GAFA Geometric And Functional Analysis

SUBADDITIVITY OF THE ENTROPY AND ITS RELATION TO BRASCAMP–LIEB TYPE INEQUALITIES

Eric A. Carlen and Dario Cordero–Erausquin

Abstract. We prove a general duality result showing that a Brascamp–Lieb type inequality is equivalent to an inequality expressing subadditivity of the entropy, with a complete correspondence of best constants and cases of equality. This opens a new approach to the proof of Brascamp–Lieb type inequalities, via subadditivity of the entropy. We illustrate the utility of this approach by proving a general inequality expressing the subadditivity property of the entropy on $\mathbb{R}^n$, and fully determining the cases of equality. As a consequence of the duality mentioned above, we obtain a simple new proof of the classical Brascamp–Lieb inequality, and also a fully explicit determination of all of the cases of equality. We also deduce several other consequences of the general subadditivity inequality, including a generalization of Hadamard's inequality for determinants. Finally, we also prove a second duality theorem relating superadditivity of the Fisher information and a sharp convolution type inequality for the fundamental eigenvalues of Schrödinger operators. Though we focus mainly on the case of random variables in $\mathbb{R}^n$ in this paper, we discuss extensions to other settings as well.

1 Introduction

Let $(\Omega, \mathcal{S}, \mu)$ be a measure space, and let f be a probability density on $(\Omega, \mathcal{S}, \mu)$. That is, f is a non-negative integrable function on Ω with $\int_\Omega f \mathrm{d}\mu = 1$. On the convex subset of probability densities

$$\left\{ f : \int_\Omega f \ln(1 + f) \, \mathrm{d}\mu < \infty \right\}, \tag{1.1}$$

the *entropy* of f, $S(f)$, is defined by

$$S(f) = \int_\Omega f(x) \ln f(x) \mathrm{d}\mu(x) \,.$$

With this sign convention for the entropy, the inequalities we derive are of superadditive type; however, the terminology "subadditivity of the entropy" is too well entrenched to use anything else.

Now let $p : \Omega \to \mathbb{R}$ be measurable. Let ν be a Borel measure on $\mathbb{R}$, and define $f_{(p)}$ to be the probability density on $(\mathbb{R}, \mathcal{B}, \nu)$ such that for all bounded continuous

functions ϕ on $\mathbb{R}$,

$$\int_\Omega \phi(p(x))f(x)\,\mathrm{d}\mu(x) = \int_\mathbb{R} \phi(t)f_{(p)}(t)\,\mathrm{d}\nu(t)\,. \qquad (1.2)$$

In other words, the measure $f_{(p)}\,\mathrm{d}\nu$ is the "push-forward" of the measure $f\,\mathrm{d}\mu$ under p:

$$p\#(f\,\mathrm{d}\mu) = f_{(p)}\,\mathrm{d}\nu\,.$$

The entropy of $S(f_{(p)})$ is defined just as $S(f)$ was, except with $(\mathbb{R}, \mathcal{B}, \nu)$ replacing $(\Omega, \mathcal{S}, \mu)$. We shall be concerned with the following two questions:

(1) Given m measurable functions $p_1, \ldots, p_m$ on Ω, and m non-negative numbers $c_1, \ldots, c_m$, is there a finite constant D such that

$$\sum_{j=1}^m c_j S(f_{(p_j)}) \le S(f) + D \qquad (1.3)$$

for all probability densities f with finite entropy (i.e. satisfying (1.1))?

(2) Given m measurable functions $p_1, \ldots, p_m$ on Ω, and m non-negative numbers $c_1, \ldots, c_m$, is there a finite constant D such that

$$\int_\Omega \prod_{j=1}^m f_j\big(p_j(x)\big)\,\mathrm{d}\mu(x) \le e^D \prod_{j=1}^m \left(\int_\mathbb{R} f_j^{1/c_j}(t)\,\mathrm{d}\nu(t) \right)^{c_j} \qquad (1.4)$$

for any m non-negative functions $f_1, \ldots, f_m$ on $\mathbb{R}$?

For example, consider the case that $\Omega = \mathbb{R}^n$ with its standard Euclidean structure, and μ is Lebesgue measure on $\mathbb{R}^n$, while ν is Lebesgue measure on $\mathbb{R}$. Suppose that $a_1, \ldots, a_m$ are m vectors that span $\mathbb{R}^n$, and define

$$p_j(x) = a_j \cdot x\,.$$

In this case, if we let X denote a random vector with values in $\mathbb{R}^n$ whose law has the density f, then $f_{(p_j)}$ is simply the density of the law of $a_j \cdot X$. If we define the entropy of a random variable to be the entropy of its density, provided it has one, we can rewrite this Euclidean version of (1.3) as

$$\sum_{j=1}^m c_j S(a_j \cdot X) \le S(X) + D\,.$$

In case $m = n$, $c_j = 1$ for all j, and $\{a_1, \ldots, a_n\}$ is an orthonormal basis of $\mathbb{R}^n$, then this inequality holds with $D = 0$, and is the classical subadditivity of the entropy inequality.

It is even easier to recognize (1.4) as a classical result in this setting: It becomes

$$\int_{\mathbb{R}^n} \prod_{j=1}^m f_j(a_j \cdot x)\,\mathrm{d}\mu(x) \le e^D \prod_{j=1}^m \left(\int_\mathbb{R} f_j^{1/c_j}(t)\,\mathrm{d}t \right)^{c_j}\,,$$

which is the classical Brascamp–Lieb inequality. A celebrated theorem of Brascamp and Lieb [BrL] says that the best constant e^D in this inequality can be computed by using only centered Gaussian functions as trial functions. A new proof based on optimal mass transport was given by Barthe [Ba] who also gave a characterization (depending on the vectors a_j and the constants c_j) of when the constant is finite together with a description of the optimizers in some situations. Carlen, Lieb and

Loss [CLL1] introduced a new approach to the Brascamp–Lieb inequalities based on heat flow (see also [BaC]). These authors also completed the gaps left by Barthe in the description of the optimizers. Bennett, Carbery, Christ and Tao [BeCCT1] used a similar approach to deal with the multi-dimensional versions of the Brascamp–Lieb inequality (see also [BeCCT2] for a direct approach to the finiteness of the constant e^D). The paper [CLL1] (and [BeCCT1] in the multi-dimensional setting) develops a "splitting procedure" that will prove useful in our situation too. But we shall see that working with entropy clarifies many technical points.

There are a number of other examples, besides the classical one in the Euclidean setting, where choices of (Ω, μ) and ν lead to inequalities of interest. For a second example, take $\Omega = S^{n-1}$, the unit sphere in $\mathbb{R}^n$, and let μ be the uniform probability measure on S^{n-1}. Then take ν to be the probability measure on $\mathbb{R}$ that is the law of $u \cdot x$, where u is any unit vector in $\mathbb{R}^n$, so that for all continuous functions ϕ,

$$\int_{S^{n-1}} \phi(u \cdot x) \, \mathrm{d}\mu(x) = \int_{\mathbb{R}} \phi(t) \, \mathrm{d}\nu(t) \,.$$

(By the rotational invariance of μ, this does not depend on the choice of u.) Now let $\{e_1, \ldots, e_n\}$ denote the standard orthonormal basis in $\mathbb{R}^n$, and define $p_j(x)$ on S^{n-1} by $p_j(x) = e_j \cdot x$. Then one has the optimal inequalities

$$\sum_{j=1}^{n} \frac{1}{2} S(f_{(p_j)}) \leq S(f) \,, \tag{1.5}$$

for any probability density f on (Ω, μ) with finite entropy, and

$$\int_{S^{n-1}} \prod_{j=1}^{n} f_j(e_j \cdot x) \, \mathrm{d}\mu(x) \leq \prod_{j=1}^{n} \left(\int_{S^{n-1}} f_j^2(e_j \cdot x) \, \mathrm{d}\mu(x) \right)^{1/2}$$

$$= \prod_{j=1}^{n} \left(\int_{[-1,1]} f_j^2(t) \, \mathrm{d}\nu(t) \right)^{1/2}, \tag{1.6}$$

for any n non-negative functions $f_1, \ldots, f_n$ on $[-1, 1]$. See [CLL1] for the original proofs of (1.5) and (1.6), in which (1.5) was deduced from (1.6). See [BaCM] for a different and direct proof of (1.5).

Since we are concerned in this paper with the relation between subadditivity of entropy and Brascamp–Lieb type inequalities, it is worth recalling the short argument from [CLL1] that provided the passage from (1.6) to (1.5): Let f be any probability density on S^{n-1}, and let $f_{(p_1)}, f_{(p_2)}, \ldots, f_{(p_n)}$ be its n marginals, as above.

Then define another probability density g on S^{n-1} by

$$g(x) := \frac{1}{C} \prod_{j=1}^{n} f_{(p_j)}^{1/2}(e_j \cdot x) \quad \text{where} \quad C := \int_{S^{n-1}} \prod_{j=1}^{n} f_{(p_j)}^{1/2}(e_j \cdot x) \, \mathrm{d}\mu(x) \,.$$

Then by positivity of the relative entropy (Jensen's inequality), we have

$$0 \leq \int_{S^{n-1}} \ln \left(\frac{f}{g} \right) f \, \mathrm{d}\mu = S(f) - \int_{S^{n-1}} \left(\sum_{j=1}^{n} \ln f_{(p_j)}^{1/2}(e_j \cdot x) \right) f(x) \, \mathrm{d}\mu(x) + \ln C$$

$$= S(f) - \frac{1}{2} \int_{\mathbb{R}} \left(\sum_{j=1}^{n} f_{(p_j)} \ln f_{(p_j)} \right) \mathrm{d}\nu + \ln C$$

$$= S(f) - \frac{1}{2}\sum_{j=1}^{n} S\big(f_{(p_j)}\big) + \ln C\,. \tag{1.7}$$

Finally, (1.6) implies that

$$C = \int_{S^{n-1}} \prod_{j=1}^{n} f_{(p_j)}^{1/2}(e_j \cdot x)\,\mathrm{d}\mu(x) \le \prod_{j=1}^{n}\left(\int_{S^{n-1}} f_{(p_j)}(e_j \cdot x)\,\mathrm{d}\mu(x)\right)^{1/2} = 1$$

since each $f_{(p_j)}$ is a probability density. Thus, $\ln(C) \le 0$, so that (1.6) now follows from (1.7). This argument may give the impression that (1.6) is a "stronger" inequality than (1.5), but as we shall see, this is not the case.

For a third example, take $\Omega = \mathcal{S}_n$, the symmetric group on n letters. Let μ be the uniform probability measure on $\mathcal{S}_n$, and take ν to be the uniform probability measure on $\{1, 2, \ldots, n\}$, so $\nu(i) = 1/n$ for all i. Define the functions $p_j : \Omega \to \{1, 2, \ldots, n\} \subset \mathbb{R}$ by $p_j(\sigma) = \sigma(j)$ for any permutation σ of $\{1, 2, \ldots, n\}$ Then one has the optimal inequalities

$$\sum_{j=1}^{n} \frac{1}{2} S\big(f_{(p_j)}\big) \le S(f)\,, \tag{1.8}$$

for any probability density f on (Ω, μ), and

$$\int_{\mathcal{S}_n} \prod_{j=1}^{n} f_j\big(p_j(\sigma)\big)\,\mathrm{d}\mu(\sigma) \le \prod_{j=1}^{n}\left(\int_{\mathcal{S}_n} f_j^2(p_j(\sigma))\,\mathrm{d}\mu(\sigma)\right)^{1/2}$$

$$= \prod_{j=1}^{n}\left(\sum_{i=1}^{n} f_j^2(i)\nu(i)\right)^{1/2}\,, \tag{1.9}$$

for any n non-negative functions $f_1, \ldots, f_n$ on $\{1, \ldots, n\}$. See [CLL2] for the proof of (1.9). One could then derive (1.8) using the exact same argument that was used to derive (1.5) from (1.6).

There are more examples of interesting specializations of (1.3) and (1.4). However, these examples suffice to illustrate the context in which the present work is set, and we now turn to the results. One basic result of this paper is the following:

The two questions concerning (1.3) *and* (1.4) *that were raised above are in fact one and the same: We shall prove here that the answer to one question is "yes" if and only if the answer to the other question is "yes" – with the same constant D, and with a complete correspondence of cases of equality.*

Thus, if one's goal is to prove a generalized Brascamp–Lieb type inequality, one possible route is to directly prove the corresponding generalized subadditivity of the entropy inequality. We shall demonstrate the utility of this approach by giving a simple proof of the classical Brascamp–Lieb inequality on $\mathbb{R}^n$, including a determination of all of the cases of equality, through a direct analysis of the entropy. We shall use rather elementary properties of the entropy (scaling properties and conditional entropy) together with geometric properties of the Fisher information. Moreover, the generalized subadditivity of the entropy inequality that we prove here is new (in its full generality), and is interesting in and of itself. As we shall see, it turns out to have a rich geometric structure. From the point of view of information

theory, it might also be of interest to use the converse implication and to reinterpret some Brascamp–Lieb inequalities (such as the sharp Young's convolution inequality) in terms of subadditivity inequalities for the entropy.

The rest of the paper is organized as follows. In section 2, we give the proof that (1.3) and (1.4) are dual to one another, so that once one has one inequality established with the cases of equality determined, one has the same for the other. We shall state this duality in a very general setting.

In section 3, we prove the sharp version of the general Euclidean subadditivity of the entropy inequality.

In section 4 we shall deduce some interesting consequences from this, including a generalization of Hadamard's inequality for the determinant.

The final section 5 gives another duality result showing that the superadditivity inequalities for Fisher information are dual to certain convolution type inequalities of ground state eigenvalues of Schrödinger operators. These inequalities appear to be new. They may be of some intrinsic interest, but our interest in them here is that a direct proof of the eigenvalue inequalities would yield a direct proof of Fisher information inequalities that would in turn yield entropy and Brascamp–Lieb inequalities.

2   Duality of the Brascamp–Lieb Inequality and Subadditivity of the Entropy

We show that the Brascamp–Lieb inequality is dual to the subadditivity of the entropy, so that once one has proved one of these inequalities with sharp constants, one has the other with sharp constants too. In fact, we shall see that there is an exact correspondence also for cases of equality, but in the next theorem, we focus on the constants.

