

Sharp Maximum-Entropy Comparisons

Efe Aras and Thomas A. Courtade

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

Abstract—We establish a family of sharp entropy inequalities with Gaussian extremizers. These inequalities hold for certain dependent random variables, namely entropy-maximizing couplings subject to information constraints. Several well-known results, such as the Zamir–Feder and Brunn–Minkowski inequalities, follow as special cases.

I. INTRODUCTION AND MAIN RESULTS

Let X be a random vector on $\mathbb{R}^n$, having density f with respect to Lebesgue measure. We define the Shannon entropy

$$h(X) = - \int_{\mathbb{R}^n} f(x) \log f(x) dx,$$

where $\log$ denotes the natural logarithm. If X has finite second moments, then the entropy of X always exists in the Lebesgue sense, and is bounded from above. If X does not admit a density, we adopt the convention that $h(X) = -\infty$.

Inequalities relating entropies of random variables have played a foundational role in information theory and its applications, dating back to Shannon’s seminal work. In recent decades, entropy inequalities have become a subject of independent investigation due in part to their close relationship with functional and geometric inequalities (see, e.g., [1], [2] and references therein). Occupying a special place in this field is the Shannon–Stam entropy power inequality (EPI) [3], which is responsible for impossibility results in information theory (see, e.g., applications in [4]) and statistics (e.g., [5]), and captures Gaussian concentration phenomena through its implication of the Gaussian log-Sobolev inequality (see, e.g., [6]). The EPI can be equivalently stated as the following comparison: If X_1, X_2 are independent random variables with finite entropies and second moments, and $\tilde{X}_1, \tilde{X}_2$ are independent Gaussian random variables with $h(\tilde{X}_i) = h(X_i)$, then

$$h(\tilde{X}_1 + \tilde{X}_2) \leq h(X_1 + X_2). \quad (1)$$

The present paper establishes a general class of such comparisons, unifying and extending the known landscape.

II. MAIN RESULTS

To state our main results, we start with some notation. For a collection of random vectors $(X_i)_{i=1}^k$ in $\mathbb{R}^n$, let $\Pi(X_1, \dots, X_k)$ denote the set of couplings of $X_1, \dots, X_k$. Although $\Pi(X_1, \dots, X_k)$ is technically a collection of probability measures on $\mathbb{R}^{n \times k}$, we write $X \in \Pi(X_1, \dots, X_k)$ to denote a $(n \times k)$ -dimensional random vector $X =$

$(X'_1, X'_2, \dots, X'_k)$ such that $X'_i = X_i$ in distribution for each $1 \leq i \leq k$ (i.e., the law of X is an element of $\Pi(X_1, \dots, X_k)$).

For jointly distributed random vectors $(X_i)_{i=1}^k$ in $\mathbb{R}^n$ and a subset $S \subseteq [k] := \{1, \dots, k\}$, define the “multi-information”

$$I(X_S) := D\left(P_{X_S} \parallel \prod_{i \in S} P_{X_i}\right),$$

where P_{X_S} denotes the joint law of $X_S := (X_i)_{i \in S}$. Note that $I(X_S) = 0$ implies $(X_i)_{i \in S}$ are independent. More generally, specifying that $I(X_S) \leq \delta$ ensures that P_{X_S} is δ -close to the independent coupling of the $(X_i)_{i \in S}$ in relative entropy. For a function $\nu : 2^{[k]} \rightarrow [0, +\infty]$, let $\Pi(X_1, \dots, X_k; \nu) \subseteq \Pi(X_1, \dots, X_k)$ denote the set of couplings of $(X_i)_{i=1}^k$ that satisfy

$$I(X_S) \leq \nu(S), \quad \forall S \subseteq [k].$$

For convenience, we adopt the convention $I(X_S) = 0$ for $S = \emptyset$ so that we do not have to persistently exclude the degenerate case $S = \emptyset$. Thus, for example, if $\nu(S) = 0$ for all $S \subseteq [k]$, then $\Pi(X_1, \dots, X_k; \nu)$ is a singleton set, whose only element is the product measure $\prod_{i=1}^k P_{X_i}$. On the other hand, if $\nu(S) = +\infty$ for all $S \subseteq [k]$, then $\Pi(X_1, \dots, X_k; \nu)$ is equal to the set of all couplings $\Pi(X_1, \dots, X_k)$. As before, we write $X \in \Pi(X_1, \dots, X_k; \nu)$ to denote a random vector $X = (X_1, \dots, X_k)$ with law in $\Pi(X_1, \dots, X_k; \nu)$.

Our first main result is a generalization of the Zamir–Feder inequality [7] to constrained maximum-entropy couplings.

