Available online at www.sciencedirect.com
SCIENCE DIRECT®
J. Math. Pures Appl. 86 (2006) 89–99
JOURNAL DE MATHÉMATIQUES PURES ET APPLIQUÉES
www.elsevier.com/locate/matpur

Entropy of spherical marginals and related inequalities ☆

F. Barthe [a,*], D. Cordero-Erausquin [b], B. Maurey [b]

[a] *Institut de Mathématiques, Laboratoire de Statistique et Probabilités, UMR C 5583, Université Paul Sabatier, 31062 Toulouse Cedex 09, France*
[b] *Laboratoire d'Analyse et de Mathématiques Appliquées, UMR 8050, Université de Marne la Vallée, 77454 Marne la Vallée Cedex 2, France*

Received 13 January 2006

Available online 5 June 2006

Abstract

We investigate relations between the entropy of a spherical density and those of its marginals, together with spherical convolution type inequalities. We extend results by Carlen, Lieb and Loss to more general configurations. Our argument involves a corresponding superadditivity result for Fisher information, which has a clear geometric meaning.

Résumé

Nous étudions les relations entre l'entropie d'une densité sphérique et celles de ses marginales, ainsi que des inégalités de type convolution sur les sphères. Nous étendons des résultats de Carlen, Lieb et Loss à des configurations plus générales. Notre argument passe par un résultat analogue de sur-additivité de l'information de Fisher, qui possède une interprétation géométrique simple.

Keywords: Entropy; Marginals; Spherical inequalities

1. Introduction

The present manuscript is motivated by the recent far-reaching work of Carlen, Lieb and Loss [5] on convolution-type inequalities on the sphere $S^{n-1} \subset \mathbb{R}^n$. These authors also studied such relations on permutation groups [6]. The motivations for pushing forward the investigation on spherical inequalities are both mathematical and physical. We would like to provide a new geometric insight on the problem and give possible extensions to the study of a system of particles with given kinetic energy. In particular, we aim at understanding how information on each particle is reflected on the whole system.

Given a random vector $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ with probability density f its entropy, defined as $H(X) = H(f) = -\int_{\mathbb{R}^n} f \log f$ is know to be subadditive, namely $H(X) \leqslant \sum_{i=1}^n H(X_i)$ (see, e.g., [7]). This classical fact can be

☆ Research supported in part by the European Network PHD, MCRN-51193. The first and second-named authors gratefully acknowledge the hospitality of the Erwin Schrödinger Institute, Vienna, where part of their contribution to this work was completed.
* Corresponding author.
E-mail addresses: barthe@math.ups-tlse.fr (F. Barthe), dario.cordero@univ-mlv.fr (D. Cordero-Erausquin), bernard.maurey@univ-mlv.fr (B. Maurey).

0021-7824/$ – see front matter
doi:10.1016/j.matpur.2006.04.003

reformulated in Gauss space as follows. Let γ_n be the standard Gaussian measure on $\mathbb{R}^n$ and let g be a probability density on $\mathbb{R}^n$ with respect to γ_n, $\int g\,\mathrm{d}\gamma_n = 1$. Let g_i denote the ith marginal of this density, given by:

$$g_i(x) = \int g(y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_n)\,\mathrm{d}\gamma_n(y),$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. This marginal can be seen either as a function of the one real variable x_i or as a function of $x \in \mathbb{R}^n$ which depends only on the ith variable (g_i is again a probability density with respect to γ_n). Then, denoting by $S(h) = \int_{\mathbb{R}^n} h \log h\,\mathrm{d}\gamma_n$ the (relative) entropy of a probability density h with respect to γ_n, one has

$$\sum_{i=1}^{n} S(g_i) \leqslant S(g).$$

Motivated by the study of particle systems which preserve kinetic energy, Carlen, Lieb and Loss established a corresponding fact for probability densities on spheres [5]. We denote by σ the uniform probability measure on the Euclidean sphere S^{n-1}. For a probability density f on S^{n-1}, let f_i denote its marginal on the ith coordinate. More precisely, this is the function so that $f_i(x)$ only depends on the ith coordinate x_i of $x \in S^{n-1}$ and for every bounded measurable φ defined on $[-1, 1]$,

$$\int_{S^{n-1}} \varphi(x_i) f(x)\,\mathrm{d}\sigma(x) = \int_{S^{n-1}} \varphi(x_i) f_i(x)\,\mathrm{d}\sigma(x).$$

The above mentioned authors established the inequality:

$$\sum_{i=1}^{n} S(f_i) \leqslant 2S(f),$$

where

$$S(g) = \int_{S^{n-1}} g \log g\,\mathrm{d}\sigma$$

is the spherical entropy of a probability density g (relative to σ). The surprise here is that the factor 2 has to be there, whatever the dimension. Indeed, the uniform measure on spheres of dimension n and radius $\sqrt{n}$ is in many aspects close to a Gaussian measure. But in this particular situation, the Gaussian inequality cannot be recovered as a limit of the spherical one. Somehow the effects on entropy of the coordinate dependence

$$|v_1|^2 + \cdots + |v_n|^2 = 1$$

of points $v = (v_1, \ldots, v_n)$ on a sphere do not vanish as the dimension increases. Carlen, Lieb and Loss deduce their entropy estimate from a beautiful new version of the Brascamp–Lieb [3] inequalities for spheres which can be viewed as a spherical convolution type inequality. It asserts that for every nonnegative measurable functions f_i on $[-1, 1]$ one has:

$$\int_{S^{n-1}} \prod_{i=1}^{n} \sqrt{f_i(x_i)}\,\mathrm{d}\sigma(x) \leqslant \prod_{i=1}^{n} \left(\int_{S^{n-1}} f_i(x_i)\,\mathrm{d}\sigma(x) \right)^{1/2}.$$

