Brascamp-Lieb Inequalities for Non-Commutative Integration

Eric A. Carlen[1] and Elliott H. Lieb[2]

Received: August 6, 2008
Revised: September 22, 2008

Communicated by Heinz Siedentop

Abstract. We formulate a non-commutative analog of the Brascamp-Lieb inequality, and prove it in several concrete settings.

2000 Mathematics Subject Classification: 47C15, 15A45, 26D15
Keywords and Phrases: inequalities, traces, non-commutative integration

1 Introduction

1.1 Young's inequality in the context of ordinary Lebesgue integration

In this paper, we shall extend the class of generalized Young's inequalities known as Brascamp-Lieb inequalities (B-L inequalities) to an operator algebra setting entailing non-commutative integration.
The original Young's inequality [40] states that for non negative measurable functions f_1, f_2 and f_3 on $\mathbb{R}$, and $1 \leq p_1, p_2, p_3 \leq \infty$, with $1/p_1 + 1/p_2 + 1/p_3 = 2$,

$$\int_{\mathbb{R}^2} f_1(x) f_2(x-y) f_3(y) \mathrm{d}x \mathrm{d}y \leq \left(\int_{\mathbb{R}} f_1^{p_1}(t)\mathrm{d}t\right)^{1/p_1} \left(\int_{\mathbb{R}} f_2^{p_2}(t)\mathrm{d}t\right)^{1/p_2} \left(\int_{\mathbb{R}} f_3^{p_3}(t)\mathrm{d}t\right)^{1/p_3} . \tag{1.1}$$

[1]Work partially supported by U.S. National Science Foundation grant DMS 06-00037.
[2]Work partially supported by U.S. National Science Foundation grant PHY 06-52854.

Thus, it provides an estimate of the integral of a product of functions in terms of a product of L^p norms of these functions. The crucial difference with a Hölder type inequality is that the integrals on the right are integrals over only $\mathbb{R}$, while the integral on the left is an integral over $\mathbb{R}^2$, and none of the three factors in the product on the left – $f(x)$, $g(x-y)$ or $h(y)$ – are integrable (to any power) on $\mathbb{R}^2$.
To frame the inequality in terms that are more amenable to the generalizations considered here, define the maps $\phi_j : \mathbb{R}^2 \to \mathbb{R}$, $j = 1, 2, 3$, by

$$\phi_1(x,y) = x \qquad \phi_2(x,y) = x - y \qquad \text{and} \qquad \phi_3(x,y) = y \ .$$

Then (1.1) can be rewritten as

$$\int_{\mathbb{R}^2} \left(\prod_{j=1}^{3} f_j \circ \phi_j \right) \mathrm{d}^2 x \le \prod_{j=1}^{3} \left(\int_{\mathbb{R}} f_j^{p_j}(t) \mathrm{d}t \right)^{1/p_j} \ . \tag{1.2}$$

There is now no particular reason to limit ourselves to products of only three functions, or to integrals over $\mathbb{R}^2$ and $\mathbb{R}$, or even any Euclidean space for that matter:

1.1 DEFINITION. Given measure spaces $(\Omega, \mathcal{S}, \mu)$ and $(M_j, \mathcal{M}_j, \nu_j)$, $j = 1, \dots, N$, not necessarily distinct, together with measurable functions $\phi_j : \Omega \to M_j$ and numbers $p_1, \dots, p_N$ with $1 \le p_j \le \infty$, $1 \le j \le N$, we say that *a B-L inequality holds for* $\{\phi_1, \dots, \phi_N\}$ *and* $\{p_1, \dots, p_N\}$ in case there is a finite constant C such that

$$\int_\Omega \prod_{j=1}^{N} f_j \circ \phi_j \mathrm{d}\mu \le C \prod_{j=1}^{N} \|f_j\|_{L^{p_j}(\nu_j)} \tag{1.3}$$

holds whenever each f_j is non-negative and measurable on M_j, $j = 1, \dots, N$.
There are by now many examples. One of the oldest is the original *discrete Young's inequality.* In the current notation, this concerns the case in which $\Omega = \mathbb{Z}^2$ equipped with counting measure, $N = 3$, and each M_j is $\mathbb{Z}$, equipped with counting measure. Then with

$$\phi_1(m,n) = m \qquad \phi_2(m,n) = m - n \qquad \text{and} \qquad \phi_3(m,n) = n \ ,$$

(1.2) holds for any three non-negative functions $f_j : \mathbb{Z} \to \mathbb{R}_+$ under the same conditions on the p_j as in the continuous case; i.e., $1/p_1 + 1/p_2 + 1/p_3 = 2$. There is a significant difference: In the discrete case, the constant $C = 1$ is sharp, and there is equality if and only if one of the f_j is identically zero, or else f_1 vanishes except at some m_0, f_3 vanishes except at some n_0, and f_2 vanishes except at $m_0 - n_0$. This discrete inequality is also due to Young [40], while the statement about cases of equality is proved in [20], where the authors also consider extensions to more than three functions.

In the continuous case, a much wider generalization to more than three functions was made by B-L in [5], where the sharp constant in Young's inequality – which is strictly less than 1 unless $p_1 = p_2 = 1$ – was obtained, with a proof that the only non-negative functions yielding equality are certain Gaussian functions. (This best constant was also obtained at the same time by Beckner [4], for three functions.)
These inequalities generalize from $\mathbb{R}$ to $\mathbb{R}^n$. The complete generalization to the case in which the M_j are all Euclidean spaces, but of different dimension, and the ϕ_j are linear transformations from $\mathbb{R}^n$ to M_j, was proved by Lieb [23]. Again, the maximizers are Gaussians. Another proof of this generalized version, together with a reverse form, was obtained by Barthe [1], who also provided a detailed analysis of the cases of equality in the original B-L inequality from [5]. The cases of equality in the higher dimensional generalization from [23] were analyzed in detail in [7, 8].
Examples in which Ω is the sphere S^{N-1} or the permutation group $\mathcal{S}^N$ were proved in [13, 14], and the above definition of B-L inequalities in arbitrary measure spaces is taken from [10], where a duality between B-L inequalities and subadditivity of entropy inequalities is proved.

1.2 A generalized Young's inequality in the context of non-commutative integration

In non-commutative integration theory, as developed by Irving Segal [31, 32, 33], the basic framework is a triple $(\mathcal{H}, \mathfrak{A}, \lambda)$ where $\mathcal{H}$ is a Hilbert space, $\mathfrak{A}$ is a W^* algebra (a von Neumann algebra) of operators on $\mathcal{H}$, and λ is a positive linear functional on the finite rank operators in $\mathfrak{A}$. In Segal's picture, the algebra $\mathfrak{A}$ corresponds to the algebra of bounded measurable functions, and applying the positive linear functional λ to a positive operator corresponds to taking the integral of a positive function. That is,

$$A \mapsto \lambda(A) \qquad \text{corresponds to} \qquad f \mapsto \int_M f \mathrm{d}\nu \ .$$

Such a triple $(\mathcal{H}, \mathfrak{A}, \lambda)$ is called a *non-commutative integration space*. Certain natural regularity properties must be imposed on λ if one is to get a well-behaved non-commutative integration theory, but we shall not go into these here as the examples that we consider are all based on the case in which λ is the *trace* on operators on $\mathcal{H}$, or some closely related functional, for which discussion of these extra conditions would be a digression.
In this operator algebra setting there are natural non-commutative analogs of the usual L^p spaces: If A is a finite rank operator in $\mathfrak{A}$, and $1 \leq q < \infty$, define

$$\|A\|_{q,\lambda} = \left(\lambda(A^*A)^{q/2}\right)^{1/q} \ .$$

This defines a norm (under appropriate conditions on λ that are obvious for the trace), and the completion of the space of finite rank operators in $\mathfrak{A}$ under

this norm defines a non-commutative L^p space. (The completion may contain unbounded operators "affiliated" to $\mathfrak{A}$.) For more on the general theory of non-commutative integration, see the early papers [15, 31, 33, 34] and the more recent work in [16, 19, 21, 25].
To frame an analog of (1.3) in an operator algebra setting, we replace the measure spaces by non-commutative integration spaces:

$$(\Omega, \mathcal{S}, \mu) \longrightarrow (\mathcal{H}, \mathfrak{A}, \lambda) \quad \text{and} \quad (M_j, \mathcal{M}_j, \nu_j) \longrightarrow (\mathcal{H}_j, \mathfrak{A}_j, \lambda_j) \quad j = 1, \dots, N \ .$$

The right hand side of (1.3) has an obvious generalization to the operator algebra setting in terms of the non-commutative L_p norms introduced above. As for the left hand side of (1.3), regard $f_j \mapsto f_j \circ \phi_j$ as a W^* algebra homomorphism (which, restricted to the W^* algebra $L^\infty(M_j)$, it is), and suppose we are given W^* homomorphisms

$$\phi_j : \mathfrak{A}_j \to \mathfrak{A} \ .$$

Then each $\phi_j(A_j)$ belongs to $\mathfrak{A}$. However in the non-commutative case, the product of the $\phi_j(A_j)$ depends on their order in the product, and need not be self-adjoint even – let alone positive – even if each of the A_j are positive.
Therefore, let us return to the left side of (1.3) and suppose that each f_j is strictly positive. Then defining

$$h_j = \ln(f_j) \qquad \text{so that} \qquad f_j \circ \phi_j = e^{h \circ \phi_j} \ ,$$

we can then rewrite (1.3) as

$$\int_\Omega \exp\left(\sum_{j=1}^N h_j \circ \phi_j\right) \mathrm{d}\mu \le C \prod_{j=1}^N \|e^{h_j}\|_{L^{p_j}(\nu_j)} \ , \tag{1.4}$$

We can now formulate our operator algebra analog of (1.3):

1.2 DEFINITION. Given non-commutative integration spaces $(\mathcal{H}, \mathfrak{A}, \lambda)$ and $(\mathcal{H}_j, \mathfrak{A}_j, \lambda_j)$, $j = 1, \dots, N$, together with W^* algebra homomorphisms $\phi_j : \mathfrak{A}_j \to \mathfrak{A}$, $j = 1, \dots, N$, and indices $1 \le p_j \le \infty$, $j = 1, \dots, N$, a *non-commutative B-L inequality holds for* $\{\phi_1, \dots, \phi_N\}$ *and* $\{p_1, \dots, p_N\}$ if there is a finite constant C so that

$$\boxed{\lambda\left(\exp\left[\sum_{j=1}^N \phi_j(H_j)\right]\right) \le C \prod_{j=1}^N \left(\lambda_j \exp\left[p_j H_j\right]\right)^{1/p_j}} \tag{1.5}$$

whenever H_j is self-adjoint in $\mathfrak{A}_j$, $j = 1, \dots, N$.

In this paper, we are concerned with determining the indices and the best constant C for which such an inequality holds, and shall focus on two examples: The first concerns *operators on tensor products of Hilbert spaces*, and the second concerns *Clifford algebras*.

