Strengthening the Entropy Power Inequality

Thomas A. Courtade
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Email: courtade@berkeley.edu

***Abstract*—We tighten the entropy power inequality (EPI) when one of the random summands is Gaussian. Our strengthening is closely related to strong data processing for Gaussian channels and generalizes the (vector extension of) Costa's EPI. This leads to a new reverse EPI and, as a corollary, sharpens Stam's inequality relating entropy power and Fisher information. Applications to network information theory are given, including a short self-contained proof of the converse for the two-encoder quadratic Gaussian source coding problem. The proof of our main result is based on weak convergence and a doubling argument for establishing Gaussian optimality via rotational-invariance.**

I. Introduction and Main Result

For a random variable X with density f, the differential entropy of X is defined by[1]

$$h(X) = -\int f(x)\log f(x)dx. \tag{1}$$

Similarly, $h(\mathbf{X})$ is defined to be the differential entropy of a random vector $\mathbf{X}$ on $\mathbb{R}^n$. Shannon's celebrated entropy power inequality (EPI) asserts that for X, W independent

$$2^{2h(X+W)} \geq 2^{2h(X)} + 2^{2h(W)}. \tag{2}$$

Under the assumption that W is Gaussian, we prove the following strengthening of (2):

Theorem 1. *Let $X \sim P_X$, and let $W \sim N(0,\sigma^2)$ be independent of X. For any V satisfying $X \to (X+W) \to V$,*

$$2^{2(h(X+W)-I(X;V))} \geq 2^{2(h(X)-I(X+W;V))} + 2^{2h(W)}. \tag{3}$$

The notation $X \to (X+W) \to V$ indicates that the random variables X, $X+W$ and V form a Markov chain, in that order. Throughout, we write $X \to Y \to V|Q$ to denote random variables X, Y, V, Q with joint distribution factoring as $P_{XYVQ} = P_{XQ}P_{Y|XQ}P_{V|YQ}$. That is, $X \to Y \to V$ form a Markov chain conditioned on Q.

To familiarize the reader with (3), we remark that the function

$$g_I(t, P_{AB}) = \sup_{V: A\to B\to V} \left\{I(V;A) : I(V;B) \leq t\right\}, \tag{4}$$

is the best-possible (or *strong*) data-processing function (cf. [1]) for the pair $(A,B) \sim P_{AB}$ since $I(V;A) \leq g_I(I(V;B), P_{AB})$ for any V satisfying $A \to B \to V$. Adopting the notation of Theorem 1 and defining $Y = X+W$, inequality (3) may be restated as

$$2^{2(h(Y)-g_I(t,P_{XY}))} \geq 2^{2(h(X)-t)} + 2^{2h(W)} \quad \text{for all } t \geq 0.$$

Hence, the slack in Shannon's EPI (2) for Gaussian W (due to non-Gaussianness of X) can be improved by choosing an appropriate point on the strong data processing curve $g_I(\cdot, P_{XY})$. In view of this, (3) might be called a *strong EPI*.

This work was supported by NSF Grants CCF-1528132 and CCF-0939370 (Center for Science of Information).

[1] When the integral (1) does not exist, or if X does not have density, then we adopt the convention that $h(X) = -\infty$.

We remark that there exist various improvements of the EPI in the literature (e.g,. [2] for log-concave densities, and [3] for subsets), however none are notably similar to that presented here. The reader is referred to the recent survey [4] for an overview.

A conditional version of the EPI is often useful in applications. Theorem 1 easily generalizes along these lines. Indeed, due to joint convexity of $\log(2^x + 2^y)$ in x, y, we obtain the following corollary of Theorem 1:

Corollary 1. *Suppose X, W are conditionally independent given Q, and moreover that W is conditionally Gaussian given Q. Then, for any V satisfying $X \to (X+W) \to V|Q$,*

$$2^{2(h(X+W|Q)-I(X;V|Q))} \geq 2^{2(h(X|Q)-I(X+W;V|Q))} + 2^{2h(W|Q)}. \tag{5}$$

As one would expect, Theorem 1 also admits a vector generalization, which may be regarded as our main result:

Theorem 2. *Suppose $\mathbf{X}, \mathbf{W}$ are n-dimensional random vectors that are conditionally independent given Q, and moreover that $\mathbf{W}$ is conditionally Gaussian given Q. Then, for any V satisfying $\mathbf{X} \to (\mathbf{X}+\mathbf{W}) \to V|Q$,*

$$2^{\frac{2}{n}(h(\mathbf{X}+\mathbf{W}|Q)-I(\mathbf{X};V|Q))} \geq 2^{\frac{2}{n}(h(\mathbf{X}|Q)-I(\mathbf{X}+\mathbf{W};V|Q))} + 2^{\frac{2}{n}h(\mathbf{W}|Q)}. \tag{6}$$

The restriction of $\mathbf{W}$ to be conditionally Gaussian in Theorem 2 should not be viewed as a severe limitation. Indeed, in typical applications of the EPI, one of the variables is Gaussian as surveyed by Rioul [5, Section I].

In the following section, we demonstrate several applications of Theorem 2. In particular, we show that Costa's EPI [6] and its generalization [7] are immediate corollaries of Theorem 2, along with a new reverse EPI and a sharpening of the Stam-Gross logarithmic Sobolev inequality. Also, we will see that Theorem 2 leads to a very brief proof of the converse for the rate region of the quadratic Gaussian two-encoder source-coding problem [8]. Finally, an application to the one-sided Gaussian interference channel is discussed. Proofs of the main results are sketched in Section III.