In this paper we extend the work of Carlen, Lieb and Loss in several ways. In particular, we give another approach to the entropy inequality on the sphere. It is very natural in the context of information theory since it relies on the Fisher information:

$$I(f) = \int_{S^{n-1}} \frac{|\nabla f|^2}{f}\,\mathrm{d}\sigma. \tag{1}$$

We show that this quantity satisfies the analogous inequality:

$$\sum_{i=1}^{n} I(f_i) \leqslant 2I(f).$$

The argument is inspired by Carlen's proof of strict superadditivity of Fisher information in $\mathbb{R}^n$ [4]. The core of the proof is a geometric observation on projections of the Euclidean basis onto tangent spaces. This new insight allows us to consider more general configurations, where the marginals are taken with respect to sets of coordinates which decompose the identity map. In particular this applies to systems of particles in $\mathbb{R}^d$ with fixed kinetic energy and total momentum. We also prove multidimensional extensions of the spherical Brascamp–Lieb inequality. For the latter, we use the heat semi-group as in the paper by Carlen, Lieb and Loss, although in a different way (motivated again by geometry).

2. Notation and geometric parameters

In this section we start with some notation and background on spherical functions depending only on a subspace of $\mathbb{R}^n$ and on marginals of a probability density. We then introduce the notions of *geometric configuration* and of *configuration constant* which will be central for the functional inequalities under study.

Throughout the paper, given a subspace $E \subset \mathbb{R}^n$, we will denote by P_E the orthogonal projection onto E:

$$P_E x \in E \quad \text{and} \quad x - P_E x \in E^\perp. \tag{2}$$

When $E = 0$, $P_E = 0$.

Let E be a vector subspace of $\mathbb{R}^n$. A function on the sphere $x \to h(x)$ *depends only on the projection onto* E, or in short, *depends only on* E, if it is of the form $h = k \circ P_E$ for some function k on the unit ball of E. An equivalent condition is that

$$h = h \circ R$$

on S^{n-1} for every (direct) isometry R fixing E. We shall denote this set of isometries as

$$SO\big(E^\perp\big) = \{R \in SO_n;\ Rx = x,\ \text{for all } x \in E\}.$$

Let f be a probability density on S^{n-1}. Its marginal on E is the function $f_E : S^{n-1} \to \mathbb{R}^+$ depending only on E, such that for all bounded measurable functions φ also depending only on E it holds

$$\int f(x)\varphi(x)\,\mathrm{d}\sigma(x) = \int f_E(x)\varphi(x)\,\mathrm{d}\sigma(x).$$

In probabilistic language, if X is a random vector on S^{n-1} with distribution $f\,\mathrm{d}\sigma$, then the law of $P_E X$ has density f_E with respect to the projected measure $P_E\sigma$. Changing variables $x = Ry$ in the former integral for $R \in SO(E^\perp)$ and using the invariance of φ and the measure σ, we get

$$\int f(Ry)\varphi(y)\,\mathrm{d}\sigma(y) = \int f_E(x)\varphi(x)\,\mathrm{d}\sigma(x).$$

Integrating in R with respect to the Haar measure μ on $SO(E^\perp)$ provides a simple representation for the marginal, namely

$$f_E(x) = \int_{SO(E^\perp)} f(Rx)\,\mathrm{d}\mu(R). \tag{3}$$

Given a element $x \in S^{n-1}$ of the sphere, the vector space $x^\perp$ is the tangent hyperplane to S^{n-1} at x. We will use the following notation for a subspace $E \subset \mathbb{R}^n$ and an element x of the sphere:

$$E(x) := P_{x^\perp} E.$$

Definition 1 *(Geometric configuration and configuration constant).* A *geometric configuration* $\mathcal{E}$ in $\mathbb{R}^n$ is a collection $(E_1, \ldots, E_k)$ of (non-zero) subspaces of $\mathbb{R}^n$ together with positive numbers $(c_1, \ldots, c_k)$ for some $k \geqslant 1$. To a geometric configuration $\mathcal{E} = ((E_i)_{\leqslant k}, (c_i)_{i \leqslant k})$ we associate a *configuration constant* $C_n(\mathcal{E}) > 0$ defined by:

$$C_n(\mathcal{E}) := \sup_{x,y \in S^{n-1},\ \langle x,y\rangle = 0} \sum_{i=1}^{k} c_i |P_{E_i(x)} y|^2. \tag{4}$$

Note that $P_{E_i(x)} = P_{P_{x^\perp} E_i}$. We will see the configuration constant as the best possible constant in the inequality

$$\forall x \in S^{n-1},\ \forall y \in x^\perp, \quad \sum_{i=1}^{k} c_i |P_{E_i(x)} y|^2 \leqslant C_n(\mathcal{E})|y|^2. \tag{5}$$

This constant will appear when comparing information theoretic quantities of a density with the ones of its marginals on the directions of the configuration. We make no a priori hypothesis on the collection of subspaces E_i. In the last section, we shall see how to estimate the constant $C_n(\mathcal{E})$ in some given geometric situations.

