

## MIT Open Access Articles

### *Euclidean Forward–Reverse Brascamp–Lieb Inequalities: Finiteness, Structure, and Extremals*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**As Published:** <https://doi.org/10.1007/s12220-020-00398-y>

**Publisher:** Springer US

**Persistent URL:** <https://hdl.handle.net/1721.1/132089>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



## Euclidean Forward–Reverse Brascamp–Lieb Inequalities: Finiteness, Structure, and Extremals

**Cite this article as:** Thomas A. Courtade and Jingbo Liu, Euclidean Forward–Reverse Brascamp–Lieb Inequalities: Finiteness, Structure, and Extremals, The Journal of Geometric Analysis <https://doi.org/10.1007/s12220-020-00398-y>

This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

Author accepted manuscript

## EUCLIDEAN FORWARD-REVERSE BRASCAMP-LIEB INEQUALITIES: FINITENESS, STRUCTURE AND EXTREMALS

THOMAS A. COURTADE AND JINGBO LIU

ABSTRACT. A new proof is given for the fact that centered gaussian functions saturate the Euclidean forward-reverse Brascamp-Lieb inequalities, extending the Brascamp-Lieb and Barthe theorems. A duality principle for best constants is also developed, which generalizes the fact that the best constants in the Brascamp-Lieb and Barthe inequalities are equal. Finally, as the title hints, the main results concerning finiteness, structure and gaussian-extremizability for the Brascamp-Lieb inequality due to Bennett, Carbery, Christ and Tao are generalized to the setting of the forward-reverse Brascamp-Lieb inequality.

### 1. INTRODUCTION AND MAIN RESULTS

We begin with notation that will prevail throughout. Let  $(E_i)_{1 \leq i \leq k}$  and  $(E^j)_{1 \leq j \leq m}$  be Euclidean spaces, i.e., finite-dimensional Hilbert spaces endowed with Lebesgue measure and the usual inner product  $\langle \cdot, \cdot \rangle$  giving rise to Euclidean length  $|\cdot|$ . We write  $E_0 = \bigoplus_{i=1}^k E_i$ , and let  $\pi_{E_i} : E_0 \rightarrow E_i$  be the orthogonal projection of  $E_0$  onto **subspace**  $E_i$ .

Let  $\mathbf{B} := (B_{ij})_{1 \leq i \leq k, 1 \leq j \leq m}$ , where each  $B_{ij} : E_i \rightarrow E^j$  is a bounded linear transformation. Because it will be referred to frequently, we define  $B_j : E_0 \rightarrow E^j$  according to

$$B_j x := \sum_{i=1}^k B_{ij} \pi_{E_i}(x), \quad x \in E_0.$$

Note that the collection  $(B_j)_{1 \leq j \leq m}$  may be regarded as an equivalent characterization of  $\mathbf{B}$ . Define  $\mathbf{B}^* := (B_{ij}^*)_{1 \leq i \leq k, 1 \leq j \leq m}$ , where  $A^*$  denotes the adjoint of  $A$ .

We let  $\mathbf{c} := (c_i)_{1 \leq i \leq k}$  and  $\mathbf{d} := (d_j)_{1 \leq j \leq m}$  be collections of positive real numbers satisfying

$$\sum_{i=1}^k c_i \dim(E_i) = \sum_{j=1}^m d_j \dim(E^j), \tag{1}$$

and we refer to the triple  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  as a *datum*. Finally,  $\mathbb{R}^+$  denotes the non-negative real numbers.

**1.1. The forward-reverse Brascamp-Lieb inequalities.** For a given datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , this paper is concerned with characterizing the best constant  $D$  in the following statement: If measurable functions  $f_i : E_i \rightarrow \mathbb{R}^+$ ,  $1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+$ ,  $1 \leq j \leq m$  satisfy

$$\prod_{i=1}^k f_i^{c_i}(x_i) \leq \prod_{j=1}^m g_j^{d_j} \left( \sum_{i=1}^k c_i B_{ij} x_i \right) \quad \forall x_i \in E_i, \quad 1 \leq i \leq k, \tag{2}$$

then

$$\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i} \leq e^D \prod_{j=1}^m \left( \int_{E^j} g_j \right)^{d_j}, \tag{3}$$

---

2010 *Mathematics Subject Classification*. Primary 26D15, 39B72; Secondary 47G10.

This work was supported by NSF grants CCF-1704967, CCF-0939370 and CCF-1750430.

where the integrals are with respect to Lebesgue measure on the respective spaces. To facilitate later referencing, we make a formal definition.

**Definition 1.1.** Given a datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , we define  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  to be the smallest constant  $D$  such that (3) holds for all nonnegative measurable functions satisfying the constraints (2).

**Remark 1.2.** If (1) does not hold, then dilating all functions by a common factor shows  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) = +\infty$ , motivating the assumption. It is easy to see that  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) > -\infty$ .

The above class of inequalities was introduced by the authors together with Cuff and Verdú, and termed *Forward-Reverse Brascamp-Lieb inequalities* [25]. This choice of terminology reflects the observation that taking  $k = 1$  and  $c_1 = 1$  specializes to the classical (forward, or direct) Brascamp-Lieb inequalities [11, 12, 23]; on the other hand, taking  $m = 1$  and  $d_1 = 1$  specializes to the reverse form of the Brascamp-Lieb inequalities introduced by Barthe [4].

The celebrated result of Lieb [23] is that in the case  $k = c_1 = 1$ , the best constant  $D(1, \mathbf{d}, \mathbf{B})$  can be computed by considering only centered gaussian functions  $f_1, g_1, \dots, g_m$ . Likewise, Barthe showed in [4] that in the case of  $m = d_1 = 1$ , the best constant  $D(\mathbf{c}, 1, \mathbf{B})$  can be computed by considering only centered gaussian functions  $f_1, \dots, f_k, g_1$ . Barthe also established a remarkable duality between the forward and reverse Brascamp-Lieb inequalities, in the sense that

$$D(\mathbf{c}, 1, \mathbf{B}) = D(1, \mathbf{c}, \mathbf{B}^*), \tag{4}$$

where, by Definition 1.1 applied to the datum  $(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$ , the quantity  $D(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$  denotes the smallest constant  $D$  in the inequality

$$\prod_{j=1}^m \left( \int_{E^j} g_j \right)^{d_j} \leq e^D \prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i},$$

holding for all measurable functions  $f_i : E_i \rightarrow \mathbb{R}^+$ ,  $1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+$ ,  $1 \leq j \leq m$  satisfying

$$\prod_{j=1}^m g_j^{d_j}(y_j) \leq \prod_{i=1}^k f_i^{c_i} \left( \sum_{j=1}^m d_j B_{ij}^* y_j \right) \quad \forall y_j \in E^j, 1 \leq j \leq m. \tag{5}$$

Perhaps surprisingly, the forward-reverse Brascamp-Lieb inequality suggests that there is no fundamental distinction between the traditional forward and reverse forms of the Brascamp-Lieb inequality. Indeed, they are each a particular instance of the inequality (3) under the hypothesis (2) for an appropriate choice of datum. Most importantly, the gaussian saturation property continues to hold for the forward-reverse Brascamp-Lieb inequality, as well as a full-fledged form of the duality relation (4). This both clarifies and unifies the general landscape of Euclidean Brascamp-Lieb-Barthe-type inequalities and the duality they enjoy. This is our first main result:

**Theorem 1.3.** *The quantities  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  and  $D(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$  can be computed by considering only centered gaussian functions  $(f_i)_{1 \leq i \leq k}$  and  $(g_j)_{1 \leq j \leq m}$  in their respective definitions. Moreover, it holds that*

$$D(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D(\mathbf{d}, \mathbf{c}, \mathbf{B}^*). \tag{6}$$

**Remark 1.4.** The sufficiency of considering gaussian functions for computing the constant  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  was already established in our previous work [25, Theorem 2]. As will be explained in Section 4, the gaussian saturation property is closely connected (in fact, formally equivalent) to a result announced by Barthe and Wolff in the note [5], and proved in their recent followup work [6]. The identity (6) has not been previously observed. **A related family of inequalities has also been considered by Cordero-Erasquin and Maurey [16]; their main result can be interpreted as the assertion  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq 0$ , under the assumption that  $\mathbf{B}$  is contractive in a certain sense (interested readers are referred to [16, Section 5] for the relevant statements).**

**Remark 1.5.** The identity (6) explains the scaling of each  $B_{ij}$  by  $c_i$  in (2) and, similarly, the scaling of each  $B_{ij}^*$  by  $d_j$  in (5). If we were not after (6), these scalar factors could be absorbed into the maps themselves without affecting the first claim of Theorem 1.3.

There are now several independent proofs of the original Brascamp-Lieb and Barthe theorems. Early proofs relied on rearrangement arguments [11,12], and Lieb appealed to rotational invariance of the extremizers [23]. Barthe came up with a clever optimal transport argument, and simultaneously proved both the forward and reverse inequality [4], further establishing equality of best constants. More recently, Lehec [22] gave a probabilistic proof of both theorems using a variational representation for functionals due to Boué and Dupuis [10]. Semigroup techniques provide yet another avenue of proof; see Bennett, Carbery, Christ and Tao [8], or Carlen, Lieb and Loss [14]. Our previous work [25] gave an information-theoretic proof of the gaussian saturation part of Theorem 1.3 (therefore extending to the classical settings as well), by way of a doubling argument similar to that employed by Geng and Nair in [18] for a different problem. This doubling argument is similar in spirit to that given by Lieb [23], but it exploited an equivalent entropic representation of the problem.

As far as applications go, it is well-known that the Brascamp-Lieb inequalities imply many other classical inequalities in analysis and geometry, such as Hölder’s inequality, Young’s inequality, and the Loomis-Whitney inequality. Likewise, Barthe’s inequality contains, for example, the Prékopa-Leindler and Brunn-Minkowski inequalities as special cases. All of these implications and more are described in Gardner’s survey of the Brunn-Minkowski inequality, which places the Brascamp-Lieb, Barthe, and reverse Young inequalities atop a hierarchy of implications [17, Figure 1], with none of the three evidently implying the others. In the accompanying discussion, Gardner asks whether stronger unifying inequalities await discovery; the content of Theorem 1.3 may be regarded as an affirmative answer. We have already described how the Brascamp-Lieb and Barthe inequalities may be immediately recognized as special cases of the forward-reverse Brascamp-Lieb inequality. It turns out that the reverse Young inequality constitutes another instance of the forward-reverse Brascamp-Lieb inequality. Further examples will be given in Section 4.

**Example 1.6.** Let  $0 < p, q, r \leq 1$  satisfy  $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ . For  $\phi, \psi$  non-negative measurable functions on  $\mathbb{R}^n$ , the reverse Young inequality for convolutions asserts

$$\|\phi * \psi\|_r \geq C^n \|\phi\|_p \|\psi\|_q, \tag{7}$$

where  $\|h\|_p := (\int_{\mathbb{R}^n} |h|^p dx)^{1/p}$  for  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $p \in \mathbb{R}$ . The sharp constant is given by  $C = C_p C_q / C_r$ , with

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

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

We may assume  $r < 1$  henceforth, else if  $r = 1$ , then we must have  $p = q = 1$ , and the claim is trivial. Under this assumption, it is easily verified using the reverse Hölder inequality and renaming functions, that (7) is equivalent to

$$\iint f_1^{1/p}(x-y) f_2^{1/q}(y) g_1^{1/r'}(x) dx dy \geq C^n \left( \int_{\mathbb{R}^n} f_1 \right)^{1/p} \left( \int_{\mathbb{R}^n} f_2 \right)^{1/q} \left( \int_{\mathbb{R}^n} g_1 \right)^{1/r'}, \tag{8}$$

holding for  $f_1, f_2, g_1$  non-negative measurable functions on  $\mathbb{R}^n$ .

To place this into the framework of the forward-reverse Brascamp-Lieb inequality, let  $E_1 = E_2 = E^1 = \mathbb{R}^n$  and  $E^2 = \mathbb{R}^{2n}$ , with  $c_1 = 1/p$ ,  $c_2 = 1/q$ ,  $d_1 = 1/r - 1$  and  $d_2 = 1$ . Further, let  $\mathbf{B}$  be such that

$$\sum_{i=1}^2 c_i B_{i1} x_i = x_1 + x_2; \quad \sum_{i=1}^2 c_i B_{i2} x_i = (x_1, x_2), \quad \forall x_1, x_2 \in \mathbb{R}^n.$$

Then, the hypothesis (2) boils down to

$$f_1^{1/p}(x_1)f_2^{1/q}(x_2) \leq g_1^{1/r-1}(x_1+x_2)g_2(x_1, x_2) \quad \forall x_1, x_2 \in \mathbb{R}^n. \quad (9)$$

For arbitrary functions  $f_1, f_2, g_1$ , the best function  $g_2$  can be computed as

$$g_2(x, y) := f_1^{1/p}(x-y)f_2^{1/q}(y)g_1^{1/r'}(x), \quad x, y \in \mathbb{R}^n,$$

where “best” is in the sense that the RHS of (3) is minimized subject to (9). On substituting this choice of  $g_2$  into (3) and rearranging, we are left precisely with (8), with best constant necessarily characterized as  $C^n = e^{-D(\mathbf{c}, \mathbf{d}, \mathbf{B})}$ , computed by considering only centered gaussian functions.

The relation (6) allows us to easily deduce an interesting “dual” to the reverse Young inequality, given in the following example. Here, we emphasize that the term dual is meant in terms of (6), which is the same sense in which the Brascamp-Lieb and Barthe inequalities are dual to one another. It bears a superficial resemblance to Maurey’s property  $(\tau)$  [26] and functional Santaló inequalities (e.g., [2, 20]).

**Example 1.7.** Let  $0 < p, q, r \leq 1$  satisfy  $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ , and let  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  be the datum of the previous example that yields the reverse Young inequality (8). By considering the dual datum  $(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$  and applying (6), we conclude after elementary simplification that

$$\int_{\mathbb{R}^{2n}} w \leq K^n \|f\|_p \|g\|_q \|h\|_{r'}, \quad (10)$$

for nonnegative measurable functions  $f, g, h, w$  satisfying

$$w(z_1, z_2) \leq \inf_{y \in \mathbb{R}^n} \left( f(z_1 - y)g(z_2 - y)h(y) \right), \quad z_1, z_2 \in \mathbb{R}^n. \quad (11)$$

The sharp constant  $K$  is given by  $K = K_p K_q K_r$ , where<sup>1</sup>

$$K_s^2 := |s|^{1/s} |s'|^{1/s'}, \quad 0 < s \leq +\infty.$$

The choice of functions

$$f(x) = e^{|x|^2/p'}, \quad g(x) = e^{|x|^2/q'}, \quad h(x) = e^{|x|^2/r}$$

saturate the inequality (10) when  $w$  is taken equal to the infimum in (11).

**Remark 1.8.** To be more precise, the instances in the above example where  $r < 1$  follow by (6), and the exceptional case  $r = 1$  ( $r' = -\infty$ ) can be easily checked directly.

**Remark 1.9.** In analogy to Example 1.7, one can derive a “dual” Young inequality. It formally reverses the inequality of the previous example: If  $p, q, r \geq 1$  satisfy  $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ , then

$$\int_{\mathbb{R}^{2n}} w \geq K^n \|f\|_p \|g\|_q \|h\|_{r'} \quad (12)$$

for nonnegative measurable functions  $f, g, h, w$  satisfying

$$w(z_1, z_2) \geq \sup_{y \in \mathbb{R}^n} \left( f(z_1 - y)g(z_2 - y)h(y) \right), \quad z_1, z_2 \in \mathbb{R}^n. \quad (13)$$

Since the standard Young inequality is a special case of the Brascamp-Lieb inequality, its dual (12) is a special case of the Barthe inequality, and should therefore be regarded as already known. In contrast, we do not know whether (10) (or equivalent) has appeared previously in the literature.

---

<sup>1</sup>If  $s \in \{1, +\infty\}$ ,  $K_s$  is defined by the limit  $K_s := \lim_{t \rightarrow s} K_t$  to avoid indeterminate forms in the definition.

To close this section, let us remark briefly on the chief difficulty encountered in proving Theorem 1.3 compared to the special cases corresponding to the Brascamp-Lieb and Barthe inequalities. In the case of the direct Brascamp-Lieb inequalities, the function  $f_1$  can be explicitly computed in terms of the  $(g_j)_{1 \leq j \leq m}$  (specifically,  $f_1 = \prod_{j=1}^m (g_j \circ B_j)^{d_j}$ ). In the reverse case, the function  $g_1$  can be computed explicitly in terms of the  $(f_i)_{1 \leq i \leq k}$ . Such a simplification is not typically possible in the more general forward-reverse inequalities; this leads us to establish a rather subtle duality principle, to be made precise in Theorem 2.11 (see Remark 2.13). Once this duality principle is established, the structure of gaussian extremizers can be distilled, and techniques previously developed for proving the Brascamp-Lieb and Barthe inequalities can be successfully adapted to forward-reverse setting.

**1.2. Finiteness and gaussian-extremizability.** A motivation of the present paper is to better understand the structural properties of the gaussian extremizers in Theorem 1.3 (when they exist), and to establish necessary and sufficient conditions for finiteness of  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . To this end, we give a new proof of Theorem 1.3, which combines ideas from Lehec’s probabilistic proof of the Brascamp-Lieb and Barthe inequalities [22], the structural viewpoint on the forward Brascamp-Lieb inequality developed by Bennett, Carbery, Christ and Tao [8], and the entropic duality of the forward-reverse Brascamp-Lieb inequality in [25]. The detailed structural results, for example, allow us to easily identify “geometric” instances of the forward-reverse Brascamp-Lieb inequality, which may be particularly useful in applications (see, e.g., Section 4.3). While the results below address the structure of gaussian extremizers, they do not characterize non-gaussian extremizers. Valdimarsson has done this for the direct Brascamp-Lieb inequalities [28], and it seems reasonable to speculate that his approach could be extended to the forward-reverse inequalities considered here. We leave this question to future work.

Let us now make precise the notions of gaussian-extremizability and extremizers that have been alluded to above. We will need some more notation, which will prevail throughout. For a Euclidean space  $E$ , we let  $S(E)$  denote the set of self-adjoint linear operators on  $E$ , and  $S^+(E)$  denote the set of self-adjoint positive-definite linear operators on  $E$ . That is,  $A \in S^+(E)$  if  $A \in S(E)$  and it further holds that  $\langle Ax, x \rangle > 0$  for all nonzero  $x \in E$ . If  $A \in S(E)$  and  $\langle Ax, x \rangle \geq 0$  for all  $x \in E$ , we say that  $A$  is positive-semidefinite. For  $A, B \in S(E)$ , we write  $A \geq B$  if  $A - B$  is positive-semidefinite. Finally, for positive-semidefinite  $A$ , we let  $A^{1/2}$  denote the unique positive-semidefinite  $M$  such that  $A = M^2$ .

