On Rényi Entropy Power Inequalities

Eshed Ram Igal Sason

Abstract

This paper gives improved Rényi entropy power inequalities (R-EPIs). Consider a sum $S_n = \sum_{k=1}^n X_k$ of n independent continuous random vectors taking values on $\mathbb{R}^d$, and let $\alpha \in [1, \infty]$. An R-EPI provides a lower bound on the order-α Rényi entropy power of S_n that, up to a multiplicative constant (which may depend in general on n, α, d), is equal to the sum of the order-α Rényi entropy powers of the n random vectors $\{X_k\}_{k=1}^n$. For $\alpha = 1$, the R-EPI coincides with the well-known entropy power inequality by Shannon. The first improved R-EPI is obtained by tightening the recent R-EPI by Bobkov and Chistyakov which relies on the sharpened Young's inequality. A further improvement of the R-EPI also relies on convex optimization and results on rank-one modification of a real-valued diagonal matrix.

Keywords: Rényi entropy, entropy power inequality, Rényi entropy power.

I. Introduction

One of the well-known inequalities in information theory is the entropy power inequality (EPI) which has been introduced by Shannon [41, Theorem 15]. Let X be a d-dimensional random vector with a probability density function, let $h(X)$ be its differential entropy, and let $N(X) = \exp\left(\frac{2}{d}\, h(X)\right)$ be the entropy power of X. The EPI states that for independent random vectors $\{X_k\}_{k=1}^n$, the following inequality holds:

$$N\left(\sum_{k=1}^n X_k\right) \geq \sum_{k=1}^n N(X_k) \tag{1}$$

with equality in (1) if and only if $\{X_k\}_{k=1}^n$ are Gaussian random vectors with proportional covariances.

The EPI has proved to be an instrumental tool in proving converse theorems for the capacity region of the Gaussian broadcast channel [6], the Gaussian wire-tap channel [30], the capacity region of the Gaussian broadcast multiple-input multiple-output (MIMO) channel [49], and a converse theorem in multi-terminal lossy compression [35]. Due to its importance, the EPI has been proved with information-theoretic tools in several insightful ways (see, e.g., [7], [18], [22], [27, Appendix D], [37], [44], [46]); e.g., the proof in [46] relies on fundamental relations between information and estimation measures ([21], [23]), together with the simple fact that for estimating a sum of two random variables, it is preferable to have access to the individual noisy measurements rather than to their sum. More studies on the theme include EPIs for discrete random variables and some analogies [24], [25], [26], [29], [40], [42], [50], generalized EPIs [31], [32], [52], reverse EPIs [10], [11], [34], [51], related inequalities to the EPI in terms of rearrangements [47], and some refined versions of the EPI for specialized distributions [15], [16], [25], [45]. An overview on EPIs is provided in [1]; we also refer the reader to a preprint

E. Ram and I. Sason are with the Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel. E-mails: {s6eshedr@tx, sason@ee}.technion.ac.il.

This work has been supported by the Israeli Science Foundation (ISF) under Grant 12/12. It has been presented in part at the 2016 IEEE International Symposium on Information Theory, Barcelona, Spain, July 10–15, 2016.

of a recent survey paper by Madiman *et al.* [34] which addresses forward and reverse EPIs with Rényi measures, and their connections with convex geometry.

The Rényi entropy and divergence have been introduced in [36], and they evidence a long track record of usefulness in information theory and its applications. Recent studies of the properties of these Rényi measures have been provided in [19], [20] and [43]. In the following, the differential Rényi entropy and the Rényi entropy power are introduced.

Definition 1 (Differential Rényi entropy): Let X be a random vector which takes values in $\mathbb{R}^d$, and assume that it has a probability density function which is designated by f_X. The differential Rényi entropy of X of order $\alpha \in (0,1) \cup (1,\infty)$, denoted by $h_\alpha(X)$, is given by

$$h_\alpha(X) = \frac{1}{1-\alpha}\ \log\left(\int_{\mathbb{R}^d} f_X^\alpha(x)\,\mathrm{d}x\right) \tag{2}$$

$$= \frac{\alpha}{1-\alpha}\ \log \|f_X\|_\alpha. \tag{3}$$

The differential Rényi entropies of orders $\alpha = 0, 1, \infty$ are defined by the continuous extension of $h_\alpha(X)$ for $\alpha \in (0,1) \cup (1,\infty)$, which yields

$$h_0(X) = \log\, \lambda\big(\mathrm{supp}(f_X)\big), \tag{4}$$

$$h_1(X) = h(X) = -\int_{\mathbb{R}^d} f_X(x)\, \log f_X(x)\,\mathrm{d}x, \tag{5}$$

$$h_\infty(X) = -\log\big(\operatorname{ess\,sup}(f_X)\big) \tag{6}$$

where λ in (4) is the Lebesgue measure in $\mathbb{R}^d$.