3. Functional inequalities

We will first prove an estimate for the Fisher information of a spherical marginal and then use this estimate in order to prove superadditivity of Fisher information and (relative) entropy for a given geometric configuration. We will also exhibit a multidimensional Brascamp–Lieb inequality associated with a geometric configuration.

A probability density is said to have finite Fisher information if $\sqrt{f} \in W^{1,2}(S^{n-1})$ and (1) is finite (the integration is on the set where $f > 0$). The gradient of a function f defined on the sphere is, of course, the spherical gradient $\nabla f(x) \in x^\perp$ for $x \in S^{n-1}$, which is also the usual gradient at x of the function $\tilde{f}$ defined by $\tilde{f}(y) = f(y/|y|)$ for $y \neq 0$. The representation formula (3) allows us to estimate the Fisher information of the marginal f_E of a probability density f.

Proposition 1. *Let E be a vector subspace of $\mathbb{R}^n$, $n \geqslant 2$. Let f be a probability density on S^{n-1} with finite Fisher information. Then the Fisher information of its marginal f_E on E verifies*:

$$I(f_E) \leqslant \int_{S^{n-1}} \frac{|P_{E(x)} \nabla f(x)|^2}{f(x)} \, \mathrm{d}\sigma(x). \tag{6}$$

Proof. We will assume that f is a continuously differentiable probability density on the sphere bounded away from zero ($f \geqslant \varepsilon$), hence with finite Fisher information. The general case follows by approximation. The marginal

$$f_E(x) = \int_{SO(E^\perp)} f(Rx) \, \mathrm{d}\mu(R)$$

is differentiable with gradient given by:

$$\nabla f_E(x) = \int_{SO(E^\perp)} {}^t R \nabla f(Rx) \, \mathrm{d}\mu(R) = \int_{SO(E^\perp)} R^{-1} \nabla f(Rx) \, \mathrm{d}\mu(R).$$

Since $f_E(x)$ only depends on $P_E x$, necessarily

$$\nabla f_E(x) \in P_{x^\perp} E = E(x). \tag{7}$$

Indeed differentiating $f_E(x) = \varphi(P_E x)$ yields for $v \in x^\perp$:

$$\langle \nabla f_E(x), v \rangle = \langle \nabla \varphi(P_E x), P_E v \rangle = \langle P_E \nabla \varphi(P_E x), v \rangle = \langle P_{x^\perp} P_E \nabla \varphi(P_E x), v \rangle,$$

where we have used $v = P_{x^\perp} v$. Therefore

$$\nabla f_E(x) = P_{E(x)} \nabla f_E(x) = \int_{SO(E^\perp)} P_{E(x)} R^{-1} \nabla f(Rx) \, \mathrm{d}\mu(R) = \int_{SO(E^\perp)} R^{-1} P_{E(Rx)} \nabla f(Rx) \, \mathrm{d}\mu(R).$$

The last equality follows from elementary calculations: recall that for a subspace $F \subset \mathbb{R}^n$ and an isometry $O \in O_n$ we have $P_F = O^{-1} P_{OF} O$. Thus

$$P_{E(x)} R^{-1} = P_{P_{x^\perp} E} R^{-1} = R^{-1} P_{R P_{x^\perp} E}$$

and $RP_{x^\perp}E = P_{R(x^\perp)}RE = P_{(Rx)^\perp}RE$. Combining these facts for $R \in SO(E^\perp)$, for which $RE = E$, we get $P_{E(x)}R^{-1} = R^{-1}P_{P_{(Rx)^\perp}E} = R^{-1}P_{E(Rx)}$ as claimed. The above expression of ∇f_E may be combined with the Cauchy–Schwarz inequality:

$$\left|\nabla f_E(x)\right|^2 \leqslant \left(\int_{SO(E^\perp)} \left|P_{E(Rx)}\nabla f(Rx)\right| \mathrm{d}\mu(R)\right)^2 \leqslant \int_{SO(E^\perp)} \frac{|P_{E(Rx)}\nabla f(Rx)|^2}{f(Rx)} \mathrm{d}\mu(R) \cdot \int_{SO(E^\perp)} f(Rx)\,\mathrm{d}\mu(R).$$

The last integral in the above formula is exactly $f_E(x)$. Finally

$$I(f_E) = \int_{S^{n-1}} \frac{|\nabla f_E(x)|^2}{f_E(x)} \mathrm{d}\sigma(x) \leqslant \int_{S^{n-1}} \int_{SO(E^\perp)} \frac{|P_{E(Rx)}\nabla f(Rx)|^2}{f(Rx)} \mathrm{d}\mu(R)\,\mathrm{d}\sigma(x) = \int_{S^{n-1}} \frac{|P_{E(x)}\nabla f(x)|^2}{f(x)} \mathrm{d}\sigma(x),$$

by Fubini and the rotational invariance of σ. □

We can now turn to the superadditivity of information and entropy.