II. Applications

A. Generalized Costa's EPI and a New Reverse EPI

Costa's EPI [6] asserts concavity of entropy power, and has been generalized to a vector setting by Liu *et al.* [7]. We demonstrate below that this generalization follows as a corollary to Theorem 2 by taking V equal to $\mathbf{X}$ contaminated by additive Gaussian noise. In this sense, Theorem 2 may be interpreted as a further generalization of Costa's EPI, where the additive noise is no longer restricted to be Gaussian. First, we have the following new EPI for three random summands (one of which is Gaussian):

978-1-5090-1805-5/16/$31.00 ©2016 IEEE

Theorem 3. *Let $\mathbf{X} \sim P_{\mathbf{X}}, \mathbf{Z} \sim P_{\mathbf{Z}}$ and $\mathbf{W} \sim N(0, \Sigma)$ be independent, n-dimensional random vectors. Then*

$$2^{\frac{2}{n}(h(\mathbf{X}+\mathbf{W})+h(\mathbf{Z}+\mathbf{W}))} \geq 2^{\frac{2}{n}(h(\mathbf{X})+h(\mathbf{Z}))} + 2^{\frac{2}{n}(h(\mathbf{X}+\mathbf{Z}+\mathbf{W})+h(\mathbf{W}))}. \quad (7)$$

Proof. This is an immediate consequence of Theorem 2 by putting $V = \mathbf{X} + \mathbf{Z} + \mathbf{W}$ and rearranging exponents. □

The vector generalization of Costa's EPI now follows:

Theorem 4. *[7] Let $\mathbf{X} \sim P_{\mathbf{X}}$ and $\mathbf{N} \sim N(0, \Sigma)$ be independent, n-dimensional random vectors. For a positive semidefinite matrix $A \preceq I$,*

$$2^{\frac{2}{n}h(\mathbf{X}+A^{1/2}\mathbf{N})} \geq |I - A|^{1/n} 2^{\frac{2}{n}h(\mathbf{X})} + |A|^{1/n} 2^{\frac{2}{n}h(\mathbf{X}+\mathbf{N})}.$$

Proof. Let $\mathbf{N}_1, \mathbf{N}_2$ be independent copies of $\mathbf{N}$, and put $\mathbf{W} = A^{1/2}\mathbf{N}_1$ and $\mathbf{Z} = (I - A)^{1/2}\mathbf{N}_2$. Since $\mathbf{N} = \mathbf{Z} + \mathbf{W}$ in distribution, the claim follows from Theorem 3. □

Theorem 3 also yields a new reverse EPI as a corollary, which sharpens Stam's inequality. Toward this end, suppose $\mathbf{X}$ has smooth density f and define the entropy power $N(\mathbf{X})$ and Fisher information $J(\mathbf{X})$ as

$$N(\mathbf{X}) = \tfrac{1}{2\pi e} 2^{\frac{2}{n}h(\mathbf{X})} \qquad J(\mathbf{X}) = \mathbb{E}\|\nabla \ln f(\mathbf{X})\|^2. \quad (8)$$

Letting $\mathbf{W} \sim N(0, tI)$ in Theorem 3 and applying de Bruijn's Identity as $t \to 0$, we obtain the following reverse EPI:

Theorem 5. *Let $\mathbf{X} \sim P_{\mathbf{X}}, \mathbf{Z} \sim P_{\mathbf{Z}}$ be independent, n-dimensional random vectors with smooth densities. Then*

$$nN(\mathbf{X} + \mathbf{Z}) \leq N(\mathbf{X})N(\mathbf{Z}) \left(J(\mathbf{X}) + J(\mathbf{Z})\right). \quad (9)$$

In contrast to other reverse EPIs that tend to be more restrictive in their assumptions (e.g., log-concavity of densities [9]), inequality (9) applies generally and bounds the entropy power $N(\mathbf{X}+\mathbf{Z})$ in terms of the marginal entropy powers and Fisher informations. We refer the reader to the survey [4] for an overview of other known reverse EPIs.

Stam's inequality [10] (or, equivalently, the Gaussian logarithmic Sobolev inequality [11]) states $N(\mathbf{X})J(\mathbf{X}) \geq n$. By taking $\mathbf{X}, \mathbf{Z}$ to be IID in (9), we have sharpened Stam's inequality:

$$N(\mathbf{X})J(\mathbf{X}) \geq n \frac{N\left(\frac{1}{\sqrt{2}}(\mathbf{X} + \mathbf{Z})\right)}{N(\mathbf{X})}. \quad (10)$$

Finally, we note that Stam's inequality controls the deficit in the classical EPI in the following sense: if both $\mathbf{X}$ and $\mathbf{Z}$ nearly saturate Stam's inequality, then (9) can be rearranged to show the classical EPI will also be nearly saturated for the sum $\mathbf{X}+\mathbf{Z}$. A similar statement holds for the convolution inequality for Fisher information. Indeed, applying Stam's inequality to the sum $\mathbf{X} + \mathbf{Z}$, inequality (9) yields:

$$\frac{1}{J(\mathbf{X} + \mathbf{Z})} \leq \left(\frac{1}{J(\mathbf{X})} + \frac{1}{J(\mathbf{Z})}\right) \mathsf{p}(\mathbf{X})\mathsf{p}(\mathbf{Z}), \quad (11)$$

where $\mathsf{p}(\mathbf{X}) := \frac{1}{n} N(\mathbf{X})J(\mathbf{Z}) \geq 1$, and $\mathsf{p}(\mathbf{Z})$ is defined similarly.

B. Two-Encoder Quadratic Gaussian Source Coding

The converse for the two-encoder quadratic Gaussian source coding problem was a longstanding open problem in the field of network information theory until its ultimate resolution by Wagner *et al.* [8]. Wagner *et al.*'s work built upon Oohama's earlier solution to the one-helper problem [12] and the independent solutions to the Gaussian CEO problem [13], [14]. Since Wagner *et al.*'s original proof of the sum-rate constraint, other proofs have been proposed (e.g., [15]), however all known proofs are quite complex. Below, we show that the converse result for the entire rate region is a direct consequence of Theorem 2, thus unifying the results of [8] and [12] under a common and succinct EPI.

