Unifying the Brascamp-Lieb Inequality and the Entropy Power Inequality

Venkat Anantharam
Dept. of EECS, UC Berkeley
ananth@berkeley.edu

Varun Jog
Dept. of ECE, UW - Madison
vjog@wisc.edu

Chandra Nair
Dept. of IE, CUHK
chandra@ie.cuhk.hk.edu

***Abstract*—The entropy power inequality (EPI) and the Brascamp-Lieb inequality (BLI) are fundamental inequalities concerning the differential entropies of linear transformations of random vectors. The EPI provides lower bounds for the differential entropy of linear transformations of random vectors with independent components. The BLI, on the other hand, provides upper bounds on the differential entropy of a random vector in terms of the differential entropies of some of its linear transformations. In this paper, we define a family of entropy functionals, which we show are subadditive. We then establish that Gaussians are extremal for these functionals by mimicking the idea in Geng and Nair (2014). As a consequence, we obtain a new entropy inequality that generalizes both the BLI and EPI. By considering a variety of independence relations among the components of the random vectors appearing in these functionals, we also obtain families of inequalities that lie between the EPI and the BLI.**

***Index Terms*—Entropy power inequality, Brascamp-Lieb inequality, subadditivity**

I. INTRODUCTION

Information inequalities provide some of the most powerful mathematical tools in an information theorist's toolbox and are therefore a vital part of information theory. Inequalities such as the non-negativity of mutual information and the data processing inequality are so fundamental to information theory that they are inseparable from information-theoretic notation. These basic inequalities, combined with Fano's inequality, are powerful enough to provide a proof of the converse of Shannon's channel coding theorem. For harder problems in network information theory, it is necessary to develop more nuanced information inequalities. Not surprisingly, it is often the case that discovering new inequalities leads to breakthroughs in network information theory problems.

On a related note "single-letter characterizations" of a capacity region or outer bounds to a capacity region in network information theory are induced by subadditive functionals that reduce the characterization of the region to one governed by a single channel use. In this paper we identify a new functional that is sub-additive and for which Gaussian distributions are extremal. Consequently, we obtain a new class of information inequalities that unifies two fundamental inequalities: the entropy power inequality (EPI) and the Brascamp-Lieb inequality (BLI). In what follows, we provide a brief introduction to the EPI and the BLI and state our main results.

As notational conventions in what follows, $:=$ and $=:$ denote equality by definition, while, for an integer $n > 0$, $[n]$ denotes $\{1, \ldots, n\}$ and $I_{n\times n}$ denotes the $n \times n$ identity matrix. All vectors are thought of as column vectors. For random vectors X and Y, we write $Z = (X, Y)$ for the random vector that would normally be written as $Z = [X^T, Y^T]^T$.

Entropy power inequality: The EPI states that for independent $\mathbb{R}^n$-valued random variables X and Y we have:

$$e^{\frac{2h(X+Y)}{n}} \geq e^{\frac{2h(X)}{n}} + e^{\frac{2h(Y)}{n}}, \tag{1}$$

where, $h(\cdot)$ refers to the differential entropy function and all the differential entropies in equation (1) are assumed to exist. Equality holds if and only if X and Y are Gaussian random variables with proportional covariance matrices. The EPI was proposed by Shannon [1] and was first proved by Stam [2]. The EPI has an equivalent statement due to Lieb [3], which is that for all $\lambda \in (0, 1)$ we have:

$$h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda h(X) + (1-\lambda)h(Y). \tag{2}$$

Equality holds in the above inequality if and only if X and Y are Gaussian random variables with identical covariance matrices. Note that $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$ may be interpreted as a linear transformation of an $\mathbb{R}^{2n}$-valued random variable $Z := (X, Y)$ with some independence constraints on the components of Z, namely $X \perp\!\!\!\perp Y$. Another result along such lines is Zamir and Feder's EPI [4] for linear transformations of random vectors with independent components. This EPI has an equivalent formulation, discovered in [5], [6], that is analogous to the one in equation (2): For an $\mathbb{R}^n$-valued random vector $X := (X_1, \ldots, X_n)$ with independent components and any $k \times n$ matrix A satisfying $AA^T = I_k$, Zamir and Feder's EPI states that

$$h(AX) \geq \sum_{j=1}^{n} \alpha_j^2 h(X_j), \tag{3}$$

where α_j^2 is the squared-norm of the j-th column of A; i.e., $\alpha_j^2 := \sum_{i=1}^{k} a_{ij}^2$.