Theorem 2 *(Superadditivity of information and entropy). Let $\mathcal{E} = ((E_i)_{i\leqslant k}, (c_i)_{i\leqslant k})$ be a geometric configuration of $\mathbb{R}^n$ $(n \geqslant 2)$ with configuration constant $C_n(\mathcal{E})$. Then for any probability density f on S^{n-1} with finite Fisher information, we have*:

$$\sum_{i=1}^k c_i I(f_{E_i}) \leqslant C_n(\mathcal{E}) I(f),$$

and for a probability density f on S^{n-1} with finite entropy,

$$\sum_{i=1}^k c_i S(f_{E_i}) \leqslant C_n(\mathcal{E}) S(f).$$

Proof. By definition (5) of $C_n(\mathcal{E})$ we have, for every $x \in S^{n-1}$,

$$\sum_{i=1}^k c_i \left|P_{E_i(x)}\nabla f(x)\right|^2 \leqslant C_n(\mathcal{E})\left|\nabla f(x)\right|^2.$$

Integrating over the sphere gives, in view of (6), the inequality for the Fisher information.

It is classical that the derivative of the entropy function along the heat flow is related to the Fisher information. This classical fact allows us to deduce the entropy inequality from the one we just proved for Fisher information. More precisely if $(P_t)_{t\geqslant 0} = (\mathrm{e}^{t\Delta})_{t\geqslant 0}$ denotes the heat semigroup on S^{n-1}, then for any probability density f we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{S^{n-1}} P_t f \log P_t f \,\mathrm{d}\sigma = \int_{S^{n-1}} (1 + \log P_t f)\Delta P_t f \,\mathrm{d}\sigma = -\int_{S^{n-1}} \langle \nabla \log P_t f, \nabla P_t f\rangle \,\mathrm{d}\sigma = -I(P_t f).$$

As a consequence,

$$S(f) = -\big(S(1) - S(f)\big) = -\int_0^{+\infty} \frac{\mathrm{d}}{\mathrm{d}t} S(P_t f)\,\mathrm{d}t = \int_0^{\infty} I(P_t f)\,\mathrm{d}t.$$

We have omitted here the technical arguments supporting this calculation. They rely on the well-understood regularizing properties of the heat equation. Let us also observe that the semigroup commutes with marginals: if $E \subset \mathbb{R}^n$ is a vector subspace, then

$$f_E = \int_{SO(E^\perp)} f \circ R \,\mathrm{d}\mu(R),$$

and since P_t commutes with rotations

$$P_t(f_E) = \int_{SO(E^\perp)} P_t(f \circ R)\,\mathrm{d}\mu(R) = \int_{SO(E^\perp)} (P_t f) \circ R\,\mathrm{d}\mu(R) = (P_t f)_E.$$

Applying the inequality for Fisher information yields

$$C_n(\mathcal{E}) I(P_t f) \geqslant \sum_{i=1}^k c_i I\big((P_t f)_{E_i}\big) = \sum_{i=1}^k c_i I\big(P_t(f_{E_i})\big),$$

and integrating in the time variable we get $C_n(\mathcal{E}) S(f) \geqslant \sum_{i=1}^k c_i S(f_{E_i})$. □

We end this section with a multidimensional version of the spherical Brascamp–Lieb inequality obtained in [5].

Theorem 3 *(Spherical Brascamp–Lieb inequality). Let $\mathcal{E} = ((E_i)_{i\leqslant k}, (c_i)_{i\leqslant k})$ be a geometric configuration of $\mathbb{R}^n$ $(n \geqslant 2)$ with configuration constant $C_n(\mathcal{E})$ and set*

$$d_i = \frac{c_i}{C_n(\mathcal{E})}.$$

If $f_1, \ldots, f_k : S^{n-1} \to \mathbb{R}_+$ are functions such that each f_i depends only on E_i, then

$$\int_{S^{n-1}} \prod_{i=1}^k f_i(x)^{d_i}\,\mathrm{d}\sigma(x) \leqslant \prod_{i=1}^k \left(\int_{S^{n-1}} f_i(x)\,\mathrm{d}\sigma(x) \right)^{d_i}.$$

Proof. We will use the heat semi-group as did Carlen, Lieb and Loss, although in a rather different way. We shall need the following dual characterization of the configuration constant:

$$\forall x \in S^{n-1},\ \forall y_1, \ldots, y_n \in x^\perp, \quad \left| \sum_{i=1}^k c_i P_{E_i(x)} y_i \right|^2 \leqslant C_n(\mathcal{E}) \sum_{i=1}^k c_i |y_i|^2. \tag{8}$$

Indeed, it is easily checked that, fixing x, the inequalities (5) and (8) on $x^\perp$ are dual one to another.