Theorem 6. *[8] Let $\mathbf{X}, \mathbf{Y} = \{X_i, Y_i\}_{i=1}^n$ be independent identically distributed pairs of jointly Gaussian random variables with correlation ρ. Let $\phi_X : \mathbb{R}^n \to \{1, \ldots, 2^{nR_X}\}$ and $\phi_Y : \mathbb{R}^n \to \{1, \ldots, 2^{nR_Y}\}$, and define*

$$d_X \triangleq \frac{1}{n}\mathbb{E}\left[\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\phi_X(\mathbf{X}), \phi_Y(\mathbf{Y})]\|^2\right] \quad (12)$$

$$d_Y \triangleq \frac{1}{n}\mathbb{E}\left[\|\mathbf{Y} - \mathbb{E}[\mathbf{Y}|\phi_X(\mathbf{X}), \phi_Y(\mathbf{Y})]\|^2\right]. \quad (13)$$

Then, for $\beta(D) \triangleq 1 + \sqrt{1 + \frac{4\rho^2 D}{(1-\rho^2)^2}}$,

$$R_X \geq \frac{1}{2}\log\left(\frac{1}{d_X}\left(1 - \rho^2 + \rho^2 2^{-2R_Y}\right)\right) \quad (14)$$

$$R_Y \geq \frac{1}{2}\log\left(\frac{1}{d_Y}\left(1 - \rho^2 + \rho^2 2^{-2R_X}\right)\right) \quad (15)$$

$$R_X + R_Y \geq \frac{1}{2}\log\frac{(1 - \rho^2)\beta(d_X d_Y)}{2d_X d_Y}. \quad (16)$$

Our proof hinges on a simple corollary[2] of Theorem 2:

Proposition 1. *For $\mathbf{X}, \mathbf{Y}$ as above,*

$$2^{-\frac{2}{n}(I(\mathbf{Y};U)+I(\mathbf{X};V|U))} \geq \rho^2\, 2^{-\frac{2}{n}(I(\mathbf{X};U)+I(\mathbf{Y};V|U))} + 1 - \rho^2$$

for any U, V satisfying $U \to \mathbf{X} \to \mathbf{Y} \to V$.

Proof. Since mutual information is invariant to scaling, we may assume without loss of generality that $Y_i = \rho X_i + Z_i$, where $X_i \sim N(0, 1)$ and $Z_i \sim N(0, 1 - \rho^2)$, independent of X_i. Now, Theorem 2 implies

$$2^{\frac{2}{n}(h(\mathbf{Y}|U)-I(\mathbf{X};V|U))} \geq 2^{\frac{2}{n}(h(\rho\mathbf{X}|U)-I(\mathbf{Y};V|U))} + 2^{\frac{2}{n}h(\mathbf{Z})}$$
$$= \rho^2 2^{\frac{2}{n}(h(\mathbf{X}|U)-I(\mathbf{Y};V|U))} + 2\pi e(1 - \rho^2).$$

Since $2^{-\frac{2}{n}h(\mathbf{Y})} = 2^{-\frac{2}{n}h(\mathbf{X})} = \frac{1}{2\pi e}$, multiplying through by $\frac{1}{2\pi e}$ establishes the claim. □

Proof of Theorem 6. For convenience, put $U = \phi_X(\mathbf{X})$ and $V = \phi_Y(\mathbf{Y})$. Using the Markov relationship $U \to \mathbf{X} \to \mathbf{Y} \to V$, we may rearrange the exponents in Proposition 1 to obtain the equivalent inequality

$$2^{-\frac{2}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V))} \geq 2^{-\frac{2}{n}I(\mathbf{X},\mathbf{Y};U,V)}\left(1 - \rho^2 + \rho^2 2^{-\frac{2}{n}I(\mathbf{X},\mathbf{Y};U,V)}\right). \quad (17)$$

The left- and right-hand sides of (17) are monotone decreasing in $\frac{1}{n}(I(\mathbf{X}; U, V) + I(\mathbf{Y}; U, V))$ and $\frac{1}{n}I(\mathbf{X}, \mathbf{Y}; U, V)$, respectively. Therefore, if $\frac{1}{n}I(\mathbf{X}, \mathbf{Y}; U, V) \leq R$ and

$$\frac{1}{n}(I(\mathbf{X}; U, V) + I(\mathbf{Y}; U, V)) \geq \frac{1}{2}\log\frac{1}{D} \quad (18)$$

for some pair (R, D), then we have $D \geq 2^{-2R}\left(1 - \rho^2 + \rho^2 2^{-2R}\right)$, which is a quadratic inequality with respect to the term 2^{-2R}. This is easily solved using the quadratic formula to obtain:

$$2^{-2R} \leq \frac{2D}{(1 - \rho^2)\beta(D)} \quad \Longrightarrow \quad R \geq \frac{1}{2}\log\frac{(1 - \rho^2)\beta(D)}{2D},$$

[2] Proposition 1 was first established by the author and Jiao in [16].

