Concavity of Entropy Power: Equivalent Formulations and Generalizations

Thomas A. Courtade
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

***Abstract*—We show that Costa's entropy power inequality, when appropriately formulated, can be precisely generalized to non-Gaussian additive perturbations. This reveals fundamental links between the Gaussian logarithmic Sobolev inequality and the convolution inequalities for entropy and Fisher information. Various consequences including a reverse entropy power inequality and information-theoretic central limit theorems are also established.**

I. Introduction

Costa's entropy power inequality (EPI) asserts that entropy power is concave with respect to the variance of additive Gaussian noise [1], and is regarded as one of the classical results in the information theory literature. Since its introduction, Costa's EPI has found applications in many converse proofs and has been generalized in several ways. Two related directions were explored by Payaró and Palomar [2] and Liu, Liu, Poor and Shamai [3], who considered concavity of entropy power with respect to a matrix-valued argument. Cheng and Geng recently investigated the higher-order derivatives of entropy power, showing that they alternate in sign through the fourth derivative [4]. Separately, Villani gave a short proof of Costa's EPI showing that it also holds for heat flow on Riemannian manifolds with nonnegative Ricci curvature [5]. Finally, an equivalent form of Costa's EPI extends to a broad class of other diffusion processes (e.g., [6]), and plays a fundamental role in their analysis. This latter point does not appear to have been explicitly noted in the literature, so we explain the connection following Proposition 1, stated below.

In the present paper, we provide generalizations of Costa's EPI, and demonstrate applications ranging from deficit estimates for functional inequalities to information-theoretic central limit theorems. Before doing so, let us first establish notation. For a random vector X on $\mathbb{R}^d$ with density f, the entropy and Fisher information are defined by[1]

$$h(X) = -\int f \log f, \qquad J(X) = \int f\,|\nabla \log f|^2, \quad (1)$$

provided the integrals exist, with $|\cdot|$ denoting Euclidean length. We adopt the convention that $J(X) = \infty$ if the defining integral does not exist. The entropy power of X is defined according to

$$N(X) = \frac{1}{2\pi e} e^{\frac{2}{d}h(X)}. \quad (2)$$

Entropy power and Fisher information famously satisfy convolution inequalities, which enjoy broad applications ranging from converse proofs in information theory (e.g., [7]) to central limit theorems (e.g., [8]). In his 1948 paper, Shannon proposed the entropy power inequality (EPI): For independent random vectors X, Y,

$$N(X+Y) \geq N(X) + N(Y). \quad (3)$$

Although Shannon's proof was incomplete, the EPI was eventually proved by Stam [9]. In the course of his proof, Stam also established the Fisher information inequality (FII)

$$1/J(X+Y) \geq 1/J(X) + 1/J(Y). \quad (4)$$

Closely related to entropy and Fisher information are their relative counterparts. For a distribution P, absolutely continuous with respect to Q, the relative entropy is defined as

$$D(P\|Q) = \mathbb{E}_P\left[\log \tfrac{\mathrm{d}P}{\mathrm{d}Q}\right], \quad (5)$$

and the relative Fisher information is defined as

$$I(P\|Q) = \mathbb{E}_P\left[\left|\nabla\left(\log \tfrac{\mathrm{d}P}{\mathrm{d}Q}\right)\right|^2\right], \quad (6)$$

provided the expectation exists. If it does not, we again adopt the convention that $I(P\|Q) = \infty$. Note that we have made use of the shorthand notation $\mathbb{E}_P[f] := \mathbb{E}[f(X)]$, where $X \sim P$. When $X \sim P$ and $Y \sim Q$, we also write $D(X\|Y)$ and $I(X\|Y)$ to denote $D(P\|Q)$ and $I(P\|Q)$, respectively.

Throughout, we let $Z \sim N(0, \mathrm{I})$ denote a standard normal random vector; its dimension will be clear from context. Completely equivalent to (3) and (4), respectively, are the inequalities

$$tD(X\|Z) + \bar{t}\,D(Y\|Z) \geq D(\sqrt{t}X + \sqrt{\bar{t}}Y\|Z) \quad (7)$$
$$tI(X\|Z) + \bar{t}\,I(Y\|Z) \geq I(\sqrt{t}X + \sqrt{\bar{t}}Y\|Z), \quad (8)$$

where X, Y are independent zero-mean random vectors with finite second moments, $t \in [0,1]$ and $\bar{t} := (1-t)$. Given their equivalence, we shall refer to inequalities (3) and (7) collectively as the entropy power inequality (EPI), and inequalities (4) and (8) collectively as the Fisher information inequality (FII).

