for this reason been studied by many authors. The case of infinite dimensional representations has only very recently come into focus, mostly in connection with the Patterson Conjecture on the divisor of the Selberg zeta function [7, 8, 9, 11, 17]. In this paper we want to show that the Patterson conjecture [7] is related to the Lewis correspondence [21], i.e., that the multiplicities of automorphic representations can be expressed in terms of cohomology groups with certain infinite dimensional coefficient spaces.

One way to put (a special case of) the Patterson conjecture for *cocompact torsion-free* Γ in a split group G is to say that the multiplicity $N_\Gamma(\pi)$ of an irreducible unitary principal series representation π in the space $L^2(\Gamma\backslash G)$ is given by

$$N_\Gamma(\pi) \ = \ \dim H^{d-r}(\Gamma, \pi^\omega),$$

where r is the rank of G and π^ω is the subspace of analytic vectors in π, finally, $d = \dim(G/K)$ is the dimension of the symmetric space attached to G, where K is a maximal compact subgroup.

Our main result states that this assertion can be generalized to all arithmetic groups provided the ordinary group cohomology is replaced by the cuspidal cohomology. It will probably also work for more general lattices, but we stick to arihmetic groups, because some of the constructions used in this paper, like the Borel-Serre compactification, or the decomposition of the regular G-representation on the space $L^2(\Gamma\backslash G)$, have in the literature only been formulated for arithmetic groups. The relation to the Lewis correspondence is as follows. In [25] Don Zagier states that the correspondence for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ can be interpreted as the identity

$$N_\Gamma(\pi) \;=\; \dim H^1_{par}(\Gamma, \pi^{\omega/2}),$$

where π is as before, $\pi^{\omega/2}$ is a slightly bigger space than π^ω and H^1_{par} is the parabolic cohomology. Since π is a unitary principal series representation it follows that $N_\Gamma(\pi)$ coincides with the multiplicity of π in the cuspidal part $L^2_{\mathrm{cusp}}(\Gamma\backslash G)$ of $L^2(\Gamma\backslash G)$. More precisely, the correspondence gives an isomorphism

$$\mathrm{Hom}_G\left(\pi, L^2_{\mathrm{cusp}}(\Gamma\backslash G)\right) \to H^1_{par}(\Gamma, \pi^{\omega/2}).$$

As a consequence of our main result we will get the following theorem.

THEOREM 0.1 *For every Fuchsian group* Γ *we have*

$$N_\Gamma(\pi) \;=\; \dim H^1_{\mathrm{cusp}}(\Gamma, \pi^\omega_{it}).$$

Here $H^\bullet_{\mathrm{cusp}}$ is the cuspidal cohomology. For finite dimensional modules the cuspidal cohomology is a subspace of the parabolic cohomology.
The following is our main theorem.

THEOREM 0.2 *Let* Γ *be a torsion-free arithmetic subgroup of a split semisimple Lie group* G. *Let* $\pi \in \hat{G}$ *be an irreducible unitary principal series representation. Then*

$$N_\Gamma(\pi) \;=\; \dim H^{d-r}_{\mathrm{cusp}}(\Gamma, \pi^\omega),$$

where $d = \dim G/K$ *and* r *is the real rank of* G.
For G *non-split the assertion remains true for a generic set of representations* π.

This raises many questions. For a finite dimensional representation E it is known that the cuspidal cohomology is a subspace of the parabolic cohomology. The same assertion for infinite dimensional E is wrong in general, see Corollary 5.2. Can one characterize those infinite dimensional E for which the cuspidal cohomology indeed injects into ordinary cohomology?
Another question suggests itself: in which sense does our construction in the case $\mathrm{PSL}_2(\mathbb{Z})$ coincide with the Lewis correspondence? To even formulate a conjecture we must assume two further conjectures. First assume that in the relevant cases cuspidal and parabolic cohomology coincide; next assume

that the parabolic cohomology with coefficients in π^ω agrees with parabolic cohomology in $\pi^{\omega/2}$. Let M_λ be the space of cusp forms of eigenvalue λ. Then our construction gives a map into the dual space of the cohomology, $\alpha : M_\lambda \to H^1_{par}(\Gamma, \pi^\omega)^*$. The Lewis construction on the other hand gives a map $\beta : M_\lambda \to H^1_{par}(\Gamma, \pi^\omega)$. Together they define a duality on M_λ. One is tempted to speculate that this duality coincides with the natural duality given by the integral on the upper half plane. If that were so, then the two maps α and β would determine each other.

1 Fuchsian groups

Let G be the group $\mathrm{SL}_2(\mathbb{R})/\pm 1$. For $s \in \mathbb{C}$ let π_s denote the principal series representation with parameter s. Recall that this representation can be viewed as the regular representation on the space of square integrable sections of a line bundle over $\mathbb{P}^1(\mathbb{R}) \cong G/P$, where P is the subgroup of upper triangular matrices. For $s \in i\mathbb{R}$ this representation will be irreducible unitary. For any admissible representation π of G let π^ω denote the space of analytic vectors in π. Then π^ω is a locally convex vector space with continuous G-representation ([18], p. 463). Let $\pi^{-\omega}$ be its continuous dual. For $\pi = \pi_s$ the space π_s^ω is the space of analytic sections of a line bundle over $\mathbb{P}^1(\mathbb{R})$. Let $\pi_s^{\omega/2}$ denote the space of sections which are smooth everywhere and analytic up to the possible exception of finitely many points. Let $\Gamma = \mathrm{SL}_2(\mathbb{Z})/\pm 1$ be the modular group. For an irreducible representation π of G let $N_\Gamma(\pi)$ be its multiplicity in $L^2(\Gamma\backslash G)$. Let $H^1_{par}(\Gamma, \pi_s^\omega)$ denote the *parabolic cohomology*, i.e., the subspace of $H^1(\Gamma, \pi_s^\omega)$ generated by all cocycles μ which vanish on parabolic elements. For the group $\Gamma = \mathrm{SL}_2(\mathbb{Z})/\pm 1$ this means that H^1_{par} consists of all cohomology classes which have a representing cocycle μ with $\mu\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 0$. In [25] D. Zagier stated that for $s \in i\mathbb{R}$,

$$N_\Gamma(\pi) \;=\; \dim H^1_{par}(\Gamma, \pi_s^{\omega/2}).$$

We will first relate this to the Patterson Conjecture for the cocompact case.

Theorem 1.1 *Let $\Gamma \subset G$ be a discrete, cocompact and torsion-free subgroup, then for $s \in i\mathbb{R}$,*

$$N_\Gamma(\pi_s) \;=\; \dim H^1_{par}(\Gamma, \pi_s^\omega) \;=\; \dim H^1(\Gamma, \pi_s^\omega).$$

Proof: Since Γ does not contain parabolic elements the parabolic cohomology coincides with the ordinary group cohomology. The Patterson Conjecture [7, 12] shows that

$$N_\Gamma(\pi_s) \;=\; \dim H^1(\Gamma, \pi_s^{-\omega}) - 2\dim H^2(\Gamma, \pi_s^{-\omega}).$$

Poincaré duality [8] implies that the dimension of the space $H^j(\Gamma, \pi_s^\omega)$ equals the dimension of $H^{2-j}(\Gamma, \pi_s^{-\omega})$. The Theorem follows from the next lemma.

Lemma 1.2 *For every Fuchsian group we have* $H^0(\Gamma, \pi_s^\omega) = 0$.

Proof: For this recall that every $f \in \pi_s^\omega$ is a continuous function on G satisfying among other things, $f(nx) = f(x)$ for every $n \in N$, where N is the unipotent group of all matrices modulo ± 1 which are upper triangular with ones on the diagonal. If f is Γ-invariant, then $f \in C(G/\Gamma)$. By Moore's Theorem ([26], Thm. 2.2.6) it follows that the action of N on G/Γ is ergodic. In particular, this implies that f must be constant. Since $s \in i\mathbb{R}$ this implies that $f = 0$. □

2 Arbitrary arithmetic groups

Throughout, let G be a semisimple Lie group with finite center and finitely many connected components.
Let Γ be an arithmetic subgroup of G and assume that Γ is torsion-free. Then Γ is the fundamental group of $\Gamma\backslash X$, where $X = G/K$ the symmetric space and every Γ-module M induces a local system or locally constant sheaf $\mathcal{M}$ on $\Gamma\backslash X$. In the étale picture the sheaf $\mathcal{M}$ equals $\mathcal{M} = \Gamma\backslash(X \times M)$, (diagonal action). Let $\overline{\Gamma\backslash X}$ denote the Borel-Serre compactification [3] of $\Gamma\backslash X$, then Γ also is the fundamental group of $\overline{\Gamma\backslash X}$ and M induces a sheaf also denoted by $\mathcal{M}$ on $\overline{\Gamma\backslash X}$. This notation is consistent as the sheaf on $\Gamma\backslash X$ is indeed the restriction of the one on $\overline{\Gamma\backslash X}$. Let $\partial(\Gamma\backslash X)$ denote the boundary of the Borel-Serre compactification. We have natural identifications

$$H^j(\Gamma, M) \ \cong \ H^j(\Gamma\backslash X, \mathcal{M}) \ \cong H^j(\overline{\Gamma\backslash X}, \mathcal{M}).$$

We define the *parabolic cohomology* of a Γ-module M to be the kernel of the restriction to the boundary, ie,

$$H^j_{par}(\Gamma, M) \ \stackrel{\text{def}}{=} \ \ker\left(H^j(\overline{\Gamma\backslash X}, \mathcal{M}) \to H^j(\partial(\Gamma\backslash X), \mathcal{M})\right).$$

The long exact sequence of the pair $(\overline{\Gamma\backslash X}, \partial(\Gamma\backslash X))$ gives rise to

$$\dots \to H^j_c(\Gamma\backslash X, \mathcal{M}) \to H^j(\Gamma\backslash X, \mathcal{M}) =$$

$$= H^j(\overline{\Gamma\backslash X}, \mathcal{M}) \to H^j(\partial(\Gamma\backslash X), \mathcal{M}) \to \dots$$

The image of the cohomology with compact supports under the natural map is called the *interior cohomology* of $\Gamma\backslash X$ and is denoted by $H^j_!(\Gamma\backslash X, \mathcal{M})$. The exactness of the above sequence shows that

$$H^j_{par}(\Gamma, M) \ \cong \ H^j_!(\Gamma\backslash X, \mathcal{M}).$$

Let E be a locally convex space. We shall write E' for its topological dual. We assume that Γ acts linearly and continuously on E. We will present a natural complex that computes the cohomology $H^\bullet(\Gamma, E)$.

Let $\mathcal{E}_\Gamma$ be the locally constant sheaf on $\Gamma\backslash X$ given by E. Then $\mathcal{E}_\Gamma$ has stalk E and $H^\bullet(\Gamma, E) = H^\bullet(X_\Gamma, \mathcal{E}_\Gamma)$.
Let $\Omega_\Gamma^0, \dots, \Omega_\Gamma^d$ be the sheaves of differential forms on X_Γ and let $\mathcal{E}_\Gamma^p$ be the sheaf locally given by

$$\mathcal{E}_\Gamma^p(U) \;=\; \Omega_\Gamma^p(U)\hat{\otimes}\mathcal{E}_\Gamma(U),$$

where $\hat{\otimes}$ denotes the completion of the algebraic tensor product $\otimes$ in the projective topology. Write $X_\Gamma = \Gamma\backslash G/K = \Gamma\backslash X$. Let d denote the exterior differential. Then $D = d \otimes 1$ is a differential on $\mathcal{E}_\Gamma^\bullet$ and

$$0 \to \mathcal{E}_\Gamma \xrightarrow{D} \mathcal{E}_\Gamma^0 \xrightarrow{D} \cdots \xrightarrow{D} \mathcal{E}_\Gamma^d \to 0$$

is a fine resolution of $\mathcal{E}_\Gamma$. Hence $H^\bullet(X_\Gamma, \mathcal{E}_\Gamma) = H^\bullet(X_\Gamma, \mathcal{E}_\Gamma^\bullet)$.
Let $\Omega^\bullet(X)$ be the space of differential forms on X. The complex $\mathcal{E}_\Gamma^\bullet(X_\Gamma)$ is isomorphic to the space of Γ-invariants $(\Omega^\bullet(X)\hat{\otimes}E)^\Gamma$. So we get

$$H^\bullet(\Gamma, E) \cong H^\bullet\left((\Omega^\bullet(X)\hat{\otimes}E)^\Gamma\right).$$

We can write

$$\Omega^p(X) \;=\; (C^\infty(G) \otimes \wedge^p\mathfrak{p}^*)^K,$$

where $\mathfrak{p}$ is the positive part in the Cartan decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, where $\mathfrak{g}$ is the complexified Lie algebra of G and $\mathfrak{k}$ is the complexified Lie algebra of K. The group K acts on $\mathfrak{p}^*$ via the coadjoint representation and on $C^\infty(G)$ via right translations and Γ, or more precisely G, acts by left translations on $C^\infty(G)$.
From now on we assume that E is not only a Γ-module but is a topological vector space that carries a continuous G-representation. We say that E is *admissible* if every K-isotype $E(\tau)$, $\tau \in \hat{K}$ is finite dimensional. Let E^∞ denote the subspace of smooth vectors. We say that E is smooth if $E = E^\infty$. We then have

$$(C^\infty(G) \otimes \wedge^p\mathfrak{p}^*)\hat{\otimes}E \cong C^\infty(G)\hat{\otimes}(E \otimes \wedge^p\mathfrak{p}^*)$$

as a $G \times K$-module, where G acts diagonally on $C^\infty(G)$ by left translations and on E by the given representation. The group K acts diagonally on $C^\infty(G)$ by right translations and on $\wedge^p\mathfrak{p}^*$ via the coadjoint action.

Lemma 2.1 *For any locally convex complete topological vector space F we have*

$$C^\infty(G)\hat{\otimes}F \;\cong\; C^\infty(G, F),$$

where the right hand side denotes the space of all smooth maps from G to F.

Proof: See [14], Example 1 after Theorem 13. □

Thus we have a $G \times K$-action on the space $C^\infty(G, \wedge^p\mathfrak{p}^* \otimes E)$ given by

$$(g, k).f \;=\; (\mathrm{Ad}^*(k) \otimes g)\, L_g\, R_k\, f,$$

where $L_g f(x) = f(g^{-1}x)$ and $R_k f(x) = f(xk)$.
The map

$$\psi\colon C^\infty(G, \wedge^p \mathfrak{p}^* \otimes E) \ \to\ C^\infty(G, \wedge^p \mathfrak{p}^* \otimes E)$$

given by

$$\psi(f)(x) = (1 \otimes x^{-1}).f(x)$$

is an isomorphism to the same space with a different the $G \times K$ structure. Indeed, one computes,

$$\begin{aligned}
\psi((g,k).f)(x) &= (1 \otimes x^{-1})(g,k).f(x) \\
&= (1 \otimes x^{-1})(\mathrm{Ad}^*(k) \otimes g) f(g^{-1}xk) \\
&= (\mathrm{Ad}^*(k) \otimes k(g^{-1}xk)^{-1}) f(g^{-1}xk) \\
&= (\mathrm{Ad}^*(k) \otimes k) R_k L_g \psi(f)(x).
\end{aligned}$$

For a smooth G-representation F we write $H^\bullet(\mathfrak{g}, K, F)$ for the cohomology of the standard complex of $(\mathfrak{g}, K)$-cohomology [5]. Then $H^\bullet(\mathfrak{g}, K, F) = H^\bullet(\mathfrak{g}, K, F_K)$, where F_K is the $(\mathfrak{g}, K)$-module of K-finite vectors in F.
If we assume that E is smooth, we get from this

$$H^\bullet(\Gamma, E) \ =\ H^\bullet(\mathfrak{g}, K, C^\infty(\Gamma\backslash G) \hat{\otimes} E).$$

In the case of finite dimensional E one can replace $C^\infty(\Gamma\backslash G)$ with the space of functions of moderate growth [6]. This is of importance, since it leads to a decomposition of the cohomology space into the cuspidal part and contributions from the parabolic subgroups. To prove this, one starts with differential forms of moderate growth and applies ψ. For infinite dimensional E this proof does not work, since it is not clear that ψ should preserve moderate growth, even if one knows that the matrix coefficients of E have moderate growth.
By the Sobolev Lemma the space of smooth vectors $L^2(\Gamma\backslash G)^\infty$ of the natural unitary representation of G on $L^2(\Gamma\backslash G)$ is a subspace of $C^\infty(\Gamma\backslash G)$. The representation $L^2(\Gamma\backslash G)$ splits as $L^2(\Gamma\backslash G) = L^2_{disc} \oplus L^2_{cont}$, where $L^2_{disc} = \bigoplus_{\pi \in \hat{G}} N_\Gamma(\pi)\pi$ is a direct Hilbert sum of irreducible representations and L^2_{cont} is a finite sum of continuous Hilbert integrals. The space of cusp forms $L^2_{\mathrm{cusp}}(\Gamma\backslash G) = \bigoplus_{\pi \in \hat{G}} N_{\Gamma,\mathrm{cusp}}(\pi)\pi$ is a subspace of L^2_{disc}. Note that $L^2_{\mathrm{cusp}}(\Gamma\backslash G)^\infty$ is a closed subspace of $C^\infty(G)$. The *cuspidal cohomology* is defined by

$$H^\bullet_{\mathrm{cusp}}(\Gamma, E) \ =\ H^\bullet(\mathfrak{g}, K, L^2_{\mathrm{cusp}}(\Gamma\backslash G)^\infty \hat{\otimes} E).$$

For finite dimensional E it turns out that $H^\bullet_{\mathrm{cusp}}(\Gamma, E)$ coincides with the image in $H^\bullet(\mathfrak{g}, K, C^\infty(\Gamma\backslash G)\hat{\otimes} E)$ under the inclusion map. This comes about as a consequence of the fact that the cohomology can also be computed using functions of uniform moderate growth and that in the space of such functions, $L^2_{\mathrm{cusp}}(\Gamma\backslash G)^\infty$ has a G-complement. The Borel-conjecture [13] is a refinement of this assertion. For infinite dimensional E this injectivity does not hold in general, see Corollary 5.2.

We define the *reduced cuspidal cohomology* to be the image $\tilde{H}^\bullet_{\rm cusp}(\Gamma, E)$ of $H^\bullet_{\rm cusp}(\Gamma, E)$ in $H^\bullet(\Gamma, E)$. Finally, let $H_{(2)}(\Gamma, E)$ be the image of the space $H^\bullet(\mathfrak{g}, K, L^2(\Gamma\backslash G)^\infty\hat{\otimes}E)$ in $H^\bullet(\mathfrak{g}, K, C^\infty(\Gamma\backslash G)\hat{\otimes}E)$.

Proposition 2.2 *We have the following inclusions of cohomology groups,*

$$\tilde{H}^\bullet_{\rm cusp}(\Gamma, E) \ \subset \ H^\bullet_{par}(\Gamma, E) \ \subset \ H^\bullet_{(2)}(\Gamma, E).$$

Proof: The cuspidal condition ensures that every cuspidal class vanishes on each homology class of the boundary. This implies the first conclusion. Since every parabolic class has a compactly supported representative, the second also follows. □

3 Gelfand Duality

Recall that a *Harish-Chandra module* is a $(\mathfrak{g}, K)$-module which is admissible and finitely generated. Every Harish-Chandra module is of finite length. For a Harish-Chandra module V write $\tilde{V}$ for its dual, ie, $\tilde{V} = (V^*)_K$, the K-finite vectors in the algebraic dual.

A *globalization* of a Harish-Chandra module V is a continuous representation of G on a complete locally convex vector space W such that V is isomorphic to the $(\mathfrak{g}, K)$-module of K-finite vectors W_K. It was shown in [18] that there is a minimal globalization $V^{\min}$ and a maximal globalization $V^{\max}$ such that for every globalization W there are unique functorial continuous linear G-maps

$$V^{\min} \ \hookrightarrow \ W \ \hookrightarrow \ V^{\max}.$$

The spaces $V^{\min}$ and $V^{\max}$ are given explicitly by

$$V^{\min} \ = \ C_c^\infty(G) \otimes_{\mathfrak{g},K} V$$

and

$$V^{\max} \ = \ \mathrm{Hom}_{\mathfrak{g},K}(\tilde{V}, C^\infty(G)).$$

The action of G on V^{max} is given by

$$g.\alpha(\tilde{v})(x) \ = \ \alpha(\tilde{v})(g^{-1}x).$$

Let $\hat{G}$ be the unitary dual of G, i.e., the set of all isomorphism classes of irreducible unitary representations of G. Note [18] that for $\pi \in \hat{G}$ we have $(\pi_K)^{\min} = \pi^\omega$ and $(\pi_K)^{\max} = \pi^{-\omega}$.

The following is a key result of this paper.

Theorem 3.1 *Let F be a smooth G-representation on a complete locally convex topological vector space. Then there is a functorial isomorphism*

$$H^\bullet(\mathfrak{g}, K, F\hat{\otimes}V^{\max}) \ \to \ \mathrm{Ext}^\bullet_{\mathfrak{g},K}(\tilde{V}, F),$$

where as usual one writes $\mathrm{Ext}^\bullet_{\mathfrak{g},K}(\tilde{V}, F)$ *for* $\mathrm{Ext}^\bullet_{\mathfrak{g},K}(\tilde{V}, F_K)$.

PROOF: We have

$$\begin{aligned} F\hat{\otimes}V^{\max} &= F\hat{\otimes}\mathrm{Hom}_{\mathfrak{g},K}(\tilde{V},C^\infty(G)) \\ &= \mathrm{Hom}_{\mathfrak{g},K}(\tilde{V},F\hat{\otimes}C^\infty(G)) \\ &= \mathrm{Hom}_{\mathfrak{g},K}(\tilde{V},C^\infty(G,F)). \end{aligned}$$

LEMMA 3.2 *The map*

$$\begin{aligned} \mathrm{Hom}_{\mathfrak{g},K}\left(\tilde{V},C^\infty(G,F)\right)^{\mathfrak{g},K} &\to \mathrm{Hom}_{\mathfrak{g},K}\left(\tilde{V},F\right) \\ \phi &\mapsto \alpha, \end{aligned}$$

with $\alpha(\tilde{v}) = \phi(\tilde{v})(1)$ *is an isomorphism.*

PROOF: Note that ϕ satisfies

$$\phi(X.\tilde{v})(x) = \frac{d}{dt}\phi(\tilde{v})(x\exp(tX))\Big|_{t=0}, \qquad X\in\mathfrak{g},$$

since it is a $(\mathfrak{g},K)$-homomorphism. Further,

$$\frac{d}{dt}\phi(\tilde{v})(\exp(tX)\,x)\Big|_{t=0} = X.\phi(\tilde{v})(x), \qquad X\in\mathfrak{g},$$

since ϕ is $(\mathfrak{g},K)$-invariant. Similar identities hold for the K-action. This implies that α is a $(\mathfrak{g},K)$-homomorphism. Note that the $(\mathfrak{g},K)$-invariance of ϕ also leads to

$$\phi(\tilde{v})(gx) = g.\phi(\tilde{v})(x), \qquad g,x\in G.$$

Hence if $\alpha = 0$ then $\phi = 0$ so the map is injective. For surjectivity let α be given and define ϕ by $\phi(\tilde{v})(x) = x.\alpha(\tilde{v})$. Then ϕ maps to α. □

By the Lemma we get an isomorphism

$$H^0(\mathfrak{g},K,F\hat{\otimes}V^{\max}) \cong \mathrm{Hom}_{\mathfrak{g},K}(\tilde{V},F) \cong \mathrm{Ext}^0_{\mathfrak{g},K}(\tilde{V},F)$$

and thus a functorial isomorphism on the zeroth level. We will show that both sides in the theorem define universal δ-functors [19]. From this the theorem will follow. Fix V and let $S^j(F) = H^j(\mathfrak{g},K,F\hat{\otimes}V^{\max})$ as well as $T^j(F) = \mathrm{Ext}^j_{\mathfrak{g},K}(\tilde{V},F)$. We will show that $S^\bullet$ and $T^\bullet$ are universal δ-functors from the category $Rep_s^\infty(G)$ defined below to the category of complex vector spaces. The objects of $Rep_s^\infty(G)$ are smooth continuous representations of G on Hausdorff locally convex topological vector spaces and the morphisms are *strong morphisms*. A continuous G-morphism $f : A \to B$ is called strong morphism or *s-morphism* if (a) $\ker f$ and $\mathrm{im} f$ are closed topological direct summands and (b) f induces an isomorphism of $A/\ker f$ onto $f(A)$. Then by [5], Chapter IX, the category $Rep_s^\infty(G)$ is an abelian category with enough injectives. In fact, for $F \in Rep_s^\infty(G)$ the map $F \to C^\infty(G,F)$ mapping f to the function $\alpha(x) = x.f$ is a monomorphism into the s-injective object $C^\infty(G,F)$ (cf. [5], Lemma IX.5.2), which is considered a G-module via $x\alpha(y) = \alpha(yx)$. Let us consider $S^\bullet$ first. By Corollary IX.5.6 of [5] we have

$$S^\bullet(F) \;=\; H^\bullet(\mathfrak{g}, K, F\hat{\otimes}V^{\max}) \;\cong\; H_d^\bullet(G, F\hat{\otimes}V^{\max}),$$

where the right hand side is the differentiable cohomology. The functor $.\hat{\otimes}V^{\max}$ is s-exact and therefore $S^\bullet$ is a δ-functor. We show that it is erasable. For this it suffices to show that $S^j(C^\infty(G,F)) = 0$ for $j > 0$. Now

$$C^\infty(G,F)\hat{\otimes}V^{\max} \;\cong\; C^\infty(G)\hat{\otimes}F\hat{\otimes}V^{\max} \;\cong\; C^\infty(G, F\hat{\otimes}V^{\max})$$

and therefore for $j > 0$,

$$S^j(C^\infty(G,F)) \;\cong\; H_d^j(G, C^\infty(G, F\hat{\otimes}V^{\max})) \;=\; 0,$$

since $C^\infty(G, F\hat{\otimes}V^{\max})$ is s-injective. Thus $S^\bullet$ is erasable and therefore universal.
Next consider $T^\bullet(F) = \mathrm{Ext}^\bullet_{\mathfrak{g},K}(\tilde{V}, F)$. Since an exact sequence of smooth representations gives an exact sequence of $(\mathfrak{g}, K)$-modules, it follows that $T^\bullet$ is a δ-functor.To show that it is erasable let $j > 0$. Then

$$\begin{aligned} T^j(C^\infty(G,F)) &= \mathrm{Ext}^j_{\mathfrak{g},K}(\tilde{V}, C^\infty(G)\hat{\otimes}F) \\ &= H^j(\mathfrak{g}, K, \mathrm{Hom}_{\mathbb{C}}(\tilde{V}, C^\infty(G))\hat{\otimes}F) \\ &= H_d^j(G, \mathrm{Hom}_{\mathbb{C}}(\tilde{V}, C^\infty(G))\hat{\otimes}F) \\ &= H_d^j(G, \mathrm{Hom}_{\mathbb{C}}(\tilde{V}, C^\infty(G)))\hat{\otimes}F \\ &= \mathrm{Ext}^j_{\mathfrak{g},K}(\tilde{V}, C^\infty(G))\hat{\otimes}F. \end{aligned}$$

By Theorem 6.13 of [18] we have $\mathrm{Ext}^j_{\mathfrak{g},K}(\tilde{V}, C^\infty(G)) = 0$. The Theorem is proven. □

Choosing $C^\infty(\Gamma\backslash G)$ and $L^2_{\mathrm{cusp}}(\Gamma\backslash G)^\infty$ for F in Theorem 3.1 gives the following Corollary.

Corollary 3.3 (i)

$$H^p(\Gamma, V^{\max}) \;\cong\; \mathrm{Ext}^p_{\mathfrak{g},K}(\tilde{V}, C^\infty(\Gamma\backslash G)).$$

For Γ cocompact and $p = 0$ this is known under the name Gelfand Duality*.*

(ii)

$$H^\bullet_{\mathrm{cusp}}(\Gamma, V^{\max}) \;\cong\; \mathrm{Ext}^\bullet_{\mathfrak{g},K}(\tilde{V}, L^2_{\mathrm{cusp}}(\Gamma\backslash G)^\infty).$$

4 The case of the maximal globalization

The space of cusp forms decomposes discretely,

$$L^2_{\mathrm{cusp}}(\Gamma\backslash G) \;=\; \bigoplus_{\pi\in\hat{G}} N_{\Gamma,\mathrm{cusp}}(\pi)\pi.$$

Suppose that V has an infinitesimal character χ. Let $\hat{G}(\chi)$ be the set of all irreducible unitary representations of G with infinitesimal character χ. It is easy to see that

$$\begin{aligned}
\mathrm{Ext}^\bullet_{\mathfrak{g},K}(\tilde{V}, L^2_{\mathrm{cusp}}(\Gamma\backslash G)^\infty) &= \mathrm{Ext}^\bullet_{\mathfrak{g},K}\left(\tilde{V}, \bigoplus_{\pi\in\hat{G}(\chi)} N_{\Gamma,\mathrm{cusp}}(\pi)\pi_K\right)\\
&= \bigoplus_{\pi\in\hat{G}(\chi)} N_{\Gamma,\mathrm{cusp}}(\pi)\,\mathrm{Ext}^\bullet_{\mathfrak{g},K}\left(\tilde{V},\pi_K\right)\\
&= \bigoplus_{\pi\in\hat{G}(\chi)} N_{\Gamma,\mathrm{cusp}}(\pi)\,\mathrm{Ext}^\bullet_{\mathfrak{g},K}\left(\tilde{\pi}_K, V\right)
\end{aligned}$$

The last line follows by dualizing.
Let P be a parabolic subgroup and $\mathfrak{m}\oplus\mathfrak{a}\oplus\mathfrak{n}$ a Langlands decomposition of its Lie algebra.

LEMMA 4.1 *For a $(\mathfrak{g},K)$-module π and a $(\mathfrak{a}\oplus\mathfrak{m},K_M)$-module U we have*

$$\mathrm{Hom}_{\mathfrak{g},K}(\pi,\mathrm{Ind}_P^G(U)) \;\cong\; \mathrm{Hom}_{\mathfrak{a}\oplus\mathfrak{m},K_M}(H_0(\mathfrak{n},\pi), U\otimes\mathbb{C}_{\rho_P}),$$

where $\mathbb{C}_{\rho_P}$ is the one dimensional A-module given by ρ_P.

PROOF: See [15] page 101. □

LEMMA 4.2 *Let $\mathcal{C}$ be an abelian category with enough injectives. Let $\mathfrak{a}$ be a finite dimensional abelian complex Lie algebra and let T be a covariant left exact functor from $\mathcal{C}$ to the category of $\mathfrak{a}$-modules. Assume that T maps injectives to $\mathfrak{a}$-acyclics and that T has finite cohomological dimension, i.e., that $R^pT=0$ for p large. Let M be an object of $\mathcal{C}$ such that $R^pT(M)$ is finite dimensional for every p. Let $H_\mathfrak{a}$ denote the functor $H^0(\mathfrak{a},\cdot)$. Then*

$$\sum_{p\geq 0}\binom{p}{r}(-1)^{p+r}\,\dim R^p(H_\mathfrak{a}\circ T)(M) \;=\; \sum_{p\geq 0}(-1)^p\,\dim H^0(\mathfrak{a},R^pT(M)),$$

where $r=\dim\mathfrak{a}$. If M is T-acyclic, then these alternating sums degenerate to

$$\dim R^r(H_\mathfrak{a}\circ T)(M) \;=\; \dim H^0(\mathfrak{a},T(M)).$$

Proof: Split $\mathfrak{a} = \mathfrak{a}_1 \oplus \mathfrak{b}_1$, where $\dim \mathfrak{a}_1 = 1$. Consider the Grothendieck spectral sequence with $E_2^{p,q} = H^p(\mathfrak{a}_1, R^q(H_{\mathfrak{b}_1} \circ T)(M))$ which abuts to $R^{p+q}(H_{\mathfrak{a}} \circ T)(M)$ (see [19], Theorem XX.9.6). Since $\mathfrak{a}_1$ is one dimensional, for any finite dimensional $\mathfrak{a}_1$-module V we have $H^0(\mathfrak{a}, V) \cong H^1(\mathfrak{a}, V)$ and $H^p(\mathfrak{a}_1, V) = 0$ if $p > 1$. This implies $E_2^{0,q} \cong E_2^{1,q}$ and $E_2^{p,q} = 0$ for $p \notin \{0,1\}$. The spectral sequence therefore degenerates and

$$\begin{aligned}
&\sum_{p\ge 0} \binom{p}{r} (-1)^{p+r} \dim R^p(H_{\mathfrak{a}} \circ T)(M) \\
=\;& \sum_{p\ge 0} \binom{p}{r} (-1)^{p+r} \dim E_2^{0,p} + \sum_{p\ge 1} \binom{p}{r} (-1)^{p+r} \dim E_2^{1,p-1} \\
=\;& \sum_{p\ge 0} \binom{p}{r} (-1)^{p+r} \dim E_2^{0,p} + \sum_{p\ge 0} \binom{p+1}{r} (-1)^{p+r-1} \dim E_2^{1,p} \\
=\;& \sum_{p\ge 0} \left(\binom{p+1}{r} - \binom{p}{r} \right) (-1)^{p+r-1} \dim E_2^{0,p} \\
=\;& \sum_{p\ge 0} \binom{p}{r-1} (-1)^{p+r-1} \dim E_2^{0,p} \\
=\;& \sum_{p\ge 0} \binom{p}{r-1} (-1)^{p+r-1} \dim H^0(\mathfrak{a}_1, R^p(H_{\mathfrak{b}_1} \circ T)(M)).
\end{aligned}$$

Next we split $\mathfrak{b}_1 = \mathfrak{a}_2 \oplus \mathfrak{b}_2$, where $\mathfrak{a}_2$ is one-dimensional. Since the $\mathfrak{a}_1$-action commutes with the a_2-action the isomorphism

$$H^0(\mathfrak{a}_2, R^p(H_{\mathfrak{b}_2} \circ T)(M)) \;\cong\; H^1(\mathfrak{a}_2, R^p(H_{\mathfrak{b}_2} \circ T)(M))$$

is an $\mathfrak{a}_1$-isomorphism. Therefore we apply the same argument to get down to

$$\sum_{p\ge 0} \binom{p}{r-2} (-1)^{p+r-2} \dim H^0(\mathfrak{a}_1 \oplus \mathfrak{a}_2, R^p(H_{\mathfrak{b}_2} \circ T)(M)).$$

Iteration gives the claim.
To get the last assertion of the lemma, note that if M is T-acyclic, then following the inductive argument above, one sees that $R^p(H_{\mathfrak{a}} \circ T(M) = 0$ for $p > r$. □

Let P be a minimal parabolic subgroup of G so that M is compact. For a unitary irreducible representation σ of M and linear functional $\nu \in i\mathfrak{a}^*$ we obtain the unitary principal series representation $\pi_{\sigma,\nu}$ of G induced from P. Let $\mathfrak{t}$ be a Cartan subalgebra of $\mathfrak{m} = \mathrm{Lie}_{\mathbb{C}}(M)$. Then $\mathfrak{h} = \mathfrak{a} \oplus \mathfrak{t}$ is a Cartan subalgebra of $\mathfrak{g}$. Let $\Lambda_\sigma \in \mathfrak{t}^*$ be a representative of the infinitesimal character of σ. Then $\Lambda_\sigma + \nu \in \mathfrak{h}^*$ is a representative of the infinitesimal character of $\pi_{\sigma,\nu}$.

We say that the parameters (σ,ν) are *generic* if $\pi_{\sigma,\nu}$ is irreducible and for any two w, w' in the Weyl group of $(\mathfrak{h},\mathfrak{g})$ the linear functional

$$w(\Lambda_\sigma+\nu)|_{\mathfrak{a}} - w'(\Lambda_\sigma+\nu)|_{\mathfrak{a}}$$

on $\mathfrak{a}$ is not a positive integer linear combination of positive roots.

Theorem 4.3 *If G is $\mathbb{R}$-split and $\pi_{\sigma,\nu}$ is irreducible, we have*

$$N_\Gamma(\pi_{\sigma,\nu}) \;=\; H^r_{\rm cusp}(\Gamma,\pi^{-\omega}_{\sigma,\nu}).$$

If G is not split, the assertions remains true if the parameters (σ,ν) are generic.

Proof: First note that since G is split, the decomposition of $L^2(\Gamma\backslash G)$ as in [20, 1] implies that for imaginary ν one has $N_\Gamma(\pi_{\sigma,\nu}) = N_{\Gamma,{\rm cusp}}(\pi_{\sigma,\nu})$, since the Eisenstein series are regular at imaginary ν. Applying Lemma 4.1 with $\pi = L^2_{\rm cusp}(\Gamma\backslash G)^\infty$ and ${\rm Ind}_P^G(U) = \pi_{\sigma,\nu}$ we find

$$\begin{aligned}
&{\rm Hom}_{\mathfrak{a}\oplus\mathfrak{m},M}\left(H_0(\mathfrak{n}, L^2_{\rm cusp}(\Gamma\backslash G)^\infty_K), \sigma\otimes(\nu+\rho_P)\right)\\
&\quad\cong\quad {\rm Hom}_{\mathfrak{g},K}\left(L^2_{\rm cusp}(\Gamma\backslash G)^\infty, \pi_{\sigma,\nu}\right)\\
&\quad\cong\quad {\rm Hom}_{\mathfrak{g},K}\left(\tilde\pi_{\sigma,\nu}, L^2_{\rm cusp}(\Gamma\backslash G)^\infty\right)
\end{aligned}$$

so that

$$N_{\Gamma,{\rm cusp}}(\pi_{\sigma,\nu}) = \dim {\rm Hom}_{\mathfrak{a}\oplus\mathfrak{m},M}\left(H_0(\mathfrak{n}, L^2_{\rm cusp}(\Gamma\backslash G)^\infty_K), \sigma\otimes(\nu+\rho_P)\right).$$

In order to calculate the latter we apply Lemma 4.2 to the category $\mathcal{C}$ of $(\mathfrak{g},K)$ modules and

$$\begin{aligned}
T(V) &= {\rm Hom}_M(H_0(\mathfrak{n},\tilde V), U\otimes\mathbb{C}_{\rho_P})\\
&= (H_0(\mathfrak{n},\tilde V)^*\otimes U\otimes\mathbb{C}_{\rho_P})^M.
\end{aligned}$$

The conditions of the Lemma 4.2 are easily seen to be satisfied since $H_0(\mathfrak{n},\cdot)$ maps injectives to injectives and $H^0(M,\cdot)$ is exact. Note that in the case of a representation π of G,

$$\begin{aligned}
H_{\mathfrak{a}}\circ T(\pi) &\cong {\rm Hom}_{\mathfrak{a}\oplus\mathfrak{m},M}(H_0(\mathfrak{n},\tilde\pi_K), U\otimes\mathbb{C}_{\rho_P})\\
&\cong {\rm Hom}_{\mathfrak{g},K}(\tilde\pi, {\rm Ind}_P^G(U))
\end{aligned}$$

by Lemma 4.1. From this we obtain

$${\rm Ext}^j_{\mathfrak{g},K}(\tilde\pi, {\rm Ind}_P^G(U)) = R^p(H_{\mathfrak{a}}\circ T)(\pi).$$

Now Lemma 4.2 shows that

$$\dim {\rm Ext}^r_{\mathfrak{g},K}(\tilde\pi, {\rm Ind}_P^G(U))$$

equals

$$\dim \mathrm{Hom}_{AM}(H_0(\mathfrak{n},\tilde{\pi}_K), U \otimes \mathbb{C}_{\rho_P}).$$

Suppose that π is an irreducible summand in $L^2_{\mathrm{cusp}}(\Gamma\backslash G)$. If G is split, the set of π which share the same infinitesimal character as $\pi_{\sigma,\nu}$ equals the set of all $\pi_{\xi,w\nu}$, where w ranges over the Weyl group and $\xi \in \hat{M}$. Then the space $\mathrm{Hom}_{AM}(H_j(\mathfrak{n},\tilde{\pi}_K), U \otimes \mathbb{C}_{\rho_P})$ is only non-zero for $\pi = \pi_{\xi,w\nu}$ for some w. But then Proposition 2.32 of [15] implies that $H_j(\mathfrak{n},\tilde{\pi}_K)$ is zero unless $j = 0$. The same conclusion is assured in the non-split case by the genericity condition. Now the proof is completed by the following calculation:

$$\begin{aligned} N_{\Gamma,\mathrm{cusp}}(\pi_{\sigma,\nu}) &= \dim \mathrm{Hom}_{\mathfrak{a}\oplus\mathfrak{m},M}(H_0(\mathfrak{n}, L^2_{\mathrm{cusp}}(\Gamma\backslash G)^\infty_K), \sigma \otimes (\nu + \rho_P)) \\ &= \dim \mathrm{Ext}^r_{\mathfrak{g},K}(L^2_{\mathrm{cusp}}(\Gamma\backslash G), \pi_{\sigma,\nu}) \\ &= \dim \mathrm{Ext}^r_{\mathfrak{g},K}(\tilde{\pi}_{\sigma,\nu}, L^2_{\mathrm{cusp}}(\Gamma\backslash G)) \\ &= \dim H^r_{\mathrm{cusp}}(\Gamma, \pi^{-\omega}_{\sigma,\nu}), \end{aligned}$$

where the last equality is a consequence of Corollary 3.3(ii), applied to $V = \pi_{\sigma,\nu}$. □

5 Poincaré duality

In order to conclude the main Theorem it suffices to prove the following Poincaré duality.

Theorem 5.1 *(Poincaré duality)*
For every Harish-Chandra module V,

$$H^j_{\mathrm{cusp}}(\Gamma, V^{\max}) \;\cong\; H^{d-j}_{\mathrm{cusp}}(\Gamma, \tilde{V}^{\min})^*,$$

and both spaces are finite dimensional.

Before we prove the theorem, we add a Corollary.

Corollary 5.2 *Let Γ be a torsion-free non-uniform lattice in $G = \mathrm{PSL}_2(\mathbb{R})$ and $\pi \in \hat{G}$ a principal series representation with $N_{\Gamma,\mathrm{cusp}}(\pi) \neq 0$. Let $E = \tilde{\pi}_K^{min}$. Then the natural map $H^\bullet_{\mathrm{cusp}}(\Gamma, E) \to H^\bullet(\Gamma, E)$ is not injective.*

Proof of the Corollary: We have $H^0_{\mathrm{cusp}}(\Gamma, \pi^{max}) \neq 0$ and by the Poincaré duality, $H^2_{\mathrm{cusp}}(\Gamma, E) \neq 0$. However, as the cohomological dimension of Γ is 1, it follows that $H^2(\Gamma, E) = 0$. □

Proof of the Theorem: A *duality* between two complex vector spaces E, F is a bilinear pairing,

$$\langle .,.\rangle : E \times F \;\to\; \mathbb{C}$$

which is *non-degenerate*, i.e., for every $e \in E$ and every $f \in F$,

$$\langle e, F\rangle = 0 \quad\Rightarrow\quad e = 0,$$
$$\langle E, f\rangle = 0 \quad\Rightarrow\quad f = 0.$$

We say that E and F are *in duality* if there is a duality between them. Note that if E and F are in duality and one of them is finite dimensional, then the other also is and their dimensions agree. The pairing is called *perfect* if it induces isomorphisms $E \cong F^*$ and $F \cong E^*$. If E and F are topological vector spaces then the pairing is called *topologically perfect* if it induces topological isomorphisms $E \cong F'$ and $F \cong E'$, where the dual spaces are equipped with the strong dual topology.

Now suppose that V and W are $\mathfrak{g}, K$-modules in duality through a $\mathfrak{g}, K$-invariant pairing. Recall the canonical complex defining $\mathfrak{g}, K$-cohomology which is given by $C^q(V) = \mathrm{Hom}_K(\wedge^q(\mathfrak{g}/\mathfrak{k}), V) = (\wedge^q(\mathfrak{g}/\mathfrak{k})^* \otimes V)^K$. Let $d = \dim G/K$. The prescription $\langle y \otimes v, y' \otimes w\rangle = (-1)^q \langle v, w\rangle\, y \wedge y'$ gives a pairing from $C^q(V) \times C^{d-q}(W)$ to $\wedge^d(\mathfrak{g}/\mathfrak{k})^* \cong \mathbb{C}$. Let $d : C^q \to C^{q+1}$ be the differential, then one sees [5], $\langle da, b\rangle = \langle a, db\rangle$.

Let π be an irreducible unitary representation of G. Then the spaces π^∞ and $\tilde\pi^{-\infty}$ are each other's strong duals [10]. The same holds for $V^{\max}$ and $\tilde V^{\min}$ [18].

Lemma 5.3 *The spaces $A = \pi^{-\infty}\hat\otimes V^{\max}$ and $B = \tilde\pi^\infty \hat\otimes \tilde V^{\min}$ are each other's strong duals. Both of them are LF-spaces.*

Proof: Since $C^\infty(G)$ is nuclear and Fréchet and $\tilde\pi$ is a Hilbert space the results of [24], §III.50, allow us to conclude that $C^\infty(G, \tilde\pi)' = C^\infty(G)\hat\otimes\tilde\pi$ is nuclear which is then true also for $C^\infty(G, \tilde\pi)$. Now the embedding of $\tilde\pi^\infty$ into $C^\infty(G, \tilde\pi)$ shows the nuclearity of $\tilde\pi^\infty$.

Since $\tilde V$ is finitely generated one can embed the space

$$V^{\max} = \mathrm{Hom}_{\mathfrak{g},K}(\tilde V, C^\infty(G))$$

into a strict inductive limit $\varinjlim_j \mathrm{Hom}(\tilde V^j, C^\infty(G))$ with finite dimensional V^j's. Then the nuclearity of $V^{\max}$ follows from the nuclearity of

$$\mathrm{Hom}(\tilde V^j, C^\infty(G)) = (\tilde V^j)^* \otimes C^\infty(G).$$

We conclude that the spaces $V^{\max}$ and $\tilde\pi^\infty$ are nuclear Fréchet spaces. Their duals $\pi^{-\infty}$ and $\tilde V^{\min}$ are LF-spaces (see [14], Introduction IV). Therefore they all are barreled ([22], p. 61). By [22], p. 119 we know that the inductive completions of the tensor products $\pi^{-\infty}\bar\otimes V^{\max}$ and $\pi^\infty\bar\otimes V^{\min}$ are barreled. Since $V^{\max}$ and π^∞ are nuclear, these inductive completions coincide with the projective completions. So A and B are barreled. By Theorem 14 of [14] it follows that the strong duals A' and B' are complete and by the Corollary to Lemma 9 of [14] it follows that $A' = B$ and $B' = A$. Finally, Lemma 9 of [14] implies that A and B are LF-spaces. $\square$

Proposition 5.4 *For every $\pi \in \hat{G}$ and every Harish-Chandra module V the vector spaces $H^q(\mathfrak{g}, K, \pi^{-\infty}\hat{\otimes}V^{\max})$ and $H^q(\mathfrak{g}, K, \tilde{\pi}^{\infty}\hat{\otimes}\tilde{V}^{\min})$ are finite dimensional. The above pairing between their canonical complexes induces a duality between them, so*

$$H^q(\mathfrak{g}, K, \pi^{-\infty}\hat{\otimes}V^{\max}) \;\cong\; H^{d-q}(\mathfrak{g}, K, \tilde{\pi}^{\infty}\hat{\otimes}\tilde{V}^{\min})^*.$$

Proof: Note that by Theorem 3.1,

$$H^q(\mathfrak{g}, K, \pi^{-\infty}\hat{\otimes}V^{\max}) \cong \mathrm{Ext}^q_{\mathfrak{g},K}(\tilde{V}, \pi^{-\infty}) \cong \mathrm{Ext}^q_{\mathfrak{g},K}(\tilde{V}, \pi_K)$$

and the latter space is finite dimensional ([5], Proposition I.2.8). The proposition will thus follow from the next lemma.

Lemma 5.5 *Let A, B be smooth representations of G. Suppose that A and B are LF-spaces and that they are in perfect topological duality through a G-invariant pairing. Assume that $H^\bullet(\mathfrak{g}, K, A)$ is finite dimensional. Then the natural pairing between $H^q(\mathfrak{g}, K, A)$ and $H^{d-q}(\mathfrak{g}, K, B)$ is perfect.*

Proof: We only have to show that the pairing is non-degenerate. We will start by showing that the induced map $H^{d-q}(\mathfrak{g}, K, B)$ to $H^q(\mathfrak{g}, K, A)^*$ is injective. So let $b \in Z^{d-q}(B) = C^{d-q}(B) \cap \ker d$ with $\langle a, b\rangle = 0$ for every $a \in Z^q(A)$. Define a map $\psi\colon d(C^q(A)) \to \mathbb{C}$ by

$$\psi(da) \;=\; \langle a, b\rangle\,.$$

We now show that the image $d(C^q(A))$ is a closed subspace of $C^{q+1}(A)$ and that the map $C^q(A)/\ker d \to d(C^q(A))$ is a topological isomorphism. For this let E be a finite dimensional subspace of $Z^{q+1}(A)$ that bijects to $H^{q+1}(\mathfrak{g}, K, A)$. Since E is finite dimensional, it is closed. The map $\eta = d + 1\colon C^q(A) \oplus E \to Z^{q+1}(A)$ is continuous and surjective. Since $C^q(A)$ and $Z^{q+1}(A)$ are LF-spaces, the map η is open (see [24], p. 78), hence it induces a topological isomorphism $(C^q(A)/\ker d) \oplus E \;\to\; Z^{q+1}(A)$. This implies that $d(C^q(A))$ is closed and $C^q(A)/\ker d \;\to\; d(C^q(A))$ is a topological isomorphism. Consequently, the map ψ is continuous. Hence it extends to a continuous linear map on $C^{q+1}(A)$. Therefore, it is given by an element f of $C^{d-q-1}(A)$, so

$$\langle a, b\rangle \;=\; \langle da, f\rangle \;=\; \langle a, df\rangle$$

for every $a \in C^q(A)$. We conclude $b = df$ and thus the non-degeneracy on one side. In particular it follows that $H^\bullet(\mathfrak{g}, K, B)$ is finite dimensional as well. The claim now follows by symmetry. □

We will now deduce Theorem 5.1. We have

$$\begin{aligned} H^q_{\mathrm{cusp}}(\Gamma, V^{\max}) &\cong \bigoplus_{\pi\in\hat{G}(\chi)} N_{\Gamma,\mathrm{cusp}}(\pi)\,\mathrm{Ext}^q_{\mathfrak{g},K}(\tilde{V}, \pi_K) \\ &\cong \bigoplus_{\pi\in\hat{G}(\chi)} N_{\Gamma,\mathrm{cusp}}(\pi)\, H^q(\mathfrak{g}, K, \pi^{-\infty}\hat{\otimes} V^{\max}) \\ &\cong \bigoplus_{\pi\in\hat{G}(\chi)} N_{\Gamma,\mathrm{cusp}}(\pi)\, H^{d-q}(\mathfrak{g}, K, \tilde{\pi}^{\infty}\hat{\otimes} \tilde{V}^{\min})^* \\ &\cong \bigoplus_{\pi\in\hat{G}(\chi)} N_{\Gamma,\mathrm{cusp}}(\pi)\, H^{d-q}(\mathfrak{g}, K, \pi^{\infty}\hat{\otimes} \tilde{V}^{\min})^* \\ &\cong H^{d-q}_{\mathrm{cusp}}(\Gamma, \tilde{V}^{\min})^*. \end{aligned}$$

In the second to last step we have used the fact that L^2_{cusp} is self-dual. Theorem 5.1 and thus Theorem 0.2 follow.

It remains to deduce Theorem 0.1. For Γ torsion-free arithmetic it follows directly from Theorem 0.2 and Lemma 1.2. Since the Borel-Serre compactification exists for arbitrary Fuchsian groups, the proof runs through and we also get Theorem 0.1 for torsion-free Fuchsian groups. An arbitrary Fuchsian group Γ has a finite index subgroup Γ' which is torsion-free. An inspection shows that all our constructions allow descent from Γ'-invariants to Γ-invariants and thus Theorem 0.1 follows. □

References

[1] Arthur, J.: *Eisenstein series and the trace formula.* Automorphic Forms, Representations, and L-Functions. Corvallis 1977; Proc. Symp. Pure Math. XXXIII, 253-274 (1979).

[2] Borel, A.; Garland, H.: *Laplacian and discrete spectrum of an arithmetic group.* Amer. J. Math. 105, 309–335 (1983).

[3] Borel, A.; Serre, J.P.: *Corners and Arithmetic Groups.* Comment. Math. Helv. 48, 436–491 (1973).

[4] Borel, A.: *Stable real cohomology of arithmetic groups.* Ann. Sci. ENS 7, 235-272 (1974).

[5] Borel, A.; Wallach,N.: *Continuous Cohomology, Discrete Groups, and Representations of Reductive Groups.* Ann. Math. Stud. 94, Princeton 1980.

[6] Borel, A.: *Stable real cohomology of arithmetic groups. II.* Manifolds and Lie groups (Notre Dame, Ind., 1980), pp. 21–55, Progr. Math., 14, Birkhäuser, Boston, Mass., 1981.

[7] Bunke, U.; Olbrich, M.: *Γ-Cohomology and the Selberg Zeta Function* J. reine u. angew. Math. 467, 199-219 (1995).

[8] Bunke, U.; Olbrich, M.: *Cohomological properties of the canonical globalizations of Harish-Chandra modules. Consequences of theorems of Kashiwara-Schmid, Casselman, and Schneider-Stuhler.* Ann. Global Anal. Geom. 15 (1997), no. 5, 401–418.

[9] Bunke, U.; Olbrich, M.: *Resolutions of distribution globalizations of Harish-Chandra modules and cohomology.* J. Reine Angew. Math. 497, 47-81, (1998).

[10] Casselman, W.: *Canonical extensions of Harish-Chandra modules to representations of G.* Can. J. Math. 41, 385-438 (1989).

[11] Deitmar, A.: *Geometric zeta-functions of locally symmetric spaces.* Am. J. Math. 122, vol 5, 887-926 (2000).

[12] Deitmar, A.: *Selberg zeta functions for spaces of higher rank.* http://arXiv.org/abs/math.NT/0209383.

[13] Franke, J.: *Harmonic analysis in weighted L_2-spaces.* Ann. Sci. École Norm. Sup. (4) 31 (1998), no. 2, 181–279.

[14] Grothendieck, A.: *Produits tensoriels topologiques et espaces nucléaires.* Mem. Amer. Math. Soc. 1955 (1955), no. 16.

[15] Hecht, H.; Schmid, W.: *Characters, asymptotics and $\mathfrak{n}$-homology of Harish-Chandra modules.* Acta Math. 151, 49-151 (1983).

[16] Iwaniec, H.: *Spectral methods of automorphic forms. Second edition.* Graduate Studies in Mathematics, 53. American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002.

[17] Juhl, A.: *Cohomological theory of dynamical zeta functions.* Progress in Mathematics, 194. Birkhuser Verlag, Basel, 2001.

[18] Kashiwara, M.; Schmid, W.: *Quasi-equivariant D-modules, equivariant derived category, and representations of reductive Lie groups.* In: Lie Theory and Geometry. In Honor of Bertram Kostant, Progr. in Math., Birkhäuser, Boston. 457-488 (1994).

[19] Lang, S.: *Algebra.* Revised third edition. Graduate Texts in Mathematics, 211. Springer-Verlag, New York, 2002.

[20] Langlands, R.: *On the Functional Equations Satisfied by Eisenstein Series.* SLNM 544, 1976.

[21] Lewis, J.; Zagier, D.: *Period functions for Maass wave forms.* I. Ann. of Math. (2) 153 (2001), no. 1, 191–258.

[22] SCHAEFER, H.: *Topological vector spaces.* The Macmillan Co., New York; Collier-Macmillan Ltd., London 1966

[23] SCHMIDT, T.: *Rational Period Functions and Parabolic Cohomology.* J. Number Theory 57, 50-65 (1996).

[24] TREVES, F.: *Topological vector spaces, distributions and kenels.* Academic Press, New York 1967.

[25] ZAGIER, D.: *New Points of View on the Selberg Zeta Function.* Proceedings of the Japanese-German Seminar "Explicit structures of modular forms and zeta functions" Hakuba, Sept. 2001. Ryushi-do, 2002.

[26] ZIMMER, R.: *Ergodic Theory and Semisimple Groups* Birkhäuser 1984.

Anton Deitmar
Mathematisches Institut
Auf der Morgenstelle 10
72076 Tübingen, Germany
deitmar@uni-tuebingen.de

Joachim Hilgert
Universität Paderborn
Fakultät für Elektrotechnik
Informatik und Mathematik
Warburger Str. 100
33098 Paderborn, Germany
hilgert@math.uni-paderborn.de

Vanishing Geodesic Distance on Spaces of Submanifolds and Diffeomorphisms

Peter W. Michor, David Mumford

Received: March 18, 2005

Communicated by Eckard Meinrenken

Abstract. The L^2-metric or Fubini-Study metric on the non-linear Grassmannian of all submanifolds of type M in a Riemannian manifold (N, g) induces geodesic distance 0. We discuss another metric which involves the mean curvature and shows that its geodesic distance is a good topological metric. The vanishing phenomenon for the geodesic distance holds also for all diffeomorphism groups for the L^2-metric.

2000 Mathematics Subject Classification: Primary 58B20, 58D15, 58E12

1. Introduction

In [10] we studied the L^2-Riemannian metric on the space of all immersions $S^1 \to \mathbb{R}^2$. This metric is invariant under the group $\operatorname{Diff}(S^1)$ and we found that it induces *vanishing geodesic distance* on the quotient space $\operatorname{Imm}(S^1, \mathbb{R}^2)/\operatorname{Diff}(S^1)$. In this paper we extend this result to the general situation $\operatorname{Imm}(M, N)/\operatorname{Diff}(M)$ for any compact manifold M and Riemannian manifold (N, g) with $\dim N > \dim M$. On the open subset $\operatorname{Emb}(M, N)/\operatorname{Diff}(M)$, which may be identified with the space of all submanifolds of diffeomorphism type M in N (the non-linear Grassmanian or differentiable 'Chow' variety) this says that the infinite dimensional analog of the Fubini Study metric induces vanishing geodesic distance. The picture that emerges for these infinite-dimensional manifolds is quite interesting: there are simple expressions for the Christoffel symbols and curvature tensor, the geodesic equations are simple and of hyperbolic type and, at least in the case of plane curves, the geodesic spray exists locally. *But* the curvature is positive and unbounded in some high frequency directions, so these spaces wrap up on themselves arbitrarily tightly, allowing the infimum of path lengths between two given points to be zero.

We also carry over to the general case the stronger metric from [10] which weights the L^2 metric using the second fundamental form. It turns out that we have only to use the mean curvature in order to get positive geodesic distances, hence a good topological metric on the space $\operatorname{Emb}(S^1,\mathbb{R}^2)/\operatorname{Diff}(S^1)$. The reason is that the first variation of the volume of a submanifold depends on the mean curvature and the key step is showing that the square root of the volume of M is Lipschitz in our stronger metric. The formula for this metric is:

$$G^A_f(h,k) := \int_M (1 + A\|\operatorname{Tr}^{f^*g}(S^f)\|^2_{g^{N(f)}})g(h,k)\operatorname{vol}(f^*g)$$

where S^f is the second fundamental form of the immersion f; section 3 contains the relevant estimates. In section 4 we also compute the sectional curvature of the L^2-metric in the hope to relate the vanishing of the geodesic distance to unbounded positivity of the sectional curvature: by going through ever more positively curved parts of the space we can find ever shorter curves between any two submanifolds.

In the final section 5 we show that the vanishing of the geodesic distance also occurs on the Lie group of all diffeomorphisms on each connected Riemannian manifold. Short paths between any 2 diffeomorphisms are constructed by using rapidly moving compression waves in which individual points are trapped for relatively long times. We compute the sectional curvature also in this case.

2. The manifold of immersions

2.1. Conventions. Let M be a compact smooth connected manifold of dimension $m \geq 1$ and let (N,g) be a connected Riemannian manifold of dimension $n > m$. We shall use the following spaces and manifolds of smooth mappings.

$\operatorname{Diff}(M)$, the regular Lie group ([8], 38.4) of all diffeomorphisms of M.

$\operatorname{Diff}_{x_0}(M)$, the subgroup of diffeomorphisms fixing $x_0 \in M$.

$\operatorname{Emb} = \operatorname{Emb}(M,N)$, the manifold of all smooth embeddings $M \to N$.

$\operatorname{Imm} = \operatorname{Imm}(M,N)$, the manifold of all smooth immersions $M \to N$. For an immersion f the tangent space with foot point f is given by $T_f \operatorname{Imm}(M,N) = C^\infty_f(M,TN) = \Gamma(f^*TN)$, the space of all vector fields along f.

$\operatorname{Imm}_f = \operatorname{Imm}_f(M,N)$, the manifold of all smooth *free* immersions $M \to N$, i.e., those with trivial isotropy group for the right action of $\operatorname{Diff}(M)$ on $\operatorname{Imm}(M,N)$.

$B_e = B_e(M,N) = \operatorname{Emb}(M,N)/\operatorname{Diff}(M)$, the manifold of submanifolds of type M in N, the base of a smooth principal bundle, see 2.2.

$B_i = B_i(M,N) = \operatorname{Imm}(M,N)/\operatorname{Diff}(M)$, an infinite dimensional 'orbifold', whose points are, roughly speaking, smooth immersed submanifolds of type M in N, see 2.4.

$B_{i,f} = B_{i,f}(M,N) = \operatorname{Imm}_f(M,\mathbb{R}^2)/\operatorname{Diff}(M)$, a manifold, the base of a principal fiber bundle, see 2.3.

For a smooth curve $f : \mathbb{R} \to C^\infty(M,N)$ corresponding to a mapping $f : \mathbb{R} \times M \to N$, we shall denote by Tf the curve of tangent mappings, so that $Tf(t)(X_x) = T_x(f(t,\quad)).X_x$. The time derivative will be denoted by either $\partial_t f = f_t : \mathbb{R} \times M \to TN$.

2.2. The principal bundle of embeddings $\operatorname{Emb}(M,N)$. We recall some basic results whose proof can be found in [8]):

(A) The set $\operatorname{Emb}(M,N)$ *of all smooth embeddings* $M \to N$ *is an open subset of the smooth Fréchet manifold* $C^\infty(M,N)$ *of all smooth mappings* $M \to N$ *with the* C^∞*-topology. It is the total space of a smooth principal bundle* $\pi : \operatorname{Emb}(M,N) \to B_e(M,N)$ *with structure group* $\operatorname{Diff}(M)$*, the smooth regular Lie group group of all diffeomorphisms of* M*, whose base* $B_e(M,N)$ *is the smooth Fréchet manifold of all submanifolds of* N *of type* M*, i.e., the smooth manifold of all simple closed curves in* N*.* ([8], 44.1)

(B) This principal bundle admits a smooth principal connection described by the horizontal bundle whose fiber $\mathcal{N}_c$ *over* c *consists of all vector fields* h *along* f *such that* $g(h,Tf) = 0$*. The parallel transport for this connection exists and is smooth.* ([8], 39.1 and 43.1)

2.3. Free immersions. The manifold $\operatorname{Imm}(M,N)$ of all immersions $M \to N$ is an open set in the manifold $C^\infty(M,N)$ and thus itself a smooth manifold. An immersion $f : M \to N$ is called *free* if $\operatorname{Diff}(M)$ acts freely on it, i.e., $f \circ \varphi = c$ for $\varphi \in \operatorname{Diff}(M)$ implies $\varphi = \operatorname{Id}$. We have the following results:

- *If* $\varphi \in \operatorname{Diff}(M)$ *has a fixed point and if* $f \circ \varphi = f$ *for some immersion* f *then* $\varphi = \operatorname{Id}$. This is ([4], 1.3).
- *If for* $f \in \operatorname{Imm}(M,N)$ *there is a point* $x \in f(M)$ *with only one preimage then* f *is a free immersion.* This is ([4], 1.4). There exist free immersions without such points.
- The manifold $B_{i,f}(M,N)$ ([4], 1.5) *The set* $\operatorname{Imm}_f(M,N)$ *of all free immersions is open in* $C^\infty(M,N)$ *and thus a smooth submanifold. The projection*

$$\pi : \operatorname{Imm}_f(M,N) \to \frac{\operatorname{Imm}_f(M,N)}{\operatorname{Diff}(M)} =: B_{i,f}(M,N)$$

  onto a Hausdorff smooth manifold is a smooth principal fibration with structure group $\operatorname{Diff}(M)$*. By* ([8], 39.1 and 43.1) *this fibration admits a smooth principal connection described by the horizontal bundle with* *fiber $\mathcal{N}_c$ consisting of all vector fields h along f such that $g(h, Tf) = 0$. This connection admits a smooth parallel transport over each smooth curve in the base manifold.*

We might view $\mathrm{Imm}_f(M,N)$ as the nonlinear Stiefel manifold of parametrized submanifolds of type M in N and consequently $B_{i,f}(M,N)$ as the nonlinear Grassmannian of unparametrized submanifolds of type M.

2.4. NON FREE IMMERSIONS. Any immersion is proper since M is compact and thus by ([4], 2.1) the orbit space $B_i(M,N) = \mathrm{Imm}(M,N)/\operatorname{Diff}(M)$ is Hausdorff. Moreover, by ([4], 3.1 and 3.2) for any immersion f the isotropy group $\operatorname{Diff}(M)_f$ is a finite group which acts as group of covering transformations for a finite covering $q_c : M \to \bar{M}$ such that f factors over q_c to a free immersion $\bar{f} : \bar{M} \to N$ with $\bar{f} \circ q_c = f$. Thus the subgroup $\operatorname{Diff}_{x_0}(M)$ of all diffeomorphisms φ fixing $x_0 \in M$ acts freely on $\mathrm{Imm}(M,N)$. Moreover, for each $f \in \mathrm{Imm}$ the submanifold $\mathcal{Q}(f)$ from 4.4, (1) is a slice in a strong sense:

- $\mathcal{Q}(f)$ is invariant under the isotropy group $\operatorname{Diff}(M)_f$.
- If $\mathcal{Q}(f)\circ\varphi\cap\mathcal{Q}(f) \neq \emptyset$ for $\varphi \in \operatorname{Diff}(M)$ then φ is already in the isotropy group $\varphi \in \operatorname{Diff}(M)_f$.
- $\mathcal{Q}(f) \circ \operatorname{Diff}(M)$ is an invariant open neigbourhood of the orbit $f \circ \operatorname{Diff}(M)$ in $\mathrm{Imm}(M,N)$ which admits a smooth retraction r onto the orbit. The fiber $r^{-1}(f \circ \varphi)$ equals $\mathcal{Q}(f \circ \varphi)$.

Note that also the action

$$\mathrm{Imm}(M,N) \times \operatorname{Diff}(M) \to \mathrm{Imm}(M,N) \times \mathrm{Imm}(M,N), \qquad (f,\varphi) \mapsto (f, f\circ\varphi)$$

is proper so that all assumptions and conclusions of Palais' slice theorem [13] hold. This results show that the orbit space $B_i(M,N)$ has only singularities of orbifold type times a Fréchet space. We may call the space $B_i(M,N)$ an infinite dimensional *orbifold*. The projection $\pi : \mathrm{Imm}(M,N) \to B_i(M,N) = \mathrm{Imm}(M,N)/\operatorname{Diff}(M)$ is a submersion off the singular points and has only mild singularities at the singular strata. The normal bundle $\mathcal{N}_f$ mentioned in 2.2 is well defined and is a smooth vector subbundle of the tangent bundle. We do not have a principal bundle and thus no principal connections, but we can prove the main consequence, the existence of horizontal paths, directly:

2.5. PROPOSITION. *For any smooth path f in $\mathrm{Imm}(M,N)$ there exists a smooth path φ in $\operatorname{Diff}(M)$ with $\varphi(t, \quad) = \mathrm{Id}_M$ depending smoothly on f such that the path h given by $h(t,\theta) = f(t,\varphi(t,\theta))$ is horizontal: $g(h_t, Th) = 0$.*

PROOF. Let us write $h = f \circ \varphi$ for $h(t,x) = f(t,\varphi(t,x))$, etc. We look for φ as the integral curve of a time dependent vector field $\xi(t,x)$ on M, given by $\partial_t\varphi = \xi \circ \varphi$. We want the following expression to vanish:

$$g\big(\partial_t(f\circ\varphi), T(f\circ\varphi)\big) = g\big((\partial_t f\circ\varphi + (Tf\circ\varphi).\partial_t\varphi, (Tf\circ\varphi).T\varphi\big)$$

$$= \big(g(\partial_t f, Tf)\circ\varphi\big).T\varphi + g\big((Tf\circ\varphi)(\xi\circ\varphi),(Tf\circ\varphi).T\varphi\big)$$
$$= \big(\big(g(\partial_t f, Tf) + g(Tf.\xi, Tf)\big)\circ\varphi\big).T\varphi$$

Since $T\varphi$ is everywhere invertible we get

$$0 = g\big(\partial_t(f\circ\varphi), T(f\circ\varphi)\big) \iff 0 = g(\partial_t f, Tf) + g(Tf.\xi, Tf)$$

and the latter equation determines the non-autonomous vector field ξ uniquely. □

2.6. Curvatures of an immersion. Consider a fixed immersion $f \in \mathrm{Imm}(M,N)$. The *normal bundle* $N(f) = Tf^\perp \subset f^*TN \to M$ has fibers $N(f)_x = \{Y \in T_{f(x)}N : g(Y, T_xf.X) = 0 \text{ for all } X \in T_xM\}$. Every vector field $h : M \to TN$ along f then splits as $h = Tf.h^\top + h^\perp$ into its tangential component $h^\top \in \mathfrak{X}(M)$ and its normal component $h^\perp \in \Gamma(N(f))$.

Let ∇^g be the Levi-Civita covariant derivative of g on N and let ∇^{f^*g} the Levi-Civita covariant derivative of the pullback metric f^*g on M. The *shape operator* or *second fundamental form* $S^f \in \Gamma(S^2T^*M \otimes N(f))$ of f is then given by

$$S^f(X,Y) = \nabla^g_X(Tf.Y) - Tf.\nabla^{f^*g}_X Y \quad \text{for } X, Y \in \mathfrak{X}(M). \tag{1}$$

It splits into the following irreducible components under the action of the group $O(T_xM)\times O(N(f)_x)$: the mean curvature $\mathrm{Tr}^{f^*g}(S^f) = \mathrm{Tr}((f^*g)^{-1}\circ S^f) \in \Gamma(N(f))$ and the trace free shape operator $S^f_0 = S^f - \mathrm{Tr}^{f^*g}(S^f)$. For $X \in \mathfrak{X}(M)$ and $\xi \in \Gamma(N(f))$, i.e., a normal vector field along f, we may also split $\nabla^g_X\xi$ into the components which are tangential and normal to $Tf.TM$,

$$\nabla^g_X\xi = -Tf.L^f_\xi(X) + \nabla^{N(f)}_X\xi \tag{2}$$

where $\nabla^{N(f)}$ is the induced connection in the normal bundle respecting the metric $g^{N(f)}$ induced by g, and where the *Weingarten* tensor field $L^f \in \Gamma(N(f)^*\otimes T^*M\otimes TM)$ corresponds to the shape operator via the formula

$$(f^*g)(L^f_\xi(X), Y) = g^{N(f)}(S^f(X,Y),\xi). \tag{3}$$

Let us also split the Riemann curvature R^g into tangential and normal parts: For $X_i \in \mathfrak{X}(M)$ or T_xM we have (*theorema egregium*):

$$\begin{aligned} &g(R^g(Tf.X_1, Tf.X_2)(Tf.X_3), Tf.X_4) = (f^*g)(R^{f^*g}(X_1,X_2)X_3, X_4)+\\ &\quad + g^{N(f)}(S^f(X_1,X_3), S^f(X_2,X_4)) - g^{N(f)}(S^f(X_2,X_3), S^f(X_1,X_4)). \end{aligned}\tag{4}$$

The normal part of R^g is then given by (*Codazzi-Mainardi equation*):

$$\begin{aligned} &(R^g(Tf.X_1, Tf.X_2)(Tf.X_3))^\perp =\\ &\quad = \big(\nabla^{N(f)\otimes T^*M\otimes T^*M}_{X_1} S^f\big)(X_2,X_3) - \big(\nabla^{N(f)\otimes T^*M\otimes T^*M}_{X_2} S^f\big)(X_1,X_3). \end{aligned}\tag{5}$$

2.7. VOLUMES OF AN IMMERSION. For an immersion $f \in \operatorname{Imm}(M,N)$, we consider the volume density $\operatorname{vol}^g(f) = \operatorname{vol}(f^*g) \in \operatorname{Vol}(M)$ on M given by the local formula $\operatorname{vol}^g(f)|_U = \sqrt{\det((f^*g)_{ij})}|du^1 \wedge \dots \wedge du^m|$ for any chart $(U, u : U \to \mathbb{R}^m)$ of M, and the induced volume function $\operatorname{Vol}^g : \operatorname{Imm}(M,N) \to \mathbb{R}_{>0}$ which is given by $\operatorname{Vol}^g(f) = \int_M \operatorname{vol}(f^*g)$. The tangent mapping of $\operatorname{vol} : \Gamma(S^2_{>0}T^*M) \to \operatorname{Vol}(M)$ is given by $d\operatorname{vol}(\gamma)(\eta) = \frac{1}{2}\operatorname{Tr}(\gamma^{-1}.\eta)\operatorname{vol}(\gamma)$. We consider the pullback mapping $P_g : f \mapsto f^*g, P_g : \operatorname{Imm}(M,N) \to \Gamma(S^2_{>0}T^*M)$. A version of the following lemma is [7], 1.6.

LEMMA. *The derivative of* $\operatorname{vol}^g = \operatorname{vol} \circ P_g : \operatorname{Imm}(M,N) \to \operatorname{Vol}(M)$ *is*

$$\begin{aligned} d\operatorname{vol}^g(h) &= d(\operatorname{vol} \circ P_g)(h) \\ &= -\operatorname{Tr}^{f^*g}(g(S^f, h^\perp))\operatorname{vol}(f^*g) + \tfrac{1}{2}\operatorname{Tr}^{f^*g}(\mathcal{L}_{h^\top}(f^*g))\operatorname{vol}(f^*g). \\ &= -(g(\operatorname{Tr}^{f^*g}(S^f), h^\perp))\operatorname{vol}(f^*g) + \operatorname{div}^{f^g}(h^\top)\operatorname{vol}(f^*g). \end{aligned}$$

PROOF. We consider a curve $t \mapsto f(t, \quad)$ in Imm with $\partial_t|_0 f = h$. We also use a chart $(U, u : U \to \mathbb{R}^m)$ on M. Then we have

$$f^*g|_U = \sum_{i,j}(f^*g)_{ij}du^i \otimes du^j = \sum_{i,j} g(Tf.\partial_{u_i}, Tf.\partial_{u^j})du^i \otimes du^j$$

$$\partial_t \operatorname{vol}^g(f)|_U = \frac{\det((f^*g)_{ij})(f^*g)^{kl}\partial_t(f^*g)_{lk}}{2\sqrt{\det((f^*g)_{ij})}}|du^1 \wedge \dots \wedge du^m|$$

where

$$\begin{aligned} \partial_t(f^*g)_{ij} &= \partial_t g(Tf.\partial_{u^i}, Tf.\partial_{u^j}) \\ &= g(\nabla^g_{\partial_t}(Tf.\partial_{u^i}), Tf.\partial_{u^j}) + g(\partial_{u^i}, \nabla^g_{\partial_t}(Tf.\partial_{u^j})) \\ g(\nabla^g_{\partial_t}Tf.\partial_{u^i}, Tf.\partial_{u^j}) &= g(\nabla^g_{\partial_{u^i}}Tf.\partial_t + Tf.\operatorname{Tor} + Tf.[\partial_t, \partial_{u^i}], Tf.\partial_{u^j}) \\ &= g(\nabla^g_{\partial_{u^i}}(Tf.\partial_t)^\perp, Tf.\partial_{u^j}) + g(\nabla^g_{\partial_{u^i}}(Tf.\partial_t)^\top, Tf.\partial_{u^j}) \\ g(\nabla^g_{\partial_{u^i}}(\partial_t f)^\perp, Tf.\partial_{u^j}) &= g(-Tf.L_{(\partial_t f)^\perp}\partial_{u^i}, Tf.\partial_{u^j}) + g(\nabla^{N(f)}_{\partial_{u^i}}(\partial_t f)^\perp, Tf.\partial_{u^j}) \\ &= -(f^*g)(L_{(\partial_t f)^\perp}\partial_{u^i}, \partial_{u^j}) \\ &= -g(S^f(\partial_{u^i}, \partial_{u^j}), (\partial_t f)^\perp) \\ g(\nabla^g_{\partial_{u^i}}(\partial_t f)^\top, Tf.\partial_{u^j}) &= (f^*g)(\nabla^{f^*g}_{\partial_{u^i}}(\partial_t f)^\top, \partial_{u^j}) + 0 \\ &= (f^*g)(\nabla^{f^*g}_{(\partial_t f)^\top}\partial_{u^i} + \operatorname{Tor} - [(\partial_t f)^\top, \partial_{u^i}], \partial_{u^j}), \\ \partial_t(f^*g)_{ij} &= -2g(S^f(\partial_{u^i}, \partial_{u^j}), (\partial_t f)^\perp) + (\mathcal{L}_{(\partial_t f)^\top}(f^*g))(\partial_{u^i}, \partial_{u^j}) \end{aligned}$$

This proves the first formula. For the second one note that

$$\tfrac{1}{2}\operatorname{Tr}((f^*g)^{-1}\mathcal{L}_{h^\top}(f^*g))\operatorname{vol}(f^*g) = \mathcal{L}_{h^\top}(\operatorname{vol}(f^*g)) = \operatorname{div}^{f^*g}(h^\top)\operatorname{vol}(f^*g). \quad \square$$

3. Metrics on spaces of mappings

3.1. The metric G^A. Let $h, k \in C^\infty_f(M, TN)$ be two tangent vectors with foot point $f \in \operatorname{Imm}(M, N)$, i.e., vector fields along f. Let the induced volume density be $\operatorname{vol}(f^*g)$. We consider the following weak Riemannian metric on $\operatorname{Imm}(M, N)$, for a constant $A \geq 0$:

$$G^A_f(h, k) := \int_M \Big(1 + A\|\operatorname{Tr}^{f^*g}(S^f)\|^2_{g^{N(f)}}\Big)\, g(h, k) \operatorname{vol}(f^*g)$$

where $\operatorname{Tr}^{f^*g}(S^f) \in N(f)$ is the mean curvature, a section of the normal bundle, and $\|\operatorname{Tr}^{f^*g}(S^f)\|_{g^{N(f)}}$ is its norm. The metric G^A is invariant for the action of $\operatorname{Diff}(M)$. This makes the map $\pi : \operatorname{Imm}(M, N) \to B_i(M, N)$ into a *Riemannian submersion* (off the singularities of $B_i(M, N)$).

Now we can determine the bundle $\mathcal{N} \to \operatorname{Imm}(M, N)$ of tangent vectors which are normal to the $\operatorname{Diff}(M)$-orbits. The tangent vectors to the orbits are $T_f(f \circ \operatorname{Diff}(M)) = \{Tf.\xi : \xi \in \mathfrak{X}(M)\}$. Inserting this for k into the expression of the metric G we see that

$$\begin{aligned}\mathcal{N}_f &= \{h \in C^\infty(M, TN) : g(h, Tf) = 0\}\\ &= \Gamma(N(f)),\end{aligned}$$

the space of sections of the normal bundle. This is independent of A.

A tangent vector $h \in T_f \operatorname{Imm}(M, N) = C^\infty_f(M, TN) = \Gamma(f^*TN)$ has an orthonormal decomposition

$$h = h^\top + h^\perp \in T_f(f \circ \operatorname{Diff}^+(M)) \oplus \mathcal{N}_f$$

into smooth tangential and normal components.

Since the Riemannian metric G^A on $\operatorname{Imm}(M, N)$ is invariant under the action of $\operatorname{Diff}(M)$ it induces a metric on the quotient $B_i(M, N)$ as follows. For any $F_0, F_1 \in B_i$, consider all liftings $f_0, f_1 \in \operatorname{Imm}$ such that $\pi(f_0) = F_0, \pi(f_1) = F_1$ and all smooth curves $t \mapsto f(t, \quad)$ in $\operatorname{Imm}(M, N)$ with $f(0, \cdot) = f_0$ and $f(1, \cdot) = f_1$. Since the metric G^A is invariant under the action of $\operatorname{Diff}(M)$ the arc-length of the curve $t \mapsto \pi(f(t, \cdot))$ in $B_i(M, N)$ is given by

$$\begin{aligned}L^{\text{hor}}_{G^A}(f) &:= L_{G^A}(\pi(f(t, \cdot))) =\\ &= \int_0^1 \sqrt{G^A_{\pi(f)}(T_f\pi.f_t, T_f\pi.f_t)}\, dt = \int_0^1 \sqrt{G^A_f(f_t^\perp, f_t^\perp)}\, dt =\\ &= \int_0^1 \Big(\int_M (1 + A\|\operatorname{Tr}^{f^*g}(S^f)\|^2_{g^{N(f)}}) g(f_t^\perp, f_t^\perp) \operatorname{vol}(f^*g)\Big)^{\frac{1}{2}} dt\end{aligned}$$

In fact the last computation only makes sense on $B_{i,f}(M, N)$ but we take it as a motivation. The metric on $B_i(M, N)$ is defined by taking the infimum of this over all paths f (and all lifts f_0, f_1):

$$\operatorname{dist}_{G^A}^{B_i}(F_1, F_2) = \inf_f L_{G^A}^{\text{hor}}(f).$$

3.2. THEOREM. *Let $A = 0$. For $f_0, f_1 \in \operatorname{Imm}(M, N)$ there exists always a path $t \mapsto f(t, \cdot)$ in $\operatorname{Imm}(M, N)$ with $f(0, \cdot) = f_0$ and $\pi(f(1, \cdot)) = \pi(f_1)$ such that $L_{G^0}^{hor}(f)$ is arbitrarily small.*

PROOF. Take a path $f(t, \theta)$ in $\operatorname{Imm}(M, N)$ from f_0 to f_1 and make it horizontal using 2.4 so that that $g(f_t, Tf) = 0$; this forces a reparametrization on f_1.

Let $\alpha : M \to [0, 1]$ be a surjective Morse function whose singular values are all contained in the set $\{\frac{k}{2N} : 0 \le k \le 2N\}$ for some integer N. We shall use integers n below and we shall use only multiples of N.

Then the level sets $M_r := \{x \in M : \alpha(x) = r\}$ are of Lebesque measure 0. We shall also need the slices $M_{r_1,r_2} := \{x \in M : r_1 \le \alpha(x) \le r_2\}$. Since M is compact there exists a constant C such that the following estimate holds uniformly in t:

$$\int_{M_{r_1,r_2}} \operatorname{vol}(f(t, \quad)^* g) \le C(r_2 - r_1) \int_M \operatorname{vol}(f(t, \quad)^* g)$$

Let $\tilde f(t, x) = f(\varphi(t, \alpha(x)), x)$ where $\varphi : [0, 1] \times [0, 1] \to [0, 1]$ is given as in [10], 3.10 (which also contains a figure illustrating the construction) by

$$\varphi(t,\alpha)=\begin{cases} 2t(2n\alpha - 2k) & \text{for } \ 0 \le t \le 1/2,\ \frac{2k}{2n} \le \alpha \le \frac{2k+1}{2n} \\ 2t(2k + 2 - 2n\alpha) & \text{for } \ 0 \le t \le 1/2,\ \frac{2k+1}{2n} \le \alpha \le \frac{2k+2}{2n} \\ 2t - 1 + 2(1 - t)(2n\alpha - 2k) & \text{for } \ 1/2 \le t \le 1,\ \frac{2k}{2n} \le \alpha \le \frac{2k+1}{2n} \\ 2t - 1 + 2(1 - t)(2k + 2 - 2n\alpha) & \text{for } \ 1/2 \le t \le 1,\ \frac{2k+1}{2n} \le \alpha \le \frac{2k+2}{2n}. \end{cases}$$

Then we get $T\tilde f = \varphi_\alpha . d\alpha . f_t + Tf$ and $\tilde f_t = \varphi_t . f_t$ where

$$\varphi_\alpha = \begin{cases} +4nt \\ -4nt \\ +4n(1-t) \\ -4n(1-t) \end{cases}, \qquad \varphi_t = \begin{cases} 4n\alpha - 4k \\ 4k + 4 - 4n\alpha \\ 2 - 4n\alpha + 4k \\ -(2 - 4n\alpha + 4k) \end{cases}.$$

We use horizontality $g(f_t, Tf) = 0$ to determine $\tilde f_t^\perp = \tilde f_t + T\tilde f(X)$ where $X \in TM$ satisfies $0 = g(\tilde f_t + T\tilde f(X), T\tilde f(\xi))$ for all $\xi \in TM$. We also use

$$d\alpha(\xi) = f^* g(\operatorname{grad}^{f^* g} \alpha, \xi) = g(Tf(\operatorname{grad}^{f^* g} \alpha), Tf(\xi))$$

and get

$$\begin{aligned} 0 &= g(\tilde f_t + T\tilde f(X), T\tilde f(\xi)) \\ &= g\Big(\varphi_t f_t + \varphi_\alpha d\alpha(X) f_t + Tf(X), \varphi_\alpha d\alpha(\xi) f_t + Tf(\xi)\Big) \\ &= \varphi_t . \varphi_\alpha . (f^* g)(\operatorname{grad}^{f^* g} \alpha, \xi) \|f_t\|_g^2 + \end{aligned}$$

$$+\varphi_\alpha^2.(f^*g)(\operatorname{grad}^{f^*g}\alpha,X).(f^*g)(\operatorname{grad}^{f^*g}\alpha,\xi)\|f_t\|_g^2+g(Tf(X),Tf(\xi))$$
$$=(\varphi_t.\varphi_\alpha+\varphi_\alpha^2.(f^*g)(\operatorname{grad}^{f^*g}\alpha,X))\|f_t\|_g^2(f^*g)(\operatorname{grad}^{f^*g}\alpha,\xi)+(f^*g)(X,\xi)$$

This implies that $X=\lambda\operatorname{grad}^{f^*g}\alpha$ for a function λ and in fact we get

$$\tilde f_t^\perp=\frac{\varphi_t}{1+\varphi_\alpha^2\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}f_t-\frac{\varphi_t\varphi_\alpha\|f_t\|_g^2}{1+\varphi_\alpha^2\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}Tf(\operatorname{grad}^{f^*g}\alpha)$$

and

$$\|\tilde f_t\|_g^2=\frac{\varphi_t^2\|f_t\|_g^2}{1+\varphi_\alpha^2\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}$$

From $T\tilde f=\varphi_\alpha.d\alpha.f_t+Tf$ and $g(f_t,Tf)=0$ we get for the volume form

$$\operatorname{vol}(\tilde f^*g)=\sqrt{1+\varphi_\alpha^2\,\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}\,\operatorname{vol}(f^*g).$$

For the horizontal length we get

$$\begin{aligned}
L^{\text{hor}}(\tilde f)&=\int_0^1\Big(\int_M\|\tilde f_t^\perp\|_g^2\operatorname{vol}(\tilde f^*g)\Big)^{\frac12}dt=\\
&=\int_0^1\Big(\int_M\frac{\varphi_t^2\|f_t\|_g^2}{\sqrt{1+\varphi_\alpha^2\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}}\operatorname{vol}(f^*g)\Big)^{\frac12}dt=\\
&=\int_0^{\frac12}\Bigg(\sum_{k=0}^{n-1}\Big(\int_{M_{\frac{2k}{2n},\frac{2k+1}{2n}}}\frac{(4n\alpha-4k)^2\|f_t\|_g^2}{\sqrt{1+(4nt)^2\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}}\operatorname{vol}(f^*g)+\\
&\qquad+\int_{M_{\frac{2k+1}{2n},\frac{2k+2}{2n}}}\frac{(4k+4-4n\alpha)^2\|f_t\|_g^2}{\sqrt{1+(4nt)^2\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}}\operatorname{vol}(f^*g)\Big)\Bigg)^{\frac12}dt+\\
&\quad+\int_{\frac12}^1\Bigg(\sum_{k=0}^{n-1}\Big(\int_{M_{\frac{2k}{2n},\frac{2k+1}{2n}}}\frac{(2-4n\alpha+4k)^2\|f_t\|_g^2}{\sqrt{1+(4n(1-t))^2\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}}\operatorname{vol}(f^*g)+\\
&\qquad+\int_{M_{\frac{2k+1}{2n},\frac{2k+2}{2n}}}\frac{(2-4n\alpha+4k)^2\|f_t\|_g^2}{\sqrt{1+(4n(1-t))^2\|d\alpha\|_{f^*g}^2\|f_t\|_g^2}}\operatorname{vol}(f^*g)\Big)\Bigg)^{\frac12}dt
\end{aligned}$$

Let $\varepsilon>0$. The function $(t,x)\mapsto\|f_t(\varphi(t,\alpha(x)),x)\|_g^2$ is uniformly bounded. On $M_{\frac{2k}{2n},\frac{2k+1}{2n}}$ the function $4n\alpha-4k$ has values in $[0,2]$. Choose disjoint geodesic balls centered at the finitely many singular values of the Morse function α of total f^*g-volume $<\varepsilon$. Restricted to the union M_{sing} of these balls the integral above is $O(1)\varepsilon$. So we have to estimate the integrals on the complement $\tilde M=M\setminus M_{\text{sing}}$ where the function $\|d\alpha\|_{f^*g}$ is uniformly bounded from below by $\eta>0$.

Let us estimate one of the sums above. We use the fact that the singular points of the Morse function α lie all on the boundaries of the sets $\tilde M_{\frac{2k}{2n},\frac{2k+1}{2n}}$ so that we can transform the integrals as follows:

$$\begin{aligned}\sum_{k=0}^{n-1}\int_{\tilde M_{\frac{2k}{2n},\frac{2k+1}{2n}}} &\frac{(4n\alpha-4k)^2\|f_t\|^2_g}{\sqrt{1+(4nt)^2\|d\alpha\|^2_{f^*g}\|f_t\|^2_g}}\operatorname{vol}(f^*g)=\\ =\sum_{k=0}^{n-1}\int_{\frac{2k}{2n}}^{\frac{2k+1}{2n}}\int_{\tilde M_r}&\frac{(4nr-4k)^2\|f_t\|^2_g}{\sqrt{1+(4nt)^2\|d\alpha\|^2_{f^*g}\|f_t\|^2_g}}\frac{\operatorname{vol}(i_r^*f^*g)}{\|d\alpha\|_{f^*g}}\,dr\end{aligned}$$

We estimate this sum of integrals: Consider first the set of all $(t,r,x\in M_r)$ such that $|f_t(\varphi(t,r),x)|<\varepsilon$. There we estimate by

$$O(1).n.16n^2.\varepsilon^2.(r^3/3)|_{r=0}^{r=1/2n}=O(\varepsilon).$$

On the complementary set where $|f_t(\varphi(t,r),x)|\ge\varepsilon$ we estimate by

$$O(1).n.16n^2.\frac{1}{4nt\eta^2\varepsilon}(r^3/3)|_{r=0}^{r=1/2n}=O(\frac{1}{nt\eta^2\varepsilon})$$

which goes to 0 if n is large enough. The other sums of integrals can be estimated similarly, thus $L^{\text{hor}}(\tilde f)$ goes to 0 for $n\to\infty$. It is clear that one can approximate φ by a smooth function whithout changing the estimates essentially. $\square$

3.3. A Lipschitz bound for the volume in G^A. We apply the Cauchy-Schwarz inequality to the derivative 2.7 of the volume $\operatorname{Vol}^g(f)$ along a curve $t\mapsto f(t,\quad)\in\operatorname{Imm}(M,N)$:

$$\begin{aligned}\partial_t\operatorname{Vol}^g(f)&=\partial_t\int_M\operatorname{vol}^g(f(t,\quad))=\int_M d\operatorname{vol}^g(f)(\partial_tf)\\ &=-\int_M\operatorname{Tr}^{f^*g}(g(S^f,f_t^\perp))\operatorname{vol}(f^*g)\le\Big|\int_M\operatorname{Tr}^{f^*g}(g(S^f,f_t^\perp))\operatorname{vol}(f^*g)\Big|\\ &\le\Big(\int_M 1^2\operatorname{vol}(f^*g)\Big)^{\frac12}\Big(\int_M\operatorname{Tr}^{f^*g}(g(S^f,f_t^\perp))^2\operatorname{vol}(f^*g)\Big)^{\frac12}\\ &\le\operatorname{Vol}^g(f)^{\frac12}\frac{1}{\sqrt A}\Big(\int_M(1+A\|\operatorname{Tr}^{f^*g}(S^f)\|^2_{g^{N(f)}})\,g(f_t^\perp,f_t^\perp)\operatorname{vol}(f^*g)\Big)^{\frac12}\end{aligned}$$

Thus

$$\begin{aligned}\partial_t(\sqrt{\operatorname{Vol}^g(f)})&=\frac{\partial_t\operatorname{Vol}^g(f)}{2\sqrt{\operatorname{Vol}^g(f)}}\le\\ &\le\frac{1}{2\sqrt A}\Big(\int_M(1+A\|\operatorname{Tr}^{f^*g}(S^f)\|^2_{g^{N(f)}})\,g(f_t^\perp,f_t^\perp)\operatorname{vol}(f^*g)\Big)^{\frac12}\end{aligned}$$

and by using (3.1) we get

$$\begin{aligned}
\sqrt{\operatorname{Vol}^g(f_1)} - \sqrt{\operatorname{Vol}^g(f_0)} &= \int_0^1 \partial_t(\sqrt{\operatorname{Vol}^g(f)})\,dt \\
&\le \frac{1}{2\sqrt{A}} \int_0^1 \Big(\int_M (1 + A\|\operatorname{Tr}^{f^*g}(S^f)\|^2_{g^{N(f)}})\, g(f_t^\perp, f_t^\perp)\operatorname{vol}(f^*g)\Big)^{\frac12} dt \\
&= \frac{1}{2\sqrt{A}} L^{\mathrm{hor}}_{G^A}(f).
\end{aligned}$$

If we take the infimum over all curves connecting f_0 with the $\operatorname{Diff}(M)$-orbit through f_1 we get:

Proposition. Lipschitz continuity of $\sqrt{\operatorname{Vol}^g} : B_i(M,N) \to \mathbb{R}_{\ge 0}$. *For F_0 and F_1 in $B_i(M,N) = \operatorname{Imm}(M,N)/\operatorname{Diff}(M)$ we have for $A > 0$:*

$$\sqrt{\operatorname{Vol}^g(F_1)} - \sqrt{\operatorname{Vol}^g(F_0)} \le \frac{1}{2\sqrt{A}} \operatorname{dist}^{B_i}_{G^A}(F_1, F_2).$$

3.4. Bounding the area swept by a path in B_i. We want to bound the area swept out by a curve starting from F_0 to any immersed submanifold F_1 nearby in our metric. First we use the Cauchy-Schwarz inequality in the Hilbert space $L^2(M, \operatorname{vol}(f(t,\quad)^*g))$ to get

$$\begin{aligned}
&\int_M 1.\|f_t\|_g \operatorname{vol}(f^*g) = \langle 1, \|f_t\|_g\rangle_{L^2} \le \\
&\qquad\qquad \le \|1\|_{L^2}\|c_t\|_{L^2} = \Big(\int_M \operatorname{vol}(f^*g)\Big)^{\frac12}\Big(\int_M |f_t| \operatorname{vol}(f^*g)\Big)^{\frac12}.
\end{aligned}$$

Now we assume that the variation $f(t,x)$ is horizontal, so that $g(f_t, Tf) = 0$. Then $L_{G^A}(f) = L^{\mathrm{hor}}_{G^A}(f)$. We use this inequality and then the intermediate value theorem of integral calculus to obtain

$$\begin{aligned}
L^{\mathrm{hor}}_{G^A}(f) = L_{G^A}(f) &= \int_0^1 \sqrt{G^A_f(f_t, f_t)}\,dt \\
&= \int_0^1 \Big(\int_M (1 + A\|\operatorname{Tr}^{f^*}(S^f)\|^2_{f^*g})\|f_t\|^2 \operatorname{vol}(f^*g)\Big)^{\frac12} dt \\
&\ge \int_0^1 \Big(\int_M \|f_t\|^2 \operatorname{vol}(f^*g)\Big)^{\frac12} dt \\
&\ge \int_0^1 \Big(\int_M \operatorname{vol}(f(t,\quad)^*g)\Big)^{-\frac12} \int_M \|f_t(t,\quad)\|_g \operatorname{vol}(f(t,\quad)^*g)\,dt \\
&= \Big(\int_M \operatorname{vol}(f(t_0,\quad)^*g)\Big)^{-\frac12} \int_0^1 \int_M \|f_t(t,\quad)\|_g \operatorname{vol}(f(t,\quad)^*g)\,dt \\
&\qquad\qquad \text{for some intermediate value } 0 \le t_0 \le 1, \\
&\ge \frac{1}{\sqrt{\operatorname{Vol}^g(f(t_0,\quad))}} \int_{[0,1]\times M} \operatorname{vol}^{m+1}(f^*g)
\end{aligned}$$

PROPOSITION. AREA SWEPT OUT BOUND. *If f is any path from F_0 to F_1, then*

$$\begin{pmatrix} (m+1) - \textit{volume of the region swept} \\ \textit{out by the variation } f \end{pmatrix} \le \max_t \sqrt{\mathrm{Vol}^g(f(t,\quad))} \cdot L^{hor}_{G^A}(f).$$

Together with the Lipschitz continuity 3.3 this shows that the geodesic distance $\inf L^{B_i}_{G^A}$ separates points, at least in the base space $B(M,N)$ of embeddings.

3.5. HORIZONTAL ENERGY OF A PATH AS ANISOTROPIC VOLUME. We consider a path $t \mapsto f(t,\quad)$ in $\mathrm{Imm}(M,N)$. It projects to a path $\pi \circ f$ in B_i whose energy is:

$$\begin{aligned} E_{G^A}(\pi \circ f) &= \tfrac{1}{2}\int_a^b G^A_{\pi(f)}(T\pi.f_t, T\pi.f_t)\,dt = \tfrac{1}{2}\int_a^b G^A_f(f_t^\perp, f_t^\perp)\,dt = \\ &= \tfrac{1}{2}\int_a^b \int_M (1 + A\|\mathrm{Tr}^{f^*g}(S^f)\|^2_{g^{N(f)}})g(f_t^\perp, f_t^\perp)\,\mathrm{vol}(f^*g)\,dt. \end{aligned}$$

We now consider the graph $\gamma_f : [a,b] \times M \ni (t,x) \mapsto (t, f(t,x)) \in [a,b] \times N$ of the path f and its image Γ_f, an immersed submanifold with boundary of $\mathbb{R} \times N$. We want to describe the horizontal energy as a functional on the space of immersed submanifolds with fixed boundary, remembering the fibration of $\mathrm{pr}_1 : \mathbb{R} \times N \to \mathbb{R}$. We get:

$$\begin{aligned} &E_{G^A}(\pi \circ f) = \\ &= \tfrac{1}{2}\int_{[a,b]\times M} \Big(1 + A\|\mathrm{Tr}^{f^*g}(S^f)\|^2_{g^{N(f)}}\Big) \frac{\|f_t^\perp\|^2}{\sqrt{1 + \|f_t^\perp\|^2_g}}\,\mathrm{vol}(\gamma_f^*(dt^2 + g)) \end{aligned}$$

Now $\|f_t^\perp\|_g$ depends only on the graph Γ_f and on the fibration over time, since any reparameterization of Γ_f which respects the fibration over time is of the form $(t,x) \mapsto (t, f(t, \varphi(t,x)))$ for some path φ in $\mathrm{Diff}(M)$ starting at the identity, and $(\partial_t|_0 f(t, \varphi(t,x)))^\perp = f_t^\perp$. So the above expression is intrinsic for the graph Γ_f and the fibration. In order to find a geodesic from the shape $\pi(f(a,\quad))$ to the shape $\pi(f(b,\quad))$ one has to find an immersed surface which is a critical point for the functional E_{G^A} above. This is a Plateau-problem with anisotropic volume.

4. THE GEODESIC EQUATION AND THE CURVATURE ON B_i

4.1. THE GEODESIC EQUATION OF G^0 IN $\mathrm{Imm}(M,N)$. The energy of a curve $t \mapsto f(t,\quad)$ in $\mathrm{Imm}(M,N)$ for G^0 is

$$E_{G^0}(f) = \tfrac{1}{2}\int_a^b \int_M g(f_t, f_t)\,\mathrm{vol}(f^*g).$$

The geodesic equation for G^0

$$\boxed{\begin{aligned} \nabla^g_{\partial_t} f_t + \operatorname{div}^{f^*g}(f_t^\top) f_t - g(f_t^\perp, \operatorname{Tr}^{f^*g}(S^f)) f_t +\\ + \tfrac12 Tf.\operatorname{grad}^{f^*g}(\|f_t\|_g^2) + \tfrac12 \|f_t\|_g^2 \operatorname{Tr}^{f^*g}(S^f) = 0 \end{aligned}} \tag{1}$$

Proof. A different proof is in [7], 2.2. For a function a on M we shall use

$$\begin{aligned} \int_M a \operatorname{div}^{f^*g}(X) \operatorname{vol}(f^*g) &= \int_M a \mathcal{L}_X(\operatorname{vol}(f^*g)) \\ &= \int_M \mathcal{L}_X(a \operatorname{vol}(f^*g)) - \int_M \mathcal{L}_X(a) \operatorname{vol}(f^*g) \\ &= -\int_M (f^*g)(\operatorname{grad}^{f^*g}(a), X) \operatorname{vol}(f^*g) \end{aligned}$$

in calculating the first variation of the energy with fixed ends:

$$\begin{aligned} \partial_s E_{G^0}(f) &= \tfrac12 \int_a^b \int_M \Big(\partial_s g(f_t,f_t) \ \operatorname{vol}(f^*g) + g(f_t,f_t)\, \partial_s \operatorname{vol}(f^*g) \Big) dt \\ &= \int_a^b \int_M \Big(g(\nabla^g_{\partial_s} f_t, f_t) \ \operatorname{vol}(f^*g) + \tfrac12 \|f_t\|_g^2 \operatorname{div}^{f^*g}(f_s^\top) \operatorname{vol}(f^*g) \\ &\qquad - \tfrac12 \|f_t\|_g^2 g(f_s^\perp, \operatorname{Tr}^{f^*g}(S^f)) \operatorname{vol}(f^*g) \Big) dt \end{aligned}$$

For the first summand we have:

$$\begin{aligned} &\int_a^b \int_M g(\nabla^g_{\partial_s} f_t, f_t) \ \operatorname{vol}(f^*g)\, dt = \int_a^b \int_M g(\nabla^g_{\partial_t} f_s, f_t) \ \operatorname{vol}(f^*g)\, dt \\ &= \int_a^b \int_M (\partial_t g(f_s,f_t) - g(f_s, \nabla^g_{\partial_t} f_t)) \operatorname{vol}(f^*g)\, dt \\ &= -\int_a^b \int_M g(f_s,f_t) \partial_t \operatorname{vol}(f^*g)\, dt - \int_a^b \int_M g(f_s, \nabla^g_{\partial_t} f_t) \operatorname{vol}(f^*g)\, dt \\ &= \int_a^b \int_M \Big(-g(f_s,f_t) \operatorname{div}^{f^*g}(f_t^\top) + g(f_s,f_t) g(f_t^\perp, \operatorname{Tr}^{f^*g}(S^f)) - \\ &\qquad - g(f_s, \nabla^g_{\partial_t} f_t) \Big) \operatorname{vol}(f^*g)\, dt \end{aligned}$$

The second summand yields:

$$\begin{aligned} &\int_a^b \int_M \tfrac12 \|f_t\|_g^2 \operatorname{div}^{f^*g}(f_s^\top) \operatorname{vol}(f^*g)\, dt \\ &\qquad = -\int_a^b \int_M \tfrac12 (f^*g)(f_s^\top, \operatorname{grad}^{f^*g}(\|f_t\|_g^2)) \operatorname{vol}(f^*g)\, dt \\ &\qquad = -\int_a^b \int_M \tfrac12 g(f_s, Tf.\operatorname{grad}^{f^*g}(\|f_t\|_g^2)) \operatorname{vol}(f^*g)\, dt \end{aligned}$$

Thus the first variation $\partial_s E_{G^0}(f)$ is:

$$\int_a^b \int_M g\Big(f_s, -\nabla^g_{\partial_t} f_t - \operatorname{div}^{f^*g}(f_t^\top) f_t + g(f_t^\perp, \operatorname{Tr}^{f^*g}(S^f)) f_t \\ - \tfrac12 Tf.\operatorname{grad}^{f^*g}(\|f_t\|_g^2) - \tfrac12 \|f_t\|_g^2 g(f_s^\perp, \operatorname{Tr}^{f^*g}(S^f))\Big) \operatorname{vol}(f^*g)\, dt. \quad \square$$

4.2. GEODESICS FOR G^0 IN $B_i(M,N)$. We restrict to geodesics $t \mapsto f(t, \quad)$ in $\operatorname{Imm}(M,N)$ which are horizontal: $g(f_t, Tf) = 0$. Then $f_t^\top = 0$ and $f_t = f_t^\perp$, so equation 4.1.(1) becomes

$$\nabla^g_{\partial_t} f_t - g(f_t, \operatorname{Tr}^{f^*g}(S^f)) f_t + \tfrac12 Tf.\operatorname{grad}^{f^*g}(\|f_t\|_g^2) + \tfrac12 \|f_t\|_g^2 \operatorname{Tr}^{f^*g}(S^f) = 0.$$

It splits into a vertical (tangential) part

$$-Tf.(\nabla_{\partial_t} f_t)^\top + \tfrac12 Tf.\operatorname{grad}^{f^*g}(\|f_t\|_g^2) = 0$$

which vanishes identically since

$$\begin{aligned}(f^*g)(\operatorname{grad}^{f^*g}(\|f_t\|_g^2), X) &= X(g(f_t,f_t)) = 2g(\nabla_X f_t, f_t) = 2g(\nabla^g_{\partial_t} Tf.X, f_t)\\ &= 2\partial_t g(Tf.X, f_t) - 2g(Tf.X, \nabla^g_{\partial_t} f_t) = -2g(Tf.X, \nabla^g_{\partial_t} f_t),\end{aligned}$$

and a horizontal (normal) part which is the geodesic equation in B_i:

$$(1) \qquad \boxed{\nabla^{N(f)}_{\partial_t} f_t - g(f_t, \operatorname{Tr}^{f^*g}(S^f)) f_t + \tfrac12 \|f_t\|_g^2 \operatorname{Tr}^{f^*g}(S^f) = 0, \quad g(Tf, f_t) = 0}$$

4.3. THE INDUCED METRIC OF G^0 IN $B_i(M,N)$ IN A CHART. Let $f_0 : M \to N$ be a fixed immersion which will be the 'center' of our chart. Let $N(f_0) \subset f_0^*TN$ be the normal bundle to f_0. Let $\exp^g : N(f_0) \to N$ be the exponential map for the metric g and let $V \subset N(f_0)$ be a neighborhood of the 0 section on which the exponential map is an immersion. Consider the mapping

$$(1) \qquad \begin{aligned}&\psi = \psi_{f_0} : \Gamma(V) \to \operatorname{Imm}(M,N), \quad , \psi(\Gamma(V)) =: \mathcal{Q}(f_0),\\ &\psi(a)(x) = \exp^g(a(x)) = \exp^g_{f_0(x)}(a(x)).\end{aligned}$$

The inverse (on its image) of $\pi \circ \psi_f : \Gamma(V) \to B_i(M,N)$ is a smooth chart on $B_i(M,N)$. Our goal is to calculate the induced metric on this chart, that is

$$((\pi \circ \psi_{f_0})^* G^0_a)(b_1, b_2)$$

for any $a \in \Gamma(V), b_1, b_2 \in \Gamma(N(f_0))$. This will enable us to calculate the sectional curvatures of B_i.

We shall fix the section a and work with the ray of points $t.a$ in this chart. Everything will revolve around the map:

$$f(t,x) = \psi(t.a)(x) = \exp^g(t.a(x)).$$

We shall also use a fixed chart $(M \supset U \xrightarrow{u} \mathbb{R}^m)$ on M with $\partial_i = \partial/\partial u^i$. Then $x \mapsto (t \mapsto f(t,x)) = \exp^g_{f_0(x)}(t.a(x))$ is a variation consisting entirely of geodesics, thus:

$$\begin{aligned}
&t \mapsto \partial_i f(t,x) = Tf.\partial_i =: Z_i(t,x,a) \textit{ is the Jacobi field along } t \mapsto f(t,x) \textit{with}\\
&Z_i(0,x,a) = \partial_i|_x \exp_{f_0(x)}(0) = \partial_i|_x f_0 = T_x f_0.\partial_i|_x, \textit{ and}\\
&(\nabla^g_{\partial_t} Z_i)(0,x) = (\nabla^g_{\partial_t} Tf.\partial_i)(0,x) = (\nabla^g_{\partial_i} Tf.\partial_t)(0,x) =\\
&\qquad = \nabla^g_{\partial_i}(\partial_t|_0 \exp_{f_0(x)}(t.a(x))) = (\nabla^g_{\partial_i} a)(x).
\end{aligned} \tag{2}$$

Then the pullback metric is given by

$$\begin{aligned}
f^*g &= \psi(ta)^*g = g(Tf,Tf) =\\
&= \sum_{i,j=1}^m g(Tf.\partial_i, Tf,\partial_j)\, du^i \otimes du^j = \sum_{i,j=1}^m g(Z_i,Z_j)\, du^i \otimes du^j.
\end{aligned} \tag{3}$$

The induced volume density is:

$$\operatorname{vol}(f^*g) = \sqrt{\det(g(Z_i,Z_j))}|du^1 \wedge \dots \wedge du^m| \tag{4}$$

Moreover we have for $a \in \Gamma(V)$ and $b \in \Gamma(N(f_0))$

$$\begin{aligned}
&(T_{ta}\psi.tb)(x) = \partial_s|_0 \exp^g_{f_0(x)}(ta(x) + stb(x)))\\
&\qquad = Y(t,x,a,b) \quad \text{for the Jacobi field } Y \text{ along } t \mapsto f(t,x) \text{ with}\\
&Y(0,x,a,b) = 0_{f_0(x)},\\
&(\nabla^g_{\partial_t} Y(\quad, x,a,b))(0) = \nabla^g_{\partial_t}\partial_s \exp^g_{f_0(x)}(ta(x) + stb(x))\\
&\qquad = \nabla^g_{\partial_s}\partial_t|_0 \exp^g_{f_0(x)}(ta(x) + stb(x))\\
&\qquad = \nabla^g_{\partial_s}(a(x) + s.b(x))|_{s=0} = b(x).
\end{aligned} \tag{5}$$

Now we want to split $T_a\psi.b$ into vertical (tangential) and horizontal parts with respect to the immersion $\psi(ta) = f(t, \quad)$. The tangential part has locally the form

$$\begin{aligned}
&Tf.(T_{ta}\psi.tb)^\top = \sum_{i=1}^m c^i\, Tf.\partial_i = \sum_{i=1}^m c^i\, Z_i \quad \text{where for all } j\\
&g(Y,Z_j) = g\left(\sum_{i=1}^m c^i\, Z_i, Z_j\right) = \sum_{i=1}^m c^i\, (f^*g)_{ij},\\
&c^i = \sum_{j=1}^m (f^*g)^{ij} g(Y,Z_j).
\end{aligned}$$

Thus the horizontal part is

$$(T_{ta}\psi.tb)^\perp = Y^\perp = Y - \sum_{i=1}^m c^i\, Z_i = Y - \sum_{i,j=1}^m (f^*g)^{ij}\, g(Y,Z_j)\, Z_i \tag{6}$$

Thus the induced metric on $B_i(M,N)$ has the following expression in the chart $(\pi\circ\psi_{f_0})^{-1}$, where $a\in\Gamma(V)$ and $b_1,b_2\in\Gamma(N(f_0))$:

$$\begin{aligned}
((\pi\circ\psi_{f_0})^*G^0)_{ta}(b_1,b_2) &= G^0_{\pi(\psi(ta))}(T_{ta}(\pi\circ\psi)b_1, T_{ta}(\pi\circ\psi)b_2)\\
&= G^0_{\psi(ta)}((T_{ta}\psi.b_1)^\perp,(T_{ta}\psi.b_2)^\perp)\\
&= \int_M g((T_{ta}\psi.b_1)^\perp,(T_{ta}\psi.b_2)^\perp)\ \mathrm{vol}(f^*g)\\
&= \int_M \frac{1}{t^2} g\Big(Y(b_1)-\sum_{i,j}(f^*g)^{ij}\, g(Y(b_1),Z_j)\,Z_i\ ,\ Y(b_2)\Big)\\
&\qquad\qquad \sqrt{\det(g(Z_i,Z_j))}|du^1\wedge\cdots\wedge du^m|
\end{aligned}\tag{7}$$

4.4. Expansion to order 2 of the induced metric of G^0 in $B_i(M,N)$ in a chart. We use the setting of 4.3, the Einstein summation convention, and the abbreviations $f_i:=\partial_i f_0=\partial_{u^i}f_0$ and $\nabla^g_i:=\nabla^g_{\partial_i}=\nabla^g_{\partial_{u^i}}$. We compute the expansion in t up to order 2 of the metric 4.3.(7). Our method is to use the Jacobi equation

$$\nabla^g_{\partial_t}\nabla^g_{\partial_t}Y=R^g(\dot c,Y)\dot c$$

which holds for any Jacobi field Y along a geodesic c. By 2.6.(2) we have:

$$g(\nabla^g_i a,f_j)=-(f_0^*g)(L^{f_0}_a(\partial_i),\partial_j)=-g(a,S^{f_0}(f_i,f_j))=g(a,S^{f_0}_{ij})\tag{1}$$

We start by expanding the pullback metric 4.3.(3) and its inverse:

$$\begin{aligned}
\partial_t(f^*g)_{ij} &= \partial_t g(Z_i,Z_j)=g(\nabla^g_{\partial_t}Z_i,Z_j)+g(Z_i,\nabla^g_{\partial_t}Z_j)\\
\partial_t^2 g(Z_i,Z_j) &= g(\nabla^g_{\partial_t}\nabla^g_{\partial_t}Z_i,Z_j)+2g(\nabla^g_{\partial_t}Z_i,\nabla^g_{\partial_t}Z_j)+g(Z_i,\nabla^g_{\partial_t}\nabla^g_{\partial_t}Z_j)\\
(f^*g)_{ij} &= (f_0^*g)_{ij}+t\big(g(\nabla^g_i a,f_j)+g(f_i,\nabla^g_j a)\big)+\\
&\quad+\tfrac12 t^2\big(g(R^g(a,f_i)a,f_j)+2g(\nabla^g_i a,\nabla^g_j a)+g(f_i,R^g(a,f_j)a)\big)\\
&\quad+O(t^3)\\
&= (f_0^*g)_{ij}-2t(f_0^*g)(L^{f_0}_a(\partial_i),\partial_j)\\
&\quad+t^2\big(g(R^g(a,f_i)a,f_j)+g(\nabla^g_i a,\nabla^g_j a)\big)+O(t^3)
\end{aligned}\tag{2}$$

We expand now the volume form $\mathrm{vol}(f^*g)=\sqrt{\det(g(Z_i,Z_j))}|du^1\wedge\cdots\wedge du^m|$. The time derivative at 0 of the inverse of the pullback metric is:

$$\partial_t(f^*g)^{ij}|_0=-(f_0^*g)^{ik}(\partial_t|_0(f^*g)_{kl})(f_0^*g)^{lj}=-(f_0^*g)^{ik}(f_0^*g)(a,S^{f_0}_{kl})(f_0^*g)^{lj}$$

Therefore,

$$\begin{aligned}
\partial_t\sqrt{\det(g(Z_i,Z_j))} &= \tfrac12(f^*g)^{ij}\partial_t(g(Z_i,Z_j))\sqrt{\det(g(Z_i,Z_j))}\\
\partial_t^2\sqrt{\det(g(Z_i,Z_j))} &= \tfrac12\partial_t(f^*g)^{ij}\partial_t(g(Z_i,Z_j))\sqrt{\det(g(Z_i,Z_j))}\\
&\quad+\tfrac12(f^*g)^{ij}\partial_t^2(g(Z_i,Z_j))\sqrt{\det(g(Z_i,Z_j))}
\end{aligned}$$

$$+ \tfrac{1}{2}(f^*g)^{ij}\partial_t(g(Z_i,Z_j))\partial_t\sqrt{\det(g(Z_i,Z_j))}$$

and so

$$\begin{aligned}
\operatorname{vol}(f^*g) &= \sqrt{\det(g(Z_i,Z_j))}|du^1\wedge\cdots\wedge du^m| \\
&= \Big(1 - t\operatorname{Tr}(L^{f_0}_a) + t^2\Big(-\operatorname{Tr}(L^{f_0}_a\circ L^{f_0}_a) + \tfrac{1}{2}(\operatorname{Tr}(L^{f_0}_a))^2 \\
&\qquad + \tfrac{1}{2}(f^*g)^{ij}\big(g(R^g(a,f_i)a,f_j) + g(\nabla^g_i a,\nabla^g_j a)\big)\Big) + O(t^3)\Big)\operatorname{vol}(f_0^*g).
\end{aligned} \tag{3}$$

Moreover, by 2.6.(2) we may split $\nabla^g_i a = -Tf_0.L^{f_0}_a(\partial_i)+\nabla^{N(f_0)}_i a$ and we write $\nabla^\perp_i a$ for $\nabla^{N(f_0)}_i a$ shortly. Thus:

$$\begin{aligned}
(f_0^*g)^{ij}g(\nabla^g_i a,\nabla^g_j a) &= (f_0^*g)^{ij}\left(g(Tf_0.L^{f_0}_a(\partial_i),Tf_0.L^{f_0}_a(\partial_j)) + g(\nabla^\perp_i a,\nabla^\perp_j a)\right) \\
&= \operatorname{Tr}(L^{f_0}_a\circ L^{f_0}_a) + (f_0^*g)^{ij}g(\nabla^\perp_i a,\nabla^\perp_j a)
\end{aligned}$$

and so that the tangential term above combines with the first t^2 term in the expansion of the volume, changing its coefficient from -1 to $-\frac{1}{2}$.

Let us now expand

$$\begin{aligned}
&g((T_{ta}\psi.tb_1)^\perp,(T_{ta}\psi.tb_2)^\perp) \\
&\qquad = g\big(Y(b_1) - (f^*g)^{ij}g(Y(b_1),Z_j)\,Z_i\ ,\ Y(b_2)\big) \\
&\qquad = g(Y(b_1),Y(b_2)) - (f^*g)^{ij}g(Y(b_1),Z_j)g(Z_i,Y(b_2)).
\end{aligned}$$

We have:

$$\begin{aligned}
\partial_t g(Y(b_1),Y(b_2)) &= g(\nabla^g_{\partial_t}Y(b_1),Y(b_2)) + g(Y(b_1),\nabla^g_{\partial_t}Y(b_2)) \\
\partial_t^2 g(Y(b_1),Y(b_2)) &= g(\nabla^g_{\partial_t}\nabla^g_{\partial_t}Y(b_1),Y(b_2)) + 2g(\nabla^g_{\partial_t}Y(b_1),\nabla^g_{\partial_t}Y(b_2)) \\
&\quad + g(Y(b_1),\nabla^g_{\partial_t}\nabla^g_{\partial_t}Y(b_2)) \\
&= 2g(R^g(a,Y(b_1))a,Y(b_2)) + 2g(\nabla^g_{\partial_t}Y(b_1),\nabla^g_{\partial_t}Y(b_2)) \\
\partial_t g(Y(b_1),Z_j) &= g(\nabla^g_{\partial_t}Y(b_1),Z_j) + g(Y(b_1),\nabla^g_{\partial_t}Z_j) \\
\partial_t^2 g(Y(b_1),Z_j) &= 2g(R^g(a,Y(b_1))a,Z_j) + 2g(\nabla^g_{\partial_t}Y(b_1),\nabla^g_{\partial_t}Z_j)
\end{aligned}$$

Note that:

$$\begin{aligned}
&Y(0,h)=0,\quad (\nabla^g_{\partial_t}Y(h))(0)=h,\quad (\nabla^g_{\partial_t}\nabla^g_{\partial_t}Y(h))(0) = R^g(a,Y(0,h))a = 0, \\
&(\nabla^g_{\partial_t}\nabla^g_{\partial_t}\nabla^g_{\partial_t}Y(h))(0) = R^g(a,\nabla^g_{\partial_t}Y(h)(0))a = R^g(a,h)a.
\end{aligned}$$

Thus:

$$\begin{aligned}
&g(Y(b_1),Y(b_2)) - (f^*g)^{ij}g(Y(b_1),Z_j)g(Z_i,Y(b_2)) \\
&= t^2g(b_1,b_2) + t^4\big(\tfrac{1}{3}g(R^g(a,b_1)a,b_2) - (f_0^*g)^{ij}g(b_1,\nabla^\perp_j a)g(\nabla^\perp_i a,b_2)\big) + O(t^5).
\end{aligned}$$

The expansion of G^0 up to order 2 is thus:

$$\begin{aligned}
&((\pi\circ\psi_{f_0})^*G^0)_{ta}(b_1,b_2) = \\
&\qquad = \int_M \frac{1}{t^2}g\Big(Y(b_1) - \sum_{i,j}(f^*g)^{ij}\,g(Y(b_1),Z_j)\,Z_i\ ,\ Y(b_2)\Big)\ \operatorname{vol}(f^*g)
\end{aligned}$$

$$
\begin{aligned}
&= \int_M \Big(g(b_1,b_2) \operatorname{vol}(f_0^* g) - t \int_M g(b_1,b_2) \operatorname{Tr}(L_a^{f_0}) \operatorname{vol}(f_0^* g) \\
&\quad + t^2 \int_M \Big(g(b_1,b_2) \Big(-\tfrac12 \operatorname{Tr}(L_a^{f_0} \circ L_a^{f_0}) + \tfrac12 \operatorname{Tr}(L_a^{f_0})^2 \\
&\qquad + \tfrac12 (f^* g)^{ij} g(R^g(a,f_i)a, f_j) + \tfrac12 (f^*g)^{ij} g(\nabla_i^\perp a, \nabla_j^\perp a) \Big) \\
&\qquad + \tfrac13 g(R^g(a,b_1)a,b_2) - (f_0^* g)^{ij} g(b_1, \nabla_j^\perp a) g(\nabla_i^\perp a, b_2) \Big) \operatorname{vol}(f_0^* g) \\
&\quad + O(t^3) \qquad (4)
\end{aligned}
$$

4.5. Computation of the sectional curvature in $B_i(M,N)$ at f_0. We use the following formula which is valid in a chart:

$$
\begin{aligned}
&2R_a(m,h,m,h) = 2G_a^0(R_a(m,h)m,h) = \\
&= -2d^2 G^0(a)(m,h)(h,m) + d^2G^0(a)(m,m)(h,h) + d^2G^0(a)(h,h)(m,m) \\
&\quad - 2G^0(\Gamma(h,m),\Gamma(m,h)) + 2G^0(\Gamma(m,m),\Gamma(h,h))
\end{aligned}
$$

The sectional curvature at the two-dimensional subspace $P_a(m,h)$ of the tangent space which is spanned by m and h is then given by:

$$
k_a(P(m,h)) = -\frac{G_a^0(R(m,h)m,h)}{\|m\|^2\|h\|^2 - G_a^0(m,h)^2}.
$$

We compute this directly for $a = 0$. From the expansion up to order 2 of $G_{ta}^0(b_1,b_2)$ in 4.4.(4) we get

$$
dG^0(0)(a)(b_1,b_2) = -\int_M g(b_1,b_2) g(a, \operatorname{Tr}^{f_0^* g} S^{f_0}) \operatorname{vol}(f_0^* g)
$$

and we compute the Christoffel symbol:

$$
\begin{aligned}
-2G_0^0(\Gamma_0(a,b),c) &= -dG^0(0)(c)(a,b) + dG^0(0)(a)(b,c) + dG^0(0)(b)(c,a) \\
&= \int_M \Big(g(a,b) g(c, \operatorname{Tr}^{f_0^* g}(S^{f_0})) - g(b,c) g(a, \operatorname{Tr}^{f_0^* g}(S^{f_0})) \\
&\qquad - g(c,a) g(b, \operatorname{Tr}^{f_0^* g}(S^{f_0})) \Big) \operatorname{vol}(f_0^* g) \\
&= \int_M g\Big(c,\ g(a,b) \operatorname{Tr}^{f_0^* g}(S^{f_0}) - \operatorname{Tr}(L_a^{f_0}) b - \operatorname{Tr}(L_b^{f_0}) a\Big) \operatorname{vol}(f_0^* g) \\
\Gamma_0(a,b) &= -\tfrac12 g(a,b) \operatorname{Tr}^{f_0^* g}(S^{f_0}) + \tfrac12 \operatorname{Tr}(L_a^{f_0}) b + \tfrac12 \operatorname{Tr}(L_b^{f_0}) a
\end{aligned}
$$

The expansion 4.4.(4) also gives:

$$
\begin{aligned}
&\tfrac{1}{2!} d^2 G_0^0(a_1,a_2)(b_1,b_2) = \\
&= \int_M \Big(g(b_1,b_2) \Big(-\tfrac12 \operatorname{Tr}(L_{a_1}^{f_0} \circ L_{a_2}^{f_0}) + \tfrac12 \operatorname{Tr}(L_{a_1}^{f_0}) \operatorname{Tr}(L_{a_2}^{f_0}) \\
&\qquad + \tfrac12 (f^*g)^{ij} g(R^g(a_1,f_i)a_2,f_j) + \tfrac12 (f^*g)^{ij} g(\nabla_i^\perp a_1, \nabla_j^\perp a_2) \Big) \\
&\qquad + \tfrac16 g(R^g(a_1,b_1)a_2,b_2) + \tfrac16 g(R^g(a_2,b_1)a_1,b_2)
\end{aligned}
$$

$$- \tfrac{1}{2}(f_0^*g)^{ij} g(b_1, \nabla_j^\perp a_1) g(\nabla_i^\perp a_2, b_2)$$
$$- \tfrac{1}{2}(f_0^*g)^{ij} g(b_1, \nabla_j^\perp a_2) g(\nabla_i^\perp a_1, b_2)\Big) \operatorname{vol}(f_0^*g) + O(t^3)$$

Thus we have:

$$\begin{aligned}
&- d^2G^0(0)(x,y)(y,x) + \tfrac{1}{2}d^2G^0(0)(x,x)(y,y) + \tfrac{1}{2}d^2G^0(0)(y,y)(x,x) = \\
&= \int_M \Big(-2g(y,x)\Big(-\tfrac{1}{2}\operatorname{Tr}(L_x^{f_0}\circ L_y^{f_0}) + \tfrac{1}{2}\operatorname{Tr}(L_x^{f_0})\operatorname{Tr}(L_y^{f_0}) \\
&\qquad\qquad + \tfrac{1}{2}(f_0^*g)^{ij} g(R^g(x,f_i)y,f_j) + \tfrac{1}{2}(f_0^*g)^{ij} g(\nabla_i^\perp x, \nabla_j^\perp y)\Big) \\
&\qquad + g(y,y)\Big(-\tfrac{1}{2}\operatorname{Tr}(L_x^{f_0}\circ L_x^{f_0}) + \tfrac{1}{2}\operatorname{Tr}(L_x^{f_0})\operatorname{Tr}(L_x^{f_0}) \\
&\qquad\qquad + \tfrac{1}{2}(f_0^*g)^{ij} g(R^g(x,f_i)x,f_j) + \tfrac{1}{2}(f_0^*g)^{ij} g(\nabla_i^\perp x, \nabla_j^\perp x)\Big) \\
&\qquad + g(x,x)\Big(-\tfrac{1}{2}\operatorname{Tr}(L_y^{f_0}\circ L_y^{f_0}) + \tfrac{1}{2}\operatorname{Tr}(L_y^{f_0})\operatorname{Tr}(L_y^{f_0}) \\
&\qquad\qquad + \tfrac{1}{2}(f_0^*g)^{ij} g(R^g(y,f_i)y,f_j) + \tfrac{1}{2}(f_0^*g)^{ij} g(\nabla_i^\perp y, \nabla_j^\perp y)\Big) \\
&\qquad + g(R^g(y,x)y,x) \\
&\qquad + (f_0^*g)^{ij}\,\big(g(y,\nabla_j^\perp y)g(\nabla_i^\perp x,x) + g(y,\nabla_j^\perp x)g(\nabla_i^\perp y,x)\big) \\
&\qquad - (f_0^*g)^{ij}\,\big(g(x,\nabla_j^\perp y)g(\nabla_i^\perp y,x) + g(y,\nabla_j^\perp x)g(\nabla_i^\perp x,y)\big)\Big) \operatorname{vol}(f_0^*g)
\end{aligned}$$

For the second part of the curvature we have

$$\begin{aligned}
-G_0(\Gamma_0(x,y),&\Gamma_0(y,x)) + G_0(\Gamma_0(y,y),\Gamma_0(x,x)) = \\
&= \tfrac{1}{4}\int_M \Big((\|x\|_g^2\|y\|_g^2 - g(x,y)^2)\|\operatorname{Tr}^{f_0^*g}(S^f)\|_g^2 \\
&\qquad - 3\|\operatorname{Tr}(L_x^{f_0})y - \operatorname{Tr}(L_y^{f_0})x\|_g^2\Big)\operatorname{vol}(f_0^*g)
\end{aligned}$$

To organize all these terms in the curvature tensor, note that they belong to *three* types: terms which involve the second fundamental form L^{f_0}, terms which involve the curvature tensor R^g of N and terms which involve the normal component of the covariant derivative $\nabla^\perp a$. There are 3 of the first type, two of the second and the ones of the third can be organized neatly into two also. The final curvature tensor is the integral over M of their sum. Here are the terms in detail:

(1) *Terms involving the trace of products of L's.* These are:

$$-\tfrac{1}{2}\Big(g(y,y)\operatorname{Tr}(L_x^{f_0}\circ L_x^{f_0}) - 2g(x,y)\operatorname{Tr}(L_x^{f_0}\circ L_y^{f_0}) + g(x,x)\operatorname{Tr}(L_y^{f_0}\circ L_y^{f_0})\Big).$$

Note that x and y are sections of the normal bundle $N(f_0)$, so we may define $x\wedge y$ to be the induced section of $\bigwedge^2 N(f_0)$. Then the expression inside the parentheses is a positive semi-definite quadratic function of $x\wedge y$. To see this, note a simple linear algebra fact – that if $Q(a,b)$ is any positive semi-definite inner product on $\mathbb{R}^n$, then

$$\widetilde{Q}(a\wedge b, c\wedge d) = \\ <a,c>Q(b,d)- <a,d> Q(b,c)+ <b,d> Q(a,c)- <b,c> Q(a,d)$$

$$\widetilde{Q}(a\wedge b, a\wedge b) = \|a\|^2 Q(b,b) - 2 <a,b> Q(a,b) + \|b\|^2 Q(a,a)$$

is a positive semi-definite inner product on $\bigwedge^2 V$. In particular, $\operatorname{Tr}(L^{f_0}_x \circ L^{f_0}_y)$ is a positive semi-definite inner product on the normal bundle, hence it defines a positive semi-definite inner product $\widetilde{\operatorname{Tr}}(L^{f_0}\circ L^{f_0})$ on $\bigwedge^2 N(f_0)$. Thus:

$$\operatorname{term}(1) = -\tfrac12 \widetilde{\operatorname{Tr}}(L^{f_0}\circ L^{f_0})(x\wedge y) \le 0.$$

(2) *Terms involving trace of one L.* We have terms both from the second and first derivatives of G, namely

$$\tfrac12\Big(g(y,y)\operatorname{Tr}(L^{f_0}_x)^2 - 2g(x,y)\operatorname{Tr}(L^{f_0}_x)\operatorname{Tr}(L^{f_0}_y) + g(x,x)\operatorname{Tr}(L^{f_0}_y)^2\Big)$$

and

$$-\tfrac34 \|\operatorname{Tr}(L^{f_0}_x)y - \operatorname{Tr}(L^{f_0}_y)x\|^2_g$$

which are the same up to their coefficients. Their sum is

$$\operatorname{term}(2) = -\tfrac14 \|\operatorname{Tr}(L^{f_0}_x)y - \operatorname{Tr}(L^{f_0}_y)x\|^2_g \le 0.$$

Note that this is a function of $x\wedge y$ also.

(3) *The term involving the norm of the second fundamental form.* Since $\|x\|^2_g\|y\|^2_g - g(x,y)^2 = \|x\wedge y\|^2_g$, this term is just:

$$\operatorname{term}(3) = +\tfrac14 \|x\wedge y\|^2_g \|\operatorname{Tr}^g(S^{f_0})\|^2_g \ge 0.$$

(4) *The curvature of N term.* This is

$$\operatorname{term}(4) = g(R^g(x,y)x,y).$$

Note that because of the skew-symmetry of the Riemann tensor, this is a function of $x\wedge y$ also.

(5) *The Ricci-curvature-like term.* The other curvature terms are

$$\tfrac12 (f_0^*g)^{ij}\Big(g(x,x)g(R^g(y,f_i)y,f_j) \\ - 2g(x,y)g(R^g(x,f_i)y,f_j) + g(y,y)g(R^g(x,f_i)x,f_j)\Big).$$

If V and W are two perpendicular subspaces of the tangent space TN_p at a point p, then we can define a 'cross Ricci curvature' $\operatorname{Ric}(V,W)$ in terms of bases $\{v_i\},\{w_j\}$ of V and W by

$$\operatorname{Ric}(V,W) = g^{ij}g^{kl}g(R^g(v_i,w_k)v_j,w_l).$$

Then this term factors as

$$\operatorname{term}(5) = \|x\wedge y\|^2_g \operatorname{Ric}(TM, \operatorname{span}(x,y)).$$

(6-7) *Terms involving the covariant derivative of a.* It is remarkable that, so far, every term in the curvature tensor of B_i vanishes if $x \wedge y \equiv 0$, e.g., if the codimension of N in M is one! Now we have the terms

$$\begin{aligned}(f_0^* g)^{ij}\Big(&g(x,y)g(\nabla_i^\perp x, \nabla_j^\perp y) - \tfrac{1}{2}g(x,x)g(\nabla_i^\perp y, \nabla_j^\perp y) - \tfrac{1}{2}g(y,y)g(\nabla_i^\perp x, \nabla_j^\perp x)\\ &- g(x, \nabla_i^\perp x)g(y, \nabla_j^\perp y) - g(x, \nabla_i^\perp y)g(y, \nabla_j^\perp x)\\ &+ g(x, \nabla_i^\perp y)g(x, \nabla_j^\perp y) + g(y, \nabla_i^\perp x)g(y, \nabla_j^\perp x)\Big).\end{aligned}$$

To understand this expression, we need a linear algebra computation, namely that if $a, b, a', b' \in \mathbb{R}^n$, then:

$$\begin{aligned}&< a,b >< a',b' > - < a,a' >< b,b' > - < a,b' >< b,a' > -\\ &- \tfrac{1}{2} < a,a >< b',b' > - \tfrac{1}{2} < b,b >< a',a' > + < a,b' >^2 + < b,a' >^2 =\\ &\qquad = \tfrac{1}{2}(< a,b' > - < b,a' >)^2 - \tfrac{1}{2}\|a \wedge b' - b \wedge a'\|^2\end{aligned}$$

Note that the term $g(x, \nabla^\perp y)$ (without an i) is a section of Ω^1_M and the sum over i and j is just the norm in Ω^1_M, so the above computation applies and the expression splits into 2 terms:

$$\begin{aligned}\text{term}(6) &= -\tfrac{1}{2}\|(g(x, \nabla^\perp y) - g(y, \nabla^\perp x)\|^2_{\Omega^1_M} \le 0\\ \text{term}(7) &= \tfrac{1}{2}\|x \wedge \nabla^\perp y - y \wedge \nabla^\perp x\|^2_{\Omega^1_M \otimes \wedge^2 N(f)} \ge 0\end{aligned}$$

Altogether, we get that the Riemann curvature of B_i is the integral over M of the sum of the above 7 terms. We have the Corollary:

Corollary. *If the codimension of M in N is one, then all sectional curvatures of B_i are non-negative. For any codimension, sectional curvature in the plane spanned by x and y is non-negative if x and y are parallel, i.e., $x \wedge y = 0$ in $\bigwedge^2 T^*N$.*

In general, the negative terms in the curvature tensor (giving positive sectional curvature) are clearly connected with the vanishing of geodesic distance: in some directions the space wraps up on itself in tighter and tighter ways. However, in codimension two or more with a flat ambient space N (so terms (4) and (5) vanish), there seem to exist conflicting tendencies making B_i close up or open up: terms (1), (2) and (6) give positive curvature, while terms (3) and (7) give negative curvature. It would be interesting to explore the geometrical meaning of these, e.g., for manifolds of space curves.

5. Vanishing geodesic distance on groups of diffeomorphisms

5.1. The H^0-metric on groups of diffeomorphisms. Let (N, g) be a smooth connected Riemannian manifold, and let $\mathrm{Diff}_c(N)$ be the group of all diffeomorphisms with compact support on N, and let $\mathrm{Diff}_0(N)$ be the subgroup of those which are diffeotopic in $\mathrm{Diff}_c(N)$ to the identity; this is the connected component of the identity in $\mathrm{Diff}_c(N)$, which is a regular Lie group in the sense of [8], section 38, see [8], section 42. The Lie algebra is $\mathfrak{X}_c(N)$, the space of all smooth vector fields with compact support on N, with the negative of the usual bracket of vector fields as Lie bracket. Moreover, $\mathrm{Diff}_0(N)$ is a simple group (has no nontrivial normal subgroups), see [5], [14], [9]. The *right invariant* H^0-metric on $\mathrm{Diff}_0(N)$ is then given as follows, where $h, k : N \to TN$ are vector fields with compact support along φ and where $X = h \circ \varphi^{-1}, Y = k \circ \varphi^{-1} \in \mathfrak{X}_c(N)$:

$$\begin{aligned} G^0_\varphi(h,k) &= \int_N g(h,k)\,\mathrm{vol}(\varphi^* g) = \int_N g(X\circ\varphi, Y\circ\varphi)\varphi^*\,\mathrm{vol}(g) \\ &= \int_N g(X,Y)\,\mathrm{vol}(g) \qquad (1) \end{aligned}$$

5.2. THEOREM. *Geodesic distance on* $\mathrm{Diff}_0(N)$ *with respect to the* H^0*-metric vanishes.*

PROOF. Let $[0,1] \ni t \mapsto \varphi(t, \quad)$ be a smooth curve in $\mathrm{Diff}_0(N)$ between φ_0 and φ_1. Consider the curve $u = \varphi_t \circ \varphi^{-1}$ in $\mathfrak{X}_c(N)$, the right logarithmic derivative. Then for the length and the energy we have:

$$L_{G^0}(\varphi) = \int_0^1 \sqrt{\int_N \|u\|_g^2\,\mathrm{vol(g)}}\; dt \tag{1}$$

$$E_{G^0}(\varphi) = \int_0^1 \int_N \|u\|_g^2\,\mathrm{vol(g)}\; dt \tag{2}$$

$$L_{G^0}(\varphi)^2 \le E_{G^0}(\varphi) \tag{3}$$

(4) Let us denote by $\mathrm{Diff}_0(N)^{E=0}$ the set of all diffeomorphisms $\varphi \in \mathrm{Diff}_0(N)$ with the following property: For each $\varepsilon > 0$ there exists a smooth curve from the identity to φ in $\mathrm{Diff}_0(N)$ with energy $\le \varepsilon$.

(5) *We claim that* $\mathrm{Diff}_0(N)^{E=0}$ *coincides with the set of all diffeomorphisms which can by reached from the identity by a smooth curve of arbitraily short* G^0*-length.* This follows by (3).

(6) *We claim that* $\mathrm{Diff}_0(N)^{E=0}$ *is a normal subgroup of* $\mathrm{Diff}_0(N)$. Let $\varphi_1 \in \mathrm{Diff}_0(N)^{E=0}$ and $\psi \in \mathrm{Diff}_0(N)$. For any smooth curve $t \mapsto \varphi(t, \quad)$ from the identity to φ_1 with energy $E_{G^0}(\varphi) < \varepsilon$ we have

$$\begin{aligned} E_{G^0}(\psi^{-1}\circ\varphi\circ\psi) &= \int_0^1\int_N \|T\psi^{-1}\circ\varphi_t\circ\psi\|_g^2\,\mathrm{vol}((\psi^{-1}\circ\varphi\circ\psi)^* g) \\ &\le \sup_{x\in N}\|T_x\psi^{-1}\|^2 \cdot \int_0^1\int_N \|\varphi_t\circ\psi\|_g^2 (\varphi\circ\psi)^*\,\mathrm{vol}((\psi^{-1})^* g) \\ &\le \sup_{x\in N}\|T_x\psi^{-1}\|^2 \cdot \sup_{x\in N}\frac{\mathrm{vol}((\psi^{-1})^* g)}{\mathrm{vol}(g)} \cdot \int_0^1\int_N \|\varphi_t\circ\psi\|_g^2\,(\varphi\circ\psi)^*\,\mathrm{vol}(g) \end{aligned}$$

$$\leq \sup_{x\in N} \|T_x\psi^{-1}\|^2 \cdot \sup_{x\in N} \frac{\mathrm{vol}((\psi^{-1})^*g)}{\mathrm{vol}(g)} \cdot E_{G^0}(\varphi).$$

Since ψ is a diffeomorphism with compact support, the two suprema are bounded. Thus $\psi^{-1}\circ\varphi_1\circ\psi \in \mathrm{Diff}_0(N)^{E=0}$.

(7) *We claim that* $\mathrm{Diff}_0(N)^{E=0}$ *is a non-trivial subgroup.* In view of the simplicity of $\mathrm{Diff}_0(N)$ mentioned in 5.1 this concludes the proof.

It remains to find a non-trivial diffeomorphism in $\mathrm{Diff}_0(N)^{E=0}$. The idea is to use compression waves. The basic case is this: take any non-decreasing smooth function $f:\mathbb{R}\to\mathbb{R}$ such that $f(x)\equiv 0$ if $x\ll 0$ and $f(x)\equiv 1$ if $x\gg 0$. Define

$$\varphi(t,x) = x + f(t-\lambda x)$$

where $\lambda < 1/\max(f')$. Note that

$$\varphi_x(t,x) = 1-\lambda f'(t-\lambda x) > 0,$$

hence each map $\varphi(t,\quad)$ is a diffeomorphism of $\mathbb{R}$ and we have a path in the group of diffeomorphisms of $\mathbb{R}$. These maps are not the identity outside a compact set however. In fact, $\varphi(x) = x+1$ if $x\ll 0$ and $\varphi(x)=x$ if $x\gg 0$. As $t\to-\infty$, the map $\varphi(t,\quad)$ approaches the identity, while as $t\to+\infty$, the map approaches translation by 1. This path is a moving compression wave which pushes all points forward by a distance 1 as it passes. We calculate its energy between two times t_0 and t_1:

$$\begin{aligned}
E_{t_0}^{t_1}(\varphi) &= \int_{t_0}^{t_1}\int_{\mathbb{R}} \varphi_t(t,\varphi(t,\quad)^{-1}(x))^2 dx\,dt = \int_{t_0}^{t_1}\int_{\mathbb{R}} \varphi_t(t,y)^2\varphi_y(t,y)dy\,dt \\
&= \int_{t_0}^{t_1}\int_{\mathbb{R}} f'(z)^2\cdot(1-\lambda f'(z))dz/\lambda\,dt \\
&\leq \frac{\max f'^2}{\lambda}\cdot(t_1-t_0)\cdot\int_{\mathrm{supp}(f')}(1-\lambda f'(z))dz
\end{aligned}$$

If we let $\lambda = 1-\varepsilon$ and consider the specific f given by the convolution

$$f(z) = \max(0,\min(1,z))\star G_\varepsilon(z),$$

where G_ε is a smoothing kernel supported on $[-\varepsilon,+\varepsilon]$, then the integral is bounded by 3ε, hence

$$E_{t_0}^{t_1}(\varphi) \leq (t_1-t_0)\tfrac{3\varepsilon}{1-\varepsilon}.$$

We next need to adapt this path so that it has compact support. To do this we have to start and stop the compression wave, which we do by giving it variable length. Let:

$$f_\varepsilon(z,a) = \max(0,\min(a,z))\star(G_\varepsilon(z)G_\varepsilon(a)).$$

The starting wave can be defined by:

$$\varphi_\varepsilon(t,x) = x + f_\varepsilon(t-\lambda x, g(x)),\quad \lambda<1,\quad g \text{ increasing}.$$

Note that the path of an individual particle x hits the wave at $t = \lambda x - \varepsilon$ and leaves it at $t = \lambda x + g(x) + \varepsilon$, having moved forward to $x + g(x)$. Calculate the derivatives:

$$\begin{aligned}
(f_\varepsilon)_z &= I_{0\le z\le a} \star (G_\varepsilon(z)G_\varepsilon(a)) \in [0,1] \\
(f_\varepsilon)_a &= I_{0\le a\le z} \star (G_\varepsilon(z)G_\varepsilon(a)) \in [0,1] \\
(\varphi_\varepsilon)_t &= (f_\varepsilon)_z(t-\lambda x, g(x)) \\
(\varphi_\varepsilon)_x &= 1 - \lambda(f_\varepsilon)_z(t-\lambda x, g(x)) + (f_\varepsilon)_a(t-\lambda x, g(x)) \cdot g'(x) > 0.
\end{aligned}$$

This gives us:

$$\begin{aligned}
E_{t_0}^{t_1}(\varphi) &= \int_{t_0}^{t_1}\int_{\mathbb{R}} (\varphi_\varepsilon)_t^2 (\varphi_\varepsilon)_x dx\, dt \\
&\le \int_{t_0}^{t_1}\int_{\mathbb{R}} (f_\varepsilon)_z^2(t-\lambda x, g(x)) \cdot (1-\lambda(f_\varepsilon)_z(t-\lambda x, g(x)))dx\, dt \\
&\quad + \int_{t_0}^{t_1}\int_{\mathbb{R}} (f_\varepsilon)_z^2(t-\lambda x, g(x)) \cdot (f_\varepsilon)_a(t-\lambda x, g(x)) g'(x) dx\, dt
\end{aligned}$$

The first integral can be bounded as in the original discussion. The second integral is also small because the support of the z-derivative is $-\varepsilon \le t - \lambda x \le g(x) + \varepsilon$, while the support of the a-derivative is $-\varepsilon \le g(x) \le t - \lambda x + \varepsilon$, so together $|g(x) - (t - \lambda x)| \le \varepsilon$. Now define x_1 and x_2 by $g(x_1) + \lambda x_1 = t + \varepsilon$ and $g(x_0) + \lambda x_0 = t - \varepsilon$. Then the inner integral is bounded by

$$\int_{|g(x)+\lambda x - t|\le\varepsilon} g'(x)dx = g(x_1) - g(x_0) \le 2\varepsilon,$$

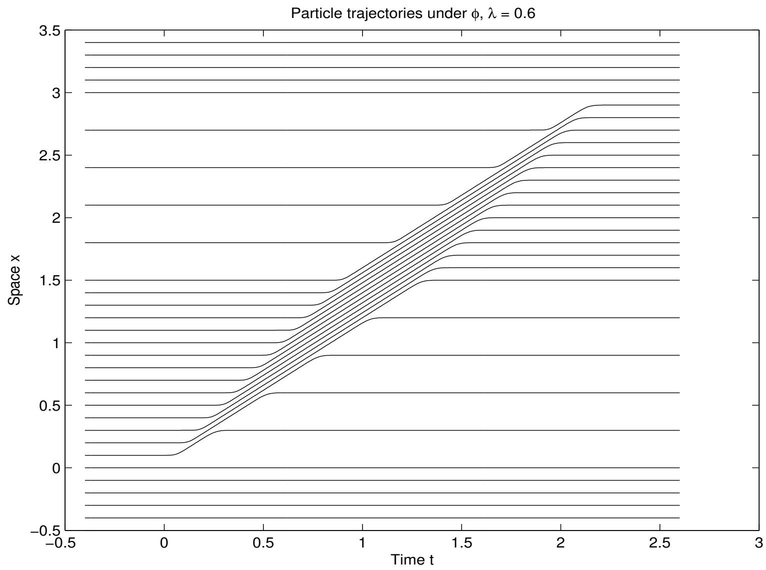
and the whole second term is bounded by $2\varepsilon(t_1 - t_0)$. Thus the length is $O(\varepsilon)$.

The end of the wave can be handled by playing the beginning backwards. If the distance that a point x moves when the wave passes it is to be $g(x)$, so that the final diffeomorphism is $x \mapsto x + g(x)$ then let $b = \max(g)$ and use the above definition of φ while $g' > 0$. The modification when $g' < 0$ (but $g' > -1$ in order for $x \mapsto x + g(x)$ to have positive derivative) is given by:

$$\varphi_\varepsilon(t, x) = x + f_\varepsilon(t - \lambda x - (1-\lambda)(b - g(x)), g(x)).$$

A figure showing the trajectories $\varphi_\varepsilon(t, x)$ for sample values of x is shown in the figure above.

It remains to show that $\mathrm{Diff}_0(N)^{E=0}$ is a nontrivial subgroup for an arbitrary Riemannian manifold. We choose a piece of a unit speed geodesic containing no conjugate points in N and Fermi coordinates along this geodesic; so we can assume that we are in an open set in $\mathbb{R}^m$ which is a tube around a piece of the u^1-axis. Now we use a small bump function in the the slice orthogonal to the u^1-axis and multiply it with the construction from above for the coordinate u^1. Then it follows that we get a nontrivial diffeomorphism in $\mathrm{Diff}_0(N)^{E=0}$ again. □



Remark. Theorem 5.2 can possibly be proved directly without the help of the simplicity of $\mathrm{Diff}_0(N)$. For $N = \mathbb{R}$ one can use the method of 5.2, (7) in the parameter space of a curve, and for general N one can use a Morse function on N to produce a special coordinate for applying the same method, as we did in the proof of theorem 3.2.

5.3. Geodesics and sectional curvature for G^0 on $\mathrm{Diff}(N)$. According to Arnold [1], see [11], 3.3, for a right invariant weak Riemannian metric G on an (possibly infinite dimensional) Lie group the geodesic equation and the curvature are given in terms of the adjoint operator (with respect to G, if it exists) of the Lie bracket by the following formulas:

$$\begin{aligned}
u_t &= -\operatorname{ad}(u)^* u, \quad u = \varphi_t \circ \varphi^{-1} \\
G(\operatorname{ad}(X)^* Y, Z) &:= G(Y, \operatorname{ad}(X) Z) \\
4G(R(X,Y)X,Y) &= 3G(\operatorname{ad}(X)Y, \operatorname{ad}(X)Y) - 2G(\operatorname{ad}(Y)^* X, \operatorname{ad}(X)Y) \\
&\quad - 2G(\operatorname{ad}(X)^* Y, \operatorname{ad}(Y)X) + 4G(\operatorname{ad}(X)^* X, \operatorname{ad}(Y)^* Y) \\
&\quad - G(\operatorname{ad}(X)^* Y + \operatorname{ad}(Y)^* X, \operatorname{ad}(X)^* Y + \operatorname{ad}(Y)^* X)
\end{aligned}$$

In our case, for $\mathrm{Diff}_0(N)$, we have $\operatorname{ad}(X)Y = -[X,Y]$ (the bracket on the Lie algebra $\mathfrak{X}_c(N)$ of vector fields with compact support is the negative of the usual one), and:

$$\begin{aligned}
G^0(X,Y) &= \int_N g(X,Y)\operatorname{vol}(g) \\
G^0(\operatorname{ad}(Y)^* X, Z) &= G^0(X, -[Y,Z]) = \int_N g(X, -\mathcal{L}_Y Z)\operatorname{vol}(g)
\end{aligned}$$

$$= \int_N g\Big(\mathcal{L}_Y X + (g^{-1}\mathcal{L}_Y g)X + \operatorname{div}^g(Y)X, Z\Big)\operatorname{vol}(g)$$
$$\operatorname{ad}(Y)^* = \mathcal{L}_Y + g^{-1}\mathcal{L}_Y(g) + \operatorname{div}^g(Y)\operatorname{Id}_T N = \mathcal{L}_Y + \beta(Y),$$

where the tensor field $\beta(Y) = g^{-1}\mathcal{L}_Y(g) + \operatorname{div}^g(Y)\operatorname{Id} : TN \to TN$ is self adjoint with respect to g. Thus the geodesic equation is

$$u_t = -(g^{-1}\mathcal{L}_u(g))(u) - \operatorname{div}^g(u)u = -\beta(u)u, \qquad u = \varphi_t \circ \varphi^{-1}.$$

The main part of the sectional curvature is given by:

$$\begin{aligned}
&4G(R(X,Y)X,Y) = \\
&= \int_N \Big(3\|[X,Y]\|_g^2 + 2g((\mathcal{L}_Y + \beta(Y))X,[X,Y]) + 2g((\mathcal{L}_X + \beta(X))Y,[Y,X]) \\
&\qquad + 4g(\beta(X)X,\beta(Y)Y) - \|\beta(X)Y + \beta(Y)X\|_g^2\Big)\operatorname{vol}(g) \\
&= \int_N \Big(-\|\beta(X)Y - \beta(Y)X + [X,Y]\|_g^2 - 4g([\beta(X),\beta(Y)]X,Y)\Big)\operatorname{vol}(g)
\end{aligned}$$

So sectional curvature consists of a part which is visibly non-negative, and another part which is difficult to decompose further.

5.4. Example: Burgers' equation. For $(N,g) = (\mathbb{R},\text{can})$ or (S^1,can) the geodesic equation is *Burgers' equation* [2], a completely integrable infinite dimensional system,

$$u_t = -3u_x\,u, \qquad u = \varphi_t \circ \varphi^{-1}$$

and we get $G^0(R(X,Y)X,Y) = -\int [X,Y]^2\,dx$ so that all sectional curvatures are non-negative.

5.5. Example: n-dimensional analog of Burgers' equation. For $(N,g) = (\mathbb{R}^n,\text{can})$ or $((S^1)^n,\text{can})$ we have:

$$(\operatorname{ad}(X)Y)^k = \sum_i((\partial_i X^k)Y^i - X^i(\partial_i Y^k))$$
$$(\operatorname{ad}(X)^*Z)^k = \sum_i\Big((\partial_k X^i)Z^i + (\partial_i X^i)Z^k + X^i(\partial_i Z^k)\Big),$$

so that the geodesic equation is given by

$$\partial_t u^k = -(\operatorname{ad}(u)^\top u)^k = -\sum_i\Big((\partial_k u^i)u^i + (\partial_i u^i)u^k + u^i(\partial_i u^k)\Big),$$

called the basic Euler-Poincaré equation (EPDiff) in [6], the n-dimensional analog of Burgers' equation.

5.6. Stronger metrics on $\mathrm{Diff}_0(N)$. A very small strengthening of the weak Riemannian H^0-metric on $\mathrm{Diff}_0(N)$ makes it into a true metric. We define the stronger right invariant semi-Riemannian metric by the formula:

$$G^A_\varphi(h,k) = \int_N (g(X,Y) + A\,\mathrm{div}_g(X).\,\mathrm{div}_g(Y))\,\mathrm{vol}(g).$$

Then the following holds:

5.7. Theorem. *For any distinct diffeomorphisms φ_0, φ_1, the infimum of the lengths of all paths from φ_0 to φ_1 with respect to G^A is positive.*

This implies that the metric G^0 induces positive geodesic distance on the subgroup of volume preserving diffeomorphism since it coincides there with the metric G^A.

Proof. Let $\psi_1 = \varphi_0 \circ \varphi_1^{-1}$. If $\varphi_0 \neq \varphi_1$, there are two functions ρ and f on N with compact support such that:

$$\int_N \rho(y) f(\psi_1(y))\,\mathrm{vol}(g)(y) \neq \int_N \rho(y) f(y)\,\mathrm{vol}(g)(y).$$

Now consider any path $\varphi(t,y)$ between the two maps with derivative $u = \varphi_t \circ \varphi^{-1}$. Inverting the diffeomorphisms (or switching from a Lagrangian to an Eulerian point of view), let $\psi(t, \quad) = \varphi(0, \quad) \circ \varphi(t, \quad)^{-1}$. Then $\psi_t = -T\psi(u)$ and we have:

$$\begin{aligned}
\int_N &\rho(y) f(\psi_1(y))\,\mathrm{vol}(g)(y) - \int_N \rho(y) f(y)\,\mathrm{vol}(g)(y) = \\
&= \int_0^1 \int_N \rho(y) \partial t f(\psi(t,y)\,\mathrm{vol}(g)(y) dt \\
&= \int_0^1 \int_N \rho(y) (df \circ \psi)(\psi_t(t,y))\,\mathrm{vol}(g)(y)\,dt \\
&= \int_0^1 \int_N \rho(y) (Tf \circ \psi)(-T\psi(u(t,y)))\,\mathrm{vol}(g)(y) dt
\end{aligned}$$

But $\mathrm{div}((f\circ\psi)\cdot \rho u) = (f\circ\psi)\cdot \mathrm{div}(\rho u) + (Tf\circ\psi)(T\psi(\rho u))$. The integral of the left hand side is 0, hence:

$$\begin{aligned}
\Big| \int_N &\rho(y) f(\psi_1(y))\,\mathrm{vol}(g)(y) - \int_N \rho(y) f(y)\,\mathrm{vol}(g)(y) \Big| \\
&= \Big| \int_0^1 \int_N (f\circ\psi)\,\mathrm{div}(\rho u)\,\mathrm{vol}(g) dt \Big| \\
&\leq \sup(|f|) \int_0^1 \sqrt{\int_N C_\rho \|u\|^2 + C'_\rho |\,\mathrm{div}(u)|^2\,\mathrm{vol}(g) dt}
\end{aligned}$$

for constants C_ρ, C'_ρ depending only on ρ. Clearly the right hand side is a lower bound for the length of any path from φ_0 to φ_1. □

5.8. GEODESICS FOR G^A ON $\mathrm{Diff}(\mathbb{R})$. See [3] and [12]. We consider the groups $\mathrm{Diff}_c(\mathbb{R})$ or $\mathrm{Diff}(S^1)$ with Lie algebras $\mathfrak{X}_c(\mathbb{R})$ or $\mathfrak{X}(S^1)$ with Lie bracket $\mathrm{ad}(X)Y = -[X,Y] = X'Y - XY'$. The G^A-metric equals the H^1-metric on $\mathfrak{X}_c(\mathbb{R})$, and we have:

$$\begin{aligned}
G^A(X,Y) &= \int_{\mathbb{R}} (XY + AX'Y')dx = \int_{\mathbb{R}} X(1 - A\partial_x^2)Y\,dx,\\
G^A(\mathrm{ad}(X)^*Y, Z) &= \int_{\mathbb{R}} (YX'Z - YXZ' + AY'(X'Z - XZ')')dx\\
&= \int_{\mathbb{R}} Z(1-\partial_x^2)(1-\partial_x^2)^{-1}(2YX' + Y'X - 2AY''X' - AY'''X)dx,\\
\mathrm{ad}(X)^*Y &= (1-\partial_x^2)^{-1}(2YX' + Y'X - 2AY''X' - AY'''X)\\
\mathrm{ad}(X)^* &= (1-\partial_x^2)^{-1}(2X' + X\partial_x)(1 - A\partial_x^2)
\end{aligned}$$

so that the geodesic equation in Eulerian representation $u = (\partial_t f)\circ f^{-1} \in \mathfrak{X}_c(\mathbb{R})$ or $\mathfrak{X}(S^1)$ is

$$\begin{aligned}
\partial_t u = -\,\mathrm{ad}(u)^*u &= -(1-\partial_x^2)^{-1}(3uu' - 2Au''u' - Au'''u), \text{ or}\\
u_t - u_{txx} &= Au_{xxx}.u + 2Au_{xx}.u_x - 3u_x.u,
\end{aligned}$$

which for $A = 1$ is the *Camassa-Holm equation* [3], another completely integrable infinite dimensional Hamiltonian system. Note that here geodesic distance is a well defined metric describing the topology.

REFERENCES

[1] V.I. Arnold. Sur la géometrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Ann. Inst. Fourier* 16 (1966), 319–361.

[2] J. Burgers. A mathematical model illustrating the theory of turbulence. *Adv. Appl. Mech.* 1 (1948), 171–199.

[3] R. Camassa and D. Holm. An integrable shallow water equation with peaked solutions. *Phys. Rev. Lett.* 71 (1993), 1661-1664.

[4] Vicente Cervera, Francisca Mascaro, and Peter W. Michor. The orbit structure of the action of the diffeomorphism group on the space of immersions. *Diff. Geom. Appl.*, 1 (1991), 391–401.

[5] D.B.A. Epstein. The simplicity of certain groups of homeomorphisms. *Compositio Math.* 22 (1970), 165–173.

[6] D. Holm, T. Ratnanather, A. Trouve and L. Younes. Soliton Dynamics in Computational Anatomy *NeuroImage*, Sept. 2004.

[7] Gerd Kainz. A metric on the manifold of immersions and its Riemannian curvature. *Monatsh. Math.*, 98 (1984), 211-217.

[8] Andreas Kriegl and Peter W. Michor. *The Convenient Setting for Global Analysis.* AMS, Providence, 1997. 'Surveys and Monographs 53'.

[9] John N. Mather. Commutators of diffeomorphisms I, II. *Comm. Math. Helv.* 49 (1974), 512–528; 50 (1975), 33–40.

[10] Peter W. Michor and David Mumford. *Riemannian geometries on spaces of plane curves*. arXiv:math.DG/0312384.
[11] Peter W. Michor and Tudor Ratiu. Geometry of the Virasoro-Bott group. *J. Lie Theory*, 8 (1998), 293–309. arXiv:math.DG/9801115
[12] G. Misiołek. A shallow water equation as a geodesic flow on the Bott-Virasoro group. *J. Geom. Phys.* 24 (1998), 203–208.
[13] Richard Palais. On the existence of slices for actions of non-compact Lie groups. *Ann. of Math.* 73 (1961), 295–323.
[14] William Thurston. Foliations and groups of diffeomorphisms. *Bull. AMS* 80 (1974), 304–307.

Peter W. Michor
Fakultät für Mathematik
Universität Wien
Nordbergstrasse 15
A-1090 Wien, Austria
and
Erwin Schrödinger Institut
für Mathematische Physik
Boltzmanngasse 9
A-1090 Wien, Austria
Peter.Michor@univie.ac.at

David Mumford
Division of Applied Mathematics
Brown University
Box F, Providence, RI 02912, USA
David_Mumford@brown.edu

The Microstates Free Entropy Dimension of any DT–operator is 2

Ken Dykema[1], Kenley Jung[2], and Dimitri Shlyakhtenko[3]

Received: January 5, 2005

Communicated by Joachim Cuntz

Abstract. Suppose that μ is an arbitrary Borel measure on $\mathbb{C}$ with compact support and $c > 0$. If Z is a $\mathrm{DT}(\mu, c)$–operator as defined by Dykema and Haagerup in [6], then the microstates free entropy dimension of Z is 2.

2000 Mathematics Subject Classification: 46L54 (28A78)

1 Introduction.

DT–operators were introduced by Dykema and Haagerup in their work on invariant subspaces of certain operators in a II_1 factor [5, 6]. A DT–operator Z is specified by two parameters, μ and c, where $c > 0$ and μ is a Borel probability measure on $\mathbb{C}$ with compact support. Roughly, the operator Z is determined by stating that its $*$–distribution is the same as the limit $*$–distribution as $N \to \infty$ of random matrices

$$Z_N = D_N + cT_N,$$

where D_N are diagonal $N \times N$ matrices whose spectral measures converge to μ in distribution, while T_N is a strictly upper triangular random $N \times N$ matrix with i.i.d. Gaussian entries. Equivalently, (see [15], [12], [6] and the appendix of [7]), Z can be viewed as a sum $Z = d + cT$, where d is a normal operator with spectral measure μ contained in a diffuse von Neumann algebra A, and T is an A-valued circular operator with a certain covariance. Finally, a result of Śniady [14] shows that a $\mathrm{DT}(\mu, c)$–operator is one whose free entropy is maximized among all those operators having Brown measure equal to μ and with a fixed off–diagonality.
If we write $Z = d+cT$ as above, it is clear that $W^*(Z) \subset W^*(d, T) \subseteq W^*(A\cup\{T\})$, while a simple computation shows $W^*(A \cup \{T\}) = L(\mathbb{F}_2)$. By Lemma 6.2 of [6], for any μ we may choose d having trace of spectral measure equal to μ and so that

[1]Research supported by the Alexander von Humboldt Foundation and NSF grant DMS–0300336.
[2]Research supported by an NSF Postdoctoral Fellowship.
[3]Research supported by a Sloan Foundation fellowship and NSF grant DMS-0355226.

$d, T \in W^*(Z)$; by [7], $A \subseteq W^*(T)$, so we always have $W^*(Z) \cong L(\mathbb{F}_2)$. Thus Z can be viewed as an interesting generator for this free group factor.

In order to test the hypothesis that Voiculescu's free entropy dimension δ_0 [16, 17, 20] is the same for any sets of generators of a von Neumann algebra, it is important to decide whether the free entropy dimension of Z is 2 ($L(\mathbb{F}_2)$ clearly has another set of generators of free entropy dimension 2).

For another version of free entropy dimension, also defined by Voiculescu, called the non-microstates free entropy dimension [18], L. Aagaard has recently shown [1] that the dimension of Z is indeed 2. It is known by [4] that the non-microstates free entropy dimension dominates δ_0 but at present it is open whether the reverse inequality holds. Thus, Aagaard's result does not solve the question for the original microstates definition.

In this paper, we show that, indeed, $\delta_0(Z) = 2$. Our proof uses an equivalent packing number formulation of the microstates free entropy dimension, due to Jung [8]. In this approach, to get the nontrivial lower bound on $\delta_0(Z)$, one must have lower bounds on the ϵ–packing numbers of spaces of matricial microstates for Z, which are in turn obtained by lower bounds on the volume of ϵ–neighborhoods of these microstate spaces. The kth microstate space is the set $\Gamma(Z; m, k, \gamma)$, for $m, k \in \mathbb{N}$ and $\gamma > 0$, of all $k \times k$ complex matrices whose $*$–moments up to order m are γ–close to the values of the corresponding $*$–moments of Z, and the volumes are for Lebesgue measure λ_k on $M_k(\mathbb{C})$ viewed as a Euclidean space of real dimension $2k^2$ with coordinates corresponding to the real and imaginary parts of the entries of a matrix.

In order to outline how we get these lower bounds on volumes, let us for convenience take Z equal to the $\mathrm{DT}(\delta_0, 1)$–operator T. A key result that we use is a recent one of Aagaard and Haagerup [2], showing that a certain ϵ–perturbation of T has Brown measure uniformly distributed on the disk of radius $r_\epsilon := 1/\sqrt{\log(1+\epsilon^{-2})}$ centered at the origin; note how slowly this disk shrinks as ϵ approaches zero. Applying a result of Śniady [13] to this situation, we find matrices $A_k \in M_k(\mathbb{C})$ that lie in ϵ–neighborhoods of microstate spaces for T, whose eigenvalues are close to uniformly distributed (as k gets large) in the disk of radius r_ϵ. Thus, in order to get a lower bound on the volume of a 2ϵ–neighborhood of a microstate space for T, it will suffice to get a lower bound on the volume of a unitary orbit of an ϵ–neighborhood of A_k.

Every element of $M_k(\mathbb{C})$ has an upper triangular matrix in its unitary orbit. Thus, letting $T_k(\mathbb{C})$ denote the set of upper triangular matrices in $M_k(\mathbb{C})$, there is a measure ν_k on $T_k(\mathbb{C})$ such that $\lambda_k(\mathcal{O}) = \nu_k(\mathcal{O} \cap T_k)$ for every $\mathcal{O} \subseteq M_k(\mathbb{C})$ invariant under unitary conjugation. Freeman Dyson identified such a measure ν_k (see Appendix 35 of [11]), and showed that if we view $T_k(\mathbb{C})$ as a Euclidean space of real dimension $k(k-1)$ with coordinates corresponding to the real and imaginary parts of the matrix entries lying on and above the diagonal, then ν_k is absolutely continuous with respect to Lebesgue measure on $T_k(\mathbb{C})$ and has density given at $B = (b_{ij})_{1 \le i,j \le k} \in T_k(\mathbb{C})$ by

$$C_k \prod_{1 \le p < q \le k} |b_{pp} - b_{qq}|^2, \tag{1}$$

where the constant is

$$C_k = \frac{\pi^{k(k-1)/2}}{\prod_{j=1}^k j!}. \tag{2}$$

We will use this measure of Dyson to find lower bound on the volume of unitary orbits of an ϵ–neighborhood of A_k, and we may take A_k to be upper triangular. However, so far we only have information about the eigenvalues of A_k, namely the diagonal part of it. Loosely speaking, in order to get a handle on the part strictly above the diagonal, we use a result of Dykema and Haagerup [6] to realize T as an upper triangular matrix

$$T = \frac{1}{\sqrt{N}} \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1N} \\ 0 & T_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & T_{N-1,N} \\ 0 & \cdots & 0 & T_{NN} \end{bmatrix}$$

of operators where each T_{ii} is a copy of T, each T_{ij} for $i < j$ is circular and the family $(T_{ij})_{1\le i\le j\le N}$ is $*$–free. Thus, A_k can be taken to be of the form

$$\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1N} \\ 0 & B_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_{N-1,N} \\ 0 & \cdots & 0 & B_{NN} \end{bmatrix}$$

where each B_{ii} is upper triangular, where we have good knowledge of the eigenvalue distributions of each B_{ii} and where the B_{ij} for $i < j$ approximate $*$–free circular elements. Using the strengthened asymptotic freeness results of Voiculescu [19], we find enough approximants for these B_{ij}. Although we still have no real knowledge about the entries of the B_{ii} lying above the diagonal, these parts are of negligibly small dimension as N gets large, and we are able to get good enough lower bounds. The techniques we use for estimating integrals of the quantity (1) over certain regions are taken from [9].

2 MICROSTATES FOR Z WITH WELL–SPACED SPECTRAL DENSITIES

The following lemma is an application of the result of Aagaard and Haagerup [2] mentioned in the introduction in order to make perturbations of general DT–operators having Brown measure that is relatively well spread out. For an element a of a non-commutative probability space $(\mathcal{M}, \tau)$, we write $\|a\|_2$ for $\tau(a^*a)^{1/2}$.

LEMMA 2.1. *Let μ be a compactly supported Borel probability measure on $\mathbb{C}$ and let $c > 0$. Let Z be a* $\mathrm{DT}(\mu, c)$*–operator in a W*–noncommutative probability space* $(\mathcal{M}, \tau)$*. Let us write*

$$\mu = \nu + \sum_{i=1}^s a_i \delta_{z_i}$$

for some $s \in \{0\} \cup \mathbb{N} \cup \{\infty\}$, $z_i \in \mathbb{C}$ and $a_i > 0$, where ν is a diffuse measure and where $z_i \neq z_j$ if $i \neq j$. Consider the W–noncommutative probability space*

$$(\widetilde{\mathcal{M}}, \tilde{\tau}) = (\mathcal{M}, \tau) * (L(\mathbb{F}_2), \tau_{\mathbb{F}_2}).$$

Then for every $\epsilon > 0$, there is $\widetilde{Z}_\epsilon \in \widetilde{\mathcal{M}}$ such that $\|\widetilde{Z}_\epsilon - Z\|_2 \leq \epsilon c$ and where the Brown measure of $\widetilde{Z}_\epsilon$ is equal to

$$\sigma_\epsilon := \nu + \sum_{i=1}^{s} a_i \rho_{i,\epsilon},$$

where $\rho_{i,\epsilon}$ is the probability measure that is uniform distribution on the disk centered at z_i and having radius

$$r_i := c\sqrt{\frac{a_i}{\log(1 + a_i \epsilon^{-2})}}.$$

Finally, if $\delta > 0$ and if

$$X_\delta = \{(w_1, w_2) \in \mathbb{C}^2 \mid |w_1 - w_2| < \delta\},$$

then

$$(\sigma_\epsilon \times \sigma_\epsilon)(X_\delta) \leq (\nu \times \nu)(X_\delta) + 2\sum_{i=1}^{s} \min(a_i, \delta^2 c^{-2} \log\big(1 + a_i \epsilon^{-2}\big)). \tag{3}$$

Proof. By results from [6], taking projections onto local spectral subspaces of Z, we find projections $p_j \in \mathcal{M}$ (for $0 \leq j < s+1$) such that

- $\sum_{j=0}^{s} p_j = 1$,
- $p_0 + p_1 + \cdots + p_k$ is Z–invariant for all integers k such that $0 \leq k < s+1$,
- $\tau(p_k) = \begin{cases} |\nu| & \text{if } k = 0 \\ a_k & \text{if } 1 \leq k < s+1, \end{cases}$
- In $(p_k \mathcal{M} p_k, \tau(p_k)^{-1} \tau\restriction_{p_k \mathcal{M} p_k})$, $p_k Z p_k$ is $\mathrm{DT}(|\nu|^{-1}\nu, c\sqrt{|\nu|})$ if $k = 0$ and is $\mathrm{DT}(\delta_{z_k}, c\sqrt{a_k})$ if $1 \leq k < s+1$.

Let $Y \in \widetilde{\mathcal{M}}$ be centered circular such that Y and Z are $*$–free and $\tilde{\tau}(Y^*Y) = 1$. Let

$$\widetilde{Z}_\epsilon = Z + \epsilon \sum_{i=1}^{s} a_i^{-1/2} c p_i Y p_i. \tag{4}$$

Then $\|\widetilde{Z}_\epsilon - Z\|_2^2 = \epsilon^2 c^2 \sum_{i=1}^{s} a_i \leq \epsilon^2 c^2$. On the other hand, $\widetilde{Z}_\epsilon$ is upper triangular with respect to the projections $p_0, p_1, \ldots$; the Brown measure of $\widetilde{Z}_\epsilon$ is, therefore, equal to the Brown measure of its diagonal part

$$p_0 Z p_0 + \sum_{i=1}^{s} \big(p_i Z p_i + \epsilon\, a_i^{-1/2} c p_i Y p_i\big). \tag{5}$$

But in $(p_i\widetilde{\mathcal{M}}p_i, a_i^{-1}\tilde{\tau}\restriction_{p_i\widetilde{\mathcal{M}}p_i})$, the operator $\epsilon\, a_i^{-1/2}cp_iYp_i$ is a centered circular operator of second moment ϵ^2c^2 that is $*$–free from the $\mathrm{DT}(\delta_{z_i}, c\sqrt{a_i})$ operator p_iZp_i. Therefore, the random variable

$$p_iZp_i + \epsilon\, a_i^{-1/2}cp_iYp_i \tag{6}$$

has the same $*$–distribution as $z_iI + c\sqrt{a_i}(T + \epsilon\, a_i^{-1/2}Y)$, where T is a $\mathrm{DT}(\delta_0, 1)$–operator that is $*$–free from Y. By [2], the Brown measure of the random variable (6) is equal to $\rho_{i,\epsilon}$. This yields σ_ϵ for the Brown measure of the operator (5), hence of $\widetilde{Z}_\epsilon$ itself.
Finally, we have

$$(\sigma_\epsilon \times \sigma_\epsilon)(X_\delta) \le (\nu \times \nu)(X_\delta) + 2\sum_{i=1}^{s} a_i(\sigma_\epsilon \times \rho_{i,\epsilon})(X_\delta) \tag{7}$$

and

$$(\sigma_\epsilon \times \rho_{i,\epsilon})(X_\delta) = \int_{\mathbb{C}} \rho_{i,\epsilon}(w + \delta\,\mathbb{D})d\sigma_\epsilon(w) \le \min(1, \delta^2 r_i^{-2}), \tag{8}$$

where $\mathbb{D}$ is the unit disk in $\mathbb{C}$. Taken together, (7) and (8) yield the inequality (3). □

The next lemma uses a result of Śniady [13] to find matrix approximants of the operators appearing in Lemma 2.1.
In the following lemma and throughout this paper, for a matrix $A \in M_k(\mathbb{C})$ we let $|A|_2 = \mathrm{tr}_k(A^*A)^{1/2}$, where tr_k is the normalized trace on $M_k(\mathbb{C})$. Moreover, by the eigenvalue distribution of $A \in M_k(\mathbb{C})$ we mean its Brown measure, which is just the probability measure that is uniformly distributed on its list of eigenvalues $\lambda_1, \ldots, \lambda_k$, where these are listed according to (general) multiplicity, i.e. a value z is listed $\dim \bigcup_{n=1}^\infty \ker((A - zI)^n)$ times.

LEMMA 2.2. *Let μ be a compactly supported Borel probability measure on $\mathbb{C}$ and let $c > 0$. Then there exists a sequence $\langle y_k\rangle_{k=1}^\infty$ such that for any $\epsilon > 0$, there exists a sequence $\langle z_{k,\epsilon}\rangle_{k=1}^\infty$ such that*

- *$y_k, z_{k,\epsilon} \in M_k(\mathbb{C})$,*
- *$\|y_k\|$ and $\|z_{k,\epsilon}\|$ remain bounded as $k \to \infty$,*
- *$\limsup_{k\to\infty} |y_k - z_{k,\epsilon}|_2 \le \epsilon c$,*
- *y_k converges in $*$–moments as $k \to \infty$ to a $\mathrm{DT}(\mu, c)$–operator,*
- *the eigenvalue distribution of $z_{k,\epsilon}$ converges weakly as $k \to \infty$ to the measure σ_ϵ described in Lemma 2.1.*

Proof. Let Z be a $\mathrm{DT}(\mu, c)$–operator, let $\widetilde{Y}$ be the operator $\sum_{i=1}^s a_i^{-1/2}cp_iYp_i$ appearing in (4) in the proof of the preceding lemma, so that $\widetilde{Z}_\epsilon = Z + \epsilon\widetilde{Y}$. Since Z can be constructed in $L(\mathbb{F}_2)$ and since free group factors can be embedded in the ultrapower R^ω of the hyperfinite II_1 factor, there are bounded sequences $\langle y_k\rangle_{k=1}^\infty$ and

$\langle d_k\rangle_{k=1}^{\infty}$ such that $y_k, d_k \in M_k(\mathbb{C})$ and such that the pair y_k, d_k converges in $*$–moments to the pair $Z, \widetilde{Y}$. Letting $\widetilde{z}_k = y_k + \epsilon d_k$, we have that $\widetilde{z}_k$ converges in $*$–moments to $\widetilde{Z}_\epsilon$ as $k \to \infty$. By Theorem 7 of [13], there is a sequence $\langle z_{k,\epsilon}\rangle_{k=1}^{\infty}$ with $z_{k,\epsilon} \in M_k(\mathbb{C})$ such that $\|z_{k,\epsilon} - \widetilde{z}_{k,\epsilon}\|$ tends to zero and the eigenvalue distribution of $z_{k,\epsilon}$ converges weakly as $k \to \infty$ to the Brown measure of $\widetilde{Z}_\epsilon$, namely, to σ_ϵ. $\square$

Suppose that $\lambda = \langle\lambda_j\rangle_{j=1}^{k}$ is a finite sequence of complex numbers. For each j, write $\lambda_j = a_j + ib_j$, $a_j, b_j \in \mathbb{R}$. Define $Q_\epsilon = \prod_{j=1}^{k}[a_j - \epsilon, a_j + \epsilon]$ and $R_\epsilon = \prod_{j=1}^{k}[b_j - \epsilon, b_j + \epsilon]$. Set

$$E_\epsilon(\lambda) = \int_{R_\epsilon}\left(\int_{Q_\epsilon}\prod_{\substack{1\le i,j\le k\\ i\neq j}}\left(|s_i - s_j| + |t_i - t_j|^2\right)^{1/2} ds\right)dt,$$

where $ds = ds_1\cdots ds_k$ and $dt = dt_1\cdots dt_k$.
The following lemma proves lower bounds for certain asymptotics of the quantities $E_\epsilon(\lambda)$. We will apply this lemma to the case when λ is the eigenvalue sequence of matrices like the $z_{k,\epsilon}$ found in Lemma 2.2.

Lemma 2.3. *Let μ and c be as in Lemma 2.1. For each $\epsilon > 0$ and $k \in \mathbb{N}$, let $\lambda^{(k,\epsilon)} = \langle\lambda_1^{(k,\epsilon)}, \ldots, \lambda_{n(k)}^{(k,\epsilon)}\rangle$ be a finite sequence of complex numbers and assume that for every $\epsilon > 0$,*

$$\sup_{k\in\mathbb{N},\, 1\le j\le n(k)} |\lambda_j^{(k,\epsilon)}| < \infty$$

and the probability measures

$$\frac{1}{n(k)}\sum_{j=1}^{n(k)}\delta_{\lambda_j^{(k,\epsilon)}} \tag{9}$$

converge weakly to the measure σ_ϵ of Lemma 2.1 as $k \to \infty$. Let

$$f(\epsilon) = \liminf_{k\to\infty} n(k)^{-2}\log(E_\epsilon(\lambda^{(k,\epsilon)})).$$

Then

$$\liminf_{\epsilon\to 0}\left(\frac{f(\epsilon)}{|\log\epsilon|}\right) \ge 0. \tag{10}$$

Proof. Note that we must have $n(k) \to \infty$ as $k \to \infty$. Given $\epsilon > 0$ small, take $1 \ge \delta > 3\epsilon$. Define

$$W_{k,\epsilon} = \{(i,j) \in \{1,\ldots,n(k)\}^2 \mid i \neq j,\, |\lambda_i^{(k,\epsilon)} - \lambda_j^{(k,\epsilon)}| < \delta\}.$$

Writing for each $1 \le j \le k$, $\lambda_j^{(k,\epsilon)} = a_j + ib_j$ where $a_j, b_j \in \mathbb{R}$ define $Q_{\epsilon,k} =$

$\prod_{j=1}^{n(k)}[a_j-\epsilon,a_j+\epsilon]$, $R_{\epsilon,k}=\prod_{j=1}^{n(k)}[b_j-\epsilon,b_j+\epsilon]$, and $K_{\epsilon,k}=Q_{\epsilon,k}\times R_{\epsilon,k}$. Now

$$\begin{aligned}
E_\epsilon(\lambda^{(k,\epsilon)}) &= \int_{K_{\epsilon,k}}\prod_{i\neq j}(|s_i-s_j|^2+|t_i-t_j|^2)^{1/2}dsdt \\
&\geq (\delta-\sqrt{8}\epsilon)^{n(k)^2-\#W_{k,\epsilon}}\int_{K_{\epsilon,k}}\prod_{(i,j)\in W_{k,\epsilon}}(|s_i-s_j|^2+|t_i-t_j|^2)^{1/2}dsdt \\
&\geq (\delta-3\epsilon)^{n(k)^2-\#W_{k,\epsilon}}\left(\int_{Q_{\epsilon,k}}\prod_{(i,j)\in W_{k,\epsilon}}|s_i-s_j|ds\right)\cdot \\
&\qquad\qquad\left(\int_{R_{\epsilon,k}}\prod_{(i,j)\in W_{k,\epsilon}}|t_i-t_j|dt\right),
\end{aligned}$$

where $ds=ds_1\cdots ds_{n(k)}$ and $dt=dt_1\cdots dt_{n(k)}$.
We now wish to find a lower bounds for the two integrals in the above expression. By Fubini's Theorem we can assume $a_1\leq a_2\leq\cdots\leq a_{n(k)}$. Let

$$[-\epsilon,\epsilon]^{n(k)}_<=\{(x_1,\ldots,x_{n(k)})\in[-\epsilon,\epsilon]^{n(k)}\mid x_1<x_2<\cdots<x_{n(k)}\}.$$

Then by the change of variables $[-\epsilon,\epsilon]^{n(k)}_<\ni(x_1,\ldots,x_{n(k)})\mapsto(a_1+x_1,\ldots,a_{n(k)}+x_{n(k)})\in Q_{\epsilon,k}$ and Selberg's Integral Formula it follows that

$$\begin{aligned}
\int_{Q_{\epsilon,k}}\prod_{(i,j)\in W_{k,\epsilon}}|s_i-s_j|ds &\geq \int_{[-\epsilon,\epsilon]^{n(k)}_<}\prod_{(i,j)\in W_{k,\epsilon}}|x_i-x_j|dx_1\cdots dx_{n(k)} \\
&\geq (2\epsilon)^{-(n(k)^2-n(k)-\#W_{k,\epsilon})}\cdot\int_{[-\epsilon,\epsilon]^{n(k)}_<}\prod_{i\neq j}|x_i-x_j|dx_1\cdots dx_{n(k)} \\
&= \frac{(2\epsilon)^{-(n(k)^2-n(k)-\#W_{k,\epsilon})}}{n(k)!}\cdot\int_{[-\epsilon,\epsilon]^{n(k)}}\prod_{i\neq j}|x_i-x_j|dx_1\cdots dx_{n(k)} \\
&= \frac{(2\epsilon)^{n(k)+\#W_{k,\epsilon}}}{n(k)!}\cdot\prod_{j=0}^{n(k)-1}\frac{\Gamma(j+2)\Gamma(j+1)^2}{\Gamma(n(k)+j+1)},
\end{aligned}$$

The same lower bound applies to $\int_{R_{\epsilon,k}}\prod_{(i,j)\in W_{k,\epsilon}}|t_i-t_j|dt$ so that combining these two we get

$$\begin{aligned}
E_\epsilon(\lambda^{(k,\epsilon)}) &\geq (\delta-3\epsilon)^{n(k)^2-\#W_{k,\epsilon}}\left(\frac{(2\epsilon)^{n(k)+\#W_{k,\epsilon}}}{n(k)!}\cdot\prod_{j=0}^{n(k)-1}\frac{\Gamma(j+2)\Gamma(j+1)^2}{\Gamma(n(k)+j+1)}\right)^2 \\
&\geq (\delta-3\epsilon)^{n(k)^2}\left(\frac{(2\epsilon)^{n(k)+\#W_{k,\epsilon}}}{n(k)!}\cdot\prod_{j=0}^{n(k)-1}\frac{\Gamma(j+2)\Gamma(j+1)^2}{\Gamma(n(k)+j+1)}\right)^2.
\end{aligned}$$

Using

$$\lim_{k\to\infty} n(k)^{-2}\log(\prod_{j=0}^{n(k)-1}\frac{\Gamma(j+2)\Gamma(j+1)^2}{\Gamma(n(k)+j+1)}) = -2\log 2,$$

we find

$$f(\epsilon) \geq \log(\delta-3\epsilon) + 2\log(2\epsilon)\limsup_{k\to\infty}\frac{\#W_{k,\epsilon}}{n(k)^2} - 4\log 2.$$

Since the measures (9) converge weakly to σ_ϵ, by standard approximation techniques one sees

$$\lim_{k\to\infty}\frac{\#W_{k,\epsilon}}{n(k)^2} = (\sigma_\epsilon\times\sigma_\epsilon)(X_\delta),$$

where X_δ is as in Lemma 2.1. As $\epsilon\to 0$ choose $\delta = \frac{1}{|\log\epsilon|}$, so that $\delta^2\log(1+a\epsilon^{-2})\to 0$ for all $a>0$ and $\frac{\epsilon}{\delta}\to 0$ and $\frac{\log\delta}{\log\epsilon}\to 0$. Using the upper bound (3) and the fact that ν is diffuse, we get

$$\lim_{\epsilon\to 0}(\sigma_\epsilon\times\sigma_\epsilon)(X_\delta) = 0.$$

Now one easily verifies that (10) holds. □

3 The Main Result

Before beginning the main result first a few comments on a packing formulation for microstates free entropy dimension are in order. If $X=\{x_1,\ldots,x_n\}$ is an n-tuple of selfadjoint elements in a tracial von Neumann algebra, then the free entropy dimension (as defined by Voiculescu [17]) is given by the formula

$$\delta_0(X) = n + \limsup_{\epsilon\to 0}\frac{\chi(x_1+\epsilon s_1,\ldots,x_n+\epsilon s_n : s_1,\ldots,s_n)}{|\log\epsilon|}$$

where $\{s_1,\ldots,s_n\}$ is a semicircular family free from X. The packing formulation found in [8] and modified slightly in [10] (to remove the norm restriction on microstates), is

$$\delta_0(X) = \limsup_{\epsilon\to 0}\frac{\mathbb{P}_\epsilon(X)}{|\log\epsilon|},$$

where

$$\mathbb{P}_\epsilon(X) = \inf_{m\in\mathbb{N},\,\gamma>0}\limsup_{k\to\infty} k^{-2}\log P_\epsilon(\Gamma(X;m,k,\gamma)). \tag{11}$$

Here, $\Gamma(X;m,k,\gamma)\subseteq(M_k(\mathbb{C})_{s.a.})^n$ is the microstate space of Voiculescu [16], but taken without norm restriction, as considered in [3], and P_ϵ is the packing number with respect to the metric arising from the normalized trace.

Let $Y=\{y_1,\ldots,y_n\}$ be an arbitrary n-tuple of (possibly nonselfadjoint) elements in a tracial von Neumann algebra. Now the definition of $\mathbb{P}_\epsilon$ makes perfect sense for the set Y if we replace the microstate space in (11) with the non-selfadjoint $*$-microstate space $\Gamma(Y;m,k,\gamma)\subseteq(M_k(\mathbb{C}))^n$, which is the set of all n–tuples of $k\times k$ matrices whose $*$–moments up to order m approximate those of Y within tolerance of γ. Let us (temporarily) denote the quantity so obtained by $\overline{\mathbb{P}_\epsilon}(Y)$ and define

$$\overline{\delta_0}(Y) = \limsup_{\epsilon\to 0} \frac{\overline{\mathbb{P}_\epsilon}(Y)}{|\log\epsilon|}. \tag{12}$$

It is easy to see that if X is a set of selfadjoints, then $\overline{\mathbb{P}_\epsilon}(X) \geq \mathbb{P}_\epsilon(X) \geq \overline{\mathbb{P}_{2\epsilon}}(X)$ and that in the nonselfadjoint setting the quantity (12) is a $*$-algebraic invariant, so that

$$\begin{aligned}
\delta_0(\mathrm{Re}(y_1),&\mathrm{Im}(y_1),\dots,\mathrm{Re}(y_n),\mathrm{Im}(y_n)) = \\
&= \limsup_{\epsilon\to 0} \frac{\mathbb{P}_\epsilon(\mathrm{Re}(y_1),\mathrm{Im}(y_1),\dots,\mathrm{Re}(y_n),\mathrm{Im}(y_n))}{|\log\epsilon|} \\
&= \limsup_{\epsilon\to 0} \frac{\overline{\mathbb{P}_\epsilon}(\mathrm{Re}(y_1),\mathrm{Im}(y_1),\dots,\mathrm{Re}(y_n),\mathrm{Im}(y_n))}{|\log\epsilon|} \\
&= \limsup_{\epsilon\to 0} \frac{\overline{\mathbb{P}_\epsilon}(Y)}{|\log\epsilon|} = \overline{\delta_0}(Y),
\end{aligned}$$

where $\mathrm{Re}(y_i)$ and $\mathrm{Im}(y_i)$ are the real and imaginary parts of y_i. Moreover, if X is set of selfadjoints, then

$$\delta_0(X) = \limsup_{\epsilon\to 0} \frac{\mathbb{P}_\epsilon(X)}{|\log\epsilon|} = \limsup_{\epsilon\to 0} \frac{\overline{\mathbb{P}_\epsilon}(X)}{|\log\epsilon|} = \overline{\delta_0}(X).$$

The following notational conventions, which will be used in the remainder of this paper, are, therefore, justified: for any finite set of operators Y (selfadjoint or otherwise) in a tracial von Neumann algebra we will write $\mathbb{P}_\epsilon(Y)$ for the packing quantity derived from the nonselfadjoint microstates (that was denoted $\overline{\mathbb{P}_\epsilon}(Y)$ above) and we will write $\delta_0(Y)$ for the free entropy dimension of Y that was denoted $\overline{\delta_0}(Y)$ above.
In the proof of the main result, we will use $E_\epsilon(A)$ for $A \in M_k(\mathbb{C})$ to mean $E_\epsilon(\lambda)$, where $\lambda = \langle\lambda_j\rangle_{j=1}^k$ are the eigenvalues of A listed according to general multiplicity (see the description immediately before Lemma 2.2). Notice that this is independent of the choice of λ since $E_\epsilon(\lambda\circ\sigma) = E_\epsilon(\lambda)$ for any permutation σ of $\{1,\dots,k\}$.

Theorem 3.1. *Let Z be a $\mathrm{DT}(\mu,c)$–operator, for any compactly supported Borel probability measure μ on the complex plane and any $c > 0$. Then $\delta_0(Z) = 2$.*

Proof. Obviously $\delta_0(Z) \leq 2$ so it suffices to show the reverse inequality.
We may without loss of generality assume $c = 1$ (see Proposition 2.12 of [6]). Fix $N \in \mathbb{N}$ with $N \geq 2$. By Theorem 4.12 of [6],

$$\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1N} \\ 0 & B_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_{N-1,N} \\ 0 & \cdots & 0 & B_{NN} \end{bmatrix} \in \mathcal{M}\otimes M_N(\mathbb{C}) \tag{13}$$

is a $\mathrm{DT}(\mu,1)$–operator where $\{B_{11},\ldots,B_{NN}\}\cup\langle B_{ij}\rangle_{1\le i<j\le N}$ is a $*$-free family in $\mathcal{M}$, the B_{ii} are $\mathrm{DT}(\mu,\frac{1}{\sqrt{N}})$–operators, and each B_{ij} is circular with $\varphi(|B_{ij}^2|)=\frac{1}{N}$. From this we see that finding microstates for Z is equivalent to finding microstates for the operator (13) in $\mathcal{M}\otimes M_N(\mathbb{C})$.

Consider the sequence $\langle y_k\rangle_{k=1}^\infty$ constructed in Lemma 3.2 and for each $\epsilon>0$ small enough, the corresponding sequence $\langle z_{k,\epsilon}\rangle_{k=1}^\infty$. Let $R>1$, $m\in\mathbb{N}$, $\gamma>0$ and take $\gamma'=\gamma/16^m(R+1)^m>0$. By Corollary 2.11 of [19] there exist $k\times k$ complex unitary matrices $u_{1k},u_{2k},\ldots,u_{kk}$ such that $\{u_{1k}y_ku_{1k}^*,\ldots,u_{Nk}y_ku_{Nk}^*\}$ is an (m,γ')–$*$–free family in $M_k(\mathbb{C})$. Also,by an application of Corollary 2.14 of [19], there exists a set $\Omega_k\subset\Gamma_R(\langle B_{ij}\rangle_{1\le i<j\le N};m,k,\gamma')$ such that for any $\langle\eta_{ij}\rangle_{1\le i<j\le N}\in\Omega_k$,

$$\{u_{1k}y_ku_{1k}^*,\ldots,u_{Nk}y_ku_{Nk}^*\}\cup\langle\eta_{ij}\rangle_{1\le i<j\le N}$$

is an (m,γ')-$*$ free family and such that

$$\begin{aligned}\liminf_{k\to\infty}\left(k^{-2}\cdot\log(\mathrm{vol}(\Omega_k))+\frac{N(N-1)}{2}\cdot\log k\right)\ge\\ \ge\chi(\langle\mathrm{Re}B_{ij}\rangle_{1\le i<j\le N},\langle\mathrm{Im}B_{ij}\rangle_{1\le i<j\le N})>-\infty,\end{aligned}$$

where the volume is computed with respect to the product of the Euclidean norm $k^{1/2}|\cdot|_2$. Since the operator (13) is a copy of Z, for any $\langle\eta_{ij}\rangle_{1\le i<j\le N}\in\Omega_k$ we have

$$\begin{bmatrix}u_{1k}y_ku_{1k}^* & \eta_{12} & \cdots & \eta_{1N}\\ 0 & u_{2k}y_2u_{2k}^* & \ddots & \vdots\\ \vdots & \ddots & \ddots & \eta_{N-1,N}\\ 0 & \cdots & 0 & u_{Nk}y_ku_{Nk}^*\end{bmatrix}\in\Gamma(Z;m,Nk,\gamma).$$

Because every complex matrix can be put into an upper-triangular form with respect to an orthonormal basis, we can find for each $1\le j\le N$, a $k\times k$ unitary matrix v_{jk} such that $v_{jk}u_{jk}z_{k,\epsilon}u_{jk}^*v_{jk}^*$ is upper triangular. Observe now that for any $\langle\eta_{ij}\rangle_{1\le i<j\le n}\in\Omega_k$, the product of matrices

$$\begin{bmatrix}v_{1k} & 0 & \cdots & 0\\ 0 & v_{2k} & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & v_{Nk}\end{bmatrix}\begin{bmatrix}u_{1k}y_ku_{1k}^* & \eta_{12} & \cdots & \eta_{1N}\\ 0 & u_{2k}y_ku_{2k}^* & \ddots & \vdots\\ \vdots & \ddots & \ddots & \eta_{N-1,N}\\ 0 & \cdots & 0 & u_{Nk}y_ku_{Nk}^*\end{bmatrix}\cdot$$

$$\cdot\begin{bmatrix}v_{1k}^* & 0 & \cdots & 0\\ 0 & v_{2k}^* & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & v_{Nk}^*\end{bmatrix}$$

is also an element of $\Gamma(Z;m,Nk,\gamma)$ and is equal to

$$\begin{bmatrix} v_{1k}u_{1k}y_ku_{1k}^*v_{1k}^* & v_{1k}\eta_{12}v_{2k}^* & \cdots & v_{1j}\eta_{1N}v_{2k}^* \\ 0 & v_{2j}u_{2k}y_ku_{2k}^*v_{2k}^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & v_{(N-1),k}\eta_{N-1,N}v_{Nk}^* \\ 0 & \cdots & 0 & v_{Nk}u_{Nk}y_ku_{Nk}^*v_{Nk}^* \end{bmatrix}.$$

Moreover,

$$|v_{jk}u_{jk}z_{k,\epsilon}u_{jk}^*v_{jk}^* - v_{jk}u_{jk}y_ku_{jk}^*v_{jk}^*|_2 = |z_{k,\epsilon} - y_k|_2$$

and $\limsup_{k\to\infty}|z_{k,\epsilon} - y_k|_2 \le \epsilon/\sqrt{N}$. Therefore, for k sufficiently large and for each $1 \le j \le N$ we have $|v_{jk}u_{jk}z_{k,\epsilon}u_{jk}^*v_{jk}^* - v_{jk}u_{jk}y_ku_{jk}^*v_{jk}^*|_2 \le \epsilon$. Set $d_{jk} = v_{jk}u_{jk}z_{k,\epsilon}u_{jk}^*v_{jk}^*$, and denote by G_k the set of all $Nk \times Nk$ matrices of the form

$$\begin{bmatrix} d_{1k} & v_{1k}\eta_{12}v_{2k}^* & \cdots & v_{1j}\eta_{1N}v_{Nk}^* \\ 0 & d_{2k} & \ddots & \vdots \\ \vdots & \ddots & \ddots & v_{(N-1),k}\eta_{N-1,N}v_{Nk}^* \\ 0 & \cdots & 0 & d_{Nk} \end{bmatrix}$$

where $\langle\eta_{ij}\rangle_{1\le i<j\le N} \in \Omega_k$. Notice that each d_{jk} is upper triangular and its eigenvalue distribution is exactly the same as that of $z_{k,\epsilon}$. For k sufficiently large, the set G_k lies in the ϵ-neighborhood of $\Gamma(Z;m,Nk,\gamma)$. Let $\theta(G_k)$ denote the unitary orbit of G_k in $M_{Nk}(\mathbb{C})$. We will now find lower bounds for the ϵ-packing numbers of $\theta(G_k)$ and thus, ones for $\Gamma(Z;m,Nk,\gamma)$.

Denote by $H_k \subset M_{Nk}(\mathbb{C})$ all matrices of the form

$$\begin{bmatrix} 0 & v_{1k}\eta_{12}v_{2k}^* & \cdots & v_{1j}\eta_{1N}v_{Nk}^* \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & v_{(N-1),k}\eta_{N-1,N}v_{Nk}^* \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

where $\langle\eta_{ij}\rangle_{1\le i<j\le N} \in \Omega_k$. Notice that H_k is isometric to the space of all matrices of the form

$$\begin{bmatrix} 0 & \eta_{12} & \cdots & \eta_{1N} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \eta_{N-1,N} \\ 0 & \cdot & \cdots & 0 \end{bmatrix}$$

where $\langle\eta_{ij}\rangle_{1\le i<j\le N} \in \Omega_k$. It follows that H_k must also have the same volume as the above subspace, computed in the obvious ambient Hilbert space of block upper triangular matrices obeying the above decomposition. Recall that for $n \in \mathbb{N}$, $T_n(\mathbb{C})$ denotes the set of uppertriangular matrices in $M_n(\mathbb{C})$; let $T_{n,<}(\mathbb{C})$ denote the matrices in $T_n(\mathbb{C})$ that have zero diagonal, i.e. the strictly upper triangular matrices in $M_n(\mathbb{C})$. Denote by W_k the subset of $T_{Nk,<}(\mathbb{C})$ consisting of all matrices x such that $|x|_2 < \epsilon$ and $x_{ij} = 0$ whenever $1 \le p < q \le N$ and $(p-1)k < i \le pk$ and $(q-1)k < j \le qk$. Thus, W_k consists of $N \times N$ diagonal matrices whose diagonal entries are strictly upper triangular $k \times k$ matrices. Denote by D_k the subset of diagonal matrices x of $M_{Nk}(\mathbb{C})$ such that $|x|_2 < \epsilon\sqrt{2}$. It follows that if f_k is the matrix

$$\begin{bmatrix} d_{1k} & 0 & \cdots & 0 \\ 0 & d_{2k} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & d_{Nk} \end{bmatrix}$$

then $f_k + D_k + W_k + H_k \subset \mathcal{N}_{3\epsilon}(G_k)$, where the 3ϵ neighborhood is taken in the ambient space $T_{Nk}(\mathbb{C})$ with respect to the metric induced by $|\cdot|_2$. Now observe that the space of diagonal $Nk \times Nk$ matrices and $T_{Nk,<}(\mathbb{C})$ are orthogonal subspaces of $T_{Nk}(\mathbb{C})$. Let $\theta_{3\epsilon}(G_k)$ denote the 3ϵ–neighborhood of the unitary orbit $\theta(G_k)$ of G_k. Thus, denoting by dX Lebesgue measure on $T_{Nk}(\mathbb{C})$ where $X = \langle x_{ij}\rangle_{1\le i\le j\le k}$, using Dyson's formula we have

$$\begin{aligned}
\mathrm{vol}(\theta_{3\epsilon}(G_k)) &\ge C_{Nk} \cdot \int_{f_k+D_k+W_k+H_k} \prod_{1\le i<j\le Nk} |x_{ii}-x_{jj}|^2 dX \\
&= C_{Nk}\cdot \mathrm{vol}(W_k+H_k)\cdot \int_{f_k+D_k} \prod_{1\le i<j\le Nk} |x_{ii}-x_{jj}|^2 dx_{11}\cdots dx_{(Nk)(Nk)} \\
&\ge C_{Nk}\cdot \mathrm{vol}(W_k+H_k)\cdot E_\epsilon(z_{k,\epsilon}\otimes I_N), \qquad (14)
\end{aligned}$$

where the constant C_{Nk} is as in 2 and where $\mathrm{vol}(\theta_{3\epsilon}(G_k))$ is computed in $M_{Nk}(\mathbb{C})$ and $\mathrm{vol}(W_k + H_k)$ is computed in $T_{Nk,<}(\mathbb{C})$, both being Euclidean volumes corresponding to the norms $(Nk)^{1/2}|\cdot|_2$. Clearly $\theta_{3\epsilon}(G_k) \subset \mathcal{N}_{4\epsilon}(\Gamma(Z; m, Nk, \gamma))$, so (14) gives a lower bound on $\mathrm{vol}(\mathcal{N}_{4\epsilon}(\Gamma(Z; m, Nk, \gamma))$ as well.
Using (14) and the standard volume comparison test, we have

$$\begin{aligned}
P_\epsilon(\Gamma(Z;m,Nk,\gamma)) &\ge \frac{\mathrm{vol}(\mathcal{N}_{4\epsilon}(\Gamma(Z;m,Nk,\gamma)))}{\mathrm{vol}(\mathcal{B}_{6\epsilon})} \\
&\ge C_{Nk}\cdot E_\epsilon(z_{k,\epsilon}\otimes I_N)\cdot \mathrm{vol}(W_k+H_k)\cdot \frac{\Gamma((Nk)^2+1)}{\pi^{(Nk)^2}(6(Nk)^{1/2}\epsilon)^{2(Nk)^2}},
\end{aligned}$$

where $\mathcal{B}_{6\epsilon}$ is a ball in $M_{Nk}(\mathbb{C})$ of radius 6ϵ with respect to $|\cdot|_2$, and we are computing volumes corresponding to the Euclidean norm $(Nk)^{1/2}|\cdot|_2$. Since W_k and H_k are orthogonal, we have $\mathrm{vol}(W_k + H_k) = \mathrm{vol}(W_k)\mathrm{vol}(H_k)$, where each volume is taken in the subspace of appropriate dimension. But W_k is a ball of radius $(Nk)^{1/2}\epsilon$ in a space of real dimension $Nk(k-1)$, so

$$\mathrm{vol}(W_k+H_k) = \frac{\pi^{\frac{Nk(k-1)}{2}}((Nk)^{1/2}\epsilon)^{Nk(k-1)}}{\Gamma(\frac{Nk(k-1)}{2}+1)}\cdot (N^{1/2})^{k^2N(N-1)}\mathrm{vol}(\Omega_k).$$

Applying Stirling's formula, we find

$$\begin{aligned}
\mathbb{P}_\epsilon(Z;m,\gamma) &\geq \liminf_{k\to\infty}(Nk)^{-2}\log P_\epsilon(\Gamma(Z;m,Nk,\gamma)) \\
&\geq \liminf_{k\to\infty}(Nk)^{-2}\log(E_\epsilon(z_{k,\epsilon}\otimes I_N)) \\
&\quad +\liminf_{k\to\infty}\Bigg((Nk)^{-2}\log(C_{Nk})+\frac{1}{2N}\log k+\frac{1}{N}\log\epsilon \\
&\qquad -\frac{1}{2N}\log(\frac{Nk(k-1)}{2})+\log((Nk)^2)-\log k \\
&\qquad -2\log\epsilon+(Nk)^{-2}\log(\mathrm{vol}(\Omega_k))\Bigg)+K_1 \\
&= \liminf_{k\to\infty}(Nk)^{-2}\log(E_\epsilon(z_{k,\epsilon}\otimes I_N)) \\
&\quad +\liminf_{k\to\infty}\left((Nk)^{-2}\log(C_{Nk})+\frac{1}{2}\log k\right) \\
&\quad +\liminf_{k\to\infty}\left((Nk)^{-2}\log(\mathrm{vol}(\Omega_k))+(\frac{1}{2}-\frac{1}{2N})\log k\right) \\
&\quad +(2-N^{-1})|\log\epsilon|+K_2 \\
&= \liminf_{k\to\infty}(Nk)^{-2}\log(E_\epsilon(z_{k,\epsilon}\otimes I_N)) \\
&\quad +N^{-2}\chi(\langle\mathrm{Re}B_{ij}\rangle_{1\leq i<j\leq N},\langle\mathrm{Im}B_{ij}\rangle_{1\leq i<j\leq N}) \\
&\quad +(2-N^{-1})|\log\epsilon|+K_3,
\end{aligned}$$

where K_1, K_2 and K_3 are constants independent of ϵ, m and γ. Taking $m\to\infty$ and $\gamma\to 0$, we get

$$\begin{aligned}
\mathbb{P}_\epsilon(Z) &\geq \liminf_{k\to\infty}(Nk)^{-2}\log(E_\epsilon(z_{k,\epsilon}\otimes I_N)) \\
&\quad +N^{-2}\chi(\langle\mathrm{Re}B_{ij}\rangle_{1\leq i<j\leq N},\langle\mathrm{Im}B_{ij}\rangle_{1\leq i<j\leq N}) \\
&\quad +(2-N^{-1})|\log\epsilon|+K_3.
\end{aligned}$$

Since the eigenvalue distribution of $z_{k,\epsilon}\otimes I_N$ converges as $k\to\infty$ to the measure σ_ϵ of Lemma 2.1, dividing by $|\log\epsilon|$ and applying Lemma 2.3 now yields

$$\delta_0(Z)=\limsup_{\epsilon\to 0}\frac{\mathbb{P}_\epsilon(Z)}{|\log\epsilon|}\geq\liminf_{\epsilon\to 0}\frac{f(\epsilon)}{|\log\epsilon|}+2-N^{-1}\geq 2-N^{-1}.$$

Since N was arbitrary, it follows that $\delta_0(Z)\geq 2$, thereby completing the proof. □

Acknowledgement. A significant part of this research was conducted during the 2004 Free Probabilty Workshop at the Banff International Research Station, and the authors would like to thank the organizers and sponsors for providing them with that opportunity to work together. K.D. would like to thank Joachim Cuntz, Siegfried Echterhoff and the Mathematics Institute at the Westfälische Wilhelms–Universität Münster for their generous hospitality during his year–long visit, which included much of the time he was working on this project.

References

[1] Aagaard, L. 'The non-microstates free entropy dimension of DT–operators' *J. Funct. Anal.* 213 (2004), 176-205.

[2] Aagaard, L. and Haagerup, U. 'Moment formulas for the quasi-nilpotent DT–operator', *Int. J. Math.* 15 (2004) 581-628.

[3] Belinschi, S.T. and Bercovici, H. 'A property of free entropy', *Pacific J. Math.* 211 (2003), 35-40.

[4] Biane, P., Capitaine, M. and Guionnet, A. 'Large deviation bounds for matrix Brownian motion', *Invent. Math.* 152 (2003), 433-459.

[5] Dykema, K. and Haagerup, U. 'Invariant subspaces of Voiculescu's circular operator', *Geom Funct. Anal.*, 11 (2001), 693-741.

[6] Dykema, K. and Haagerup, U. 'DT-operators and decomposability of Voiculescu's circular operator'. *Amer. J. Math*, 126 (2004), 121-189.

[7] Dykema, K. and Haagerup, U. 'Invariant subspaces of the quasinilpotent DT-operator', *J. Funct. Anal.* 209 (2004), 332-366.

[8] Jung, Kenley 'A free entropy dimension lemma'. *Pacific J. Math.* 177 (2003), 265-271.

[9] Jung, Kenley 'Fractal dimensions and entropies for microstate spaces', *J. Funct. Anal.* 222 (2005), 217-251.

[10] Jung, Kenley 'A hyperfinite inequality for free entropy dimension', to appear in *Proc. Amer. Math. Soc.*

[11] Mehta, M.L. *Random Matrices, second edition,* Academic Press, 1991.

[12] Shlyakhtenko, D. 'Random Gaussian band matrices and freeness with amalgamation', *Internat. Math. Res. Notices* 1996, 1013-1026.

[13] Śniady, P. 'Random regularization of Brown spectral measure', *J. Funct. Anal.* 193 (2002), 291-313.

[14] Śniady, P. 'Inequality for Voiculescu's free entropy in terms of Brown measure', *Int. Math. Res. Not.* 2003, 51–64.

[15] Śniady, P. 'Multinomial identities arising from the free probability', *J. Comb. Theory A* 101 (2003), 1-19.

[16] Voiculescu, D. 'The analogues of entropy and of Fisher's information measure in free probability theory, II'. *Invent. Math.* 118, (1994), 411-440.

[17] Voiculescu, D. 'The analogues of entropy and of Fisher's information measure in free probability theory III: The absence of Cartan subalgebras', *Geom. Funct. Anal.* 6 (1996), 172-199.

[18] Voiculescu, D. 'The analogues of entropy and of Fisher's information measure in free probabilility, V', *Invent. Math.* 132 (1998), 189–227.

[19] Voiculescu, D. 'A strengthened asymptotic freeness result for random matrices with applications to free entropy' *Internat. Math. Res. Notices* 1998, 41-64.

[20] Voiculescu, D. 'Free entropy'. *Bull. London Math. Soc.* 34 (2002), 257-332.

K. Dykema
Mathematisches Institut
Westfälische
Wilhelms–Universität Münster
Einsteinstr. 62,
48149 Münster, Germany
dykema@math.uni-muenster.de
PERMANENT ADDRESS:
Department of Mathematics
Texas A&M University
College Station, TX 77843-3368,
USA
kdykema@math.tamu.edu

K. Jung and D. Shlyakhtenko
Department of Mathematics
University of California
Los Angeles, CA 90095-3840,
USA
kjung@math.ucla.edu
shlyakht@math.ucla.edu

CM Points and Quaternion Algebras

C. Cornut and V. Vatsal

Received: December 23, 2004

Communicated by Peter Schneider

Abstract. This paper provides a proof of a technical result (Corollary 2.10 of Theorem 2.9) which is an essential ingredient in our proof of Mazur's conjecture over totally real number fields [3].

2000 Mathematics Subject Classification: 14G35, 11G18, 11G15
Keywords and Phrases: Shimura curves, CM points

Contents

1 Introduction

Let F be a totally real number field of degree d, and let B denote a quaternion algebra over F. For the purposes of this introduction, we assume that either:

- B is *definite*, meaning that $B_v = B \otimes F_v$ is non-split for all real places v of F, or
- B is *indefinite*, meaning that B_v is split for precisely one real v.

We shall write G to denote the algebraic group over $\mathbf{Q}$ whose points over a $\mathbf{Q}$-algebra A are the set $(B \otimes A)^\times$.

Now let K be an imaginary quadratic extension of F. We suppose that there is given an embedding $K \to B$. Then associated to the data of B and K, one can define a collection of points, the so-called CM points. The natural habitat for these points depends on whether B is definite or indefinite: in the former case, the CM points are just an infinite discrete set, whereas in the latter, they inhabit certain canonical algebraic curves, the *Shimura curves*, associated to the indefinite algebra B. Our goal in this paper is to study the distribution of these CM points in certain auxiliary spaces. The main result proven here is the key ingredient in our proof in [3] of certain non-vanishing theorems for certain automorphic L-functions over F and their derivatives. The theorems of [3] may be regarded as generalizations of Mazur's conjectures in [12] when $F = \mathbf{Q}$.

Our original intention was simply to write a single paper proving the non-vanishing theorems for the L-functions, using the connection between L-functions and CM points, and proving a basic nontriviality theorem for the latter. However, in the course of doing this, we realized that although the CM points in the definite and indefinite cases are *a priori* very different, the proof of the main nontriviality result on CM points runs along parallel lines. In light of this, it seemed somewhat artificial to give essentially the same arguments twice, once in each of the two cases. The present paper therefore presents a rather general result about CM points on quaternion algebras, which allows us to obtain information about CM points in both the definite and indefinite cases. The former case follows trivially, but the latter requires us to develop a certain amount of foundational material on Shimura curves, their various models, and the associated CM points.

Since this paper is neccessarily rather technical, we want to give an overview of the contents. The first part deals with the abstract results. The main theorems are given in Theorem 2.9 and Corollary 2.10. Although the statements are somewhat complicated, they are not hard to prove, in view of our earlier results [2], [18], [19], where all the main ideas are already present. As before, the basic ingredient is Ratner's theorem on unipotent flows on p-adic Lie groups.

The second part is concerned with the applications of the abstract result to CM points on Shimura curves. We start with basic theory of Shimura curves, especially their integral models and reduction. In Section 3.1.1, we define the CM points and supersingular points, and establish the basic fact that the reduction of a CM point at an inert prime is a supersingular point. The basic result on CM points on Shimura curves is stated in Theorem 3.5. Section 3.2 gives a series of group theoretic descriptions of the various sets and maps which appear in Theorem 3.5, thus reducing its proof to a purely group theoretical statement, which may be deduced from the results in the first part of this paper.