A centered gaussian function (or kernel)  $g : E \rightarrow \mathbb{R}^+$  is a function of the form

$$g(x) = \exp\left(-\frac{1}{2}\langle A^{-1}x, x \rangle\right), \quad A \in S^+(E),$$

where  $A$  is said to be the *covariance* of the gaussian kernel  $g$ . We remark that a centered gaussian random vector on  $E$  with covariance  $A$  has density (with respect to Lebesgue measure on  $E$ ) proportional to  $g$ .

Restricting attention to centered gaussian functions

$$\begin{aligned} f_i &: x \in E_i \mapsto \exp(-\langle V_i x, x \rangle), \quad V_i \in S^+(E_i), 1 \leq i \leq k, \\ g_j &: x \in E^j \mapsto \exp(-\langle U_j x, x \rangle), \quad U_j \in S^+(E^j), 1 \leq j \leq m, \end{aligned}$$

and defining

$$\Lambda_{\mathbf{c}} := \bigoplus_{i=1}^k c_i \text{id}_{E_i},$$

we see that the hypothesis (2) boils down to

$$\sum_{j=1}^m d_j \langle U_j B_j \Lambda_{\mathbf{c}} x, B_j \Lambda_{\mathbf{c}} x \rangle \leq \sum_{i=1}^k c_i \langle V_i \pi_{E_i}(x), \pi_{E_i}(x) \rangle \quad \forall x \in E_0. \quad (14)$$

Additionally, using (1), we may compute

$$\frac{\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i}}{\prod_{j=1}^m \left( \int_{E^j} g_j \right)^{d_j}} = \frac{\prod_{j=1}^m \det(U_j)^{d_j/2}}{\prod_{i=1}^k \det(V_i)^{c_i/2}}.$$

Collecting the above and comparing to (3), this motivates definition of the quantity

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) := \frac{1}{2} \sup \left( \sum_{j=1}^m d_j \log \det U_j - \sum_{i=1}^k c_i \log \det V_i \right), \quad (15)$$

where the supremum is over all  $V_i \in S^+(E_i), 1 \leq i \leq k$  and  $U_j \in S^+(E^j), 1 \leq j \leq m$  satisfying (14). In words,  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is the smallest possible constant  $D$  in (3), holding for all gaussian kernels satisfying the constraints (2). By definition,  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . The first part of Theorem 1.3 asserts that this is always met with equality.

**Remark 1.10.** Even if (1) does not hold, it remains true that  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  as defined above will equal the smallest possible constant  $D$  in (3), holding for all gaussian kernels satisfying the constraints (2). Indeed, scaling the  $(V_i)_{1 \leq i \leq k}$  and  $(U_j)_{1 \leq j \leq m}$  by a common factor shows  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = +\infty$ , while dilating functions by a common factor will show that the best constant  $D$  in (3) will also be equal to  $+\infty$ .

**Definition 1.11.** A datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is *gaussian-extremizable* if the supremum in (15) is attained and is finite for some  $V_i \in S^+(E_i), 1 \leq i \leq k$  and  $U_j \in S^+(E^j), 1 \leq j \leq m$  satisfying (14). Any such operators are said to be *gaussian extremizers*.

The constraint (14) is a bit cumbersome to write out. So, henceforth, we adopt some notation to make statements more compact; for given  $V_i \in S^+(E_i), 1 \leq i \leq k$  and  $U_j \in S^+(E^j), 1 \leq j \leq m$ , we define  $V_{\mathbf{c}} : E_0 \rightarrow E_0$  and  $U_{\mathbf{d}} : \bigoplus_{j=1}^m E^j \rightarrow \bigoplus_{j=1}^m E^j$  according to

$$V_{\mathbf{c}} := \bigoplus_{i=1}^k c_i V_i; \quad U_{\mathbf{d}} := \bigoplus_{j=1}^m d_j U_j.$$

Note that this does not cause any ambiguity since  $V_i \in S^+(E_i), 1 \leq i \leq k$  defines  $V_{\mathbf{c}}$ , and vice versa. Similarly for  $U_{\mathbf{d}}$ . The constraints (14) may now be concisely written as the operator inequality

$$\Lambda_{\mathbf{c}} B^* U_{\mathbf{d}} B \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}}, \quad (16)$$

where  $B : E_0 \rightarrow \bigoplus_{j=1}^m E^j$  is the linear operator defined according to

$$B : x \mapsto B_1 x + \dots + B_m x; \quad x \in E_0.$$

We introduce one last piece of notation before our characterization of gaussian-extremizability:

**Definition 1.12.** For  $A_i \in S^+(E_i), i = 1, \dots, k$ , we let  $\Pi(A_1, \dots, A_k)$  denote the set of positive-semidefinite  $A \in S(E_0)$  satisfying

$$\langle A x_i, x_i \rangle = \langle A_i x_i, x_i \rangle \text{ for all } x_i \in E_i, 1 \leq i \leq k.$$

**Remark 1.13.** The above definition has a natural interpretation in terms of couplings. Consider a gaussian random vector  $X_i$  taking values in  $E_i$ , with covariance  $A_i, 1 \leq i \leq k$ . If a jointly gaussian coupling of  $(X_i)_{1 \leq i \leq k}$  has covariance  $A$ , then  $A \in \Pi(A_1, \dots, A_k)$ . Conversely, each  $A \in \Pi(A_1, \dots, A_k)$  corresponds to the covariance of a jointly gaussian coupling of  $(X_i)_{1 \leq i \leq k}$ .



**Theorem 1.14** (Structure of gaussian extremizers). *A datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable if and only if (1) holds and there are  $V_i \in S^+(E_i), 1 \leq i \leq k$  and  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  satisfying*

$$\sum_{j=1}^m d_j \Lambda_{\mathbf{c}} B_j^* (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1} B_j \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}}. \quad (17)$$

Moreover, any such  $(V_i)_{1 \leq i \leq k}$  together with  $U_j = (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1}, 1 \leq j \leq m$ , are gaussian extremizers.

**Remark 1.15.** Implicit in (17) is the assertion that the stated inverses exist; it is therefore necessary for each  $(B_j)_{1 \leq j \leq m}$  to be surjective in order for the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  to be gaussian extremizable. Moreover, after left- and right-multiplying both sides of (17) by  $\Pi^{1/2}$ , the traces of the respective sides will be equal by (1). Hence, (17) is met with equality on restriction to the subspace  $\Pi E_0$ . See also Remark 2.14 and Proposition 2.15 for more along these lines.

**Remark 1.16.** In view of the previous remark, for the classical setting of  $k = c_1 = 1$ , gaussian-extremizability reduces to (1) and existence of  $V \in S^+(E_1)$  such that

$$\sum_{j=1}^m d_j B_j^* (B_j V^{-1} B_j^*)^{-1} B_j = V.$$

This has been repeatedly observed in previous proofs of the Brascamp-Lieb and Barthe inequalities.

The *geometric Brascamp-Lieb inequalities* proposed by Ball [3] and later generalized by Barthe [4] are a special case of the (forward) Brascamp-Lieb inequalities for which the linear maps are isometric and  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) = 0$ . The class of geometric Brascamp-Lieb inequalities are useful in applications (see, e.g., the volume ratio inequalities due to K. Ball), and are formally equivalent to the class of gaussian-extremizable Brascamp-Lieb inequalities [8, Proposition 3.6]. The following is a generalization to the forward-reverse setting:

**Corollary 1.17** (Geometric forward-reverse Brascamp-Lieb inequality (I)). *Assume (1) holds, and let linear maps  $Q_j : E_0 \rightarrow E^j$  and  $\Sigma \in \Pi(\text{id}_{E_1}, \dots, \text{id}_{E_k})$  satisfy*

$$Q_j \Sigma Q_j^* = \text{id}_{E^j} \text{ for each } 1 \leq j \leq m, \quad \text{and} \quad \sum_{j=1}^m d_j Q_j^* Q_j \leq \Lambda_{\mathbf{c}}.$$

If measurable functions  $f_i : E_i \rightarrow \mathbb{R}^+, 1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+, 1 \leq j \leq m$  satisfy

$$\prod_{i=1}^k f_i^{c_i}(\pi_{E_i}(x)) \leq \prod_{j=1}^m g_j^{d_j}(Q_j x) \quad \forall x \in E_0, \quad (18)$$

then

$$\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i} \leq \prod_{j=1}^m \left( \int_{E^j} g_j \right)^{d_j}. \quad (19)$$

*Proof.* By defining the maps  $B_j$  (and therefore  $\mathbf{B}$ ) via  $B_j \Lambda_{\mathbf{c}} = Q_j$ , the hypothesis (18) coincides with (2). For the corresponding datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , we see that  $V_i = \text{id}_{E_i}, 1 \leq i \leq k$ , and  $\Pi = \Sigma$  satisfy (17) by using the assumptions on the  $(Q_j)_{1 \leq j \leq m}$ . Now, it is a matter of plugging in the asserted extremizers in Theorem 1.14 into the definition of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  to see that  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = 0$ , and therefore  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) = 0$  by Theorem 1.3. This gives (19) by definition.  $\square$

**Remark 1.18.** In general,  $\Sigma$  in Corollary 1.17 does not need to be of full rank. An illustrative example is Barthe's inequality, in which  $m = d_1 = 1, E^1 = \mathbb{R}^n$  and  $E_i = \mathbb{R}$  for each  $1 \leq i \leq k$ , with  $k \geq n$ . In this setting  $Q_1$ , viewed as a matrix, has columns  $c_i q_i \in \mathbb{R}^n$ , where  $|q_i| = 1, 1 \leq i \leq k$ .

The reader can check that for the choice  $\Sigma = \Lambda_{\mathbf{c}}^{-1} Q_1^* Q_1 \Lambda_{\mathbf{c}}^{-1} \in S^+(\mathbb{R}^k)$ , the assumptions of the corollary are equivalent to Barthe's frame condition

$$|q_i| = 1, \quad 1 \leq i \leq k; \quad \text{and} \quad \sum_{i=1}^k c_i q_i \otimes q_i = \mathbf{I}_{n \times n}.$$

Note that this  $\Sigma$  has rank at most  $n \leq k$ , so can be rank-deficient.

Although it is a special case, a more symmetric formulation of the geometric forward-reverse Brascamp-Lieb inequality may be stated as follows, and explains the role of  $\Sigma$  in the previous as a transformation of coordinates when it is assumed to be of full rank. By specializing to  $k = c_1 = 1$ , the geometric case of the Brascamp-Lieb inequality is easily recognized. It will also be useful for the applications in Section 4.3.

**Corollary 1.19** (Geometric forward-reverse Brascamp-Lieb inequality (II)). *Let linear maps  $U_i : E_0 \rightarrow E_i$  and  $V_j : E_0 \rightarrow E^j$  satisfy  $U_i U_i^* = \text{id}_{E_i}$  and  $V_j V_j^* = \text{id}_{E^j}$ , for all  $1 \leq i \leq k$  and  $1 \leq j \leq m$ . Assume further that the map*

$$x \in E_0 \mapsto \sum_{i=1}^k U_i x \in E_0$$

is a bijection, and that the following frame condition holds

$$\sum_{i=1}^k c_i U_i^* U_i = \sum_{j=1}^m d_j V_j^* V_j.$$

If measurable functions  $f_i : E_i \rightarrow \mathbb{R}^+$ ,  $1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+$ ,  $1 \leq j \leq m$  satisfy

$$\prod_{i=1}^k f_i^{c_i}(U_i x) \leq \prod_{j=1}^m g_j^{d_j}(V_j x) \quad \forall x \in E_0, \tag{20}$$

then

$$\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i} \leq \prod_{j=1}^m \left( \int_{E^j} g_j \right)^{d_j}. \tag{21}$$

**Remark 1.20.** The map  $\sum_{i=1}^k U_i : E_0 \rightarrow E_0$  being a bijection is equivalent to  $\sum_{i=1}^k c_i U_i^* U_i > 0$ .

*Proof.* By taking traces, the frame condition ensures that (1) holds. Next, view  $U_i$  as a linear map from  $E_0$  into itself, so that  $\ker(U_i^*) = E_i^\perp$  for each  $1 \leq i \leq k$ . Since  $\sum_{i=1}^k U_i$  is a bijection, it is invertible, and therefore we are justified in defining  $Q_j := V_j (\sum_{i=1}^k U_i)^{-1}$  and  $\Sigma := \left( \sum_{i=1}^k U_i \right) \left( \sum_{i=1}^k U_i \right)^*$ . Evidently,  $\Sigma$  is positive-definite, and for  $x_i \in E_i$ , we have  $\langle \Sigma x_i, x_i \rangle = \langle x_i, x_i \rangle$  using the assumption  $U_i U_i^* = \text{id}_{E_i}$  and identification of  $\ker(U_i^*) = E_i^\perp$ . Therefore,  $\Sigma \in \Pi(\text{id}_{E_1}, \dots, \text{id}_{E_k})$ . Furthermore, it follows from definitions that  $Q_j \Sigma Q_j^* = \text{id}_{E^j}$  for each  $1 \leq j \leq m$ . Using again the fact that  $\sum_{i=1}^k U_i$  is a bijection, we find that (20) and (18) are equivalent by a change of variables. Thus, the hypotheses of Corollary 1.17 are fulfilled, and the conclusion follows.  $\square$

The first formulation of the geometric forward-reverse Brascamp-Lieb inequality motivates the following definitions:

**Definition 1.21.** A datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is said to be *geometric* if (1) holds, and the maps  $(Q_j)_{1 \leq j \leq m}$  defined via  $Q_j := B_j \Lambda_{\mathbf{c}}$  satisfy  $\sum_{j=1}^m d_j Q_j^* Q_j \leq \Lambda_{\mathbf{c}}$  and  $Q_j \Sigma Q_j^* = \text{id}_{E^j}$ ,  $1 \leq j \leq m$ , for some  $\Sigma \in \Pi(\text{id}_{E_1}, \dots, \text{id}_{E_k})$ .

**Definition 1.22.** Data  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  and  $(\mathbf{c}', \mathbf{d}', \mathbf{B}')$  are said to be equivalent if  $\mathbf{c} = \mathbf{c}'$ ,  $\mathbf{d} = \mathbf{d}'$ , and there exist invertible linear transformations  $C : E_0 \rightarrow E_0$  and  $C_j : E^j \rightarrow E^j$  such that  $B'_j = C_j^{-1} B_j C^{-1}$  for each  $1 \leq j \leq k$ .

The following characterization of gaussian extremizability extends [8, Proposition 3.6] to the forward-reverse setting.

**Theorem 1.23.** A datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable if and only if it is equivalent to a geometric datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B}')$ .

*Proof.* Suppose  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable. This is equivalent to the existence of  $V_i \in S^+(E_i)$ ,  $1 \leq i \leq k$  and  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  satisfying (17). Define  $V := \Lambda_{\mathbf{c}}^{-1} V_{\mathbf{c}} = \bigoplus_{i=1}^k V_i$ , and note that  $\Sigma := V^{1/2} \Pi V^{1/2} \in \Pi(\text{id}_{E_1}, \dots, \text{id}_{E_k})$ , and moreover, that  $V^{-1/2}$  commutes with  $\Lambda_{\mathbf{c}}$ . Define  $B'_j := C_j^{-1} B_j C^{-1}$  via the transformations  $C_j := (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{1/2}$  and  $C := V^{1/2}$ . Then, (17) is expressed in terms of  $\mathbf{B}' := (B'_j)_{1 \leq j \leq m}$  and  $\Sigma$  as

$$\sum_{j=1}^m d_j \Lambda_{\mathbf{c}} B_j'^* (B'_j \Lambda_{\mathbf{c}} \Sigma \Lambda_{\mathbf{c}} B_j'^*)^{-1} B'_j \Lambda_{\mathbf{c}} \leq \Lambda_{\mathbf{c}}.$$

In particular, this easily implies  $(\mathbf{c}, \mathbf{d}, \mathbf{B}')$  is geometric since  $B'_j \Lambda_{\mathbf{c}} \Sigma \Lambda_{\mathbf{c}} B_j'^* = \text{id}_{E_j}$  by construction. Moreover,  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  and  $(\mathbf{c}, \mathbf{d}, \mathbf{B}')$  are equivalent by definition. If  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is equivalent to a geometric datum, then the argument can be reversed to conclude gaussian-extremizability via Theorem 1.14.  $\square$

To state our main result on conditions for finiteness of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  and gaussian-extremizability, we define product-form subspaces, followed by two definitions analogous to those in [8].

**Definition 1.24.** A subspace  $T$  is said to be of *product-form* if

$$T = \bigoplus_{i=1}^k \pi_{E_i}(T)$$

**Definition 1.25.** A *critical subspace* for  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is a non-zero proper subspace  $T \subset E_0$  of product-form satisfying

$$\sum_{i=1}^k c_i \dim(\pi_{E_i} T) = \sum_{j=1}^m d_j \dim(B_j T).$$

**Definition 1.26.** The datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is *simple* if it has no critical subspaces.

We remark that the definition of criticality in [8] does not include the restriction to product-form subspaces. However, the two definitions are still consistent, because in their setting  $E_0 = E_1$ , so that any subspace is trivially of product-form.

Our final main result generalizes [8, Theorem 1.13] to the setting of the forward-reverse Brascamp-Lieb inequality:

**Theorem 1.27** (Conditions for finiteness and gaussian-extremizability). *For a datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , the quantity  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is finite if and only if we have the scaling condition*

$$\sum_{i=1}^k c_i \dim(E_i) = \sum_{j=1}^m d_j \dim(E^j) \tag{22}$$

and the dimension condition

$$\sum_{i=1}^k c_i \dim(\pi_{E_i} T) \leq \sum_{j=1}^m d_j \dim(B_j T) \text{ for all product-form subspaces } T \subseteq E_0. \tag{23}$$

In particular, these conditions imply that each  $(B_j)_{1 \leq j \leq m}$  must be surjective. Moreover, if  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is simple, then it is gaussian-extremizable.