where $\beta(D) \triangleq 1+\sqrt{1+\frac{4\rho^2 D}{(1-\rho^2)^2}}$. Note that Jensen's inequality and the maximum-entropy property of Gaussians imply $\frac{1}{n}I(\mathbf{X};U,V) \geq \frac{1}{2}\log\frac{1}{d_X}$ and $\frac{1}{n}I(\mathbf{Y};U,V) \geq \frac{1}{2}\log\frac{1}{d_Y}$, so that

$$\frac{1}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V)) \geq \frac{1}{2}\log\frac{1}{d_X d_Y}, \tag{19}$$

establishing (16) since $\frac{1}{n}I(\mathbf{X},\mathbf{Y};U,V) \leq \frac{1}{n}\left(H(U)+H(V)\right) \leq R_X + R_Y$. Similarly, Proposition 1 implies

$$2^{2R_X+\log d_X} \geq 2^{\frac{2}{n}(I(\mathbf{X};U|V)-I(\mathbf{X};U,V))} = 2^{-\frac{2}{n}I(\mathbf{X};V)} \tag{20}$$

$$\geq (1-\rho^2)+\rho^2 2^{-\frac{2}{n}I(\mathbf{Y};V)} \tag{21}$$

$$\geq (1-\rho^2)+\rho^2 2^{-2R_Y}. \tag{22}$$

Rearranging (and symmetry) yields (14)-(15). □

C. One-sided Gaussian Interference Channel

We now briefly discuss how Theorem 2 might be applied to the interference channel[3]. Recall that the one-sided Gaussian interference channel (GIC) is a memoryless channel, with input-output relationship given by

$$Y_1 = X_1 + W \tag{23}$$

$$Y_2 = \alpha Y_1 + X_2 + W_2, \tag{24}$$

where X_i and Y_i are the channel inputs and observations corresponding to Encoder i and Decoder i, respectively, for $i=1,2$. Here, $W \sim N(0,1)$ and $W_2 \sim N(0,1-\alpha^2)$ are independent of each other and of the channel inputs X_1, X_2. We assume $|\alpha| < 1$ since the capacity is known in the strong interference regime of $|\alpha| > 1$. Observe that we have expressed the one-sided GIC in *degraded form*, which has capacity region identical to the corresponding non-degraded version [18]. The capacity region of the one-sided GIC remains unknown in the regime of $|\alpha| < 1$.

For convenience, define $Y_0 = \alpha Y_1 + W_2$, and let $\mathscr{C}(\alpha, P_1, P_2)$ denote the set of achievable rate pairs (R_1, R_2) for the one-sided GIC described above, when user i is subject to power constraint P_i, $i=1,2$. See [19] for formal definitions.

Theorem 7. *$(R_1,R_2) \in \mathscr{C}(\alpha,P_1,P_2)$ only if*

$$2^{-2R_2+o(1)} \geq 2^{-\frac{2}{n}I(X_1^n,X_2^n;Y_2^n)} \times \sup_{V:Y_1^n\to Y_0^n\to V}\left\{\alpha^2 2^{2R_1-\frac{2}{n}I(Y_0^n;V|Y_1^n)}+(1-\alpha^2)2^{\frac{2}{n}I(Y_1^n;V)}\right\},$$

for some $P_{X_1^n X_2^n} = P_{X_1^n}P_{X_2^n}$ satisfying $\mathbb{E}[\|X_i^n\|^2] \leq nP_i$, $i=1,2$.

Proof. The proof is a direct consequence of Fano's inequality and Theorem 2 applied to any V such that $Y_1^n \to Y_2^n \to V|X_2^n$. Since $I(Y_2^n;V|X_2^n) = I(Y_0^n;V,X_2^n)$ and $I(Y_1^n;V|X_2^n) = I(Y_1^n;V,X_2^n)$, the supremum can be taken over $V\colon Y_1^n \to Y_0^n \to V$ instead of $Y_1^n \to Y_2^n \to V|X_2^n$. □

The Han-Kobayashi achievable region [19], [20] evaluated for Gaussian inputs (without power control) can be expressed as the set of rate pairs (R_1,R_2) satisfying $R_i \leq \frac{1}{2}\log(1+P_i)$, $i=1,2$ and

$$2^{-2R_2} \geq \frac{\alpha^2 P_2 2^{2R_1}}{(P_2+1-\alpha^2)(1+\alpha^2 P_1+P_2)} + \frac{1-\alpha^2}{P_2+1-\alpha^2}. \tag{25}$$

[3]Costa's EPI plays a role in GIC converse results (e.g., [17]). Since we have seen that Theorem 2 subsumes Costa's EPI, the GIC is a natural application.

Interestingly, (25) this takes a similar form to the outer bound of Theorem 7; however, it is known that transmission without power control is suboptimal for the Gaussian Z-interference channel in general [21], [22]. Nevertheless, it may be possible to identify a random variable V in the supremum in Theorem 7, possibly depending on X_2^n, which ultimately improves known bounds.

III. Proof Sketch

Here we sketch the proofs of Theorems 1 and 2, complete details can be found in [23]. For random variables $X,Y \sim P_{XY}$, we write $X|\{Y=y\}$ to denote the random variable X conditional on $\{Y=y\}$. A sequence of random variables $\{X_n, n\geq 1\}$ will be denoted by the shorthand $\{X_n\}$, and convergence of $\{X_n\}$ in distribution to a random variable X_* is written $X_n \xrightarrow{\mathcal{D}} X_*$.

We begin with an optimization problem: For a random variable $X \sim P_X$, let Y be defined via the additive Gaussian noise channel $P_{Y|X}$ given by $Y = \sqrt{\mathsf{snr}}X + Z$, where $Z \sim N(0,1)$, and define the family of functionals

$$\mathsf{s}_\lambda(X,\mathsf{snr}) = -h(X)+\lambda h(Y) + \inf_{V:X\to Y\to V}\Big\{I(Y;V)-\lambda I(X;V)\Big\} \tag{26}$$

parameterized by $\lambda \geq 1$. For $(X,Y,Q) \sim P_{XQ}P_{Y|X}$, define the functional of P_{XQ}

$$\mathsf{s}_\lambda(X,\mathsf{snr}|Q) = -h(X|Q)+\lambda h(Y|Q) + \inf_{V:X\to Y\to V|Q}\Big\{I(Y;V|Q)-\lambda I(X;V|Q)\Big\}. \tag{27}$$

We consider the optimization problem

$$\mathsf{V}_\lambda(\mathsf{snr}) = \inf_{P_{XQ}:\mathbb{E}[X^2]\leq 1}\mathsf{s}_\lambda(X,\mathsf{snr}|Q). \tag{28}$$

In (28), it suffices to consider $Q \in \mathcal{Q}$ with $|\mathcal{Q}| \leq 2$. By Fenchel-Caratheodory-Bunt, this is sufficient to preserve the values of $\mathbb{E}[X^2] = \sum_q p(q)\mathbb{E}[X^2|Q=q]$ and $\mathsf{s}_\lambda(X,\mathsf{snr}|Q) = \sum_q p(q)\mathsf{s}_\lambda(X,\mathsf{snr}|Q=q)$.