1.3 A generalized Young's inequality for tensor products

1.3 EXAMPLE. Let $\mathcal{H}_j$, $j = 1, \ldots, n$ be separable Hilbert spaces, and let Let $\mathcal{K}$ denote the tensor product

$$\mathcal{K} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n \ .$$

Define $\mathfrak{A}$ to be $\mathfrak{B}(\mathcal{K})$, the algebra of bounded operators on $\mathcal{K}$, and define λ to be Tr, the trace Tr on $\mathcal{K}$, so that $(\mathcal{H}, \mathfrak{A}, \lambda) = (\mathcal{K}, \mathfrak{B}(\mathcal{K}), \mathrm{Tr})$.
For any non-empty subset J of $\{1, \ldots, n\}$, let $\mathcal{K}_J$ denote the tensor product

$$\mathcal{K}_J = \bigotimes_{j \in J} \mathcal{H}_j \ .$$

Define $\mathfrak{A}_J$ to be $\mathfrak{B}(\mathcal{K}_J)$, the algebra of bounded operators on $\mathcal{K}_J$, and define λ_J to be Tr_J, the trace on $\mathcal{K}_J$, so that $(\mathcal{H}_J, \mathfrak{A}_J, \lambda_J) = (\mathcal{K}_J, \mathfrak{B}(\mathcal{K}_J), \mathrm{Tr}_J)$.
There are natural homomorphisms ϕ_J embedding the $2^n - 1$ algebras $\mathfrak{A}_J$ into $\mathfrak{A}$. For instance, if $J = \{1, 2\}$,

$$\phi_{\{1,2\}}(A_1 \otimes A_2) = A_1 \otimes A_2 \otimes I_{\mathcal{H}_3} \otimes \cdots \otimes I_{\mathcal{H}_N} \ , \tag{1.6}$$

and $\phi_{\{1,2\}}$ is extended linearly.
It is obvious that in case $J \cap K = \emptyset$ and $J \cup K = \{1, \ldots, n\}$, then for all $H_J \in \mathfrak{A}_J$ and $H_K \in \mathfrak{A}_K$,

$$\mathrm{Tr}\left(e^{H_J + H_K}\right) = \mathrm{Tr}_J\left(e^{H_J}\right) \mathrm{Tr}_K\left(e^{H_K}\right) \ , \tag{1.7}$$

but things are more interesting when $J \cap K \neq \emptyset$ and J and K are both proper subsets of $\{1, \ldots, n\}$. If H_J and H_K do not commute, which is the typical situation for $J \cap K \neq \emptyset$, one can estimate the left hand side of (1.7) by first applying the Golden–Thompson inequality [17, 36], which says that for self-adjoint operators H_J and H_K,

$$\mathrm{Tr}\left(e^{H_J + H_K}\right) \leq \mathrm{Tr}\left(e^{H_J} e^{H_K}\right) \ .$$

One might then apply Hölder's inequality – but if J and K are proper subsets of $\{1, \ldots, n\}$, this will yield a finite bound if and only if all of the Hilbert spaces whose indices are not included in *both* J and K are finite dimensional. Even then, the bound depends on the dimension in an unpleasant way. The non-commutative B-L Inequalities provided by the next theorem do not have this defect.

1.4 THEOREM. *Let $J_1, \ldots, J_N$ be N non-empty subsets of $\{1, \ldots, n\}$ For each $i \in \{1, \ldots, n\}$, let $p(i)$ denote the number of the sets $J_1, \ldots, J_N$ that contain i, and let p denote the minimum of the $p(i)$. Then, for self-adjoint operators H_j on $\mathcal{K}_{J_j}$, $j = 1, \ldots, N$,*

$$\boxed{\mathrm{Tr}\left(\exp\left[\sum_{j=1}^{N} \phi_{J_j}(H_j)\right]\right) \leq \prod_{j=1}^{N} \left(\mathrm{Tr}_{J_j} \ e^{qH_j}\right)^{1/q}} \tag{1.8}$$

for $q = p$ (and hence all $1 \le q \le p$), while for all $q > p$, it is possible for the left hand side to be infinite, while the right hand side is finite.

Note that in Theorem 1.4, the constant C in Definition (1.2) is 1. The fact that the constant $C = 1$ is best possible, and that the inequality cannot hold for $q > p = \min\{p(1), \dots, p(N)\}$ is easy to see by considering the case that each $\mathcal{H}_j$ has finite dimension d_j, and $H_j = 0$ for each j. Then

$$\mathrm{Tr}\left(\exp\left[\sum_{j=1}^{N} \phi_{J_j}(H_j)\right]\right) = \prod_{j=1}^{n} d_j$$

and

$$\prod_{j=1}^{N} \left(\mathrm{Tr}_{J_j} e^{qH_j}\right)^{1/q} = \prod_{j=1}^{N} \prod_{k \in J_j} d_k^{1/q} = \prod_{j=1}^{n} d_j^{p(j)/q} \ .$$

We will prove inequality (1.8) for $q = p$ in Section 3.

As an example, consider the case in which $n = 6$, $N = 3$ and

$$J_1 = \{1,2,3\} \quad J_2 = \{3,4,5\} \qquad \text{and} \quad J_3 = \{5,6,1\} \ .$$

Here, $p = 1$, and hence

$$\mathrm{Tr}\left(\exp\left[\sum_{j=1}^{3} \phi_{J_j}(H_j)\right]\right) \le \prod_{j=1}^{3} \left(\mathrm{Tr}_{J_j}\ e^{H_j}\right) \ . \tag{1.9}$$

Theorem 1.4 says that the inequality (1.9) can be extended to larger tensor products, and this obviously has an statistical mechanical interpretation as a bound on the *partition function* of a collection of interacting spins in terms of a product of partition functions of simple constituent sub-systems. For more background on this, see Thirring's book [35].

To estimate the left side of (1.9) without using Theorem 1.4, one might use the Golden-Thompson inequality and then Schwarz's inequality to write

$$\begin{aligned}\mathrm{Tr}\left(\exp\left[\sum_{j=1}^{3} \phi_{J_j}(H_j)\right]\right) &\le \mathrm{Tr}\left(e^{\phi_1(H_1)+\phi_3(H_3)} e^{\phi_2(H_2)}\right)\\ &\le \left(\mathrm{Tr}\ e^{2[\phi_1(H_1)+\phi_3(H_3)]}\right)^{1/2} \left(\mathrm{Tr}\ e^{2\phi_2(H_2)}\right)^{1/2} \ .\end{aligned}$$

While the L^2 norms are an improvement over the L^1 norms in (1.9), the traces are now over the entire tensor product space. Thus, for example,

$$\left(\mathrm{Tr}\ e^{2\phi_2(H_2)}\right)^{1/2} = (d_1 d_2 d_6)^{1/2} \left(\mathrm{Tr}_{J_2}\ e^{2H_2}\right)^{1/2}$$

where d_j is the dimension of Hilbert space $\mathcal{H}_j$. This dimension dependence may be unfavorable if any of the dimensions is large.

1.4 A generalized Young's inequality in Clifford algebras

Our next example concerns Clifford algebras, which, as Segal emphasized [32], allow one to represent fermion Fock space as an L^2 space – albeit a non-commutative L^2 space, but still with many of the advantages of having a Hilbert space represented as a function space, as in the usual Schrödinger representation in quantum mechanics.

In the finite dimensional setting, with n degrees of freedom, one starts with n operators $Q_1, \dots, Q_n$ on some Hilbert space $\mathcal{H}$ that satisfy the *canonical anticommutation relations*

$$Q_i Q_j + Q_j Q_i = 2\delta_{i,j} I \ .$$

One can concretely construct such operators acting on $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, the n–fold tensor product of $\mathbb{C}^2$ with itself; see [6]. The Clifford algebra $\mathfrak{C}$ is the operator algebra on $\mathcal{H}$ that is generated by $Q_1, \dots, Q_n$.

The Clifford algebra $\mathfrak{C}$ itself is 2^n dimensional. In fact, let $\alpha = (\alpha_1, \dots, \alpha_n)$ be a *fermionic multi-index*, which means that each α_j is either 0 or 1. Then define

$$Q^\alpha = Q_1^{\alpha_1} Q_2^{\alpha_2} \cdots Q_n^{\alpha_n} \ . \tag{1.10}$$

it is easy to see that the 2^n operators Q^α are a basis for the Clifford algebra, so that any operator A in $\mathfrak{C}$ has a unique expression

$$A = \sum_\alpha x_\alpha Q^\alpha \ .$$

The linear functional τ on $\mathfrak{C}$ is defined by

$$\tau\left(\sum_\alpha x_\alpha Q^\alpha\right) = x_{(0,\dots,0)} \ . \tag{1.11}$$

That is, τ acting on A picks out the coefficient of the identity in $A = \sum_\alpha x_\alpha Q^\alpha$. It turns out that when the Clifford algebra is constructed in the way described here, as an algebra operators on the 2^n dimensional space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, τ is nothing other than the normalized trace:

$$\tau(A) = \frac{1}{2^n}\mathrm{Tr}_{\mathcal{H}}(A) \ .$$

Hence τ is a positive linear functional, and $((\mathbb{C}^2)^{\otimes n}, \mathfrak{C}, \tau)$ is a non-commutative integration space in the sense of Segal.

Clifford algebras have infinitely many subalgebras that are also Clifford algebras of lower dimension. This is in contrast to the setting described in Example 1.3, in which the only natural subalgebras are the $2^n - 1$ subalgebras corresponding to the $2^n - 1$ non-empty subsets of the index set $\{1, \dots, n\}$.

To describe these subalgebras, let $\mathcal{J}$ be the *canonical injection* of $\mathbb{R}^n$ into $\mathfrak{C}$, which is given by

$$\mathcal{J}((x_1, \ldots, x_n)) = \sum_{j=1}^{n} x_j Q_j \ . \tag{1.12}$$

If x and y are any two vectors in $\mathbb{R}^n$, it is easy to see from the canonical anticommutation relations that

$$(\mathcal{J}(x))(\mathcal{J}(y)) = 2(x \cdot y)I \ .$$

Hence if V is *any* m dimensional subspace of $\mathbb{R}^n$, and $\{u_1, \ldots, u_m\}$ is *any* orthonormal basis for V, the m operators

$$\mathcal{J}(u_1), \ldots, \mathcal{J}(u_m)$$

again satisfy the canonical anticommutation relations, and generate a subalgebra of $\mathfrak{C}$ that is denoted by $\mathfrak{C}(V)$, and referred to as *the Clifford algebra over* V. In the same vein, it is convenient to refer to $\mathfrak{C}$ itself as the *Clifford algebra over* $\mathbb{R}^n$. Obviously, $\mathfrak{C}(V)$ is naturally isomorphic to $\mathfrak{C}(\mathbb{R}^m)$, and for $A \in \mathfrak{C}(V)$ one may compute $\tau(A)$ using either the normalized trace τ inherited from $\mathfrak{C}$, or the normalized trace τ_V induced by the identification of $\mathfrak{C}(V)$ with $\mathfrak{C}(\mathbb{R}^m)$. As Segal emphasized, $((\mathbb{C}^2)^{\otimes n}, \mathfrak{C}, \tau)$ is, in many ways, a non-commutative analog of the Gaussian measure space $(\mathbb{R}^n, \gamma(x)\mathrm{d}x)$ where

$$\gamma(x) = \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2} \ . \tag{1.13}$$

In fact, just as orthogonality implies independence in $(\mathbb{R}^n, \gamma(x)\mathrm{d}x)$, if V and W are two orthogonal subspaces of $\mathbb{R}^n$, and if $A \in \mathfrak{C}(V)$ and $B \in \mathfrak{C}(W)$, then

$$\tau(AB) = \tau(A)\tau(B) \ .$$

The results we prove here reinforce this analogy. We are now ready to introduce our next example:

