Brascamp-Lieb inequalities

Summer School[*], Kopp

September 26 - October 1, 2021

Organizers:

Christoph Thiele
University of Bonn, Germany

Pavel Zorin-Kranich
University of Bonn, Germany

[*]supported by Hausdorff Center for Mathematics, University of Bonn and CRC 1060: "The Mathematics of Emergent Effects", University of Bonn

Contents

1 Geometric Brascamp-Lieb inequalities and Gaussian Extremizers

After J.Bennett, A.Carbery, M.Christ and T.Tao [BCCT08]

A summary written by Michele Ferrante

Abstract

We introduce the notion of Brascamp-Lieb inequalities and prove that gaussians are extremizers for the geometric Brascamp-Lieb inequalities.

1.1 Definitions and Examples

In their paper [BL76], Brascamp and Lieb were interested in the study of the sharp constant for the Young Convolution inequality

$$\|f * g\|_r \leq \|f\|_p \|g\|_q, \tag{1}$$

where $1 + 1/r = 1/p + 1/q,\ 1 \leq p, q \leq \infty$. It is easy to see that, by duality, this is equivalent to

$$\left| \int f(x) g(x-y) h(y) dx dy \right| \leq \|f\|_p \|g\|_q \|h\|_t, \tag{2}$$

where $2 = 1/t + 1/p + 1/q$. They observed this was a particular case of a more general class of multilinear inequalities which we call nowadays Brascamp-Lieb inequalities.

Definition 1. *We define an Euclidean space H to be a real Hilbert space of finite dimension, endowed with the usual Lebesgue measure. If $m \geq 0$ is an integer, we define an m-transformation $\mathbf{B}$ to be a triple*

$$\mathbf{B} := (H, (H_j)_{1 \leq j \leq m}, (B_j)_{1 \leq j \leq m})$$

where $H, H_1, \dots, H_m$ are Euclidean spaces and, for each j, $B_j : H \to H_j$ is a linear transformation. We define a Brascamp-Lieb datum to be a couple $(\mathbf{B}, \mathbf{p})$, where $\mathbf{B}$ is an m-transformation and $\mathbf{p}$ is an m-tuple $(p_j)_{1 \leq j \leq m} \in \mathbb{R}_+^m$. For each Brascamp-Lieb datum $(\mathbf{B}, \mathbf{p})$, we can consider the m-linear Brascamp-Lieb inequality

$$\int_H \prod_{j=1}^{m} (f_j \circ B_j)^{p_j} \leq \mathrm{BL}(\mathbf{B}, \mathbf{p}) \prod_{j=1}^{m} (\int_{H_j} f_j)^{p_j}, \tag{3}$$

where $f_j : H_j \to \mathbb{R}_+$ are non-negative measurable functions and $\mathrm{BL}(\mathbf{B}, \mathbf{p}) \in (0, +\infty]$ *is the best constant for which the inequality holds.*

The Young inequality (2) is a Brascamp-Lieb inequality for the m-transformation

$$\mathbf{B} = (\mathbb{R}^d \times \mathbb{R}^d, (\mathbb{R}^d)_{1\le j\le 3}, (B_j)_{1\le j\le 3}),$$

where

$$B_1(x,y) = x, \quad B_2(x,y) = y, \quad B_3(x,y) = x-y.$$

In particular we know that $\mathrm{BL}(\mathbf{B},\mathbf{p})$ is finite if and only if $p_1+p_2+p_3 = 2,\ 0 \le p_1,p_2,p_3 \le 1$, and, in these cases, it equals to

$$\mathrm{BL}(\mathbf{B},\mathbf{p}) = \left(\prod_{j=1}^{3} \frac{(1-p_j)^{1-p_j}}{p_j^{p_j}}\right)^{d/2}.$$

The Hölder inequality is also a particular case of Brascamp-Lieb inequality, which is obtained with

$$\mathbf{B} = (H, (H)_{1\le j\le m}, (\mathrm{Id}_H)_{1\le j\le m}).$$

In this case $\mathrm{BL}(\mathbf{B},\mathbf{p}) = 1$ if $\sum_{j=1}^m p_j = 1,\ 0 \le p_j \le 1$ for each j, and is infinite otherwise.

Another classical example is the Loomis-Whitney inequality. Here we take

$$\mathbf{B} = (\mathbb{R}^n, (e_j^\perp)_{1\le j\le n}, (P_j)_{1\le j\le n})$$

where $e_1,\dots,e_n$ is the standard basis of $\mathbb{R}^n$, $e_j^\perp \subset \mathbb{R}^n$ is the orthogonal complement of e_j, and P_j is the orthogonal projection onto $e_j^\perp$. The Loomis-Whitney inequality states that $\mathrm{BL}(\mathbf{B},\mathbf{p}) = 1$ if $\mathbf{p} = (\frac{1}{n-1},\dots,\frac{1}{n-1})$, and is infinite otherwise.

Both Hölder inequality and Loomis-Whitney inequality are in fact cases of what are usually called geometric Brascamp-Lieb inequalities.

Definition 2. *We say that a Brascamp-Lieb datum $(\mathbf{B},\mathbf{p})$ is a geometric Brascamp-Lieb datum if*

$$B_j B_j^* = \mathrm{Id}_{H_j}, \tag{4}$$

for every j, and

$$\sum_{j=1}^m p_j B_j^* B_j = \mathrm{Id}_H. \tag{5}$$

Since (4) implies that B_j^* is an isometry, it is easy to see that this is equivalent to ask that each H_j is a subspace of H and the B_j are the projection maps from H to H_j.

It can be proved that the Young inequality is equivalent to a geometric Brascamp-Lieb inequality, in the sense that there exist some linear changes of coordinates that make $(\mathbf{B}, \mathbf{p})$ into a geometric Brascamp-Lieb datum.

1.2 Gaussian Extremizers via Heat Flow Method

We will now prove a sharp result for geometric Brascamp-Lieb inequalities and, in particular, that gaussians are extremizers. The main idea is that, to prove $A \leq B$, we want to generate a quantity $Q(t)$ which is monotone and

$$\lim_{t \to -\infty} Q(t) = A, \quad \lim_{t \to +\infty} Q(t) = B.$$

To do so, we want to use a property of supersolutions of the transport equation.

Proposition 3. *Let $I \subseteq \mathbb{R}$, let H be a Euclidean space, let $u : I \times H \to \mathbb{R}_+$ be a smooth non-negative function, and $\vec{v} : I \times H \to H$ be a smooth vector field, such that $\vec{v}u(t, x)$ is rapidly decreasing for $x \to \infty$ locally uniformly in t. Suppose that we have the transport inequality*

$$\partial_t u(t, x) + \nabla(\vec{v}(t, x)u(t, x)) \geq 0 \tag{6}$$

for all $(t, x) \in I \times H$, where ∇ is the divergence on the Euclidean space H. Then the quantity

$$Q(t) := \int_H u(t, x)\ dx \in [0, +\infty]$$

is non-decreasing.

Proof. Let $t_1 < t_2$. From Stokes' theorem we can write

$$\begin{aligned}\int_H u(t_2, x)\psi(x)\ dx &- \int_H u(t_1, x)\psi(x)\ dx \\ &= \int_{t_1}^{t_2} \int_H (\partial_t u(t, x)\psi(x) + \nabla(\psi(x)\vec{v}(t, x)u(t, x))\ dxdt\end{aligned}$$

for any non-negative smooth cutoff function ψ. Using the product rule and (6) we conclude

$$\int_H u(t_2, x)\psi(x)\ dx - \int_H u(t_1, x)\psi(x)\ dx \geq \int_{t_1}^{t_2} \int_H \langle \nabla\psi(x), \vec{v}(t, x)u(t, x)\rangle\ dxdt.$$

Letting ψ approach the constant function 1 and since $\vec{v}u$ is rapidly decreasing, uniformly in $[t_1, t_2]$, the right-hand side tends to 0 and we obtain the claim. $\square$

A multilinear version of the previous result is the following.

Proposition 4. *Let $p_1, \dots, p_m > 0$, let H be a Euclidean space, and for each $1 \le j \le m$ let $u_j : \mathbb{R}_+ \times H \to \mathbb{R}_+$ be a smooth strictly positive function, and $\vec{v}_j : \mathbb{R}_+ \times H \to H$ be a smooth vector field. Suppose we have a smooth vector field $\vec{v} : \mathbb{R}_+ \times H \to H$, such that $\vec{v}\prod_{j=1}^m u_j^{p_j}(t,x)$ is rapidly decreasing for $x \to \infty$ locally uniformly in t, and we have the inequalities*

$$\partial_t u_j(t,x) + \nabla(\vec{v}_j u_j(t,x)) \ge 0 \text{ for all } 1 \le j \le m \tag{7}$$

$$\nabla(\vec{v} - \sum_{j=1}^m p_j \vec{v}_j) \ge 0 \tag{8}$$

$$\sum_{j=1}^m p_j \langle \vec{v} - \vec{v}_j, \nabla \log u_j \rangle_H \ge 0. \tag{9}$$

Then the quantity

$$Q(t) := \int_H \prod_{j=1}^m u_j(t,x)^{p_j} \; dx \tag{10}$$

is non-decreasing.

The idea of the proof is to apply Proposition 3 by proving

$$\partial_t \prod_{j=1}^m u_j^{p_j} + \nabla(\vec{v} \prod_{j=1}^m u_j^{p_j}) \ge 0.$$

Theorem 5. *Let $(\mathbf{B}, \mathbf{p})$ be a geometric Brascamp-Lieb datum. Then* $\mathrm{BL}(\mathbf{B}, \mathbf{p}) = 1$. *Moreover* $\mathrm{BL}(\mathbf{B}, \mathbf{p})$ *is achieved by a gaussian.*

Sketch of Proof. Assume, without loss of generality, that $H_j \le H$ and that B_j are the orthogonal projections from H to H_j. By choosing $f_j = \exp(-\pi\|x\|_{H_j}^2)$, we see that we must have $\mathrm{BL}(\mathbf{B}, \mathbf{p}) \ge 1$. So we only need to prove $\mathrm{BL}(\mathbf{B}, \mathbf{p}) \le 1$ to obtain our claim.

It thus suffices to show that, for any $f_j : H_j \to \mathbb{R}_+$, we have

$$\int_H \prod_{j=1}^m (f_j \circ B_j)^{p_j} \le \prod_{j=1}^m (\int_{H_j} f_j)^{p_j}. \tag{11}$$

Now let $u_j : \mathbb{R}_+ \times H \to \mathbb{R}_+$ be the solution to the heat equation Cauchy problem

$$\begin{aligned} \partial_t u_j(t,x) &= \Delta_H u_j(t,x) \\ u_j(0,x) &= f_j \circ B_j(x) \end{aligned}$$

where Δ_H is the Laplacian on H.

In order to apply Proposition 4, we rewrite the heat equation as a transport equation

$$\partial_t u_j + \nabla(\vec{v}_j u_j) = 0$$

where $\vec{v}_j := -\nabla \log u_j$; thus (7) is trivially satisfied. Next we set

$$\vec{v} := \sum_{j=1}^{m} p_j \vec{v}_j$$

so that (8) is also trivially satisfied.

One has to verify (9) and the technical condition that $\vec{v} \prod_{j=1}^{m} u_j^{p_j}$ is rapidly decreasing in space.

By invoking Proposition 4 we conclude that the quantity

$$Q(t) := \int_H \prod_{j=1}^{m} u_j^{p_j}(t,x)\ dx$$

is non-decreasing for $0 < t < \infty$. From Fatou's lemma we have

$$\int_H \prod_{j=1}^{m} (f_j \circ B_j)^{p_j} \leq \limsup_{t \to 0^+} Q(t).$$

The result follows by showing that

$$\liminf_{t \to \infty} Q(t) \leq \prod_{j=1}^{m} \left(\int_{H_j} f_j \right)^{p_j}.$$

□

References

[BCCT08] J. Bennett, A. Carbery, M. Christ, and T. Tao, *The Brascamp-Lieb inequalities: finiteness, structure and extremals*, Geom. Funct. Anal. 17.5 (2008), pp. 13431415.

[BL76] H. J. Brascamp and E. H. Lieb, *Best constants in Young's inequality, its converse, and its generalization to more than three functions*, Adv. Math. 20 (1976), pp. 151-173.

Michele Ferrante, University of Birmingham
email: m.ferrante@bham.ac.uk

2 On Gaussian Brunn-Minkowski inequalities

After F. Barthe and N. Huet [BH]

A summary written by Marco Fraccaroli

Abstract

We prove versions of the Brunn-Minkowski (BM) and the Prékopa-Leindler (PL) inequalities in the case of a Gaussian measure, as well as the equivalence between them, by means of an evolution lemma. We use the same lemma to recover the Brascamp-Lieb (BL) inequality result of [BCCT], as well as the reverse Brascamp-Lieb one.

2.1 Lebesgue BM and PL inequalities

Let $n \geq 1$. For the Lebesgue measure μ_n on $\mathbb{R}^n$, we have the BM inequality.

Theorem 1 (Lebesgue BM)**.** *For all Borel sets $A, B \subseteq \mathbb{R}^n$, we have*

$$\mu_n(A+B)^{\frac{1}{n}} \geq \mu_n(A)^{\frac{1}{n}} + \mu_n(B)^{\frac{1}{n}},$$

where $A+B$ is the Minkowski sum of the two sets

$$A + B \coloneqq \{a + b \colon a \in A, b \in B\}.$$

An equivalent reformulation is the following one.

Theorem 2 (Logarithmic Lebesgue BM)**.** *For every $\lambda \in [0,1]$, for all Borel sets $A, B \subseteq \mathbb{R}^n$, we have*

$$\mu_n(\lambda A + (1-\lambda)B) \geq \mu_n(A)^{\lambda}\mu_n(B)^{1-\lambda}.$$

In fact, the BM inequality is equivalent also to its a priori stronger functional version, the PL inequality.

Theorem 3 (Lebesgue PL)**.** *Let $\lambda \in [0,1]$. For all Borel functions $f, g, h \colon \mathbb{R}^n \to [0,\infty)$ if*

$$\forall y, z \in \mathbb{R}^n, \quad h(\lambda y + (1-\lambda)z) \geq f(y)^{\lambda} g(z)^{1-\lambda},$$

then

$$\int_{\mathbb{R}^n} h \,\mathrm{d}\mu_n \geq \Big(\int_{\mathbb{R}^n} f \,\mathrm{d}\mu_n\Big)^{\lambda}\Big(\int_{\mathbb{R}^n} g \,\mathrm{d}\mu_n\Big)^{1-\lambda}.$$

Setting

$$H = h, \quad F = f^{\lambda}, \quad G = g^{1-\lambda}, \quad \lambda = p^{-1}, \quad \text{for } p \in [1,\infty],$$

the previous inequality can be interpreted as a reverse Hölder's inequality, the prototype of the reverse BL one.

2.2 Heat flow

Let γ_n be the Gaussian measure associated with the density

$$\mathrm{d}\gamma_n(x) = (2\pi)^{-\frac{n}{2}} e^{-\frac{|x|^2}{2}} \,\mathrm{d}x,$$

and let $\Phi\colon \mathbb{R} \to [0,1]$ be the cumulative distribution associated with γ_1

$$\Phi(x) = \int_{-\infty}^{x} \mathrm{d}\gamma_1(y).$$

For every non-negative Borel function f on $\mathbb{R}^n$, let $P_t f$ be its *heat flow* defined on $[0,\infty) \times \mathbb{R}^n$ by

$$P_t f(x) = f * D^1_{\sqrt{t}}\gamma_n(x) = \int_{\mathbb{R}^n} f(x-y)(2\pi t)^{-\frac{n}{2}} e^{-\frac{|y|^2}{2t}} \,\mathrm{d}y.$$

Since the rescaled Gaussian $D^1_{\sqrt{t}}\gamma_n$ satisfies the heat equation

$$2\partial_t u = \Delta_x u,$$

then, for f smooth enough, integrating by parts, $P_t f$ does too. Moreover, $P_t f$ reproduces both the pointwise evaluation of f and its integral

$$\forall x \in \mathbb{R}^n, \qquad \begin{cases} P_0 f(x) = f(x), \\ \sqrt{t}^n P_t f(x) \to C \displaystyle\int_{\mathbb{R}^n} f \,\mathrm{d}\mu_n, \quad t \to \infty. \end{cases}$$

2.3 Gaussian BM and PL inequalities

As in the Lebesgue case, for the Gaussian measure γ_n on $\mathbb{R}^n$, we have a BM and a PL inequality, and they are equivalent.

Theorem 4 ([BH], Thm 1). *Let $m \geq 1$, $\alpha_i \in [0,\infty)$, $i \in \{1,\dots,m\}$. TFAE*

1. $\displaystyle\sum_{i=1}^{m} \alpha_i \geq 1$, *and* $\forall j \notin I_{conv}$, $\displaystyle\alpha_j - \sum_{i\neq j} \alpha_i \leq 1$.
2. *(Gaussian BM) For all Borel sets $A_i \subseteq \mathbb{R}^n$ such that A_i is convex when $i \in I_{conv}$, we have*

$$\Phi^{-1} \circ \gamma_n\Big(\sum_{i=1}^{m} \alpha_i A_i\Big) \geq \sum_{i=1}^{m} \alpha_i \Phi^{-1} \circ \gamma_n(A_i).$$

3. *(Gaussian PL) For all Borel functions $h, f_i \colon \mathbb{R}^n \to [0,1]$ such that $\Phi^{-1} \circ f_i$ is concave when $i \in I_{conv}$, if*

$$\forall x_1, \ldots, x_m \in \mathbb{R}^n, \quad \Phi^{-1} \circ h\Big(\sum_{i=1}^m \alpha_i x_i\Big) \geq \sum_{i=1}^m \alpha_i \Phi^{-1} \circ f_i(x_i),$$

then

$$\Phi^{-1}\Big(\int_{\mathbb{R}^n} h \, \mathrm{d}\gamma_n\Big) \geq \sum_{i=1}^m \alpha_i \Phi^{-1}\Big(\int_{\mathbb{R}^n} f_i \, \mathrm{d}\gamma_n\Big).$$

4. *(HF Gaussian PL) For all Borel functions $h, f_i \colon \mathbb{R}^n \to [0,1]$, $i \in \{1, \ldots, m\}$, such that $\Phi^{-1} \circ f_i$ is concave when $i \in I_{conv}$, if*

$$\forall x_1, \ldots, x_m \in \mathbb{R}^n, \quad \Phi^{-1} \circ h\Big(\sum_{i=1}^m \alpha_i x_i\Big) \geq \sum_{i=1}^m \alpha_i \Phi^{-1} \circ f_i(x_i),$$

then, for every $t \in [0, \infty)$,

$$\forall x_1, \ldots, x_m \in \mathbb{R}^n, \quad \Phi^{-1} \circ P_t h\Big(\sum_{i=1}^m \alpha_i x_i\Big) \geq \sum_{i=1}^m \alpha_i \Phi^{-1} \circ P_t f_i(x_i).$$

It is easy to see that

$$(4) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1).$$

In particular, in the last case it is enough to consider sets A_i given by either balls with centre in the origin and appropriate radii or their complements. In [BH], to complete the proof of the Theorem, the authors show the following chain of implications

$$(1) \Rightarrow (4) \text{ for ``smooth'' functions} \Rightarrow (2),$$
$$(2) \text{ in } \mathbb{R}^{n+1} \Rightarrow (4) \text{ for arbitrary functions on } \mathbb{R}^n.$$

"Smooth" functions are twice continuously differentiable functions h, f_i such that for $f = h$ or $f = f_i$, we have

$$\forall t > 0, \forall x \in \mathbb{R}^n, \quad |\nabla f(x + \sqrt{t}y)| e^{-\frac{|y|^2}{2}} \to 0 \quad \text{for } |y| \to \infty.$$

Moreover, they satisfy another technical condition guaranteeing that C defined below satisfies the second condition in Lemma 5. Such functions provide arbitrarily close approximations of characteristic functions of Borel sets.

For all such Borel functions $h, f_i\colon \mathbb{R}^n \to [0,1]$, $i \in \{1,\dots,m\}$, we define the auxiliary function C on $[0,\infty)\times\mathbb{R}^{nm}$ by

$$C(t,x) = C(t,x_1,\dots,x_m) = H_t\Big(\sum_{i=1}^m \alpha_i x_i\Big) - \sum_{i=1}^m \alpha_i F_{i,t}(x_i),$$

where, for all Borel functions $f\colon \mathbb{R}^n \to [0,1]$, we set

$$F_t(x) = \Phi^{-1}\circ P_t f(x).$$

The auxiliary function C satisfies the following properties

$$C(0,x) \geq 0, \quad \forall x \in \mathbb{R}^{nm},$$
$$2\partial_t C = \Big(\Delta H_t - \sum_{i=1}^m \alpha_i \Delta F_{i,t}\Big) + \Big(-H_t|\nabla H_t|^2 + \sum_{i=1}^m \alpha_i F_{i,t}|\nabla F_{i,t}|^2\Big).$$

Therefore, we obtain the desired inequality, at least for "smooth" functions, via the following evolution lemma.

Lemma 5. *Let $C\colon [0,\infty)\times\mathbb{R}^{nm}\to\mathbb{R}$ be twice differentiable such that*

- *(Evolution conditions) For every $(t,x)\in[0,\infty)\times\mathbb{R}^{nm}$, we have*

$$\begin{cases} \mathrm{Hess}_x(C)(t,x) \geq 0 \\ \nabla_x C(t,x) = 0 \\ C(t,x) \leq 0 \end{cases} \qquad \Rightarrow \partial_t C(t,x) \geq 0$$

- *For some $T>0$, we have*

$$\liminf_{|x|\to\infty}\Big(\inf_{t\in[0,T]} C(t,x)\Big) \geq 0.$$

- *(Initial condition) For every $x\in\mathbb{R}^{nm}$, we have $C(0,x)\geq 0$.*

Then, for all $t\in[0,T]$, $x\in\mathbb{R}^{nm}$, we have

$$C(t,x) \geq 0.$$

We extend the result to arbitrary functions upon the following observations. The integral of a function f on $\mathbb{R}^n$ with respect to the measure γ_n can be represented as the γ_{n+1}-measure of a certain set in $\mathbb{R}^{n+1}$. In particular, the association between functions and sets is well-behaved with respect to linear combinations.

Finally, we observe that a version of Theorem 4 above with the function Φ^{-1} replaced by the logarithm follows from the reverse BL statement in Theorem 6 below for an appropriate choice of the parameters.

2.4 BL and reverse BL inequalities via evolution lemma

Theorem 6 ([BH], Thm 4 & 5)**.** *Let $m, N \geq 1$. Let $n_i \geq 1$, $B_i\colon \mathbb{R}^N \to \mathbb{R}^{n_i}$ linear map, $\alpha_i \in [0,\infty)$, $i \in \{1,\dots,m\}$. TFAE*

1. $\displaystyle\sum_{i=1}^m \alpha_i B_i^* B_i = \mathrm{Id}_N$, *and* $B_i B_i^* = \mathrm{Id}_{n_i}$, $\forall i \in \{1,\dots,m\}$.

2. *(HF BL) For all Borel functions $h\colon \mathbb{R}^N \to [0,\infty)$, $f_i\colon \mathbb{R}^{n_i} \to [0,\infty)$ if*

$$\forall x \in \mathbb{R}^N, \quad \log(h(x)) \leq \sum_{i=1}^m \alpha_i \log(f_i(B_i x)),$$

   then, for every $t \in [0,\infty)$,

$$\forall x \in \mathbb{R}^N, \quad \log(P_t h(x)) \leq \sum_{i=1}^m \alpha_i \log(P_t f_i(B_i x)).$$

3. *(HF reverse BL) For all Borel functions $h\colon \mathbb{R}^N \to [0,\infty)$, $f_i\colon \mathbb{R}^{n_i} \to [0,\infty)$ if*

$$\forall x_i \in \mathbb{R}^{n_i}, \quad \log\Big(h\Big(\sum_{i=1}^m \alpha_i B_i^* x_i\Big)\Big) \geq \sum_{i=1}^m \alpha_i \log(f_i(x_i)),$$

   then, for every $t \in [0,\infty)$,

$$\forall x_i \in \mathbb{R}^{n_i}, \quad \log\Big(P_t h\Big(\sum_{i=1}^m \alpha_i B_i^* x_i\Big)\Big) \geq \sum_{i=1}^m \alpha_i \log(P_t f_i(x_i)).$$

Taking $t \to \infty$ and $x = 0$ in the second statement of Theorem 6, we recover the BL inequality

$$\int_{\mathbb{R}^N} \prod_{i=1}^m f_i(B_i x)^{\alpha_i}\, \mathrm{d}\mu_N(x) \leq \prod_{i=1}^m \Big(\int_{\mathbb{R}^{n_i}} f_i(x_i)\, \mathrm{d}\mu_{n_i}(x_i)\Big)^{\alpha_i}.$$

Taking $t \to \infty$ and $x = 0$ in the third statement of Theorem 6, we recover the reverse BL inequality

$$\int_{\mathbb{R}^N} \sup\Big\{\prod_{i=1}^m f_i(x_i)^{\alpha_i} \colon x_i \in \mathbb{R}^{n_i}, \sum_{i=1}^m \alpha_i B_i^* x_i = x\Big\}\, \mathrm{d}\mu_N(x) \geq$$
$$\geq \prod_{i=1}^m \Big(\int_{\mathbb{R}^{n_i}} f_i(x_i)\, \mathrm{d}\mu_{n_i}(x_i)\Big)^{\alpha_i}.$$

The proof follows via the evolution lemma as before. For example in the case of reverse BL inequality, for every Borel function $f\colon \mathbb{R}^n \to [0,1]$, we set

$$F_t(x) = \log(P_t f(x)),$$

and we define the auxiliary function C on $[0,\infty) \times \mathbb{R}^{nm}$ by

$$C(t,x) = C(t,x_1,\ldots,x_m) = H_t\Big(\sum_{i=1}^m \alpha_i B_i^* x_i\Big) - \sum_{i=1}^m \alpha_i F_{i,t}(x_i).$$

It is worth noting that in both case the auxiliary function C satisfies stronger properties, namely for every $(t,x) \in [0,\infty)$, we have

$$\begin{cases} \mathrm{Hess}_x(C)(t,x) \geq 0 \\ \nabla_x C(t,x) = 0 \end{cases} \qquad \Rightarrow \partial_t C(t,x) \geq 0$$

We conclude with a comparison between the proof of BL inequality by the evolution lemmata appearing in [BH] and [BCCT].
Let B_i be linear maps, α_i be coefficients satisfying the algebraic conditions stated in Theorem 6. Let $f_i\colon \mathbb{R}^{n_i} \to [0,\infty)$ be Borel functions.
In [BH], the authors prove that the quantity

$$\prod_{i=1}^m P_t f_i(B_i x)^{\alpha_i} - P_t\Big(\prod_{i=1}^m f_i(B_i\,\cdot\,)^{\alpha_i}\Big)(x),$$

is non-negative on $[0,\infty)\times \in \mathbb{R}^N$. In [BCCT], the authors prove that the quantity

$$Q(t) = \int_{\mathbb{R}^N} \prod_{i=1}^m P_t f_i(B_i x)^{\alpha_i}\, \mathrm{d}\mu_N(x),$$

is non-decreasing in $t \in [0,\infty)$.

References

[BCCT] J. Bennett, A. Carbery, M. Christ, and T. Tao, *The Brascamp-Lieb inequalities: finiteness, structure and extremals.* Geom. Funct. Anal. 17.5 (2008), pp. 13431415.

[BH] F. Barthe and N. Huet, *On Gaussian Brunn-Minkowski inequalities.* Studia Math. 191.3 (2009), pp. 283304.

Marco Fraccaroli, Universität Bonn
email: `mfraccar@math.uni-bonn.de`

3 Inequalities in the Euclidean and Gaussian settings

After R. J. Gardner and R. Latała

A summary written by Georgios Dosidis

Abstract

We obtain the Prékopa-Leindler and Shannon-Stam inequalities as limits of the sharp Young inequality. We also deduce Brunn–Minkowski and isoperimetric inequalities. Moreover, we prove the Gaussian isoperimetric, Bobkov, and Gross log-Sobolev inequalities as limits of Ehrhards inequality. The aim is to expose relations between the various inequalities and to compare the Euclidean and the Gaussian cases.

3.1 The Euclidean case

3.1.1 The isoperimetric and Brunn-Minkowski inequalities

The isoperimetric inequality states that amongst appropriately well behaved sets of a given volume, balls minimize the surface area. Minkowskis definition of the surface area $S(M)$ of a suitable set M in $\mathbb{R}^n$ is

$$S(M) = \lim_{\varepsilon \to 0+} \frac{V(M+\varepsilon B) - V(M)}{\varepsilon}.$$

Here $V(X)$ is the n-dimensional (Lebesgue) volume of X, $B = B^n$ is the ball in $\mathbb{R}^n$, the sum $X + Y = \{x + y : x \in X, y \in Y\}$ stands for the Minkowski sum and $rX = \{rx : x \in X\}$ is a dilation. The set $M + \varepsilon B$ is the ε-enlargement of M. We will use this definition for the surface area when M is a convex body (compact convex set with nonempty interior) or a compact domain with piecewise C^1 boundary.

Theorem 1 (Isoperimetric inequality for convex bodies in $\mathbb{R}^n$)**.** *Let K be a convex body in $\mathbb{R}^n$. Then*

$$\left(\frac{V(K)}{V(B)}\right)^{1/n} \leq \left(\frac{S(K)}{S(B)}\right)^{1/(n-1)}, \tag{1}$$

with equality if and only if K is a ball.

About a century ago, not long after the first complete proof of the classical isoperimetric inequality was found, Minkowski proved the following inequality.

Theorem 2 (Brunn-Minkowski, standard form)**.** *Let K and L be convex bodies in $\mathbb{R}^n$, $0 < \lambda < 1$. Then*

$$V\big((1-\lambda)K + \lambda L\big)^{1/n} \geq (1-\lambda)V(K)^{1/n} + \lambda V(L)^{1/n}. \tag{2}$$

Equality holds in (2) *if and only if K and L are homothetic (i.e., equal up to translation and dilatation).*

More generally, for (2) to hold it is enough to assume that K, L are bounded, nonempty measurable sets such that $(1-\lambda)K + \lambda L$ is also measurable for all λ. The inequality (2) had been proved for $n = 3$ earlier by Brunn, and now it is known as the Brunn-Minkowski. Using the homogeneity of the volume, we see that for all $t, s > 0$,

$$V\big(sK + tL\big)^{1/n} \geq sV(K)^{1/n} + tV(L)^{1/n}.$$

The isoperimetric inequality (Theorem 1) follows from Brunn-Minkowski.

Proof of Theorem 1. From the Brunn-Minkowski we have that

$$V(K + \varepsilon B) \geq \big(V(K)^{1/n} + \varepsilon V(B)^{1/n}\big)^n \geq V(K)\left(1 + n\varepsilon\left(\frac{V(B)}{V(K)}\right)^{1/n}\right)$$

since $(1+x)^n \geq 1 + nx$ for all $x \geq 0$. Thus

$$S(K) = \lim_{\varepsilon \to 0+} \frac{V(K+\varepsilon B) - V(K)}{\varepsilon} \geq nV(K)\left(\frac{V(B)}{V(K)}\right)^{1/n},$$

which, along with the familiar identity $S(B) = nV(B)$ for the ball yields

$$\frac{S(K)}{S(B)} = \frac{S(K)}{nV(B)} \geq \frac{V(K)\left(\frac{V(B)}{V(K)}\right)^{1/n}}{V(B)}.$$

Rearranging we obtain (1). If K is not homothetic to a ball, the Brunn-Minkowski is a strict inequality, which yields the equality condition. $\square$

3.1.2 The Prékopa-Leindler inequality

The Brunn-Minkowski inequality can be stated in the following form: For K, L convex bodies and $0 < \lambda < 1$,

$$V\big((1-\lambda)K + \lambda L\big) \geq V(K)^{1-\lambda} V(L)^{\lambda}. \tag{3}$$

The two forms are equivalent. Indeed, the standard form implies the multiplicative form via the weighted arithmetic-geometric means inequality (AM-GM), whereas the inverse is also immediate using standard techniques (see [Ga, Corollary 5.3]).

The Prékopa-Leindler inequality is a functional generalization of this form of Brunn-Minkowski.

Theorem 3 (The Prékopa-Leindler inequality)**.** *Let $0 < \lambda < 1$ and let f, g, and h be nonnegative integrable functions on $\mathbb{R}^n$ satisfying*

$$h\left((1-\lambda)x + \lambda y\right) \geq f(x)^{1-\lambda} g(y)^{\lambda}, \tag{4}$$

for all $x, y \in \mathbb{R}^n$. Then

$$\int_{\mathbb{R}^n} h(x)dx \geq \left(\int_{\mathbb{R}^n} f(x)dx\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g(x)dx\right)^{\lambda}. \tag{5}$$

Proof of Brunn-Minkowski. Set $h = 1_{(1-\lambda)K+\lambda L}$, $f = 1_K$ and $g = 1_L$. If $x, y \in R^n$, then $f(x)^{1-\lambda} g(y)^{\lambda} > 0$ (and in fact equals 1) if and only if $x \in X$ and $y \in Y$. The latter implies $(1-\lambda)x + \lambda y \in (1-\lambda)X + \lambda Y$, which is true if and only if $h((1-\lambda)x + \lambda y) = 1$. Therefore (4) holds. Thus

$$\begin{aligned} V((1-\lambda)X + \lambda Y) &= \int_{\mathbb{R}^n} h(x)dx \\ &\geq \left(\int_{\mathbb{R}^n} f(x)dx\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g(x)dx\right)^{\lambda} = V(K)^{1-\lambda} V(L)^{\lambda}. \end{aligned}$$

$\square$

3.1.3 Young's inequality

Theorem 4 (Young's inequality)**.** *Let $0 < p, q, r$ satisfying $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, and let $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$ be nonnegative. Then*

$$\textit{(Young's inequality)} \qquad \|f * g\|_r \leq \|f\|_p \|g\|_q, \qquad \textit{for } p, q, r \geq 1 \tag{6}$$

and

$$\textit{(Reverse Young's inequality)} \quad \|f * g\|_r \geq \|f\|_p \|g\|_q, \qquad \textit{for } p, q, r \leq 1. \tag{7}$$

Here $C = C_pC_q/C_r$, where

$$C_s^2 = \frac{|s|^{1/s}}{|s'|^{1/s'}}$$

for $1/s + 1/s' = 1$ (that is, s and s' are Hölder conjugates).