Brascamp-Lieb inequality: The BLI [7] is actually a family of functional inequalities that lies, in some sense, at the intersection of information and functional inequalities. Many well-known and commonly used inequalities are special cases of the BLI, including Hölder's inequality, the Loomis-Whitney inequality, the Prékopa-Leindler inequality, and sharp forms of Young's convolution inequalities [8], [9].

978-1-5386-9291-2/19/$31.00 ©2019 IEEE

Theorem 1 (Functional form of the BLI): For $j \in [m]$, let E, E_j be Euclidean spaces, $A_j : E \to E_j$ be linear maps, c_j be positive real numbers, and f_j be nonnegative integrable functions on E_j. Define the function $\mathcal{F}$ via

$$\mathcal{F}(f_1, \ldots, f_m) := \frac{\int_E \prod_{j=1}^m f_j^{c_j}(A_j x)dx}{\prod_{j=1}^m \left(\int_{E_j} f_j(x_j)dx_j\right)^{c_j}}.$$

Then the supremum of $\mathcal{F}$ over all nonnegative and integrable f_j is equal to the supremum of $\mathcal{F}$ when f_j are centered Gaussian functions; i.e., for all $j \in [m]$, we have $f_j(x_j) \propto e^{-x_j^T B_j x_j}$ for some positive semidefinite B_j.

Surprisingly, a direct connection exists between the functional form of the BLI and a generalized subadditivity result for differential entropy. This link was first discovered in Carlen et al. [10], and has since led to new proofs and generalizations of the original BLI [11]–[15]. The information-theoretic form of the BLI is the following:

Theorem 2 (Information-theoretic form of the BLI): For $i \in [m]$, let E, E_i, A_i, and c_i be as in Theorem 1. For a random variable X on E with a well-defined differential entropy (see Definition 1) and satisfying $E[\|X\|_2]^2 < \infty$, define $f(X)$ as

$$f(X) := h(X) - \sum_{j=1}^m c_j h(A_j X). \quad (4)$$

Then the supremum of f over all such random variables X is equal to the supremum of f over all Gaussian random variables.

This information-theoretic form is completely equivalent to the functional form: For a fixed choice of the A_j and c_j, the suprema in both problems have a direct relationship, and the cases of equality are also in correspondence [11, Theorem 2.1]. For this reason, we will only consider the information-theoretic form of the BLI in this paper.

Our contributions: The classical EPI and the EPI of Zamir and Feder are valid only under certain independence assumptions. To be precise, for an $\mathbb{R}^{2n}$-valued random vector Z, the EPI requires independence of $Z_{1:n}$ and $Z_{n+1:n}$ and considers the sum of these two vectors, whereas Zamir and Feder's EPI requires all the components to be independent and considers linear transformations of Z. Here we use the notation $Z_{a:b}$ to denote the random vector $(Z_a, Z_{a+1}, \ldots, Z_b)$. It is natural to consider more general "mixed" independence constraints, for instance independence of $Z_{1:k_1}, Z_{k_1+1:k_2}, \ldots, Z_{k_r+1:n}$ for suitable choices of k_i, and establish lower bounds on $h(AZ)$ for a matrix A. This is indeed a special case of the setting considered in our work.

Consider an $\mathbb{R}^n$-valued random vector $X := (X_1, \ldots, X_k)$, where $k \leq n$ and X_i are mutually independent $\mathbb{R}^{r_i}$-valued random variables. Note that $\sum_{i=1}^k r_i = n$. We consider the following function:

$$f(X) := \sum_{i=1}^k d_i h(X_i) - \sum_{j=1}^m c_j h(A_j X), \quad (5)$$

for positive constants d_i and c_j where $i \in [k]$ and $j \in [m]$ for some $m \geq 1$, and surjective linear transformations A_j from $\mathbb{R}^n$ to $\mathbb{R}^{n_j}$. Our main result in Theorem 3 states that the supremum of $f(\cdot)$ over all random variables X satisfying the stated independence constraints is the same as the supremum evaluated over Gaussian random variables. Moreover, we identify necessary and sufficient conditions on n, k, m and the r_i, d_i, c_j, n_j and A_j such that this supremum is finite. Theorem 3 also provides a generalization of Zamir and Feder's result to certain kinds of dependent random variables.