**Remark 1.28.** The special case where  $k = c_1 = 1$  reduces to [8, Theorem 1.13].

**Remark 1.29.** By Theorem 1.3, finiteness of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is equivalent to finiteness of  $D_g(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$ . As expected, it can also be verified directly that the conditions of Theorem 1.27 are invariant under considering the dual datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B}) \rightarrow (\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$ . Indeed, let  $E^0 := \bigoplus_{j=1}^m E^j$ , and define  $B^i : E^0 \rightarrow E_i$  via the map

$$B^i y := \sum_{j=1}^m B_{ij}^* \pi_{E^j}(y), \quad y \in E^0.$$

For any  $W_j \subseteq E^j$ ,  $1 \leq j \leq m$ , define the product-form subspace  $W = \bigoplus_{j=1}^m W_j \subseteq E^0$ , and consider  $T = \bigoplus_{i=1}^k T_i$ , with  $T_i := E_i / B^i W$ . By the inclusion  $\bigoplus_{i=1}^k B^i W \supseteq B^* W \supseteq B_j^* W_j$ , we have

$$B_j T \subseteq B_j(E_0 / (B^* W)) \subseteq B_j(E_0 / (B_j^* W_j)) \subseteq E^j / W_j,$$

where the final equality follows again by set inclusion and the observation  $B_j^* W_j \subseteq \ker(B_j)^\perp$ . Hence, (23) implies

$$\sum_{i=1}^k c_i \dim(E_i) - \sum_{i=1}^k c_i \dim(B^i W) \leq \sum_{j=1}^m d_j \dim(E^j) - \sum_{j=1}^m d_j \dim(W_j).$$

Cancelling terms using (22) (which is trivially invariant to considering dual datums), we are left with the desired dual counterpart to (23).

**1.3. Outline of the paper.** Section 2 of this paper proves Theorem 1.3 under the assumption that the datum is gaussian-extremizable. This relies on establishing the structure of gaussian extremizers given in Theorem 1.14, and then adapting Lehec's stochastic proof of the Brascamp-Lieb and Barthe inequalities to exploit this information.

It turns out that the analysis of the gaussian-extremizable case more or less suffices to prove Theorem 1.3 in its full generality. To do this, we develop a machinery for iteratively decomposing a datum that is not gaussian-extremizable. This is the general focus of Section 3, which parallels the development of Bennett, Carbery, Christ and Tao for the special case of the direct Brascamp-Lieb inequality [8]. The conditions for finiteness and gaussian-extremizability articulated in Theorem 1.27 are a product of these arguments.

Connections between the forward-reverse Brascamp-Lieb inequality and other results in the literature are detailed in Section 4.

**Acknowledgment.** The first author is grateful for discussions with Franck Barthe, Michael Christ, Varun Jog, Joseph Lehec and Paweł Wolff. In particular, he thanks Franck Barthe for drawing our attention to references [5, 6], and Paweł Wolff for explaining the connections to the forward-reverse Brascamp-Lieb inequality.

## 2. THE GAUSSIAN-EXTREMIZABLE CASE

The goal of this section is to establish Theorem 1.3 under the assumption of gaussian-extremizability. Specifically, we aim to prove the following two results:

**Theorem 2.1.** *If a datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable, then  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ .*

**Theorem 2.2.** *If a datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable, then so is  $(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$ . Moreover,*

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

*regardless of whether the data are gaussian-extremizable.*

**Remark 2.3.** If  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = +\infty$ , then the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  does not satisfy the definition of gaussian-extremizability. However, we will clearly have  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  in this case also due to the general relation  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) \geq D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ .

**Remark 2.4.** The combination of the above implies the assertion of Theorem 1.3 if (i)  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable; or (ii)  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = +\infty$ .

**2.1. Proof of Theorems 1.14 and 2.1.** Let us **admit** the following preliminary version of Theorem 1.14, the proof of which is deferred to Section 2.3.

**Proposition 2.5.** *If a datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable, then (1) holds and there are  $V_i \in S^+(E_i)$ ,  $1 \leq i \leq k$  and  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  such that*

$$\sum_{j=1}^m d_j \Lambda_{\mathbf{c}} B_j^* (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1} B_j \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}}. \quad (24)$$

We now describe the variational representation for gaussian integrals due to Boué and Dupuis [10] (see also Borell [9], and historical remarks by Lehec [21]). For a given time horizon  $T$  and a Euclidean space  $E$ , a Brownian motion  $(W_t)_{0 \leq t \leq T}$  (starting from 0) taking values in  $E$  is said to have covariance  $K \in S^+(E)$  if  $\text{Cov}(W_1) = K$ . Let  $\mathbb{H}(E, K)$  denote the Hilbert space of all absolutely continuous paths  $u : [0, T] \rightarrow E$  starting from 0, equipped with norm

$$\|u\|_{\mathbb{H}(E, K)}^2 := \int_0^T \langle K^{-1} \dot{u}_s, \dot{u}_s \rangle ds.$$

A *drift*  $U$  is any process adapted to the Brownian filtration which has sample paths belonging to  $\mathbb{H}(E, K)$  almost surely.

**Proposition 2.6.** *Let  $g : E \rightarrow \mathbb{R}$  be measurable and bounded from below, and let  $(W_t)_{0 \leq t \leq T}$  be a Brownian motion with covariance  $K \in S^+(E)$ . It holds that*

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

where the supremum is taken over all drifts  $U$ .

We now prove Theorems 1.14 and 2.1, on the basis of Proposition 2.5. The argument is an adaptation of Lehec's **stochastic** proof of the Brascamp-Lieb and Barthe inequalities [22], with the main difference being the incorporation of the optimality conditions of Proposition 2.5. **A stochastic approach is also employed by Cordero-Erasquin and Maurey in [16]; however, their argument does not immediately yield sharp constants, nor does it identify the structure of extremizers.**

*Proof of Theorems 1.14 and 2.1.* Assume  $(f_i)_{1 \leq i \leq k}$  and  $(g_j)_{1 \leq j \leq m}$  satisfy (2). For purposes of proving the theorem, we may assume that each  $f_i$  is supported on some compact  $K_i \subset E_i$ , and is bounded from above, say by  $M$ . We may also assume that each  $g_j$  is bounded from below, say by  $M^{-1}$  on the compact set  $\sum_{i=1}^k c_k B_{ij} K_i \subset E^j$ . The general result will follow by dominated convergence. As a result, it is easy to see that we may now assume each  $g_j$  is bounded from above by some  $M' = M'(M, \mathbf{c}, \mathbf{d})$ , still preserving (2). Indeed, this modification can only reduce the product in the RHS of (3), making our task more difficult.

With the above assumptions, fix any  $\delta > 0$  and introduce the auxiliary functions

$$u_i = \log(f_i + \delta)$$

which are of course bounded from below. Using the assumption that the  $f_i$ 's and  $g_j$ 's are uniformly bounded and the hypothesis (2), there are constants  $C, c$  (depending on  $M, M', \mathbf{c}, \mathbf{d}$ , but not on  $\delta$ ) such that taking

$$v_j = \log(g_j + C\delta^c)$$

we will have

$$\sum_{i=1}^k c_i u_i(\pi_{E_i}(x)) \leq \sum_{j=1}^m d_j v_j(B_j \Lambda_{\mathbf{c}} x) \quad \forall x \in E_0.$$

Moreover, the  $v_j$ 's are also bounded from below.

Using the assumption of gaussian-extremizability, we invoke Proposition 2.5 to select  $V_i \in S^+(E_i)$ ,  $i = 1, \dots, k$  and  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  satisfying (24). In particular, this implies

$$\sum_{j=1}^m d_j \|B_j \Lambda_{\mathbf{c}} u\|_{\mathbb{H}(E^j, B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)}^2 \leq \sum_{i=1}^k c_i \|\pi_{E_i}(u)\|_{\mathbb{H}(E_i, V_i^{-1})}^2 \quad (25)$$

for any absolutely continuous path  $u : [0, T] \rightarrow E_0$ .

Now, let  $(W_t)_{0 \leq t \leq T}$  be a Brownian motion taking values in  $E_0$  with covariance  $\Pi$ . For each  $i = 1, \dots, k$ , define  $W_t^i = \pi_{E_i}(W_t)$ , which is a Brownian motion on  $E_i$  with covariance  $V_i^{-1}$ .

Fix  $\epsilon > 0$ , and for each  $i = 1, \dots, k$ , let  $U^i$  be an  $E_i$ -valued drift belonging to  $\mathbb{H}(E_i, V_i^{-1})$  such that

$$\log \mathbb{E} e^{u_i(W_T^i)} - c_i^{-1} \epsilon \leq \mathbb{E} \left[ u_i(W_T^i + U_T^i) - \frac{1}{2} \|U^i\|_{\mathbb{H}(E_i, V_i^{-1})}^2 \right],$$

the existence of which follows from Proposition 2.6. Define the  $E_0$ -valued process  $U = \sum_{i=1}^k U^i$ , and note that  $B_j \Lambda_{\mathbf{c}} U$  is a drift belonging to  $\mathbb{H}(E^j, B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)$  by (25).

Hence, the above estimates and another application of Proposition 2.6 give

$$\begin{aligned} & \sum_{i=1}^k c_i \log \mathbb{E} e^{u_i(W_T^i)} - \epsilon \\ & \leq \mathbb{E} \left[ \sum_{i=1}^k c_i u_i(W_T^i + U_T^i) - \frac{1}{2} \sum_{i=1}^k c_i \|U^i\|_{\mathbb{H}(E_i, V_i^{-1})}^2 \right] \\ & \leq \mathbb{E} \left[ \sum_{j=1}^m d_j v_j(B_j \Lambda_{\mathbf{c}} W_T + B_j \Lambda_{\mathbf{c}} U_T) - \frac{1}{2} \sum_{j=1}^m d_j \|B_j \Lambda_{\mathbf{c}} U\|_{\mathbb{H}(E^j, B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)}^2 \right] \\ & \leq \sum_{j=1}^m d_j \log \mathbb{E} e^{v_j(B_j \Lambda_{\mathbf{c}} W_T)}. \end{aligned}$$

Since  $f_i \leq e^{u_i}$  and  $e^{v_j} = g_j + C\delta^c$ , we conclude by arbitrary choice of  $\epsilon, \delta$  that

$$\prod_{i=1}^k (\mathbb{E}[f_i(W_T^i)])^{c_i} \leq \prod_{j=1}^m (\mathbb{E}[g_j(B_j \Lambda_{\mathbf{c}} W_T)])^{d_j}.$$

In particular, writing out the expectations as integrals and canceling common factors using (1), the above may be rewritten as

$$\begin{aligned} & \prod_{i=1}^k \left( \int_{E_i} f_i(x) e^{-\langle V_i x, x \rangle / 2T} dx \right)^{c_i} \\ & \leq \frac{\prod_{i=1}^k \det(V_i^{-1})^{1/2}}{\prod_{j=1}^m \det(B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{1/2}} \prod_{j=1}^m \left( \int_{E^j} g_j(x) e^{-\langle (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1} x, x \rangle / 2T} dx \right)^{d_j}. \end{aligned}$$

Letting  $T \rightarrow +\infty$ , monotone convergence yields

$$\prod_{i=1}^k \left( \int_{E_i} f_i(x) dx \right)^{c_i} \leq \left( \frac{\prod_{j=1}^m \det((B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1})^{d_j}}{\prod_{i=1}^k \det(V_i)^{c_i}} \right)^{1/2} \prod_{j=1}^m \left( \int_{E^j} g_j(x) dx \right)^{d_j}.$$

The consequences of this are two-fold: defining  $U_j = (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1}$ ,  $1 \leq j \leq m$ , we see that (24) implies  $(V_i)_{1 \leq i \leq k}, (U_j)_{1 \leq j \leq m}$  satisfy (14). Therefore, the ratio of determinants is at most  $\exp(D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}))$ , proving Theorem 2.1. In fact, the ratio of determinants must be precisely equal to  $\exp(D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}))$ , since we have freedom in choosing the functions  $(f_i)_{1 \leq i \leq k}, (g_j)_{1 \leq j \leq m}$  subject to (2). Thus, we have also shown that if (1) holds and there exist  $V_i \in S^+(E_i)$ ,  $i = 1, \dots, k$  and  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  satisfying (24), then the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable with  $(V_i)_{1 \leq i \leq k}$  and  $U_j = (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1}$ ,  $1 \leq j \leq m$  being gaussian extremizers. This proves the converse of Proposition 2.5 (and therefore Theorem 1.14 on the basis of Proposition 2.5).  $\square$

**Remark 2.7.** Using the structure of gaussian extremizers in the gaussian-extremizable case, other proof techniques such as optimal transport or heat flow should also work to establish the above.

**2.2. Proof of Theorem 2.2.** For  $V_i \in S^+(E_i)$ ,  $1 \leq i \leq k$  and  $U_j \in S^+(E^j)$ ,  $1 \leq j \leq m$ , we first note that the Schur complement condition for positive-semidefiniteness implies

$$\Lambda_{\mathbf{c}}^{1/2} B^* \Lambda_{\mathbf{d}}^{1/2} \left( \bigoplus_{j=1}^m U_j \right) \Lambda_{\mathbf{d}}^{1/2} B \Lambda_{\mathbf{c}}^{1/2} \leq \left( \bigoplus_{i=1}^k V_i \right)$$

if and only if

$$\Lambda_{\mathbf{d}}^{1/2} B \Lambda_{\mathbf{c}}^{1/2} \left( \bigoplus_{i=1}^k V_i^{-1} \right) \Lambda_{\mathbf{c}}^{1/2} B^* \Lambda_{\mathbf{d}}^{1/2} \leq \left( \bigoplus_{j=1}^m U_j^{-1} \right).$$

Noting that the first inequality is precisely (16), it follows immediately that  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D_g(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$  (regardless of gaussian-extremizability). Now, suppose the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable. If  $V_i \in S^+(E_i)$ ,  $1 \leq i \leq k$  and  $U_j \in S^+(E^j)$ ,  $1 \leq j \leq m$  are gaussian extremizers for  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , then it is immediate from the above observation that  $U_j^{-1} \in S^+(E^j)$ ,  $1 \leq j \leq m$  and  $V_i^{-1} \in S^+(E_i)$ ,  $1 \leq i \leq k$  and are gaussian extremizers for the datum  $(\mathbf{d}, \mathbf{c}, \mathbf{B}^*)$ .

**2.3. Proof of Proposition 2.5.** The goal of this section is to prove the optimality conditions asserted in Proposition 2.5, which was the core assumption needed to prove Theorems 1.14 and 2.1. The basic argument boils down to a strong min-max theorem. This is given in the first subsection. The second subsection leverages this min-max identity to complete the proof of Proposition 2.5. The arguments of this section follow those appearing in our previous work [25], however it suffices to restrict attention to a particular duality enjoyed by positive-definite operators.

**2.3.1. A strong min-max theorem.**

**Theorem 2.8.** For any  $K_i \in S^+(E_i)$ ,  $i = 1, \dots, k$ , it holds that

$$\begin{aligned} & \max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det(B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) + \sum_{j=1}^m d_j \dim(E^j) \\ &= \inf_{(V_i, U_j)_{1 \leq i \leq k, 1 \leq j \leq m}} \left( \sum_{i=1}^k c_i \langle V_i, K_i \rangle_{\text{HS}} - \sum_{j=1}^m d_j \log \det U_j \right), \end{aligned} \quad (26)$$

where the infimum is over  $V_i \in S^+(E_i)$ ,  $1 \leq i \leq k$  and  $U_j \in S^+(E^j)$ ,  $1 \leq j \leq m$  satisfying  $\Lambda_{\mathbf{c}} B^* U_{\mathbf{d}} B \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}}$ .

The critical ingredient in the proof is the Fenchel-Rockafellar duality theorem [27], stated here as it appears in [29, Theorem 1.9]:

**Theorem 2.9.** *Let  $X$  be a normed vector space,  $X^*$  its topological dual space, and  $\Theta, \Xi$  two proper convex functions. Let  $\Theta^*, \Xi^*$  be the Legendre-Fenchel transforms of  $\Theta, \Xi$  respectively. Assume that there is some  $x_0 \in X$  such that*

$$\Theta(x_0) < +\infty, \quad \Xi(x_0) < +\infty,$$

and  $\Theta$  is continuous at  $x_0$ . Then,

$$\inf_{x \in X} [\Theta(x) + \Xi(x)] = \max_{x^* \in X^*} [-\Theta^*(-x^*) - \Xi^*(x^*)].$$

**Remark 2.10.** It is a part of both theorems above that the stated maximum is attained.

*Proof of Theorem 2.8.* In our application of the Fenchel-Rockafellar theorem, we will take  $X = S(E_0)$ , regarded as a Hilbert space with respect to the usual Hilbert-Schmidt inner product. In this case, we also identify  $X^* = S(E_0)$ . So, for  $K_i \in S^+(E_i)$ ,  $i = 1, \dots, k$ , given and  $F \in S(E_0)$ , define the functionals

$$\Xi(F) := \inf_{\substack{V_1 \in S^+(E_1), \dots, V_k \in S^+(E_k): \\ \langle V_i x, x \rangle \geq \langle Fx, x \rangle, \forall x \in E_0}} \sum_{i=1}^k c_i \langle V_i, K_i \rangle_{\text{HS}}$$

and

$$\Theta(F) := \inf_{\substack{U_1 \in S^+(E^1), \dots, U_m \in S^+(E^j): \\ \langle U_j B \Lambda_{\mathbf{c}} x, B \Lambda_{\mathbf{c}} x \rangle \leq \langle Fx, x \rangle, \forall x \in E_0}} \sum_{j=1}^m d_j \log \det U_j^{-1},$$

with the convention that  $\Theta(F) = +\infty$  if  $F \notin S^+(E_0)$  (since the infimum will be over an empty set).

It is easy to see that both  $\Theta : S(E_0) \rightarrow \mathbb{R}$  and  $\Xi : S(E_0) \rightarrow \mathbb{R}$  are proper convex functions, with  $\Theta(\text{id}_{E_0}) < +\infty$  and  $\Xi(\text{id}_{E_0}) < +\infty$ . It is straightforward to check the continuity of  $\Theta$  at  $\text{id}_{E_0}$ , so the hypotheses of Theorem 2.9 are fulfilled.

