This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2024.3385728

Accepted Manuscript - Version of Record at: https://doi.org/10.1109/TIT.2024.3385728

An Entropy Power Inequality for Dependent Variables

Mokshay Madiman, *Senior Member IEEE,* James Melbourne, and Cyril Roberto,
.

***Abstract*—The entropy power inequality for independent random variables is a foundational result of information theory, with deep connections to probability and geometric functional analysis. Very few extensions of the entropy power inequality have been developed for settings with dependence. We address this gap in the literature by developing entropy power inequalities for dependent random variables. In particular, we highlight the role of log-supermodularity in delivering sufficient conditions for an entropy power inequality stated using conditional entropies.**

***Index Terms*—Entropy Power Inequality; Dependent Variables; log-supermodular; submodular functions; Fisher information.**

I. INTRODUCTION

When a random vector X has density f, the *entropy* of X is

$$h(X) = h(f) := -\int f(x)\log f(x)dx = \mathbb{E}[-\log f(X)]. \tag{1}$$

This quantity is sometimes called the Boltzmann-Shannon or Boltzmann-Gibbs entropy, or the differential entropy. In dimension d, the *entropy power* of X is $N(X) = e^{\frac{2h(X)}{d}}$. As is usual, we abuse notation and write $h(X)$ and $N(X)$, even though these are functionals depending only on the density of X and not on its random realization. The entropy power $N(X) \in [0,\infty]$ can be thought of as a "measure of randomness". It is an (inexact) analogue of volume: if U_A is uniformly distributed on a bounded Borel set A, then it is easily checked that $h(U_A) = \log|A|$ and hence $N(U_A) = |A|^{2/d}$. The reason for this particular definition of entropy power is that the "correct" comparison is not to uniforms but to Gaussians: observe that when $Z \sim N(0,\sigma^2 I)$ (i.e., Z has the Gaussian distribution with mean 0 and covariance matrix that is a multiple of the identity), the entropy power of Z is $N(Z) = (2\pi e)\sigma^2$. Thus the entropy power of X is– up to a universal constant– the variance of the isotropic normal that has the same entropy as X, i.e., if $Z \sim N(0,\sigma_Z^2 I)$ and $h(Z) = h(X)$, then

$$N(X) = N(Z) = (2\pi e)\sigma_Z^2.$$

Looked at this way, entropy power is the "effective variance" of a random variable.

The entropy power inequality (EPI) states that for any two independent random vectors X and Y in $\mathbb{R}^d$ such that the entropies of X, Y and $X+Y$ exist,

$$N(X+Y) \geqslant N(X) + N(Y).$$

The EPI was stated by Shannon [50] with an incomplete proof; the first complete proof was provided by Stam in his Ph.D. thesis [51] (see also [52]). The EPI plays an important role in Information Theory, where it first arose and was used (first by Shannon, and later by many others) to prove statements about the fundamental limits of communication over various models of communication channels. It has also been recognized as a very useful inequality in Probability Theory, with close connections to the logarithmic Sobolev inequality for the Gaussian distribution as well as to the Central Limit Theorem. The history of the EPI and its connections to many other inequalities are described in the surveys [18], [19], [21], [39] and the connections to the Central Limit Theorem in [25], [33].

The EPI has been generalized and extended in several directions. For instance, there has been considerable recent progress in the understanding of Rényi entropy analogs of the EPI (see [8], [9], [11], [31], [44], [48], [49], [57]), using which fruitful connections have been made with other branches of mathematics [4], [30], [36], [40], [45], [46]. Discrete analogs have been explored in [1], [12], [22], [23], [27], [37], [42], [43], [58]. Matrix extensions and connections of the EPI with powerful functional inequalities like the Brascamp-Lieb inequality have been discovered [2], [15], [32], [59]. Further strengthenings, sharpenings, and generalizations can be found in [3], [16], [17], [34], [35], [41], [56], while so-called reverse entropy power inequalities for special classes of distributions have been obtained in [6], [7], [10].

The aim of this paper is to deliver an EPI for dependent summands. Several previous attempts have been made towards this same goal, see [14], [24], [26], [53], [54], which we now briefly recollect. This paper borrows the framework of [14], where Carlen and Soffer– among other results– derive a stability version of the entropy power inequality by obtaining a monotonicity result along the Ornstein-Uhlenbeck semigroup (in contrast to Stam [52], whose original proof of the EPI used similar monotonicity result for the heat semigroup), and in the same paper derived CLT convergence even for some dependent variables using entropic methods. The first EPI stated and proved for potentially dependent random variables, as far as we are aware, comes from Takano [53], [54], who shows that

---

M. Madiman is with the University of Delaware, Department of Mathematical Sciences, 501 Ewing Hall, Newark, DE 19716, USA. email: madiman@udel.edu

J. Melbourne e-mail: james.melbourne@cimat.mx, .

C. Roberto is with MODAL'X, UPL, University Paris Nanterre, CNRS, F92000 Nanterre France. email: croberto@math.cnrs.fr. Research supported by the Labex MME-DII funded by ANR, reference ANR-11-LBX-0023-01, the Fondation Simone et Cino del Duca and the FP2M federation (CNRS FR 2036).

Accepted Manuscript - Version of Record at: https://doi.org/10.1109/TIT.2024.3385728

an EPI holds for some pairs of dependent $\mathbb{R}$-valued random variables. Takano's argument, which also uses monotonicity in heat flow, requires weak dependence as measured by a constant δ_4, which is the 4-th moment of the averaged normalized difference between the joint density and the product of the marginals. Based on similar machinery as Takano, a breakthrough in perspective was achieved by Johnson [24], who pursued a *conditional* entropy power inequality (i.e., one expressed in terms of conditional entropies) for a pair of real-valued random variables. However, as we will demonstrate through counterexample, the main result of this paper [24, Theorem 6] is not correct.

In this paper, we present some new EPIs for dependent summands, with specific attention paid to demonstrating and exploring "log-supermodularity" as a sufficient condition for the veracity of EPIs. In particular we have the following theorem.

Theorem I.1. *Let X, Y be $\mathbb{R}^d$-valued random variables such that their joint density convolved with the standard Gaussian in $\mathbb{R}^{2d}$ is log-supermodular. Then*

$$e^{\frac{2}{d}h(X+Y)} \geqslant e^{\frac{2}{d}h(X|Y)} + e^{\frac{2}{d}h(Y|X)}.$$

The definition of log-supermodularity will be recalled in subsequent sections, where results will be stated and derived in greater generality. Let us mention that when X and Y are independent, (X, Y) has a log-supermodular density if both X and Y have log-supermodular densities. Thus when $d = 1$, Theorem I.1 contains the classical EPI since every density on the real line is trivially log-supermodular. We emphasize that although Theorem I.1 will be derived from a more general inequality that does contain the classical EPI, as stated above the conditional log-supermodular EPI does not generalize the classical one for $d \geqslant 2$.

We will also highlight a recent conjecture of Zartash and Robeva [60], that log-supermodularity is stable under standard Gaussian convolution. In very recent work [38], we have investigated inequalities for log-supermodular densities, and in particular verified their conjecture. Consequently, one can relax[1] the hypothesis of Theorem I.1 to simply asking that the joint density of (X, Y) is log-supermodular; we refer to [38] for details.

Let us outline the rest of the paper. In Section II we fix notations and definitions requisite for the analysis. In Section III, we extend Johnson's [24] (see also [53]) Fisher information inequality for pairs of random variables to arbitrary number of summands. In Section IV, we derive a general EPI for dependent random variables via the Ornstein-Uhlenbeck flow, with a quantitative error term given by the integral of a Fisher information quantity along the flow. In Section V we derive corollaries of the main result, and explain the usefulness of log-supermodularity. We conclude the paper in Section VI by correcting the record and discussing some erroneous statements in [24]; in particular, we give there a counterexample to [24, Theorem 6].

II. NOTATION AND DEFINITIONS

We first recall the definition of the conditional entropy. Given X an $\mathbb{R}^n$-valued random variable and Y an $\mathbb{R}^k$-valued random variable with a joint density, the conditional entropy of X given Y is defined as

$$h(X|Y) := h(X, Y) - h(Y).$$

Throughout the remaining paper we denote by $X = (X_1, \ldots, X_n)$ the $(\mathbb{R}^d)^n$-valued random variable, with components $X_i \in \mathbb{R}^d$ and by $p\colon (\mathbb{R}^d)^n \to \mathbb{R}^+$ the joint density of X and we assume that p is sufficiently smooth. We may often use the shorthand notation $x_i^j := (x_i, x_{i+1}, \ldots, x_j)$, $i, j \in \{1, \ldots, n\}$, $i \leqslant j$. Also, p is function of $x = (x_1, \ldots, x_n)$, with, for $i = 1, \ldots, n$, $x_i = (x_{i,1}, \ldots, x_{i,d}) \in \mathbb{R}^d$.