For X independent of $Z \sim N(0, \mathrm{I})$, we write $X + \sqrt{t}Z$ to denote the usual Gaussian perturbation of X. Unless otherwise specified, we similarly let $X_t := e^{-t}X + (1-e^{-2t})^{1/2}Z$ denote the Ornstein-Uhlenbeck process at time t, which starts with distribution X. Fisher information and entropy are intrinsically linked along these (equivalent) processes via de Bruijn's identity, which states $\frac{\mathrm{d}}{\mathrm{d}t}h(X+\sqrt{t}Z) = \frac{1}{2}J(X+\sqrt{t}Z)$. This can be equivalently stated in terms of the Ornstein-Uhlenbeck process as $\frac{\mathrm{d}}{\mathrm{d}t}D(X_t\|Z) = -I(X_t\|Z)$.

This work was supported in part by NSF grants CCF-1528132 and CCF-0939370.

[1] Throughout, we adopt the convention that all logarithms have base e

978-1-5090-4096-4/17/$31.00 ©2017 IEEE

With basic notation established, we now establish equivalent formulations of Costa's EPI. A proof is given in the appendix.

Proposition 1. *The following are equivalent:*

1) *(Costa's EPI) For any $0 \le t \le 1$ and any random vector X having density on $\mathbb{R}^d$ with $\mathbb{E}|X|^2 < \infty$,*
$$N(X + \sqrt{t}Z) \ge (1-t)N(X) + tN(X+Z). \quad (9)$$
2) *For any $t \ge 0$ and any random vector X having density on $\mathbb{R}^d$ with $\mathbb{E}|X|^2 < \infty$,*
$$N(X + \sqrt{t}Z) \le N(X)\left(1 + \frac{t}{d}J(X)\right). \quad (10)$$
3) *For any $t \ge 0$ and any random vector X having density on $\mathbb{R}^d$,*
$$D(X\|Z) \le \frac{1-e^{-2t}}{2} I(X\|Z) + D(X_t\|Z). \quad (11)$$

Costa's original proof of (9) was accomplished by explicitly establishing the inequality $\frac{\mathrm{d}^2}{\mathrm{d}t^2}N(X+\sqrt{t}Z) \le 0$. In light of the equivalences presented in Proposition 1, a proof of Costa's EPI is immediate from the convolution inequality for Fisher information by integrating de Bruijn's identity[2]. Indeed,

$$D(X\|Z) - D(X_t\|Z) = \int_0^t I(X_s\|Z)\,\mathrm{d}s \le \int_0^t e^{-2s} I(X\|Z)\,\mathrm{d}s = \frac{1-e^{-2t}}{2} I(X\|Z).$$

This argument reveals how Costa's EPI may be interpreted as a precise interpolation between a tautology and the EPI. Indeed, letting $t \downarrow 0$ yields the trivially sharp inequality $0 \le 0$. On the other hand, letting $t \to \infty$ yields the Gaussian logarithmic Sobolev inequality (LSI) $D(X\|Z) \le \frac{1}{2}I(X\|Z)$, which is completely equivalent to the special case of (7) when Y is standard normal (e.g., [6]); i.e., $D(X_t\|Z) \le e^{-2t}D(X\|Z)$.

Reformulation of Costa's EPI as (11) provides a point of departure that leads to a rich generalization in the context of Markov processes. In particular, let $e^{-\varphi}$ be a probability density, where φ is smooth and satisfies $\nabla^2\varphi \ge \rho\,\mathrm{Id}$ for some $\rho \in \mathbb{R}$. Now, let $\mathbf{P} = (P_t)_{t\ge 0}$ denote the Markov semigroup with generator $L = \Delta - \nabla\varphi \cdot \nabla$. If $(X_t)_{t\ge 0}$ is the associated diffusion process with initial distribution $X_0 \sim X$, then a similar argument as above gives

$$D(X\|V) \le \frac{1-e^{-2\rho t}}{2\rho} I(X\|V) + D(X_t\|V), \quad (12)$$

where V has density $e^{-\varphi}$, corresponding to the stationary distribution for this process. In the special case where $(X_t)_{t\ge0}$ is the standard Ornstein-Uhlenbeck process, then $\rho = 1$ and $V \sim N(0, \mathrm{I})$, so that (12) coincides precisely with (11). Interested readers are referred to [6] for a full treatment.