Let $f_1, \ldots, f_k$ be functions as in the theorem and let $f_i(t,x) := P_t(f_i)(x) = (\mathrm{e}^{t\Delta} f_i)(x)$ denote the evolution of f_i along the heat semi-group:

$$f_i(0,x) = f_i(x) \quad \text{and} \quad \partial_t f_i(t,x) = \Delta_x f_i(t,x).$$

We can assume that the f_i's are C^∞-smooth, which guarantees smoothness of $(t,x) \to f_i(t,x)$ on $[0,+\infty) \times S^{n-1}$ (this simplifies nonessential continuity issues at $t = 0$) and that the f_i's are bounded away from 0 (this property always holds for f_t when $t > 0$). The function $f_i(t,\cdot)$ depends also only on E_i since the invariance of the semi-group (or equivalently of the spherical Laplacian) under rotations shows that

$$f_i(t, Rx) = P_t(f_i)(Rx) = P_t(f_i \circ R)(x) = P_t(f_i)(x)$$

for every $R \in SO(E_i^\perp)$. Introduce the smooth function:

$$\alpha(t) := \int_{S^{n-1}} \prod_{i=1}^k f_i(t,x)^{d_i}\,\mathrm{d}\sigma(x).$$

Setting $h_i(t,x) = \log f_i(t,x)$ and $H(t,x) := \sum_{i=1}^k d_i h_i(t,x)$, we find:

$$\partial_t h_i = \Delta_x h_i + |\nabla_x h_i|^2$$

and

$$\alpha'(t) = \int_{S^{n-1}} \left(\sum_{i=1}^k d_i \big[\Delta_x h_i + |\nabla_x h_i|^2\big] \right) \mathrm{e}^H\,\mathrm{d}\sigma.$$

Integration by parts of $\Delta_x h_i \mathrm{e}^H$ gives:

$$\alpha'(t) = \int_{S^{n-1}} \left(-\sum_{i,j \leqslant k} d_i d_j \nabla_x h_i \cdot \nabla_x h_j + \sum_{i=1}^{k} d_i |\nabla_x h_i|^2 \right) \mathrm{e}^H \, \mathrm{d}\sigma$$
$$= \int_{S^{n-1}} \left(-\left| \sum_{i}^{k} d_i \nabla_x h_i(t,x) \right|^2 + \sum_{i=1}^{k} d_i \left| \nabla_x h_i(t,x) \right|^2 \right) \mathrm{e}^{H(t,x)} \, \mathrm{d}\sigma(x).$$

Since $h_i(t,\cdot)$ depends only on E_i, the argument after (7) yields $\nabla_x h_i(t,x) \in E_i(x)$. Hence we can use (8) in the form

$$\left| \sum_{i=1}^{k} d_i \nabla_x h_i(t,x) \right|^2 \leqslant \sum_{i=1}^{k} d_i \left| \nabla_x h_i(t,x) \right|^2.$$

This shows that $\alpha'(t) \geqslant 0$. Passing to the limit when $t \to +\infty$ provides the claimed inequality $\alpha(0) \leqslant \alpha(+\infty)$. □

4. Geometric parameters and applications

In this section we show how to estimate the configuration constant $C_n(\mathcal{E})$ in some relevant geometric or physical situations and how to recover and extend the results by Carlen, Lieb and Loss. The idea of considering decompositions of the identity is inspired by the geometric version of the Euclidean Brascamp–Lieb inequality put forward by Ball in [1].

Lemma 4. *Let $k, n \geqslant 1$ be integers and $c_1, \ldots, c_k > 0$. Assume that $E_1, \ldots, E_k$ are subspaces of $\mathbb{R}^n$ for which the following decomposition of the identity map holds*:

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

Let $S \subset \mathbb{R}^n$ be a subspace of codimension m, and for each $i \leqslant k$ denote by Q_i the orthogonal projection onto $P_S E_i$. Then for all $y \in S$:

$$|y|^2 \leqslant \sum_{i=1}^{k} c_i |Q_i y|^2 \leqslant (1+m)|y|^2. \tag{9}$$

In other words, if we consider the Q_i's as maps on S,

$$\mathrm{Id}_S \leqslant \sum_{i=1}^{k} c_i Q_i \leqslant (1+m)\mathrm{Id}_S.$$

Proof. We start with the lower bound, which is easy. Indeed for any subspace $E \subset \mathbb{R}^n$ and for all $y \in S$ we have $|P_E y| \leqslant |Q_E y|$ where Q_E denotes the orthogonal projection onto $P_S E$. To see this we first use the description of the projection of a point x onto a subspace as the closest point to x:

$$|y - P_E y| = \inf_{e \in E} |y - e| \quad \text{and} \quad |y - Q_E y| = \inf_{f \in P_S E} |y - f| = \inf_{e \in E} |y - P_S e|.$$

Next for $y \in S$, we have $|y - e| \geqslant |P_S(y - e)| = |y - P_S e|$ and therefore:

$$|y - Q_E y| \leqslant |y - P_E y|.$$

The orthogonality relations (2) for Q_E and P_E then yield the claimed inequality $|P_E y| \leqslant |Q_E y|$. The decomposition of the identity in $\mathbb{R}^n$ then gives for $y \in S$

$$|y|^2 = \sum_{i=1}^{k} c_i |P_{E_i} y|^2 \leqslant \sum_{i=1}^{k} c_i |Q_i y|^2.$$

Now, we address the upper bound. Let $y \in S$. Assume that $Q_i y \neq 0$. We set $v_{i,1} := \frac{Q_i y}{|Q_i y|} \in P_S E_i$ and we choose a unit vector $e_{i,1} \in E_i$ such that there exists $\lambda > 0$ with $P_S e_{i,1} = \lambda v_{i,1}$. Next we choose an orthonormal basis of E_i starting with $e_{i,1}$ and denoted by $(e_{i,j})_{j \leqslant n_i}$ where $n_i = \dim(E_i)$. In the case when $Q_i y = 0$ we just choose any orthonormal basis in E_i and give the same name to it.