Let  $M_0$  denote the infimum in the RHS of (26), and observe that definitions easily imply

$$\inf_{F \in S(E_0)} [\Theta(F) + \Xi(F)] = M_0.$$

So, we turn our attention toward computing

$$\max_{H \in S(E_0)} [-\Theta^*(-H) - \Xi^*(H)].$$

To this end, we claim that

$$\Xi^*(H) \geq \begin{cases} 0 & \text{if } H \in \Pi(K_1, \dots, K_k) \\ +\infty & \text{otherwise.} \end{cases} \quad (27)$$

Indeed,

$$\begin{aligned} \Xi^*(H) &= \sup_{F \in S(E_0)} (\langle H, F \rangle - \Xi(F)) \geq \sup_{F \in S^+(E_0)} (\langle H, -F \rangle - \Xi(-F)) \\ &= \sup_{F \in S^+(E_0)} -\langle H, F \rangle = \begin{cases} 0 & \text{if } H \in S^+(E_0) \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Next, define  $H_i := \pi_{E_i} H \pi_{E_i}^*$  and note that

$$\begin{aligned} \Xi^*(H) &= \sup_{F \in S(E_0)} (\langle H, F \rangle - \Xi(F)) \geq \sup_{F_i \in S(E_i)} (\langle H, \pi_{E_i}^* F_i \pi_{E_i} \rangle - \Xi(\pi_{E_i}^* F_i \pi_{E_i})) \\ &= \sup_{F_i \in S(E_i)} \langle H_i - K_i, F_i \rangle = \begin{cases} 0 & \text{if } H_i = K_i \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$



This proves (27). Hence, we may conclude

$$\begin{aligned}
 & \max_{H \in S(E_0)} \left[ -\Theta^*(-H) - \Xi^*(H) \right] \\
 & \leq \sup_{K \in \Pi(K_1, \dots, K_k)} \left[ -\Theta^*(-K) \right] \\
 & = \sup_{K \in \Pi(K_1, \dots, K_k)} \inf_{F \in S^+(E_0)} [\langle K, F \rangle + \Theta(F)] \\
 & = \sup_{K \in \Pi(K_1, \dots, K_k)} \inf_{(U_j \in S^+(E^j))_{1 \leq j \leq m}} \left( \sum_{j=1}^m d_j \langle K, \Lambda_{\mathbf{c}} B_j^* U_j B_j \Lambda_{\mathbf{c}} \rangle - \sum_{j=1}^m d_j \log \det U_j \right) \\
 & = \sup_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \left( \inf_{(U_j \in S^+(E^j))_{1 \leq j \leq m}} \left( \langle B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*, U_j \rangle - \sum_{j=1}^m d_j \log \det U_j \right) \right) \\
 & = \sup_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) + \sum_{j=1}^m d_j \dim(E^j).
 \end{aligned}$$

It is straightforward to argue that the supremum is attained using the facts that  $\Pi(K_1, \dots, K_k)$  is compact and that  $B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*$  is uniformly bounded over all  $K \in \Pi(K_1, \dots, K_k)$ .

Therefore, invoking Theorem 2.9, we have shown

$$M_0 \leq \max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) + \sum_{j=1}^m d_j \dim(E^j).$$

The reverse direction is considerably simpler; consider any  $K \in \Pi(K_1, \dots, K_k)$ . Then,

$$\begin{aligned}
 & \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) + \sum_{j=1}^m d_j \dim(E^j) \\
 & \leq \inf_{(U_j \in S^+(E^j))_{1 \leq j \leq m}} \sum_{j=1}^m d_j \langle B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*, U_j \rangle - \sum_{j=1}^m d_j \log \det U_j.
 \end{aligned} \tag{28}$$

Note that, for any  $U_j \in S^+(E^j)$ ,  $1 \leq j \leq m$ , if  $V_i \in S^+(E_i)$ ,  $1 \leq i \leq k$  satisfy

$$\langle U_{\mathbf{d}} B \Lambda_{\mathbf{c}} x, B \Lambda_{\mathbf{c}} x \rangle \leq \langle V_{\mathbf{c}} x, x \rangle \quad \forall x \in E_0, \tag{29}$$

then it will hold that

$$\sum_{j=1}^m d_j \langle B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*, U_j \rangle_{\text{HS}} \leq \sum_{i=1}^k c_i \langle V_i, K_i \rangle_{\text{HS}}. \tag{30}$$

To see that this is indeed the case, let  $x$  be a centered gaussian random vector in  $E_0$  with covariance  $K$ , and take expectations of both sides in (29). So, combining estimates yields

$$\sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) + \sum_{j=1}^m d_j \dim(E^j) \leq M_0,$$

completing the proof of the theorem.  $\square$

**2.3.2. Completion of proof of Proposition 2.5.** The first step in completing the proof of Proposition 2.5 is to note an equivalent dual formulation of the optimization problem defining  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . This formulation relies on the strong min-max identity of the previous subsection. Through an analysis of the equality cases, we ultimately arrive at Proposition 2.5.

**Theorem 2.11.** Fix  $-\infty < C \leq +\infty$ . The following statements are equivalent:

1) For all  $V_i \in S^+(E_i), 1 \leq i \leq k$  and  $U_j \in S^+(E^j), 1 \leq j \leq m$  satisfying

$$\Lambda_{\mathbf{c}} B^* U_{\mathbf{d}} B \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}} \tag{31}$$

it holds that

$$\sum_{j=1}^m d_j \log \det U_j \leq C + \sum_{i=1}^k c_i \log \det V_i. \tag{32}$$

2) For all  $K_i \in S^+(E_i), 1 \leq i \leq k$ ,

$$\sum_{i=1}^k c_i \log \det K_i \leq C + \max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*). \tag{33}$$

**Remark 2.12.** By definition of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  and the asserted equivalence, the best constant  $C$  in each of the above inequalities is equal to  $2D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ .

**Remark 2.13.** To appreciate the difference between the forward-reverse Brascamp-Lieb inequality and the classical forward and reverse inequalities, we invite the reader to prove Theorem 2.11 in the special case of  $k = 1$  (which is symmetric with  $m = 1$ ). This is a simple exercise, requiring only a few lines of elementary linear algebra. Since the maximum in (33) becomes trivial, the major difficulty of the characterization (i.e., the strong min-max theorem of the previous section) is avoided.

*Proof.* For purposes of the proof, we assume each  $(B_j)_{1 \leq j \leq m}$  is surjective, ensuring invertibility of  $(B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*)$  for  $K \in S^+(E_0)$ . If any of the  $B_j$  fail to be surjective, it is easy to see that the best constant  $C$  in both 1) and 2) will be equal to  $+\infty$ , thereby handling this exceptional case. Moreover, we may assume that (1) holds. Again, if this is not the case, then by rescaling the various operators by a common factor, we see that the best constant  $C$  in both 1) and 2) will be equal to  $+\infty$ .

*Proof of 1)  $\Rightarrow$  2).* Fix any  $\epsilon > 0$ . By Theorem 2.8, there are  $V_i \in S^+(E_i), 1 \leq i \leq k$  and  $U_j \in S^+(E^j), 1 \leq j \leq m$  satisfying (31) such that

$$\begin{aligned} & \max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) + \sum_{j=1}^m d_j \dim(E^j) \\ & \geq \sum_{i=1}^k c_i \langle V_i, K_i \rangle_{\text{HS}} - \sum_{j=1}^m d_j \log \det U_j - \epsilon. \end{aligned}$$

Hence, we have

$$\begin{aligned} & \sum_{i=1}^k c_i \log \det K_i - \max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) \\ & \leq \sum_{i=1}^k c_i \log \det K_i + \sum_{j=1}^m d_j \dim(E^j) + \sum_{j=1}^m d_j \log \det U_j - \sum_{i=1}^k c_i \langle V_i, K_i \rangle_{\text{HS}} + \epsilon \\ & \leq \sum_{i=1}^k c_i \log \det K_i + \sum_{j=1}^m d_j \dim(E^j) + C + \sum_{i=1}^k c_i \log \det V_i - \sum_{i=1}^k c_i \langle V_i, K_i \rangle_{\text{HS}} + \epsilon \\ & \leq C + \epsilon, \end{aligned}$$

where the penultimate inequality is (32), and the final inequality is due to the elementary inequality  $\log \det M \leq \text{tr}(M) - \dim(E_i)$  for  $M \in S^+(E_i)$  and the scaling condition (1). Since  $\epsilon$  was arbitrary, it follows that 1)  $\Rightarrow$  2).

*Proof of 2)⇒1).* Fix any  $V_i \in S^+(E_i), i = 1, \dots, k$ . By the method of Lagrange multipliers and the weak max-min inequality, we have

$$\begin{aligned} & \sup_{\substack{(U_j \in S^+(E^j))_{1 \leq j \leq m}: \\ \Lambda_{\mathbf{c}} B^* U_{\mathbf{d}} B \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}}}} \sum_{j=1}^m d_j \log \det U_j \\ & \leq \inf_{M \in S^+(E_0)} \sup_{(U_j \in S^+(E^j))_{1 \leq j \leq m}} \left[ \sum_{j=1}^m d_j \log \det U_j + \langle M, V_{\mathbf{c}} \rangle_{\text{HS}} - \langle M, \Lambda_{\mathbf{c}} B^* U_{\mathbf{d}} B \Lambda_{\mathbf{c}} \rangle_{\text{HS}} \right] \\ & = \inf_{M \in S^+(E_0)} \sup_{(U_j \in S^+(E^j))_{1 \leq j \leq m}} \left[ \sum_{j=1}^m d_j \log \det U_j + \langle M, V_{\mathbf{c}} \rangle_{\text{HS}} - \sum_{j=1}^m d_j \langle B_j \Lambda_{\mathbf{c}} M \Lambda_{\mathbf{c}} B_j^*, U_j \rangle_{\text{HS}} \right]. \end{aligned}$$

Now, we note that the gradient of the objective above with respect to  $U_j$  is given by  $d_j U_j^{-1} - B_j \Lambda_{\mathbf{c}} M \Lambda_{\mathbf{c}} B_j^*$ . So, taking  $U_j = (B_j \Lambda_{\mathbf{c}} M \Lambda_{\mathbf{c}} B_j^*)^{-1}$  achieves the maximum in the inner optimization problem. Making this substitution, we may continue as

$$\begin{aligned} & \sup_{\substack{(U_j \in S^+(E^j))_{1 \leq j \leq m}: \\ \Lambda_{\mathbf{c}} B^* U_{\mathbf{d}} B \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}}}} \sum_{j=1}^m d_j \log \det U_j \\ & = \inf_{M \in S^+(E_0)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} M \Lambda_{\mathbf{c}} B_j^*)^{-1} - \sum_{j=1}^m d_j \dim(E^j) + \langle M, V_{\mathbf{c}} \rangle_{\text{HS}} \quad (34) \end{aligned}$$

$$\leq - \max_{M \in \Pi(V_1^{-1}, \dots, V_k^{-1})} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} M \Lambda_{\mathbf{c}} B_j^*) \quad (35)$$

$$\leq C - \sum_{i=1}^k c_i \log \det V_i^{-1} = C + \sum_{i=1}^k c_i \log \det V_i$$

where the first inequality follows since we are optimizing over a smaller set, and for  $M \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  it holds that

$$\langle M, V_{\mathbf{c}} \rangle_{\text{HS}} = \sum_{i=1}^k c_i \langle V_i^{-1}, V_i \rangle_{\text{HS}} = \sum_{i=1}^k c_i \dim(E_i) = \sum_{j=1}^m d_j \dim(E^j).$$

The second inequality is an application of (33).  $\square$

Finally, we are in a position to complete the proof of Proposition 2.5.

*Proof of Proposition 2.5.* If the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable, then (1) must hold, as remarked at the start of the proof of Theorem 2.11. Furthermore, for extremizers  $V_i \in S^+(E_i), 1 \leq i \leq k$ , we get equality throughout the second part of the proof of Theorem 2.11 for the optimal constant  $C = 2D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ ; in particular, equality will be attained in (35). Therefore, we remark that any

$$\Pi \in \arg \max_{M \in \Pi(V_1^{-1}, \dots, V_k^{-1})} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} M \Lambda_{\mathbf{c}} B_j^*)$$

is positive-semidefinite and achieves the minimum in (34), after replacing the infimum over positive-definite operators with a minimum over positive-semidefinite operators. Letting  $\Pi$  be as defined above, consider any positive-semidefinite  $A \in S(E_0)$ . For  $\varepsilon > 0$ , Taylor expansion and assumed

optimality of  $\Pi$  gives

$$\begin{aligned}
 & \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} (\Pi + \varepsilon A) \Lambda_{\mathbf{c}} B_j^*)^{-1} - \sum_{j=1}^m d_j \dim(E^j) + \langle (\Pi + \varepsilon A), V_{\mathbf{c}} \rangle_{\text{HS}} \\
 &= \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1} - \sum_{j=1}^m d_j \dim(E^j) + \langle \Pi, V_{\mathbf{c}} \rangle_{\text{HS}} \\
 &+ \varepsilon \left( - \sum_{j=1}^m d_j \left\langle A, \Lambda_{\mathbf{c}} B_j^* (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1} B_j \Lambda_{\mathbf{c}} \right\rangle_{\text{HS}} + \langle A, V_{\mathbf{c}} \rangle_{\text{HS}} \right) + o(\varepsilon) \\
 &\geq \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1} - \sum_{j=1}^m d_j \dim(E^j) + \langle \Pi, V_{\mathbf{c}} \rangle_{\text{HS}}.
 \end{aligned}$$

By sending  $\varepsilon \downarrow 0$  and letting  $A \geq 0$  vary, we find that it must hold that

$$\sum_{j=1}^m d_j \Lambda_{\mathbf{c}} B_j^* (B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*)^{-1} B_j \Lambda_{\mathbf{c}} \leq V_{\mathbf{c}} \tag{36}$$

as claimed.

**Remark 2.14.** If we consider  $A$  such that  $\ker(A) \subseteq \ker(\Pi)$ , then the same inequalities above hold for all  $\varepsilon \in \mathbb{R}$  sufficiently small, since this ensures  $\Pi + \varepsilon A$  will remain positive-semidefinite. As a result, we have equality in (36) on the restriction of both sides to  $\ker(\Pi)^\perp$ . This provides an alternate way to establish the conclusion of Remark 1.15.

□

Although it will not be needed for our purposes, there are several equivalent characterizations of gaussian extremizers which can be stated in analogy to [8, Proposition 3.6] for the direct Brascamp-Lieb inequality. The statement is provided below to give a comprehensive description of the structure of gaussian extremizers. It provides a convenient means of certifying the optimality of gaussian extremizers, which is required to compute  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ .

**Proposition 2.15.** Fix a datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , and  $(U_j \in S^+(E^j))_{1 \leq j \leq m}, (V_i \in S^+(E_i))_{1 \leq i \leq k}$ . The following statements are equivalent:

- (i)  $(U_j)_{1 \leq j \leq m}, (V_i)_{1 \leq i \leq k}$  is a global maximum in (15) subject to (14);
- (ii)  $(U_j)_{1 \leq j \leq m}, (V_i)_{1 \leq i \leq k}$  is a local maximum in (15) subject to (14);
- (iii) There exists  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  such that

$$U_j^{-1} = B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*, \quad 1 \leq j \leq m. \tag{37}$$

Moreover, for

$$\Theta := V_{\mathbf{c}} - \sum_{j=1}^m d_j \Lambda_{\mathbf{c}} B_j^* U_j B_j \Lambda_{\mathbf{c}}, \tag{38}$$

we have  $\Theta \geq 0$  and  $\Theta x = 0$  for any  $x \in \Pi E_0$ .

- (iv) The dimension condition (1) is satisfied, and there exists  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  such that (37) holds. Moreover,  $\Theta \geq 0$ , with  $\Theta$  defined as in (38).
- (v) The dimension condition (1) is satisfied, and there exists  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$  such that

$$U_j^{-1} \geq B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*, \quad 1 \leq j \leq m. \tag{39}$$

Moreover,  $\Theta \geq 0$  and  $\Pi^{1/2} \Theta \Pi^{1/2} = 0$ , with  $\Theta$  defined as in (38).

*Proof.* (i) $\implies$ (ii) is trivial.

(iii) $\implies$ (iv): By the assumption of (iii) and the fact that  $\Pi^{1/2}E_0 = \Pi E_0$ , we have  $\text{tr}(\Pi^{1/2}\Theta\Pi^{1/2}) = 0$ . Using (37) and (38) and the cyclic property of the trace, we find

$$\sum_{i=1}^k c_i \dim(E_i) = \sum_{j=1}^m d_j \dim(E^j). \quad (40)$$

(iv) $\implies$ (iii): It suffices to show that  $\Pi^{1/2}\Theta\Pi^{1/2} = 0$ . As argued above, the dimension condition is equivalent to  $\text{tr}(\Pi^{1/2}\Theta\Pi^{1/2}) = 0$ . Moreover, the assumption of (iv) requires that  $\Pi^{1/2}\Theta\Pi^{1/2} \geq 0$ . Since a nonnegative matrix has trace 0 only if it is the zero matrix, we have  $\Pi^{1/2}\Theta\Pi^{1/2} = 0$  as desired.

(iii) $\implies$ (i): This is the conclusion of Section 2.1.

(iii) $\implies$ (v): The proof of the dimension condition is the same as the (iii) $\implies$ (iv) part.

(v) $\implies$ (iii): By the assumption of (v) we have

$$\begin{aligned} 0 = \text{tr}(\Pi^{1/2}\Theta\Pi^{1/2}) &= \sum_{i=1}^k c_i \dim(E_i) - \sum_{j=1}^m d_j \text{tr}(U_j(B_j\Lambda_{\mathbf{c}}\Pi\Lambda_{\mathbf{c}}B_j^*)) \\ &\geq \sum_{i=1}^k c_i \dim(E_i) - \sum_{j=1}^m d_j \text{tr}(U_jU_j^{-1}) \\ &= 0 \end{aligned} \quad (41)$$

where (41) follows from (39). If (39) is not equality for some  $j$ , then (41) cannot achieve equality. Therefore (39) must achieve equality for all  $j$ .