Our main technical contribution is identifying new entropic functionals that satisfy a certain subadditivity property, allowing us to prove Gaussian optimality for these functionals. This proof strategy first appeared in [16], where the authors determined the capacity of a Gaussian vector broadcast channel by establishing the subadditivity property for certain information-theoretic expressions. This proof strategy has since been applied to a wide variety of problems [17]–[22].

Related work: The EPI may be thought of as a limiting special case of the BLI. Indeed, Dembo et al. [23] showed that the EPI follows from the sharp form of Young's inequality, which in turn is a special case of the BLI. A related but more geometric approach may be found in Cordero-Erausquin and Ledoux [13].

Various information-theoretic analogues of hypercontractive inequalities and reverse Brascamp-Lieb inequalities have been studied in [14], [15], [24]. Liu et al. [15], define a function F of the marginal densities of $\mathbb{R}^n$-valued random variables X as:

$$F(X) := \inf_{\{Y | Y_i \overset{d}{=} X_i,\ i \in [n]\}} \sum_{i=1}^n d_i h(Y_i) - \sum_{j=1}^m c_j h(A_j Y). \quad (6)$$

Here, by $Y_i \overset{d}{=} X_i$ we mean that the distribution of Y_i is identical to that of X_i. Theorem 8 in [15] states that the supremum of F is obtained when each X_i is a centered Gaussian random variable, in which case the infimum in definition (6) is attained when the optimal coupling Y is a jointly Gaussian random vector. Expressions (5) and (6) look very similar. However, the main difference is that (6) has an infimum over all possible couplings Y, whereas our definition in (5) enforces the unique coupling where the components Y_i are mutually independent. In particular, the EPI results as a special case of our formulation but is not directly implied using the formulation in [15].

II. Preliminaries and notation

Definition 1: For $n > 0$, let X be an $\mathbb{R}^n$-valued random variable with density f_X that lies in the convex set of probability densities

$$\left\{ f \;\middle|\; \int_{\mathbb{R}^n} f(x) \log(1 + f(x))dx < \infty \right\}. \quad (7)$$

Then we define the differential entropy of X as

$$h(X) := -\int_{\mathbb{R}^n} f_X(x) \log f_X(x)dx. \quad (8)$$

The differential entropy of a 0-dimensional random variable is defined to be 0. The integral in (7) is well-defined since both components are non-negative. The condition (7) implies that differential entropy integral is well-defined and lower-bounded away from negative infinity.

Definition 2 (Brascamp-Lieb datum): For an integer $m > 0$, define an m-transformation as a triple

$$\mathbf{A} := (n, \{n_j\}_{j\in[m]}, \{A_j\}_{j\in[m]}),$$

where for each $j \in [m]$, $A_j : \mathbb{R}^n \to \mathbb{R}^{n_j}$ is a surjective linear transformation, and $n_j \geq 0$. An m-exponent is defined as an m-tuple $\mathbf{c} = \{c_j\}_{j\in[m]}$, such that $c_j \geq 0$ for $j \in [m]$. A Brascamp-Lieb (BL) datum is defined as a pair $(\mathbf{A}, \mathbf{c})$ where $\mathbf{A}$ is an m-transformation and $\mathbf{c}$ is an m-exponent, for an integer $m > 0$.

Definition 3 (EPI datum): For an integer $k > 0$, define a k-partition of n as $\mathbf{r} = \{r_i\}_{i\in[k]}$, such that $r_i > 0$ are integers and $\sum_{i\in[k]} r_i = n$. Let $\mathbf{d} = \{d_i\}_{i\in[k]}$ such that $d_i \geq 0$ for all i be called a k-exponent. An EPI datum is the pair $(\mathbf{r}, \mathbf{d})$ where $\mathbf{r}$ is a k-partition and $\mathbf{d}$ is a k-exponent, for an integer $k > 0$.

Definition 4 (BL-EPI datum): For an integer $n > 0$, a BL-EPI datum is defined as $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ where $(\mathbf{A}, \mathbf{c})$ is a BL datum for an integer $m > 0$, and $(\mathbf{r}, \mathbf{d})$ is an EPI datum for an integer $k > 0$.

Definition 5: Let $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ be a BL-EPI datum where $\mathbf{r}$ is a k-partition of n. Define $\mathcal{P}(\mathbf{r})$ to be the set of all $\mathbb{R}^n$-valued random variables $X := (X_1, X_2, \dots, X_k)$ such that:

1) For $i \in [k]$, the random variables X_i take values in $\mathbb{R}^{r_i}$ respectively and their densities satisfy the condition in equation (7);
2) $X_1, X_2, \dots, X_k$ are independent;
3) $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|_2^2 < \infty$.