Theorem 1. *Let $(\alpha_j)_{j=1}^m \subset (0, +\infty)$ and $(Q_j : \mathbb{R}^k \rightarrow \mathbb{R}^{n_j})_{j=1}^m$ be surjective linear maps. Let $(X_i)_{i=1}^k$ be real-valued random variables with finite entropies and second moments, and let $(\tilde{X}_i)_{i=1}^k$ be Gaussian random variables with $h(\tilde{X}_i) = h(X_i)$. For any $\nu : 2^{[k]} \rightarrow [0, +\infty]$, it holds that*

$$\begin{aligned} & \sup_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k; \nu)} \sum_{j=1}^m \alpha_j h(Q_j \tilde{X}) \\ & \leq \sup_{X \in \Pi(X_1, \dots, X_k; \nu)} \sum_{j=1}^m \alpha_j h(Q_j X). \end{aligned} \quad (2)$$

Remark 2. *If $\nu(S) = 0$ for all $S \subseteq [k]$, then each set of couplings is a singleton containing only the independent coupling, thus recovering the Zamir–Feder inequality.*

Remark 3. *By the max-entropy property of Gaussians, it suffices to consider jointly Gaussian couplings in (2). For $m = 1$, (2) has the following interpretation in terms of I -projections [8]: Lebesgue measure is closer to its I -projection onto $Q_1 \Pi(X_1, \dots, X_k; \nu)$ than to that onto*

Email: {efearas96, courtade}@berkeley.edu. Research supported in part by NSF grants CCF-1704967, CCF-1750430, and CCF-0939370.

$Q_{1\#}\Pi(\tilde{X}_1, \dots, \tilde{X}_k; \nu)$, where $\#$ denotes pushforward. This may have applications to bounding large-deviation probabilities in Schrödinger-type problems (cf. [9]).

We may extend Theorem 1 to a setting where the X_i 's are random vectors of the same dimension when the linear maps have a certain product structure (similar structure appears in [10, Theorems 1.1 and 1.4] for order-infinity Rényi entropy inequalities). Ultimately, this allows recovery of results such as the multi-dimensional EPI and the Brunn–Minkowski inequality. Toward this end, let $A : \mathbb{R}^k \rightarrow \mathbb{R}^m$ be a linear map, expressed as a matrix with real-valued entries $[A]_{ij} = a_{ij}$, and let I_n denote the $n \times n$ identity matrix. Recall that the Kronecker product $A \otimes I_n$ is a linear map from $\mathbb{R}^{n \times k}$ to $\mathbb{R}^{n \times m}$ defined by

$$(A \otimes I_n)(x_1, \dots, x_k) = \begin{bmatrix} \sum_{i=1}^k a_{1i} x_i \\ \sum_{i=1}^k a_{2i} x_i \\ \vdots \\ \sum_{i=1}^k a_{mi} x_i \end{bmatrix}, \quad x_i \in \mathbb{R}^n, 1 \leq i \leq k.$$

Theorem 4. Let $(\alpha_j)_{j=1}^m \subset (0, +\infty)$ and $(Q_j : \mathbb{R}^k \rightarrow \mathbb{R}^{n_j})_{j=1}^m$ be surjective linear maps. Let $(X_i)_{i=1}^k$ be random vectors on $\mathbb{R}^n$ with finite entropies and second moments, and let $\tilde{X}_i \sim N(0, \sigma_i^2 I_n)$ be Gaussian random vectors with variance parameters chosen so that $h(\tilde{X}_i) = h(X_i)$ for each $i = 1, \dots, k$. For any $\nu : 2^{[k]} \rightarrow [0, +\infty]$, it holds that

$$\begin{aligned} & \sup_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k; \nu)} \sum_{j=1}^m \alpha_j h((Q_j \otimes I_n) \tilde{X}) \\ & \leq \sup_{X \in \Pi(X_1, \dots, X_k; \nu)} \sum_{j=1}^m \alpha_j h((Q_j \otimes I_n) X). \end{aligned}$$

To illustrate how the above results imply some of those that are known and considered classical, we establish the following EPI for information-constrained max-entropy couplings. To state it, recall that we define the entropy-power of an n -dimensional random vector as $N(X) := e^{2h(X)/n}$.

Corollary 5. Let X and Z be random vectors in $\mathbb{R}^n$ with finite second moments. For any $\zeta \in [0, +\infty]$, it holds that

$$\begin{aligned} & N(X) + N(Z) + 2\sqrt{(1 - e^{-2\zeta/n})N(X)N(Z)} \\ & \leq \sup_{(X,Z) \in \Pi(X,Z); I(X,Z) \leq \zeta} N(X+Z). \end{aligned} \quad (3)$$

Equality holds for Gaussian X, Z with proportional covariances.

Taking $\zeta = 0$ in (3) recovers the Shannon–Stam EPI, since the only admissible coupling is the independent one. Hence, (3) may be regarded as an extension of the EPI for certain dependent random variables with a sharp correction term. We remark that Takano [11] and Johnson [12] have established that the EPI holds for dependent random variables which have positively correlated scores. Given the different hypotheses, their results are not directly comparable to ours.