The decomposition of the identity of $\mathbb{R}^n$ then reads as

$$\mathrm{Id}_{\mathbb{R}^n} = \sum_{i=1}^{k} c_i P_{E_i} = \sum_{i=1}^{k} c_i \sum_{j=1}^{n_i} e_{i,j} \otimes e_{i,j}.$$

In order to avoid heavy notations we write P for P_S in the remaining of this proof. For $j \leqslant n_i$, we set $d_{i,j} := c_i |P e_{i,j}|^2 \leqslant c_i$. We also define $v_{i,j} = \frac{P e_{i,j}}{|P e_{i,j}|}$ if $|P e_{i,j}| \neq 0$ and $v_{i,j} = 0$ otherwise (note that this definition is consistent with the one of $v_{i,1}$). Multiplying the above decomposition by P on the left and on the right yields:

$$P = \sum_{i=1}^{k} c_i \sum_{j=1}^{n_i} P e_{i,j} \otimes P e_{i,j} = \sum_{i=1}^{k} c_i \sum_{j=1}^{n_i} |P e_{i,j}|^2 v_{i,j} \otimes v_{i,j} = \sum_{i \leqslant k;\, j \leqslant n_i} d_{i,j} v_{i,j} \otimes v_{i,j}.$$

Taking traces, the decomposition of $\mathrm{Id}_{\mathbb{R}^n}$ and the one of P give:

$$n = \sum_{i=1}^{k} c_i n_i = \sum_{i \leqslant k;\, j \leqslant n_i} c_i \quad \text{and} \quad \mathrm{Trace}(P) = \dim(S) = \sum_{i \leqslant k;\, j \leqslant n_i} d_{i,j}.$$

The definition of $v_{i,1}$ ensures that $|Q_i y| = \langle y, v_{i,1} \rangle$. Therefore

$$\begin{aligned} \sum_{i=1}^{k} c_i |Q_i y|^2 &= \sum_{i=1}^{k} c_i \langle y, v_{i,1} \rangle^2 = \sum_{i=1}^{k} d_{i,1} \langle y, v_{i,1} \rangle^2 + \sum_{i=1}^{k} (c_i - d_{i,1}) \langle y, v_{i,1} \rangle^2 \\ &\leqslant \sum_{i \leqslant k;\, j \leqslant n_i} d_{i,j} \langle y, v_{i,j} \rangle^2 + \sum_{i \leqslant k;\, j \leqslant n_i} (c_i - d_{i,j}) |y|^2 = |y|^2 + \big(n - \dim(S)\big) |y|^2, \end{aligned}$$

where the last equality comes from the decomposition of P and the trace relations. □

Remark. Equality conditions in (9) may be derived from our arguments. They show that the upper bound involving $(1 + m)$ is achieved only in specific situations. This constant is not optimal in general.

If we apply the previous result to the case $S = x^{\perp}$ $(m = 1)$ for some $x \in S^{n-1}$, since $Q_i = P_{E_i(x)}$ we obtain a bound on the configuration constant (4) for a decomposition of the identity.

Corollary 5 *(Configuration constant for a decomposition of the identity). Let $\mathcal{E} = ((E_i)_{i \leqslant k}, (c_i)_{i \leqslant k})$ be a geometric configuration of $\mathbb{R}^n$ which satisfies*

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

then $1 \leqslant C_n(\mathcal{E}) \leqslant 2$.

If one only assumes $\mathrm{Id}_{\mathbb{R}^n} \geqslant \sum_{i=1}^{k} c_i P_{E_i}$, it is still true that $C_n(\mathcal{E}) \leqslant 2$, as readily checked by writing $\mathrm{Id}_{\mathbb{R}^n} - \sum_{i=1}^{k} c_i P_{E_i}$ as a sum of rank one orthogonal projections and applying the above corollary. We conclude this section with two physically relevant examples.

Example 1. Let us consider the case when $\mathbb{R}^n$ decomposes into a direct orthogonal sum of subspaces:

$$\mathbb{R}^n = E_1 \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} E_k.$$

Let $\mathcal{E}$ be the geometric configuration given by the E_i's and $c_i = 1$ for $i = 1, \dots, k$. Since $\sum_i P_{E_i} = \mathrm{Id}_{\mathbb{R}^n}$ we know that $C_n(\mathcal{E}) \leqslant 2$. When $k \geqslant 2$ it is worth noticing that

$$C_n(\mathcal{E}) = 2.$$

To see this, choose two unit vectors $e_1 \in E_1$ and $e_2 \in E_2$ and an angle $\theta \notin \frac{\pi}{2}\mathbb{Z}$. Set $x = (\cos\theta)e_1 + (\sin\theta)e_2$ and $y = (\sin\theta)e_1 - (\cos\theta)e_2 \in x^\perp$. Direct calculations show that

$$P_{x^\perp}e_1 = e_1 - \langle x, e_1\rangle x = (\sin\theta)y, \qquad P_{x^\perp}e_2 = e_2 - \langle x, e_2\rangle x = -(\cos\theta)y.$$