We shall state the result in a more general setting than the one described in the introduction. We consider a reference measure space $(\Omega, \mathcal{S}, \mu)$ and a family of measure spaces $(M_j, \mathcal{M}_j, \nu_j)$ together with measurable functions $p_j : \Omega \to M_j$, $j \leq m$. For a probability density f on Ω (with respect to μ), the marginal $f_{(p_j)}$ is thus defined as the probability density on M_j (with respect to ν_j) such that

$$\int_{\Omega} \phi\big(p_j(x)\big) f(x) \, \mathrm{d}\mu(x) = \int_{M_j} \phi(t) f_{(p_j)}(t) \, \mathrm{d}\nu_j(t) \,. \tag{2.1}$$

for all bounded measurable functions ϕ on M_j ; accordingly the entropies are given by

$$S(f) = \int_{\Omega} f \ln(f) d\mu \quad \text{and} \quad S\big(f_{(p_j)}\big) = \int_{M_j} f_{(p_j)} \ln \big(f_{(p_j)}\big) \, \mathrm{d}\nu_j \,.$$

As explained in the introduction, we are mainly interested in the case $(M_j, \mathcal{M}_j, \nu_j) = (\mathbb{R}, \mathcal{B}, \nu)$ for all $j \leq m$, where ν is the Lebesgue measure on $\mathbb{R}$.

Theorem 2.1.   *Let $(\Omega, \mathcal{S}, \mu)$ be a measure space, $m \geq 1$ and for $j \leq m$, let $(M_j, \mathcal{M}_j, \nu_j)$ be a measure space together with a measurable function p_j from Ω to M_j. For any probability density f on Ω, let $f_{(p_j)}$ the probability density on M_j be defined as in (2.1). Finally, let $\{c_1, \ldots, c_m\}$ be any set of m non-negative numbers.*

Then for any $D \in \mathbb{R}$, the following two assertions are equivalent:

1. *For any m non-negative functions $f_j : M_j \to \mathbb{R}_+$, $j \leq m$, we have*

$$\int_\Omega \prod_{j=1}^m f_j\big(p_j(x)\big) \, \mathrm{d}\mu(x) \leq e^D \prod_{j=1}^m \left(\int_{M_j} f_j^{1/c_j}(t) \, \mathrm{d}\nu_j(t) \right)^{c_j}. \tag{2.2}$$

2. *For every probability density f on $(\Omega, \mathcal{S}, \mu)$ with finite entropy, we have*

$$\sum_{j=1}^m c_j S\big(f_{(p_j)}\big) \leq S(f) + D. \tag{2.3}$$

The proof depends an a well-known expression for the entropy as a Legendre transform: For any probability density f in Ω, and any function ϕ such that e^ϕ is integrable,

$$\int_\Omega f \ln \left(\frac{e^\phi}{f} \right) \mathrm{d}\mu = \int_\Omega f\phi \, \mathrm{d}\mu - \int_\Omega f \ln f \, \mathrm{d}\mu.$$

On the other hand, by Jensen's inequality,

$$\ln \left(\int_\Omega e^\phi \, \mathrm{d}\mu \right) \geq \int_\Omega f \ln \left(\frac{e^\phi}{f} \right) \mathrm{d}\mu.$$

Therefore,

$$\int_\Omega f \ln f \, \mathrm{d}\mu + \ln \left(\int_\Omega e^\phi \, \mathrm{d}\mu \right) \geq \int_\Omega f\phi \, \mathrm{d}\mu, \tag{2.4}$$

and there is equality if and only if e^ϕ is a constant multiple of f on the support of f. We shall use that this Legendre duality nicely combines with the operation of taking marginals.

Proof of Theorem 2.1. First, assume (2.2). Consider any probability density f on Ω, and any m functions ϕ_j on M_j, $j \leq m$. Using (2.4) with ϕ defined on Ω by

$$\phi(x) := \sum_{j=1}^m c_j \phi_j\big(p_j(x)\big) \tag{2.5}$$

and (2.1) we get

$$\int_\Omega f(x) \ln f(x) \, \mathrm{d}\mu \geq \int_\Omega f(x) \left(\sum_{j=1}^m c_j \phi_j\big(p_j(x)\big) \right) \mathrm{d}\mu - \ln \left(\int_\Omega \prod_{j=1}^m e^{c_j \phi_j(p_j(x))} \, \mathrm{d}\mu(x) \right)$$

$$= \sum_{j=1}^m c_j \int_{M_j} f_{(p_j)}(t)\phi_j(t) \, \mathrm{d}\nu_j(t) - \ln \left(\int_\Omega \prod_{j=1}^m e^{c_j \phi_j(p_j(x))} \, \mathrm{d}\mu(x) \right). \tag{2.6}$$

Then from the assumption (2.2) applied with $f_j = e^{\phi_j}$,

$$\int_\Omega \prod_{j=1}^m e^{c_j \phi_j(p_j(x))} \, \mathrm{d}\mu(x) \leq e^D \prod_{j=1}^n \left(\int_{M_j} e^{\phi_j(t)} \, \mathrm{d}\nu_j(t) \right)^{c_j}.$$

Therefore, (2.6) becomes

$$\int_\Omega f(x) \ln f(x) \, \mathrm{d}\mu(x)$$

$$\geq \sum_{j=1}^m c_j \left(\int_{M_j} f_{(p_j)}(t)\phi_j(t) \, \mathrm{d}\nu_j(t) - \ln \left(\int_{M_j} e^{\phi_j(t)} \, \mathrm{d}\nu_j(t) \right) \right) - D. \tag{2.7}$$

Now the optimal choice $\phi_j = \ln f_{(p_j)}$ leads to (2.3).

Conversely, suppose that (2.3) is true. Consider m functions ϕ_j on M_j, $j \leq m$, and define ϕ on Ω as in (2.5). Suppose that e^ϕ is integrable, and choose f to be the probability density

$$f(x) = \left(\int_\Omega e^{\phi(x)} \, d\mu(x) \right)^{-1} e^{\phi(x)}, \tag{2.8}$$

so that there is equality in (2.4). Then we have from (2.4) that

$$\ln \left(\int_\Omega \prod_{j=1}^m e^{c_j \phi_j(p_j(x))} \, d\mu(x) \right)$$

$$= \int_\Omega f(x) \left(\sum_{j=1}^n c_j \phi_j(p_j(x)) \right) d\mu(x) - \int_\Omega f(x) \ln f(x) \, d\mu(x)$$

$$= \sum_{j=1}^m c_j \int_{M_j} f_{(p_j)}(t) \phi_j(t) \, d\nu_j(t) - \int_\Omega f(x) \ln f(x) \, d\mu(x) . \tag{2.9}$$

On the other hand, (2.3) reads as

$$\int_\Omega f(x) \ln f(x) \, d\mu(x) \geq \sum_{j=1}^n c_j \int_{M_j} f_{(p_j)}(t) \ln f_{(p_j)}(t) \, d\nu_j(t) - D , \tag{2.10}$$

and so (2.9), and then (2.4) applied on (M_j, ν_j) with the probability density $f_{(p_j)}$ and the function ϕ_j for each $j \leq m$, imply

$$\ln \left(\int_\Omega \prod_{j=1}^m e^{c_j \phi_j(p_j(x))} \, d\mu \right)$$

$$\leq \sum_{j=1}^m c_j \left(\int_{M_j} f_{(p_j)}(t) \phi_j(t) \, d\nu_j(t) - \int_{M_j} f_{(p_j)}(t) \ln f_{(p_j)}(t) \, d\nu_j(t) \right) + D$$

$$\leq \sum_{j=1}^m c_j \ln \left(\int_{M_j} e^{\phi_j(t)} \, d\nu_j(t) \right) + D . \tag{2.11}$$

Exponentiating both sides, we obtain (2.2). $\qquad \square$

We next examine the relation between cases of equality in the two inequalities.

Theorem 2.2. *Using the notation of the previous theorem, suppose that f is a probability density on Ω for which equality holds in the subadditivity inequality (2.3). Then the marginals $f_{(p_1)}, f_{(p_2)}, \ldots, f_{(p_m)}$ of f yield equality in the Brascamp–Lieb inequality (2.2), and moreover, f and its marginals satisfy*

$$f = e^{-D} \prod_{j=1}^m \left(f_{(p_j)}(p_j(x)) \right)^{c_j} . \tag{2.12}$$

Conversely, suppose that $f_1, \ldots, f_m$ are m probability densities (on M_j with respect to ν_j for $j = 1, \ldots, m$, respectively) for which equality holds in the Brascamp–Lieb inequality (2.2). Then the probability density f defined on Ω by

$$f(x) := e^{-D} \prod_{j=1}^n \left(f_j(p_j(x)) \right)^{c_j}$$

yields equality in the subadditivity inequality (2.3) *and moreover* f_j *is the jth marginal of* f; *i.e.* $f_j = f_{(p_j)}$ *for* $j \leq m$.

Proof. Suppose that for some probability density f, $\sum_{i=1}^{m} c_i S(f_{(p_i)}) - S(f) = D$. Then with this f, we must have equality in the first inequality in (2.6), which comes from (2.4). By what we have said about the cases of equality in (2.4), this means that ϕ, defined in (2.5) is a constant multiple of $\ln f$. Moreover, to get equality in (2.7), we were forced to choose $\phi_j = \ln(f_{(p_j)})$. This ensures that (2.12) is true.

Furthermore, to get equality in our intermediate application of the Brascamp–Lieb inequality, we must have that $\{f_{(p_1)}, \ldots, f_{(p_n)}\}$ is a set of extremals for the Brascamp–Lieb inequality.

The other assertion follows in the same way.   □

By what we have just established, one could try to prove the classical Brascamp–Lieb inequality by first proving a general subadditivity of the entropy inequality for random variables in $\mathbb{R}^n$. We do this in the next section, and shall see that the determination of all of the cases of equality is particularly transparent via this route. While the Brascamp–Lieb inequality and subadditivity inequality are equivalent, there is an extra richness to the investigation of the cases of equality in the subadditivity inequality, as this involves statistical independence in a crucial way. Some hint of this can be seen in the following simple example, which sets the stage for the next section:

Let $m = n$, $c_j = 1$ for all j, and $\{a_1, \ldots, a_n\}$ be an orthonormal basis of $\mathbb{R}^n$. Take all reference measures to be Lebesgue measure. Then the Brascamp-Lieb inequality reduces to an equality, by Fubini's theorem, with $D = 0$, and *any* set of non-negative integrable functions $\{f_1, \ldots, f_n\}$ provides a case of equality.

On the other hand the dual inequality, is the *classical subadditivity* of the entropy inequality

$$\sum_{i=1}^{m} S(X \cdot a_i) \leq S(X) ,$$

and equality occurs *exactly* when the coordinates $\{X \cdot a_1, \ldots, X \cdot a_n\}$ form a set of *independent* random variables.

In this example, it may appear that the entropy inequality is the more complicated of the two inequalities. However, the fact that statistical independence enters the picture on the entropy side is quite helpful: We will make much use of simple entropy inequalities that are saturated only for independent random variables in our investigation of the cases of equality in the next section.

3   The General Subadditivity of the Entropy Inequality in $\mathbb{R}^n$

Let $\mathbb{R}^n$ be equipped with its standard Euclidean structure. Let X denote a random vector (or variable if $n = 1$) with values in $\mathbb{R}^n$, and suppose that X has a density f. We denote this correspondence between the random variable X and its density f by

writing $X \sim f$ and set

$$S(X) = S(f) = \int_{\mathbb{R}^n} f(x) \ln f(x) \, \mathrm{d}^n x \,.$$

Thus, in this section, we are specializing the general context of the introduction to the case in which Ω is $\mathbb{R}^n$, and μ is Lebesgue measure. We shall also take ν to be Lebesgue measure on $\mathbb{R}$.

Given a non-zero vector a on $\mathbb{R}^n$, identify a with the linear functional $a(x) = a \cdot x$. Then, if $f \sim X$ is a probability density on $\mathbb{R}^n$, $f_{(a)}$, as defined by (1.2), is the density of $a \cdot X$, that is $f_{(a)} \sim a \cdot X$, and

$$S(X \cdot a) = S(f_{(a)}) = \int_{\mathbb{R}} f_{(a)}(t) \ln f_{(a)}(t) \, \mathrm{d}t \,.$$

Note that (1.2) specializes to the requirement that for every bounded and continuous $\phi : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} \phi(x \cdot a) f(x) \, \mathrm{d}^n x = \int_{\mathbb{R}} \phi(t) f_{(a)}(t) \, \mathrm{d}t \,. \tag{3.1}$$

It follows that, for all $t \in \mathbb{R}$, $f_{(a)}(t) = \frac{1}{|a|} \int_{\{a \cdot x = t\}} f(x) \, \mathrm{d}^{n-1} x$. It is a direct consequence of (3.1) that for all $\lambda > 0$,

$$f_{(\lambda a)}(t) = \lambda^{-1} f_{(a)}(\lambda^{-1} t) \,. \tag{3.2}$$

With these preliminaries out of the way, we turn to the main question to be addressed in this section: Consider m non-zero vectors $a_1, \dots, a_m$ in $\mathbb{R}^n$, and m numbers $c_1, \dots, c_m$ with $c_j > 0$ for all j. Then, we ask

Is there a finite constant $D \in \mathbb{R}$ so that

$$\sum_{j=1}^{m} c_j S(a_j \cdot X) \leq S(X) + D \tag{3.3}$$

for all random vectors X in $\mathbb{R}^n$, and if so, what is the least such value of D, and what are the cases of equality?

In general there is no finite constant D for which (3.3) is true for all X. There are some simple requirements on $\{a_1, \dots, a_m\}$ and $\{c_1, \dots, c_m\}$ for this to be the case.