The reverse Young implies the following (stronger) form of the Prékopa-Leindler.

Theorem 5 (Prékopa-Leindler inequality, essential form)**.** *Let $0 < \lambda < 1$ and let $f, g \in L^1(\mathbb{R}^n)$ be nonnegative. Let*

$$s(x) = \operatorname{ess\,sup}_y f\left(\frac{x-y}{1-\lambda}\right)^{1-\lambda} g\left(\frac{y}{\lambda}\right)^{\lambda}.$$

Then s is measurable and

$$\|s\|_1 \geq \|f\|_1^{1-\lambda}\|g\|^{\lambda}. \tag{8}$$

To see that Theorem 5 implies the usual form note that if h is any integrable function satisfying

$$h((1-\lambda)x + \lambda y) \geq f(x)^{1-\lambda} g(y)^{\lambda},$$

we must have $h \geq s$ almost everywhere and thus $\|h\|_1 \geq \|s\|_1 \geq \|f\|_1^{1-\lambda}\|g\|^{\lambda}$.

We will prove the essential form of the Prékopa-Leindler using the limiting case $r \to 0$ of the reverse Young's inequality, thus linking the chain of proofs all the way to the isopemetric inequality.

Proof of Theorem 5. Using a standard limiting argument, it suffices to prove the theorem when f and g are bounded measurable functions with compact support. Assuming this, note that $s(x) = \lim_{m\to\infty} S_m(x)$, where

$$S_m(x) = \left(\int_{\mathbb{R}^n} f\left(\frac{x-y}{(1-\lambda)m}\right)^{1-\lambda} g\left(\frac{y}{\lambda}\right)^{\lambda m}\right)^{1/(m-1)}.$$

Note that $\|s\|_1 = \lim_{m\to\infty}\|S_m\|_1$ because the S_m's are uniformly bounded with compact supports.

Applying the reverse Young inequality to S_m with $m > \max\{(1-\lambda)^{-1}, \lambda^{-1}\}$, $p = 1/((1-\lambda)m), q = 1/(\lambda m)$, and $r = 1/(m-1)$, we obtain

$$
\begin{aligned}
\|S_m\|_1 &= \int_{\mathbb{R}^n} \left(\left(\int_{\mathbb{R}^n} f\left(\frac{x-y}{(1-\lambda)m} \right)^{1-\lambda} g\left(\frac{y}{\lambda}\right)^{\lambda m} \right)^{1/(m-1)} dx \\
&\geq \left(C^n \left(\int_{\mathbb{R}^n} f\left(\frac{x}{1-\lambda} \right) dx \right)^{(1-\lambda)m} \left(\int_{\mathbb{R}^n} g\left(\frac{y}{\lambda}\right) dy \right)^{\lambda m} \right)^{1/(m-1)} \\
&\to \|f\|_1^{1-\lambda} \|g\|_1^{\lambda}.
\end{aligned}
$$

as $m \to \infty$, since $\lim_{m\to\infty} C^{n/(m-1)} = (1-\lambda)^{-(1-\lambda)}\lambda^{-\lambda}$. $\square$

3.1.4 The Shannon–Stam inequality

Suppose that X is a discrete random variable taking possible values $x_1, \ldots, x_m$ with probabilities $p_1, \ldots, p_m$, respectively, where $\sum_{i=1}^m \frac{1}{p_i} = 1$. Shannon introduced a measure of the average uncertainty removed by revealing the value of X. This quantity,

$$
H_m(p_1, \ldots, p_m) = -\sum_{i=1}^m p_i \log p_i,
$$

is called the entropy of X. It can also be regarded as a measure of the missing information; indeed, the function H_m is concave and achieves its maximum when $p_1 = \cdots = p_m = \frac{1}{m}$, that is, when all outcomes are equally likely.

If X is a random vector in $\mathbb{R}^n$ with probability density f, the entropy $h_1(X)$ of X is defined analogously:

$$
h_1(X) = h_1(f) = -\int_{\mathbb{R}^n} f(x) \log f(x) dx.
$$

The entropy power $N(X)$ of X is

$$
N(X) = \frac{1}{2\pi e} \exp\left(\frac{2}{n} h_1(X)\right).
$$

Theorem 6 (Entropy power inequality)**.** *Let X and Y be independent random vectors in $\mathbb{R}^n$ with probability densities in $L^p(\mathbb{R}^n)$ for some $p > 1$. Then*

$$
N(X+Y) \geq N(X) + N(Y). \tag{9}
$$

This can be proved via the following lemma

Lemma 7. *Let f and g be nonnegative functions in $L^s(\mathbb{R}^n)$ for some $s > 1$, such that*

$$\int_{\mathbb{R}^n} f(x)dx = \int_{\mathbb{R}^n} g(x)dx = 1.$$

Then for $0 < \lambda < 1$,

$$h_1(f * g) - (1-\lambda)h_1(f) - \lambda h_1(g) \geq -\frac{n}{2}\left((1-\lambda)\log(1-\lambda) + \lambda \log \lambda\right). \quad (10)$$

The proof follows from Young's inequality (see [Ga, Lemma 18.2]) Putting $\lambda = \dfrac{N(Y)}{N(X)+N(Y)}$ and simplifying the resulting inequality leads directly to (9).

3.2 The Gaussian case

3.2.1 Gaussian isoperimetry

A Gaussian measure μ is a probability measure on $\mathbb{R}^n$ that is the affine image of the canonical Gaussian measure γ_n with probability density $d\gamma_n(x) = (2\pi)^{-n/2} exp(-|x|^2/2)dx$. Linear images of γ_n are called centered Gaussian measures. For a Banach space F, μ is called a centered Gaussian measure on F if there are $g_1, g_2, \ldots$ independent $\mathcal{N}(0,1)$ random variables (r.v.) and vectors $x_1, x_2, \ldots$ in F such that the series $X = \sum_{i=1}^{\infty} x_i g_i$ is convergent almost surely and in every L^p, $0 < p < \infty$, and is distributed as μ.

We will denote by Φ the distribution function of the standard normal $\mathcal{N}(0,1)$ r.v., that is

$$\Phi(x) = \gamma_1(-\infty, x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt, \quad -\infty \leq x \leq \infty,$$

and by $\phi(x) = \Phi'(x) = (2\pi)^{-1/2} e^{-x^2/2}$ its derivative.

We saw that in the Euclidean case, balls are the minimizers of the surface area among sets of equal volume. In the Gaussian case, amongst Borel sets of given Gaussian measure, the minimizers of the Gaussian boundary measure are the half-spaces. For a Borel set A let $A_t = A + tB$ be its t-enlargement.

Theorem 8 (Gaussian isoperimetric inequality)**.** *Let A be a Borel set in $\mathbb{R}^n$ and let H be an affine halfspace such that $\gamma_n(A) = \gamma_n(H) = \Phi(a)$ for some $a \in \mathbb{R}$. Then*

$$\gamma_n(A_t) \geq \gamma_n(H_t) = \Phi(a+t) \quad (11)$$

for all $t \geq 0$.

Let $I(t) = \phi \circ \Phi^{-1}(t)$, $t \in [0,1]$ be the Gaussian isoperimetric function. The equivalent form of Theorem 8 is that for all Borel sets A in $\mathbb{R}^n$

$$\gamma_n^+(A) := \liminf_{t \to 0^+} \frac{\gamma_n(A_t) - \gamma_n(A)}{t} \geq I(\gamma_n(A)). \tag{12}$$

The equality in (12) holds for any affine halfspace.

3.2.2 Ehrhard's inequality

Theorem 9 (Ehrhard's inequality)**.** *If μ is a centered Gaussian measure on a separable Banach space F and A, B are Borel sets in F, with at least one of them convex, then*

$$\Phi^{-1}\big(\mu(\lambda A + (1-\lambda)B)\big) \geq \lambda \Phi^{-1}(\mu(A)) + (1-\lambda)\Phi^{-1}(\mu(B)) \tag{13}$$

for $\lambda \in [0,1]$.

Theorem 9 implies the isoperimetric inequality (11). Indeed let A be a Borel set in $\mathbb{R}^n$ with $\gamma_n(A) = \Phi(a)$ for some $a \in \mathbb{R}$. Then

$$\begin{aligned}
\Phi^{-1}(\gamma_n(A_t)) &= \Phi^{-1}(\gamma_n(\lambda(\lambda^{-1}A) + (1-\lambda)((1-\lambda)^{-1}tB))) \\
&\geq \lambda\Phi^{-1}(\gamma_n(\lambda^{-1}A)) + (1-\lambda)\Phi^{-1}(\gamma_n((1-\lambda)^{-1}tB)) \\
&\to \Phi^{-1}(\gamma_n(A)) + t
\end{aligned}$$

as $\lambda \to 1^{-1}$ and thus $\gamma_n(A_t) \geq \Phi(a+t)$.

Ehrhards inequality has the following Prékopa-Leindler type functional version. Suppose that $\lambda \in (0,1)$ and $f, g, h : \mathbb{R}^n \to [0,1]$ are such that for all $x, y \in \mathbb{R}^n$

$$\Phi^{-1}(h(\lambda x + (1-\lambda)y)) \geq \lambda\Phi^{-1}(f(x)) + (1-\lambda)\Phi^{-1}(g(y)),$$

then

$$\Phi^{-1}\Big(\int_{\mathbb{R}^n} h\,d\gamma_n\Big) \geq \lambda\Phi^{-1}\Big(\int_{\mathbb{R}^n} f\,d\gamma_n\Big) + (1-\lambda)\Phi^{-1}\Big(\int_{\mathbb{R}^n} g\,d\gamma_n\Big).$$

3.2.3 Gross-Gauss and Bobkov's inequalities

Gross showed that the Gaussian measures γ_n satisfy the logarithmic Sobolev inequality

$$\int_{\mathbb{R}^n} g^2 \log g^2 d\gamma_n - \int_{\mathbb{R}^n} g^2 d\gamma_n \log\left(\int_{\mathbb{R}^n} g^2 d\gamma_n\right) \leq 2\int_{\mathbb{R}^n} |\nabla g|^2 d\gamma_n \tag{14}$$

for all smooth functions $g : \mathbb{R}^n \to \mathbb{R}$. It can be shown that (14) implies the concentration inequality

$$\gamma_n \left(h \geq \int_{\mathbb{R}^n} h d\gamma_n + t \right) \leq e^{-t^2/2}, \quad t \geq 0 \tag{15}$$

for all Lipschitz functions h.

Theorem 10 (Bobkov's inequality)**.** *For any locally Lipschitz function $f : \mathbb{R}^n \to [0,1]$ we have*

$$I \left(\int_{\mathbb{R}^n} f d\gamma_n \right) \leq \int_{\mathbb{R}^n} \sqrt{I(f)^2 + |\nabla f|^2} d\gamma_n. \tag{16}$$

Theorem 10 easily implies the isoperimetric inequality (12) by approximating the indicator function 1_A by Lipschitz functions. On the other hand if we apply (12) to the set $A = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : \Phi(y) < f(x)\}$ in $\mathbb{R}^{n+1}$ we get (16). It is also not hard to derive the logarithmic Sobolev inequality (14) as a limit case of Bobkovs inequality. One should use (16) for $f = \varepsilon g^2$ (with g bounded) and let ε tend to 0, using that $I(t) \sim t\sqrt{2\log(1/t)}$ as $t \to 0^+$.

References

[Ga] Gardner, R. J. , *The Brunn-Minkowski inequality.* Bull. Amer. Math. Soc. **39.3** (2002), no. 1, 355–405. mr: 1898210.

[La] Latała, R., *On some inequalities for Gaussian measures.* Proceedings of the International Vol. II (Beijing, 2002). Higher Ed. Press, Beijing, (2002), pp. 813–822. mr: 1957087.

Georgios Dosidis, Charles University
email: dosidis@karlin.mff.cuni.cz

4 Dimension conditions for non-geometric BL inequalities

After J. Bennett, A. Carbery, M. Christ, and T. Tao [BCCT08] and D. Maldague [M19]

A summary written by Aleksandar Bulj

Abstract

We present the dimension conditions for non-geometric Brascamp Lieb inequalities using two different approaches. The first approach, from [BCCT08], uses Gaussian extremizers for reduction to geometric Brascamp - Lieb inequalities and heat flow method for the proof of geometric Brascamp Lieb inequalities, while the second approach, form [M19] uses induction on dimension and Hölder's inequality/complex interpolation.

4.1 Introduction

We are interested in the question of determining necessary and sufficient conditions for the Brascamp - Lieb inequality to hold. Let us quickly recall the definitions we need.

Definition 1. *Let $m \geq 1$ be an integer. Let $H, H_1, \dots, H_m$ be finite dimensional real Hilbert spaces with Lebesgue measure dx, let $(B_j)_{1\leq j\leq m}$ be a m-tuple of surjective linear transformations $B_j : H \to H_j$, let $(p_j)_{1\leq j\leq m} \in \mathbb{R}_+^m$ be a m-tuple of nonnegative real numbers and let $(f_j)_{1\leq j\leq m}$ be a m-tuple of nonnegative measurable functions. We denote:*

$$\boldsymbol{B} := (H, (H_j)_{1\leq j\leq m}, (B_j)_{1\leq j\leq m}), \quad \boldsymbol{p} := (p_j)_{1\leq j\leq m}, \quad \boldsymbol{f} := (f_j)_{1\leq j\leq m}$$

and we call pair $(\boldsymbol{B}, \boldsymbol{p})$ Brascamp - Lieb datum.
We define:

$$BL(\boldsymbol{B}, \boldsymbol{p}; \boldsymbol{f}) := \frac{\int_H \prod_{j=1}^m f_j(B_j(x))^{p_j}}{\prod_{j=1}^m \left(\int_{H_j} f_j(x_j)dx_j\right)^{p_j}}.$$

Finally, we define the Brascamp - Lieb constant, $BL(\boldsymbol{B}, \boldsymbol{p})$ as:

$$BL(\boldsymbol{B}, \boldsymbol{p}) := \sup\left\{BL(\boldsymbol{B}, \boldsymbol{p}; \boldsymbol{f}) : f_j \geq 0,\, 0 < \int_{H_j} f_j < \infty\right\}. \tag{1}$$

Definition 2. *Brascamp - Lieb datum* $(\boldsymbol{B}, \boldsymbol{p})$ *is called geometric if* $B_j B_j^* = \mathrm{id}_{H_j}$ *for* $j = 1, \ldots, m$ *and:*

$$\sum_{j=1}^{m} p_j B_j^* B_j = \mathrm{id}_H .$$

Restricting the supremum in (1) to gaussian inputs $\boldsymbol{f} = (\exp(-\pi\langle A_j x, x\rangle))_{1\le j\le m}$, where $A_j : H_j \to H_j$ is positive definite transformation and explicitly calculating the expression, we arrive to the following definition.

Definition 3. *For Brascamp - Lieb datum* $(\boldsymbol{B}, \boldsymbol{p})$ *we define* $BL_g(\boldsymbol{B}, \boldsymbol{p})$ *as:*

$$BL_g(\boldsymbol{B}, \boldsymbol{p}) = \sup \left\{ \left(\frac{\prod_{j=1}^{m} (\det_{H_j} A_j)^{p_j}}{\det_H(\sum_{j=1}^{m} p_j B_j^* A_j B_j)} \right)^{\frac{1}{2}} : A_j > 0 \right\}, \tag{2}$$

where $A_j > 0$ *means that* $A_j : H_j \to H_j$ *is positive definite transformation.*

We say that the $(\boldsymbol{B}, \boldsymbol{p})$ *is gaussian-extremisable if there exists* m*-tuple of positive definite matrices* $(A_j)_{1\le j\le m}$ *for which the supremum is attained in* (2)*.*

It is obvious from the definition that $BL_g(\boldsymbol{B}, \boldsymbol{p}) \le BL(\boldsymbol{B}, \boldsymbol{p})$. The theorem 1.9 in [BCCT08] (first observed by Lieb) claims that actually the equality holds, while the theorem 1.15 in [BCCT08] gives the necessary and sufficient conditions for Brascamp - Lieb constant to be finite. We state both theorems simultaneously.

Theorem 4 (1.9 and 1.15 in [BCCT08])**.** *Let* $(\boldsymbol{B}, \boldsymbol{p})$ *be Brascamp - Lieb datum. Then:*

$$BL(\boldsymbol{B}, \boldsymbol{p}) = BL_g(\boldsymbol{B}, \boldsymbol{p})$$

and $BL(\boldsymbol{B}, \boldsymbol{p})$ *is finite if and only if the following two conditions hold.*

$$\dim(H) = \sum_{j=1}^{m} p_j \dim(H_j) \tag{3}$$

$$\dim(V) \le \sum_{j=1}^{m} p_j \dim(B_j(V)) \quad \textit{for all } V \le H \tag{4}$$

For the other approach we need the following "local" definition of the Brascamp - Lieb constant.

Definition 5. *Let* $\boldsymbol{B} = (\mathbb{R}^n, (\mathbb{R}^{n_j})_{1\le j\le m}, (B_j)_{1\le j\le m})$ *with* $\boldsymbol{p} \in \mathbb{R}^m_+$ *be Bras-camp - Lieb datum. We define*

$$BL(\boldsymbol{B}, \boldsymbol{p}, R) := \sup_{\boldsymbol{f}} \frac{\int_{[-R,R]^n} \prod_{j=1}^m f_j(B_j(x))^{p_j}}{\prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} f_j(x_j)dx_j\right)^{p_j}}, \tag{5}$$

where the supremum is taken over m*-tuples* $\boldsymbol{f} = (f_j)_{1\le j\le m}$*, of nonnegative functions that are constant on the cubes* $v + [0,1)^{n_j}$*,* $v \in \mathbb{Z}^{n_j}$*.*

In [M19], Maldague quantified the growth rate of $BL(\boldsymbol{B}, \boldsymbol{p}, R)$ as a function of R.

Theorem 6 (1 in [M19])**.** *Let* $(\boldsymbol{B}, \boldsymbol{p})$ *be a Brascamp - Lieb datum with* B_j *surjective for* $j = 1, \dots, m$*. Then*

$$BL(\boldsymbol{B}, \boldsymbol{p}, R) \asymp_{(\boldsymbol{B},\boldsymbol{p})} \sup_{V\le H} R^{\max\{\dim V - \sum_{j=1}^m p_j \dim B_j(V),\ 0\}} \tag{6}$$

4.2 Proofs of theorems 4 and 6 (outline)

4.2.1 Proof of theorem 4

Necessary conditions for finiteness of $BL(\boldsymbol{B}, \boldsymbol{p})$

Necessity of both (3) and (4) can be showed by testing (1) against Gaussian m-tuples $\boldsymbol{f} = (\exp(-\pi\langle A_j x, x\rangle))_{1\le j\le m}$.

To see that $BL(\boldsymbol{B}, \boldsymbol{p}) < \infty$ implies (3), we set $A_j = \lambda \,\mathrm{id}_{H_j}$, where $\lambda > 0$. Letting $\lambda \to 0$ and $\lambda \to \infty$, we conclude that (3) holds.

To see that $BL(\boldsymbol{B}, \boldsymbol{p}) < \infty$ implies (4), for arbitrary $V \le H$, we set $A_j = \epsilon\, \mathrm{id}_{B_j V} \oplus \mathrm{id}_{(B_j V)^\perp}$, where $0 < \epsilon < 1$. Bounding the denominator of (1) above with $C\epsilon^{\frac{\dim(V)}{2}}$ and letting $\epsilon \to 0$, (4) follows.

Sufficient conditions for finiteness of $BL(\boldsymbol{B}, \boldsymbol{p})$

If (3) or (4) doesn't hold, it follows from the previous paragraph that $BL_g(\boldsymbol{B}, \boldsymbol{p}) = \infty$, so because $BL_g(\boldsymbol{B}, \boldsymbol{p}) \le BL(\boldsymbol{B}, \boldsymbol{p})$, the equality holds in this case. Therefore, it remains to prove that if (3) and (4) hold, then

$$BL(\boldsymbol{B}, \boldsymbol{p}) = BL_g(\boldsymbol{B}, \boldsymbol{p}) < \infty \tag{7}$$

We need the following definition.

Definition 7. *A proper, non-zero subspace* $V \le H$ *is called critical with respect to datum* $(\boldsymbol{B}, \boldsymbol{p})$ *if*

$$\dim(V) = \sum_{j=1}^m p_j \dim(B_j(V))$$

A datum $(\boldsymbol{B}, \boldsymbol{p})$ *is called simple if* H *has no critical subspaces.*

Now we prove the statement (7) using induction on $\dim(H)$. We consider two cases.

In the first case, when $(\boldsymbol{B}, \boldsymbol{p})$ is simple (this case covers the basis of the indution because every 1-dimensional space is simple), one can show (5.2 in [BCCT08]) that $BL_g(\boldsymbol{B}, \boldsymbol{p}) < \infty$ and $(\boldsymbol{B}, \boldsymbol{p})$ is gaussian - extremisable. After that, one can show (3.6 in [BCCT08]) that there exists a geometric Brascamp - Lieb datum $(\boldsymbol{B}', \boldsymbol{p}')$ such that:

$$BL(\boldsymbol{B}, \boldsymbol{p}) = BL(\boldsymbol{B}', \boldsymbol{p}')BL_g(\boldsymbol{B}, \boldsymbol{p})$$

At last, one uses knowledge that Brascamp - Lieb constant is equal to 1 for geometric Brascamp - Lieb datum (2.8 in [BCCT08]) so the step induction step is proved in this case.

In the second case, if $(\boldsymbol{B}, \boldsymbol{p})$ has critical subspace V, one can show (4.7 and 4.8 in [BCCT08]) that the Brascamp - Lieb constants split:

$$BL(\boldsymbol{B}, \boldsymbol{p}) = BL(\boldsymbol{B}_V, \boldsymbol{p})BL(\boldsymbol{B}_{H/V}, \boldsymbol{p})$$

and the same for BL_g, where $\boldsymbol{B}_V = (V, (B_j(V))_{1\leq j\leq m}\ B_{j,V})$ and $\boldsymbol{B}_{H/V} = (H/V, (H_j/(B_j(V)))_{1\leq j\leq m}, B_{j,H/V})$ are appropriately defined Brascamp - Lieb data that both satisfy conditions (3) and (4). Since dimensions of V and H/V are strictly less than dimension of H, the claim (7) in this case follows inductively and the theorem 4 is proved.

4.2.2 Proof of theorem 6

Lower bound

The lower bound follows by discreztizing functions used to prove necessity of the condition (4) in the previous proof. Formally, for any subspace V, one tests inequality (5) against m-tuple of functions f_j that are constant on cubes that intersect the super-level set of the function $x \mapsto -\langle A_j x, x\rangle$, where $A_j = \frac{1}{R}\,\mathrm{id}_{B_jV} \oplus \mathrm{id}_{(B_jV)^\perp}$.

Upper bound

First we define the discretized quantity suitable for induction. Let $H \leq \mathbb{R}^n$ be a subspace and $B_j : H \to \mathbb{R}^{n_j}$ a linear operator. Let $\mathcal{L}_j^0 \subset \mathbb{Z}^{n_j}$ be a subset such that $B_j(H) \subset \cup_{v\in\mathcal{L}_j^0}(v + [0,1)^{n_j})$. For $A > 0$ we define:

$$\|f\|_{A,\mathcal{L}_j^0} := \sum_{v\in\mathcal{L}_j^0} \|f_j\|_{L^\infty((v+A[0,1)^{n_j})\cap B_j(H))}.$$

The theorem will follow from the following lemma by choosing $H = \mathbb{R}^n$ and noting that the product of the norms on the right hand side is comparable to the denominator of (5) when all B_j are surjective.

Proposition 8. *Let $H \leq \mathbb{R}^n$ be arbitrary subspace and let $B_j : H \to \mathbb{R}^{n_j}$ be linear operators. There exist parameters $A_j \geq 1$, $j = 1, \dots, m$ such that*

$$\int_{\{x \in H, |x| \leq R\}} \prod_{j=1}^{m} f_j(B_j(x))^{p_j} \lesssim \sup_{V \leq H} R^{\max\{\dim V - \sum_{j=1}^{m} p_j \dim B_j(V),\ 0\}} \prod_{j=1}^{m} \|f_j\|_{A_j, \mathcal{L}_j^0}^{p_j}$$

for all $R \geq 1$ and nonnegative measurable functions $f_j : H_j \to \mathbb{R}_+$

In this proof we say that a proper, nonzero subspace $V \leq H$ is *critical* with respect to datum $(\boldsymbol{B}, \boldsymbol{p})$ if

$$\dim(V) - \sum_{j=1}^{m} p_j \dim(B_j(V)) = \sup_{V \leq H} \left[\dim(V) - \sum_{j=1}^{m} p_j \dim(B_j(V)) \right]$$

and we call Brascamp - Lieb datum *simple* if H has no critical subsets. Now, we prove the proposition by induction on $m + \dim(H)$. When $m = 1$ the statement follows from change of variables and when $\dim(H) = 1$, the statement follows form Hölder's inequality and monotonicity of $l^p(\mathbb{Z})$-norms. Now, the step of the induction is proved considering two cases.

In the first case, when $(\boldsymbol{B}, \boldsymbol{p})$ is not simple with critical subset W, one can see that the supremum of exponent of R in the assumption for W equals to the wanted exponent and the exponent of R for $W^{\perp}$ equals to 0 so we can use Fubini's theorem and induction assumption (because $\dim(W), \dim(W^{\perp}) < \dim(H)$) to prove the inequality in this case.

In the second case, when $(\boldsymbol{B}, \boldsymbol{p})$ is simple, one considers set of all m-tuples $\boldsymbol{p}$ for which the datum is simple. One can easily see that the set is intersection of finitely many half spaces so it has finitely many extreme points. It is enough to prove the inequality for extreme points because of complex interpolation. Therefore, if $\boldsymbol{p}'$ is an extreme point, there are two possibilities - either there exists a critical subset for $(\boldsymbol{B}, \boldsymbol{p}')$ and we are in the first case of the induction or $\boldsymbol{p}'$ has some coordinate equal to zero in which case the inequality is reduced to the case with $m - 1$ functions. Therefore, the theorem is proved.

References

[BCCT08] J. Bennett, A. Carbery, M. Christ, and T. Tao. "The Brascamp-Lieb inequalities: finiteness, structure and extremals". In: *Geom. Funct. Anal.* **17.5.** (2008), pp. 1343-1415. arXiv:1811.11052. MR:2377493

[M19] D. Maldague. "Regularized Brascamp-Lieb inequalities and an application. Preprint. 2019. arXiv:1904.06450.

Aleksandar Bulj, University of Zagreb
email: aleksandar.bulj@math.hr

5 Duality and geometry of optimal transportation problems

After C. Villani [V]

A summary written by Constantin Bilz

Abstract

We introduce the optimal transportation problems of Kantorovich and Monge, discuss the Kantorovich duality and apply it to characterise optimal transference plans in an important special case.

5.1 Introduction

Let (X, μ) and (Y, ν) be topological spaces equipped with Borel probability measures and let $c : X \times Y \to [0, \infty]$ be a measurable *cost function.* The optimal transportation problem asks for the most efficient way to transfer all the mass from μ to ν, given that one unit of mass can be moved from a point $x \in X$ to a point $y \in Y$ at cost $c(x, y)$.

More precisely, a *transference plan* is a probability measure π on $X \times Y$ with marginals $\pi_X = \mu$ and $\pi_Y = \nu$, i.e.

$$\pi[A \times Y] = \mu(A) \quad \text{and} \quad \pi[X \times B] = \nu(B)$$

for any measurable sets $A \subseteq X$ and $B \subseteq Y$. Hence $\pi[A \times B]$ determines how much mass we transfer from the set A to the set B and the marginal conditions ensure that all of the mass in A is transferred to Y and that B receives the prescribed amount of mass. There always exists at least one transference plan, namely the tensor product $\mu \otimes \nu$. The *total transportation cost* associated to a transference plan is

$$I(\pi) = \int_{X \times Y} c(x, y) \, d\pi(x, y).$$

Let $\Pi(\mu, \nu)$ be the nonempty set of all transference plans. *Kantorovich's optimal transportation problem* from the 1940s is the following minimisation problem:

$$\text{Minimise } I(\pi) \text{ over all } \pi \in \Pi(\mu, \nu). \tag{1}$$

If the *optimal transportation cost* $\inf_{\pi \in \Pi(\mu,\nu)} I(\pi)$ is attained for some transference plan π, then we say that π is *optimal.*

Note that the statement of (1) does not involve the topological structure of X and Y. We will however need this structure for the results in the following sections.

Historically, Kantorovich's problem originated from a stronger and in general harder problem first considered by Monge in 1781. Monge's additional assumption is that no mass can be split, meaning that the only allowed transference plans are those of the form $\pi = \mu \circ (\mathrm{id} \times T)^{-1}$ for some measurable map $T\colon X \to Y$. In other words, *Monge's optimal transportation problem* is the following:

$$\text{Minimise } \int_X c(x, T(x))\, d\mu(x) \text{ over all measurable } T \text{ with } \nu = \mu \circ T^{-1}. \quad (2)$$

Since Kantorovich's problem is linear while Monge's problem is highly nonlinear, we will mostly focus on Kantorovich's problem. However, under certain circumstances, the solutions to both problems turn out to coincide. This is the case for example if $X = Y = \mathbb{R}^n$, the cost function is strictly convex and μ and ν assign no mass to sets of Hausdorff dimension at most $n-1$, see e.g. Theorem 4 below and [V, p. 6f.].

In Subsection 5.2 we will introduce the duality theory of the Kantorovich problem, which we will then use in Subsection 5.3 to characterise optimal transportation plans in an important special case.

5.2 Kantorovich duality via Fenchel–Rockafellar duality

That Kantorovich's problem admits a useful dual formulation may be expected because of the linearity of both I and the conditions defining $\Pi(\mu, \nu)$. In fact, this duality is related to the duality of finite-dimensional linear programming, a subject that was, a few years previously, also founded by Kantorovich.

We now describe the dual problem to Kantorovich's problem. Let Φ_c be the set of pairs $(\phi, \psi) \in L^1(X, \mu) \times L^1(Y, \nu)$ such that

$$\phi(x) + \psi(y) \le c(x, y)$$

for μ-almost every x and ν-almost every y. Define

$$J(\phi, \psi) = \int_X \phi\, d\mu + \int_Y \psi\, d\nu.$$

The following result contains the duality between (1) and the maximisation problem for J on Φ_c. The assumptions are very general, but a typical case for applications is the Euclidean space $X = Y = \mathbb{R}^n$ together with the quadratic cost function $c(x, y) = |x - y|^2$.

A *Polish space* is a separable completely metrisable topological space.

Theorem 1 (Kantorovich duality)**.** *Let (X, μ) and (Y, ν) be Polish spaces equipped with Borel probability measures. If c is lower semi-continuous, then Kantorovich's problem* (1) *admits an optimal transference plan and*

$$\min_{\pi \in \Pi(\mu,\nu)} I(\pi) = \sup_{(\phi,\psi)\in\Phi_c} J(\phi, \psi).$$

Just like the duality of linear programming, Theorem 1 can be proved by using a minimax principle, in this case the Fenchel–Rockafellar duality from convex analysis. For this, let E be a normed space and let E^* be its topological dual. For our purposes, E will be the space of bounded continuous functions on $X \times Y$. Hence if X and Y are compact, then E^* will be the space of Radon measures on $X \times Y$.

Given a convex function Θ on E with values in $\mathbb{R} \cup \{+\infty\}$, its *Legendre–Fenchel transform* (or *convex conjugate*) is the convex function Θ^* on E^* with values in $\mathbb{R} \cup \{+\infty\}$ given by

$$\Theta^*(z^*) = \sup_{z\in E}\langle z^*, z\rangle - \Theta(z).$$

Theorem 2 (Fenchel–Rockafellar duality)**.** *Let Θ and Ξ be convex functions on E with values in $\mathbb{R} \cup \{+\infty\}$ and let $z_0 \in E$ be such that $\Theta(z_0)$ and $\Xi(z_0)$ are finite and Θ is continuous at z_0. Then,*

$$\inf_{z\in E} \Theta(z) + \Xi(z) = \sup_{z^*\in E^*} -\Theta^*(-z^*) - \Xi^*(z^*)$$

and the supremum is attained.

The proof of this result is rather short and uses the Hahn–Banach theorem. If X and Y are compact and c is continuous, then Theorem 1 is an easy consequence of Theorem 2. The general case follows from this by a technical approximation argument.

5.3 Optimality criteria

We will apply the Kantorovich duality to characterise optimal transference plans in the case of the quadratic cost function on Euclidean space.

In order to formulate the results in this section, we need the notion of the subdifferential of a convex function. Let ϕ be a convex function on $\mathbb{R}^n$

with values in $\mathbb{R} \cup \{+\infty\}$. By Rademacher's theorem, ϕ is differentiable almost everywhere on $\operatorname{dom}(\phi) = \{x \in \mathbb{R}^n \mid \phi(x) < \infty\}$. At every point x of differentiability it holds that for $y = \nabla\phi(x)$,

$$\phi(z) \geq \phi(x) + \langle y, z - x \rangle \quad \text{for any } z \in \mathbb{R}^n. \tag{3}$$

While the gradient $\nabla\phi$ is only defined almost everywhere on $\operatorname{dom}(\phi)$, the *subdifferential* $\partial\phi$ is a set-valued function defined everywhere as follows: For $y \in \mathbb{R}^n$ we let $y \in \partial\phi(x)$ if and only if (3) holds. One can show that $\partial\phi(x) = \{y\}$ if and only if ϕ is differentiable at x with $y = \nabla\phi(x)$. In the following it may be helpful to identify the graph of the subdifferential with a subset of $X \times Y = \mathbb{R}^n \times \mathbb{R}^n$.