Define $\mathcal{P}_g(\mathbf{r}) \subseteq \mathcal{P}(\mathbf{r})$ as the set of random variables X that satisfy the properties above, in addition to each X_i, $i \in [k]$, being Gaussian.

Definition 6: For a BL-EPI datum $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$, define $M(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ as

$$M(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d}) := \sup_{X\in\mathcal{P}(\mathbf{r})} \sum_{i=1}^{k} d_i h(X_i) - \sum_{j=1}^{m} c_j h(A_j X).$$

Similarly, define $M_g(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ as the above supremum taken over Gaussian inputs $X \in \mathcal{P}_g(\mathbf{r})$. When the BL-EPI datum is fixed, we shall omit the $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ argument and use the simiplified notation M and M_g.

The following two concepts are required for the statement of Theorem 4.

Definition 7: Let $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ be a BL-EPI datum. Define a subspace $V \subseteq \mathbb{R}^n$ as being of $\mathbf{r}$-product form if V may be written as $V = V_1 \times V_2 \times \cdots \times V_k$ for subspaces $V_i \subseteq \mathbb{R}^{r_i}$, for $i \in [k]$.

III. Main results

Our main result are as follows:

Theorem 3 (Unified EPI and BLI): Let $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ be a BL-EPI datum. Recall the definition

$$M_g := \sup_{Z\in\mathcal{P}_g(\mathbf{r})} \sum_{i=1}^{k} d_i h(Z_i) - \sum_{j=1}^{m} c_j h(A_j Z). \tag{9}$$

Then for any $X \in \mathcal{P}(\mathbf{r})$, the following inequality holds:

$$\sum_{i=1}^{k} d_i h(X_i) - \sum_{j=1}^{m} c_j h(A_j X) \leq M_g. \tag{10}$$

Naturally, we have $M \geq M_g$. Thus, if M_g is $+\infty$, then so is M. If $M_g < \infty$, then the above result implies $M \leq M_g$, and thus $M = M_g$. An equivalent way of stating the above result is asserting $M = M_g$.

Theorem 3 yields a useful result only if $M < \infty$. We show the following necessary and sufficient conditions for this to hold:

Theorem 4: For a BL-EPI datum $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$, we have $M(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d}) < \infty$ if and only if $\sum_{i=1}^{k} d_i r_i = \sum_{j=1}^{m} c_j n_j$ and for all $\mathbf{r}$-product form V,

$$\sum_{i=1}^{k} d_i \dim(V_i) \leq \sum_{j=1}^{m} c_j \dim(A_j V).$$

Due to space constraints, we shall not present the complete proof of Theorem 3. In Section IV we distill the novel contributions of this work by proving the critical subadditivity step for a new entropy functional. We shall omit the proof of Theorem 4 in this paper. The detailed proofs of both theorems may be found in [25].

IV. Subadditivity lemma

As noted before, our proof mimics the strategy developed in [16] for solving optimization problems of the following form: $\sup_{\mathrm{Cov}(X)\preceq\Sigma} s(X)$. The upper-concave envelope of s, denoted by S, is defined as

$$S(X) = \sup_U s(X|U) = \sup_U \sum_{u\in\mathcal{U}} s(X|U = u) p_U(u),$$

where the supremum is taken over auxiliary random variables U taking values in finite sets $\mathcal{U}$ of arbitrary cardinality. The most crucial step in this proof strategy is establishing a certain "subadditivity" result for the S function. The ingredients for establishing the subadditivity result developed here stems from the ideas to establish converses to coding theorems and outer bounds in network information theory. An argument with the flavor of the argument employed here can be found outlined in [26]. The main idea is to exploit the chain rule for entropy in two separate ways. Given a random vector (X_1, X_2), we use the two expansions for the joint differential entropy $h(X_1, X_2)$:

(A) $h(X_1, X_2) = h(X_1) + h(X_2) - I(X_1; X_2)$,
(B) $h(X_1, X_2) = h(X_1|X_2) + h(X_2|X_1) + I(X_1; X_2)$.

