Linearized Brascamp–Lieb Inequalities

International Zurich Seminar on Information and Communication (IZS), March 6 – 8, 2024

Linearized Brascamp–Lieb Inequalities

Thomas A. Courtade
University of California, Berkeley
Department of Electrical Engineering and Computer Sciences
email: courtade@berkeley.edu

***Abstract*—Combining Valdimarsson's characterization of extremizers for the Brascamp–Lieb inequalities together with their dual entropic form, a linearization argument reveals that several well-known inequalities in probability can be viewed as consequences of the Brascamp–Lieb inequalities. The resulting "linearized Brascamp–Lieb inequalities" admit interpretation as a sharp spectral gap inequality for a simple physical process.**

I. Introduction

Fix a Euclidean space E, linear subspaces $E_i \subset E$, $i = 1, \dots, k$, a collection of linear maps $\mathbf{B} = (B_i : E \to E_i)_{i=1}^k$, and non-negative real numbers $\mathbf{c} = (c_i)_{i=1}^k \subset (0, \infty)$. The pair $(\mathbf{B}, \mathbf{c})$ is called a *(Brascamp–Lieb) datum*. The *Brascamp–Lieb inequalities* [3], [4] take the form

$$\int_E \prod_{i=1}^k (f_i \circ B_i)^{c_i} \le e^{\mathrm{BL}(\mathbf{B},\mathbf{c})} \prod_{i=1}^k \left(\int_{E_i} f_i \right)^{c_i}, \qquad (1)$$

where the *Brascamp–Lieb constant* $\mathrm{BL}(\mathbf{B}, \mathbf{c})$ is defined to be the smallest constant such that (1) holds for all non-negative $f_i \in L^1(E_i)$, $i = 1, \dots, k$. Here, the integrals are with respect to Lebesgue measure, and a theorem of Lieb [15] is that $\mathrm{BL}(\mathbf{B}, \mathbf{c})$ can be computed by considering centered Gaussian functions $(f_i)_{i=1}^k$.

For a linear subspace $V \subset E$, we let $P_V : E \to E$ denote the orthogonal projection of E onto V. A datum $(\mathbf{B}, \mathbf{c})$ is said to be *geometric* if $B_i^* B_i = P_{E_i}$ for each $i = 1, \dots, k$, and the following *frame condition* holds:

$$\sum_{i=1}^k c_i P_{E_i} = \mathrm{id}_E \,. \qquad (2)$$

When $(\mathbf{B}, \mathbf{c})$ is geometric, we have $\mathrm{BL}(\mathbf{B}, \mathbf{c}) = 0$ [1].

For a given datum $(\mathbf{B}, \mathbf{c})$, inequality (1) is said to be *extremizable* if there exist admissible $(f_i)_{i=1}^k$ such that (1) is met with equality. Modulo an equivalence relation that amounts to a linear change of variables, it is known that all extremizable data are equivalent to geometric data [1], and the extremizers in this case have been completely characterized by Valdimarsson [18].

For a Euclidean space E, let $\mathcal{M}(E)$ denote the set of Borel probability measures on E, absolutely continuous with respect to Lebesgue measure. For $\mu \in \mathcal{M}(E)$ with density $d\mu = f dx$, We define the *(Shannon) entropy*

$$h(\mu) = -\int_E f \log f dx,$$

provided the integral exists in the Lebesgue sense. Carlen and Cordero-Erausquin [8] observed the following dual formulation of the Brascamp–Lieb inequalities: For every $\mu \in \mathcal{M}(E)$ with finite entropy,

$$h(\mu) \le \sum_{i=1}^k c_i h(B_i \# \mu) + \mathrm{BL}(\mathbf{B}, \mathbf{c}), \qquad (3)$$

where # denotes the usual pushforward operation. We say that (3) is *extremizable* if there exists $\mu \in \mathcal{M}(E)$ such that (3) is an equality, and all entropies therein are finite; such a μ is called an *extremizer*. As one expects, (3) is extremizable if and only if (1) is extremizable. Hence, we can simply refer to the datum $(\mathbf{B}, \mathbf{c})$ as being extremizable without confusion.