First of all, for (3.3) to hold for any finite constant D, the set of vectors $\{a_1, \dots, a_m\}$ must span $\mathbb{R}^n$. The following construction is useful for this and other purposes: Let V be any proper subspace of $\mathbb{R}^n$, and let $V^\perp$ be its orthogonal complement. Then for any number $\lambda > 0$, let $X_{V,\lambda}$ denote the centered Gaussian random vector (see below for definition) such that

$$\forall u \in V, \quad \mathbb{E}\big((u \cdot X_{V,\lambda})^2\big) = \lambda \quad \text{and} \quad \forall u \in V^\perp, \quad \mathbb{E}\big((u \cdot X_{V,\lambda})^2\big) = 1 \,. \tag{3.4}$$

Then

$$S(X_{V,\lambda}) = -\frac{n}{2} \ln(2\pi e) - \frac{\dim(V)}{2} \ln(\lambda) \tag{3.5}$$

while, for any a in $\mathbb{R}^n$,

$$S(a \cdot X_{V,\lambda}) = -\tfrac{1}{2} \ln(2\pi e) - \tfrac{1}{2} \ln\big(\lambda |Pa|^2 + |P^\perp a|^2\big) \,, \tag{3.6}$$

where P is the orthogonal projection onto V, and $P^\perp = I - P$.

Now take V to be the orthogonal complement of the span of $\{a_1, \ldots, a_m\}$. If the latter is a proper subspace of $\mathbb{R}^n$, then $\dim(V) \geq 1$, and we see that for any finite D, (3.3) would be violated for sufficiently large λ, since then $|Pa_j|^2 = 0$ for each j.

Beyond this spanning condition, there are some simple compatibility conditions that must be satisfied by the vectors a_j and the numbers c_j. First of all, it follows from (3.2) that for all $\lambda > 0$,

$$S(\lambda X) = S(X) - n\ln(\lambda) \quad \text{and} \quad S(a \cdot \lambda X) = S(a \cdot X) - \ln(\lambda).$$

Therefore, (3.3) can only hold when

$$\sum_{j=1}^m c_j = n. \tag{3.7}$$

There is a further necessary condition that is somewhat less obvious. The key observation to make is that the right-hand side of (3.6) tends to infinity as λ tends to zero if and only if $|P^\perp a|^2 = 0$.

Consider any subset J of $\{1, \ldots, m\}$, and let

$$V_J := \operatorname{span}\{a_j \, ; \; j \in J\}.$$

Let G_J denote the Gaussian random variable $X_{V_J, \lambda}$ defined by (3.4) when $V = V_J$. Note that for each $j \in J$, $|P^\perp a_j|^2 = 0$, so that, for such j,

$$S(a_j \cdot G_J) = -\tfrac{1}{2}\ln(2\pi e) - \tfrac{1}{2}\ln\left(|a_j|^2\right) - \tfrac{1}{2}\ln(\lambda),$$

which tends to infinity as λ tends to zero. Therefore, letting λ approach zero, we see that the leading term in $\sum_{j=1}^m c_j S(a_j \cdot G_J) - S(G_J)$ is at least

$$\tfrac{1}{2}\Big(\dim(V_J) - \sum_{j \in J} c_j\Big)\ln(\lambda).$$

(It is exactly this unless for some $i \notin J$, $a_i \in V_J$, in which case we could have taken an even "worse" set J.) Hence, if $\dim(V_J) - \sum_{j \in J} c_j < 0$, there can be no upper bound on $\sum_{j=1}^m c_j S(a_j \cdot G) - S(G)$. Therefore, (3.3) can only hold when it is the case that, for all J,

$$\sum_{j \in J} c_j \leq \dim(V_J). \tag{3.8}$$

In particular, we must have $c_j \leq 1$ for all j.

We shall give a simple proof that these necessary conditions are sufficient. The following notation shall be used throughout the proof: Given any family $\{a_1, \ldots, a_m\}$ of vectors spanning $\mathbb{R}^n$, let

$$A = [a_1, \ldots, a_m]$$

denote $n \times m$ matrix whose jth column is a_j. We shall also use A to denote the family $\{a_1, \ldots, a_m\}$ of spanning vectors. Thinking of A as the matrix of a linear transformation, computed in the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, will be useful in the proofs of several lemmas below. Note that A has full (row) rank. Next, let c denote the vector in $\mathbb{R}^m$ whose jth entry is c_j. Finally, define the quantity $D(A, c)$ by

$$D(A, c) := \sup_X \left\{ \sum_{j=1}^m c_j S(a_j \cdot X) - S(X) \right\}, \tag{3.9}$$

where the supremum is taken over all random vectors X with values in $\mathbb{R}^n$ and with finite entropy. A random vector X for which this supremum is attained will be said to be *extremal* and will be called an *extremizer*.

Notice that with A fixed, $D(A, \cdot)$ is the pointwise supremum of a set of affine functions, and as such, it is convex. We introduce

$$K_A := \left\{ c \in [0,1]^m ;\ c \text{ verifies } (3.7) \text{ and } (3.8)\ \forall J \subset \{1, \ldots, m\} \right\}, \qquad (3.10)$$

which is clearly a convex subset of the hyperplane of $\mathbb{R}^m$ defined by (3.7). As we have seen, $D(A, c)$ is infinite outside K_A. We shall also need later to distinguish the interior of K_A relative to the intersection of $[0,1]^m$ and the hyperplane specified by (3.7):

$$K_A^\circ := \left\{ c \in K_A ;\ \sum_{j \in J} c_j < \dim(V_J),\ \forall J \subsetneq \{1, \ldots, m\},\ J \neq \emptyset \right\}. \qquad (3.11)$$

We shall make extensive use of the fact that K_A and K_A° are invariant under linear transformation, in the sense that for any invertible linear operator T on $\mathbb{R}^n$, we obviously have $K_{TA} = K_A$ and $K_{TA}^\circ = K_A^\circ$ with the notation $TA = [Ta_1, \ldots, Ta_m]$ when $A = [a_1, \ldots, a_m]$.

Also define $D_{\mathcal{G}}(A, c)$, the Gaussian analog of (3.9), by

$$D_{\mathcal{G}}(A, c) := \sup_G \left\{ \sum_{j=1}^m c_j S(a_j \cdot G) - S(G) \right\}. \qquad (3.12)$$

in which the supremum is taken over all centered Gaussian random vectors G with values in $\mathbb{R}^n$. By a *centered Gaussian random vector*, we mean one that has a density of the form

$$\frac{1}{|\det(C)|} \left(\frac{1}{2\pi} \right)^{n/2} e^{-|C^{-1}x|^2/2}$$

for some symmetric invertible matrix C on $\mathbb{R}^n$. More generally, a *Gaussian random vector* is a random vector of the form $x_0 + G$ with $x_0 \in \mathbb{R}^n$ and G a centered Gaussian random vector. We can restrict ourselves to centered random vectors because the entropy is invariant under translation. A Gaussian random vector is said to be *isotropic* if its covariance matrix is a multiple of the identity; it is said to be *standard* if it is centered and if its covariance matrix is the identity (i.e. it is a $\mathcal{N}(0, \mathrm{Id})$ Gaussian vector).

At this point, it is important to note that all the definitions made so far make sense more generally on a finite-dimensional Euclidean space $(E, \cdot)$. We have made the identification $E = \mathbb{R}^n$, which has the advantage of allowing us to work with matrices. Later, we shall also need to work on subspaces of $\mathbb{R}^n$, which are then canonically equipped with the Euclidean structure inherited from $\mathbb{R}^n$; we then need to work with the corresponding Euclidean versions of the notions introduced above.

It is clear that $D_{\mathcal{G}}(A, c)$ is also a convex function of c, and that $D_{\mathcal{G}}(A, c) \leq D(A, c)$. Also, since our proof that $D(A, c) = \infty$ for $c \notin K_A$ used a centered Gaussian random vector, it shows also that $D_{\mathcal{G}}(A, c) = \infty$ for $c \notin K_A$. In fact, we have the following:

Theorem 3.1. *For every family $A = \{a_1, \ldots, a_m\}$ of m vectors spanning $\mathbb{R}^n$ and every vector c in $\mathbb{R}^m$ with $0 \leq c_j \leq 1$ for all j, we have*

$$D(A, c) = D_{\mathcal{G}}(A, c) \,,$$

and furthermore $D(A, c)$ is finite if and only if $c \in K_A$.

The proof will be accomplished in three steps:

Step 1: We shall first consider the case in which the vectors a_j are all unit vectors u_j satisfying the following special condition, put forward by K. Ball in the setting of Brascamp–Lieb inequalities (see, e.g. [B]):

$$\sum_{j=1}^{m} c_j \, u_j \otimes u_j = \mathrm{Id}_{\mathbb{R}^n} \,, \tag{3.13}$$

with $c_j \geq 0$. (Note that (3.7) automatically holds, as it can be seen by taking the trace, and that $c_j \leq 1$ for all $j \leq m$.) Under this condition, we give a simple proof of Theorem 3.1 using an elementary superadditivity property of the Fisher information and integration along the heat flow. The proof here draws on ideas from [BaCM].

Step 2: We shall show that for $c \in K_A^\circ$, there is a linear change of variables that reduces this case to the one considered in the first step. While the lemma that provides the existence of the change of variables would appear to be a simple statement about linear algebra, the existence of this change of variables is intimately connected with the existence of Gaussian optimizers for the subadditivity (and hence the Brascamp–Lieb) inequality.

Step 3: We show that on $K_A \backslash K_A^\circ$, the variational problem in (3.9) may be "split" into two problems of the same type, but each involving only a subsets of the original vectors, and integration over a proper subspace of $\mathbb{R}^n$. Repeating this splitting operation, one eventually reduces to variational problems of the type considered in the second step. This step is modeled after a similar splitting argument developed in [CLL1], but as we shall see, the entropic version has advantages that will help us determine all of the cases of equality.

REMARK. If one is content to prove only that $D(A, c)$ is finite if and only if $c \in K_A$, there is a very expeditious route: One can easily check the finiteness of $D(A, c)$ at the extreme points of $c \in K_A$ (where, as shown by Barthe, each c_j is either 0 or 1). Then the convexity of $D(A, c)$ implies finiteness on all of K_A, and we know it is infinite outside. Proving the equality $D(A, c) = D_{\mathcal{G}}(A, c)$ on all of K_A is more subtle: The values of $D(A, c)$ and $D_{\mathcal{G}}(A, c)$ *do* jump as one crosses the boundary of K_A, and we see nothing to preclude $D(A, c)$ from jumping up more than $D_{\mathcal{G}}(A, c)$ *on* the boundary. Thus, it is not only for the classification of the cases of equality that we argue as we do in the third step: we do not know of any quick way to "pass to the boundary" of K_A and wrap up the proof of Theorem 3.1 after the second step without developing the splitting argument.

We now begin with the first step. Here we shall use a simple superadditivity result for the *Fisher information*: If $X \sim f$ is a random vector with a differentiable

density f, define the Fisher information of X or of f by

$$I(X) = I(f) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \,. \tag{3.14}$$

This quantity is related to the entropy through the heat flow as follows: Let Δ denote the Laplacian on $\mathbb{R}^n$, and let G denote a standard Gaussian random vector on $\mathbb{R}^n$ independent of X so that, if $f \sim X$,

$$e^{t\Delta} f \sim X + \sqrt{t} G \,.$$

Then we have the identity

$$\tfrac{\mathrm{d}}{\mathrm{d}t} S(e^{t\Delta} f) = -I(e^{t\Delta} f) \,,$$

and in particular, the right-hand side is finite for all $t > 0$.

The basic inequality concerning the Fisher information that will yield us our subadditivity result is the fact that for any unit vector u,

$$I(f_{(u)}) = I(u \cdot X) \le \int_{\mathbb{R}^n} \frac{|u \cdot \nabla f|^2}{f} \,, \tag{3.15}$$

with equality if and only if f is the product of $f_{(u)}$ and a probability density g on the orthogonal complement of u. This was proved in [C]; see Theorem 2 there with $p = 2$. Let us include here for completeness a different proof taken from [BaCLM] (were more abstract settings are studied). This proof requires more regularity than the one in [C], but that is fine for our purpose, as we shall apply the inequality along the heat flow.

Using the definition of the marginal (3.1) twice and Hölder's inequality, we have

$$\begin{aligned}
I\big(f_{(u)}\big) &= -\int_{\mathbb{R}} f_{(u)} \big(\ln f_{(u)}\big)'' \,\mathrm{d}t = -\int_{\mathbb{R}^n} f(x) \big(\ln f_{(u)}\big)''(x \cdot u) \,\mathrm{d}^n x \\
&= \int_{\mathbb{R}^n} \frac{(f_{(u)})'(x \cdot u)(u \cdot \nabla f(x))}{f_{(u)}(x \cdot u)} \,\mathrm{d}^n x \\
&\le \sqrt{\int_{\mathbb{R}^n} \frac{[(f_{(u)})'(x \cdot u)]^2}{(f_{(u)}(x \cdot u))^2} \, f(x) \,\mathrm{d}^n x \int_{\mathbb{R}^n} \frac{(u \cdot \nabla f(x))^2}{f(x)} \,\mathrm{d}^n x} \\
&= \sqrt{\int_{\mathbb{R}} \frac{[(f_{(u)})']^2}{(f_{(u)})^2} f_{(u)} \,\mathrm{d}t \int_{\mathbb{R}^n} \frac{(u \cdot \nabla f)^2}{f} \,\mathrm{d}^n x} \\
&= \sqrt{I(f_{(u)}) \int_{\mathbb{R}^n} \frac{(u \cdot \nabla f)^2}{f} \,\mathrm{d}^n x} \,.
\end{aligned}$$

This proves (3.15). Equality in (3.15) requires equality in Hölder's inequality and so for some $\lambda \in \mathbb{R}$ we have $(u \cdot \nabla) \log f(x) = \lambda (\log f_{(u)})'(x \cdot u)$ for all $x \in \mathbb{R}^n$; this λ has to be 1 for equality to hold in (3.15) and therefore $f(x) = f_{(u)}(x \cdot u) h(x - (x \cdot u)u)$ for some probability density h on $u^\perp$.

From (3.15), we immediately deduce the superadditivity of information. But before stating the result, let us state a definition needed to discuss the cases of equality.