The final two sections in the paper are meant to shed some light on related topics: section 3.3.1 investigates the dependence of Shimura curves on a certain parameter $\epsilon = \pm 1$, while section 3.3.2 provides some insight on a certain subgroup of $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ which plays a prominent role in the statements of Theorem 3.5 and also appears in the André-Oort conjecture.

In conclusion, we mention that a fuller discussion of the circle of ideas and theorems that are the excuse for this paper may be found in the introduction of [3], where the main arithmetical applications are also spelled out. We would also like to thank Hee Oh for a number of useful conversations, and Nimish Shah for providing us with the proof of the crucial Lemma 2.30.

2 CM points on quaternion algebras

2.1 CM points, special points and reduction maps

We keep the following notations: F is a totally real number field, K is a totally imaginary quadratic extension of F and B is *any* quaternion algebra over F which is split by K. At this point we make no assumption on B at infinity. We fix once and for all, an F-embedding $\iota : K \hookrightarrow B$ and a prime P of F where B is split. We denote by $\varpi_P \in F_P^\times$ a local uniformizer at P.
For any quaternion algebra B' over F, we denote by $\mathrm{Ram}(B')$, $\mathrm{Ram}_f(B')$ and $\mathrm{Ram}_\infty(B')$ the set of places (resp. finite places, resp. archimedean places) of F where B' ramifies.

2.1.1 Quaternion algebras.

Let S be a finite set of finite places of F such that

S1 $\forall v \in S$, B is unramified at v.

S2 $|S| + |\mathrm{Ram}_f(B)| + [F : \mathbf{Q}]$ is even.

S3 $\forall v \in S$, v is inert or ramifies in K.

The first two assumptions imply that there exists a totally definite quaternion algebra B_S over F such that $\mathrm{Ram}_f(B_S) = \mathrm{Ram}_f(B) \cup S$. The third assumption implies that there exists an F-embedding $\iota_S : K \to B_S$. We choose such a pair (B_S, ι_S).

2.1.2 Algebraic groups

We put

$$G = \mathrm{Res}_{F/\mathbf{Q}}(B^\times),\ G_S = \mathrm{Res}_{F/\mathbf{Q}}(B_S^\times),$$

$$T = \mathrm{Res}_{F/\mathbf{Q}}(K^\times) \text{ and } Z = \mathrm{Res}_{F/\mathbf{Q}}(F^\times).$$

These are algebraic groups over $\mathbf{Q}$. We identify Z with the center of G and G_S. We use ι and ι_S to embed T as a maximal subtorus in G and G_S. We denote by $\mathrm{nr} : G \to Z$ and $\mathrm{nr}_S : G_S \to Z$ the algebraic group homomorphisms induced by the reduced norms $\mathrm{nr} : B^\times \to F^\times$ and $\mathrm{nr}_S : B_S^\times \to F^\times$.

2.1.3 Adelic groups

Let $\mathbf{A}_f$ denote the finite adeles of $\mathbf{Q}$. We shall consider the following locally compact, totally discontinuous groups:

- $G(\mathbf{A}_f) = (B \otimes_{\mathbf{Q}} \mathbf{A}_f)^\times$, $G_S(\mathbf{A}_f) = (B_S \otimes_{\mathbf{Q}} \mathbf{A}_f)^\times$, $T(\mathbf{A}_f) = (K \otimes_{\mathbf{Q}} \mathbf{A}_f)^\times$ and $Z(\mathbf{A}_f) = (F \otimes_{\mathbf{Q}} \mathbf{A}_f)^\times$ with their usual topology.

- $G(S) = \prod_{v \notin S} B_{S,v}^\times \times \prod_{v \in S} F_v^\times$ where $\prod_{v \notin S} B_{S,v}^\times$ is the restricted product of the $B_{S,v}^\times$'s over all finite places of F not in S, with respect to the compact subgroups $R_v^\times \subset B_{S,v}^\times$, where R_v is the closure in $B_{S,v}$ of some fixed $\mathcal{O}_F$-order R in B_S.

These groups are related by a commutative diagram of continuous morphisms:

$$\begin{tikzcd}
 & G(\mathbf{A}_f) \arrow[dr, "\phi_S"'] \arrow[drr, bend left, "\mathrm{nr}"] & & \\
T(\mathbf{A}_f) \arrow[ur] \arrow[dr] & & G(S) \arrow[r, "\mathrm{nr}'_S"] & Z(\mathbf{A}_f) \\
 & G_S(\mathbf{A}_f) \arrow[ur, "\pi_S"] \arrow[urr, bend right, "\mathrm{nr}_S"'] & &
\end{tikzcd}$$

In this diagram,

- $T(\mathbf{A}_f) \to G(\mathbf{A}_f)$ and $T(\mathbf{A}_f) \to G_S(\mathbf{A}_f)$ are the closed embeddings induced by ι and ι_S.
- $\mathrm{nr} : G(\mathbf{A}_f) \to Z(\mathbf{A}_f)$ and $\mathrm{nr}_S : G_S(\mathbf{A}_f) \to Z(\mathbf{A}_f)$ are the continuous, open and surjective group homomorphisms induced by nr and nr_S.
- $\mathrm{nr}'_S : G(S) \to Z(\mathbf{A}_f)$ is the continuous, open and surjective group homomorphism induced by $\mathrm{nr}_{S,v} : B_{S,v}^\times \to F_v^\times$ for $v \notin S$ and by the identity on the remaining factors.
- $\pi_S : G_S(\mathbf{A}_f) = \prod_{v \notin S} B_{S,v}^\times \times \prod_{v \in S} B_{S,v}^\times \to G(S) = \prod_{v \notin S} B_{S,v}^\times \times \prod_{v \in S} F_v^\times$ is the continuous, open and surjective group homomorphism induced by the identity on $\prod_{v \notin S} B_{S,v}^\times$ and by the reduced norms $\mathrm{nr}_{S,v} : B_{S,v}^\times \to F_v^\times$ on the remaining factors. It induces an isomorphism of topological groups between $G_S(\mathbf{A}_f)/\ker(\pi_S)$ and $G(S)$. Since $\ker(\pi_S) \simeq \prod_{v \in S} B_{S,v}^1$ is compact, π_S is also a closed map.

The definition of

$$\phi_S : G(\mathbf{A}_f) = \prod_{v \notin S} B_v^\times \times \prod_{v \in S} B_v^\times \to G(S) = \prod_{v \notin S} B_{S,v}^\times \times \prod_{v \in S} F_v^\times$$

is more involved. By construction, B_v and $B_{S,v}$ are isomorphic for $v \notin S$. We shall construct a collection of isomorphisms $(\phi_v : B_v \to B_{S,v})_{v \notin S}$ such that (1) $\forall v \notin S$, $\phi_v \circ \iota = \iota_S$ on K_v, and (2) the product of the ϕ_v's yields a continuous isomorphism between $\prod_{v \notin S} B_v^\times$ and $\prod_{v \notin S} B_{S,v}^\times$. Note that any two such families are conjugated by an element of $\prod_{v \notin S} K_v^\times$. Once such a family has been chosen, we may define the morphism ϕ_S by taking $\prod_{v \notin S} \phi_v$ on $\prod_{v \notin S} B_v^\times$ and $\mathrm{nr}_v : B_v^\times \to F_v^\times$ on the remaining factors. It is then a continuous, open and surjective group homomorphism which makes the above diagram commute.

We first fix a maximal $\mathcal{O}_F$-order R in B (respectively R_S in B_S). For all but finitely many v's, (a) $R_v \simeq M_2(\mathcal{O}_{F_v}) \simeq R_{S,v}$ and (b) $\iota^{-1}(R_v)$ and $\iota_S^{-1}(R_{S,v})$ are the maximal order of K_v. For such v's we may choose the isomorphism $\phi_v : R_v \xrightarrow{\simeq} R_{S,v}$ in such a way that $\phi_v \circ \iota = \iota_S$ on K_v. Indeed, starting with any isomorphism $\phi_v^? : R_v \to R_{S,v}$, we obtain two optimal embeddings $\phi_v^? \circ \iota$ and ι_S of $\mathcal{O}_{K_v}$ in $R_{S,v}$. By [20, Théorème 3.2 p. 44], any two such embeddings are conjugated by an element of $R_{S,v}^\times$: the corresponding conjugate of $\phi_v^?$ has the required property.
For those v's that satisfy (a) and (b), we thus obtain an isomorphism $\phi_v : B_v \to B_{S,v}$ such that $\phi_v(R_v) = R_{S,v}$ and $\phi_v \circ \iota = \iota_S$ on K_v. For the remaining v's not in S, we only require the second condition: $\phi_v \circ \iota = \iota_S$ on K_v. Such ϕ_v's do exists by the Skolem-Noether theorem [20, Théorème 2.1 p. 6]. The resulting collection $(\phi_v)_{v\notin S}$ satisfies (1) and (2).

2.1.4 Main objects

Definition 2.1 We define the space CM of *CM points*, the space $\mathcal{X}(S)$ of *special points at* S and the space $\mathcal{Z}$ of *connected components* by

$$\begin{aligned} \mathrm{CM} &= \overline{T(\mathbf{Q})}\backslash G(\mathbf{A}_f) \\ \mathcal{X}(S) &= \overline{G(S,\mathbf{Q})}\backslash G(S) \\ \mathcal{Z} &= \overline{Z(\mathbf{Q})^+}\backslash Z(\mathbf{A}_f) \end{aligned}$$

where $\overline{T(\mathbf{Q})}$ is the closure of $T(\mathbf{Q})$ in $T(\mathbf{A}_f)$, $\overline{G(S,\mathbf{Q})}$ is the closure of $G(S,\mathbf{Q}) = \pi_S(G_S(\mathbf{Q}))$ in $G(S)$ and $\overline{Z(\mathbf{Q})^+}$ is the closure of $Z(\mathbf{Q})^+ = F^{>0}$ in $Z(\mathbf{A}_f)$.

These are locally compact totally discontinuous Hausdorff spaces equipped with a right, continuous and transitive action of $G(\mathbf{A}_f)$ (with $G(\mathbf{A}_f)$ acting on $\mathcal{X}(S)$ through ϕ_S and on $\mathcal{Z}$ through nr). By [20, Théorème 1.4 pp. 61–64], $\mathcal{X}(S)$ and $\mathcal{Z}$ are *compact* spaces.

Definition 2.2 The *reduction map* RED_S *at* S, the *connected component map* c_S and their composite

$$c : \mathrm{CM} \xrightarrow{\mathrm{RED}_S} \mathcal{X}(S) \xrightarrow{c_S} \mathcal{Z}.$$

are respectively induced by

$$\mathrm{nr} : G(\mathbf{A}_f) \xrightarrow{\phi_S} G(S) \xrightarrow{\mathrm{nr}'_S} Z(\mathbf{A}_f).$$

Remark 2.3 Since $\phi_S(T(\mathbf{Q})) = \pi_S(T(\mathbf{Q})) \subset \pi_S(G_S(\mathbf{Q})) = G(S,\mathbf{Q})$, ϕ_S maps $\overline{T(\mathbf{Q})}$ to $\overline{G(S,\mathbf{Q})}$ and indeed induces a map $\mathrm{CM} \to \mathcal{X}(S)$. Similarly, c_S is well-defined since $\mathrm{nr}'_S(G(S,\mathbf{Q})) = \mathrm{nr}_S(G_S(\mathbf{Q})) = Z(\mathbf{Q})^+$ (by the norm theorem [20, Théorème 4.1 p. 80]).

It follows from the relevant properties of nr, ϕ_S and nr'_S that c, RED_S and c_S are continuous, open and surjective $G(\mathbf{A}_f)$-equivariant maps. Since $\mathcal{X}(S)$ is compact, c_S is also a closed map.

Remark 2.4 The terminology *CM points*, *special points* and *connected components* is motivated by the example of Shimura curves: see the second part of this paper, especially section 3.2.

2.1.5 Galois actions

The profinite commutative group $\overline{T(\mathbf{Q})}\backslash T(\mathbf{A}_f)$ acts continuously on CM, by multiplication on the *left*. This action is faithful and commutes with the right action of $G(\mathbf{A}_f)$. Using the inverse of Artin's reciprocity map $\mathrm{rec}_K : \overline{T(\mathbf{Q})}\backslash T(\mathbf{A}_f) \xrightarrow{\simeq} \mathrm{Gal}_K^{\mathrm{ab}}$, we obtain a continuous, $G(\mathbf{A}_f)$-equivariant and faithful action of $\mathrm{Gal}_K^{\mathrm{ab}}$ on CM.[1]
Similarly, Artin's reciprocity map $\mathrm{rec}_F : \overline{Z(\mathbf{Q})^+}\backslash Z(\mathbf{A}_f) \xrightarrow{\simeq} \mathrm{Gal}_F^{\mathrm{ab}}$ allows one to view $\mathcal{Z}$ as a principal homogeneous $\mathrm{Gal}_F^{\mathrm{ab}}$-space. From this point of view, $c : \mathrm{CM} \to \mathcal{Z}$ is a $\mathrm{Gal}_K^{\mathrm{ab}}$-equivariant map in the sense that for $x \in \mathrm{CM}$ and $\sigma \in \mathrm{Gal}_K^{\mathrm{ab}}$,

$$c(\sigma \cdot x) = \sigma\mid_{F^{\mathrm{ab}}} \cdot c(x).$$

2.1.6 Further objects

For technical purposes, we will also need to consider the following objects:

- $\mathcal{X}_S = \overline{G_S(\mathbf{Q})}\backslash G_S(\mathbf{A}_f)$, where $\overline{G_S(\mathbf{Q})}$ is the closure of $G_S(\mathbf{Q})$ in $G_S(\mathbf{A}_f)$.
- $q_S : \mathcal{X}_S \to \mathcal{X}(S)$ is induced by $\pi_S : G_S(\mathbf{A}_f) \to G(S)$.

The composite map $c_S \circ q_S : \mathcal{X}_S \to \mathcal{Z}$ is induced by $\mathrm{nr}_S : G_S(\mathbf{A}_f) \to Z(\mathbf{A}_f)$. By [20, Théorème 1.4 p. 61], $\mathcal{X}_S$ is compact. Note that q_S is indeed well defined since $\pi_S(G_S(\mathbf{Q})) = G(S,\mathbf{Q})$. In fact, $\pi_S(\overline{G_S(\mathbf{Q})}) = \overline{G(S,\mathbf{Q})}$ since π_S is a closed map: the fibers of q_S are the $\ker(\pi_S)$-orbits in $\mathcal{X}_S$. In particular, q_S yields a $G(S)$-equivariant homeomorphism between $\mathcal{X}_S/\ker(\pi_S)$ and $\mathcal{X}(S)$.

2.1.7 Measures

The group $G^1(S) = \ker(\mathrm{nr}'_S)$ (resp. $G^1_S(\mathbf{A}_f) = \ker(\mathrm{nr}_S)$) acts on the fibers of c_S (resp. $c_S \circ q_S$). In section 2.4.1 below, we shall prove the following proposition. Recall that a Borel probability measure on a topological space is a measure defined on its Borel subsets which assigns voume 1 to the total space.

[1] This action extends to a continuous, $G(\mathbf{A}_f)$-equivariant action of $\mathrm{Gal}(K^{\mathrm{ab}}/F)$ as follows. By the Skolem-Noether theorem, there exists an element $b \in B^\times$ such that $x \mapsto x^b = b^{-1}xb$ induces the non-trivial F-automorphism of K. In particular, b^2 belongs to $T(\mathbf{Q})$. Multiplication on the left by b induces an involution ι on CM such that for all $x \in \mathrm{CM}$ and $\sigma \in \mathrm{Gal}_K^{\mathrm{ab}}$, $\iota(\sigma x) = \sigma^\iota \iota x$ where $\sigma \mapsto \sigma^\iota$ is the involution on $\mathrm{Gal}_K^{\mathrm{ab}}$ which is induced by the nontrivial element of $\mathrm{Gal}(K/F)$.

Proposition 2.5 *The above actions are transitive and for each* $z \in \mathcal{Z}$, (1) *there exists a unique* $G_S^1(\mathbf{A}_f)$*-invariant Borel probability measure* μ_z *on* $(c_S \circ q_S)^{-1}(z)$*, and* (2) *there exists a unique* $G^1(S)$*-invariant Borel probability measure* μ_z *on* $c_S^{-1}(z)$.

The uniqueness implies that these two measures are compatible, in the sense that the (proper) map $q_S : (c_S \circ q_S)^{-1}(z) \to c_S^{-1}(z)$ maps one to the other: this is why we use the same notation μ_z for both measures. Similarly, for any $g \in G_S^1(\mathbf{A}_f)$ (resp. $G(S)$), the measure $\mu_{z \cdot g}(\star g)$ equals μ_z on $(c_S \circ q_S)^{-1}(z)$ (resp. on $c_S^{-1}(z)$).

2.1.8 Level structures

For a compact open subgroup H of $G(\mathbf{A}_f)$, we denote by CM_H, $\mathcal{X}_H(S)$ and $\mathcal{Z}_H$ the quotients of CM, $\mathcal{X}(S)$ and $\mathcal{Z}$ by the right action of H. We still denote by c, RED_S and c_S the induced maps on these quotient spaces:

$$c : \mathrm{CM}_H \xrightarrow{\mathrm{RED}_S} \mathcal{X}_H(S) \xrightarrow{c_S} \mathcal{Z}_H.$$

Note that $\mathcal{X}_H(S)$ and $\mathcal{Z}_H$ are *finite* spaces, being discrete and compact. We have

$$\mathcal{Z}_H = \mathcal{Z}/\mathrm{nr}(H) \quad \text{and} \quad \mathcal{X}_H(S) = \mathcal{X}(S)/H(S) \simeq \mathcal{X}_S/H_S$$

where $H(S) = \phi_S(H) \subset G(S)$ and $H_S = \pi_S^{-1}(H(S)) \subset G_S$. The Galois group $\mathrm{Gal}_K^{\mathrm{ab}}$ still acts continuously on the (now discrete) spaces CM_H and $\mathcal{Z}_H$, and c is a $\mathrm{Gal}_K^{\mathrm{ab}}$-equivariant map.

2.2 Main theorems: the statements

2.2.1 Simultaneous reduction maps

Let $\mathfrak{S}$ be a nonempty finite collection of finite sets of non-archimedean places of F not containing P and satisfying conditions S1 to S3 of section 2.1.1. That is: each element of $\mathfrak{S}$ is a finite set S of finite places of F such that $\forall v \in S$, v is not equal to P, K_v is a field, and B_v is split, and $|S| + |\mathrm{Ram}_f(B)| + [F : \mathbf{Q}]$ is even. For each S in $\mathfrak{S}$, we choose a totally definite quaternion algebra B_S over F with $\mathrm{Ram}_f(B_S) = \mathrm{Ram}_f(B) \cup S$, an embedding $\iota_S : K \to B_S$ and a collection of isomorphisms $(\phi_v : B_v \to B_{S,v})_{v \notin S}$ as in section 2.1.3.
For each S in $\mathfrak{S}$, we thus obtain (among other things) an algebraic group G_S over $\mathbf{Q}$, two locally compact and totally discontinuous adelic groups $G_S(\mathbf{A}_f)$ and $G(S)$, a commutative diagram of continuous homomorphisms as in Section 2.1.3, a *special set* $\mathcal{X}(S) = \overline{G(S,\mathbf{Q})}\backslash G(S)$, a *reduction map* $\mathrm{RED}_S : \mathrm{CM} \to \mathcal{X}(S)$ and a *connected component* map $c_S : \mathcal{X}(S) \to \mathcal{Z}$ with the property that each fiber $c_S^{-1}(z)$ of c_S has a unique Borel probability measure μ_z which is right invariant under $G^1(S) = \ker(\mathrm{nr}'_S)$ (we refer the reader to section 2.1 for all notations).

Let $\mathfrak{R}$ be a nonempty finite subset of $\mathrm{Gal}_K^{\mathrm{ab}}$ and consider the sequence

$$\mathrm{CM} \xrightarrow{\mathrm{RED}} \mathcal{X}(\mathfrak{S},\mathfrak{R}) \xrightarrow{C} \mathcal{Z}(\mathfrak{S},\mathfrak{R})$$

where

- $\mathcal{X}(\mathfrak{S}) = \prod_{S\in\mathfrak{S}} \mathcal{X}(S)$ and $\mathcal{X}(\mathfrak{S},\mathfrak{R}) = \prod_{\sigma\in\mathfrak{R}} \mathcal{X}(\mathfrak{S}) = \prod_{S,\sigma} \mathcal{X}(S)$;
- $\mathcal{Z}(\mathfrak{S}) = \prod_{S\in\mathfrak{S}} \mathcal{Z}$ and $\mathcal{Z}(\mathfrak{S},\mathfrak{R}) = \prod_{\sigma\in\mathfrak{R}} \mathcal{Z}(\mathfrak{S}) = \prod_{S,\sigma} \mathcal{Z}$;
- $C : \mathcal{X}(\mathfrak{S},\mathfrak{R}) \to \mathcal{Z}(\mathfrak{S},\mathfrak{R})$ maps $x = (x_{S,\sigma})$ to $C(x) = (c_S(x_{S,\sigma}))$;
- $\mathrm{RED} : \mathrm{CM} \to \mathcal{X}(\mathfrak{S},\mathfrak{R})$ is the *simultaneous reduction map* which sends x to $\mathrm{RED}(x) = (\mathrm{RED}_S(\sigma \cdot x))$.

We also put $G(\mathfrak{S},\mathfrak{R}) = \prod_{S,\sigma} G(S)$ and $G^1(\mathfrak{S},\mathfrak{R}) = \prod_{S,\sigma} G^1(S)$, so that $G(\mathfrak{S},\mathfrak{R})$ acts on $\mathcal{X}(\mathfrak{S},\mathfrak{R})$ and $\mathcal{Z}(\mathfrak{S},\mathfrak{R})$, C is equivariant for these actions and its fibers are the $G^1(\mathfrak{S},\mathfrak{R})$-orbits in $\mathcal{X}(\mathfrak{S},\mathfrak{R})$. For $z = (z_{S,\sigma})$ in $\mathcal{Z}(\mathfrak{S},\mathfrak{R})$, the measure $\mu_z = \prod_{S,\sigma} \mu_{z_{S,\sigma}}$ is a $G^1(\mathfrak{S},\mathfrak{R})$-invariant Borel probability measure on $C^{-1}(z) = \prod_{S,\sigma} c_S^{-1}(z_{S,\sigma})$. If $g \in G(\mathfrak{S},\mathfrak{R})$ and $z \in \mathcal{Z}(\mathfrak{S},\mathfrak{R})$, $\mu_{z\cdot g}(\star g) = \mu_z$ on $C^{-1}(z)$.
The Galois group $\mathrm{Gal}_K^{\mathrm{ab}}$ acts diagonally on $\mathcal{Z}(\mathfrak{S},\mathfrak{R}) = \prod_{S,\sigma} \mathcal{Z}$ (through its quotient $\mathrm{Gal}_F^{\mathrm{ab}}$) and the composite map $C \circ \mathrm{RED} : \mathrm{CM} \to \mathcal{Z}(\mathfrak{S},\mathfrak{R})$ is $\mathrm{Gal}_K^{\mathrm{ab}}$-equivariant. For $x \in \mathrm{CM}$, we shall frequently write $\bar{x} = C \circ \mathrm{RED}(x)$. Explicitly:

$$\bar{x} = C \circ \mathrm{RED}(x) = (\sigma \cdot c(x))_{S,\sigma} \in \mathcal{Z}(\mathfrak{S},\mathfrak{R}) = \prod_{S,\sigma} \mathcal{Z}.$$

2.2.2 Main theorem

In this section, we state the main results, without proofs. The proofs are long, and will be given later.

Definition 2.6 A *P-isogeny class* of CM points is a $B_P^\times$-orbit in CM. If $\mathcal{H} \subset \mathrm{CM}$ is a P-isogeny class and f is a $\mathbf{C}$-valued function on CM, we say that *$f(x)$ goes to $a \in \mathbf{C}$ as x goes to infinity in $\mathcal{H}$* if the following holds: for any $\epsilon > 0$, there exists a compact subset $C(\epsilon)$ of CM such that $|f(x) - a| \le \epsilon$ for all $x \in \mathcal{H} \setminus C(\epsilon)$.

Remark 2.7 This definition can be somewhat clarified if we introduce the Alexandroff "one point" compactification $\widehat{\mathrm{CM}} = \mathrm{CM} \cup \{\infty\}$ of the locally compact space CM. It is easy to see that the point $\infty \in \widehat{\mathrm{CM}}$ lies in the closure of any P-isogeny class $\mathcal{H}$ (simply because P-isogeny classes are not relatively compact in CM). Our definition of "$f(x)$ goes to $a \in \mathbf{C}$ as x goes to infinity in $\mathcal{H}$" is then equivalent to the assertion that the limit of $f\,|_{\mathcal{H}}$ at ∞ exists and equals a.

Definition 2.8 An element $\sigma \in \mathrm{Gal}_K^{\mathrm{ab}}$ is *P-rational* if $\sigma = \mathrm{rec}_K(\lambda)$ for some $\lambda \in \widehat{K}^\times$ whose P-component λ_P belongs to the subgroup $K^\times \cdot F_P^\times$ of $K_P^\times$. We denote by $\mathrm{Gal}_K^{P-\mathrm{rat}} \subset \mathrm{Gal}_K^{\mathrm{ab}}$ the subgroup of all P-rational elements.

In the above definition, $\mathrm{rec}_K : \widehat{K}^\times \twoheadrightarrow \mathrm{Gal}_K^{\mathrm{ab}}$ is Artin's reciprocity map. We normalize the latter by specifying that it sends local uniformizers to *geometric* Frobeniuses.

Theorem 2.9 *Suppose that the finite subset $\mathfrak{R}$ of $\mathrm{Gal}_K^{\mathrm{ab}}$ consists of elements which are pairwise distinct modulo $\mathrm{Gal}_K^{P-\mathrm{rat}}$. Let $\mathcal{H} \subset \mathrm{CM}$ be a P-isogeny class and let $\mathcal{G}$ be a compact open subgroup of $\mathrm{Gal}_K^{\mathrm{ab}}$ with Haar measure dg. Then for every continuous function $f : \mathcal{X}(\mathfrak{S}, \mathfrak{R}) \to \mathbf{C}$,*

$$x \mapsto \int_{\mathcal{G}} f \circ \textsc{Red}(g \cdot x) dg - \int_{\mathcal{G}} dg \int_{C^{-1}(g \cdot \bar{x})} f d\mu_{g \cdot \bar{x}}$$

goes to 0 as x goes to infinity in $\mathcal{H}$.

2.2.3 Surjectivity

Let H be a compact open subgroup of $G(\mathbf{A}_f)$. Replacing CM, $\mathcal{X}$ and $\mathcal{Z}$ by CM_H, $\mathcal{X}_H$ and $\mathcal{Z}_H$ in the constructions of section 2.2.1, we obtain a sequence

$$\mathrm{CM}_H \xrightarrow{\textsc{Red}} \mathcal{X}_H(\mathfrak{S}, \mathfrak{R}) \xrightarrow{C} \mathcal{Z}_H(\mathfrak{S}, \mathfrak{R})$$

where

- $\mathcal{X}_H(\mathfrak{S}, \mathfrak{R}) = \prod_{S,\sigma} \mathcal{X}_H(S) = \mathcal{X}(\mathfrak{S}, \mathfrak{R})/H(\mathfrak{S}, \mathfrak{R})$ and
- $\mathcal{Z}_H(\mathfrak{S}, \mathfrak{R}) = \prod_{S,\sigma} \mathcal{Z}_H = \mathcal{Z}(\mathfrak{S}, \mathfrak{R})/H(\mathfrak{S}, \mathfrak{R})$ with
- $H(\mathfrak{S}, \mathfrak{R}) = \prod_{S,\sigma} H(S)$, a compact open subgroup of $G(\mathfrak{S}, \mathfrak{R})$.

Applying the main theorem to the characteristic functions of the (finitely many) elements of $\mathcal{X}_H(\mathfrak{S}, \mathfrak{R})$, we obtain the following surjectivity result. Let $\overline{\mathcal{H}}$ be the image of $\mathcal{H}$ in CM_H.

Corollary 2.10 *For all but finitely many x in $\overline{\mathcal{H}}$,*

$$\textsc{Red}(\mathcal{G} \cdot x) = C^{-1}(\mathcal{G} \cdot \bar{x}) \quad \text{in } \mathcal{X}_H(\mathfrak{S}, \mathfrak{R})$$

where $\bar{x} = C \circ \textsc{Red}(x) \in \mathcal{Z}_H(\mathfrak{S}, \mathfrak{R})$.

2.2.4 Equidistribution

When $H = \widehat{R}^\times$ for some Eichler order R in B, we can furthermore specify the asymptotic behavior (as x varies inside $\overline{\mathcal{H}}$) of

$$\mathrm{Prob}\left\{\mathrm{Red}(\mathcal{G}\cdot x) = s\right\} \stackrel{\mathrm{def}}{=} \frac{1}{|\mathcal{G}\cdot x|}\left|\{g\cdot x;\, \mathrm{Red}(g\cdot x) = s,\, g\in\mathcal{G}\}\right|$$

where s is a fixed point in $\mathcal{X}_H(\mathfrak{S},\mathfrak{R})$. To state our result, we first need to define a few constants.
Let $\mathcal{N} = \prod_Q Q^{n_Q}$ be the level of R. By construction, the compact open subgroup $H_S = \pi_S^{-1}\phi_S(H)$ of $G_S(\mathbf{A}_f)$ equals $\widehat{R}_S^\times$ for some Eichler order $R_S \subset B_S$ whose level $\mathcal{N}_S$ is the "prime-to-S" part of $\mathcal{N}$: $\mathcal{N}_S = \prod_{Q\notin S} Q^{n_Q}$. For $g \in G_S(\mathbf{A}_f)$ and $x = G_S(\mathbf{Q})gH_S$ in

$$\mathcal{X}_H(S) \simeq \mathcal{X}_S/H_S = \overline{G_S(\mathbf{Q})}\backslash G_S(\mathbf{A}_f)/H_S = G_S(\mathbf{Q})\backslash G_S(\mathbf{A}_f)/H_S$$

put $\mathcal{O}(g) = g\widehat{R}_S g^{-1}\cap B_S$. This is an $\mathcal{O}_F$-order in B_S whose $B_S^\times$-conjugacy class depends only upon x. The isomorphism class of the group $\mathcal{O}(g)^\times/\mathcal{O}_F^\times$ also depends only upon x and since B_S is totally definite, this group is finite [20, p. 139]. The *weight* $\omega(x)$ of x is the order of this group: $\omega(x) = [\mathcal{O}(g)^\times : \mathcal{O}_F^\times]$. The *weight* of an element $s = (x_{S,\sigma})$ in $\mathcal{X}_H(\mathfrak{S},\mathfrak{R})$ is then given by $\omega(s) = \prod_{S,\sigma}\omega(x_{S,\sigma})$.
Finally, we put

$$\Omega = \frac{1}{\Omega(\mathcal{G})}\cdot\left(\prod_{S\in\mathfrak{S}}\frac{\Omega(F)}{\Omega(B_S)\cdot\Omega(\mathcal{N}_S)}\right)^{|\mathfrak{R}|}$$

where

- $\Omega(F) = 2^{2[F:\mathbf{Q}]-1}[\mathcal{O}_F^\times : \mathcal{O}_F^{>0}]^{-1}\,|\zeta_F(-1)|^{-1}$,
- $\Omega(B_S) = \prod_{Q\in\mathrm{Ram}_f(B_S)}(\|Q\|-1)$,
- $\Omega(\mathcal{N}_S) = \|\mathcal{N}_S\|\cdot\prod_{Q|\mathcal{N}_S}(\|Q\|^{-1}+1)$ and
- $\Omega(\mathcal{G})$ is the order of the image of $\mathcal{G}$ in the Galois group $\mathrm{Gal}(F_1^+/F)$ of the narrow Hilbert class field F_1^+ of F.

Here $\|\cdot\|$ denotes the absolute norm.

Corollary 2.11 *For all $\epsilon > 0$, there exists a finite set $\overline{\mathcal{C}}(\epsilon)\subset\overline{\mathcal{H}}$ such that for all $s\in\mathcal{X}_H(\mathfrak{S},\mathfrak{R})$ and $x\in\overline{\mathcal{H}}\setminus\overline{\mathcal{C}}(\epsilon)$,*

$$\left|\mathrm{Prob}\left\{\mathrm{Red}(\mathcal{G}\cdot x) = s\right\} - \frac{\Omega}{\omega(s)}\right| \le \epsilon$$

if s belongs to $C^{-1}(\mathcal{G}\cdot\bar{x})$ and $\mathrm{Prob}\left\{\mathrm{Red}(\mathcal{G}\cdot x) = s\right\} = 0$ *otherwise.*

The remainder of this first part of the paper is devoted to the proofs of Proposition 2.5, Theorem 2.9, Corollary 2.10, Corollary 2.11.

2.3 Proof of the main theorems: first reductions

Notations

For a continuous function $f : \mathcal{X}(\mathfrak{S},\mathfrak{R}) \to \mathbf{C}$ and $x \in \mathrm{CM}$, we put

$$A(f,x) = \int_{\mathcal{G}} f \circ \mathrm{Red}(g\cdot x)dg \quad \text{and} \quad B(f,x) = B(f,\bar{x}) = \int_{\mathcal{G}} I(f, g\cdot\bar{x})dg$$

where $\bar{x} = C \circ \mathrm{Red}(x)$, with $I(f,z) = \int_{C^{-1}(z)} f d\mu_z$ for $z \in \mathcal{Z}(\mathfrak{S},\mathfrak{R})$.
Then the theorem says that for all $\epsilon > 0$, there exists a compact subset $\mathcal{C}(\epsilon) \subset \mathrm{CM}$ such that,

$$\forall x \in \mathcal{H},\ x \notin \mathcal{C}(\epsilon): \quad |A(f,x) - B(f,x)| \leq \epsilon.$$

We claim that the functions $x \mapsto A(f,x)$ and $x \mapsto B(f,x)$ are well-defined. This is clear for $A(f,x)$, as $g \mapsto f \circ \mathrm{Red}(g\cdot x)$ is continuous on $\mathcal{G}$. For $B(f,x)$, we claim that $g \mapsto I(f, g\cdot\bar{x})$ is also continuous. Since $g \mapsto g\cdot\bar{x}$ is continuous, it is sufficient to show that $z \mapsto I(f,z)$ is continuous on $\mathcal{Z}(\mathfrak{S},\mathfrak{R})$. Note that for $u \in G(\mathfrak{S},\mathfrak{R})$,

$$I(f,z\cdot u) - I(f,z) = \int_{C^{-1}(z\cdot u)} f d\mu_{z\cdot u} - \int_{C^{-1}(z)} f d\mu_z = \int_{C^{-1}(z)} \left(f(\star u) - f\right) d\mu_z.$$

Since f is continuous and $\mathcal{X}(\mathfrak{S},\mathfrak{R})$ is compact, f is uniformly continuous. It follows that $I(f,z\cdot u) - I(f,z)$ is small when u is small and $z \mapsto I(f,z)$ is indeed continuous.
To prove the theorem, we may assume that f is locally constant. Indeed, there exists a locally constant function $f' : \mathcal{X}(\mathfrak{S},\mathfrak{R}) \to \mathbf{C}$ such that $\|f - f'\| \leq \epsilon/3$. If the theorem were known for f', we could find a compact subset $\mathcal{C}(\epsilon) \subset \mathrm{CM}$ such that $|A(f',x) - B(f',x)| \leq \epsilon/3$ for all $x \in \mathcal{H}$ with $x \notin \mathcal{C}(\epsilon)$, thus obtaining

$$\begin{aligned} &|A(f,x) - B(f,x)| \\ &\quad\leq\ |A(f,x) - A(f',x)| + |A(f',x) - B(f',x)| + |B(f',x) - B(f,x)| \\ &\quad\leq\ \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

A decomposition of $\mathcal{G}\cdot\mathcal{H}\cdot H$

From now on, we shall thus assume that f is locally constant. Let H be a compact open subgroup of $G(\mathbf{A}_f)$ such that f factors through $\mathcal{X}(\mathfrak{S},\mathfrak{R})/H(\mathfrak{S},\mathfrak{R})$, where $H(\mathfrak{S},\mathfrak{R}) = \prod_{S,\sigma} H(S)$ with $H(S) = \phi_S(H)$. Then

- $x \mapsto A(x) = \int_{\mathcal{G}} f \circ \mathrm{Red}(g\cdot x)dg$ factors through $\mathcal{G}\backslash\mathrm{CM}/H$,
- $z \mapsto I(z) = \int_{C^{-1}(z)} f d\mu_z$ factors through $\mathcal{Z}(\mathfrak{S},\mathfrak{R})/H(\mathfrak{S},\mathfrak{R})$, hence
- $x \mapsto B(x) = B(\bar{x}) = \int_{\mathcal{G}} I(g\cdot\bar{x})dg$ factors through $\mathcal{G}\backslash\mathrm{CM}/H$

(where $\bar{x} = C \circ \textsc{Red}(x) \in \mathcal{Z}(\mathfrak{S}, \mathfrak{R})$ as usual).
For a nonzero nilpotent element $N \in B_P$, the formula $u(t) = 1 + tN$ defines a group isomorphism $u : F_P \to U = u(F_P) \subset B_P^\times$. We say that $U = \{u(t)\}$ is a one parameter unipotent subgroup of $B_P^\times$.

Proposition 2.12 *There exists:* (1) *a* finite *set* $\mathcal{I}$, (2) *for each* $i \in \mathcal{I}$, *a point* $x_i \in \mathcal{H}$ *and a one parameter unipotent subgroup* $U_i = \{u_i(t)\}$ *of* $B_P^\times$, *and* (3) *a compact open subgroup* κ *of* $F_P^\times$ *such that*

1. $\mathcal{G} \cdot \mathcal{H} \cdot H = \bigcup_{i \in \mathcal{I}} \bigcup_{n \geq 0} \mathcal{G} \cdot x_i \cdot u_i(\kappa_n) \cdot H$, *and*
2. $\forall i \in \mathcal{I}$ *and* $\forall n \geq 0$, $\mathcal{G} \cdot x_i \cdot u_i(\kappa_n) \cdot H = \mathcal{G} \cdot x_i u_{i,n} \cdot H$,

where $\kappa_n = \varpi_P^{-n} \kappa \subset F_P^\times$ *and* $u_{i,n} = u_i(\varpi_P^{-n}) \in u_i(\kappa_n)$.

Proof. Section 2.6.

Unipotent orbits: reduction of Theorem 2.9

This decomposition allows us to switch from Galois (=toric) orbits to unipotent orbits of CM points. To deal with the latter, we have the following proposition. We fix a CM point $x \in \mathcal{H}$ and a one parameter unipotent subgroup $U = \{u(t)\}$ in $B_P^\times$. We also choose a Haar measure $\lambda = dt$ on F_P. Then Theorem 2.9 follows from Proposition 2.12 and

Proposition 2.13 *Under the assumptions of Theorem 2.9, for almost all* $g \in \mathrm{Gal}_K^{ab}$,

$$\lim_{n \to \infty} \frac{1}{\lambda(\kappa_n)} \int_{\kappa_n} f \circ \textsc{Red}(g \cdot x \cdot u(t)) dt = \int_{C^{-1}(g \cdot \bar{x})} f d\mu_{g \cdot \bar{x}}.$$

Proof. Section 2.5.

To deduce Theorem 2.9, we may argue as follows. By taking the integral over $g \in \mathcal{G}$ and using (a) Lebesgue's dominated convergence theorem to exchange $\int_{\mathcal{G}}$ and $\lim_n$, and (b) Fubini's theorem to exchange $\int_{\mathcal{G}}$ and $\int_{\kappa_n}$, we obtain:

$$\lim_{n \to \infty} \frac{1}{\lambda(\kappa_n)} \int_{\kappa_n} A(x \cdot u(t)) dt = B(x).$$

This holds for all x and u.
Then for $x = x_i$ and $u = u_i$, we also know from part (2) of Proposition 2.12 that $t \mapsto A(x_i \cdot u_i(t))$ is constant on κ_n, equal to $A(x_i u_{i,n})$. In particular,

$$\forall i \in \mathcal{I} : \quad \lim_{n \to \infty} A(x_i u_{i,n}) = B(x_i).$$

Fix $\epsilon > 0$ and choose $N \geq 0$ such that

$$\forall n > N, \forall i \in \mathcal{I} : \quad |A(x_i u_{i,n}) - B(x_i)| \leq \epsilon.$$

Put $\mathcal{C}(\epsilon) = \bigcup_{i\in\mathcal{I}} \bigcup_{n=0}^{N} \mathcal{G}\cdot x_i u_i(\kappa_n)\cdot H$, a compact subset of CM.
For any $x \in \mathcal{H}$, there exists $i \in \mathcal{I}$ and $n \geq 0$ such that x belongs to $\mathcal{G}\cdot x_i u_{i,n} H$, so that $A(x) = A(x_i u_{i,n})$ and $B(x) = B(x_i u_{i,n}) = B(x_i)$. If $x \notin \mathcal{C}(\epsilon)$, $n > N$ and $|A(x) - B(x)| \leq \epsilon$, QED.

Reduction of Corollaries 2.10 and 2.11

Let H be a compact open subgroup of $G(\mathbf{A}_f)$ and let $f : \mathcal{X}(\mathfrak{S},\mathfrak{R}) \to \{0,1\}$ be the characteristic function of some $s \in \mathcal{X}_H(\mathfrak{S},\mathfrak{R})$, say $s = \widetilde{s}\cdot H(\mathfrak{S},\mathfrak{R})$ with $\widetilde{s} \in \mathcal{X}(\mathfrak{S},\mathfrak{R})$. The function $z \mapsto I(f,z) = \int_{C^{-1}(z)} f d\mu_z$ factors through $\mathcal{Z}_H(\mathfrak{S},\mathfrak{R})$ and equals 0 outside $C(s) = C(\widetilde{s})\cdot H(\mathfrak{S},\mathfrak{R})$. Let $I(s)$ be its value on $C(s)$.
For $\widetilde{x} \in \mathrm{CM}$, we easily obtain:

- $A(f,\widetilde{x}) = \mathrm{Prob}\,\{\mathrm{Red}(\mathcal{G}\cdot x) = s\}$ where x is the image of $\widetilde{x}$ in CM_H,
- $B(f,\widetilde{x}) = 0$ if $\bar{x} = C\circ \mathrm{Red}(x)$ does not belong to $\mathcal{G}\cdot C(s)$, and
- $B(f,\widetilde{x}) = I(s)/\Omega(\mathcal{G},H)$ otherwise, where $\Omega(\mathcal{G},H)$ is the common size of all $\mathcal{G}$-orbits in $\mathcal{Z}(\mathfrak{S},\mathfrak{R})/H(\mathfrak{S},\mathfrak{R}) \simeq \prod_{S,\sigma} \mathcal{Z}/\mathrm{nr}(H)$, which is also the size of the $\mathcal{G}$-orbits in $\mathcal{Z}/\mathrm{nr}(H)$.

If $\mathrm{nr}(H)$ is the maximal compact subgroup $\widehat{\mathcal{O}}_F^{\times}$ of $Z(\mathbf{A}_f)$ (which occurs when $H = \widehat{R}^{\times}$ for some Eichler order $R \subset B$), $\mathcal{Z}/\mathrm{nr}(H) \simeq \mathrm{Gal}(F_1^+/F)$ and $\Omega(\mathcal{G},H)$ is the order of the image of $\mathcal{G}$ in $\mathrm{Gal}(F_1^+/F)$: $\Omega(\mathcal{G},H) = \Omega(\mathcal{G})$.
The main theorem asserts that for all $\epsilon > 0$, there exists a compact subset $\mathcal{C}(\epsilon)$ of CM such that for all $x \in \overline{\mathcal{H}} \setminus \overline{\mathcal{C}}(\epsilon)$ (where $\overline{\mathcal{H}}$ and $\overline{\mathcal{C}}(\epsilon)$ are the images of $\mathcal{H}$ and $\mathcal{C}(\epsilon)$ in CM_H),

$$|\mathrm{Prob}\,\{\mathrm{Red}(\mathcal{G}\cdot x) = s\} - I(s)/\Omega(\mathcal{G},H)| \leq \epsilon$$

if $s \in C^{-1}(\mathcal{G}\cdot\bar{x})$ and $\mathrm{Prob}\,\{\mathrm{Red}(\mathcal{G}\cdot x) = s\} = 0$ otherwise. Note that $\overline{\mathcal{C}}(\epsilon)$ is finite, being compact and discrete. To prove the corollaries, it remains to (1) show that $I(s)$ is nonzero and (2) compute $I(s)$ exactly when H arises from an Eichler order in B.
Write $s = (x_{S,\sigma})$ with $x_{S,\sigma} = \widetilde{x}_{S,\sigma} H_S$ in $\mathcal{X}(S)/H(S) \simeq \mathcal{X}_S/H_S$ ($S \in \mathfrak{S}$, $\sigma \in \mathfrak{R}$ and $\widetilde{x}_{S,\sigma} \in \mathcal{X}_S$). Then $I(s) = \prod_{S,\sigma} I(s)_{S,\sigma}$ with

$$I(s)_{S,\sigma} = \int_{(c_S\circ q_S)^{-1}(z_{S,\sigma})} f_{S,\sigma} d\mu_{z_{S,\sigma}}$$

where $z_{S,\sigma} = c_S \circ q_S(\widetilde{x}_{S,\sigma}) \in \mathcal{Z}$ and $f_{S,\sigma} : \mathcal{X}_S \to \{0,1\}$ is the characteristic function of $x_{S,\sigma}$.

Proposition 2.14 (1) *For all $S \in \mathfrak{S}$ and $\sigma \in \mathfrak{R}$, $I(s)_{S,\sigma} > 0$.*

(2) *If $H = \widehat{R}^\times$ for some Eichler order $R \subset B$ of level $\mathcal{N}$,*

$$I(s)_{S,\sigma} = \frac{1}{\omega(x_{S,\sigma})} \cdot \frac{\Omega(F)}{\Omega(B_S) \cdot \Omega(\mathcal{N}_S)}$$

with $\omega(\star)$, $\Omega(F)$, $\Omega(B_S)$ and $\Omega(\mathcal{N}_S)$ as in section 2.2.4.

Proof. See section 2.4.2, especially Proposition 2.18.

In particular, $I(s) > 0$ and if $H = \widehat{R}^\times$ with R as above,

$$I(s) = \frac{1}{\omega(s)} \cdot \left(\prod_{S \in \mathfrak{S}} \frac{\Omega(F)}{\Omega(B_S) \cdot \Omega(\mathcal{N}_S)} \right)^{|\mathfrak{R}|}.$$

Thus we obtain Corollaries 2.10 and 2.11.

2.4 Further reductions

The arguments of the last section have reduced our task to proving Propositions 2.5, 2.12, 2.13, and 2.14. In this section, we make some further steps in this direction. Section 2.4.1 gives the proof of Proposition 2.5. Section 2.4.2 gives the proof of Proposition 2.14. Finally, Section 2.4.3 is a step towards Ratner's theorem and the proof of Proposition 2.14.
Throughout this section, S is a finite set of finite places of F subject to the condition S1 to S3 of section 2.1.1.

2.4.1 Existence of a measure and proof of Proposition 2.5

We shall repeatedly apply the following principle:

Lemma 2.15 *[20, Lemme 1.2, p. 105] Suppose that L and C are topological groups with L locally compact and C compact. If Λ is a discrete and cocompact subgroup of $L \times C$, the projection of Λ to L is a discrete and cocompact subgroup of L.*

By [20, Théorème 1.4, p. 61], $G^1_S(\mathbf{Q})$ diagonally embedded in $G^1_S(\mathbf{A}_f) \times G^1_S(\mathbf{A}_\infty)$ is a discrete and cocompact subgroup. Since $G^1_S(\mathbf{A}_\infty)$ is compact, $G^1_S(\mathbf{Q})$ is also discrete and cocompact in $G^1_S(\mathbf{A}_f)$. Since the sequence

$$1 \to \ker(\pi_S) \to G^1_S(\mathbf{A}_f) \to G^1(S) \to 1$$

is split exact with $\ker(\pi_S)$ compact, $G^1(S, \mathbf{Q}) = \pi_S(G^1_S(\mathbf{Q}))$ is again a discrete and cocompact subgroup of $G^1(S)$.

Lemma 2.16 *The fibers of $c_S \circ q_S$ are the $G^1_S(\mathbf{A}_f)$-orbits in $\mathcal{X}_S$. For $g \in G_S(\mathbf{A}_f)$ and $x = \overline{G_S(\mathbf{Q})}g$ in $\mathcal{X}_S$, the stabilizer of x in $G^1_S(\mathbf{A}_f)$ is a discrete and cocompact subgroup of $G^1_S(\mathbf{A}_f)$ given by $\mathrm{Stab}_{G^1_S(\mathbf{A}_f)}(x) = g^{-1}G^1_S(\mathbf{Q})g$.*

Proof. Fix $x = \overline{G_S(\mathbf{Q})}g$ in $\mathcal{X}_S$ and put $z = c_S \circ q_S(x) = \overline{Z(\mathbf{Q})^+}\mathrm{nr}_S(g) \in \mathcal{Z}$. The fiber of $c_S \circ q_S$ above z is the image of $L = \mathrm{nr}_S^{-1}\left(\overline{Z(\mathbf{Q})^+}\mathrm{nr}_S(g)\right)$ in $\mathcal{X}_S$ and the stabilizer of x in $G_S^1(\mathbf{A}_f)$ equals $M = G_S^1(\mathbf{A}_f) \cap g^{-1}\overline{G_S(\mathbf{Q})}g$. We have to show that $L = \overline{G_S(\mathbf{Q})}gG_S^1(\mathbf{A}_f)$ and $M = g^{-1}G_S^1(\mathbf{Q})g$.
We break this up into a series of steps.

Step 1: *$\overline{G_S(\mathbf{Q})}gG_S^1(\mathbf{A}_f)$ is closed in $G_S(\mathbf{A}_f)$.* This is equivalent to saying that the $G_S^1(\mathbf{A}_f)$-orbit of x is closed in $\mathcal{X}_S$. Since M contains $g^{-1}G_S^1(\mathbf{Q})g$ which is cocompact in $G_S^1(\mathbf{A}_f)$, M itself is cocompact in $G_S^1(\mathbf{A}_f)$. It follows that $x \cdot G_S^1(\mathbf{A}_f)$ is compact, hence closed in $\mathcal{X}_S$.

Step 2: $L = \overline{G_S(\mathbf{Q})}gG_S^1(\mathbf{A}_f)$. Since $\mathrm{nr}_S : G_S(\mathbf{A}_f) \to Z(\mathbf{A}_F)$ is open, L is the closure of $\mathrm{nr}_S^{-1}(Z(\mathbf{Q})^+\mathrm{nr}_S(g))$ in $G_S(\mathbf{A}_f)$. The norm theorem [20, Théorème 4.1 p. 80] implies that $\mathrm{nr}_S^{-1}(Z(\mathbf{Q})^+\mathrm{nr}_S(g)) = G_S(\mathbf{Q})gG_S^1(\mathbf{A}_f)$ and then $L = \overline{G_S(\mathbf{Q})}gG_S^1(\mathbf{A}_f)$ by (1).

Step 3: $\overline{G_S(\mathbf{Q})} = \overline{Z(\mathbf{Q})}G_S(\mathbf{Q})$. This is easy. See for instance the proof of Corollary 3.10.

Step 4: $M = g^{-1}G_S^1(\mathbf{Q})g$. Suppose that γ belongs to $M = G_S^1(\mathbf{A}_f) \cap g^{-1}\overline{G_S(\mathbf{Q})}g$. By (3), $\gamma = g^{-1}\lambda g_{\mathbf{Q}}g$ for some $\lambda \in \overline{Z(\mathbf{Q})}$ and $g_{\mathbf{Q}} \in G_S(\mathbf{Q})$ with $\mathrm{nr}_S(\gamma) = 1$. Then $\alpha = \lambda^2 = \mathrm{nr}(g_{\mathbf{Q}}^{-1})$ belongs to $\overline{Z(\mathbf{Q})}^2 \cap Z(\mathbf{Q})^+ \subset Z(\mathbf{A}_f)^2 \cap Z(\mathbf{Q})^+$. Since α belongs to $Z(\mathbf{Q})^+ \subset F^\times$, we may form the abelian extension $F(\sqrt{\alpha})$ of F. Since α also belongs to $Z(\mathbf{A}_f)^2$, this extension splits everywhere and is therefore trivial: $\alpha = \lambda_0^2$ for some $\lambda_0 \in F^\times$. Then λ/λ_0 is an element of order 2 in $\overline{Z(\mathbf{Q})} \cap \widehat{\mathcal{O}}_F^\times$. Since $\overline{Z(\mathbf{Q})} \cap \widehat{\mathcal{O}}_F^\times = \overline{\mathcal{O}_F^\times}$ is isomorphic to the profinite completion of $O_F^\times$ (a finite type $\mathbf{Z}$-module), λ/λ_0 actually belongs to $\{\pm 1\}$, the torsion subgroup of $\mathcal{O}_F^\times$. We have shown that λ belongs to $Z(\mathbf{Q})$, hence $\gamma = g^{-1}\lambda g_{\mathbf{Q}}g$ belongs to $g^{-1}G_S(\mathbf{Q})g \cap G_S^1(\mathbf{A}_f) = g^{-1}G_S^1(\mathbf{Q})g$.

Since (1) q_S identifies $\mathcal{X}_S/\ker(\pi_S)$ with $\mathcal{X}(S)$ and (2) $G_S^1(\mathbf{A}_f) \simeq G^1(S) \times \ker(\pi_S)$ with $\ker(\pi_S)$ compact, we obtain:

Lemma 2.17 *The fibers of c_S are the $G^1(S)$-orbits in $\mathcal{X}(S)$. For $g \in G(S)$ and $x = \overline{G(S,\mathbf{Q})}g$ in $\mathcal{X}(S)$, the stabilizer of x in $G^1(S)$ is a discrete and cocompact subgroup of $G^1(S)$ given by $\mathrm{Stab}_{G^1(S)}(x) = g^{-1}G^1(S,\mathbf{Q})g$.*

For $z \in \mathcal{Z}$ and $x \in (c_S \circ q_S)^{-1}(z)$, the map $g \mapsto x \cdot g$ induces a $G_S^1(\mathbf{A}_f)$-equivariant homeomorphism between $\mathrm{Stab}(x)\backslash G_S^1(\mathbf{A}_f)$ and $(c_S \circ q_S)^{-1}(z)$. Similarly, any $x \in c_S^{-1}(z)$ defines a $G^1(S)$-equivariant homeomorphism between $\mathrm{Stab}(x)\backslash G^1(S)$ and $c_S^{-1}(z)$. Proposition 2.5 easily follows.

2.4.2 A computation.

Any Haar measure μ^1 on $G_S^1(\mathbf{A}_f)$ induces a collection of $G_S^1(\mathbf{A}_f)$-invariant Borel measures μ_z^1 on the fibers $(c_S \circ q_S)^{-1}(z)$ of $c_S \circ q_S : \mathcal{X}_S \to \mathcal{Z}$. These measures are characterized by the fact that for any compact open subgroup H_S^1 of $G_S^1(\mathbf{A}_f)$ and any $x \in (c_S \circ q_S)^{-1}(z)$,

$$\mu_z^1(x \cdot H_S^1) = \frac{\mu^1(H_S^1)}{\left|\mathrm{Stab}_{H_S^1}(x)\right|}$$

($\mathrm{Stab}_{H_S^1}(x) = \mathrm{Stab}_{G_S^1(\mathbf{A}_f)}(x) \cap H_S^1$ is indeed finite since $\mathrm{Stab}_{G_S^1(\mathbf{A}_f)}(x)$ is discrete while H_S^1 is compact). One easily checks that $\mu_{z\cdot g}^1(\star g)$ equals μ_z^1 on $(c_S \circ q_S)^{-1}(z)$ for any $g \in G_S(\mathbf{A}_f)$. It follows that these measures assign the same volume λ to each fiber of $c_S \circ q_S$, and $\mu_z^1 = \lambda\mu_z$ on $(c_S \circ q_S)^{-1}(z)$.
We shall now simultaneously determine λ (or find out which normalization of μ^1 yields $\lambda = 1$) and compute a formula for

$$\varphi_z(x) = \mu_z\left(xH_S \cap (c_S \circ q_S)^{-1}(z)\right) \qquad (x \in \mathcal{X}_S, z \in \mathcal{Z})$$

where H_S is a compact open subgroup of $G_S(\mathbf{A}_f)$. The map $z \mapsto \varphi_z(x)$ factors through $\mathcal{Z}/\mathrm{nr}_S(H_S)$ and equals 0 outside $c_S \circ q_S(xH_S) = c_S \circ q_S(x) \cdot \mathrm{nr}_S(H_S)$. Let $z_1, \cdots, z_n$ be a set of representatives for $\mathcal{Z}/\mathrm{nr}_S(H_S)$ and for $1 \leq i \leq n$, let $x_{i,1}, \cdots, x_{i,n_i}$ be a set of representatives in $(c_S \circ q_S)^{-1}(z_i)$ of

$$(c_S \circ q_S)^{-1}(z_i \mathrm{nr}_S(H_S))/H_S = (c_S \circ q_S)^{-1}(z_i) \cdot H_S/H_S.$$

The $x_{i,j}$'s then form a set of representatives for $\mathcal{X}_S/H_S$ and

$$\textstyle\sum_{i,j}\varphi_{z_i}(x_{i,j}) = \sum_i \mu_{z_i}\left(\cup_{j=1}^{n_i} x_{i,j}H_S \cap (c_S \circ q_S)^{-1}(z_i)\right) = \sum_i 1 = n \tag{1}$$

since $(x_{i,j}H_S)_{j=1}^{n_i}$ covers $(c_S \circ q_S)^{-1}(z_i)$.
To compute $\varphi_z(x)$, we may assume that $z = c_S \circ q_S(x)$. Choose $g \in G_S(\mathbf{A}_f)$ such that $x = \overline{G_S(\mathbf{Q})}g$ and put $H_S^1 = H_S \cap G_S^1(\mathbf{A}_f)$. By Lemma 2.16, the map $b \mapsto x \cdot b$ yields a bijection

$$g^{-1}G_S^1(\mathbf{Q})g\backslash(g^{-1}\overline{G_S(\mathbf{Q})}g \cdot H_S) \cap G_S^1(\mathbf{A}_f)/H_S^1 \xrightarrow{\simeq} xH_S \cap (c_S \circ q_S)^{-1}(z)/H_S^1. \tag{2}$$

Note that $g^{-1}\overline{G_S(\mathbf{Q})}g \cdot H_S = g^{-1}G_S(\mathbf{Q})g \cdot H_S$. Let $(a_k b_k)_{k=1}^m$ be a set of representatives for the left hand side of (2), with a_k in $g^{-1}G_S(\mathbf{Q})g$, b_k in H_S and $\mathrm{nr}_S(a_k b_k) = 1$. Since $x \cdot a_k = x$ and b_k normalizes H_S^1,

$$\varphi_z(x) = \sum_{k=1}^m \mu_z\left(x \cdot a_k b_k H_S^1\right) = \frac{m}{|H_S^1 \cap g^{-1}G_S^1(\mathbf{Q})g|} \times \frac{\mu^1(H_S^1)}{\lambda} \cdot .$$

On the other hand, the map $a_k b_k \mapsto \mathrm{nr}_S(a_k) = \mathrm{nr}_S(b_k)^{-1}$ yields a bijection between the left hand side of (2) and

$$\mathrm{nr}_S\left(H_S \cap g^{-1}G_S(\mathbf{Q})g\right)\backslash \mathrm{nr}_S\left(H_S\right) \cap \mathrm{nr}_S(G_S(\mathbf{Q})).$$

Since $\mathrm{nr}_S(G_S(\mathbf{Q})) = Z(\mathbf{Q})^+$, we obtain

$$\varphi_z(x) = \frac{|q(g, H_S)|}{|k(g, H_S)|} \times \frac{\mu^1(H_S^1)}{\lambda} \tag{3}$$

where $k(g, H_S)$ and $q(g, H_S)$ are respectively the kernel and cokernel of

$$gH_Sg^{-1} \cap G_S(\mathbf{Q}) \xrightarrow{\mathrm{nr}_S} \mathrm{nr}_S(H_S) \cap Z(\mathbf{Q})^+.$$

When $H_S = \widehat{R}_S^\times$ for some *Eichler* order R_S in B_S, the following simplifications occur:

- $\mathrm{nr}_S(H_S) = \widehat{\mathcal{O}}_F^\times$, so that $n = |\mathcal{Z}/\mathrm{nr}_S(H_S)| = \left|\widehat{F}^\times/F^{>0}\widehat{\mathcal{O}}_F^\times\right| = h_F^+$ is the order of the narrow class group of F. Note that $h_F^+ = h_F \cdot [\mathcal{O}_F^{>0} : (\mathcal{O}_F^\times)^2]$, where h_F is the class number of F and $(\mathcal{O}_F^\times)^2 = \{x^2 \mid x \in \mathcal{O}_F^\times\}$.
- The map $g \mapsto L(g) = g \cdot \widehat{R}_S \cap B_S$ yields a bijection between $\mathcal{X}_S/H_S = G_S(\mathbf{Q})\backslash G_S(\mathbf{A}_f)/H_S$ and the set of isomorphism classes of nonzero right R-ideals in B_S. Moreover, the left order $\mathcal{O}(g)$ of $L(g)$ equals $g\widehat{R}_Sg^{-1} \cap B_S$, so that $\mathcal{O}(g)^\times = gH_Sg^{-1} \cap G_S(\mathbf{Q})$.
- The following commutative diagram with exact rows

$$\begin{array}{ccccccccc} 1 & \to & \mathcal{O}_F^\times & \to & \mathcal{O}(g)^\times & \to & \mathcal{O}(g)^\times/\mathcal{O}_F^\times & \to & 1 \\ & & {\scriptstyle 2}\downarrow & & {\scriptstyle \mathrm{nr}_S}\downarrow & & \downarrow & & \\ 1 & \to & \mathcal{O}_F^{>0} & \to & \mathcal{O}_F^{>0} & \to & 1 & & \end{array}$$

  yields an exact sequence

$$1 \to \{\pm 1\} \to k(g, H_S) \to \mathcal{O}(g)^\times/\mathcal{O}_F^\times \to \mathcal{O}_F^{>0}/(\mathcal{O}_F^\times)^2 \to q(g, H_S) \to 1.$$

  In particular,

$$\frac{|q(g, H_S)|}{|k(g, H_S)|} = \frac{[\mathcal{O}_F^{>0} : (\mathcal{O}_F^\times)^2]}{2 \cdot [\mathcal{O}(g)^\times : \mathcal{O}_F^\times]}.$$

Combining this, (1), (3) and [20, Corollaire 2.3 p. 142], we obtain:

$$\frac{\mu^1(H_S^1)}{\lambda} = \frac{2^{[F:\mathbf{Q}]}\left|\zeta_F(-1)\right|^{-1}}{\|\mathcal{N}_S\| \cdot \prod_{Q \in \mathrm{Ram}_f(B_S)}(\|Q\| - 1) \cdot \prod_{Q|\mathcal{N}_S}(\|Q\|^{-1} + 1)}$$

where $\mathcal{N}_S$ is the level of R_S. This tells us how to normalize μ^1 in order to have $\lambda = 1$. We have proven:

Proposition 2.18 *Let H_S be a compact open subgroup of $G_S(\mathbf{A}_f)$. For $x \in \mathcal{X}_S$ and $z \in \mathcal{Z}$,*

$$\mu_z\left(xH_S \cap (c_S \circ q_S)^{-1}(z)\right) = \begin{cases} \frac{|q(g,H_S)|}{|k(g,H_S)|}\mu^1(H_S^1) & \text{if } z \in c_S \circ q_S(xH_S) \\ 0 & \text{otherwise} \end{cases}$$

where $x = \overline{G_S(\mathbf{Q})g}$, $k(g, H_S)$ and $q(g, H_S)$ are as above, $H_S^1 = H_S \cap G_S^1(\mathbf{A}_f)$ and μ^1 is the unique Haar measure on $G_S^1(\mathbf{A}_f)$ such that

$$\mu^1(H_S^1) = \frac{2^{[F:\mathbf{Q}]}\left|\zeta_F(-1)\right|^{-1}}{\prod_{Q \in \mathrm{Ram}_f(B_S)}(\|Q\| - 1)}$$

when $H_S = \widehat{R}_S^\times$ *for some* maximal *order* $R_S \subset B_S$. *Moreover, if* $H_S = \widehat{R}_S^\times$ *for some Eichler order* $R_S \subset B_S$ *of level* $\mathcal{N}_S$,

$$\frac{|q(g,H_S)|}{|k(g,H_S)|}\mu^1(H_S^1) = \frac{1}{[\mathcal{O}(g)^\times : \mathcal{O}_F^\times]} \times \frac{\Omega(F)}{\Omega(B_S)\cdot\Omega(\mathcal{N}_S)}$$

with $\Omega(F)$, $\Omega(B_S)$ *and* $\Omega(\mathcal{N}_S)$ *as in section 2.2.4.*

2.4.3 P-adic uniformization.

Suppose moreover that P does not belong to S (this is the case for all $S \in \mathfrak{S}$). Since B splits at P, so does B_S.
Let H be a compact open subgroup of $G_S^1(\mathbf{A}_f)^P = \left\{x \in G_S^1(\mathbf{A}_f) \mid x_P = 1\right\}$. For $z \in \mathcal{Z}$, the right action of $G_S^1(\mathbf{A}_f)$ on $c_S \circ q_S^{-1}(z)$ induces a right action of $B_{S,P}^1 = \{b \in B_{S,P}^\times \mid \mathrm{nr}_S(b) = 1\}$ on $c_S \circ q_S^{-1}(z)/H$.

Lemma 2.19 *This action is transitive and the stabilizer of* $x \in c_S \circ q_S^{-1}(z)/H$ *is a discrete and cocompact subgroup* $\Gamma_S(x)$ *of* $B_{S,P}^1$. *For* $x = \overline{G_S(\mathbf{Q})}gH$ *(with* $g \in G_S(\mathbf{A}_f)$*),* $\Gamma_S(x) = g_P^{-1}\Gamma_S g_P$ *where* $g_P \in B_{S,P}^\times$ *is the* P*-component of* g *and* $\Gamma_S = \Gamma_S(gHg^{-1})$ *is the projection to* $B_{S,P}^1$ *of* $G_S^1(\mathbf{Q}) \cap \left\{gHg^{-1}\cdot B_{S,P}^1\right\} \subset G_S^1(\mathbf{A}_f)$. *The commensurator of* Γ_S *in* $B_{S,P}^\times$ *equals* $F_P^\times B_S^\times$.

Proof. The stabilizer of $\widetilde{x} = \overline{G_S(\mathbf{Q})}g$ in $G_S^1(\mathbf{A}_f)$ equals $\mathrm{Stab}(\widetilde{x}) = g^{-1}G_S^1(\mathbf{Q})g$ (by Lemma 2.16). The strong approximation theorem [20, Théorème 4.3, p. 81] implies that $\mathrm{Stab}(\widetilde{x})B_{S,P}^1 H = G_S^1(\mathbf{A}_f)$. Using Lemma 2.16 again, we obtain

$$(c_S\circ q_S)^{-1}(z) = \widetilde{x}\cdot G_S^1(\mathbf{A}_f) = \widetilde{x}\cdot B_{S,P}^1 H = \widetilde{x}\cdot HB_{S,P}^1 = x\cdot B_{S,P}^1.$$

In particular, $B_{S,P}^1$ acts transitively on $(c_S\circ q_S)^{-1}(z)/H$. An easy computation shows that $\Gamma_S(x) = g_P^{-1}\Gamma_S g_P$ with Γ_S as above.
Put $U = gHg^{-1}\cdot B_{S,P}^1$. The continuous map $U \cap G_S^1(\mathbf{Q})\backslash U \hookrightarrow G_S^1(\mathbf{Q})\backslash G_S^1(\mathbf{A}_f)$ is (1) open since U is open in $G_S^1(\mathbf{A}_f)$ and (2) surjective by the strong approximation theorem. In particular, $U \cap G_S^1(\mathbf{Q})$ is a discrete and cocompact subgroup of U. Since $U = gHg^{-1}\times B_{S,P}^1$ (with gHg^{-1} compact), the projection Γ_S of $U\cap G_S^1(\mathbf{Q})$ to $B_{S,P}^1$ is indeed discrete and cocompact in $B_{S,P}^1$.
Finally, since the compact open subgroups of $G_S^1(\mathbf{A}_f)^P$ are all commensurable, neither the commensurability class of Γ_S nor its commensurator in $B_{S,P}^\times$ depends upon g or H. When $g = 1$ and $H = \widehat{R}^\times \cap G_S^1(\mathbf{A}_f)^P$ for some Eichler order $R \subset B_S$, Γ_S is the image in $B_{S,P}^1$ of the subgroup $\{x \in R[1/P]^\times \mid \mathrm{nr}_S(x) = 1\}$ of $B_S^\times$. The commensurator of Γ_S in $B_{S,P}^\times$ then equals $F_P^\times B_S^\times$ by [20, Corollaire 1.5, p. 106].

Similarly, let H be a compact open subgroup of $G^1(S)^P = \{x \in G^1(S) \mid x_P = 1\}$. Then $B_{S,P}^1$ acts on $c_S^{-1}(z)/H$ and we have the following lemma.

Lemma 2.20 *This action is transitive and the stabilizer of $x \in c_S^{-1}(z)/H$ is a discrete and cocompact subgroup $\Gamma_S(x)$ of $B^1_{S,P}$. For $x = \overline{G(S,\mathbf{Q})}gH$ with g in $G(S)$, $\Gamma_S(x) = g_P^{-1}\Gamma_S g_P$ where $\Gamma_S = \Gamma_S(gHg^{-1})$ is the projection to $B^1_{S,P}$ of $G^1(S,\mathbf{Q}) \cap \left\{gHg^{-1} \cdot B^1_{S,P}\right\} \subset G^1(S)$. The commensurator of Γ_S in $B^\times_{S,P}$ equals $F_P^\times B_S^\times$.*

Proof. The proof is similar, using Lemma 2.17 instead of 2.16. Alternatively, we may deduce the results for c_S from those for $c_S \circ q_S$ as follows. Put $H' = \pi_S^{-1}(H)$. Then H' is a compact open subgroup of $G^1_S(\mathbf{A}_f)$ and q_S induces a $B^1_{S,P}$-equivariant homeomorphism between $(c_S \circ q_S)^{-1}(z)/H'$ and $c_S^{-1}(z)/H$.

In particular, the map $b \mapsto x \cdot b$ induces a $B^1_{S,P}$-equivariant homeomorphism

$$\Gamma_S(x)\backslash B^1_{S,P} \xrightarrow{\simeq} c_S^{-1}(z)/H.$$

Since $\Gamma_S(x)$ is discrete and cocompact in $B^1_{S,P} \simeq \mathrm{SL}_2(F_P)$, there exists a unique $B^1_{S,P}$-invariant Borel probability measure on the left hand side. It corresponds on the right hand side to the image of the measure μ_z through the (proper) map $c_S^{-1}(z) \to c_S^{-1}(z)/H$: the latter is indeed yet another $B^1_{S,P}$-invariant Borel probability measure.

2.5 Reduction of Proposition 2.13 to Ratner's theorem

Let us fix a point $x \in \mathrm{CM}$, a one parameter unipotent subgroup $U = \{u(t)\}$ in $B_P^\times$, a compact open subgroup κ in $F_P^\times$ and a Haar measure $\lambda = dt$ on F_P. For $n \geq 0$, we put $\kappa_n = \varpi_P^{-n}\kappa$ so that $\lambda(\kappa_n) \to \infty$ as $n \to \infty$. For $\gamma \in \mathrm{Gal}_K^{\mathrm{ab}}$ and $t \in F_P$,

$$C \circ \mathrm{Red}(\gamma \cdot x \cdot u(t)) = \gamma \cdot \bar{x} \quad \text{with } \bar{x} = C \circ \mathrm{Red}(x) \in \mathcal{Z}(\mathfrak{S}, \mathfrak{R})$$

where $\mathfrak{S}, \mathfrak{R}$, C and Red are as in section 2.2.1. Our aim is to prove the following two propositions, which together obviously imply Proposition 2.13.

Proposition 2.21 *Suppose that $\mathrm{Red}(x \cdot U)$ is dense in $C^{-1}(\bar{x})$. Then for any continuous function $f : C^{-1}(\bar{x}) \to \mathbf{C}$,*

$$\lim_{n\to\infty} \frac{1}{\lambda(\kappa_n)} \int_{\kappa_n} f \circ \mathrm{Red}(x \cdot u(t))dt = \int_{C^{-1}(\bar{x})} f d\mu_{\bar{x}}.$$

Proposition 2.22 *Under the assumptions of Theorem 2.9, $\mathrm{Red}(\gamma \cdot x \cdot U)$ is dense in $C^{-1}(\gamma \cdot \bar{x})$ for almost all $\gamma \in \mathrm{Gal}_K^{ab}$.*

2.5.1 Reduction of Proposition 2.21

We may assume that f is locally constant (by the same argument that we already used in section 2.3). In this case, there exists a compact open subgroup H of $G^1(\mathbf{A}_f)$ such that f factors through $C^{-1}(\bar{x})/H(\mathfrak{S}, \mathfrak{R})$. For our

purposes, it will be sufficient to assume that f is right $H(\mathfrak{S},\mathfrak{R})$-invariant when H is a compact open subgroup of $G^1(\mathbf{A}_f)^P = \left\{x \in G^1(\mathbf{A}_f) \mid x_P = 1\right\}$. Here, $H(\mathfrak{S},\mathfrak{R}) = \prod_{S,\sigma} H(S)$ with $H(S) = \phi_S(H)$ as usual.
For such an H, the right action of $G^1(\mathfrak{S},\mathfrak{R})$ on $C^{-1}(\bar{x})$ induces a right action of $\prod_{S,\sigma} B^1_{S,P}$ on $C^{-1}(\bar{x})/H(\mathfrak{S},\mathfrak{R})$ which together with the isomorphism

$$\textstyle\prod_{S,\sigma} \phi_{S,P} : (B^1_P)^{\mathfrak{S}\times\mathfrak{R}} \xrightarrow{\simeq} \prod_{S,\sigma} B^1_{S,P}$$

yields a right action of $(B^1_P)^{\mathfrak{S}\times\mathfrak{R}}$ on $C^{-1}(\bar{x})/H(\mathfrak{S},\mathfrak{R})$.
By Lemma 2.20, the map $(b_{S,\sigma}) \mapsto \text{Red}(x) \cdot (\phi_{S,P}(b_{S,\sigma}))$ yields a $(B^1_P)^{\mathfrak{S}\times\mathfrak{R}}$-equivariant homeomorphism

$$\Gamma(x,H)\backslash(B^1_P)^{\mathfrak{S}\times\mathfrak{R}} \xrightarrow{\simeq} C^{-1}(\bar{x})/H(\mathfrak{S},\mathfrak{R}) \tag{4}$$

where $\Gamma(x,H)$ is the stabilizer of $\text{Red}(x) \cdot H(\mathfrak{S},\mathfrak{R})$ in $(B^1_P)^{\mathfrak{S}\times\mathfrak{R}}$. Note that $\Gamma(x,H)$ equals $\prod_{S,\sigma} \Gamma_{S,\sigma}(x,H)$ where for each $S \in \mathfrak{S}$ and $\sigma \in \mathfrak{R}$,

$$\Gamma_{S,\sigma}(x,H) = \phi^{-1}_{S,P}\left\{\text{Stab}_{B^1_{S,P}}\left(\text{RED}_S(\sigma \cdot x) \cdot H(S)\right)\right\}$$

is a discrete and cocompact subgroup of $B^1_P \simeq \text{SL}_2(F_P)$.
Under this equivariant homeomorphism,

- the image of $t \mapsto \text{Red}(x \cdot u(t))$ in $C^{-1}(\bar{x})/H(\mathfrak{S},\mathfrak{R})$ corresponds to the image of $t \mapsto \Delta \circ u(t)$ in $\Gamma(x,H)\backslash(B^1_P)^{\mathfrak{S}\times\mathfrak{R}}$, where $\Delta : B^1_P \to (B^1_P)^{\mathfrak{S}\times\mathfrak{R}}$ is the diagonal map;
- the image of $\mu_{\bar{x}}$ on $C^{-1}(\bar{x})/H(\mathfrak{S},\mathfrak{R})$ corresponds to the (unique) $(B^1_P)^{\mathfrak{S}\times\mathfrak{R}}$-invariant Borel probability measure on $\Gamma(x,H)\backslash(B^1_P)^{\mathfrak{S}\times\mathfrak{R}}$.

Writing $\mu_{\Gamma(x,H)}$ for the latter measure, the above discussion shows that Proposition 2.21 is a consequence of the following purely P-adic statement, itself a special case of a theorem of Ratner, Margulis, and Tomanov.

Proposition 2.23 *Suppose that $\Gamma(x,H) \cdot \Delta(U)$ is dense in $(B^1_P)^{\mathfrak{S}\times\mathfrak{R}}$. Then for any continuous function $f : \Gamma(x,H)\backslash(B^1_P)^{\mathfrak{S}\times\mathfrak{R}} \to \mathbf{C}$,*

$$\lim_{n\to\infty} \frac{1}{\lambda(\kappa_n)} \int_{\kappa_n} f(\Delta \circ u(t))dt = \int_{\Gamma(x,H)\backslash(B^1_P)^{\mathfrak{S}\times\mathfrak{R}}} f d\mu_{\Gamma(x,H)}.$$

Proof. See section 2.7.

2.5.2 Reduction of Proposition 2.22

We keep the above notations and choose:

- an element $g \in G(\mathbf{A}_f)$ such that $x = \overline{T(\mathbf{Q})g}$ in CM;

- for each $\sigma \in \mathfrak{R}$, an element $\lambda_\sigma \in T(\mathbf{A}_f)$ such that $\sigma = \mathrm{rec}_K(\lambda_\sigma) \in \mathrm{Gal}_K^{\mathrm{ab}}$.

For $S \in \mathfrak{S}$ and $\sigma \in \mathfrak{R}$, we thus obtain (using Lemma 2.20):

- $\mathrm{RED}_S(\sigma \cdot x) = \overline{G(S,\mathbf{Q})}\phi_S(\lambda_\sigma g)$ and
- $\Gamma_{S,\sigma}(x,H) = g_P^{-1}\lambda_{\sigma,P}^{-1}\Gamma_{S,\sigma}^0(x,H)\lambda_{\sigma,P}g_P$

where $\lambda_{\sigma,P}$ and g_P are the P-components of λ_σ and g while $\Gamma^0_{S,\sigma}(x,H)$ is the inverse image (through $\phi_{S,P}: B_P^1 \to B_{S,P}^1$) of the projection to $B^1_{S,P}$ of

$$G^1(S,\mathbf{Q}) \cap \left\{\phi_S\left(\lambda_\sigma g H g^{-1}\lambda_\sigma^{-1}\right)\cdot B^1_{S,P}\right\} \subset G^1(S).$$

For a subgroup Γ of B_P^1, we denote by $[\Gamma]$ the commensurability class of Γ in B_P^1, namely the set of all subgroups of B_P^1 which are commensurable with Γ. The group $B_P^\times$ acts on the right on the set of all commensurability classes (by $[\Gamma]\cdot b = [b^{-1}\Gamma b]$) and the stabilizer of $[\Gamma]$ for this action is nothing but the commensurator of Γ in $B_P^\times$.
Since the compact open subgroups of $G^1(\mathbf{A}_f)^P$ are all commensurable, the commensurability class $[\Gamma_S^0]$ of $\Gamma^0_{S,\sigma}(x,H)$ does not depend upon H, x or σ (but it does depend on S). Similarly, the commensurability class $[\Gamma_{S,\sigma}(x)]$ of $\Gamma_{S,\sigma}(x,H)$ does not depend upon H and $[\Gamma_{S,\sigma}(x)] = [\Gamma^0_S]\cdot\lambda_{\sigma,P}g_P$. Changing x to $\gamma\cdot x$ ($\gamma \in \mathrm{Gal}_K^{\mathrm{ab}}$) changes g to λg, where λ is an element of $T(\mathbf{A}_f)$ such that $\gamma = \mathrm{rec}_K(\lambda)$. In particular,

$$[\Gamma_{S,\sigma}(\gamma\cdot x)] = [\Gamma_S^0]\cdot\lambda_{\sigma,P}\lambda_P g_P = [\Gamma_S^0]\cdot\lambda_P\lambda_{\sigma,P}g_P$$

where λ_P is the P-component of λ. On the other hand, the stabilizer of $[\Gamma_S^0]$ in $B_P^\times$ equals $F_P^\times\phi_S^{-1}(B_S^\times)$ by Lemma 2.20. Since $K_P^\times \cap F_P^\times\phi_S^{-1}(B_S^\times) = F_P^\times K^\times$,

$$[\Gamma_{S,\sigma}(\gamma\cdot x)] = [\Gamma_{S,\sigma}(\gamma'\cdot x)] \Longleftrightarrow \gamma \equiv \gamma' \mod \mathrm{Gal}_K^{P-\mathrm{rat}}.$$

With these notations, we have

Proposition 2.24 *Under the assumptions of Theorem 2.9, for (S,σ) and (S',σ') in $\mathfrak{S}\times\mathfrak{R}$ with $(S,\sigma)\neq(S',\sigma')$, the set*

$$\mathfrak{B}((S,\sigma),(S',\sigma')) = \left\{\gamma \in \mathrm{Gal}_K^{ab};\ [\Gamma_{S,\sigma}(\gamma\cdot x)]\cdot U = [\Gamma_{S',\sigma'}(\gamma\cdot x)]\cdot U\right\}$$

is the disjoint union of countably many cosets of Gal_K^{P-rat} in Gal_K^{ab}.

Proof. Fix $(S,\sigma)\neq(S',\sigma')$ in $\mathfrak{S}\times\mathfrak{R}$. We have to show that (under the assumptions of Theorem 2.9) the image of

$$\mathfrak{B}' = \left\{\lambda_P \in K_P^\times;\ [\Gamma_S^0]\cdot\lambda_{\sigma,P}\lambda_P g_P\cdot U = [\Gamma^0_{S'}]\cdot\lambda_{\sigma',P}\lambda_P g_P\cdot U\right\}$$

in $K_P^\times/F_P^\times K^\times$ is at most countable. For that purpose we may as well consider the image of $\mathfrak{B}'$ in $K_P^\times/F_P^\times$.

We first consider the case where $S \neq S'$. In this case, we claim that $\mathfrak{B}'$ is empty. In fact: For $S \neq S'$, we claim that

$$[\Gamma_S^0] \cdot B_P^\times \neq [\Gamma_{S'}^0] \cdot B_P^\times.$$

To see this, suppose that $[\Gamma_{S'}^0] = [\Gamma_S^0] \cdot b$ for some $b \in B_P^\times$. Then $b^{-1} F_P^\times \phi_{S,P}^{-1}(B_S^\times) b = F_P^\times \phi_{S',P}^{-1}(B_{S'}^\times)$, so that $F_P^\times B_{S'}^\times = F_P^\times \phi(B_S^\times)$ in $B_{S',P}^\times$, where $\phi : B_{S,P} \to B_{S',P}$ is the isomorphism of F_P-algebras which sends α to $\phi_{S',P}(b^{-1}\phi_{S,P}^{-1}(\alpha) b)$. Since $F_P B_S = F_P^\times B_S^\times \cup \{0\}$ and similarly for $B_{S'}$, $F_P B_{S'} = F_P \phi(B_S)$.
We contend that ϕ maps B_S to $B_{S'}$. Indeed, suppose that α belongs to B_S and choose $\eta \in F$ such that $\mathrm{Tr}(\alpha+\eta) = \mathrm{Tr}(\alpha) + 2\eta \neq 0$. Since $\alpha + \eta$ belongs to B_S, there exists $\mu \in F_P$ and $\beta \in B_{S'}$ such that $\phi(\alpha + \eta) = \mu\beta$. Taking traces on both sides we obtain $\mu = \frac{\mathrm{Tr}(\alpha+\eta)}{\mathrm{Tr}(\beta)} \in F$, so that $\phi(\alpha + \eta) = \phi(\alpha) + \eta$ belongs to $B_{S'}$, and so does $\phi(\alpha)$.
By symmetry, $\phi^{-1}(B_{S'}) \subset B_S$ and ϕ yields an isomorphism of F-algebras between B_S and $B_{S'}$. This is a contradiction, since B_S and $B_{S'}$ are non-isomorphic quaternion algebras over F when $S \neq S'$. This proves the proposition when $S \neq S'$.
Next we consider the case where $S = S'$ but $\sigma \neq \sigma'$. In this case, an element λ_P in $K_P^\times$ belongs to $\mathfrak{B}'$ if and only if there exists $t \in F_P$ such that

$$b(t) = \lambda_P w(t) \lambda_P^{-1} (\lambda_{\sigma,P} \lambda_{\sigma',P}^{-1}) \in F_P^\times B_S^\times, \tag{5}$$

for $w(t) = \lambda_{\sigma,P} \phi_S(g_P u(t) g_P^{-1}) \lambda_{\sigma,P}^{-1}$. We contend that this condition can only be satisfied for countably many λ_P modulo $F_P^\times$.
Suppose first that $K_P^\times$ normalizes the unipotent subgroup W of elements of the form $w(t)$, for all $t \in F_P$. In this situation, $K_P^\times$ is a split torus, and we claim that (5) never holds for any λ_P and t. To see this, observe that if $k \in K_p$ is arbitrary, then, in view of the representation of elements of $K_P^\times$ and W by triangular matrices, the commutator $[k, b(t)]$ is unipotent. (This also follows from standard facts about Borel subgroups.) Since $b(t) \in F_P^\times B_S^\times$, we can apply this to elements of $K_P \cap B_S = K$, to conclude that either $B_S^\times$ contains nontrivial unipotent elements, or that $[k, b(t)]$ is trivial for all k. The former is impossible, since B_S is a definite quaternion algebra, so we conclude that $b(t)$ commutes with $K^\times$ which implies that $b(t) \in K_P$. It follows that $b(t) \in F_P^\times B_S^\times \cap K_P = F_P^\times K^\times$. But now looking at the form of $b(t)$ shows that $w(t) = 1$ and $(\lambda_{\sigma,P} \lambda_{\sigma',P}^{-1}) \in F_P^\times K^\times$, which contradicts the fact that $\sigma \not\equiv \sigma'$ mod $\mathrm{Gal}_K^{P-\mathrm{rat}}$.
It remains to dispose of the situation where K_P fails to normalize W. In this case, we may argue as follows. Since $w(t)$ is unipotent, the left-hand-side of (5) has norm independent of λ_P. On the other hand, the set $F_P^\times B_S^\times$ contains only countably many elements of given norm. It follows that there are only countably many possibilities for the left-hand-side of (5). Thus consider a given element α in $F_P^\times B_S^\times$. We want to count the number of cosets $\lambda_P F_P^\times \in K_P^\times / F_P^\times$

such that there there exists $t \in F_P$ with

$$\lambda_P w(t) \lambda_P^{-1} = \alpha(\lambda_{\sigma',P} \lambda_{\sigma,P}^{-1}). \tag{6}$$

Note that since $\lambda_{\sigma',P}\lambda_{\sigma,P}^{-1}$ is not an element of $\mathrm{Gal}_K^{P-\mathrm{rat}}$ by assumption, any such t is neccesarily nontrivial. But since the normalizer of W in $K_P^\times$ is precisely $F_P^\times$, we see that if λ_P and λ'_P belong to different $F_P^\times$-cosets, then the conjugates of W by λ_P and λ'_P have trivial intersection. It follows that for each α, there is at most a unique coset in $\lambda_P F_P^\times \in K_P^\times / F_P^\times$ such that (6) holds for some t. Since there are only countably many possibilities for α, our contention follows. □

We may now prove Proposition 2.22. Put

$$\mathfrak{B} = \bigcup_{(S,\sigma)\neq(S',\sigma')} \mathfrak{B}((S,\sigma),(S',\sigma'))$$

so that $\mathfrak{B}$ is again the disjoint union of countably many cosets of $\mathrm{Gal}_K^{P-\mathrm{rat}}$ in $\mathrm{Gal}_K^{\mathrm{ab}}$. Since any such coset is *negligible*, so is $\mathfrak{B}$. We claim that $\mathrm{Red}(\gamma \cdot x \cdot U)$ is dense in $C^{-1}(\gamma \cdot \bar{x})$ if γ belongs to $\mathrm{Gal}_K^{\mathrm{ab}} \setminus \mathfrak{B}$. In fact:

Lemma 2.25 *For $\gamma \in \mathrm{Gal}_K^{ab}$, the following conditions are equivalent:*

1. $\mathrm{Red}(\gamma \cdot x \cdot U)$ *is dense in* $C^{-1}(\gamma \cdot \bar{x})$.
2. $\Gamma(\gamma \cdot x, H) \cdot \Delta(U)$ *is dense in* $(B_P^1)^{\mathfrak{S}\times\mathfrak{R}}$ *($\forall H$ compact open in $G^1(\mathbf{A}_f)^P$).*
3. $\Gamma(\gamma \cdot x, H) \cdot \Delta(U)$ *is dense in* $(B_P^1)^{\mathfrak{S}\times\mathfrak{R}}$ *($\exists H$ compact open in $G^1(\mathbf{A}_f)^P$).*
4. γ *does not belong to* $\mathfrak{B}$.

Proof. Lemma 2.17 implies that a subset Z of $C^{-1}(\bar{\gamma} \cdot x)$ is everywhere dense if and only if for every compact open subgroup H of $G^1(\mathbf{A}_f)^P$, the image of Z in the quotient $C^{-1}(\bar{\gamma} \cdot x)/H(\mathfrak{S},\mathfrak{R})$ is everywhere dense. Applying this to $Z = \mathrm{Red}(\gamma \cdot x \cdot U)$ yields (1) ⇔ (2). But now (2) implies (3) and (4) is equivalent to (3) for any H by Proposition 2.35 below. □

In summary, the arguments of this section show that Proposition 2.13 follows from Proposition 2.23, together with Proposition 2.35 below.

2.6 Proof of Proposition 2.12.

Let V be a simple left B_P-module, so that $V \simeq F_P^2$ as an F_P-module and $V \simeq K_P$ as a K_P-module. We fix a K_P-basis e of V and an $\mathcal{O}_{F_P}$-basis $(1,\omega)$ of $\mathcal{O}_{K_P}$. Then $(e, \omega e)$ is an F_P-basis of V, which we use to identify $B_P \simeq \mathrm{End}_{F_P}(V)$ with $M_2(F_P)$. Under this identification, the element $x = \alpha + \beta\omega$ of K_P corresponds to the matrix $\left(\begin{smallmatrix} \alpha & -\beta\theta \\ \beta & \alpha+\beta\tau \end{smallmatrix}\right)$ with $\theta = \mathrm{nr}(\omega)$ and $\tau = \mathrm{Tr}(\omega)$.

Let $\mathcal{L}$ be the set of all $\mathcal{O}_{F_P}$-lattices in V. To each L in $\mathcal{L}$, we may attach an integer $n(L)$ as follows. The set $\mathcal{O}(L) = \{\lambda \in K_P; \lambda L \subset L\}$ is an $\mathcal{O}_{F_P}$-order in K_P and therefore equals $\mathcal{O}_n = \mathcal{O}_{F_P} + P^n \mathcal{O}_{K_P}$ for a unique integer n: we take $n(L) = n$. From a matrix point of view, $n(L)$ is the smallest integer $n \geq 0$ such that $\varpi_P^n \left(\begin{smallmatrix} 0 & -\theta \\ 1 & \tau \end{smallmatrix}\right) L \subset L$.

Lemma 2.26 *The map $L \mapsto n(L)$ induces a bijection $K_P^\times \backslash \mathcal{L} \to \mathbf{N}$.*

Proof. For $\lambda \in K_P^\times$ and $L \in \mathcal{L}$, $\mathcal{O}(\lambda L) = \mathcal{O}(L)$, so that $n(\lambda \cdot L) = n(L)$: our map is well-defined. Conversely, suppose that $n(L) = n(L') = n$ for $L, L' \in \mathcal{L}$. Since both L and L' are free, rank one $\mathcal{O}_n$-submodules of $V = K_P \cdot e$, there exists $\lambda \in K_P^\times$ such that $\lambda \cdot L = L'$: our map is injective. It is also surjective, since $n(\mathcal{O}_n \cdot e) = n$ for all $n \in \mathbf{N}$.

Put $L_0 = \mathcal{O}_0 \cdot e$, $R = \mathrm{End}(L_0) = M_2(\mathcal{O}_{F_P})$, $\delta = \left(\begin{smallmatrix} \varpi_P & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $u(t) = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$ for $t \in F_P$. Then:

Lemma 2.27 *For $n \geq 0$ and $t \in \mathcal{O}_{F_P}^\times$,*

$$n\big(u(\varpi_P^{-n} t) \cdot L_0\big) = 2n \quad \textit{and} \quad n\big(u(\varpi_P^{-n} t) \cdot \delta L_0\big) = 2n + 1.$$

Proof. Left to the reader.

Let us consider a P-isogeny class $\mathcal{H} \subset \mathrm{CM}$, a compact open subgroup $\mathcal{G} \subset \mathrm{Gal}_K^{\mathrm{ab}}$ and a compact open subgroup $H \subset G(\mathbf{A}_f)$. We choose an element $x_0 \in \mathcal{H}$ such that $x_0 = \overline{T(\mathbf{Q})} \cdot g_0$ for some $g_0 \in G(\mathbf{A}_f)$ whose P-component equals 1. Let $K_P^{\mathcal{G}}$ be the kernel of

$$K_P^\times \hookrightarrow T(\mathbf{A}_f) \xrightarrow{\mathrm{rec}_K} \mathrm{Gal}_K^{\mathrm{ab}} \to \mathrm{Gal}_K^{\mathrm{ab}}/\mathcal{G}.$$

This is an open subgroup of finite index in $K_P^\times$. Let N be a positive integer such that

- $1 + P^N \mathcal{O}_{K_P}^\times \subset K_P^{\mathcal{G}}$ and
- H contains the image of $R(N)^\times = 1 + P^N R \subset B_P^\times$ in $G(\mathbf{A}_f)$.

We denote by $\mathcal{I}_1 \subset K_P^\times$ (resp. $\mathcal{I}_2 \subset R^\times$) a chosen set of representatives for $K_P^\times / K_P^{\mathcal{G}}$ (resp. $R^\times / R(N)^\times$) and put $\mathcal{I} = \mathcal{I}_1 \times \{0, 1\} \times \mathcal{I}_2$. For $i = (\lambda, \epsilon, r) \in \mathcal{I}$, we put

$$x_i = x_0 \cdot \lambda \delta^\epsilon r \in \mathcal{H} \qquad \text{and} \qquad u_i(t) = (\delta^\epsilon r)^{-1} u(t) (\delta^\epsilon r) \ (t \in F_P).$$

We finally put $\kappa = 1 + P^{N+1} \mathcal{O}_{F_P}$, a compact open subgroup of $F_P^\times$.
The following result gives the proof of Proposition 2.12.

Proposition 2.28 *With notations as above,*

1. *$\mathcal{G}\cdot\mathcal{H}\cdot H = \cup_{i\in\mathcal{I}}\cup_{n\geq 0}\mathcal{G}\cdot x_i u_i(\kappa_n)\cdot H$ and*

2. *$\forall i\in\mathcal{I}$ and $\forall n\geq 0$, $\mathcal{G}\cdot x_i u_i(\kappa_n)\cdot H = \mathcal{G}\cdot x_i u_{i,n}\cdot H$*

where $\kappa_n = \varpi_P^{-n}\kappa$ and $u_{i,n} = u(\varpi_P^{-n})\subset u_i(\kappa_n)$.

Proof. Note that $x_i\cdot u_i(t) = x_0\cdot\lambda u(t)\delta^\epsilon r$.
(1) We have to show that any element x of $\mathcal{H}$ belongs to $\mathcal{G}\cdot x_i u_i(\kappa_n)\cdot H$ for some $i\in\mathcal{I}$ and $n\geq 0$. Write $x = x_0\cdot b$ with $b\in B_P^\times$.
Consider the lattice $b\cdot L_0\subset V$ and write $n(b\cdot L_0) = 2n+\epsilon$ with $\epsilon\in\{0,1\}$. By Lemma 2.26 and 2.27, there exists $\lambda_0\in K_P^\times$ such that $b\cdot L_0 = \lambda_0\cdot u(\varpi_P^{-n})\delta^\epsilon\cdot L_0$, hence $b = \lambda_0\cdot u(\varpi_P^{-n})\delta^\epsilon\cdot r_0$ for some $r_0\in R^\times$. By definition of $\mathcal{I}_1$ and $\mathcal{I}_2$, there exists $\lambda\in\mathcal{I}_1$, $k\in K_P^{\mathcal{G}}$, $r\in\mathcal{I}_2$ and $h\in R(N)^\times$ such that $\lambda_0 = k\cdot\lambda$ and $r_0 = rh$. Put $i = (\lambda,\epsilon,r)\in\mathcal{I}$, $t = \varpi_P^{-n}\in\kappa_n$ and $\sigma = \mathrm{rec}_K(k)\in\mathcal{G}$. Since $x_0\cdot k = \sigma\cdot x_0$, we obtain

$$x = x_0\cdot b = \sigma\cdot x_0\cdot\lambda u(t)\delta^\epsilon rh = \sigma\cdot(x_i u_i(t))\cdot h\in\mathcal{G}\cdot x_i\cdot u_i(\kappa_n)\cdot H.$$

(2) We have to show that for $i = (\lambda,\epsilon,r)\in\mathcal{I}$, $n\geq 0$ and $a\in\kappa$,

$$x_i\cdot u_i(\varpi_P^{-n}a)\in\mathcal{G}\cdot x_i u_{i,n}\cdot H.$$

Put $y_a = 1-a^{-1}$, $\lambda_{n,a} = 1+\varpi_P^n y_a\omega\in K_P^\times$ and $\sigma_{n,a} = \mathrm{rec}_K(\lambda_{n,a})$. Since a belongs to $\kappa = 1+P^{N+1}O_{F_P}^\times$, y_a belongs to P^{N+1}, $\lambda_{n,a}$ belongs to $K_P^{\mathcal{G}}$ and $\sigma_{n,a}$ belongs to $\mathcal{G}$. As a matrix,

$$\lambda_{n,a} = \left(\begin{array}{cc} 1 & -\theta\varpi_P^n y_a \\ \varpi_P^n y_a & 1+\tau\varpi_P^n y_a\end{array}\right)\in K_P^\times\subset\mathrm{GL}_2(F_P).$$

In particular,

$$\delta^{-\epsilon}u(-\varpi_P^{-n})\cdot\lambda_{n,a}\cdot u(\varpi_P^{-n}a)\delta^\epsilon = \begin{pmatrix} 1-y_a & -(\theta\varpi_P^n+\tau)\varpi_P^{-\epsilon}y_a \\ \varpi_P^{n+\epsilon}y_a & 1+(a+\tau\varpi_P^n)y_a\end{pmatrix}\equiv 1 \bmod P^N.$$

In other words: there exists $r'\in R(N)^\times$ such that $\lambda_{n,a}u(\varpi_P^{-n}a)\delta^\epsilon = u(\varpi_P^{-n})\delta^\epsilon r'$. We thus obtain:

$$\begin{array}{rcl}\sigma_{n,a}\cdot x_i\cdot u_i(\varpi^{-n}a) & = & x_0\cdot\lambda\lambda_{n,a}u(\varpi_P^{-n}a)\delta^\epsilon r \\ & = & x_0\cdot\lambda u(\varpi_P^{-n})\delta^\epsilon r' r \\ & = & x_i\cdot u_{i,n}\cdot h\end{array}$$

with $h = r^{-1}r'r\in R(N)^\times\subset H$. QED.

2.7 An application of Ratner's Theorem

In this section, we study the distribution of certain unipotent flows on $X = \Gamma\backslash G^r$, where $G = \mathrm{SL}_2(F_P)$, r is a positive integer and $\Gamma = \Gamma_1\times\cdots\times\Gamma_r$ is a product of cocompact lattices in G. Our key tool is the following special case of a theorem of Margulis and Tomanov [11, Theorem 11.2] (see also Ratner's Theorem 3 in [15]). We fix a Haar measure λ on F_P.

Theorem 2.29 (Uniform Distribution) *Let $V = \{v(t)\}$ be a one-parameter unipotent subgroup of G^r.*

1. *For every $x \in X$, there exists*
   - *a closed subgroup $L \supset V$ of G^r such that $\overline{x \cdot V} = x \cdot L$, and*
   - *an L-invariant Borel probability measure μ on X supported on $\overline{x \cdot V}$.*
2. *With x and μ as above, for every continuous function f on X and every compact set κ of F_P with positive measure, we have*

$$\lim_{|s|\to\infty} \frac{1}{\lambda(s\cdot\kappa)} \int_{s\cdot\kappa} f(x\cdot v(t))d\lambda(t) = \int_X f(y)d\mu(y).$$

Here λ denotes a choice of Haar measure on F_P.

The measure μ is uniquely determined by x and V. On the other hand, we may replace the closed subgroup $L \supset V$ of G^r by $\Sigma = \{g \in G^r \mid \mu \text{ is } g\text{-invariant}\}$. Indeed, Σ is a closed subgroup of G^r which contains L and therefore also V. Since μ is Σ-invariant, so is its support $\overline{x \cdot V} = x \cdot L$. In particular, $\overline{x \cdot V} = x \cdot \Sigma$. Suppose now that $V = \Delta(U)$, where $\Delta : G \to G^r$ is the diagonal map and $U = \{u(t)\}$ is a (non-trivial) one-parameter unipotent subgroup of G. In this case, a result of M. Ratner shows that Σ contains some "twisted" diagonal:

Lemma 2.30 *There exists an element $c \in U^r$ such that $c\Delta(G)c^{-1} \subset \Sigma$.*

Proof. This is Corollary 4 of Theorem 6 in [15] when $F_P = \mathbf{Q}_p$ (note that the centralizer of $\Delta(U)$ in G^r equals $\{\pm U\}^r$). The case of general F_P seems to be well-known to the experts, see for instance the notes of N. Shah [17].

This leaves only finitely many possible values for $\Omega = c^{-1}\Sigma c$. Indeed:

Lemma 2.31 *For any subgroup Ω of G^r such that $\Delta(G) \subset \Omega$, there exists a partition $\{I_\alpha\}$ of $\{1, \cdots, r\}$ such that*

$$\prod_\alpha \Delta^{I_\alpha}(G) \subset \Omega \subset \{\pm 1\}^r \cdot \prod_\alpha \Delta^{I_\alpha}(G)$$

where $\Delta^{I_\alpha}(G)$ is the diagonal subgroup of $\{(g_i) \in G^r; \forall i \notin I_\alpha,\, g_i = 1\}$.

Proof. This is a slight generalization of Proposition 3.10 of [2]. According to the latter, there exists a partition $\{I_\alpha\}$ of $\{1, \cdots, r\}$ such that

$$\{\pm 1\}^r \cdot \Omega = \{\pm 1\}^r \cdot \textstyle\prod_\alpha \Delta^{I_\alpha}(G).$$

Taking the derived group on both sides gives $\prod_\alpha \Delta^{I_\alpha}(G) = [\Omega : \Omega] \subset \Omega$.

The equivalence relation $\sim$ on $\{1, \cdots, r\}$ which is defined by the above partition can easily be retrieved from $\overline{x \cdot \Delta(U)} = x \cdot \Sigma$ by the following rule: for $1 \leq i, j \leq r$, $i \sim j$ if and only if the projection

$$\overline{x \cdot \Delta(U)} \subset X \to \Gamma_i \backslash G \times \Gamma_j \backslash G$$

is *not* surjective. On the other hand, this equivalence relation can also be used to characterize those Ω-orbits which are closed subsets of X:

Lemma 2.32 *For $g = (g_i) \in G^r$, the Ω-orbit of $y = \Gamma \cdot g$ is closed in X if and only for all $1 \leq i, j \leq r$ with $i \sim j$, $g_i^{-1}\Gamma_i g_i$ and $g_j^{-1}\Gamma_j g_j$ are commensurable in G.*

Proof. The map $\omega \mapsto y \cdot \omega$ induces a continuous bijection $\theta : g^{-1}\Gamma g \cap \Omega \backslash \Omega \to y \cdot \Omega$. We first claim that $y \cdot \Omega$ is closed in X if and only if $g^{-1}\Gamma g \cap \Omega \backslash \Omega$ is compact. The if part is trivial: if $g^{-1}\Gamma g \cap \Omega \backslash \Omega$ is compact, so is $\theta(g^{-1}\Gamma g \cap \Omega \backslash \Omega) = y \cdot \Omega$. To prove the converse, it is sufficient to show that θ is an homeomorphism when $y \cdot \Omega$ is closed (hence compact). Now if $y \cdot \Omega$ is a closed subset of X, $\Gamma \cdot g \cdot \Omega$ is a closed subset of G^r and $g^{-1}\Gamma g \cdot \Omega$ is a Baire space. Since Γ is countable (being discrete in a σ-compact space), it follows that Ω is open in $g^{-1}\Gamma g \cdot \Omega$ and θ is indeed an homeomorphism.
Put $\Omega' = \prod_\alpha \Delta^{I_\alpha}(G)$. Since $\Omega' \subset \Omega \subset \{\pm 1\}^r \cdot \Omega'$, $g^{-1}\Gamma g \cap \Omega$ is cocompact in Ω if and only if $g^{-1}\Gamma g \cap \Omega'$ is cocompact in Ω'. Note that $g^{-1}\Gamma g \cap \Omega' \backslash \Omega' \simeq \prod_\alpha \Gamma_\alpha \backslash G$ where $\Gamma_\alpha = \cap_{i \in I_\alpha} g_i^{-1}\Gamma_i g_i$, and $\Gamma_\alpha \backslash G$ is compact if and only if $g_i^{-1}\Gamma_i g_i$ is commensurable with $g_j^{-1}\Gamma_j g_j$ for all $i, j \in I_\alpha$. This finishes the proof of the lemma.

We thus obtain a second characterization of the equivalence relation $\sim$.

Definition 2.33 We say that two subgroups Γ and Γ' of G are U-*commensurable* if there exists $u \in U$ such that Γ and $u^{-1}\Gamma u$ are commensurable.

Corollary 2.34 *Write $x = \Gamma \cdot g$ with $g = (g_i) \in G^r$. For $1 \leq i, j \leq r$, $i \sim j$ if and only if $g_i^{-1}\Gamma_i g_i$ and $g_j^{-1}\Gamma_j g_j$ are U-commensurable in G.*

Proof. Write $c = (c_i) \in U^r$ and put $y = x \cdot c = \Gamma \cdot gc$. Then $y \cdot \Omega = x \cdot \Sigma = \overline{x \cdot V}$ is a closed subset of X. The lemma implies that $(g_i c_i)^{-1}\Gamma_i (g_i c_i)$ and $(g_j c_j)^{-1}\Gamma_j (g_j c_j)$ are commensurable in G when $i \sim j$. Conversely, suppose that $g_i^{-1}\Gamma_i g_i$ and $\alpha^{-1} g_j^{-1}\Gamma_j g_j \alpha$ are commensurable for some $\alpha \in U$. Put $\Gamma' = \Gamma_i \times \Gamma_j$, $X' = \Gamma' \backslash G^2$, $c' = (1, \alpha)$ and $\Delta'(g) = (g, g)$ for $g \in G$. Let $p : X \to X'$ be the obvious projection. The lemma implies that $p(x) \cdot c'\Delta'(G)c'^{-1}$ is closed in X', so that

$$p(\overline{x \cdot \Delta(U)}) \subset \overline{p(x) \cdot \Delta'(U)} \subset p(x) \cdot c'\Delta'(G)c'^{-1}.$$

In particular, $p(\overline{x \cdot \Delta(U)}) \neq X'$ and $i \sim j$.

Proposition 2.35 *The following conditions are equivalent:*

1. $\overline{x \cdot \Delta(U)} = X$.
2. *For all* $1 \leq i \neq j \leq r$, $g_i^{-1}\Gamma_i g_i$ *and* $g_j^{-1}\Gamma_j g_j$ *are* not *U-commensurable.*

The measure μ of Theorem 2.29 is then the (unique) G^r-invariant Borel probability measure on X.

Proof. Both conditions are equivalent to the assertion that the partition $\{I_\alpha\}$ of $\{1, \cdots, r\}$ is trivial. In that case, $\Omega = G^r = \Sigma$ and μ is G^r-invariant.

3 The case of Shimura curves

3.1 Shimura Curves

3.1.1 Definitions

We start by defining the Shimura curves. Let $\{\tau_1, \cdots, \tau_d\} = \mathrm{Hom}_{\mathbf{Q}}(F, \mathbf{R})$ be the set of real embeddings of F. We shall always view F as a subfield of $\mathbf{R}$ or $\mathbf{C}$ through τ_1. Let S be a set of finite primes such that $|S| + d$ is odd, and let B denote the quaternion algebra over F which ramifies precisely at $S \cup \{\tau_2, \cdots, \tau_d\}$ (a finite set of even order). Let G be the reductive group over $\mathbf{Q}$ whose set of points on a commutative $\mathbf{Q}$-algebra A is given by $G(A) = (B \otimes A)^\times$.
In particular, $G_{\mathbf{R}} \simeq G_1 \times \cdots \times G_d$ where $B_{\tau_i} = B \otimes_{F,\tau_i} \mathbf{R}$ and G_i is the algebraic group over $\mathbf{R}$ whose set of points on a commutative $\mathbf{R}$-algebra A is given by $G_i(A) = (B_{\tau_i} \otimes_{\mathbf{R}} A)^\times$. Fix $\epsilon = \pm 1$ and let X be the $G(\mathbf{R})$-conjugacy class of the morphism from $\mathbb{S} \stackrel{\text{def}}{=} \mathrm{Res}_{\mathbf{C}/\mathbf{R}}(\mathbb{G}_{m,\mathbf{C}})$ to $G_{\mathbf{R}}$ which maps $z = x + iy \in \mathbb{S}(\mathbf{R}) = \mathbf{C}^\times$ to

$$\left[\begin{pmatrix} x & y \\ -y & x \end{pmatrix}^{\epsilon}, 1, \cdots, 1\right] \in G_1(\mathbf{R}) \times \cdots \times G_d(\mathbf{R}) \simeq G(\mathbf{R}).$$

We have used an isomorphism of $\mathbf{R}$-algebras $B_{\tau_1} \simeq M_2(\mathbf{R})$ to identify G_1 and $\mathrm{GL}_{2,\mathbf{R}}$; the resulting conjugacy class X does not depend upon this choice (but it does depend on ϵ, cf. section 3.3.1 below).
For every compact open subgroup H of $G(\mathbf{A}_f)$, the quotient of $G(\mathbf{A}_f)/H \times X$ by the diagonal left action of $G(\mathbf{Q})$ is a Riemann surface

$$M_H^{\mathrm{an}} \stackrel{\text{def}}{=} G(\mathbf{Q}) \backslash (G(\mathbf{A}_f)/H \times X)$$

which is compact unless $d = 1$ ($F = \mathbf{Q}$) and $S = \emptyset$ ($G = \mathrm{GL}_2$). The *Shimura curve* M_H is Shimura's canonical model for M_H^{an}. It is a smooth curve over F (the reflex field) whose underlying Riemann surface $M_H(\mathbf{C})$ equals M_H^{an}.

3.1.2 CM points

Among the models of M_H^{an}, the Shimura curve M_H is characterized by specifying the action of Galois (the "reciprocity law") on certain *special points*. A morphism $h : \mathbb{S} \to G_{\mathbf{R}}$ in X is special if it factors through the real locus of some $\mathbf{Q}$-rational subtorus of G and a point x in M_H^{an} is special if $x = [g, h]$ with h special (and g in $G(\mathbf{A}_f)$).
Now let K be an imaginary quadratic extension of F such that there exists some embedding $K \to B$. Put $T = \mathrm{Res}_{K/\mathbf{Q}}(\mathbb{G}_{m,K})$. Any embedding $K \hookrightarrow B$ yields an embedding $T \hookrightarrow G$. In the sequel, we shall fix an embedding of K in B, and study those special points in X or M_H^{an} for which $h : \mathbb{S} \to G_{\mathbf{R}}$ factors through the morphism $T_{\mathbf{R}} \hookrightarrow G_{\mathbf{R}}$ which is induced by the fixed F-embedding $K \hookrightarrow B$. We shall refer to such points as *CM points.* We denote by CM_H the set of CM points in $M_H^{\mathrm{an}} = M_H(\mathbf{C})$. It is clear that this set is nonempty. Furthermore, Shimura's theory implies that any CM point is algebraic, defined over the maximal abelian extension K^{ab} of K (see section 3.2.4 below).

3.1.3 Integral models and supersingular points

Let v be a finite place of F where B is split and put $\mathcal{S} = \mathrm{Spec}\, \mathcal{O}(v)$ where $\mathcal{O}(v)$ is the local ring of F at v. We denote by F_v and $\mathcal{O}_v$ the completion of F at v and its ring of integers. For simplicity, we shall only consider level structures $H \subset G(\mathbf{A}_f)$ which decompose as $H = H^v H_v$ where H^v (resp. H_v) is a compact open subgroup of

$$G(\mathbf{A}_f)^v = \{g \in G(\mathbf{A}_f) \mid g_v = 1\} \quad (\text{resp. } B_v^\times \subset G(\mathbf{A}_f)).$$

In the non-compact (classical) case where $F = \mathbf{Q}$ and $G = \mathrm{GL}_2$, it is well-known that M_H is a coarse moduli space which classifies elliptic curves (with level structures) over extensions of $\mathbf{Q}$. Extending the moduli problem to elliptic curves over $\mathcal{S}$-schemes, we obtain a regular model $\mathbf{M}_H/\mathcal{S}$ of M_H. A geometric point in the special fiber of $\mathbf{M}_H$ is *supersingular* if it corresponds to (the class of) a supersingular elliptic curve.
In the general (compact) case, the Shimura curve M_H may not be a moduli space. However, provided that H^v is *sufficiently small* (a condition depending upon H_v), Carayol describes in [1] a proper and regular model $\mathbf{M}_H/\mathcal{S}$ of M_H, which is smooth when H_v is a maximal compact open subgroup of B_v^*. When H^v fails to be sufficiently small in the sense of [1], we let $\mathbf{M}_H/\mathcal{S}$ be the quotient of $\mathbf{M}_{H'}$ by the $\mathcal{S}$-linear right action of H/H', where $H' = H'^v H_v$ for a sufficiently small compact, open and normal subgroup H'^v of H^v. Then $\mathbf{M}_H/\mathcal{S}$ is again a proper and regular model of M_H which is smooth when H_v is maximal (cf. [9, p. 508]), and it does not depend upon the choice of H'^v.
These models form a projective system $\{\mathbf{M}_H\}_H$ of proper $\mathcal{S}$-schemes with finite flat transition maps, whose limit $\mathbf{M} = \varprojlim \mathbf{M}_H$ has a right action of $G(\mathbf{A}_f)$ and carries an $\mathcal{O}_v$-divisible module $\mathbf{E}$ of height 2 (cf. [1, Appendice] for the definition and basic properties of $\mathcal{O}_v$-divisible modules). A geometric point

x in the special fiber of $\mathbf{M}$ is said to be *ordinary* if $\mathbf{E} \mid x$ is isomorphic to the product of the $\mathcal{O}_v$-divisible constant module $F_v/\mathcal{O}_v$ with Σ_1, the unique $\mathcal{O}_v$-formal module of height 1. Otherwise, x is *supersingular* and $\mathbf{E} \mid x$ is isomorphic to Σ_2, the unique $\mathcal{O}_v$-formal module of height 2. A supersingular point in the special fiber of $\mathbf{M}_H$ is one which lifts to a supersingular point in $\mathbf{M}$.
In the classical case, the supersingular points also have such a description, with $\mathbf{E}$ equal to the relevant Barsotti-Tate group in the universal elliptic curve on $\mathbf{M} = \varprojlim \mathbf{M}_H$.

3.1.4 Reduction maps

Let us choose a place $\bar{v}$ of K^{ab} above v, with ring of integers $\mathcal{O}(\bar{v}) \subset K^{\mathrm{ab}}$ and residue field $\mathbf{F}(\bar{v})$, an algebraic closure of the residue field $\mathbf{F}(v)$ of v. Consider the specialization maps:

$$M_H(K^{\mathrm{ab}}) = \mathbf{M}_H(K^{\mathrm{ab}}) \leftarrow \mathbf{M}_H(\mathcal{O}(\bar{v})) \rightarrow \mathbf{M}_H(\mathbf{F}(\bar{v}))$$

In the compact case, $\mathbf{M}_H$ is proper over $\mathcal{S}$ and the first of these two maps is a bijection by the valuative criterion of properness. In the classical non-compact case, the first map is still injective ($\mathbf{M}_H$ is separated over $\mathcal{S}$); by [16, Theorem 6], its image contains CM_H. In both cases, we obtain a reduction map

$$\mathrm{RED}_v : \mathrm{CM}_H \rightarrow \mathbf{M}_H(\mathbf{F}(\bar{v})).$$

Let $\mathbf{M}_H^{ss}(v)$ be the set of supersingular points in $\mathbf{M}_H(\mathbf{F}(\bar{v}))$.

Lemma 3.1 *If v does not split in K,* $\mathrm{RED}_v(\mathrm{CM}_H) \subset \mathbf{M}_H^{ss}(v)$.

Proof. (Sketch) Let $\mathbf{E}^0$ be the $\mathcal{O}_v$-divisible module $\mathbf{E}$ "up to isogeny". There is an F_v-linear right action of $G(\mathbf{A}_f)$ on $\mathbf{E}^0$ covering the right action of $G(\mathbf{A}_f)$ on $\mathbf{M}$ (see [1, 7.5] for the compact case). For any point x on $\mathbf{M}$, we thus obtain an F_v-linear right action of $\mathrm{Stab}_{G(\mathbf{A}_f)}(x)$ on $\mathbf{E}^0 \mid x$. If x is a CM point, say $x = [g,h]$ for some $g \in G(\mathbf{A}_f)$ and $h : \mathbb{S} \rightarrow T_{\mathbf{R}} \hookrightarrow G_{\mathbf{R}}$ in X, $g^{-1}T(\mathbf{Q})g \subset \mathrm{Stab}_{G(\mathbf{A}_f)}(x)$ and the induced F_v-linear action of $T(\mathbf{Q}) = K^*$ on $\mathbf{E}^0 \mid x$ (or its inverse, depending upon ϵ) arises from an F_v-linear right K_v-module structure on $\mathbf{E}^0 \mid x$. The connected part of the special fiber $\mathbf{E}^0 \mid \mathrm{RED}_v(x)$ therefore inherits a K_v-module structure. Since $\mathrm{End}_{F_v}(\Sigma_1^0) \simeq F_v$, this connected part can not be isomorphic to Σ_1^0 unless v splits in K.

3.1.5 Connected components

We now want to define yet another type of "reduction map". Recall from Shimura's theory that the natural map from M_H^{an} to its set of connected components $\pi_0(M_H^{\mathrm{an}})$ corresponds to an F-morphism $c : M_H \rightarrow \mathcal{M}_H$ between the Shimura curve M_H and a zero-dimensional Shimura variety $\mathcal{M}_H$ over F

whose finitely many points are algebraic over the maximal abelian extension $F^{\mathrm{ab}} \subset K^{\mathrm{ab}}$ of F. Since $\mathbf{M}_H$ is regular (hence normal), this morphism extends over $\mathcal{S}$ to a morphism $c : \mathbf{M}_H \to \boldsymbol{\mathcal{M}}_H$ between $\mathbf{M}_H$ and the normalization $\boldsymbol{\mathcal{M}}_H$ of $\mathcal{S}$ in $\mathcal{M}_H$, a finite and regular $\mathcal{S}$-scheme. With $\mathcal{Z}_H \stackrel{\mathrm{def}}{=} \pi_0(M_H^{\mathrm{an}}) = \mathcal{M}_H(K^{\mathrm{ab}})$ and $\mathcal{Z}_H(v) \stackrel{\mathrm{def}}{=} \boldsymbol{\mathcal{M}}_H(\mathbf{F}(\overline{v}))$, the following diagram is commutative:

$$\begin{array}{ccc} \mathrm{CM}_H & \stackrel{\mathrm{RED}_v}{\longrightarrow} & \mathbf{M}_H^{\mathrm{ss}}(v) \\ c\downarrow & & \downarrow c \\ \mathcal{Z}_H & \stackrel{\mathrm{RED}_v}{\longrightarrow} & \mathcal{Z}_H(v) \end{array}$$

Put $\mathcal{X}_H(v) = \mathbf{M}_H^{ss}(v) \times_{\mathcal{Z}_H(v)} \mathcal{Z}_H$. If v does not split in K, we thus obtain a *reduction map* and a *connected component map*

$$\begin{array}{ccccc} \mathrm{CM}_H & \stackrel{\mathrm{RED}_v}{\longrightarrow} & \mathcal{X}_H(v) & \stackrel{c_v}{\longrightarrow} & \mathcal{Z}_H \\ x & \longmapsto & (\mathrm{RED}_v(x), c(x)) & \longmapsto & c(x) \end{array} \tag{7}$$

The composite map $c = c_v \circ \mathrm{RED}_v : \mathrm{CM}_H \to \mathcal{Z}_H$ does not depend upon v and commutes with the action of $\mathrm{Gal}_K^{\mathrm{ab}}$ on both sides.

Remark 3.2 In the compact case, $\mathbf{M}_H \xrightarrow{c} \boldsymbol{\mathcal{M}}_H \to \mathcal{S}$ is the Stein factorization

$$\mathbf{M}_H \to \mathrm{Spec}(\Gamma(\mathbf{M}_H, \mathcal{O}_{\mathbf{M}_H})) \to \mathcal{S}$$

of the proper morphism $\mathbf{M}_H \to \mathcal{S}$. Indeed, since $\boldsymbol{\mathcal{M}}_H$ is affine, c factors through an S-morphism $\alpha : \mathrm{Spec}(\Gamma(\mathbf{M}_H, \mathcal{O}_{\mathbf{M}_H})) \to \boldsymbol{\mathcal{M}}_H$. Over the generic point of $\mathcal{S}$, $\alpha_{/F} : \mathrm{Spec}(\Gamma(M_H, \mathcal{O}_{M_H})) \to \mathcal{M}_H$ is a morphism between finite étale F-schemes which induces a bijection on complex points: it is therefore an isomorphism. Since $\mathbf{M}_H$ is a regular scheme which is proper and flat over $\mathcal{S}$, $\mathrm{Spec}(\Gamma(\mathbf{M}_H, \mathcal{O}_{\mathbf{M}_H}))$ is a normal scheme which is finite and flat over $\mathcal{S}$. It follows that α is an isomorphism.

3.1.6 Simultaneous reduction maps

Let $\mathfrak{S}$ be a finite set of finite places of F which are non-split in K and away from S: for each $v \in \mathfrak{S}$, K_v is a field and $B_v \simeq M_2(F_v)$. Let also $\mathfrak{R}$ be a finite set of Galois elements in $\mathrm{Gal}_K^{\mathrm{ab}}$. We put

$$\mathcal{X}_H(\mathfrak{S}, \mathfrak{R}) \stackrel{\mathrm{def}}{=} \prod_{v \in \mathfrak{S}, \sigma \in \mathfrak{R}} \mathcal{X}_H(v) \quad \text{and} \quad \mathcal{Z}_H(\mathfrak{S}, \mathfrak{R}) \stackrel{\mathrm{def}}{=} \prod_{v \in \mathfrak{S}, \sigma \in \mathfrak{R}} \mathcal{Z}_H(v)$$

and define *a simultaneous reduction map* and *a connected component map*

$$\mathrm{Red} : \mathrm{CM}_H \to \mathcal{X}_H(\mathfrak{S}, \mathfrak{R}) \quad \text{and} \quad C : \mathcal{X}_H(\mathfrak{S}, \mathfrak{R}) \to \mathcal{Z}_H(\mathfrak{S}, \mathfrak{R})$$

by $\mathrm{Red}(x) = (\mathrm{RED}_v(\sigma x))_{v \in \mathfrak{S}, \sigma \in \mathfrak{R}}$ and $C(x_{v,\sigma}) = (c_v(x_{v,\sigma}))_{v \in \mathfrak{S}, \sigma \in \mathfrak{R}}$. For $x \in \mathrm{CM}_H$ and $\tau \in \mathrm{Gal}_K^{\mathrm{ab}}$, we have

$$C \circ \mathrm{Red}(\tau x) = (\tau \sigma c(x))_{v \in \mathfrak{S}, \sigma \in \mathfrak{R}}.$$

3.1.7 Main theorem

Let P be a maximal ideal of $\mathcal{O}_F$ and suppose that $P \notin S \cup \mathfrak{S}$. In particular, $B_P \simeq M_2(F_P)$. We make no assumptions on P relative to K: P may either split, ramify or be inert in K. Then the following definitions reprise the ones already made in the first section of the paper: we have chosen to repeat them here for the convenience of the reader.

Definition 3.3 We say that two points x and $y \in M_H^{\mathrm{an}}$ are *P-isogeneous* if $x = [g, h]$ and $y = [g', h]$ for some $h \in X$ and $g, g' \in G(\mathbf{A}_f)$ such that $g_w = g'_w$ for every finite place $w \neq P$ of F.

Note that a point which is P-isogeneous to a CM point is again a CM point.

Definition 3.4 An element $\sigma \in \mathrm{Gal}_K^{\mathrm{ab}}$ is *P-rational* if $\sigma = \mathrm{rec}_K(\lambda)$ for some $\lambda \in \widehat{K}^\times$ whose P-component λ_P belongs to the subgroup $K^\times \cdot F_P^\times$ of $K_P^\times$. We denote by $\mathrm{Gal}_K^{P-\mathrm{rat}} \subset \mathrm{Gal}_K^{\mathrm{ab}}$ the subgroup of all P-rational elements.

In the above definition, $\mathrm{rec}_K : \widehat{K}^\times \twoheadrightarrow \mathrm{Gal}_K^{\mathrm{ab}}$ is Artin's reciprocity map. We normalize the latter by specifying that it sends local uniformizers to *geometric* Frobeniuses.

Theorem 3.5 *Suppose that the finite subset $\mathfrak{R}$ of Gal_K^{ab} consists of elements which are pairwise distinct modulo Gal_K^{P-rat}. Let $\mathcal{H} \subset \mathrm{CM}_H$ be a P-isogeny class of CM points and let $\mathcal{G}$ be a compact open subgroup of Gal_K^{ab}. Then for all but finitely many points $x \in \mathcal{H}$,*

$$\mathrm{Red}(\mathcal{G} \cdot x) = C^{-1}(C \circ \mathrm{Red}(\mathcal{G} \cdot x)).$$

Remark 3.6 When the level structure H arises from an Eichler order in B, our proof of this surjectivity statement yields a little bit more: for any $y \in \mathcal{X}_H(\mathfrak{S}, \mathfrak{R})$, we can compute the asymptotic behavior of the probability that $\mathrm{Red}(g \cdot x) = y$ for some $g \in \mathcal{G}$, as x goes to infinity inside $\mathcal{H}$ (see Corollary 2.11).

3.2 Uniformization

Write CM, $\mathbf{M}^{ss}(v)$, $\mathcal{Z}$, $\mathcal{Z}(v)$ and $\mathcal{X}(v) = \mathbf{M}^{ss}(v) \times_{\mathcal{Z}(v)} \mathcal{Z}$ for the projective limits of $\{\mathrm{CM}_H\}$, $\{\mathbf{M}_H^{ss}(v)\}$, $\{\mathcal{Z}_H\}$, $\{\mathcal{Z}_H(v)\}$ and $\{\mathcal{X}_H(v)\}$. These sets now have a right action of $G(\mathbf{A}_f)$. For $X \in \{\mathrm{CM}, \mathbf{M}^{ss}(v), \mathcal{Z}, \mathcal{Z}(v)\}$, the natural map $X/H \to X_H$ is a bijection while $\mathcal{X}(v)/H \to \mathcal{X}_H(v)$ is surjective. The projective limit of (7) yields $G(\mathbf{A}_f)$-equivariant maps

$$\mathrm{CM} \xrightarrow{\mathrm{RED}_v} \mathcal{X}(v) \xrightarrow{c_v} \mathcal{Z}$$

which we shall now compute.

3.2.1 CM points

From [4, Proposition 2.1.10] or [14, Theorem 5.27],

$$M^{\mathrm{an}} \stackrel{\mathrm{def}}{=} \varprojlim M_H^{\mathrm{an}} = G(\mathbf{Q})\backslash \left(G(\mathbf{A}_f)/\overline{Z(\mathbf{Q})} \times X \right)$$

where $Z = \mathrm{Res}_{F/\mathbf{Q}}(\mathbb{G}_{m,F})$ is the center of G and $\overline{Z(\mathbf{Q})}$ is the closure of $Z(\mathbf{Q})$ in $Z(\mathbf{A}_f)$. Inside M^{an}, $\mathrm{CM} \stackrel{\mathrm{def}}{=} \varprojlim \mathrm{CM}_H$ corresponds to those elements which can be represented by (g,h) with $g \in G(\mathbf{A}_f)$ and h a CM point in X. Let us construct such an h and show that any other CM point belongs to the same $G(\mathbf{Q})$-orbit.
Since K_v is a field for all $v \in S \cup \{\tau_2, \cdots, \tau_d\}$ (the set of places of F where B ramifies), there exists an F-embedding $\iota : K \hookrightarrow B$. Moreover, any other F-embedding $K \hookrightarrow B$ is conjugated to ι by an element of $B^\times = G(\mathbf{Q})$. We use ι to identify T as a $\mathbf{Q}$-rational subtorus of G and also chose an extension $\tau_1 : K \hookrightarrow \mathbf{C}$ of our distinguished embedding $\tau_1 : F \hookrightarrow \mathbf{R}$. In the sequel, we shall always view K as a subfield of $\mathbf{C}$ through τ_1.
Put $T_i = \mathrm{Res}_{K_{\tau_i}/\mathbf{R}}(\mathbb{G}_{m,K_{\tau_i}})$ (with $K_{\tau_i} = K \otimes_{F,\tau_i} \mathbf{R}$), so that $T_{\mathbf{R}} \simeq T_1 \times \cdots \times T_d$ and this decomposition is compatible with the decomposition $G_{\mathbf{R}} \simeq G_1 \times \cdots \times G_d$ of section 3.1.1. Moreover, $\tau_1 : K \hookrightarrow \mathbf{C}$ induces an isomorphism between K_{τ_1} and $\mathbf{C}$ which allows us to identify T_1 and $\mathbb{S}$. There are exactly two morphisms s and $\bar{s} : \mathbb{S} \to T_{\mathbf{R}}$ whose composite with $\iota_{\mathbf{R}} : T_{\mathbf{R}} \hookrightarrow G_{\mathbf{R}}$ belongs to X. They are characterized by

$$s(z) = (z^\epsilon, 1, \cdots, 1) \quad \text{and} \quad \bar{s}(z) = (\bar{z}^\epsilon, 1, \cdots, 1) \quad \text{for } z \in \mathbf{C}^\times = \mathbb{S}(\mathbf{R}).$$

Finally, there exists an element $b \in B^\times = G(\mathbf{Q})$ such that $b\iota(\lambda)b^{-1} = \iota(\bar{\lambda})$ for all $\lambda \in K$ (where $\lambda \mapsto \bar{\lambda}$ is the non-trivial F-automorphism of K). But then $b(\iota_{\mathbf{R}} \circ \bar{s})b^{-1} = \iota_{\mathbf{R}} \circ s$, so that $h = \iota_{\mathbf{R}} \circ s$ and $\bar{h} = \iota_{\mathbf{R}} \circ \bar{s}$ belong to the same $G(\mathbf{Q})$-orbit in X. Since the centralizer of h in $G(\mathbf{Q})$ equals $T(\mathbf{Q})$, we obtain:

Lemma 3.7 *The map $g \mapsto [1,h] \cdot g = [g,h]$ induces a bijection*

$$\overline{T(\mathbf{Q})}\backslash G(\mathbf{A}_f) \xrightarrow{\simeq} \mathrm{CM}$$

where $\overline{T(\mathbf{Q})}$ is the closure of $T(\mathbf{Q})$ in $T(\mathbf{A}_f)$.

Proof. The above discussion gives a bijection $T(\mathbf{Q})\backslash G(\mathbf{A}_f)/\overline{Z(\mathbf{Q})} \simeq \mathrm{CM}$. We claim that $T(\mathbf{Q})\overline{Z(\mathbf{Q})} = \overline{T(\mathbf{Q})}$. Indeed, $\overline{Z(\mathbf{Q})}$ is the product of $Z(\mathbf{Q}) = F^\times$ with the closure of $\mathcal{O}_F^\times$ in $\widehat{\mathcal{O}}_F^\times \subset Z(\mathbf{A}_f) = \widehat{F}^\times$ (this holds more generally for any number field). Therefore, $T(\mathbf{Q})\overline{Z(\mathbf{Q})} = K^\times \overline{\mathcal{O}_F^\times}$ and

$$T(\mathbf{Q})\overline{Z(\mathbf{Q})} \cap \widehat{\mathcal{O}}_K^\times = \mathcal{O}_K^\times \overline{\mathcal{O}_F^\times} = \cup_{\alpha \in O_K^\times/O_F^\times} \alpha \overline{\mathcal{O}_F^\times}.$$

Since $[\mathcal{O}_K^\times : \mathcal{O}_F^\times]$ is finite, $T(\mathbf{Q})\overline{Z(\mathbf{Q})}$ is a locally closed, hence closed subgroup of $T(\mathbf{A}_f)$. Our claim easily follows.

3.2.2 Connected components

Let $G(\mathbf{R})^+$ and $Z(\mathbf{R})^+$ be the identity components of $G(\mathbf{R})$ and $Z(\mathbf{R})$ and put $G(\mathbf{Q})^+ = G(\mathbf{Q}) \cap G(\mathbf{R})^+$ and $Z(\mathbf{Q})^+ = Z(\mathbf{Q}) \cap Z(\mathbf{R})^+$. Thus, $G(\mathbf{R})^+$ is the set of elements in $G(\mathbf{R})$ whose projection to $G_1(\mathbf{R}) \simeq GL_2(\mathbf{R})$ has a positive determinant while $Z(\mathbf{Q})^+$ is the subgroup of totally positive elements in $Z(\mathbf{Q}) = F^\times$. Let $X^+ = G(\mathbf{R})^+ \cdot h$ be the connected component of h in X. Since $G(\mathbf{Q})$ is dense in $G(\mathbf{R})$, $G(\mathbf{Q}) \cdot X^+ = X$ and

$$M_H^{\mathrm{an}} \simeq G(\mathbf{Q})^+ \backslash \left(G(\mathbf{A}_f)/H \times X^+ \right).$$

It follows that $\mathcal{Z}_H = \pi_0(M_H^{\mathrm{an}}) \simeq G(\mathbf{Q})^+ \backslash G(\mathbf{A}_f)/H$.
On the other hand, the reduced norm $\mathrm{nr} : B \to F$ induces a surjective morphism $\mathrm{nr} : G \to Z$ whose kernel $G^1 \subset G$ is the derived group of G. The norm theorem ($\mathrm{nr}(G(\mathbf{Q})^+) = Z(\mathbf{Q})^+$, [20, p. 80]) and the strong approximation theorem ($G^1(\mathbf{Q})$ is dense in $G^1(\mathbf{A}_f)$, [20, p. 81]) together imply that the reduced norm induces a bijection between $G(\mathbf{Q})^+ \backslash G(\mathbf{A}_f)/H$ and $Z(\mathbf{Q})^+ \backslash Z(\mathbf{A}_f)/\mathrm{nr}(H)$. With $\mathcal{Z} \stackrel{\mathrm{def}}{=} \varprojlim \mathcal{Z}_H$, we thus obtain:

Lemma 3.8 *The map $g \mapsto c([1,h]) \cdot g = c([g,h])$ factors through the reduced norm and yields a bijection*

$$\overline{Z(\mathbf{Q})^+} \backslash Z(\mathbf{A}_f) \xrightarrow{\simeq} \mathcal{Z}$$

where $\overline{Z(\mathbf{Q})^+}$ is the closure of $Z(\mathbf{Q})^+$ in $Z(\mathbf{A}_f)$.

3.2.3 Supersingular points

Proposition 3.9 (1) *The right action of $G(\mathbf{A}_f)$ on $\mathcal{X}(v) \stackrel{\mathrm{def}}{=} \varprojlim \mathcal{X}_H(v)$ is transitive and factors through the surjective group homomorphism*

$$(\mathbf{1}, \mathrm{nr}_v) : G(\mathbf{A}_f) = G(\mathbf{A}_f)^v \times B_v^\times \to G(\mathbf{A}_f)^v \times F_v^\times$$

where $G(\mathbf{A}_f)^v = \{g \in G(\mathbf{A}_f); g_v = 1\}$.
(2) *For any point $x \in \mathcal{X}(v)$ (such as $x = \mathrm{RED}_v([1,h])$ if v does not split in K), the stabilizer of x in $G(\mathbf{A}_f)^v \times F_v^\times$ may be computed as follows.*
Let B' be the quaternion algebra over F which is obtained from B by changing the invariants at v and τ_1: B' is totally definite and $\mathrm{Ram}_f B' = \mathrm{Ram}_f B \cup \{v\}$. Put $G' = \mathrm{Res}_{F/\mathbf{Q}}(B'^\times)$, a reductive group over $\mathbf{Q}$ with center Z. There exists an isomorphism $\phi_x^v : G(\mathbf{A}_f)^v \xrightarrow{\simeq} G'(\mathbf{A}_f)^v$ such that $(\phi_x^v, \mathbf{1}) : G(\mathbf{A}_f)^v \times F_v^\times \xrightarrow{\simeq} G'(\mathbf{A}_f)^v \times F_v^\times$ maps $\mathrm{Stab}(x)$ to the image of $G'(\mathbf{Q})\overline{Z(\mathbf{Q})} \subset G'(\mathbf{A}_f)$ through the (surjective) map

$$(\mathbf{1}, \mathrm{nr}_v) : G'(\mathbf{A}_f) = G'(\mathbf{A}_f)^v \times B'^\times_v \to G'(\mathbf{A}_f)^v \times F_v^\times.$$

Proof. In the compact case, this is exactly how Carayol describes the action of $G(\mathbf{A}_f)$ on a set which he denotes by S, cf. Proposition 11.2 of [1]. The fact that Carayol's set S equals our $\mathcal{X}(v)$ follows from the discussion of [1, Section 10.1]. The non-compact case is similar.

Define

$$G(v) \stackrel{\text{def}}{=} G'(\mathbf{A}_f)^v \times F_v^\times,$$

$$\phi_x(v) \stackrel{\text{def}}{=} (\phi_x^v, \mathrm{nr}_v) : G(\mathbf{A}_f) \twoheadrightarrow G(v),$$

and let $G(\mathbf{Q}, v)$ be the image of $G'(\mathbf{Q})$ in $G(v)$:

$$G(\mathbf{Q}, v) = (\mathbf{1}, \mathrm{nr}_v)(G'(\mathbf{Q})).$$

Corollary 3.10 *The map $g \mapsto x \cdot g$ factors through $\phi_x(v)$ and induces a bijection*

$$\overline{G(\mathbf{Q}, v)} \backslash G(v) \xrightarrow{\simeq} \mathcal{X}(v)$$

where $\overline{G(\mathbf{Q}, v)}$ is the closure of $G(\mathbf{Q}, v)$ in $G(v)$.

Proof. We have to show that $\overline{G(\mathbf{Q}, v)} = (\mathbf{1}, \mathrm{nr}_v)(G'(\mathbf{Q})\overline{Z(\mathbf{Q})})$. We first claim that $G'(\mathbf{Q})\overline{Z(\mathbf{Q})}$ is locally closed (hence closed) in $G'(\mathbf{A}_f)$. Indeed, let R be a maximal $\mathcal{O}_F$-order in B'. As $\overline{Z(\mathbf{Q})} = F^\times \overline{\mathcal{O}_F^\times}$,

$$\left(G'(\mathbf{Q})\overline{Z(\mathbf{Q})}\right) \cap \widehat{R}^\times = \left(B'^\times \overline{\mathcal{O}_F^\times}\right) \cap \widehat{R}^\times = R^\times \overline{\mathcal{O}_F^\times} = \cup_{\alpha \in R^\times / O_F^\times} \alpha \overline{\mathcal{O}_F^\times}$$

is closed because $[R^\times : \mathcal{O}_F^\times]$ is finite (use [20, p. 139]). The map $\mathrm{nr}_v : B'^\times_v \to F_v^\times$ is open and surjective with a *compact* kernel: it is therefore a closed map, and so is $(\mathbf{1}, \mathrm{nr}_v) : G'(\mathbf{A}_f) \to G(v)$. In particular, $(\mathbf{1}, \mathrm{nr}_v)(G'(\mathbf{Q})\overline{Z(\mathbf{Q})})$ is closed in $G(v)$, so that $\overline{G(\mathbf{Q}, v)} \subset (\mathbf{1}, \mathrm{nr}_v)(G'(\mathbf{Q})\overline{Z(\mathbf{Q})})$ and the other inclusion is trivial.

3.2.4 Reciprocity laws

We now want to describe the reciprocity laws for CM points and connected components, following [14] instead of [4] (see the remark at the end of [14, §12]). In particular: (1) reciprocity maps send uniformizers to geometric Frobenius; (2) Galois actions on geometric points are *left* actions.

Let $\mu : \mathbb{G}_{m,\mathbf{C}} \to T_\mathbf{C}$ be the cocharacter which is defined by $\mu(z) = s \circ r(z)$, where $r : \mathbb{G}_{m,\mathbf{C}} \to \mathbb{S}_\mathbf{C} \simeq \mathbb{G}_{m,\mathbf{C}} \times \mathbb{G}_{m,\mathbf{C}}$ maps z to $(z, 1)$ (and $\mathbb{S}_\mathbf{C} \simeq \mathbb{G}_{m,\mathbf{C}} \times \mathbb{G}_{m,\mathbf{C}}$ is induced by $z \otimes_\mathbf{R} a \mapsto (za, \bar{z}a)$ for $z \in \mathbf{C}$ and a in some $\mathbf{C}$-algebra A). The isomorphism

$$T_\mathbf{C} \xrightarrow{\lambda \otimes a \mapsto (\tau(\lambda)a)_\tau} \mathbb{G}_{m,\mathbf{C}}^{\mathrm{Hom}(K,\mathbf{C})}$$

yields a bijection between the set of cocharacters of T and $\mathbf{Z}^{\mathrm{Hom}(K,\mathbf{C})}$, with $\sigma \in \mathrm{Aut}(\mathbf{C})$ acting on the latter set by $(n_\tau)_\tau \cdot \sigma = (n_{\sigma\tau})_\tau$. The cocharacter μ

corresponds to $n_\tau = \epsilon$ if $\tau = \tau_1$ and $n_\tau = 0$ otherwise. In particular, the field of definition of μ equals $\tau_1(K) \simeq K$ and the morphism

$$T = \mathrm{Res}_{K/\mathbf{Q}}(\mathbb{G}_{m,K}) \xrightarrow{\mathrm{Res}_{K/\mathbf{Q}}(\mu)} \mathrm{Res}_{K/\mathbf{Q}}(T_K) \xrightarrow{\mathrm{Norm}_{K/\mathbf{Q}}} T$$

sends z to z^ϵ. We thus obtain:

Lemma 3.11 *The CM points are algebraic, defined over the maximal abelian extension K^{ab} of K. For $\sigma = \mathrm{rec}_K(\lambda)$ with $\lambda \in T(\mathbf{A}_f) = \widehat{K}^\times$, the action of σ on* $\mathrm{CM} \simeq \overline{T(\mathbf{Q})}\backslash G(\mathbf{A}_f)$ *is given by multiplication on the left by λ^ϵ.*

Similarly:

Lemma 3.12 *The connected components are defined over the maximal abelian extension F^{ab} of F. For $\sigma = \mathrm{rec}_F(\lambda)$ with $\lambda \in Z(\mathbf{A}_f) = \widehat{F}^\times$, the action of σ on $\mathcal{Z} \simeq \overline{Z(\mathbf{Q})}^+\backslash Z(\mathbf{A}_f)$ is given by multiplication by λ^ϵ.*

In particular, the pro-étale F-scheme $\mathcal{M} \stackrel{\text{def}}{=} \varprojlim \mathcal{M}_H$ together with its right action of $G(\mathbf{A}_f)$ is (non-canonically) isomorphic to $\mathrm{Spec}(F^{\mathrm{ab}})$ on which $G(\mathbf{A}_f)$ acts through $g \mapsto \mathrm{Spec}(\sigma)$ with $\sigma = \mathrm{rec}_F(\mathrm{nr}(g)^\epsilon)$, while $\boldsymbol{\mathcal{M}} \stackrel{\text{def}}{=} \varprojlim \boldsymbol{\mathcal{M}}_H$ is (non-canonically) isomorphic to the spectrum of the ring of v-integers in F^{ab}. It follows that the reduction map $\mathcal{Z} = \mathcal{M}(F^{\mathrm{ab}}) \to \mathcal{Z}(v) = \boldsymbol{\mathcal{M}}(\mathbf{F}(\overline{v}))$ identifies $\mathcal{Z}(v)$ with $\mathcal{Z}/\mathcal{O}_v^\times$ (viewing $\mathcal{O}_v^\times$ as a subgroup of $Z(\mathbf{A}_f) = Z(\mathbf{A}_f)^v \times F_v^\times$). Since $\mathcal{O}_v^\times$ also acts trivially on $\mathbf{M}^{ss}(v) = \varprojlim \mathbf{M}_H^{ss}(v)$ (cf. [1, section 11.2]), the projection $\mathcal{X}(v) \to \mathbf{M}^{ss}(v)$ also identifies $\mathbf{M}^{ss}(v)$ with $\mathcal{X}(v)/\mathcal{O}_v^\times$ (viewing now $\mathcal{O}_v^\times$ as a subgroup of $G(v) = G'(\mathbf{A}_f)^v \times F_v^\times$).

Corollary 3.13 *If* $\mathrm{nr}(H_v) = \mathcal{O}_v^\times$ (1) *$\boldsymbol{\mathcal{M}}_H$ is a finite étale $\mathcal{S}$-scheme,* (2) *$\mathcal{Z}_H \simeq \mathcal{Z}_H(v)$ and* (3) *$\mathcal{X}(v)/H \simeq \mathcal{X}_H(v) \simeq \mathbf{M}_H^{ss}(v)$.*

Proof. In general, $\boldsymbol{\mathcal{M}}_H$ is isomorphic to the spectrum of the ring of v-integers in the abelian extension F_H of F which is cut out by $\mathrm{rec}_K(\mathrm{nr}(H))$. If $\mathcal{O}_v^\times \subset \mathrm{nr}(H)$, F_H is unramified at v and $\boldsymbol{\mathcal{M}}_H$ is therefore a finite étale $\mathcal{S}$-scheme. This proves (1) and (2), and (2) implies that $\mathcal{X}_H(v) = \mathbf{M}_H^{ss}(v) \times_{\mathcal{Z}_H(v)} \mathcal{Z}_H \simeq \mathbf{M}_H^{ss}(v)$. Finally, since $\mathcal{X}(v)/\mathcal{O}_v^\times \simeq \mathbf{M}^{ss}(v)$, $\mathcal{X}(v)/H \simeq \mathbf{M}^{ss}(v)/H \simeq \mathbf{M}_H^{ss}(v)$.

Remark 3.14 The assumption $\mathrm{nr}(H_v) = \mathcal{O}_v^\times$ holds true when $H = \widehat{R}^\times$ for some Eichler order $R \subset B$.

3.2.5 Conclusion

Putting lemmas 3.7, 3.8 and Corollary 3.10 together, we obtain a commutative diagram

$$\begin{array}{ccccc}
\overline{T(\mathbf{Q})}\backslash G(\mathbf{A}_f) & \xrightarrow{(1)} & \overline{G(\mathbf{Q},v)}\backslash G(v) & \xrightarrow{(2)} & \overline{Z(\mathbf{Q})^+}\backslash Z(\mathbf{A}_f) \\
\simeq\downarrow & & \simeq\downarrow & & \simeq\downarrow \\
\mathrm{CM} & \xrightarrow{\mathrm{RED}_v} & \mathcal{X}(v) & \xrightarrow{c_v} & \mathcal{Z}
\end{array}$$

where (1) is induced by $\phi_x(v) : G(\mathbf{A}_f) \to G(v)$ (with $x = \mathrm{RED}_v([1,h])$) while (2) is induced by the morphism

$$\begin{array}{rcl}
G'(\mathbf{A}_f)^v \times F_v^\times = G(v) & \longrightarrow & Z(\mathbf{A}_f) = Z(\mathbf{A}_f)^v \times F_v^\times \\
(g^v, \lambda_v) & \longmapsto & (\mathrm{nr}(g^v), \lambda_v)
\end{array}$$

For a compact open subgroup H of $G(\mathbf{A}_f)$, put $H(v) = \phi_x(v)(H) \subset G(v)$. We thus obtain a diagram

$$\begin{array}{ccccc}
T(\mathbf{Q})\backslash G(\mathbf{A}_f)/H & \longrightarrow & G(\mathbf{Q},v)\backslash G(v)/H(v) & \longrightarrow & Z(\mathbf{Q})^+\backslash Z(\mathbf{A}_f)/\mathrm{nr}(H) \\
\simeq\downarrow & & \downarrow & & \simeq\downarrow \\
\mathrm{CM}_H & \xrightarrow{\mathrm{RED}_v} & \mathcal{X}_H(v) & \xrightarrow{c_v} & \mathcal{Z}_H
\end{array}$$

in which the middle vertical arrow is surjective (and a bijection when $H = \widehat{R}^\times$ for some Eichler order $R \subset B$). Theorem 3.5 is therefore a consequence of a special case (S) of Theorem 2.9, corresponding to the situation where $\mathfrak{S}$ (in the notations of Theorem 2.9) equals $\{\{v\}, v \in \mathfrak{S}\}$ (in the notations of Theorem 3.5).

3.3 Complements

3.3.1 On the parameter $\epsilon = \pm 1$

Let us fix an isomorphism of $\mathbf{R}$-algebras between B_{τ_1} and $M_2(\mathbf{R})$, thus obtaining an isomorphism of group schemes over $\mathbf{R}$ between G_1 and $\mathrm{GL}_2(\mathbf{R})$. Let X_ϵ be the $G(\mathbf{R})$-conjugacy class of the morphism $h_\epsilon : \mathbb{S} \to G_{\mathbf{R}}$ which sends $z = x + iy$ to

$$h_\epsilon(z) = \left[\begin{pmatrix} x & y \\ -y & x \end{pmatrix}^\epsilon, 1, \cdots, 1\right] \in G_1(\mathbf{R}) \times \cdots \times G_d(\mathbf{R}) \simeq G(\mathbf{R})$$

and let $\{M_H(\epsilon)\}$ be the corresponding collection of Shimura curves. We thus have a compatible system of isomorphisms $\psi_H(\epsilon) : M_H(\epsilon) \times_F \mathbf{C} \to M_H^{\mathrm{alg}}(\epsilon)$, where $M_H^{\mathrm{alg}}(\epsilon)$ is the algebraic curve over $\mathbf{C}$ whose underlying Riemann surface equals

$$M_H^{\mathrm{an}}(\epsilon) = G(\mathbf{Q})\backslash \left(G(\mathbf{A}_f)/H \times X_\epsilon\right).$$

The topology, the differentiable structure and the real analytic structure of X_ϵ are induced from those of $G(\mathbf{R})$ through the map $g \mapsto g h_\epsilon g^{-1}$. For $h \in X_\epsilon$ and

$z \in \mathbf{C}^\times = \mathbb{S}(\mathbf{R})$, the map $x \mapsto h(z)xh(z)^{-1}$ fixes h and therefore induces an $\mathbf{R}$-linear map $T_h(\mathrm{ad}h(z))$ on the tangent space T_hX_ϵ of X_ϵ at h. The almost complex structure on X_ϵ is characterized by the fact that $T_h(\mathrm{ad}h(z))$ acts by multiplication by $z/\bar{z}$ on T_hX_ϵ for all $h \in X_\epsilon$ and $z \in \mathbf{C}^\times$. This almost complex structure is known to be integrable.

Remark 3.15 Most authors replace X_ϵ by $\mathbf{C}-\mathbf{R}$ with $G(\mathbf{R})$ acting through the projection on the first component $G_1(\mathbf{R}) \simeq \mathrm{GL}_2(\mathbf{R})$, by $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \lambda = \frac{a\lambda+b}{c\lambda+d}$ ($\lambda \in \mathbf{C}-\mathbf{R}$). This corresponds to $\epsilon = 1$. Indeed, the map $gh_1g^{-1} \mapsto g \cdot i$ yields a diffeomorphism between X_1 and $\mathbf{C}-\mathbf{R}$ and for $z \in \mathbf{C}^\times$, the derivative of $\lambda \mapsto gh_1(z)g^{-1} \cdot \lambda$ at $\lambda = g \cdot i$ equals $z/\bar{z}$. On the other hand, our main reference [1] on Shimura curves very explicitly uses $\epsilon = -1$. While it seems clear that Carayol's constructions could easily be transfered to the $\epsilon = 1$ case, we will show below that the choice of ϵ is, in fact, irrelevant.

From the above discussion, we know that the $G(\mathbf{R})$-equivariant map $\Phi : X_\epsilon \to X_{-\epsilon}$ which sends h to h^{-1} is an antiholomorphic diffeomorphism. For any compact open subgroup H of $G(\mathbf{A}_f)$, Φ therefore induces an antiholomorphic diffeomorphism between $M_H^{\mathrm{an}}(\epsilon)$ and $M_H^{\mathrm{an}}(-\epsilon)$ and an antilinear isomorphism between $M_H^{\mathrm{alg}}(\epsilon)$ and $M_H^{\mathrm{alg}}(-\epsilon)$, namely an isomorphism of *schemes* $\Phi : M_H^{alg}(\epsilon) \to M_H^{alg}(-\epsilon)$ such that the diagram

$$\begin{array}{ccc} M_H^{\mathrm{alg}}(\epsilon) & \xrightarrow{\Phi} & M_H^{\mathrm{alg}}(-\epsilon) \\ \downarrow & & \downarrow \\ \mathrm{Spec}(\mathbf{C}) & \xrightarrow{\mathrm{Spec}(\tau)} & \mathrm{Spec}(\mathbf{C}) \end{array}$$

is commutative (τ=complex conjugation).
For any scheme X over $\mathrm{Spec}(\mathbf{C})$, we denote by $\tau X \to \mathrm{Spec}(\mathbf{C})$ the pull-back of $X \to \mathrm{Spec}(\mathbf{C})$ through $\mathrm{Spec}(\tau) : \mathrm{Spec}(\mathbf{C}) \to \mathrm{Spec}(\mathbf{C})$. The above diagram thus yields an isomorphism of complex curves between $M_H^{\mathrm{alg}}(\epsilon)$ and $\tau M_H^{\mathrm{alg}}(-\epsilon)$ which together with $\psi_H(\epsilon)$ and $\psi_H(-\epsilon)$ induces an isomorphism

$$\Phi' : M_H(\epsilon) \times_F \mathbf{C} \to M_H(-\epsilon) \times_F \mathbf{C} \simeq \tau(M_H(-\epsilon) \times_F \mathbf{C})$$

(recall that F is embedded in $\mathbf{C}$ through $\tau_1 : F \hookrightarrow \mathbf{R}$). In other words, $M_H(-\epsilon)$ is a twist of $M_H(\epsilon)$. We shall now determine this twist.
For $\sigma \in \mathrm{Aut}(\mathbf{C}/F)$, let $\rho(\sigma)$ be the F-automorphism of $M_H(\epsilon)$ such that $\rho(\sigma) \cdot [g, h] = [g\lambda, h]$ for $g \in G(\mathbf{A}_f)$ and $h \in X_\epsilon$, where λ is any element of $Z(\mathbf{A}_f)$ (the center of $G(\mathbf{A}_f)$) such that $\mathrm{rec}_F(\lambda) = \sigma$ in $\mathrm{Gal}_F^{\mathrm{ab}}$. One easily checks that $\sigma \mapsto \rho(\sigma)$ is a well-defined group homomorphism $\rho : \mathrm{Aut}(\mathbf{C}/F) \to \mathrm{Aut}_F(M_H(\epsilon))$ which factors through $\mathrm{Gal}(F'_H/F)$ where F'_H is the abelian extension of F corresponding to the subgroup $F^\times \cdot (Z(\mathbf{A}_f) \cap H)$ of $Z(\mathbf{A}_f) = \widehat{F}^\times$.

Lemma 3.16 *Φ' realizes $M_H(-\epsilon)$ as the twist of $M_H(\epsilon)$ by $\rho^{-\epsilon}$.*

Proof. On the level of complex points, Φ' is the composite of Φ with the action of complex conjugation. The latter is described by a conjecture of Langlands [10], proven in [13]. We obtain: for $x = [g, h] \in M_H^{\mathrm{an}}(\epsilon)$ (with $g \in G(\mathbf{A}_f)$ and $h \in X_\epsilon$), $\Phi'(x) = [g, \bar{h}^{-1}] \in M_H^{\mathrm{an}}(-\epsilon)$ where $\bar{h} : \mathbb{S} \to G_{\mathbf{R}}$ maps z to $h(\bar{z})$. Note that $h \mapsto \bar{h}$ is indeed an involution of X_ϵ since

$$\bar{h}_\epsilon(x+iy) = \left[\begin{pmatrix} x & -y \\ y & x \end{pmatrix}^\epsilon, 1, \cdots, 1 \right] = \omega h_\epsilon(x+iy)\omega^{-1}$$

where $\omega = \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, 1, \cdots, 1 \right] \in G(\mathbf{R})$.

If h is a special point of X_ϵ, $\bar{h}^{-1}$ is a special point of $X_{-\epsilon}$. More precisely, suppose that $h : \mathbb{S} \to G_{\mathbf{R}}$ factors through $T'_{\mathbf{R}}$ for some *maximal* $\mathbf{Q}$-rational subtorus $T' \subset G$. Let $\mu_h : \mathbb{G}_{m,\mathbf{C}} \to T'_{\mathbf{C}}$ be the induced cocharacter ($\mu_h(z) = h(z,1)$), let $E_h \subset \mathbf{C}$ be the field of definition of μ_h (so that $F \subset E_h$) and put $\mathrm{rec}_h = \mathrm{Norm}_{E_h/\mathbf{Q}} \circ \mathrm{Res}_{E_h/\mathbf{Q}}(\mu_h)$:

$$\mathrm{rec}_h : \mathrm{Res}_{E_h/\mathbf{Q}}(\mathbb{G}_{m,E_h}) \to \mathrm{Res}_{E_h/\mathbf{Q}}(T'_{E_h}) \to T'.$$

Let also $\mu_0 : \mathbb{G}_{m,\mathbf{C}} \to Z_{\mathbf{C}} \subset T'_{\mathbf{C}}$ be the cocharacter defined by

$$\mu_0(z) = \left[\begin{pmatrix} z & \\ & z \end{pmatrix}, 1, \cdots, 1 \right] \in Z(\mathbf{R}) \subset G(\mathbf{R}) \simeq G_1(\mathbf{R}) \times \cdots \times G_d(\mathbf{R}).$$

Then μ_0 is defined over F and

$$Z = \mathrm{Res}_{F/\mathbf{Q}}(\mathbb{G}_{m,F}) \overset{\mathrm{Res}_{F/\mathbf{Q}}(\mu_0)}{\longrightarrow} \mathrm{Res}_{F/\mathbf{Q}}(Z_F) \overset{\mathrm{Norm}_{F/\mathbf{Q}}}{\longrightarrow} Z$$

is the identity map. Since $\mu_h \cdot \mu_{\bar{h}} = \mu_0^\epsilon$, $\mu_{\bar{h}}$ is also defined over E_h and

$$\mathrm{rec}_h \cdot \mathrm{rec}_{\bar{h}} = \mathrm{Norm}^\epsilon_{E_h/F} : \mathrm{Res}_{E_h/\mathbf{Q}}(\mathbb{G}_{m,E_h}) \to \mathrm{Res}_{F/\mathbf{Q}}(\mathbb{G}_{m,F}) = Z \subset T'.$$

It follows that (1) for $g \in G(\mathbf{A}_f)$, both $x = [g, h]$ and $\Phi'(x) = [g, \bar{h}^{-1}]$ are defined over the maximal abelian extension $E_h^{\mathrm{ab}} \subset \mathbf{C}$ of E_h; (2) for $\lambda \in \widehat{E}_h^\times = \mathrm{Res}_{E_h/\mathbf{Q}}(\mathbf{A}_f)$ and $\sigma = \mathrm{rec}_{E_h}(\lambda) \in \mathrm{Gal}^{\mathrm{ab}}_{E_h}$,

$$\begin{aligned} \Phi'(\rho(\sigma)^{-\epsilon}(\sigma \cdot x)) &= [\mathrm{rec}_h(\lambda) g \mathrm{Norm}^{-\epsilon}_{E_h/F}(\lambda), \bar{h}^{-1}] \\ &= [\mathrm{rec}_h(\lambda) \mathrm{Norm}^{-\epsilon}_{E_h/F}(\lambda) g, \bar{h}^{-1}] \\ &= [\mathrm{rec}_{\bar{h}^{-1}}(\lambda) g, \bar{h}^{-1}] \\ &= \sigma \cdot \Phi'(x). \end{aligned}$$

Our claim now easily follows from the uniqueness of canonical models.

As a scheme over F, the twist $M_H(\epsilon)'$ of $M_H(\epsilon)$ by $\rho^{-\epsilon}$ may be constructed as the quotient of $M_H(\epsilon) \times_{\mathrm{Spec}(F)} \mathrm{Spec}(F'_H)$ by the (right) action of $\mathrm{Gal}(F'_H/F)$ which maps σ to the F-automorphism $\alpha(\sigma) = (\rho(\sigma)^\epsilon, \mathrm{Spec}(\sigma))$ of $M_H(\epsilon) \times_F \mathrm{Spec}(F'_H)$.

Lemma 3.17 *Suppose that $H = \bar{H}$ where $g \mapsto \bar{g}$ is the anticommutative involution of $G(\mathbf{A}_f)$ which is induced by the canonical involution of B. Then $M_H(\epsilon)'$ is isomorphic to $M_H(\epsilon)$. In particular, $M_H(\epsilon) \simeq M_H(-\epsilon)$.*

Proof. We shall construct an involution θ of $M_H(\epsilon)$ with the property that for all $\sigma \in \Gamma_H \overset{\text{def}}{=} \mathrm{Gal}(F'_H/F)$,

$$\theta \circ \alpha(\sigma) = (1 \times \mathrm{Spec}(\sigma)) \circ \theta \quad \text{on } M_H(\epsilon) \times_F \mathrm{Spec}(F'_H).$$

Such a θ induces an F-isomorphism between $M_H(\epsilon)'$ and $M_H(\epsilon)$.
Recall that $\mathcal{M}_H(\epsilon) = \mathrm{Spec}\left(\Gamma(\mathcal{O}_{M_H(\epsilon)}, M_H(\epsilon))\right)$ is non-canonically isomorphic to $\mathrm{Spec}(F_H)$ where F_H is the abelian extension of F cut out by $F^\times_{>0} \cdot \mathrm{nr}(H) \subset \widehat{F}^\times$. Since $(F^\times \cdot H \cap \widehat{F}^\times)^2 \subset F^\times_{>0} \cdot \mathrm{nr}(H)$, there is a well defined group homomorphism $\kappa : \Gamma_H \to \mathrm{Gal}(F_H/F)$ given by $\kappa(\mathrm{rec}_F(\lambda)) = \mathrm{rec}_F(\lambda^2)$ for $\lambda \in \widehat{F}^\times$. It follows from the discussion after Lemma 3.12 that for $\sigma \in \Gamma_H$ and $x \in M_H(\epsilon)(\mathbf{C})$,

$$c(\rho(\sigma)(x)) = \kappa(\sigma)^\epsilon \cdot c(x) \quad \text{in } \mathcal{M}_H(\epsilon)(\mathbf{C}) = \mathcal{M}_H(\epsilon)(F_H).$$

On the other hand, F'_H is a subfield of F_H whenever $F^\times_{>0} \cdot \mathrm{nr}(H) \subset F^\times \cdot H \cap \widehat{F}^\times$. This is indeed the case when $\bar{H} = H$. In particular, we may choose an F-morphism $\mathcal{M}_H(\epsilon) \to \mathrm{Spec}(F'_H)$, thus obtaining an F-morphism $c' : M_H(\epsilon) \to \mathrm{Spec}(F'_H)$ such that

$$\forall \sigma \in \Gamma_H, \quad c' \circ \rho(\sigma) = \mathrm{Spec}(\sigma^{2\epsilon}) \circ c'.$$

Let A be an F-algebra and let $z = (x, y)$ be an A-valued point of the F-scheme $M_H(\epsilon) \times_F \mathrm{Spec}(F'_H)$. Then $c'(x)$ and y are A-valued points of $\mathrm{Spec}(F'_H)$. If $\mathrm{Spec}(A)$ is connected, there exists a unique element $\gamma \overset{\text{def}}{=} \gamma(z)$ in Γ_H such that $c'(x) = \mathrm{Spec}(\gamma^{-\epsilon}) \circ y$. This defines an F-morphism $z \mapsto \gamma(z)$ from $M_H(\epsilon) \times_F \mathrm{Spec}(F'_H)$ to the constant F-scheme Γ_H. For $z = (x, y)$ as above, we put $\theta(z) = (\rho(\gamma(z))(x), y)$. One easily checks that θ has the required properties.

When $H = \bar{H}$, we thus obtain an F-isomorphism between $M_H(\epsilon)$ and $M_H(-\epsilon)$. On the level of complex points, such an isomorphism is given by

$$[g, h] \in M_H^{\mathrm{an}}(\epsilon) \mapsto [\bar{g}^{-1}, \bar{h}^{-1}] \in M_H^{\mathrm{an}}(-\epsilon).$$

Note that the condition $H = \bar{H}$ defines a cofinal subset of the set of all compact open subgroups H of $G(\mathbf{A}_f)$. Also, $H = \bar{H}$ when $H = \widehat{R}^\times$ for some Eichler order R in B, in which case F_H and F'_H are respectively the Hilbert class field and the narrow Hilbert class field of F.

3.3.2 P-rational elements of $\mathrm{Gal}_K^{\mathrm{ab}}$.

It may seem rather surprising that the bizarre subgroup $\mathrm{Gal}_K^{P-\mathrm{rat}}$ of P-rational elements in $\mathrm{Gal}_K^{\mathrm{ab}}$ should play any role in the theory of CM points. For instance,

$\mathrm{Gal}_K^{P-\mathrm{rat}}$ is *not* a closed subgroup of $\mathrm{Gal}_K^{\mathrm{ab}}$, although it contains the closed subgroup

$$\mathrm{Gal}(K^{\mathrm{ab}}/K[P^\infty]) = \mathrm{rec}_K \left\{ \lambda \in \widehat{O}_K^\times, \lambda_P \in O_{F_P}^\times \right\}.$$

The Galois group $\mathrm{Gal}(K[P^\infty]/K)$ is topologically isomorphic to $G_0 \times \mathbf{Z}_p^{[F_P:\mathbf{Q}_p]}$ where p is the residue characteristic of P and G_0 is a finite group, the torsion subgroup of $\mathrm{Gal}(K[P^\infty]/K)$. The subfield of $K[P^\infty]$ which is fixed by G_0 is the composite of all $\mathbf{Z}_p$-extensions of K which are unramified outside P and Galois and dihedral over F. The image of $\mathrm{Gal}_K^{P-\mathrm{rat}}$ in $\mathrm{Gal}(K[P^\infty]/K)$ is a *dense* but *countable* subgroup which is generated by the Frobeniuses of those primes of K which are not above P (the intersection of this subgroup with G_0 plays a key role in [3], where it is denoted by G_1). In particular, $\mathrm{Gal}_K^{P-\mathrm{rat}}$ is a *dense* but *negligible* (i.e. measurable with trivial measure) subgroup of $\mathrm{Gal}_K^{\mathrm{ab}}$. The map $\sigma = \mathrm{rec}_K(\lambda) \mapsto \lambda_P$ yields a bijection between $\mathrm{Gal}_K^{\mathrm{ab}}/\mathrm{Gal}_K^{P-\mathrm{rat}}$ and $K_P^\times/K^\times F_P^\times$.

However, the appearance of rational elements is perhaps less surprising when one recalls that the present work originated in the study of elliptic curves over anticyclotomic towers of number fields, since the distinction between suitably defined rational and irrational elements of Galois groups occurs quite frequently in the context of Iwasawa theory. For instance, the celebrated theorems of Ferrero and Washington on the growth of class numbers in $\mathbf{Z}_p$ extensions of abelian fields rely crucially on the fact that nontrivial roots of unity are irrational. Another example of this occurs in recent work of Hida [7], [8] on anticyclotomic families of Hecke characters, where the key observation is the irrationality of certain Galois actions on Serre-Tate deformation spaces. In fact, the irrationality arguments given by Ferrero and Washington were the original motivation for the introduction in [18] of rational and irrational elements to the study of CM points.

In this section, we shall provide some further evidence for the relevance of P-rational elements by relating them to the André-Oort conjecture:

Proposition 3.18 *For $\sigma \in \mathrm{Gal}_K^{ab}$ and $x \in \mathrm{CM}_H$, put $\delta(x) = (x, \sigma x) \in M_H(\mathbf{C})^2$. The following conditions are equivalent.*

1. *σ is a P-rational element.*
2. *For any collection $\mathcal{E} \subset \mathrm{CM}_H$ of P-isogeneous CM points, the Zariski closure of $\delta(\mathcal{E})$ in $(M_H \times_F \mathbf{C})^2$ has dimension ≤ 1.*
3. *For some collection $\mathcal{E} \subset \mathrm{CM}_H$ of P-isogeneous CM points, the Zariski closure of $\delta(\mathcal{E})$ has dimension 1.*

For the proof of this proposition, we may and do assume that $H = \widehat{R}^\times$ for some maximal order $R \subset B$. For any CM point

$$x = [g] \in \mathrm{CM}_H = T(\mathbf{Q})\backslash G(\mathbf{A}_f)/H,$$

the stabilizer of x in $\mathrm{Gal}_K^{\mathrm{ab}}$ then equals $\mathrm{rec}_K(K^\times \mathcal{O}(x)^\times)$ where $\mathcal{O}(x) = K \cap gHg^{-1}$ is an $\mathcal{O}_F$-order in B. Moreover, there exists a unique integral ideal $\mathcal{C} \subset \mathcal{O}_F$ such that $\mathcal{O}(x) = \mathcal{O}_{K,\mathcal{C}} \stackrel{\mathrm{def}}{=} \mathcal{O}_F + \mathcal{C}\mathcal{O}_K$. We refer to $\mathcal{C} \stackrel{\mathrm{def}}{=} \mathcal{C}(x)$ as *the conductor* of x and denote by $\ell_P(x) \geq 0$ the exponent of P in $\mathcal{C}(x)$, so that $\mathcal{C}(x) = \mathcal{C}_0(x)P^{\ell_P(x)}$ for some integral ideal $\mathcal{C}_0(x)$ which is relatively prime to P. By construction, $x \mapsto \mathcal{C}(x)$ is constant on $\mathrm{Gal}_K^{\mathrm{ab}}$-orbits while $x \mapsto \mathcal{C}_0(x)$ is constant on P-isogeny classes. It follows from [20, pp. 42-44] that the fibers of $\mathcal{C}$ are finite. In particular:

Lemma 3.19 *The function $x \mapsto \ell_P(x)$ has finite fibers on any P-isogeny class.*

This function is related to the usual distance d on the Bruhat-Tits tree

$$\mathcal{T} = F_P^\times \backslash B_P^\times / R_P^\times \simeq F_P^\times \backslash \mathrm{GL}_2(F_P)/\mathrm{GL}_2(\mathcal{O}_{F_P}).$$

Indeed, the group $K_P^\times$ acts on the left on $\mathcal{T}$ by isometries, and for $v = [b] \in \mathcal{T}$ (with $b \in B_P^\times$), the stabilizer of v in $K_P^\times$ equals $F_P^\times \mathcal{O}(v)^\times$where $\mathcal{O}(v) = K_P \cap bR_Pb^{-1}$ is an $\mathcal{O}_{F_P}$-order in K_P. Just as above, there exists a unique integer $n \stackrel{\mathrm{def}}{=} n(v) \in \mathbf{N}$ such that $\mathcal{O}(v) = \mathcal{O}_n$ with $\mathcal{O}_n \stackrel{\mathrm{def}}{=} \mathcal{O}_{F_P} + P^n \mathcal{O}_{K_P}$ ($\mathcal{O}_n$ is the completion of $\mathcal{O}_{K,\mathcal{C}_0 P^n}$ at P for any integral ideal $\mathcal{C}_0 \subset \mathcal{O}_F$ which is relatively prime to P). It is clear that for a CM point $x = [g] \in \mathrm{CM}_H$, $\ell_P(x) = n(v)$ where $v = [g_P]$ ($g_P \in B_P^\times$ is the P-component of $g \in G(\mathbf{A}_f)$).
It is well-known that

- The map $v \mapsto n(v)$ yields a bijection between $K_P^\times \backslash \mathcal{T}$ and $\mathbf{N}$.
- The subset $\mathcal{T}_0 = \{v \in \mathcal{T};\, n(v) = 0\}$ of $\mathcal{T}$ consists of a vertex, two adjacent vertices or the set of vertices on a line in $\mathcal{T}$, depending upon whether P is inert, ramifies or splits in K.
- For any $v \in \mathcal{T}$, $n(v)$ is also the distance between v and $\mathcal{T}_0$.

In particular, suppose that $(v_n, v_{n-1}, \cdots, v_0)$ and $(w_m, w_{m-1}, \cdots, w_0)$ are geodesics in $\mathcal{T}$ from $v = v_n$ and $w = w_m$ to $\mathcal{T}_0$. Then $n(v_i) = i$ for $0 \leq i \leq n$ and $n(w_j) = j$ for $0 \leq j \leq m$. The geodesic γ between v and w may then be computed as follows:

- if $v_0 \neq w_0$, $\gamma = (v_n, v_{n-1}, \cdots, v_0, u_1, \cdots, u_{r-1}, w_0, w_1, \cdots, w_m)$ where $(v_0, u_1, \cdots, u_{r-1}, w_0)$ is the geodesic between v_0 and w_0 inside the connected subtree $\mathcal{T}_0$ of $\mathcal{T}$.
- if $v_0 = w_0$, $\gamma = (v_n, v_{n-1}, \cdots, v_c = w_c, w_{c+1}, \cdots, w_m)$ where c is the largest integer $\leq n, m$ such that $v_c = w_c$.

In the special case where $w = \lambda v$ for some $\lambda \in K_P^\times$, $n = m = n(v)$ and $w_i = \lambda v_i$ for $0 \leq i \leq n$. If moreover $d(v, \lambda v) \leq 2n$, it thus must be that $v_0 = w_0$. With c as above, $d(v, w) = 2(n - c)$ and $v_c = w_c = \lambda v_c$, so that λ belongs to $F_P^\times \mathcal{O}_c^\times$.
We have obtained:

Lemma 3.20 *Suppose that $d(v,\lambda v)\leq 2n(v)$ (with $v\in\mathcal{T}$ and $\lambda\in K_P^\times$). Then $d(v,\lambda v)=2k$ for some $k\in\{0,\cdots,n(v)\}$ and λ belongs to $F_P^\times\mathcal{O}_{n(v)-k}^\times$.*

We may now sketch the proof of Proposition 3.18. Of course, (2) implies (3).

Proof that (1) implies (2)

We have to show that for any P-isogeny class $\mathcal{H}\subset CM_H$, $\delta(\mathcal{H})$ is contained in a one dimensional subscheme of $(M_H\times_F\mathbf{C})^2$ if $\sigma=\mathrm{rec}_K(\lambda)$ for some $\lambda\in T(\mathbf{A}_f)=\widehat{K}^\times$ whose P-component λ_P belongs to $K^\times F_P^\times$. Choose $g_0\in G(\mathbf{A}_f)$ such that

$$\mathcal{H}=T(\mathbf{Q})\backslash T(\mathbf{Q})B_P^\times g_0H/H \quad \text{inside} \quad \mathrm{CM}_H=T(\mathbf{Q})\backslash G(\mathbf{A}_f)/H.$$

Using Lemma 3.11, we find that

$$\begin{aligned}\delta(\mathcal{H}) &= \left\{([bg_0],[\lambda^\epsilon bg_0])\,;\,b\in B_P^\times\right\}\\ &= \left\{([bg_0],[bg_0\gamma])\,;\,b\in B_P^\times\right\}\end{aligned}$$

where $\gamma=g_0^{-1}\lambda^\epsilon g_0$. Indeed, $bg_0\gamma=b\lambda^\epsilon g_0=\lambda^\epsilon bg_0$ for any $b\in B_P^\times$ as λ_P^ϵ belongs to $F_P^\times$. In particular, $\delta(\mathcal{H})$ is contained in the 1-dimensional image of the (algebraic!) morphism $M_{H\cap\gamma H\gamma^{-1}}\to M_H^2$ which sends $[g,h]$ to $([g,h],[g\gamma,h])$ $(g\in G(\mathbf{A}_f),\,h\in X)$.

Proof that (3) implies (1)

Write $\sigma=\mathrm{rec}_K(\lambda)$ with $\lambda\in\widehat{K}^\times$. Suppose that the Zariski closure of $\delta(\mathcal{E})$ in $(M_H\times_F\mathbf{C})^2$ has dimension 1 for some (infinite) collection $\mathcal{E}$ of P-isogeneous CM points. We have to show that the P-component $\lambda_P\in K_P^\times$ of λ belongs to $F_P^\times K^\times$.

By a proven case of the André-Oort conjecture [6, Theorem 1.2] there exists an infinite subset $\mathcal{E}'\subset\mathcal{E}$ and some element $\gamma\in G(\mathbf{A}_f)$ such that $\delta(\mathcal{E}')$ is contained in the image of a morphism $M_{H\cap\gamma H\gamma^{-1}}\to M_H^2$ as above. Fix $x=[g_0]\in\mathcal{E}'$ and let $\{g_1,\cdots,g_s\}\subset H$ be a set of representatives for $H/H\cap\gamma H\gamma^{-1}$. For each $x'=[bg_0]\in\mathcal{E}'$ (with $b\in B_P^\times$), we know that $x=[bg_0g_i]$ for any $i\in\{1,\cdots,s\}$ while

$$\sigma\cdot x\in\{[bg_0g_1\gamma],\cdots,[bg_0g_s\gamma]\}\,.$$

Replacing g_0 by g_0g_i for a suitable $1\leq i\leq s$ and using lemmas 3.11 and 3.19, we obtain: there exists a sequence $b_n\in B_P^\times$ such that (a) $\varphi(n)\overset{\text{def}}{=}\ell_P([b_ng_0])$ goes to infinity with n and (b) $[\lambda^\epsilon b_ng_0]=[b_ng_0\gamma]$ for all $n\geq 0$. By (b), there exists $\lambda_n\in T(\mathbf{Q})=K^\times$ and $h_n\in H$ such that for all $n\geq 0$,

$$\lambda_n\lambda^\epsilon b_ng_0=b_ng_0\gamma h_n \quad \text{in } G(\mathbf{A}_f).$$

Put $v_n = [b_n g_{0,P}] \in \mathcal{T}$ and $\mu_n = \lambda_n \lambda^\epsilon$. Since $\mu_{n,P} \cdot v_n = [b_n g_{0,P} \gamma_P]$, $d(v_n, \mu_{n,P} \cdot v_n) = d_0$ does not depend upon n. Pick $N \geq 0$ such that $\forall n \geq N$, $d_0 \leq 2\varphi(n)$. By Lemma 3.20, $d_0 = 2k$ and

$$\forall n \geq N: \quad \mu_{n,P} \in F_P^\times \mathcal{O}_{\varphi(n)-k}^\times. \tag{8}$$

On the other hand, $\mu_n \mu_N^{-1} = b_n g_0 \gamma h_n g_0^{-1} b_n^{-1} b_N g_0 h_N^{-1} \gamma^{-1} g_0^{-1} b_N^{-1}$. Away from P, this equation simplifies to $(\mu_n \mu_N^{-1})^P = (g_0 \gamma h_n h_N^{-1} \gamma^{-1} g_0^{-1})^P$, so that

$$(\mu_n \mu_N^{-1})_Q \in K_Q^\times \cap (g_0\gamma)_Q R_Q^\times (g_0\gamma)_Q^{-1} \subset \mathcal{O}_{K_Q}^\times \tag{9}$$

for all $Q \neq P$.
Let $U_F \subset U_K$ be the groups of all elements $z \in F^\times$ (resp. $z \in K^\times$) which are units away from P. Since K is a totally imaginary quadratic extension of F, $\mathrm{rank}_{\mathbf{Z}} U_K = \mathrm{rank}_{\mathbf{Z}} U_F$ if P does not split in K and $\mathrm{rank}_{\mathbf{Z}} U_K = \mathrm{rank}_{\mathbf{Z}} U_F + 1$ otherwise. Let U'_K be the subgroup of U_K defined by $U'_K = U_K \cap F_P^\times O_{K_P}^\times$. Then $U_F \subset U'_K \subset U_K$, and $[U_F : U'_K]$ is finite. Let $\mathcal{R} \subset U'_K$ be a set of representatives for U'_K/U_F.
By (9), $\mu_n \mu_N^{-1} = \lambda_n \lambda_N^{-1}$ belongs to U_K for all $n \geq 0$. Then (8) shows that $\mu_n \mu_N^{-1}$ belongs to U'_K for all $n \geq N$. For such n's, we may thus write

$$\lambda_n = \lambda_N r(n) u(n) \quad \text{with } r(n) \in \mathcal{R} \text{ and } u(n) \in U_F.$$

Using (8) again, we find that $\lambda_N r(n) \lambda_P^\epsilon$ belongs to $F_P^\times \mathcal{O}_{\varphi(n)-k}^\times$ for all $n \geq N$. Choosing a subsequence on which $r(n) = r$ is constant, we finally obtain:

$$\lambda_N r \lambda_P^\epsilon \in F_P^\times = \cap_{n \geq 0} F_P^\times \mathcal{O}_n^\times.$$

Since $\lambda_N r$ belongs to $K^\times$, λ_P indeed belongs to $K^\times F_P^\times$.

Remark 3.21 More generally, it may be shown that for any infinite collection $\mathcal{E} \subset \mathrm{CM}_H$ of P-isogeneous CM points and any finite subset $\mathcal{R} = \{\sigma_1, \cdots, \sigma_r\}$ of $\mathrm{Gal}_K^{\mathrm{ab}}$, the Zariski closure of $\{(\sigma_1 x, \cdots, \sigma_r x); x \in \mathcal{E}\}$ in $V = (M_H \times_F \mathbf{C})^r$ contains a connected component of V if and only if the σ_i's are pairwise distinct modulo $\mathrm{Gal}_K^{P-\mathrm{rat}}$ (Hint: use section 7.3 of [6] and a variant of Proposition 2.1 of [5]).

References

[1] H. Carayol. Sur la mauvaise réduction des courbes de Shimura. *Compositio Math.*, 59(2):151–230, 1986.

[2] C. Cornut. Mazur's conjecture on higher Heegner points. *Invent. Math.*, 148(3):495–523, 2002.

[3] C. Cornut and V. Vatsal. Nontriviality of Rankin-Selberg L-functions and CM points. Preprint, 2004.

[4] P. Deligne. Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques. In *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2*, Proc. Sympos. Pure Math., XXXIII, pages 247–289. Amer. Math. Soc., Providence, R.I., 1979.

[5] B. Edixhoven. Special points on product of modular curves. *Duke Math J.*, 126(2):325–348, 2005.

[6] B. Edixhoven and A. Yafaev. Subvarieties of Shimura varieties. *Ann. of Math. (2)*, 157(2):621–645, 2003.

[7] H. Hida. The mu-invariant of p-adic Hecke L-functions. Available from http://www.math.ucla.edu/, 2001.

[8] H. Hida. Non-vanishing modulo p of Hecke L-values. In *Geometric Aspects of Dwork Theory*, pages 731–780. DeGruyter, 2004.

[9] N. Katz and B. Mazur. *Arithmetic moduli of elliptic curves*, volume 108 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1985.

[10] R. Langlands. Some contemporary problems with origins in the Jugendtraum. In *Mathematical developments arising from Hilbert problems (Proc. Sympos. Pure Math., Vol. XXVIII, Northern Illinois Univ., De Kalb, Ill., 1974)*, pages 401–418. Amer. Math. Soc., Providence, R. I., 1976.

[11] G. A. Margulis and G. M. Tomanov. Invariant measures for actions of unipotent groups over local fields on homogeneous spaces. *Invent. Math.*, 116(1-3):347–392, 1994.

[12] B. Mazur. Modular curves and arithmetic. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*, pages 185–211. PWN, 1984.

[13] J. Milne. The action of an automorphism of $\mathbf{C}$ on a Shimura variety and its special points. In *Arithmetic and geometry, Vol. I*, volume 35 of *Progr. Math.*, pages 239–265. Birkhäuser Boston, Boston, MA, 1983.

[14] J. Milne. Introduction to Shimura varieties. Available from http://www.jmilne.org/math/Preprints/, 2004.

[15] M. Ratner. Raghunathan's conjectures for Cartesian products of real and p-adic Lie groups. *Duke Math. J.*, 77(2):275–382, 1995.

[16] J.-P. Serre and J. Tate. Good reduction of abelian varieties. *Ann. of Math. (2)*, 88:492–517, 1968.

[17] N. Shah. Uniformly distributed unipotent trajectories on products of $SL_2(K)/\Gamma$. To appear, 2004.

[18] V. Vatsal. Uniform distribution of Heegner points. *Invent. Math.*, 148:1–46, 2002.

[19] V. Vatsal. Special values of anticylotomic L-functions. *Duke Math J.*, 116(2):219–261, 2003.

[20] M.-F. Vignéras. *Arithmétique des algèbres de quaternions*, volume 800 of *Springer Lecture Notes*. Springer-Verlag, 1980.

Arakelov Invariants of Riemann Surfaces

Robin de Jong

Received: August 25, 2004
Revised: June 16, 2005

Communicated by Thomas Peternell

Abstract. We derive closed formulas for the Arakelov-Green function and the Faltings delta-invariant of a compact Riemann surface.

2000 Mathematics Subject Classification: 14G40, 14H42, 14H55.
Keywords and Phrases: Arakelov theory, delta-invariant, Green function, Weierstrass points.

1. Introduction

The main goal of this paper is to give closed formulas for the Arakelov-Green function G and the Faltings delta-invariant δ of a compact Riemann surface. Both G and δ are of fundamental importance in the Arakelov theory of arithmetic surfaces [2] [8] and it is a central problem in this theory to relate these difficult invariants to more accessible ones. For example, in [8] Faltings gives formulas which relate G and δ for elliptic curves directly to theta functions and to the discriminant modular form. Formulas of a similar explicit nature were derived by Bost in [3] for Riemann surfaces of genus 2. As to the case of general genus, less specific but still quite explicit formulas are known due to Bost [3] (for the Arakelov-Green function) and to Bost and Gillet-Soulé [4] [10] (for the delta-invariant). We recall these results in Sections 2 and 4 below.
In the present paper we express G and δ in terms of two new invariants S and T. Both S and T are initially defined as the norms of certain isomorphisms between line bundles, but eventually we find that they admit a very explicit description in terms of theta functions. They are intimately related to the divisor $\mathcal{W}$ of Weierstrass points. Of these new invariants, the T is certainly the easiest one. We are able to calculate it for hyperelliptic Riemann surfaces [13], where it is essentially a power of the Petersson norm of the discriminant modular form. The invariant S is less easy and involves a certain integral over the Riemann surface. We believe that the approach using S and T is very suitable for obtaining numerical results. An example at the end of this paper, where we compute δ and a special value of G for a certain hyperelliptic Riemann surface of genus 3, is meant to illustrate this belief.

We start our discussion by recalling the definitions of G and δ. From now on until the end of section 4, we fix a compact Riemann surface X. Let g be its genus, which we assume to be positive. The space of holomorphic differentials $H^0(X,\Omega^1_X)$ carries a natural hermitian inner product $(\omega,\eta) \mapsto \frac{i}{2}\int_X \omega \wedge \overline{\eta}$. We fix this inner product once and for all. Let $\{\omega_1,\ldots,\omega_g\}$ be an orthonormal basis with respect to this inner product. We have then a fundamental (1,1)-form μ on X given by $\mu = \frac{i}{2g}\sum_{k=1}^g \omega_k \wedge \overline{\omega}_k$. It is verified immediately that the form μ does not depend on the choice of orthonormal basis, and hence is canonical. Using this form, one defines the Arakelov-Green function G on $X \times X$. This function gives the local intersections "at infinity" of two divisors in Arakelov theory [2].

Theorem 1.1. *(Arakelov) There exists a unique function $G : X \times X \to \mathbb{R}_{\geq 0}$ satisfying the following properties for all $P \in X$:*

(i) *the function $\log G(P,Q)$ is C^∞ for $Q \neq P$;*
(ii) *we can write $\log G(P,Q) = \log|z_P(Q)| + f(Q)$ locally about P, where z_P is a local coordinate about P and where f is C^∞ about P;*
(iii) *we have $\partial_Q\overline{\partial}_Q \log G(P,Q)^2 = 2\pi i \mu(Q)$ for $Q \neq P$;*
(iv) *we have $\int_X \log G(P,Q)\mu(Q) = 0$.*

Theorem 1.1 is proved in [2]. We call the function G the Arakelov-Green function of X. We note that by an application of Stokes' theorem one can prove the symmetry relation $G(P,Q) = G(Q,P)$ for any $P,Q \in X$.

Importantly, the Arakelov-Green function gives rise to certain canonical metrics on line bundles on X. First, consider line bundles of the form $O_X(P)$ with P a point on X. Let s be the canonical generating section of $O_X(P)$. We then define a smooth hermitian metric $\|\cdot\|_{O_X(P)}$ on $O_X(P)$ by putting $\|s\|_{O_X(P)}(Q) = G(P,Q)$ for any $Q \in X$. By property (iii) of the Arakelov-Green function, the curvature form of $O_X(P)$ is equal to μ. Second, it is clear that the function G can be used to put a hermitian metric on the line bundle $O_{X\times X}(\Delta_X)$, where Δ_X is the diagonal on $X \times X$, by putting $\|s\|(P,Q) = G(P,Q)$ for the canonical generating section s of $O_{X\times X}(\Delta_X)$. Restricting to the diagonal, we have a canonical adjunction isomorphism $O_{X\times X}(-\Delta_X)|_{\Delta_X} \xrightarrow{\sim} \Omega^1_X$. We define a hermitian metric $\|\cdot\|_{\mathrm{Ar}}$ on Ω^1_X by insisting that this adjunction isomorphism be an isometry. It is proved in [2] that this gives a smooth hermitian metric on Ω^1_X, and that its curvature form is a multiple of μ. For the rest of the paper we shall take these metrics on $O_X(P)$ and Ω^1_X (as well as on tensor product combinations of them) for granted and refer to them as Arakelov metrics.

Next we introduce the Faltings delta-invariant. Let $\mathcal{H}_g$ be the generalised Siegel upper half plane of complex symmetric $g \times g$-matrices with positive definite imaginary part. Let $\tau \in \mathcal{H}_g$ be a period matrix associated to a symplectic basis of $H_1(X,\mathbb{Z})$ and consider the analytic jacobian $\mathrm{Jac}(X) = \mathbb{C}^g/\mathbb{Z}^g + \tau\mathbb{Z}^g$ associated to τ. We fix τ for the rest of our discussion. On $\mathbb{C}^g$ one has a theta function $\vartheta(z;\tau) = \sum_{n\in\mathbb{Z}^g} \exp(\pi i {}^t n \tau n + 2\pi i {}^t n z)$, giving rise to an effective divisor Θ_0 and a line bundle $O(\Theta_0)$ on $\mathrm{Jac}(X)$. Now consider on the other hand the set $\mathrm{Pic}_{g-1}(X)$ of divisor classes of degree $g-1$ on X. It comes with a special subset Θ given by the classes of effective divisors. A fundamental theorem of Abel-Jacobi-Riemann says that there is a canonical bijection $\mathrm{Pic}_{g-1}(X) \xrightarrow{\sim} \mathrm{Jac}(X)$ mapping Θ onto Θ_0. As a result, we can equip $\mathrm{Pic}_{g-1}(X)$ with the structure of a compact complex manifold, together with a divisor Θ and a line bundle $O(\Theta)$. We fix this structure for the rest of the discussion.

The function ϑ is not well-defined on $\mathrm{Pic}_{g-1}(X)$ or $\mathrm{Jac}(X)$. We can remedy this by putting $\|\vartheta\|(z;\tau) = (\det \mathrm{Im}\,\tau)^{1/4} \exp(-\pi {}^t y (\mathrm{Im}\,\tau)^{-1} y)|\vartheta(z;\tau)|$, with $y = \mathrm{Im}\, z$. One can check that $\|\vartheta\|$ descends to a function on $\mathrm{Jac}(X)$. By our identification $\mathrm{Pic}_{g-1}(X) \xrightarrow{\sim} \mathrm{Jac}(X)$ we obtain $\|\vartheta\|$ as a function on $\mathrm{Pic}_{g-1}(X)$. It can be checked that this function is independent of the choice of τ.

The delta-invariant is the constant appearing in the following theorem, due to Faltings (cf. [8], p. 402).

Theorem 1.2. *(Faltings) There is a constant $\delta = \delta(X)$ depending only on X such that the following holds. Let $\{\omega_1, \ldots, \omega_g\}$ be an orthonormal basis of $H^0(X, \Omega^1_X)$. Let $P_1, \ldots, P_g, Q$ be points on X with $P_1, \ldots, P_g$ pairwise distinct. Then the formula*

$$\|\vartheta\|(P_1 + \cdots + P_g - Q) = \exp(-\delta(X)/8) \cdot \frac{\|\det \omega_k(P_l)\|_{\mathrm{Ar}}}{\prod_{k<l} G(P_k, P_l)} \cdot \prod_{k=1}^{g} G(P_k, Q)$$

holds.

The definition of the delta-invariant may seem quite complicated, yet it plays an important role in Arakelov intersection theory and in the function theory of the moduli space $\mathcal{M}_g$ of Riemann surfaces of genus g. In fact, as has become clear from certain asymptotic results [14] [20], the function $\exp(-\delta(X))$ can be interpreted as a natural "distance" function on $\mathcal{M}_g$ measuring the distance to the Deligne-Mumford boundary. As to Arakelov theory, the delta-invariant plays the role of an archimedean contribution in the Noether formula for arithmetic surfaces [8] [18]. The idea that $\delta(X)$ gives a distance to the boundary is supported by this formula.

The plan of this paper is as follows. In Section 2 we state a proposition and observe that it leads quickly to a formula for G. In Section 3 we prove this proposition. In Section 4 we derive our closed formula for δ. Some applications of our results to Arakelov intersection theory are given in Section 5. We conclude with a numerical example in Section 6.

2. The Arakelov-Green function

As was mentioned in the Introduction, the Weierstrass points of X play an important role in our approach to G and δ. The idea of considering Weierstrass points in the context of Arakelov theory is not new, cf. [6] and [14] for example. We start by recalling how we obtain the divisor of Weierstrass points using a Wronskian differential on X. Let $\{\psi_1, \ldots, \psi_g\}$ be an (arbitrary) basis of

$H^0(X,\Omega_X^1)$. Let P be a point on X and let z be a local coordinate about P. Write $\psi_k = f_k(z)dz$ for $k=1,\ldots,g$. We have a holomorphic function

$$W_z(\psi) = \det\left(\frac{1}{(l-1)!}\frac{d^{l-1}f_k}{dz^{l-1}}\right)_{1\leq k,l\leq g}$$

locally about P from which we build the $g(g+1)/2$-fold holomorphic differential

$$\tilde{\psi} = W_z(\psi)(dz)^{\otimes g(g+1)/2}.$$

We call $\tilde{\psi}$ the Wronskian differential about P and it is readily checked that $\tilde{\psi}$ is independent of the choice of the local coordinate. In fact, and this is less trivial, the differential $\tilde{\psi}$ extends over X to give a non-zero global section of the line bundle $\Omega_X^{\otimes g(g+1)/2}$. A change of basis changes the Wronskian differential by a non-zero scalar factor and hence the divisor of a Wronskian differential $\tilde{\psi}$ on X is unique. We denote this divisor by $\mathcal{W}$, the divisor of Weierstrass points. This divisor is effective and we have $\deg \mathcal{W} = g^3 - g$. Writing $\mathcal{W} = \sum_{P\in X} w(P)\cdot P$ we call the integer $w(P)$ the *weight* at P. This weight can be calculated using gap sequences, but we shall not need this.
Now fix for the moment a $Q \in X$. We consider the map $\phi_Q : X \to \mathrm{Pic}_{g-1}(X)$ given by sending $P \mapsto [gP - Q]$. We put a smooth hermitian metric on $O(\Theta)$ by setting $\|s\| = \|\vartheta\|$ where s is the canonical generating section of $O(\Theta)$. We shall refer to this metric as the Arakelov metric on $O(\Theta)$. It can be verified by a short calculation using Riemann's bilinear relations that $\phi_Q^* O(\Theta)$ is a line bundle on X of degree g^3 and with curvature form a multiple of μ. In fact we can say more. It is a classical result (cf. for example [9], p. 31) that $\phi_Q^*(\Theta) = \mathcal{W} + g\cdot Q$. Hence we obtain the first statement of the next proposition.

Proposition 2.1. *We have a canonical isomorphism*

$$\sigma_Q : \phi_Q^*(O(\Theta)) \xrightarrow{\sim} O_X(\mathcal{W} + g\cdot Q)$$

of line bundles on X. When both sides are equipped with their Arakelov metrics, the isomorphism σ_Q has constant norm on X. This norm is independent of the choice of Q.

The proposition will be proven in the next section. Meanwhile, we observe that it leads quite quickly to a closed formula for G.

Definition 2.2. We define $S(X)$ to be the norm of σ_Q for any $Q \in X$.

In more concrete terms we have the following formula.

Corollary 2.3. *For any P, Q on X we have*

$$G(P,Q)^g \cdot \prod_{W\in\mathcal{W}} G(P,W) = S(X)\cdot \|\vartheta\|(gP - Q)\,,$$

where the Weierstrass points are counted with their weights.

It follows from this corollary that the function $\prod_{W\in\mathcal{W}} \|\vartheta\|(gP - W)$ does not vanish if P is not a Weierstrass point. Hence the following formula makes sense.

Theorem 2.4. *For any P, Q on X with P not a Weierstrass point we have*

$$G(P,Q)^g = S(X)^{1/g^2} \cdot \frac{\|\vartheta\|(gP-Q)}{\prod_{W\in\mathcal{W}} \|\vartheta\|(gP-W)^{1/g^3}} \,.$$

Here the Weierstrass points are counted with their weights.

Proof. This follows by applying the formula from Corollary 2.3 two times. First, take the (weighted) product over Q running through $\mathcal{W}$. This gives

$$\prod_{W\in\mathcal{W}} G(P,W)^{g^3} = S(X)^{g^3-g} \cdot \prod_{W\in\mathcal{W}} \|\vartheta\|(gP-W) \,.$$

Plug this in again in the formula from Corollary 2.3. This gives

$$G(P,Q)^g \cdot S(X)^{\frac{g^3-g}{g^3}} \cdot \prod_{W\in\mathcal{W}} \|\vartheta\|(gP-W)^{1/g^3} = S(X) \cdot \|\vartheta\|(gP-Q) \,,$$

and a little rewriting gives the result. □

Taking logarithms in Corollary 2.3 and then integrating against μ with respect to the variable P immediately gives the following explicit formula for $S(X)$.

Theorem 2.5. *For any fixed Q, the function $\log\|\vartheta\|(gP-Q)$ is integrable against μ, and the formula*

$$\log S(X) = -\int_X \log\|\vartheta\|(gP-Q) \cdot \mu(P)$$

holds.

The invariant $S(X)$ is readily calculated in the case $g = 1$.

Example 2.6. Suppose that $X = \mathbb{C}/\mathbb{Z} + \tau\mathbb{Z}$ with $\operatorname{Im}\tau > 0$ is an elliptic curve. The form μ is given by $\mu = \frac{i}{2}(dz \wedge d\bar{z})/\operatorname{Im}\tau$ and we have

$$\log S(X) = -\int_X \log\|\vartheta\| \cdot \mu \,.$$

A calculation (see [15], p. 45 or for a different approach [14], p. 250) yields

$$\log S(X) = -\log((\operatorname{Im}\tau)^{1/4}|\eta(\tau)|) \,,$$

where $\eta(\tau)$ is the usual Dedekind eta-function $\eta(q) = q^{1/24}\prod_{n=1}^{\infty}(1-q^n)$ in $q = \exp(2\pi i\tau)$.

In the case that $P = W$ is a Weierstrass point of X, the formula in Theorem 2.4 is still correct, provided that on the right hand side we take a limit for P approaching W. That this limit exists and that it indeed gives $G(W,Q)^g$ follows easily from the proof of Theorem 2.4.
We finish this section by discussing very shortly several other approaches to G that we know from the literature. First of all, it is quite natural to develop G in terms of the eigenvalues and eigenfunctions of a Laplacian associated to μ on X. This is the approach taken in [8], see especially Section 3 of that paper. Second, it is possible to express G in terms of abelian differentials of the second and third kind, see for example [15], Chapter II. Third, and this is

perhaps most close to our approach since it also involves theta functions quite explicitly, there is an integral formula for G derived by Bost, cf. [3], Proposition 1. This interesting result reads as follows: let ν be the curvature form of $O(\Theta)$ on $\mathrm{Pic}_{g-1}(X)$. Then there is an invariant $A(X)$ of X such that for every P, Q on X the formula

$$\log G(P,Q) = \frac{1}{g!} \int_{\Theta+P-Q} \log \|\vartheta\| \cdot \nu^{g-1} + A(X)$$

holds. It would be interesting to have results that relate $A(X)$ and $S(X)$ to each other in a natural, conceptual way.

3. Proof of Proposition 2.1

Proposition 2.1 follows directly from Lemmas 3.1 and 3.2 below. We will be dealing, among other things, with the line bundle $\wedge^g H^0(X,\Omega_X^1) \otimes_{\mathbb{C}} O_X$ on X. We equip this line bundle with the constant metric deriving from the hermitian inner product $(\omega,\eta) \mapsto \frac{i}{2}\int_X \omega \wedge \overline{\eta}$ on $H^0(X,\Omega_X^1)$ that we introduced in Section 1. From now on, this metric will be taken for granted and we shall also refer to it as an Arakelov metric.

Lemma 3.1. *There is a canonical isomorphism of line bundles*

$$\rho : \Omega_X^{\otimes g(g+1)/2} \otimes \left(\wedge^g H^0(X,\Omega_X^1) \otimes_{\mathbb{C}} O_X\right)^{\vee} \xrightarrow{\sim} O_X(\mathcal{W})$$

on X. When both sides are equipped with their Arakelov metrics, the norm of this isomorphism is constant on X.

Proof. The Wronskian differential $\tilde{\psi}$ formed on an arbitrary basis $\{\psi_1,\ldots,\psi_g\}$ of $H^0(X,\Omega_X^1)$ leads to a morphism of line bundles

$$\wedge^g H^0(X,\Omega_X^1) \otimes_{\mathbb{C}} O_X \longrightarrow \Omega_X^{g(g+1)/2}$$

by setting

$$\xi_1 \wedge \ldots \wedge \xi_g \mapsto \frac{\xi_1 \wedge \ldots \wedge \xi_g}{\psi_1 \wedge \ldots \wedge \psi_g} \cdot \tilde{\psi} \,.$$

This gives a canonical section in $\Omega_X^{\otimes g(g+1)/2} \otimes \left(\wedge^g H^0(X,\Omega_X^1) \otimes_{\mathbb{C}} O_X\right)^{\vee}$ whose divisor is $\mathcal{W}$ and thus we obtain the required isomorphism. The norm is constant on X because both sides have the same curvature form, and the divisors of their canonical sections are equal. □

Lemma 3.2. *Let Q be an arbitrary point of X. There is a canonical isomorphism of line bundles*

$$\phi_Q^*(O(\Theta)) \xrightarrow{\sim} \left(\Omega_X^{\otimes g(g+1)/2} \otimes \left(\wedge^g H^0(X,\Omega_X^1) \otimes_{\mathbb{C}} O_X\right)^{\vee}\right) \otimes O_X(g \cdot Q)$$

on X. When both sides are equipped with their Arakelov metrics, the norm of this isomorphism is constant on X and equal to $\exp(\delta(X)/8)$.

Proof. We are done once we prove that

$$\exp(\delta(X)/8) \cdot \|\vartheta\|(gP-Q) = \|\tilde{\omega}\|_{\mathrm{Ar}}(P) \cdot G(P,Q)^g$$

for arbitrary P,Q on X, where $\tilde{\omega}$ is the Wronskian differential formed out of an orthonormal basis $\{\omega_1,\ldots,\omega_g\}$ of $H^0(X,\Omega_X^1)$ and where the norm $\|\tilde{\omega}\|_{\mathrm{Ar}}$ of $\tilde{\omega}$ is taken in the line bundle $\Omega_X^{\otimes g(g+1)/2}$ equipped with its Arakelov metric. The required formula follows from the formula in Theorem 1.2, by a computation which is also performed in [14], p. 233 and which runs as follows. Let P be a point on X, and choose a local coordinate z about P. By definition of the Arakelov metric on Ω_X^1 we have that $\lim_{Q\to P}|z(Q)-z(P)|/G(Q,P) = \|dz\|_{\mathrm{Ar}}(P)$. Letting $P_1,\ldots,P_g$ approach P in Theorem 1.2 we obtain

$$\begin{aligned}
\lim_{P_l\to P}\frac{\|\det\omega_k(P_l)\|_{\mathrm{Ar}}}{\prod_{k<l}G(P_k,P_l)} &= \lim_{P_l\to P}\left\{\frac{\|\det\omega_k(P_l)\|_{\mathrm{Ar}}}{\prod_{k<l}|z(P_k)-z(P_l)|}\cdot\frac{\prod_{k<l}|z(P_k)-z(P_l)|}{\prod_{k<l}G(P_k,P_l)}\right\}\\
&= \left\{\lim_{P_l\to P}\frac{|\det\omega_k(P_l)|}{\prod_{k<l}|z(P_k)-z(P_l)|}\right\}\cdot\|dz\|_{\mathrm{Ar}}^{g+g(g-1)/2}(P)\\
&= |W_z(\omega)(P)|\cdot\|dz\|_{\mathrm{Ar}}^{g(g+1)/2}(P)\\
&= \|\tilde{\omega}\|_{\mathrm{Ar}}(P)\,.
\end{aligned}$$

The required formula is therefore just a limiting case of Theorem 1.2 where all P_k approach P. □

4. The Faltings delta-invariant

In the present section we will express the Faltings delta-invariant $\delta(X)$ in terms of $S(X)$ and a second invariant $T(X)$. The significance of our formula is that the constant $T(X)$ is in a sense "classical" and easy to calculate numerically.
To start our discussion, we observe that it follows from the previous sections that (multiples of) the divisor $\mathcal{W}$ of Weierstrass points appear as a divisor of a section of a line bundle in various different situations. We will take advantage of this fact and take combinations until we obtain an isomorphism of line bundles whose norm is easy to measure.
First of all, recall (this is Proposition 2.1) that we have for any Q on X a canonical isomorphism

$$\sigma_Q : \phi_Q^*(O(\Theta)) \xrightarrow{\sim} O_X(\mathcal{W}+g\cdot Q)\,.$$

Taking the (weighted) tensor product over the Weierstrass points of X, we obtain a canonical isomorphism

$$\sigma_{\mathcal{W}} : \bigotimes_{W\in\mathcal{W}}\phi_W^*(O(\Theta)) \xrightarrow{\sim} O_X(g^3\cdot\mathcal{W})\,.$$

Second, recall that by Lemma 3.1 we have a canonical isomorphism

$$\rho : \Omega_X^{\otimes g(g+1)/2}\otimes\left(\wedge^g H^0(X,\Omega_X^1)\otimes_{\mathbb{C}} O_X\right)^\vee \xrightarrow{\sim} O_X(\mathcal{W})\,.$$

Thirdly, taking a closer look at Lemma 3.2 we see that the proof in fact implies that we have on $X \times X$ a canonical isomorphism

$$\sigma : \Phi^*(O(\Theta)) \xrightarrow{\sim} O_{X\times X}(\mathcal{W} \cdot X + g \cdot \Delta_X)$$

where $\Phi : X \times X \to \mathrm{Pic}_{g-1}(X)$ is the map sending $(P,Q) \mapsto [gP - Q]$ and where again Δ_X is the diagonal on $X \times X$. Restricting σ to the diagonal, and using the adjunction isomorphism, we obtain a canonical isomorphism

$$\sigma|_\Delta : \Phi^*(O(\Theta))|_{\Delta_X} \otimes \Omega_X^{\otimes g} \xrightarrow{\sim} O_X(\mathcal{W}) \,.$$

Taking suitable combinations of $\sigma_{\mathcal{W}}, \rho$ and $\sigma|_\Delta$ we obtain

Proposition 4.1. *There is a canonical isomorphism of (fractional) line bundles*

$$\tau : \left(\Phi^*(O(\Theta))|_{\Delta_X} \otimes \Omega_X^{\otimes g}\right)^{\otimes(g+1)} \xrightarrow{\sim}$$
$$\left(\bigotimes_W \phi_W^*(O(\Theta))\right)^{\otimes(g-1)/g^3} \otimes \left(\Omega_X^{\otimes g(g+1)/2} \otimes \left(\wedge^g H^0(X,\Omega_X^1) \otimes_{\mathbb{C}} O_X\right)^\vee\right)^{\otimes 2}$$

on X.

Our results thus far imply that τ has a constant norm for the Arakelov metrics on both sides.

Definition 4.2. We define $T(X)$ to be the norm of τ on X.

The constant $T(X)$ admits the following concrete description using a local coordinate.

Proposition 4.3. *Let $P \in X$ not a Weierstrass point and let z be a local coordinate about P. Define $\|F_z\|(P)$ as*

$$\|F_z\|(P) = \lim_{Q \to P} \frac{\|\vartheta\|(gP-Q)}{|z(P)-z(Q)|^g} \,.$$

This limit exists and is non-zero. Further, let $\{\omega_1, \ldots, \omega_g\}$ be an orthonormal basis of $H^0(X,\Omega_X^1)$. Then the formula

$$T(X) = \|F_z\|(P)^{-(g+1)} \cdot \prod_{W \in \mathcal{W}} \|\vartheta\|(gP - W)^{(g-1)/g^3} \cdot |W_z(\omega)(P)|^2$$

holds, where $W_z(\omega)$ is the determinant of the Wronskian of $\{\omega_1, \ldots, \omega_g\}$ with respect to z.

In particular, the evaluation of $T(X)$ for a given X only involves the evaluation of certain classical functions at an arbitrary (non-Weierstrass) point of X.

Proof. Let F be the canonical section of $\Phi^*(O(\Theta))|_{\Delta_X} \otimes \Omega_X^{\otimes g}$ coming from the canonical section in $\Phi^*(O(\Theta))$ and the canonical generating section of $O_{X\times X}(\Delta_X)$ using the adjunction isomorphism. For its norm we have $\|F\| = \|F_z\| \cdot \|dz\|_{\mathrm{Ar}}^g$ in the local coordinate z. We see from the isomorphism $\sigma|_\Delta$

that $\|F\|(P)$ does not vanish if P is not a Weierstrass point. Next, the canonical section of $\bigotimes_{W\in\mathcal{W}}\phi_W^* O(\Theta)$ has norm $\prod_{W\in\mathcal{W}}\|\vartheta\|(gP-W)$ at P. Finally, the canonical section of $\Omega_X^{\otimes g(g+1)/2}\otimes\left(\wedge^g H^0(X,\Omega_X^1)\otimes_{\mathbb{C}} O_X\right)^\vee$ has norm $\|\tilde{\omega}\|_{\mathrm{Ar}}=|W_z(\omega)|\cdot\|dz\|_{\mathrm{Ar}}^{g(g+1)/2}$. The proposition follows by taking the appropriate combinations of these norms. □

Considering the norms of the three isomorphisms $\sigma_{\mathcal{W}},\rho$ and $\sigma|_\Delta$ one sees that they are directly expressible in terms of $\exp(\delta)$ and $S(X)$. Hence the same holds for the norm $T(X)$ of τ. Viewing things a little differently, we obtain a formula for $\exp(\delta)$ in terms of $S(X)$ and $T(X)$.

Theorem 4.4. *The formula*

$$\exp(\delta(X)/4)=S(X)^{-(g-1)/g^3}\cdot T(X)$$

holds.

Proof. The norm of $\sigma_{\mathcal{W}}$ is equal to $S(X)^{g^3-g}$. The norm of ρ is equal to $S(X)\exp(-\delta(X)/8)$ as becomes clear by decomposing again the isomorphism from Proposition 2.1, which has norm $S(X)$, into the isomorphisms from Lemmas 3.1 and 3.2. Lastly, the norm of $\sigma|_\Delta$ is equal to $S(X)$ since σ has this norm and the restriction to the diagonal using the adjunction isomorphism is an isometry. We obtain the required formula by just combining. □

We want to finish this section with a second formula for $T(X)$, involving only first order derivatives of the theta function. It is based on a function $\|J\|$ on $\mathrm{Sym}^g X$ introduced by Guàrdia in [11].

Let $\tau\in\mathcal{H}_g$ be a period matrix associated to a symplectic basis of $H_1(X,\mathbb{Z})$ and consider again the analytic jacobian $\mathrm{Jac}(X)=\mathbb{C}^g/\mathbb{Z}^g+\tau\mathbb{Z}^g$. For $w_1,\ldots,w_g\in\mathbb{C}^g$ we put

$$J(w_1,\ldots,w_g)=\det\left(\frac{\partial\vartheta}{\partial z_k}(w_l)\right)$$

and

$$\|J\|(w_1,\ldots,w_g)=(\det\operatorname{Im}\tau)^{\frac{g+2}{4}}\exp(-\pi\sum_{k=1}^g {}^t y_k(\operatorname{Im}\tau)^{-1}y_k)\cdot|J(w_1,\ldots,w_g)|\,.$$

Here in the latter formula $y_k=\operatorname{Im}w_k$ for $k=1,\ldots,g$. It can be checked that the function $\|J\|(w_1,\ldots,w_g)$ depends only on the classes in $\mathrm{Jac}(X)$ of the vectors w_k. Now let $P_1,\ldots,P_g$ be a set of g points on X. We take vectors $w_1,\ldots,w_g\in\mathbb{C}^g$ such that for all $k=1,\ldots,g$ the class $[w_k]\in\mathrm{Jac}(X)$ corresponds to $[\sum_{l=1, l\neq k}^g P_l]\in\mathrm{Pic}_{g-1}(X)$ under the Abel-Jacobi-Riemann correspondence $\mathrm{Pic}_{g-1}(X)\leftrightarrow\mathrm{Jac}(X)$. We then put $\|J\|(P_1,\ldots,P_g)=\|J\|(w_1,\ldots,w_g)$. One may check that this does not depend on the choice of the period matrix τ. The function $\|J\|$ has the following geometrical property: we have $\|J\|(P_1,\ldots,P_g)=0$ if and only if $P_1,\ldots,P_g$ are linearly dependent on the image of X under the canonical map $X\to\mathbb{P}(H^0(X,\Omega_X^1)^\vee)$. We refer to [11] for a proof of the following theorem.