(ii) $\implies$ (iii): By definition, we have

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) := \frac{1}{2} \sup \left( \sum_{j=1}^m d_j \log \det U_j - \sum_{i=1}^k c_i \log \det V_i \right), \quad (42)$$

where the supremum is over  $V_i \in S^+(E_i)$  and  $U_j \in S^+(E^j)$  satisfying the constraint

$$\Theta := V_{\mathbf{c}} - \Lambda_{\mathbf{c}}B^*U_{\mathbf{d}}B\Lambda_{\mathbf{c}} \geq 0. \quad (43)$$

We need to show that if given  $(V_i)_{1 \leq i \leq k}$  and  $(U_j)_{1 \leq j \leq m}$  is a local maximum of (42), then there exists  $\Pi$  with the claimed properties. To this end, let  $S_0^+(E_0)$  denote the set of positive semidefinite operators on  $E_0$ , and define the convex cone:

$$\mathcal{C} := \{(c_1\Pi_1, \dots, c_k\Pi_k, -d_1B_1\Lambda_{\mathbf{c}}\Pi\Lambda_{\mathbf{c}}B_1^*, \dots, -d_mB_m\Lambda_{\mathbf{c}}\Pi\Lambda_{\mathbf{c}}B_m^*) : \Pi \in S_0^+(E_0)\}, \quad (44)$$

where  $\Pi_i : E_i \rightarrow E_i$  is defined as  $\Pi_i = \pi_{E_i}\Pi\pi_{E_i}^*$ . Recall that the dual cone  $\mathcal{C}^*$  of  $\mathcal{C}$  is defined as the set of all vectors whose inner product with any element in  $\mathcal{C}$  is nonnegative. Here, we have

$$\mathcal{C}^* := \{(V_1, \dots, V_k, U_1, \dots, U_m) : V_i \in S_0^+(E_i), U_j \in S_0^+(E^j), V_{\mathbf{c}} - \Lambda_{\mathbf{c}}B^*U_{\mathbf{d}}B\Lambda_{\mathbf{c}} \geq 0\}. \quad (45)$$

Note that this is indeed the dual cone since the constraint  $V_{\mathbf{c}} - \Lambda_{\mathbf{c}}B^*U_{\mathbf{d}}B\Lambda_{\mathbf{c}} \geq 0$  can be rewritten as

$$\langle \Pi, V_{\mathbf{c}} - \Lambda_{\mathbf{c}}B^*U_{\mathbf{d}}B\Lambda_{\mathbf{c}} \rangle_{\text{HS}} \geq 0, \quad \text{for all } \Pi \in S_0^+(E_0), \quad (46)$$

or equivalently,  $\sum_{i=1}^k \langle c_i\Pi_i, V_i \rangle_{\text{HS}} + \sum_{j=1}^m \left\langle -d_j B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^*, U_j \right\rangle_{\text{HS}} \geq 0$ , for all  $\Pi \in S_0^+(E_0)$ . Note that the constraint defining  $\mathcal{C}^*$  is the same as the constraint (43) for the optimization (42). Now, from (42), we see that the gradient of the objective function with respect to  $(V_1, \dots, V_k, U_1, \dots, U_m)$  is equal to  $(-\frac{c_1}{2}V_1^{-1}, \dots, -\frac{c_k}{2}V_k^{-1}, \frac{d_1}{2}U_1^{-1}, \dots, \frac{d_m}{2}U_m^{-1})$ . The assumed local optimality implies that the inner product of the gradient at  $(V_i)_{1 \leq i \leq k}, (U_j)_{1 \leq j \leq m}$  with any element in  $\mathcal{C}^*$  is non-positive, hence the negative gradient belongs to the dual of  $\mathcal{C}^*$ . Recall that the double-dual of a closed

convex cone is itself, hence the negative gradient belongs to  $\mathcal{C}$ , and therefore (upon absorbing a factor 2 in  $\Pi$ ) we find  $\Pi \in S_0^+(E_0)$  such that

$$\Pi_i = V_i^{-1}, \quad i = 1, \dots, k, \tag{47}$$

$$B_j \Lambda_{\mathbf{c}} \Pi \Lambda_{\mathbf{c}} B_j^* = U_j^{-1}, \quad j = 1, \dots, m. \tag{48}$$

Note that (47) together with  $\Pi \in S_0^+(E_0)$  is equivalent to  $\Pi \in \Pi(V_1^{-1}, \dots, V_k^{-1})$ , so it only remains to show that  $\Theta x = 0$  for  $x \in \Pi E_0$ , or equivalently  $\Pi^{1/2} \Theta \Pi^{1/2} = 0$ . The assumed local optimality implies that the dimension condition (40) must hold, else scaling  $(V_i)_{1 \leq i \leq k}, (U_j)_{1 \leq j \leq m}$  by an appropriate common factor will increase the value of the functional being optimized. As noted in the proof of (iv)  $\implies$  (iii), the dimension condition together with (43) (i.e., nonnegativity of  $\Theta$ ) implies  $\Pi^{1/2} \Theta \Pi^{1/2} = 0$ , as desired.  $\square$

To conclude this section, we record the following observation which will be needed later.

**Lemma 2.16.** *Assume the  $(B_j)_{1 \leq j \leq m}$  are surjective. The map*

$$(K_1, \dots, K_k) \longmapsto \left( \sum_{i=1}^k c_i \log \det K_i - \max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) \right)$$

*is upper-semicontinuous on  $\prod_{i=1}^k S^+(E_i)$  with respect to the norm topology.*

*Proof.* The map  $(K_1, \dots, K_k) \longmapsto \sum_{i=1}^k c_i \log \det K_i$  is continuous on  $\prod_{i=1}^k S^+(E_i)$ , and we may write

$$\max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) = \sup_{K \in \Pi(K_1, \dots, K_k) \cap S^+(E_0)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*),$$

which is lower-semicontinuous in  $(K_1, \dots, K_k)$  on  $\prod_{i=1}^k S^+(E_i)$ , since it is the pointwise supremum of continuous functions.  $\square$

### 3. DECOMPOSABILITY AND CONDITIONS FOR FINITENESS

The aim of this section is to complete the proof of Theorem 1.3 by successively decomposing data which are not gaussian-extremizable into ones that are either gaussian-extremizable or degenerate. A consequence of the arguments will be the necessary and sufficient conditions for finiteness given by Theorem 1.27. The development closely parallels the treatment of the forward Brascamp-Lieb inequality in [8], however the modifications are significant enough that it warrants explicitly giving the details. Our starting point will be to state a characterization of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  in terms of Shannon entropies, denoted by  $h$ . This characterization will be exploited to facilitate the various computations later on. Basic properties of the Shannon entropy (subadditivity, scaling, etc.) will be taken for granted here; the unfamiliar reader can find the needed properties collected in Appendix A, or any standard text on information theory.

**3.1. Entropic characterization of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ .** We first introduce some notation that will be needed below and again in Section 4.4. For a collection of random vectors  $(X_i)_{1 \leq i \leq k}$  taking values in  $(E_i)_{1 \leq i \leq k}$ , respectively, we denote the marginal law of each  $X_i$  as  $P_{X_i}$ , and denote their joint law by  $P_{\mathbf{X}}$ . The set of couplings of  $(X_i)_{1 \leq i \leq k}$  (i.e., joint laws of  $(X_i)_{1 \leq i \leq k}$  with  $X_i$ -marginal equal to  $P_{X_i}$  for each  $1 \leq i \leq k$ ) is denoted by  $\Pi(P_{X_1}, \dots, P_{X_k})$ . Since elements of  $S^+(E_i)$  are in one-to-one correspondence with centered gaussian probability measures on  $E_i$  (see Remark 1.13), this notation is consistent with the earlier definition of  $\Pi(K_1, \dots, K_k)$  for  $K_i \in S^+(E_i)$ ,  $1 \leq i \leq k$ .

If  $Z$  is a gaussian random vector in  $E$  with covariance  $\Sigma \in S^+(E)$ , we have the identity

$$h(Z) = \frac{1}{2} \log \left( (2\pi e)^{\dim(E)} \det(\Sigma) \right). \tag{49}$$

So, a reinterpretation of Theorem 2.11 and Remark 2.12 is the following entropic characterization of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . A similar characterization also holds for  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ ; see Theorem 4.12 and accompanying remarks in Section 4.4.

**Proposition 3.1.** *If  $(Z_i)_{1 \leq i \leq k}$  are gaussian random vectors in  $(E_i)_{1 \leq i \leq k}$ , respectively, then*

$$\sum_{i=1}^k c_i h(c_i^{-1} Z_i) - \max_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} Z_i\right) \leq D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}), \quad (50)$$

where the maximum is over all couplings of the  $(Z_i)_{1 \leq i \leq k}$ . Moreover, the constant  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is best possible.

**Remark 3.2.** Although not explicitly stated, it suffices to consider jointly gaussian couplings in (50), giving equivalence to (33). This is a consequence of the fact that gaussians maximize entropy for a given covariance.

**Remark 3.3.** Since any choice of gaussian  $(Z_i)_{1 \leq i \leq k}$  in (50) have finite second moments by definition, Proposition A.2 in Appendix A ensures that the entropies  $h\left(\sum_{i=1}^k B_{ij} Z_i\right)$  exist and satisfy

$$h\left(\sum_{i=1}^k B_{ij} Z_i\right) < +\infty \text{ for each } 1 \leq j \leq m.$$

Moreover, if  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) < +\infty$ , then there must exist a coupling of the  $(Z_i)_{1 \leq i \leq k}$  (in particular, the one achieving the maximum in (50)) such that

$$h\left(\sum_{i=1}^k B_{ij} Z_i\right) > -\infty \text{ for each } 1 \leq j \leq m. \quad (51)$$

In other words, under this optimal coupling, the entropy  $h\left(\sum_{i=1}^k B_{ij} Z_i\right)$  exists and is finite for each  $1 \leq j \leq m$ .

**3.2. Decomposability of  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  and  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  for non-simple data.** Let  $T$  be a subspace of  $E_0$  having product form, and define  $B_{j,T}$  to be the restriction of  $B_j$  to  $T$ . Note that if  $T_i := \pi_{E_i} T$ , and  $B_{ij,T_i}$  is the restriction of  $B_{ij}$  to  $T_i$ , then the product-form assumption implies

$$B_{j,T} x = \sum_{i=1}^k B_{ij,T_i} \pi_{E_i}(x), \quad x \in T.$$

Now, let  $B_{ij,T_i^\perp}$  denote the restriction of  $(\pi_{(B_j T)^\perp} B_{ij})$  to  $T_i^\perp$ , the orthogonal complement of  $T_i$  in  $E_i$ , and define the collections of linear maps

$$\begin{aligned} \mathbf{B}_T &:= \{B_{ij,T_i} : T_i \longrightarrow (B_j T)\}_{1 \leq i \leq k, 1 \leq j \leq m}, \text{ and} \\ \mathbf{B}_{E_0/T} &:= \{B_{ij,T_i^\perp} : T_i^\perp \longrightarrow (B_j T)^\perp\}_{1 \leq i \leq k, 1 \leq j \leq m}, \end{aligned}$$

where  $(B_j T)^\perp$  denotes the orthogonal complement of  $B_j T$  in  $E^j$ ,  $1 \leq j \leq m$ .

**Lemma 3.4.** *For any product-form subspace  $T \subseteq E_0$ , it holds that*

$$D(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq D(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T})$$

and

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}).$$

**Remark 3.5.** It may happen in the decomposition of  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  into  $(\mathbf{c}, \mathbf{d}, \mathbf{B}_T)$  and  $(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T})$  that we encounter subspaces of dimension zero. Just as argued in [8], these subspaces can be safely disregarded in the following and subsequent computations. In particular, the entropy of a random variable on a subspace of dimension zero (a degenerate situation) is defined to be equal to zero in subsequent computations.

*Proof.* We assume both  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T)$  and  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T})$  are finite, else the claim is trivial.

Fix  $M > 0$ . Let  $f_i : E_i \rightarrow \mathbb{R}^+$ ,  $1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+$ ,  $1 \leq j \leq m$  be non-negative measurable functions, bounded from above by  $M$  and satisfying

$$\prod_{i=1}^k f_i^{c_i}(z_i) \leq \prod_{j=1}^m g_j^{d_j} \left( \sum_{i=1}^k c_i B_{ij} z_i \right) \quad \forall z_i \in E_i, 1 \leq i \leq k. \quad (52)$$

Define  $T_i := \pi_{E_i} T$ . For  $z_i \in E_i$ , define  $x_i := \pi_{T_i}(z_i)$  and  $y_i := \pi_{T_i^\perp}(z_i)$ . In this notation, the hypothesis (52) implies

$$\prod_{i=1}^k f_i^{c_i}(x_i + y_i) \leq \prod_{j=1}^m g_j^{d_j} \left( \sum_{i=1}^k c_i (B_{ij, T_i} x_i + \pi_{B_j T} B_{ij} y_i) + \sum_{i=1}^k c_i B_{ij, T_i^\perp} y_i \right). \quad (53)$$

By the Fubini-Tonelli theorem, the map  $x_i \in T_i \mapsto f_i(x_i + y_i)$  is measurable for almost every  $y_i \in T_i^\perp$ ; define  $N_i \subset T_i^\perp$  to be the (null) set of  $y_i \in T_i^\perp$  for which  $x_i \in T_i \mapsto f_i(x_i + y_i)$  is not measurable. By defining

$$\tilde{f}_i(x_i + y_i) = f(x_i + y_i) 1_{T_i^\perp \setminus N_i}(y_i), \quad x_i \in T_i, y_i \in T_i^\perp,$$

we have that  $x_i \in T_i \mapsto \tilde{f}_i(x_i + y_i)$  is measurable for all  $y_i \in T_i^\perp$ . Moreover, since  $f$  was only modified on a null set,  $\int_{E_i} \tilde{f}_i = \int_{E_i} f_i$ . Similarly, the map  $u_j \in B_j T \mapsto g_j(u_j + v_j)$  is measurable for almost every  $v_j \in (B_j T)^\perp$ ; define  $N^j \subset (B_j T)^\perp$  to be the (null) set of  $v_j \in (B_j T)^\perp$  for which  $u_j \in B_j T \mapsto g_j(u_j + v_j)$  is not measurable. Almost as before, we define

$$\tilde{g}_j(u_j + v_j) = g_j(u_j + v_j) 1_{(B_j T)^\perp \setminus N^j}(v_j) + M 1_{N^j}(v_j), \quad u_j \in B_j T, v_j \in (B_j T)^\perp,$$

and are guaranteed that  $u_j \in B_j T \mapsto \tilde{g}_j(u_j + v_j)$  is measurable for all  $v_j \in (B_j T)^\perp$ , and  $\int_{E^j} \tilde{g}_j = \int_{E^j} g_j$ .

Since  $\tilde{f}_i \leq f_i$  and  $g_j \leq \tilde{g}_j$  by construction, we have that (53) holds with  $f_i$  (resp.  $g_j$ ) replaced by  $\tilde{f}_i$  (resp.  $\tilde{g}_j$ ). So, by definition of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T)$  and translation invariance of the Lebesgue integral, we consider the hypothesis (53) for fixed  $(y_i)_{1 \leq i \leq k}$  to conclude

$$\prod_{i=1}^k \left( \int_{T_i} \tilde{f}_i(x + y_i) dx \right)^{c_i} \leq e^{D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T)} \prod_{j=1}^m \left( \int_{B_j T} \tilde{g}_j \left( u + \sum_{i=1}^k B_{ij, T_i^\perp} y_i \right) du \right)^{d_j} \quad \forall y_i \in T_i^\perp, 1 \leq i \leq k.$$

By Fubini-Tonelli, the map  $y_i \in T_i^\perp \mapsto \int_{T_i} \tilde{f}_i(u + y_i) du$  is measurable. Similarly,  $v_j \in (B_j T)^\perp \mapsto \int_{B_j T} \tilde{g}_j(u + v_j) du$  is measurable. Therefore, by definition of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T})$ , we find

$$\prod_{i=1}^k \left( \int_{T_i^\perp} \int_{T_i} \tilde{f}_i(x + y) dx dy \right)^{c_i} \leq e^{D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T})} \prod_{j=1}^m \left( \int_{(B_j T)^\perp} \int_{B_j T} \tilde{g}_j(u + v) dudv \right)^{d_j}.$$

By an application of Tonelli's theorem combined with the previously noted identities  $\int_{E_i} \tilde{f}_i = \int_{E_i} f_i$  and  $\int_{E^j} \tilde{g}_j = \int_{E^j} g_j$ , we conclude

$$\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i} \leq e^{D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T})} \prod_{j=1}^m \left( \int_{E^j} g_j \right)^{d_j}$$



for any non-negative, bounded measurable functions  $(f_i)_{1 \leq i \leq k}$  and  $(g_j)_{1 \leq j \leq m}$  satisfying (52). It was already noted in the proof of Theorems 1.14 and 2.1 that bounded functions saturate the definition of  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , so the first claim is proved.

The statement for  $D_g$  follows by an identical argument, considering only centered gaussian functions. In fact, it is even easier since there are no measurability considerations to deal with.  $\square$

**Lemma 3.6.** *Let  $T \subset E_0$  be a critical subspace for the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . It holds that*

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}).$$

**Remark 3.7.** The same conclusion also holds for  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , though we do not need to prove it separately here. It will follow from subsequent results, and is stated later as Corollary 3.11.

*Proof.* We take advantage of the entropic characterization of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  in Proposition 3.1 to give a simple proof, though it is also possible to appeal to the functional formulation. We assume  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) < +\infty$ , since otherwise the corresponding claims follow from Lemma 3.4 and the fact that  $D_g > -\infty$  for any datum.

Recall that critical subspaces are of product form by definition. Define  $T_i := \pi_{E_i} T$ , and let  $X_i, Y_i$  be independent, gaussian random vectors in  $T_i, T_i^\perp$ , respectively (each having finite entropies by definition). Define  $Z_i = \epsilon^{-1} X_i + Y_i$ , which is a gaussian random vector in  $E_i$ , and note that

$$\begin{aligned} h(c_i^{-1} Z_i) &= h(\epsilon^{-1} c_i^{-1} X_i, c_i^{-1} Y_i) = h(\epsilon^{-1} c_i^{-1} X_i) + h(c_i^{-1} Y_i) \\ &= \dim(T_i) \log(\epsilon^{-1}) + h(c_i^{-1} X_i) + h(c_i^{-1} Y_i), \quad 1 \leq i \leq k. \end{aligned}$$

Now, for any coupling of the  $(Z_i)_{1 \leq i \leq k}$  satisfying (51), it follows by subadditivity of entropy (Proposition A.3) and the scaling property (Proposition A.1) that

$$h\left(\sum_{i=1}^k B_{ij} Z_i\right) \tag{54}$$

$$= h\left(\epsilon^{-1} \sum_{i=1}^k B_{ij} X_i + \pi_{B_j T} \sum_{i=1}^k B_{ij} Y_i, \pi_{(B_j T)^\perp} \sum_{i=1}^k B_{ij} Y_i\right) \tag{55}$$

$$\leq h\left(\epsilon^{-1} \sum_{i=1}^k B_{ij} X_i + \pi_{B_j T} \sum_{i=1}^k B_{ij} Y_i\right) + h\left(\pi_{(B_j T)^\perp} \sum_{i=1}^k B_{ij} Y_i\right) \tag{56}$$

$$= \dim(B_j T) \log(\epsilon^{-1}) + h\left(\sum_{i=1}^k B_{ij} X_i + \epsilon \pi_{B_j T} \sum_{i=1}^k B_{ij} Y_i\right) + h\left(\pi_{(B_j T)^\perp} \sum_{i=1}^k B_{ij} Y_i\right). \tag{57}$$

Recalling Remark 3.3, we note that all terms are finite.