Theorem 3 (Knott–Smith criterion)**.** *Let μ and ν be Borel probability measures on $\mathbb{R}^n$ with finite second moments and consider the quadratic cost function $c(x, y) = |x - y|^2$. A transference plan π is optimal if and only if there exists a convex lower semi-continuous function ϕ such that*

$$y \in \partial\phi(x) \quad \textit{for } \pi\textit{-almost every pair of points } (x, y).$$

In this case, the dual Kantorovich problem is maximised by the pair

$$\left(\frac{|x|^2}{2} - \phi(x), \frac{|y|^2}{2} - \phi^*(y) \right),$$

where ϕ^* is the convex conjugate defined in the last subsection.

Uniqueness is ensured by the next theorem if we make an additional assumption on the regularity of μ.

Theorem 4 (Brenier's theorem)**.** *Under the assumptions of Theorem 3, if additionally μ gives no mass to sets of Hausdorff dimension at most $n - 1$, then there exists a unique optimal transference plan π and it holds that*

$$\pi = \mu \circ (\operatorname{id} \times \nabla\phi)^{-1},$$

where $\nabla\phi$ is the μ-almost everywhere unique gradient of a convex function such that $\nu = \mu \circ (\nabla\phi)^{-1}$.

In this case, $\nabla\phi$ is called the *Brenier map* pushing μ forward to ν. It also follows from the theorem that $\nabla\phi$ is the unique solution to the Monge transportation problem (2).

References

[V] Villani, C., *Topics in optimal transportation.* Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence, RI, 2003, xvi+370 pp.

Constantin Bilz, University of Birmingham
email: C.Bilz@pgr.bham.ac.uk

6 Existence and regularity of the Brenier's map

After R. J. McCann [McC95] and S. Alesker, S. Dar, and V. Milman [ADM99].

A summary written by Mateus Sousa

Abstract

We present the construction of the Brenier map by McCann and its regularities properties proved by Caffareli.

6.1 Introduction

Given a pair of probability measures μ and ν in $\mathbb{R}^d$, a natural question one might ask is if there is a canonical way of transporting one measure to another. More precisely, what is the "natural" way can one produce a map $T : (\mathbb{R}^d, \mu) \to (\mathbb{R}^d, \nu)$ such that for every Borel set $M \subset \mathbb{R}^d$

$$T_{\#}\mu(M) := \mu(T^{-1}(M)) = \nu(M).$$

In dimension $d = 1$, as long as the measures μ and ν are free from atoms, one can choose a function T satisfying

$$\mu((-\infty, x]) = \nu((-\infty, T(x)]), \tag{1}$$

where T is non-decreasing and $T(x) \in \mathbb{R} \cup \{\pm\infty\}$, and is uniquely determined μ-almost everywhere, and one could say such transformation is indeed very natural and a good answer to the aforementioned question. In higher dimensions, such a natural map might not be clear at first glance, but an answer to this questions was given by Brenier [Bre87, Bre91] under certain restrictions on μ and ν. In the class of measures which Brenier was working with, he proved that one can produce a transformation $T : (\mathbb{R}^d, \mu) \to (\mathbb{R}^d, \nu)$ such that $T = \nabla\psi$, where ψ is a convex function, and such transformation T is unique μ-almost everywhere. Brenier's result was later improved by McCann [McC95], in the form of the following result.

Theorem 1. *Let μ and ν be probability measures on $\mathbb{R}^d$, such that μ vanishes on Borel sets of Hausdorff dimension $d - 1$. Then there exists a convex function ψ on $\mathbb{R}^d$ such that $(\nabla\psi)_{\#}\mu = \nu$. Although ψ might not be unique, the map $\nabla\psi$ is uniquely determined μ-almost everywhere.*

One interesting face of this problem happens when the measures μ and ν are both absolutely continuous with respect to the Lebesgue measure, i.e, if there are nonnegative functions f and g such that $\mathrm{d}\mu(x) = f(x)\mathrm{d}x$ and $\mathrm{d}\nu(x) = g(x)\mathrm{d}x$. In the one-dimensional case, assuming enough regularity on f and g, one can formally differentiate equation (1) to obtain

$$T'(x)g(T(x)) = f(x).$$

In the higher dimension setting, when in the presence of the Brenier map $\nabla\psi$, this formally generalizes to the following equation

$$\det(D^2\psi(x))g(\nabla\psi(x)) = f(x), \tag{2}$$

and one can found results in the literature about how smoothness of the densities f and g ensures regularity of ψ, and is one the techniques to produce convex solutions to the Monge-Ampère equation (2). One instance of such regularity results was proven by Caffarelli [ADM99, Theorem 1.3]. Assume that

(i) f is locally bounded and locally bounded away from zero, i.e, for every $R > 0$ and $|x| \geq R$

$$0 < c(R) \leq f(x) \leq C(R).$$

(ii) The measure ν is supported in (the closure of) a bounded open convex set Γ, and there are constants $\lambda, \Lambda > 0$ such that

$$\lambda \leq g(y) \leq \Lambda$$

for every $y \in \Gamma$.

Theorem 2 ([ADM99])**.** *Under conditions (i) and (ii), the Brenier map*

$$\nabla\psi(\mathbb{R}^d, \mu) \to (\mathbb{R}^d, \nu)$$

is continuous. Moreover, $\nabla\psi$ belongs to the Hölder class C^α, for some $\alpha > 0$. Furthermore, if f and g are locally Hölder, then ψ is $C^{2.\alpha}$, for some $\alpha > 0$.

The proof of Theorem 2 relies on several geometric observations about of convex functions and affine invariance properties of solutions to the Monge-Ampère equation (2); see the appendix in [ADM99] and the references in there for more details.

On the remainder of this summary, we focus on the main ingredients behind the proof of existence of the map $\nabla\psi$ in Theorem 1.

6.2 McCann's result

6.2.1 Preliminaries

In the context of Theorem 1, a convex function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a function such that

$$\psi(tx + (1-t)y) \leq t\psi(x) + (1-t)\psi(y),$$

whenever the right-hand side above is finite, and $t \in (0,1)$. Any such function is known to be continuous on the interior of the convex set $\text{dom}\psi := \{\psi < \infty\}$, and differentiable outside of set of Hausdorff dimension at most $d-1$. This clearly justifies the very natural hypothesis that μ vanishes on sets of dimensions $d-1$ in Theorem 1. Although the gradient might not be defined everywhere, for every point in the interior of $\text{dom}\psi$ there is a $y \in \mathbb{R}^d$ such

$$\psi(z) - \psi(x) \geq \langle y, z - x \rangle \tag{3}$$

for all $z \in \mathbb{R}^d$. Such a point y is called a subgradient of ψ at x. Whenever $\nabla\psi(x)$ exists, it is a subgradient of ψ at x. This motivates the following definition:

Definition 3. *The subdifferential of a convex function ψ on $\mathbb{R}^d$ is the subset $\partial\psi \subset \mathbb{R}^d \times \mathbb{R}^d$ of pairs (x, y) such that y is a subgradient of ψ at x.*

If one sums inequality (3) over a cyclic sequence $\{x_i\}$ on $\text{dom}\psi$ such that $x_1 = x_{n+1}$ and choose $z_i = x_{i+1}$ one has

$$\langle y_1, x_2 - x_1 \rangle + \cdots + \langle y_n, x_1 - x_n \rangle \leq 0. \tag{4}$$

Definition 4. *A set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ is said to be cyclically monotone if any finite number of points $(x_i, y_i) \in S$, $i = 1, \ldots, n$ satisfies the inequality* (4)*.*

By the previous observation, it is obvious any subset of the subdifferential of a convex function is cyclic monotone. Rockafeller's theorem [Vil03, Theorem 2.27] provides a converse to this principle: any cyclic monotone set has to be contained in the subdifferential of some convex function. This fact is the key ingredient in order to prove existence of ψ in Theorem 1.

Definition 5. *A measure γ in $\mathbb{R}^d \times \mathbb{R}^d$ is said to have μ and ν as its left and right marginals respectively if $\mu(M) = \gamma(M \times \mathbb{R}^d)$ and $\nu(M) = \gamma(\mathbb{R}^d \times M)$ for any borel set $M \subset \mathbb{R}^d$. We denote the set of all such measures as $\Gamma(\mu, \nu)$.*

6.2.2 A glimpse of the proof

The following proposition explicits the connection between measures supported in cyclic monotone sets with marginals μ and ν and the Brenier map. Here $\mathrm{id} \times \nabla\psi$ denotes the map $x \mapsto (x, \nabla\psi(x)$.

Proposition 6. *Supposed a probability measure γ in $\mathbb{R}^d \times \mathbb{R}^d$ is supported on a subset of a subdifferential $\partial\psi$ of a convex function ψ. Let $\gamma \in \Gamma(\mu, \nu)$. If μ vanishes on Borel sets of Hausdorff dimension $d-1$, then the gradient $\nabla\psi$ pushes μ into ν, i.e, $(\nabla\psi)_{\#}\mu = \nu$. Furthermore, $\gamma = (\mathrm{id} \times \nabla\psi)_{\#}\mu$.*

Given Rockafeller's theorem, Proposition 6 implies directly the desired result in case $\Gamma(\mu.\nu)$ contained at least one measure γ supported on a subset of a subdifferential a convex function, and that is the content of the following result.

Theorem 7 (Existence of monotone correlations)**.** *There is a a measure $\gamma \in \Gamma(\mu, \nu)$ supported in a cyclic monotone set.*

This result follows in three steps. First, in the case where μ and ν are special sums of point masses, this result is straightforward. Second, the property in the result is preserved by weak-* limits. Finally, the aforementioned set of special sums of point masses will form a dense set in the weak-* topology. Explicitly, this is the content of the following three lemmas.

Lemma 8 (Cyclical monotonicity of correlated pairs)**.** *Fix n points $x_i \in \mathbb{R}^d$ and n points $y_i \in \mathbb{R}^d$. There is a permutation σ of the set $\{1, \ldots, n\}$ such that the set $S = \{(x_{\sigma(i)}, y_i),\ i = 1, \ldots, n\}$ is cyclic monotone.*

Lemma 9 (Weak-* density of point mass sums)**.** *The set of measures of the form*

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$

with $x_1, \ldots, x_n \in \mathbb{R}^d$ forms a weak- dense set of the set all probability measures in $\mathbb{R}^d$.*

Lemma 10 (Weak-* limits preserve monotone correlations)**.** *Consider a sequence γ_k of probability measures in $\mathbb{R}^d \times \mathbb{R}^d$ that converges in the weak-* to γ. Then*

(i) γ will have cyclic monotone support if all the γ_n have cyclic monotone support;

(ii) If the left and right marginal of γ_n converge in the weak-* sense to limits μ and ν respectively, then $\gamma \in \Gamma(\mu, \nu)$.

And now we can conclude the existence of the desired $\nabla\psi$ such that $(\nabla\psi)_{\#}\mu = \nu$ in Theorem 1, as we wished.

References

[Bre87] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs.* C. R. Acad. Sci. Paris Sér. Math. 305 (1987), 805-808.

[Bre91] Y. Brenier, *Polar factorization and monotone rearranoement of vector-valued functions.* Comm. Pure Appl. Math. 44 (1991), 375-417.

[ADM99] S. Alesker, S. Dar, and V. Milman, *A remarkable measure preserving diffeomorphism between two convex bodies in* $\mathbb{R}^n$. Geom. Dedicata 74.2 (1999), pp. 201?212.

[McC95] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps.* Duke Math. J. 80.2 (1995), pp. 309?323.

[Vil03] C. Villani. *Topics in optimal transportation.* Vol. 58. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003, pp. xvi+370.

Mateus Sousa, Basque Center for Applied Mathematics
email: mcosta@bcamath.org

7 BL and reverse BL via mass transport

After Franck Barthe [BAR1]

A summary written by Aswin Govindan Sheri

Abstract

We prove that multi-dimensional Brascamp–Lieb inequality and its reverse form are exhausted by centred gaussians. The proof is based on a theorem of Brenier on mass-preserving maps between measure spaces.

7.1 Introduction

We begin by describing the problem of Brascamp–Lieb inequality in its multi-dimensional form. Let $m \geq n$, $\mathbf{p} = (p_i)_{1 \leq i \leq m} \in (0, \infty)^{(m)}$ and $\mathbf{B} = (B_i)_{1 \leq i \leq m}$ be chosen such that $B_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$ are linear maps. The Brascamp–Lieb constant $\mathrm{BL}(\mathbf{B}, \mathbf{p})$ is defined as

$$\mathrm{BL}(\mathbf{B}, \mathbf{p}) := \sup_{f_i \in L^1_+(\mathbb{R}^{n_i})} \left\{ B : \int \prod_{j=1}^{m} (f_j \circ B_j(x))^{p_j} \mathrm{d}x \leq B \prod_{j=1}^{m} \left(\int f_j \right)^{p_j} \right\}. \tag{1}$$

We are interested in the extremisers of (1). Moving in that direction, let us restrict our focus to the centred gaussians. Suppose $\mathcal{S}_+(\mathbb{R}^k)$ denote the space of all $k \times k$ symmetric, positive definite matrices. For $A \in \mathcal{S}_+(\mathbb{R}^k)$, let g_A denote the centred gaussian function on $\mathbb{R}^n$ defined by

$$g_A(x) = e^{-\langle Ax, x \rangle}.$$

Slightly modifying (1), we can introduce the constant $\mathrm{BL}_g(\mathbf{B}, \mathbf{p})$ that takes the optimal value of B in (1), now restricted to all the centred gaussians. i.e.

$$\begin{aligned} \mathrm{BL}_g(\mathbf{B}, \mathbf{p}) : &= \sup_{A_i \in \mathcal{S}^+(\mathbb{R}^{n_i})} \left[\int \prod_{j=1}^{m} (g_{A_j} \circ B_j(x))^{p_j} \mathrm{d}x \right] \left[\prod_{j=1}^{m} \left(\int g_{A_j} \right)^{-p_j} \right] \\ &= \pi^{\frac{1}{2}(n - \sum_{i=1}^{m} p_i n_i)} \sup_{A_i \in \mathcal{S}^+(\mathbb{R}^{n_i})} \left[\det \left(\sum_{j=1}^{m} p_j B_j^* A_j B_j \right)^{-\frac{1}{2}} \right] \left[\prod_{j=1}^{m} (\det A_j)^{\frac{p_j}{2}} \right] \end{aligned} \tag{2}$$

where we have made use of the the identity

$$\int_{\mathbb{R}^n} e^{-\langle Ax, x \rangle} \mathrm{d}x = \pi^{\frac{n}{2}} \cdot (\det A)^{-\frac{1}{2}}, \quad \text{whenever } A \in \mathcal{S}_+(\mathbb{R}^n).$$

By its very definition, $\mathrm{BL}_g(\mathbf{B},\mathbf{p}) \le \mathrm{BL}(\mathbf{B},\mathbf{p})$. The question that interests us is the validity of its converse. In other words, is the BL inequality (1) exhausted by centred gaussians? This question was first answered positively by a classical theorem of Lieb[LIEB] and later by Barthe[BAR1].

In [BAR1], the proof begins from a dual formulation of (1), called the reverse BL inequality. The philosophy of dualising comes from convexity theory in order to attack optimisation problems with convex constraints. For $f_i \in L^1_+(\mathbb{R}^{n_i})$, consider the function

$$I((f_i)_{1\le i\le m}) := \int^*_{\mathbb{R}^n} \sup\left\{ \prod_{i=1}^m f_i^{p_i}(y_i) : \sum_{i=1}^m p_i B_i^* y_i = x \text{ and } y_i \in \mathbb{R}^{n_i} \right\} \mathrm{d}x \quad (3)$$

where $\int^*$ represents an outer integral. The reverse BL constant is defined as

$$\mathrm{RBL}(\mathbf{B},\mathbf{p}) := \inf_{f_i \in L^1_+(\mathbb{R}^{n_i})} \frac{I((f_i)_{1\le i\le m})}{\prod_{j=1}^m \left(\int f_j\right)^{p_j}}. \quad (4)$$

As before, we shall restrict the space of all functions where the infimum is taken in (4) to define

$$\mathrm{RBL}_g(\mathbf{B},\mathbf{p}) := \inf_{A_i \in \mathcal{S}_+(\mathbb{R}^{n_i})} \frac{I((g_{A_i})_{1\le i\le m})}{\prod_{j=1}^m \left(\int g_{A_j}\right)^{p_j}}.$$

Having all the basic definitions in place, we are ready to state the result of Barthe.

Theorem 1 ([BAR1])**.** *Let a BL datum $(\boldsymbol{B},\boldsymbol{p})$ be chosen such that*

$$\sum_{i=1}^m p_i n_i = n, \quad \bigcap_{i\le m} \ker B_i = \{0\},$$

and B_i's are linear surjections onto $\mathbb{R}^{n_i}$ for each i. Then,

$$\mathrm{BL}(\boldsymbol{B},\boldsymbol{p}) = \mathrm{BL}_g(\boldsymbol{B},\boldsymbol{p}) \text{ and } \mathrm{RBL}(\boldsymbol{B},\boldsymbol{p}) = \mathrm{RBL}_g(\boldsymbol{B},\boldsymbol{p}). \quad (5)$$

Moreover, they are related to each other by the relation

$$\mathrm{BL}(\boldsymbol{B},\boldsymbol{p}) = [\mathrm{RBL}(\boldsymbol{B},\boldsymbol{p})]^{-1}. \quad (6)$$

7.2 Proof of Theorem 1

We begin by proving a statement of duality that relates the forward and reverse BL constants for the gaussians.

Lemma 2. *Let* $(\boldsymbol{B}, \boldsymbol{p})$ *satisfies the assumptions in Theorem 1. Then,*

$$\mathrm{BL}_g(\boldsymbol{B}, \boldsymbol{p}) \cdot \mathrm{RBL}_g(\boldsymbol{B}, \boldsymbol{p}) = 1 \text{ and } \mathrm{RBL}_g(\boldsymbol{B}, \boldsymbol{p}) = 0 \iff \mathrm{BL}_g(\boldsymbol{B}, \boldsymbol{p}) = \infty. \tag{7}$$

Proof. Fix an m-tuple $(A_i)_{1\le i\le m}$, where $A_i \in \mathcal{S}_+(\mathbb{R}^{n_i})$. Consider a quadratic form

$$Q(y) := \left\langle \sum_{i=1}^m p_i B_i^* A_i B_i y, y \right\rangle, \quad y \in \mathbb{R}^n.$$

Note that the assumptions on $(\mathbf{B}, \mathbf{p})$ make sure that $\det Q \neq 0$. The dual quadratic form of Q is defined by the relation

$$Q^*(x) := \sup\{|\langle x, y\rangle|^2 : Q(y) \le 1\}, \quad x \in \mathbb{R}^n.$$

As it turns out, one can give an alternate expression for Q^*. We claim that

$$Q^*(x) = \inf_{A_i\in\mathcal{S}_+(\mathbb{R}^{n_i})} \left\{ \sum_{i=1}^m \left\langle p_i A_i^{-1} x_i, x_i \right\rangle : x = \sum_{i=1}^m p_i B_i^* x_i \right\} \tag{8}$$

Indeed, if $x = \sum_{i=1}^m p_i B_i^* x_i$, Cauchy–Schwartz tells us that

$$\langle x, y\rangle^2 = \left\langle \sum_{i=1}^m p_i B_i^* x_i, y \right\rangle^2 = \left\langle \sum_{i=1}^m p_i A_i^{-\frac{1}{2}} x_i, A_i^{\frac{1}{2}} B_i y \right\rangle^2 \le \left\langle \sum_{i=1}^m p_i A_i^{-1} x_i, x_i \right\rangle Q(y).$$

In fact, we will have an equality above if $x_i = A_i B_i y$ and $y = (\sum_{i=1}^m p_i B_i^* A_i B_i)^{-1} x$. Thus the supremum in (8) and the infimum in the definition of Q^* coincide.

Recalling the definition of the operator I in (3), we can write

$$\begin{aligned}
I((g_{A_i})_{1\le i\le m}) &= \int_{\mathbb{R}^n} \sup\left\{ e^{-\sum_{i=1}^m \left\langle p_i A_i^{-1} x_i, x_i \right\rangle} : x = \sum_{i=1}^m p_i B_i^* x_i \right\} \mathrm{d}x \\
&= \int_{\mathbb{R}^n} e^{-Q^*(x)} \mathrm{d}x = \pi^{\frac{n}{2}} (\det Q^*)^{-\frac{1}{2}}.
\end{aligned}$$

However, $\det Q^* \cdot \det Q = 1$ by the classical duality relation. Thus,

$$\begin{aligned}
\mathrm{RBL}_g(\mathbf{B}, \mathbf{p}) &= \inf_{A_j\in\mathcal{S}_+(\mathbb{R}^{n_i})} \pi^{\frac{n}{2}} (\det Q^*)^{-\frac{1}{2}} \prod_{j=1}^m \left(\int g_{A_j} \right)^{-\frac{p_j}{2}} \\
&= \inf_{A_j\in\mathcal{S}_+(\mathbb{R}^{n_i})} \pi^{\frac{1}{2}(n-\sum_{i=1}^m p_i n_i)} (\det Q)^{\frac{1}{2}} \prod_{j=1}^m (\det A_j)^{-\frac{p_j}{2}} \\
&= [\mathrm{BL}_g(\mathbf{B}, \mathbf{p})]^{-1}.
\end{aligned}$$

□

The next lemma uses techniques from mass transport theory to find a relation between $\mathrm{RBL}(\mathbf{B},\mathbf{p})$ and $\mathrm{BL}(\mathbf{B},\mathbf{p})$.

Lemma 3.
$$\mathrm{BL}(\boldsymbol{B},\boldsymbol{p}) \leq (\mathrm{BL}_g(\boldsymbol{B},\boldsymbol{p}))^2\mathrm{RBL}(\boldsymbol{B},\boldsymbol{p}). \tag{9}$$

Before proceeding to its proof, let us see how Theorem 1 follows now. Indeed, using (9) and (7), one can create the chain

$$\begin{aligned}\mathrm{BL}(\mathbf{B},\mathbf{p}) &\leq (\mathrm{BL}_g(\mathbf{B},\mathbf{p}))^2\mathrm{RBL}(\mathbf{B},\mathbf{p}) \\ &\leq (\mathrm{BL}_g(\mathbf{B},\mathbf{p}))^2\mathrm{RBL}_g(\mathbf{B},\mathbf{p}) = \mathrm{BL}_g(\mathbf{B},\mathbf{p}) \leq \mathrm{BL}(\mathbf{B},\mathbf{p}).\end{aligned}$$

Thus, all inequalities in this chain are in fact equalities, which concludes the proof of (5) and (6).

Proof of Lemma 3. Let us begin with certain reductions in the problem. Of-course, we can assume that $0 < \mathrm{BL}_g(\mathbf{B},\mathbf{p}) < \infty$. Let $f_j, h_j \in L^1_+(\mathbb{R}^{n_i})$ for $1 \leq i \leq m$ such that $\int f_j = \int h_j = 1$. Define a multi-linear operator J by

$$J(h_1,\cdots,h_m) := \int \prod_{j=1}^{m}(h_j \circ B_j(x))^{p_j}\mathrm{d}x.$$

By (1) and (4), it is enough to prove that

$$J(h_1,\cdots,h_m) \leq (\mathrm{BL}_g(\mathbf{B},\mathbf{p}))^2 I(f_1,\cdots,f_m) \tag{10}$$

for any such choice of functions. Let $\mathfrak{C}_L(\mathbb{R}^n)$ denote the subset of $L^1_+(\mathbb{R}^n)$ whose elements are restrictions to some open Euclidean ball of positive lipchitz function on $\mathbb{R}^n$. Using standard density arguments and the monotonicity of I and J, one can see that it is enough to prove (10) when $f_i, h_i \in \mathfrak{C}_L(\mathbb{R}^{n_i})$.

To proceed from here, we need to invoke a result from the mass transport theory about measure preserving maps. Its usage is motivated by an earlier result of Barthe[BAR2] where the one dimensional version of this lemma was proved using measure-preserving maps in $\mathbb{R}$. In the multi-dimensional case, Brenier[BRE] proved the existence of a mass preserving map, deriving out of a convex potential provided some strong integral assumptions on the moments of boundary measures are met. Later, these extra assumptions were removed by McCann[McC]. Moreover, A result by Caffarelli provides an insight into the regularity properties of such Brenier mappings. For the purpose of our proof, all these results are combined and written as a single theorem below.

Theorem 4 ([CAR],[BRE],[McC]). *For $i = 1, 2$, let Ω_i be a bounded region of $\mathbb{R}^k$ with Ω_2, convex. Let $f_i \geq 0$ be integrable functions supported on Ω_i with $\int f_1 = \int f_2$. Assume also that f_i's are lipchitz and that both f_i and f_i^{-1} are bounded on Ω_i. There exists a convex, twice continuously differentiable function $\Phi : \mathbb{R}^k \to \mathbb{R}$ such that $\nabla\Phi$ is a measure preserving map from $(\Omega_1, f_1 \mathrm{d}x)$ to $(\Omega_2, f_2 \mathrm{d}x)$. Moreover, Φ satisfies the Monge–Ampére equation*

$$\det(\nabla^2\Phi(x)) f_2 \circ \nabla\Phi(x) = f_1(x) \text{ for all } x \in \Omega_1.$$

In order to apply this theorem in our setup, let Ω_{h_i} denote the region where h_i is positive. By Theorem 4 we obtain a continuously differentiable map $T_i = \nabla\Phi_i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$ such that

$$\det(dT_i(x)) f_i \circ T_i(x) = h_i(x) \text{ for all } x \in \Omega_{h_i}. \tag{11}$$

Since Φ is convex and h_i is non-vanishing on Ω_{h_i}, $dT_i(x)$ lies in $\mathcal{S}_+(\mathbb{R}^{n_i})$. Define a function $\Theta : S := \bigcap_{i=1}^m B_i^{-1}(\Omega_{h_i}) \to \mathbb{R}^n$ by

$$\Theta(y) := \sum_{i=1}^m p_i B_i^* T_i(B_i y).$$

Clearly, $d\Theta_i(y)$ is positive semi-definite. Also, (2) tells us that

$$\det(d\Theta(y)) \geq [\mathrm{BL}_g(\mathbf{B}, \mathbf{p})]^{-1} \prod_{i=1}^m (\det dT_i(B_i y))^{p_i} > 0$$

for any $y \in S$. Thus, $d\Theta(y) \in \mathcal{S}_+(\mathbb{R}^n)$ with which we can conclude that Θ is injective. Using (2) and (11), we see that

$$\begin{aligned}
&(\mathrm{BL}_g(\mathbf{B}, \mathbf{p}))^{-1} J((h_i)_{1\leq i\leq m}) \\
&\quad= (\mathrm{BL}_g(\mathbf{B}, \mathbf{p}))^{-1} \int_S \prod_{i=1}^m h_i^{p_i}(B_i y) \mathrm{d}y \\
&\quad= (\mathrm{BL}_g(\mathbf{B}, \mathbf{p}))^{-1} \int_S \prod_{i=1}^m [\det(dT_i \circ B_i(y)) f_i(T_i \circ B_i y)]^{p_i} \,\mathrm{d}y \\
&\quad\leq \int_S \det(d\Theta(y)) \prod_{i=1}^m [f_i(T_i \circ B_i y)]^{p_i} \,\mathrm{d}y \\
&\quad\leq \int_S \det(d\Theta(y)) \sup\left\{ \prod_{i=1}^m f_i(y_i)^{p_i} : \Theta(y) = \sum_{i=1}^m p_i B_i^* y_i \right\} \mathrm{d}y \\
&\quad= \int_{\mathbb{R}^n} \sup\left\{ \prod_{i=1}^m f_i(y_i)^{p_i} : x = \sum_{i=1}^m p_i B_i^* y_i \right\} \mathrm{d}x \\
&\quad= I((f_i)_{1\leq i\leq m}).
\end{aligned}$$

By taking the supremum over all h_i's and infimum over all f_i's, one can finally obtain (9). $\square$

References

[BAR1] Barthe, Franck., *On a reverse form of the Brascamp-Lieb inequality* Invent. Math. 137 (1998), 335-361.

[LIEB] Lieb, E. H., *Gaussian kernels have only Gaussian maximizers*, Invent. Math. 102 (1990), 179 -208.

[CAR] Caffarelli,L., *The regularity of mappings with a convex potential*, J. Amer. Math. Soc., 4 (1992), 99 -104.

[BRE] Brenier, Yann, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44 (1991), 375–417.

[BAR2] Barthe, Franck, *Inégalités de Brascamp-Lieb et convexité*, C. R. Acad. Sci. Paris Sér. I Math., 324 (1997), 885–888.

[McC] McCann, Robert J., *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J., 80 (1995), 309–323.

Aswin Govindan Sheri, University of Edinburgh
email: `aswin.govindan-sheri@sms.ed.ac.uk`

8 Positive Gaussian kernels also have Gaussian minimizers

After F. Barthe and P. Wolff [BW]

A summary written by Valentina Ciccone

Abstract

Following the results of [BW], lower bounds for certain operators with Gaussian kernels are discussed and conditions under which the sharpest constant can be computed by considering centered Gaussian functions only are provided.

8.1 Introduction

The celebrated work of Lieb "Gaussian kernels have only Gaussian maximizers" [L] studies operators with Gaussian kernels from L^p to L^q and provides conditions under which the operator norm can be computed by considering centered Gaussian functions only. For the particular case of multilinear operators with real valued Gaussian kernel Lieb's result [L] reads as follows.

Theorem 1. *Let m be a positive integer and for $i = 1, ..., m$ let $p_i \geq 1$ and $B_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$ be linear surjective maps. Let $\mathcal{Q}$ be a positive semi-definite quadratic form on $\mathbb{R}^n$. For functions $f_i \in L^{p_i}(\mathbb{R}^{n_i}, \mathbb{R})$ non-identically zero define the functional*

$$H(f_1, ..., f_m) = \frac{\int_{\mathbb{R}^n} e^{-\mathcal{Q}(x)} \prod_{i=1}^m f_i(B_i x) dx}{\prod_{i=1}^m \|f_i\|_{p_i}}.$$

Then the supremum of H over all $m-$tuples of such functions is equal to the supremum of H over $m-$tuples of centered Gaussian functions only.

Observe that by setting $c_i = \frac{1}{p_i}$ and substituting f_i with $f_i^{c_i}$ one gets the same result for the following functional on non-negative integrable functions:

$$I(f_1, ..., f_m) = \frac{\int_{\mathbb{R}^n} e^{-\mathcal{Q}(x)} \prod_{i=1}^m f_i(B_i x)^{c_i} dx}{\prod_{i=1}^m \left(\int_{\mathbb{R}^{n_i}} f_i \right)^{c_i}}.$$

In the particular case in which $\mathcal{Q} = 0$ one may recover from the above theorem the celebrated Brascamp-Lieb inequalities.

Several related inverse inequalities have been studied in the literature over the years.

In particular, in [BW] inequalities of the following form are considered

$$\int_H e^{-\mathcal{Q}(x)} \prod_{k=1}^m f_i^{c_k}(B_k x)dx \geq C \prod_{k=1}^m \left(\int_{H_k} f_i \right)^{c_k},$$

where:

- $0 \leq m^+ \leq m$ are integers;
- $H, H_1, ..., H_m$ are Euclidean spaces endowed with the usual Lebesgue measure;
- for $k = 1, ..., m$, $c_k \in \mathbb{R}$ are such that $c_i > 0$ for $i \leq m^+$ and $c_i < 0$ for $i > m^+$;
- $B_k : H \to H_k$ are surjective linear maps;
- $\mathcal{Q} : H \to \mathbb{R}$ is a quadratic form with signature $(s^+(\mathcal{Q}), s^-(\mathcal{Q}))$;
- $f_k : H_k \to [0, +\infty]$ are measurable functions satisfying $0 < \int_{H_k} f_k < +\infty$.

We refer to this inequalities as inverse Brascamp-Lieb inequalities.

The objective is to provide a counterpart to Lieb's result and to answer questions like: When is it possible to compute the optimal constant C by considering only (centered) Gaussian functions? Under what conditions is the inequality non-trivial, i.e. when is $C > 0$?

To address these questions we introduce the linear functional

$$J(f_1, ..., f_m) = \frac{\int_H e^{-\mathcal{Q}(x)} \prod_{k=1}^m f_i^{c_k}(B_k x)dx}{\prod_{k=1}^m \left(\int_{H_k} f_i \right)^{c_k}},$$

and assume the convention $0 \cdot \infty = 0$ for the product $\prod_{k=1}^m f_i^{c_k}(B_k x)$. Then, the objective is to study a minimization problem for the functional J. We denote by $\inf_{\mathcal{G}} J$ the infimum of J over $m-$tuples of Gaussian functions (not necessarily centered) and by $\inf_{\mathcal{CG}} J$ the infimum of J over $m-$tuples of centered Gaussian functions.

In this summary we focus on the Gaussian extremizability of geometric versions of inverse Brascamp Lieb inequalities, see the forthcoming Theorem 5.

8.2 Non-degeneracy conditions and Gaussian infimizers

Let B_+ denote the linear map $(B_1, ..., B_{m^+})$, namely

$$\begin{aligned} B_+ : H &\to H_1 \times ... \times H_{m^+}, \\ x &\mapsto (B_1 x, ..., B_{m^+} x). \end{aligned}$$

We introduce the following non-degeneracy assumptions:

(i) $\mathcal{Q}$ is positive definite on $\ker B_+$;

(ii) $\dim H \geq s^+(\mathcal{Q}) + \dim H_1 + ... + \dim H_{m^+}$.