Since $\cos\theta\sin\theta \neq 0$ it follows that $y \in P_{x^\perp}E_1 \cap P_{x^\perp}E_2 = E_1(x) \cap E_2(x)$. In particular $P_{E_1(x)}y = P_{E_2(x)}y = y$. For $j \geqslant 3$, $E_j \subset x^\perp$ and therefore $E_j(x) = E_j$. The latter spaces are orthogonal to y, so $P_{E_j(x)}y = 0$ for $j \geqslant 3$. As a conclusion y is a unit vector of $x^\perp$ with

$$\sum_{i=1}^{k} |P_{E_i(x)}y|^2 = 2|y|^2.$$

Note that the constant 2 can also be attained in the limit, for instance when $x \to E_1$ and $y \in E_1(x)$.

It is physically relevant to assume all the E_i's have the same dimension d (the physical dimension) and $n = kd$. Indeed, let us consider a system of $k \geqslant 2$ particles of equal mass, say one, in $\mathbb{R}^d$, with total kinetic energy 1. Their speeds $(V_1, \dots, V_k)$, which satisfy:

$$|V_1|^2 + \cdots + |V_k|^2 = 1, \tag{10}$$

are assumed to be distributed according to some probability density $f\,\mathrm{d}\sigma$ where σ is uniform on S^{kd-1}. Let $R_i^d := \{(x_1, \dots, x_k) \in (\mathbb{R}^d)^k;\ x_j = 0 \text{ for } j \neq i\}$ be the copy of $\mathbb{R}^d$ in the ith place. Denoting by f_i the marginal of f on R_i^d, our results imply the multidimensional version of the entropy inequality of Carlen, Lieb and Loss:

$$\sum_{i=1}^{k} S(f_i) \leqslant 2S(f),$$

or identifying random variables and their densities:

$$\sum_{i=1}^{k} S(V_i) \leqslant 2S(V_1, \dots, V_k).$$

We may also state the corresponding multidimensional versions of the spherical Brascamp–Lieb inequality.

Remark. As shown in [5] the factor 2 in the above inequalities is best possible. It can however be improved under additional assumptions on the functions f_i. Let us illustrate this in the case of functions depending on one variable. With the abuse of notation $f_i(x) = f_i(x_i)$, and provided the functions f_i are *even* and *non-increasing on* $[0, 1]$ then one has:

$$\int_{S^{n-1}} \prod_{i=1}^{n} f_i(x_i)\,\mathrm{d}\sigma(x) \leqslant \prod_{i=1}^{n} \int_{S^{n-1}} f_i(x_i)\,\mathrm{d}\sigma(x) = 1.$$

See [8] and the references therein for this functional version of subindependence of coordinates, put forward in a different form in [2]. In this case, if a probability density f has marginals f_i verifying the above assumptions, the variational characterization of the entropy and the definition of f_i as marginals of f yield:

$$\begin{aligned} S(f) &= \sup\left\{ \int_{S^{n-1}} f \log h\,\mathrm{d}\sigma;\ h \geqslant 0 \text{ and } \int_{S^{n-1}} h\,\mathrm{d}\sigma \leqslant 1 \right\} \geqslant \int_{S^{n-1}} f \log\left(\prod_{i=1}^{n} f_i\right)\mathrm{d}\sigma \\ &= \sum_{i=1}^{n} \int_{S^{n-1}} f \log f_i\,\mathrm{d}\sigma = \sum_{i=1}^{n} \int_{S^{n-1}} f_i \log f_i\,\mathrm{d}\sigma = \sum_{i=1}^{n} S(f_i). \end{aligned}$$

Example 2. Let us now consider a system of k particles of unit mass in $\mathbb{R}^d$ with speeds $(V_1, \dots, V_k)$, having as before, kinetic energy one (10), and now also fixed momentum $m \in \mathbb{R}^d$

$$V_1 + \cdots + V_k = m,$$

with $|m| \in [0, \sqrt{k})$. If $d = 1$ we require $k \geqslant 3$, otherwise $k \geqslant 2$ is enough. The speeds of these particles are thus forced to stay in a sphere of radius $\sqrt{1 - |m|^2/k}$ with center $(m/k, \dots, m/k)$ in the affine space going through this point with direction:

$$F := \left\{ (x_1, \dots, x_k) \in \left(\mathbb{R}^d\right)^k;\ \sum_{i=1}^k x_i = 0 \right\}.$$

Our hypotheses ensure that $\dim(F) = (k-1)d \geqslant 2$. The decomposition of the identity of $\mathbb{R}^{kd}$, $\mathrm{Id}_{\mathbb{R}^{kd}} = \sum_{i=1}^k P_{R_i^d}$, can be projected onto the space F as

$$P_F = \sum_{i=1}^k P_F P_{R_i^d} P_F.$$

It is easily checked that $P_F P_{R_i^d} P_F = \frac{k-1}{k} P_{E_i}$ where

$$E_i := P_F R_i^d = \left\{ \left(-x, \dots, -x, \underbrace{(k-1)x}_{i}, -x, \dots, -x\right);\ x \in \mathbb{R}^d \right\}.$$