Definition 2 (Rényi entropy power): For a d-dimensional random vector X with density, the Rényi entropy power of order $\alpha \in [0,\infty]$ is given by

$$N_\alpha(X) = \exp\left(\tfrac{2}{d}\, h_\alpha(X)\right). \tag{7}$$

Since $h_\alpha(X)$ is specialized to the Shannon entropy $h(X)$ for $\alpha = 1$, the possibility of generalizing the EPI with Rényi entropy powers has emerged, leading to the following question:

Question 1: Let $\{X_k\}$ be independent d-dimensional random vectors with probability density functions, and let $\alpha \in [0,\infty]$ and $n \in \mathbb{N}$. Does a Rényi entropy power inequality (R-EPI) of the form

$$N_\alpha\left(\sum_{k=1}^n X_k\right) \geq c_\alpha^{(n,d)}\ \sum_{k=1}^n N_\alpha(X_k) \tag{8}$$

hold for some positive constant $c_\alpha^{(n,d)}$ (which may depend on the order α, dimension d, and number of summands n) ?

In [28, Theorem 2.4], a sort of an R-EPI for the Rényi entropy of order $\alpha \geq 1$ has been derived with some analogy to the classical EPI; this inequality, however, does not apply the usual convolution unless $\alpha = 1$. In [47, Conjectures 4.3, 4.4], Wang and Madiman conjectured an R-EPI for an arbitrary finite number of independent random vectors in $\mathbb{R}^d$ for $\alpha > \frac{d}{d+2}$.

Question 1 has been recently addressed by Bobkov and Chistyakov [9], showing that (8) holds with

$$c_\alpha = \tfrac{1}{e}\, \alpha^{\frac{1}{\alpha-1}}, \quad \forall\, \alpha > 1 \tag{9}$$

independently of the values of n, d. It is the purpose of this paper to derive some improved R-EPIs for $\alpha > 1$ (the case of $\alpha = 1$ refers to the EPI (1)). A study of Question 1 for $\alpha \in (0,1)$ is currently an open problem (see [9, p. 709]).

In view of the close relation in (3) between the (differential) Rényi entropy and the L_α norm, the sharpened version of Young's inequality plays a key role in [9] for the derivation of an R-EPI, as well as in our paper for the derivation of some improved R-EPIs. The sharpened version of Young's inequality was also used by Dembo *et al.* [18] for proving the EPI.

For $\alpha \in (1, \infty)$, let $\alpha' = \frac{\alpha}{\alpha-1}$ be Hölder's conjugate. For $\alpha > 1$, Theorem 1 provides a new tighter constant in comparison to (9) which gets the form

$$c_\alpha^{(n)} = \alpha^{\frac{1}{\alpha-1}} \left(1 - \frac{1}{n\alpha'}\right)^{n\alpha'-1} \tag{10}$$

independently of the dimension d. The new R-EPI with the constant in (10) asymptotically coincides with the tight bound by Rogozin [38] when $\alpha \to \infty$ and $n = 2$, and it also asymptotically coincides with the R-EPI in [9] when $n \to \infty$. Moreover, the R-EPI with the new constant in (10) is further improved in Theorem 2 by a more involved analysis which relies on convex analysis and some interesting results from matrix theory; the latter result yields a closed-form solution for $n = 2$.

This paper is organized as follows: In Section II, preliminary material and notation are introduced. A new R-EPI is derived in Section III for $\alpha > 1$, and special cases of this improved bound are studied. Section IV derives a strengthened R-EPI for a sum of $n \geq 2$ random variables; for $n = 2$, it is specialized to a bound which is expressed in a closed form; its computation for $n > 2$ requires a numerical optimization which is easy to perform. Section V exemplifies numerically the tightness of the new R-EPIs in comparison to some previously reported bounds, and finally Section VI summarizes the paper.

II. Analytical Tools

This section includes notation and tools which are essential to the analysis in this paper. It starts with the sharpened Young's inequality, followed by results on rank-one modification of a symmetric eigenproblem [14]. We also include here some properties of the differential Rényi entropy and Rényi entropy power which are useful to the analysis in this paper.

A. Basic Inequalities

The derivation of the R-EPIs in this work partially relies on the sharpened Young's inequality and the monotonicity of the Rényi entropy in its order. For completeness, we introduce these results in the following.