Toward the other extreme, taking $\zeta = +\infty$ allows for unconstrained optimization over couplings, and completing the square gives the inequality

$$e^{h(X)/n} + e^{h(Z)/n} \leq \sup_{(X,Z) \in \Pi(X,Z)} e^{h(X+Z)/n},$$

where we emphasize the change in exponent from 2 to 1. If X, Z are uniform on compact subsets $K, L \subset \mathbb{R}^n$, respectively, we obtain the celebrated Brunn–Minkowski inequality

$$|K|^{1/n} + |L|^{1/n} \leq \sup_{(X,Z) \in \Pi(X,Z)} N(X+Z)^{1/2} \leq |K+L|^{1/n},$$

where $K+L$ denotes the Minkowski sum of K and L , and $|\cdot|$ denotes the n -dimensional Lebesgue volume. Here, the second inequality follows since $X+Z$ is supported on the Minkowski sum $K+L$, and hence the entropy is upper bounded by that of the uniform distribution on that set. It is known that equality is attained when K, L are positive homothetic convex bodies, which highlights that the stated conditions for equality in Corollary 5 are sufficient, but not always necessary. Indeed, for X, Z equal in distribution, Cover and Zhang [13] showed

$$h(2X) \leq \sup_{(X,Z) \in \Pi(X,Z)} h(X+Z),$$

with equality if and only if X is log-concave. This implies that for X, Z identically distributed and $\zeta = +\infty$, equality is achieved in (3) if and only if X is log-concave.

To lend some historical perspective, we note that it has long been observed that there is a striking similarity between the Brunn–Minkowski inequality and the EPI (see, e.g., [14] and citing works). It is well-known that each can be obtained from convolution inequalities involving Rényi entropies (e.g., the sharp Young inequality [15]–[17], or rearrangement inequalities analogous to (1) [18]), when the orders of the involved Rényi entropies are taken to the limit 0 or 1, respectively. Quantitatively linking both inequalities using only Shannon entropies has proved elusive, and has been somewhat of a looming question. In this sense, Corollary 5 provides an answer. Again, the Brunn–Minkowski inequality and EPI are obtained as logical endpoints, but this time the family of inequalities involves only Shannon entropies instead of Rényi entropies of varying orders. In contrast to derivations involving Rényi entropies where summands are always independent (corresponding to the convolution of densities), the key idea here is to allow dependence between the random summands, subject to a mutual information constraint.

III. PROOFS

Before proving the main results, let us first set some notation and then explain what is known. To this end, for a k -tuple of positive reals $(a_1, a_2, \dots, a_k)$, let $\Pi(a_1, a_2, \dots, a_k)$ denote the set of positive semidefinite $k \times k$ matrices A with diagonal entry $[A]_{ii} = a_i$ for each $i = 1, \dots, k$. This is consistent with the notation of Π for couplings; indeed, A may be thought of as the covariance of a k -dimensional Gaussian vector that couples Gaussian random variables with individual variances $a_1, \dots, a_k$. Also, let $\text{diag}(a_1, \dots, a_k)$ denote the diagonal

matrix A with diagonal entry $[A]_{ii} = a_i$ for each $i = 1, \dots, k$. We denote the set of real $n \times n$ positive definite matrices by $S^+(\mathbb{R}^n)$, and let $\langle \cdot, \cdot \rangle$ denote the trace inner product.

Now, to explain what is already known in the context of our results, let $\mathbf{Q} := (Q_j : \prod_{i=1}^k \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{n_j})_{j=1}^m$ be a collection of surjective linear maps, and let non-negative reals $\mathbf{c} := (c_i)_{1 \leq i \leq k} \subset (0, +\infty)$ and $\mathbf{d} := (d_j)_{1 \leq j \leq m} \subset (0, +\infty)$ satisfy the dimension condition

$$\sum_{i=1}^k c_i m_i = \sum_{j=1}^m d_j n_j.$$

Define $D_g(\mathbf{Q}, \mathbf{c}, \mathbf{d})$ to be the smallest constant $D \in \mathbb{R} \cup \{+\infty\}$ such that

$$\sum_{i=1}^k c_i h(X_i) \leq \sup_{X \in \Pi(X_1, \dots, X_k)} \sum_{j=1}^m d_j h(Q_j X) + D$$

for any choice of Gaussian random vectors X_i on $\mathbb{R}^{m_i}$, $1 \leq i \leq k$. With this notation set, [2, Theorem 1.14] and [2, Theorem 4.2] together imply the following entropic dual of the “forward-reverse” Brascamp–Lieb inequalities, together with a characterization of the structure of the extremizers.