Let $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ be a BL-EPI datum. Let $X := (X_1, X_2, \ldots, X_k) \in \mathcal{P}(\mathbf{r})$, where $X_i \sim p_{X_i}$. A natural definition for $s(X)$ is

$$s(X) := \sum_{i=1}^{k} d_i h(X_i) - \sum_{j=1}^{m} c_j h(A_j X). \tag{11}$$

In [25], we use an alternate definition where we use Gaussian perturbed random variables to make the analysis rigorous in some technical aspects. However, the main strategy for proving the subadditivity lemma is identical in both cases. Let S be the upper-concave envelope of s. Consider the space $\mathcal{P}(2\mathbf{r})$, defined as the set of all random variables (X_1, X_2) such that $X_1 := (X_{11}, \ldots, X_{k1})$ and $X_2 := (X_{12}, \ldots, X_{k2})$ are $\mathbb{R}^n$-valued random vectors such that the random vectors $(X_{i1}, X_{i2}) \in \mathbb{R}^{2r_i}$, $i \in [k]$ are mutually independent, (X_{i1}, X_{i2}), $i \in [k]$ satisfy the condition in equation (7), and condition (3) of Definition 5 holds for (X_1, X_2). Define the extension of s to this space as

$$s(X_1, X_2) := \sum_{i=1}^{k} d_i h(X_{i1}, X_{i2}) - \sum_{j=1}^{m} c_j h(A_j X_1, A_j X_2). \tag{12}$$

As before, the upper-concave envelope of this extension is denoted by $S(X_1, X_2)$.

Lemma 1 (Subadditivity lemma): The function S is subadditive; i.e., if $(X_1, X_2) \in \mathcal{P}(2\mathbf{r})$ then

$$S(X_1, X_2) \leq S(X_1) + S(X_2). \tag{13}$$

Proof of Lemma 1:

Let U be an arbitrary auxiliary random variable such that $p_{X_1 X_2|U} \in \mathcal{P}(2\mathbf{r})$. Consider the following expansion, motivated by expansion (A):

$$\begin{aligned} s(X_1, X_2 \mid U) = &\left[\sum_{i=1}^{k} d_i h(X_{i1}|U) - \sum_{j=1}^{m} c_j h(A_j X_1|U)\right] \\ &+ \left[\sum_{i=1}^{k} d_i h(X_{i2}|U) - \sum_{j=1}^{m} c_j h(A_j X_2|U)\right] \\ &+ \left[-\sum_{i=1}^{k} d_i I(X_{i1}; X_{i2}|U) + \sum_{j=1}^{m} c_j I(A_j X_1; A_j X_2|U)\right]. \end{aligned} \tag{14}$$

For simplicity, denote the terms in the square brackets by $T_1(U)$, $T_2(U)$, and $T_3(U)$, respectively. Since $p_{X_1|U}(\cdot|U), p_{X_2|U}(\cdot|U) \in \mathcal{P}(\mathbf{r})$, we have $T_1(U) \leq S(X_1)$ and $T_2(U) \leq S(X_2)$. Substituting these inequalities, we arrive at

$$s(X_1, X_2|U) \leq S(X_1) + S(X_2) + T_3(U). \tag{15}$$

We now expand $s(X_1, X_2 \mid U)$ in a different way, motivated by expansion (B):

$$\begin{aligned} s(X_1, X_2 \mid U) = &\Big[\sum_{i=1}^{k} d_i h(X_{i1}|U, X_{i2}) - \sum_{j=1}^{m} c_j h(A_j X_1|U, A_j X_2)\Big] \\ &+ \Big[\sum_{i=1}^{k} d_i h(X_{i2}|U, X_{i1}) - \sum_{j=1}^{m} c_j h(A_j X_2|U, A_j X_1\Big] \\ &+ \Big[\sum_{i=1}^{k} d_i I(X_{i1}; X_{i2}|U) - \sum_{j=1}^{m} c_j I(A_j X_1; A_j X_2|U)\Big]. \end{aligned} \tag{16}$$