Then, the score function of the vector $X = (X_1, \ldots, X_n)$ is defined for $x \in (\mathbb{R}^d)^n$ as

$$\rho(x) := -\frac{\nabla p}{p}(x) = (\rho_1(x), \ldots, \rho_n(x)) \in (\mathbb{R}^d)^n,$$

where ∇ denotes the gradient vector. For simplicity of notations set $\nabla_i = (\frac{\partial}{\partial x_{i,1}}, \ldots, \frac{\partial}{\partial x_{i,d}})$ so that

$$\rho_i = -\frac{\nabla_i p}{p} \in \mathbb{R}^d.$$

Next we define the Fisher information matrix $I = (I_{ij})_{i,j}$ on $\mathbb{R}^n \times \mathbb{R}^n$ as

$$I_{ij} = I_{ij}(X) := \mathbb{E}_X \langle \rho_i, \rho_j \rangle = \mathbb{E} \langle \rho_i(X), \rho_j(X) \rangle$$

for $i, j = 1, \ldots, n$, where $\langle \cdot, \cdot \rangle$ stands for the usual scalar product on $\mathbb{R}^d$.

Now denote by $f\colon \mathbb{R}^d \to \mathbb{R}^+$ the density of the sum $W := X_1 + \cdots + X_n$ and define similarly its score function as

$$\rho_W := -\frac{\nabla f}{f} \in \mathbb{R}^d.$$

It will be useful to relate the score function of W to that of X. For simplicity of notations set $\bar{x}^i$ for the vector x without the coordinates of x_i, and the corresponding differential form $d\bar{x}^i := dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$. Then, we observe that, with the notation $x_w^{(i)} = (x_1^{i-1}, w - \sum_{j \neq i} x_j, x_{i+1}^n)$ for any $i = 1, \ldots, n$,

$$\begin{aligned} \rho_W(w) &= -\int \cdots \int \frac{\nabla_i p(x_w^{(i)})}{f(w)} d\bar{x}^i \\ &= -\int \cdots \int \frac{\nabla_i p(x_w^{(i)})}{p(x_w^{(i)})} \frac{p(x_w^{(i)})}{f(w)} d\bar{x}^i \\ &= \mathbb{E}\left(\rho_i(X) | W = w\right). \end{aligned} \tag{2}$$

In particular, for any $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, it holds

$$\left(\sum_{i=1}^n \lambda_i\right) \rho_W(w) = \mathbb{E}\left(\sum_{i=1}^n \lambda_i \rho_i(X) \middle| W = w\right), \qquad w \in \mathbb{R}^d. \tag{3}$$

---

[1] Log-supermodularity under convolution with a standard Gaussian implies log-supermodularity. Indeed, by a homogeneity argument, (X, Y) log-supermodular under convolution with a standard Gaussian Z implies that (X, Y) is log-supermodular under convolution with tZ for any $t > 0$, taking $t \to 0$ will show that (X, Y) are log-supermodular.

Accepted Manuscript - Version of Record at: https://doi.org/10.1109/TIT.2024.3385728

In general, for a random variable X with density p, its Fisher information is

$$I(X) = \int \frac{|\nabla p|^2}{p}.$$

Finally, we denote by J (or $I(W) = I(X_1 + \cdots + X_n)$) the Fisher information of W, *i.e.*

$$J := \mathbb{E}_W \left(\langle \rho_W, \rho_W \rangle \right) = \mathbb{E}|\rho_W(W)|^2.$$

III. FISHER INFORMATION INEQUALITY FOR DEPENDENT RANDOM VARIABLES.

In this section, we extend the Fisher information inequality for dependent variables obtained by Johnson [24] in the case $n = 2$ to arbitrary n. Denote by $\mathbf{1} := (1, \ldots, 1)$ the vector of $\mathbb{R}^n$ with all ones, and recall that $\langle \cdot, \cdot \rangle$ denotes the scalar product (of $\mathbb{R}^d$ or $\mathbb{R}^n$, depending on the context).

Proposition III.1. *For $\mathbb{R}^d$-valued random vectors, X_1, $\ldots, X_n$ with sufficiently smooth joint density, for any $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, it holds*

$$\left(\sum_{i=1}^n \lambda_i \right)^2 J \leqslant \sum_{i,j=1}^n \lambda_i \lambda_j I_{ij}. \tag{4}$$

Further,

$$\frac{1}{J} \geqslant \langle \mathbf{1}, I^{-1} \mathbf{1} \rangle \tag{5}$$

when I^{-1}, the inverse matrix of I exists.

When $d = 1$ and the X_i are assumed independent (4) yields Stam [52, Equation 2.9] as $I_{ij} = 0$ for $i \neq j$. Taking $d = 1$ and $n = 2$, (5) recovers the inequality of Johnson [24, Theorem 4], while the additional assumption of independence one recovers Blachman [5, Equation 3]. The equivalence between (4) and (5) in the independent case was attributed to De Bruijn by Stam in a footnote of [52].

Proof. Fix $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$. Since

$$\left\| \left(\sum_{i=1}^n \lambda_i \right) \rho_W(W) - \sum_{i=1}^n \lambda_i \rho_i(X) \right\|^2 \geqslant 0$$

expanding the product and taking the expectation, we have that

$$\left(\sum_{i=1}^n \lambda_i \right)^2 J + \sum_{i,j=1}^n \lambda_i \lambda_j I_{ij} - 2 \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j \mathbb{E}\left(\langle \rho_W, \rho_j(X) \rangle \right),$$

is non-negative. But (2) guarantees that

$$\begin{aligned} \mathbb{E}\left(\langle \rho_W, \rho_j(X) \rangle \right) &= \mathbb{E}\left(\langle \rho_W, \mathbb{E}\left(\rho_j(X) | W \right) \rangle \right) \\ &= \mathbb{E}\left(\langle \rho_W, \rho_W \rangle \right) \\ &= J. \end{aligned}$$

Hence, for any $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ it holds

$$\left(\sum_{i=1}^n \lambda_i \right)^2 J \leqslant \sum_{i,j=1}^n \lambda_i \lambda_j I_{ij} = \lambda^T I \lambda \tag{6}$$

as expected, where the last equality is just a rewriting using the matrix representation (here $\lambda := (\lambda_1, \ldots, \lambda_n)$ and λ^T is the vector λ transposed).

Our aim is now to optimize (6) over all $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$. Indeed, (6) implies that

$$\frac{1}{J} \geqslant \sup_{\lambda_1, \ldots, \lambda_n} \frac{\left(\sum_{i=1}^n \lambda_i \right)^2}{\sum_{i,j=1}^n \lambda_i \lambda_j I_{ij}} = \sup_{\lambda_1, \ldots, \lambda_n} \frac{\langle \mathbf{1}, \lambda \rangle^2}{\sum_{i,j=1}^n \lambda_i \lambda_j I_{ij}}$$

where we recall that $\mathbf{1} := (1, \ldots, 1)$. Since I is positive semi-definite it has a square root we denote by $\sqrt{I}$ that is invertible when I is. Then, setting $\alpha = \sqrt{I} \lambda$, the supremum above equals

$$\sup_{\alpha_1, \ldots, \alpha_n} \frac{\langle (\sqrt{I})^{-1} \alpha, \mathbf{1} \rangle^2}{\langle \alpha, \alpha \rangle}.$$

By the Cauchy–Schwarz Inequality and symmetry,

$$\begin{aligned} \langle (\sqrt{I})^{-1} \alpha, \mathbf{1} \rangle^2 &= \langle \alpha, (\sqrt{I})^{-1} \mathbf{1} \rangle^2 \\ &\leqslant \langle \alpha, \alpha \rangle \langle (\sqrt{I})^{-1} \mathbf{1}, (\sqrt{I})^{-1} \mathbf{1} \rangle \\ &= \langle \alpha, \alpha \rangle \langle \mathbf{1}, I^{-1} \mathbf{1} \rangle \end{aligned}$$

which ends the proof using the equality case in the Cauchy–Schwarz Inequality. ☐

In some situations it will be useful to deal with the Fisher information of the weighted sum $\sum_{i=1}^n \lambda_i X_i$ instead of that of $W = \sum_{i=1}^n X_i$. We claim that

$$I\left(\sum_{i=1}^n \lambda_i X_i \right) \leqslant \sum_{i,j=1}^n \lambda_i \lambda_j I_{ij}(X) \tag{7}$$

for any $\lambda_1, \ldots, \lambda_n \in (0, 1)$ with $\sum_{i=1}^n \lambda_i^2 = 1$, where as usual $X = (X_1, \ldots, X_n)$.