Theorem 6. *Let the above notation prevail. For any random vectors X_i on $\mathbb{R}^{m_i}$, $1 \leq i \leq k$, with finite entropies and finite second moments, it holds that*

$$\sum_{i=1}^k c_i h(X_i) \leq \sup_{X \in \Pi(X_1, \dots, X_k)} \sum_{j=1}^m d_j h(Q_j X) + D_g(\mathbf{Q}, \mathbf{c}, \mathbf{d}). \quad (4)$$

Moreover, if $K_i \in S^+(\mathbb{R}^{m_i})$ and $K \in \Pi(K_1, \dots, K_k)$ satisfy

$$\sum_{j=1}^m d_j Q_j^T (Q_j K Q_j^T)^{-1} Q_j \leq \text{diag}(c_1 K_1^{-1}, \dots, c_k K_k^{-1}),$$

then $D_g(\mathbf{Q}, \mathbf{c}, \mathbf{d})$ is finite, and equality is achieved for $X_i \sim N(0, K_i)$, $1 \leq i \leq k$.

Remark 7. *Inequalities of the form (4) were considered in [19] under an independence assumption. The results therein can be realized as a special case of Theorem 6 (or the earlier work [20]). See [2, Section 4.4] for details.*

We note that (4) constitutes a family of inequalities with Gaussian extremizers, similar to our main result. However, the key point to be made is that Theorem 1 provides a precise comparison between certain entropies evaluated for (marginally) specified random variables, and those for Gaussian random variables with the same (marginal) entropies. This is precisely in the same spirit as the Shannon–Stam inequality and the Zamir–Feder inequality. In contrast, (4) does not directly yield such a comparison, because the marginal entropies of the Gaussian extremizers are a function of the triple $(\mathbf{Q}, \mathbf{c}, \mathbf{d})$, and are therefore not determined by the entropies of the X_i ’s we select to appear in (4). To achieve the desired comparison, we must turn Theorem 6 around in the following sense: we fix $\mathbf{Q}, \mathbf{d}$ and $(K_i)_{1 \leq i \leq k}$, and then show that there is a choice

of $\mathbf{c}$ for which the Gaussians $X_i \sim N(0, K_i)$, $1 \leq i \leq k$ are extremal in (4). In order to accomplish this, we will at some point require that the K_i ’s are positive reals, and not matrices (n.b. this implies that each $Q_j : \mathbb{R}^k \rightarrow \mathbb{R}^{n_j}$ in the definition of $\mathbf{Q}$). This is the reason that Theorem 1 is stated in terms of random variables, despite Theorem 6 applying to random vectors.

The crux of the above argument is contained in the following technical lemma.

Lemma 8. *Fix $\mathbf{Q}, \mathbf{d}$ and $(K_i)_{1 \leq i \leq k} \subset (0, +\infty)$. Assume for each natural basis vector $e_i \in \mathbb{R}^k$, there is $j \in [m]$ (possibly depending on i) such that $Q_j e_i \neq 0$. There exists $\mathbf{c} := (c_i)_{1 \leq i \leq k} \subset (0, +\infty)$ satisfying*

$$\sum_{i=1}^k c_i = \sum_{j=1}^m d_j n_j, \quad (5)$$

and $K \in \Pi(K_1, \dots, K_k)$ satisfying

$$\sum_{j=1}^m d_j Q_j^T (Q_j K Q_j^T)^{-1} Q_j \leq \text{diag}(c_1 K_1^{-1}, \dots, c_k K_k^{-1}).$$

Proof. To start, recall the Legendre duality for log det, stated as follows: For $A \in S^+(\mathbb{R}^n)$, we have

$$n + \log \det A = \min_{B \in S^+(\mathbb{R}^n)} (\langle A, B \rangle - \log \det B), \quad (6)$$

where the minimum is uniquely achieved by $B = A^{-1}$. Now, note that for any $K \in \Pi(K_1, \dots, K_k)$ and $U_j \in S^+(\mathbb{R}^{n_j})$, $1 \leq j \leq m$ and $V_i \in (0, +\infty)$, $1 \leq i \leq k$ satisfying the operator inequality

$$\sum_{j=1}^m d_j Q_j^T U_j Q_j \leq \text{diag}(V_1, \dots, V_k), \quad (7)$$

we have

$$\begin{aligned} & \sum_{j=1}^m d_j \log \det(Q_j K Q_j^T) + \sum_{j=1}^m d_j n_j \\ & \leq \sum_{j=1}^m d_j \langle (Q_j K Q_j^T), U_j \rangle - \sum_{j=1}^m d_j \log \det(U_j) \end{aligned} \quad (8)$$

$$\leq \sum_{i=1}^k \langle K_i, V_i \rangle - \sum_{j=1}^m d_j \log \det(U_j). \quad (9)$$