For ease of notation, call the three terms in the square brackets $R_1(U)$, $R_2(U)$, and $R_3(U) = -T_3(U)$, respectively. Similar to inequality (15), we would like to upper bound $R_1(U)$ and $R_2(U)$ by $S(X_1)$ and $S(X_2)$ respectively. However, the conditioning in each of the two differential entropy terms in each $R_a(U), a = 1, 2$ is not the same. Using the chain rule of mutual information and data processing relations, we may make the conditioning in $R_1(U)$ and $R_2(U)$ uniform by introducing some extra mutual information terms:

$$\begin{aligned} R_1(U) &= \Big[\sum_{i=1}^{k} d_i h(X_{i1}|U, X_{i2}) - \sum_{j=1}^{m} c_j h(A_j X_1|U, A_j X_2)\Big] \\ &= \Big[\sum_{i=1}^{k} d_i h(X_{i1}|U, X_2) - \sum_{j=1}^{m} c_j h(A_j X_1|U, X_2)\Big] \\ &\quad - \Big[\sum_{j=1}^{m} c_j I(A_j X_1; X_2|U, A_j X_2)\Big] \\ &=: \tilde{R}_1(U) - I_1(U). \end{aligned}$$

The above steps are justified as follows. First, it is easy to check that $X_{i1} \perp\!\!\!\perp \{X_{l2}\}_{l \neq i}$ conditioned on (U, X_{i2}). This means that, for all $1 \leq i \leq k$, $h(X_{i1}|U, X_{i2}) = h(X_{i1}|U, X_2)$. Also, we may verify the Markov chain (conditioned on U) $A_j X_2 \to X_2 \to A_j X_1$, which gives the equality

$$h(A_j X_1|U, A_j X_2) = h(A_j X_1|U, X_2) + I(A_j X_1; X_2|U, A_j X_2).$$

Similar reasoning for R_2 gives

$$\begin{aligned} R_2(U) &= \Big[\sum_{i=1}^{k} d_i h(X_{i2}|U, X_{i1}) - \sum_{j=1}^{m} c_j h(A_j X_2|U, A_j X_1)\Big] \\ &= \Big[\sum_{i=1}^{k} d_i h(X_{i2}|U, X_1) - \sum_{j=1}^{m} c_j h(A_j X_2|U, X_1)\Big] \\ &\quad - \Big[\sum_{j=1}^{m} c_j I(A_j X_2; X_1|U, A_j X_1)\Big] \\ &=: \tilde{R}_2(U) - I_2(U). \end{aligned}$$

Substituting the expressions for $R_1(U)$ and $R_2(U)$ in the expansion (16), we arrive at

$$\begin{aligned} s(X_1, X_2 \mid U) &= \tilde{R}_1(U) + \tilde{R}_2(U) - T_3(U) - I_1(U) - I_2(U) \\ &\overset{(a)}{\leq} S(X_1) + S(X_2) - T_3(U) - I_1(U) - I_2(U) \\ &\overset{(b)}{\leq} S(X_1) + S(X_2) - T_3(U). \end{aligned} \tag{17}$$

Here, Step (a) uses $p_{X_1|U}(\cdot|U), p_{X_2|U}(\cdot|U) \in \mathcal{P}(\mathbf{r})$. Step (b) uses the non-negativity of the c_j and of $I_1(U)$ and $I_2(U)$.

Combining equations (15) and (17) gives

$$s(X_1, X_2 \mid U) \leq S(X_1) + S(X_2).$$

Taking the supremum on the left hand side over auxiliary random variables U yields the claimed subadditivity result. ∎

Proof of Theorem 3: Having proved the key subadditivity step, the proof follows the steps outlined in [16, Appendix II]. Start with two independent copies X_1^*, X_2^*, of the maximizing distribution; form the new pair $\left(\frac{X_1^*+X_2^*}{\sqrt{2}}, \frac{X_1^*-X_2^*}{\sqrt{2}}\right)$ of (potentially dependent) distributions; and argue from the proof of subadditivity and the maximality of X_1^*, X_2^* that $\frac{X_1^*+X_2^*}{\sqrt{2}}$ and $\frac{X_1^*-X_2^*}{\sqrt{2}}$ are independent, implying that X_1^*, X_2^* have to be Gaussians. To make the technical aspects of these steps go through, we need to use perturbed versions of the functions and distributions considered. The detailed proof can be found in [25].

Remark 1: In [25] we present some new forms of inequalities that arise from Theorem 3. An example of a new inequality of this form is a non-trivial lower bound for $h(X_1+Y, X_2+Y)$ when $(X_1, X_2) \perp\!\!\!\perp Y$, under some mild assumptions.

V. Conclusion

We established a new inequality that unifies the BLI and the EPI by identifying a new class of entropic functionals for which a subadditivity property holds. There are several interesting research directions that are worth pursuing. We did not address the questions of uniqueness of extremizers for the family of functionals that we considered. Finally, although our results generalize the BLI and the EPI to vector random variables with more general independence properties, these independence properties are still quite restrictive. It would be interesting to establish similar entropy inequalities under weaker independence conditions.