DEFINITION 3.2 (Reducible spanning set). Let $\{a_1, \ldots, a_m\}$ be any set of m vectors spanning $\mathbb{R}^n$. It is a *reducible spanning set* in case there are two proper subspaces

V_1 and V_2 of $\mathbb{R}^n$ such that $\mathbb{R}^n = V_1 \oplus V_2$, and such that each a_j belongs to either V_1 or to V_2. Otherwise, $\{a_1, \ldots, a_m\}$ is called an *irreducible spanning set*.

PROPOSITION 3.3.   *Consider any set* $\{u_1, \ldots, u_m\}$ *of* m *unit vectors in* $\mathbb{R}^n$, *such that there are numbers* $\{c_1, \ldots, c_m\}$, *with* $0 \leq c_j \leq 1$ *for each* $j \leq m$, *so that the decomposition of the identity* (3.13) *is satisfied. Let* G *denote a standard Gaussian random vector.*

Then for all random vectors X *with finite Fisher information,*

$$\sum_{j=1}^{m} c_j \, I(u_j \cdot X) \leq I(X), \tag{3.16}$$

with equality if $X = G$, *and for all random vectors* X *with finite entropy*

$$\sum_{j=1}^{m} c_j \, S(u_j \cdot X) - S(X) \leq \sum_{j=1}^{m} c_j \, S(u_j \cdot G) - S(G) = 0. \tag{3.17}$$

Moreover there is equality in these inequalities if and only if for each $j \leq m$, $u_j \cdot X$ *and* $X - (u_j \cdot X)u_j$ *are independent. Under the condition that* $n \geq 2$ *and that* $\{u_1, \ldots, u_m\}$ *is an irreducible spanning set, then there is equality in these inequalities if and only if* X *is an isotropic Gaussian random vector.*

Note that this proposition in particular implies that $D(U, c) = D_{\mathcal{G}}(U, c) = 0$ when $U = [u_1, \ldots, u_m]$ are unit vectors of $\mathbb{R}^n$ and $c = (c_1, \ldots, c_m)$ non-negative real numbers satisfying (3.13).

The proof of (3.16) and (3.17) is elementary and follows [BaCM]. The determination of the cases of equality requires a bit more work, but it remains quiet direct (compared to analogous result on the side of the Brascamp–Lieb inequality).

Proof. Inequality (3.16) follows immediately from (3.15) and condition (3.13) rewritten in the form

$$\forall x \in \mathbb{R}^n, \quad \sum_{j=1}^{m} c_j \, (x \cdot u_j)^2 = |x|^2.$$

Equality for $X = G$ is obvious as $G \cdot u_i$ is a standard Gaussian variable and so the computation boils down to the equality $\sum c_j = n$. (For the same reason the right-hand side of the inequality (3.17) is zero.)

As we have noted, the Fisher information of f is related to the entropy of f through $\frac{\mathrm{d}}{\mathrm{d}t} S(e^{t\Delta} f) = -I(e^{t\Delta} f)$. It is also easy to see (using that Δ commutes with translations) that if u is any unit vector, then $f_{(u)}$, the marginal of f along u has the property that $(e^{t\Delta} f)_{(u)} = e^{t\Delta} f_{(u)}$, where we keep the same notation of the 1-dimensional heat semi-group ($\Delta g = g''$ in dimension 1); we again have (in dimension 1) that

$$\frac{\mathrm{d}}{\mathrm{d}t} S\big((e^{t\Delta} f)_{(u)} \big) = -I\big((e^{t\Delta} f)_{(u)} \big).$$

Then since $e^{t\Delta} f \sim X + \sqrt{t}G$, and because $\sum_{j=1}^{m} c_j S(u_j \cdot X) - S(X)$ is invariant under dilation, i.e. under the substitution $X \to \lambda X$, we get

$$\left[\sum_{j=1}^{m} c_j S(u_j \cdot G) - S(G) \right] - \left[\sum_{j=1}^{m} c_j S(u_j \cdot X) - S(X) \right]$$
$$= \int_0^\infty \left[I(e^{t\Delta} f) - \sum_{j=1}^{m} c_j I\big((e^{t\Delta} f)_{(u_j)}\big) \right] \mathrm{d}t \, .$$

By Theorem 3.3, the integrand above is non-negative for all t, and so (3.17) is proved.

The condition for cases of equality in (3.15) tell us that there is equality in (3.16) for a random vector X with finite Fisher information if and only if X verifies the following property $(\mathcal{P})$:

$$(\mathcal{P}) \quad \forall i \le m \, , \quad X \cdot u_i \text{ and } X - (X \cdot u_i) u_i \text{ are independent} \, .$$

If G is a standard Gaussian random vector independent of X, then X verifies $(\mathcal{P})$ if and only if, for all $t > 0$, $X + \sqrt{t} G$ verifies $(\mathcal{P})$. Thus for a random vector with finite entropy, there is equality in (3.17) if and only if X verifies $(\mathcal{P})$.

Our goal is now to characterize, when $n \ge 2$, the random vectors verifying $(\mathcal{P})$ under the assumption that $\{u_1, \ldots, u_m\}$ is an irreducible spanning set of unit vectors. First note that if we prove that $X + \sqrt{t} G$ is an isotropic Gaussian for all $t > 0$, then so is X. Therefore, using again the stability of the property $(\mathcal{P})$, we need only consider random vectors X with smooth and strictly positive density. Secondly, we can assume that no two vectors of the family $\{u_i\}_{i \le m}$ are linearly dependent. Indeed, by keeping only one representative for the subspaces $\mathbb{R} u_j$, we construct a subfamily of the vectors $\{u_i\}_{i \le m}$ which span $\mathbb{R}^n$ and which remains irreducible.

So from now on let $\{u_1, \ldots, u_m\}$ be an irreducible spanning set of unit vectors of $\mathbb{R}^n$ ($n \ge 2$), with no two vectors linearly dependent, and X a random vector verifying $(\mathcal{P})$ and with a smooth density $f > 0$. Thus for every $i \le m$ there exists two probability densities g_i and h_i, on $\mathbb{R}$ and $u_i^\perp \simeq \mathbb{R}^{n-1}$ respectively, such that

$$f(x) = g_i(x \cdot u_i) h_i\big(x - (x \cdot u_i) x\big)$$

Writing $F = \log f$, $G_i = \log g_i$ and $H_i = \log h_i$ for each $i \le m$, we have

$$F(x) = G_i(x \cdot u_i) + H_i\big(x - (x \cdot u_i) x\big) \, ,$$

so that

$$(u_i \cdot \nabla) F(x) = G_i'(u_i \cdot x) \, .$$

Hence, for any $j \ne i$,

$$(u_j \cdot \nabla)(u_i \cdot \nabla) F(x) = (u_i \cdot u_j) G_i''(u_i \cdot x) \, .$$

Interchanging the roles of i and j,

$$(u_j \cdot \nabla)(u_i \cdot \nabla) F(x) = (u_i \cdot u_j) G_j''(u_j \cdot x) \, .$$

Evidently the left-hand side depends on x only thorough $u_i \cdot x$ and only through $u_j \cdot x$. But since u_i and u_j are linearly independent, this means that the left-hand side is constant. Hence,

$$\text{for every } \ i \ne j, \ (u_i \cdot \nabla)(u_i \cdot \nabla) F \ \text{is constant} \, .$$

Furthermore, under the condition that $\{u_1, \ldots, u_m\}$ is an irreducible spanning set, if any one vector u_i is removed from $\{u_1, \ldots, u_m\}$, the remaining vectors still span $\mathbb{R}^n$. For otherwise, since $m \ge n \ge 2$, we could take V_1 to be the span of $\{u_i\}$, and take

V_2 to be the span of $\{u_1, \ldots, u_m\} \setminus u_i$, and we would have $\mathbb{R}^n = V_1 \oplus V_2$. Thus each u_i decomposes in the generating family $\{u_j\}_{j \neq i}$ and therefore,

$$\text{for every } \ i, j \leq m, \ (u_i \cdot \nabla)(u_i \cdot \nabla)F \ \text{ is constant.}$$

But this implies that the Hessian of F is constant. Thus, X is Gaussian. To prove that this Gaussian is isotropic, let C be the covariance matrix of X. Then property $(\mathcal{P})$ implies that each u_i is an eigenvector of C. Since eigenvectors of symmetric matrices are orthogonal if they have distinct eigenvalues, all of the eigenvalues must be the same unless there is such a "splitting" of $\mathbb{R}^n$ into at least two (orthogonal) subspaces that together contain all of the vectors u_j. This would contradict the hypothesis that $\{u_1, \ldots, u_m\}$ is an irreducible spanning set.   □

The following lemma will facilitate the application of the statement concerning the cases of equality in Proposition 3.3:

LEMMA 3.4.   Let $A = \{a_1, \ldots, a_m\}$ be any family of m vectors spanning $\mathbb{R}^n$. If $\{a_1, \ldots, a_m\}$ is a reducible spanning set and $D(A, c)$ is finite, then $c \notin K_A^\circ$.

Proof. Let $\mathbb{R}^n = V_1 \oplus V_2$ be a decomposition of $\mathbb{R}^n$ into two proper subspaces such that each a_j is contained in one of them or the other. Let V be the *orthogonal complement* of V_1, $\mathbb{R}^n = V_1 \overset{\perp}{\oplus} V$ and let $X_{V,\lambda}$ be the Gaussian random variable defined as in (3.4). Then by (3.5) and (3.6), with P denoting the orthogonal projection onto V,

$$\sum_{j=1}^m c_j S(a_j \cdot X_{V,\lambda}) - S(X_{V,\lambda})$$

$$= -\tfrac{1}{2} \Big(\sum_{j: a_j \in V_1} c_j \ln \big(|a_j|^2\big) + \sum_{j: a_j \in V_2} c_j \ln \big(\lambda |Pa_j|^2 + |P^\perp a_j|^2\big) \Big)$$

$$+ \tfrac{1}{2} \dim(V) \ln(\lambda), \quad (3.18)$$

with $Pa_j \neq 0$ for $j \in V_2$, since $Px = 0 \Rightarrow x \in V_1$. Then, using that $\dim(V) = \dim(V_2)$, this expression (in λ) has the form

$$\tfrac{1}{2} \Big(\dim(V_2) - \sum_{j: a_j \in V_2} c_j \Big) \ln(\lambda) + (\text{terms bounded in } \lambda > 1) \,,$$

which is unbounded for large λ unless

$$\sum_{j: a_j \in V_2} c_j = \dim(V_2) \,.$$

This must be the case since by hypothesis that $D(A, c) < \infty$. Thus, $c \notin K_A^\circ$.   □

We have now completed the first step. We start the second by showing that the change of variables matrix R does exist for $c \in K_A^\circ$. The existence of such a change of variables can be deduced from results of Bennett–Carbery–Christ–Tao [BeCCT1]. However, the flow of logic in their deduction (and in [CLL1]) runs counter to ours: They first show that such a change of variables exists whenever there are Gaussian optimizers for the Brascamp–Lieb problem, and then show that Gaussian optimizers exist for $c \in K_A^\circ$. Here, we need the change of variables at the outset of our analysis,

and hence need a direct proof of this result. We now provide one, using a geometric result of Barthe.

LEMMA 3.5.   Let $A = \{a_1, \ldots, a_m\}$ be any family of m vectors that span $\mathbb{R}^n$. Let $\{c_1, \ldots, c_m\}$ be any m numbers verifying $0 \leq c_j \leq 1$ and satisfying (3.8). If $c \in K_A^\circ$, then there exists an invertible symmetric $n \times n$ matrix R so that

$$\sum_{j=1}^m c_j \left(\frac{R a_j}{|R a_j|} \right) \otimes \left(\frac{R a_j}{|R a_j|} \right) = \mathrm{Id}_{\mathbb{R}^n} \, . \tag{3.19}$$

When $n \geq 2$, there is exactly one such matrix R satisfying the further requirements that R be positive definite, and that $\mathrm{trace}(R^2) = n$. On the other hand, for $c \notin K_A$ no such matrix R exists.

REMARK.   After settling the cases of equality in Theorem 3.1 we shall derive *necessary and sufficient conditions* for the existence of such a matrix R. Though the conditions are simple and explicit, it turns out that the matrix R exists if and only if the supremum in (3.12) is attained at some centered Gaussian G, and our proof that the conditions we give are necessary and sufficient depends on this.

Proof. Take any diagonal $m \times m$ matrix S with positive diagonal entries s_j, $j \leq m$, and define the $n \times n$ matrix R_S by

$$R_S = \left((AS)(AS)^t \right)^{-1/2} .$$

This makes sense since $(AS)(AS)^t$ is a positive definite $n \times n$ matrix. Notice that $(R_S A S)(R_S A S)^t = \mathrm{Id}_{\mathbb{R}^n}$, or, what is the same

$$\sum_{j=1}^m s_j^2 R_S a_j \otimes R_S a_j = \mathrm{Id}_{\mathbb{R}^n} \, .$$

Therefore,

$$\sum_{j=1}^m c_j \left(\frac{s_j}{\sqrt{c_j}} R_S a_j \right) \otimes \left(\frac{s_j}{\sqrt{c_j}} R_S a_j \right) = I \, .$$

We have what we seek if and only if, for each j, $\frac{s_j}{\sqrt{c_j}} R_S a_j$ is a unit vector, which is the case if and only if, for each j, $c_j = s_j^2 |R_S a_j|^2$. By the definition of R_S, this means

$$c_j = e_j \cdot \left[(AS)^t ((AS)(AS)^t)^{-1} (AS) \right] e_j \tag{3.20}$$

where $\{e_1, \ldots, e_m\}$ denotes the standard orthonormal basis in $\mathbb{R}^m$. Note that $e_j \cdot (AS)^t ((AS)(AS)^t)^{-1} (AS) e_j$ is also the jth diagonal entry of the orthogonal projection in $\mathbb{R}^m$ onto the image of $(AS)^t$.

It has been shown [Ba] (see [CLL1] for another proof and a statement in this formulation) that there exists positive numbers $s_1, \ldots, s_m$ for which (3.20) is true whenever $c \in K_A^\circ$, and that in this case, when $n \geq 2$, the set of numbers is unique up to a common multiple. Thus, for $c \in K_A^\circ$, such an R exists.