The first inequality is (6), and the second inequality follows from (7). Indeed, if we let $X \sim N(0, K)$, we see that (7) gives

$$\begin{aligned} \sum_{j=1}^m d_j \langle (Q_j K Q_j^T), U_j \rangle &= \mathbb{E} \left\langle \sum_{j=1}^m d_j Q_j^T U_j Q_j X, X \right\rangle \\ &\leq \mathbb{E} \langle \text{diag}(V_1, \dots, V_k) X, X \rangle \\ &= \sum_{i=1}^k \langle K_i, V_i \rangle. \end{aligned} \quad (10)$$

Now, [2, Theorem 2.8] asserts the min-max principle

$$\max_{K \in \Pi(K_1, \dots, K_k)} \sum_{j=1}^m d_j \log \det(Q_j K Q_j^T) + \sum_{j=1}^m d_j n_j \quad (11)$$

$$= \inf_{(V_i)_{i=1}^k, (U_j)_{j=1}^m} \left(\sum_{i=1}^k \langle V_i, K_i \rangle - \sum_{j=1}^m d_j \log \det U_j \right). \quad (12)$$

where the infimum is over $U_j \in S^+(\mathbb{R}^{n_j})$, $1 \leq j \leq m$ and $V_i \in (0, +\infty)$, $1 \leq i \leq k$ satisfying (7). The fact that the maximum in (11) is achieved (i.e., it is not a supremum) is part of the quoted result.

Letting K^* denote the maximizer in (11), we must have $\det(Q_j K^* Q_j^T) > 0$ for each $1 \leq j \leq m$, otherwise the maximum would be equal to $-\infty$, which is a contradiction (indeed, consider $K = \text{diag}(K_1, \dots, K_k)$ and use surjectivity of the Q_j 's to conclude that (11) is finite). Hence, we deduce from (6) and the string of inequalities (8)-(9) that

$$\sum_{j=1}^m d_j \log \det(Q_j K^* Q_j^T) + \sum_{j=1}^m d_j n_j = \inf_{(V_i)_{i=1}^k} \left(\sum_{i=1}^k \langle V_i, K_i \rangle - \sum_{j=1}^m d_j \log \det U_j^* \right),$$

where $U_j^* = (Q_j K^* Q_j^T)^{-1}$, and the infimum is over $(V_i)_{i=1}^k$ satisfying the operator inequality

$$\sum_{j=1}^m d_j Q_j^T U_j^* Q_j \leq \text{diag}(V_1, \dots, V_k). \quad (13)$$

By compactness, it is evident that this infimum will be achieved by some $(V_i^*)_{i=1}^k \subset [0, +\infty)$. Now, we define $c_i := K_i V_i^*$, so $V_i^* = c_i K_i^{-1}$ since $K_i > 0$ is scalar. Since $(U_j^*)_{j=1}^m$ and $(V_i^*)_{i=1}^k$ must satisfy (10) with equality, we have

$$\sum_{i=1}^k c_i = \sum_{i=1}^k K_i V_i^* = \sum_{j=1}^m d_j \langle (Q_j K Q_j^T), U_j^* \rangle = \sum_{j=1}^m d_j n_j.$$

Note that if $c_i = 0$, then $V_i^* = 0$, which means that $Q_j e_i = 0$ for all $j \in [m]$ by (13) and the previous observation that $U_j^* = (Q_j K^* Q_j^T)^{-1}$ is positive definite for each j . This contradicts our assumption, so we conclude each $c_i > 0$ as desired. $\square$

The combination of Lemma 8 and Theorem 6 allow us to turn Theorem 6 on its head, as was our goal. Namely, we fix arbitrary $\mathbf{Q}, \mathbf{d}$ and positive reals $(K_i)_{1 \leq i \leq k}$, and then show that there is a choice of $\mathbf{c}$ for which the Gaussian random variables $X_i \sim N(0, K_i)$, $1 \leq i \leq k$ are extremizers of the inequality (4). In other words, we specify the extremizers first, and then construct an inequality of the form (4) for which they are extremal. This is the key idea needed to prove the following preliminary version of Theorem 1, where the couplings are unconstrained.

Theorem 9. *Let $(X_i)_{i=1}^k$ be real-valued random variables with finite entropies and finite second moments. Let $(d_j)_{j=1}^m$ be positive real numbers, and let $(Q_j : \mathbb{R}^k \rightarrow \mathbb{R}^{n_j})_{j=1}^m$*

be surjective linear maps. If $(\tilde{X}_i)_{i=1}^k$ are Gaussian random variables with $h(\tilde{X}_i) = h(X_i)$ for each $1 \leq i \leq k$, then

$$\sup_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k)} \sum_{j=1}^m d_j h(Q_j \tilde{X}) \leq \sup_{X \in \Pi(X_1, \dots, X_k)} \sum_{j=1}^m d_j h(Q_j X).$$