When assumptions (i) and (ii) are not verified, $\inf J$ can only be 0 or $+\infty$, we refer to [BW, Section 2] for a detailed case by case analysis. Hence, the cases in which (i) and (ii) hold are, to some extent, the only non-degenerate cases.

Theorem 2. ([BW, Th. 2.9]) *Assume that* (i), (ii) *hold. Then* $\inf J = \inf_{\mathcal{CG}} J$.

Therefore, assumptions (i) and (ii) allow to compute the sharpest constant in the inverse Brascamp-Lieb inequalities by considering centered Gaussian functions only.

We briefly sketch the main steps of the proof of Theorem 2 following [BW, Subsection 3.3]. We refer to [BW, Section 3] for the precise arguments.

Sketch of the proof. The first step consists in introducing a decomposition of the Gaussian kernel $e^{-\mathcal{Q}}$. This is achieved by relying on the following lemma [BW, Lemma 3.1].

Lemma 3. *Assumptions* (i) *and* (ii) *hold if and only if there exist vector spaces* H_0, H_{m+1}, *linear surjective maps* $B_0 : H \to H_0$, $B_{m+1} : H \to H_{m+1}$, *and positive definite quadratic forms* $\mathcal{Q}_+$ *on* H_0 *and* $\mathcal{Q}_-$ *on* H_{m+1} *satisfying:*

- $(B_0, B_+) : H \to H_0 \times ... \times H_{m^+}$ *is bijective;*
- $\ker B_+ \subset \ker B_{m+1}$*;*
- *for all* $x \in H$, $\mathcal{Q}(x) = \mathcal{Q}_+(B_0 x) - \mathcal{Q}_-(B_{m+1} x)$.

With this notation, let $Q_+ : H_0 \to H_0$, $Q_- : H_{m+1} \to H_{m+1}$ be such that for all $x \in H_0$ $\mathcal{Q}_+(x) = \pi\langle Q_+x, x\rangle$, and for all $y \in H_{m+1}$ $\mathcal{Q}_-(y) = \pi\langle Q_+y, y\rangle$. Let $f_0 : H_0 \to [0, +\infty]$, $f_{m+1} : H_{m+1} \to [0, +\infty]$ be defined as

$$f_0(x) = \sqrt{\det Q_+} e^{-\pi\langle Q_+x,x\rangle}, \qquad f_{m+1}(x) = \sqrt{\det Q_-} e^{-\pi\langle Q_-x,x\rangle}.$$

Fix $c_0 = 1$, $c_{m+1} = -1$. For $k = 1, ..., m$, fix $f_k : H_k \to [0, +\infty]$ to be measurable functions of integral one. Then

$$J(f_1, ..., f_m) = \sqrt{\frac{\det Q_-}{\det Q_+}} \int_H \prod_{k=0}^{m+1} f_k^{c_k}(B_k x) dx. \tag{1}$$

For technical reasons throughout the proof the functions f_k, $k = 1, ..., m$, need to be chosen from some suitable classes of test functions. Generalization to measurable functions is achieved by approximation arguments in the very last step of the proof.
Next, a tuple of centered Gaussian functions on H_k of integral one, g_k, $k = 0, ..., m + 1$, is introduced.
Then, using tools from optimal transport theory one can construct transportation maps T_k which push forward $f_k(x)dx$ onto $f_k(y)dy$. Such transportation maps are used to construct a change of variable $\theta : H \to H$ which is surjective and which is used to rewrite (1) as an integral involving Gaussian functions only.
Then, after some computation the desired sharp lower bound on $J(f_1, ..., f_m)$ is obtained.

8.3 Minimizers for geometric inverse Brascamp-Lieb inequalities

We introduce the following geometric conditions:

(iii) $B_k B_k^* = \mathrm{Id}_{H_k}$, $k = 1, ..., m$;

(iv) $Q + \sum_{k=1}^m B_k^* B_k = \mathrm{Id}_H$;

where Q denotes the self-adjoint map on H such that $\mathcal{Q}(x) = \pi\langle x, Qx\rangle$ for all $x \in H$.

Theorem 4. *Assume that the non-degeneracy assumptions (i) and (ii) are satisfied and that the geometric conditions (iii) and (iv) hold. Then*

$$\inf_{\mathcal{CG}} J = 1.$$

For the proof we refer to [BW, Theorem 4.5] and its proof.

We are now ready to introduce the desired result [BW, Theorem 4.7].

Theorem 5. *Let c_k, $k = 1, ..., m$ and B_k, $k = 1, ..., m$, be as defined in the Introduction and assume that the geometric condition* (iii) *holds. Let $Q : H \to H$ be a symmetric operator. Assume that also the geometric condition* (iv) *and the non-degeneracy assumption* (ii) *hold. Then for all non-negative integrable functions $h_k : H_k \to [0, +\infty]$ such that $\int h_k > 0$ it holds that*

$$\int_H e^{-\pi\langle x,Qx\rangle} \prod_{k=1}^m h_k^{c_k}(B_k x)dx \geq \prod_{k=1}^m \left(\int_{H_k} f_k \right)^{c_k}.$$

Equality holds when for all k and all $y \in H_k$, $h_k(y) = \exp(-\pi|y|^2)$.

Proof. The geometric condition (iv) implies that $Q + \sum_{i=1}^{m^+} c_i B_i^* B_i$ is positive definite and in particular that the restriction of Q to $\ker B_+$ is positive definite. Then, in view of Theorem 2 $\inf J = \inf_{\mathcal{CG}} J$ and Theorem 4 ensures that $\inf_{\mathcal{CG}} J = 1$. $\square$

References

[BW] Barthe, F. and Wolff, P., *Positive Gaussian kernels also have Gaussian minimizers.* Preprint, 2018, arXiv: 1805.02455;

[L] Lieb, E. H., *Gaussian kernels have only Gaussian maximizers.* Invent. Math. 102 (1990), no. 1, pp. 179208.

Valentina Ciccone, University of Bonn
email: `ciccone@math.uni-bonn.de`

9 Near-extremizers for the isoperimetric inequality

10 The Borell-Boué-Dupuis formula

After Joseph Lehec [Leh13] (see also [LehHDR])

A summary written by Giuseppe Negro

Abstract

We discuss the dual approach to the formulas of Borell and of Boué-Dupuis in terms of the relative entropy, due to J. Lehec.

10.1 Introduction

In [Bor00], Borell established the following formula, for arbitrary measurable and bounded below $f\colon \mathbb{R}^n \to \mathbb{R}$, and pointed out its applications to various functional inequalities;

$$\log \int_{\mathbb{R}^n} e^{f(x)}\, d\gamma_n(x) = \sup \left\{ \mathsf{E}\left[f\left(B_1 + \int_0^1 u_s\, ds \right) - \frac{1}{2}\int_0^1 |u_s|^2\, ds \right] \right\}. \quad (1)$$

Here, $d\gamma_n$ is the Gaussian measure on $\mathbb{R}^n$ and $(B_t)_{t\geq 0}$ is a standard n-dimensional Brownian motion; note that, following the probabilistic convention, we denote with subscript t or s the time variable of a stochastic process. The supremum in (1) is taken over all n-dimensional processes $(u_t)_{t\in[0,1]}$ that are progressively measurable[1] and that satisfy $\int_0^1 |u_s|^2\, ds < \infty$ almost surely.

In this note we discuss Lehec's unified approach to (1) and to the more general formula of Boué and Dupuis [BD98], which we will introduce in a moment after establishing the necessary notation.

We consider the *Wiener probability space* $(\mathbb{W}, \mathcal{B}, \gamma)$, where $\mathbb{W}$ denotes the space of those $w \in C([0,1] \to \mathbb{R}^n)$ such that $w(0) = 0$, equipped with the topology of uniform convergence, $\mathcal{B}$ is the Borel σ-algebra and γ is the Wiener measure, defined as the law of B; that is, for all functionals $F\colon \mathbb{W} \to [0,\infty]$,

$$\int_{\mathbb{W}} F(w)\, d\gamma(w) = \mathsf{E}\left[F \circ B\right]. \quad (2)$$

(Note that $t \to B_t$ is almost surely in $\mathbb{W}$, hence the composition $F \circ B$ makes sense). An embedded subspace of $\mathbb{W}$ is the Cameron-Martin space $\mathbb{H}$, whose elements are absolutely continuous paths $u\colon [0,1] \to \mathbb{R}^n$, with $u(0) = 0$ and

$$\|u\|_{\mathbb{H}}^2 := \int_0^1 |\dot{u}_s|^2\, ds < \infty. \quad (3)$$

[1] We won't give the technical definition of this; intuitively, a *progressively measurable* process is one that only depends on past values of the Brownian motion. For such processes it is possible to define a meaningful notion of *stochastic integral*.

We can now state the formula of Boué and Dupuis. For all functionals $F\colon \mathbb{W} \to \mathbb{R}$ that are measurable and bounded below,

$$\log \int_{\mathbb{W}} e^F \, d\gamma = \sup \left\{ \mathsf{E}\left[F(B+U) - \frac{1}{2}\|U\|_{\mathbb{H}}^2 \right] \right\}, \tag{4}$$

where the supremum is taken over all progressively measurable processes U that belong to $\mathbb{H}$ almost surely; such processes are called *drifts*. Borell's formula (1) follows from (4) by letting

$$U_t := \int_0^t u_s \, ds \quad \text{and} \quad F(w) := f(w(1)).$$

We end the introduction by remarking that all these results have a counterpart on the infinite time interval $[0, \infty)$.

10.2 Dual formulation; the entropy

Given a probability measure μ that is absolutely continuous with respect to γ, that is, $d\mu = \rho \, d\gamma$ where ρ is a nonnegative function on $\mathbb{W}$, the *relative entropy* of μ with respect to γ is defined as

$$H(\mu|\gamma) = -\int_{\mathbb{W}} \log(\rho(w)^{-1})\rho(w) \, d\gamma(w); \tag{5}$$

note that $H(\mu|\gamma) \geq 0$ by the Jensen inequality. Conventionally, $H(\mu|\gamma) = \infty$ for those probabilities μ that are not γ-absolutely continuous.

The formula (4) is equivalent to the following theorem. A sketch of its proof will be given in the next section.

Theorem 1. *Let μ be a γ-absolutely continuous probability. Then*

$$H(\mu|\gamma) \leq \frac{1}{2}\mathsf{E}[\|U\|_{\mathbb{H}}^2], \tag{6}$$

for all drifts U such that $B + U$ has law μ. With some technical assumptions on μ, there is a drift that attains the equality.

Theorem 1 is, arguably, the main result of [Leh13]. Its relation to the formula (4) is due to the following convex duality:

$$\log \int_{\mathbb{W}} e^F \, d\gamma = \sup \left\{ \int_{\mathbb{W}} F \, d\mu - H(\mu|\gamma) \right\}, \tag{7}$$

the supremum being over all Borel probabilities μ on $\mathbb{W}$. Thus, for all drift U such that $B+U$ has law μ, Theorem 1 yields

$$\begin{aligned}\log \int_{\mathbb{W}} e^F \, d\gamma &\geq \int_{\mathbb{W}} F \, d\mu - \frac{1}{2}\mathsf{E}\|U\|_{\mathbb{H}}^2 \\ &= \mathsf{E}\left[F(B+U) - \frac{1}{2}\|U\|_{\mathbb{H}}^2\right],\end{aligned} \tag{8}$$

with equality for those U that attain equality in (6). We conclude that the formula (4) follows from Theorem 1.

10.3 Some elements of the proof of Theorem 1.

For arbitrary Borel probability measures μ and λ, and for every measurable map $T\colon \mathbb{W} \to \mathbb{W}$,

$$H(\mu \circ T^{-1} | \gamma \circ T^{-1}) \leq H(\mu|\gamma); \tag{9}$$

this is proved via the Jensen inequality for conditional expectations.

To apply this to the proof of Theorem 1, the tool to use is the formula of Girsanov, which, given an arbitrary drift U, allows us to construct a new probability Q such that $X = B + U$ has law γ under Q; recall that B is a standard Brownian motion under the probability P. We will give more details in a moment. Letting μ denote the law of X under P,

$$H(\mu|\gamma) = H(P \circ X^{-1} | Q \circ X^{-1}) \leq H(P|Q). \tag{10}$$

The probability Q is constructed in terms of the process

$$D_t = \exp\left(-\int_0^t \langle \dot{U}_s, dB_s\rangle - \frac{1}{2}\int_0^t |\dot{U}_s|^2 \, ds\right), \tag{11}$$

with some suitable integrability condition on U; having $\|U\|_{\mathbb{H}}$ bounded suffices. The probability Q is given by $dQ = D_1 dP$; in particular, letting E denote the expectation with respect to P,

$$H(P|Q) = -\mathsf{E}[\log D_1] = \frac{1}{2}\mathsf{E}[\|U\|_{\mathbb{H}}^2],$$

since the stochastic integral $\int_0^t \langle \dot{U}_s, dB_s\rangle$ has expectation 0. Combining this with (10) proves the inequality (6) of Theorem 1. We will not discuss the cases of equality in this short note.

References

[Bor00] C. Borell, *Diffusion equations and geometric inequalities.* Potential Anal. **12** (2000), 49–71;

[BD98] M. Boué and P. Dupuis, *A variational representation for certain functionals of Brownian motion.* Ann. Probab., **26** (1998), pp. 1641–1659.

[Leh13] J. Lehec, *Representation formula for the entropy and functional inequalities.* Ann. Inst. H. Poincaré Probab. Statist. **49** (3) 885 - 899, August 2013.

[LehHDR] J. Lehec, *Processus stocastiques, convexité et inégalités fonctionnelles.* Thése HDR, 2016. `https://hal.archives-ouvertes.fr/tel-01428644`

Giuseppe Negro, University of Birmingham, United Kingdom
email: `g.negro@bham.ac.uk`

11 Short probabilistic proofs of BrascampLieb and Barthe theorems

After J. Lehec [Leh14]

A summary written by Víctor Olmos

Abstract

We give a short proof of the BrascampLieb theorem using a representation formula for certain functionals of the Brownian motion due to Boué and Dupuis. A similar argument is then applied to the Barthe theorem regarding the reversed Brascamp-Lieb inequality.

11.1 Boué-Dupuis formula

Fix a finite time horizon $T > 0$ and let $W = (W_t)_{t\in[0,T]}$ be an n-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in[0,T]})$. Let A be the covariance matrix of W_1 and let $\mathbb{H}$ be the Cameron-Martin space associated with W, that is, the Hilbert space of absolutely continuous functions $u : [0,T] \longrightarrow \mathbb{R}^n$ with

$$\|u\|_{\mathbb{H}} := \left(\int_0^T \langle A^{-1}u'(t), u'(t)\rangle \, dt \right)^{1/2} < \infty.$$

An adapted process U such that almost every path belongs to $\mathbb{H}$ is called a *drift*. In 1998, Boué and Dupuis proved a very useful representation formula that allows us to compute the expectation of certain functionals of W.

Proposition 1. *Let $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a Borel measurable function bounded from below. Then, with the notation above,*

$$\log \mathbb{E}\big[e^{g(W_T)}\big] = \sup_{U \ drift} \mathbb{E}\left[g(W_T + U_T) - \frac{1}{2}\|U\|_{\mathbb{H}} \right].$$

The important detail of this formula that will be useful for us is that the right hand side, once a drift has been chosen, is linear in g. Hence appropriately decomposing g as a sum of certain g_i and then applying the proposition again we will obtain a product of expectations, precisely what is needed for the Brascamp-Lieb inequality.

11.2 The direct inequality

Let $(c_1, B_1), \ldots, (c_m, B_m)$ be a Brascamp-Lieb datum on $\mathbb{R}^n$, that is, each c_i is a positive number and each B_i is a surjective linear mapping from $\mathbb{R}^n$ onto $\mathbb{R}^{n_i}$ for some $n_i \in \mathbb{N}$. The Brascamp-Lieb constant associated with that datum is the smallest constant C such that

$$\int_{\mathbb{R}^n} \prod_{i=1}^m f_i(B_i x)^{c_i}\, dx \leq C \prod_{i=1}^m \left(\int_{\mathbb{R}^{n_i}} f_i(x)\, dx \right)^{c_i} \tag{1}$$

for all non-negative integrable functions $f_i : \mathbb{R}^{n_i} \longrightarrow \mathbb{R}$. From a standard scaling argument it is easy to see that a necessary condition to have finiteness of the constant C is that

$$\sum_{i=1}^m c_i n_i = n.$$

We will assume this condition without further mention. The Brascamp-Lieb theorem states that this inequality is saturated by Gaussian functions.

Theorem 2. *For a Brascamp-Lieb datum as above, assume that there exists a positive definite matrix A such that*

$$A^{-1} = \sum_{i=1}^m c_i B_i^* (B_i A B_i^*)^{-1} B_i. \tag{2}$$

Then the Brascamp-Lieb constant associated with the datum is

$$C = \left(\frac{\det(A)}{\prod_{i=1}^m \det(B_i A B_i^*)^{c_i}} \right)^{1/2},$$

and equality in (1) *is obtained for the Gaussian functions*

$$f_i(x) = \exp\left(-\frac{1}{2} \langle (B_i A B_i^*)^{-1} x, x \rangle \right), \quad 1 \leq i \leq m.$$

Sketch of the proof. The equality case follows by simply computing both terms of (1). Then fix non-negative measurable functions $f_1, \ldots, f_n$ and let W be a Brownian motion with covariance matrix A as in the previous section. Define the functions $g_i = \log(f_i + \delta)$ for some $\delta > 0$ fixed and $g = \sum_{i=1}^m c_i g_i \circ B_i$. By Proposition 1, for $\varepsilon > 0$ there exists a drift U such that

$$\log \mathbb{E}\left[e^{g(W_T)} \right] \leq \sum_{i=1}^m c_i \mathbb{E}[g_i(B_i W_T + B_i U_T)] - \frac{1}{2} \mathbb{E}[\|U\|_{\mathbb{H}}] + \varepsilon.$$

Since the process $B_i W$ is again a Brownian motion on $\mathbb{R}^{n_i}$ with covariance matrix $A_i = B_i A B_i^*$, we can define the respective Cameron-Martin spaces $\mathbb{H}_i$ and express $\|\cdot\|_{\mathbb{H}}$ with respect to the $\|\cdot\|_{\mathbb{H}_i}$. Using again Proposition 1 we can show that

$$\log \mathbb{E}\left[e^{g(W_T)}\right] \leq \sum_{i=1}^{m} c_i \log \mathbb{E}\left[e^{g_i(B_i W_T)}\right] + \varepsilon.$$

Finally we have $\prod_{i=1}^m (f_i \circ B_i)^{c_i} \leq e^g$, so sending $\varepsilon, \delta \to 0$ we get

$$\mathbb{E}[f(W_T)] \leq \prod_{i=1}^{m} \mathbb{E}[f_i(B_i W_T)]^{c_i}.$$

It only remains to use that W_T is a Gaussian random variable with covariance TA to compute both sides of the above expression, and then let $T \to \infty$. $\square$

Example 3. *The Brascamp-Lieb inequality can be used for example to find the best constant in Young's convolution inequality, that is, the minimum positive constant C such that*

$$\|f * g\|_{L^r(\mathbb{R})} \leq C \|f\|_{L^p(\mathbb{R})} \|g\|_{L^q(\mathbb{R})}$$

for all $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$, with $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$.

By duality, Young's inequality is equivalent to

$$\int_{\mathbb{R}^2} f(x+y)^{1/p} g(y)^{1/q} h(x)^{1-1/r}\, dy\, dx \leq C \left(\int_{\mathbb{R}} f\right)^{1/p} \left(\int_{\mathbb{R}} g\right)^{1/q} \left(\int_{\mathbb{R}} h\right)^{1-1/r},$$

so C is the Brascamp-Lieb constant associated with the data $(1/p, (1,1))$, $(1/q, (0,1))$, $(1-1/r, (1,0))$. A computation shows that

$$C = \left(\frac{p^{1/p}\; q^{1/q}\; r'^{1/r'}}{p'^{1/p'}\; q'^{1/q'}\; r^{1/r}}\right)^{1/2}.$$

11.3 The reversed inequality

Again, given a Brascamp-Lieb datum as before, the reversed Brascamp-Lieb constant associated with it is the smallest C_r such that for every family of non-negative measurable functions $f_1, \dots, f_m, f$ satisfying

$$\prod_{i=1}^{m} f_i(x_i)^{c_i} \leq f\left(\sum_{i=1}^{m} c_i B_i^* x_i\right) \tag{3}$$

for every $(x_1, \ldots, x_m) \in \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_m}$, we have

$$\prod_{i=1}^{m} \left(\int_{\mathbb{R}^{n_i}} f_i(x_i)\, dx_i \right)^{c_i} \leq C_r \int_{\mathbb{R}^n} f(x)\, dx. \tag{4}$$

Again, under the same conditions as in the previous theorem, Gaussian functions saturate this inequality.

Lemma 4. *Let $A_1, \ldots, A_m$ be positive definite matrices on $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_m}$, and define $A = \left(\sum_{i=1}^{n} c_i B_i^* A_i^{-1} B_i\right)^{-1}$. Then for all $x \in \mathbb{R}^n$ we have*

$$\langle Ax, x \rangle = \inf \left\{ \sum_{i=1}^{m} c_i \langle A_i x_i, x_i \rangle : \sum_{i=1}^{m} c_i B_i^* x_i = x \right\}.$$

Theorem 5. *Let $(c_1, B_1), \ldots, (c_m, B_m)$ be a Brascamp-Lieb datum such that there exists a positive definite matrix A satisfying* (2)*. Then the associated reversed Brascamp-Lieb constant is*

$$C_r = \left(\frac{\det(A)}{\prod_{i=1}^{m} \det(B_i A B_i^*)^{c_i}} \right)^{1/2}.$$

Moreover, we have equality in (4) *for the Gaussian functions*

$$f_i(x_i) = \exp\left(-\frac{1}{2} \langle B_i A B_i^* x_i, x_i \rangle \right), \quad x_i \in \mathbb{R}^{n_i}, \quad 1 \leq i \leq m,$$

$$f(x) = \exp\left(-\frac{1}{2} \langle Ax, x \rangle \right), \quad x \in \mathbb{R}^n.$$

Sketch of the proof. The proof is similar to the one of the previous theorem. Pick non-negative measurable $f_1, \ldots, f_m, f$ satisfying (3) and assume that the f_i are bounded, and let $g_i = \log(f_i + \delta)$ as before. One can choose $c, C > 0$ such that $g = \log(f + C\delta^c)$ satisfies

$$\sum_{i=1}^{m} c_i g_i(x_i) \leq g\left(\sum_{i=1}^{m} c_i B_i^* x_i \right).$$

Again, take a Brownian motion $(W_t)_{t \in [0,T]}$ with covariance matrix A. Using the Cameron-Martin spaces associated with $(B_i A B_i^*)^{-1}$ and applying Proposition 1 and Lemma 4, we can prove as before that

$$\sum_{i=1}^{m} c_i \log \mathbb{E}\left[e^{g_i(A_i^{-1} B_i W_T)} \right] \leq \log \mathbb{E}\left[e^{g(A^{-1} W_T)} \right].$$

Finally, since $f_i \leq e^{g_i}$ and $e^g = f + C\delta^c$, letting $\delta \to 0$ we obtain

$$\prod_{i=1}^{m} \mathbb{E}[f_i(A_i^{-1}B_iW_T)]^{c_i} \leq \mathbb{E}[f(A^{-1}W_T)].$$

Letting $T \to \infty$ proves the first part of the theorem. The equality case is proven by a simple computation using Lemma 4. □

References

[Leh14] Lehec, J., *Short probabilistic proof of the Brascamp-Lieb and Barthe theorems.* Canad. Math. Bull. 57.3 (2014), pp. 585597.

[BD98] Boué, M. and Dupuis, P., *A variational representation for certain functionals of Brownian motion.* Ann. Probab. 26.4 (1998), pp. 16411659.

Víctor Olmos, Universidad Nacional de Educación a Distancia
email: volmos21@alumno.uned.es

12 Euclidean forwardreverse BrascampLieb inequalities: finiteness, structure, and extremals

After T. Courtade, J. Liu [TL]

A summary written by Jaume de Dios Pont

Abstract

We describe and provide a proof a proof of the forward-reverse Brascamp-Lieb inequality using the Boue-Dupuis formula for Gaussian integrals. We provide necesssary and sufficient condigions for Extremizability of the forward-reverse Brascamp-Lieb inequality.

12.1 Introduction

Let $E = \bigoplus_{i=1}^k E_i$ and $F = \bigoplus_{j=1}^m F^j$ Euclidean spaces, and $B_{ij} : E_i \to F_j$ linear maps. Given non-negative lists of real numbers $\mathbf{c} := (c_1, \dots c_k), \mathbf{d} := (d_1, \dots d_m)$ we are concerned with the best constant D such that:

$$\prod_{i=1}^k \left(\int_{E_i} f_i \right)^{c_i} \le e^D \prod_{j=1}^m \left(\int_{F^j} g_j(x_i) \right)^{d_j} \tag{1}$$

for all measurable functions $f_i : E_i \to \mathbb{R}^+$, $g_i : F^i \to \mathbb{R}^+$ that satisfy the inequality

$$\prod_{i=1}^k f_i^{c_i}(x_i) \le \prod_{j=1}^m g_j^{d_j} \left(\sum_{i=1}^k c_i B_{ij} x_i \right) \tag{2}$$

for all $(x_1, \dots x_k) \in \bigoplus_{i=1}^k E_i$.

We will call $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ a *datum* for the forward-reverse BL inequality, and define $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ as the smallest constant D such that (1) holds whenever (2) holds. A necessary condition for $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ to be finite is the scaling condition

$$\sum_{i=1}^k c_i \dim E_i = \sum_{j=1}^m d_j \dim E_j. \tag{3}$$

As in the Brascamp-Lieb case, the necessity comes from considering dilations of functions f_i, g_i satisfying (1) none of which are zero a.e.

This family of inequalities extend simultaneously the Brascamp-Lieb inequalities (when $k = 1$) and Barthe's reverse Brascamp-Lieb inequality (when $m = 1$).

Finiteness and Gaussian extremisability

The first result concerns the situations in which Gaussians extremize forward-reverse Brascamp-Lieb inequalities. It states that whenever there is an extremizer amongst all Gaussian functions, this must be an extremizer amongst all functions as well.

Theorem 1. *Let $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ be the largest constant $\mathbf{D}$ in equation (2) under the further constraint that f_i, g_j are restricted to be centered gaussians. If a datum $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ is Gaussian extremizable (in the sense that $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ is attained by specific f_i, g_j) then*

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D(\mathbf{c}, \mathbf{d}, \mathbf{B})$$

It han been long known that the sharp constants for the usual Brascamp-Lieb inequalities and for Barthe's reverse were the same (at least when written in the formulation of 1). This has a generalization to the forward-reverse Brascamp-Lieb inequality by defining $(\mathbf{B}^*)_{ij} = B_{j,i}$, in the form of:

Theorem 2. *If $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ is gaussian-extremizable so is $(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$. Moreover*

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D_g(\mathbf{d}, \mathbf{c}, \mathbf{B}^*),$$

regardless of whether the data are gaussian-extremizable.

Gaussian extremizability can be deduced from the operators B_{ij} themselves. Before doing so, however, we must introduce some notation:

We will denote by $\Lambda_{\mathbf{c}}$ the matrix $\bigotimes_{i=1}^k \mathbf{c}_i I_{E_i}$. Analogously, $\Lambda_{\mathbf{d}}$ will denote the matrix $\bigotimes_{i=1}^k \mathbf{d}_i I_{V_i}$. Given $V_1, \dots V_j$ symmetric matrices in $E_i \otimes E_i$, we will denote by $V_{\mathbf{c}}$ the matrix $\bigotimes_{i=1}^k \mathbf{c}_i I_{E_i}$.

Last, given positive definite matrices $A_1, \dots A_k$ in $A_i : E_i \times E_i \to \mathbb{R}$, we define $\Pi(A_1, \dots A_k)$ as the set of positive-definite bilinear forms in $E = \bigoplus_{i=1}^k E_i$ that are equal to A_i that when restricted to each E_i.

Theorem 3. *Let $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ be a Brascamp-Lieb datum. The following are equivalent:*

1. $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ *is Gaussian-extremizable.*

2. *There exist invertible linear maps* $(\alpha_i)_{i=1\ldots k}$*,* $(\beta_j)_{j=1\ldots m}$ *such that with* $\tilde{\mathbf{B}} = (\beta_j B_{ij} \alpha_i)_{ij}$*, the datum* $(\mathbf{c}, \mathbf{d}, \tilde{\mathbf{B}})$ *is geometric.*

3. *There exist positive definite matrices* $(V_i)_{i=1,\ldots k} \in S^+(E_i)$ *and* $\Pi \in \Pi(V_1^{-1}, \ldots V_k^{-1})$ *such that:*

$$\sum_{j=1}^{m} d_j \Lambda_{\mathbf{c}} B_j^* (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1} B_j \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}} \tag{4}$$

Moreover, from the last statement there an explicit construction of the extremizers and optimal constant from Π, V_i can be deduced.

Key ideas of the proofs

Theorem 1 depends on the assertion [1. $\implies$ 3] of Theorem 3, and is proven in a similar fashion to the proof of forward and reverse Brascamp-Lieb using the Boué-Dupuis formula. The existence of Π given in Theorem 3, Item 3. gives the covariance of the Brownian motion to which the Boué-Dupuis formula is applied.

The proof of Theorem 2 is self contained: One shows that if $f_i = \exp(-x \cdot U_i \cdot x), g_j = \exp(-x \cdot V_j \cdot x)$ are an admissible family of Gaussians for $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ in the sense that (1) holds, then $f_i = \exp(-x \cdot V_i^{-1} \cdot x), g_j = \exp(-x \cdot U_i^{-1} \cdot x)$ is admissible for $(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$, with the same constant in (2). This is shwon by explicitly writing condition 1 for Gaussians explicitly, which becomes a problem about positive definiteness of certain matrices.

The proof of Theorem 3, on the other hand, is significantly more involved in the forward-reverse scenario.

- [2. $\implies$ 3.] By definition, all Geometric data satisfy 3., by directly setting $V_i = \Pi = Id$.

- [3. $\implies$ 2.] can be shown by constructing an explicit transformation using (4).

- [3. $\implies$ 1.] In fact follows from the proof of Theorem 3: If (4) holds, one may use the Boué-Dupuis formula to construct the explicit Gaussian extremizers.

- The proof of [1. $\implies$ 3.] is done by a careful transformation of the constrained problem (minimize D in (2) for all gaussians with the constraint given by (1)) into an equivalent unconstrained problem. The result then follows by setting the gradient of the unconstrained problem equal to zero. The main ingredients of the transformation is a max-min inequality derived using the Fenchel-Rockafellar duality.

References

[TL] Courtade , T. and Liu, J. *Euclidean forwardreverse BrascampLieb inequalities: finiteness, structure, and extremals* In: J. Geom. Anal. 31.4 (2021), pp. 33003350.

Jaume de Dios Pont, UCLA
email: jdedios@math.ucla.edu

13 Functional Erhard inequality

After C. Borell [Bor03],[Bor07]

A summary written by Lukas Mauth

Abstract

If C is a domain in $\mathbb{R}^n$, we denote the Brownian exit time of C by T_C. For given domains C and D in $\mathbb{R}^n$ we will establish an upper bound for the distribution function of T_{C+D}, when the distribution functions of T_C and T_D are known. Furthermore, we will give a proof of the classical functional erhard inequality.

13.1 Introduction

For two subsets C and D of $\mathbb{R}^n$ we define the Minkowski sum as

$$C + D := \{x + y | x \in C \text{ and } y \in D\}.$$

Moreover, we define for $\alpha > 0$ the dilation $\alpha C = \{\alpha x | x \in C\}$. Throughout $W = (W(t))_{t \geq 0}$ denotes Brownian in $\mathbb{R}^n$ and if C is a domain in $\mathbb{R}^n$,

$$T_C = T_C^W := \inf\{t > 0 | W(t) \notin C\}$$

is called the exit time from C. Below the notation $P_x[\cdot]$ or $E_x[\cdot]$ indicates that Brownian motion starts at the point x at time zero.

If H is an open affine half-space in $\mathbb{R}^n$, the Bachelier formula (see [KaSh]) for the distribution of the maximum of real-valued Brownian motion yields

$$P_x[T_H > t] = \Psi\left(\frac{d(x, H^c)}{\sqrt{t}}\right), t > 0, x \in H,$$

where $d(x, H^c) = \min_{y \in \mathbb{R}^n \setminus H} |x - y|$ and

$$\Psi(r) = 2\int_0^r e^{-\frac{\lambda^2}{2}} \frac{d\lambda}{\sqrt{2\pi}}.$$

We the following the notation for the expected Brownian exit times.

$$E_x[T_C] = \int_0^\infty P_x[T_C > t] dt.$$

The result on the distribution function of the Brownian exit time regarding Minkowski sums, which we aim to prove, reads as follows.

Theorem 1. *Let C and D be domains in $\mathbb{R}^n$ and let $f : C \to [0,1]$, $g : D \to [0,1]$ and $h : C + D \to [0,1]$ be continuous functions such that for all $x \in C$ and $y \in D$*

$$\Psi^{-1}(h(x+y)) \geq \Psi^{-1}(f(x)) + \Psi^{-1}(g(y)).$$

Then, for all $x \in C, y \in D$ and $t > 0$

$$\Psi^{-1}(E_x[h(W(t)); T_{C+D} > t])$$
$$\geq \Psi^{-1}(E_x[f(W(t)); T_C > t]) + \Psi^{-1}(E_y[g(W(t)); T_D > t]).$$

In particular

$$\Psi^{-1}(P_{x+y}[T_{C+D} > t]) \geq \Psi^{-1}(P_x[T_C > t]) + \Psi^{-1}(P_y([T_D > t])),$$

where the equality gets exhausted if C and D are parallel affine half-spaces.