1.5 EXAMPLE. For some $n > 1$, let $\mathfrak{A}$ be the Clifford algebra over $\mathbb{R}^n$ with its usual inner product, and let $\mathfrak{A}$ be equipped with its unique tracial state τ, which is simply the normalized trace.
For each $j = 1, \ldots, N$, let V_j be a subspace of $\mathbb{R}^n$, and let $\mathfrak{A}_j$ be $\mathfrak{C}(V_j)$, the Clifford algebra over V_j with the inner product that V_j inherits from $\mathbb{R}^n$. Let $\mathfrak{A}_j$ be equipped with its unique tracial state τ_j. The natural embedding of V_j into $\mathbb{R}^n$ induces a homomorphism of $\mathfrak{A}_j$ into $\mathfrak{A}$, and we define this to be ϕ_j. In this setting, we shall prove

1.6 THEOREM. *Let $V_1, \ldots, V_N$ be N subspaces of $\mathbb{R}^n$, and let $\mathfrak{A}_j$ be the Clifford algebra over V_j with the inner product that V_j inherits from $\mathbb{R}^n$, and let $\mathfrak{A}_j$*

be equipped with its unique tracial state τ_j. Let ϕ_j be the natural homomorphism of $\mathfrak{A}_j$ into $\mathfrak{A}$ induced by the natural embedding of V_j into $\mathbb{R}^n$. Then

$$\boxed{\tau\left(\exp\left[\sum_{j=1}^N \phi_j(H_j)\right]\right) \le \prod_{j=1}^N \left(\tau_j\ e^{p_j H_j}\right)^{1/p_j}} \tag{1.14}$$

for all self-adjoint operators $H_j \in \mathfrak{A}_j$ if and only if

$$\sum_{j=1}^N \frac{1}{p_j} P_j \le I_{\mathbb{R}^n}\ . \tag{1.15}$$

where P_j is the orthogonal projection onto V_j in $\mathbb{R}^n$.

In the special case in which $\dim(V_j) = 1$ for each j, (1.14) reduces to an interesting inequality for the hyperbolic cosine. Indeed, let u_j be one of the two unit vectors in V_j.
Then, with $u_j \otimes u_j$ denoting the orthogonal projection onto the span of u_j, (1.15) reduces to

$$\sum_{j=1}^N \frac{1}{p_j} u_j \otimes u_j \le I_{\mathbb{R}^n}\ . \tag{1.16}$$

The greater simplification, however, is that in this case, the space of self-adjoint operators in each $\mathfrak{A}_j$ is just two dimensional, and, with $\mathcal{J}$ denoting the canonical injection defined in (1.12),

$$H_j = a_j I + b_j \mathcal{J}(u_j)$$

for some real numbers a_j and b_j. Then

$$\sum_{j=1}^N H_j = \left(\sum_{j=1}^N a_j\right) I + \mathcal{J}\left(\sum_{j=1}^N b_j u_j\right)\ .$$

This operator has exactly two eigenvalues,

$$\left(\sum_{j=1}^N a_j\right) \pm \left|\sum_{j=1}^N b_j u_j\right|$$

with equal multiplicities.
Likewise, $p_j H_j$ has exactly two eigenvalues $p_j a_j \pm p_j b_j$ with equal multiplicities.
Hence, in this case, (1.14) reduces to

$$\cosh\left(\left|\sum_{j=1}^N b_j u_j\right|\right) \le \prod_{j=1}^N \left(\cosh(p_j b_j)\right)^{1/p_j} \quad \text{for all} \quad (b_1, \dots, b_N) \in \mathbb{R}^N\ , \tag{1.17}$$

which, according to Theorem 1.6, must hold whenever (1.16) is satisfied. (The a_j's make the same contribution to both sides, and may be cancelled away.) Taking the logarithm of both sides, this can be rewritten as

$$\ln\cosh\left(\left|\sum_{j=1}^N b_j u_j\right|\right) \le \sum_{j=1}^N \frac{1}{p_j}\ln\cosh(p_j b_j) \quad \text{for all} \quad (b_1,\dots,b_N)\in\mathbb{R}^N\ , \tag{1.18}$$

and this inequality must hold whenever the unit vectors $\{u_1,\dots,u_N\}$ and the positive numbers $\{p_1,\dots,p_N\}$ satisfy (1.16).

Later on, we shall give an elementary proof of this inequality, and hence an elementary proof of Theorem 1.6 when each V_j is one dimensional. Our proof of the other cases is less than elementary, and even our elementary proof of (1.18) is less than direct.

2 Subadditivty of Entropy and Generalized Young's Inequalities

In the examples we have introduced in the previous section, the positive linear functionals λ under consideration are either traces or normalized traces. Throughout this section, we assume that our non-commutative integration spaces $(\mathcal{H},\mathfrak{A},\lambda)$ are based on *tracial* positive linear functionals λ. That is, we require that for all $A,B\in\mathfrak{A}$,

$$\lambda(AB) = \lambda(BA)\ .$$

In such a non-commutative integration space $(\mathcal{H},\mathfrak{A},\lambda)$, a *probability density* is a non-negative element ρ of $\mathfrak{A}$ such that $\lambda(\rho)=1$. Indeed, the tracial property of λ ensures that

$$\lambda(\rho A) = \lambda(A\rho) = \lambda(\rho^{1/2}A\rho^{1/2})$$

so that $A\mapsto\lambda(\rho A)$ is a positive linear functional that equals 1 on the identity.

Now suppose we have N non-commutative integration spaces $(\mathcal{H}_j,\mathfrak{A}_j,\lambda_j)$ and W^* homomorphisms $\phi_j:\mathfrak{A}_j\to\mathfrak{A}$. Then these homomorphisms induce maps from the space of probability densities on $\mathfrak{A}$ to the spaces of probability densities on the $\mathfrak{A}_j$:

For any probability density ρ on $(\mathfrak{A},\lambda)$, let ρ_j be the probability density on $(\mathfrak{A}_j,\lambda_j)$ defined by

$$\lambda_j(\rho_j A) = \lambda(\rho\phi_j(A))$$

for all $A\in\mathfrak{A}_j$.

For example, in the setting of Example 1.3, ρ_{J_j} is just the partial trace of ρ over $\bigotimes_{k\in J_j^c}\mathcal{H}_k$, leaving an operator on $\bigotimes_{k\in J_j}\mathcal{H}_k$. In the Clifford algebra setting of Example 1.5, ρ_j is simply the orthogonal projection of ρ in $L^2(\mathfrak{C},\tau)$ onto $\mathfrak{C}(V_j)$, which is also known as the conditional expectation [38] of ρ given $\mathfrak{C}(V_j)$.

In this section, we are concerned with the relations between the *entropies* of ρ and the $\rho_1,\dots,\rho_N$. The entropy of a probability density ρ, $S(\rho)$, is defined by

$$S(\rho) = -\lambda(\rho\ln\rho)\ .$$

Evidently, the entropy functional is concave on the set of probability densities.

2.1 DEFINITION. Given tracial non-commutative integration spaces $(\mathcal{H}, \mathfrak{A}, \lambda)$ and $(\mathcal{H}_j, \mathfrak{A}_j, \lambda_j)$, $j = 1, \ldots, N$, together with W^* algebra homomorphisms $\phi_j : \mathfrak{A}_j \to \mathfrak{A}$, $j = 1, \ldots, N$, and numbers $1 \leq p_j \leq \infty$, $j = 1, \ldots, N$, a *generalized subadditivity of entropy inequality* holds if there is a finite constant C so that

$$\boxed{\sum_{j=1}^{N} \frac{1}{p_j} S(\rho_j) \geq S(\rho) - \ln C} \tag{2.1}$$

for all probability densities ρ in $\mathfrak{A}$.

It turns out that for tracial non-commutative integration spaces, generalized subadditivity of entropy inequalities and B-L inequalities are dual to one another, just as they are in the commutative case [10], so that if one holds, so does the other, with the same values of $p_1, \ldots, p_N$ and C. The following is, in fact, a direct non-commutative analog of the main theorem of [10].

2.2 THEOREM. *Let $(\mathcal{H}, \mathfrak{A}, \lambda)$ and $(\mathcal{H}_j \mathfrak{A}_j, \lambda_j)$, $j = 1, \ldots, N$, be tracial non-commutative integration spaces. Let $\phi_j : \mathfrak{A}_j \to \mathfrak{A}$, $j = 1, \ldots, N$ be W^* algebra homomorphisms. Then for any numbers $1 \leq p_j \leq \infty$, $j = 1, \ldots, N$, and any finite constant C, the generalized subadditivity of entropy inequality (2.1) is true for all probability densities ρ on $\mathfrak{A}$ if and only if the non-commutative B-L inequality (1.5) is true for all self-adjoint $H_j \in \mathfrak{A}_j$, $j = 1, \ldots, N$, with the same $p_1, \ldots, p_N$ and the same C.*

As a consequence of Theorem 2.2, one strategy for proving a non-commutative B-L inequality is to prove the corresponding generalized subadditivity of entropy inequality. We shall see in our examples that this is an effective strategy; indeed, this is how we prove Theorems 1.4 and 1.6.

In the current tracial context, the proof of Theorem 2.2 is a direct adaptation of the proof of the corresponding result in the context of Lebesgue integration given in [10]. It turns on a well–known formula for the Legendre transform of the entropy. For completeness, we give this formula in Lemma 2.3 below. Except for the $S(A) = -\infty$ case in (2.2) below, it is the same as [26, Prop. 1.10]. The Legendre transform is usually defined for functions on a dual pair of real linear spaces. Therefore, it is convenient to extend the definition of S to all of $\mathfrak{A}_{\mathrm{h},1}$, the (real) subspace of self-adjoint elements A of $\mathfrak{A}$ with $\lambda(|A|) < \infty$. as follows:

$$S(A) = \begin{cases} -\lambda(A \ln A) & \text{if } A \geq 0 \text{ and } \lambda(A) = 1, \\ -\infty & \text{otherwise.} \end{cases} \tag{2.2}$$

2.3 LEMMA. *Let $\mathfrak{A}$ be $\mathfrak{B}(\mathcal{H})$, the algebra of bounded operators on a separable Hilbert space $\mathcal{H}$. Let λ denote either the trace* Tr *on $\mathcal{H}$, or, if $\mathcal{H}$ is finite* *dimensional, the normalized trace τ. Then for all $A \in \mathfrak{A}_{\mathrm{h},1}$,*

$$-S(A) = \sup_{H \text{ self adjoint, semibounded above}} \left\{\lambda(AH) - \ln\left(\lambda\left(e^H\right)\right)\right\} \ . \tag{2.3}$$

The supremum is an attained maximum if and only if A is a strictly positive probability density, in which case it is attained at H if and only if $H = \ln A + cI$ for some $c \in \mathbb{R}$. Consequently, for all self adjoint H with H semibounded above,

$$\ln\left(\lambda\left(e^H\right)\right) = \sup_{A \in \mathfrak{A}_{\mathrm{h},1}} \left\{\lambda(AH) + S(A)\right\} \ . \tag{2.4}$$

The supremum is a maximum at all points of the domain of $\ln\left(\lambda\left(e^H\right)\right)$, in which case it is attained only at the single point $A = e^H/(\lambda(e^H))$.