Proof. Let K_i denote the variance of $\tilde{X}_i$. We can assume without loss of generality that for each natural basis vector $e_i \in \mathbb{R}^k$, there is $j \in [m]$ such that $Q_j e_i \neq 0$. Indeed, if this is not the case, then the desired inequality does not involve X_i or $\tilde{X}_i$ at all, so that this coordinate can be disregarded. Thus, by Lemma 8, there exists $K \in \Pi(K_1, \dots, K_k)$ and $\mathbf{c} := (c_i)_{1 \leq i \leq k} \subset (0, +\infty)$ satisfying (5) and

$$\sum_{j=1}^m d_j Q_j^T (Q_j K Q_j^T)^{-1} Q_j \leq \text{diag}(c_1 K_1^{-1}, \dots, c_k K_k^{-1}).$$

By Theorem 6, we have

$$\begin{aligned} & \sup_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k)} \sum_{j=1}^m d_j h(Q_j \tilde{X}) + D_g(\mathbf{Q}, \mathbf{c}, \mathbf{d}) \\ &= \sum_{i=1}^k c_i h(\tilde{X}_i) \\ &= \sum_{i=1}^k c_i h(X_i) \leq \sup_{X \in \Pi(X_1, \dots, X_k)} \sum_{j=1}^m d_j h(Q_j X) + D_g(\mathbf{Q}, \mathbf{c}, \mathbf{d}). \end{aligned}$$

Since $D_g(\mathbf{Q}, \mathbf{c}, \mathbf{d})$ is finite (by Theorem 6), we may subtract it from both sides to complete the proof. $\square$

Finally, we are in a position to prove Theorem 1. The strategy will be to convert the constrained optimization problem into an unconstrained one by the method of Lagrange multipliers, and then deduce the desired result as a corollary of Theorem 9.

Proof of Theorem 1. For real-valued random variables $(Z_i)_{i=1}^k$ having finite entropies and second moments, define the functional

$$F(\lambda, P_Z) := \left(- \sum_{j=1}^m \alpha_j h(Q_j Z) + \sum_{S: \nu(S) < +\infty} \lambda(S) (I(Z_S) - \nu(S)) \right)$$

for arguments $\lambda : 2^{[k]} \rightarrow \mathbb{R}$ and $P_Z \in \Pi(Z_1, \dots, Z_k)$. It follows by Sion's minimax theorem that

$$\inf_{Z \in \Pi(Z_1, \dots, Z_k)} \sup_{\lambda \geq 0} F(\lambda, P_Z) = \sup_{\lambda \geq 0} \inf_{Z \in \Pi(Z_1, \dots, Z_k)} F(\lambda, P_Z). \quad (14)$$

Indeed, for any fixed $P_Z \in \Pi(Z_1, \dots, Z_k)$, the function $\lambda \mapsto F(\lambda, P_Z)$ is linear in λ . On the other hand, $\Pi(Z_1, \dots, Z_k)$ is a convex subset of probability measures on $\mathbb{R}^k$ that is closed with respect to the weak topology; it is also tight due to the assumption of finite second moments. Hence, by Prokhorov's theorem, it is compact with respect to the weak topology. For fixed $\lambda \geq 0$, the functional $P_Z \mapsto F(\lambda, P_Z)$ is convex on $\Pi(Z_1, \dots, Z_k)$ by the usual convexity properties of

entropy. Moreover, it is weakly lower semicontinuous. Indeed, if a sequence $(P_n)_{n \geq 1} \subset \Pi(Z_1, \dots, Z_k)$ has weak limit $P^* \in \Pi(Z_1, \dots, Z_k)$, then we also have convergence of the moments $\mathbb{E}|Q_j Z^{(n)}|^2 \rightarrow \mathbb{E}|Q_j Z^*|^2$, where $Z^{(n)} \sim P_n$ and $Z^* \sim P^*$. This follows since $Q_j Z^{(n)} \rightarrow Q_j Z^*$ in distribution, and $(|Q_j Z^{(n)}|^2)_{n \geq 1}$ can be verified to be uniformly integrable since each $P_n \in \Pi(Z_1, \dots, Z_k)$, and the Z_i 's have finite second moments. The terms $-h(Q_j Z^{(n)})$ can be written as a relative entropy with respect to a Gaussian having the same covariance as that of $Q_j Z^*$, plus an affine function of the difference $\mathbb{E}|Q_j Z^{(n)}|^2 - \mathbb{E}|Q_j Z^*|^2$. Thus, the claimed lower semicontinuity follows immediately due to the previously established convergence of second moments together with lower semicontinuity of relative entropy. Hence, all hypotheses of Sion's minimax theorem are satisfied, and (14) holds.