As for the uniqueness, note that given any such matrix R, we can change variables, replacing $X \to R^{-1} X$ and $a_j \to u_j := |R a_j|^{-1} R a_j$. Then Proposition 3.3 may be applied to deduce that the only extremizers for the new problem are isotropic Gaussians. Undoing the change of variables, we see that the only extremizers of the

original problem are Gaussians whose covariance is a multiple of R^2. Thus, under the further condition that R is positive definite (instead of simply symmetric), and that the trace of R^2 is fixed, R is uniquely determined.

The same change of variables argument (which is exploited systematically in Lemma 3.6 below) shows, through Proposition 3.3, that if such a matrix R exists, then $D(A, c) < \infty$. As we have seen, this is impossible when $c \notin K_A$. □

REMARK. The first proof that there exists a solution, essentially unique, to (3.20) whenever $c \in K_A^\circ$ is due to Barthe [Ba]. However, he used a different characterization of K_A, and did not mention the condition (3.8). Another proof of this, based directly on (3.8) was given in [CLL1], together with a proof that the characterization of K_A in Barthe's paper is equivalent to the one based on (3.8).

With the change of variables provided by the previous lemma, we can finish the second step and describe what happens when $c \in K_A^\circ$.

LEMMA 3.6.   For any family $A = \{a_1, \ldots, a_m\}$ of m vectors spanning $\mathbb{R}^n$, and all vectors c in K_A°,
$$D(A, c) = D_{\mathcal{G}}(A, c),$$
and there exist a Gaussian optimizer. Moreover, if $n \geq 2$, then $\sum_{j=1}^{m} c_j S(a_j \cdot X) - S(X)$ $= D(A, c)$ if and only if X is Gaussian and its covariance is a constant multiple of R^2 where R is the unique positive definite matrix verifying (3.19) with $\mathrm{Tr}(R^2) = n$.

REMARK.   The condition "$n \geq 2$", which has already appeared several times, is present because in one dimension, the subadditivity problem is trivial, so that Gaussians play no special role. Indeed, assume we are given $c_1, \ldots, c_m \geq 0$ with the condition that $\sum c_j = 1$ and $A = \{a_1, \ldots, a_m\}$ a family of non-zero real numbers. Then, setting
$$D := -\sum_{j=1}^{m} c_j \log |a_j|$$
we have, for every random variable X on $\mathbb{R}$ with finite entropy,
$$\sum_{j=1}^{m} c_j S(a_j X) - S(X) = D.$$
Therefore $D(A, c) = D$ and every random variable X is an extremizer.

Proof. Let R be an invertible symmetric matrix verifying (3.19) provided by the Lemma 3.5. Since for any random vector X with finite entropy, we have
$$S(X \cdot a_j) = S\left(\frac{Ra_j}{|Ra_j|} \cdot R^{-1}X \right) - \ln\left(|Ra_j| \right) \quad \text{and} \quad S(X) = S(R^{-1}X) - \ln\left(|\det(R)| \right),$$
we obtain
$$\sum_{j=1}^{m} c_j S(a_j \cdot X) - S(X) = \sum_{j=1}^{m} c_j S\left(\frac{Ra_j}{|Ra_j|} \cdot R^{-1}X \right) - S(R^{-1}X)$$
$$- \sum_{j=1}^{m} c_j \ln\left(|Ra_j| \right) + \ln\left(|\det(R)| \right).$$

Introduce the family of vectors $u_j := Ra_j/|Ra_j|$ for $j \leq m$, and set $U = [u_1, \ldots, u_m]$. The previous equality implies that

$$D(A, c) = D(U, c) - \sum_{j=1}^{m} c_j \ln \left(|Ra_j| \right) + \ln \left(|\det(R)| \right). \qquad (3.21)$$

Thus we are reduced to studying the problem determining $D(U, c)$ and the extremizers there (noting that X is an extremizer for $D(A, c)$ if and only if $R^{-1}X$ is extremizer an for $D(U, c)$). Note also that since the vectors $\{u_1, \ldots, u_m\}$ are obtained from the vectors $\{a_1, \ldots, a_m\}$ by a non-singular linear transformation, they span $\mathbb{R}^n$, and we have $K_U^\circ = K_A^\circ \ni c$.

Since $U = [u_1, \ldots, u_m]$ is a family of *unit* vectors verifying the decomposition of the identity (3.13), we can apply Proposition 3.3 and get that

$$D(U, c) = D_{\mathcal{G}}(U, c) = 0 < \infty, \qquad (3.22)$$

and every isotropic Gaussian vector is an extremizer. To prove that all optimizers are Gaussian when $n \geq 2$, note first that, by Lemma 3.4, $c \in K_U^\circ$ implies that $\{u_1, \ldots, u_m\}$ is an irreducible spanning set. Therefore any optimizer of the variational problem defining $D(U, c)$ is an isotropic Gaussian. (Then every optimizer for $D(A, c)$ is Gaussian whose covariance is a multiple of R^2.) □

REMARK.   Note that the proof above also gives the following statement: If there exists an invertible matrix R verifying (3.19) then (with no further assumptions on c and A) we have that $D(A, c) < +\infty$ and that RG is an extremizer for every standard Gaussian vector G.

We now turn to the third step. When $c \in K_A \backslash K_A^\circ$, we will pick a non-empty proper subset J of $\{1, \ldots, m\}$ of least cardinality among subsets for which equality holds in (3.8). We shall now show that the variational problem defining $D(A, c)$ splits into two such problems involving fewer vectors and random variables in a lower-dimensional space. Repeated splittings, and what we have already proved, will enable us to settle all questions concerning the variational problem defining $D(A, c)$. The splitting argument presented here is patterned on one developed in [CLL1] for the Brascamp–Lieb inequality. However, as we shall see, in the subadditivity setting, the argument leads to a clear and simple analysis of cases of equality. It relies on properties of the conditional entropy.

As mentioned at the beginning of this section, we shall need to work on subspaces of $\mathbb{R}^n$ and thus make use of the definition made above in the setting of Euclidean spaces. For a given family $A = \{v_1, \ldots v_k\}$ of vectors on $\mathbb{R}^n$, we introduce the Euclidean subspace $E := \mathrm{span}(v_1, \ldots, v_k)$ equipped with the induced Euclidean structure from $\mathbb{R}^n$ (i.e. the scalar product is the same). For real numbers $c_1, \ldots, c_k$ with $0 \leq c_j \leq 1$, the quantities $D(A, c)$ and $D_G(A, c)$ are then implicitly assumed to be defined on the Euclidean subspace E (the random vectors live on E and the entropies are computed with respect to the Lebesgue measure on E, where the laws of the vectors live). Accordingly, the set K_A is to be understood as

$$K_A := \left\{ c \in [0, 1]^k ; \sum_{j=1}^{k} c_j = \dim(E) \text{ and (3.8) holds } \forall J \subset \{1, \ldots, k\} \right\}.$$

Let us fix the following notation. Let $A = \{a_1, \ldots, a_m\}$ be a family of of $m \geq 1$ vectors spanning an Euclidean space E,

$$E = \mathrm{span}\big(\{a_j \, ; \; j \in I\}\big)$$

(in a first step we shall have $E = \mathbb{R}^n$). For family of m real numbers $c \in K_A$ and a non-empty proper subset J of $\{1, \ldots, m\}$ for which equality holds in (3.8), denote by P_J the orthogonal projection onto $V_J = \mathrm{span}\{a_j; j \in J\}$ and let $P_J^\perp = \mathrm{Id}_E - P_J$ be the complementary projection. Define, for $j \in J^c := \{i \in \{1, \ldots, m\} \, ; i \notin J\}$ the vector $b_j = P_J^\perp a_j$ and

$$A_J = [a_j \, ; \; j \in J] \quad \text{and} \quad B_{J^c} = [b_j \, ; \; j \in J^c]$$

the ordered (by ordering J and J^c as increasing subsequences of $1, \ldots, m$) families of vectors $(a_j)_{j \in J}$ and $(b_j)_{j \in J^c}$. For any subset K of $\{1, \ldots, m\}$, and $c \in \mathbb{R}^m$, let c_K denote the vector of $\mathbb{R}^{|K|}$ whose coordinate are the $(c_j)_{j \in K}$ (K being written as an increasing subsequence of $1, \ldots, m$). Since there is equality in (3.8) for J, we have

$$c_J \in K_{A_J}.$$

Note that $V_J + V_{J^c} = E$ (*a priori* this sum is not direct) and so $V_J^\perp = P_J^\perp V_{J^c}$. Thus we have $V_J^\perp = \mathrm{span}(\{b_j : j \in J^c\}t)$, i.e.

$$E = \mathrm{span}\big(\{a_j \, ; \; j \in J\}\big) \overset{\perp}{\oplus} \mathrm{span}\big(\{b_j \, ; \; j \in J^c\}\big). \tag{3.23}$$

And we also have

$$c_{J^c} \in K_{B_{J^c}}.$$

Indeed, using (3.23) and equality in (3.8) for J, we have $\sum_{j \in J^c} c_j = \dim(\mathrm{span}\{b_j; j \in J^c\})$, and also for $\tilde{J} \subset J^c$, since $P_J^\perp a_j = 0$ for $j \in J$,

$$\sum_{j \in \tilde{J}} c_j + \dim(V_J) = \sum_{j \in J \cup \tilde{J}} c_j$$

$$\leq \dim\big(\mathrm{span}\{a_j \, ; \; j \in J \cup \tilde{J}\}\big)$$

$$= \dim\big(\mathrm{span}\{P_J a_j + P_J^\perp a_j \, ; \; j \in J \cup \tilde{J}\}\big)$$

$$\leq \dim\big(\mathrm{span}\{P_J a_j \, ; \; j \in J \cup \tilde{J}\}\big) + \dim\big(\mathrm{span}\{P_J^\perp a_j \, ; \; j \in \tilde{J}\}\big)$$

$$= \dim(V_J) + \dim\big(\mathrm{span}\{b_j \, ; \; j \in \tilde{J}\}\big)$$

For an invertible operator T on $\mathbb{R}^n$ we shall use the standard notation

$$T^{-*} := (T^{-1})^* = (T^*)^{-1}$$

where $T^* x \cdot y = x \cdot Ty$ for all $x, y \in \mathbb{R}^n$. With these definitions, we now state the splitting lemma. Only the first part of the statement is needed to complete the proof of Theorem 3.1; the rest will be used for the characterization of extremizers.

LEMMA 3.7.   *Given any family* $A = \{a_1, \ldots, a_m\}$ *of* m *vectors spanning* $\mathbb{R}^n$ *and* $c \in K_A \setminus K_A^\circ$ *with* $c_j > 0$ *for all* $j \leq m$, *let* J *be a non-empty proper subset of* $\{1, \ldots, m\}$ *for which equality holds in (3.8), and suppose that* J *has the least cardinality among all such subsets. Then with* A_J, c_J, B_{J^c} *and* c_{J^c} *defined as above, we have*

$$D(A, c) = D(A_J, c_J) + D(B_{J^c}, c_{J^c}), \tag{3.24}$$

and if $D_{\mathcal{G}}(B_{J^c}, c_{J^c}) = D(B_{J^c}, c_{J^c})$, *then* $D_{\mathcal{G}}(A, c) = D(A, c)$.

Suppose next that there exists an extremizing random vector X; i.e. a random vector X such that

$$\sum_{j=1}^{m} c_j\, S(a_j \cdot X) - S(X) = D(A,c)\,. \tag{3.25}$$

Then

$$\mathbb{R}^n = V_J \oplus V_{J^c}\,, \tag{3.26}$$

and this direct sum is an orthogonal decomposition in the inner product given by the covariance matrix of X; i.e. $\langle x,y \rangle = \mathbb{E}[(x \cdot (X - \mathbb{E}X))(y \cdot (X - \mathbb{E}X))]$.

Moreover, if T is an (invertible) operator on $\mathbb{R}^n$ such that one has the orthogonal decomposition

$$\mathbb{R}^n = TV_J \overset{\perp}{\oplus} TV_{J^c}$$

(for instance $T = H_X^{1/2}$ where H_X is the covariance matrix of an extremizer X, so that $\langle x,y \rangle = x \cdot H_X y$), then X is an extremizer (3.25) if and only if $T^{-*}X$ decomposes as $T^{-*}X = Y + Z$ where Y and Z are independent random vectors with values in TV_J and TV_{J^c}, and which are extremizer for $([Ta_j;\, j \in J], c_J)$ and $([Ta_j;\, j \in J^c], c_{J^c})$, respectively.

The proof of this lemma relies on some well-known identities and inequalities concerning conditional entropy that we now recall.

Let E and F be two Euclidean spaces (equipped with the Lebesgue measure). If W and Y are two random vectors with values in E and F respectively, with a joint density $\rho(w,y)$ on $E \times F$, let $\rho_Y(y) = \int_E \rho(w,y)\,\mathrm{d}w$ and $\rho_W(w) = \int_F \rho(w,y)\,\mathrm{d}y$ be the two marginal densities on F and E, which are of course the densities of W and Y respectively.

Then the *conditional density of W given Y* is $\rho(w|y) = \rho(w,y)/\rho_Y(y)$. The *conditional entropy of W given $Y = y$* is then defined to be

$$S(W|Y=y) = \int_E \rho(w|y)\ln \rho(w|y)\,\mathrm{d}w\,.$$

Since the entropy of (W,Y), $S(W,Y)$, is given by

$$S(W,Y) = \int_{E \times F} \rho(w,y)\ln \rho(w,y)\,\mathrm{d}w\,\mathrm{d}y\,,$$

the identity

$$S(W,Y) = \int_F S(W|Y=y)\rho_Y(y)\,\mathrm{d}y + S(Y) \tag{3.27}$$

follows directly from the definitions. Furthermore, by Jensen's inequality

$$S(W) \leq \int_E S(W|Y=y)\rho_Y(y)\,\mathrm{d}y\,, \tag{3.28}$$

and there is equality if and only if W and Y are independent.