Thus the identity map of F verifies the decomposition:

$$\mathrm{Id}_F = \frac{k-1}{k} \sum_{i=1}^k P_{E_i}. \tag{11}$$

Denote by S_F the unit sphere in F. The vector $(V_1, \dots, V_k)$ has no density (and therefore no well defined entropy) on S^{kd-1} because the distribution is concentrated on the set

$$\mathcal{S} := \left(\frac{m}{k}, \dots, \frac{m}{k}\right) + \sqrt{1 - \frac{|m|^2}{k}} S_F.$$

Let f be the density distribution of $(V_1, \dots, V_k)$ with respect to $\sigma_{\mathcal{S}}$, the uniform probability measure on the sphere $\mathcal{S}$. We shall denote by $\widetilde{S}(h) = \int h \log h \, \mathrm{d}\sigma_{\mathcal{S}}$ the entropy of a probability density with respect to $\sigma_{\mathcal{S}}$. We may also identify $\mathcal{S}$ with S_F itself; of course, all the results stated for S^{n-1} hold for arbitrary Euclidean spheres modulo the obvious modifications. Let us now observe the following property of marginals:

$$f_{R_i^d} = f_{E_i}.$$

Indeed, a function depends only on the ith component v_i of its argument $v \in S_F$ if and only if it depends only on

$$P_{E_i} v = \frac{1}{k-1}\left(-v_i, \dots, -v_i, \underbrace{(k-1)v_i}_{i}, -v_i, \dots, -v_i\right).$$

In view of (11), our entropy estimate applied to the sphere of F then gives:

$$\sum_{i=1}^k \frac{k-1}{k} \widetilde{S}(f_{E_i}) \leqslant 2\widetilde{S}(f),$$

or equivalently in terms of random variables:

$$\sum_{i=1}^k \widetilde{S}(V_i) \leqslant \frac{2k}{k-1} \widetilde{S}(V_1, \dots, V_k).$$

We have used that the geometric configuration in F composed of the subspaces E_i and the scalars $(k-1)/k$ satisfying (11) has a configuration constant upper bounded by 2. It is actually equal to 2 as seen by considering, for $k \geqslant 3$, a unit vector $v \in \mathbb{R}^d$, together with $x = (v, -v, 0, \ldots, 0)/\sqrt{2} \in F$ and

$$y = \frac{1}{\sqrt{k(k-1)}}\left(\frac{k-2}{2}v, \frac{k-2}{2}v, -v, \ldots, -v\right) \in x^{\perp} \cap F.$$

In this case

$$\sum_{i=1}^{k} \frac{k-1}{k} |P_{E_i(x)} y|^2 = 2|y|^2.$$

Remark 3. Let us point out a slightly different approach to the latter example. Starting from $x \in S_F$, we can apply Lemma 4 to the decomposition $\mathrm{Id}_{\mathbb{R}^{kd}} = \sum_{i=1}^{k} P_{R_i^d}$ when the subspace S is chosen to be the tangent space to S_F at x, of codimension $d+1$. The lemma yields

$$\sum_{i=1}^{k} P_{E_i(x)} \leqslant (2+d)\mathrm{Id}_S,$$

where we have used that $P_S R_i^d = P_S P_F R_i^d = P_S E_i = E_i(x)$. Our arguments then imply that

$$\sum_{i=1}^{k} \widetilde{S}(V_i) \leqslant (2+d)\widetilde{S}(V_1, \ldots, V_k).$$

In our case $2 + d \geqslant 2k/(k-1)$ so this estimate is less precise than our previous one. However it is valid for arbitrary subspaces of codimension d whereas the $2k/(k-1)$-bound relies on the fact that $P_F P_{R_i^d} P_F = \frac{k-1}{k} P_{E_i}$ which is not true in general.

Acknowledgements

We thank Dominique Bakry, Eric Carlen and Michel Ledoux for discussions on related topics.

References

[1] K. Ball, Volumes of sections of cubes and related problems, in: Geometric Aspects of Functional Analysis (1987–88), in: Lecture Notes in Math., vol. 1376, Springer-Verlag, Berlin, 1989, pp. 251–260.
[2] K. Ball, I. Perissinaki, The subindependence of coordinate slabs in ℓ_p^n balls, Israel J. Math. 107 (1998) 289–299.
[3] H.J. Brascamp, E.H. Lieb, Best constants in Young's inequality, its converse and its generalization to more than three functions, Adv. Math. 20 (1976) 151–173.
[4] E.A. Carlen, Superadditivity of Fisher's information and logarithmic Sobolev inequalities, J. Funct. Anal. 101 (1) (1991) 194–211.
[5] E.A. Carlen, E.H. Lieb, M. Loss, A sharp analog of Young's inequality on S^N and related entropy inequalities, J. Geom. Anal. 14 (3) (2004) 487–520.
[6] E.A. Carlen, E.H. Lieb, M. Loss, An inequality of Hadamard type for permanents, Preprint, 2005.
[7] A. Dembo, T.M. Cover, J.A. Thomas, Information-theoretic inequalities, IEEE Trans. Inform. Theory 37 (6) (1991) 1501–1518.
[8] A. Naor, The surface measure and cone measure on the sphere of ℓ_p^n, Trans. Amer. Math. Soc., in press.