Now, let $(X_i)_{i=1}^k$ and $(\tilde{X}_i)_{i=1}^k$ be as defined in the statement to prove. Using the above observation together with Theorem 9, we have

$$\begin{aligned} & \inf_{X \in \Pi(X_1, \dots, X_k; \nu)} - \sum_{j=1}^m \alpha_j h(Q_j X) \\ &= \inf_{X \in \Pi(X_1, \dots, X_k)} \sup_{\lambda \geq 0} F(\lambda, P_X) \\ &= \sup_{\lambda \geq 0} \inf_{X \in \Pi(X_1, \dots, X_k)} F(\lambda, P_X) \\ &\leq \sup_{\lambda \geq 0} \inf_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k)} F(\lambda, P_{\tilde{X}}) \\ &= \inf_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k)} \sup_{\lambda \geq 0} F(\lambda, P_{\tilde{X}}) \\ &= \inf_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k; \nu)} - \sum_{j=1}^m \alpha_j h(Q_j \tilde{X}). \end{aligned} \quad (15)$$

To see why inequality (15) follows from Theorem 9, write

$$\begin{aligned} & \sum_{S: \nu(S) < +\infty} \lambda(S) I(X_S) \\ &= \sum_{S: \nu(S) < +\infty} \lambda(S) \sum_{i \in S} h(X_i) - \sum_{S: \nu(S) < +\infty} \lambda(S) h(\pi_S X), \end{aligned}$$

where π_S is the projection $\pi_S : (x_1, \dots, x_k) \mapsto (x_i)_{i \in S}$ (a surjective linear map). As a result, Theorem 9 implies

$$\begin{aligned} & \sup_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k)} \left(\sum_{j=1}^m \alpha_j h(Q_j \tilde{X}) + \sum_{S: \nu(S) < +\infty} \lambda(S) h(\pi_S \tilde{X}) \right) \\ &\leq \sup_{X \in \Pi(X_1, \dots, X_k)} \left(\sum_{j=1}^m \alpha_j h(Q_j X) + \sum_{S: \nu(S) < +\infty} \lambda(S) h(\pi_S X) \right). \end{aligned}$$

Taken together with $h(\tilde{X}_i) = h(X_i)$, we obtain (15). Thus, the proof is complete. $\square$

Proof of Theorem 4. Let $\mathbf{Q}$ denote the collection of linear maps $(Q_j)_{1 \leq j \leq m}$. For any choice of positive reals $\mathbf{d} = (d_j)_{1 \leq j \leq m}$, we can repeat the argument of Theorem 9 to conclude that there is a valid choice of $\mathbf{c} = (c_i)_{1 \leq i \leq k}$ for which

the Gaussian random variables $\tilde{Z}_1 \sim N(0, \sigma_1^2), \dots, \tilde{Z}_k \sim N(0, \sigma_k^2)$ saturate the inequality

$$\sum_{i=1}^k c_i h(Z_i) \leq \sup_{Z \in \Pi(Z_1, \dots, Z_k)} \sum_{j=1}^m d_j h(Q_j Z) + D_g(\mathbf{Q}, \mathbf{c}, \mathbf{d}), \quad (16)$$

holding (by definition) for all collections of random variables $(Z_i)_{1 \leq i \leq k}$ with finite entropies and second moments. Note that a simple consequence of subadditivity of entropy is that

$$\begin{aligned} & \sup_{\tilde{X} \in \Pi(\tilde{X}_1, \dots, \tilde{X}_k)} \sum_{j=1}^m \alpha_j h((Q_j \otimes I_n) \tilde{X}) \\ &\leq \sup_{\tilde{Z} \in \Pi(\tilde{Z}_1, \dots, \tilde{Z}_k)} n \sum_{j=1}^m \alpha_j h(Q_j \tilde{Z}), \end{aligned} \quad (17)$$

where $\tilde{X}_1 \sim N(0, \sigma_1^2 I_n), \dots, \tilde{X}_k \sim N(0, \sigma_k^2 I_n)$ as in the statement of the theorem. However, we have equality in (17) by considering only the couplings in $\Pi(\tilde{X}_1, \dots, \tilde{X}_k)$ that are the n -fold products of the optimal coupling in the RHS.

By the tensorization properties enjoyed by D_g (see [2], [20]), it holds that $D_g(\mathbf{Q}^n, \mathbf{c}, \mathbf{d}) = n D_g(\mathbf{Q}, \mathbf{c}, \mathbf{d})$, where $\mathbf{Q}^n := (Q_j \otimes I_n)_{1 \leq j \leq m}$. Hence, we conclude that $(\tilde{X}_i)_{i=1}^k$ saturate the inequality

$$\begin{aligned} & \sum_{i=1}^k c_i h(X_i) \\ &\leq \sup_{X \in \Pi(X_1, \dots, X_k)} \sum_{j=1}^m d_j h((Q_j \otimes I_n) X) + D_g(\mathbf{Q}^n, \mathbf{c}, \mathbf{d}), \end{aligned}$$

which (by definition) holds for all collections of random vectors $(X_i)_{i=1}^k$ in $\mathbb{R}^n$ with finite entropies and second moments. The rest proceeds exactly as in the one-dimensional setting. $\square$

Finally, we are in a position to prove our last result.