Recall that for two probability measures $\nu, \mu \in \mathcal{M}(E)$, the *relative entropy* is defined as

$$D(\nu \| \mu) := \begin{cases} \int_E \log(\frac{d\nu}{d\mu}) d\nu & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise.} \end{cases}$$

Having recalled all of the above, we can now state a variation of the Brascamp–Lieb inequalities involving relative entropies, for reference measure equal to an extremizer of (3).

Theorem 1. *Let $(\mathbf{B}, \mathbf{c})$ be extremizable, and $\mu \in \mathcal{M}(E)$ an extremizer in* (3)*. For any $\nu \in \mathcal{M}(E)$, we have*

$$\sum_{i=1}^k c_i D(B_i \# \nu \| B_i \# \mu) \le D(\nu \| \mu). \qquad (4)$$

A linearization argument leads to the following family of variance inequalities, which is the subject of this note.

Theorem 2. *Let $(\mathbf{B}, \mathbf{c})$ be extremizable, and $\mu \in \mathcal{M}(E)$ an extremizer in* (3)*. For $X \sim \mu$ and integrable $\varphi : E \to \mathbb{R}$,*

$$\sum_{i=1}^k c_i \operatorname{Var}(\mathbb{E}[\varphi(X) | B_i X]) \le \operatorname{Var}(\varphi(X)). \qquad (5)$$

In order to apply (5) in practice, we need two things: (i) a characterization of extremizable data; and (ii) a characterization of extremal $\mu \in \mathcal{M}(E)$ in (3). The first has been already addressed, and in particular, it suffices to consider geometric data, which are concisely characterized by the frame condition (2). The second can also be addressed easily enough. In particular, Valdimarsson's characterization of extremal $(f_i)_{i=1}^k$ in (1) can be translated to a neat characterization of extremal μ in (3). To state it, let $\mu \in \mathcal{M}(E)$, and let μ_{E_i} (resp. $\mu_{E_i^\perp}$) denote the marginal of μ on E_i (resp. $E_i^\perp$). We say that μ *splits*

The author acknowledges support from NSF-CCF 1750430, the hospitality of the Laboratoire de Probabilités, Statistique et Modélisation, and the Invited Professor Program of the Fondation Sciences Mathématiques de Paris.

along $(E_i, E_i^{\perp})$ if we have the decomposition $\mu = \mu_{E_i} \otimes \mu_{E_i^{\perp}}$. In other words, μ splits along $(E_i, E_i^{\perp})$, if it is product with respect to the orthogonal decomposition $E = E_i \oplus E_i^{\perp}$.

The following can be distilled from Valdimarsson's characterization of extremal $(f_i)_{i=1}^k$ in (1), and provides a satisfactory answer to the second issue noted above.

Proposition 1. *Let* $(\mathbf{B}, \mathbf{c})$ *be geometric, and let* $\mu \in \mathcal{M}(E)$ *have finite entropy. The following are equivalent:*
1) μ *is an extremizer in* (3)*;*
2) μ *splits along* $(E_i, E_i^{\perp})$ *for each* $i = 1, \dots, k$.

We remark that Valdimarsson [18] actually leads to a more explicit characterization of extremal μ than above (roughly speaking, an extremal μ has a rigid factorization into independent components, with some factors chosen freely, and others isotropic Gaussians). However, for our purposes, the characterization in Proposition 1 suffices, and is easily stated.

We thus arrive at the following simple and explicit statement, which we call *linearized Brascamp–Lieb inequalities*.

Corollary 1 (Linearized Brascamp–Lieb inequalities)**.** *Let* $\mathbf{c}$ *and* $(E_i)_{i=1}^k$ *satisfy the frame condition* (2)*. If* X *has law that splits along* $(E_i, E_i^{\perp})$ *for each* $i = 1, \dots, k$*, then for all integrable* $\varphi : E \to \mathbb{R}$*,*

$$\sum_{i=1}^k c_i \operatorname{Var}(\mathbb{E}[\varphi(X)|P_{E_i}X]) \le \operatorname{Var}(\varphi(X)). \tag{6}$$

The remainder of this note is organized as follows. Section II illustrates a few applications of (6) to inequalities in probability. Section III explains how (6) may be interpreted as a sharp spectral gap inequality. Section IV contains the proofs, and Section V gives some brief concluding remarks.