To see this, observe first that if p is the joint density of $X = (X_1, \ldots, X_n)$, then for $x = (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$

$$q(x) = \frac{1}{\prod_{i=1}^n \lambda_i^d} p\left(\frac{x_1}{\lambda_1}, \ldots, \frac{x_n}{\lambda_n} \right),$$

is the density of $(\lambda_1 X_1, \ldots, \lambda_n X_n)$. Therefore, changing variables, and using the notation $\frac{x}{\lambda} := (\frac{x_1}{\lambda_1}, \ldots, \frac{x_n}{\lambda_n})$

$$\begin{aligned} & I_{ij}(\lambda_1 X_1, \ldots, \lambda_n X_n) \\ &= \iint \frac{1}{q(x)} \langle \nabla_i q(x), \nabla_j q(x) \rangle dx_1 \ldots dx_n \\ &= \iint \frac{\frac{1}{\lambda_i \lambda_j} \langle \nabla_i p(\frac{x}{\lambda}), \nabla_j p(\frac{x}{\lambda}) \rangle / (\lambda_1^d \ldots \lambda_n^d)^2}{p\left(\frac{x}{\lambda} \right) / \lambda_1^d \ldots \lambda_n^d} dx_1 \ldots dx_n \\ &= \frac{1}{\lambda_i \lambda_j} \iint \frac{1}{p(x)} \langle \nabla_i p(x), \nabla_j p(x) \rangle dx_1 \ldots dx_n \\ &= \frac{1}{\lambda_i \lambda_j} I_{ij}(X). \end{aligned} \tag{8}$$

By (4) for t_i such that $\sum_{i=1}^n t_i = 1$ and $Y = (Y_1, \ldots, Y_n)$,

$$I(Y_1 + \cdots + Y_n) \leqslant \sum_{i,j} t_i t_j I_{ij}(Y).$$

Applying this to λ such that $\sum_{i=1}^n \lambda_i^2 = 1$ and $X = (X_1, \ldots, X_n)$ for $Y_i = \lambda_i X$ and $t_i = \lambda_i^2$, gives

$$\begin{aligned} I(\lambda_1 X_1 + \cdots + \lambda_n X_n) &\leqslant \sum_{i,j} \lambda_i^2 \lambda_j^2 I_{ij}(\lambda_1 X_1, \ldots, \lambda_n X_n) \\ &= \sum_{i,j} \lambda_i \lambda_j I_{ij}(X), \end{aligned}$$

Accepted Manuscript - Version of Record at: https://doi.org/10.1109/TIT.2024.3385728

where the equality follows from (8). We note that under of the assumption of independence (7) yields [14, Lemma 1.3], which also appears as [13, Theorem 7].

We end this section with a simple but useful remark related to the Fisher information of a single random variable X_i in comparison to that of $X = (X_1, \dots, X_n)$. Given the joint density p of X (recall that $d\bar{x}^i := dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$), the i-th marginal is

$$p_i(x_i) := \int p(x_1, \dots, x_n) d\bar{x}_i, \qquad x_i \in \mathbb{R}^d$$

so that the Fisher information of X_i is given by

$$I(X_i) = \int_{\mathbb{R}^d} \frac{|\nabla p_i|^2}{p_i}.$$

Now, by the Cauchy–Schwarz Inequality,

$$\begin{aligned} I(X_i) &= \int_{\mathbb{R}^d} \frac{\left|\int \nabla_i p(x) d\bar{x}_i\right|^2}{\int p(x) d\bar{x}_i} dx_i \\ &= \int_{\mathbb{R}^d} \frac{\left|\int \frac{\nabla_i p(x)}{\sqrt{p(x)}} \sqrt{p(x)} d\bar{x}_i\right|^2}{\int p(x) d\bar{x}_i} dx_i \\ &\leqslant \iint\limits_{\mathbb{R}^d \times (\mathbb{R}^d)^{(n-1)}} \frac{|\nabla_i p(x)|^2}{p(x)} d\bar{x}_i dx_i \\ &= I_{ii}(X). \end{aligned} \tag{9}$$

Similarly (details are left to the reader), for any $i = 1, \dots, n$,

$$\begin{aligned} I(X_j, j \neq i) &= \int_{(\mathbb{R}^d)^{(n-1)}} \frac{\sum_{j \neq i} \left|\int \nabla_j p(x) dx_i\right|^2}{\int p(x) d\bar{x}_i} d\bar{x}_i \\ &\leqslant \sum_{j \neq i} I_{jj}(X). \end{aligned} \tag{10}$$

IV. EPI FOR DEPENDENT VARIABLES VIA THE ORNSTEIN-UHLENBECK FLOW

In this section we will use the Ornstein-Uhlenbeck approach of Carlen and Soffer [14] to the entropy power inequality to get some general EPI valid for non independent variables.

Given a density $f \colon \mathbb{R}^m \to \mathbb{R}_+$ define the adjoint (for the Lebesgue measure) operator of the Orstein-Uhlenbeck semigroup P_t^*, as the solution of the following PDE in $\mathbb{R}^m$

$$\frac{d}{dt} P_t^* f = (\Delta + x \cdot \nabla) P_t^* f + m P_t^* f$$

where the dot sign stands for the scalar product, Δ is the Laplacian, ∇ the gradient operator (a vector) and initial condition $P_t^* f|_{t=0} = f$. This semi-group has an explicit representation formula (Mehler-type formula). For X with density f and $Z \sim N(0, I)$, it holds

$$\begin{aligned} P_t^* f(x) &= e^t \mathbb{E}[f(e^t x - \sqrt{e^{2t} - 1} Z)] \\ &= \mathbb{E}[g_{\sqrt{1 - e^{-2t}}}(x - e^{-t} X)] \end{aligned}$$

with $g_s(x) = \frac{e^{-|x|^2/2s}}{(2\pi s)^{m/2}}$ the m-dimensional centered Gaussian density of variance s. Furthermore, $e^{-t} X + \sqrt{1 - e^{-2t}} Z$ (with X and $Z \sim N(0, I)$ independent) has density $P_t^* f$.

One key property is the following relation between the Shannon entropy and the Fisher information (see [14, Lemma 1.2]):

$$\begin{aligned} &h(e^{-t} X + \sqrt{1 - e^{-2t}} Z) \\ &\qquad = h(X) + \int_0^t I(e^{-s} X + \sqrt{1 - e^{-2s}} Z) ds - mt. \end{aligned} \tag{11}$$

We are in position to prove the following proposition which constitutes a linearized form of the entropy power inequality for general random variables.

Proposition IV.1. *Let $X_1, \dots, X_n$ be $\mathbb{R}^d$-valued random variables and $\lambda_1, \dots, \lambda_n \in (0, 1)$ so that $\sum_{i=1}^n \lambda_i^2 = 1$. Set $Y_i = Y_i(t) := e^{-t} X_i + \sqrt{1 - e^{-2t}} Z_i$, $i = 1, \dots, n$, where $Z_i \sim N(0, I)$ are independent standard Gaussian on $\mathbb{R}^d$. Then for all $t \geqslant 0$ it holds*

$$\begin{aligned} &h\left(\sum_{i=1}^n \lambda_i Y_i(t)\right) - \sum_{i=1}^n \lambda_i^2 h(Y_i(t)) \\ &\qquad \leqslant h\left(\sum_{i=1}^n \lambda_i X_i\right) - \sum_{i=1}^n \lambda_i^2 h(X_1) + R_t \end{aligned}$$

with

$$\begin{aligned} R_t &= R_t(X_1, \dots, X_n) \\ &= \int_0^t \left(\sum_{i=1}^n \lambda_i^2 (I_{ii}(s) - I(Y_i(s))) + \sum_{\substack{i,j=1 \\ i \neq j}}^n \lambda_i \lambda_j I_{ij}(s)\right) ds \end{aligned}$$

where we set for simplicity $I_{ij}(s) := I_{ij}(Y_1(s), \dots, Y_n(s))$. In particular,

$$h\left(\sum_{i=1}^n \lambda_i X_i\right) \geqslant \sum_{i=1}^n \lambda_i^2 h(X_i) - \limsup_{t \to \infty} R_t.$$

Remark IV.1. *Note that for independent random variables $X_1, \dots, X_n$, $R_t(X_1, \dots, X_n) = 0$ for all t, therefore recovering Carlen-Soffer's inequality [14, Theorem 1.1] and Stam's linearized version of the entropy power inequality.*