Organization: Proposition 1 is generalized to non-Gaussian perturbations in Section II. Several applications are presented in Section III. These include links between the EPI, FII and LSI as well as connections between the LSI and convergence in information-theoretic central limit theorems on short time-scales. We close with two problems for future work in Section IV.

[2]Dembo [10] gave another proof of Costa's EPI using the FII which is different from that given here. We are not aware of any place the present proof appears in the literature (in particular, the content of Proposition 1), but suspect it may be known to some, even if unpublished. Also, Costa's EPI in the form of (10) was rediscovered in [11].

II. Main Results

We will establish that the three equivalent forms of concavity of entropy power in Proposition 1 may be suitably generalized to non-Gaussian additive perturbations. The starting point of our investigation is the following inequality for entropy powers, established by the author in [12]:

Theorem 1. *Let X, Y, W be independent random vectors on $\mathbb{R}^d$, with W being Gaussian. Then*
$$N(X+W)N(Y+W) \ge N(X)N(Y) + N(X+Y+W)N(W). \quad (13)$$

We will see shortly that this inequality extends (9) in a way that subsumes the results we have discussed thus far. As a first demonstration, it was remarked in [12] that (13) immediately implies the vector generalization of Costa's EPI due to Liu, Liu, Poor and Shamai by taking Y, W to be appropriately scaled Gaussians and simplifying:

Corollary 1. *[3] Let X and $G \sim N(0,\Sigma)$ be independent random vectors on $\mathbb{R}^d$, with Σ positive definite. For positive semidefinite $A \preceq \mathrm{I}$ commuting[3] with Σ,*
$$N(X + A^{1/2}G) \ge |\mathrm{I} - A|^{1/d}N(X) + |A|^{1/d}N(X+G).$$

From the above example, we see Theorem 1 evidently generalizes Costa's EPI (9) and its vector extension. The second and third equivalent formulations of Proposition 1 can be similarly extended. Specifically,

Theorem 2. *The following are true and equivalent:*

1) *For independent random vectors X, Y having density on $\mathbb{R}^d$ with finite second moments,*
$$dN(X+Y) \le N(X)N(Y)\left(J(X) + J(Y)\right). \quad (14)$$
2) *For independent centered random vectors X, Y having density on $\mathbb{R}^d$, and all $0 \le t \le 1$,*
$$D(X\|Z) + D(Y\|Z) \le \frac{\bar{t}}{2}I(X\|Z) + \frac{t}{2}I(Y\|Z) + D(\sqrt{t}X + \sqrt{\bar{t}}Y\|Z). \quad (15)$$

Specializing (14) by taking $Y = \sqrt{t}Z$ recovers (10). Similarly, letting Y be standard normal in (15) recovers (11). Taken together with the fact that Y need not be Gaussian in Theorem 1, we see that Costa's EPI (in each of its equivalent forms) can be suitably generalized to non-Gaussian additive perturbations.

Unlike the three equivalent formulations of Costa's EPI in Proposition 1, the inequality of Theorem 1 appears stronger than the two equivalent forms presented in Theorem 2. Indeed, (14) is equivalent to, for any $s \ge 0$,

$$N(X+Y+\sqrt{2s}Z) \le \frac{\mathrm{d}}{\mathrm{d}s}\Big(N(X+\sqrt{s}Z)N(Y+\sqrt{s}Z)\Big).$$

[3]The commutativity hypothesis was not stated in [3], but the inequality can fail without it [13].

Integrating both sides over $s \in [0,t]$ gives

$$N(\sqrt{t}Z)\frac{1}{t}\int_0^t N(X+Y+\sqrt{2s}Z)\,\mathrm{d}s \leq N(X+\sqrt{t}Z)N(Y+\sqrt{t}Z) - N(X)N(Y). \quad (16)$$

This would be capable of recovering (1) if it were true that $\frac{1}{t}\int_0^t N(X+Y+\sqrt{2s}Z)\,\mathrm{d}s \geq N(X+Y+\sqrt{t}Z)$. However, this inequality actually goes in the reverse direction due to Jensen's inequality and concavity of entropy power.