II. Applications

It's well-known that the Brascamp–Lieb inequalities (1) contain many classical analytic and geometric inequalities (e.g., the Hölder, Young, and Loomis–Whitney inequalities), and their dual formulation (3) can be seen as generalizing the information-theoretic inequality known as subadditivity of entropy. All of these applications require only the evaluation of $\mathrm{BL}(\mathbf{B}, \mathbf{c})$, which can be accomplished in practice due to the Gaussian saturation property. By incorporating the characterization of extremizers into the picture, we obtain (4) and (5). As a consequence, we find that a variety of probabilistic inequalities may also be obtained from the Brascamp–Lieb inequalities. Toward that end, let us now demonstrate some special cases of the linearized Brascamp–Lieb inequalities.

Example 1 (Efron–Stein inequality [13], [17])**.** *Let* $X = (X_i)_{i=1}^k$ *be a random vector with independent components* $(X_i)_{i=1}^k$*, and define*

$$X^{(i)} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots X_k).$$

For any measurable φ *with* $\operatorname{Var}(\varphi(X)) < \infty$*,*

$$\operatorname{Var}(\varphi(X)) \le \sum_{i=1}^k \mathbb{E}[\operatorname{Var}(\varphi(X)|X^{(i)})]. \tag{7}$$

Proof. We can assume X takes values in E, and choose E_i such that X_i is the component of X in $E_i^{\perp}$. This implies the orthogonal decomposition $E = \oplus_{i=1}^k E_i^{\perp}$, which yields the frame condition

$$\frac{1}{k-1} \sum_{i=1}^k P_{E_i} = \mathrm{id}_E\,.$$

By the independence hypothesis, the law of X splits along $(E_i, E_i^{\perp})$ for each $i = 1, \dots, k$, and therefore (7) follows from (6) by invoking the classical variance decomposition $\operatorname{Var}(\varphi(X)) = \mathbb{E}[\operatorname{Var}(\varphi(X)|Y)] + \operatorname{Var}(\mathbb{E}[\varphi(X)|Y])$ with $Y = X^{(i)}$. ☐

More generally, the classical variance decomposition can be applied directly to (6) to obtain a generalized version of the Efron–Stein inequality. We'll return to this in our interpretation of (5) as a spectral gap inequality in Section III.

Example 2 (Dembo–Kagan–Shepp inequality [12])**.** *Let* $(X_i)_{i\ge 1}$ *be a sequence of i.i.d. random vectors, and define* $S_n = \sum_{j=1}^n X_j$*. If function* g *satisfies* $\mathbb{E}[g(S_n)] < \infty$*, then*

$$\operatorname{Var}(\mathbb{E}[g(S_n)|S_m]) \le \frac{m}{n} \operatorname{Var}(g(S_n)), \qquad n \ge m \ge 1. \tag{8}$$

Proof. For simplicity of notation, we'll assume each X_i is one-dimensional. Consider the random vector $X = (X_1, \dots, X_n)$ taking values in $E := \mathbb{R}^n$, with X_j the projection of X along natural basis vector e_j, $j = 1, \dots, n$. Take $(E_i)_{i=1}^k$ be an enumeration of all $k = \binom{n}{m}$ subspaces of E, equal to the linear span of exactly m natural basis vectors. By construction, X splits along $(E_i, E_i^{\perp})$, and the frame condition (2) holds with $c_i = \frac{n}{m}/\binom{n}{m}$ for each i. By symmetry, $\mathbb{E}[g(S_n)|P_{E_i}X]$ are equal in law for each $i = 1, \dots, k$. So, an application of (6) with $\varphi(X) = g(S_n)$ gives

$$\operatorname{Var}(\mathbb{E}[g(S_n)|X_1, \dots, X_m]) \le \frac{m}{n} \operatorname{Var}(g(S_n)), \qquad n \ge m \ge 1.$$

The claim follows since S_m is a sufficient statistic of $(X_1, \dots, X_m)$ for S_n. ☐

By L^2 duality, (5) is equivalent to the following "variance drop" inequality.