Acknowledgements

VA was supported by the NSF grants CNS-1527846 and CCF-1618145, the NSF Science & Technology Center grant CCF-0939370 (Science of Information), and the William and Flora Hewlett Foundation supported Center for Long Term Cybersecurity at Berkeley. VJ is supported by the NSF grant CCF-1841190, and is grateful to the Department of Information Engineering at CUHK for hosting him in July 2018, when a part of this work was done. The research of CN was supported by GRF grants 14303714, 14231916, 14206518 and a discretionary fund of the Vice Chancellor of CUHK.

References

[1] C. E. Shannon. A mathematical theory of communication, I and II. *Bell Syst. Tech. J*, 27:379–423, 1948.

[2] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959.

[3] E. H. Lieb. Proof of an entropy conjecture of Wehrl. In *Inequalities*, pages 359–365. Springer, 2002.

[4] R. Zamir and M. Feder. A generalization of the entropy power inequality with applications. *IEEE Transactions on Information Theory*, 39(5):1723–1728, 1993.

[5] O. Rioul. Information theoretic proofs of entropy power inequalities. *IEEE Transactions on Information Theory*, 57(1):33–55, 2011.

[6] R. Zamir and M. Feder. A generalization of information theoretic inequalities to linear transformations of independent vector. In *Proceedings of the 6-th Joint Swedish-Russian International Workshop on Information Theory*, pages 254–258, 1993.

[7] 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.

[8] R. Gardner. The Brunn-Minkowski inequality. *Bulletin of the American Mathematical Society*, 39(3):355–405, 2002.

[9] 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.

[10] 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.

[11] 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.

[12] F. Barthe, D. Cordero-Erausquin, and B. Maurey. Entropy of spherical marginals and related inequalities. *Journal de Mathématiques Pures et Appliquées*, 86(2):89–99, 2006.

[13] D. Cordero-Erausquin and M. Ledoux. The geometry of Euclidean convolution inequalities and entropy. *Proceedings of the American Mathematical Society*, 138(8):2755–2769, 2010.

[14] C. Nair. *Equivalent formulations of hypercontractivity using information measures*. International Zurich Seminar on Communications (IZS), 2014. Available at http://chandra.ie.cuhk.edu.hk/pub/papers/manuscripts/IZS14.pdf.

[15] J. Liu, T. Courtade, P. Cuff, and S. Verdú. A forward-reverse Brascamp-Lieb inequality: Entropic duality and Gaussian optimality. *Entropy*, 20(6):418, 2018.

[16] Y. Geng and C. Nair. The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages. *IEEE Transactions on Information Theory*, 60(4):2087–2104, 2014.

[17] T. Courtade and J. Jiao. An extremal inequality for long Markov chains. In *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, pages 763–770. IEEE, 2014.

[18] H. Kim, B. Nachman, and A. El Gamal. Superposition coding is almost always optimal for the Poisson broadcast channel. *IEEE Transactions on Information Theory*, 62(4):1782–1794, 2016.

[19] Z. Goldfeld. MIMO Gaussian broadcast channels with common, private and confidential messages. In *Information Theory Workshop (ITW), 2016 IEEE*, pages 41–45. IEEE, 2016.

[20] T. Yang, N. Liu, W. Kang, and S. S. Shitz. An upper bound on the sum capacity of the downlink multicell processing with finite backhaul capacity. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2053–2057. IEEE, 2017.

[21] X. Zhang, V. Anantharam, and Y. Geng. Gaussian optimality for derivatives of differential entropy using linear matrix inequalities. *Entropy*, 20(3):182, 2018.

[22] T. A. Courtade. A strong entropy power inequality. *IEEE Transactions on Information Theory*, 64(4):2173–2192, 2018.

[23] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *Information Theory, IEEE Transactions on*, 37(6):1501–1518, 1991.

[24] 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.

[25] V. Anantharam, V. Jog, and C. Nair. Unifying the Brascamp-Lieb inequality and the entropy power inequality. *arXiv preprint arXiv:1901.06619*, 2019.

[26] C. Nair. *An extremal inequality related to hypercontractivity of Gaussian random variables*. Information Theory and Applications Workshop (2014), 2014. Available at http://chandra.ie.cuhk.edu.hk/pub/papers/manuscripts/ITA14.pdf.