The essence of Theorem 1 is the explicit characterization:

Theorem 8.

$$\mathsf{V}_\lambda(\mathsf{snr}) = \begin{cases} \frac{1}{2}\Big[\lambda\log\left(\frac{\lambda 2\pi e}{\lambda-1}\right)-\log\left(\frac{2\pi e}{\lambda-1}\right)+\log(\mathsf{snr})\Big] & \mathsf{snr}\geq\frac{1}{\lambda-1}\\ \frac{1}{2}\Big[\lambda\log\left(2\pi e(1+\mathsf{snr})\right)-\log\left(2\pi e\right)\Big] & \mathsf{snr}\leq\frac{1}{\lambda-1}.\end{cases}$$

The idea behind proving Theorem 8 is that we only need to consider Gaussian random variables in optimization problem (28). Our argument is based on weak convergence and draws inspiration from a technique employed by Geng and Nair for establishing Gaussian optimality via rotational-invariance [24], which has roots in a doubling trick applied successfully in the study of functional inequalities (e.g., [25]–[27]). The critical ingredients can be summarized as follows:

Lemma 1. *There exists a sequence $\{X_n, Q_n\}$ satisfying*

$$\lim_{n\to\infty}\mathsf{s}_\lambda(X_n,\mathsf{snr}|Q_n) = \mathsf{V}_\lambda(\mathsf{snr}) \tag{29}$$

$$\mathbb{E}[X_n^2] \leq 1 \tag{30}$$

and $(X_n,Q_n) \xrightarrow{\mathcal{D}} (X_*,Q_*)$, with $X_*|\{Q_*=q\} \sim N(\mu_q,\sigma_X^2)$ for P_{Q_*}-a.e. q, with $\sigma_X^2 \leq 1$ not depending on q.

Lemma 2. *If $X_n \xrightarrow{\mathcal{D}} X_* \sim N(\mu,\sigma_X^2)$ and $\sup_n\mathbb{E}[X_n^2] < \infty$, then $\liminf_{n\to\infty}\mathsf{s}_\lambda(X_n,\mathsf{snr}) \geq \mathsf{s}_\lambda(X_*,\mathsf{snr})$.*

The proof of Lemma 1 is sketched in Section III-A. The proof of Lemma 2 is more mechanical (though, nontrivial) and is omitted due to space constraint.

With the above Lemmas in hand, the proof of Theorem 8 follows from calculus and the classical EPI. We require the following proposition, which is an easy corollary of the conditional EPI.

Proposition 2. *Let $X \sim N(0,\gamma)$ and $Z \sim N(0,1)$ be independent, and define $Y = \sqrt{\mathsf{snr}}X + Z$. Then for $\lambda \geq 1$,*

$$\inf_{V:X\to Y\to V}\Big(I(Y;V) - \lambda I(X;V)\Big) = \begin{cases} \frac{1}{2}\left[\log\left((\lambda-1)\gamma\,\mathsf{snr}\right) - \lambda\log\left(\frac{\lambda-1}{\lambda}\left(1+\gamma\,\mathsf{snr}\right)\right)\right] & \gamma\,\mathsf{snr} \geq \frac{1}{\lambda-1} \\ 0 & \gamma\,\mathsf{snr} \leq \frac{1}{\lambda-1}. \end{cases}$$

Proof of Theorem 8. Noting that $\mathsf{s}_\lambda(X,\mathsf{snr})$ is invariant to translations of $\mathbb{E}[X]$, it follows from Lemmas 1 and 2 that

$$\mathsf{V}_\lambda(\mathsf{snr}) = \inf_{0\leq\gamma\leq 1}\mathsf{s}_\lambda(X_\gamma,\mathsf{snr}), \qquad \text{where } X_\gamma \sim N(0,\gamma).$$

Recalling the definition of $\mathsf{s}_\lambda(\,\cdot\,,\mathsf{snr})$, Proposition 2 implies

$$\mathsf{s}_\lambda(X_\gamma,\mathsf{snr}) = \begin{cases} \frac{1}{2}\left[\lambda\log\left(\frac{\lambda 2\pi e}{\lambda-1}\right) - \log\left(\frac{2\pi e}{\lambda-1}\right) + \log(\mathsf{snr})\right] & \text{if } \gamma\,\mathsf{snr} \geq \frac{1}{\lambda-1} \\ \frac{1}{2}\left[\lambda\log\left(2\pi e\left(1+\gamma\,\mathsf{snr}\right)\right) - \log\left(2\pi e\gamma\right)\right] & \text{if } \gamma\,\mathsf{snr} \leq \frac{1}{\lambda-1}. \end{cases}$$

Differentiating with respect to the quantity γ, we find that $\frac{1}{2}\left[\lambda\log\left(2\pi e\left(1+\gamma\,\mathsf{snr}\right)\right) - \log\left(2\pi e\gamma\right)\right]$ is decreasing in γ provided $\gamma\,\mathsf{snr} \leq \frac{1}{\lambda-1}$. Therefore, taking $\gamma = 1$ minimizes $\mathsf{s}_\lambda(X_\gamma,\mathsf{snr})$ over the interval $\gamma \in [0,1]$, proving the claim. □

Given the explicit characterization of $\mathsf{V}_\lambda(\mathsf{snr})$, a dual form of (3), we are now in a position to prove Theorem 1.