Corollary 2 (Variance Drop[1])**.** *Let the notation and assumptions of Theorem 2 prevail. For any real-valued* $\psi_i : B_i X \mapsto \psi_i(B_i X)$ *with finite variance,*

$$\operatorname{Var}\left(\sum_{i=1}^k c_i \psi_i(B_i X)\right) \le \sum_{i=1}^k c_i \operatorname{Var}\left(\psi_i(B_i X)\right). \tag{9}$$

Moreover, this is equivalent to (5)*.*

It is tempting to regard (9) as a consequence of Jensen's inequality applied to convexity of variance. To see that it is not, assume without loss of generality that $(\mathbf{B}, \mathbf{c})$ is geometric. Taking traces of the frame condition implies $\sum_{i=1}^k c_i \le 1$, with

[1]Inequality (9) can be obtained by applying (1) to functions $f_i = e^{\epsilon \psi_i} \tilde{f}_i$, for extremal $(\tilde{f}_i)_{i=1}^k$ and vanishing ϵ. However, the interpretation of $\tilde{f}_i$ as the marginal density of $B_i \# \mu$ for some meaningful μ only becomes apparent upon inspection of passage between (1) and (3) via duality.

equality only in the case where $E_i = E$ for every i. In this latter case, every X is an admissible extremizer. Hence, (9) is a strict improvement of Jensen's inequality except in degenerate cases.

Proof. To see that (5) implies (9), put $\varphi := \sum_{i=1}^k c_i \psi_i \circ B_i$. Applying Cauchy-Schwarz twice followed by (5), we have

$$\begin{aligned}
&\mathrm{Var}(\varphi(X)) \\
&= \sum_{i=1}^k c_i \,\mathrm{Cov}(\mathbb{E}[\varphi(X)|B_iX], \psi_i(B_iX)) \\
&\leq \sum_{i=1}^k c_i \,\mathrm{Var}\left(\mathbb{E}[\varphi(X)|B_iX]\right)^{1/2} \mathrm{Var}\left(\psi_i(B_iX)\right)^{1/2} \\
&\leq \left(\sum_{i=1}^k c_i \,\mathrm{Var}\left(\mathbb{E}[\varphi(X)|B_iX]\right)\right)^{1/2} \left(\sum_{i=1}^k c_i \,\mathrm{Var}\left(\psi_i(B_iX)\right)\right)^{1/2} \\
&\leq \mathrm{Var}\left(\varphi(X)\right)^{1/2} \left(\sum_{i=1}^k c_i \,\mathrm{Var}\left(\psi_i(B_iX)\right)\right)^{1/2}.
\end{aligned}$$

To see the reverse implication (9) $\Rightarrow$(5), observe that

$$\begin{aligned}
&\sum_{i=1}^k c_i \,\mathrm{Var}(\mathbb{E}[\varphi(X)|B_iX]) \\
&= \sum_{i=1}^k c_i \,\mathrm{Cov}(\varphi(X), \mathbb{E}[\varphi(X)|B_iX]) \\
&\leq \mathrm{Var}(\varphi(X))^{1/2} \,\mathrm{Var}\left(\sum_{i=1}^k c_i \mathbb{E}[\varphi(X)|B_iX]\right)^{1/2} \\
&\leq \mathrm{Var}\left(\varphi(X)\right)^{1/2} \left(\sum_{i=1}^k c_i \,\mathrm{Var}\left(\mathbb{E}[\varphi(X)|B_iX]\right)\right)^{1/2},
\end{aligned}$$

where the first inequality is Cauchy–Schwarz, and the second follows from (9) with $\psi_i(B_iX) = \mathbb{E}[\varphi(X)|B_iX]$. ☐

As a special case, we recover an inequality due to Madiman and Barron [16], which is itself a generalization of a classical result on U-statistics due to Hoeffding [14]. To state it, recall that $\mathcal{T} \subset 2^{[n]}$ is said to be an r-cover of $[n] := \{1, \ldots, n\}$ if each element of $[n]$ is contained in exactly r members of $\mathcal{T}$.

Example 3 (Madiman–Barron inequality [16])**.** *Let $X = (X_m)_{m=1}^n$ be a collection of n independent random random vectors, and let $(S_i)_{i=1}^k \subset 2^{[n]}$ be an r-cover of $[n]$. For any real-valued $\psi_i : B_iX \mapsto \psi_i(B_iX)$ with finite variance,*

$$\mathrm{Var}\left(\sum_{i=1}^k \psi_i(X_{S_i})\right) \leq r \sum_{i=1}^k \mathrm{Var}\left(\psi_i(X_{S_i})\right), \tag{10}$$

where $X_{S_i} := (X_m)_{m \in S_i}$.