Proof: We consider first the case that $\lambda = \mathrm{Tr}$, and $\mathcal{H}$ has finite dimension d. To see that the supremum is ∞ unless $0 \le A \le I$, let c be any real number, and let u be any unit vector. Then let H be c times the orthogonal projection onto u. For this choice of H,

$$\lambda(AH) - \ln\left(\lambda\left(e^H\right)\right) = c\langle u, Au\rangle - \ln(e^c + (d-1)) \ .$$

If $\langle u, Au\rangle < 0$, this tends to ∞ as c tends to $-\infty$. If $\langle u, Au\rangle > 1$, this tends to ∞ as c tends to ∞. Hence we need only consider $0 \le A \le I$. Next, taking $H = cI$, $c \in \mathbb{R}$,

$$\lambda(AH) - \ln\left(\lambda\left(e^H\right)\right) = c\lambda(A) - c - \ln(d) \ .$$

Unless $\lambda(A) = 1$, this tends to ∞ as c tends to ∞. Hence we need only consider the case that A is a density matrix ρ.
Let ρ be any density matrix on $\mathcal{H}$ and let H be any self-adjoint operator such that $\mathrm{Tr}(e^H) < \infty$, and define the density matrix σ by

$$\sigma = \frac{e^H}{\mathrm{Tr}(e^H)} \ .$$

Then, by the positivity of the relative entropy,

$$\mathrm{Tr}(\rho\ln\rho - \rho\ln\sigma) \ge 0 \ ,$$

with equality if and only if $\sigma = \rho$. But by the definition of σ, this reduces to

$$\mathrm{Tr}(\rho\ln\rho) \ge \mathrm{Tr}(\rho H) - \ln\left(\mathrm{Tr}\left(e^H\right)\right) \ ,$$

with equality if and only if $H = \ln\rho$. From here, the rest is very simple, including the treatment of the normalized trace. ■

Petz [27] has shown how to extend Lemma 2.3 to a much more general context, and his result can be used to extend the validity Theorem 2.2 beyond the tracial case. However, since the examples in which we prove the generalized subadditivity inequality here are tracial, we shall not go into this.

Proof of Theorem 2.2: Suppose, first, that the non-commutative B-L inequality (1.5) holds. Then, for any probability density ρ in $\mathfrak{A}$, and any self-adjoint $H_j \in \mathfrak{A}_j$, $j = 1, \dots, N$, apply (2.3) with $A = \rho$ and $H = \sum_{j=1}^N \phi_j(H_j)$ to obtain

$$\begin{aligned} -S(\rho) &\geq \lambda\left(\rho\left[\sum_{j=1}^N \phi_j(H_j)\right]\right) - \ln\left[\lambda\left(\exp\left[\sum_{j=1}^N \phi_j(H_j)\right]\right)\right] \\ &= \sum_{j=1}^N \lambda_j(\rho_j H_j) - \ln\left[\lambda\left(\exp\left[\sum_{j=1}^N \phi_j(H_j)\right]\right)\right] \\ &\geq \sum_{j=1}^N \lambda_j(\rho_j H_j) - \ln\left[C \prod_{j=1}^N \lambda_j\left(e^{p_j H_j}\right)^{1/p_j}\right] \\ &= \sum_{j=1}^N \frac{1}{p_j}\left[\lambda_j(\rho_j[p_j H_j]) - \ln\left(\lambda_j\left(e^{[p_j H_j]}\right)\right)\right] - \ln C \ . \end{aligned} \tag{2.5}$$

The first inequality here is (2.3), and the second is the non-commutative B-L inequality (1.5).

Now choosing $p_j H_j$ to maximize $\lambda_j(\rho_j[p_j H_j]) - \ln\left(\lambda_j\left(e^{[p_j H_j]}\right)\right)$, we get from (2.3) once more that

$$\lambda_j(\rho_j[p_j H_j]) - \ln\left(\lambda_j\left(e^{[p_j H_j]}\right)\right) = -S(\rho_j) = \lambda_j(\rho_j \ln \rho_j) \ .$$

Thus, we have proved (2.1) with the same $p_1, \dots, p_N$ and C that we had in (1.5).

Next, suppose that (2.1) is true. We shall show that in this case the non-commutative B-L inequality (1.5) holds with the same $p_1, \dots, p_N$ and C. To do this, let the self-sadjoint operators $H_1, \dots, H_N$ be given, and define

$$\rho = \left[\lambda\left(\exp\left[\sum_{j=1}^N \phi_j(H_j)\right]\right)\right]^{-1} \exp\left[\sum_{j=1}^N \phi_j(H_j)\right] \ .$$

Then by Lemma 2.3,

$$
\begin{aligned}
\ln\left[\lambda\left(\exp\left[\sum_{j=1}^N \phi_j(H_j)\right]\right)\right] &= \lambda\left(\rho\left[\sum_{j=1}^N \phi_j(H_j)\right]\right) + S(\rho)\\
&= \sum_{j=1}^N \lambda_j\left[\rho_j H_j\right] + S(\rho)\\
&\le \sum_{j=1}^N \frac{1}{p_j}\left[\lambda_j\left[\rho_j(p_j H_j)\right] + S(\rho_j)\right] + \ln C\\
&\le \sum_{j=1}^N \frac{1}{p_j}\ln\left[\lambda_j\left(\exp(p_j H_j)\right)\right] + \ln C
\end{aligned}
\tag{2.6}
$$

The first inequality is the generalized subadditivity of entropy inequality (2.1), and the second is (2.4).
Exponentiating both sides of (2.6), we obtain (1.5) with the same $p_1, \dots, p_N$ and C that we had in (2.1). ■

3 Proof of the generalized subadditivity of entropy inequality for tensor products of Hilbert spaces

The crucial tool that we use in this section is the *strong subadditivity of entropy* [24], which we now recall in a formulation that is suited to our purposes.
Suppose, as in Example 1.3, that we are given n separable Hilbert spaces $\mathcal{H}_1, \dots, \mathcal{H}_n$. As before, let $\mathcal{K}$ denote their tensor product, and for any non-empty subset J of $\{1, \dots, n\}$, let $\mathcal{K}_J$ denote $\bigotimes_{j\in J}\mathcal{H}_j$.
For a density matrix ρ on $\mathcal{K}$, and any non-empty subset J of $\{1, \dots, n\}$, define $\rho_J = \mathrm{Tr}_{J^c}\rho$ to be the density matrix on $\mathcal{K}_J$ induced by the natural injection of $\mathfrak{B}(\mathcal{K}_J)$ into $\mathfrak{B}(\mathcal{K})$. As noted above, ρ_J is nothing other than the partial trace of ρ over the complementary product of Hilbert spaces, $\bigotimes_{j\notin J}\mathcal{H}_j$.
The strong subadditivity of entropy is expressed by the inequality stating that for all nonempty $J, K \subset \{1, \dots, n\}$,

$$S(\rho_J) + S(\rho_K) \ge S(\rho_{J\cup K}) + S(\rho_{J\cap K})\ . \tag{3.1}$$

In case $\mathcal{J} \cap K = \emptyset$, it reduce to the ordinary subadditivity of entropy, which is the elementary inequality

$$S(\rho_J) + S(\rho_K) \ge S(\rho_{J\cup K}) \qquad \text{for} \quad J\cap K = \emptyset\ . \tag{3.2}$$

Combining these, we have

$$
\begin{aligned}
S(\rho_{\{1,2\}}) + S(\rho_{\{2,3\}}) + S(\rho_{\{3,1\}}) &\ge S(\rho_{\{1,2,3\}}) + S(\rho_{\{2\}}) + S(\rho_{\{1,3\}})\\
&\ge 2S(\rho_{\{1,2,3\}})\ ,
\end{aligned}
\tag{3.3}
$$

where the first inequality is the strong subadditivity (3.1) and the second is the ordinary subadditivity (3.2). Thus, for $n = 3$ and $J_1 = \{1,2\}$, $J_2 = \{2,3\}$ and $J_3 = \{3,1\}$, we obtain

$$\frac{1}{2}\sum_{j=1}^{N} S(\rho_{J_j}) \geq S(\rho) \ .$$

In this example, each index $i \in \{1,1,3\}$ belonged to exactly two of the set J_1, J_2 and J_3, and this is the source of the factor of $1/2$ in the inequality. The same procedure leads to the following result:

3.1 THEOREM. *Let $J_1,\dots,J_N$ be N non-empty subsets of $\{1,\dots,n\}$ For each $i \in \{1,\dots,n\}$, let $p(i)$ denote the number of the sets $J_1,\dots,J_N$ that contain i, and let p denote the minimum of the $p(i)$. Then*

$$\frac{1}{p}\sum_{j=1}^{N} S(\rho_{J_j}) \geq S(\rho) \tag{3.4}$$

for all density matrices ρ on $\mathcal{K} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$.

Proof: Simply use strong subadditivty to combine overlapping sets to produce as many "complete" sets as possible, as in the example above. Clearly, there can be no more than p of these. If $p(i) > p$ for some indices i, there will be "left over" partial sets. The entropy is always non-negative, and therefore, discarding the corresponding entropies gives us $\sum_{j=1}^{N} S(\rho_{J_j}) \geq pS(\rho)$, and hence the inequality. ∎

It is now a very simple matter to prove Theorem 1.4:

Proof of Theorem 1.4: By the remarks made after the statement of the theorem, all that remains to be proved is the inequality (1.8) for $q = p$. However, this follows directly from Theorem 2.2 and Theorem 3.4. ∎

4 On the generalized Young's inequality with a Gaussian reference measure

Before turning to the proof of our non-commutative B-L inequality in Clifford algebras, we discuss the commutative case in which the reference measures is Gaussian. We do this here for two reasons: First, as noted, a Clifford algebra $\mathfrak{C}$ with its normalized trace τ is a non-commutative analog of a Gaussian measure space. This analogy is strong enough that we shall be able to pattern our analysis in the Clifford algebra case on an analysis of the Gaussian case.

Second, the Gaussian inequality is of interest in itself, and seems not to have been fully studied before. Suppose that $V_1,\dots,V_N$ are N non-zero subspaces of $\mathbb{R}^n$, and for each j, define $\phi_j = P_j$ to be the orthogonal projection of $\mathbb{R}^n$ onto V_j. Equip $\mathbb{R}^n$ and equip each V_j with Lebesgue measure. Then the problem of determining for which sets of indices $\{p_1,\dots,p_N\}$ there exists a finite constant C so that (1.3) holds for all non-negative measurable functions f_j on V_j, $j = 1, \dots, N$ is highly non trivial, and has only recently been fully solved [7, 8]. Moreover, determining the value of the best constant C for those choices of $\{p_1, \dots, p_N\}$ is still a challenging finite dimensional variational problem for which there is no general explicit solution.

In contrast, suppose we are given a non-degenerate Gaussian measure on $\mathbb{R}^n$. It will be convenient to take the covariance matrix of the Gaussian to define the inner product, so that the Gaussian becomes a unit Gaussian. For each positive integer m, define $\gamma_m(x) = (2\pi)^{-m/2}e^{-|x|^2/2}$ on $\mathbb{R}^m$. Then equipping $\mathbb{R}^n$ with the measure $\gamma_n(x)\mathrm{d}x$ and equipping each V_j with the $\gamma_{d_j}(x)\mathrm{d}x$, d_j being the dimension of V_j, it turns out that there is a *very simple necessary and sufficient condition* on the indices $\{p_1, \dots, p_N\}$ for the constant C to be finite, and better yet, *the best constant C is always 1 whenever it is finite*:

4.1 THEOREM. *Let $V_1, \dots, V_N$ be N non-zero subspaces of $\mathbb{R}^n$, and for each j, and let d_j denote the dimension of V_j. Define P_j to be the orthogonal projection of $\mathbb{R}^n$ onto V_j. Given the numbers p_j, $1 \le p_j < \infty$ for $j = 1, \dots, N$, there exists a finite constant C such that*

$$\int_{\mathbb{R}^n} \prod_{j=1}^N f_j \circ P_j(x)\gamma_n(x)\mathrm{d}x \le C \prod_{j=1}^N \left(\int_{V_j} f_j^{p_j}(y)\gamma_{d_j}(y)\mathrm{d}y \right)^{1/p_j} \tag{4.1}$$

holds for all non-negative f_j on V_j, $j = 1, \dots, N$, if and only if

$$\sum_{j=1}^N \frac{1}{p_j} P_j \le \mathrm{Id}_{\mathbb{R}^n} \tag{4.2}$$

and in this case, $C = 1$.