Proof of Theorem 1. We first establish (3) under the additional assumption that $\mathbb{E}[X^2] < \infty$. Toward this goal, since mutual information is invariant to scaling, it is sufficient to prove that, for $Y = \sqrt{\mathsf{snr}}X + Z$ with $\mathbb{E}[X^2] \leq 1$ and $Z \sim N(0,1)$ independent of X, we have

$$2^{2(h(Y)-I(X;V))} \geq \mathsf{snr}\,2^{2(h(X)-I(Y;V))} + 2^{2h(Z)} \tag{31}$$

for V satisfying $X \to Y \to V$. Multiplying both sides by σ^2 and choosing $\mathsf{snr} := \frac{\mathrm{Var}(X)}{\sigma^2}$ gives the desired inequality (3) when $\mathbb{E}[X^2] < \infty$. Thus, to prove (31), observe by definition of $\mathsf{V}_\lambda(\mathsf{snr})$ that

$$-h(X) + I(Y;V) \geq \lambda(I(X;V) - h(Y)) + \mathsf{V}_\lambda(\mathsf{snr}). \tag{32}$$

Minimizing the RHS over λ proves the inequality. In particular, elementary calculus shows that the RHS of (32) is minimized when λ satisfies $\frac{\lambda}{\lambda-1} = \frac{1}{2\pi e}2^{-2(I(X;V)-h(Y))}$. Substituting into (32) and recalling $2^{2h(Z)} = 2\pi e$ proves (31).

The assumption that $\mathbb{E}[X^2] < \infty$ can be eliminated via a truncation argument. See [23]. □

A. Proof of Lemma 1

We now sketch the proof of Lemma 1, the most significant technical ingredient needed to establish Theorem 8. We begin with two lemmas, stated without proof (see [23] for details). The first is a superadditive property enjoyed by $\mathsf{s}_\lambda(X,\mathsf{snr}|Q)$ on doubling, and the second is a characterization of the normal distribution in the context of weak convergence.

Lemma 3. *Let $P_{Y|X}$ be the Gaussian channel $Y = \sqrt{\mathsf{snr}}X + Z$, where $Z \sim N(0,1)$ is independent of X. Now, suppose $(X,Y,Q) \sim P_{XQ}P_{Y|X}$, and let (X_1,Y_1,Q_1) and (X_2,Y_2,Q_2) denote two independent copies of (X,Y,Q). Define*

$$X_+ = \frac{1}{\sqrt{2}}\left(X_1 + X_2\right) \qquad X_- = \frac{1}{\sqrt{2}}\left(X_1 - X_2\right), \tag{33}$$

and in a similar manner, define Y_+, Y_-. Letting $\mathbf{Q} = (Q_1,Q_2)$, we have for $\lambda \geq 1$

$$2\mathsf{s}_\lambda(X,\mathsf{snr}|Q) \geq \mathsf{s}_\lambda(X_+,\mathsf{snr}|X_-,\mathbf{Q}) + \mathsf{s}_\lambda(X_-,\mathsf{snr}|Y_+,\mathbf{Q})$$
$$2\mathsf{s}_\lambda(X,\mathsf{snr}|Q) \geq \mathsf{s}_\lambda(X_+,\mathsf{snr}|Y_-,\mathbf{Q}) + \mathsf{s}_\lambda(X_-,\mathsf{snr}|X_+,\mathbf{Q}).$$

Lemma 4. *Suppose $(X_{1,n},X_{2,n}) \xrightarrow{\mathcal{D}} (X_{1,*},X_{2,*})$ with $\sup_n \mathbb{E}[X_{i,n}^2] < \infty$ for $i = 1,2$. Let $(Z_1,Z_2) \sim N(0,\sigma^2 I)$ be pairwise independent of $(X_{1,n},X_{2,n})$ and, for $i = 1,2$, define $Y_{i,n} = X_{i,n} + Z_i$. If $X_{1,n}, X_{2,n}$ are independent and*

$$\liminf_{n\to\infty} I(X_{1,n} + X_{2,n}; X_{1,n} - X_{2,n}|Y_{1,n},Y_{2,n}) = 0, \tag{34}$$

then $X_{1,*}, X_{2,*}$ are independent Gaussian random variables with identical variances.

The proof of Lemma 4 is omitted due to space constraint. However, we point out that it is similar in spirit to a famous result of Bernstein: If X_1, X_2 are independent random variables such that $X_1 + X_2$ and $X_1 - X_2$ are independent, then X_1 and X_2 are normal, with identical variances.

Proof of Lemma 1. For convenience, we will refer to any sequence $\{X_n,Q_n\}$ satisfying (29)-(30) as *admissible*. Since $\mathsf{s}_\lambda(X_n,\mathsf{snr}|Q_n)$ is invariant to translations of the mean of X_n, we may restrict our attention to admissible sequences satisfying $\mathbb{E}[X_n] = 0$ without any loss of generality.

Begin by letting $\{X_n,Q_n\}$ be an admissible sequence with the property that

$$\lim_{n\to\infty}\left(h(Y_n|Q_n) - h(X_n|Q_n)\right) \leq \liminf_{n\to\infty}\left(h(Y_n'|Q_n') - h(X_n'|Q_n')\right) \tag{35}$$

for any other admissible sequence $\{X_n',Q_n'\}$. Such a sequence can always be constructed by a diagonalization argument, and therefore exists. Moreover, we can show that the LHS of (35) is finite due to $\mathsf{V}_\lambda(\mathsf{snr}) < \infty$ and conditioning reduces entropy.