Proof. Let E be the space in which the random vector $X = (X_1, \ldots, X_n)$ takes values. Consider the geometric datum with $c_i = 1/r$ and E_i equal to the subspace of E in which the coordinate X_{S_i} lives, and apply (9). ☐

We've focused this section on implications of (5), but we emphasize that the relative entropy inequalities (4) also contain useful results. To give a quick example, we note that Shearer's inequality corresponds to the case where $(\mathbf{B}, \mathbf{c})$ is geometric, and μ has suitable product structure.

Example 4 (Shearer's inequality [10])**.** *Let E admit an orthogonal decomposition $E = \oplus_{m=1}^n V_m$, and let $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ enjoy product structure with respect to this decomposition (μ_m is a probability measure on V_m, $m = 1, \ldots, n$). Fix a collection of subsets $(S_i)_{i=1}^k \subset 2^{[n]}$. If $\mathbf{c} = (c_i) \subset (0, \infty)_{i=1}^k$ satisfies $\sum_{i: S_i \ni m} c_i = 1$ for each $m = 1, \ldots, n$, then for all probability measures ν*

$$\sum_{i=1}^k c_i D(\nu_{S_i} \| \mu_{S_i}) \leq D(\nu \| \mu), \tag{11}$$

where μ_{S_i} (resp. ν_{S_i}) denotes the marginal of μ (resp. ν) on $\oplus_{m \in S_i} V_i$.

Proof. Put $E_i = \oplus_{m \in S_i} V_i$, and note that $\sum_{i: S_i \ni m} c_i = 1$ coincides with the frame condition (2). Thus, the claim follows from (4). ☐

Remark 1. *The most common statement of Shearer's inequality assumes $(S_i)_{i=1}^k$ is an r-cover of $[n]$, and has all $(c_i)_{i=1}^k$ equal to $1/r$. However, inequality* (11) *can be regarded as a simple self-strengthening obtained by iteration. A weighted version of* (10) *also appears in [16].*

Remark 2. *In the terminology of Valdimarsson [18], Shearer's inequality* (11) *corresponds to* (4) *in the special case of a geometric datum* $(\mathbf{B}, \mathbf{c})$ *with no "dependent subspace".*

The author has previously observed that the Dembo–Kagan–Shepp inequality and the Madiman–Barron inequality can be derived directly by linearizing Shearer's inequality [11], as can be the Efron–Stein inequality. Of course, each of these classical inequalities has its own direct proof by ad hoc arguments. Nevertheless, these examples are worth repeating to emphasize their interpretation as special cases of linearized Brascamp–Lieb inequalities. The following is a simple explicit example of a linearized Brascamp–Lieb inequality that is not a linearization of Shearer's inequality.

Example 5. *Let $X \sim N(0, \mathrm{id}_{\mathbb{R}^2})$, and let $(u_i)_{i=1}^3 \subset \mathbb{R}^2$ be equiangular unit vectors (i.e., $u_i^T u_i = 1$ and $u_i^T u_{i'} = \cos(2\pi/3) = -1/2$ for $i \neq i'$). For any integrable φ,*

$$\sum_{i=1}^3 \mathrm{Var}(\mathbb{E}[\varphi(X)|u_i^T X]) \leq \frac{3}{2} \mathrm{Var}(\varphi(X)).$$

III. Spectral gap interpretation

We've seen how several inequalities in probability follow as special cases of the linearized Brascamp–Lieb inequalities. Now, we turn attention to the most general statement of the linearized Brascamp–Lieb inequalities and give a simple physical interpretation, inspired by the folklore interpretation of the Efron–Stein inequality as a Poincaré (or, spectral gap) inequality. Toward this end, for a linear subspace $V \subset E$

and $x \in E$, write $x = (x_V, x_{V^\perp})$, where $x_V := P_V x$, and $x_{V^\perp} := P_{V^\perp} x = (\mathrm{id}_E - P_V)x$.

