Barthe, F. *et al.* (2011) "Correlation and Brascamp–Lieb Inequalities for Markov Semigroups," International Mathematics Research Notices, Vol. 2011, No. 10, pp. 2177–2216
Advance Access publication June 9, 2010
doi:10.1093/imrn/rnq114

Correlation and Brascamp–Lieb Inequalities for Markov Semigroups

Franck Barthe[1], Dario Cordero-Erausquin[2], Michel Ledoux[1], and Bernard Maurey[3]

[1]Institut de Mathématiques de Toulouse (CNRS UMR 5219), Université Paul Sabatier, 31062 Toulouse Cedex 9, France, [2]Institut de Mathématiques de Jussieu (CNRS UMR 7586), Équipe d'Analyse Fonctionnelle, Université Pierre et Marie Curie, 4, place Jussieu, 75252 Paris Cedex 05, France, and [3]Laboratoire d'Analyse et de Mathématiques Appliquées (CNRS UMR 8050), Université de Marne-la-Vallée, 77454 Marne-la-Vallée Cedex 2, France

Correspondence to be sent to: cordero@math.jussieu.fr

This paper builds upon several recent works, where semigroup proofs of Brascamp–Lieb inequalities are provided in various settings (Euclidean space, spheres, and symmetric groups). Our aim is two-fold. Firstly, we provide a general, unifying, framework based on Markov generators, in order to cover a variety of examples of interest going beyond previous investigations. Secondly, we put forward the combinatorial reasons for which unexpected exponents occur in these inequalities. Related superadditivity of information and entropy inequalities are also studied.

1 Introduction

A celebrated inequality of Brascamp and Lieb [8, 20] asserts that given linear surjective maps between Euclidean spaces $B_i : H \to H_i$, $i = 1, \ldots, m$, and given positive coefficients

Received July 16, 2009; Revised May 10, 2010; Accepted May 12, 2010

$(c_i)_{i=1}^m$, the best constant C such that for all non-negative measurable functions $f_i : H_i \to \mathbb{R}$ it holds

$$\int_H \prod_{i=1}^m f_i(B_i x)^{c_i} \mathrm{d}x \le C \prod_{i=1}^m \left(\int_{H_i} f_i(y) \, \mathrm{d}y \right)^{c_i}$$

can be computed by requiring the inequality on centered Gaussian functions only (i.e., of the form $f_i = e^{-Q_i}$ where Q_i is a positive definite quadratic form). The result was first established in [8] for one-dimensional spaces H_i, later extended in [20] to the multi-dimensional case. This far-reaching extension of Hölder's inequality found applications in harmonic analysis but also in convex geometry. Indeed, a particular case called the geometric Brascamp–Lieb inequality, put forward by Ball [2] when $\dim(H_i) = 1$, leads to many precise volume estimates. The general *geometric* version corresponds to the case when for all $i = 1, \ldots, m$, $B_i B_i^* = \mathrm{Id}_{H_i}$ and $\sum_i c_i B_i^* B_i = \mathrm{Id}_H$, where B_i^* is the adjoint of B_i. Under these hypotheses, the optimal constant in the Brascamp–Lieb inequality is $C = 1$. More concretely: let $E_1, \ldots, E_m$ be vector subspaces of $\mathbb{R}^n$ with its canonical Euclidean structure. Denoting by P_{E_i} the orthogonal projection on to E_i, if $\sum_i c_i P_{E_i} = \mathrm{Id}_{\mathbb{R}^n}$ then for all measurable functions $f_i : \mathbb{R}_i \to \mathbb{R}^+$ it holds

$$\int_{\mathbb{R}^n} \prod_{i=1}^m f_i(P_{E_i} x)^{c_i} \mathrm{d}x \le \prod_{i=1}^m \left(\int_{E_i} f_i \right)^{c_i}.$$

There exist by now many different proofs of the Brascamp–Lieb theorem: symmetrization when $\dim(H_i) = 1$ [8], study of Gaussian kernels [20], and optimal transport [3]. Heat flow derivation was presented in the recent works [11] for $\dim(H_i) = 1$ and [7] in general: the geometric Brascamp–Lieb inequality is established by interpolating between the left- and right-hand sides of the inequality, thanks to the Heat semigroup. As developed in these works (see Remark 3), and central to the approach, the case when optimal Gaussian functions exist follows from the geometric case by a clever change of variables and turns out to be generic (the non-trivial remaining cases are in a sense "boundary" cases and can be decomposed into simpler ones). So the geometric case is also essential from a theoretical viewpoint. The Heat flow proofs required a more precise study of the structure of the problem, since the finiteness of the constant and the existence of Gaussian maximizers have to be treated beforehand. They lead to a complete treatment of the equality cases [7, 11, 22]. They were also flexible enough to adapt to other ambient spaces, as observed by Carlen, Lieb, and Loss [11] who discovered the following Young type inequality on the Euclidean sphere $\mathbb{S}^{n-1}$: for all measurable functions $f_i : [-1, 1] \to \mathbb{R}^+$, it holds

$$\int_{\mathbb{S}^{n-1}} \prod_{i=1}^{n} f_i(x_i)\,\mathrm{d}\sigma(x) \leq \prod_{i=1}^{m} \left(\int_{\mathbb{S}^{n-1}} f_i(x_i)^2 \mathrm{d}\sigma(x) \right)^{\frac{1}{2}}, \tag{1}$$

where σ is the uniform probability measure on $\mathbb{S}^{n-1}$. This inequality can be understood as a correlation inequality: the coordinates of a uniform random vector on the sphere are not independent, so there is no Fubini equality. Instead, inequality (1) holds and is a lot better than Hölder's inequality, which would involve L^n norms of the functions. In a sense, the exponent 2, which turns out to be optimal, shows that the coordinate functions are not too far from being independent. The above inequality was extended to a spherical version of the geometric Brascamp–Lieb inequality in [5]. Carlen, Lieb, and Loss also proved a similar inequality for the set of permutations of a finite set and coordinate functions [12].

In this paper, we provide a general framework based on Markov generators that allows us to unify the existing results, derive extensions, and clarify the conditions that are required to prove correlation inequalities. Decompositions of the identity as (10) play an important role. In the case of functions depending on blocks of coordinates, we put forward a general set of conditions, which is similar to the hypotheses of Finner's theorem for product probability spaces [17], but applies to particular non-product spaces. See, for example, Propositions 11, 21, and Section 4.2.4.

The structure of the exposition is as follows. The abstract framework is described in Section 2 where a general condition is stated. The next sections provide concrete illustrations of Proposition 2. Section 3 deals with the case where our Markov generator is a diffusion, as it is the case in some classical geometric and probabilistic situations. In particular, we shall put forward the algebraic content of our condition in the case of Riemannian Lie groups (with emphasis on the orthogonal group $SO(n)$) and their quotients. We study discrete models and their combinatorics in Section 4, and the case where the generator is a sum of squares in Section 5. The final section is devoted to related entropy inequalities for the marginals of a probability distribution. Since these inequalities are dual (and equivalent) to the Brascamp–Lieb (BL) inequalities, it gives a different way of obtaining the above-mentioned inequalities. The entropic inequalities are consequences of superadditive inequalities for the associated Fisher information that are directly derived from the general condition for the Markov generator, in both continuous and discrete situations.

2 The Abstract Argument: Commuting Maps and BL-condition

The basic input is a measurable space E and a Markov semigroup $(P_t)_{t\geq 0}$ acting on functions on E, with generator L. We do not discuss here the various questions related to the underlying domain of L and its associated carré du champ operator (see below) as well as the classes of functions under consideration. When a given inequality on functions is stated, it is always understood relatively to the suitable domains of $(P_t)_{t\geq 0}$, L or Γ. These are clear in all the continuous or discrete illustrations in this work. We refer to [1] for an introduction and further details in this respect and to [15] for the discrete setting.

The general framework of our study is the following. We introduce $m \geq 1$ measurable spaces E_i and maps $T_i : E \to E_i$, $i = 1, \dots, m$. We assume that, for each $i = 1, \dots, m$, the map T_i *commutes* with P_t or L in the sense that for every $g : E_i \to \mathbb{R}$, $L(g \circ T_i)$ factors through T_i:

$$L(g \circ T_i) = \tilde{g} \circ T_i \tag{2}$$

for some $\tilde{g} : E_i \to \mathbb{R}$. In other words, L (or P_t) leaves invariant the algebra of functions on E of the form $g \circ T_i$. This means that P_t or L may be projected on E_i and there exists a Markov generator L_i on E_i such that

$$L(g \circ T_i) = (L_i g) \circ T_i.$$

We denote below by $(P_t^i)_{t\geq 0}$ the semigroup with generator L_i. It follows that $P_t(g \circ T_i) = (P_t^i g) \circ T_i$.

We aim at understanding how the “geometry” or the “combinatorics” of the T_i’s and the choice of constants $c_i > 0$ ensure that

$$P_t\left(\prod_{i=1}^m f_i^{c_i} \circ T_i\right) \leq \prod_{i=1}^m \big(P_t(f_i \circ T_i)\big)^{c_i}$$

for all $f_i : E_i \to \mathbb{R}^+$, $i = 1, \dots, m$. Since $(P_t(F^{1/c}))^c \leq (P_t(F^{1/d}))^d$ for $c \geq d > 0$, we would like to pick the largest possible constants c_i’s. Also, for obvious reasons (pick all the f_i

but one to be identically 1), the c_i's will belong to $(0, 1]$ and the inequalities we consider can be rewritten in terms of L^{p_i} norms for $p_i = 1/c_i$.

This problem is of course reminiscent of the Brascamp–Lieb convolution inequalities described in the introduction, and it can as well be interpreted as a correlation problem. This correlation problem has many ramifications, as we shall see.

We will, in this general framework, be dealing with inequalities that are valid for the measures $P_t(.)(x)$, uniformly on the point x.

Definition 1 (The BL-condition). Let $(P_t)_{t\geq 0}$ be a Markov semigroup on E with generator L. Let c_i be non-negative reals and $T_i : E \to E_i$ maps commuting with L, for $i = 1, \dots, m$. We say that $\{c_i, T_i\}$ satisfy the **BL**-*condition* if: for all functions $F_i : E \to \mathbb{R}$, $i = 1, \dots, m$, of the form $F_i = g_i \circ T_i$, setting $H = \sum_{i=1}^m c_i F_i$, it holds

$$e^{-H} L(e^H) \leq \sum_{i=1}^m c_i\, e^{-F_i} L(e^{F_i}). \tag{3}$$

□

This definition is motivated by following the main equivalence that is implicit in [11] and [4].

Proposition 2. With the notation of the previous definition, the following statements are equivalent:

- For all non-negative functions $f_i : E_i \to \mathbb{R}$, $i = 1, \dots, m$, and every $t \geq 0$,

$$P_t\Big(\prod_{i=1}^m f_i^{c_i} \circ T_i\Big) \leq \prod_{i=1}^m \big(P_t(f_i \circ T_i)\big)^{c_i}. \tag{4}$$

- The $\{c_i, T_i\}$ satisfy the **BL**-condition. □

Proof. Let $f_i : E_i \to \mathbb{R}$, $i = 1, \dots, m$, be bounded positive functions. Let $t \geq 0$ and consider

$$\alpha(s) = P_s\Big(\exp\Big(\sum_{i=1}^m c_i \log P_{t-s}(f_i \circ T_i)\Big)\Big), \quad 0 \leq s \leq t.$$

Set $F_i = \log P_{t-s}(f_i \circ T_i)$, $i = 1, \dots, m$, and $H = \sum_{i=1}^m c_i F_i$. Direct calculations give

$$\alpha'(s) = P_s\Big(L(e^H) - e^H \sum_{i=1}^m c_i\, e^{-F_i} L(e^{F_i})\Big).$$

Next, by the commutation property (2), $F_i = \log P_{t-s}(f_i \circ T_i)$ is a function of T_i so that, under (3), $\alpha'(s) \le 0$ and thus $\alpha(0) \ge \alpha(t)$. Hence, (4) follows from (3). The converse implication is obtained by differentiating (4) at $t = 0$. ■

Remark 3. Given maps $T_i : E \to E_i$, $i = 1, \dots, m$, one may not always be able to check the **BL**-condition (3). It might be necessary to consider further bijective maps $R : E \to E$, $R_i : E_i \to E_i$ and to deal with $\tilde{T}_i = R_i \circ T_i \circ R : E \to E_i$, $i = 1, \dots, m$, (still assumed to commute with P_t) instead of T_i. This is exemplified by the paper [7] where the Gaussian-extremizable cases of the Euclidean Brascamp–Lieb inequality are reduced to the geometric Brascamp–Lieb inequality. Actually, this change of variables is also implicit in [11] where the functions f_i are evolving according to different semigroups. When, among centered Gaussian functions of integral one, the functional $\int \prod (f_i \circ B_i)^{c_i}$ admits a maximizer, differentiating around this maximum yields an equality between linear symmetric maps, which can be used to change variables and reduce to the geometric situation. □