Proof. Fix $t > 0$. Applying (7) to $Y_i(s) = e^{-s} X_i + \sqrt{1 - e^{-2s}} Z_i$ with $s \leqslant t$, we have

$$I\left(\sum_{i=1}^n \lambda_i Y_i(s)\right) \leqslant \sum_{i,j=1}^n \lambda_i \lambda_j I_{ij}(Y_1(s), \dots, Y_n(s)).$$

Therefore, by (11), it holds

$$\begin{aligned} h\left(\sum_{i=1}^n \lambda_i Y_i(t)\right) &- h\left(\sum_{i=1}^n \lambda_i X_i\right) \\ &= \int_0^t I\left(\sum_{i=1}^n \lambda_i Y_i(s)\right) ds - t\ d \\ &\leqslant \int_0^t \left(\sum_{i,j=1}^n \lambda_i \lambda_j I_{ij}(Y_1(s), \dots, Y_n(s))\right) ds - td \\ &= \sum_{i=1}^n \lambda_i^2 \left(\int_0^t I(Y_i(s)) ds - dt\right) + R_t \\ &= \sum_{i=1}^n \lambda_i^2 \left(h(Y_i(t)) - h(X_i)\right) + R_t. \end{aligned}$$

This leads to the first part of the proposition.

For the second part it is enough to take the limit $t \to \infty$ and to observe that

$$h\left(\sum_{i=1}^n \lambda_i Y_i(t)\right) - \sum_{i=1}^n \lambda_i^2 h(Y_i(t))$$

converges to $h(Z_{n+1}) - \sum_{i=1}^n \lambda_i^2 h(Z_i)$ where $Z_i \sim N(0, I)$ are independent. Since $h(Z_i) = \frac{d}{2}\log(2\pi e)$ and $\sum \lambda_i^2 = 1$, $h(Z_{n+1}) - \sum_{i=1}^n \lambda_i^2 h(Z_i) = 0$, ending the proof of the proposition. □

As a direct consequence, we obtain the following entropy power inequality for general random variables.

Corollary IV.2. *Let $X_1, \dots, X_n$ be $\mathbb{R}^d$-valued random variables and, for $i = 1, \dots, n$,*

$$\lambda_i^2 := \frac{e^{\frac{2}{d}h(X_i)}}{e^{\frac{2}{d}h(X)} + \cdots + e^{\frac{2}{d}h(X_n)}}$$

Set $\bar{X}_i = X_i/\lambda_i$ and $\bar{Y}_i(s) = e^{-s}\bar{X}_i + \sqrt{1 - e^{-2s}} Z_i$, $s \geqslant 0$, where $Z_1, \dots, Z_n \sim N(0, I)$ are independent standard Gaussian (on $\mathbb{R}^d$). Finally denote

$$\begin{aligned} \bar{R} &= \bar{R}(X_1, \dots, X_n) \\ &= \frac{2}{d}\int_0^\infty \left(\sum_{i=1}^n \lambda_i^2 (\bar{I}_{ii}(s) - I(\bar{Y}_i(s))) + \sum_{\substack{i,j=1 \\ i \neq j}}^n \lambda_i \lambda_j \bar{I}_{ij}(s)\right) ds \end{aligned}$$

where we set for simplicity $\bar{I}_{ij}(s) := I_{ij}(\bar{Y}_1(s), \dots, \bar{Y}_n(t))$. Then,

$$e^{\frac{2}{d}h(X_1 + \cdots + X_n)} \geqslant \left(\sum_{i=1}^n e^{\frac{2}{d}h(X_i)}\right) e^{-\bar{R}}.$$

Remark IV.2. *As for the previous proposition, if X_i are independent $\bar{R}(X_1, \dots, X_n) = 0$ and we recover the usual entropy power inequality.*

Proof. It follows by Proposition IV.1 applied to $\bar{X}_i = X_i/\lambda_i$, $i = 1, \dots, n$ that

$$\begin{aligned} e^{\frac{2}{d}h(X_1 + \cdots + X_n)} &= e^{\frac{2}{d}h(\lambda_1 \bar{X}_1 + \cdots + \lambda_n \bar{X}_n)} \\ &\geqslant \exp\left\{\frac{2\lambda_1^2}{d} h(\bar{X}_1) + \cdots + \frac{2\lambda_n^2}{d} h(\bar{X}_n)\right\} e^{-R(\bar{X}_1, \dots, \bar{X}_n)} \\ &= \prod_{i=1}^n \left(\frac{e^{\frac{2}{d}h(X_i)}}{\lambda_i^2}\right)^{\lambda_i^2} e^{-\bar{R}(X_1, \dots, X_n)} \end{aligned}$$

since $h(\bar{X}_i) = h(X_i/\lambda_i) = h(X_i) - d\log\lambda_i$ and $R(\bar{X}_1, \dots, \bar{X}_n) = \bar{R}(X_1, \dots, X_n)$. The expected result follows. □

V. ENTROPY POWER INEQUALITY FOR CONDITIONAL ENTROPY

Our next aim is to deal with entropy power inequality for conditional entropy. We start with a linearized form similar to Proposition IV.1.

Proposition V.1. *Let $X_1, \dots, X_n$ be $\mathbb{R}^d$-valued random variables and $\lambda_1, \dots, \lambda_n \in (0, 1)$ so that $\sum_{i=1}^n \lambda_i^2 = 1$. Set $Y_i(s) = e^{-s}X_i + \sqrt{1 - e^{-2s}} Z_i$, $s \geqslant 0$, $i = 1, \dots, n$, where $Z_1, \dots, Z_n \sim N(0, I)$ are independent standard Gaussian (on $\mathbb{R}^d$). Then it holds*

$$\begin{aligned} h\left(\sum_{i=1}^n \lambda_i Y_i\right) &- \sum_{i=1}^n \lambda_i^2 h(Y_i | Y_j, j \neq i) \\ &\leqslant h\left(\sum_{i=1}^n \lambda_i X_i\right) - \sum_{i=1}^n \lambda_i^2 h(X_i | X_j, j \neq i) + S_t \end{aligned}$$

where

$$S_t = S_t(X_1, \dots, X_n) = \int_0^t \mathcal{S}_s ds$$

for

$$\begin{aligned} \mathcal{S}_s := &\sum_{i,j=1}^n \lambda_i \lambda_j I_{ij}(Y_1(s), \dots, Y_n(s)) \\ &- \sum_{i=1}^n I_{ii}(Y_1(s), \dots, Y_n(s)) + \sum_{i=1}^n \lambda_i^2 I(Y_j(s), j \neq i). \end{aligned}$$

In particular,

$$h\left(\sum_{i=1}^n \lambda_i X_i\right) \geqslant \sum_{i=1}^n \lambda_i^2 h(X_i | X_j, j \neq i) - \limsup_{t \to \infty} S_t.$$

Proof. Observe that, for any $i = 1, \dots, n$,

$$h(Y_i(t) | Y_j(t), j \neq i) = h(Y_1(t), \dots, Y_n(t)) - h(Y_j(t), j \neq i). \tag{12}$$

Therefore,

$$\begin{aligned} h\left(\sum_{i=1}^n \lambda_i Y_i(t)\right) &- \sum_{i=1}^n \lambda_i^2 h(Y_i(t) | Y_j(t), j \neq i) \\ &= h\left(\sum_{i=1}^n \lambda_i Y_i(t)\right) - h(Y_1(t), \dots, Y_n(t)) \\ &\quad + \sum_{i=1}^n \lambda_i^2 h(Y_j(t), j \neq i). \end{aligned}$$