Theorem 4.5. *Let $P_1, \ldots, P_g, Q$ be points on X with $P_1, \ldots, P_g$ pairwise distinct. Then the formula*

$$\|\vartheta\|(P_1 + \cdots + P_g - Q)^{g-1} = \exp(\delta(X)/8) \cdot \|J\|(P_1, \ldots, P_g) \cdot \frac{\prod_{k=1}^g G(P_k, Q)^{g-1}}{\prod_{k<l} G(P_k, P_l)}$$

holds.

A combination of Theorems 2.4, 4.4 and 4.5 yields the following formula for $T(X)$.

Proposition 4.6. *Let $P_1, \ldots, P_g, Q$ be generic points on X. Then the formula*

$$T(X) = \left(\frac{\|\vartheta\|(P_1 + \cdots + P_g - Q)}{\prod_{k=1}^g \|\vartheta\|(gP_k - Q)^{1/g}} \right)^{2g-2} \cdot \\ \cdot \left(\frac{\prod_{k \neq l} \|\vartheta\|(gP_k - P_l)^{1/g}}{\|J\|(P_1, \ldots, P_g)^2} \right) \cdot \prod_{W \in \mathcal{W}} \prod_{k=1}^{g} \|\vartheta\|(gP_k - W)^{(g-1)/g^4}$$

holds. Again the Weierstrass points are counted with their weights.

Let us make the invariant $T(X)$ explicit in the case that X is an elliptic curve. Writing $X = \mathbb{C}/\mathbb{Z} + \tau\mathbb{Z}$ with $\mathrm{Im}\,\tau > 0$ we obtain from either Proposition 4.3 or 4.6 that

$$T(X) = (\mathrm{Im}\,\tau)^{-3/2} \exp(\pi \mathrm{Im}\,\tau/2) \cdot |\frac{\partial\vartheta}{\partial z}(\frac{1+\tau}{2};\tau)|^{-2}.$$

By Jacobi's derivative formula (cf. [19], Chapter I, §13) we can rewrite this as

$$T(X) = (2\pi)^{-2} \cdot ((\mathrm{Im}\,\tau)^6 |\Delta(\tau)|)^{-1/4}$$

where Δ is the discriminant modular form $\Delta(q) = \eta(q)^{24} = q \prod_{n=1}^{\infty} (1-q^n)^{24}$ in $q = \exp(2\pi i \tau)$. Using Theorem 4.4 we obtain

$$\delta(X) = -\log((\mathrm{Im}\,\tau)^6 |\Delta(\tau)|) - 8\log(2\pi)$$

which is well-known, see [8], p. 417.
In [13] we obtain a generalisation of the above formula for $T(X)$ to the case where X is a hyperelliptic Riemann surface of genus $g \geq 2$. The result is expressed in terms of the discriminant modular form φ_g on the generalised Siegel upper half plane $\mathcal{H}_g$ as defined in [17], Section 3. This is a modular form on $\Gamma_g(2) = \{\gamma \in \mathrm{Sp}(2g, \mathbb{Z}) : \gamma \equiv I_{2g} \mod 2\}$ of weight $4r$, where $r = \binom{2g+1}{g+1}$, generalising the usual discriminant modular form Δ in genus 1.

Theorem 4.7. *Let X be a hyperelliptic Riemann surface of genus $g \geq 2$. Choose an ordering of the Weierstrass points on X and construct a canonical symplectic basis of $H_1(X, \mathbb{Z})$ starting with this ordering (cf.* [19], *Chapter IIIa, §5). Let $\tau \in \mathcal{H}_g$ be a period matrix of X associated to this canonical basis and put $\Delta_g(\tau) = 2^{-(4g+4)n} \cdot \varphi_g(\tau)$ where $n = \binom{2g}{g+1}$. Then $\Delta_g(\tau)$ is non-zero and the formula*

$$T(X) = (2\pi)^{-2g} \cdot ((\mathrm{Im}\,\tau)^{2r} |\Delta_g(\tau)|)^{-\frac{3g-1}{8ng}}$$

holds.

It is an intriguing question whether the invariant $T(X)$ admits of a simple description in terms of modular forms for a *general* Riemann surface X of genus g.

We finish this section by remarking that a closed formula of a quite different type can be given for δ using work of Bost [4] and Gillet-Soulé [10]. The point of view leading to this formula is the Riemannian manifold structure on X deriving from μ. Let ds^2 be the metric on X given in conformal coordinates by $ds^2 = 2h_{z\bar{z}}dzd\bar{z}$ with $h_{z\bar{z}} = \|dz\|_{\mathrm{Ar}}^{-2}$. Let $\det' \Delta_h$ be the zeta regularised determinant of the Laplace operator with respect to this metric, and let $\mathrm{vol}(X,h)$ be the volume of X. Then the formula

$$\delta(X) = c(g) - 6 \log \frac{\det \Delta_h'}{\mathrm{vol}(X,h)}$$

holds, where $c(g)$ is a constant depending only on g. It would be interesting to know whether the terms occurring in this formula can be naturally related to the constants $S(X)$ and $T(X)$ which are the subject of this paper.

5. Applications to intersection theory

In this section we discuss several applications of our results to Arakelov intersection theory. Let $p : \mathcal{X} \to B$ be an arithmetic surface over the spectrum B of the ring of integers of a number field K. For us this means that $\mathcal{X}$ is a regular scheme and that p is a proper and flat relative curve whose generic fiber is smooth and geometrically connected. We denote this generic fiber by $\mathcal{X}_K$. We assume that the reader is familiar with the basic notions and statements in the Arakelov intersection theory on $\mathcal{X}$, as explained in [2] or [8].

We let g be the genus of $\mathcal{X}_K$, and assume that it is positive. We fix a K-basis $\{\psi_1, \ldots, \psi_g\}$ of regular differential 1-forms on $\mathcal{X}_K$. Looking back at the discussion at the beginning of Section 2, which was purely algebraic, we note that a non-zero Wronskian differential $\tilde{\psi}$ can be formed out of this basis. Its divisor $\mathrm{div}\, \tilde{\psi}$ is an effective K-divisor on $\mathcal{X}_K$ and we have, completely analogous to Lemma 3.1, a canonical isomorphism

$$\Omega_{\mathcal{X}_K}^{\otimes g(g+1)/2} \otimes_{O_{\mathcal{X}_K}} \left(\wedge^g H^0(\mathcal{X}_K, \Omega^1_{\mathcal{X}_K}) \otimes_K O_{\mathcal{X}_K} \right)^\vee \xrightarrow{\sim} O_{\mathcal{X}_K}(\mathrm{div}\, \tilde{\psi})$$

of invertible sheaves on $\mathcal{X}_K$. We denote by $\mathcal{W}$ the Zariski closure of $\mathrm{div}\, \tilde{\psi}$ in $\mathcal{X}$. Let $\omega_{\mathcal{X}/B}$ be the relative dualising sheaf of $p : \mathcal{X} \to B$.

Lemma 5.1. *The above isomorphism extends to a canonical isomorphism*

$$\rho : \omega_{\mathcal{X}/B}^{\otimes g(g+1)/2} \otimes_{O_\mathcal{X}} \left(p^*(\det p_* \omega_{\mathcal{X}/B}) \right)^\vee \xrightarrow{\sim} O_\mathcal{X}(\mathcal{V} + \mathcal{W})$$

of invertible sheaves on $\mathcal{X}$, for some effective divisor $\mathcal{V}$ whose support is entirely contained in the fibers of p.

Proof. The idea for the proof is taken from [1], p. 1298, where an analogous result is proven for the function field case. We recall that $\tilde{\psi}$ is given in a local coordinate z by $W_z(\psi)(dz)^{\otimes g(g+1)/2}$ where

$$W_z(\psi) = \det \left(\frac{1}{(l-1)!} \frac{d^{l-1} f_k}{dz^{l-1}} \right)_{1 \leq k,l \leq g}$$

if the ψ_k are locally written as $\psi_k = f_k(z)dz$ for $k = 1, \ldots, g$. On $\mathcal{X}_K$ this gives rise to a morphism of invertible sheaves

$$\wedge^g H^0(\mathcal{X}_K, \Omega^1_{\mathcal{X}_K}) \otimes_K O_{\mathcal{X}_K} \longrightarrow \Omega^{\otimes g(g+1)/2}_{\mathcal{X}_K}$$

by setting

$$\xi_1 \wedge \ldots \wedge \xi_g \mapsto \frac{\xi_1 \wedge \ldots \wedge \xi_g}{\psi_1 \wedge \ldots \wedge \psi_g} \cdot \tilde{\psi}$$

(cf. the proof of Lemma 3.1). Now note that the construction of $\tilde{\psi}$ is valid for smooth proper curves over any base scheme. As a result, by modifying the basis $\{\psi_1, \ldots, \psi_g\}$ if necessary, the above morphism extends canonically at least over the open dense subscheme of $\mathcal{X}$ where p is smooth. Automatically it extends then further to give a canonical morphism $p^*(\det p_* \omega_{\mathcal{X}/B}) \to \omega^{\otimes g(g+1)/2}_{\mathcal{X}/B}$ on the whole of $\mathcal{X}$. Multiplying by $(p^*(\det p_* \omega_{\mathcal{X}/B}))^\vee$ we obtain from this a morphism

$$O_{\mathcal{X}} \longrightarrow \omega^{\otimes g(g+1)/2}_{\mathcal{X}/B} \otimes_{O_{\mathcal{X}}} \left(p^*(\det p_* \omega_{\mathcal{X}/B}) \right)^\vee .$$

The image of 1 is a section whose divisor is a divisor $\mathcal{V} + \mathcal{W}$ where $\mathcal{V}$ is effective and has support entirely contained in the fibers of p. This gives the lemma. $\square$

The divisor $\mathcal{V}$ is an invariant of the arithmetic surface $p : \mathcal{X} \to B$ and we shall use it without further mention in the sequel.

Example 5.2. In the case that $g = 1$, the morphism $p^* p_* \omega_{\mathcal{X}/B} \to \omega_{\mathcal{X}/B}$ in the above proof is just the natural morphism, as is readily checked. According to [16], Corollary 3.27, if $p : \mathcal{X} \to B$ is a minimal arithmetic surface, then the natural morphism $p^* p_* \omega_{\mathcal{X}/B} \to \omega_{\mathcal{X}/B}$ is in fact an isomorphism. Hence we find $\mathcal{V} = \varnothing$ in this case.

We want to translate the isomorphism ρ of Lemma 5.1 into an equality of Arakelov divisors on $\mathcal{X}$. For this we need a notation for the norm of ρ at the various complex embeddings of K.

Definition 5.3. Let X be a compact Riemann surface of positive genus. We denote by $R(X)$ the norm of the isomorphism ρ from Lemma 3.1.

It follows from our discussion so far that $R(X) = S(X) \cdot \exp(-\delta(X)/8)$. Now let's turn back to our arithmetic surface $p : \mathcal{X} \to B$. We recall from [2] [8] that both sides of the isomorphism ρ from Lemma 5.1 come equipped with a canonical structure of metrised invertible sheaf, and that to each non-zero rational section of such a sheaf we can associate its Arakelov divisor. For each complex embedding σ of K we denote by X_σ the compact Riemann surface $(\mathcal{X}_K \otimes_{K,\sigma} \mathbb{C})(\mathbb{C})$ obtained from base changing $\mathcal{X}_K$ to $\mathbb{C}$ along σ. We denote by F_σ the corresponding Arakelov fiber. The next proposition follows easily from Lemma 5.1 and from the fact that ρ has constant norm $R(X_\sigma)$ on X_σ.

Proposition 5.4. *We have an equality*

$$\frac{1}{2}g(g+1)\omega_{\mathcal{X}/B} = \mathcal{V} + \mathcal{W} + \sum_{\sigma} \log R(X_\sigma) \cdot F_\sigma + p^*(\det p_* \omega_{\mathcal{X}/B})$$

of Arakelov divisors on $\mathcal{X}$. Here the sum runs over the complex embeddings of K.

This proposition can be used to deduce some interesting formulas involving Arakelov intersection numbers.

Definition 5.5. We define a function R on the set of closed fibers of $p : \mathcal{X} \to B$ as follows. Let s be a closed point of B. If $g = 1$, we put $R(\mathcal{X}_s) = 0$. If $g \geq 2$, then we define $R(\mathcal{X}_s)$ by the equality $(2g-2) \cdot \log R(\mathcal{X}_s) = (\mathcal{V}_s, \omega_{\mathcal{X}/B}) \cdot \log \# k(s)$, where $(\mathcal{V}_s, \omega_{\mathcal{X}/B})$ is the usual intersection number of the divisors $\mathcal{V}$ and $\omega_{\mathcal{X}/B}$ above s, and where $k(s)$ is the residue field at s.

As the next proposition implies, the function R can be seen as an analogue of the previously defined R for compact Riemann surfaces. The quantity $\widehat{\deg} \det p_* \omega_{\mathcal{X}/B}$ is the usual Arakelov degree of the metrised invertible sheaf $\det p_* \omega_{\mathcal{X}/B}$ on B (i.e., $[K : \mathbb{Q}]$ times the Faltings height of $p : \mathcal{X} \to B$).

Proposition 5.6. *Assume that $p : \mathcal{X} \to B$ is a semi-stable arithmetic surface. Then for the self-intersection of the relative dualising sheaf we have a lower bound*

$$(\omega, \omega) \geq \frac{8(g-1)}{(2g-1)(g+1)} \cdot \left(\sum_s \log R(\mathcal{X}_s) + \sum_\sigma \log R(X_\sigma) + \widehat{\deg} \det p_* \omega_{\mathcal{X}/B} \right) .$$

Here the first sum runs over the closed points $s \in B$, and the second sum runs over the complex embeddings of K.

Proof. In the case $g = 1$, the lower bound is trivially satisfied since we have $(\omega_{\mathcal{X}/B}, \omega_{\mathcal{X}/B}) = 0$ in this case by [8], Theorem 7. So assume that $g \geq 2$. We take the equality from Proposition 5.1 and intersect the divisors on both sides with $\omega_{\mathcal{X}/B}$. This gives that $\frac{1}{2}g(g+1)(\omega_{\mathcal{X}/B}, \omega_{\mathcal{X}/B})$ can be written as

$$(\mathcal{W}, \omega_{\mathcal{X}/B}) + (2g-2) \cdot \left(\sum_s \log R(\mathcal{X}_s) + \sum_\sigma \log R(X_\sigma) + \widehat{\deg} \det p_* \omega_{\mathcal{X}/B} \right) .$$

For the term $(\mathcal{W}, \omega_{\mathcal{X}/B})$ we have by [8], Theorem 5 the lower bound

$$(\mathcal{W}, \omega_{\mathcal{X}/B}) \geq \frac{g^3 - g}{2g(2g-2)} (\omega_{\mathcal{X}/B}, \omega_{\mathcal{X}/B}) = \frac{1}{4}(g+1)(\omega_{\mathcal{X}/B}, \omega_{\mathcal{X}/B})$$

since the generic degree of $\mathcal{W}$ is $g^3 - g$. Using this in the first equality gives the required lower bound. □

We remark that for a semi-stable arithmetic surface $p : \mathcal{X} \to B$ the numbers $\log R(\mathcal{X}_s)$ are always non-negative. Lower bounds of a similar type for $(\omega_{\mathcal{X}/B}, \omega_{\mathcal{X}/B})$ can be found in [6]. The problem with the above proposition is that the right hand side may be negative, and then the lower bound becomes useless in view of the fundamental inequality $(\omega_{\mathcal{X}/B}, \omega_{\mathcal{X}/B}) \geq 0$ proved by Faltings [8]. However, for any fixed $g \geq 2$ the invariant $\log R(X)$ can become arbitrarily large, as the next proposition shows.

Proposition 5.7. *Let X_t be a holomorphic family of compact Riemann surfaces of genus $g \geq 2$ over the punctured disk, degenerating as $t \to 0$ to the union of two Riemann surfaces of positive genera g_1, g_2 with two points identified. Suppose that neither of these two points was a Weierstrass point. Then the formula*

$$\log R(X_t) = -\frac{g_1 g_2}{2g} \log |t| + O(1)$$

holds as $t \to 0$.

For a proof we refer to the author's thesis [12].
The next application we have in mind is a formula for the self-intersection of a point. In order to derive this formula it is convenient to use the machinery of the determinant of cohomology $\det Rp_*(\cdot)$ and the Deligne bracket $\langle \cdot, \cdot \rangle$, for which we refer to [7]. We will use that for any section $P : B \to \mathcal{X}$ and any invertible sheaf L on $\mathcal{X}$ we have canonical isomorphisms $\langle O_{\mathcal{X}}(P), L \rangle \xrightarrow{\sim} P^* L$ and $\langle P, \omega_{\mathcal{X}/B} \rangle \xrightarrow{\sim} \langle P, P \rangle^{\otimes -1}$. The latter is sometimes called the adjunction formula. Moreover, we have a Riemann-Roch theorem in the following form: for each invertible sheaf L on $\mathcal{X}$ there is a canonical isomorphism $(\det Rp_* L)^{\otimes 2} \xrightarrow{\sim} \langle L, L \otimes \omega_{\mathcal{X}/B}^{-1} \rangle \otimes (\det p_* \omega_{\mathcal{X}/B})^{\otimes 2}$.

Lemma 5.8. *Let P be a section of p, not a Weierstrass point on the generic fiber. Then we have a canonical isomorphism*

$$\upsilon : P^*(O_{\mathcal{X}}(\mathcal{V} + \mathcal{W}))^{\otimes 2} \xrightarrow{\sim} (\det Rp_* O_{\mathcal{X}}(gP))^{\otimes -2}$$

of line bundles on B. When restricted to the generic fiber, the left hand side gets identified with $O_{\mathrm{Spec} K}^{\otimes 2}$ and the right hand side gets identified with $H^0(\mathcal{X}_K, O_{\mathcal{X}_K}(gP))^{\otimes -2}$. The latter has a canonical trivialising section 1 and the isomorphism υ, when restricted to the generic fiber, sends the 1 of $O_{\mathrm{Spec} K}^{\otimes 2}$ to the 1 of $H^0(\mathcal{X}_K, O_{\mathcal{X}_K}(gP))^{\otimes -2}$.

Proof. The Riemann-Roch theorem applied to the invertible sheaf $O_{\mathcal{X}}(gP)$ gives a canonical isomorphism

$$(\det Rp_* O_{\mathcal{X}}(gP))^{\otimes 2} \xrightarrow{\sim} \langle O_{\mathcal{X}}(gP), O_{\mathcal{X}}(gP) \otimes \omega_{\mathcal{X}/B}^{-1} \rangle \otimes (\det p_* \omega_{\mathcal{X}/B})^{\otimes 2} .$$

By the adjunction formula, the right hand side can be canonically identified with $\langle P, P \rangle^{\otimes g(g+1)} \otimes (\det p_* \omega_{\mathcal{X}/B})^{\otimes 2}$, giving a canonical isomorphism

$$(\det Rp_* O_{\mathcal{X}}(gP))^{\otimes 2} \xrightarrow{\sim} \langle P, P \rangle^{\otimes g(g+1)} \otimes (\det p_* \omega_{\mathcal{X}/B})^{\otimes 2} .$$

On the other hand, pulling back the isomorphism ρ from Lemma 5.1 along P and using once more the adjunction formula we find a canonical isomorphism

$$\langle P, P \rangle^{\otimes -g(g+1)/2} \xrightarrow{\sim} \langle \mathcal{V} + \mathcal{W}, P \rangle \otimes \det p_* \omega_{\mathcal{X}/B}$$

and hence

$$\langle P,P\rangle^{\otimes g(g+1)} \xrightarrow{\sim} (P^* O_{\mathcal{X}}(\mathcal{V}+\mathcal{W}))^{\otimes -2} \otimes (\det p_* \omega_{\mathcal{X}/B})^{\otimes -2} .$$

The isomorphism v follows by combining these isomorphisms. Since P is not a Weierstrass point on the generic fiber, we have that $H^0(\mathcal{X}_K, gP)$ is 1-dimensional and hence is generated by its canonical section 1. The last statement of the lemma follows then by carefully spelling out all the isomorphisms. □

Proposition 5.9. *Let P be a section of p, not a Weierstrass point on the generic fiber. Then $R^1 p_* O_{\mathcal{X}}(gP)$ is a torsion module on B and the self-intersection $-\frac{1}{2}g(g+1)(P,P)$ is given by*

$$-\sum_\sigma \log G(P_\sigma, \mathcal{W}_\sigma) + \log \# R^1 p_* O_{\mathcal{X}}(gP) + \sum_\sigma \log R(X_\sigma) + \widehat{\deg} \det p_* \omega_{\mathcal{X}/B} .$$

Here σ runs through the complex embeddings of K.

Proof. That $R^1 p_* O_{\mathcal{X}}(gP)$ is a torsion module on B follows since we have $H^1(\mathcal{X}_K, gP) = 0$ on the generic fiber. As to the formula, we take the equality from Proposition 5.1 and intersect the divisors on both sides with P. By the Arakelov adjunction formula $(\omega, P) = -(P,P)$ we obtain

$$-\frac{1}{2}g(g+1)(P,P) = (\mathcal{V}+\mathcal{W}, P) + \sum_\sigma \log R(X_\sigma) + \widehat{\deg} \det p_* \omega_{\mathcal{X}/B} .$$

It remains therefore to see that $(\mathcal{V}+\mathcal{W}, P)_{\mathrm{fin}} = \log \# R^1 p_* O_{\mathcal{X}}(gP)$. For this we invoke Lemma 5.8. It follows from the description of v on the generic fiber that in fact v is the natural isomorphism over the open dense subscheme of B where P does not meet $\mathcal{V}+\mathcal{W}$. Now for any closed point s of B denote by e_s the length at s of $R^1 p_* O_{\mathcal{X}}(gP)$. Then if we let $D = \sum_s e_s \cdot s$, the invertible sheaf $\det R^1 p_* O_{\mathcal{X}}(gP)$ gets identified with $O_B(D)$ and, since $\det R^0 p_* O_{\mathcal{X}}(gP)$ is trivialised by the section 1, the determinant of cohomology $\det R p_* O_{\mathcal{X}}(gP)$ gets identified with $O_B(-D)$. By Lemma 5.8, for any closed point s the length e_s coincides with the intersection multiplicity of P and $\mathcal{V}+\mathcal{W}$ at s and consequently $(\mathcal{V}+\mathcal{W}, P)_{\mathrm{fin}} = \sum_s e_s \log \# k(s) = \log \# R^1 p_* O_{\mathcal{X}}(gP)$. □

6. A numerical example

In this section we use the results of Sections 2 and 4 to calculate the Faltings height and the self-intersection of the relative dualising sheaf of a certain hyperelliptic curve of genus 3 defined over the rationals. We start with two theoretical results, both of which can be proved by methods similar to those used in [5], Section 3.

Let K be a number field, and let O_K be its ring of integers. For a non-zero element $a \in O_K$ and a prime ideal $\wp$ of O_K we denote by $v_\wp(a)$ the exponent of $\wp$ in the prime ideal decomposition of $a \cdot O_K$. Let $f \in O_K[x]$ be a monic polynomial of degree 5 with $f(0)$ and $f(1)$ units in O_K and put $g(x) = x(x-1) + 4f(x)$.

Proposition 6.1. *Suppose that the discriminant Δ of g is non-zero, that we have $v_\wp(\Delta) = 0$ or 1 for each prime ideal $\wp$ of residue characteristic $\neq 2$, and that g mod $\wp$ has a unique multiple root of multiplicity 2 for prime ideals $\wp$ with residue characteristic $\neq 2$ and with $v_\wp(\Delta) = 1$. Then the equation*

$$C_f \, : \, y^2 = x(x-1)f(x)$$

defines a hyperelliptic curve of genus 3 over K. It extends to a semi-stable arithmetic surface $p : \mathcal{X} \to B = \mathrm{Spec}(O_K)$. We have that $\mathcal{X}$ has bad reduction at $\wp$ if and only if $\wp$ has residue characteristic $\neq 2$ and $v_\wp(\Delta) = 1$. For such $\wp$, the fiber at $\wp$ is an irreducible curve with a single double point. The differentials $dx/y, xdx/y, x^2dx/y$ form a basis of the free O_B-module $p_*\omega_{\mathcal{X}/B}$. The points W_0, W_1 on C_f given by $x = 0$ and $x = 1$ extend to disjoint sections of p.

As to the Faltings height and the self-intersection of the relative dualising sheaf of C_f we have the following. For a complex embedding σ of K we denote by $X_{f,\sigma}$ the compact Riemann surface $(C_f \otimes_{K,\sigma} \mathbb{C})(\mathbb{C})$ obtained by base changing C_f to $\mathbb{C}$ along σ. For each σ, we choose a symplectic basis of $H_1(X_{f,\sigma}, \mathbb{Z})$ and form the period matrix $\Omega_\sigma = (\Omega_{1,\sigma}|\Omega_{2\sigma})$ for $dx/y, xdx/y$ and x^2dx/y on this basis. We further put $\tau_\sigma = \Omega_{1\sigma}^{-1}\Omega_{2\sigma}$.

Proposition 6.2. *The degree of* $\det p_*\omega_{\mathcal{X}/B}$ *satisfies*

$$\widehat{\deg} \det p_*\omega_{\mathcal{X}/B} = -\frac{1}{2} \sum_\sigma \log \left(|\det \Omega_{1\sigma}|^2 (\det \mathrm{Im}\, \tau_\sigma) \right) .$$

For the self-intersection of the relative dualising sheaf we have the formula

$$(\omega_{\mathcal{X}/B}, \omega_{\mathcal{X}/B}) = 24 \sum_\sigma \log G_\sigma(W_0, W_1) \, ,$$

where G_σ denotes the Arakelov-Green function on $X_{f,\sigma}$.

We apply these propositions to a concrete example. We choose $K = \mathbb{Q}$ and $f(x) = x^5 + 6x^4 + 4x^3 - 6x^2 - 5x - 1$. It can be checked that $g(x) = x(x-1) + 4f(x)$ satisfies the conditions of Proposition 6.1. The corresponding hyperelliptic curve C_f has bad reduction at the primes $p = 37, p = 701$ and $p = 14717$. Let X_f be the compact Riemann surface obtained from base changing C_f to the complex numbers. We choose an ordering of the Weierstrass points of X and as in [19], Chapter IIIa, §5 this gives us a canonical way to construct a symplectic basis for $H_1(X_f, \mathbb{Z})$. We have computed the periods with respect to this basis of the differentials $dx/y, xdx/y$ and x^2dx/y. Using Proposition 6.2 we easily obtain

$$\widehat{\deg} \det p_*\omega_{\mathcal{X}/B} = -1.280295247656532068\ldots$$

which is the Faltings height of C_f. Next we take a look at the self-intersection of the relative dualising sheaf. According to Proposition 6.2 we need to calculate $G(W_0, W_1)$. We apply Theorem 2.4 where we carefully take a limit for P approaching W_0. Using theory as developed for example in [19], Chapter IIIa it is possible to make the Abel-Jacobi-Riemann correspondence $\mathrm{Pic}_2(X_f) \leftrightarrow$

$\mathrm{Jac}(X_f)$ completely explicit. This makes it easy to carry out the theta function evaluations that are needed to compute $G(W_0, W_1)$. The calculation of $S(X_f)$ is, however, considerably harder. We recall that in our case $S(X_f)$ is given by

$$\log S(X_f) = -\int_{X_f} \log \|\vartheta\|(3P - Q) \cdot \mu(P)$$

where μ is the Arakelov metric and where Q is any point on X_f. We want to express μ in terms of the coordinates x, y but then it immediately becomes clear that the integrand will diverge at the Weierstrass point at infinity. However, by taking logarithms in Theorem 2.4 and integrating against $\mu(Q)$ we find the alternative formula

$$\log S(X_f) = -9 \int_X \log \|\vartheta\|(3P - Q) \cdot \mu(Q) + \frac{1}{3} \sum_{W \in \mathcal{W}} \log \|\vartheta\|(3P - W)\,,$$

valid for any non-Weierstrass point P on X, in which the integrand behaves better. In fact, the integrand now only has a singularity at $Q = P$. Let $\Omega = (\Omega_1 | \Omega_2)$ be the period matrix of X_f referred to above and put $\tau = \Omega_1^{-1}\Omega_2$. Let (μ_{kl}) be the matrix given by $(\mu_{kl}) = \left(\overline{\Omega}_1 (\mathrm{Im}\,\tau)^t \Omega_1\right)^{-1}$. An application of Riemann's bilinear relations yields $\mu = \frac{i}{6} \sum \mu_{kl}\, \psi_k \wedge \overline{\psi_l}$ with $\psi_1 = dx/y, \psi_2 = xdx/y$ and $\psi_3 = x^2 dx/y$. Writing $x = u + iv$ with $u, v \in \mathbb{R}$ we can rewrite this as the real 2-form

$$\begin{aligned} \mu = &\frac{1}{3} \left(\mu_{11} + 2\mu_{12} u + 2\mu_{13}(u^2 - v^2) + \mu_{22}(u^2 + v^2) \right. \\ &\left. + 2\mu_{23} u (u^2 + v^2) + \mu_{33}(u^2 + v^2)^2 \right) \cdot \frac{du dv}{|h(u + iv)|}\,, \end{aligned}$$

where $h(x) = x(x-1)g(x)$. Using a computer algebra package, we have evaluated the integral. This is a slow process, because one has to take care of the logarithmic singularity. On the other hand, it is possible to check the answers by trying various choices of P. We found that within reasonable time limits we can only reach an accuracy within $\pm$ 0.005. The end result is

$$\log S(X_f) = 17.57\ldots$$

Using this we find the approximation

$$G(W_0, W_1) = 2.33\ldots$$

and finally

$$(\omega_{\mathcal{X}/B}, \omega_{\mathcal{X}/B}) = 20.32\ldots$$

It is almost no extra effort to compute also the delta-invariant of X_f. Using Theorem 4.7 we obtain, first of all,

$$\log T(X_f) = -4.44361200473681284\ldots$$

With Theorem 4.4 and our value above for $\log S(X_f)$ we get as a result

$$\delta(X_f) = -33.40\ldots$$

The reader may check that the Noether formula [18] is verified by our numerical results.

ACKNOWLEDGMENTS. The author wishes to thank Gerard van der Geer for his encouragement and helpful remarks.

References

[1] S. Y. Arakelov, *Families of algebraic curves with fixed degeneracies*, Math. USSR Izvestija 5 (1971), 1277–1302.

[2] S. Y. Arakelov, *An intersection theory for divisors on an arithmetic surface*, Math. USSR Izvestija 8 (1974), 1167–1180.

[3] J.-B. Bost, *Fonctions de Green-Arakelov, fonctions thêta et courbes de genre 2*, C.R. Acad. Sci. Paris Ser. I 305 (1987), 643–646.

[4] J.-B. Bost, *Conformal and holomorphic anomalies on Riemann surfaces and determinant line bundles.* In: M. Mebkhout, R. Seneor (eds.), 8th International Congress on Mathematical Physics, World Scientific 1987.

[5] J.-B. Bost, J.-F. Mestre, L. Moret-Bailly, *Sur le calcul explicite des "classes de Chern" des surfaces arithmétiques de genre 2.* In: Séminaire sur les pinceaux de courbes elliptiques, Astérisque 183 (1990), 69–105.

[6] J.-F. Burnol, *Weierstrass points on arithmetic surfaces*, Inv. Math. 107 (1992), 421–432.

[7] P. Deligne, *Le déterminant de la cohomologie.* In: Contemporary Mathematics vol. 67, American Mathematical Society (1987), 93–177.

[8] G. Faltings, *Calculus on arithmetic surfaces*, Ann. of Math. 119 (1984), 387–424.

[9] J.D. Fay, *Theta functions on Riemann surfaces.* Lect. Notes in Math. vol. 352, Springer-Verlag 1973.

[10] H. Gillet, C. Soulé, *An arithmetic Riemann-Roch theorem*, Inv. Math. 110 (1992), 473–543.

[11] J. Guàrdia, *Analytic invariants in Arakelov theory for curves*, C.R. Acad. Sci. Paris Ser. I 329 (1999), 41–46.

[12] R. de Jong, *Explicit Arakelov geometry.* Thesis University of Amsterdam, 2004.

[13] R. de Jong, *Faltings' delta-invariant of a hyperelliptic Riemann surface.* In: G. van der Geer, B. Moonen, R. Schoof (eds.), Proceedings of the Texel Conference "The analogy between number fields and function fields", Birkhäuser Verlag 2005.

[14] J. Jorgenson, *Asymptotic behavior of Faltings's delta function*, Duke Math. J. 61 (1990), 1, 303–328.

[15] S. Lang, *Introduction to Arakelov theory.* Springer-Verlag 1988.

[16] Q. Liu, *Algebraic Geometry and Arithmetic Curves.* Oxford Graduate Texts in Mathematics 6, Oxford Science Publications 2002.

[17] P. Lockhart, *On the discriminant of a hyperelliptic curve*, Trans. Amer. Math. Soc. 342 (1994), 2, 729–752.

[18] L. Moret-Bailly, *La formule de Noether pour les surfaces arithmétiques*, Inv. Math. 98 (1989), 491–498.

[19] D. Mumford, *Tata Lectures on Theta I,II.* Progress in Mathematics vol. 28, 43, Birkhäuser Verlag 1984.

[20] R. Wentworth, *The asymptotics of the Arakelov-Green's function and Faltings' delta invariant*, Commun. Math. Phys. 137 (1991), 427–459.

Robin de Jong
Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden
The Netherlands
rdejong@math.leidenuniv.nl

The Projected Single-Particle Dirac Operator for Coulombic Potentials

D. H. Jakubassa-Amundsen

Received: January 22, 2003
Revised: March 24, 2004

Communicated by Ulf Rehmann

Abstract. A sequence of unitary transformations is applied to the one-electron Dirac operator in an external Coulomb potential such that the resulting operator is of the form $\Lambda_+ A\,\Lambda_+ + \Lambda_- A\,\Lambda_-$ to any given order in the potential strength, where Λ_+ and Λ_- project onto the positive and negative spectral subspaces of the free Dirac operator. To first order, $\Lambda_+ A\,\Lambda_+$ coincides with the Brown-Ravenhall operator. Moreover, there exists a simple relation to the Dirac operator transformed with the help of the Foldy-Wouthuysen technique. By defining the transformation operators as integral operators in Fourier space it is shown that they are well-defined and that the resulting transformed operator is p-form bounded. In the case of a modified Coulomb potential, $V = -\gamma x^{-1+\epsilon}$, $\epsilon > 0$, one can even prove subordinacy of the n-th order term in γ with respect to the $n-1$st order term for all $n > 1$, as well as their p-form boundedness with form bound less than one.

2000 Mathematics Subject Classification: 81Q10, 81Q15, 81V45

1 Introduction

Consider a relativistic electron in the Coulomb field V, described by the Dirac operator (in relativistic units, $\hbar = c = 1$)

$$H \;=\; D_0 + V, \qquad\qquad D_0 := -i\boldsymbol{\alpha}\,\partial/\partial\mathbf{x} + \beta m, \quad V(x) := -\frac{\gamma}{x} \qquad (1.1)$$

where D_0 is the free Dirac operator defined in the Hilbert space $L_2(\mathbb{R}^3)\otimes\mathbb{C}^4$. D_0 is self-adjoint on the Sobolev space $H_1(\mathbb{R}^3)\otimes\mathbb{C}^4$ (with $H_\sigma(\mathbb{R}^3) := \{\varphi \in L_2(\mathbb{R}^3) : \int d\mathbf{p}\,|\hat{\varphi}(\mathbf{p})|^2\,(1+p^2)^\sigma < \infty\}$), and its form domain is $H_{1/2}(\mathbb{R}^3)\otimes\mathbb{C}^4$. The potential strength of V is $\gamma := Ze^2$, Z is the nuclear charge number,

$e^2 = (137.04)^{-1}$ the fine structure constant, $\boldsymbol{\alpha}$ and β the Dirac matrices and $x := |\mathbf{x}|$ [14]. A hat on a function denotes its Fourier transform.
It is well-known that H is not bounded from below. As long as pair creation is neglected, the conventional way to circumvent this deficiency is the introduction of the semibounded operator P_+HP_+ where P_+ projects onto the positive spectral subspace of H. As a first approximation, Brown and Ravenhall [1] introduced the operator

$$B := \Lambda_+ H \Lambda_+, \qquad \Lambda_\pm := \frac{1}{2}\left(1 \pm \frac{D_0}{|D_0|}\right) \tag{1.2}$$

with Λ_+ projecting onto the positive spectral subspace of the free Dirac operator D_0, and $|D_0| = \sqrt{D_0^2}$ is the free energy. In momentum space one has

$$\tilde{D}_0(\mathbf{p}) := \left(\frac{D_0}{|D_0|}\right)(\mathbf{p}) = \frac{\boldsymbol{\alpha}\mathbf{p} + \beta m}{E_p}, \qquad E_p := \sqrt{p^2 + m^2} \tag{1.3}$$

with the electron mass m. By construction, the Brown-Ravenhall operator B is of first order in the potential V and has been shown to be bounded from below for subcritical potential strength γ [5].
An alternative way to derive a semibounded operator from H has been suggested by Douglas and Kroll [4], using the Foldy-Wouthuysen transformation technique [6]. The decoupling of the positive and negative spectral subspaces of H to order n in V is achieved by means of $n+1$ successive unitary transformations U_j', $j = 0, 1, ..., n$

$$(U_n' \cdots U_1' \cdot U_0')\; H\; (U_n' \cdots U_1' \cdot U_0')^{-1} \;=:\; H_n' \;+\; R_{n+1} \tag{1.4}$$

which cast the tranformed operator into a block-diagonal contribution H_n' plus an error term R_{n+1} with potential strength given by the $n+1$st power of γ. U_0' is the free Foldy-Wouthuysen transformation which block-diagonalises D_0 exactly [14],

$$U_0' := A\left(1 \;+\; \beta\,\frac{\boldsymbol{\alpha}\mathbf{p}}{E_p + m}\right), \qquad A := \left(\frac{E_p + m}{2E_p}\right)^{\frac{1}{2}}, \tag{1.5}$$

and for U_j', Douglas and Kroll [4] use

$$U_j' \;=\; (1 + W_j^2)^{\frac{1}{2}} \;+\; W_j, \qquad j = 1, ..., n \tag{1.6}$$

with antisymmetric operators W_j. It should be noted that the choice of U_j' is not unique, and neither is the resulting operator H_n' from (1.4) for $n > 4$ as has been shown by Wolf, Reiher, and Hess [15]. Applying the free Foldy-Wouthuysen transformation (1.5) one obtains [4, 9]

$$U_0' H U_0'^{-1} \;=\; \beta E_p \;+\; \mathcal{E}_1 \;+\; \mathcal{O}_1$$

$$\mathcal{E}_1 := A\left(V + \frac{\boldsymbol{\alpha}\mathbf{p}}{E_p + m} V \frac{\boldsymbol{\alpha}\mathbf{p}}{E_p + m}\right) A, \quad \mathcal{O}_1 := \beta A\left(\frac{\boldsymbol{\alpha}\mathbf{p}}{E_p + m} V - V \frac{\boldsymbol{\alpha}\mathbf{p}}{E_p + m}\right) A \tag{1.7}$$

where the transformed potential has been split into an even term $\mathcal{E}_1$ (commuting with β) and an odd term $\mathcal{O}_1$ (anticommuting with β, since $\alpha_k\beta = -\beta\alpha_k$, $k = 1, 2, 3$). For an exponential unitary transformation,

$$U_j' := e^{-iS_j}, \qquad j = 1, ..., n \tag{1.8}$$

with a symmetric operator S_j, the next transformation gives in agreement with [9]

$$e^{-iS_1}\, U_0' H U_0'^{-1}\, e^{iS_1} = \beta E_p + \mathcal{E}_1 + \mathcal{O}_1 + i[\beta E_p, S_1] + i[\mathcal{E}_1 + \mathcal{O}_1, S_1]$$
$$-\frac{1}{2}\,[\,[\beta E_p, S_1],\, S_1] + R_3. \tag{1.9}$$

S_1 is defined from the requirement that $\mathcal{O}_1$ is eliminated,

$$i\,[\beta E_p, S_j] = -\mathcal{O}_j, \qquad j = 1 \tag{1.10}$$

hence S_1 is odd and of first order in the potential like $\mathcal{O}_1$. After each transformation U_j', the $j+1$st order term in γ of R_{j+1} is decomposed into even ($\mathcal{E}_{j+1}$) and odd ($\mathcal{O}_{j+1}$) contributions, and the successive transformation $U_{j+1}' = e^{-iS_{j+1}}$ is chosen to eliminate $\mathcal{O}_{j+1}$, which is achieved by the condition (1.10) for the $j > 1$ under consideration. With this procedure one arrives at the even (and hence block-diagonal) operator

$$H_n' = \beta E_p + \mathcal{E}_1 + ... + \mathcal{E}_n. \tag{1.11}$$

The physical quantity of interest is the expectation value of the transformed Dirac operator. For the Brown-Ravenhall operator, consider the expectation value formed with 4-spinors φ in the positive spectral subspace of D_0 which in momentum space can be expressed in terms of Pauli spinors $u \in H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^2$,

$$\hat{\varphi}(\mathbf{p}) = \frac{1}{\sqrt{2E_p\,(E_p + m)}} \begin{pmatrix} (E_p + m)\ \hat{u}(\mathbf{p}) \\ \mathbf{p}\boldsymbol{\sigma}\ \hat{u}(\mathbf{p}) \end{pmatrix} \tag{1.12}$$

where $\boldsymbol{\sigma}$ is the vector of the three Pauli matrices. Then, an operator b_m acting on $H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^2$ may be defined by [5]

$$(\varphi, B\varphi) =: (u, b_m u). \tag{1.13}$$

On the other hand, in case of the Douglas-Kroll transformed operator H_n', its upper block corresponds to the particle states (having positive energy) and therefore the expectation value has to be formed with the four-spinor $\psi := \binom{u}{0}$ with $u \in H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^2$ as above.

For the first-order term, H_1', it is easy to show [2] that its expectation value agrees with the expectation value of the Brown-Ravenhall operator, i.e.

$$(u, b_m u) \;=\; \left(\binom{u}{0}, H_1' \binom{u}{0} \right), \qquad\qquad u \in H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^2, \tag{1.14}$$

such that H_n' may be considered as the natural continuation of B to higher order in V. While H_n' is known explicitly up to $n = 5$ [15], the spectral properties, in particular the boundedness from below, have only been investigated for the Jansen-Hess operator (i.e. $n = 2$) [2, 8].
The aim of this work is to prove two theorems.

Theorem 1.1. *Let $H = D_0 + V$ be the one-particle Dirac operator acting on $\mathcal{S}(\mathbb{R}^3) \otimes \mathbb{C}^4$ with $\mathcal{S}$ the Schwartz space of smooth strongly localised functions. Let γ be the strength of the Coulomb potential V and $p := |\mathbf{p}|$. Then there exists a sequence of unitary transformations $U_k = e^{iB_k}$, $k = 1, ..., n$, such that the transformed Dirac operator can be written in the following way*

$$(U_1 \cdots U_n)^{-1} H\, U_1 \cdots U_n \;=:\; H^{(n)} \;+\; R^{(n+1)}$$

$$H^{(n)} := \;\Lambda_+ \left(\sum_{k=0}^{n} H_k \right) \Lambda_+ \;+\; \Lambda_- \left(\sum_{k=0}^{n} H_k \right) \Lambda_-. \tag{1.15}$$

Here, Λ_+ projects onto the positive spectral subspace of D_0, $\Lambda_- = 1 - \Lambda_+$, and H_k is a p-form bounded operator, its form bound being proportional to γ^k, $k = 1, ..., n$. The remainder $R^{(n+1)}$ which still couples the spectral subspaces of D_0 is p-form bounded with form bound $O(\gamma^{n+1})$ when γ tends to zero. The operators B_k are symmetric and bounded, extending to self-adjoint operators on $L_2(\mathbb{R}^3) \otimes \mathbb{C}^4$.

An operator H_k with the properties stated in the theorem is said to be *of order* γ^k.
It is shown below that $H_0 := D_0$ and $H_1 := V$ such that to first order, the Brown-Ravenhall operator B is recovered. (1.15) implies that the transformed Dirac operator can be expressed in terms of projectors to arbitrary order in the potential strength. Similar transformation schemes are known for bounded operators on lattices, see e.g. [3] and [13].
The next theorem states the unitary equivalence of the transformed Dirac operators obtained with either the transformation scheme from Theorem 1.1 or the Douglas-Kroll transformation scheme (1.8) - (1.11).

Theorem 1.2. *Let $\varphi \in \Lambda_+(\mathcal{S}(\mathbb{R}^3) \otimes \mathbb{C}^4)$ be a 4-spinor in the positive spectral subspace of D_0, which defines a Pauli spinor u by means of (1.12). Let H_n' be the Douglas-Kroll transformed Dirac operator to n-th order in the potential strength, obtained with exponential unitary operators U_j'. Then to any order n, its expectation value agrees with the one of the transformed Dirac operator* *from (1.15),*

$$(\varphi, H^{(n)}\,\varphi) \;=\; \left(\varphi, \sum_{k=0}^{n} H_k\,\varphi\right) \;=\; \left(\begin{pmatrix} u \\ 0 \end{pmatrix}, H'_n \begin{pmatrix} u \\ 0 \end{pmatrix}\right), \qquad n = 1, 2, \ldots\,. \tag{1.16}$$

This extends the first-order equation (1.14) to arbitrary order n. Actually (1.16) holds also for other types of unitary operators, provided the same type is used in all transformations U_k and U'_k, $k \geq 1$.

2 Proof of Theorem 1.1

2.1 Derivation of unitary transformations

The sequence of unitary operators $U_k = e^{iB_k}$ is constructed with the help of an iteration scheme. Following Sobolev [13] we consider U_k as an element of the group $U_k(t) = e^{iB_k t}$, $t \in \mathbb{R}$.

Let A be an arbitrary t-independent operator. The derivative of the transformed operator is given by

$$\frac{d}{dt}A(t) := \frac{d}{dt}\left(e^{-iB_k t}\, A\, e^{iB_k t}\right) \;=\; i\, U_k(-t)\; [A, B_k]\; U_k(t) \tag{2.1}$$

where the commutator $[A, B_k] := AB_k - B_k A$. This equation is easily integrated, noting that $A(0) = A$,

$$A(t) \;=\; U_k(-t)\, A\, U_k(t) \;=\; A \;+\; i \int_0^t d\tau\; U_k(-\tau)\; [A, B_k]\; U_k(\tau). \tag{2.2}$$

Iterating once, i.e. replacing A by the operator $[A, B_k]$ in (2.2) and inserting the resulting equation into the r.h.s. of (2.2), one obtains for $t = 1$

$$A(1) \;=\; A \;+\; i[A, B_k] \;+\; i^2 \int_0^1 d\tau \int_0^\tau dt'\, U_k(-t')\, [[A, B_k], B_k]\, U_k(t'). \tag{2.3}$$

After n iterations the following representation of A_1 is obtained,

$$A(1) \;=\; A + i[A, B_k] \;+\; \frac{1}{2!} i^2\; [[A, B_k], B_k] \;+ \ldots +\; \frac{1}{n!} i^n\; [[\ldots[A, B_k], \ldots, B_k] \;+\; R \tag{2.4}$$

where the n-th term contains n commutators with B_k, and the remainder R is an $(n+1)$-fold integral.

Let us apply this scheme inductively to the Dirac operator $H = D_0 + V$. Assume that to order $n - 1$ the transformation has been achieved with a resulting operator of the form given in Theorem 1.1,

$$(U_1 \cdots U_{n-1})^{-1}\; H\; U_1 \cdots U_{n-1} \;=\; H^{(n-1)} \;+\; H_n \;+\; \tilde{R}^{(n+1)} \tag{2.5}$$

where H_n and $\tilde{R}^{(n+1)}$ are respectively of order γ^n and γ^{n+1}, and still couple the spectral subspaces. Decompose H_n into

$$H_n \;=\; V_n \;+\; W_n, \qquad\qquad V_n := \;\Lambda_+ H_n \Lambda_+ \;+\; \Lambda_- H_n \Lambda_-$$
$$W_n := \;\Lambda_+ H_n \Lambda_- \;+\; \Lambda_- H_n \Lambda_+ \,. \tag{2.6}$$

The next transformation, $U_n = e^{iB_n}$, aims at eliminating the term W_n which, in contrast to V_n, couples the spectral subspaces. This condition will fix B_n. We note that from (2.4), the transformation reproduces the operator itself, such that the term $H^{(n-1)}$, already in the desired form, is preserved. From this it follows that $H^{(n-1)}$ contains the zero-order term $\Lambda_+ D_0 \Lambda_+ + \Lambda_- D_0 \Lambda_- \;= D_0$ (note that $\Lambda_\pm$ commutes with D_0 and $\Lambda_+^2 + \Lambda_-^2 = 1$).
We obtain

$$U_n^{-1}\,(H^{(n-1)} \;+\; H_n)\,U_n \;=\; H^{(n-1)} \;+\; V_n \;+\; W_n \;+\; i[D_0, B_n] \tag{2.7}$$
$$+\; i\,[(\Lambda_+ \sum_{k=1}^{n-1} H_k \Lambda_+ \;+\; \Lambda_- \sum_{k=1}^{n-1} H_k \Lambda_-),\; B_n] \;+\; \tilde{R},$$

where $\tilde{R}$ collects the terms containing at least two commutators with B_n. B_n is determined from the requirement

$$W_n \;+\; i\,[D_0, B_n] \;=\; 0. \tag{2.8}$$

Since W_n is of order γ^n, B_n is proportional to γ^n (the boundedness of B_n is shown later). Moreover, the commutators of the type $[(\Lambda_+ H_k \Lambda_+ + \Lambda_- H_k \Lambda_-), B_n]$ are of order γ^{n+k} with $k \geq 1$, and $\tilde{R}$ is of order γ^{2n}. Hence, these terms are disregarded (together with the remainder $\tilde{R}^{(n+1)}$ from (2.5)) in constructing the transformed operator to order n,

$$H^{(n)} \;=\; H^{(n-1)} \;+\; V_n \;=\; D_0 \;+\; V_1 \;+\; V_2 \;+ \ldots +\; V_n. \tag{2.9}$$

Particularly interesting are the cases $n = 1$ and $n = 2$. For $n = 1$, we have

$$H^{(1)} \;=\; D_0 \;+\; V_1 \;=\; \Lambda_+\,(D_0 + V)\,\Lambda_+ \;+\; \Lambda_-\,(D_0 + V)\,\Lambda_-. \tag{2.10}$$

Restricting $H^{(1)}$ to the positive spectral subspace $\Lambda_+(\mathcal{S}(\mathbb{R}^3) \otimes \mathbb{C}^4)$, the second term on the r.h.s. of (2.10) vanishes and the remaining term agrees with the Brown-Ravenhall operator.
Let us now consider $n = 2$. From (2.9) it follows that the transformed Dirac operator in second order is determined by the first transformation, $U_1 = e^{iB_1}$, only. However, the existence of the second transformation, $U_2 = e^{iB_2}$, has to be established to show that $H^{(2)}$ is indeed the transformed operator, with a remainder of order γ^3. We have

$$U_1^{-1}\,H\,U_1 \;=\; D_0 + V_1 + W_1 \;+\; i\,[D_0, B_1] \;+\; i\,[V, B_1] \;-\; \frac{1}{2}[[D_0, B_1], B_1] \;+\; R,$$

$$R \;=\; -\int_0^1 d\tau \int_0^\tau dt'\; U_1(-t')\;[[V,B_1],B_1]\;U_1(t') \tag{2.11}$$

$$-\,i\int_0^1 d\tau \int_0^\tau dt' \int_0^{t'} d\tau'\; U_1(-\tau')\;[[[D_0,B_1],B_1],B_1]\;U_1(\tau').$$

Making use of the defining relation for B_1, $W_1 + i[D_0,B_1] = 0$, the operator $H^{(2)}$ takes the form

$$H^{(2)} \;=\; D_0 + V_1 \;+\; \Lambda_+ H_2 \Lambda_+ \;+\; \Lambda_- H_2 \Lambda_- \tag{2.12}$$

$$H_2 := \; i\,[V_1,B_1] \;+\; \frac{i}{2}[W_1,B_1].$$

2.2 Integral operators in Fourier space and the determination of B_1

Since D_0 is a multiplication operator in momentum space, it is convenient to set up the calculus in Fourier space. Let $\varphi \in \mathcal{S}(\mathbb{R}^3) \otimes \mathbb{C}^4$. We define an integral operator W acting on φ by means of

$$(W\,\varphi)(\mathbf{x}) := \frac{1}{(2\pi)^{3/2}} \int d\mathbf{p}\; e^{i\mathbf{p}\mathbf{x}}\, w(\mathbf{x},\mathbf{p})\; \hat{\varphi}(\mathbf{p}) \tag{2.13}$$

where here and in the following the (three-dimensional) momentum integrals extend over the whole $\mathbb{R}^3$. This agrees with the formal definition of a pseudodifferential operator [13] and we will call $w(\mathbf{x},\mathbf{p})$ the symbol of W. Introducing the Fourier transform $\hat{w}(\mathbf{q},\mathbf{p})$, $W\varphi$ takes the form

$$(W\,\varphi)(\mathbf{x}) \;=\; \frac{1}{(2\pi)^3} \int d\mathbf{p}\; e^{i\mathbf{p}\mathbf{x}} \int d\mathbf{q}\; e^{i\mathbf{q}\mathbf{x}}\; \hat{w}(\mathbf{q},\mathbf{p})\; \hat{\varphi}(\mathbf{p}). \tag{2.14}$$

From this, the Fourier transform of $W\varphi$ is found

$$(\widehat{W\varphi})(\mathbf{p}') \;=\; \frac{1}{(2\pi)^{3/2}} \int d\mathbf{p}\; \hat{w}(\mathbf{p}'-\mathbf{p},\mathbf{p})\; \hat{\varphi}(\mathbf{p}). \tag{2.15}$$

With φ in (2.14) replaced by $G\varphi$, the symbol of a product WG of two integral operators is derived,

$$(W\,G\,\varphi)(\mathbf{x}) \;=\; \frac{1}{(2\pi)^3} \int d\mathbf{p}'\; e^{i\mathbf{p}'\mathbf{x}} \int d\mathbf{p}\; \hat{w}(\mathbf{p}'-\mathbf{p},\mathbf{p})\; \widehat{G\varphi}(\mathbf{p}). \tag{2.16}$$

Using (2.15) for $\widehat{G\varphi}(\mathbf{p})$, as well as the definition (2.14) of the Fourier transformed symbol $\widehat{wg}$ of WG, one gets

$$\widehat{wg}(\mathbf{q},\mathbf{p}) \;=\; \frac{1}{(2\pi)^{3/2}} \int d\mathbf{p}'\; \hat{w}(\mathbf{q}-\mathbf{p}',\mathbf{p}+\mathbf{p}')\; \hat{g}(\mathbf{p}',\mathbf{p}). \tag{2.17}$$

For the goal of determining the transformation U_1, its exponent B_1 is considered as an integral operator. One has to solve (2.8) for $n = 1$, using $W_1 = \Lambda_+ V \Lambda_- + \Lambda_- V \Lambda_+$ and (1.2),

$$-i\,[D_0, B_1] \;=\; W_1 \;=\; \frac{1}{2}\left(V \;-\; \frac{D_0}{|D_0|}\,V\,\frac{D_0}{|D_0|}\right). \tag{2.18}$$

Let ϕ_1 be the symbol of B_1. From (2.14) and with D_0 from (1.1) one has

$$(D_0\,B_1\,\varphi)(\mathbf{x}) \;=\; \frac{1}{(2\pi)^3}\int d\mathbf{p}\,d\mathbf{q}\,D_0\,e^{i(\mathbf{p}+\mathbf{q})\mathbf{x}}\,\hat{\phi}_1(\mathbf{q},\mathbf{p})\,\hat{\varphi}(\mathbf{p})$$

$$=\;\frac{1}{(2\pi)^3}\int d\mathbf{p}\,d\mathbf{q}\,[\boldsymbol{\alpha}(\mathbf{p}+\mathbf{q}) \;+\; \beta m]\,e^{i(\mathbf{p}+\mathbf{q})\mathbf{x}}\,\hat{\phi}_1(\mathbf{q},\mathbf{p})\,\hat{\varphi}(\mathbf{p}). \tag{2.19}$$

The Fourier transforms of $D_0\varphi$ and of $V\varphi$ are, respectively, obtained from

$$(\widehat{D_0\varphi})(\mathbf{p}) \;=\; (\boldsymbol{\alpha}\mathbf{p} \;+\; \beta m)\,\hat{\varphi}(\mathbf{p})$$

$$(V\,\varphi)(\mathbf{x}) \;=\; \frac{1}{(2\pi)^{3/2}}\int d\mathbf{q}\,e^{i\mathbf{q}\mathbf{x}}\left(-\frac{\gamma}{2\pi^2 q^2}\right)\int d\mathbf{p}\,e^{i\mathbf{p}\mathbf{x}}\,\hat{\varphi}(\mathbf{p}) \tag{2.20}$$

such that the symbol v of V is defined by $\hat{v}(\mathbf{q},\mathbf{p}) \;=\; -\sqrt{2/\pi}\,\gamma/q^2$. Acting (2.18) on φ and equating the respective symbols leads to the following algebraic equation for $\hat{\phi}_1$:

$$[\boldsymbol{\alpha}(\mathbf{p}+\mathbf{q}) \;+\; \beta m]\,\hat{\phi}_1(\mathbf{q},\mathbf{p}) \;-\; \hat{\phi}_1(\mathbf{q},\mathbf{p})\,[\boldsymbol{\alpha}\mathbf{p} + \beta m] \;=\; i\,\hat{w}_1(\mathbf{q},\mathbf{p}) \tag{2.21}$$

$$=\; -\frac{i\gamma_0}{q^2}\,[1 \;-\; \tilde{D}_0(\mathbf{q}+\mathbf{p})\cdot\tilde{D}_0(\mathbf{p})]$$

with $\gamma_0 \;:=\; \gamma/\sqrt{2\pi}$ and $\tilde{D}_0(\mathbf{p})$ the operator from (1.3) with norm unity. $\hat{w}_1(\mathbf{q},\mathbf{p})$, behaving like q^{-1} for $q \to 0$, is less singular than $\hat{v}(\mathbf{q},\mathbf{p})$, such that the prescription (2.6) for W_1 implies a regularisation of the potential V.

Lemma 2.1. *A solution $\hat{\phi}_1(\mathbf{q},\mathbf{p})$ for the symbol of B_1 is given by*

$$\hat{\phi}_1(\mathbf{q},\mathbf{p}) \;=\; -\frac{i\gamma_0}{q^2}\,\frac{1}{E_p + E_{|\mathbf{q}+\mathbf{p}|}}\,\left(\tilde{D}_0(\mathbf{q}+\mathbf{p}) \;-\; \tilde{D}_0(\mathbf{p})\right) \tag{2.22}$$

which satisfies the condition for symmetry of B_1 [13],

$$\hat{\phi}_1(-\mathbf{q},\mathbf{p}+\mathbf{q})^* \;=\; \hat{\phi}_1(\mathbf{q},\mathbf{p}). \tag{2.23}$$

It is estimated by

$$|\hat{\phi}_1(\mathbf{q},\mathbf{p})| \;\leq\; \frac{c}{q}\,\frac{1}{(q+p+1)^2} \tag{2.24}$$

with some constant $c \in \mathbb{R}_+$. B_1 is a bounded operator on $L_2(\mathbb{R}^3)\otimes\mathbb{C}^4$.

Proof. a) Calculation of $\hat{\phi}_1$.

In order to solve (2.21) the ansatz is made

$$\hat{\phi}_1(\mathbf{q},\mathbf{p}) \;=\; -\frac{i\gamma_0}{q^2}\,(c_1\boldsymbol{\alpha}\mathbf{q} \;+\; c_2\boldsymbol{\alpha}\mathbf{p} \;+\; c_3\beta) \tag{2.25}$$

and from the properties of the Dirac matrices $\beta^2 = 1$, $\alpha_i^2 = 1$, $\beta\alpha_i = -\alpha_i\beta$, $i = 1,2,3$, $\alpha_i\alpha_k = -\alpha_k\alpha_i$ $(i \neq k)$, the following identities are derived

$$\boldsymbol{\alpha}\mathbf{p}\cdot\boldsymbol{\alpha}\mathbf{p} \;=\; p^2, \qquad\qquad \boldsymbol{\alpha}\mathbf{q}\cdot\boldsymbol{\alpha}\mathbf{p} \;=\; 2\mathbf{pq} \;-\; \boldsymbol{\alpha}\mathbf{p}\cdot\boldsymbol{\alpha}\mathbf{q}. \tag{2.26}$$

Insertion of (2.25) into (2.21) then leads to an equation of the type

$$\lambda_1\,\boldsymbol{\alpha}\mathbf{p}\cdot\boldsymbol{\alpha}\mathbf{q} \;+\; \lambda_2\,\boldsymbol{\alpha}\mathbf{q}\cdot\beta \;+\; \lambda_3\,\boldsymbol{\alpha}\mathbf{p}\cdot\beta \;+\; \lambda_4 \;=\; 0 \tag{2.27}$$

where the λ_k, $k = 1,...,4$, are scalars depending on $\mathbf{p}$ and $\mathbf{q}$. (2.27) must hold for $\mathbf{p},\mathbf{q} \in \mathbb{R}^3$ whence $\lambda_k = 0$, $k = 1,...,4$. The resulting system of 4 equations for the c_i, $i = 1,2,3$ has a unique solution,

$$c_1\,(q^2{+}2\mathbf{pq}) \;=\; 1 - \frac{E_p}{E_{|\mathbf{q}+\mathbf{p}|}}, \qquad c_2 \;=\; 2c_1 - \frac{1}{E_p\,E_{|\mathbf{q}+\mathbf{p}|}}, \qquad c_3 \;=\; c_2 m \tag{2.28}$$

such that

$$\hat{\phi}_1(\mathbf{q},\mathbf{p}) = -\frac{i\gamma_0}{q^2}\left[[(\mathbf{q}{+}2\mathbf{p})\boldsymbol{\alpha}{+}2\beta m]\frac{1}{q^2{+}2\mathbf{pq}}\left(1 - \frac{E_p}{E_{|\mathbf{q}+\mathbf{p}|}}\right) - \frac{\mathbf{p}\boldsymbol{\alpha} + \beta m}{E_p E_{|\mathbf{q}+\mathbf{p}|}}\right]. \tag{2.29}$$

It is readily verified that (2.29) can be cast into the form (2.22), proving that $\hat{\phi}_1(\mathbf{q},\mathbf{p})$ is continuous in both variables except for $q = 0$.

b) The symmetry condition (2.23) follows immediately from (2.22) using the self-adjointness of $\boldsymbol{\alpha}$ and β.

c) We define the class of our integral operators (2.14) by means of the estimate of their symbols in the six-dimensional space $(\mathbf{q},\mathbf{p}) \in \mathbb{R}^6$. This estimate determines the convergence properties of integrals without the precise knowledge of the symbols themselves, and it is an easy way to deal with products of integral operators in proofs of boundedness or p-form boundedness.

In order to estimate a symbol by its asymptotic behaviour for $q, p \to 0$ and $q, p \to \infty$, it must be a continuous function of the two variables in $\mathbb{R}_+ \times \mathbb{R}_+$. This condition is fulfilled for $\hat{\phi}_1(\mathbf{q},\mathbf{p})$. By inspection of (2.22) one finds that $\hat{\phi}_1(\mathbf{q},\mathbf{p})$ is finite $\neq 0$ for $p = 0$ while it behaves $\sim 1/q$, $q \to 0$, $\sim 1/q^3$, $q \to \infty$ and $\sim 1/p^2$, $p \to \infty$. Taken into consideration that $\phi_1(\mathbf{x},\mathbf{p})$ is dimensionless as is B_1 (cf. $U_1 = e^{iB_1}$ and (2.13)) whence $\hat{\phi}_1(\mathbf{q},\mathbf{p})$ is of dimension (momentum)$^{-3}$, the estimate

$|\hat{\phi}_1(\mathbf{q},\mathbf{p})| \leq c/q \cdot (q+p+1)^{-2}$ is obtained. The constant c is determined by the maximum of $q \cdot |\hat{\phi}_1(\mathbf{q},\mathbf{p})|$ in $q,p \in \mathbb{R}_+$ (which exists due to continuity) and by the decay constants of $\hat{\phi}_1(\mathbf{q},\mathbf{p})$ for $q \to \infty$ and $p \to \infty$, respectively (maximised in the second variable).

d) We present here only the proof of the form boundedness of B_1; the operator boundedness can be shown along the same lines.

The basic ingredient is the *Lieb and Yau formula* which is a consequence of the Schur test [12] and which also can be derived from Schwarz's inequality [11]. We give it in a slightly generalised form,

$$\left| \int d\mathbf{q}\, d\mathbf{p}\, \overline{\hat{\varphi}(\mathbf{q})}\, |K(\mathbf{q},\mathbf{p})|\, \hat{\varphi}(\mathbf{p}) \right| \tag{2.30}$$

$$\leq \left(\int d\mathbf{q} d\mathbf{p} |\hat{\varphi}(\mathbf{p})|^2 |K(\mathbf{q},\mathbf{p})| \left| \frac{f(p)}{f(q)} \right|^2 \right)^{\frac{1}{2}} \cdot \left(\int d\mathbf{q} d\mathbf{p} |\hat{\varphi}(\mathbf{p})|^2 |K(\mathbf{p},\mathbf{q})| \left| \frac{f(p)}{f(q)} \right|^2 \right)^{\frac{1}{2}}$$

with $f(p) > 0$ for $p > 0$ a smooth convergence generating function. For a symmetric kernel, $K(\mathbf{p},\mathbf{q}) = K(\mathbf{q},\mathbf{p})$, (2.30) simplifies to the conventional form [11, 5].

From the condition (2.23) we have the following symmetry with respect to interchange of $\mathbf{q}$ and $\mathbf{p}$,

$$|\hat{\phi}_1^*(\mathbf{q}-\mathbf{p},\mathbf{p})| \;=\; |\hat{\phi}_1(\mathbf{p}-\mathbf{q},\mathbf{q})| \;\overset{\mathbf{q}\leftrightarrow\mathbf{p}}{\mapsto}\; |\hat{\phi}_1(\mathbf{q}-\mathbf{p},\mathbf{p})|. \tag{2.31}$$

One then obtains for $\varphi \in L_2(\mathbb{R}^3) \otimes \mathbb{C}^4$ applying (2.15) and subsequently (2.30) and (2.31)

$$|(\varphi, B_1\varphi)| = |(\hat{\varphi}, \widehat{B_1\varphi})| \leq \frac{1}{(2\pi)^{\frac{3}{2}}} \int d\mathbf{q}\, |\overline{\hat{\varphi}(\mathbf{q})}| \int d\mathbf{p}\, |\hat{\phi}_1(\mathbf{q}-\mathbf{p},\mathbf{p})|\, |\hat{\varphi}(\mathbf{p})| \tag{2.32}$$

$$\leq \frac{1}{(2\pi)^{\frac{3}{2}}} \left(\int d\mathbf{q}\, d\mathbf{p}\, |\hat{\varphi}(\mathbf{p})|^2\, |\hat{\phi}_1(\mathbf{q}-\mathbf{p},\mathbf{p})| \left| \frac{f(p)}{f(q)} \right|^2 \right)^{\frac{1}{2}}$$

$$\cdot \left(\int d\mathbf{q}\, d\mathbf{p}\, |\hat{\varphi}(\mathbf{p})|^2\, |\hat{\phi}_1^*(\mathbf{q}-\mathbf{p},\mathbf{p})| \left| \frac{f(p)}{f(q)} \right|^2 \right)^{\frac{1}{2}}.$$

Both symbol and its adjoint can be estimated by the same expression (2.24) (from (2.22) one even has $\hat{\phi}_1^*(\mathbf{q},\mathbf{p}) = -\hat{\phi}_1(\mathbf{q},\mathbf{p})$), yielding

$$|(\varphi, B_1\varphi)| \leq \frac{c}{(2\pi)^{\frac{3}{2}}} \int d\mathbf{p}\, |\hat{\varphi}(\mathbf{p})|^2 \cdot I_1(p)$$

$$I_1(p) := \int d\mathbf{q}\, \frac{1}{|\mathbf{q}-\mathbf{p}|}\, \frac{1}{(|\mathbf{q}-\mathbf{p}|+p+1)^2} \left| \frac{f(p)}{f(q)} \right|^2. \tag{2.33}$$

For the form boundedness of B_1 it remains to prove that $I_1(p)$ is bounded for $p \in \mathbb{R}_+$. The angular integration is performed with the help of the formula

$$\int_{-1}^{1} dx \, \frac{1}{\sqrt{b+ax}} \, \frac{1}{(\sqrt{b+ax}\,+p+1)^2} \;=\; \frac{2}{a} \int_{\sqrt{b-a}}^{\sqrt{b+a}} \frac{dz}{(z+p+1)^2}$$

$$= \; \frac{2}{a} \left(\frac{1}{\sqrt{b-a}\;+p+1} \;-\; \frac{1}{\sqrt{b+a}\;+p+1} \right) \tag{2.34}$$

identifying $|\mathbf{q}-\mathbf{p}| \;=\; \sqrt{q^2+p^2-2qpx} \;=: \sqrt{b+ax}, \quad x := \cos\vartheta_{\mathbf{q},\mathbf{p}}$. Choosing $f(p) := p^{\frac{1}{2}}$, one obtains

$$I_1(p) = 2\pi \int_0^\infty dq \left(\frac{1}{|q-p|+p+1} - \frac{1}{q+2p+1} \right) = 4\pi \ln \frac{2p+1}{p+1} < \infty. \tag{2.35}$$

□

As a consequence of the boundedness on $L_2(\mathbb{R}^3) \otimes \mathbb{C}^4$ and the symmetry, B_1 is self-adjoint and $U_1 = e^{iB_1}$ is unitary.

2.3 Existence of B_2 and the transformations of higher order

Let ϕ_n denote the symbol of B_n. It is briefly indicated how ϕ_2 can be obtained explicitly, but for B_n, $\;n \geq 2$, the calculus with operator classes is applied instead. B_2 is defined by

$$-i[D_0, B_2] = W_2 = i\Lambda_+ \left[[V_1, B_1] + \frac{1}{2}[W_1, B_1] \right] \Lambda_- + i\Lambda_- \left[[V_1, B_1] + \frac{1}{2}[W_1, B_1] \right] \Lambda_+. \tag{2.36}$$

With W_1 from (2.18) and $V_1 = V - W_1$ one obtains

$$W_2 \;=\; \frac{i}{8} \left(3\,[V, B_1] \;+\; [\tilde{D}_0, V\tilde{D}_0\, B_1] \;+\; [\tilde{D}_0, B_1 \tilde{D}_0 V] \;+\; 3\,\tilde{D}_0\,[B_1, V]\,\tilde{D}_0 \right). \tag{2.37}$$

(2.36) is, like the corresponding equation for B_1, solved in momentum space by introducing the respective symbols of the operators. The Fourier transformed symbol $\hat{w}_2$ of W_2 is composed of expressions of the type (cf (2.17))

$$\widehat{v\phi_1}(\mathbf{q},\mathbf{p}) \;=\; -\frac{\gamma}{2\pi^2} \int d\mathbf{p}' \, \frac{1}{|\mathbf{q}-\mathbf{p}'|^2} \, \hat{\phi}_1(\mathbf{p}',\mathbf{p}) \tag{2.38}$$

where $\widehat{v\phi_1}$ is the Fourier transformed symbol of the product VB_1. This implies that — in contrast to (2.21) — the equation for $\hat{\phi}_2(\mathbf{p},\mathbf{q})$,

$$[\boldsymbol{\alpha}\,(\mathbf{p}+\mathbf{q}) \;+\; \beta m]\; \hat{\phi}_2(\mathbf{q},\mathbf{p}) \;-\; \hat{\phi}_2(\mathbf{q},\mathbf{p})\;[\boldsymbol{\alpha}\mathbf{p} \;+\beta m] \;=\; i\,\hat{w}_2(\mathbf{q},\mathbf{p}), \tag{2.39}$$

involves an extra integral on the r.h.s. and is solved with the ansatz (using (2.22) for $\hat{\phi}_1(\mathbf{p}',\mathbf{p})$)

$$\hat{\phi}_2(\mathbf{q},\mathbf{p}) \;=\; -\frac{i\gamma\gamma_0}{16\pi^2}\int d\mathbf{p}'\,\frac{1}{p'^2}\,\frac{1}{|\mathbf{q}-\mathbf{p}'|^2}\;(c_1 \;+\; c_2\,\boldsymbol{\alpha}\mathbf{p}\cdot\beta m$$

$$+\, c_3\,\boldsymbol{\alpha}\mathbf{q}\cdot\beta m \;+\; c_4\,\boldsymbol{\alpha}\mathbf{p}'\cdot\beta m \;+\; c_5\,\boldsymbol{\alpha}\mathbf{p}'\cdot\boldsymbol{\alpha}\mathbf{p} \;+\; c_6\,\boldsymbol{\alpha}\mathbf{q}\cdot\boldsymbol{\alpha}\mathbf{p} \;+\; c_7\,\boldsymbol{\alpha}\mathbf{q}\cdot\boldsymbol{\alpha}\mathbf{p}')\,. \tag{2.40}$$

The scalar coefficients $c_j,\;\; j=1,...,7$ (depending on $\mathbf{q},\mathbf{p}$ and $\mathbf{p}'$) are uniquely determined.

The matter of interest is, however, not the explicit form of B_2 or generally, of $B_n,\;\; n\geq 2,$ but the existence of the potentials W_n and V_n on the form domain $H_{1/2}(\mathbb{R}^3)\otimes\mathbb{C}^4$ of the free Dirac operator, such that the expectation value of $H^{(n)}$ from (2.9) is well defined.

Lemma 2.2. *Let $U_n = e^{iB_n},\;\; n\geq 1,$ be the transformations from Theorem 1.1. Let ϕ_n be the symbol of B_n and W_n the potential in the defining equation for ϕ_n. Then W_n is p-form bounded on $H_{1/2}(\mathbb{R}^3)\otimes\mathbb{C}^4$ by means of*

$$|(\varphi,W_n\varphi)|\;\leq\; c\,(\varphi,p\,\varphi) \tag{2.41}$$

with $c\in\mathbb{R}_+$, and B_n extends to a bounded operator on $L_2(\mathbb{R}^3)\otimes\mathbb{C}^4$.

From the inductive definition of the transformation scheme it is easily seen that W_n is symmetric. As a consequence of (2.8), B_n is symmetric and hence self-adjoint, such that U_n is unitary for all $n\geq 1$.

Proof.

a) p-form boundedness of W_n

The proof is by induction. The p-form boundedness of W_1, eq. (2.18), follows from Kato's inequality [10] and from the self-adjointness and boundedness of $\tilde{D}_0$,

$$|(\varphi,W_1\varphi)|\;\leq\;\frac{1}{2}\,|(\varphi,V\varphi)|\;+\;\frac{1}{2}\,|(\tilde{D}_0\varphi,V\tilde{D}_0\varphi)|\;\leq\;\frac{\gamma\pi}{4}\,(\varphi,p\varphi)\;+\;\frac{\gamma\pi}{4}\,(\varphi,p\varphi) \tag{2.42}$$

where in the second term, $\tilde{D}_0 p\tilde{D}_0 = p$ has been used.

Let $n\geq 2$. According to the transformation scheme outlined in section 2.1, W_n is composed of multiple commutators of V with $B_k,\;\; k<n$. In particular for $n=2,\;\; [V,B_1]$ contributes (see (2.37)) and for $n=3$, one needs $[[V,B_1],B_1]$ and $[V,B_2]$. The additionally occurring factors $\tilde{D}_0$ can be disregarded in the context of p-form boundedness since $\tilde{D}_0$ is a bounded multiplication operator in momentum space. In the general case, the product of all factors $B_k,\;\; k\leq n-1$, which enter into a given commutator contributing to W_n must be proportional to γ^{n-1} since W_n is of the order γ^n.

By induction hypothesis, W_k is p-form bounded on $H_{1/2}(\mathbb{R}^3)\otimes\mathbb{C}^4$ for $k\leq n-1$. This means that all commutators of smaller order than n in γ are

p-form bounded. In the induction step one has to show that $[V, B_{n-1}]$ and $[[\cdot], B_k]$, $k < n-1$, are p-form bounded where $[\cdot]$ denotes a p-form bounded multiple commutator.

Without loss of generality one may assume that $[\cdot]$ is symmetric. This guarantees the symmetry property (2.31) for the Fourier transformed symbol $\widehat{[\cdot]}(\mathbf{p}, \mathbf{q})$. We estimate $|\widehat{[\cdot]}| \leq |\widehat{[\cdot]}| + |\widehat{[\cdot]}^*|$ and apply the Lieb and Yau formula for this symmetrised kernel. Then, using that a symbol can be estimated by its adjoint and vice versa, p-form boundedness can be expressed in the following way

$$|(\varphi, [\cdot]\varphi)| \leq \frac{1}{(2\pi)^{\frac{3}{2}}} \int d\mathbf{p}\, |\hat{\varphi}(\mathbf{p})|^2 \int d\mathbf{q} \Big(\Big|\widehat{[\cdot]}(\mathbf{q}-\mathbf{p}, \mathbf{p})\Big| + \Big|\widehat{[\cdot]}^*(\mathbf{q}-\mathbf{p}, \mathbf{p})\Big| \Big) \left| \frac{f(p)}{f(q)} \right|^2$$

$$\leq \frac{1}{(2\pi)^{3/2}} \int d\mathbf{p}\ |\hat{\varphi}(\mathbf{p})|^2\ c\, p \ = \ c'\ (\varphi, p\, \varphi) \tag{2.43}$$

with some constant $c > 0$. The inequality in the second line of (2.43) restricts the convergence generating function to $f(p) := p^\lambda$ with $\frac{1}{2} < \lambda < \frac{3}{2}$. This is true because both $\widehat{[\cdot]}(\mathbf{q}-\mathbf{p}, \mathbf{p})$ and $\widehat{[\cdot]}(\mathbf{p}-\mathbf{q}, \mathbf{q})$ are regular for $q \to 0$ (since all operators of which $[\cdot]$ is composed have symbols which are regular when the second variable tends to zero), restricting $\lambda < \frac{3}{2}$, and because $\widehat{[\cdot]}$ is of dimension (momentum)$^{-2}$, decreasing like q^{-2} for $q \to \infty$, such that $\lambda > \frac{1}{2}$ is required. These properties hold also for the symmetric operator W_k. Thus for $k < n$ when W_k is p-form bounded, (2.43) therefore is also valid for its Fourier transformed symbol $\hat{w}_k$ in place of $\widehat{[\cdot]}$.

Another ingredient in the proof of p-form boundedness of W_n is the fact that the symbol ϕ_n can be estimated by w_n. First note that the estimate of $\hat{\phi}_n$ is related to the estimate of $\hat{w}_n$ by an equation of the type (2.39), derived from the defining equation (2.8). This equation implies that the behaviour of $\hat{\phi}_n$ for $p \to 0$ and $q \to 0$ is that of $\hat{w}_n$, while there occurs an extra power of q^{-1} and p^{-1} for $q \to \infty$ and $p \to \infty$, respectively. Therefore

$$|\hat{\phi}_n(\mathbf{q}, \mathbf{p})| \ \leq \ \frac{c}{q+p+1}\, |\hat{w}_n(\mathbf{q}, \mathbf{p})|. \tag{2.44}$$

i) p-form boundedness of $[V, B_{n-1}]$

From the symmetry of V, B_{n-1} and (2.15) we get

$$|(\varphi, [V, B_{n-1}]\varphi)| \ \leq \ |(\widehat{V\varphi}, \widehat{B_{n-1}\varphi})| \ + \ |(\widehat{B_{n-1}\varphi}, \widehat{V\varphi})|$$

$$\leq \frac{1}{(2\pi)^3} \int d\mathbf{p}' d\mathbf{p}\, d\mathbf{q}\, |\hat{\varphi}(\mathbf{p})| \Big\{ |\hat{v}^*(\mathbf{p}'-\mathbf{p}, \mathbf{p})|\, |\hat{\phi}_{n-1}(\mathbf{p}'-\mathbf{q}, \mathbf{q})|$$

$$+ |\hat{\phi}^*_{n-1}(\mathbf{p}'-\mathbf{p}, \mathbf{p})|\, |\hat{v}(\mathbf{p}'-\mathbf{q}, \mathbf{q})| \Big\} |\hat{\varphi}(\mathbf{q})|. \tag{2.45}$$

Due to the symmetry in $\mathbf{p}$ and $\mathbf{q}$, the Lieb and Yau formula can be applied in the same way as in (2.32) and in (2.43). Using $f(p) := p$ for the convergence generating function and estimating $\hat{v}$ one gets

$$|(\varphi, [V, B_{n-1}]\,\varphi)| \;\leq\; c \int d\mathbf{p}\;\; |\hat{\varphi}(\mathbf{p})|^2 \;\left\{ \int d\mathbf{p}'\;\frac{1}{|\mathbf{p}'-\mathbf{p}|^2}\;\frac{p^2}{p'^2} \cdot \int d\mathbf{q} \right.$$

$$\left. \left|\hat{\phi}_{n-1}(\mathbf{p}'-\mathbf{q},\mathbf{q})\right| \frac{p'^2}{q^2} \; + \int d\mathbf{p}' \left|\hat{\phi}_{n-1}(\mathbf{p}-\mathbf{p}',\mathbf{p}')\right| \frac{p^2}{p'^2} \cdot \int d\mathbf{q} \frac{1}{|\mathbf{p}'-\mathbf{q}|^2}\,\frac{p'^2}{q^2} \right\}. \tag{2.46}$$

The last $\mathbf{q}$-integral is evaluated with the substitution $\mathbf{q}' := \mathbf{q}/p'$ and with $\mathbf{e}_{p'} := \mathbf{p}'/p'$,

$$\int d\mathbf{q} \frac{1}{|\mathbf{q}-\mathbf{p}'|^2}\,\frac{p'^2}{q^2} = p' \int_0^\infty dq' \int_{S^2} d\Omega_{q'} \frac{1}{|\mathbf{q}'-\mathbf{e}_{p'}|^2} \;=\; 2\pi p' \int_0^\infty \frac{dq'}{q'} \ln \left| \frac{1+q'}{1-q'} \right|$$

$$= \; \pi^3 p'. \tag{2.47}$$

Also $\hat{\phi}_{n-1}$ is estimated by $\hat{w}_{n-1}$ via (2.44). One obtains for the second term in curly brackets of (2.46) with the help of the p-form boundedness of W_{n-1} by means of (2.43)

$$\int d\mathbf{p}' |\hat{\phi}_{n-1}(\mathbf{p}-\mathbf{p}',\mathbf{p}')| \frac{p^2}{p'^2} \cdot \pi^3 p' \leq \tilde{c} \int d\mathbf{p}' \frac{1}{|\mathbf{p}-\mathbf{p}'| + p' + 1}\, |\hat{w}_{n-1}(\mathbf{p}-\mathbf{p}',\mathbf{p}')| \frac{p^2}{p'}$$

$$\leq c'' p \tag{2.48}$$

where $|\mathbf{p}-\mathbf{p}'| + p' + 1 \;\geq p'$ has been used. Further we note that the factor $(|\mathbf{p}'-\mathbf{q}| + q + 1)^{-1}$ is bounded for all $q \geq 0$ and hence can be estimated by its value at $q = 0$. We thus get for the other $\mathbf{q}$-integral

$$\int d\mathbf{q}\; |\hat{\phi}_{n-1}(\mathbf{p}'-\mathbf{q},\mathbf{q})|\; \frac{p'^2}{q^2} \;\leq\; \frac{\tilde{c}}{p'+1} \int d\mathbf{q}\; |\hat{w}_{n-1}(\mathbf{p}'-\mathbf{q},\mathbf{q})|\; \frac{p'^2}{q^2}$$

$$\leq\; \frac{\tilde{c}}{p'+1} \cdot c p' \;\leq\; \tilde{c} c \tag{2.49}$$

such that by means of (2.47), the first term in the curly brackets of (2.46) is estimated by $c'p$ with some constant $c' \in \mathbb{R}_+$. This proves the p-form boundedness of $[V, B_{n-1}]$.

ii) p-form boundedness of $[[\cdot], B_k]$

In (2.45), V and B_{n-1} are replaced with $[\cdot]$ and B_k, respectively, and the expression in curly brackets (integrated over $\mathbf{p}'$) is taken as the kernel $K(\mathbf{q},\mathbf{p})$ in the Lieb and Yau formula (2.30). Then, estimating the symbol by its adjoint and vice versa, one arrives at

$$|(\varphi, [\,[\cdot], B_k]\,\varphi)| \;\leq\; \frac{\tilde{c}}{(2\pi)^3} \int d\mathbf{p}\;\; |\hat{\varphi}(\mathbf{p})|^2 \;\cdot \left\{ \int d\mathbf{p}'\;\; \left|\widehat{[\cdot]}(\mathbf{p}-\mathbf{p}',\mathbf{p}')\right|\; \frac{p^2}{p'^2} \right.$$

$$\cdot \int d\mathbf{q} \left|\hat{\phi}_k(\mathbf{p}'-\mathbf{q},\mathbf{q})\right| \frac{p'^2}{q^2} + \int d\mathbf{p}' \left|\hat{\phi}_k(\mathbf{p}-\mathbf{p}',\mathbf{p}')\right| \frac{p^2}{p'^2} \int d\mathbf{q} \left|\widehat{[\cdot]}(\mathbf{p}'-\mathbf{q},\mathbf{q})\right| \frac{p'^2}{q^2} \Big\}. \tag{2.50}$$

By (2.43), the last $\mathbf{q}$-integral is bounded by cp' such that the second term in the curly brackets can be estimated by

$$\int d\mathbf{p}' |\hat{\phi}_k(\mathbf{p}-\mathbf{p}',\mathbf{p}')| \frac{p^2}{p'} \le c \int d\mathbf{p}' \frac{1}{|\mathbf{p}-\mathbf{p}'| + p' + 1} |\hat{w}_k(\mathbf{p}-\mathbf{p}',\mathbf{p}')| \frac{p^2}{p'} \le c' p \tag{2.51}$$

according to (2.48) because $\hat{w}_k$ is p-form bounded. By (2.49), the other $\mathbf{q}$-integral is estimated by a constant $\tilde{c}$, such that the first term of (2.50) in curly brackets is with (2.43) estimated by $p \cdot c''$ with some constant c''. This proves the p-form boundedness of $[[\cdot], B_k]$, $k < n-1$, and hence together with (i) the p-form boundedness of W_n.

From the p-form boundedness of W_n, $n \ge 1$, on $H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^4$, proven above, follows immediately the p-form boundedness of V_n, $n \ge 1$ since both operators differ only by factors $\tilde{D}_0$. We have therefore proven that to arbitrary order n,

$$|(\varphi, (V_1 + ... + V_n)\, \varphi)| \le c\, (\varphi, p\, \varphi) \tag{2.52}$$

for $\varphi \in H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^4$.

b) Boundedness of B_n

This is a consequence of the p-form boundedness of W_n. From (2.43) with $[\cdot]$ replaced by B_n one gets

$$|(\varphi, B_n \varphi)| \le \frac{c}{(2\pi)^{3/2}} \int d\mathbf{p}\, |\hat{\varphi}(\mathbf{p})|^2 \int d\mathbf{q}\, |\hat{\phi}_n^*(\mathbf{q}-\mathbf{p},\mathbf{p})| \frac{p^2}{q^2} \tag{2.53}$$

where the convergence generating function is again chosen as $f(p) := p$. From (2.49) the $\mathbf{q}$-integral is estimated by a constant. Hence,

$$|(\varphi, B_n \varphi)| \le \text{const}\, (\varphi, \varphi). \tag{2.54}$$

□

Remark. Due to logarithmic divergencies occurring in the estimates of $\hat{w}_n(\mathbf{q},\mathbf{p})$, $n \ge 1$, the proof of boundedness of B_n cannot be based on the algebra of symbol estimates, a powerful method in the case of periodic potentials [13].

2.4 The Remainder $R^{(n+1)}$ and its p-form boundedness

From its definition as remainder after multiple iterations of (2.3)-type equations (see e.g. (2.11)), $R^{(n+1)}$ is composed of a finite number of compact integrals over a unitary transform of the same multiple commutators $[\cdot]$ which would contribute to the $n+1$st order term V_{n+1} after one additional transformation (for the commutator involving D_0, use (2.8)). These commutators are p-form bounded according to the proof of Lemma 2.2, and it remains to show that the unitary transform preserves the p-form boundedness. Consider

$$|(\varphi,\, U_k(-\tau)\,[\cdot]\,U_k(\tau)\,\varphi)| \;=\; |(U_k(\tau)\,\varphi,\, [\cdot]\,U_k(\tau)\,\varphi)|$$
$$\leq\; c\,(U_k(\tau)\,\varphi,\, p\,U_k(\tau)\,\varphi) \;=\; c\,(\varphi,\, U_k(-\tau)\,p\,U_k(\tau)\,\varphi). \tag{2.55}$$

Since $U_k(\tau)=e^{iB_k\tau}$ with B_k a bounded operator, we can Taylor expand

$$(\varphi, e^{-iB_k\tau}\,p\,e^{iB_k\tau}\,\varphi) \;\leq\; \sum_{n,m=0}^{\infty}\frac{\tau^n}{n!}\,\frac{\tau^m}{m!}\,|(\varphi, B_k^n\,p\,B_k^m\,\varphi)|. \tag{2.56}$$

The sum on the r.h.s. represents a symmetric operator such that its kernel has the required symmetry property to apply the Lieb and Yau formula (with convergence generating function $f(p)=p$). Our proof proceeds in 4 steps: We prove p-form boundedness of (i) pB_k, (ii) pB_k^m (by induction), (iii) $B_kpB_k^m$, (iv) $B_k^npB_k^m$.

According to (2.32) we establish boundedness of an operator A by means of boundedness of the integral I_A over its Fourier transformed symbol $\hat{s}_A$. For B_k, we have boundedness from (2.49),

$$I_{B_k} := \frac{1}{(2\pi)^{3/2}}\int d\mathbf{q}\,|\hat{\phi}_k(\mathbf{p}-\mathbf{q},\mathbf{q})|\,\frac{p^2}{q^2} \;\leq\; c_k. \tag{2.57}$$

p-form boundedness is proven by showing that the integrals I_A (with $A := B_k^npB_k^m$) are proportional to p.

(i)

$$I_{pB_k} \;=\; \frac{1}{(2\pi)^{3/2}}\int d\mathbf{q}\,p\,|\hat{\phi}_k(\mathbf{p}-\mathbf{q},\mathbf{q})|\,\frac{p^2}{q^2} \;\leq\; p\,c_k. \tag{2.58}$$

(ii) Our induction hypothesis is $I_{pB_k^m} \leq pc_k^m$. We decompose $pB_k^{m+1} = pB_k^m\cdot B_k$ and use (2.17) for the symbol of a product of operators. Then with (2.57),

$$I_{pB_k^{m+1}} \;=\; \frac{1}{(2\pi)^{3/2}}\int d\mathbf{q}'\,|\hat{s}_{pB_k^{m+1}}(\mathbf{p}-\mathbf{q}',\mathbf{q}')|\,\frac{p^2}{q'^2} \tag{2.59}$$

$$\leq \frac{1}{(2\pi)^3}\int d\mathbf{q}\,|\hat{s}_{pB_k^m}(\mathbf{p}-\mathbf{q},\mathbf{q})|\,\frac{p^2}{q^2}\cdot\int d\mathbf{q}'\,|\hat{\phi}_k(\mathbf{q}-\mathbf{q}',\mathbf{q}')|\,\frac{q^2}{q'^2} \leq p\,c_k^m\cdot c_k = p\,c_k^{m+1}.$$

(iii) Decomposing $B_kpB_k^m = B_k\cdot pB_k^m$, one has from (2.51)

$$I_{B_kpB_k^m} \;\leq\; \frac{1}{(2\pi)^3}\int d\mathbf{q}\,|\hat{\phi}_k(\mathbf{p}-\mathbf{q},\mathbf{q})|\,\frac{p^2}{q^2}\cdot\int d\mathbf{q}'\,|\hat{s}_{pB_k^m}(\mathbf{q}-\mathbf{q}',\mathbf{q}')|\,\frac{q^2}{q'^2}$$
$$\leq\; c_k^m\,\frac{1}{(2\pi)^{3/2}}\int d\mathbf{q}\,|\hat{\phi}_k(\mathbf{p}-\mathbf{q},\mathbf{q})|\,\frac{p^2}{q} \;\leq\; c_k^m\,c_k'\,p. \tag{2.60}$$

(iv) We claim $I_{B_k^n p B_k^m} \le p\, c_k^{'n}\, c_k^m$. Then, using (2.60)

$$I_{B_k^{n+1} p B_k^m} \;\le\; \frac{1}{(2\pi)^3} \int d\mathbf{q}\; |\hat{\phi}_k(\mathbf{p}-\mathbf{q},\mathbf{q})|\; \frac{p^2}{q^2} \cdot \int d\mathbf{q}'\; |\hat{s}_{B_k^n p B_k^m}(\mathbf{q}-\mathbf{q}',\mathbf{q}')|\; \frac{q^2}{q'^2}$$

$$\le\; c_k^{'n}\, c_k^m \cdot c_k'\, p \;=\; p\, c_k^{'n+1}\, c_k^m. \tag{2.61}$$

Thus we obtain from the Lieb and Yau formula applied to (2.56)

$$\sum_{n,m=0}^{\infty} \frac{\tau^n}{n!}\, \frac{\tau^m}{m!}\, |(\varphi, B_k^n\, p\, B_k^m\, \varphi)| \;\le\; c \sum_{n,m=0}^{\infty} \frac{\tau^n}{n!}\, \frac{\tau^m}{m!}\, c_k^{'n}\, c_k^m\, (\varphi, p\,\varphi)$$

$$=\; c\, e^{c_k'\tau + c_k\tau}\, (\varphi, p\,\varphi) \tag{2.62}$$

with c a constant resulting from using the same estimate for symbol and its adjoint. This shows that $(\varphi, U_k(-\tau) p U_k(\tau)\varphi)$ is p-form bounded and completes the proof since $\exp(c_k'\tau + c_k\tau)$ is a continuous function of τ. With the same reasoning, any multiple finite-dimensional compact integral over multiple unitary transforms of p-form bounded commutators is therefore again p-form bounded.

3 Subordinacy of the higher-order contributions

Since to any order n the p-form bound of $H^{(n)}$ is proportional to γ^n while for the remainder $R^{(n+1)}$ it is proportional to γ^{n+1}, one gets convergence of the expansion in the strength of the Coulomb field for $\gamma \to 0$. However, for larger $\gamma < 1$, the p-form bounds of V_n obtained with the above estimates can in general not be restricted to numbers less than 1. In this section we will consider a slight modification of the Coulomb potential,

$$V(x) := -\frac{\gamma}{x^{1-\epsilon}}, \qquad \hat{v}(\mathbf{q}) \;=\; -\gamma \sqrt{\frac{2}{\pi}}\, \frac{f_\epsilon}{q^{2+\epsilon}}, \quad \epsilon > 0 \tag{3.1}$$

where $\hat{v}(\mathbf{q})$ is the Fourier transform and $f_\epsilon := \cos\frac{\pi\epsilon}{2} \cdot \Gamma(1+\epsilon) \;\to 1$ for $\epsilon \to 0$. All quantities defined previously will now pertain to the modified potential (3.1).
Our results are collected in the following proposition.

Proposition 3.1. *For the modified Coulomb potential (3.1) we have*

(i) For every $k \in \mathbb{N}$, $\epsilon < \frac{1}{k+1}$, the k-th order potential term V_k is p-form bounded with form bound less than 1.

(ii) Let $\varphi \in \mathcal{S}(\mathbb{R}^3) \otimes \mathbb{C}^4$. Let $\mu_k > 0$ for $k \in \mathbb{N}$ be the infimum of the constant c in the estimate $|(\varphi, V_k\varphi)| \le c\,(\varphi, p^{1-k\epsilon}\varphi)$. Then V_{k+1} is subordinate to V_k in the sense

$$|(\varphi, V_{k+1}\,\varphi)| \;\le\; \delta\, |(\varphi, V_k\varphi)| \;+\; C\,(\varphi,\varphi) \tag{3.2}$$

with $0 < \delta < 1$ and $C \in \mathbb{R}_+$ a constant depending on δ.

(iii) Let $R^{(n+1)} := U_n^* \cdots U_1^* H U_1 \cdots U_n - H^{(n)}$ *be the remainder of order* $n+1$ *in the potential strength. Then* $R^{(n+1)}$ *is subordinate to* V_n.

For the proof, a lemma is needed.

Lemma 3.2. *For* $0 < (n+1)\epsilon < 1$, $c_0 \in \mathbb{R}_+$, $n \in \mathbb{N}$ *and every* $\varphi \in H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^4$ *one has*

$$c_0\,(\varphi, p^{1-(n+1)\epsilon}\varphi) \;\le\; c\,(\varphi, p^{1-n\epsilon}\varphi) \;+\; C\,(\varphi,\varphi) \tag{3.3}$$

with $c < 1$ *and* $C \in \mathbb{R}_+$. *For* $n = 0$ *this implies* p*-form boundedness with form bound* < 1,

$$c_0\,(\varphi, p^{1-\epsilon}\varphi) \;\le\; c\,(\varphi, p\,\varphi) \;+\; C\,(\varphi,\varphi). \tag{3.4}$$

Proof. We use an elementary inequality from analysis,

$$a\cdot b \;\le\; \frac{a^\lambda}{\lambda} \;+\; \frac{b^\mu}{\mu}, \qquad a,b>0, \quad \frac{1}{\lambda} \;+\; \frac{1}{\mu} \;=1, \tag{3.5}$$

choose $\lambda := \frac{1-n\epsilon}{1-(n+1)\epsilon} > 1$, $\mu = \frac{1-n\epsilon}{\epsilon}$ and $0 < \delta < 1$ to be specified later. We decompose

$$p^{1-(n+1)\epsilon} = \left(\delta p^{1-(n+1)\epsilon}\right)\cdot\frac{1}{\delta} \le \frac{1-(n+1)\epsilon}{1-n\epsilon}\,\delta^{\frac{1-n\epsilon}{1-(n+1)\epsilon}}\,p^{1-n\epsilon} + \frac{\epsilon}{1-n\epsilon}\left(\frac{1}{\delta}\right)^{\frac{1-n\epsilon}{\epsilon}}. \tag{3.6}$$

Then, estimating further (using $\delta^\lambda < \delta$),

$$c_0\,(\varphi, p^{1-(n+1)\epsilon}\varphi) \;\le\; c_0\,\delta\,(\varphi, p^{1-n\epsilon}\varphi) \;+\; c_0\,\frac{\epsilon}{1-n\epsilon}\,\delta^{-\frac{1-n\epsilon}{\epsilon}}\,(\varphi,\varphi). \tag{3.7}$$

With the choice $\delta := \min\{\frac{1}{2c_0}, \frac{1}{2}\}$, (3.3) is verified. □

Proof of Proposition.

We start by showing that $|(\varphi, V_k\varphi)| \le c\,(\varphi, p^{1-k\epsilon}\varphi)$ with some constant $c > 0$, such that the definition of μ_k in (ii) makes sense. Since $0 < (\varphi, p^{1-k\epsilon}\varphi) < \infty$ (for $\varphi \ne 0$), $\mu_k = 0$ implies $(\varphi, V_k\varphi) = 0$ which means that in this case V_k does not contribute to the expectation value of the transformed Dirac operator (2.9) and hence can be disregarded.
First we estimate the expectation value of V_1. According to (2.21), the symbol of V_1 is given by
$\hat{v}_1(\mathbf{q},\mathbf{p}) = \hat{v}(\mathbf{q}) - \hat{w}_1(\mathbf{q},\mathbf{p}) = -\frac{\gamma_0}{q^{2+\epsilon}}\, f_\epsilon\,(1 + \tilde{D}_0(\mathbf{q}+\mathbf{p})\cdot\tilde{D}_0(\mathbf{p}))$. Since the multiplier of $q^{-(2+\epsilon)}$ is a bounded operator which is estimated by a constant, one finds according to (2.32) and (2.33) with $f(p) := p$

$$|(\varphi, V_1\,\varphi)| \;\le\; \frac{1}{(2\pi)^{3/2}} \int d\mathbf{p}\; |\hat{\varphi}(\mathbf{p})|^2\; I_1(p), \tag{3.8}$$

$$I_1(p) := \int d\mathbf{q}\, |\hat{v}_1(\mathbf{q}-\mathbf{p},\mathbf{p})| \left|\frac{f(p)}{f(q)}\right|^2 \le c_1 \int d\mathbf{q}\, \frac{1}{|\mathbf{q}-\mathbf{p}|^{2+\epsilon}} \cdot \frac{p^2}{q^2}.$$

With the substitution $p\,\mathbf{q}' := (\mathbf{q}-\mathbf{p})$ one obtains for the latter integral

$$p^3 \int d\mathbf{q}'\, \frac{1}{(pq')^{2+\epsilon}}\, \frac{p^2}{|p\mathbf{q}'+\mathbf{p}|^2} = 2\pi\, p^{1-\epsilon} \int_0^\infty \frac{dq'}{q'^{1+\epsilon}}\, \ln\left|\frac{q'+1}{q'-1}\right| \le c\cdot p^{1-\epsilon} \tag{3.9}$$

resulting in $|(\varphi, V_1\varphi)| \le c_0\,(\varphi, p^{1-\epsilon}\varphi)$ where $c_1, c, c_0 \in \mathbb{R}_+$ are constants. Equation (3.4) completes the proof of (i) for $k=1$. Note that since $\mu_1 = \inf c_0 > 0$, we have $|(\varphi, V_1\varphi)| > \frac{\mu_1}{2}\,(\varphi, p^{1-\epsilon}\varphi)$.

The proof of (ii) is by induction. First we show (ii) for $k=1$. From (2.12) we have $V_2 = i[V_1, B_1] + \frac{i}{2}\,[W_1, B_1] - W_2$ with W_2 from (2.37). Following the argumentation given in section 2.3 one can disregard the bounded operators $\tilde{D}_0$ in the estimates of p-form boundedness and consider V_2 as being represented by the commutator $[V, B_1]$.

With the modified Coulomb potential the symbol of B_1 which is proportional to $\hat{w}_1(\mathbf{q},\mathbf{p})$ according to (2.21), is estimated by $|\hat{\phi}_1(\mathbf{q},\mathbf{p})| \le \frac{c}{q^{1+\epsilon}}\,\frac{1}{(q+p+1)^2}$.

The estimate of $|(\varphi, [V, B_1]\,\varphi)|$ is obtained from (2.46) by means of interchanging $\hat{\phi}_{n+1}$ with $\hat{\phi}_1$ and $|\mathbf{p}'-\mathbf{p}|^2$, $|\mathbf{p}'-\mathbf{q}|^2$ with $|\mathbf{p}'-\mathbf{p}|^{2+\epsilon}$, $|\mathbf{p}'-\mathbf{q}|^{2+\epsilon}$. Recalling that $|\hat{\phi}_1|$ can be replaced by its adjoint $|\hat{\phi}_1^*|$ and substituting $\mathbf{q}' := \mathbf{q}/p'$ in the first integral one obtains

$$|(\varphi, [V, B_1]\varphi)| \le c \int d\mathbf{p}\, |\hat{\varphi}(\mathbf{p})|^2\, \{I_{11}(p) + I_{12}(p)\} \tag{3.10}$$

$$I_{11}(p) = \int d\mathbf{p}' \frac{1}{|\mathbf{p}'-\mathbf{p}|^{2+\epsilon}} \frac{p^2}{p'^2} \cdot \int d\mathbf{q}' \frac{1}{q'^2} \frac{1}{p'^{\epsilon}|\mathbf{q}'-\mathbf{e}_{p'}|^{1+\epsilon}(|\mathbf{q}'-\mathbf{e}_{p'}|+1+\frac{1}{p'})^2}$$

$$I_{12}(p) = \int d\mathbf{p}'\, \frac{1}{|\mathbf{p}'-\mathbf{p}|^{1+\epsilon}(|\mathbf{p}'-\mathbf{p}|+p+1)^2}\, \frac{p^2}{p'^2} \cdot \int d\mathbf{q}\, \frac{1}{|\mathbf{p}'-\mathbf{q}|^{2+\epsilon}}\, \frac{p'^2}{q^2}.$$

For I_{11}, the $\mathbf{q}'$-integral is estimated by dropping $\frac{1}{p'}$ in the last factor of the denominator. Using (2.47) together with the substitution $\mathbf{k} := \mathbf{q}' - \mathbf{e}_{p'}$, one finds

$$\int \frac{d\mathbf{q}'}{q'^2} \frac{1}{|\mathbf{q}'-\mathbf{e}_{p'}|^{1+\epsilon}} \frac{1}{(|\mathbf{q}'-\mathbf{e}_{p'}|+1)^2} = \int_0^\infty dk \frac{k^{1-\epsilon}}{(k+1)^2} \cdot \frac{2\pi}{k} \ln\left|\frac{k+1}{k-1}\right| < \infty. \tag{3.11}$$

The $\mathbf{p}'$-integral results with the substitution $p\,\mathbf{p}'' := \mathbf{p}'$ and with the same techniques for the angular integration as applied in (2.34), in

$$p^2 \int d\mathbf{p}' \frac{1}{|\mathbf{p}'-\mathbf{p}|^{2+\epsilon}} \frac{1}{p'^{2+\epsilon}} = p^{1-2\epsilon} \int_0^\infty dp'' \frac{1}{p''^{\epsilon}} \frac{2\pi}{\epsilon p''} \left(\frac{1}{|p''-1|^{\epsilon}} - \frac{1}{|p''+1|^{\epsilon}}\right)$$
$$\le c\cdot p^{1-2\epsilon} \tag{3.12}$$

with a constant c.

I_{12} is treated in a similar way. The $\mathbf{q}$-integral is the one from (3.8) which is estimated by $c\,p'^{1-\epsilon}$. The remaining integral is estimated by replacing $|\mathbf{p}'-\mathbf{p}|+p+1$ with $|\mathbf{p}'-\mathbf{p}|+p$. With the substitution $\mathbf{p}' =: p\mathbf{k}+\mathbf{p}$ and with the techniques from (2.34) one obtains

$$p^2 \int d\mathbf{p}' \, \frac{1}{|\mathbf{p}'-\mathbf{p}|^{1+\epsilon}\,(|\mathbf{p}'-\mathbf{p}|+p)^2\,p'^{1+\epsilon}}$$

$$= p^{1-2\epsilon}\int_0^\infty \frac{k^{1-\epsilon}dk}{(k+1)^2}\cdot\frac{2\pi}{k(1-\epsilon)}\,\left[(k+1)^{1-\epsilon} - |k-1|^{1-\epsilon}\right] \le c\,p^{1-2\epsilon} \tag{3.13}$$

with some constant c. Using (3.7) and the definition of μ_k one thus obtains

$$|(\varphi,V_2\varphi)| \le c_0\,(\varphi,p^{1-2\epsilon}\varphi) \le c_0\,\delta\,(\varphi,p^{1-\epsilon}\varphi) + C\,(\varphi,\varphi) \tag{3.14}$$

$$< \frac{2c_0\,\delta}{\mu_1}\,|(\varphi,V_1\varphi)| + C(\varphi,\varphi)$$

with $2c_0\delta/\mu_1 < 1$ for a suitably chosen δ. This proves (ii) of Proposition 3.1 for $k=1$.

The proof of the induction step from k to $k+1$ proceeds along the same lines as applied in section 2.3 to show the p-form boundedness of V_n and W_n. By induction hypothesis commutators of order $m \le k$ in the potential strength, denoted by $[\cdot]_m$, have the following symbol estimates (compare (2.43))

$$\int d\mathbf{q}\,\left(\left|\widehat{[\cdot]}_m(\mathbf{q}-\mathbf{p},\mathbf{p})\right| + \left|\widehat{[\cdot]}^*_m(\mathbf{q}-\mathbf{p},\mathbf{p})\right|\right)\left(\frac{p}{q}\right)^{2\lambda} \le c\,p^{1-m\epsilon} \tag{3.15}$$

where λ can be chosen in the interval $]\frac{1}{2},\frac{3}{2}[$. We demonstrate the proof for the commutator $[[\cdot]_m, B_{k-m+1}]$ which contributes to V_{k+1}. For the commutator $[V,B_k]$ which also contributes to V_{k+1} the proof is similar. Since the symbol classes of W_m and $[\cdot]_m$ are equal, it follows from (2.44)

$$|\hat{\phi}_m(\mathbf{q},\mathbf{p})| \le \frac{c}{q+p+1}\left|\widehat{[\cdot]}_m(\mathbf{q},\mathbf{p})\right| \le \frac{c}{q+1}\left|\widehat{[\cdot]}_m(\mathbf{q},\mathbf{p})\right|. \tag{3.16}$$

Then from (2.50), one has with some $c_0 \in \mathbb{R}_+$,

$$|(\varphi,[[\cdot]_m,B_{k-m+1}]\,\varphi)| \le \frac{c_0}{(2\pi)^3}\int d\mathbf{p}\,|\hat{\varphi}(\mathbf{p})|^2\,(I_{00}+I_{01}) \tag{3.17}$$

$$I_{00} := \int d\mathbf{p}'|\widehat{[\cdot]}_m(\mathbf{p}-\mathbf{p}',\mathbf{p}')|\frac{p^2}{p'^2}\cdot\int d\mathbf{q}\frac{1}{|\mathbf{q}-\mathbf{p}'|+p'+1}|\widehat{[\cdot]}_{k-m+1}(\mathbf{q}-\mathbf{p}',\mathbf{p}')|\frac{p'^2}{q^2}$$

$$I_{01} := \int d\mathbf{p}'\frac{1}{|\mathbf{p}-\mathbf{p}'|+p'+1}|\widehat{[\cdot]}_{k-m+1}(\mathbf{p}-\mathbf{p}',\mathbf{p}')|\frac{p^2}{p'^2}\cdot\int d\mathbf{q}\,|\widehat{[\cdot]}_m(\mathbf{p}'-\mathbf{q},\mathbf{q})|\frac{p'^2}{q^2}$$

where in the term denoted by I_{00}, $|\hat{\phi}_{k-m+1}(\mathbf{p'}-\mathbf{q},\mathbf{q})|$ was estimated by its adjoint before applying (3.16). In I_{01}, the $\mathbf{q}$-integral is by (3.15) estimated by $c\,p'^{\,1-m\epsilon}$. Further one has with $a \geq 0$ and $\delta > 0$

$$\int d\mathbf{q}\;\frac{1}{a+q+1}\;|\widehat{[\cdot]}_n(\mathbf{p}-\mathbf{q},\mathbf{q})|\;\frac{p^2}{q^2}\cdot q^{1-\delta}\;\leq\;p^{-\delta}\int d\mathbf{q}\;|\widehat{[\cdot]}_n(\mathbf{p}-\mathbf{q},\mathbf{q})|\;\frac{p^{2+\delta}}{q^{2+\delta}}$$

$$\leq\;c\,p^{1-n\epsilon-\delta} \tag{3.18}$$

if $1+\frac{\delta}{2}<\frac{3}{2}$. Then with $\delta := m\epsilon$, $\quad I_{01}\;\leq\tilde{c}\,p^{1-(k-m+1)\epsilon}\,p^{-m\epsilon}\;=\;\tilde{c}\,p^{1-(k+1)\epsilon}$. In I_{00} we estimate in the denominator $|\mathbf{q}-\mathbf{p}'|+p'+1$ by p' and subsequently use (3.15) to estimate the $\mathbf{q}$-integral by $c\,p'^{\,-(k-m+1)\epsilon}$. With $\lambda := 1+\frac{k-m+1}{2}\,\epsilon$ (for $(k-m+1)\epsilon<1$) in (3.15) we obtain $I_{00}\;\leq\;\tilde{c}\,p^{-(k-m+1)\epsilon}.p^{1-m\epsilon}\;=\;\tilde{c}\,p^{1-(k+1)\epsilon}$. Therefore

$$|(\varphi,[[\cdot]_m,B_{k-m+1}]\,\varphi)|\;\leq\;c_0\,(\varphi,p^{1-(k+1)\epsilon}\varphi) \tag{3.19}$$

which proves (3.15) for $k+1$. The same estimate can be shown for $|(\varphi,[V,B_k]\,\varphi)|$. Hence

$$|(\varphi,V_{k+1}\varphi)|\;\leq\;c'\,(\varphi,p^{1-(k+1)\epsilon}\varphi)\;\leq c'\,\delta\,\frac{2}{\mu_k}\,|(\varphi,V_k\varphi)|\;+\;C\,(\varphi,\varphi) \tag{3.20}$$

which completes the proof of Proposition 3.1 (ii).

The proof of (i) for $k>1$ is again by induction. Assume V_k is p-form bounded with form bound $c_1<1$. Then we have from (ii)

$$|(\varphi,V_{k+1}\varphi)|\leq\delta|(\varphi,V_k\varphi)|+C(\varphi,\varphi)\leq\delta\,(c_1(\varphi,p\,\varphi)+C_1(\varphi,\varphi))+C(\varphi,\varphi). \tag{3.21}$$

Since δ can be chosen arbitrarily small, one has $\delta c_1<1$. A consequence of (3.21) is the p-form boundedness (with form bound <1) of every finite sum $V_1+...+V_n$.

For the proof of (iii) we have to show that all B_k are bounded operators. Then we can proceed as in section 2.4 to show that a unitary transform $U_k=e^{iB_k\tau}$ preserves the $p^{1-(n+1)\epsilon}$-form boundedness of the commutators of order $n+1$ in the potential strength of which $R^{(n+1)}$ is consisting. Consequently, one has with $(\varphi,p^{1-k\epsilon}\varphi)<\frac{2}{\mu_k}\,|(\varphi,V_k\varphi)|$ for $k=n+1$ and with (3.2),

$$|(\varphi,R^{(n+1)}\,\varphi)|\;\leq\;\mathrm{const}\cdot|(\varphi,V_{n+1}\varphi)|\;\leq\;\mathrm{const}\cdot\delta\,|(\varphi,V_n\varphi)|+C'\,(\varphi,\varphi) \tag{3.22}$$

with $\mathrm{const}\cdot\delta<1$ for a suitably chosen δ. This shows the subordinacy with respect to V_n.

It remains to prove the boundedness of B_k. We will show this directly by using the algebra of symbol estimates. For B_1, from (2.57) with the substitution $\mathbf{q}' := \mathbf{q}-\mathbf{p}$,

$$I_{B_1}\;\leq\;\frac{c}{(2\pi)^{3/2}}\int d\mathbf{q}\;\frac{1}{|\mathbf{p}-\mathbf{q}|^{1+\epsilon}}\;\frac{1}{(|\mathbf{p}-\mathbf{q}|+q+1)^2}\cdot\frac{p^{1-\epsilon}}{q^{1-\epsilon}}$$

$$\leq \frac{c}{(2\pi)^{3/2}} \int d\mathbf{q}' \, \frac{1}{q'^{1+\epsilon}} \, \frac{1}{(q'+1)^2} \cdot \frac{p^{1-\epsilon}}{|\mathbf{p}+\mathbf{q}'|^{1-\epsilon}} \leq c' \tag{3.23}$$

since the integral is finite for $p \to 0$ and for $p \to \infty$ and the singularity of the last factor at $\mathbf{p} = -\mathbf{q}'$ is integrable. The convergence generating function $f(p) = p^{\frac{1-\epsilon}{2}}$ was chosen to allow for a (2.60)-type estimate when showing that the presence of U_k plays no role (but to prove boundedness of I_{B_1}, one can also take $f(p) = 1$).

For B_2, we use the estimate (2.44) of $\hat{\phi}_2$ by $\hat{w}_2$ and recall that W_2 is determined from the commutator $[V, B_1]$. Consider the symbol of VB_1 via (2.17),

$$|\widehat{v\phi}_1(\mathbf{q},\mathbf{p})| \leq \frac{c}{(2\pi)^{3/2}} \int d\mathbf{p}' \, \frac{1}{|\mathbf{q}-\mathbf{p}'|^{2+\epsilon}} \cdot \frac{1}{p'^{1+\epsilon}(p'+p+1)^2}. \tag{3.24}$$

It is found that $|\widehat{v\phi}_1(\mathbf{q},\mathbf{p})| = \text{const}$ for $p = 0$, $\sim 1/p^2$ for $p \to \infty$ and $\sim 1/q^{2+\epsilon}$ for $q \to \infty$ while it diverges for $q \to 0$. The behaviour near $q = 0$ is obtained by performing the angular integration with the help of a (2.34)-type formula such that one gets for $q \neq 0$, $\epsilon \neq 0$,

$$|\widehat{v\phi}_1(\mathbf{q},\mathbf{p})| \leq \frac{\tilde{c}}{q} \int_0^\infty \frac{dp'}{p'^\epsilon} \, \frac{1}{(p'+p+1)^2} \left(\frac{1}{|q-p'|^\epsilon} - \frac{1}{|q+p'|^\epsilon} \right). \tag{3.25}$$

Since the divergence at $q = 0$ results from the behaviour of the integral near $p' = 0$, it is sufficient to reduce the integration region to $[0,1]$ and estimate $(p'+p+1)^{-2} \leq 1$. The resulting integral can be performed analytically with the help of hypergeometric functions [7, p.284], and it behaves $\sim q^{1-2\epsilon}$ for $q \to 0$. $B_1 V$ is in the same operator class such that we obtain

$$|\hat{w}_2(\mathbf{q},\mathbf{p})| \leq c \, \frac{1+q^\epsilon}{q^{2\epsilon}(q+p+1)^2}. \tag{3.26}$$

By induction, one can show that for $k > 2$, one has $|\hat{w}_k(\mathbf{q},\mathbf{p})| \leq \frac{c}{(q+p+1)^{2+\epsilon}}$. Thus one obtains regularisation upon increasing k, resulting in bounded operators B_k, $k > 1$. $\square$

Proposition 3.1 provides justification for representing the transformed Dirac operator in terms of a series expansion in the potential strength. Note, however, that the limit $\epsilon \to 0$ cannot be carried out since in (3.7), $\frac{\epsilon}{1-n\epsilon} \, \delta^{-\frac{1-n\epsilon}{\epsilon}} \to \infty$ as $\epsilon \to 0$, which implies $C \to \infty$ in (3.3). Therefore, this limit cannot be used to prove for the Coulomb potential the p-form boundedness of V_k, $1 \leq k \leq n$ with form bound < 1. On the other hand, it has been shown with different tools that this property holds for V_1 and $V_1 + V_2$ in case of the Coulomb field [2, 8].

4 Proof of Theorem 1.2

The link between the two transformation schemes under consideration is provided by the following lemma.

Lemma 4.1. *Let* $H = D_0 + V$ *and* $U_j = e^{iB_j}$, $j = 1, ..., n$ *be the transformation scheme from section 2.1, where the potential term of* k*-th order in* γ *is decomposed into* $V_k + W_k$. *Let* U_0', $U_j' = e^{-iS_j}$, $j = 1, ..., n$ *be the Douglas-Kroll transformation scheme with the decomposition* $\mathcal{E}_k + \mathcal{O}_k$ *of the* k*-th order potential term. Then one has the identification*

$$\beta E_p = U_0' \, D_0 \, U_0^{'-1}, \qquad \mathcal{E}_k = U_0' \, V_k \, U_0^{'-1}, \qquad \mathcal{O}_k = U_0' \, W_k \, U_0^{'-1}$$

$$S_k = U_0' \, B_k \, U_0^{'-1}, \qquad k = 1, ..., n \tag{4.1}$$

with U_0' *from (1.5).*

The key observation is the relation between the spinor $\psi = \binom{u}{0}$ and the spinor φ in the positive spectral subspace of D_0,

$$\varphi = U_0^{'-1} \, \psi. \tag{4.2}$$

In momentum space, this equation is easily verified from $U_0^{'-1} = (\frac{\boldsymbol{\alpha}\mathbf{p}}{E_p+m}\beta + 1)\, A$ and from the explicit form (1.12) of $\hat{\varphi}(\mathbf{p})$. Then with the help of (2.6), the assertion (1.16) of Theorem 2.1 reads

$$\left(\varphi, \, (D_0 + \sum_{k=1}^{n} V_k) \, \varphi\right) = \left(U_0^{'-1} \, \psi, \sum_{k=0}^{n} H_k \, U_0^{'-1} \, \psi\right) = (\psi, H_n' \, \psi). \tag{4.3}$$

Identifying terms of fixed order $k \leq n$ and using (1.11), the assertion (4.3) is a consequence of

$$U_0' \, V_k \, U_0^{'-1} = \mathcal{E}_k, \qquad k = 1, 2... \quad \text{and} \quad U_0' \, D_0 \, U_0^{'-1} = \beta E_p \tag{4.4}$$

and hence of Lemma 4.1.

Proof of Lemma 4.1.

a) Verification of (4.1) up to first order in γ

The equality $U_0' D_0 U_0^{'-1} = \beta E_p$ is a consequence of (1.7) for zero potential. By means of explicit calculation (which for the sake of simplicity is only presented for the massless case $m = 0$), one gets from (2.18)

$$U_0' W_1 U_0^{'-1} = \frac{1}{\sqrt{2}} \left(1 + \beta \frac{\boldsymbol{\alpha}\mathbf{p}}{p}\right) \cdot \frac{1}{2} \left(V - \frac{\boldsymbol{\alpha}\mathbf{p}}{p} V \frac{\boldsymbol{\alpha}\mathbf{p}}{p}\right) \cdot \frac{1}{\sqrt{2}} \left(1 + \frac{\boldsymbol{\alpha}\mathbf{p}}{p} \beta\right)$$

$$= \frac{1}{2} \left[\beta \, \frac{\boldsymbol{\alpha}\mathbf{p}}{p} \, V \; - \; \beta V \, \frac{\boldsymbol{\alpha}\mathbf{p}}{p}\right] = \mathcal{O}_1 \tag{4.5}$$

and similarly, $U_0' V_1 U_0^{'-1} = \mathcal{E}_1$ with $V_1 = \frac{1}{2} \left(V + \frac{\boldsymbol{\alpha}\mathbf{p}}{p} \, V \, \frac{\boldsymbol{\alpha}\mathbf{p}}{p}\right)$. For the $m \neq 0$ case, one needs the relation $\frac{m}{E_p} + \frac{p^2}{E_p(E_p+m)} = 1$.

In order to prove $S_1 = U_0' B_1 U_0'^{-1}$ we first show that S_1 is uniquely determined by (1.10). Representing S_1 and $\mathcal{O}_1$ by their respective symbols s_1 and o_1 via (2.14) and noting that βE_p is a multiplication operator in Fourier space, one obtains from (1.10)

$$-\hat{o}_1(\mathbf{q},\mathbf{p}) \;=\; i\beta E_{|\mathbf{p}+\mathbf{q}|}\,\hat{s}_1(\mathbf{q},\mathbf{p}) \;-\; i\hat{s}_1(\mathbf{q},\mathbf{p})\,\beta E_p$$

$$= \; i\beta\left(E_{|\mathbf{p}+\mathbf{q}|} \;+\; E_p\right)\hat{s}_1(\mathbf{q},\mathbf{p}) \tag{4.6}$$

which can be uniquely solved for $\hat{s}_1(\mathbf{q},\mathbf{p})$. We now transform the defining equation (2.8) for B_1 with U_0'

$$U_0'\,W_1\,U_0'^{-1} \;=\; -i\,(U_0'D_0U_0'^{-1})(U_0'B_1U_0'^{-1}) \;+\; i\,(U_0'B_1U_0'^{-1})(U_0'D_0U_0'^{-1})$$

$$\Longleftrightarrow \qquad \mathcal{O}_1 \;=\; -i\,\left[\beta E_p, U_0'B_1\,U_0'^{-1}\right]. \tag{4.7}$$

From the uniqueness of the solution it follows from (4.7) and (1.10) that $U_0'B_1U_0'^{-1} = S_1$ and hence the uniqueness of the operator B_1.

b) Proof of (4.1) by induction for arbitrary order n in γ

We assume that to order $n-1$ the assertion of Lemma 4.1 holds. Then, the relation between the Dirac operators transformed to $n-1$st order, asserted by Theorem 1.2, is also true, i.e. (including the n-th order terms)

$$\beta E_p \;+\; \mathcal{E}_1 \;+\; \mathcal{E}_2 \;+ \ldots +\; \mathcal{E}_{n-1} \;+\; \mathcal{E}_n \;+\; \mathcal{O}_n$$

$$= \; U_0'\;(D_0 \;+\; V_1 \;+ \ldots + V_{n-1} \;+\; V_n \;+\; W_n)\;U_0'^{-1} \tag{4.8}$$

since $\mathcal{E}_n$ and $\mathcal{O}_n$ only depend on $\beta E_p,\ \mathcal{E}_j,\ \mathcal{O}_j,\ S_j,\ \ j = 1,\ldots,n-1,$ with the identical dependence of V_n and W_n on $D_0,\ V_j,\ W_j,\ B_j,\ \ j = 1,\ldots,n-1$. From (4.1) for $j = 1,\ldots,n-1$ it therefore follows that $\mathcal{E}_n = U_0'V_nU_0'^{-1}$ and $\mathcal{O}_n = U_0'W_nU_0'^{-1}$. Carrying out the n-th transformation one gets

$$U_n' \cdots U_0'HU_0'^{-1} \cdots U_n'^{-1} = \beta E_p + \mathcal{E}_1 + \ldots + \mathcal{E}_{n-1} + \mathcal{E}_n + \mathcal{O}_n + i[\beta E_p, S_n] + R_{n+1}$$

$$U_n^* \cdots U_1^* H U_1 \cdots U_n = D_0 + V_1 + \ldots + V_{n-1} + V_n + W_n + i[D_0, B_n] + R^{(n+1)}. \tag{4.9}$$

B_n is obtained from $W_n = -i\,[D_0, B_n]$, or transformed with U_0',

$$U_0'\,W_n\,U_0'^{-1} \;=\; \mathcal{O}_n \;=\; -iU_0'\,[D_0, B_n]\,U_0'^{-1} \;=\; -i\,[\beta E_p, U_0'B_nU_0'^{-1}]. \tag{4.10}$$

Since the solution S_n to $\mathcal{O}_n = -i[\beta E_p, S_n]$ is unique, one gets $U_0'B_nU_0'^{-1} = S_n$. $\square$

From the correspondence of the two transformation schemes it follows that for $n = 2$, $H^{(2)}$ from (2.12) when acting on the positive spectral subspace of D_0, reduces to $\Lambda_+(D_0 + V + \frac{i}{2}\,[W_1, B_1])\Lambda_+$ since $i[V_1, B_1]$ corresponds to an odd operator which vanishes upon projection.

Acknowledgment

I should like to thank H. Kalf and H. Siedentop for enlightening discussions. I am particularly grateful to A. Sobolev for suggesting the unitary transformation sequence applied in this work and for his openness and patience during many discussions. I should also like to thank A. Sobolev and the University of Sussex for their hospitality during my visits in the course of this work. Partial support by the Volkswagen Foundation is gratefully acknowledged.

References

[1] G. E. Brown and D. G. Ravenhall. On the interaction of two electrons. *Proc. Roy. Soc. London A,* 208: 552-559, 1952.

[2] R. Brummelhuis, H. Siedentop, and E. Stockmeyer. The ground state energy of relativistic one-electron atoms according to Jansen and Hess. *Doc. Math.,* 7: 167-182, 2002.

[3] N. Datta, R. Fernández and J. Fröhlich. Effective hamiltonians and phase diagrams for tight-binding models. *J. Stat. Phys.,* 96: 545-611, 1999.

[4] M. Douglas and N. M. Kroll. Quantum electrodynamical corrections to the fine structure of helium. *Ann. Phys. (N.Y.),* 82: 89-155, 1974.

[5] W. D. Evans, P. Perry, and H. Siedentop. The spectrum of relativistic one-electron atoms according to Bethe and Salpeter. *Commun. Math. Phys.,* 178: 733-746, 1996.

[6] L. L. Foldy and S. A. Wouthuysen. On the Dirac theory of spin $\frac{1}{2}$ particles and its non-relativistic limit. *Phys. Rev.,* 78: 29-36, 1950.

[7] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products.* Academic Press, New York, 1965.

[8] D. H. Jakubassa-Amundsen. The essential spectrum of relativistic one-electron ions in the Jansen-Hess model. *Math. Phys. Electron. J.,* 8(3): 1-30, 2002

[9] G. Jansen and B. A. Hess. Revision of the Douglas-Kroll transformation. *Phys. Rev. A,* 39: 6016-6017, 1989.

[10] T. Kato. *Perturbation Theory for Linear Operators,* volume 132 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1966.

[11] E. H. Lieb and H.-T. Yau. The stability and instability of relativistic matter. *Commun. Math. Phys.,* 118: 177-213, 1988.

[12] P. R. Halmos and V. S. Sunder. *Bounded Integral Operators on L^2 Spaces.* Springer-Verlag, Berlin, 1976.

[13] A. V. Sobolev. Integrated density of states for the periodic Schrödinger operator in dimension two. *Ann. Henri Poincaré,* 6: 31-84, 2005.

[14] B. Thaller. *The Dirac Equation,* Springer-Verlag, Berlin, 1992.

[15] A. Wolf, M. Reiher, and B. A. Hess. The generalised Douglas-Kroll transformation. *J. Chem. Phys.,* 117: 9215-9226, 2002.

D. H. Jakubassa-Amundsen
Mathematics Institute
University of Munich
Theresienstr. 39
80333 Munich
Germany
doris.jaku@lrz.uni-muenchen.de

Algebraic K-Theory and Sums-of-Squares Formulas

Daniel Dugger, Daniel C. Isaksen

Received: January 20, 2005
Revised: July 25, 2005

Communicated by Stefan Schwede

Abstract. We prove a result about the existence of certain 'sums-of-squares' formulas over a field F. A classical theorem uses topological K-theory to show that if such a formula exists over $\mathbb{R}$, then certain powers of 2 must divide certain binomial coefficients. In this paper we use algebraic K-theory to extend the result to all fields not of characteristic 2.

1. Introduction

Let F be a field. A classical problem asks for which values of r, s, and n does there exist an identity of the form

$$\left(x_1^2+\cdots+x_r^2\right)\left(y_1^2+\cdots+y_s^2\right)=z_1^2+\cdots+z_n^2$$

in the polynomial ring $F[x_1,\ldots,x_r,y_1,\ldots,y_s]$, where the z_i's are bilinear expressions in the x's and y's. Such an identity is called a sums-of-squares formula of type $[\boldsymbol{r},\boldsymbol{s},\boldsymbol{n}]$. For the history of this problem, see the expository papers [L, Sh].

The main theorem of this paper is the following:

Theorem 1.1. *Assume F is not of characteristic 2. If a sums-of-squares formula of type $[r,s,n]$ exists over F, then $2^{\lfloor\frac{s-1}{2}\rfloor-i+1}$ divides $\binom{n}{i}$ for $n-r<i\leq\lfloor\frac{s-1}{2}\rfloor$.*

As one specific application, the theorem shows that a formula of type $[13,13,16]$ cannot exist over any field of characteristic not equal to 2. Previously this had only been known in characteristic zero. (Note that the case $\mathrm{char}(F)=2$, which is not covered by the theorem, is trivial: formulas of type $[r,s,1]$ always exist). In the case $F=\mathbb{R}$, the above theorem was essentially proven by Atiyah [At] as an early application of complex K-theory; the relevance of Atiyah's paper to the sums-of-squares problem was only later pointed out by Yuzvinsky [Y]. The result for characteristic zero fields can be deduced from the case $F=\mathbb{R}$ by an algebraic argument due to K. Y. Lam and T. Y. Lam (see [Sh]). Thus, our contribution is the extension to fields of non-zero characteristic. In this sense the present paper is a natural sequel to [DI], which extended another classical condition about sums-of-squares. We note that sums-of-squares formulas in characteristic p were first seriously investigated in [Ad1, Ad2].

Our proof of Theorem 1.1, given in Section 2, is a modification of Atiyah's original argument. The existence of a sums-of-squares formula allows one to make conclusions about the geometric dimension of certain algebraic vector bundles. A computation of algebraic K-theory (in fact just algebraic K^0), given in Section 3, determines restrictions on what that geometric dimension can be—and this yields the theorem.

Atiyah's result for $F = \mathbb{R}$ is actually slightly better than our Theorem 1.1. The use of topological KO-theory rather than complex K-theory yields an extra power of 2 dividing some of the binomial coefficients. It seems likely that this stronger result holds in non-zero characteristic as well and that it could be proved with Hermitian algebraic K-theory.

1.2. Restatement of the main theorem. The condition on binomial coefficients from Theorem 1.1 can be reformulated in a slightly different way. This second formulation surfaces often, and it's what arises naturally in our proof. We record it here for the reader's convenience. Each of the following observations is a consequence of the previous one:

- By repeated use of Pascal's identity $\binom{c}{d} = \binom{c-1}{d-1} + \binom{c-1}{d}$, the number $\binom{n+i-1}{k+i}$ is a $\mathbb{Z}$-linear combination of the numbers $\binom{n}{k+1}, \binom{n}{k+2}, \ldots, \binom{n}{k+i}$. Similarly, $\binom{n}{k+i}$ is a $\mathbb{Z}$-linear combination of $\binom{n}{k+1}, \binom{n+1}{k+2}, \ldots, \binom{n+i-1}{k+i}$.
- An integer b is a common divisor of $\binom{n}{k+1}, \binom{n}{k+2}, \ldots, \binom{n}{k+i}$ if and only if it is a common divisor of $\binom{n}{k+1}, \binom{n+1}{k+2}, \ldots, \binom{n+i-1}{k+i}$.
- The series of statements
$$2^N \,\Big|\, \binom{n}{k+1}, 2^{N-1} \,\Big|\, \binom{n}{k+2}, \ldots, 2^{N-i+1} \,\Big|\, \binom{n}{k+i}$$
is equivalent to the series of statements
$$2^N \,\Big|\, \binom{n}{k+1}, 2^{N-1} \,\Big|\, \binom{n+1}{k+2}, \ldots, 2^{N-i+1} \,\Big|\, \binom{n+i-1}{k+i}.$$
- If N is a fixed integer, then 2^{N-i+1} divides $\binom{n}{i}$ for $n-r < i \leq N$ if and only if 2^{N-i+1} divides $\binom{r+i-1}{i}$ for $n-r < i \leq N$.

The last observation shows that Theorem 1.1 is equivalent to the theorem below. This is the form in which we'll actually prove the result.

Theorem 1.3. *Suppose that F is not of characteristic 2. If a sums-of-squares formula of type $[r,s,n]$ exists over F, then $2^{\lfloor \frac{s-1}{2} \rfloor - i+1}$ divides the binomial coefficient $\binom{r+i-1}{i}$ for $n-r < i \leq \lfloor \frac{s-1}{2} \rfloor$.*

1.4. Notation. Throughout this paper $K^0(X)$ denotes the Grothendieck group of locally free coherent sheaves on the scheme X. This group is usually denoted $K_0(X)$ in the literature.

2. The main proof

In this section we fix a field F not of characteristic 2. Let q_k be the quadratic form on $\mathbb{A}^k$ defined by $q_k(x) = \sum_{i=1}^k x_i^2$. A sums-of-squares formula of type $[r,s,n]$ gives a bilinear map $\phi\colon \mathbb{A}^r \times \mathbb{A}^s \to \mathbb{A}^n$ such that $q_r(x)q_s(y) = q_n(\phi(x,y))$. We begin with a simple lemma:

Lemma 2.1. *Let $F \hookrightarrow E$ be a field extension, and let $y \in E^s$ be such that $q_s(y) \neq 0$. Then for $x \in E^r$ one has $\phi(x,y) = 0$ if and only if $x = 0$.*

Proof. Let $\langle -,- \rangle$ denote the inner product on E^k corresponding to the quadratic form q_k. Note that the sums-of-squares identity implies that

$$\langle \phi(x,y), \phi(x',y) \rangle = q_s(y)\langle x, x' \rangle$$

for any x and x' in E^r. If one had $\phi(x,y) = 0$ then the above formula would imply that $q_s(y)\langle x, x'\rangle = 0$ for every x'; but since $q_s(y) \neq 0$, this can only happen when $x = 0$. $\square$

Let V_q be the subvariety of $\mathbb{P}^{s-1}$ defined by $q_s(y) = 0$. Let ξ denote the restriction to V_q of the tautological line bundle $\mathcal{O}(-1)$ of $\mathbb{P}^{s-1}$.

Proposition 2.2. *If a sums-of-squares formula of type $[r,s,n]$ exists over F, then there is an algebraic vector bundle ζ on $\mathbb{P}^{s-1} - V_q$ of rank $n - r$ such that*

$$r[\xi] + [\zeta] = n$$

as elements of the Grothendieck group $K^0(\mathbb{P}^{s-1} - V_q)$ of locally free coherent sheaves on $\mathbb{P}^{s-1} - V_q$.

Proof. We'll write $q = q_s$ in this proof, for simplicity. Let $S = F[y_1, \dots, y_s]$ be the homogeneous coordinate ring of $\mathbb{P}^{s-1}$. By [H, Prop. II.2.5(b)] one has $\mathbb{P}^{s-1} - V_q = \operatorname{Spec} R$, where R is the subring of the localization S_q that consists of degree 0 homogeneous elements. The group $K^0(\mathbb{P}^{s-1} - V_q)$ is naturally isomorphic to the Grothendieck group of finitely-generated projective R-modules. Let P denote the subset of S_q consisting of homogeneous elements of degree -1, regarded as a module over R. Then P is projective and is the module of sections of the vector bundle ξ. To see explicitly that P is projective of rank 1, observe that there is a split-exact sequence $0 \to R^{s-1} \to R^s \xrightarrow{\pi} P \to 0$ where $\pi(p_1, \dots, p_s) = \sum p_i \cdot \frac{y_i}{q}$ and the splitting $\chi\colon P \to R^s$ is $\chi(f) = (y_1 f, y_2 f, \dots, y_s f)$.

From our bilinear map $\phi\colon \mathbb{A}^r \times \mathbb{A}^s \to \mathbb{A}^n$ we get linear forms $\phi(e_i, y) \in S^n$ for $1 \leq i \leq r$. Here e_i denotes the standard basis for F^r, and $y = (y_1, \dots, y_s)$ is the vector of indeterminates from S. If f belongs to P, then each component of $f \cdot \phi(e_i, y)$ is homogeneous of degree 0—hence lies in R.

Define a map $\alpha\colon P^r \to R^n$ by

$$(f_1, \dots, f_r) \mapsto f_1\phi(e_1, y) + f_2\phi(e_2, y) + \cdots + f_r\phi(e_r, y).$$

We can write $\alpha(f_1, \dots, f_r) = \phi((f_1, \dots, f_r), y)$, where the expression on the right means to formally substitute each f_i for x_i in the defining formula for ϕ. If $R \to E$ is any map of rings where E is a field, we claim that $\alpha \otimes_R E$ is an injective map $E^r \to E^n$. To see this, note that $R \to E$ may be extended to a map $u\colon S_q \to E$ (any map $\operatorname{Spec} E \to \mathbb{P}^{s-1} - V_q$ lifts to the affine variety $q \neq 0$, as the projection map from the latter to the former is a Zariski locally trivial bundle). One obtains an isomorphism $P \otimes_R E \to E$ by sending $f \otimes 1$ to $u(f)$. Using this, $\alpha \otimes_R E$ may be readily identified with the map $x \mapsto \phi(x, u(y))$. Now apply Lemma 2.1.

Since R is a domain, we may take E to be the quotient field of R. It follows that α is an inclusion. Let M denote its cokernel. The module M will play the role of ζ in the statement of the proposition, so to conclude the proof we only need show that M is projective. An inclusion of finitely-generated projectives $P_1 \hookrightarrow P_2$ has projective cokernel if and only if $P_1 \otimes_R E \to P_2 \otimes_R E$ is injective for every map $R \to E$ where E is a field (that is to say, the map has constant rank on the fibers)—this follows at once using [E, Ex. 6.2(iii),(v)]. As we have already verified this property for α, we are done. $\square$

Remark 2.3. The above algebraic proof hides some of the geometric intuition behind Proposition 2.2. We outline a different approach more in the spirit of [At].

Let $Gr_r(\mathbb{A}^n)$ denote the Grassmannian variety of r-planes in affine space $\mathbb{A}^n$. We claim that ϕ induces a map $f\colon \mathbb{P}^{s-1} - V_q \to Gr_r(\mathbb{A}^n)$ with the following behavior on points. Let $[y]$ be a point of $\mathbb{P}^{s-1}$ represented by a point y of $\mathbb{A}^s$ such that $q_s(y) \neq 0$. Then the map $\phi_y : x \mapsto \phi(x, y)$ is a linear inclusion by Lemma 2.1. Let $f([y])$ be the r-plane that is the image of ϕ_y. Since ϕ is bilinear, we get that $\phi_{\lambda y} = \lambda \cdot \phi_y$ for any scalar y. This shows that $f([y])$ is well-defined. We leave it as an exercise for the reader to carefully construct f as a map of schemes.

The map f has a special property related to bundles. If η_r denotes the tautological r-plane bundle over the Grassmannian, we claim that ϕ induces a map of bundles $\tilde{f}\colon r\xi \to \eta_r$ covering the map f. To see this, note that the points of $r\xi$ (defined over some field E) correspond to equivalence classes of pairs $(y, a) \in \mathbb{A}^s \times \mathbb{A}^r$ with $q(y) \neq 0$, where $(\lambda y, a) \sim (y, \lambda a)$ for any λ in the field. The pair (y, a) gives us a line $\langle y \rangle \subseteq \mathbb{A}^s$ together with r points $a_1 y, a_2 y, \ldots, a_r y$ on the line.

One defines $\tilde{f}$ so that it sends (y, a) to the element of η_r represented by the vector $\phi(a, y)$ lying on the r-plane spanned by $\phi(e_1, y), \ldots, \phi(e_r, y)$. This respects the equivalence relation, as $\phi(\lambda a, y) = \phi(a, \lambda y)$. So we have described our map $\tilde{f}\colon r\xi \to \eta_r$. We again leave it to the reader to construct f as a map of schemes.

One readily checks that $\tilde{f}$ is a linear isomorphism on geometric fibers, using Lemma 2.1. So $\tilde{f}$ gives an isomorphism $r\xi \cong f^*\eta_r$ of bundles on $\mathbb{P}^{s-1} - V_q$.

The bundle η_r is a subbundle of the rank n trivial bundle, which we denote by n. Consider the quotient n/η_r, and set $\zeta = f^*(n/\eta_r)$. Since $n = [\eta_r] + [n/\eta_r]$ in $K^0(Gr_r(\mathbb{A}^n))$, application of f^* gives $n = [f^*\eta_r] + [\zeta]$ in $K^0(\mathbb{P}^{s-1} - V_q)$. Now recall that $f^*\eta_r \cong r\xi$. This gives the desired formula in Proposition 2.2.

The next task is to compute the Grothendieck group $K^0(\mathbb{P}^{s-1} - V_q)$. This becomes significantly easier if we assume that F contains a square root of -1. The reason for this is made clear in the next section.

Proposition 2.4. *Suppose that F contains a square root of -1 and is not of characteristic 2. Let $c = \lfloor \frac{s-1}{2} \rfloor$. Then $K^0(\mathbb{P}^{s-1} - V_q)$ is isomorphic to $\mathbb{Z}[\nu]/(2^c\nu, \nu^2 = -2\nu)$, where $\nu = [\xi] - 1$ generates the reduced Grothendieck group $\tilde{K}^0(\mathbb{P}^{s-1} - V_q) \cong \mathbb{Z}/2^c$.*

The proof of the above result will be deferred until the next section. Note that $K^0(\mathbb{P}^{s-1} - V_q)$ has the same form as the complex K-theory of real projective space $\mathbb{R}P^{s-1}$ [A, Thm. 7.3]. To complete the analogy, we point out that when $F = \mathbb{C}$ the space $\mathbb{C}P^{s-1} - V_q(\mathbb{C})$ is actually homotopy equivalent to $\mathbb{R}P^{s-1}$ [Lw, 6.3]. We also mention that for the special case where F is contained in $\mathbb{C}$, the above proposition was proved in [GR, Theorem, p. 303].

By accepting the above proposition for the moment, we can finish the

Proof of Theorem 1.3. Recall that one has operations γ^i on $\tilde{K}^0(X)$ for any scheme X [SGA6, Exp. V] (see also [AT] for a very clear explanation). If $\gamma_t = 1 + \gamma^1 t + \gamma^2 t^2 + \cdots$ denotes the generating function, then their basic properties are:

(i) $\gamma_t(a + b) = \gamma_t(a)\gamma_t(b)$.

(ii) For a line bundle L on X one has $\gamma_t([L] - 1) = 1 + t([L] - 1)$.

(iii) If E is an algebraic vector bundle on X of rank k then $\gamma^i([E] - k) = 0$ for $i > k$.

The third property follows from the preceding two via the splitting principle.

If a sums-of-squares identity of type $[r, s, n]$ exists over a field F, then it also exists over any field containing F. So we may assume F contains a square root of -1. If we write $X = \mathbb{P}^{s-1} - V_q$, then by Proposition 2.2 there is a rank $n - r$ bundle ζ on X such that $r[\xi] + [\zeta] = n$ in $K^0(X)$. This may also be written as $r([\xi] - 1) + ([\zeta] - (n - r)) = 0$ in $\tilde{K}^0(X)$. Setting $\nu = [\xi] - 1$ and applying the operation γ_t we have

$$\gamma_t(\nu)^r \cdot \gamma_t([\zeta] - (n - r)) = 1$$

or

$$\gamma_t([\zeta] - (n - r)) = \gamma_t(\nu)^{-r} = (1 + t\nu)^{-r}.$$

The coefficient of t^i on the right-hand-side is $(-1)^i\binom{r+i-1}{i}\nu^i$, which is the same as $-2^{i-1}\binom{r+i-1}{i}\nu$ using the relation $\nu^2 = -2\nu$. Finally, since ζ has rank $n - r$ we know that $\gamma^i([\zeta] - (n - r)) = 0$ for $i > n - r$. In light of Proposition 2.4, this means that 2^c divides $2^{i-1}\binom{r+i-1}{i}$ for $i > n - r$, where $c = \lfloor \frac{s-1}{2} \rfloor$. When $i - 1 < c$, we can rearrange the powers of 2 to conclude that 2^{c-i+1} divides $\binom{r+i-1}{i}$ for $n - r < i \leq c$. $\square$

3. K-theory of deleted quadrics

The rest of the paper deals with the K-theoretic computation stated in Proposition 2.4. This computation is entirely straightforward, and could have been done in the 1970's. We do not know of a reference, however.

Let $Q_{n-1} \hookrightarrow \mathbb{P}^n$ be the split quadric defined by one of the equations

$$a_1b_1+\cdots+a_kb_k = 0\ (n=2k-1) \qquad \text{or} \qquad a_1b_1+\cdots+a_kb_k+c^2 = 0\ (n=2k).$$

Beware that in general Q_{n-1} is not the same as the variety V_q of the previous section. However, if F contains a square root i of -1 then one can write $x^2+y^2=(x+iy)(x-iy)$. After a change of variables the quadric V_q becomes isomorphic to Q_{n-1}. These 'split' quadrics Q_{n-1} are simpler to compute with, and we can analyze the K-theory of these varieties even if F does not contain a square root of -1.

Write $DQ_n = \mathbb{P}^n - Q_{n-1}$, and let ξ be the restriction to DQ_n of the tautological line bundle $\mathcal{O}(-1)$ of $\mathbb{P}^n$. In this section we calculate $K^0(DQ_n)$ over any ground field F not of characteristic 2. Proposition 2.4 is an immediate corollary of this more general result:

Theorem 3.1. *Let F be a field of characteristic not 2. The ring $K^0(DQ_n)$ is isomorphic to $\mathbb{Z}[\nu]/(2^c\nu, \nu^2=-2\nu)$, where $\nu=[\xi]-1$ generates the reduced group $\tilde{K}^0(DQ_n)\cong \mathbb{Z}/2^c$ and $c=\lfloor \frac{n}{2}\rfloor$.*

Remark 3.2. We remark again that we are writing $K^0(X)$ for what is usually denoted $K_0(X)$ in the algebraic K-theory literature. We prefer this notation partly because it helps accentuate the relationship with topological K-theory.

3.3. Basic facts about K-theory. Let X be a scheme. As usual $K^0(X)$ denotes the Grothendieck group of locally free coherent sheaves, and $G_0(X)$ (also called $K_0'(X)$) is the Grothendieck group of coherent sheaves [Q, Section 7]. Topologically speaking, $K^0(-)$ is the analog of the usual complex K-theory functor $KU^0(-)$ whereas G_0 is something like a Borel-Moore version of *KU-homology*.

Note that there is an obvious map $\alpha\colon K^0(X)\to G_0(X)$ coming from the inclusion of locally free coherent sheaves into all coherent sheaves. When X is nonsingular, α is an isomorphism whose inverse $\beta\colon G_0(X)\to K^0(X)$ is constructed in the following way [H, Exercise III.6.9]. If $\mathcal{F}$ is a coherent sheaf on X, there exists a resolution

$$0\to \mathcal{E}_n\to\cdots\to\mathcal{E}_0\to\mathcal{F}\to 0$$

in which the $\mathcal{E}_i$'s are locally free and coherent. One defines $\beta(\mathcal{F})=\sum_i(-1)^i[\mathcal{E}_i]$. This does not depend on the choice of resolution, and now $\alpha\beta$ and $\beta\alpha$ are obviously the identities. This is 'Poincare duality' for K-theory.

Since we will only be dealing with smooth schemes, we are now going to blur the distinction between G_0 and K^0. If $\mathcal{F}$ is a coherent sheaf on X, we will write $[\mathcal{F}]$ for the class that it represents in $K^0(X)$, although we more literally mean $\beta([\mathcal{F}])$. As an easy exercise, check that if $i\colon U\hookrightarrow X$ is an open immersion then the image of $[\mathcal{F}]$ under $i^*\colon K^0(X) \to K^0(U)$ is the same as $[\mathcal{F}\,|_U]$. We will use this fact often.

If $j\colon Z \hookrightarrow X$ is a smooth embedding and $i\colon X - Z \hookrightarrow X$ is the complement, there is a Gysin sequence [Q, Prop. 7.3.2]

$$\cdots \to K^{-1}(X-Z) \longrightarrow K^0(Z) \xrightarrow{j_!} K^0(X) \xrightarrow{i^*} K^0(X-Z) \longrightarrow 0.$$

(Here $K^{-1}(X-Z)$ denotes the group usually called $K_1(X-Z)$, and i^* is surjective because X is regular). The map $j_!$ is known as the Gysin map. If $\mathcal{F}$ is a coherent sheaf, then $j_!([\mathcal{F}])$ equals the class of its pushforward $j_*(\mathcal{F})$ (also known as extension by zero). Note that the pushforward of coherent sheaves is exact for closed immersions.

3.4. BASIC FACTS ABOUT $\mathbb{P}^n$. If Z is a degree d hypersurface in $\mathbb{P}^n$, then the structure sheaf $\mathcal{O}_Z$ can be pushed forward to $\mathbb{P}^n$ along the inclusion $Z \to \mathbb{P}^n$; we will still write this pushforward as $\mathcal{O}_Z$. It has a very simple resolution of the form $0 \to \mathcal{O}(-d) \to \mathcal{O} \to \mathcal{O}_Z \to 0$, where $\mathcal{O}$ is the trivial rank 1 bundle on $\mathbb{P}^n$ and $\mathcal{O}(-d)$ is the d-fold tensor power of the tautological line bundle $\mathcal{O}(-1)$ on $\mathbb{P}^n$. So $[\mathcal{O}_Z]$ equals $[\mathcal{O}] - [\mathcal{O}(-d)]$ in $K^0(\mathbb{P}^n)$. From now on we'll write $[\mathcal{O}] = 1$.

Now suppose that $Z \hookrightarrow \mathbb{P}^n$ is a complete intersection, defined by the regular sequence of homogeneous equations $f_1, \ldots, f_r \in k[x_0, \ldots, x_n]$. Let f_i have degree d_i. The module $k[x_0, \ldots, x_n]/(f_1, \ldots, f_r)$ is resolved by the Koszul complex, which gives a locally free resolution of $\mathcal{O}_Z$. It follows that

$$[\mathcal{O}_Z] = \bigl(1 - [\mathcal{O}(-d_1)]\bigr)\bigl(1 - [\mathcal{O}(-d_2)]\bigr)\cdots\bigl(1 - [\mathcal{O}(-d_r)]\bigr) \tag{3.4}$$

in $K^0(\mathbb{P}^n)$. In particular, note that for a linear subspace $\mathbb{P}^i \hookrightarrow \mathbb{P}^n$ one has

$$[\mathcal{O}_{\mathbb{P}^i}] = \Bigl(1 - [\mathcal{O}(-1)]\Bigr)^{n-i}$$

because $\mathbb{P}^i$ is defined by $n-i$ linear equations.

One can compute that $K^0(\mathbb{P}^n) \cong \mathbb{Z}^{n+1}$, with generators $[\mathcal{O}_{\mathbb{P}^0}], [\mathcal{O}_{\mathbb{P}^1}], \ldots, [\mathcal{O}_{\mathbb{P}^n}]$ (see [Q, Th. 8.2.1], as one source). If $t = 1 - [\mathcal{O}(-1)]$, then the previous paragraph tells us that $K^0(\mathbb{P}^n) \cong \mathbb{Z}[t]/(t^n)$ as rings. Here t^k corresponds to $[\mathcal{O}_{\mathbb{P}^{n-k}}]$.

3.5. COMPUTATIONS. Let $n = 2k$. Recall that Q_{2k-1} denotes the quadric in $\mathbb{P}^{2k}$ defined by $a_1b_1 + \cdots + a_kb_k + c^2 = 0$. The Chow ring $\mathrm{CH}^*(Q_{2k-1})$ consists of a copy of $\mathbb{Z}$ in every dimension (see [DI, Appendix A] or [HP, XIII.4–5], for example). The generators in dimensions k through $2k-1$ are represented by subvarieties of Q_{2k-1} which correspond to linear subvarieties $\mathbb{P}^{k-1}, \mathbb{P}^{k-2}, \ldots, \mathbb{P}^0$ under the embedding $Q_{2k-1} \hookrightarrow \mathbb{P}^{2k}$. In terms of equations, $\mathbb{P}^{k-i}$ is defined by $c = b_1 = \cdots = b_k = 0$ together with $0 = a_k = a_{k-1} = \cdots = a_{k-i+2}$. The generators of the Chow ring in degrees 0 through $k-1$ are represented by subvarieties $Z_i \hookrightarrow \mathbb{P}^{2k}$ ($k \le i \le 2k-1$), where Z_i is defined by the equations

$$0 = b_1 = b_2 = \cdots = b_{2k-1-i}, \qquad a_1b_1 + \cdots + a_kb_k + c^2 = 0.$$

Note that $Z_{2k-1} = Q_{2k-1}$.

The following result is proven in [R, pp. 128-129] (see especially the first paragraph on page 129):

PROPOSITION 3.6. *The group $K^0(Q_{2k-1})$ is isomorphic to $\mathbb{Z}^{2k}$, with generators $[\mathcal{O}_{\mathbb{P}^0}], \dots, [\mathcal{O}_{\mathbb{P}^{k-1}}]$ and $[\mathcal{O}_{Z_k}]$, ..., $[\mathcal{O}_{Z_{2k-1}}]$.*

It is worth noting that to prove Theorem 3.1 we don't actually need to know that $K^0(Q_{2k-1})$ is free—all we need is the list of generators.

Proof of Theorem 3.1 when n is even. Set $n = 2k$. To calculate $K^0(DQ_{2k})$ we must analyze the localization sequence

$$\cdots \to K^0(Q_{2k-1}) \xrightarrow{j_!} K^0(\mathbb{P}^{2k}) \to K^0(DQ_{2k}) \to 0.$$

The image of $j_! : K^0(Q_{2k-1}) \to K^0(\mathbb{P}^{2k})$ is precisely the subgroup generated by $[\mathcal{O}_{\mathbb{P}^0}], \dots, [\mathcal{O}_{\mathbb{P}^{k-1}}]$ and $[\mathcal{O}_{Z_k}]$, ..., $[\mathcal{O}_{Z_{2k-1}}]$. Since $\mathbb{P}^i$ is a complete intersection defined by $2k - i$ linear equations, formula (3.4) tells us that $[\mathcal{O}_{\mathbb{P}^i}] = t^{2k-i}$ for $0 \le i \le k - 1$.
Now, Z_{2k-1} is a degree 2 hypersurface in $\mathbb{P}^{2k}$, and so $[\mathcal{O}_{Z_{2k-1}}]$ equals $1-[\mathcal{O}(-2)]$. Note that

$$1 - [\mathcal{O}(-2)] = 2(1 - [\mathcal{O}(-1)]) - (1 - [\mathcal{O}(-1)])^2 = 2t - t^2.$$

In a similar way one notes that Z_i is a complete intersection defined by $2k-1-i$ linear equations and one degree 2 equation, so formula (3.4) tells us that

$$[\mathcal{O}_{Z_i}] = (1 - [\mathcal{O}(-1)])^{2k-1-i} \cdot (1 - [\mathcal{O}(-2)]) = t^{2k-1-i}(2t - t^2).$$

The calculations in the previous two paragraphs imply that the kernel of the map $K^0(\mathbb{P}^{2k}) \to K^0(DQ_{2k})$ is the ideal generated by $2t-t^2$ and t^{k+1}. This ideal is equal to the ideal generated by $2t - t^2$ and $2^k t$, so $K^0(DQ_{2k})$ is isomorphic to $\mathbb{Z}[t]/(2^k t, 2t - t^2)$. If we substitute $\nu = [\xi] - 1 = -t$, we find $\nu^2 = -2\nu$.
To find $\tilde{K}^0(DQ_{2k})$, we just have to take the additive quotient of $K^0(DQ_{2k})$ by the subgroup generated by 1. This quotient is isomorphic to $\mathbb{Z}/2^k$ and is generated by ν. □

This completes the proof of Theorem 3.1 in the case where n is even. The computation when n is odd is very similar:

Proof of Theorem 3.1 when n is odd. In this case Q_{n-1} is defined by the equation $a_1 b_1 + \cdots + a_k b_k = 0$ with $k = \frac{n+1}{2}$. The Chow ring $\mathrm{CH}^*(Q_{n-1})$ consists of $\mathbb{Z}$ in every dimension except for $k - 1$, which is $\mathbb{Z} \oplus \mathbb{Z}$. The generators are the Z_i's ($k - 1 \le i \le 2k - 2$) defined analogously to before, together with the linear subvarieties $\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^{k-1}$. By [R, pp. 128–129], the group $K^0(Q_{n-1})$ is again free of rank $2k$ on the generators $[\mathcal{O}_{Z_i}]$ and $[\mathcal{O}_{\mathbb{P}^i}]$. One finds that $K^0(DQ_n)$ is isomorphic to $\mathbb{Z}[t]/(2t - t^2, t^k) = \mathbb{Z}[t]/(2t - t^2, 2^{k-1}t)$. Everything else is as before. □

References

[A] J. F. Adams, *Vector fields on spheres*, Ann. of Math. (2) 75 (1962), 603–632.

[Ad1] J. Adem, *On the Hurwitz problem over an arbitrary field I*, Bol. Soc. Mat. Mexicana (2) 25 (1980), no. 1, 29–51.

[Ad2] J. Adem, *On the Hurwitz problem over an arbitrary field II*, Bol. Soc. Mat. Mexicana (2) 26 (1981), no. 1, 29–41.

[At] M. F. Atiyah, *Immersions and embeddings of manifolds*, Topology 1 (1962), 125–132.

[AT] M. F. Atiyah and D. O. Tall, *Group representations, λ-rings and the J-homomorphism*, Topology 8 (1969), 253–297.

[DI] D. Dugger and D. C. Isaksen, *The Hopf condition for bilinear forms over arbitrary fields*, preprint, 2003.

[E] D. Eisenbud, *Commutative algebra, with a view toward algebraic geometry*, Graduate Texts in Mathematics 150, Springer, 1995.

[GR] A. V. Geramita and L. G. Roberts, *Algebraic vector bundles on projective space*, Invent. Math. 10 (1970), 298–304.

[H] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics 52, Springer, 1977.

[HP] W. V. D. Hodge and D. Pedoe, *Methods of algebraic geometry*, Vol. II, Cambridge University Press, 1952.

[L] K. Y. Lam, *Topological methods for studying the composition of quadratic forms*, Quadratic and Hermitian Forms, (Hamilton, Ont., 1983), pp. 173–192, Canadian Mathematical Society Conference Proceedings 4, Amer. Math. Soc., 1984.

[Lw] P. S. Landweber, *Fixed point free conjugations on complex manifolds*, Ann. Math. (2) 86 (1967), 491–502.

[Q] D. Quillen, *Higher algebraic K-theory I*, Algebraic K-theory, I: Higher K-theories (Proc. Conf. Battelle Memorial Inst., Seattle, 1972), pp. 85–147, Lecture Notes in Mathematics 341, Springer, 1973.

[R] L. Roberts, *Base change for K_0 of algebraic varieties*, Algebraic K-theory, II: "Classical" algebraic K-theory and connections with arithmetic (Proc. Conf., Battelle Memorial Inst., Seattle, 1972), pp. 122–134, Lecture Notes in Mathematics 342, Springer, 1973.

[SGA6] *Théorie des intersections et théorème de Riemann-Roch*, Séminaire de Géométrie Algébrique du Bois-Marie 1966–1967 (SGA 6), by P. Berthelot, A. Grothendieck, and L. Illusie, Lecture Notes in Mathematics 225, Springer, 1971.

[Sh] D. B. Shapiro, *Products of sums of squares*, Expo. Math. 2 (1984), 235–261.

[Y] S. Yuzvinsky, *Orthogonal pairings of Euclidean spaces*, Michigan Math. J. 28 (1981), 131–145.

Daniel Dugger
Department of Mathematics
University of Oregon
Eugene, OR 97403, USA
ddugger@math.uoregon.edu

Daniel C. Isaksen
Department of Mathematics
Wayne State University
Detroit, MI 48202, USA
isaksen@math.wayne.edu

Some Remarks on Morphisms Between Fano Threefolds: Erratum

cf. Documenta Math. 9, p. 471–486

Ekaterina Amerik

Received: July 5, 2005

Communicated by Thomas Peternell

Abstract. It was pointed out to me by James McKernan that there is a mistake in the proof of Proposition 2.7. Namely, it is stated (and "used") there that the complement to a smooth conic in $\mathbb{P}^2$ is simply-connected. This is obviously false; indeed, the fundamental group is cyclic of order two. This mistake is, however, easy to deal with.

2000 Mathematics Subject Classification: 14J45
Keywords and Phrases: Fano threefolds, connectedness

It was pointed out to me by James McKernan that there is a mistake in the proof of Proposition 2.7. Namely, it is stated (and "used") there that the complement to a smooth conic in $\mathbb{P}^2$ is simply-connected. This is obviously false; indeed, the fundamental group is cyclic of order two.

This mistake is, however, easy to deal with. What I propose below is the corrected version of the end of the proof of Proposition 2.7, starting from the line 20. Instead of "But the latter is simply-connected..." and so forth, read: "Remark that the fundamental group of the latter is cyclic of order two. In the case $deg(X) = 4$ and $m = 5$, the degree of p is 10; this means that H has at least 5 irreducible components. Each of them gives at least one conic through a general point of X. These conics are mapped to different lines on V_5, because they intersect. This is a contradiction because on V_5 there are only three lines through a general point.

In the other case $deg(X) = 10$ and $m = 2$, the degree of p is 4, and this means that H cannot be irreducible. To derive a contradiction, we need another easy observation, which is made for example in [HM] (Proposition 12). It says that (thanks to the fact that $Pic(X) = \mathbb{Z}$ and that the normal bundle of a general conic on X is trivial) for H_0 an irreducible component of H and $\mathcal{C}_0$ the universal family of conics over H_0, the evaluation map $\mathcal{C}_0 \to X$ cannot be generically one-to-one. So in fact each irreducible component of H provides at least two conics through a general point of X and therefore at least two lines through a general point of V_5; H must thus be irreducible, a contradiction."

References

[HM] J.M. Hwang, N. Mok, Finite morphisms onto Fano manifolds of Picard number 1 which have rational curves with trivial normal bundles, J. Algebraic Geom. 12 (2003), no. 4, 627–651.

Ekaterina Amerik
Université Paris-Sud
Laboratoire des Mathématiques
Bâtiment 425
91405 Orsay
France
Ekaterina.Amerik@math.u-psud.fr

Moduli Schemes of Generically Simple Azumaya Modules

Norbert Hoffmann, Ulrich Stuhler

Received: December 16, 2004

Communicated by Ulf Rehmann

Abstract. Let A be an Azumaya algebra over a smooth projective variety X or more generally, a torsion free coherent sheaf of algebras over X whose generic fiber is a central simple algebra. We show that generically simple torsion free A-module sheaves have a projective coarse moduli scheme; it is smooth and even symplectic if X is an abelian or K3 surface and A is Azumaya. We explain a relation to the classical theory of the Brandt groupoid.

2000 Mathematics Subject Classification: 14J60 (Primary), 16H05
Keywords and Phrases: Moduli space, torsion free sheaf, Azumaya module, Brandt groupoid

Introduction

Let X be a smooth projective variety, e. g. a surface, over an algebraically closed field k. Let $\mathcal{A}$ be a sheaf of Azumaya algebras over X or more generally, a torsion free coherent sheaf of algebras over X whose generic fiber $\mathcal{A}_\eta$ is a central simple algebra over the function field of X. This paper is about moduli schemes of generically simple, locally projective $\mathcal{A}$-module sheaves E.

These moduli schemes are in close analogy to the Picard variety of X. In fact, our main result says that we do not need any stability condition for our sheaves E to construct coarse moduli schemes parameterizing them, say with fixed Hilbert polynomial or Chern classes. We find that these schemes are in general not proper over k, but they have natural compactifications: Working with torsion free sheaves E instead of only locally projective ones, we obtain projective moduli schemes.

This gives lots of interesting moduli spaces, which certainly deserve further study. For example, we show in section 3 that they are smooth projective and even symplectic if X is an abelian or K3 surface and $\mathcal{A}$ is an Azumaya algebra. They are also related to classifying isomorphism types of Azumaya algebras

$\mathcal{A}$ in a given central division algebra $\mathcal{A}_\eta = D$, a topic already present in the classical literature on algebras; this relation is explained in section 4.
We construct these moduli schemes in section 2. We use standard techniques from geometric invariant theory (GIT) and a boundedness result, which has been known for some time in the case of characteristic $\mathrm{char}(k) = 0$, but is one of the deep results of A. Langer in [9] for $\mathrm{char}(k) = p > 0$. Our construction works for any integral projective scheme X over k; the precise setup is formulated in section 1. As M. Lieblich has pointed out to us, this construction can also be seen as a special case of Simpson's general result [16, Theorem 4.7], at least if $\mathrm{char}(k) = 0$; for $\mathrm{char}(k) = p > 0$ see also [10].
In section 3, we apply deformation theory to our $\mathcal{A}$-modules, mainly in the case where X is a surface and $\mathcal{A}$ is an Azumaya algebra. Besides the smoothness mentioned above, we also show here that torsion free sheaves are really necessary to obtain projective moduli schemes, because locally projective sheaves of $\mathcal{A}$-modules can degenerate to torsion free ones.
During the final preparations of this paper, we were informed about the MIT-thesis of M. Lieblich [13, 12]. This thesis works much more systematically and abstractly and contains several results similar to ours in the language of algebraic stacks. We thank A. J. de Jong and M. Lieblich for informations concerning their work. Similar results have also been obtained independently by K. Yoshioka [17]; they have been used by D. Huybrechts and P. Stellari [8] to prove a conjecture of Caldararu. We thank F. Heinloth, J. Heinloth, Y. Holla and A. Langer for useful comments and discussions.

1 Families of $\mathcal{A}$-modules

Let X be an integral projective scheme over the algebraically closed field k. Throughout this paper, $\mathcal{A}$ denotes a sheaf of associative $\mathcal{O}_X$-algebras satisfying the following properties:

1. As a sheaf of $\mathcal{O}_X$-modules, $\mathcal{A}$ is coherent and torsion free.
2. The stalk $\mathcal{A}_\eta$ of $\mathcal{A}$ over the generic point $\eta \in X$ is a central simple algebra over the function field $F = k(X) = \mathcal{O}_{X,\eta}$.

For example, X could be a smooth projective variety over k, and $\mathcal{A}$ could be a sheaf of Azumaya algebras over X.

Remark 1.1. If $\dim X = 1$, then $\mathcal{A}_\eta$ is a matrix algebra over $k(X)$ by Tsen's theorem. So the first interesting case is $\dim X = 2$.

Our main objects will be generically simple torsion free $\mathcal{A}$-modules, i. e. sheaves E of left $\mathcal{A}$-modules over X which are torsion free and coherent as $\mathcal{O}_X$-modules and whose generic fiber E_η is a simple module over the central simple algebra $\mathcal{A}_\eta$. By Wedderburn's structure theorem, we have $A_\eta \cong \mathrm{Mat}(n \times n; D)$ for a division algebra D, say of dimension r^2 over $k(X)$; that E_η is simple means that

it is Morita equivalent to a one-dimensional vector space over D. In particular, E has rank $r^2 n$ over $\mathcal{O}_X$.
Note that any such $\mathcal{A}$-module E has only scalar endomorphisms: Indeed, $\mathrm{End}_{\mathcal{A}}(E)$ is a finite-dimensional k-algebra; it has no zero-divisors because it embeds into the division algebra $\mathrm{End}_{\mathcal{A}_\eta}(E_\eta) \cong D^{\mathrm{op}}$. This implies $\mathrm{End}_{\mathcal{A}}(E) = k$ because k is algebraically closed.

Lemma 1.2. *Suppose that $k \subseteq K$ is a field extension, and let $K(X_K)$ be the function field of $X_K := X \times_k \operatorname{Spec} K$. If D is a finite-dimensional division algebra over $k(X)$, then $D_K := D \otimes_{k(X)} K(X_K)$ is a division algebra, too.*

Proof. Since k is algebraically closed, $k(X) \otimes_k K$ is an integral domain; its quotient field is $K(X_K)$. Suppose that D_K contains zero divisors. Clearing denominators, we can then construct zero divisors in $D \otimes_{k(X)} (k(X) \otimes_k K)$, which is clearly isomorphic to $D \otimes_k K$. Consequently, there is a finitely generated k-algebra $A \subseteq K$ such that $D \otimes_k A$ contains zero divisors. These zero divisors are automatically nonzero modulo some maximal ideal $\mathfrak{m} \subset A$, so $D \otimes_k A/\mathfrak{m}$ also contains zero divisors. But $A/\mathfrak{m} \cong k$ by Hilbert's Nullstellensatz; hence D contains zero divisors. This contradiction shows that D_K has to be a division algebra if D is. $\square$

Corollary 1.3. *If E is a generically simple torsion free $\mathcal{A}$-module, then the pullback E_K of E to X_K is a generically simple torsion free module under the pullback $\mathcal{A}_K$ of $\mathcal{A}$.*

Proof. E_K is clearly torsion free and coherent over $\mathcal{O}_{X_K}$. Since the generic fiber of E is Morita equivalent to a one-dimensional D-vector space, the generic fiber of E_K is Morita equivalent to a one-dimensional D_K-vector space; hence E_K is generically simple. $\square$

Definition 1.4. A family of generically simple torsion free $\mathcal{A}$-modules *over a k-scheme S is a sheaf $\mathcal{E}$ of left modules under the pullback $\mathcal{A}_S$ of $\mathcal{A}$ to $X \times_k S$ with the following properties:*

1. *$\mathcal{E}$ is coherent over $\mathcal{O}_{X \times_k S}$ and flat over S.*
2. *For every point $s \in S$, the fiber $\mathcal{E}_s$ is a generically simple torsion free $\mathcal{A}_{k(s)}$-module.*

Here $k(s)$ is the residue field of S at s, and the fiber $\mathcal{E}_s$ is by definition the pullback of $\mathcal{E}$ to $X \times_k \operatorname{Spec} k(s)$.

We denote the corresponding moduli functor by

$$\mathcal{M} = \mathcal{M}_{\mathcal{A}/X} : \underline{Sch_k} \longrightarrow \underline{Sets};$$

it sends a k-scheme S to the set of isomorphism classes of families $\mathcal{E}$ of generically simple torsion free $\mathcal{A}$-modules over S. Our main goal is to construct and study coarse moduli schemes for this functor.

If $\mathcal{E}$ is a family of generically simple torsion free $\mathcal{A}$-modules over S, then there is an open subset of S above which these $\mathcal{A}$-modules are locally projective. However, we work with all torsion free $\mathcal{A}$-modules because they satisfy the following valuative criterion for properness:

Proposition 1.5. *Let V be a discrete valuation ring over k with quotient field K. Then the restriction map*

$$\mathcal{M}(\mathrm{Spec}\, V) \longrightarrow \mathcal{M}(\mathrm{Spec}\, K)$$

is bijective.

Proof. Let $\pi \in V$ be a uniformising element, and let $l = V/(\pi)$ be the residue field of V. We denote by

$$X_K \xrightarrow{j} X_V \xleftarrow{i} X_l$$

the open embedding of the generic fiber and the closed embedding of the special fiber; here $X_A := X \times_k \mathrm{Spec}\, A$ for any k-algebra A. Let η (resp. ξ) be the generic point of X_K (resp. of X_l), and let

$$j_\eta : \mathrm{Spec}\, \mathcal{O}_{X_V,\eta} \longrightarrow X_V \quad (\text{resp. } j_\xi : \mathrm{Spec}\, \mathcal{O}_{X_V,\xi} \longrightarrow X_V)$$

be the 'inclusion' morphism of the subset $\{\eta\}$ (resp. $\{\xi, \eta\}$) into X_V. Let $E \in \mathcal{M}(\mathrm{Spec}\, K)$ be an $\mathcal{A}_K$-module.

Assume given an extension $\mathcal{E} \in \mathcal{M}(\mathrm{Spec}\, V)$ of E. Then $\mathcal{E}$ embeds canonically into j_*E; in particular, the stalk $\mathcal{E}_\xi$ over the discrete valuation ring $\mathcal{O}_{X_V,\xi}$ embeds into the generic fiber E_η. $\mathcal{E}$ is uniquely determined by E and $\mathcal{E}_\xi$ because

$$\mathcal{E} = j_*E \cap j_{\xi,*}\mathcal{E}_\xi \subseteq j_{\eta,*}E_\eta; \tag{1}$$

this equation follows easily from the assumption that the special fiber $i^*\mathcal{E}$ is torsion free, cf. the proof of [11, Proposition 6].

Moreover, the $\mathcal{A}$-stable $\mathcal{O}_{X_V,\xi}$-lattice $\mathcal{E}_\xi \subset E_\eta$ is unique up to powers of π because its quotient modulo π is a simple module under the generic fiber of $\mathcal{A}_l$ by corollary 1.3. This implies that $\mathcal{E}$ is determined by E up to isomorphism, thereby proving injectivity.

To prove surjectivity, we construct an extension $\mathcal{E}$ of E as follows: The simple $\mathcal{A}_\eta$-module E_η is Morita equivalent to a one-dimensional vector space over the division algebra $D_K = D \otimes_{k(X)} \mathcal{O}_{X_V,\eta}$. Inside this vector space, we choose a free module of rank one over $D_V := D \otimes_{k(X)} \mathcal{O}_{X_V,\xi}$ and denote by $\mathcal{E}_\xi$ the Morita equivalent submodule of E_η. Then we define $\mathcal{E}$ by (1); this clearly defines a sheaf of $\mathcal{A}_V$-modules over X_V which is flat over V, whose generic fiber $j^*\mathcal{E}$ is E, and whose special fiber $i^*\mathcal{E}$ is generically simple. According to the proof of [11, Proposition 6] again, $\mathcal{E}$ is coherent over $\mathcal{O}_{X_V}$, and its special fiber $i^*\mathcal{E}$ is torsion free. This shows $\mathcal{E} \in \mathcal{M}(\mathrm{Spec}\, V)$. $\square$

Remark 1.6. Suppose that X is smooth. In the trivial case $\mathcal{A} = \mathcal{O}_X$, generically simple locally projective $\mathcal{A}$-modules are just line bundles, so they also satisfy the valuative criterion for properness; here locally projective and only torsion free modules lie in different connected components of the moduli space. This is no longer true for nontrivial $\mathcal{A}$, even if $\mathcal{A}$ is a sheaf of Azumaya algebras over an abelian or K3 surface X: If $\mathcal{A}_\eta$ is not just a full matrix algebra over $k(X)$, then every generically simple torsion free $\mathcal{A}$-module is a degeneration of locally projective ones by theorem 3.6.iii below; in particular, the latter do not satisfy the valuative criterion for properness.

2 Construction of the coarse moduli schemes

We choose an ample line bundle $\mathcal{O}_X(1)$ on X and put $d := \dim(X)$. As usual, the Hilbert polynomial $P(E)$ of a coherent sheaf E on X with respect to this choice is given by

$$P(E; m) := \chi(E(m)) = \sum_{i=0}^{d} (-1)^i \dim_k \mathrm{H}^i(X; E(m))$$

where $E(m) := E \otimes \mathcal{O}_X(1)^{\otimes m}$. See [7, Chapter 1.2] for details about $P(E)$, in particular for the fact that it is a polynomial of degree $d = \dim(X)$ if E is torsion free.

Recall that the Hilbert polynomial is locally constant in flat families. Keeping it fixed, we denote by

$$\mathcal{M}_{\mathcal{A}/X;P} : \underline{Sch}_k \longrightarrow \underline{Sets}$$

the subfunctor of $\mathcal{M}_{\mathcal{A}/X}$ that parameterizes families $\mathcal{E}$ all of whose fibers $\mathcal{E}_s$ have Hilbert polynomial P.

Our first aim is to prove that the class $\mathcal{M}_{\mathcal{A}/X;P}(\operatorname{Spec} k)$ of generically simple torsion free $\mathcal{A}$-modules E with Hilbert polynomial P is bounded (in the sense of [7, Definition 1.7.5]); this will follow easily from the following apparently weaker statement:

Proposition 2.1. *The class of coherent $\mathcal{O}_X$-modules E with the following two properties is bounded:*

1. *E is torsion free and has Hilbert polynomial P.*
2. *E admits at least one $\mathcal{A}$-module structure for which it is generically simple.*

Proof. Suppose that E is such a sheaf on X. We recall a few concepts, which are all in [7, Chapter 1.2].

$$\mathrm{rk}(E) := \dim_F(E_\eta), \quad \deg(E) := c_1(E) \cdot \mathcal{O}_X(1)^{d-1}, \qquad \mu(E) := \frac{\deg(E)}{\mathrm{rk}(E)}$$

denote the rank, degree and slope of E. Writing the Hilbert polynomial as

$$P(E;m) = \sum_{i=0}^{d} \alpha_i(E)\binom{m+i-1}{i}$$

with integral coefficients α_i [7, Chapter 1.2], one has

$$\deg(E) = \alpha_{d-1}(E) - \mathrm{rk}(E)\alpha_{d-1}(\mathcal{O}_X),$$

cf. [7, Definition 1.2.11]. As $P(E)$ is fixed, it follows in particular that the slope $\mu(E) = \mu$ is fixed. We denote

$$\mu_{\max}(E) := \max\{\mu(E')\big|0 \neq E' \subseteq E \text{ a coherent } \mathcal{O}_X\text{-subsheaf of } E\}.$$

$\mu_{\max}(E)$ is in fact the slope of the first subsheaf $E_{\max} \subseteq E$ in the μ-Harder-Narasimhan filtration of E [7, Section 1.6].
According to a deep result of A. Langer [9, Theorem 4.2], our class of $\mathcal{O}_X$-modules E is bounded if the numbers $\mu_{\max}(E)$ are bounded from above. To check the latter, we choose an integer $m \in \mathbb{Z}$ such that the coherent sheaf $\mathcal{A}(m)$ is generated by its global sections. Since E is generically simple, the multiplication map

$$\mathcal{A} \otimes_{\mathcal{O}_X} E_{\max} \longrightarrow E$$

is generically surjective. Consequently, the induced map

$$\mathrm{H}^0(\mathcal{A}(m)) \otimes_k E_{\max} \longrightarrow E(m)$$

is generically surjective, too. But E certainly has a μ-semistable torsion free quotient E'' with $\mu(E'') \leq \mu$, e.g. the last quotient from the μ-Harder-Narasimhan filtration of E. It is easy to see that there is a nonzero map $E_{\max} \to E''(m)$, obtained by composing

$$E_{\max} \longrightarrow E(m) \longrightarrow E''(m).$$

Since $E_{\max}$ and $E''(m)$ are μ-semistable, this implies

$$\mu_{\max}(E) = \mu(E_{\max}) \leq \mu(E''(m)) = m\mu(\mathcal{O}_X(1)) + \mu(E'') \leq m\mu(\mathcal{O}_X(1)) + \mu.$$

This is the required bound for $\mu_{\max}(E)$. □

According to the proposition, there is an integer $m = m_{\mathcal{A}/X;P}$ with the following property: For every generically simple torsion free $\mathcal{A}$-module E with Hilbert polynomial P, $E(m)$ is generated by global sections, and $\mathrm{H}^i(E(m)) = 0$ for all $i > 0$. We keep this m fixed in the sequel and denote by $N := P(m)$ the common dimension of all the vector spaces $\mathrm{H}^0(E(m))$.

Proposition 2.2. *i) There is a fine moduli scheme R of finite type over k that parameterizes generically simple torsion free $\mathcal{A}$-modules E with Hilbert polynomial P together with a basis of* $\mathrm{H}^0(E(m))$.

ii) *For $l \gg m$, there is an ample line bundle L_l on R whose fibre at E is canonically isomorphic to* $\det \mathrm{H}^0(E(l))$.

iii) *The algebraic group* $\mathrm{GL}(N)$ *over k acts on R by changing the chosen bases of the* $\mathrm{H}^0(E(m))$.

iv) *There is a natural action of* $\mathrm{GL}(N)$ *on L_l that lifts the action in iii.*

v) *The scheme-theoretic stabilizer of every point in $R(k)$ coincides with the scalars $\mathbb{G}_m \subseteq \mathrm{GL}(N)$.*

Proof. i) Let $\mathrm{Quot}_P(\mathcal{A}(-m)^N)$ be Grothendieck's Quot-scheme parameterizing coherent quotients E with Hilbert polynomial P of the $\mathcal{O}_X$-module sheaf $\mathcal{A}(-m)^N$. We can take for R the locally closed subscheme of $\mathrm{Quot}_P(\mathcal{A}(-m)^N)$ defined by the following conditions:

1. The quotient sheaf E is torsion free.
2. E is an $\mathcal{A}$-module, i. e. the kernel of $\mathcal{A}(-m)^N \to E$ is an $\mathcal{A}$-submodule.
3. As an $\mathcal{A}$-module, E is generically simple.
4. The following composed map is an isomorphism:

$$k^N \longrightarrow \mathrm{H}^0(\mathcal{A}^N) \longrightarrow \mathrm{H}^0(E(m)).$$

   Here the left map is given by the unit of the algebra $\mathcal{A}$.

(In particular, this proves that the class of $\mathcal{A}$-modules E in question is bounded.)
ii) By Grothendieck's construction of Quot-schemes, the L_l are a fortiori ample line bundles on $\mathrm{Quot}_P(\mathcal{A}(-m)^N)$.
iii) and iv) also hold for the whole Quot-scheme, cf. [7, 4.3], and its subscheme R is clearly $\mathrm{GL}(N)$-invariant.
v) Let E be a generically simple torsion free $\mathcal{A}$-module with Hilbert polynomial P. Choose a basis of $\mathrm{H}^0(E(m))$ and let $G \subseteq \mathrm{GL}(N)$ be the scheme-theoretic stabilizer of the corresponding point in $R(k)$. It suffices to show that G and $\mathbb{G}_m$ have the same set of points with values in k and in $k[\varepsilon]$ where $\varepsilon^2 = 0$.
Every point in $G(k)$ corresponds to an automorphism of E; hence $G(k) = k^*$ because E has only scalar endomorphisms.
Similarly, every point in $G(k[\varepsilon])$ corresponds to an automorphism of the constant family $E[\varepsilon]$ over $\mathrm{Spec}\, k[\varepsilon]$. Restricting from $\mathrm{Spec}\, k[\varepsilon]$ to $\mathrm{Spec}\, k$, we get an exact sequence

$$0 \longrightarrow \mathrm{End}_{\mathcal{A}}(E) \xrightarrow{\cdot\varepsilon} \mathrm{End}_{\mathcal{A}[\varepsilon]}(E[\varepsilon]) \longrightarrow \mathrm{End}_{\mathcal{A}}(E) \longrightarrow 0;$$

again because E has only scalar endomorphisms, it implies $\mathrm{End}(E[\varepsilon]) = k[\varepsilon]$ and hence $G(k[\varepsilon]) = k[\varepsilon]^*$.
This proves $G = \mathbb{G}_m$. □

Theorem 2.3. *If $l \gg m$, then every point of R is GIT-stable for the action of* $\mathrm{SL}(N) \subset \mathrm{GL}(N)$ *with respect to the linearization L_l.*

Proof. We carry the necessary parts of [7, Chapter 4.4] over to our situation. Put $V := k^N$; then the points of R correspond to quotients

$$\rho : V \otimes_k \mathcal{A}(-m) \to E.$$

We fix such a point.
Let $V' \subset V$ be a proper vector subspace, and put

$$E' := \rho(V' \otimes_k \mathcal{A}(-m)) \subseteq E.$$

Then E' is an $\mathcal{A}$-submodule of E with nonzero generic fibre; since E_η contains no proper $\mathcal{A}_\eta$-submodules, this implies $E'_\eta = E_\eta$. So E and E' have the same rank, i. e. their Hilbert polynomials have the same leading coefficient. Hence

$$\dim(V) \cdot \chi(E'(l)) > \dim(V') \cdot \chi(E(l)) \tag{2}$$

if l is sufficiently large. (We can find one l uniformly for all V' because the family of vector subspaces $V' \subset V$ is bounded.)
After these preliminaries, we can check that our point ρ in R satisfies the Hilbert-Mumford criterion for GIT-stability, cf.[7, Theorem 4.2.11]. So consider a nontrivial one-parameter subgroup $\lambda : \mathbb{G}_m \to \mathrm{SL}(V) = \mathrm{SL}(N)$; we will rather work with the associated eigenspace decomposition $V = \bigoplus_{n\in\mathbb{Z}} V_n$ where $\mathbb{G}_m$ acts on V_n with weight n.
Let $\bar{R}$ be the closure of R in the projective embedding given by L_l; this is also the closure of R in $\mathrm{Quot}_P(V \otimes_k \mathcal{A}(-m))$ because the L_l are ample on the whole Quot-scheme. We have to look at the limit $\lim_{t\to 0} \lambda(t) \cdot \rho$ in $\bar{R}$. This is a fixed point for the $\mathbb{G}_m$-action, so $\mathbb{G}_m$ acts on the fibre of L_l over it, necessarily with some weight $-\mu^{L_l}(\rho, \lambda) \in \mathbb{Z}$; what we have to show for stability is $\mu^{L_l}(\rho, \lambda) > 0$.
First we describe the limit point $\lim_{t\to 0} \lambda(t) \cdot \rho$ as a point in the Quot-scheme. We put

$$V_{\leq n} := \bigoplus_{\nu \leq n} V_\nu \subseteq V \quad \text{and} \quad E_{\leq n} := \rho(V_{\leq n} \otimes_k \mathcal{A}(-m)) \subseteq E.$$

Then $V_n = V_{\leq n}/V_{\leq n-1}$; we put $E_n := E_{\leq n}/E_{\leq n-1}$, thus obtaining surjections

$$\rho_n : V_n \otimes_k \mathcal{A}(-m) \longrightarrow E_n.$$

Then

$$\bar{\rho} := \bigoplus_{n\in\mathbb{Z}} \rho_n : V \otimes_k \mathcal{A}(-m) \longrightarrow \bigoplus_{n\in\mathbb{Z}} E_n$$

is also a point in $\mathrm{Quot}_P(V \otimes_k \mathcal{A}(-m))$; it is the limit we are looking for:

$$\bar{\rho} = \lim_{t\to 0} \lambda(t) \cdot \rho.$$

To prove this, just copy the proof of [7, Lemma 4.4.3], replacing $\mathcal{O}_X(-m)$ by $\mathcal{A}(-m)$ everywhere.
The second step is to consider the fibre of L_l over $\bar{\rho}$. It is by definition

$$\det H^0(\bigoplus_{n\in\mathbb{Z}} E_n(l));$$

this is canonically isomorphic to the tensor product of the determinant of cohomology of the $E_n(l)$. Now $\mathbb{G}_m$ acts on V_n with weight n, so it acts on the fiber in question with weight

$$-\mu^{L_l}(\rho,\lambda) = \sum_{n\in\mathbb{Z}} n\cdot\chi(E_n(l)),$$

cf. [7, Lemma 4.4.4].
Finally, we use the preliminaries above to estimate this sum. If we apply (2) to $V' = V_{\le n}$, $E' = E_{\le n}$ and sum up, we get

$$\sum_{n\in\mathbb{Z}} \big(\dim(V)\cdot\chi(E_{\le n}(l)) - \dim(V_{\le n})\cdot\chi(E(l))\big) > 0; \tag{3}$$

note that only finitely many summands are nonzero because $V_{\le n}$ is zero or V for almost all n since almost all V_n are zero. Put

$$a_n := \dim(V)\cdot\chi(E_n(l)) - \dim(V_n)\cdot\chi(E(l));$$

again, all but finitely many of these integers are nonzero, and

$$\sum_{n\in\mathbb{Z}} a_n = \dim(V)\cdot\chi(E(l)) - \dim(V)\cdot\chi(E(l)) = 0.$$

If we write $a_{\le n} := \sum_{\nu\le n} a_\nu$, then

$$\sum_n a_{\le n} + \sum_n n a_n = 0$$

because the sum of the a_n is zero. Hence (3) is equivalent to

$$\sum_{n\in\mathbb{Z}} n\big(\dim(V)\cdot\chi(E_n(l)) - \dim(V_n)\cdot\chi(E(l))\big) < 0.$$

But $\sum_n n\dim V_n = 0$ because $\mathbb{G}_m$ acts on V with determinant 1. Thus we obtain

$$\sum_{n\in\mathbb{Z}} n\big(\dim(V)\cdot\chi(E_n(l))\big) < 0,$$

i. e. $-\mu^{L_l}(\rho,\lambda)\cdot\dim(V) < 0$. This proves that the Hilbert-Mumford criterion for GIT-stability is satisfied here. □

Now we can state our main result. See [15] for the concepts 'geometric quotient' and 'coarse moduli scheme'.

Theorem 2.4. *i) The action of* $\mathrm{GL}(N)$ *on* R *described in proposition 2.2.iii above admits a geometric quotient*

$$\mathrm{M}_{\mathcal{A}/X;P} := \mathrm{GL}(N)\backslash R$$

which is a separated scheme of finite type over k.

ii) The quotient morphism

$$R \longrightarrow \mathrm{M}_{\mathcal{A}/X;P}$$

is a principal $\mathrm{PGL}(N)$*-bundle (locally trivial in the fppf-topology).*

iii) $\mathrm{M}_{\mathcal{A}/X;P}$ *is a coarse moduli scheme for the moduli functor* $\mathcal{M}_{\mathcal{A}/X;P}$.

iv) $\mathrm{M}_{\mathcal{A}/X;P}$ *is projective over* k.

Proof. i) According to geometric invariant theory [15, Theorem 1.10 and Appendix 1.C] and theorem 2.3 above, the action of $\mathrm{SL}(N)$ on R admits a geometric quotient $\mathrm{M}_{\mathcal{A}/X;P}$ which is quasiprojective over k. Then $\mathrm{M}_{\mathcal{A}/X;P}$ is also a geometric quotient for the action of $\mathrm{GL}(N)$ on R because both groups act via $\mathrm{PGL}(N) = \mathrm{PSL}(N)$.
ii) This morphism is affine by its GIT-construction. According to [15, Proposition 0.9], it suffices to show that the action of $\mathrm{PGL}(N)$ on R is free, i. e. that

$$\psi : \mathrm{PGL}(N) \times R \longrightarrow R \times R, \qquad (g, r) \mapsto (g \cdot r, r)$$

is a closed immersion. The fibers of ψ over k-points of $R \times R$ are either empty or isomorphic to $\operatorname{Spec} k$ by proposition 2.2.v above; furthermore, ψ is proper due to [15, Proposition 0.8]. Hence ψ is indeed a closed immersion.
iii) Part i implies that $\mathrm{M}_{\mathcal{A}/X;P}$ is a coarse moduli scheme for the functor

$$\underline{\mathrm{GL}(N)\backslash R} : \underline{Sch_k} \longrightarrow \underline{Sets}$$

that sends a k-scheme S to the set of $\mathrm{GL}(N)(S)$-orbits in $R(S)$. However, this functor is very close to the moduli functor $\mathcal{M} = \mathcal{M}_{\mathcal{A}/X;P}$:
We have a morphism from the functor represented by R to $\mathcal{M}$; it simply forgets the extra structure. If S is a scheme over k, then two S-valued points of R have the same image in $\mathcal{M}(S)$ if and only if they are in the same $\mathrm{GL}(N)(S)$-orbit. Thus we get a morphism of functors

$$\phi : \underline{\mathrm{GL}(N)\backslash R} \longrightarrow \mathcal{M}$$

which is injective for every scheme S over k. The image of ϕ consists of all sheaves $\mathcal{E} \in \mathcal{M}(S)$ for which the vector bundle $pr_*\mathcal{E}(m)$ over S is trivial, where $pr : X \times_k S \to S$ is the canonical projection.

In particular, ϕ is bijective whenever S is the spectrum of a field, and it induces an isomorphism between the Zariski sheafifications of both functors. It follows that $\mathrm{M}_{\mathcal{A}/X;P}$ is also a coarse moduli scheme for the functor $\mathcal{M}$.
iv) We already know that $\mathrm{M}_{\mathcal{A}/X;P}$ is quasiprojective. Furthermore, it satisfies the valuative criterion for properness by proposition 1.5. $\square$

In particular, the moduli functor $\mathcal{M}_{\mathcal{A}/X}$ of *all* generically simple torsion free $\mathcal{A}$-modules has a coarse moduli scheme

$$\mathrm{M}_{\mathcal{A}/X} = \coprod_P \mathrm{M}_{\mathcal{A}/X;P}$$

which is a disjoint sum of projective schemes over k. If X is smooth of dimension d, then we have another such decomposition

$$\mathrm{M}_{\mathcal{A}/X} = \coprod_{c_1,\ldots,c_d} \mathrm{M}_{\mathcal{A}/X;c_1,\ldots,c_d}$$

given by fixing the Chern classes $c_i \in \mathrm{CH}^i(X)$, the Chow group of cycles modulo *algebraic* equivalence. Indeed, each $\mathrm{M}_{\mathcal{A}/X;c_1,\ldots,c_d}$ is open and closed in some $\mathrm{M}_{\mathcal{A}/X;P}$ where P is given by Hirzebruch-Riemann-Roch.
If X is a smooth projective surface, then this decomposition reads

$$\mathrm{M}_{\mathcal{A}/X} = \coprod_{\substack{c_1\in \mathrm{NS}(X)\\ c_2\in\mathbb{Z}}} \mathrm{M}_{\mathcal{A}/X;c_1,c_2}.$$

3 Deformations and smoothness

We introduce the usual cohomology classes that describe deformations of a coherent $\mathcal{A}$-module E, following Artamkin [1]. By definition, a *deformation* $\mathcal{E}$ of E over a local artinian k-algebra $(A,\mathfrak{m})$ with residue field k is a (flat) family $\mathcal{E}$ of coherent $\mathcal{A}$-modules parameterized by $\operatorname{Spec} A$ together with an isomorphism $k\otimes_A \mathcal{E} \cong E$.
Consider first the special case $A = k[\varepsilon]$ with $\varepsilon^2 = 0$. Then we have an exact sequence of A-modules

$$0 \longrightarrow k \xrightarrow{\cdot\varepsilon} k[\varepsilon] \longrightarrow k \longrightarrow 0. \tag{4}$$

By definition, the *Kodaira-Spencer class* of the deformation $\mathcal{E}$ over $k[\varepsilon]$ is the Yoneda extension class

$$\mathrm{ks}(\mathcal{E}) := [0 \longrightarrow E \xrightarrow{\cdot\varepsilon} \mathcal{E} \longrightarrow E \longrightarrow 0] \in \operatorname{Ext}^1_{\mathcal{A}}(E,E)$$

obtained by tensoring (4) over A with $\mathcal{E}$.

Lemma 3.1. *The Kodaira-Spencer map* ks *is a bijection between isomorphism classes of deformations of E over $k[\varepsilon]$ and elements of* $\operatorname{Ext}^1_{\mathcal{A}}(E,E)$.

Proof. Let $\mathcal{E}$ be an $\mathcal{A}$-module extension of E by E. Then $\mathcal{E}$ becomes an $\mathcal{A}[\varepsilon]$-module if we let ε act via the composition $\mathcal{E} \twoheadrightarrow E \hookrightarrow \mathcal{E}$. According to the local criterion for flatness [4, Theorem 6.8], $\mathcal{E}$ is flat over $k[\varepsilon]$ and hence a deformation of E. This defines the required inverse map. □

Now let $(A, \mathfrak{m})$ be arbitrary again, and let $\tilde{A}$ be a minimal extension of A. In other words, $(\tilde{A}, \tilde{\mathfrak{m}})$ is another local artinian k-algebra with residue field k, and $A \cong \tilde{A}/(\nu)$ where $\nu \in \tilde{A}$ is annihilated by $\tilde{\mathfrak{m}}$. Then we have an exact sequence of A-modules

$$0 \longrightarrow k \xrightarrow{\cdot\nu} \tilde{\mathfrak{m}} \longrightarrow A \longrightarrow k \longrightarrow 0. \tag{5}$$

By definition, the *obstruction class* of the deformation $\mathcal{E}$ over A is the Yoneda extension class

$$\mathrm{ob}(\mathcal{E}; k \overset{\cdot\nu}{\hookrightarrow} \tilde{A} \twoheadrightarrow A) := [0 \longrightarrow E \xrightarrow{\cdot\nu} \tilde{\mathfrak{m}} \otimes_A \mathcal{E} \longrightarrow \mathcal{E} \longrightarrow E \longrightarrow 0] \in \mathrm{Ext}^2_{\mathcal{A}}(E, E)$$

obtained by tensoring (5) over A with $\mathcal{E}$. Whenever we want to mention $\mathcal{A}$, we write $\mathrm{ob}_{\mathcal{A}}$ instead of ob; on the other hand, we may omit $k \overset{\cdot\nu}{\hookrightarrow} \tilde{A} \twoheadrightarrow A$ if they are clear from the context.

Lemma 3.2. *The obstruction class* $\mathrm{ob}(\mathcal{E}; k \overset{\cdot\nu}{\hookrightarrow} \tilde{A} \twoheadrightarrow A)$ *vanishes if and only if* $\mathcal{E}$ *can be extended to a deformation* $\tilde{\mathcal{E}}$ *over* $\tilde{A}$.

Proof. If $\tilde{\mathcal{E}}$ is a deformation over $\tilde{A}$ extending $\mathcal{E}$, then we can tensor it over $\tilde{A}$ with the diagram

$$\begin{tikzcd}
0 \arrow[r] & k \arrow[r, "\cdot\nu"] \arrow[d, equal] & \tilde{\mathfrak{m}} \arrow[r] \arrow[d] & \mathfrak{m} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & k \arrow[r, "\cdot\nu"] & \tilde{A} \arrow[r] & A \arrow[r] & 0;
\end{tikzcd}$$

this gives us a morphism of short exact sequences of $\mathcal{A}$-modules

$$\begin{tikzcd}
0 \arrow[r] & E \arrow[r, "\cdot\nu"] \arrow[d, equal] & \tilde{\mathfrak{m}} \otimes_A \mathcal{E} \arrow[r] \arrow[d] & \mathfrak{m} \otimes_A \mathcal{E} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & E \arrow[r] & \tilde{\mathcal{E}} \arrow[r] & \mathcal{E} \arrow[r] & 0.
\end{tikzcd}$$

The existence of such a morphism implies $\mathrm{ob}(\mathcal{E}; k \overset{\cdot\nu}{\hookrightarrow} \tilde{A} \twoheadrightarrow A) = 0$ due to the standard exact sequence

$$\ldots \mathrm{Ext}^1_{\mathcal{A}}(\mathcal{E}, E) \longrightarrow \mathrm{Ext}^1_{\mathcal{A}}(\mathfrak{m} \otimes_A \mathcal{E}, E) \longrightarrow \mathrm{Ext}^2_{\mathcal{A}}(E, E) \ldots$$

Conversely, $\mathrm{ob}(\mathcal{E}; k \overset{\cdot\nu}{\hookrightarrow} \tilde{A} \twoheadrightarrow A) = 0$ implies the existence of such a morphism of short exact sequences of $\mathcal{A}$-modules. Then $\tilde{\mathcal{E}}$ becomes an $\tilde{A} \otimes_k \mathcal{A}$-module if we let any $a \in \tilde{\mathfrak{m}} \subset \tilde{A}$ act via the composition $\tilde{\mathcal{E}} \twoheadrightarrow \mathcal{E} \xrightarrow{a \otimes -} \tilde{\mathfrak{m}} \otimes_A \mathcal{E} \hookrightarrow \tilde{\mathcal{E}}$. According to the local criterion for flatness [4, Theorem 6.8], $\tilde{\mathcal{E}}$ is flat over $\tilde{A}$ and hence a deformation of E extending $\mathcal{E}$. □

In the special case $A = k[\varepsilon]$ and $\tilde{A} = k[\delta]$ with $\delta^3 = 0$, we note that (5) is the Yoneda product of (4) with itself and hence

$$\mathrm{ob}(\mathcal{E}; k \overset{\cdot\delta^2}{\hookrightarrow} k[\delta] \twoheadrightarrow k[\varepsilon]) = \mathrm{ks}(\mathcal{E}) \times \mathrm{ks}(\mathcal{E}). \tag{6}$$

From now on, we assume that X is smooth. Then we have a trace map

$$\mathrm{tr}_{\mathcal{O}_X} : \mathrm{Ext}^i_{\mathcal{O}_X}(E, E) \longrightarrow \mathrm{H}^i(X, \mathcal{O}_X)$$

for every coherent $\mathcal{O}_X$-module E: it is defined using a finite locally free resolution of E, cf. [7, 10.1.2]. If E is a coherent $\mathcal{A}$-module, then we define the trace map $\mathrm{tr} = \mathrm{tr}_{\mathcal{A}/\mathcal{O}_X}$ as the composition

$$\mathrm{tr} = \mathrm{tr}_{\mathcal{A}/\mathcal{O}_X} : \mathrm{Ext}^i_{\mathcal{A}}(E, E) \longrightarrow \mathrm{Ext}^i_{\mathcal{O}_X}(E, E) \xrightarrow{\mathrm{tr}_{\mathcal{O}_X}} \mathrm{H}^i(X, \mathcal{O}_X)$$

where the first map is induced by the forgetful functor from $\mathcal{A}$-modules to $\mathcal{O}_X$-modules. Similarly, one can define a trace map

$$\mathrm{tr}^{\omega_X}_{\mathcal{A}/\mathcal{O}_X} : \mathrm{Ext}^i_{\mathcal{A}}(E, E \otimes_{\mathcal{O}_X} \omega_X) \longrightarrow \mathrm{H}^i(X, \omega_X)$$

where ω_X is the canonical line bundle on X; cf. [7, p. 218].
Now suppose that $\mathcal{E}$ is a deformation of E over A. Then we have a line bundle $\det \mathcal{E}$ over $X_A = X \times_k \operatorname{Spec} A$: it is defined using a finite locally free resolution of $\mathcal{E}$ as an $\mathcal{O}_{X_A}$-module, cf. [7, 1.1.17 and Proposition 2.1.10].

Proposition 3.3. *If $\tilde{A}$ is a minimal extension of A, then*

$$\mathrm{tr}(\mathrm{ob}_{\mathcal{A}}(\mathcal{E}; k \hookrightarrow \tilde{A} \twoheadrightarrow A)) = \mathrm{ob}_{\mathcal{O}_X}(\det \mathcal{E}; k \hookrightarrow \tilde{A} \twoheadrightarrow A) \in \mathrm{H}^2(X, \mathcal{O}_X).$$

In particular, $\mathrm{tr}(\mathrm{ob}_{\mathcal{A}}(\mathcal{E})) = 0$ *if the Picard variety* $\mathrm{Pic}(X)$ *is smooth.*

Proof. The forgetful map $\mathrm{Ext}^2_{\mathcal{A}}(E, E) \to \mathrm{Ext}^2_{\mathcal{O}_X}(E, E)$ maps $\mathrm{ob}_{\mathcal{A}}(\mathcal{E})$ to $\mathrm{ob}_{\mathcal{O}_X}(\mathcal{E})$ by definition. It is known that $\mathrm{tr}_{\mathcal{O}_X}$ maps the latter to $\mathrm{ob}_{\mathcal{O}_X}(\det \mathcal{E})$; cf. Artamkin's paper [1] for the computation. □

For the rest of this section, we assume that $\mathcal{A}$ is even a sheaf of Azumaya algebras over the smooth projective variety X of dimension d.

Proposition 3.4. *Every coherent sheaf E of $\mathcal{A}$-modules has a resolution of length $\leq d = \dim(X)$ by locally projective sheaves of $\mathcal{A}$-modules.*

Proof. If m is sufficiently large, then the twist $E(m)$ is generated by its global sections; this gives us a surjection ∂_0 of $E_0 := \mathcal{A}(-m)^N$ onto E for some N. Applying the same procedure to the kernel of ∂_0 and iterating, we obtain an infinite resolution by locally free $\mathcal{A}$-modules

$$\ldots E_d \xrightarrow{\partial_d} E_{d-1} \ldots E_1 \xrightarrow{\partial_1} E_0 \xrightarrow{\partial_0} E.$$

We claim that the image of ∂_d is locally projective over $\mathcal{A}$; then we can truncate there, and the proposition follows.
It suffices to check this claim over the complete local rings $\hat{\mathcal{O}}_{X,x}$ at the closed points x of X; there $\mathcal{A}$ becomes a matrix algebra $\hat{\mathcal{A}}_x$, so the resulting $\hat{\mathcal{A}}_x$-modules $\hat{E}_{i,x}$ are Morita equivalent to $\hat{\mathcal{O}}_{X,x}$-modules. Since $\hat{\mathcal{O}}_{X,x}$ has homological dimension d, the image of $\partial_{d,x} : \hat{E}_{d,x} \to \hat{E}_{d-1,x}$ is projective over $\hat{\mathcal{A}}_x$. Hence the image of ∂_d is indeed locally projective over $\mathcal{A}$. $\square$

Our main tool to control the extension classes introduced above will be the following variant of Serre duality. To state it, we fix an isomorphism $\mathrm{H}^d(X, \omega_X) \cong k$.

Proposition 3.5. *We still assume that X is smooth of dimension d and that $\mathcal{A}$ is a sheaf of Azumaya algebras. If E and E' are coherent $\mathcal{A}$-modules, then the Yoneda product*

$$\mathrm{Ext}^i_{\mathcal{A}}(E, E') \otimes \mathrm{Ext}^{d-i}_{\mathcal{A}}(E', E \otimes \omega_X) \longrightarrow \mathrm{Ext}^d_{\mathcal{A}}(E, E \otimes \omega_X)$$

followed by the trace map

$$\mathrm{tr}^{\omega_X}_{\mathcal{A}/\mathcal{O}} : \mathrm{Ext}^d_{\mathcal{A}}(E, E \otimes \omega_X) \longrightarrow \mathrm{H}^d(X, \omega_X) \cong k$$

defines a perfect pairing of finite-dimensional vector spaces over k.

Proof. We start with the special case that E and E' are locally projective over $\mathcal{A}$. Then the Ext-groups in question are Zariski cohomology groups of the locally free $\mathcal{O}_X$-module sheaves $\mathcal{H}om_{\mathcal{A}}(E, E')$ and $\mathcal{H}om_{\mathcal{A}}(E', E) \otimes \omega_X$. But $\mathcal{H}om_{\mathcal{A}}(E, E')$ and $\mathcal{H}om_{\mathcal{A}}(E', E)$ are dual to each other by means of an appropriate local trace map, using the fact that the trace map $\mathcal{A} \otimes_{\mathcal{O}_X} \mathcal{A} \to \mathcal{O}_X$ is nowhere degenerate because $\mathcal{A}$ is Azumaya. Hence this special case follows from the usual Serre duality theorem for locally free $\mathcal{O}_X$-modules.
If E and E' are not necessarily locally projective over $\mathcal{A}$, then we choose finite locally projective resolutions, using proposition 3.4. Induction on their length reduces us to the case where E and E' have resolutions of length one by $\mathcal{A}$-modules for which the duality in question holds. Now $\mathrm{Ext}^i_{\mathcal{A}}(E, E')$ and $\mathrm{Ext}^{d-i}_{\mathcal{A}}(E', E \otimes \omega)^{\mathrm{dual}}$ are δ-functors in both variables E and E', and the pairing defines a morphism between them. An application of the five lemma to the resulting morphisms of long exact sequences proves the required induction step. $\square$

Theorem 3.6. *Let X be an abelian or K3 surface over k, and let $\mathcal{A}$ be a sheaf of Azumaya algebras over X. Suppose $\mathcal{A}_\eta \cong \mathrm{Mat}(n \times n; D)$ for a central division algebra D of dimension r^2 over the function field $k(X)$.*

i) The moduli space $\mathrm{M}_{\mathcal{A}/X}$ of generically simple torsion free $\mathcal{A}$-modules E is smooth.

ii) There is a nowhere degenerate alternating 2*-form on the tangent bundle of* $\mathrm{M}_{\mathcal{A}/X}$*.*

iii) If $r \geq 2$*, then the open locus* $\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X}$ *of locally projective* $\mathcal{A}$*-modules* E *is dense in* $\mathrm{M}_{\mathcal{A}/X}$*.*

iv) If we fix the Chern classes $c_1 \in \mathrm{NS}(X)$ *and* $c_2 \in \mathbb{Z}$ *of* E*, then*

$$\dim \mathrm{M}_{\mathcal{A}/X;c_1,c_2} = \Delta/(nr)^2 - c_2(\mathcal{A})/n^2 - r^2\chi(\mathcal{O}_X) + 2$$

where $\Delta = 2r^2nc_2 - (r^2n-1)c_1^2$ *is the discriminant of* E*.*

Proof. i) We have to check that all obstruction classes

$$\mathrm{ob}_{\mathcal{A}}(\mathcal{E}; k \hookrightarrow \tilde{A} \twoheadrightarrow A) \in \mathrm{Ext}^2_{\mathcal{A}}(E,E)$$

vanish. $\mathrm{Pic}(X)$ is known to be smooth; using proposition 3.3, it suffices to show that the trace map

$$\mathrm{tr}_{\mathcal{A}/\mathcal{O}_X} : \mathrm{Ext}^2_{\mathcal{A}}(E,E) \longrightarrow \mathrm{H}^2(X, \mathcal{O}_X)$$

is injective. But it is straightforward to check that this map is Serre-dual to the natural map

$$\mathrm{H}^0(X, \omega_X) \longrightarrow \mathrm{Hom}_{\mathcal{A}}(E, E \otimes \omega_X),$$

which is an isomorphism because ω_X is trivial and E has only scalar endomorphisms.

ii) Mukai's argument in [14] carries over to our situation as follows. We fix an isomorphism $\omega_X \cong \mathcal{O}_X$. The Kodaira-Spencer map identifies the tangent space $T_{[E]}\mathrm{M}_{\mathcal{A}/X}$ with $\mathrm{Ext}^1_{\mathcal{A}}(E,E)$. On this vector space, the Serre duality 3.5 defines a nondegenerate bilinear form. Indeed, this form is just the Yoneda product

$$\mathrm{Ext}^1_{\mathcal{A}}(E,E) \otimes \mathrm{Ext}^1_{\mathcal{A}}(E,E) \longrightarrow \mathrm{Ext}^2_{\mathcal{A}}(E,E);$$

the right hand side is isomorphic to k by Serre duality again. Equation (6) implies that this bilinear form is alternating because all obstruction classes vanish here.

iii) Let E be a generically simple torsion free $\mathcal{A}$-module, and let $\mathrm{Quot}_l(E/\mathcal{A})$ be the moduli scheme of quotients $E \twoheadrightarrow T$ where T is a coherent $\mathcal{A}$-module of finite length l. This is a closed subscheme of Grothendieck's Quot-scheme $\mathrm{Quot}_{lnr}(E)$ parameterizing those exact sequences of coherent sheaves $0 \to E' \to E \to T \to 0$ for which E' is an $\mathcal{A}$-submodule, i. e. the composition

$$\mathcal{A} \otimes E' \hookrightarrow \mathcal{A} \otimes E \stackrel{\cdot}{\longrightarrow} E \twoheadrightarrow T$$

vanishes; here $(nr)^2 = \mathrm{rk}(\mathcal{A})$. In particular, $\mathrm{Quot}_l(E/\mathcal{A})$ is projective over k. We show by induction that $\mathrm{Quot}_l(E/\mathcal{A})$ is connected; cf. [7, 6.A.1].

Let $\mathrm{Drap}_{l_1,l_2}(E/\mathcal{A})$ be the moduli scheme of iterated quotients $E \twoheadrightarrow T_1 \twoheadrightarrow T_2$ where T_i is a coherent $\mathcal{A}$-module of finite length l_i for $i = 1, 2$; this is again a closed subscheme of some Flag-scheme [7, 2.A.1] and hence projective over k. Sending such an iterated quotient to T_1 and to the pair $(T_2, \mathrm{supp}(T_1/T_2))$ defines two morphisms

$$\mathrm{Quot}_{l+1}(E/\mathcal{A}) \xleftarrow{\theta_1} \mathrm{Drap}_{l+1,l}(E/\mathcal{A}) \xrightarrow{\theta_2} \mathrm{Quot}_l(E/\mathcal{A}) \times X.$$

Using Morita equivalence over the complete local rings at the support of torsion sheaves, it is easy to see that θ_1 and θ_2 are both surjective; moreover, the fibers of θ_2 are projective spaces and hence connected. This shows that $\mathrm{Quot}_{l+1}(E/\mathcal{A})$ is connected if $\mathrm{Quot}_l(E/\mathcal{A})$ is; thus they are all connected.
Let E still be a generically simple torsion free $\mathcal{A}$-module; we have to show that its connected component in $\mathrm{M}_{\mathcal{A}/X}$ contains a locally projective $\mathcal{A}$-module. Let

$$E^* := \mathcal{H}om_{\mathcal{O}_X}(E, \mathcal{O}_X)$$

be the dual of E; this is a sheaf of right $\mathcal{A}$-modules. The double dual E^{**} is a sheaf of left $\mathcal{A}$-modules again; it is locally free over $\mathcal{O}_X$ and hence locally projective over $\mathcal{A}$. We have an exact sequence

$$0 \longrightarrow E \xrightarrow{\iota} E^{**} \xrightarrow{\pi} T \longrightarrow 0$$

where T is a coherent $\mathcal{A}$-module of finite length l. There is a natural map

$$\mathrm{Quot}_l(E^{**}/\mathcal{A}) \longrightarrow \mathrm{M}_{\mathcal{A}/X}$$

that sends a quotient to its kernel; since $\mathrm{Quot}_l(E^{**}/\mathcal{A})$ is connected, we may assume that T is as simple as possible, i. e. that its support consists of l distinct points $x_1, \dots, x_l \in X$ where the stalks T_{x_i} are Morita-equivalent to coherent skyscraper sheaves of length one.
In this situation, we adapt an argument of Artamkin [2] to show that E can be deformed to a locally projective $\mathcal{A}$-module if $r \geq 2$. We consider the diagram

$$\begin{tikzcd}
\mathrm{Ext}^1_{\mathcal{A}}(E,E) \arrow[r, "\delta"] & \mathrm{Ext}^2_{\mathcal{A}}(T,E) \arrow[r, "\pi^*"] \arrow[d, "\iota_*"] & \mathrm{Ext}^2_{\mathcal{A}}(E^{**},E) \\
 & \mathrm{Ext}^2_{\mathcal{A}}(T,E^{**}) \arrow[r, equal] & \bigoplus_{i=1}^{l} \mathrm{Ext}^2_{\mathcal{A}}(T_{x_i},E^{**}).
\end{tikzcd}$$

Here π^* is Serre-dual to $\pi_* : \mathrm{Hom}_{\mathcal{A}}(E, E^{**}) \to \mathrm{Hom}_{\mathcal{A}}(E, T)$ because $\omega_X \cong \mathcal{O}_X$. But the only morphisms from E to E^{**} are the multiples of ι; hence $\pi^* = 0$, and the connecting homomorphism δ from the long exact sequence is surjective. ι_* corresponds under Serre duality and Morita equivalence to the direct sum of the restriction maps

$$\mathrm{Hom}_{\mathcal{O}_{x_i}}(\mathcal{O}_{x_i}^r, k_{x_i}) \longrightarrow \mathrm{Hom}_{\mathcal{O}_{x_i}}(\mathfrak{m}_{x_i} \oplus \mathcal{O}_{x_i}^{r-1}, k_{x_i})$$

where $\mathcal{O}_x = \hat{\mathcal{O}}_{X,x}$ is the complete local ring of X at x, $\mathfrak{m}_x \subseteq \mathcal{O}_x$ is its maximal ideal, and $k_x = \mathcal{O}_x/\mathfrak{m}_x$ is the residue field. Assuming $r \geq 2$, these restriction maps are obviously nonzero.