Proof of Theorem 2. We first establish (14) as a consequence of Theorem 1. To this end, let $W = \sqrt{t}Z$ in (13), so that (13) particularizes to

$$\frac{N(X+\sqrt{t}Z)N(Y+\sqrt{t}Z) - N(X)N(Y)}{t} \geq N(X+Y+\sqrt{t}Z) \geq N(X+Y).$$

Letting $t \downarrow 0$, an application of de Bruijn's identity and the chain rule for derivatives yields (14). Note that in applying de Bruijn's identity, we may assume the Fisher informations are finite, else (14) is trivially true.

We now prove equivalence of (14) and (15), which is inspired by Carlen's proof of equivalence between Stam's inequality and Gross' logarithmic Sobolev inequality [14] (see also Raginsky and Sason [15]). Before doing so, we recall the scaling properties $N(tX) = t^2 N(X)$ and $t^2 J(tX) = J(X)$. Also, if $G_s \sim N(0, s\mathrm{I})$, the relative entropy $D(X\|G_s)$ and Fisher information $I(X\|G_s)$ are related to $h(X)$ and $J(X)$ via

$$h(X) - \frac{d}{2}\log(2\pi e s) = -D(X\|G_s) + \frac{1}{2s}\mathbb{E}|X|^2 - \frac{d}{2} \quad (17)$$

$$J(X) = I(X\|G_s) + \frac{2}{s}d - \frac{1}{s^2}\mathbb{E}|X|^2, \quad (18)$$

holding for any random vector X on $\mathbb{R}^d$ with $\mathbb{E}|X|^2 < \infty$. Further, $N(\cdot), J(\cdot)$ are translation invariant, so we may assume without loss of generality that all random vectors are centered.

• Proof of (14) $\Rightarrow$ (15): We assume $I(X\|Z), I(Y\|Z) < \infty$, else (15) is a tautology. Now, finiteness of $I(X\|Z)$ implies $\mathbb{E}|X|^2 < \infty$, and similarly for Y (see, e.g., [16, Proof of Thm. 5]). Using the scaling properties of $N(\cdot)$ and $J(\cdot)$, (14) implies

$$N\left(\sqrt{t}X + \sqrt{\bar{t}}Y\right) \leq N(X)N(Y)\left(\frac{\bar{t}\,J(X) + t\,J(Y)}{d}\right),$$

Now, taking logarithms, multiplying through by $d/2$ and recalling $\log x \leq x - 1$, we have:

$$\begin{aligned} &h\left(\sqrt{t}X + \sqrt{\bar{t}}Y\right) - \frac{d}{2}\log(2\pi e) \\ &\leq h(X) - \frac{d}{2}\log(2\pi e) + h(Y) - \frac{d}{2}\log(2\pi e) \quad (19) \\ &\quad + \frac{d}{2}\log\left(\frac{\bar{t}\,J(X) + t\,J(Y)}{d}\right) \\ &\leq h(X) - \frac{d}{2}\log(2\pi e) + h(Y) - \frac{d}{2}\log(2\pi e) \quad (20) \\ &\quad + \frac{1}{2}\left(\bar{t}\,J(X) + t\,J(Y)\right) - \frac{d}{2}. \end{aligned}$$

Now, (15) follows from the identities (17) and (18) for $s = 1$.

• Proof of (15) $\Rightarrow$ (14): We may assume X, Y have finite Fisher information and second moments. With this assumption in place, consider any $s > 0$ and observe via rescaling using the identities (17)-(18) that (15) is equivalent to

$$D(X\|G_s) + D(Y\|G_s) \leq s\left(\frac{\bar{t}}{2}I(X\|G_s) + \frac{t}{2}I(Y\|G_s)\right) + D\left(\sqrt{t}X + \sqrt{\bar{t}}Y\middle\|G_s\right). \quad (21)$$

Hence, using (17)-(18) again and rearranging, we find that this is the same as

$$\log N\left(\sqrt{t}X + \sqrt{\bar{t}}Y\right) \leq \log N(X)N(Y) + s\frac{\bar{t}\,J(X) + t\,J(Y)}{d} - \log s - 1.$$

Recalling $1 + \log a = \inf_{s>0}(as - \log s)$, we may minimize the RHS over $s > 0$ to obtain

$$\begin{aligned} dN\left(\sqrt{t}X + \sqrt{\bar{t}}Y\right) &\leq N(X)N(Y)\left(\bar{t}\,J(X) + t\,J(Y)\right) \\ &= N(\sqrt{t}X)N(\sqrt{\bar{t}}Y)\left(J(\sqrt{t}X) + J(\sqrt{\bar{t}}Y)\right), \end{aligned}$$

where the last equality follows via the scaling properties of $N(\cdot)$ and $J(\cdot)$. A simple rescaling recovers (14) and completes the proof. $\square$

III. Consequences

In this section, we explore several consequences of Theorem 2. Further examples may be found in [17].