It is usually of more interest to state Brascamp–Lieb type inequalities with respect to the invariant measure μ of the semigroup $(P_t)_{t\ge 0}$. When $(P_t)_{t\ge 0}$ is ergodic with invariant probability measure μ, we may let $t \to \infty$ in the local inequality (4) and get inequalities of the type

$$\int \prod_{i=1}^m f_i^{c_i} \circ T_i \,\mathrm{d}\mu \le \prod_{i=1}^m \Big(\int f_i \circ T_i \,\mathrm{d}\mu\Big)^{c_i}. \tag{5}$$

Actually, this can be viewed directly by studying $\beta(t) = \int \prod_i P_t(f_i \circ T_i)^{c_i}\,\mathrm{d}\mu$. Indeed with the notation in the above proof

$$\beta'(t) = \int e^H \Big(\sum_{i=1}^m c_i\, e^{-F_i} L(e^{F_i})\Big)\,\mathrm{d}\mu = -\int \Big(L(e^H) - e^H \sum_{i=1}^m c_i\, e^{-F_i} L(e^{F_i})\Big)\mathrm{d}\mu.$$

Hence integrating from 0 to ∞, the **BL**-condition (3) yields (5). Note that the condition $\beta'(t) \geq 0$ may be rewritten in terms of the Dirichlet form $\mathcal{E}(f,g) := \int f(-Lg)\mathrm{d}\mu$ as

$$\sum_{i=1}^{m} c_i\, \mathcal{E}\big(e^{H-F_i}, e^{F_i}\big) \leq 0.$$

Remark 4. If $(P_t)_{t\geq 0}$ has an infinite invariant measure μ, more hypotheses are needed to get a meaningful limit to the local bounds as $t \to \infty$. Assume that $(P_t)_{t\geq 0}$ is of dimension n, and size $\kappa > 0$, in the sense that for every μ-integrable function $f : E \to \mathbb{R}$, at any point,

$$\lim_{t\to\infty} t^{n/2} P_t f = \kappa \int f \mathrm{d}\mu.$$

If the semigroups $(P_t^i)_{t\geq 0}$ have invariant measures μ_i, dimensions n_i, and sizes κ_i, $i = 1, \ldots, m$, and if in addition $\sum_{i=1}^m c_i n_i = n$, we may use $P_t(f_i \circ T_i) = P_t^i(f_i) \circ T_i$ and let $t \to \infty$ in (4) to get

$$\int \prod_{i=1}^{m} f_i^{c_i} \circ T_i \,\mathrm{d}\mu \leq \kappa^{-1} \prod_{i=1}^{m} \left(\kappa_i \int f_i \,\mathrm{d}\mu_i\right)^{c_i}.$$

□

3 Examples of Diffusion Semigroups

This section is devoted to several examples of illustration of the preceding abstract scheme in case the generator L satisfies a chain rule formula. Recall that the *carré du champ* of the generator L is defined on some suitable algebra of functions by

$$\Gamma(f,g) = \frac{1}{2}(L(fg) - fLg - gLf). \tag{6}$$

For simplicity, one writes $\Gamma(f)$ for $\Gamma(f,f)$. If L is a diffusion generator (i.e., a linear differential operator of order 2 without constant term), then the chain rule yields $L(e^f) = e^f\big(Lf + \Gamma(f)\big)$. So for $H = \sum_{i=1}^m c_i F_i$,

$$e^{-H} L(e^H) - \sum_{i=1}^{m} c_i\, e^{-F_i} L(e^{F_i}) = \Gamma(H) - \sum_{i=1}^{m} c_i \Gamma(F_i).$$

Hence, we have:

Fact 5 (BL-condition in the diffusion case). If L is a diffusion operator, then the **BL**-condition (3) is equivalent to saying that for every function $f_i : E_i \to \mathbb{R}$, $i = 1, \dots, m$,

$$\Gamma\left(\sum_{i=1}^m c_i\, f_i \circ T_i\right) = \sum_{i,j=1}^m c_i c_j\, \Gamma(f_i \circ T_i,\, f_j \circ T_j) \le \sum_{i=1}^m c_i\, \Gamma(f_i \circ T_i). \tag{7}$$

□

Depending on the structure, this condition may be expressed more intrinsically in terms of the operators T_i. We investigate several instances below.

3.1 Riemannian manifolds

Let us assume that E is a Riemannian manifold and that $\Gamma(f) = |\nabla f|^2$. This is in particular the case if P_t is the Heat equation on E associated to the Riemannian Laplacian Δ. We also assume that the maps T_i are differentiable. Then Condition (7) amounts to the fact that for every $x \in E$, and for all smooth functions f_i,

$$\left|\sum_{i=1}^m c_i \nabla(f_i \circ T_i)(x)\right|^2 \le \sum_{i=1}^m c_i \left|\nabla(f_i \circ T_i)(x)\right|^2. \tag{8}$$

For each $x \in E$, we introduce the subspace of $T_x E$, the tangent space at x,

$$\mathcal{E}_i(x) := \left\{\nabla(f_i \circ T_i)(x);\ f_i : E_i \to \mathbb{R}\right\} \subset T_x E. \tag{9}$$

This is the orthogonal complement of the kernel of $DT_i(x)$, so it is orthogonal to the tangent directions of the level set $\{y \in E;\ T_i(y) = T_i(x)\}$. We denote by $P_{\mathcal{E}_i(x)}$ the orthogonal projection on $\mathcal{E}_i(x)$ in the Euclidean space $T_x E$. We can reformulate (8) using the following well-known equivalence, which relies on the fact that a linear map and its adjoint have the same norm: for $\mathcal{E}$ a Euclidean space, $\mathcal{E}_i$, $i = 1, \dots, m$, Euclidean subspaces of $\mathcal{E}$ and $c_1, \dots, c_m > 0$ we have:

$$\forall v_i \in \mathcal{E}_i, \quad \left|\sum_{i=1}^m c_i\, v_i\right|^2 \le \sum_{i=1}^m c_i\, |v_i|^2 \quad \Longleftrightarrow \quad \forall v \in \mathcal{E}, \quad \sum_{i=1}^m c_i \left|P_{\mathcal{E}_i} v\right|^2 \le |v|^2$$

writing $P_{\mathcal{E}_i}$ for the orthogonal projection on to $\mathcal{E}_i$. More concisely, denoting the identity map by $\mathrm{Id}_{\mathcal{E}}$, the latter condition rewrites as an inequality between symmetric maps: $\sum_{i=1}^m c_i P_{\mathcal{E}_i} \le \mathrm{Id}_{\mathcal{E}}$.

Therefore, we see that **BL**-condition amounts here to a "moving decomposition of the identity" inequality in all tangent spaces.

Fact 6 (BL-condition in the Riemannian case). In the setting described above, the **BL**-condition (3) is equivalent to saying that for all $x \in E$,

$$\sum_{i=1}^{m} c_i P_{\mathcal{E}_i(x)} \leq \mathrm{Id}_{T_x E}. \tag{10}$$

□

Next, we present instances of such decompositions in the case of model spaces.

Geometric Brascamp–Lieb inequality in Euclidean space. In $\mathbb{R}^n$, let, for $i = 1, \ldots, m$, E_i be vector subspaces of dimension $n_i \geq 1$ and let $c_i \geq 0$, such that

$$\sum_{i=1}^{m} c_i P_{E_i} = \mathrm{Id}_{\mathbb{R}^n}.$$

We take of course $T_i : \mathbb{R}^n \to E_i$ such that $T_i(x) = P_{E_i} x$, $x \in \mathbb{R}^n$, $i = 1, \ldots, m$.

If B is a linear map, $\nabla(f \circ B)(x) = {}^t B \nabla f(Bx)$ and $\Delta(f \circ B)(x) = \mathrm{Tr}\big({}^t B \mathrm{Hess}\, f(Bx) B\big)$. It is then clear that the generator $L = \Delta - x \cdot \nabla$ of the Ornstein–Uhlenbeck semigroup commutes with the T_i's. Also for all $x \in \mathbb{R}^n$, the spaces $\mathcal{E}_i(x)$ are simply E_i. Hence, (10) is guaranteed by the decomposition of the identity induced by the E_i's. Thus, we get a Brascamp–Lieb inequality for the standard Gaussian measure, which is ergodic for the Ornstein–Uhlenbeck semigroup:

$$\begin{aligned}\int_{\mathbb{R}^n} \prod_{i=1}^{m} f_i(P_{E_i} x)^{c_i} e^{-|x|^2/2} \frac{\mathrm{d}x}{(2\pi)^{n/2}} &\leq \prod_{i=1}^{m} \left(\int_{\mathbb{R}^n} f_i(P_{E_i} x) e^{-|x|^2/2} \frac{\mathrm{d}x}{(2\pi)^{n/2}} \right)^{c_i} \\ &= \prod_{i=1}^{m} \left(\int_{E_i} f_i(y) e^{-|y|^2/2} \frac{\mathrm{d}y}{(2\pi)^{n_i/2}} \right)^{c_i}.\end{aligned}$$

The Brascamp–Lieb inequality with a Gaussian measure was already mentioned in [8, 20]. Note that the decomposition of identity rewrites as $\sum_{i=1}^{m} c_i |P_{E_i} x|^2 = |x|^2$. Hence setting $g_i(y) = f_i(y) \exp(-|y|^2/2)$ and using the condition $n = \sum c_i n_i$ (take traces in the decomposition of the identity), we obtain the Euclidean inequality

$$\int_{\mathbb{R}^n} \prod_{i=1}^{m} g_i(P_{E_i} x)^{c_i} \mathrm{d}x \leq \prod_{i=1}^{m} \left(\int_{E_i} g_i(y)\, \mathrm{d}y \right)^{c_i}.$$

Alternatively, we could have used the Heat semigroup (with generator Δ) to get a local inequality and pass to the limit using the dimension of this semigroup, as explained in Remark 4.

Further investigations on the connections between decompositions of the identity of $\mathbb{R}^n$ and functional inequalities (such as Young's convolution inequality, Shannon's inequality, and hypercontractivity of the Ornstein–Uhlenbeck semigroup) can be found in [14].

Geometric Brascamp–Lieb inequality on the sphere. The first inequality of this type was established by Carlen, Lieb, and Loss [11] for coordinate functions on the sphere. It involves an unexpected exponent 2. A natural extension in the spirit of the latter Euclidean inequality was given in [5]. It reads as follows: if $x \in \mathbb{S}^{n-1} \subset \mathbb{R}^n$ (the standard $(n-1)$-sphere), set as before $T_i(x) = P_{E_i}(x)$, $i = 1, \ldots, m$, where $E_i \subset \mathbb{R}^n$ are subspaces for which we have

$$\sum_{i=1}^{m} c_i P_{E_i} \leq \mathrm{Id}_{\mathbb{R}^n}.$$

Then, whenever f_i are non-negative measurable functions on the sphere, such that f_i depends only on E_i (i.e., $f_i(x) = g_i(P_{E_i}(x))$, for the uniform probability measure σ on $\mathbb{S}^{n-1}$ we have,

$$\int_{\mathbb{S}^{n-1}} \prod_{i=1}^{m} f_i^{c_i/2} \mathrm{d}\sigma \leq \prod_{i=1}^{m} \left(\int_{\mathbb{S}^{n-1}} f_i \,\mathrm{d}\sigma \right)^{c_i/2}.$$

It is easy to see that the Laplacian on $\mathbb{S}^{n-1}$ commutes with the operators T_i. The strategy in [5] is to derive decompositions of the identity in all tangent hyperplanes to the sphere, thus fulfilling Condition (10). Another approach based on analysis on the orthogonal group will be given next.

Hyperbolic space. It is natural to ask for an hyperbolic analogue of the previous statement. Let us explain, in two dimensions, why the method does not give any interesting correlation inequality. The natural functionals T_i to consider are the Busemann functions (which basically are the coordinates in the direction of a point at infinity), they commute with the Laplace operator. In the disk model, choose $b_1, \ldots, b_m$ on the unit circle and let T_i be the corresponding Busemann functions. At a point x in the disk, the directions $\mathcal{E}_i(x)$ are simply the lines spanned by the gradients of the T_i's (the tangent to the geodesic passing through x and going to b_i). When x tends to a point at infinity b, which is not one of the b_i's, it is clear that the lines $\mathcal{E}_i(x)$ become asymptotically parallel to the line $\mathbb{R}b$. Hence if a decomposition of the identity exists in all tangent planes, we get that $\sum c_i \leq 1$. But in this case, the decomposition (10) is trivial since $P_{\mathcal{E}_i(x)} \leq \mathrm{Id}$, and the inequality that we get is nothing else than Hölder's inequality.

3.2 Riemannian Lie groups

In the case of Lie groups (and their quotients), the geometric structure required to have Brascamp–Lieb type inequalities is very clear and elegant.

The algebraic structure of the problem appears clearly when functions depending only on some variables are seen as functions invariant under the (right) action of subgroups of isometries. For instance, a function $f(x)$ on $\mathbb{R}^n$ is a function of x_1 if and only if f is invariant under all translation leaving $e_1 = (1, 0, \dots, 0)$ invariant. Note also that a function $f(x)$ on the sphere $S^{n-1} \subset \mathbb{R}^n$ is a function of x_1 if and only if f is invariant under all rotations leaving e_1 invariant. In this section, we shall extensively use this point of view in the case of compact Riemannian Lie group.