Consider an experiment where two particles of equal mass and respective velocities $x, x' \in E$ undergo an elastic collision. By conservation of energy and momentum, the particles necessarily exchange velocity components on some subspace V. That is, the post-collision velocities of the first and second particles are, respectively:

$$x_+ = (x'_V, x_{V^\perp}), \quad \text{and} \quad x'_+ = (x_V, x'_{V^\perp}).$$

Suppose we now adopt a probabilistic collision model in which the subspace V is randomly chosen from some set $\{V_1, \dots, V_k\}$, with respective probabilities $p_1, \dots, p_m$. Then, given pre-collision velocities x, x', the expected change of velocity imparted to the first particle through collision is

$$\Delta v(x, x') = \sum_{i=1}^{k} p_j P_{V_i}(x' - x).$$

If the incoming velocities x, x' undergo a common orthogonal transformation, then a natural physical constraint imposed on the model is that the expected change in velocity $\Delta v(x, x')$ should undergo the same orthogonal transformation. That is, we require Δv to satisfy $\Delta v(Ux, Ux') = U \Delta v(x, x')$ for all $x, x' \in E$ and orthogonal $U : E \to E$. Using definitions, this invariance implies that there must exist some $\lambda \in \mathbb{R}$ such that

$$\sum_{i=1}^{k} p_i P_{V_i} = \lambda \, \mathrm{id}_E \, .$$

Moreover, it is easy to check that $0 \le \lambda \le 1$, with equality only in the trivial cases where $V_i = \{0\}$ for every i (non-interacting particles), or where $V_i = E$ for every i (particles completely exchange velocities).

Now, let μ be a probability measure on E, and consider a stochastic process $(X(t); t \ge 0)$ where a particle with initial velocity $X(0)$ is placed in contact with a bath containing particles with velocities distributed i.i.d. according to μ, and collisions between our particle and particles in the bath occur at rate 1, according to a Poisson point process. Note that if a collision happens at time t, the post-collision velocity of our particle will be

$$X(t+) = (X'_{V_i}, X_{V_i^\perp}(t-)) \quad \text{with probability } p_i, \ 1 \le i \le k,$$

where $X' \sim \mu$ is independent of the pre-collision velocity $X(t-)$ of the particle of interest. Assuming the bath is in equilibrium, the background measure μ must be invariant under these dynamics, which is true if and only if it splits along each $(V_i, V_i^\perp)$, $i = 1, \dots, k$.

The linearized Brascamp–Lieb inequalities can be interpreted as a spectral gap inequality for this stochastic process. Indeed, define $E_i := V_i^\perp$ and $c_i := \frac{p_i}{1-\lambda}$, which can be checked to satisfy the frame condition (2). For $X \sim \mu$, the linearized Brascamp–Lieb inequalities can be rewritten as

$$\mathrm{Var}(\varphi(X)) \le \frac{1}{\lambda} \sum_{i=1}^{k} p_i \mathbb{E}[\mathrm{Var}(\varphi(X) | X_{V_i^\perp})],$$

by the classical variance decomposition. Thus, in general, the linearized Brascamp–Lieb inequalities coincide with the sharp Poincaré inequality for the described dynamics.

The inequality (4) can similarly be interpreted as governing convergence to equilibrium, but in the stronger sense of relative entropy. In our setting, (4) can be written as

$$\sum_{i=1}^{k} p_i D(\mu_{V_i} \otimes \nu_{V_i^\perp} \| \mu) \le (1 - \lambda) D(\nu \| \mu), \tag{12}$$

where μ_{V_i} and $\nu_{V_i^\perp}$ denote the marginals of μ and ν on V_i and $V_i^\perp$, respectively. If our particle has pre-collision velocity with law ν, then the post-collision velocity of the particle will have $\mu_{V_i} \otimes \nu_{V_i^\perp}$ with probability p_i, and therefore the law of the post-collision velocity averaged over the collision model is the mixture $\nu_+ := \sum_{i=1}^{k} \mu_{V_i} \otimes \nu_{V_i^\perp}$. By convexity of relative entropy, the above inequality implies $D(\nu_+ \| \mu) \le (1 - \lambda) D(\nu \| \mu)$, demonstrating a strict trend to equilibrium in relative entropy with each collision. Since we assume collisions occur at rate 1, if our particle has initial velocity with law ν_0 and $(\nu_t)_{t \ge 0}$ denotes the evolution of ν_0 along these dynamics, an application of Grönwall's lemma yields the exponential decay of entropy

$$D(\nu_t \| \mu) \le e^{-\lambda t} D(\nu_0 \| \mu), \quad t \ge 0.$$

Remark 3. *There seems to be no fundamental reason to limit ourselves to a discrete set of collision possibilities. For example, if $E = \mathbb{R}^n$, we could take the frame to be $\{P_{\mathrm{span}\{\sigma\}}; \sigma \in \mathbb{S}^{n-1}\}$, equipped with the uniform measure on $\mathbb{S}^{n-1}$. This would give spectral gap $\lambda = 1/n$, and the unique invariant measures are the isotropic Gaussians.*

IV. Proofs

The hard work has already been done by Bennett, Carbery, Christ and Tao [1], Valdimarsson [18], and Carlen and Cordero-Erausquin [8]. We only need to point out how the ingredients fit neatly together. We only sketch the proofs due to space constraints.