Proof of Lemma 3.7. Fix any random vector X with values in $\mathbb{R}^n$ and suppose that $S(X)$ is finite. We shall use the definition and notation given before the lemma. Let P_J denote the orthogonal projection onto V_J, and recall that we have the decomposition (3.23), so that $P_J^\perp = \mathrm{Id}_{\mathbb{R}^n} - P_J$ is also the orthogonal projection onto $\mathrm{span}(\{b_j : j \in J^c\})$ where $b_j = P_J^\perp(a_j)$ for all $j \in J^c$. Let us introduce

$$Y = P_J X \quad \text{and} \quad Z = P_J^\perp X$$

so that $X = Y + Z$. Then $S(X) = S(Y, Z)$ and so from (3.27),

$$S(X) = \int_{V_J} S(Z|Y = y)\rho_Y(y)\,\mathrm{d}y + S(Y)\,. \tag{3.29}$$

For each $j \in J$, we have $a_j \cdot X = a_j \cdot Y$, so that

$$S(a_j \cdot X) = S(a_j \cdot Y) \quad \text{for } j \in J\,. \tag{3.30}$$

Note that for $j \in J^c$, $b_j \neq 0$, or else $a_j \in V_J$; but this is impossible since $c_j > 0$, and we already have $\sum_{j=1}^{\ell} c_j = \dim(V_J)$. We have, using the invariance of the entropy under translation,

$$S(a_j \cdot X|Y = y) = S(a_j \cdot Z + a_j \cdot y|Y = y) = S(b_j \cdot Z|Y = y) \quad \text{for } j \in J^c\,. \tag{3.31}$$

Therefore, by applying (3.28) to $(X \cdot a_j, Y)$ on $\mathbb{R} \times V_J$, we get

$$S(a_j \cdot X) \leq \int_{V_J} S(b_j \cdot Z|Y = y)\rho_Y(y)\,\mathrm{d}y \quad \text{for } j \in J^c\,. \tag{3.32}$$

Now combining (3.29), (3.30) and (3.32), we have that

$$\sum_{j=1}^{m} c_j S(a_j \cdot X) - S(X) \leq \int_{V_J} \Big[\sum_{j \in J^c} c_j S(b_j \cdot Z|Y = y) - S(Z|Y = y) \Big] \rho_Y(y)\,\mathrm{d}y$$
$$+ \sum_{j \in J} c_j S(a_j \cdot Y) - S(Y)\,. \tag{3.33}$$

It is clear from (3.33) and the definition of $D(B_{J^c}, c_{J^c})$ that

$$D(A, c) \leq D(A_J, c_J) + D(B_{J^c}, c_{J^c})\,.$$

To see that there is actually equality here, we use the fact that J is a critical set of minimal cardinality. This implies that $c_J \in K_{A_J}^\circ$, and by Lemma 3.6, there is a centered Gaussian random vector Y for which

$$\sum_{j \in J} c_j S(a_j \cdot Y) - S(Y) = D(A_J, c_J)\,. \tag{3.34}$$

Pick $\epsilon > 0$ and let Z be any random variable with values in $V_J^\perp$ that is independent of Y and such that

$$\sum_{j \in J^c} c_j S(b_j \cdot Z) - S(Z) > D(B_{J^c}, c_{J^c}) - \epsilon\,. \tag{3.35}$$

For $\delta > 0$, form the $\mathbb{R}^n$ valued random vector $X = \delta Y + Z$. Since Y and Z are orthogonal and independent, $S(X) = S(\delta Y, Z) = S(\delta Y) + S(Z)$. The scaling invariance implies that (3.34) holds when Y is replaced by δY. Also, for $j \in J^c$, as δ approaches zero, $S(a_j \cdot X) = S(b_j \cdot Z + \delta a_j \cdot Y)$ approaches $S(b_j \cdot Z)$. (Note that by the independence of Y and Z, $b_j \cdot Z + \delta a_j Y$ is simply a standard Gaussian regularization of $b_j \cdot Z$.) It now follows that, for δ sufficiently small,

$$\sum_{j=1}^{m} c_j S(a_j \cdot X) - S(X) \geq D(A_J, c_J) + D(B_{J^c}, c_{J^c}) - 2\epsilon\,.$$

This implies that $D(A, c) \geq D(A_J, c_J) + D(B_{J^c}, c_{J^c})$. We have implicitly assumed that $D(B_{J^c}, c_{J^c}) < +\infty$ (we shall later only need this case, actually), but the argument remains valid if $D(B_{J^c}, c_{J^c}) = +\infty$. Thus (3.24) is established.

Now suppose that $D_{\mathcal{G}}(B_{J^c}, c_{J^c}) = D(B_{J^c}, c_{J^c})$. Then we may further assume that the random variable Z in the previous paragraph is a centered Gaussian random variable. Combining this with the independent extremal centered Gaussian random variable Y, provided by Lemma 3.6, we see that we may take the random variable X in the previous paragraph to be a centered Gaussian. Hence, in this case, $D_{\mathcal{G}}(A, c) = D(A, c)$.

It remains to prove the last statements concerning the cases of equality.

We first assume that we are given a finite entropy random variable X for which (3.25) is satisfied. By making a translation, we may assume that X is centered; i.e. $E(X) = 0$. Furthermore, the covariance matrix is non-degenerate or else the law of X would be concentrated on a proper subspace and this is inconsistent with finite entropy. Since X satisfies (3.25), there must be equality in (3.33), and it must be the case that

$$\sum_{j \in J} c_j S(a_j \cdot Y) - S(Y) = D(A_J, c_J) \tag{3.36}$$

and that for each $y \in V_J$,

$$\sum_{j \in J^c} c_j S(b_j \cdot Z | Y = y) - S(Z | Y = y) = D(B_{J^c}, c_{J^c}). \tag{3.37}$$

And since X is centered, so is Y. Next, in addition to equality in (3.37), we must have equality in (3.33). Since the only inequality used in deriving (3.33) was (3.32), this in turn requires equality in (3.32) for each $j \in J^c$. By (3.31), this means that for $j \in J^c$,

$$S(a_j \cdot X) = \int_{V_J} S(a_j \cdot X | Y = y) \rho_Y(y) \, \mathrm{d}y.$$

By the condition for equality in (3.28), this implies that for $j \in J^c$, $a_j \cdot X$ and Y are independent random variables. But then for any $y \in V_J$, by independence

$$\langle y, a_j \rangle = \mathbb{E}\big[(y \cdot Y)(a_j \cdot X)\big] = \mathbb{E}(y \cdot Y)\mathbb{E}(a_j \cdot X) = 0.$$

This shows that V_J and V_{J^c} are orthogonal subspaces in the inner product defined in terms of the covariance. Thus their dimension sums exactly to n and so (3.26) holds.

We now prove the final statement describing how extremizers split.

Note that, given an invertible operator T on $\mathbb{R}^n$, a random vector X is extremal (3.25) for (A, c) if and only if $T^{-*}X$ is extremal for (TA, c) with the notation $TA = [Ta_1, \ldots, Ta_m]$. Indeed, since $a_j \cdot X = Ta_j \cdot T^{-*}X$ and $S(T^{-*}X) = S(X) + \ln(|\det(T)|)$ we have that (3.25) is equivalent to

$$\sum_{j=1}^{m} c_j \, S(Ta_j \cdot T^{-*}X) - S(T^{-*}X) = D(TA, c)$$

and $D(TA, c) = D(A, c) - \ln(|\det(T)|)$.

As in the statement of the lemma, let T be an invertible operator on $\mathbb{R}^n$ such that $\mathbb{R}^n = TV_J \overset{\perp}{\oplus} TV_{J^c}$. The previous remark explains the mechanism of replacing A by TA and X by $T^{-*}X$. So after this transformation we are reduced to proving the statement in the case $T = \mathrm{Id}$. Therefore we assume from now on that

$$\mathbb{R}^n = V_J \overset{\perp}{\oplus} V_{J^c}.$$

We go back to the beginning of the proof and note that $b_j = a_j$ for all $j \in J^c$: the orthogonal projection does nothing in this case ($P_J^\perp = P_{J^c}$).

Assume X is an extremizer (3.25) which is decomposed as before as $X = Y + Z$. Then as in the argument above we must have that

$$\sum_{j \in J} c_j S(a_j \cdot Y) - S(Y) = D(A_J, c_J) \tag{3.38}$$

and that, for each $y \in V_J$,

$$\sum_{j \in J^c} c_j S(a_j \cdot Z | Y = y) - S(Z | Y = y) = D(A_{J^c}, c_{J^c}) \,, \tag{3.39}$$

with Y and $a_j \cdot X$ independent for every $j \in J^c$. Since $a_j \cdot X = a_j \cdot Z$ for every $j \in J^c$ we have that $a_j \cdot Z$ is independent of Y for $j \in J^c$ and so $S(a_j \cdot Z | Y = y) = S(a_j \cdot Z)$. Using this together with (3.28) for $W = Z$, we get, after integrating (3.39) with respect to $\rho_Y(y)\, dy$, and applying (3.28),

$$D(A_{J^c}, c_{J^c}) \leq \sum_{j \in J^c} c_j S(a_j \cdot Z) - S(Z) \,.$$

By the definition of $D(A_{J^c}, c_{J^c})$ this inequality must be an equality, i.e.

$$\sum_{j \in J^c} c_j S(a_j \cdot Z) - S(Z) = D(A_{J^c}, c_{J^c}) \,, \tag{3.40}$$

and therefore, there must be equality in the application of (3.28) that we just made. This implies that Z and Y are independent, as claimed.

Conversely, let X be a random vector such that $X = Y + Z$ in the decomposition $\mathbb{R}^n = V_J \overset{\perp}{\oplus} V_{J^c}$ with Y and Z independent and such that (3.38) and (3.40) holds. Then we have (3.39) and we readily check that There is equality at every step. So X is indeed an extremizer (3.25). $\qquad\square$

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. By Lemma 3.6, whenever $c \in K_A^\circ$, $D_{\mathcal{G}}(A, c) = D(A, c)$, and there is a Gaussian optimizer.

Hence it remains to consider the case $c \in K_A \setminus K_A^\circ$. Then taking J to be a proper non-empty subset of $\{1, \ldots, m\}$ of least cardinality for which there is equality in (3.8), we may "peel off" $|J|$ vectors from our set, as in the first part of Lemma 3.7, and reduce matters to the consideration of $D(B_{J^c}, c_{J^c})$. By that lemma, $D_{\mathcal{G}}(A, c) = D(A, c)$ whenever $D_{\mathcal{G}}(B_{J^c}, c_{J^c}) = D(B_{J^c}, c_{J^c})$. Now, if B_{J^c} and c_{J^c} are such that, for every proper subset of the remaining indices, strict inequality holds in the analog of (3.8), i.e. $c_{J^c} \in K_{A_{J^c}}^\circ$, then $D_{\mathcal{G}}(B_{J^c}, c_{J^c}) = D(B_{J^c}, c_{J^c})$ follows from Lemma 3.6. Otherwise, we "peel off" another proper subset of indices for which equality holds in (3.8), and reduce to a problem with a strictly smaller number of vectors. In a finite number of steps, this process must end. $\qquad\square$

Our next theorem concerns the cases of equality in the subadditivity inequality. As we have seen in Lemma 3.7, when there is equality, and no c_j is zero, then either $c \in K_A^\circ$, or the variational problem can be split into two problems of the same type, but involving reduced number of vectors, and for random variables taking values in subspaces of a reduced dimension.

Of course, each of these reduced problems must also have an optimizer, and so we can apply the same dichotomy to each of them. This leads to the following definition:

DEFINITION 3.8 (Totally reducible for c). Let $A = \{a_1, \ldots, a_m\}$ be a family of vectors that spans $\mathbb{R}^n$ and $\{c_1, \ldots, c_m\}$ a set of real numbers with $0 \leq c_j \leq 1$ for $j \leq m$. We say that $\{a_1, \ldots, a_m\}$ is *totally reducible for c* if $c \in K_A$ and in case for some $k \geq 1$ there is a decomposition (possibly with $k = 1$)

$$\{1, \ldots, m\} = J_0 \cup J_1 \cup \ldots \cup J_k$$

where $j \in J_0$ if and only if $c_j = 0$, and

$$\mathbb{R}^n = V_{J_1} \oplus \cdots \oplus V_{J_k} \quad \text{with} \quad V_{J_i} = \text{span}\big(\{a_\ell : \ell \in J_i\}\big),$$

such that, for each $1 \leq i \leq k$, there is *no* non-empty proper subset of J_i that yields equality in (3.8). Here, J_0 may be empty, but for $1 \leq i \leq k$, J_i is to be non-empty.

Note that, if $\{a_1, \ldots, a_m\}$ is totally reducible for c, then we have, with the notation of the definition, that for $1 \leq k \leq m$,

$$c_{J_i} \in K_{J_i}^\circ.$$

The analysis made so far proves the following theorem, which gives a complete analysis of the cases of equality in the subadditivity inequality.

Theorem 3.9. *Consider a family $A = \{a_1, \ldots, a_m\}$ of vectors spanning $\mathbb{R}^n$. Then for any $c \in K_A$, there exists a finite entropy random variable X for which*

$$\sum_{j=1}^m c_j S(a_j \cdot X) - S(X) = D(A, c), \tag{3.41}$$

if and only if $\{a_1, \ldots, a_m\}$ is totally reducible for c. In this case, if $\mathbb{R}^n = V_{J_1} \oplus \cdots \oplus V_{J_k}$ is the corresponding decomposition of $\mathbb{R}^n$ from Definition 3.8, let T be any symmetric positive operator on $\mathbb{R}^n$ such that the following orthogonal decomposition holds

$$\mathbb{R}^n = TV_{J_1} \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} TV_{J_k}. \tag{3.42}$$

Then the extremizers (3.41) are exactly the random vectors X such that $T^{-*}X$ decompose as

$$T^{-*}X = X_1 + \cdots + X_k$$

where $\{X_1, \ldots, X_k\}$ is an independent set of random variables with each X_i taking values in TV_{J_i} and extremal for the corresponding problem $([Ta_j; j \in J_i], c_{J_i})$. More precisely, for each $i \leq k$, if $\dim(V_{J_i}) = 1$, then X_i can be any finite entropy random variable with values in TV_{J_i}. However, if $\dim(V_{J_i}) > 1$, then X_i is necessarily Gaussian, and its covariance is a constant multiple of R_i^2, where R_i is the unique positive definite linear transformation on TV_{J_i} such that

$$\sum_{j \in J_i} c_j \left(\frac{R_i Ta_j}{|R_i Ta_j|} \right) \otimes \left(\frac{R_i Ta_j}{|R_i Ta_j|} \right) = \text{Id}_{TV_{J_i}} \quad \text{and} \quad \text{trace}(R_i^2) = \dim(TV_{J_i}) = \dim(V_{J_i}).$$

Finally, if X is an extremizer (3.41) then the symmetric positive operator T defined by $x \cdot T^2 y = \mathbb{E}[(x \cdot (X - \mathbb{E}X))(y \cdot (X - \mathbb{E}X))]$ satisfies the required condition (3.42).