Our second goal is to establish the classical functional Erhard inequality. To that end we denote by γ_n the centered gaussian measure on $\mathbb{R}^n$. If there is no ambiguity about the dimension n, we will drop it from the notation. We set for $r \in \mathbb{R}$

$$\Phi(r) := \int_0^\infty e^{-\frac{\lambda^2}{2}} \frac{d\lambda}{\sqrt{2\pi}}.$$

Theorem 2 (Functional Erhard inequality)**.** *Let C and D be any Borel sets in $\mathbb{R}^n$. Then we have for all $0 < \theta < 1$*

$$\Phi^{-1}(\gamma(\theta C + (1-\theta)D)) \geq \theta\Phi^{-1}(\gamma(C)) + (1-\theta)\Phi^{-1}(\gamma(D)).$$

We adopt here the convention that $\infty - \infty = -\infty + \infty = -\infty$.

The Erhard inequality lies at the heart of a large hierarchy of inequalities in the Gaussian setting. For instance it implies the well known Gaussian isoperimetric inequality, which in turn imples numerous geometric and analytic inequalities for Gaussian measures.

13.2 Methods

These two theorems go particularly well together, because their proofs are based on the same method. We will just illustrate the proof of the functional Erhard-inequality from which a proof of Theorem 1 will follow by simple modfications.

At the root of this is a geometric proof of the Erhard inequality in the special case, where C and D are convex. Erhards original proof used a Gaussian analogue of Steiner symmetrization, a classical tool in measure theory, which is for example used to prove the isodiametric inequality. The downside of this proof is that it relies heavily on the convexity of the sets and is thus in nature rigid. Hence, the question if the inequality holds for general Borel sets remained open for a long time.

It was then C. Borell who came up with an entirely different approach relying more on analytic properties than geometric ones.

The starting point of the argument is the functional form of Theorem 2, which states that if given functions f, g, and h satisfy

$$\Phi^{-1}(h(\lambda x + \mu y)) \geq \lambda \Phi^{-1}(f(x)) + \mu \Phi^{-1}(g(y)), \tag{1}$$

then we find

$$\Phi^{-1}\Big(\int_{\mathbb{R}^n} h d\gamma_n\Big) \geq \lambda \Phi^{-1}\Big(\int_{\mathbb{R}^n} f d\gamma_n\Big) + \mu \Phi^{-1}\Big(\int_{\mathbb{R}^n} g d\gamma_n\Big). \tag{2}$$

The statement of Theorem 2 can be recovered by setting $f = \chi_C, g = \chi_D$ and $h = \chi_{C+D}$.

Thus, its proof is reduced to establishing (2). Borell now considers the function

$$C(t, x, y) := \Phi^{-1}(u_h(t, \lambda x + \mu y)) - \lambda \Phi^{-1}(u_f(t, x)) - \mu \Phi^{-1}(u_g(t, y)),$$

where $u_f(t, \cdot)$ denotes the heat semigroup

$$u_f(t, x) := \int_{\mathbb{R}^n} f(x + \sqrt{t} z) \gamma_n(dz).$$

Our assumption (1) is now equivalent to $C(0, x, y) \geq 0$, while the erhard inequality (2) can be rewritten as $C(1, 0, 0) \geq 0$.

The crucial point is that the function $C(t, x, y)$ is the solution of a certain parabolic equation in the domain $[0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n$. Hence, we can employ the weak parabolic maximum principle (see [Evans]), implying that $\min C = \min C(0, \cdot, \cdot) \geq 0$ and thus proving (2) as a direct consequence.

13.2.1 Further results

From the last proof one could expect to exhibit information when equality in (1) is achieved, since a approach to substitute the weak maximum principle by the strong version seems natural and in reach. Indeed, with a little more care we can prove the following more general result, which proof relies on the same method used to prove Theorem 2.

Let F be a seperable Fréchet space and γ a centered Gaussian measure on F such that each bounded linear functional on F has a centered Gaussian distribution. Let $\mathcal{B}(F)$ the Borel $\sigma-$algebra on F. The definition of Minkowski sum carries over canonically to this more general setting.

Theorem 3. *Suppose $\alpha, \beta > 0$ are given. Then the inequality*

$$\Phi^{-1}(\gamma(\theta C + (1-\theta)D)) \geq \theta\Phi^{-1}(\gamma(C)) + (1-\theta)\Phi^{-1}(\gamma(D)). \tag{3}$$

is valid for all $C, D \in \mathcal{B}(F)$ if

$$\alpha + \beta \geq 1 \text{ and } |\alpha - \beta| \leq 1. \tag{4}$$

Moreover, if γ is not a Dirac measure at origin and (3) *is valid for all $C, D \in \mathcal{B}(F)$, then* (4) *holds.*

Equality occurs in (3) *if C and D are parallel affine half-spaces. If in addition $\alpha + \beta = 1$, then equality in* (3) *occurs if C is convex and $D = C$.*

References

[Bor03] C. Borell, *The Ehrhard inequality.* In: C. R. Math. Acad. Sci. Paris 337.10 (2003), pp. 663 666.

[Bor07] C. Borell, *Minkowski sums and Brownian exit times.* In: Ann. Fac. Sci. Toulouse Math. (6) 16.1 (2007), pp. 3747.

[Evans] L. C. Evans, *Partial Differential Equations,* volume 19 of *Graduate Studies in Mathematics.* AMS, 1997.

[KaSh] I. Karatzas, S. E. Shreve, *Brownian Motion and Stochastic Calculus.* Second Edition, Springer 1991.

Lukas Mauth, University of Bonn
email: `s6lumaut@uni-bonn.de`

14 Functional Ehrhard inequality via stochastic minimax

After "The Borell-Ehrhard game" by Ramon van Handel

A summary written by Ilseok Lee

Abstract

This summary of [HR] gives a proof of functional Ehrhard inequality using a game-theoretic mechanism:a minimax variational principle for Brownian motion.

14.1 The Borell-Ehrhard Game

14.1.1 Setting and main result

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ be a probability space with a complete and right-continuous filtration, and let $\{W_t\}$ be a standard n-dimensional Brownian motion adapted to the filtration $\{\mathcal{F}_t\}$. Denote the standard Gaussian measure γ_n on $\mathbb{R}^n$ as

$$\gamma_n(dx) = e^{-|x|^2/2} \frac{dx}{(2\pi)^{n/2}},$$

and $\Phi(x) := \gamma_1((-\infty, x])$.

Definition 1. *A control is a progressively measurable n-dimensional process $\beta = \{\beta_t\}_{t\in[0,1]}$. Denote $\mathcal{C}$ the family of all controls such that $\|\beta\|_\infty < \infty$*

Definition 2. *A (Elliott-Kalton) strategy is a map $\alpha : \mathcal{C} \to \mathcal{C}$ such that for every $t \in [0,1]$ and $\beta, \beta' \in \mathcal{C}$ such that $\beta_s(\omega) = \beta'_s(\omega)$ for a.e $(s,\omega) \in [0,t]\times\Omega$, we have $\alpha_s(\beta)(\omega) = \alpha_s(\beta')(\omega)$ for a.e $(s,\omega) \in [0,t] \times \Omega$. Denote by $\mathcal{S}$ the family of all strategies such that* $\sup\{\|\alpha(\beta)\|_\infty : \|\beta\|_\infty \le R\} < \infty$ *for all $R < \infty$*

Theorem 3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be bounded and uniformly continuous, and define*

$$J_f[\alpha, \beta] := \mathbf{E}\left[\int_0^1 e^{-\frac{1}{2}\int_0^t \|\beta_s\|^2 ds} \langle \alpha_t, \beta_t\rangle dt + e^{-\frac{1}{2}\int_0^1 \|\beta_t\|^2 dt} f\left(W_1 + \int_0^1 \alpha_t dt\right)\right]$$

for $\alpha, \beta \in \mathcal{C}$. Then

$$\Phi^{-1}\left(\int \Phi(f) d\gamma_n\right) = \sup_{\alpha\in\mathcal{S}} \inf_{\beta\in\mathcal{C}} J_f[\alpha(\beta), \beta] = \inf_{\alpha\in\mathcal{S}} \sup_{\beta\in\mathcal{C}} J_f[\alpha(\beta), \beta].$$

The process $\alpha(\beta)$ is progressively measurable by construction, so the above theorem does not have measurability issues.

14.1.2 The Borell PDE

Throughout the proof, we will assume without loss of generality that f is bounded, smooth, and has bounded derivatives of all orders. Once the result is proved in this case, the conclusion is readily extended to functions f that are only bounded and uniformly continuous (see section 8.2 in [FG])

For $(t,x) \in [0,1] \times \mathbb{R}^n$, define $u(t,x) := \mathbf{E}[\Phi(f(W_1 - W_t + x))]$ and $v(t,x) := \Phi^{-1}(u(t,x))$ where $u(t,x)$ and $v(t,x)$ solve the following PDEs respectively:

$$\frac{\partial u}{\partial t} + \frac{1}{2}\Delta u = 0, u(1,x) = \Phi(f(x))$$

and

$$\frac{\partial v}{\partial t} + \frac{1}{2}\Delta v - \frac{1}{2}v\|\nabla v\|^2 = 0, v(1,x) = f(x)$$

where the second PDE above is called Borell's PDE.

Lemma 4. *Let c be a constant such that $2c \geq \sup_x f(x)$. Then*

$$-\frac{1}{2}v\|\nabla v\|^2 = \sup_{a\in\mathbb{R}^n} \inf_{b\in\mathbb{R}^n} \left\{ \langle a + cb, \nabla v + b\rangle - \frac{1}{2}v\|b\|^2 \right\},$$

where the optimizer $a^* = (c-v)\nabla v$, $b^* = -\nabla v$ is a saddle point.

Proof. Denote $H(a,b) := \langle a+cb, \nabla v+b\rangle - \frac{1}{2}v\|b\|^2$ for $a,b\in\mathbb{R}^n$. Then simple computations imply that $H(a,b^*) = -\frac{1}{2}v\|\nabla v\|^2$ and $H(a^*,b) = \frac{1}{2}(2c-v)\|b+\nabla v\|^2 - \frac{1}{2}v\|\nabla v\|^2$. But note that as $2c \geq f$, we have $2c - v \geq 0$ by the definition of v. Therefore $\sup_a \inf_b H(a,b) \leq \sup_a H(a,b^*) = -\frac{1}{2}v\|\nabla v\|^2 = \inf_b H(a^*,b) \leq \sup_a \inf_b H(a,b)$. □

Proof of Theorem 3. Fix $2c \geq f$, and consider the stochastic differential equation $dX_t^\beta = (c - v(t,X_t^\beta))\nabla v(t,X_t^\beta)dt + c\beta_t dt + dW_t$ and $X_0^\beta = 0$ for $\beta \in \mathcal{C}$. The classical Picard scheme(Theorem 4.8 in [LR]) implies that since $(c-v)\nabla v$ is smooth with bounded derivatives, there exists a unique progressively measurable map $F : \mathbb{W} \to \mathbb{W}$ such that $X^\beta = F[\{\int_0^t c\beta_s ds + W_t\}]$, where $\mathbb{W}$ denotes the space of continuous paths with its canonical filtration. Consider $\alpha_t^*(\beta) := (c - v(t,X_t^\beta))\nabla v(t,X_t^\beta)$ which depends causally on β by construction and is uniformly bounded. Then $\alpha^* \in \mathcal{S}$(see Definition 2).

Applying Itô's formula to the process $t \mapsto e^{-\frac{1}{2}\int_0^t \|\beta_s\|^2 ds} v(t,X_t^\beta)$ gives

$$\int_0^1 e^{-\frac{1}{2}\int_0^t \|\beta_s\|^2 ds}\langle \alpha_t^*(\beta) + c\beta_t, \beta_t\rangle dt + e^{-\frac{1}{2}\int_0^1 \|\beta_t\|^2 dt} f(X_1^\beta)$$

$$= v(0,0) + \int_0^1 e^{-\frac{1}{2}\int_0^t \|\beta_s\|^2 ds}\langle \nabla v(t,X_t^\beta), dW_t\rangle$$

$$+\int_0^1 e^{-\frac{1}{2}\int_0^t\|\beta_s\|^2ds}\Big\{\frac{\partial v}{\partial t}(t,X_t^\beta)+\frac{1}{2}\Delta v(t,X_t^\beta)$$
$$+\langle\alpha_t^*(\beta)+c\beta_t,\nabla v(t,X_t^\beta)+\beta_t\rangle-\frac{1}{2}v(t,X_t^\beta)\|\beta_t\|^2\Big\}dt.$$

Lemma 4 and Borell's PDE imply that the last integral in the above expression is nonnegative. Since ∇v is bounded, the Brownian integral is a martingale, so for every $\beta\in\mathcal{C}$

$$\begin{aligned}v(0,0)&\le \mathbf{E}\left[\int_0^1 e^{-\frac{1}{2}\int_0^t\|\beta_s\|^2ds}\langle\alpha_t^*(\beta)+c\beta_t,\beta_t\rangle dt+e^{-\frac{1}{2}\int_0^1\|\beta_t\|^2dt}f(X_1^\beta)\right]\\&=J_f[\alpha^*(\beta)+c\beta,\beta]\end{aligned}$$

Thus $\Phi^{-1}\left(\int\Phi(f)d\gamma_n\right)=v(0,0)\le\sup_{\alpha\in\mathcal{S}}\inf_{\beta\in\mathcal{C}}J_f[\alpha(\beta),\beta]$.

Fix any $\alpha\in\mathcal{S}$ and a time step $\delta=N^{-1}(N\ge 1)$. For $t\in[0,\delta)$, let $\beta_t:=-\nabla v(0,0)$. Extend β on $[0,1]$ as follows: suppose that β has been defined on $[0,k\delta)$. Let $\beta_t^k:=\beta_t 1_{[0,k\delta)}(t)$. Define $\beta_t:=-\nabla v\left(k\delta,W_{k\delta}+\int_0^{k\delta}\alpha_s(\beta^k)ds\right)$ for $t\in[k\delta,(k+1)\delta)$. Thus $\beta\in\mathcal{C}$(arbitrarily choose $\beta_1:=0$).

Consider a process $\{X_t\}$ defined by $X_t:=W_t+\int_0^t\alpha_s(\beta)ds$. Definition 2 implies that since $\beta_t=\beta_t^k$ for all $t\in[0,k\delta)$, $\alpha_s(\beta)(\omega)=\alpha_s(\beta^k)(\omega)$ for a.e. $(s,\omega)\in[0,k\delta)\times\Omega$. Also $\beta_t=-\nabla v(k\delta,X_{k\delta})$ for every $t\in[k\delta,(k+1)\delta)$ a.s.

Applying Itô's formula to $t\mapsto e^{-\frac{1}{2}\int_0^t\|\beta_s\|^2ds}v(t,X_t)$ as in above gives $J_f[\alpha(\beta),\beta]=v(0,0)+\mathbf{E}[\Gamma]$ where

$$\Gamma:=\int_0^1 e^{-\frac{1}{2}\int_0^t\|\beta_s\|^2ds}\Big\{\frac{1}{2}v(t,X_t)(\|\nabla v(t,X_t)\|^2-\|\beta_t\|^2)+\langle\alpha_t(\beta),\nabla v(t,X_t)+\beta_t\rangle\Big\}dt.$$

As v is bounded and has bounded derivatives of all orders and $\|\nabla v(t,X_t)\|^2-\|\beta_t\|^2=\langle\nabla v(t,X_t)-\beta_t,\nabla v(t,X_t)+\beta_t\rangle$,

$$\Gamma\le C\sum_{k=0}^{N-1}\int_{k\delta}^{(k+1)\delta}(1+\|\alpha_t(\beta)\|)(\delta+\|X_t-X_{k\delta}\|)dt$$

for a constant C depending on f only. Since $\mathbf{E}\|X_t-X_{k\delta}\|\le\mathbf{E}\|W_t-W_{k\delta}\|+\delta\|\alpha(\beta)\|_\infty\le\sqrt{n\delta}+\delta\|\alpha(\beta)\|_\infty$ for $t\in[k\delta,(k+1)\delta]$,

$$\mathbf{E}[\Gamma]\le C(1+K)(\sqrt{n\delta}+(1+K)\delta)\le C'(1+K)^2\sqrt{\delta},$$

where $K:=\sup\{\|\alpha(\beta')\|_\infty:\|\beta'\|_\infty\le\|\nabla v\|_\infty\}<\infty$ by definition as $\alpha\in\mathcal{S}$, and where C' depends only on f. Thus $\inf_{\beta'\in\mathcal{C}}J_f[\alpha(\beta'),\beta']\le J_f[\alpha(\beta),\beta]\le v(0,0)+C'(1+K)^2\sqrt{\delta}$ which implies

$$\sup_{\alpha\in\mathcal{S}}\inf_{\beta\in\mathcal{C}}J_f[\alpha(\beta),\beta]\le v(0,0)=\Phi^{-1}\left(\int\Phi(f)d\gamma_n\right)$$

Thus combining above results gives

$$\Phi^{-1}\left(\int \Phi(f)d\gamma_n\right) = \sup_{\alpha\in\mathcal{S}}\inf_{\beta\in\mathcal{C}} J_f[\alpha(\beta),\beta]$$

Also

$$\Phi^{-1}\left(\int \Phi(f)d\gamma_n\right) = -\Phi^{-1}\left(\int \Phi(-f)d\gamma_n\right) = -\sup_{\alpha\in\mathcal{S}}\inf_{\beta\in\mathcal{C}} J_{-f}[\alpha(\beta),\beta]$$

$$= \inf_{\alpha\in\mathcal{S}}\sup_{\beta\in\mathcal{C}}(-J_{-f}[\alpha(\beta),\beta]) = \inf_{\alpha\in\mathcal{S}}\sup_{\beta\in\mathcal{C}} J_f[\alpha(\beta),-\beta] = \inf_{\alpha\in\mathcal{S}}\sup_{\beta\in\mathcal{C}} J_f[\alpha(\beta),\beta].$$

□

14.2 The Functional Ehrhard inequality

Corollary 5 (Functional Ehrhard inequality)**.** *Let $\lambda \in [0,1]$, and let f,g,h be uniformly continuous functions with values in $[\epsilon, 1-\epsilon]$ for some $\epsilon > 0$. Suppose that for all $x, y \in \mathbb{R}^n$*

$$\lambda\Phi^{-1}(f(x)) + (1-\lambda)\Phi^{-1}(g(y)) \le \Phi^{-1}(h(\lambda x + (1-\lambda)y)).$$

Then

$$\lambda\Phi^{-1}\Big(\int f d\gamma_n\Big) + (1-\lambda)\Phi^{-1}\Big(\int g d\gamma_n\Big) \le \Phi^{-1}\Big(\int h d\gamma_n\Big).$$

Proof. Fix $\delta > 0$, and choose $\alpha_f, \alpha_g \in \mathcal{S}$ and $\beta_h \in \mathcal{C}$ such that

$$\sup_{\alpha\in\mathcal{S}}\inf_{\beta\in\mathcal{C}} J_{\Phi^{-1}(f)}[\alpha(\beta),\beta] \le \inf_{\beta\in\mathcal{C}} J_{\Phi^{-1}(f)}[\alpha_f(\beta),\beta] + \delta,$$
$$\sup_{\alpha\in\mathcal{S}}\inf_{\beta\in\mathcal{C}} J_{\Phi^{-1}(g)}[\alpha(\beta),\beta] \le \inf_{\beta\in\mathcal{C}} J_{\Phi^{-1}(g)}[\alpha_g(\beta),\beta] + \delta,$$
$$J_{\Phi^{-1}(h)}[\lambda\alpha_f(\beta_h)+(1-\lambda)\alpha_g(\beta_h),\beta_h] \le \inf_{\beta\in\mathcal{C}} J_{\Phi^{-1}(h)}[\lambda\alpha_f(\beta)+(1-\lambda)\alpha_g(\beta),\beta]+\delta.$$

Then Theorem 3 implies that

$$\lambda\Phi^{-1}\Big(\int f d\gamma_n\Big) + (1-\lambda)\Phi^{-1}\Big(\int g d\gamma_n\Big)$$

$$\le \lambda J_{\Phi^{-1}(f)}[\alpha_f(\beta_h),\beta_h] + (1-\lambda)J_{\Phi^{-1}(g)}[\alpha_g(\beta_h),\beta_h] + 2\delta$$
$$\le J_{\Phi^{-1}(h)}[\lambda\alpha_f(\beta_h) + (1-\lambda)\alpha_g(\beta_h),\beta_h] + 2\delta$$
$$\le \Phi^{-1}\Big(\int h d\gamma_n\Big) + 3\delta$$

Letting $\delta \downarrow 0$ gives the result. □

References

[HR] van Handel, R., *The Borell-Ehrhard game.* Probab. Theory Relat. Fields 170, 555585 (2018). https://doi.org/10.1007/s00440-017-0762-4;

[FG] Folland,G.B.*Real Analysis.* Pure and Applied Mathematics (New York), 2nd edn. Wiley, New York (1999)

[LR] Liptser, R.S., Shiryaev, A.N.*Statistics of Random Processes. I,* Applications of Mathematics (New York), vol. 5, Expanded edn. Springer, Berlin (2001)

Ilseok Lee, University of Bonn
email: s6illeee@uni-bonn.de

15 Unification of Ehrhard and Prékopa-Leindler inequalities

After P. Ivanisvili [1]

A summary written by Olli Saari

Abstract

Inequalities of Ehrhard and Prékopa-Leindler are obtained as special cases of a further functional inequality involving an auxiliary function H. This is in the Gaussian case. Validity of the functional inequality is shown to be equivalent to H being a solution to a certain partial differential equation.

15.1 Inequalities

Let $n \geq 1$ be the dimension. Denote the Gaussian measure by

$$d\gamma_n(x) = (2\pi)^{-n/2} e^{-|x|^2/2}\, dx.$$

We set

$$\Phi(x) = \int_{-\infty}^{x} d\gamma_1(x).$$

We consider the following two inequalities.

- *Prékopa-Leindler inequality:* Let h, f, g be positive measurable functions and $\lambda \in (0,1)$. If

  $$h(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda} g(y)^{1-\lambda}$$

  Then

  $$\int_{\mathbb{R}^n} h(x)\, d\gamma_n(x) \geq \left(\int_{\mathbb{R}^n} f(x)\, d\gamma_n(x) \right)^{\lambda} \left(\int_{\mathbb{R}^n} g(x)\, d\gamma_n(x) \right)^{1-\lambda}.$$

- *Ehrhard's inequality:* Let $h, f, g : \mathbb{R}^n \to [0,1]$ be functions such that for all $x, y \in \mathbb{R}^n$

  $$\Phi^{-1}(h(\lambda x + \mu y)) \geq \lambda \Phi^{-1}(f(x)) + \mu \Phi^{-1}(g(y))$$

  where $\lambda, \mu \geq 0$, $\lambda + \mu \geq 1$ and $|\lambda - \mu| \leq 1$. Then

  $$\Phi^{-1}\left(\int_{\mathbb{R}^n} h(x)\, d\gamma_n(x) \right) \geq \lambda \Phi^{-1}\left(\int_{\mathbb{R}^n} f(x)\, d\gamma_n(x) \right) + \mu \Phi^{-1}\left(\int_{\mathbb{R}^n} g(x)\, d\gamma_n(x) \right).$$

Notice that setting $\lambda \in (0,1)$ and $\mu = 1-\lambda$ and replacing $\Phi(x)$ by e^x, one sees that Prékopa-Leindler inequality is of the same form as Ehrhard's inequality. Only the choice of parameters and range of functions must be changed. Further, one sees that a minimal function to always satisfy the assumptions imposed on h must be of the form

$$\sup_{\lambda x+\mu y=t} H(f(x), g(y)),$$

and hence one natural generalization of the inequalities reads

$$\int_{\mathbb{R}^n} \sup_{y\in\mathbb{R}^n} H(f((x-y)/a), g(y/b))\, d\gamma_n(t) \geq H\left(\int_{\mathbb{R}^n} f(x)\, d\gamma_n(x), \int_{\mathbb{R}^n} f(y)\, d\gamma_n(y)\right) \quad (1)$$

for $a, b > 0$ and a suitable real valued function H. Certain choices of H yielding the inequalites of Ehrhard and Prékopa-Leindler as special cases. The main result is the following.

Theorem 1. *Let $I, J \subset \mathbb{R}$ be closed intervals and let $H \in C^3(I\times J;\mathbb{R})$ with $H_x(x,y)H_y(x,y) \neq 0$ for all $(x,y) \in I\times J$. Let $a, b > 0$ be such that $|1-a^2-b^2| \leq 2ab$ and $n \geq 1$. Then (1) holds if and only if*

$$a^2 H_{xx}H_y^2 + (1-a^2-b^2)H_xH_yH_{xy} + b^2 H_{yy}H_x^2 \geq 0. \quad (2)$$

The necessity holds for probability measures much more general than γ_n. Indeed, a probability measure $\mu \ll \mathcal{L}^1$ with finite moments can satisfy (1) only if (2) holds.

15.2 The proof

Necessity

The necessity of the differential inequality follows by testing the inequality (1) with suitable f and g. Set

$$\varphi_{\epsilon,\delta}(t) = \begin{cases} -\delta\epsilon^{-\alpha}, & t \leq -\delta\epsilon^{-\alpha} \\ t, & -\delta\epsilon^{-\alpha} \leq t \leq \epsilon^{-\alpha} \\ \epsilon^{-\alpha}, & t \geq \epsilon^{-\alpha} \end{cases}$$

and

$$f(x) = u + \epsilon\frac{\varphi_{\epsilon,\delta}(ax)}{H_u(u,v)} + \epsilon^2\frac{p\varphi_{\epsilon,\delta}(ax)}{2H_u(u,v)}$$
$$g(y) = u + \epsilon\frac{\varphi_{\epsilon,\delta}(by)}{H_v(u,v)} + \epsilon^2\frac{q\varphi_{\epsilon,\delta}(ay)}{2H_u(u,v)}$$

with all parameters ϵ, δ, α, p and q to be specified later. Feeding in the test functions f and g, one derives the necessary condition from (1) after a rather lengthy computation.

Sufficiency, first step

Consider positive integers k, k_1, k_2, k_3. Let each A_j with $j = 1, 2, 3$ be a $k \times k_j$ matrix of full rank. Let C be a positive definite $k \times k$ matrix. Fix a real valued twice differentiable function B defined on a closed rectangle $I_1 \times I_2 \times I_3 \subset \mathbb{R}^3$ and define the block matrix

$$D(x) = \{(A_i^T C A_j)\partial_{ij} B(x)\}_{i,j=1}^3$$

for all $x \in \mathbb{R}^3$. Denote

$$d\gamma_C(x) = \frac{1}{\sqrt{(2\pi)^k \det C}} e^{-|C^{-1/2}x|^2/2}\, dx.$$

Theorem 1 follows from Theorem 2 by essentially setting $k = 2n$, $k_i = n$ and choosing the matrices and the function B carefully.

Theorem 2. *The matrix $D(x)$ is positive definite for all $x \in I_1 \times I_2 \times I_3$ if and only if*

$$\int_{\mathbb{R}^k} B(u_1(xA_1), u_2(xA_2), u_3(xA_3))\, d\gamma_C(x) \\ \geq B\left[\left(\int_{\mathbb{R}^{k_i}} u_i(y\sqrt{A_i^T C A_i})\, d\gamma_{k_i}\right)_{i=1}^3\right] \quad (3)$$

holds for all Borel measurable $u_i : \mathbb{R}^{k_i} \to I_i$, $i = 1, 2, 3$.

Sufficiency, second step

It remains to prove Theorem 2. Writing $\tilde{A}_i = C^{1/2} A_i$, one may assume without loss of generality that C is the identity matrix. Denote $\tilde{u}_i(x) = u_i(xA_i)$ and $\vec{u} = (\tilde{u}_i)_{i=1}^3$. Let $P_t = e^{-t\Delta}$ be the heat extension and set $P_t\vec{u} = (P_t\tilde{u}_i)_{i=1}^3$ where the heat extension is understood to detect the dimension of the input function.

Note that $P_{1/2}$ evaluated at $x = 0$ is then the expectation with respect to $d\gamma$. By changing variables, we see that (3) is equivalent to

$$V(x, t) := B(P_t\vec{u}(x)) - P_t B(\vec{u}(x)) \leq 0 \quad (4)$$

with this choice. By a computation, we verify

$$(\partial_t - \Delta)V(x,t) = -[P_t \nabla \vec{u}(x)]D(P_t \vec{u}(x))[P_t \nabla \vec{u}(x)]^T.$$

The assumption on $D(x)$ being positive definite translates to $V(x,t)$ being a subsolution of the heat equation

If $D(x)$ is positive definite and $V(x,t)$ is accordingly subsolution, the parabolic maximum principle implies $V(0,1/2) \leq V(0,0) = 0$ as claimed. Conversely, if (3) holds for all test functions, then $V(x,t) \leq 0$ for all (x,t). It follows

$$0 \geq \lim_{t \to 0+} \frac{V(x,t) - V(x,0)}{t} = -[P_0 \nabla \vec{u}(x)]D(P_0 \vec{u}(x))[P_0 \nabla \vec{u}(x)]^T$$

and as u is arbitrary, the claim follows.

15.3 Extensions

Similar method is applied to a number of other inequalities in [2]. Several versions of Theorem 2 are proved there. The number of input functions need not be restricted to two. Choosing the function B and and the matrices differently, one can prove the inequalities here, both Gaussian and Lebesgueian, as well as many related inequalities. The method can be seen as a systematic reformulation in terms of a "Bellman PDEs" of other heat flow methods.

References

[1] Ivanisvili, P., *A boundary value problem and the Ehrhard inequality.* Studia. Math. **246** (2019), no. 3, 257–293;

[2] Ivanisvili, P., Volberg, A., *Bellman partial differential equation and the hill property for classical isoperimetric problems.* arXiv:1506.03409v2.

Olli Saari, University of Bonn
email: saari@math.uni-bonn.de

16 Capacity of Rank Decreasing Operators

After Gurvits [5]

A summary written by Jacob Denson

Abstract

We describe the theory of the capacity of rank-decreasing positive operators, and Gurvit's operator rescaling algorithm to compute the capacity of such operators, with a brief discussion of the connection between this theory and the computation of Brascamp-Lieb constants.

Recall a theorem of Lieb [7], which shows that any inequality of the form

$$\int_{\mathbf{R}^n} \prod_{i=1}^m |f_i(B_i x)|^{p_i} \, dx \leq \mathrm{BL}(B,p) \cdot \prod_{i=1}^m \|f_i\|_{L^1(\mathbf{R}^{n_i})}^{p_i},$$

has an optimal Brascamp-Lieb constant $\mathrm{BL}(B,p)$ satisfying

$$\mathrm{BL}(B,p)^2 = \sup_{X_1,\dots,X_m \succ 0} \frac{\det(\sum_{i=1}^m p_i B_i^* X_i B_i)}{\prod_{i=1}^m \det(X_i)^{p_i}}. \tag{1}$$

Here and in what follows, for a square matrix A, we write $A \succ 0$ and $A \succeq 0$ to mean A is positive definite or semidefinite.

The focus of these notes is to introduce analysts to a setting, originally studied by computing scientists to understand problems in combinatorial optimization and quantum information theory, which can provide insight into the computation of the Brascamp-Lieb constant. Utilizing this theory, Garg, Gurvits, and Wigderson [4] formulated a polynomial time algorithm for approximating the Brascamp-Lieb constant, a fact that has important theoretical consequences to the general theory of Brascamp-Lieb inequalities independent of practical application to computation of particular Brascamp-Lieb constants. This connection is more fully explored in the following talk. In this summary, we introduce the original setting studied, only noting some similarities to Brascamp Lieb as we proceed.

16.1 Operator Scaling

Let $M(N)$ denote the space of all $N \times N$ complex-valued matrices. The main object of study in these notes are *positive operators* $T : M(N) \to M(N)$, i.e. linear transformations between spaces of matrices such that if $X \succeq 0$, then $T(X) \succeq 0$. There is a rich theory of these maps, connected to representation theory and free probability theory. For more background, the textbook [2] provides a thorough introduction to this theory. A useful example to keep in mind, given any matrices $B_1, \ldots, B_m$ and $p_1, \ldots, p_m > 0$, is the operator

$$T(X) = \sum_{i=1}^{m} p_i (B_i^* X B_i). \tag{2}$$

With any positive operator T, we associate a quantity $\mathrm{Cap}(T) \geq 0$, known as the *capacity* of T, defined by setting

$$\mathrm{Cap}(T) = \inf_{X \succ 0} \frac{\det(TX)}{\det(X)}. \tag{3}$$

The connection between capacity and Brascamp-Lieb may be hinted at by comparing (3) to (1) when T is of the form given in (2). Being defined by a non-convex optimization, it seems difficult to explicitly compute capacity. In this talk, we discuss a result of Gurvits [5], who found an efficient algorithm to compute, for each $\varepsilon > 0$, a value $\mathrm{Cap}_{\mathrm{approx}}(T)$ such that

$$\mathrm{Cap}(T) \leq \mathrm{Cap}_{\mathrm{approx}}(T) \leq (1 + \varepsilon) \cdot \mathrm{Cap}(T).$$

We obtain this approximation by applying the technique of *operator scaling*, which we describe in this section.

Given any invertible matrices $A, B \in M(N)$, and any positive operator $T : M(N) \to M(N)$, we define a *scaled operator* $T_{A,B}(X) = BT(AXA^*)B^*$. For any $X \succ 0$, if we write $Y = AXA^*$, then

$$\frac{\det(T_{A,B}X)}{\det(X)} = \det(A)^2 \det(B)^2 \cdot \frac{\det(TY)}{\det(Y)}. \tag{4}$$

Taking infima over both sides of (4) for all choices of input X shows that $\mathrm{Cap}(T_{A,B}) = \det(A)^2 \det(B^2)\mathrm{Cap}(T)$. Thus computing $\mathrm{Cap}(T_{A,B})$ immediately gives $\mathrm{Cap}(T)$. The idea of operator is to rescale an operator into something whose capacity we can more easily compute. This motivates the introduction of *doubly stochastic operators*.