So, for any  $\epsilon > 0$ , we use the assumption that  $T$  was critical to cancel the  $\log(\epsilon^{-1})$  terms to find

$$\begin{aligned} D(\mathbf{c}, \mathbf{d}, \mathbf{B}) &\geq \sum_{i=1}^k c_i h(c_i^{-1} X_i) - \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} X_i + \epsilon \pi_{B_j T} \sum_{i=1}^k B_{ij} Y_i\right) \\ &\quad + \sum_{i=1}^k c_i h(c_i^{-1} Y_i) - \sum_{j=1}^m d_j h\left(\pi_{(B_j T)^\perp} \sum_{i=1}^k B_{ij} Y_i\right) \end{aligned}$$

for some coupling of the  $(Z_i)_{1 \leq i \leq k} \equiv (X_i, Y_i)_{1 \leq i \leq k}$ . Next, since gaussians have bounded second moments by definition, weak upper semicontinuity of entropy (Proposition A.4) implies

$$\limsup_{\epsilon \rightarrow 0} h\left(\sum_{i=1}^k B_{ij} X_i + \epsilon \pi_{B_j T} \sum_{i=1}^k B_{ij} Y_i\right) \leq h\left(\sum_{i=1}^k B_{ij} X_i\right), \quad 1 \leq j \leq m.$$

In combination with the previous estimate, we have

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) \geq \sum_{i=1}^k c_i h(c_i^{-1} X_i) - \max_{P_{\mathbf{X}} \in \Pi(P_{X_1}, \dots, P_{X_k})} \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij, T_i} X_i\right) \\ + \sum_{i=1}^k c_i h(c_i^{-1} Y_i) - \max_{P_{\mathbf{Y}} \in \Pi(P_{Y_1}, \dots, P_{Y_k})} \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij, T_i^\perp} Y_i\right).$$

Since we chose  $(X_i)_{1 \leq i \leq k}$  and  $(Y_i)_{1 \leq i \leq k}$  to be arbitrary gaussians on their respective subspaces, it follows that

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) \geq D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}).$$

Comparing to Lemma 3.4, the claim is proved. □

### 3.3. Necessary conditions for finiteness.

**Proposition 3.8.** *If  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) < +\infty$  or  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) < +\infty$ , then we must have (22) and*

$$\sum_{i=1}^k c_i \dim(\pi_{E_i} T) \leq \sum_{j=1}^m d_j \dim(B_j T) \text{ for all product-form subspaces } T \subseteq E_0. \quad (58)$$

*In particular, each  $B_j$  must be surjective.*

*Proof.* Since  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  by definition, it suffices to establish necessary conditions for  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  to be finite. The condition (22) can be easily seen using the scaling property of entropy (Proposition A.1) by multiplying all random variables in (50) by a common scalar factor.

The necessity of (58) follows immediately from the proof of Lemma 3.6, but without cancelling the  $\sum_{i=1}^k c_i \dim(T_i) \log(\epsilon^{-1})$  and  $\sum_{j=1}^m d_j \dim(B_j T) \log(\epsilon^{-1})$  terms. These terms cancelled previously under the assumption that  $T$  was critical, but this will not be the case if we assume  $T$  is such that

$$\sum_{i=1}^k c_i \dim(\pi_{E_i} T) > \sum_{j=1}^m d_j \dim(B_j T),$$

leading to an arbitrarily large lower bound on  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  as  $\epsilon$  vanishes.

To see that each  $B_j$  must be surjective, we take  $T = E_0$  and compare (1) to (58). □

**3.4. Sufficient conditions for finiteness and gaussian-extremizability.** The goal of this section is to establish the sufficiency of the conditions in Theorem 1.27 for finiteness and gaussian-extremizability. We start with a technical lemma, which is the counterpart of [8, Lemma 5.1] for our setting.

**Lemma 3.9.** *Define  $N := \dim(E_0)$  and let  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  be a datum such that (22) holds and*

$$\sum_{j=1}^m d_j \dim(B_j T) \geq \sum_{i=1}^k c_i \dim(\pi_{E_i} T) \text{ for all product-form subspaces } T \subseteq E_0. \quad (59)$$

*In particular, this implies each  $B_j$  is surjective. Then, there is a real number  $c > 0$  such that, for every orthonormal basis  $(e_n)_{1 \leq n \leq N}$  of  $E_0$  with the property that each  $e_n \in E_i$  for some  $1 \leq i \leq k$ , there exists a set  $I_j \subseteq \{1, \dots, N\}$  for each  $1 \leq j \leq m$  with  $|I_j| = \dim(E^j)$  such that*

$$\sum_{j=1}^m d_j |I_j \cap \{n+1, \dots, N\}| \geq \sum_{i=1}^k c_i |S_i \cap \{n+1, \dots, N\}| \text{ for all } 0 \leq n \leq N, \quad (60)$$

where  $S_i := \{n : e_n \in E_i\}$ ,  $1 \leq i \leq k$  and

$$\left\| \bigwedge_{n \in I_j} B_j e_n \right\|_{E_j} \geq c \text{ for all } 1 \leq j \leq m. \quad (61)$$

Moreover, if there are no critical subspaces, then there is a constant  $\delta > 0$  depending only on the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  such that

$$\sum_{j=1}^m d_j |I_j \cap \{n+1, \dots, N\}| \geq \sum_{i=1}^k c_i |S_i \cap \{n+1, \dots, N\}| + \delta \text{ for all } 0 < n < N. \quad (62)$$

*Proof.* Since the space of all orthonormal bases is compact, and the number of possible  $I_j$  is finite, it follows by continuity and compactness that (61) may be replaced by the weaker assumption that  $(B_j e_n)_{n \in I_j}$  are linearly independent in  $E_j$  for each  $1 \leq j \leq m$ .

Now, we construct  $I_j$  by a backwards greedy algorithm. Specifically, we set  $I_j$  equal to those indices  $n$  for which  $B_j e_n$  is not in the linear span of  $\{B_j e_{n'}; n < n' \leq \dim(E_0)\}$ . Since  $B_j$  is surjective, we will have  $|I_j| = \dim(E_j)$ . To prove (60), we first fix  $n$  satisfying  $0 < n < N$ , and apply (59) with  $T$  equal to the span of  $\{e_{n+1}, \dots, e_{\dim(E_0)}\}$ , which is of product form by the assumption that each  $e_n \in \bigcup_{i=1}^k E_i$ . Specifically, due to construction of  $I_j$  we have

$$\dim(B_j T) = |I_j \cap \{n+1, \dots, N\}|.$$

On the other hand,

$$\dim(\pi_{E_i} T) = |S_i \cap \{n+1, \dots, N\}|,$$

establishing (60) when  $0 < n < N$ . The case of  $n = N$  is trivial, and the case of  $n = 0$  follows from equality in (59) for  $T = E_0$  since  $|S_i \cap \{1, \dots, N\}| = |S_i| = \dim(E_i)$ , and  $|I_j \cap \{1, \dots, N\}| = |I_j| = \dim(E_j)$ .

Now, if there are no critical subspaces, then there is  $\delta > 0$  depending only on the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  such that (59) can be refined to

$$\sum_{j=1}^m d_j \dim(B_j T) \geq \sum_{i=1}^k c_i \dim(\pi_{E_i} T) + \delta \quad (63)$$

for all non-zero proper subspaces  $T \subset E_0$ . Indeed, this easily follows since the left and right sides of (59) only take finitely many values. Incorporating this into the previous analysis gives (62).  $\square$

**Proposition 3.10** (Sufficient conditions for finiteness and gaussian-extremizability). *If the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is such that (22) and (23) hold, then  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is finite. If it further holds that  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is simple, then  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable.*

*Proof.* The argument follows the strategy of proof for [8, Proposition 5.2], but is recast in terms of entropies which we find more convenient in the present setting<sup>2</sup>. Define  $N := \dim(E_0)$  and consider gaussian random vectors  $Z_i$  in  $E_i$ ,  $1 \leq i \leq k$ . It is trivially true that

$$\max_{P_{Z_1, \dots, Z_k}} \sum_{j=1}^m d_j h \left( \sum_{i=1}^k c_i B_{ij} Z_i \right) \geq \sum_{j=1}^m d_j h \left( \sum_{i=1}^k c_i B_{ij} Z'_i \right), \quad (64)$$

where  $Z'_i = Z_i$  in distribution for each  $1 \leq i \leq k$ , and  $Z'_1, \dots, Z'_k$  are independent. Without loss of generality, we may write

$$Z'_i = \sum_{n \in S_i} W_n e_n, \quad 1 \leq i \leq k$$

<sup>2</sup>Note that we work exclusively with gaussian random vectors here, so all computations can be stated in terms of determinants, using the identity (49).

where  $(e_n)_{n \in S_i} \subset E_i \subset E_0$  is an orthonormal basis for  $E_i$ ,  $1 \leq i \leq k$ ,  $(S_i)_{1 \leq i \leq k}$  is a partition of  $\{1, \dots, N\}$ , and  $(W_n)_{1 \leq n \leq N}$  is a collection of independent one-dimensional gaussian random variables. We may further assume that the indices are chosen to satisfy  $h(W_1) \leq \dots \leq h(W_N)$ . Now, we invoke Lemma 3.9 and it follows from the scaling property for entropy (Proposition A.1) that

$$h\left(\sum_{i=1}^k B_{ij} Z_i'\right) = h\left(\sum_{n=1}^N W_n B_j e_n\right) \geq h\left(\sum_{n \in I_j} W_n B_j e_n\right) \geq h((W_n)_{n \in I_j}) + C, \quad (65)$$

for some constant  $C$  depending only on the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  since  $(B_j e_n)_{n \in I_j}$  form a basis of  $E^j$  with a lower bound on degeneracy.

Now, by telescoping and Lemma 3.9, we may write

$$\begin{aligned} & \sum_{j=1}^m d_j h((W_n)_{n \in I_j}) \\ &= \sum_{n=1}^N h(W_n) \sum_{j=1}^m d_j |I_j \cap \{n\}| \\ &= \left(\sum_{j=1}^m d_j |I_j|\right) h(W_1) + \sum_{n=1}^{N-1} (h(W_{n+1}) - h(W_n)) \sum_{j=1}^m d_j |I_j \cap \{n+1, \dots, N\}| \\ &= \left(\sum_{i=1}^k c_i |S_i|\right) h(W_1) + \sum_{n=1}^{N-1} (h(W_{n+1}) - h(W_n)) \sum_{j=1}^m d_j |I_j \cap \{n+1, \dots, N\}| \\ &\geq \left(\sum_{i=1}^k c_i |S_i|\right) h(W_1) + \sum_{n=1}^{N-1} (h(W_{n+1}) - h(W_n)) \left(\sum_{i=1}^k c_i |S_i \cap \{n+1, \dots, N\}| + \delta\right) \\ &= \sum_{n=1}^N h(W_n) \sum_{i=1}^k c_i |S_i \cap \{n\}| + \delta \sum_{n=1}^{N-1} (h(W_{n+1}) - h(W_n)) \\ &= \sum_{i=1}^k c_i \left(\sum_{n \in S_i} h(W_n)\right) + \delta (h(W_N) - h(W_1)) \\ &= \sum_{i=1}^k c_i h(Z_i) + \delta (h(W_N) - h(W_1)). \end{aligned}$$

So, we conclude

$$\sum_{i=1}^k c_i h(c_i^{-1} Z_i) - \max_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} Z_i\right) \leq C' - \delta (h(W_N) - h(W_1)) \quad (66)$$

for constants  $C', \delta \geq 0$  depending only on the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . In particular,  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is finite.

Now, if  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is simple, then the last claim of Lemma 3.9 implies  $\delta > 0$ . Since the LHS of (66) is invariant to scaling each  $Z_i$  by a common factor (due to the scaling condition (22)), it easily follows that there are constants  $c_1, c_2 > 0$  depending only on the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  such that we may restrict our attention to  $Z_i$  satisfying

$$c_1 \leq \text{Var}(W_1) \leq \lambda_{\min}(\text{Cov}(Z_i)) \leq \lambda_{\max}(\text{Cov}(Z_i)) \leq \text{Var}(W_N) \leq c_2 \text{ for each } 1 \leq i \leq k.$$

Thus, in supremizing the LHS of (50), it suffices to consider gaussian  $Z_i$  with covariances in a compact set, with eigenvalues uniformly bounded away from zero. Equivalently, in supremizing the

functional

$$(K_1, \dots, K_k) \mapsto \left( \sum_{i=1}^k c_i \log \det K_i - \max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det (B_j \Lambda_{\mathbf{c}} K \Lambda_{\mathbf{c}} B_j^*) \right)$$

over  $\prod_{i=1}^k S^+(E_i)$ , it suffices to consider each  $K_i$  in a compact set, with eigenvalues bounded away from zero. It therefore follows by upper-semicontinuity (i.e., Lemma 2.16) that an extremizer exists. Thus,  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is gaussian-extremizable.  $\square$

*Proof of Theorem 1.27.* The claim is an immediate corollary of Propositions 3.8 and 3.10.  $\square$

We may now also prove Theorem 1.3, which proceeds just as in [8, Proof of Theorems 1.9 and 1.15]:

*Proof of Theorem 1.3.* In view of Theorem 2.2, we only need to prove that  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . To do this, we induct on the dimension  $\dim(E_0)$ . The case  $\dim(E_0) = 0$  is trivial, so assume the claim holds for smaller values of  $\dim(E_0)$ .

We may assume  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) < +\infty$ , else the claim is trivial since  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) \geq D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  by definition. Thus, we assume that (22) and (23) hold, since these are necessary conditions for finiteness of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  by Theorem 1.27. If  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is simple, then it is also gaussian-extremizable by Theorem 1.27, so the desired claim follows by Theorem 2.1. On the other hand, if  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is not simple, then by Lemma 3.6 and the definition of simple, there exists a critical subspace  $T \subset E_0$  for which

$$D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}).$$

By Lemma 3.4, we also have

$$D(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq D(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}).$$

By the induction hypothesis,

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

and

$$D(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}) = D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}).$$

Combining the above estimates, we have

$$D(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq D(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}) = D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}) = D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}).$$

Taken together with the trivial inequality  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , we must have equality. This closes the induction and completes the proof.  $\square$

In analogy to Lemma 3.6, the following corollary is now immediate. It is not needed elsewhere, but we state it for completeness.

**Corollary 3.11.** *Let  $T \subset E_0$  be a critical subspace for the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . It holds that*

$$D(\mathbf{c}, \mathbf{d}, \mathbf{B}) = D(\mathbf{c}, \mathbf{d}, \mathbf{B}_T) + D(\mathbf{c}, \mathbf{d}, \mathbf{B}_{E_0/T}).$$

#### 4. CONNECTIONS TO OTHER BRASCAMP-LIEB-TYPE INEQUALITIES

**4.1. The Brascamp-Lieb inequality.** It is clear by now that the Brascamp-Lieb inequality is a special case of the forward-reverse inequality. On the other hand, if we assume (22) holds and that (23) holds for *all* subspaces, not just those of product form, then finiteness of  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  be established as a consequence of the finiteness conditions for the forward Brascamp-Lieb inequality. The argument is as follows, and is due to Michael Christ.

Assume (22) and further assume that (23) holds for all subspaces  $T$  (not just those of product form). Define the index sets  $I := \{1, \dots, k\}$  and  $J = \{1, \dots, m\}$ . Since the statement and conclusion are invariant to rescaling  $\mathbf{c}, \mathbf{d}$  by the same constant, we assume without loss of generality that

$\max_{i \in I} c_i < 1$  and  $\max_{j \in J} d_j < 1$ . Now, assuming  $I, J$  are disjoint index sets, we define the augmented index set  $J^* = I \cup J$ . For  $j \in J^* \setminus J$ , define  $d_j = (1 - c_i)$ ,  $E^j = E_i$ , and  $B_j = \pi_{E_i} \Lambda_{\mathbf{c}}^{-1}$ . Now, if  $(f_i)_{i \in I}$  and  $(g_j)_{j \in J}$  satisfy (2), then defining  $g_j = f_i$  for  $j \in J^* \setminus J$ , it follows that

$$\prod_{i \in I} f_i(\pi_{E_i}(x)) \leq \prod_{j \in J} g_j^{d_j}(B_j \Lambda_{\mathbf{c}} x) \prod_{i \in I} f_i^{1-c_i}(\pi_{E_i} x) = \prod_{j \in J^*} g_j^{d_j}(B_j \Lambda_{\mathbf{c}} x).$$

Integrating over both sides and using the fact that  $E_0 = \bigoplus_{i=1}^k E_i$ , we obtain

$$\prod_{i \in I} \int_{E_i} f_i \leq \int_{E_0} \prod_{i \in I} f_i(\pi_{E_i}(x)) dx \leq \int_{E_0} \prod_{j \in J^*} g_j^{d_j}(B_j \Lambda_{\mathbf{c}} x) dx. \quad (67)$$

By the finiteness criteria for the forward Brascamp-Lieb inequality [8, Theorem 1.13],

$$\int_{E_0} \prod_{j \in J^*} g_j^{d_j}(B_j \Lambda_{\mathbf{c}} x) dx \leq e^D \prod_{j \in J^*} \left( \int_{E^j} g_j \right)^{d_j} = e^D \prod_{j \in J} \left( \int_{E^j} g_j \right)^{d_j} \prod_{i \in I} \left( \int_{E_i} f_i \right)^{1-c_i}, \quad (68)$$

where  $D < +\infty$  provided

$$\sum_{j \in J^*} d_j \dim(B_j T) \geq \dim(T) \text{ for all subspaces } T \subseteq E_0, \quad (69)$$

and further holding with equality when  $T = E_0$ . Assuming this is true for the moment, we combine (67) and (68) to conclude  $D(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq D < +\infty$ , as desired.

So, to verify (69) and therefore justify the application (68), observe that, since we assumed (23) for all subspaces  $T \subseteq E_0$ ,

$$\sum_{j \in J^*} d_j \dim(B_j T) = \sum_{j \in J} d_j \dim(B_j T) + \sum_{i \in I} (1 - c_i) \dim(\pi_{E_i} T) \quad (70)$$

$$\geq \sum_{i \in I} \dim(\pi_{E_i} T) \geq \dim(T), \quad (71)$$

with equality holding when  $T = E_0$  by (22).

**Remark 4.1.** The disadvantage of the above argument is that it will not, in general, recover the sharp constant (and therefore the gaussian saturation property) for the forward-reverse Brascamp-Lieb inequality, even under the stronger condition that (23) holds for all subspaces  $T \subseteq E_0$ .

**Remark 4.2.** Example 1.6 (reverse Young inequality) provides an important counterpoint to the above discussion. The reader can check that (23) is verified for all product-form subspaces, however it fails to hold for some non-product form subspaces. Hence, bootstrapping the direct Brascamp-Lieb inequality as above would fail to give a finite constant in the reverse Young inequality.

**4.2. The Barthe-Wolff inverse Brascamp-Lieb inequality.** The following ‘‘inverse’’ Brascamp-Lieb inequality was announced by Barthe and Wolff in the note [5] and proved rigorously in [6]. We write it in a form to emphasize the connection to Theorem 1.3.

**Theorem 4.3.** *Let  $C \in (-\infty, +\infty]$  be any constant, and let previously introduced notation prevail. For any measurable functions  $f_i : E_i \rightarrow \mathbb{R}^+$ ,  $1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+$ ,  $1 \leq j \leq m$ ,*

$$\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i} \prod_{j=1}^m \left( \int_{E^j} g_j \right)^{-d_j} \leq e^C \int_{E_0} \prod_{i=1}^k f_i^{c_i}(\pi_{E_i} x) \prod_{j=1}^m g_j^{-d_j}(B_j x) dx \quad (72)$$

*if and only if (72) holds for all centered gaussian functions  $(f_i)_{1 \leq i \leq k}$  and  $(g_j)_{1 \leq j \leq m}$ .*

For sake of comparison, we restate the gaussian saturation part of Theorem 1.3 here in equivalent form as follows:

**Theorem 4.4.** *Let  $D \in (-\infty, +\infty]$  be any constant, and let previously introduced notation prevail. For any measurable functions  $f_i : E_i \rightarrow \mathbb{R}^+$ ,  $1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+$ ,  $1 \leq j \leq m$  satisfying*

$$\prod_{i=1}^k f_i^{c_i}(\pi_{E_i} x) \leq \prod_{j=1}^m g_j^{d_j}(B_j x) \quad \forall x \in E_0, \quad (73)$$

we have

$$\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i} \leq e^D \prod_{j=1}^m \left( \int_{E^j} g_j \right)^{d_j}, \quad (74)$$

if and only if (74) holds for all centered gaussian functions  $(f_i)_{1 \leq i \leq k}$  and  $(g_j)_{1 \leq j \leq m}$  satisfying (73).