We hasten to point out that this theorem is partially known. In the special case that each of the subspaces V_j is one dimensional, Barthe and Cordero-Erausquin [2], have proved that it is *sufficient* to have equality in (4.2); in this one dimensional context, their sufficient condition then is

$$\sum_{j=1}^N \frac{1}{p_j} u_j \otimes u_j = \mathrm{Id}_{\mathbb{R}^n} \tag{4.3}$$

with each u_j being a unit vector spanning V_j. They did this as an intermediate step in a short proof of the *Lebesgue measure* version of the B-L inequality under the condition (4.3) – the so-called *geometric case*. Perhaps because their main focus was the Lebesgue measure case, in which (4.3) is not a necessary condition for finiteness of the constant C, they did not address the necessity of this condition in the Gaussian case.

Indeed, under the condition that equality holds in (4.2), the inequality (4.1) is equivalent to its Lebesgue measure analog, which is known to hold with the constant $C = 1$ under this same condition [7, 8]. To see this, define $g_1, \ldots, g_N$ by

$$g_j(y) = f_j(y)(\gamma_j(y))^{1/d_j} \qquad j = 1, \ldots, N \ .$$

As noted in [2] for the one dimensional case, this change of variable allows one to pass back and forth between the Gaussian and Lebesgue measure version of the B-L inequality – under the condition (1.15).

It is also perhaps worth remarking that once one has proved the sufficiency of equality in (4.2), the inequality case follows easily: By the spectral theorem, one can write

$$I - \sum_{j=1}^{N} \frac{1}{p_j} P_j = \sum_{k=1}^{n} \nu_k Q_k$$

where the Q_k are rank one projections, and each ν_k satisfies $0 \leq \nu_k \leq 1$. For each k such that $\nu_k > 0$, define $q_k = 1/\nu_k$, and augment the sets of p_j's and P_j's by adding in all such q_k's and Q_k's. The augmented sets now satisfy (4.2) with equality, but the augmentation is "invisible" in the augmented version of (4.1) if we specialize to $f_k = 1$ identically for each index k in the augmentation. Hence, the equality case discussed in [2] really is the heart of the matter.

Nonetheless, it is worthwhile to give a proof of Theorem 4.1 here for two reasons: First, it may be surprising that the condition (1.15) is necessary for the inequality to hold with any finite constant at all. Second, the proof we will give of sufficiency of the condition (1.15) serves as a model for the proof of the corresponding theorem in the Clifford algebra case that we consider in the next section.

In proving Theorem 4.1, our first step is to pass to the problem of proving a generalized subadditivity inequality. Because the commutative version of Theorem 2.2 has been proved in [10], Theorem 4.2 below on subbadditivity of entropy with respect to a Gaussian reference measure is equivalent to Theorem 4.1. Hence, it suffices to prove one or the other.

Before stating and proving the subadditivty theorem, we first recall that for any probability density ρ on $(\mathbb{R}^m, \mathrm{d}\gamma_m)$, the entropy of ρ, is defined by

$$S(\rho) = -\int_{\mathbb{R}^m} \rho(y) \ln \rho(y) \gamma_m(y) \mathrm{d}y \ .$$

Note that the relative entropy of $\rho(y)\gamma_m(y)\mathrm{d}y$ to $\gamma_m(y)\mathrm{d}y$ is $-S(\rho)$; in the convention employed here, the entropy S is concave, and the relative entropy is convex.

4.2 THEOREM. *Let $V_1, \ldots, V_N$ be N non-zero subspaces of $\mathbb{R}^n$, and for each j, let d_j denote the dimension of V_j. Define P_j to be the orthogonal projection of $\mathbb{R}^n$ onto V_j. For any probability density ρ on $(\mathbb{R}^n, \mathrm{d}\gamma_n)$, let ρ_{V_j} denote the marginal density on $(V_j, \mathrm{d}\gamma_{d_j})$. Then, given the numbers p_j, $1 \leq p_j < \infty$ for $j = 1, \ldots, N$, there exists a finite constant C such that*

$$\sum_{j=1}^{N} \frac{1}{p_j} S(\rho_{V_j}) \geq S(\rho) - \ln(C) \tag{4.4}$$

holds for all probability densities ρ *on* $(\mathbb{R}^n, \mathrm{d}\gamma_n)$*, if and only if*

$$\sum_{j=1}^{N} \frac{1}{p_j} P_j \leq I \tag{4.5}$$

and in this case, $\ln(C) = 0$.

We first prove necessity of the condition (4.5):

4.3 LEMMA. *The condition (4.5) in Theorem 4.2 is necessary.*

PROOF: It suffices to consider densities of the form

$$\rho(x) = \exp(b \cdot x - |b|^2/2) \ ,$$

for $b \in \mathbb{R}^n$. Then

$$\rho_{V_j}(x) = \exp(P_j b \cdot y - |P_j b|^2/2) \ ,$$

and we compute:

$$S(\rho) = -\frac{|b|^2}{2} \qquad \text{and} \qquad S(\rho_{V_j}) = -\frac{|P_j b|^2}{2} \ .$$

Thus

$$\sum_{j=1}^{N} \frac{1}{p_j} S(\rho_{V_j}) - S(\rho) = b \cdot \left[Id_{\mathbb{R}^n} - \sum_{j=1}^{N} \frac{1}{p_j} P_j \right] b \ ,$$

and evidently this is bounded below if and only if (4.5) is satisfied. ■

4.1 Proof of sufficiency

The sufficiency of the condition (4.5) will be proved using an interpolation between an arbitrary density ρ and the uniform density that is provided by the Mehler semigroup. (Indeed, Barthe and Coredero-Erausquin used the Mehler semigroup in their work [2] mentioned above, but used it in a direct proof of the Gaussian B-L inequality inspired by the heat-flow method introduced in [13]. The heat flow approach to prove subadditivity inequalities was developed in [3] and [10].)

The Mehler semigroup is the strongly continuous semigroup of positivity preserving contractions on $L^2(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$ whose generator $-\mathcal{N}$ is given by the Dirichlet form

$$\mathcal{E}(f, g) = \int_{\mathbb{R}^n} \nabla f^*(x) \cdot \nabla g(x) \gamma_n(x) \mathrm{d}x \tag{4.6}$$

through $\langle f, \mathcal{N} g\rangle_{L^2(\gamma_n)} = \mathcal{E}(f, g)$, where f^* is the complex conjugate of f. Integrating by parts, one finds

$$\mathcal{N} = -(\Delta - x \cdot \nabla) \ ,$$

The eigenvalues of $\mathcal{N}$ are the non-negative integers, and the eigenfunctions are the Hermite polynomials. (In certain physical contexts, the eigenvalues count occupancy of a quantum state and $\mathcal{N}$ is called the *Boson number operator*.) There is a simple explicit formula for the $e^{-t\mathcal{N}}$:

$$e^{-t\mathcal{N}}f(x) = \int_{\mathbb{R}^n} f\left(e^{-t}x + \sqrt{1-e^{2t}}y\right)\gamma_n(y)\mathrm{d}y \ , \tag{4.7}$$

which is easily checked.
Since evidently $\mathcal{N}1 = 0$, and $e^{-t\mathcal{N}}$ is self-adjoint, it also preserves integrals against $\gamma_n(x)\mathrm{d}x$, and hence, if ρ is any probability density, so is each $\rho_t := e^{-t\mathcal{N}}$. As one sees from (4.7),

$$\lim_{t\to\infty} e^{-t\mathcal{N}}\rho(x) = 1 \ , \tag{4.8}$$

the uniform probability density on $(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$, and thus the Mehler semigroup provides us with an interpolation between any probability density ρ and the uniform density 1.
This interpolation is well-behaved with respect to the operation of taking marginals: Consider any probability density ρ on $(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$, and any m dimensional subspace V of $\mathbb{R}^n$. Let ρ_V be the marginal density of ρ as in Theorem 4.2. Then of course, we may regard ρ_V as a probability density on $(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$ that is constant along directions in $V^\perp$. (Simply compose ρ_V with P_V.) Interpreted this way, so that both ρ and ρ_V are functions on $\mathbb{R}^N$,

$$\left(e^{-t\mathcal{N}}\rho\right)_V = e^{-t\mathcal{N}}\left(\rho_V\right) \ . \tag{4.9}$$

That is, the process of taking marginals commutes with the action of the Mehler semigroup.
The next point to note is that the entropy is monotone increasing along this interpolation: Differentiating, with $\rho_t = e^{-t\mathcal{N}}\rho$,

$$\frac{\mathrm{d}}{\mathrm{d}t}S(\rho_t) = -\int_{\mathbb{R}^n} \ln(\rho_t)(\Delta - x\cdot\nabla)\rho_t\gamma_n\mathrm{d}x = \int_{\mathbb{R}^n} \nabla\ln\rho_t\cdot\nabla\rho_t\gamma_n\mathrm{d}x = \mathcal{E}(\ln\rho_t, \rho_t) \ .$$

For any smooth density ρ, $\mathcal{E}(\ln\rho, \rho) = \int_{\mathbb{R}^n}\nabla\ln\rho\cdot\nabla\rho\gamma_n\mathrm{d}x = \int_{\mathbb{R}^n}|\nabla\ln\rho|^2\rho\gamma_n\mathrm{d}x$, and hence $S(\rho_t)$ is strictly increasing for all t. Moreover, since $(x,t)\mapsto |x|^2/t$ is jointly convex on $\mathbb{R}^n\times\mathbb{R}_+$, $\rho\mapsto\mathcal{E}(\ln\rho,\rho)$ has a unique extension as a convex functional on the set of all probability densities on $(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$.