Let G be a connected compact Riemannian Lie group with unit element denoted by e. Let $\mathcal{G} = T_e M$ be the associated Lie algebra; by assumption, $\mathcal{G}$ is a Euclidean space. Let μ be the normalized bi-invariant Haar measure on G. Here we will work with the Laplace–Beltrami operator Δ as Markov generator, for which we indeed have that

$$\Gamma(f) = |\nabla f|^2,$$

as required in the previous section.

Let G_i be a connected Lie subgroup of G, with Lie algebra $\mathcal{G}_i \subset \mathcal{G}$. A function $f : G \to \mathbb{R}$ is said to be G_i-right invariant if

$$f(xg) = f(x), \quad \forall g \in G_i, \quad \forall x \in G.$$

Equivalently, f is of the form $g \circ T_i$ where $T_i : G \to G/G_i$ is the canonical projection on to the right quotient, defined by $T_i(x) = xG_i$. In other words, using notation (9), we are interested in the case where, for $x \in G$,

$$\mathcal{E}_i(x) = \{\nabla f(x); \quad f : G \to \mathbb{R} \text{ is } G_i\text{-right invariant}\}.$$

If f is G_i-right invariant, then for all $v \in \mathcal{G}_i$ and all $t \in \mathbb{R}$,

$$f\big(x \exp(tv)\big) = f(x), \quad \forall x \in G.$$

If f is differentiable, we get that

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} f\big(x\exp(tv)\big) = \langle \nabla f(x), \mathrm{d}(L_x)_e v\rangle, \quad \forall v \in \mathcal{G}_i,$$

where $L_x : G \to G$ is the left multiplication by x. Since L_x is an isometry of G, its differential at e, $\mathrm{d}(L_x)_e$, is an isometry between the Euclidean spaces $T_eG = \mathcal{G}$ and T_xG. In particular, we will exploit the invariance property in the following form

$$(\mathrm{d}(L_x)_e)^{-1}\nabla f(x) \in \mathcal{G}_i^{\perp}, \quad \forall x \in G. \tag{11}$$

Roughly speaking, a G_i-right-invariant function f "depends" only on $\mathcal{G}_i^{\perp}$ in the sense that the gradient $\nabla f(x)$ is in the direction $\mathcal{G}_i^{\perp}$ transported on T_xM:

$$\mathcal{E}_i(x) = \mathrm{d}(L_x)_e\,\mathcal{E}_i,$$

setting $\mathcal{E}_i := \mathcal{G}_i^{\perp}$. With this formalism, the condition to have a Brascamp–Lieb inequality boils down to the existence of a decomposition of the identity in the Lie algebra.

Theorem 7. Let G be a connected compact Riemannian Lie group. Let $(G_i)_{i=1}^m$ be connected Lie subgroups and let $\mathcal{E}_i := \mathcal{G}_i^{\perp}$be the orthogonal complements in the Lie algebra $\mathcal{G}$ of G of their Lie algebras $(\mathcal{G}_i)_{i=1}^m$. Assume that for given $d_1, \dots, d_m > 0$ the following inequality holds between symmetric linear maps of $\mathcal{G}$:

$$\sum_{i=1}^m d_i P_{\mathcal{E}_i} \le \mathrm{Id}_{\mathcal{G}}. \tag{12}$$

Then the **BL**-condition (3) is satisfied. In particular, if for $i = 1, \dots, m$, $f_i : G \to \mathbb{R}^+$ is G_i-right invariant, it holds

$$\int_G \prod_{i=1}^m f_i^{d_i}\,\mathrm{d}\mu \le \prod_{i=1}^m \left(\int_G f_i\,\mathrm{d}\mu\right)^{d_i}. \tag{13}$$

□

Proof. We consider the Heat kernel on G. The Laplace–Beltrami operator commutes with right multiplication by the elements of the group so that the commutation relation is verified, in particular $P_t f_i$ is again G_i invariant. Next let us check Condition (3) in the form (8) put forward in the beginning of the Riemannian case. If for $i \le n$, h_i is a differentiable G_i-invariant function then, rewriting (11) as

$$\mathrm{d}(L_{x^{-1}})_e \nabla h_i(x) \in \mathcal{E}_i$$

we get $P_{\mathcal{E}_i} \mathrm{d}(L_{x^{-1}})_e \nabla h_i(x) = \mathrm{d}(L_{x^{-1}})_e \nabla h_i(x)$. Using the fact that $\mathrm{d}(L_{x^{-1}})_e$ is an isometry between $T_x M$ and $\mathcal{G}$ and the decomposition of the identity in $\mathcal{G}$, we see that

$$\Big\| \sum_i d_i \, \nabla h_i(x) \Big\|^2 = \Big\| (\mathrm{d}L_{x^{-1}}) \Big(\sum_i d_i \, \nabla h_i(x) \Big) \Big\|^2 = \Big\| \sum_i d_i \, (\mathrm{d}L_{x^{-1}}) \nabla h_i(x) \Big\|^2$$
$$\leq \sum_i d_i \, \big\| (\mathrm{d}L_{x^{-1}}) \nabla h_i(x) \big\|^2 = \sum_i d_i \, \big\| \nabla h_i(x) \big\|^2.$$

The result follows. Equivalently, we could have said that the isometry $\mathrm{d}L_x$ pushes forward the decomposition (12) from $\mathcal{G} = T_e G$ to the decomposition (10) on $T_x G$. ■

3.2.1 *Calculations in $SO(n)$*

We consider subgroups related to the natural action of $SO(n)$ on $\mathbb{R}^n$ and study the relationship between decompositions of the identity of $\mathbb{R}^n$ and the ones induced on $\mathcal{A}_n = so(n)$, the set of antisymmetric $n \times n$ matrices, which is the Lie algebra of $SO(n)$. The Euclidean structure on $\mathcal{A}_n$ is given by the Hilbert–Schmidt norm and the corresponding scalar product $\langle A, B \rangle = \mathrm{Tr}({}^t AB) = -\mathrm{Tr}(AB)$.

We will consider as before functions on $SO(n)$, which are right invariant with respect to subgroups. There exists two natural subgroups associated to a subspace $E \subset \mathbb{R}^n$: $\mathrm{Fix}(E)$ and $\mathrm{Stab}(E)$.

Lemma 8. Let E be a vector subspace of $\mathbb{R}^n$. Consider the group

$$H = \mathrm{Fix}(E) := \{U \in SO(n);\ U_{|E} = \mathrm{Id}\}$$

and let $\mathcal{H}$ be its Lie algebra. We have $\mathcal{H} = \{A \in \mathcal{A}_n;\ A_{|E} = 0\}$ and if $P_{\mathcal{E}} : \mathcal{A}_n \to \mathcal{A}_n$ denotes the orthogonal projection on to $\mathcal{E} := \mathcal{H}^\perp$, we have that

$$\|P_{\mathcal{E}}(A)\|^2 = 2\|P_E A\|^2 - \|P_E A P_E\|^2, \quad \forall A \in \mathcal{A}_n.$$

Moreover a function $f : G \to \mathbb{R}$ is H-right invariant means that $f(U)$ is actually a function of $U_{|E}$. □

Proof. The equality $\mathcal{H} = \{A \in \mathcal{A}_n;\ A_{|E} = 0\}$ is obvious. Let us check that the orthogonal projection of $A \in \mathcal{A}_n$ on to $\mathcal{H}$ is $P_{E^\perp} A P_{E^\perp}$. Indeed, the latter is clearly antisymmetric and vanishes on vectors of E, so it belongs to $\mathcal{H}$. It remains to check the orthogonality condition: if $B \in \mathcal{H}$,

$$\begin{aligned}-\langle B, A - P_{E^\perp} A P_{E^\perp}\rangle &= \mathrm{Tr}\Big(B\big(A - P_{E^\perp} A P_{E^\perp}\big)\Big)\\ &= \mathrm{Tr}(BA) - \mathrm{Tr}(BP_{E^\perp} A P_{E^\perp}).\end{aligned}$$

Since B vanishes on E, $B = B(P_E + P_{E^\perp}) = BP_{E^\perp}$ and taking adjoints $P_{E^\perp} B = B$. It is then clear that $\mathrm{Tr}(BP_{E^\perp} AP_{E^\perp}) = \mathrm{Tr}(BA)$. The orthogonality follows.

Since $\mathcal{E} = \mathcal{H}^\perp$ and denoting for shortness P instead of P_E, and I instead of $\mathrm{Id}_{\mathbb{R}^n}$, we have

$$P_{\mathcal{E}}(A) = A - P_{E^\perp} AP_{E^\perp} = A - (I - P)A(I - P) = PA + AP - PAP.$$

Eventually, since $P_{\mathcal{E}}$ is a self-adjoint involution

$$\begin{aligned}\|P_{\mathcal{E}}(A)\|^2 &= \langle A, P_{\mathcal{E}} A\rangle = -\mathrm{Tr}(A(PA + AP - PAP))\\ &= -2\mathrm{Tr}(A^2 P) + \mathrm{Tr}(APAP) = 2\|PA\|^2 - \|PAP\|^2.\end{aligned}$$

The statement on H-right-invariant functions is easy. Such a function can be viewed as a function on $SO(n)/H \approx SO(n)/SO(E^\perp)$, which can be identified to the Stiefel manifold of orthogonal frames of size $\dim(E)$ in $\mathbb{R}^n$. More explicitly, $U_1 H = U_2 H$ is equivalent to $U_2^{-1} U_1 \in H$, that is for all $x \in E$, $U_1(x) = U_2(x)$. Hence, the restriction of U to E characterizes the class of U in the quotient. ■

Lemma 9. Let E be a vector subspace of $\mathbb{R}^n$. Consider the group

$$H = \mathrm{Stab}(E) := \{U \in SO(n);\ U(E) \subset E\}$$

and let $\mathcal{H}$ be its Lie algebra. If $P_{\mathcal{E}} : \mathcal{A}_n \to \mathcal{A}_n$ denotes the orthogonal projection on to $\mathcal{H}^\perp$, it holds

$$\|P_{\mathcal{E}}(A)\|^2 = 2\|P_E A\|^2 - 2\|P_E AP_E\|^2, \quad \forall A \in \mathcal{A}_n.$$

Moreover a function $f : G \to \mathbb{R}$ is H-right invariant means that $f(U)$ is actually a function of $U(E)$. □

Proof. The argument is very similar to the one of the previous lemma. First, note that

$$H = \{U \in SO(n);\ U(E) = E\} = \{U \in SO(n);\ U(E) \subset E \text{ and } U(E^{\perp}) \subset E^{\perp}\}.$$

For a H-right-invariant function f, $f(U)$ depends only on UH. Since $U_1 H = U_2 H$ is equivalent to $U_1(E) = U_2(E)$, the quantity $f(U)$ depends on $U(E)$. In other words, f factors through the Grassmann manifold of spaces of dimension $\dim(E)$ in $\mathbb{R}^n$.

One easily checks that $\mathcal{H} = \{A \in \mathcal{A}_n;\ A(E) \subset E \text{ and } A(E^{\perp}) \subset E^{\perp}\}$. The orthogonal projection for $A \in \mathcal{A}_n$ on to $\mathcal{H}$ is $P_E A P_E + P_{E^{\perp}} A P_{E^{\perp}}$. Indeed, this is clearly an antisymmetric map for which E and $E^{\perp}$ are stable. Moreover for $B \in \mathcal{H}$, it is clear that $B = P_E B P_E + P_{E^{\perp}} B P_{E^{\perp}}$. Hence

$$-\langle B, A - P_E A P_E + P_{E^{\perp}} A P_{E^{\perp}} \rangle = \mathrm{Tr}(BA) - \mathrm{Tr}(B P_E A P_E) - \mathrm{Tr}(B P_{E^{\perp}} A P_{E^{\perp}}) = 0.$$

Eventually, since $\mathcal{E} = \mathcal{H}^{\perp}$, $P_{\mathcal{E}}(A) = A - P_E A P_E + P_{E^{\perp}} A P_{E^{\perp}}$. So calculating as in the previous lemma, we have $P_{\mathcal{E}}(A) = PA + AP - 2PAP$ and

$$\begin{aligned} \|P_{\mathcal{E}}(A)\|^2 &= \langle A, P_{\mathcal{E}} A\rangle = -\mathrm{Tr}(A(PA + AP - 2PAP)) \\ &= -2\mathrm{Tr}(A^2 P) + 2\mathrm{Tr}(APAP) = 2\|PA\|^2 - 2\|PAP\|^2. \end{aligned}$$ ■

The connection between decompositions of identity of $\mathbb{R}^n$ and of $\mathcal{A}_n$ is explained next.