Notation 1: For $\alpha > 0$, let $\alpha' = \frac{\alpha}{\alpha-1}$, i.e., $\frac{1}{\alpha} + \frac{1}{\alpha'} = 1$; if $\alpha = 1$, we define $\alpha' = \infty$. Note that $\alpha \in [1, \infty)$ if and only if $\alpha' \in [0, \infty]$. This notation is known as Hölder's conjugate.

Fact 1 (Monotonicity of the Rényi entropy): The Rényi entropy, $h_\alpha(X)$, is monotonically non-increasing in α.

From (3), it follows that for $\alpha \in (0,1) \cup (1, \infty)$, if f is a probability density function of a d-dimensional vector X, then

$$h_\alpha(X) = -\log\bigl(\|f\|_\alpha^{\alpha'}\bigr). \tag{11}$$

A useful consequence of Fact 1 and (11) is the following result (a weaker version of it is given in [9, Lemma 1]):

Corollary 1: Let $\alpha \in (0,1) \cup (1,\infty)$, and let $f \in L^{\alpha}(\mathbb{R}^d)$ be a probability density function (i.e., f is a non-negative function with $\|f\|_1 = 1$). Then, for every $\beta \in (0,\alpha)$ with $\beta \neq 1$,

$$\|f\|_{\beta}^{\beta'} \leq \|f\|_{\alpha}^{\alpha'}. \tag{12}$$

Notation 2: For every $t \in (0,1) \cup (1,\infty)$, let

$$A_t = t^{\frac{1}{t}} \, |t'|^{-\frac{1}{|t'|}} \tag{13}$$

and let $A_1 = A_\infty = 1$. Note that for $t \in (0,\infty]$

$$A_{t'} = \frac{1}{A_t}. \tag{14}$$

The sharpened Young's inequality, first derived by Beckner [4] and re-derived with alternative proofs in, e.g., [3] and [13] is given as follows:

Fact 2 (Sharpened Young's inequality): Let $p, q, r > 0$ satisfy

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}, \tag{15}$$

let $f \in L^p(\mathbb{R}^d)$ and $g \in L^q(\mathbb{R}^d)$ be non-negative functions, and let $f * g$ denote their convolution.

- If $p, q, r > 1$, then

$$\|f * g\|_r \leq \left(\frac{A_p A_q}{A_r}\right)^{\frac{d}{2}} \|f\|_p \, \|g\|_q. \tag{16}$$

- If $p, q, r < 1$, then

$$\|f * g\|_r \geq \left(\frac{A_p A_q}{A_r}\right)^{\frac{d}{2}} \|f\|_p \, \|g\|_q. \tag{17}$$

Furthermore, (16) and (17) hold with equalities if and only if f and g are Gaussian probability density functions (up to multiplicative constants).

Note that the condition in (15) can be expressed in terms of the Hölder's conjugates as follows:

$$\frac{1}{p'} + \frac{1}{q'} = \frac{1}{r'}. \tag{18}$$

By using (18) and mathematical induction, the sharpened Young's inequality can be extended to more than two functions as follows:

Corollary 2: Let $\nu, \{\nu_k\}_{k=1}^n > 0$ satisfy $\sum_{k=1}^n \frac{1}{\nu_k'} = \frac{1}{\nu'}$, let

$$A = \left(\frac{1}{A_\nu} \prod_{k=1}^n A_{\nu_k}\right)^{\frac{d}{2}} \tag{19}$$

where the right side in (19) is defined by (13), and let $f_k \in L^{\nu_k}(\mathbb{R}^d)$ be non-negative functions.

- If $\nu, \{\nu_k\}_{k=1}^n > 1$, then

$$\|f_1 * \ldots * f_n\|_\nu \leq A \prod_{k=1}^n \|f_k\|_{\nu_k}. \tag{20}$$

- If $\nu, \{\nu_k\}_{k=1}^n < 1$, then

$$\|f_1 * \ldots * f_n\|_\nu \geq A \prod_{k=1}^n \|f_k\|_{\nu_k} \tag{21}$$

with equalities in (20) and (21) if and only if f_k are scaled versions of Gaussian probability densities (up to multiplicative constants) for all k.