Hence, by (11), it holds

$$\begin{aligned}
&h\left(\sum_{i=1}^n \lambda_i Y_i(t)\right) - \sum_{i=1}^n \lambda_i^2 h(Y_i(t)|Y_j(t), j\neq i)\\
&= h\left(\sum_{i=1}^n \lambda_i X_i\right) - h(X_1,\dots,X_n) + \sum_{i=1}^n \lambda_i^2 h(X_j, j\neq i)\\
&\quad + \int_0^t \Bigg[I\left(\sum_{i=1}^n \lambda_i Y_i(s)\right) - I(Y_1(s),\dots,Y_n(s))\\
&\qquad\qquad + \sum_{i=1}^n \lambda_i^2 I(Y_j(s), j\neq i)\Bigg] ds\\
&= h\left(\sum_{i=1}^n \lambda_i X_i\right) - \sum_{i=1}^n \lambda_i^2 h(X_i|X_j, j\neq i)\\
&\quad + \int_0^t \Bigg[I\left(\sum_{i=1}^n \lambda_i Y_i(s)\right) - I(Y_1(s),\dots,Y_n(s))\\
&\qquad\qquad + \sum_{i=1}^n \lambda_i^2 I(Y_j(s), j\neq i)\Bigg] ds
\end{aligned}$$

where in the last equality we used (12) at time $t = 0$ to reconstruct the conditional entropies. Applying the Fisher information inequality (7), we obtain

$$I\left(\sum_{i=1}^n \lambda_i Y_i(s)\right) \leqslant \sum_{i,j=1}^n \lambda_i\lambda_j I_{ij}(s)$$

where we set for simplicity $I_{ij}(s) := I_{ij}(Y_1(s),\dots,Y_n(s))$. On the other hand, if we denote p_s the density of $(Y_1(s),\dots,Y_n(s))$,

$$I(Y_1(s),\dots,Y_n(s)) = \int_{(\mathbb{R}^d)^n} \frac{|\nabla p_s|^2}{p_s} = \sum_{i=1}^n I_{ii}(s).$$

It follows that

$$\begin{aligned}
&h\left(\sum_{i=1}^n \lambda_i Y_i(t)\right) - \sum_{i=1}^n \lambda_i^2 h(Y_i(t)|Y_j(t), j\neq i)\\
&\leqslant h\left(\sum_{i=1}^n \lambda_i X_i\right) - \sum_{i=1}^n \lambda_i^2 h(X_i|X_j, j\neq i)\\
&\quad + \int_0^t \Bigg[\sum_{i,j=1}^n \lambda_i\lambda_j I_{ij}(s) - \sum_{i=1}^n I_{ii}(s)\\
&\qquad\qquad + \sum_{i=1}^n \lambda_i^2 I(Y_j(s), j\neq i)\Bigg] ds
\end{aligned}$$

as expected.

The second part of the theorem follows from the fact that, in the limit $t\to\infty$,

$$\begin{aligned}
h\left(\sum_{i=1}^n \lambda_i Y_i\right) - \sum_{i=1}^n \lambda_i^2 h(Y_i|Y_j, j\neq i) \to&\\
h\left(Z_{n+1}\right) - \sum_{i=1}^n \lambda_i^2 h(Z_i),&
\end{aligned}$$

where Z_i, $i = 1,\dots,n+1$, are *i.i.d.* standard Gaussian variables in $\mathbb{R}^d$, for which it is known that $h(Z_i) = \frac{d}{2}\log(2\pi e)$. In particular $h\left(Z_{n+1}\right) - \sum_{i=1}^n \lambda_i^2 h(Z_i) = 0$ leading to the desired conclusion. □

As a direct consequence, we deduce a general entropy power inequality for conditional entropy.

Corollary V.2. *Let $X_1,\dots,X_n$ be $\mathbb{R}^d$-valued random variables and, for $i = 1,\dots,n$,*

$$\lambda_i^2 := \frac{e^{\frac{2}{d}h(X_i|X_j,j\neq i)}}{\sum_{i=1}^n e^{\frac{2}{d}h(X_i|X_j,j\neq i)}}.$$

Set $\bar X_i = X_i/\lambda_i$ and $\bar Y_i(s) = e^{-s}\bar X_i + \sqrt{1-e^{-2s}}Z_i$, $s\geqslant 0$, where $Z_1,\dots,Z_n \sim N(0,I)$ are independent standard Gaussian (on $\mathbb{R}^d$). Finally denote

$$\begin{aligned}
\bar S &= \bar S(X_1,\dots,X_n)\\
&= \frac{2}{d}\int_0^\infty \Bigg[\sum_{i,j=1}^n \lambda_i\lambda_j \bar I_{ij}(s) - \sum_{i=1}^n \bar I_{ii}(s)\\
&\qquad\qquad + \sum_{i=1}^n \lambda_i^2 I(\bar Y_j(s), j\neq i)\Bigg] ds
\end{aligned}$$

where we set for simplicity $\bar I_{ij}(s) := I_{ij}(\bar Y_1(s),\dots,\bar Y_n(s))$. Then it holds

$$e^{\frac{2}{d}h(X_1\cdots+X_n)} \geqslant \left(\sum_{i=1}^n e^{\frac{2}{d}h(X_i|X_j,j\neq i)}\right) e^{-\bar S}. \tag{13}$$

In particular, if $\bar I_{ij}(s)\leqslant 0$ for all $i\neq j$ and all $s>0$, then

$$e^{\frac{2}{d}h(X_1\cdots+X_n)} \geqslant \sum_{i=1}^n e^{\frac{2}{d}h(X_i|X_j,j\neq i)}.$$

Proof. By Proposition V.1 applied to $\bar X_i$, it holds

$$\begin{aligned}
&e^{\frac{2}{d}h(X_1+\cdots+X_n)} = e^{\frac{2}{d}h(\lambda_1\bar X_1+\cdots+\lambda_2\bar X_2)}\\
&\geqslant \exp\left\{\frac{2}{d}\sum_{i=1}^n \lambda_i^2 h(\bar X_i|\bar X_j, j\neq i)\right\} e^{-\limsup_{t\to\infty} S_t(\bar X_1,\dots,\bar X_n)}\\
&= \prod_{i=1}^n \left(\frac{e^{\frac{2}{d}h(X_i|X_j,j\neq i)}}{\lambda_i^2}\right)^{\lambda_i^2} e^{-\bar S}
\end{aligned}$$

since $h(\bar X_i|\bar X_j, j\neq i) = h(X_i|X_j, j\neq i) - \log\lambda_i$ for all $i = 1,\dots,n$, and

$$\limsup_{t\to\infty} S_t(\bar X_1,\dots,\bar X_n) = \bar S(X_1,\dots,X_n).$$

The first result follows.

For the second conclusion, by (10) applied to $\bar Y_1(s),\dots,\bar Y_n(s)$, for all $i = 1,\dots,n$, it holds

$$\begin{aligned}
I(\bar Y_j(s), j\neq i) &\leqslant \sum_{j\neq i} \bar I_{jj}(s)\\
&= \left(\sum_{j=1}^n \bar I_{jj}(s)\right) - \bar I_{ii}(s).
\end{aligned}$$

Therefore, since $\sum_{i=1}^n \lambda_i^2 = 1$,

$$\sum_{i=1}^n \lambda_i^2 I(\bar Y_j, j\neq i) \leqslant \sum_{i=1}^n \bar I_{ii}(s) - \sum_{i=1}^n \lambda_i^2 \bar I_{ii}(s).$$

Accepted Manuscript - Version of Record at: https://doi.org/10.1109/TIT.2024.3385728

In turn,

$$\begin{aligned}\bar{S} &= 2\int_0^\infty \Big[\sum_{i,j=1}^n \lambda_i\lambda_j \bar{I}_{ij}(s) - \sum_{i=1}^n \bar{I}_{ii}(s) \\ &\qquad\qquad + \sum_{i=1}^n \lambda_i^2 I(\bar{Y}_j(s), j\neq i)\Big] ds \\ &\leqslant 2\int_0^\infty \sum_{\substack{i,j=1\\ i\neq j}}^n \lambda_i\lambda_j \bar{I}_{ij}(s) ds.\end{aligned}$$

The assumption $\bar{I}_{ij}(s) \leqslant 0$ ensures that $\bar{S} \leqslant 0$ leading to the desired entropy power inequality for conditional entropy. □

Remark V.1. *The reader might be surprised by the opposite expression of $\bar{R}$ in Corollary IV.2 and $\bar{S}$ in Corollary V.2. Indeed, for $n = 2$, $\bar{R}$ and $\bar{S}$ take the form (after some algebra and using that $\lambda_1^2 + \lambda_2^2 = 1$)*

$$\begin{aligned}\bar{R} = \frac{d}{2}\int_0^\infty \Big[&\lambda_1^2(\bar{I}_{11}(s) - I(\bar{Y}_1(s))) + \lambda_2^2(\bar{I}_{22}(s) - I(\bar{Y}_2(s))) \\ &+ 2\lambda_1\lambda_2 \bar{I}_{12}(t)\Big] dt\end{aligned}$$

while

$$\begin{aligned}\bar{S} = \frac{d}{2}\int_0^\infty \Big[&-\lambda_1^2(\bar{I}_{22}(s) - I(\bar{Y}_2(s))) - \lambda_2^2(\bar{I}_{11}(s) - I(\bar{Y}_1(s))) \\ &+ 2\lambda_1\lambda_2 \bar{I}_{12}(s)\Big] ds\end{aligned}$$

We emphasize the terms $\bar{I}_{ii}(t) - I(\bar{Y}_i(t))$ appear in both $\bar{R}$ and $\bar{S}$, but with opposite signs. This change arises from the decomposition formula of the entropy (12)*. The conditional entropy power inequality (Corollary V.2), leverages that the non-negativity of $\bar{I}_{ii}(t) - I(\bar{Y}_i(t))$ (see* (9)*) can be combined combined the assumption $I_{12}(s) \leqslant 0$, to directly yield $\bar{S} \leqslant 0$. Such an argument cannot be applied directly to guarantee the negativity of $\bar{R}$.*

In [24], it is claimed that, under the condition[2] *$I_{12}(s) \geqslant 0$, the following conditional entropy power inequality*

$$e^{2h(X_1+X_2)} \geqslant e^{2h(X_1|X_2)} + e^{2h(X_2|X_1)}$$

holds. This goes in the opposite direction of our condition $I_{12}(s) \leqslant 0$. In fact, as we will show in the appendix, with a counter-example, the statement in [24] is incorrect.