To see the connection between the two results, we first note that Theorem 4.4 implies Theorem 4.3 by augmenting the datum  $(\mathbf{c}, \mathbf{d}, \mathbf{B})$  with  $E^{m+1} = E_0$ ,  $d_{m+1} = 1$  and  $B_{m+1} = \text{id}_{E_0}$ . By choosing the function  $g_{m+1} : E^{m+1} \rightarrow \mathbb{R}^+$  according to

$$g_{m+1}(x) = \prod_{i=1}^k f_i^{c_i}(\pi_{E_i} x) \prod_{j=1}^m g_j^{-d_j}(B_j x),$$

the hypothesis (73) is satisfied, and therefore (72) follows from (74). For given functions  $(f_i)_{1 \leq i \leq k}$  and  $(g_j)_{1 \leq j \leq m}$ , the above choice of  $g_{m+1}$  is clearly best-possible, so the best constant  $C$  in (72) must be equal to the best constant  $D$  in Theorem 4.4 for the augmented datum, which can be computed by considering only centered gaussian functions.

In fact, the reverse is also true. That is, Theorem 4.4 may be derived from Theorem 4.3. The argument is a bit less straightforward in comparison, but nevertheless brief. The idea is to apply Theorem 4.3 with exponents  $c'_i = 1 + tc_i$  and  $d'_j = td_j$ , where  $t > 0$  is a parameter that will tend to  $+\infty$ . For this choice of exponents, we apply the pointwise inequality (73) to see that the RHS of (72) can be upper bounded as

$$\int_{E_0} \prod_{i=1}^k f_i^{c'_i}(\pi_{E_i} x) \prod_{j=1}^m g_j^{-d'_j}(B_j x) dx \leq \int_{E_0} \left( \prod_{i=1}^k f_i^{c'_i}(\pi_{E_i} x) \right) \left( \prod_{i=1}^k f_i^{-tc_i}(\pi_{E_i} x) \right) dx = \prod_{i=1}^k \left( \int_{E_i} f_i \right).$$

Invoking (72) itself and dividing exponents by  $t$ , we find that (73) implies

$$\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i} \leq e^{C_t/t} \prod_{j=1}^m \left( \int_{E^j} g_j \right)^{d_j},$$

where  $C_t$  denotes the best constant in the inequality (72) for the exponents  $(c'_i)_{1 \leq i \leq k}$  and  $(d'_j)_{1 \leq j \leq m}$ . In particular, for  $D$  the best constant in (74), we have  $D \leq C_t/t$  for all  $t > 0$ . By the gaussian saturation claim of Theorem 4.3 and direct computation (see [6, Section 2.2]), one may calculate

$$e^{2C_t} = \sup \frac{\det \left( \sum_{i=1}^k c'_i \pi_{E_i}^* C_i \pi_{E_i} - \sum_{j=1}^m d'_j B_j^* A_j B_j \right)}{\prod_{i=1}^k (\det C_i)^{c'_i} \prod_{j=1}^m (\det A_j)^{-d'_j}}, \quad (75)$$

where the supremum is over all  $C_i \in S^+(E_i)$  and  $A_j \in S^+(E^j)$  satisfying

$$\sum_{i=1}^k c'_i \pi_{E_i}^* C_i \pi_{E_i} \geq \sum_{j=1}^m d'_j B_j^* A_j B_j. \quad (76)$$

The set of  $(C_i)_{1 \leq i \leq k}$  and  $(A_j)_{1 \leq j \leq m}$  satisfying (76) are monotone decreasing in  $t$  (with respect to inclusion), so in calculating  $\liminf_{t \rightarrow \infty} C_t/t$ , we need only consider positive-definite  $(C_i)_{1 \leq i \leq k}$  and

$(A_j)_{1 \leq j \leq m}$  in the intersection of all such sets; i.e., those satisfying

$$\sum_{i=1}^k c_i \pi_{E_i}^* C_i \pi_{E_i} \geq \sum_{j=1}^m d_j B_j^* A_j B_j. \tag{77}$$

Assuming (77) holds, we bound

$$\begin{aligned} \frac{\det \left( \sum_{i=1}^k c'_i \pi_{E_i}^* C_i \pi_{E_i} - \sum_{j=1}^m d'_j B_j^* A_j B_j \right)^{1/t}}{\prod_{i=1}^k (\det C_i)^{1/t+c_i} \prod_{j=1}^m (\det A_j)^{-d_j}} &\leq \frac{\det \left( \sum_{i=1}^k c'_i \pi_{E_i}^* C_i \pi_{E_i} \right)^{1/t}}{\prod_{i=1}^k (\det C_i)^{1/t+c_i} \prod_{j=1}^m (\det A_j)^{-d_j}} \\ &= \frac{\prod_{i=1}^k (1 + t c_i)^{\dim(E_i)/t}}{\prod_{i=1}^k (\det C_i)^{c_i} \prod_{j=1}^m (\det A_j)^{-d_j}}. \end{aligned}$$

Hence,

$$\liminf_{t \rightarrow \infty} C_t/t \leq \sup \left( \frac{1}{2} \sum_{j=1}^m d_j \log \det A_j - \frac{1}{2} \sum_{i=1}^k c_i \log \det C_i \right),$$

where the supremum is over all  $C_i \in S^+(E_i)$  and  $A_j \in S^+(E^j)$  satisfying (77). This is precisely the best constant  $D$  obtained in (74) by considering only centered gaussian functions, so the proof is complete.

**Remark 4.5.** The above argument showing equivalence of Theorems 4.3 and 4.4 is due to Pawel Wolff. Despite their formal equivalence, both results have their merits, and the techniques used to derive each are complimentary. In particular, Barthe and Wolff use an optimal transport argument, while our proof relies primarily on duality and structural decomposition. Additionally, the different formulations of the results have their respective advantages. For example, Theorem 1.3 highlights the unification of the Brascamp-Lieb and Barthe inequalities, together with the duality of best constants (6). On the other hand, Barthe and Wolff’s formulation emphasizes an inverse principle to Lieb’s [23]. Our proof is perhaps simpler since it avoids the detailed case analysis encountered in [6], but preference may depend on the reader’s taste.

**Remark 4.6.** Theorem 4.3 is a particular case of the general inverse inequality by Barthe and Wolff which allows for integration against a nontrivial gaussian kernel in the RHS of (72), and for which the gaussian saturation property remains valid. As we will see in the next section, their geometric inequality can be recovered in full generality with nontrivial gaussian kernel as a consequence of the geometric forward-reverse Brascamp-Lieb inequality. Hence, there is a formal equivalence between the geometric Barthe-Wolff inequalities stated with (i) trivial gaussian kernel; or (ii) nontrivial gaussian kernel. Although we do not pursue it here, it is an interesting question whether this formal equivalence continues to hold for non-geometric instances of the Barthe-Wolff inequality.

**4.3. Inequalities with Gaussian kernels.** In this section, we establish inequalities for integrals against gaussian kernels as applications of our main results. They are easy corollaries of the geometric forward-reverse Brascamp-Lieb inequality. Similar results could be stated for the general forward-reverse inequality, but we restrict attention to the geometric case to simplify the discussion.

**Definition 4.7.** For a Euclidean space  $E$ , we let  $\gamma_E$  denote the standard gaussian measure on  $E$ . That is,

$$d\gamma_E(x) = \frac{1}{(2\pi)^{\dim(E)/2}} e^{-\frac{1}{2}|x|^2} dx.$$

**Theorem 4.8.** Let  $H$  be a Euclidean space, and  $Q \in S(H)$  with signature  $(s^+(Q), s^-(Q))$ . Consider linear maps  $U_i : H \rightarrow E_i$  and  $V_j : H \rightarrow E^j$  satisfying  $U_i U_i^* = \text{id}_{E_i}$  and  $V_j V_j^* = \text{id}_{E^j}$ , for all



$1 \leq i \leq k$  and  $1 \leq j \leq m$ . Let  $(c_i)_{1 \leq i \leq k}$  and  $(d_j)_{1 \leq j \leq m}$  be positive numbers, and suppose that

$$Q + \sum_{i=1}^k c_i U_i^* U_i = \sum_{j=1}^m d_j V_j^* V_j > 0, \text{ and } \dim(H) \geq s^+(Q) + \sum_{i=1}^k \dim(E_i). \quad (78)$$

If measurable functions  $f_i : E_i \rightarrow \mathbb{R}^+$ ,  $1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+$ ,  $1 \leq j \leq m$  satisfy

$$\prod_{i=1}^k f_i^{c_i}(U_i x) \leq \prod_{j=1}^m g_j^{d_j}(V_j x) \quad \forall x \in H, \quad (79)$$

then

$$\prod_{i=1}^k \left( \int_{E_i} f_i d\gamma_{E_i} \right)^{c_i} \leq \prod_{j=1}^m \left( \int_{E^j} g_j d\gamma_{E^j} \right)^{d_j}. \quad (80)$$

*Proof.* Decompose  $Q = Q^+ - Q^-$ , with  $Q^+, Q^- \in S^+(H)$ . By spectral decomposition, write

$$Q^+ = \sum_{\ell=1}^{s^+(Q)} \lambda_\ell u_\ell^* u_\ell; \quad Q^- = \sum_{\ell=1}^{s^-(Q)} \mu_\ell v_\ell^* v_\ell,$$

where  $\lambda_\ell > 0$  (resp.  $\mu_\ell > 0$ ) and  $u_\ell u_\ell^* = \text{id}_{\mathbb{R}}$  (resp.  $v_\ell v_\ell^* = \text{id}_{\mathbb{R}}$ ). Thus, the first assumption in (78) can be written as

$$\sum_{i=1}^k c_i U_i^* U_i + \sum_{\ell=1}^{s^+(Q)} \lambda_\ell u_\ell^* u_\ell = \sum_{j=1}^m d_j V_j^* V_j + \sum_{\ell=1}^{s^-(Q)} \mu_\ell v_\ell^* v_\ell > 0 \quad (81)$$

In particular the LHS is a linear map with rank equal to  $\dim(H)$  by the positive-definiteness assumption, so by subadditivity of rank, it holds that  $\dim(H) \leq s^+(Q) + \sum_{i=1}^k \dim(E_i)$ . By (78), we must have equality. Thus, we can consider the map

$$x \in H \rightarrow (U_1 x, \dots, U_k x, u_1 x, \dots, u_{s^+(Q)} x)$$

as a bijective linear map from  $H$  to  $H$ . Now, if (79) is satisfied, then (81) implies that we also have

$$\begin{aligned} & \prod_{i=1}^k \left( f_i(U_i x) \frac{1}{(2\pi)^{\dim(E_i)/2}} e^{-\frac{1}{2}|U_i x|^2} \right)^{c_i} \times \prod_{\ell=1}^{s^+(Q)} \phi^{\lambda_\ell}(u_\ell x) \\ & \leq \prod_{j=1}^m \left( g_j(V_j x) \frac{1}{(2\pi)^{\dim(E^j)/2}} e^{-\frac{1}{2}|V_j x|^2} \right)^{d_j} \times \prod_{\ell=1}^{s^-(Q)} \phi^{\mu_\ell}(v_\ell x), \quad \forall x \in H, \end{aligned}$$

where  $\phi$  denotes the standard gaussian density

$$\phi(z) := \frac{1}{2\pi} e^{-\frac{1}{2}|z|^2}, \quad z \in \mathbb{R}.$$

Since  $\int_{\mathbb{R}} \phi = 1$ , the inequality (80) follows from an application of Corollary 1.19.  $\square$

An important consequence of Theorem 4.8 is the following geometric inverse Brascamp-Lieb inequality proved by Barthe and Wolff [6, Theorem 4.7], recovered here in full generality. We remark that the reverse Hölder-type inequality for gaussian random vectors due to Chen, Dafnis and Paouris [15, Theorem 1(ii)] follows as a direct consequence [6, Section 4.3], so should be considered as yet another example. The direct Chen-Dafnis-Paouris inequality [15, Theorem 1(i)] is a consequence of the forward Brascamp-Lieb inequality.

**Corollary 4.9.** *Let  $H$  be a Euclidean space, and consider linear maps  $U_i : H \rightarrow E_i$  and  $V_j : H \rightarrow E^j$  satisfying  $U_i U_i^* = \text{id}_{E_i}$  and  $V_j V_j^* = \text{id}_{E^j}$ , for all  $1 \leq i \leq k$  and  $1 \leq j \leq m$ . Let  $(c_i)_{1 \leq i \leq k}$  and  $(d_j)_{1 \leq j \leq m}$  be positive numbers, and suppose that*

$$Q + \sum_{i=1}^k c_i U_i^* U_i = \sum_{j=1}^m d_j V_j^* V_j + \text{id}_H, \text{ and } \dim(H) \geq s^+(Q) + \sum_{i=1}^k \dim(E_i)$$

for  $Q : H \rightarrow H$  a symmetric operator. For all non-negative measurable functions  $f_i : E_i \rightarrow \mathbb{R}^+$ ,  $1 \leq i \leq k$  and  $g_j : E^j \rightarrow \mathbb{R}^+$ ,  $1 \leq j \leq m$ , it holds that

$$\prod_{i=1}^k \left( \int_{E_i} f_i \right)^{c_i} \prod_{j=1}^m \left( \int_{E^j} g_j \right)^{-d_j} \leq \int_H e^{-\pi \langle Qx, x \rangle} \prod_{i=1}^k f_i^{c_i}(U_i x) \prod_{j=1}^m g_j^{-d_j}(V_j x) dx.$$

*Proof.* Define  $E^{m+1} = H$ ,  $V_{m+1} = \text{id}_H$  and  $d_{m+1} = 1$ . Put

$$\tilde{f}_i(x) := f_i \left( (2\pi)^{-1/2} x \right) e^{\frac{1}{2}|x|^2}, 1 \leq i \leq k; \quad \tilde{g}_j(x) := g_j \left( (2\pi)^{-1/2} x \right) e^{\frac{1}{2}|x|^2}, 1 \leq j \leq m$$

and define  $\tilde{g}_{m+1} : H \rightarrow H$  defined according to

$$\begin{aligned} \tilde{g}_{m+1}(x) &:= \prod_{i=1}^k \tilde{f}_i^{c_i} \left( (2\pi)^{-1/2} U_i x \right) \prod_{j=1}^m \tilde{g}_j^{-d_j} \left( (2\pi)^{-1/2} V_j x \right) \\ &= \left( \prod_{i=1}^k f_i^{c_i} \left( (2\pi)^{-1/2} U_i x \right) \prod_{j=1}^m g_j^{-d_j} \left( (2\pi)^{-1/2} V_j x \right) \right) \exp \left( \frac{1}{2}|x|^2 - \frac{1}{2} \langle Qx, x \rangle \right). \end{aligned}$$

Now, by change of variables  $u \leftarrow (2\pi)^{-1/2} x$ ,

$$\int_{E_i} \tilde{f}_i(x) d\gamma_{E_i}(x) = \int_{E_i} f_i(u) du, \quad 1 \leq i \leq k; \quad \int_{E^j} \tilde{g}_j(x) d\gamma_{E^j}(x) = \int_{E^j} g_j(u) du, \quad 1 \leq j \leq m$$

and

$$\int_H \tilde{g}_{m+1}(x) d\gamma_H(x) = \int_H e^{-\pi \langle Qu, u \rangle} \prod_{i=1}^k f_i^{c_i}(U_i u) \prod_{j=1}^m g_j^{-d_j}(V_j u) du.$$

So, the claim follows from Theorem 4.8. □

**Remark 4.10.** We have seen above that the geometric Barthe-Wolff inequality [6, Theorem 4.7] follows as a consequence of Corollary 1.19, the latter being a special case of the complete characterization of geometric instances of the Forward-Reverse Brascamp-Lieb inequality given in Corollary 1.17. So, it appears *prima facie* that the class of geometric Forward-Reverse Brascamp-Lieb inequalities is more extensive than the geometric instances of the Barthe-Wolff inequality.

**4.4. The Anantharam-Jog-Nair inequality.** In a recent paper [1], Anantharam, Jog and Nair established the following entropic inequality:

**Theorem 4.11.** *Consider independent random vectors  $(Z_i)_{1 \leq i \leq k}$  taking values in  $(E_i)_{1 \leq i \leq k}$ , respectively, each having density with respect to Lebesgue measure, finite entropies, and finite second moments. It holds that*

$$\sum_{i=1}^k c_i h(Z_i) - \sum_{j=1}^m d_j h \left( \sum_{i=1}^k B_{ij} Z_i \right) \leq M_g(\mathbf{c}, \mathbf{d}, \mathbf{B}), \tag{82}$$

where the constant  $M_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is defined as the supremum of the LHS, taken over independent gaussian  $(Z_i)_{1 \leq i \leq k}$ . Moreover, the quantity  $M_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is finite if and only if we have the scaling condition

$$\sum_{i=1}^k c_i \dim(E_i) = \sum_{j=1}^m d_j \dim(E^j) \tag{83}$$

and the dimension condition

$$\sum_{i=1}^k c_i \dim(\pi_{E_i} T) \leq \sum_{j=1}^m d_j \dim(B_j T) \text{ for all product-form subspaces } T \subseteq E_0. \tag{84}$$

The inequality (82) is of interest because it simultaneously expresses both the entropy power inequality and the (entropic formulation [13] of) the Brascamp-Lieb inequality. The former is generally considered a consequence of the latter, obtained by considering a limiting case of parameters. As such, an inequality encompassing both simultaneously was previously not known. Analogously, it turns out that Theorem 4.11 can be derived as a corollary of Theorem 1.3 by considering a limiting case of parameters.

To give the argument, we first state an entropic characterization of  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , which directly parallels Proposition 3.1 (here, the reader is reminded of the notation introduced in Section 3.1). Specifically, the following entropic characterization of  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is a special case of [25, Theorem 1]<sup>3</sup>, which generalizes to abstract settings and extends the entropic formulation of the forward Brascamp-Lieb inequality due to Carlen and Cordero-Erasquin [13], as well as the entropic formulation of the reverse Brascamp-Lieb inequality independently put forth in [24] and [7] (the latter being specific to discrete spaces).