By the same logic as in the remark following (28), we may assume that $Q_n \in \mathcal{Q}$, where $|\mathcal{Q}| = 3$, since this is sufficient to preserve the values of $\mathbb{E}[X_n^2]$, $\mathsf{s}_\lambda(X_n,\mathsf{snr}|Q_n)$ and $(h(Y_n|Q_n) - h(X_n|Q_n))$. Thus, since $\mathcal{Q}$ is finite and $\mathbb{E}[X_n^2] \leq 1$, the sequence $\{X_n,Q_n\}$ is tight. By Prokhorov's theorem, we may assume that there is some (X_*,Q_*) for which $(X_n,Q_n) \xrightarrow{\mathcal{D}} (X_*,Q_*)$ by restricting our attention to a subsequence of $\{X_n,Q_n\}$ if necessary. Moreover, $\mathbb{E}[X_*^2] \leq \liminf_{n\to\infty}\mathbb{E}[X_n^2] \leq 1$ by Fatou's lemma.

Next, for a given n, let $(X_{1,n},Q_{1,n})$ and $(X_{2,n},Q_{2,n})$ denote two independent copies of (X_n,Q_n). Define

$$X_{+,n} = \frac{1}{\sqrt{2}}(X_{1,n} + X_{2,n}) \quad X_{-,n} = \frac{1}{\sqrt{2}}(X_{1,n} - X_{2,n}),$$

In a similar manner, define $Y_{+,n}, Y_{-,n}$, and put $\mathbf{Q}_n = (Q_{1,n},Q_{2,n})$. Applying Lemma 3, we obtain

$$2\mathsf{s}_\lambda(X_n,\mathsf{snr}|Q_n) \geq \mathsf{s}_\lambda(X_{+,n},\mathsf{snr}|X_{-,n}\mathbf{Q}_n) + \mathsf{s}_\lambda(X_{-,n},\mathsf{snr}|Y_{+,n}\mathbf{Q}_n), \tag{36}$$

and the symmetric inequality

$$2\mathsf{s}_\lambda(X_n,\mathsf{snr}|Q_n) \geq \mathsf{s}_\lambda(X_{+,n},\mathsf{snr}|Y_{-,n}\mathbf{Q}_n) + \mathsf{s}_\lambda(X_{-,n},\mathsf{snr}|X_{+,n}\mathbf{Q}_n). \tag{37}$$

By independence of $X_{1,n}$ and $X_{2,n}$ and the assumption that $\mathbb{E}[X_n] = 0$, we have

$$\mathbb{E}[X_{+,n}^2] = \mathbb{E}[X_{-,n}^2] = \frac{1}{2}\mathbb{E}[X_{1,n}^2] + \frac{1}{2}\mathbb{E}[X_{2,n}^2] = \mathbb{E}[X_n^2] \le 1.$$

Hence, it follows that the terms in the RHS of (36) and the RHS of (37) are each lower bounded by $\mathsf{V}_\lambda(\mathsf{snr})$. Since $\lim_{n\to\infty} \mathsf{s}_\lambda(X_n, \mathsf{snr}|Q_n) = \mathsf{V}_\lambda(\mathsf{snr})$ by definition, we find that

$$\mathsf{V}_\lambda(\mathsf{snr}) = \lim_{n\to\infty} \frac{1}{2}\Big(\mathsf{s}_\lambda(X_{+,n}, \mathsf{snr}|Y_{-,n}\mathbf{Q}_n) + \mathsf{s}_\lambda(X_{-,n}, \mathsf{snr}|Y_{+,n}\mathbf{Q}_n)\Big).$$

In particular, by letting the random pair (X_n', Q_n') correspond to equal time-sharing between the pairs $(X_{+,n}, (Y_{-,n}\mathbf{Q}_n))$ and $(X_{-,n}, (Y_{+,n}\mathbf{Q}_n))$, we have constructed an admissible sequence $\{X_n', Q_n'\}$ which satisfies

$$\lim_{n\to\infty} \mathsf{s}_\lambda(X_n', \mathsf{snr}|Q_n') = \mathsf{V}_\lambda(\mathsf{snr}). \tag{38}$$

The following identity can be shown by standard manipuation

$$h(Y_n|Q_n) - h(X_n|Q_n) = h(Y_n'|Q_n') - h(X_n'|Q_n') + \frac{1}{2} I(X_{+,n}; X_{-,n}|Y_{+,n}, Y_{-,n}, \mathbf{Q}_n). \tag{39}$$

Since the sequence $\{X_n', Q_n'\}$ is admissible, it must also satisfy (35). Therefore, in view of (39) and the fact that the LHS of (35) is finite, this implies that

$$\liminf_{n\to\infty} I(X_{1,n} + X_{2,n}; X_{1,n} - X_{2,n}|Y_{1,n}, Y_{2,n}, \mathbf{Q}_n) = 0.$$

This completes the proof since an application of Lemma 4 guarantees that, for P_{Q_*}-a.e. q, the random variable $X_*|\{Q_* = q\}$ is normal with variance not depending on q, and moreover we have already observed that $\mathbb{E}[X_*^2] \le 1$, so the variance of $X_*|\{Q_* = q\}$ is at most unity as claimed. □

B. Extension to Random Vectors

The vector generalization of Shannon's EPI is proved by a combination of conditioning, Jensen's inequality and induction. The same argument does not appear to readily apply in generalizing Theorem 1 to its vector version due to complications arising from the Markov constraint $\mathbf{X} \to (\mathbf{X}+\mathbf{W}) \to V$. However, the desired generalization may be established by noting an additivity property enjoyed by the dual form.