In what follows we give a sufficient condition on the density p of the random vector $(X_1, \ldots, X_n)$ for the condition $\bar{I}_{ij}(s) \leqslant 0$ in Corollary V.2 to hold for all $s > 0$. To that aim we need to introduce the notion of log-supermodular functions on $(\mathbb{R}^d)^n$.

Definition V.3. *A function $u\colon (\mathbb{R}^d)^n \to (0,\infty)$ is said to be log-supermodular if for all $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in (\mathbb{R}^d)^n$ it holds*

$$u(x)u(y) \leqslant u(x\wedge y)u(x\vee y)$$

where $x \vee y \in (\mathbb{R}^d)^n$ denotes the componentwise maximum of x and y and $x \wedge y \in (\mathbb{R}^d)^n$ denotes the componentwise minimum of x and y.

To be precise, if $x_i = (x_{i,1}, \ldots, x_{i,d})$, $i = 1, \ldots, n$, and similarly for y,

$$x \wedge y = (\min(x_{1,1}, y_{1,1}), \ldots, \min(x_{n,d}, y_{n,d}))$$

and

$$x \vee y = (\max(x_{1,1}, y_{1,1}), \ldots, \max(x_{n,d}, y_{n,d})).$$

The class of log-supermodular densities is widely studied in various field of mathematics; we refer to the introduction of [60] for an account of the literature and discussion. Note that, in some older literature, log-supermodular densities are called multivariate totally positive of order 2 (MTP$_2$).

It is known (see, e.g., [55] or [20, Proposition 2.5]) that a continuously twice-differentiable function u is log-supermodular if and only if $\frac{\partial^2}{\partial x_{i,k}\partial x_{j,l}} \log u \geqslant 0$ for all distinct $(i,k), (j,l)$, $i,j = 1, \ldots, n$, $k,l = 1\ldots, d$.

We will need to deal with log-supermodular densities that remain log-supermodular after convolution with a Gaussian with covariance matrix proportional to the identity matrix (*i.e.* of the form κI, with $\kappa > 0$ and I the identity matrix). We will call this class of densities $\mathcal{C}$. Namely, if g_s stands for the centered Gaussian density with covariance matrix sI_{nd} ($s > 0$) on $\mathbb{R}^{nd}$, and let $\mathcal{L}$ denote the space of log-supermodular density functions we denote by

$$\mathcal{C} := \{p \in \mathcal{L} : p * g_s \in \mathcal{L}, \forall s > 0\}.$$

As discussed in Karlin and Rinott [28] convolution of any two log-supermodular densities need not be log-supermodular (see [28, page 486] for a counterexample with two Gaussian densities). In [29] it is proved that a Gaussian density with covariance matrix Σ is log-supermodular if and only if the off-diagonal entries of Σ^{-1} are non-positive. In [60, Theorem 6] Zartash and Robeva give some conditions for a log-supermodular density p to belong to $\mathcal{C}$. Note that by a straightforward scaling argument, having $p * g_s$ log-supermodular for all $s > 0$ in the definition of $\mathcal{C}$ is equivalent to having $p * g_{s_o}$ log-supermodular for any specific $s_o > 0$. In particular $\mathcal{C} \neq \varnothing$. Moreover, they conjecture (see [60, Conjecture 10]) that, in fact, $\mathcal{C}$ coincides with the class of all log-supermodular densities.

We will prove the following Corollary.

Corollary V.4. *Let $p \in \mathcal{C}$ be a log-supermodular density of a random vector $X = (X_1, \ldots, X_n) \in (\mathbb{R}^d)^n$. Then it holds*

$$e^{\frac{2}{d}h(X_1+\cdots+X_n)} \geqslant e^{\frac{2}{d}h(X_1|X_i, i\neq 1)} + \cdots + e^{\frac{2}{d}h(X_n|X_i, i\neq n)}.$$

Proof. We observe first that, if $u \in \mathcal{C}$ is twice differentiable on $(\mathbb{R}^d)^n$, by definition,

$$\begin{aligned}P_t^* u((x)) &= \mathbb{E}[g_{\sqrt{1-e^{-2t}}}(x - e^{-t}X)] \\ &= \int_{\mathbb{R}^{nd}} u(y) g_{\sqrt{1-e^{-2t}}}(x - e^{-t}y) dy \\ &= e^{nt}\int_{\mathbb{R}^{nd}} u(z_1e^t, \ldots, z_ne^t) g_{\sqrt{1-e^{-2t}}}(x - z) dz \\ &= u_t * g_{\sqrt{1-e^{-2t}}}(x),\end{aligned}$$

---

[2] Strictly speaking, in [24], the author is dealing with the heat flow and not the Ornstein-Uhlenbeck flow. However this is just a matter of scaling and reformulation as there is a correspondence between the two flows.

where we set $u_t(z) := e^{nt}u(z_1e^t, \ldots, z_ne^t)$. Therefore, if $u \in \mathcal{C}$, u_t is log-supermodular, and by definition of $\mathcal{C}$, the convolution $u_t * g_{\sqrt{1-e^{-2t}}}$ is also log-supermodular. Equivalently this can be rephrased as: if $u \in \mathcal{C}$ is smooth enough, $P_t^* u$ is log-supermodular.

Our aim is now to prove that, with the notations of Corollary V.2, it holds

$$I_{ij}(\bar{Y}_1(s), \ldots, Y_n(s)) \leqslant 0$$

for all $i \neq j$, and to apply Corollary V.2. Since p is the density of $(X_1, \ldots, X_n)$, for $(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$,

$$\bar{p}(x_1, \ldots, x_n) = \lambda_1 \ldots \lambda_n p(\lambda_1 x_1, \ldots, \lambda_n x_n),$$

is the density of the random vector $(\bar{X}_1, \ldots, \bar{X}_n) = (\frac{X_1}{\lambda_1}, \ldots, \frac{X_n}{\lambda_n})$ (the λ_i's are defined in Corollary V.2, they satisfy $\lambda_i \in (0,1)$ and $\sum \lambda_i^2 = 1$). Denote by $\bar{p}_s$ the density of $(\bar{Y}_1(s), \ldots, \bar{Y}_n(s))$ where we recall that $\bar{Y}_i(s) = e^{-s}\bar{X}_i + \sqrt{1-e^{-2s}}Z_i$, with $Z_i \sim N(0, I)$ an independent standard Gaussian. We know that $\bar{p}_s = P_s^* \bar{p}$ and by the above observation that $\bar{p}_s$ is log-supermodular.

Recall that a smooth density u is log-supermodular if and only if $\frac{\partial^2}{\partial x_{i,k} \partial x_{j,l}} \log u \geqslant 0$ for all distinct $(i,k), (j,l)$, $i, j = 1, \ldots, n$, $k, l = 1, \ldots, d$. Hence, $\frac{\partial^2}{\partial x_{i,k} \partial x_{j,l}} \log \bar{p}_s \geqslant 0$ for all distinct $(i,k), (j,l)$.