Hence there is a class $\xi \in \mathrm{Ext}^1_{\mathcal{A}}(E, E)$ whose image in $\mathrm{Ext}^2_{\mathcal{A}}(T_{x_i}, E^{**})$ is nonzero for all i. Since all obstruction classes vanish, we can find a deformation $\mathcal{E}$ of E over a smooth connected curve whose Kodaira-Spencer class is ξ; it remains to show that a general fiber E' of $\mathcal{E}$ is locally projective over $\mathcal{A}$.

Forming the double dual, we get an exact sequence

$$0 \longrightarrow E' \longrightarrow (E')^{**} \longrightarrow T' \longrightarrow 0.$$

An explicit computation using Morita equivalence shows that the forgetful map

$$k^r \cong \mathrm{Ext}^2_{\mathcal{A}}(T_{x_i}, E^{**}) \longrightarrow \mathrm{Ext}^2_{\mathcal{O}_X}(T_{x_i}, E^{**}) \cong k^{n^2 r^3}$$

is injective. Hence the Kodaira-Spencer class of $\mathcal{E}$ in $\mathrm{Ext}^1_{\mathcal{O}_X}(E, E)$ also has nonzero image in $\mathrm{Ext}^2_{\mathcal{O}_X}(T_{x_i}, E^{**})$ for all i. According to [2, Corollary 1.3 and the proof of Lemma 6.2], this implies that E' is less singular than E, i. e. the length of T' as an $\mathcal{O}_X$-module is strictly less than nr at every point of its support. But T' is an $\mathcal{A}$-module, so these lengths are all divisible by nr; hence $T' = 0$, and E' is locally projective over $\mathcal{A}$.

iv) Using i and iii, it suffices to compute the dimension of

$$T_{[E]} M_{\mathcal{A}/X} \cong \mathrm{Ext}^1_{\mathcal{A}}(E, E) \cong \mathrm{H}^1(\mathit{End}_{\mathcal{A}}(E))$$

for a generically simple locally projective $\mathcal{A}$-module E. Note that H^0 and H^2 are Serre-dual to each other and hence both one-dimensional here.

The endomorphism sheaf $\mathit{End}_{\mathcal{A}}(E)$ is an Azumaya algebra of rank r^2 over X, and the natural map

$$\mathcal{A} \otimes_{\mathcal{O}_X} \mathit{End}_{\mathcal{A}}(E) \longrightarrow \mathit{End}_{\mathcal{O}_X}(E)$$

is an isomorphism; this is easily checked by reducing to the case that $\mathcal{A}$ is a matrix algebra. Furthermore, $c_1(\mathcal{A})$ is numerically equivalent to zero because $\mathcal{A} \cong \mathcal{A}^{\mathrm{dual}}$ using the trace over $\mathcal{O}_X$; similarly, $c_1(\mathit{End}_{\mathcal{A}}(E))$ and $c_1(\mathit{End}_{\mathcal{O}_X}(E))$ are also numerically equivalent to zero. Using this, the formalism of Chern classes yields

$$r^2 c_2(\mathcal{A}) + (nr)^2 c_2(\mathit{End}_{\mathcal{A}}(E)) = \Delta \tag{7}$$

where Δ is the discriminant of E as above. Hence

$$\chi(\mathit{End}_{\mathcal{A}}(E)) = -\Delta/(nr)^2 + c_2(\mathcal{A})/n^2 + r^2\chi(\mathcal{O}_X)$$

by Hirzebruch-Riemann-Roch. □

4 Left and right orders

We still assume that X is smooth projective of dimension d and that $\mathcal{A}$ is a sheaf of Azumaya algebras over X; furthermore, we suppose that the generic fiber $\mathcal{A}_\eta$ is a division algebra D of dimension r^2 over $k(X)$. In this case, generically simple locally projective $\mathcal{A}$-modules are just locally free $\mathcal{A}$-modules of rank one; of course every such $\mathcal{A}$-module E can be embedded into D. The endomorphism sheaf of such a left $\mathcal{A}$-module E is then just an order in D acting by right multiplication on E, exactly as in the classical picture of the Brandt groupoid [3, VI, §2, Satz 14].
The Picard group $\mathrm{Pic}(X)$ acts on the moduli scheme $\mathrm{M}_{\mathcal{A}/X}$ by tensor product: a line bundle $L \in \mathrm{Pic}(X)$ acts as $E \mapsto E \otimes_{\mathcal{O}_X} L$. The projective group scheme $\mathrm{Pic}^0(X)$ of line bundles L algebraically equivalent to zero acts on the individual pieces $\mathrm{M}_{\mathcal{A}/X;c_1,\dots,c_d}$ because $c_i(E \otimes L) = c_i(E)$ for all i. The same remarks hold for

$$\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X} = \coprod_{c_1,\dots,c_d} \mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X;c_1,\dots,c_d}$$

where the superscript lp denotes the open locus of locally projective (and hence locally free) $\mathcal{A}$-modules.

Proposition 4.1. *There is a geometric quotient of $\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X}$ by the action of $\mathrm{Pic}(X)$; it is a disjoint sum of separated schemes of finite type over k. Its closed points correspond bijectively to isomorphism classes of Azumaya algebras over X with generic fiber D.*

Proof. $\mathrm{Pic}(X)$ acts with finite stabilizers; this follows from

$$\det(E \otimes L) \cong \det(E) \otimes L^{\otimes r^2}. \tag{8}$$

For fixed Chern classes $c_1, \dots, c_d$, let $G \subseteq \mathrm{Pic}(X)$ be the subgroup of all line bundles L that map $\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X;c_1,\dots,c_d}$ to $\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X;c_1,\dots,c_d}$. Then G contains $\mathrm{Pic}^0(X)$, and its image in $\mathrm{NS}(X) = \mathrm{Pic}(X)/\mathrm{Pic}^0(X)$ is contained in the r^2-torsion and hence finite, so G is a projective group scheme. Therefore a geometric quotient of $\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X;c_1,\dots,c_d}$ by G exists and is separated and of finite type over k, according to [6, Exposé V, Théorème 7.1].
It remains to construct the announced bijection. As

$$\mathcal{E}nd_{\mathcal{A}}(E)^{\mathrm{op}} \cong \mathcal{E}nd_{\mathcal{A}}(E \otimes L)^{\mathrm{op}} =: \mathcal{A}',$$

which is again an Azumaya algebra with generic fiber D, we obtain a well defined map from closed points of the quotient to isomorphism classes of such $\mathcal{A}'$. Conversely, given $\mathcal{A}, \mathcal{A}'$, the possible locally free $\mathcal{A}$-modules E of rank one with $\mathcal{E}nd_{\mathcal{A}}(E)^{\mathrm{op}} \cong \mathcal{A}'$ all differ only by tensoring with line bundles. This can be seen as follows:
Suppose that E and E' are locally free $\mathcal{A}$-modules of rank one with

$$\mathcal{E}nd_{\mathcal{A}}(E)^{\mathrm{op}} \cong \mathcal{E}nd_{\mathcal{A}}(E')^{\mathrm{op}} \cong \mathcal{A}'.$$

We choose embeddings of E and E' into $\mathcal{A}_\eta = D$; this also embeds $End_{\mathcal{A}}(E)^{\mathrm{op}}$ and $End_{\mathcal{A}}(E')^{\mathrm{op}}$ into D. The given isomorphism between them induces an automorphism of D, i. e. conjugation with an element of D; altering the embedding $E' \hookrightarrow D$ by a right multiplication with this element, we may assume that $End_{\mathcal{A}}(E)^{\mathrm{op}} = End_{\mathcal{A}}(E')^{\mathrm{op}} =: \mathcal{A}'$ as subalgebras of D.
There is an open subscheme $U \subseteq X$ such that $\mathcal{A}|_U = E|_U = E'|_U = \mathcal{A}'|_U$. Furthermore, $X \setminus U$ is a finite union of divisors $D_1, \dots D_l$; it is enough to study the question at the generic points x_i of the D_i. There the local ring $\mathcal{O}_{X,x_i}$ is a discrete valuation ring; over its completion $\hat{\mathcal{O}}_{X,x_i}$, $\mathcal{A}$ becomes isomorphic to a matrix algebra, so we can describe the situation using Morita equivalence as follows:
$\hat{E}_{x_i}$, $\hat{E}'_{x_i}$ correspond to lattices over $\hat{\mathcal{O}}_{X,x_i}$ in $F^r_{x_i}$ (F_{x_i} the completion of F at x_i) such that $\mathrm{End}_{\hat{\mathcal{O}}_{x_i}}(\hat{E}_{x_i})^{\mathrm{op}} = \mathrm{End}_{\hat{\mathcal{O}}_{x_i}}(\hat{E}'_{x_i})^{\mathrm{op}} = \hat{\mathcal{A}}'_{x_i}$. But then it is an easy exercise to see that $\hat{E}'_{x_i} = \pi_i^{N_i}\hat{E}_{x_i}$ for some $N_i \in \mathbb{Z}$ where $\pi_i \in \hat{\mathcal{O}}_{x_i}$ is a uniformising element. From this our claim follows.
This shows that the map above is injective. For the surjectivity, we consider two Azumaya algebras $\mathcal{A}, \mathcal{A}' \subseteq D$ and define $E(U) := \{f \in D : \mathcal{A}|_U \cdot f \subseteq \mathcal{A}'|_U\}$. Using Morita equivalence as above, it is easy to check that E is a locally free $\mathcal{A}$-module of rank one with $End_{\mathcal{A}}(E)^{\mathrm{op}} = \mathcal{A}'$; this proves the surjectivity. □

Remark 4.2. If $\mathcal{A}'$ is another sheaf of Azumaya algebras over X with generic fiber $\mathcal{A}'_\eta \cong D = \mathcal{A}_\eta$, then the moduli spaces $\mathrm{M}_{\mathcal{A}/X}$ and $\mathrm{M}_{\mathcal{A}'/X}$ are isomorphic. Indeed, the preceeding proof shows that there is a locally free $\mathcal{A}$-module E of rank one with $End_{\mathcal{A}}(E)^{\mathrm{op}} \cong \mathcal{A}'$; then E is a right $\mathcal{A}'$-module, and one checks easily that the functor $E \otimes_{\mathcal{A}'} \_$ defines an equivalence from left $\mathcal{A}'$-modules to left $\mathcal{A}$-modules.

Remark 4.3. If X is a surface, then this quotient can be decomposed explicitly into pieces of finite type as follows:
The action of $\mathrm{Pic}(X)$ preserves the discriminant $\Delta(E) \in \mathbb{Z}$, so we get a decomposition

$$\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X}/\mathrm{Pic}(X) = \coprod_{\Delta \in \mathbb{Z}} \mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X;\Delta}/\mathrm{Pic}(X).$$

Now the first Chern class $c_1(E) \in NS(X)$ decomposes $\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X;\Delta}$ into pieces of finite type. But $c_1(E \otimes L) = c_1(E) + r^2 c_1(L)$, and $r^2\mathrm{NS}(X)$ has finite index in $\mathrm{NS}(X)$, so $\mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X;\Delta}/\mathrm{Pic}(X)$ is indeed of finite type over k.
According to equation (7), fixing $\Delta(E)$ corresponds to fixing $c_2(\mathcal{A}') \in \mathbb{Z}$ where $\mathcal{A}' = End_{\mathcal{A}}(E)^{\mathrm{op}}$. If X is an abelian or K3 surface, then theorem 3.6.iv yields

$$\dim \mathrm{M}^{\mathrm{lp}}_{\mathcal{A}/X;\Delta}/\mathrm{Pic}(X) = c_2 - (r^2 - 1)\chi(\mathcal{O}_X)$$

where $c_2 = c_2(\mathcal{A}') = \Delta/r^2 - c_2(\mathcal{A}) \in \mathbb{Z}$ is the second Chern class of the Azumaya algebras $\mathcal{A}'$ that this quotient parameterizes.

Remark 4.4. M. Lieblich [13, 12] has compactified such moduli spaces of Azumaya algebras using his generalized Azumaya algebras, i. e. algebra objects in a derived category corresponding to endomorphism algebras of torsion free rank one $\mathcal{A}$-modules.

Remark 4.5. Since the automorphism group of the matrix algebra $\mathrm{Mat}(r \times r)$ is $\mathrm{PGL}(r)$, Azumaya algebras of rank r^2 correspond to principal $\mathrm{PGL}(r)$-bundles. Moduli spaces for the latter have recently been constructed and compactified by T. Gomez, A. Langer, A. Schmitt and I. Sols [5].

References

[1] I. V. Artamkin. On the deformation of sheaves. *Izv. Akad. Nauk SSSR*, 52(3):660–665, 1988. Translation: *Math. USSR Izv.*, 32(3):663–668, 1989.

[2] I. V. Artamkin. Deformation of torsion-free sheaves on an algebraic surface. *Izv. Akad. Nauk SSSR*, 54(3):435–468, 1990. Translation: *Math. USSR Izv.*, 36(3):449–485, 1991.

[3] M. Deuring. *Algebren.* Zweite, korrigierte Auflage. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 41. Springer-Verlag, Berlin, 1968.

[4] D. Eisenbud. *Commutative algebra. With a view toward algebraic geometry.* Springer-Verlag, New York, 1995.

[5] T. Gomez, A. Langer, A. Schmitt and I. Sols. Moduli Spaces for Principal Bundles in Arbitrary Characteristic. preprint math.AG/0506511.

[6] A. Grothendieck and M. Demazure. *SGA 3: Schémas en groupes. I: Propriétés générales des schémas en groupes.* Springer-Verlag, Berlin, 1970.

[7] D. Huybrechts and M. Lehn. *The geometry of moduli spaces of sheaves.* Vieweg-Verlag, Braunschweig, 1997.

[8] D. Huybrechts and P. Stellari. Proof of Caldararu's conjecture. An appendix to a paper by Yoshioka. preprint math.AG/0411541.

[9] A. Langer. Semistable sheaves in positive characteristic. *Ann. of Math. (2)*, 159(1):251–276, 2004.

[10] A. Langer. Moduli spaces of sheaves in mixed characteristic. *Duke Math. J.*, 124(3):571–586, 2004.

[11] S. G. Langton. Valuative criteria for families of vector bundles on algebraic varieties. *Ann. of Math. (2)*, 101:88–110, 1975.

[12] M. Lieblich. Moduli of twisted sheaves. preprint math.AG/0411337.

[13] M. Lieblich. *Moduli of Twisted Sheaves and Generalized Azumaya Algebras.* PhD thesis, Massachusetts Institute of Technology, 2004.

[14] S. Mukai. Symplectic structure of the moduli space of sheaves on an abelian or $K3$ surface. *Invent. Math.*, 77(1):101–116, 1984.

[15] D. Mumford and J. Fogarty. *Geometric invariant theory. Second Edition.* Springer-Verlag, Berlin, 1982.

[16] C. T. Simpson. Moduli of representations of the fundamental group of a smooth projective variety. I. *Publ. Math. IHES*, 79:47–129, 1994.

[17] K. Yoshioka. Moduli spaces of twisted sheaves on a projective variety. preprint math.AG/0411538.

Norbert Hoffmann
Mathematisches Institut
Bunsenstraße 3–5
D-37073 Göttingen
Germany
hoffmann@uni-math.gwdg.de

Ulrich Stuhler
Mathematisches Institut
Bunsenstraße 3–5
D-37073 Göttingen
Germany
stuhler@uni-math.gwdg.de

Parametrized Braid Groups of Chevalley Groups

Jean-Louis Loday[1] and Michael R. Stein[2]

Received: February 17, 2005

Communicated by Ulf Rehmann

Abstract. We introduce the notion of a braid group parametrized by a ring, which is defined by generators and relations and based on the geometric idea of painted braids. We show that the parametrized braid group is isomorphic to the semi-direct product of the Steinberg group (of the ring) with the classical braid group. The technical heart of the proof is the Pure Braid Lemma, which asserts that certain elements of the parametrized braid group commute with the pure braid group. This first part treats the case of the root system A_n; in the second part we prove a similar theorem for the root system D_n.

2000 Mathematics Subject Classification: Primary 20F36; Secondary 19Cxx; 20F55.
Keywords and Phrases: Braid group, Steinberg group, parametrized braid group, root system.

1 Introduction

Suppose that the strands of a braid are painted and that the paint from a strand spills onto the strand beneath it, modifying the color of the lower strand as in the picture below.

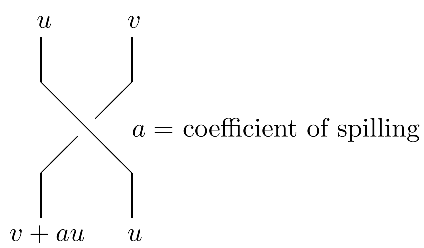


Figure 1

[1]Partially supported by Northwestern University in 2002.
[2]Partially supported by Université Louis Pasteur in 1998-99.

This gives rise, for any ring A, to the *parametrized braid group* $Br_n(A)$, which is generated by elements y_i^a, where i is an integer, $1 \leq i \leq n-1$, and a is an element of A, subject to the relations

$$\begin{array}{rcl} (A1) & y_i^a y_i^0 y_i^b = y_i^0 y_i^0 y_i^{a+b} & \\ (A1 \times A1) & y_i^a y_j^b = y_j^b y_i^a & \text{if } |i-j| \geq 2, \\ (A2) & y_i^a y_{i+1}^b y_i^c = y_{i+1}^c y_i^{b+ac}\, y_{i+1}^a & \end{array}$$

for any $a, b, c \in A$

A variation of this group first appeared in [L]. The choice of names for these relations will be explained in section 2 below. The derivation of these relations from the painted braid model can be seen in Figures 2 and 3 below.)
Observe that when $A = \{0\}$ (the zero ring), one obtains the classical Artin braid group Br_n, whose presentation is by generators y_i, $1 \leq i \leq n-1$, and relations

$$\begin{array}{rcl} (A1 \times A1) & y_i y_j = y_j y_i & \text{if } |i-j| \geq 2, \\ (A2) & y_i y_{i+1} y_i = y_{i+1} y_i\, y_{i+1} & \end{array}$$

A question immediately comes to mind: does Figure 1 correctly reflect the elements of the parametrized braid group? Up to equivalence, a picture would be completely determined by a braid and a linear transformation of the set of colors. This linear transformation lies in the subgroup $E_n(A)$ of elementary matrices. Hence if the elements of the parametrized braid group correspond exactly to the pictures, then the group should be the semi-direct product of $E_n(A)$ by Br_n. We will show that this is almost the case: we only need to replace $E_n(A)$ by the Steinberg group $St_n(A)$ (*cf.* [St] [Stb]).

Theorem. *For any ring A there is an isomorphism*

$$Br_n(A) \cong St_n(A) \rtimes Br_n \ ,$$

where the action of Br_n is via the symmetric group $\mathcal{S}_n$.

The quotient of $Br_n(A)$ by the relation $y_i^0 y_i^0 = 1$ is the group studied by Kassel and Reutenauer [K-R] (in this quotient group, our relation (A1) becomes $y_i^a y_i^0 y_i^b = y_i^{a+b}$, which is exactly the relation used in [K-R] in place of (A1)). They show that this quotient is naturally isomorphic to the semi-direct product $St_n(A) \rtimes \mathcal{S}_n$ of the Steinberg group with the symmetric group. So our theorem is a lifting of theirs.

The proof consists in constructing maps both ways. The key point about their existence is a technical result called the Pure Braid Lemma (cf. 2.2.1). It says that a certain type of parametrized braid commute with the pure braid group.

In the first part of the paper we give a proof of the above theorem which corresponds to the family of Coxeter groups A_n. In the second part we prove a similar theorem for the family D_n, cf. section 3. The Pure Braid Lemma in the D_n case is based on a family of generators of the pure braid group found by Digne and Gomi [D-G]. We expect to prove similar theorems for the other Coxeter groups.
A related result has been announced in [Bon].

2 The parametrized braid group

We introduce the parametrized braid group of the family of Coxeter groups A_{n-1}.

Definition 2.0.1. *Let A be a ring (not necessarily unital nor commutative). The parametrized braid group $Br_n(A)$ is generated by the elements y_i^a, where i is an integer, $1 \leq i \leq n-1$, and a is an element of A, subject to the relations*

$$\begin{array}{lll} (A1) & y_i^a y_i^0 y_i^b = y_i^0 y_i^0 y_i^{a+b} & \\ (A1 \times A1) & y_i^a y_j^b = y_j^b y_i^a & \text{if } |i-j| \geq 2, \\ (A2) & y_i^a y_{i+1}^b y_i^c = y_{i+1}^c y_i^{b+ac}\, y_{i+1}^a & \end{array}$$

for any $a, b, c \in A$.

The geometric motivation for the defining relations of this group, and its connection with braids, can be seen in the following figures in which u, v, w (the colors) are elements of A, and a, b, c are the *coefficients of spilling*. Relation $(A1)$ comes from Figure 2.
Relation $(A1 \times A1)$ arises because the actions of y_i^a and of y_j^b on the strands of the braid are disjoint when $|i-j| \geq 2$, so that these two elements commute.
Relation $(A2)$ derives from Figure 3.

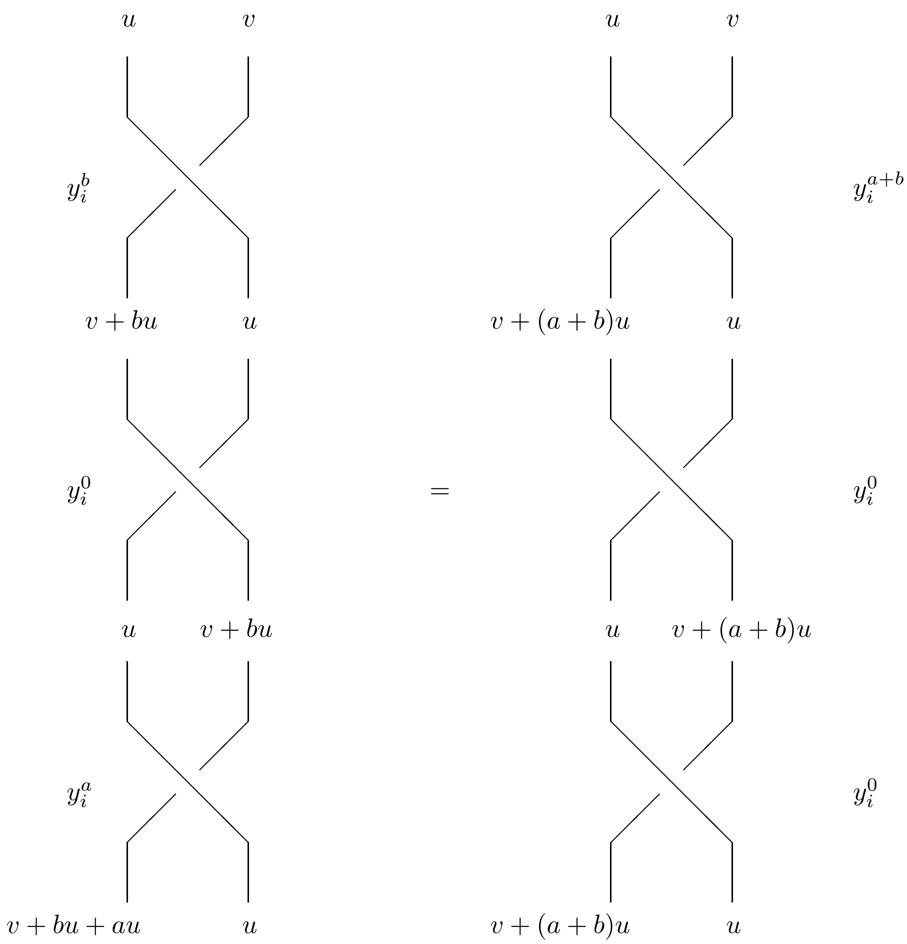

u
v
u
v
y_i^b
y_i^{a+b}
$v+bu$
u
$v+(a+b)u$
u
y_i^0
$=$
y_i^0
u
$v+bu$
u
$v+(a+b)u$
y_i^a
y_i^0
$v+bu+au$
u
$v+(a+b)u$
u

Figure 2

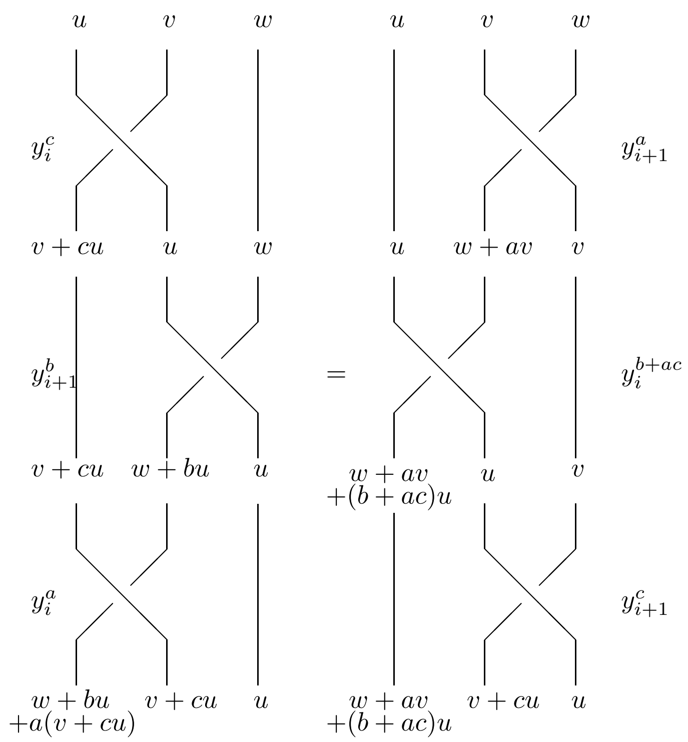


Figure 3

2.1 Technical lemmas

Let Φ be the root system A_n or D_n and let A be a ring (not necessarily unital) which is supposed to be commutative in the D_n case. Let Δ be a simple subsystem of the root system Φ. Elements of Δ are denoted by α or α_i. The image of $b \in Br(\Phi)$ in the Weyl group $W(\Phi)$ is denoted $\bar{b}$. The parametrized braid group $Br(D_n, A)$ is defined in 3.1.1.

Lemma 2.1.1. *The following relations in $Br(\Phi, A)$ are consequences of relation $(A1)$:*

$$\begin{array}{rcl} (y_\alpha^0 y_\alpha^0) y_\alpha^a &=& y_\alpha^a (y_\alpha^0 y_\alpha^0), \\ y_\alpha^a (y_\alpha^0)^{-1} y_\alpha^b &=& y_\alpha^{a+b}, \\ y_\alpha^{-a} &=& y_\alpha^0 (y_\alpha^a)^{-1} y_\alpha^0. \end{array}$$

Proof. Replacing b by 0 in $(A1)$ shows that $y_\alpha^0 y_\alpha^0$ commutes with y_α^a. From this follows

$$\begin{array}{rcl} y_\alpha^a (y_\alpha^0)^{-1} y_\alpha^b &=& (y_\alpha^0 y_\alpha^0)^{-1} y_\alpha^a (y_\alpha^0 y_\alpha^0)(y_\alpha^0)^{-1} y_\alpha^b \\ &=& (y_\alpha^0 y_\alpha^0)^{-1} y_\alpha^a y_\alpha^0 y_\alpha^b \\ &=& (y_\alpha^0 y_\alpha^0)^{-1} y_\alpha^0 y_\alpha^0 y_\alpha^{a+b} \\ &=& y_\alpha^{a+b}. \end{array}$$

Putting $b = -a$ in the second relation yields the third relation. □

Lemma 2.1.2. *Assume that the Pure Braid Lemma holds for the root system Φ. Suppose that $\alpha \in \Delta$ and $b \in Br(\Phi)$ are such that $\bar{b}(\alpha) \in \Delta$ (where $\bar{b} \in W(\Phi)$ denotes the image of b). Then for any $a \in A$ one has*

$$by^a_\alpha (y^0_\alpha)^{-1} b^{-1} = y^a_{\bar{b}(\alpha)} (y^0_{\bar{b}(\alpha)})^{-1} \quad \textit{in } Br(\Phi, A).$$

Proof. First let us show that there exists $b' \in Br(\Phi)$ such that $b' y^a_\alpha b'^{-1} = y^a_{\bar{b}(\alpha)}$. The two roots α and $\bar{b}(\alpha)$ have the same length, hence they are connected, in the Dynkin diagram, by a finite sequence of edges with $m = 3$ (cf. [Car, Lemma 3.6.3]). Therefore it is sufficient to prove the existence of b' when α and $\bar{b}(\alpha)$ are adjacent. In that case α and $\bar{b}(\alpha)$ generate a subsystem of type A_2; we may assume $\alpha = \alpha_1$ and $\bar{b}(\alpha) = \alpha_2 \in A_2$; and we can use the particular case of relation $(A2)$, namely

$$y^0_{\alpha_1} y^0_{\alpha_2} y^a_{\alpha_1} = y^a_{\alpha_2} y^0_{\alpha_1} y^0_{\alpha_2}$$

to show that

$$y^0_{\alpha_1} y^0_{\alpha_2} y^a_{\alpha_1} (y^0_{\alpha_2})^{-1} (y^0_{\alpha_1})^{-1} = y^a_{\alpha_2}.$$

(Here $b' = y^0_{\alpha_1} y^0_{\alpha_2}$, $\alpha = \alpha_1 = -\epsilon_1 + \epsilon_2, \bar{b'}(\alpha) = \alpha_2 = -\epsilon_2 + \epsilon_3$.)
To conclude the proof of the Lemma it is sufficient to show that

$$by^a_\alpha (y^0_\alpha)^{-1} b^{-1} = y^a_\alpha (y^0_\alpha)^{-1}$$

whenever $b(\alpha) = \alpha$. According to [H, Theorem, p. 22], $\bar{b}$ is a product of simple reflections σ_{α_i} for $\alpha_i \in \Delta$ which are not connected to α in the Dynkin diagram of Δ. Hence we can write b as the product of an element in the pure braid group and generators y_{α_i} which commute with y^a_α by relation $(A1 \times A1)$. Since we have assumed that the Pure Braid Lemma holds for Φ, we can thus conclude that $by^a_\alpha (y^0_\alpha)^{-1} b^{-1} = y^a_\alpha (y^0_\alpha)^{-1}$ as desired. □

2.2 Braid group and pure braid group

The group $Br_n(0) = Br_n$ is the classical Artin braid group with generators $y_i\,, 1 \leq i \leq n-1$, and relations

$$\begin{aligned} y_i\, y_j &= y_j\, y_i\,, \qquad |i-j| \geq 2, \\ y_i\, y_{i+1}\, y_i &= y_{i+1}\, y_i\, y_{i+1}\,. \end{aligned}$$

The quotient of Br_n by the relations $y_i y_i = 1\,, 1 \leq i \leq n-1$ is the symmetric group $\mathcal{S}_n$; the image of $b \in Br_n$ in $\mathcal{S}_n$ is denoted by $\bar{b}$. The kernel of the surjective homomorphism $Br_n \to \mathcal{S}_n$ is the *pure braid group*, denoted PBr_n. It is generated by the elements

$$\mathbf{a}_{j,i} := y_j y_{j-1} \cdots y_i y_i \cdots y_{j-1} y_j,$$

for $n \geq j \geq i \geq 1$, ([Bir]; see Figure 4 below).

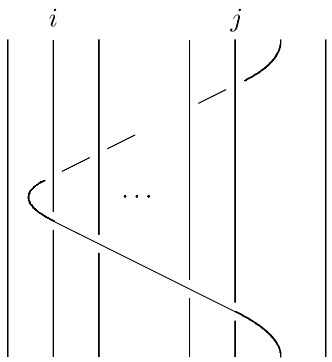


Figure 4: the pure braid $\mathbf{a}_{j,i}$

Lemma 2.2.1 (Pure Braid Lemma for A_{n-1}). *Let y_k^a be a generator of $Br_n(A)$ and let $\omega \in PBr_n = PBr_n(0)$. Then there exists $\omega' \in Br_n$, independent of a, such that*

$$y_k^a\ \omega = \omega'\ y_k^a.$$

Hence for any integer k and any element $a \in A$, the element $y_k^a(y_k^0)^{-1} \in PBr_n(A)$ commutes with every element of the pure braid group PBr_n.

Notation. Before beginning the proof of Lemma 2.2.1, we want to simplify our notation. We will abbreviate

$$\begin{array}{rcl} y_k^0 & = & \mathbf{k} \\ (y_k^0)^{-1} & = & \mathbf{k}^{-1} \\ y_k^a & = & \mathbf{k}^a \end{array}$$

Note that $\mathbf{k}^{-1}$ does *not* mean $\mathbf{k}^a$ for $a = -1$.
If there exist ω' and $\omega'' \in Br_n$ such that $\mathbf{k}^a\omega = \omega'\mathbf{j}^a\omega''$, we will write

$$\mathbf{k}^a\omega \sim \mathbf{j}^a\omega''.$$

Observe that $\sim$ is *not* an equivalence relation, but it is compatible with multiplication on the right by elements of Br_n: if $\eta \in Br_n$,

$$\mathbf{k}^a\omega \sim \mathbf{j}^a\omega'' \Leftrightarrow \mathbf{k}^a\omega = \omega'\mathbf{j}^a\omega'' \Leftrightarrow \mathbf{k}^a\omega\eta = \omega'\mathbf{j}^a\omega''\eta \Leftrightarrow \mathbf{k}^a\omega\eta \sim \mathbf{j}^a\omega''\eta$$

For instance

$$\begin{array}{rcll} \mathbf{k}^a\ \mathbf{k}\ \mathbf{k} & \sim & \mathbf{k}^a & \text{by Lemma 2.1.1,} \\ \mathbf{k}^a\ (\mathbf{k-1})\ \mathbf{k} & \sim & (\mathbf{k-1})^a & \text{by relation (A2),} \\ \mathbf{k}^a\ (\mathbf{k-1}) & \sim & (\mathbf{k-1})^a\ \mathbf{k} & \text{by relation (A2).} \end{array}$$

Proof of Lemma 2.2.1. It is clear that the first assertion implies the second one: since ω' is independent of a, we can set $a = 0$ to determine that $\omega' = y_k^0\ \omega\ (y_k^0)^{-1}$. Substituting this value of ω' in the expression $y_k^a\ \omega = \omega'\ y_k^a$ completes the proof.

To prove the first assertion, we must show, in the notation just introduced, that $\mathbf{k}^a\omega \sim \mathbf{k}^a$ for every $\omega \in PBr_n$, and it suffices to show this when ω is one of the generators $\mathbf{a}_{j,i} = \mathbf{j}\ (\mathbf{j-1})\ \cdots \mathbf{i}\ \mathbf{i} \cdots\ (\mathbf{j-1})\ \mathbf{j}$ above. Since y_k^a commutes with y_i^0 for all $i \neq k-1, k, k+1$, it commutes with $\mathbf{a}_{j,i}$ whenever $k < i-1$ or whenever $k > j+1$. So we are left with the following 3 cases:

$$\begin{array}{llcll}
(1a) & (\mathbf{j+1})^a\mathbf{a}_{j,i} & \sim & (\mathbf{j+1})^a & \\
(1b) & \mathbf{j}^a\mathbf{a}_{j,i} & \sim & \mathbf{j}^a & \\
(1c) & \mathbf{k}^a\mathbf{a}_{j,i} & \sim & \mathbf{k}^a & i-1 \leq k \leq j-1
\end{array}$$

Our proof of case (1a) is by induction on the half-length of $\omega = \mathbf{a}_{j,i}$. When $i = k-1$, $\omega = (\mathbf{k-1})\ (\mathbf{k-1})$, and we have

$$\begin{aligned}
\mathbf{k}^a\ (\mathbf{k-1})\ (\mathbf{k-1}) &= \mathbf{k}^a\ (\mathbf{k-1})\ \mathbf{k}^{-1}\,\mathbf{k}\ (\mathbf{k-1}) \\
&\sim (\mathbf{k-1})^a\ \mathbf{k}\ (\mathbf{k-1}) \\
&\sim \mathbf{k}^a\ .
\end{aligned}$$

and more generally,

$$\begin{aligned}
\mathbf{k}^a\ (\mathbf{k-1})\ (\mathbf{k-2}) \cdots (\mathbf{k-2})\ (\mathbf{k-1}) & \\
&\sim (\mathbf{k-1})^a\ \mathbf{k}\ (\mathbf{k-2}) \cdots (\mathbf{k-2})(\mathbf{k-1}) \\
&= (\mathbf{k-1})^a\ (\mathbf{k-2}) \cdots (\mathbf{k-2})\ \mathbf{k}\ (\mathbf{k-1}) \\
&\sim (\mathbf{k-1})^a\ \mathbf{k}\ (\mathbf{k-1}) \qquad \text{(by induction)}, \\
&\sim \mathbf{k}^a\ .
\end{aligned}$$

The proof of case (1b) is also by induction on the half-length of $\omega = \mathbf{a}_{j,i}$. When $i = k$, $\omega = \mathbf{k}\ \mathbf{k}$, and we have

$$\mathbf{k}^a\mathbf{k}\ \mathbf{k} \sim \mathbf{k}^a \qquad \text{by Lemma 2.1.1.}$$

Then

$$\begin{aligned}
\mathbf{k}^a\mathbf{k}\ (\mathbf{k-1}) \cdots (\mathbf{k-1})\ \mathbf{k} & \\
&= \mathbf{k}^a\mathbf{k}\ (\mathbf{k-1})\ \mathbf{k}\ \mathbf{k}^{-1}\ (\mathbf{k-2}) \cdots (\mathbf{k-2})\ (\mathbf{k-1})\ \mathbf{k} \\
&\sim \mathbf{k}^a(\mathbf{k-1})\ \mathbf{k}\ (\mathbf{k-1})\ (\mathbf{k-2}) \cdots (\mathbf{k-2})\ \mathbf{k}^{-1}\,(\mathbf{k-1})\ \mathbf{k} \\
&\sim (\mathbf{k-1})^a(\mathbf{k-1})\ (\mathbf{k-2}) \cdots (\mathbf{k-2})\ (\mathbf{k-1})\ \mathbf{k}\ (\mathbf{k-1})^{-1} \\
&\sim (\mathbf{k-1})^a\ \mathbf{k}\ (\mathbf{k-1})^{-1} \qquad \text{(by induction)} \\
&\sim \mathbf{k}^a.
\end{aligned}$$

For case (1c) it is sufficient to check the cases

$$\begin{array}{llcl}
(2a) & \omega & = & (\mathbf{k+1})\ (\mathbf{k+1}) \\
(2b) & \omega & = & (\mathbf{k+1})\ \mathbf{k}\ \mathbf{k}\ (\mathbf{k+1}) \\
(2c) & \omega & = & (\mathbf{k+1})\ \mathbf{k}\ (\mathbf{k-1}) \cdots (\mathbf{k-1})\ \mathbf{k}\ (\mathbf{k+1})
\end{array}$$

which are proved as follows:

$$\begin{array}{rcl}\mathbf{k}^a(\mathbf{k+1})\ (\mathbf{k+1}) & = & \mathbf{k}^a(\mathbf{k+1})\ \mathbf{k}^{-1}\,\mathbf{k}\ (\mathbf{k+1}) \\ & \sim & (\mathbf{k+1})^a\ \mathbf{k}\ (\mathbf{k+1}) \\ & \sim & \mathbf{k}^a \qquad \text{Case (2a)}\end{array}$$

$$\begin{array}{rcl}\mathbf{k}^a(\mathbf{k+1})\mathbf{kk}(\mathbf{k+1}) & \sim & (\mathbf{k+1})^a\mathbf{k}\ (\mathbf{k+1}) \\ & \sim & \mathbf{k}^a \qquad \text{Case (2b)}\end{array}$$

$$\begin{array}{l}\mathbf{k}^a(\mathbf{k+1})\ \mathbf{k}\ (\mathbf{k-1})\cdots(\mathbf{k-1})\ \mathbf{k}\ (\mathbf{k+1}) \\ \qquad \sim\ (\mathbf{k+1})^a(\mathbf{k-1})\cdots(\mathbf{k-1})\ \mathbf{k}\ (\mathbf{k+1}) \\ \qquad \sim\ (\mathbf{k+1})^a\mathbf{k}\ (\mathbf{k+1}) \\ \qquad \sim\ \mathbf{k}^a \qquad \text{Case (2c)}\end{array}$$

□

Proposition 2.2.2. *For any $\omega \in Br_n(0)$ the element $\omega y_k^a (y_k^0)^{-1}\omega^{-1}$ depends only on the class $\bar{\omega}$ of ω in $\mathcal{S}_n$. Moreover if $\bar{\omega}(j) = k$ and $\bar{\omega}(j+1) = k+1$, then $\omega y_k^a (y_k^0)^{-1}\omega^{-1} = y_j^a (y_j^0)^{-1}$.*

Proof. The first statement is a consequence of Lemma 2.2.1 since the Weyl group $W(A_{n-1}) = \mathcal{S}_n$ is the quotient of Br_n by PBr_n.
The second part is a consequence of relation $(A1 \times A1)$ and the following computation:

$$\begin{array}{rcl}\mathbf{2}^{-1}\mathbf{1}^{-1}\mathbf{2}^a\mathbf{12} & = & \mathbf{1}^a\mathbf{2}^{-1}\mathbf{1}^{-1}\mathbf{2}^{-1}\mathbf{12} \\ & = & \mathbf{1}^a\mathbf{1}^{-1}\mathbf{2}^{-1}\mathbf{1}^{-1}\mathbf{12} \\ & = & \mathbf{1}^a\mathbf{1}^{-1}\ .\end{array}$$

□

2.3 The Steinberg group and an action of the Weyl group

When $\Phi = A_{n-1}$, the Steinberg group of the ring A is well-known and customarily denoted $St_n(A)$. In that case it is customary to write $x_{ii+1}^a = x_{ii+1}(a)$ for the element $x_\alpha(a), \alpha = \epsilon_i - \epsilon_{i+1} \in \Delta$, and, more generally, x_{ij}^a when $\alpha = \epsilon_i - \epsilon_j \in A_{n-1}$.

Definition 2.3.1 ([Stb]). *The Steinberg group of the ring A, denoted $St_n(A)$, is presented by the generators $x_{ij}^a, 1 \le i,j \le n, i \ne j, a \in A$ subject to the relations*

$$\begin{array}{lll}(St0) & x_{ij}^a x_{ij}^b = x_{ij}^{a+b} & \\ (St1) & x_{ij}^a x_{kl}^b = x_{kl}^b x_{ij}^a, & i \ne l, j \ne k \\ (St2) & x_{ij}^a x_{jk}^b = x_{jk}^b x_{ik}^{ab} x_{ij}^a, & i \ne k.\end{array}$$

We should make two observations about this definition. First, it follows from $(St0)$ that $x^0_{ij} = 1$. Second, relation $(St2)$ is given in a perhaps unfamiliar form. We have chosen this form, which is easily seen to be equivalent to $(R2)$, because of its geometric significance (*cf.* [K-S] for the relationship with the Stasheff polytope), and for the simplification it brings in computation.

The Weyl group $W(A_{n-1})$ is the symmetric group $\mathcal{S}_n$. Its action on the Steinberg group is induced by the formula

$$(3) \qquad \sigma \cdot x^a_{ij} := x^a_{\sigma(i)\sigma(j)}, \quad \sigma \in \mathcal{S}_n, a \in A$$

2.4 The main result

Theorem 2.4.1. *For any (not necessarily unital) ring A the map*

$$\phi : Br_n(A) \to St_n(A) \rtimes Br_n$$

from the parametrized braid group to the semi-direct product of the Artin braid group with the Steinberg group induced by $\phi(y^a_i) = x^a_{i\,i+1}\, y_i$ is a group isomorphism.

Proof. Step (a). We show that ϕ is a well-defined group homomorphism.

• Relation (A1):

$$\begin{array}{rcll} \phi(y^a_i y^0_i y^b_i) & = & x^a_{i\,i+1} y_i\, y_i\, x^b_{i\,i+1} y_i\,, & \text{since} \quad x^0_{i\,i+1} = 1, \\ & = & y_i\, y_i\, x^a_{i\,i+1} x^b_{i\,i+1} y_i & \text{since} \quad \overline{y_i\, y_i} = 1 \in \mathcal{S}_n, \\ & = & y_i\, y_i\, x^{a+b}_{i\,i+1} y_i & \text{by } (St0), \\ & = & \phi(y^0_i y^0_i y^{a+b}_i). & \end{array}$$

• Relation $(A1 \times A1)$ follows immediately from $(St1)$.
• Relation $(A2)$ is proved by using the relations of Br_n and the 3 relations $(St0), (St1), (St2)$ as follows:

$$\begin{array}{rcl} \phi(y^a_i y^b_{i+1} y^c_i) & = & x^a_{i\,i+1} y_i\, x^b_{i+1\,i+2}\, \underbrace{y_{i+1}\, x^c_{i\,i+1}}\, y_i \\ & = & x^a_{i\,i+1} y_i\, \underbrace{x^b_{i+1\,i+2} x^c_{i\,i+2}}\, y_{i+1}\, y_i\,, \\ & = & x^a_{i\,i+1}\, \underbrace{y_i\, x^c_{i\,i+2}}\, x^b_{i+1\,i+2}\, y_{i+1}\, y_i\,, \\ & = & \underbrace{x^a_{i\,i+1} x^c_{i+1\,i+2}}\, \underbrace{y_i\, x^b_{i+1\,i+2}}\, y_{i+1}\, y_i\,, \\ & = & x^c_{i+1\,i+2} x^{ac}_{i\,i+2}\, \underbrace{x^a_{i\,i+1} x^b_{i\,i+2}}\, \underbrace{y_i\, y_{i+1}\, y_i}\,, \\ & = & x^c_{i+1\,i+2} x^{ac+b}_{i\,i+2}\, \underbrace{x^a_{i\,i+1} y_{i+1}}\, y_i\, y_{i+1}\,, \\ & = & x^c_{i+1\,i+2}\, \underbrace{x^{ac+b}_{i\,i+2}\, y_{i+1}}\, \underbrace{x^a_{i\,i+2}\, y_i\, y_{i+1}}\,, \\ & = & \underbrace{x^c_{i+1\,i+2} y_{i+1}}\, \underbrace{x^{ac+b}_{i\,i+1} y_i}\, \underbrace{x^a_{i+1\,i+2}\, y_{i+1}}\,, \\ & = & \phi(y^c_{i+1}\, y^{b+ac}_i\, y^a_{i+1}). \end{array}$$

Step (b). This is the Pure Braid Lemma 2.2.1 for A_n.

Step (c). We construct a homomorphism $\psi : St_n(A) \rtimes Br_n \to Br_n(A)$. We first construct $\psi : St_n(A) \to \operatorname{Ker} \pi$, where π is the surjection $Br_n(A) \to Br_n$, by setting

$$\psi(x_{ij}^a) \quad := \quad \omega\ y_k^a (y_k^0)^{-1}\ \omega^{-1}$$

where ω is an element of Br_n such that $\bar{\omega}(k) = i$ and $\bar{\omega}(k+1) = j$ (for instance, $\psi(x_{12}^a) = y_1^a (y_1^0)^{-1}$ and $\psi(x_{13}^a) = y_2^0 (y_1^a (y_1^0)^{-1})(y_2^0)^{-1}$). Observe that this definition does not depend on the choice of $\bar{\omega}$ (by Lemma 2.1.2), and does not depend on how we choose a lifting ω of $\bar{\omega}$ (by the Pure Braid Lemma 2.2.1). In order to show that ψ is a homomorphism, we must demonstrate that the Steinberg relations are preserved.

• Relation $(St0)$: it suffices to show that $\psi(x_{12}^a x_{12}^b) = \psi(x_{12}^{a+b})$,

$$\begin{array}{lll}
\psi(x_{12}^a x_{12}^b) & = y_1^a (y_1^0)^{-1} y_1^b (y_1^0)^{-1} = y_1^a (y_1^0)^{-2} y_1^0 y_1^b (y_1^0)^{-1} & \\
& = (y_1^0)^{-2} y_1^a y_1^0 y_1^b (y_1^0)^{-1} & \text{by 2.2.1,} \\
& = y_1^{a+b} (y_1^0)^{-1} & \text{by (A1),} \\
& = \psi(x_{12}^{a+b}). &
\end{array}$$

• Relation $(St1)$: it suffices to show that $\psi(x_{12}^a x_{34}^b) = \psi(x_{34}^b x_{12}^a)$ and that $\psi(x_{12}^a x_{13}^b) = \psi(x_{13}^b x_{12}^a)$. The first case is an immediate consequence of the Pure Braid Lemma 2.2.1 and of relation $(A1 \times A1)$. Let us prove the second case, which relies on the Pure Braid Lemma 2.2.1 and relation $(A2)$:

$$\begin{array}{rcl}
\psi(x_{12}^a x_{13}^b) & = & \underbrace{y_1^a (y_1^0)^{-1}}_{} y_2^0 y_1^b (y_1^0)^{-1} (y_2^0)^{-1} \\
& = & (y_1^0)^{-2} y_1^a \underbrace{y_1^0 y_2^0 y_1^b}_{} (y_1^0)^{-1} (y_2^0)^{-1} \\
& = & (y_1^0)^{-2} \underbrace{y_1^a y_2^b y_1^0}_{} \underbrace{y_2^0 (y_1^0)^{-1} (y_2^0)^{-1}}_{} \\
& = & (y_1^0)^{-2} y_2^0 y_1^b \underbrace{y_2^a (y_1^0)^{-1} (y_2^0)^{-1}}_{} y_1^0 \\
& = & (y_1^0)^{-2} y_2^0 y_1^b (y_1^0)^{-1} (y_2^0)^{-1} \underbrace{y_1^a y_1^0}_{} \\
& = & \underbrace{(y_1^0)^{-2} y_2^0 y_1^b (y_1^0)^{-1} (y_2^0)^{-1} (y_1^0)^2}_{} \underbrace{y_1^a (y_1^0)^{-1}}_{} \\
& = & \psi(x_{13}^b)\psi(x_{12}^a).
\end{array}$$

• Relation $(St2)$: it suffices to show that $\psi(x_{12}^a x_{23}^b) = \psi(x_{23}^b x_{13}^{ab} x_{12}^a)$.

$$\begin{array}{rcl}
\psi(x_{12}^a x_{23}^b) & = & y_1^a (y_1^0)^{-1} \underbrace{y_2^b}_{} (y_2^0)^{-1} \\
& = & y_1^a \underbrace{(y_1^0)^{-1} y_2^b y_1^0}_{} (y_1^0)^{-1} (y_2^0)^{-1} \\
& = & \underbrace{y_1^a y_2^0 y_1^b}_{} \underbrace{(y_2^0)^{-1} (y_1^0)^{-1} (y_2^0)^{-1}}_{} \\
& = & y_2^b \underbrace{\qquad}_{} y_1^{ab} \underbrace{\qquad}_{} y_2^a (y_1^0)^{-1} (y_2^0)^{-1} (y_1^0)^{-1} \\
& = & y_2^b (y_2^0)^{-1} y_2^0 y_1^{ab} (y_1^0)^{-1} \underbrace{y_1^0 y_2^a (y_1^0)^{-1} (y_2^0)^{-1}}_{} (y_1^0)^{-1} \\
& = & \underbrace{y_2^b (y_2^0)^{-1}}_{} \underbrace{y_2^0 y_1^{ab} (y_1^0)^{-1} (y_2^0)^{-1}}_{} \underbrace{y_1^a (y_1^0)^{-1}}_{} \\
& = & \psi(x_{23}^b)\psi(x_{13}^{ab})\psi(x_{12}^a),
\end{array}$$

as a consequence of relation $(A2)$.
From 2.2.2 it follows that the action of an element of Br_n by conjugation on Ker π depends only on its class in $\mathcal{S}_n$. The definition of ψ on $St_n(A)$ makes clear that it is an $\mathcal{S}_n$-equivariant map.
Defining ψ on Br_n by $\psi(y_\alpha) = y^0_\alpha \in Br_n(0)$ yields a group homomorphism

$$\psi : St_n(A) \rtimes Br_n \to \text{Ker } \pi \rtimes Br_n = Br_n(A).$$

The group homomorphisms ϕ and ψ are clearly inverse to each other since they interchange y^a_α and $x^a_\alpha y_\alpha$. Hence they are both isomorphisms, as asserted. $\square$

Corollary 2.4.2 (Kassel-Reutenauer [K-R]). *The group presented by generators y^a_i, $1 \leq i \leq n-1$, $a \in A$, and relations*

$$\begin{aligned} (y^0_i)^2 &= 1 \\ y^a_i (y^0_i)^{-1} y^b_i &= y^{a+b}_i \\ y^a_i y^b_j &= y^b_j y^a_i \qquad \text{if } |i-j| \geq 2 \\ y^a_i y^b_{i+1} y^c_i &= y^c_{i+1} y^{b+ac}_i \, y^a_{i+1} \end{aligned}$$

$a, b, c \in A$, is isomorphic to the semi-direct product $St_n(A) \rtimes \mathcal{S}_n$.

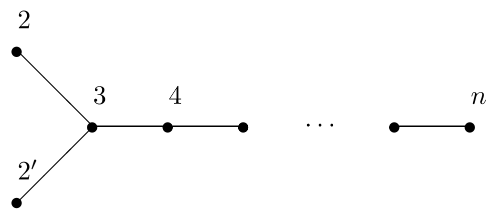
Observe that when the first relation in this Corollary is deleted, the second relation has several possible non-equivalent liftings. The one we have chosen, $(A1)$, is what allows us to prove Theorem 2.4.1.

3 The parametrized braid group in the D_n case

In this section we discuss the parametrized braid group $Br(D_n, A)$ for a commutative ring A and prove that it is isomorphic to the semi-direct product of the Steinberg group $St(D_n, A)$ by the braid group $Br(D_n, 0)$.

3.1 The braid group and the parametrized braid group for D_n

Let $\Delta = \{\alpha_2, \alpha_{2'}, \alpha_3, \ldots, \alpha_n\}$ be a fixed simple subsystem of a root system of type $D_n, n \geq 3$. We adopt the notation of [D-G] in which the simple roots on the fork of D_n are labeled $\alpha_2, \alpha_{2'}$. The system D_n contains 2 subsystems of type A_{n-1} generated by the simple subsystems $\{\alpha_2, \alpha_3, \ldots, \alpha_n\}$ and $\{\alpha_{2'}, \alpha_3, \ldots, \alpha_n\}$, and, for $n \geq 4$, a subsystem of type D_{n-1} generated by the simple subsystem $\{\alpha_2, \alpha_{2'}, \alpha_3, \ldots, \alpha_{n-1}\}$.



Dynkin diagram of D_n

The *Weyl group* $W(D_n)$ is generated by the simple reflections $\{\sigma_i = \sigma_{\alpha_i} | \alpha_i \in \Delta\}$, with defining relations

$$\begin{aligned} \sigma_i^2 &= 1 \\ (\sigma_i\sigma_j)^{m(i,j)} &= 1 \end{aligned}$$

for $i, j \in \{2, 2', 3, \ldots, n\}$, where

$$m(i,j) = \begin{cases} 2 & \text{if } \alpha_i, \alpha_j \text{ are not connected in the Dynkin diagram,} \\ 3 & \text{if } \alpha_i, \alpha_j \text{ are connected in the Dynkin diagram} \end{cases}$$

Since the only values for $m(\alpha, \beta)$ are $1, 2$ and 3, the group $Br(D_n, A)$ involves only relations $(A1), (A1 \times A1)$ and $(A2)$.

Definition 3.1.1. *The parametrized braid group of type D_n with parameters in the commutative ring A, denoted $Br(D_n, A)$, is generated by the elements y_α^a, where $\alpha \in \Delta$ and $a \in A$. The relations are, for $a, b \in A$ and $\alpha, \beta \in \Delta$*

$$\begin{array}{rll} (A1) & y_\alpha^a y_\alpha^0 y_\alpha^b = y_\alpha^0 y_\alpha^0 y_\alpha^{a+b} & \\ (A1 \times A1) & y_\alpha^a y_\beta^b = y_\beta^b y_\alpha^a & \textit{if } m(\alpha,\beta) = 2 \\ (A2) & y_\alpha^a y_\beta^b y_\alpha^c = y_\beta^c y_\alpha^{b+ac} y_\beta^a & \textit{if } m(\alpha,\beta) = 3 \textit{ and } \alpha < \beta \end{array}$$

Note that the simple roots in D_n are ordered so that $\alpha_{2'} < \alpha_3$.

3.2 The Steinberg group of D_n and the main result

The roots of D_n are $\{\pm\epsilon_i \pm \epsilon_j | 1 \leq i \neq j \leq n\}$, [Car],[H]. The Weyl group $W(D_n) \cong (\mathbb{Z}/2)^{n-1} \rtimes \mathcal{S}_n$ [Bour, p. 257, (X)] acts on the roots by permuting the indices (action of $\mathcal{S}_n$) and changing the signs (action of $(\mathbb{Z}/2)^{n-1}$). For the simple subsystem Δ we take $\alpha_i = -\epsilon_{i-1} + \epsilon_i$ for $i = 2, \cdots, n$, and $\alpha_{2'} = \epsilon_1 + \epsilon_2$. If u and v are positive integers and α, β two roots, the linear combination $u\alpha + v\beta$ is a root if and only if $u = 1 = v$, $\alpha = \pm\epsilon_i \pm \epsilon_j, \beta \mp \epsilon_j \pm \epsilon_k$ and $\pm\epsilon_i \pm \epsilon_k \neq 0$. In this case the definition of the Steinberg group is as follows.

Definition 3.2.1 ([Stb][St]). *The Steinberg group of type D_n with parameters in the commutative ring A, denoted $St(D_n, A)$, is generated by elements x^a_α, where $\alpha \in \Phi$ and $a \in A$, subject to the relations (for $a, b \in A$ and $\alpha, \beta \in \Phi$)*

$$\begin{array}{lll} (St0) & x^a_\alpha x^b_\alpha = x^{a+b}_\alpha & \\ (St1) & x^a_\alpha x^b_\beta = x^b_\beta x^a_\alpha & \text{if } \alpha+\beta \notin D_n \text{ and } \alpha+\beta \neq 0, \\ (St2) & x^a_\alpha x^b_\beta = x^b_\beta x^{ab}_{\alpha+\beta} y^a_\alpha & \text{if } \alpha+\beta \in D_n. \end{array}$$

The Weyl group $W(D_n)$ acts on $St(D_n, A)$ by $\sigma \cdot x^a_\alpha = x^a_{\sigma(\alpha)}$, and we can construct the semi-direct product $St(D_n, A) \rtimes Br(D_n)$ with respect to this action.

Theorem 3.2.2. *For any commutative ring A the map*

$$\phi : Br(D_n, A) \to St(D_n, A) \rtimes Br(D_n)$$

induced by $\phi(y^a_\alpha) = x^a_\alpha\, y_\alpha$ is a group isomorphism.

Corollary 3.2.3. *The group presented by generators y^a_i, $i = 2', 2, 3, \cdots, n$, $a \in A$ and relations*

$$\begin{array}{rll} (y^0_i)^2 &= 1 & \\ y^a_i (y^0_i)^{-1} y^b_i &= y^{a+b}_i & \\ y^a_i y^b_j &= y^b_j y^a_i & \text{if } |i-j| \geq 2, \text{ or } i = 2, j = 2', \\ y^a_i y^b_{i+1} y^c_i &= y^c_{i+1} y^{b+ac}_i\, y^a_{i+1} & \text{where } i+1 = 3 \text{ when } i = 2' \end{array}$$

for $a, b, c \in A$, is isomorphic to the semi-direct product $St(D_n, A) \rtimes W(D_n)$.

Proof of Corollary. For each simple root $\alpha_i \in D_n$, write y^a_i for $y^a_{\alpha_i}$. □

Proof of Theorem 3.2.2.
Step (a). Since the relations involved in the definitions of $Br(D_n, A)$ and $St(D_n, A)$ are the same as the relations in the case of A_{n-1}, the map ϕ is well-defined (*cf.* Theorem 2.4.1).
Step (b). The proof of the Pure Braid Lemma in the D_n case will be given below in 3.3.
Step (c). Let $\pi : Br(D_n, A) \to Br(D_n)$ be the projection which sends each $a \in A$ to 0 (as usual we identify $Br(D_n, 0)$ with $Br(D_n)$). We define

$$\psi : St(D_n, A) \rtimes Br(D_n) \to Br(D_n, A) \cong \operatorname{Ker} \pi \rtimes Br(D_n)$$

on the first component by $\psi(x^a_\alpha) = y^a_\alpha (y^0_\alpha)^{-1} \in \operatorname{Ker} \pi$ for $\alpha \in \Delta$. For any $\alpha \in D_n$ there exists $\sigma \in W(D_n)$ such that $\sigma(\alpha) \in \Delta$. Let $\tilde\sigma \in Br(D_n)$ be a lifting of σ, and define $\psi(x^a_\alpha) = \tilde{\sigma}^{-1}\psi(x^a_{\sigma(\alpha)})\tilde{\sigma} \in \operatorname{Ker}\pi$. This element is well-defined since it does not depend on the lifting of σ by the Pure Braid Lemma for D_n (Lemma 3.3.2), and does not depend on the choice of σ by Lemma 2.1.2. In order to show that ψ is a well-defined group homomorphism, it suffices to show that the Steinberg relations are preserved. But this is the same verification as in the A_{n-1} case, (*cf.* Theorem 2.4.1.)

The group homomorphisms ϕ and ψ are inverse to each other since they interchange y^a_α and $x^a_\alpha y_\alpha$. Hence they are both isomorphisms. $\square$

3.3 The Pure Braid Lemma for D_n

3.3.1 Generators for the Pure Braid Group of D_n

In principle, the method of Reidemeister-Schreier [M-K-S] is available to deduce a presentation of $PBr(D_n)$ from that of $Br(D_n)$. The details have been worked out by Digne and Gomi [D-G], although not in the specificity we need here. From their work we can deduce that the group $PBr(D_n)$ is generated by the elements $y^2_\alpha, \alpha \in \Delta$, together with a very small set of their conjugates. For example, $PBr(D_4)$ is generated by the 12 elements

$$\mathbf{2}^2, \mathbf{2'}^2, \mathbf{3}^2, {}^{\mathbf{3}}\mathbf{2}^2, {}^{\mathbf{3}}\mathbf{2'}^2, {}^{\mathbf{322'}}\mathbf{3}^2, \mathbf{4}^2, {}^{\mathbf{4}}\mathbf{3}^2, {}^{\mathbf{43}}\mathbf{2'}^2, {}^{\mathbf{43}}\mathbf{2}^2, {}^{\mathbf{4322'}}\mathbf{3}^2, {}^{\mathbf{4322'3}}\mathbf{4}^2$$

where a prefixed exponent indicates conjugation: ${}^h g = hgh^{-1}$. Here (and throughout) we use the simplified notations $\mathbf{k}^a = y^a_{\alpha_k}$ and $\mathbf{k} = y^0_{\alpha_k}$ similar to those of 2.2.

PROPOSITION 3.3.1. *For $n \geq 4$, $PBr(D_n)$ is generated by the elements*
- $\mathbf{a}_{j,i} = \mathbf{j}(\mathbf{j}-\mathbf{1})\ldots(\mathbf{i}+\mathbf{1})\mathbf{ii}(\mathbf{i}+\mathbf{1})\ldots(\mathbf{j}-\mathbf{1})\mathbf{j}$, $n \geq j \geq i \geq 2$, *and*
- $\mathbf{b}_{j,i} = \mathbf{j}(\mathbf{j}-\mathbf{1})\ldots\mathbf{322'3}\ldots(\mathbf{i}-\mathbf{1})\mathbf{ii}(\mathbf{i}-\mathbf{1})\ldots\mathbf{32'23}(\mathbf{j}-\mathbf{1})\mathbf{j}$, $n \geq j \geq i \geq 3$,

where $i + 1 = 3$ when $i = 2'$.

Note. Since the notation can be confusing, let us be clear about the definition of these generators in certain special cases:
- When $i = j, \mathbf{a}_{j,i} = \mathbf{i}^2$.
- When $i = 3, \mathbf{b}_{j,3} = \mathbf{j}(\mathbf{j}-\mathbf{1})\ldots\mathbf{322'332'2}\ldots(\mathbf{j}-\mathbf{1})\mathbf{j}$.

Proof of (3.3.1). We work in the case where $W = W(D_n)$ in the notation of [D-G]. In the proof of [D-G, Corollary 2.7], we see that $P_W = U_n \rtimes P_{W_{I_{n-1}}}$; taking $I_n = \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_{2'}, \ldots, \mathbf{s}_{n-1}\}$, $n \geq 4$, as on [D-G, p. 10], we see that their P_W is equal to (our) $PBr(D_n)$ and their $P_{W_{I_{n-1}}}$ is equal to (our) $PBr(D_{n-1})$. It follows that a set of generators for $PBr(D_n)$ can be obtained as the union of a set of generators for $PBr(D_{n-1})$ with a set of generators for U_n. This sets the stage for an inductive argument, since $D_3 = A_3$ (with $\{\alpha_2, \alpha_3, \alpha_{2'}\} \subset D_3$ identified with $\{\alpha_1, \alpha_2, \alpha_3\} \subset A_3$). Because $W(D_n)$ is a finite Weyl group, it follows from [D-G, Proposition 3.6], that U_n is generated (not just normally generated) by the elements $\mathbf{a_{b,s}}$, and a list of these generators in our case is given on [D-G, p. 10].

The calculations necessary to prove the Pure Braid Lemma for D_n are simpler if we replace the Digne-Gomi generators by the equivalent set in which conjugation is replaced by *reflection*; that is, we replace a generator ${}^h g = hgh^{-1}$ by hgh', where if $h = y_{i_1} \dots y_{i_k}$, $h' = y_{i_k} \dots y_{i_1}$. (We already used this trick in the case of A_{n-1}.) For D_4, this procedure yields as generators of $PBr(D_4)$ the set

$$\mathbf{2}^2, \mathbf{2}'^2, \mathbf{3}^2, \mathbf{32^23}, \mathbf{32'^23}, \mathbf{322'3^22'23}, \mathbf{4}^2, \mathbf{43^24}, \mathbf{432^234}, \mathbf{432'^234},$$
$$\mathbf{4322'3^22'234}, \mathbf{4322'34^232'234},$$

and, more generally, $PBr(D_n), n \geq 4$ is generated by the elements stated in the Proposition. $\square$

Lemma 3.3.2 (Pure Braid Lemma for D_n). *Let $\alpha_k \in D_n$, let y_k^a be a generator of $Br(D_n, A)$, and let $\omega \in PBr(D_n)$. Then there exists $\omega' \in Br(D_n)$, independent of a, such that*

$$y_k^a \ \omega = \omega' \ y_k^a.$$

Hence for any integer k and any element $a \in A$, the element $y_k^a(y_k^0)^{-1} \in Br(D_n, A)$ commutes with every element of the pure braid group $PBr(D_n)$.

Proof. Let us show that the first assertion implies the second one. Let $\omega \in PBr(D_n) \subset PBr(D_n, A)$. By the first assertion of the Lemma we have

$$y_k^a \omega = \omega' y_k^a$$

for some $\omega' \in Br(D_n, 0)$, independent of a. Setting $a = 0$ tells us that $\omega' = y_k^0 \omega (y_k^0)^{-1}$. Thus

$$y_k^a (y_k^0)^{-1} \omega = \omega y_k^a (y_k^0)^{-1}$$

as desired.

Before beginning the proof of the first assertion, we recall some notation introduced in 2.2. We abbreviate $y_{\alpha_k}^a$ by $\mathbf{k}^a$ and $y_{\alpha_k}^0$ by $\mathbf{k}$. Whenever there exist ω' and $\omega'' \in Br(D_n)$ such that $\mathbf{k}^a \omega = \omega' \mathbf{j}^a \omega''$, we will write $\mathbf{k}^a \omega \sim \mathbf{j}^a \omega''$. This is *not* an equivalence relation, but it is compatible with multiplication on the right by elements of $Br(D_n)$: if $\eta \in Br(D_n)$,

$$\mathbf{k}^a \omega \sim \mathbf{j}^a \omega'' \Leftrightarrow \mathbf{k}^a \omega = \omega' \mathbf{j}^a \omega'' \Leftrightarrow \mathbf{k}^a \omega \eta = \omega' \mathbf{j}^a \omega'' \eta \Leftrightarrow \mathbf{k}^a \omega \eta \sim \mathbf{j}^a \omega'' \eta$$

From defining relations $(A1), (A1 \times A1)$, and $(A2)$ of 2, we can deduce the following:

$$\begin{array}{llcl}
\text{(4a)} & \mathbf{k}^a\mathbf{k}\mathbf{k} & \sim & \mathbf{k}^a \\
\text{(4b)} & \mathbf{k}^a\ (\mathbf{k-1})\ \mathbf{k} & \sim & (\mathbf{k-1})^a \\
\text{(4c)} & \mathbf{k}^a\ (\mathbf{k-1})\ \mathbf{k}^{-1} & \sim & (\mathbf{k-1})^a \\
\text{(4d)} & \mathbf{k}^a(\mathbf{k+1})\mathbf{k} & \sim & (\mathbf{k+1})^a \\
\text{(4e)} & (\mathbf{k+1})^a\mathbf{k}(\mathbf{k+1}) & \sim & \mathbf{k}^a \\
\text{(4f)} & (\mathbf{k+1})^a\mathbf{k}(\mathbf{k+1})^{-1} & \sim & \mathbf{k}^a \\
\text{(4g)} & \mathbf{k}^a(\mathbf{k+1})\mathbf{k}^{-1} & \sim & (\mathbf{k+1})^a \\
\text{(4h)} & \mathbf{k}^a(\mathbf{k+1})^{-1}\mathbf{k}^{-1} & \sim & (\mathbf{k+1})^a \\
\text{(4i)} & (\mathbf{k+1})^a\mathbf{k}^{-1}(\mathbf{k+1})^{-1} & \sim & \mathbf{k}^a \\
\text{(4j)} & \mathbf{k}^a\mathbf{k}(\mathbf{k-1}) & \sim & (\mathbf{k-1})^a(\mathbf{k-1})\mathbf{k}^{-1}
\end{array}$$

Proof, continued. In the notation just introduced, we must show, for every $\omega \in PBr(D_n)$, that $\mathbf{k}^a\omega \sim \mathbf{k}^a$, and it suffices to show this when ω is one of the generators $\mathbf{a}_{j,i}$ or $\mathbf{b}_{j,i}$ of 3.3.1. That is, we must show

$$\begin{array}{llcll}
\text{(5a)} & \mathbf{k}^a\mathbf{a}_{j,i} & \sim & \mathbf{k}^a & n \geq j \geq i \geq 2, \quad 1 \leq k \leq n \\
\text{(5b)} & \mathbf{k}^a\mathbf{b}_{j,i} & \sim & \mathbf{k}^a & n \geq j \geq i \geq 3, \quad 1 \leq k \leq n
\end{array}$$

The proofs of (5a) for $i \geq 3$ are exactly the same as the corresponding proofs for A_{n-1} (see section 2); the additional case $i = 2'$ presents no new issues. Thus we shall concentrate on proving (5b); the proof proceeds by induction on n.

The case $n = 3$ is the case of the root system $D_3 = A_3$, which is part of the Pure Braid Lemma 2.2.1 for A_{n-1}. Hence we may assume $n \geq 4$, and that (5b) holds whenever $j, k \leq n-1$. That is, we must prove (5b) in these cases:

$$k = n,\ j \leq n-1; \quad k \leq n-1,\ j = n; \quad k = n,\ j = n$$

which further subdivide into the cases

$$\begin{array}{lll}
\text{(6)} & k = n, & j \leq n-2 \\
\text{(7)} & k = n, & j = n-1 \\
\text{(8)} & k \leq n-2, & j = n \\
\text{(9)} & k = n-1, & j = n \\
\text{(10)} & k = n, & j = n
\end{array}$$

• *Case* (6) *$k = n$ and $j \leq n-2$.* Since $i \leq j \leq n-2$, it follows from relation $(A1 \times A1)$ that $\mathbf{n}^a$ commutes with every generator which occurs in the expression for $\mathbf{b}_{j,i}$; hence

$$\begin{aligned}
\mathbf{n}^a\mathbf{b}_{j,i} &= \mathbf{n}^a\mathbf{j}(\mathbf{j-1})\ldots\mathbf{3 2 2' 3}\ldots(\mathbf{i-1})\mathbf{i}\mathbf{i}(\mathbf{i-1})\ldots\mathbf{3 2' 2 3}(\mathbf{j-1})\mathbf{j} \\
&= \mathbf{b}_{j,i}\mathbf{n}^a \\
&\sim \mathbf{n}^a
\end{aligned}$$

as desired.

• *Case* (7) $k = n$ *and* $j = n-1$.

If $i \leq n-2$, then

$$\begin{aligned}
&\mathbf{n}^a\mathbf{b}_{n-1,i} \\
&= \mathbf{n}^a\mathbf{(n-1)(n-2)\ldots 322'3\ldots(i-1)ii(i-1)\ldots 32'23\ldots(n-2)(n-1)} \\
&= \underbrace{\mathbf{n}^a\mathbf{(n-1)n}^{-1}}\mathbf{n(n-2)\ldots 322'3\ldots(i-1)ii(i-1)\ldots 32'23\ldots(n-2)(n-1)} \\
&\sim \mathbf{(n-1)}^a\mathbf{n(n-2)\ldots 322'3\ldots(i-1)ii(i-1)\ldots 32'23\ldots(n-2)(n-1)} \\
&= \underbrace{\mathbf{(n-1)}^a\mathbf{(n-2)\ldots 322'3\ldots(i-1)ii(i-1)\ldots 32'23\ldots(n-2)}}\mathbf{n(n-1)} \\
&\sim \mathbf{(n-1)}^a\mathbf{n(n-1)} \\
&\sim \mathbf{n}^a
\end{aligned}$$

as desired.

The case $k = n, i = j = n-1$, is considerably more complicated. We first prove some preliminary lemmas.

LEMMA 3.3.3.

$$\mathbf{n}^a\mathbf{(n-1)(n-2)\ldots 322'} \sim \mathbf{2}^a\mathbf{32'4\ldots(n-2)(n-1)n}$$

Proof.

$$\begin{aligned}
\mathbf{n}^a\mathbf{(n-1)(n-2)\ldots 322'} &= \mathbf{n}^a\mathbf{(n-1)n}^{-1}\mathbf{n(n-2)\ldots 322'} \\
&\sim \mathbf{(n-1)}^a\mathbf{n(n-2)\ldots 322'} \\
&= \mathbf{(n-1)}^a\mathbf{(n-2)\ldots 322'n} \\
&\vdots \\
&\sim \mathbf{3}^a\mathbf{22'4\ldots(n-2)(n-1)n} \\
&= \mathbf{3}^a\mathbf{23}^{-1}\mathbf{32'4\ldots(n-2)(n-1)n} \\
&\sim \mathbf{2}^a\mathbf{32'4\ldots(n-2)(n-1)n}
\end{aligned}$$

□

LEMMA 3.3.4.

$$\mathbf{32'4354\ldots(n-3)(n-1)(n-2)nn} = \mathbf{345\ldots(n-1)nn2'34\ldots(n-2)}$$

Proof.

$$\begin{aligned}
&\mathbf{32'4354\ldots(n-3)(n-1)(n-2)nn} \\
&\qquad = \mathbf{32'4354\ldots(n-3)(n-1)nn(n-2)} \\
&\qquad = \mathbf{32'4354\ldots(n-1)nn(n-3)(n-2)} \\
&\qquad \vdots \\
&\qquad = \mathbf{345\ldots(n-1)nn2'34\ldots(n-2)}
\end{aligned}$$

$\square$

We now complete the case $k = n, i = j = n-1$.

$$
\begin{aligned}
&\mathbf{n}^a \mathbf{b}_{n-1,n-1} \\
&\quad = \mathbf{n}^a(\mathbf{n-1})(\mathbf{n-2})\dots\mathbf{322'3}\dots(\mathbf{n-2})(\mathbf{n-1})(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad \sim \mathbf{2}^a\mathbf{32'4}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n3}\dots(\mathbf{n-2})(\mathbf{n-1})(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \quad \text{(by Lemma 3.3.3)} \\
&\quad = \mathbf{2}^a\mathbf{32'4}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{3}\dots(\mathbf{n-2})\underbrace{\mathbf{n}(\mathbf{n-1})\mathbf{n}^{-1}}\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad = \mathbf{2}^a\mathbf{32'4}\dots(\mathbf{n-2})\underbrace{(\mathbf{n-1})\mathbf{3}\dots}(\mathbf{n-2})(\mathbf{n-1})^{-1}\mathbf{n}(\mathbf{n-1})\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad = \mathbf{2}^a\mathbf{32'4}\dots(\mathbf{n-2})\mathbf{3}\dots\underbrace{(\mathbf{n-1})(\mathbf{n-2})(\mathbf{n-1})^{-1}}\mathbf{n}(\mathbf{n-1})\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad = \mathbf{2}^a\mathbf{32'4}\dots(\mathbf{n-2})\mathbf{3}\dots\underbrace{(\mathbf{n-2})^{-1}(\mathbf{n-1})(\mathbf{n-2})}\mathbf{n}(\mathbf{n-1})\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\qquad \vdots \\
&\quad = \mathbf{2}^a\mathbf{32'}\underbrace{\mathbf{434}^{-1}}\mathbf{54}\dots(\mathbf{n-1})(\mathbf{n-2})\mathbf{n}(\mathbf{n-1})\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad = \mathbf{2}^a\underbrace{\mathbf{32'3}^{-1}}\mathbf{4354}\dots(\mathbf{n-1})(\mathbf{n-2})\mathbf{n}(\mathbf{n-1})\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad = \mathbf{2}^a\underbrace{\mathbf{2}'^{-1}\mathbf{32'}}\mathbf{4354}\dots(\mathbf{n-1})(\mathbf{n-2})\mathbf{n}(\mathbf{n-1})\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad \sim \mathbf{2}^a\mathbf{32'4354}\dots(\mathbf{n-1})(\mathbf{n-2})\mathbf{n}\underbrace{(\mathbf{n-1})\mathbf{n}(\mathbf{n-1})}(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad = \mathbf{2}^a\mathbf{32'4354}\dots(\mathbf{n-1})(\mathbf{n-2})\mathbf{n}\underbrace{\mathbf{n}(\mathbf{n-1})\mathbf{n}}(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \\
&\quad = \mathbf{2}^a\mathbf{345}\dots(\mathbf{n-1})\mathbf{nn2'34}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n}(\mathbf{n-2})\dots \\
&\qquad \dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1}) \quad \text{(by Lemma 3.3.4)}
\end{aligned}
$$

We now manipulate part of this expression so that we can apply induction.

$$
\begin{aligned}
&\mathbf{2}^a\mathbf{345}\dots(\mathbf{n-1})\mathbf{nn2'} \\
&\qquad\qquad = \mathbf{2}^a\mathbf{345}\dots(\mathbf{n-1})\mathbf{nn}\underbrace{(\mathbf{n-1})\dots\mathbf{5433}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots(\mathbf{n-1})^{-1}}\mathbf{2'} \\
&\qquad\qquad \sim \mathbf{2}^a\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots(\mathbf{n-1})^{-1}\mathbf{2'}
\end{aligned}
$$

(by the case of $A_{n-1} = \{\alpha_2, \alpha_3, \dots, \alpha_n\}$). Hence

$$
\begin{aligned}
&\mathbf{n}^a\mathbf{b}_{n-1,n-1}\\
&\quad= \mathbf{2}^a\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots(\mathbf{n-1})^{-1}\mathbf{2'34}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n}(\mathbf{n-2})(\mathbf{n-3})\dots\\
&\qquad\dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad= \mathbf{2}^a\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots(\mathbf{n-1})^{-1}\mathbf{2'34}\dots(\mathbf{n-2})(\mathbf{n-1})(\mathbf{n-2})\mathbf{n}(\mathbf{n-3})\dots\\
&\qquad\dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad= \mathbf{2}^a\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots(\mathbf{n-1})^{-1}\mathbf{2'34}\dots(\mathbf{n-1})(\mathbf{n-2})(\mathbf{n-1})\mathbf{n}(\mathbf{n-3})\dots\\
&\qquad\dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad= \mathbf{2}^a\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots(\mathbf{n-2})^{-1}\mathbf{2'34}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n}(\mathbf{n-3})\dots\\
&\qquad\dots\mathbf{32'23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\qquad\vdots\\
&\quad= \mathbf{2}^a\mathbf{3}^{-1}\mathbf{2'34}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n2'23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad= \mathbf{2}^a\mathbf{3}^{-1}\mathbf{2'32'4}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad= \mathbf{2}^a\mathbf{3}^{-1}\mathbf{32'34}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad= \mathbf{2}^a\mathbf{2'34}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad\sim \mathbf{2}^a\mathbf{34}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n23}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad= \mathbf{2}^a\mathbf{324}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n3}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\quad\sim \mathbf{3}^a\mathbf{4}\dots(\mathbf{n-2})(\mathbf{n-1})\mathbf{n3}\dots(\mathbf{n-2})(\mathbf{n-1})\\
&\qquad\vdots\\
&\quad\sim (\mathbf{n-1})^a\mathbf{n}(\mathbf{n-1})\\
&\quad\sim \mathbf{n}^a
\end{aligned}
$$

as desired. □

- *Case* (8) $k \le n-2$ *and* $j = n$.

$$
\begin{aligned}
&\mathbf{k}^a\mathbf{b}_{n,i}\\
&\quad= \mathbf{k}^a\mathbf{n}(\mathbf{n-1})\dots\mathbf{322'3}\dots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\dots\mathbf{32'23}\dots(\mathbf{n-1})\mathbf{n}\\
&\quad\sim \underbrace{\mathbf{k}^a(\mathbf{k+1})\dots\mathbf{322'3}\dots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\dots\mathbf{32'23}\dots(\mathbf{k+1})}(\mathbf{k+2})\dots(\mathbf{n-1})\mathbf{n}\\
&\quad\sim \mathbf{k}^a(\mathbf{k+2})\dots\mathbf{n}\quad\text{(by induction)}\\
&\quad\sim \mathbf{k}^a
\end{aligned}
$$

as desired.

- *Case* (9) $k = n-1$ *and* $j = n$.

$$
\begin{aligned}
&(\mathbf{n-1})^a\mathbf{b}_{n,i}\\
&\quad= (\mathbf{n-1})^a\mathbf{n}(\mathbf{n-1})\dots\mathbf{322'3}\dots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\dots\mathbf{32'23}\dots(\mathbf{n-1})\mathbf{n}\\
&\quad= \underbrace{(\mathbf{n-1})^a\mathbf{n}(\mathbf{n-1})}\dots\mathbf{322'3}\dots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\dots\mathbf{32'}\,223\dots(\mathbf{n-1})\mathbf{n}\\
(\dagger)&\quad\sim \mathbf{n}^a(\mathbf{n-2})\dots\mathbf{322'3}\dots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\dots\mathbf{32'23}\dots(\mathbf{n-1})\mathbf{n}
\end{aligned}
$$

Suppose first that $i \leq n-2$. Then

$$\begin{aligned}
(\dagger) \quad &= \mathbf{n}^a(\mathbf{n-2})\ldots\mathbf{322'3}\ldots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}\\
&= (\mathbf{n-2})\ldots\mathbf{322'3}\ldots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\ldots\mathbf{32'23}\ldots(\mathbf{n-2})\mathbf{n}^a(\mathbf{n-1})\mathbf{n}\\
&\sim \mathbf{n}^a(\mathbf{n-1})\mathbf{n}\\
&\sim (\mathbf{n-1})^a
\end{aligned}$$

as desired.

If $i = n-1$

$$\begin{aligned}
(\dagger) \quad &= \mathbf{n}^a(\mathbf{n-2})\ldots\mathbf{322'3}\ldots(\mathbf{n-2})(\mathbf{n-1})(\mathbf{n-1})(\mathbf{n-2})\ldots\\
&\quad \ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}\\
&\sim \underbrace{\mathbf{n}^a(\mathbf{n-1})(\mathbf{n-1})}(\mathbf{n-2})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}\\
&\sim \mathbf{n}^a(\mathbf{n-2})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n} \quad \text{(by the case } A_2 = \{\alpha_{n-1},\alpha_n\})\\
&\sim \mathbf{n}^a(\mathbf{n-1})\mathbf{n} \quad \text{(by relation } (A1\times A1))\\
&\sim (\mathbf{n-1})^a
\end{aligned}$$

as desired.

If $i = n$

$$\begin{aligned}
(\dagger) \quad &= \mathbf{n}^a(\mathbf{n-2})\ldots\mathbf{322'3}\ldots(\mathbf{n-1})\mathbf{nn}(\mathbf{n-1})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}\\
&\sim \mathbf{n}^a(\mathbf{n-1})\mathbf{nn}(\mathbf{n-1})(\mathbf{n-2})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}\\
&\sim (\mathbf{n-1})^a\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}\\
&\sim \mathbf{n}^a(\mathbf{n-2})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}\\
&\sim \mathbf{n}^a(\mathbf{n-1})\mathbf{n}\\
&\sim (\mathbf{n-1})^a
\end{aligned}$$

as desired.

- *Case* (10) $k = n$ *and* $j = n$.

$$\begin{aligned}
&\mathbf{n}^a\mathbf{b}_{n,i}\\
&= \mathbf{n}^a\mathbf{n}(\mathbf{n-1})(\mathbf{n-2})\ldots\mathbf{322'3}\ldots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}\\
(\dagger)\ &\sim (\mathbf{n-1})^a(\mathbf{n-1})\mathbf{n}^{-1}(\mathbf{n-2})\ldots\mathbf{322'3}\ldots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})\mathbf{n}
\end{aligned}$$

If $i \leq n-2$, then

$$\begin{aligned}
&(\dagger)\\
&= (\mathbf{n-1})^a(\mathbf{n-1})(\mathbf{n-2})\ldots\mathbf{322'3}\ldots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\ldots\mathbf{32'23}\ldots\underbrace{\mathbf{n}^{-1}(\mathbf{n-1})\mathbf{n}}\\
&= \underbrace{(\mathbf{n-1})^a(\mathbf{n-1})(\mathbf{n-2})\ldots\mathbf{322'3}\ldots(\mathbf{i-1})\mathbf{ii}(\mathbf{i-1})\ldots\mathbf{32'23}\ldots(\mathbf{n-1})}\mathbf{n}(\mathbf{n-1})^{-1}\\
&\qquad \sim (\mathbf{n-1})^a\mathbf{n}(\mathbf{n-1})^{-1} \quad \text{(by induction)}\\
&\qquad \sim \mathbf{n}^a
\end{aligned}$$

as desired.

If $i = n-1$, then

$$
\begin{aligned}
(\dagger) \quad &= \mathbf{(n-1)}^a\mathbf{(n-1)(n-2)}\ldots \\
&\quad \ldots\mathbf{322'3}\ldots\mathbf{(n-2)n}^{-1}\mathbf{(n-1)(n-1)(n-2)}\ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&\quad \vdots \\
&\sim \mathbf{3}^a\mathbf{322'4}^{-1}\mathbf{35}^{-1}\mathbf{4}\ldots \\
&\quad \ldots\mathbf{(n-1)}^{-1}\mathbf{(n-2)n}^{-1}\mathbf{(n-1)(n-1)(n-2)}\ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&= \mathbf{3}^a\mathbf{322'4}^{-1}\mathbf{35}^{-1}\mathbf{4}\ldots\mathbf{(n-1)}^{-1}\mathbf{(n-2)}\underbrace{\mathbf{n}^{-1}\mathbf{(n-1)n}}\mathbf{n}^{-1}\mathbf{(n-1)(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&= \mathbf{3}^a\mathbf{322'4}^{-1}\mathbf{35}^{-1}\mathbf{4}\ldots \\
&\quad \ldots\underbrace{\mathbf{(n-1)}^{-1}\mathbf{(n-2)(n-1)}}\,\mathbf{n}\,\underbrace{\mathbf{(n-1)}^{-1}\mathbf{n}^{-1}\mathbf{(n-1)}}\mathbf{(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&= \mathbf{3}^a\mathbf{322'4}^{-1}\mathbf{35}^{-1}\mathbf{4}\ldots \\
&\quad \ldots\mathbf{(n-3)}\underbrace{\mathbf{(n-2)(n-1)(n-2)}^{-1}}\,\mathbf{n}\,\underbrace{\mathbf{n(n-1)}^{-1}}\mathbf{n}^{-1}\mathbf{(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&= \mathbf{3}^a\mathbf{322'4}^{-1}\mathbf{35}^{-1}\mathbf{4}\ldots \\
&\quad \ldots\mathbf{(n-3)(n-2)(n-1)nn(n-2)}^{-1}\mathbf{(n-1)}^{-1}\mathbf{n}^{-1}\mathbf{(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&\quad \vdots \\
&= \mathbf{3}^a\mathbf{322'}\underbrace{\mathbf{4}^{-1}\mathbf{34}}\ldots\mathbf{(n-2)(n-1)nn4}^{-1}\mathbf{5}^{-1}\ldots \\
&\quad \ldots\mathbf{(n-1)}^{-1}\mathbf{n}^{-1}\mathbf{(n-2)}\ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&\sim \mathbf{2}^a\mathbf{2}\underbrace{\mathbf{3}^{-1}\mathbf{2'3}}\mathbf{43}^{-1}\mathbf{5}\ldots\mathbf{(n-1)nn4}^{-1}\mathbf{5}^{-1}\ldots\mathbf{(n-1)}^{-1}\mathbf{n}^{-1}\mathbf{(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&= \mathbf{2}^a\mathbf{22'32'}^{-1}\mathbf{4}\ldots\mathbf{(n-1)nn3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\ldots\mathbf{(n-1)}^{-1}\mathbf{n}^{-1}\mathbf{(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&= \mathbf{2}^a\mathbf{22'34}\ldots\mathbf{(n-1)nn2'}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\ldots\mathbf{(n-1)}^{-1}\mathbf{n}^{-1}\mathbf{(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&\sim \mathbf{2}^a\mathbf{234}\ldots\mathbf{(n-1)nn2'}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\ldots\mathbf{(n-1)}^{-1}\mathbf{n}^{-1}\mathbf{(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n} \\
&= \mathbf{2}^a\mathbf{234}\ldots \\
&\quad \ldots\mathbf{(n-1)nn}\underbrace{\mathbf{(n-1)}\ldots\mathbf{4322}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\ldots\mathbf{(n-1)}^{-1}}\mathbf{2'}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\ldots \\
&\quad \ldots\mathbf{(n-1)}^{-1}\mathbf{n}^{-1}\mathbf{(n-2)}\ldots \\
&\quad \ldots\mathbf{32'23}\ldots\mathbf{(n-1)n}
\end{aligned}
$$

$$
\begin{aligned}
&\sim \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots\\
&\quad \dots(\mathbf{n-1})^{-1}\mathbf{2}'^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots(\mathbf{n-1})^{-1}\mathbf{n}^{-1}(\mathbf{n-2})\dots\mathbf{32}'\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&\quad \text{(by the case of } A_{n-1}=\{\alpha_2,\alpha_3,\dots,\alpha_n\})\\
&\sim \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots\\
&\quad \dots(\mathbf{n-1})^{-1}\mathbf{2}'^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots\\
&\quad \dots(\mathbf{n-3})^{-1}\underbrace{(\mathbf{n-2})^{-1}(\mathbf{n-1})^{-1}(\mathbf{n-2})}\mathbf{n}^{-1}(\mathbf{n-3})\dots\\
&\quad \dots\mathbf{32}'\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&= \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots(\mathbf{n-1})^{-1}\mathbf{2}'^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots(\mathbf{n-3})^{-1}\\
&\quad \underbrace{(\mathbf{n-1})(\mathbf{n-2})^{-1}(\mathbf{n-1})^{-1}}\mathbf{n}^{-1}(\mathbf{n-3})\dots\\
&\quad \dots\mathbf{32}'\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&= \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots\\
&\quad \dots(\mathbf{n-1})^{-1}(\mathbf{n-1})\mathbf{2}'^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots\\
&\quad \dots(\mathbf{n-3})^{-1}(\mathbf{n-2})^{-1}(\mathbf{n-1})^{-1}\mathbf{n}^{-1}(\mathbf{n-3})\dots\\
&\quad \dots\mathbf{32}'\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&= \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots(\mathbf{n-2})\mathbf{2}'^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\mathbf{5}^{-1}\dots\\
&\quad \dots(\mathbf{n-3})^{-1}(\mathbf{n-2})^{-1}(\mathbf{n-1})^{-1}\mathbf{n}^{-1}(\mathbf{n-3})\dots\\
&\quad \dots\mathbf{32}'\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&\quad \vdots\\
&= \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{2}'^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots\mathbf{n}^{-1}\mathbf{2}'\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&= \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\underbrace{\mathbf{2}'^{-1}\mathbf{3}^{-1}\mathbf{2}'}\mathbf{4}^{-1}\dots\mathbf{n}^{-1}\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&= \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{32}'^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots\mathbf{n}^{-1}\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&\sim \mathbf{2}^a\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots\mathbf{n}^{-1}\mathbf{23}\dots(\mathbf{n-1})\mathbf{n}\\
&= \mathbf{2}^a\underbrace{\mathbf{2}^{-1}\mathbf{3}^{-1}\mathbf{2}}\mathbf{4}^{-1}\dots\mathbf{n}^{-1}\mathbf{3}\dots(\mathbf{n-1})\mathbf{n}\\
&= \underbrace{\mathbf{2}^a\mathbf{32}^{-1}}\mathbf{3}^{-1}\mathbf{4}^{-1}\dots\mathbf{n}^{-1}\mathbf{3}\dots(\mathbf{n-1})\mathbf{n}\\
&\sim \mathbf{3}^a\mathbf{3}^{-1}\mathbf{4}^{-1}\dots\mathbf{n}^{-1}\mathbf{3}\dots(\mathbf{n-1})\mathbf{n}\qquad\text{(by (4g))}\\
&\quad \vdots\\
&\sim (\mathbf{n-1})^a(\mathbf{n-1})^{-1}\underbrace{\mathbf{n}^{-1}(\mathbf{n-1})\mathbf{n}}\qquad\text{(by (4g))}\\
&= (\mathbf{n-1})^a(\mathbf{n-1})^{-1}(\mathbf{n-1})\mathbf{n}(\mathbf{n-1})^{-1}\\
&= (\mathbf{n-1})^a\mathbf{n}(\mathbf{n-1})^{-1}\\
&\sim \mathbf{n}^a
\end{aligned}
$$

as desired.

If $i = n$, then

$$
\begin{aligned}
(\dagger) \quad = \quad & \mathbf{(n-1)}^{a}\mathbf{(n-1)n}^{-1}\mathbf{(n-2)}\dots\mathbf{322'3}\dots\mathbf{(n-1)nn(n-1)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n} \\
= \quad & \mathbf{(n-1)}^{a}\mathbf{(n-1)(n-2)}\dots\mathbf{322'3}\dots\underbrace{\mathbf{n}^{-1}\mathbf{(n-1)n}}\mathbf{n(n-1)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n} \\
= \quad & \mathbf{(n-1)}^{a}\mathbf{(n-1)(n-2)}\dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)n}\underbrace{\mathbf{(n-1)}^{-1}\mathbf{n(n-1)}}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n} \\
= \quad & \mathbf{(n-1)}^{a}\mathbf{(n-1)(n-2)}\dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-1)n}^{-1}\mathbf{(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n} \\
= \quad & \mathbf{(n-1)}^{a}\mathbf{(n-1)(n-2)}\dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{n}^{-1}\mathbf{(n-1)n} \\
= \quad & \mathbf{(n-1)}^{a}\mathbf{(n-1)(n-2)}\dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n(n-1)}^{-1} \\
\sim \quad & \mathbf{(n-2)}^{a}\mathbf{(n-2)(n-1)}^{-1}\mathbf{(n-3)}\dots \\
& \dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n(n-1)}^{-1} \qquad \text{(by (4g))} \\
= \quad & \mathbf{(n-2)}^{a}\mathbf{(n-2)(n-3)}\dots \\
& \dots\mathbf{322'3}\dots\mathbf{(n-1)}^{-1}\mathbf{(n-2)(n-1)nn(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n(n-1)}^{-1} \\
= \quad & \mathbf{(n-2)}^{a}\mathbf{(n-2)(n-3)}\dots \\
& \dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)(n-2)}^{-1}\mathbf{nn(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n(n-1)}^{-1} \\
= \quad & \mathbf{(n-2)}^{a}\mathbf{(n-2)(n-3)}\dots \\
& \dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-2)}^{-1}\mathbf{(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n(n-1)}^{-1} \\
= \quad & \mathbf{(n-2)}^{a}\mathbf{(n-2)(n-3)}\dots \\
& \dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-1)(n-2)(n-1)}^{-1}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-1)n(n-1)}^{-1} \\
= \quad & \mathbf{(n-2)}^{a}\mathbf{(n-2)(n-3)}\dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-3)(n-1)}^{-1}\mathbf{(n-2)(n-1)n(n-1)}^{-1} \\
= \quad & \mathbf{(n-2)}^{a}\mathbf{(n-2)(n-3)}\dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-3)(n-2)(n-1)(n-2)}^{-1}\mathbf{n(n-1)}^{-1} \\
= \quad & \mathbf{(n-2)}^{a}\mathbf{(n-2)(n-3)}\dots\mathbf{322'3}\dots\mathbf{(n-2)(n-1)nn(n-1)(n-2)}\dots \\
& \dots\mathbf{32'23}\dots\mathbf{(n-3)(n-2)(n-1)n(n-2)}^{-1}\mathbf{(n-1)}^{-1} \\
& \vdots
\end{aligned}
$$

$$\begin{aligned}
&\sim \mathbf{2}^a\mathbf{2}\mathbf{2}'\mathbf{3}\dots(\mathbf{n}-\mathbf{1})\mathbf{n}\mathbf{n}(\mathbf{n}-\mathbf{1})\dots\\
&\quad \dots\mathbf{3}\mathbf{2}'\mathbf{2}\mathbf{3}\dots(\mathbf{n}-\mathbf{1})\mathbf{n}\mathbf{2}^{-1}\mathbf{3}^{-1}\dots(\mathbf{n}-\mathbf{1})^{-1} \quad \text{(by (4g))}\\
&\sim \underbrace{\mathbf{2}^a\mathbf{2}\mathbf{3}\dots(\mathbf{n}-\mathbf{1})\mathbf{n}\mathbf{n}(\mathbf{n}-\mathbf{1})\dots\mathbf{3}\mathbf{2}}\,\mathbf{2}'\mathbf{3}\dots(\mathbf{n}-\mathbf{1})\mathbf{n}\mathbf{2}^{-1}\mathbf{3}^{-1}\dots(\mathbf{n}-\mathbf{1})^{-1}\\
&\sim \mathbf{2}^a\mathbf{2}'\mathbf{3}\dots(\mathbf{n}-\mathbf{1})\mathbf{n}\mathbf{2}^{-1}\mathbf{3}^{-1}\dots(\mathbf{n}-\mathbf{1})^{-1}\\
&\quad \text{(by the case of } A_{n-1}=\{\alpha_2,\alpha_3,\dots,\alpha_n\})\\
&\sim \mathbf{2}^a\mathbf{3}\dots(\mathbf{n}-\mathbf{1})\mathbf{n}\mathbf{2}^{-1}\mathbf{3}^{-1}\dots(\mathbf{n}-\mathbf{1})^{-1}\\
&= \mathbf{2}^a\mathbf{3}\mathbf{2}^{-1}\dots(\mathbf{n}-\mathbf{1})\mathbf{n}\mathbf{3}^{-1}\dots(\mathbf{n}-\mathbf{1})^{-1}\\
&= \mathbf{3}^a\mathbf{4}\dots(\mathbf{n}-\mathbf{1})\mathbf{n}\mathbf{3}^{-1}\dots(\mathbf{n}-\mathbf{1})^{-1}\\
&\quad \vdots\\
&= (\mathbf{n}-\mathbf{1})^a\mathbf{n}(\mathbf{n}-\mathbf{1})^{-1}\\
&= \mathbf{n}^a
\end{aligned}$$

as desired. □

References

[Bir] Birman, J. S., *Braids, links, and mapping class groups.* Annals of Mathematics Studies, No. 82. Princeton University Press, Princeton, N.J., University of Tokyo Press, Tokyo, 1974.

[Bon] Bonis, R., *Groupes d'Artin et formes réelles des relations de Yang-Baxter.* C. R. Acad. Sci. Paris Sér. I Math. 329 (1999), no. 8, 663–666.

[Bour] Bourbaki, N., Groupes et algèbres de Lie, Chapitres IV, V et VI. Actualités Scientifiques et Industrielles, No. 1337 Hermann, Paris 1968.

[Car] Carter, Roger W., Simple groups of Lie type. John Wiley & Sons, Inc., New York, 1989.

[D-G] Digne, F., Gomi, Y., *Presentation of pure braid groups.* J. Knot Theory Ramifications 10 (2001), no. 4, 609–623.

[H] Humphreys, J. E., Reflection groups and Coxeter groups. Cambridge Studies in Advanced Mathematics, 29. Cambridge University Press, Cambridge, 1990.

[K-S] Kapranov, M.M., Saito, M., *Hidden Stasheff polytopes in algebraic K-theory and in the space of Morse functions.* in Higher homotopy structures in topology and mathematical physics (Poughkeepsie, NY, 1996), 191–225, Contemp. Math., 227, Amer. Math. Soc., Providence, RI, 1999.

[K-R] Kassel, Ch. et Reutenauer, Ch., *Une variante à la Coxeter du groupe de Steinberg.* K-Theory 14 (1998), no. 4, 305–318.

[L] Loday, J.-L., *Homotopical syzygies*, in "Une dégustation topologique: Homotopy theory in the Swiss Alps", Contemporary Mathematics no 265 (AMS) (2000), 99–127.

[M-K-S] Magnus, W., Karrass, A., Solitar, D., Combinatorial group theory: Presentations of groups in terms of generators and relations. Interscience Publishers [John Wiley & Sons, Inc.], New York-London-Sydney 1966.

[St] Stein, M. R., *Generators, relations and coverings of Chevalley groups over commutative rings*. Amer. J. Math. 93 (1971), 965–1004.

[Stb] Steinberg, R., *Générateurs, relations et revêtements de groupes algébriques*. 1962 Colloq. Théorie des Groupes Algébriques (Bruxelles, 1962) pp. 113–127 Librairie Universitaire, Louvain, Gauthier-Villars, Paris.

Jean-Louis Loday
Institut de Recherche Mathématique Avancée,
CNRS et Université Louis Pasteur
7 rue R. Descartes,
67084 Strasbourg Cedex, France
loday@math.u-strasbg.fr

Michael R. Stein
Department of Mathematics
Northwestern University
2033 Sheridan Road
Evanston IL 60208–2730 USA
mike@math.northwestern.edu

Localization of the Essential Spectrum for Relativistic N-Electron Ions and Atoms

D. H. Jakubassa-Amundsen

Received: October 12, 2005
Revised: November 4, 2005

Communicated by Heinz Siedentop

Abstract. The HVZ theorem is proven for the pseudorelativistic N-electron Jansen-Hess operator ($2 \leq N \leq Z$) which acts on the spinor Hilbert space $\mathcal{A}(H_1(\mathbb{R}^3) \otimes \mathbb{C}^4)^N$ where $\mathcal{A}$ denotes antisymmetrization with respect to particle exchange. This 'no pair' operator results from the decoupling of the electron and positron degrees of freedom up to second order in the central potential strength $\gamma = Ze^2$.

2000 Mathematics Subject Classification: 81Q10, 81V45

Introduction

We consider N interacting electrons in a central Coulomb field generated by a point nucleus of charge number Z which is infinitely heavy and located at the origin. For stationary electrons where the radiation field and pair creation can be neglected, the $N+1$ particle system is described by the Coulomb-Dirac operator, introduced by Sucher [23]. The Jansen-Hess operator used in the present work, which acts on the positive spectral subspace of N free electrons, is derived from the Coulomb-Dirac operator by applying a unitary transformation scheme [12, 13] which is equivalent to the Douglas-Kroll transformation scheme [6]. The transformed operator is represented as an infinite series of operators which do not couple the electron and positron degrees of freedom. For $N = 1$, each successive term in this series is of increasing order in the strength γ of the central field. The series has been shown to be convergent for subcritical potential strength ($\gamma < \gamma_c = 0.3775$, corresponding to $Z < 52$ [21]). For $N > 1$ the expansion parameter is e^2, which comprises the central field strength Ze^2 and the strength e^2 of the electron-electron interaction. A numerical investigation of the cases $N = 1,\ Z-1$ and Z across the periodic table has revealed [27] that the ground-state energy of an N-electron system is already quite well represented if the series is truncated after the second-order term. This approximation defines the Jansen-Hess operator (see (3.1) below). In the present work we provide the localization of the essential spectrum of this operator. Recently [14] we have proven the HVZ theorem (which dates back to Hunziker [10], van Winter [26] and Zhislin [28] for the Schrödinger operator and to Lewis, Siedentop and Vugalter [16] for the scalar pseudorelativistic Hamiltonian) for the two-particle Brown-Ravenhall operator [2] which is the first-order term in the above mentioned series of operators. Now we extend this proof successively to the multiparticle Brown-Ravenhall operator (section 1), to the two-electron Jansen-Hess operator (section 2) and finally to the N-electron Jansen-Hess operator. We closely follow the earlier work [14] where the details can be found. A quite different proof of the HVZ theorem for the multiparticle Brown-Ravenhall operator is presently under investigation [18].

1 Multiparticle Brown-Ravenhall case

For N electrons of mass m in a central field, generated by a point nucleus which is infinitely heavy and fixed at the origin, the Brown-Ravenhall operator is given by (in relativistic units, $\hbar = c = 1$)

$$H^{BR} \;=\; \Lambda_{+,N}\,(\sum_{k=1}^{N}(D_0^{(k)} \,+\, V^{(k)}) \,+\, \sum_{k>l=1}^{N} V^{(kl)})\,\Lambda_{+,N} \tag{1.1}$$

where $D_0^{(k)} = \boldsymbol{\alpha}^{(k)}\mathbf{p}_k + \beta^{(k)}m$ is the free Dirac operator of electron k, $V^{(k)} = -\gamma/x_k$ is the central potential with strength $\gamma = Ze^2$, and $V^{(kl)} = e^2/|\mathbf{x}_k - \mathbf{x}_l|$ is the electron-electron interaction, $e^2 \approx 1/137.04$ being the fine structure constant and $x_k = |\mathbf{x}_k|$ the distance of electron k from the origin. Further, $\Lambda_{+,N} = \Lambda_+^{(1)} \cdots \Lambda_+^{N}$ (as shorthand for $\bigotimes_{k=1}^{N} \Lambda_+^{(k)}$) is the (tensor) product of the single-particle projectors $\Lambda_+^{(k)} = \frac{1}{2}(1 + D_0^{(k)}/E_{p_k})$ onto the positive spectral subspace of $D_0^{(k)}$. H^{BR} acts in the Hilbert space $\mathcal{A}(L_2(\mathbb{R}^3) \otimes \mathbb{C}^4)^N$, and is well-defined in the form sense and positive on $\mathcal{A}(H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^4)^N$ for $\gamma < \gamma_{BR} = \frac{2}{\pi/2+2/\pi} \approx 0.906$ (see (1.10) below). For the multi-nucleus case the Brown-Ravenhall operator was shown to be positive if $\gamma < 0.65$ [9].

An equivalent operator, which is defined in a reduced spinor space by means of $(\psi_+, H^{BR}\psi_+) = (\varphi, h^{BR}\varphi)$ with $\psi_+ \in \Lambda_{+,N}(\mathcal{A}(H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^4)^N)$ and $\varphi \in \mathcal{A}(H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^2)^N$, is [7]

$$h^{BR} \;=\; \sum_{k=1}^{N}(T^{(k)} \,+\, b_{1m}^{(k)}) \,+\, \sum_{k>l=1}^{N} v^{(kl)}. \tag{1.2}$$

Explicitly, with $A_k := A(p_k) = \left(\frac{E_{p_k}+m}{2E_{p_k}}\right)^{1/2}$ and $G_k := \boldsymbol{\sigma}^{(k)}\mathbf{p}_k g(p_k)$,

$g(p_k) = (2E_{p_k}(E_{p_k}+m))^{-1/2}$, one has [14]

$$T^{(k)} := E_{p_k} \;=\; \sqrt{p_k^2+m^2}, \qquad b_{1m}^{(k)} \;=\; -\gamma\,(A_k\,\frac{1}{x_k}\,A_k \;+\; G_k\,\frac{1}{x_k}\,G_k)$$

$$\begin{aligned} v^{(kl)} \;&=\; A_kA_l\,\frac{e^2}{|\mathbf{x}_k-\mathbf{x}_l|}\,A_kA_l \;+\; A_kG_l\,\frac{e^2}{|\mathbf{x}_k-\mathbf{x}_l|}\,A_kG_l \\ &+ G_kA_l\,\frac{e^2}{|\mathbf{x}_k-\mathbf{x}_l|}\,G_kA_l \;+\; G_kG_l\,\frac{e^2}{|\mathbf{x}_k-\mathbf{x}_l|}\,G_kG_l. \end{aligned} \tag{1.3}$$

Let us consider the two-cluster decompositions $\{C_{1j}, C_{2j}\}$ of the N-electron atom, obtained by moving electron j far away from the atom or by separating the nucleus from all electrons. Denote by C_{1j} the cluster located near the origin (containing the nucleus), while C_{2j} contains either one electron ($j = 1, ..., N$) or all electrons ($j = 0$). Correspondingly, h^{BR} is split into

$$h^{BR} \;=\; T \;+\; a_j \;+\; r_j, \qquad j = 0, 1, ..., N, \tag{1.4}$$

with $T := \sum_{k=1}^{N} T^{(k)}$, while a_j denotes the interaction of the particles located all in cluster C_{1j} or all in C_{2j}. The remainder r_j collects the interactions between particles sitting in different clusters and is supposed to vanish when C_{2j} is moved to infinity.
Define for $j \in \{0, 1, ..., N\}$

$$\Sigma_0 := \min_j \inf\,\sigma(T + a_j). \tag{1.5}$$

Then we have

Theorem 1 (HVZ theorem for the multiparticle Brown-Ravenhall operator).
Let h^{BR} be the Brown-Ravenhall operator for $N > 2$ electrons in a central field of strength $\gamma < \gamma_{BR} = \frac{2}{\pi/2+2/\pi}$, and let (1.4) be its two-cluster decompositions. Then the essential spectrum of h^{BR} is given by

$$\sigma_{ess}(h^{BR}) \;= [\Sigma_0, \infty). \tag{1.6}$$

In fact, the assertion (1.6) holds even in a more general case. For $K \geq 2$ introduce K-cluster decompositions $d := \{C_1, ..., C_K\}$ of the $N+1$ particles, and split $h^{BR} = T + a_d + r_d$ accordingly (where $T + a_d$ describes the infinitely separated clusters while r_d comprises all interactions between particles sitting in two different clusters). Let

$$\Sigma_1 := \min_{\#d\geq 2} \inf\,\sigma(T + a_d). \tag{1.7}$$

Then $\sigma_{ess}(h^{BR}) \;=\; [\Sigma_1, \infty)$ with $\Sigma_1 \;=\; \Sigma_0$. This result, known from the Schrödinger case [20, p.122], relies on the fact that the electron-electron interaction is repulsive ($V^{(kl)} \geq 0$ respective $v^{(kl)} \geq 0$) and can be proved as follows.

First consider K-cluster decompositions of the form $\#C_1 = N+1-(K-1)$ and $\#C_i = 1, \quad i = 2,...,K$ (i.e. one ion and $K-1$ separated electrons). For any $j \in \{1,...,N\}$ we use for the two-cluster decompositions the notation $T + a_j = h_{N-1}^{BR} + T^{(j)}$, where the subscript on h^{BR} denotes the number of electrons in the central field, and assume (1.6) to hold. Then

$$\inf \sigma(h_N^{BR}) \leq \inf \sigma_{ess}(h_N^{BR}) \leq \inf \sigma(h_{N-1}^{BR} + T^{(j)})$$
$$= \inf \sigma(h_{N-1}^{BR}) + m. \tag{1.8}$$

By induction (corresponding to successive removal of an electron) we get

$$\inf \sigma(h_{N-1}^{BR}) \leq \inf \sigma(h_{N-1-N'}^{BR}) + N'm, \qquad 0 \leq N' < N-1. \tag{1.9}$$

Since for a K-cluster decomposition of this specific form one has $T + a_d = h_{N-(K-1)}^{BR} + T^{(1)} + ... + T^{(K-1)}$, it follows that $\inf \sigma(T + a_d) = \inf \sigma(h_{N-(K-1)}^{BR}) + (K-1)m \geq \inf \sigma(h_{N-1}^{BR}) + m \geq \Sigma_0$.
Assume now cluster decompositions d with $\#C_1 = N+1-(K-1)$ fixed ($K \in \{3,...,N\}$) but where $\#C_i > 1$ for at least one $i > 1$. Then $T + a_d$ is increased by (nonnegative) electron-electron interaction terms $v^{(kl)}$ as compared to the K-cluster decompositions considered above, such that $\inf \sigma(T + a_d)$ is higher (or equal) than for the case $\#C_i = 1, \quad i = 2,...,K$. Therefore, cluster decompositions with $\#C_i > 1$ (for some $i > 1$) do not contribute to Σ_1, such that, together with (1.9), $\Sigma_1 = \Sigma_0$ is proven.

Let us embark on the proof of Theorem 1. The required lemmata will bear the same numbers as in [14].
We say that an operator $\mathcal{O}$ is $\frac{1}{R}$-bounded if $\mathcal{O}$ is bounded by $\frac{c}{R}$ with some constant $c > 0$.

(a) In order to prove the 'hard part' of the HVZ theorem, $\sigma_{ess}(h^{BR}) \subset [\Sigma_0, \infty)$, we start by noting that the potential of h^{BR} is T-form bounded with form bound $c < 1$ if $\gamma < \gamma_{BR}$. With $\psi_+ \in \Lambda_{+,N}\mathcal{A}(H_{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^4)^N$, this follows from the estimates [4, 25, 13] (using that $V^{(k)} \leq 0$ and $V^{(kl)} \geq 0$),

$$(\psi_+, \left(\sum_{k=1}^N V^{(k)} + \sum_{k>l=1}^N V^{(kl)}\right) \psi_+) \leq \sum_{k>l=1}^N \frac{e^2}{\gamma_{BR}} (\psi_+, E_{p_1} \psi_+)$$
$$= \frac{N-1}{2} \frac{e^2}{\gamma_{BR}} (\psi_+, T \psi_+)$$
$$(\psi_+, \left(\sum_{k=1}^N V^{(k)} + \sum_{k>l=1}^N V^{(kl)}\right) \psi_+) \geq -\frac{\gamma}{\gamma_{BR}} \sum_{k=1}^N (\psi_+, E_{p_1} \psi_+) \tag{1.10}$$
$$= -\frac{\gamma}{\gamma_{BR}} (\psi_+, T \psi_+)$$

such that $c := \max\{\frac{\gamma}{\gamma_{BR}}, \frac{N-1}{2} \frac{e^2}{\gamma_{BR}}\}$. $c < 1$ requires $\gamma < \gamma_{BR}$ for all physical values of N ($N < 250$). From (1.10), $h^{BR} \geq 0$ for $\gamma \leq \gamma_{BR}$.
In order to establish Persson's theorem (proven in [5] for Schrödinger operators and termed Lemma 2 in [14]),

$$\inf \sigma_{ess}(h^{BR}) \;=\; \lim_{R\to\infty} \inf_{\|\varphi\|=1} (\varphi, h^{BR}\varphi) \tag{1.11}$$

if $\varphi \in \mathcal{A}(C_0^\infty(\mathbb{R}^{3N}\backslash B_R(0)) \otimes \mathbb{C}^{2N})$ where $B_R(0) \subset \mathbb{R}^{3N}$ is a ball of radius R centered at the origin, we need the fact that the Weyl sequence φ_n for a λ in the essential spectrum of h^{BR} can be chosen such that it is supported outside a ball $B_n(0)$:

Lemma 1. *Let $h^{BR} = T + V$, let V be relatively form bounded with respect to T. Then $\lambda \in \sigma_{ess}(h^{BR})$ iff there exists a sequence of functions $\varphi_n \in \mathcal{A}(C_0^\infty(\mathbb{R}^{3N}\backslash B_n(0)) \otimes \mathbb{C}^{2N})$ with $\|\varphi_n\| = 1$ such that*

$$\|(h^{BR} - \lambda)\, \varphi_n\| \;\longrightarrow\; 0 \qquad \text{as } n \to \infty. \tag{1.12}$$

If such φ_n exist they form a Weyl sequence because φ_n converge weakly to zero [14]. For the proof of the converse direction, let $\lambda \in \sigma_{ess}(h^{BR})$ be characterized by a Weyl sequence $\psi_n \in \mathcal{A}(C_0^\infty(\mathbb{R}^3) \otimes \mathbb{C}^2)^N$, $\|\psi_n\| = 1$, with $\psi_n \xrightarrow{w} 0$ and $\|(h^{BR} - \lambda)\psi_n\| \to 0$ as $n \to \infty$. Let $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_N) \in \mathbb{R}^{3N}$ be the coordinates of the N electrons and define a smooth symmetric auxiliary function $\chi_0 \in C_0^\infty(\mathbb{R}^{3N})$ mapping to $[0,1]$ by means of

$$\chi_0(\frac{\mathbf{x}}{n}) \;=\; \begin{cases} 1, & x \leq n \\ 0, & x > 2n \end{cases} \tag{1.13}$$

where $x = |\mathbf{x}| = \sqrt{x_1^2 + ... + x_N^2}$. Then we set $\chi_n(\mathbf{x}) := 1 - \chi_0(\mathbf{x}/n)$ and claim that a subsequence of the sequence $\varphi_n := \psi_n \chi_n \in \mathcal{A}(C_0^\infty(\mathbb{R}^{3N}\backslash B_n(0)) \otimes \mathbb{C}^{2N})$ satisfies the requirements of Lemma 1.
In order to show that $\|(h^{BR} - \lambda)\varphi_n\| = \|\chi_n(h^{BR} - \lambda)\psi_n + [h^{BR}, \chi_0]\psi_n\| \to 0$ for $n \to \infty$, we have to estimate the single-particle contributions $\|[T^{(k)}, \chi_0]\psi_n\|$ and $\|[b_{1m}^{(k)}, \chi_0]\psi_n\|$. With $b_{1m}^{(k)}$ of the form $B_k \frac{1}{x_k} B_k$ where $B_k \in \{A_k, G_k\}$ is a bounded multiplication operator in momentum space, we have to consider commutators of the type $p_k[B_k, \chi_0]$ which are multiplied by bounded operators. These commutators are shown to be $\frac{1}{n}$-bounded in the same way as for $N = 2$ [14], by working in momentum space and introducing the N-dimensional Fourier transform (marked by a hat) of the Schwartz function χ_0,

$$\left(\widehat{\chi_0(\frac{\cdot}{n})}\right)(\mathbf{p}) \;=\; \frac{1}{(2\pi)^{3N/2}} \int_{\mathbb{R}^{3N}} d\mathbf{x}\, e^{-i\mathbf{p}\mathbf{x}}\, \chi_0(\frac{\mathbf{x}}{n}) \;=\; n^{3N}\, \hat{\chi}_0(\mathbf{p}_1 n, ..., \mathbf{p}_N n), \tag{1.14}$$

where $\mathbf{p} = (\mathbf{p}_1, ..., \mathbf{p}_N)$, and by using the mean value theorem to estimate the difference $|B_k(\mathbf{p}_k) - B_k(\mathbf{p}_{k'})|$ respective $|T^{(k)}(\mathbf{p}_k) - T^{(k)}(\mathbf{p}_{k'})|$. The two-particle contributions $\|[v^{(kl)}, \chi_0]\, \psi_n\|$ can, according to the representation (1.3) of $v^{(kl)}$, also be split into single-particle commutators $p_k[B_k,\chi_0]$ multiplied by bounded operators. Their estimate as well as the remaining parts of the proof of Lemma 1 for $N>2$ (in particular the normalizability of φ_n for sufficiently large n which relies on the relative form boundedness of the total potential) can be mimicked from the case of $N=2$.

Our aim is a generalization of the localization formula of Lewis et al [16] to the operator h^{BR}. We introduce the Ruelle-Simon [22] partition of unity $(\phi_j)_{j=0,\ldots,N} \in C^\infty(\mathbb{R}^{3N})$ which is subordinate to the two-cluster decompositions (1.4). It is defined on the unit sphere in $\mathbb{R}^{3N}$ and has the following properties (see e.g. [5, p.33], [24])

$$\sum_{j=0}^{N} \phi_j^2 \;=\; 1, \qquad \phi_j(\lambda\mathbf{x}) \;=\; \phi_j(\mathbf{x}) \quad \text{for } x=1 \text{ and } \lambda \geq 1,$$

$$\begin{aligned} \operatorname{supp}\phi_j \,\cap\, \mathbb{R}^{3N}\backslash B_1(0) \,\subseteq \{\mathbf{x}\in\mathbb{R}^{3N}\backslash B_1(0) :\, |\mathbf{x}_k-\mathbf{x}_l| \geq Cx \text{ for all } k\in C_{1j} \\ \text{and } l\in C_{2j}, \;\text{ and } x_k \geq Cx \text{ for all } k\in C_{2j}\}, \;\; j=0,1,...,N \end{aligned} \tag{1.15}$$

where C is a constant and it is again assumed that the nucleus belongs to cluster C_{1j}. Then we have

Lemma 3. *Let $h^{BR}=T+a_j+r_j$, $(\phi_j)_{j=0,\ldots,N}$ be the Ruelle-Simon partition of unity and $\varphi\in\mathcal{A}(C_0^\infty(\mathbb{R}^{3N}\backslash B_R(0))\otimes\mathbb{C}^{2N})$ with $R>1$. Then, with some constant c,*

$$|(\phi_j\varphi, r_j\,\phi_j\varphi)| \;\leq\; \frac{c}{R}\,\|\varphi\|^2, \qquad j=0,..,N. \tag{1.16}$$

There are two possibilities. r_j may (a) consist of terms $b_{1m}^{(k)}$ for some $k\in C_{2j}$, or (b) of terms $v^{(kl)}$ with particles k and l in different clusters. For the proof, all summands of r_j are estimated separately. For each summand of r_j (to a given cluster decomposition j), a specific smooth auxiliary function χ mapping to $[0,1]$ is introduced which is unity on the support of $\phi_j\varphi$, such that $\phi_j\varphi\chi=\phi_j\varphi$. In case (a) we have $\operatorname{supp}\phi_j\varphi \subset \mathbb{R}^{3N}\backslash B_R(0)\cap\{x_k\geq Cx\}$, i.e. $x_k \geq CR$. Therefore we define the (single-particle) function

$$\chi_k(\frac{\mathbf{x}_k}{R}) := \begin{cases} 0, & x_k < CR/2 \\ 1, & x_k \geq CR \end{cases}. \tag{1.17}$$

With $b_{1m}^{(k)}$ of the form $B_k\frac{1}{x_k}B_k$ we have to consider
$(\phi_j\varphi, B_k\frac{1}{x_k}B_k\,\chi_k\phi_j\varphi) = (\phi_j\varphi, B_k\frac{1}{x_k}\chi_kB_k\,\phi_j\varphi) + (\phi_j\varphi, B_k\frac{1}{x_k}\,[B_k,1-\chi_k]\,\phi_j\varphi)$. The first term is uniformly $2/R$-bounded by the choice (1.17) of χ_k, whereas the second term can be estimated in momentum space as in the two-electron case (respective in the proof of Lemma 1).

In case (b) we have $\operatorname{supp}\phi_j\varphi\subset\mathbb{R}^{3N}\backslash B_R(0)\cap\{|\mathbf{x}_k-\mathbf{x}_l|\geq Cx\}$, i.e. $|\mathbf{x}_k-\mathbf{x}_l| \geq CR$. Accordingly, we take

$$\chi_{kl}(\frac{\mathbf{x}_k-\mathbf{x}_l}{R}) := \begin{cases} 0, & |\mathbf{x}_k-\mathbf{x}_l| \;<\; CR/2 \\ 1, & |\mathbf{x}_k-\mathbf{x}_l| \;\geq\; CR \end{cases}. \tag{1.18}$$

With the representation (1.3) of $v^{(kl)}$, we have to estimate commutators of the type $p_k[B_k\, B_l, 1-\chi_{kl}]$. The proof of their uniform $1/R$-boundedness can be copied from the two-electron case.
The second ingredient of the localization formula is an estimate for the commutator of ϕ_j with h^{BR}:

Lemma 4. *Let h^{BR} from (1.2) and $(\phi_j)_{j=0,...,N}$ be the Ruelle-Simon partition of unity. Then for $\varphi \in \mathcal{A}(C_0^\infty(\mathbb{R}^{3N}\backslash B_R(0)) \otimes \mathbb{C}^{2N})$ and $R>2$ one has*

$$\begin{array}{llcl} (a) & |\sum\limits_{j=0}^{N}(\phi_j\varphi, [T, \phi_j]\,\varphi)| & \leq & \dfrac{c}{R^2}\,\|\varphi\|^2 \\ (b) & |(\phi_j\varphi, [b_{1m}^{(k)}, \phi_j]\,\varphi)| & \leq & \dfrac{c}{R}\,\|\varphi\|^2 \\ (c) & |(\phi_j\varphi, [v^{(kl)}, \phi_j]\,\varphi)| & \leq & \dfrac{c}{R}\,\|\varphi\|^2 \end{array} \tag{1.19}$$

where c is a generic constant.

Item (a) is proven in [16]. For items (b) and (c) we define the smooth auxiliary N-particle function χ mapping to $[0,1]$,

$$\chi(\frac{\mathbf{x}}{R}) := \left\{ \begin{array}{ll} 0, & x < R/2 \\ 1, & x \geq R \end{array} \right. . \tag{1.20}$$

Then $\phi_j\varphi = \phi_j\varphi\chi$ on $\operatorname{supp}\varphi$, and therefore $(\phi_j\varphi, [b_{1m}^{(k)}, \phi_j]\,\varphi) = (\phi_j\varphi, [b_{1m}^{(k)}, \phi_j\chi]\,\varphi)$. The $\frac{1}{R}$-estimate, claimed in (1.19), relies on the scaling property $\phi_j(\mathbf{x})\chi(\frac{\mathbf{x}}{R}) = \phi_j(\frac{\mathbf{x}}{R/2})\chi(\frac{\mathbf{x}}{R})$ which holds for $R>2$ since $\operatorname{supp}\chi$ (and hence $\operatorname{supp}\phi_j\chi$) is outside $B_{R/2}(0)$. Thus, working in coordinate space and using the mean value theorem, we get the estimate

$$|(\phi_j\chi)(\mathbf{x}_1, ..., \mathbf{x}_k, ..., \mathbf{x}_N) - (\phi_j\chi)(\mathbf{x}_1, ..., \mathbf{x}_k', ..., \mathbf{x}_N)| \leq |\mathbf{x}_k - \mathbf{x}_k'|\,\frac{c_0}{R} \tag{1.21}$$

(only the k-th coordinate in the second entry of the l.h.s. is primed). Since (1.21) holds for arbitrary $k \in \{1, ..., N\}$, the proof of (b) and (c) can be carried out in the same way as done in the two-electron case, by estimating the kernel of $B_k \in \{A_k, G_k\}$ in coordinate space by $c/|\mathbf{x}_k - \mathbf{x}_k'|^3$ (using asymptotic analysis [19]) and subsequently proving the uniform $\frac{1}{R}$-boundedness of $[B_k, \phi_j\chi]\,\frac{1}{x_k}$ respective $[B_k, \phi_j\chi]\,\frac{1}{|\mathbf{x}_k - \mathbf{x}_l|}$.
With Lemmata 3 and 4 we obtain the desired localization formula for h^{BR},

$$(\varphi, h^{BR}\varphi) \;=\; \sum_{j=0}^{N}(\phi_j\varphi, (T+a_j)\,\phi_j\varphi)$$

$$+\; \sum_{j=0}^{N}(\phi_j\varphi, r_j\,\phi_j\varphi) \;-\; \sum_{j=0}^{N}(\phi_j\varphi, [h^{BR}, \phi_j]\,\varphi) \tag{1.22}$$

$$= \sum_{j=0}^{N} (\phi_j \varphi, (T + a_j)\, \phi_j \varphi) \; + \; O(\frac{1}{R})\; \|\varphi\|^2$$

for $R > 2$. From Persson's theorem (1.11) and the definition (1.5) of Σ_0 we therefore get

$$\inf \sigma_{ess}(h^{BR}) \;=\; \lim_{R\to\infty} \inf_{\|\varphi\|=1} \sum_{j=0}^{N} (\phi_j \varphi, (T + a_j)\, \phi_j \varphi)$$

$$\geq \; \Sigma_0 \sum_{j=0}^{N} (\phi_j \varphi, \phi_j \varphi) \;=\; \Sigma_0 \tag{1.23}$$

which proves the inclusion $\sigma_{ess}(h^{BR}) \subset [\Sigma_0, \infty)$.

(b) We now turn to the 'easy part' of the proof where we have to verify $[\Sigma_0, \infty) \subset \sigma_{ess}(h^{BR})$.
We start by showing that for every $j \in \{0, 1, ..., N\}$, $\sigma(T + a_j)$ is continuous, i.e. for any $\lambda \in [\inf \sigma(T + a_j), \infty)$ one has $\lambda \in \sigma(T + a_j)$. If the cluster C_{2j} consists of a single electron j, then $T + a_j = T^{(j)} + h_{N-1}^{BR}$ where h_{N-1}^{BR} does not contain any interaction with electron j. The continuity of $\sigma(T^{(j)} + h_{N-1}^{BR})$ then follows from the continuity of $\sigma(T^{(j)})$ in the same way as for $N = 2$.
In the case $j = 0$ where C_{2j} contains N electrons, the total momentum $\mathbf{p}_0$ of C_{2j} is well-defined and commutes with its Hamiltonian $h_0 := T + \sum_{k>l=1}^{N} v^{(kl)} = T + a_0$. This follows from the absence of any central potential in h_0 and from the symmetry of $v^{(kl)}$,

$$[(-i\nabla_{\mathbf{x}_k} - i\nabla_{\mathbf{x}_l}), \frac{1}{|\mathbf{x}_k - \mathbf{x}_l|}]\psi(\mathbf{x}) = (-i\nabla_{\mathbf{x}_k} - i\nabla_{\mathbf{x}_l})(\frac{1}{|\mathbf{x}_k - \mathbf{x}_l|})\; \psi(\mathbf{x}) = 0 \cdot \psi(\mathbf{x}) = 0. \tag{1.24}$$

Thus the eigenfunctions to h_0 can be chosen as eigenfunctions of $\mathbf{p}_0$. For $p_0 \geq 0$ the associated center of mass energy of C_{2j} is continuous. Therefore, $\inf \sigma(h_0)$ is attained for $p_0 = 0$ and $\sigma(h_0)$ is continuous.
Let $\lambda \in [\Sigma_0, \infty)$. We have $\Sigma_0 = \inf \sigma(T + a_j)$ for a specific $j \in \{0, ..., N\}$. Then $\lambda \in \sigma(T + a_j)$, i.e. there exists a defining sequence $\varphi_n(\mathbf{x}) \in C_0^\infty(\mathbb{R}^{3N}) \otimes \mathbb{C}^{2N}$ with $\|\varphi_n\| = 1$ and $\|(T + a_j - \lambda)\, \varphi_n\| \to 0$ for $n \to \infty$.
Assume that l electrons belong to cluster C_{2j} which we will enumerate by $N - l + 1, ..., N$, and follow [10] to define the unitary translation operator $T_{\mathbf{a}}$ by means of

$$T_{\mathbf{a}}\, \psi(\mathbf{x}_1, ..., \mathbf{x}_N) \;=\; \psi(\mathbf{x}_1, ..., \mathbf{x}_{N-l}, \mathbf{x}_{N-l+1} - \mathbf{a}, ..., \mathbf{x}_N - \mathbf{a}) \tag{1.25}$$

with $|\mathbf{a}| = a$ and let $\mathbf{a}_l := (\mathbf{a}, ..., \mathbf{a}) \in \mathbb{R}^{3l}$. Hence cluster C_{2j} moves to infinity as $a \to \infty$.

Let $\psi_n^{(a)} := T_{\mathbf{a}}\varphi_n$ and $\mathcal{A}\psi_n^{(a)}$ be the antisymmetric function constructed from $\psi_n^{(a)}$. We claim that $\mathcal{A}\psi_n^{(a)}$ is a defining sequence for $\lambda \in \sigma(h^{BR})$. It is sufficient (as shown below) to prove that $\psi_n^{(a)}$ has this property. We have trivially $\|\psi_n^{(a)}\| = \|\varphi_n\|$ and we have to show that $\|(h^{BR} - \lambda)\,\psi_n^{(a)}\| \to 0$ for $n \to \infty$ and a suitably large a. We have

$$\|(h^{BR} - \lambda)\,\psi_n^{(a)}\| \;\leq\; \|(T + a_j - \lambda)\,\psi_n^{(a)}\| \;+\; \|r_j\,\psi_n^{(a)}\|. \tag{1.26}$$

T commutes with $T_{\mathbf{a}}$ because T is a multiplication operator in momentum space. Since the central potentials contained in a_j are not affected by $T_{\mathbf{a}}$ (because $T_{\mathbf{a}}$ does not act on the particle coordinates of cluster C_{1j}), we also have $[T_{\mathbf{a}}, a_j] = 0$. In fact, assuming e.g. that electrons k and l are in cluster C_{2j} and using the representation (1.3) for $v^{(kl)}$ we have with $T_{\mathbf{a}}^* T_{\mathbf{a}} = 1$ and $B_k, \tilde{B}_k \in \{A_k, G_k\}$,

$$\begin{aligned} T_{\mathbf{a}}^* B_k \tilde{B}_l \,\frac{1}{|\mathbf{x}_k - \mathbf{x}_l|}\, B_k \tilde{B}_l\, T_{\mathbf{a}} \;&=\; B_k \tilde{B}_l T_{\mathbf{a}}^* \,\frac{1}{|\mathbf{x}_k - \mathbf{x}_l|}\, T_{\mathbf{a}}\, B_k \tilde{B}_l \\ =\; B_k \tilde{B}_l \,\frac{1}{|\mathbf{x}_k + \mathbf{a} - (\mathbf{x}_l + \mathbf{a})|}\, B_k \tilde{B}_l \;&=\; B_k \tilde{B}_l \,\frac{1}{|\mathbf{x}_k - \mathbf{x}_l|}\, B_k \tilde{B}_l \end{aligned} \tag{1.27}$$

such that $[T_{\mathbf{a}}, v^{(kl)}] = 0$. Then, given some $\epsilon > 0$, the first term of (1.26) reduces to

$$\|(T + a_j - \lambda)\; T_{\mathbf{a}}\varphi_n\| \;=\; \|T_{\mathbf{a}}\,(T + a_j - \lambda)\;\varphi_n\| \;\leq\; \|T_{\mathbf{a}}\|\; \|(T + a_j - \lambda)\;\varphi_n\| \;<\; \epsilon/2 \tag{1.28}$$

if $n > N_0$ for N_0 sufficiently large.

For the second term in (1.26) we note that r_j consists of terms $b_{1m}^{(k)}$ with $k \notin C_{1j}$ and terms $v^{(kk')}$ with $k \in C_i,\;\; k' \in C_{i'},\;\; i \neq i'$. Moreover, since a_j does not contain any intercluster interactions, we can choose $\varphi_n = \varphi_1^{(n)} \cdot \varphi_2^{(n)}$ as a product of functions $(\varphi_1^{(n)} \in C_0^\infty(\mathbb{R}^{3(N-l)} \otimes \mathbb{C}^{2(N-l)}),\;\; \varphi_2^{(n)} \in C_0^\infty(\mathbb{R}^{3l} \otimes \mathbb{C}^{2l}))$ each of which describing the electrons in cluster C_{1j} respective C_{2j}. Let $\operatorname{supp}\varphi_i^{(n)} \subset B_{R_i}(0)$ for a suitable R_i.

Consider $\|b_{1m}^{(k)} \psi_n^{(a)}\|$ with $k \in C_{2j}$. We have $\operatorname{supp} T_{\mathbf{a}}\varphi_2^{(n)} \subset B_{R_2}(\mathbf{a}_l)$. Let $a > 2R_2$. For all $k' \in C_{2j}$, on the support of $T_{\mathbf{a}}\varphi_2^{(n)}$ we have $R_2 > |\mathbf{x}_{k'} - \mathbf{a}| \geq a - x_{k'}$ and thus $x_{k'} > a - R_2$. Therefore we can write $\operatorname{supp} T_{\mathbf{a}}\varphi_2^{(n)} \subset \mathbb{R}^{3l} \backslash B_{|\mathbf{a}_l| - R_2}(0) \cap \{x_{k'} > a - R_2 \;\forall\, k' \in C_{2j}\}$. Assume we can prove

Lemma 5. *Let $\varphi \in C_0^\infty(\Omega) \otimes \mathbb{C}^{2l}$ with $\Omega := \{\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_l) \in \mathbb{R}^{3l} :\;\; x_i > R \;\forall\, i = 1, ..., l\}$ and $R > 1$. Then for $k \in \{1, ..., l\}$ and some constant c,*

$$\|b_{1m}^{(k)}\,\varphi\| \;\leq\; \frac{c}{R}\,\|\varphi\|. \tag{1.29}$$

Since $b_{1m}^{(k)}$ acts only on $T_{\mathbf{a}}\varphi_2^{(n)}$ we obtain

$$\|b_{1m}^{(k)}\; T_{\mathbf{a}}\varphi_n\| \;=\; \|\varphi_1^{(n)}\|\; \|b_{1m}^{(k)}\; T_{\mathbf{a}}\varphi_2^{(n)}\| \;\leq\; \frac{c}{a - R_2}\; \|T_{\mathbf{a}}\,\varphi_2^{(n)}\| \;<\; \frac{2c}{a} \tag{1.30}$$

because the $\varphi_i^{(n)}$ are normalized. As a consequence, for any $k \in C_{2j}$, the l.h.s. of (1.30) can be made smaller than $\epsilon/4l$ for sufficiently large a.

For the proof of Lemma 5 or, equivalently, of $|(\phi, b_{1m}^{(k)}\varphi)| \le \frac{c}{R}\,\|\varphi\|\,\|\phi\|$ for all $\phi \in (C_0^\infty(\mathbb{R}^3)\otimes\mathbb{C}^2)^l$, we note that the basic difference to the respective assertion for $N=2$ lies in the possible multiparticle nature of ϕ and φ. However, the property of the domain Ω of φ allows for the introduction of the (single-particle) smooth auxiliary function (mapping to $[0,1]$),

$$\chi(\frac{\mathbf{x}_k}{R}) := \begin{cases} 0, & x_k < R/2 \\ 1, & x_k \ge R \end{cases}, \tag{1.31}$$

such that $\varphi\chi = \varphi$. Then the proof can be copied from the two-electron case. For the two-particle interaction contained in r_j, one has

Lemma 6. *Let $\psi_n^{(a)} = T_{\mathbf{a}}\varphi_n = T_{\mathbf{a}}\,\varphi_1^{(n)}\varphi_2^{(n)}$ as defined above. Then for all $\varphi \in (C_0^\infty(\mathbb{R}^3)\otimes\mathbb{C}^2)^N$ and $a > 4R$,*

$$|(\varphi, v^{(kk')}\psi_n^{(a)})| \le \frac{c_0}{a-2R}\,\|\varphi\|\,\|\psi_n^{(a)}\| \tag{1.32}$$

with some positive constants c_0 and R, provided particles k and k' belong to two different clusters.

For the proof of Lemma 6, we need again a suitable auxiliary function χ. Let $k' \in C_{1j}$, $k \in C_{2j}$. We have $\operatorname{supp}\varphi_1^{(n)}\varphi_2^{(n)} \subset B_{R_1}(0)\times B_{R_2}(0)$ and $\operatorname{supp} T_{\mathbf{a}}\varphi_1^{(n)}\varphi_2^{(n)} \subset B_{R_1}(0)\times B_{R_2}(\mathbf{a}_l)$. Hence $x_{k'} < R_1$ and $x_k > a - R_2$. So the inter-electron separation can be estimated by $|\mathbf{x}_k - \mathbf{x}_{k'}| \ge x_k - x_{k'} > a - R_2 - R_1$. Let $R := \max\{R_1, R_2\}$ and $\tilde a := a - 2R$. Define

$$\chi_{kk'}(\frac{\mathbf{x}_k - \mathbf{x}_{k'}}{\tilde a}) := \begin{cases} 0, & |\mathbf{x}_k - \mathbf{x}_{k'}| < \tilde a/2 \\ 1, & |\mathbf{x}_k - \mathbf{x}_{k'}| \ge \tilde a \end{cases}. \tag{1.33}$$

Then $\chi_{kk'}$ is unity on the support of $\psi_n^{(a)}$, such that $\chi_{kk'}\psi_n^{(a)} = \psi_n^{(a)}$. With this function, the proof of Lemma 6 is done exactly as in the two-electron case. Collecting results, we obtain for $n > N_0$ and $a > 4R$ sufficiently large

$$\begin{aligned}\|(h^{BR}-\lambda)\,\psi_n^{(a)}\| \;\le\; \|(T + a_j - \lambda)\,\varphi_n\| \;+\; l\,\frac{2c}{a} \\ +\;\tilde N\,\frac{2c_0}{a}\,\|\psi_n^{(a)}\| \;<\; \epsilon\end{aligned} \tag{1.34}$$

where $\tilde N$ is the total number of two-electron intercluster interactions. This proves that $\lambda \in \sigma(h^{BR})$. Since $\lambda \in [\Sigma_0,\infty)$ was chosen arbitrarily, we therefore have $[\Sigma_0,\infty) \subset \sigma(h^{BR})$, indicating that $\sigma(h^{BR})$ has to be continuous in $[\Sigma_0,\infty)$. Consequently, $[\Sigma_0,\infty) \subset \sigma_{ess}(h^{BR})$ which completes the proof of Theorem 1.

We are left to show that the defining sequence for λ can be chosen to be antisymmetric. We write $\mathcal{A}\psi_n^{(a)} = c_1 \sum\limits_{\sigma\in\mathcal{P}} \operatorname{sign}(\sigma)\,\psi_{n,\sigma}^{(a)}$ where $\psi_{n,\sigma}^{(a)} = \psi_n^{(a)}(\mathbf{x}_{\sigma(1)},...,\mathbf{x}_{\sigma(N)})$ with $\mathcal{P}$ the permutation group of the numbers $1,...,N$, and c_1 is a normalization constant. Since h^{BR} is symmetric upon particle exchange we have

$$\|(h^{BR}-\lambda)\,\mathcal{A}\psi_n^{(a)}\| \;\le\; c_1 \sum_{\sigma\in\mathcal{P}} \|(h^{BR}-\lambda)\,\psi_{n,\sigma}^{(a)}\| \;=\; c_1\,(\#\sigma)\,\|(h^{BR}-\lambda)\,\psi_n^{(a)}\|. \tag{1.35}$$

By (1.34) this can be made smaller than ϵ since the number $\#\sigma$ of permutations is finite.
It remains to prove that $\mathcal{A}\psi_n^{(a)}$ is normalizable. Without restriction we can assume in the factorization $\varphi_n = \varphi_n^{(1)}\cdot\varphi_n^{(2)}$ that $\varphi_n^{(1)}$ and $\varphi_n^{(2)}$ are antisymmetric, such that σ can be restricted to the permutation of coordinates relating to different clusters. We claim that scalar products of the form

$$\begin{aligned}(\varphi_1^{(n)}(\mathbf{x}_1,...,\mathbf{x}_{N-l})\,\varphi_2^{(n)}(\mathbf{x}_{N-l+1}-\mathbf{a},...,\mathbf{x}_N-\mathbf{a}),\;\varphi_1^{(n)}(\mathbf{x}_{\sigma(1)},...,\mathbf{x}_{\sigma(N-l)})\\ \cdot\,\varphi_2^{(n)}(\mathbf{x}_{\sigma(N-l+1)}-\mathbf{a},...,\mathbf{x}_{\sigma(N)}-\mathbf{a})\,)\end{aligned} \tag{1.36}$$

where $\exists\, k\in\{1,...,N-l\}$ and $k'\in\{N-l+1,...,N\}$ such that $\sigma(k)\in\{N-l+1,...,N\}$ and $\sigma(k')\in\{1,...,N-l\}$, can be made arbitrarily small for a suitably large a. In fact, since $x_{\sigma(k')} < R_1$ on $\operatorname{supp}\varphi_1^{(n)}$ and $|\mathbf{x}_{\sigma(k')}-\mathbf{a}| < R_2$ on $\operatorname{supp}\varphi_2^{(n)}$, we have $\int\limits_{\mathbb{R}^3} d\mathbf{x}_{\sigma(k')}\overline{\varphi}_1^{(n)}(...,\mathbf{x}_{\sigma(k')},...)\,\varphi_2^{(n)}(...,\mathbf{x}_{\sigma(k')}-\mathbf{a},...) = 0$ if $a > R_1+R_2$. Thus we get

$$\|A\psi_n^{(a)}\|^2 \;=\; c_1^2 \sum_{\sigma\in\mathcal{P}} (\psi_{n,\sigma}^{(a)},\psi_{n,\sigma}^{(a)}\,) \tag{1.37}$$

since all cross terms vanish for sufficiently large a. This guarantees the normalizability of $\mathcal{A}\psi_n^{(a)}$.

2 The two-electron Jansen-Hess operator

The Jansen-Hess operator includes the terms which are quadratic in the fine structure constant e^2. We restrict ourselves in this section to the two-electron ion and write the Jansen-Hess operator $H^{(2)}$ in the following form [11]

$$H^{(2)} \;=\; H_2^{BR} \;+\; \Lambda_{+,2}\left(\sum_{k=1}^{2} B_{2m}^{(k)} \;+\; C^{(12)}\right)\,\Lambda_{+,2} \tag{2.1}$$

$$B_{2m}^{(k)} := \frac{\gamma^2}{8\pi^2}\left\{\frac{1}{x_k}\left(1-\frac{\boldsymbol{\alpha}^{(k)}\mathbf{p}_k+\beta^{(k)}m}{E_{p_k}}\right)V_{10,m}^{(k)} + V_{10,m}^{(k)}\left(1-\frac{\boldsymbol{\alpha}^{(k)}\mathbf{p}_k+\beta^{(k)}m}{E_{p_k}}\right)\frac{1}{x_k}\right\}$$

$$V_{10,m}^{(k)} := 2\pi^2 \int_0^\infty dt\, e^{-tE_{p_k}}\, \frac{1}{x_k}\, e^{-tE_{p_k}}$$

where H_2^{BR} is the Brown-Ravenhall operator from (1.1) indexed by 2 (for $N = 2$), $\Lambda_{+,2} = \Lambda_+^{(1)} \Lambda_+^{(2)}$ and $V_{10,m}^{(k)}$ is a bounded single-particle integral operator. The two-particle second-order contribution $C^{(12)}$ is given by

$$C^{(12)} := \sum_{k=1}^{2} (V^{(12)}\, \Lambda_-^{(k)}\, F_0^{(k)} \;+\; F_0^{(k)}\, \Lambda_-^{(k)}\, V^{(12)}) \tag{2.2}$$

$$F_0^{(k)} := -\frac{1}{2\pi} \int_{-\infty}^{\infty} d\eta\, \frac{1}{D_0^{(k)} + i\eta}\, V^{(k)}\, \frac{1}{D_0^{(k)} + i\eta}$$

and $\Lambda_-^{(k)} = 1 - \Lambda_+^{(k)}$. Also $F_0^{(k)}$ is a bounded single-particle integral operator. In the same way as for the Brown-Ravenhall operator, an equivalent operator $h^{(2)}$ acting on the reduced spinor space $\mathcal{A}(L_2(\mathbb{R}^3) \otimes \mathbb{C}^2)^2$, can be defined,

$$h^{(2)} \;=\; h_2^{BR} \;+\; \sum_{k=1}^{2} b_{2m}^{(k)} \;+\; c^{(12)} \tag{2.3}$$

with

$$\begin{aligned} b_{2m}^{(k)} \;=\; \frac{\gamma^2}{8\pi^2} \Big\{ & A_k \frac{1}{x_k} V_{10,m}^{(k)} A_k \;-\; G_k \frac{1}{x_k} \frac{\boldsymbol{\sigma}^{(k)}\mathbf{p}_k}{E_{p_k}} V_{10,m}^{(k)} A_k \;-\; A_k \frac{1}{x_k} \frac{m}{E_{p_k}} V_{10,m}^{(k)} A_k \\ & + G_k \frac{1}{x_k} V_{10,m}^{(k)} G_k \;-\; A_k \frac{1}{x_k} \frac{\boldsymbol{\sigma}^{(k)}\mathbf{p}_k}{E_{p_k}} V_{10,m}^{(k)} G_k \;+\; G_k \frac{1}{x_k} \frac{m}{E_{p_k}} V_{10,m}^{(k)} G_k \;+\; \text{h.c.} \Big\} \end{aligned} \tag{2.4}$$

where A_k, G_k are defined below (1.2) and h.c. means Hermitean conjugate (such that $b_{2m}^{(k)}$ is a symmetric operator). Note that, due to the presence of the projector $\Lambda_+^{(k)}$ in (2.1), $b_{2m}^{(k)}$ contains only even powers in $\boldsymbol{\sigma}^{(k)}$. In a similar way, $c^{(12)}$ is derived from $C^{(12)}$. The particle mass m is assumed to be nonzero throughout (for $m = 0$, the spectrum of the single-particle Jansen-Hess operator is absolutely continuous with infimum zero [11]).

For potential strength $\gamma < 0.89$ (slightly smaller than γ_{BR}), it was shown [13] that the total potential of $H^{(2)}$ (and hence also of $h^{(2)}$) is relatively form bounded (with form bound smaller than 1) with respect to the kinetic energy operator. Therefore, $h^{(2)}$ is well-defined in the form sense and is a self-adjoint operator by means of the Friedrichs extension of the restriction of $h^{(2)}$ to $\mathcal{A}(C_0^\infty(\mathbb{R}^3) \otimes \mathbb{C}^2)^2$. The above form boundedness guarantees the existence of a $\mu > 0$ such that $h^{(2)} + \mu > 0$ for $\gamma < 0.89$. If $\gamma < 0.825$, one can even choose $\mu = 0$ [13].

Let us introduce the operator $\tilde{h}^{(2)}$ by means of $h^{(2)} =: \tilde{h}^{(2)} + c^{(12)}$ and define in analogy to (1.4) the two-cluster decompositions of $\tilde{h}^{(2)}$ for j=0,1,2,

$$\tilde{h}^{(2)} \;=\; T \;+\; a_j \;+\; r_j \tag{2.5}$$

as well as

$$\Sigma_0 := \min_j \inf \sigma(T + a_j). \tag{2.6}$$

The aim of this section is to prove

Theorem 2 (HVZ theorem for the two-electron Jansen-Hess operator).
Let $h^{(2)} = \sum_{k=1}^{2} (T^{(k)} + b_{1m}^{(k)} + b_{2m}^{(k)}) + v^{(12)} + c^{(12)} = \tilde{h}^{(2)} + c^{(12)}$ *be the two-electron Jansen-Hess operator with potential strength* $\gamma < 0.66$ $(Z \leq 90)$. *Let (2.5) be the two-cluster decompositions of* $\tilde{h}^{(2)}$ *and* Σ_0 *from (2.6). Then the essential spectrum of* $h^{(2)}$ *is given by*

$$\sigma_{ess}(h^{(2)}) \;=\; [\Sigma_0, \infty). \tag{2.7}$$

We start by noting that the two-particle second-order potential $c^{(12)}$ does not change the essential spectrum of $h^{(2)}$:

Proposition 1. *Let* $h^{(2)} = \tilde{h}^{(2)} + c^{(12)}$ *be the two-electron Jansen-Hess operator with potential strength* $\gamma < 0.66$. *Then one has*

$$\sigma_{ess}(h^{(2)}) \;=\; \sigma_{ess}(\tilde{h}^{(2)}). \tag{2.8}$$

Proof.
The proof is performed for the equivalent operator $H^{(2)} =: \tilde{H}^{(2)} + \Lambda_{+,2} C^{(12)} \Lambda_{+,2}$.
The resolvent difference

$$R_\mu := \; (H^{(2)} + \mu)^{-1} \; - \; (\tilde{H}^{(2)} + \mu)^{-1} \tag{2.9}$$

is bounded for $\mu \geq 0$ since $H^{(2)}$ as well as $\tilde{H}^{(2)}$ are positive for $\gamma < 0.825$ which exceeds the critical γ of Proposition 1. We will show that R_μ is compact. Then, following the argumentation of [7], one can use Lemma 3 of [20, p.111] together with the strong spectral mapping theorem ([20, p.109]) to prove that the essential spectra of $H^{(2)}$ and $\tilde{H}^{(2)}$ coincide.

Let $T_0 := \Lambda_{+,2} \sum_{k=1}^{2} D_0^{(k)} \Lambda_{+,2}$ which is a positive operator (for $m \neq 0$) on the positive spectral subspace $\Lambda_{+,2} \mathcal{A}(H_1(\mathbb{R}^3) \otimes \mathbb{C}^4)^2$. (The negative spectral subspace is disregarded throughout because $H^{(2)} = 0$ on that subspace.) With the help of the second resolvent identity, one decomposes R_μ into

$$R_\mu \;=\; -(\tilde{H}^{(2)} + \mu)^{-1} \, \Lambda_{+,2} \, C^{(12)} \, \Lambda_{+,2} \, (H^{(2)} + \mu)^{-1} \tag{2.10}$$

$$= - \left[(\tilde{H}^{(2)} + \mu)^{-1} (T_0 + \mu) \right] \cdot \left\{ (T_0 + \mu)^{-1} \, \Lambda_{+,2} \, C^{(12)} \, \Lambda_{+,2} \, (T_0 + \mu)^{-1} \right\}$$

$$\cdot \left[(T_0 + \mu)(H^{(2)} + \mu)^{-1} \right].$$

One can show (see [11, proof b) of Theorem II.1 with T replaced by T_0]) that for $\gamma < 0.66$, the two operators in square brackets are bounded. This relies on the relative boundedness of the total potential of $H^{(2)}$ (respective $\tilde{H}^{(2)}$) with respect to T_0, with (operator) bound less than one for $m = 0$ ([11]; Appendix B). Due to scaling (for $\mu = 0$), the boundedness of the operators in square brackets holds for all m. The operator in curly brackets is shown to be compact. To this aim it is written as

$$(T_0 + \mu)^{-1} \Lambda_{+,2} \, C^{(12)} \Lambda_{+,2} \, (T_0 + \mu)^{-1} \tag{2.11}$$

$$= (T_0 + \mu)^{-1} (T + \mu) \, \Lambda_{+,2} W_2 \, \Lambda_{+,2} \, (T + \mu)(T_0 + \mu)^{-1}$$

with $W_2 := (T + \mu)^{-1} C^{(12)} (T + \mu)^{-1}$. According to Herbst [8], W_2 is decomposed into $W_{2n} + R_n$ where $(W_{2n})_{n \in \mathbb{N}}$ is a sequence of Hilbert-Schmidt operators satisfying $\|W_{2n} - W_2\| \to 0$ for $n \to \infty$. It follows that W_{2n} is compact such that also W_2 is compact (see e.g. [15, III.4.2,V.2.4]). W_{2n} is defined by regularizing the Coulomb potential by means of $\frac{1}{x} \, e^{-\epsilon x}$ and by introducing convergence generating functions $e^{-\epsilon p}$ in momentum space, where $\epsilon := \frac{1}{n} > 0$ is a small quantity. Details of the proof are found in [11]. The adjacent factors of W_2 in (2.11) are easily seen to be bounded for $\mu = 0$. Since $\Lambda_{+,2} = \Lambda_{+,2}^2$ commutes with T, one has e.g. $\Lambda_{+,2} T T_0^{-1} = (\Lambda_{+,2} T \Lambda_{+,2}) T_0^{-1} = T_0 T_0^{-1} = 1$. Therefore, the operator in curly brackets and hence R_μ is proven to be compact for $\mu = 0$. $\square$

Proof of Theorem 2. With Proposition 1 at hand, it remains to prove the HVZ theorem for the operator $\tilde{h}^{(2)}$, which in fact holds for all $\gamma < \gamma_{BR}$.
We proceed along the same lines as done in the proof of the HVZ theorem for the Brown-Ravenhall operator. It is thus only necessary to extend Lemmata 1,3,4 and 5 to the operator $\tilde{h}^{(2)}$ which is obtained from h^{BR} by including the single-particle second-order potentials $b_{2m}^{(k)}$, $k = 1, 2$. We start with the lemmata required for the 'hard part' of the proof.

a) In the formulation of Lemma 1 we simply replace h^{BR} by $\tilde{h}^{(2)}$ throughout (and take $N = 2$).
In order to prove $\|[\tilde{h}^{(2)}, \chi_0] \, \psi_n\| \leq \frac{c}{n} \, \|\psi_n\|$ with $\psi_n \in \mathcal{A}(C_0^\infty(\mathbb{R}^3) \otimes \mathbb{C}^2)^2$ a Weyl sequence for $\lambda \in \sigma_{ess}(\tilde{h}^{(2)})$ and χ_0 from (1.13) with $\mathbf{x} := (\mathbf{x}_1, \mathbf{x}_2)$, we have to show in addition to the Brown-Ravenhall case,

$$|(\phi, \, [b_{2m}^{(1)}, \chi_0] \, \psi_n)| \, \leq \, \frac{c}{n} \, \|\phi\| \, \|\psi_n\| \tag{2.12}$$

for all $\phi \in C_0^\infty(\mathbb{R}^6) \otimes \mathbb{C}^4$. Due to the symmetry property of ψ_n, the same bound holds also for $[b_{2m}^{(2)}, \chi_0]$. The operator $b_{2m}^{(1)}$ defined in (2.4) is a sum of terms of the structure $B(\mathbf{p}_1) \frac{1}{x_1} B_\lambda(\mathbf{p}_1) \, V_{10,m}^{(1)} B_\mu(\mathbf{p}_1)$ where $B(\mathbf{p}_1) \in \{A_1, G_1\}$ like for $b_{1m}^{(1)}$ whereas $B_\lambda(\mathbf{p}_1), B_\mu(\mathbf{p}_1) \in \{1, \frac{\boldsymbol{\sigma}^{(1)} \mathbf{p}_1}{E_{p_1}}, \frac{m}{E_{p_1}}, A_1, G_1\}$ are all analytic, bounded multiplication operators in momentum space. We pick for the sake of demonstration the second term of (2.4) and decompose

$$[G_1 \frac{1}{x_1} \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}} V_{10,m}^{(1)} A_1,\, \chi_0] \;=\; [G_1, \chi_0]\, p_1 \cdot \frac{1}{p_1 x_1} \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}} V_{10,m}^{(1)} A_1$$

$$+\, G_1 \frac{1}{x_1 p_1} \cdot p_1\, [\frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}}, \chi_0]\, V_{10,m}^{(1)} A_1 +\, G_1 \frac{1}{x_1 p_1} \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}} \cdot p_1\, [V_{10,m}^{(1)}, \chi_0]\, A_1$$

$$+\, G_1 \frac{1}{x_1 p_1} \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}} \cdot p_1\, V_{10,m}^{(1)} \frac{1}{p_1} \cdot p_1\, [A_1, \chi_0]. \tag{2.13}$$

We will show that the commutators (including the factor p_1) are $\frac{1}{n}$-bounded and the adjacent factors bounded. The latter is trivial (since also $\frac{1}{x_1 p_1}$ is bounded, $\|\frac{1}{x_1 p_1}\| = 2$, see e.g. [8]) except for the operator $p_1 V_{10,m}^{(1)} \frac{1}{p_1}$ in the last term. The boundedness of this operator is readily proved by invoking its kernel in momentum space. From (2.1) we have

$$\|\, p_1\, V_{10,m}^{(1)} \frac{1}{p_1} \| \;=\; 2\pi^2\, \| \int_0^\infty dt\; e^{-tE_{p_1}}\, p_1\, \frac{1}{x_1 p_1}\, e^{-tE_{p_1}}\, \|$$

$$\leq\; 2\pi^2\, \| \int_0^\infty dt\; e^{-tE_{p_1}}\, p_1 \,\| \cdot \| \frac{1}{x_1 p_1} \| \cdot \| e^{-tE_{p_1}} \| \;\leq\; 4\pi^2 \tag{2.14}$$

where $\int\limits_0^\infty dt\, e^{-tE_{p_1}} = 1/E_{p_1}$ has been used.

The commutators $p_1\,[A_1, \chi_0]$ and $p_1\,[G_1, \chi_0]$ have already been dealt with in the context of the Brown-Ravenhall operator. $p_1\,[\frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}}, \chi_0]$ is of the same type, because for any B_λ, one has the estimate $|B_\lambda(\mathbf{p}_1) - B_\lambda(\mathbf{p}_1')| = |\mathbf{p}_1 - \mathbf{p}_1'|\, |\nabla_{\mathbf{p}_1} B_\lambda(\boldsymbol{\xi})| \;\leq\; |\mathbf{p}_1 - \mathbf{p}_1'|\, \frac{c_0}{1+p_1}$ from the mean value theorem, where $\boldsymbol{\xi}$ is some point between $\mathbf{p}_1$ and $\mathbf{p}_1'$. For the commutator with $V_{10,m}^{(1)}$ we have

$$p_1\,[V_{10,m}^{(1)}, \chi_0] \;=\; 2\pi^2 \int_0^\infty dt\; p_1\, [e^{-tE_{p_1}}, \chi_0]\, \frac{1}{x_1}\, e^{-tE_{p_1}}$$

$$+\, 2\pi^2 \int_0^\infty dt\; p_1\, e^{-tE_{p_1}}\, \frac{1}{x_1}\, [e^{-tE_{p_1}}, \chi_0]. \tag{2.15}$$

The proof of its $\frac{1}{n}$-boundedness proceeds with the help of the Lieb and Yau formula [17], derived from the Schwarz inequality (see also [14, Lemma 7]), in momentum space. Explicitly, in the estimate

$$|(\hat{\phi}, \widehat{\mathcal{O}\,\psi_n})| \;\leq\; \left(\int_{\mathbb{R}^6} d\mathbf{p}\; |\hat{\phi}(\mathbf{p})|^2\; I(\mathbf{p}) \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^6} d\mathbf{p}'\; |\hat{\psi}_n(\mathbf{p}')|^2\; J(\mathbf{p}') \right)^{\frac{1}{2}} \tag{2.16}$$

where $\mathcal{O} := p_1[V_{10,m}^{(1)}, \chi_0]$ and $k_{\mathcal{O}}$ its kernel, one has to show that the integrals I and J obey

$$I(\mathbf{p}) := \int_{\mathbb{R}^6} d\mathbf{p}'\; |k_{\mathcal{O}}(\mathbf{p}, \mathbf{p}')|\; \frac{f(\mathbf{p})}{f(\mathbf{p}')} \;\leq\; \frac{c}{n}$$

$$J(\mathbf{p}') := \int_{\mathbb{R}^6} d\mathbf{p}\, |k_{\mathcal{O}}(\mathbf{p},\mathbf{p}')|\, \frac{f(\mathbf{p}')}{f(\mathbf{p})} \;\leq\; \frac{c}{n} \tag{2.17}$$

with some constant c (independent of $\mathbf{p},\mathbf{p}'$) for a suitably chosen nonnegative convergence generating function f.

We use the two-dimensional ($N = 2$) Fourier transform (1.14) of χ_0 and the momentum representation of $\frac{1}{x_1}$ to write for the first term in (2.15),

$$\Big(\widehat{\int_0^\infty dt p_1 [e^{-tE_{p_1}}, \chi_0] \frac{1}{x_1} e^{-tE_{p_1}} \varphi}\Big)(\mathbf{p}_1,\mathbf{p}_2) = \int_{\mathbb{R}^6} d\mathbf{q} d\mathbf{p}_2' k_1(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\mathbf{p}_2')\hat{\varphi}(\mathbf{q},\mathbf{p}_2')$$

$$k_1(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\mathbf{p}_2') := \frac{1}{(2\pi)^3}\,\frac{1}{2\pi^2}\, p_1\, n^6 \int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{p}_1'\; \hat{\chi}_0(n(\mathbf{p}_1-\mathbf{p}_1'), n(\mathbf{p}_2-\mathbf{p}_2'))$$
$$\cdot (e^{-tE_{p_1}} - e^{-tE_{p_1'}})\, \frac{1}{|\mathbf{p}_1'-\mathbf{q}|^2}\, e^{-tE_q}. \tag{2.18}$$

From the mean value theorem we get

$$|e^{-tE_{p_1}} - e^{-tE_{p_1'}}| \;=\; |\mathbf{p}_1-\mathbf{p}_1'|\, |te^{-tE_\xi}\, \frac{\boldsymbol{\xi}}{E_\xi}| \;\leq\; |\mathbf{p}_1-\mathbf{p}_1'|\, t\, e^{-tE_\xi} \tag{2.19}$$

with $\boldsymbol{\xi} = \lambda \mathbf{p}_1' + (1-\lambda)\mathbf{p}_1$ for some $\lambda \in [0,1]$. We have to show that the integral over the modulus of the kernel of (2.18), with a suitable convergence generating function f, is $\frac{1}{n}$-bounded. We choose $f(p) = p$ and make the substitution $\mathbf{y}_i := n(\mathbf{p}_i - \mathbf{p}_i')$ for $\mathbf{p}_i'$, $i = 1,2$. Then

$$I(\mathbf{p}_1,\mathbf{p}_2) := \int_{\mathbb{R}^6} d\mathbf{q}\, d\mathbf{p}_2'\; |k_1(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\mathbf{p}_2')|\; \frac{f(p_1)}{f(q)}$$
$$\leq \frac{1}{(2\pi)^4\pi}\, p_1 \int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{q} \int_{\mathbb{R}^6} d\mathbf{y}_1\, d\mathbf{y}_2\; |\hat{\chi}_0(\mathbf{y}_1,\mathbf{y}_2)| \tag{2.20}$$
$$\cdot\, \frac{y_1}{n}\, t\, e^{-tE_\xi}\; \frac{1}{|\mathbf{q}-(\mathbf{p}_1-\mathbf{y}_1/n)|^2}\; e^{-tE_q} \cdot \frac{p_1}{q}.$$

The t-integral can be carried out, $\int\limits_0^\infty dt\, te^{-(E_\xi+E_q)} = (E_\xi + E_q)^{-2}$ with $\boldsymbol{\xi} = \mathbf{p}_1 - \frac{\lambda}{n}\mathbf{y}_1$. Define $\mathbf{q}_1 := \mathbf{p}_1 - \mathbf{y}_1/n$ and consider

$$S := p_1^2 \int_{\mathbb{R}^3} d\mathbf{q}\; \frac{1}{|\mathbf{q}-\mathbf{q}_1|^2}\; \frac{1}{q}\; \frac{1}{(E_\xi+E_q)^2}. \tag{2.21}$$

Estimating the last factor by $\frac{1}{E_\xi} \cdot \frac{1}{q}$ and performing the angular integration, one obtains

$$S \;\leq\; p_1^2\, \frac{2\pi}{q_1}\, \frac{1}{E_\xi} \int_0^\infty \frac{dq}{q}\, \ln\left|\frac{q+q_1}{q-q_1}\right| \;=\; \pi^3\, \frac{p_1}{|\mathbf{p}_1-\mathbf{y}_1/n|}\; \frac{p_1}{\sqrt{(\mathbf{p}_1-\frac{\lambda}{n}\mathbf{y}_1)^2+m^2}}. \tag{2.22}$$

Insertion into (2.20) gives

$$I(\mathbf{p}_1,\mathbf{p}_2) \;\leq\; \frac{1}{(4\pi)^2}\,\frac{1}{n}\int_{\mathbb{R}^6} d\mathbf{y}_1\,d\mathbf{y}_2\,|\hat{\chi}_0(\mathbf{y}_1,\mathbf{y}_2)|\,y_1$$

$$\cdot\frac{p_1}{|\mathbf{p}_1-\mathbf{y}_1/n|}\,\frac{p_1}{\sqrt{(\mathbf{p}_1-\frac{\lambda}{n}\mathbf{y}_1)^2+m^2}} \;\leq\; \frac{c}{n} \tag{2.23}$$

because the singularity at $\mathbf{y}_1 = n\mathbf{p}_1$ is integrable and the integral is finite for all $p_1 \geq 0$ due to $\hat{\chi}_0 \in \mathcal{S}(\mathbb{R}^6)$. Since the kernel k_1 is not symmetric in $\mathbf{p}_1, \mathbf{q}$, the estimate of $J(\mathbf{q},\mathbf{p}_2') := \int_{\mathbb{R}^6} d\mathbf{p}_1\,d\mathbf{p}_2\,|k_1(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\mathbf{p}_2')|\,\frac{f(q)}{f(p_1)}$ is needed too. The $\frac{1}{n}$-boundedness of $J(\mathbf{q},\mathbf{p}_2')$ can be shown along the same lines, using $(E_{\xi'}+E_q)^{-2} \leq \xi'^{-2}$ with $\boldsymbol{\xi}' := \mathbf{p}_1' + \lambda(\mathbf{p}_1-\mathbf{p}_1')$.
We still have to estimate the second term in (2.15). Its kernel is

$$k_2(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\mathbf{p}_2') := \frac{1}{(2\pi)^3}\,\frac{1}{2\pi^2}\,p_1\,n^6\int_0^\infty dt\;e^{-tE_{p_1}}$$

$$\int_{\mathbb{R}^3} d\mathbf{p}_1'\;\frac{1}{|\mathbf{p}_1-\mathbf{p}_1'|^2}\;\hat{\chi}_0(n(\mathbf{p}_1'-\mathbf{q}),n(\mathbf{p}_2-\mathbf{p}_2'))\;(e^{-tE_{p_1'}}\;-\;e^{-tE_q}). \tag{2.24}$$

With (2.19) the t-integral can be carried out as before. Making the substitution $\mathbf{y}_1 := n(\mathbf{p}_1'-\mathbf{q})$, $\mathbf{y}_2 := n(\mathbf{p}_2-\mathbf{p}_2')$ for $\mathbf{q}$ and $\mathbf{p}_2'$, respectively, one gets with the choice $f(p) = p^{\frac{1}{2}}$,

$$\tilde{I}(\mathbf{p}_1,\mathbf{p}_2) := \int_{\mathbb{R}^6} d\mathbf{q}\,d\mathbf{p}_2'\;|k_2(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\mathbf{p}_2')|\;\frac{f(p_1)}{f(q)} \tag{2.25}$$

$$\leq \frac{1}{(2\pi)^4\pi}p_1\int_{\mathbb{R}^6}d\mathbf{y}_1d\mathbf{y}_2|\hat{\chi}_0(\mathbf{y}_1,\mathbf{y}_2)|\int_{\mathbb{R}^3}d\mathbf{p}_1'\frac{1}{|\mathbf{p}_1-\mathbf{p}_1'|^2}\,\frac{y_1}{n}\,\frac{1}{(E_{p_1}+E_{\tilde{\xi}})^2}\cdot\frac{p_1^{\frac{1}{2}}}{|\mathbf{p}_1'-\frac{\mathbf{y}_1}{n}|^{\frac{1}{2}}}$$

with $\tilde{\boldsymbol{\xi}} := \lambda\mathbf{q}+(1-\lambda)\mathbf{p}_1' = \mathbf{p}_1'-\frac{\lambda}{n}\mathbf{y}_1$. We estimate $(E_{p_1}+E_{\tilde{\xi}})^{-2} \leq p_1^{-\frac{1}{2}}E_{\tilde{\xi}}^{-\frac{3}{2}}$.
Then the integral over $\mathbf{p}_1'$ reduces to

$$\tilde{S} := p_1\int_{\mathbb{R}^3} d\mathbf{p}_1'\;\frac{1}{|\mathbf{p}_1-\mathbf{p}_1'|^2}\;\frac{1}{|\mathbf{p}_1'-\frac{\mathbf{y}_1}{n}|^{\frac{1}{2}}}\;\frac{1}{[(\mathbf{p}_1'-\frac{\lambda}{n}\mathbf{y}_1)^2+m^2]^{\frac{3}{4}}}. \tag{2.26}$$

Even when the two singularities coincide (for $\mathbf{y}_1 = n\mathbf{p}_1$), they are integrable. Since the integrand behaves like $p_1'^{-2}$ for $p_1' \to \infty$, $\tilde{S}$ is finite for all $0 \leq p_1 < \infty$. It remains to estimate $\tilde{S}$ for $p_1 \to \infty$. We substitute $p_1\mathbf{x} := \mathbf{p}_1'-\mathbf{p}_1$, such that with $\mathbf{e}_{p_1} := \mathbf{p}_1/p_1$,

$$\tilde{S} \;=\; \int_{\mathbb{R}^3}\frac{d\mathbf{x}}{x^2}\;\frac{1}{|\mathbf{x}+\mathbf{e}_{p_1}-\frac{\mathbf{y}_1}{np_1}|^{\frac{1}{2}}}\;\frac{1}{[(\mathbf{x}+\mathbf{e}_{p_1}-\frac{\lambda}{np_1}\mathbf{y}_1)^2+\frac{m^2}{p_1^2}]^{\frac{3}{4}}}$$

$$\longrightarrow\;\int_{\mathbb{R}^3}\frac{d\mathbf{x}}{x^2}\;\frac{1}{|\mathbf{x}+\mathbf{e}_{p_1}|^2}\;=\;\pi^3 \qquad \text{as } p_1 \to \infty. \tag{2.27}$$

Therefore, $\tilde{I}$ is $\frac{1}{n}$-bounded for all $p_1 \geq 0$. It is easy to prove that also $\tilde{J}(\mathbf{q},\mathbf{p}_2') := \int_{\mathbb{R}^6} d\mathbf{p}_1\, d\mathbf{p}_2\, |k_2(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\mathbf{p}_2')|\, \frac{q^{\frac{1}{2}}}{p_1^{\frac{1}{2}}}$ is $\frac{1}{n}$-bounded, using the estimate $(E_{p_1} + E_{|\mathbf{q}+\lambda(\mathbf{p}_1'-\mathbf{q})|})^{-2} \leq p_1^{-2}$.

Collecting results, this shows the $\frac{1}{n}$-boundedness of $p_1[V_{10,m}^{(1)},\chi_0]$. With the same tools, the $\frac{1}{n}$-boundedness of the commutator of χ_0 with the remaining contributions from (2.4) to $b_{2m}^{(1)}$ is established.

The second item of Lemma 1, the normalizability of the sequence $\varphi_n := (1-\chi_0)\psi_n$, follows immediately from the proof concerning the Brown-Ravenhall operator, because of the relative form boundedness of the total potential of $\tilde{h}^{(2)}$ with form bound smaller than one for $\gamma < \gamma_{BR}$ (see [13] and Lemma 7).

b) In the formulation of Lemma 3, the only change is again the replacement of h^{BR} with $\tilde{h}^{(2)}$ (and $N=2$).

We consider the case $j=1$ where $r_1 = b_{1m}^{(1)} + b_{2m}^{(1)} + v^{(12)}$, and we have to show in addition to the Brown-Ravenhall case that

$$|(\phi_1\varphi, b_{2m}^{(1)}\,\phi_1\varphi)| \;\leq\; \frac{c}{R}\,\|\varphi\|^2 \tag{2.28}$$

provided $\varphi \in \mathcal{A}(C_0^\infty(\mathbb{R}^6\backslash B_R(0)) \otimes \mathbb{C}^4)$ and $R>1$.

We note that every summand of $b_{2m}^{(1)}$ in (2.4) is of the form $B_1 \frac{1}{x_1} W_1$ or $W_1 \frac{1}{x_1} B_1$ where B_1 is a bounded multiplication operator in momentum space, while W_1 is a bounded integral operator. For operators of the first type we take the smooth auxiliary function $\chi_1(\frac{\mathbf{x}_1}{R})$ from (1.17) which is unity on the support of $\phi_1\varphi$ and decompose

$$(\chi_1\phi_1\varphi, B_1\frac{1}{x_1}W_1\,\phi_1\varphi) \;=\; (\phi_1\varphi, B_1\chi_1\frac{1}{x_1}W_1\,\phi_1\varphi) \;+\; (\phi_1\varphi, [\chi_1, B_1]\,\frac{1}{x_1}\,W_1\,\phi_1\varphi). \tag{2.29}$$

Since $\operatorname{supp}\chi_1 \subset \mathbb{R}^3\backslash B_{CR/2}(0)$ we have

$$|(B_1\phi_1\varphi, \chi_1\frac{1}{x_1}W_1\phi_1\varphi)| \leq \frac{2}{CR}\int_{\mathbb{R}^6} d\mathbf{x}_1 d\mathbf{x}_2 |(B_1\phi_1\varphi)(\mathbf{x}_1,\mathbf{x}_2)|\;\chi_1|(W_1\phi_1\varphi)(\mathbf{x}_1,\mathbf{x}_2)|$$

$$\leq\; \frac{2}{CR}\,\|B_1\|\;\|\varphi\|\;\|W_1\|\;\|\varphi\| \;\leq\; \frac{c_0}{R}\,\|\varphi\|^2. \tag{2.30}$$

For the second contribution to (2.29), we have to estimate $[\chi_{1,0},B_1]p_1$ with $\chi_{1,0} := 1-\chi_1$ in momentum space. Since $B_1 \in \{A_1, G_1\}$ we use the relation $|(\tilde{\varphi}, [\chi_{1,0},B_1]p_1\,\tilde{\psi})| = |(\tilde{\psi}, p_1[B_1,\chi_{1,0}]\,\tilde{\varphi})|$ (for suitable $\tilde{\varphi},\tilde{\psi}$), the uniform $\frac{1}{R}$-boundedness of which has already been proven in the context of the Brown-Ravenhall case. The second operator, $W_1\frac{1}{x_1}B_1$ is treated in the same way, using $W_1\frac{1}{x_1}B_1\chi_1\,\phi_1\varphi = W_1\frac{1}{x_1}\chi_1 B_1\,\phi_1\varphi + W_1\frac{1}{x_1}[B_1,\chi_1]\,\phi_1\varphi$.

In the case $j=0$ we have $r_0 = \sum_{k=1}^{2}(b_{1m}^{(k)} + b_{2m}^{(k)})$, and since $\operatorname{supp}\phi_0$ requires $x_1 \geq Cx$ as well as $x_2 \geq Cx$, $\mathbf{x} = (\mathbf{x}_1,\mathbf{x}_2)$, the auxiliary function can be taken from (1.17) for $k = 1$ or $k = 2$. The further proof of the lemma is identical to the one for $j = 1$.

c) Lemma 4 (formulated for $\tilde{h}^{(2)}$ in place of h^{BR}) which is needed for the localization formula, has to be supplemented with the following estimate

$$(d) \qquad |(\phi_j \varphi, [b_{2m}^{(k)}, \phi_j]\, \varphi)| \;\leq\; \frac{c}{R}\, \|\varphi\|^2 \tag{2.31}$$

for $\varphi \in \mathcal{A}(C_0^\infty(\mathbb{R}^6 \backslash B_R(0)) \otimes \mathbb{C}^4)$ and $R > 2$.

The proof is carried out in coordinate space as are the proofs of the Brown-Ravenhall items of Lemma 4. We split the commutator in the same way as in the proof of Lemma 1. In order to show how to proceed, we pick again the second term of (2.4), take $k = 1$ and decompose

$$[G_1 \frac{1}{x_1} \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}} V_{10,m}^{(1)} A_1, \phi_j] \;=\; [G_1, \phi_j]\, \frac{1}{x_1} \cdot \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}} V_{10,m}^{(1)} A_1$$

$$+\, G_1 \frac{1}{x_1} [\frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}}, \phi_j]\, V_{10,m}^{(1)} A_1 \;+\; G_1 \frac{1}{x_1} \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}}\, x_1 \cdot \frac{1}{x_1}\, [V_{10,m}^{(1)}, \phi_j]\, A_1 \tag{2.32}$$

$$+\, G_1 \frac{1}{x_1} \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}} V_{10,m}^{(1)}\, x_1 \cdot \frac{1}{x_1}\, [A_1, \phi_j].$$

We have to prove the $\frac{1}{R}$-boundedness of the commutators (including the factor $\frac{1}{x_1}$) and to assure the boundedness of the adjacent operators. The commutators with G_1 and A_1 have already been dealt with in the Brown-Ravenhall case. As concerns $[\frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}}, \phi_j]\frac{1}{x_1}$, we have to show that its kernel obeys the estimate

$$|\check{k}_{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1 \frac{1}{E_{p_1}}}(\mathbf{x}_1, \mathbf{x}_1')| \;\leq\; \frac{c}{|\mathbf{x}_1 - \mathbf{x}_1'|^3} \tag{2.33}$$

with some constant c. When dealing with the Brown-Ravenhall operator, we have shown the corresponding estimate for the kernel of the operator $\boldsymbol{\sigma}^{(1)}\mathbf{p}_1 g(p_1)$ with $g(p_1) = [2(p_1^2 + m^2 + m\sqrt{p_1^2 + m^2}\,)]^{-\frac{1}{2}}$. Replacing $g(p_1)$ with $(p_1^2 + m^2)^{-\frac{1}{2}}$ does neither change the analyticity property of the kernel nor its behaviour as $|\mathbf{x}_1 - \mathbf{x}_1'|$ tends to 0 or infinity, from which (2.33) is established [14].

For the further proof of the $\frac{1}{R}$-boundedness of the commutator, we can substitute ϕ_j with $\phi_j \chi$ where $\chi(\frac{\mathbf{x}}{R})$ is defined in (1.20) with $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ (see the discussion below (1.20)). Thus we can use the estimate (1.21) (for $k = 1$ and $N = 2$) derived from the mean value theorem and mimic the proof of the two-electron Brown-Ravenhall case.

For the treatment of the remaining commutator, $[V_{10,m}^{(1)}, \phi_j \chi]\frac{1}{x_1}$, we set $\psi_j := \phi_j \chi$ and decompose

$$[V_{10,m}^{(1)}, \psi_j]\, \frac{1}{x_1} \;=\; [V_{10,m}^{(1)}\, \frac{1}{x_1}, \psi_j] \tag{2.34}$$

$$= \; 2\pi^2 \int_0^\infty dt\,[e^{-tE_{p_1}},\psi_j]\,\frac{1}{x_1}\,e^{-tE_{p_1}}\,\frac{1}{x_1} \,+\, 2\pi^2 \int_0^\infty dt\,e^{-tE_{p_1}}\,\frac{1}{x_1}\,[e^{-tE_{p_1}},\psi_j]\,\frac{1}{x_1}.$$

The kernel of $e^{-tE_{p_1}}$ in coordinate space is given by [17]

$$\check{k}_{e^{-tE_{p_1}}}(\mathbf{x}_1,\mathbf{x}_1',t) \;=\; \check{k}_{e^{-tE_{p_1}}}(\tilde{\mathbf{x}},t) \;=\; \frac{t}{2\pi^2}\,\frac{m^2}{\tilde{x}^2+t^2}\,K_2(m\sqrt{\tilde{x}^2+t^2}\,) \tag{2.35}$$

where K_2 is a modified Bessel function of the second kind and $\tilde{\mathbf{x}} := \mathbf{x}_1 - \mathbf{x}_1'$. Making use of the analyticity of $K_2(z)$ for $z>0$ and its behaviour $K_2(z) \sim \frac{2}{z^2}$ for $z \to 0$ and $K_2(z) \sim \sqrt{\frac{\pi}{2z}}\,e^{-z}$ for $z \to \infty$ [1, p.377] we have

$$|K_2(z)| \;\le\; \frac{2c_1}{z^2} \tag{2.36}$$

and therefore we can estimate $\check{k}_{e^{-tE_{p_1}}}$ by the corresponding kernel for $m=0$,

$$|\check{k}_{e^{-tE_{p_1}}}(\tilde{\mathbf{x}},t)| \;\le\; \frac{t}{\pi^2}\,\frac{c_1}{(\tilde{x}^2+t^2)^2} \;=\; c_1\,\check{k}_{e^{-tp_1}}(\tilde{\mathbf{x}},t). \tag{2.37}$$

Thus we obtain for the kernel of the second contribution to (2.34), using (2.37) and (1.21),

$$S_0 := \left|\int_0^\infty dt\;\check{k}_{e^{-tE_{p_1}}\frac{1}{x_1}[e^{-tE_{p_1}},\psi_j]\frac{1}{x_1}}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2)\right| \tag{2.38}$$

$$= \left|\int_0^\infty dt \int_{\mathbb{R}^3} d\mathbf{x}_1'\,\frac{t}{2\pi^2}\,\frac{m^2}{(\mathbf{x}_1-\mathbf{x}_1')^2+t^2}\,K_2(m\sqrt{(\mathbf{x}_1-\mathbf{x}_1')^2+t^2}\,)\,\frac{1}{x_1'}\right.$$

$$\left.\cdot\frac{t}{2\pi^2}\,\frac{m^2}{(\mathbf{x}_1'-\mathbf{y}_1)^2+t^2}\;\;K_2(m\sqrt{(\mathbf{x}_1'-\mathbf{y}_1)^2+t^2}\,)\,\frac{1}{y_1}\;(\psi_j(\mathbf{y}_1,\mathbf{x}_2)-\psi_j(\mathbf{x}_1',\mathbf{x}_2))\right|$$

$$\le\; \frac{c_1^2}{\pi^4}\int_0^\infty t^2\,dt\int_{\mathbb{R}^3} d\mathbf{x}_1'\;\frac{1}{[(\mathbf{x}_1-\mathbf{x}_1')^2+t^2]^2}\,\frac{1}{x_1'}\;\frac{1}{[(\mathbf{x}_1'-\mathbf{y}_1)^2+t^2]^2}\,\frac{1}{y_1}\cdot|\mathbf{x}_1'-\mathbf{y}_1|\,\frac{c_0}{R}.$$

With the help of the estimate $\frac{1}{[(\mathbf{x}_1'-\mathbf{y}_1)^2+t^2]^2} \le \frac{1}{t}\,\frac{1}{|\mathbf{x}_1'-\mathbf{y}_1|^3}$, the t-integral can be carried out,

$$\int_0^\infty dt\;\frac{t}{[(\mathbf{x}_1-\mathbf{x}_1')^2+t^2]^2} \;=\; \frac{1}{2\,|\mathbf{x}_1-\mathbf{x}_1'|^2}. \tag{2.39}$$

According to the Lieb and Yau formula (2.16) in coordinate space, the $\frac{1}{R}$-boundedness of S_0 integrated over $\mathbf{y}_1$, respectively over $\mathbf{x}_1$, with a suitably chosen convergence generating function f, has to be shown (in analogy to (2.17)).

With the choice $f(x) = x^\alpha$ and (2.39) we have

$$I(\mathbf{x}_1) := \int_{\mathbb{R}^3} d\mathbf{y}_1\,S_0\,\frac{f(x_1)}{f(y_1)}$$

$$\leq \frac{\tilde{c}}{R} \int_{\mathbb{R}^3} d\mathbf{x}_1' \, \frac{1}{|\mathbf{x}_1 - \mathbf{x}_1'|^2} \, \frac{1}{x_1'} \int_{\mathbb{R}^3} d\mathbf{y}_1 \, \frac{1}{y_1} \, \frac{1}{|\mathbf{x}_1' - \mathbf{y}_1|^2} \cdot \frac{x_1^\alpha}{y_1^\alpha}. \tag{2.40}$$

With the substitutions $\mathbf{y}_1 =: x_1'\mathbf{z}$ and then $\mathbf{x}_1' =: x_1\boldsymbol{\xi}$ the two integrals separate such that (with $\mathbf{e}_x := \mathbf{x}/x$)

$$I(\mathbf{x}_1) \;\leq\; \frac{\tilde{c}}{R} \int_{\mathbb{R}^3} \frac{d\boldsymbol{\xi}}{\xi^{1+\alpha}} \, \frac{1}{|\mathbf{e}_{x_1} - \boldsymbol{\xi}|^2} \cdot \int_{\mathbb{R}^3} \frac{d\mathbf{z}}{z^{1+\alpha}} \, \frac{1}{|\mathbf{e}_{x_1'} - \mathbf{z}|^2} \;\leq\; \frac{C}{R} \tag{2.41}$$

if $0 < \alpha < 2$ [3]. In the same way it is shown that $J(\mathbf{y}_1) := \int\limits_{\mathbb{R}^3} d\mathbf{x}_1 \, S_0 \, \frac{y_1^\alpha}{x_1^\alpha} \leq \frac{C}{R}$ for $1 < \alpha < 3$. Thus $\alpha = 3/2$ assures the $1/R$-boundedness of I and J. The first contribution to (2.34) is treated along the same lines. This proves the $1/R$-boundedness of $[V_{10,m}^{(1)}, \psi_j] \frac{1}{x_1}$.

Finally the boundedness of the two operators occurring in (2.32), $\frac{1}{x_1} \frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}} x_1$ and $\frac{1}{x_1} V_{10,m}^{(1)} x_1$, has to be shown. We use the fact that for any bounded operator $\mathcal{O}$, $\frac{1}{x_1} \mathcal{O} x_1 = \frac{1}{x_1} [\mathcal{O}, x_1] + \mathcal{O}$, such that for the first operator, only the boundedness of $\frac{1}{x_1} [\frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}}, x_1]$ has to be established. We use the estimate (2.33) to write

$$|\check{k}_{\frac{1}{x_1}[\frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}}, x_1]}(\mathbf{x}_1, \mathbf{x}_1')|$$

$$= \left| \frac{1}{x_1} \, \check{k}_{\frac{\boldsymbol{\sigma}^{(1)}\mathbf{p}_1}{E_{p_1}}}(\mathbf{x}_1, \mathbf{x}_1') \cdot (x_1 - x_1') \right| \;\leq\; \frac{\tilde{c}}{x_1} \, \frac{1}{|\mathbf{x}_1 - \mathbf{x}_1'|^2} \tag{2.42}$$

and with the choice $f(x) = x^{3/2}$, (2.42) multiplied by $f(x_1)/f(x_1')$ and integrated over $d\mathbf{x}_1'$, respective multiplied by $f(x_1')/f(x_1)$ and integrated over $d\mathbf{x}_1$, is finite. This proves the desired boundedness.

Concerning the operator $\frac{1}{x_1} V_{10,m}^{(1)} x_1$ we decompose

$$\frac{1}{x_1} \, V_{10,m}^{(1)} \, x_1 \;=\; 2\pi^2 \int_0^\infty dt \, \frac{1}{x_1} \, e^{-tE_{p_1}} \, \frac{1}{x_1} \left\{ [e^{-tE_{p_1}}, x_1] + x_1 e^{-tE_{p_1}} \right\}. \tag{2.43}$$

In the second contribution the t-integral can be carried out, $\int\limits_0^\infty dt \, \frac{1}{x_1} \, e^{-2tE_{p_1}} = \frac{1}{2x_1 E_{p_1}}$ which is a bounded operator. For the first contribution, we can again use the estimate (2.36) for the Bessel function together with the estimate for the t-dependence, resulting in (2.39), such that

$$\tilde{S}_0 := \left| \int_0^\infty dt \; \check{k}_{\frac{1}{x_1} e^{-tE_{p_1}} \, \frac{1}{x_1} [e^{-tE_{p_1}}, x_1]}(\mathbf{x}_1, \mathbf{y}_1) \right|$$

$$= \left| \int_0^\infty dt \, \frac{1}{x_1} \int_{\mathbb{R}^3} d\mathbf{x}_1' \, \frac{t}{2\pi^2} \, \frac{m^2}{|\mathbf{x}_1 - \mathbf{x}_1'|^2 + t^2} \, K_2(m\sqrt{(\mathbf{x}_1 - \mathbf{x}_1')^2 + t^2}\,) \, \frac{1}{x_1'} \right. \tag{2.44}$$

$$\cdot \frac{t}{2\pi^2}\,\frac{m^2}{(\mathbf{x}_1'-\mathbf{y}_1)^2+t^2}\,K_2(m\sqrt{(\mathbf{x}_1'-\mathbf{y}_1)^2+t^2}\,)\,(y_1-x_1')\Bigg|$$

$$\leq \tilde{c}_0\,\frac{1}{x_1}\int_{\mathbb{R}^3} d\mathbf{x}_1'\;\frac{1}{|\mathbf{x}_1-\mathbf{x}_1'|^2}\;\frac{1}{x_1'}\;\frac{1}{|\mathbf{x}_1'-\mathbf{y}_1|^2}.$$

With $\alpha = 3/2$, in the same way as shown in the step from (2.40) to (2.41), one obtains $\tilde{I}(\mathbf{x}_1) := \int\limits_{\mathbb{R}^3} d\mathbf{y}_1\,\tilde{S}_0\frac{x_1^\alpha}{y_1^\alpha} \leq c$ and $\tilde{J}(\mathbf{y}_1) := \int\limits_{\mathbb{R}^3} d\mathbf{x}_1\,\tilde{S}_0\frac{y_1^\alpha}{x_1^\alpha} \leq c$. Thus the boundedness of $\frac{1}{x_1}\,V_{10,m}^{(1)}x_1$ is shown.

In the remaining contributions to $[b_{2m}^{(1)},\phi_j\chi]$ the terms not treated so far are $[\frac{m}{E_{p_1}},\phi_j\chi]\frac{1}{x_1}$, the $\frac{1}{R}$-boundedness of which follows from the estimate $|\check{k}_{\frac{m}{E_{p_1}}}(\mathbf{x}_1,\mathbf{x}_1')| \leq c/|\mathbf{x}_1-\mathbf{x}_1'|^3$ (which is proven in the same way as the corresponding Brown-Ravenhall estimate for $\tilde{g}(p_1) := \frac{m}{\sqrt{2}}\,(E_{p_1}+\sqrt{E_{p_1}^2+mE_{p_1}}\,)^{-1}$ in place of m/E_{p_1} [14]). The boundedness of the additional term $\frac{1}{x_1}\,\frac{m}{E_{p_1}}x_1$ (respective $\frac{1}{x_1}[\frac{m}{E_{p_1}},x_1]$) follows from (2.42) formulated for $\check{k}_{\frac{m}{E_{p_1}}}$.

This completes the proof of Lemma 4.

We now turn to the 'easy part' of the HVZ theorem, where we have to assure that $[\Sigma_0,\infty)\subset\sigma_{ess}(\tilde{h}^{(2)})$. We use the method of proof applied to the multi-particle Brown-Ravenhall operator (see section 1 (b)). The proof of continuity of $\sigma(T+a_j)$ for $j\in\{0,...,N\}$ with a_j from (2.5) does not depend on the choice of the single-particle potential and hence also holds true for the Jansen-Hess operator. With $T_{\mathbf{a}}$ from (1.25) for $N=2$ and $\varphi_n\in C_0^\infty(\mathbb{R}^6)\otimes\mathbb{C}^4$ a defining sequence for $\lambda\in\sigma(T+a_j)$ we have (according to (1.26)) to show that $\|r_j\,T_{\mathbf{a}}\varphi_n\| < \epsilon$ for n and a sufficiently large, where r_j now includes the terms $b_{2m}^{(k)}$ with $k\notin C_{1j}$.

(d) Lemma 5 has therefore to be supplemented with the conjecture

$$\|b_{2m}^{(k)}\,\varphi\| \;\leq\; \frac{c}{R}\,\|\varphi\| \tag{2.45}$$

where $\varphi\in C_0^\infty(\Omega)\otimes\mathbb{C}^{2l}$ with $\Omega := \{\mathbf{x}=(\mathbf{x}_1,...,\mathbf{x}_l)\in\mathbb{R}^{3l} : x_i > R\;\forall i = 1,...,l\}$, $R>1$ and $k\in\{1,...,l\}$ where l is the number of electrons in cluster C_{2j}. The domain Ω allows for the introduction of the auxiliary function χ from (1.31) which is unity on the support of φ.

As discussed in the proof of Lemma 3, $b_{2m}^{(k)}$ consists (for $k=1$) of terms like $W_1\frac{1}{x_1}B_1$ (and its Hermitean conjugate), such that the idea of (2.29) can be used,

$$|(\phi, W_1\,\frac{1}{x_1}\,B_1\,\varphi)| \;\leq\; |(W_1\phi,\frac{1}{x_1}\,\chi\,B_1\,\varphi)| \;+\; |(W_1\phi,\frac{1}{x_1}\,[\chi,B_1]\,\varphi)|. \tag{2.46}$$

Therefore, the proof of Lemma 3 establishes the validity of (2.45), too.
Thus the proof of Theorem 2 is complete. □

We remark that the two-particle potentials of $\tilde{h}^{(2)}$ coincide with those of h^{BR} and hence are nonnegative. Therefore, as demonstrated in section 1 (below (1.9)), $j = 0$ (corresponding to the cluster decomposition where the nucleus is separated from all electrons) can be omitted in the determination of Σ_0. Thus the infimum of the essential spectrum of $h^{(2)}$ is given by the first ionization threshold (i.e. the infimum of the spectrum of the operator describing an ion with one electron less) increased by the electron's rest energy m.

3 The multiparticle Jansen-Hess operator

Let

$$H_N^{(2)} := H^{BR} + \Lambda_{+,N} \, (\sum_{k=1}^N B_{2m}^{(k)} + \sum_{k>l=1}^N C^{(kl)}) \, \Lambda_{+,N} \tag{3.1}$$

$$=: \tilde{H}_N^{(2)} + \Lambda_{+,N} \sum_{k>l=1}^N C^{(kl)} \, \Lambda_{+,N}$$

with H^{BR} from (1.1) and the second-order potentials from (2.1) and (2.2).
According to section 1, the proofs of the required lemmata to assure the HVZ theorem for $\tilde{H}_N^{(2)}$ are easily generalized to the N-electron case (with the exception of Lemma 1). For Lemma 1 to hold, we have to establish the form boundedness of the total potential $\tilde{W}_0$ of $\tilde{H}_N^{(2)}$ with respect to the multiparticle kinetic energy T_0. We can prove (see Appendix A)

Lemma 7. *Let* $\tilde{H}_N^{(2)} =: T_0 + \tilde{W}_0$ *be (as defined in (3.1) with* $T_0 := \Lambda_{+,N} \sum_{k=1}^N D_0^{(k)} \Lambda_{+,N}$*) the* N*-electron Jansen-Hess operator without the second-order two-electron interaction terms, acting on* $\mathcal{A}(H_1(\mathbb{R}^3) \otimes \mathbb{C}^4)^N$. *Then* $\tilde{W}_0$ *is relatively form bounded with respect to the kinetic energy operator* T_0,

$$|(\psi, \tilde{W}_0 \psi)| \leq c_1 \, (\psi, T_0 \, \psi) + C_1 \, (\psi, \psi) \tag{3.2}$$

with $c_1 < 1$ *for* $\gamma < \gamma_{BR}$ *irrespective of the electron number* N *(for* $N \leq Z$*).*

We remark that the relative form boundedness of the total potential $W_0 := H_N^{(2)} - T_0$ holds only for a smaller critical γ. Using the estimate $|(\psi_+, \sum_{k>l=1}^N C^{(kl)} \psi_+)| \leq \gamma \frac{e^2 \pi^2}{2} \frac{N-1}{2} (\psi, T_0 \psi)$ with $\psi_+ := \Lambda_{+,N} \psi$ [13], we found $\gamma < 0.454$ $(Z \leq 62)$ for $N = Z$.
The proof of Lemma 1 for the N-electron operator $\tilde{H}_N^{(2)}$ is then done in the same way as for the Brown-Ravenhall operator in section 1 (using the estimates for the second-order single-particle interaction from section 2 (a)).
For the proof of Proposition 1 formulated for the N-electron case we note that the resolvent $R_{N,\mu} := (H_N^{(2)} + \mu)^{-1} - (\tilde{H}_N^{(2)} + \mu)^{-1}$ can be written as a finite sum of compact operators of the type (2.10). In place of W_2, we have

$W_{kl} := (T+\mu)^{-1} C^{(kl)} (T+\mu)^{-1}$ with the N-particle kinetic energy T. Since, however, $(T+\mu)^{-1} \leq (T^{(k)} + T^{(l)} + \mu)^{-1}$, the compactness proof for W_{kl} can be copied from the $N = 2$ case. In addition, we have to assure the relative operator boundedness of the total potential of $H_N^{(2)}$:

Lemma 8. *Let $H_N^{(2)} =: T_0 + W_0$ be the N-electron Jansen-Hess operator. For $\gamma < \gamma_1$ the total potential W_0 is bounded by the kinetic energy operator,*

$$\|W_0\, \psi\| \;\leq\; c_0\, \|T_0\, \psi\| \tag{3.3}$$

with $c_0 < 1$. For $N = Z$ (and $m = 0$), $\gamma_1 = 0.285$ ($Z \leq 39$).

The proof is given in Appendix B. A consequence of this proof is the relative boundedness of the total potential of $\tilde{H}_N^{(2)}$ (with bound < 1) for $\gamma < \gamma_1$. We note that the critical potential strength may well be improved by using more refined techniques for the estimate of $W_0 \psi$ in the case of large N.
Collecting results, we have shown that the HVZ theorem holds also for the N-electron Jansen-Hess operator, provided γ is below a critical potential strength ($\gamma < 0.285$ if $N = Z$).

Appendix A (Proof of Lemma 7)

When showing the relative form boundedness of the potential $\tilde{W}_0$, we can disregard the projectors $\Lambda_{+,N}$ in (1.1) and (3.1). In fact, define the potential $\tilde{W}$ by $\tilde{H}_N^{(2)} = T_0 + \tilde{W}_0 =: \Lambda_{+,N} (\sum\limits_{k=1}^{N} D_0^{(k)} + \tilde{W})\, \Lambda_{+,N}$. Assume we prove for $\psi_+ := \Lambda_{+,N}\, \psi \in \Lambda_{+,N} (\mathcal{A}(H_1(\mathbb{R}^3) \otimes \mathbb{C}^4)^N)$ an N-particle function in the positive spectral subspace and $T = E_{p_1} + ... + E_{p_N}$,

$$|(\psi_+, \tilde{W} \psi_+)| \;\leq\; c_1\, (\psi_+, T \psi_+) \;+\; C_1\, (\psi_+, \psi_+) \tag{A.1}$$

with constants $c_1 < 1$ and $C_1 \geq 0$. Then we get

$$(\psi, \tilde{W}_0 \psi) \;=\; (\psi, \Lambda_{+,N} \tilde{W} \Lambda_{+,N}\, \psi) \;=\; (\psi_+, \tilde{W}\, \psi_+). \tag{A.2}$$

Noting that $(\psi_+, T\psi_+) \;=\; (\psi_+, \sum\limits_{k=1}^{N} D_0^{(k)} \psi_+) \;=\; (\psi, T_0\, \psi)$ and $\|\Lambda_{+,N}\, \psi\| \;\leq\; \|\Lambda_{+,N}\|\, \|\psi\| \;\leq\; \|\psi\|$ because $\|\Lambda_{+,N}\| \;=\; \|\Lambda_+^{(1)}\| \cdots \|\Lambda_+^{(n)}\| \;=\; 1$, Lemma 7 is verified with the help of (A.1).
In order to show (A.1) we start by estimating from below. We use $V^{(kl)} \geq 0$, $|(\psi_+, V^{(k)} \psi_+)| \leq \frac{\gamma}{\gamma_{BR}}\, (\psi_+, E_{p_1} \psi_+)$ (for $\gamma \leq \gamma_{BR}$; [4, 13]) as well as $(\psi_+, B_{2m}^{(k)} \psi_+) \;\geq -m d_0 \gamma^2\, (\psi_+, \psi_+)$ (for $\gamma \leq 4/\pi$) with $d_0 := 8 + 12\sqrt{2}$ [3, 13]. Then

$$(\psi_+, \tilde{W} \psi_+) \;\geq\; -\,\frac{\gamma}{\gamma_{BR}} \sum_{k=1}^{N} (\psi_+, E_{p_1}\, \psi_+) \;-\; m d_0 \gamma^2 \sum_{k=1}^{N} (\psi_+, \psi_+)$$

$$= -\frac{\gamma}{\gamma_{BR}} (\psi_+, T\psi_+) - md_0 N\gamma^2 (\psi_+, \psi_+). \tag{A.3}$$

For the estimate from above we use $(\psi_+, (V^{(k)} + B_{2m}^{(k)})\psi_+) \le m(d_0\gamma^2 + \frac{3}{2}\gamma)(\psi_+, \psi_+)$ for $\gamma \le 4/\pi$ [3], [11, Lemma II.8] as well as $(\psi_+, V^{(kl)}\psi_+) \le \frac{e^2}{\gamma_{BR}}(\psi_+, E_{p_1}\psi_+)$ (for $\gamma \le \gamma_{BR}$) which is an immediate consequence of the estimate of $V^{(k)}$. Then

$$(\psi_+, \tilde{W}\psi_+) \le m(d_0\gamma^2 + \frac{3}{2}\gamma) N (\psi_+, \psi_+) + \frac{N-1}{2} \frac{e^2}{\gamma_{BR}} (\psi_+, T\psi_+) \tag{A.4}$$

such that (A.1) holds with $c_1 := \max\{\frac{\gamma}{\gamma_{BR}}, \frac{N-1}{2}\frac{e^2}{\gamma_{BR}}\}$. For $N \le Z$, one has $c_1 = \frac{\gamma}{\gamma_{BR}}$ which is smaller than one if $\gamma < \gamma_{BR}$.

Appendix B (Proof of Lemma 8)

For the proof of the relative boundedness of the total potential W_0, let $H_N^{(2)} = T_0 + W_0 =: \Lambda_{+,N}(\sum_{k=1}^N D_0^{(k)} + W)\Lambda_{+,N}$ where W denotes the total potential from (3.1). Assume that

$$\|W\psi_+\| \le c_0 \, \|\sum_{k=1}^N D_0^{(k)}\psi_+\| = c_0 \, \|T\psi_+\| \tag{B.1}$$

with $\psi_+ = \Lambda_{+,N}\,\psi$. Then

$$\|W_0\psi\| = \|\Lambda_{+,N} W \Lambda_{+,N}\psi\| \le \|\Lambda_{+,N}\| \, \|W\psi_+\| \le c_0 \, \|\sum_{k=1}^N D_0^{(k)}\psi_+\|$$

$$= c_0 \, \|\Lambda_{+,N} \sum_{k=1}^N D_0^{(k)} \Lambda_{+,N}\psi\| = c_0 \, \|T_0\psi\| \tag{B.2}$$

since $\Lambda_{+,N} = \Lambda_{+,N}^2$ commutes with $D_0^{(k)}$.
In order to verify (B.1) we set $W^{(k)} := V^{(k)} + B_{2m}^{(k)}$ and estimate

$$\|W\psi_+\| \le \|\sum_{k=1}^N W^{(k)}\psi_+\| + \|\sum_{k>l=1}^N V^{(kl)}\psi_+\| + 2\|\sum_{k>l=1}^N C_a^{(kl)}\psi_+\| + 2\|\sum_{k>l=1}^N C_b^{(kl)}\psi_+\| \tag{B.3}$$

where according to (2.2), $C_a^{(kl)} := V^{(kl)}\Lambda_-^{(l)} F_0^{(l)}$ and $C_b^{(kl)} := F_0^{(l)}\Lambda_-^{(l)} V^{(kl)} = C_a^{(kl)*}$, and the antisymmetry of ψ_+ with respect to particle exchange was used to reduce the four contributions to $C^{(kl)}$ to two.
From [11] it follows that $\|\sum_{k=1}^N W^{(k)}\psi_+\| \le \sqrt{c_w}\,\|T\psi_+\|$ with $c_w := (\frac{4}{3}\gamma + \frac{2}{9}\gamma^2)^2$ and $\|V^{(kl)}\psi_+\| \le \sqrt{c_v}\,\|E_{p_k}\psi_+\|$ with $c_v := 4e^4$. Likewise, using $\|\Lambda_-^{(l)}\| = 1$

and $\|F_0^{(l)}\| \leq \frac{\gamma}{\pi}(\frac{\pi^2}{4}-1)$ [11], one has $\|C_a^{(kl)}\psi_+\| \leq \sqrt{c_v}\,\|E_{p_k}(\Lambda_-^{(l)}F_0^{(l)}\psi_+)\| \leq \sqrt{c_v}\,\|F_0^{(l)}\|\,\|E_{p_k}\psi_+\| \leq \sqrt{\tilde{c}_s}\;\|E_{p_k}\psi_+\|$ and the same estimate for $\|C_b^{(kl)}\psi_+\|$, with $\tilde{c}_s := (\frac{\gamma}{\pi}(\frac{\pi^2}{4}-1))^2 c_v$ (for $m=0$). For the cross terms $V^{(kl)}V^{(kl')}$ $(l \neq l')$ we substitute $\mathbf{y}_l := \mathbf{x}_l - \mathbf{x}_k$ and $\mathbf{y}_{l'} := \mathbf{x}_{l'} - \mathbf{x}_k$ for $\mathbf{x}_l$ and $\mathbf{x}_{l'}$, respectively, and get

$$(\psi_+, V^{(kl)}V^{(kl')}\psi_+) \;=\; \int_{\mathbb{R}^{3N}} (\prod_{\substack{k'=1\\ k'\neq l,l'}}^{N} d\mathbf{x}_{k'})\, d\mathbf{y}_l\, d\mathbf{y}_{l'}\; \frac{e^2}{y_l}\,\overline{\psi}_+(...,\mathbf{y}_l+\mathbf{x}_k,\mathbf{y}_{l'}+\mathbf{x}_k,...)$$

$$\cdot \frac{e^2}{y_{l'}}\,\psi_+(...,\mathbf{y}_l+\mathbf{x}_k,\mathbf{y}_{l'}+\mathbf{x}_k,...). \tag{B.4}$$

Keeping for the moment $\mathbf{x}_{k'}$ fixed and using the Fourier representation with respect to $\mathbf{y}_l$ and $\mathbf{y}_{l'}$ (setting $\varphi_+(\mathbf{y}_l,\mathbf{y}_{l'}) := \psi_+(...,\mathbf{y}_l+\mathbf{x}_k,\mathbf{y}_{l'}+\mathbf{x}_k,...)$),

$$\widehat{(\frac{1}{y_{l'}}\varphi_+)}(\mathbf{p}_l,\mathbf{p}_{l'}) \;=\; \frac{1}{2\pi^2}\int_{\mathbb{R}^3} d\mathbf{p}'\;\frac{1}{|\mathbf{p}_{l'}-\mathbf{p}'|^2}\,\hat{\varphi}_+(\mathbf{p}_l,\mathbf{p}'), \tag{B.5}$$

the Lieb and Yau formula (2.16) with the convergence generating function $f(p) = p^{3/2}$ gives

$$\left|\int_{\mathbb{R}^6} d\mathbf{y}_l\, d\mathbf{y}_{l'}\;\frac{e^2}{y_l}\,\overline{\varphi}_+\;\frac{e^2}{y_{l'}}\,\varphi_+\right| \;\leq\; 4e^4\int_{\mathbb{R}^6} d\mathbf{p}_l\, d\mathbf{p}_{l'}\;|\hat{\varphi}_+(\mathbf{p}_l,\mathbf{p}_{l'})|^2\;p_l p_{l'} \tag{B.6}$$

such that, using $p \leq E_p$,
$|(\psi_+, V^{(kl)}V^{(kl')}\psi_+)| \leq c_v\,(\psi_+, p_l p_{l'}\psi_+) \leq c_v\,(\psi_+, E_{p_l}E_{p_{l'}}\psi_+)$. The same estimate holds for $V^{(kl)}V^{(k'l')}$ with distinct indices. The symmetry of ψ_+ with respect to particle exchange and $\sum\limits_{k>l=1}^{N} 1 = \frac{N(N-1)}{2}$ then leads to the result $\|\sum\limits_{k>l=1}^{N} V^{(kl)}\psi_+\|^2 \leq c_v \cdot \max\{\frac{N-1}{2}, \frac{1}{2}[\frac{N(N-1)}{2}-1]\}\;\|T\psi_+\|^2$.
The remaining contribution to (B.3) can partly be reduced to the estimate of $V^{(kl)}$. Let k,l,k',l' be distinct indices and set $\varphi_l := \Lambda_-^{(l)}F_0^{(l)}\psi_+$. Then we obtain for the cross terms of $(\sum\limits_{k>l=1}^{N} C_a^{(kl)})^2$,

$$|(C_a^{(kl)}\psi_+, C_a^{(k'l')}\psi_+)| \;=\; |(\Lambda_-^{(l)}F_0^{(l)}\psi_+, V^{(kl)}V^{(k'l')}\Lambda_-^{(l')}F_0^{(l')}\psi_+)|$$

$$= \;|(\varphi_l, V^{(kl)}V^{(k'l')}\varphi_{l'})| \;\leq\; c_v\,(\varphi_l, p_k p_{k'}\varphi_l)^{\frac{1}{2}}\,(\varphi_{l'}, p_k p_{k'}\varphi_{l'})^{\frac{1}{2}}. \tag{B.7}$$

Since $l \neq k'$, $p_{k'}$ commutes with $\Lambda_-^{(l)}F_0^{(l)}$ such that

$$(\varphi_l, p_k p_{k'}\varphi_l) \;=\; \|(p_k p_{k'})^{\frac{1}{2}}\varphi_l\|^2 \;=\; \|\Lambda_-^{(l)}F_0^{(l)}(p_k p_{k'})^{\frac{1}{2}}\varphi_+\|^2$$

$$\leq\; \|F_0^{(l)}\|^2\;(\psi_+, p_k p_{k'}\psi_+). \tag{B.8}$$

The cross terms of $(\sum\limits_{k>l=1}^{N} C_b^{(kl)})^2$ have the same estimate. In fact,

$$|(C_b^{(kl)}\psi_+, C_b^{(k'l')}\psi_+)| \;=\; |(\psi_+, V^{(kl)}\Lambda_-^{(l)}F_0^{(l)}\,F_0^{(l')}\Lambda_-^{(l')}V^{(k'l')}\psi_+)|$$

$$=\; |(\varphi_{l'}, V^{(kl)}V^{(k'l')}\varphi_l)| \;\leq\; c_v\,\|F_0^{(l)}\|^2\,(\psi_+, p_k p_{k'}\psi_+). \tag{B.9}$$

If any two indices coincide, we use for simplicity a weaker estimate, e.g.

$$|(C_b^{(kl)}\psi_+, C_b^{(k'l)}\psi_+)| \leq \|C_b^{(kl)}\psi_+\|\|C_b^{(k'l)}\psi_+\| \leq \tilde{c}_s\|E_{p_k}\psi_+\|^2 \leq \tilde{c}_s\frac{1}{N}\|T\psi_+\|^2 \tag{B.10}$$

and similarly for $C_a^{(kl)}$.