For a random vector $\mathbf{X} \sim P_{\mathbf{X}}$, let $\mathbf{Y}$ be defined via the additive Gaussian noise channel $\mathbf{Y} = \Gamma^{1/2}\mathbf{X} + \mathbf{Z}$, where $\mathbf{Z} \sim N(0, I)$ is independent of $\mathbf{X}$ and Γ is a diagonal matrix with nonnegative diagonal entries. Analogous to the scalar case, define for $\lambda \ge 1$

$$\mathsf{s}_\lambda(\mathbf{X}, \Gamma|Q) = -h(\mathbf{X}|Q) + \lambda h(\mathbf{Y}|Q) + \inf_{V:\mathbf{X}\to\mathbf{Y}\to V|Q} \Big\{ I(\mathbf{Y}; V|Q) - \lambda I(\mathbf{X}; V|Q) \Big\}, \tag{40}$$

and consider the optimization problem

$$\mathsf{V}_\lambda(\Gamma) = \inf_{P_{\mathbf{X}Q}\,:\,\mathbb{E}[X_i^2]\le 1, i\in[n]} \mathsf{s}_\lambda(\mathbf{X}, \Gamma|Q). \tag{41}$$

Theorem 9. *If* $\Gamma = \mathrm{diag}(\mathsf{snr}_1, \mathsf{snr}_2, \ldots, \mathsf{snr}_n)$*, then*

$$\mathsf{V}_\lambda(\Gamma) = \sum_{i=1}^n \mathsf{V}_\lambda(\mathsf{snr}_i).$$

Proof. Let Γ be a block diagonal matrix with blocks given by $\Gamma = \mathrm{diag}(\Gamma_1, \Gamma_2)$. Partition $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$ such that $\mathbf{Y}_i = \Gamma_i^{1/2}\mathbf{X}_i + \mathbf{Z}_i$ for $i = 1, 2$. Then, for any V such that $\mathbf{X} \to \mathbf{Y} \to V|Q$, it is shown in [23] that

$$\mathsf{s}_\lambda(\mathbf{X}, \Gamma|Q) \ge \mathsf{s}_\lambda(\mathbf{X}_1, \Gamma_1|\mathbf{X}_2, Q) + \mathsf{s}_\lambda(\mathbf{X}_2, \Gamma_2|\mathbf{Y}_1, Q). \tag{42}$$

Hence, $\mathsf{V}_\lambda(\Gamma) \ge \sum_{i=1}^2 \mathsf{V}_\lambda(\Gamma_i)$. Induction proves the claim. □

To finish, the proof of Theorem 2 follows similarly to that of Theorem 1, but by first whitening $\mathbf{W}$ and employing Theorem 9 in place of Theorem 8.

References

[1] F. du Pin Calmon, Y. Polyanskiy, and Y. Wu, "Strong data processing inequalities in power-constrained Gaussian channels." in *International Symposium on Information Theory*. IEEE, 2015, pp. 2558–2562.

[2] G. Toscani, "A strengthened entropy power inequality for log-concave densities," *Information Theory, IEEE Transactions on*, vol. 61, no. 12, pp. 6550–6559, 2015.

[3] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *Information Theory, IEEE Transactions on*, vol. 53, no. 7, pp. 2317–2329, 2007.

[4] M. Madiman, J. Melbourne, and P. Xu, "Forward and reverse entropy power inequalities in convex geometry," *arXiv preprint arXiv:1604.04225*, 2016.

[5] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Transactions on Inf. Theory*, vol. 57, no. 1, pp. 33–55, Jan 2011.

[6] M. Costa, "A new entropy power inequality," *IEEE Transactions on Inf. Theory*, vol. 31, no. 6, pp. 751–760, Nov 1985.

[7] R. Liu, T. Liu, H. Poor, and S. Shamai, "A vector generalization of Costa's entropy-power inequality with applications," *IEEE Transactions on Inf. Theory*, vol. 56, no. 4, pp. 1865–1879, April 2010.

[8] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Transactions on Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[9] S. Bobkov and M. Madiman, "Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures," *Journal of Functional Analysis*, vol. 262, no. 7, pp. 3309–3339, 2012.

[10] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959.

[11] L. Gross, "Logarithmic Sobolev inequalities," *American Journal of Mathematics*, vol. 97, no. 4, pp. 1061–1083, 1975.

[12] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Inf. Theory*, vol. 43, no. 6, pp. 1912–1923, Nov 1997.

[13] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic Gaussian CEO problem," in *International Symposium on Information Theory*, June 2004, pp. 119–124.

[14] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Transactions on Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, July 2005.

[15] J. Wang, J. Chen, and X. Wu, "On the sum rate of Gaussian multiterminal source coding: New proofs and results," *IEEE Transactions on Inf. Theory*, vol. 56, no. 8, pp. 3946–3960, 2010.

[16] T. Courtade and J. Jiao, "An extremal inequality for long Markov chains," in *52nd Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2014, pp. 763–770.

[17] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *arXiv preprint arXiv:1504.04419*, 2015.

[18] M. Costa, "On the Gaussian interference channel," *IEEE Transactions on Inf. Theory*, vol. 31, no. 5, pp. 607–615, Sep 1985.

[19] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.

[20] H. Te Sun and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE transactions on information theory*, vol. 27, no. 1, pp. 49–60, 1981.

[21] M. H. Costa, "Noisebergs in Z-Gaussian interference channels," in *Information Theory and Applications Workshop (ITA)*. IEEE, 2011, pp. 1–6.

[22] C. Nair and M. H. Costa, "Gaussian Z-interference channel: Around the corner," in *Information Theory and Applications Workshop (ITA), 2016*. IEEE, 2016.

[23] T. A. Courtade, "Strengthening the entropy power inequality," *arXiv preprint arXiv:1602.03033*, 2016.

[24] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IEEE Transactions on Inf. Theory*, vol. 60, no. 4, pp. 2087–2104, April 2014.

[25] E. H. Lieb, "Gaussian kernels have only Gaussian maximizers," *Inventiones mathematicae*, vol. 102, no. 1, pp. 179–208, 1990.

[26] E. A. Carlen, "Superadditivity of Fisher's information and logarithmic Sobolev inequalities," *Journal of Functional Analysis*, vol. 101, no. 1, pp. 194–211, 1991.

[27] F. Barthe, "Optimal Young's inequality and its converse: a simple proof," *Geometric & Functional Analysis GAFA*, vol. 8, no. 2, pp. 234–242, 1998.