A. Links between deficits in information inequalities

There has been a growing literature on quantitative deficit estimates for functional inequalities (see, e.g., [18], [19]). Here, we show that the deficits of the LSI, EPI and FII — three of the most fundamental functional inequalities in information theory — are intrinsically linked to one another.

For independent centered random vectors X, Y on $\mathbb{R}^d$, define the following (non-negative) quantities, which capture associated deficits in the LSI, EPI and FII respectively:

$$\begin{aligned} \delta_{\mathsf{LSI}}(X) &= \frac{1}{2}I(X\|Z) - D(X\|Z) \\ \delta_{\mathsf{EPI},t}(X,Y) &= tD(X\|Z) + \bar{t}\,D(Y\|Z) - D(\sqrt{t}X + \sqrt{\bar{t}}Y\|Z) \\ \delta_{\mathsf{FII},t}(X,Y) &= tI(X\|Z) + \bar{t}\,I(Y\|Z) - I(\sqrt{t}X + \sqrt{\bar{t}}Y\|Z). \end{aligned}$$

With these definitions in hand, (15) may be concisely rewritten as

$$\delta_{\mathsf{EPI},t}(X,Y) \leq \bar{t}\,\delta_{\mathsf{LSI}}(X) + t\delta_{\mathsf{LSI}}(Y), \quad (22)$$

holding for any centered random vectors X, Y. Definitions yield the following equivalent form:

Theorem 3. *For centered random vectors X, Y,*

$$\delta_{\mathsf{LSI}}(\sqrt{t}X + \sqrt{\bar{t}}Y) + \frac{1}{2}\delta_{\mathsf{FII},t}(X,Y) \leq \delta_{\mathsf{LSI}}(X) + \delta_{\mathsf{LSI}}(Y).$$

Two interesting consequences follow. The first is a convolution inequality for the deficit in the LSI which may be of independent interest. Indeed, since $\delta_{\mathsf{FII},t}(X,Y) \geq 0$, it follows immediately that

$$\delta_{\mathsf{LSI}}(\sqrt{t}X + \sqrt{\bar{t}}Y) \leq \delta_{\mathsf{LSI}}(X) + \delta_{\mathsf{LSI}}(Y), \quad (23)$$

showing that not only are relative entropy and Fisher information well-behaved under convolution of densities, but so is their difference (as captured by δ_{LSI}).

Second, when expressed in the form of Theorem 3, we observe that our estimates are essentially best-possible. Indeed, from Theorem 3, it follows that for any $t \in [0,1]$

$$\begin{aligned}&\delta_{\mathsf{LSI}}(X) + \delta_{\mathsf{LSI}}(Y)\\ &\geq \frac{1}{2}\left(\delta_{\mathsf{LSI}}(\sqrt{t}X + \sqrt{\bar{t}}Y) + \frac{1}{2}\delta_{\mathsf{FII},t}(X,Y)\right)\\ &\quad + \frac{1}{2}\left(\delta_{\mathsf{LSI}}(\sqrt{\bar{t}}X + \sqrt{t}Y) + \frac{1}{2}\delta_{\mathsf{FII},\bar{t}}(X,Y)\right). \end{aligned} \tag{24}$$

However, definitions and the fact that $\delta_{\mathsf{EPI},t} \geq 0$ imply

$$\begin{aligned}\delta_{\mathsf{LSI}}(X) + \delta_{\mathsf{LSI}}(Y) &\leq \delta_{\mathsf{LSI}}(\sqrt{t}X + \sqrt{\bar{t}}Y) + \frac{1}{2}\delta_{\mathsf{FII},t}(X,Y)\\ &\quad + \delta_{\mathsf{LSI}}(\sqrt{\bar{t}}X + \sqrt{t}Y) + \frac{1}{2}\delta_{\mathsf{FII},\bar{t}}(X,Y),\end{aligned}$$

establishing that the lower bound (24) derived from Theorem 3 is tight within constant factor of 2 for all $t \in [0,1]$.