Counting terms in the sum $\sum\limits_{k>l=1}^{N} C_a^{(kl)} \sum\limits_{k'>l'=1}^{N} C_a^{(k'l')}$ we have $\frac{N(N-1)}{2}$ square terms, $\frac{1}{4}N(N-1)(N-2)(N-3)$ terms with four distinct indices and $N(N-1)(N-2)$ terms where two of the four indices agree (while the other two are distinct). For all terms of the last type, the estimate (B.10) is used whereas for the other terms we proceed as in the case of $(\sum\limits_{k>l=1}^{N} V^{(kl)})^2$. This leads to

$$\|\sum_{k>l=1}^{N} C_a^{(kl)}\psi_+\|^2 \;\leq\; \tilde{c}_s\cdot\max\{\frac{N-1}{2}, \frac{(N-2)(N-3)}{4}\}\;\|T\psi_+\|^2$$

$$+\;\tilde{c}_s\;(N-1)(N-2)\;\|T\psi_+\|^2. \tag{B.11}$$

Inserting our results into (B.3) we find $\|W\psi_+\| \leq c_0\|T\psi_+\|$ with

$$c_0 := \;\sqrt{c_w}\;+\;\sqrt{c_v\cdot\max\{\frac{N-1}{2}, \frac{1}{2}\,[\frac{N(N-1)}{2}-1]\}} \tag{B.12}$$

$$+\,4\,\sqrt{\tilde{c}_s}\,\sqrt{\max\{\frac{N-1}{2}, \frac{(N-2)(N-3)}{4}\} + (N-1)(N-2)}.$$

For $N = Z$ we get $c_0 < 1$ for $\gamma < 0.285$ which corresponds to $Z \leq 39$. For $N = 2$, we need $\gamma < 0.66\;\; (Z \leq 90)$ which slightly improves on our earlier estimate ($Z \leq 89$ [11]), obtained by using (B.10)-type estimates for all two-particle interaction cross terms.

Acknowledgment

I would like to thank A. Sobolev, H. Siedentop and S. Morozov for stimulating discussions and critical comments. Partial support by the EU network Analysis and Quantum (contract HPRN-CT-2002-00277) is gratefully acknowledged.

References

[1] Abramowitz M. and Stegun I. A.: *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* New York, Dover Publications, 1965

[2] Brown G. E. and Ravenhall D. G.: On the interaction of two electrons. *Proc. Roy. Soc. London* A208, 552-559 (1951)

[3] Brummelhuis R., Siedentop H. and Stockmeyer E.: The ground-state energy of relativistic one-electron atoms according to Jansen and Hess. *Doc. Math.* 7, 167-182 (2002)

[4] Burenkov V. I. and Evans W. D.: On the evaluation of the norm of an integral operator associated with the stability of one-electron atoms. *Proc. Roy. Soc. (Edinburgh) 128A*, 993-1005 (1998)

[5] Cycon H. L., Froese R. G., Kirsch W. and Simon B.: *Schrödinger Operators with Application to Quantum Mechanics and Global Geometry.* Text and Monographs in Physics, 1st Edition. Berlin, Springer-Verlag, 1987

[6] Douglas M. and Kroll N. M.: Quantum electrodynamical corrections to the fine structure of helium. *Ann. Phys. (N. Y.)* 82, 89-155 (1974)

[7] Evans W. D., Perry P. and Siedentop H.: The spectrum of relativistic one-electron atoms according to Bethe and Salpeter. *Commun. Math. Phys.* 178, 733-746 (1996)

[8] I. W. Herbst, Spectral theory of the operator $(p^2+m^2)^{\frac{1}{2}}-Ze^2/r$. *Commun. Math. Phys.* 53, 285-294 (1977).

[9] Hoever G. and Siedentop S.: Stability of the Brown-Ravenhall operator. *Math. Phys. Electron. J.* 5(6), 1-11 (1999)

[10] Hunziker W.: On the spectra of Schrödinger multiparticle Hamiltonians. *Helv. Phys. Acta* 39, 451-462 (1966)

[11] Jakubaßa-Amundsen D. H.: Spectral theory of the atomic Dirac operator in the no-pair formalism. Ph. D. thesis, University of Munich (2004)

[12] Jakubaßa-Amundsen D. H.: The projected single-particle Dirac operator for Coulombic potentials. *Doc. Math.* 10, 331-356 (2005)

[13] Jakubaßa-Amundsen D. H.: Pseudorelativistic operator for a two-electron ion. *Phys. Rev.* A71, 032105, 1-8 (2005)

[14] Jakubaßa-Amundsen D. H.: The HVZ theorem for a pseudo-relativistic operator. Submitted to *Ann. Henri Poincaré* (2005)

[15] Kato T.: *Perturbation Theory for Linear Operators.* Berlin, Springer-Verlag, 1980

[16] Lewis R. T., Siedentop H. and Vugalter S.: The essential spectrum of relativistic multi-particle operators. *Ann. Inst. Henri Poincaré* 67, No. 1, 1-28 (1997)

[17] Lieb E. H. and Yau H.-T.: The stability and instability of relativistic matter. *Commun. Math. Phys.* 118, 177-213 (1988)

[18] Morozov S. and Vugalter S.: Stability of atoms in the Brown-Ravenhall model. arXiv:math-ph/0511033 (2005)

[19] Murray J. D.: *Asymptotic Analysis.* Oxford, Clarendon Press, 1974, §2, 5

[20] Reed M. and Simon B.: *Analysis of Operators.* Vol. IV of Methods of Modern Mathematical Physics. New York, Academic Press, 1978

[21] Siedentop H. and Stockmeyer E.: An analytic Douglas-Kroll-Hess method. *Phys. Lett.* A341 473-478 (2005)

[22] Simon B.: Geometric methods in multiparticle quantum systems. *Commun. Math. Phys.* 55, 259-274 (1977)

[23] Sucher J.: Foundations of the relativistic theory of many-electron atoms. *Phys. Rev.* A22, 348-362 (1980)

[24] Teschl G.: Lecture Notes on *Schrödinger Operators.* Section 9.4, 2004.
http://www.mat.univie.ac.at/~gerald/ftp/index.html

[25] Tix C.: Strict positivity of a relativistic Hamiltonian due to Brown and Ravenhall. *Bull. London Math. Soc.* 30, 283-290 (1998)

[26] van Winter C.: Theory of finite systems of particles I. The Green function. *Mat. Fys. Dan. Vid. Selsk.* 2 No. 8, 1-60 (1964)

[27] Wolf A., Reiher M. and Hess B. A.: The generalized Douglas-Kroll transformation. *J. Chem. Phys.* 117, 9215-9226 (2002)

[28] Zhislin G. M.: A study of the spectrum of the Schrödinger operator for a system of several particles. *Trudy Moskov. Mat. Obsc.* 9, 81-120 (1960)

D. H. Jakubassa-Amundsen
Mathematics Institute
University of Munich
Theresienstr. 39
80333 Munich
Germany
dj@mathematik.uni-muenchen.de

Slope Filtrations Revisited

Kiran S. Kedlaya[1]

Received: July 24, 2005

Communicated by Peter Schneider

Abstract. We give a "second generation" exposition of the slope filtration theorem for modules with Frobenius action over the Robba ring, providing a number of simplifications in the arguments. Some of these are inspired by parallel work of Hartl and Pink, which points out some analogies with the formalism of stable vector bundles.

2000 Mathematics Subject Classification: 14F30.

Contents

[1]Thanks to Olivier Brinon for providing detailed feedback on [19, Chapter 3], to Francesco Baldassarri and Frans Oort for some suggestions on the exposition of this paper, to Mark Kisin for suggesting the statement of Theorem 6.3.3, and to Bruno Chiarellotto for pointing out an error in a prior version of Proposition 3.5.4. Thanks also to Baldassarri and to Pierre Berthelot for organizing two workshops on F-isocrystals and rigid cohomology in December 2004 and June 2005, which helped spark the aforementioned discussions. The author was supported by NSF grant number DMS-0400727.

1 Introduction

This paper revisits the slope filtration theorem given by the author in [19]. Its main purpose is expository: it provides a simplified and clarified presentation of the theory of slope filtrations over rings of Robba type. In the process, we generalize the theorem in a fashion useful for certain applications, such as the semistable reduction problem for overconvergent F-isocrystals [24].

In the remainder of this introduction, we briefly describe the theorem and some applications, then say a bit more about the nature and structure of this particular paper.

1.1 The slope filtration theorem

The Dieudonné-Manin classification [18], [29] describes the category of finite free modules equipped with a Frobenius action, over a complete discrete valuation ring with algebraically closed residue field, loosely analogous to the eigenspace decomposition of a vector space over an algebraically closed field equipped with a linear transformation. When the residue field is unrestricted, the classification no longer applies, but one does retrieve a canonical filtration whose successive quotients are all isotypical of different types if one applies the Dieudonné-Manin classification after enlarging the residue field.
The results of [19] give an analogous pair of assertions for finite free modules equipped with a Frobenius action over the Robba ring over a complete discretely valued field of mixed characteristic. (The Robba ring consists of those formal Laurent series over the given coefficient field converging on some open annulus of outer radius 1.) Namely, over a suitable "algebraic closure" of the Robba ring, every such module admits a decomposition into the same sort of standard pieces as in Dieudonné-Manin ([19, Theorem 4.16] and Theorem 4.5.7 herein), and the analogous canonical slope filtration descends back down to the original module ([19, Theorem 6.10] and Theorem 6.4.1 herein).

1.2 Applications

The original application of the slope filtration theorem was to the p-adic local monodromy theorem on quasi-unipotence of p-adic differential equations with Frobenius structure over the Robba ring. (The possibility of, and need for, such a theorem first arose in the work of Crew [11], [12] on the rigid cohomology of curves with coefficients, and so the theorem is commonly referred to as "Crew's conjecture".) Specifically, the slope filtration theorem reduces the pLMT to its unit-root case, established previously by Tsuzuki [35]. We note, for now in passing, that Crew's conjecture has also been proved by André [1] and by Mebkhout [30], using the Christol-Mebkhout index theory for p-adic differential equations; for more on the relative merits of these proofs, see Remark 7.2.8.
In turn, the p-adic local monodromy theorem is already known to have several applications. Many of these are in the study of rigid p-adic cohomology of varieties over fields of characteristic p: these include a full faithfulness theorem for restriction between the categories of overconvergent and convergent F-isocrystals [20], a finiteness theorem with coefficients [22], and an analogue of Deligne's "Weil II" theorem [23]. The pLMT also gives rise to a proof of Fontaine's conjecture that every de Rham representation (of the absolute Galois group of a mixed characteristic local field) is potentially semistable, via a construction of Berger [3] linking the theory of (ϕ, Γ)-modules to the theory of

p-adic differential equations.

Subsequently, other applications of the slope filtration theorem have come to light. Berger [4] has used it to give a new proof of the theorem of Colmez-Fontaine that weakly admissible (ϕ, Γ)-modules are admissible. A variant of Berger's proof has been given by Kisin [26], who goes on to give a classification of crystalline representations with nonpositive Hodge-Tate weights in terms of certain Frobenius modules; as corollaries, he obtains classification results for p-divisible groups conjectured by Breuil and Fontaine. Colmez [10] has used the slope filtration theorem to construct a category of "trianguline representations" involved in a proposed p-adic Langlands correspondence. André and di Vizio [2] have used the slope filtration theorem to prove an analogue of Crew's conjecture for q-difference equations, by establishing an analogue of Tsuzuki's theorem for such equations. (The replacement of differential equations by q-difference equations does not affect the Frobenius structure, so the slope filtration theorem applies unchanged.) We expect to see additional applications in the future.

1.3 Purpose of the paper

The purpose of this paper is to give a "second generation" exposition of the proof of the slope filtration theorem, using ideas we have learned about since [19] was written. These ideas include a close analogy between the theory of slopes of Frobenius modules and the formalism of semistable vector bundles; this analogy is visible in the work of Hartl and Pink [17], which strongly resembles our Dieudonné-Manin classification but takes place in equal characteristic $p > 0$. It is also visible in the theory of filtered (ϕ, N)-modules, used to study p-adic Galois representations; indeed, this theory is directly related to slope filtrations via the work of Berger [4] and Kisin [26].

In addition to clarifying the exposition, we have phrased the results at a level of generality that may be useful for additional applications. In particular, the results apply to Frobenius modules over what might be called "fake annuli", which occur in the context of semistable reduction for overconvergent F-isocrystals (a higher-dimensional analogue of Crew's conjecture). See [25] for an analogue of the p-adic local monodromy theorem in this setting.

1.4 Structure of the paper

We conclude this introduction with a summary of the various chapters of the paper.

In Chapter 2, we construct a number of rings similar to (but more general than) those occurring in [19, Chapters 2 and 3], and prove that a certain class of these are Bézout rings (in which every finitely generated ideal is principal).

In Chapter 3, we introduce σ-modules and some basic terminology for dealing with them. Our presentation is informed by some strongly analogous work (in equal characteristic p) of Hartl and Pink.

In Chapter 4, we give a uniform presentation of the standard Dieudonné-Manin decomposition theorem and of the variant form proved in [19, Chapter 4], again using the Hartl-Pink framework.

In Chapter 5, we recall some results mostly from [19, Chapter 5] on σ-modules over the bounded subrings of so-called analytic rings. In particular, we compare the "generic" and "special" polygons and slope filtrations.

In Chapter 6, we give a streamlined form of the arguments of [19, Chapter 6], which deduce from the Dieudonné-Manin-style classification the slope filtration theorem for σ-modules over arbitrary analytic rings.

In Chapter 7, we make some related observations. In particular, we explain how the slope filtration theorem, together with Tsuzuki's theorem on unit-root σ-modules with connection, implies Crew's conjecture. We also explain the relevance of the terms "generic" and "special" to the discussion of Chapter 5.

2 The basic rings

In this chapter, we recall and generalize the ring-theoretic setup of [19, Chapter 3].

Convention 2.0.1. Throughout this chapter, fix a prime number p and a power $q = p^a$ of p. Let K be a field of characteristic p, equipped with a valuation v_K; we will allow v_K to be trivial unless otherwise specified. Let K_0 denote a subfield of K on which v_K is trivial. We will frequently do matrix calculations; in so doing, we apply a valuation to a matrix by taking its minimum over entries, and write I_n for the $n \times n$ identity matrix over any ring. See Conventions 2.2.2 and 2.2.6 for some further notations.

2.1 Witt rings

Convention 2.1.1. Throughout this section only, assume that K and K_0 are perfect.

Definition 2.1.2. Let $W(K)$ denote the ring of p-typical Witt vectors over K. Then W gives a covariant functor from perfect fields of characteristic p to complete discrete valuation rings of characteristic 0, with maximal ideal p and perfect residue field; this functor is in fact an equivalence of categories, being a quasi-inverse of the residue field functor. In particular, the absolute (p-power) Frobenius lifts uniquely to an automorphism σ_0 of $W(K)$; write σ for the $\log_p(q)$-th power of σ_0. Use a horizontal overbar to denote the reduction map from $W(K)$ to K. In this notation, we have $\overline{u^\sigma} = \overline{u}^q$ for all $u \in W(K)$.

We will also want to allow some ramified extensions of Witt rings.

Definition 2.1.3. Let $\mathcal{O}$ be a finite totally ramified extension of $W(K_0)$, equipped with an extension of σ; let π denote a uniformizer of $\mathcal{O}$. Write $W(K, \mathcal{O})$ for $W(K) \otimes_{W(K_0)} \mathcal{O}$, and extend the notations $\sigma, \overline{x}$ to $W(K, \mathcal{O})$ in the natural fashion.

Definition 2.1.4. For $\overline{z} \in K$, let $[\overline{z}] \in W(K)$ denote the Teichmüller lift of K; it can be constructed as $\lim_{n\to\infty} y_n^{p^n}$ for any sequence $\{y_n\}_{n=0}^\infty$ with $\overline{y}_n = \overline{z}^{1/p^n}$. (The point is that this limit is well-defined: if $\{y'_n\}_{n=0}^\infty$ is another such sequence, we have $y_n^{p^n} \equiv (y'_n)^{p^n} \pmod{p^n}$.) Then $[\overline{z}]^\sigma = [\overline{z}]^q$, and if $\overline{z}' \in K$, then $[\overline{z}\overline{z}'] = [\overline{z}][\overline{z}']$. Note that each $x \in W(K, \mathcal{O})$ can be written uniquely as $\sum_{i=0}^\infty [\overline{z_i}]\pi^i$ for some $\overline{z_0}, \overline{z_1}, \dots \in K$; similarly, each $x \in W(K, \mathcal{O})[\pi^{-1}]$ can be written uniquely as $\sum_{i\in\mathbb{Z}} [\overline{z_i}]\pi^i$ for some $z_i \in K$ with $\overline{z_i} = 0$ for i sufficiently small.

Definition 2.1.5. Recall that K was assumed to be equipped with a valuation v_K. Given $n \in \mathbb{Z}$, we define the "partial valuation" v_n on $W(K, \mathcal{O})[\pi^{-1}]$ by

$$v_n\left(\sum_i [\overline{z_i}]\pi^i\right) = \min_{i\le n}\{v_K(\overline{z_i})\}; \tag{2.1.6}$$

it satisfies the properties

$$\begin{aligned}
v_n(x+y) &\ge \min\{v_n(x), v_n(y)\} \qquad (x, y \in W(K, \mathcal{O})[\pi^{-1}], n \in \mathbb{Z})\\
v_n(xy) &\ge \min_{m\in\mathbb{Z}}\{v_m(x) + v_{n-m}(y)\} \qquad (x, y \in W(K, \mathcal{O})[\pi^{-1}], n \in \mathbb{Z})\\
v_n(x^\sigma) &= q v_n(x) \qquad (x \in W(K, \mathcal{O})[\pi^{-1}], n \in \mathbb{Z})\\
v_n([\overline{z}]) &= v_K(\overline{z}) \qquad (\overline{z} \in K, n \ge 0).
\end{aligned}$$

In each of the first two inequalities, one has equality if the minimum is achieved exactly once. For $r > 0$, $n \in \mathbb{Z}$, and $x \in W(K, \mathcal{O})[\pi^{-1}]$, put

$$v_{n,r}(x) = r v_n(x) + n;$$

for $r = 0$, put $v_{n,r}(x) = n$ if $v_n(x) < \infty$ and $v_{n,r}(x) = \infty$ if $v_n(x) = \infty$. For $r \ge 0$, let $W_r(K, \mathcal{O})$ be the subring of $W(K, \mathcal{O})$ consisting of those x for which $v_{n,r}(x) \to \infty$ as $n \to \infty$; then σ sends $W_{qr}(K, \mathcal{O})$ onto $W_r(K, \mathcal{O})$. (Note that there is no restriction when $r = 0$.)

Lemma 2.1.7. *Given $x, y \in W_r(K, \mathcal{O})[\pi^{-1}]$ nonzero, let i and j be the smallest and largest integers n achieving $\min_n\{v_{n,r}(x)\}$, and let k and l be the smallest and largest integers n achieving $\min_n\{v_{n,r}(y)\}$. Then $i + k$ and $j + l$ are the smallest and largest integers n achieving $\min_n\{v_{n,r}(xy)\}$, and this minimum equals $\min_n\{v_{n,r}(x)\} + \min_n\{v_{n,r}(y)\}$.*

Proof. We have

$$v_{m,r}(xy) \ge \min_n\{v_{n,r}(x) + v_{m-n,r}(y)\},$$

with equality if the minimum on the right is achieved only once. This means that:

- for all m, the minimum is at least $v_{i,r}(x) + v_{k,r}(y)$;

- for $m = i+k$ and $m = j+l$, the value $v_{i,r}(x)+v_{k,r}(y)$ is achieved exactly once (respectively by $n = i$ and $n = j$);
- for $m < i+k$ or $m > j+l$, the value $v_{i,r}(x) + v_{k,r}(y)$ is never achieved.

This implies the desired results. □

Definition 2.1.8. Define the map $w_r : W_r(K, \mathcal{O})[\pi^{-1}] \to \mathbb{R} \cup \{\infty\}$ by

$$w_r(x) = \min_n \{v_{n,r}(x)\}; \tag{2.1.9}$$

also write w for w_0. By Lemma 2.1.7, w_r is a valuation on $W_r(K, \mathcal{O})[\pi^{-1}]$; moreover, $w_r(x) = w_{r/q}(x^\sigma)$. Put

$$W_{\mathrm{con}}(K, \mathcal{O}) = \cup_{r>0} W_r(K, \mathcal{O});$$

note that $W_{\mathrm{con}}(K, \mathcal{O})$ is a discrete valuation ring with residue field K and maximal ideal generated by π, but is not complete if v_K is nontrivial.

Remark 2.1.10. Note that u is a unit in $W_r(K, \mathcal{O})$ if and only if $v_{n,r}(u) > v_{0,r}(u)$ for $n > 0$. We will generalize this observation later in Lemma 2.4.7.

Remark 2.1.11. Note that w is a p-adic valuation on $W(K, \mathcal{O})$ normalized so that $w(\pi) = 1$. This indicates two discrepancies from choices made in [19]. First, we have normalized $w(\pi) = 1$ instead of $w(p) = 1$ for internal convenience; the normalization will not affect any of the final results. Second, we use w for the p-adic valuation instead of v_p (or simply v) because we are using v's for valuations in the "horizontal" direction, such as the valuation on K, and the partial valuations of Definition 2.1.5. By contrast, decorated w's denote "nonhorizontal" valuations, as in Definition 2.1.8.

Lemma 2.1.12. *The (noncomplete) discrete valuation ring $W_{\mathrm{con}}(K, \mathcal{O})$ is henselian.*

Proof. It suffices to verify that if $P(x)$ is a polynomial over $W_{\mathrm{con}}(K, \mathcal{O})$ and $y \in W_{\mathrm{con}}(K, \mathcal{O})$ satisfies $P(y) \equiv 0 \pmod{\pi}$ and $P'(y) \not\equiv 0 \pmod{\pi}$, then there exists $z \in W_{\mathrm{con}}(K, \mathcal{O})$ with $z \equiv y \pmod{\pi}$ and $P(z) = 0$. To see this, pick $r > 0$ such that $w_r(P(y)/P'(y)^2) > 0$; then the usual Newton iteration gives a series converging under w to a root z of P in $W(K, \mathcal{O})$ with $z \equiv y \pmod{\pi}$. However, the iteration also converges under w_r, so we must have $z \in W_r(K, \mathcal{O})$. (Compare [19, Lemma 3.9].) □

2.2 Cohen rings

Remember that Convention 2.1.1 is no longer in force, i.e., K_0 and K no longer need be perfect.

Definition 2.2.1. Let C_K denote a *Cohen ring* of K, i.e., a complete discrete valuation ring with maximal ideal pC_K and residue field K. Such a ring necessarily exists and is unique up to noncanonical isomorphism [15, Proposition 0.19.8.5]. Moreover, any map $K \to K'$ can be lifted, again noncanonically, to a map $C_K \to C_{K'}$.

Convention 2.2.2. For the remainder of the chapter, assume chosen and fixed a map (necessarily injective) $C_{K_0} \to C_K$. Let $\mathcal{O}$ be a finite totally ramified extension of C_{K_0}, and let π denote a uniformizer of $\mathcal{O}$. Write Γ^K for $C_K \otimes_{C_{K_0}} \mathcal{O}$; we write Γ for short if K is to be understood, as it will usually be in this chapter.

Definition 2.2.3. By a *Frobenius lift* on Γ, we mean any endomorphism $\sigma : \Gamma \to \Gamma$ lifting the absolute q-power Frobenius on K. Given σ, we may form the completion of the direct limit

$$\Gamma^K \xrightarrow{\sigma} \Gamma^K \xrightarrow{\sigma} \cdots ; \tag{2.2.4}$$

for $K = K_0$, this ring is a finite totally ramified extension of $\Gamma^{K_0} = \mathcal{O}$, which we denote by $\mathcal{O}^{\mathrm{perf}}$. In general, if σ is a Frobenius lift on Γ^K which maps $\mathcal{O}$ into itself, we may identify the completed direct limit of (2.2.4) with $W(K^{\mathrm{perf}}, \mathcal{O}^{\mathrm{perf}})$; we may thus use the induced embedding $\Gamma^K \hookrightarrow W(K^{\mathrm{perf}}, \mathcal{O}^{\mathrm{perf}})$ to define $v_n, v_{n,r}, w_r, w$ on Γ.

Remark 2.2.5. In [19], a Frobenius lift is assumed to be a power of a p-power Frobenius lift, but all calculations therein work in this slightly less restrictive setting.

Convention 2.2.6. For the remainder of the chapter, assume chosen and fixed a Frobenius lift σ on Γ which carries $\mathcal{O}$ into itself.

Definition 2.2.7. Define the *levelwise topology* on Γ by declaring that a sequence $\{x_l\}_{l=0}^{\infty}$ converges to zero if and only if for each n, $v_n(x_l) \to \infty$ as $l \to \infty$. This topology is coarser than the usual π-adic topology.

Definition 2.2.8. For L/K finite separable, we may view Γ^L as a finite unramified extension of Γ^K, and σ extends uniquely to Γ^L; if L/K is Galois, then $\mathrm{Gal}(L/K)$ acts on Γ^L fixing Γ^K. More generally, we say L/K is *pseudo-finite separable* if $L = M^{1/q^n}$ for some M/K finite separable and some nonnegative integer n; in this case, we define Γ^L to be a copy of Γ^M viewed as a Γ^M-algebra via σ^n. In particular, we have a unique extension of v_K to L, under which L is complete, and we have a distinguished extension of σ to Γ^L (but only because we built the choice of σ into the definition of Γ^L).

Remark 2.2.9. One can establish a rather strong functoriality for the formation of the Γ^L, as in [19, Section 2.2]. One of the simplifications introduced here is to avoid having to elaborate upon this.

Definition 2.2.10. For $r > 0$, put $\Gamma_r = \Gamma \cap W_r(K^{\mathrm{perf}}, \mathcal{O})$. We say Γ *has enough r-units* if every nonzero element of K can be lifted to a unit of Γ_r. We say Γ *has enough units* if Γ has enough r-units for some $r > 0$.

Remark 2.2.11. (a) If K is perfect, then Γ has enough r-units for any $r > 0$, because a nonzero Teichmüller element is a unit in every Γ_r.

(b) If Γ^K has enough r-units, then $\Gamma^{K^{1/q}}$ has enough qr-units, and vice versa.

Lemma 2.2.12. *Suppose that Γ^K has enough units, and let L be a pseudo-finite separable extension of K. Then Γ^L has enough units.*

Proof. It is enough to check the case when L is actually finite separable. Put $d = [L : K]$. Apply the primitive element theorem to produce $\overline{x} \in L$ which generates L over K, and apply Lemma 2.1.12 to produce $x \in \Gamma^L_{\mathrm{con}}$ lifting $\overline{x}$.
Recall that any two Banach norms on a finite dimensional vector space over a complete normed field are equivalent [32, Proposition 4.13]. In particular, if we let v_L denote the unique extension of v_K to L, then there exists a constant $a > 0$ such that whenever $\overline{y} \in L$ and $\overline{c_0}, \dots, \overline{c_{d-1}} \in K$ satisfy $\overline{y} = \sum_{i=0}^{d-1} \overline{c_i x^i}$, we have $v_L(\overline{y}) \leq \min_i \{v_K(\overline{c_i x^i})\} + a$.
Pick $r > 0$ such that Γ^K has enough r-units and x is a unit in Γ^L_r, and choose $s > 0$ such that $1 - s/r > sa$. Given $\overline{y} \in L$, lift each $\overline{c_i}$ to either zero or a unit in Γ_r, and set $y = \sum_{i=0}^{d-1} c_i x^i$. Then for all $n \geq 0$,

$$\begin{aligned} v_{n,r}(y) &\geq \min_i \{v_{n,r}(c_i x^i)\} \\ &\geq \min_i \{r v_K(\overline{c_i}) + ri v_L(\overline{x})\} \\ &\geq r v_L(\overline{y}) - ra. \end{aligned}$$

In particular, $v_{n,s}(y) > v_{0,s}(y)$ for $n > 0$, so y is a unit in Γ^L_s. Since s does not depend on y, we conclude that Γ^L has enough s-units, as desired. □

Definition 2.2.13. Suppose that Γ has enough units. Define $\Gamma_{\mathrm{con}} = \cup_{r>0} \Gamma_r = \Gamma \cap W_{\mathrm{con}}(K, \mathcal{O})$; then Γ_{con} is again a discrete valuation ring with maximal ideal generated by π. Although Γ_{con} is not complete, it is henselian thanks to Lemma 2.1.12. For L/K pseudo-finite separable, we may view Γ^L_{con} as an extension of Γ^K_{con}, which is finite unramified if L/K is finite separable.

Remark 2.2.14. Remember that v_K is allowed to be trivial, in which case the distinction between Γ and Γ_{con} collapses.

Proposition 2.2.15. *Let L be a finite separable extension of K. Then for any $x \in \Gamma^L_{\mathrm{con}}$ such that $\overline{x}$ generates L over K, we have $\Gamma^L_{\mathrm{con}} \cong \Gamma^K_{\mathrm{con}}[x]/(P(x))$, where $P(x)$ denotes the minimal polynomial of x.*

Proof. Straightforward. □

Convention 2.2.16. For L the *completed* perfect closure or algebraic closure of K, we replace the superscript L by "perf" or "alg", respectively, writing Γ^{perf} or Γ^{alg} for Γ^L and so forth. (Recall that these are obtained by embedding Γ^K into $W(K^{\mathrm{perf}}, \mathcal{O})$ via σ, and then embedding the latter into $W(L, \mathcal{O})$ via Witt vector functoriality.) Beware that this convention disagrees with a convention of [19], in which $\Gamma^{\mathrm{alg}} = W(K^{\mathrm{alg}}, \mathcal{O})$, without the completion; we will comment further on this discrepancy in Remark 2.4.13.

The next assertions are essentially [13, Proposition 8.1], only cast a bit more generally; compare also [20, Proposition 4.1].

Definition 2.2.17. By a *valuation p-basis* of K, we mean a subset $S \subset K$ such that the set U of monomials in S of degree $< p$ in each factor (and degree 0 in almost all factors) is a valuation basis of K over K^p. That is, each $\overline{x} \in K$ has a unique expression of the form $\sum_{\overline{u} \in U} \overline{c_u}\,\overline{u}$, with each $\overline{c_u} \in K^p$ and almost all zero, and one has

$$v_K(\overline{x}) = \min_{\overline{u} \in U}\{v_K(\overline{c_u}\,\overline{u})\}.$$

Example 2.2.18. For example, $K = k((t))$ admits a valuation p-basis consisting of t plus a p-basis of k over k^p. In a similar vein, if $[v(K^*) : v((K^p)^*)] = [K : K^p] < \infty$, then one can choose a valuation p-basis for K by selecting elements of K^* whose images under v generate $v(K^*)/v((K^p)^*)$. (See also the criterion of [27, Chapter 9].)

Lemma 2.2.19. *Suppose that Γ has enough units and that K admits a valuation p-basis S. Then there exists a Γ-linear map $f : \Gamma^{\mathrm{perf}} \to \Gamma$ sectioning the inclusion $\Gamma \to \Gamma^{\mathrm{perf}}$, which maps $\Gamma^{\mathrm{perf}}_{\mathrm{con}}$ to Γ_{con}.*

Proof. Choose $r > 0$ such that Γ has enough r-units, and, for each $\overline{s} \in S$, choose a unit s of Γ_r lifting $\overline{s}$. Put $U_0 = \{1\}$. For n a positive integer, let U_n be the set of products

$$\prod_{\overline{s} \in S} (s^{e_s})^{\sigma^{-n}}$$

in which each $e_s \in \{0, \dots, q^n - 1\}$, all but finitely many e_s are zero (so the product makes sense), and the e_s are not all divisible by q. Put $V_n = U_0 \cup \cdots \cup U_n$; then the reductions of V_n form a basis of $K^{q^{-n}}$ over K. We can thus write each element of $\Gamma^{\sigma^{-n}}$ uniquely as a sum $\sum_{u \in V_n} x_u u$, with each $x_u \in \Gamma$ and for any integer $m > 0$, only finitely many of the x_u nonzero modulo π^m. Define the map $f_n : \Gamma^{\sigma^{-n}} \to \Gamma$ sending $x = \sum_{u \in V_n} x_u u$ to x_1.
Note that each element of each U_n is a unit in $(\Gamma^{\sigma^{-n}})_r$. Since S is a valuation p-basis, it follows (by induction on m) that if we write $x = \sum_{u \in V_n} x_u u$, then

$$\min_{j \le m}\{v_{j,r}(x)\} = \min_{u \in V_n} \min_{j \le m}\{v_{j,r}(x_u u)\}.$$

Hence for any $r' \in (0, r]$, f_n sends $(\Gamma^{\sigma^{-n}})_{r'}$ to $\Gamma_{r'}$. That means in particular that the f_n fit together to give a function f that extends by continuity to all of Γ^{perf}, sections the map $\Gamma \to \Gamma^{\mathrm{perf}}$, and carries $\Gamma^{\mathrm{perf}}_{\mathrm{con}}$ to Γ_{con}. $\square$

Remark 2.2.20. It is not clear to us whether it should be possible to loosen the restriction that K must have a valuation p-basis, e.g., by imitating the proof strategy of Lemma 2.2.12.

Proposition 2.2.21. *Suppose that Γ has enough units and that K admits a valuation p-basis. Let $\mu : \Gamma \otimes_{\Gamma_{\mathrm{con}}} \Gamma^{\mathrm{alg}}_{\mathrm{con}} \to \Gamma^{\mathrm{alg}}$ denote the multiplication map, so that $\mu(x \otimes y) = xy$.*

(a) If $x_1, \dots, x_n \in \Gamma$ are linearly independent over Γ_{con}, and $\mu(\sum_{i=1}^n x_i \otimes y_i) = 0$, then $y_i = 0$ for $i = 1, \dots, n$.

(b) If $x_1, \dots, x_n \in \Gamma$ are linearly independent over Γ_{con}, and $\mu(\sum_{i=1}^n x_i \otimes y_i) \in \Gamma$, then $y_i \in \Gamma_{\mathrm{con}}$ for $i = 1, \dots, n$.

(c) The map μ is injective.

Proof. (a) Suppose the contrary; choose a counterexample with n minimal. We may assume without loss of generality that $w(y_1) = \min_i \{w(y_i)\}$; we may then divide through by y_1 to reduce to the case $y_1 = 1$, where we will work hereafter.

Any $g \in \mathrm{Gal}(K^{\mathrm{alg}}/K^{\mathrm{perf}})$ extends uniquely to an automorphism of Γ^{alg} over Γ^{perf}, and to an automorphism of $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$ over $\Gamma^{\mathrm{perf}}_{\mathrm{con}}$. Then

$$0 = \sum_{i=1}^n x_i y_i = \sum_{i=1}^n x_i y_i^g = \sum_{i=2}^n x_i (y_i^g - y_i);$$

by the minimality of n, we have $y_i^g = y_i$ for $i = 2, \dots, n$. Since this is true for any g, we have $y_i \in \Gamma^{\mathrm{perf}}_{\mathrm{con}}$ for each i.

Let f be the map from Lemma 2.2.19; then

$$0 = \sum x_i y_i = f\left(\sum x_i y_i\right) = \sum x_i f(y_i) = \sum x_i (y_i - f(y_i)),$$

so again $y_i = f(y_i)$ for $i = 2, \dots, n$. Hence $x_1 = -\sum_{i=2}^n x_i y_i$, contradicting the linear independence of the x_i over Γ_{con}.

(b) For g as in (a), we have $0 = \sum x_i (y_i^g - y_i)$; by (a), we have $y_i^g = y_i$ for all i and g, so $y_i \in \Gamma^{\mathrm{perf}}_{\mathrm{con}}$. Now $0 = \sum x_i (y_i - f(y_i))$, so $y_i = f(y_i) \in \Gamma_{\mathrm{con}}$.

(c) Suppose on the contrary that $\sum_{i=1}^n x_i \otimes y_i \neq 0$ but $\sum_{i=1}^n x_i y_i = 0$; choose such a counterexample with n minimal. By (a), the x_i must be linearly dependent over Γ_{con}; without loss of generality, suppose we can write $x_1 = \sum_{i=2}^n c_i x_i$ with $c_i \in \Gamma_{\mathrm{con}}$. Then $\sum_{i=1}^n x_i \otimes y_i = \sum_{i=2}^n x_i \otimes (y_i + c_i)$ is a counterexample with only $n-1$ terms, contradicting the minimality of n.

□

2.3 Relation to the Robba ring

We now recall how the constructions in the previous section relate to the usual Robba ring.

Convention 2.3.1. Throughout this section, assume that $K = k((t))$ and $K_0 = k$; we may then describe Γ as the ring of formal Laurent series $\sum_{i\in\mathbb{Z}} c_i u^i$ with each $c_i \in \mathcal{O}$, and $w(c_i) \to \infty$ as $i \to -\infty$. Suppose further that the Frobenius lift is given by $\sum c_i u^i \mapsto \sum c_i^\sigma (u^\sigma)^i$, where $u^\sigma = \sum a_i u^i$ with $\liminf_{i\to-\infty} w(a_i)/(-i) > 0$.

Definition 2.3.2. Define the naïve partial valuations v_n^{naive} on Γ by the formula

$$v_n^{\text{naive}}\left(\sum c_i u^i\right) = \min\{i : w(c_i) \leq n\}.$$

These functions satisfy some identities analogous to those in Definition 2.1.5:

$$\begin{aligned} v_n^{\text{naive}}(x+y) &\geq \min\{v_n^{\text{naive}}(x), v_n^{\text{naive}}(y)\} \qquad (x, y \in \Gamma[\pi^{-1}], n \in \mathbb{Z}) \\ v_n^{\text{naive}}(xy) &\geq \min_{m\leq n}\{v_m^{\text{naive}}(x) + v_{n-m}^{\text{naive}}(y)\} \qquad (x, y \in \Gamma[\pi^{-1}], n \in \mathbb{Z}). \end{aligned}$$

Again, equality holds in each case if the minimum on the right side is achieved exactly once. Put

$$v_{n,r}^{\text{naive}}(x) = r v_n^{\text{naive}}(x) + n.$$

For $r > 0$, let Γ_r^{naive} be the set of $x \in \Gamma$ such that $v_{n,r}^{\text{naive}}(x) \to \infty$ as $n \to \infty$. Define the map w_r^{naive} on Γ_r^{naive} by

$$w_r^{\text{naive}}(x) = \min_n\{v_{n,r}^{\text{naive}}(x)\};$$

then w_r^{naive} is a valuation on Γ_r^{naive} by the same argument as in Lemma 2.1.7. Put

$$\Gamma_{\text{con}}^{\text{naive}} = \cup_{r>0}\Gamma_r^{\text{naive}}.$$

By the hypothesis on the Frobenius lift, we can choose $r > 0$ such that u^σ/u^q is a unit in Γ_r^{naive}.

Lemma 2.3.3. *For $r > 0$ such that u^σ/u^q is a unit in Γ_r^{naive}, and $s \in (0, qr]$, we have*

$$\min_{j\leq n}\{v_{j,s}^{\text{naive}}(x)\} = \min_{j\leq n}\{v_{j,s/q}^{\text{naive}}(x^\sigma)\} \tag{2.3.4}$$

for each $n \geq 0$ and each $x \in \Gamma$.

Proof. The hypothesis ensures that (2.3.4) holds for $x = u^i$ for any $i \in \mathbb{Z}$ and any n. For general x, write $x = \sum_i c_i u^i$; then on one hand,

$$\begin{aligned} \min_{j\leq n}\{v_{j,s/q}^{\text{naive}}(x^\sigma)\} &\geq \min_{i\in\mathbb{Z}}\{\min_{j\leq n}\{v_{j,s/q}^{\text{naive}}(c_i^\sigma(u^\sigma)^i)\}\} \\ &= \min_{i\in\mathbb{Z}}\{\min_{j\leq n}\{v_{j,s}^{\text{naive}}(c_i u^i)\}\} \\ &= \min_{j\leq n}\{v_{j,s}^{\text{naive}}(x)\}. \end{aligned}$$

On the other hand, if we take the smallest j achieving the minimum on the left side of (2.3.4), then the minimum of $v_{j,s}^{\text{naive}}(c_i u^i)$ is achieved by a *unique* integer i. Hence the one inequality in the previous sequence is actually an equality. □

Lemma 2.3.5. *For $r > 0$ such that u^σ/u^q is a unit in $\Gamma_r^{\mathrm{naive}}$, and $s \in (0, qr]$, we have*

$$\min_{j \leq n}\{v_{j,s}(x)\} = \min_{j \leq n}\{v_{j,s}^{\mathrm{naive}}(x)\} \tag{2.3.6}$$

for each $n \geq 0$ and each $x \in \Gamma$. In particular, $\Gamma_s^{\mathrm{naive}} = \Gamma_s$, and $w_s(x) = w_s^{\mathrm{naive}}(x)$ for all $x \in \Gamma_s$.

Proof. Write $x = \sum_{i=0}^{\infty}[\overline{x_i}]\pi^i$ with each $\overline{x_i} \in K^{\mathrm{perf}}$. Choose an integer l such that $\overline{x_i}^{q^l} \in K$ for $i = 0, \dots, n$, and write $\overline{x_i}^{q^l} = \sum_{h \in \mathbb{Z}} \overline{c_{hi}} t^h$ with $\overline{c_{hi}} \in k$. Choose $c_{hi} \in \mathcal{O}$ lifting $\overline{c_{hi}}$, with $c_{hi} = 0$ whenever $\overline{c_{hi}} = 0$, and put $y_i = \sum_h c_{hi} u^h$.
Pick an integer $m > n$, and define

$$x' = \sum_{i=0}^{n} y_i^{q^m} (\pi^i)^{\sigma^{l+m}};$$

then $w(x' - x^{\sigma^{l+m}}) > n$. Hence for $j \leq n$, $v_j(x') = v_j(x^{\sigma^{l+m}}) = q^{l+m} v_j(x)$ and $v_j^{\mathrm{naive}}(x') = v_j^{\mathrm{naive}}(x^{\sigma^{l+m}})$.
From the way we chose the y_i, we have

$$v_j^{\mathrm{naive}}(y_i^{q^m}(\pi^i)^{\sigma^{l+m}}) = q^{l+m} v_0(\overline{x_i}) \qquad (j \geq i).$$

It follows that $v_j^{\mathrm{naive}}(x') = q^{l+m} v_j(x)$ for $j \leq n$; that is, we have $v_j^{\mathrm{naive}}(x^{\sigma^{l+m}}) = q^{l+m} v_j(x)$ for $j \leq n$. In particular, we have

$$\min_{j \leq n}\{v_{j,s}(x)\} = \min_{j \leq n}\{v_{j,s/q^{l+m}}^{\mathrm{naive}}(x^{\sigma^{l+m}})\}.$$

By Lemma 2.3.3, this yields the desired result. (Compare [19, Lemmas 3.6 and 3.7].) ☐

Corollary 2.3.7. *For $r > 0$ such that u^σ/u^q is a unit in $\Gamma_r^{\mathrm{naive}}$, Γ has enough qr-units, and $\Gamma_{\mathrm{con}} = \Gamma_{\mathrm{con}}^{\mathrm{naive}}$.*

Remark 2.3.8. The ring $\Gamma_r^{\mathrm{naive}}[\pi^{-1}]$ is the ring of bounded rigid analytic functions on the annulus $|\pi|^r \leq |u| < 1$, and the valuation w_s^{naive} is the supremum norm on the circle $|u| = |\pi|^s$. This geometric interpretation motivates the subsequent constructions, and so is worth keeping in mind; indeed, much of the treatment of analytic rings in the rest of this chapter is modeled on the treatment of rings of functions on annuli given by Lazard [28], and our results generalize some of the results in [28] (given Remark 2.3.9 below).

Remark 2.3.9. In the context of this section, the ring $\Gamma_{\mathrm{an,con}}$ is what is usually called the *Robba ring* over K. The point of view of [19], maintained here, is that the Robba ring should always be viewed as coming with the "equipment" of a Frobenius lift σ; this seems to be the most convenient angle from which to approach σ-modules. However, when discussing a statement about $\Gamma_{\mathrm{an,con}}$

that depends only on its underlying topological ring (e.g., the Bézout property, as in Theorem 2.9.6), one is free to use any Frobenius, and so it is sometimes convenient to use a "standard" Frobenius lift, under which $u^\sigma = u^q$ and $v_n^{\mathrm{naive}} = v_n$ for all n. In general, however, one cannot get away with only standard Frobenius lifts because the property of standardness is not preserved by passing to $\Gamma^L_{\mathrm{an,con}}$ for L a finite separable extension of $k((t))$.

Remark 2.3.10. It would be desirable to be able to have it possible for $\Gamma_r^{\mathrm{naive}}$ to be the ring of rigid analytic functions on an annulus over a p-adic field whose valuation is not discrete (e.g., the completed algebraic closure $\mathbb{C}_p$ of $\mathbb{Q}_p$), since the results of Lazard we are analogizing hold in that context. However, this seems rather difficult to accommodate in the formalism developed above; for instance, the v_n cannot be described in terms of Teichmüller elements, so an axiomatic characterization is probably needed. There are additional roadblocks later in the story; we will flag some of these as we go along.

Remark 2.3.11. One can carry out an analogous comparison between naïve and true partial valuations when K is the completion of $k(x_1, \dots, x_n)$ for a "monomial" valuation, in which $v(x_1), \dots, v(x_n)$ are linearly independent over $\mathbb{Q}$; this gives additional examples in which the hypothesis "Γ has enough units" can be checked, and hence additional examples in which the framework of this paper applies. See [25] for details.

2.4 Analytic rings

We now proceed roughly as in [19, Section 3.3]; however, we will postpone certain "reality checks" on the definitions until the next section.

Convention 2.4.1. Throughout this section, and for the rest of the chapter, assume that the field K is complete with respect to the valuation v_K, and that Γ^K has enough r_0-units for some fixed $r_0 > 0$. Note that the assumption that K is complete ensures that Γ_r is complete under w_r for any $r \in [0, r_0)$.

Definition 2.4.2. Let I be a subinterval of $[0, r_0)$ bounded away from r_0, i.e., $I \subseteq [0, r]$ for some $r < r_0$. Let Γ_I be the Fréchet completion of $\Gamma_{r_0}[\pi^{-1}]$ for the valuations w_s for $s \in I$; note that the functions $v_n, v_{n,s}, w_s$ extend to Γ_I by continuity, and that σ extends to a map $\sigma : \Gamma_I \to \Gamma_{q^{-1}I}$. For $I \subseteq J$ subintervals of $[0, r_0)$ bounded away from 0, we have a natural map $\Gamma_J \to \Gamma_I$; this map is injective with dense image. For $I = [0, s]$, note that $\Gamma_I = \Gamma_s[\pi^{-1}]$. For $I = (0, s]$, we write $\Gamma_{\mathrm{an},s}$ for Γ_I.

Remark 2.4.3. In the context of Section 2.3, Γ_I is the ring of rigid analytic functions on the subspace of the open unit disc defined by the condition $\log_{|\pi|} |u| \in I$; compare Remark 2.3.8.

Definition 2.4.4. For I a subinterval of $[0, r_0)$ bounded away from r_0, and for $x \in \Gamma_I$ nonzero, define the *Newton polygon* of x to be the lower convex hull of the set of points $(v_n(x), n)$, minus any segment the negative of whose slope is not in I. Define the *slopes* of x to be the negations of the slopes of the Newton polygon of x. Define the *multiplicity* of $s \in (0, r]$ as a slope of x to be the difference in vertical coordinates between the endpoints of the segment of the Newton polygon of x of slope $-s$, or 0 if no such segment exists. If x has only finitely many slopes, define the *total multiplicity* of x to be the sum of the multiplicities of all slopes of x. If x has only one slope, we say x is *pure* of that slope.

Remark 2.4.5. The analogous definition of total multiplicity for $\Gamma_r^{\mathrm{naive}}$ counts the total number of zeroes (with multiplicities) that a function has in the annulus $|\pi|^r \leq |u| < 1$.

Remark 2.4.6. Note that the multiplicity of any given slope is always finite. More generally, for any closed subinterval $I = [r', r]$ of $[0, r_0)$, the total multiplicity of any $x \in \Gamma_I$ is finite. Explicitly, the total multiplicity equals $i - j$, where i is the largest n achieving $\min_n\{v_{n,r}(x)\}$ and j is the smallest n achieving $\min_n\{v_{n,r'}(x)\}$. In particular, if $x \in \Gamma_{\mathrm{an},r}$, the slopes of x form a sequence decreasing to zero.

Lemma 2.4.7. *For $x, y \in \Gamma_I$ nonzero, the multiplicity of each $s \in I$ as a slope of xy is the sum of the multiplicities of s as a slope of x and of y. In particular, Γ_I is an integral domain.*

Proof. For $x, y \in \Gamma_r[\pi^{-1}]$, this follows at once from Lemma 2.1.7. In the general case, note that the conclusion of Lemma 2.1.7 still holds, by approximating x and y suitably well by elements of $\Gamma_r[\pi^{-1}]$. □

Definition 2.4.8. Let $\Gamma_{\mathrm{an,con}}^K$ be the union of the $\Gamma_{\mathrm{an},r}^K$ over all $r \in (0, r_0)$; this ring is an integral domain by Lemma 2.4.7. Remember that we are allowing v_K to be trivial, in which case $\Gamma_{\mathrm{an,con}} = \Gamma_{\mathrm{con}}[\pi^{-1}] = \Gamma[\pi^{-1}]$.

Example 2.4.9. In the context of Section 2.3, the ring $\Gamma_{\mathrm{an,con}}$ consists of formal Laurent series $\sum_{n\in\mathbb{Z}} c_n u^n$ with each $c_n \in \mathcal{O}[\pi^{-1}]$, $\liminf_{n\to-\infty} w(c_n)/(-n) > 0$, and $\liminf_{n\to\infty} w(c_n)/n \geq 0$. The latter is none other than the Robba ring over $\mathcal{O}[\pi^{-1}]$.

We make a few observations about finite extensions of $\Gamma_{\mathrm{an,con}}$.

Proposition 2.4.10. *Let L be a finite separable extension of K. Then the multiplication map*

$$\mu : \Gamma_{\mathrm{an,con}}^K \otimes_{\Gamma_{\mathrm{con}}^K} \Gamma_{\mathrm{con}}^L \to \Gamma_{\mathrm{an,con}}^L$$

is an isomorphism. More precisely, for any $x \in \Gamma_{\mathrm{con}}^L$ such that $\overline{x}$ generates L over K, we have $\Gamma_{\mathrm{an,con}}^L \cong \Gamma_{\mathrm{an,con}}^K[x]/(P(x))$.

Proof. For $s > 0$ sufficiently small, we have $\Gamma_s^L \cong \Gamma_s^K[x]/(P(x))$ by Lemma 2.1.12, from which the claim follows. □

Corollary 2.4.11. *Let L be a finite Galois extension of K. Then the fixed subring of* $\operatorname{Frac}\Gamma^L_{\mathrm{an,con}}$ *under the action of* $G = \operatorname{Gal}(L/K)$ *is equal to* $\operatorname{Frac}\Gamma^K_{\mathrm{an,con}}$.

Proof. By Proposition 2.4.10, the fixed subring of $\Gamma^L_{\mathrm{an,con}}$ under the action of G is equal to $\Gamma^K_{\mathrm{an,con}}$. Given $x/y \in \operatorname{Frac}\Gamma^L_{\mathrm{an,con}}$ fixed under G, put $x' = \prod_{g \in G} x^g$; since x' is G-invariant, we have $x' \in \Gamma^K_{\mathrm{an,con}}$. Put $y' = x'y/x \in \Gamma^L_{\mathrm{an,con}}$; then $x'/y' = x/y$, and both x' and x'/y' are G-invariant, so y' is as well. Thus $x/y \in \operatorname{Frac}\Gamma^K_{\mathrm{an,con}}$, as desired. □

Lemma 2.4.12. *Let I be a subinterval of $(0, r_0)$ bounded away from r_0. Then the union $\cup\Gamma^L_I$, taken over all pseudo-finite separable extensions L of K, is dense in Γ^{alg}_I.*

Proof. Let M be the algebraic closure (not completed) of K. Then $\cup\Gamma^L$ is clearly dense in Γ^M for the p-adic topology. By Remark 2.2.11 and Lemma 2.2.12, the set of pseudo-finite separable extensions L such that Γ^L has enough r_0-units is cofinal. Hence the set U of $x \in \cup\Gamma^L_{r_0}$ with $w_{r_0}(x) \geq 0$ is dense in the set V of $x \in \Gamma^M_{r_0}$ with $w_{r_0}(x) \geq 0$ for the p-adic topology. On these sets, the topology induced on U or V by any one w_s with $s \in (0, r_0)$ is coarser than the p-adic topology. Thus U is also dense in V for the Fréchet topology induced by the w_s for $s \in I$. It follows that $\cup\Gamma^L_I$ is dense in Γ^M_I; however, the condition that $0 \notin I$ ensures that $\Gamma^M_I = \Gamma^{\mathrm{alg}}_I$, so we have the desired result. □

Remark 2.4.13. Recall that in [19] (contrary to our present Convention 2.2.16), the residue field of Γ^{alg} is the algebraic closure of K, rather than the completion thereof. However, the definition of $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$ comes out the same, and our convention here makes a few statements a bit easier to make. For instance, in the notation of [19], an element x of $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$ can satisfy $v_n(x) = \infty$ for all $n < 0$ without belonging to $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$. (Thanks to Francesco Baldassarri for suggesting this change.)

2.5 Reality checks

Before proceeding further, we must make some tedious but necessary "reality checks" concerning the analytic rings. This is most easily done for K perfect, where elements of Γ_I have canonical decompositions (related to the "strong semiunit decompositions" of [19, Proposition 3.14].)

Definition 2.5.1. For K perfect, define the functions $f_n : \Gamma[\pi^{-1}] \to K$ for $n \in \mathbb{Z}$ by the formula $x = \sum_{n \in \mathbb{Z}} [f_n(x)]\pi^n$, where the brackets again denote Teichmüller lifts. Then

$$v_n(x) = \min_{m \leq n}\{v_K(f_m(x))\} \leq v_K(f_n(x)),$$

which implies that f_n extends uniquely to a continuous function $f_n : \Gamma_I \to K$ for any subinterval $I \subseteq [0, \infty)$, and that the sum $\sum_{n \in \mathbb{Z}} [f_n(x)]\pi^n$ converges to

x in Γ_I. We call this sum the *Teichmüller presentation* of x. Let x_+, x_-, x_0 be the sums of $[f_n(x)]\pi^n$ over those n for which $v_K(f_n(x))$ is positive, negative, or zero; we call the presentation $x = x_+ + x_- + x_0$ the *plus-minus-zero presentation* of x.

From the existence of Teichmüller presentations, it is obvious that for instance, if $x \in \Gamma_{\mathrm{an},r}$ satisfies $v_n(x) = \infty$ for all $n < 0$, then $x \in \Gamma_r$. In order to make such statements evident in case K is not perfect, we need an approximation of the same technique.

Definition 2.5.2. Define a *semiunit* to be an element of Γ_{r_0} which is either zero or a unit. For $I \subseteq [0, r_0)$ bounded away from r_0 and $x \in \Gamma_I$, a *semiunit presentation* of x (over Γ_I) is a convergent sum $x = \sum_{i \in \mathbb{Z}} u_i \pi^i$, in which each u_i is a semiunit.

Lemma 2.5.3. *Suppose that $u_0, u_1, \dots$ are semiunits.*

(a) For each $i \in \mathbb{Z}$ and $r \in (0, r_0)$,

$$w_r(u_i \pi^i) \geq \min_{n \leq i} \left\{ v_{n,r} \left(\sum_{j=0}^{i} u_j \pi^j \right) \right\}.$$

(b) Suppose that $\sum_{i=0}^{\infty} u_i \pi^i$ converges π-adically to some x such that for some $r \in (0, r_0)$, $v_{n,r}(x) \to \infty$ as $n \to \infty$. Then $w_r(u_i \pi^i) \to \infty$ as $i \to \infty$, so that $\sum_i u_i \pi^i$ is a semiunit presentation of x over Γ_r.

Proof. (a) The inequality is evident for $i = 0$; we prove the general claim by induction on i. If $w_r(u_i \pi^i) \geq w_r(u_j \pi^j)$ for some $j < i$, then the induction hypothesis yields the claim. Otherwise, $w_r(u_i \pi^i) < w_r(\sum_{j<i} u_j \pi^j)$, so $v_{n,r}(\sum_{j=0}^{i} u_j \pi^j) = v_{n,r}(u_i \pi^i)$, again yielding the claim.

(b) Choose $r' \in (r, r_0)$; we can then apply (a) to deduce that

$$\begin{aligned} w_{r'}(u_i \pi^i) &\geq \min_{n \leq i} \{ v_{n,r'}(x) \} \\ &= \min_{n \leq i} \{ (r'/r) v_{n,r}(x) + (1 - r'/r) n \}. \end{aligned}$$

It follows that

$$\begin{aligned} w_r(u_i \pi^i) &\geq \min_{n \leq i} \{ v_{n,r}(x) + (r/r' - 1) n \} + (1 - r/r') i \\ &= \min_{n \leq i} \{ v_{n,r}(x) + (1 - r/r')(i - n) \}. \end{aligned}$$

Since $v_{n,r}(x) \to \infty$ as $n \to \infty$, the right side tends to ∞ as $n \to \infty$. □

Lemma 2.5.4. *Given a subinterval I of $[0, r_0)$ bounded away from r_0, and $r \in I$, suppose that $x \in \Gamma_{[r,r]}$ has the property that for any $s \in I$, $v_{n,s}(x) \to \infty$ as $n \to \pm\infty$. Suppose also that $\sum_i u_i \pi^i$ is a semiunit presentation of x over $\Gamma_{[r,r]}$. Then $\sum_i u_i \pi^i$ converges in Γ_I; in particular, $x \in \Gamma_I$.*

Proof. By applying Lemma 2.5.3(a) to $\sum_{i=-N}^{N} u_i \pi^i$ and using continuity, we deduce that $w_r(u_i\pi^i) \geq \min_{n \leq i}\{v_{n,r}(x)\}$. For $s \in I$ with $s \geq r$, we have $w_s(u_i\pi^i) \geq (s/r)w_r(u_i\pi^i) + (s/r-1)(-i)$, so $w_s(u_i\pi^i) \to \infty$ as $i \to -\infty$. On the other hand, for $s \in I$ with $s < r$, we have

$$\begin{aligned}
w_s(u_i\pi^i) &= (s/r)w_r(u_i\pi^i) + (1-s/r)i \\
&\geq (s/r)\min_{n \leq i}\{v_{n,r}(x)\} + (1-s/r)i \\
&= (s/r)\min_{n \leq i}\{(r/s)v_{n,s}(x) + (1-r/s)n\} + (1-s/r)i \\
&= \min_{n \leq i}\{v_{n,s}(x) + (s/r-1)(n-i)\} \\
&\geq \min_{n \leq i}\{v_{n,s}(x)\};
\end{aligned}$$

by the hypothesis that $v_{n,s}(x) \to \infty$ as $n \to \pm\infty$, we have $w_s(u_i\pi^i) \to \infty$ as $i \to -\infty$ also in this case.

We conclude that $\sum_{i<0} u_i\pi^i$ converges in Γ_I; put $y = x - \sum_{i<0} u_i\pi^i$. Then $\sum_{i=0}^{\infty} u_i\pi^i$ converges to y under w_r, hence also π-adically. By Lemma 2.5.3(b), $\sum_{i=0}^{\infty} u_i\pi^i$ converges in Γ_r, so we have $x \in \Gamma_I$, as desired. □

One then has the following variant of [19, Proposition 3.14].

Proposition 2.5.5. *For I a subinterval of $[0, r_0)$ bounded away from r_0, every $x \in \Gamma_I$ admits a semiunit presentation.*

Proof. We first verify that for $r \in (0, r_0)$, every element of Γ_r admits a semiunit presentation. Given $x \in \Gamma_r$, we can construct a sum $\sum_i u_i\pi^i$ converging π-adically to x, in which each u_i is a semiunit. By Lemma 2.5.3(b), this sum actually converges under w_s for each $s \in [0, r]$, hence yields a semiunit presentation.

We now proceed to the general case; by Lemma 2.5.4, it is enough to treat the case $I = [r, r]$. Choose a sum $\sum_{i=0}^{\infty} x_i$ converging to x in $\Gamma_{[r,r]}$, with each $x_i \in \Gamma_r[\pi^{-1}]$. We define elements $y_{il} \in \Gamma_r[\pi^{-1}]$ for $i \in \mathbb{Z}$ and $l \geq 0$, such that for each l, there are only finitely many i with $y_{il} \neq 0$, as follows. By the vanishing condition on the y_{il}, $x_0 + \cdots + x_l - \sum_{j<l}\sum_i y_{ij}\pi^i$ belongs to $\Gamma_r[\pi^{-1}]$ and so admits a semiunit presentation $\sum_i u_i\pi^i$ by virtue of the previous paragraph. For each i with $w_r(u_i\pi^i) < w_r(x_{l+1})$ (of which there are only finitely many), put $y_{il} = u_i$; for all other i, put $y_{il} = 0$. Then

$$w_r\left(x_0 + \cdots + x_l - \sum_{j \leq l}\sum_i y_{ij}\pi^i\right) \geq w_r(x_{l+1}).$$

In particular, the doubly infinite sum $\sum_{i,l} y_{il}\pi^i$ converges to x under w_r. If we set $z_i = \sum_l y_{il}$, then the sum $\sum_i z_i\pi^i$ converges to x under w_r.
Note that whenever $y_{il} \neq 0$, $w_r(x_l) \leq w_r(y_{il}\pi^i)$ by Lemma 2.5.3, whereas $w_r(y_{il}\pi^i) < w_r(x_{l+1})$ by construction. Thus for any fixed i, the values of $w_r(y_{il}\pi^i)$, taken over all l such that $y_{il} \neq 0$, form a strictly increasing sequence. Since each such y_{il} is a unit in Γ_{r_0}, we have $w_{r_0}(y_{il}\pi^i) = (r_0/r)w_r(y_{il}\pi^i) + (1 - r_0/r)i$; hence the values of $w_{r_0}(y_{il}\pi^i)$ also form an increasing sequence. Consequently, the sum $\sum_l y_{il}$ converges in Γ_{r_0} (not just under w_r) and its limit z_i is a semiunit. Thus $\sum_i z_i\pi^i$ is a semiunit presentation of x over $\Gamma_{[r,r]}$, as desired. □

Corollary 2.5.6. *For $r \in (0, r_0)$ and $x \in \Gamma_{[r,r]}$, we have $x \in \Gamma_r$ if and only if $v_n(x) = \infty$ for all $n < 0$.*

Proof. If $x \in \Gamma_r$, then $v_n(x) = \infty$ for all $n < 0$. Conversely, suppose that $v_n(x) = \infty$ for all $n < 0$. Apply Proposition 2.5.5 to produce a semiunit presentation $x = \sum_i u_i\pi^i$. Suppose there exists $j < 0$ such that $u_j \neq 0$; pick such a j minimizing $w_r(u_j\pi^j)$. Then $v_{j,n}(x) = w_r(u_j\pi^j) \neq \infty$, contrary to assumption. Hence $u_j = 0$ for $j < 0$, and so $x = \sum_{i=0}^{\infty} u_i\pi^i \in \Gamma_r$. □

Corollary 2.5.7. *Let $I \subseteq J$ be subintervals of $[0, r_0)$ bounded away from r_0. Suppose $x \in \Gamma_I$ has the property that for each $s \in J$, $v_{n,s}(x) \to \infty$ as $n \to \pm\infty$. Then $x \in \Gamma_J$.*

Proof. Produce a semiunit presentation of x over Γ_J using Proposition 2.5.5, then apply Lemma 2.5.4. □

The numerical criterion provided by Corollary 2.5.7 in turn implies a number of results that are evident in the case of K perfect (using Teichmüller presentations).

Corollary 2.5.8. *For $K \subseteq K'$ an extension of complete fields such that Γ^K and $\Gamma^{K'}$ have enough r_0-units, and $I \subseteq J \subseteq [0, r_0)$ bounded away from r_0, we have*

$$\Gamma_I^K \cap \Gamma_J^{K'} = \Gamma_J^K.$$

Corollary 2.5.9. *Let $I = [a, b]$ and $J = [c, d]$ be subintervals of $[0, r_0)$ bounded away from r_0 with $a \leq c \leq b \leq d$. Then the intersection of Γ_I and Γ_J within $\Gamma_{I\cap J}$ is equal to $\Gamma_{I\cup J}$. Moreover, any $x \in \Gamma_{I\cap J}$ with $w_s(x) > 0$ for $s \in I \cap J$ can be written as $y + z$ with $y \in \Gamma_I$, $z \in \Gamma_J$, and*

$$\begin{aligned} w_s(y) &\geq (s/c)w_c(x) \qquad (s \in [a, c]) \\ w_s(z) &\geq (s/b)w_b(x) \qquad (s \in [b, d]) \\ \min\{w_s(y), w_s(z)\} &\geq w_s(x) \qquad (s \in [c, b]). \end{aligned}$$

Proof. The first assertion follows from Corollary 2.5.7. For the second assertion, apply Proposition 2.5.5 to obtain a semiunit presentation $x = \sum u_i\pi^i$. Put $y = \sum_{i\leq 0} u_i\pi^i$ and $z = \sum_{i>0} u_i\pi^i$; these satisfy the claimed inequalities. □

Remark 2.5.10. The notion of a semiunit presentation is similar to that of a "semiunit decomposition" as in [19], but somewhat less intricate. In any case, we will have only limited direct use for semiunit presentations; we will mostly exploit them indirectly, via their role in proving Lemma 2.5.11 below.

Lemma 2.5.11. *Let I be a closed subinterval of $[0,r]$ for some $r \in (0,r_0)$, and suppose $x \in \Gamma_I$. Then there exists $y \in \Gamma_r$ such that*

$$w_s(x-y) \geq \min_{n<0}\{v_{n,s}(x)\} \qquad (s \in I).$$

Proof. Apply Proposition 2.5.5 to produce a semiunit presentation $\sum_i u_i \pi^i$ of x. Then we can choose $m > 0$ such that $w_s(u_i\pi^i) > \min_{n<0}\{v_{n,s}(x)\}$ for $s \in I$ and $i > m$. Put $y = \sum_{i=0}^m u_i \pi^i$; then the desired inequality follows from Lemma 2.5.3(a). □

Corollary 2.5.12. *A nonzero element x of Γ_I is a unit in Γ_I if and only if it has no slopes; if $I = (0,r]$, this happens if and only if x is a unit in $\Gamma_r[\pi^{-1}]$.*

Proof. If x is a unit in Γ_I, it has no slopes by Lemma 2.4.7. Conversely, suppose that x has no slopes; then there exists a single m which minimizes $v_{m,s}(x)$ for all $s \in I$. Without loss of generality we may assume that $m = 0$; we may then apply Lemma 2.5.11 to produce $y \in \Gamma_r$ such that $w_s(x-y) \geq \min_{n<0}\{v_{n,s}(x)\}$ for all $s \in I$. Since Γ has enough r-units, we can choose a unit $z \in \Gamma_r$ such that $w(y-z) > 0$; then $w_s(1 - xz^{-1}) > 0$ for all $s \in I$. Hence the series $\sum_{i=0}^\infty (1-xz^{-1})^i$ converges in Γ_I, and its limit u satisfies $uxz^{-1} = 1$. This proves that x is a unit.
In case $I = (0,r]$, x has no slopes if and only if there is a unique m which minimizes $v_{m,s}(x)$ for all $s \in (0,r]$; this is only possible if $v_n(x) = \infty$ for $n < m$. By Corollary 2.5.6, this implies $x \in \Gamma_r[\pi^{-1}]$; by the same argument, $x^{-1} \in \Gamma_r[\pi^{-1}]$. □

2.6 Principality

In Remark 2.3.8, the annulus of which $\Gamma_r^{\mathrm{naive}}$ is the rigid of rigid analytic functions is affinoid (in the sense of Berkovich in case the endpoints are not rational) and one-dimensional, and so $\Gamma_r^{\mathrm{naive}}$ is a principal ideal domain. This can be established more generally.
Before proceeding further, we mention a useful "positioning lemma", which is analogous to but not identical with [19, Lemma 3.24].

Lemma 2.6.1. *For $r \in (0,r_0)$ and $x \in \Gamma_{[r,r]}$ nonzero, there exists a unit $u \in \Gamma_{r_0}$ and an integer i such that, if we write $y = u\pi^i x$, then:*

(a) $w_r(y) = 0$*;*

(b) $v_0(y-1) > 0$*;*

(c) $v_{n,r}(y) > 0$ *for* $n < 0$*.*

Proof. Define i to be the largest integer minimizing $v_{-i,r}(x)$. Apply Lemma 2.5.11 to find $z \in \Gamma_r$ such that $w_r(\pi^i x - z) \geq \min_{n<0}\{v_{n,r}(\pi^i x)\}$. Since Γ has enough r_0-units, we can choose a unit u of Γ_{r_0} such that $u^{-1} \equiv z \pmod{\pi}$; then u and i have the desired properties. □

Definition 2.6.2. For $x \in \Gamma_r$ nonzero, define the *height* of x as the largest n such that $w_r(x) = v_{n,r}(x)$; it can also be described as the p-adic valuation of x plus the total multiplicity of x. By convention, we say 0 has height $-\infty$.

Lemma 2.6.3 (Division algorithm). *For $r \in (0, r_0)$ and $x, y \in \Gamma_r$ with x nonzero, there exists $z \in \Gamma_r$ such that $y - z$ is divisible by x, and z has height less than that of x. Moreover, we can ensure that $w_r(z) \geq w_r(y)$.*

Proof. Let m be the height of x. Apply Proposition 2.5.5 to choose a semiunit presentation $\sum_{i=0}^{\infty} u_i \pi^i$ of x, and put $x' = x - \sum_{i=0}^{m-1} u_i \pi^i$; then $x'\pi^{-m}$ is a unit in Γ_r, and by Lemma 2.5.3,

$$w_r(x - x') \geq w_r(x) + (1 - r/r_0).$$

Define a sequence $\{y_l\}_{l=0}^{\infty}$ as follows. Put $y_0 = y$. Given y_l with $y_l - y$ divisible by x and $w_r(y_l) \geq w_r(y)$, if y_l has height less than m, we may take $z = y_l$ and be done with the proof of the lemma. So we may assume that y_l has height at least m, which means that $\min_n\{v_{n,r}(y_l)\}$ is achieved by at least one $n \geq m$. Pick $y'_l \in \Gamma_{r_0}$ with $w_r(y'_l - y_l) \geq w_r(y_l) + (1 - r/r_0)$, and apply Lemma 2.5.11 to $y'_l \pi^{-m}$ to produce $z_l \in \Gamma_{r_0}$ such that $w_{r_0}(z_l - y'_l \pi^{-m}) \geq \min_{n<0}\{v_{n,r_0}(y'_l \pi^{-m})\}$. Put

$$\begin{aligned} y_{l+1} &= y_l - z_l(\pi^m/x')x \\ &= (y_l - y'_l) + (y'_l - z_l\pi^m) + z_l\pi^m(1 - x'/x). \end{aligned}$$

By construction, we have $w_r(y_l - y'_l) \geq w_r(y_l) + (1 - r/r_0)$ and $w_r(z_l\pi^m(1 - x'/x)) \geq w_r(y_l) + (1 - r/r_0)$. Moreover, for $n \geq m$, we have

$$\begin{aligned} v_{n,r}(y'_l - z_l\pi^m) &= (r/r_0)v_{n,r_0}(y'_l - z_l\pi^m) + (1 - r/r_0)n \\ &\geq (r/r_0)w_{r_0}(y'_l - z_l\pi^m) + (1 - r/r_0)m \\ &\geq (r/r_0)\min_{j<m}\{v_{j,r_0}(y'_l)\} + (1 - r/r_0)m \\ &= \min_{j<m}\{v_{j,r}(y'_l) + (1 - r/r_0)(m - j)\} \\ &\geq \min_{j<m}\{v_{j,r}(y'_l)\} + (1 - r/r_0) \\ &\geq w_r(y'_l) + (1 - r/r_0) = w_r(y_l) + (1 - r/r_0). \end{aligned}$$

It follows that for $n \geq m$, we have $v_{n,r}(y_{l+1}) \geq w_r(y_l) + (1 - r/r_0)$. We may assume that y_{l+1} also has height at least m, in which case $w_r(y_{l+1}) \geq w_r(y_l) + (1 - r/r_0)$. Hence (unless the process stops at some finite l, in which case we already know that we win) the y_l converge to zero under w_r, and

$$y = x\sum_{l=0}^{\infty}(y_l - y_{l+1})/x \in \Gamma_r$$

is divisible by x, so we may take $z = 0$. $\square$

Remark 2.6.4. Note how we used the fact that Γ^K has enough r_0-units, not just enough r-units. Also, note that the discreteness of the valuation on K was essential to ensuring that the sequence $\{y_l\}$ converges to zero.

This division algorithm has the usual consequence.

Proposition 2.6.5. *For $r \in (0, r_0)$, Γ_r is a principal ideal domain.*

Proof. Let J be a nonzero ideal of Γ_r, and pick $x \in J$ of minimal height. Then for any $y \in J$, apply Lemma 2.6.3 to produce z of height less than x with $y - z$ divisible by x. Then $z \in J$, so we must have $z = 0$ by the minimality in the choice of x. In other words, every $y \in J$ is divisible by x, as claimed. $\square$

Remark 2.6.6. Here is one of the roadblocks mentioned in Remark 2.3.10: if $\mathcal{O}$ is not discretely valued, then it is not even a PID itself, so the analogue of Γ_r cannot be one either.

To extend Proposition 2.6.5 to more Γ_I, we use the following factorization lemma (compare [19, Lemma 3.25]). We will refine this lemma a bit later; see Lemma 2.9.1.

Lemma 2.6.7. *For $I = [r', r] \subseteq [0, r_0)$ and $x \in \Gamma_I$, there exists a unit u of Γ_I such that $ux \in \Gamma_r$, and all of the slopes of ux in $[0, r]$ belong to I.*

Proof. By applying Lemma 2.6.1, we may reduce to the case where $w_{r'}(x) = 0$, $v_0(x - 1) > 0$, and $v_{n,r'}(x) > 0$ for $n < 0$; then for $n < 0$, we must have $v_n(x) > 0$ and so $v_{n,s}(x) > 0$ for all $s \in I$. Put

$$c = \min_{s \in I}\{\min_{n \leq 0}\{v_{n,s}(x - 1)\}\} > 0.$$

Define the sequence $u_0, u_1, \dots$ of units of Γ_I as follows. First set $u_0 = 1$. Given u_l such that $\min_{n \leq 0}\{v_{n,s}(u_l x - 1)\} \geq c$ for all $s \in I$, apply Lemma 2.5.11 to produce $y_l \in \Gamma_r$ such that $w_s(y_l - u_l x) \geq \min_{n<0}\{v_{n,s}(u_l x)\}$ for all $s \in I$. We may thus take $u_{l+1} = u_l(1 - y_l + u_l x)$, because $w_s(y_l - u_l x) \geq c$; moreover, for $n < 0$,

$$\begin{aligned} v_{n,r'}(u_{l+1}x) &= v_{n,r'}(y_l - u_{l+1}x) \\ &= v_{n,r'}((y_l - u_l x)(1 - u_l x)) \\ &\geq \min_m\{v_{m,r'}(y_l - u_l x) + v_{n-m,r'}(1 - u_l x)\}. \end{aligned}$$

This last minimum is at least $\min_{n<0}\{v_{n,r'}(u_l x)\}$. Moreover, if it is ever less than $\min_{n<0}\{v_{n,r'}(u_l x)\} + c$, then the smallest value of n achieving $\min_n\{v_{n,r'}(u_{l+1}x)\}$ is strictly greater than the smallest value of n achieving $\min_n\{v_{n,r'}(u_l x)\}$ (since in that case, terms in the last minimum above with $m \leq 0$ cannot affect the minimum of the left side over all $n < 0$).

In other words, for every l, there exists $l' > l$ such that

$$\min_{n<0}\{v_{n,r'}(u_{l'}x)\} \geq \min_{n<0}\{v_{n,r'}(u_l x)\} + c.$$

Hence in case $s = r'$, we have $\min_{n<0}\{v_{n,s}(u_l x)\} \to \infty$ as $l \to \infty$; consequently, the same also holds for $s \in I$. It follows that the sequence $\{u_l\}$ converges to a limit $u \in \Gamma_I$, and that $v_{n,s}(ux) = \infty$ for $n < 0$, so that $ux \in \Gamma_r$ by Corollary 2.5.6. Moreover, by construction, $\min_n\{v_{n,r'}(ux)\}$ is achieved by $n = 0$, so all of the slopes of ux are at least r'. □

Proposition 2.6.8. *Let I be a closed subinterval of $[0, r_0)$. Then Γ_I is a principal ideal domain.*

Proof. Put $I = [r', r]$, and let J be a nonzero ideal of Γ_I. By Lemma 2.6.7, each element x of J can be written (nonuniquely) as a unit u of Γ_I times an element y of Γ_r. Let J' be the ideal of Γ_r generated by all such y; by Proposition 2.6.8, J' is principal, generated by some z. Since $J' \subseteq J \cap \Gamma_r$, we have $z \in J$; on the other hand, each $x \in J$ has the form uy with $u \in \Gamma_I$ and $y \in \Gamma_r$, and y is a multiple of z in Γ_r, so x is a multiple of z in Γ_I. Hence z generates J, as desired. □

Remark 2.6.9. Proposition 2.6.8 generalizes Lazard's [28, Corollaire de Proposition 4].

2.7 Matrix approximations and factorizations

We need a matrix approximation lemma similar to [19, Lemma 6.2]; it is in some sense a matricial analogue of Lemma 2.6.1.

Lemma 2.7.1. *Let I be a closed subinterval of $[0, r]$ for some $r \in (0, r_0)$, and let M be an invertible $n \times n$ matrix over Γ_I. Then there exists an invertible $n \times n$ matrix U over $\Gamma_r[\pi^{-1}]$ such that $w_s(MU - I_n) > 0$ for $s \in I$. Moreover, if $w_s(\det(M) - 1) > 0$, we can ensure that $\det(U) = 1$.*

Proof. By applying Lemma 2.6.1 to $\det(M)$ (and then multiplying a single row of U by the resulting unit), we can ensure that $w_s(\det(M) - 1) > 0$ for $s \in I$. With this extra hypothesis, we proceed by induction on n.
Let C_i denote the cofactor of M_{ni} in M, so that $\det(M) = \sum_{i=1}^n C_i M_{ni}$, and in fact $C_i = (M^{-1})_{in} \det(M)$. Put $\alpha_i = \det(M)^{-1} M_{ni}$, so that $\sum_{i=1}^n \alpha_i C_i = 1$. Choose $\beta_1, \dots, \beta_{n-1}, \beta'_n \in \Gamma_r[\pi^{-1}]$ such that for $s \in I$ and $i = 1, \dots, n-1$,

$$w_s(\beta_i - \alpha_i) > -w_s(C_i) \qquad w_s(\beta'_n - \alpha_n) > -w_s(C_n).$$

Note that for $c \in \mathcal{O}$ with $w(c)$ sufficiently large, $\beta_n = \beta'_n + c$ satisfies $w_s(\beta_n - \alpha_n) > -w_s(C_n)$ for $s \in I$. Moreover, by Proposition 2.6.8, we can find $\gamma \in \Gamma_r[\pi^{-1}]$ generating the ideal generated by $\beta_1, \dots, \beta_{n-1}$; then the $\beta'_n + c$ are pairwise coprime for different $c \in \mathcal{O}$, so only finitely many of them can have a nontrivial common factor with γ. In particular, for $w(c)$ sufficiently large, $\beta_1, \dots, \beta_n$ generate the unit ideal in $\Gamma_r[\pi^{-1}]$.

With $\beta_1, \dots, \beta_n$ so chosen, we can choose a matrix A over $\Gamma_r[\pi^{-1}]$ of determinant 1 such that $A_{ni} = \beta_i$ for $i = 1, \dots, n$ (because $\Gamma_r[\pi^{-1}]$ is a PID, again by Proposition 2.6.8). Put $M' = MA^{-1}$, and let C'_n be the cofactor of M'_{nn} in M'. Then

$$\begin{aligned} C'_n &= (AM^{-1})_{nn} \det(M) \\ &= \sum_{i=1}^n A_{ni}(M^{-1})_{in} \det(M) = \sum_{i=1}^n \beta_i C_i, \end{aligned}$$

so that

$$C'_n = 1 + \sum_{i=1}^n (\beta_i - \alpha_i) C_i$$

and so $w_s(C'_n - 1) > 0$ for $s \in I$. In particular, C'_n is a unit in Γ_I.

Apply the induction hypothesis to the upper left $(n-1) \times (n-1)$ submatrix of M', and extend the resulting $(n-1) \times (n-1)$ matrix V to an $n \times n$ matrix by setting $V_{ni} = V_{in} = 0$ for $i = 1, \dots, n-1$ and $V_{nn} = 1$. Then we have $\det(M'V) = \det(M)$, so $w_s(\det(M'V) - 1) > 0$ for $s \in I$, and

$$w_s((M'V - I_n)_{ij}) > 0 \qquad (i = 1, \dots, n-1;\ j = 1, \dots, n-1;\ s \in I).$$

We now perform an "approximate Gaussian elimination" over Γ_I to transform $M'V$ into a new matrix N with $w_s(N - I_n) > 0$ for $s \in I$. First, define a sequence of matrices $\{X^{(h)}\}_{h=0}^\infty$ by $X^{(0)} = M'V$ and

$$X_{ij}^{(h+1)} = \begin{cases} X_{ij}^{(h)} & i < n \\ X_{nj}^{(h)} - \sum_{m=1}^{n-1} X_{nm}^{(h)} X_{mj}^{(h)} & i = n; \end{cases}$$

note that $X^{(h+1)}$ is obtained from $X^{(h)}$ by subtracting $X_{nm}^{(h)}$ times the m-th row from the n-th row for $m = 1, \dots, n-1$ in succession. At each step, for each $s \in I$, $\min_{1 \leq j \leq n-1}\{w_s(X_{nj}^{(h)})\}$ increases by at least $\min_{1 \leq i,j \leq n-1}\{w_s((M'V - I_n)_{ij})\}$; the latter is bounded away from zero over all $s \in I$, because I is closed and $w_s(x)$ is a continuous function of s. Thus for h sufficiently large, we have

$$w_s(X_{nj}^{(h)}) > \max\{0, \max_{1 \leq i \leq n-1}\{-w_s(X_{in}^{(h)})\}\} \qquad (s \in I;\ j = 1, \dots, n-1).$$

Pick such an h and set $X = X^{(h)}$; note that $\det(X) = \det(M'V)$, so $w_s(\det(X) - 1) > 0$ for $s \in I$. For $s \in I$,

$$\begin{aligned} w_s((X - I_n)_{ij}) &> 0 \qquad (i = 1, \dots, n;\ j = 1, \dots, n-1) \\ w_s(X_{in} X_{nj}) &> 0 \qquad (i = 1, \dots, n-1;\ j = 1, \dots, n-1) \end{aligned}$$

and hence also $w_s(X_{nn} - 1) > 0$.

Next, we perform "approximate backsubstitution". Define a sequence of matrices $\{W^{(h)}\}_{h=0}^{\infty}$ by setting $W^{(0)} = X$ and

$$W_{ij}^{(h+1)} = \begin{cases} W_{ij}^{(h)} - W_{in}^{(h)} W_{nj}^{(h)} & i < n \\ W_{ij}^{(h)} & i = n; \end{cases}$$

note that $W^{(h+1)}$ is obtained from $W^{(h)}$ by subtracting $W_{in}^{(h)}$ times the n-th row from the i-th row for $i = 1, \dots, n-1$. At each step, for $s \in I$, $w_s(W_{in}^{(h)})$ increases by at least $w_s(X_{nn} - 1)$; again, the latter is bounded away from zero over all $s \in I$ because I is closed and $w_s(x)$ is continuous in s. Thus for h sufficiently large,

$$w_s(W_{in}^{(h)}) > 0 \qquad (s \in I;\ 1 \leq i \leq n-1).$$

Pick such an h and set $W = W_h$; then $w_s(W - I_n) > 0$ for $s \in I$. (Note that the inequality $w_s(X_{in} X_{nj}) > 0$ for $i = 1, \dots, n-1$ and $j = 1, \dots, n-1$ ensures that the second set of row operations does not disturb the fact that $w_s(W_{ij}^{(h)}) > 0$ for $i = 1, \dots, n-1$ and $j = 1, \dots, n-1$.)
To conclude, note that by construction, $(M'V)^{-1}W$ is a product of elementary matrices over Γ_I, each consisting of the diagonal matrix plus one off-diagonal entry. By suitably approximating the off-diagonal entry of each matrix in the product by an element of Γ_r, we get an invertible matrix Y over Γ_r such that $w_s(M'VY - I_n) > 0$ for $s \in I$. We may thus take $U = A^{-1}VY$ to obtain the desired result. □

We also need a factorization lemma in the manner of [19, Lemma 6.4].

Lemma 2.7.2. *Let $I = [a, b]$ and $J = [c, d]$ be subintervals of $[0, r_0)$ bounded away from r_0, with $a \leq c \leq b \leq d$, and let M be an $n \times n$ matrix over $\Gamma_{I \cap J}$ with $w_s(M - I_n) > 0$ for $s \in I \cap J$. Then there exist invertible $n \times n$ matrices U over Γ_I and V over Γ_J such that $M = UV$.*

Proof. We construct sequences of matrices U_l and V_l over Γ_I and Γ_J, respectively, with

$$\begin{aligned} w_s(U_l - I_n) &\geq (s/c) w_c(M - I_n) \qquad (s \in [a, c]) \\ w_s(V_l - I_n) &\geq (s/b) w_b(M - I_n) \qquad (s \in [b, d]) \\ \min\{w_s(U_l - I_n), w_s(V_l - I_n)\} &\geq w_s(M - I_n) \qquad (s \in [c, b]) \\ w_s(U_l^{-1} M V_l^{-1} - I_n) &\geq 2^l w_s(M - I_n) \qquad (s \in [c, b]), \end{aligned}$$

as follows. Start with $U_0 = V_0 = I_n$. Given U_l, V_l, put $M_l = U_l^{-1} M V_l^{-1}$. Apply Corollary 2.5.9 to split $M_l - I_n = Y_l + Z_l$ with Y_l defined over Γ_I, Z_l defined over Γ_J, and

$$\begin{aligned} w_s(Y_l) &\geq (s/c) w_c(M_l - I_n) \geq (s/c) w_c(M - I_n) \qquad (s \in [a, c]) \\ w_s(Z_l) &\geq (s/b) w_b(M_l - I_n) \geq (s/b) w_b(M - I_n) \qquad (s \in [b, d]) \\ \min\{w_s(Y_l), w_s(Z_l)\} &\geq w_s(M_l - I_n) \geq 2^l w_s(M - I_n) \qquad (s \in [c, b]). \end{aligned}$$

Put $U_{l+1} = U_l(I+Y_l)$ and $V_{l+1} = (I+Z_l)V_l$; then one calculates that $w_s(M_{l+1} - I_n) \geq 2^{l+1} w_s(M - I_n)$ for $s \in [c, b]$.
We deduce that the sequences $\{U_l\}$ and $\{V_l\}$ each converge under w_s for $s \in [c, b]$, and the limits U and V satisfy $\min\{w_s(U - I_n), w_s(V - I_n)\} \geq w_s(M - I_n)$ for $s \in [c, b]$, and $M = UV$. However, the subset $x \in \Gamma_I$ on which

$$w_s(x) \geq \begin{cases} (s/c)w_c(M - I_n) & s \in [a, c] \\ w_s(M - I_n) & s \in [c, b] \end{cases}$$

is complete under any one w_s, so U has entries in Γ_I and $w_s(U - I_n) \geq (s/c)w_c(M - I_n)$ for $s \in [a, c]$. Similarly, V has entries in Γ_J and $w_s(V - I_n) \geq (s/b)w_b(M - I_n)$ for $s \in [b, d]$. In particular, U and V are invertible over Γ_I and Γ_J, and $M = UV$, yielding the desired factorization. □

2.8 Vector bundles

Over an open rigid analytic annulus, one specifies a vector bundle by specifying a vector bundle (necessarily freely generated by global sections) on each closed subannulus and providing glueing data; if the field of coefficients is spherically complete, it can be shown that the result is again freely generated by global sections. Here we generalize the discretely valued case of this result to analytic rings. (For rank 1, the annulus statement can be extracted from results of [28]; the general case can be found in [21, Theorem 3.4.3]. In any case, it follows from our Theorem 2.8.4 below.)

Definition 2.8.1. Let I be a subinterval of $[0, r_0)$ bounded away from r_0, and let S be a collection of closed subintervals of I closed under finite intersections, whose union is all of I. Define an *S-vector bundle* over Γ_I to be a collection consisting of one finite free Γ_J-module M_J for each $J \in S$, plus isomorphisms

$$\iota_{J_1,J_2} : M_{J_1} \otimes_{\Gamma_{J_1}} \Gamma_{J_2} \cong M_{J_2}$$

whenever $J_2 \subseteq J_1$, satisfying the compatibility condition $\iota_{J_2,J_3} \circ \iota_{J_1,J_2} = \iota_{J_1,J_3}$ whenever $J_3 \subseteq J_2 \subseteq J_1$. These may be viewed as forming a category in which a morphism between the collections $\{M_J\}$ and $\{N_J\}$ consists of a collection of morphisms $M_J \to N_J$ of Γ_J-modules which commute with the isomorphisms ι_{J_1,J_2}.

This definition obeys the analogue of the usual glueing property for coherent sheaves on an affinoid space (i.e., the theorem of Kiehl-Tate).

Lemma 2.8.2. *Let I be a subinterval of $[0, r_0)$ bounded away from r_0, and let $S_1 \subseteq S_2$ be two collections of closed subintervals of I as in Definition 2.8.1. Then the natural functor from the category of S_2-vector bundles over Γ_I to S_1-vector bundles over Γ_I is an equivalence.*

Proof. We define a quasi-inverse functor as follows. Given $J \in S_2$, by compactness we can choose $J_1, \dots, J_m \in S_1$ with $J \subseteq J' = J_1 \cup \cdots \cup J_m$; it is enough to consider the case where $m = 2$ and $J_1 \cap J_2 \neq \emptyset$, as we can repeat the construction to treat the general case.

Define $M_{J'}$ to be the $\Gamma_{J'}$-submodule of $M_{J_1} \oplus M_{J_2}$ consisting of those pairs $(\mathbf{v}_1, \mathbf{v}_2)$ such that

$$\iota_{J_1, J_1 \cap J_2}(\mathbf{v}_1) = \iota_{J_2, J_1 \cap J_2}(\mathbf{v}_2).$$

Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be a basis of M_{J_1} and let $\mathbf{w}_1, \dots, \mathbf{w}_n$ be a basis of M_{J_2}. Then there is an invertible $n \times n$ matrix A over $M_{J_1 \cap J_2}$ given by $\mathbf{w}_j = \sum_i A_{ij} \mathbf{v}_i$. By Lemma 2.7.2, A can be factored as UV, where U is invertible over Γ_{J_1} and V is invertible over Γ_{J_2}. Set

$$\mathbf{e}_j = \left(\sum_i U_{ij} \mathbf{v}_i, \sum_i (V^{-1})_{ij} \mathbf{w}_i \right);$$

then $\mathbf{e}_1, \dots, \mathbf{e}_n$ form a basis of $M_{J'}$, since the first components form a basis of M_{J_1}, the second components form a basis of M_{J_2}, and the intersection of Γ_{J_1} and Γ_{J_2} within $\Gamma_{J_1 \cap J_2}$ equals $\Gamma_{J_1 \cup J_2}$ (by Corollary 2.5.9). In particular, the natural maps $M_{J'} \otimes_{\Gamma_{J'}} \Gamma_{J_i} \to M_{J_i}$ for $i = 1, 2$ are isomorphisms. We may thus set $M_J = M_{J'} \otimes_{\Gamma_{J'}} \Gamma_J$. □

Definition 2.8.3. By Lemma 2.8.2, the category of S-vector bundles over Γ_I is canonically independent of the choice of S. We thus refer to its elements simply as *vector bundles over* Γ_I.

It follows that for I closed, any vector bundle over Γ_I is represented by a free module; a key result for us is that one has a similar result over $\Gamma_{\mathrm{an},r}$.

Theorem 2.8.4. *For $r \in (0, r_0)$, the natural functor from finite free $\Gamma_{\mathrm{an},r}$-modules to vector bundles over $\Gamma_{\mathrm{an},r} = \Gamma_{(0,r]}$ is an equivalence.*

Proof. To produce a quasi-inverse functor, let $J_1 \subseteq J_2 \subseteq \cdots$ be an increasing sequence of closed intervals with right endpoints r, whose union is $(0, r]$; for ease of notation, write Γ_i for Γ_{J_i}. We can specify a vector bundle over Γ_i by specifying a finite free Γ_i-module E_i for each i, plus identifications $E_{i+1} \otimes_{\Gamma_{i+1}} \Gamma_i \cong E_i$.

Choose a basis $\mathbf{v}_{1,1}, \dots, \mathbf{v}_{1,n}$ of E_1. Given a basis $\mathbf{v}_{i,1}, \dots, \mathbf{v}_{i,n}$ of E_i, we choose a basis $\mathbf{v}_{i+1,1}, \dots, \mathbf{v}_{i+1,n}$ of E_{i+1} as follows. Pick any basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of E_{i+1}, and define an invertible $n \times n$ matrix M_i over Γ_i by writing $\mathbf{e}_l = \sum_j (M_i)_{jl} \mathbf{v}_{i,j}$. Apply Lemma 2.7.1 to produce an invertible $n \times n$ matrix U_i over Γ_r such that $w_s(M_i U_i - I_n) > 0$ for $s \in J_i$. Apply Lemma 2.5.11 to produce an $n \times n$ matrix V_i over Γ_r with $w_s(M_i U_i - I_n - V_i) \geq \min_{m<0}\{v_{m,s}(M_i U_i - I_n)\}$ for $s \in J_i$; then $w_r(V_i) > 0$, so $I_n + V_i$ is invertible over Γ_r. Put $W_i = M_i U_i (I_n + V_i)^{-1}$, and define $\mathbf{v}_{i+1,1}, \dots, \mathbf{v}_{i+1,n}$ by $\mathbf{v}_{i+1,l} = \sum_j (W_i)_{jl} \mathbf{v}_{i,j}$; these form another basis of E_{i+1} because we changed basis over Γ_r.

If we write $J_i = [r_i, r]$, then for any fixed $s \in (0, r]$, we have

$$\begin{aligned} w_s(W_i - I_n) &= w_s((M_i U_i - I_n - V_i)(I_n + V_i)^{-1}) \\ &\geq \min_{m<0}\{v_{m,s}(M_i U_i - I_n)\} \\ &= \min_{m<0}\{(s/r_i) v_{m,r_i}(M_i U_i - I_n) + (1 - s/r_i)m\} \\ &\geq \min_{m<0}\{v_{m,r_i}(M_i U_i - I_n)\} + (s/r_i - 1) \\ &> (s/r_i - 1), \end{aligned}$$

which tends to ∞ as $i \to \infty$. Thus the product $W_1 W_2 \cdots$ converges to an invertible matrix W over $\Gamma_{\mathrm{an},r}$, and the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of E_1 defined by

$$\mathbf{e}_l = \sum_j W_{jl} \mathbf{v}_{1,j}$$

actually forms a basis of each E_i. Hence the original vector bundle can be reconstructed from the free $\Gamma_{\mathrm{an},r}$-module generated by $\mathbf{e}_1, \dots, \mathbf{e}_n$; this yields the desired quasi-inverse. □

Corollary 2.8.5. *For $r \in (0, r_0)$, let M be a finite free $\Gamma_{\mathrm{an},r}$-module. Then every closed submodule of M is free; in particular, every closed ideal of $\Gamma_{\mathrm{an},r}$ is principal.*

Proof. A submodule is closed if and only if it gives rise to a sub-vector bundle of the vector bundle associated to M; thus the claim follows from Theorem 2.8.4. □

Remark 2.8.6. One might expect that more generally every vector bundle over Γ_I is represented by a finite free Γ_I-module; we did not verify this.

2.9 The Bézout property

One pleasant consequence of Theorem 2.8.4 is the fact that the ring $\Gamma_{\mathrm{an},r}$ has the Bézout property, as we verify in this section. We start by refining the conclusion of Lemma 2.6.7 (again, compare [19, Lemma 3.25]).

Lemma 2.9.1. *For $r, s, s' \in (0, r_0)$ with $s' < s < r$, and $f \in \Gamma_{[s',r]}$, there exists $g \in \Gamma_r[\pi^{-1}]$ with the following properties.*

(a) The ideals generated by f and g in $\Gamma_{[s,r]}$ coincide.

(b) The slopes of g in $[0, r]$ are all contained in $[s, r]$.

Moreover, any such g also has the following property.

(c) f is divisible by g in $\Gamma_{[s',r]}$.

Proof. By Lemma 2.6.7, we can find a unit u of $\Gamma_{[s',r]}$ such that $uf \in \Gamma_r[\pi^{-1}]$ and the slopes of uf in $[0,r]$ are all contained in $[s,r]$. We may thus take $g = uf$ to obtain at least one $g \in \Gamma_r[\pi^{-1}]$ satisfying (a) and (b); hereafter, we let g be any element of $\Gamma_r[\pi^{-1}]$ satisfying (a) and (b). Then the multiplicity of each element of $[s,r]$ as a slope of g is equal to its multiplicity as a slope of f. Since $\Gamma_r[\pi^{-1}]$ is a PID by Proposition 2.6.8, we can find an element $h \in \Gamma_r[\pi^{-1}]$ generating the ideal generated by uf and g in $\Gamma_r[\pi^{-1}]$; in particular, the multiplicity of each element of $[s,r]$ as a slope of h is less than or equal to its multiplicity as a slope of g. However, h must also generate the ideal generated by f and g in $\Gamma_{[s,r]}$, which is generated already by f alone; in particular, the multiplicity of each element of $[s,r]$ as a slope of f is equal to its multiplicity as a slope of h.
We conclude that each element of $[s,r]$ occurs as a slope of f, g, h all with the same multiplicity. Since g only has slopes in $[s,r]$, g/h must be a unit in $\Gamma_r[\pi^{-1}]$; hence uf is already divisible by g in $\Gamma_r[\pi^{-1}]$, so f is divisible by g in $\Gamma_{[s',r]}$ as desired. □

Lemma 2.9.2. *Given $r \in (0, r_0)$ and $x \in \Gamma_r[\pi^{-1}]$ with greatest slope $s_0 < r$, choose $r' \in (s_0, r)$. Then for any $y \in \Gamma_r[\pi^{-1}]$ and any $c > 0$, there exists $z \in \Gamma_r[\pi^{-1}]$ with $y - z$ divisible by x in $\Gamma_{\mathrm{an},r}$, such that $w_s(z) > c$ for $s \in [r', r]$.*

Proof. As in the proof of Lemma 2.6.1, we can find a unit $u \in \Gamma_r$ and an integer i such that $\min_n\{v_{n,r}(ux\pi^i)\}$ is achieved by $n = 0$ but not by any $n > 0$, and that $v_0(ux\pi^i - 1) > 0$. Since $s_0 < r$, in fact $\min_n\{v_{n,r}(ux\pi^i)\}$ is only achieved by $n = 0$, so $w_r(ux\pi^i - 1) > 0$. Similarly, $w_s(ux\pi^i - 1) > 0$ for $s \in [r', r]$; since $[r', r]$ is a closed interval, we can choose $d > 0$ such that $w_s(ux\pi^i - 1) \geq d$ for $s \in [r', r]$. Now simply take

$$z = y(1 - ux\pi^i)^N$$

for some integer N with $Nd + w_s(y) > c$ for $s \in [r', r]$. □

We next introduce a "principal parts lemma" (compare [19, Lemma 3.31]).

Lemma 2.9.3. *For $r \in (0, r_0)$, let $I_1 \subset I_2 \subset \cdots$ be an increasing sequence of closed subintervals of $(0, r]$ with right endpoints r, whose union is all of $(0, r]$, and put $\Gamma_i = \Gamma_{I_i}$. Given $f \in \Gamma_{\mathrm{an},r}$ and $g_i \in \Gamma_i$ such that for each i, $g_{i+1} - g_i$ is divisible by f in Γ_i, there exists $g \in \Gamma_{\mathrm{an},r}$ such that for each i, $g - g_i$ is divisible by f in Γ_i.*

Proof. Apply Lemma 2.9.1 to produce $f_i \in \Gamma_r$ dividing f in $\Gamma_{\mathrm{an},r}$, such that f and f_i generate the same ideal in Γ_i, f_i only has slopes in I_i, and f/f_i has no slopes in I_i; put $f_0 = 1$. By Lemma 2.9.1 again (with s' varying), f_i is divisible by f_{i-1} in $\Gamma_{\mathrm{an},r}$, hence also in $\Gamma_r[\pi^{-1}]$; put $h_i = f_i/f_{i-1} \in \Gamma_r[\pi^{-1}]$ and $h_0 = 1$. Set $x_0 = 0$. Given $x_i \in \Gamma_r[\pi^{-1}]$ with $x_i - g_i$ divisible by f_i in Γ_i, note that the ideal generated by h_{i+1} and f_i in $\Gamma_r[\pi^{-1}]$ is principal by Proposition 2.6.8. Moreover, any generator has no slopes by Lemma 2.4.7 and so must be a unit in $\Gamma_r[\pi^{-1}]$ by Corollary 2.5.12. That is, we can find $a_{i+1}, b_{i+1} \in \Gamma_r[\pi^{-1}]$ with $a_{i+1}h_{i+1} + b_{i+1}f_i = 1$. Moreover, by applying Lemma 2.9.2, we may choose a_{i+1}, b_{i+1} with $w_s(b_{i+1}(g_{i+1} - x_i)f_i) \geq i$ for $s \in I_i$. (More precisely, apply Lemma 2.9.2 with the roles of x and y therein played by h_{i+1} and b_{i+1}, respectively; this is valid because h_{i+1} has greatest slope less than any element of I_i.)

Now put $x_{i+1} = x_i + b_{i+1}(g_{i+1} - x_i)f_i$; then $x_{i+1} - g_i$ is divisible by f_i in Γ_i, as then is $x_{i+1} - g_{i+1}$ since $g_{i+1} - g_i$ is divisible by f_i in Γ_i. By Lemma 2.9.1, $x_{i+1} - g_{i+1}$ is divisible by f_i also in Γ_{i+1}. Since $x_{i+1} - g_{i+1}$ is also divisible by h_{i+1} in Γ_{i+1}, $x_{i+1} - g_{i+1}$ is divisible by f_{i+1} in Γ_{i+1}.

For any given s, we have $w_s(x_{i+1} - x_i) \geq i$ for i large, so the x_i converge to a limit g in $\Gamma_{\mathrm{an},r}$. Since $g - g_i$ is divisible by f_i in Γ_i, it is also divisible by f in Γ_i. This yields the desired result. $\square$

Remark 2.9.4. The use of the f_i in the proof of Lemma 2.9.3 is analogous to the use of "slope factorizations" in [19]. Slope factorizations (convergent products of pure elements converging to a specified element of $\Gamma_{\mathrm{an},r}$), which are inspired by the comparable construction in [28], will not be used explicitly here; see [19, Lemma 3.26] for their construction.

We are finally ready to analyze the Bézout property.

Definition 2.9.5. A *Bézout ring/domain* is a ring/domain in which every finitely generated ideal is principal. Such rings look like principal ideal rings from the point of view of finitely generated modules. For instance:

- Every n-tuple of elements of a Bézout domain which generate the unit ideal is unimodular, i.e., it occurs as the first row of a matrix of determinant 1 [19, Lemma 2.3]. (Beware that [19, Lemma 2.3] is stated for a Bézout ring but is only valid for a Bézout domain.)
- The saturated span of any subset of a finite free module over a Bézout domain is a direct summand [19, Lemma 2.4].
- Every finitely generated locally free module over a Bézout domain is free [19, Proposition 2.5], as is every finitely presented torsion-free module [12, Proposition 4.9].
- Any finitely generated submodule of a finite free module over a Bézout ring is free (straightforward).

Theorem 2.9.6. *For $r \in (0, r_0)$, the ring $\Gamma_{\mathrm{an},r}$ is a Bézout domain (as then is $\Gamma_{\mathrm{an,con}}$). More precisely, if J is an ideal of $\Gamma_{\mathrm{an},r}$, the following are equivalent.*

(a) The ideal J is closed.

(b) The ideal J is finitely generated.

(c) The ideal J is principal.

Proof. Clearly (c) implies both (a) and (b). Also, (a) implies (c) by Theorem 2.8.4. It thus suffices to show that (b) implies (a); by induction, it is enough to check in case J is generated by two nonzero elements x, y. Moreover, we may form the closure of J, find a generator z, and then divide x and y by z; in other words, we may assume that 1 is in the closure of J, and then what we are to show is that $1 \in J$.

Let $I_1 \subset I_2 \subset \cdots$ be an increasing sequence of closed subintervals of $(0, r]$, with right endpoints r, whose union is all of $(0, r]$. Then x and y generate the unit ideal in $\Gamma_i = \Gamma_{I_i}$ for each i; that is, we can choose $a_i, b_i \in \Gamma_i$ with $a_i x + b_i y = 1$. Note that $b_{i+1} - b_i$ is divisible by x in Γ_i; by Lemma 2.9.3, we can choose $b \in \Gamma_{\mathrm{an},r}$ with $b - b_i$ divisible by x in Γ_i for each i. Then $by - 1$ is divisible by x in each Γ_i, hence also in $\Gamma_{\mathrm{an},r}$ (by Corollary 2.5.7); that is, x and y generate the unit ideal in $\Gamma_{\mathrm{an},r}$, as desired.

We have thus shown that (a), (b), (c) are equivalent, proving that $\Gamma_{\mathrm{an},r}$ is a Bézout ring. Since $\Gamma_{\mathrm{an,con}}$ is the union of the $\Gamma_{\mathrm{an},r}$ for $r \in (0, r_0)$, it is also a Bézout ring because any finitely generated ideal is generated by elements of some $\Gamma_{\mathrm{an},r}$. $\square$

Remark 2.9.7. In Lazard's theory (in which Γ_I becomes the ring of rigid analytic functions on the annulus $\log_{|\pi|} |u| \in I$), the implication (b) $\Longrightarrow$ (a) is [28, Proposition 11], and holds without restriction on the coefficient field. The implication (a) $\Longrightarrow$ (b) is equivalent to spherical completeness of the coefficient field [28, Théorème 2]; however, the analogue here would probably require K also to be spherically complete (compare Remark 2.8.6), which is an undesirable restriction. For instance, it would complicate the process of descending the slope filtration in Chapter 6.

3 σ-modules

We now introduce modules equipped with a semilinear endomorphism (σ-modules) and study their properties, specifically over $\Gamma_{\mathrm{an,con}}$. In order to highlight the parallels between this theory and the theory of stable vector bundles (see for instance [33]), we have shaped our presentation along the lines of that of Hartl and Pink [17]; they study vector bundles with a Frobenius structure on a punctured disc over a complete nonarchimedean field of *equal* characteristic p, and prove results very similar to our results over $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$.

Beware that our overall sign convention is "arithmetic" and not "geometric"; it thus agrees with the sign conventions of [18] (and of [19]), but disagrees with the sign convention of [17] and with the usual convention in the vector bundle setting.

Remark 3.0.1. We retain all notation from Chapter 2, except that we redefine the term "slope"; see Definition 3.4.1. In particular, K is a field complete with respect to the valuation v_K, and Γ^K is assumed to have enough r_0-units for some $r_0 > 0$. Remember that v_K is allowed to be trivial unless otherwise specified; that means any result about $\Gamma_{\mathrm{an,con}}$ also applies to $\Gamma[\pi^{-1}]$, unless its statement explicitly requires v_K to be nontrivial.

3.1 σ-modules

Definition 3.1.1. For a ring R containing $\mathcal{O}$ in which π is not a zero divisor, equipped with a ring endomorphism σ, a *σ-module* over a ring R is a finite locally free R-module M equipped with a map $F : \sigma^* M \to M$ (the *Frobenius action*) which becomes an isomorphism after inverting π. (Here $\sigma^* M = M \otimes_{R,\sigma} R$; that is, view R as a module over itself via σ, and tensor it with M over R.) We can view M as a left module over the twisted polynomial ring $R\{\sigma\}$; we can also view F as a σ-linear additive endomorphism of M. A homomorphism of σ-modules is a module homomorphism equivariant with respect to the Frobenius actions.

Remark 3.1.2. We will mostly consider σ-modules over Bézout rings like $\Gamma_{\mathrm{an,con}}$, in which case there is no harm in replacing "locally free" by "free" in the definition of a σ-module.

Remark 3.1.3. The category of σ-modules is typically not abelian (unless v_K is trivial), because we cannot form cokernels thanks to the requirement that the underlying modules be locally free.

Remark 3.1.4. For any positive integer a, σ^a is also a Frobenius lift, so we may speak of σ^a-modules. This will be relevant when we want to perform "restriction of scalars" in Section 3.2. However, there is no loss of generality in stating definitions and theorems in the case $a = 1$, i.e., for σ-modules.

Definition 3.1.5. Given a σ-module M of rank n and an integer c (which must be nonnegative if $\pi^{-1} \notin R$), define the *twist* $M(c)$ of M by c to be the module M with the Frobenius action multiplied by π^c. (Beware that this definition reflects an earlier choice of normalization, as in Remark 2.1.11, and a choice of a sign convention.) If π is invertible in R, define the *dual* $M^\vee$ of M to be the σ-module $\mathrm{Hom}_R(M, R) \cong (\wedge^{n-1} M) \otimes (\wedge^n M)^{\otimes -1}$ and the *internal hom* of M, N as $M^\vee \otimes N$.

Definition 3.1.6. Given a σ-module M over a ring R, let $H^0(M)$ and $H^1(M)$ denote the kernel and cokernel, respectively, of the map $F - 1$ on M; note that if N is another σ-module, then there is a natural bilinear map $H^0(M) \times H^1(N) \to H^1(M \otimes N)$. Given two σ-modules M_1 and M_2 over R, put $\mathrm{Ext}(M_1, M_2) = H^1(M_1^\vee \otimes M_2)$; a standard homological calculation (as in [17, Proposition 2.4]) shows that $\mathrm{Ext}(M_1, M_2)$ coincides with the Yoneda Ext^1 in this category. That is, $\mathrm{Ext}(M_1, M_2)$ classifies short exact sequences $0 \to M_2 \to M \to M_1 \to 0$ of σ-modules over R, up to isomorphisms

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M_2 & \longrightarrow & M & \longrightarrow & M_1 & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & M_2 & \longrightarrow & M & \longrightarrow & M_1 & \longrightarrow & 0 \end{array}$$

which induce the identity maps on M_1 and M_2.

3.2 Restriction of Frobenius

We now introduce two functors analogous to those induced by the "finite maps" in [17, Section 7]. Beware that the analogy is not perfect; see Remark 3.2.2.

Definition 3.2.1. Fix a ring R equipped with an endomorphism σ. For a a positive integer, let $[a] : R\{\sigma^a\} \to R\{\sigma\}$ be the natural inclusion homomorphism of twisted polynomial rings. Define the *a-pushforward functor* $[a]_*$, from σ-modules to σ^a-modules, to be the restriction functor along $[a]$. Define the *a-pullback functor* $[a]^*$, from σ^a-modules to σ-modules, to be the extension of scalars functor

$$M \mapsto R\{\sigma\} \otimes_{R\{\sigma^a\}} M.$$

Note that $[a]^*$ and $[a]_*$ are left and right adjoints of each other. Also, $[a]^*[a']^* = [aa']^*$ and $[a]_*[a']_* = [aa']_*$. Furthermore, $[a]_*(M(c)) = ([a]_*M)(ac)$.

Remark 3.2.2. There are some discrepancies in the analogy with [17], due to the fact that there the corresponding map $[a]$ is actually a homomorphism of the underlying ring, rather than a change of Frobenius. The result is that some (but not all!) of the properties of the pullback and pushforward are swapped between here and [17]. For an example of this mismatch in action, see Proposition 3.4.4.

Remark 3.2.3. The functors $[a]$ will ultimately serve to rescale the slopes of a σ-module; using them makes it possible to avoid the reliance in [19, Chapter 4] on making extensions of $\mathcal{O}$. Among other things, this lets us get away with normalizing w in terms of the choice of $\mathcal{O}$, since we will not have to change that choice at any point except in Lemma 5.2.4.

Lemma 3.2.4. *For any positive integer a and any integer c, $[a]_*[a]^*(R(c)) \cong R(c)^{\oplus a}$.*

Proof. We can write

$$[a]_*[a]^*(R(c)) \cong \oplus_{i=0}^{a-1}\{\sigma^i\}(R(c)),$$

where on the right side $R\{\sigma^a\}$ acts separately on each factor. Hence the claim follows. (Compare [17, Proposition 7.4].) □

Lemma 3.2.5. *Suppose that the residue field of $\mathcal{O}$ contains an algebraic closure of $\mathbb{F}_q$. For i a positive integer, let L_i be the fixed field of $\mathcal{O}[\pi^{-1}]$ under σ^i.*

(a) *For any σ-module M over $\Gamma_{\mathrm{an,con}}$ and any positive integer a, $H^0([a]_*M) = H^0(M) \otimes_{L_1} L_a$.*

(b) *For any σ-modules M and N over $\Gamma_{\mathrm{an,con}}$ and any positive integer a, $M \cong N$ if and only if $[a]_*M \cong [a]_*N$.*

Proof. (a) It suffices to show that $H^0([a]_*M)$ admits a basis invariant under the induced action of σ. Since σ generates $\mathrm{Gal}(L_a/L_1)$, this follows from Hilbert's Theorem 90.

(b) A morphism $[a]_*M \to [a]_*N$ corresponds to an element of $V = H^0(([a]_*M)^\vee \otimes [a]_*N) \cong H^0([a]_*(M^\vee \otimes N))$, which by (a) coincides with $H^0(M^\vee \otimes N) \otimes_{L_1} L_a$. If there is an isomorphism $[a]_*M \cong [a]_*N$, then the determinant locus on V is not all of V; hence the same is true on $H^0(M^\vee \otimes N)$. We can thus find an F-invariant element of $M^\vee \otimes N$ corresponding to an isomorphism $M \cong N$. (Compare [17, Propositions 7.3 and 7.5].)

□

3.3 σ-modules of rank 1

In this section, we analyze some σ-modules of rank 1 over $\Gamma_{\mathrm{an,con}}$; this amounts to solving some simple equations involving σ, as in [19, Proposition 3.19] (compare also [17, Propositions 3.1 and 3.3]).

Definition 3.3.1. Define the twisted powers $\pi^{\{m\}}$ of π by the two-way recurrence

$$\pi^{\{0\}} = 1, \qquad \pi^{\{m+1\}} = (\pi^{\{m\}})^\sigma \pi.$$

First, we give a classification result.

Proposition 3.3.2. *Let M be a σ-module of rank 1 over R, for R one of $\Gamma^{\mathrm{alg}}_{\mathrm{con}}, \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}], \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$. Then there exists a unique integer n, which is nonnegative if $R = \Gamma^{\mathrm{alg}}_{\mathrm{con}}$, such that $M \cong R(n)$.*

Proof. Let $\mathbf{v}$ be a generator of M, and write $F\mathbf{v} = x\mathbf{v}$. Then x must be a unit in R, so by Corollary 2.5.12, $x \in \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$ and so $w(x)$ is defined. If $M \cong R(n)$, we must then have $n = w(x)$; hence n is unique if it exists.
Since the residue field of $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$ is algebraically closed, we can find a unit u in $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$ such that $u^\sigma \pi^{-n} x \equiv u \pmod{\pi}$. Choose $r > 0$ such that $w_r(u^\sigma \pi^{-n} x/u - 1) > 0$; then there exists a unit $y \in \Gamma^{\mathrm{alg}}$ with $u^\sigma \pi^{-n} x/u = y^\sigma/y$, and a direct calculation (by induction on m) shows that $v_{m,r}(y) > 0$ for all $m > 0$. (For details, see the proof of Lemma 5.4.1.) Hence y is a unit in $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$, and so $\mathbf{w} = (u/y)\mathbf{v}$ is a generator of M satisfying $F\mathbf{w} = \pi^n \mathbf{w}$. Thus there exists an isomorphism $M \cong R(n)$. □

We next compute some instances of H^0.

Lemma 3.3.3. *Let n be a nonnegative integer. If $x \in \Gamma_{\mathrm{an,con}}$ and $x - \pi^n x^\sigma \in \Gamma_{\mathrm{con}}[\pi^{-1}]$, then $x \in \Gamma_{\mathrm{con}}[\pi^{-1}]$.*

Proof. Suppose the contrary; put $y = x - \pi^n x^\sigma$. We can find m with $v_m(y) = \infty$ and $0 < v_m(x) < \infty$, since both hold for m sufficiently small by Corollary 2.5.6. Then

$$v_m(x) > q^{-1} v_m(x) = q^{-1} v_m(\pi^n x^\sigma) = q^{-1} v_{m-n}(x^\sigma) = v_{m-n}(x) \geq v_m(x),$$

contradiction. □

Proposition 3.3.4. *Let n be an integer.*

(a) If $n = 0$, then $H^0(\Gamma_{\mathrm{con}}[\pi^{-1}](n)) = H^0(\Gamma_{\mathrm{an,con}}(n)) \neq 0$; moreover, any nonzero element of $H^0(\Gamma_{\mathrm{an,con}})$ is a unit in $\Gamma_{\mathrm{an,con}}$.

(b) If $n > 0$, then $H^0(\Gamma_{\mathrm{an,con}}(n)) = 0$.

(c1) If $n < 0$ and v_K is trivial, then $H^0(\Gamma_{\mathrm{an,con}}(n)) = 0$.

(c2) If $n < 0$, v_K is nontrivial, and K is perfect, then $H^0(\Gamma_{\mathrm{an,con}}(n)) \neq 0$.

Proof. The group $H^0(R(n))$ consists of those $x \in R$ with

$$\pi^n x^\sigma = x, \tag{3.3.5}$$

so our assertions are all really about the solvability of this equation.

(a) If $n = 0$, then $H^0(\Gamma_{\mathrm{con}}[\pi^{-1}](n)) = H^0(\Gamma_{\mathrm{an,con}}(n))$ by Lemma 3.3.3, and the former equals the fixed field of $\mathcal{O}[\pi^{-1}]$ under σ.

(b) By Lemma 3.3.3, any solution x of (3.3.5) over $\Gamma_{\mathrm{an,con}}$ actually belongs to $\Gamma_{\mathrm{con}}[\pi^{-1}]$. In particular, $w(x) = w(x^\sigma)$ is well-defined. But (3.3.5) yields $w(x^\sigma) + n = w(x)$, which for $n > 0$ forces $x = 0$.

(c1) If $n < 0$ and v_K is trivial, then $w(x) = w(x^\sigma)$ is well-defined, but (3.3.5) yields $w(x^\sigma) + n = w(x)$, which forces $x = 0$.

(c2) If $n < 0$, v_K is nontrivial, and K is perfect, we may pick $\overline{u} \in K$ with $v_K(\overline{u}) > 0$ (since v_K is nontrivial), and then set x to be the limit of the convergent series
$$\sum_{m \in \mathbb{Z}} (\pi^{\{m\}})^n [\overline{u}^{q^m}]$$
to obtain a nonzero solution of (3.3.5).

□

Remark 3.3.6. If $n < 0$, v_K is nontrivial, and K is not perfect, then the size of $H^0(\Gamma_{\mathrm{an,con}}(n))$ depends on the particular choice of the Frobenius lift σ on $\Gamma_{\mathrm{an,con}}$. For instance, in the notation of Section 2.3, if σ is a so-called "standard Frobenius lift" sending u to u^q, then any solution $x = \sum x_i u^i$ of (3.3.5) must have $x_i = 0$ whenever i is not divisible by q. By the same token, $x_i = 0$ whenever i is not divisible by q^2, or by q^3, and so on; hence we must have $x \in \mathcal{O}[\pi^{-1}]$, which as in (b) above is impossible for $n > 0$. On the other hand, if $u^\sigma = (u+1)^q - 1$, then $x = \log(1+u) \in \Gamma_{\mathrm{an,con}}$ satisfies $x^\sigma = qx$; indeed, the existence of such an x is a backbone of the theory of (Φ, Γ)-modules associated to p-adic Galois representations, as in [4].

We next consider H^1.

Proposition 3.3.7. *Let n be an integer.*

(a) If $n = 0$ and K is separably closed, then $H^1(\Gamma_{\mathrm{con}}[\pi^{-1}](n)) = H^1(\Gamma_{\mathrm{an,con}}(n)) = 0$.

(b1) If $n \geq 0$, then the map $H^1(\Gamma_{\mathrm{con}}[\pi^{-1}](n)) \to H^1(\Gamma_{\mathrm{an,con}}(n))$ is injective.

(b2) If $n > 0$ and v_K is trivial, then $H^1(\Gamma_{\mathrm{an,con}}(n)) = 0$.

(b3) If $n > 0$, v_K is nontrivial, and K is perfect, then $H^1(\Gamma_{\mathrm{an,con}}(n)) \neq 0$, with a nonzero element given by $[\overline{x}]$ for any $\overline{x} \in K$ with $v_K(\overline{x}) < 0$.

(c) If $n < 0$ and K is perfect, then $H^1(\Gamma_{\mathrm{con}}[\pi^{-1}](n)) = H^1(\Gamma_{\mathrm{an,con}}(n)) = 0$.

Proof. The group $H^1(R(n))$ consists of the quotient of the additive group of R by the subgroup of those $x \in R$ for which the equation

$$x = y - \pi^n y^\sigma \tag{3.3.8}$$

has a solution $y \in R$, so our assertions are all really about the solvability of this equation.

(a) If $n = 0$ and K is separably closed, then for each $x \in \Gamma$, there exists $y \in \Gamma$ such that $x \equiv y - y^q \pmod{\pi}$. By iterating this construction, we can produce for any $x \in \Gamma[\pi^{-1}]$ an element $y \in \Gamma[\pi^{-1}]$ satisfying (3.3.8), such that

$$v_m(y) \geq \min\{v_m(x), v_m(x)/q\};$$

in particular, if $x \in \Gamma_{\mathrm{con}}[\pi^{-1}]$, then $y \in \Gamma_{\mathrm{con}}[\pi^{-1}]$. Moreover, given $x \in \Gamma_{\mathrm{an,con}}$, we can write x as a convergent series of elements of $\Gamma_{\mathrm{con}}[\pi^{-1}]$, and thus produce a solution of (3.3.8). Hence $H^1(\Gamma_{\mathrm{con}}(n)) = H^1(\Gamma_{\mathrm{an,con}}(n)) = 0$.

(b1) This follows at once from Lemma 3.3.3.

(b2) If $n > 0$ and v_K is trivial, then for any $x \in \Gamma_{\mathrm{an,con}}$, the series

$$y = \sum_{m=0}^{\infty} (\pi^{\{m\}})^n x^{\sigma^m}$$

converges π-adically to a solution of (3.3.8).

(b3) By (b1), it suffices to show that $x = [\overline{x}]$ represents a nonzero element of $H^1(\Gamma_{\mathrm{con}}[\pi^{-1}](n))$. By (b2), there exists a unique $y \in \Gamma[\pi^{-1}]$ satisfying (3.3.8); however, we have $v_{mn}(y) = q^m v_K(\overline{x})$ for all $m \geq 0$, and so $y \notin \Gamma_{\mathrm{con}}[\pi^{-1}]$.

(c) Let $x = \sum_i [\overline{x_i}]\pi^i$ be the Teichmüller presentation of x. Pick $c > 0$, and let z_1 and z_2 be the sums of $[\overline{x_i}]\pi^i$ over those i for which $v_K(\overline{x_i}) < c$ and $v_K(\overline{x_i}) \geq c$, respectively. Then the sums

$$y_1 = \sum_{m=0}^{\infty} -(\pi^{\{-m-1\}})^n z_1^{\sigma^{-m-1}}$$
$$y_2 = \sum_{m=0}^{\infty} (\pi^{\{m\}})^n z_2^{\sigma^m}$$

converge to solutions of $z_1 = y_1 - \pi^n y_1^\sigma$ and $z_2 = y_2 - \pi^n y_2^\sigma$, respectively. Hence $y = y_1 + y_2$ is a solution of (3.3.8). Moreover, if $x \in \Gamma_{\mathrm{con}}[\pi^{-1}]$, then $\overline{x_i} = 0$ for i sufficiently small, so we can choose c to ensure $z_2 = 0$; then $y = y_1 \in \Gamma_{\mathrm{con}}[\pi^{-1}]$.

□

3.4 Stability and semistability

As in [17], we can set up a formal analogy between the study of σ-modules over $\Gamma_{\mathrm{an,con}}$ and the study of stability of vector bundles.

Definition 3.4.1. For M a σ-module of rank 1 over $\Gamma_{\mathrm{an,con}}$ generated by some $\mathbf{v}$, define the *degree* of M, denoted $\deg(M)$, to be the unique integer n such that $M \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}} \cong \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}(n)$, as provided by Proposition 3.3.2; concretely, $\deg(M)$ is the valuation of the unit via which F acts on a generator of M. For a σ-module M over $\Gamma_{\mathrm{an,con}}$ of rank n, define $\deg(M) = \deg(\wedge^n M)$. Define $\mu(M) = \deg(M)/\operatorname{rank}(M)$; we refer to $\mu(M)$ as the *slope* of M (or as the *weight* of M, per the terminology of [17, Section 6]).

Lemma 3.4.2. *Let M be a σ-module over $\Gamma_{\mathrm{an,con}}$, and let N be a σ-submodule of M with $\operatorname{rank}(M) = \operatorname{rank}(N)$. Then $\deg(N) \geq \deg(M)$, with equality if and only if $M = N$; moreover, equality must hold if v_K is trivial.*

Proof. By taking exterior powers, it suffices to check this for $\operatorname{rank} M = \operatorname{rank} N = 1$; also, there is no harm in assuming that K is algebraically closed. By Proposition 3.3.2, $M \cong \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}(c)$ and $N \cong \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}(d)$ for some integers c, d. By twisting, we may reduce to the case $d = 0$. Then by Proposition 3.3.4, we have $0 \geq c$, with equality forced if v_K is trivial; moreover, if $c = 0$, then N contains a generator which also belongs to $H^0(M)$. But every nonzero element of the latter also generates M, so $c = 0$ implies $M = N$. (Compare [17, Proposition 6.2].) □

Lemma 3.4.3. *If $0 \to M_1 \to M \to M_2 \to 0$ is a short exact sequence of σ-modules over $\Gamma_{\mathrm{an,con}}$, then $\deg(M) = \deg(M_1) + \deg(M_2)$.*

Proof. Put $n_1 = \mathrm{rank}(M_1)$ and $n_2 = \mathrm{rank}(M_2)$. Then the claim follows from the existence of the isomorphism

$$\wedge^{n_1+n_2} M \cong (\wedge^{n_1} M_1) \otimes (\wedge^{n_2} M_2)$$

of σ-modules. □

Proposition 3.4.4. *Let a be a positive integer, let M be a σ-module over $\Gamma_{\mathrm{an,con}}$, and let N be a σ^a-module over $\Gamma_{\mathrm{an,con}}$.*

(a) $\deg([a]_* M) = a \deg(M)$ *and* $\deg([a]^* N) = \deg(N)$.

(b) $\mathrm{rank}([a]_* M) = \mathrm{rank}(M)$ *and* $\mathrm{rank}([a]^* N) = a\, \mathrm{rank}(N)$.

(c) $\mu([a]_* M) = a\mu(M)$ *and* $\mu([a]^* N) = \frac{1}{a}\mu(N)$.

Proof. Straightforward (compare [17, Proposition 7.1], but note that the roles of the pullback and pushforward are interchanged here). □

Definition 3.4.5. We say a σ-module M over $\Gamma_{\mathrm{an,con}}$ is *semistable* if $\mu(M) \leq \mu(N)$ for any nonzero σ-submodule N of M. We say M is *stable* if $\mu(M) < \mu(N)$ for any nonzero proper σ-submodule N of M. Note that the direct sum of semistable σ-modules of the same slope is also semistable. By Proposition 3.4.4, for any positive integer a, if $[a]_* M$ is (semi)stable, then M is (semi)stable.

Remark 3.4.6. As noted earlier, the inequalities are reversed from the usual definitions of stability and semistability for vector bundles, because of an overall choice of sign convention.

Remark 3.4.7. Beware that this use of the term "semistable" is only distantly related to its use to describe p-adic Galois representations!

Lemma 3.4.8. *For any integer c and any positive integer n, the σ-module $\Gamma_{\mathrm{an,con}}(c)^{\oplus n}$ is semistable of slope c.*

Proof. There is no harm in assuming that K is algebraically closed, and that v_K is nontrivial. Let N be a nonzero σ-submodule of M of rank d' and degree c'. Then

$$\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}(c') \cong \wedge^{d'} N \subseteq \wedge^{d'} \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}(c)^{\oplus n} \cong \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}(cd')^{\oplus \binom{n}{d'}}.$$

In particular, $H^0(\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}(cd' - c')) \neq 0$; by Proposition 3.3.4(b), this implies $c' \geq cd'$, yielding semistability. (Compare [17, Proposition 6.3(b)].) □

Lemma 3.4.9. *For any positive integer a and any integer c, the σ-module $[a]^*(\Gamma_{\mathrm{an,con}}(c))$ over $\Gamma_{\mathrm{an,con}}$ is semistable of rank a, degree c, and slope c/a. Moreover, if a and c are coprime, then $[a]^*(\Gamma_{\mathrm{an,con}}(c))$ is stable.*

Proof. By Lemma 3.2.4 and Lemma 3.4.8, $[a]_*[a]^*(\Gamma_{\mathrm{an,con}}(c))$ is semistable of rank a and slope c; as noted in Definition 3.4.5, it follows that $[a]^*(\Gamma_{\mathrm{an,con}}(c))$ is semistable.
Let M be a nonzero σ-submodule of $[a]^*(\Gamma_{\mathrm{an,con}}(c))$. If a and c are coprime, then $\deg(M) = (c/a)\,\mathrm{rank}(M)$ is an integer; since $\mathrm{rank}(M) \leq a$, this is only possible for $\mathrm{rank}(M) = a$. But then Lemma 3.4.2 implies that $\mu(M) > c/a$ unless $M = [a]^*(\Gamma_{\mathrm{an,con}}(c))$. We conclude that $[a]^*(\Gamma_{\mathrm{an,con}}(c))$ is stable. (Compare [17, Proposition 8.2].) □

3.5 Harder-Narasimhan filtrations

Using the notions of degree and slope, we can make the usual formal construction of Harder-Narasimhan filtrations, with its usual properties.

Definition 3.5.1. Given a multiset S of n real numbers, define the *Newton polygon* of S to be the graph of the piecewise linear function on $[0, n]$ sending 0 to 0, whose slope on $[i-1, i]$ is the i-th smallest element of S; we refer to the point on the graph corresponding to the image of n as the *endpoint* of the polygon. Conversely, given such a graph, define its *slope multiset* to be the slopes of the piecewise linear function on $[i-1, i]$ for $i = 1, \dots, n$. We say that the Newton polygon of S *lies above* the Newton polygon of S' if no vertex of the polygon of S lies below the polygon of S', and the two polygons have the same endpoint.

Definition 3.5.2. Let M be a σ-module over $\Gamma_{\mathrm{an,con}}$. A *semistable filtration* of M is an exhaustive filtration $0 = M_0 \subset M_1 \subset \cdots \subset M_l = M$ of M by saturated σ-submodules, such that each successive quotient M_i/M_{i-1} is semistable of some slope s_i. A *Harder-Narasimhan filtration* (or *HN-filtration*) of M is a semistable filtration with $s_1 < \cdots < s_l$. An HN-filtration is unique if it exists, as M_1 can then be characterized as the unique maximal σ-submodule of M of minimal slope, and so on.

Definition 3.5.3. Let M be a σ-module over $\Gamma_{\mathrm{an,con}}$. Given a semistable filtration $0 = M_0 \subset M_1 \subset \cdots \subset M_l = M$ of M, form the multiset consisting of, for $i = 1, \dots, l$, the slope $\mu(M_i/M_{i-1})$ with multiplicity $\mathrm{rank}(M_i/M_{i-1})$. We call this the *slope multiset* of the filtration, and we call the associated Newton polygon the *slope polygon* of the filtration. If M admits a Harder-Narasimhan filtration, we refer to the slope multiset as the *Harder-Narasimhan slope multiset* (or *HN-slope multiset*) of M, and to the Newton polygon as the *Harder-Narasimhan polygon* (or *HN-polygon*) of M.

Proposition 3.5.4. *Let M be a σ-module over $\Gamma_{\mathrm{an,con}}$ admitting an HN-filtration. Then the HN-polygon lies above the slope polygon of any semistable filtration of M.*

Proof. Let $0 = M_0 \subset M_1 \subset \cdots \subset M_l = M$ be an HN-filtration, and let $0 = M'_0 \subset M'_1 \subset \cdots \subset M'_m = M$ be a semistable filtration. To prove the

inequality, it suffices to prove that for each of $i = 1, \dots, l$, we can choose $\mathrm{rank}(M/M_i)$ slopes from the slope multiset of the semistable filtration whose sum is greater than or equal to the sum of the greatest $\mathrm{rank}(M/M_i)$ HN-slopes of M; note that the latter is just $\deg(M/M_i)$.
For $l = 1, \dots, m$, put

$$d_l = \mathrm{rank}(M'_l + M_i) - \mathrm{rank}(M'_{l-1} + M_i) \leq \mathrm{rank}(M'_l/M'_{l-1}).$$

Since M'_l/M'_{l-1} is semistable and $(M'_l + M_i)/(M'_{l-1} + M_i)$ is a quotient of M'_l/M'_{l-1}, we have

$$\mu(M'_l/M'_{l-1}) \geq \mu((M'_l + M_i)/(M'_{l-1} + M_i)).$$

However, we also have

$$\sum_{i=1}^{l} d_l \mu((M'_l + M_i)/(M'_{l-1} + M_i)) = \deg(M/M_i),$$

yielding the desired inequality. □

Remark 3.5.5. We will ultimately prove that every σ-module over $\Gamma_{\mathrm{an,con}}$ admits a Harder-Narasimhan filtration (Proposition 4.2.5), and that the successive quotients become isomorphic over $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$ to direct sums of "standard" σ-modules of the right slope (Theorems 6.3.3 and 6.4.1). Using the formalism of Harder-Narasimhan filtrations makes it a bit more convenient to articulate the proofs of these assertions.

3.6 Descending subobjects

We will ultimately be showing that the formation of a Harder-Narasimhan filtration of a σ-module over $\Gamma_{\mathrm{an,con}}$ commutes with base change. In order to prove this sort of statement, it will be useful to have a bit of terminology.

Definition 3.6.1. Given an injection $R \hookrightarrow S$ of integral domains equipped with compatible endomorphisms σ, a σ-module M over R, and a saturated σ-submodule N_S of $M_S = M \otimes_R S$, we say that N_S *descends to* R if there is a saturated σ-submodule N of M such that $N_S = N \otimes_R S$; note that N is unique if it exists, because it can be characterized as $M \cap N_S$. Likewise, given a filtration of M_S, we say the filtration descends to R if it is induced by a filtration of M.

The following lemma lets us reduce most descent questions to consideration of submodules of rank 1.

Lemma 3.6.2. *With notation as in Definition 3.6.1, suppose that R is a Bézout domain, and put $d = \mathrm{rank}\, N_S$. Then N_S descends to R if and only if $\wedge^d N_S \subseteq (\wedge^d M)_S$ descends to R.*

Proof. If $N_S = N \otimes_R S$, then $\wedge^d N_S = (\wedge^d N) \otimes_R S$ descends to R. Conversely, if $\wedge^d N_S = (N') \otimes_R S$, let N be the σ-submodule of M consisting of those $\mathbf{v} \in M$ such that $\mathbf{v} \wedge \mathbf{w} = 0$ for all $\mathbf{w} \in N'$. Then N is saturated and $N \otimes_R S = N_S$, since N is defined by linear conditions which in M_S cut out precisely N_S. Since R is a Bézout domain, this suffices to ensure that N is free; since N is visibly stable under F, N is in fact a σ-submodule of M, and so N_S descends to R. $\square$

4 Slope filtrations of σ-modules

In this chapter, we give a classification of σ-modules over $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$, as in [19, Chapter 4]. However, this presentation looks somewhat different, mainly because of the formalism introduced in the previous chapter. We also have integrated into a single presentation the cases where v_K is nontrivial and where v_K is trivial; these are presented separately in [19] (in Chapters 4 and 5 respectively). These two cases do have different flavors, which we will point out as we go along.

Beware that although we have mostly made the exposition self-contained, there remains one notable exception: we do not repeat the key calculation made in [19, Lemma 4.12]. See Lemma 4.3.3 for the relevance of this calculation.

Convention 4.0.1. To lighten the notational load, we write $\mathcal{R}$ for $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$. Whenever working over $\mathcal{R}$, we also make the harmless assumption that π is σ-invariant.

4.1 Standard σ-modules

Following [17, Section 8], we introduce the standard building blocks into which we will decompose σ-modules over $\mathcal{R}$.

Definition 4.1.1. Let c, d be coprime integers with $d > 0$. Define the σ-module $M_{c,d} = [d]^*(\mathcal{R}(c))$ over $\mathcal{R}$; that is, $M_{c,d}$ is freely generated by $\mathbf{e}_1, \dots, \mathbf{e}_d$ with

$$F\mathbf{e}_1 = \mathbf{e}_2, \quad \dots, \quad F\mathbf{e}_{d-1} = \mathbf{e}_d, \quad F\mathbf{e}_d = \pi^c \mathbf{e}_1.$$

This σ-module is stable of slope c/d by Lemma 3.4.9. We say a σ-module M is *standard* if it is isomorphic to some $M_{c,d}$; in that case, we say a basis $\mathbf{e}_1, \dots, \mathbf{e}_d$ as above is a *standard basis* of M.

Lemma 4.1.2. *(a) $M_{c,d} \otimes M_{c',d'} \cong M_{c'',d''}^{\oplus dd'/d''}$, where $c/d + c'/d' = c''/d''$ in lowest terms.*

(b) $M_{c,d}(c') \cong M_{c+c'd,d}$.

(c) $M_{c,d}^\vee \cong M_{-c,d}$.

Proof. To verify (a), it is enough to do so after applying $[dd']_*$ thanks to Lemma 3.2.5. Then the desired isomorphism follows from Lemma 3.2.4. Assertion (b) follows from (a), and (c) follows from the explicit description of $M_{c,d}$ given above. (Compare [17, Proposition 8.3].) $\square$

Proposition 4.1.3. *Let c, d be coprime integers with $d > 0$.*

(a) The group $H^0(M_{c,d})$ is nonzero if and only if v_K is nontrivial and $c/d \leq 0$, or v_K is trivial and $c/d = 0$.

(b) The group $H^1(M_{c,d})$ is nonzero if and only if v_K is nontrivial and $c/d > 0$.

Proof. These assertions follow from Propositions 3.3.4 and 3.3.7, plus the fact that $H^i([d]^*M) \cong H^i(M)$ for $i = 0, 1$. $\square$

Corollary 4.1.4. *Let c, c', d, d' be integers, with d, d' positive and* $\gcd(c, d) = \gcd(c', d') = 1$.

(a) We have $\mathrm{Hom}(M_{c',d'}, M_{c,d}) \neq 0$ *if and only if v_K is nontrivial and $c'/d' \geq c/d$, or v_K is trivial and $c'/d' = c/d$.*

(b) We have $\mathrm{Ext}(M_{c',d'}, M_{c,d}) \neq 0$ *if and only if v_K is nontrivial and $c'/d' < c/d$.*

Remark 4.1.5. One can show that $\mathrm{End}(M_{c,d})$ is a division algebra (this can be deduced from the fact that $M_{c,d}$ is stable) and even describe it explicitly, as in [17, Proposition 8.6]. For our purposes, it will be enough to check that $\mathrm{End}(M_{c,d})$ is a division algebra after establishing the existence of Dieudonné-Manin decompositions; see Corollary 4.5.9.

4.2 Existence of eigenvectors

In classifying σ-modules over $\mathcal{R}$, it is useful to employ the language of "eigenvectors".

Definition 4.2.1. Let d be a positive integer. A *d-eigenvector* (or simply *eigenvector* if $d = 1$) of a σ-module M over $\mathcal{R}$ is a nonzero element $\mathbf{v}$ of M such that $F^d\mathbf{v} = \pi^c\mathbf{v}$ for some integer c. We refer to the quotient c/d as the *slope* of $\mathbf{v}$.

Proposition 4.2.2. *Suppose that v_K is nontrivial. Then every nontrivial σ-module over $\mathcal{R}$ contains an eigenvector.*

Proof. The calculation is basically that of [19, Proposition 4.8]: use the fact that F makes things with "very positive partial valuations" converge better whereas F^{-1} makes things with "very negative partial valuations" converge better. However, one can simplify the final analysis a bit, as is done in [17, Theorem 4.1].

We first set some notation as in [19, Proposition 4.8]. Let M be a nontrivial σ-module over $\mathcal{R}$. Choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of M, and define the invertible $n \times n$ matrix A over $\mathcal{R}$ by the equation $F\mathbf{e}_j = \sum_i A_{ij}\mathbf{e}_i$. Choose $r > 0$ such that A and its inverse have entries in $\Gamma^{\mathrm{alg}}_{\mathrm{an},r}$. Choose $\epsilon > 0$, and choose an integer c with $c \leq \min\{w_r(A), w_r((A^{-1})^{\sigma^{-1}})\} - \epsilon$. Choose an integer $m > c$ such that the interval

$$\left(\frac{-c+m}{(q-1)r}, \frac{q(c+m)}{(q-1)r} \right)$$

is nonempty (true for m sufficiently large), and choose d in the intersection of that interval with the image of v_K. (The choice of d is possible because v_K is nontrivial and the residue field of $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$ is algebraically closed, so the image of v_K contains a copy of $\mathbb{Q}$ and hence is dense in $\mathbb{R}$.)

For an interval $I \subseteq (0, r]$, define the functions $a, b : \Gamma_I \to \Gamma_I$ as follows. For $x \in \Gamma_I$, let $\sum_{j \in \mathbb{Z}} [\overline{x_j}]\pi^j$ be the Teichmüller presentation of x (as in Definition 2.5.1). Let $a(x)$ and $b(x)$ be the sums of $[\overline{x_j}]\pi^j$ over those j with $v_K(\overline{x_j}) < d$ and $v_K(\overline{x_j}) \geq d$, respectively. We may think of a and b as splitting x into "negative" and "positive" terms. (This decomposition shares its canonicality with the corresponding decomposition in [17, Theorem 4.1] but not with the one in [19, Proposition 4.8].)

For $I \subseteq (0, r]$, let M_I denote the Γ_I-span of the $\mathbf{e}_i$. For $\mathbf{v} \in M_I$, write $\mathbf{v} = \sum_i x_i \mathbf{e}_i$, and for $s \in I$, define

$$\begin{aligned} v_{n,s}(\mathbf{v}) &= \min_i \{v_{n,s}(x_i)\} \\ w_s(\mathbf{v}) &= \min_i \{w_s(x_i)\}. \end{aligned}$$

Put $a(\mathbf{v}) = \sum_i a(x_i)\mathbf{e}_i$ and $b(\mathbf{v}) = \sum_i b(x_i)\mathbf{e}_i$; then by the choice of d, we have for $\mathbf{v} \in M_{(0,r]}$,

$$\begin{aligned} w_r(\pi^m F^{-1}(a(\mathbf{v}))) &\geq w_r(a(\mathbf{v})) + \epsilon \\ w_r(\pi^{-m} F(b(\mathbf{v}))) &\geq w_r(b(\mathbf{v})) + \epsilon. \end{aligned}$$

Put

$$f(\mathbf{v}) = \pi^{-m} b(\mathbf{v}) - F^{-1}(a(\mathbf{v})).$$

If $\mathbf{v} \in M_{(0,r]}$ is such that $\mathbf{w} = F\mathbf{v} - \pi^m \mathbf{v}$ also lies in $M_{(0,r]}$, then

$$\begin{aligned} &F(\mathbf{v} + f(\mathbf{w})) - \pi^m(\mathbf{v} + f(\mathbf{w})) \\ &\qquad = F(f(\mathbf{w})) - \pi^m f(\mathbf{w}) + \mathbf{w} \\ &\qquad = F(\pi^{-m} b(\mathbf{w})) - a(\mathbf{w}) - b(\mathbf{w}) + \pi^m F^{-1}(a(\mathbf{w})) + \mathbf{w} \\ &\qquad = \pi^{-m} F(b(\mathbf{w})) + \pi^m F^{-1}(a(\mathbf{w})) \end{aligned}$$

lies in $M_{(0,r]}$ as well, and

$$\begin{aligned} w_r(f(\mathbf{w})) &\geq w_r(\pi^{-m}\mathbf{w}) \\ w_r(F(\mathbf{v} + f(\mathbf{w})) - \pi^m(\mathbf{v} + f(\mathbf{w}))) &\geq w_r(\mathbf{w}) + \epsilon. \end{aligned}$$

Now define a sequence $\{\mathbf{v}_l\}_{l=0}^{\infty}$ in $M_{(0,r]}$ as follows. Pick $\overline{x} \in K$ with $v_K(\overline{x}) = d$, and set

$$\mathbf{v}_0 = \pi^{-m}[\overline{x}]\mathbf{e}_1 + [\overline{x}^{1/q}]F^{-1}\mathbf{e}_1.$$

Given $\mathbf{v}_l \in M_{(0,r]}$, set

$$\mathbf{w}_l = F\mathbf{v}_l - \pi^m \mathbf{v}_l, \qquad \mathbf{v}_{l+1} = \mathbf{v}_l + f(\mathbf{w}_l).$$

We calculated above that $\mathbf{w}_l \in M_{(0,r]}$ implies $\mathbf{w}_{l+1} \in M_{(0,r]}$; since $\mathbf{w}_0 \in M_{(0,r]}$ evidently, we have $\mathbf{w}_l \in M_{(0,r]}$ for all l, so $\mathbf{v}_{l+1} \in M_{(0,r]}$ and the iteration continues. Moreover,

$$\begin{aligned} w_r(\mathbf{v}_{l+1} - \mathbf{v}_l) &\geq w_r(\pi^{-m}\mathbf{w}_l) \\ w_r(\mathbf{w}_l) &= w_r(F(\mathbf{v}_{l-1} + f(\mathbf{w})) - \pi^m(\mathbf{v}_{l-1} + f(\mathbf{w}))) \\ &\geq w_r(\mathbf{w}_{l-1}) + \epsilon. \end{aligned}$$

Hence $\mathbf{w}_l \to 0$ and $f(\mathbf{w}_l) \to 0$ under w_r as $l \to \infty$, so the $\mathbf{v}_l$ converge to a limit $\mathbf{v} \in M_{[r,r]}$.

We now check that $\mathbf{v} \neq 0$. Since $w_r(\mathbf{w}_{l+1}) \geq w_r(\mathbf{w}_l) + \epsilon$ for all l, we certainly have $w_r(\mathbf{w}_l) \geq w_r(\mathbf{w}_0)$ for all l, and hence $w_r(\mathbf{v}_{l+1} - \mathbf{v}_l) \geq w_r(\pi^{-m}\mathbf{w}_0)$. To compute $w_r(\mathbf{w}_0)$, note that $w_r(\pi^{-m}[\overline{x}]\mathbf{e}_1) = dr - m$, whereas

$$\begin{aligned} w_r([\overline{x}^{1/q}]F^{-1}\mathbf{e}_1) &\geq dr/q + c \\ &> dr - m \end{aligned}$$

by the choice of d. Hence $w_r(\mathbf{v}_0) = dr - m$. On the other hand,

$$\begin{aligned} w_r(\pi^{-m}\mathbf{w}_0) &= w_r(\pi^{-m}F\mathbf{v}_0 - \mathbf{v}_0) \\ &= w_r(\pi^{-2m}[\overline{x}^q]F\mathbf{e}_1 - [\overline{x}^{1/q}]F^{-1}\mathbf{e}_1) \\ &\geq \min\{dqr + c - 2m, dr/q + c\}. \end{aligned}$$

The second term is strictly greater than $dr - m$ as above, while the first term equals $dr - m$ plus $dr(q-1) + c - m$, and the latter is positive again by the choice of d. Thus $w_r(\pi^{-m}\mathbf{w}_0) > w_r(\mathbf{v}_0)$, and so $w_r(\mathbf{v}_{l+1} - \mathbf{v}_l) > w_r(\mathbf{v}_0)$; in particular, $w_r(\mathbf{v}) = w_r(\mathbf{v}_0)$ and so $\mathbf{v} \neq 0$.

Since the $\mathbf{v}_l$ converge to $\mathbf{v}$ in $M_{[r,r]}$, the $F\mathbf{v}_l$ converge to $F\mathbf{v}$ in $M_{[r/q,r/q]}$. On the other hand, $F\mathbf{v}_l = \mathbf{w}_l + \pi^m\mathbf{v}_l$, so the $F\mathbf{v}_l$ converge to $\pi^m\mathbf{v}$ in $M_{[r,r]}$. In particular, the $F\mathbf{v}_l$ form a Cauchy sequence under w_s for $s = r/q$ and $s = r$, hence also for all $s \in [r/q, r]$, and the limit in $M_{[r/q,r]}$ must equal both $F\mathbf{v}$ and $\pi^m\mathbf{v}$. Therefore $\mathbf{v} \in M_{[r/q,r]}$ and $F\mathbf{v} = \pi^m\mathbf{v}$ in $M_{[r/q,r/q]}$. But now by induction on i, $\mathbf{v} \in M_{[r/q^i,r]}$ for all i, so $\mathbf{v} \in M_{(0,r]} \subset M$ is a nonzero eigenvector, as desired. □

Lemma 4.2.3. *Suppose that v_K is nontrivial. Let M be a σ-module of rank n over $\mathcal{R}$.*

(a) There exists an integer c_0 such that for any integer $c \geq c_0$, there exists an injection $\mathcal{R}(c)^{\oplus n} \hookrightarrow M$.

(b) There exists an integer c_1 such that for any integer $c \leq c_1$, there exists an injection $M \hookrightarrow \mathcal{R}(c)^{\oplus n}$.

Proof. By taking duals, we may reduce (b) to (a). We prove (a) by induction on n, with empty base case $n = 0$. By Proposition 4.2.2, there exists an eigenvector of M; the saturated span of this eigenvector is a rank 1 σ-submodule of M, necessarily isomorphic to some $\mathcal{R}(m)$ by Proposition 3.3.2. By the induction hypothesis, we can choose $c_0 \geq m$ so that $\mathcal{R}(c_0)^{n-1}$ injects into $M/\mathcal{R}(m)$. Let N be the preimage of $\mathcal{R}(c_0)^{n-1}$ in M; then there exists an exact sequence

$$0 \to \mathcal{R}(m) \to N \to \mathcal{R}(c_0)^{n-1} \to 0,$$

which splits by Corollary 4.1.4. Thus $N \cong \mathcal{R}(m) \oplus \mathcal{R}(c_0)^{n-1} \subseteq M$, and $\mathcal{R}(m)$ contains a copy of $\mathcal{R}(c_0)$ by Corollary 4.1.4. This yields the desired result by Corollary 4.1.4 again. □

Proposition 4.2.4. *For any nonzero σ-module M over $\Gamma_{\mathrm{an,con}}$, the slopes of all nonzero σ-submodules of N are bounded below. Moreover, there is a nonzero σ-submodule of N of minimal slope, and any such σ-submodule is semistable.*

Proof. To check the first assertion, we may assume (by enlarging K as needed) that K is algebraically closed, so that $\Gamma_{\mathrm{an,con}} = \mathcal{R}$, and that v_K is nontrivial. By Lemma 4.2.3, there exists an injection $M \hookrightarrow \mathcal{R}(c)^{\oplus n}$ for some c, where $n = \operatorname{rank} M$. By Lemma 3.4.8, it follows that $\mu(N) \geq c$ for any σ-submodule N of M, yielding the first assertion.
As for the second assertion, the slopes of σ-submodules of M form a discrete subset of $\mathbb{Q}$, because their denominators are bounded above by n. Hence this set has a least element, yielding the remaining assertions. □

Proposition 4.2.5. *Every σ-module over $\Gamma_{\mathrm{an,con}}$ admits a Harder-Narasimhan filtration.*

Proof. Let M be a nontrivial σ-module over $\Gamma_{\mathrm{an,con}}$. By Proposition 4.2.4, the set of slopes of nonzero σ-submodules of M has a least element s_1. Suppose that N_1, N_2 are σ-submodules of M with $\mu(N_1) = \mu(N_2) = s_1$; then the internal sum $N_1 + N_2$ is a quotient of the direct sum $N_1 \oplus N_2$. By Proposition 4.2.4, each of N_1 and N_2 is semistable, as then is $N_1 \oplus N_2$. Hence $\mu(N_1 + N_2) \leq s_1$; by the minimality of s_1, we have $\mu(N_1 + N_2) = s_1$. Consequently, the set of σ-submodules of M of slope s_1 has a maximal element M_1. Repeating this argument with M replaced by M/M_1, and so on, yields a Harder-Narasimhan filtration. □

Remark 4.2.6. Running this argument is a severe obstacle to working with spherically complete coefficients (as suggested by Remark 2.3.10), as it is no longer clear that there exists a minimum slope among the σ-submodules of a given σ-module.

Remark 4.2.7. It may be possible to simplify the calculations in this section by using Lemma 6.1.1; we have not looked thoroughly into this possibility.

4.3 More eigenvectors

We now give a crucial refinement of the conclusion of Proposition 4.2.2 by extracting an eigenvector of a specific slope in a key situation. This is essentially [19, Proposition 4.15]; compare also [17, Proposition 9.1]. Beware that we are omitting one particularly unpleasant part of the calculation; see Lemma 4.3.3 below.

We start by identifying $H^0(\mathcal{R}(-1))$.

Lemma 4.3.1. *The map*

$$\overline{y} \mapsto \sum_{i\in\mathbb{Z}} [\overline{y}^{q^{-i}}]\pi^i$$

induces a bijection $\mathfrak{m}_K \to H^0(\mathcal{R}(-1))$, where $\mathfrak{m}_K$ denotes the subset of K on which v_K is positive.

Proof. On one hand, for $v_K(\overline{y}) > 0$, the sum $y = \sum_{i\in\mathbb{Z}}[\overline{y}^{q^{-i}}]\pi^i$ converges and satisfies $y^\sigma = \pi y$. Conversely, if $y^\sigma = \pi y$, comparing the Teichmüller presentations of y^σ and of πy forces y to assume the desired form. ∎

We next give a "positioning argument" for elements of $H^1(\mathcal{R}(m))$, following [19, Lemmas 4.13 and 4.14].

Lemma 4.3.2. *For m a positive integer, every nonzero element of $H^1(\mathcal{R}(m))$ is represented by some $x \in \Gamma^{\mathrm{alg}}_{\mathrm{con}}$ with $v_n(x) = v_{m-1}(x)$ for $n \geq m$. Moreover, we can ensure that for each $n \geq 0$, either $v_n(x) = \infty$ or $v_n(x) < 0$.*

Proof. We first verify that each element of $H^1(\mathcal{R}(m))$ is represented by an element of $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$. If $v_n(x) \geq 0$ for all $n \in \mathbb{Z}$, then the sum $y = \sum_{i=0}^\infty x^{\sigma^i}\pi^{mi}$ converges in $\mathcal{R}$ and satisfies $y - \pi^m y^\sigma = x$, so x represents the zero class in $H^1(\mathcal{R}(m))$. In other words, if $x \in \mathcal{R}$ has plus-minus-zero representation $x_+ + x_- + x_0$, then x and x_- represent the same class in $H^1(\mathcal{R}(m))$, and visibly $x_- \in \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$.

We next verify that each element of $H^1(\mathcal{R}(m))$ is represented by an element x of $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$ with $\inf_n\{v_n(x)\} > -\infty$. Given any $x \in \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$, let $x = \sum_i [\overline{x_i}]\pi^i$ be the Teichmüller representation of x. For each i, let c_i be the smallest nonnegative integer such that $q^{-c_i}v_K(\overline{x_i}) \geq -1$, and put

$$y_i = \sum_{j=1}^{c_i} [\overline{x_i}^{q^{-j}}]\pi^{i-mj};$$

since c_i grows only logarithmically in i, $i - mc_i \to \infty$ and the sum $\sum_i y_i$ converges π-adically. Moreover,

$$\limsup_{i\to\infty} \min_{1\leq j\leq c_i} \{-q^{-j}v_0(\overline{x_i})/(i-mj)\}$$

is finite, because the same is true for each of q^{-j} (clear), $-v_0(\overline{x_i})/i$ (by the definition of $\mathcal{R}$), and $i/(i-mj)$ (because c_i grows logarithmically in i). Hence the sum $y = \sum_i y_i$ actually converges in $\mathcal{R}$; if we set $x' = x - \pi^m y^\sigma + y$, then x and x' represent the same class in $H^1(\mathcal{R}(m))$. However,

$$x' = \sum_i [\overline{x_i}^{q^{-c_i}}]\pi^{i-mc_i}$$

satisfies $v_n(x') \geq -1$ for all n.

Since x and $\pi^m x^\sigma$ represent the same element of $H^1(\mathcal{R}(m))$, we can also say that each element of $H^1(\mathcal{R}(m))$ is represented by an element x of $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$ with $\inf_n\{v_n(x)\} > -\infty$. Given such an x, put $h = \inf_n\{v_n(x)\}$; we attempt to construct a sequence $x_0, x_1, \dots$ with the following properties:

(a) $x_0 = x$;

(b) each x_l generates the same element of $H^1(\mathcal{R}(m))$ as does x;

(c) $x_l \equiv 0 \pmod{\pi^{lm}}$;

(d) for each n and l, $v_n(x_l) \geq h$.

We do this as follows. Given x_l, let $\sum_{i=lm}^\infty [\overline{x_{l,i}}]\pi^i$ be the Teichmüller presentation of x_l, and put

$$u_l = \sum_{i=lm}^{lm+m-1} [\overline{x_{l,i}}]\pi^i.$$

If $v_{lm+m-1}(x_l) \geq h/q$, put $x_{l+1} = x_l - u_l + \pi^m u_l^\sigma$; otherwise, leave x_{l+1} undefined.

If x_l is defined for each l, then set $y = \sum_{l=0}^\infty u_l$; this sum converges in $\mathcal{R}$, and its limit satisfies $x - y + \pi^m y^\sigma = 0$. Hence in this case, x represents the trivial class in $H^1(\mathcal{R}(m))$.

On the other hand, if we are able to define x_l but not x_{l+1}, put

$$\begin{aligned} y &= x_l - u_l \\ z &= \pi^{-(l+1)m} y^{\sigma^{-l-1}} + \pi^{-lm} u_l^{\sigma^{-l}}; \end{aligned}$$

then x and z represent the same class in $H^1(\mathcal{R}(m))$. Moreover, $h/q > v_n(u)$ and $v_n(y) \geq h$ for all n, so $hq^{-l-1} > v_n(\pi^{-lm}u_l^{\sigma^{-l}})$ and $v_n(\pi^{-(l+1)m}y^{\sigma^{-l-1}}) \geq hq^{-l-1}$; consequently $v_n(z) = v_{m-1}(z)$ for $n \geq m$. In addition, we can pass from z to the minus part z_- of its plus-minus-zero representation (since z and z_- represent the same class in $H^1(\mathcal{R}(m))$, as noted above) to ensure that $v_n(z_-)$ is either infinite or negative for all $n \geq 0$. We conclude that every nonzero class in $H^1(\mathcal{R}(m))$ has a representative of the desired form. $\square$

Lemma 4.3.3. *Let d be a positive integer. For any $x \in H^1([d]^*(\mathcal{R}(d+1)))$, there exists $y \in H^0(\mathcal{R}(-1))$ nonzero such that x and y pair to zero in $H^1(\mathcal{R}(-1) \otimes [d]^*(\mathcal{R}(d+1))) = H^1([d]^*(\mathcal{R}(1)))$.*

Proof. Identify $H^1([d]^*(\mathcal{R}(d+1)))$ and $H^1([d]^*(\mathcal{R}(1)))$ with $H^1(\mathcal{R}(d+1))$ and $H^1(\mathcal{R}(1))$, respectively, where the modules in the latter cases are σ^d-modules. Then the question can be stated as follows: for any $x \in \mathcal{R}$, there exist $y, z \in \mathcal{R}$ with y nonzero, such that

$$y^\sigma = \pi y, \qquad xy = z - \pi z^{\sigma^d}.$$

By Lemma 4.3.2, we may assume that $x \in \Gamma_{\mathrm{con}}^{\mathrm{alg}}$ and that $v_n(x) = v_d(x) < 0$ for $n > d$. In this case, the claim follows from a rather involved calculation [19, Lemma 4.12] which we will not repeat here. (For a closely related calculation, see [17, Proposition 9.5].) □

Proposition 4.3.4. *Assume that v_K is nontrivial. For d a positive integer, suppose that*

$$0 \to M_{1,d} \to M \to M_{-1,1} \to 0$$

is a short exact sequence of σ-modules over $\mathcal{R}$. Then M contains an eigenvector of slope 0.

Proof. The short exact sequence corresponds to a class in

$$\begin{aligned}\mathrm{Ext}(M_{-1,1}, M_{1,d}) &\cong H^1(M_{1,1} \otimes M_{1,d}) \\ &\cong H^1(M_{d+1,d}) \\ &\cong H^1([d]^*(\mathcal{R}(d+1))).\end{aligned}$$

From the snake lemma, we obtain an exact sequence

$$H^0(M) \to H^0(M_{-1,1}) = H^0(\mathcal{R}(-1)) \to H^1(M_{1,d}) = H^1([d]^*\mathcal{R}(1)),$$

in which the second map (the connecting homomorphism) coincides with the pairing with the given class in $H^1([d]^*(\mathcal{R}(d+1)))$. By Lemma 4.3.3, this homomorphism is not injective; hence $H^0(M) \neq 0$, as desired. □

Corollary 4.3.5. *Assume that v_K is nontrivial. For d a positive integer, suppose that*

$$0 \to M_{1,1} \to M \to M_{-1,d} \to 0$$

is a short exact sequence of σ-modules over $\mathcal{R}$. Then $M^\vee$ contains an eigenvector of slope 0.

Proof. Dualize Proposition 4.3.4. □

Corollary 4.3.6. *Assume that v_K is nontrivial. For c, c', c'' integers with $c + c' \leq 2c''$, suppose that*

$$0 \to M_{c,1} \to M \to M_{c',1} \to 0$$

is a short exact sequence of σ-modules over $\mathcal{R}$. Then M contains an eigenvector of slope c''.

Proof. By twisting, we may reduce to the case $c'' = 0$. If $c \leq 0$, then already $M_{c,1}$ contains an eigenvector of slope 0 by Corollary 4.1.4, so we may assume $c \geq 0$. Since $c + c' \leq 0$, by Corollary 4.1.4, we can find a copy of $M_{-c,1}$ within $M_{c',1}$; taking the preimage of $M_{-c,1}$ within M allows us to reduce to the case $c' = -c$.

We treat the cases $c \geq 0$ by induction, with base case $c = 0$ already treated. If $c > 0$, then twisting yields an exact sequence

$$0 \to M_{c-1,1} \to M(-1) \to M_{-c-1,1} \to 0.$$

By Corollary 4.1.4, we can choose a submodule of $M_{-c-1,1}$ isomorphic to $M_{-c+1,1}$; let N be its inverse image in $M(-1)$. Applying the induction hypothesis to the sequence

$$0 \to M_{c-1,1} \to N \to M_{-c+1,1} \to 0$$

yields an eigenvector of N of slope 0, and hence an eigenvector of M of slope 1. Let P be the saturated span of that eigenvector; it is isomorphic to $M_{m,1}$ for some m by Proposition 3.3.2, and we must have $m \leq 1$ by Corollary 4.1.4. If $m \leq 0$, M has an eigenvector of slope 0, so suppose instead that $m = 1$. We then have an exact sequence

$$0 \to P \cong M_{1,1} \to M \to M/P \to 0$$

in which M/P, which has rank 1 and degree -1 (by Lemma 3.4.3), is isomorphic to $M_{-1,1}$ by Proposition 3.3.2. Applying Proposition 4.3.4 now yields the desired result. (Compare [17, Corollary 9.2].) □

4.4 Existence of standard submodules

We now run the induction setup of [17, Theorem 11.1] to produce standard submodules of a σ-module of small slope.

Definition 4.4.1. For a given integer $n \geq 1$, let (A_n), (B_n) denote the following statements about n.

(A_n) Let a be any positive integer, and let M be any σ^a-module over $\mathcal{R}$. If $\mathrm{rank}(M) \leq n$ and $\deg(M) \leq 0$, then M contains an eigenvector of slope 0.

(B_n) Let a be any positive integer, and let M be any σ^a-module over $\mathcal{R}$. If $\mathrm{rank}(M) \leq n$, then M contains a saturated σ^a-submodule which is standard of slope $\leq \mu(M)$.

Note that if v_K is nontrivial, both (A_1) and (B_1) hold thanks to Proposition 3.3.2.

Lemma 4.4.2. *Assume that v_K is nontrivial. For $n \geq 2$, if (A_{n-1}) and (B_{n-1}) hold, then (A_n) holds.*

Proof. It suffices to show that if M is a σ^a-module over $\mathcal{R}$ with $\operatorname{rank}(M) = n$ and $\deg(M) \leq 0$, then M contains an eigenvector of slope 0; after twisting, we may reduce to the case $1 - n \leq \deg(M) \leq 0$. Suppose on the contrary that no such eigenvector exists; by Corollary 4.1.4, M then contains no eigenvector of any nonpositive slope.

On the other hand, by Proposition 4.2.2, M contains an eigenvector; in particular, M contains a saturated σ^a-submodule of rank 1. By Proposition 3.3.2, we thus have an exact sequence

$$0 \to M_{c,1} \to M \to N \to 0$$

for some integer c, which by hypothesis must be positive.

Choose c as small as possible; then $\deg(N) = \deg(M) - c \leq 0$ by Lemma 3.4.3, so by (A_{n-1}), N contains an eigenvector of slope 0. That is, N contains a σ^a-submodule isomorphic to $M_{0,1}$; let M' be the preimage in M of that submodule. We then have an exact sequence

$$0 \to M_{c,1} \to M' \to M_{0,1} \to 0.$$

By Corollary 4.3.6, M' contains an eigenvector of slope $\lceil c/2 \rceil$. By the minimality of the choice of c, we must have $c \leq \lceil c/2 \rceil$, or $c = 1$.

Put $c' = \deg(M) - 1 = \deg(N)$, so that $c' < 0$. By (B_{n-1}), N contains a saturated σ^a-submodule P which is standard of slope $\leq c'/(n-1)$; let P' be the preimage in M of that σ^a-submodule. If $\operatorname{rank}(P) < n-1$, then $\deg(P') \leq 1 + \operatorname{rank}(P)(c'/(n-1)) < 1$; since $\deg(P')$ is an integer, we have $\deg(P') \leq 0$. By (A_{n-1}), P' contains an eigenvector of slope 0, contradicting our hypothesis. If $\operatorname{rank}(P) = n - 1$, we have an exact sequence

$$0 \to M_{1,1} \to P' \to P \cong M_{c'',n-1} \to 0$$

for some $c'' \leq -1$. By Corollary 4.1.4, there is a nonzero homomorphism $M_{-1,n-1} \to M_{c'',n-1}$; if its image has rank $< n-1$, then again (A_{n-1}) forces P' to contain an eigenvector of slope 0, contrary to assumption. Hence $M_{c'',n-1}$ contains a copy of $M_{-1,n-1}$; choose such a copy and let P'' be its inverse image in P'. By Corollary 4.3.5, $(P'')^\vee$ contains an eigenvector of slope 0, and hence a primitive eigenvector of slope at most 0; this eigenvector corresponds to a rank $n-1$ submodule of P'' of slope at most 0. By (A_{n-1}), P'' contains an eigenvector of slope 0, contradicting our hypothesis.

In any case, our hypothesis that M contains no eigenvector of slope 0 has been contradicted, yielding the desired result. □

Lemma 4.4.3. *Assume that v_K is nontrivial. For $n \geq 2$, if (A_n) and (B_{n-1}) hold, then (B_n) holds.*

Proof. Let M be a σ^a-module of rank n and degree c. Put $b = n/\gcd(n,c)$; then by (A_n) applied after twisting, $[b]_* M$ (which has rank n and degree bc,

by Proposition 3.4.4) has an eigenvector of slope bc/n. That is, M has a b-eigenvector $\mathbf{v}$ of slope c/n; this gives a nontrivial map $f : M_{bc/n,b} \to M$ sending a standard basis to $\mathbf{v}, F\mathbf{v}, \dots, F^{b-1}\mathbf{v}$. Let N be the saturated span of the image of f, and put $m = \operatorname{rank} N$. Then $\wedge^m N$ admits a b-eigenvector of slope cm/n, so by Corollary 4.1.4, the slope of N is at most c/n. If $m < n$, we may apply (B_{n-1}) to N to obtain the desired result.
Suppose instead that $m = n$, which also implies $b = n$ since necessarily $m \leq b \leq n$. Then the map f is injective, so its image has slope c/n. By Lemma 3.4.2, f must in fact be surjective; thus $M \cong M_{c,n}$, as desired. □

4.5 Dieudonné-Manin decompositions

Definition 4.5.1. A *Dieudonné-Manin decomposition* of a σ-module M over $\mathcal{R}$ is a direct sum decomposition $M = \oplus_{i=1}^m M_{c_i,d_i}$ of M into standard σ-submodules. The *slope multiset* of such a decomposition is the union of the multisets consisting of c_i/d_i with multiplicity d_i for $i = 1, \dots, m$.

Remark 4.5.2. If M admits a Dieudonné-Manin decomposition, then M admits a basis of n-eigenvectors for $n = (\operatorname{rank} M)!$; more precisely, any basis of $H^0([n]_* M)$ over the fixed field of σ^n gives a basis of $[n]_* M$ over $\mathcal{R}$. The slopes of these n-eigenvectors coincide with the slope multiset of the decomposition.

Proposition 4.5.3. *Assume that v_K is nontrivial. Then every σ-module M over $\mathcal{R}$ admits a Dieudonné-Manin decomposition.*

Proof. We first show that every semistable σ-module M over $\mathcal{R}$ is isomorphic to a direct sum of standard σ-submodules of slope $\mu(M)$. We see this by induction on $\operatorname{rank}(M)$; by Lemmas 4.4.2 and 4.4.3, we have (B_n) for all n, so M contains a saturated σ-submodule N which is standard of some slope $\leq \mu(M)$. Since M has been assumed semistable, we have $\mu(N) \geq \mu(M)$; hence $\mu(N) = \mu(M)$, and M/N is also semistable. By the induction hypothesis, M/N splits as a direct sum of standard σ-submodules of slope $\mu(M)$; then by Corollary 4.1.4, the exact sequence

$$0 \to N \to M \to M/N \to 0$$

splits. This yields the desired result.
In the general case, by Proposition 4.2.5, M has an HN-filtration $0 = M_0 \subset M_1 \subset \cdots \subset M_l = M$, with each successive quotient M_i/M_{i-1} semistable of slope s_i, and $s_1 < \cdots < s_l$. By the above, M_i/M_{i-1} admits a Dieudonné-Manin decomposition with all slopes s_i; the filtration then splits thanks to Corollary 4.1.4. Hence M admits a Dieudonné-Manin decomposition. □

We will use the case of v_K nontrivial to establish the existence of Dieudonné-Manin decompositions also when v_K is trivial.

Definition 4.5.4. For R a (commutative) ring, let $R((t^{\mathbb{Q}}))$ denote the Hahn-Mal'cev-Neumann algebra of generalized power series $\sum_{i\in\mathbb{Q}} c_i t^i$, where each

$c_i \in R$ and the set of $i \in \mathbb{Q}$ with $c_i \neq 0$ is well-ordered (has no infinite decreasing subsequence); these series form a ring under formal series multiplication, with a natural valuation v given by $v(\sum_i c_i t^i) = \min\{i : c_i \neq 0\}$. For R an algebraically closed field, $R((t^{\mathbb{Q}}))$ is also algebraically closed; see [31, Chapter 13] for this and other properties of these algebras.

Lemma 4.5.5. *Suppose that k is algebraically closed and that $K = k((t^{\mathbb{Q}}))$. Let M be a σ-module over $\mathcal{O}[\pi^{-1}]$ such that $M \otimes \mathcal{R}$ admits a basis of eigenvectors. Then any such basis is a basis of M.*

Proof. We may identify Γ^K with the π-adic completion of $\mathcal{O}((t^{\mathbb{Q}}))$. In so doing, elements of $\mathcal{R}$ can be viewed as formal sums $\sum_{i\in\mathbb{Q}} c_i t^i$ with $c_i \in \mathcal{O}[\pi^{-1}]$.
Suppose $\mathbf{v} \in M \otimes \mathcal{R}$ nonzero satisfies $F\mathbf{v} = \pi^m \mathbf{v}$. We can then formally write $\mathbf{v} = \sum_{i\in\mathbb{Q}} \mathbf{v}_i t^i$ with $\mathbf{v}_i \in M$, and then we have $F\mathbf{v}_i = \pi^m \mathbf{v}_{qi}$ for each i. If $\mathbf{v}_i \neq 0$ for some $i < 0$, we then have $\mathbf{v}_{q^l i} = \pi^{-lm} F^l \mathbf{v}_i$, but this violates the convergence condition defining $\mathcal{R}$. Hence $\mathbf{v}_i = 0$ for $i < 0$.
Let $\mathcal{R}^+$ be the subring of $\mathcal{R}$ consisting of series $\sum_i c_i t^i$ with $c_i = 0$ for $i < 0$. Now if $M \otimes \mathcal{R}$ admits a basis of eigenvectors, then we have just shown that each basis element belongs to $M \otimes \mathcal{R}^+$, and likewise for the dual basis of $M^\vee \otimes \mathcal{R}^+$. We can then reduce modulo the ideal of $\mathcal{R}^+$ consisting of series with constant coefficient zero, to produce a basis of eigenvectors of M. $\square$

Remark 4.5.6. Beware that in the proof of Lemma 4.5.5, there do exist nonzero eigenvectors in $M \otimes \mathcal{R}^+$ with constant coefficient zero; however, these eigenvectors cannot be part of a basis.

Theorem 4.5.7. *Let M be a σ-module over $\mathcal{R}$.*

(a) There exists a Dieudonné-Manin decomposition of M.

(b) For any Dieudonné-Manin decomposition $M = \oplus_{j=1}^m M_{c_j,d_j}$ of M, let $s_1 < \cdots < s_l$ be the distinct elements of the slope multiset of the decomposition. For $i = 1, \dots, l$, let M_i be the direct sum of M_{c_j,d_j} over all j for which $c_j/d_j \leq s_i$. Then the filtration $0 \subset M_1 \subset \cdots \subset M_l = M$ coincides with the HN-filtration of M.

(c) The slope multiset of any Dieudonné-Manin decomposition of M consists of the HN-slopes of M. In particular, the slope multiset does not depend on the choice of the decomposition.

Proof. (a) For v_K nontrivial, this is Proposition 4.5.3, so we need only treat the case of v_K trivial. Another way to say this is every σ-module M over $\mathcal{O}[\pi^{-1}]$ is isomorphic to a direct sum of standard σ-submodules.

Set notation as in Lemma 4.5.5. By Proposition 4.5.3, $M \otimes \mathcal{R}$ is isomorphic to a direct sum of standard σ-modules. That direct sum has the form $N \otimes \mathcal{R}$, where N is the corresponding direct sum of standard σ-modules over $\mathcal{O}[\pi^{-1}]$. The isomorphism $M \otimes \mathcal{R} \to N \otimes \mathcal{R}$ corresponds to an element of $H^0((M^\vee \otimes N) \otimes \mathcal{R})$, which extends to a basis of eigenvectors of $[n]_*(M^\vee \otimes N) \otimes \mathcal{R}$ for some n. By Lemma 4.5.5, this basis consists of elements of $[n]_*(M^\vee \otimes N)$; hence $M \cong N$, as desired.

(b) By Lemma 3.4.9, any standard σ-module is stable; hence a direct sum of standard σ-modules of a single slope is semistable. Thus the described filtration is indeed an HN-filtration.

(c) This follows from (b).

□

REMARK 4.5.8. The case of v_K trivial in Theorem 4.5.7(a) is precisely the standard Dieudonné-Manin classification of σ-modules over a complete discretely valued field with algebraically closed residue field. It is more commonly derived on its own, as in [14], [29], [18], or [19, Theorem 5.6].

COROLLARY 4.5.9. *For any coprime integers c, d with $d > 0$, $\mathrm{End}(M_{c,d})$ is a division algebra.*

Proof. Suppose $\phi \in \mathrm{End}(M_{c,d})$ is nonzero. Decompose $\mathrm{im}(\phi)$ according to Theorem 4.5.7; then each standard summand of $\mathrm{im}(\phi)$ must have slope $\leq c/d$ by Corollary 4.1.4. On the other hand, each summand is a σ-submodule of $M_{c,d}$, so must have slope $\geq c/d$ again by Corollary 4.1.4. Thus each standard summand of $\mathrm{im}(\phi)$ must have slope exactly c/d. In particular, there can be only one such summand, it must have rank d, and by Lemma 3.4.2, $\mathrm{im}(\phi) = M_{c,d}$. Hence ϕ is surjective; since ϕ is a linear map between free modules of the same finite rank, it is also injective. We conclude that $\mathrm{End}(M_{c,d})$ is indeed a division algebra, as desired. □

PROPOSITION 4.5.10. *A σ-module M over $\mathcal{R}$ is semistable (resp. stable) if and only if $M \cong M_{c,d}^{\oplus n}$ for some c, d, n (resp. $M \cong M_{c,d}$ for some c, d).*

Proof. This is an immediate corollary of Theorem 4.5.7. (Compare [17, Corollary 11.6].) □

REMARK 4.5.11. By Theorem 4.5.7, every σ-module over $\mathcal{R}$ decomposes as a direct sum of semistable σ-modules, i.e., the Harder-Narasimhan filtration splits. However, when v_K is nontrivial, this decomposition/splitting is not canonical, so it does not make sense to try to prove any descent results for such decompositions. (When v_K is trivial, the splitting is unique by virtue of Corollary 4.1.4.) Of course, the number and type of summands in a Dieudonné-Manin decomposition are unique, since they are determined by the HN-polygon; indeed, they constitute complete invariants for isomorphism of σ-modules over $\mathcal{R}$ (compare [17, Corollary 11.8]).

PROPOSITION 4.5.12. *Let M be a σ-module over $\mathcal{R}$, let M_1 be the first step in the Harder-Narasimhan filtration, and put $d = \mathrm{rank}(M_1)$. Then $\wedge^d M_1$ is the first step in the Harder-Narasimhan filtration of $\wedge^d M$.*

Proof. Decompose M according to Theorem 4.5.7, so that M_1 is the direct sum of the summands of minimum slope s_1. Take the d-th exterior power of this decomposition (i.e., apply the Künneth formula); by Lemma 4.1.2, the minimum slope among the new summands is ds_1, achieved only by $\wedge^d M_1$. $\square$

Remark 4.5.13. More generally, the first step in the Harder-Narasimhan filtration of $\wedge^i M$ is $\wedge^i M_j$, for the smallest j such that $\operatorname{rank}(M_j) \geq i$; the argument is similar.

Proposition 4.5.14. *Let M be a σ-module over $\mathcal{R}$, and let $M \cong \oplus_{i=1}^l M_{c_i,d_i}$ be a Dieudonné-Manin decomposition of M.*

- *(a) If v_K is nontrivial, then there exists a nonzero homomorphism $f : M_{c,d} \to M$ of σ-modules if and only if $c/d \geq \min_i\{c_i/d_i\}$, and there exists a nonzero homomorphism $f : M \to M_{c,d}$ of σ-modules if and only if $c/d \leq \max_i\{c_i/d_i\}$.*
- *(b) If v_K is trivial, then there exists a nonzero homomorphism $f : M_{c,d} \to M$ or $f : M \to M_{c,d}$ of σ-modules if and only if $c/d \in \{c_1/d_1, \dots, c_l/d_l\}$.*

Proof. Apply Corollary 4.1.4. $\square$

4.6 The calculus of slopes

Theorem 4.5.7 affords a number of consequences for the calculus of slopes.

Definition 4.6.1. Let M be a σ-module over $\Gamma_{\mathrm{an,con}}$. Define the *absolute HN-slopes* and *absolute HN-polygon* of M to be the HN-slopes and HN-polygon of $M \otimes \mathcal{R}$, and denote the latter by $P(M)$. We say M is *pure* (or *isoclinic*) of slope s if the absolute HN-slopes of M are all equal to s. By Proposition 4.5.10, M is isoclinic if and only if $M \otimes \mathcal{R}$ is semistable. We use the adjective *unit-root* to mean "isoclinic of slope 0".

Remark 4.6.2. We will show later (Theorem 6.4.1) that the HN-filtration of $M \otimes \mathcal{R}$ coincides with the base extension of the HN-filtration of M, which will mean that the absolute HN-slopes of M coincide with the HN-slopes of M.

Proposition 4.6.3. *Let M and M' be σ-modules over $\Gamma_{\mathrm{an,con}}$. Let $c_1, \dots, c_m$ and $c'_1, \dots, c'_n$ be the absolute HN-slopes of M and M', respectively.*

- *(a) The absolute HN-slopes of $M \oplus M'$ are $c_1, \dots, c_m, c'_1, \dots, c'_n$.*
- *(b) The absolute HN-slopes of $M \otimes M'$ are $c_i c'_j$ for $i = 1, \dots, m$, $j = 1, \dots, n$.*
- *(c) The absolute HN-slopes of $\wedge^d M$ are $c_{i_1} + \cdots + c_{i_d}$ for all $1 \leq i_1 < \cdots < i_d \leq m$.*
- *(d) The absolute HN-slopes of $[a]_* M$ are $ac_1, \dots, ac_m$.*
- *(e) The absolute HN-slopes of $M(b)$ are $c_1 + b, \dots, c_m + b$.*

Proof. There is no harm in tensoring up to $\mathcal{R}$, or in applying $[a]_*$ for some positive integer a. In particular, using Theorem 4.5.7, we may reduce to the case where M and M' admit bases of eigenvectors, whose slopes must be the c_i and the c'_j. Then we obtain bases of eigenvectors of $M \oplus M'$, $M \otimes M'$, $\wedge^d M$, $[a]_* M$, $M(b)$, and thus may read off the claims. □

Proposition 4.6.4. *Let M_1, M_2 be σ-modules over $\Gamma_{\mathrm{an,con}}$ such that each absolute HN-slope of M_1 is less than each absolute HN-slope of M_2. Then $\mathrm{Hom}(M_1, M_2) = 0$.*

Proof. Tensor up to $\mathcal{R}$, then apply Theorem 4.5.7 and Proposition 4.5.14. □

4.7 Splitting exact sequences

As we have seen already (e.g., in Proposition 4.3.4), a short exact sequence of σ-modules over $\Gamma_{\mathrm{an,con}}$ may or may not split. Whether or not it splits depends very much on the Newton polygons involved. For starters, we have the following.

Definition 4.7.1. Given Newton polygons $P_1, \dots, P_m$, define the *sum* $P_1 + \cdots + P_m$ of these polygons to be the Newton polygon whose slope multiset is the union of the slope multisets of $P_1, \dots, P_m$. Also, write $P_1 \geq P_2$ to mean that P_1 lies above P_2.

Proposition 4.7.2. *Let $0 \to M_1 \to M \to M_2 \to 0$ be an exact sequence of σ-modules over $\Gamma_{\mathrm{an,con}}$.*

(a) We have $P(M) \geq P(M_1) + P(M_2)$.

(b) We have $P(M) = P(M_1) + P(M_2)$ if and only if the exact sequence splits over $\mathcal{R}$.

Proof. For (a), note that from the HN-filtrations of $M_1 \otimes \mathcal{R}$ and $M_2 \otimes \mathcal{R}$, we obtain a semistable filtration of $M \otimes \mathcal{R}$ whose Newton polygon is $P(M_1) + P(M_2)$. The claim now follows from Proposition 3.5.4.

For (b), note that if the sequence splits, then $P(M) = P(M_1) + P(M_2)$ by Proposition 4.6.3. Conversely, suppose that $P(M) = P(M_1) + P(M_2)$; we prove by induction on rank that if $\Gamma_{\mathrm{an,con}} = \mathcal{R}$, then the exact sequence splits. Our base case is where M_1 and M_2 are standard. If $\mu(M_1) \leq \mu(M_2)$, then the exact sequence splits by Corollary 4.1.4, so assume that $\mu(M_1) > \mu(M_2)$. By Theorem 4.5.7, we have an isomorphism $M \cong M_1 \oplus M_2$, which gives a map $M_2 \to M$. By Corollary 4.5.9, the composition $M_2 \to M \to M_2$ is either zero or an isomorphism; in the former case, by exactness the image of $M_2 \to M$ must land in $M_1 \subseteq M$. But that violates Corollary 4.1.4, so the composition $M_2 \to M \to M_2$ is an isomorphism, and the exact sequence splits.

We next treat the case of M_1 nonstandard. Apply Theorem 4.5.7 to obtain a decomposition $M_1 \cong N \oplus N'$ with N standard. We have

$$\begin{aligned} P(M) &\geq P(N) + P(M/N) \qquad \text{[by (a)]} \\ &\geq P(N) + P(M_1/N) + P(M_2) \qquad \text{[by (a)]} \\ &= P(M_1) + P(M_2) \qquad \text{[because } N \text{ is a summand of } M_1\text{]} \\ &= P(M) \qquad \text{[by hypothesis].} \end{aligned}$$

Hence all of the inequalities must be equalities; in particular, $P(M/N) = P(M_1/N) + P(M_2)$. By the induction hypothesis, the exact sequence $0 \to M_1/N \to M/N \to M_2 \to 0$ splits; consequently, the exact sequence $0 \to N' \to M \to M/N' \to 0$ splits. But we have an exact sequence $0 \to N \to M/N' \to M_2 \to 0$ and as above, we have $P(M/N') = P(N) + P(M_2)$, so this sequence also splits by the induction hypothesis. This yields the claim.

To conclude, note that the case of M_2 nonstandard follows from the case of M_1 nonstandard by taking duals. Hence we have covered all cases. □

Corollary 4.7.3. *Let $0 \to M_1 \to M \to M_2 \to 0$ be an exact sequence of σ-modules over $\mathcal{R}$, such that every slope of M_1 is less than or equal to every slope of M_2. Then the exact sequence splits; in particular, the HN-multiset of M is the union of the HN-multiset of M_1 and M_2.*

Proof. With the assumption on the slopes, the filtration induced by the HN-filtrations of M_1 and M_2 becomes an HN-filtration after possibly removing one redundant step in the middle (in case the highest slope of M_1 coincides with the lowest slope of M_2). Thus its Newton polygon coincides with the HN-polygon, so Proposition 4.7.2 yields the claim. □

Corollary 4.7.4. *Let $0 = M_0 \subset M_1 \subset \cdots \subset M_l = M$ be a filtration of a σ-module M over $\mathcal{R}$ by saturated σ-submodules with isoclinic quotients. Suppose that the Newton polygon of the filtration coincides with the HN-polygon of M. Then the filtration splits.*

Remark 4.7.5. In certain contexts, one can obtain stronger splitting theorems; for instance, the key step in [20] is a splitting theorem for σ-modules with connection over Γ_{con} (in the notation of Section 2.3).

5 Generic and special slope filtrations

Given a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$, we have two paradigms for constructing slopes and HN-polygons: the "generic" paradigm, in which we pass to $\Gamma[\pi^{-1}]$ as if v_K were trivial, and the "special" paradigm, in which we pass to $\Gamma_{\mathrm{an,con}}$. (See Section 7.3 for an explanation of the use of these adjectives.) In this chapter, we compare these paradigms: our main results are that the special HN-polygon lies above the generic one (Proposition 5.5.1), and that when the two polygons coincide, one obtains a common HN-filtration over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$ (Theorem 5.5.2). This last result is a key tool for constructing slope filtrations in general.

Convention 5.0.1. We continue to retain notations as in Chapter 2. We again point out that when working over $\Gamma_{\mathrm{con}}[\pi^{-1}]$, the adjective "generic" will imply passage to $\Gamma[\pi^{-1}]$, while the adjective "special" will imply passage to $\Gamma_{\mathrm{an,con}}$. We also abbreviate such expressions as "generic absolute HN-slopes" to "generic HN-slopes". (Keep in mind that the modifier "absolute" will ultimately be rendered superfluous anyway by Theorem 6.4.1.)

5.1 Interlude: lattices

Besides descending subobjects, we will also have need to descend entire σ-modules; this matter is naturally discussed in terms of lattices.

Definition 5.1.1. Let $R \hookrightarrow S$ be an injection of domains, and let M be a finite locally free S-module. An *R-lattice* in M is an R-submodule N of M such that the induced map $N \otimes_R S \to M$ is a bijection. If M is a σ-module, an R-lattice in the category of σ-modules is a module-theoretic R-lattice which is stable under F.

The existence of a Γ-lattice for a σ-module defined over $\Gamma[\pi^{-1}]$ is closely tied to nonnegativity of the slopes.

Proposition 5.1.2. *Let M be a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ with nonnegative generic slopes. Then M contains an F-stable Γ_{con}-lattice N. Moreover, if the generic slopes of M are all zero, then N can be chosen so that $F : \sigma^* M \to M$ is an isomorphism.*

Proof. Put $M' = M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$; then by Theorem 4.5.7, we can write M' as a direct sum of standard submodules, whose slopes by hypothesis are non-negative. From this presentation, we immediately obtain a Γ^{alg}-lattice of M' (generated by standard basis vectors of the standard submodules); its intersection with M gives the desired lattice. □

Proposition 5.1.2 also has the following converse.

Proposition 5.1.3. *Let M be a σ-module over Γ. Then the generic HN-slopes of M are all nonnegative; moreover, they are all zero if and only if $F : \sigma^* M \to M$ is an isomorphism.*

Proof. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of M, and define the $n \times n$ matrix A over Γ by $F\mathbf{e}_j = \sum_i A_{ij}\mathbf{e}_i$. Suppose $\mathbf{v}$ is an eigenvector of M, and write $\mathbf{v} = \sum_i c_i \mathbf{e}_i$ and $F\mathbf{v} = \sum_i d_i \mathbf{e}_i$; then $\min_i\{w(d_i)\} - \min_i\{w(c_i)\}$ is the slope of $\mathbf{v}$. But $w(A_{ij}) \geq 0$ for all i, j, so $\min_i\{w(d_i)\} \geq \min_i\{w(c_i)\}$. This yields the first claim.

For the second claim, note on one hand that if $F : \sigma^* M \to M$ is an isomorphism, then $M^\vee$ is also a σ-module over Γ, and so the generic HN-slopes of both M and $M^\vee$ are nonnegative. Since these slopes are negatives of each other by Proposition 4.6.3, they must all be zero. On the other hand, if the generic HN-slopes of M are all zero, then by Theorem 4.5.7, $M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ admits a basis

$\mathbf{e}_1, \dots, \mathbf{e}_n$ of eigenvectors of slope 0. Put $M' = M \otimes \Gamma^{\mathrm{alg}}$; for $i = 0, \dots, n$, let M'_i be the intersection of M' with the $\Gamma^{\mathrm{alg}}[\pi^{-1}]$-span of $\mathbf{e}_1, \dots, \mathbf{e}_i$. Then each M'_i is F-stable; moreover, M'_i/M'_{i-1} is spanned by the image of $\pi^{c_i}\mathbf{e}_i$ for some c_i, and so $F : \sigma^*(M'_i/M'_{i-1}) \to (M'_i/M'_{i-1})$ is an isomorphism for each i. It follows that $F : \sigma^* M' \to M'$ is an isomorphism, as then is $F : \sigma^* M \to M$. $\square$

Remark 5.1.4. The results in this section can also be proved using cyclic vectors, as in [19, Proposition 5.8]; compare Lemma 5.2.4 below.

5.2 The generic HN-filtration

Since the distinction between v_K trivial and nontrivial was not pronounced in the previous chapter, it is worth taking time out to clarify some phenomena specific to the "generic" (v_K trivial) setting.

Proposition 5.2.1. *For any σ-module M over $\Gamma^{\mathrm{alg}}[\pi^{-1}]$, there is a unique decomposition $M = P_1 \oplus \cdots \oplus P_l$, where each P_i is isoclinic, and the generic slopes $\mu(P_1), \dots, \mu(P_l)$ are all distinct.*

Proof. The existence of such a decomposition follows from Theorem 4.5.7; the uniqueness follows from repeated application of Corollary 4.1.4. $\square$

Definition 5.2.2. Let M be a σ-module over $\Gamma^{\mathrm{alg}}[\pi^{-1}]$. Define the *slope decomposition* of M to be the decomposition $M = P_1 \oplus \cdots \oplus P_l$ given by Proposition 5.2.1.

For the rest of this section, we catalog some routine methods for identifying the generic slopes of a σ-module.

Definition 5.2.3. Let M be a σ-module over $\Gamma[\pi^{-1}]$ of rank n. A *cyclic vector* of M is an element $\mathbf{v} \in M$ such that $\mathbf{v}, F\mathbf{v}, \cdots, F^{n-1}\mathbf{v}$ form a basis of M.

Lemma 5.2.4. *Let M be a σ-module over $\Gamma[\pi^{-1}]$ of rank n. Let $\mathbf{v}$ be a cyclic vector of M, and define $a_0, \dots, a_{n-1} \in \Gamma[\pi^{-1}]$ by the equation*

$$F^n\mathbf{v} + a_{n-1}F^{n-1}\mathbf{v} + \cdots + a_0\mathbf{v} = 0.$$

Then the generic HN-polygon of M coincides with the Newton polygon of the polynomial $x^n + a_{n-1}x^{n-1} + \cdots + a_0$.

Proof. There is no harm in assuming that K is algebraically closed, that π is fixed by σ, or that $\mathcal{O}$ is large enough that the slopes of M are all integers. Then M admits a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ with $F\mathbf{e}_i = \pi^{c_i}\mathbf{e}_i$ for some integers c_i.
Put $\mathbf{v}_0 = \mathbf{v}$. Given $\mathbf{v}_l$, write $\mathbf{v}_l = x_{l,1}\mathbf{e}_1 + \cdots + x_{l,n}\mathbf{e}_n$, put $b_l = \pi^{c_l}x_{l,l}^\sigma/x_{l,l}$, and put $\mathbf{v}_{l+1} = F\mathbf{v}_l - b_l\mathbf{v}_l$. Then $\mathbf{v}_l$ lies in the span of $\mathbf{e}_{l+1}, \dots, \mathbf{e}_n$; in particular, $\mathbf{v}_n = 0$.
We then have

$$(F - b_{n-1}) \cdots (F - b_1)(F - b_0)\mathbf{v} = 0;$$

since $\mathbf{v}$ is a cyclic vector, there is a unique way to write $F^n\mathbf{v}$ as a linear combination of $\mathbf{v}, F\mathbf{v}, \cdots, F^{n-1}\mathbf{v}$. Hence we have an equality of operators

$$(F - b_{n-1}) \cdots (F - b_1)(F - b_0) = F^n + a_{n-1}F^{n-1} + \cdots + a_0,$$

from which the equality of polygons may be read off directly. $\square$

Remark 5.2.5. One can turn Lemma 5.2.4 around and use it to prove the existence of Dieudonné-Manin decompositions in the case of v_K trivial; for instance, this is the approach in [19, Theorem 5.6]. One of the essential difficulties in [19] is that there is no analogous way to "read off" the HN-polygon of a σ-module over $\Gamma_{\mathrm{an,con}}$; this forces the approach to constructing the slope filtration over $\Gamma_{\mathrm{an,con}}$ to be somewhat indirect.

Lemma 5.2.4 is sometimes inconvenient to apply, because the calculus of cyclic vectors is quite "nonlinear". The following criterion will prove to be more useful for our purposes.

Lemma 5.2.6. *Let M be a σ-module over $\Gamma^{\mathrm{alg}}[\pi^{-1}]$. Suppose that there exists a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of M with the property that the matrix A given by $F\mathbf{e}_j = \sum_i A_{ij}\mathbf{e}_i$ satisfies $w(AD^{-1} - I_n) > 0$ for some $n \times n$ diagonal matrix D over $\Gamma^{\mathrm{alg}}[\pi^{-1}]$. Then the generic slopes of M are equal to the valuations of the diagonal entries of D. Moreover, there exists an invertible matrix U over Γ^{alg} with $w(U - I_n) > 0$, $w(DU^\sigma D^{-1} - I_n) > 0$, and $U^{-1}AU^\sigma = D$.*

Proof. One can directly solve for U; see [19, Proposition 5.9] for this calculation. Note that it does not matter whether the D^{-1} appears to the left or to the right of A, as a change of basis will flip it over to the other side; the entries of D will get hit by σ or its inverse, but their valuations will not change. Alternatively, the existence of U also follows from Proposition 5.4.5 below. $\square$

Remark 5.2.7. Lemmas 5.2.4 and 5.2.6 suggest that one can read off the generic HN-polygon of a σ-module over $\Gamma[\pi^{-1}]$ by computing the slopes of eigenvectors of a matrix via which F acts on some basis. This does not work in general, as observed by Katz [18].

5.3 Descending the generic HN-filtration

In the generic setting (v_K trivial), we have the following descent property for Harder-Narasimhan filtrations.

Proposition 5.3.1. *Let M be a σ-module over $\Gamma[\pi^{-1}]$. Then the Harder-Narasimhan filtration of M, tensored up to $\Gamma^{\mathrm{alg}}[\pi^{-1}]$, gives the Harder-Narasimhan filtration of $M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$.*

Proof. One can prove this by Galois descent, as in [19, Proposition 5.10]; here is an alternate argument. It suffices to check that the first step M_1' of the Harder-Narasimhan filtration of $M' = M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ descends to $\Gamma[\pi^{-1}]$; by

Lemma 3.6.2 and Proposition 4.5.12, we may by taking exterior powers reduce to the case where M_1' has rank 1. In particular, the least slope s_1 must be an integer. By twisting, we may assume that $s_1 = 0$.

By Proposition 5.1.2, we can find a σ-stable Γ-lattice N of M; put $N' = N \otimes \Gamma^{\mathrm{alg}}$. Then $N' \cap M_1'$ may be characterized as the set of limit points, for the π-adic topology, of sequences of the form $\{F^l \mathbf{v}_l\}_{l=0}^{\infty}$ with $\mathbf{v}_l \in N'$ for each l. (This may be verified on a basis of d-eigenvectors for appropriate d thanks to Theorem 4.5.7, where it is evident.)

The characterization of $N' \cap M_1'$ we just gave is linear, so it cuts out a rank one submodule of N already over Γ. This yields the desired result. $\square$

5.4 de Jong's reverse filtration

We now consider the case of σ-modules over $\Gamma_{\mathrm{con}}[\pi^{-1}]$, in which case we have a "generic" HN-filtration defined over $\Gamma^{\mathrm{alg}}[\pi^{-1}]$, and a "special" HN-filtration defined over $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$. These two filtrations are not directly comparable, because they live over incompatible overrings of $\Gamma_{\mathrm{con}}[\pi^{-1}]$. To compare them, we must use a "reverse filtration" that meets both halfway; the construction is due to de Jong [13, Proposition 5.8], but our presentation follows [19, Proposition 5.11] (wherein it is called the "descending generic filtration").

We first need a lemma that descends some eigenvectors from Γ^{alg} to $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$; besides de Jong's [13, Proposition 5.8], this generalizes a lemma of Tsuzuki [34, 4.1.1].

Lemma 5.4.1. *Let M be a σ-module over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$ all of whose generic slopes are nonpositive. Let $\mathbf{v} \in M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ be an eigenvector of slope 0. Then $\mathbf{v} \in M$.*

Proof. By Proposition 5.1.2, we can find an F-stable $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$-lattice $N^\vee$ of $M^\vee$; the dual lattice N is an F^{-1}-stable $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$-lattice of M. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of N, define the $n \times n$ matrix A over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$ by the equation $F^{-1}\mathbf{e}_j = \sum_i A_{ij} \mathbf{e}_i$, and choose $r > 0$ such that A is invertible over Γ^{alg}_r.

Let $\mathbf{v} \in M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ be an eigenvector of slope 0; in showing that $\mathbf{v} \in M$, there is no harm in assuming (by multiplying by a power of π as needed) that $\mathbf{v} \in N \otimes \Gamma^{\mathrm{alg}}$. Write $\mathbf{v} = \sum x_i \mathbf{e}_i$ with $x_i \in \Gamma^{\mathrm{alg}}$. Then for each $l \geq 0$, we have

$$\begin{aligned}\min_i \min_{m \leq l} \{v_{m,r}(x_i)\} &\geq w_r(A) + \min_i \min_{m \leq l} \{v_{m,r}(x_i^{\sigma^{-1}})\} \\ &\geq w_r(A) + q^{-1} \min_i \min_{m \leq l} \{v_{m,r}(x_i)\}.\end{aligned}$$

It follows that

$$\min_i \min_{m \leq l} \{v_{m,r}(x_i)\} \geq q w_r(A)/(q-1),$$

and so $x_i \in \Gamma^{\mathrm{alg}}_{\mathrm{con}}$. Hence $\mathbf{v} \in M$, as desired. $\square$

We now proceed to construct the reverse filtration.

Definition 5.4.2. Let M be a σ-module over $\Gamma^{\mathrm{alg}}[\pi^{-1}]$ with slope decomposition $P_1 \oplus \cdots \oplus P_l$, labeled so that $\mu(P_1) > \cdots > \mu(P_l)$. Define the *reverse filtration* of M as the semistable filtration $0 = M_0 \subset M_1 \subset \cdots \subset M_l = M$ with $M_i = P_1 \oplus \cdots \oplus P_i$ for $i = 1, \dots, l$. By construction, its Newton polygon coincides with the generic Newton polygon of M.

Proposition 5.4.3. *Let M be a σ-module over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$. Then the reverse filtration of $M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ descends to $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$.*

Proof. Put $M' = M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$, and let $0 = M'_0 \subset M'_1 \subset \cdots \subset M'_l = M'$ be the reverse filtration of M'. It suffices to show that M'_1 descends to $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$; by Lemma 3.6.2, we may reduce to the case where $\operatorname{rank} M'_1 = 1$ by passing from M to an exterior power. By twisting, we may then reduce to the case $\mu(M'_1) = 0$. By Proposition 3.3.2, M'_1 is then generated by an eigenvector of slope 0; by Lemma 5.4.1, that eigenvector belongs to M. Hence M'_1 descends to M, proving the claim. (Compare [19, Proposition 5.11].) $\square$

Remark 5.4.4. The reverse filtration actually descends all the way to $\Gamma_{\mathrm{con}}[\pi^{-1}]$ whenever K is perfect, but we will not need this.

It may also be useful for some applications to have a quantitative version of Proposition 5.4.3.

Proposition 5.4.5. *Let M be a σ-module over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$. Suppose that for some $r > 0$, there exists a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of M with the property that the $n \times n$ matrix A given by $F\mathbf{e}_j = \sum_i A_{ij}\mathbf{e}_i$ has entries in $\Gamma^{\mathrm{alg}}_r[\pi^{-1}]$. Suppose further that $w(AD^{-1} - I_n) > 0$ and $w_r(AD^{-1} - I_n) > 0$ for some $n \times n$ diagonal matrix D over $\mathcal{O}[\pi^{-1}]$, with $w(D_{11}) \geq \cdots \geq w(D_{nn})$. Then there exists an invertible $n \times n$ matrix U over $\Gamma^{\mathrm{alg}}_{qr}$ with $w(U - I_n) > 0$, $w_r(U - I_n) > 0$, $w(DU^\sigma D^{-1} - I_n) > 0$, $w_r(DU^\sigma D^{-1} - I_n) > 0$, such that $U^{-1}AU^\sigma D^{-1} - I_n$ is upper triangular nilpotent.*

Proof. Put $c_0 = \min\{w(AD^{-1} - I_n), w_r(AD^{-1} - I_n)\}$ and $U_0 = I_n$. Given U_l, put $A_l = U_l^{-1}AU_l^\sigma$, and write $A_lD^{-1} - I_n = B_l + C_l$ with B_l upper triangular nilpotent and C_l lower triangular. Suppose that $\min\{w(A_lD^{-1} - I_n), w_r(A_lD^{-1} - I_n)\} \geq c_0$; put $c_l = \min\{w(C_l), w_r(C_l)\}$. Choose a matrix X_l over Γ_{qr} with

$$\begin{gathered}
C_l + X_l = DX_l^\sigma D^{-1}, \\
\min\{w(X_l), w(DX_l^\sigma D^{-1})\} \geq w(C_l), \\
\min\{w_r(X_l), w_r(DX_l^\sigma D^{-1})\} \geq c_l.
\end{gathered}$$

(This amounts to solving a system of equations of the form $c + x = \lambda^{-1}x^\sigma$ for $\lambda \in \mathcal{O}$; the analysis is as in Proposition 3.3.7.)
Put $U_{l+1} = U_l(I_n + X_l)$. We then have

$$A_{l+1}D^{-1} = (I_n - X_l + X_l^2 - \cdots)(I_n + B_l + C_l)(I_n + DX_l^\sigma D^{-1}),$$

whence $w(C_{l+1}) \geq w(C_l) + c_0$ and $w_r(C_{l+1}) \geq c_l + c_0$. Consequently $c_l \geq (l+1)c_0$, so the U_l converge under w_{qr} to a limit U such that $U^{-1}AU^{\sigma}D^{-1} - I_n$ is upper triangular nilpotent, as desired. □

Remark 5.4.6. For more on de Jong's original application of the reverse filtration, see Section 7.5.

5.5 Comparison of polygons

Using the reverse filtration, we obtain the fundamental comparison between the generic and special Newton polygons of a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$.

Proposition 5.5.1. *Let M be a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$. Then the special HN-polygon of M (i.e., the HN-polygon of $M \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$) lies above the generic HN-polygon (i.e., the HN-polygon of $M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$).*

Proof. Tensor up to $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$, then apply Proposition 3.5.4 to the reverse filtration (Proposition 5.4.3). □

The case where the polygons coincide is especially pleasant, and will be crucial to our later results.

Theorem 5.5.2. *Let M be a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ whose generic and special HN-polygons coincide. Then the generic and special HN-filtrations of $M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ and $M \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$, respectively, are both obtained by base change from an exhaustive filtration of M.*

Proof. It suffices to check that the first steps of the generic and special HN-filtrations descend and coincide; by Lemma 3.6.2, we may reduce to the case where the least slope of the common polygon occurs with multiplicity 1. Choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of M, let $\mathbf{v} \in M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ be a generator of the first step of the HN-filtration of $M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$, and write $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. By Proposition 5.3.1, $a_i/a_j \in \Gamma[\pi^{-1}]$ for any i, j with $a_j \neq 0$.
By Corollary 4.7.4, the reverse filtration splits over $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$; by Proposition 3.3.7(b1), it also splits over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$. Hence $a_i/a_j \in \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$ for any i, j with $a_j \neq 0$. Since $\Gamma[\pi^{-1}] \cap \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}] = \Gamma_{\mathrm{con}}[\pi^{-1}]$, we have $a_i/a_j \in \Gamma_{\mathrm{con}}[\pi^{-1}]$ for any i, j with $a_j \neq 0$. Hence the first step of the generic HN-filtration descends to $\Gamma_{\mathrm{con}}[\pi^{-1}]$; let M_1 be the corresponding σ-submodule of $M \otimes \Gamma_{\mathrm{con}}[\pi^{-1}]$. Let M_1' be the first step of the HN-filtration of $M \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$; then as in the proof of Proposition 4.2.5, M_1' is the maximal σ-submodule of $M \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$ of slope $\mu(M_1') = \mu(M_1)$. In particular, $M_1 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}} \subseteq M_1'$, and by Lemma 3.4.2, $M_1 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$ and M_1' actually coincide. Hence the first step of the special HN-filtration is also a base extension of M_1. This yields the claim. (Compare [19, Proposition 5.16].) □

Remark 5.5.3. The conditions of Theorem 5.5.2 may look restrictive, and indeed they are: many "natural" examples of σ-modules over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ do not have the same special and generic HN-polygons (e.g., the example of Section 7.3). However, Theorem 5.5.2 represents a critical step in the descent process for slope filtrations, as it allows us to move information from the generic paradigm into the special paradigm: specifically, the descent from algebraically closed K down to general K is much easier in the generic setting. Of course, in order to use this link, we must be able to force ourselves into the setting of Theorem 5.5.2; this is done in the next chapter.

Note that Theorem 5.5.2 implies that the generic and special HN-polygons of M coincide if and only if the generic HN-filtration descends to $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$. In some applications, it may be more useful to have a quantitative refinement of that statement; we can give one (in imitation of the proof of Proposition 4.3.4) as follows.

Proposition 5.5.4. *Let M be a σ-module over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$. Suppose that for some $r > 0$, there exists a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of M with the property that the $n \times n$ matrix A given by $F\mathbf{e}_j = \sum_i A_{ij}\mathbf{e}_i$ has entries in $\Gamma^{\mathrm{alg}}_r[\pi^{-1}]$. Suppose further that $w(AD^{-1} - I_n) > 0$ and $w_r(AD^{-1} - I_n) > 0$ for some $n \times n$ diagonal matrix D over $\mathcal{O}[\pi^{-1}]$, with $w(D_{11}) \geq \cdots \geq w(D_{nn})$. Let U be an invertible $n \times n$ matrix over Γ^{alg} such that $w(U - I_n) > 0$, $w(DU^\sigma D^{-1} - I_n) > 0$, and $U^{-1}AU^\sigma = D$ (as in Lemma 5.2.6). Then the generic and special HN-polygons of M coincide if and only if*

$$w_r(U - I_n) > 0, \qquad w_r(DU^\sigma D^{-1} - I_n) > 0. \tag{5.5.5}$$

Proof. The conditions on U imply that M admits a basis of eigenvectors over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$; the slopes of these eigenvectors then give both the generic and special HN-slopes. Conversely, if the generic and special HN-polygons of M coincide, then U is invertible over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$ by Theorem 5.5.2. It thus remains to prove that if U has entries in $\Gamma^{\mathrm{alg}}_{\mathrm{con}}$, then in fact (5.5.5) holds.
We start with a series of reductions. By Proposition 5.4.5, we may reduce to the case where $AD^{-1} - I_n$ is upper triangular nilpotent. Considering all matrices now as block matrices by grouping rows and columns where the diagonal entries of D have the same valuation, we see that the matrix U must now be block upper triangular, with diagonal blocks invertible over $\mathcal{O}$. Neither this property nor (5.5.5) is disturbed by multiplying U by a block diagonal matrix over $\mathcal{O}$, so we may reduce to the case where U is block upper triangular with identity matrices on the diagonal. Finally, note that at this point it suffices to check the case where $n = 2$, $w(D_{11}) > w(D_{22})$, and

$$AD^{-1} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix},$$

as the general case follows by repeated application of this case.
Put $\lambda \in D_{11}D_{22}^{-1} \in \pi\mathcal{O}$. Then what we are to show is that given $a \in \Gamma^{\mathrm{alg}}_r$, $u \in \Gamma^{\mathrm{alg}}_{\mathrm{con}}$ with

$$w_r(a) > 0, \qquad w(u) > 0, \qquad \lambda a + u = \lambda u^\sigma, \tag{5.5.6}$$

we must have $u \in \Gamma_r^{\mathrm{alg}}$ and $w_r(u) > 0$. Before verifying this, we make one further simplifying reduction, using the fact that there is no harm in replacing a by $a - b + \lambda^{-1} b^{\sigma^{-1}}$ as long as $w_r(b) > 0$ and $w_r(\lambda^{-1} b^{\sigma^{-1}}) > 0$.
Note that for $b = \pi^m [\overline{x}]$ with $\overline{x} \in K^{\mathrm{alg}}$, the condition that

$$w_r(b) \leq w_r(\lambda^{-1} b^{\sigma^{-1}})$$

is equivalent to

$$v_K(\overline{x}) \leq -\frac{q}{(q-1)r} w(\lambda).$$

Moreover, this condition and the bound $w_r(b) \geq 0$ together imply that $w(\lambda^{-1} b^{\sigma^{-1}}) > 0$.
Now write $a = \sum_{i=0}^{\infty} \pi^i [\overline{a_i}]$, and put $c = qw(\lambda)/(q-1)r$. For each i, let j_i be the smallest nonnegative integer such that $q^{-j_i} v_K(\overline{a_i}) > -c$. We may then replace a by

$$a' = \sum_{i=0}^{\infty} \lambda^{-1} \lambda^{-\sigma^{-1}} \cdots \lambda^{-\sigma^{j_i - 1}} (\pi^i [\overline{a_i}])^{\sigma^{-j_i}}$$

without disturbing the truth of (5.5.6). In particular, we may reduce to the case where $v_n(a) > -c$ for all $n \geq 0$.
Under these conditions, put $m = w(\lambda) > 0$, and note that if $v_n(u) \leq -c/q$ for some u, then $v_{n+m}(\lambda u^\sigma) = q v_n(u) \leq -c$, so the equation $\lambda a + u = \lambda u^\sigma$ implies $v_{n+m}(u) = q v_n(u)$. By induction on l, we have

$$v_{n+lm}(u) = q^l v_n(u)$$

for all nonnegative integers l, but this contradicts the hypothesis that $u \in \Gamma_{\mathrm{con}}^{\mathrm{alg}}$. Consequently, we must have $v_n(u) > -c/q$ for all n.
Since $-c/q \geq -c(q-1)/q = -w(\lambda)/r$, the bound $v_n(u) > -c/q$ implies that $v_{n,r}(u) > 0$ for $n \geq m$. Since $u \equiv 0 \pmod{\lambda}$, we have $w_r(u) > 0$, as desired. $\square$

6 Descents

We now show that the formation of the Harder-Narasimhan filtration commutes with base change, thus establishing the slope filtration theorem; the strategy is to show that a σ-module over $\Gamma_{\mathrm{an,con}}$ admits a model over Γ_{con} whose special and generic Newton polygons coincide, then invoke Theorem 5.5.2. The material here is derived from [19, Chapter 6], but our presentation here is much more streamlined.

6.1 A matrix lemma

The following lemma is analogous to [19, Proposition 6.8], but in our new approach, we can prove much less and still eventually get the desired conclusion; this simplifies the matrix calculation considerably.