4.4 DEFINITION (Entropy Production). The *entropy production* functional is the convex functional $D(\rho)$ on probability densities on $(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$ given by

$$D(\rho) = \int_{\mathbb{R}^n} \ln\rho(x)\mathcal{N}\rho(x)\gamma_n(x)\mathrm{d}x = \mathcal{E}(\ln\rho, \rho) \ . \tag{4.10}$$

With this definition,

$$\frac{\mathrm{d}}{\mathrm{d}t}S(e^{-t\mathcal{N}}\rho) = D(e^{-t\mathcal{N}}\rho) \ .$$

Now because of (4.9), for any subspace V of $\mathbb{R}^n$,

$$\frac{\mathrm{d}}{\mathrm{d}t}S([e^{-t\mathcal{N}}\rho]_V) = D([e^{-t\mathcal{N}}\rho]_V) \ .$$

Since $[e^{-t\mathcal{N}}\rho]_V$ is constant along directions orthogonal to V, the derivatives in those directions that figure in $D([e^{-t\mathcal{N}}\rho]_V)$ are irrelevant; we need only take derivatives along directions in V. This consideration leads to the definitions of the *restricted number operator*, and the *restricted entropy production*:
Given an m dimensional subspace V of $\mathbb{R}^n$, let P_V be the orthogonal projection onto V. The restricted number operator $\mathcal{N}_V$ is the self-adjoint operator on $L^2(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$ defined through

$$\langle f, \mathcal{N}_V g\rangle_{L^2(\gamma_n)} = \int_{\mathbb{R}^n} \nabla f^*(x)\cdot P_V\nabla g(x)\gamma_n(x)\mathrm{d}x \ , \tag{4.11}$$

and the *restricted entropy production functional* $D_V(\rho)$ is the convex functional given by

$$D_V(\rho) = \int_{\mathbb{R}^n} \left(\mathcal{N}_V \ln\rho(x)\right)\rho(x)\gamma_n(x)\mathrm{d}x \ . \tag{4.12}$$

With this definition, $D(\rho_V) = D_V(\rho_V)$. However, there is a crucial difference between $D_V(\rho)$ and $D(\rho_V)$:

4.5 LEMMA. *For any smooth probability density ρ on $(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$, and any m dimensional subspace V of $\mathbb{R}^n$, let ρ_V be the corresponding marginal density regarded as a probability density on $(\mathbb{R}^n, \gamma_n(x)\mathrm{d}x)$. Then*

$$D(\rho_V) \le D_V(\rho) \ . \tag{4.13}$$

Proof: Regard ρ_V as a function on $\mathbb{R}^n$ (by composing it with P_V). Assume that ρ is smooth and bounded above and below by strictly positive numbers. Notice that since ρ_V is constant along directions in $V^\perp$,

$$\mathcal{N}\ln\rho_V = \mathcal{N}_V\ln\rho_V \ .$$

Then, integrating by parts, and using the definition of ρ_V and the Schwarz inequality, we obtain:

$$D(\rho_V) = \int_{\mathbb{R}^n}\left[\mathcal{N}_V\ln\rho_V(x)\right]\rho_V(x)\gamma_n(x)\mathrm{d}x = \int_{\mathbb{R}^n}\left[\mathcal{N}_V\ln\rho_V(x)\right]\rho(x)\gamma_n(x)\mathrm{d}x \ ,$$

where we have used the definition of ρ_V to replace the second ρ_V by ρ itself. Then, by the definition of $\mathcal{N}_V$, and the Schwarz inequality,

$$
\begin{aligned}
D(\rho_V) &= \int_{\mathbb{R}^n} (\nabla \ln \rho_V(x)) \cdot P_V \nabla \rho(x) \gamma_n \mathrm{d}x \\
&= \int_{\mathbb{R}^n} (\nabla \ln \rho_V(x)) \cdot P_V \left(\nabla \ln \rho(x)\right) \rho(x) \gamma_n(x) \mathrm{d}x \\
&\leq \left(\int_{\mathbb{R}^n} |\nabla \ln \rho_V(x)|^2 \rho(x)\gamma_n(x)\mathrm{d}x\right)^{1/2} \left(\int_{\mathbb{R}^n} |P_V \nabla \ln \rho(x)|^2 \rho\gamma_n \mathrm{d}x\right)^{1/2}
\end{aligned}
\tag{4.14}
$$

In the first factor in the last line, we may replace ρ by ρ_V since $|\nabla \ln \rho_V(x)|^2$ depends on x only through $P_V x$. Hence this factor is simply $\sqrt{D(\rho_V)}$, and the second factor is $\sqrt{D_V(\rho)}$. ■

The proof we have just given is patterned on the proof of an analogous result in the Lebesgue measure case in [10], which in turn is based on similar arguments in [9] and [3]. It is somewhat more complicated to adapt the argument to the Clifford algebra setting, but this is what we shall do in the next section. We are now ready to prove the sufficiency of condition (4.5):

4.6 LEMMA. *The condition (4.5) in Theorem 4.2 is sufficient.*

Proof: For a probability density ρ on $(\mathbb{R}^n, \mathrm{d}\gamma_n)$ with $S(\rho) > -\infty$, it is easy to see that

$$\lim_{t\to\infty} S(e^{-t\mathcal{N}}\rho) = S(1) = 0$$

and hence, $\lim_{t\to\infty} S(e^{-t\mathcal{N}}(\rho_{V_j})) = 0$ for each $j = 1, \ldots, N$. Therefore, it suffices to show that

$$a(t) := \left[\sum_{j=1}^N \frac{1}{p_j} S(e^{-t\mathcal{N}}\rho_{V_j}) - S(e^{-t\mathcal{N}}\rho)\right]$$

is monotone decreasing in t.

Differentiating, and using (4.9), and then Lemma 4.5,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} a(t) &= \left[\sum_{j=1}^N \frac{1}{p_j} D((e^{-t\mathcal{N}}\rho)_{V_j}) - D(e^{-t\mathcal{N}}\rho)\right] \\
&\leq \left[\sum_{j=1}^N \frac{1}{p_j} D_{V_j}(e^{-t\mathcal{N}}\rho) - D(e^{-t\mathcal{N}}\rho)\right]
\end{aligned}
\tag{4.15}
$$

Now note that by (4.12), whenever (4.5) is satisfied,

$$\sum_{j=1}^N \frac{1}{p_j} D_{V_j}(\sigma) \leq D(\sigma)$$

for any smooth density σ. Hence the derivative of $\alpha(t)$ is negative for all $t > 0$. ■

5 Generalized subadditivity of entropy in Clifford algebras

In this section we shall prove

5.1 THEOREM. *Let $V_1, \dots, V_N$ be N subspaces of $\mathbb{R}^n$, and let $\mathfrak{A}_j$ be the Clifford algebra over V_j with the inner product that V_j inherits from $\mathbb{R}^n$, and let $\mathfrak{A}_j$ be equipped with its unique tracial state τ_j. For any probability density $\rho \in \mathfrak{A}$, let ρ_{V_j} be the induced probability density in $\mathfrak{A}_j$. Let $S(\rho) = -\tau(\rho \ln \rho)$ and $S(\rho_{V_j}) = -\tau_j(\rho_{V_j} \ln \rho_{V_j})$*
Then

$$\sum_{j=1}^N \frac{1}{p_j} S(\rho_{V_j}) \geq S(\rho) \tag{5.1}$$

for all probability densities $\rho \in \mathfrak{A}$ if and only if

$$\sum_{j=1}^N \frac{1}{p_j} P_j \leq I_{\mathbb{R}^n} \ . \tag{5.2}$$

where P_j is the orthogonal projection onto V_j in $\mathbb{R}^n$.

Granted this result, we have:
Proof of Theorem 1.6: Theorem 2.2 and Theorem 5.1 together prove Theorem 1.6. ■
We shall now prove Theorem 5.1. As before, we begin by proving the necessity of (5.2).

5.2 LEMMA. *The condition (5.2) in Theorem 5.1 is necessary.*

Proof: For any vector $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, define

$$\rho_a = I + \sum_{j=1}^n a_j Q_j = I + a \cdot Q \ .$$

Then ρ_a is a probability density if and only if $|a| \leq 1$. Indeed, ρ_a has only two eigenvalues, $1 \pm |a|$, with equal multiplicity.
Then $(\rho_a)_{V_j} = I + (P_j a) \cdot Q$, and so $(\rho_a)_{V_j}$ has only two eigenvalues, $1 \pm |P_j a|$, with equal multiplicity. Therefore,

$$S(\rho_a) = -\psi(|a|) \qquad \text{and} \qquad S((\rho_a)_{V_j}) = -\psi(|P_j a|) \ . \tag{5.3}$$

where $\psi(x)$ is the convex function defined by

$$\psi(x) = \begin{cases} \frac{1}{2}\left[(1+x)\ln(1+x) + (1-x)\ln(1-x)\right] & \text{if } |x| \leq 1. \\ \infty & \text{otherwise,} \end{cases} \tag{5.4}$$

Thus, for (5.1) to hold for each ρ_a, $|a| \leq 1$, it must be the case that

$$\sum_{j=1}^{N} \frac{1}{p_j} \psi(|P_j a|) \leq \psi(|a|) \qquad \text{for all } a \text{ with } |a| \leq 1 \ . \tag{5.5}$$

Then, since $\psi(x) = x^2 + \mathcal{O}(x^4)$, replacing a by ta, $0 < t < 1$, we see that (5.2) must hold. ∎

Because of (5.3), once we have proved Theorem 5.1, we will have a proof of (5.5). However, it is of interest to prove this inequality directly, and we do that next.

5.3 PROPOSITION. *The inequality (5.5) holds whenever (5.2) is satisfied.*

Proof: An easy calculation of derivatives shows that

$$\psi'(x) = \operatorname{arctanh}(x) \qquad \text{and} \qquad \psi''(x) = \frac{1}{1-x^2}$$

for $|x| < 1$.

Now fix any a with $|a| < 1$. Then, for $t > 0$, define

$$\eta(t) = \psi(e^{-t}|a|) - \sum_{j=1}^{N} \frac{1}{p_j} \psi(e^{-t}|P_j a|) \ .$$

We have to show that $\eta(t) > 0$ for all $t > 0$. Since evidently $\lim_{t\to\infty} \eta(t) = 0$, it suffices to show that $\eta'(t) < 0$ for all $t > 0$.

Differentiating, we find

$$\eta'(t) = -e^{-t}\left[|a| \ \operatorname{arctanh}(e^{-t}|a|) - \sum_{j=1}^{N} \frac{1}{p_j} |P_j a| \ \operatorname{arctanh}(e^{-t}|P_j a|) \right] := e^{-t}\theta(t).$$

Hence, it suffices to show that $\theta(t) \geq 0$ for all $t > 0$. Once again, since $\lim_{t\to\infty} \theta(t) = 0$, it suffices to show that $\theta'(t) < 0$ for all $t > 0$. Differentiating, we find

$$\theta'(t) = -e^{-t} \left[\frac{|a|^2}{1 - e^{-2t}|a|^2} - \sum_{j=1}^{N} \frac{1}{p_j} \frac{|P_j a|^2}{1 - e^{-2t}|P_j a|^2} \right] \ .$$

Multiplying through by e^{-t}, and absorbing a factor of e^{-t} into a, it suffices to show that

$$\frac{|a|^2}{1-|a|^2} \geq \sum_{j=1}^{N} \frac{1}{p_j} \frac{|P_j a|^2}{1 - |P_j a|^2} \tag{5.6}$$

for all $|a| \leq 1$. However, since $|a| \geq |P_j a|$,

$$\frac{|P_j a|^2}{1-|a|^2} \geq \frac{|P_j a|^2}{1-|P_j a|^2} \ ,$$

and thus (5.6) follows from (5.2). ■

We are now in a position to give an elementary proof of Theorem 1.6 in the special case that each V_j is one dimensional. As explained in Example 1.5, it suffices in this case to prove the following:

5.4 PROPOSITION. *Suppose $\{u_1,\dots,u_N\}$ is any set of N unit vectors in $\mathbb{R}^n$, and $\{p_1,\dots,p_N\}$ is any set of N positive numbers such that*

$$\sum_{j=1}^{N} c_j u_j \otimes u_j = I_{\mathbb{R}^n} \ . \tag{5.7}$$

Then for any $b=(b_1,\dots,b_N)$ in$\mathbb{R}^N$,

$$\ln\cosh\left(\left|\sum_{j=1}^{N} b_j u_j\right|\right) \le \sum_{j=1}^{N} \frac{1}{p_j}\ln\cosh(p_j b_j) \ . \tag{5.8}$$

Proof: Let ψ^* denote the function $\psi^*(x)=\ln\cosh(x)$, $x\in\mathbb{R}$. The notation is meant to indicate the well known fact, easily checked, that ψ^* is the Legendre transform of the function ψ defined in (5.4).