B. Reverse Entropy Power and Fisher Information Inequalities

Due to its fundamental role in information theory, there has been sustained interest in obtaining reverse forms of the entropy power inequality (3). As shown by Bobkov and Chistyakov [20], the EPI cannot be reversed in general, at least not up to a constant factor. Nevertheless, progress has been made. A notable example of a reverse EPI is due to Bobkov and Madiman [21], who show that for independent random vectors X, Y with log-concave densities, there exist linear volume preserving maps u, v such that

$$N(u(X) + v(Y)) \leq C(N(X) + N(Y)), \tag{25}$$

where C is an absolute constant. Bobkov and Madiman's result mirrors Milman's reverse Brunn-Minkowski inequality, which is pleasant since the EPI itself mirrors the Brunn-Minkowski inequality. A similar statement holds for a more general class of convex measures. Another example of a reverse EPI is due to Ball, Nayar and Tkocz [22], who also restrict attention to the class of log-concave densities and show the EPI can be reversed in this setting if all terms are exponentiated by a constant factor. See also the recent survey by Madiman, Melbourne and Xu [23] for related results.

An immediate implication of Theorem 2 is the following reverse EPI, noted previously by the author in [12]:

Theorem 4. *Let X, Y be independent random vectors on $\mathbb{R}^d$ with finite second moment, and choose λ to satisfy $\lambda/(1-\lambda) = N(Y)/N(X)$. Then*

$$N(X+Y) \leq (N(X) + N(Y))\left(\lambda \mathsf{p}(X) + (1-\lambda)\mathsf{p}(Y)\right),$$

where $\mathsf{p}(X) := \frac{1}{d}N(X)J(X)$.

Stam [9] observed that $\mathsf{p}(X) \geq 1$ (now referred to as *Stam's inequality*), with equality iff $X \sim N(0, \sigma^2 \mathrm{I})$ for $\sigma^2 > 0$. As a consequence of Theorem 2, if both X and Y each nearly saturate Stam's inequality, then the EPI (and also FII) will be nearly saturated. A remarkable aspect of Theorem 4 in comparison to the above reverse EPIs is that no regularity assumptions (e.g., log-concave densities) are imposed.

Finally, since Theorems 3 and 4 are equivalent in light of Theorem 2, this reverse EPI is essentially best possible in the context of the discussion of the previous section.

C. Short-term convergence in information-theoretic CLTs

Let X be a centered random vector on $\mathbb{R}^d$ with $\mathrm{Cov}(X) = \mathrm{I}$, and define the normalized sums $U_n = \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$, where $X_1, X_2, \dots, X_n$ are independent copies of X. The entropic central limit theorem (CLT) due to Barron [24] asserts that $D(U_n\|Z) \to 0$, provided $D(U_{n_0}\|Z) < \infty$ for some n_0. Likewise, the CLT for Fisher information, due to Barron and Johnson [8], asserts that $I(U_n\|Z) \to 0$, provided $I(U_{n_0}\|Z) < \infty$ for some n_0.

Recently, Bobkov et al. [25] have settled a longstanding conjecture and shown that under moment conditions

$$D(U_n\|Z) = O(1/n), \tag{26}$$

which is consistent with the convergence rates predicted by the Berry-Esseen theorem. Although (26) provides good long-term estimates on convergence in the entropic CLT, it does not immediately provide any information about the short-term behavior of $D(U_n\|Z)$.

The next result partially addresses this issue by establishing a dimension-free *lower* bound on $D(U_n\|Z)$ in terms of $\delta_{\mathsf{LSI}}(X)$ and n. Roughly speaking, if $\delta_{\mathsf{LSI}}(X) \ll D(X\|Z)$, then $D(U_n\|Z)$ is assured to decay slowly on short time scales. A similar result holds for Fisher information. That is, if $\delta_{\mathsf{LSI}}(X) \ll I(X\|Z)$, then $I(U_n\|Z)$ will decay slowly on short time scales. Essentially, each of these quantities decay at most linearly in n, with slope $\delta_{\mathsf{LSI}}(X)$.