Proof. The proof relies on successive applications of the Lemmas 3.7 and 3.6. First of all, note that the vectors a_j for the indices j such that $c_j = 0$ play no role in the inequality, and so without loss of generality, we may discard these indices without changing $D(A, c)$, the extremizers and K_A. So we will assume that $c_j > 0$ for all $j \leq m$ (this means $J_0 = \emptyset$ in Definition 3.8).

Assume there exists an extremizer X, which, after translation, can be assumed to be symmetric, and let T be the symmetric positive operator on $\mathbb{R}^n$ defined by $Tx \cdot Ty = \mathbb{E}[(x \cdot X)(y \cdot X)]$. As explained in the proof of the Lemma 3.7 the change of vectors $X \to T^{-*}X$ and $a_j \to Ta_j$ reduces the problem to the case $T = \mathrm{Id}$, which means that X has unit covariance. Then from Lemma 3.7 we have $\mathbb{R}^n = V_{J_1} \overset{\perp}{\oplus} V_{J_1^c}$ for some set of indices J_1, with $c_{J_1} \in K^\circ_{A_{J_1}}$ and $c_{J_1^c} \in K_{A_{J_1^c}}$, and moreover in this orthogonal decomposition $X = X_1 + Z$ with X_1 and Z extremal for (A_{J_1}, c_{J_1}) and $(A_{J_1^c}, c_{J_1^c})$, respectively. We then apply Lemma 3.7 on the space $V_{J_1^c}$ to the vector with unit covariance Z which is extremal. This gives for some $J_2 \subset J_1^c$ another orthogonal decomposition $V_{J_1^c} = V_{J_2} \overset{\perp}{\oplus} V_{J_2^c}$ where $J_2^c = \{j \in J_1^c; j \notin J_2\}$ with $c_{J_2} \in K^\circ_{A_{J_2}}$. After a finite number k of step this process must end and we have

$$\mathbb{R}^n = V_{J_1} \oplus \cdots \oplus V_{J_k}$$

with $c_{J_i} \in K^\circ_{A_{J_i}}$ for $i \leq k$. This shows that there exists an extremizer only when $\{a_1, \ldots, a_m\}$ is totally reducible for c. Note that we have also shown that this sum is orthogonal w.r.t. the scalar product given by the covariance of an extremizer.

We assume from now that $\{a_1, \ldots, a_m\}$ is totally reducible for c and that $\mathbb{R}^n = V_{J_1} \oplus \cdots \oplus V_{J_k}$ is the corresponding decomposition of $\mathbb{R}^n$ from Definition 3.8. We can assume that $|J_1| \leq |J_2| \leq \ldots \leq |J_k|$. Let T be any symmetric positive operator on $\mathbb{R}^n$ such that the following orthogonal decomposition holds:

$$\mathbb{R}^n = TV_{J_1} \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} TV_{J_k}.$$

Of course, there always exists such a linear map T. As before the change of vectors $X \to T^{-*}X$ and $a_j \to Ta_j$ reduces the problem to the case $T = \mathrm{Id}$ and

$$\mathbb{R}^n = V_{J_1} \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} V_{J_k}.$$

With this orthogonal decomposition in hand, we can use Lemma 3.7 to successively "peel-off" orthogonal blocks. We first apply this lemma to J_1 and $J_1^c = J_2 \cup \ldots \cup J_k$, and then on the space $V_{J_1^c} = V_{J_1}^\perp = V_{J_2} \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} V_{J_k}$ to J_2, and so on. After k steps we get that $D(A, c) = D(A_{J_1}, c_{J_1}) + \ldots + D(A_{J_k}, c_{J_k})$ and that a random vector X is an extremizer if and only if it can be written as

$$X = X_1 + \ldots + X_k$$

where X_i has values in V_{J_i} and is extremal for (A_{J_i}, c_{J_i}), and with the property that

$$X_i \text{ is independent of } (X_{i+1}, \ldots, X_k), \quad \text{for } i = 1, \ldots, k-1. \tag{3.43}$$

(Note that in order to construct and extremizer X we start with an extremizer X_k on V_{J_k} and, then add an extremal independent X_{k-1} on V_{J_k} in order to get an extremizer on $V_{J_{k-1}} \overset{\perp}{\oplus} V_{J_k}$, and so on by repeated applications of Lemma 3.7).

Observe that the independence property (3.43) is equivalent to the independence of the set of random vectors $\{X_1, \ldots, X_k\}$. Next remember that for each $i \leq m$ we have $c_{J_i} \in K_{J_i}^\circ$. Thus Lemma 3.6 applies and when $\dim(V_{J_i}) > 1$ then X_i is Gaussian and its variance is imposed as stated. Recall that in dimension 1 the problem is trivial and all random variables are extremal (in particular Gaussian variables are extremal). □

Note that the previous theorem tells us in particular that when optimizers exist, Gaussian optimizers exist (however this was not a necessary step in our approach).

Of course, by Theorems 2.1 and 2.2, we now also know that optimizers for the classical Brascamp–Lieb inequality exist under the exact same conditions for optimality described in Theorem 3.9, and that moreover, the optimizers Brascamp–Lieb inequality are exactly the marginals of the optimizing probability densities for the subadditivity inequality. The full description of optimizers (in one-dimensional Brascamp–Lieb inequalities) was given in [CLL1], building on a previous characterization by Barthe [Ba]. In the multi-dimensional case, building on Barthe's work too, Bennett–Carbery–Christ–Tao [BeCCT1] obtained some description, but the problem was completely solved only recently by Valdimarsson [V].

4   Consequences of the General Subadditivity Inequality in $\mathbb{R}^n$

There are several interesting consequences of Theorems 3.1 and 3.9. The first is a generalization of Hadamard's inequality for determinants.

Theorem 4.1. *Consider any family $A = \{a_1, \ldots, a_m\}$ of m vectors that span $\mathbb{R}^n$, any set of numbers $\{c_1, \ldots, c_m\}$ with $0 \leq c_i \leq 1$. Then with $D(A, c)$ as above, for any linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^n$,*

$$|\det(T)| \leq e^{D(A,c)} \left(\prod_{j=1}^m |T(a_j)|^{c_j} \right), \tag{4.1}$$

and this inequality is sharp in that the constant $e^{D(A,c)}$ cannot be decreased. Moreover, for $c \in K_A^\circ$, there is transformation T with $\det(T) = 1$ for which equality holds in (4.1), and, when $n \geq 2$, if we take T to be positive, then T is unique (up to multiplication by a positive scalar).

REMARK. In the case that $m = n$, and the vectors $\{a_1, \ldots, a_m\}$ are an orthonormal basis, and $c_1 = \cdots = c_n = 1$, this reduces Hadamard's inequality for determinants. In the special case $\sum_{m=1}^m c_j a_j \otimes a_j = \mathrm{Id}_{\mathbb{R}^n}$, so that $D(A, c) = 0$, this result has been proved by Ball [B], with a very simple proof.

For simplicity we have stated the existence of an extremal T only when $c \in K_A^\circ$, but the right condition is that A is totally reducible for c, just as in Theorem 3.9.

Proof. By making a polar decomposition, we may assume without loss of generality that T is positive definite. Let G_T be the centered Gaussian random variable with $\mathrm{E}(u \cdot G_T) = |T(u)|^2$ for all vectors u in $\mathbb{R}^n$. Then simply evaluating the left-hand side of $\sum_{j=1}^m c_j S(a \cdot G_T) - S(G_T) \leq D(A, c)$, we obtain (4.1). Then Theorem 3.1 provides the rest. □

Theorem 4.1 gives us one simple variational expression for $D(A, c)$, namely
$$D(A, c) = \sup \left\{ \ln \left(\frac{|\det(T)|}{\prod_{j=1}^{m} |T(a_j)|^{c_j}} \right) : T \text{ positive definite} \right\}.$$
There is however a simpler variational formula for $D(A, c)$ over an even lower-dimensional space, as suggested by the fact that $e^{D(A,c)}$ is also the sharp constant in the Brascamp–Lieb inequality. By the classical theorem of Brascamp and Lieb, $e^{D(A,c)}$ may be computing by taking the functions $\{f_1, \ldots, f_m\}$ in the Brascamp–Lieb inequality to be centered Gaussians; i.e.
$$\left\{ f_1(t), \ldots, f_m(t) \right\} = \left\{ e^{-(s_1 t)^2}, \ldots, e^{-(s_m t)^2} \right\},$$
and varying the m numbers $s_1, \ldots, s_m$. This leads directly to the variational expression (4.2) for $D(A, c)$. Let us recall that the existence of optimizers for this problems was proved by Brascamp and Lieb [BrL] under the hypothesis that every set of n vectors chosen from $\{a_1, \ldots, a_m\}$ is linearly independent and later proved by Barthe [Ba] for $c \in K_A^\circ$. The next theorem gives the complete result. Although the variational formula (4.2) can be deduced by duality, we give a direct proof of it starting from the subadditivity inequality.

Theorem 4.2. *Consider any set $\{a_1, \ldots, a_m\}$ of m vectors that span $\mathbb{R}^n$, $n \geq 2$. Let $\{c_1, \ldots, c_m\}$ be any set of numbers with $0 \leq c_i \leq 1$ verifying (3.7). Let T denote the $m \times m$ diagonal matrix whose jth diagonal entry is t_j, and define the function $\Phi_A(t_1, \ldots, t_m)$ by*
$$\Phi_A(t_1, \ldots, t_m) = \ln \det(A e^T A^t).$$
This is a convex function on $\mathbb{R}^m$, and
$$D(A, c) + \sum_{j=1}^{m} c_j \ln(c_j) = \frac{1}{2} \sup_{\{t_1, \ldots, t_m\}} \left(\sum_{j=1}^{m} c_j t_j - \Phi_A(t_1, \ldots, t_m) \right). \qquad (4.2)$$
The supremum in (4.2) is attained if and only if $\{a_1, \ldots, a_m\}$ is totally reducible for c. Moreover,
$$\Phi_A(t_1, \ldots, t_m) = \sup_{\{c_1, \ldots, c_m\}} \left(\sum_{j=1}^{m} c_j t_j - 2 \left(D(A, c) + \sum_{j=1}^{m} c_j \ln(c_j) \right) \right). \qquad (4.3)$$

Proof. For an $m \times m$ diagonal matrix S with positive entries s_j, introduce $R_S := ((AS)(AS)^t)^{-1/2}$ as in the proof of Lemma 3.5. Let G be a standard Gaussian random vector on $\mathbb{R}^n$ (i.e. $G \in \mathcal{N}(0, \mathrm{Id})$), and set $G_S = R_S G$. Then
$$\sum_{j=1}^{m} c_j S(a_j \cdot G_S) - S(G_S) = -\ln \left(\det(R_S^{-1}) \right) - \frac{1}{2} \sum_{j=1}^{m} c_j \ln \left(|R_S a_j|^2 \right).$$
However,
$$|R_S a_j|^2 = s_j^{-2} |R_S(s_j a_j)|^2 = s_j^{-2} |R_S(SA)e_j|^2 = s_j^{-2} e_j \cdot (AS)^t \left((AS)(AS)^t \right)^{-1} (AS)e_j,$$
where e_j is the jth standard basis vector in $\mathbb{R}^m$. Recall that $e_j(AS)^t((AS)(AS)^t)^{-1} \cdot (AS)e_j$ is the jth diagonal entry of the orthogonal projection in $\mathbb{R}^m$ onto the image of $(AS)^t$. Since this orthogonal projection has rank n, its trace is n. Therefore, if we define $c_j(S) = |R_S(SA)e_j|^2$, we have
$$\sum_{j=1}^{m} c_j(S) = n$$

for all S. Thus, by Jensen's inequality,

$$\sum_{j=1}^{m} c_j \ln(c_j) \geq \sum_{j=1}^{m} c_j \ln\big(c_j(S)\big),$$

with equality exactly when $c_j(S) = c_j$ for all j. Therefore, for all S,

$$D(A,c) \geq \sum_{j=1}^{m} c_j S(a_j \cdot G_S) - S(G_S) \geq -\ln\big(\det(R_S^{-1})\big) + \frac{1}{2}\sum_{j=1}^{m} c_j \ln(s_j^2) - \frac{1}{2}\sum_{j=1}^{m} c_j \ln(c_j).$$

so that

$$D(A,c) + \sum_{j=1}^{m} c_j \ln(c_j) \geq \frac{1}{2}\bigg(\sum_{j=1}^{m} c_j \ln(s_j^2) - \ln \det(R_S^{-2})\bigg).$$

Moreover, as we see from the proof of Lemma 3.5 (based on an observation by Barthe) and Lemma 3.6 and the remarks made just above, there is equality when $c \in K_A^\circ$ and $S = S_0$ is the choice of S (unique up to a multiple) for which (3.20) is true. Let T denote the $m \times m$ diagonal matrix whose jth diagonal entry is $t_j = \ln s_j^2$. Then $\ln(\det(R_S^{-2})) = \ln(\det(Ae^T A^t))$ and therefore, if we define the function Φ_A by

$$\Phi_A(t_1, \ldots, t_m) = \ln \det(Ae^T A^t),$$

we have, for every $t_1, \ldots, t_m \in \mathbb{R}$,

$$2\bigg(D(A,c) + \sum_{j=1}^{m} c_j \ln(c_j)\bigg) + \Phi_A(t_1, \ldots, t_m) \geq \sum_{j=1}^{m} c_j t_j \qquad (4.4)$$

with equality, when $c \in K_A^\circ$ for some choice of t_j's. The function $c \to 2D(A,c) + 2\sum_{j=1}^{m} c_j \ln(c_j)$ is convex (because, as mentioned at the beginning of the previous section, the function $c \to D(a,c)$ is convex by definition), and its domain (i.e. where it is $< +\infty$) is K_A. Therefore we get that

$$\Phi_A(t_1, \ldots, t_m) = \sup_{c \in K_A^\circ} \bigg\{ \sum_{j=1}^{m} c_j t_j - 2\bigg(D(A,c) + \sum_{j=1}^{m} c_j \ln(c_j)\bigg)\bigg\} = \sup_{c \in K_A} \{\ldots\}$$

$$= \sup_{c \in \mathbb{R}^m} \{\ldots\}.$$

This shows that Φ_A is convex on $\mathbb{R}^m$ and that it is the Legendre transform of the convex function $c \to 2D(A,c) + 2\sum_{j=1}^{m} c_j \ln(c_j)$.