Proposition 10. For $i = 1, \dots, m$, let $c_i > 0$, E_i be a vector subspace of $\mathbb{R}^n$ and let G_i be either $\mathrm{Fix}(E_i)$ or $\mathrm{Stab}(E_i)$. Denote by $\mathcal{E}_i = \mathcal{G}_i^{\perp}$ the orthogonal of $\mathcal{G}_i$ (the Lie algebra of G_i) in $\mathcal{A}_n$. We have

$$\sum_{i=1}^{m} c_i P_{E_i} \leq \mathrm{Id}_{\mathbb{R}^n} \implies \sum_{i=1}^{m} \frac{c_i}{2} P_{\mathcal{E}_i} \leq \mathrm{Id}_{\mathcal{A}_n}.$$

As a consequence, if $\sum_{i=1}^{m} c_i P_{E_i} \leq \mathrm{Id}_{\mathbb{R}^n}$ then inequality (13) holds on $G = SO(n)$ (equipped with its uniform probability measure μ) whenever each $f_i(U)$ is a function of $U(E_i)$ or of $U_{|E_i}$, $i = 1, \dots, m$. □

Proof. By Lemma 8 and Lemma 9, for any $A \in \mathcal{A}_n$, $\|P_{\mathcal{E}_i}(A)\|^2 \leq 2\|P_{E_i}A\|^2$. Hence

$$\sum_{i=1}^m \frac{c_i}{2}\|P_{\mathcal{E}_i}(A)\|^2 \leq \sum_{i=1}^m c_i\|P_{E_i}A\|^2 = \sum_{i=1}^m c_i \mathrm{Tr}({}^tAP_{E_i}A)$$
$$= \mathrm{Tr}\Big({}^tA\big(\sum_{i=1}^m c_i P_{E_i}\big)A\Big) \leq \mathrm{Tr}({}^tAA) = \|A\|^2.$$

■

Note that we have not used the full strength of Lemmata 8 and 9, since we have discarded the terms $\|P_{E_i}AP_{E_i}\|^2$. However, in the case where the E_i's are one-dimensional subspaces of $\mathbb{R}^n$, these terms vanish, since in this particular case we have

$$P_{E_i}AP_{E_i} = 0.$$

So, if $E_i = \mathbb{R}u_i$ where the u_i's are norm 1 vectors satisfying the decomposition of the identity

$$\sum_{i=1}^m c_i\, u_i \otimes u_i = \mathrm{Id}_{\mathbb{R}^n} \tag{14}$$

where $u_i \otimes u_i = P_{E_i}$, then we have, with the notation of the proposition,

$$\sum_{i=1}^m \frac{c_i}{2} P_{\mathcal{E}_i} = \mathrm{Id}_{\mathcal{A}_n}.$$

We do not loose in the passage to the Lie algebra. A particular case of interest is when $m = n$, $c_1 = \ldots = c_n = 1$ and $(u_1, \ldots, u_n)$ is an orthonormal basis of $\mathbb{R}^n$.

For higher dimensional E_i's, it is possible, in some specific situations, to recombine the terms $\|P_{E_i}AP_{E_i}\|^2$ to recover a multiple of $\|A\|^2$ and to improve the exponents in the correlation inequality. This is easily seen for coordinate subspaces, that is, spaces spanned by vectors of the canonical basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$ (or of any given orthonormal basis, of course). The following proposition puts forward a typical set of conditions in order that **BL**-condition (3) is fulfilled. It will appear later in similar forms.

Proposition 11. Let $\mathcal{I}$ be a collection of subsets of $\{1, \ldots, n\}$. Assume that it is written as a disjoint union $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$. For each non-empty subset $I \in \mathcal{I}$, let $c_I \geq 0$, $E_I :=$ span$(e_i;\ i \in I)$, and $f_I : SO(n) \to \mathbb{R}^+$ such that

- if $I \in \mathcal{I}_1$ then for all U, $f_I(U)$ only depends on $U_{|E_I}$,
- if $I \in \mathcal{I}_2$ then for all U, $f_I(U)$ only depends on $U(E_I)$.

If for all $1 \le i, j \le n$ with $i \neq j$ it holds:

$$\sum_{\substack{I \in \mathcal{I}_1 \\ I \cap \{i,j\} \neq \emptyset}} c_I + \sum_{\substack{I \in \mathcal{I}_2 \\ \mathrm{card}(I \cap \{i,j\}) = 1}} c_I \le 1,$$

then **BL**-condition(3) is satisfied and in particular

$$\int_{SO(n)} \prod_{I \in \mathcal{I}} f_I^{c_I} \,\mathrm{d}\mu \le \prod_{I \in \mathcal{I}} \left(\int_{SO(n)} f_I \,\mathrm{d}\mu \right)^{c_I}.$$

□

Proof. Simply note that for $A = (a_{i,j})_{1 \le i,j \le n} \in \mathcal{A}_n$, $\|P_{E_I} A P_{E_I}\|^2 = \sum_{i,j \in I} a_{i,j}^2$ and

$$\|P_{E_I} A\|^2 = \mathrm{Tr}({}^t A P_{E_i} A) = \mathrm{Tr}\Big(\sum_{i \in I} Ae_i \otimes Ae_i \Big) = \sum_{i \in I} \|Ae_i\|^2 = \sum_{i \in I} \sum_{j=1}^{n} a_{i,j}^2.$$

Let us set $\lambda_I := 1$ if $I \in \mathcal{I}_1$, $\lambda_I := 2$ if $I \in \mathcal{I}_2$. Using Lemmata 8 and 9, and the antisymmetry of $A \in \mathcal{A}_n$, we have

$$\begin{aligned}
&\sum_I c_I \|P_{\mathcal{E}_I}(A)\|^2 = \sum_I c_I \left(2\|P_{E_i} A\|^2 - \lambda_I \|P_{E_I} A P_{E_I}\|^2 \right) \\
&= \sum_I c_I \left(2 \sum_{i \in I} \sum_{j=1}^{n} a_{i,j}^2 - \lambda_I \sum_{i,j \in I} a_{i,j}^2 \right) = \sum_{i,j=1}^{n} a_{i,j}^2 \left(2 \sum_{I;\, i \in I} c_I - \sum_{I;\, i,j \in I} \lambda_I c_I \right) \\
&= 2 \sum_{1 \le i < j \le n} a_{i,j}^2 \left(\sum_{I;\, i \in I} c_I + \sum_{I;\, j \in I} c_I - \sum_{I;\, i,j \in I} \lambda_I c_I \right).
\end{aligned}$$

The latter is upper bounded by $\|A\|^2$ as soon as for all $i \neq j$,

$$\sum_I c_I \left(\mathbf{1}_{i \in I} + \mathbf{1}_{j \in I} - \lambda_I \mathbf{1}_{i,j \in I} \right) \le 1,$$

which is exactly our hypothesis on the coefficients $(c_I)_{I \in \mathcal{I}}$. Hence, $\sum_I c_I P_{\mathcal{E}_I} \le \mathrm{Id}_{\mathcal{A}_n}$ and Theorem 7 yields the claim. ■

Let us restate the previous result for the case where all the c_i's are identical.

Proposition 12. Let $\mathcal{I}$ be a family of subsets of $\{1, \dots, n\}$, and consider

$$p := \max_{1 \le i < j \le n} \operatorname{card}\big\{I \in \mathcal{I};\ I \cap \{i, j\} \ne \emptyset\big\},$$
$$q := \max_{1 \le i < j \le n} \operatorname{card}\big\{I \in \mathcal{I};\ \operatorname{card}\big(I \cap \{i, j\}\big) = 1\big\},$$

then for all non-negative functions g_I, h_I defined on suitable spaces,

$$\int \prod_{I \in \mathcal{I}} g_I(U_{|E_I}) \,\mathrm{d}\mu(U) \le \prod_{I \in \mathcal{I}} \left(\int g_I(U_{|E_I})^p \,\mathrm{d}\mu(U) \right)^{\frac{1}{p}},$$
$$\int \prod_{I \in \mathcal{I}} h_I(U(E_I)) \,\mathrm{d}\mu(U) \le \prod_{I \in \mathcal{I}} \left(\int h_I(U(E_I))^q \,\mathrm{d}\mu(U) \right)^{\frac{1}{q}}.$$

□

Let us put forward two particular cases of application of the previous result:

- Blocks of coordinates: if $\mathcal{I}$ is a non-trivial partition of $\{1, \dots, n\}$, then each pair $\{i, j\}$ meets at most two sets in the family and we get $p = q = 2$.
- Loomis–Whitney inequality: if $\mathcal{I}$ is the family of all subsets of $\{1, \dots, n\}$ of size k, then any pair meets $\binom{n}{k} - \binom{n-2}{k}$ sets. Hence, we have

$$p = \binom{n}{k} - \binom{n-2}{k} = \binom{n-1}{k-1} + \binom{n-2}{k-1}.$$

  However, the number of sets of cardinality k that intersect a given pair in exactly one point is $\binom{n}{k} - \binom{n-2}{k} - \binom{n-2}{k-2} = 2\binom{n-2}{k-1}$. So we get a smaller exponent

$$q = 2\binom{n-2}{k-1}.$$

It is worth noting that a direct application of Proposition 10 would have given *worst* estimates (when $k \ge 2$), in both cases. Indeed, if we denote by P^I the projection on to a subspace spanned by $\{e_i,\ i \in I\}$ for $I \subset \{1, \dots n\}$, we have

$$\sum_{|I|=k} \frac{n}{k\binom{n}{k}} P^I = \mathrm{Id}_{\mathbb{R}^n}$$

and therefore we would get exponents p and q equal to $2\frac{k}{n}\binom{n}{k} = 2\binom{n-1}{k-1}$.

Remark 13. One can take advantage of the terms $\|P_E A P_A\|^2$ in more general situations. They have to be rather symmetric though. Letting $2 \le k \le n-1$, one instance is given by the family of all the spaces spanned by any k vertices of a regular simplex in $\mathbb{R}^n$ with center of mass at the origin. □

3.2.2 *Passing to quotients*

So far, we have taken advantage of right invariances of the functions f_i. Plainly, similar results hold if all the functions are left invariant instead. It would be very interesting to get better inequalities when the functions f_i enjoy left and right invariances together (this would encompass functions on $SO(n)$ depending on matrices U only through submatrices). Unfortunately, our approach does not give interesting general results in this direction (nothing better than what one gets by applying first Hölder's inequality in order to get two integrals; each of these integrals is then upper-bounded by using only one-sided invariance). In the specific case when the functions have different right invariances and a common left invariance, our results can be stated instead on the left quotient. This is a way to get inequalities for homogeneous spaces corresponding to a compact Riemannian Lie group.

Let us illustrate this remark for the sphere: if E_i is a subspace of $\mathbb{R}^n$ and $f_i : \mathbb{S}^{n-1} \to \mathbb{R}^+$ is of the form $f_i(x) = g_i(P_{E_i} x)$, we may introduce $F_i : SO(n) \to \mathbb{R}^+$ defined by $F_i(U) = g_i(P_{E_i}{}^t U e_1) = g_i({}^t(U P_{E_i}) e_1)$. Then F_i is $\mathrm{Fix}(E_i)$-right invariant and also $\mathrm{Fix}(\mathbb{R}e_1)$-left invariant. Applying our results on $SO(n)$ and using the fact that the law of ${}^t U e_1$ under the Haar probability measure on $SO(n)$ is the uniform distribution on the sphere we recover the main result of [5], which extends inequality (1): if $\sum_i c_i P_{E_i} \le \mathrm{Id}_{\mathbb{R}^n}$ then

$$\int_{\mathbb{S}^{n-1}} \prod_i f_i^{c_i/2} \mathrm{d}\sigma \le \prod_i \left(\int_{\mathbb{S}^{n-1}} f_i \, \mathrm{d}\sigma \right)^{c_i/2}.$$

Moreover, if $f : \mathbb{S}^{n-1} \to \mathbb{R}^+$ is of the form $f(x) = g\big(|P_E x|\big)$, then the function $F : SO(n) \to \mathbb{R}^+$ defined by $F(U) = g\big(|P_E{}^t U e_1|\big)$ is $\mathrm{Stab}(E_i)$-right invariant and $\mathrm{Fix}(\mathbb{R}e_1)$-left invariant. This allows us to transfer all of our $SO(n)$ results to the sphere.

Actually, a more general route is to note that **BL**-condition, in the form (12), passes to quotient.