Finally, integrating by parts, we get

$$\begin{aligned} I_{ij}(\bar{Y}_1(s), \ldots, \bar{Y}_n(s)) &= \int_{(\mathbb{R}^d)^n} \frac{\langle \nabla_i \bar{p}_s, \nabla_j \bar{p}_s \rangle}{\bar{p}_s} \\ &= \sum_{k=1}^{d} \int_{(\mathbb{R}^d)^n} \frac{\partial \log \bar{p}_s}{\partial x_{i,k}} \frac{\partial \bar{p}_s}{\partial x_{j,k}} \\ &= -\sum_{k=1}^{d} \int \frac{\partial^2 \log \bar{p}_s}{\partial x_{i,k} \partial x_{j,k}} \bar{p}_s \leqslant 0, \end{aligned}$$

which is the expected result. This ends the proof of the corollary by means of Corollary V.2. ☐

VI. ON THE CONDITIONAL ENTROPY POWER INEQUALITY OF JOHNSON

We work in dimension $d = 1$ and with $n = 2$. Let (X, Y) be a random vector and $(X_t, Y_t) = (X, Y) + (Z_1, Z_2)$ with $Z_1 \sim N(0, f(t))$, $Z_2 \sim N(0, g(t))$ are independent normal. In [24] f and g are implicitly defined in the proof of Theorem 6 and given by the relation

$$f'(t) = e^{2h(X_t|Y_t)}, \qquad g'(t) = e^{2h(Y_t|X_t)}. \tag{14}$$

Under the condition that, for all t, $I_{12}(X_t, Y_t) \geqslant 0$, [24, Theorem 6] states that

$$e^{2h(X+Y)} \geqslant e^{2h(X|Y)} + e^{2h(Y|X)}.$$

Writing (14) explicitly does not give a clear path towards proving the existence of solutions and we will not pursue their existence or non-existence of solutions. However, as we now demonstrate, this theorem is incorrect. For this, assume that (X, Y) is a random vector of Gaussian law $N(0, \Sigma)$ with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & -a \\ -a & 1 \end{pmatrix}, \qquad a \in (0,1).$$

Then,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N(0, \Sigma_t),$$

with

$$\Sigma_t := \begin{pmatrix} 1 + f(t) & -a \\ -a & 1 + g(t) \end{pmatrix},$$

where f and g satisfy (14), and where $Z_1 \sim N(0, f(t))$, $Z_2 \sim N(0, g(t))$ are independent. Note that by coordinate symmetry of the random variables $f(t) = g(t)$. In fact an explicit solution $f(t) = g(t) = \sqrt{a^2 + (1-a^2)e^{4\pi e t}} - 1$. is easily obtained, but we will not make use of this expression. Since (X_t, Y_t) is Gaussian a direct computation[3] yields the Fisher information matrix to be

$$\begin{aligned} \Sigma_t^{-1} &= \begin{pmatrix} I_{11}(X_t, Y_t) & I_{12}(X_t, Y_t) \\ I_{12}(X_t, Y_t) & I_{22}(X_t, Y_t) \end{pmatrix} \\ &= \frac{1}{(1 + f(t))^2 - a^2} \begin{pmatrix} 1 + f(t) & a \\ a & 1 + f(t) \end{pmatrix}. \end{aligned}$$

In particular, with our choice of $a \in (0,1)$, it holds

$$I_{12}(X_t, Y_t) = \frac{a}{(1 + f(t))^2 - a^2} > 0$$

for all $t \geqslant 0$. According to [24, Theorem 6] this should imply that

$$e^{2h(X+Y)} \geqslant e^{2h(X|Y)} + e^{2h(Y|X)}.$$

However since for a d-dimensional Gaussian Z with covariance Σ one has $h(Z) = \frac{d}{2} \log(2\pi e \det(\Sigma)^{\frac{1}{d}})$, we have

$$h(Y|X) = h(X|Y) = h(X, Y) - h(Y) = \frac{1}{2} \log(2\pi e(1 - a^2)),$$

and hence

$$\begin{aligned} e^{2h(X+Y)} &= 4\pi e(1 - a) < 4\pi e(1 - a)(1 + a) \\ &= e^{2h(X|Y)} + e^{2h(Y|X)}, \end{aligned}$$

and contradiction.

Acknowledgements: The authors thank Oliver Johnson for fruitful discussion and helpful comments, as well as two anonymous reviewers whose careful reading and suggestions improved the quality of this paper.

---

[3] We mention that this identity corresponds to equality in the multivariate Cramer-Rao (see for example [47, Equation 29.4]), that

$$\Sigma_\theta(\hat{\theta}) \succeq I^{-1}(\theta)$$

where $\succeq$ refers to order as positive definite matrices and $\Sigma_\theta(\hat{\theta})$ is the covariance of $\hat{\theta}$ an estimator of a parameter $\theta \in \Theta \subseteq \mathbb{R}^d$, and I^{-1} is the inverse of the usual Fisher information matrix associated to the parameter θ. Taking a location model, so that $\theta \in \mathbb{R}^d$ is the mean of a variable X and $\hat{\theta} = X$, the Fisher information matrix (and the covariance matrix as well) is independent of θ and we recover the notion defined here, the inequality reads as the covariance of X, $\Sigma \succeq I^{-1}$. Knowing that the Gaussian location model satisfies an exact equality in Cramer-Rao is $\Sigma = I^{-1}$, or as utilized above, one has the Fisher information matrix $I = \Sigma^{-1}$.

Accepted Manuscript - Version of Record at: https://doi.org/10.1109/TIT.2024.3385728

REFERENCES

[1] E. Abbe, J. Li, and M. Madiman. Entropies of weighted sums in cyclic groups and an application to polar codes. *Entropy, Special issue on "Entropy and Information Inequalities" (edited by V. Jog and J. Melbourne)*, 19(9), September 2017.

[2] V. Anantharam, V. Jog, and C. Nair. Unifying the Brascamp-Lieb inequality and the entropy power inequality. *IEEE Trans. Inform. Theory*, 68(12):7665–7684, 2022.

[3] S. Artstein, K. M. Ball, F. Barthe, and A. Naor. Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc.*, 17(4):975–982 (electronic), 2004.

[4] K. Ball, P. Nayar, and T. Tkocz. A reverse entropy power inequality for log-concave random vectors. *Studia Math.*, 235(1):17–30, 2016.

[5] N.M. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Information Theory*, IT-11:267–271, 1965.

[6] S. Bobkov and M. Madiman. Dimensional behaviour of entropy and information. *C. R. Acad. Sci. Paris Sér. I Math.*, 349:201–204, Février 2011.

[7] S. Bobkov and M. Madiman. Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures. *J. Funct. Anal.*, 262:3309–3339, 2012.

[8] S. G. Bobkov and G. P. Chistyakov. Bounds for the maximum of the density of the sum of independent random variables. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 408(Veroyatnost i Statistika. 18):62–73, 324, 2012.

[9] S. G. Bobkov and G. P. Chistyakov. Entropy power inequality for the Rényi entropy. *IEEE Trans. Inform. Theory*, 61(2):708–714, February 2015.

[10] S. G. Bobkov and M. M. Madiman. On the problem of reversibility of the entropy power inequality. In *Limit theorems in probability, statistics and number theory*, volume 42 of *Springer Proc. Math. Stat.*, pages 61–74. Springer, Heidelberg, 2013.

[11] S. G. Bobkov and A. Marsiglietti. Variants of the entropy power inequality. *IEEE Trans. Inform. Theory*, 63(12):7747–7752, 2017.

[12] S. G. Bobkov, A. Marsiglietti, and J. Melbourne. Concentration functions and entropy bounds for discrete log-concave distributions. *Combinatorics, Probability and Computing*, 31(1):54–72, 2022.

[13] E. A. Carlen. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 101(1):194–211, 1991.

[14] E. A. Carlen and A. Soffer. Entropy production by block variable summation and central limit theorems. *Comm. Math. Phys.*, 140, 1991.

[15] D. Cordero-Erausquin and M. Ledoux. The geometry of Euclidean convolution inequalities and entropy. *Proc. Amer. Math. Soc.*, 138(8):2755–2769, 2010.

[16] T. A. Courtade. A strong entropy power inequality. *IEEE Trans. Inform. Theory*, 64(4, part 1):2173–2192, 2018.

[17] T. A. Courtade, M. Fathi, and A. Pananjady. Quantitative stability of the entropy power inequality. *IEEE Trans. Inform. Theory*, 64(8):5691–5703, 2018.

[18] A. Dembo, T.M. Cover, and J.A. Thomas. Information-theoretic inequalities. *IEEE Trans. Inform. Theory*, 37(6):1501–1518, 1991.

[19] M. Fradelizi, M. Madiman, A. Marsiglietti, and A. Zvavitch. The convexification effect of Minkowski summation. *EMS Surveys in Mathematical Sciences*, 5(1/2):1–64, 2019.

[20] M. Fradelizi, M. Madiman, and A. Zvavitch. Sumset estimates in convex geometry. *Preprint,* `arXiv:2206.01565`, 2022.

[21] R. J. Gardner. The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, 39(3):355–405 (electronic), 2002.

[22] S. Haghighatshoar, E. Abbe, and E. Telatar. A new entropy power inequality for integer-valued random variables. *IEEE Trans. Inform. Th.*, 60(7):3787–3796, July 2014.