Lemma 6.1.1. *For $r > 0$, suppose that $\Gamma = \Gamma^K$ contains enough r_0-units for some $r_0 > qr$. Let D be an invertible $n \times n$ matrix over $\Gamma_{[r,r]}$, and put $h = -w_r(D) - w_r(D^{-1})$. Let A be an $n \times n$ matrix over $\Gamma_{[r,r]}$ such that $w_r(AD^{-1} - I_n) > h/(q-1)$. Then there exists an invertible $n \times n$ matrix U over $\Gamma_{[r,qr]}$ such that $U^{-1}AU^\sigma D^{-1} - I_n$ has entries in $\pi\Gamma_r$ and $w_r(U^{-1}AU^\sigma D^{-1} - I_n) > 0$.*

Proof. Put $c_0 = w_r(AD^{-1} - I_n) - h/(q-1)$. Define sequences of matrices $U_0, U_1, \dots$ and $A_0, A_1, \dots$ as follows. Start with $U_0 = I_n$. Given an invertible $n \times n$ matrix U_l over $\Gamma_{[r,qr]}$, put $A_l = U_l^{-1} A U_l^\sigma$. Suppose that $w_r(A_l D^{-1} - I_n) \geq c_0 + h/(q-1)$; put

$$c_l = \min_{m \leq 0} \{v_{m,r}(A_l D^{-1} - I_n)\} - h/(q-1),$$

so that $c_l \geq c_0 > 0$.
Let $\sum u_{ijlm}\pi^m$ be a semiunit presentation of $(A_l D^{-1} - I_n)_{ij}$. Let X_l be the $n \times n$ matrix with $(X_l)_{ij} = \sum_{m \leq 0} u_{ijlm}\pi^m$, and put $U_{l+1} = U_l(I_n + X_l)$. By Lemma 2.5.3, for $m \leq 0$ and $s \in [r, qr]$,

$$\begin{aligned} w_s(u_{ijlm}\pi^m) &\geq (s/r) w_r(u_{ijlm}\pi^m) \\ &\geq (s/r) \min_{k \leq m}\{v_{k,r}(A_l D^{-1} - I_n)\} \\ &\geq (s/r)(c_l + h/(q-1)); \end{aligned}$$

hence U_{l+1} is also invertible over $\Gamma_{[r,qr]}$. Moreover,

$$\begin{aligned} w_r(DX_l^\sigma D^{-1}) &\geq w_r(D) + w_r(X_l^\sigma) + w_r(D^{-1}) \\ &= w_{qr}(X_l) - h \\ &\geq q(c_l + h/(q-1)) - h \\ &\geq qc_l + h/(q-1). \end{aligned}$$

Since

$$A_{l+1}D^{-1} = (I_n + X_l)^{-1}(A_l D^{-1})(I_n + DX_l^\sigma D^{-1}),$$

we then have $w_r(A_{l+1}D^{-1} - I_n) \geq c_0 + h/(q-1)$, so the iteration may continue. We now prove by induction that $c_l \geq (l+1)c_0$ for $l = 0, 1, \dots$; this is clearly true for $l = 0$. Given the claim for l, write

$$\begin{aligned} &(I_n + X_l)^{-1}A_l D^{-1} = \\ &\qquad I_n + (A_l D^{-1} - I_n - X_l) - X_l(A_l D^{-1} - I_n) + X_l^2(I_n + X_l)^{-1}A_l D^{-1}. \end{aligned}$$

Note that

$$\begin{aligned} v_m(A_l D^{-1} - I_n - X_l) &= \infty \qquad (m \leq 0) \\ w_r(X_l(A_l D^{-1} - I)) &\geq (l+2)c_0 + 2h/(q-1) \\ w_r(X_l^2(I_n + X_l)^{-1}A_l D^{-1}) &\geq 2(l+1)c_0 + 2h/(q-1). \end{aligned}$$

Putting this together, this means

$$v_{m,r}((I_n + X_l)^{-1} A_l D^{-1} - I_n) \geq (l+2)c_0 + h/(q-1) \qquad (m \leq 0).$$

Since $w_r(DX_l^\sigma D^{-1}) \geq qc_l + h/(q-1) \geq (l+2)c_0 + h/(q-1)$, we have $v_{m,r}(A_{l+1}D^{-1} - I_n) \geq (l+2)c_0 + h/(q-1)$ for $m \leq 0$, i.e., $c_{l+1} \geq (l+2)c_0$. Since $w_s(X_l) \geq (s/r)(c_l + h/(q-1))$ for $s \in [r, qr]$, and $c_l \to \infty$ as $l \to \infty$, the sequence $\{U_l\}$ converges to a limit U, which is an invertible $n \times n$ matrix over $\Gamma_{[r,qr]}$ satisfying $w_r(U^{-1}AU^\sigma D^{-1} - I_n) \geq c_0 + h/(q-1) > 0$. Moreover, by construction, we have $v_m(U^{-1}AU^\sigma D^{-1} - I_n) = \infty$ for $m \leq 0$; by Corollary 2.5.6, $U^{-1}AU^\sigma D^{-1} - I_n$ has entries in $\pi\Gamma_r$, as desired. □

6.2 Good models of σ-modules

We now give a highly streamlined version of some arguments from [19, Chapter 6], that produce "good integral models" of σ-modules over $\Gamma_{\mathrm{an,con}}$.

Lemma 6.2.1. *In Lemma 6.1.1, suppose that A and D are both invertible over $\Gamma_{\mathrm{an},r}$. Then U is invertible over $\Gamma_{\mathrm{an},qr}$.*

Proof. Put $B = U^{-1}AU^\sigma D^{-1}$, so that B is invertible over Γ_r. In the equation

$$U^{-1}AU^\sigma = BD,$$

the matrices A, U^σ, B, D are all invertible over $\Gamma_{[r/q,r]}$, so U must be as well. Since the entries of U and U^{-1} already belong to $\Gamma_{[r,qr]}$, they in fact belong to $\Gamma_{[r/q,qr]}$ by Corollary 2.5.7. Repeating this argument, we see that U is invertible over $\Gamma_{[r/q^i,qr]}$ for all positive integers i, yielding the desired result. □

Proposition 6.2.2. *Let M be a σ-module over $\Gamma^K_{\mathrm{an,con}}$. Then there exists a finite separable extension L of K and a σ-module N over $\Gamma^L_{\mathrm{con}}[\pi^{-1}]$ such that $N \otimes \Gamma^L_{\mathrm{an,con}} \cong M \otimes \Gamma^K_{\mathrm{an,con}}$ and the generic and special HN-polygons of N coincide.*

Proof. Note that it is enough to do this with L pseudo-finite separable, since applying F allows to pass from L to L^q. Also, there is no harm in assuming that the slopes of M are integers, after applying $[a]_*$ as necessary.

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of M, and define the invertible $n \times n$ matrix A over $\Gamma^K_{\mathrm{an,con}}$ by $F\mathbf{e}_j = \sum_i A_{ij}\mathbf{e}_i$. By Theorem 4.5.7, there exists an invertible $n \times n$ matrix U over $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$ and a diagonal $n \times n$ matrix D whose diagonal entries are powers of π, such that $U^{-1}AU^\sigma = D$. Put $h = \max_{i,j}\{w(D_{ii}) - w(D_{jj})\} = -w(D) - w(D^{-1})$.

Choose $r > 0$ such that Γ has enough r_0-units for some $r_0 > qr$, A is defined and invertible over $\Gamma^K_{\mathrm{an},r}$, and U is defined and invertible over $\Gamma^{\mathrm{alg}}_{\mathrm{an},qr}$. Let M_r be the $\Gamma^K_{\mathrm{an},qr}$-span of the $\mathbf{e}_i$. By Lemma 2.4.12, we can choose L pseudo-finite separable over K such that Γ^L has enough r_0-units, and such that there exists an $n \times n$ matrix V over $\Gamma^L_{[r,qr]}$ with $w_s(V - U) > -w_s(U^{-1}) + qh/(q-1)$ for $s \in [r, qr]$. Since

$$V^{-1}AV^{\sigma}D^{-1} = (U^{-1}V)^{-1}D(U^{-1}V)^{\sigma}D^{-1},$$

it follows that $w_r(V^{-1}AV^{\sigma}D^{-1} - I_n) > h/(q-1)$.
By Lemma 6.1.1, there exists an invertible $n \times n$ matrix W over $\Gamma^L_{[r,qr]}$ for which the matrix $B = (VW)^{-1}A(VW)^{\sigma}$ is such that $BD^{-1} - I_n$ has entries in $\pi\Gamma^L_r$ and $w_r(BD^{-1} - I_n) > 0$. By Lemma 6.2.1, the matrix VW is actually invertible over $\Gamma^L_{\mathrm{an},qr}$.
Let N be the σ-module over $\Gamma^L_{\mathrm{con}}[\pi^{-1}]$-module generated by $\mathbf{v}_1, \dots, \mathbf{v}_n$ with Frobenius action defined by $F\mathbf{v}_j = \sum_i B_{ij}\mathbf{v}_i$; then by what we have just shown, $N \otimes \Gamma^L_{\mathrm{an,con}} \cong M \otimes \Gamma^L_{\mathrm{an,con}}$. By Lemma 5.2.6, the generic HN-slopes of N are the $w(D_{ii})$, which by construction are also the special HN-slopes of N. (Compare [19, Proposition 6.9].) □

Remark 6.2.3. It should also be possible to establish Proposition 6.2.2 without the finite extension L of K.

6.3 Isoclinic σ-modules

Before proceeding to the general descent problem for HN-filtrations, we analyze the isoclinic case.

Definition 6.3.1. Let M be a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$. We say that M is *isoclinic* (of slope $\mu(M)$) if $M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ is isoclinic. By Proposition 5.5.1, this implies that $M \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$ is also isoclinic.

Proposition 6.3.2. *Let M be a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ which is unit-root (isoclinic of slope 0). Then*

$$H^0(M) = H^0(M \otimes \Gamma[\pi^{-1}]) = H^0(M \otimes \Gamma_{\mathrm{an,con}}[\pi^{-1}]);$$

in particular, if K is algebraically closed, then M admits a basis of eigenvectors.

Proof. There is no harm (thanks to Corollary 2.5.8) in assuming from the outset that K is algebraically closed. In this case, the eigenvectors of $M \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ all have slope 0 by Corollary 4.1.4; by Lemma 5.4.1, they all belong to M. Hence M admits a basis of eigenvectors; in particular, $M \cong M_{0,1}^{\oplus n}$. Then $H^0(M) = H^0(M \otimes \Gamma_{\mathrm{an,con}}[\pi^{-1}])$ by Proposition 3.3.4. □

Theorem 6.3.3. *(a) The base change functor, from isoclinic σ-modules over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ of some prescribed slope s to isoclinic σ-modules over $\Gamma[\pi^{-1}]$ of slope s, is fully faithful.*

(b) The base change functor, from isoclinic σ-modules over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ of slope s to isoclinic σ-modules over $\Gamma_{\mathrm{an,con}}$ of slope s, is an equivalence of categories.

In particular, any isoclinic σ-module over $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$ is isomorphic to a direct sum of standard σ-modules of the same slope (and hence is semistable by Proposition 4.5.10).

Proof. To see that the functors are fully faithful, let M and N be isoclinic σ-modules over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ of the same slope s. Then $M^\vee \otimes N$ is unit-root, and $\mathrm{Hom}(M, N) = H^0(M^\vee \otimes N)$, so the full faithfulness assertion follows from Proposition 6.3.2.
To see that the functor in (b) is essentially surjective, apply Proposition 6.2.2 to produce an F-stable $\Gamma^L_{\mathrm{con}}[\pi^{-1}]$-lattice N in $M \otimes \Gamma^L_{\mathrm{an,con}}$ for some finite separable extension L of K, which we may take to be Galois. Note that N is unique by full faithfulness of the base change functor (from $\Gamma^L_{\mathrm{con}}[\pi^{-1}]$ to $\Gamma^L_{\mathrm{an,con}}$); hence the action of $G = \mathrm{Gal}(L/K)$ on $M \otimes \Gamma^L_{\mathrm{an,con}}$ induces an action on N. By ordinary Galois descent, there is a unique $\Gamma_{\mathrm{con}}[\pi^{-1}]$-lattice M_b of N such that $N = M_b \otimes \Gamma^L_{\mathrm{con}}[\pi^{-1}]$; because of the uniqueness, M_b is F-stable. This yields the desired result. (Compare [19, Theorem 6.10].) □

Remark 6.3.4. The functor in (a) is not essentially surjective in general. For instance, if $K = k((t))$ with k perfect, M is an isoclinic σ-module over Γ_{con}, and b_n is the highest break of the representation of $\mathrm{Gal}(K^{\mathrm{sep}}/K)$ on the images modulo π^n of the eigenvectors of $M \otimes \Gamma^{\mathrm{alg}}$, then b_n/n is bounded. By contrast, if M is an isoclinic σ-module over Γ, b_n/n need not be bounded.

Incidentally, Theorem 6.3.3 allows us to give a more succinct characterization of isoclinic σ-modules, which one could take as an alternate definition.

Proposition 6.3.5. *Let c, d be integers with $d > 0$. Then a σ-module M over $\Gamma_{\mathrm{an,con}}$ is isoclinic of slope $s = c/d$ if and only if M admits a Γ_{con}-lattice N such that $\pi^{-c}F^d$ maps some (any) basis of N to another basis of N.*

Proof. If such a lattice exists, then we may apply Proposition 5.1.3 to $([d]_*M)(-c)$ to deduce that its generic HN-slopes are all zero. By Proposition 4.6.3, M is isoclinic of slope c/d.
Conversely, suppose M is isoclinic of slope s; then $([d]_*M)(-c)$ is isoclinic of slope 0. By Theorem 6.3.3, $([d]_*M)(-c)$ admits a unique F-stable $\Gamma_{\mathrm{con}}[\pi^{-1}]$-lattice N'. By Proposition 5.1.2, N' in turn admits an F-stable Γ_{con}-lattice N; by Proposition 5.1.3, the Frobenius on N carries any basis to another basis. □

6.4 Descent of the HN-filtration

At last, we are ready to establish the slope filtration theorem [19, Theorem 6.10].

Theorem 6.4.1. *Let M be a σ-module over $\Gamma_{\mathrm{an,con}}$. Then there exists a unique filtration $0 = M_0 \subset M_1 \subset \cdots \subset M_l = M$ of M by saturated σ-submodules with the following properties.*

(a) For $i = 1, \dots, l$, the quotient M_i/M_{i-1} is isoclinic of some slope s_i.

(b) $s_1 < \cdots < s_l$.

Moreover, this filtration coincides with the Harder-Narasimhan filtration of M.

Proof. Since isoclinic σ-modules are semistable by Theorem 6.3.3, any filtration as in (a) and (b) is a Harder-Narasimhan filtration. In particular, the filtration is unique if it exists.
To prove existence, it suffices to show that the HN-filtration of $M' = M \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$ descends to $\Gamma_{\mathrm{an,con}}$. Let M'_1 be the first step of that filtration; by Lemma 3.6.2, it is enough to check that M'_1 descends to $\Gamma_{\mathrm{an,con}}$ in the case $\operatorname{rank}(M'_1) = 1$.
By Proposition 6.2.2, there exists a finite separable extension L of K and a σ-module N over $\Gamma^L_{\mathrm{con}}[\pi^{-1}]$ such that $M \otimes \Gamma^L_{\mathrm{an,con}} \cong N \otimes \Gamma^L_{\mathrm{an,con}}$, and N has the same generic and special HN-polygons. Of course there is no harm in assuming L is Galois with $\mathrm{Gal}(L/K) = G$. By Theorem 5.5.2, M'_1 descends to $\Gamma^L_{\mathrm{an,con}}$; let M_1 be the descended submodule of $M \otimes \Gamma^L_{\mathrm{an,con}}$.
Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of M, let $\mathbf{v}$ be a generator of M_1, and write $\mathbf{v} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$ with $a_i \in \Gamma^L_{\mathrm{an,con}}$. Then for each i, j with $a_j \neq 0$, $a_i/a_j \in \operatorname{Frac}\Gamma^L_{\mathrm{an,con}}$ is invariant under G. By Corollary 2.4.11, $a_i/a_j \in \operatorname{Frac}\Gamma^K_{\mathrm{an,con}}$ for each i, j with $a_j \neq 0$; clearing denominators, we obtain a nonzero element of $M_1 \cap M$. Hence M_1 descends to $\Gamma_{\mathrm{an,con}}$, as desired. (Compare [19, Theorem 6.10].) $\square$

Corollary 6.4.2. *For any extension K' of K (complete with respect to a valuation $v_{K'}$ extending v_K, such that $\Gamma^{K'}$ has enough units), and any σ-module M over $\Gamma^K_{\mathrm{an,con}}$, the HN-filtration of $M \otimes \Gamma^{K'}_{\mathrm{an,con}}$ coincides with the result of tensoring the HN-filtration of M with $\Gamma^{K'}_{\mathrm{an,con}}$. In other words, the formation of the HN-filtration commutes with base change.*

Proof. The characterization of the HN-filtration given by Theorem 6.4.1 is stable under base change. $\square$

Corollary 6.4.3. *A σ-module over $\Gamma_{\mathrm{an,con}}$ is semistable if and only if it is isoclinic.*

Proof. If M is an isoclinic σ-module over $\Gamma_{\mathrm{an,con}}$, then M is semistable by Proposition 4.5.10. Conversely, if M is not isoclinic, then by Theorem 6.4.1, M admits a nonzero σ-submodule M_1 with $\mu(M_1) < \mu(M)$, so M is not semistable. $\square$

7 Complements

7.1 Differentials and the slope filtration

The slope filtration turns out to behave nicely with respect to differentials; this is what allows the application to Crew's conjecture.

Definition 7.1.1. Let S/R be an extension of topological rings. A *module of continuous differentials* is a topological S-module $\Omega^1_{S/R}$ equipped with a continuous R-linear derivation $d : S \to \Omega^1_{S/R}$, having the following universal property: for any topological S-module M equipped with a continuous R-linear derivation $D : S \to M$, there exists a unique morphism $\phi : \Omega^1_{S/R} \to M$ of topological S-modules such that $D = \phi \circ d$. Since the definition is via a universal property, the module of continuous differentials is unique up to unique isomorphism if it exists at all.

Convention 7.1.2. For the remainder of this section, assume that Γ_{con}, viewed as a topological $\mathcal{O}$-algebra via the levelwise topology, has a module of continuous differentials $\Omega^1_{\Gamma_{\mathrm{con}}/\mathcal{O}}$ which is finite free over Γ_{con}. In this case, for any finite separable extension L of K, $\Omega^1_{\Gamma_{\mathrm{con}}/\mathcal{O}} \otimes_{\Gamma_{\mathrm{con}}} \Gamma^L_{\mathrm{an,con}}$ is a module of continuous differentials of $\Gamma^L_{\mathrm{an,con}}$ over $\mathcal{O}[\pi^{-1}]$.

Example 7.1.3. If $K = k((t))$ as in Section 2.3, we have a module of continuous differentials for Γ_{con} over $\mathcal{O}$ given by $\Omega^1_{\Gamma_{\mathrm{con}}/\mathcal{O}} = \Gamma_{\mathrm{con}}\, dt$ with the formal derivation d sending $\sum c_i t^i$ to $(\sum i c_i t^{i-1}) dt$.

Remark 7.1.4. Note that $\Omega^1_{\Gamma_{\mathrm{con}}/\mathcal{O}}$ may be viewed as a σ-module, via the action of $d\sigma$. Since $d\sigma(\omega) \equiv 0 \pmod{\pi}$ for any $\omega \in \Omega^1_{\Gamma_{\mathrm{con}}/\mathcal{O}}$, by Proposition 5.1.3 the generic slopes of $\Omega^1_{\Gamma_{\mathrm{con}}/\mathcal{O}}$ as a σ-module are all positive. By Proposition 5.5.1, the special slopes of $\Omega^1_{\Gamma_{\mathrm{an,con}}/\mathcal{O}}$ as a σ-module are also nonnegative.

Definition 7.1.5. For $S = \Gamma_{\mathrm{con}}, \Gamma_{\mathrm{con}}[\pi^{-1}], \Gamma_{\mathrm{an,con}}$, define a *$\nabla$-module* over S to be a finite free S-module equipped with an integrable connection $\nabla : M \to M \otimes \Omega^1_{S/\mathcal{O}}$. (Integrability here means that the composition of ∇ with the induced map $M \otimes \Omega^1_{S/\mathcal{O}} \to M \otimes \wedge^2_S \Omega^1_{S/\mathcal{O}}$ is the zero map.) Define a *(σ, ∇)-module* over S to be a finite free S-module M equipped with the structures of both a σ-module and a ∇-module, which commute as in the following diagram:

$$\begin{tikzcd} M \arrow[r, "\nabla"] \arrow[d, "F"] & M \otimes \Omega^1_{S/\mathcal{O}} \arrow[d, "F \otimes d\sigma"] \\ M \arrow[r, "\nabla"] & M \otimes \Omega^1_{S/\mathcal{O}}. \end{tikzcd}$$

Proposition 7.1.6. *Let M be a (σ, ∇)-module over $\Gamma_{\mathrm{an,con}}$. Then each step of the HN-filtration of M is a (σ, ∇)-submodule of M.*

Proof. Let M_1 be the first step of the HN-filtration, which is isoclinic by Theorem 6.4.1; it suffices to check that M_1 is a (σ, ∇)-submodule of M. To simplify notation, write N for $\Omega^1_{\Gamma_{\mathrm{an,con}}/\mathcal{O}}$. Then the map $M_1 \to (M/M_1) \otimes N$ induced by ∇ is a morphism of σ-modules, and the slopes of $(M/M_1) \otimes N$ are all strictly greater than the slope of M_1 by Remark 7.1.4. By Proposition 4.6.4, the map $M_1 \to (M/M_1) \otimes N$ must be zero; that is, M_1 is a ∇-submodule of M. This proves the desired result. □

Proposition 7.1.7. *Let M be an isoclinic σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ such that $M \otimes \Gamma_{\mathrm{an,con}}$ admits the structure of a (σ, ∇)-module. Then M admits a corresponding structure of a (σ, ∇)-module.*

Proof. Write N for $\Omega^1_{\Gamma_{\mathrm{con}}[\pi^{-1}]/\mathcal{O}}$, so that ∇ induces an additive map $M \otimes \Gamma_{\mathrm{an,con}} \to (M \otimes N) \otimes \Gamma_{\mathrm{an,con}}$. Pick a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of M, and define the map $f : M \to M \otimes N$ by

$$f\left(\sum c_i \mathbf{e}_i\right) = \sum_i \mathbf{e}_i \otimes dc_i.$$

Then $\nabla - f$ is $\Gamma_{\mathrm{an,con}}$-linear, so we may view it as an element $\mathbf{v}$ of $M^\vee \otimes M \otimes N \otimes \Gamma_{\mathrm{an,con}}$. That element satisfies $F\mathbf{v} - \mathbf{v} = \mathbf{w}$ for some $\mathbf{w} \in M^\vee \otimes M \otimes N$ by the commutation relation between F and ∇. However, $M^\vee \otimes M$ is unit-root, so the (generic and special) slopes of $M^\vee \otimes M \otimes N$ are all positive by Remark 7.1.4. By Theorem 4.5.7 and Proposition 3.3.7(b1), it follows that $\mathbf{v} \in M^\vee \otimes M \otimes N$, so ∇ acts on M, as desired. (Compare [19, Theorem 6.12] and [3, Lemme V.14].) $\square$

Remark 7.1.8. We suspect there is a more conceptual way of saying this in terms of splitting a certain exact sequence of σ-modules.

7.2 The p-adic local monodromy theorem

We recall briefly how the slope filtration theorem, plus a theorem of Tsuzuki, yields the p-adic local monodromy theorem (formerly Crew's conjecture).

Convention 7.2.1. Throughout this section, retain notation as in Section 2.3, i.e., $\Gamma_{\mathrm{an,con}}$ is the Robba ring. Note that in this case, the integrability condition in Definition 7.1.5 is vacuous, since $\Omega^1_{\Gamma_{\mathrm{an,con}}/\mathcal{O}}$ is free of rank 1.

Definition 7.2.2. We say a ∇-module M over $\Gamma_{\mathrm{an,con}}$ is *constant* if it is spanned by the kernel of ∇; equivalently, M is isomorphic to a direct sum of trivial ∇-modules. (The "trivial" ∇-module here means $\Gamma_{\mathrm{an,con}}$ itself with the connection given by d.) We say M is *quasi-constant* if $M \otimes \Gamma^L_{\mathrm{an,con}}$ is constant for some finite separable extension L of K. We say a (σ, ∇)-module is (quasi-)constant if its underlying ∇-module is (quasi-)constant.

The key external result we need here is the following result essentially due to Tsuzuki [35]. A simplified proof of Tsuzuki's result has been given by Christol [5]; however, see [2] for the corrections of some errors in [5].

Proposition 7.2.3. *Let M be a unit-root (σ, ∇)-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$. Then $M \otimes \Gamma_{\mathrm{an,con}}$ is quasi-constant; in fact, for some finite separable extension L of K, $M \otimes \Gamma^L_{\mathrm{con}}[\pi^{-1}]$ admits a basis in the kernel of ∇.*

Proof. Suppose first that $M = M_0 \otimes \Gamma_{\mathrm{con}}[\pi^{-1}]$ for a unit-root (σ, ∇)-module M_0 over Γ_{con}. Then our desired statement is Tsuzuki's theorem [35, Theorem 4.2.6]. To reduce to this case, apply Proposition 5.1.2 to obtain a σ-stable Γ_{con}-lattice M_0; by applying Frobenius repeatedly, we see that M_0 must also be ∇-stable. Thus Tsuzuki's theorem applies to give the claimed result. $\square$

Definition 7.2.4. We say a ∇-module M over $\Gamma_{\mathrm{an,con}}$ is said to be *unipotent* if it admits an exhaustive filtration by ∇-submodules whose successive quotients are constant. We say M is *quasi-unipotent* if $M \otimes \Gamma^L_{\mathrm{an,con}}$ is unipotent for some finite separable extension L of K. We say a (σ, ∇)-module is (quasi-)unipotent if its underlying ∇-module is (quasi-)unipotent.

Theorem 7.2.5 (p-adic local monodromy theorem). *With notations as in Section 2.3 (i.e., $\Gamma_{\mathrm{an,con}}$ is the Robba ring), let M be a (σ, ∇)-module over $\Gamma_{\mathrm{an,con}}$. Then M is quasi-unipotent.*

Proof. By Proposition 7.1.6, each step of the HN-filtration of M is ∇-stable. It is thus enough to check that each successive quotient is quasi-unipotent; in other words, we may assume that M itself is isoclinic.
Note that the definition of unipotence does not depend on the Frobenius lift, so there is no harm in either applying the functor $[b]_*$ or in twisting. We may thus reduce to the case where M is isoclinic of slope zero (i.e., is unit-root). Applying Proposition 7.2.3 then yields the desired result. □

Example 7.2.6. In Theorem 7.2.5, it can certainly happen that M fails to be quasi-constant. For instance, if $u^\sigma = qu$, and M has rank two with a basis $\mathbf{e}_1, \mathbf{e}_2$ such that

$$F\mathbf{e}_1 = \mathbf{e}_1, \qquad F\mathbf{e}_2 = q\mathbf{e}_2$$
$$\nabla \mathbf{e}_1 = 0, \qquad \nabla \mathbf{e}_2 = \mathbf{e}_1 \otimes \frac{du}{u},$$

then M is a (σ, ∇)-module which is unipotent but not quasi-constant.

Remark 7.2.7. The fact that quasi-unipotence follows from the existence of a slope filtration as in Theorem 6.4.1 was originally pointed out by Tsuzuki [36, Theorem 5.2.1]. Indeed, that observation was the principal motivation for the construction of slope filtrations of σ-modules in [19].

Remark 7.2.8. We remind the reader that Theorem 7.2.5 has also been proved (independently) by André [1] and by Mebkhout [30], using the index theory for p-adic differential equations developed in a series of papers by Christol and Mebkhout [6], [7], [8], [9]. This represents a completely orthogonal approach to ours, as it primarily involves the structure of the connection rather than the Frobenius. The different approaches seem to have different strengths. For example, on one hand, the Christol-Mebkhout approach seems to say more about p-adic differential equations on annuli over p-adic fields which are not discretely valued. On the other hand, our approach has a certain flexibility that the Christol-Mebkhout approach lacks; for instance, it carries over directly to the q-difference situation considered by André and di Vizio in [1], whereas the analogue of the Christol-Mebkhout theory seems much more difficult to develop. It also carries over to the setting of "fake annuli" arising in the problem of semistable reduction for overconvergent F-isocrystals: in this setting, one replaces $k((t))$ by the completion of $k[t_1, t_1^{-1}, \dots, t_n, t_n^{-1}]$ for a valuation which totally orders monomials (i.e., the valuations of $t_1, \dots, t_n$ are linearly independent over $\mathbb{Q}$). See [25] for further details.

7.3 Generic versus special revisited

The adjectives "generic" and "special" were introduced in Chapter 5 to describe the two paradigms for attaching slopes to σ-modules over $\Gamma_{\mathrm{con}}[\pi^{-1}]$. Here is a bit of clarification as to why this was done. Throughout this section, retain notation as in Section 2.3.

Let $X \to \operatorname{Spec} k[\![t]\!]$ be a smooth proper morphism, for k a field of characteristic $p > 0$. Then the crystalline cohomology of X, equipped with the action of the absolute Frobenius, gives a (σ, ∇)-module M over the ring $\mathcal{O}[\![u]\!]$; the crystalline cohomology of the generic fibre corresponds to $M \otimes \Gamma$, whereas the crystalline cohomology of the special fibre corresponds to $M \otimes (\mathcal{O}[\![u]\!]/u\mathcal{O}[\![u]\!])$. However, the latter is isomorphic to $M \otimes (\Gamma_{\mathrm{an,con}} \cap \mathcal{O}[\pi^{-1}][\![u]\!])$ by "Dwork's trick" [13, Lemma 6.3]. Thus the generic and special HN-polygons correspond precisely to the Newton polygons of the generic and special fibres; the fact that the special HN-polygon lies above the generic HN-polygon (i.e., Proposition 5.5.1) in this case follows from Grothendieck's specialization theorem.

This gives a theoretical explanation for why Proposition 5.5.1 holds, but a more computationally explicit example may also be useful. (Thanks to Frans Oort for suggesting this presentation.) Suppose that σ is chosen so that $u^\sigma = u^p$. Let M be the rank 2 σ-module over Γ_{con} defined by

$$F\mathbf{v}_1 = \mathbf{v}_2, \qquad F\mathbf{v}_2 = p\mathbf{v}_1 + u\mathbf{v}_2.$$

Then $\mathbf{v}_1$ is a cyclic vector and $F^2\mathbf{v}_1 - uF\mathbf{v}_1 - p\mathbf{v}_1 = 0$, so by Lemma 5.2.4, the generic HN-polygon of M has the same slopes as the Newton polygon of the polynomial $x^2 - ux - p$, namely 0 and 1. On the other hand, Dwork's trick implies that the special HN-polygon of M has slopes 1/2 and 1/2.

7.4 Splitting exact sequences (again)

For reference, we collect here some more results about computing H^1 of σ-modules.

Proposition 7.4.1. *For any σ-modules M_1, M_2 over $\Gamma_{\mathrm{con}}[\pi^{-1}]$, the map $\operatorname{Ext}(M_1, M_2) \to \operatorname{Ext}(M_1 \otimes \Gamma_{\mathrm{an,con}}, M_2 \otimes \Gamma_{\mathrm{an,con}})$ is surjective.*

Proof. Let

$$0 \to M_2 \otimes \Gamma_{\mathrm{an,con}} \to M \to M_1 \otimes \Gamma_{\mathrm{an,con}} \to 0$$

be a short exact sequence of σ-modules. Choose a basis of M_2, then lift to M a basis of M_1; the result is a basis of M. Let A be the matrix via which F acts on this basis; after rescaling the basis of M_2 suitably, we can put ourselves into the situation of Lemma 6.1.1. We can now perform the iteration of Lemma 6.1.1 in such a way as to respect the short exact sequence (i.e., take $u_{ijlm} = 0$ whenever the pair (i,j) falls in the lower left block); as in the proof of Proposition 6.2.2, we end up with a model M_b of M over $\Gamma_{\mathrm{con}}[\pi^{-1}]$, which by construction sits in an exact sequence $0 \to M_2 \to M_b \to M_1 \to 0$. This yields the desired surjectivity. $\square$

We next give some generalizations of parts of Proposition 3.3.7.

Proposition 7.4.2. *Let M be a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ whose generic HN-slopes are all nonnegative. Then the map $H^1(M) \to H^1(M \otimes \Gamma_{\mathrm{an,con}})$ is injective.*

Proof. By Proposition 5.1.2, we can choose an F-stable Γ_{con}-lattice M_0 of M. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of M_0, and define the matrix A over Γ_{con} by $F\mathbf{e}_j = \sum_i A_{ij}\mathbf{e}_i$. Choose $r > 0$ such that A has entries in Γ_r and $w_r(A) \geq 0$.
Suppose $\mathbf{v} \in M$ and $\mathbf{w} \in M \otimes \Gamma_{\mathrm{an,con}}$ satisfy $\mathbf{v} = \mathbf{w} - F\mathbf{w}$, and write $\mathbf{v} = \sum_i x_i \mathbf{e}_i$ and $\mathbf{w} = \sum_i y_i \mathbf{e}_i$ with $x_i \in \Gamma_{\mathrm{con}}[\pi^{-1}]$ and $y_i \in \Gamma_{\mathrm{an,con}}$; then $x_i = y_i - \sum_j A_{ij} y_j^\sigma$.
If $\mathbf{w} \notin M$, we can choose $m < 0$ such that $v_m(x_i) = \infty$ and $0 < \min_i \min_{l \leq m}\{v_{l,r}(y_i)\} < \infty$. Then

$$\begin{aligned}
\min_{l \leq m}\{v_{l,r}(y_i)\} &> q^{-1} \min_{l \leq m}\{v_{l,r}(y_i)\} \\
&\geq q^{-1} \min_{l \leq m} \min_j \{v_{l,r}(A_{ij} y_j^\sigma)\} \\
&\geq q^{-1} \min_{l \leq m} \min_j \{v_{l,r}(y_j^\sigma)\} \\
&= q^{-1} \min_{l \leq m} \min_j \{r v_l(y_j^\sigma) + l\} \\
&\geq \min_{l \leq m} \min_j \{r v_l(y_j) + l\} \\
&= \min_{l \leq m} \min_j \{v_{l,r}(y_j)\};
\end{aligned}$$

taking the minimum over all i yields a contradiction. Hence $\mathbf{w} \in M$, yielding the injectivity of the map $H^1(M) \to H^1(M \otimes \Gamma_{\mathrm{an,con}})$. $\square$

Proposition 7.4.3. *Let M be a σ-module over $\Gamma[\pi^{-1}]$ whose HN-slopes are all positive. Then $F - 1$ is a bijection on M, i.e., $H^0(M) = H^1(M) = 0$.*

Proof. By Proposition 5.1.2, we can choose an F-stable Γ-lattice M_0 of M such that $F(M_0) \subseteq \pi M_0$. Then $F - 1$ is a bijection on $M_0/\pi M_0$, hence also on M_0 and M. $\square$

Lemma 7.4.4. *For L/K a finite Galois extension, and M a σ-module over $\Gamma_{\mathrm{an,con}}$, the map $H^1(M) \to H^1(M \otimes \Gamma^L_{\mathrm{an,con}})$ is injective.*

Proof. Put $G = \mathrm{Gal}(L/K)$. Given $\mathbf{w} \in M$, suppose that there exists $\mathbf{v} \in M \otimes \Gamma^L_{\mathrm{an,con}}$ such that $\mathbf{w} = \mathbf{v} - F\mathbf{v}$. Then we also have $\mathbf{w} = \mathbf{v}' - F\mathbf{v}'$ for

$$\mathbf{v}' = \frac{1}{\#G} \sum_{g \in G} \mathbf{v}^g,$$

which is G-invariant and hence belongs to M. □

Theorem 7.4.5. *Let M be a σ-module over $\Gamma_{\mathrm{an,con}}$ whose HN-slopes are all positive. Then the exact sequence*

$$0 \to M \to N \to \Gamma_{\mathrm{an,con}} \to 0 \tag{7.4.6}$$

splits if and only if N has smallest HN-slope zero.

Proof. If the sequence splits, then $P(N) = P(M) + P(\Gamma_{\mathrm{an,con}})$ by Proposition 4.7.2, so N has smallest HN-slope zero. We verify the converse first in the case where K is algebraically closed, so that $\Gamma_{\mathrm{an,con}} = \Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$.
We proceed by induction on $\mathrm{rank}(M)$. In case M is isoclinic, then the inequality $P(N) \geq P(M) + P(\Gamma_{\mathrm{an,con}})$ from Proposition 4.7.2 and the hypothesis that N has smallest HN-slope zero together force $P(N) = P(M) + P(\Gamma_{\mathrm{an,con}})$; by Proposition 4.7.2 again, (7.4.6) splits. In case M is not isoclinic, let M_1 be the first step in the HN-filtration of M. By Proposition 4.7.2, we have

$$P(N) \geq P(M_1) + P(N/M_1) \geq P(M_1) + P(M/M_1) + P(\Gamma_{\mathrm{an,con}});$$

since $P(N)$ has smallest slope 0, so does $P(M_1) + P(N/M_1)$. Since $P(M_1)$ has positive slope, $P(N/M_1)$ must have smallest slope zero. Hence by the induction hypothesis, the exact sequence $0 \to M/M_1 \to N/M_1 \to \Gamma_{\mathrm{an,con}} \to 0$ splits, say as $N/M_1 \cong M/M_1 \oplus M'$ with $M' \cong \Gamma_{\mathrm{an,con}}$. Let N' be the inverse image of M' under the surjection $N \to N/M_1$; it follows now that (7.4.6) splits if and only if $0 \to M_1 \to N' \to M' \to 0$ splits. By Proposition 4.7.2 again, $P(N) \geq P(N') + P(M/M_1) \geq P(M_1) + P(M/M_1) + P(\Gamma_{\mathrm{an,con}})$, and $P(M/M_1)$ has all slopes positive, so $P(N')$ has smallest slope zero. Again by the induction hypothesis, $0 \to M_1 \to N' \to M' \to 0$ splits, yielding the splitting of (7.4.6).
To summarize, we have proved that if N has smallest HN-slope zero, then (7.4.6) splits after tensoring with $\Gamma^{\mathrm{alg}}_{\mathrm{an,con}}$; it remains to descend this splitting back to $\Gamma_{\mathrm{an,con}}$. To do this, apply Proposition 6.2.2 to produce a σ-module M_0 over $\Gamma^L_{\mathrm{con}}[\pi^{-1}]$, for some finite Galois extension L of K, with $M_0 \otimes \Gamma^L_{\mathrm{an,con}} \cong M \otimes \Gamma^L_{\mathrm{an,con}}$. Then apply Proposition 7.4.1 to descend the given exact sequence, after tensoring up to $\Gamma^L_{\mathrm{an,con}}$, to an exact sequence $0 \to M_0 \to N_0 \to \Gamma^L_{\mathrm{con}}[\pi^{-1}] \to 0$. We have shown that the exact sequence $0 \to M_0 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}} \to N_0 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{an,con}} \to \Gamma^{\mathrm{alg}}_{\mathrm{an,con}} \to 0$ splits; by Proposition 7.4.2, the exact sequence $0 \to M_0 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}] \to N_0 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}] \to \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}] \to 0$ splits.
Choose $\mathbf{w} \in M_0$ whose image in $H^1(M_0)$ corresponds to the exact sequence $0 \to M_0 \to N_0 \to \Gamma^L_{\mathrm{con}}[\pi^{-1}] \to 0$; we have now shown that the class of $\mathbf{w}$

in $H^1(M_0 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}])$ vanishes. By Proposition 7.4.3, there exists a unique $\mathbf{v} \in M_0 \otimes \Gamma^L[\pi^{-1}]$ with $\mathbf{w} = \mathbf{v} - F\mathbf{v}$. Since the class of $\mathbf{w}$ in $H^1(M_0 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}])$ vanishes, we have $\mathbf{v} \in M_0 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$; since M_0 is free over $\Gamma^L_{\mathrm{con}}[\pi^{-1}]$ and $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}] \cap \Gamma^L[\pi^{-1}] = \Gamma^L_{\mathrm{con}}[\pi^{-1}]$, we have $\mathbf{v} \in M_0$. Hence the sequence (7.4.6) splits after tensoring with $\Gamma^L_{\mathrm{an,con}}$. By Lemma 7.4.4, (7.4.6) also splits, as desired. □

7.5 Full faithfulness

Here is de Jong's original application of the reverse filtration [13, Proposition 8.2].

Proposition 7.5.1. *Suppose that K admits a valuation p-basis. Let M be a σ-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ admitting an injective F-equivariant $\Gamma_{\mathrm{con}}[\pi^{-1}]$-linear morphism $\phi : M \to \Gamma[\pi^{-1}](m)$ for some integer m. Then $\phi^{-1}(\Gamma_{\mathrm{con}}[\pi^{-1}])$ is a σ-submodule of M of rank 1, and its extension to $\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$ is equal to the first step of the reverse filtration of M. In particular, M has highest generic slope m with multiplicity 1.*

Proof. We first suppose K is algebraically closed (this case being [13, Corollary 5.7]). Let M_1 be the first step of the reverse filtration of M. Then $M_1 \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}]$ is isomorphic to a direct sum of standard σ-modules of some slope $s_1 = c/d$; by Lemma 5.4.1 (applied to $([d]_*(M_1 \otimes \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]))(-c)$), M_1 itself is isomorphic to a direct sum of standard σ-modules of slope s_1.
The map ϕ induces a nonzero F-equivariant map $M_1 \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}] \to \Gamma^{\mathrm{alg}}[\pi^{-1}](m)$, and hence a nonzero element of $H^0(M_1^\vee \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}](m))$. By Corollary 4.1.4, we must have $m = s_1$, and so $M_1 \cong \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}](m)^{\oplus n}$ for some n. By Proposition 3.3.4, we have $H^0(M_1^\vee(m)) = H^0(M_1^\vee \otimes \Gamma^{\mathrm{alg}}[\pi^{-1}](m))$, and so ϕ actually induces an injective F-equivariant map $M_1 \to \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}](m)$.
To summarize, M has highest generic slope m, and the first step of the reverse filtration is contained in $\phi^{-1}(\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}])$. Since the latter is a σ-submodule of M of rank no more than 1, we have the desired result.
We now suppose K is general. Put $M' = M \otimes_{\Gamma_{\mathrm{con}}[\pi^{-1}]} \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$; by Proposition 2.2.21(c), the composite map

$$\psi : M' \stackrel{\phi \otimes 1}{\to} \Gamma[\pi^{-1}] \otimes_{\Gamma_{\mathrm{con}}[\pi^{-1}]} \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}] \stackrel{\mu}{\to} \Gamma^{\mathrm{alg}}[\pi^{-1}]$$

is also injective. By the above, $\psi^{-1}(\Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}])$ is a σ-submodule of M' of rank 1, and coincides with the first step of the reverse filtration of M'. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of M, put $\mathbf{v} = \psi^{-1}(1)$, and write $\mathbf{v} = \sum x_i \mathbf{e}_i$ with $x_i \in \Gamma^{\mathrm{alg}}_{\mathrm{con}}[\pi^{-1}]$. Then $1 = \psi(\mathbf{v}) = \sum x_i \phi(\mathbf{e}_i)$; by Proposition 2.2.21(b), we have $x_i \in \Gamma_{\mathrm{con}}[\pi^{-1}]$ for $i = 1, \dots, n$. Hence $\mathbf{v} \in \phi^{-1}(\Gamma_{\mathrm{con}}[\pi^{-1}])$, so the latter is a σ-submodule of M of rank 1. This yields the desired result. □

Remark 7.5.2. Proposition 7.5.1 can be used to reduce instances of showing $H^0(M) = H^0(M \otimes \Gamma[\pi^{-1}])$, for M an F-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$, to showing that a certain class in $H^1(N)$, where N is a related F-module over $\Gamma_{\mathrm{con}}[\pi^{-1}]$ with positive HN-slopes, vanishes. Thanks to Proposition 7.4.2, this in turn reduces to checking vanishing of the class in $H^1(N \otimes \Gamma_{\mathrm{an,con}})$, where either Dwork's trick, in the case of [13], or the p-adic local monodromy theorem, in the case of [20], can be brought to bear. Note that by Theorem 7.4.5, it is enough to check vanishing of a class in $H^1(N \otimes \Gamma_{\mathrm{an,con}})$ after replacing K by a finite separable extension.

References

[1] Y. André, Filtrations de type Hasse-Arf et monodromie p-adique, *Invent. Math.* 148 (2002), 285–317.

[2] Y. André and L. di Vizio, q-difference equations and p-adic local monodromy, in Analyse complexe, systèmes dynamiques, sommabilité des séries divergentes et théories galoisiennes (I) — Volume en l'honneur de Jean-Pierre Ramis (M. Loday-Richaud, ed.), *Astérisque* 296 (2004), 55–111.

[3] L. Berger, Représentations p-adiques et équations différentielles, *Invent. Math.* 148 (2002), 219–284.

[4] L. Berger, Équations différentielles p-adiques et (ϕ, N)-modules filtrés, preprint available at `arXiv: math.NT/0406601` (version of 29 Jun 2004).

[5] G. Christol, About a Tsuzuki theorem, in *p-adic functional analysis (Ioannina, 2000)*, Lecture Notes in Pure and Appl. Math. 222, Dekker, New York, 2001, 63–74.

[6] G. Christol and Z. Mebkhout, Sur le théorème de l'indice des équations différentielles p-adiques. I, *Ann. Inst. Fourier (Grenoble)* 43 (1993), 1545–1574.

[7] G. Christol and Z. Mebkhout, Sur le théorème de l'indice des équations différentielles p-adiques. II, *Ann. of Math.* 146 (1997), 345–410.

[8] G. Christol and Z. Mebkhout, Sur le théorème de l'indice des équations différentielles p-adiques. III, *Ann. of Math.* 151 (2000), 385–457.

[9] G. Christol and Z. Mebkhout, Sur le théorème de l'indice des équations différentielles p-adiques. IV, *Invent. Math.* 143 (2001), 629–672.

[10] P. Colmez, Série principale unitaire pour $\mathrm{GL}_2(\mathbb{Q}_p)$ et représentations triangulines de dimension 2, preprint available at `http://www.math.jussieu.fr/~colmez/`.

[11] R. Crew, F-isocrystals and p-adic representations, in *Algebraic geometry, Bowdoin, 1985 (Brunswick, Maine, 1985)*, Proc. Sympos. Pure Math. 46, Amer. Math. Soc., Providence, 1987, 111–138.

[12] R. Crew, Finiteness theorems for the cohomology of an overconvergent isocrystal on a curve, *Ann. Sci. École Norm. Sup.* 31 (1998), 717–763.

[13] A.J. de Jong, Homomorphisms of Barsotti-Tate groups and crystals in positive characteristic, *Invent. Math.* 134 (1998), 301–333.

[14] J. Dieudonné, Lie groups and Lie hyperalgebras over a field of characteristic $p > 0$. IV, *Amer. J. Math.* 77 (1955), 429–452.

[15] A. Grothendieck, Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas. I, *Publ. Math. IHÉS* 20 (1964).

[16] G. Harder and M.S. Narasimhan, On the cohomology groups of moduli spaces of vector bundles on curves, *Math. Ann.* 212 (1974/75), 215–248.

[17] U. Hartl and R. Pink, Vector bundles with a Frobenius structure on the punctured unit disc, *Compos. Math.* 140 (2004), 689–716.

[18] N.M. Katz, Slope filtration of F-crystals, in Journées de Géométrie Algébrique de Rennes (Rennes, 1978), Vol. I, *Astérisque* 63 (1979), 113–163.

[19] K.S. Kedlaya, A p-adic local monodromy theorem, *Annals of Math.* 160 (2004), 93–184.

[20] K.S. Kedlaya, Full faithfulness for overconvergent F-isocrystals, in *Geometric aspects of Dwork theory*, de Gruyter, Berlin, 2004, 819–835.

[21] K.S. Kedlaya, Local monodromy for p-adic differential equations: an overview, *Intl. J. of Number Theory* 1 (2005), 109–154.

[22] K.S. Kedlaya, Finiteness of rigid cohomology with coefficients, *Duke Math. J.*, to appear; preprint available at `arXiv: math.AG/0208027` (version of 2 Nov 2005).

[23] K.S. Kedlaya, Fourier transforms and p-adic "Weil II", preprint available at `arXiv: math.NT/0210149` (version of 20 Jul 2005).

[24] K.S. Kedlaya, Semistable reduction for overconvergent F-isocrystals, I: Unipotence and logarithmic extensions, preprint available at `arXiv: math.NT/0405069` (version of 24 Jul 2005).

[25] K.S. Kedlaya, The p-adic local monodromy theorem for fake annuli, I: The monomial case, preprint available at `arXiv: math.NT/0507496` (version of 8 Aug 2005).

[26] M. Kisin, Crystalline representations and F-crystals, preprint available at `http://math.uchicago.edu/~kisin`.

[27] F.-V. Kuhlmann, untitled book in progress, available at `http://math.usask.ca/~fvk/Fvkbook.htm` (references to version of 16 Mar 2005).

[28] M. Lazard, Les zéros des fonctions analytiques d'une variable sur un corps valué complet, *Publ. Math. IHÉS* 14 (1962), 47–75.

[29] Ju.I. Manin, Theory of commutative formal groups over fields of finite characteristic (Russian), *Uspehi Mat. Nauk* 18 (1963), 3–90.

[30] Z. Mebkhout, Analogue p-adique du théorème de Turrittin et le théorème de la monodromie p-adique, *Invent. Math.* 148 (2002), 319–351.

[31] D.S. Passman, *The algebraic structure of group rings*, Robert E. Krieger Publishing Co. Inc., Melbourne, FL, 1985.

[32] P. Schneider, *Nonarchimedean functional analysis*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2002.

[33] S.S. Shatz, The decomposition and specialization of algebraic families of vector bundles, *Compos. Math.* 35 (1977), 163–187.

[34] N. Tsuzuki, The overconvergence of morphisms of étale ϕ-∇-spaces on a local field, *Compos. Math.* 103 (1996), 227–239.

[35] N. Tsuzuki, Finite local monodromy of overconvergent unit-root F-isocrystals on a curve, *Amer. J. Math.* 120 (1998), 1165–1190.

[36] N. Tsuzuki, Slope filtration of quasi-unipotent overconvergent F-isocrystals, *Ann. Inst. Fourier (Grenoble)* 48 (1998), 379–412.

Kiran S. Kedlaya
Dept. of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139
USA
kedlaya@mit.edu

Time-Like Isothermic Surfaces Associated to Grassmannian Systems

M. P. Dussan and M. A. Magid

Received: March 3, 2005
Revised: August 10, 2005

Communicated by Patrick Delorme

Abstract. We establish that, as is the case with space-like isothermic surfaces, time-like isothermic surfaces in pseudo-riemannian space $\mathbb{R}^{n-j,j}$ are associated to the $O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$-system.

2000 Mathematics Subject Classification: Primary 53A35; Secondary53C40.
Keywords and Phrases: Isothermic surfaces, integrable systems, Grassmannian systems

1 Introduction

There is no doubt that the recent renaissance in interest about isothermic surfaces is principally due to the fact that they constitute an integrable system, as can be seen in several new works where it is shown, for instance, that the theory of isothermic surfaces in $\mathbb{R}^3$ can be reformulated within the modern theory of soliton theory [4], or can be analyzed as curved flats in the symmetric space $O(4,1)/O(3)\times O(1,1)$ [3]. Additionally, in a recent work of Burstall [1], we find an account of the theory of isothermic surfaces in $\mathbb{R}^n$ from both points of view: of classic surfaces geometry as well as from the perspective of the modern theory of integrable systems and loop groups.
The key point of this class of surfaces, as well as of the classic pseudospherical surfaces and those with constant mean curvature, is that the Gauss-Codazzi equations are soliton equations, they have a zero-curvature formulation, i.e.,

the equations should amount to the flatness of a family of connections depending on an auxiliary parameter. It is well known that this special property allows actions of an infinite dimensional group on the space of solutions, called the "dressing action" in the soliton theory. For instance, the geometric transformations found for the surfaces above such as Backlund, Darboux and Ribaucour, arise as the dressing action of some simple elements.

More recently, in 1997, Terng in [12] defined a new integrable system, the U/K-system (or n-dimensional system associated to U/K), which is very closely related to that of curved flats discovered by Ferus and Pedit [8]. Terng, in [12], showed that the U/K-system admits a Lax connection and initiated the project to study the geometry associated with these systems. In fact, using the existence of this Lax connection, in 2002 Bruck-Du-Park-Terng ([2]) studied the geometry involved in two particular cases of U/K-systems: $O(m+n)/O(m)\times O(n)$ and $O(m+n,1)/O(m)\times O(n,1)$-systems. For these cases, they found that the isothermic surfaces, submanifolds with constant sectional curvatures and submanifolds admitting principal curvature coordinates are associated to them, and, that the dressing actions of simple elements on the space of solutions corresponded to Backlund, Darboux and Ribaucour transformations for submanifolds.

Later, looking for a relation between space-like isothermic surfaces in pseudo-riemannian space and the U/K-systems, the first author found in [6] that the class of space-like isothermic surfaces in pseudo-riemannian space $\mathbb{R}^{n-j,j}$ for any signature j, were associated to the $O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$-system. The principal point in this study was the suitable choice of a one maximal abelian subalgebra, which allows one to obtain elliptic Gauss equations, which are appropriate for space-like surfaces.

The main goal of this note is to show that time-like isothermic surfaces in the pseudo-riemannian space $\mathbb{R}^{n-j,j}$ are also associated to the $O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$-systems, defined by other two maximal abelian subalgebras, that are not conjugate under the $Ad(K)$-action, where $K=O(n-j,j)\times O(1,1)$. We study the class of time-like surfaces both with diagonal and non-diagonal second fundamental form, in the cases when its principal curvatures are real and distinct and when they are complex conjugates. We show that an isothermic pair i.e, two isothermic time-like surfaces which are dual, in the diagonal or non-diagonal case, are associated to our systems. Additionally, in this paper we present a review of the principal results recently obtained in [7], about the geometric transformations associated to the dressing action of certain elements with two simple poles on the space of solutions of the complex $O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$-system, corresponding to the timelike isothermic surfaces whose second fundamental forms are non-diagonal. The geometric transformations associated to real case of timelike isothermic surfaces with second fundamental forms are diagonal, were already studied in [14].

Finally, we note that all time-like surfaces of constant mean curvature, all time-like rotation surfaces and all time-like members of Bonnet families are examples of time-like isothermic surfaces [11].

2 The U/K-systems

In this section, we introduce the definition of U/K-system given by Terng in [12]. Let U be a semi-simple Lie group, σ an involution on U and K the fixed point set of σ. Then U/K is a symmetric space. The Lie algebra $\mathcal{K}$ is the fixed point set of the differential σ_* of σ at the identity, in others words, it is the $+1$ eigenspace of σ_*. Let now $\mathcal{P}$ denote the -1 eigenspace of σ_*. Then we have the Lie algebra of U, $\mathcal{U} = \mathcal{K} \oplus \mathcal{P}$ and

$$[\mathcal{K},\mathcal{K}] \subset \mathcal{K},\ [\mathcal{K},\mathcal{P}] \subset \mathcal{P},\ [\mathcal{P},\mathcal{P}] \subset \mathcal{K}.$$

Let $\mathcal{A}$ be a non-degenerate maximal abelian subalgebra in $\mathcal{P}$, $a_1, a_2, ..., a_n$ a basis for $\mathcal{A}$ and $\mathcal{A}^\perp$ the orthogonal complement of $\mathcal{A}$ in the algebra $\mathcal{U}$ with respect to the Killing form $<,>$. Then the U/K-system is the following first order system of non-linear partial differential equations for $v : \mathbb{R}^n \to \mathcal{P} \cap \mathcal{A}^\perp$.

$$[a_i, v_{x_j}] - [a_j, v_{x_i}] = [[a_i, v], [a_j, v]], \qquad 1 \leq i \neq j \leq n, \tag{1}$$

where $v_{x_j} = \frac{\partial v}{\partial x_j}$.

The first basic result established in [12] is the existence of one-parameter family of connections whose flatness condition is exactly the U/K-system.

Theorem 2.1. *([12]) The following statements are equivalent for a map $v : \mathbb{R}^n \to \mathcal{P} \cap \mathcal{A}^\perp$:*
i) v is solution of the U/K-system (1).
ii)

$$[\frac{\partial}{\partial x_i} + \lambda a_i + [a_i, v], \frac{\partial}{\partial x_j} + \lambda a_j + [a_j, v]] = 0 \ for\ all\ \lambda \in \mathbb{C}, \tag{2}$$

iii) θ_λ is a flat $\mathcal{U}_\mathbb{C} = \mathcal{U} \otimes \mathbb{C}$-connection 1-form on $\mathbb{R}^n$ for all $\lambda \in \mathbb{C}$, where

$$\theta_\lambda = \sum (a_i\lambda + [a_i, v])dx_i. \tag{3}$$

iv) There exists E so that $E^{-1}dE = \theta_\lambda$.

The one-parameter family of flat connections θ_λ given by (3) is called the *Lax connection* of the U/K-system (1).
It is well known that for a flat connection $\theta = \sum_{i=1}^n A_i(x)dx_i$, the *trivialization* of θ, is a solution E for the following linear system:

$$E_{x_i} = EA_i. \tag{4}$$

Or equivalently of $E^{-1}dE = \theta$.

3 Main Results

In the next two subsections we establish our results that time-like isothermic surfaces are associated to the Grassmannian system $O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$. In fact, using the existence of another two maximal abelian subalgebras in the subspace $\mathcal{P}$, different from that of the space-like case given in [6] in which the first author obtained elliptic Gauss equations, we associate to each of these maximal abelian subalgebras one $O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$-system. As we will see, these systems are not equivalent and for each of these maximal abelian subalgebra we obtain hyperbolic Gauss equations, which are correct for time-like surfaces.
Let $U/K = O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$, where

$$O(n-j+1,j+1) = \\ \{X \in GL(n+2) | X^t \begin{pmatrix} I_{n-j,j} & 0 \\ 0 & J' \end{pmatrix} X = \begin{pmatrix} I_{n-j,j} & 0 \\ 0 & J' \end{pmatrix}\},$$

$$I_{n-j,j} = \begin{pmatrix} I_{n-j} & 0 \\ 0 & -I_j \end{pmatrix} \quad \text{and} \quad J' = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Let $\mathcal{U} = o(n-j+1,j+1)$ be the Lie algebra of U and $\sigma : \mathcal{U} \to \mathcal{U}$ be an involution defined by $\sigma(X) = I_{n,2}^{-1} X I_{n,2}$. Denote by $\mathcal{K}, \mathcal{P}$ the +1, -1 eigenspaces of σ respectively, i.e.,

$$\mathcal{K} = \{\begin{pmatrix} Y_1 & 0 \\ 0 & Y_2 \end{pmatrix} \Big| Y_1 \in o(n-j,j), Y_2 \in o(1,1)\} = o(n-j,j)\times o(1,1),$$

and

$$\mathcal{P} = \{\begin{pmatrix} 0 & \xi \\ -J'\xi^t I_{n-j,j} & 0 \end{pmatrix} \Big| \xi \in \mathcal{M}_{n\times 2}\}.$$

3.1 Time-like case with diagonal second fundamental form

Here we assume the elements $a_1, a_2 \in \mathcal{M}_{(n+2)\times(n+2)}$, where

$$a_1 = e_{n,n+1} + e_{n,n+2} + e_{n+1,n} + e_{n+2,n}$$

$$a_2 = -e_{1,n+1} + e_{1,n+2} - e_{n+1,1} + e_{n+2,1},$$

and e_{ij} is the $(n+2)\times(n+2)$ elementary matrix whose only non-zero entry is 1 in the ij^{th} place.
Then it is easy to see that the subalgebra $\mathcal{A} =< a_1, a_2 >$ is maximal abelian in $\mathcal{P}$, that $Tr[a_1^2]Tr[a_2^2] - Tr[a_1a_2]^2 = 16$ with $Tr[a_1^2] = 4$, so the induced metric on $\mathcal{A}$ is positive definite and finally that

$$\mathcal{P}\cap\mathcal{A}^\perp = \{\begin{pmatrix} 0 & \xi \\ -J'\xi^t I_{n-j,j} & 0 \end{pmatrix} \Big| \xi \in \mathcal{M}_{n\times 2},\ \xi_{11} = \xi_{12},\ \xi_{n1} = -\xi_{n2}\}.$$

So using this basis $\{a_1, a_2\}$, the U/K-system (1) for this symmetric space is the following PDE for

$$\xi = \begin{pmatrix} \xi_1 & \xi_1 \\ r_{1,1} & r_{1,2} \\ \vdots & \vdots \\ r_{n-2,1} & r_{n-2,2} \\ \xi_2 & -\xi_2 \end{pmatrix} : \mathbb{R}^2 \to \mathcal{M}_{n\times 2},$$

$$\begin{cases} (r_{i,2})_{x_1} - (r_{i,1})_{x_1} = -2(r_{i,1} + r_{i,2})\xi_2, \quad i = 1, ..., n-2 \\ (r_{i,2})_{x_2} + (r_{i,1})_{x_2} = 2(r_{i,2} - r_{i,1})\xi_1, \quad i = 1, ..., n-2 \\ 2((\xi_1)_{x_2} + (\xi_2)_{x_1}) = \sum_{i=1}^{n-2} \sigma_i (r_{i1}^2 - r_{i2}^2) \\ (\xi_2)_{x_2} + (\xi_1)_{x_1} = 0, \end{cases} \tag{5}$$

where $\sigma_i = 1,\ i = 1, ..., n-j-1$ and $\sigma_i = -1,\ i = n-j, ..., n-2$.
We now denote the entries of ξ by:

$$\begin{pmatrix} \xi_1 & \xi_1 \\ \xi_2 & -\xi_2 \end{pmatrix} = F \quad \text{and} \quad \begin{pmatrix} r_{1,1} & r_{1,2} \\ \vdots & \vdots \\ r_{n-2,1} & r_{n-2,2} \end{pmatrix} = G.$$

For convenience, we call the U/K-system (5) the real $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system, because this system will correspond to time-like surfaces in $\mathbb{R}^{n-j,j}$ whose shape operators are diagonalizable.
Continuing with the same notation used in [2], the real $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II is the PDE for $(F, G, B) : \mathbb{R}^2 \to gl_*(2) \times \mathcal{M}_{(n-2)\times 2} \times O(1,1)$, where $gl_*(2) = \{N \in \mathcal{M}_{2\times 2} | N_{11} = N_{12},\ N_{21} = -N_{22}\}$

$$\begin{cases} (r_{i,2})_{x_1} - (r_{i,1})_{x_1} = -2\xi_2(r_{i,1} + r_{i,2}), \quad i = 1, ..., n-2 \\ (r_{i,1})_{x_2} + (r_{i,2})_{x_2} = 2\xi_1(r_{i,2} - r_{i,1}), \quad i = 1, ..., n-2 \\ 2((\xi_1)_{x_2} + (\xi_2)_{x_1}) = \sum_{i=1}^{n-2} \sigma_i (r_{i1}^2 - r_{i2}^2) \\ (b_{11})_{x_1} - (b_{12})_{x_1} = 2\xi_2(b_{11} + b_{12}) \\ (b_{21})_{x_1} - (b_{22})_{x_1} = 2\xi_2(b_{22} + b_{21}) \\ (b_{11})_{x_2} + (b_{12})_{x_2} = -2\xi_1(b_{11} - b_{12}) \\ (b_{21})_{x_2} + (b_{22})_{x_2} = -2\xi_1(b_{21} - b_{22}) \end{cases} \tag{6}$$

where the matrix $B = (b_{ij}) \in O(1,1)$. Now we recall that if we take $g = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ solution of $g^{-1}dg = \theta_0$ and B being the particular case $B = \begin{pmatrix} b & 0 \\ 0 & \frac{1}{b} \end{pmatrix} = \begin{pmatrix} e^{2u} & 0 \\ 0 & e^{-2u} \end{pmatrix}$, we obtain the relation

$$B^{-1}dB = \begin{pmatrix} -2\xi_1 dx_2 + 2\xi_2 dx_1 & 0 \\ 0 & -2\xi_2 dx_1 + 2\xi_1 dx_2 \end{pmatrix},$$

which implies that $\xi_1 = -u_{x_2}$ and $\xi_2 = u_{x_1}$, hence the matrix ξ becomes:

$$\xi = \begin{pmatrix} -u_{x_2} & -u_{x_2} \\ r_{1,1} & r_{1,2} \\ \vdots & \vdots \\ r_{n-2,1} & r_{n-2,2} \\ u_{x_1} & -u_{x_1} \end{pmatrix}.$$

So the real $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II is the set of partial differential equations for $(u, r_{1,1}, r_{1,2}, \ldots, r_{n-2,1}, r_{n-2,2})$:

$$\begin{cases} (r_{i,2})_{x_1} - (r_{i,1})_{x_1} = -2(r_{i,1} + r_{i,2})u_{x_1}, & i = 1, ..., n-2 \\ (r_{i,1})_{x_2} + (r_{i,2})_{x_2} = -2(r_{i,2} - r_{i,1})u_{x_2}, & i = 1, ..., n-2 \\ 2(u_{x_1x_1} - u_{x_2x_2}) = \sum_{i=1}^{n-2} \sigma_i (r_{i1}^2 - r_{i2}^2) \end{cases} \tag{7}$$

We observe that the next proposition follows from Proposition 2.5 in [2].

Proposition 3.1. *the following statements are equivalent for map (F, G, B) : $\mathbb{R}^2 \to gl_*(2) \times \mathcal{M}_{(n-2)\times 2} \times O(1,1)$:*
(1) (F, G, B) is solution of (6).
(2) $\theta_\lambda^{II} := g_2 \theta_\lambda g_2^{-1} - dg_2 g_2^{-1}$ is a flat connection on $\mathbb{R}^2$ for all $\lambda \in \mathbb{C}$, where θ_λ is the Lax connection associated to the solution ξ of the system (5) and $g_2 = \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix}$ is the $O(1,1)$-part of the trivialization $g = (g_1, g_2)$ of θ_0.
(3) $\theta_\lambda^{II} := g_2 \theta_\lambda g_2^{-1} - dg_2 g_2^{-1}$ is flat for $\lambda = 1$, where g_2 is the same as in item (2).

Before showing the relationship between the Grassmannian system and isothermic surfaces we give the definition of a time-like isothermic surface with shape operators diagonalized over $\mathbb{R}$.

Definition 3.1. *(*Real isothermic surface*) Let $\mathcal{O}$ be a domain in $\mathbb{R}^{1,1}$. An immersion $X : \mathcal{O} \to \mathbb{R}^{n-j,j}$ is called a real time-like isothermic surface if it has flat normal bundle and the two fundamental forms are:*

$$I = e^{2v}(-dx_1^2 + dx_2^2), \ II = e^v \sum_{i=2}^{n-1} (g_{i-1,2} dx_2^2 - g_{i-1,1} dx_1^2) e_i,$$

with respect to some parallel normal frame $\{e_i\}$. Or equivalently $(x_1, x_2) \in \mathcal{O}$ is conformal and line of curvature coordinate system for X.

We note that each isothermic surface has a dual surface ([11]) and make the following related definition.

Definition 3.2. *(*Real isothermic time-like dual pair in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$*). Let $\mathcal{O}$ be a domain in $\mathbb{R}^{1,1}$ and $X_i : \mathcal{O} \to \mathbb{R}^{n-j,j}$ an immersion with*

flat and non-degenerate normal bundle for $i = 1, 2$. (X_1, X_2) *is called a real isothermic time-like dual pair in* $\mathbb{R}^{n-j,j}$ *of type* $O(1,1)$ *if :*
(i) The normal plane of $X_1(x)$ *is parallel to the normal plane of* $X_2(x)$ *and* $x \in \mathcal{O}$,
(ii) there exists a common parallel normal frame $\{e_2, ..., e_{n-1}\}$, *where* $\{e_i\}_2^{n-j}$ *and* $\{e_i\}_{n-j+1}^{n-1}$ *are space-like and time-like vectors resp.*
(iii) $x \in \mathcal{O}$ *is a conformal line of curvature coordinate system with respect to* $\{e_2, ..., e_{n-1}\}$ *for each* X_k *such that the fundamental forms of* X_k *are:*

$$\begin{cases} I_1 = b^{-2}(-dx_1^2 + dx_2^2), \\ II_1 = b^{-1} \sum_{i=2}^{n-1} [-(g_{i-1,1} + g_{i-1,2})dx_1^2 + (g_{i-1,2} - g_{i-1,1})dx_2^2] e_i, \\ I_2 = b^2(-dx_1^2 + dx_2^2), \\ II_2 = b \sum_{i=2}^{n-1} [-(g_{i-1,1} + g_{i-1,2})dx_1^2 - (g_{i-1,2} - g_{i-1,1})dx_2^2] e_i, \end{cases} \tag{8}$$

where $B = \begin{pmatrix} b & 0 \\ 0 & b^{-1} \end{pmatrix}$ *is in* $O(1,1)$ *and a* $\mathcal{M}_{(n-2)\times 2}$*-valued map* $G = (g_{ij})$.

Our first result, whose proof follows the same lines of Theorem 6.8 or 7.4 in [2], gives us the relationship between the dual pair of real isothermic timelke surfaces in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$ and the solutions of the real $O(n-j+1, j+1)/O(n-j,j) \times O(1,1)$-system II (6):

THEOREM 3.1. *Suppose* $(u, r_{1,1}, r_{1,2}, \ldots, r_{n-2,1}, r_{n-2,2})$ *is solution of (7) and* F, B *are given by*

$$F = \begin{pmatrix} \xi_1 & \xi_1 \\ \xi_2 & -\xi_2 \end{pmatrix} = \begin{pmatrix} -u_{x_2} & -u_{x_2} \\ u_{x_1} & -u_{x_1} \end{pmatrix}, \quad B = \begin{pmatrix} e^{2u} & 0 \\ 0 & e^{-2u} \end{pmatrix}.$$

Then: (*a*)
$\omega =$

$$\begin{pmatrix} 0 & \epsilon_1\beta_1 dx_2 & \ldots & \epsilon_{n-2}\beta_{n-2} dx_2 & 2(-\xi_1 dx_1 + \xi_2 dx_2) \\ -\beta_1 dx_2 & 0 & \ldots & 0 & -\eta_1 dx_1 \\ -\beta_2 dx_2 & 0 & \ldots & 0 & -\eta_2 dx_1 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ -\beta_{n-2} dx_2 & 0 & \ldots & 0 & -\eta_{n-2} dx_1 \\ 2(\xi_2 dx_2 - \xi_1 dx_1) & -\epsilon_1\eta_1 dx_1 & \ldots & -\epsilon_{n-2}\eta_{n-2} dx_1 & 0 \end{pmatrix} \tag{9}$$

where $\epsilon_i = 1$ *for* $i < n-j$ *and* $\epsilon_i = -1$ *for* $i \geq n-j$, *and where* $\beta_i = (r_{i,2} - r_{i,1})$, $\eta_i = (r_{i,1} + r_{i,2})$, $i = 1, ..., n-2$, *is a flat* $o(n-j,j)$*-valued connection 1-form. Hence there exists* $A : \mathbb{R}^2 \to O(n-j,j)$ *such that*

$$A^{-1} dA = \omega, \tag{10}$$

where ω *is given by (9).*

(*b*)

$$A\begin{pmatrix}-dx_2 & 0 & \dots & 0 & dx_1\\ dx_2 & 0 & \dots & 0 & dx_1\end{pmatrix}^t B^{-1}$$

is exact. So there exists a map $X:\mathbb{R}^2\to\mathcal{M}_{n\times 2}$ *such that*

$$dX = A\begin{pmatrix}-dx_2 & 0 & \dots & 0 & dx_1\\ dx_2 & 0 & \dots & 0 & dx_1\end{pmatrix}^t B^{-1} \tag{11}$$

(*c*) *Let* $X_j:\mathbb{R}^2\to\mathbb{R}^{n-j,j}$ *denote the* j*-th column of* X *(solution of 11) and* e_i *denote the* i*-th column of* A*. Then* (X_1,X_2) *is a dual pair of real isothermic timelike surfaces in* $\mathbb{R}^{n-j,j}$ *of type* $O(1,1)$*. I.e.* (X_1,X_2) *have the following properties:*
(1) e_1, e_n *are resp. space-like and time-like tangent vectors to* X_1 *and* X_2*, i.e, the tangent planes of* X_1, X_2 *are parallel.*
(2) $\{e_2,...,e_{n-1}\}$ *is a parallel normal frame for* X_1 *and* X_2*, with* $\{e_2,...,e_{n-j}\}$ *and* $\{e_{n-j+1},...,e_{n-1}\}$ *being resp. space-like and time-like vectors.*
(3) the two fundamental forms for the immersion X_k *are:*

$$\begin{cases} I_1 = e^{-4u}(dx_2^2 - dx_1^2)\\ II_1 = e^{-2u}\sum_{i=2}^{n-1}[(r_{i-1,2}-r_{i-1,1})dx_2^2 - (r_{i-1,1}+r_{i-1,2})dx_1^2]e_i\\ I_2 = e^{4u}(dx_2^2-dx_1^2)\\ II_2 = e^{2u}\sum_{i=2}^{n-1}[-(r_{i-1,2}-r_{i-1,1})dx_2^2 - (r_{i-1,1}+r_{i-1,2})dx_1^2]e_i\end{cases}$$

Remark 3.1. *We observe that we can prove a theorem like Theorem (3.1) for a general solution* (F,G,B) *of system (6) by taking a generic* $F=(f_{ij})$ *and* $B=(b_{ij})=\begin{pmatrix}b & 0\\ 0 & b^{-1}\end{pmatrix}\in O(1,1)$*, i.e, we conclude that if* (F,G,B) *is a solution of system (6), we obtain a real isothermic timelike dual pair in* $\mathbb{R}^{n-j,j}$ *of type* $O(1,1)$ *with I and II fundamental forms like in (8).*

Now for the converse, we have the following result.

Theorem 3.2. *Let* (X_1,X_2) *be a real isothermic time-like dual pair in* $\mathbb{R}^{n-j,j}$ *of type* $O(1,1)$*,* $\{e_2,...,e_{n-1}\}$ *a common parallel normal frame and* (x_1,x_2) *a common isothermal line of curvature coordinates for* X_1 *and* X_2*, such that the two fundamental forms* I_k*,* II_k *for* X_k *are given by (8). Set* $f_{11}=-\frac{b_{x_2}}{2b}=f_{12}$*,* $f_{22}=-\frac{b_{x_1}}{2b}=-f_{21}$*, and* $F=(f_{ij})_{2\times 2}$*. Then if all entries of* G *and the* $g_{i-1,1}+g_{i-1,2}$*,* $g_{i-1,2}-g_{i-1,1}$ *are non-zero, then* (F,G,B) *is a solution of (6).*

Proof. From the definition of real isothermic time-like pair in $\mathbb{R}^{n-j,j}$, we have

$$\omega_1^{(1)} = -b^{-1}dx_2, \quad \omega_n^{(1)} = b^{-1}dx_1 \quad \omega_1^{(2)} = b\,dx_2, \quad \omega_n^{(2)} = b\,dx_1$$

is a dual 1-frame for X_k and $\omega_{1\alpha}^{(k)} = l_\alpha(g_{\alpha-1,2}-g_{\alpha-1,1})dx_2$, $\omega_{n\alpha}^{(k)} = -l_\alpha(g_{\alpha-1,1}+g_{\alpha-1,2})dx_1$ for each X_k, where $l_\alpha = 1$ if $\alpha = 2,...,n-j$ and $l_\alpha = -1$ if

$\alpha = n-j+1,...,n-1$. We observe that $\omega^{(k)}_{i\alpha}$, $i = 1, n$, $\alpha = 2,...,n-1$ are independent of k. We find that the Levi-civita connection 1-form for the metric I_k is:

$$\omega^{(k)}_{1n} = \frac{b_{x_1}}{b}dx_2 + \frac{b_{x_2}}{b}dx_1 = 2(-f^{(k)}_{22}dx_2 - f^{(k)}_{11}dx_1),$$

which are independent from k. Hence $\omega^{(k)}_{1n} = \omega^{(1)}_{1n} = 2(-f_{22}dx_2 - f_{11}dx_1) = 2(\xi_2 dx_2 - \xi_1 dx_1)$. So the structure equations and the Gauss-Codazzi equations for X_1, X_2 imply that (F, G, B) is a solution of system (6). ■

So, from Theorems (3.1), (3.2) and Remark (3.1), it follows that there exists a correspondence between the solutions (F, G, B) of system (6) and a dual pair of real isothermic timelike surfaces in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$.

Theorem 3.3. *The real $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II (6) is the Gauss-Codazzi equation for a time-like surface in $\mathbb{R}^{n-j,j}$ such that:*

$$\begin{cases} I = e^{4u}(dx_2^2 - dx_1^2) \\ II = e^{2u}\sum_{i=2}^{n-1}[-(r_{i-1,2} - r_{i-1,1})dx_2^2 - (r_{i-1,1} + r_{i-1,2})dx_1^2]e_i \end{cases} \tag{12}$$

Proof. We can read from I and II that: $\omega_1 = e^{2u}dx_2, \omega_n = e^{2u}dx_1$, $\omega_{1,i} = \eta_i(r_{i-1,2} - r_{i-1,1})dx_2$, and $\omega_{n,i} = -\eta_i(r_{i-1,2} + r_{i-1,1})dx_1$, where $\eta_i = 1$ if $i = 2,...,n-j$ and $\eta_i = -1$ if $i = n-j+1,...,n-1$. Now use the structure equations: $d\omega_1 = \omega_n \wedge \omega_{1n}$ and $d\omega_n = \omega_1 \wedge \omega_{n1}$, to obtain:

$$\omega_{1n} = 2(u_{x_1}dx_2 + u_{x_2}dx_1).$$

Now from the Gauss equation: $d\omega_{1n} = -\sum_{i=2}^{n-j}\omega_{1,i}\wedge\omega_{n,i} + \sum_{i=n-j+1}^{n-1}\omega_{1,i}\wedge\omega_{n,i}$, we have that

$$u_{x_1x_1} - u_{x_2x_2} = \frac{1}{2}[\sum_{i=1}^{n-2}\sigma_i(r_{i1}^2 - r_{i2}^2)].$$

The Codazzi equations: $d\omega_{1,i} = -\omega_{1n} \wedge \omega_{n,i}$ and $d\omega_{n,i} = -\omega_{n1} \wedge \omega_{1,i}$ for $i = 2,...,n-1$, yield, for these values of i,

$$\begin{aligned} (r_{i-1,2})_{x_1} - (r_{i-1,1})_{x_1} &= -2(r_{i-1,2} + r_{i-1,1})u_{x_1} \\ (r_{i-1,1})_{x_2} + (r_{i-1,2})_{x_2} &= -2(r_{i-1,2} - r_{i-1,1})u_{x_2}. \end{aligned}$$

Collecting our information we see that the Gauss-Codazzi equation is the following system for $(u, r_{1,1}, r_{1,2}, \ldots, r_{n-2,1}, r_{n-2,2})$:

$$\begin{cases} (r_{i-1,2})_{x_1} - (r_{i-1,1})_{x_1} = -2(r_{i-1,2} + r_{i-1,1})u_{x_1}, & i = 2,...,n-1 \\ (r_{i-1,1})_{x_2} + (r_{i-1,2})_{x_2} = -2(r_{i-1,2} - r_{i-1,1})u_{x_2}, & i = 2,...,n-1 \\ 2(u_{x_1x_1} - u_{x_2x_2}) = \sum_{i=1}^{n-2}\sigma_i(r_{i1}^2 - r_{i2}^2) \end{cases} \tag{13}$$

Hence if we put

$$B = \begin{pmatrix} e^{2u} & 0 \\ 0 & e^{-2u} \end{pmatrix}, \; F = \begin{pmatrix} -u_{x_2} & -u_{x_2} \\ u_{x_1} & -u_{x_1} \end{pmatrix}, \; G = \begin{pmatrix} r_{1,1} & r_{1,2} \\ \vdots & \vdots \\ r_{n-2,1} & r_{n-2,2} \end{pmatrix}, \tag{14}$$

we see that (F, G, B) is solution of the real $O(n-j+1, j+1)/O(n-j, j)\times O(1,1)$-system II. Conversely, if (F, G, B) is solution of the real $O(n-j+1, j+1)/O(n-j, j)\times O(1,1)$-system II (6), and we assume B being as in (14), then the fourth and sixth equation of system (6), imply that

$$\xi_2 = u_{x_1}, \;\; \xi_1 = -u_{x_2}$$

ie, (F, G, B) is the form (14). Finally writing the real $O(n-j+1, j+1)/O(n-j, j)\times O(1,1)$-system II for this (F, G, B) in terms of u and r_{ij} we get equation (13). ■

The next result follows from Theorem (3.1) and Theorem (3.3).

Theorem 3.4. *Let $\mathcal{O}$ be a domain of $\mathbb{R}^{1,1}$, and $X_2 : \mathcal{O} \to \mathbb{R}^{n-j,j}$ an immersion with flat normal bundle and $(x_1, x_2) \in \mathcal{O}$ an isothermal line of curvature coordinate system with respect to a parallel normal frame $\{e_2, ..., e_{n-1}\}$, such that I and II fundamental forms are given by (12). Then there exists an immersion X_1, unique up to translation, such that (X_1, X_2) is a real isothermic timelike dual pair in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$. Moreover, the fundamental forms of X_1, X_2 are respectively:*

$$\begin{cases} I_1 = e^{-4u}(dx_2^2 - dx_1^2) \\ II_1 = e^{-2u}\sum_{i=2}^{n-1}[(r_{i-1,2} - r_{i-1,1})dx_2^2 - (r_{i-1,1} + r_{i-1,2})dx_1^2]e_i \\ I_2 = e^{4u}(dx_2^2 - dx_1^2) \\ II_2 = e^{2u}\sum_{i=2}^{n-1}[-(r_{i-1,2} - r_{i-1,1})dx_2^2 - (r_{i-1,1} + r_{i-1,2})dx_1^2]e_i \end{cases} \tag{15}$$

It follows from Gauss equation that the Gaussian curvatures of X_1 and X_2 of the real isothermic timelike dual pair (15), denoted by $K_G^{(1)}$, $K_G^{(2)}$, and the mean curvatures, denoted by $\eta^{(1)}$ and $\eta^{(2)}$, are given by

$$K_G^{(1)} = -e^{4u}\sum_{i=1}^{n-2}\sigma_i(r_{i,1}^2 - r_{i,2}^2), \qquad K_G^{(2)} = e^{-4u}\sum_{i=1}^{n-2}\sigma_i(r_{i,1}^2 - r_{i,2}^2),$$

$$\eta^{(1)} = e^{2u}\sum_{i=1}^{n-2} r_{i,2}e_{i+1}, \qquad \eta^{(2)} = e^{-2u}\sum_{i=1}^{n-2} r_{i,1}e_{i+1},$$

where $\sigma_i = 1,\; i = 1, ..., n-j-1$ and $\sigma_i = -1,\; i = n-j, ..., n-2$.

3.2 Timelike case with non-diagonal second fundamental form

We continue with the same notational convention used in the subsection above. For this new case, we take the elements $a_1, a_2 \in \mathcal{M}_{(n+2)\times(n+2)}$, to be

$$a_1 = e_{1,n+1} + e_{n,n+2} + e_{n+1,n} - e_{n+2,1}$$

$$a_2 = -e_{1,n+2} + e_{n,n+1} + e_{n+1,1} + e_{n+2,n}.$$

We note that $Tr[a_1^2]Tr[a_2^2] - Tr[a_1a_2]^2 = -16$ and $Tr[a_1^2] = 0$, so that the induced metric on $\mathcal{A}$ is time-like.
One can see easily that the space $\mathcal{A}$ spanned by a_1 and a_2 is a maximal abelian subalgebra contain in $\mathcal{P}$, and that

$$\mathcal{A}^\perp \cap \mathcal{P} = \{\begin{pmatrix} 0 & \xi \\ -J'\xi^t I_{n-j,j} & 0 \end{pmatrix} |\xi \in \mathcal{M}_{n\times 2}, \xi_{11} = -\xi_{n2}, \xi_{12} = \xi_{n1}\}.$$

So the matrix $v \in \mathcal{A}^\perp \cap \mathcal{P}$ if and only if

$$v = \begin{pmatrix} 0 & \dots & \dots & & 0 & \xi_1 & \xi_2 \\ 0 & \dots & \dots & & 0 & r_{1,1} & r_{1,2} \\ \vdots & \dots & \dots & & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & & 0 & r_{n-2,1} & r_{n-2,2} \\ 0 & \dots & \dots & & 0 & \xi_2 & -\xi_1 \\ -\xi_2 & -r_{1,2} & \dots & r_{n-2,2} & -\xi_1 & 0 & 0 \\ -\xi_1 & -r_{1,1} & \dots & r_{n-2,1} & \xi_2 & 0 & 0 \end{pmatrix}.$$

Then using this basis $\{a_1, a_2\}$, the U/K-system (1) for this symmetric space is the following PDE for

$$\xi = \begin{pmatrix} \xi_1 & \xi_2 \\ r_{1,1} & r_{1,2} \\ \vdots & \vdots \\ r_{n-2,1} & r_{n-2,2} \\ \xi_2 & -\xi_1 \end{pmatrix} : \mathbb{R}^2 \to \mathcal{M}_{n\times 2},$$

$$\begin{cases} -r_{i,2_{x_2}} - r_{i,1_{x_1}} & = 2(r_{i,2}\xi_1 - r_{i,1}\xi_2), \quad i = 1, ..., n-2 \\ -r_{i,1_{x_2}} + r_{i,2_{x_1}} & = -2(r_{i,1}\xi_1 + r_{i,2}\xi_2), \quad i = 1, ..., n-2 \\ (-2\xi_1)_{x_2} + (2\xi_2)_{x_1} & = \sum_{i=1}^{n-2} \sigma_i(r_{i,1}^2 + r_{i,2}^2) \\ (2\xi_2)_{x_2} - (2\xi_1)_{x_1} & = 0. \end{cases} \tag{16}$$

We now denote the entries of ξ by:

$$\begin{pmatrix} \xi_1 & \xi_2 \\ \xi_2 & -\xi_1 \end{pmatrix} = F \quad \text{and} \quad \begin{pmatrix} r_{1,1} & r_{1,2} \\ \vdots & \vdots \\ r_{n-2,1} & r_{n-2,2} \end{pmatrix} = G.$$

For convenience, we call the U/K-system (16) the complex $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system, because this system will correspond to time-like surfaces in $\mathbb{R}^{n-j,j}$ whose shape operators have complex eigenvalues.
Now, the complex $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II is the following PDE for $(F, G, B) : \mathbb{R}^2 \to gl_*(2) \times \mathcal{M}_{(n-2)\times 2} \times O(1,1)$, where $gl_*(2)$

is the set of matrices 2×2 such that $-f_{11}=f_{22},\ f_{21}=f_{12}$,

$$\begin{cases} -r_{i,2_{x_2}} - r_{i,1_{x_1}} & = 2(r_{i,2}\xi_1 - r_{i,1}\xi_2), \\ -r_{i,1_{x_2}} + r_{i,2_{x_1}} & = -2(r_{i,1}\xi_1 + r_{i,2}\xi_2), \\ (-2\xi_1)_{x_2} + (2\xi_2)_{x_1} & = \sum_{i=1}^{n-2}\sigma_i(r_{i,1}^2 + r_{i,2}^2) \\ b_{22_{x_2}} + b_{21_{x_1}} & = -2b_{22}\xi_1 + 2b_{21}\xi_2, \\ b_{12_{x_2}} + b_{11_{x_1}} & = -2b_{12}\xi_1 + 2b_{11}\xi_2, \\ b_{21_{x_2}} - b_{22_{x_1}} & = 2b_{21}\xi_1 + 2b_{22}\xi_2, \\ b_{11_{x_2}} - b_{12_{x_1}} & = 2b_{11}\xi_1 + 2b_{12}\xi_2, \end{cases} \tag{17}$$

where the matrix $B=(b_{ij})\in O(1,1)$ and $1\le i\le n-2$. Now taking $B=\begin{pmatrix} e^{2u} & 0 \\ 0 & e^{-2u}\end{pmatrix}$, and using the fact that

$$B^{-1}dB = \begin{pmatrix} 2\xi_2 dx_1 + 2\xi_1 dx_2 & 0 \\ 0 & -2\xi_2 dx_1 - 2\xi_1 dx_2 \end{pmatrix},$$

we have

$$\xi = \begin{pmatrix} u_{x_2} & u_{x_1} \\ r_{1,1} & r_{1,2} \\ \vdots & \vdots \\ r_{n-2,1} & r_{n-2,2} \\ u_{x_1} & -u_{x_2} \end{pmatrix}.$$

So the complex $O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$-system II is the PDE for $(u, r_{1,1}, r_{1,2}, \ldots, r_{n-2,1}, r_{n-2,2})$:

$$\begin{cases} -r_{i,2_{x_2}} - r_{i,1_{x_1}} & = 2(r_{i,2}\xi_1 - r_{i,1}\xi_2), \\ -r_{i,1_{x_2}} + r_{i,2_{x_1}} & = -2(r_{i,1}\xi_1 + r_{i,2}\xi_2), \\ -2u_{x_2x_2} + 2u_{x_1x_1} & = \sum_{i=1}^{n-2}\sigma_i(r_{i,1}^2 + r_{i,2}^2). \end{cases} \tag{18}$$

Remark 3.2. *We recall that the complex $O(n-j+1,j+1)/O(n-j,j)\times O(1,1)$-system II is the flatness condition for the family:*

$$\theta_\lambda^{II} = \begin{pmatrix} \omega & MB^{-1} \\ BN & 0 \end{pmatrix}$$

where $B=(b_{ij})\in O(1,1)$ and the matrices $\omega\in\mathcal{M}_{n\times n}, M\in\mathcal{M}_{n\times 2}, N\in\mathcal{M}_{2\times n}$ are given by:

$$\omega = \begin{pmatrix} 0 & \vec{a} & \vec{b} & c \\ -\vec{a}^{\,t} & 0 & 0 & \vec{d}^{\,t} \\ \vec{b}^{\,t} & 0 & 0 & \vec{e}^{\,t} \\ c & \vec{d} & -\vec{e} & 0 \end{pmatrix}, \quad M = \lambda\begin{pmatrix} dx_1 & -dx_2 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ dx_2 & dx_1 \end{pmatrix} \tag{19}$$

$$N = \lambda\begin{pmatrix} dx_2 & 0 & \ldots & 0 & dx_1 \\ -dx_1 & 0 & \ldots & 0 & dx_2 \end{pmatrix}$$

where

$$\begin{aligned}
\vec{a} &= (a_1, \ldots, a_{n-j-1}) \text{ and } a_k = r_{k,1}dx_2 - r_{k,2}dx_1, \text{ for } 1 \le k \le n-j-1,\\
\vec{b} &= (b_{n-j}, \ldots, b_{n-2}) \text{ and } b_q = -r_{q,1}dx_2 + r_{q,2}dx_1, \text{ for } n-j \le q \le n-2,\\
c &= -2\xi_1 dx_1 - 2\xi_2 dx_2\\
\vec{d} &= (d_1, \ldots, d_{n-j-1}) \text{ and } d_k = -r_{k,1}dx_1 - r_{k,2}dx_2, \text{ for } 1 \le k \le n-j-1,\\
\vec{e} &= (e_q, \ldots, e_{n-2}) \text{ and } e_q = -r_{q,1}dx_1 - r_{q,2}dx_2, \text{ for } n-j \le q \le n-2.
\end{aligned}$$

We note that a proposition similar to Proposition (3.1), can be proven in this new case.
At this point we need the appropriate definition of a complex isothermic surface, i.e., one that has an isothermal coordinate system with respect to which all the shape operators are diagonalized over $\mathbb{C}$.

Definition 3.3. (Complex isothermic surface) *Let $\mathcal{O}$ be a domain in $\mathbb{R}^{1,1}$. An immersion $X : \mathcal{O} \to \mathbb{R}^{n-j,j}$ is called a complex time-like isothermic surface if it has flat normal bundle and the two fundamental forms are:*

$$I = \pm e^{2v}(-dx_1^2 + dx_2^2),\ II = \sum_{i=2}^{n-1} e^v(g_{i1}(dx_2^2 - dx_1^2) - 2g_{i2}dx_1dx_2)e_i,$$

with respect to some parallel normal frame $\{e_i\}$.

Remark 3.3. *We note that given any complex isothermic surface there is a dual isothermic surface with parallel normal space ([11]). The U/K system generates this pair of dual surfaces, making it clear that they should be considered essentially as a single unit.*

Definition 3.4. (Complex isothermic time-like dual pair in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$). *Let $\mathcal{O}$ be a domain in $\mathbb{R}^{1,1}$ and $X_i : \mathcal{O} \to \mathbb{R}^{n-j,j}$ an immersion with flat and non-degenerate normal bundle for $i = 1, 2$. (X_1, X_2) is called a complex isothermic timelike dual pair in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$ if :*
(i) The normal plane of $X_1(x)$ is parallel to the normal plane of $X_2(x)$ and $x \in \mathcal{O}$,
(ii) there exists a common parallel normal frame $\{e_2, ..., e_{n-1}\}$, where $\{e_i\}_2^{n-j}$ and $\{e_i\}_{n-j+1}^{n-1}$ are space-like and time-like vectors resp.
(iii) $x \in \mathcal{O}$ is a isothermal coordinate system with respect to $\{e_2, ..., e_{n-1}\}$, for each X_k, such that the fundamental forms of X_k are diagonalizable over $\mathbb{C}$. Namely,

$$\begin{cases} I_1 = b^{-2}(dx_1^2 - dx_2^2),\\ II_1 = -b^{-1}\sum_{i=1}^{n-2}[g_{i,2}(dx_2^2 - dx_1^2) + 2g_{i,1}dx_1dx_2]e_{i+1},\\ I_2 = b^2(-dx_1^2 + dx_2^2),\\ II_2 = b\sum_{i=1}^{n-2}[g_{i,1}(dx_2^2 - dx_1^2) - 2g_{i,2}dx_1dx_2]e_{i+1}, \end{cases} \tag{20}$$

where $B = \begin{pmatrix} b & 0 \\ 0 & b^{-1} \end{pmatrix}$ *is in* $O(1,1)$ *and a* $\mathcal{M}_{(n-2)\times 2}$*-valued map* $G = (g_{ij})$.

Theorem 3.5. *Suppose* $(u, r_{1,1}, r_{1,2}, \dots, r_{n-2,1}, r_{n-2,2})$ *is solution of (18) and* F, B *are given by*

$$F = \begin{pmatrix} u_{x_2} & u_{x_1} \\ u_{x_1} & -u_{x_2} \end{pmatrix}, B = \begin{pmatrix} e^{2u} & 0 \\ 0 & e^{-2u} \end{pmatrix}.$$

Then: (a) *The* ω *defined by (19) is a flat* $o(n-j,j)$*-valued connection 1-form. Hence there exists* $A : \mathbb{R}^2 \to O(n-j,j)$ *such that*

$$A^{-1} dA = \omega. \tag{21}$$

(b)

$$A \begin{pmatrix} dx_1 & 0 & \dots & 0 & dx_2 \\ -dx_2 & 0 & \dots & 0 & dx_1 \end{pmatrix}^t B^{-1}$$

is exact. So there exists a map $X : \mathbb{R}^2 \to \mathcal{M}_{n\times 2}$ *such that*

$$dX = A \begin{pmatrix} dx_1 & 0 & \dots & 0 & dx_2 \\ -dx_2 & 0 & \dots & 0 & dx_1 \end{pmatrix}^t B^{-1}. \tag{22}$$

(c) *Let* $X_i : \mathbb{R}^2 \to \mathbb{R}^{n-j,j}$ *denote the i-th column of* X *(solution of (22)) and* e_i *denote the i-th column of* A. *Then* X_1 *and* X_2 *are a dual pair of isothermic time-like surfaces in* $\mathbb{R}^{n-j,j}$ *with common isothermal coordinates and second fundamental forms diagonalized over* $\mathbb{C}$, *so that:*
(1) e_1, e_n *are space-like and time-like tangent vectors to* X_1 *and* X_2, *i.e, the tangent planes of* X_1, X_2 *are parallel.*
(2) $\{e_2, \dots, e_{n-1}\}$ *form a parallel normal frame for* X_1 *and* X_2 *of signature* $\{n-j-1, j-1\}$.
(3) the two fundamental forms for the immersion X_i *are:*

$$\begin{cases} I_1 = e^{-4u}(dx_1^2 - dx_2^2) \\ II_1 = -e^{-2u} \sum_{i=1}^{n-2} [r_{i,2}(dx_2^2 - dx_1^2) + 2r_{i,1} dx_1 dx_2] e_{i+1} \\ I_2 = e^{4u}(dx_2^2 - dx_1^2) \\ II_2 = e^{2u} \sum_{i=1}^{n-2} [r_{i,1}(dx_2^2 - dx_1^2) - 2r_{i,2} dx_1 dx_2] e_{i+1}. \end{cases} \tag{23}$$

Remark 3.4. *We observe that we can prove a theorem like Theorem (3.5) for a general solution* (F, G, B) *of system (17) by taking a generic* $F = (f_{ij})$ *and* $B = (b_{ij}) = \begin{pmatrix} b & 0 \\ 0 & b^{-1} \end{pmatrix} \in O(1,1)$, *i.e, we conclude that if* (F, G, B) *is a solution of system (17), we obtain a complex isothermic timelike dual pair in* $\mathbb{R}^{n-j,j}$ *of type* $O(1,1)$ *with I and II fundamental forms like in (20).*

Now for the converse, we have the following result.

Theorem 3.6. *Let (X_1, X_2) be a complex isothermic timelike dual pair in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$, $\{e_2, ..., e_{n-1}\}$ a common parallel normal frame and (x_1, x_2) a common isothermal coordinates for X_1 and X_2, such that the two fundamental forms I_k, II_k for X_k are given by (20). Set $f_{11} = \frac{b_{x_2}}{2b} = -f_{22}$, $f_{12} = \frac{b_{x_1}}{2b} = f_{21}$, and $F = (f_{ij})_{2\times 2}$. Then if all entries of G are non-zero, then (F, G, B) is a solution of (17).*

Proof. From the definition of complex isothermic timelike dual pair in $\mathbb{R}^{n-j,j}$, we have

$$\omega_1^{(1)} = b^{-1}dx_1, \quad \omega_n^{(1)} = b^{-1}dx_2, \quad \omega_1^{(2)} = -bdx_2, \quad \omega_n^{(2)} = bdx_1$$

is a dual 1-frame for X_k and $\omega_{1\alpha}^{(k)} = l_\alpha(-g_{\alpha-1,2}dx_1 + g_{\alpha-1,1}dx_2)$, $\omega_{n\alpha}^{(k)} = -l_\alpha(g_{\alpha-1,1}dx_1 + g_{\alpha-1,2}dx_2)$ for each X_k, where $l_\alpha = 1$ if $\alpha = 2, ..., n-j$ and $l_\alpha = -1$ if $\alpha = n-j+1, ..., n-1$. We observe that $\omega_{i\alpha}^{(k)}$, $i = 1, n$, $\alpha = 2, ..., n-1$ are independent of k. We find that the Levi-civita connection 1-form for the metric I_k is:

$$\omega_{1n}^{(k)} = -\frac{b_{x_2}}{b}dx_1 - \frac{b_{x_1}}{b}dx_2,$$

which are independent from k. Hence $\omega_{1n}^{(k)} = \omega_{1n}^{(1)} = 2(f_{22}dx_1 - f_{12}dx_2) = -2(\xi_1 dx_1 + \xi_2 dx_2)$. So the structure equations and the Gauss-Codazzi equations for X_1, X_2 imply that (F, G, B) is a solution of system (17). ■

So, from Theorems (3.5), (3.6) and Remark (3.4), follows that exists a correspondence between the solutions (F, G, B) of system (17) and a dual pair of complex isothermic timelike surfaces in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$.

Theorem 3.7. *The complex $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II (17) is the Gauss-Codazzi equation for a time-like surface in $\mathbb{R}^{n-j,j}$ such that:*

$$\begin{cases} I = e^{4u}(dx_2^2 - dx_1^2) \\ II = e^{2u}\sum_{i=1}^{n-2}[r_{i,1}(dx_2^2 - dx_1^2) - 2r_{i,2}dx_1dx_2]e_{i+1}. \end{cases} \tag{24}$$

Proof. For this surface we can read off from the fundamental forms I and II that

$$\begin{cases} \omega_1 = -e^{2u}dx_2 \\ \omega_n = e^{2u}dx_1 \\ \omega_{1\,i} = \sigma_i(r_{i-1,1}dx_2 - r_{i-1,2}dx_1) \quad \text{for} \quad 2 \le i \le n-1 \\ \omega_{n\,i} = -\sigma_i(r_{i-1,1}dx_1 + r_{i-1,2}dx_2) \quad \text{for} \quad 2 \le i \le n-1. \end{cases}$$

Using the structure equations, we can see that

$$\omega_{1n} = -2u_{x_2}dx_1 - 2u_{x_1}dx_2,$$

and that the Gauss and Codazzi equations are the same as (18), since we have

$$u_{x_1} = \xi_2, \qquad u_{x_2} = \xi_1.$$

Hence if we put

$$B = \begin{pmatrix} e^{2u} & 0 \\ 0 & e^{-2u} \end{pmatrix}, \quad F = \begin{pmatrix} u_{x_2} & u_{x_1} \\ u_{x_1} & -u_{x_2} \end{pmatrix}, \tag{25}$$
$$G = \begin{pmatrix} r_{1,1} & r_{1,2} \\ \vdots & \vdots \\ r_{n-2,1} & r_{n-2,2} \end{pmatrix}$$

we have that (F, G, B) is solution of the complex $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II (17). Conversely, if (F, G, B) is solution of the complex $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II (17), and B is as in (25), then the fourth and sixth equation from system (17), imply that

$$\xi_2 = u_{x_1}, \;\; \xi_1 = u_{x_2}$$

i.e., (F, G, B) has the form (25). Finally writing the $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II (17) for this (F, G, B), in terms of u and r_i we get equation (18). ■

The next result follows from Theorem (3.5) and Theorem (3.7).

Theorem 3.8. *Let $\mathcal{O}$ be a domain of $\mathbb{R}^{1,1}$, and $X_2 : \mathcal{O} \to \mathbb{R}^{n-j,j}$ an immersion with flat normal bundle and $(x_1, x_2) \in \mathcal{O}$ a isothermal coordinates system with respect to a parallel normal frame $\{e_2, ..., e_{n-1}\}$, such that I and II fundamental forms are given by (24). Then there exists an immersion X_1, unique up to translation, such that (X_1, X_2) is a complex isothermic timelike dual pair in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$. Moreover, the fundamental forms of X_1, X_2 are respectively:*

$$\begin{cases} I_1 = e^{-4u}(dx_1^2 - dx_2^2) \\ II_1 = -e^{-2u} \sum_{i=1}^{n-2} [r_{i,2}(dx_2^2 - dx_1^2) + 2r_{i,1} dx_1 dx_2] e_{i+1} \\ I_2 = e^{4u}(dx_2^2 - dx_1^2) \\ II_2 = e^{2u} \sum_{i=1}^{n-2} [r_{i,1}(dx_2^2 - dx_1^2) - 2r_{i,2} dx_1 dx_2] e_{i+1}. \end{cases} \tag{26}$$

Finally, it follows from Gauss equation that the Gaussian curvatures of X_1 and X_2 of a complex isothermic timelike dual pair (26), denoted by $K_G^{(1)}$, $K_G^{(2)}$, are given by

$$K_G^{(1)} = e^{4u} \sum_{i=1}^{n-2} \sigma_i (r_{i,1}^2 + r_{i,2}^2), \qquad K_G^{(2)} = e^{-4u} \sum_{i=1}^{n-2} \sigma_i (r_{i,1}^2 + r_{i,2}^2),$$

where $\sigma_i = 1, \; i = 1, ..., n-j-1$ and $\sigma_i = -1, \; i = n-j, ..., n-2$.

Example: Next we give an explicit example of a dual pair of complex timelike isothermic surfaces in $\mathbb{R}^{2,1}$ and the associated solution to the complex $O(3,2)/O(2,1) \times O(1,1)$-system II.

We consider first the Lorentzian helicoid

$$X(x_1, x_2) = (x_2, \sinh(x_1)\sinh(x_2), \cosh(x_2)\sinh(x_1))$$

with normal vector:

$$N(x_1, x_2) = \frac{1}{\cosh(x_1)}(-\sinh(x_1), \cosh(x_2), \sinh(x_2)).$$

The dual surface to this surface is:

$$\widehat{X}(x_1, x_2) = \frac{1}{\cosh(x_1)}(\sinh(x_1), -\cosh(x_2), -\sinh(x_2)),$$

which is a parametrization of part of the standard immersion of the Lorenztian sphere (see[11]). They constitute a dual pair of complex timelike isothermic surfaces in $\mathbb{R}^{2,1}$, with first and second fundamental forms given resp. by

$$I^1 = \cosh^2(x_1)[-dx_1^2 + dx_2^2], \quad II^1 = 2dx_1dx_2,$$

$$I^2 = (1/\cosh^2(x_1))[dx_1^2 - dx_2^2], \quad II^2 = (1/\cosh^2(x_1))[dx_2^2 - dx_1^2].$$

Here

$$B = \begin{pmatrix} \cosh x_1 & 0 \\ 0 & \cosh^{-1} x_1 \end{pmatrix}, \ F = \begin{pmatrix} 0 & \frac{\tanh x_1}{2} \\ \frac{\tanh x_1}{2} & 0 \end{pmatrix},$$

and

$$G = \begin{pmatrix} 0, & -\cosh^{-1} x_1 \end{pmatrix},$$

are a solution of the complex $O(3,2)/O(2,1) \times O(1,1)$-system II. More specifically, taking $e^{2u} = \cosh x_1$, we have $(u, 0, -\cosh^{-1} x_1)$ is a solution of the complex $O(3,2)/O(2,1) \times O(1,1)$-system II.

4 Appendix: Associated Geometric transformations

The first part of this appendix concerns the geometric transformations on surfaces in the pseudo-euclidean space $\mathbb{R}^{n-j,j}$ corresponding to the action of an element with two simple poles on the space of local solutions of our complex $O(n-j+1, j+1)/O(n-j, j) \times O(1,1)$-system II (17). In particular, the results which will be established here were proved by the authors in [7], hence we invite the reader to see in [7] the proof's details. In addition, the reader will find in [7], an explicit example of an isothermic timelike dual pair in $\mathbb{R}^{2,1}$ of type $O(1,1)$ constructed by applying the Darboux transformation to the trivial solution of complex system II (18). We note that the study of the geometric transformations associated to the real case, was already considered in [14].

In the second part of this appendix, we establish the moving frame formulas for timelike surfaces in $\mathbb{R}^{n-j,j}$.

Initially in [7], we made a natural extension of the Ribaucour transformation definition given in [5], and of the definition of Darboux transformation for surfaces in $\mathbb{R}^m$ for our case of complex timelike surfaces. Later, we found the rational element $g_{s,\pi}$ whose action corresponds to the Ribaucour and Darboux transformations just as we defined them. We now review the principal results of [7].

We start by defining Ribaucour and Darboux transformations for timelike surfaces in $\mathbb{R}^{n-j,j}$ whose shape operators have conjugate eigenvalues as follows:
For $x \in \mathbb{R}^{n-j,j}$ and $v \in (T\mathbb{R}^{n-j,j})_x$, where let $\gamma_{x,v}(t) = x + tv$ denote the geodesic starting at x in the direction of v.

Definition 4.1. *Let M^m and $\widetilde{M}^m$ be Lorentzian submanifolds whose shape operators are all diagonalizable over $\mathbb{R}$ or $\mathbb{C}$ immersed in the pseudo-riemannian space $\mathbb{R}^{n-j,j}$, $0 \leq j < n$. A sphere congruence is a vector bundle isomorphism $P : \mathcal{V}(M) \to \mathcal{V}(\widetilde{M})$ that covers a diffeomorphism $\phi : M \to \widetilde{M}$ with the following conditions:*
(1) If ξ is a parallel normal vector field of M, then $P \circ \xi \circ \phi^{-1}$ is a parallel normal field of $\widetilde{M}$.
(2) For any nonzero vector $\xi \in \mathcal{V}_x(M)$, the geodesics $\gamma_{x,\xi}$ and $\gamma_{\phi(x),P(\xi)}$ intersect at a point that is the same parameter value t away from x and $\phi(x)$.

For the following definition we assume that each shape operator is diagonalized over the real or complex numbers. We note that there are submanifolds for which this does not hold.

Definition 4.2. *A sphere congruence $P : \mathcal{V}(M) \to \mathcal{V}(\widetilde{M})$ that covers a diffeomorphism $\phi : M \to \widetilde{M}$ is called a Ribaucour transformation if it satisfies the following additional properties:*
(1) If e is an eigenvector of the shape operator A_ξ of M, corresponding to a real eigenvalue then $\phi_*(e)$ is an eigenvector of the shape operator $A_{P(\xi)}$ of $\widetilde{M}$ corresponding to a real eigenvalue.
If $e_1 + ie_2$ is an eigenvector of A_ξ on $(TM)^{\mathbb{C}}$ corresponding to the complex eigenvalue $a + ib$ (so that $e_1 - ie_2$ corresponds to the eigenvalue $a - ib$), then $\phi_*(e_1) + i\phi_*(e_2)$ is an eigenvector corresponding to a complex eigenvalue for $A_{P(\xi)}$.
(2) The geodesics $\gamma_{x,e}$ and $\gamma_{\phi(x),\phi_*(e)}$ intersect at a point that is equidistant to x and $\phi(x)$ for real eigenvectors e, and γ_{x,e_j} and $\gamma_{\phi(x),\phi_*(e_j)}$ meet for the real and imaginary parts of complex eigenvectors $e_1 + ie_2$, i.e., for $j = 1, 2$.

Definition 4.3. *Let $M, \widetilde{M}$ be two timelike surfaces in $\mathbb{R}^{n-j,j}$ with flat and non-degenerate normal bundle, shape operators that are diagonalizable over $\mathbb{C}$ and $P : \mathcal{V}(M) \to \mathcal{V}(\widetilde{M})$ a Ribaucour transformation that covers the map $\phi : M \to \widetilde{M}$. If, in addition, ϕ is a sign-reversing conformal diffeomorphism then P is called a Darboux transformation.*

In definition (4.3), by a sign-reversing conformal diffeomorphism we mean that the time-like and space like vectors are interchanged and the conformal coordinates remain conformal.
Next we define the rational element

$$g_{s,\pi}(\lambda) = (\pi + \frac{\lambda - is}{\lambda + is}(I - \pi))(\overline{\pi} + \frac{\lambda + is}{\lambda - is}(I - \overline{\pi})) \tag{27}$$

where $0 \neq s \in \mathbb{R}$, π is the orthogonal projection of $\mathbb{C}^{n+2}$ onto the span of $\begin{pmatrix} W \\ iZ \end{pmatrix}$ with respect to the bi-linear form $\langle , \rangle_2$ given by

$$\langle U, V \rangle_2 =$$
$$\overline{u}_1 v_1 + ... + \overline{u}_{n-j} v_{n-j} - \overline{u}_{n-j+1} v_{n-j+1} - ... - \overline{u}_n v_n + \overline{u}_{n+1} v_{n+2} + \overline{u}_{n+2} v_{n+1},$$

for $W \in \mathbb{R}^{n-j,j}$, $Z \in \mathbb{R}^{1,1}$ unit vectors.
It is easy to see that $g_{s,\pi}$ belongs to the group:

$$G_- = \{g : S^2 \to U_{\mathbb{C}} \mid g \text{ is meromorphic, } g(\infty) = I \text{ and satisfies the reality conditions}\},$$

where $U_{\mathbb{C}} = O(n-j+1, j+1; \mathbb{C})$ and the reality conditions are the following, for a map $g : \mathbb{C} \to U_{\mathbb{C}}$:

$$\begin{cases} \overline{g(\overline{\lambda})} = g(\lambda) \\ I_{n,2}\ g(-\lambda)\ I_{n,2} = g(\lambda) \\ g(\lambda)^t \begin{pmatrix} I_{n-j,j} & 0 \\ 0 & J' \end{pmatrix} g(\lambda) = \begin{pmatrix} I_{n-j,j} & 0 \\ 0 & J' \end{pmatrix}. \end{cases} \tag{28}$$

With this, we have:

Theorem 4.1. *Let (X_1, X_2) be a complex isothermic timelike dual pair in $\mathbb{R}^{n-j,j}$ of type $O(1,1)$ corresponding to the solution (u, G) of the system (18), and let $\xi = \begin{pmatrix} F \\ G \end{pmatrix}$ the corresponding solution of the system (16), where*

$$F = \begin{pmatrix} u_{x_2} & u_{x_1} \\ u_{x_1} & -u_{x_2} \end{pmatrix}, B = \begin{pmatrix} e^{2u} & 0 \\ 0 & e^{-2u} \end{pmatrix}.$$

Let $g_{s,\pi}$ defined in (27), and $\widehat{W}$, $\widehat{Z}$ as in Main Lemma 4.1 (see below), for the solution ξ of the system (16). Let $(\widetilde{E^\sharp}^{II}, \widetilde{A}^\sharp, \widetilde{B}^\sharp) = g_{s,\pi}.(E^{II}, A, B)$ the action of $g_{s,\pi}$ over (E^{II}, A, B) where $A, B, \widetilde{A}^\sharp, \widetilde{B}^\sharp$ are the entries of

$$E(x, 0) = \begin{pmatrix} A(x) & 0 \\ 0 & B(x) \end{pmatrix}, \quad \widetilde{E^\sharp}(x, 0) = \begin{pmatrix} \widetilde{A}^\sharp(x) & 0 \\ 0 & \widetilde{B}^\sharp(x) \end{pmatrix}$$

and E^{II} is the frame corresponding to the solution (F,G,B) of the complex system II (18). Write $A=(e_1,...,e_n)$ and $\widetilde{A}^\sharp=(\widetilde{e}_1,...,\widetilde{e}_n)$. Set

$$\begin{cases}\widetilde{X}_1=X_1+\frac{2}{s}\widehat{z}_2e^{-2u}\sum_{i=1}^n\widehat{w}_ie_i,\\ \widetilde{X}_2=X_2+\frac{2}{s}\widehat{z}_1e^{2u}\sum_{i=1}^n\widehat{w}_ie_i,\end{cases}\tag{29}$$

Then
(i) $(\widetilde{u},\widetilde{G})$ is the solution of system (18), corresponding to $\widetilde{X}=(\widetilde{X}_1,\widetilde{X}_2)$, where $e^{4\widetilde{u}}=\frac{4\widehat{z}_2^4}{e^{4u}}$ and $\widetilde{G}=(\widetilde{r}_{ij})$ is defined by Main Lemma 4.1, for the new solution $\widetilde{\xi}$ of the system (16).
(ii) The fundamental forms of pair $(\widetilde{X}_1,\widetilde{X}_2)$ are respectively

$$\begin{cases}\widetilde{I}_1=e^{4\widetilde{u}}(-dx_1^2+dx_2^2)\\ \widetilde{II}_1=e^{2\widetilde{u}}\sum_{i=1}^{n-2}[\widetilde{r}_{i,1}(dx_2^2-dx_1^2)-2\widetilde{r}_{i,2}dx_1dx_2]\widetilde{e}_{i+1}.\\ \widetilde{I}_2=e^{-4\widetilde{u}}(dx_1^2-dx_2^2)\\ \widetilde{II}_2=-e^{-2\widetilde{u}}\sum_{i=1}^{n-2}\widetilde{r}_{i,2}(dx_2^2-dx_1^2)+2\widetilde{r}_{i,1}dx_1dx_2]\widetilde{e}_{i+1}.\end{cases}$$

(iii) The bundle morphism $P(e_k(x))=\widetilde{e}_k(x)$, $k=2,...,n-1$ covering the map $X_i\to\widetilde{X}_i$ is a Darboux transformation for each $i=1,2$.

Proof. For (i) and (ii) we just observe that

$$d\widetilde{X}=\widetilde{A}^\sharp\begin{pmatrix}dx_1&0&\dots&0&dx_2\\-dx_2&0&\dots&0&dx_1\end{pmatrix}^t\widetilde{B^\sharp}^{-1},$$

and calculate.
For (iii) we observe that the map $\phi:X_i\to\widetilde{X}_i$ is sign-reversing conformal because the coordinates (x_1,x_2) are isothermic for X_i and $\widetilde{X}_i$ but timelike and spacelike vectors are interchanged. The rest of the properties of Darboux transformation follows from Lemma 4.2 below. ■

Lemma 4.1. *(Main Lemma) Let $\xi=\begin{pmatrix}F\\G\end{pmatrix}$ be a solution of the system (16), and $E(x,\lambda)$ a frame of ξ such that $E(x,\lambda)$ is holomorphic for $\lambda\in\mathbb{C}$. Let $g_{s,\pi}$ the map defined by (27) and $\widetilde{\pi}(x)$ the orthogonal projection onto $\mathbb{C}\begin{pmatrix}\widetilde{W}\\i\widetilde{Z}\end{pmatrix}(x)$ with respect to $\langle,\rangle_2$, where*

$$\begin{pmatrix}\widetilde{W}\\i\widetilde{Z}\end{pmatrix}(x)=E(x,-is)^{-1}\begin{pmatrix}W\\iZ\end{pmatrix}.\tag{30}$$

Let $\widehat{W}=\frac{\widetilde{W}}{\|\widetilde{W}\|_{n-j,j}}$ and $\widehat{Z}=\frac{\widetilde{Z}}{\|\widetilde{Z}\|_{1,1}}$, $\widetilde{E}(x,\lambda)=g_{s,\pi}(\lambda)E(x,\lambda)g_{s,\widetilde{\pi}(x)}(\lambda)^{-1}$,

$$\widetilde{\xi}=\xi-2s(\widehat{W}\widehat{Z}^tJ')_*,\tag{31}$$

where (ϑ_*) is the projection onto the span of $\{a_1,a_2\}^\perp$. Then $\widetilde{\xi}$ is a solution of system (16), $\widetilde{E}$ is a frame for $\widetilde{\xi}$ and $\widetilde{E}(x,\lambda)$ is holomorphic in $\lambda\in\mathbb{C}$.

For the Proof of the Main Lemma see ([7]).

Writing the new solution given by Main Lemma 4.1 as $\widetilde{\xi} = \begin{pmatrix} \widetilde{F} \\ \widetilde{G} \end{pmatrix}$, one sees the components of $\widetilde{\xi}$ are:

$$\begin{cases} \widetilde{f}_{11} = -\widetilde{f}_{22} = f_{11} - s(\widehat{w}_1\widehat{z}_2 - \widehat{w}_n\widehat{z}_1), \\ \widetilde{f}_{12} = \widetilde{f}_{21} = f_{12} - s(\widehat{w}_1\widehat{z}_1 + \widehat{w}_n\widehat{z}_2), \\ \widetilde{r}_{i1} = r_{i1} - 2s\widehat{w}_{1+i}\widehat{z}_2 \\ \widetilde{r}_{i2} = r_{i2} - 2s\widehat{w}_{1+i}\widehat{z}_1, \end{cases} \tag{32}$$

for $F = (f_{ij})_{2\times 2},\ G = (r_{ij})_{(n-2)\times 2},\ \widetilde{F} = (\widetilde{f}_{ij})_{2\times 2},\ \widetilde{G} = (\widetilde{r}_{ij})_{(n-2)\times 2}$.

Lemma 4.2. *Let* $\xi = \begin{pmatrix} F \\ G \end{pmatrix}$ *solution of (16),* E *frame of* ξ*,* $E(x,0) = \begin{pmatrix} A(x) & 0 \\ 0 & B(x) \end{pmatrix}$*,* (F,G,B) *a solution corresponding to complex* $O(n-j+1, j+1)/O(n-j,j)\times O(1,1)$*-system II, and*

$$(\widetilde{F}, \widetilde{G}, \widetilde{B}^\sharp, \widetilde{E^\sharp}^{II}) = g_{s,\pi}.(F,G,B,E^{II}), \quad \widetilde{A}^\sharp = g_{s,\pi}.A$$

the action of $g_{s,\pi}$ *over the solution* (F,G,B) *and the matrix* A*, resp.. Let* $e_i, \widetilde{e}_i$ *denote the i-th column of* A *and* $\widetilde{A}^\sharp$ *resp. Then we have*
(i) $X = (X_1, X_2)$ *and* $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2)$ *are complex isothermic timelike dual pairs in* $\mathbb{R}^{n-j,j}$ *of type* $O(1,1)$ *such that* $\{e_2, ...e_{n-1}\}$ *and* $\{\widetilde{e}_2, ..., \widetilde{e}_{n-1}\}$ *are parallel normal frames for* X_j *and* $\widetilde{X}_j$ *respectively for* $j = 1, 2$*, where* $\{e_\alpha\}_{\alpha=2}^{n-j}$ *and* $\{e_\alpha\}_{\alpha=n-j+1}^{n-1}$ *are spacelike and timelike vectors resp.*
(ii) The solutions of the complex $O(n-j+1, j+1)/O(n-j,j)\times O(1,1)$*-system II corresponding to* X *and* $\widetilde{X}$ *are* (F,G,B) *and* $(\widetilde{F}, \widetilde{G}, \widetilde{B}^\sharp)$ *resp.*
(iii) The bundle morphism $P(e_k(x)) = \widetilde{e}_k(x)$ $k = 2, ..., n-1$*, is a Ribaucour Transformation covering the map* $X_j(x) \mapsto \widetilde{X}_j(x)$ *for each* $j = 1, 2$*.*
(iv) There exist smooth functions ψ_{ik} *such that* $X_i + \psi_{ik}e_k = \widetilde{X}_i + \psi_{ik}\widetilde{e}_k$ *for* $1 \le i \le 2$ *and* $1 \le k \le n$*.*

For the proof of Lemma 4.2, see ([7]).

Now we begin the second part of this appendix, where we review the method of moving frames for time-like surfaces in the Lorentz space $\mathbb{R}^{n-j,j}$. Set

$$e_A \cdot e_B = \sigma_{AB} = I_{n-j,j} = \begin{pmatrix} I_{n-j} & 0 \\ 0 & -I_j \end{pmatrix}.$$

We also let $\sigma_i := \sigma_{ii}$.

For the time-like immersion X set $dX = \omega_1 e_1 + \omega_n e_n$, so that a space-like unit tangent vector to the surface is e_1, a time-like unit vector to the surface is e_n and the normal space is spanned by e_α, for $2 \leq \alpha \leq n-1$. Define

$$de_B = \sum_A \omega_{AB} e_A. \tag{33}$$

This gives $\omega_{AB} = \sigma_A e_A \cdot de_B$ and

$$\omega_{AB}\sigma_A + \omega_{BA}\sigma_B = 0. \tag{34}$$

From $d(dX) = 0$ we get:

$$d\omega_1 = \omega_n \wedge \omega_{1n} \tag{35}$$

$$d\omega_n = \omega_1 \wedge \omega_{n1} \tag{36}$$

$$\omega_1 \wedge \omega_{\alpha 1} + \omega_n \wedge \omega_{\alpha n} = 0, \tag{37}$$

for α as above.
In addition, by Cartan's Lemma we have:

$$\omega_{1\alpha} = h^\alpha_{11}\omega_1 + h^\alpha_{1n}\omega_n, \quad \omega_{n\alpha} = h^\alpha_{n1}\omega_1 + h^\alpha_{nn}\omega_n.$$

This makes the first fundamental form:

$$I : \omega_1^2 - \omega_n^2 \tag{38}$$

and the second fundamental form is:

$$\begin{aligned} II : \quad & -\textstyle\sum_{k=1,n} \sum_\alpha \omega_{k\alpha}\sigma_k\,\omega_k\,\sigma_\alpha\, e_\alpha = \\ & -\textstyle\sum_\alpha (h^\alpha_{11}\omega_1 + h^\alpha_{1n}\omega_n)\,\omega_1\,\sigma_\alpha\, e_\alpha + \sum_\alpha (h^\alpha_{n1}\omega_1 + h^\alpha_{nn}\omega_n)\,\omega_n\,\sigma_\alpha\, e_\alpha. \end{aligned} \tag{39}$$

We also have: $d\omega_{CA} = -\sum_B \omega_{CB} \wedge \omega_{BA}$, which yield the Gauss and Codazzi equations. The Gauss equation comes from examining $d\omega_{1n}$, while the Codazzi equations are from $d\omega_{1\alpha}$ and $d\omega_{n\alpha}$.

References

[1] Burstall, F. *Isothermic surfaces: conformal geometry, Clifford algebras and integrable systems*, preprint, math-dg/0003096.

[2] Bruck, M., Du, X., Park, J., Terng, C-L., *The Submanifold Geometries associated to Grassmannian Systems*, Memoirs of A.M.S. no. 735 (2002).

[3] Burstall, F., Hertrich-J U., Pedit F., Pinkall U., *Curved flats and isothermic surfaces*, Math. Z. no. 2, 225 (1997).

[4] Cieslinski, J. Goldstein, P. Sym, A., *Isothermic surfaces in* $\mathbb{R}^3$ *as soliton surfaces*, Phys. Lett. A. 205 (1995), 37-43.

[5] Dajczer, M., Tojeiro, R., *Commuting codazzi tensors and the ribaucour transformations for submanifolds*, Results in MatH. 44 (2003), 258-278.

[6] Dussan, M.P., *Space-like isothermic surfaces and Grassmannian systems.* Submitted for publication.

[7] Dussan, M.P., Magid, M., *Complex Timelike Isothermic Surfaces and their Geometric transformations.* to appear in Balkan Journal of Geometry and its Applications.

[8] Ferus D., Pedit F., *Curved flats in symmetric space*, Manuscripta Math. 91 (1996), 445-454.

[9] C.H. Gu, H.S. Hu, Jun-Ichi Inoguchi, *On time-like surfaces of positive constant Gaussian curvature and imaginary principal curvatures*, J. Geom. and Phys. 41 (2002), 296-311.

[10] Inoguchi, J-I.,*Darboux transformations on timelike constant mean curvature surfaces*, J. Geom. and Phys. 32 (1999), 57-78.

[11] Magid, M. A., *Lorentzian Isothermic Surfaces in* $\mathbf{R}^n_j$, to appear in Rocky Mountain J. M.

[12] Terng, C.-L. *Solitons equations and differential geometry*, J. Differential Geom. 45, (1997), no. 2, 407-445.

[13] Terng, C-L., Uhlenbeck, K., *Backlund transformations and loop group actions*, Comm. Pure. Appl. Math. 53, (2000), 1-75. 407-445.

[14] Zuo, D., Chen, Q., Cheng, Y., $G^{p,q}_{m,n}$*-System II and diagonalizable timelike immersions in* $R^{p,m}$, to appear in IOP.

M. P. Dussan
Department of Mathematics
IME - USP
São Paulo, Brazil
dussan@ime.usp.br

M. A. Magid
Department of Mathematics
Wellesley College
Wellesley, MA
mmagid@wellesley.edu

On Triangulated Orbit Categories

Dedicated to Claus Michael Ringel
on the occasion of his sixtieth birthday

Bernhard Keller

Received: June 8, 2005

Communicated by Peter Schneider

Abstract. We show that the category of orbits of the bounded derived category of a hereditary category under a well-behaved autoequivalence is canonically triangulated. This answers a question by Aslak Buan, Robert Marsh and Idun Reiten which appeared in their study [8] with M. Reineke and G. Todorov of the link between tilting theory and cluster algebras (*cf.* also [16]) and a question by Hideto Asashiba about orbit categories. We observe that the resulting triangulated orbit categories provide many examples of triangulated categories with the Calabi-Yau property. These include the category of projective modules over a preprojective algebra of generalized Dynkin type in the sense of Happel-Preiser-Ringel [29], whose triangulated structure goes back to Auslander-Reiten's work [6], [44], [7].

2000 Mathematics Subject Classification: Primary 18E30; Secondary 16G20.
Keywords and Phrases: Derived category, Triangulated category, Orbit category, Calabi-Yau category

1 Introduction

Let $\mathcal{T}$ be an additive category and $F : \mathcal{T} \to \mathcal{T}$ an automorphism (a standard construction allows one to replace a category with autoequivalence by a category with automorphism). Let $F^{\mathbf{Z}}$ denote the group of automorphisms generated by F. By definition, the *orbit category* $\mathcal{T}/F = \mathcal{T}/F^{\mathbf{Z}}$ has the same objects as $\mathcal{T}$ and its morphisms from X to Y are in bijection with

$$\bigoplus_{n\in\mathbf{Z}} \mathsf{Hom}_{\mathcal{T}}(X, F^n Y).$$

The composition is defined in the natural way (*cf.* [17], where this category is called the skew category). The canonical projection functor $\pi : \mathcal{T} \to \mathcal{T}/F$ is endowed with a natural isomorphism $\pi \circ F \xrightarrow{\sim} \pi$ and 2-universal among such functors. Clearly $\mathcal{T}/F$ is still an additive category and the projection is an additive functor. Now suppose that $\mathcal{T}$ is a triangulated category and that F is a triangle functor. Is there a triangulated structure on the orbit category such that the projection functor becomes a triangle functor? One can show that in general, the answer is negative. A closer look at the situation even gives one the impression that quite strong assumptions are needed for the answer to be positive. In this article, we give a sufficient set of conditions. Although they are very strong, they are satisfied in certain cases of interest. In particular, one obtains that the cluster categories of [8], [16] are triangulated. One also obtains that the category of projective modules over the preprojective algebra of a generalized Dynkin diagram in the sense of Happel-Preiser-Ringel [29] is triangulated, which is also immediate from Auslander-Reiten's work [6], [44], [7]. More generally, our method yields many easily constructed examples of triangulated categories with the Calabi-Yau property.
Our proof consists in constructing, under quite general hypotheses, a 'triangulated hull' into which the orbit category $\mathcal{T}/F$ embeds. Then we show that under certain restrictive assumptions, its image in the triangulated hull is stable under extensions and hence equivalent to the hull.
The contents of the article are as follows: In section 3, we show by examples that triangulated structures do not descend to orbit categories in general. In section 4, we state the main theorem for triangulated orbit categories of derived categories of hereditary algebras. We give a first, abstract, construction of the triangulated hull of an orbit category in section 5. This construction is based on the formalism of dg categories as developped in [32] [21] [57]. Using the natural t-structure on the derived category of a hereditary category we prove the main theorem in section 6.
We give a more concrete construction of the triangulated hull of the orbit category in section 7. In some sense, the second construction is 'Koszul-dual' to the first: whereas the first construction is based on the tensor algebra

$$T_A(X) = \oplus_{n=0}^{\infty} X^{\otimes_A n}$$

of a (cofibrant) differential graded bimodule X over a differential graded algebra A, the second one uses the 'exterior algebra'

$$A \oplus X^{\wedge}[-1]$$

on its dual $X^{\wedge} = \mathsf{RHom}_A(X_A, A)$ shifted by one degree. In the cases considered by Buan et al. [8] and Caldero-Chapoton-Schiffler [16], this also yields an

interesting new description of the orbit category itself in terms of the stable category [40] of a differential graded algebra.
In section 8, we observe that triangulated orbit categories provide easily constructed examples of triangulated categories with the Calabi-Yau property. Finally, in section 9, we characterize our constructions by universal properties in the 2-category of enhanced triangulated categories. This also allows us to examine their functoriality properties and to formulate a more general version of the main theorem which applies to derived categories of hereditary categories which are not necessarily module categories.

2 Acknowledgments

The author thanks C. M. Ringel for his interest and encouragement throughout the years. He is indebted to A. Buan, R. Marsh and I. Reiten for the question which motivated the main theorem of this article and to H. Asashiba for suggesting a first generalization of its original formulation. He is grateful to H. Lenzing for pointing out the need to weaken its hypotheses to include orbit categories of general hereditary categories. He expresses his thanks to C. Geiss and J.-A. de la Peña for the hospitality he enjoyed at the Instituto de Matemáticas, UNAM, in August 2004 while the material of section 8 was worked out. He thanks R. Bocklandt, W. Crawley-Boevey, K. Erdmann, C. Geiss, A. Neeman, I. Reiten, A. Rudakov, J. Schröer, G. Tabuada, M. Van den Bergh, Feng Ye, Bin Zhu and an anonymous referee for helpful comments and corrections.

3 Examples

Let $\mathcal{T}$ be a triangulated category, $F : \mathcal{T} \to \mathcal{T}$ an autoequivalence and $\pi : \mathcal{T} \to \mathcal{T}/F$ the projection functor. In general, a morphism

$$u : X \to Y$$

of $\mathcal{T}/F$ is given by a morphism

$$X \to \bigoplus_{i=1}^{N} F^{n_i} Y$$

with N non vanishing components $u_1, \ldots, u_N$ in $\mathcal{T}$. Therefore, in general, u does not lift to a morphism in $\mathcal{T}$ and it is not obvious how to construct a 'triangle'

$$X \xrightarrow{u} Y \longrightarrow Z \longrightarrow SZ$$

in $\mathcal{T}/F$. Thus, the orbit category $\mathcal{T}/F$ is certainly not trivially triangulated. Worse, in 'most' cases, it is impossible to endow $\mathcal{T}/F$ with a triangulated structure such that the projection functor becomes a triangle functor. Let us consider three examples where $\mathcal{T}$ is the bounded derived category $\mathcal{D}^b(A) = \mathcal{D}^b(\mathsf{mod}A)$ of the category of finitely generated (right) modules $\mathsf{mod}A$ over an algebra A of finite dimension over a field k. Thus the objects of $\mathcal{D}^b(A)$ are the complexes

$$M = (\ldots \to M^p \to M^{p+1} \to \ldots)$$

of finite-dimensional A-modules such that $M^p = 0$ for all large $|p|$ and morphisms are obtained from morphisms of complexes by formally inverting all quasi-isomorphisms. The suspension functor S is defined by $SM = M[1]$, where $M[1]^p = M^{p+1}$ and $d_{M[1]} = -d_M$, and the triangles are constructed from short exact sequences of complexes.

Suppose that A is hereditary. Then the orbit category $\mathcal{D}^b(A)/S^2$, first introduced by D. Happel in [27], is triangulated. This result is due to Peng and Xiao [43], who show that the orbit category is equivalent to the homotopy category of the category of 2-periodic complexes of projective A-modules.

On the other hand, suppose that A is the algebra of dual numbers $k[X]/(X^2)$. Then the orbit category $\mathcal{D}^b(A)/S^2$ is not triangulated. This is an observation due to A. Neeman (unpublished). Indeed, the endomorphism ring of the trivial module k in the orbit category is isomorphic to a polynomial ring $k[u]$. One checks that the endomorphism $1+u$ is monomorphic. However, it does not admit a left inverse (or else it would be invertible in $k[u]$). But in a triangulated category, each monomorphism admits a left inverse.

One might think that this phenomenon is linked to the fact that the algebra of dual numbers is of infinite global dimension. However, it may also occur for algebras of finite global dimension: Let A be such an algebra. Then, as shown by D. Happel in [27], the derived category $\mathcal{D}^b(A)$ has Auslander-Reiten triangles. Thus, it admits an autoequivalence, the Auslander-Reiten translation τ, defined by

$$\mathsf{Hom}(?, S\tau M) \xrightarrow{\sim} D\mathsf{Hom}(M, ?) ,$$

where D denotes the functor $\mathsf{Hom}_k(?, k)$. Now let Q be the Kronecker quiver

$$1 \rightrightarrows 2 \ .$$

The path algebra $A = kQ$ is finite-dimensional and hereditary so Happel's theorem applies. The endomorphism ring of the image of the free module A_A in the orbit category $\mathcal{D}^b(A)/\tau$ is the preprojective algebra $\Lambda(Q)$ (*cf.* section 7.3). Since Q is not a Dynkin quiver, it is infinite-dimensional and in fact contains a polynomial algebra (generated by any non zero morphism from the simple projective P_1 to $\tau^{-1}P_1$). As above, it follows that the orbit category does not admit a triangulated structure.

4 The main theorem

Assume that k is a field, and $\mathcal{T}$ is the bounded derived category $\mathcal{D}^b(\mathsf{mod}A)$ of the category of finite-dimensional (right) modules $\mathsf{mod}A$ over a finite-dimensional k-algebra A. Assume that $F : \mathcal{T} \to \mathcal{T}$ is a standard equivalence

[48], *i.e.* F is isomorphic to the derived tensor product

$$? \overset{L}{\otimes}_A X : \mathcal{D}^b(\mathsf{mod} A) \to \mathcal{D}^b(\mathsf{mod} A)$$

for some complex X of A-A-bimodules. All autoequivalences with an 'algebraic construction' are of this form, *cf.* section 9.

Theorem 1 *Assume that the following hypotheses hold:*

1) *There is a hereditary abelian k-category $\mathcal{H}$ and a triangle equivalence*

$$\mathcal{D}^b(\mathsf{mod} A) \xrightarrow{\sim} \mathcal{D}^b(\mathcal{H})$$

In the conditions 2) and 3) below, we identify $\mathcal{T}$ with $\mathcal{D}^b(\mathcal{H})$.

2) *For each indecomposable U of $\mathcal{H}$, only finitely many objects F^iU, $i \in \mathbf{Z}$, lie in $\mathcal{H}$.*

3) *There is an integer $N \geq 0$ such that the F-orbit of each indecomposable of $\mathcal{T}$ contains an object S^nU, for some $0 \leq n \leq N$ and some indecomposable object U of $\mathcal{H}$.*

Then the orbit category $\mathcal{T}/F$ admits a natural triangulated structure such that the projection functor $\mathcal{T} \to \mathcal{T}/F$ is triangulated.

The triangulated structure on the orbit category is (most probably) not unique. However, as we will see in section 9.6, the orbit category is the associated triangulated category of a dg category, the *exact (or pretriangulated) dg orbit category*, and the exact dg orbit category is unique and functorial (in the homotopy category of dg categories) since it is the solution of a universal problem. Thus, although perhaps not unique, the triangulated structure on the orbit category is at least canonical, insofar as it comes from a dg structure which is unique up to quasi-equivalence.
The construction of the triangulated orbit category $\mathcal{T}/F$ via the exact dg orbit category also shows that there is a triangle equivalence between $\mathcal{T}/F$ and the stable category $\underline{\mathcal{E}}$ of some Frobenius category $\mathcal{E}$.
In sections 7.2, 7.3 and 7.4, we will illustrate the theorem by examples. In sections 5 and 6 below, we prove the theorem. The strategy is as follows: First, under very weak assumptions, we embed $\mathcal{T}/F$ in a naturally triangulated ambient category $\mathcal{M}$ (whose intrinsic interpretation will be given in section 9.6). Then we show that $\mathcal{T}/F$ is closed under extensions in the ambient category $\mathcal{M}$. Here we will need the full strength of the assumptions 1), 2) and 3).
If $\mathcal{T}$ is the derived category of an abelian category which is not necessarily a module category, one can still define a suitable notion of a standard equivalence $\mathcal{T} \to \mathcal{T}$, *cf.* section 9. Then the analogue of the above theorem is true, *cf.* section 9.9.

5 Construction of the triangulated hull $\mathcal{M}$

The construction is based on the formalism of dg categories, which is briefly recalled in section 9.1. We refer to [32], [21] and [57] for more background.

5.1 The dg orbit category

Let $\mathcal{A}$ be a dg category and $F : \mathcal{A} \to \mathcal{A}$ a dg functor inducing an equivalence in $H^0\mathcal{A}$. We define the *dg orbit category* $\mathcal{B}$ to be the dg category with the same objects as $\mathcal{A}$ and such that for $X, Y \in \mathcal{B}$, we have

$$\mathcal{B}\,(X,Y) \xrightarrow{\sim} \mathsf{colim}_p \bigoplus_{n\geq 0} \mathcal{A}\,(F^n X, F^p Y)\ ,$$

where the transition maps of the colim are given by F. This definition ensures that $H^0\mathcal{B}$ is isomorphic to the orbit category $(H^0\mathcal{A})/F$.

5.2 The projection functor and its right adjoint

¿From now on, we assume that for all objects X, Y of $\mathcal{A}$, the group

$$(H^0\mathcal{A})\,(X, F^n Y)$$

vanishes except for finitely many $n \in \mathbf{Z}$. We have a canonical dg functor $\pi : \mathcal{A} \to \mathcal{B}$. It yields an $\mathcal{A}$-$\mathcal{B}$-bimodule

$$(X,Y) \mapsto \mathcal{B}\,(\pi X, Y).$$

The standard functors associated with this bimodule are

- the derived tensor functor (=induction functor)

$$\pi_* : \mathcal{DA} \to \mathcal{DB}$$

- the derived Hom-functor (right adjoint to π_*), which equals the restriction along π:

$$\pi_\rho : \mathcal{DB} \to \mathcal{DA}$$

For $X \in \mathcal{A}$, we have

$$\pi_*(X^\wedge) = (\pi X)^\wedge\ ,$$

where $X^\wedge$ is the functor represented by X. Moreover, we have an isomorphism in $\mathcal{DA}$

$$\pi_\rho \pi_*(X^\wedge) = \bigoplus_{n\in\mathbf{Z}} F^n(X^\wedge),$$

by the definition of the morphisms of $\mathcal{B}$ and the vanishing assumption made above.

5.3 Identifying objects of the orbit category

The functor $\pi_\rho : \mathcal{DB} \to \mathcal{DA}$ is the restriction along a morphism of dg categories. Therefore, it detects isomorphisms. In particular, we obtain the following: Let $E \in \mathcal{DB}$, $Z \in \mathcal{DA}$ and let $f : Z \to \pi_\rho E$ be a morphism. Let $g : \pi_* Z \to E$ be the morphism corresponding to f by the adjunction. In order to show that g is an isomorphism, it is enough to show that $\pi_\rho g : \pi_\rho \pi_* Z \to \pi_\rho E$ is an isomorphism.

5.4 The ambient triangulated category

We use the notations of the main theorem. Let X be a complex of A-A-bimodules such that F is isomorphic to the total derived tensor product by X. We may assume that X is bounded and that its components are projective on both sides. Let $\mathcal{A}$ be the dg category of bounded complexes of finitely generated projective A-modules. The tensor product by X defines a dg functor from $\mathcal{A}$ to $\mathcal{A}$. By abuse of notation, we denote this dg functor by F as well. The assumption 2) implies that the vanishing assumption of subsection 5.2 is satisfied. Thus we obtain a dg category $\mathcal{B}$ and an equivalence of categories

$$\mathcal{D}^b(\mathsf{mod}A)/F \xrightarrow{\sim} H^0\mathcal{B}.$$

We let the ambient triangulated category $\mathcal{M}$ be the triangulated subcategory of $\mathcal{DB}$ generated by the representable functors. The Yoneda embedding $H^0\mathcal{B} \to \mathcal{DB}$ yields the canonical embedding $\mathcal{D}^b(\mathsf{mod}A)/F \to \mathcal{M}$. We have a canonical equivalence $\mathcal{D}(\mathsf{Mod}A) \xrightarrow{\sim} \mathcal{DA}$ and therefore we obtain a pair of adjoint functors (π_*, π_ρ) between $\mathcal{D}(\mathsf{Mod}A)$ and $\mathcal{DB}$. The functor π_* restricts to the canonical projection $\mathcal{D}^b(\mathsf{mod}A) \to \mathcal{D}^b(\mathsf{mod}A)/F$.

6 The orbit category is closed under extensions

Consider the right adjoint π_ρ of π_*. It is defined on $\mathcal{DB}$ and takes values in the unbounded derived category of all A-modules $\mathcal{D}(\mathsf{Mod}A)$. For $X \in \mathcal{D}^b(\mathsf{mod}A)$, the object $\pi_*\pi_\rho X$ is isomorphic to the sum of the translates F^iX, $i \in \mathbf{Z}$, of X. It follows from assumption 2) that for each fixed $n \in \mathbf{Z}$, the module $H^n_{\mathsf{mod}A} F^iX$ vanishes for almost all $i \in \mathbf{Z}$. Therefore the sum of the F^iX lies in $\mathcal{D}(\mathsf{mod}A)$. Consider a morphism $f : \pi_*X \to \pi_*Y$ of the orbit category $\mathcal{T}/F = \mathcal{D}^b(\mathsf{mod}A)/F$. We form a triangle

$$\pi_*X \to \pi_*Y \to E \to S\pi_*X$$

in $\mathcal{M}$. We apply the right adjoint π_ρ of π_* to this triangle. We get a triangle

$$\pi_\rho\pi_*X \to \pi_\rho\pi_*Y \to \pi_\rho E \to S\pi_\rho\pi_*X$$

in $\mathcal{D}(\mathsf{Mod}A)$. As we have just seen, the terms $\pi_\rho\pi_*X$ and $\pi_\rho\pi_*Y$ of the triangle belong to $\mathcal{D}(\mathsf{mod}A)$. Hence so does $\pi_\rho E$. We will construct an object $Z \in \mathcal{D}^b(\mathsf{mod}A)$ and an isomorphism $g : \pi_*Z \to E$ in the orbit category. Using proposition 6.1 below, we extend the canonical t-structure on

$\mathcal{D}^b(\mathcal{H}) \xrightarrow{\sim} \mathcal{D}^b(\mathsf{mod}A)$ to a t-structure on all of $\mathcal{D}(\mathsf{mod}A)$. Since $\mathcal{H}$ is hereditary, part b) of the proposition shows that each object of $\mathcal{D}(\mathsf{mod}A)$ is the sum of its $\mathcal{H}$-homology objects placed in their respective degrees. In particular, this holds for $\pi_\rho E$. Each of the homology objects is a finite sum of indecomposables. Thus $\pi_\rho E$ is a sum of shifted copies of indecomposable objects of $\mathcal{H}$. Moreover, $F\pi_\rho E$ is isomorphic to $\pi_\rho E$, so that the sum is stable under F. Hence it is a sum of F-orbits of shifted indecomposable objects. By assumption 3), each of these orbits contains an indecomposable direct factor of

$$\bigoplus_{0\leq n\leq N} (H^n \pi_\rho E)[-n].$$

Thus there are only finitely many orbits involved. Let $Z_1, \ldots, Z_M$ be shifted indecomposables of $\mathcal{H}$ such that $\pi_\rho E$ is the sum of the F-orbits of the Z_i. Let f be the inclusion morphism

$$Z = \bigoplus_{i=1}^{M} Z_i \to \pi_\rho E$$

and $g : \pi_* Z \to E$ the morphism corresponding to f under the adjunction. Clearly $\pi_\rho g$ is an isomorphism. By subsection 5.3, the morphism g is invertible and we are done.

6.1 Extension of t-structures to unbounded categories

Let $\mathcal{T}$ be a triangulated category and $\mathcal{U}$ an aisle [35] in $\mathcal{T}$. Denote the associated t-structure [10] by $(\mathcal{U}_{\leq 0}, \mathcal{U}_{\geq 0})$, its heart by $\mathcal{U}_0$, its homology functors by $H^n_{\mathcal{U}} : \mathcal{T} \to \mathcal{U}_0$ and its truncation functors by $\tau_{\leq n}$ and $\tau_{>n}$. Suppose that $\mathcal{U}$ is *dominant*, *i.e.* the following two conditions hold:

1) a morphism s of $\mathcal{T}$ is an isomorphism iff $H^n_{\mathcal{U}}(s)$ is an isomorphism for all $n \in \mathbf{Z}$ and

2) for each object $X \in \mathcal{T}$, the canonical morphisms

$$\mathsf{Hom}(?, X) \to \mathsf{lim}\mathsf{Hom}(?, \tau_{\leq n} X) \text{ and } \mathsf{Hom}(X, ?) \to \mathsf{lim}\mathsf{Hom}(\tau_{>n} X, ?)$$

are surjective.

Let $\mathcal{T}^b$ be the full triangulated subcategory of $\mathcal{T}$ whose objects are the $X \in \mathcal{T}$ such that $H^n_{\mathcal{U}}(X)$ vanishes for all $|n| \gg 0$. Let $\mathcal{V}^b$ be an aisle on $\mathcal{T}^b$. Denote the associated t-structure on $\mathcal{T}^b$ by $(\mathcal{V}_{\leq n}, \mathcal{V}_{>n})$, its heart by $\mathcal{V}_0$, the homology functor by $H^n_{\mathcal{V}^b} : \mathcal{B} \to \mathcal{V}_0$ and its truncation functors by $(\sigma_{\leq 0}, \sigma_{>0})$.
Assume that there is an $N \gg 0$ such that we have

$$H^0_{\mathcal{V}^b} \xrightarrow{\sim} H^0_{\mathcal{V}^b} \tau_{>-n} \text{ and } H^0_{\mathcal{V}^b} \tau_{\leq n} \xrightarrow{\sim} H^0_{\mathcal{V}^b}$$

for all $n \geq N$. We define $H^0_{\mathcal{V}} : \mathcal{T} \to \mathcal{V}_0$ by

$$H^0_{\mathcal{V}}(X) = \mathsf{colim} H^0_{\mathcal{V}^b} \tau_{>-n} \tau_{\leq m} X$$

and $H^n_{\mathcal{V}}(X) = H^0_{\mathcal{V}} S^n X$, $n \in \mathbf{Z}$. We define $\mathcal{V} \subset \mathcal{T}$ to be the full subcategory of $\mathcal{T}$ whose objects are the $X \in \mathcal{T}$ such that $H^n_{\mathcal{V}}(X) = 0$ for all $n > 0$.

Proposition 1 *a) $\mathcal{V}$ is an aisle in $\mathcal{T}$ and the associated t-structure is dominant.*

b) If $\mathcal{V}^b$ is hereditary, i.e. *each triangle*

$$\sigma_{\leq 0} X \to X \to \sigma_{>0} X \to S\sigma_{\leq 0} X \ , \ X \in \mathcal{T}^b$$

splits, then $\mathcal{V}$ is hereditary and each object $X \in \mathcal{T}$ is (non canonically) isomorphic to the sum of the $S^{-n} H^n_{\mathcal{V}}(X)$, $n \in \mathbf{Z}$.

The proof is an exercise on t-structures which we leave to the reader.

7 Another construction of the triangulated hull of the orbit category

7.1 The construction

Let A be a finite-dimensional algebra of finite global dimension over a field k. Let X be an A-A-bimodule complex whose homology has finite total dimension. Let F be the functor

$$? \overset{L}{\otimes}_A X : \mathcal{D}^b(\mathsf{mod} A) \to \mathcal{D}^b(\mathsf{mod} A).$$

We suppose that F is an equivalence and that for all L, M in $\mathcal{D}^b(\mathsf{mod} A)$, the group

$$\mathsf{Hom}(L, F^n M)$$

vanishes for all but finitely many $n \in \mathbf{Z}$. We will construct a triangulated category equivalent to the triangulated hull of section 5.

Consider A as a dg algebra concentrated in degree 0. Let B be the dg algebra with underlying complex $A \oplus X[-1]$, where the multiplication is that of the trivial extension:

$$(a, x)(a', x') = (aa', ax' + xa').$$

Let $\mathcal{D}B$ be the derived category of B and $\mathcal{D}^b(B)$ the *bounded derived category*, *i.e.* the full subcategory of $\mathcal{D}B$ formed by the dg modules whose homology has finite total dimension over k. Let $\mathsf{per}(B)$ be the *perfect* derived category of B, *i.e.* the smallest subcategory of $\mathcal{D}B$ containing B and stable under shifts, extensions and passage to direct factors. By our assumption on A and X, the perfect derived category is contained in $\mathcal{D}^b(B)$. The obvious morphism $B \to A$ induces a restriction functor $\mathcal{D}^b A \to \mathcal{D}^b B$ and by composition, we obtain a functor

$$\mathcal{D}^b A \to \mathcal{D}^b B \to \mathcal{D}^b(B)/\mathsf{per}(B)$$

Theorem 2 *The category $\mathcal{D}^b(B)/\mathsf{per}(B)$ is equivalent to the triangulated hull* (cf. *section 5) of the orbit category of $\mathcal{D}^b(A)$ under F and the above functor identifies with the projection functor.*

Proof. If we replace X by a quasi-isomorphic bimodule, the algebra B is replaced by a quasi-isomorphic dg algebra and its derived category by an equivalent one. Therefore, it is no restriction of generality to assume, as we will do, that X is cofibrant as a dg A-A-bimodule. We will first compute morphisms in $\mathcal{D}B$ between dg B-modules whose restrictions to A are cofibrant. For this, let C be the dg submodule of the bar resolution of B as a bimodule over itself whose underlying graded module is

$$C = \coprod_{n\geq 0} B \otimes_A X^{\otimes_A n} \otimes_A B.$$

The bar resolution of B is a coalgebra in the category of dg B-B-bimodules (*cf. e.g.* [32]) and C becomes a dg subcoalgebra. Its counit is

$$\varepsilon : C \to B \otimes_A B \to B$$

and its comultiplication is given by

$$\Delta(b_0, x_1, \ldots, x_n, b_{n+1}) = \sum_{i=0}^{n} (b_0, x_1, \ldots, x_i) \otimes 1 \otimes 1 \otimes (x_{i+1}, \ldots, b_{n+1}).$$

It is not hard to see that the inclusion of C in the bar resolution is an homotopy equivalence of left (and of right) dg B-modules. Therefore, the same holds for the counit $\varepsilon : C \to B$. For an arbitrary right dg B-module L, the counit ε thus induces a quasi-isomorphism $L \otimes_B C \to L$. Now suppose that the restriction of L to A is cofibrant. Then $L \otimes_A B \otimes_A B$ is cofibrant over B and thus $L \otimes_B C \to L$ is a cofibrant resolution of L. Let $\mathcal{C}_1$ be the dg category whose objects are the dg B-modules whose restriction to A is cofibrant and whose morphism spaces are the

$$\mathsf{Hom}_B(L \otimes_B C, M \otimes_B C).$$

Let $\mathcal{C}_2$ be the dg category with the same objects as $\mathcal{C}_1$ and whose morphism spaces are

$$\mathsf{Hom}_B(L \otimes_B C, M).$$

By definition, the composition of two morphisms f and g of $\mathcal{C}_2$ is given by

$$f \circ (g \otimes \mathbf{1}_C) \circ (\mathbf{1}_L \otimes \Delta).$$

We have a dg functor $\Phi : \mathcal{C}_2 \to \mathcal{C}_1$ which is the identity on objects and sends $g : L \to M$ to

$$(g \otimes \mathbf{1}_C) \circ (\mathbf{1}_L \otimes \Delta) : L \otimes_B C \to M \otimes_B C.$$

The morphism

$$\mathsf{Hom}_B(L \otimes_B C, M) \to \mathsf{Hom}_B(L \otimes_B C, M \otimes_B C)$$

given by Φ is left inverse to the quasi-isomorphism induced by $M \otimes_B C \to M$. Therefore, the dg functor Φ yields a quasi-isomorphism between $\mathcal{C}_2$ and $\mathcal{C}_1$ so that we can compute morphisms and compositions in $\mathcal{D}B$ using $\mathcal{C}_2$. Now suppose that L and M are cofibrant dg A-modules. Consider them as dg B-modules via restriction along the projection $B \to A$. Then we have natural isomorphisms of complexes

$$\mathsf{Hom}_{\mathcal{C}_2}(L, M) = \mathsf{Hom}_B(L \otimes_B C, M) \xrightarrow{\sim} \mathsf{Hom}_A(\coprod_{n \geq 0} L \otimes_A X^{\otimes_A n}, M).$$

Moreover, the composition of morphisms in $\mathcal{C}_2$ translates into the natural composition law for the right hand side. Now we will compute morphisms in the quotient category $\mathcal{D}B/\mathsf{per}(B)$. Let M be as above. For $p \geq 0$, let $C_{\leq p}$ be the dg subbimodule with underlying graded module

$$\coprod_{n=0}^{p} B \otimes_A X^{\otimes_A n} \otimes_A B.$$

Then each morphism

$$P \to M \otimes_B C$$

of $\mathcal{D}B$ from a perfect object P factors through

$$M \otimes_B C_{\leq p} \to M \otimes_B C$$

for some $p \geq 0$. Therefore, the following complex computes morphisms in $\mathcal{D}B/\mathsf{per}(B)$:

$$\mathsf{colim}_{p \geq 1} \mathsf{Hom}_B(L \otimes_B C, M \otimes_B C/C_{\leq p-1}).$$

Now it is not hard to check that the inclusion

$$M \otimes X^{\otimes_A p} \otimes_A A \to M \otimes_B (C/C_{\leq p-1})$$

is a quasi-isomorphism of dg B-modules. Thus we obtain quasi-isomorphisms

$$\mathsf{Hom}_B(L \otimes_B C, M \otimes X^{\otimes_A p} \otimes_A A) \to \mathsf{Hom}_B(L \otimes_B C, M \otimes_B C/C_{\leq p-1})$$

and

$$\prod_{n \geq 0} \mathsf{Hom}_A(L \otimes_A X^{\otimes_A n}, M \otimes_A X^{\otimes_A p}) \to \mathsf{Hom}_B(L \otimes_B C, M \otimes_B C/C_{\leq p-1}).$$

Moreover, it is not hard to check that if we define transition maps

$$\prod_{n \geq 0} \mathsf{Hom}_A(L \otimes_A X^{\otimes_A n}, M \otimes_A X^{\otimes_A p}) \to \prod_{n \geq 0} \mathsf{Hom}_A(L \otimes_A X^{\otimes_A n}, M \otimes_A X^{\otimes_A (p+1)})$$

by sending f to $f \otimes_A \mathbf{1}_X$, then we obtain a quasi-isomorphism of direct systems of complexes. Therefore, the following complex computes morphisms in $\mathcal{D}B/\mathsf{per}(B)$:

$$\mathsf{colim}_{p\geq 1} \coprod_{n\geq 0} \mathsf{Hom}_A(L\otimes_A X^{\otimes_A n}, M\otimes_A X^{\otimes_A p}).$$

Let $\mathcal{C}_3$ be the dg category whose objects are the cofibrant dg A-modules and whose morphisms are given by the above complexes. If L and M are cofibrant dg A-modules and belong to $\mathcal{D}^b(\mathsf{mod}A)$, then, by our assumptions on F, this complex is quasi-isomorphic to its subcomplex

$$\mathsf{colim}_{p\geq 1} \coprod_{n\geq 0} \mathsf{Hom}_A(L\otimes_A X^{\otimes_A n}, M\otimes_A X^{\otimes_A p}).$$

Thus we obtain a dg functor

$$\mathcal{B} \to \mathcal{C}_3$$

(where $\mathcal{B}$ is the dg category defined in 5.1) which induces a fully faithful functor $H^0(\mathcal{B}) \to \mathcal{D}B/\mathsf{per}(B)$ and thus a fully faithful functor $\mathcal{M} \to \mathcal{D}^b(B)/\mathsf{per}(B)$. This functor is also essentially surjective. Indeed, every object in $\mathcal{D}^b(B)$ is an extension of two objects which lie in the image of $\mathcal{D}^b(\mathsf{mod}A)$. The assertion about the projection functor is clear from the above proof.

7.2 The motivating example

Let us suppose that the functor F is given by

$$M \mapsto \tau S^{-1} M\ ,$$

where τ is the Auslander-Reiten translation of $\mathcal{D}^b(A)$ and S the shift functor. This is the case considered in [8] for the construction of the cluster category. The functor F^{-1} is isomorphic to

$$M \mapsto S^{-2}\nu M$$

where ν is the Nakayama functor

$$\nu = ? \overset{L}{\otimes}_A DA\ ,\ DA = \mathsf{Hom}_k(A,k).$$

Thus F^{-1} is given by the bimodule $X = (DA)[-2]$ and $B = A \oplus (DA)[-3]$ is the trivial extension of A with a non standard grading: A is in degree 0 and DA in degree 3. For example, if A is the quiver algebra of an alternating quiver whose underlying graph is A_n, then the underlying ungraded algebra of B is the quadratic dual of the preprojective algebra associated with A_n, *cf.* [15]. The algebra B viewed as a differential graded algebra was investigated by Khovanov-Seidel in [36]. Here the authors show that $\mathcal{D}^b(B)$ admits a canonical action by the braid group on $n+1$ strings, a result which was obtained independently in a similar context by Zimmermann-Rouquier [52]. The canonical generators of the braid group act by triangle functors T_i endowed with morphisms $\phi_i : T_i \to \mathbf{1}$. The cone on each ϕ_i belongs to $\mathsf{per}(B)$ and $\mathsf{per}(B)$ is in fact equal to its smallest triangulated subcategory stable under direct factors and containing these cones. Thus, the action becomes trivial in $\mathcal{D}^b(B)/\mathsf{per}(B)$ and in a certain sense, this is the largest quotient where the ϕ_i become invertible.

7.3 Projectives over the preprojective algebra

Let A be the path algebra of a Dynkin quiver, *i.e.* a quiver whose underlying graph is a Dynkin diagram of type A, D or E. Let C be the associated preprojective algebra [26], [20], [50]. In Proposition 3.3 of [7], Auslander-Reiten show that the category of projective modules over C is equivalent to the stable category of maximal Cohen-Macaulay modules over a representation-finite isolated hypersurface singularity. In particular, it is triangulated. This can also be deduced from our main theorem: Indeed, it follows from D. Happel's description [27] of the derived category $\mathcal{D}^b(A)$ that the category $\mathsf{proj}C$ of finite dimensional projective C-modules is equivalent to the orbit category $\mathcal{D}^b(A)/\tau$, *cf.* also [25]. Moreover, by the theorem of the previous section, we have an equivalence

$$\mathsf{proj}C \xrightarrow{\sim} \mathcal{D}^b(B)/\mathsf{per}(B)$$

where $B = A \oplus (DA)[-2]$. This equivalence yields in fact more than just a triangulated structure: it shows that $\mathsf{proj}C$ is endowed with a canonical Hochschild 3-cocycle m_3, *cf.* for example [11]. It would be interesting to identify this cocycle in the description given in [23].

7.4 Projectives over $\Lambda(L_n)$

The category of projective modules over the algebra $k[\varepsilon]/(\varepsilon^2)$ of dual numbers is triangulated. Indeed, it is equivalent to the orbit category of the derived category of the path algebra of a quiver of type A_2 under the Nakayama autoequivalence ν. Thus, we obtain examples of triangulated categories whose Auslander-Reiten quiver contains a loop. It has been known since Riedtmann's work [49] that this cannot occur in the stable category (*cf.* below) of a self-injective finite-dimensional algebra. It may therefore seem surprising, *cf.* [58], that loops do occur in this more general context. However, loops already do occur in stable categories of finitely generated reflexive modules over certain non commutative generalizations of local rings of rational double points, as shown by Auslander-Reiten in [6]. These were completely classified by Reiten-Van den Bergh in [46]. In particular, the example of the dual numbers and its generalization below are among the cases covered by [46].
The example of the dual numbers generalizes as follows: Let $n \geq 1$ be an integer. Following [30], the *generalized Dynkin graph* L_n is defined as the graph

$$1 \text{ --- } 2 \text{ --- } \cdots \text{ --- } n-1 \text{ --- } n$$

Its edges are in natural bijection with the orbits of the involution which exchanges each arrow α with $\overline{\alpha}$ in the following quiver:

$$\varepsilon=\overline{\varepsilon} \circlearrowright 1 \underset{\overline{a_1}}{\overset{a_1}{\rightleftarrows}} 2 \underset{\overline{a_2}}{\overset{a_2}{\rightleftarrows}} \quad \cdots \quad \underset{\overline{a_{n-2}}}{\overset{a_{n-2}}{\rightleftarrows}} n-1 \underset{\overline{a_{n-1}}}{\overset{a_{n-1}}{\rightleftarrows}} n \, .$$

The associated preprojective algebra $\Lambda(L_n)$ of generalized Dynkin type L_n is defined as the quotient of the path algebra of this quiver by the ideal generated by the relators

$$r_v = \sum \alpha \, \overline{\alpha} \, ,$$

where, for each $1 \leq v \leq n$, the sum ranges over the arrows α with starting point v. Let A be the path algebra of a Dynkin quiver with underlying Dynkin graph A_{2n}. Using D. Happel's description [27] of the derived category of a Dynkin quiver, we see that the orbit category $\mathcal{D}^b(A)/(\tau^n S)$ is equivalent to the category of finitely generated projective modules over the algebra $\Lambda(L_n)$. By the main theorem, this category is thus triangulated. Its Auslander-Reiten quiver is given by the ordinary quiver of $\Lambda(L_n)$, *cf.* above, endowed with $\tau = \mathbf{1}$: Indeed, in $\mathcal{D}^b(A)$, we have $S^2 = \tau^{-(2n+1)}$ so that in the orbit category, we obtain

$$\mathbf{1} = (\tau^n S)^2 = \tau^{2n} S^2 = \tau^{-1}.$$

8 On the Calabi-Yau property

8.1 Serre functors and localizations

Let k be a field and $\mathcal{T}$ a k-linear triangulated category with finite-dimensional Hom-spaces. We denote the suspension functor of $\mathcal{T}$ by S. Recall from [47] that a *right Serre functor* for $\mathcal{T}$ is the datum of a triangle functor $\nu : \mathcal{T} \to \mathcal{T}$ together with bifunctor isomorphisms

$$D\mathsf{Hom}_{\mathcal{T}}(X, ?) \xrightarrow{\sim} \mathsf{Hom}_{\mathcal{T}}(?, \nu X) \, , \; X \in \mathcal{T} \, ,$$

where $D = \mathsf{Hom}_k(?, k)$. If ν exists, it is unique up to isomorphism of triangle functors. Dually, a *left Serre functor* is the datum of a triangle functor $\nu' : \mathcal{T} \to \mathcal{T}$ and isomorphisms

$$D\mathsf{Hom}_{\mathcal{T}}(?, X) \xrightarrow{\sim} \mathsf{Hom}_{\mathcal{T}}(\nu', ?) \, , \; X \in \mathcal{T}.$$

The category $\mathcal{T}$ *has Serre duality* if it has both a left and a right Serre functor, or equivalently, if it has a onesided Serre functor which is an equivalence, *cf.* [47] [13]. The following lemma is used in [9].

Lemma 1 *Suppose that $\mathcal{T}$ has a left Serre functor ν'. Let $\mathcal{U} \subset \mathcal{T}$ be a thick triangulated subcategory and $L : \mathcal{T} \to \mathcal{T}/\mathcal{U}$ the localization functor.*

a) If L admits a right adjoint R, then $L\nu' R$ is a left Serre functor for $\mathcal{T}/\mathcal{U}$.

b) *More generally, if the functor $\nu' : \mathcal{T} \to \mathcal{T}$ admits a total right derived functor $\mathbf{R}\nu' : \mathcal{T}/\mathcal{U} \to \mathcal{T}/\mathcal{U}$ in the sense of Deligne [19] with respect to the localization $\mathcal{T} \to \mathcal{T}/\mathcal{U}$, then $\mathbf{R}\nu'$ is a left Serre functor for $\mathcal{T}/\mathcal{U}$.*

Proof. a) For X, Y in $\mathcal{T}$, we have

$$\begin{aligned}\mathsf{Hom}_{\mathcal{T}/\mathcal{U}}(L\nu' RX, Y) &= \mathsf{Hom}_{\mathcal{T}}(\nu' RX, RY)\\ &= D\mathsf{Hom}_{\mathcal{T}}(RY, RX) = D\mathsf{Hom}_{\mathcal{T}/\mathcal{U}}(Y, X).\end{aligned}$$

Here, for the last isomorphism, we have used that R is fully faithful (L is a localization functor).
b) We assume that $\mathcal{T}$ is small. Let $\mathsf{Mod}\mathcal{T}$ denote the (large) category of functors from $\mathcal{T}^{op}$ to the category of abelian groups and let $h : \mathcal{T} \to \mathsf{Mod}\mathcal{T}$ denote the Yoneda embedding. Let L^* be the unique right exact functor $\mathsf{Mod}\mathcal{T} \to \mathsf{Mod}(\mathcal{T}/\mathcal{U})$ which sends hX to hLX, $X \in \mathcal{T}$. By the calculus of (right) fractions, L^* has a right adjoint R which takes an object Y to

$$\mathsf{colim}_{\Sigma_Y} hY' ,$$

where the colim ranges over the category Σ_Y of morphisms $s : Y \to Y'$ which become invertible in $\mathcal{T}/\mathcal{U}$. Clearly L^*R is isomorphic to the identity so that R is fully faithful. By definition of the total right derived functor, for each object $X \in \mathcal{T}/\mathcal{U}$, the functor

$$\mathsf{colim}_{\Sigma_X} h(L\nu' X') = L^* \nu'^* Rh(X)$$

is represented by $\mathbf{R}\nu'(X)$. Therefore, we have

$$\begin{aligned}\mathsf{Hom}_{\mathcal{T}/\mathcal{U}}(\mathbf{R}\nu'(X), Y) &= \mathsf{Hom}_{\mathsf{Mod}\mathcal{T}/\mathcal{U}}(L^*\nu'^* Rh(X), h(Y))\\ &= \mathsf{Hom}_{\mathsf{Mod}\mathcal{T}}(\nu'^* Rh(X), Rh(Y)).\end{aligned}$$

Now by definition, the last term is isomorphic to

$$\mathsf{Hom}_{\mathsf{Mod}\mathcal{T}}(\mathsf{colim}_{\Sigma_X} h(\nu' X'), \mathsf{colim}_{\Sigma_Y} hY') = \mathsf{lim}_{\Sigma_X} \mathsf{colim}_{\Sigma_Y} \mathsf{Hom}_{\mathcal{T}}(\nu' X', Y')$$

and this identifies with

$$\begin{aligned}\mathsf{lim}_{\Sigma_X} \mathsf{colim}_{\Sigma_Y} D\mathsf{Hom}_{\mathcal{T}}(Y', X') &= D(\mathsf{colim}_{\Sigma_X} \mathsf{lim}_{\Sigma_Y} \mathsf{Hom}_{\mathcal{T}}(Y', X'))\\ &= D\mathsf{Hom}_{\mathcal{T}/\mathcal{U}}(Y, X).\end{aligned}$$

8.2 Definition of the Calabi-Yau property

Keep the hypotheses of the preceding section. By definition [39], the triangulated category $\mathcal{T}$ is *Calabi-Yau of CY-dimension* d if it has Serre duality and there is an isomorphism of triangle functors

$$\nu \xrightarrow{\sim} S^d.$$

By extension, if we have $\nu^e \xrightarrow{\sim} S^d$ for some integer $e > 0$, one sometimes says that $\mathcal{T}$ is Calabi-Yau of fractional dimension d/e. Note that $d \in \mathbf{Z}$ is only determined up to a multiple of the order of S. It would be interesting to link the CY-dimension to Rouquier's [51] notion of dimension of a triangulated category.
The terminology has its origin in the following example: Let X be a smooth projective variety of dimension d and let $\omega_X = \Lambda^d T_X^*$ be the canonical bundle. Let $\mathcal{T}$ be the bounded derived category of coherent sheaves on X. Then the classical Serre duality

$$D\mathsf{Ext}^i_X(\mathcal{F},\mathcal{G}) \xrightarrow{\sim} \mathsf{Ext}^{d-i}(\mathcal{G},\mathcal{F}\otimes\omega_X)\,,$$

where $\mathcal{F}$, $\mathcal{G}$ are coherent sheaves, lifts to the isomorphism

$$D\mathsf{Hom}_{\mathcal{T}}(\mathcal{F},\mathcal{G}) \xrightarrow{\sim} \mathsf{Hom}_{\mathcal{T}}(\mathcal{G},\mathcal{F}\otimes\omega_X[d])\,,$$

where $\mathcal{F}$, $\mathcal{G}$ are bounded complexes of coherent sheaves. Thus $\mathcal{T}$ has Serre duality and $\nu = ?\otimes\omega_X[d]$. So the category $\mathcal{T}$ is Calabi-Yau of CY-dimension d iff ω_X is isomorphic to $\mathcal{O}_X$, which means precisely that the variety X is Calabi-Yau of dimension d.
If $\mathcal{T}$ is a Calabi-Yau triangulated category 'of algebraic origin' (for example, the derived category of a category of modules or sheaves), then it often comes from a Calabi-Yau A_∞-category. These are of considerable interest in mathematical physics, since, as Kontsevich shows [38], [37], *cf.* also [18], a topological quantum field theory is associated with each Calabi-Yau A_∞-category satisfying some additional assumptions[1].

8.3 Examples

(1) If A is a finite-dimensional k-algebra, then the homotopy category $\mathcal{T}$ of bounded complexes of finitely generated projective A-modules has a Nakayama functor iff DA is of finite projective dimension. In this case, the category $\mathcal{T}$ has Serre duality iff moreover A_A is of finite injective dimension, *i.e.* iff A is Gorenstein, *cf.* [28]. Then the category $\mathcal{T}$ is Calabi-Yau (necessarily of CY-dimension 0) iff A is symmetric.
(2) If Δ is a Dynkin graph of type A_n, D_n, E_6, E_7 or E_8 and h is its Coxeter number (*i.e.* $n+1$, $2(n-1)$, 12, 18 or 30, respectively), then for the bounded derived category of finitely generated modules over a quiver with underlying graph Δ, we have isomorphisms

$$\nu^h = (S\tau)^h = S^h\tau^h = S^{(h-2)}.$$

[1]Namely, the associated triangulated category should admit a generator whose endomorphism A_∞-algebra B is compact (*i.e.* finite-dimensional), smooth (*i.e.* B is perfect as a bimodule over itself), and whose associated Hodge-de Rham spectral sequence collapses (this property is conjectured to hold for all smooth compact A_∞-algebras over a field of characteristic 0).

Hence this category is Calabi-Yau of fractional dimension $(h-2)/h$.
(3) Suppose that A is a finite-dimensional algebra which is selfinjective (*i.e.* A_A is also an injective A-module). Then the category $\mathsf{mod}A$ of finite-dimensional A-modules is Frobenius, *i.e.* it is an abelian (or, more generally, an exact) category with enough projectives, enough injectives and where an object is projective iff it is injective. The stable category $\underline{\mathsf{mod}}\,A$ obtained by quotienting $\mathsf{mod}A$ by the ideal of morphisms factoring through injectives is triangulated, *cf.* [27]. The inverse of its suspension functor sends a module M to the kernel ΩM of an epimorphism $P \to M$ with projective P. Let $\mathcal{N}M = M \otimes_A DA$. Then $\underline{\mathsf{mod}}\,A$ has Serre duality with Nakayama functor $\nu = \Omega \circ \mathcal{N}$. Thus, the stable category is Calabi-Yau of CY-dimension d iff we have an isomorphism of triangle functors

$$\Omega^{(d+1)} \circ \mathcal{N} = \mathbf{1}.$$

For this, it is clearly sufficient that we have an isomorphism

$$\Omega^{d+1}_{A^e}(A) \otimes_A DA \xrightarrow{\sim} A\,,$$

in the stable category of A-A-bimodules, *i.e.* modules over the selfinjective algebra $A \otimes A^{op}$. For example, we deduce that if A is the path algebra of a cyclic quiver with n vertices divided by the ideal generated by all paths of length $n-1$, then $\underline{\mathsf{mod}}\,A$ is Calabi-Yau of CY-dimension 3.
(4) Let A be a dg algebra. Let $\mathsf{per}(A) \subset \mathcal{D}(A)$ be the subcategory of perfect dg A-modules, *i.e.* the smallest full triangulated subcategory of $\mathcal{D}(A)$ containing A and stable under forming direct factors. For each P in $\mathsf{per}(A)$ and each $M \in \mathcal{D}(A)$, we have canonical isomorphisms

$$D\mathsf{RHom}_A(P, M) \xrightarrow{\sim} \mathsf{RHom}_A(M, D\mathsf{RHom}_A(P, A))$$

and

$$P \overset{L}{\otimes}_A DA \xrightarrow{\sim} D\mathsf{RHom}_A(P, A).$$

So we obtain a canonical isomorphism

$$D\mathsf{RHom}_A(P, M) \xrightarrow{\sim} \mathsf{RHom}_A(M, P \overset{L}{\otimes}_A DA).$$

Thus, if we are given a quasi-isomorphism of dg A-A-bimodules

$$\phi : A[n] \to DA\,,$$

we obtain

$$D\mathsf{RHom}_A(P, M) \xrightarrow{\sim} \mathsf{RHom}_A(M, P[n])$$

and in particular $\mathsf{per}(A)$ is Calabi-Yau of CY-dimension n.
(5) To consider a natural application of the preceding example, let B be the symmetric algebra on a finite-dimensional vector space V of dimension n and $\mathcal{T} \subset \mathcal{D}(B)$ the localizing subcategory generated by the trivial B-module k (*i.e.* the smallest full triangulated subcategory stable under infinite sums and

containing the trivial module). Let $\mathcal{T}^c$ denote its subcategory of compact objects. This is exactly the triangulated subcategory of $\mathcal{D}(B)$ generated by k, and also exactly the subcategory of the complexes whose total homology is finite-dimensional and supported in 0. Then $\mathcal{T}^c$ is Calabi-Yau of CY-dimension n. Indeed, if

$$A = \mathsf{RHom}_B(k, k)$$

is the Koszul dual of B (thus, A is the exterior algebra on the dual of V concentrated in degree 1; it is endowed with $d = 0$), then the functor

$$\mathsf{RHom}_B(k, ?) : \mathcal{D}(B) \to \mathcal{D}(A)$$

induces equivalences from $\mathcal{T}$ to $\mathcal{D}(A)$ and $\mathcal{T}^c$ to $\mathsf{per}(A)$, *cf.* for example [32]. Now we have a canonical isomorphism of A-A-bimodules $A[n] \xrightarrow{\sim} DA$ so that $\mathsf{per}(A)$ and $\mathcal{T}$ are Calabi-Yau of CY-dimension n. As pointed out by I. Reiten, in this case, the Calabi-Yau property even holds more generally: Let $M \in \mathcal{D}(B)$ and denote by $M_{\mathcal{T}} \to M$ the universal morphism from an object of $\mathcal{T}$ to M. Then, for $X \in \mathcal{T}^c$, we have natural morphisms

$$\begin{aligned} \mathsf{Hom}_{\mathcal{D}(B)}(M, X[n]) \to \mathsf{Hom}_{\mathcal{T}}(M_{\mathcal{T}}, X[n]) &\xrightarrow{\sim} D\mathsf{Hom}_{\mathcal{T}}(X, M_{\mathcal{T}}) \\ &\xrightarrow{\sim} D\mathsf{Hom}_{\mathcal{D}(B)}(X, M). \end{aligned}$$

The composition

$$(*) \qquad \mathsf{Hom}_{\mathcal{D}(B)}(M, X[n]) \to D\mathsf{Hom}_{\mathcal{D}(B)}(X, M)$$

is a morphism of (co-)homological functors in $X \in \mathcal{T}^c$ (resp. $M \in \mathcal{D}(B)$). We claim that it is an isomorphism for $M \in \mathsf{per}(B)$ and $X \in \mathcal{T}^c$. It suffices to prove this for $M = B$ and $X = k$. Then one checks it using the fact that

$$\mathsf{RHom}_B(k, B) \xrightarrow{\sim} k[-n].$$

These arguments still work for certain non-commutative algebras B: If B is an Artin-Schelter regular algebra [2] [1] of global dimension 3 and type A and $\mathcal{T}$ the localizing subcategory of the derived category $\mathcal{D}(B)$ of non graded B-modules generated by the trivial module, then $\mathcal{T}^c$ is Calabi-Yau and one even has the isomorphism $(*)$ for each perfect complex of B-modules M and each $X \in \mathcal{T}^c$, *cf.* for example section 12 of [41].

8.4 Orbit categories with the Calabi-Yau property

The main theorem yields the following

Corollary 1 *If $d \in \mathbf{Z}$ and Q is a quiver whose underlying graph is Dynkin of type A, D or E, then*

$$\mathcal{T} = \mathcal{D}^b(kQ)/\tau^{-1}S^{d-1}$$

is Calabi-Yau of CY-dimension d. In particular, the cluster category $\mathcal{C}_{kQ}$ is Calabi-Yau of dimension 2 and the category of projective modules over the preprojective algebra $\Lambda(Q)$ is Calabi-Yau of CY-dimension 1.

The category of projective modules over the preprojective algebra $\Lambda(L_n)$ of example 7.4 does not fit into this framework. Nevertheless, it is also Calabi-Yau of CY-dimension 1, since we have $\tau = \mathbf{1}$ in this category and therefore $\nu = S\tau = S$.

8.5 Module categories over Calabi-Yau categories

Calabi-Yau triangulated categories turn out to be 'self-reproducing': Let $\mathcal{T}$ be a triangulated category. Then the category $\mathsf{mod}\mathcal{T}$ of finitely generated functors from $\mathcal{T}^{op}$ to $\mathsf{Mod}k$ is abelian and Frobenius, *cf.* [24], [42]. If we denote by Σ the exact functor $\mathsf{mod}\mathcal{T} \to \mathsf{mod}\mathcal{T}$ which takes $\mathsf{Hom}(?, X)$ to $\mathsf{Hom}(?, SX)$, then it is not hard to show [24] [25] that we have

$$\Sigma \xrightarrow{\sim} S^3$$

as triangle functors $\underline{\mathsf{mod}}\mathcal{T} \to \underline{\mathsf{mod}}\mathcal{T}$. One deduces the following lemma, which is a variant of a result which Auslander-Reiten [7] obtained using dualizing R-varieties [4] and their functor categories [3], *cf.* also [5] [45]. A similar result is due to Geiss [25].

Lemma 2 *If $\mathcal{T}$ is Calabi-Yau of CY-dimension d, then the stable category $\underline{\mathsf{mod}}\,\mathcal{T}$ is Calabi-Yau of CY-dimension $3d-1$. Moreover, if the suspension of $\mathcal{T}$ is of order n, the order of the suspension functor of $\underline{\mathsf{mod}}\,\mathcal{T}$ divides $3n$.*

For example, if A is the preprojective algebra of a Dynkin quiver or equals $\Lambda(L_n)$, then we find that the stable category $\underline{\mathsf{mod}}\,A$ is Calabi-Yau of CY-dimension $3 \times 1 - 1 = 2$. This result, with essentially the same proof, is due to Auslander-Reiten [7]. For the preprojective algebras of Dynkin quivers, it also follows from a much finer result due to Ringel and Schofield (unpublished). Indeed, they have proved that there is an isomorphism

$$\Omega^3_{A^e}(A) \xrightarrow{\sim} DA$$

in the stable category of bimodules, *cf.* Theorems 4.8 and 4.9 in [15]. This implies the Calabi-Yau property since we also have an isomorphism

$$DA \otimes_A DA \xrightarrow{\sim} A$$

in the stable category of bimodules, by the remark following definition 4.6 in [15]. For the algebra $\Lambda(L_n)$, the analogous result follows from Proposition 2.3 of [12].

9 Universal properties

9.1 The homotopy category of small dg categories

Let k be a field. A *differential graded (=dg) k-module* is a $\mathbf{Z}$-graded vector space

$$V = \bigoplus_{p \in \mathbf{Z}} V^p$$

endowed with a differential d of degree 1. The *tensor product* $V \otimes W$ of two dg k-modules is the graded space with components

$$\bigoplus_{p+q=n} V^p \otimes W^q \,,\ n \in \mathbf{Z}\,,$$

and the differential $d \otimes \mathbf{1} + \mathbf{1} \otimes d$, where the tensor product of maps is defined using the Koszul sign rule. A *dg category* [32] [21] is a k-category $\mathcal{A}$ whose morphism spaces are dg k-modules and whose compositions

$$\mathcal{A}(Y,Z) \otimes \mathcal{A}(X,Y) \to \mathcal{A}(X,Z)$$

are morphisms of dg k-modules. For a dg category $\mathcal{A}$, the *category* $H^0(\mathcal{A})$ has the same objects as $\mathcal{A}$ and has morphism spaces $H^0\mathcal{A}\,(X,Y)$, $X,Y \in \mathcal{A}$. A *dg functor* $F : \mathcal{A} \to \mathcal{B}$ between dg categories is a functor compatible with the grading and the differential on the morphism spaces. It is a *quasi-equivalence* if it induces quasi-isomorphisms in the morphism spaces and an equivalence of categories from $H^0(\mathcal{A})$ to $H^0(\mathcal{B})$. We denote by dgcat the category of small dg categories. The *homotopy category* of small dg categories is the localization $\mathsf{Ho}(\mathsf{dgcat})$ of dgcat with respect to the class of quasi-equivalences. According to [56], the category dgcat admits a structure of Quillen model category (*cf.* [22], [31]) whose weak equivalences are the quasi-equivalences. This implies in particular that for $\mathcal{A}, \mathcal{B} \in \mathsf{dgcat}$, the morphisms from $\mathcal{A}$ to $\mathcal{B}$ in the localization $\mathsf{Ho}(\mathsf{dgcat})$ form a set.

9.2 The bimodule bicategory

For two dg categories $\mathcal{A}$, $\mathcal{B}$, we denote by $\mathsf{rep}(\mathcal{A},\mathcal{B})$ the full subcategory of the derived category $\mathcal{D}(\mathcal{A}^{op} \otimes \mathcal{B})$, *cf.* [32], whose objects are the dg $\mathcal{A}$-$\mathcal{B}$-bimodules X such that $X(?,A)$ is isomorphic to a representable functor in $\mathcal{D}(\mathcal{B})$ for each object A of $\mathcal{A}$. We think of the objects of $\mathsf{rep}(\mathcal{A},\mathcal{B})$ as 'representations up to homotopy' of $\mathcal{A}$ in $\mathcal{B}$. The *bimodule bicategory* rep, *cf.* [32] [21], has as objects all small dg categories; the morphism category between two objects $\mathcal{A}$, $\mathcal{B}$ is $\mathsf{rep}(\mathcal{A},\mathcal{B})$; the composition bifunctor

$$\mathsf{rep}(\mathcal{B},\mathcal{C}) \times \mathsf{rep}(\mathcal{A},\mathcal{B}) \to \mathsf{rep}(\mathcal{A},\mathcal{B})$$

is given by the derived tensor product $(X,Y) \mapsto X \overset{L}{\otimes}_{\mathcal{B}} Y$. For each dg functor $F : \mathcal{A} \to \mathcal{B}$, we have the dg bimodule

$$X_F : A \mapsto \mathcal{B}\,(?, FA)\,,$$

which clearly belongs to $\mathsf{rep}(\mathcal{A},\mathcal{B})$. One can show that the map $F \mapsto X_F$ induces a bijection, compatible with compositions, from the set of morphisms from $\mathcal{A}$ to $\mathcal{B}$ in $\mathsf{Ho}(\mathsf{dgcat})$ to the set of isomorphism classes of bimodules X in $\mathsf{rep}(\mathcal{A},\mathcal{B})$. In fact, a much stronger result by B. Toën [57] relates rep to the Dwyer-Kan localization of dgcat.

9.3 Dg orbit categories

Let $\mathcal{A}$ be a small dg category and $F \in \mathsf{rep}(\mathcal{A}, \mathcal{A})$. We assume, as we may, that F is given by a cofibrant bimodule. For a dg category $\mathcal{B}$, define $\widetilde{\mathsf{eff}}_0(\mathcal{A}, F, \mathcal{B})$ to be the category whose objects are the pairs formed by an $\mathcal{A}$-$\mathcal{B}$-bimodule P in $\mathsf{rep}(\mathcal{A}, \mathcal{B})$ and a morphism of dg bimodules

$$\phi : P \to PF.$$

Morphisms are the morphisms of dg bimodules $f : P \to P'$ such that we have $\phi' \circ f = (fF) \circ \phi$ in the category of dg bimodules. Define $\mathsf{eff}_0(\mathcal{A}, F, \mathcal{B})$ to be the localization of $\widetilde{\mathsf{eff}}_0(\mathcal{A}, F, \mathcal{B})$ with respect to the morphisms f which are quasi-isomorphisms of dg bimodules. Denote by $\mathsf{eff}(\mathcal{A}, F, \mathcal{B})$ the full subcategory of $\mathsf{eff}_0(\mathcal{A}, F, \mathcal{B})$ whose objects are the (P, ϕ) where ϕ is a quasi-isomorphism. It is not hard to see that the assignments

$$\mathcal{B} \mapsto \mathsf{eff}_0(\mathcal{A}, F, \mathcal{B}) \quad \text{and} \quad \mathcal{B} \mapsto \mathsf{eff}(\mathcal{A}, F, \mathcal{B})$$

are 2-functors from rep to the category of small categories.

Theorem 3 *a) The 2-functor $\mathsf{eff}_0(\mathcal{A}, F, ?)$ is 2-representable,* i.e. *there is small dg category $\mathcal{B}_0$ and a pair (P_0, ϕ_0) in $\mathsf{eff}(\mathcal{A}, F, \mathcal{B}_0)$ such that for each small dg category $\mathcal{B}$, the functor*

$$\mathsf{rep}(\mathcal{B}_0, \mathcal{B}) \to \mathsf{eff}_0(\mathcal{A}, F, \mathcal{B}_0)\,,\ G \mapsto G \circ P_0$$

is an equivalence.

b) The 2-functor $\mathsf{eff}(\mathcal{A}, F, ?)$ is 2-representable.

c) For a dg category $\mathcal{B}$, a pair (P, ϕ) is a 2-representative for $\mathsf{eff}_0(\mathcal{A}, F, ?)$ iff $H^0(P) : H^0(\mathcal{A}) \to H^0(\mathcal{B})$ is essentially surjective and, for all objects A, B of $\mathcal{A}$, the canonical morphism

$$\bigoplus_{n \in \mathbf{N}} \mathcal{A}\,(F^n A, B) \to \mathcal{B}\,(PA, PB)$$

is invertible in $\mathcal{D}(k)$.

d) For a dg category $\mathcal{B}$, a pair (P, ϕ) is a 2-representative for $\mathsf{eff}(\mathcal{A}, F, ?)$ iff $H^0(P) : H^0(\mathcal{A}) \to H^0(\mathcal{B})$ is essentially surjective and, for all objects A, B of $\mathcal{A}$, the canonical morphism

$$\bigoplus_{c \in \mathbf{Z}} \mathsf{colim}_{r \gg 0} \mathcal{A}\,(F^{p+r} A, F^{p+c+r} B) \to \mathcal{B}\,(PA, PB)$$

is invertible in $\mathcal{D}(k)$.

We *define* $\mathcal{A}/F$ to be the 2-representative of $\mathsf{eff}(\mathcal{A}, F, ?)$. For example, in the notations of 5.1, $\mathcal{A}/F$ is the dg orbit category $\mathcal{B}$. It follows from part d) of the theorem that we have an equivalence

$$H^0(\mathcal{A})/H^0(F) \to H^0(\mathcal{A}/F).$$

Proof. We only sketch a proof and refer to [54] for a detailed treatment. Define $\mathcal{B}_0$ to be the dg category with the same objects as $\mathcal{A}$ and with the morphism spaces

$$\mathcal{B}_0\,(A, B) = \bigoplus_{n \in \mathbf{N}} \mathcal{A}\,(F^n A, B).$$

We have an obvious dg functor $P_0 : \mathcal{A} \to \mathcal{B}_0$ and an obvious morphism $\phi : P_0 \to P_0 F$. The pair (P_0, ϕ_0) is then 2-universal in rep. This yields a) and c). For b), one adjoins a formal homotopy inverse of ϕ to $\mathcal{B}_0$. One obtains d) by computing the homology of the morphism spaces in the resulting dg category.

9.4 Functoriality in $(\mathcal{A}, F)$

Let a square of rep

$$\begin{tikzcd} \mathcal{A} \arrow[r, "G"] \arrow[d, "F"'] & \mathcal{A}' \arrow[d, "F'"] \\ \mathcal{A} \arrow[r, "G"] & \mathcal{A}' \end{tikzcd}$$

be given and an isomorphism

$$\gamma : F'G \to GF$$

of $\mathsf{rep}(\mathcal{A}, \mathcal{A}')$. We assume, as we may, that $\mathcal{A}$ and $\mathcal{A}'$ are cofibrant in dgcat and that F, F' and G are given by cofibrant bimodules. Then $F'G$ is a cofibrant bimodule and so $\gamma : F'G \to GF$ lifts to a morphism of bimodules

$$\widetilde{\gamma} : F'G \to GF.$$

If $\mathcal{B}$ is another dg category and (P, ϕ) an object of $\mathsf{eff}_0(\mathcal{A}', F', \mathcal{B})$, then the composition

$$PG \xrightarrow{\phi G} PF'G \xrightarrow{P\widetilde{\gamma}} PGF$$

yields an object $PG \to PGF$ of $\mathsf{eff}_0(\mathcal{A}, F, \mathcal{B})$. Clearly, this assignment extends to a functor, which induces a functor

$$\mathsf{eff}(\mathcal{A}', F', \mathcal{B}) \to \mathsf{eff}(\mathcal{A}, F, \mathcal{B}).$$

By the 2-universal property of section 9.3, we obtain an induced morphism

$$\overline{G} : \mathcal{A}/F \to \mathcal{A}'/F'.$$

One checks that the composition of two pairs (G, γ) and (G', γ') induces a functor isomorphic to the composition of $\overline{G}$ with $\overline{G'}$.

9.5 The bicategory of enhanced triangulated categories

We refer to [33, 2.1] for the notion of an *exact dg category*. We also call these categories *pretriangulated* since if $\mathcal{A}$ is an exact dg category, then $H^0(\mathcal{A})$ is triangulated. More precisely, $\mathcal{E} = Z^0(\mathcal{A})$ is a Frobenius category and $H^0(\mathcal{A})$ is its associated stable category $\underline{\mathcal{E}}$ (*cf.* example (3) of section 8.3 for these notions). The inclusion of the full subcategory of (small) exact dg categories into $\mathsf{Ho}(\mathsf{dgcat})$ admits a left adjoint, namely the functor $\mathcal{A} \mapsto \mathsf{pretr}(\mathcal{A})$ which maps a dg category to its 'pretriangulated hull' defined in [14], *cf.* also [33, 2.2]. More precisely, the adjunction morphism $\mathcal{A} \to \mathsf{pretr}(\mathcal{A})$ induces an equivalence of categories

$$\mathsf{rep}\,(\mathsf{pretr}(\mathcal{A}), \mathcal{B}) \to \mathsf{rep}(\mathcal{A}, \mathcal{B})$$

for each exact dg category $\mathcal{B}$, *cf.* [55].
The *bicategory* enh *of enhanced [14] triangulated categories*, *cf.* [32] [21], has as objects all small exact dg categories; the morphism category between two objects $\mathcal{A}$, $\mathcal{B}$ is $\mathsf{rep}(\mathcal{A}, \mathcal{B})$; the composition bifunctor

$$\mathsf{rep}(\mathcal{B}, \mathcal{C}) \times \mathsf{rep}(\mathcal{A}, \mathcal{B}) \to \mathsf{rep}(\mathcal{A}, \mathcal{C})$$

is given by the derived tensor product $(X, Y) \mapsto X \overset{L}{\otimes}_{\mathcal{B}} Y$.

9.6 Exact dg orbit categories

Now let $\mathcal{A}$ be an exact dg category and $F \in \mathsf{rep}(\mathcal{A}, \mathcal{A})$. Then $\mathcal{A}/F$ is the dg orbit category of subsection 5.1 and $\mathsf{pretr}(\mathcal{A}/F)$ is an exact dg category such that $H^0\mathsf{pretr}(\mathcal{A}/F)$ is the triangulated hull of section 5. In particular, we obtain that the triangulated hull is the stable category of a Frobenius category. From the construction, we obtain the universal property:

Theorem 4 *For each exact dg category $\mathcal{B}$, we have an equivalence of categories*

$$\mathsf{rep}(\mathsf{pretr}(\mathcal{A}/F), \mathcal{B}) \to \mathsf{eff}(F, \mathcal{B}).$$

9.7 An example

Let A be a finite-dimensional algebra of finite global dimension and TA the trivial extension algebra, *i.e.* the vector space $A \oplus DA$ endowed with the multiplication defined by

$$(a, f)(b, g) = (ab, ag + fb)\ ,\ (a, f), (b, g) \in TA\ ,$$

and the grading such that A is in degree 0 and DA in degree 1. Let $F : \mathcal{D}^b(A) \to \mathcal{D}^b(A)$ equal $\tau S^2 = \nu S$ and let $\widetilde{F}$ be the dg lift of F given by $? \otimes_A R[1]$, where R is a projective bimodule resolution of DA. Let $\mathcal{D}^b(A)_{dg}$ denote a dg category quasi-equivalent to the dg category of bounded complexes of finitely generated projective A-modules.

Theorem 5 *The following are equivalent*

(i) *The k-category $\mathcal{D}^b(A)/F$ is naturally equivalent to its 'triangulated hull'*

$$H^0(\mathsf{pretr}(\mathcal{D}^b(A)_{dg}/\widetilde{F})).$$

(ii) *Each finite-dimensional TA-module admits a grading.*

Proof. We have a natural functor

$$\mathsf{mod}A \to \mathsf{grmod}TA$$

given by viewing an A-module as a graded TA-module concentrated in degree 0. As shown by D. Happel [27], *cf.* also [34], this functor extends to a triangle equivalence Φ from $\mathcal{D}^b(A)$ to the stable category $\underline{\mathsf{grmod}}\,TA$, obtained from $\mathsf{grmod}TA$ by killing all morphisms factoring through projective-injectives. We would like to show that we have an isomorphism of triangle functors

$$\Phi \circ \tau S^2 \xrightarrow{\sim} \Sigma \circ \Phi$$

where Σ is the grading shift functor for graded TA-modules: $(\Sigma M)^p = M^{p+1}$ for all $p \in \mathbf{Z}$. ¿From [27], we know that $\tau S \xrightarrow{\sim} \nu$, where $\nu = ? \overset{L}{\otimes}_A DA$. Thus it remains to show that

$$\Phi \circ \nu S \xrightarrow{\sim} \Sigma \circ \Phi.$$

As shown in [34], the equivalence Φ is given as the composition

$$\begin{aligned}\mathcal{D}^b(\mathsf{mod}A) \to \quad & \mathcal{D}^b(\mathsf{grmod}TA) \to \\ & \mathcal{D}^b(\mathsf{grmod}TA)/\mathsf{per}(\mathsf{grmod}TA) \to \underline{\mathsf{grmod}}\,TA\,,\end{aligned}$$

where the first functor is induced by the above inclusion, the notation $\mathsf{per}(\mathsf{grmod}TA)$ denotes the triangulated subcategory generated by the projective-injective TA-modules and the last functor is the 'stabilization functor' *cf.* [34]. We have a short exact sequence of graded TA-modules

$$0 \to \Sigma^{-1}(DA) \to TA \to A \to 0.$$

We can also view it as a sequence of left A and right graded TA-modules. Let P be a bounded complex of projective A-modules. Then we obtain a short exact sequence of complexes of graded TA-modules

$$0 \to \Sigma^{-1}(P \otimes_A DA) \to P \otimes_A TA \to P \to 0$$

functorial in P. It yields a functorial triangle in $\mathcal{D}^b(\mathsf{grmod}A)$. The second term belongs to $\mathsf{per}(\mathsf{grmod}TA)$. Thus in the quotient category

$$\mathcal{D}^b(\mathsf{grmod}TA)/\mathsf{per}(\mathsf{grmod}TA)\,,$$

the triangle reduces to a functorial isomorphism

$$P \xrightarrow{\sim} S\Sigma^{-1}\nu P.$$

Thus we have a functorial isomorphism

$$\Phi(P) \xrightarrow{\sim} S\Sigma^{-1}\Phi(\nu P).$$

Since A is of finite global dimension, $\mathcal{D}^b(\mathsf{mod}A)$ is equivalent to the homotopy category of bounded complexes of finitely generated projective A-modules. Thus we get the required isomorphism

$$\Sigma\Phi \xrightarrow{\sim} \Phi S\nu.$$

More precisely, one can show that $\underline{\mathsf{grmod}}\,TA$ has a canonical dg structure and that there is an isomorphism

$$(\mathcal{D}^b(A))_{dg} \xrightarrow{\sim} (\underline{\mathsf{grmod}}\,TA)_{dg}$$

in the homotopy category of small dg categories which induces Happel's equivalence and under which Σ corresponds to the lift $\widetilde{F}$ of $F = S\nu$. Hence the orbit categories $\mathcal{D}^b(\mathsf{mod}A)/\tau S^2$ and $\underline{\mathsf{grmod}}\,TA/\Sigma$ are equivalent and we are reduced to determining when $\underline{\mathsf{grmod}}\,TA/\Sigma$ is naturally equivalent to its triangulated hull. Clearly, we have a full embedding

$$\underline{\mathsf{grmod}}\,TA/\Sigma \to \underline{\mathsf{mod}}\,TA$$

and its image is formed by the TA-modules which admit a grading. Now $\underline{\mathsf{mod}}\,TA$ is naturally equivalent to the triangulated hull. Therefore, condition (i) holds iff the embedding is an equivalence iff each finite-dimensional TA-module admits a grading.

In [53], A. Skowroński has produced a class of examples where condition (ii) does not hold. The simplest of these is the algebra A given by the quiver

$$\begin{tikzcd}
& \bullet \arrow[dl, "\beta"'] \arrow[dd] \\
\bullet \arrow[dr, "\alpha"'] & \\
& \bullet
\end{tikzcd}$$

with the relation $\alpha\beta = 0$. Note that this algebra is of global dimension 2.

9.8 Exact categories and standard functors

Let $\mathcal{E}$ be a small exact k-category. Denote by $\mathcal{C}^b(\mathcal{E})$ the category of bounded complexes over $\mathcal{E}$ and by $\mathcal{A}c^b(\mathcal{E})$ its full subcategory formed by the acyclic bounded complexes. The categories with the same objects but whose morphisms are given by the morphism *complexes* are denoted respectively by $\mathcal{C}^b(\mathcal{E})_{dg}$ and $\mathcal{A}c^b(\mathcal{E})_{dg}$. They are exact dg categories and so is the dg quotient [33] [21]

$$\mathcal{D}^b(\mathcal{E})_{dg} = \mathcal{C}^b(\mathcal{E})_{dg}/\mathcal{A}c^b(\mathcal{E})_{dg}.$$

Let $\mathcal{E}'$ be another small exact k-category. We call a triangle functor $F : \mathcal{D}^b(\mathcal{E}) \to \mathcal{D}^b(\mathcal{E}')$ a *standard functor* if it is isomorphic to the triangle functor induced by a morphism

$$\widetilde{F} : \mathcal{D}^b(\mathcal{E})_{dg} \to \mathcal{D}^b(\mathcal{E}')_{dg}$$

of $\mathsf{Ho}(\mathsf{dgcat})$. Slightly abusively, we then call $\widetilde{F}$ a *dg lift* of F. Each exact functor $\mathcal{E} \to \mathcal{E}'$ yields a standard functor; a triangle functor is standard iff it admits a lift to an object of $\mathsf{rep}(\mathcal{D}^b(\mathcal{E})_{dg}, \mathcal{D}^b(\mathcal{E}')_{dg})$; compositions of standard functors are standard; an adjoint (and in particular, the inverse) of a standard functor is standard.
If $F : \mathcal{D}^b(\mathcal{E}) \to \mathcal{D}^b(\mathcal{E})$ is a standard functor with dg lift $\widetilde{F}$, we have the dg orbit category $\mathcal{D}^b(\mathcal{E})_{dg}/\widetilde{F}$ and its pretriangulated hull

$$\mathcal{D}^b(\mathcal{E})_{dg}/\widetilde{F} \to \mathsf{pretr}(\mathcal{D}^b(\mathcal{E})_{dg}/\widetilde{F}).$$

The examples in section 3 show that this functor is not an equivalence in general.

9.9 Hereditary categories

Now suppose that $\mathcal{H}$ is a small hereditary abelian k-category with the Krull-Schmidt property (indecomposables have local endomorphism rings and each object is a finite direct sum of indecomposables) where all morphism and extension spaces are finite-dimensional. Let

$$F : \mathcal{D}^b(\mathcal{H}) \to \mathcal{D}^b(\mathcal{H})$$

be a standard functor with dg lift $\widetilde{F}$.

Theorem 6 *Suppose that F satisfies assumptions 2) and 3) of the main theorem in section 4. Then the canonical functor*

$$\mathcal{D}^b(\mathcal{H})/F \to H^0(\mathsf{pretr}(\mathcal{D}^b(\mathcal{H})_{dg}/\widetilde{F}))$$

is an equivalence of k-categories. In particular, the orbit category $\mathcal{D}^b(\mathcal{H})/F$ admits a triangulated structure such that the projection functor becomes a triangle functor.

The proof is an adaptation, left to the reader, of the proof of the main theorem.

Suppose for example that $\mathcal{D}^b(\mathcal{H})$ has a Serre functor ν. Then ν is a standard functor since it is induced by the tensor product with bimodule

$$(A, B) \mapsto D\mathsf{Hom}_{\mathcal{D}^b(\mathcal{H})_{dg}}(B, A)\ ,$$

where $D = \mathsf{Hom}_k(?, k)$. The functor $\tau^{-1} = S\nu^{-1}$ induces equivalences $\mathcal{I} \xrightarrow{\sim} S\mathcal{P}$ and $\mathcal{H}_{ni} \to \mathcal{H}_{np}$, where $\mathcal{P}$ is the subcategory of projectives, $\mathcal{I}$ the subcategory of injectives, $\mathcal{H}_{np}$ the subcategory of objects without a projective direct summand and $\mathcal{H}_{ni}$ the subcategory of objects without an injective direct summand. Now let $n \geq 2$ and consider the autoequivalence $F = S^n\nu^{-1} = S^{n-1}\tau^{-1}$ of $\mathcal{D}^b(\mathcal{H})$. Clearly F is standard. It is not hard to see that F satisfies the hypotheses 2) and 3) of the main theorem in section 4. Thus the orbit category

$$\mathcal{D}^b(\mathcal{H})/F = \mathcal{D}^b(\mathcal{H})/S^n\nu^{-1}$$

is triangulated. Note that we have excluded the case $n = 1$ since the hypotheses 2) and 3) are not satisfied in this case, in general, as we see from the last example in section 3.

References

[1] M. Artin, J. Tate, and M. Van den Bergh, *Some algebras associated to automorphisms of elliptic curves*, The Grothendieck Festschrift, Vol. I, Progr. Math., vol. 86, Birkhäuser Boston, Boston, MA, 1990, pp. 33–85.

[2] Michael Artin and William F. Schelter, *Graded algebras of global dimension* 3, Adv. in Math. 66 (1987), no. 2, 171–216.

[3] Maurice Auslander and Idun Reiten, *Stable equivalence of Artin algebras*, Proceedings of the Conference on Orders, Group Rings and Related Topics (Ohio State Univ., Columbus, Ohio, 1972) (Berlin), Springer, 1973, pp. 8–71. Lecture Notes in Math., Vol. 353.

[4] ———, *Stable equivalence of dualizing R-varieties*, Advances in Math. 12 (1974), 306–366.

[5] ———, *Stable equivalence of dualizing R-varieties. II. Hereditary dualizing R-varieties*, Advances in Math. 17 (1975), no. 2, 93–121.

[6] ———, *Almost split sequences for rational double points*, Trans. Amer. Math. Soc. 302 (1987), no. 1, 87–97.

[7] ———, *$D\,Tr$-periodic modules and functors*, Representation theory of algebras (Cocoyoc, 1994), CMS Conf. Proc., vol. 18, Amer. Math. Soc., Providence, RI, 1996, pp. 39–50.

[8] Aslak Bakke Buan, Robert J. Marsh, Markus Reineke, Idun Reiten, and Gordana Todorov, *Tilting theory and cluster combinatorics*, Advances in Mathematics, to appear.

[9] Aslak Bakke Buan, Robert J. Marsh, and Idun Reiten, *Cluster-tilted algebras*, Trans. Amer. Math. Soc., to appear.

[10] Alexander A. Beilinson, Joseph Bernstein, and Pierre Deligne, *Analyse et topologie sur les espaces singuliers*, Astérisque, vol. 100, Soc. Math. France, 1982 (French).

[11] David Benson, Henning Krause, and Stefan Schwede, *Realizability of modules over Tate cohomology*, Trans. Amer. Math. Soc. 356 (2004), no. 9, 3621–3668 (electronic).

[12] Jerzy Białkowski, Karin Erdmann, and Andrzej Skowroński, *Deformed preprojective algebras of generalized Dynkin type*, Preprint, 2004.

[13] A. Bondal and M. van den Bergh, *Generators and representability of functors in commutative and noncommutative geometry*, Mosc. Math. J. 3 (2003), no. 1, 1–36, 258.

[14] A. I. Bondal and M. M. Kapranov, *Enhanced triangulated categories*, Math. USSR Sbornik 70 (1991), 93–107.

[15] Sheila Brenner, Michael C. R. Butler, and Alastair D. King, *Periodic algebras which are almost Koszul*, Algebr. Represent. Theory 5 (2002), no. 4, 331–367.

[16] Philippe Caldero, Frédéric Chapoton, and Ralf Schiffler, *Quivers with relations arising from clusters (A_n case)*, arXiv:math.RT/0401316.

[17] Claude Cibils and Eduardo Marcos, *Skew category, Galois covering and smash product of a k-category*, arXiv:math.RA/0312214, to appear in the Proceedings of the AMS.

[18] Kevin Costello, *Topological conformal field theories and Calabi-Yau categories*, arXiv:math.QA/0412149.

[19] Pierre Deligne, *Cohomologie à supports propres*, pp. vi+640, Springer-Verlag, Berlin, 1973, Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964 (SGA 4), Dirigé par M. Artin, A. Grothendieck et J. L. Verdier. Avec la collaboration de P. Deligne et B. Saint-Donat, Lecture Notes in Mathematics, Vol. 305.

[20] Vlastimil Dlab and Claus Michael Ringel, *The preprojective algebra of a modulated graph*, Representation theory, II (Proc. Second Internat. Conf., Carleton Univ., Ottawa, Ont., 1979), Lecture Notes in Math., vol. 832, Springer, Berlin, 1980, pp. 216–231.

[21] Vladimir Drinfeld, *DG quotients of DG categories*, J. Algebra 272 (2004), no. 2, 643–691.

[22] W. G. Dwyer and J. Spaliński, *Homotopy theories and model categories*, Handbook of algebraic topology, North-Holland, Amsterdam, 1995, pp. 73–126.

[23] Karin Erdmann and Nicole Snashall, *On Hochschild cohomology of preprojective algebras. I, II*, J. Algebra 205 (1998), no. 2, 391–412, 413–434.

[24] Peter Freyd, *Stable homotopy*, Proc. Conf. Categorical Algebra (La Jolla, Calif., 1965), Springer, New York, 1966, pp. 121–172.

[25] Christof Geiss, *The universal cover of a preprojective algebra*, Talk at the 11th International Conference on Representations of Algebras, Pátzcuaro, August 2004.

[26] I. M. Gel′fand and V. A. Ponomarev, *Model algebras and representations of graphs*, Funktsional. Anal. i Prilozhen. 13 (1979), no. 3, 1–12.

[27] Dieter Happel, *On the derived category of a finite-dimensional algebra*, Comment. Math. Helv. 62 (1987), no. 3, 339–389.

[28] Dieter Happel, *On Gorenstein algebras*, Representation theory of finite groups and finite-dimensional algebras (Bielefeld, 1991), Progr. Math., vol. 95, Birkhäuser, Basel, 1991, pp. 389–404.

[29] Dieter Happel, Udo Preiser, and Claus Michael Ringel, *Binary polyhedral groups and Euclidean diagrams*, Manuscripta Math. 31 (1980), no. 1-3, 317–329.

[30] ———, *Vinberg's characterization of Dynkin diagrams using subadditive functions with application to DTr-periodic modules*, Representation theory, II (Proc. Second Internat. Conf., Carleton Univ., Ottawa, Ont., 1979), Lecture Notes in Math., vol. 832, Springer, Berlin, 1980, pp. 280–294.

[31] Mark Hovey, *Model categories*, Mathematical Surveys and Monographs, vol. 63, American Mathematical Society, Providence, RI, 1999.

[32] Bernhard Keller, *Deriving DG categories*, Ann. Sci. École Norm. Sup. (4) 27 (1994), no. 1, 63–102.

[33] ———, *On the cyclic homology of exact categories*, J. Pure Appl. Algebra 136 (1999), no. 1, 1–56.

[34] Bernhard Keller and Dieter Vossieck, *Sous les catégories dérivées*, C. R. Acad. Sci. Paris Sér. I Math. 305 (1987), no. 6, 225–228.

[35] ———, *Aisles in derived categories*, Bull. Soc. Math. Belg. Sér. A 40 (1988), no. 2, 239–253.

[36] Mikhail Khovanov and Paul Seidel, *Quivers, Floer cohomology, and braid group actions*, J. Amer. Math. Soc. 15 (2002), no. 1, 203–271 (electronic).

[37] Maxim Kontsevich, *Topological field theory for triangulated categories*, Talk at the conference on K-Theory and Noncommutative Geometry, Institut Henri Poincaré, Paris, June 2004.

[38] ———, *Feynman diagrams and low-dimensional topology*, First European Congress of Mathematics, Vol. II (Paris, 1992), Progr. Math., vol. 120, Birkhäuser, Basel, 1994, pp. 97–121.

[39] ———, *Triangulated categories and geometry*, Course at the École Normale Supérieure, Paris, Notes taken by J. Bellaïche, J.-F. Dat, I. Marin, G. Racinet and H. Randriambololona, 1998.

[40] Henning Krause, *The stable derived category of a Noetherian scheme*, Compos. Math. 141 (2005), no. 5, 1128–1162.

[41] D. M. Lu, J. H. Palmieri, Q. S. Wu, and J. J. Zhang, *A_∞-algebras for ring theorists*, Proceedings of the International Conference on Algebra, vol. 11, 2004, pp. 91–128.

[42] Amnon Neeman, *Triangulated Categories*, Annals of Mathematics Studies, vol. 148, Princeton University Press, Princeton, NJ, 2001.

[43] Liangang Peng and Jie Xiao, *Root categories and simple Lie algebras*, J. Algebra 198 (1997), no. 1, 19–56.

[44] Idun Reiten, *Finite-dimensional algebras and singularities*, Singularities, representation of algebras, and vector bundles (Lambrecht, 1985), Lecture Notes in Math., vol. 1273, Springer, Berlin, 1987, pp. 35–57.

[45] ———, *Hereditary categories with Serre duality*, Representations of algebra. Vol. I, II, Beijing Norm. Univ. Press, Beijing, 2002, pp. 109–121.

[46] Idun Reiten and Michel Van den Bergh, *Two-dimensional tame and maximal orders of finite representation type*, Mem. Amer. Math. Soc. 80 (1989), no. 408, viii+72.

[47] ———, *Noetherian hereditary abelian categories satisfying Serre duality*, J. Amer. Math. Soc. 15 (2002), no. 2, 295–366 (electronic).

[48] Jeremy Rickard, *Derived equivalences as derived functors*, J. London Math. Soc. (2) 43 (1991), no. 1, 37–48.

[49] C. Riedtmann, *Algebren, Darstellungsköcher, Überlagerungen und zurück*, Comment. Math. Helv. 55 (1980), no. 2, 199–224.

[50] Claus Michael Ringel, *The preprojective algebra of a quiver*, Algebras and modules, II (Geiranger, 1996), CMS Conf. Proc., vol. 24, Amer. Math. Soc., Providence, RI, 1998, pp. 467–480.

[51] Raphaël Rouquier, *Dimensions of triangulated categories*, arXiv:math.CT/0310134.

[52] Raphaël Rouquier and Alexander Zimmermann, *Picard groups for derived module categories*, Proc. London Math. Soc. (3) 87 (2003), no. 1, 197–225.

[53] Andrzej Skowroński, *Generalization of Yamagata's theorem on trivial extensions*, Arch. Math. (Basel) 48 (1987), no. 1, 68–76.

[54] Gonçalo Tabuada, *On the homotopy theory of dg categories*, Ph. D. thesis in preparation.

[55] ______, *Invariants additifs de dg-catégories*, Int. Math. Res. Notices 53 (2005), 3309–3339.

[56] ______, *Une structure de catégorie de modèles de Quillen sur la catégorie des dg-catégories*, C. R. Math. Acad. Sci. Paris 340 (2005), no. 1, 15–19.

[57] Bertrand Toën, *The homotopy theory of dg-categories and derived Morita theory*, arXiv:math.AG/0408337.

[58] Jie Xiao and Bin Zhu, *Relations for the Grothendieck groups of triangulated categories*, J. Algebra 257 (2002), no. 1, 37–50.