Proof of Corollary 5. By Theorem 4,

$$\sup_{(\tilde{X}, \tilde{Z}) \in \Pi(\tilde{X}, \tilde{Z}; \nu)} N(\tilde{X} + \tilde{Z}) \leq \sup_{(X, Z) \in \Pi(X, Z; \nu)} N(X + Z), \quad (18)$$

where $\tilde{X} \sim N(0, \sigma_1^2 I_n)$ and $\tilde{Z} \sim N(0, \sigma_2^2 I_n)$ are such that $h(\tilde{X}) = h(X)$ and $h(\tilde{Z}) = h(Z)$. Now, straightforward computations reveal

$$\begin{aligned} & N(\tilde{X}) + N(\tilde{Z}) + 2\sqrt{(1 - e^{-2\zeta/n})N(\tilde{X})N(\tilde{Z})} \\ &= \sup_{(\tilde{X}, \tilde{Z}) \in \Pi(\tilde{X}, \tilde{Z}; I(\tilde{X}; \tilde{Z}) \leq \zeta} N(\tilde{X} + \tilde{Z}). \end{aligned}$$

Combining with (18) completes the proof since $N(\tilde{X}) = N(X)$ and $N(\tilde{Z}) = N(Z)$. $\square$

REFERENCES

- [1] Mokshay Madiman, James Melbourne, and Peng Xu. Forward and reverse entropy power inequalities in convex geometry. In *Convexity and Concentration*, pages 427–485. Springer, 2017.
- [2] Thomas A Courtade and Jingbo Liu. Euclidean forward-reverse Brascamp–Lieb inequalities: Finiteness, structure, and extremals. *The Journal of Geometric Analysis*, 31(4):3300–3350, 2021.
- [3] Aart J Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959.
- [4] Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge university press, 2011.
- [5] Efe Aras, Kuan-Yun Lee, Ashwin Pananjady, and Thomas A Courtade. A family of Bayesian Cramér–Rao bounds, and consequences for log-concave priors. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 2699–2703. IEEE, 2019.
- [6] Eric A Carlen. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 101(1):194–211, 1991.
- [7] Ram Zamir and Meir Feder. A generalization of the entropy power inequality with applications. *IEEE Transactions on Information Theory*, 39(5):1723–1728, 1993.
- [8] Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158, 1975.
- [9] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete & Continuous Dynamical Systems-A*, 34(4):1533, 2014.
- [10] Mokshay Madiman, James Melbourne, and Peng Xu. Rogozin’s convolution inequality for locally compact groups. *arXiv preprint arXiv:1705.00642*, 2017.
- [11] Seiji Takano. The inequalities of Fisher information and entropy power for dependent variables. In S. Watanabe, M. Fukushima, Y. Prohorov, and A. Shiryaev, editors, *Proceedings of the 7th Japan-Russia Symposium on Probability Theory and Mathematical Statistics, Tokyo*, pages 460–470. World Scientific, 1995.
- [12] Oliver Johnson. A conditional entropy power inequality for dependent variables. *IEEE transactions on information theory*, 50(8):1581–1583, 2004.
- [13] Thomas M Cover and Zhen Zhang. On the maximum entropy of the sum of two dependent random variables. *IEEE Transactions on Information Theory*, 40(4):1244–1246, 1994.
- [14] Max Costa and Thomas Cover. On the similarity of the entropy power inequality and the Brunn–Minkowski inequality (corresp.). *IEEE Transactions on Information Theory*, 30(6):837–839, 1984.
- [15] Amir Dembo, Thomas M Cover, and Joy A Thomas. Information theoretic inequalities. *IEEE Transactions on Information theory*, 37(6):1501–1518, 1991.
- [16] Herm Jan Brascamp and Elliott H Lieb. On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976.
- [17] Elliott H Lieb. Proof of an entropy conjecture of Wehrl. *Communications in Mathematical Physics*, 62(1):35–41, 1978.
- [18] Liyao Wang and Mokshay Madiman. Beyond the entropy power inequality, via rearrangements. *IEEE Transactions on Information Theory*, 60(9):5116–5137, 2014.
- [19] Venkat Anantharam, Varun Jog, and Chandra Nair. Unifying the Brascamp–Lieb inequality and the entropy power inequality. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 1847–1851. IEEE, 2019.
- [20] Jingbo Liu, Thomas A Courtade, Paul W Cuff, and Sergio Verdú. A forward-reverse Brascamp–Lieb inequality: Entropic duality and Gaussian optimality. *Entropy*, 20(6):418, 2018.