[23] P. Harremoës and C. Vignat. An entropy power inequality for the binomial family. *J. Inequal. Pure Appl. Math.*, 4(5):Article 93, 6 pp. (electronic), 2003.

[24] O. Johnson. A conditional entropy power inequality for dependent variables. *IEEE Trans. Inform. Theory*, 50(8):1581–1583, 2004.

[25] O. Johnson. *Information theory and the central limit theorem*. Imperial College Press, London, 2004.

[26] O. Johnson. An information-theoretic central limit theorem for finitely susceptible FKG systems. *Teor. Veroyatn. Primen.*, 50(2):331–343, 2005.

[27] O. Johnson and Y. Yu. Monotonicity, thinning, and discrete versions of the entropy power inequality. *IEEE Trans. Inform. Theory*, 56(11):5387–5395, 2010.

[28] S. Karlin and Y. Rinott. Classes of orderings of measures and related correlation inequalities. II. Multivariate reverse rule distributions. *J. Multivariate Anal.*, 10(4):499–516, 1980.

[29] S. Karlin and Y. Rinott. M-matrices as covariance matrices of multinormal distributions. *Linear Algebra Appl.*, 52/53:419–438, 1983.

[30] J. Li. Rényi entropy power inequality and a reverse. *Studia Math.*, 242:303–319, 2018.

[31] J. Li, A. Marsiglietti, and J. Melbourne. Further investigations of rényi entropy power inequalities and an entropic characterization of s-concave densities. In B. Klartag and E. Milman, editors, *Geometric aspects of functional analysis: Israel Seminar (GAFA) 2017-2019*, volume 2266 of *Lecture Notes in Mathematics*, pages 95–123. Springer, 2020.

[32] J. Liu, T. A. Courtade, P. W. Cuff, and S. Verdú. A forward-reverse Brascamp-Lieb inequality: entropic duality and Gaussian optimality. *Entropy*, 20(6):Paper No. 418, 32, 2018.

[33] M. Madiman. A primer on entropic limit theorems. *Preprint*, 2024.

[34] M. Madiman and A.R. Barron. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory*, 53(7):2317–2329, July 2007.

[35] M. Madiman and F. Ghassemi. Combinatorial entropy power inequalities: A preliminary study of the Stam region. *IEEE Trans. Inform. Theory*, 65(3):1375–1386, March 2019.

[36] M. Madiman and I. Kontoyiannis. Entropy bounds on abelian groups and the Ruzsa divergence. *IEEE Trans. Inform. Theory*, 64(1):77–92, January 2018.

[37] M. Madiman, J. Melbourne, and C. Roberto. Bernoulli sums and Rényi entropy inequalities. *Bernoulli*, 29(2):1578–1599, May 2023.

[38] M. Madiman, J. Melbourne, and C. Roberto. Preservation of log-supermodularity on convolution. *Preprint*, 2024.

[39] M. Madiman, J. Melbourne, and P. Xu. Forward and reverse entropy power inequalities in convex geometry. In E. Carlen, M. Madiman, and E. M. Werner, editors, *Convexity and Concentration*, volume 161 of *IMA Volumes in Mathematics and its Applications*, pages 427–485. Springer, 2017.

[40] M. Madiman, J. Melbourne, and P. Xu. Rogozin's convolution inequality for locally compact groups. *Preprint,* `arXiv:1705.00642`, 2017.

[41] M. Madiman, P. Nayar, and T. Tkocz. Two remarks on generalized entropy power inequalities. In B. Klartag and E. Milman, editors, *Geometric aspects of functional analysis: Israel Seminar (GAFA) 2017-2019*, volume 2266 of *Lecture Notes in Mathematics*, pages 169–185. Springer, 2020.

[42] M. Madiman, L. Wang, and J. O. Woo. Majorization and Rényi entropy inequalities via Sperner theory. *Discrete Math.*, 342(10):2911–2923, October 2019.

[43] M. Madiman, L. Wang, and J. O. Woo. Rényi entropy inequalities for sums in prime cyclic groups. *SIAM J. Discrete Math.*, 35(3):1628–1649, 2021.

[44] A. Marsiglietti and J. Melbourne. On the entropy power inequality for the Rényi entropy of order [0, 1]. *IEEE Trans. Inform. Theory*, 65(3):1387–1396, 2019.

[45] J. Melbourne and G. Palafox-Castillo. A discrete complement of Lyapunov's inequality and its information theoretic consequences. *Ann. Appl. Probab.*, 33(6A):4322–4340, 2023.

[46] J. Melbourne and T. Tkocz. Reversal of Rényi entropy inequalities under log-concavity. *IEEE Transactions on Information Theory*, 67(1):45–51, 2020.

[47] Y. Polyanskiy and Y. Wu. Information theory from coding to learning. Textbook draft available at https://people.lids.mit.edu/yp/homepage/papers.html, 2022+.

[48] E. Ram and I. Sason. On Rényi entropy power inequalities. *IEEE Trans. Inform. Theory*, 62(12):6800–6815, 2016.

[49] O. Rioul. Rényi entropy power inequalities via normal transport and rotation. *Entropy*, 20(9):641, 2018.

[50] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[51] A. J. Stam. *Some mathematical properties of quantities of information*. PhD thesis, Technische Universiteit Delft, April 1959.

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

[53] S. Takano. The inequalities of Fisher information and entropy power for dependent variables. In *Probability theory and mathematical statistics (Tokyo, 1995)*, pages 460–470. World Sci. Publ., River Edge, NJ, 1996.

[54] S. Takano. Entropy and a limit theorem for some dependent variables. In *Proceedings of Prague Stochastics '98*, volume 2, pages 549–552. Union of Czech Mathematicians and Physicists, 1998.

[55] D. M. Topkis. *Supermodularity and complementarity*. Frontiers of Economic Research. Princeton University Press, Princeton, NJ, 1998.

[56] G. Toscani. A Strengthened Entropy Power Inequality for Log-Concave Densities. *IEEE Trans. Inform. Theory*, 61(12):6550–6559, 2015.

[57] L. Wang and M. Madiman. Beyond the entropy power inequality, via rearrangements. *IEEE Trans. Inform. Theory*, 60(9):5116–5137, September 2014.

[58] J. O. Woo and M. Madiman. A discrete entropy power inequality for uniform distributions. In *Proc. IEEE Intl. Symp. Inform. Theory*, pages 1625–1629, Hong Kong, China, June 2015.

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

[60] A. Zartash and E. Robeva. Convolutions of totally positive distributions with applications to kernel density estimation. *Algebr. Stat.*, 13(1):57–79, 2022.

Mokshay Madiman received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 1999, and the Sc.M. and Ph.D. degrees in applied mathematics from Brown University, Providence, RI, USA, in 2001 and 2005, respectively. From 2005 to 2012, he was with the Department of Statistics, Yale University, New Haven, CT, USA, first as a Gibbs Assistant Professor, then as an Assistant Professor, and finally as an Associate Professor of statistics and applied mathematics. He has spent a semester each visiting the School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, the National Mathematics Initiative, Indian Institute of Science, Bengaluru, the Department of Operations Research and Financial Engineering, Princeton University, and the Institute for Mathematics and Its Applications, Minneapolis. Since January 2013, he has been an Associate Professor with the Department of Mathematical Sciences, University of Delaware, and from 2014 to 2017, he was also an Adjunct Professor of Mathematics with the Tata Institute of Fundamental Research, India. His research interests include probability theory, information theory, combinatorics, functional analysis, and convex geometry. He was awarded the NSF CAREER Award in 2011.

James Melbourne received his PhD in Mathematics in 2015 at the University of Minnesota. He was a postdoctoral researcher in the University of Delaware Mathematics department from 2015 to 2017, and a postdoctoral researcher at the University of Minnesota in Electrical and Computer Engineering from 2017-2020. He is currently an assistant professor at CIMAT in the department of Probability and Statistics. His research focuses on convexity theory in the context of probabilistic, geometric, and information theoretic inequalities as well as their applications.

Cyril Roberto received his PhD in Mathematics in 2001 at the University of Toulouse. He was postdoctoral researcher in Rome 3 University in 2002. From 2002 to 2011 he was assistant professor with the mathematics department, Marne la Vallée University, and is Professor at Paris Nanterre University since 2011. He has spent a semester each visiting the Imperial College of London (2006), Roma 3 University under a ERC grant (2008 and 2009) and CIMAT in Guanajuato, Mexico (2022). His research interests range from Probability Theory to Functional Analysis, Convex Geometry and Statistical Mechanics (interacting particles systems out of equilibrium).