Lemma 14. Let E be a Riemannian homogeneous space and G a compact Riemannian Lie group of isometries acting transitively on E. Assume we are in the situation of Theorem 7. A function $f : E \to \mathbb{R}$ is said G_i invariant if $f(g \cdot x) = f(x)$ for every

$x \in E$ and $g \in G_i$. We can consider the associated $\tilde{T}_i : E \to E/G_i$ or more simply, with the notation (9),

$$\mathcal{E}_i(x) = \{\nabla f(x); \quad f : E \to \mathbb{R} \text{ is } G_i \text{ invariant}\}.$$

If Condition (12) holds on G, then the **BL**-condition holds in E in the equivalent form (10). □

Proof. Fix $x \in E$ and let $G_x = \{g \in G; \ g \cdot x = x\}$. Then, if we decompose the algebra $\mathcal{G} = T_{\mathrm{Id}} G$ (equipped with its Euclidean structure) as an orthonormal sum $\mathcal{G} = \mathcal{G}_x \oplus \mathcal{G}_x^\perp$ where $\mathcal{G}_x$ is the Lie algebra associated to G_x, we have that $\mathcal{G}_x^\perp$ is isometric to $T_x M$ by the isometry map

$$\pi = \pi_x \, : \, A \longrightarrow \pi(A) = \frac{\mathrm{d}}{\mathrm{d}t}_{|t=0} \exp(tA) \cdot x.$$

We see that $\pi(\mathcal{G}_i) \subset \mathcal{E}_i(x)^\perp$ and therefore $\mathcal{E}_i(x) \subset \pi(\mathcal{E}_i)$. Note that $G_x \subset G_i$ and $\mathcal{E}_i \subset \mathcal{G}_x^\perp$. Since $P_{\pi(\mathcal{E}_i)} = \pi P_{\mathcal{E}_i} \pi^{-1}$, we get from (12) that

$$\sum_{i=1}^m c_i \, P_{\mathcal{E}_i(x)} \leq \mathrm{Id}_{T_x E}. \qquad ■$$

It is sometimes necessary to work directly on quotients, in particular for quotients of finite measure with a cover of infinite measure. We briefly discuss the example of the flat torus $(\mathbb{R}/\mathbb{Z})^n$. We consider for $i = 1, \dots, m$, rational vectors $u_i \in \mathbb{Q}^n$. For each i, let ℓ_i be the largest common divisor of the numbers $\langle u_i, e_1 \rangle, \dots, \langle u_i, e_n \rangle$. In order to define the map $x \mapsto \langle x, u_i \rangle$ on the torus, one has to identify $\langle u_i, e_k \rangle$ to 0 for all k. This amounts to quotient $\mathbb{R}$ by $\sum_{k=1}^m \langle u_i, e_k \rangle \mathbb{Z} = \ell_i \mathbb{Z}$. Let $T_i : (\mathbb{R}/\mathbb{Z})^n \to \mathbb{R}/\ell_i\mathbb{Z}$ be the map defined by $T_i(x) = \langle x, u_i \rangle \mod \ell_i$. One easily checks that the Laplacian commutes with T_i (same calculation as in $\mathbb{R}^n$). Since for every x, $\nabla(f_i \circ T_i)(x)$ is a multiple of u_i, if $\sum_{i=1}^m c_i u_i \otimes u_i \leq \mathrm{Id}_{\mathbb{R}^n}$ it follows that

$$\int_{(\mathbb{R}/\mathbb{Z})^n} \prod_i f_i(\langle x, u_i \rangle)^{c_i} \mathrm{d}x \leq \prod_{i=1}^m \left(\int_{(\mathbb{R}/\mathbb{Z})^n} f_i(\langle x, u_i \rangle) \mathrm{d}x \right)^{c_i} = \prod_{i=1}^m \left(\int_{\mathbb{R}/\ell_i\mathbb{Z}} f_i \right)^{c_i}.$$

3.3 Dirichlet distributions and their relatives

For $x \in \mathbb{R}^n$, we set $S(x) = x_1 + \cdots + x_n$. Let $\alpha \in (0, +\infty)^n$, then by definition the Dirichlet law $D_{n-1}(\alpha)$ is the distribution of

$$\frac{(X_1, \ldots, X_{n-1})}{X_1 + \cdots + X_n}$$

where $X_1, \ldots, X_n$ are independent random variables such that for each i, X_i is $Gamma(\alpha_i)$ distributed. More precisely, it is supported on $T_{n-1} = \{y \in \mathbb{R}_+^{n-1};\ y_1 + \cdots + y_{n-1} \le 1\}$ and

$$D_{n-1}(\alpha)(\mathrm{d}y) = \frac{\Gamma(S(\alpha))}{\prod_{i\le n}\Gamma(\alpha_i)}\Big(\prod_{i\le n-1} y_i^{\alpha_i-1}\Big)\Big(1 - \sum_{i\le n-1} y_i\Big)^{\alpha_n-1}\mathbf{1}_{T_{n-1}}(y)\,\mathrm{d}y.$$

In order to get more symmetric results, we prefer to work with another representation: we consider the law $\widetilde{D}_{n-1}(\alpha)$ of

$$\frac{(X_1, \ldots, X_n)}{X_1 + \cdots + X_n}.$$

It is supported on the regular simplex $\Delta_{n-1} = \{y \in \mathbb{R}_+^n;\ y_1 + \cdots + y_n = 1\}$ and its density with respect to Lebesgue measure on Δ_{n-1} is proportional to $y \mapsto \prod_{i\le n} y_i^{\alpha_i-1}$. Recall that some Dirichlet distributions are closely related to uniform spherical measures. Indeed if G_i are independent variables with distribution $\exp(-t^2)\mathrm{d}t/\sqrt{\pi}$, then the uniform measure on $\mathbb{S}^N$ coincides with the law of

$$\frac{(G_1, \ldots, G_N)}{\sqrt{G_1^2 + \cdots + G_N^2}}.$$

Note that G_i^2 has distribution $Gamma(1/2)$. Write $N = k_1 + \cdots + k_n$. It is then clear that the image of the uniform probability on $\mathbb{S}^{N-1}$ by the map

$$x \mapsto (x_1^2 + \cdots + x_{k_1}^2,\ x_{k_1+1}^2 + \cdots + x_{k_1+k_2}^2,\ \ldots,\ x_{k_1+\cdots+k_{n-1}+1}^2 + \cdots + x_N^2).$$

is $\widetilde{D}_{n-1}(k_1/2, \ldots, k_n/2)$. This allows us to transfer some of our spherical results, but only to Dirichlet laws with half-integer coefficients. In order to deal with general coefficients, the following direct study is needed.

The measure $D_{n-1}(\alpha)$ is known (see [16, 21]) to be reversible and ergodic for the following Fleming–Viot operator

$$L_\alpha f = \sum_{i\le n-1} x_i \partial^2_{i,i} f - \sum_{i,j\le n-1} x_i x_j \partial^2_{i,j} f + \sum_{i\le n-1} \big(\alpha_i - S(\alpha) x_i\big)\partial_i f.$$

In the symmetric representation associated to $\widetilde{D}_{n-1}(\alpha)$, it is natural to consider the operator $\widetilde{L}_\alpha$ defined for smooth functions $f : \mathbb{R}^n \to \mathbb{R}^+$ and for $x \in \Delta_{n-1}$ by

$$\widetilde{L}_\alpha f(x) = \sum_{i\le n} x_i \partial^2_{i,i} f(x) - \sum_{i,j\le n} x_i x_j \partial^2_{i,j} f(x) + \sum_{i\le n} \big(\alpha_i - S(\alpha) x_i\big)\partial_i f(x).$$

It is not hard to check that $\widetilde{L}_\alpha f$ only depends on the restriction of f to Δ_{n-1} (in the intrinsic formulation $\partial_i g$ is to be understood as $Dg \cdot P_H e_i = Dg \cdot (e_i - \mathbf{1}/n)$, where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$ and $H = \mathbf{1}^\perp$). However, it is convenient to be able to apply $\widetilde{L}_\alpha f$ to functions f defined on the whole space. For example, if we write $f(y) = g(y_1, \dots, y_{n-1})$, $y \in \Delta_{n-1}$ then it is clear that $\widetilde{L}_\alpha f(y) = L_\alpha g(y_1, \dots, y_{n-1})$; hence the properties of L_α will pass to $\widetilde{L}_\alpha f$ (in particular $\widetilde{D}_{n-1}(\alpha)$ is reversible and ergodic for the semigroup generated by $\widetilde{L}_\alpha f$).

The carré du champ of $\widetilde{L}_\alpha$ can be expressed in the following convenient form, for $x \in \Delta_{n-1}$:

$$\begin{aligned}\Gamma(f) &= \sum_{i\le n} x_i (\partial_i f)^2 - \sum_{i,j\le n} x_i x_j \partial_i f \partial_j f \\ &= \sum_{i\le n} x_i (\partial_i f)^2 - \Big(\sum_{i\le n} x_i \partial_i f\Big)^2 \\ &= \frac{1}{2}\sum_{i\ne j} x_i x_j (\partial_i f - \partial_j f)^2,\end{aligned}$$

where we have noted that $\Gamma(f)$ is actually a variance with respect to the probability measure $\sum x_i \delta_i$. The last formula comes from the representation $\mathrm{Var}(X) = (1/2)E((X - X')^2)$ where X' is an independent copy of X. We are ready to establish

Proposition 15. Let $\mathcal{I}$ be a collection of subsets of $\{1, \dots, n\}$. Assume that it is written as a disjoint union $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$. For each non-empty subset $I \in \mathcal{I}$, let $c_I \ge 0$, and $f_I : \Delta_{n-1} \to \mathbb{R}^+$ such that

- if $I \in \mathcal{I}_1$ then for all x, $f_I(x)$ only depends on $(x_k)_{k\in I}$,
- if $I \in \mathcal{I}_2$ then for all x, $f_I(x)$ only depends on $\sum_{k\in I} x_k$.

If for all $1 \le i, j \le n$ with $i \neq j$ it holds:

$$\sum_{\substack{I\in\mathcal{I}_1\\ I\cap\{i,j\}\neq\emptyset}} c_I + \sum_{\substack{I\in\mathcal{I}_2\\ \mathrm{card}(I\cap\{i,j\})=1}} c_I \le 1,$$

then the **BL**-condition (3) is satisfied and if X is $\widetilde{D}_{n-1}(\alpha)$ distributed

$$E\Big(\prod_{I\in\mathcal{I}} f_I^{c_I}(X)\Big) \le \prod_{I\in\mathcal{I}} \Big(Ef_I(X)\Big)^{c_I}.$$

□

Proof. First, we check the commutation relations. Since the coordinates play symmetric roles, we may assume that $I = \{1, \dots, k\}$. Also, we may extend our functions to $\mathbb{R}_+^n$. If for all x, $g(x) = f(x_1 + \cdots + x_n)$ it is obvious that

$$\partial_i g(x) = \begin{cases} 0 & \text{if } i > k \\ f'(x_1+\cdots+x_k) & \text{if } i \le k \end{cases} \quad \text{and} \quad \partial^2_{i,j} g(x) = \begin{cases} 0 & \text{if } i \text{ or } j > k \\ f''(x_1+\cdots+x_k) & \text{if } i, j \le k. \end{cases}$$

It is then clear that $\widetilde{L}_\alpha g(x)$ is a function of $x_1 + \cdots + x_k$. Similarly, if $g(x) = h(x_1, \dots, x_k)$ then $\widetilde{L}_\alpha g(x)$ is a function of $(x_i)_{i\le k}$.

Next, we have to check the analogue of Condition (7), namely

$$\Gamma\Big(\sum_I c_I f_I\Big) \le \sum_I c_I \Gamma(f_I).$$

In view of the above expression of Γ, this amounts to show that for all $x \in \Delta_{n-1}$,

$$\sum_{1\le i\neq j\le n} x_i x_j \Big(\sum_I c_I \partial_i f_I - \sum_I c_I \partial_j f_I\Big)^2 \le \sum_I c_I \sum_{1\le i\neq j\le n} x_i x_j \Big(\partial_i f_I - \partial_j f_I\Big)^2.$$

Hence, it is sufficient to show that for all $i \neq j$, it holds

$$\Big(\sum_I c_I\big(\partial_i f_I - \partial_j f_I\big)\Big)^2 \le \sum_I c_I \Big(\partial_i f_I - \partial_j f_I\Big)^2.$$

If $f_I(x)$ only depends on $(x_k)_{k\in I}$, then $\partial_i f_I - \partial_j f_I = 0$ if $\{i, j\} \cap I = \emptyset$. Moreover if $f_I(x) = g(\sum_{k\in I} x_k)$, then $\partial_i f_I - \partial_j f_I = 0$ also if $\{i, j\} \subset I$. Hence, the summations on I actually only involve the sets $I \in \mathcal{I}_1$ such that $\{i, j\} \cap I \neq \emptyset$ and the sets $I \in \mathcal{I}_2$ such that $\mathrm{card}(\{i, j\} \cap I) = 1$. By hypothesis, the sum of the corresponding coefficients c_I is at most one, so that the required inequality is a mere consequence of the convexity of the square function. Hence, Condition (7) holds true and we get the local inequality. By ergodicity, the inequality passes to the measure $\widetilde{D}_{n-1}(\alpha)$. ■

Let $p > 0$. Let $B_p^n = \{x \in \mathbb{R}^n;\ \sum_i |x_i|^p \le 1\}$ be the unit ball for the ℓ_p norm on $\mathbb{R}^n$. On the corresponding unit sphere $\partial B_p^n = \{x \in \mathbb{R}^n;\ \sum_i |x_i|^p = 1\}$, one often considers the cone measure μ_p^n defined by $\mu_p^n(A) = \mathrm{Vol}_n([0,1].A)/\mathrm{Vol}_n(B_p^n)$, $A \subset \partial B_p^n$. Here $[0,1]\cdot A$ is the intersection of B_p^n with the cone of apex at the origin spanned by A.