Proof of Proposition 1. Let $(\mathbf{B}, \mathbf{c})$ be geometric. As observed in [8, Theorem 2.2], inspection of the duality argument that allows passage between (1) and (3) reveals that μ is an extremizer in (3) if and only if it admits a density f satisfying

$$f = \prod_{i=1}^{k} (f_i \circ B_i)^{c_i}, \tag{13}$$

where f_i denotes the density of $B_i \# \mu$. Moreover, if (13) holds, the marginal densities $(f_i)_{i=1}^k$ will be extremizers in (1). Now, the asserted splitting property can be obtained from the splitting property in Valdimarsson's characterization of extremizers for (1) in geometric settings [18]. ☐

With the identity (13) already noted, Theorem 1 follows easily.

Proof of Theorem 1. All extremizable data are equivalent to geometric data by a linear change of variables. Hence, by the data processing property of relative entropy, we may assume $(\mathbf{B}, \mathbf{c})$ is geometric without any loss of generality.

To prove (4), it clearly suffices to assume $D(\nu\|\mu) < \infty$, since otherwise the claim is trivial; note that this implies $\nu \ll \mu$, and also $D(B_i\#\nu\|B_i\#\mu) < \infty$ for each i by the data processing inequality. Now, let $d\mu = f dx$, and write

$$\begin{aligned} D(\nu\|\mu) &= -h(\nu) + \int \log f d\nu \\ &\geq -\sum_{i=1}^{k} c_i \left(h(B_i\#\nu) + \int \log(f_i \circ B_i) d\nu \right) \\ &= -\sum_{i=1}^{k} c_i \left(h(B_i\#\nu) + \int \log(f_i) d(B_i\#\nu) \right) \\ &= \sum_{i=1}^{k} c_i D(B_i\#\nu\|B_i\#\mu), \end{aligned}$$

where the first and last lines are definitions, the inequality follows from (3) and (13), and the penultimate line follows from the definition of pushforward. □

The standard program for deriving a spectral gap inequality from an entropy inequality is to linearize it to reveal the local behavior (see, e.g., [9]). Toward that end, recall that the relative entropy of $P \ll Q$ can be written as $D(P\|Q) = \int \frac{dP}{dQ} \log\left(\frac{dP}{dQ}\right) dQ$. Therefore, if P is a perturbation of Q in the sense that $dP = (1+\epsilon\varphi)dQ$ for a bounded function φ and ϵ sufficiently small, then Taylor expansion of $x \in \mathbb{R}^+ \mapsto x \log x$ about $x = 1$ gives the local behavior of relative entropy

$$D(P\|Q) = \frac{\epsilon^2}{2} \operatorname{Var}_Q(\varphi) + o(\epsilon^2),$$

where the first-order term is absent since φ necessarily satisfies $\int \varphi dQ = 0$ for P to be a probability measure.

Proof of Theorem 2. It suffices to assume φ is bounded, since the general statement follows by localization. Thus, let $X \sim \mu$ be an extremizer in (3), assume $\int \varphi d\mu = 0$ and define $d\mu_\epsilon := (1 + \epsilon\varphi)d\mu$, which is a valid probability measure for all ϵ sufficiently small. Definitions imply

$$d(B_i\#\mu_\epsilon) = (1 + \epsilon\mathbb{E}[\varphi(X)|B_iX])d(B_i\#\mu),$$

where $\mathbb{E}[\varphi(X)|B_iX]$ is the conditional expectation of $\varphi(X)$ with respect to the σ-algebra generated by B_iX. Thus, by linearization and Theorem 1, we have

$$\begin{aligned} &\frac{\epsilon^2}{2} \sum_{i=1}^{k} c_i \operatorname{Var}(\mathbb{E}[\varphi(X)|B_iX]) + o(\epsilon^2) \\ &= \sum_{i=1}^{k} c_i D\left(B_i\#\mu_\epsilon \middle\| B_i\#\mu\right) \\ &\leq D(\mu_\epsilon\|\mu) = \frac{\epsilon^2}{2} \operatorname{Var}(\varphi(X)) + o(\epsilon^2). \end{aligned}$$