Moreover, for given A and c, equality in (4.4) for some $t_1, \ldots, t_m$ means that for the corresponding values $s_1, \ldots, s_m$, the Gaussian G_S is an extremizer for the variational problem defining $D(A,c)$. By Theorem 3.9, tis means that $\{a_1, \ldots, a_m\}$ is totally reducible for c.

Conversely, if $\{a_1, \ldots, a_m\}$ is totally reducible for c, then the variational problem in (4.2) splits into a sum of independent and orthogonal (after a suitable linear transformation T) such problems, but of the interior type (i.e. $c \in K_{TA}^\circ$) for which Barthe showed optimizers to exist. Equivalently, the next Theorem 4.3 ensures that we can find a positive operator R for which the decomposition of the identity (3.19) holds. Then, as mentioned in the remark after the proof of Lemma 3.6, the random vector RG is extremal for $D(A,c)$ and setting $s_j^2 = c_j/|Ra_j|^2$ we have that $R = R_S$ and $c_j(S) = c_j$ by construction (see the proof of Lemma 3.5). This guarantees equality at all steps of our computation above and thus ensures equality in (4.4). $\quad\square$

REMARK. We have proved that

$$D(A,c) + \sum_{j=1}^{m} c_j \ln(c_j) = \frac{1}{2}\Phi_A^*(c)$$

where Φ_A^* denotes the Legendre transform of Φ_A. Since $\nabla\Phi_A^*(\nabla\Phi_A(0)) = 0$, the choice $c = \nabla\Phi_A(0)$ minimizes $\Phi_A^*(c)$, and hence $D(A,c) + \sum_{j=1}^{m} c_j \ln(c_j)$. There is a misprint in [CLL1] in which it is stated (in slightly different notation) that this choice of c minimizes $D(A,c)$ itself.

We finally return to Lemma 3.5, as we are now in a position to give necessary and sufficient conditions for the existence of the change of variables provided there.

Let $A = \{a_1,\ldots,a_m\}$ be a family of m vectors spanning $\mathbb{R}^n$, and let c be any vector in $\mathbb{R}^m$ with $0 \le c_j \le 1$ for all j. Theorem 3.9 gives us necessary and sufficient conditions for the existence of an extremal X for the subadditivity inequality. By Theorem 2.2, these conditions are also necessary and sufficient for the existence of extremals for the Brascamp–Lieb inequality. Moreover, we see that extremals for the latter exist if and only if centered Gaussian extremals exist.

From here, it is easy to prove the following theorem which supersedes Lemma 3.5, and gives necessary and sufficient conditions for the existence of the change of variables considered there. This result was obtained (in the more general multi-dimensional setting) by Bennett–Carbery–Christ–Tao [BeCCT1] along their study of the Brascamp–Lieb extremizers; here we use the extremizers to the subadditivity of entropy inequality. Though this theorem concerns a problem in linear algebra, we do not know a direct proof of it in a purely linear algebra context, though there may be one.

Theorem 4.3. *Let $\{a_1,\ldots,a_m\}$ be any collection of m vectors that span $\mathbb{R}^n$ for $n \ge 2$. Let $\{c_1,\ldots,c_m\}$ be any m numbers satisfying $0 \le c_j \le 1$ for each j. Then there exists an an invertible symmetric matrix $n \times n$ matrix R so that*

$$\sum_{i=1}^{m} c_j \left(\frac{Ra_j}{|Ra_j|}\right) \otimes \left(\frac{Ra_j}{|Ra_j|}\right) = \mathrm{Id}_{\mathbb{R}^n}$$

if and only if the set $\{a_1,\ldots,a_m\}$ is totally reducible for c

Proof. The proof of Lemma 3.6 shows that whenever such a matrix R exists, there exists an optimizer for the subadditivity inequality. Thus, by Theorem 3.9, the condition that $\{a_1,\ldots,a_m\}$ is totally reducible for c is necessary.

Conversely assume that $\{a_1,\ldots,a_m\}$ is totally reducible for c and that $\mathbb{R}^n = V_{J_1} \oplus \cdots \oplus V_{J_k}$ is the corresponding decomposition of $\mathbb{R}^n$ from Definition 3.8. We can then find an invertible operator T on $\mathbb{R}^n$ such that the following orthogonal decomposition holds

$$\mathbb{R}^n = TV_{J_1} \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} TV_{J_k}\,.$$

Since we have $c_{J_i} \in K_{A_{J_i}}^\circ = K_{TA_{J_i}}^\circ$ for $i \le m$ (with $TA_{J_i} = [Ta_j, j \in J_i]$), we may use Lemma 3.5 on each of the reduced orthogonal subspaces TV_{J_i}; this gives us some symmetric invertible operator R_i on TV_{J_i} and putting all these operators together

we get a symmetric invertible operator R on $\mathbb{R}^n$ such that

$$\sum_{i=1}^{m} c_j \left(\frac{RTa_j}{|RTa_j|} \right) \otimes \left(\frac{RTa_j}{|RTa_j|} \right) = \mathrm{Id}_{\mathbb{R}^n} \, .$$

Then the positive symmetric operator $\tilde{R} = \sqrt{T^* R^2 T}$ satisfies the desired property

$$\sum_{i=1}^{m} c_j \left(\frac{\tilde{R}a_j}{|\tilde{R}a_j|} \right) \otimes \left(\frac{\tilde{R}a_j}{|\tilde{R}a_j|} \right) = \mathrm{Id}_{\mathbb{R}^n} \, . \qquad \square$$

5 A Convolution Inequality for Eigenvalues

We investigate here the dual of the superadditivity of Fisher information inequality (3.16) in Proposition 3.3.

In section 2 we have shown that the Legendre transform of the entropy provides an equivalence between subadditivity of the entropy and Brascamp–Lieb inequalities. It turns out that the Fisher information is also a convex functional and its Legendre transform is known to be the smallest eigenvalue of a Schrödinger operator. (This is used extensively in the theory of large deviations, for example). We shall use this fact to derive a subadditivity of the smallest eigenvalues of Schrödinger operators.

For any continuous bounded function V on $\mathbb{R}^n$, define

$$\lambda(V) = \sup \left\{ \int_{\mathbb{R}^n} V(x)\phi^2(x)\,\mathrm{d}x - 4 \int_{\mathbb{R}^n} \left| \nabla \phi(x) \right|^2 : \int_{\mathbb{R}^n} \phi^2(x)\,\mathrm{d}x = 1 \right\} . \qquad (5.1)$$

Then $-\lambda(V)$ is the "ground state" eigenvalue of

$$-4\Delta - V \, ,$$

provided the bottom of the spectrum is an eigenvalue, and in any case, it is the bottom of the spectrum.

Then since

$$I(f) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f}\,\mathrm{d}x = 4 \int_{\mathbb{R}^n} \left| \nabla \sqrt{f} \right|^2 \mathrm{d}x \, ,$$

we can rewrite (5.1) as

$$\lambda(V) = \sup \left\{ \int_{\mathbb{R}^n} V(x)f(x)\,\mathrm{d}x - I(f) \right\} ,$$

where the supremum is taken over all probability densities f. This gives us the analog of (2.4) for Fisher information:

$$\int_{\mathbb{R}^n} V(x)f(x)\,\mathrm{d}x \le \lambda(V) + I(f) \, , \qquad (5.2)$$

with equality if and only if $f = \phi^2$ where $(-4\Delta - V)\phi = -\lambda(V)\phi$. (Here, by the definition (5.1) of $\lambda(V)$, ϕ is the "ground state" eigenfunction.

Now let $V_1, \dots V_n$ be continuous functions on $\mathbb{R}$, and define

$$V = \sum_{j=1}^{n} V_j(u_j \cdot x)$$

where $\{u_1, \dots, u_n\}$ is any orthonormal basis for $\mathbb{R}^n$. Then

$$-4\Delta - V = \sum_{j=1}^{n} \left(-4(u_j \cdot \nabla)^2 - V_j(u_j \cdot x) \right) ,$$

so that, by separation of variables,

$$\lambda(V) = \sum_{j=1}^{n} \lambda(V_j) \,.$$

The following result generalizes this to the case in which we have m unit vectors $\{u_1, \ldots, u_m\}$ satisfying (3.13).

Theorem 5.1. *Let $\{u_1, \ldots, u_m\}$ be any m unit vectors in $\mathbb{R}^n$ such that there are positive numbers $c_1, \ldots, c_m$ satisfying*

$$\sum_{j=1}^{m} c_j u_j \otimes u_j = \mathrm{Id}_{\mathbb{R}^n} \,.$$

For any m continuous bounded functions $V_1, \ldots, V_m$ on $\mathbb{R}$, define on $\mathbb{R}^n$

$$V(x) = \sum_{j=1}^{m} V(u_j \cdot x) \,.$$

Then

$$\lambda(V) \leq \sum_{j=1}^{m} c_j \lambda\left(\frac{1}{c_j} V_j\right) . \tag{5.3}$$

Proof. Choose an $\epsilon > 0$ and a probability density $f = \phi^2$ such that

$$\int_{\mathbb{R}^n} V(x)\phi^2(x)\,\mathrm{d}x - 4 \int_{\mathbb{R}^n} \left|\nabla\phi(x)\right|^2 \geq \lambda(V) - \epsilon \,.$$

Then using (3.16),

$$\begin{aligned}
\lambda(V) - \epsilon &\leq \int_{\mathbb{R}^n} f(x) \left(\sum_{j=1}^{m} V_j(u_j \cdot x)\right) \mathrm{d}x - I(f) \\
&= \sum_{j=1}^{m} \int_{\mathbb{R}} f_{(u_j)}(t) V_j(t)\,\mathrm{d}t - I(f) \\
&\leq \sum_{j=1}^{m} \int_{\mathbb{R}} f_{(u_j)}(t) V_j(t)\,\mathrm{d}t - \sum_{j=1}^{m} c_j I\left(f_{(u_j)}\right) \\
&= \sum_{j=1}^{m} c_j \left(\int_{\mathbb{R}} f_{(u_j)}(t) \left(\frac{1}{c_j} V_j\right)(t)\,\mathrm{d}t - I\left(f_{(u_j)}\right)\right) \\
&\leq \sum_{j=1}^{m} c_j \lambda\left(\frac{1}{c_j} V_j\right) ,
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, this proves the result.   □

The inequality (5.3) is sharp since one can use another Legendre transform, as in the proof of Theorem 2.1, and see that it implies the sharp inequality (3.16). Inequality (5.3) could also be proved using a semi-group (or stochastic) method inspired by the one used by Borell [Bo] in his study of Brunn–Minkowski type inequalities (which, somehow, are the *converse* of the inequalities considered here); this would be more complicated than starting from the inequality (3.16) for the Fisher information, though.

An analogous result for functions on the sphere could be given using the sharp superadditivity of Fisher information inequality proved in [BaCM].

References

[B]   K. Ball, Convex geometry and functional analysis, in "Handbook of the Geometry of Banach Spaces, Vol. I, (W. Johnson, J. Lindenstrauss, eds.), North-Holland, Amsterdam (2001), 161–194.

[Ba]   F. Barthe, On a reverse form of the Brascamp–Lieb inequality, Invent. Math. 134:2 (1998), 235–361.

[BaC]   F. Barthe, D. Cordero–Erausquin, Inverse Brascamp–Lieb inequalities along the heat equation, Geometric Aspects of Functional Analysis (2002–2003) (V. Milman, G. Schechtman, eds.), Springer Lecture Notes in Mathematics 1850 (2004), 65–71.

[BaCM]   F. Barthe, D. Cordero–Erausquin, B. Maurey, Entropy of spherical marginals and related inequalities, J. Math. Pures Appl. 86:2 (2006), 89–99.

[BaCLM] F. Barthe, D. Cordero–Erausquin, M. Ledoux, B. Maurey, work in progress).

[BeCCT1] J. Bennett, A. Carbery, M. Christ, T. Tao, The Brascamp–Lieb inequalities: finiteness, structure, and extremals, GAFA, Geom. funct. anal. 17:5 (2007), 1343–1415.

[BeCCT2] J. Bennett, A. Carbery, M. Christ, T. Tao, Finite bounds for Holder–Brascamp–Lieb multilinear inequalities, Math. Res. Lett, to appear.

[Bo]   C. Borell, Diffusion equations and geometric inequalities, Potential Anal. 12:1 (2000), 49–71.

[BrL]   H. Brascamp, E. Lieb, The best constant in Young's inequality and its generalization to more than three functions, Advances in Math. 20:2 (1976), 151–173.

[C]   E. Carlen, Superadditivity of Fisher information and logarithmic Sobolev inequalities, J. Funct. Anal. 101 (1991), 194–211.

[CLL1]   E. Carlen, M. Loss, E. Lieb, A sharp form of Young's inequality on S^N and related entropy inequalities, Jour. Geom. Anal. 14 (2004), 487–520.

[CLL2]   E. Carlen, M. Loss, E. Lieb, A inequality of Hadamard type for permanents, Meth. and Appl. of Anal. 13:1 (2006), 1–17.

[V]   S.I. Valdimarsson, Optimisers for the Brascamp–Lieb inequality, Israel J. Math. 168 (2008), 253–274,

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

Dario Cordero–Erausquin, Institut de Mathématiques de Jussieu, Université Pierre et Marie Curie - Paris 6, 4 Place Jussieu, 75252 Paris Cedex 05, France
                                             cordero@math.jussieu.fr