Corollary 16. Let X be a random vector on $\mathbb{R}^n$. Assume that it is either uniformly distributed on B_p^n or distributed according to the cone measure on ∂B_p^n. Then for all even functions $f_i : [-1,1] \to \mathbb{R}^+$

$$E\Big(\prod_{i=1}^n f_i(X_i)\Big) \le \prod_{i=1}^n E\Big(f_i(X_i)^2\Big)^{\frac{1}{2}}.$$

□

Proof. This is deduced from a particular case of the previous result on Dirichlet distributions, which ensures that for Y distributed according to $\widetilde{D}_{n-1}(\alpha)$, and $g_i : [0,1] \to \mathbb{R}^+$, a similar inequality holds: $E\prod g_i(Y_i) \le \prod \big(Eg_i^2(Y_i)\big)^{1/2}$. Indeed, the uniform measure on B_p^n and the cone measure on ∂B_p^n can be viewed as symmetrized versions of the images of Dirichlet laws by maps of the form $T(x_1,\dots,x_n) = (T_1(x_1),\dots,T_n(x_n))$. Hence if we choose $g_i = f_i \circ T_i$ in the latter inequality, we get the claim. Let us make this strategy explicit in the case of the cone measure. Let ε_i, G_i, $i = 1,\dots,n$, be independent random variables. Assume that ε_1 is uniform on $\{-1,1\}$ and G_i distributed according to $e^{-t^p}\mathrm{d}t/\Gamma(1+1/p)$. Then it is known that the vector

$$X = \frac{(\varepsilon_1 G_1,\dots,\varepsilon_n G_n)}{\left(G_1^p + \cdots + G_n^p\right)^{\frac{1}{p}}}$$

is distributed according to the cone measure. Hence, $|X_i|^p = G_i^p/(G_1^p + \cdots + G_n^p)$ where G_i^p is *Gamma*$(1/p)$ distributed. So applying the Brascamp–Lieb inequality for $f_i(x) = g_i(x_i^{1/p})$ yields the claim.

A similar approach is possible for the uniform distribution on B_p^n thanks to the representation provided in [6]. ■

Remark 17. The cone measure on ∂B_2^n is simply the uniform measure on $\mathbb{S}^{n-1}$, for which a similar inequality holds for general functions f_i (i.e., it is not necessary to assume that they are even). Hence, one may ask whether the symmetry assumption in the previous corollary is really needed. In order to remove it, one would need a result for symmetrized Dirichlet laws, namely for measures on ∂B_1^n with density with respect to Lebesgue measure proportional to $\prod_i |x_i|^{\alpha_i - 1}$. At first sight, there does not seem to be any problem to extend our approach. However, the ergodicity of these measures is a delicate issue. Indeed, the fact that the density vanishes inside the domain may, in terms of the corresponding random process, create potential barriers that may not be crossed or potential wells into which the process may get stuck. On the technical level, the domain of the operator may be too small to contain enough non-symmetric functions. □

Remark 18. Proposition 15 and many results of this work involve two kinds of functions, which depend only on some coordinates $(x_k)_{k\in I}$ (some depend on all these coordinates and some depend on them only through their sum). It is possible to consider more general dependencies. We have not tried to reach the highest generality in this respect. Let us briefly mention a quite general extension of Proposition 15: we could consider functions f_I where $I = (I_1, \dots, I_K)$ is a collection of disjoint subsets of $\{1, \dots, n\}$, such that $f_I(x)$ only depends on

$$T_I(x) := \left(\sum_{i\in I_1} x_i, \dots, \sum_{i\in I_K} x_i \right).$$

One can check that the map T_I commutes with the Fleming–Viot operator (this uses the disjointness of $I_1, \dots, I_K$). If one considers now a collection of functions $(f_I)_{I\in\mathcal{I}}$ and corresponding coefficients $(f_I)_{I\in\mathcal{I}}$, then a Brascamp–Lieb inequality holds provided for all $i \neq j$ in $\{1, \dots, n\}$, $\sum_{I\in A_{i,j}} c_I \leq 1$, where

$$A_{i,j} = \Big\{ I \in \mathcal{I};\ \exists \ell,\ \mathrm{card}(I_\ell \cap \{i, j\}) = 1 \Big\}.$$

The proof follows the same arguments as the one of Proposition 15. We omit the details. Note that several results of this paper can be extended in an analogous way. □

4 Discrete Models

In this section, we deal with discrete models, and in particular we have to use the **BL**-condition in its brute form (3) since we are no longer working with diffusion generators.

We nevertheless provide a simple criterion that can be worked out for a number of discrete models of interest.

4.1 Abstract criterion

Throughout this paragraph, E will thus be a finite or countable state space. Let K be a Markov kernel on E, that is, $K : E \times E \to [0, \infty)$ is such that for every $x \in E$, $\sum_{y\in E} K(x, y) = 1$. If $f : E \to \mathbb{R}$ is bounded, set $Kf(x) = \sum_{y\in E} K(x, y) f(y)$, $x \in E$. As before, for given maps $T_i : E \to E_i$, $i = 1, \ldots, m$, we say they commute with K if for any function $f : E \to \mathbb{R}$, $K(f \circ T_i)$ is a function of T_i. Again, this amounts to the existence of a Markov kernel K_i on E_i such that $K(f \circ T_i) = K_i(f) \circ T_i$. This definition is of course equivalent to abstract one of Section 2 in terms of the associated Markov generator

$$L = K - \mathrm{Id}.$$

The next proposition provides a simple equivalent criterion for the **BL**-condition (3) in this context.

Proposition 19 (BL-condition in the discrete case). For distinct $x, y \in E$ such that $K(x, y) > 0$, set

$$I_{x,y} = \big\{i \in \{1, \ldots, m\};\ T_i(x) \neq T_i(y)\big\}.$$

Let $c_i \geq 0$, $i = 1, \ldots, m$. Then the **BL**-condition (3) holds if and only if

$$\sum_{i\in I_{x,y}} c_i \leq 1, \qquad \text{for all } x \neq y \text{ in } E \text{ such that } K(x, y) > 0. \tag{15}$$

Therefore, under this condition, for every non-negative function $f_i : E_i \to \mathbb{R}$, $i = 1, \ldots, m$, and every $t \geq 0$,

$$P_t\Big(\prod_{i=1}^m f_i^{c_i} \circ T_i\Big) \leq \prod_{i=1}^m \big(P_t(f_i \circ T_i)\big)^{c_i}.$$

In particular, if K has an ergodic invariant probability measure μ and if for all $x, y \in E$ distinct with $K(x, y) > 0$, it holds $\mathrm{card}\{i = 1, \dots, m;\ T_i(x) \neq T_i(y)\} \leq p$, then choosing $c_i = \frac{1}{p}$, $i = 1, \dots, m$, we have that

$$\int \prod_{i=1}^{m} f_i \circ T_i \,\mathrm{d}\mu \leq \prod_{i=1}^{m} \left(\int (f_i \circ T_i)^p \mathrm{d}\mu \right)^{\frac{1}{p}}.$$

□

Proof. At fixed $x \in E$, Condition (3) may be written as

$$\sum_{y\in E} K(x, y)\big(e^{\sum_{i=1}^{m} c_i [f_i\circ T_i(y) - f_i\circ T_i(x)]} - 1\big) \leq \sum_{i=1}^{m} c_i \sum_{y\in E} K(x, y)\big(e^{f_i\circ T_i(y) - f_i\circ T_i(x)} - 1\big). \tag{16}$$

The sums over i on both sides only run over $i \in I_{x,y}$ so that the preceding inequality is equivalent to saying that

$$\sum_{y\in E} K(x, y)\varphi\Big(\sum_{i\in I_{x,y}} c_i [f_i \circ T_i(y) - f_i \circ T_i(x)]\Big) \leq \sum_{y\in E} K(x, y) \sum_{i\in I_{x,y}} c_i \varphi\Big(f_i \circ T_i(y) - f_i \circ T_i(x)\Big),$$

where $\varphi(u) = e^u - 1$. Since $\varphi(0) = 0$, we can restrict the previous sum over $y \in E \setminus \{x\}$, and of course we can ask that $K(x, y) \neq 0$. Now, for fixed $x, y \in E$ with $x \neq y$ and $K(x, y) \neq 0$, we argue that Condition (15) on the $c_i's$ combines with the convexity of φ to give (pointwise) the desired inequality.

Conversely, if (16) holds for all choices of f_i, $i = 1, \dots, m$, we choose $f_i(z) = \theta \mathbf{1}_{z\neq T_i(x)}$ where $\theta \in \mathbb{R}^+$. Letting $\theta \to +\infty$ and comparing the orders of the terms in (16) show that for each $y \neq x$ with $K(x, y) \neq 0$, we must have $\sum_{i\in I_{x,y}} c_i \leq 1$. ■

Remark 20 (Extension to non-finite settings). The careful reader has probably noticed that the finiteness (or countability) of E is not central in the argument. All the argument works as soon as we can express $L + I =: K$ in terms of a Markov kernel. Indeed, this allows us to reduce the problem to a pointwise inequality. □

We next illustrate instances of the preceding result.

4.2 Examples

4.2.1 *Homomorphisms of finitely generated groups*

Let for example G, G_i, $i = 1, \dots, m$, be finite or countable groups and $T_i : G \to G_i$ be homomorphisms. Let K be a Markov kernel on G. It is clear that each T_i commutes with K.

Assume furthermore that K is left invariant in the sense that $K(gx, gy) = K(x, y)$ for all $x, y, g \in G$. We may let for example G be finitely generated with generating set S, and $K(x, y) = \mathrm{card}\,(S)^{-1} 1_S(y^{-1}x)$, $x, y \in G$. Then, Condition (15) of Proposition 19 amounts to

$$\sum_{i \in I_z} c_i \le 1$$

for every $z \in S$ where $I_z = \{i = 1, \dots, m;\ z \notin \mathrm{Ker}\,(T_i)\}$.

4.2.2 *Coordinates of the symmetric group*

Let E be the symmetric group $\mathcal{S}_n$ over n elements $\{1, \dots, n\}$, $n \ge 2$. This set is the discrete analogue of $SO(n)$. Unlike the continuous setting, there are several possible choices for the kernel K. However in view of the latter proposition, where each couple (x, y) with $K(x, y) > 0$ leads to a linear constraint on the exponents c_i, it is natural to take a small (or even minimal) generating set S and to consider:

$$K(x, y) = \frac{1}{\mathrm{card}(S)} \text{ if there is } \tau \in S \text{ with } y = \tau x.$$

We choose for S the set of all transpositions. The following calculation will show that it is the best choice, since it minimizes the size of the support $\mathrm{supp}(\tau) = \{j;\ \tau(j) \ne j\}$.

The normalized counting measure μ is invariant for K. Actually, S being stable by inverse is also reversible:

$$\begin{aligned}\int (Kf)\, g \,\mathrm{d}\mu &= \int \frac{1}{\mathrm{card}(S)} \sum_{\tau \in S} f(\tau x) g(x) \,\mathrm{d}\mu(x) \\ &= \int \frac{1}{\mathrm{card}(S)} \sum_{\tau \in S} f(y) g(\tau^{-1} y) \,\mathrm{d}\mu(y) = \int (Kg)\, f \,\mathrm{d}\mu.\end{aligned}$$

Let I be a subset of $\{1, \dots, n\}$. We consider the map T_I defined by

$$T_I(x) = x_{|I} = (x(i))_{i \in I}, \quad \forall x \in \mathcal{S}_n.$$

Then T_I commutes with K; indeed

$$K(f\circ T_I)(x) = \frac{2}{n(n-1)}\sum_{\tau\in S}(f\circ T_I)(\tau x)$$

and $T_I(\tau x) = (\tau\circ x)_{|I} = \tau\circ x_{|I}$ depends only on $T_I(x)$. The result of Proposition 19 involves the condition $T_I(x)\neq T_I(y)$ for $K(x,y)>0$. Let us formulate it in a more concrete manner:

$$T_I(x)\neq T_I(\tau x) \Longleftrightarrow \exists\, i\in I,\ x(i)\neq \tau x(i) \Longleftrightarrow I\cap x^{-1}(\mathrm{supp}(\tau))\neq\emptyset.$$

Note that since the proposition involves this condition for all $x\in\mathcal{S}_n$, the set $x^{-1}(\mathrm{supp}(\tau))$ can be any set with the size of the support of τ. Choosing transpositions then clearly appears as the most economical choice.

For $I\subset\{1,\dots,n\}$, we may also consider the map R_I defined by

$$R_I(x) = x(I) = \{x(i),\ i\in I\},\quad \forall x\in\mathcal{S}_n.$$

Then R_I also commutes with K and for any x and any transposition τ, $R_I(x)\neq R_I(\tau x)$ happens if and only if τ moves one point in $x(I)$ outside $x(I)$. Hence

$$R_I(x)\neq R_I(\tau x) \Longleftrightarrow \mathrm{card}\big(I\cap x^{-1}(\mathrm{supp}(\tau))\big) = 1.$$

Combining these observations with Proposition 19 yields a discrete analogue to Proposition 11:

Proposition 21. Let $\mathcal{I}$ be a collection of subsets of $\{1,\dots,n\}$. Assume that it is written as a disjoint union $\mathcal{I}=\mathcal{I}_1\cup\mathcal{I}_2$. For each non-empty subset $I\in\mathcal{I}$, let $c_I\geq 0$ and $f_I:\mathcal{S}_n\to\mathbb{R}^+$ such that

- if $I\in\mathcal{I}_1$ then for all x, $f_I(x)$ only depends on $x_{|I}$,
- if $I\in\mathcal{I}_2$ then for all x, $f_I(x)$ only depends on $x(I)$.