Now, given a set of N orthogonal projections $\{P_1,\dots,P_N\}$ satisfying (5.2), we may make any choice of a unit vector u_j from the range of P_j, and then the N unit vectors $\{u_1,\dots,u_N\}$ will satisfy (5.7). Conversely, given any set of N unit vectors $\{u_1,\dots,u_N\}$ that satisfy (5.7), we may take $P_j=u_j\otimes u_j$, and then (5.2) is satisfied. Hence, we suppose we are given a set of N orthogonal projections $\{P_1,\dots,P_N\}$ satisfying (5.2), and for each j, u_j is a unit vector in the range of P_j.

Then for any $b\in\mathbb{R}^n$,

$$\begin{aligned}
\psi^*\left(\left|\sum_{j=1}^{N} b_j u_j\right|\right) &= \sup_{a\in\mathbb{R}^n}\left\{a\cdot\sum_{j=1}^{N} b_j u_j - \psi(|a|)\right\}\\
&= \sup_{|a|\le 1}\left\{\sum_{j=1}^{N} P_j a\cdot b_j u_j - \psi(|a|)\right\}\\
&\le \sup_{|a|\le 1}\left\{\sum_{j=1}^{N} P_j a\cdot b_j u_j - \sum_{j=1}^{N}\frac{1}{p_j}\psi(|P_j a|)\right\}\\
&\le \sup_{|a|\le 1}\left\{\sum_{j=1}^{N} |P_j a||b_j| - \sum_{j=1}^{N}\frac{1}{p_j}\psi(|P_j a|)\right\}\\
&= \sup_{|a|\le 1}\left\{\sum_{j=1}^{N}\frac{1}{p_j}\big[|P_j a|p_j|b_j| - \psi(|P_j a|)\big]\right\} \ ,
\end{aligned} \tag{5.9}$$

where the first inequality is from (5.5), and the second is from Schwarz. Then, by the definition of the Legendre transform, for any a,

$$\psi^*(p_j b_j) \geq |P_j a|(p_j|b_j|) - \psi(|P_j a|) \ ,$$

we obtain

$$\psi^*\left(\left|\sum_{j=1}^N b_j u_j\right|\right) \leq \sum_{j=1}^N \frac{1}{p_j}\psi^*(p_j b_j) \ ,$$

which is (5.8). ■

We now prove Theorem 5.1 in full generality. This gives another proof of the last two propositions, but by less elementary means. The proof will follow the basic pattern of the proof of Theorem 4.2 and use the Clifford algebra analog of the Mehler semigroup. This is the so-called Clifford–Mehler semigroup, about which we now recall a few relevant facts.

5.1 About the Clifford–Mehler semigroup

There is also a differential calculus in the Clifford algebra. Let $Q_1, \ldots, Q_n$ be any set of n generators for the Clifford algebra $\mathfrak{C}$ over $\mathbb{R}^n$. For $A \in \mathfrak{C}$, define

$$\nabla_i(A) = \frac{1}{2}\left[Q_i A - \Gamma(A) Q_i\right] \ ,$$

where Γ is the *grading operator* on $\mathfrak{C}$: That is, using the notation in (1.10),

$$\Gamma(Q^\alpha) = (-1)^{|\alpha|} Q^\alpha \ .$$

One computes that $\nabla_i(Q^\alpha) = 0$ if $\alpha(i) = 0$, and otherwise, $\nabla_i(Q^\alpha)$ is what one gets by anti-commuting the factor of Q_i through to the left, and then removing it. In this sense it is like a differentiation operator, and what is more, it is a skew derivation on $\mathfrak{C}$, which means that for all and A and B in $\mathfrak{C}$, $\nabla_j(AB) = \nabla_j(A)B + \Gamma(A)\nabla_j(B)$.

The Clifford algebra analog of the Gaussian energy integral (4.6) is given by

$$\mathcal{E}(A, B) = \tau\left(\sum_{j=1}^n \nabla_j A^* \nabla_j B\right) \ , \tag{5.10}$$

for all $A, B \in \mathfrak{C}$. This is the *Clifford Dirichlet form* studied by Gross [18]. Then, the fermionic number operator, also denoted $\mathcal{N}$, is defined by

$$\mathcal{E}(A, B) = \tau(A^* \mathcal{N}(B)) \ .$$

It is easy to see that the spectrum of $\mathcal{N}$ consists of the non negative integers $\{0, 1, \ldots, n\}$ and that

$$\mathcal{N} Q^\alpha = |\alpha| Q^\alpha \ . \tag{5.11}$$

The Clifford Mehler semigroup is then given by $e^{-t\mathcal{N}}$. It is clear from this definition, (1.11) and (5.11) that for any $A \in \mathfrak{C}$, $\lim_{t\to\infty} e^{-t\mathcal{N}}(A) = \tau(A)I$. Thus for any probability density ρ in $\mathfrak{C}$,

$$t \mapsto \rho_t = e^{-t\mathcal{N}}(\rho)$$

provides an interpolation between ρ and I, and each ρ_t is a probability density. This corresponds exactly to the Mehler semigroup interpolation that was used to prove Theorem 4.2, and we shall use it here in the same way, though some additional complications shall arise.

The operator $\mathcal{N}$ does not depend on the choice of the set of generators $Q_1, \ldots, Q_n$. Indeed, if $\{u_1, \ldots, u_n\}$ is any orthonormal basis of $\mathbb{R}^n$, and we define $\widetilde{Q}_j = u_j \cdot Q$ for $j = 1, \ldots, n$, then the Clifford Dirichlet form that one obtains using this basis to define the derivatives is the same as the original.

In particular, given an m dimensional subspace V of $\mathbb{R}^n$, we may choose $\{u_1, \ldots, u_n\}$ so that $\{u_1, \ldots, u_m\}$ is an orthonormal basis for V, and then the first m generators will be a set of generators for $\mathfrak{C}_V$. We then define the *reduced Clifford Dirichlet form* $\mathcal{E}_V$ by

$$\mathcal{E}_V(A, B) = \tau\left(\sum_{i,j=1}^{m} \nabla_i A^* [P_V]_{i,j} \nabla_j B\right) \ , \tag{5.12}$$

where $[P_V]_{i,j}$ is the i,jth entry of the $n \times n$ matrix for P_V. The restricted number operator $\mathcal{N}_V$ is then the self-adjoint operator on $L^2(\mathfrak{C})$ given by $\tau(A^*\mathcal{N}_V(B)) = \mathcal{E}_V(A, B)$.

Now, for any probability density ρ in $\mathfrak{C}$ let ρ_V be the corresponding marginal density regarded as an operator in $\mathfrak{C}$ by identifying it with $\phi_V(\rho_V)$, where ϕ_V is the canonical embedding of $\mathfrak{C}(V)$ into $\mathfrak{C}(\mathbb{R}^n)$. Then it is an easy consequence of the definitions that

$$\left(e^{-t\mathcal{N}}\rho\right)_V = e^{-t\mathcal{N}}\left(\rho_V\right) = e^{-t\mathcal{N}_V}\left(\rho_V\right) \ . \tag{5.13}$$

Also, under the condition (5.2), it is easy to see that

$$\sum_{j=1}^{N} \frac{1}{p_j} \mathcal{N}_{V_j} \leq \mathcal{N} \ . \tag{5.14}$$

Finally, we introduce *entropy production* $D(\rho)$: With $\rho_t := e^{-t\mathcal{N}}\rho$, we differentiate and find

$$\frac{\mathrm{d}}{\mathrm{d}t} S\left(\rho_t\right) = \tau\left(\ln(\rho_t)\mathcal{N}(\rho_t)\right) = \mathcal{E}(\ln(\rho_t), \rho_t) \ .$$

5.5 DEFINITION (Entropy Production). The *entropy production* functional at a probability density ρ is the functional defined by

$$D(\rho) = \tau\left(\ln(\rho)\mathcal{N}(\rho)\right) = \mathcal{E}(\ln(\rho), \rho) \ . \tag{5.15}$$

Given an m dimensional subspace V of $\mathbb{R}^n$, the *restricted entropy production* functional at a probability density ρ is the functional defined by

$$D_V(\rho) = \tau\left(\ln(\rho)\mathcal{N}_V(\rho)\right) = \mathcal{E}_V(\ln(\rho), \rho) \ . \tag{5.16}$$

The following lemma is the basis of our proof of the sufficiency of (5.2). In the course of proving it, we shall see that both $D(\rho)$ and $D_V(\rho)$ are convex functionals, which is somewhat less obvious than in the Gaussian case.

5.6 LEMMA. *For any any probability density ρ in $\mathfrak{C}(\mathbb{R}^n)$, and any m dimensional subspace V of $\mathbb{R}^n$, let ρ_V be the corresponding marginal probability density regarded as a probability density in $\mathfrak{C}(\mathbb{R}^n)$. Then*

$$D(\rho_V) \leq D_V(\rho) \ .$$

Proof: We choose an orthonormal basis $\{u_1, \dots, u_n\}$ for $\mathbb{R}^n$ such that $\{u_1, \dots, u_m\}$ is an orthonormal basis for V. Without loss of generality, we may suppose that $\{u_1, \dots, u_n\}$ is the standard basis so that $\{Q_1, \dots, Q_m\}$ is a set of generators for $\mathfrak{C}(V)$. Then,

$$\mathcal{E}(A, B) = \tau\left(\sum_{j=1}^n \nabla_j A^* \nabla_j B\right) \qquad \text{and} \qquad \mathcal{E}_V(A, B) = \tau\left(\sum_{j=1}^m \nabla_j A^* \nabla_j B\right) \ . \tag{5.17}$$

It will be convenient to define $\mathcal{N}_j = \nabla_j^* \nabla_j \qquad j = 1, \dots, n$. Then we have

$$\mathcal{N} = \sum_{j=1}^n \mathcal{N}_j \qquad \text{and} \qquad \mathcal{N}_V = \sum_{j=1}^m \mathcal{N}_j \ , \tag{5.18}$$

and so

$$D_V(\rho) = \sum_{j=1}^m \tau\left(\ln\rho, \mathcal{N}_j \rho\right) \ . \tag{5.19}$$

Since $\mathcal{N}_j Q^\alpha = \begin{cases} Q^\alpha & \text{if } \alpha(j) = 1, \\ 0 & \alpha(j) = 0 \end{cases}$, each $\mathcal{N}_j$ is an orthogonal projection, and so (5.19) can be rewritten as

$$D_V(\rho) = \sum_{j=1}^m \tau\left(\mathcal{N}_j(\ln\rho), \mathcal{N}_j \rho\right) \ . \tag{5.20}$$