Bernhard Keller
UFR de Mathématiques
Institut de Mathématiques
UMR 7586 du CNRS
Case 7012
Université Paris 7
2, place Jussieu
75251 Paris Cedex 05
France
keller@math.jussieu.fr

A Common Recursion For Laplacians of Matroids and Shifted Simplicial Complexes

Art M. Duval

Received: June 10, 2004
Revised: March 31, 2005

Communicated by Günter Ziegler

Abstract. A recursion due to Kook expresses the Laplacian eigenvalues of a matroid M in terms of the eigenvalues of its deletion $M-e$ and contraction M/e by a fixed element e, and an error term. We show that this error term is given simply by the Laplacian eigenvalues of the pair $(M-e, M/e)$. We further show that by suitably generalizing deletion and contraction to arbitrary simplicial complexes, the Laplacian eigenvalues of shifted simplicial complexes satisfy this exact same recursion. We show that the class of simplicial complexes satisfying this recursion is closed under a wide variety of natural operations, and that several specializations of this recursion reduce to basic recursions for natural invariants.

We also find a simple formula for the Laplacian eigenvalues of an arbitrary pair of shifted complexes in terms of a kind of generalized degree sequence.

2000 Mathematics Subject Classification: Primary 15A18; Secondary 05B35, 05E99
Keywords and Phrases: Laplacian, spectra, matroid complex, shifted simplicial complex, Tutte polynomial

1. Introduction

The independence complex of matroids and shifted simplicial complexes are two of only four types of simplicial complexes whose combinatorial Laplacians $L = \partial\partial^* + \partial^*\partial$ are known to have only integer eigenvalues (see Kook, Reiner, and Stanton [27], and [16], respectively). The other two types, which will not concern us further, are matching complexes of complete graphs [14] and chessboard complexes [21]. More information and background about the combinatorial Laplacian and its eigenvalues may be found in Section 2 and [16, 20, 27]. Our main result (Theorems 3.18 and 4.23) is another, more striking, similarity between the Laplacian eigenvalues of matroids and shifted complexes: they satisfy the exact same recursion, which we call the *spectral recursion*, equation (2). This recursion is stated in terms of the *spectrum polynomial*, a natural generating function for Laplacian eigenvalues, defined in equation (1).

The Tutte polynomial T_M of a matroid M satisfies the recursion $T_M = T_{M-e} + T_{M/e}$, when e is neither a loop nor an isthmus, and where $M-e$ and M/e denote the deletion and contraction, respectively, of M with respect to ground element e. When Kook, Reiner, and Stanton proved that the Laplacian spectrum of a matroid is integral, they also speculated on the existence of a Tutte polynomial-like recursion for the spectrum polynomial of a matroid M, though possibly with a third "error" term, besides the deletion and contraction, on the right-hand side [27, Question 3]. Kook [26] found such a recursion, but the error term in his formulation is somewhat complicated to state, with two cases depending on whether or not the ground element e is a closed element in M. Subsequently, Kook and Reiner (private communication) asked if this error term might be just the spectrum polynomial of the matroid pair $(M-e, M/e)$.

One of our main results (Theorem 3.18) is that Kook and Reiner's conjecture is true, that is, the spectrum polynomial of M can be expressed simply in terms of the spectrum polynomials of $M-e$, M/e, and $(M-e, M/e)$. This is the spectral recursion. We show, furthermore, by suitably generalizing the definitions of deletion and contraction from matroids to arbitrary simplicial complexes (Section 2), that shifted complexes also satisfy the spectral recursion (Theorem 4.23).

This raises the natural question: What is the largest class of simplicial complexes, necessarily a common generalization of matroids and shifted complexes, satisfying the spectral recursion? We will see that this class is closed under the operations of join, skeleta, Alexander dual, and disjoint union (Corollaries 4.5, 4.19, 6.8, and 6.11, respectively). We might hope that it is closed also under deletion and contraction, as matroids and shifted complexes each are. In the same vein, it may be worthwhile to restrict our attention to those complexes that are also Laplacian integral. Unfortunately, no hint to determining this common generalization is apparent in the proofs of either Laplacian integrality or the spectral recursion, which are each rather different for matroids and shifted complexes.

Jarrah and Laubenbacher [23] examined another property shared by matroids and shifted complexes. Klivans [24] has characterized simplicial complexes that are simultaneously shifted *and* the matroid complex of some matroid; this is, in some sense, the reverse of finding a natural common generalization of matroids and shifted complexes.

The common generalization includes neither of the other known types of Laplacian integral simplicial complexes. Direct computations show that the matching complex of the complete graph on 5 vertices and the 2×3 chessboard complex both fail to satisfy the spectral recursion with respect to any vertex. Also excluded is the 3-edge path (Example 2.5), which rules out as the common generalization such otherwise likely candidates as vertex-decomposable [32][9, Section 11] or shellable complexes [8, 9].

A key piece of the proof that matroids satisfy the spectral recursion is a decomposition of the Laplacian of $(M-e, M/e)$ into a direct sum of Laplacians of M/C's for all circuits C containing e (Lemma 3.3). We may combine this with the spectral recursion to express the spectrum polynomial of a matroid completely in terms of spectrum polynomials of smaller matroids (with no matroid pairs), which permits a truly recursive way of computing Laplacian eigenvalues for matroids (Remark 3.19).

Unfortunately, we are unable to state any formula for the Laplacian eigenvalues of an *arbitrary* matroid pair (*i.e.*, besides $(M-e, M/e)$). We are able, however, to use tools developed in the proof of the spectral recursion for shifted complexes to find a simple formula for the Laplacian eigenvalues of an arbitrary shifted simplicial pair (Theorem 5.7). This naturally generalizes a formula for a single shifted complex [16]; the graph case goes back to Merris [29]. Similarly, we generalize a related conjectured inequality on the Laplacian spectrum of an arbitrary simplicial complex [16] to an arbitrary simplicial pair (Conjecture 5.8); the graph case was conjectured by Grone and Merris [22]. Passing from graphs to simplicial complexes in [16] required generalizing the well-known notion of degree sequences for graphs. Now passing to simplicial *pairs*, we introduce a less than obvious, but perfectly natural, further generalization of degree sequence (Subsection 5.2).

The Tutte polynomial is arguably the most important invariant of matroid theory (see, *e.g.*, [12]). The spectrum polynomial shares several nice features with the Tutte polynomial, such as being well-behaved under join (Corollary 4.3), disjoint union (Lemma 6.9), and several dual operators (equations (29) and (32)). Furthermore, specializations obtained by plugging in particular values for one or the other of the variables of the spectrum polynomial reduce it to well-known invariants. Consequently (and now going beyond matroids and the Tutte polynomial), in each of these specializations, the spectral recursion holds for all simplicial complexes Δ (not just matroids and shifted complexes), because it reduces to a basic recursion expressing the relevant invariant for Δ in terms of that invariant for $\Delta - e$ and Δ/e (Theorem 2.4 and Corollary 4.8).

In contrast to the Tutte polynomial recursion, the spectral recursion does not need to exclude loops and isthmuses as special cases. Indeed, the spectral recursion holds for all complexes (not just matroids and shifted complexes) when e is a loop (Proposition 2.3) or an isthmus (Proposition 2.2 and Theorem 2.4).

Section 2 contains more information about Laplacians and the spectral recursion, including some special cases. Sections 3 and 4 are devoted to the proofs that matroids and shifted complexes, respectively, satisfy the spectral recursion. The formula for eigenvalues of arbitrary shifted simplicial pairs is developed in Section 5. Finally, in Section 6, we show that disjoint union and several duality operators, including Alexander duality, all preserve the property of satisfying the spectral recursion.

2. Laplacians of simplicial pairs

For further background on simplicial complexes, their boundary maps and homology groups, see, *e.g.*, [30, Chapter 1]. If Δ and Δ' are simplicial complexes on the same ground set of vertices, then we will say (Δ, Δ') is a *simplicial pair*, but we set $(\Delta, \Delta') = (\Gamma, \Gamma')$ when the set differences $\Delta \backslash \Delta'$ and $\Gamma \backslash \Gamma'$ are equal as subsets of the power set of the ground set of vertices (here $A \backslash B$ denotes the set difference $\{a \in A : a \notin B\}$ between sets A and B); more formally, then, a simplicial pair is an equivalence class on ordered pairs of simplicial complexes. In all cases, definitions applying to a simplicial pair (Δ, Δ') may be specialized to a single simplicial complex Δ, by letting $\Delta' = \emptyset$, the empty simplicial complex.

As usual, let $C_i = C_i(\Delta, \Delta'; \mathbb{R}) := C_i(\Delta; \mathbb{R})/C_i(\Delta'; \mathbb{R})$ denote the i-dimensional oriented $\mathbb{R}$-chains of (Δ, Δ'), *i.e.*, the formal $\mathbb{R}$-linear sums of oriented i-dimensional faces $[F]$ such that $F \in \Delta_i \backslash \Delta'_i$, where Δ_i denotes the set of i-dimensional faces of Δ. Let $\partial_{(\Delta,\Delta');i} = \partial_i \colon C_i \to C_{i-1}$ denote the usual (signed) *boundary operator*. Via the natural bases $\Delta_i \backslash \Delta'_i$ and $\Delta_{i-1} \backslash \Delta'_{i-1}$ for $C_i(\Delta, \Delta'; \mathbb{R})$ and $C_{i-1}(\Delta, \Delta'; \mathbb{R})$, respectively, the boundary map ∂_i has an adjoint map $\partial_i^* \colon C_{i-1}(\Delta, \Delta'; \mathbb{R}) \to C_i(\Delta, \Delta'; \mathbb{R})$; *i.e.*, the matrices representing ∂ and ∂^* in the natural bases are transposes of one another.

Definition. Let $L'_i = \partial_{i+1}\partial^*_{i+1}$ and $L''_i = \partial^*_i \partial_i$. Then the (*$i$-dimensional*) *Laplacian* of (Δ, Δ') is the map $L_i(\Delta, \Delta') \colon C_i(\Delta, \Delta'; \mathbb{R}) \to C_i(\Delta, \Delta'; \mathbb{R})$ defined by

$$L_i = L_i(\Delta, \Delta') := L'_i + L''_i = \partial_{i+1}\partial^*_{i+1} + \partial^*_i \partial_i.$$

For more information, see, *e.g.*, [16, 20, 27]. Laplacians of pairs of graphs were considered in [13]. Each of L'_i and L''_i is positive semidefinite, since each is the composition of a linear map and its adjoint. Therefore, their sum L_i is also positive semidefinite, and so has only non-negative real eigenvalues. (See also Proposition 4.6 and [20, Proposition 2.1].) These eigenvalues do not depend on the arbitrary ordering of the vertices of Δ, and are thus invariants of (Δ, Δ'); see, *e.g.*, [16, Remark 3.2]. Define $\mathbf{s}_i(\Delta, \Delta')$ to be the multiset of eigenvalues of $L_i(\Delta, \Delta')$, and define $m_\lambda(L_i(\Delta, \Delta'))$ to be the multiplicity of λ in $\mathbf{s}_i(\Delta, \Delta')$. The single complex case ($\Delta' = \emptyset$) of the following proposition is the first result of combinatorial Hodge theory, which goes back to Eckmann [18].

Proposition 2.1. *The multiplicity of 0 as an eigenvalue of the i-dimensional Laplacian L_i of (Δ, Δ') is the ith reduced Betti number of (Δ, Δ'),* i.e.*,*

$$m_0(L_i(\Delta, \Delta')) = \tilde{\beta}_i(\Delta, \Delta') = \dim_{\mathbb{R}} \tilde{H}_i(\Delta, \Delta'; \mathbb{R}).$$

Proof. A nice summary is given in the proof of [20, Proposition 2.1]. The usual setup is for just a single simplicial complex (*i.e.*, the special case $\Delta' = \emptyset$), but only depends on the C_i's and ∂_i's forming a chain complex ($\partial^2 = 0$), which still holds even when $\Delta' \neq \emptyset$. (*Cf.* Proposition 4.6.) $\square$

A natural generating function for the Laplacian eigenvalues of a simplicial pair (Δ, Δ') is

$$(1) \qquad S_{(\Delta,\Delta')}(t,q) := \sum_{i\geq 0} t^i \sum_{\lambda\in \mathbf{s}_{i-1}(\Delta,\Delta')} q^\lambda = \sum_{i,\lambda} m_\lambda(L_{i-1}(\Delta,\Delta'))t^i q^\lambda.$$

We call $S_{(\Delta,\Delta')}$ the *spectrum polynomial* of (Δ, Δ'). Although $S_{(\Delta,\Delta')}$ is defined for any simplicial pair (Δ, Δ'), it is only truly a polynomial when the Laplacian eigenvalues are not only non-negative, but integral as well. This will be true for the cases we are concerned with, primarily matroids [27], shifted complexes [16], and shifted simplicial pairs (Theorem 5.7 and Remark 5.9). For the special case of a matroid, a "spectrum polynomial" Spec was defined, differently, in [27], but we will see later that the two definitions agree in this case up to simple changes in indexing (see Lemma 3.6 and [27, Corollary 18]). Letting $\lambda \in \mathbf{s}_{i-1}$ instead of $\lambda \in \mathbf{s}_i$ simplifies the statement of some later results, notably Corollary 4.3.
Recall (*e.g.*, [5, Section 7.3]) the *independence complex* $IN(M)$ of a matroid M on ground set E is the simplicial complex whose faces are the independent sets of M and whose vertex set is E. (For background about matroids, see, *e.g.*, [31, 34, 35].) We will sometimes use M and $IN(M)$ interchangeably, so, for instance, $L_i(M) := L_i(IN(M)) = L_i(IN(M), \emptyset)$ and $S_M := S_{IN(M)} = S_{(IN(M),\emptyset)}$. Similarly, if N is another matroid on the same ground set such that $IN(N) \subseteq IN(M)$ (*i.e.*, $N \leq M$ in the weak order on matroids), then $L_i(M, N) = L_i(IN(M), IN(N))$ and $S_{(M,N)} = S_{(IN(M),IN(N))}$. In this case, we say (M, N) is a *matroid pair*.
We now naturally generalize the notion of deletion and contraction for matroids (see *e.g.*, [11]) to arbitrary simplicial complexes.

Definition. Let Δ be a simplicial complex on vertex set V, and $e \in V$. Then the *deletion* of Δ with respect to e is the simplicial complex

$$\Delta - e = \{F \in \Delta \colon e \notin F\}$$

on vertex set $V - e$, and the *contraction* of Δ with respect to e is the simplicial complex

$$\Delta/e = \{F - e \colon F \in \Delta,\ e \in F\}$$

on vertex set $V - e$. Note that $\Delta/e = \mathrm{lk}_\Delta\, e$, the usual simplicial complex link [30, Section 2]; we use the term "contraction" to highlight similarities to matroid theory.

It is easy to verify that $IN(M - e) = IN(M) - e$ as long as e is not an isthmus of M, and that $IN(M/e) = IN(M)/e$ as long as e is not a loop of M. There is thus no confusion in the notational shortcuts $S_{M-e} := S_{IN(M-e)} = S_{IN(M)-e}$ and $S_{M/e} := S_{IN(M/e)} = S_{IN(M)/e}$ as long as e is not an isthmus or a loop, respectively.
Since e is an isthmus of M precisely when e is a vertex of every facet of $IN(M)$, define e to be an *isthmus* of a simplicial complex Δ if e is a vertex of every facet of Δ (so Δ is a cone with apex e – see Subsection 4.1). Similarly, since e is a loop of M precisely when e is not a vertex of any face of $IN(M)$, define

e to be a *loop* of a simplicial complex Δ if e is in the vertex set of Δ, but in no face of Δ (even the singleton $\{e\}$ is not a face, contrary to usual simplicial complex conventions).

Our definitions mean that if e is an isthmus of simplicial complex Δ, then the deletion $\Delta - e$ equals Δ/e. (When e is an isthmus of a matroid M, the *matroid* deletion $M - e$ is left undefined in *e.g.*, Brylawski [11], though $M - e = M/e$ in Welsh [34, Section 4.2] and Oxley [31, Corollary 3.1.25].) If e is a loop of simplicial complex Δ, then the contraction Δ/e is $\emptyset$, the empty simplicial complex. (When e is a loop of a matroid M, the *matroid* contraction M/e equals $M - e$.)

Definition. We will say that a simplicial complex Δ *satisfies the spectral recursion with respect to* e if e is a vertex of Δ and

$$S_\Delta(t,q) = qS_{\Delta-e}(t,q) + qtS_{\Delta/e}(t,q) + (1-q)S_{(\Delta-e,\Delta/e)}(t,q). \tag{2}$$

We will say Δ *satisfies the spectral recursion* if Δ satisfies the spectral recursion with respect to every vertex in its vertex set. (Note that Proposition 2.3 below means we need not be too particular about the vertex set of Δ.)

Our main result is that Δ satisfies the spectral recursion when Δ is either the independence complex of a matroid (Theorem 3.18) or a shifted simplicial complex (Theorem 4.23), and e is any vertex of Δ. We illustrate now a few special cases of the spectral recursion, which are easy to verify, and some of which are used in later sections.

Proposition 2.2. *The simplicial complex whose sole facet is a single vertex satisfies the spectral recursion.*

Proposition 2.3. *If e is a loop of simplicial complex Δ, then Δ satisfies the spectral recursion with respect to e.*

Proposition 2.2 and Theorem 4.4 will show that, if e is an isthmus of Δ, then Δ satisfies the spectral recursion with respect to e.

Theorem 2.4. *If Δ is any simplicial complex, and e is any vertex of Δ, then the spectral recursion holds when $q = 0$, $q = 1$, $t = 0$, or $t = -1$.*

Proof. Plugging $q = 0$ into S immediately yields $S_{(\Delta,\Delta')}(t,0) = \sum_i t^i \tilde{\beta}_{i-1}(\Delta,\Delta')$, by Proposition 2.1. Proving the spectral recursion in this case then reduces to showing

$$\tilde{\beta}_{i-1}(\Delta) = \tilde{\beta}_{i-1}(\Delta - e, \Delta/e), \tag{3}$$

for all i. This, in turn, is a consequence of the basic topology facts $\tilde{\beta}_{i-1}(\Delta) = \tilde{\beta}_{i-1}(\Delta, \operatorname{st}_\Delta e)$ and $(\Delta, \operatorname{st}_\Delta e) = (\Delta - e, \Delta/e)$, where $\operatorname{st}_\Delta e$ denotes the usual star of e in Δ, the simplicial complex whose facets are the facets of Δ containing e. Setting $q = 1$, we see $S_{(\Delta,\Delta')}(t,1) = \sum_i (f_{i-1}(\Delta) - f_{i-1}(\Delta'))t^i$, where f_i is the number of i-dimensional faces of Δ, since there are as many eigenvalues of $L_{i-1}(\Delta,\Delta')$ as there are faces in $\Delta_{i-1} \setminus \Delta'_{i-1}$ (assuming $\Delta' \subseteq \Delta$). It is then an easy exercise to verify that, when $q = 1$, the t^{i+1} coefficient of the spectral recursion reduces to the easy observation

$$f_i(\Delta) = f_i(\Delta - e) + f_{i-1}(\Delta/e). \tag{4}$$

If we set $t = 0$, it is easy to see that $S_\Delta(0, q) = q^{v(\Delta)}$, where $v(\Delta)$ denotes the number of non-loop vertices of Δ. The spectral recursion in this case reduces to the trivial observation that $v(\Delta) = 1 + v(\Delta - e)$ if e is not a loop, but $v(\Delta) = v(\Delta - e)$ if e is a loop.
We will also see in Corollary 4.8 that, when $t = -1$, the spectral recursion reduces to an easy identity about Euler characteristic. □

In the special case where Δ is a near-cone (see Subsection 4.5) and e is its apex, it is not hard to verify that the $t^{\dim \Delta + 1}$ coefficient of the spectral recursion reduces to [16, Lemma 5.3].
The following complex is the simplest and smallest counterexample to both Laplacian integrality and the spectral recursion.

Example 2.5. Let Δ be the 1-dimensional simplicial complex with vertices a, b, c, d and facets (maximal faces) $\{a, b\}$,$\{b, c\}$, and $\{c, d\}$. It is easy to check directly that $\Delta - e$, Δ/e, and $(\Delta - e, \Delta/e)$ are all Laplacian integral for any choice of e, while Δ is not integral. It then follows immediately that Δ does not satisfy the spectral recursion for any choice of e.

3. Matroids

In this section, we show that the independence complex of a matroid satisfies the spectral recursion, equation (2). The key step of the section is a simple trick in Subsection 3.1 to reduce the problem of computing $S_{(M-e,M/e)}$ to computing $S_{M/C}$ for all circuits C containing e. Subsection 3.2 shows how an algorithm due to Kook, Reiner, and Stanton [27] allows us to compute the spectrum polynomial of a matroid from its combinatorial information; we also compare what this algorithm computes for M, $M - e$, M/e, and M/C. The final steps of the calculation, which largely consist of translating to generating functions the results of the previous subsections, are in Subsection 3.3.
We first set our notation for matroids; for further background, and any terms not defined here, see [35]. Let $M = M(E)$ be a matroid on ground set E. We will let $\mathcal{B} = \mathcal{B}(M)$, $\mathcal{I} = \mathcal{I}(M)$, $\mathcal{C} = \mathcal{C}(M)$, and $\mathcal{F} = \mathcal{F}(M)$ denote the sets of bases, independent sets, circuits, and flats (closed sets) of M, respectively. If $A \subseteq E$, let $\mathrm{rk}_M(A) = \mathrm{rk}(A)$ denote the *rank* of A (with respect to M), and let $\overline{A} = \mathrm{cl}_M(A)$ denote the *closure* of A (with respect to M). We will often write V for $M(V)$ in the special case when V is a flat of M. When $A \subseteq V$, the *set* $V - A$ may be considered to be the *matroid* V/A in matroid M/A, but considered to be the *matroid* $V - A$ in matroid $M - A$. We will also use the notions of internal and external activity as in, *e.g.*, [5].

3.1. A partition. If Δ is a simplicial complex and A is a set disjoint from the vertices of Δ, then let $A \circ \Delta$ denote

$$A \circ \Delta := \{A \dot\cup F \colon F \in \Delta\}.$$

It will soon be important to note that $A \circ \Delta$ is a simplicial pair; in fact $A \circ \Delta = (2^A * \Delta, (2^A \backslash \{A\}) * \Delta)$, where 2^A denotes the simplicial complex consisting of all subsets of A, and $*$ denotes the usual join, as defined in Section 4.

Lemma 3.1. *If Δ is a simplicial complex and A a finite set disjoint from the vertices of Δ, then*

$$S_{A \circ \Delta}(t, q) = t^{|A|} S_\Delta(t, q).$$

Proof. Under the natural bijection between Δ and $A \circ \Delta$, given by $\phi : F \mapsto A \dot\cup F$, the boundary operators ∂_Δ and $\partial_{A \circ \Delta}$ are the same. That is, $\partial_{A \circ \Delta}[A \dot\cup F] = [A]\partial_\Delta[F]$, simply by numbering the vertices of $A \circ \Delta$ so that the elements of A all come last. Since the boundary operators are the same, so are the Laplacians, but the dimension shift in ϕ means $\mathbf{s}_i(\Delta) = \mathbf{s}_{i+|A|}(A \circ \Delta)$. The lemma now follows readily. $\square$

If I is independent in M and $p \in \overline{I} - I$, we will let $\mathrm{ci}(p, I) = \mathrm{ci}_M(p, I) = \mathrm{ci}_{\overline{I}}(p, I)$ be the unique circuit of $\overline{I}$ contained in $I \dot\cup p$. Dually, if $b \in I$, we will let $\mathrm{bo}(b, I) = \mathrm{bo}_M(b, I) = \mathrm{bo}_{\overline{I}}(b, I)$ be the unique bond of $\overline{I}$ contained in $(\overline{I} - I) \dot\cup b$. It is easy to see that if $p \notin \overline{I}$, then $I \in \mathcal{I}(M/p)$. Therefore we may safely refer to $\mathrm{ci}_M(p, I)$ for any $I \in \mathcal{I}(M - p) - \mathcal{I}(M/p)$.

Lemma 3.2. *If $I', I \in \mathcal{I}(M - e) - \mathcal{I}(M/e)$ and $I' \subseteq I$, then $\mathrm{ci}_M(e, I') = \mathrm{ci}_M(e, I)$.*

Proof. From $\mathrm{ci}_M(e, I') \subseteq I' \dot\cup e \subseteq I \dot\cup e$ it follows that $\mathrm{ci}_M(e, I')$ is a circuit in $I \dot\cup e$, and thus the unique circuit in $I \dot\cup e$, *i.e.*, $\mathrm{ci}_M(e, I)$. $\square$

The following lemma is the key step to proving that matroids satisfy the spectral recursion.

Lemma 3.3. *Let $M(E)$ be a matroid, and $e \in E$. If e is not a loop, then*

$$L_i(M - e, M/e) = \bigoplus_{\substack{C \in \mathcal{C}(M) \\ e \in C}} L_i((C - e) \circ IN(M/C)).$$

Proof. For any $C \in \mathcal{C}(M)$ such that $e \in C$, let

$$M_C = \{I \in \mathcal{I}(M - e) - \mathcal{I}(M/e) \colon \mathrm{ci}_M(e, I) = C\};$$

we will see shortly that this is a simplicial pair. By Lemma 3.2,

$$\partial_{(M-e, M/e)}[I] = \partial_C[I]$$

for any $I \in \mathcal{I}(M - e) - \mathcal{I}(M/e)$, where $C = \mathrm{ci}_M(e, I)$. Thus removing M/e from $M - e$ partitions $L_i(M - e, M/e)$ into

$$L_i(M - e, M/e) = \bigoplus_{\substack{C \in \mathcal{C}(M) \\ e \in C}} L_i(M_C).$$

Furthermore, it is easy to see that

$$\begin{aligned}M_C &= \{I \in \mathcal{I}(M-e) \colon C-e \subseteq I\} = (C-e) \circ IN((M-e)/(C-e)) \\ &= (C-e) \circ IN(M/C).\end{aligned}$$

□

3.2. The Kook-Reiner-Stanton algorithm. The decomposition in Proposition 3.4 below was first discovered by Etienne and Las Vergnas [19, Theorem 5.1], but we will rely upon Algorithm 3.5, due to Kook, Reiner, and Stanton [27, proof of Theorem 1], for producing this decomposition.

Proposition 3.4. *Given a base B of matroid M, there is a unique disjoint decomposition $B = B_1 \dot\cup B_2$ into two (necessarily) independent sets such that:*

- *B_1 has internal activity 0; and*
- *B_2 has external activity 0, with respect to the matroid M/V, where $V = \overline{B_1}$.*

Algorithm 3.5. This algorithm produces the decomposition guaranteed by the previous theorem. It takes the base B as input, and outputs the pair (B_1, B_2).

Step 1: Set $B_1 = B$, $B_2 = \emptyset$.
Step 2: Let $V = \overline{B_1}$.
Step 3: Find an internally active element b for B_1 as a base of the flat V.

- If no such element b exists, then stop and output the pair (B_1, B_2).
- If such a b exists, then set $B_1 := B_1 - b$, $B_2 := B_2 \dot\cup b$ (we call this step a *removal*), and return to Step 2.

Notation. If the decomposition of base B in matroid M produced by the above algorithm is $B = B_1 \dot\cup B_2$, then let $\pi(B) = \pi_M(B) = B_1$. If $I \in \mathcal{I}(M)$, then let $\overline{\pi}_M(I) = \mathrm{cl}_V(\pi_V(I)) = \mathrm{cl}_M(\pi_V(I))$, where $V = \mathrm{cl}_M(I)$. If W is any closed set containing I (equivalently, containing $V = \mathrm{cl}_M(I)$), then $\mathrm{cl}_W(I) = \mathrm{cl}_M(I) = V$, and so $\overline{\pi}_W(I) = \mathrm{cl}_V(\pi_V(I)) = \overline{\pi}_M(I)$. In particular, $\overline{\pi}_V(I) = \overline{\pi}_M(I)$.

The following lemma, which is little more than a recasting of [27, Corollary 18] in language tailored to our purposes, reduces computations of the spectrum polynomial to computations of $\overline{\pi}$.

Lemma 3.6. *For any matroid $M(E)$,*

$$S_M(t,q) = q^{|E|} \sum_{I \in \mathcal{I}(M)} t^{\mathrm{rk}(\overline{I})} (q^{-1})^{|\overline{\pi}_M(I)|} = q^{|E|} \sum_{V \in \mathcal{F}(M)} t^{\mathrm{rk}(V)} \sum_{I \in \mathcal{B}(V)} (q^{-1})^{|\overline{\pi}_V(I)|}.$$

Let $\tilde\chi(\Delta) := \sum (-1)^i f_i(\Delta)$ denote the (reduced) Euler characteristic of simplicial complex Δ; we also use the shorthand $\tilde\chi(M) = \tilde\chi(IN(M))$. If $V \subseteq W$ are flats of matroid M, let $\mu(W,V) = \mu_M(W,V)$ denote the Möbius function of the sublattice $[W, V]$ in the lattice of flats of M. The proof of [27, equation (2.2)] shows that

$$\sum_{B\in\mathcal{B}(M)} x^{|\overline{\pi}_M(B)|} = \sum_{V\in\mathcal{F}(M)} |\tilde{\chi}(V)||\mu(V,M)|x^{|V|}. \tag{5}$$

We use the same techniques to do something similar.

Lemma 3.7. *For any matroid $M(E)$, and any $e \in E$,*

$$\sum_{\substack{B\in\mathcal{B}(M)\\ e\in\overline{\pi}_M(B)}} x^{|\overline{\pi}_M(B)|} = \sum_{\substack{V\in\mathcal{F}(M)\\ e\in V}} |\tilde{\chi}(V)||\mu(V,M)|x^{|V|}.$$

In particular, this sum is independent of the linear order on E.

Proof. By Algorithm 3.5 (see also its proof in [27]), there is a bijection between:

- the set $\mathcal{V}$ of triples (V, B_1, B_2) where V is a flat of M, B_1 is a base of internal activity 0 for V (in particular, $V = \overline{B_1}$), and B_2 is a base of external activity 0 for M/V; and
- the set $\mathcal{B}$ of bases B of M.

Furthermore, $B = B_1 \dot\cup B_2$ and $\pi_M(B) = B_1$. Thus

$$\sum_{\substack{B\in\mathcal{B}(M)\\ e\in\overline{\pi}_M(B)}} x^{|\overline{\pi}_M(B)|} = \sum_{\substack{(V,B_1,B_2)\in\mathcal{V}\\ e\in V}} x^{|V|}.$$

We must then determine how many triples (V, B_1, B_2) there are in $\mathcal{V}$ for a fixed flat V. Mimicking an argument from the proof of [27, Theorem 1], we recall from [5, Theorem 7.8.4] that there are $|\tilde{\chi}(V)|$ bases of internal activity 0 for V, and from [5, Proposition 7.4.7] that there are $|\mu(V, M)|$ bases of external activity 0 for M/V. So for every V, there are $|\tilde{\chi}(V)|$ choices for B_1, and, independently, $|\mu(V, M)|$ choices for B_2. Thus,

$$\sum_{\substack{(V,B_1,B_2)\in\mathcal{V}\\ e\in V}} x^{|V|} = \sum_{\substack{V\in\mathcal{F}(M)\\ e\in V}} |\tilde{\chi}(V)||\mu(V,M)|x^{|V|},$$

completing the proof. □

We now see how Algorithm 3.5 works on $M-e$ (Lemma 3.11) and M/e (Lemma 3.13), and on M/C when C is a circuit containing e (Lemma 3.15). We first need three technical lemmas whose easy proofs are omitted. We abuse set difference notation slightly to let $A\backslash x$ denote $\{a \in A\colon a \neq x\}$, when A is a set that may or may not contain element x.

Lemma 3.8. *Let I be an independent set in matroid M, let e be last in the linear order, and assume that $e \notin I$ and that e is not an isthmus of M. Then b is internally active in I (with respect to M) iff b is internally active in I (with respect to $M - e$).*

Lemma 3.9. *Let I be an independent set in matroid M, and let $e, b \in I$. Then b is internally active in I (with respect to M) iff b is internally active in $I - e$ (with respect to M/e).*

Lemma 3.10. *Let I be an independent set in matroid M, and let i be an isthmus in $\overline{I}$. Then $b \neq i$ is internally active in I (with respect to M) iff b is internally active in $I - i$ (with respect to M).*

Lemma 3.11. *Let B be a base of $M - e$, so B is also a base of M and $e \notin B$. Also assume e is last in the linear order. Then $\pi_{M-e}(B) = \pi_M(B)$.*

Proof. Use Algorithm 3.5 to compute $\pi_M(B)$. By Lemma 3.8, every step of the algorithm can be copied in $M - e$; that is, when element b is removed from B_1 in M, we can remove b from B_1 in $M - e$. And also by Lemma 3.8, when there are no more elements to remove from B_1 in M, then there are also no more elements to remove from B_1 in $M - e$. $\square$

Corollary 3.12. *Let I be an independent set of $M-e$, so I is also independent in M and $e \notin I$. Also assume e is last in the linear order. Then*

$$\overline{\pi}_{M-e}(I) = \overline{\pi}_M(I) \backslash e.$$

Lemma 3.13. *Let B be a base of M such that $e \in B$, so $B - e$ is a base of M/e. Also assume e is last in the linear order. Then*

$$\pi_{M/e}(B - e) = \pi_M(B) \backslash e$$

Proof. Again use Algorithm 3.5 to compute $\pi_M(B)$, except do not remove e unless it is the only element that can be removed. As in Lemma 3.11, every step can be copied in M/e, this time by Lemma 3.9, as long as we are not removing e, and have not yet removed e. Also by Lemma 3.9, if we never remove e, then when there are no more elements to remove in M, there are no more elements to remove in M/e. Thus, if e is never removed (*i.e.*, if $e \in \pi_M(B)$), then $\pi_{M/e}(B - e) = \pi_M(B) - e$.
If e is eventually removed in M, it must be when e is an isthmus, since e is ordered last (so it can be the minimal element of $\mathrm{bo}(e, I)$ only if it is the only element – *i.e.*, if it is an isthmus). Since we put off removing e until there were no other possible removals, Lemma 3.10 guarantees that there are no new removals possible after e is removed. Since the removals were identical in M and M/e until e was removed, $\pi_{M/e}(B - e) = \pi_M(B)$. $\square$

Corollary 3.14. *Let I be an independent set of M such that $e \in I$, so $I - e$ is independent in M/e. Also assume e is last in the linear order. Then*

$$\overline{\pi}_{M/e}(I - e) = \overline{\pi}_M(I) \backslash e.$$

Proof. Let $V = \mathrm{cl}_M(I)$. Then $\mathrm{cl}_{M/e}(I-e) = V-e$ as *sets*, so $\mathrm{cl}_{M/e}(I-e) = V/e$ as *matroids*. Thus, by the definition of $\overline{\pi}$, we have

$$\overline{\pi}_{M/e}(I - e) = \mathrm{cl}_{M/e}(\pi_{V/e}(I - e)).$$

If $e \in \pi_V(I)$, then simply $\mathrm{cl}_{M/e}(\pi_{V/e}(I-e)) = \mathrm{cl}_{M/e}(\pi_V(I)-e) = \mathrm{cl}_M(\pi_V(I))- e = \overline{\pi}_M(I) \backslash e$; the first equality is by Lemma 3.13, the second equality is a routine exercise using $e \in \pi_V(I)$, and the last equality is from the definition of $\overline{\pi}$.

If $e \notin \pi_V(I)$, then the proof of Lemma 3.13 shows that e is an isthmus in $\mathrm{cl}_V(\pi_V(I) \cup e)$. Then, since $\mathrm{cl}(A \dot\cup i) = (\mathrm{cl}\, A) \dot\cup i$ for any A and any isthmus $i \notin A$,

$$\mathrm{cl}_M(\pi_V(I) \cup e) = \mathrm{cl}_V(\pi_V(I) \cup e) = \mathrm{cl}_V(\pi_V(I)) \cup e = \overline{\pi}_M(I) \cup e. \tag{6}$$

Now, also in this case,

$$\begin{aligned} \mathrm{cl}_{M/e}(\pi_{V/e}(I - e)) = \mathrm{cl}_{M/e}(\pi_V(I)) &= \mathrm{cl}_M(\pi_V(I) \cup e) - e = (\overline{\pi}_M(I) \cup e) - e \\ &= \overline{\pi}_M(I) \backslash e; \end{aligned}$$

the first equality is by Lemma 3.13, the second equality is from the definition of $\mathrm{cl}_{M/e}$, and the third equality is equation (6). □

Lemma 3.15. *Let B be a base of matroid $M(E)$, let e be first in the linear order on E, and assume that $e \notin B$ and e is not a loop. Let $C = \mathrm{ci}(e, B)$, so $B - (C - e)$ is a base of M/C. Then*

$$\pi_{M/C}(B - (C - e)) = \pi_M(B) - (C - e).$$

Proof. It is an easy exercise to check that $\mathrm{bo}_{M/C}(b, B - (C - e)) = \mathrm{bo}_M(b, B)$ for any $b \in B - (C - e)$. It then follows that b is internally active in B (with respect to M) iff b is minimal in $\mathrm{bo}_M(b, B) = \mathrm{bo}_M(b, B - (C - e))$ iff b is internally active in $B - (C - e)$ (with respect to M/C).

Now, as in Lemmas 3.11 and 3.13, use Algorithm 3.5 to compute $\pi_{M/C}(B - (C - e))$. Once again, every step can be copied in M, computing $\pi_M(B)$. Furthermore, when there are no more elements in $B - (C - e)$ to remove in computing $\pi_{M/C}(B - (C - e))$, the only elements of B that could possibly be removed in computing $\pi_M(B)$ must be in $C - e$. We now show that any $c \in C - e$ is not internally active, and thus that the removals in M and M/C are identical, which will complete the proof.

It is easy to see that $C = \mathrm{ci}_{\overline{B_1}}(e, B_1)$, where B_1 is what remains of B after performing all the removals in M corresponding to the removals in M/C. Thus $c \in C - e \subseteq \mathrm{ci}_{\overline{B_1}}(e, B_1)$ implies, by *e.g.*, [5, Lemma 7.3.1], that $e \in \mathrm{bo}_{\overline{B_1}}(c, B_1)$. Since e is first in the linear order, c is, as desired, not internally active. □

3.3. The spectral recursion for matroids. We now prove that matroids satisfy the spectral recursion (Theorem 3.18), by comparing $qtS_{M/e} + qS_{M-e} - S_M$ and $S_{(M-e,M/e)}$. In each case, we get two expressions, one in terms of $\tilde\chi$ and μ, the other in terms of $\overline{\pi}$. The expressions in terms of $\tilde\chi$ and μ lead to a quick proof, by reducing a key piece of the equation to the $q = 0$ case for a flat. The expressions in terms of $\overline{\pi}$ suggest a more bijective proof, which is not hard to prove either. Both proofs are given.

Lemma 3.16. *If $M(E)$ is a matroid, and $e \in E$ is neither an isthmus nor a loop, then*

$$\begin{aligned} qS_{M-e}(t,q) &+ qtS_{M/e}(t,q) - S_M(t,q) \\ &= (q-1)q^{|E|} \sum_{V \in \mathcal{F}(M)} t^{\mathrm{rk}_M(V)} \sum_{\substack{I \in \mathcal{B}(V) \\ e \in \overline{\pi}_V(I)}} (q^{-1})^{|\overline{\pi}_V(I)|} \\ &= (q-1) \sum_{V \in \mathcal{F}(M)} t^{\mathrm{rk}_M(V)} \sum_{\substack{W \in \mathcal{F}(V) \\ e \in W}} |\tilde{\chi}(W)||\mu(W,V)| q^{|E|-|W|}. \end{aligned}$$

Proof. We compute each of S_{M-e} and $S_{M/e}$ using Lemma 3.6. First,

$$\begin{aligned} S_{M-e}(t,q) &= q^{|E-e|} \sum_{I \in \mathcal{I}(M-e)} t^{\mathrm{rk}_{M-e}(\mathrm{cl}_{M-e}(I))} (q^{-1})^{|\overline{\pi}_{M-e}(I)|} \\ &= q^{|E-e|} \sum_{\substack{I \in \mathcal{I}(M) \\ e \notin I}} t^{\mathrm{rk}_M(\mathrm{cl}_M(I))} (q^{-1})^{|\overline{\pi}_M(I) \backslash e|}, \end{aligned} \tag{7}$$

since: $I \in \mathcal{I}(M-e)$ iff $I \in \mathcal{I}(M)$ and $e \notin I$; $|\overline{\pi}_{M-e}(I)| = |\overline{\pi}_M(I)\backslash e|$, by Corollary 3.12; and $\mathrm{rk}_{M-e}(\mathrm{cl}_{M-e}(I)) = \mathrm{rk}_M(\mathrm{cl}_M(I)\backslash e)$ is an easy matroid exercise. Similarly,

$$\begin{aligned} S_{M/e}(t,q) &= q^{|E-e|} \sum_{I' \in \mathcal{I}(M/e)} t^{\mathrm{rk}_{M/e}(\mathrm{cl}_{M/e}(I'))} (q^{-1})^{|\overline{\pi}_{M/e}(I')|} \\ &= q^{|E-e|} \sum_{\substack{I \in \mathcal{I}(M) \\ e \in I}} t^{\mathrm{rk}_M(\mathrm{cl}_M(I))-1} (q^{-1})^{|\overline{\pi}_M(I) \backslash e|}, \end{aligned} \tag{8}$$

where $I = I' \dot\cup e$ for $I' \in \mathcal{I}(M/e)$, since: $|\overline{\pi}_{M/e}(I')| = |\overline{\pi}_{M/e}(I-e)| = |\overline{\pi}_M(I)\backslash e|$, by Corollary 3.14; and $\mathrm{rk}_{M/e}(\mathrm{cl}_{M/e}(I')) = \mathrm{rk}_M(\mathrm{cl}_M(I))-1$ is a routine exercise, using $e \in \mathrm{cl}_M(I)$.
Combining equations (7) and (8), and then sorting independent sets by their closures, we get

$$\begin{aligned} qS_{M-e}(t,q) + qtS_{M/e}(t,q) &= q^{|E|} \sum_{I \in \mathcal{I}(M)} t^{\mathrm{rk}_M(\mathrm{cl}_M(I))} (q^{-1})^{|\overline{\pi}_M(I) \backslash e|} \\ &= q^{|E|} \sum_{V \in \mathcal{F}(M)} t^{\mathrm{rk}_M(V)} \sum_{I \in \mathcal{B}(V)} (q^{-1})^{|\overline{\pi}_V(I) \backslash e|}. \end{aligned} \tag{9}$$

Furthermore

$$\begin{aligned} \sum_{I \in \mathcal{B}(V)} (q^{-1})^{|\overline{\pi}_V(I) \backslash e|} &= \sum_{\substack{I \in \mathcal{B}(V) \\ e \notin \overline{\pi}_V(I)}} (q^{-1})^{|\overline{\pi}_V(I) \backslash e|} + \sum_{\substack{I \in \mathcal{B}(V) \\ e \in \overline{\pi}_V(I)}} (q^{-1})^{|\overline{\pi}_V(I) \backslash e|} \\ &= \sum_{I \in \mathcal{B}(V)} (q^{-1})^{|\overline{\pi}_V(I)|} + (q-1) \sum_{\substack{I \in \mathcal{B}(V) \\ e \in \overline{\pi}_V(I)}} (q^{-1})^{|\overline{\pi}_V(I)|}; \end{aligned} \tag{10}$$

plugging into equation (10) into equation (9) readily leads to the first equation of the lemma. The second equation then follows directly from Lemma 3.7. $\square$

Lemma 3.17. *If $M(E)$ is a matroid, and $e \in E$ is neither an isthmus nor a loop, then*

$$S_{(M-e,M/e)}(t,q) = q^{|E|} \sum_{V \in \mathcal{F}(M)} t^{\mathrm{rk}_M(V)} \sum_{\substack{C \in \mathcal{C}(V) \\ e \in C}} q^{-|C|} \sum_{I \in \mathcal{B}(V/C)} (q^{-1})^{|\overline{\pi}_{V/C}(I)|}$$
$$= \sum_{V \in \mathcal{F}(M)} t^{\mathrm{rk}_M(V)} \sum_{\substack{W \in \mathcal{F}(V) \\ e \in W}} \sum_{\substack{C \in \mathcal{C}(W) \\ e \in W}} |\tilde{\chi}(W/C)||\mu(W,V)| q^{|E|-|W|}.$$

Proof. By Lemmas 3.6, 3.1, and 3.3,

$$S_{(M-e,M/e)}(t,q) = \sum_{\substack{C \in \mathcal{C}(M) \\ e \in C}} t^{\mathrm{rk}_M(C)} S_{M/C}(t,q)$$
$$= \sum_{\substack{C \in \mathcal{C}(M) \\ e \in C}} t^{\mathrm{rk}_M(C)} q^{|E-C|} \sum_{W \in \mathcal{F}(M/C)} t^{\mathrm{rk}_{M/C}(W)} \sum_{I \in \mathcal{B}(W)} (q^{-1})^{|\overline{\pi}_W(I)|}.$$

Now, the flats of M/C are $V - C$ as *sets*, and thus V/C as *matroids*, for all flats V of M containing C. Therefore,

$$S_{(M-e,M/e)}(t,q)$$
$$= q^{|E|} \sum_{\substack{C \in \mathcal{C}(M) \\ e \in C}} t^{\mathrm{rk}_M(C)} q^{-|C|} \sum_{\substack{V \in \mathcal{F}(M) \\ C \subseteq V}} t^{\mathrm{rk}_M(V)-\mathrm{rk}_M(C)} \sum_{I \in \mathcal{B}(V/C)} (q^{-1})^{|\overline{\pi}_{V/C}(I)|}$$
$$= q^{|E|} \sum_{V \in \mathcal{F}(M)} t^{\mathrm{rk}_M(V)} \sum_{\substack{C \in \mathcal{C}(M) \\ e \in C \subseteq V}} q^{-|C|} \sum_{I \in \mathcal{B}(V/C)} (q^{-1})^{|\overline{\pi}_{V/C}(I)|},$$

which is the first equation of the lemma, once we note that $C \in \mathcal{C}(V)$ iff $C \in \mathcal{C}(M)$ and $C \subseteq V$.

The second equation of the lemma then follows from

$$\sum_{\substack{C \in \mathcal{C}(V) \\ e \in C}} q^{-|C|} \sum_{I \in \mathcal{B}(V/C)} (q^{-1})^{|\overline{\pi}_{V/C}(I)|}$$
$$= \sum_{\substack{C \in \mathcal{C}(V) \\ e \in C}} q^{-|C|} \sum_{W/C \in \mathcal{F}(V/C)} |\tilde{\chi}(W/C)||\mu_{V/C}(W/C, V/C)|(q^{-1})^{|W/C|}$$
$$= \sum_{\substack{C \in \mathcal{C}(V) \\ e \in C}} \sum_{\substack{W \in \mathcal{F}(V) \\ C \subseteq W}} |\tilde{\chi}(W/C)||\mu(W,V)|(q^{-1})^{|W|}$$
$$= \sum_{\substack{W \in \mathcal{F}(V) \\ e \in W}} \sum_{\substack{C \in \mathcal{C}(V) \\ e \in C \subseteq W}} |\tilde{\chi}(W/C)||\mu(W,V)|(q^{-1})^{|W|}.$$

The first equation above is from equation (5); we are also using the same characterization of flats of a contraction as in the previous paragraph. The second equation is since the interval $[W/C, V/C]$ in the lattice of flats of V/C is isomorphic to the interval $[W, V]$ in the lattice of flats of V, again by that same characterization of flats in a contraction. It only remains to again note that $C \in \mathcal{C}(W)$ iff $C \in \mathcal{C}(M)$ and $C \subseteq W$. $\square$

Theorem 3.18. *If M is a matroid, then its independence complex $IN(M)$ satisfies the spectral recursion, equation* (2).

Proof. By Proposition 2.3, we may assume e is not a loop. By Lemma 2.2 and Theorem 4.4 below (which does not depend on anything in this section), we may assume e is not an isthmus. As discussed at the beginning of the subsection, there are now two ways to finish off the proof, one using the $q = 0$ case, the other using a bijection.

$q = 0$ *proof.* By Theorem 2.4, we know that the spectral recursion holds, for any matroid, with $q = 0$. By Lemmas 3.16 and 3.17, this means

$$|\tilde{\chi}(M)| = \sum_{\substack{C \in \mathcal{C}(M) \\ e \in C}} |\tilde{\chi}(M/C)| \tag{11}$$

for any matroid M, since only terms with $W = E$ survive when $q = 0$. (Equation (11) is also, as noted by Kook [25], dual to Crapo's complementation theorem (*e.g.*, [1, Theorem 4.33]) applied to the dual matroid of M.) Thus, simply by plugging in the flat W, as a matroid, for the matroid M in equation (11),

$$|\tilde{\chi}(W)| = \sum_{\substack{C \in \mathcal{C}(W) \\ e \in C}} |\tilde{\chi}(W/C)|$$

whenever W is a flat of M containing e. By Lemmas 3.16 and 3.17 again, we are done.

Bijective proof. By Lemmas 3.16 and 3.17, it suffices to show

$$\sum_{\substack{I \in \mathcal{B}(M) \\ e \in \overline{\pi}_W(I)}} (q^{-1})^{|\overline{\pi}_M(I)|} = \sum_{\substack{C \in \mathcal{C}(M) \\ e \in C}} \sum_{I \in \mathcal{B}(M/C)} (q^{-1})^{|\overline{\pi}_{M/C}(I)| + |C|}. \tag{12}$$

Further, Lemma 3.7 shows that the sum on the left-hand side of equation (12) is independent of the ordering of the ground set. Similarly, Lemma 3.17 itself shows the same thing for the sum on the right-hand side. So we now assume, for the remainder of this proof, that e is ordered first in the linear order on E. Equation (12) would follow naturally from a bijection

$$\phi\colon \{B \in \mathcal{B}(M) \colon e \in \overline{\pi}_M(B)\} \to \{(C, I) \colon C \in \mathcal{C}(M), I \in \mathcal{B}(M/C), e \in C\}$$

such that

$$\overline{\pi}_{M/C}(I) \,\dot{\cup}\, C = \overline{\pi}_M(B), \tag{13}$$

where $\phi(B) = (C, I)$. Such a bijection is given by, as we now show, $C = \mathrm{ci}(e, B)$ and $I = B - (C - e)$ in one direction, and $B = I \,\dot{\cup}\, C - e$ in the other.

First note that, since e is ordered first, if $e \in B$ then e is internally active in B, and so $e \notin \pi_M(B)$. It is then easy to see in this case that $e \notin \overline{\pi}_M(B)$. We may therefore safely assume $e \notin B$, and so $C = \mathrm{ci}(e, B)$ is well-defined. It then follows that ϕ is well-defined.

It is easy to see that ϕ is injective. Showing that ϕ is surjective reduces to verifying that $e \in \overline{\pi}_M(B)$ when $B = I \dot{\cup} C - e$; by Lemma 3.15, $C - e \subseteq \pi_M(B)$, so $e \in C = \overline{C - e} \subseteq \overline{\pi}_M(B)$.

Finally, to verify equation (13), by Lemma 3.15 and the definition of closure in a matroid contraction, $\overline{\pi}_{M/C}(B - (C - e)) = \mathrm{cl}_M(\pi_M(B) \cup e) - C$. Also, $\mathrm{cl}_M(\pi_M(B) \cup e) = \mathrm{cl}_M(\pi_M(B)) = \overline{\pi}_M(B)$, since $e \in \overline{\pi}_M(B)$, which completes the proof of equation (13). $\square$

Remark 3.19. The spectral recursion does not provide a truly recursive way to compute S_M, due to the presence of $S_{(M-e,M/e)}$, since the recursion only applies to a single matroid, and not a matroid pair like $(M - e, M/e)$. We can however, combine it with Lemmas 3.1 and 3.3 for a recursion that is truly recursive, albeit with more terms than the spectral recursion:

$$S_M(t,q) = qS_{M-e}(t,q) + qtS_{M/e}(t,q) + (1-q) \sum_{\substack{C \in \mathcal{C}(M) \\ e \in C}} t^{\mathrm{rk}_M(C)} S_{M/C}(t,q).$$

I am grateful to E. Babson for this observation.

4. Shifted complexes

We postpone until Subsection 4.5 the actual definition of shifted complexes, but we will see there that a shifted complex is a skeleton of a cone of a smaller shifted complex (Lemmas 4.21 4.22. To prove that shifted complexes satisfy the spectral recursion, equation (2), then, it suffices to show that taking skeleta and taking cones each preserve the property of satisfying the spectral recursion – which are interesting results in their own right.

We will prove in Subsection 4.1 that the property of satisfying the spectral recursion is preserved by taking joins (Corollary 4.5), and thus by taking cones (*cf.* Proposition 2.2). The key step is that a simple formula [16, Theorem 4.10] for the eigenvalues of the join generalizes straightforwardly from single simplicial complexes to simplicial pairs (Corollaries 4.2 and 4.3).

Proving that taking skeleta preserves the property of satisfying the spectral recursion is harder, and is the focus of Subsections 4.2–4.4. The key facts about Laplacians, established in Subsections 4.2 and 4.3, respectively, are that the non-zero eigenvalues come in pairs in consecutive dimensions (Lemma 4.7), and that taking $(d-1)$-skeleta preserves non-zero eigenvalues of the finer Laplacians in dimension $d-1$ and below (Lemma 4.11).

The only eigenvalues in dimension $d-1$ and below that are changed by taking $(d-1)$-skeleta, then, are some $(d-1)$-dimensional eigenvalues that become 0 when their counterparts (in the sense of Lemma 4.7) in dimension d are removed. It is auspicious that these replaced $(d-1)$-dimensional eigenvalues must line up properly in the spectral recursion (since their counterparts in dimension d, the only non-zero eigenvalues in that dimension, do as well) and that the 0's that replace them also line up properly (since the spectral recursion is true with $q = 0$ for both the original complex and its skeleton, by Theorem 2.4). But it turns out that we are better off with f-vectors ($q = 1$, also a good case by Theorem 2.4) than with homology ($q = 0$), in part because the change in f-vectors resulting from taking skeleta is much easier to describe than the change in homology.

In Subsection 4.4, we will see that the difference between the spectrum polynomials of the skeleton and the original complex can be described largely in terms of the f-vector (Lemma 4.14), allowing us to describe the difference in the spectral recursion between the skeleton and the original complex in a particularly useful form (Lemma 4.15). From there, simple generating function manipulations lead to Theorem 4.18, which states that a d-dimensional simplicial complex satisfies the spectral recursion with respect to a vertex if and only if its $(d-1)$-skeleton and pure d-skeleton (the complex generated by its facets) do as well.

4.1. Joins and cones. Define the *join* $(\Delta, \Delta') * (\Gamma, \Gamma')$ of two simplicial pairs on disjoint vertex sets to be

$$(\Delta, \Delta') * (\Gamma, \Gamma') := \{F \mathbin{\dot\cup} G \colon F \in \Delta\backslash\Delta', G \in \Gamma\backslash\Gamma'\}$$

(here, $\dot\cup$ denotes disjoint union), which equals the simplicial pair

$$(\Delta * \Gamma, (\Delta' * \Gamma) \cup (\Delta * \Gamma')). \tag{14}$$

When $\Delta' = \Gamma' = \emptyset$, this reduces to the usual join $\Delta * \Gamma$. When, further, Δ is a single vertex, say v, the join is written as $v * \Gamma$, the *cone* over Γ with *apex* v.

The proofs of the following two results on simplicial pairs are identical (modulo some indexing changes) to those of the analogous statements for single simplicial complexes [16, Section 4].

Proposition 4.1. *For any two simplicial pairs (Δ, Δ') and (Γ, Γ') and every k, the map defined $\mathbb{R}$-linearly by $[F] \otimes [G] \mapsto [F \mathbin{\dot\cup} G]$ identifies the vector spaces*

$$\bigoplus_{i+j=k} C_{i-1}((\Delta, \Delta'); \mathbb{R}) \otimes C_{j-1}((\Gamma, \Gamma'); \mathbb{R}) \cong C_{k-1}((\Delta, \Delta') * (\Gamma, \Gamma'); \mathbb{R})$$

and has the following property with respect to the Laplacians L of the appropriate dimensions in $(\Delta, \Delta'), (\Gamma, \Gamma')$, and $(\Delta, \Delta') * (\Gamma, \Gamma')$:

$$L((\Delta, \Delta') * (\Gamma, \Gamma')) = L(\Delta, \Delta') \otimes \mathrm{id} + \mathrm{id} \otimes L(\Gamma, \Gamma'). \tag{15}$$

Corollary 4.2. *If (Δ, Δ') and (Γ, Γ') are two simplicial pairs, then*

$$\mathbf{s}_{k-1}((\Delta, \Delta') * (\Gamma, \Gamma')) = \bigcup_{\substack{i+j=k \\ \lambda \in \mathbf{s}_{i-1}(\Delta,\Delta'),\ \mu \in \mathbf{s}_{j-1}(\Gamma,\Gamma')}} \lambda + \mu.$$

It is then an easy exercise in generating functions to verify the following corollary.

Corollary 4.3. *If (Δ,Δ') and (Γ,Γ') are two simplicial pairs, then*

$$S_{(\Delta,\Delta')*(\Gamma,\Gamma')} = S_{(\Delta,\Delta')}S_{(\Gamma,\Gamma')}.$$

Theorem 4.4. *If Δ satisfies the spectral recursion with respect to e, and Γ is any simplicial complex whose vertex set is disjoint from the vertex set of Δ, then the join $\Delta * \Gamma$ satisfies the spectral recursion with respect to e.*

Proof. By Corollary 4.3 twice, and our hypothesis,

$$\begin{aligned} S_{\Delta*\Gamma} = S_\Delta S_\Gamma &= (qS_{\Delta-e} + qtS_{\Delta/e} + (1-q)S_{(\Delta-e,\Delta/e)})S_\Gamma \\ &= qS_{(\Delta-e)*\Gamma} + qtS_{(\Delta/e)*\Gamma} + (1-q)S_{(\Delta-e,\Delta/e)*\Gamma}. \end{aligned}$$

This last expression is exactly what we need, since it is easy to verify that join commutes with deletion and contraction, *i.e.*, $(\Delta - e) * \Gamma = (\Delta * \Gamma) - e$ and $(\Delta/e) * \Gamma/e = (\Delta * \Gamma)/e$, and also since equation (14) with $\Delta' = \emptyset$ then yields

$$(\Delta - e, \Delta/e) * \Gamma = ((\Delta - e) * \Gamma, (\Delta/e) * \Gamma) = ((\Delta * \Gamma) - e, (\Delta * \Gamma)/e).$$

□

Corollary 4.5. *If Δ and Γ each satisfy the spectral recursion, then so does their join $\Delta * \Gamma$.*

4.2. Finer Laplacians. Recall from Section 2 that $L'_i = L'_i(\Delta,\Delta') := \partial_{i+1}\partial^*_{i+1}$ and $L''_i = L''_i(\Delta,\Delta') := \partial^*_i\partial_i$, so that $L_i = L'_i + L''_i$. Define $\mathbf{s}'_i(\Delta,\Delta')$ and $\mathbf{s}''_i(\Delta,\Delta')$ to be the multiset of eigenvalues of $L'_i(\Delta,\Delta')$ and $L''_i(\Delta,\Delta')$, respectively, arranged in weakly decreasing order.
Following [16], let the equivalence relation $\boldsymbol{\lambda} \mathrel{\mathring{=}} \boldsymbol{\mu}$ on multisets $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ denote that $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ agree in the multiplicities of all of their *non-zero parts*, *i.e.*, that they coincide except for possibly their number of zeroes. Also let $\boldsymbol{\lambda} \cup \boldsymbol{\mu}$ denote the $\mathring{=}$-equivalence class whose non-zero parts are the multiset union of the non-zero parts of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$.

Proposition 4.6. *If (Δ,Δ') is a simplicial pair, then*

$$\mathbf{s}_i(\Delta,\Delta') \mathrel{\mathring{=}} \mathbf{s}''_i(\Delta,\Delta') \cup \mathbf{s}''_{i+1}(\Delta,\Delta').$$

Proof. The proof is identical to the single simplicial complex ($\Delta' = \emptyset$) case in [16, Equation (3.6)], and depends only upon $\partial^2 = 0$ and routine eigenvalue calculations involving adjoints. □

If (Δ,Δ') is a simplicial pair, let

$$S''_{(\Delta,\Delta'),i}(q) := \sum_{\substack{\lambda \in \mathbf{s}''_i(\Delta,\Delta') \\ \lambda \neq 0}} q^\lambda, \quad \text{and}$$

$$S''_{(\Delta,\Delta')}(t,q) := \sum_i S''_{(\Delta,\Delta'),i-1}(q)t^i.$$

Zero eigenvalues are omitted from these definitions of S'' in order to more naturally encode Proposition 4.6 into the language of generating functions, in Lemma 4.7, below. Also let

$$B_{(\Delta,\Delta')}(t) := \sum_i \tilde{\beta}_{i-1}(\Delta,\Delta')t^i = \sum_i m_0(L_{i-1}(\Delta,\Delta'))t^i = S_{(\Delta,\Delta')}(t,0).$$

These three definitions of B are equivalent by Proposition 2.1.
From now on, when there is no confusion about the variables t and q, we will often omit them for clarity.

Lemma 4.7. *If (Δ,Δ') is a simplicial pair, then*

$$S_{(\Delta,\Delta')} = (1+t^{-1})S''_{(\Delta,\Delta')} + B_{(\Delta,\Delta')}.$$

Proof. Combine Propositions 4.6 and 2.1. □

Corollary 4.8. *If Δ is any simplicial complex, and e is any vertex of Δ, then the spectral recursion holds when $t = -1$.*

Proof. By Lemma 4.7, for any simplicial pair (Δ,Δ'),

$$S_{(\Delta,\Delta')}(-1,q) = B_{(\Delta,\Delta')}(-1,q) = \sum_i (-1)^i \tilde{\beta}_i(\Delta,\Delta') = \chi(\Delta,\Delta'),$$

where $\chi(\Delta,\Delta')$ denotes the Euler characteristic of the simplicial pair (Δ,Δ') (see *e.g.*, [30]). The identity $\chi(\Delta,\Delta') = \chi(\Delta) - \chi(\Delta')$, which holds as long as $\Delta' \subseteq \Delta$, immediately reduces the $t = -1$ instance of the spectral recursion to $\chi(\Delta) = \chi(\Delta - e) - \chi(\Delta/e)$. This, in turn, follows from $\chi(\Delta) = \sum_i (-1)^i f_i(\Delta)$ and equation (4). □

If Δ is a simplicial complex, define

$$F_\Delta(t) := \sum_i f_{i-1}(\Delta)t^i.$$

If $\phi(q)$ is a function of q, define

$$D_q\phi := \phi(q) - \phi(1).$$

The point of D_q is that it helps us convert from B and homology (the effect on which of taking skeleta is hard to describe) to F and f-vectors (the effect on which of taking skeleta is easy to describe) in the following lemma.

Lemma 4.9. *If $\Delta \subseteq \Delta'$ are simplicial complexes, then*

$$S_{(\Delta,\Delta')} = (1+t^{-1})D_q S''_{(\Delta,\Delta')} + F_\Delta - F_{\Delta'}.$$

Proof. By Lemma 4.7,

$$F_\Delta(t) - F_{\Delta'}(t) = S_{(\Delta,\Delta')}(t,1) = (1+t^{-1})S''_{(\Delta,\Delta')}(t,1) + B_{(\Delta,\Delta')}(t).$$

Thus

$$B_{(\Delta,\Delta')}(t) = -(1+t^{-1})S''_{(\Delta,\Delta')}(t,1) + F_\Delta(t) - F_{\Delta'}(t),$$

which, when plugged back into Lemma 4.7, yields the desired result. □

4.3. Skeleta. Recall the *s-skeleton* of a simplicial complex Δ is

$$\Delta^{(s)} := \{F \in \Delta\colon\ \dim F \leq s\}.$$

Also recall that a simplicial complex is *pure* if all its facets have the same dimension. The *pure s-skeleton* of a simplicial complex Δ is

$$\Delta^{[s]} := \{F \in \Delta\colon F \subseteq G, G \in \Delta, \dim G = s\}.$$

In other words, $\Delta^{[s]}$ is the subcomplex of Δ consisting of the s-dimensional faces of Δ, and all their subfaces. (See [8, Definition 2.8].) The results of the following lemma are easy exercises.

Lemma 4.10. *If Δ is a simplicial complex and e is a vertex of Δ, then*

1. $(\Delta - e)^{(s)} = \Delta^{(s)} - e$;
2. $(\Delta/e)^{(s-1)} = \Delta^{(s)}/e$;
3. $\Delta_s = (\Delta^{[s]})_s$;
4. $(\Delta - e)_s = (\Delta^{[s]} - e)_s$; *and*
5. $(\Delta/e)_{s-1} = (\Delta^{[s]}/e)_{s-1}$.

Lemma 4.11. *If $\dim \Delta' \leq d-1$, then*

$$\mathbf{s}''_{d-1}(\Delta, \Delta') \mathrel{\mathring{=}} \mathbf{s}''_{d-1}(\Delta^{(d-1)}, {\Delta'}^{(d-2)}).$$

Proof. Since Δ and $\Delta^{(d-1)}$ agree in dimensions $d-1$ and below,

$$\mathbf{s}''_{d-1}(\Delta, \Delta') = \mathbf{s}''_{d-1}(\Delta^{(d-1)}, \Delta').$$

Next, replacing Δ' by ${\Delta'}^{(d-2)}$ in $(\Delta^{(d-1)}, \Delta')$ has the effect of adding $(d-1)$-dimensional faces (in fact, all the $(d-1)$-dimensional faces of Δ') to the simplicial pair, all of whose boundary faces are still not present in the simplicial pair, since $\dim \Delta' \leq d-1$. Thus

$$\partial_{(\Delta^{(d-1)}, {\Delta'}^{(d-2)});d-1} = \partial_{(\Delta^{(d-1)}, \Delta');d-1} \oplus 0$$

(equivalently, the matrices representing the two boundary operators differ only in some additional zero columns); *cf.* proof of Lemma 5.1. It is then easy to check that, since $L''_{d-1} = \partial^*_{d-1}\partial_{d-1}$,

$$L''_{d-1}(\Delta^{(d-1)}, {\Delta'}^{(d-2)}) = L''_{d-1}(\Delta^{(d-1)}, \Delta') \oplus 0,$$

and so

$$s''_{d-1}(\Delta^{(d-1)}, {\Delta'}^{(d-2)}) \mathrel{\mathring{=}} s''_{d-1}(\Delta^{(d-1)}, \Delta').$$

□

Corollary 4.12. *If $\dim \Delta' \leq d-1$, then*

$$S''_{(\Delta,\Delta'),d-1} = S''_{(\Delta^{(d-1)}, {\Delta'}^{(d-2)}),d-1}$$

Corollary 4.13. *If $\dim \Delta \leq d$ and $\dim \Delta' \leq d-1$, then*

$$S''_{(\Delta^{(d-1)}, {\Delta'}^{(d-2)})} = S''_{(\Delta,\Delta')} - S''_{(\Delta,\Delta'),d} t^{d+1}.$$

Proof. Clearly, (Δ, Δ') and $(\Delta^{(d-1)}, \Delta'^{(d-2)})$ agree in dimensions $d-2$ and below. Corollary 4.12 thus ensures $S''_{(\Delta^{(d-1)}, \Delta'^{(d-2)})} = \sum_{i \le d} S''_{(\Delta,\Delta'),i-1} t^i$. Then simply note, since $\dim \Delta \le d$, that $S''_{(\Delta,\Delta'),d} t^{d+1}$ is the only remaining term from $S''_{(\Delta,\Delta')}$ not found in $S''_{(\Delta^{(d-1)}, \Delta'^{(d-2)})}$. □

4.4. The spectral recursion and skeleta.

Lemma 4.14. *If* $\dim \Delta \le d$, $\dim \Delta' \le d-1$, *and* $\Delta' \subseteq \Delta$, *then*

$$\begin{aligned}&S_{(\Delta^{(d-1)}, \Delta'^{(d-2)})}\\ &= S_{(\Delta,\Delta')} - (f_d(\Delta) + f_{d-1}(\Delta))t^{d+1} - (D_q S''_{(\Delta,\Delta'),d} - f_{d-1}(\Delta'))(t^{d+1} + t^d).\end{aligned}$$

Proof. First use the definition of D_q and Corollary 4.13 to get

$$D_q S''_{(\Delta^{(d-1)}, \Delta'^{(d-2)})} = D_q S''_{(\Delta,\Delta')} - (D_q S''_{(\Delta,\Delta'),d}) t^{d+1}. \tag{16}$$

Then apply Lemma 4.9 (twice) and equation (16) to compute

(17)

$$\begin{aligned}&S_{(\Delta^{(d-1)}, \Delta'^{(d-2)})}\\ &\qquad = S_{(\Delta,\Delta')} - (F_\Delta - F_{\Delta'}) - (t^d + t^{d+1}) D_q S''_{(\Delta,\Delta'),d}\\ &\qquad\qquad + (F_{\Delta^{(d-1)}} - F_{\Delta'^{(d-2)}})\end{aligned}$$

$$= S_{(\Delta,\Delta')} - (f_d(\Delta) t^{d+1} - f_{d-1}(\Delta') t^d) - (t^d + t^{d+1}) D_q S''_{(\Delta,\Delta'),d}. \tag{18}$$

The lemma now follows by adding the quantity $(t^d + t^{d+1}) f_{d-1}(\Delta')$ to the middle term of the right hand side of equation (18), while subtracting it from the last term. □

If Δ is a simplicial complex and e is a vertex of Δ, let

$$\begin{aligned}\mathcal{S}_{\Delta,e} &:= S_\Delta - (qS_{\Delta-e} + qtS_{\Delta/e} + (1-q)S_{(\Delta-e,\Delta/e)}),\\ \mathcal{S}^d_{\Delta,e} &:= S''_{\Delta,d} - (qS''_{\Delta-e,d} + qS''_{\Delta/e,d-1} + (1-q)S''_{(\Delta-e,\Delta/e),d}), \text{ and}\\ \mathcal{D}^d_{\Delta,e} &:= D_q \mathcal{S}^d_{\Delta,e} + (1-q) f_{d-1}(\Delta/e).\end{aligned}$$

We have defined $\mathcal{S}_{\Delta,e}$ precisely so that Δ satisfies the spectral recursion with respect to e if and only if $\mathcal{S}_{\Delta,e} = 0$, and we have defined $\mathcal{S}^d_{\Delta,e}$ to be the d-dimensional finer Laplacian version of $\mathcal{S}_{\Delta,e}$. The significance of $\mathcal{D}$ is made apparent by the next lemma, which is the last key step to proving Theorem 4.18.

Lemma 4.15. *If* $\dim \Delta \le d$ *and* e *is a vertex of* Δ, *then*

$$\mathcal{S}_{\Delta^{(d-1)},e} = \mathcal{S}_{\Delta,e} - (t^d + t^{d+1}) \mathcal{D}^d_{\Delta,e}.$$

Proof. Since $\dim\Delta \le d$, then $\dim\Delta - e \le d$ and $\dim\Delta/e \le d-1$. Therefore

$$\begin{aligned}
&\mathcal{S}_{\Delta^{(d-1)},e}\\
&= S_{\Delta^{(d-1)}} - qS_{(\Delta-e)^{(d-1)}} - qtS_{(\Delta/e)^{(d-2)}} - (1-q)S_{((\Delta-e)^{(d-1)},(\Delta/e)^{(d-2)})}\\
&= S_{\Delta} - qS_{\Delta-e} - qtS_{\Delta/e} - (1-q)S_{(\Delta-e,\Delta/e)}\\
&\quad - f_d(\Delta)t^{d+1} + qf_d(\Delta-e)t^{d+1} + qtf_{d-1}(\Delta/e)t^d\\
&\qquad + (1-q)(f_d(\Delta-e) + f_{d-1}(\Delta/e))t^{d+1}\\
&\quad - D_qS''_{\Delta,d}(t^{d+1}+t^d) + qD_qS''_{\Delta-e,d}(t^{d+1}+t^d) + qtD_qS''_{\Delta/e,d-1}(t^d+t^{d-1})\\
&\qquad + (1-q)(D_qS''_{(\Delta-e,\Delta/e),d} - f_{d-1}(\Delta/e))(t^{d+1}+t^d).
\end{aligned}$$

The first equation above is by the definition of $\mathcal{S}$ and Lemma 4.10. The second equation involves expanding each term of the left-hand side by Lemma 4.14, and then regrouping like terms. Now, the second line and third lines of this last expression add up to zero, by equation (4). The lemma then follows from the definitions of $\mathcal{S}$ and $\mathcal{S}^d$. □

Lemma 4.16. *If $\dim\Delta \le d$ and e is a vertex of Δ, then $\mathcal{S}_{\Delta,e} = 0$ implies $\mathcal{D}^d_{\Delta,e} = 0$.*

Proof. It is easy to see that $\mathcal{S}_{\Delta^{(d-1)},e}$ has no power of t higher than $\dim\Delta^{(d-1)} + 1 = d$. But since $\mathcal{S}_{\Delta,e} = 0$, Lemma 4.15 implies that $0 = [t^{d+1}]\mathcal{S}_{\Delta^{(d-1)},e} = \mathcal{D}^d_{\Delta,e}$. Here, we are using the coefficient notation $[t^i](\sum_j a_jt^j) := a_i$. □

Lemma 4.17. *If Δ is a simplicial complex and e is a vertex of Δ, then*

$$\mathcal{D}^d_{\Delta,e} = \mathcal{D}^d_{\Delta^{[d]},e}.$$

Proof. By expanding $\mathcal{D}^d_{\Delta,e}$ we need only show that we may replace Δ by $\Delta^{[d]}$ in each of $S''_{\Delta,d}$, $S''_{\Delta-e,d}$, $S''_{\Delta/e,d-1}$, $S''_{(\Delta-e,\Delta/e),d}$, and $f_{d-1}(\Delta/e)$. But this follows from Lemma 4.10 and the definition of S''. □

Theorem 4.18. *If $\dim\Delta \le d$, and e is a vertex of Δ, then Δ satisfies the spectral recursion with respect to e iff $\Delta^{(d-1)}$ and $\Delta^{[d]}$ do as well.*

Proof. First assume Δ satisfies the spectral recursion with respect to e. Then $0 = \mathcal{S}_{\Delta,e}$. By Lemma 4.16, then, $\mathcal{D}^d_{\Delta,e} = 0$. And then by Lemma 4.15, $\mathcal{S}_{\Delta^{(d-1)},e} = 0$. Furthermore, $\mathcal{D}^d_{\Delta^{[d]},e} = \mathcal{D}^d_{\Delta,e} = 0$, by Lemma 4.17.
Conversely, assume $\Delta^{(d-1)}$ and $\Delta^{[d]}$ satisfy the spectral recursion with respect to e. By Lemma 4.16, then $\mathcal{D}^d_{\Delta^{[d]},e} = 0$. And then by Lemmas 4.17 and 4.15,

$$\mathcal{S}_{\Delta,e} = \mathcal{S}_{\Delta^{(d-1)},e} + (t^d + t^{d+1})\mathcal{D}^d_{\Delta,e} = \mathcal{S}_{\Delta^{(d-1)},e} + (t^d + t^{d+1})\mathcal{D}^d_{\Delta^{[d]},e} = 0.$$

□

Corollary 4.19. *If $\dim\Delta \le d$, then Δ satisfies the spectral recursion iff $\Delta^{(d-1)}$ and $\Delta^{[d]}$ do as well.*

4.5. Shifted complexes. Recall a *k-set* is a set with k elements, and a *k-family* over ground set E is a collection of k-subsets of E. For a k-set F, let $\operatorname{bd} F$ denote the $(k-1)$-family of all $(k-1)$-subsets of F. For a k-family $\mathcal{K}$, its *unsigned boundary* $\operatorname{bd}\mathcal{K}$ is the $(k-1)$-family $\cup_{F\in\mathcal{K}} \operatorname{bd} F$.
If $F = \{f_1 < \cdots < f_k\}$ and $G = \{g_1 < \cdots < g_k\}$ are k-subsets of integers, then $F \leq_P G$ under the *componentwise partial order* if $f_p \leq g_p$ for all p. A k-family $\mathcal{K}$ is *shifted* if $F \leq_P G$ and $G \in \mathcal{K}$ together imply that $F \in \mathcal{K}$. A simplicial complex Δ is *shifted* if Δ_i is shifted for every i. The useful properties of shifted families in the following lemma are easy to verify.

Lemma 4.20. *If $\mathcal{K}_1$ and $\mathcal{K}_2$ are shifted families, then so are $\operatorname{bd}\mathcal{K}_1$ and $\mathcal{K}_1 \cap \mathcal{K}_2$.*

We say that Δ is a *near-cone with apex* 1 if $\operatorname{bd}(\Delta - 1) \subseteq \Delta/1$, where bd denotes the usual unsigned boundary complex consisting of all faces that are not facets. Equivalently, Δ is a near-cone with apex 1 if $F - v \,\dot\cup\, 1 \in \Delta$ whenever $F \in \Delta$, $1 \notin F$, and $v \in F$. (See, *e.g.*, [7] for more on near-cones.) We omit the easy proofs of the following two lemmas.

Lemma 4.21. *Let Δ be a simplicial complex on $[n]$. Then Δ is shifted if and only if Δ is a near-cone with apex 1, and both $\Delta - 1, \Delta/1$ are shifted with respect to the ordered vertex set $[2, n]$.*

Lemma 4.22. *If Δ is a pure d-dimensional near-cone with apex 1, then*

$$\Delta = (1 * (\Delta - 1))^{(d)}$$

Theorem 4.23. *If Δ is a shifted simplicial complex, then Δ satisfies the spectral recursion, equation (2).*

Proof. The proof is by induction on the dimension and number of vertices of Δ. The base cases, when $\dim\Delta = 0$ or Δ has one vertex (a special case of $\dim\Delta = 0$, anyway) are easy to check.
Assume $\dim\Delta = d \geq 1$. By induction, $\Delta^{(d-1)}$ satisfies the spectral recursion. By Corollary 4.19, it remains to show that $\Delta^{[d]}$ satisfies the spectral recursion as well.
To this end, first note that Δ_d, the family of facets of $\Delta^{[d]}$, is shifted; then, by Lemma 4.20 and reverse induction on dimension, $\Delta^{[d]}$ is shifted. By definition, $\Delta^{[d]}$ is also pure, so Lemma 4.22 implies

$$\Delta^{[d]} = (1 * (\Delta^{[d]} - 1))^{(d)}.$$

Since $\Delta^{[d]}$ is shifted, $\Delta^{[d]} - 1$ is also shifted, with one less vertex, and so satisfies the spectral recursion, by induction. Thus $1 * (\Delta^{[d]} - 1)$ also satisfies the spectral recursion by Proposition 2.2 and Corollary 4.5. Then Corollary 4.19 guarantees that $\Delta^{[d]}$ satisfies the spectral recursion. □

5. Arbitrary shifted simplicial pairs

Merris [29] found a simple description of the Laplacian spectrum of a shifted graph (2-family), in terms of the degree sequence of the graph. This was generalized in [16] to shifted families, by suitably generalizing the notion of degree sequence. In this section, we extend both the theorem, and the notion of degree sequence, to shifted family *pairs* (Theorem 5.7). As in [16], the technique is to find identical recursive formulas, similar to those in [16], for the Laplacian spectrum (Corollary 5.4) and the generalized degree sequence (Lemma 5.6), in Subsections 5.1 and 5.2, respectively. The two threads are tied together with the proof of Theorem 5.7 in Subsection 5.3. Along the way, we rely upon tools developed in Section 4.

Grone and Merris [22] conjectured that Merris' description of the spectrum of a shifted graph becomes a majorization inequality for an *arbitrary* graph. This was also generalized from graphs to families (though still not proved) in [16]. In Subsection 5.3, we also further extend this conjecture from families to family pairs (Conjecture 5.8).

5.1. Laplacians. Recall the definition of family in Subsection 4.5. If (for some k), $\mathcal{K}$ and $\mathcal{K}'$ are a k-family and $(k-1)$-family, respectively, on the same ground set of vertices, then we will say $(\mathcal{K},\mathcal{K}')$ is a *family pair*, but we set $(\mathcal{K},\mathcal{K}') = (\mathcal{K},\mathcal{K}'')$ when $(\operatorname{bd}\mathcal{K}) \cap \mathcal{K}' = (\operatorname{bd}\mathcal{K}) \cap \mathcal{K}''$ (more formally, then, a family pair is an equivalence class on ordered pairs of families). We will say $(\mathcal{K},\mathcal{K}')$ is a *shifted family pair* when $\mathcal{K}$ is shifted and $(\mathcal{K},\mathcal{K}') = (\mathcal{K},\mathcal{K}'')$ for some $\mathcal{K}''$ that is shifted on the same ordered ground set as $\mathcal{K}$.

Let $C(\mathcal{K};\mathbb{R})$ denote the *oriented chains* of k-family $\mathcal{K}$, *i.e.*, the formal $\mathbb{R}$-linear sums of oriented faces $[F]$ such that $F \in \mathcal{K}$. If $(\mathcal{K},\mathcal{K}')$ is a family pair, then the *boundary operator* $\partial_{(\mathcal{K},\mathcal{K}')}\colon C(\mathcal{K};\mathbb{R}) \to C((\operatorname{bd}\mathcal{K})\backslash\mathcal{K}';\mathbb{R})$ is defined as it is for simplicial complexes, except that the sum is now restricted to faces in $(\operatorname{bd}\mathcal{K})\backslash\mathcal{K}'$. Equivalently, $\partial_{(\mathcal{K},\mathcal{K}')} = \partial_{(\Delta(\mathcal{K}),\Delta(\mathcal{K}'));k-1}$, when $\mathcal{K}$ is a k-family and $\mathcal{K}'$ is a $(k-1)$-family. As with simplicial complexes, the boundary operator has an adjoint $\partial^*_{(\mathcal{K},\mathcal{K}')}$, so the matrices representing ∂ and ∂^* in the natural bases are transposes of one another.

Definition. The *Laplacian* of $(\mathcal{K},\mathcal{K}')$ is the map $L(\mathcal{K},\mathcal{K}')\colon C(\mathcal{K};\mathbb{R}) \to C(\mathcal{K};\mathbb{R})$ defined by

$$L(\mathcal{K},\mathcal{K}') := \partial^*_{(\mathcal{K},\mathcal{K}')}\partial_{(\mathcal{K},\mathcal{K}')}.$$

It immediately follows that

$$L(\mathcal{K},\mathcal{K}') = L''_{k-1}(\Delta(\mathcal{K}),\Delta(\mathcal{K}')), \tag{19}$$

where $\Delta(\mathcal{K})$ denote the pure $(k-1)$-dimensional simplicial complex whose facets are the members of k-family $\mathcal{K}$.

It should be clear that $\partial_{(\mathcal{K},\mathcal{K}')}$, and hence $L(\mathcal{K},\mathcal{K}')$, is well-defined on family pairs; that is, $\partial_{(\mathcal{K},\mathcal{K}')} = \partial_{(\mathcal{K},\mathcal{K}'')}$ and $L(\mathcal{K},\mathcal{K}') = L(\mathcal{K},\mathcal{K}'')$, when $(\mathcal{K},\mathcal{K}') = (\mathcal{K},\mathcal{K}'')$. Of course, we may always specialize to a single family by letting $\mathcal{K}' = \emptyset$.

Recall that Δ_i denotes the $(i+1)$-family of i-dimensional faces of simplicial complex Δ.

Lemma 5.1. *If* $\dim\Delta' \le d-1$, *then*

$$L''_d(\Delta,\Delta') = L(\Delta_d,\Delta'_{d-1}).$$

Proof. The boundary maps $\partial_{(\Delta,\Delta');d}$ and $\partial_{(\Delta_d,\Delta'_{d-1})}$ used to define $L''_d(\Delta,\Delta')$ and $L(\Delta_d,\Delta'_{d-1})$, respectively, both act on $C_d(\Delta;\mathbb{R})$. By the definitions of L and L''_d, then, it will suffice to show that, for any $F\in\Delta_d$,

$$\partial_{(\Delta,\Delta');d}[F]=\partial_{(\Delta_d,\Delta'_{d-1})}[F]. \tag{20}$$

Now, the only difference between the left-hand and right-hand sides of this equation is that the left-hand side is a sum restricted to faces in the set difference $\Delta_{d-1}\backslash\Delta'_{d-1}$, and the right-hand side is a sum restricted to faces in $(\operatorname{bd}\Delta_d)\backslash\Delta'_{d-1}$. Since Δ is a simplicial complex, $\operatorname{bd}\Delta_d\subseteq\Delta_{d-1}$, so the only difference between the two sums is provided by faces in $\Delta_{d-1}\backslash(\operatorname{bd}\Delta_d)$. But any such face will not be in $\operatorname{bd}F$, the unsigned boundary of F, and thus not appear in the expression for the signed boundary map, anyway. (Equivalently, the matrices representing $\partial_{(\Delta,\Delta');d}$ and $\partial_{(\Delta_d,\Delta'_{d-1})}$ differ only in extra 0 rows indexed by $(d-1)$-dimensional faces of Δ not contained in any d-dimensional face of Δ, and these extra 0 rows do not affect $L=\partial^*\partial$.) This establishes equation (20), and hence the lemma. (*Cf.* the proof of Lemma 4.11). □

Lemma 5.1, Proposition 4.6, and equation (19) allow us to go back and forth between families and complexes.

Lemma 5.2. *If (Γ,Γ') is a simplicial pair, then*

$$qtS_{(\Gamma,\Gamma')}=S''_{(1*\Gamma,1*\Gamma')}.$$

Proof. We compute $S_{(1*\Gamma,1*\Gamma')}$ in two different ways. Since $1*\Gamma$ and $1*\Gamma'$ are cones, $(1*\Gamma,1*\Gamma')$ has trivial homology, so $B_{(1*\Gamma,1*\Gamma')}=0$. Thus, by Lemma 4.7,

$$S_{(1*\Gamma,1*\Gamma')}=(1+t^{-1})S''_{(1*\Gamma,1*\Gamma')}+B_{(1*\Gamma,1*\Gamma')}=t^{-1}(1+t)S''_{(1*\Gamma,1*\Gamma')}.$$

On the other hand, by Corollary 4.3,

$$S_{(1*\Gamma,1*\Gamma')}=S_{1*(\Gamma,\Gamma')}=q(1+t)S_{(\Gamma,\Gamma')}.$$

The lemma now follows immediately. □

Define $\mathbf{s}(\mathcal{K},\mathcal{K}')$ to be the multiset of eigenvalues of $L(\mathcal{K},\mathcal{K}')$, arranged in weakly decreasing order. When $\mathbf{s}(\mathcal{K},\mathcal{K}')$ consists of non-negative integers, it is a partition. We will use the notation of [28] for partitions, except that we will denote the *conjugate* or *transpose* of partition $\boldsymbol{\lambda}$ by $\boldsymbol{\lambda}^T$. In particular, $1^m=(m)^T$ denotes the partition consisting of m 1's. Recall from Subsection 4.2 the definitions of $\overset{\circ}{=}$ and $\cup$ for multisets, which apply equally well to partitions and weakly decreasing sequences.
Recall the definition of near-cone from subsection 4.5.

Lemma 5.3. *If $\Delta'\subseteq\Delta$ are pure near-cones with apex* 1*, and* $\dim\Delta=d$ *and* $\dim\Delta'=d-1$*, then, as partitions,*

$$\mathbf{s}''_d(\Delta,\Delta')\overset{\circ}{=}1^{f_{d-1}(\Delta/1)-f_{d-1}(\Delta'-1)}+(\mathbf{s}''_d(\Delta-1,\Delta'-1)\cup\mathbf{s}''_{d-1}(\Delta/1,\Delta'/1)).$$

Proof. Recall the coefficient notation $[t^i](\sum_j a_j t^j) := a_i$. First note

$$[t^i]S''_{(\Gamma,\Gamma')} = S''_{(\Gamma,\Gamma'),i-1} \tag{21}$$

$$[t^i]B_{(\Gamma,\Gamma')} = \tilde{\beta}_{i-1}(\Gamma,\Gamma') \tag{22}$$

for any simplicial pair (Γ,Γ') and for any i. Then, by Lemmas 4.12, 4.22, 5.2, and equation (21),

$$S''_{(\Delta,\Delta'),d} = q[t^d]S_{(\Delta-1,\Delta'-1)},$$

so $\mathbf{s}''_d(\Delta,\Delta')$ has just as many non-zero parts as there are terms in $q[t^d]S_{(\Delta-1,\Delta'-1)}$. Lemma 4.7 and equations (21) and (22) now imply

$$\begin{aligned} S''_{(\Delta,\Delta'),d} &= q[t^d]S_{(\Delta-1,\Delta'-1)} \\ &= q(S''_{(\Delta-1,\Delta'-1),d-1} + S''_{(\Delta-1,\Delta'-1),d} + \tilde{\beta}_{d-1}(\Delta-1,\Delta'-1)), \end{aligned}$$

so the non-zero parts of $\mathbf{s}''_d(\Delta,\Delta')$ are given by adding 1 to every element of the multiset union of three partitions: $\mathbf{s}''_{d-1}(\Delta-1,\Delta'-1)$; $\mathbf{s}''_d(\Delta-1,\Delta'-1)$; and the partition consisting of $\tilde{\beta}_{d-1}(\Delta-1,\Delta'-1)$ zeros. This means

$$\mathbf{s}''_d(\Delta,\Delta') \mathrel{\mathring{=}} 1^m + (\mathbf{s}''_{d-1}(\Delta-1,\Delta'-1) \cup \mathbf{s}''_d(\Delta-1,\Delta'-1)), \tag{23}$$

where m is the number of terms in $q[t^d]S_{(\Delta-1,\Delta'-1)}$, since we established above that $\mathbf{s}''_d(\Delta,\Delta')$ has m non-zero parts. But $\Delta' \subseteq \Delta$ easily implies $\Delta'-1 \subseteq \Delta-1$, and so there are

$$m = f_{d-1}(\Delta-1) - f_{d-1}(\Delta'-1)$$

terms in $q[t^d]S_{(\Delta-1,\Delta'-1)}$.

It is easy to verify that, since Δ and Δ' are pure near-cones (of dimensions d and $d-1$, respectively) with apex 1,

$$(\Delta-1)^{(d-1)} = \Delta/1;\ \text{and} \tag{24}$$

$$(\Delta'-1)^{(d-2)} = \Delta'/1. \tag{25}$$

From equation (24), we conclude

$$m = f_{d-1}(\Delta-1) - f_{d-1}(\Delta'-1) = f_{d-1}(\Delta/1) - f_{d-1}(\Delta'-1). \tag{26}$$

From equations (24) and (25), and Lemma 4.11, we conclude

$$\mathbf{s}''_{d-1}(\Delta-1,\Delta'-1) \mathrel{\mathring{=}} \mathbf{s}''_{d-1}(\Delta/1,\Delta'/1). \tag{27}$$

The lemma now follows from equations (23), (26), and (27). □

Definition. Let $\mathcal{K}$ be a k-family on ground set E, and $e \in E$. Then the *deletion* of $\mathcal{K}$ with respect to e is the k-family

$$\mathcal{K} - e = \{F \in \mathcal{K} \colon e \notin F\}$$

on ground set $E - e$, and the *contraction* of $\mathcal{K}$ with respect to e is the $(k-1)$-family

$$\mathcal{K}/e = \{F - e \colon F \in \mathcal{K},\ e \in F\}$$

on ground set $E - e$.

The following identities are immediate: $(\Delta(\mathcal{K})-e)_{k-1} = \mathcal{K}-e$, $(\Delta(\mathcal{K})/e)_{k-2} = \mathcal{K}/e$, and $\Delta(\mathcal{K})_{k-1} = \mathcal{K}$.
Define a k-family to be a *near-cone with apex* 1 when $\mathrm{bd}(\mathcal{K}-1) \subseteq \mathcal{K}/1$. It is an easy exercise to verify that $\mathcal{K}$ is a near-cone iff $\Delta(\mathcal{K})$ is a near-cone. Also, as with simplicial complexes (Lemma 4.21), $\mathcal{K}$ is shifted iff $\mathcal{K}$ is a near-cone with apex 1 such that $\mathcal{K}-1$ and $\mathcal{K}/1$ are shifted. The following corollary generalizes [16, Lemma 5.3].

Corollary 5.4. *If $\mathcal{K}$ and $\mathcal{K}'$ are near-cone families with apex* 1 *such that $\mathcal{K}' \subseteq \mathrm{bd}\,\mathcal{K}$, then*

$$\mathbf{s}(\mathcal{K},\mathcal{K}') \mathrel{\mathring{=}} 1^{|\mathcal{K}/1|-|\mathcal{K}'-1|} + (\mathbf{s}(\mathcal{K}-1,\mathcal{K}'-1) \cup \mathbf{s}(\mathcal{K}/1,\mathcal{K}'/1)).$$

Proof. Say $\mathcal{K}$ is a k-family, so $\mathcal{K}'$ is a $(k-1)$-family. Let $\Delta = \Delta(\mathcal{K})$ and $\Delta' = \Delta(\mathcal{K}')$. From $\mathcal{K}' \subseteq \mathrm{bd}\,\mathcal{K}$, it follows that $\Delta' \subseteq \Delta$. Then, by Lemmas 5.1 and 5.3,

$$\begin{aligned}
\mathbf{s}(\mathcal{K},\mathcal{K}') &= \mathbf{s}(\Delta_{k-1},\Delta'_{k-2}) = \mathbf{s}''_{k-1}(\Delta,\Delta') \\
&\mathrel{\mathring{=}} 1^{f_{k-2}(\Delta/1)-f_{k-2}(\Delta'-1)} + (\mathbf{s}''_{k-1}(\Delta-1,\Delta'-1) \cup \mathbf{s}''_{k-2}(\Delta/1,\Delta'/1)) \\
&= 1^{f_{k-2}(\Delta/1)-f_{k-2}(\Delta'-1)} \\
&\quad + (\mathbf{s}((\Delta-1)_{k-1},(\Delta'-1)_{k-2}) \cup \mathbf{s}((\Delta/1)_{k-2},(\Delta'/1)_{k-3})) \\
&= 1^{|\mathcal{K}/1|-|\mathcal{K}'-1|} + (\mathbf{s}(\mathcal{K}-1,\mathcal{K}'-1) \cup \mathbf{s}(\mathcal{K}/1,\mathcal{K}'/1)).
\end{aligned}$$

□

5.2. Degree sequences.

Notation. We will write $F-\lambda$ to denote the set difference $F\backslash\{\lambda\}$, with the implicit assumption that $\lambda \in F$, just as writing $F \mathbin{\dot\cup} \mu$ carries the implicit assumption that $\mu \notin F$. For instance, $\{F \in \mathcal{K} \colon F-\lambda \notin \mathcal{K}'\}$ in the following definition is shorthand for $\{F \in \mathcal{K} \colon \lambda \in F, F\backslash\{\lambda\} \notin \mathcal{K}'\}$.

Definition. Let $(\mathcal{K},\mathcal{K}')$ be a family pair on ground set E. Define the *degree* of λ in $(\mathcal{K},\mathcal{K}')$ by

$$d_\lambda(\mathcal{K},\mathcal{K}') := |\{F \in \mathcal{K} \colon F-\lambda \notin \mathcal{K}'\}|.$$

It is easy to see that d_λ is well-defined on family pairs; that is, $d_\lambda(\mathcal{K},\mathcal{K}') = d_\lambda(\mathcal{K},\mathcal{K}'')$ when $(\mathcal{K},\mathcal{K}') = (\mathcal{K},\mathcal{K}'')$. The *degree sequence* $\mathbf{d} = \mathbf{d}(\mathcal{K},\mathcal{K}')$ is the partition whose parts are $\{d_\lambda \colon \lambda \in E\}$.

In other words, to find the degree sequence of $(\mathcal{K},\mathcal{K}')$, label all the edges in the Hasse diagram of $\Delta(\mathcal{K})$ in the natural way, by the vertex being added; then d_λ counts the number of edges in the Hasse diagram labelled λ, and connecting a face in $\mathcal{K}$ with a face in $(\mathrm{bd}\,\mathcal{K})\backslash\mathcal{K}'$. When $\mathcal{K}' = \emptyset$, then $\mathbf{d}(\mathcal{K}) = \mathbf{d}(\mathcal{K},\emptyset)$ is the generalized degree sequence of family $\mathcal{K}$ defined in [16, Section 2]. It is also easy to see that $d_\lambda(\mathcal{K}) = |\mathcal{K}/\lambda|$. When $\mathcal{K}$ is the set of edges of a graph, then $\mathbf{d}(\mathcal{K})$ is the usual degree sequence of a graph.

Lemma 5.5. *If* $(\mathcal{K}, \mathcal{K}')$ *is a shifted family pair on* $[1, n]$ *and* $1 \leq \lambda < \mu \leq n$, *then* $d_\lambda(\mathcal{K}, \mathcal{K}') \geq d_\mu(\mathcal{K}, \mathcal{K}')$*;* i.e.

$$\mathbf{d}(\mathcal{K}, \mathcal{K}') = (d_1(\mathcal{K}, \mathcal{K}'), d_2(\mathcal{K}, \mathcal{K}'), \ldots, d_n(\mathcal{K}, \mathcal{K}')).$$

In other words, the ordering of the degrees of the degree sequence of a shifted family pair is given by the linear ordering of their vertices.

Proof. It will suffice to find an injection from $\{F \in \mathcal{K} \colon F - \mu \notin \mathcal{K}'\}$, a set whose cardinality equals $d_\mu(\mathcal{K}, \mathcal{K}')$, into $\{F \in \mathcal{K} \colon F - \lambda \notin \mathcal{K}'\}$, a set whose cardinality equals $d_\lambda(\mathcal{K}, \mathcal{K}')$. It is easy to verify, using that $\mathcal{K}$ and $\mathcal{K}'$ are shifted, that such an injection ϕ is given by

$$\phi(F) = \begin{cases} F & \text{if } \lambda \in F \\ F - \mu \mathbin{\dot\cup} \lambda & \text{if } \lambda \notin F. \end{cases}$$

□

The following lemma generalizes [16, Lemma 5.2]

Lemma 5.6. *If* $\mathcal{K}$ *and* $\mathcal{K}'$ *are shifted families on ground set* $[1, n]$, *and* $\mathcal{K}' \subseteq \operatorname{bd} \mathcal{K}$, *then, as partitions,*

$$\mathbf{d}(\mathcal{K}, \mathcal{K}')^T = 1^{|\mathcal{K}/1| - |\mathcal{K}' - 1|} + (\mathbf{d}(\mathcal{K} - 1, \mathcal{K}' - 1)^T \cup \mathbf{d}(\mathcal{K}/1, \mathcal{K}'/1)^T).$$

Proof. By standard partition arguments, this reduces to showing

$$\mathbf{d}(\mathcal{K}, \mathcal{K}') = (|\mathcal{K}/1| - |\mathcal{K}' - 1|) \cup (\mathbf{d}(\mathcal{K} - 1, \mathcal{K}' - 1) + \mathbf{d}(\mathcal{K}/1, \mathcal{K}'/1)),$$

which is a direct consequence of the following two facts:

- $d_1(\mathcal{K}, \mathcal{K}') = |\mathcal{K}/1| - |\mathcal{K}' - 1|$; and
- if $\lambda > 1$, then $d_\lambda(\mathcal{K}, \mathcal{K}') = d_\lambda(\mathcal{K} - 1, \mathcal{K}' - 1) + d_\lambda(\mathcal{K}/1, \mathcal{K}'/1)$.

The indexing on the second fact is indeed what is necessary, thanks to Lemma 5.5, because $\mathcal{K} - 1$, $\mathcal{K}' - 1$, $\mathcal{K}/1$, and $\mathcal{K}'/1$ each have ground set $[2, n]$. Each fact is an easy exercise, the first of which depends upon $\mathcal{K}$ being shifted. □

5.3. A relative generalized Merris theorem. Merris [29, Theorem 2] showed that when $\mathcal{K}$ is the 2-family of edges of a shifted graph, then $\mathbf{s}(\mathcal{K}) \doteq \mathbf{d}(\mathcal{K})^T$. This was generalized in [16, Theorem 1.1] to allow $\mathcal{K}$ to be any shifted family. The main result of this section, below, further generalizes this to shifted family *pairs*. The proof is similar to that of [16, Theorem 1.1].

Theorem 5.7. *If* $(\mathcal{K}, \mathcal{K}')$ *is a shifted family pair, then*

$$\mathbf{s}(\mathcal{K}, \mathcal{K}') \doteq \mathbf{d}(\mathcal{K}, \mathcal{K}')^T$$

Proof. By Lemmas 4.20 and 4.21, $(\mathcal{K} - 1, \mathcal{K}' - 1) = (\mathcal{K} - 1, (\mathcal{K}' - 1) \cap \operatorname{bd}(\mathcal{K} - 1))$ and $(\mathcal{K}/1, \mathcal{K}'/1) = (\mathcal{K}/1, (\mathcal{K}'/1) \cap \operatorname{bd}(\mathcal{K}/1))$ are shifted family pairs. Then the result is immediate from Corollary 5.4, Lemmas 4.21 and 5.6, and induction on the number of vertices. □

Grone and Merris [22, Conjecture 2] conjectured that when $\mathcal{K}$ is the 2-family of edges of an *arbitrary* graph, then the equality (modulo zeros) $\mathbf{s}(\mathcal{K}) \mathrel{\mathring{=}} \mathbf{d}(\mathcal{K})^T$ above becomes a majorization inequality $\mathbf{s}(\mathcal{K}) \trianglelefteq \mathbf{d}(\mathcal{K})^T$, *i.e.*, $\sum_{j=1}^k s_j \leq \sum_{j=1}^k d_j^T$ for all k, where $\mathbf{s}(\mathcal{K}) = (s_1, s_2, \ldots)$ and $\mathbf{d}(\mathcal{K})^T = (d_1^T, d_2^T, \ldots)$ are written as weakly decreasing sequences. This majorization inequality was also conjectured (but not proved) to hold when $\mathcal{K}$ is any family, in [16, Conjecture 1.2]. Based on no more than a few examples, and that [16, Theorem 1] successfully extends to pairs in Theorem 5.7 above, we extend this conjecture to family pairs as well.

Conjecture 5.8. *If $(\mathcal{K}, \mathcal{K}')$ is a family pair, then*

$$\mathbf{s}(\mathcal{K}, \mathcal{K}') \trianglelefteq \mathbf{d}(\mathcal{K}, \mathcal{K}')^T.$$

Stephen [33, Theorem 4.3.1] has shown that if the Grone-Merris conjecture is true for all graphs, then Conjecture 5.9 holds for graph pairs ($\mathcal{K}$ is a 2-family and $\mathcal{K}'$ is a 1-family).

Remark 5.9. Theorem 5.7 suffices to find the spectrum of a shifted simplicial pair (that is, a simplicial pair (Δ, Δ'), where Δ and Δ' are each shifted on the same ordered ground set), not just a shifted family pair. To see this, first note that by Proposition 4.6, finding $\mathbf{s}''_i(\Delta, \Delta')$ for all i determines the spectrum of the simplicial pair (Δ, Δ'). Since $\mathbf{s}''_i$ depends only on i- and $(i-1)$-dimensional faces, $\mathbf{s}''_i(\Delta, \Delta') = \mathbf{s}''_i(\Delta^{(i)}, \Delta'^{(i)})$. Finally, then, $\mathbf{s}''_i(\Delta, \Delta') = \mathbf{s}''_i(\Delta^{(i)}, \Delta'^{(i)}) = \mathbf{s}''_i(\Delta^{(i)}, \Delta'^{(i-1)}) = \mathbf{s}(\Delta_i, \Delta'_{i-1})$, by Lemmas 4.11 and 5.1.

6. Operations that preserve the spectral recursion

In this section, we see how the spectral recursion, equation (2), and the spectrum polynomial behave with respect to some natural operators on simplicial complexes. Each operator has significance for, or motivation from, matroids and/or shifted complexes. Our main results are that the property of satisfying the spectral recursion is preserved by disjoint union (Corollary 6.11), Alexander duality (Corollary 6.8), and, with a slight modification allowing order filters as well as simplicial complexes, two other dual operators (Theorems 6.3 and 6.6).

6.1. Duals.

The Tutte polynomial for matroids (see, *e.g.*, [12]) whose recursion ($T_M = T_{M-e} + T_{M/e}$) inspired and resembles the spectral recursion, is well-behaved with respect to matroid duals ($T_M(x, y) = T_{M^*}(y, x)$), so it is natural to ask what duality does to the spectrum polynomial and the spectral recursion. There are three natural involutions on simplicial complexes that are each appropriate generalizations of matroid duality. How these involutions affect the Laplacians of *families* has already been considered in [16, Section 4]. Recall that an *order filter* Ψ with vertices V is a collection of subsets of V, closed under taking supersets; that is, $F \in \Psi$ and $F \subseteq G \subseteq V$ together imply $G \in \Psi$.

Definition. Let Δ be a simplicial complex (respectively, order filter) with vertex set V. The *dual* of Δ is the order filter (respectively, simplicial complex)

$$\Delta^* = \{V - F \colon F \in \Delta\}.$$

The *complement* of Δ is the order filter (respectively, simplicial complex)

$$\Delta^c = \{F \subseteq V \colon F \notin \Delta\}.$$

The *Alexander dual* of Δ is the simplicial complex (respectively, order filter)

$$\Delta^\vee = \Delta^{*c} = \Delta^{c*}.$$

The Alexander dual has received attention lately in combinatorial topology (see, *e.g.*, [2, 6]) and in combinatorial commutative algebra (see, *e.g.*, [3, 4, 10, 17]).

It is easy to see that $\Delta^{**} = \Delta^{cc} = \Delta^{\vee\vee} = \Delta$ for every simplicial complex Δ, and similarly for order filters. If we define an order filter Ψ to be shifted when its every family Ψ_i of i-dimensional faces is shifted, then it is easy to see that duality and complementation preserve being shifted, though with the reverse vertex order. Consequently, Alexander duality preserves being shifted.

If Ψ and Ψ' are order filters on the same ground set of vertices, we define the *order filter pair* (Ψ, Ψ') to be the simplicial pair (Ψ'^c, Ψ^c), as defined in Section 2. (This means that, more formally, an order filter pair is an equivalence class on ordered pairs of order filters.) Thus $(\Psi, \Psi') = (\Omega, \Omega')$ when the set differences $\Psi \backslash \Psi'$ and $\Omega \backslash \Omega'$ are equal as subsets of the power set of the ground set of vertices. As with simplicial complexes, results and definitions about order filter pairs (Ψ, Ψ') may be specialized to a single order filter, by letting $\Psi' = \emptyset$, the empty order filter.

The definitions of deletion and contraction extend naturally to order filters. The deletion and contraction $\Psi - e$ and Ψ/e of an order filter Ψ on vertex set V are still order filters, though on vertex set $V - e$. In contrast to simplicial complexes, Ψ/e is not necessarily a subset of Ψ (though $\Psi - e \subseteq \Psi$, still), and $\Psi - e \subseteq \Psi/e$ (whereas, for simplicial complexes, $\Delta/e \subseteq \Delta - e$).

We now borrow a trick from [27, Proposition 6] (see also [16, Proposition 4.2]) to investigate how the dual affects Laplacians and the spectral recursion. Let (Δ, Δ') be a simplicial pair with vertex set $[n]$; it is easy to specialize from pairs of duals to a single dual, since the dual of the empty simplicial complex is again empty, so $\Delta^* = (\Delta^*, \emptyset) = (\Delta^*, \emptyset^*)$. Define $\phi_i(\Delta, \Delta') \colon C_i(\Delta, \Delta'; \mathbb{R}) \to C_{n-i-2}(\Delta^*, \Delta'^*; \mathbb{R})$ to be the $\mathbb{R}$-linear isomorphism induced by

$$\phi_i(\Delta, \Delta') \colon [F] \mapsto \sigma(F)[\overline{F}],$$

where $\sigma(F) = (-1)^{\sum_{j \in F} j}$, and $\overline{F} = [n] - F$.

Lemma 6.1. *Let (Δ, Δ') be a simplicial pair with vertex set $[n]$, and let $\phi_j = \phi_j(\Delta, \Delta')$ for any j. Then*

(1) $\phi_{i+1}^{-1} \partial_{(\Delta^*, \Delta'^*); n-i-2} \phi_i = -\partial^*_{(\Delta, \Delta'); i+1}$, *and*
(2) $\phi_i \partial_{(\Delta, \Delta'); i+1} \phi_{i+1}^{-1} = -\partial^*_{(\Delta^*, \Delta'^*); n-i-2}$.

Proof. These are each a routine check of signs. □

Corollary 6.2. *Let* (Δ, Δ') *be a simplicial pair with vertex set* $[n]$*, and let* $\phi_j = \phi_j(\Delta, \Delta')$ *for any* j*. Then*

$$L_i(\Delta, \Delta') = \phi^{-1} L_{n-i-2}(\Delta^*, \Delta'^*)\phi.$$

An immediate corollary is that, as first conjectured by V. Reiner (personal communication),

$$\mathbf{s}_i(\Delta, \Delta') = \mathbf{s}_{n-i-2}(\Delta^*, \Delta'^*), \tag{28}$$

which translates into generating functions as

$$S_{(\Delta^*,\Delta'^*)}(t,q) = t^n S_{(\Delta,\Delta')}(t^{-1}, q). \tag{29}$$

We might hope that, if simplicial complex Δ satisfies the spectral recursion with respect to a vertex e, then Δ^* would, too, but this is not quite true. Routine calculations using equation (29), and duality identitites $(\Delta - e)^* = \Delta^*/e$ and $(\Delta/e)^* = \Delta^* - e$, show that

$$S_{\Delta^*}(t,q) = qtS_{\Delta^*/e}(t,q) + qS_{\Delta^*-e}(t,q) + (1-q)tS_{(\Delta^*/e,\Delta^*-e)}(t,q). \tag{30}$$

We thus call

$$S_\Psi = qS_{\Psi-e}(t,q) + qtS_{\Psi/e}(t,q) + (1-q)tS_{(\Psi/e,\Psi-e)}(t,q) \tag{31}$$

the spectral recursion for order filters. Theorem 6.6 below provides further evidence that this is the right formulation for order filters. A unified approach to the spectral recursions for simplicial complexes and order filters is to develop a spectral recursion for simplicial complex pairs (which includes simplicial complexes and order filters as special cases), which is explored in [15].

Theorem 6.3. *If* Δ *is a simplicial complex and* e *is an element of its vertex set, then* Δ *satisfies the spectral recursion with respect to* e *iff* Δ^* *satisfies the spectral recursion for order filters, equation* (31)*, with respect to* e*.*

Proof. The forward implication follows from equation (30) above. The proof of the reverse implication is similar. □

The following proposition is a restatement of [16, Corollary 4.7].

Proposition 6.4. *Let* Δ *be a simplicial complex with vertex set* $[n]$*. If* $\lambda \neq n$*, then* $m_\lambda(L_i(\Delta)) = m_\lambda(L_{n-i-3}(\Delta^\vee))$*.*

The following corollary was first conjectured by V. Reiner (personal communication).

Corollary 6.5. *If* Δ *is a simplicial complex with vertex set* $[n]$*, then* $\mathbf{s}_{i-1}(\Delta)$ *and* $\mathbf{s}_i(\Delta^c)$ *agree, except for the multiplicity of* n*.*

Proof. By equation (28), $\mathbf{s}_i(\Delta^c) = \mathbf{s}_i(\Delta^{\vee *}) = \mathbf{s}_{n-i-2}(\Delta^\vee)$, so, if $\lambda \neq n$, then

$$m_\lambda(L_i(\Delta^c)) = m_\lambda(L_{n-i-2}(\Delta^\vee) = m_\lambda(L_{n-3-(n-i-2)}(\Delta)) = m_\lambda(L_{i-1})$$

by Proposition 6.4. □

The preceding proof is not as simple as it seems. The proof of Proposition 6.4 in [16, Corollary 4.7] is somewhat involved, and gets to the Alexander dual via the complement. Especially in light of the simplicity of the statement of Corollary 6.5, we might hope it would have a more direct proof that does not call upon the Alexander dual.

Corollary 6.5 translates into generating functions as

$$S_{\Delta^c}(t,q) = tS_\Delta(t,q) + q^n A_\Delta(t), \tag{32}$$

which we may rewrite as

$$S_\Delta(t,q) = t^{-1}S_{\Delta^c}(t,q) - q^n t^{-1} A_\Delta(t), \tag{33}$$

where $A_\Delta(t)$ is a polynomial in t that depends on Δ.

Theorem 6.6. *If e is a vertex of simplicial complex Δ, then Δ satisfies the spectral recursion with respect to e iff Δ^c satisfies the spectral recursion for order filters, equation* (31)*, with respect to e.*

Proof. First assume Δ^c satisfies the spectral recursion for order filters with respect to e. Then, we may use equations (32) and (33), and the complement identities $(\Delta - e)^c = \Delta^c - e$ and $(\Delta/e)^c = \Delta^c/e$, to compute

$$S_\Delta = qS_{\Delta-e} + qtS_{\Delta/e} + (1-q)S_{(\Delta-e,\Delta/e)} + q^n t^{-1}(A_{\Delta-e} + A_{\Delta/e} - A_\Delta). \tag{34}$$

By Lemma 2.4, all simplicial complexes satisfy the spectral recursion when $q = 1$, so plugging $q = 1$ into the above equation yields

$$S_\Delta(1,t) = S_\Delta(1,t) + 1^n t^{-1}(A_{\Delta-e} + A_{\Delta/e} - A_\Delta).$$

Therefore $A_{\Delta-e} + A_{\Delta/e} - A_\Delta = 0$, which, when plugged back into equation (34), proves Δ satisfies the spectral recursion with respect to e.

The reverse implication is proved similarly. □

Theorems 6.3 and 6.6 together imply the corresponding result for Alexander duality:

Theorem 6.7. *If e is a vertex of simplicial complex Δ, then Δ satisfies the spectral recursion with respect to e iff $\Delta^\vee$ satisfies the spectral recursion with respect to e.*

Corollary 6.8. *If Δ is a simplicial complex, then Δ satisfies the spectral recursion iff $\Delta^\vee$ does as well.*

6.2. Union. The Tutte polynomial is well-behaved with respect to matroid direct sum ($T_{M\oplus N} = T_M + T_N$), which corresponds to the union of simplicial complexes with disjoint vertex sets ($IN(M \oplus N) = IN(M) \cup IN(N)$). So it is natural to ask what disjoint union does to the spectrum polynomial and the spectral recursion.

Lemma 6.9. *If Δ and Γ are two non-empty simplicial complexes with disjoint vertex sets, then*

$$S_{\Delta\cup\Gamma} = S_\Delta + S_\Gamma + (t^0 + t^1)(q^{n+m} - (q^n + q^m)) + t^1 q^0,$$

where Δ and Γ have $n = f_0(\Delta)$ and $m = f_0(\Gamma)$ non-loop vertices, respectively.

Proof. For $i > 1$, it is clear that $L_{i-1}(\Delta \cup \Gamma) = L_{i-1}(\Delta) \oplus L_{i-1}(\Gamma)$, since no $(i-1)$-dimensional face of Δ has any boundary in Γ, and vice versa. Thus

$$\mathbf{s}_{i-1}(\Delta \cup \Gamma) = \mathbf{s}_{i-1}(\Delta) \cup \mathbf{s}_{i-1}(\Gamma)$$

for $i > 1$.

The vertices of Δ and Γ are disjoint, but they *share* the empty face in their boundary. It is easy to see that $\mathbf{s}''_0(\Sigma) \mathrel{\mathring{=}} (f_0(\Sigma))$ for any simplicial complex Σ, so $\mathbf{s}''_0(\Delta\cup\Gamma) = (n+m)$, while $\mathbf{s}''_0(\Delta)\cup\mathbf{s}''_0(\Gamma) = (n, m)$. Also, since $\dim L_0(\Delta\cup\Gamma) = f_0(\Delta \cup \Gamma) = n + m = f_0(\Delta) + f_0(\Gamma) = \dim L_0(\Delta) + \dim L_0(\Gamma)$, then $\mathbf{s}_0(\Delta \cup \Gamma)$ and $\mathbf{s}_0(\Delta) \cup \mathbf{s}_0(\Gamma)$ have the same number of parts. By Proposition 4.6, it then follows that

$$\mathbf{s}_0(\Delta \cup \Gamma) \cup (n, m) = \mathbf{s}_0(\Delta) \cup \mathbf{s}_0(\Gamma) \cup (n + m, 0).$$

(In other words, to change $\mathbf{s}_0(\Delta) \cup \mathbf{s}_0(\Gamma)$ into $\mathbf{s}_0(\Delta \cup \Gamma)$, replace (n, m) in $\mathbf{s}_0(\Delta)\cup\mathbf{s}_0(\Gamma)$ by $(n+m, 0)$ in $\mathbf{s}_0(\Delta\cup\Gamma)$.) Similarly, since $\Delta\cup\Gamma$, Δ, and Γ each have exactly one empty face, $\mathbf{s}_{-1}(\Delta\cup\Gamma)$ has one element, and $\mathbf{s}_{-1}(\Delta)\cup\mathbf{s}_{-1}(\Gamma)$ has two elements, and so

$$\mathbf{s}_{-1}(\Delta \cup \Gamma) = (n + m),$$

while

$$\mathbf{s}_{-1}(\Delta) \cup \mathbf{s}_{-1}(\Gamma) = (n, m).$$

The lemma now follows immediately. □

We continue to assume Δ and Γ are non-empty simplicial complexes with disjoint vertex sets, and that Γ has m non-loop vertices. By arguments similar to those in the proof of Lemma 6.9,

$$S_{(\Gamma,\emptyset)} = S_\Gamma - (t^0 + t^1)q^m + t^1 q^0,$$

and so

$$S_{(\Delta\cup\Gamma,\Delta')} = S_{(\Delta,\Delta')} + S_\Gamma - (t^0 + t^1)q^m + t^1 q^0. \tag{35}$$

Theorem 6.10. *If Δ satisfies the spectral recursion with respect to e, and Γ is any simplicial complex whose vertex set is disjoint from the vertex set of Δ, then $\Delta \cup \Gamma$ satisfies the spectral recursion with respect to e.*

Proof. If $\Gamma = \emptyset$, then the theorem is trivially true. Otherwise, it is a routine calculation with Lemma 6.9 and equation (35). □

Corollary 6.11. *If Δ and Γ each satisfy the spectral recursion, then so does their disjoint union $\Delta \cup \Gamma$.*

The following example shows that the *arbitrary* union of two simplicial complexes satisfying the spectral recursion does not itself necessarily satisfy the spectral recursion, even if both complexes are pure.

Example 6.12. Let Δ be the pure 1-dimensional simplicial complex on vertex set $\{a, b, c, d, e\}$ with facets $\{ab, ac, ad, ae, bc, bd\}$. (We omit brackets and commas from each face for clarity.) Let Γ be the pure 1-dimensional simplicial complex on the same vertex set with facets $\{ab, ac, ad, ae, de\}$. Now, Δ is

shifted with vertices ordered $a < b < c < d < e$, and Γ is shifted with vertices ordered $a < d < e < b < c$, so each satisfies the spectral recursion.
On the other hand, we can easily show $\Delta \cup \Gamma$ does not satisfy the spectral recursion with respect to vertex d. First check directly that $\Delta \cup \Gamma$ is not Laplacian integral. (Note that $\Delta \cup \Gamma$ is the 1-dimensional skeleton of the cone over Example 2.5.) Next, since $(\Delta \cup \Gamma) - d$ and $(\Delta \cup \Gamma)/d$ are each isomorphic to shifted complexes (with different vertex orders), they are each Laplacian integral. It is also easy to directly verify that $((\Delta \cup \Gamma) - d, (\Delta \cup \Gamma)/d)$ is Laplacian integral as well. Thus, the right-hand side of the spectral recursion in this instance has all integer exponents, but the left-hand side does not.

7. Acknowledgements

I am grateful to Vic Reiner, Eric Babson, Woong Kook, and Tamon Stephen for many helpful comments and suggestions, to Vic Reiner also for introducing me to this problem, and to several referees for their insightful advice.

References

[1] M. Aigner, *Combinatorial theory*, Grundlehren der Mathematischen Wissenschaften, vol. 234, Springer-Verlag, Berlin-New York, 1979.

[2] E. Babson, A. Björner, S. Linusson, J. Shareshian, and V. Welker, "Complexes of not i-connected graphs", *Topology* 38 (1999), 271–299.

[3] D. Bayer, "Monomial ideals and duality", unpublished lecture notes, 1996, available at
`http://www.math.columbia.edu/~bayer/papers/Duality_B96.pdf`.

[4] D. Bayer, H. Charalambous, and S. Popescu, "Extremal Betti numbers and applications to monomial ideals", *J. Alg.* 221 (1999), 497–512.

[5] A. Björner, "The homology and shellability of matroids and geometric lattices", in *Matroid applications*, pp. 226–283, Encyclopedia Math. Appl., vol. 40, Cambridge Univ. Press, Cambridge, 1992.

[6] A. Björner, L. M. Butler, and A. O. Matveev "Note on a combinatorial application of Alexander duality", *J. Combin. Theory Ser. A* 80 (1997), 163–165.

[7] A. Björner and G. Kalai, "An extended Euler-Poincaré theorem", *Acta Math.* 161 (1988), 279–303.

[8] A. Björner and M. L. Wachs, "Shellable nonpure complexes and posets. I", *Trans. Amer. Math. Soc.* 348 (1996), 1299–1327.

[9] A. Björner and M. L. Wachs, "Shellable nonpure complexes and posets. II", *Trans. Amer. Math. Soc.* 349 (1997), 3945–3975.

[10] W. Bruns and J. Herzog, "Semigroup rings and simplicial complexes", *J. Pure Appl. Algebra* 122 (1997), 185–208.

[11] T. Brylawski, "Constructions", in *Theory of matroids*, pp. 127–223, Encyclopedia Math. Appl., vol. 26, Cambridge Univ. Press, Cambridge, 1986.

[12] T. Brylawski and J. Oxley, "The Tutte polynomial and its applications", in *Matroid applications*, pp. 123–225, Encyclopedia Math. Appl., vol. 40, Cambridge Univ. Press, Cambridge, 1992.

[13] F. Chung and R. B. Ellis, "A chip-firing game and Dirichlet eigenvalues", *Discrete Math.* 257 (2002), 341–355.

[14] X. Dong and M. L. Wachs, "Combinatorial Laplacian of the matching complex", *Electron. J. Combin.* 9 (2002), #R17, 11 pp.

[15] A. M. Duval, "A Relative Laplacian spectral recursion", in preparation.

[16] A. M. Duval and V. Reiner, "Shifted simplicial complexes are Laplacian integral", *Trans. Amer. Math. Soc.* 354 (2002), 4313–4344.

[17] J. A. Eagon and V. Reiner, "Resolutions of Stanley-Reisner rings and Alexander duality", *J. Pure Appl. Algebra* 130 (1998), 265–275.

[18] B. Eckmann, "Harmonische Funktionen und Randwertaufgaben in einem Komplex", *Comment. Math. Helv.* 17 (1945), 240-255.

[19] G. Etienne and M. Las Vergnas, "External and internal elements of a matroid basis", *Discrete Math.* 179 (1998), 111–119.

[20] J. Friedman, "Computing Betti numbers via combinatorial Laplacian", in *Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing*, (Philadelphia, 1996), pp. 386–391, ACM, New York, 1996.

[21] J. Friedman and P. Hanlon, "On the Betti numbers of chessboard complexes", *J. Alg. Comb.* 8 (1998), 193–203.

[22] R. Grone and R. Merris, "The Laplacian spectrum of a graph II", *SIAM J. Disc. Math.* 7 (1994), 221–229.

[23] A. S. Jarrah and R. Laubenbacher, "Generic Cohen-Macaulay monomial ideals", *Ann. Comb.* 8 (2004), 45–61.

[24] C. J. Klivans, "Shifted matroid complexes", preprint, 2003.

[25] W. Kook, "A circuit formula for homology of matroids", preprint, 2000.

[26] W. Kook, "Recurrence relations for the spectrum polynomial of a matroid", *Disc. Appl. Math.* 143 (2004), 312–317.

[27] W. Kook, V. Reiner, and D. Stanton, "Combinatorial Laplacians of matroid complexes", *J. Amer. Math. Soc.* 13 (2000), 129–148.

[28] I. G. Macdonald, *Symmetric functions and Hall polynomials*, Oxford University Press, New York, 1979.

[29] R. Merris, "Degree maximal graphs are Laplacian integral", *Lin. Alg. Appl.* 199 (1994), 381–389.

[30] J. R. Munkres, *Elements of algebraic topology*, Addison-Wesley, Menlo Park CA, 1984.

[31] J. G. Oxley, *Matroid theory*, Oxford University Press, New York, 1992.

[32] J. S. Provan and L. J. Billera, "Decompositions of simplicial complexes related to diameters of convex polyhedra", *Math. Oper. Res.* 5 (1980), 576–594.

[33] T. Stephen, "A Majorization bound for the eigenvalues of some graph Laplacians", preprint, 2004; `arXiv:math.CO/0411153`.

[34] D. J. A. Welsh, *Matroid theory*, Academic Press, New York, 1976.

[35] N. White (ed.), *Theory of matroids*, Encyclopedia Math. Appl., vol. 26, Cambridge Univ. Press, Cambridge, 1986.

Art M. Duval
Department of Mathematical Sciences
University of Texas at El Paso
El Paso, TX 79968-0514
USA
artduval@math.utep.edu

Arithmetic Characteristic Classes of Automorphic Vector Bundles

J. I. Burgos Gil[1], J. Kramer, U. Kühn

Received: June 23, 2005

Communicated by Peter Schneider

Abstract. We develop a theory of arithmetic characteristic classes of (fully decomposed) automorphic vector bundles equipped with an invariant hermitian metric. These characteristic classes have values in an arithmetic Chow ring constructed by means of differential forms with certain log-log type singularities. We first study the cohomological properties of log-log differential forms, prove a Poincaré lemma for them and construct the corresponding arithmetic Chow groups. Then, we introduce the notion of log-singular hermitian vector bundles, which is a variant of the good hermitian vector bundles introduced by Mumford, and we develop the theory of arithmetic characteristic classes. Finally we prove that the hermitian metrics of automorphic vector bundles considered by Mumford are not only good but also log-singular. The theory presented here provides the theoretical background which is required in the formulation of the conjectures of Maillot-Roessler in the semi-abelian case and which is needed to extend Kudla's program about arithmetic intersections on Shimura varieties to the non-compact case.

Contents

[1]Partially supported by Grant DGI BFM2000-0799-C02-01 and DGI BFM2003-02914

1 Introduction

The main goal. The main purpose of this article is to extend the arithmetic intersection theory and the theory of arithmetic characteristic classes à la Gillet, Soulé to the category of (fully decomposed) automorphic vector bundles equipped with the natural equivariant hermitian metric on Shimura varieties of non-compact type. In order to achieve our main goal, an extension of the formalism by Gillet, Soulé taking into account vector bundles equipped with hermitian metrics allowing a certain type of singularities has to be provided. The main prerequisite for the present work is the article [10], where the foundations of cohomological arithmetic Chow groups are given. Before continuing to explain our main results and the outline of the paper below, let us fix some basic notations for the sequel.

Let B denote a bounded, hermitian, symmetric domain. By definition, $B = G/K$, where G is a semi-simple adjoint group and K a maximal compact subgroup of G with non-discrete center. Let Γ be a neat arithmetic subgroup of G; it acts properly discontinuously and fixed-point free on B. The quotient space $X = \Gamma\backslash B$ has the structure of a smooth, quasi-projective, complex variety. The complexification $G_{\mathbb{C}}$ of G yields the compact dual $\check{B}$ of B given by

$\check{B} = G_{\mathbb{C}}/P_+ \cdot K_{\mathbb{C}}$, where $P_+ \cdot K_{\mathbb{C}}$ is a suitable parabolic subgroup of G equipped with the Cartan decomposition of $\mathrm{Lie}(G)$ and P_+ is the unipotent radical of this parabolic subgroup. Every $G_{\mathbb{C}}$-equivariant holomorphic vector bundle $\check{E}$ on $\check{B}$ defines a holomorphic vector bundle E on X; E is called an *automorphic vector bundle*. An automorphic vector bundle E is called *fully decomposed*, if $E = E_\sigma$ is associated to a representation $\sigma : P_+ \cdot K_{\mathbb{C}} \longrightarrow \mathrm{GL}_n(\mathbb{C})$, which is trivial on P_+. Since K is compact, every fully decomposed automorphic vector bundle E admits a G-equivariant hermitian metric h.

Let us recall the following basic example. Let $\pi : \mathcal{B}_g^{(N)} \longrightarrow \mathcal{A}_g^{(N)}$ denote the universal abelian variety over the moduli space of principally polarized abelian varieties of dimension g with a level-N structure ($N \geq 3$); let $e : \mathcal{A}_g^{(N)} \longrightarrow \mathcal{B}_g^{(N)}$ be the zero section, and $\Omega = \Omega^1_{\mathcal{B}_g^{(N)}/\mathcal{A}_g^{(N)}}$ the relative cotangent bundle. The Hodge bundle $e^*\Omega$ is an automorphic vector bundle on $\mathcal{A}_g^{(N)}$, which is equipped with a natural hermitian metric h. Another example of an automorphic vector bundle on $\mathcal{A}_g^{(N)}$ is the determinant line bundle $\omega = \det(e^*\Omega)$; the corresponding hermitian automorphic line bundle $(\det(e^*\Omega), \det(h))$ is denoted by $\overline{\omega}$.

Background results. Let (E, h) be an automorphic hermitian vector bundle on $X = \Gamma \backslash B$, and $\overline{X}$ a smooth toroidal compactification of X. In [34], D. Mumford has shown that the automorphic vector bundle E admits a canonical extension E_1 to $\overline{X}$ characterized by a suitable extension of the hermitian metric h to E_1. However, the extension of h to E_1 is no longer a smooth hermitian metric, but inherits singularities of a certain type. On the other hand, it is remarkable that this extended hermitian metric behaves in many aspects like a smooth hermitian metric. In this respect, we will now discuss various definitions which were made in the past in order to extract basic properties for these extended hermitian metrics.

In [34], D. Mumford introduced the concept of good forms and good hermitian metrics. The good forms are differential forms, which are smooth on the complement of a normal crossing divisor and have certain singularities along this normal crossing divisor; the singularities are modeled by the singularities of the Poincaré metric. The good forms have the property of being locally integrable with zero residue. Therefore, they define currents, and the map from the complex of good forms to the complex of currents is a morphism of complexes. The good hermitian metrics are again smooth hermitian metrics on the complement of a normal crossing divisor and have logarithmic singularities along the divisor in question. Moreover, the entries of the associated connection matrix are good forms. The Chern forms for good hermitian vector bundles, i.e., of vector bundles equipped with good hermitian metrics, are good forms, and the associated currents represent the Chern classes in cohomology. Thus, in this sense, the good hermitian metrics behave like smooth hermitian metrics. In the same paper, D. Mumford proves that automorphic hermitian vector bundles are good hermitian vector bundles.

In [14], G. Faltings introduced the concept of a hermitian metric on line bundles with logarithmic singularities along a closed subvariety. He showed that the heights associated to line bundles equipped with singular hermitian metrics of this type have the same finiteness properties as the heights associated to line bundles equipped with smooth hermitian metrics. The Hodge bundle $\overline{\omega}$ on $\mathcal{A}_g^{(N)}$ equipped with the Petersson metric provides a prominent example of such a hermitian line bundle; it plays a crucial role in Faltings's proof of the Mordell conjecture. Recall that the height of an abelian variety A with respect to $\overline{\omega}$ is referred to as the Faltings height of A. It is a remarkable fact that, if A has complex multiplication of abelian type, its Faltings height is essentially given by a special value of the logarithmic derivative of a Dirichlet L-series. It is conjectured by P. Colmez that in the general case the Faltings height is essentially given by a special value of the logarithmic derivative of an Artin L-series.

In [30], the third author introduced the concept of logarithmically singular hermitian line bundles on arithmetic surfaces. He provided an extension of arithmetic intersection theory (on arithmetic surfaces) adapted to such logarithmically singular hermitian line bundles. The prototype of such a line bundle is the automorphic hermitian line bundle $\overline{\omega}$ on the modular curve $\mathcal{A}_1^{(N)}$. J.-B. Bost and, independently, U. Kühn calculated its arithmetic self-intersection number $\overline{\omega}^2$ to

$$\overline{\omega}^2 = d_N \cdot \zeta_{\mathbb{Q}}(-1)\left(\frac{\zeta'_{\mathbb{Q}}(-1)}{\zeta_{\mathbb{Q}}(-1)} + \frac{1}{2}\right);$$

here $\zeta_{\mathbb{Q}}(s)$ denotes the Riemann zeta function and d_N equals the degree of the classifying morphism of $\mathcal{A}_1^{(N)}$ to the coarse moduli space $\mathcal{A}_1^{(1)}$.

In [10], an abstract formalism was developed, which allows to associate to an arithmetic variety $\mathcal{X}$ arithmetic Chow groups $\widehat{\mathrm{CH}}^*(\mathcal{X},\mathcal{C})$ with respect to a cohomological complex $\mathcal{C}$ of a certain type. This formalism is an abstract version of the arithmetic Chow groups introduced in [8]. In [10], the arithmetic Chow ring $\widehat{\mathrm{CH}}^*(\mathcal{X},\mathcal{D}_{\mathrm{pre}})_{\mathbb{Q}}$ was introduced, where the cohomological complex $\mathcal{D}_{\mathrm{pre}}$ in question is built from pre-log and pre-log-log differential forms. This ring allows us to define arithmetic self-intersection numbers of automorphic hermitian line bundles on arithmetic varieties associated to $X = \Gamma\backslash B$. It is expected that these arithmetic self-intersection numbers play an important role for possible extensions of the Gross-Zagier theorem to higher dimensions (cf. conjectures of S. Kudla).

In [6], J. Bruinier, J. Burgos, and U. Kühn use the theory developed in [10] to obtain an arithmetic generalization of the Hirzebruch-Zagier theorem on the generating series for cycles on Hilbert modular varieties. Recalling that Hilbert modular varieties parameterize abelian surfaces with multiplication by the ring of integers $\mathcal{O}_K$ of a real quadratic field K, a major result in [6] is the following formula for the arithmetic self-intersection number of the automorphic hermitian line bundle $\overline{\omega}$ on the moduli space of abelian surfaces with multiplication by $\mathcal{O}_K$ with a fixed level-N structure

$$\overline{\omega}^3 = -d_N \cdot \zeta_K(-1)\left(\frac{\zeta_K'(-1)}{\zeta_K(-1)} + \frac{\zeta_{\mathbb{Q}}'(-1)}{\zeta_{\mathbb{Q}}(-1)} + \frac{3}{2} + \frac{1}{2}\log(D_K)\right);$$

here D_K is the discriminant of $\mathcal{O}_K$, $\zeta_K(s)$ is the Dedekind zeta function of K, and, as above, d_N is the degree of the classifying morphism obtained by forgetting the level-N structure.

As another application of the formalism developed in [10], we derived a height pairing with respect to singular hermitian line bundles for cycles in any codimension. Recently, G. Freixas in [15] has proved finiteness results for our height pairing, thus generalizing both Faltings's results mentioned above and the finiteness results of J.-B. Bost, H. Gillet and C. Soulé in [4] in the smooth case.

The main achievement of the present paper is to give constructions of arithmetic intersection theories, which are suited to deal with all of the above vector bundles equipped with hermitian metrics having singularities of a certain type such as the automorphic hermitian vector bundles on Shimura varieties of non-compact type.

For a perspective view of applications of the theory developed here, we refer to the conjectures of V. Maillot and D. Roessler [31], K. Köhler [26], and the program due to S. Kudla [28], [29], [27].

Arithmetic characteristic classes. We recall from [36] that the arithmetic K-group $\widehat{\mathrm{K}}_0(\mathcal{X})$ of an arithmetic variety $\mathcal{X}$ à la Gillet, Soulé is defined as the free group of pairs $(\overline{E},\eta)$ of a hermitian vector bundle $\overline{E}$ and a smooth differential form η modulo the relation

$$(\overline{S},\eta') + (\overline{Q},\eta'') = (\overline{E},\eta' + \eta'' + \widetilde{\mathrm{ch}}(\overline{\mathcal{E}})),$$

for every short exact sequence of vector bundles (equipped with arbitrary smooth hermitian metrics)

$$\overline{\mathcal{E}} : 0 \longrightarrow \overline{S} \longrightarrow \overline{E} \longrightarrow \overline{Q} \longrightarrow 0,$$

and for any smooth differential forms η',η''; here $\widetilde{\mathrm{ch}}(\overline{\mathcal{E}})$ denotes the (secondary) Bott-Chern form of $\overline{\mathcal{E}}$.

In [36], H. Gillet and C. Soulé attached to the elements of $\widehat{\mathrm{K}}_0(\mathcal{X})$, represented by hermitian vector bundles $\overline{E} = (E,h)$, arithmetic characteristic classes $\widehat{\phi}(\overline{E})$, which lie in the "classical" arithmetic Chow ring $\widehat{\mathrm{CH}}^*(\mathcal{X})_{\mathbb{Q}}$. A particular example of such an arithmetic characteristic class is the arithmetic Chern character $\widehat{\mathrm{ch}}(\overline{E})$, whose definition also involves the Bott-Chern form $\widetilde{\mathrm{ch}}(\overline{\mathcal{E}})$.

In order to be able to carry over the concept of arithmetic characteristic classes to the category of vector bundles E over an arithmetic variety $\mathcal{X}$ equipped with a hermitian metric h having singularities of the type considered in this paper, we proceed as follows: Letting h_0 denote an arbitrary smooth hermitian metric on E, we have the obvious short exact sequence of vector bundles

$$\overline{\mathcal{E}}:\ 0 \longrightarrow 0 \longrightarrow (E,h) \longrightarrow (E,h_0) \longrightarrow 0,$$

to which is attached the Bott-Chern form $\widetilde{\phi}(\overline{\mathcal{E}})$ being no longer smooth, but having certain singularities. Formally, we then set

$$\widehat{\phi}(E,h) := \widehat{\phi}(E,h_0) + \mathrm{a}\left(\widetilde{\phi}(\overline{\mathcal{E}})\right),$$

where a is the morphism mapping differential forms into arithmetic Chow groups. In order to give meaning to this definition, we need to know the singularities of $\widetilde{\phi}(\overline{\mathcal{E}})$; moreover, we have to show the independence of the (arbitrarily chosen) smooth hermitian metric h_0.

Once we can control the singularities of $\widetilde{\phi}(\overline{\mathcal{E}})$, the abstract formalism developed in [10] reduces our task to find a cohomological complex $\mathcal{C}$, which contains the elements $\widetilde{\phi}(\overline{\mathcal{E}})$, and has all the properties we desire for a reasonable arithmetic intersection theory. Once the complex $\mathcal{C}$ is constructed, we obtain an arithmetic K-theory with properties depending on the complex $\mathcal{C}$, of course.

The most naive way to construct an arithmetic intersection theory for automorphic hermitian vector bundles would be to only work with good forms and good metrics. This procedure is doomed to failure for the following two reasons: First, the complex of good forms is not a Dolbeault complex. However, this first problem can be easily solved by imposing that it is also closed under the differential operators ∂, $\bar{\partial}$, and $\partial\bar{\partial}$. The second problem is that the complex of good forms is not big enough to contain the singular Bott-Chern forms which occur. For example, if $\mathcal{L}$ is a line bundle, h_0 a smooth hermitian metric and h a singular metric, which is good along a divisor D (locally, in some open coordinate neighborhood, given by the equation $z = 0$), the Bott-Chern form (associated to the first Chern class) $\widetilde{\mathrm{c}}_1(\mathcal{L}; h, h_0)$ encoding the change of metrics grows like $\log\log(1/|z|)$, whereas the good functions are bounded.

The solution of these problems led us to consider the $\mathcal{D}_{\log}$-complexes $\mathcal{D}_{\mathrm{pre}}$ made by pre-log and pre-log-log forms and its subcomplex $\mathcal{D}_{l,ll}$ consisting of log and log-log forms. We emphasize that neither the complex of good forms nor the complex of pre-log-log forms are contained in each other. We also note that if one is interested in arithmetic intersection numbers, the results obtained by both theories agree.

Discussion of results. The $\mathcal{D}_{\log}$-complex $\mathcal{D}_{\mathrm{pre}}$ made out of pre-log and pre-log-log forms could be seen as the complex that satisfies the minimal requirements needed to allow log-log singularities along a fixed divisor as well as to have an arithmetic intersection theory with arithmetic intersection numbers in the proper case (see [10]). As we will show in theorem 4.55, the Bott-Chern forms associated to the change of metrics between a smooth hermitian metric and a good metric belong to the complex of pre-log-log forms. Therefore, we can define arithmetic characteristic classes of good hermitian vector bundles in the arithmetic Chow groups with pre-log-log forms. If our arithmetic variety is proper, we can use this theory to calculate arithmetic Chern numbers of automorphic hermitian vector bundles of arbitrary rank. However, the main disadvantage of $\mathcal{D}_{\mathrm{pre}}$ is that we do not know the size of the associated cohomology groups.

The $\mathcal{D}_{\log}$-complex $\mathcal{D}_{l,ll}$ made out of log and log-log forms is a subcomplex of $\mathcal{D}_{\mathrm{pre}}$. The main difference is that *all* the derivatives of the component functions of the log and log-log forms have to be bounded, which allows us to use an inductive argument to prove a Poincaré lemma, which implies that the associated Deligne complex computes the usual Deligne-Beilinson cohomology (see theorem 2.42). For this reason we have better understanding of the arithmetic Chow groups with log-log forms (see theorem 3.17).

Since a good form is in general not a log-log form, it is not true that the Chern forms for a good hermitian vector bundle are log-log forms. Hence, we introduce the notion of log-singular hermitian metrics, which have, roughly speaking, the same relation to log-log forms as the good hermitian metrics to good forms. We then show that the Bott-Chern forms associated to the change of metrics between smooth hermitian metrics and log-singular hermitian metrics are log-log forms. As a consequence, we can define the Bott-Chern forms for short exact sequences of vector bundles equipped with log-singular hermitian metrics. These Bott-Chern forms have an axiomatic characterization similar to the Bott-Chern forms for short exact sequences of vector bundles equipped with smooth hermitian metrics. The Bott-Chern forms are the main ingredients in order to extend the theory of arithmetic characteristic classes to log-singular hermitian vector bundles.

The price we have to pay in order to use log-log forms is that it is more difficult to prove that a particular form is log-log: we have to bound all derivatives. Note however that most pre-log-log forms which appear are also log-log forms (see for instance section 6). On the other hand, we point out that the theory of log-singular hermitian vector bundles is not optimal for several other reasons. The most important one is that it is not closed under taking sub-objects, quotients and extensions. For example, let

$$0 \longrightarrow (E', h') \longrightarrow (E, h) \longrightarrow (E'', h'') \longrightarrow 0$$

be a short exact sequence of hermitian vector bundles such that the metrics h' and h'' are induced by h. Then, the assumption that h is a log-singular hermitian metric does not imply that the hermitian metrics h' and h'' are log-singular, and vice versa. In particular, automorphic hermitian vector bundles that are not fully decomposed can always be written as successive extensions of fully decomposed automorphic hermitian vector bundles, whose metrics are in general not log-singular. A related question is that the hermitian metric of a unipotent variation of polarized Hodge structures induced by the polarization is in general not log-singular. These considerations suggest that one should further enlarge the notion of log-singular hermitian metrics.

Since the hermitian vector bundles defined on a quasi-projective variety may have arbitrary singularities at infinity, we also consider differential forms with arbitrary singularities along a normal crossing divisor. Using these kinds of differential forms we are able to recover the arithmetic Chow groups à la Gillet, Soulé for quasi-projective varieties.

Finally, another technical difference between this paper and [10] is the fact that in the previous paper the complex $\mathcal{D}_{\log}(X,p)$ is defined by applying the Deligne complex construction to the Zariski sheaf $E_{\log}$, which, in turn, is defined as the Zariski sheaf associated to the pre-sheaf $E^{\circ}_{\log}$. In theorem 3.6, we prove that the pre-sheaf $E^{\circ}_{\log}$ is already a sheaf, which makes it superfluous to take the associated sheaf. Moreover, the proof is purely geometric and can be applied to other similar complexes like $\mathcal{D}_{\mathrm{pre}}$ or $\mathcal{D}_{l,ll}$.

Outline of paper. The set-up of the paper is as follows. In section 2, we introduce several complexes of singular differential forms and discuss their relationship. Of particular importance are the complexes of log and log-log forms for which we prove a Poincaré lemma allowing us to characterize their cohomology by means of their Hodge filtration. In section 3, we introduce and study arithmetic Chow groups with differential forms which are log-log along a fixed normal crossing divisor D. We also consider differential forms having arbitrary singularities at infinity; in particular, we prove that for D being the empty set, the arithmetic Chow groups defined by Gillet, Soulé are recovered. In section 5, we discuss several classes of singular hermitian metrics; we prove that the Bott-Chern forms associated to the change of metrics between a smooth hermitian metric and a log-singular hermitian metric are log-log forms. We also show that the Bott-Chern forms associated to the change of metrics between a smooth hermitian metric and a good hermitian metric are pre-log-log. This allows us to define arithmetic characteristic classes of log-singular hermitian vector bundles. Finally, in section 6, after having given a brief recollection of the basics of Shimura varieties, we prove that the fully decomposed automorphic vector bundles equipped with an equivariant hermitian metric are log-singular hermitian vector bundles. In this respect many examples are provided to which the theory developed in this paper can be applied.

Acknowledgements: In course of preparing this manuscript, we had many stimulating discussions with many colleagues. We would like to thank them all. In particular, we would like to express our gratitude to J.-B. Bost, J. Bruinier, P. Guillen, W. Gubler, M. Harris, S. Kudla, V. Maillot, D. Roessler, C. Soulé, J. Wildeshaus. Furthermore, we would like to thank EAGER, the Arithmetic Geometry Network, the Newton Institute (Cambridge), and the Institut Henri Poincaré (Paris) for partial support of our work.

2 Log and log-log differential forms

In this section, we will introduce several complexes of differential forms with singularities along a normal crossing divisor D, and we will discuss their basic properties.

The first one $\mathscr{E}_X^*\langle D\rangle$ is a complex with logarithmic growth conditions in the spirit of [22]. Unlike in [22], where the authors consider only differential forms of type $(0,q)$, we consider here the whole Dolbeault complex and we show that it is an acyclic resolution of the complex of holomorphic forms with logarithmic poles along the normal crossing divisor D, i.e., this complex computes the cohomology of the complement of D. Another difference with [22] is that, in order to be able to prove the Poincaré lemma for such forms, we need to impose growth conditions to all derivatives of the functions. Note that a similar condition has been already considered in [24].

The second complex $\mathscr{E}_X^*\langle\langle D\rangle\rangle$ contains differential forms with singularities of log-log type along a normal crossing divisor D, and is related with the complex of good forms in the sense of [34]. As the complex of good forms, it contains the Chern forms for fully decomposed automorphic hermitian vector bundles and is functorial with respect to certain inverse images. Moreover all the differential forms belonging to this complex are locally integrable with zero residue. The new property of this complex is that it satisfies a Poincaré lemma that implies that this complex is quasi-isomorphic to the complex of smooth differential forms, i.e., this complex computes the cohomology of the whole variety. The main interest of this complex, as we shall see in subsequent sections, is that it contains also the Bott-Chern forms associated to fully decomposed automorphic vector bundles. Note that neither the complex of good forms in the sense of [34] nor the complex of log-log forms are contained in each other.

The third complex $\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle$ that we will introduce is a mixture of the previous complexes. It is formed by differential forms which are log along a normal crossing divisor D_1 and log-log along another normal crossing divisor D_2. This complex computes the cohomology of the complement of D_1.

By technical reasons we will introduce several other complexes.

2.1 Log forms

General notations. Let X be a complex manifold of dimension d. We will denote by $\mathscr{E}_X^*$ the sheaf of complex smooth differential forms over X.

Let D be a normal crossing divisor on X. Let V be an open coordinate subset of X with coordinates $z_1,\dots,z_d$; we put $r_i=|z_i|$. We will say that V is adapted to D, if the divisor $D\cap V$ is given by the equation $z_1\cdots z_k=0$, and the coordinate neighborhood V is small enough; more precisely, we will assume that all the coordinates satisfy $r_i\leq 1/e^e$, which implies that $\log(1/r_i)>e$ and $\log(\log(1/r_i))>1$.

We will denote by $\Delta_r\subseteq\mathbb{C}$ the open disk of radius r centered at 0, by $\overline{\Delta}_r$ the

closed disk, and we will write $\Delta_r^* = \Delta_r \setminus \{0\}$ and $\overline{\Delta}_r^* = \overline{\Delta}_r \setminus \{0\}$.
If f and g are two functions with non-negative real values, we write $f \prec g$, if there exists a real constant $C > 0$ such that $f(x) \leq C \cdot g(x)$ for all x in the domain of definition under consideration.

MULTI-INDICES. We collect here all the conventions we will use about multi-indices.

NOTATION 2.1. For any multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{Z}_{\geq 0}^d$, we write

$$|\alpha| = \sum_{i=1}^{d} \alpha_i, \qquad z^\alpha = \prod_{i=1}^{d} z_i^{\alpha_i}, \qquad \bar{z}^\alpha = \prod_{i=1}^{d} \bar{z}_i^{\alpha_i},$$
$$r^\alpha = \prod_{i=1}^{d} r_i^{\alpha_i}, \qquad (\log(1/r))^\alpha = \prod_{i=1}^{d} (\log(1/r_i))^{\alpha_i},$$
$$\frac{\partial^{|\alpha|}}{\partial z^\alpha} f = \frac{\partial^{|\alpha|}}{\prod_{i=1}^{d} \partial z_i^{\alpha_i}} f, \qquad \frac{\partial^{|\alpha|}}{\partial \bar{z}^\alpha} f = \frac{\partial^{|\alpha|}}{\prod_{i=1}^{d} \partial \bar{z}_i^{\alpha_i}} f.$$

If α and β are multi-indices, we write $\beta \geq \alpha$, if, for all $i = 1, \dots, d$, $\beta_i \geq \alpha_i$. We denote by $\alpha + \beta$ the multi-index with components $\alpha_i + \beta_i$. If $1 \leq i \leq d$, we will denote by γ^i the multi-index with all the entries zero except the i-th entry that takes the value 1. More generally, if I is a subset of $\{1, \dots, d\}$, we will denote by γ^I the multi-index

$$\gamma_i^I = \begin{cases} 1, \&i \in I, \\ 0, \&i \notin I. \end{cases}$$

We will denote by $\underline{n}$ the constant multi-index

$$\underline{n}_i = n.$$

In particular, $\underline{0}$ is the multi-index $\underline{0} = (0, \dots, 0)$.
If α is a multi-index and $k \geq 1$ is an integer, we will denote by $\alpha^{\leq k}$ the multi-index

$$\alpha_i^{\leq k} = \begin{cases} \alpha_i, & i \leq k, \\ 0, \&i > k. \end{cases}$$

For a multi-index α, the order function associated to α,

$$\Phi_\alpha : \{1, \dots, |\alpha|\} \longrightarrow \{1, \dots, d\}$$

is given by

$$\Phi_\alpha(i) = k, \text{ if } \sum_{j=1}^{k-1} \alpha_j < i \leq \sum_{j=1}^{k} \alpha_j.$$

Log forms. We introduce now a complex of differential forms with logarithmic growth along a normal crossing divisor. This complex can be used to compute the cohomology of a non-compact algebraic complex manifold with its usual Hodge filtration. It contains the C^∞ logarithmic Dolbeault complex defined in [7], but it is much bigger and, in particular, it contains also the log-log differential forms defined later. In contrast to the pre-log forms introduced in [10], in the definition given here we impose growth conditions to the differential forms and to all their derivatives.

The problem of the weight filtration of the complex of log forms will not be treated here.

Let X be a complex manifold of dimension d, D a normal crossing divisor, $U = X \setminus D$, and $\iota : U \longrightarrow X$ the inclusion.

Definition 2.2. Let V be a coordinate neighborhood adapted to D. For every integer $K \geq 0$, we say that a smooth complex function f on $V \setminus D$ has *logarithmic growth along D of order K*, if there exists an integer N_K such that, for every pair of multi-indices $\alpha, \beta \in \mathbb{Z}^d_{\geq 0}$ with $|\alpha + \beta| \leq K$, it holds the inequality

$$\left| \frac{\partial^{|\alpha|}}{\partial z^\alpha} \frac{\partial^{|\beta|}}{\partial \bar{z}^\beta} f(z_1, \dots, z_d) \right| \prec \frac{\left| \prod_{i=1}^k \log(1/r_i) \right|^{N_K}}{|z^{\alpha^{\leq k}} \bar{z}^{\beta^{\leq k}}|}. \tag{2.3}$$

We say that f has *logarithmic growth along D of infinite order*, if it has logarithmic growth along D of order K for all $K \geq 0$. The *sheaf of differential forms on X with logarithmic growth of infinite order along D*, denoted by $\mathscr{E}^*_X\langle D\rangle$, is the subalgebra of $\iota_*\mathscr{E}^*_U$ generated, in each coordinate neighborhood adapted to D, by the functions with logarithmic growth of infinite order along D and the differentials

$$\begin{aligned} &\frac{\mathrm{d}\, z_i}{z_i}, \frac{\mathrm{d}\, \bar{z}_i}{\bar{z}_i}, &&\text{for } i = 1, \dots, k, \\ &\mathrm{d}\, z_i, \mathrm{d}\, \bar{z}_i, &&\text{for } i = k+1, \dots, d. \end{aligned} \tag{2.4}$$

As a shorthand, a differential form with logarithmic growth of infinite order along D is called log along D or, if D is understood, a log form.

The Dolbeault algebra of log forms. The sheaf $\mathscr{E}^*_X\langle D\rangle$ inherits from $\iota_*\mathscr{E}^*_U$ a real structure and a bigrading. Moreover, it is clear that, if ω is a log form, then $\partial\omega$ and $\bar{\partial}\omega$ are also log forms. Therefore, $\mathscr{E}^*_X\langle D\rangle$ is a sheaf of Dolbeault algebras. We will use all the notations of [10], §5, concerning Dolbeault algebras. For the convenience of the reader we will recall these notations in section 3.1. In particular, from the structure of Dolbeault algebra, there is a well defined Hodge filtration denoted by F.

Pre-log forms. Recall that, in [10], section 7.2, there is introduced the sheaf of pre-log forms denoted $\mathscr{E}^*_X\langle D\rangle_{\mathrm{pre}}$. It is clear that there is an inclusion of sheaves

$$\mathscr{E}^*_X\langle D\rangle \subseteq \mathscr{E}^*_X\langle D\rangle_{\mathrm{pre}}.$$

The cohomology of the complex of log forms. Let $\Omega^*_X(\log D)$ be the sheaf of holomorphic forms with logarithmic poles along D (see [12]). Then, the more general theorem 2.42 implies

Theorem 2.5. *The inclusion*

$$\Omega^*_X(\log D) \longrightarrow \mathscr{E}^*_X\langle D\rangle$$

is a filtered quasi-isomorphism with respect to the Hodge filtration.

In other words, this complex is a resolution of the sheaf of holomorphic forms with logarithmic poles along D, $\Omega^*_X(\log D)$. Thus, if X is a compact Kähler manifold, the complex of global sections $\Gamma(X, \mathscr{E}^*_X\langle D\rangle)$ computes the cohomology of the open complex manifold $U = X \setminus D$ with its Hodge filtration.
Note that corollary 2.5 implies that there is an isomorphism in the derived category $R\iota_*\underline{\mathbb{C}}_U \longrightarrow \mathscr{E}^*_X\langle D\rangle$. This isomorphism is compatible with the real structures. Hence, the complex $\mathscr{E}^*_X\langle D\rangle$ also provides the real structure of the cohomology of U.

Inverse images. The complex of log forms is functorial with respect to inverse images. More precisely, we have the following result.

Proposition 2.6. *Let $f : X \longrightarrow Y$ be a morphism of complex manifolds of dimension d and d'. Let D_X, D_Y be normal crossing divisors on X, Y, respectively, satisfying $f^{-1}(D_Y) \subseteq D_X$. If η is a section of $\mathscr{E}^*_Y\langle D_Y\rangle$, then $f^*\eta$ is a section of $\mathscr{E}^*_X\langle D_X\rangle$.*

Proof. Let p be a point of X. Let V and W be open coordinate neighborhoods of p and $f(p)$, respectively, adapted to D_X and D_Y, and such that $f(V) \subseteq W$. Let k and k' be the number of components of $V \cap D_X$ and $W \cap D_Y$, respectively. Then, the condition $f^{-1}(D_Y) \subseteq D_X$ implies that f can be written as

$$f(x_1, \dots, x_d) = (z_1, \dots, z_{d'}) \tag{2.7}$$

with

$$z_i = \begin{cases} x_1^{a_{i,1}} \cdots x_k^{a_{i,k}} u_i, & \text{if } i \leq k', \\ w_i, & \text{if } i > k', \end{cases}$$

where $u_1, \dots, u_{k'}$ are holomorphic functions that do not vanish in V, the $a_{i,j}$ are non negative integers and $w_{k'+1}, \dots, w_{d'}$ are holomorphic functions. Shrinking V, if necessary, we may assume that the functions u_j are holomorphic and do not vanish in a neighborhood of the adherence of V.
For $1 \leq i \leq k'$, we have

$$f^*\left(\frac{\mathrm{d}\, z_i}{z_i}\right) = \sum_{j=1}^{k} a_{i,j} \frac{\mathrm{d}\, x_j}{x_j} + \frac{\mathrm{d}\, u_i}{u_i}.$$

Since the function $1/u_i$ is holomorphic in a neighborhood of the adherence of V, the function $1/u_i$ and all its derivatives are bounded. If follows that $f^*(\mathrm{d}\, z_i/z_i)$

is a log form (along D_X). The same argument shows that $f^*(\mathrm{d}\,\bar{z}_i/\bar{z}_i)$ is a log form.

If a function g on W satisfies

$$|g(z_1,\dots,z_{d'})| \prec \left|\prod_{i=1}^{k'} \log(1/|z_i|)\right|^N,$$

then f^*g satisfies

$$\begin{aligned}|f^*g(x_1,\dots,x_d)| &\prec \left|\prod_{i=1}^{k'}\left(\sum_{j=1}^{k} a_{i,j}\log(1/|x_j|)+\log(1/|u_i|)\right)\right|^N\\ &\prec \left|\prod_{j=1}^{k}\log(1/|x_j|)\right|^{Nk'}.\end{aligned}$$

Therefore, f^*g has logarithmic growth. It remains to bound the derivatives of f^*g. To ease notation, we will bound only the derivatives with respect to the holomorphic coordinates, the general case being analogous.

For any multi-index $\alpha \in \mathbb{Z}^d_{\geq 0}$, the function $\partial^{|\alpha|}/\partial x^{\alpha}(f^*g)$ is a linear combination of the functions

$$\left\{\frac{\partial^{|\beta|}}{\partial z^{\beta}}g\prod_{i=1}^{|\beta|}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_{\Phi_\beta(i)}\right\}_{\beta,\{\alpha_i\}}, \tag{2.8}$$

where β runs over all multi-indices $\beta \in \mathbb{Z}^{d'}_{\geq 0}$ such that $|\beta| \leq |\alpha|$, and $\{\alpha_i\}$ runs over all families of multi-indices $\alpha^i \in \mathbb{Z}^d_{\geq 0}$ such that

$$\sum_{i=1}^{|\beta|}\alpha^i = \alpha.$$

The function Φ_α is the order function introduced in 2.1.

Then, since g is a log function,

$$\begin{aligned}\left|\frac{\partial^{|\beta|}}{\partial z^{\beta}}g\prod_{i=1}^{|\beta|}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_{\Phi_\beta(i)}\right| &\prec \frac{\left|\prod_{j=1}^{k'}\log(1/|z_j|)\right|^{N_{|\beta|}}}{|z^{\beta^{\leq k'}}|}\left|\prod_{i=1}^{|\beta|}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_{\Phi_\beta(i)}\right|\\ &\prec \left|\prod_{j=1}^{k}\log(1/|x_j|)\right|^{N_{|\beta|}k'}\prod_{i=1}^{|\beta^{\leq k'}|}\left|\frac{1}{z_{\Phi_\beta(i)}}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_{\Phi_\beta(i)}\right|.\end{aligned}$$

But, by the assumption on the map f, it is easy to see that, for $1 \leq j \leq k'$, we have

$$\left|\frac{1}{z_j}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_j\right| \prec \frac{1}{|x^{(\alpha^i)^{\leq k}}|}, \tag{2.9}$$

which implies the proposition. □

Polynomial growth in the local universal cover. We can characterize log forms as differential forms that have polynomial growth in a local universal cover. Let $M > 1$ be a real number and let $U_M \subseteq \mathbb{C}$ be the subset given by

$$U_M = \{x \in \mathbb{C} \mid \operatorname{Im} x > M\}.$$

Let K be an open subset of $\mathbb{C}^{d-k}$. We consider the space $(U_M)^k \times K$ with coordinates $(x_1, \dots, x_d)$.

Definition 2.10. A function f on $(U_M)^k \times K$ is said to have *imaginary polynomial growth*, if there is a sequence of integers $\{N_n\}_{n\geq 0}$ such that for every pair of multi-indices $\alpha, \beta \in \mathbb{Z}^d_{\geq 0}$, the inequality

$$\left|\frac{\partial^{|\alpha|}}{\partial x^{\alpha}} \frac{\partial^{|\beta|}}{\partial \bar{x}^{\beta}} f(x_1, \dots, x_d)\right| \prec \left|\prod_{i=1}^{k} \operatorname{Im} x_i\right|^{N_{|\alpha|+|\beta|}} \tag{2.11}$$

holds. The space of differential forms on $(U_M)^k \times K$ with *imaginary polynomial growth* is generated by the functions with imaginary polynomial growth and the differentials

$$\mathrm{d}\, x_i,\ \mathrm{d}\, \bar{x}_i, \qquad \text{for } i = 1, \dots, d.$$

Let X, D, U, and ι be as in definition 2.2.

Definition 2.12. Let W be an open subset of X and $\omega \in \Gamma(W, \iota_*(\mathscr{E}^*_U))$ be a differential form. For every point $p \in W$, there is an open coordinate neighborhood $V \subseteq W$, which is adapted to D and such that the coordinates of V induce an identification $V \cap U = (\Delta_r^*)^k \times K$. We choose $M > \log(1/r)$ and denote by $\pi : (U_M)^k \times K \longrightarrow V$ the covering map given by

$$\pi(x_1, \dots, x_d) = (e^{2\pi i x_1}, \dots, e^{2\pi i x_k}, x_{k+1}, \dots, x_d).$$

We say that ω has *polynomial growth in the local universal cover*, if for every V and π as above, $\pi^*\omega$ has imaginary polynomial growth.

It is easy to see that the differential forms with polynomial growth in the local universal cover form a sheaf of Dolbeault algebras.

Theorem 2.13. *A differential form has polynomial growth in the local universal cover, if and only if, it is a log form.*

Proof. We start with the case of a function. So let f be a function with polynomial growth in the local universal cover and let V be a coordinate neighborhood as in definition 2.12. Let $g = \pi^* f$. By definition, g satisfies

$$g(\dots, x_i + 1, \dots) = g(\dots, x_i, \dots), \qquad \text{for } 1 \leq i \leq k. \tag{2.14}$$

We write formally

$$f(z_1,\dots,z_d)=g(x_1(z_1),\dots,x_d(z_d))$$

with

$$x_i(z_i)=\begin{cases}\frac{1}{2\pi i}\log z_i, &\text{for } i\le k,\\ z_i, &\text{for } i>k.\end{cases}$$

Note that this makes sense because of the periodicity properties (2.14). Then, we have

$$\frac{\partial^{|\alpha|}}{\partial z^{\alpha}}\frac{\partial^{|\beta|}}{\partial \bar{z}^{\beta}}f(z_1,\dots,z_d)=\sum_{\substack{\alpha'\le\alpha\\ \beta'\le\beta}}C^{\alpha',\beta'}_{\alpha,\beta}\frac{\partial^{|\alpha'|}}{\partial x^{\alpha'}}\frac{\partial^{|\beta'|}}{\partial \bar{x}^{\beta'}}g(x_1,\dots,x_d)\cdot$$
$$\cdot\frac{\partial^{|\alpha-\alpha'|}}{\partial z^{\alpha-\alpha'}}\left(\frac{\partial x}{\partial z}\right)^{\alpha'}\frac{\partial^{|\beta-\beta'|}}{\partial \bar{z}^{\beta-\beta'}}\left(\frac{\partial \bar{x}}{\partial \bar{z}}\right)^{\beta'}, \quad (2.15)$$

for certain constants $C^{\alpha',\beta'}_{\alpha,\beta}$. But the estimates

$$\left|\frac{\partial^{|\alpha'|}}{\partial x^{\alpha'}}\frac{\partial^{|\beta'|}}{\partial \bar{x}^{\beta'}}g(x_1,\dots,x_d)\right|\prec\left|\prod_{i=1}^{k}|x_i|\right|^{N_{\alpha',\beta'}}\prec\left|\prod_{i=1}^{k}\log(1/|z_i|)\right|^{N_{\alpha',\beta'}}$$

and

$$\frac{\partial^{|\alpha-\alpha'|}}{\partial z^{\alpha-\alpha'}}\left(\frac{\partial x}{\partial z}\right)^{\alpha'}\frac{\partial^{|\beta-\beta'|}}{\partial \bar{z}^{\beta-\beta'}}\left(\frac{\partial \bar{x}}{\partial \bar{z}}\right)^{\beta'}\prec\frac{1}{|z^{\alpha^{\le k}}\bar{z}^{\beta^{\le k}}|} \quad (2.16)$$

imply the bounds of f and its derivatives. The converse is proven in the same way.

To prove the theorem for differential forms, observe that, for $1\le i\le k$,

$$\pi^*\left(\frac{\mathrm{d}\,z_i}{z_i}\right)=2\pi i\,\mathrm{d}\,x_i.$$

□

2.2 Log-log forms

log-log growth forms. Let X, D, U, and ι be as in definition 2.2.

Definition 2.17. Let V be a coordinate neighborhood adapted to D. For every integer $K\ge 0$, we say that a smooth complex function f on $V\setminus D$ has *log-log growth along D of order K*, if there exists an integer N_K such that, for every pair of multi-indices $\alpha,\beta\in\mathbb{Z}^d_{\ge 0}$ with $|\alpha+\beta|\le K$, it holds the inequality

$$\left|\frac{\partial^{|\alpha|}}{\partial z^{\alpha}}\frac{\partial^{|\beta|}}{\partial \bar{z}^{\beta}}f(z_1,\dots,z_d)\right|\prec\frac{\left|\prod_{i=1}^{k}\log(\log(1/r_i))\right|^{N_K}}{|z^{\alpha^{\le k}}\bar{z}^{\beta^{\le k}}|}. \quad (2.18)$$

We say that f has *log-log growth along D of infinite order*, if it has log-log growth along D of order K for all $K \geq 0$. The *sheaf of differential forms on X with log-log growth along D of infinite order* is the subalgebra of $\iota_* \mathscr{E}_U^*$ generated, in each coordinate neighborhood V adapted to D, by the functions with log-log growth along D and the differentials

$$\begin{aligned} &\frac{\mathrm{d}\, z_i}{z_i \log(1/r_i)}, \frac{\mathrm{d}\, \bar{z}_i}{\bar{z}_i \log(1/r_i)}, && \text{for } i = 1, \dots, k, \\ &\mathrm{d}\, z_i, \mathrm{d}\, \bar{z}_i, && \text{for } i = k+1, \dots, d. \end{aligned}$$

A differential form with log-log growth along D of infinite order will be called a *log-log growth form*. The sheaf of differential forms on X with log-log growth along D of infinite order will be denoted by $\mathscr{E}_X^*\langle\langle D \rangle\rangle_{\mathrm{gth}}$.

The following characterization of differential forms with log-log growth of infinite order is left to the reader.

Lemma 2.19. *Let V be an open coordinate subset adapted to D and let I, J be two subsets of $\{1, \dots, d\}$. Then, the form $f \,\mathrm{d}\, z_I \wedge \mathrm{d}\, \bar{z}_J$ is a log-log growth form of infinite order, if and only if, for every pair of multi-indices $\alpha, \beta \in \mathbb{Z}_{\geq 0}^d$, there is an integer $N_{\alpha,\beta} \geq 0$ such that*

$$\left| \frac{\partial^{|\alpha|}}{\partial z^\alpha} \frac{\partial^{|\beta|}}{\partial \bar{z}^\beta} f(z_1, \dots, z_d) \right| \prec \frac{\left| \prod_{i=1}^k \log(\log(1/r_i)) \right|^{N_{\alpha,\beta}}}{r^{(\gamma^I + \gamma^J + \alpha + \beta)^{\leq k}} (\log(1/r))^{(\gamma^I + \gamma^J)^{\leq k}}}. \tag{2.20}$$

□

Definition 2.21. A function that satisfies the bound (2.20) for any pair of multi-indices α, β with $\alpha + \beta \leq K$ will be called a *(I, J)-log-log growth function of order K*. If it satisfies the bound (2.20) for any pair multi-indices α, β, it will be called a *(I, J)-log-log growth function of infinite order.*

Log-log forms. Unlike the case of log growth forms, the fact that ω is a log-log growth form does not imply that its differential $\partial\omega$ is a log-log growth form.

Definition 2.22. We say that a smooth complex differential form ω is *log-log along D*, if the differential forms ω, $\partial\omega$, $\bar{\partial}\omega$, and $\partial\bar{\partial}\omega$ have log-log growth along D of infinite order. The sheaf of differential forms log-log along D will be denoted by $\mathscr{E}_X^*\langle\langle D \rangle\rangle$. As a shorthand, if D is clear from the context, a differential form which is log-log along D, will be called a log-log form.

From the definition, it is clear that the sheaf of log-log forms is contained in the sheaf of log forms.

Let V be a coordinate subset adapted to D. For $i = 1, \dots, k$, the function $\log(\log(1/r_i))$ is a log-log function and the differential forms

$$\frac{\mathrm{d}\, z_i}{z_i \log(1/r_i)}, \frac{\mathrm{d}\, \bar{z}_i}{\bar{z}_i \log(1/r_i)}, \qquad \text{for } i = 1, \dots, k,$$

are log-log forms.

The Dolbeault algebra of log-log forms. As in the case of log forms, the sheaf $\mathscr{E}^*_X\langle\langle D\rangle\rangle$ inherits a real structure and a bigrading. Moreover, we have forced the existence of operators ∂ and $\bar{\partial}$. Therefore, $\mathscr{E}^*_X\langle\langle D\rangle\rangle$ is a sheaf of Dolbeault algebras (see section 3.1). In particular, there is a well defined Hodge filtration, denoted by F.

Pre-log-log forms Recall that, in [10], section 7.1, there is introduced the sheaf of pre-log-log forms, denoted by $\mathscr{E}^*_X\langle\langle D\rangle\rangle_{\text{pre}}$. It is clear that there is an inclusion of sheaves

$$\mathscr{E}^*_X\langle\langle D\rangle\rangle \subseteq \mathscr{E}^*_X\langle\langle D\rangle\rangle_{\text{pre}}.$$

The cohomology of the complex of log-log differential forms. Let Ω^*_X be the sheaf of holomorphic forms. Then, theorem 2.42, which will be proved later, implies

Theorem 2.23. *The inclusion*

$$\Omega^*_X \longrightarrow \mathscr{E}^*_X\langle\langle D\rangle\rangle$$

is a filtered quasi-isomorphism with respect to the Hodge filtration.

In other words, this complex is a resolution of Ω^*_X, the sheaf of holomorphic differential forms on X. Therefore, if X is a compact Kähler manifold, the complex of global sections $\Gamma(X, \mathscr{E}^*_X\langle\langle D\rangle\rangle)$ computes the cohomology of X with its Hodge filtration. As in the case of log forms it also provides the usual real structure of the cohomology of X. One may say that the singularities of the log-log complex are so mild that they do not change the cohomology.

Inverse images. As in the case of pre-log-log forms, the sheaf of log-log forms is functorial with respect to inverse images. More precisely, we have the following result.

Proposition 2.24. *Let $f : X \longrightarrow Y$ be a morphism of complex manifolds of dimension d and d'. Let D_X, D_Y be normal crossing divisors on X, Y, respectively, satisfying $f^{-1}(D_Y) \subseteq D_X$. If η is a section of $\mathscr{E}^*_Y\langle\langle D_Y\rangle\rangle$, then $f^*\eta$ is a section of $\mathscr{E}^*_X\langle\langle D_X\rangle\rangle$.*

Proof. Since the differential operators ∂ and $\bar{\partial}$ are compatible with inverse images, we have to show that the pre-image of a form with log-log growth of infinite order has log-log growth of infinite order. We may assume that, locally, f can be written as in equation (2.7). If a function g satisfies

$$|g(z_1, \dots, z_{d'})| \prec \left|\prod_{i=1}^{k'} \log(\log(1/|z_i|))\right|^N,$$

we then estimate

$$\begin{aligned}
|(f^*g)(x_1,\dots,x_d)| &\prec \left|\prod_{i=1}^{k'} f^*\log(\log(1/|z_i|))\right|^N\\
&\prec \left|\prod_{i=1}^{k'}\sum_{j=1}^{k}\log(\log(1/|x_j|))\right|^N\\
&\prec \left|\sum_{j=1}^{k}\log(\log(1/|x_j|))\right|^{Nk'}.
\end{aligned}$$

Therefore, f^*g has log-log growth.
Next we have to bound the derivatives of f^*g. As in the proof of proposition 2.6, we will bound only the derivatives with respect to the holomorphic coordinates. Again, we observe that, for any multi-index $\alpha \in \mathbb{Z}^d_{\geq 0}$, the function $\partial^{|\alpha|}/\partial x^\alpha(f^*g)$ is a linear combination of the functions (2.8). But, using that g is a log-log growth function, we can further estimate

$$\begin{aligned}
\left|\frac{\partial^{|\beta|}}{\partial z^\beta}g\prod_{i=1}^{|\beta|}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_{\Phi_\beta(i)}\right| &\prec \frac{\left|\prod_{j=1}^{k'}\log(\log(1/|z_j|))\right|^{N_{|\beta|}}}{|z^{\beta^{\leq k'}}|}\left|\prod_{i=1}^{|\beta|}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_{\Phi_\beta(i)}\right|\\
&\prec \left|\prod_{j=1}^{k}\log(\log(1/|x_j|))\right|^{N_{|\beta|}k'}\prod_{i=1}^{|\beta^{\leq k'}|}\left|\frac{1}{z_{\Phi_\beta(i)}}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_{\Phi_\beta(i)}\right|.
\end{aligned}$$

By (2.9), this yields

$$\left|\frac{\partial^{|\beta|}}{\partial z^\beta}g\prod_{i=1}^{|\beta|}\frac{\partial^{|\alpha^i|}}{\partial x^{\alpha^i}}z_{\Phi_\beta(i)}\right| \prec \left|\prod_{j=1}^{k}\log(\log(1/|x_j|))\right|^{N_{|\beta|}k'}\prod_{i=1}^{|\beta^{\leq k'}|}\frac{1}{|x^{\alpha^{\leq k}}|}.$$

Thus, f^*g has log-log growth of infinite order.
Finally, we are led to study the inverse image of the differential forms

$$\frac{\mathrm{d}\, z_i}{z_i\log(1/|z_i|)},\ \frac{\mathrm{d}\,\bar z_i}{\bar z_i\log(1/|z_i|)},\ \text{for } i=1,\dots,k'.$$

We have

$$f^*\left(\frac{\mathrm{d}\, z_i}{z_i\log(1/z_i\bar z_i)}\right)=\frac{1}{\log(1/z_i\bar z_i)}\left(\sum_{i=j}^{k}a_{i,j}\frac{\mathrm{d}\, x_j}{x_j}+\frac{\mathrm{d}\, u_i}{u_i}\right).$$

Since we have assumed that u_i is a non-vanishing holomorphic function in a neighborhood of the adherence of V (see the proof of proposition 2.6), the function $1/u_i$ and all its derivatives are bounded. Therefore, it only remains to show that the functions

$$f^*\left(\frac{1}{\log(1/|z_i|)}\right) \text{ and } \log(1/|x_j|)f^*\left(\frac{1}{\log(1/|z_i|)}\right), \text{ for } a_{i,j} \neq 0, \tag{2.25}$$

have log-log growth of infinite order, which is left to the reader. □

Integrability. Since the sheaf of log-log forms is contained in the sheaf of pre-log-log forms, then [10], proposition 7.6, implies

Proposition 2.26. *(i) Any log-log form is locally integrable.*

(ii) If η is a log-log form, and $[\eta]_X$ is the associated current, then

$$[\mathrm{d}\,\eta]_X = \mathrm{d}[\eta]_X.$$

The same holds true for the differential operators ∂, $\bar\partial$, and $\partial\bar\partial$.

□

Logarithmic growth in the local universal cover. We will define a new class of singular forms closely related to the log-log forms. The discussion will be parallel to the one at the end of the previous section.
Let U_M, K, and $(x_1, \dots, x_d)$ be as in definition 2.10.

Definition 2.27. A function f on $(U_M)^k \times K$ is said to have *imaginary logarithmic growth*, if there is a sequence of integers $\{N_n\}_{n\geq 0}$ such that for every pair of multi-indices $\alpha, \beta \in \mathbb{Z}^d_{\geq 0}$, the inequality

$$\left|\frac{\partial^{|\alpha|}}{\partial x^\alpha}\frac{\partial^{|\beta|}}{\partial \bar x^\beta} f(x_1,\dots,x_d)\right| \prec \frac{\left|\prod_{i=1}^k \log(\operatorname{Im} x_i)\right|^{N_{|\alpha|+|\beta|}}}{|x^{\alpha^{\leq k}}\bar x^{\beta^{\leq k}}|} \tag{2.28}$$

holds. The space of differential forms with *imaginary logarithmic growth* is generated by the functions with imaginary logarithmic growth and the differentials

$$\begin{aligned} &\frac{\mathrm{d}\,x_i}{\operatorname{Im} x_i}, \frac{\mathrm{d}\,\bar x_i}{\operatorname{Im} x_i}, && \text{for } i = 1,\dots,k,\\ &\mathrm{d}\,x_i,\ \mathrm{d}\,\bar x_i, && \text{for } i = k+1,\dots,d. \end{aligned}$$

Let X, D, U, and ι be as in definition 2.2.

Definition 2.29. Let W be an open subset of X and let ω be a differential form in $\Gamma(W, \iota_*(\mathscr{E}_U)^*)$. For every point $p \in W$, there is an open coordinate neighborhood $V \subseteq W$, which is adapted to D and such that the coordinates of

V induce an identification $V \cap U = (\Delta_r^*)^k \times K$. We choose $M > \log(1/r)$ and denote by $\pi : (U_M)^k \times K \longrightarrow V$ the covering map given by

$$\pi(x_1, \dots, x_d) = (e^{2\pi i x_1}, \dots, e^{2\pi i x_k}, x_{k+1}, \dots, x_d).$$

We say that ω has *logarithmic growth in the local universal cover*, if, for every V and π as above, $\pi^*\omega$ has imaginary logarithmic growth.

It is easy to see that the differential forms with logarithmic growth in the local universal cover form a sheaf of Dolbeault algebras.

Theorem 2.30. *The sheaf of differential forms with logarithmic growth in the local universal cover is contained in the sheaf of log-log forms.*

Proof. Since the forms with logarithmic growth in the local universal cover form a Dolbeault algebra, it is enough to check that a differential form with logarithmic growth in the local universal cover has log-log growth of infinite order. We start with the case of a function. So let f and g be as in the proof of theorem 2.13. To bound the derivatives of f we use equation (2.15). But in this case

$$\begin{aligned}\left|\frac{\partial^{|\alpha'|}}{\partial x^{\alpha'}}\frac{\partial^{|\beta'|}}{\partial \bar{x}^{\beta'}} g(x_1, \dots, x_d)\right| &\prec \frac{\left|\prod_{i=1}^k \log(|x_i|)\right|^{N_{\alpha',\beta'}}}{|x^{\alpha'^{\leq k}}\bar{x}^{\beta'^{\leq k}}|} \\ &\prec \frac{\left|\prod_{i=1}^k \log(\log(1/|z_i|))\right|^{N_{\alpha',\beta'}}}{|\log(1/|z|)^{\alpha'^{\leq k}+\beta'^{\leq k}}|}.\end{aligned} \tag{2.31}$$

Note that now the different terms of equation (2.15) have slightly different bounds. If we combine the worst bounds of (2.31) with (2.16), we obtain

$$\left|\frac{\partial^{|\alpha|}}{\partial z^{\alpha}}\frac{\partial^{|\beta|}}{\partial \bar{z}^{\beta}} f(z_1, \dots, z_d)\right| \prec \frac{\left|\prod_{i=1}^k \log(\log(1/r_i))\right|^{N_K}}{|z^{\alpha^{\leq k}}\bar{z}^{\beta^{\leq k}}|\prod_{i=1}^k |\log(1/r_1)|^{\min(\alpha_i,1)+\min(\beta_i,1)}}, \tag{2.32}$$

which implies the bounds of f and its derivatives.
To prove the statement for differential forms, we observe that for $1 \leq i \leq k$,

$$\pi^* \frac{\mathrm{d}\, z_i}{z_i \log(1/|z_i|)} = \frac{i\, \mathrm{d}\, x_i}{\operatorname{Im} x_i}.$$

□

Remark 2.33. The differential forms that interest us are the forms with logarithmic growth in the local universal cover. We have introduced the log-log forms because it is easier to work with bounds of the function and its derivatives in usual coordinates than with the condition of logarithmic growth in the local universal cover. This is particularly true in the proof of the Poincaré lemma. Note however that theorem 2.30 provides us only with an inclusion of sheaves and does not give us a characterization of differential forms with logarithmic growth in the local universal cover. This can be seen by the fact that the bounds (2.32) are sharper than the bounds of definition 2.21. We have chosen the bounds of definition 2.21 because the sharper bounds (2.32) are not functorial. Moreover, they do not characterize forms with logarithmic growth in the local universal cover. In fact, it does not exist a characterization of forms with logarithmic growth in the local universal cover in terms of bounds of the function and its derivatives in usual coordinates.

2.3 Log and log-log mixed forms

For the general situation which we are interested in, we need a combination of the concepts of log and log-log forms.

Mixed growth forms. Let X, D, U, and ι be as in the previous section. Let D_1 and D_2 be normal crossing divisors, which may have common components, and such that $D = D_1 \cup D_2$. We denote by D_2' the union of the components of D_2 that are not contained in D_1. We say that an open coordinate subset V, with coordinates $z_1, \dots, z_d$, is adapted to D_1 and D_2, if $D_1 \cap V$ has equation $z_1 \cdots z_k = 0$ and $D_2' \cap V$ has equation $z_{k+1} \cdots z_l = 0$; we put $r_i := |z_i| < 1/e^e$ for $i = 1, \dots, d$.

Definition 2.34. Let V be a coordinate neighborhood adapted to D_1 and D_2. For every integer $K \geq 0$, we say that a smooth complex function f on $V \setminus D$ has *log growth along D_1 and log-log growth along D_2 of order K*, if there exists an integer $N_K \geq 0$ such that, for every pair of multi-indices $\alpha, \beta \in \mathbb{Z}^d_{\geq 0}$, with $|\alpha + \beta| \leq K$, it holds the inequality

$$\left| \frac{\partial^{|\alpha|}}{\partial z^{\alpha}} \frac{\partial^{|\beta|}}{\partial \bar{z}^{\beta}} f(z_1, \dots, z_d) \right| \prec \frac{\left| \prod_{i=1}^{k} \log(1/r_i) \prod_{j=k+1}^{l} \log(\log(1/r_j)) \right|^{N_K}}{|z^{\alpha^{\leq l}} \bar{z}^{\beta^{\leq l}}|}. \tag{2.35}$$

We say that f has *log growth along D_1 and log-log growth along D_2 of infinite order*, if it has log growth along D_1 and log-log growth along D_2 of order K for all $K \geq 0$. The *sheaf of differential forms on X with log growth along D_1 and log-log growth along D_2 of infinite order* is the subalgebra of $\iota_* \mathscr{E}^*_U$ generated, in each coordinate neighborhood V adapted to D_1 and D_2, by the functions with log growth along D_1 and log-log growth along D_2, and the differentials

$$\begin{aligned}
&\frac{\mathrm{d}\, z_i}{z_i}, \frac{\mathrm{d}\, \bar{z}_i}{\bar{z}_i}, && \text{for } i = 1, \dots, k, \\
&\frac{\mathrm{d}\, z_i}{z_i \log(1/r_i)}, \frac{\mathrm{d}\, \bar{z}_i}{\bar{z}_i \log(1/r_i)}, && \text{for } i = k+1, \dots, l, \\
&\mathrm{d}\, z_i, \mathrm{d}\, \bar{z}_i, && \text{for } i = l+1, \dots, d.
\end{aligned}$$

When the normal crossing divisors D_1 and D_2 are clear from the context, a differential form with log growth along D_1 and log-log growth along D_2 of infinite order will be called a *mixed growth form.* The sheaf of differential forms on X with log growth along D_1 and log-log growth along D_2 of infinite order will be denoted $\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}$.

It is clear that

$$\mathscr{E}_X^*\langle D_1\rangle \wedge \mathscr{E}_X^*\langle\langle D_2\rangle\rangle_{\mathrm{gth}} \subseteq \mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}. \tag{2.36}$$

Observe moreover that, by definition,

$$\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}} = \mathscr{E}_X^*\langle D_1\langle D_2'\rangle\rangle_{\mathrm{gth}}.$$

We leave it to the reader to state the analogue of lemma 2.19.

Partial differentials. Let V be an open coordinate system adapted to D_1 and D_2. In this coordinate system we may decompose the operators ∂ and $\bar\partial$ as

$$\partial = \sum_j \partial_j \quad \text{and} \quad \bar\partial = \sum_j \bar\partial_j, \tag{2.37}$$

where ∂_j and $\bar\partial_j$ contain only the derivatives with respect to z_j.
The following lemma follows directly from the definition.

Lemma 2.38. *Let E_j denote the divisor given by $z_j = 0$. If $\omega \in \mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(V)$, then*

$$\partial_j\omega \in \begin{cases} \mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(V), & \text{if } j \leq k \text{ or } j > l, \\ \mathscr{E}_X^*\langle D_1 \cup E_j\langle D_2\rangle\rangle_{\mathrm{gth}}(V), & \text{if } k < j \leq l, \end{cases}$$

and the same is true for the operator $\bar\partial_j$.

□

Mixed Forms.

Definition 2.39. We say that a section ω of $\iota_*\mathscr{E}_U^*$ is *log along D_1 and log-log along D_2*, if the differential forms ω, $\partial\omega$, $\bar\partial\omega$, and $\partial\bar\partial\omega$ are sections of $\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}$. The sheaf of differential forms log along D_1 and log-log along D_2 will be denoted by $\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle$. As a shorthand, a differential form which is log along D_1 and log-log along D_2, will be called a mixed form.

As the complexes we have defined in the previous sections, the complex $\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle$ is a sheaf of Dolbeault algebras.

Inverse images. We can generalize propositions 2.6 and 2.24, with the same technique, to the case of mixed forms.

Proposition 2.40. *Let $f : X \longrightarrow Y$ be a morphism of complex manifolds. Let D_1, D_2 and E_1, E_2 be normal crossing divisors on X and Y respectively, such that $D_1 \cup D_2$ and $E_1 \cup E_2$ are also normal crossing divisors. Furthermore, assume that $f^{-1}(E_1) \subseteq D_1$ and $f^{-1}(E_2) \subseteq D_1 \cup D_2$. If η is a section of $\mathscr{E}_Y^*\langle E_1\langle E_2\rangle\rangle$, then $f^*\eta$ is a section of $\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle$.* □

Integrability. Let X be a complex manifold and D a normal crossing divisor. Let y be a p-codimensional cycle of X and let $Y = \operatorname{supp} y$. Let $\pi : \widetilde{X} \longrightarrow X$ be an embedded resolution of singularities of Y, with normal crossing divisors $D_Y = \pi^{-1}(Y)$ and $\widetilde{D} = \pi^{-1}(D)$ and such that $D_Y \cup \widetilde{D}$ is also a normal crossing divisor.

Lemma 2.41. *Assume that $g \in \Gamma(\widetilde{X}, \mathscr{E}^n_{\widetilde{X}}\langle D_Y\langle \widetilde{D}\rangle\rangle)$. Then, the following statements hold:*

(i) If $n < 2p$, then g is locally integrable on the whole of X. We denote by $[g]_X$ the current associated to g.

(ii) If $n < 2p - 1$, then $\mathrm{d}[g]_X = [\mathrm{d}\, g]_X$.

Proof. This is a particular case of [10], lemma 7.13. □

The cohomology of the complex of mixed forms. We are now in position to state the main result of this section.

Theorem 2.42. *The inclusion*

$$\Omega^*_X(\log D_1) \longrightarrow \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle$$

is a filtered quasi-isomorphism with respect to the Hodge filtration.

Proof. To prove the theorem we will use the classical argument for proving the Poincaré lemma in several variables. We will state here the general argument and we will delay the specific analytic lemmas that we need until the next section.

The theorem is equivalent to the exactness of the sequence of sheaves

$$0 \longrightarrow \Omega^p_X(\log D_1) \xrightarrow{\ i\ } \mathscr{E}^{p,0}_X\langle D_1\langle D_2\rangle\rangle \xrightarrow{\ \bar\partial\ } \mathscr{E}^{p,1}_X\langle D_1\langle D_2\rangle\rangle \xrightarrow{\ \bar\partial\ } \dots$$

The exactness in the first step is clear because a holomorphic form on $X \setminus (D_1 \cup D_2)$ that satisfies the growth conditions imposed in the definitions can only have logarithmic poles along D_1.

For the exactness in the other steps we choose a point $x \in X$. Let V be a coordinate neighborhood of x adapted to D_1 and D_2, and such that x has coordinates $(0, \dots, 0)$.

Let $0 < \epsilon \ll 1$, we denote by $\Delta^d_{x,\epsilon}$ the poly-cylinder

$$\Delta^d_{x,\epsilon} = \{(z_1, \dots, z_d) \in V \mid r_i < \epsilon,\ i = 1, \dots, d\}.$$

In the next section we will prove that, for $j = 1, \dots, d$ and $0 < \epsilon' < \epsilon \ll 1$, there exist operators

$$\begin{aligned} K^{\epsilon',\epsilon}_j &: \mathscr{E}^{p,q}_X\langle D_1\langle D_2\rangle\rangle(\Delta^d_{x,\epsilon}) \longrightarrow \mathscr{E}^{p,q-1}_X\langle D_1\langle D_2\rangle\rangle(\Delta^d_{x,\epsilon'}),\\ P^{\epsilon',\epsilon}_j &: \mathscr{E}^{p,q}_X\langle D_1\langle D_2\rangle\rangle(\Delta^d_{x,\epsilon}) \longrightarrow \mathscr{E}^{p,q}_X\langle D_1\langle D_2\rangle\rangle(\Delta^d_{x,\epsilon'}),\end{aligned}$$

that satisfy the following conditions

(A) If the form ω does not contain any term with $\mathrm{d}\,\bar{z}_i$ for $i > j$, then $K_j^{\epsilon',\epsilon}\omega$ and $P_j^{\epsilon',\epsilon}\omega$ do not contain any term with $\mathrm{d}\,\bar{z}_i$ for $i \geq j$.

(B) $\bar{\partial} K_j^{\epsilon',\epsilon} + K_j^{\epsilon',\epsilon}\bar{\partial} + P_j^{\epsilon',\epsilon} = \mathrm{id}$.

Let $q > 0$ and let $\omega \in \mathscr{E}_X^{p,q}\langle D_1 \langle D_2\rangle\rangle_x$ be a germ of a closed form. Assume that ω is defined in a poly-cylinder $\Delta^d_{x,\epsilon}$. By abuse of notation we will not distinguish between sections and germs. Therefore, ω will denote also a closed differential form over $\Delta^d_{x,\epsilon}$ that represents this germ. Moreover, as the open set of definition of each section will be clear from the context, we will not make it explicit. We choose real numbers $0 < \epsilon_1 < \ldots < \epsilon_d < \epsilon$. Then, by property (B), we have

$$\omega = \bar{\partial} K_d^{\epsilon,\epsilon_d}(\omega) + P_d^{\epsilon,\epsilon_d}(\omega).$$

We write $\omega_1 = P_d^{\epsilon,\epsilon_d}(\omega)$. Then, ω_1 does not contain $\mathrm{d}\,\bar{z}_d$ and $\omega - \omega_1$ is a boundary. We define inductively $\omega_{j+1} = P_{d-j}^{\epsilon_{d-j+1},\epsilon_{d-j}}(\omega_j)$. Then, for all j, $\omega - \omega_j$ is a boundary and ω_j does not contain $\mathrm{d}\,\bar{z}_i$ for $i > d - j$. Therefore, $\omega_{d-q+1} = 0$ and ω is a boundary. $\square$

2.4 Analytic lemmas

In this section we will prove the analytic lemmas needed to prove theorem 2.42 and we will define the operators K and P that appear in the proof of this theorem.

Primitive functions with growth conditions. Let f be a smooth function on Δ^*_ϵ, which is integrable on any compact subset of Δ_ϵ. Then, for $\epsilon' < \epsilon$ and $z \in \Delta^*_{\epsilon'}$, we write

$$I_{\epsilon'}(f)(z) = \frac{1}{2\pi\sqrt{-1}} \int_{\Delta_{\epsilon'}} f(w) \frac{\mathrm{d}\,w \wedge \mathrm{d}\,\bar{w}}{w - z}.$$

We denote $r = |z|$.

Lemma 2.43. *(i) If f is a smooth function on Δ^*_ϵ such that*

$$|f(z)| \prec \frac{|\log(\log(1/r))|^N}{(r\log(1/r))^2},$$

then f is integrable in each compact subset of Δ_ϵ and

$$\frac{\partial}{\partial \bar{z}} I_{\epsilon'}(f)(z) = f(z).$$

(ii) If f is a smooth function on Δ^*_ϵ such that

$$|f(z)| \prec \frac{|\log(\log(1/r))|^N}{r\log(1/r)} \quad \textit{and} \quad \left|\frac{\partial}{\partial \bar{z}} f(z)\right| \prec \frac{|\log(\log(1/r))|^N}{(r\log(1/r))^2},$$

then

$$2\pi\sqrt{-1}f(z) = \int_{\partial\Delta_{\epsilon'}} f(w)\frac{\mathrm{d}\,w}{w-z} + \int_{\overline{\Delta}_{\epsilon'}} \frac{\partial}{\partial \bar{w}} f(w)\frac{\mathrm{d}\,w\wedge \mathrm{d}\,\bar{w}}{w-z}.$$

(iii) If f is a smooth function on Δ^*_ϵ such that

$$|f(z)| \prec \frac{|\log(\log(1/r))|^N}{r\log(1/r)} \quad \textit{and} \quad \left|\frac{\partial}{\partial z}f(z)\right| \prec \frac{|\log(\log(1/r))|^N}{(r\log(1/r))^2},$$

then

$$\frac{\partial}{\partial z}\int_{\overline{\Delta}_{\epsilon'}} f(w)\frac{\mathrm{d}\,w\wedge \mathrm{d}\,\bar{w}}{w-z} = -\int_{\partial\Delta_{\epsilon'}} f(w)\frac{\mathrm{d}\,\bar{w}}{w-z} + \int_{\overline{\Delta}_{\epsilon'}} \frac{\partial}{\partial w} f(w)\frac{\mathrm{d}\,w\wedge \mathrm{d}\,\bar{w}}{w-z}.$$

Proof. We start by proving the integrability of f. Viewed as a function of ϵ, we estimate

$$\begin{aligned}\left|\int_{\Delta_\epsilon} \frac{(\log(\log(1/r)))^N}{r^2(\log(1/r))^2}\,\mathrm{d}\,z\wedge \mathrm{d}\,\bar{z}\right| &\prec \left|\int_0^\epsilon \frac{(\log(\log(1/r)))^N}{r^2(\log(1/r))^2} r\,\mathrm{d}\,r\right| \\ &\prec \left|\int_0^\epsilon \frac{1}{r(\log(1/r))^{3/2}}\,\mathrm{d}\,r\right| \\ &\prec \frac{1}{(\log(1/\epsilon))^{1/2}},\end{aligned}$$

which proves the integrability. Then, the claimed formulas are proven as in [20], pp. 24–26. The only new point one has to care about is that the singularities at $z = 0$ do not contribute to Stokes theorem. □

Lemma 2.44. *Let $0 < \epsilon \ll 1$ be a real number and let f be a smooth function on Δ^*_ϵ. Let $\epsilon' < \epsilon$.*

(i) If $\omega = f\,\mathrm{d}\,\bar{z} \in \mathscr{E}^{0,1}_{\Delta_\epsilon}\langle 0\rangle(\Delta_\epsilon)$, then the function f is integrable on any compact subset of Δ_ϵ. We write $g = I_{\epsilon'}(f)$. Then, $g \in \mathscr{E}^{0,0}_{\Delta_{\epsilon'}}\langle 0\rangle(\Delta_{\epsilon'})$ and

$$\bar{\partial} g = \omega. \tag{2.45}$$

(ii) If, moreover, $\omega \in \mathscr{E}^{0,1}_{\Delta_\epsilon}\langle\langle 0\rangle\rangle_{\mathrm{gth}}(\Delta_\epsilon)$, then $g \in \mathscr{E}^{0,0}_{\Delta_{\epsilon'}}\langle\langle 0\rangle\rangle_{\mathrm{gth}}(\Delta_{\epsilon'})$.

(iii) If $\omega = f\,\mathrm{d}\,\bar{z}\wedge \mathrm{d}\,z \in \mathscr{E}^{1,1}_{\Delta_\epsilon}\langle\langle 0\rangle\rangle_{\mathrm{gth}}(\Delta_\epsilon)$, then the function f is integrable on any compact subset of Δ_ϵ. If we write $g = I_{\epsilon'}(f)$ as before, then $g\,\mathrm{d}\,z \in \mathscr{E}^{1,0}_{\Delta_{\epsilon'}}\langle\langle 0\rangle\rangle_{\mathrm{gth}}(\Delta_{\epsilon'})$ and

$$\bar{\partial}(g\wedge \mathrm{d}\,z) = \omega. \tag{2.46}$$

Proof. The integrability in the three cases and equations (2.45) and (2.46) are in lemma 2.43. Therefore, it remains only to prove the necessary bounds.
Proof of (i). The condition on ω is equivalent to the inequalities

$$\left|\frac{\partial^{\alpha+\beta}}{\partial z^{\alpha}\partial\bar{z}^{\beta}}f(z)\right| \prec \frac{|\log(1/r)|^{N_{\alpha+\beta}}}{r^{\alpha+\beta+1}} \tag{2.47}$$

for a certain family of integers $\{N_n\}_{n\in\mathbb{Z}_{\geq 0}}$. We may assume that these integers satisfy for $a \leq b$ the inequality $N_a \leq N_b$. We can apply [22], lemma 1, to conclude that g is smooth on $\Delta^*_{\epsilon'}$ and that

$$|g(z)| \prec |\log(1/r)|^{N'_0}$$

for some integer N'_0.
Thus, to prove statement (i), it remains to bound the derivatives of g. Equation (2.45) implies the bound for the derivatives

$$\frac{\partial^{\alpha+\beta}}{\partial z^{\alpha}\partial\bar{z}^{\beta}}g,$$

when $\beta \geq 1$. Therefore, we may assume $\beta = 0$ and $\alpha \geq 1$.
Let $\rho : \mathbb{C} \longrightarrow [0,1]$ be a smooth function such that

$$\rho|_{B(0,1)} = 1, \qquad \rho|_{\mathbb{C}\setminus B(0,2)} = 0,$$

where $B(p,\delta)$ is the open ball of center p and radius δ. Fix $z_0 \in \Delta^*_{\epsilon'}$. Since we want to bound the derivatives of $g(z)$ as z goes to zero, we may assume $z_0 \in \Delta^*_{\epsilon'/2}$. Write $r_0 = |z_0|$, and put

$$\rho_{z_0}(z) = \rho\left(3\frac{z-z_0}{r_0}\right).$$

Then, we have

$$\rho|_{B(z_0,r_0/3)} = 1, \qquad \rho|_{\mathbb{C}\setminus B(z_0,2r_0/3)} = 0.$$

Moreover, we have

$$\frac{\partial^{\alpha}}{\partial z^{\alpha}}\rho_{z_0}(z) \leq \frac{C_\alpha}{r_0^{\alpha}} \tag{2.48}$$

for some constants C_α.
By the choice of z_0, we have that $\operatorname{supp}(\rho_{z_0}) \subseteq \Delta^*_{\epsilon'}$. We write $f_1 = \rho_{z_0}f$ and $f_2 = (1-\rho_{z_0})f$. Then, for $z \in B(z_0, r_0/3)$, we introduce the auxiliary functions

$$g_1(z) = \frac{1}{2\pi\sqrt{-1}}\int_{\Delta_\epsilon} f_1(w)\frac{\mathrm{d}\,w\wedge\mathrm{d}\,\bar{w}}{w-z}, \quad g_2(z) = \frac{1}{2\pi\sqrt{-1}}\int_{\Delta_\epsilon} f_2(w)\frac{\mathrm{d}\,w\wedge\mathrm{d}\,\bar{w}}{w-z}.$$

These functions satisfy

$$g = g_1 + g_2.$$

Therefore, we can bound the derivatives of g_1 and g_2 separately. We first bound the derivatives of g_1.

$$\begin{aligned}\frac{\partial^\alpha}{\partial z^\alpha}\int_{\Delta_\epsilon} f_1(w)\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{w-z} &= \frac{\partial^\alpha}{\partial z^\alpha}\int_{\mathbb{C}} f_1(w)\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{w-z}\\ &= \frac{\partial^\alpha}{\partial z^\alpha}\int_{\mathbb{C}} f_1(u+z)\frac{\mathrm{d}\, u\wedge \mathrm{d}\,\bar u}{u}\\ &= \int_{\mathbb{C}}\frac{\partial^\alpha}{\partial z^\alpha} f_1(u+z)\frac{\mathrm{d}\, u\wedge \mathrm{d}\,\bar u}{u}\\ &= \int_{\mathbb{C}}\frac{\partial^\alpha}{\partial w^\alpha} f_1(w)\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{w-z}\\ &= \int_{B(z_0,2r_0/3)}\frac{\partial^\alpha}{\partial w^\alpha} f_1(w)\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{w-z}.\end{aligned}$$

Hence, using the bounds for the derivatives of f and equation (2.48), we find the inequality

$$\begin{aligned}\left|\frac{\partial^\alpha}{\partial z^\alpha} g_1(z_0)\right| &\prec \frac{|\log(1/r_0)|^{N_\alpha}}{r_0^{\alpha+1}}\left|\int_{B(z_0,2r_0/3)}\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{w-z_0}\right|\\ &\prec \frac{|\log(1/r_0)|^{N_\alpha}}{r_0^{\alpha}}.\end{aligned}$$

Now we bound the derivatives of g_2. Since for $z\in B(z_0,r_0/3)$, the function $f_2(w)$ is identically zero in a neighborhood of the point $w=z$, we have

$$\frac{\partial^\alpha}{\partial z^\alpha}\int_{\Delta_\epsilon} f_2(w)\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{w-z} = \int_{\Delta_\epsilon} f_2(w)\alpha!\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{(w-z)^{\alpha+1}}.$$

Let $A=B(0,r_0/2)$. Then, for $w\in A$, we have $|w-z_0|\geq r_0/2$. Thus, we obtain

$$\begin{aligned}\left|\int_A f_2(w)\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{(w-z_0)^{\alpha+1}}\right| &\prec \frac{1}{r_0^{\alpha+1}}\int_0^{r_0/2}\frac{|\log(1/\rho)|^{N_0}}{\rho}\rho\,\mathrm{d}\,\rho\\ &\prec \frac{1}{r_0^{\alpha}}|\log(1/r_0)|^{N_0}.\end{aligned}$$

Here we use that

$$\int(\log x)^N\,\mathrm{d}\, x = x\sum_{i=0}^{N}(-1)^i\frac{N!}{(N-i)!}(\log x)^{N-i}.$$

We write $B=\Delta_{\epsilon'}\setminus(A\cup B(z_0,r_0/3))$. In this region $|w-z_0|\geq|w/4|$. Therefore, we get

$$\left|\int_B f_2(w)\frac{\mathrm{d}\, w\wedge \mathrm{d}\,\bar w}{(w-z_0)^{\alpha+1}}\right|\prec\int_{r_0/2}^{\epsilon'}\frac{|\log(1/\rho)|^{N_0}}{\rho}\frac{\rho\,\mathrm{d}\,\rho}{\rho^{\alpha+1}}\prec\frac{1}{r_0^{\alpha}}|\log(1/r_0)|^{N_0+1}.$$

Here we use that

$$\int (\log x)^n \frac{1}{x^m} \,\mathrm{d}\, x = \begin{cases} \frac{1}{n+1}(\log x)^{n+1} & \text{if } m = 1, \\ \frac{1}{x^{m-1}} P_{n,m}(\log x) & \text{if } m > 1, \end{cases}$$

where $P_{n,m}$ is a polynomial of degree n. Summing up, we obtain

$$\left| \frac{\partial^\alpha}{\partial z^\alpha} g(z_0) \right| \prec \frac{|\log(1/r_0)|^{N_\alpha + 1}}{r_0^\alpha}$$

Observe that, for $\alpha = 0$, this is the proof of [22], lemma 1.
Proof of (ii). In this case, by lemma 2.19, the condition on ω is equivalent to the inequalities

$$\left| \frac{\partial^{\alpha+\beta}}{\partial z^\alpha \partial \bar{z}^\beta} f(z) \right| \prec \frac{|\log(\log(1/r))|^{N_{\alpha+\beta}}}{r^{\alpha+\beta+1} \log(1/r)} \tag{2.49}$$

for a certain increasing family of integers $\{N_n\}_{n \in \mathbb{Z}_{\geq 0}}$. Again, by lemma 2.19, to prove statement (ii), we have to show

$$\left| \frac{\partial^{\alpha+\beta}}{\partial z^\alpha \partial \bar{z}^\beta} g(z) \right| \prec \frac{|\log(\log(1/r))|^{N'_{\alpha+\beta}}}{r^{\alpha+\beta}} \tag{2.50}$$

for a certain family of integers $\{N'_n\}_{n \in \mathbb{Z}_{\geq 0}}$.
By (2.45), the functions

$$\frac{\partial^{\alpha+\beta}}{\partial z^\alpha \partial \bar{z}^\beta} g,$$

when $\beta \geq 1$, satisfy the required bounds. Thus, it remains to bound $\partial^\alpha / \partial z^\alpha g$ for $\alpha \geq 0$. As in the proof of statement (i), we fix z_0 and write $g = g_1 + g_2$. For g_1, we work as before and get for $\alpha \geq 0$

$$\left| \frac{\partial^\alpha}{\partial z^\alpha} g_1(z_0) \right| \prec \frac{|\log(\log(1/r_0))|^{N_\alpha}}{r_0^\alpha \log(1/r_0)}.$$

To bound g_2, we integrate over the regions A and B as before. We first bound the integral over the region $A = B(0, r_0/2)$.

$$\left| \int_A f_2(w) \frac{\mathrm{d}\, w \wedge \mathrm{d}\, \bar{w}}{(w - z_0)^{\alpha+1}} \right| \prec \frac{1}{r_0^{\alpha+1}} \int_0^{r_0/2} \frac{|\log(\log(1/\rho))|^{N_0}}{\log(1/\rho)} \,\mathrm{d}\, \rho.$$

Since, for $\rho < 1/e^{e^{N_0}}$, the function

$$\frac{(\log(\log(1/\rho)))^{N_0}}{\log(1/\rho)}$$

is an increasing function, we have

$$\begin{aligned} \frac{1}{r_0^{\alpha+1}} \int_0^{r_0/2} \frac{|\log(\log(1/\rho))|^{N_0}}{\log(1/\rho)} \,\mathrm{d}\, \rho &\prec \frac{|\log(\log(1/r_0))|^{N_0}}{r_0^{\alpha+1} \log(1/r_0)} \int_0^{r_0/2} \mathrm{d}\, \rho \\ &\prec \frac{|\log(\log(1/r_0))|^{N_0}}{r_0^\alpha \log(1/r_0)} \end{aligned}$$

in the domain $0 < r_0 \leq 2/e^{e^{N_0}}$.
If f and g are two continuous functions with g strictly positive, defined on a compact set, then $f \prec g$. Therefore, the above inequality extends to the domain $0 \leq r_0 \leq \epsilon'/2$.
We now bound the integral over the region $B = \Delta_{\epsilon'} \setminus (A \cup B(z_0, r_0/3))$. By the bound of the function f, we have

$$\left| \int_B f_2(w) \frac{\mathrm{d}\, w \wedge \mathrm{d}\, \bar{w}}{(w - z_0)^{\alpha+1}} \right| \prec \int_{r_0/2}^{\epsilon'} \frac{|\log(\log(1/\rho))|^{N_0}}{\rho^{\alpha+1} \log(1/\rho)} \,\mathrm{d}\,\rho.$$

Thus, in the case $\alpha = 0$, we have

$$\int_{r_0/2}^{\epsilon'} \frac{|\log(\log(1/\rho))|^{N_0}}{\rho \log(1/\rho)} \,\mathrm{d}\,\rho \prec |\log(\log(1/r_0))|^{N_0+1}.$$

In the case $\alpha > 0$, since, for $\rho < 1/e^e$, the function

$$\frac{(\log(\log(1/\rho)))^{N_0}}{\rho^{1/2} \log(1/\rho)}$$

is a decreasing function, we have

$$\begin{aligned}\int_{r_0/2}^{\epsilon'} \frac{|\log(\log(1/\rho))|^{N_0}}{\rho^{\alpha+1} \log(1/\rho)} \,\mathrm{d}\,\rho &\prec \frac{|\log(\log(1/r_0))|^{N_0}}{r_0^{1/2} \log(1/r_0)} \int_{r_0/2}^{\epsilon'} \frac{1}{\rho^{\alpha+1/2}} \,\mathrm{d}\,\rho \\ &\prec \frac{|\log(\log(1/r_0))|^{N_0}}{r_0^{\alpha} \log(1/r_0)}.\end{aligned}$$

Summing up, we obtain

$$\left| \frac{\partial^\alpha}{\partial z^\alpha} g(z_0) \right| \prec \frac{|\log(\log(1/r_0))|^{N_\alpha+1}}{r_0^\alpha}.$$

This finishes the proof of the second statement.
Proof of (iii). In this case, again by lemma 2.19, the condition on ω is equivalent to the conditions

$$\left| \frac{\partial^{\alpha+\beta}}{\partial z^\alpha \partial \bar{z}^\beta} f(z) \right| \prec \frac{|\log(\log(1/r))|^{N_{\alpha+\beta}}}{r^{\alpha+\beta+2} (\log(1/r))^2}$$

for a certain increasing family of integers $\{N_n\}_{n \in \mathbb{Z}_{\geq 0}}$, and the inequalities we have to prove are

$$\left| \frac{\partial^{\alpha+\beta}}{\partial z^\alpha \partial \bar{z}^\beta} g(z) \right| \prec \frac{|\log(\log(1/r))|^{N'_{\alpha+\beta}}}{r^{\alpha+\beta+1} \log(1/r)}$$

for a certain family of integers $\{N'_n\}_{n \in \mathbb{Z}_{\geq 0}}$.

First we note that, by equation (2.46), for $\beta \geq 1$, the functions

$$\frac{\partial^{\alpha+\beta}}{\partial z^{\alpha}\partial \bar{z}^{\beta}} g$$

satisfy the required bounds. Thus, it remains to bound the functions $\partial^{\alpha} g/\partial z^{\alpha}$ for $\alpha \geq 0$. The proof is similar as before. We decompose again $g = g_1 + g_2$. In this case

$$\left|\frac{\partial^{\alpha}}{\partial z^{\alpha}} g_1(z_0)\right| \prec \frac{|\log(\log(1/r_0))|^{N_{\alpha}}}{\log(1/r_0)^2 r_0^{\alpha+1}}.$$

Whereas the integral of g_2 over A is bounded as

$$\begin{aligned}\left|\int_A f_2(w)\frac{\mathrm{d}\, w \wedge \mathrm{d}\, \bar{w}}{(w-z_0)^{\alpha+1}}\right| &\prec \frac{1}{r_0^{\alpha+1}}\int_0^{r_0/2}\frac{|\log(\log(1/\rho))|^{N_0}}{\rho\log(1/\rho)^2}\,\mathrm{d}\,\rho \\ &\prec \frac{|\log(\log(1/r_0))|^{N_0}}{r_0^{\alpha+1}\log(1/r_0)},\end{aligned}$$

and the integral of g_2 over B is bounded as

$$\begin{aligned}\left|\int_B f_2(w)\frac{\mathrm{d}\, w \wedge \mathrm{d}\, \bar{w}}{(w-z_0)^{\alpha+1}}\right| &\prec \int_{r_0/2}^{\epsilon'}\frac{|\log(\log(1/\rho))|^{N_0}}{\rho^{\alpha+2}\log(1/\rho)^2}\,\mathrm{d}\,\rho \\ &\prec \frac{|\log(\log(1/r_0))|^{N_0}}{r_0^{\alpha+1}\log(1/r_0)^2}.\end{aligned}$$

Summing up, we obtain for $\alpha \geq 0$

$$\left|\frac{\partial^{\alpha}}{\partial z^{\alpha}} g(z_0)\right| \prec \frac{|\log(\log(1/r_0))|^{N_{\alpha}}}{r_0^{\alpha+1}\log(1/r_0)}.$$

This finishes the proof of the lemma. □

Remark 2.51. Observe that, in general, a section of $\mathscr{E}^{1,1}_{\Delta_\epsilon}\langle 0\rangle(\Delta_\epsilon)$ is not locally integrable (see remark 2.55). Therefore, the analogue of lemma 2.44 (iii) is not true for log forms.

The operators K and P. Let X, U, D, ι, D_1, and D_2 be as in definition 2.39.