Theorem 5. *Let X be a centered random vector on $\mathbb{R}^d$ with finite second moments and define the normalized sums $U_n = \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$, where $X_1, X_2, \dots, X_n$ are independent copies of X. The following hold for all $n \geq 1$:*

$$D(U_n\|Z) \geq D(X\|Z) - (n-1)\delta_{\mathsf{LSI}}(X) \tag{27}$$

$$\frac{1}{2}I(U_n\|Z) \geq \frac{1}{2}I(X\|Z) - n\,\delta_{\mathsf{LSI}}(X) + \delta_{\mathsf{LSI}}(U_n). \tag{28}$$

Proof. We apply Theorem 2 identifying $t \leftarrow \frac{n}{n+m}$, $X \leftarrow U_n$ and $Y \leftarrow U_m$. This yields the inequality

$$\begin{aligned}D(U_{n+m}\|Z) &\geq D(U_n\|Z) + D(U_m\|Z)\\ &\quad - \frac{m}{2(m+n)}I(U_n\|Z) - \frac{n}{2(m+n)}I(U_m\|Z).\end{aligned} \tag{29}$$

The proof of (27) and (28) now follows by induction on $n+m$. See [17] for details. □

D. A 'discrete-time' variation on Costa's EPI

Let X be a centered random vector on $\mathbb{R}^d$ with finite second moments, put $X_0 = X$, and for $n \geq 0$ inductively define $X_{n+1} = \frac{1}{\sqrt{2}}(X_n + X_n')$, where X_n, X_n' are independent copies of X_n. Note that there is some resemblance between the discrete time process $\{X_n, n \geq 0\}$ and the continuous time Ornstein-Uhlenbeck process $(X_t)_{t\geq 0}$. Indeed, both sequences tend to Gaussian (the former by the CLT). Additionally, as a function of the time index, both sequences generally converge to the stationary distribution exponentially quickly. This can be made precise in the case of the Ornstein-Uhlenbeck process where $D(X_t\|Z) \leq e^{-2t} D(X\|Z)$. Likewise, as discussed in the previous section, we expect that $D(X_n\|N(0, \mathrm{Cov}(X))) = O(2^{-n})$ under moment conditions. As we have witnessed in (11), Costa's EPI makes a (dimension-free) quantitative statement about the decay of relative entropy in the Ornstein-Uhlenbeck process. Below, we find that an analogous statement holds for the decay of relative entropy corresponding to the process $\{X_n, n \geq 0\}$. The proof follows by iterating (22), and is omitted due to space constraint.

Theorem 6. *Let X, $\{X_n, n \geq 0\}$ be as above. For any $n \geq 0$,*

$$\frac{1}{4}\sum_{k=0}^{n-1} 2^{-k} I(X_k\|Z) \geq D(X\|Z) - 2^{-n} D(X_n\|Z). \qquad (30)$$

In essence, Costa's EPI makes a precise statement about the behavior of entropy as one interpolates between an arbitrary random vector X and Gaussian Z along the Ornstein-Uhlenbeck process. The discrete time process $\{X_n, n \geq 0\}$ may be thought of as another way of interpolating between X and Gaussian, viz-à-vis repeated convolution. Remarkably, Theorem 6 shows that this latter method of interpolation bears strong resemblance to the de Bruijn identity for the Ornstein-Uhlenbeck process:

$$\int_0^t I(X_s\|Z)\,\mathrm{d}s = D(X\|Z) - D(X_t\|Z) \qquad t \geq 0.$$

Of course, we have already seen this implies Costa's EPI since $\frac{1-e^{-2t}}{2} I(X\|Z) \geq \int_0^t I(X_s\|Z)\,\mathrm{d}s$.

IV. Closing Remarks

We have seen that, when appropriately formulated, the concavity of entropy power can be generalized to non-Gaussian additive perturbations. On this note, we ask whether the Gaussian assumption on W is necessary in Theorem 1. Despite some effort in searching, no counterexample has been found.

Along a different line, we know that there is a strong analogy between the EPI and the Brunn-Minkowski inequality. In light of this, we are moved to speculate that a geometric analogue to Theorem 1 may hold. In particular, for K, L, B bounded convex sets in $\mathbb{R}^d$, does it hold that:

$$|K+L+B|^{1/d}|B|^{1/d} + |K|^{1/d}|L|^{1/d} \leq |K+B|^{1/d}|L+B|^{1/d},$$

where $|\cdot|$ denotes d-dimensional volume? Convexity of the sets is necessary, else L, B could be taken to be balls of specific radii, implying the Costa-Cover conjecture [26] for non-convex K. This was recently shown to be false [27].