A positive operator T is *doubly stochastic* if $T(I) = T^*(I) = I$, where T^* is the adjoint of T with respect to the inner product $(X, Y) \mapsto \mathrm{Tr}(XY^*)$. What interests us about this definition is that $\mathrm{Cap}(T) = 1$ for *any* doubly stochastic operator T. The proof of this statement can be found as Theorem 4.32 of [8], with relevant background information about doubly stochastic matrices found in Section 8.5 of [9].

Thus, given a positive operator T, if we can find $A, B \in M(N)$ such that $T_{A,B}$ is doubly stochastic, then it follows that $\mathrm{Cap}(T) = \det(A)^{-2} \det(B)^{-2}$.

This is clearly only possible if $\text{Cap}(T) > 0$, but not *quite* possible for all such T. In fact, in [3], it is shown that this rescaling is only possible if the equation defining capacity has extremizers, i.e. $\text{Cap}(T) = \det(TX)/\det(X)$ for some particular input X. Nonetheless, if $\text{Cap}(T) > 0$ then we will find doubly stochastic operators approximating T arbitrarily closely. Since capacity is a continuous function of the input, this suffices to approximate $\text{Cap}(T)$ up to an arbitrary error.

Gurvits' approximation algorithm is quite simple, an application of a similar method first utilized by Sinkhorn [10]. Given a positive operator T, it is easy to scale the operator to an operator S with $S(I) = I$; we simply consider the operator $T_{I,T(I)^{-1/2}}$. Similarily, we can scale T to an operator S with $S^*(I) = I$ by taking the scaling $T_{T*(I)^{-1/2},I}$. The challenge is to obtain scalings for which both properties are approximately true. Sinkhorn's trick is to iteratively apply each of the rescalings to our operators, obtaining a sequence of matrices $T_0, T_1, T_2, \dots$ with $T_0 = T$, $T_i(I) = I$ for odd i, and $T_i^*(I) = I$ for even i. If this sequence converges, the limit will be doubly stochastic, as desired.

Convergence to a single stochastic matrix does not occur for all T with $\text{Cap}(T) > 0$, but we will show that if $\text{Cap}(T) > 0$, then the distance between the elements of the sequence to the family of *all* stochastic operators converges to zero, which is sufficient for our purposes. To analyze the convergence, we rely on the capacity as a *potential* for the analysis of the algorithm (this was the main reason Gurvits first defined the capacity in [5]). To determine how close we are to a doubly stochastic matrix, we use the measure

$$\text{DS}(T) = \|T(I) - I\|^2 + \|T^*(I) - I\|^2.$$

The following properties then hold:

1. If $T(I) = I$ or $T^*(I) = I$, then $A_N \lesssim_N \text{Cap}(T) \leq 1$ for some constant $A_N > 0$ depending on N. Moreover, for $\text{Cap}(T) \geq 1/2$,
$$\text{DS}(T_n) \leq 6 \log(1/\text{Cap}(T_n)).$$
2. $\text{Cap}(T_n)$ is increasing in $n > 0$. More precisely,
$$\text{Cap}(T_{n+1}) \geq e^{\min(1,\text{DS}(T_n))/6} \cdot \text{Cap}(T_n).$$
3. If $T(I) = I$ or $T^*(I) = I$ and $\text{DS}(T) < 1/(N+1)$, then $\text{Cap}(T) > 0$.

We claim that from these properties, we can detect in $M_0 = 6(N+1)\log(1/A_N)$ iterations of the algorithm whether $\text{Cap}(T) > 0$. Indeed, if $\text{Cap}(T) > 0$, then

Property 1 implies that $\mathrm{Cap}(T_1) \geq A_N$. For any δ and M, repeated applications of Property 3 imply that there either exists $i < M$ such that $\mathrm{DS}(T_i) \leq \delta$, or $\mathrm{Cap}(T_M) \geq e^{\varepsilon M/6} A_N$. In particular, if $\delta = 1/(N+1)$, and $M = M_0$, then this calculation, combined with Property 1, implies that if $\mathrm{Cap}(T) > 0$, then there must exist $i \leq M_0$ with $\mathrm{DS}(T_i) < 1/(N+1)$. On the other hand, Property 3 implies that if $\mathrm{Cap}(T) = 0$, then $\mathrm{DS}(T_n) \geq 1/(N+1)$ for all n. Checking $\mathrm{DS}(T_i)$ for $i \leq M_0$ thus gives a simple way to check whether $\mathrm{Cap}(T) > 0$. Using very similar techniques, we leave it as an exercise to check these properties imply that if $\mathrm{Cap}(T) > 0$, then in $(6/\varepsilon)\log(1/A_N)$ iterations of the algorithm, one can find i such that $\mathrm{DS}(T_i) \leq \varepsilon$. Thus we have a reliable way to approximate the capacity of a matrix, which [4] extends to find efficient ways to approximate Brascamp-Lieb constants. More advanced techniques, given in [3], show that we only actually need to run the algorithm for $\mathrm{Poly}(N, \log(1/\varepsilon))$ iterations to obtain a ε-approximation.

16.2 Capacity of Rank-Decreasing Operators

We conclude these notes by discussing an analogy between Brascamp-Lieb and the study of capacity. Bennett, Carbery, Christ, and Tao [BCCT08] showed that $\mathrm{BL}(B,p)$ is finite in (1) if and only if $\sum_{j=1}^m p_j n_j = 0$, and

$$\dim(V) \leq \sum_{j=1}^m p_j \dim(B_j V) \qquad \text{for all subspaces } V \subset \mathbf{R}^n. \tag{5}$$

Thus the finiteness of (1) acts as a guarantee for the mapping properties of the matrices $B_1, \dots, B_m$ given in (5), and vice versa. A useful property of a positive operator T is that it is *rank non-decreasing*, i.e. for any $X \succeq 0$,

$$\mathrm{Rank}(TX) \geq \mathrm{Rank}(X). \tag{6}$$

Condition (6) seems somewhat similar to (5). And indeed, analogous to the equivalence between (5) and the finiteness of (1), one has an equivalence between (6) and the non-vanishing of (3).

Theorem 1. *T is rank non-decreasing if and only if $\mathrm{Cap}(T) > 0$.*

Proof. A simple family of positive operators are those of the form

$$TX = X_{11}A_1 + \dots + X_{NN}A_N, \tag{7}$$

where $A_1, \dots, A_N \succeq 0$. For such an operator, we can write

$$\mathrm{Cap}(T) = \inf_{\gamma_1,\dots,\gamma_N > 0} \frac{\det(\sum_{j=1}^N \gamma_j A_j)}{\gamma_1 \cdots \gamma_N}.$$

Results from a previous paper of Gurvits and Samorodnitsky [6] imply Theorem 1 in the special case of an operator defined by (7). Assuming this result, we indicate how this implies the general case.

For each orthonormal basis $U = \{u_1, \dots, u_N\}$, we define the decoherence operator $D_U(X) = \sum \langle Xu_i, u_i \rangle \cdot u_i u_i^*$, and then consider the operator T_U defined such that $T_U(X) = (T \circ D_U)(X) = \sum_{i=1}^{N} \langle Xu_i, u_i \rangle \cdot T(u_i u_i^*)$. This operator is, up to a change of basis in $M(N)$, described in the form (7). Thus T_U is rank non-decreasing if and only if $\mathrm{Cap}(T_U) > 0$. The theorem then follows from the following two properties of this construction:

1. T is rank non-decreasing if and only if T_U is as well, for all bases U.

2. $\mathrm{Cap}(T) = \inf_U \mathrm{Cap}(T_U)$.

If $\mathrm{Cap}(T) > 0$, then Property 2 implies $\mathrm{Cap}(T_U) > 0$ for all U, so T_U is rank non-decreasing for all U, and thus Property 1 implies T is rank non-decreasing. The converse is similar.

The proof of Properties 1 and 2 both rely on a simple trick. We will prove Property 1 here. Given any $X \succeq 0$, we can find an orthonormal basis U diagonalizing X, and then for such U we have $T(X) = T_U(X)$. This immediately implies T is rank non-decreasing if T_U is rank non-decreasing for all U. The converse follows because the composition of rank non-decreasing operators is rank non-decreasing, and D_U is rank non-decreasing (all doubly stochastic operators are rank non-decreasing). $\square$

References

[1] J.M. Bennett, A. Carbery, M. Christ and T. Tao, *The Brascamp-Lieb Inequalities: Finiteness, Structure and Extremals*, Geom. Funct. Anal. **17** (2008), 1343-1415.

[2] Rajendra Bhatia, *Positive Definite Matrices*, Princeton University Press, 2007.

[3] Ankit Garg, Leonid Gurvits, Rafael Oliveira, Avi Wigderson, *Operator Scaling: Theory and Applications*, Foundations of Computational Mathematics, **20** (2020), 223–290.

[4] Ankit Garg, Leonid Gurvits, Avi Wigderson, *Algorithmic and Optimization Aspects of Brascamp-Lieb Inequalities, via Operator Scaling*, Geom. Funct. Anal, **28** (2018), 100–145.

[5] Leonid Gurvits, *Classical Complexity and Quantum Entanglement*, J. Comput. Syst. Sci. **69** (2004), 448–484.

[6] L. Gurvits, A. Samorodnitsky, *a Deterministic Algorithm Approximating the Mixed Discriminant and Mixed Volume, and a Combinatorial Corollary*, Discrete Comput. Geom. **27** (2002), 531–550.

[7] E.H. Lieb, *Gaussian Kernels have only Gaussian Maximizers* Invent. Math. **102** (1990), 179–208.

[8] John Watrous, *the Theory of Quantum Information* Cambridge University Press, 2018.

[9] Denis Serre, *Matrices*, Springer, 2010.

[10] Richard Sinkhorn, *A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices*, Ann. Math. Statist. **35** (1964), 876–879.

Jacob Denson, UW Madison
email: `jcdenson@wisc.edu`

17 Algorithmic and optimization aspects of the Brascamp-Lieb constant

After A. Garg, L. Gurvits, R. Oliveira and A. Wigderson [GGOW18]

A summary written by Felipe Gonçalves

Abstract

The authors of [GGOW18] produce an efficient algorithm, embedded in the computational framework of operator scaling, that decides in polynomial time when the Brascamp-Lieb constant is finite, delivering an arbitrarily close approximation of its value or else exhibiting a counter-example to the subspace inequality criteria.

17.1 Introduction

The Brascamp-Lieb inequality [BL76, L90, BCCT08] gives a condition for the existence of a constant $C > 0$ such that

$$\int_{\mathbb{R}^n} \prod_{j=1}^{m} f_j(B_j x)^{p_j} dx \leq C \prod_{j=1}^{m} \left(\int_{\mathbb{R}^{n_j}} f_j(x) dx \right)^{p_j} \quad \text{for all} \quad f_j : \mathbb{R}^{n_j} \to \mathbb{R}_+,$$

where $B_j : \mathbb{R}^n \to \mathbb{R}^{n_j}$ are given linear surjective maps, $1 \leq n_j \leq n$ and $p_j \in \mathbb{R}_+$. We refer to the collection $(\mathbf{B}, \mathbf{p}) = ((B_1, ..., B_m), (p_1, ..., p_m))$ as the BL-*datum* and C will depend on $(\mathbf{B}, \mathbf{p})$. Given a BL-datum we then define the optimal constant by

$$\mathrm{BL}(\mathbf{B}, \mathbf{p}) := \sup \frac{\int_{\mathbb{R}^n} \prod_{j=1}^{m} f_j(B_j x)^{p_j} dx}{\prod_{j=1}^{m} \left(\int_{\mathbb{R}^{n_j}} f_j(x) dx \right)^{p_j}}.$$

A criteria for the finiteness of the Brascamp-Lieb constant is given in [BCCT08]. Namely, $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$ iff the following hold:

- $n = \sum_j p_j n_j$;
- $\dim V \leq \sum_j p_j \dim B_j V$ for any subspace $V \subset \mathbb{R}^n$.

For now on we assume all $(\mathbf{B}, \mathbf{p})$ satisfy $n = \sum_j p_j n_j$. We say that $(\mathbf{B}, \mathbf{p})$ is *isotropy-normalized* if $\sum_j p_j B_j^T B_j = I_n$. Otherwise, we can isotropy-normalize it by the transformation

$$B_j \leftarrow B_j \left(\sum_{j=1}^{m} p_j B_j^T B_j \right)^{-1/2}.$$

We say that $(\mathbf{B}, \mathbf{p})$ is *projection-normalized* if each B_j is a projection, namely $B_j B_j^T = I_{n_j}$. We can projection-normalize it by the transformation

$$B_j \leftarrow (B_j B_j^T)^{-1/2} B_j.$$

Observe however that one normalization destroys the other. We say $(\mathbf{B}, \mathbf{p})$ is *geometric* if it is isotropy-normalized and projection-normalized. Projection-normalization is always possible since B_j is surjective, and so $B_j B_j^T$ is positive definite. Lieb [L90] showed that if $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$ then

$$\mathrm{BL}(\mathbf{B}, \mathbf{p})^2 = \sup_{X_j \succ 0} \frac{\prod_{j=1}^m (\det X_j)^{p_j}}{\det\left(\sum_{j=1}^m p_j B_j^T X_j B_j\right)},$$

where the supremum is taken over all positive definite matrices $X_j \in GL_{n_j}(\mathbb{R})$. In particular, if $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$ then $\sum_j p_j B_j^T B_j$ is invertible (hence positive definite) because we can take $X_j = I_{n_j}$. Hence we can always apply isotropy-normalization.

Lemma 1 ([GGOW18])**.** *Let* $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$ *and* $(\mathbf{B}, \mathbf{p})$ *be either isotropy-normalized or projection-normalized. Then* $\mathrm{BL}(\mathbf{B}, \mathbf{p}) \geq 1$ *and equality holds if and only if* $(\mathbf{B}, \mathbf{p})$ *is geometric.*

17.2 Main Results

In [GGOW18] the authors study the computational aspects of the $\mathrm{BL}(\mathbf{B}, \mathbf{p})$ constant. More specifically they ask, given a rational datum $(\mathbf{B}, \mathbf{p})$, is there an *efficient algorithm* that answers the following questions:

(a) Decide if $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$?

(b) If $\mathrm{BL}(\mathbf{B}, \mathbf{p}) = \infty$, find $V \subset \mathbb{R}^n$ such that $\dim V > \sum_j p_j \dim B_j V$.

(c) If $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$, can we numerically approximate the value $\mathrm{BL}(\mathbf{B}, \mathbf{p})$ with arbitrary precision?

Note that if $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$ and $B'_j = C_j^{-1/2} B_j C^{-1/2}$ then

$$\mathrm{BL}(\mathbf{B}', \mathbf{p})^2 = \det C \prod_{j=1}^m (\det C_j)^{p_j} \mathrm{BL}(\mathbf{B}, p)^2.$$

They propose the following *alternate minimization algorithm*.

Algorithm: Input a BL-datum $(\mathbf{B}, \mathbf{p})$ and $\epsilon > 0$.
Step 1: We first isotropy-normalize $(\mathbf{B}, \mathbf{p})$.
Step 2: If $(\mathbf{B}, \mathbf{p})$ was just isotropy-normalized, then projection-normalize it. Otherwise, then isotropy-normalize it.
Step 3: If $\mathrm{BL}(\mathbf{B}, \mathbf{p}) \leq 1 + \epsilon$ then halt, otherwise go back to Step 2.

Such greedy iterative scaling is common in the computational context of operator scaling and have been used widely. The general idea is the following: If we are given a metric space (M, d), two sets $X, Y \subset M$ and we want to find $(x^*, y^*) = \arg\min_{x \in X, y \in Y} d(x, y)$, then one can iteratively use the alternate minimization $y_1 = \arg\min_{y \in Y} d(x_0, y)$, $x_1 = \arg\min_{x \in X} d(x, y_1)$, $y_2 = \arg\min_{y \in Y} d(x_1, y)$, $x_2 = \arg\min_{x \in X} d(x, y_2)$ and so on. The hope is that $(x_n, y_n) \to (x^*, y^*)$ as $n \to \infty$. Such algorithm has been employed for instance in the search for doubly stochastic scaling for completely positive operators [G04, GGOW16, IGS18] (more details see [GGOW18, Section 3]).

We say that a BL-datum $(\mathbf{B}, \mathbf{p})$ has total binary length $b \in \mathbb{N}$ if all B_j's have rational coordinates and we have $b = \sum_{j,k,l} \mathrm{bin.length}(\langle B_j e_k, e_l \rangle)$. We say that $(\mathbf{B}, \mathbf{p})$ has common denominator d if $p_j \in \frac{1}{d}\mathbb{Z}_+$ for all j.

Theorem 2. *Assume that the datum $(\mathbf{B}, \mathbf{p})$ has binary length b and common denominator d. Assume that $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$. Then the proposed algorithm above computes a $(1+\epsilon)$-approximation of $\mathrm{BL}(\mathbf{B}, \mathbf{p})$ in time $\mathrm{poly}(d, b, 1/\epsilon)$. Moreover, the algorithm outputs $\mathbf{B}'$ which is almost geometric in the sense $\mathbf{B}'$ is either isotropy-normalized or projection-normalized, and $\mathrm{BL}(\mathbf{B}, \mathbf{p}) \leq 1+\epsilon$.*

Building up on previous work [G04, GGOW16] and using the results of [IGS18] as a black box, the authors were able construct a bridge connecting the world of operator scaling and Brascamp-Lieb inequalities, where the following result then follows immediately from [IGS18].

Theorem 3. *There is an algorithm that on input $(\mathbf{B}, \mathbf{p})$ of binary length b and common denominator d runs in time $\mathrm{poly}(b, d)$ and tests if $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$. Moreover, if $\mathrm{BL}(\mathbf{B}, \mathbf{p}) = \infty$ it then provides a subspace $V \subset \mathbb{R}^n$ such that $\dim V > \sum_j p_j \dim B_j V$.*

References

[BCCT08] J. Bennett, A. Carbery, M. Christ and T. Tao. The Brascam-Lieb inequalities: finiteness, structure, and extremals. Geom. and Funct. Anal., 17(5) (2008), 1343-1415.

[BL76] H. Brascamp and E. Lieb. Best constants in Youngs inequality, its converse and its generalization to more than three functions. Advances in Mathematics 20 (1976), 151-172.

[GGOW16] A. Garg, L. Gurvits, R. Oliveira and A. Wigderson, A Deterministic Polynomial Time Algorithm for Non-commutative Rational Identity Testing, 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS), 109-117.

[GGOW18] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson. Algorithmic and optimization aspects of Brascamp-Lieb inequalities, via operator scaling. Geom. Funct. Anal. 28(1) (2018), 100-145.

[G04] L. Gurvits. Classical complexity and quantum entanglement. In: J. Comput. System Sci. 69(3) (2004), 448-484.

[IGS18] G. Ivanyos, Y. Qiao and K. V. Subrahmanyam. Constructive non-commutative rank computation is in deterministic polynomial time. Comput. Complex. 27 (2018), 561-593.

[L90] E. Lieb. Gaussian kernels have only Gaussian maximizers. Inventiones Mathematicae 102 (1990), 179-208.

F. Gonçalves, University of Bonn
email: `goncalve@math.uni-bonn.de`

18 The Kempf–Ness theorem

After Böhm and Lafuente [BL]

A summary written by Gianmarco Brocchi

Abstract

The Brascamp–Lieb constant is related to the length of minimal vectors in the sense of the Kempf–Ness theorem. We present the real version of the theorem and main ideas of the proof.

18.1 Introduction

Given a collection of surjective maps $\pi_j \colon \mathbb{R}^d \to \mathbb{R}^{d_j}$ and numbers $s_j > 0$ for $j \in \{1, \ldots, m\}$, with m, d and $d_j \in \mathbb{N}$, with $d_j < d$, we consider the Brascamp–Lieb inequality:

$$\int_{\mathbb{R}^d} \prod_j |f_j(\pi_j x)|^{s_j} dx \leq \mathsf{BL}(\{\pi_j, s_j\}) \prod_{j=1}^{m} \left(\int_{\mathbb{R}^{d_j}} f_j(y) dy \right)^{s_j} . \tag{1}$$

The inequality (1) is maximised by Gaussian. Let g_j be a Gaussian on $\mathbb{R}^{d_j}$. By plugging g_j into (1) we obtain

$$\mathsf{BL}(\{\pi_j, s_j\}) \geq \frac{\int_{\mathbb{R}^d} \prod_j |g_j(\pi_j x)|^{s_j} dx}{\prod_{j=1}^{m} \left(\int_{\mathbb{R}^{d_j}} g_j(y) dy \right)^{s_j}} = \left(\frac{\det(\sum_j s_j \pi_j^* A_j^* A_j \pi_j)}{\prod_{j=1}^{m} \det(A_j^* A_j)^{s_j}} \right)^{-1/2} .$$

So the optimal constant $\mathsf{BL}(\{\pi_j, s_j\})$ is achieved by taking the supremum over all matrices $A_j \in \mathsf{GL}(d_j)$. In [Gr], the right hand side of the expression above is written by using the Hilbert–Schmidt norm, so that

$$\mathsf{BL}(\{\pi_j, s_j\})^{-1} = \inf_{\substack{A_j \in \mathsf{SL}(d_j) \\ A \in \mathsf{SL}(d)}} \prod_{j=1}^{m} \left(d_j^{-1/2} \| A_j \pi_j A^* \|_{\mathrm{HS}} \right)^{s_j d_j} . \tag{2}$$

Equation (2) gives a way to approximate Brascamp–Lieb constant by minimising a "distance function" under the action of the group $G \subset \mathsf{GL}(d_1) \otimes \cdots \otimes \mathsf{GL}(d_m) \otimes \mathsf{GL}(d)$ on the vector space $(V, \langle \cdot, \cdot \rangle)$ where the projections π_j live. The quantity $\mathsf{BL}(\{\pi_j, s_j\})^{-1}$ is the length of the minimal vector in a given orbit.

A classical theorem by George Kempf and Linda Ness relates closed orbits and minimal vectors.

18.2 Kempf–Ness theorem

Definition 1. *Let G be a group acting on a vector space V endowed with inner product $\langle \cdot, \cdot \rangle$, and let $d \colon V \to \mathbb{R}_+$ be a given function. For $v \in V$, a* minimal *vector $\bar{v}$ in the orbit $G \cdot v$ is a vector that minimises $d(\cdot)$.*

Let $\mathscr{M}$ be the set of minimal vectors in V.

Remark 2 (Closed orbits intersect $\mathscr{M}$)**.** *If the orbit $G \cdot v$ is closed (as a set), the intersection with closed balls $B_R(0) \coloneqq \{w : d(w) \leq R\}$ for R large enough is not empty and is compact. In particular, $d(\cdot)$ has a minimum on $G \cdot v \cap B_R(0)$, so $G \cdot v$ contains a minimal vector.*

The converse, for real reductive Lie groups, is the (real version of the) Kempf–Ness theorem. We briefly introduce reductive Lie groups.

Let G be a Lie group and let $\mathfrak{g}$ be the Lie algebra of G. We will consider the symmetric part of the algebra given by the Cartan decomposition.

Remark 3 (Cartan decomposition)**.** *Let $G \subset \mathsf{GL}(d)$ be a Lie group and let $\mathfrak{g}$ be its Lie algebra. Then $\mathfrak{g}$ can be decomposed in symmetric and anti-symmetric part: $\mathfrak{g} = \mathfrak{s} \oplus \mathfrak{a}$, where $\mathfrak{s} = \mathfrak{g} \cap \mathsf{Sym}(V)$. If $[\cdot, \cdot]$ is the Poisson bracket: $[a, b] = ab - ba$, then $[\mathfrak{s}, \mathfrak{s}] \subset \mathfrak{a}$ and $[\mathfrak{s}, \mathfrak{a}] \subset \mathfrak{s}$.*

Definition 4. *A Lie group G is called* reductive *if can be written as $G = K \cdot \exp(\mathfrak{s})$, where K is a maximal compact subgroup of G.*

We will be interested in subgroups of $\mathsf{GL}(d, \mathbb{R})$.

Theorem 5 (Real Kempf–Ness Theorem)**.** *Let $G \subset \mathsf{GL}(d, \mathbb{R})$ be a reductive Lie group with a maximal subgroup $K = G \cap O(\mathbb{R}^d)$. For $v \in V$ the orbit $G \cdot v$ contains a minimal vector if and only if is closed. Moreover $G \cdot v \cap \mathscr{M} = K \cdot v$.*

18.3 Proof of the theorem

The proof is based on two main facts:

1. If the orbit $G \cdot v$ is not closed, the elements in the closure can be reached with a one-parameter subgroup. This is proved by contradiction: assuming that all such orbits are separated leads to an absurd.

2. If the distance function $d(\cdot)$ is strictly convex, its critical points are minima and there are not such points on non-closed orbits.

We start by considering the simpler case of abelian groups.

18.4 Abelian case

Let T be a abelian, non compact, connected Lie group[2] and let $\mathfrak{t}$ be its Lie algebra. Let $K \subset T$ be a maximal, compact subgroup. In the complex case, one can think of K as the elements of T with modulus 1.

Representation

The elements in T can be written as $\exp(t\alpha) \in T$ for $\alpha \in \mathfrak{t}$ and $t \in \mathbb{R}$. In particular, there is $(v_1, \ldots, v_N)$ basis of V which diagonalises the action of T. Then for $\lambda \in \mathfrak{t}$ we have

$$e^{\lambda} \cdot v = (e^{\langle \lambda, \alpha_1 \rangle} v_1, \ldots, e^{\langle \lambda, \alpha_N \rangle} v_N), \quad \text{for} \quad \alpha_1, \ldots, \alpha_N \in \mathfrak{t}.$$

For an abelian group T, the representation of its action is enough to show that, given any $\bar{v} \in \overline{T \cdot v} \setminus T \cdot v$, there is a one-parameter semigroup intersecting the orbit $T \cdot \bar{v}$.

Lemma 6 (Hilbert–Mumford for abelian groups)**.** *For any $\bar{v} \in \overline{T \cdot v} \setminus T \cdot v$ there exists $g \in T$ and $\alpha \in \mathfrak{t}$ such that $\lim_{s \to \infty} \exp(s\alpha) \cdot v = g \cdot \bar{v}$.*

Convexity of the distance function

For $\alpha \in \mathfrak{s}$ and $t \in \mathbb{R}$, consider the distance function

$$d(t) \coloneqq d_{\alpha,v}(t) \coloneqq \|\exp(t\alpha) \cdot v\|^2.$$

This is the square of the distance to the origin along the curve $\exp(t\alpha) \cdot v$ in G. The function $d(t)$ is convex, so its critical points are minima.

Lemma 7 (Convexity)**.** *For $A \in \mathfrak{s}$ and $v \in V$ the function $d_{\alpha,v}(t)$ is convex, in particular $d''(t) = 4\|A \cdot \exp(tA) \cdot v\|^2$.*

Let $\alpha \in \mathfrak{s}$ and assume that $\lim_{t \to \infty} \exp(t\alpha) \cdot v = \bar{v}$ exists. Then, by convexity of $d(t)$, we have that

$$\|e^{t\alpha} \cdot v\| > \|\bar{v}\|, \quad \forall t \in \mathbb{R}.$$

Thus the function $d(t)$ cannot achieve its minimum on a non-closed orbit.

[2] the notation comes from T being an algebraic torus in the complex case.

18.5 Real reductive groups

For general real reductive groups, one can write $G = KTK$, where K is compact and T is abelian. It is enough to show that the limit of the one-parameter subgroup exists.

Lemma 8 (Hilbert–Mumford for real reductive groups)**.** *Let G be a real reductive group and let $v \in V$. If the orbit $G \cdot v$ is not closed then there exists $\alpha \in \mathfrak{s}$ such that $\lim_{s\to\infty} \exp(s\alpha) \cdot v$ exists.*

Idea of the proof. Let $\mathfrak{t} \subset \mathfrak{s}$ be the maximal abelian subalgebra. Since $G = KTK$, with $T = \exp(\mathfrak{t})$, it is enough to show that given $\bar{v} \in \overline{G \cdot v} \setminus G \cdot v$ there exists $g \in G$, $k \in K$ and $\alpha \in \mathfrak{t}$ such that $\lim_{s\to\infty} \exp(s\alpha) \cdot (k \cdot v) = g \cdot \bar{v}$.

Suppose by contradiction that the two orbits $\overline{G \cdot \bar{v}}$ and $\overline{T \cdot k \cdot v}$ are disjoint for all $k \in K$. Assume we can separate any of these closed orbits with a function f_k. Exploiting the compactness of K, we can extract finitely many functions for the job and construct a single function f which separates $\overline{TK \cdot v}$ and $\overline{G \cdot \bar{v}}$. Since $K \cdot \bar{v} \subset G \cdot \bar{v}$, we can then separate the orbits $\overline{TK \cdot v}$ and $K \cdot \bar{v}$. But this implies that $\bar{v} \notin K(\overline{TK \cdot v})$ and so $\bar{v} \notin \overline{G \cdot v}$, which is absurd. $\square$

We discuss separation of orbits in the next subsection.

18.5.1 Separation of closed orbits

Consider a subset of coordinate indices $I \subset \{1, \ldots, N\}$ and let U_I be the subset of vectors whose non-zero coordinates belongs to I: $U_I = \{v \in V : \ v_i \neq 0$ if and only if $i \in I\}$.

Lemma 9. *The orbit $T \cdot v$ is closed if and only if there is a convex combination $\{\theta_i\}$ of $\{\alpha_i\}$ such that $\sum_i \theta_i \alpha_i = 0$.*

Given a closed orbit $\mathcal{O}_1$, consider the corresponding $\theta := \{\theta_i\}$ given by the above lemma. Define the function $f_\theta \colon V \to \mathbb{R}$ as

$$f_\theta(v) := \begin{cases} \prod_{i=1}^N v_i^{\theta_i} & \text{if } i \in I \\ 0 & \text{otherwise .} \end{cases}$$

The function f_θ is continuous. Moreover, by using the representation of the action of T, we see that f_θ is also T-invariant, indeed

$$f_\theta(\exp(\lambda) \cdot v) = \prod_{i=1}^N (e^{\langle \lambda, \alpha_i \rangle} v_i)^{\theta_i} = e^{\langle \lambda, \sum_i \alpha_i \theta_i \rangle} f_\theta(v) = f_\theta(v).$$

Intuitively, if two closed orbits are distinct, there must exist a zero combination of α_i for one orbit that is not zero for the other one. In other words, there exists θ such that the map f_θ separates the two orbits.

Lemma 10. *Let $\mathcal{O}_1, \mathcal{O}_2$ be two distinct, closed T-orbits. Then there exists $\theta = \{\theta_i\}$ such that $f_\theta(\mathcal{O}_1) \neq f_\theta(\mathcal{O}_2)$.*

References

[KN] Kempf, G. and Ness, L., *The length of vectors in representation spaces.* Algebraic geometry. Springer, Berlin, Heidelberg, 1979. 233–243.

[BL] Böhm, C. and Lafuente, R. A., *Real geometric invariant theory.* arXiv:1701.00643 (2017).

[Gr] Gressman, P. T., *L^p-improving estimates for Radon-like operators and the Kakeya-Brascamp-Lieb inequality.* Advances in Mathematics, **387**, (2021), 107831.

Gianmarco Brocchi, University of Birmingham
email: `gianmarcobrocchi@gmail.com`

19 The Brascamp-Lieb constant via invariant polynomials

After P. T. Gressman [Gre21]

A summary written by Rajula Srivastava

Abstract

The goal of the section is to establish the existence of certain invariant polynomials, which yield quantitative information about the Brascamp-Lieb constant. Further, a concrete characterization of such polynomials is also given.

19.1 Introduction

For each $j = 1, \ldots, m$, let $\pi_j : \mathbb{R}^n \to \mathbb{R}^{n_j}$ denote an arbitrary linear map and let $p_j \in [0, 1]$ be a real number. The Brascamp-Lieb constant $\mathrm{BL}(\{\pi_j, p_j\}_{j=1}^m)$ is defined to be the smallest non-negative real number such that

$$\int_{\mathbb{R}^n} \prod_{j=1}^{m} (f_j(\pi_j(x)))^{p_j} \mathrm{d}x \lesssim \mathrm{BL}(\{\pi_j, p_j\}_{j=1}^m) \prod_{j=1}^{m} \left(\int_{\mathbb{R}^{n_j}} f_j \right)^{p_j} \tag{1}$$

for all non-negative measurable functions $f_j \in L^1(\mathbb{R}^{n_j})$. Throughout this discussion, we shall assume that the dimensions n_j, n and the exponents p_j satisfy the relation

$$\sum_{j=1}^{m} \frac{p_j n_j}{n} = 1, \tag{2}$$

which, via a scaling argument, can be seen to be a necessary condition for the Brascamp-Lieb constant in (1) to be finite. It was shown by Lieb [Lie90] that any Brascamp-Lieb inequality has an extremizing sequence of Gaussians, which implies that

$$[\mathrm{BL}(\{\pi_j, p_j\}_{j=1}^m)]^{-1} = \inf_{A_1 \in \mathrm{GL}_{n_1}, \ldots, A_m \in \mathrm{GL}_{n_m}} \left[\frac{\det \left(\sum_{j=1}^m p_j \pi_j^* A_j A_j^* \pi_j \right)}{\prod_{j=1}^m (\det A_j^* A_j)^{p_j}} \right]^{\frac{1}{2}}, \tag{3}$$

where GL_{n_j} denotes the Lie group of invertible $n_j \times n_j$ matrices.

The following lemma establishes an analogous relationship between the Brascamp-Lieb constant and an infimum over the product of the *Hilbert-Schmidt norm* of operators of the form $A_j \pi_j A^*$, where the matrices A_j, A now vary over the *Special Linear* Lie groups SL_{n_j} and SL_n respectively (the Lie group formed by matrices of determinant equal to 1).