Notation 2.52. Let $x \in X$. Let V be an open coordinate neighborhood of x with coordinates $z_1, \dots, z_d$, adapted to D_1 and D_2, such that x has coordinates $(0, \dots, 0)$. Thus, D_1 has equation $z_1 \cdots z_k = 0$ and D_2' has equation $z_{k+1} \cdots z_l = 0$. Once this coordinate neighborhood is chosen, we put

$$\begin{aligned}\zeta_i &= \frac{\mathrm{d}\, z_i}{z_i}, &&\text{if } 1 \leq i \leq k,\\ \zeta_i &= \mathrm{d}\, z_i, &&\text{if } i > k.\end{aligned}$$

For any subset $I \subseteq \{1, \dots, d\}$, we denote

$$\zeta_I = \bigwedge_{i \in I} \zeta_i, \quad \mathrm{d}\,\bar{z}_I = \bigwedge_{i \in I} \mathrm{d}\,\bar{z}_i.$$

Given any differential form ω, let

$$\omega = \sum_{I,J} f_{I,J} \zeta_I \wedge \mathrm{d}\,\bar{z}_J$$

be the decomposition of ω in monomials. Then, we write

$$\omega_{I,J} = f_{I,J} \zeta_I \wedge \mathrm{d}\,\bar{z}_J.$$

For any subset $I \subseteq \{1, \dots, d\}$ and $i \in I$, we will write

$$\sigma(I,i) = \sharp\{j \in I \mid j < i\} \qquad \text{and} \qquad I_i = I \setminus \{i\}.$$

Definition 2.53. Let $0 < \epsilon' < \epsilon \ll 1$. Let $\Delta^d_{x,\epsilon}$ be the poly-cylinder centered at x of radius ϵ. Let $\omega \in \mathscr{E}^{p,q}_X \langle D_1 \langle D_2 \rangle \rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon})$, and let

$$\omega = \sum_{I,J} f_{I,J} \zeta_I \wedge \mathrm{d}\,\bar{z}_J \tag{2.54}$$

be the decomposition of ω into monomials. We define

$$\begin{aligned} K^{\epsilon',\epsilon}_j(\omega) = \sum_I (-1)^{\# I} \zeta_I \wedge \\ \sum_{J \mid j \in J} \frac{(-1)^{\sigma(J,j)}}{2\pi\sqrt{-1}} \int_{\Delta_{\epsilon'}} f_{I,J}(\dots, z_{j-1}, w, z_{j+1}, \dots) \frac{\mathrm{d}\,w \wedge \mathrm{d}\,\bar{w}}{w - z_j} \,\mathrm{d}\,\bar{z}_{J_j}, \end{aligned}$$

$$\begin{aligned} P^{\epsilon',\epsilon}_j(\omega) = \sum_I \zeta_I \wedge \\ \sum_{J \mid j \notin J} \frac{1}{2\pi\sqrt{-1}} \int_{\partial\Delta_{\epsilon'}} f_{I,J}(\dots, z_{j-1}, w, z_{j+1}, \dots) \frac{\mathrm{d}\,w}{w - z_j} \,\mathrm{d}\,\bar{z}_J. \end{aligned}$$

To ease notation, if ϵ and ϵ' are clear from the context, we will drop them and write K_j, resp. P_j instead of $K^{\epsilon',\epsilon}_j$, resp. $P^{\epsilon',\epsilon}_j$.

Remark 2.55. The reason why we use the differentials ζ_I instead of $\mathrm{d}\,z_I$ in the definition of K and P, is that, in general, a log form is not locally integrable. For instance, if $d = k = 1$ and $\omega = f\ \mathrm{d}\,z \wedge \mathrm{d}\,\bar{z}$ is a section of $\mathscr{E}^{1,1}_{\Delta_\epsilon} \langle 0 \rangle (\Delta_\epsilon)$, then f satisfies

$$|f(z)| \prec \frac{|\log(1/r)|^N}{r^2},$$

and the integral

$$\int_{\overline{\Delta}_{\epsilon'}} \frac{|\log(1/|w|)|^N}{|w|^2} \frac{\mathrm{d}\, w \wedge \mathrm{d}\, \bar{w}}{w-z}$$

does not converge. But, by the definition we have adopted, $K^{\epsilon',\epsilon}(\omega) = g \,\mathrm{d}\, z$, where

$$g(z) = \frac{1}{z} I_{\epsilon'}(z \cdot f) = \frac{1}{2\pi\sqrt{-1}} \frac{1}{z} \int_{\overline{\Delta}_{\epsilon'}} w f(w) \frac{\mathrm{d}\, w \wedge \mathrm{d}\, \bar{w}}{w-z}.$$

This integral is absolutely convergent and

$$\frac{\partial}{\partial \bar{z}} g(z) = \frac{1}{z} \frac{\partial}{\partial \bar{z}} I_{\epsilon'}(z \cdot f)(z) = \frac{z f(z)}{z} = f(z).$$

This trick will force us to be careful when studying the compatibility of K with the operator ∂ because, for a log form ω, the definitions of $K(\omega)$ and of $K(\partial\omega)$ use different kernels in the integral operators.

Theorem 2.56. *Let $\omega \in \mathscr{E}^{p,q}_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon})$. Then, we have*

$$K^{\epsilon',\epsilon}_j(\omega) \in \mathscr{E}^{p,q-1}_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon'}), \text{ and}$$
$$P^{\epsilon',\epsilon}_j(\omega) \in \mathscr{E}^{p,q}_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon'}).$$

These operators satisfy

(i) If the form ω does not contain any term with $\mathrm{d}\,\bar{z}_i$ for $i > j$, then $K_j\omega$ and $P_j\omega$ do not contain any term with $\mathrm{d}\,\bar{z}_i$ for $i \geq j$.

(ii) If $\omega \in \mathscr{E}^{p,q}_X\langle D_1\langle D_2\rangle\rangle(\Delta_{x,\epsilon})$, then

$$K^{\epsilon',\epsilon}_j(\omega) \in \mathscr{E}^{p,q-1}_X\langle D_1\langle D_2\rangle\rangle(\Delta_{x,\epsilon'}), \text{ and}$$
$$P^{\epsilon',\epsilon}_j(\omega) \in \mathscr{E}^{p,q}_X\langle D_1\langle D_2\rangle\rangle(\Delta_{x,\epsilon'}).$$

(iii) In this case, $\bar{\partial} K_j + K_j \bar{\partial} + P_j = \mathrm{id}$.

Proof. By lemma 2.44 and the theorem of taking derivatives under the integral sign, we have that $K_j(\omega) \in \mathscr{E}^{p,q-1}_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon'})$, and it is clear that $P_j(\omega) \in \mathscr{E}^{p,q}_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon'})$. Then, property (i) follows from the definition and it is easy to see that, if $\bar{\partial}\omega$, $\partial\omega$, and $\partial\bar{\partial}\omega$ belong to $\mathscr{E}^{*}_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon})$, the same is true for $\bar{\partial} P_j(\omega)$, $\partial P_j(\omega)$, and $\partial\bar{\partial} P_j(\omega)$.
In the sequel of the proof, we will denote by E_m the divisor given by $z_m = 0$. Assume now that $\bar{\partial}\omega \in \mathscr{E}^{*}_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon})$. We will prove property (iii). We

write

$$\omega = \sum_{I,J} f_{I,J}\zeta_I \wedge \mathrm{d}\,\bar{z}_J,$$
$$\omega_1 = \sum_{I,j\in J} f_{I,J}\zeta_I \wedge \mathrm{d}\,\bar{z}_J,$$
$$\omega_2 = \sum_{I,j\notin J} f_{I,J}\zeta_I \wedge \mathrm{d}\,\bar{z}_J.$$

Recall that we have introduced the operator $\bar{\partial}_j$ in equation (2.37). We write $\bar{\partial}_{\neq j} = \bar{\partial} - \bar{\partial}_j$, and we decompose

$$\bar{\partial}K_j(\omega) = \bar{\partial}K_j(\omega_1) = \bar{\partial}_{\neq j}K_j(\omega_1) + \bar{\partial}_j K_j(\omega_1),$$

and

$$K_j(\bar{\partial}\omega) = K_j(\bar{\partial}_{\neq j}\omega_1 + \bar{\partial}_j\omega_2)$$

The difficulty at this point is that, when $k < j \leq l$, the form ω is log-log along E_j but, according to lemma 2.38, $\bar{\partial}_j\omega$ only needs to be log along E_j, and the integral operator K_j for log-log forms may diverge when applied to log forms. The key point is to observe that the extra hypothesis about $\bar{\partial}\omega$ allows us to apply the operator K_j to the differential forms $\bar{\partial}_{\neq j}\omega_1$ and $\bar{\partial}_j\omega_2$ individually: Fix I and J with $j \in J$ and $m \neq j$. We consider first the problematic case $k < j \leq l$. By lemma 2.38, we have

$$\bar{\partial}_m \omega_{I,J_m} \in \begin{cases} \mathscr{E}^*_X\langle D_1 \cup E_m\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon}), & \text{if } k < m \leq l, \\ \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon}), & \text{otherwise.} \end{cases}$$

Therefore, if we denote by D' the union of all the components of D different from E_j, then

$$(\bar{\partial}_{\neq j}\omega_1)_{I,J} \in \mathscr{E}^*_X\langle D'\langle E_j\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon}).$$

Since, by hypothesis, $(\bar{\partial}\omega)_{I,J} \in \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon})$ and $(\bar{\partial}_j\omega_2)_{I,J} = (\bar{\partial}\omega - \bar{\partial}_{\neq j}\omega_1)_{I,J}$, then

$$(\bar{\partial}_j\omega_2)_{I,J} \in \mathscr{E}^*_X\langle D'\langle E_j\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon}),$$

and we can apply the operator K_j for log-log forms to the differential forms $\bar{\partial}_{\neq j}\omega_1$ and $\bar{\partial}_j\omega_2$ individually. If $j \leq k$, then ω is log along E_j; the same is true for the differential forms $\bar{\partial}_{\neq j}\omega_1$ and $\bar{\partial}_j\omega_2$. But in this case the operator K_j is the operator for log forms and can be applied to $\bar{\partial}_{\neq j}\omega_1$ and $\bar{\partial}_j\omega_2$ individually. The case $j > l$ is similar. Thus, we can write

$$K_j(\bar{\partial}_{\neq j}\omega_1 + \bar{\partial}_j\omega_2) = K_j(\bar{\partial}_{\neq j}\omega_1) + K_j(\bar{\partial}_j\omega_2).$$

But by the theorem of taking derivatives under the integral sign, we now obtain

$$\bar{\partial}_{\neq j} K_j(\omega_1) + K_j(\bar{\partial}_{\neq j}\omega_1) = 0.$$

By lemma 2.44, we have

$$\bar{\partial}_j K_j(\omega_1) = \omega_1,$$

and by the generalized Cauchy integral formula (lemma 2.43 (ii)), we note

$$K_j(\bar{\partial}_j\omega_2) = \omega_2 - P_j(\omega_2) = \omega_2 - P_j(\omega).$$

Summing up, we obtain

$$\bar{\partial} K_j(\omega) + K_j(\bar{\partial}\omega) = \omega - P_j(\omega). \tag{2.57}$$

By (2.57) and the fact that $K_j(\bar{\partial}\omega), P_j(\omega) \in \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon'})$, we obtain that

$$\bar{\partial} K_j(\omega) \in \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon'}).$$

Assume now that $\partial\omega \in \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta_{x,\epsilon})$. We fix $I, J \subseteq \{1,\dots,d\}$, with $j \in J$. If $j \notin I$, then

$$(\partial K_j(\omega))_{I,J_j} = \sum_{m\neq j} \partial_m K_j(\omega_{I_m,J}) = K_j\left(\sum_{m\neq j} \partial_m \omega_{I_m,J}\right) = K_j((\partial\omega)_{I,J}).$$

Therefore, it belongs to $\mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon})$. If $j \in I$, we write

$$(\partial K_j(\omega))_{I,J_j} = \sum_{m\neq j} \partial_m K_j(\omega_{I_m,J}) + \partial_j K_j(\omega_{I_j,I}). \tag{2.58}$$

The theorem of taking derivatives under the integral sign implies for $m \neq j$

$$\partial_m K_j(\omega_{I_m,J}) = -K_j(\partial_m \omega_{I_m,J}).$$

Note that the term on the right hand side is well defined by lemma 2.38. We first treat the case $j \leq k$. We have to be careful because the integral kernels appearing in the expressions $\partial_j K_j(\omega_{I_j,J})$ and $K_j(\partial_j \omega_{I_j,J})$ are different in each term.

Again by lemma 2.38,

$$\partial_j \omega_{I_j,J} \in \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon}.$$

Since, moreover, $\partial\omega \in \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon})$,

$$\sum_{m\neq j} \partial_m \omega_{I_m,J} = (\partial\omega)_{I,J} - \partial_j \omega_{I_j,J} \in \mathscr{E}^*_X\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon}).$$

Hence, by lemma 2.44

$$K_j\left(\sum_{m\neq j}\partial_m\omega_{I_m,J}\right)\in\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon}).$$

By the same lemma it follows that

$$\partial_j K_j(\omega_{I_j,I})\in\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon}).$$

Now we treat the case $j>k$. In this case the expressions $\partial_j K_j(\omega_{I_j,J})$ and $K_j(\partial_j\omega_{I_j,J})$ use the same integral kernel. By lemma 2.43 (iii), we have

$$\partial_j K_j(\omega_{I_j,J})=-K_j(\partial_j\omega_{I_j,J})+\frac{(-1)^{\#I+\sigma(J,j)+\sigma(I,j)}}{2\pi\sqrt{-1}}\int_{\gamma_{\epsilon'}}f_{I_j,J}\frac{\mathrm{d}\,\bar w}{w-z}\zeta_I\wedge\mathrm{d}\,\bar z_{J_j}.$$

Hence, we arrive at

$$(\partial K_j(\omega))_{I,J_j}=-(K_j(\partial\omega))_{I,J_j}+\frac{(-1)^{\#I+\sigma(J,j)+\sigma(I,j)}}{2\pi\sqrt{-1}}\int_{\gamma_{\epsilon'}}f_{I_j,J}\frac{\mathrm{d}\,\bar w}{w-z}\zeta_I\wedge\mathrm{d}\,\bar z_{J_j}.$$

Thus, it belongs to $\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon})$.
Finally, assume that $\partial\omega,\bar\partial\omega,\partial\bar\partial\omega\in\mathscr{E}_X^*\langle D_1\langle D_2\rangle\rangle_{\mathrm{gth}}(\Delta^d_{x,\epsilon})$. By equation (2.57), we have

$$\partial\bar\partial K_j(\omega)=-\partial K_j(\bar\partial\omega)+\partial\omega-\partial P_j(\omega);$$

therefore, the result follows from the previous cases. □

2.5 Good forms.

In this section we recall the definition of good forms in the sense of [34]. We introduce the complex of Poincaré singular forms that is contained in both, the complex of good forms and the complex of log-log forms.

Poincaré growth. Let X, D, U, and ι be as in definition 2.2.

Definition 2.59. Let V be a coordinate neighborhood adapted to D. We say that a smooth complex function f on $V\setminus D$ has *Poincaré growth (along D)*, if it is bounded. We say that it has *Poincaré growth (along D) of infinite order*, if for all multi-indices $\alpha,\beta\in\mathbb{Z}^d_{\geq 0}$

$$\left|\frac{\partial^{|\alpha|}}{\partial z^\alpha}\frac{\partial^{|\beta|}}{\partial\bar z^\beta}f(z_1,\dots,z_d)\right|\prec\frac{1}{|z^{\alpha^{\leq k}}\bar z^{\beta^{\leq k}}|}.\tag{2.60}$$

The *sheaf of differential forms on X with Poincaré growth (resp. of infinite order)* is the subalgebra of $\iota_*\mathscr{E}_U^*$ generated, in each coordinate neighborhood V

adapted to D, by the functions with Poincaré growth (resp. of infinite order) and the differentials

$$\begin{aligned} &\frac{\mathrm{d}\, z_i}{z_i \log(1/r_i)}, \frac{\mathrm{d}\, \bar{z}_i}{\bar{z}_i \log(1/r_i)}, && \text{for } i = 1, \dots, k, \\ &\mathrm{d}\, z_i, \mathrm{d}\, \bar{z}_i, && \text{for } i = k+1, \dots, d. \end{aligned}$$

Good forms. We recall that a smooth form ω on $X \setminus D$ is *good (along D)*, if ω and $\mathrm{d}\,\omega$ have Poincaré growth along D (see [34]). Observe that, since the operator d is not bi-homogeneous, the sheaf of good forms is not bigraded. Although good forms are very similar to pre-log-log forms, there is no inclusion between both sheaves. Nevertheless, we have the following easy

Lemma 2.61. *If ω is a good form of pure bidegree, then it is a pre-log-log form, if and only if, $\partial\bar{\partial}\omega$ has log-log growth of order 0.* □

Poincaré singular forms.

Definition 2.62. We will say that ω is *Poincaré singular (along D)*, if ω, $\partial\omega$, $\bar{\partial}\omega$, and $\partial\bar{\partial}\omega$ have Poincaré growth of infinite order.

Note that the sheaf of Poincaré singular forms is contained in both, the sheaf of good forms and the sheaf of log-log forms. Observe moreover that we cannot expect to have a Poincaré lemma for the complex of Poincaré singular forms, precisely due to the absence of the functions $\log(\log(1/r_i))$.

Functoriality. The complex of Poincaré singular forms share some of the properties of the complex of log-log forms. For instance, we have the following compatibility with respect to inverse images which is proven as in proposition 2.24.

Proposition 2.63. *Let $f : X \longrightarrow Y$ be a morphism of complex manifolds of dimension d and d'. Let D_X, D_Y be normal crossing divisors on X, Y, respectively, satisfying $f^{-1}(D_Y) \subseteq D_X$. If η is a Poincaré singular form on Y, then $f^*\eta$ is a Poincaré singular form on X.* □

3 Arithmetic Chow rings with log-log growth conditions

In this section we use the theory of abstract cohomological arithmetic Chow rings developed in [10] to obtain a theory of arithmetic Chow rings with log-log forms. Since we have computed the cohomology of the complex of log-log forms, we have a more precise knowledge of the size of these arithmetic Chow rings than of the arithmetic Chow rings with pre-log-log forms considered in [10].

3.1 Dolbeault algebras and Deligne algebras

In this section we recall the notion of Dolbeault algebra and the properties of the associated Deligne algebra.

Dolbeault algebras.

Definition 3.1. A *Dolbeault algebra* $A = (A^*_{\mathbb{R}}, \mathrm{d}_A, \wedge)$ is a real differential graded commutative algebra which is bounded from below and equipped with a bigrading on $A_{\mathbb{C}} := A_{\mathbb{R}} \otimes \mathbb{C}$,

$$A^n_{\mathbb{C}} = \bigoplus_{p+q=n} A^{p,q},$$

satisfying the following properties:

(i) The differential d_A can be decomposed as the sum of operators $\mathrm{d}_A = \partial + \bar{\partial}$ of type $(1,0)$, resp. $(0,1)$.

(ii) It satisfies the symmetry property $\overline{A^{p,q}} = A^{q,p}$, where $\overline{}$ denotes complex conjugation.

(iii) The product induced on $A_{\mathbb{C}}$ is compatible with the bigrading:

$$A^{p,q} \wedge A^{p',q'} \subseteq A^{p+p',q+q'}.$$

By abuse of notation, we will also denote by A^* the complex differential graded commutative algebra $A^*_{\mathbb{C}}$.

Notation 3.2. Given a Dolbeault algebra A we will use the following notations. The Hodge filtration F of A^* is the decreasing filtration given by

$$F^p A^n_{\mathbb{C}} = \bigoplus_{p' \geq p} A^{p',n-p'}.$$

The filtration $\overline{F}$ is the complex conjugate of F, i.e.,

$$\overline{F}^p A^n = \overline{F^p A^n}.$$

For an element $x \in A$, we write $x^{i,j}$ for its component in $A^{i,j}$. For $k, k' \geq 0$, we define an operator $F^{k,k'} : A \longrightarrow A$ by the rule

$$F^{k,k'}(x) := \sum_{l \geq k, l' \geq k'} x^{l,l'}.$$

We note that the operator $F^{k,k'}$ is the projection of A^* onto the subspace $F^k A^* \cap \overline{F}^{k'} A^*$. We will write $F^k = F^{k,-\infty}$.
We denote by $A^n_{\mathbb{R}}(p)$ the subgroup $(2\pi i)^p \cdot A^n_{\mathbb{R}} \subseteq A^n$, and we define the operator

$$\pi_p : A \longrightarrow A_{\mathbb{R}}(p)$$

by setting $\pi_p(x) := \frac{1}{2}(x + (-1)^p \bar{x})$.

The Deligne complex.

Definition 3.3. Let A be a Dolbeault algebra. Then, the *Deligne complex* $(\mathcal{D}^*(A,*),\mathrm{d}_{\mathcal{D}})$ *associated to* A is the graded complex given by

$$\mathcal{D}^n(A,p)=\begin{cases} A_{\mathbb{R}}^{n-1}(p-1)\cap F^{n-p,n-p}A^{n-1}, & \text{if} \quad n\leq 2p-1,\\ A_{\mathbb{R}}^{n}(p)\cap F^{p,p}A^{n}, & \text{if} \quad n\geq 2p,\end{cases}$$

with differential given by $(x\in\mathcal{D}^n(A,p))$

$$\mathrm{d}_{\mathcal{D}}\,x=\begin{cases} -F^{n-p+1,n-p+1}\,\mathrm{d}_A\,x, & \text{if} \quad n<2p-1,\\ -2\partial\bar{\partial}x, & \text{if} \quad n=2p-1,\\ \mathrm{d}_A\,x, & \text{if} \quad n\geq 2p.\end{cases}$$

The Deligne algebra.

Definition 3.4. Let A be a Dolbeault algebra. The *Deligne algebra associated to* A is the Deligne complex $\mathcal{D}^*(A,*)$ together with the graded commutative product $\bullet:\mathcal{D}^n(A,p)\times\mathcal{D}^m(A,q)\longrightarrow\mathcal{D}^{n+m}(A,p+q)$, given by

$$x\bullet y=$$
$$\begin{cases} (-1)^n r_p(x)\wedge y+x\wedge r_q(y), & \text{if } n<2p,\ m<2q,\\ F^{l-r,l-r}(x\wedge y), & \text{if } n<2p,\ m\geq 2q,\ l<2r,\\ F^{r,r}(r_p(x)\wedge y)+2\pi_r(\partial(x\wedge y)^{r-1,l-r}), & \text{if } n<2p,\ m\geq 2q,\ l\geq 2r,\\ x\wedge y, & \text{if } n\geq 2p,\ m\geq 2q,\end{cases}$$

where we have written $l=n+m$, $r=p+q$, and $r_p(x)=2\pi_p(F^p\,\mathrm{d}_A\,x)$.

Specific degrees. In the sequel we will be interested in some specific degrees, where we can give simpler formulas. Namely, we consider

$$\begin{aligned}\mathcal{D}^{2p}(A,p)&=A_{\mathbb{R}}^{2p}(p)\cap A^{p,p},\\ \mathcal{D}^{2p-1}(A,p)&=A_{\mathbb{R}}^{2p-2}(p-1)\cap A^{p-1,p-1},\\ \mathcal{D}^{2p-2}(A,p)&=A_{\mathbb{R}}^{2p-3}(p-1)\cap(A^{p-2,p-1}\oplus A^{p-1,p-2}).\end{aligned}$$

The corresponding differentials are given by

$$\begin{aligned}\mathrm{d}_{\mathcal{D}}\,x&=\mathrm{d}_A\,x, && \text{if } x\in\mathcal{D}^{2p}(A,p),\\ \mathrm{d}_{\mathcal{D}}\,x&=-2\partial\bar{\partial}x, && \text{if } x\in\mathcal{D}^{2p-1}(A,p),\\ \mathrm{d}_{\mathcal{D}}(x,y)&=-\partial x-\bar{\partial}y, && \text{if } (x,y)\in\mathcal{D}^{2p-2}(A,p).\end{aligned}$$

Moreover, the product is given as follows: for $x\in\mathcal{D}^{2p}(A,p)$, $y\in\mathcal{D}^{2q}(A,q)$ or $y\in\mathcal{D}^{2q-1}(A,q)$, we have

$$x\bullet y=x\wedge y,$$

and for $x \in \mathcal{D}^{2p-1}(A,p)$, $y \in \mathcal{D}^{2q-1}(A,q)$, we have

$$x \bullet y = -\partial x \wedge y + \bar{\partial} x \wedge y + x \wedge \partial y - x \wedge \bar{\partial} y.$$

Deligne complexes and Deligne-Beilinson cohomology. The main interest in Deligne complexes is expressed by the following theorem which is proven in [8] in a particular case, although the proof is valid in general.

Theorem 3.5. *Let X be a complex algebraic manifold, $\overline{X}$ a smooth compactification of X with $D = \overline{X} \setminus X$ a normal crossing divisor, and denote by $j : X \longrightarrow \overline{X}$ the natural inclusion. Let $\mathscr{A}^*$ be a sheaf of Dolbeault algebras over $\overline{X}^{\mathrm{an}}$ such that, for every n, p the sheaves $\mathscr{A}^*$ and $F^p\mathscr{A}^*$ are acyclic, $\mathscr{A}^*_{\mathbb{R}}$ is a multiplicative resolution of $Rj_*\mathbb{R}$ and $(\mathscr{A}^*, F)$ is a multiplicative filtered resolution of $(\Omega^*_{\overline{X}}(\log D), F)$. Putting $A^* = \Gamma(\overline{X}, \mathscr{A}^*)$, we have a natural isomorphism of graded algebras*

$$H^*_{\mathcal{D}}(X, \mathbb{R}(p)) \cong H^*(\mathcal{D}(A,p)).$$

□

Notation. In the sequel we will use the following notation. The sheaves of differential forms will be denoted by the italic letter $\mathscr{E}$, and the corresponding spaces of global sections will be denoted by the same letter in roman typography E. For instance, we have

$$E^n_X\langle D_1\langle D_2\rangle\rangle = \Gamma(X, \mathscr{E}^n_X\langle D_1\langle D_2\rangle\rangle).$$

Logarithmic singularities at infinity Let X be a quasi-projective complex manifold. Let $E_{\log}(X)$ be the Dolbeault algebra of differential forms with logarithmic singularities at infinity (see [10], §5). Recall that in [10], $E_{\log}$ is defined as the Zariski sheaf associated to the pre-sheaf $E_{\log}{}^\circ$, which associates to any quasi-projective complex manifold X

$$E^*_{\log}(X)^\circ = \varinjlim E^*_{\overline{X}_\alpha}(\log D_\alpha),$$

where the limit is taken over all possible compactifications $\overline{X}_\alpha$ of X with $D_\alpha = \overline{X}_\alpha \setminus X$ a normal crossing divisor. Nevertheless, the step of taking the associated Zariski sheaf is not necessary by the following result. See [10], definition 3.1, for the definition of a totally acyclic sheaf.

Theorem 3.6. *For every pair of integers p, q, the pre-sheaf $E^{p,q}_{\log}{}^\circ$ is a totally acyclic sheaf.*

Proof. Let U and V be two open subsets of X. We have to prove the exactness of the sequence

$$0 \longrightarrow E^{p,q}_{\log}(U \cup V)^\circ \xrightarrow{\phi} E^{p,q}_{\log}(U)^\circ \oplus E^{p,q}_{\log}(V)^\circ \xrightarrow{\psi} E^{p,q}_{\log}(U \cap V)^\circ \longrightarrow 0.$$

The injectivity of ϕ and the fact that $\psi \circ \phi = 0$ are obvious.
Put $Y = X \setminus U$ and $Z = X \setminus V$. Let $\pi_{Y\cap Z} : \widetilde{X}_{Y\cap Z} \longrightarrow X$ be an embedded resolution of singularities of $Y \cap Z$ such that the strict transform of Y, denoted by $\widehat{Y}$, and the strict transform of Z, denoted by $\widehat{Z}$, do not meet. Let $\{\sigma_{Y,Z}, \sigma_{Z,Y}\}$ be a partition of unity subordinate to the open cover $\{\widetilde{X} \setminus \widehat{Z}, \widetilde{X} \setminus \widehat{Y}\}$. If $\omega \in E^{p,q}_{\log}(U \cap V)^{\circ}$, then $\sigma_{Y,Z}\omega \in E^{p,q}_{\log}(U)^{\circ}$ and $\sigma_{Z,Y}\omega \in E^{p,q}_{\log}(V)^{\circ}$. Therefore, we get

$$\omega = \psi(-\sigma_{Y,Z}\omega, \sigma_{Z,Y}\omega),$$

which proves the surjectivity of ψ.
Let now $(\omega, \eta) \in E^{p,q}_{\log}(U)^{\circ} \oplus E^{p,q}_{\log}(V)^{\circ}$ be such that $\psi(\omega, \eta) = 0$. Then, ω and η agree on $U \cap V$. Therefore, they define a smooth form on $U \cup V$; by abuse of notation, we denote it by ω. The subtle point here is to know that, after some blow-ups with centers contained in Y, ω will have logarithmic singularities along the exceptional divisor, and the same is true after some blow-ups with centers contained in Z. We have to prove that ω has logarithmic singularities after blowing-up only centers contained in $Y \cap Z$.
To this end we need the following easy lemma, which follows from Hironaka's resolution of singularities.

Lemma 3.7. *Let X be a regular variety over a field of characteristic zero and let C_1 and C_2 be two closed subsets. Let $\pi : \widetilde{X} \longrightarrow X$ be a proper birational morphism, which is an isomorphism in the complement of $C_1 \cup C_2$. Then, there is a factorization*

$$\begin{tikzcd}
 & \widetilde{X}_2 \arrow[dl, "\pi_2"'] \arrow[dr] & \\
\widetilde{X}_1 \arrow[dr, "\pi_1"'] & & \widetilde{X} \arrow[dl, "\pi"] \\
 & X &
\end{tikzcd}$$

where $\widetilde{X}_1$ and $\widetilde{X}_2$ are regular varieties, π_1 and π_2 are proper birational morphisms, π_1 is an isomorphism over the complement of C_1 and π_2 is an isomorphism over the complement of the strict transform of C_2 in $\widetilde{X}_1$. Moreover, it is possible to choose the factorization in such a way that $\pi_1^{-1}(C_1)$ and $(\pi_2^{-1} \circ \pi_1^{-1})(C_1 \cup C_2)$ are normal crossing divisors. □

Let $\pi_{Y\cap Z} : \widetilde{X}_{Y\cap Z} \longrightarrow X$ be as before, and denote by $D_{Y\cap Z}$ the exceptional divisor. Since $\omega \in E^{p,q}_{\log}(U)^{\circ}$, there exists an embedded resolution of singularities $\widetilde{X}_Y$ of Y with exceptional divisor D_Y, which we can assume to factor through a proper birational morphism $\widetilde{X}_Y \longrightarrow \widetilde{X}_{Y\cap Z}$, and $\omega \in E^{p,q}_{\widetilde{X}_Y}(\log D_Y)$. We apply the previous lemma to the morphism $\widetilde{X}_Y \longrightarrow \widetilde{X}_{Y\cap Z}$ and the closed subsets

$D_{Y\cap Z}$ and $\widehat{Y}$. In this way we obtain a diagram

$$\begin{tikzcd} & \widetilde{X}'_Y \arrow[dl] \arrow[dr] & \\ \widetilde{X}'_{Y\cap Z} \arrow[dr, "\pi'_{Y\cap Z}"'] & & \widetilde{X}_Y \arrow[dl] \\ & \widetilde{X}_{Y\cap Z} & \end{tikzcd}$$

In $\widetilde{X}'_{Y\cap Z}$ we denote by $\widehat{Y}'$ and by $\widehat{Z}'$ the strict transforms of Y and Z, respectively, and by $D'_{Y\cap Z}$ the exceptional divisor. Now, since $\omega \in E^{p,q}_{\log}(V)^{\circ}$, we can repeat the process. There exists an embedded resolution of singularities $\widetilde{X}'_Z$ of Z in X with exceptional divisor D'_Z that factors through a proper birational morphism $\widetilde{X}'_Z \longrightarrow \widetilde{X}'_{Y\cap Z}$. Then, $\omega \in E^{p,q}_{\widetilde{X}'_Z}(\log D'_Z)$. We apply the previous lemma to this last morphism and the closed subsets $D'_{Y\cap Z}$ and $\widehat{Z}'$ to obtain the diagram

$$\begin{tikzcd} & \widetilde{X}''_Z \arrow[dl] \arrow[dr] & \\ \widetilde{X}''_{Y\cap Z} \arrow[dr] & & \widetilde{X}'_Z \arrow[dl] \\ & \widetilde{X}'_{Y\cap Z} & \end{tikzcd}$$

In $\widetilde{X}''_{Y\cap Z}$ we denote by $\widehat{Y}''$ and $\widehat{Z}''$ the strict transforms of Y and Z, respectively, and by $D''_{Y\cap Z}$ the exceptional divisor. To conclude the proof of the theorem, it is enough to show that

$$\omega \in E^{p,q}_{\widetilde{X}''_{Y\cap Z}}(\log D''_{Y\cap Z}).$$

This condition can be checked locally.
If $x \notin D''_{Y\cap Z}$, by hypothesis, ω_x is the germ of a smooth form.
Assume now that $x \in D''_{Y\cap Z} \setminus \widehat{Z}''$. We write D'_Z and D''_Z for the preimages of Z in $\widetilde{X}'_Z$ and $\widetilde{X}''_Z$, respectively. By construction, both are normal crossing divisors. By hypothesis, $\omega \in E^{p,q}_{\widetilde{X}'}(\log D'_Z)$. By the functoriality of logarithmic singularities, $\omega \in E^{p,q}_{\widetilde{X}''}(\log D''_Z)$. Let W be a neighborhood of x, whose intersection with $\widehat{Z}''$ is empty. Therefore, it is isomorphic to an open subset of $\widetilde{X}''_Z$, hence

$$\omega|_W \in \Gamma(W, \mathscr{E}^{p,q}_{\widetilde{X}''_{Y\cap Z}}(\log D''_{Y\cap Z})) = \Gamma(W, \mathscr{E}^{p,q}_{\widetilde{X}''_Z}(\log D''_Z)).$$

Finally, if $x \in D''_{Y\cap Z} \cap \widehat{Z}''$, we use a similar argument. $\square$

Remark 3.8. The argument of the previous theorem applies also to the complex $E_{\mathrm{pre}}(\underline{X})$ of [10], definition 7.16. Therefore, it that case, the morphism between the pre-sheaf and the associated sheaf is an isomorphism. Observe moreover that the same argument will apply to all the Zariski sheaves that we will introduce in this paper.

The Deligne complex with logarithmic singularities. We will denote

$$\mathcal{D}^*_{\log}(X,p)=\mathcal{D}^*(E_{\log}(X),p).$$

Then, theorem 3.5 implies that

$$H^*_{\mathcal{D}}(X,\mathbb{R}(p))\cong H^*(\mathcal{D}_{\log}(X,p)).$$

3.2 The $\mathcal{D}_{\log}$-complex of log-log forms

$\mathcal{D}_{\log}$-complexes. Recall that, to define the arithmetic Chow groups of an arithmetic variety X as in [10], we need first an auxiliary complex of graded abelian sheaves on the Zariski site of smooth real schemes that satisfies Gillet axioms. As in [10], we will use the complex of sheaves $\mathcal{D}_{\log}$. This sheaf is given, for any smooth real scheme $U_{\mathbb{R}}$, by

$$\mathcal{D}_{\log}(U_{\mathbb{R}},p)=\mathcal{D}_{\log}(U_{\mathbb{C}},p)^{\sigma},$$

where σ is the involution that acts as complex conjugation on the space and on the coefficients (see [10], §5.3).
Then, we need to choose a $\mathcal{D}_{\log}$-complex over $X_{\mathbb{R}}$. That is, a complex $\mathcal{C}^*_{X_{\mathbb{R}}}(*)$ of graded abelian sheaves on the Zariski topology of $X_{\mathbb{R}}$ together with a morphism

$$\mathcal{D}_{\log,X_{\mathbb{R}}}\longrightarrow\mathcal{C}_{X_{\mathbb{R}}}$$

such that all the sheaves $\mathcal{C}^n_{X_{\mathbb{R}}}(p)$ are totally acyclic (see [10], definitions 3.1 and 3.4). The $\mathcal{D}_{\log}$-complex $\mathcal{C}$ plays the role of the fiber over the archimedean places of the arithmetic ring A. The aim of this section is to construct a $\mathcal{D}_{\log}$-complex by mixing log and log-log forms.

Varieties with a fixed normal crossing divisor. We will follow the notations of [10], §7.4, that we recall shortly. Let X be a complex algebraic manifold of dimension d, and D a normal crossing divisor. We will denote by $\underline{X}$ the pair (X,D). If $W\subseteq X$ is an open subset, we will write $\underline{W}=(W,D\cap W)$. In the sequel we will consider all operations adapted to the pair $\underline{X}$. For instance, if $Y\subsetneq X$ is a closed algebraic subset and $W=X\setminus Y$, then an embedded resolution of singularities of Y in $\underline{X}$ is a proper modification $\pi:\widetilde{X}\longrightarrow X$ such that $\pi\big|_{\pi^{-1}(W)}:\pi^{-1}(W)\longrightarrow W$ is an isomorphism, and

$$\pi^{-1}(Y),\ \pi^{-1}(D),\ \pi^{-1}(Y\cup D)$$

are normal crossing divisors on $\widetilde{X}$. Using Hironaka's theorem on the resolution of singularities [25], one can see that such an embedded resolution of singularities exists.
Analogously, a normal crossing compactification of $\underline{X}$ will be a smooth compactification $\overline{X}$ such that the adherence $\overline{D}$ of D, the subsets $B_{\overline{X}} = \overline{X} \setminus X$ and $B_{\overline{X}} \cup \overline{D}$ are normal crossing divisors.

Logarithmic growth along infinity. Given a diagram of normal crossing compactifications of $\underline{X}$

$$\begin{tikzcd} \overline{X}' \arrow[r, "\varphi"] & \overline{X} \\ & X \arrow[ul] \arrow[u] \end{tikzcd}$$

with divisors $B_{\overline{X}'}$ and $B_{\overline{X}}$ at infinity, respectively, proposition 2.40 gives rise to an induced morphism

$$\varphi^* : \mathscr{E}^*_{\overline{X}}\langle B_{\overline{X}}\langle \overline{D}\rangle\rangle \longrightarrow \mathscr{E}^*_{\overline{X}'}\langle B_{\overline{X}'}\langle \overline{D}'\rangle\rangle.$$

In order to have a complex that is independent of the choice of a particular compactification we take the limit over all possible compactifications.

Definition 3.9. Let $\underline{X} = (X, D)$ be as above. Then, we define the *complex* $E^*_{\mathrm{l,ll}}(\underline{X})$ *of differential forms on* X *log along infinity and log-log along* D as

$$E^*_{\mathrm{l,ll}}(\underline{X}) = \varinjlim \Gamma(\overline{X}, \mathscr{E}^*_{\overline{X}}\langle B_{\overline{X}}\langle \overline{D}\rangle\rangle),$$

where the limit is taken over all normal crossing compactifications $\overline{X}$ of $\underline{X}$.

A $\mathcal{D}_{\log}$-complex. Let X be a smooth real variety and D a normal crossing divisor defined over $\mathbb{R}$; as before, we write $\underline{X} = (X, D)$. For any $U \subseteq X$, the complex $E^*_{\mathrm{l,ll}}(\underline{U}_{\mathbb{C}})$ is a Dolbeault algebra with respect to the wedge product.

Definition 3.10. For any Zariski open subset $U \subseteq X$, we put

$$\mathcal{D}^*_{\mathrm{l,ll},\underline{X}}(U,p) = (\mathcal{D}^*_{\mathrm{l,ll},\underline{X}}(U,p), \mathrm{d}_{\mathcal{D}}) = (\mathcal{D}^*(E_{\mathrm{l,ll}}(\underline{U}_{\mathbb{C}}), p)^{\sigma}, \mathrm{d}_{\mathcal{D}}),$$

where the operator $\mathcal{D}$ is as in definition 3.4 and σ is the involution that acts as complex conjugation in the space and in the coefficients (see [10], 5.55). When the pair $\underline{X}$ is understood, we write $\mathcal{D}^*_{\mathrm{l,ll}}$ instead of $\mathcal{D}^*_{\mathrm{l,ll},\underline{X}}$. The complex $\mathcal{D}^*_{\mathrm{l,ll}}$ will be called *the* $\mathcal{D}_{\log}$*-complex of log-log forms* or just *the complex of log-log forms.*

Then, the analogue of [10], theorem 7.18, holds.

Theorem 3.11. *The complex* $\mathcal{D}_{\mathrm{l,ll},\underline{X}}$ *is a* $\mathcal{D}_{\log}$*-complex on* X*. Moreover, it is a pseudo-associative and commutative* $\mathcal{D}_{\log}$*-algebra.* □

The cohomology of the complex $\mathcal{D}_{\mathrm{l,ll},\underline{X}}$. The main advantage of the complex $\mathcal{D}_{\mathrm{l,ll},\underline{X}}$ over the complex $\mathcal{D}_{\mathrm{pre},\underline{X}}$ of [10] is the following result that is a consequence of theorem 2.42 and theorem 3.5 (see [10], theorem 5.19, and [8]).

Theorem 3.12. *The inclusion $\mathcal{D}_{\log,X} \longrightarrow \mathcal{D}_{\mathrm{l,ll},\underline{X}}$ is a quasi-isomorphism. Therefore, the hypercohomology over X of the complex of sheaves $\mathcal{D}_{\mathrm{l,ll},\underline{X}}$, as well as the cohomology of its complex of global sections, is naturally isomorphic to the Deligne-Beilinson cohomology of X.* □

3.3 Properties of Green objects with values in $\mathcal{D}_{\mathrm{l,ll}}$.

We start by noting that theorem 3.11 together with [10], section 3, provides us with a theory of Green objects with values in $\mathcal{D}_{\mathrm{l,ll},\underline{X}}$.

Mixed forms representing the class of a cycle. Since we know the cohomology of the complex of mixed forms, we obtain the analogue of proposition 5.48 in [10], which is more precise than the analogue of proposition 7.20 in [10]. In particular, we have

Proposition 3.13. *Let X be a smooth real variety and D a normal crossing divisor. Put $\underline{X} = (X, D)$. Let y be a p-codimensional cycle on X with support Y. Then, we have that the class of the cycle (ω, g) in $H^{2p}_{\mathcal{D}_{\mathrm{l,ll}},Y}(X,p)$ is equal to the class of y, if and only if*

$$-2\partial\bar{\partial}[g]_X = [\omega] - \delta_y. \tag{3.14}$$

Proof. The proof is completely analogous to the proof of [10], 5.48, using theorem 3.12 and lemma 2.41. □

Inverse images.

Proposition 3.15. *Let $f : X \longrightarrow Y$ be a morphism of smooth real varieties, let D_X, D_Y be normal crossing divisors on X, Y respectively, satisfying $f^{-1}(D_Y) \subseteq D_X$. Put $\underline{X} = (X, D_X)$ and $\underline{Y} = (Y, D_Y)$. Then, there exists a contravariant f-morphism*

$$f^{\#} : \mathcal{D}_{\mathrm{l,ll},\underline{Y}} \longrightarrow f_*\mathcal{D}_{\mathrm{l,ll},\underline{X}}.$$

Proof. By proposition 2.40, the pull-back of differential forms induces a morphism of the corresponding Dolbeault algebras of mixed forms. This morphism is compatible with the involution σ. Thus, this morphism gives rise to an induced morphism between the corresponding Deligne algebras. □

Push-forward. We will only state the most basic property concerning direct images, which is necessary to define arithmetic degrees. Note however that we expect that the complex of log-log forms will be useful in the study of non smooth, proper, surjective morphisms. By proposition 2.26, we have

Proposition 3.16. *Let $\underline{X} = (X, D)$ be a proper, smooth real variety with fixed normal crossing divisor D. Let $f : X \longrightarrow \operatorname{Spec}(\mathbb{R})$ denote the structural morphism. Then, there exists a covariant f-morphism*

$$f_\# : f_* \mathcal{D}_{\mathrm{l,ll},\underline{X}} \longrightarrow \mathcal{D}_{\log,\operatorname{Spec}(\mathbb{R})}.$$

In particular, if X has dimension d, we obtain a well defined morphism

$$f_\# : \widehat{H}^{2d+2}_{\mathcal{D}_{\mathrm{l,ll}},\mathcal{Z}^{d+1}}(X, d) \longrightarrow \widehat{H}^{2}_{\mathcal{D}_{\log},\mathcal{Z}^{1}}(\operatorname{Spec}(\mathbb{R}), 1) = \mathbb{R}.$$

Note that, by dimension reasons, we have $\mathcal{Z}^{d+1} = \emptyset$, and

$$\widehat{H}^{2d+2}_{\mathcal{D}_{\mathrm{l,ll}},\mathcal{Z}^{d+1}}(X, d) = H^{2d+1}(\mathcal{D}_{\mathrm{l,ll}}(X, d+1)) = H^{2d+1}_{\mathcal{D}}(X, \mathbb{R}(d+1).$$

Thus, every element of $\widehat{H}^{2d+2}_{\mathcal{D}_{\mathrm{l,ll}},\mathcal{Z}^{d+1}}(X, d)$ is represented by a pair $\mathfrak{g} = (0, \widetilde{g})$. The morphism $f_\#$ mentioned above, is then given by

$$\mathfrak{g} = (0, \widetilde{g}) \longmapsto \left(0, \frac{1}{(2\pi i)^d} \int_X g\right).$$

3.4 Arithmetic Chow rings with log-log forms

Arithmetic Chow groups. We are now in position to apply the machinery of [10]. Let (A, Σ, F_∞) be an arithmetic ring and let X be a regular arithmetic variety over A. Let D be a fixed normal crossing divisor of X_Σ stable under F_∞. As in the previous section, we will denote by $\underline{X}$ the pair $(X_{\mathbb{R}}, D)$. The natural inclusion $\mathcal{D}_{\log} \longrightarrow \mathcal{D}_{\mathrm{l,ll}}$ induces a $\mathcal{D}_{\log}$-complex structure in $\mathcal{D}_{\mathrm{l,ll}}$. Then, $(X, \mathcal{D}_{\mathrm{l,ll}})$ is a $\mathcal{D}_{\log}$-arithmetic variety. Therefore, applying the theory of [10], section 4, we define the arithmetic Chow groups $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll}})$. These groups will be called log-log arithmetic Chow groups.

Exact sequences. We start the study of these arithmetic Chow groups by writing the exact sequences of [10], theorem 4.13. Observe that, since we have better control on the cohomology of $\mathcal{D}_{\mathrm{l,ll}}$, we obtain better results than in [10], §7.

Theorem 3.17. *The following sequences are exact:*

$$\mathrm{CH}^{p-1,p}(X) \xrightarrow{\rho} \widetilde{\mathcal{D}}^{2p-1}_{\mathrm{l,ll}}(X, p) \xrightarrow{\mathrm{a}} \widehat{\mathrm{CH}}^p(X, \mathcal{D}_{\mathrm{l,ll}}) \xrightarrow{\zeta} \mathrm{CH}^p(X) \longrightarrow 0,$$

$$\mathrm{CH}^{p-1,p}(X) \xrightarrow{\rho} H^{2p-1}_{\mathcal{D}}(X_{\mathbb{R}}, \mathbb{R}(p)) \xrightarrow{\mathrm{a}} \widehat{\mathrm{CH}}^p(X, \mathcal{D}_{\mathrm{l,ll}}) \xrightarrow{(\zeta,-\omega)}$$
$$\mathrm{CH}^p(X) \oplus \mathrm{Z}\mathcal{D}^{2p}_{\mathrm{l,ll}}(X, p) \xrightarrow{\mathrm{cl}+h} H^{2p}_{\mathcal{D}}(X_{\mathbb{R}}, \mathbb{R}(p)) \longrightarrow 0,$$

$$\mathrm{CH}^{p-1,p}(X) \xrightarrow{\rho} H^{2p-1}_{\mathcal{D}}(X_{\mathbb{R}}, \mathbb{R}(p)) \xrightarrow{\mathrm{a}} \widehat{\mathrm{CH}}^p(X, \mathcal{D}_{\mathrm{l,ll}})_0 \xrightarrow{\zeta} \mathrm{CH}^p(X)_0 \longrightarrow 0.$$

□

Multiplicative properties. Since $\mathcal{D}_{\mathrm{l,ll}}$ is a pseudo-associative and commutative $\mathcal{D}_{\log}$-algebra, we have

Theorem 3.18. *The abelian group*

$$\widehat{\mathrm{CH}}^{*}(X,\mathcal{D}_{\mathrm{l,ll}})_{\mathbb{Q}}=\bigoplus_{p\geq 0}\widehat{\mathrm{CH}}^{p}(X,\mathcal{D}_{\mathrm{l,ll}})\otimes\mathbb{Q}$$

is an associative and commutative $\mathbb{Q}$-algebra with a unit. □

Inverse images. By proposition 2.40, there are some cases, where we can define the inverse image for the log-log arithmetic Chow groups.

Theorem 3.19. *Let $f : X \longrightarrow Y$ be a morphism of arithmetic varieties over A. Let E be a normal crossing divisor on $Y_{\mathbb{R}}$ and D a normal crossing divisor on $X_{\mathbb{R}}$ such that $f^{-1}(E)\subseteq D$. Write $\underline{X}=(X_{\mathbb{R}},D)$ and $\underline{Y}=(Y_{\mathbb{R}},E)$. Then, there is defined an inverse image morphism*

$$f^{*}:\widehat{\mathrm{CH}}^{*}(Y,\mathcal{D}_{\mathrm{l,ll}})\longrightarrow\widehat{\mathrm{CH}}^{*}(X,\mathcal{D}_{\mathrm{l,ll}}).$$

Moreover, it is a morphism of rings after tensoring with $\mathbb{Q}$. □

Push-forward. We will state only the consequence of the integrability of log-log forms.

Theorem 3.20. *If X is projective over A, then there is a direct image morphism of groups*

$$f_{*}:\widehat{\mathrm{CH}}^{d+1}(X,\mathcal{D}_{\mathrm{l,ll}})\longrightarrow\widehat{\mathrm{CH}}^{1}(\operatorname{Spec}A),$$

where d is the relative dimension of X. □

Relationship with other arithmetic Chow groups. Since we know the cohomology of the complex $\mathcal{D}_{\mathrm{l,ll}}$, we can make a comparison statement more precise than in [10], theorem 6.23.

Theorem 3.21. *The structural morphism*

$$\mathcal{D}_{\log,X}\longrightarrow\mathcal{D}_{\mathrm{l,ll},\underline{X}}$$

induces a morphism

$$\widehat{\mathrm{CH}}^{*}(X,\mathcal{D}_{\log})\longrightarrow\widehat{\mathrm{CH}}^{*}(X,\mathcal{D}_{\mathrm{l,ll}})$$

that is compatible with inverse images, intersection products and arithmetic degrees. If X is projective, the isomorphism between $\widehat{\mathrm{CH}}^{*}(X,\mathcal{D}_{\log})$ and the *arithmetic Chow groups defined by Gillet and Soulé (denoted by* $\widehat{\mathrm{CH}}^*(X)$*) induce morphisms*

$$\widehat{\mathrm{CH}}^*(X) \longrightarrow \widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll}}) \tag{3.22}$$

also compatible with inverse images, intersection products and arithmetic degrees. Moreover, if D is empty and X is projective, then the above morphisms are isomorphisms. □

3.5 The $\mathcal{D}_{\log}$-complex of log-log forms with arbitrary singularities at infinity

The arithmetic Chow groups defined by Gillet and Soulé for quasi-projective varieties use differential forms with arbitrary singularities in the boundary. Therefore, in order to be able to recover the arithmetic Chow groups of Gillet and Soulé, we have to introduce another variant of arithmetic Chow groups, where we allow the differential forms to have arbitrary singularities in certain directions.

Mixing log, log-log and arbitrary singularities. Let X be a complex algebraic manifold and D a fixed normal crossing divisor of X. We write $\underline{X} = (X, D)$.

Definition 3.23. For every Zariski open subset U of X, we write

$$E^*_{\mathrm{l,ll,a},\underline{X}}(U) = \varinjlim_{\overline{U}} \Gamma(\overline{U}, \mathscr{E}^*_{\overline{U}}\langle B_{\overline{U}}\langle \overline{D}\rangle\rangle),$$

where the limit is taken over all diagrams

$$\begin{tikzcd} U \arrow[r, "\bar{\iota}"] \arrow[dr, "\iota"'] & \overline{U} \arrow[d, "\beta"] \\ & X \end{tikzcd}$$

such that $\bar{\iota}$ is an open immersion, β is a proper morphism and $B_{\overline{U}} = \overline{U} \setminus U$, $\overline{D} = \beta^{-1}(D)$, $B_{\overline{U}} \cup \overline{D}$ are normal crossing divisors.

Definition 3.24. Let X be a complex algebraic manifold and D a fixed normal crossing divisor of X. We write $\underline{X} = (X, D)$ as before. For any Zariski open subset $U \subseteq X$, we put

$$\mathcal{D}^*_{\mathrm{l,ll,a},\underline{X}}(U,p) = (\mathcal{D}^*_{\mathrm{l,ll,a},\underline{X}}(U,p), \mathrm{d}_{\mathcal{D}}) = (\mathcal{D}^*(E_{\mathrm{l,ll,a},\underline{X}}(U_{\mathbb{C}}), p), \mathrm{d}_{\mathcal{D}}).$$

If X is a smooth algebraic variety over $\mathbb{R}$, and D, U are defined over $\mathbb{R}$, we put

$$\mathcal{D}^*_{\mathrm{l,ll,a},\underline{X}}(U,p) = (\mathcal{D}^*_{\mathrm{l,ll,a},\underline{X}}(U,p), \mathrm{d}_{\mathcal{D}}) = (\mathcal{D}^*(E_{\mathrm{l,ll,a},\underline{X}}(U_{\mathbb{C}}), p)^{\sigma}, \mathrm{d}_{\mathcal{D}}),$$

where σ is as in section 3.2.

Note that, when X is quasi-projective, the varieties $\overline{U}$ of definition 3.23 are not compactifications of U, but only *partial compactifications*. Therefore, the sections of $\mathcal{D}^{*}_{\mathrm{l,ll,a},\underline{X}}(U,p)$ have three different kinds of singularities. We can see this more concretely as follows. Let Y be a closed subset of X with $U = X \setminus Y$, and let $\overline{X}$ be a smooth compactification of X with $Z = \overline{X} \setminus X$. Let η be a section of $\mathcal{D}^{*}_{\mathrm{l,ll,a},\underline{X}}(U,p)$. If we consider η as a singular form on $\overline{X}$, then η is log along Y (in the sense that it is log along a certain resolution of singularities of Y), log-log along D and has arbitrary singularities along Z. Therefore, in general, we have

$$\mathcal{D}^{*}_{\mathrm{l,ll,a},\underline{X}}(U,p) \neq \mathcal{D}^{*}_{\mathrm{l,ll,a},\underline{U}}(U,p).$$

Nevertheless, when $\underline{X}$ is clear from the context, we will drop it from the notation.

Remark 3.25. If X is projective, the complexes of sheaves $\mathcal{D}^{*}_{\mathrm{l,ll,a},\underline{X}}$ and $\mathcal{D}^{*}_{\mathrm{l,ll},\underline{X}}$ agree. In contrast, they do not agree, when X is quasi-projective. Note, moreover, that, when X is quasi-projective, the complex $\mathcal{D}^{*}_{\mathrm{l,ll,a},\underline{X}}$ does not compute the Deligne-Beilinson cohomology of X, but a mixture between Deligne-Beilinson cohomology and analytic Deligne cohomology. Nevertheless, as we will see, the local nature of the purity property of Deligne-Beilinson cohomology implies also a purity property for these complexes.

Logarithmic singularities and Blow-ups. Let X be a complex manifold, $D \subseteq X$ a normal crossing divisor, and $Y \subseteq X$ an e-codimensional smooth subvariety such that the pair (D, Y) has normal crossings. Let $\pi : \widetilde{X} \longrightarrow X$ be the blow-up of X along Y. Write $\widetilde{D} = \pi^{-1}(D)$ and $\widetilde{Y} = \pi^{-1}(Y)$. Let $i : Y \longrightarrow X$ and $j : \widetilde{Y} \longrightarrow \widetilde{X}$ denote the inclusions, and let $g : \widetilde{Y} \longrightarrow Y$ denote the induced morphism. Observe that g is a projective bundle.

Proposition 3.26. *Let $p \geq 0$ be an integer. Then, we have:*

(i) If $Y \subseteq D$, then the morphism $\Omega^p_X(\log D) \longrightarrow R\pi_*\Omega^p_{\widetilde{X}}(\log \widetilde{D})$ is a quasi-isomorphism, i.e.,

$$\begin{aligned} \pi_*\Omega^p_{\widetilde{X}}(\log \widetilde{D}) &\cong \Omega^p_X(\log D), \quad \textit{and} \\ R^q\pi_*\Omega^p_{\widetilde{X}}(\log \widetilde{D}) &= 0, \qquad\qquad \textit{for } q > 0. \end{aligned}$$

(ii) If $Y \not\subseteq D$ and $e > 1$, then

$$\begin{aligned} \pi_*\Omega^p_{\widetilde{X}}(\log \widetilde{D} \cup \widetilde{Y}) &\cong \Omega^p_X(\log D), \\ R^q\pi_*\Omega^p_{\widetilde{X}}(\log \widetilde{D} \cup \widetilde{Y}) &= 0, \qquad \textit{for } q \neq 0, e-1, \textit{ and} \\ R^{e-1}\pi_*\Omega^p_{\widetilde{X}}(\log \widetilde{D} \cup \widetilde{Y}) &\cong i_*(R^{e-1}g_*\Omega^{p-1}_{\widetilde{Y}}(\log \widetilde{D} \cap \widetilde{Y})) \\ &\cong i_*(\Omega^{p-e}_Y(\log D \cap Y) \otimes R^{e-1}g_*\Omega^{e-1}_{\widetilde{Y}/Y}). \end{aligned}$$

(iii) If $Y \not\subseteq D$ and $e = 1$, then $\pi = \mathrm{id}$, and there is a short exact sequence

$$0 \longrightarrow \Omega^p_X \longrightarrow \Omega^p_X(\log Y) \longrightarrow i_* \Omega^{p-1}_Y \longrightarrow 0.$$

Proof. The third statement is standard; the first statement is [19], Proposition 4.4 (ii).
Using [19], Proposition 4.4 (i), the fact that i_* is an exact functor and that g is a projective bundle, we obtain

$$\begin{aligned}
\pi_* \Omega^p_{\widetilde{X}}(\log \widetilde{D}) &\cong \Omega^p_X(\log D),\\
\pi_* j_* \Omega^{p-1}_{\widetilde{Y}}(\log \widetilde{D} \cap \widetilde{Y}) &\cong i_* g_* \Omega^{p-1}_{\widetilde{Y}}(\log \widetilde{D} \cap \widetilde{Y})\\
&\cong i_* \Omega^{p-1}_Y(\log D \cap Y),\\
R^q \pi_* \Omega^p_{\widetilde{X}}(\log \widetilde{D}) &\cong R^q(\pi \circ j)_* \Omega^p_{\widetilde{Y}}(\log \widetilde{D} \cap \widetilde{Y})\\
&\cong i_* R^q g_* \Omega^p_{\widetilde{Y}}(\log \widetilde{D} \cap \widetilde{Y})\\
&\cong \begin{cases} i_*(\Omega^{p-q}_Y(\log D \cap Y) \otimes R^q g_* \Omega^q_{\widetilde{Y}/Y}), & \text{if } 1 \leq q < e,\\ 0, & \text{if } g \geq e. \end{cases}
\end{aligned}$$

Let $\mathscr{O}(1)$ be the ideal sheaf of $\widetilde{Y}$ in $\widetilde{X}$. We consider the exact sequence

$$0 \longrightarrow \Omega^p_{\widetilde{X}}(\log \widetilde{D}) \longrightarrow \Omega^p_{\widetilde{X}}(\log \widetilde{D} \cup \widetilde{Y}) \xrightarrow{\mathrm{Res}} j_* \Omega^{p-1}_{\widetilde{Y}}(\log \widetilde{D} \cap \widetilde{Y}) \longrightarrow 0$$

and the corresponding long exact sequence obtained by applying the functor $R\pi_*$. The connecting morphism of this long exact sequence

$$\begin{aligned}
R^{q-1}\pi_* j_* \Omega^{p-1}_{\widetilde{Y}}(\log \widetilde{D} \cap \widetilde{Y}) &\cong i_*(\Omega^{p-q}_Y(\log D \cap Y) \otimes R^{q-1} g_* \Omega^{q-1}_{\widetilde{Y}/Y})\\
\longrightarrow R^q \pi_* \Omega^p_{\widetilde{X}}(\log \widetilde{D}) &\cong i_*(\Omega^{p-q}_Y(\log D \cap Y) \otimes R^q g_* \Omega^q_{\widetilde{Y}/Y})
\end{aligned}$$

can be identified with the product by $\mathrm{c}_1(\mathscr{O}_{\widetilde{X}}(1))$, which is an isomorphism for $0 < q \leq e - 1$. The result now follows from this exact sequence. □

This proposition has the following consequence.

Corollary 3.27. *Let X be a complex algebraic manifold and Y a complex subvariety of codimension e. Let $\widetilde{X} \longrightarrow X$ be an embedded resolution of singularities of Y obtained as in [25]. Then, we have*

$$R^q \pi_* \Omega^p_{\widetilde{X}}(\log D) \cong \begin{cases} \Omega^p_X, & \textit{if } q = 0,\\ 0, & \textit{if } p < e \textit{ or } 0 < q < e - 1. \end{cases}$$

Proof. According to [25], $\widetilde{X}$ is obtained by a series of elementary steps

$$\widetilde{X} = \widetilde{X}_N \longrightarrow \widetilde{X}_{N-1} \longrightarrow \dots \longrightarrow \widetilde{X}_0 = X,$$

where $\widetilde{X}_k$ is the blow-up of $\widetilde{X}_{k-1}$ along a smooth subvariety W_{k-1}, contained in the strict transform of Y, therefore of codimension greater or equal than e. Moreover, if D_k is the union of exceptional divisors up to the step k, then the pair (D_k, W_k) has normal crossings. The result follows by applying proposition 3.26 to each blow-up. □

The following theorem implies in particular the weak purity condition for the complex $\mathcal{D}_{\mathrm{l,ll,a},\underline{X}}$.

Theorem 3.28. *Let $\underline{X} = (X, D)$ be as above. Let $Y \subseteq X$ be a Zariski closed subset of codimension greater or equal than p. Let c be the number of connected components of Y of codimension p. Then, the natural morphisms*

$$H^n_{\mathcal{D}_{\mathrm{l,ll,a}},Y}(X,p) \longrightarrow H^n_Y(X,\mathbb{R}(p))$$

are isomorphisms for all integers n. Therefore, we have

$$\begin{aligned} &H^n_{\mathcal{D}_{\mathrm{l,ll,a}},Y}(X,p) = 0, \qquad \text{for } n < 2p\,,\\ &H^{2p}_{\mathcal{D}_{\mathrm{l,ll,a}},Y}(X,p) \cong \mathbb{R}(p)^c. \end{aligned}$$

Proof. We fix a diagram

$$\begin{tikzcd} U \arrow[r, "\bar\iota"] \arrow[dr, "\iota"'] & \overline{U} \arrow[d, "\beta"] \\ & X \end{tikzcd}$$

such that $\bar\iota$ is an open immersion, β is a proper morphism, and $B = \overline{U} \setminus U$, $\overline{D} = \beta^{-1}(D)$, $B \cup \overline{D}$ are normal crossing divisors. Hence, $\overline{U}$ is an embedded resolution of singularities of Y. We assume moreover that $\overline{U}$ is obtained from X as $\widetilde{X}$ is obtained from X in corollary 3.27.

By theorem 2.42, the complexes $\mathcal{D}^*_{\mathrm{l,ll,a},\underline{X}}(X,p)$ and $\mathcal{D}^*_{\mathrm{l,ll,a},\underline{X}}(U,p)$ are quasi-isomorphic to the complexes $\mathcal{D}^*(E^*_{\underline{X}},p)$ and $\mathcal{D}^*(E^*_{\overline{U}}\langle B\rangle,p)$, respectively.

By the definition of the Deligne complex and theorem 2.6.2 in [8], there are quasi-isomorphisms

$$\begin{aligned} \mathcal{D}^*(E^*_{\underline{X}},p) &\longrightarrow s\left(E^*_{\underline{X},\mathbb{R}}(p) \to E^*_{\underline{X}} \,/ F^p E^*_{\underline{X}}\right),\\ \mathcal{D}^*(E^*_{\overline{U}}\langle B\rangle,p) &\longrightarrow s\left(E^*_{\overline{U}}\langle B\rangle_{\mathbb{R}}(p) \to E^*_{\overline{U}}\langle B\rangle \,/ F^p E^*_{\overline{U}}\langle B\rangle\right). \end{aligned}$$

By corollary 3.27 and theorem 2.5, the natural morphism

$$E^*_{\underline{X}} \,/ F^p E^*_{\underline{X}} \longrightarrow E^*_{\overline{U}}\langle B\rangle \,/ F^p E^*_{\overline{U}}\langle B\rangle$$

is a quasi-isomorphism. Hence, the morphism

$$s\left(\mathcal{D}^*(E^*_{\underline{X}},p) \to \mathcal{D}^*(E^*_{\overline{U}}\langle B\rangle,p)\right) \longrightarrow s\left(E^*_{\underline{X},\mathbb{R}}(p) \to E^*_{\overline{U}}\langle B\rangle_{\mathbb{R}}(p)\right)$$

is a quasi-isomorphism. Since the left hand complex computes $H^n_{\mathcal{D}_{\mathrm{l,ll,a}},Y}(X,p)$ and the right hand complex computes $H^n_Y(X,\mathbb{R}(p))$, we obtain the first statement of the theorem. The second statement follows form the purity of singular cohomology. □

Summing up the properties of the complex $\mathcal{D}_{\mathrm{l,ll,a},\underline{X}}$, we obtain

Theorem 3.29. *The complex $\mathcal{D}_{\mathrm{l,ll,a},\underline{X}}$ is a $\mathcal{D}_{\log}$-complex on X. Moreover, it is a pseudo-associative and commutative $\mathcal{D}_{\log}$-algebra and satisfies the weak purity condition (see [10], definition 3.1).* □

3.6 Arithmetic Chow rings with arbitrary singularities at infinity

Let A, X, D, and $\underline{X}$ be as at the beginning of section 3.4. Applying [10], section 4, we define the arithmetic Chow groups $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}})$. Then, theorems 3.18, 3.19, and 3.21 are also true for these groups. For theorem 3.20 to be true, we need X to be projective, but in this case there is no difference between $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}})$ and $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll}})$.
Since $\mathcal{D}_{\mathrm{l,ll,a},\underline{X}}$ satisfies the weak purity property, the analogue of theorem 3.17 reads as follows.

Theorem 3.30. *The following sequence is exact:*

$$\mathrm{CH}^{p-1,p}(X) \xrightarrow{\rho} \widetilde{\mathcal{D}}^{2p-1}_{\mathrm{l,ll,a}}(X,p) \xrightarrow{\mathrm{a}} \widehat{\mathrm{CH}}^p(X, \mathcal{D}_{\mathrm{l,ll,a}}) \xrightarrow{\zeta} \mathrm{CH}^p(X) \longrightarrow 0.$$

□

Another consequence of theorem 3.28 is the analogue of proposition 3.13, which is proved in the same way.

Proposition 3.31. *Let X be a smooth real variety and D a normal crossing divisor. Put $\underline{X} = (X, D)$. Let y be a p-codimensional cycle on X with support Y. Then, the class of the cycle (ω, g) in $H^{2p}_{\mathcal{D}_{\mathrm{l,ll,a}},Y}(X,p)$ is equal to the class of y, if and only if*

$$-2\partial\bar{\partial}[g]_X = [\omega] - \delta_y. \tag{3.32}$$

□

From this proposition and theorem 3.30, we obtain the analogue of theorem 6.23 in [10]:

Theorem 3.33. *Let $\widehat{\mathrm{CH}}^p(X)$ be the arithmetic Chow groups defined by Gillet and Soulé. If $D = \emptyset$, the assignment*

$$[y, (\omega_y, \widetilde{g}_y)] \mapsto [y, 2(2\pi i)^{d-p+1}[g_y]_X]$$

induces a well defined isomorphism

$$\Psi : \widehat{\mathrm{CH}}^p(X, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{CH}}^p(X),$$

which is compatible with products and pull-backs.

Remark 3.34. Note that, if $f : X \longrightarrow Y$ is a proper morphism between arithmetic varieties over A and such that $f_{\mathbb{R}} : X_{\mathbb{R}} \longrightarrow Y_{\mathbb{R}}$ is smooth, then there is a covariant f-pseudo morphism (see [10], definition 3.71) that induces a push-forward morphism

$$f_* : \widehat{\mathrm{CH}}^p(X, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{CH}}^p(Y, \mathcal{D}_{\mathrm{l,ll,a}}).$$

This push-forward is compatible with the push-forward defined by Gillet and Soulé.

Remark 3.35. We can define $\mathcal{D}_{\mathrm{l,ll,a,pre},\underline{X}}$ in the same way as $\mathcal{D}_{\mathrm{l,ll,a},\underline{X}}$ by replacing pre-log and pre-log-log forms for log and log-log forms. We then obtain a theory of arithmetic Chow groups $\widehat{\mathrm{CH}}^p(X, \mathcal{D}_{\mathrm{l,ll,a,pre}})$ with analogous properties. Note however that since we have not established the weak purity property of pre-log forms, we do not have the analogue of theorem 3.33.

4 Bott-Chern forms for log-singular hermitian vector bundles

The arithmetic intersection theory of Gillet and Soulé is complemented by an arithmetic K-theory and a theory of characteristic classes. A main ingredient of the theory of arithmetic characteristic classes are the Chern forms and Bott-Chern forms for hermitian vector bundles. In this section, after defining the class of singular metrics considered in this paper, we will generalize the theory of Chern forms and Bott-Chern forms to include this class of singular metrics.

4.1 Chern forms for hermitian metrics

Here we recall the Chern-Weil theory of characteristic classes for hermitian vector bundles. By a hermitian metric we will always mean a smooth hermitian metric.

Chern forms. Let $B \subseteq \mathbb{R}$ be a subring, let $\phi \in B[[T_1, \dots, T_n]]$ be any symmetric power series in n variables, and let $M_n(\mathbb{C})$ be the algebra of $n \times n$ complex matrices. For every $k \geq 0$, let $\phi^{(k)}$ be the homogeneous component of ϕ of degree k. We will denote also by $\phi^{(k)} : M_n(\mathbb{C}) \longrightarrow \mathbb{C}$ the unique polynomial map which is invariant under conjugation by $\mathrm{GL}_n(\mathbb{C})$ and whose value in the diagonal matrix $\mathrm{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_i \in \mathbb{C}$, is $\phi^{(k)}(\lambda_1, \dots, \lambda_n)$. More generally, if A is any B-algebra, $\phi^{(k)}$ defines a map $\phi^{(k)} : M_n(A) \longrightarrow A$, and if $I \subseteq A$ is a nilpotent subalgebra, we can define $\phi = \sum_k \phi^{(k)} : M_n(I) \longrightarrow A$.
Let $\overline{E} = (E, h)$ be a hermitian vector bundle of rank n on a complex manifold X. Let $\xi = \{\xi_1, \dots, \xi_n\}$ be a frame for E in an open subset $V \subseteq X$. We denote by $h(\xi) = (h_{ij}(\xi))$ the matrix of h in the frame ξ. Let $K(\xi)$ be the curvature matrix $K(\xi) = \bar{\partial}(\partial h(\xi) \cdot h(\xi)^{-1})$. The Chern form associated to ϕ and $\overline{E}$ is the form

$$\phi(\overline{E}) = \phi(-K(\xi)) \in E_V^*.$$

Basic properties. The following properties of the Chern forms are well known.

Theorem 4.1. *(i) By the invariance of the $\phi^{(k)}$, the Chern form $\phi(\overline{E})$ is independent of the choice of the frame ξ. Therefore, it globalizes to a differential form $\phi(\overline{E}) \in E^*_X$.*

(ii) The Chern forms are closed.

(iii) The component $\phi^{(k)}$ belongs to $\mathcal{D}^{2k}(E_X, k) = E^{k,k}_X \cap E^{2k}_{X,\mathbb{R}}(k)$.

(iv) If $X_{\mathbb{R}} = (X, F_\infty)$ is a real manifold, the vector bundle E is defined over $\mathbb{R}$, and the hermitian metric h is invariant under F_∞, then $\phi^{(k)}(E,h) \in \mathcal{D}^{2k}(E_X, k)^\sigma$, where σ is as in definition 3.10.

$\square$

Chern classes. Since the Chern forms are closed, they represent cohomology classes $\phi(E) = [\phi(E,h)] \in \bigoplus_k H^{2k}(\mathcal{D}(E_X,k))$. If X is projective, then $\bigoplus_k H^{2k}(\mathcal{D}(E_X,k)) = \bigoplus_k H^{2k}_{\mathcal{D}}(X, \mathbb{R}(k))$, hence we obtain classes in Deligne-Beilinson cohomology

$$\phi(E) \in \bigoplus_k H^{2k}_{\mathcal{D}}(X, \mathbb{R}(k)).$$

Note that, to simplify notations, the function ϕ will have different meanings according to its arguments. For instance, $\phi(E,h) = \phi(\overline{E})$ will mean the Chern form that depends on the bundle and the metric, whereas $\phi(E)$ will mean the Chern class that depends only on the bundle.

When X is quasi-projective, by means of smooth at infinity hermitian metrics, the Chern-Weil theory also allows us to construct Chern classes in Deligne-Beilinson cohomology.

Let E be an algebraic vector bundle on the quasi-projective complex manifold X. By [11], proposition 2.2, there exists a compactification $\widetilde{X}$ of X and a vector bundle $\widetilde{E}$ on $\widetilde{X}$ such that $\widetilde{E}|_X = E$. Let $\widetilde{h}$ be a smooth hermitian metric on $\widetilde{E}$ and let h be the induced metric on E. The hermitian metric h is said to be smooth at infinity.

With these notation, we write

$$\phi(E,h) = \phi(\widetilde{E}, \widetilde{h})|_X.$$

By [11], the class represented by $\phi(E,h)$ does not depend on the choice of $\widetilde{X}$, $\widetilde{E}$, nor $\widetilde{h}$.

Recall that there are Chern classes defined in the Chow ring $\phi(E)_{\mathrm{CH}} \in \mathrm{CH}^*(X)$; they are compatible with the Chern classes in cohomology. More precisely, we have

Proposition 4.2. *The composition*

$$\mathrm{CH}^k(X) \xrightarrow{\mathrm{cl}} H^{2k}_{\mathcal{D}}(X, \mathbb{R}(k)) \longrightarrow H^{2k}(\mathcal{D}(E_X, k))$$

sends $\phi^{(k)}(E)_{\mathrm{CH}}$ *to* $\phi^{(k)}(E)$.

Proof. If X is projective, then $H^{2k}_{\mathcal{D}}(X, \mathbb{R}(k)) = H^{k,k}(X, \mathbb{C}) \cap H^{2k}(X, \mathbb{R}(k))$. Therefore, the result follows from the compatibility of Chern classes on the Chow ring and on ordinary cohomology (see, e.g., [16], §19). If X is quasi-projective, the result follows from the projective case by functoriality. □

4.2 Bott-Chern forms for hermitian metrics

Here we recall the theory of Bott-Chern forms. For more details we refer to [36], [11], [9].

Bott-Chern forms. Let

$$\overline{\mathcal{E}} : 0 \longrightarrow (E', h') \longrightarrow (E, h) \longrightarrow (E'', h'') \longrightarrow 0$$

be a short exact sequence of hermitian vector bundles; by this we mean a short exact sequence of vector bundles, where each vector bundle is equipped with an arbitrarily chosen hermitian metric. Let ϕ be as in 4.1 and assume E has rank n.

The Chern classes behave additively with respect to exact sequences, i.e.,

$$\phi(E) = \phi(E' \oplus E'').$$

In general, this is not true for the Chern forms. This lack of additivity on the level of Chern forms is measured by the Bott-Chern forms.

The fundamental result of the theory of Bott-Chern forms is the following theorem (see [5], [2], [17]).

Theorem 4.3. *There is a unique way to attach to every sequence* $\overline{\mathcal{E}}$ *as above, a form* $\widetilde{\phi}(\overline{\mathcal{E}})$ *in*

$$\bigoplus_k \widetilde{\mathcal{D}}^{2k-1}(E_X, k) = \bigoplus_k \mathcal{D}^{2k-1}(E_X, k) / \operatorname{Im}(\mathrm{d}_{\mathcal{D}})$$

satisfying the following properties

(i) $\mathrm{d}_{\mathcal{D}}\, \widetilde{\phi}(\overline{\mathcal{E}}) = \phi(E' \oplus E'', h' \oplus h'') - \phi(E, h)$.

(ii) $f^* \widetilde{\phi}(\overline{\mathcal{E}}) = \widetilde{\phi}(f^* \overline{\mathcal{E}})$, *for every holomorphic map* $f : X \longrightarrow Y$.

(iii) If $(E, h) = (E', h') \overset{\perp}{\oplus} (E'', h'')$, *then* $\widetilde{\phi}(\overline{\mathcal{E}}) = 0$.

There are different methods to construct Bott-Chern forms. We will introduce a variant of the method used in [17] and that is the dual of the construction used in [11].

THE FIRST TRANSGRESSION BUNDLE. Let $\mathcal{O}(1)$ be the dual of the tautological line bundle of $\mathbb{P}^1$ with the standard metric. If $(x:y)$ are projective coordinates of $\mathbb{P}^1_{\mathbb{C}}$, then x and y are generating global sections of $\mathcal{O}(1)$ with norms

$$\|x\|^2 = \frac{x\bar{x}}{x\bar{x}+y\bar{y}} \quad \text{and} \quad \|y\|^2 = \frac{y\bar{y}}{x\bar{x}+y\bar{y}}.$$

Let

$$\overline{\mathcal{E}} : 0 \longrightarrow (E', h') \longrightarrow (E, h) \longrightarrow (E'', h'') \longrightarrow 0$$

be a short exact sequence of hermitian vector bundles such that h' is induced by h.

Let p_1, p_2, be the first and the second projection of $X \times \mathbb{P}^1_{\mathbb{C}}$, respectively. We write $E(n) = p_1^*E \otimes p_2^*\mathcal{O}(n)$. On this vector bundle we consider the metric induced by h and the standard metric of $\mathcal{O}(n)$, and we denote by $\overline{E}(n)$ this hermitian vector bundle. Analogously, we write $E''(n) = p_1^*E'' \otimes p_2^*\mathcal{O}(n)$ and denote by $\overline{E}''(n)$ the corresponding hermitian vector bundle.

DEFINITION 4.4. The *first transgression bundle* $\operatorname{tr}_1(\overline{\mathcal{E}})$ is the kernel of the morphism

$$\begin{array}{ccc} \overline{E}(1) \oplus \overline{E}''(1) & \longrightarrow & E''(2) \\ (s,t) & \longmapsto & s \otimes x - t \otimes y \end{array}$$

with the induced metric.

Note that the definition of $\operatorname{tr}_1(\overline{\mathcal{E}})$ includes the metric; therefore, the expression $\phi(\operatorname{tr}_1(\overline{\mathcal{E}}))$ means the Chern form of the hermitian vector bundle $\operatorname{tr}_1(\overline{\mathcal{E}})$ and not its Chern class.

The key property of the first transgression bundle is the following. We denote by i_0 and i_∞ the morphisms $X \longrightarrow X \times \mathbb{P}^1$ given by

$$i_0(p) = (p, (0:1)),$$
$$i_\infty(p) = (p, (1:0)).$$

Then, $i_0^*(\operatorname{tr}_1(\overline{\mathcal{E}}))$ is isometric to (E, h) and $i_\infty^*(\operatorname{tr}_1(\overline{\mathcal{E}}))$ is isometric to $(E', h') \overset{\perp}{\oplus} (E'', h'')$.

THE CONSTRUCTION OF BOTT-CHERN FORMS. Let $t = x/y$ be the absolute coordinate of $\mathbb{P}^1$. Let us consider the current $W_1 = \left[-\frac{1}{2}\log(t\bar{t})\right]$ on $\mathbb{P}^1$ given by

$$W_1(\eta) = \left[-\frac{1}{2}\log(t\bar{t})\right](\eta) = -\frac{1}{2\pi i}\int_{\mathbb{P}^1} \frac{\eta}{2}\log(t\bar{t}).$$

By the Poincaré Lelong equation

$$-2\partial\bar{\partial}\left[-\frac{1}{2}\log(t\bar{t})\right] = \delta_{(1:0)} - \delta_{(0:1)}. \tag{4.5}$$

Definition 4.6. Let X be a complex manifold, $\overline{\mathcal{E}}$ an exact sequence of hermitian vector bundles

$$\overline{\mathcal{E}} : 0 \longrightarrow (E', h') \longrightarrow (E, h) \longrightarrow (E'', h'') \longrightarrow 0,$$

such that the metric h' is induced by the metric h. The *Bott-Chern form* associated to the exact sequence $\overline{\mathcal{E}}$ is the differential form over X given by

$$\phi(\overline{\mathcal{E}}) = W_1(\phi(\mathrm{tr}_1(\overline{\mathcal{E}}))) = -\frac{1}{2\pi i}\int_{\mathbb{P}^1} \phi(\mathrm{tr}_1(\overline{\mathcal{E}}))\frac{1}{2}\log(t\bar{t}).$$

Note that we use also the letter ϕ to denote the Bott-Chern form associated to a power series ϕ because the meaning of $\phi(\overline{\mathcal{E}})$ is determined again by the argument $\overline{\mathcal{E}}$, which, in this case, is an exact sequence of hermitian vector bundles.

Definition 4.7. If $\overline{\mathcal{E}}$ is an exact sequence as above, but such that h' is not the metric induced by h, then we consider the exact sequences

$$\lambda^1\overline{\mathcal{E}} : 0 \longrightarrow (E', \widetilde{h'}) \longrightarrow (E, h) \longrightarrow (E'', h'') \longrightarrow 0,$$

where $\widetilde{h'}$ is the hermitian metric induced by h, and

$$\lambda^2\overline{\mathcal{E}} : 0 \longrightarrow 0 \longrightarrow (E' \oplus E'', \widetilde{h'} \oplus h'') \longrightarrow (E' \oplus E'', h' \oplus h'') \longrightarrow 0.$$

The *Bott-Chern form* associated to the exact sequence $\overline{\mathcal{E}}$ is

$$\phi(\overline{\mathcal{E}}) = \phi(\lambda^1\overline{\mathcal{E}}) + \phi(\lambda^2\overline{\mathcal{E}}).$$

Proposition 4.8. *If $\overline{\mathcal{E}}$ is an exact sequence as above with h' induced by h, then the Bott-Chern forms obtained from definition 4.6 and definition 4.7 agree.*

Proof. In this case we have $\lambda^1\overline{\mathcal{E}} = \overline{\mathcal{E}}$. Thus, we have to show that $\phi(\lambda^2\overline{\mathcal{E}}) = 0$. But $\mathrm{tr}_1(\lambda^2(\overline{\mathcal{E}}))$ is the bundle $p_1^*(E' \oplus E'')$ with the hermitian metric $h' \oplus h''$, which does not depend on the coordinate of $\mathbb{P}^1$. Therefore, we have

$$\phi(\lambda^2\overline{\mathcal{E}}) = -\frac{1}{2\pi i}\int_{\mathbb{P}^1} \phi(E' \oplus E'', h' \oplus h'')\frac{1}{2}\log(t\bar{t}) = 0.$$

□

It is easy to see that the forms $\phi(\overline{\mathcal{E}})$ belong to $\bigoplus_k \mathcal{D}^{2k-1}(E_X, k)$. We will denote by $\widetilde{\phi}(\overline{\mathcal{E}})$ the class of $\phi(\overline{\mathcal{E}})$ in the group

$$\bigoplus_k \widetilde{\mathcal{D}}^{2k-1}(E_X, k) = \bigoplus_k \mathcal{D}^{2k-1}(E_X, k)/\operatorname{Im}(\mathrm{d}_{\mathcal{D}}).$$

Proposition 4.9. *The classes $\widetilde{\phi}(\overline{\mathcal{E}})$ satisfy the properties of theorem 4.3.*

Proof. The first property follows from the Poincaré lemma (see, e.g., [36]). The second property is clear, because all the ingredients of the construction are functorial. We prove the third property. If $\overline{\mathcal{E}}$ is a split exact sequence with $(E,h) = (E',h') \overset{\perp}{\oplus} (E'',h'')$ and the obvious morphisms, then

$$\operatorname{tr}_1(\overline{\mathcal{E}}) = \overline{E}'(1) \oplus \overline{E}''(0)$$

with the induced metrics. Let ω be the first Chern form of the line bundle $\overline{\mathcal{O}}_{\mathbb{P}^1}(1)$. Then, we find

$$\phi(\overline{E}'(1) \oplus \overline{E}''(0)) = p_1^*(a) + p_1^*(b) \wedge p_2^*(\omega),$$

where a and b are suitable forms on X. Now we get

$$\begin{aligned}\phi(\overline{\mathcal{E}}) &= -\frac{1}{2\pi i}\int_{\mathbb{P}^1}(p_1^*(a) + p_1^*(b)\wedge p_2^*(\omega))\frac{1}{2}\log(t\bar{t})\\ &= -\frac{1}{2\pi i}\, a \wedge \int_{\mathbb{P}^1}\frac{1}{2}\log(t\bar{t}) - \frac{1}{2\pi i}\, b \wedge \int_{\mathbb{P}^1}\frac{\omega}{2}\log(t\bar{t}) = 0.\end{aligned}$$

□

Change of metrics. Of particular importance is the Bott-Chern form associated to a change of hermitian metrics. Let E be a holomorphic vector bundle of rank n with two hermitian metrics h and h'. We denote by $\operatorname{tr}_1(E,h,h')$ the first transgression bundle associated to the short exact sequence

$$0 \longrightarrow 0 \longrightarrow (E,h) \longrightarrow (E,h') \longrightarrow 0.$$

Explicitly, $\operatorname{tr}_1(E,h,h')$ is isomorphic to p_1^*E with the embedding

$$\begin{array}{ccc} p_1^*E & \longrightarrow & \overline{E}(1) \oplus \overline{E}'(1)\\ s & \longmapsto & (s\otimes y, s\otimes x);\end{array}$$

here $\overline{E} = (E,h)$ and $\overline{E}' = (E,h')$. Therefore, if ξ is a local frame for E on an open set U, it determines a local frame for $\operatorname{tr}(E,h,h')$, also denoted by ξ, on $U \times \mathbb{P}^1$. In this frame the metric is given by the matrix

$$\frac{y\bar{y}h(\xi) + x\bar{x}h'(\xi)}{x\bar{x} + y\bar{y}}. \tag{4.10}$$

Definition 4.11. Let X be a complex manifold, E be a complex vector bundle of rank n, h, h' two hermitian metrics on E, and ϕ as in section 4.1. The *Bott-Chern form* associated to the change of metric (E,h,h') is the Bott-Chern form associated to the short exact sequence

$$0 \longrightarrow 0 \longrightarrow (E,h) \longrightarrow (E,h') \longrightarrow 0.$$

We will denote this form by $\phi(E,h,h')$ or, if E is understood, by $\phi(h,h')$. This form satisfies

$$\mathrm{d}_{\mathcal{D}}\,\phi(E,h,h') = -2\partial\bar{\partial}\phi(E,h,h') = \phi(E,h') - \phi(E,h). \tag{4.12}$$

4.3 Iterated Bott-Chern forms for hermitian metrics

The theory of Bott-Chern forms can be iterated defining higher Bott-Chern forms for exact k-cubes of hermitian vector bundles. This theory provides explicit representatives of characteristic classes for higher K-theory (see [11], [9]).

Exact squares. Let $\langle -1,0,1\rangle$ be the category associated to the ordered set $\{-1,0,1\}$.

Definition 4.13. A *square of vector bundles* over X is a functor from the category $\langle -1,0,1\rangle^2$ to the category of vector bundles over X. Given a square of vector bundles $\mathcal{F}$ and numbers $i\in\{1,2\}$, $j\in\{-1,0,1\}$, then the (i,j)*-face* of $\mathcal{F}$, denoted by $\partial_i^j\mathcal{F}$, is the sequence

$$\begin{aligned}\partial_1^j\mathcal{F}&:\mathcal{F}_{j,-1}\longrightarrow\mathcal{F}_{j,0}\longrightarrow\mathcal{F}_{j,1},\\ \partial_2^j\mathcal{F}&:\mathcal{F}_{-1,j}\longrightarrow\mathcal{F}_{0,j}\longrightarrow\mathcal{F}_{1,j}.\end{aligned}$$

A square of vector bundles is called *exact*, if all the faces are short exact sequences. A *hermitian exact square* $\overline{\mathcal{F}}$ is an exact square $\mathcal{F}$ such that the vector bundles $\mathcal{F}_{i,j}$ are equipped with arbitrarily chosen hermitian metrics. If $\overline{\mathcal{F}}$ is a hermitian exact square, then the faces of $\overline{\mathcal{F}}$ are equipped with the induced hermitian metrics. The reader is referred to [11] for the definition of exact n-cubes.

Let ϕ be as before and let $\overline{\mathcal{F}}$ be a hermitian exact square of vector bundles over X such that $\mathcal{F}_{0,0}$ has rank n. Then, the form

$$\phi(\partial_1^{-1}\overline{\mathcal{F}}\oplus\partial_1^{1}\overline{\mathcal{F}})-\phi(\partial_1^{0}\overline{\mathcal{F}})-\phi(\partial_2^{-1}\overline{\mathcal{F}}\oplus\partial_2^{1}\overline{\mathcal{F}})+\phi(\partial_2^{0}\overline{\mathcal{F}})$$

is closed in the complex $\bigoplus_p\mathcal{D}^*(E_X,p)$. The iterated Bott-Chern form is a differential form

$$\phi(\overline{\mathcal{F}})\in\bigoplus_p\mathcal{D}^{2p-2}(E_X,p)$$

satisfying

$$\mathrm{d}_{\mathcal{D}}\,\phi(\overline{\mathcal{F}})=\phi(\partial_1^{-1}\overline{\mathcal{F}}\oplus\partial_1^{1}\overline{\mathcal{F}})-\phi(\partial_1^{0}\overline{\mathcal{F}})-\phi(\partial_2^{-1}\overline{\mathcal{F}}\oplus\partial_2^{1}\overline{\mathcal{F}})+\phi(\partial_2^{0}\overline{\mathcal{F}}).$$

The second transgression bundle.

Definition 4.14. Let $\overline{\mathcal{F}}$ be a hermitian exact square such that for $j=-1,0,1$, the hermitian metrics of the vector bundles $\mathcal{F}_{j,-1}$ and $\mathcal{F}_{-1,j}$ are induced by the metrics of $\mathcal{F}_{j,0}$ and $\mathcal{F}_{0,j}$, respectively. The *second transgression bundle* associated to $\overline{\mathcal{F}}$ is the hermitian vector bundle on $X\times\mathbb{P}^1\times\mathbb{P}^1$ given by

$$\mathrm{tr}_2(\overline{\mathcal{F}})=\mathrm{tr}_1\left(\mathrm{tr}_1(\partial_2^{-1}\overline{\mathcal{F}})\longrightarrow\mathrm{tr}_1(\partial_2^{0}\overline{\mathcal{F}})\longrightarrow\mathrm{tr}_1(\partial_2^{1}\overline{\mathcal{F}})\right).$$

The second transgression bundle satisfies

$$\begin{aligned} \operatorname{tr}_2(\overline{\mathcal{F}})|_{X\times\mathbb{P}^1\times(0:1)} &= \operatorname{tr}_1(\partial_2^0\overline{\mathcal{F}}),\\ \operatorname{tr}_2(\overline{\mathcal{F}})|_{X\times\mathbb{P}^1\times(1:0)} &= \operatorname{tr}_1(\partial_2^{-1}\overline{\mathcal{F}}) \overset{\perp}{\oplus} \operatorname{tr}_1(\partial_2^1\overline{\mathcal{F}}),\\ \operatorname{tr}_2(\overline{\mathcal{F}})|_{X\times(0:1)\times\mathbb{P}^1} &= \operatorname{tr}_1(\partial_1^0\overline{\mathcal{F}}),\\ \operatorname{tr}_2(\overline{\mathcal{F}})|_{X\times(1:0)\times\mathbb{P}^1} &= \operatorname{tr}_1(\partial_1^{-1}\overline{\mathcal{F}}) \overset{\perp}{\oplus} \operatorname{tr}_1(\partial_1^1\overline{\mathcal{F}}). \end{aligned} \tag{4.15}$$

The second Wang current. On $\mathbb{P}^1\times\mathbb{P}^1$ we put homogeneous coordinates $((x_1:y_1),(x_2:y_2))$; let $t_1=x_1/y_1$ and $t_2=x_2/y_2$.

Definition 4.16. The *second Wang current* is the current on $\mathbb{P}^1\times\mathbb{P}^1$ given by

$$W_2=\frac{1}{4}\left[\log(t_1\bar t_1)\left(\frac{\mathrm{d}\,t_2}{t_2}-\frac{\mathrm{d}\,\bar t_2}{\bar t_2}\right)-\log(t_2\bar t_2)\left(\frac{\mathrm{d}\,t_1}{t_1}-\frac{\mathrm{d}\,\bar t_1}{\bar t_1}\right)\right].$$

Observe that $W_2\in\mathcal{D}^2(D^*_{(\mathbb{P}^1)^2},2)$, where $D^*_{(\mathbb{P}^1)^2}$ is the Dolbeault complex of currents on $\mathbb{P}^1\times\mathbb{P}^1$. Moreover, we can write

$$W_2=\left[\left(-\frac{1}{2}\log(t_1\bar t_1)\right)\bullet\left(-\frac{1}{2}\log(t_2\bar t_2)\right)\right], \tag{4.17}$$

where $\bullet$ is the product in the Deligne complex (see definition 3.4).
For $p=(x_0:y_0)\in\mathbb{P}^1$, $i=1,2$, let $\iota_{i,p}:\mathbb{P}^1\longrightarrow\mathbb{P}^1\times\mathbb{P}^1$ be the inclusion given by

$$\begin{aligned}\iota_{1,p}(x:y)&=(x_0:y_0)\times(x:y),\\ \iota_{2,p}(x:y)&=(x:y)\times(x_0:y_0).\end{aligned}$$

Proposition 4.18. *We have the equality*

$$\mathrm{d}_{\mathcal{D}}\,W_2=(\iota_{1,(1:0)})_*W_1-(\iota_{1,(0:1)})_*W_1-(\iota_{2,(1:0)})_*W_1+(\iota_{2,(0:1)})_*W_1.$$

Proof. This proposition follows easily from a residue computation. Formally, we can interpret it as the Leibniz rule for the Deligne complex and equations (4.5), (4.17). □

The iterated Bott-Chern form.

Definition 4.19. Let $\overline{\mathcal{F}}$ be a hermitian exact square satisfying the condition of definition 4.14. The *iterated Bott-Chern form* associated to $\overline{\mathcal{F}}$ is the differential form given by

$$\begin{aligned}\phi(\overline{\mathcal{F}})=W_2(\phi(\overline{\mathcal{F}}))=&\frac{1}{(4\pi i)^2}\int_{\mathbb{P}^1\times\mathbb{P}^1}\phi(\operatorname{tr}_2(\overline{\mathcal{F}}))\wedge\log(t_1\bar t_1)\left(\frac{\mathrm{d}\,t_2}{t_2}-\frac{\mathrm{d}\,\bar t_2}{\bar t_2}\right)-\\ &\frac{1}{(4\pi i)^2}\int_{\mathbb{P}^1\times\mathbb{P}^1}\phi(\operatorname{tr}_2(\overline{\mathcal{F}}))\wedge\log(t_2\bar t_2)\left(\frac{\mathrm{d}\,t_1}{t_1}-\frac{\mathrm{d}\,\bar t_1}{\bar t_1}\right).\end{aligned}$$

When $\overline{\mathcal{F}}$ does not satisfy the condition of definition 4.14 we proceed as follows. Let $\lambda_i^k \overline{\mathcal{F}}$, $i = 1, 2$, $k = 1, 2$, be the hermitian exact square determined by

$$\partial_i^j(\lambda_i^k \overline{\mathcal{F}}) = \lambda^k(\partial_i^j \overline{\mathcal{F}}) \qquad (j = -1, 0, 1);$$

here $\lambda^k(\cdot)$ is as in definition 4.7.

Definition 4.20. Let $\overline{\mathcal{F}}$ be a hermitian exact square. Then, the *iterated Bott-Chern form* associated to $\overline{\mathcal{F}}$ is the differential form given by

$$\phi(\overline{\mathcal{F}}) = \phi(\lambda_1^1 \lambda_2^1 \overline{\mathcal{F}}) + \phi(\lambda_1^1 \lambda_2^2 \overline{\mathcal{F}}) + \phi(\lambda_1^2 \lambda_2^1 \overline{\mathcal{F}}) + \phi(\lambda_1^2 \lambda_2^2 \overline{\mathcal{F}}).$$

As in the case of exact sequences, if $\overline{\mathcal{F}}$ satisfies the condition of definition 4.14, then the iterated Bott-Chern forms obtained by means of definition 4.19 and definition 4.20 agree.

Theorem 4.21. *The second iterated Bott-Chern form satisfies*

$$\mathrm{d}_{\mathcal{D}}\, \phi(\overline{\mathcal{F}}) = \phi(\partial_1^{-1} \overline{\mathcal{F}} \oplus \partial_1^1 \overline{\mathcal{F}}) - \phi(\partial_1^0 \overline{\mathcal{F}}) - \phi(\partial_2^{-1} \overline{\mathcal{F}} \oplus \partial_2^1 \overline{\mathcal{F}}) + \phi(\partial_2^0 \overline{\mathcal{F}}).$$

Proof. This follows from (4.15) and proposition 4.18. □

The case of three different metrics. Let X be a complex manifold, E a holomorphic vector bundle on X and h, h', h'' smooth hermitian metrics on E. We will denote by $\overline{\mathcal{F}}(E, h, h', h'')$ the hermitian exact square

$$\begin{tikzcd}
0 \arrow[r] \arrow[d] & 0 \arrow[r] \arrow[d] & 0 \arrow[d] \\
0 \arrow[r] \arrow[d] & (E, h) \arrow[r] \arrow[d] & (E, h'') \arrow[d] \\
0 \arrow[r] & (E, h') \arrow[r] & (E, h'')
\end{tikzcd}$$

where the faces ∂_1^j are the rows and the faces ∂_2^j are the columns. As a shorthand, we will denote the hermitian vector bundle $\mathrm{tr}_2(\overline{\mathcal{F}}(E, h, h', h''))$ by $\mathrm{tr}_2(E, h, h', h'')$, or simply by $\mathrm{tr}_2(h, h', h'')$, if E is understood.

Definition 4.22. The *iterated Bott-Chern form* associated to the metrics h, h', h'' is the differential form given by

$$\phi(E, h, h', h'') = \phi(\overline{\mathcal{F}}(E, h, h', h'')).$$

Proposition 4.23. *The iterated Bott-Chern form satisfies*

$$\mathrm{d}_{\mathcal{D}}\, \phi(E, h, h', h'') = \phi(E, h, h') + \phi(E, h', h'') + \phi(E, h'', h).$$

Proof. By theorem 4.21, we have

$$\mathrm{d}_{\mathcal{D}}\, \phi(E, h, h', h'') = \phi(E, h', h'') - \phi(E, h, h'') - \phi(E, h'', h'') + \phi(E, h, h').$$

A direct computation shows that $\phi(E, h'', h'') = 0$ and that $\phi(E, h, h'') = -\phi(E, h'', h)$, which implies the result. □

4.4 Chern forms for singular hermitian metrics

There are various successful concepts of singular metrics in Arithmetic and Diophantine Geometry, see [3], [14], [33], and [34]. For our purposes the most important are: Faltings's notion of a metric with logarithmic singularities along a divisors with normal crossings (see [14]) and Mumford's notion of a good metric (see [34]). Both concepts have in common nature that automorphic vector bundles (equipped with their natural metrics) have the required local behavior. And, in fact, the application to automorphic vector bundles was the driving motivation to establish these definitions. For our purposes we will need a more precise description of the kind of metrics that appear when studying automorphic vector bundles.

Faltings's logarithmic singular metric. Let X be a complex manifold and let D be a normal crossing divisor. Put $U = X \setminus D$, and let $j : U \longrightarrow X$ be the inclusion. Let L be a line bundle on X and L_0 the restriction to U. A smooth metric h on L_0 is said to have logarithmic singularities along D, if, for any coordinate open subset V adapted to D and every non vanishing local section s, there exists a number $N \in \mathbb{N}$ such that

$$\max\{h(s), h^{-1}(s)\} \prec \left| \min_{j=1,..,k} \{\log |r_j|\} \right|^N . \tag{4.24}$$

Observe that this definition does not give any information on the behavior of the Chern form associated to the metric.

Good metrics in the sense of Mumford. We recall the notion of a good metric in the sense of Mumford, see [34].

Definition 4.25. Let E be a rank n vector bundle on X and E_0 the restriction to U. A smooth metric h on E_0 is said to be *good* on X, if, for all $x \in D$, there exist a neighborhood V adapted D and a holomorphic frame $\xi = \{e_1, \dots, e_n\}$ such that, writing $h(\xi)_{ij} = h(e_i, e_j)$, we have:

(i) $|h(\xi)_{ij}|, \det(h)^{-1} \prec \left(\prod_{i=1}^{k} \log(r_i)\right)^N$ for some $N \in \mathbb{N}$.

(ii) The 1-forms $(\partial h(\xi) \cdot h(\xi)^{-1})_{ij}$ are good.

A vector bundle provided with a good hermitian metric will be called a *good hermitian vector bundle.*

Lemma 4.26. *If (E, h) is a good hermitian vector bundle, then the 1-forms $(\partial h(\xi) \cdot h(\xi)^{-1})_{ij}$ are pre-log-log forms.*

Proof. Since a differential form with Poincaré growth has log-log growth (see [10], §7.1), we have that $(\partial h(\xi) \cdot h(\xi)^{-1})_{ij}$ and $\mathrm{d}(\partial h(\xi) \cdot h(\xi)^{-1})_{ij}$ have log-log growth. Since the condition of having log-log growth is bihomogeneous and

$(\partial h(\xi)\cdot h(\xi)^{-1})_{ij}$ has pure bidegree $(1,0)$, we have that $\partial(\partial h(\xi)\cdot h(\xi)^{-1})_{ij}$ and $\bar{\partial}(\partial h(\xi)\cdot h(\xi)^{-1})_{ij}$ have log-log growth. Finally, since

$$\partial(\partial h(\xi)\cdot h(\xi)^{-1}) = \partial h(\xi)\cdot h(\xi)^{-1}\wedge \partial h(\xi)\cdot h(\xi)^{-1},$$

the form $\partial\bar{\partial}(\partial h(\xi)\cdot h(\xi)^{-1})_{i,j}$ also has log-log growth. □

A fundamental property of the concept of good metrics is the following result of Mumford, see [34].

Proposition 4.27. *Let X, D, and U be as before.*

(i) Let (E_0,h) be a vector bundle over U. Then, it has at most one extension to a vector bundle E to X such that h is good along D.

(ii) If (E,h) is a good hermitian vector bundle, then, for any power series ϕ, the Chern form $\phi(E,h)$ is good. Moreover, its associated current $\left[\phi^{(k)}(E,h)\right]_X$ represents the Chern class $\phi(E)$ of E. □

Good Metrics of infinite order. Note that with the concept of good metric we have control on the local behavior of the Chern forms and of the cohomology class represented by its associated currents. As we will see later, we can also control the local behavior of the Bott-Chern forms. In order to have control on the cohomology classes represented by the Chern forms we need a slightly stronger definition, that is the analogue of our definition 2.62 of Poincaré singular forms.

Definition 4.28. Let X, D, and U be as before. Let E be a rank n vector bundle on X and let E_0 be the restriction of E to U. A smooth metric on E_0 is said to be *good of infinite order (along D)*, if, for every $x\in D$, there exist a trivializing open coordinate neighborhood V adapted to D and a holomorphic frame $\xi=\{e_1,\dots,e_n\}$ such that, writing $h(\xi)_{ij}=h(e_i,e_j)$, we have:

(i) The functions $h(\xi)_{ij}$, $\det(h(\xi))^{-1}$ belong to $\Gamma(V,\mathscr{E}^0_X\langle D\rangle)$.

(ii) The 1-forms $(\partial h(\xi)\cdot h(\xi)^{-1})_{ij}$ are Poincaré singular.

A vector bundle equipped with a good hermitian metric of infinite order will be called a ∞-*good hermitian vector bundle.*

Log-singular hermitian metrics. Although the hermitian metrics we are interested in, the automorphic hermitian metrics, are ∞-good, we will consider a slightly bigger set of singular metrics, the log-singular metrics, for which we will be able to define arithmetic characteristic classes.

Definition 4.29. Let X, D, and U be as before. Let E be a rank n vector bundle on X and let E_0 be the restriction of E to U. A smooth metric on E_0 is said to be *log-singular (along D)*, if, for every $x\in D$, there exist a trivializing open coordinate neighborhood V adapted to D and a holomorphic frame $\xi=\{e_1,\dots,e_n\}$ such that, writing $h(\xi)_{ij}=h(e_i,e_j)$, we have

(i) The functions $h(\xi)_{ij}$, $\det(h(\xi))^{-1}$ belong to $\Gamma(V, \mathscr{E}^0_X\langle D\rangle)$.

(ii) The 1-forms $(\partial h(\xi) \cdot h(\xi)^{-1})_{ij}$ belong to $\Gamma(V, \mathscr{E}^{1,0}_X\langle\langle D\rangle\rangle)$.

A vector bundle equipped with a log-singular hermitian metric will be called a *log-singular hermitian vector bundle.*

Note that, if a smooth metric on E_0 is log-singular, then the conditions of definition 4.29 are satisfied in every holomorphic frame in every trivializing open coordinate neighborhood V adapted to D.

Remark 4.30. By the very definition of log-singular metrics, the Chern forms $\phi(E, h)$ belong to the group $\oplus_k \mathcal{D}^{2k}(E_X\langle\langle D\rangle\rangle, k)$, if (E, h) is a log-singular hermitian vector bundle. Moreover, as we will see in proposition 4.61, the form $\phi(E, h)$ represents the Chern class $\phi(E)$ in $H^*_{\mathcal{D}}(X, \mathbb{R}(*))$.

Basic properties of log-singular hermitian metrics. The following properties are easily verified.

Proposition 4.31. *Let X, D, and U be as before. Let E and F be vector bundles on X, and let E_0 and F_0 be their restrictions to U. Let h_E and h_F be smooth hermitian metrics on E_0 and F_0. Write $\overline{E} = (E, h_E)$ and $\overline{F} = (F, h_F)$.*

(i) The hermitian vector bundle $\overline{E} \overset{\perp}{\oplus} \overline{F}$ is log-singular along D, if and only if, $\overline{E}$ and $\overline{F}$ are log-singular along D.

(ii) If $\overline{E}$ and $\overline{F}$ are log-singular along D, then the tensor product $\overline{E} \otimes \overline{F}$, the exterior and symmetric powers $\Lambda^n\overline{E}$, $S^n\overline{E}$, the dual bundle $\overline{E}^\vee$, and the bundle of homomorphisms $\operatorname{Hom}(\overline{E}, \overline{F})$, with their induced metrics, are log-singular along D.

□

Remark 4.32. Note however that the condition of being log-singular is not stable under taking general quotients and subbundles. That is, if (E, h) is a hermitian vector bundle, log-singular along a normal crossing divisor D, and E' is a subbundle or a quotient bundle, then the induced metric on E' need not be log-singular along D. For instance, let $X = \mathbb{A}^2$ with coordinates (t, z). Let $E = \mathcal{O}_X \oplus \mathcal{O}_X$ be the trivial rank two vector bundle with hermitian metric given, in the frame $\{e_1, e_2\}$, by the matrix

$$H = \begin{pmatrix} (\log(1/|z|))^{-1} & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.33}$$

This hermitian metric is log-singular along the divisor $D = \{z = 0\}$. But the subbundle generated by the section $e_1 + te_2$ with the induced metric does not satisfy the second condition of definition 4.29. Namely, let $h(t,z)=\|e_1+te_2\|^2$. Then, we find

$$\begin{aligned} h(t,z) &= t\bar{t}+(\log(1/|z|))^{-1}, \\ \partial h/h &= \frac{\bar{t}\,\mathrm{d}\,t}{t\bar{t}+(\log(1/|z|))^{-1}}+\frac{\mathrm{d}\,z}{z(\log(1/|z|))^{2}(t\bar{t}+(\log(1/|z|))^{-1})}. \end{aligned}$$

But the function $\bar{t}/(t\bar{t}+(\log(1/|z|))^{-1})$ is not log-log along D, as can be seen by considering the set of points

$$t=\sqrt{(\log(1/|z|))^{-1}}.$$

In this concrete case, the induced metric is not far from being log-singular: If $\widetilde{X}$ is the blow-up of X along the point $(0,0)$ and $\widetilde{D}$ is the pre-image of D, then the metric h is log-singular along $\widetilde{D}$. See also proposition 4.59 for a related example.

Remark 4.34. The condition of being log-singular is also not stable under extensions. That is, let

$$0\longrightarrow(E',h')\longrightarrow(E,h)\longrightarrow(E'',h'')\longrightarrow 0$$

be a short exact sequence with h' and h'' the hermitian metrics induced by h. If h' and h'' are log-singular, then h need not be log-singular.

Functoriality of log-singular metrics. The following result is a direct consequence of the definition and the functoriality of log forms and log-log forms.

Proposition 4.35. *Let X, X' be complex manifolds and let D, D' be normal crossing divisors of X, X', respectively. If $f: X'\longrightarrow X$ is a holomorphic map such that $f^{-1}(D)\subseteq D'$ and (E,h) is a log-singular hermitian vector bundle on X, then $(f^{*}E,f^{*}h)$ is a log-singular hermitian vector bundle on X'.* $\square$

4.5 Bott-Chern forms for singular hermitian metrics

Bott-Chern forms for log-singular hermitian metrics. In order to define characteristic classes of log-singular hermitian metrics with values in the log-log arithmetic Chow groups, we have to show that the Bott-Chern forms associated to a change of metric between a smooth metric and a log-singular metric is a log-log form. By the proof of the next theorem, it is clear that, even if we restrict ourselves to ∞-good hermitian metrics, the Bott-Chern forms are not necessarily Poincaré singular, but log-log. Therefore, the log-log forms are an essential ingredient of the theory and not only a technical addition to have the Poincaré lemma.

Theorem 4.36. *Let X be a complex manifold and let D be a normal crossing divisor. Put $U = X \setminus D$. Let E be a vector bundle on X.*

(i) If h is a smooth hermitian metric on E and h' is a smooth hermitian metric on $E|_U$, which is log-singular along D, then the Bott-Chern form $\phi(E, h, h')$ belongs to the group $\bigoplus_k \mathcal{D}^{2k-1}(E_X\langle\langle < D\rangle\rangle, k)$.

(ii) If h and h' are smooth hermitian metrics on E and h'' is a smooth hermitian metric on $E|_U$, which is log-singular along D, the iterated Bott-Chern form $\phi(E, h, h', h'')$ belongs to the group $\bigoplus_k \mathcal{D}^{2k-2}(E_X\langle\langle D\rangle\rangle, k)$.

Proof. Let V be a trivializing coordinate subset adapted to D with coordinates $(z_1, \dots, z_d)$. Thus, D has equation $z_1 \cdots z_k = 0$; we put $r_i = |z_i|$. We may also assume that V is contained in a compact subset of X. Let $\xi = \{e_i\}$ be a local holomorphic frame for E. Let g be the hermitian metric of $\operatorname{tr}_1(E, h, h')$. Since the vector bundle $\operatorname{tr}_1(E, h, h')$ is isomorphic to p_1^*E, the holomorphic frame ξ induces a holomorphic frame (also denoted by ξ) of $\operatorname{tr}_1(E, h, h')$.
For the rest of the proof the frame ξ will be fixed. Therefore, we drop it from the notation and we write

$$H = h(\xi), \qquad H_{ij} = h(\xi)_{ij} = h(e_i, e_j).$$

We use the same notation for the metrics h' and g.
Let $(x : y)$ be homogeneous coordinates of $\mathbb{P}^1$. Write $t = x/y$. We decompose $\mathbb{P}^1$ into two closed sets

$$\mathbb{P}^1_+ = \{(x : y) \in \mathbb{P}^1 \mid |x| \geq |y|\} \text{ and } \mathbb{P}^1_- = \{(x : y) \in \mathbb{P}^1 \mid |x| \leq |y|\}.$$

Then, we write

$$\phi(h, h') = \phi_+(h, h') + \phi_-(h, h'),$$

with

$$\phi_\pm(h, h') = \frac{-1}{4\pi i} \int_{\mathbb{P}^1_\pm} \phi(\operatorname{tr}_1(h, h')) \log(t\bar{t}). \tag{4.37}$$

We first show that the form $\phi_-(h, h')$ is log-log along D. One technical difficulty that we have to solve at this point is that the differential form $\phi(\operatorname{tr}_1(h, h'))$ is, in general, not a log-log form along $D \times \mathbb{P}^1$, because the vector bundle $\operatorname{tr}_1(h, h')$ need not be log-singular along $D \times \mathbb{P}^1$. This is the reason why we have to introduce a new class of singular functions.

Definition 4.38. For any pair of subsets $I, J \subseteq \{1, \dots, d\}$ and integers $n, K \geq 0$, we say that a smooth complex function f on $(V \setminus D) \times \mathbb{P}^1_-$ has *singularities of type* (n, α, β) *of order* K if there is an integer $N \geq 0$ such that, for any pair of multi-indices $\alpha, \beta \in \mathbb{Z}^d_{\geq 0}$ and integers $a, b \geq 0$ with $|\alpha + \beta| + a + b \leq K$, it holds the estimate

$$\left|\frac{\partial^{|\alpha|}}{\partial z^{\alpha}}\frac{\partial^{|\beta|}}{\partial \bar{z}^{\beta}}\frac{\partial^{a}}{\partial t}\frac{\partial^{b}}{\partial \bar{t}}f(z_1,\dots,z_d,t)\right| \prec \left(\frac{1}{|t|+(\prod_{i=1}^{k}\log(1/r_i))^{-N}}\right)^{n+a+b} \cdot \\ \cdot \frac{\left|\prod_{i=1}^{k}\log(\log(1/r_i))\right|^{N}}{r^{(\gamma^I+\gamma^J+\alpha+\beta)^{\leq k}}(\log(1/r))^{(\gamma^I+\gamma^J)^{\leq k}}}.$$

We say that f has *singularities of type* (n,α,β) *of infinite order*, if it has singularities of type (n,α,β) of order K for all $K\geq 0$.

The singularities of the differential form $\phi(\operatorname{tr}_1(h,h'))$ are controlled by the following result.

Lemma 4.39. *Let*

$$\phi(\operatorname{tr}_1(h,h')) = \sum_{\substack{0\leq a,b\leq 1\\ I,J}} f_{I,J,a,b}\, \mathrm{d}\, z^I\wedge \mathrm{d}\,\bar{z}^J\wedge \mathrm{d}\, t^a\wedge \mathrm{d}\,\bar{t}^b$$

be the decomposition of $\phi(\operatorname{tr}_1(h,h'))$ *into monomials over* $V\times\mathbb{P}^1_-$. *Then, the function* $f_{I,J,a,b}$ *has singularities of type* $(a+b,I,J)$ *of infinite order.*

Proof. On $V\times\mathbb{P}^1_-$, the matrix of g in the holomorphic frame ξ is

$$G=\frac{1}{1+t\bar{t}}(H+t\bar{t}H').$$

We write $G_1=H+t\bar{t}H'$. The differential form $\phi(\operatorname{tr}_1(h,h'))$ is a polynomial in the entries of the matrix $\bar{\partial}(\partial G G^{-1})$. Since

$$\partial G G^{-1}=\frac{-\bar{t}\,\mathrm{d}\,t}{1+t\bar{t}}\operatorname{id}+\partial G_1 G_1^{-1},$$

and the first summand of the right term is smooth, we are led to study the singularities of the matrices $\partial G_1G_1^{-1}$ and $\bar{\partial}(\partial G_1 G_1^{-1})$. This will be done in the subsequent lemmas.
We write $G_2=(H'^{-1}+t\bar{t}H^{-1})$. The following lemma is easy.

Lemma 4.40. *The matrices* H, H', G_1 *and* G_2 *satisfy the rules*

(1) $HG_1^{-1}=G_2^{-1}H'^{-1}$, (2) $G_1^{-1}H=H'^{-1}G_2^{-1}$,
(3) $H'G_1^{-1}=G_2^{-1}H^{-1}$, (4) $G_1^{-1}H'=H^{-1}G_2^{-1}$.

□

In order to bound the entries of $\partial G_1G_1^{-1}$ and the other matrices, we need the following estimates.

Lemma 4.41. *(i) The entries of the matrix G_1^{-1} are bounded. In particular, they have singularities of type $(0,\emptyset,\emptyset)$ of order 0.*

(ii) The entries of the matrix G_2^{-1} have singularities of type $(2,\emptyset,\emptyset)$ of order 0. Therefore, the entries of the matrices $t\,G_2^{-1}$ and $\bar{t}\,G_2^{-1}$ have singularities of type $(1,\emptyset,\emptyset)$ of order 0 and the entries of the matrix $t\bar{t}\,G_2^{-1}$ are bounded.

Proof. Let $H = T^{+}T$ be the Cholesky decomposition of H, where $(\bullet)^{+}$ denotes conjugate-transpose. Since H is smooth, the same is true for T. We can write

$$G_1^{-1} = T^{-1}(\mathrm{id} + t\bar{t}(T^{-1})^{+}H'T^{-1})^{-1}(T^{-1})^{+}.$$

But for any symmetric definite positive matrix A, the entries of $(\mathrm{id} + A)^{-1}$ have absolute value less that one. Therefore, the entries of the matrix G_1^{-1} are bounded. This proves the first statement.
To prove the second statement, we write

$$G_2^{-1} = T^{+}(TH'^{-1}T^{+} + t\bar{t}\ \mathrm{id})^{-1}T.$$

By the first condition of a log-singular metric, we can decompose

$$TH'^{-1}T^{+} = U^{+}DU$$

with U unitary and D diagonal with all the diagonal elements bounded from above by $(\prod_{i=1}^{k}\log(1/r_i))^{N}$ and bounded from below by $(\prod_{i=1}^{k}\log(1/r_i))^{-N}$ for some integer N. Then, we find

$$G_2^{-1} = (UT)^{+}(D + t\bar{t}\ \mathrm{id})^{-1}(UT).$$

Now the lemma follows from the fact that the norm of any entry of a unitary matrix is less or equal than one. $\square$

The remainder of the proof of lemma 4.39 is based on lemma 4.41.

Lemma 4.42. *Let $\sum \psi_{I,J,a,b}\,\mathrm{d}\,z^{I}\wedge\mathrm{d}\,\bar{z}^{J}\wedge\mathrm{d}\,t^{a}\wedge\mathrm{d}\,\bar{t}^{b}$ be the decomposition into monomials of an entry of any of the matrices $\partial G_1G_1^{-1}$, $\bar{\partial}(\partial G_1G_1^{-1})$, $\partial(\partial G_1G_1^{-1})$, and $\partial\bar{\partial}(\partial G_1G_1^{-1})$. Then, $\psi_{I,J,a,b}$ has singularities of type $(a+b,I,J)$ of order 0.*

Proof. We start with the entries of $\partial G_1G_1^{-1}$. Using lemma 4.40, we have

$$\begin{aligned}\partial G_1G_1^{-1} &= \partial HG_1^{-1} + \bar{t}\,\mathrm{d}\,tH'G_1^{-1} + t\bar{t}\partial H'G_1^{-1}\\ &= \partial HG_1^{-1} + (\bar{t}\,\mathrm{d}\,t + t\bar{t}\partial H'H'^{-1})G_2^{-1}H^{-1}.\end{aligned} \tag{4.43}$$

Therefore, the bound of the entries of $\partial G_1G_1^{-1}$ follows from lemma 4.41 and the fact that h' is log-singular.
The bound of the entries of $\partial(\partial G_1G_1^{-1})$ follows from the previous case and the formula

$$\partial(\partial G_1G_1^{-1}) = \partial G_1G_1^{-1}\wedge\partial G_1G_1^{-1}. \tag{4.44}$$

Before bounding $\bar{\partial}(\partial G_1 G_1^{-1})$, we compute

$$\bar{\partial} G_1^{-1} = -G_1^{-1} \bar{\partial} G_1 G_1^{-1} = -(\partial G_1 G_1^{-1})^{+} G_1^{-1}$$

and

$$\begin{aligned}
\bar{\partial} G_2^{-1} &= -G_2^{-1} \bar{\partial} G_2 G_2^{-1} \\
&= -G_2^{-1} (\bar{\partial} H'^{-1} + t \,\mathrm{d}\, \bar{t} H^{-1} + t\bar{t} \bar{\partial} H^{-1}) G_2^{-1} \\
&= G_2^{-1} H'^{-1} \bar{\partial} H' H'^{-1} G_2^{-1} - G_2^{-1} (t \,\mathrm{d}\, \bar{t} H^{-1} + t\bar{t} \bar{\partial} H^{-1}) G_2^{-1} \\
&= G_2^{-1} (\partial H' H'^{-1})^{+} G_1^{-1} H - G_2^{-1} (t \,\mathrm{d}\, \bar{t} H^{-1} + t\bar{t} \bar{\partial} H^{-1}) G_2^{-1}.
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
\bar{\partial}(\partial G_1 G_1^{-1}) =& \bar{\partial} \partial H G_1^{-1} + \partial H \wedge (\partial G_1 G_1^{-1})^{+} G_1^{-1} \\
&+ \bar{\partial}(\bar{t} \,\mathrm{d}\, t + t\bar{t} \partial H' H'^{-1}) G_2^{-1} H^{-1} \\
&- (\bar{t} \,\mathrm{d}\, t + t\bar{t} \partial H' H'^{-1}) G_2^{-1} \wedge (\partial H' H'^{-1})^{+} G_1^{-1} \\
&+ (\bar{t} \,\mathrm{d}\, t + t\bar{t} \partial H' H'^{-1}) G_2^{-1} \wedge (t \,\mathrm{d}\, \bar{t} H^{-1} + t\bar{t} \bar{\partial} H^{-1}) G_2^{-1} H^{-1} \\
&- (\bar{t} \,\mathrm{d}\, t + t\bar{t} \partial H' H'^{-1}) G_2^{-1} \wedge \bar{\partial} H^{-1}. \qquad (4.45)
\end{aligned}$$

Thus, the bound for the entries of $\bar{\partial}(\partial G_1 G_1^{-1})$ follows again by lemma 4.41 and the assumptions on H and H'.
Finally, the case of $\partial\bar{\partial}(\partial G_1 G_1^{-1})$ follows from the formula

$$\partial\bar{\partial}(\partial G_1 G_1^{-1}) = -\bar{\partial}(\partial G_1 G_1) \wedge \partial G_1 G_1^{-1} + \partial G_1 G_1^{-1} \wedge \bar{\partial}(\partial G_1 G_1). \qquad (4.46)$$

□

As a direct consequence of the previous lemma, we obtain that the functions $f_{I,J,a,b}$ of lemma 4.39 have singularities of type $(a+b, I, J)$ of order 0. But we have to show that they have singularities of type $(a+b, I, J)$ of infinite order. Thus, we have to bound all of their derivatives. As before, it is enough to bound the derivatives of the components of the entries of the matrix $\bar{\partial}(\partial G_1 G_1^{-1})$. By the formulas (4.43) and (4.45), it is enough to bound the derivatives of the entries of the matrices G_1^{-1} and G_2^{-1}. The idea to accomplish this task is to use induction, because the derivatives of these matrices can be written in terms of the same matrices and the derivatives of H and H', which we can control. The inductive step is provided by the next lemmas.

Lemma 4.47. *If the entries of the matrices G_1^{-1} and G_2^{-1} have singularities of type $(0,\emptyset,\emptyset)$ and $(2,\emptyset,\emptyset)$, respectively, of order K, then, for every $i = 1, \ldots, d$, the entries of $\frac{\partial}{\partial z_i} G_1 G_1^{-1}$ have singularities of type $(0, \{i\}, \emptyset)$ of order K and the entries of $\frac{\partial}{\partial t} G_1 G_1^{-1}$ have singularities of type $(1,\emptyset,\emptyset)$ of order K.*

Proof. The result is a consequence of the formulas

$$\frac{\partial}{\partial z_i} G_1 G_1^{-1} = \frac{\partial}{\partial z_i} H G_1^{-1} + t\bar{t}\left(\frac{\partial}{\partial z_i} H' H'^{-1}\right) G_2^{-1} H^{-1}, \tag{4.48}$$

$$\frac{\partial}{\partial t} G_1 G_1^{-1} = \bar{t}\,\mathrm{d}\,t G_2^{-1} H^{-1}, \tag{4.49}$$

which follow from equation (4.43). □

Lemma 4.50. *If the entries of the matrix G_1^{-1} have singularities of type $(0, \emptyset, \emptyset)$ of order K for all $i = 1, \ldots, d$, the entries of the matrix $\frac{\partial}{\partial z_i} G_1 G_1^{-1}$ have singularities of type $(0, \{i\}, \emptyset)$ of order K, and the entries of the matrix $\frac{\partial}{\partial t} G_1 G_1^{-1}$ have singularities of type $(1, \emptyset, \emptyset)$ of order K, then the entries of the matrix G_1^{-1} have singularities of type $(0, \emptyset, \emptyset)$ of order $K + 1$.*

Proof. The result follows from formulas

$$\frac{\partial}{\partial z_i} G_1^{-1} = -G_1^{-1}\left(\frac{\partial}{\partial z_i} G_1 G_1^{-1}\right), \qquad \frac{\partial}{\partial \bar{z}_i} G_1^{-1} = -\left(\frac{\partial}{\partial z_i} G_1 G_1^{-1}\right)^{+} G_1^{-1},$$

$$\frac{\partial}{\partial t} G_1^{-1} = -G_1^{-1}\left(\frac{\partial}{\partial t} G_1 G_1^{-1}\right), \qquad \frac{\partial}{\partial \bar{t}} G_1^{-1} = -\left(\frac{\partial}{\partial t} G_1 G_1^{-1}\right)^{+} G_1^{-1}.$$

□

Lemma 4.51. *If the entries of the matrices G_1^{-1} and G_2^{-1} have singularities of type $(0, \emptyset, \emptyset)$ and $(2, \emptyset, \emptyset)$, respectively, of order K, then the entries of the matrix G_2^{-1} have singularities of type $(2, \emptyset, \emptyset)$ of order $K + 1$.*

Proof. This result is consequence of the equations

$$\begin{aligned}
\frac{\partial}{\partial z_i} G_2^{-1} &= -t\bar{t} G_2^{-1} \frac{\partial}{\partial z_i} H^{-1} G_2^{-1} + H G_1^{-1}\left(\frac{\partial}{\partial z_i} H' H'^{-1}\right) G_2^{-1}, \\
\frac{\partial}{\partial \bar{z}_i} G_2^{-1} &= -t\bar{t} G_2^{-1} \frac{\partial}{\partial \bar{z}_i} H^{-1} G_2^{-1} + G_2^{-1}\left(\frac{\partial}{\partial z_i} H' H'^{-1}\right)^{+} G_1^{-1} H \\
\frac{\partial}{\partial t} G_2^{-1} &= -G_2^{-1}(\bar{t}\,\mathrm{d}\,t H^{-1}) G_2^{-1}, \\
\frac{\partial}{\partial \bar{t}} G_2^{-1} &= -G_2^{-1}(t\,\mathrm{d}\,\bar{t} H^{-1}) G_2^{-1}.
\end{aligned}$$

□

Summing up lemmas 4.41, 4.42, 4.47, 4.50, 4.51 and equations (4.43), (4.44), (4.45), (4.46), we obtain

Lemma 4.52. *Let $\sum \psi_{I,J,a,b}\,\mathrm{d}\,z^I \wedge \mathrm{d}\,\bar{z}^J \wedge \mathrm{d}\,t^a \wedge \mathrm{d}\,\bar{t}^b$ be the decomposition into monomials of an entry of any of the matrices $\partial G_1 G_1^{-1}$, $\bar{\partial}(\partial G_1 G_1^{-1})$, $\partial(\partial G_1 G_1^{-1})$, and $\partial\bar{\partial}(\partial G_1 G_1^{-1})$. Then, $\psi_{I,J,a,b}$ has singularities of type $(a + b, I, J)$ of infinite order.* □

End of proof of lemma 4.39. This finishes the proof of lemma 4.39. □

Once we have bounded the components of $\phi(\mathrm{tr}_1(h,h'))$ over $V \times \mathbb{P}^1_-$, in order to bound the components of $\phi_-(h,h')$, we have to estimate the integral (4.37).

Lemma 4.53. *Let $0 \leq a \leq 1$ be a real number. Then, we have*

$$\int_0^1 \frac{\log(1/r)}{r+a} \,\mathrm{d}\, r \leq 1 + \log(1/a) + \frac{1}{2}\log^2(1/a),$$
$$\int_0^1 \frac{r\log(1/r)}{(r+a)^2} \,\mathrm{d}\, r \leq 1 + \log(1/a) + \frac{1}{2}\log^2(1/a).$$

Proof. We have the following estimates

$$\begin{aligned}
\int_0^1 \frac{r\log(1/r)}{(r+a)^2} \,\mathrm{d}\, r &\leq \int_0^1 \frac{\log(1/r)}{r+a}\,\mathrm{d}\, r \\
&= \int_0^a \frac{\log(1/r)}{r+a}\,\mathrm{d}\, r + \int_a^1 \frac{\log(1/r)}{r+a}\,\mathrm{d}\, r \\
&\leq \int_0^a \frac{\log(1/r)}{a}\,\mathrm{d}\, r + \int_a^1 \frac{\log(1/r)}{r}\,\mathrm{d}\, r \\
&= \left.\frac{r\log(1/r)+r}{a}\right|_0^a - \left.\frac{1}{2}\log^2(1/r)\right|_a^1 \\
&= \log(1/a) + 1 + \frac{1}{2}\log^2(a).
\end{aligned}$$

□

We are now in position to bound the components of $\phi_-(h,h')$. Let

$$\phi_-(h,h') = \sum_{I,J} g_{I,J}\,\mathrm{d}\, z^I \wedge \mathrm{d}\,\bar{z}^J$$

be the decomposition of $\phi_-(h,h')$ into monomials. Then, using lemma 4.39 and lemma 4.53, we have

$$\begin{aligned}
|g_{I,J}| &= \left|\frac{1}{4\pi i}\int_{\mathbb{P}^1_-} f_{I,J,1,1}\log(t\bar{t})\,\mathrm{d}\, t \wedge \mathrm{d}\,\bar{t}\right| \\
&\prec \frac{\left|\prod_{i=1}^k \log(\log(1/r_i))\right|^N}{r^{(\gamma^I+\gamma^J)^{\leq k}}(\log(1/r))^{(\gamma^I+\gamma^J)^{\leq k}}} \cdot \\
&\qquad \cdot \int_{\mathbb{P}^1_-}\left(\frac{1}{|t| + (\prod_{i=1}^k \log(1/r_i))^{-N}}\right)^2 \log(t\bar{t})\,\mathrm{d}\, t \wedge \mathrm{d}\,\bar{t} \\
&\prec \frac{\left|\prod_{i=1}^k \log(\log(1/r_i))\right|^{N'}}{r^{(\gamma^I+\gamma^J)^{\leq k}}(\log(1/r))^{(\gamma^I+\gamma^J)^{\leq k}}}.
\end{aligned}$$

The derivatives of $g_{I,J}$ are bounded in the same way using the theorem of taking derivatives under the integral sign. The components of $\partial\phi_{-}(h,h')$ and $\bar{\partial}\phi_{-}(h,h')$ and their derivatives are bounded in a similar way using that

$$\partial\phi_{-}(h,h') = \frac{-1}{4\pi i}\int_{\mathbb{P}^1_{-}} \phi(\mathrm{tr}_1(h,h')) \wedge \frac{\mathrm{d}\,t}{t},$$

and

$$\bar{\partial}\phi_{-}(h,h') = \frac{-1}{4\pi i}\int_{\mathbb{P}^1_{-}} \phi(\mathrm{tr}_1(h,h'))\frac{\mathrm{d}\,\bar{t}}{\bar{t}}.$$

To bound the components of $\phi_{+}(h,h')$, $\partial\phi_{+}(h,h')$ and $\bar{\partial}\phi_{+}(h,h')$ and their derivatives, we will use the same technique. Let $s = 1/t$ be a local coordinate in $\mathbb{P}^1_{+}$. In these coordinates, we have

$$G = \frac{1}{1+s\bar{s}}(H' + s\bar{s}H).$$

We write

$$G_3 = (H' + s\bar{s}H), \qquad G_4 = (H^{-1} + s\bar{s}H'^{-1}).$$

In this case, using the adequate variant of definition 4.38, the analogue of lemma 4.41 is

Lemma 4.54. *(i) The entries of the matrix G_3^{-1} have singularities of type $(2,\emptyset,\emptyset)$ of order* 0. *Therefore, the entries of the matrices $s\,G_3^{-1}$ and $\bar{s}\,G_3^{-1}$ have singularities of type $(1,\emptyset,\emptyset)$ of order* 0 *and the entries of the matrix $t\bar{s}\,G_3^{-1}$ are bounded.*

(ii) The entries of the matrix G_4^{-1} are bounded. In particular, they have singularities of type $(0,\emptyset,\emptyset)$ of order 0.

□

Note that the bounds for G_3^{-1} and G_4^{-1} are not the same as the bounds for G_1^{-1} and G_2^{-1}, but they are switched. To bound the entries of $\partial G_3 G_3^{-1}$, we use

$$\partial G_3 G_3^{-1} = \partial H' H'^{-1} G_4^{-1} H^{-1} + \bar{s}\,\mathrm{d}\,sHG_3^{-1} + s\bar{s}\partial H G_3^{-1}.$$

We leave the remaining details to the reader.
Finally, to bound $\partial\bar{\partial}\phi(h,h')$, we use equation (4.12). This completes the proof of the first statement.
We now prove the second statement. By definition, we have

$$\mathrm{tr}_2(h,h',h'') = \mathrm{tr}_1(0 \longrightarrow \mathrm{tr}_1(h,h') \longrightarrow \mathrm{tr}_1(h'',h'')).$$

But $\mathrm{tr}_1(h,h')$ is a smooth hermitian vector bundle on $X \times \mathbb{P}^1$ and $\mathrm{tr}_1(h'',h'')$ is isometric to $p_1^*(E,h'')$ and, in consequence, log-singular along $D \times \mathbb{P}^1$. Therefore, we can apply lemma 4.39 to $\phi(\mathrm{tr}_2(h,h',h''))$; by lemma 4.53, the form $\phi(h,h',h'')$ has log-log growth of infinite order. To conclude that it is a log-log form we still have to control $\partial\phi(h,h',h'')$, $\bar{\partial}\phi(h,h',h'')$, and $\partial\bar{\partial}\phi(h,h',h'')$. A residue computation shows

$$\begin{aligned}\partial\phi(h,h',h'') =& \frac{1}{2}(\phi(h,h')+\phi(h',h'')+\phi(h'',h))\\ &+\frac{2}{(4\pi i)^2}\int_{\mathbb{P}^1\times\mathbb{P}^1}\phi(\mathrm{tr}_2(h,h',h''))\wedge\frac{\mathrm{d}\,t_1}{t_1}\wedge\frac{\mathrm{d}\,t_2}{t_2}\\ &-\frac{1}{(4\pi i)^2}\int_{\mathbb{P}^1\times\mathbb{P}^1}\phi(\mathrm{tr}_2(h,h',h''))\wedge\left(\frac{\mathrm{d}\,\bar{t}_1}{\bar{t}_1}\wedge\frac{\mathrm{d}\,t_2}{t_2}+\frac{\mathrm{d}\,t_1}{t_1}\wedge\frac{\mathrm{d}\,\bar{t}_2}{\bar{t}_2}\right),\end{aligned}$$

and

$$\begin{aligned}\bar{\partial}\phi(h,h',h'') =& \frac{1}{2}(\phi(h,h')+\phi(h',h'')+\phi(h'',h))\\ &-\frac{2}{(4\pi i)^2}\int_{\mathbb{P}^1\times\mathbb{P}^1}\phi(\mathrm{tr}_2(h,h',h''))\wedge\frac{\mathrm{d}\,\bar{t}_1}{\bar{t}_1}\wedge\frac{\mathrm{d}\,\bar{t}_2}{\bar{t}_2}\\ &+\frac{1}{(4\pi i)^2}\int_{\mathbb{P}^1\times\mathbb{P}^1}\phi(\mathrm{tr}_2(h,h',h''))\wedge\left(\frac{\mathrm{d}\,\bar{t}_1}{\bar{t}_1}\wedge\frac{\mathrm{d}\,t_2}{t_2}+\frac{\mathrm{d}\,t_1}{t_1}\wedge\frac{\mathrm{d}\,\bar{t}_2}{\bar{t}_2}\right).\end{aligned}$$

Hence, again by lemma 4.53, the forms $\partial\phi(h,h',h'')$ and $\bar{\partial}\phi(h,h',h'')$ have log-log growth of infinite order. Finally, since

$$\partial\bar{\partial}\phi(h,h',h'') = (\partial-\bar{\partial})(\phi(h,h')+\phi(h',h'')+\phi(h'',h))$$

by the first statement, the form $\partial\bar{\partial}\phi(h,h',h'')$ also has log-log growth of infinite order; therefore, $\phi(h,h',h'')$ is a log-log form.

End of proof of theorem 4.36. This finishes the proof of theorem 4.36. □

Bott-Chern forms for good hermitian metrics. All the theory we have developed so far is also valid for good hermitian vector bundles with the obvious changes. For instance, if the hermitian vector bundle is good instead of log-singular, we obtain that the Bott-Chern forms are pre-log-log instead of log-log.

Theorem 4.55. *Let X be a complex manifold and let D be a normal crossing divisor. Put $U = X\setminus D$. Let E be a vector bundle on X. If h and h' are smooth hermitian metrics on E and h'' is a smooth hermitian metric on $E|_U$, which is good along D, then the Bott-Chern form $\phi(E,h,h'')$ and the iterated Bott-Chern form $\phi(E,h,h',h'')$ are pre-log-log forms.*

Proof. Observe that lemma 4.40, lemma 4.41, and lemma 4.42 are true in the case of good hermitian metrics by lemma 4.26, and these results are enough to prove that $\phi(E,h,h')$ and $\phi(E,h,h',h'')$ are pre-log-log forms by the same arguments as before. □

The singularities of the first transgression bundle. With the notation of theorem 4.36, observe that the hermitian vector bundle $\operatorname{tr}_1(E,h,h')$ need not be log-singular along the divisor $D\times\mathbb{P}^1$ (see remark 4.32). Nevertheless, as we will see in the following results, it is close to be log-singular. For instance, it is log-singular along $D\times\mathbb{P}^1\cup X\times\{(0:1),(1:0)\}$, or it can be made a log-singular hermitian vector bundle after some blow-ups. This second statement will be useful in the axiomatic characterization of Bott-Chern classes.

Lemma 4.56. *Let a,b be real numbers with $a>0$ and $b>e^{1/e}$. Then, we have*

$$\frac{\log(a)}{\log(b)}<1+\frac{a}{b}.$$

Proof. If $b\geq a$, then the statement is obvious. If $a>b$, we write $a=cb$ with $c>1$. Then, the inequality of the lemma is equivalent to

$$\frac{\log(c)}{c}<\log(b).$$

But the function $\log(c)/c$ is a bounded function that has a maximum at $c=e$ with value $1/e$. Therefore, the result is a consequence of the condition on b. □

Corollary 4.57. *With the notation of theorem 4.36, the first transgression hermitian vector bundle $\operatorname{tr}_1(E,h,h')$ is log-singular along the divisor*

$$D\times\mathbb{P}^1\cup X\times\{(0:1),(1:0)\}.$$

Proof. The first condition of definition 4.29 is easy to prove. We will prove the second condition. Lemma 4.56 implies that, for $a,b\gg 0$, the inequality

$$\frac{1}{1/a+1/b}<\frac{\log(b)}{(1/a)\log(a)} \tag{4.58}$$

holds. Applying this equation to $a=1/|t|$ and $b=(\prod_{i=1}^k\log(1/r_i))^N$, we obtain

$$\frac{1}{|t|+(\prod_{i=1}^k\log(1/r_i))^{-N}}<\frac{\log((\prod_{i=1}^k\log(1/r_i))^N)}{|t|\log(1/|t|)}\prec\frac{\sum_{i=1}^k\log(\log(1/r_i))}{|t|\log(1/|t|)}.$$

Therefore, lemma 4.52 implies that on $V\times\mathbb{P}^1_-$ the entries of ∂GG^{-1} are log-log along $D\times\mathbb{P}^1\cup X\times\{(0:1),(1:0)\}$. The proof for the bound on $V\times\mathbb{P}^1_+$ is analogous. □

Proposition 4.59. *With the same hypothesis of theorem 4.36, let $D=D_1\cup\ldots\cup D_n$ be the decomposition of D in smooth irreducible components. Let $\widetilde{Z}$ be the variety obtained from $X\times\mathbb{P}^1$ by blowing-up $D_1\times(1:0)$ and then, successively, the strict transforms of $D_2\times(1:0),\ldots,D_n\times(1:0),D_1\times(0:1),\ldots,D_n\times(0:1)$. Let $\pi:\widetilde{Z}\longrightarrow X\times\mathbb{P}^1$ be the morphism induced by the blow-ups and let $C\subseteq\widetilde{Z}$ be the pre-image by π of $D\times\mathbb{P}^1$. Then, we have* *(i)* *C is a normal crossing divisor.*

(ii) *The closed immersions $i_0, i_\infty : X \longrightarrow X \times \mathbb{P}^1$, given by*

$$i_0(p) = (p, (0:1)), \qquad i_\infty(p) = (p, (1:0)),$$

can be lifted to closed immersions $j_0, j_\infty : X \longrightarrow \widetilde{Z}$.

(iii) *The hermitian vector bundle $\pi^* \operatorname{tr}_1(E, h, h')$ is log-singular along the divisor C.*

Proof. The first statement is obvious and the second is a direct consequence of the universal property of the blow-up and the fact that the intersection of the center of every blow-up with the transform of $X \times (1:0)$ or $X \times (0:1)$ is either empty or a divisor.

To prove the third statement, we will use the same notations as in the proof of theorem 4.36. Let U be the subset of $V \times \mathbb{P}^1_-$, where $|t| < 1/e^e$. For simplicity, we assume that the components of D meeting V are $D_1, \dots, D_k$ and that the component D_i has equation $z_i = 0$. Then, U, with coordinates $(z_1, \dots, z_d, t)$, is a coordinate neighborhood adapted to $D \times \mathbb{P}^1$. The open subset $\pi^{-1}(U)$ can be covered by $k+1$ coordinate neighborhoods, denoted by $\widetilde{U}_1, \dots, \widetilde{U}_{k+1}$. The coordinates of these subsets, the expression of π and the equation of C in these coordinates are given in the following table:

Subset	Coordinates	π	
U_1	$(u, x_1, \dots, x_n)$	$t = u$ $z_1 = ux_1$ $z_i = x_i\ (i \neq 1)$	$ux_1 \cdots x_k = 0$
U_j $(1 < j < k+1)$	$(u, x_1, \dots, x_n)$	$t = ux_1 \cdots x_{j-1}$ $z_j = ux_j$ $z_i = x_i\ (i \neq j)$	$ux_1 \cdots x_k = 0$
U_{k+1}	$(u, x_1, \dots, x_n)$	$t = ux_1 \cdots x_k$ $z_i = x_i\ (i = 1, \dots, d)$	$x_1 \cdots x_k = 0$

Since, for $j = 1, \dots, k$, we have

$$\pi^{-1}(D \times \mathbb{P}^1 \cup X \times \{(0:1), (1:0)\}) \cap U_j = C \cap U_j,$$

we know by corollary 4.57 and the functoriality of log-singular metrics that the hermitian vector bundle $\pi^* \operatorname{tr}_1(E, h, h')|_{U_j}$ is log-singular. Hence, we only have to prove that $\pi^* \operatorname{tr}_1(E, h, h')|_{U_{k+1}}$ is log-singular. The first condition of definition 4.29 follows easily from the definition of the metric g. To prove the second condition of definition 4.29, we can proceed in two ways. The first method is to derive this result directly from lemma 4.52 applying the chain rule. But, since we have to bound all derivatives, this is a notational nightmare. The second method is to bound the derivatives inductively mimicking the proof of lemma 4.53. To this end, instead of lemma 4.41, we use the following substitute.

Lemma 4.60. *(i) The entries of the matrix $\pi^* G_1^{-1}|_{U_{k+1}}$ are bounded in every compact subset of U_{k+1}. In particular, they are $(\emptyset,\emptyset)$-log-log growth functions of order* 0 *(see definition 2.21).*

(ii) If ψ is an entry of the matrix G_2^{-1}, then we have

$$|(\pi^*\psi|_{U_{k+1}})(x_1,\dots,x_d,u)| \prec \left|\prod_{i=1}^{k}\log(1/|x_i|)\right|^N$$

for some integer N. Therefore, $\pi^*(t\psi)$ and $\pi^*(\bar{t}\psi)$ are bounded in any compact subset of U_{k+1} and, for $i=1,\dots,k$, the function

$$\prod_{j\neq i}|x_j|\pi^*\psi$$

is a $(\{i\},\emptyset)$-log-log growth function of order 0.

□

We leave it to the reader to make explicit the analogues of lemmas 4.47, 4.50, and 4.51 in this case.

The proof that it is also log-singular in the pre-image of an open subset of $\mathbb{P}^1_+$ is analogous. □

Chern forms for log-singular hermitian bundles.

Proposition 4.61. *Let X be a complex projective manifold, D a normal crossing divisor of X, (E,h) a hermitian vector bundle log-singular along D. Let ϕ be any symmetric power series. Then, the Chern form $\phi(E,h)$ represents the Chern class $\phi(E)$ in $H^*_{\mathcal{D}}(X,\mathbb{R}(*))$.*

Proof. By theorem 2.42 and theorem 3.5, the inclusion

$$\mathcal{D}^*(E_X,*) \longrightarrow \mathcal{D}^*(E_X\langle\langle D\rangle\rangle,*)$$

is a quasi-isomorphism. Moreover, if h' is a smooth hermitian metric on E, then, in the complex $\mathcal{D}^*(E_X\langle\langle D\rangle\rangle,*)$, we have

$$\phi(E,h)-\phi(E,h') = \mathrm{d}_{\mathcal{D}}\,\phi(E,h',h).$$

Therefore, both forms represent the same class. □

Bott-Chern classes.

Definition 4.62. Let X be a complex manifold and D a normal crossing divisor. Let

$$\overline{\mathcal{E}}: 0 \longrightarrow (E',h') \longrightarrow (E,h) \longrightarrow (E'',h'') \longrightarrow 0$$

be an exact sequence of hermitian vector bundles log-singular along D. Let h'_s, h_s, and h''_s be smooth hermitian metrics on E', E, and E'', respectively. We denote by $\overline{\mathcal{E}}_s$ the corresponding exact sequence of smooth vector bundles. Let ϕ be a symmetric power series. Then, the *Bott-Chern class* associated to $\overline{\mathcal{E}}$ is the class represented by

$$\phi(\overline{\mathcal{E}}_s) + \phi(E' \oplus E'', h'_s \oplus h''_s, h' \oplus h') - \phi(E, h_s, h)$$

in the group

$$\bigoplus_k \widetilde{\mathcal{D}}^{2k-1}(E_X\langle\langle D\rangle\rangle, k) = \bigoplus_k \mathcal{D}^{2k-1}(E_X\langle\langle D\rangle\rangle, k) \,/\mathrm{d}_{\mathcal{D}}\, \mathcal{D}^{2k-2}(E_X\langle\langle D\rangle\rangle, k)\ .$$

This class is denoted by $\widetilde{\phi}(\overline{\mathcal{E}})$.

Proposition 4.63. *The Bott-Chern classes are well defined.*

Proof. The fact that the Bott-Chern forms belong to the group

$$\bigoplus_k \mathcal{D}^{2k-1}(E_X\langle\langle D\rangle\rangle)$$

is proven in theorem 4.36.
Let h'_{sa}, h_{sa} and h''_{sa} be another choice of smooth hermitian metrics and let $\overline{\mathcal{E}}_{sa}$ be the corresponding exact sequence. We denote by $\overline{\mathcal{C}}$ the exact square of smooth hermitian vector bundles

$$0 \longrightarrow 0 \longrightarrow \overline{\mathcal{E}}_{sa} \longrightarrow \overline{\mathcal{E}}_s \longrightarrow 0.$$

Then, we have

$$\begin{aligned}
&\phi(\overline{\mathcal{E}}_s) + \phi(E' \oplus E'', h'_s \oplus h''_s, h' \oplus h') - \phi(E, h_s, h)\\
&\qquad - \phi(\overline{\mathcal{E}}_{sa}) - \phi(E' \oplus E'', h'_{sa} \oplus h''_{sa}, h' \oplus h') + \phi(E, h_{sa}, h) =\\
&\mathrm{d}_{\mathcal{D}}\, \phi(\overline{\mathcal{C}}) - \mathrm{d}_{\mathcal{D}}\, \phi(E' \oplus E'', h'_s \oplus h''_s, h'_{sa} \oplus h''_{sa}, h' \oplus h'') + \mathrm{d}_{\mathcal{D}}\, \phi(E, h_s, h_{sa}, h).
\end{aligned}$$

Therefore, the Bott-Chern classes do not depend on the choice of the smooth metrics. □

Axiomatic characterization of Bott-Chern classes.

Theorem 4.64. *The Bott-Chern classes satisfy the following properties. If X is a complex manifold, D is a normal crossing divisor, and*

$$\overline{\mathcal{E}} : 0 \longrightarrow (E', h') \longrightarrow (E, h) \longrightarrow (E'', h'') \longrightarrow 0$$

is a short exact sequence of hermitian vector bundles, log-singular along D, then we have *(i)* $\mathrm{d}_{\mathcal{D}}\,\widetilde{\phi}(\overline{\mathcal{E}})=\phi(E'\oplus E'',h'\oplus h'')-\phi(E,h)$.

(ii) If $(E,h)=(E'\oplus E'',h'\oplus h'')$, *then* $\widetilde{\phi}(\overline{\mathcal{E}})=0$.

(iii) If X' *is another complex manifold,* D' *is a normal crossing divisor in* X', *and* $f:X'\longrightarrow X$ *is a holomorphic map such that* $f^{-1}(D)\subseteq D'$, *then* $\widetilde{\phi}(f^{*}\overline{\mathcal{E}})=f^{*}\widetilde{\phi}(\overline{\mathcal{E}})$.

(iv) If $\overline{\mathcal{F}}$ *is a hermitian exact square of vector bundles on* X, *log-singular along* D, *then*

$$\widetilde{\phi}(\partial_1^{-1}\overline{\mathcal{F}}\oplus\partial_1^{1}\overline{\mathcal{F}})-\widetilde{\phi}(\partial_1^{0}\overline{\mathcal{F}})-\widetilde{\phi}(\partial_2^{-1}\overline{\mathcal{F}}\oplus\partial_2^{1}\overline{\mathcal{F}})+\widetilde{\phi}(\partial_2^{0}\overline{\mathcal{F}})=0.$$

Moreover, these properties determine the Bott-Chern classes.

Proof. First we prove the unicity. By [17], 1.3.2 (see also [36], IV.3.1) properties (1) to (3) characterize the Bott-Chern classes in the case $D=\emptyset$. By functoriality, the Bott-Chern classes are determined for short exact sequences, when the three metrics are smooth. Let E be a vector bundle, h a smooth hermitian metric on E and h' a hermitian metric log-singular along D. The vector bundle $\widetilde{E}=\mathrm{tr}_1(E,h,h')$ over $X\times\mathbb{P}^1$ is isomorphic (as a vector bundle) to $p_1^{*}E$. Let h_1 be the hermitian metric on $\widetilde{E}$ induced by h and this isomorphism. Then, h_1 is a smooth hermitian metric. Let h_2 be the metric of definition 4.4. Let $\pi:Z\longrightarrow X\times\mathbb{P}^1$ and C be as in proposition 4.59. By this proposition $\pi^{*}(\widetilde{E},h_2)$ is log-singular along C. Therefore, we can assume the existence of the Bott-Chern class $\widetilde{\phi}(\widetilde{E},h_1,h_2)$. Write $\pi'=p_1\circ\pi$. We consider the integral

$$I=-\frac{1}{2\pi i}\int_{\pi'}-2\partial\bar{\partial}\widetilde{\phi}(\widetilde{E},h_1,h_2)\pi^{*}(\frac{1}{2}\log(t\bar{t})).$$

By property (1), we have

$$\begin{aligned}I&=-\frac{1}{2\pi i}\int_{\pi'}\phi(\widetilde{E},h_2)\pi^{*}(\frac{1}{2}\log(t\bar{t}))+\frac{1}{2\pi i}\int_{\pi'}\phi(\widetilde{E},h_1)\pi^{*}(\frac{1}{2}\log(t\bar{t}))\\&=-\frac{1}{2\pi i}\int_{\mathbb{P}^1}\phi(\mathrm{tr}_1(E,h,h'))\frac{1}{2}\log(t\bar{t}),\end{aligned}$$

because the second integral vanishes. But using Stokes theorem and properties (2) and (3) as in [17], 1.3.2, or [36], IV.3.1, we get

$$\begin{aligned}I&\sim j_{\infty}^{*}\widetilde{\phi}(\widetilde{E},h_1,h_2)-j_0^{*}\widetilde{\phi}(\widetilde{E},h_1,h_2)\\&=\widetilde{\phi}(E,h,h')-\widetilde{\phi}(E,h,h)\\&=\widetilde{\phi}(E,h,h'),\end{aligned}$$

where the symbol $\sim$ means equality up to the image of $\mathrm{d}_{\mathcal{D}}$. Therefore, the class $\widetilde{\phi}(E,h,h')$ is also determined by properties (1) to (3). Finally, for an arbitrary

exact sequence $\overline{\mathcal{E}}$ of hermitian vector bundles log-singular along D, property (4) implies that $\widetilde{\phi}(\overline{\mathcal{E}})$ is given by definition 4.62.
Next we prove the existence. By proposition 4.63 it only remains to show that the Bott-Chern classes defined by 4.62 satisfy properties (1) to (4). Property (1) is known for smooth metrics. If E is a vector bundle, h is a smooth hermitian metric and h' a hermitian metric log-singular along D, then, since two differential forms that agree in an open dense subset are equal, by the smooth case

$$\mathrm{d}_{\mathcal{D}}\, \widetilde{\phi}(E,h,h') = \phi(E,h') - \phi(E,h).$$

The general case follows from these two cases. Property (2) follows directly from the case of smooth metrics and definition 4.62. Property (3) is obvious from the functoriality of the definition. To prove property (4), we consider $\overline{\mathcal{F}}'$, an exact square with the same vector bundles as $\overline{\mathcal{F}}$, but with smooth metrics. Then, if we use the definition of Bott-Chern classes in the expression

$$\widetilde{\phi}(\partial_1^{-1}\overline{\mathcal{F}} \oplus \partial_1^1\overline{\mathcal{F}}) - \widetilde{\phi}(\partial_1^0\overline{\mathcal{F}}) - \widetilde{\phi}(\partial_2^{-1}\overline{\mathcal{F}} \oplus \partial_2^1\overline{\mathcal{F}}) + \widetilde{\phi}(\partial_2^0\overline{\mathcal{F}}) = 0,$$

all the change of metric terms appear twice with opposite sign. Therefore, this property follows from the smooth case. □

Real varieties. The following result follows easily.

Proposition 4.65. *Let $X_{\mathbb{R}} = (X, F_\infty)$ be a real variety, D a normal crossing divisor on $X_{\mathbb{R}}$, E a complex vector bundle defined over $\mathbb{R}$, h, h' (resp. h") smooth hermitian metrics (resp. log-singular hermitian metric) on E invariant under complex conjugation. Then, the forms $\phi(E,h'')$, $\widetilde{\phi}_1(E,h,h'')$ and $\widetilde{\phi}_2(E,h,h',h'')$ belong to the group*

$$\bigoplus_k \mathcal{D}^{2k-1}(E_{X_{\mathbb{R}}}\langle\langle D\rangle\rangle, k) = \bigoplus_k \mathcal{D}^{2k-1}(E_{X_{\mathbb{R}}}\langle\langle D\rangle\rangle, k)^{\sigma},$$

where σ is the involution that acts as complex conjugation on the space and on the coefficients. □

5 Arithmetic K-theory of log-singular hermitian vector bundles

The arithmetic intersection theory of Gillet and Soulé is complemented by an arithmetic K-theory and a theory of characteristic classes. In this section we will generalize both theories to cover the kind of singular hermitian metrics that appear naturally when considering (fully decomposed) automorphic vector bundles. If E be a vector bundle over a quasi-projective complex manifold X, then a hermitian metric h on E may have arbitrary singularities near the boundary of X. Therefore, the associated Chern forms will also have arbitrary singularities "at infinity". Thus, in order to define arithmetic characteristic classes for this kind of hermitian vector bundles, we are led to use the complex $\mathcal{D}_{\mathrm{l,ll,a}}$.

5.1 Arithmetic Chern classes of log-singular herm. vector bundles

Arithmetic Chow groups with coefficients. Let A be an arithmetic ring. Let $\widehat{X} = (X, \mathcal{C})$ be a $\mathcal{D}_{\log}$-arithmetic variety over A. Let B be a subring of $\mathbb{R}$. We will define the arithmetic Chow groups of $\widehat{X}$ with coefficients in B using the same method as in [3]. We follow the notations of [10], §4.2.
For an integer $p \geq 0$, let $\mathrm{Z}^p(X)_B = \mathrm{Z}^p(X) \otimes B$ be the group of algebraic cycles of X with coefficients in B. Then, the *group of p-codimensional arithmetic cycles of $\widehat{X} = (X, \mathcal{C})$ with coefficients in B* is given by

$$\widehat{\mathrm{Z}}^p_B(X, \mathcal{C}) = \left\{ (y, \mathfrak{g}_y) \in \mathrm{Z}^p_B(X) \oplus \widehat{H}^{2p}_{\mathcal{C}, \mathcal{Z}^p}(X, p) \,\middle|\, \mathrm{cl}(y) = \mathrm{cl}(\mathfrak{g}_y) \right\}.$$

Let $\widehat{\mathrm{Rat}}^p_B(X, \mathcal{C})$ be the B-submodule of $\widehat{\mathrm{Z}}^p_B(X, \mathcal{C})$ generated by $\widehat{\mathrm{Rat}}^p(X, \mathcal{C})$. We define *the p-th arithmetic Chow group of $\widehat{X} = (X, \mathcal{C})$ with coefficients in B* by

$$\widehat{\mathrm{CH}}^p_B(X, \mathcal{C}) = \widehat{\mathrm{Z}}^p_B(X, \mathcal{C}) \big/ \widehat{\mathrm{Rat}}^p_B(X, \mathcal{C}).$$

There is a canonical morphism

$$\widehat{\mathrm{CH}}^p_B(X, \mathcal{C}) \longrightarrow \widehat{\mathrm{CH}}^p(X, \mathcal{C}) \otimes B.$$

For instance, if $B = \mathbb{Q}$, this morphism an isomorphism, but in general, if $B = \mathbb{R}$, it is not an isomorphism.

The main theorem. Let X be a regular scheme, flat and quasi-projective over $\mathrm{Spec}(A)$. Let D be a normal crossing divisor on $X_{\mathbb{R}}$. Write $\underline{X} = (X, D)$. Then, $(X, \mathcal{D}_{\mathrm{l,ll,a},\underline{X}})$ is a quasi-projective $\mathcal{D}_{\log}$-arithmetic variety over A. A log-singular hermitian vector bundle over X is a vector bundle E over X together with a metric on E_∞, which is smooth over $X_\infty \setminus D_\infty$, log-singular along D_∞, and invariant under complex conjugation.

Theorem 5.1. *Let $\phi \in B[[T_1, \dots, T_n]]$ be a symmetric power series with coefficients in a subring B of $\mathbb{R}$. Then, there is a unique way to attach to every log-singular hermitian vector bundle $\overline{E} = (E, h)$ of rank n over $\underline{X} = (X, D)$ a characteristic class*

$$\widehat{\phi}(\overline{E}) \in \widehat{\mathrm{CH}}^*_B(X, \mathcal{D}_{\mathrm{l,ll,a}})$$

having the following properties:

(i) Functoriality. *When $f : Y \longrightarrow X$ is a morphism of regular schemes, flat and quasi-projective over A, D' a normal crossing divisor on $Y_{\mathbb{R}}$ with $f^{-1}(D) \subseteq D'$, and $\overline{E}$ a log-singular hermitian vector bundle on X, then*

$$f^*(\widehat{\phi}(\overline{E})) = \widehat{\phi}(f^*\overline{E}).$$

(ii) Normalization. *When $\overline{E} = \overline{L}_1 \oplus \dots \oplus \overline{L}_n$ is an orthogonal direct sum of hermitian line bundles, then*

$$\widehat{\phi}(\overline{E}) = \phi(\widehat{\mathrm{c}}_1(\overline{L}_1), \dots, \widehat{\mathrm{c}}_1(\overline{L}_n)).$$

(iii) Twist by a line bundle. *Let*

$$\phi(T_1 + T, \dots, T_n + T) = \sum_{i \geq 0} \phi_i(T_1, \dots, T_n) T^i.$$

Let $\overline{L}$ be a log-singular hermitian line bundle, then

$$\widehat{\phi}(\overline{E} \otimes \overline{L}) = \sum_i \widehat{\phi}_i(\overline{E})\, \widehat{\mathrm{c}}_1(\overline{L}).$$

(iv) Compatibility with characteristic forms.

$$\omega(\widehat{\phi}(\overline{E})) = \phi(E, h).$$

(v) Compatibility with the change of metrics. *If h' is another log-singular hermitian metric, then*

$$\widehat{\phi}(E, h) = \widehat{\phi}(E, h') + \mathrm{a}(\widetilde{\phi}_1(E, h', h)).$$

(vi) Compatibility with the definition of Gillet and Soulé. *If D is empty, let ψ be the isomorphism $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{CH}}^*(X)$ of theorem 3.33 and let $\widehat{\phi}_{\mathrm{GS}}(\overline{E}) \in \widehat{\mathrm{CH}}^*(X)$ be the characteristic class defined in [17]. Then*

$$\psi(\widehat{\phi}(\overline{E})) = \widehat{\phi}_{\mathrm{GS}}(\overline{E}).$$

Proof. If D is empty, we define $\widehat{\phi}(\overline{E}) = \psi^{-1}(\widehat{\phi}_{\mathrm{GS}}(\overline{E}))$. If D is not empty, but h is smooth on the whole of $X_{\mathbb{R}}$, then we define $\widehat{\phi}(\overline{E})$ by functoriality, using the tautological morphism $(X, D) \longrightarrow (X, \emptyset)$.
If D is not empty and $\overline{E} = (E, h)$ is a log-singular hermitian vector bundle, we choose any smooth metric h', invariant under F_∞. Then, we define

$$\widehat{\phi}(\overline{E}) = \widehat{\phi}(E, h') + \mathrm{a}(\widetilde{\phi}(E, h', h)).$$

This definition is independent of the choice of the metric h', because, if h'' is another smooth F_∞-invariant metric, then

$$\begin{aligned} \widehat{\phi}(E, h') + {} & \mathrm{a}(\widetilde{\phi}(E, h', h)) - \widehat{\phi}(E, h'') - \mathrm{a}(\widetilde{\phi}(E, h'', h)) \\ &= \mathrm{a}(\widetilde{\phi}(E, h'', h')) + \mathrm{a}(\widetilde{\phi}(E, h', h)) + \mathrm{a}(\widetilde{\phi}(E, h, h'')) \\ &= \mathrm{a}(\mathrm{d}_{\mathcal{D}}(E, h'', h', h)) \\ &= 0. \end{aligned}$$

All the properties stated in the theorem can be checked as in [17]. □

Remark 5.2. If X is projective, the groups $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll}})$ and $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}})$ agree. Therefore, the arithmetic characteristic classes also belong to the former group. When X is quasi-projective, in order to define characteristic classes in the group $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll}})$, we have to impose conditions on the behavior of the hermitian metrics at infinity. For instance, one may consider smooth at infinity hermitian metrics (see [11]).

Remark 5.3. If we replace good hermitian vector bundle by log hermitian vector bundle and pre-log-log forms by log-log forms (implicit in the definition of $\mathcal{D}_{\mathrm{l,ll,a}}$) in theorem 5.1, the result remains true.

5.2 Arithmetic K-theory of log-singular hermitian vector bundles

Log-singular arithmetic K-theory. We want to generalize the definition of arithmetic K-theory given by Gillet and Soulé in [17] to cover log-singular hermitian metrics.
We write

$$\widetilde{\mathcal{D}}_{\mathrm{l,ll,a}}(X) = \bigoplus_p \widetilde{\mathcal{D}}^{2p-1}_{\mathrm{l,ll,a}}(X,p),$$

$$\mathrm{Z}\mathcal{D}_{\mathrm{l,ll,a}}(X) = \bigoplus_p \mathrm{Z}\mathcal{D}^{2p}_{\mathrm{l,ll,a}}(X,p).$$

Let ch be the power series associated with the Chern character. In particular, it induces Bott-Chern forms $\widetilde{\mathrm{ch}}$ and arithmetic characteristic classes $\widehat{\mathrm{ch}}$.

Definition 5.4. Let $\underline{X}$ be as in theorem 5.1. Then, the group $\widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}})$ is the group generated by pairs $(\overline{E}, \eta)$, where $\overline{E}$ is a log-singular hermitian metric on X and $\eta \in \widetilde{\mathcal{D}}_{\mathrm{l,ll,a}}(X)$ satisfying the relations

$$(\overline{S}, \eta') + (\overline{Q}, \eta'') = (\overline{E}, \eta' + \eta'' + \widetilde{\mathrm{ch}}(\overline{\mathcal{E}}))$$

for every $\eta', \eta'' \in \widetilde{\mathcal{D}}_{\mathrm{l,ll,a}}(X)$ and every short exact sequence of log-singular hermitian vector bundles

$$\overline{\mathcal{E}} : 0 \longrightarrow \overline{S} \longrightarrow \overline{E} \longrightarrow \overline{Q} \longrightarrow 0.$$

If D is empty, then this definition agrees with the definition of Gillet and Soulé in [17].

Basic properties. The following theorem summarizes the basic properties of the arithmetic K-theory groups. They are a consequence of the corresponding results of [17] together with theorem 5.1.

Theorem 5.5. *Let $\underline{X} = (X, D)$ be an arithmetic variety over A with a fixed normal crossing divisor. Then, we have*

(i) *There are natural maps*

$$\begin{aligned}
\mathrm{a} : \widetilde{\mathcal{D}}_{\mathrm{l,ll,a}}(X) &\longrightarrow \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}}),\\
\mathrm{ch} : \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}}) &\longrightarrow \mathrm{Z}\mathcal{D}_{\mathrm{l,ll,a}}(X),\\
\mathrm{v} : \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}}) &\longrightarrow \mathrm{K}_0(X),\\
\widehat{\mathrm{ch}} : \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}}) &\longrightarrow \bigoplus \widehat{\mathrm{CH}}^p_{\mathbb{Q}}(X, \mathcal{D}_{\mathrm{l,ll,a}}),
\end{aligned}$$

given by

$$\begin{aligned}
\mathrm{a}(\eta) &= (0, \eta),\\
\mathrm{ch}([\overline{E}, \eta]) &= \mathrm{ch}(\overline{E}) + \mathrm{d}_{\mathcal{D}}\, \eta,\\
\mathrm{v}([\overline{E}, \eta]) &= [E],\\
\widehat{\mathrm{ch}}([\overline{E}, \eta]) &= \widehat{\mathrm{ch}}(\overline{E}) + \mathrm{a}(\eta).
\end{aligned}$$

(ii) *The product*

$$(\overline{E}, \eta) \otimes (\overline{E}', \eta') = \Big(\overline{E} \otimes \overline{E}', (\mathrm{ch}(\overline{E}) + \mathrm{d}_{\mathcal{D}}\, \eta) \bullet \eta' + \eta \bullet \mathrm{ch}(\overline{E}')\Big)$$

induces a commutative and associative ring structure on $\widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}})$. *The maps* v, ch, *and* $\widehat{\mathrm{ch}}$ *are compatible with this ring structure.*

(iii) *If* $\underline{Y} = (Y, D')$ *is another arithmetic variety over* A *with a fixed normal crossing divisor and* $f : X \longrightarrow Y$ *is a morphism such that* $f^{-1}(D') \subseteq D$, *then there is a pull-back morphism*

$$f^* : \widehat{\mathrm{K}}_0(Y, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}}),$$

compatible with the maps a, ch, v *and* $\widehat{\mathrm{ch}}$.

(iv) *There are exact sequences*

$$\mathrm{K}_1(X) \xrightarrow{\rho} \widetilde{\mathcal{D}}_{\mathrm{l,ll,a}}(X) \xrightarrow{\mathrm{a}} \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}}) \xrightarrow{\mathrm{v}} \mathrm{K}_0(X) \to 0, \tag{5.6}$$

and

$$\begin{aligned}
\mathrm{K}_1(X) \xrightarrow{\rho} \bigoplus_p H^{2p-1}_{\mathcal{D}_{\mathrm{l,ll,a}}}(X, p) \xrightarrow{\mathrm{a}} \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}}) \xrightarrow{\mathrm{v} + \mathrm{ch}}\\
\mathrm{K}_0(X) \oplus \mathrm{Z}\mathcal{D}_{\mathrm{l,ll,a}}(X) \to \bigoplus_p H^{2p}_{\mathcal{D}_{\mathrm{l,ll,a}}}(X, p) \to 0.
\end{aligned} \tag{5.7}$$

In these exact sequences the map ρ *is the composition*

$$\mathrm{K}_1(X) \to \bigoplus_p H^{2p-1}_{\mathcal{D}}(X_{\mathbb{R}}, \mathbb{R}(p)) \to \bigoplus_p H^{2p-1}_{\mathcal{D}_{\mathrm{l,ll,a}}}(X, p) \subseteq \widetilde{\mathcal{D}}_{\mathrm{l,ll,a}}(X),$$

where the first map is Beilinson's regulator.

(v) The Chern character

$$\widehat{\mathrm{ch}}:\widehat{\mathrm{K}}_0(X,\mathcal{D}_{\mathrm{l,ll,a}})\otimes\mathbb{Q}\longrightarrow\bigoplus\widehat{\mathrm{CH}}^p_{\mathbb{Q}}(X,\mathcal{D}_{\mathrm{l,ll,a}})$$

is a ring isomorphism. □

5.3 Variant for non regular arithmetic varieties

Since there is no general theorem of resolution of singularities, it is useful to extend the theory of arithmetic Chow groups to the case of non regular arithmetic varieties.

Arithmetic Chow groups for non regular arithmetic varieties. Let (A,Σ,F_∞) be an arithmetic ring with fraction field F. We will assume that A is equidimensional and Jacobson. In contrast to the rest of the paper, in this section, an arithmetic variety over A will be a scheme X that is quasi-projective and flat over $\mathrm{Spec}(A)$, and such that the generic fiber X_F is smooth, but that not need be regular. Since X_F is smooth, the analytic variety X_Σ is a disjoint union of connected components X_i that are equidimensional of dimension d_i. For every cohomological complex of sheaves $\mathcal{F}^*(*)$ on X_Σ we write

$$\mathcal{F}_n(p)(U)=\bigoplus_i\mathcal{F}^{2d_i-n}(d_i-p)(U\cap X_i).$$

Then, the definition of Green objects and of arithmetic Chow groups of [10] can easily be adapted to the grading by dimension.
In this way we can define, for X regular, homological Chow groups with respect to any $\mathcal{D}_{\log}$-complex $\mathcal{C}$. These homological Chow groups will be denoted by $\widehat{\mathrm{CH}}_*(X,\mathcal{C})$. In particular, we are interested in the groups $\widehat{\mathrm{CH}}_*(X,\mathcal{D}_{\mathrm{l,ll,a}})$. But now we can proceed as in [18] and we can extend the definition to the case of non regular arithmetic varieties.

Basic properties of homological Chow groups. Following [18], we can extend some of the properties of the arithmetic Chow groups to the non regular case. The proof of the next results are as in [18], 2.2.7, 2.3.1, and 2.4.2 for the algebraic cycles, but using the techniques of [10] for the Green objects.

Theorem 5.8. *Let $f:X\longrightarrow Y$ be a morphism of irreducible arithmetic varieties over A which is flat or l.c.i. Let D_Y be a normal crossing divisor on $Y_\mathbb{R}$ and D_X a normal crossing divisor on $X_\mathbb{R}$ such that $f^{-1}(D_Y)\subseteq D_X$. Write $\underline{X}=(X_\mathbb{R},D_X)$ and $\underline{Y}=(Y_\mathbb{R},D_Y)$. Then, there is defined an inverse image morphism*

$$f^*:\widehat{\mathrm{CH}}_p(Y,\mathcal{D}_{\mathrm{l,ll,a}})\longrightarrow\widehat{\mathrm{CH}}_{p+d}(X,\mathcal{D}_{\mathrm{l,ll,a}}),$$

where d is the relative dimension. □

Theorem 5.9. *Let $f:X\longrightarrow Y$ be a map of arithmetic varieties with Y regular. Let D_Y be a normal crossing divisor on $Y_\mathbb{R}$ and D_X a normal crossing divisor* *on $X_{\mathbb{R}}$ such that $f^{-1}(D_Y) \subseteq D_X$. Write $\underline{X} = (X_{\mathbb{R}}, D_X)$ and $\underline{Y} = (Y_{\mathbb{R}}, D_Y)$. Then, there is a cap product*

$$\widehat{\mathrm{CH}}^p(Y, \mathcal{D}_{\mathrm{l,ll,a}}) \otimes \widehat{\mathrm{CH}}_q(X, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{CH}}_{q-p}(X, \mathcal{D}_{\mathrm{l,ll,a}});$$

for $x \in \widehat{\mathrm{CH}}^p(Y, \mathcal{D}_{\mathrm{l,ll,a}})$ and $y \in \widehat{\mathrm{CH}}_q(X, \mathcal{D}_{\mathrm{l,ll,a}})$ we denote it by $x._f y$. This cap product turns $\widehat{\mathrm{CH}}_*(X, \mathcal{D}_{\mathrm{l,ll,a}})$ into a graded $\widehat{\mathrm{CH}}^*(Y, \mathcal{D}_{\mathrm{l,ll,a}})$-module. Moreover, it is compatible with inverse images (when defined). □

Arithmetic K-theory. The definition of arithmetic K-theory carries over to the case of non regular arithmetic varieties without modification (see [18], 2.4.2). Thus, we obtain a contravariant functor $(X, D) \longmapsto \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}})$ from arithmetic varieties with a fixed normal crossing divisor to rings.

Theorem 5.10. *Let X be an arithmetic variety. Let D_X a normal crossing divisor on $X_{\mathbb{R}}$. Write $\underline{X} = (X_{\mathbb{R}}, D_X)$. Then, there is a biadditive pairing*

$$\widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}}) \otimes \widehat{\mathrm{CH}}_*(X, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{CH}}_*(X, \mathcal{D}_{\mathrm{l,ll,a}})_{\mathbb{Q}},$$

which we write as $\alpha \otimes x \mapsto \widehat{\mathrm{ch}}(\alpha) \cap x$, with the following properties

(i) Let $f : X \longrightarrow Y$ be a morphism of arithmetic varieties, with Y regular. Let D_Y be a normal crossing divisor on $Y_{\mathbb{R}}$ such that $f^{-1}(D_Y) \subseteq D_X$. Write $\underline{Y} = (Y_{\mathbb{R}}, D_Y)$. If $\alpha \in \widehat{\mathrm{K}}_0(Y, \mathcal{D}_{\mathrm{l,ll,a}})$ and $x \in \widehat{\mathrm{CH}}_*(X, \mathcal{D}_{\mathrm{l,ll,a}})$, then

$$\widehat{\mathrm{ch}}(f^*\alpha) \cap x = \widehat{\mathrm{ch}}(\alpha)._f x.$$

(ii) If $(0, \eta) \in \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}})$ and $x \in \widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}})$, then

$$\widehat{\mathrm{ch}}((0, \eta)) \cap x = \mathrm{a}(\eta\omega(x)).$$

(iii) If $\alpha \in \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}})$ and $x \in \widehat{\mathrm{CH}}_*(X, \mathcal{D}_{\mathrm{l,ll,a}})$, then

$$\omega(\widehat{\mathrm{ch}}(\alpha) \cap x) = \mathrm{ch}(\alpha) \wedge \omega(x).$$

(iv) The pairing makes $\widehat{\mathrm{CH}}_*(X, \mathcal{D}_{\mathrm{l,ll,a}})_{\mathbb{Q}}$ into a $\widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}})$-module, i.e., for all $\alpha, \beta \in \widehat{\mathrm{K}}_0(X)$, and $x \in \widehat{\mathrm{CH}}^*(X)$, we have

$$\widehat{\mathrm{ch}}(\alpha) \cap (\widehat{\mathrm{ch}}(\beta) \cap x) = \widehat{\mathrm{ch}}(\alpha\beta) \cap x.$$

(v) If $f : X \longrightarrow Y$ is a flat or l.c.i. morphism of arithmetic varieties, let $\alpha \in \widehat{\mathrm{K}}_0(X, \mathcal{D}_{\mathrm{l,ll,a}})$ and $x \in \widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}})$. Then

$$\widehat{\mathrm{ch}}(f^*\alpha) \cap f^*x = f^*(\widehat{\mathrm{ch}}(\alpha) \cap x).$$

Proof. Follow [18], 2.4.2, but using theorem 4.64 to prove the independence of the choices. □

5.4 Some remarks on the properties of $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}})$

In [31], V. Maillot and D. Roessler have announced a preliminary version of the theory developed in this paper. The final theory has some minor differences that do not affect the heart of [31]. The aim of this section is to compare both theories.

We fix an arithmetic ring (A, Σ, F_∞), and we consider pairs $\underline{X} = (X, D)$, where X is an arithmetic variety over A and D is a normal crossing divisor of X_Σ, invariant under F_∞.

A log-singular hermitian vector bundle $\overline{E}$ is a pair (E, h), where E is a vector bundle over X and h is a hermitian metric on E_Σ, invariant under F_∞ and log-singular along D. Observe that the notion of log-singular hermitian metric is not the same as the notion of good hermitian metric. This is not important by two reasons. First, as we will see in the next section, the main examples of good hermitian vector bundles, the fully decomposed automorphic vector bundles, are good and log-singular. Second, if one insists in using good hermitian vector bundles, one can replace pre-log and pre-log-log forms by log and log-log forms to obtain an analogous theory. This alternative theory has worse cohomological properties (we have not proven the Poincaré lemma for pre-log and pre-log-log forms), but the arithmetic intersection numbers computed by both theories agree.

To each pair $\underline{X} = (X, D)$, we have assigned an $\mathbb{N}$-graded abelian group $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}})$ that satisfies, among others, the following properties:

(i) The group $\widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}})$ is equipped with an associative, commutative and unitary ring structure, compatible with the grading.

(ii) If X is proper over $\operatorname{Spec} A$, there is a direct image group homomorphism $f_* : \widehat{\mathrm{CH}}^{d+1}(X, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{CH}}^1(\operatorname{Spec} A)$, where d is the relative dimension.

(iii) For every integer $r \geq 0$ and every log-singular hermitian vector bundle there is defined the arithmetic r-th Chern class $\widehat{\mathrm{c}}_r(\overline{E}) \in \widehat{\mathrm{CH}}^r(X, \mathcal{D}_{\mathrm{l,ll,a}})$.

(iv) Let $g : X \longrightarrow Y$ be a morphism of arithmetic varieties over A, and let D and E be normal crossing divisors on $X_\mathbb{R}$ and $Y_\mathbb{R}$, respectively, such that $g^{-1}(E) \subseteq D$. Write $\underline{X} = (X_\mathbb{R}, D)$ and $\underline{Y} = (Y_\mathbb{R}, E)$. Then, there is defined an inverse image morphism
$$g^* : \widehat{\mathrm{CH}}^*(Y, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}}).$$
Moreover, it is a morphism of rings after tensoring with $\mathbb{Q}$.

(v) For every $r \geq 0$, it holds the equality $g^*(\widehat{\mathrm{c}}_r(\overline{E})) = \widehat{\mathrm{c}}_r(g^*(\overline{E}))$.

(vi) There is a forgetful morphism $\zeta : \widehat{\mathrm{CH}}^*(X, \mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \mathrm{CH}^*(X)$, compatible with inverse images and Chern classes.

(vii) There is a complex of groups

$$H^{2p-1}_{\mathcal{D}}(X_{\mathbb{R}},\mathbb{R}(p)) \xrightarrow{\mathrm{a}} \widehat{\mathrm{CH}}^{p}(X,\mathcal{D}_{\mathrm{l,ll,a}}) \xrightarrow{(\zeta,\omega)} \mathrm{CH}^{p}(X)\oplus \mathrm{Z}\mathcal{D}^{2p}_{\mathrm{l,ll,a}}(X,p)$$

that is an exact sequence when X_{Σ} is projective. Observe that the group $\mathrm{Z}\mathcal{D}^{2p}_{\mathrm{l,ll,a}}(X,p)$ does not agree with the group denoted by $\mathrm{Z}^{p,p}(X(\mathbb{C}),D)$ in [31], §1 (7). The former is made of forms that are log-log along D and the latter by forms that are good along D. Again, this is not important by two reasons. First, the image by ω of the arithmetic Chern classes of fully decomposed automorphic vector bundles lies in the intersection of the good and log-log forms. Second, the complex of log-log forms shares all the important properties of the complex of good forms (see proposition 2.26).

(viii) The morphism (ζ,ω) is a ring homomorphism; the image of a is a square zero ideal. Moreover, it holds the equality

$$\mathrm{a}(x)\cdot y = \mathrm{a}(x\cdot \mathrm{cl}(\zeta(y))),$$

where $x \in H^{2p-1}_{\mathcal{D}}(X_{\mathbb{R}},\mathbb{R}(p))$, $y \in \widehat{\mathrm{CH}}^{p}(X,\mathcal{D}_{\mathrm{l,ll,a}})$, cl is the class map, the product on the left hand side is the product in the arithmetic Chow groups and the product on the right hand side is the product in Deligne-Beilinson cohomology.

(ix) When D is empty, there is a canonical isomorphism $\widehat{\mathrm{CH}}^{*}(X,\mathcal{D}_{\mathrm{l,ll,a}}) \longrightarrow \widehat{\mathrm{CH}}^{*}(X)$, compatible with the previously discussed structures. Note that we have dropped the projectivity assumption in [31], §1 (9). Observe, moreover, that, if we use the alternative theory with pre-log-log forms, then this property is not established.

6 Automorphic vector bundles

6.1 Automorphic bundles and log-singular hermitian metrics

Fully decomposed automorphic vector bundles. Let B be a bounded, hermitian, symmetric domain. Then, by definition $B = G/K$, where G is a semi-simple adjoint group and K is a maximal compact subgroup. Inside the complexification $G_{\mathbb{C}}$ of G, there is a suitable parabolic subgroup of the form $P_{+}\cdot K_{\mathbb{C}}$, with P_{+} its unipotent radical and such that $K = G\cap P_{+}\cdot K_{\mathbb{C}}$ and $G\cdot(P_{+}\cdot K_{\mathbb{C}})$ are open in $G_{\mathbb{C}}$. This induces an open G-equivariant immersion

$$\begin{tikzcd} B \arrow[r, hook, "\iota"] \arrow[d, equal] & \check{B} \arrow[d, equal] \\ G/K \arrow[r, hook] & G_{\mathbb{C}}/P_{+}\cdot K_{\mathbb{C}}. \end{tikzcd}$$

Here, $\check{B} = G_{\mathbb{C}}/P_+ \cdot K_{\mathbb{C}}$ is a projective rational variety, and the immersion ι is compatible with the complex structure of B.
Let $\sigma : K \longrightarrow \mathrm{GL}(n, \mathbb{C})$ be a representation of K. Then, σ defines a G-equivariant vector bundle E_0 on B. We complexify σ and extend it trivially to $P_+ \cdot K_{\mathbb{C}}$ by letting it kill P_+. Then, σ defines a holomorphic $G_{\mathbb{C}}$-equivariant vector bundle $\check{E}_0$ on $\check{B}$ with $E_0 = \iota^*(\check{E}_0)$. This induces a holomorphic structure on E_0. Observe that different extensions of σ to $P_+ \cdot K_{\mathbb{C}}$ will define different holomorphic structures on E_0.
Let Γ be a neat arithmetic subgroup of G acting on B. Then, $X = \Gamma\backslash B$ is a smooth quasi-projective complex variety, and E_0 defines a holomorphic vector bundle E on X. Following [24], the vector bundles obtained in this way (with σ extended trivially) will be called *fully decomposed automorphic vector bundles*. Since we will not treat more general automorphic vector bundles in this paper, we will just call them automorphic vector bundles.
Let h_0 be a G-equivariant hermitian metric on E_0. Such metrics exist by the compactness of K. Then, h_0 determines a hermitian metric h on E.

Definition 6.1. A hermitian vector bundle (E, h) as above will be called an *automorphic hermitian vector bundle*.

Let $\overline{X}$ be a smooth toroidal compactification of X with $D = \overline{X} \setminus X$ a normal crossing divisor. We recall the following result of Mumford (see [34], theorem 3.1).

Theorem 6.2. *The automorphic vector bundle E admits a unique extension to a vector bundle E_1 over $\overline{X}$ such that h is a singular hermitian metric which is good along D.* □

By abuse of notation, the extension (E_1, h) will also be called an automorphic hermitian vector bundle.
Our task now is to improve slightly Mumford's theorem.

Theorem 6.3. *The automorphic hermitian vector bundle (E_1, h) is a ∞-good hermitian vector bundle; therefore, it is log-singular along D.*

Proof. The proof of this result will take the rest of this section. The technique of proof used follows closely the proof of theorem 3.1 in [34]. Instead of repeating the whole proof of Mumford, we will only point out the results needed to bound all the derivatives of the functions involved.

Cones and Jordan algebras. Let V be a real vector space and let $C \subseteq V$ be a homogeneous self-adjoint cone. We refer to [1] for the theory of homogeneous self-adjoint cones and their relationship with Jordan algebras. We will recall here only some basic facts.
Let $G \subseteq \mathrm{GL}(V)$ be the group of linear maps that preserve C. Since C is homogeneous, G acts transitively on C. We will denote by $\mathfrak{g}$ the Lie algebra of

G. For any point $x \in C$, let $K_x = \mathrm{Stab}(x)$. It is a maximal compact subgroup of G. Let $\mathfrak{k}_x$ be the Lie algebra of K_x and let

$$\mathfrak{g} = \mathfrak{k}_x \oplus \mathfrak{p}_x$$

be the associated Cartan decomposition. Let σ_x be the Cartan involution. Let us choose a point $e \in C$. Let $\langle\, , \rangle = \langle\, , \rangle_e$ be a positive definite scalar product such that $\sigma_e(g) = {}^t g^{-1}$ for all $g \in G$. Then, C is self-adjoint with respect to this inner product. For any point $x \in C$, let us choose $g \in G$ such that $x = ge$. We will identify V with $T_{C,x}$. For $t_1, t_2 \in V$, we will write

$$\langle t_1, t_2 \rangle_x = \langle g^{-1} t_1, g^{-1} t_2 \rangle_e.$$

The right hand side is independent of g because $\langle\, , \rangle_e$ is K_e-invariant. These products define a G-invariant Riemannian metric on C, which is denoted by ds^2_C.

The elements of $\mathfrak{g}$ act on V by endomorphisms. This action can be seen as the differential of the G action at $e \in V$, or given by the inclusion $\mathfrak{g} \subseteq \mathfrak{gl}(V)$. For any $x \in C$ there are isomorphisms

$$\mathfrak{p}_x \xrightarrow{\cong} \mathfrak{p}_x.x = V \quad \text{and} \quad P_x = \exp(\mathfrak{p}_x) \xrightarrow{\cong} P_x.x = C.$$

The elements of $\mathfrak{p}_x$ act on V by self-adjoint endomorphisms with respect to $\langle\, , \rangle_x$.

Every $\mathfrak{p}_x$ has a structure of Jordan algebra defined by

$$(\pi.\pi').x = \pi.(\pi'.x).$$

The isomorphism $\mathfrak{p}_x \longrightarrow V$ defines a Jordan algebra structure on V, which we denote by $t_1 \underset{x}{\cdot} t_2$. Observe that x is the unit element for this Jordan algebra structure.

We summarize the compatibility relations between the objects defined so far and the action of the group. Let $x = g.e$:

$$\begin{aligned} K_x &= \mathrm{Ad}(g) K_e = g K_e g^{-1}, \\ \mathfrak{k}_x &= \mathrm{ad}(g) \mathfrak{k}_e = g \mathfrak{k}_e g^{-1}, \\ \mathfrak{p}_x &= \mathrm{ad}(g) \mathfrak{p}_e = g \mathfrak{p}_e g^{-1}. \end{aligned}$$

There is a commutative diagram

$$\begin{tikzcd} \mathfrak{p}_e \arrow[r, "\mathrm{ad}(g)"] \arrow[d, ".e"'] & \mathfrak{p}_x \arrow[d, ".x"] \\ V \arrow[r, "g."] & V \end{tikzcd}$$

The horizontal arrows in the above diagram are morphisms of Jordan algebras. In particular

$$g.(t_1 \underset{e}{\cdot} t_2) = g t_1 \underset{x}{\cdot} g t_2.$$

When a unit element e is chosen, we will write $t_1.t_2$ and $\langle\, ,\, \rangle$ instead of $t_1 \underset{e}{.} t_2$ and $\langle\, ,\, \rangle_e$.

Derivatives with respect to the base point. We now study the derivatives of the scalar product and the Jordan algebra product when we move the base point.

Lemma 6.4. *Let $t_1, t_2, t_3 \in V$. Then, we have*

(i) $D_{t_1}(\langle t_2, x^{-1}\rangle_e) = -\langle t_2, t_1\rangle_x$.

(ii) $D_{t_3}\langle t_1, t_2\rangle_x = -(\langle t_3 \underset{x}{.} t_1, t_2\rangle_x + \langle t_1, t_3 \underset{x}{.} t_2\rangle_x) = -2\langle t_1, t_3 \underset{x}{.} t_2\rangle_x$.

(iii) $D_{t_3}(t_1 \underset{x}{.} t_2) = -((t_3 \underset{x}{.} t_1) \underset{x}{.} t_2 + t_1 \underset{x}{.} (t_3 \underset{x}{.} t_2))$.

Proof. The proof of 1 is in [34], p. 244. To prove 2, write $t_3 = M.x$ with $M \in \mathfrak{p}_x$. Then, $\alpha(\delta) = \exp(\delta M).x$ is a curve with $\alpha(0) = x$ and $\alpha'(0) = t_3$. Therefore, we find

$$\begin{aligned} D_{t_3}\langle t_1, t_2\rangle_x &= \frac{\mathrm{d}}{\mathrm{d}\,\delta}\langle t_1, t_2\rangle_{\exp(\delta M).x}|_{\delta=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}\,\delta}\langle \exp(\delta M)^{-1}.t_1, \exp(\delta M)^{-1}.t_2\rangle_x|_{\delta=0} \\ &= -(\langle M.t_1, t_2\rangle_x + \langle t_1, M.t_2\rangle_x) \\ &= -(\langle t_3.t_1, t_2\rangle_x + \langle t_1, t_3.t_2\rangle_x). \end{aligned}$$

The second equality of 2 follows from the fact that M acts by an endomorphism which is self-adjoint with respect to $\langle\, ,\, \rangle_x$.
The proof of 3 is completely analogous. □

We will denote by $\|\,\|_x$ the norm associated to the inner product $\langle\, ,\, \rangle_x$.

Lemma 6.5. *There is a constant $K > 0$ such that, for all $x \in C$ and $t_1, t_2 \in V$,*

$$\|t_1 \underset{x}{.} t_2\|_x \leq K\|t_1\|_x\|t_2\|_x.$$

Proof. On $\mathfrak{p}_e$ we may define the norm

$$\|M\|'_e = \sup_{t\in V}\frac{\|M.t\|_e}{\|t\|_e}.$$

Via the isomorphism $\mathfrak{p}_e \longrightarrow V$ it induces a norm on V given by

$$\|t_1\|'_e = \sup_{t\in V}\frac{\|t_1 \underset{e}{.} t\|_e}{\|t\|_e}.$$

Since any two norms in a finite dimensional vector space are equivalent, there is a constant $K > 0$ such that

$$\|t\|'_e \leq K\|t\|_e$$

for all t. Therefore, we get

$$\|t_1 \underset{e}{.} t_2\|_e \leq \|t_1\|'_e \|t_2\|_e \leq K \|t_1\|_e \|t_2\|_e.$$

But for any $x = ge$, we have

$$\|t_1 \underset{x}{.} t_2\|_x = \|g^{-1} t_1 \underset{e}{.} g^{-1} t_2\|_e \leq K \|g^{-1} t_1\|_e \|g^{-1} t_2\|_e = K \|t_1\|_x \|t_2\|_x.$$

□

Maximal $\mathbb{R}$-split torus. We fix a unit element $e \in C$. This fixes also the Jordan algebra structure of V, and we write $K = K_e$ and $\mathfrak{p} = \mathfrak{p}_e$. Let $A \subseteq \exp(\mathfrak{p})$ be a maximal $\mathbb{R}$-*split* torus with $A = \exp(\mathfrak{a})$. Then, $\exp(\mathfrak{p}) = K.A.K^{-1}$ and $C = K.A.e$. A useful result, which is proven in [1], II, §3, is the following

Proposition 6.6. *There exist a maximal set of mutually orthogonal idempotents $\epsilon_1, \ldots, \epsilon_r$ of V with $e = \epsilon_1 + \ldots + \epsilon_r$ such that*

$$\mathfrak{a}.e = \sum_{i=1}^{r} \mathbb{R}\epsilon_i \text{ and } A.e = \sum_{i=1}^{r} \mathbb{R}^+ \epsilon_i.$$

Moreover, $C \cap \mathfrak{a}.e = A.e$. □

On A, we can introduce the coordinates given by

$$A \cong A.e = \sum_{i=1}^{r} \mathbb{R}^+ \epsilon_i \cong (\mathbb{R}^+)^r.$$

As an application of the previous result we prove a bound for the norm of x^{-1}.

Lemma 6.7. *Let $\sigma \in C$. There exists a constant K such that $\|x^{-1}\| \leq K$ for all $x \in \sigma + C$.*

Proof. Since $\bigcup_{\lambda > 0}(\lambda e + C) = C$, we may assume that $\sigma = \lambda e$ for some $\lambda > 0$. Since K is compact and

$$\lambda e + C = K(\lambda e + A.e),$$

it is enough to bound x^{-1} for $x \in \lambda e + A.e$. If $x \in \lambda e + A.e$, then we can write, using the above coordinates of A, $x = a.e$ with $a = (a_1, \ldots, a_r)$ and all $a_i \geq \lambda$. Then, $x^{-1} = a^{-1}.e$. Since on a finite dimensional vector space any two norms are equivalent, we obtain

$$\|x^{-1}\|^2 \leq K_1(a_1^{-2} + \ldots + a_k^{-2}) \leq K_2/\lambda^2.$$

□

equivariant symmetric representations. Let C_n be the cone of positive definite $n \times n$ hermitian matrices. An equivariant symmetric representation of dimension n is a pair (ρ, H), where $\rho : G \longrightarrow \mathrm{GL}(n, \mathbb{C})$ is a representation and $H : C \longrightarrow C_n$ is a map such that

(i) (Equivariance) $H(gx) = \rho(g) H(x) {}^t\overline{\rho(g)}$ for all $x \in C$, $g \in G$.

(ii) (Symmetry) $\rho(g^*) = H(e).{}^t\overline{\rho(g)}^{-1}.H(e)^{-1}$ for all $g \in G$.

We will consider an equivariant symmetric representation (ρ, H_t) with H_t depending differentiably on a parameter $t \in T$ with T compact as in [34], pp. 245, 246.

Bounds of H and $\det H^{-1}$. The first step is to bound the entries of H_t and $\det H(t)^{-1}$. This is done in [34], proposition 2.3.

Proposition 6.8. *For all $\sigma \in C$, there is a constant $K > 0$ and an integer N such that*

$$\|H_t(x)\|, \, |\det H_t(x)|^{-1} \leq K \langle x, x\rangle^N \text{ for all } x \in \sigma + C.$$

□

The following results of Mumford (see [34], propositions 2.4 and 2.5) are the starting point to bound the entries of $D_v H_t.H_t^{-1}$; they will also be used to bound the derivatives of H_t.

Proposition 6.9. *Let $\xi \in V$. For all $1 \leq \alpha, \beta \leq n$, let $(D_\xi H_t.H_t^{-1})_{\alpha,\beta}$ be the (α, β)-th entry of this matrix. There is a linear map*

$$C_{\alpha\beta,t} : V \longrightarrow V$$

depending differentiably on t such that

$$(D_\xi H_t.H^{-1})_{\alpha\beta}(x) = \langle C_{\alpha\beta,t}(\xi), x^{-1}\rangle.$$

Moreover, $C_{\alpha\beta,t}$ has the property

$$\left.\begin{matrix} \xi, \eta \in \overline{C} \\ \langle \xi, \eta\rangle = 0 \end{matrix}\right\} \Rightarrow \langle C_{\alpha\beta,t}(\xi), \eta\rangle = 0.$$

□

Proposition 6.10. *For all vector fields δ on T, $\delta H_t.H_t^{-1}(x)$ is independent of x.* □

Proposition 6.11. *Let $\sigma \in C$, let P be a differential operator on T and let $\xi_1, \dots, \xi_d \in V$. Then, there is a constant $K > 0$ and an integer N such that*

$$\|D_{\xi_1} \dots D_{\xi_d} P H_t(x)\|, \, |D_{\xi_1} \dots D_{\xi_d} P \det H_t(x)| \leq K \langle x, x\rangle^N \quad (x \in \sigma + C).$$

Proof. In view of proposition 6.10 and since T is compact, it is enough to consider the case $P = \mathrm{id}$. Now, by proposition 6.9 and the fact that

$$D_{\xi_i} x^{-1} = -x^{-1}.(x^{-1}.\xi_i),$$

we can prove by induction that

$$D_{\xi_1} \dots D_{\xi_d} H_t(x) = M(C(\xi_1, \dots, \xi_d, x)).H_t(x),$$

where $M : V \longrightarrow M_n(\mathbb{C})$ is linear and $C : V \longrightarrow V$ is linear on $\xi_1, \dots, \xi_d$ and polynomial in x^{-1}.
Then, the proposition follows from proposition 6.8 and lemma 6.7. □

Bounds of $\delta H.H^{-1}$. Let $e = \epsilon_1 + \dots + \epsilon_r$ be a maximal set of orthogonal idempotents, and let A be the corresponding $\mathbb{R}$-split maximal torus. Let C_i be the boundary component containing $\epsilon_{i+1} + \dots + \epsilon_r$ (see [1], II, §3). Let $\widetilde{C} = C \cup C_1 \cup \dots \cup C_r \cup 0$ and let P be the parabolic subgroup stabilizing the flag $\{C_i\}$.
In order to be able to use proposition 6.9 to bound $D_v H_t.H_t^{-1}$ and its derivatives, we will need the following result (see [34], proposition 2.6).

Proposition 6.12. *Let $\xi_1, \xi_2 \in \widetilde{C}$, and let $\xi_1' \in V$ satisfy*

$$\left.\begin{matrix} \eta \in \overline{C} \\ \langle \xi_1, \eta \rangle = 0 \end{matrix}\right\} \Rightarrow \langle \xi_1', \eta \rangle = 0.$$

Then, for every compact subset $\omega \subseteq P$, there is a constant $K > 0$ such that

(i) $|\langle \xi_1', x^{-1} \rangle| \leq K \|\xi_1\|_x$ *for all* $x \in \omega.A.e.$

(ii) $|\langle \xi_1', \xi_2 \rangle| \leq K \|\xi_1\|_x \|\xi_2\|_x$ *for all* $x \in \omega.A.e.$

Now we can bound the derivatives of $D_v H_t.H_t^{-1}$ in terms of the Riemannian metric ds^2_C. Let $N = \dim V$ and let $\xi_1, \dots, \xi_N \in \widetilde{C}$ span V.

Proposition 6.13. *Let δ be a vector field in T, let P be a differential operator, which is a product of vector fields in T, let $(j_i)_{i=1}^n$ be a finite sequence of elements of $\{1, \dots, N\}$, and let ω be a compact subset of P. Then, there is a constant $K > 0$ such that*

$$\|D_{\xi_{j_1}} \dots D_{\xi_{j_n}} P(D_\delta H_t.H_t^{-1})\| \leq K \|\xi_{j_1}\|_x \dots \|\xi_{j_n}\|_x,$$
$$\|D_{\xi_{j_1}} \dots D_{\xi_{j_{n-1}}} P(D_{\xi_{j_n}} H_t.H_t^{-1})\| \leq K \|\xi_{j_1}\|_x \dots \|\xi_{j_n}\|_x$$

for all $x \in \omega.A.e.$

Proof. Since T is compact and in view of proposition 6.10, it is enough to prove the second inequality for $P = \mathrm{id}$. In this case, the lemma follows from propositions 6.12 and 6.9, and lemmas 6.4 and 6.5. □

Let $\sigma \subseteq C$ be the simplicial cone

$$\sigma = \sum_{i=1}^{N} \mathbb{R}^+ \xi_i .$$

Let $\{l_i\}$ be the dual basis of $\{\xi_i\}$.

Proposition 6.14. *Let δ be a vector field in T, let P be a differential operator, which is a product of vector fields in T, let $(i_j)_{j=1}^n$ be a finite sequence of elements of $\{1,\dots,N\}$, and let $a \in \overline{C}$. Then, there is a constant $K > 0$ such that*

$$\left| \prod_{j=2}^{r} D_{\xi_{i_j}} P(D_{\xi_{i_1}} H_t . H_t^{-1}(x))_{\alpha,\beta} \right| \leq \frac{K}{\prod_{j=1}^{r} l_{i_j}(x) - l_{i_j}(a)},$$

$$\left| \prod_{j=1}^{r} D_{\xi_{i_j}} P(\delta H_t . H_t^{-1}(x))_{\alpha,\beta} \right| \leq K$$

for all integers $1 \leq \alpha, \beta \leq n$ and $x \in \operatorname{Int}(\sigma + a)$.

Proof. The proof is as in [34], proposition 2.7, but using proposition 6.13 to estimate the higher derivatives. □

End of the proof. Now the proof of theorem 6.3 goes exactly as the proof of [34], theorem 3.1, but using propositions 6.11 and 6.14 to bound the higher derivatives. □

Remark 6.15. Observe that we really have proven that, if $\{e_1,\dots,e_r\}$ is a holomorphic frame of E_1 and $H = (h_{e_i,e_j})$ is the matrix of h in this frame, then the entries of H and $\det H^{-1}$ are of polynomial growth in the local universal cover (which, by theorem 2.13, is equivalent of being log forms) and that the entries of $\partial H \cdot H^{-1}$ are of logarithmic growth in the local universal cover (which, by theorem 2.30, is stronger than being log-log forms).

6.2 Shimura varieties and automorphic vector bundles

A wealth of examples where the theory developed in this paper can be applied is provided by non-compact Shimura varieties. In fact, the concrete examples developed so far are modular curves (see [30]) and Hilbert modular surfaces (see [6]), which are examples of Shimura varieties of non-compact type.
For an algebraic group G, $G(\mathbb{R})^+$ is the identity component of the topological group $G(\mathbb{R})$ and $G(\mathbb{R})_+$ is the inverse image of $G^{\mathrm{ad}}(\mathbb{R})^+$ in $G(\mathbb{R})$; also $G(\mathbb{Q})^+ = G(\mathbb{Q}) \cap G(\mathbb{R}))^+$ and $G(\mathbb{Q})_+ = G(\mathbb{Q}) \cap G(\mathbb{R}))^+$.

Definition of Shimura varieties. Let $\underline{S}$ be the real algebraic torus $\operatorname{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m$. Following Deligne [13] (see also [32]) one considers the data:

(1) G a connected reductive group defined over $\mathbb{Q}$,

(2) X a $G(\mathbb{R})$-conjugacy class of morphisms $h_x : \underline{S} \longrightarrow G_{\mathbb{R}}$ of real algebraic groups ($x \in X$),

satisfying the properties:

(a) The Hodge structure on $\operatorname{Lie} G_{\mathbb{R}}$ defined by $\operatorname{Ad} \circ h_x$ is of type

$$\{(-1,1),(0,0),(1,-1)\}.$$

(b) The involution $\operatorname{int} h_x(i)$ induces a Cartan involution on the adjoint group $G^{\mathrm{ad}}(\mathbb{R})$.

(c) Let $w : \mathbb{G}_{m,\mathbb{R}} \longrightarrow \underline{S}$ be the canonical conorm map. The weight map $h_x \circ w$ (whose image is central by (a)) is defined over $\mathbb{Q}$.

(d) Let Z'_G be the maximal $\mathbb{Q}$-split torus of Z_G, the center of G. Then, $Z_G(\mathbb{R})/Z'_G(\mathbb{R})$ is compact.

Under the above assumptions X is a product of hermitian symmetric domains corresponding to the simple non-compact factors of $G^{\mathrm{ad}}(\mathbb{R})$. Denote by $\mathbb{A}^f$ the finite adèles of $\mathbb{Q}$ and let $K \subseteq G(\mathbb{A}^f)$ be a neat (see, e.g., [35] for the definition of neat) open compact subgroup. With these data the Shimura variety $M_K(\mathbb{C})$ is defined by

$$M_K(\mathbb{C}) = M_K(G,X)(\mathbb{C}) := G(\mathbb{Q})\backslash X \times G(\mathbb{A}^f)/K.$$

Connected components of Shimura varieties. Let X^+ be a connected component of X, and for each $x \in X^+$, let h'_x be the composite of h_x with $G_{\mathbb{R}} \longrightarrow G^{\mathrm{ad}}_{\mathbb{R}}$. Then, $x \longmapsto h'_x$, identifies X^+ with a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class of morphisms $\underline{S} \longrightarrow G^{\mathrm{ad}}_{\mathbb{R}}$ that satisfy the axioms of a connected Shimura variety. In particular, X^+ is a bounded symmetric domain and X is a finite disjoint union of bounded symmetric domains (indexed by $G(\mathbb{R})/G(\mathbb{R})_+$).
Let $\mathcal{C}$ be a set of representatives of the finite set $G(\mathbb{R})_+\backslash G(\mathbb{A}^f)/K$ and, for each $g \in \mathcal{C}$, let Γ_g be the image in $G^{\mathrm{ad}}(\mathbb{R})^+$ of the subgroup $\Gamma'_g = gKg^{-1} \cap G(\mathbb{Q})_+$ of $G(\mathbb{Q})_+$. Then, Γ_g is a torsion free arithmetic subgroup of $G^{\mathrm{ad}}(\mathbb{R})^+$ and $M_K(\mathbb{C})$ is a finite disjoint union

$$M_K(\mathbb{C}) = \coprod_{g \in \mathcal{C}} \Gamma_g \backslash X^+.$$

The connected component $\Gamma_g\backslash X^+$ will be denoted by M_{Γ_g}.

Algebraic models of Shimura varieties. Every Shimura variety is a quasi-projective variety. It has a "minimal" compactification, the Baily-Borel compactification, which is highly singular. The theory of toroidal compactifications provides us with various other compactifications; among them we can choose non-singular ones whose boundaries are normal crossing divisors. Moreover, it has a model over a number field E, called the reflex field, and the toroidal compactifications are also defined over E (see [35]). This model can be extended to a proper regular model defined over $\mathcal{O}_E[N^{-1}]$, where $\mathcal{O}_E$ is the ring of integers of E and N is a suitable natural number.

Automorphic vector bundles. Let K_x be the subgroup of $G(\mathbb{R})$ stabilizing a point $x \in X$ and let P_x be the parabolic subgroup of $G(\mathbb{C})$ arising from the Cartan decomposition of $\mathrm{Lie}(G)$ associated to K_x. Let $\lambda : K_x \longrightarrow \mathrm{GL}_n$ be a finite dimensional representation of K_x. It can be extended trivially to a representation of P_x and defines a $G(\mathbb{C})$-equivariant vector bundle $\check{V}$ on the compact dual $\check{M}(\mathbb{C}) = G(\mathbb{C})/P_x$. Let $\beta : X \longrightarrow \check{M}(\mathbb{C})$ be the Borel embedding, then $V = \beta^*(\check{V})$ is a $G(\mathbb{R})$-equivariant vector bundle on X. For any neat open compact subgroup $K \subseteq G(\mathbb{A}^f)$ it defines a vector bundle

$$V_K = G(\mathbb{Q})\backslash V \times G(\mathbb{A}^f)/K$$

on the Shimura variety M_K. This vector bundle is algebraic and it is defined over the reflex field E. Following [23], the vector bundles obtained in this way, will be called *fully decomposed automorphic vector bundles.*
The restriction to any component M_{Γ_g} will be denoted by V_{Γ_g}. It is a fully decomposed automorphic vector bundle in the sense of the previous section.

Canonical extensions. Let $M_{K,\Sigma}$ be a smooth toroidal compactification of M_K and let V_K be an automorphic vector bundle on M_K. Then, there exists a canonical extension of V_K to a vector bundle $V_{K,\Sigma}$ over $M_{K,\Sigma}$ (see [34], [32], [21]). This canonical extension can be characterized in terms of an invariant hermitian metric on V.

Let M_K be a Shimura variety defined over the reflex field E. Let $M_{K,\Sigma}$ be a smooth toroidal compactification of M_K defined over E such that $D_E = M_{K,\Sigma} \setminus M_K$ is a normal crossing divisor. Let V_K be an automorphic vector bundle defined over E with canonical extension $V_{K,\Sigma}$. Let h be a $G^{\mathrm{der}}(\mathbb{R})$-invariant hermitian metric on V; it induces a hermitian metric on V_K, also denoted by h. We denote again by h the singular hermitian metric induced on $V_{K,\Sigma}$. Let $\mathcal{M}_{K,\Sigma}$ be a regular model of $M_{K,\Sigma}$ over $\mathcal{O}_E[N^{-1}]$. Assume that $V_{K,\Sigma}$ can be extended to a vector bundle $\mathcal{V}_{K,\Sigma}$ over $\mathcal{M}_{K,\Sigma}$. Then, theorem 6.3 implies

Theorem 6.16. *The pair $(\mathcal{V}_{K,\Sigma}, h)$ is a log-singular hermitian vector bundle on $\mathcal{M}_{K,\Sigma}$.* □

References

[1] A. Ash, D. Mumford, M. Rapoport, and Y. Tai, *Smooth compactifications of locally symmetric varieties*, Lie Groups: History Frontiers and Applications, vol. IV, Math. Sci. Press, 1975.

[2] J.M. Bismut, H. Gillet, and C. Soulé, *Analytic torsion and holomorphic determinant bundles I: Bott-Chern forms and analytic torsion*, Comm. Math. Phys. 115 (1988), 49–78.

[3] J.-B. Bost, *Potential theory and Lefschetz theorems for arithmetic surfaces*, Ann. Sci. Éc. Norm. Sup. 32 (1999), 241–312.

[4] J.-B. Bost, H. Gillet, and C. Soulé, *Heights of projective varieties and positive Green forms*, J. Amer. Math. Soc. 7 (1994), 903–1027.

[5] R. Bott and S.S. Chern, *Hermitian vector bundles and the equidistribution of the zeroes of their holomorphic sections*, Acta Math. 114 (1968), 71–112.

[6] J. Bruinier, J.I. Burgos Gil, and U. Kühn, *Borcherds products and arithmetic intersection theory on Hilbert modular surfaces*, Preprint 2003.

[7] J.I. Burgos Gil, *A C^∞-logarithmic Dolbeault complex*, Compositio Math. 92 (1994), 61–86.

[8] ______, *Arithmetic Chow rings and Deligne-Beilinson cohomology*, J. Alg. Geom. 6 (1997), 335–377.

[9] ______, *Hermitian vector bundles and characteristic classes*, The arithmetic and geometry of algebraic cycles (Banff 1998), CRM Proc. Lecture Notes, vol. 24, AMS, 2000, pp. 155–182.

[10] J.I. Burgos Gil, J. Kramer, and U. Kühn, *Cohomological arithmetic Chow rings*, Preprint 2003, math.AG/0404122, to appear in J. Math. Inst. Jussieu.

[11] J.I. Burgos Gil and S. Wang, *Higher Bott-Chern forms and Beilinson's regulator*, Invent. Math. 132 (1998), 261–305.

[12] P. Deligne, *Théorie de Hodge II*, Publ. Math. IHES 40 (1972), 5–57.

[13] ______, *Variétés de Shimura: Interprétation modulaire et techniques de construction de modèles canoniques*, Automorphic forms, representations and L-functions, vol. XXXIII, Proc. Symp. Pure Math., no. part. 2, AMS, 1979, pp. 247–290.

[14] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. 73 (1983), 349–366.

[15] G. Freixas, *Singular heights*, In preparation.

[16] W. Fulton, *Intersection theory*, Ergeb. der Math. und ihrer Grenzgebiete, vol. 3, Springer-Verlag, 1984.

[17] H. Gillet and C. Soulé, *Characteristic classes for algebraic vector bundles with hermitian metric I, II*, Annals of Math. 131 (1990), 163–203, 205–238.

[18] ———, *An arithmetic Riemann-Roch theorem*, Invent. Math. 110 (1992), 473–543.

[19] F. Guillén and V. Navarro Aznar, *Un critère d'extension des foncteurs définis sur les schémas lisses*, Publ. Math. IHES 95 (2002), 1–91.

[20] R.C. Gunning and H. Rossi, *Analytic functions of several complex variables*, Prentice-Hall, 1965.

[21] M. Harris, *Functorial properties of toroidal compactifications of locally symmetric varieties*, Proc. Lon. Math. Soc. 59 (1989), 1–22.

[22] M. Harris and D.H. Phong, *Cohomologie de Dolbeault à croissance logarithmique à l'infini*, C. R. Acad. Sci. Paris 302 (1986), no. 8, 307–310.

[23] M. Harris and S. Zucker, *Boundary cohomology of Shimura varieties I. coherent cohomology on toroidal compactifications*, Ann. Scient. Éc. Norm. Sup. 27 (1994), 249–344.

[24] ———, *Boundary cohomology of Shimura varieties III. coherent cohomology on higher-rank boundary strata and applications to Hodge theory*, Mém. Soc. Math. Fr. (N.S.), vol. 85, Soc. Math. Fr., 2001.

[25] H. Hironaka, *Resolution of singularities of an algebraic variety over a field of characteristic zero*, Annals of Math. 79 (1964), 109–326.

[26] K. Köhler, *A Hirzebruch proportionality principle in Arakelov geometry*, Preprint, 2001.

[27] S.S. Kudla, *Special cycles and derivatives of Eisenstein series*, to appear in MSRI Proceedings on Rankin L-values, arXiv.org/math.NT/0308295.

[28] ———, *Derivatives of Eisenstein series and generating functions for arithmetic cycles*, Séminaire Bourbaki 1999/2000, Exposés 865-879, Astérisque, vol. 276, pp. 341–368, SMF, 2002.

[29] ———, *Eisenstein series and arithmetic geometry*, vol. II, pp. 173–183, Proceedings of the ICM, Beijing, 2002.

[30] U. Kühn, *Generalized arithmetic intersection numbers*, J. Reine Angew. Math. 534 (2001), 209–236.

[31] V. Maillot and D. Roessler, *Conjectures sur les dérivées logarithmiques des fonctions L d'Artin aux entiers négatifs*, Math. Res. Lett. 9 (2002), 715–724.

[32] J.S. Milne, *Canonical models of (mixed) Shimura varieties and automorphic vector bundles*, Automorphic forms, Shimura varieties, and L-functions,, Perspect. in Math., vol. 10, Academic Press, 1990, Ann Arbor, MI, 1988, pp. 283–414.

[33] A. Moriwaki, *Intersection pairing for arithmetic cycles with degenerate Green currents*, E-print AG/9803054.

[34] D. Mumford, *Hirzebruch's proportionality theorem in the non-compact case*, Invent. Math. 42 (1977), 239–272.

[35] R. Pink, *Arithmetical compactifications of mixed Shimura varieties*, Ph.D. thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 1990.

[36] C. Soulé, D. Abramovich, J.-F. Burnol, and J. Kramer, *Lectures on Arakelov Geometry*, Cambridge Studies in Advanced Math., vol. 33, Cambridge University Press, 1992.