If for all $1\leq i,j\leq n$ with $i\neq j$

$$\sum_{\substack{I\in\mathcal{I}_1\\ I\cap\{i,j\}\neq\emptyset}} c_I + \sum_{\substack{I\in\mathcal{I}_2\\ \mathrm{card}(I\cap\{i,j\})=1}} c_I \leq 1,$$

then the **BL**-condition(3) is satisfied and

$$\int_{\mathcal{S}_n} \prod_{I\in\mathcal{I}} f_I^{c_I}\,\mathrm{d}\mu \le \prod_{I\in\mathcal{I}} \left(\int_{\mathcal{S}_n} f_I\,\mathrm{d}\mu\right)^{c_I}.$$

□

The examples given after Proposition 11 transfer to $\mathcal{S}_n$. For a family $\mathcal{I}$ of subsets of $\{1,\ldots,n\}$, introduce the exponents:

$$p = \max_{i\neq j} \operatorname{card}\big(\{I\in\mathcal{I};\ i\in I,\ \text{or } j\in I\}\big) \quad \text{and} \quad q = \max_{i\neq j} \operatorname{card}\big(\{I\in\mathcal{I};\ \operatorname{card}(I\cap\{i,j\}) = 1\}\big).$$

Then, for functions g_I and h_I defined on suitable sets, we have

$$\int_{\mathcal{S}_n} \prod_{I\in\mathcal{I}} g_I(\sigma_{|I})\,\mathrm{d}\mu(\sigma) \le \prod_{I\in\mathcal{I}} \left(\int_{\mathcal{S}_n} g_I(\sigma_{|I})^p\,\mathrm{d}\mu(\sigma)\right)^{\frac{1}{p}},$$

$$\int_{\mathcal{S}_n} \prod_{I\in\mathcal{I}} h_I(\sigma(I))\,\mathrm{d}\mu(\sigma) \le \prod_{I\in\mathcal{I}} \left(\int_{\mathcal{S}_n} h_I(\sigma(I))^q\,\mathrm{d}\mu(\sigma)\right)^{\frac{1}{q}}.$$

A particular case of interest (where these two cases coincide) is when $\mathcal{I} = \big\{\{1\},\ldots,\{n\}\big\}$. Then, $p = q = 2$ and we recover the inequality on permanents given in [12].

4.2.3 *Slices of the discrete cube and multivariate hypergeometric distributions*

For $n \ge k \ge 0$, let

$$\Omega_{n,k} = \big\{x\in\{0,1\}^n;\ x_1+\cdots+x_n = k\big\}$$

equipped with uniform measure. These sets are discrete analogues of the sphere S^{n-1}. Two elements x, y in $\Omega_{n,k}$ are neighbors if and only if they differ by exactly two of the coordinates, a relation written as $x\sim y$. Let K be the nearest neighbor random walk on $\Omega_{n,k}$ (known as the Bernoulli–Laplace model) defined by

$$Kf(x) = \frac{1}{k(n-k)} \sum_{y\sim x} f(y).$$

It is easy to check that $Kf(x)$ only depends on the ith coordinate x_i of x if this is the case for f. Indeed, the number of neighbors y of x such that $y_i = x_i$ is equal to $(k-x_i)(n-1-k+x_i)$, whereas when $y_i = 1-x_i$, this number is equal to the number of coordinates x_j,

$j \neq i$, such that $x_j = 1 - x_i$. For the coordinate maps $T_i(x) = x_i$, $1 \leq i \leq n$, we are thus in the preceding setting of commuting operators so that Proposition 19 applies with $p = 2$.

Alternatively, one can use the following observation, which was pointed out to us by P. Caputo. The uniform probability measure on $\Omega_{n,k}$ is the image of the uniform probability measure on the permutation group $\mathcal{S}_n$ by the map $x \in \mathcal{S}_n \mapsto (\mathbf{1}_{x(i) \leq k})_{1 \leq i \leq n}$. Consequently, the correlation inequalities derived on $\mathcal{S}_n$ for functions depending on blocks of coordinates pass to $\Omega_{n,k}$ to yield the same result. Such a reasoning may be extended in order to encompass more general distributions. Consider integer numbers $K \leq M$ and $m = (m_i)_{1 \leq i \leq n}$ such that $\sum_i m_i = M$. The multivariate hypergeometric distribution $\mathcal{H}(m, K)$ is defined on $\mathbb{N}^n$ by

$$\mathcal{H}(m, K)(\{(k_1, \ldots, k_n)\}) = \frac{\prod_{i=1}^m \binom{m_i}{k_i}}{\binom{M}{K}}$$

if $k_1 + \cdots + k_n = K$ and for all i, $k_i \leq m_i$ and $H(m, K)(\{(k_1, \ldots, k_n)\}) = 0$ otherwise. Given an urn containing M balls of n different colors, and more precisely m_i of the ith color, if one draws K balls (uniformly) at random then the n-tuple $(X_1, \ldots, X_n)$ consisting of the numbers of balls of each color in the sample is $\mathcal{H}(m, K)$ distributed. It is not hard to check that $\mathcal{H}(m, K)$ coincides with the image of the uniform probability law on the permutation group $\mathcal{S}_M$ by the map

$$\sigma \in \mathcal{S}_M \mapsto T(\sigma) := \left(\operatorname{card}\left\{j \in \left[1 + \sum_{\ell \leq i-1} m_\ell, \sum_{\ell \leq i} m_\ell\right];\ \sigma(j) \leq K\right\}\right)_{i=1}^n.$$

This observation can be used to show that Proposition 15 remains valid if one replaces the Dirichlet laws by multivariate hypergeometric distributions. We only outline the proof. Starting from functions f_I defined on the support of $\mathcal{H}(m, K)$, we consider the functions $g_I := f_I \circ T$. Note that $g_I(\sigma)$ depends on the images by σ of several intervals of $\{1, \ldots, M\}$. Applying Proposition 21 directly would not give the right result, since it only deals with simpler forms of dependencies. Hence, we need to go back to Proposition 19, in the spirit of the proof of Proposition 21 (this is actually related to Remark 18). We omit the details.

4.2.4 *Product spaces and Finner's theorem*

Let us go back to more general distributions (including continuous distributions on non-finite spaces) but in the context of product structures. The hypotheses in Propositions 11, 21 or 19 are reminiscent of Finner's theorem [17], which expresses that if $E = X_1 \times \cdots \times X_n$ is a product space with product probability measure $\mu = \nu_1 \otimes \cdots \otimes \nu_n$, and if, for $i = 1, \ldots, m$, $T_i : E \to E_i$ is the coordinate projection on the space $E_i := \prod_{j \in S_i} X_j$ determined by $S_i \subset \{1, \ldots, n\}$, then for any non-negative functions $f_i : E_i \to \mathbb{R}$, $i = 1, \ldots, m$,

$$\int \prod_{i=1}^{m} f_i^{c_i} \circ T_i \, d\mu \leq \prod_{i=1}^{m} \left(\int f_i \circ T_i d\mu \right)^{c_i}$$

provided that

$$\sum_{i; S_i \ni j} c_i \leq 1 \quad \text{for every} \quad j = 1, \ldots, n.$$

This statement is actually contained in Proposition 19 for a suitable choice of the kernel K. Without loss of generality, we may assume that, for each i, X_i is a finite set equipped with a probability measure ν_i that charges all points. Consider the kernels K_i on X_i given by $K_i(x_i, y_i) = \nu_i(y_i)$, and tensorize them to the product space $E = X_1 \times \cdots \times X_n$ by

$$\mathcal{K} = \frac{1}{n} \sum_{i=1}^{n} \tilde{I} \otimes \cdots \otimes \tilde{I} \otimes K_i \otimes \tilde{I} \otimes \cdots \otimes \tilde{I}$$

where $\tilde{I}$ is defined on E_j by $\tilde{I}(x_j, y_j) = 1_{x_j = y_j}$ (in other words, the associated Markov operator is the identity). The commutation property of the projection operators T_i is obvious. Moreover, for distinct elements x, y in E, $\mathcal{K}(x, y) > 0$ if and only if $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ differ at exactly one coordinate, say j. Now the set of i's such that $T_i(x) \neq T_i(y)$ is exactly the set of i's such that $S_i \ni j$.

In particular, the preceding kernel provides a proof of the classical Hölder inequality on the finite space X equipped with the probability measure ν, and by approximation on any finite measure space.

5 Sums of Squares

In this short paragraph, we briefly illustrate how the ideas developed in the preceding discrete setting may also be of interest for classes of diffusion generators. Assume the generator L is a sum of squares of vector fields on a manifold E,

$$L = \sum_{\ell} X_\ell^2.$$

Let for example $T_i : E \to \mathbb{R}^{k_i}$, $i = 1, \ldots, m$, be commuting (with L) maps. We interpret $X_\ell T_i$ coordinate by coordinate. The criterion put forward in Proposition 19 then adapts to this setting:

Proposition 22. For every ℓ, let $I_\ell := \big\{i \in \{1, \ldots, m\};\, X_\ell T_i \neq 0\big\}$. Let $c_i \geq 0$, $i = 1, \ldots, m$, be such that

$$\sum_{i \in I_\ell} c_i \leq 1 \quad \text{for every} \ \ \ell.$$

Then, for every non-negative function $f_i : E_i \to \mathbb{R}$, $i = 1, \ldots, m$, and every $t \geq 0$,

$$P_t\bigg(\prod_{i=1}^m f_i^{c_i} \circ T_i\bigg) \leq \prod_{i=1}^m \big(P_t(f_i \circ T_i)\big)^{c_i}.$$

In particular, if for all ℓ, $\operatorname{card}\{i = 1, \ldots, m;\, X_\ell T_i \neq 0\} \leq p$, we may choose $c_i = \frac{1}{p}$, $i = 1, \ldots, m$. □

Proof. Since $\Gamma(f) = \sum_\ell (X_\ell f)^2$, according to Fact 1, the **BL**-condition (3) takes the form

$$\sum_\ell (X_\ell H)^2 \leq \sum_{i=1}^m c_i \sum_\ell \big(X_\ell(f_i \circ T_i)\big)^2, \tag{17}$$

where we recall that $H = \sum_{i=1}^m c_i f_i \circ T_i$. Hence, we are done if we can prove that for every ℓ,

$$\bigg(\sum_{i=1}^m c_i X_\ell(f \circ T_i)\bigg)^2 \leq \sum_{i=1}^m c_i \big(X_\ell(f \circ T_i)\big)^2. \tag{18}$$

If f_i is a function on $\mathbb{R}^{k_i}$, then $X_\ell(f_i \circ T_i) = \langle X_\ell T_i, \nabla f_i(T_i)\rangle$ is zero when $i \notin I_\ell$. Hence, the summations in the above inequality only hold on $i \in I_\ell$. Since by hypothesis $\sum_{i \in I_\ell} c_i \leq 1$, inequality (18) is valid by convexity of the square function. The conclusion follows. ■

We illustrate this result in the context of the Loomis–Whitney inequalities on the sphere. Consider

$$\Delta = \frac{1}{2} \sum_{k,\ell} X_{k\ell}^2 = \frac{1}{2} \sum_{k,\ell} [x_k \partial_\ell - x_\ell \partial_k]^2$$

the Laplace operator on the sphere $S^{n-1} \subset \mathbb{R}^n$. Let A be a subset of $\{1, \ldots, n\}$ with d elements, and consider $T : \mathbb{R}^n \to \mathbb{R}^d$ defined by $T(x) = (x_i)_{i \in A}$. Then $X_{k\ell} T_A = 0$ if and only if $\{k, \ell\} \cap A = \emptyset$. Thus, for every k, ℓ,

$$\begin{aligned} p &= \operatorname{card}\left\{A, |A| = d;\ X_{k,\ell} T_A \neq 0\right\} \\ &= \binom{n}{d} - \binom{n-2}{d} \\ &= \binom{n-1}{d-1} + \binom{n-2}{d-1}. \end{aligned}$$

One instance of application is $d = 1$ (for which $p = 2$) from which we recover inequality (1) involving functions of $T_i(x) = x_i$. The approach here is indeed very close to the one of Carlen, Lieb, and Loss [11].

Remark 23. This viewpoint best explains the analogy between the results on $SO(n)$ and $\mathcal{S}_n$. Indeed, the infinitesimal rotation $x_k \partial_\ell - x_\ell \partial_k$ in $\operatorname{vect}(e_k, e_\ell)$ is the analogue of the transposition $\tau_{k,\ell}$. □

6 Superadditivity of Information for Markov Generators and Entropy of Marginals

In this section, we investigate, from the abstract Markov operator point of view, descriptions of the Brascamp–Lieb inequalities and entropy inequalities for marginals following [10, 11]. As in Section 2, we do not make precise the classes of functions under consideration.

Let (E, μ) be a probability space and $T_i : E \to E_i$ be measurable maps. Given a probability density f on E with respect to μ, denote by f_i its conditional expectation with respect to T_i. In other words, f_i is the unique probability density on E with respect to μ such that, for every bounded measurable $\varphi : E_i \to \mathbb{R}$,

$$\int f \varphi \circ T_i \, \mathrm{d}\mu = \int f_i \, \varphi \circ T_i \, \mathrm{d}\mu. \tag{19}$$

(Since $f_i = h_i \circ T_i$ for some $h_i : E_i \to \mathbb{R}$, h_i may be thought of as the "marginal" of f in the direction of T_i.) As shown in [11], the Brascamp–Lieb inequality (5) may be used, by standard arguments, to prove the entropy inequality for the probability density f

$$\sum_{i=1}^{m} c_i \int f_i \log f_i \mathrm{d}\mu \leq \int f \log f \mathrm{d}\mu. \tag{20}$$

A recent work by Carlen and Cordero-Erausquin [10] shows that there is a full equivalence:

Proposition 24. The following are equivalent.