Lemma 1. *We have*

$$[\mathrm{BL}(\{\pi_j, p_j\}_{j=1}^m)]^{-1} = \inf_{\substack{A_1 \in \mathrm{SL}_{n_1}, \ldots, A_m \in \mathrm{SL}_{n_m}, \\ A \in \mathrm{SL}_n}} \prod_{j=1}^m n_j^{-\frac{p_j n_j}{2}} |||A_j \pi_j A^*|||^{p_j n_j},$$

where $|||\cdot|||$ *denotes the Hilbert-Schmidt norm computed with respect to the standard basis.*

Proof Sketch. For any $A \in \mathrm{SL}_n$ and $A_j \in \mathrm{GL}_{n_j}$ we can relate the Hilbert-Schmidt norm

$$\sum_{j=1}^m p_j |||A_j \pi_j A^*|||^2 = \sum_{j=1}^m p_j \mathrm{tr}\left(A \pi_j^* A_j^* A_j \pi_j A^*\right) = \mathrm{tr}\left(\sum_{j=1}^m p_j A \pi_j^* A_j^* A_j \pi_j A^*\right)$$

to the determinant of the matrix $\sum_{j=1}^m p_j A \pi_j^* A_j^* A_j \pi_j A^*$ via the AM-GM inequality, applied to its eigen values (all non-negative), as follows

$$\left|\sum_{j=1}^m \frac{p_j}{n} |||A_j \pi_j A^*|||^2\right|^n \geq \det \sum_{j=1}^m p_j A \pi_j^* A_j^* A_j \pi_j A^* = \det \sum_{j=1}^m p_j \pi_j^* A_j^* A_j \pi_j.$$

Taking an infimum of the left hand side over $A \in \mathrm{SL}_n$ converts the inequality to an equality (we can see this by choosing certain specific values for A in terms of π_j, A_j). Using (3), we conclude that

$$[\mathrm{BL}(\{\pi_j, p_j\}_{j=1}^m)]^{-1} = \inf_{\substack{A_1 \in \mathrm{GL}_{n_1}, \ldots, A_m \in \mathrm{GL}_{n_m}, \\ A \in \mathrm{SL}_n}} \left[\frac{\sum_{j=1}^m p_j |||A_j \pi_j A^*|||^2}{n \prod_{j=1}^m |\det A_j|^{\frac{2p_j}{n}}}\right]^{\frac{n}{2}}.$$

To convert an infimum over $A_j \in \mathrm{GL}_{n_j}$ to one over SL_{n_j}, we express each matrix A_j as a non-zero constant t_j times a matrix of determinant 1. Another application of the AM-GM inequality, along with the fact that $\sum_{j=1}^m \frac{p_j n_j}{n} = 1$, then yields

$$\inf_{t_1 > 0, \ldots, t_m > 0} t_1^{-\frac{2p_1 n_1}{n}} \ldots t_m^{-\frac{2p_m n_m}{n}} \sum_{j=1}^m \frac{p_j n_j}{n} t_j^2 \frac{|||A_j \pi_j A^*|||^2}{n_j} = \prod_{j=1}^m \left(\frac{|||A_j \pi_j A^*|||^2}{n_j}\right)^{\frac{p_j n_j}{2}}.$$

This proves the lemma. ∎

19.2 How do invariant polynomials enter the picture?

Consider a group representation ρ of $SL_{n_1} \times \ldots \times SL_{n_m} \times SL_n$ defined by

$$\rho_{(A_1,\ldots,A_m,A)}(\{\pi_j\}_{j=1}^m) := \{A_j \pi_j A^*\}_{j=1}^m.$$

We now turn our attention to a polynomial function Φ of the matrices $\{\pi_j\}_{j=1}^m$, which is homogeneous of degree $d_j > 0$ in each π_j, that is,

$$\Phi(\{\lambda_j \pi_j\}_{j=1}^m) = \lambda_1^{d_1} \ldots \lambda_m^{d_m} \Phi(\{\pi_j\}_{j=1}^m) \text{ for all } \lambda_1, \ldots, \lambda_m \in \mathbb{R}, \tag{4}$$

and is $\rho-$invariant, i.e.,

$$\Phi(\{A_j \pi_j A^*\}_{j=1}^m) = \Phi(\{\pi_j\}_{j=1}^m) \tag{5}$$

for all $A_j \in \mathrm{SL}_{n_j}$ and $A \in \mathrm{SL}_n$. We also assume that the degrees d_j satisfy the relation

$$\frac{p_1 n_1}{d_1} = \ldots = \frac{p_m n_m}{d_m} = \frac{1}{s_\Phi} \tag{6}$$

for some $s_\Phi \in \mathbb{R}$. Define $|||\Phi|||$ to be the maximum of $|\Phi|$ over all $m-$tuples $\{\tilde{\pi}_j\}_{j=1}^m$ such that $|||\tilde{\pi}_j||| \leq 1$ for all $j = 1, \ldots, m$. Then for all inputs $\{\pi_j\}_{j=1}^m$, using (4) and (5), we have

$$|\Phi(\{\pi_j\}_{j=1}^m)| = |\Phi(\{A_j \pi_j A^*\}_{j=1}^m)| \leq |||\Phi||| \prod_{j=1}^m |||A_j \pi_j A^*|||^{d_j}.$$

Lemma 1 then implies that

$$[\mathrm{BL}(\{\pi_j, p_j\}_{j=1}^m)]^{-1} \geq \prod_{j=1}^m n_j^{-\frac{p_j n_j}{2}} |||\Phi|||^{-\frac{1}{s_\Phi}} |\Phi(\{\pi_j\}_{j=1}^m)|^{\frac{1}{s_\Phi}}, \tag{7}$$

thus relating the Brascamp-Lieb constant to invariant polynomials satisfying properties (4), (5) and (6). The following lemma establishes that the collection of all such invariant polynomials can in fact be used to compute the order of magnitude of the Brascamp-Lieb constant, thus strengthening inequality (7) above.

Lemma 2. *Let* $\{p_j\}_{j=1}^m \in (0,1]^m$ *be rational exponents satisfying* (2). *Let* IP *denote the collection of all non-zero invariant polynomials satisfying* (4), (5) *and* (6). *Then*

$$[\mathrm{BL}(\{\pi_j, p_j\}_{j=1}^m)]^{-1} \approx \sup_{\Phi \in \mathrm{IP}} |||\Phi|||^{-\frac{1}{s_\Phi}} |\Phi(\{\pi_j\}_{j=1}^m)|^{\frac{1}{s_\Phi}}, \tag{8}$$

with implicit constants independent of $\{\pi_j\}_{j=1}^m$*. Further, there exists a finite subset* $\mathrm{IP}_0 \subset \mathrm{IP}$ *such that*

$$\sup_{\Phi\in\mathrm{IP}} |||\Phi|||^{-\frac{1}{s_\Phi}} |\Phi(\{\pi_j\}_{j=1}^m)|^{\frac{1}{s_\Phi}} \approx \sup_{\Phi\in\mathrm{IP}_0} |||\Phi|||^{-\frac{1}{s_\Phi}} |\Phi(\{\pi_j\}_{j=1}^m)|^{\frac{1}{s_\Phi}}.$$

Proof Sketch. Due to (7), it suffices to show the required upper bound and to assume without loss of generality that the left hand side of (8) is positive. The proof proceeds by contradiction. Suppose that the required upper bound does not hold. Then, using the homogeneity of both sides of (8) in $\{\pi_j^N\}_{j=1}^m$ and the ρ-invariance of Φ, one can show that for each positive integer N, there exist data $\{\pi_j^N\}_{j=1}^m$ such that

$$1 = [\mathrm{BL}(\{\pi_j^N, p_j\}_{j=1}^m)]^{-1} > N \sup_{\Phi\in\mathrm{IP}} |||\Phi|||^{-\frac{1}{s_\Phi}} |\Phi(\{\pi_j^N\}_{j=1}^m)|^{\frac{1}{s_\Phi}}, \tag{9}$$

and $\lim_{N\to\infty} |||\pi_j^N||| = n_j^{\frac{1}{2}}$ for each $j = 1, \ldots, m$. Passing to a subsequence in N, π_j^N converges to some limiting data π_j^∞ for each j. It is not hard to see that $\mathrm{BL}(\{\pi_j^\infty, p_j\}_{j=1}^m) = 1$ as well. On the other hand, by the continuity of each $\Phi \in \mathrm{IP}$ and (9), we conclude that $\sup_{\Phi\in\mathrm{IP}} |||\Phi|||^{-\frac{1}{s_\Phi}} |\Phi(\{\pi_j^\infty\}_{j=1}^m)|^{\frac{1}{s_\Phi}} = 0$.

The desired contradiction is then derived by coming up with a polynomial $\Phi \in \mathrm{IP}$ such that $\Phi(\{\pi_j^\infty\}_{j=1}^m) \neq 0$. Instead of constructing such a Φ directly, we use the theory of invariant polynomials on the vector space V of multilinear maps of a specific form, as discussed in the next subsection. Further, the finite set IP_0 would be taken to be the polynomials associated to any finite generating set of the invariant polynomial algebra on V. $\square$

19.3 Invariant polynomials acting on multilinear maps

We first describe the basic setup. Since each exponent p_j in Lemma 2 is rational, there exist positive integers $q_1, \ldots, q_m, q$ such that $p_j n_j = \frac{q_j}{q}$ for $j = 1, \ldots, m$. We shall be interested in maps of the form

$$\Pi(\{x_i^1, y_i^1\}_{i=1}^{q_1}, \ldots, \{x_i^m, y_i^m\}_{i=1}^{q_m}),$$

linear in each $x_j^i \in \mathbb{R}^{n_j}$ and each $y_j^i \in \mathbb{R}^n$ for $i = 1, \ldots, q_j$ and $j = 1, \ldots, m$. The group $\mathrm{SL}_{n_1} \times \ldots \times \mathrm{SL}_{n_m} \times \mathrm{SL}_n$ acts on the vector space of all such Π as under

$$\begin{aligned}\rho_{(A_1,\ldots,A_m,A)}&\Pi(\{x_i^1, y_i^1\}_{i=1}^{q_1}, \ldots, \{x_i^m, y_i^m\}_{i=1}^{q_m}) \\ &= \Pi(\{A_1^* x_i^1, A^* y_i^1\}_{i=1}^{q_1}, \ldots, \{A_m^* x_i^m, A^* y_i^m\}_{i=1}^{q_m}).\end{aligned}$$

The advantage of working with a multilinear map Π of the above form is that we can make use of the following fundamental result from Geometric Invariant Theory, which gives a sufficient condition for the existence of a homogeneous, ρ-invariant polynomial P such that $P(\Pi) = 1$.

Proposition 3. *Let V be the vector space of all multilinear maps*

$$\Pi(\{x_i^1, y_i^1\}_{i=1}^{q_1}, \ldots, \{x_i^m, y_i^m\}_{i=1}^{q_m})$$

as defined above. If $\Pi \in V$ has the property that

$$\inf_{g \in G} |||\rho_g \Pi||| = |||\Pi||| > 0,$$

where $G := \mathrm{SL}_{n_1} \times \ldots \times \mathrm{SL}_{n_m} \times \mathrm{SL}_n$, ρ is as defined above and $|||\cdot|||$ is the Hilbert-Schmidt norm on V with respect to the standard basis, then there exists a homogeneous, nonconstant ρ-invariant polynomial P on V such that $P(\Pi) = 1$.

Recall that the last step in the proof of Lemma 2 is to show the existence of a polynomial $\Phi \in \mathrm{IP}$ such that $\Phi(\{\pi_j^\infty\}_{j=1}^m) = 0$. To exploit the proposition above, we define a multilinear map $\Pi^\infty \in V$ associated to the data $\{\pi_j^\infty\}_{j=1}^m$ by

$$\Pi^\infty(\{x_i^1, y_i^1\}_{i=1}^{q_1}, \ldots, \{x_i^m, y_i^m\}_{i=1}^{q_m}) := \prod_{j=1}^{m} \prod_{i=1}^{q_j} \langle x_i^j, \pi_j^\infty y_i^j \rangle. \tag{10}$$

Then $|||\rho_{(A_1,\ldots,A_m,A)}\Pi^\infty||| = \prod_{j=1}^m \prod_{i=1}^{q_j} |||A_j \pi_j^\infty A^*|||^{q_j}$ and, using the properties of $\{\pi_j^\infty\}_{j=1}^m$ and the fact that $[\mathrm{BL}(\{\pi_j^\infty, p_j\}_{j=1}^m)]^{-1} = 1$, we obtain that

$$|||\Pi^\infty||| = \inf_{\substack{A_1 \in \mathrm{SL}_{n_1}, \ldots, A_m \in \mathrm{SL}_{n_m}, \\ A \in \mathrm{SL}_n}} |||\rho_{(A_1,\ldots,A_m,A)}\Pi^\infty||| = \prod_{j=1}^{m} n_j^{\frac{n_j p_j q}{2}} > 0.$$

Thus, Proposition 3 is applicable and we conclude that there exists a nonconstant, homogeneous ρ-invariant polynomial P, say of degree d, such that $P(\Pi^\infty) = 1$. Since each entry of Π^∞ is itself a product of the entries of π_j^∞, we may view $P(\Pi^\infty)$ as a polynomial in the entries of the π_j^∞. More precisely, regarding $\pi_1, \ldots, \pi_j$ as matrices of indeterminates, we define Π as in (10), replacing $\{\pi_j^\infty\}_{j=1}^m$ by $\{\pi_j\}_{j=1}^m$. Finally, we define $\Phi(\{\pi_j\}_{j=1}^m) := P(\Pi)$. It is then straightforward to check that $\Phi \in \mathrm{IP}$ (with $d_j = dq_j$) and $\Phi(\{\pi_j^\infty\}_{j=1}^m) = 1$, thus yielding the desired contradiction in the proof of Lemma 2. Further, the finite subset IP_0 can be taken to be consisting of only polynomials of the form $P(\Pi)$, for P belonging to any finite generating set of the ρ-invariant algebra on V.

19.4 Concrete characterization of the invariant polynomials

The final result is a concrete characterization of the class IP in terms of polynomials expressible as determinants of block-form matrices, with each block entry being a constant multiple of π_j for some $j = 1, \ldots, m$.

Lemma 4. *Let* $\{p_j\}_{j=1}^m \in (0,1]^m$ *be rational exponents satisfying* (2)*,* s *be a positive integer such that* $p_j s$ *is an integer for all* $j = 1, \ldots, m$*, and let* V_s *be the vector space of all polynomials* Φ *satisfying* (4)*,* (5) *and* (6) *for* $s_\Phi = s$*. Then* V_s *is spanned by polynomials of the form* $\det M(\{\pi_j\}_{j=1}^m)$*, where* $M(\{\pi_j\}_{j=1}^m)$ *is an* $ns \times ns$ *matrix consisting of block elements of size* $n_j \times n$ *for* $j = 1, \ldots, m$*, arranged in the following way:*

- *Each block entry is a constant multiple of* π_j *for some* j*.*
- *For each* j*, there are* $p_j s$ *block rows of height* n_j*. In each such block row, all block entries are multiples of* π_j*. At most* n_j *of these block entries are non-zero.*
- *There are* s *block columns of width* n*. In each block column, there are at most* n *non-zero block entries.*

For an illustration of the structure of the matrices M, we refer the reader to Fig 1 in [Gre21]. The proof of the above lemma makes use of polynomials acting on multilinear maps, as defined in the previous section, along with an application of the "Cayley Ω process" used to generate the invariants associated to the group SL_n.

References

[Gre21] Gressman, P.T., *L^p-improving estimates for Radon-like operators and the Kakeya-Brascamp-Lieb inequality.* Adv. Math. 387 (2021), 57 pp.

[Lie90] Lieb E.H., *Gaussian kernels have only Gaussian maximizers.* Invent. Math. 02(1) (1990), 179–208.

Rajula Srivastava, UW-Madison
email: rsrivastava9@wisc.edu

20 Quantum Brascamp-Lieb dualities

After M. Berta, D. Sutter, and M. Walter [BSW]

A summary written by Sean Sovine

Abstract

Brascamp-Lieb inequalities have a dual formulation as entropy inequalities. In this work, the authors introduce a fully quantum version of this duality, relating quantum relative entropy inequalities to matrix exponential inequalities of Young type, and demonstrate applications of this duality by means of examples from quantum information theory.

20.1 Introduction

The Brascamp-Lieb (BL) problem asks for the optimal constant $C \in \mathbb{R}$ such that

$$\int_{\mathbb{R}^m} \prod_{k=1}^{n} f_k(L_K x) dx \leq \exp(C) \prod_{k=1}^{n} \|f_k\|_{1/q_k}, \tag{1}$$

holds for all non-negative functions $f_k \colon \mathbb{R}^{m_k} \to \mathbb{R}_+$, where $L_k \colon \mathbb{R}^m \to \mathbb{R}^{m_k}$ are given surjective linear maps and $q_k \in \mathbb{R}_+$ for $k = 1, \ldots, n$. The (BL) problem has a dual entropic formulation in terms of the differential entropy, which is defined for a probability distribution g as

$$H(g) := -\int g(x) \log g(x) dx.$$

Regarding the duality for Brascamp-Lieb, the inequality (1) holds for all such functions f_k, for a given $C \in \mathbb{R}$, if and only if

$$H(g) \leq \sum_{k=1}^{n} q_k H(g_k) + C$$

holds for all probability distributions g on $\mathbb{R}^m$ with finite differential entropy. Here g_k denotes the marginal probability density on $\mathbb{R}^{m_k}$ corresponding to L_k, defined by

$$\int_{\mathbb{R}^m} \phi(L_k x) g(x) dx = \int_{\mathbb{R}^{m_k}} \phi(y) g_k(y) dy$$

for all bounded continuous functions ϕ on $\mathbb{R}^{m_k}$.

Berta, Sutter, and Walter prove the following quantum version of the Brascamp-Lieb duality, which is the main theorem of their paper.

Theorem 1. *Let $n \in \mathbb{N}$, $\vec{q} = (q_1, \dots, q_n) \in \mathbb{R}_+^n$, $\vec{\mathcal{E}} = (\mathcal{E}_1, \dots, \mathcal{E}_n)$ with $\mathcal{E}_k \in TPP(A, B_k)$ for $k \in 1, \dots, n$, $\sigma \in P_\succ(A)$, $\vec{\sigma} = (\sigma_1, \dots, \sigma_n)$ with $\sigma_k \in P_\succ(B_k)$ for $k \in 1, \dots, n$, and $C \in \mathbb{R}$. Then the following two statements are equivalent:*

$$\sum_{k=1}^{n} q_k D(\mathcal{E}_k(\rho)||\sigma_k) \leq D(\rho||\sigma) + C \quad \forall \rho \in S(A), \tag{2}$$

$$\operatorname{tr}\exp\left(\log\sigma + \sum_{k=1}^{n} \mathcal{E}_k^\dagger(\log\omega_k)\right) \leq \exp(C) \prod_{k=1}^{n} \|\exp(\log\omega_k + q_k \log\sigma_k)\|_{1/q_k} \;\; \forall \omega_k \in P_\succ(B_k). \tag{3}$$

Here A, B_k are Hilbert spaces; $P_\succ(A)$ ($P_\succeq(A)$) is the set of positive-definite (resp. positive-semidefinite) operators on A; $S(A)$ is the set of bounded, positive-definite Hermitian operators on A with trace 1, which are also called density operators and represent quantum states; and $TPP(A, B_k)$ is the class of trace-preserving, positive maps from $L(A)$ to $L(B_k)$, which represent operations between two quantum systems. For each $\mathcal{E} \in TPP(A, B_k)$ we define the *adjoint map* $\mathcal{E}^\dagger$ by duality for the Hilbert-Schmidt inner product:

$$\operatorname{tr}\mathcal{E}(X)^\dagger Y = \operatorname{tr} X^\dagger \mathcal{E}^\dagger(Y) \quad \text{for all} \quad X \in T(A),$$

where $T(A)$ is the set of trace-class operators on A. For a density operator $\omega \in S(A)$ and $\tau \in P_\succeq(A)$, the *quantum relative entropy* of ω with respect to τ is defined by functional calculus as

$$D(\omega||\tau) = \operatorname{tr}\omega(\log\omega - \log\tau) \;\text{ if } \omega \ll \tau \text{ and } +\infty \text{ otherwise},$$

where $\omega \ll \tau$ means that the support of ω is contained in the support of τ. For comparison the *von Neumann entropy* of a density operator $\rho \in S(A)$ is defined as

$$H(\rho) := -\operatorname{tr}\rho\log\rho.$$

The main tool that is used in the proof of Theorem 1 is the following variational formula for the quantum relative entropy proved by Dénes Petz:

Theorem 2 ([Petz])**.** *Let $\sigma \in P_\succ(A)$. Then:*

- *For all $\rho \in S(A)$ we have*

$$D(\rho||\sigma) = \sup_{\omega \in \mathrm{P}_\succeq(A)} \{\operatorname{tr}\rho\log\omega - \log\operatorname{tr}\exp(\log\omega + \log\sigma)\}.$$

  Furthermore the supremum is attained for

$$\omega = \exp(\log\rho - \log\sigma)/\operatorname{tr}\exp(\log\rho - \log\sigma).$$

- *For all* $H \in \mathrm{Herm}(A)$*, we have*

$$\log \operatorname{tr} \exp(H + \log \sigma) = \sup_{\omega \in \mathrm{S}(A)} \{\operatorname{tr} H\omega - D(\omega || \sigma)\}.$$

  Furthermore the supremum is attained for

$$\omega = \exp(H + \log \sigma) / \operatorname{tr} \exp(H + \log \sigma).$$

The authors mainly consider the case where A, B_k are finite-dimensional Hilbert spaces, which is the case of most relevance to quantum information theory. However, the authors point out that Petz's variational formula for the relative entropy holds in the more general setting of a von Neumann algebra. Thus their main theorem can also be proved analogously in the setting of a general Hilbert space, with appropriate definitions and hypotheses.

20.2 Applications

The authors present a variety of applications of their main result. One example of these applications is a dual form of the data processing inequality for the quantum relative entropy. The *data-processing inequality* states that the quantum relative entropy is not increased by the application of a quantum channel $\mathcal{E} \in TPP(A, B)$:

$$D(\mathcal{E}(\rho) || \mathcal{E}(\sigma)) \leq D(\rho || \sigma),$$

for $\rho \in \mathrm{S}(A)$ and $\sigma \in \mathcal{P}_{\succ}(A)$. Application of the authors' quantum Brascamp-Lieb duality gives the following duality relation for the data-processing inequality:

Corollary 3. *For* $\sigma \in P_{\succ}(A)$ *and* $\mathcal{E} \in TPP(A, B)$ *the following inequalities hold and are equivalent:*

$$D(\mathcal{E}(\rho) || \mathcal{E}(\sigma)) \leq D(\rho || \sigma), \tag{4}$$

$$\operatorname{tr} \exp(\log \sigma + \mathcal{E}^{\dagger}(\log \omega)) \leq \operatorname{tr} \exp(\log \omega + \log \mathcal{E}(\sigma)) \quad \forall \omega \in \mathrm{P}_{\succ}(B). \tag{5}$$

References

[BSW] Berta, M., Sutter, D., and Walter, M., *Quantum Brascamp-Lieb Dualities.* arXiv preprint, arXiv:2002.08055;

[Petz] Petz, D., *A variational expression for the relative entropy.* Communications in Mathematical Physics, 114(2) (1988), pp. 345-349.

Sean Sovine, Virginia Tech
email: sovine5@vt.edu

21 Nonlinear Brascamp-Lieb Inequality

After J. Bennet, N. Bez, S. Buschenhenke, M. G. Cowling, and T. C. Flock. [1] and after J. Duncan. [2]

A summary written by Yu-Hsiang Lin

Abstract

We show a local nonlinear Brascamp-Lieb inequality through the existence of gaussian near-extremisers with input in controlled(effective version of Lieb's theorem) and nonlinear versions of Ball's inequality.(The main difference between [1] and [2] is that they expand the two different implications of Ball's inequality to the nonlinear case.)

21.1 Introduction

Although there are some other applications in [1] and [2], we will focus on two kinds of proof of the following theorem of local nonlinear Brascamp-Lieb inequality in this talk.

Theorem 1 (Main Theorem)**.** *Let* $(\boldsymbol{L}, \boldsymbol{p})$ *be a Brascamp-Lieb datum. Suppose that* $B_j : \mathbb{R}^n \longrightarrow \mathbb{R}^{n_j}$ *are* C^2 *submersions in a neighborhood of* x_0*, and* $dB_j(x_0) = L_j$ *for* $1 \leq j \leq m$*. Then for each* $\varepsilon > 0$*, there exists a neighborhood* U *of* x_0 *such that*

$$\int_U \prod_{j=1}^{m} f_j^{p_j}(B_j(x))dx \leq (1+\varepsilon)BL(\boldsymbol{L}, \boldsymbol{p}) \prod_{j=1}^{m} (\int_{\mathbb{R}^{n_j}} f_j(x_j)dx_j)^{p_j}$$

To show the above theorem, we need the concept of **near-extremisers** and the associate effective Lieb theorem which a quantitative description telling us in what range can we find an input closed enough to be an extremiser. We say an input $\mathbf{f} = (f_j)_{1\leq j\leq m}$ is a δ-near-extremiser if

$$BL(\mathbf{L}, \mathbf{p}, \mathbf{f}) := \frac{\int_{\mathbb{R}^n} \prod_{j=1}^{m} f_j^{p_j}(L_j x)dx}{\prod_{j=1}^{m} (\int_{\mathbb{R}^{n_j}} f_j(x_j)dx_j)^{p_j}} \geq (1-\delta)BL(\mathbf{L}, \mathbf{p})$$

We have the following theorem for δ-near-extremiser.

Theorem 2 (Effective Lieb's Theorem)**.** *Suppose that* $(\boldsymbol{L}_0, \boldsymbol{p})$ *is a Brascamp-Lieb datum with finite Brascamp-Lieb constant. Then there exist* $N \in \mathbb{N}$ *and* $\delta_0 > 0$ *such that for* $\delta \in (0, \delta_0)$*, we have*

$$\sup_{\|A_j\|, \|A_j^{-1}\| \leq \delta^{-N}} BL_g(\boldsymbol{L}, \boldsymbol{p}, \boldsymbol{A}) \geq (1-\delta)BL(\boldsymbol{L}, \boldsymbol{p})$$

for all $\boldsymbol{L}$ *sufficiently close to* $\boldsymbol{L}_0$*.*

By Ball's inequality, we have

$$BL(\mathbf{L}, \mathbf{p}, \mathbf{f})BL(\mathbf{L}, \mathbf{p}, f) \leq \sup_x BL(\mathbf{L}, \mathbf{p}, \mathbf{h}^x)BL(\mathbf{L}, \mathbf{p}, \mathbf{f} * \mathbf{g})$$

where $h_j^x(z) := f_j(z)g_j(L_jx - z)$, and $\mathbf{h}^x := (h_j^x)_{j=1}^m$, $\mathbf{f} * \mathbf{g} := (f_j * g_j)_{j=1}^m$. If $\mathbf{g}$ is a δ-near-extremiser, then we have

$$BL(\mathbf{L}, \mathbf{p}, \mathbf{f}) \leq (1 + O(\delta)) \sup_x BL(\mathbf{L}, \mathbf{p}, \mathbf{h}^x)$$

$$BL(\mathbf{L}, \mathbf{p}, \mathbf{f}) \leq (1 + O(\delta))BL(\mathbf{L}, \mathbf{p}, \mathbf{f} * \mathbf{g})$$

where the nonlinear variant of the above two inequality will play a crucial role in the proof of **Theorem** 1 in [1] and [2] respectively. Now we first follow the method in [1].

21.2 Proof of Main Theorem in [1]

Here we first introduce the concept of **local constant function**. For $\kappa > 1, \mu > 0, \Omega \subseteq \mathbb{R}^d$, a measurable subset, we say a nonnegative function f is κ-constant at scale μ on Ω if $f(x) \leq \kappa f(y)$ for $x \in \Omega$ and $y \in \mathbb{R}^d$ with $|x - y| \leq \mu$. We denote by $L^1(\Omega, \mu, \kappa)$ the subset of $L^1(\mathbb{R}^d)$ with this property. Note that the d-dimension Poisson kernel $P_t(x) := c_n \frac{t}{(t^2+\|x\|^2)^{\frac{n+1}{2}}}$ is κ-constant on scale μ on all $\mathbb{R}^d$ for μ small enough. Then $f * P_t(x)$ inherits this regularity property, and we can see that we can approximate $f \in L^1(\Omega)$ almost everywhere by functions in $L^1(\Omega, \mu, \kappa)$. Also note that we have the property, if $f \in L^1(\Omega, \mu, \kappa), g \in L^1(\Omega, \mu, \lambda)$, then $fg \in L^1(\Omega, \mu, \kappa\lambda)$. In the following, for $\delta \in (0, 1)$, we define $U_\delta(y) := \{x \in \mathbb{R}^n \mid |x - y| \leq \delta\}$.

Definition 3 (localized and regularized nonlinear BL constant)**.** *Let $C(u, \delta, \mu, \kappa)$ denote the best constant in the inequality*

$$\int_{U_\delta(u)} \prod_{j=1}^m f_j^{p_j}(B_j(y))dy \leq C(u, \delta, \mu, \kappa) \prod_{j=1}^m (\int_{\mathbb{R}^{n_j}} f_j(x_j)dx_j)^{p_j}$$

over all inputs $f_j \in L^1(B_j(U_{2\delta}(u)), \mu, \kappa)$.

Proposition 4 (Base case)**.** *There exists a constant $\nu > 0$ such that for $u \in U_\nu(0)$, $\delta \in (0, \nu)$, $\delta^{\alpha+\beta} \leq \mu$(Here $1 < \alpha < 2$ and $0 < \beta < 2 - \alpha$), we have*

$$C(u, \delta, \mu, \kappa) \leq \kappa^\rho BL(d\boldsymbol{B}(u), \boldsymbol{p})$$

where $\rho = \sum_{j=1}^m p_j$.

Proposition 5 (Recursive Inequality)**.** *There exists a constant $\nu > 0$ such that for $u \in U_\nu(0)$, $\delta \in (0, \nu)$, $\delta^{\alpha+\beta} > \mu$, we have*

$$C(u, \delta, \mu, \kappa) \leq (1 + \delta^\sigma) \max_{x \in U_{2\delta(u)}} C(x, \delta^\alpha, \mu, \kappa \exp(\delta^\sigma))$$

where the constant σ will be clear later.

The proof of **Proposition** 5 is one of the technical parts in this paper, where we will use **Theorem** 2, and one of the nonlinear variant of Ball's inequality.

With the propositions above, now we can prove **Theorem** 1. Only to show there exists a constant δ_0 depending on ε such that

$$C(0, \delta_0, \mu, 1 + \varepsilon) \leq (1 + C\varepsilon) BL(d\mathbf{B}(0), \mathbf{p})$$

uniformly in $\mu > 0$. If $\delta_0^{\alpha+\beta} \leq \mu$, then by **Proposition** 4

$$C(0, \delta_0, \mu, 1 + \varepsilon) \leq (1 + \varepsilon)^\rho BL(d\mathbf{B}(0), \mathbf{p}) \leq (1 + C_\rho \varepsilon) BL(d\mathbf{B}(0), \mathbf{p})$$

Now if $\delta_0^{\alpha+\beta} \geq \mu$, let $\delta_k := \delta_0^{(\alpha^k)}$, then using **Proposition** 5 iteratively.(**Induction on scale**.)

$$\begin{aligned} C(0, \delta_0, \mu, 1 + \varepsilon) &\leq \prod_{k=0}^{k_*-1} (1 + \delta_k^\sigma) \max_{u \in \widetilde{U}_{k_*}} C(u, \delta_{k_*}, \mu, (1 + \varepsilon) \prod_{k=0}^{k_*-1} \exp(\delta_k^\sigma)) \\ &\leq \prod_{k=0}^{k_*-1} (1 + \delta_k^\sigma) \max_{u \in \widetilde{U}_{k_*}} [(1 + \varepsilon) \prod_{k=0}^{k_*-1} \exp(\delta_k^\sigma)]^\rho BL(d\mathbf{B}(u), \mathbf{p}) \end{aligned}$$

where $\widetilde{U}_{k_*} = U_{2\delta_0 + \ldots 2\delta_{k_*-1}}(0)$, K_* is the smallest integer such that $\delta_{k_*}^{\alpha+\beta} \leq \mu$. Hence we have

$$C(0, \delta_0, \mu, 1 + \varepsilon) \leq (1 + \varepsilon)^{\rho+2} BL(d\mathbf{B}(0), \mathbf{p}) \leq (1 + C_\rho \varepsilon) BL(d\mathbf{B}(0), \mathbf{p})$$

21.3 Proof of Main Theorem in [2]

Now we let $M, M_1, ..., M_m$ be complete Riemannian manifolds and $e_x : T_xN \to N$ be the exponential map on manifold N based at point x. By a **regularized version** of effective Lieb's theorem, for $B_j : M \overset{C^2}{\to} M_j$, for δ sufficient small, we may find positive definite matrix $A_{\delta,j}$ with $\|A_{\delta,j}\|_{W^{1,\infty}(M)}, \|detA_{\delta,j}\|_{W^{1,\infty}(M)} \leq \delta^{-\varepsilon}$ such that the associate L^1 normalized gaussians

$$G_{x,\delta,j}(z) := \delta^{-n_j} det(A_{\delta,j}(x))^{\frac{1}{2}} \exp(-\pi\delta^{-2}\langle A_{\delta,j}(x)z, z\rangle)$$

is δ^ε near extremiser for $(\mathbf{dB}(x), \mathbf{p})$. Now we define the associate flow operator. $H_{x,\delta,j} : L^1(M_j) \to L^1(U_{\rho-\delta^\gamma}(B_j(x)))$

$$H_{x,\delta,j}f_j(z) := \int_{U_{\delta^\gamma,j}(z)} f_j(w)G_{x,\delta,j}(e^{-1}_{B_j(x)}(z)e^{-1}_{B_j(x)}(w))dw$$

Theorem 6 (Nonlinear Ball's Inequality)**.** *For $B_j : M \overset{C^2}{\to} M_j$, with $\|\boldsymbol{dB}\|_{W^{1,\infty}}$, $\|BL(\boldsymbol{dB}, \boldsymbol{p})\|_{L^\infty} \leq C$ and δ sufficiently small, we have*

$$\int_M \prod_{j=1}^m f_j^{p_j} \circ B_j(x)dx \leq (1+\delta^\beta) \int_M \prod_{j=1}^m (H_{x,\delta,j}f_j)^{p_j} \circ B_j(x)dx$$

With the above theorem, now we can prove **Theorem** 1 again.

$$\begin{aligned}
\int_{U_{\delta^\gamma}(x_0)} \prod_{j=1}^m f_j^{p_j} \circ B_j(x)dx &\leq (1+\delta^\beta) \int_M \prod_{j=1}^m (H_{x,\delta,j}(f_j\chi_{U_{\delta^\gamma,j}(x_0)}))^{p_j} \circ B_j(x)dx \\
&\leq (1+\delta^\beta) \int_{2U_{\delta^\gamma}(x_0)} \prod_{j=1}^m (H_{x,\delta,j}f_j)^{p_j} \circ B_j(x)dx \\
&\leq (1+\delta^\beta)(1+\delta^\eta)^{2P} \int_{2U_{\delta^\gamma}(x_0)} \prod_{j=1}^m (\widetilde{H}_{x_0,\delta,j}f_j)^{p_j} \circ L_j^{x_0}(x)dx \\
&\leq (1+\delta^\beta)(1+\delta^\eta)^{2P} BL(\mathbf{L}, \mathbf{p}) \prod_{j=1}^m (\int_{2U_{\delta^\gamma}(x_0)} \widetilde{H}_{x_0,\delta,j}f_j \circ e_{B_j(x_0)})^{p_j} \\
&\leq (1+\delta^\beta)(1+\delta^\eta)^{2P} BL(\mathbf{L}, \mathbf{p}) \prod_{j=1}^m (\int_{M_j} f_j)^{p_j}
\end{aligned}$$

where $L_j^{x_0} := e_{B_j(x_0)} \circ dB_j(x_0)$, $\mathbf{L} := \mathbf{dB}(x_0)$ and $\widetilde{H}$ is an operator enlarging the integrand of H. In the third inequality, with the pertubations, we have some δ power loss. The fourth inequality is just the linear BL inequality.