Dividing through by ϵ^2 and letting $\epsilon \downarrow 0$ completes the proof. □

In view of Proposition 1 and Theorem 2, Corollary 1 holds whenever μ is absolutely continuous with respect to Lebesgue measure and has finite entropy. It is straightforward to extend the statement to the case when either (or both) of these qualifications do not hold.

V. Closing Remarks

The duality between functional Brascamp–Lieb inequalities and their entropic form continues to hold in abstract settings. The transference principle of extremizers introduced here to obtain inequalities of the type (4) continues to apply. This suggests many interesting questions. For example, do the Shearer-type inequalities for non-product measures in [2], [5]–[7] fit into the context of Brascamp–Lieb-type inequalities on suitable spaces, as happens with $\mathbb{S}^n$ [9]? Does an approximate form of (4) hold when μ is a near-extremizer in a quantitative sense? Answers could lead to a systematic development of spectral gap inequalities for interesting classes of processes.

Acknowledgement: The author thanks Efe Aras, Pietro Caputo, and Max Fathi for stimulating conversations.

References

[1] J. Bennett, A. Carbery, M. Christ, and T. Tao. The Brascamp-Lieb inequalities: finiteness, structure and extremals. *Geometric and Functional Analysis*, 17(5):1343–1415, 2008.

[2] A. Blanca, P. Caputo, Z. Chen, D. Parisi, D. Štefankovič, E Vigoda. On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization. *Proc. of 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 3670-3692.

[3] H. J. Brascamp, E. H. Lieb, and J. M. Luttinger. A general rearrangement inequality for multiple integrals. *Journal of functional analysis*, 17(2):227–237, 1974.

[4] H. J. Brascamp and E. H. Lieb. Best constants in Young's inequality, its converse, and its generalization to more than three functions. *Advances in Mathematics*, 20(2):151–173, 1976.

[5] A. Bristiel and P. Caputo. Entropy inequalities for random walks and permutations. *arXiv preprint* arXiv:2109.06009 (2021).

[6] P. Caputo and D. Parisi. Block factorization of the relative entropy via spatial mixing. *Comm. in Mathematical Physics*, 388.2 (2021): 793-818.

[7] P. Caputo and A. Sinclair. Entropy production in nonlinear recombination models. *Bernoulli* 24(4B): 3246-3282, Nov. 2018.

[8] E. A. Carlen and D. Cordero-Erausquin. Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities. *Geometric and Functional Analysis*, 19(2):373–405, 2009.

[9] E. A. Carlen, E. H. Lieb, and M. Loss. A sharp analog of Young's inequality on S^N and related entropy inequalities. *The Journal of Geometric Analysis*, 14(3):487–520, 2004.

[10] F. R. Chung, R. L. Graham, P. Frankl, and J. B. Shearer. Some intersection theorems for ordered sets and graphs. *Journal of Combinatorial Theory, Series A*, 43(1):23–37, 1986

[11] T. A. Courtade. Bounds on the Poincaré constant for convolution measures. *Ann. Inst. H. Poincaré Probab. Statist.* 56 (1):566-579, 2020.

[12] A. Dembo, A. Kagan, and L. A. Shepp. Remarks on the maximum correlation coefficient. *Bernoulli*, pages 343–350, 2001.

[13] B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics* 9:586-596, 1981.

[14] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The annals of mathematical statistics*, pages 293–325, 1948

[15] E. H. Lieb. Gaussian kernels have only Gaussian maximizers. *Inventiones mathematicae*, 102(1):179–208, 1990.

[16] M. Madiman and A. Barron. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Transactions on Information Theory*, 53(7):2317–2329, 2007.

[17] M. J. Steele, J. Michael. An Efron–Stein inequality for nonsymmetric statistics. *The Annals of Statistics* 14, no. 2: 753-758, 1986.

[18] S. I. Valdimarsson. Optimisers for the Brascamp-Lieb inequality. *Israel J. Math.*, 168:253–274, 2008.