(i) For every non-negative function $g_i : E_i \to \mathbb{R}$, $i = 1, \dots, m$,

$$\int \prod_{i=1}^{m} g_i^{c_i} \circ T_i \mathrm{d}\mu \leq \prod_{i=1}^{m} \left(\int g_i \circ T_i \mathrm{d}\mu \right)^{c_i}.$$

(ii) For every probability density f with respect to μ,

$$\int f \log f \mathrm{d}\mu \geq \sum_{i=1}^{m} c_i \int f_i \log f_i \mathrm{d}\mu.$$

□

Since semigroup proofs are available for Brascamp–Lieb inequalities, it is natural to hope for semigroup proofs of entropy inequalities. Such an approach was suggested in [5] for spherical measures, on the basis of the corresponding inequality for the Fisher information.

In the remainder of this section, we discuss the extension of this argument to the abstract general framework, encompassing both the continuous and the discrete (non-diffusion) cases.

Let L be a Markov generator on E with semigroup $(P_t)_{t\geq 0}$. We require that L be invariant, symmetric, and ergodic for μ. Denote by Γ the carré du champ operator of L as defined in (6). Hence, the Dirichlet form is expressed as follows

$$\mathcal{E}(f, g) = \int \Gamma(f, g) \, \mathrm{d}\mu = -\int f L g \, \mathrm{d}\mu = -\int g L f \, \mathrm{d}\mu.$$

It is classical that, under suitable domain assumptions,

$$\int f \log f \mathrm{d}\mu = \int_0^\infty \mathrm{d}t \int \Gamma(P_t f, \log P_t f) \, \mathrm{d}\mu. \tag{21}$$

Definition 25. The (Fisher) information associated to (L, μ) of a suitable function $f > 0$ is defined by

$$J(f) := \mathcal{E}(f, \log f) = -\int f L(\log f) \, \mathrm{d}\mu.$$

□

Here "suitable" means that $f, \log f$ belongs to the domain of L in $L^2(\mu)$. Equality (21) becomes

$$\int f \log f \mathrm{d}\mu = \int_0^\infty J(P_t f)\,\mathrm{d}t$$

and so, in view of the commutation between T_i and L, which ensures that

$$P_t(f_i) = (P_t f)_i,$$

we see that the entropy inequality (20) may be derived from its analogue for the Fisher information.

The next result shows that such inequality for Fisher information can indeed be derived directly from the **BL**-condition in our abstract setting. In view of the previous discussion, this therefore provides a different route for proving Brascamp–Lieb inequalities.

Theorem 26 (Superadditivity of Fisher information). Assume that L is a Markov generator on E, which commutes with the maps T_i and that the **BL**-condition (3) holds. Then, for every probability density f on E with respect to μ, under the preceding notation,

$$\sum_{i=1}^m c_i\, J(f_i) \le J(f). \tag{22}$$

□

Before proving this result in full generality, let us note that in the case where L is a diffusion, this theorem can be derived easily, following ideas from [5]. Indeed, when L is a diffusion we have

$$J(f) = \int \frac{\Gamma(f)}{f}\,\mathrm{d}\mu.$$

Using the definition of the conditional density (19) and the chain rule formula for L we see that, for each $i \le m$,

$$J(f_i) = -\int f_i L(\log f_i)\,\mathrm{d}\mu = -\int fL(\log f_i)\,\mathrm{d}\mu = \int \frac{\Gamma(f, f_i)}{f_i}\,\mathrm{d}\mu.$$

Using the Cauchy–Schwarz inequality and (19) again, we get

$$J(f_i)^2 \le \int \frac{\Gamma(f, f_i)^2}{f\,\Gamma(f_i)}\,\mathrm{d}\mu \int \frac{\Gamma(f_i)\,f}{f_i^{\,2}}\mathrm{d}\mu = \int \frac{\Gamma(f, f_i)^2}{f\,\Gamma(f_i)}\,\mathrm{d}\mu \int \frac{\Gamma(f, f_i)}{f_i}\mathrm{d}\mu,$$

which means that

$$J(f_i) \le \int \frac{\Gamma(f, f_i)^2}{f\,\Gamma(f_i)}\,\mathrm{d}\mu.$$

We conclude to (22) after noticing that Condition (3) (in the form (7)) can be expressed in dual form as

$$\sum_{i=1}^m c_i\,\frac{\Gamma(f, f_i)^2}{\Gamma(f_i)} \le \Gamma(f).$$

Similar strategy however does not work in the non-diffusion case, essentially because we don't have a chain rule formula for computing $L(\log f)$. This is also what happens for the similar non-commutative inequalities recently proved by Carlen and Lieb [13]. (A challenging question is whether the approach we propose below applies to the non-commutative setting).

We present here a new method that allows us to treat the general case of a Markov generator. It relies on the following observation, which is of independent interest.

Lemma 27. Assume L is a Markov generator invariant and symmetric for μ. Then for functions $f > 0$ and H of arbitrary sign on E, we have

$$\mathcal{E}(f, H) \le \mathcal{E}(f, \log f) + \int f e^{-H} L\big(e^H\big)\,\mathrm{d}\mu. \tag{23}$$

□

In other words, we have the following dual formulation of Fisher information:

$$J(f) = \sup_H \left\{\mathcal{E}(f, H) - \int f e^{-H} L\big(e^H\big)\,\mathrm{d}\mu\right\}.$$

Proof. Since $\mathcal{E}(f, H) = \int f(-LH)\mathrm{d}\mu$ and $Lg = \lim_{t\to 0} t^{-1}[P_t g - g]$ (for g in the suitable domain), it is enough to establish the following inequality for $P = P_t$ for every $t > 0$:

$$\int f[H - PH]\mathrm{d}\mu \le \int f\log f\mathrm{d}\mu - \int (Pf)\log f\mathrm{d}\mu - \int f\mathrm{d}\mu + \int f e^{-H} P(e^H)\mathrm{d}\mu. \tag{24}$$

By assumption, $P = P_t$ is Markovian and μ is invariant and symmetric for P. By symmetry, the left-hand side is equal to $\int [P(fH) - HPf] \mathrm{d}\mu$. By Young's inequality $ab \leq a \log a - a + e^b$, $a > 0$, $b \in \mathbb{R}$, we get that for every $\lambda > 0$,

$$P(fH) = \lambda P\Big(\frac{f}{\lambda} H\Big) \leq P(f \log f) - (Pf) \log \lambda - Pf + \lambda P(e^H).$$

Hence, choosing $\lambda = fe^{-H}$,

$$P(fH) - HPf \leq P(f \log f) - (Pf) \log f - Pf + fe^{-H} P(e^H).$$

The desired inequality (24) follows after integration, since for every g we have $\int Pg \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu$. ■

With the previous lemma in hand, we can easily complete the proof of the theorem.

Proof of Theorem 26. Note that the conditional expectation property yields, for every $i = 1, \ldots, m$,

$$J(f_i) = \mathcal{E}(f_i, \log f_i) = -\int f_i \, L(\log f_i) \, \mathrm{d}\mu = -\int f \, L(\log f_i) \, \mathrm{d}\mu = \mathcal{E}(f, \log f_i). \tag{25}$$

Hence

$$\sum_{i=1}^{m} c_i J(f_i) = \sum_{i=1}^{m} c_i \, \mathcal{E}(f, \log f_i) = \mathcal{E}(f, H),$$

where $H = \sum_{i=1}^{m} c_i \log f_i$. Combining Lemma 27 and **BL**-condition (3) (written for $F_i = \log f_i$, which is a function of T_i), we get

$$\begin{aligned}
\mathcal{E}(f, H) &\leq \mathcal{E}(f, \log f) + \int fe^{-H} L\big(e^H\big) \, \mathrm{d}\mu \\
&\leq J(f) + \int f \sum_i c_i \frac{1}{f_i} \, L(f_i) \, \mathrm{d}\mu \\
&= J(f) + \sum_i c_i \int L(f_i) \, \mathrm{d}\mu = J(f),
\end{aligned}$$

where we have used in the last step that $L(f_i)/f_i$ is a function of T_i and the conditional expectation property (19). ■

Superadditive inequalities for Fisher information were considered on the sphere $S^{n-1} \subset \mathbb{R}^n$ in [5] in the case of $T_i = P_{E_i}$ with the E_i for subspaces $E_i \subset \mathbb{R}^n$ satisfying $\sum_i c_i P_{E_i} \le \mathrm{Id}_{\mathbb{R}^n}$. As explained in Section 3.2.2, the **BL**-condition (3) is verified for $d_i = c_i/2$ and we recover by the previous proposition the inequality from [5]. For applications of the superadditivity of information to classical Euclidean convolution and Brascamp–Lieb inequalities, we refer to [10, 14].

In the discrete case, some examples of superadditive inequalities for Fisher information were implicitly obtained in the papers [9, 18, 19]. The goal of these papers is to prove modified log-Sobolev inequalities of the form

$$\forall f : E \to \mathbb{R}^+ \text{ with } \int f \, \mathrm{d}\mu = 1, \quad \rho_0 \int f \log f \, \mathrm{d}\mu \le \mathcal{E}(f, \log f).$$

As pointed out to us by Eric Carlen, one can extract from their proofs (which is by induction) superadditive inequalities for Fisher information, which constitute a central technical ingredient. The main examples considered in these papers are the symmetric group and slices of the discrete cube. There, the marginals are considered with respect to maps T_i, which belong to the family studied in the previous section, for which we have proved that **BL**-condition (3) holds, and for which we therefore have the desired superadditive inequalities.

References

[1] Bakry, D. "L'hypercontractivité et son utilisation en théorie des semigroupes." In *Lectures on Probability Theory (Saint-Flour, 1992)*, 1–114. Lecture Notes in Mathematics 1581. Berlin: Springer, 1994.

[2] Ball, K. M. "Volumes of Sections of Cubes and Related Problems." In *Israel Seminar on Geometric Aspects of Functional Analysis*, edited by J. Lindenstrauss and V. D. Milman. Lecture Notes in Mathematics 1376. Berlin: Springer, 1989.

[3] Barthe, F. "On a reverse form of the Brascamp–Lieb inequality." *Inventiones Mathematicae* 134 (1998): 335–61.

[4] Barthe, F., and D. Cordero-Erausquin. "Inverse Brascamp–Lieb Inequalities along the Heat Equation." In *Geometric Aspects of Functional Analysis*, 65–71. Lecture Notes in Mathematics 1850. Berlin: Springer, 2004.

[5] Barthe, F., D. Cordero-Erausquin, and B. Maurey. "Entropy of spherical marginals and related inequalities." *Journal de Mathématiques Pures et Appliquées* 86 (2006): 89–99.

[6] Barthe, F., O. Guédon, S. Mendelson, and A. Naor. "A probabilistic approach to the geometry of the ℓ_p^n ball." *Annals of Probability* 33, no. 2 (2005): 480–513.

[7] Bennett, J., A. Carbery, M. Christ, and T. Tao. "The Brascamp–Lieb inequalities: finiteness, structure and extremals." *Geometric and Functional Analysis* 17, no. 5 (2008): 1343–415.

[8] Brascamp, H. J., and E. H. Lieb. "Best constants in Young's inequality, its converse and its generalization to more than three functions." *Advances in Mathematics* 20 (1976): 151–73.

[9] Bobkov, S., and P. Tetali. "Modified logarithmic Sobolev inequalities in discrete settings." *Journal of Theoretical Probability* 19, no. 2 (2006): 289–336.

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

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

[12] Carlen, E., E. Lieb, and M. Loss. "An inequality of Hadamard type for permanents." *Methods and Applications of Analysis* 13 (2006): 1–17.

[13] Carlen, E., and E. Lieb. "Brascamp–Lieb inequalities for non-commutative integration." preprint.

[14] Cordero-Erausquin, D., and M. Ledoux. "The geometry of Euclidean convolution inequalities and entropy." *Proceedings of the American Mathematical Society* 138 (2010): 2755–69.

[15] Diaconis, P., and L. Saloff-Coste. "Logarithmic Sobolev inequalities for finite Markov chains." *Annals of Applied Probability* 6, no. 3 (1996): 695–750.

[16] Ethier, S., and T. Kurtz. "Fleming–Viot processes in population genetics." *SIAM Journal on Control and Optimization* 31 (1993): 345–86.

[17] Finner, H. "A generalization of Hölder's inequality and some probability inequalities." *Annals of Probability* 20, no. 4 (1992): 1893–901.

[18] Gao, F., and J. Quastel. "Exponential decay of entropy in the random transposition and Bernoulli–Laplace models." *Annals of Applied Probability* 13, no. 4 (2003): 1591–600.

[19] Goel, S. "Modified logarithmic Sobolev inequalities for some models of random walk." *Stochastic Processes and Their Applications* 114 (2004): 51–79.

[20] Lieb, E. H. "Gaussian kernels have only Gaussian maximizers." *Inventiones Mathematicae* 102 (1990): 179–208.

[21] Ma, Z., and M. Röckner. *Introduction to the Theory of (Non-symmetric) Dirichlet Forms*. Berlin: Springer, 1992.

[22] Valdimarsson, S. "Optimizers for the Brascamp–Lieb inequality." *Israel Journal of Mathematics* 168 (2008): 253–74.