Now reduce to show **Theorem** 6. Let $C(s,t)$ be the best constant such that

$$\int_M \prod_{j=1}^m (H_{x,s,j}f_j)^{p_j} \circ B_j(x)dx \leq C(s,t) \int_M \prod_{j=1}^m (H_{x,t,j}f_j)^{p_j} \circ B_j(x)dx$$

$C(s,t)$ has the submultiplicity property $C(s,t) \leq C(s,r)C(r,t)$. The technical part in this paper is to show there exist β and ν such that for $\delta \in (0,\nu)$, we have $C(\delta, 2^{\frac{1}{2}}\delta) \leq (1+\delta^\beta)$. Once we have this controll, now let $\delta_k := 2^{-\frac{k}{2}}\delta_0$, then

$$C(\delta_K, \delta_0) \leq \prod_{k=1}^K C(\delta_k, \delta_{k-1}) \leq \prod_{k=1}^K (1+\delta_k^\beta)$$

Then we have the estimation $C(\delta_K, \delta_0) \leq 1 + (\frac{\delta}{2})^\beta$. Hence

$$\begin{aligned}\int_M \prod_{j=1}^m f_j^{p_j} \circ B_j(x)dx &\leq \liminf_{K\to\infty} \int_M \prod_{j=1}^m (H_{x,\delta_K,j}f_j)^{p_j} \circ B_j(x)dx \\ &\leq \liminf_{K\to\infty} C(\delta_K, \delta) \int_M \prod_{j=1}^m (H_{x,\delta,j}f_j)^{p_j} \circ B_j(x)dx \\ &\leq (1+(\frac{\delta}{2})^\beta) \int_M \prod_{j=1}^m (H_{x,\delta,j}f_j)^{p_j} \circ B_j(x)dx\end{aligned}$$

which comletes the proof of **Theorem** 6.

References

[1] Bennett, Jonathan and Bez, Neal and Buschenhenke, Stefan and Cowling, Michael G and Flock, Taryn C, *On the nonlinear Brascamp–Lieb inequality.* Duke Mathematical Journal **169** (2020), no. 17, 3291-3338;

[2] Duncan, Jennifer, *A Nonlinear Variant of Ball's Inequality.* arXiv preprint arXiv:2101.07672.

Yu-Hsiang Lin, University of Bonn
email: `fred841211@gmail.com`

22 Subset entropy power inequality and its relation to a conjectural fractional Young inequality

After S. Bobkov, M. Madiman, and L. Wang [BMW11], M. Madiman and A. Barron [MB07], M. Madiman and F. Ghassemi [MG19]

A summary written by Julian Weigt

Abstract

The classical entropy power inequality by Shannon and Stam can be seen as a limiting case of the sharp Young's inequality. We discuss a recently proven fractional generalization of the classical entropy power inequality and its connection to a conjectured fractional Young inequality.

22.1 The entropy

Let X be a discrete random variable. Its *entropy* is

$$H(X) = -\sum_{i=1}^{n} P(X = x_i) \log P(X = x_i),$$

where $\{x_1, \dots, x_n\}$ is the range of X. The entropy represents the expected number of bits one needs in order to encode one sample of X. Differently phrased, the entropy of X is the expected amount of information one sample of X provides.

Let X be a real valued random variable and denote by f its probability density. We define the *differential entropy* of X by

$$h_1(X) = -\int_{\mathbb{R}} f(x) \log f(x)\ dx.$$

The differential entropy can also be defined for $\mathbb{R}^n$-valued random variables, but for simplicity we only consider the one-dimensional case $n = 1$ here. Even though the differential entropy might look similar to the entropy H, it is not the same and does not allow for such an information theoretical interpretation. A $\mathbb{Z}$-valued random variable has differential entropy $-\infty$, in particular random variables can have negative differential entropy. Any continuous random variable X should have positive infinite entropy, because when one approximates X by a sequence of discrete random variables their entropy tends to infinity. And lastly, the integral $h_1(X)$ does not make sense from a physical point of view, because the probability distribution f of random variable should have the inverse unit of the random variable, and we usually cannot take the logarithm of quantity with a unit. For brevity we refer to the differential entropy as entropy for the rest of the presentation.

The Shannon entropy power is defined by

$$V_1(X) = e^{2h_1(X)}.$$

The classical entropy power inequality of Shannon and Stam states

$$V_1(X_1 + \ldots + X_M) \geq V_1(X_1) + \ldots + V_1(X_M) \tag{1}$$

for independent random variables $X_1, \ldots, X_M$.

22.2 Connection to Young's inequality

One can deduce (1) from the sharp Young's inequality. If the random variables $X_1, \ldots, X_M$ have probability densities $f_1, \ldots, f_M$ then $X_1 + \ldots + X_M$ has probability density $f_1 \star \ldots \star f_M$. For $p > 1$ define

$$h_p(X) = \frac{p}{p-1} \log \|f\|_p.$$

Then $h_p(X) \to h_1(X)$ and hence for $V_p(X) = e^{2h_p(X)}$ we have $V_p(X) \to V_1(X)$. For $p > 1$ we have

$$V_p(X) = \|f\|_p^{-2p'}.$$

For $\frac{p'}{q_1'} + \ldots + \frac{p'}{q_M'} = 1$ the sharp Young's inequality states

$$C_p \|f_1 \star \ldots \star f_M\|_p \leq C_{q_1} \ldots C_{q_M} \|f_1\|_{q_1} \cdot \ldots \cdot \|f_M\|_{q_M}, \tag{2}$$

where $C_p^2 = \frac{p^{\frac{1}{p}}}{p'^{\frac{1}{p'}}}$. Note that $p \to 1$ implies $q_1, \ldots, q_M \to 1$. Applying the sharp Youngs inequality and taking the limit $p \to 1$ we obtain

$$\begin{aligned} V_p(X_1 + \ldots + X_M)^{-1} &= \|f_1 \star \ldots \star f_M\|_p^{2p'} \\ &\leq \prod_{i=1}^{M} \Big(\frac{p'}{q_i'} \|f_i\|_{q_i}^{2q_i'/n} \Big)^{\frac{p'}{q_i'}} \qquad \text{for } p \to 1 \\ &= \prod_{i=1}^{M} \Big(\frac{p'}{q_i'} / V_{q_i}(X_i) \Big)^{\frac{p'}{q_i'}} \\ &\to \Big(\sum_{i=1}^{M} V_1(X_i) \Big)^{-1} \qquad \text{for } p \to 1, \end{aligned}$$

where the last limit holds for taking q_i such that $\frac{p'}{q_i'} = \frac{V_1(X_i)}{\sum_{i=1}^M V_1(X_i)}$. We can conclude (1). The sharp constants of Young's inequality contribute the factors $\left(\frac{p'}{q_i'}\right)^{\frac{p'}{q_i'}}$ which are necessary for the proof.

22.3 The subset entropy power inequality

In [MB07] they prove the subset entropy power inequality, which is the entropy power inequality (1) with a more general expression on the right hand side. A hypergraph G on $\{1, \dots, M\}$ is an arbitrary set of subsets of $\{1, \dots, M\}$. The subset entropy power inequality states that for independent random variables $X_1, \dots, X_M$ and any hypergraph G on $\{1, \dots, M\}$ we have

$$V_1(X_1 + \ldots + X_M) \geq \frac{1}{d} \sum_{S \in \mathcal{G}} V_1\Big(\sum_{i \in S} X_i\Big), \tag{3}$$

where d is the degree of G, i.e. the smallest number such that every index $i = 1, \dots, M$ occurs in at most d sets in G. The subset entropy power inequality further generalizes. For a hypergraph G on $\{1, \dots, M\}$ we say that $(\beta_S)_{S \in G}$ is a fractional partition of G if for any $i = 1, \dots, M$ we have $\sum_{S \in G : i \in S} \beta_S = 1$. In [MG19] they prove

$$V_1(X_1 + \ldots + X_M) \geq \sum_{S \in G} \beta_S V_1\Big(\sum_{i \in S} X_i\Big) \tag{4}$$

as a consequence of (3). Clearly, (4) continues to hold if $\sum_{S \in G : i \in S} \beta_S \leq 1$. Thus (3) is a special case of (4) with $\beta_S = 1/d$.

22.4 The conjecture

Let G be a d-regular hypergraph, i.e. every index $i = 1, \dots, M$ is in d sets in G. In [BMW11] they conjecture the fractional Young inequality

$$\|f_1 \star \ldots \star f_M\|_r \leq \frac{1}{C_r} \prod_{S \in G} \left[C_{p_S} \| \star_{j \in S} f_j \|_{p_S}\right]^{\frac{1}{d}}, \tag{5}$$

with $\sum_{S \in G} \frac{r'}{p_S'} = 1$.

In [BMW11] they show that (5) implies (4) the same way as one can deduce the original entropy power inequality (1) from the sharp Young's inequality. It is not evident how one could show the reverse implication, because the deduction of (5) from (4) works the same way as the deduction of the entropy power inequality (1) from the sharp Young's inequality (2), which means it involves taking the limit $p \to 1$ and choosing the coefficients p'/q'_i suitably.

For $r \geq 2$ and $1 \leq p_S \leq 2$ the conjecture (5) follows from the Hausdorff-Young inequality $\|\hat{f}\|_r \leq \|f\|_{r'}$, and Hölders inequality on the Fourier side.

22.5 Proof of the subset entropy power inequality (3)

This is a rough summary of the proof in [MB07]. It suffices to consider the case that each $i = 1, \ldots, M$ occurs in precisely d sets in G because by (1) the entropy is monotone in the sense that $h_1(X+Y) \geq h_1(X)$ for any independent random variables X, Y. Let X be a random variable with probability density f. Its score function $\rho : \mathbb{R} \to [0, \infty)$ is defined by

$$\rho(x) = \frac{d}{dx} \log f(x) = \frac{f'(x)}{f(x)}.$$

The Fisher information of X is $I(X) = E(\rho(X)^2)$. It is related to the entropy via

$$h_1(X) = \frac{1}{2}\log(2\pi e) - \frac{1}{2}\int_0^\infty I(X + \sqrt{t}Z) - \frac{1}{1+t}\, dt, \tag{6}$$

where Z is a Gaussian random variable independent of X. In order to prove (3) they show the following bound for the Fisher information. For any weights $(w_S)_{S\in G}$ with $\sum_{S\in G} w_S = 1$ we have

$$I(X_1 + \ldots + X_M) = E(\rho(X_1 + \ldots + X_M)^2) \leq E\Big[\Big(\sum_{S\in G} w_S \rho\big(\sum_{i\in S} X_i\big)\Big)^2\Big]. \tag{7}$$

A *variance drop lemma* implies

$$E\Big[\Big(\sum_{S\in G} w_S^2 \rho\big(\sum_{i\in S} X_i\big)\Big)^2\Big] \leq d \sum_{S\in G} w_S E\big[\rho\big(\sum_{i\in S} X_i\big)^2\big]. \tag{8}$$

Note that (8) already follows from the Cauchy-Schwarz inequality with M instead of r, and that in the case $r = 1$ it follows from the independence of the random variables and from $E\rho(X_i) = 0$. Combining (7) and (8) we obtain

$$I(X_1 + \ldots + X_M) \leq d \sum_{S\in G} w_S^2 I\big(\sum_{i\in S} X_i\big),$$

which due to the homogeneity of I is equivalent to

$$I(X_1 + \ldots + X_M) \leq \sum_{S\in G} w_S I\Big(\sum_{i\in S} \frac{X_i}{\sqrt{d w_S}}\Big). \tag{9}$$

However, in view of (6) this is not exactly what we need. Instead, we would like a bound of the form

$$I(X_1 + \ldots + X_M + \sqrt{t}Z) \leq \sum_{S \in G} w_S I\Big(\sum_{i \in S} \frac{X_i}{\sqrt{d w_S}} + \sqrt{t} Z_S\Big), \tag{10}$$

where Z, Z_S are Gaussian variables independent from $X_1 + \ldots + X_M$ and from $\sum_{i \in S} X_i$. Turns out one can deduce (10) by applying (9) to the random variables $X_1, \ldots, X_M, Z_1, \ldots, Z_{M'}$, where $Z_1, \ldots, Z_{M'}$ are independent Gaussians, and constructing a hypergraph G' appropriately. Then plugging (10) into (6) it follows that

$$h_1(X_1 + \ldots + X_M) \leq \sum_{S \in G} w_S h_1\big(\sum_{i \in S} X_i\big) + \frac{H(w)}{2} - \frac{\log d}{2},$$

where $H(w)$ is the discrete entropy of the tuple $w = (w_S)_{S \in G}$. Exponentiating and taking $w_S = \frac{V_1(\sum_{i \in S} X_i)}{\sum_{T \in G} V_1(\sum_{i \in T} X_i)}$, the generalized entropy power inequality (3) follows.

References

[BMW11] S. Bobkov, M. Madiman, and L. Wang. *Fractional generalizations of Young and Brunn-Minkowski inequalities.* Concentration, functional inequalities and isoperimetry. Vol. 545. Contemp. Math. Amer. Math. Soc., Providence, RI, 2011, pp. 3553. `arXiv: 1006.2884`. `MR:2858464`.

[MB07] M. Madiman and A. Barron. *Generalized Entropy Power Inequalities and Monotonicity Properties of Information.* IEEE Transactions on Information Theory 53.7 (2007), pp. 2317-2329

[MG19] M. Madiman and F. Ghassemi. *Combinatorial entropy power inequalities: a preliminary study of the Stam region.* IEEE Trans. Inform. Theory 65.3 (2019), pp. 13751386. `arXiv: 1704.01177`. `MR: 3923175`.

Julian Weigt, Aalto University
email: `julian.weigt@aalto.fi`

23 Bounds on the Poincaré constant for convolution measures

After Thomas A. Courtade

A summary written by Jennifer Duncan

Abstract

In this paper, several results are established concerning the best constant for Poincaré inequalities satisfied by functions in $W^{1,2}(\mathbb{R}^d, \mu)$, where μ is a convolution of Borel probability measures.

We let $\mathcal{P}(\mathbb{R}^d)$ denote the set of Borel probability measures on $\mathbb{R}^d$. For a given $\mu \in \mathcal{P}(\mathbb{R}^d)$ and Borel measurable $f : \mathbb{R}^n \to \mathbb{R}$, let $\mathrm{Var}(f)$ denote the *variation* of f with respect to μ.

$$\mathrm{Var}_\mu(f) := \int_{\mathbb{R}^d} |f - \mathbb{E}_\mu(f)|^2 d\mu \tag{1}$$

A measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is said to satisfy a *Poincaré inequality* with constant $C > 0$ if

$$\mathrm{Var}_\mu(f) \leq C \int |\nabla f|^2 d\mu \tag{2}$$

for all locally Lipschitz functions $f \in \mathbb{R}^d \to \mathbb{R}$. We define C_p to be the smallest constant $C > 0$ such that (2) holds. Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, we define their *convolution* $\mu * \nu \in \mathcal{P}(\mathbb{R}^d)$ as the following measure: given a Borel set $A \subset \mathbb{R}^d$,

$$\mu * \nu(A) := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \chi_A(x + y) d\mu(x) d\nu(y)$$

where χ_A is the indicator function associated to A. One may derive from a classical variance decomposition and convexity of $t \mapsto t^2$ that

$$C_p(\mu * \nu) \leq C_p(\mu) + C_p(\nu). \tag{3}$$

The results of this paper may be viewed as refined versions of this inequality

Theorem 1. *Let $(\mu_i)_{1 \leq i \leq n} \subset \mathcal{P}(\mathbb{R}^d)$. For a set $S \in [n] := \{1, 2, ..., n\}$, let μ_S denote the convolution of $(\mu_i)_{i \in S}$. If $\mathcal{C}$ is a collection of distinct subsets of $[n]$, then*

$$C_p(\mu_{[n]}) \leq \frac{1}{t} \sum_{S \in \mathcal{C}} C_p(\mu_S), \tag{4}$$

where $t := \min_{i \in [n]} |\{S \in \mathcal{C} : i \in S\}|$.

Although it is not stated here, the authors in fact prove a more general result in the context of abelian groups. One may interpret this result in the context of a central limit theorem. Let $(\Omega, \mathbb{P}, \mu)$ be a probability space, and let $X : \Omega \to \mathbb{R}^d$ be a random variable (i.e. a measurable function on Ω). We define the *law* of X, $\mathcal{L}_X \in \mathcal{P}(\mathbb{R}^d)$, to be the pushforward measure of μ under X. Explicitly, for a given Borel set $A \subset \mathbb{R}^d$,

$$\mathcal{L}_X(A) := \mu(X^{-1}(A)).$$

We then write $X \sim \mu$ to denote that $\mu = \mathcal{L}_X(A)$.

Corollary 2. *Let (X_i) be identical independently distributed (i.i.d.) random variables with law ν_1. For $n \geq 1$, let ν_n denote the law of the standardised sum $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$. Then,*

$$C_p(\nu_n) \leq C_p(\nu_{n-1}) \tag{5}$$

Hence, the Poincaré constant is non-increasing along the central limit theorem.

Corollary 3. *Let ν_n be as in the previous corollary, and let γ_{δ^2} denote the law of the normal distribution $N(0, \delta I)$. Then,*

$$C_p(\nu_n * \gamma_{\delta^2/n}) \leq C_p(\nu_1 * \gamma_{\delta^2}). \tag{6}$$

This result is surprising in the sense that the degree of the gaussian regularisation on the left is much less than on the right, but the Poincare constant is no worse. It also implies a certain quantitative central limit theorem, where the rate of convergence to a gaussian distribution is explicitly bounded. In order to make sense of this, we first need to define a suitable metric on $\mathcal{P}(\mathbb{R}^d)$. First of all, given a joint probability measure μ over a product of measure spaces $\Omega_1 \times ... \times \Omega_n$ and a subset $S \in [n]$, we define the *marginal* μ_S over $\Omega_S := \prod_{i \in S} \Omega_i$ to be the pushforward measure of μ with respect to the natural projection $\pi_S : \Omega_1 \times ... \times \Omega_n \to \Omega_S$, i.e.

$$\mu_S(A) := \mu(\pi_S^{-1}(A)).$$

Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, define $\Gamma(\mu, \nu)$ to be the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ such that μ and ν are their marginals with respect to the natural projections onto the first and second factors respectively. We then define the L^2 *Wasserstein metric* on $\mathcal{P}(\mathbb{R}^d)$ as

$$W_2(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) \right)^{1/2}.$$

There is of course a natural generalisation to the setting of Borel metric spaces with general exponents. We say that a random variable $X : \Omega \to \mathbb{R}^d$ is *isotropic* if its covariance matrix is the identity, i.e. that

$$\mathbb{E}_\mu((X_i - \mathbb{E}_\mu(X_i))(X_j - \mathbb{E}_\mu(X_j))) = \delta_{ij}.$$

Corollary 4 (Quantitative Central Limit Theorem)**.** *Let (X_i) be i.i.d. centred isotropic random vectors in $\mathbb{R}^d$ with law ν_1. If $C_p(\nu_1 * \gamma_{\delta^2}) \le C_{\delta^2}$ for some $\delta > 0$ then*

$$W_2(\nu_n, \gamma)^2 \le d\frac{2(\delta^2 + C_{\delta^2})}{\delta^2 + \sqrt{n} - 1}, \tag{7}$$

where $\nu_n \sim \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$.

For $\mu \in \mathcal{P}(\mathbb{R}^d)$, define

$$\sigma := \max_{\alpha \in \mathbb{R}^d : |\alpha| = 1} \mathrm{Var}_\mu(x \mapsto \alpha \cdot x)$$

to be the largest variance of μ in any direction (equivalently, the largest eigenvalue of the covariance matrix for μ). It is known that, if $\alpha^* \in \mathbb{R}^d$ extremises $\mathrm{Var}_\mu(x \mapsto \alpha \cdot x)$ with respect to the constraint $|\alpha^*| = 1$, then

$$C_p(\mu) - \sigma^2(\mu) \ge W_2(\mu_{\alpha^*}, \gamma_{\sigma^2(\mu)}), \tag{8}$$

where μ_{α^*} denotes the pushforward of μ under the projection $x \mapsto \alpha^* \cdot x$. The following inequality refines (3) by bounding the difference between the right and left hand sides above in terms of $C_p(\mu * \nu) - \sigma^2(\mu)$.

Theorem 5. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, and define $\sigma^2 = \sigma^2(\mu * \nu)$ for convenience. Then,*

$$C_p(\mu * \nu) \le C_p(\mu) + C_p(\nu) - \frac{C_p(\mu)C_p(\nu)}{C_p(\mu) + C_p(\nu)} \frac{(C_p(\mu * \nu) - \sigma^2)^2}{(C_p(\mu * \nu) - \sigma^2)^2 + C_p(\mu * \nu)\sigma^2} \tag{9}$$

Iterating this theorem with a fixed μ, we obtain an $\mathcal{O}(1/n)$ bound on the rate of convergence of $C(\nu_n) \to 1$, where ν_n is as in Corollary 2.

Corollary 6. *Let $X_1, X_2, \ldots$ be i.i.d. isotropic vectors in $\mathbb{R}^d$, and define ν_n to be the law of the standardised sum $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$. Then,*

$$C_p(\nu_{2^n}) - \frac{3}{2} \le \left(\frac{3}{4}\right)^n \left(C_p(\nu_1) - \frac{3}{2}\right) \tag{10}$$

and, if $C_p(\nu_1) \le 2$, it further holds that

$$C_p(\nu_{2^n}) - 1 \le \frac{7}{n+7}. \tag{11}$$

Jennifer Duncan, University of Birmingham
email: RDD710@student.bham.ac.uk

24 Unifying the Brascamp-Lieb Inequality and the Entropy Power Inequality

After Venkat Anantharam, Varun Jog and Chandra Nair [AJN]

A summary written by Joao P. G. Ramos

Abstract

In this talk we will prove an inequality generalizing information theoretical versions of the Brascamp–Lieb Inequality (BLI) and the Entropy Power Inequality. The proof of the main result uses a version of the doubling trick. In addition, we will discuss when the quantities considered in the statement of the main result are finite, and possible open questions in this direction.

24.1 Introduction

Let $f : \mathbb{R}^n \mapsto \mathbb{R}_{\geq 0}$ be a function so that

$$\int_{\mathbb{R}^n} f(x) \log(1 + f(x))\, dx < +\infty.$$

Let X be a $\mathbb{R}^n-$valued absolutely continuous random variable, so that its density f_X satisfies the condition above. We define the *differential entropy* of X as

$$h(X) = -\int_{\mathbb{R}^n} f_X(x) \log f_X(x)\, dx.$$

As discovered by Lieb [L], a way to formulate the *Entropy Power Inequality* (EPI) is that, whenever X, Y are two independent $\mathbb{R}^n-$valued random variables and $\lambda \in (0,1)$, then

$$h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda h(X) + (1-\lambda)h(Y),$$

and equality holds if and only if X, Y are Gaussian random variables with identical covariance matrices. Analogously, one may write a version of the *Brascamp–Lieb Inequality* (BLI) in terms of random variables: let X be a $\mathbb{R}^n-$valued random variable with well-defined differential entropy, and so that $E\|X\|^2 < +\infty$. Define

$$f(X) := h(X) - \sum_{j=1}^{m} c_j h(a_j \cdot X),$$

where $c_1, ..., c_m$ are positive numbers and $a_1, ..., a_m$ are vectors which span $\mathbb{R}^n$. Then the supremum of f over the class of all such absolutely continuous random variables is the same as its supremum over all *centered Gaussian* random variables.

As previously stated, we will prove, in this talk, an inequality unifying both the EPI and the BLI. In order to state our main result, we define the class $\mathcal{P}(\mathbf{r}), \mathbf{r} = (r_1, ..., r_k)$ a set of integer indices summing to n, as the set of all $\mathbb{R}^n-$valued random variables $X := (X_1, X_2, ..., X_k)$ so that

1. X_i takes values in $\mathbb{R}^{r_i}$ and each has finite entropy;
2. $X_1, ..., X_k$ are independent;
3. $EX = 0, E\|X\|_2^2 < +\infty$.

We further define the class $\mathcal{P}_g(\mathbf{r})$ as the random variables satisfying the conditions above, with X_i being Gaussian for each $i \in \{1, ..., k\}$.

Fix a set of surjective linear mappings $A_j : \mathbb{R}^n \to \mathbb{R}^{n_j}$, positive numbers $c_1, ..., c_m$ and $d_1, ..., d_k$, and $\{r_j\}_{j=1,...,k} = \mathbf{r}$ of indices as in the previous definition. We denote such a set of parameters by $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$, and call it a *BL-EP datum.*

Theorem 1 (Unified EPI and BLI)**.** *Let* $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ *be a BL-EP datum. Let*

$$M(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d}) := \sup_{X \in \mathcal{P}(\mathbf{r})} \left(\sum_{i=1}^{k} d_i h(X_i) - \sum_{j=1}^{m} c_j h(A_j X) \right),$$

and $M_g(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ *denote the supremum over the restricted class* $X \in \mathcal{P}_g(\mathbf{r})$*. Then we have*

$$M(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d}) = M_g(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d}),$$

in the sense that, if both quantities are finite, then they are equal, and if one is infinite, so is the other.

In order to see that this result is in fact a generalization of both the Entropy Power Inequality and the Brascamp–Lieb Inequality, notice the following:

1. If $n = 2d, k = 2, r_1 = r_2 = d, d_1 = \lambda, d_2 = 1 - \lambda, c_1 = 1$, and $A_1 = [\sqrt{\lambda} I_{d\times d}, \sqrt{1-\lambda} I_{d\times d}]$, then one recovers the expression $\lambda h(X) + (1 - \lambda)h(Y) - h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y)$ in the supremum defining M and M_g. Assume theorem 1 to hold. In order to prove that this expression is

≤ 0, it suffices to prove that the associated M_g is non-positive. This follows, however, from a direct computation using that $\log\det$ is a concave function.

2. If $k = 1, r_1 = n$ and $d_1 = 1$ one recovers directly the Brascamp–Lieb case for zero-mean random variables X with finite second moment.

24.2 Proof of the main result

For the proof of Theorem 1, we follow a sketch of the so-called doubling trick. We have the following steps:

1. *Finding a concave envelope.* First, we define the function

$$s(X) = \sum_{i=1}^{k} d_i h(X_i) - \sum_{j=1}^{m} c_j h(A_j X).$$

In this step, we would like to find a suitable *concave envelope* of this function. This is achieved by defining its version conditioned on a random variable U which takes finitely many values:

$$s(X|U) = \sum_{i=1}^{k} d_i h(X_i|U) - \sum_{j=1}^{m} c_j h(A_j X|U),$$

and then letting $S(X) = \sup_U s(X|U)$. These choices must be corrected/smoothened due to technical reasons, but this is the main idea behind the construction.

2. *Lifting s and S.* The next part is defining both s and S on pairs of random variables; that is, if $(X_1, X_2) \in \mathcal{P}(\mathbf{2r})$, we define the lifting of s to $\mathcal{P}(\mathbf{2r})$ as simply

$$s(X_1, X_2) = \sum_{i=1}^{k} d_i h(X_{1i}, X_{2i}) - \sum_{j=1}^{m} c_j h(A_j X_1, A_j X_2).$$

In analogy to the previous construction, we let

$$S(X_1, X_2) = \sup_U s(X_1, X_2|U).$$

3. *Subadditivity of S.* This is the crucial, and most technical, step of the proof. We prove the following Lemma:

Lemma 2. *The function S is subadditive; that is, if $(X_1, X_2) \in \mathcal{P}(2\mathbf{r})$, then*

$$S(X_1, X_2) \leq S(X_1) + S(X_2).$$

The proof of such a result is technical, but the main novelty of this work is to use two different expansions for the differential entropy of a pair of random variables: we have

$$(A)\ h(X_1, X_2) = h(X_1) + h(X_2) - I(X_1, X_2),$$

$$(B)\ h(X_1, X_2) = h(X_1|X_2) + h(X_2|X_1) + I(X_1, X_2),$$

where $I(X_1, X_2)$ denotes the *mutual information* between X_1, X_2; i.e.,

$$I(X_1, X_2) = \int_{\mathbb{R}^n \times \mathbb{R}^n} f_{(X_1,X_2)}(x, y) \log \left(\frac{f_{(X_1,X_2)}(x, y)}{f_X(x) f_Y(y)} \right) dx\, dy.$$

The main feature of such expansions is that the information appears with opposite signs in each with respect to the other. By rearranging and using the independence conditions, and using both inequalities combined together, we may employ the positivity of the mutual information in order to conclude the inequality in Lemma 2 above. As a by-product of this proof, we have that S actually *tensorizes*: if $X_1, X_2 \in \mathcal{P}(\mathbf{r})$, and X_1 and X_2 are independent, then

$$S(X_1, X_2) = S(X_1) + S(X_2).$$

4. *Optimizers of S and s and conclusion.* Finally, in order to conclude, one shows that, truncating the sets of random variables, then $\sup_X S(X)$ is attained as $s(X^*|U^*)$ for some suitably truncated X^*, and some finite-valued random variable U^*. Following this, one considers two i.i.d. copies of (X^*, U^*) as (X_1, U_2) and (X_2, U_2), and then the random variables
$$X_+ = \frac{X_1 + X_2}{\sqrt{2}},\ X_- = \frac{X_1 - X_2}{\sqrt{2}}.$$
Let $U = (U_1, U_2)$. The new pairs (X_+, U) and (X_-, U) still optimize suitably $S(X)$, in the sense that $S(X) = s(X_\pm|U)$, and X_+ and X_- are independent when conditioned on U. But so are X_1, X_2, and thus an argument characterizing Gaussians shows that X_1 and X_2 have to be Gaussians when conditioned on U. By independence of (X_1, U_1) and (X_2, U_2), one may conclude that there is a Gaussian G^* that attains a truncated version of $\sup_X S(X)$, and thus, by the considerations above, also a suitably truncated version of $\sup_X s(X)$. This finishes the proof, by taking the truncation to infinity.

24.3 Finiteness conditions and open problems

Theorem 1 gives us an equivalence between M_g and M as defined before, but some questions remain. In particular, a first question is about when the quantities M and M_g are finite. The following result is a complete characterization of the affirmative answer to this question.

Theorem 3. *Let* $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ *be a BL-EP datum. Then* $M(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d}) < \infty$ *if and only if both following conditions are fulfilled:*

1.

$$\sum_{i=1}^{k} d_i r_i = \sum_{j=1}^{m} c_j n_j;$$

2.

$$\sum_{i=1}^{k} d_i dim(V_i) \leq \sum_{j=1}^{m} c_j dim(A_j V),$$

whenever $V = V_1 \times \cdots \times V_k$ *s a subspace with* $V_i \subset \mathbb{R}^{r_i}, i = 1, ..., k$.

In spite of this result, there are still many open questions which are related to the classical Brascamp–Lieb and the Entropy Power inequalities. We summarize them in the following open problem:

Conjecture 4. *(1) Characterize when* $M(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ *admits an extremizer, and analogously, when* $M_g(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ *admits an extremizer; (2) If extremizers exist for* $M(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$*, do they exist for* $M_g(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$*? (3) Assuming extremizers exist, characterize when they are unique, up to symmetries.*

References

[AJN] V. Anantharam, V. Jog, and C. Nair. “Unifying the Brascamp-Lieb Inequality and the Entropy Power Inequality”. 2019. arXiv: 1901.06619.

[L] E. H. Lieb. “Proof of an entropy conjecture of Wehrl". In *Inequalities*, pages 359365. Springer, 2002.

Joao P. G. Ramos, ETH Zürich
email: `joao.ramos@math.ethz.ch`