To proceed, we use a formula of Gross [18] for $\mathcal{N}_j f(A)$ where $A \in \mathfrak{C}(\mathbb{R}^n)$, and f is a continuous function. To write down Gross's formula, first write $A = B + Q_j C$ where both B and C are linear combinations of the Q^α with $\alpha(j) = 0$. Then define $\widehat{A} = B - Q_j C$. Notice that if ρ is a probability density, then $\widehat{\rho}$ is again a probability density. Gross's formula is

$$\mathcal{N}_j f(A) = \frac{1}{2}\left[f(A) - f(\widehat{A}\ )\right] \ .$$

To prove this formula, notice that there is a unitary operator U such that $\widehat{A} = UAU^*$. (If the dimension n is odd, one can take U to be the product, in some order, of all of the Q_k for $k \neq j$; if the dimension is even, one can add another generator.) Therefore,

$$\widehat{f(A)} = Uf(A)U^* = f(UAU^*) = f(\widehat{A}\ ) \ .$$

Using this together with the fact that for any $A \in \mathfrak{A}$, $\mathcal{N}_j A = (1/2)[A - \widehat{A}]$, we obtain Gross's formula, which we now apply as follows:

$$\begin{aligned} \tau\left(\mathcal{N}_j(\ln\rho)\mathcal{N}_j\rho\right) &= \frac{1}{4}\tau\left([\ln(\rho) - \ln(\widehat{\rho})]\,[\rho - \widehat{\rho}]\right) \\ &= \frac{1}{4}\tau\left(\ln(\rho)\,[\rho - \widehat{\rho}]\right) + \frac{1}{4}\tau\left(\ln(\widehat{\rho})\,[\widehat{\rho} - \rho]\right) \\ &= \frac{1}{4}H[\rho|\widehat{\rho}\ ] + \frac{1}{4}H[\widehat{\rho}\ |\rho] \end{aligned} \tag{5.21}$$

where $H[\rho|\sigma] = \tau\rho(\ ln\rho - \ln\sigma)$ is the relative entropy of ρ with respect to σ. As is well known, $(\rho,\sigma) \mapsto H(\rho|\sigma)$ is jointly convex, and hence

$$\rho \mapsto \tau\left((\ln\rho)\mathcal{N}_j\rho\right)$$

is convex. Furthermore, by the fundamental monotonicity property of the relative entropy under conditional expectations [37],

$$H(\rho_V|\sigma_V) \leq H(\rho|\sigma)$$

for any two probability densities ρ and σ. It follows that $\tau\left((\ln\rho_V)\mathcal{N}_j\rho_V\right) \leq \tau\left((\ln\rho)\mathcal{N}_j\rho\right)$. Summing on j from 1 to m, we find

$$D(\rho_V) = D_V(\rho_V) = \sum_{j=1}^m \tau\left((\ln\rho_V)\mathcal{N}_j\rho_V\right) \leq \sum_{j=1}^m \tau\left((\ln\rho)\mathcal{N}_j\rho\right) = D_V(\rho) \ .$$

■

5.2 Proof of the sufficiency

5.7 LEMMA. *The condition (4.5) in Theorem 4.2 is sufficient.*

Proof: For a probability density ρ in $\mathfrak{C}(\mathbb{R}^n)$ it is easy to see that

$$\lim_{t\to\infty} S(e^{-t\mathcal{N}}\rho) = S(1) = 0$$

and hence, $\lim_{t\to\infty} S(e^{-t\mathcal{N}}(\rho_{V_j})) = 0$ for each $j = 1,\ldots,N$. Therefore, it suffices to show that

$$a(t) := \left[\sum_{j=1}^N \frac{1}{p_j} S(e^{-t\mathcal{N}}\rho_{V_j}) - S(e^{-t\mathcal{N}}\rho)\right]$$

is monotone decreasing in t.
Differentiating, and using (5.13), and then Lemma 5.6,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}a(t) &= \left[\sum_{j=1}^{N}\frac{1}{p_j}D((e^{-t\mathcal{N}}\rho)_{V_j}) - D(e^{-t\mathcal{N}}\rho)\right] \\ &\leq \left[\sum_{j=1}^{N}\frac{1}{p_j}D_{V_j}(e^{-t\mathcal{N}}\rho) - D(e^{-t\mathcal{N}}\rho)\right] \end{aligned} \tag{5.22}$$

Now note that by (4.12), whenever (4.5) is satisfied, $\sum_{j=1}^{N}\frac{1}{p_j}D_{V_j}(\sigma) \leq D(\sigma)$ for any smooth density σ. Hence the derivative of $\alpha(t)$ is negative for all $t > 0$. ■

Notice that the proof is almost identical, symbol for symbol, with that of the corresponding proof in the Gaussian case. The main difference of course is that the proof of the main lemma, Lemma 5.6, is considerably more intricate than that of its Gaussian counterpart.

Proof of Theorem 5.1: This now follows immediately from Lemma 5.2 and 5.7. ■

References

[1] F. Barthe, *On a reverse form of the Brascamp-Lieb inequality*, Invent. Math. 134 (1998), no. 2, pp. 335–361.

[2] F. Barthe and D. Cordero–Erausquin *Inverse Brascamp-Lieb inequalities along the heat equation*, in *Geometric Aspects of Functional Analysis* (2002-2003), LNM 1850, Springer, 2004, pp. 65-71.

[3] F. Barthe, D. Cordero–Erausquin and B. Maurey: *Entropy of spherical marginals and related inequalities*, Adv. in Math. 20, 1976, pp. 151-173

[4] W. Beckner, *Inequalities in Fourier series*, Ann. of Math. 102, 1975, pp. 159-182.

[5] H.J. Brascamp and E.H. Lieb: *The best constant in Young's inequality and its generalization to more than three functions*,. J. Math. Pures Appl. 86, no. 2, 2006, pp. 89-99.

[6] R. Brauer and H. Weyl: *Spinors in n dimensions*, Am. Jour. Math., 57, 1935 pp. 425-449.

[7] J. Bennett, A. Carbery, M. Christ and T. Tao: *The Brascamp-Lieb inequalities: finiteness, structure, and extremals*, Geom. and Funct. Anal. 17 , 2008, pp. 1343-1415

[8] J. Bennett, A. Carbery, M. Christ and T. Tao: *Finite bounds for Hölder-Brascamp-Lieb multilinear inequalities*, Preprint.

[9] E.A. Carlen: *Superadditivity of Fisher information and logarithmic Sobolev inequalities*, Jour. of Func. Analysis, 101, 1991, pp. 194–211.

[10] E.A. Carlen and D Cordero–Erausquin: *Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities*, to appear in Geom. Funct. Anal. (2008).

[11] E.A. Carlen and E.H. Lieb: *Optimal Hypercontractivity for Fermi Fields and Related Non-commutative Integration Inequalities*, Comm. Math. Phys., 155, 1993 pp. 27-46.

[12] E.A. Carlen and E.H. Lieb: *Optimal two-uniform convexity and fermion hypercontractivity*, Proceedings of the Kyoto Conference in Honor of Araki's 60th birthday, Kluwer Academic Publishers, Amsterdam, 1993

[13] E.A. Carlen, E.H. Lieb and M. Loss: *A sharp form of Young's inequality on S^N and related entropy inequalities*, Jour. Geom. Analysis 14, 2004, pp. 487-520

[14] E.A. Carlen, E.H. Lieb and M. Loss: *A inequality of Hadamard type for permanents*, Meth. and Appl. of Analysis, 13, no. 1, 2006 pp. 1-17

[15] J. Dixmier: *Formes linéaires sur un anneau d'opérateurs*, Bull. Soc. Math. France 81, 1953, pp. 222–245

[16] T. Fak and H. Kosaki: *Generalized s-numbers of τ–measurable operators.* Pacific J. Math., 123, 1986, pp. 269-300,

[17] S. Golden: *Lower bounds for Helmholtz functions.* Phys. Rev. 137B, 1965, pp. 1127-1128

[18] L. Gross: *Hypercontractivity and logarithmic Sobolev inequalities for the Clifford-Dirichlet form*, Duke Math. J., 43, 1975, pp. 383-396

[19] U. Haagerup: *L^p spaces associated with an arbitrary von Neumann algebra*, Algèbras d'opeêrateurs et leurs applications en physique mathêmatique (Colloque CNRS, Marseille, juin 1977) Editions du CNRS, Paris, 1979, pp. 383-396.

[20] G.H. Hardy, J.E. Littlewood and G. Pólya: *Inequalities.* Cambridge University Press, Cambridge, 1934

[21] H. Kosaki: *Applications of the complex interpolation method to a von Neumann algebra: noncommutative L^p -spaces.* J. Funct. Anal., 56, 1984, no. 1, pp. 29-78.

[22] E.H. Lieb: *Some convexity and subadditivity properties of entropy*, Bull. Amer. Math. Soc. 81, 1975, pp. 1-13.

[23] E.H. Lieb: *Gaussian kernels have only Gaussian maximizers*, Invent. Math. 102, 1980, pp. 179-208

[24] E.H. Lieb and M.B. Ruskai: *Proof of the strong subadditivity of quantum-mechanical entropy*, J. Math. Phys. 14, 1973, pp. 1938-1941.

[25] E. Nelson: *Notes on non-commutative integration*, Jour. Funct. Analysis, 15, 1974, pp. 103-116

[26] M. Ohya and D. Petz, *Quantum entropy and its uses*, Text and Monographs in Physics, Springer, New York, (1993).

[27] D. Petz: *A variational expression for the relative entropy*, Comm. Math. Phys., 114, 1988, pp. 345–349

[28] G. Pisier. *Non-commutative vector valued L^p–spaces and completely p–summing maps* Astérisque, 247, Math. Soc. of France, Paris, 1998.

[29] T. Rockafellar: *Conjugate duality and optimization*, Vol. 16, *Regional conference series in applied mathematics*, SIAM, Philadelphia, 1974

[30] M.B. Ruskai, *Inequalities for quantum entropy: A review with conditions for equality*, J. Math. Phys. 43, 2005, pp. 4358-4375 (2002). Erratum *ibid* 46, pp. 019901

[31] I.E. Segal: *A non-commutative extension of abstract integration*, Annals of Math., 57, 1953, pp. 401–457

[32] I.E. Segal.: *Tensor algebras over Hilbert spaces II*, Annals of Math., 63, 1956, pp. 160–175

[33] I.E. Segal.: *Algebraic integration theory*, Bull. Am. Math. Soc.., 71, 1965, no. 3, pp. 419-489

[34] W. Stinespring: *Integration theorems for gages and duality for unimodular groups*, Trans. Am. Math. Soc., ., 90, 1958, pp. 15-26

[35] W. Thirring: *Quantum Mechanics of Large Systems, Volume IV of A Course in Mathematical Physics* Springer, New York (1983)

[36] C. Thompson: *Inequality with application in statistical mechanics* J. Math. Phys. 6, 1812–1813 (1965)

[37] A. Uhlmann: *Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory* Commun. Math. Phys., 54, 1977, pp. 21-32

[38] H. Umegaki: *Conditional expectation in operator algebras I*, Tohoku Math. J., 6, 1954, pp. 177-181

[39] D. Voiculescu, K. Dykema and A. Nica: *Free random variables*, CRM Monograph Series, 1, Am. Math. Soc., Providence, R.I., 1992)

[40] W.H.. Young: *On the multiplication of successions of Fourier constants*, Proc. Royal soc. A., 97, 1912, pp. 331-339

[41] W.H.. Young: *Sur la généralisation du théoreme du Parseval*, *Comptes rendus.*, 155, 1912, pp. 30-33

Eric A. Carlen
Department of Mathematics
Hill Center
Rutgers University
110 Frelinghuysen Road
Piscataway, NJ 08854-8019
USA

Elliott H. Lieb
Departments of
Mathematics and Physics
Jadwin Hall
Princeton University
P. O. Box 708
Princeton, NJ 08544
USA