Appendix

Proof of Proposition 1. We will show that (9) and (10) are equivalent. The remaining claims follow as a special case of our proof of Theorem 2. To this end, since $J(X) < \infty$, it follows that $J(X + \sqrt{t}Z) < \infty$ by the convolution inequality for Fisher information (4). By de Bruijn's identity, (10) is equivalent to $N(X+\sqrt{t+s}Z) \leq N(X+\sqrt{t}Z) + s\frac{\mathrm{d}}{\mathrm{d}t}N(X+\sqrt{t}Z)$ for $t > 0$ and $s \geq 0$. Stated another way, $N(X+\sqrt{t}Z)$ lies below its tangent lines for $t > 0$, which is equivalent to $N(X+\sqrt{t}Z)$ being concave on $t \in (0, \infty)$. The conclusion can be immediately extended to $t \in [0, \infty)$ since $N(X) \leq N(X + \sqrt{t}Z)$ for any $t > 0$. □

References

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

[2] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3613–3628, 2009.

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

[4] F. Cheng and Y. Geng, "Higher order derivatives in Costa's entropy power inequality," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 5892–5905, 2015.

[5] C. Villani, "A short proof of the 'concavity of entropy power'," *IEEE Transactions on Info. Theory*, vol. 46, no. 4, pp. 1695–1696, 2000.

[6] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*. Springer Sci. & Bus. Media, 2013, vol. 348.

[7] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise (corresp.)," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 279–280, Mar 1974.

[8] O. Johnson and A. Barron, "Fisher information inequalities and the central limit theorem," *Probability Theory and Related Fields*, vol. 129, no. 3, pp. 391–409, 2004.

[9] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inf. and Ctrl.*, vol. 2, no. 2, pp. 101–112, 1959.

[10] A. Dembo, "Simple proof of the concavity of the entropy power with respect to added Gaussian noise," *IEEE Transactions on Information Theory*, vol. 35, no. 4, pp. 887–888, 1989.

[11] J. Fahs and I. Abou-Faycal, "Information measures, inequalities and performance bounds for parameter estimation in impulsive noise environments," *arXiv preprint arXiv:1609.00832*, 2016.

[12] T. A. Courtade, "Strengthening the entropy power inequality," in *2016 IEEE Intl Symp. on Info. Theory (ISIT)*. IEEE, 2016, pp. 2294–2298.

[13] T. A. Courtade, G. Han, and Y. Wu, "A counterexample to the vector generalization of Costa's EPI, and partial resolution," *arXiv preprint arXiv:1704.06164*, 2017.

[14] E. A. Carlen and A. Soffer, "Entropy production by block variable summation and central limit theorems," *Communications in mathematical physics*, vol. 140, no. 2, pp. 339–371, 1991.

[15] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications, and coding," *Foundations and Trends in Comm. and Info. Theory*, vol. 10, no. 1-2, pp. 1–246, 2013.

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

[17] T. A. Courtade, "Links between the logarithmic Sobolev inequality and the convolution inequalities for entropy and Fisher information," *arXiv preprint arXiv:1608.05431*, 2016.

[18] A. Figalli, F. Maggi, and A. Pratelli, "A mass transportation approach to quantitative isoperimetric inequalities," *Inventiones mathematicae*, vol. 182, no. 1, pp. 167–211, 2010.

[19] M. Fathi, E. Indrei, and M. Ledoux, "Quantitative logarithmic Sobolev inequalities and stability estimates," *preprint arXiv:1410.6922*, 2014.

[20] S. G. Bobkov and G. P. Chistyakov, "Entropy power inequality for the Rényi entropy," *IEEE Trans. Inf. Thy.*, vol. 61, no. 2, pp. 708–714, 2015.

[21] 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.

[22] K. Ball, P. Nayar, and T. Tkocz, "A reverse entropy power inequality for log-concave random vectors," *arXiv preprint arXiv:1509.05926*, 2015.

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

[24] A. R. Barron, "Entropy and the central limit theorem," *The Annals of probability*, pp. 336–342, 1986.

[25] S. G. Bobkov, G. P. Chistyakov, and F. Götze, "Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem," *The Annals of Probability*, vol. 41, no. 4, pp. 2479–2512, 2013.

[26] M. Costa and T. Cover, "On the similarity of the entropy power inequality and the Brunn-Minkowski inequality (corresp.)," *IEEE Transactions on Information Theory*, vol. 30, no. 6, pp. 837–839, 1984.

[27] M. Fradelizi and A. Marsiglietti, "On the analogue of the concavity of entropy power in the Brunn-Minkowski theory," *Advances in Applied Mathematics*, vol. 57, pp. 1–20, 2014.