**Theorem 4.12.** *If  $(Z_i)_{1 \leq i \leq k}$  are compactly supported random vectors in  $(E_i)_{1 \leq i \leq k}$ , respectively, each having density with respect to Lebesgue measure and finite entropies, then*

$$\sum_{i=1}^k c_i h(c_i^{-1} Z_i) - \max_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} Z_i\right) \leq D(\mathbf{c}, \mathbf{d}, \mathbf{B}), \tag{85}$$

where the (always attained) maximum is over all couplings of the  $(Z_i)_{1 \leq i \leq k}$ . Moreover, the constant  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  is best possible.

Theorem 4.12 is proved similarly to Theorem 2.11, except that the Fenchel-Rockafellar theorem is applied to the topological vector space  $X = C_c(E_0)$ , and equivalence is shown to the functional formulation of  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$ . Readers can fill in the details as an exercise, or refer to the proof of the more general [25, Theorem 1]. Despite the similarity of statements and proof strategies, it is not immediate to derive Proposition 3.1 as a special case due to several subtle technical issues that need to be dealt with.

The following proof provides a nice example of where the entropic characterization of  $D(\mathbf{c}, \mathbf{d}, \mathbf{B})$  can be useful.

*Proof of Theorem 4.11.* First, we note that (83) and (84) are necessary conditions for finiteness, which can be checked by testing on gaussian  $(Z_i)_{1 \leq i \leq k}$  which put different variances in directions  $\pi_{E_i} T$  and  $(\pi_{E_i} T)^\perp$ . So, we assume henceforth that (83) and (84) hold.

Let  $M(\mathbf{c}, \mathbf{d}, \mathbf{B})$  denote the supremum of the LHS of (82) over all independent  $(Z_i)_{1 \leq i \leq k}$  with finite entropies and finite second moments. Note that in taking this supremum, it suffices to consider compactly supported  $(Z_i)_{1 \leq i \leq k}$ . Indeed, if  $(Z_i)_{1 \leq i \leq k}$  are not compactly supported and have finite second moments, then letting  $Z_{i,R}$  be the restriction of  $Z_i$  to the ball of radius  $R$ , we

---

<sup>3</sup>n.b. Eq. (12) in [25] should read as  $\log g_i$  (resp.  $\log f_j$ ) instead of  $g_i$  (resp.  $f_j$ ).

have  $\lim_{R \rightarrow \infty} h(Z_{i,R}) = h(Z_i)$  by dominated convergence, and  $\limsup_{R \rightarrow \infty} h\left(\sum_{i=1}^k B_{ij} Z_{i,R}\right) \leq h\left(\sum_{i=1}^k B_{ij} Z_i\right)$  by weak upper semicontinuity of Shannon entropy under a second moment constraint.

We will consider an application of the forward-reverse Brascamp-Lieb inequality with modified coefficients  $c'_i = (c_i + t)$ ,  $1 \leq i \leq k$ ,  $d'_j = d_j$ ,  $1 \leq j \leq m$  and augmented datum having  $E^{m+1} := E_0$ ,  $B_{m+1} := \text{id}_{E_0}$  and  $d'_{m+1} := t$ , where  $t$  is a parameter that will tend to  $+\infty$ . Denote this augmented datum by  $(\mathbf{c} + t, (\mathbf{d}, t), \mathbf{B} \cup \{B_{m+1}\})$ .

Considering independent gaussian  $(Z_i)_{1 \leq i \leq k}$ , let  $D(P\|Q) := \int dP \log\left(\frac{dP}{dQ}\right) \geq 0$  denote the relative entropy between probability measures  $P$  and  $Q$  satisfying  $P \ll Q$ , and for any  $t \geq 0$  observe

$$\begin{aligned} & \sum_{i=1}^k (c_i + t) h(Z_i) - \sup_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \left( \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} \tilde{Z}_i\right) + t h(\tilde{Z}_1, \dots, \tilde{Z}_k) \right) \\ &= \sum_{i=1}^k c_i h(Z_i) - \sup_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \left( \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} \tilde{Z}_i\right) - t D(P_{\mathbf{Z}} \| P_{Z_1} \times \dots \times P_{Z_k}) \right) \\ &\leq \sum_{i=1}^k c_i h(Z_i) - \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} Z_i\right), \end{aligned}$$

where the equality follows by definition of relative entropy, and the inequality follows by considering the independent coupling as one element in the set the supremum is taken over. By definitions and Proposition 3.1, we conclude

$$D_g(\mathbf{c} + t, (\mathbf{d}, t), \mathbf{B} \cup \{B_{m+1}\}) + \sum_{i=1}^k (c_i + t) \dim(E_i) \log(c_i + t) \leq M_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) \text{ for all } t \geq 0. \quad (86)$$

As for finiteness of  $M_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$ , this was already taken care of in the proof of Proposition 3.10. Indeed, using the the relaxation (64), we established finiteness of  $D_g(\mathbf{c}, \mathbf{d}, \mathbf{B})$  by showing

$$\sum_{i=1}^k c_i h(Z_i) - \sup_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} \tilde{Z}_i\right) \leq \sum_{i=1}^k c_i h(Z_i) - \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} Z_i\right) \leq C < +\infty$$

for independent gaussian  $(Z_i)_{1 \leq i \leq k}$  when (83) and (84) hold.

Now, fix arbitrary  $(Z_i)_{1 \leq i \leq k}$  with compact support and finite entropies, and for each  $n \geq 0$ , consider a coupling  $(Z_1^{(n)}, \dots, Z_k^{(n)}) \sim P_{\mathbf{Z}^{(n)}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})$  satisfying

$$\begin{aligned} & \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} Z_i^{(n)}\right) - n D(P_{\mathbf{Z}^{(n)}} \| P_{Z_1} \times \dots \times P_{Z_k}) \\ & \geq \sup_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \left( \sum_{j=1}^m d_j h\left(\sum_{i=1}^k B_{ij} \tilde{Z}_i\right) - n D(P_{\mathbf{Z}} \| P_{Z_1} \times \dots \times P_{Z_k}) \right) - \frac{1}{n}. \end{aligned}$$

Since the RHS is bounded from below by selecting the independent coupling and the entropies  $h\left(\sum_{i=1}^k B_{ij} Z_i^{(n)}\right)$  can be uniformly bounded from above in terms of the second moments of the marginals  $(Z_i)_{1 \leq i \leq k}$ , it is clear that we must have

$$\lim_{n \rightarrow \infty} n D(P_{\mathbf{Z}^{(n)}} \| P_{Z_1} \times \dots \times P_{Z_k}) = 0.$$

In particular, by weak upper semicontinuity of Shannon entropy under second moment constraint, we conclude

$$\limsup_{n \rightarrow \infty} \sup_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \left( \sum_{j=1}^m d_j h \left( \sum_{i=1}^k B_{ij} \tilde{Z}_i \right) - nD(P_{\mathbf{Z}} \| P_{Z_1} \times \dots \times P_{Z_k}) \right) \leq \sum_{j=1}^m d_j h \left( \sum_{i=1}^k B_{ij} Z_i \right).$$

Combining definitions with the above, Theorem 4.12, and Theorem 1.3, we have

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \left( D_g(\mathbf{c} + n, (\mathbf{d}, n), \mathbf{B} \cup \{B_{m+1}\}) + \sum_{i=1}^k (c_i + n) \dim(E_i) \log(c_i + n) \right) \\ & \geq \sum_{i=1}^k c_i h(Z_i) - \limsup_{n \rightarrow \infty} \sup_{P_{\mathbf{Z}} \in \Pi(P_{Z_1}, \dots, P_{Z_k})} \left( \sum_{j=1}^m d_j h \left( \sum_{i=1}^k B_{ij} \tilde{Z}_i \right) - nD(P_{\mathbf{Z}} \| P_{Z_1} \times \dots \times P_{Z_k}) \right) \\ & \geq \sum_{i=1}^k c_i h(Z_i) - \sum_{j=1}^m d_j h \left( \sum_{i=1}^k B_{ij} Z_i \right). \end{aligned}$$

Since the  $(Z_i)_{1 \leq i \leq k}$  were arbitrary, it follows from (86) and the subsequent remarks that

$$M(\mathbf{c}, \mathbf{d}, \mathbf{B}) \leq M_g(\mathbf{c}, \mathbf{d}, \mathbf{B}) < +\infty,$$

completing the proof. □

#### REFERENCES

- [1] V. Anantharam, V. Jog, and C. Nair. Unifying the Brascamp-Lieb inequality and the entropy power inequality. *arXiv preprint arXiv:1901.06619*, 2019.
- [2] S. Artstein-Avidan, B. Klartag and V. Milman. The Santaló point of a function, and a functional form of Santaló inequality. *Mathematika*, 51 (2005) 33–48.
- [3] K. Ball. Volumes of sections of cubes and related problems *Geometric aspects of functional analysis*, ed. by J. Lindenstrauss and V. D. Milman, Lecture Notes in Math. 1376, Springer, Heidelberg, 1989, pp. 251–260.
- [4] F. Barthe. On a reverse form of the Brascamp-Lieb inequality. *Inventiones mathematicae*, 134(2):335–361, 1998.
- [5] F. Barthe and P. Wolff. Positivity improvement and Gaussian kernels. *Comptes Rendus Mathematique*, 352(12):1017–1021, 2014.
- [6] F. Barthe and P. Wolff. Positive Gaussian kernels also have Gaussian minimizers. *arXiv preprint arXiv:1805.02455*, 2018.
- [7] S. Beigi and C. Nair. Equivalent characterization of reverse Brascamp-Lieb-type inequalities using information measures. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1038–1042. IEEE, 2016.
- [8] J. Bennett, A. Carbery, M. Christ, and T. Tao. The Brascamp-Lieb inequalities: finiteness, structure and extremals. *Geometric and Functional Analysis*, 17(5):1343–1415, 2008.
- [9] C. Borell. Diffusion equations and geometric inequalities. *Potential Analysis*, 12(1):49–71, 2000.
- [10] M. Boué and P. Dupuis. A variational representation for certain functionals of Brownian motion. *The Annals of Probability*, 26(4):1641–1659, 1998.
- [11] H. J. Brascamp, E. H. Lieb, and J. M. Luttinger. A general rearrangement inequality for multiple integrals. *Journal of functional analysis*, 17(2):227–237, 1974.
- [12] H. J. Brascamp and E. H. Lieb. Best constants in Young’s inequality, its converse, and its generalization to more than three functions. *Advances in Mathematics*, 20(2):151–173, 1976.
- [13] E. A. Carlen and D. Cordero-Erausquin. Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities. *Geometric and Functional Analysis*, 19(2):373–405, 2009.
- [14] E. A. Carlen, E. H. Lieb, and M. Loss. A sharp analog of Young’s inequality on  $S^N$  and related entropy inequalities. *The Journal of Geometric Analysis*, 14(3):487–520, 2004.
- [15] W.-K. Chen, N. Dafnis and G. Paouris. Improved Hölder and reverse Hölder inequalities for Gaussian random vectors. *Advances in Mathematics*, 280: 643–689, 2015.
- [16] D. Cordero-Erausquin and B. Maurey. Some extensions of the Prékopa–Leindler inequality using Borell’s stochastic approach. *Studia Mathematica*, 238(3): 201–233, 2017.
- [17] R. Gardner. The Brunn-Minkowski inequality. *Bulletin of the American Mathematical Society*, 39(3):355–405, 2002.

- [18] Y. Geng and C. Nair. The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages. *Information Theory, IEEE Transactions on*, 60(4):2087–2104, April 2014.
- [19] R. M. Gray. *Entropy and information theory*. Springer-Verlag, New York, 1990.
- [20] J. Lehec. A direct proof of the functional Santaló inequality. *Comptes Rendus Mathématique*, 347(1-2): 55–58, 2009.
- [21] J. Lehec. Representation formula for the entropy and functional inequalities. In *Annales de l’IHP Probabilités et statistiques*, volume 49, pages 885–899, 2013.
- [22] J. Lehec. Short probabilistic proof of the Brascamp-Lieb and Barthe theorems. *Canadian Mathematical Bulletin*, 57(3):585–597, 2014.
- [23] E. H. Lieb. Gaussian kernels have only Gaussian maximizers. *Inventiones mathematicae*, 102(1):179–208, 1990.
- [24] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú. Brascamp-Lieb inequality and its reverse: An information theoretic view. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1048–1052. IEEE, 2016.
- [25] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú. A forward-reverse Brascamp-Lieb inequality: Entropic duality and Gaussian optimality. *Entropy (special issue on information inequalities)*, 20(6):418, 2018.
- [26] B. Maurey. Some deviation inequalities. *Geometric and Functional Analysis*, 1(2): 188–197, 1991.
- [27] R. T. Rockafellar. Extension of Fenchel’s duality theorem for convex functions. *Duke mathematical journal*, 33(1):81–89, 1966.
- [28] S. I. Valdimarsson. **Optimisers for the Brascamp-Lieb inequality.** *Israel J. Math.*, 168:253–274, 2008.
- [29] C. Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

APPENDIX A. DEFINITION AND BASIC PROPERTIES OF SHANNON ENTROPY

This appendix provides an overview of the basic definitions and properties of Shannon entropy, with the goal of allowing an unfamiliar reader to follow the entropy computations in the body of this manuscript. In the interest of keeping things brief and self-contained, we focus on the case of random vectors in a Euclidean space with finite second moments, since this is sufficient for our purposes. The interested reader can find a general treatment in [19, Chapter 5].

Let  $E$  be a Euclidean space, and assume  $X$  is a random vector taking values in  $E$ , having density  $f_X : E \rightarrow [0, +\infty)$  with respect to Lebesgue measure. The Shannon entropy (referred to henceforth as simply *entropy*) associated to  $X$  is defined as

$$h(X) := - \int_E f_X(x) \log f_X(x) dx.$$

The entropy is said to exist if the integral is well-defined in the Lebesgue sense. Here, we adopt the convention that  $0 \cdot \log 0 = 0$ . The reader should note that the entropy  $h(X)$  is a functional of the density  $f_X$ , and is not a function of the realization of the random vector  $X$ . If  $X$  does not have density with respect to Lebesgue measure, we adopt the convention that  $h(X) = -\infty$ .

By a simple change of variables, we have the following elementary property of entropy:

**Proposition A.1** (Scaling property for Shannon entropy). *If the entropy of  $X$  exists and  $A : E \rightarrow E$  is an invertible linear transformation, then we have the scaling property*

$$h(AX) = \log |\det A| + h(X).$$

In this paper, we work exclusively with random vectors having finite second moments. In this context, we note that the inequality  $-\log(z) \geq 1 - z$  on  $z \in [0, +\infty)$  yields

$$0 \leq \int_E f_X(x) \log \frac{f_X(x)}{\phi(x)} dx,$$

where  $\phi(x) := (2\pi)^{-\dim(E)/2} e^{-|x|^2/2}$  is the standard gaussian density on  $E$ . We may conclude:

**Proposition A.2** (Entropy of random vectors with bounded second moments). *Let  $X$  have density on  $E$ . If  $\mathbb{E}|X|^2 < \infty$ , then the entropy  $h(X)$  exists, and is bounded from above as*

$$h(X) \leq \frac{1}{2} \log \left( (2\pi e)^{\dim(E)} \mathbb{E}|X|^2 \right).$$

If  $E_1, E_2$  are Euclidean spaces and  $(X, Y)$  is a pair of random vectors taking values in  $E_1 \times E_2$  with joint density  $f_{XY} : E_1 \times E_2 \rightarrow [0, +\infty)$ , the *joint entropy* of  $(X, Y)$  is defined as

$$h(X, Y) = - \iint_{E_1 \times E_2} f_{XY}(x, y) \log f_{XY}(x, y) dx dy.$$

Of course, this is consistent with the original definition of entropy, applied to the  $(E_1 \times E_2)$ -valued random vector  $(X, Y)$ . A simple property is as follows:

**Proposition A.3** (Subadditivity of entropy). *Let  $(X, Y)$  be a pair of random vectors taking values in  $E_1 \times E_2$  with joint density  $f_{XY} : E_1 \times E_2 \rightarrow [0, +\infty)$ . If  $(X, Y)$  have finite second moments, then the joint and marginal entropies exist and satisfy*

$$h(X, Y) \leq h(X) + h(Y).$$

*This is met with equality if  $X, Y$  are independent.*

*Proof.* Since all entropies exist and are bounded from above, we may assume  $h(X, Y)$  is finite, else the claim is trivial. Similar to before, we note the following inequality:

$$0 \leq \iint_{E_1 \times E_2} f_{XY}(x, y) \log \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} dx dy.$$

Using existence and finiteness of the joint entropy, we apply linearity of the integral to conclude

$$- \iint_{E_1 \times E_2} f_{XY}(x, y) \log (f_X(x)f_Y(y)) dx dy \geq h(X, Y). \quad (87)$$

Now, let us decompose the density

$$f_{XY}(x, y) = f_X(x)f_{Y|X}(y|x), \quad x \in E_1, y \in E_2,$$

where, for  $x \in E_1$ , the function  $f_{Y|X}(\cdot|x)$  denotes the density of a conditional regular probability  $P_{Y|X=x}$  on  $E_2$ , and coincides with the ratio  $f_{XY}(x, y)/f_X(x)$  for almost every  $x$  in the support of  $f_X$ . Similar to how we argued entropy was bounded from above under a second moment constraint, we may show that the integral

$$- \iint_{E_1 \times E_2} f_{XY}(x, y) \log (f_X(x)) dx dy$$

exists and is bounded from above. The same conclusion holds with  $f_X(x)$  replaced by  $f_Y(y)$ . From this and (87), we conclude using the assumed finiteness of  $h(X, Y)$  that the function  $(x, y) \mapsto f_{XY}(x, y) \log (f_X(x))$  is integrable, and is equal to  $-h(X)$  by the Fubini-Tonelli theorem. Similar for  $h(Y)$ , so linearity of the integral applied to (87) proves the claim.  $\square$

Finally, we note the following consequence of lower semicontinuity of relative entropy, which follows from the Donsker-Varadhan variational formula. See, e.g., [25, Lemma A2] for details.

**Proposition A.4** (Upper semicontinuity of entropy under second moment constraint). *Let  $(X_n)_{n \geq 1}$  be a sequence of random vectors on a Euclidean space  $E$ , such that  $X_n \rightarrow X$  weakly. If  $\sup_{n \geq 1} \mathbb{E}|X_n|^2 < +\infty$ , then*

$$\limsup_{n \rightarrow \infty} h(X_n) \leq h(X).$$

UNIVERSITY OF CALIFORNIA, BERKELEY, DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCES, BERKELEY, CA 94720, USA

*Email address:* courtade@berkeley.edu

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY (IDSS), CAMBRIDGE, MA 02139, USA

*Email address:* jingbo@mit.edu