B. Rank-One Modification of a Symmetric Eigenproblem

This section is based on a paper by Bunch *et al.* [14] which addresses the eigenvectors and eigenvalues (a.k.a. eigensystem) of rank-one modification of a real-valued diagonal matrix. We use in this paper the following result [14]:

Fact 3: Let $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix with the eigenvalues $d_1 \leq d_2 \leq \ldots \leq d_n$. Let $z \in \mathbb{R}^n$ such that $\|z\|_2 = 1$ and let $\rho \in \mathbb{R}$. Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be the eigenvalues of the rank-one modification of D which is given by $C = D + \rho z z^T$. Then,

1) $\lambda_i = d_i + \rho \mu_i$, where $\sum_{i=1}^n \mu_i = 1$ and $\mu_i \geq 0$ for all $i \in \{1, \ldots, n\}$.
2) If $\rho > 0$, then the following interlacing property holds:

$$d_1 \leq \lambda_1 \leq d_2 \leq \lambda_2 \leq \ldots \leq d_n \leq \lambda_n \tag{22}$$

and, if $\rho < 0$, then

$$\lambda_1 \leq d_1 \leq \lambda_2 \leq d_2 \leq \ldots \leq \lambda_n \leq d_n. \tag{23}$$

3) If all the eigenvalues of D are different, all the entries of z are non-zero, and $\rho \neq 0$, then inequalities (22) and (23) are strict. For $i \in \{1, \ldots, n\}$, the eigenvalue λ_i is a zero of

$$W(x) = 1 + \rho \sum_{j=1}^n \frac{z_i^2}{d_j - x}. \tag{24}$$

Note that the requirement $\|z\|_2 = 1$ can be relaxed to $z \neq 0$ by letting $\hat{z} = \frac{z}{\|z\|_2}$ and $\hat{\rho} = \rho \|z\|_2^2$.

C. Rényi Entropy Power

We present some properties of the differential Rényi entropy and Rényi entropy power which are useful in this paper.

- In view of (3) and (7), for $\alpha \in (0,1) \cup (1, \infty)$,

$$N_\alpha(X) = \left(\|f_X\|_\alpha\right)^{-\frac{2\alpha'}{d}}. \tag{25}$$

- The differential Rényi entropy $h_\alpha(X)$ is monotonically non-increasing in α, and so is $N_\alpha(X)$.
- If $Y = AX + b$ where $A \in \mathbb{R}^{d \times d}, |A| \neq 0$, $b \in \mathbb{R}^d$, then for all $\alpha \in [0, \infty]$

$$h_\alpha(Y) = h_\alpha(X) + \log |A|, \tag{26}$$

$$N_\alpha(Y) = |A|^{\frac{2}{d}} \, N_\alpha(X). \tag{27}$$

This implies that the Rényi entropy power is a homogeneous functional of order 2 and it is translation invariant, i.e.,

$$N_\alpha(\lambda X) = \lambda^2 \, N_\alpha(X), \quad \forall \, \lambda \in \mathbb{R}, \tag{28}$$

$$N_\alpha(X + b) = N_\alpha(X), \quad \forall \, b \in \mathbb{R}^d. \tag{29}$$

In view of (28) and (29), $N_\alpha(X)$ has some similar properties to the variance of X. However, if we consider a sum of independent random vectors then $\mathrm{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \mathrm{Var}(X_k)$ whereas the Rényi entropy power of a sum of independent random vectors is not equal, in general, to the sum of the Rényi entropy powers of the individual random vectors (unless these independent vectors are Gaussian with proportional covariances).

The continuation of this paper considers R-EPIs for orders $\alpha \in (1, \infty]$. The case where $\alpha = 1$ refers to the EPI by Shannon [41, Theorem 15].

III. A New Rényi EPI

In the following, a new R-EPI is derived. This inequality, which is expressed in closed-form, is tighter than the R-EPI in [9, Theorem I.1].

Theorem 1: Let $\{X_k\}_{k=1}^n$ be independent random vectors with densities defined on $\mathbb{R}^d$, and let $n \in \mathbb{N}$, $\alpha > 1$, $\alpha' = \frac{\alpha}{\alpha-1}$ and $S_n = \sum_{k=1}^n X_k$. Then, the following R-EPI holds:

$$N_\alpha(S_n) \geq c_\alpha^{(n)} \sum_{k=1}^n N_\alpha(X_k) \tag{30}$$

with

$$c_\alpha^{(n)} = \alpha^{\frac{1}{\alpha-1}} \left(1 - \frac{1}{n\alpha'}\right)^{n\alpha'-1}. \tag{31}$$

Furthermore, the R-EPI in (30) has the following properties:

1) Eq. (30) improves the R-EPI in [9, Theorem I.1] for every $\alpha > 1$ and $n \in \mathbb{N}$,
2) For all $\alpha > 1$, it asymptotically coincides with the R-EPI in [9, Theorem I.1] as $n \to \infty$,
3) In the other limiting case where $\alpha \downarrow 1$, it coincides with the EPI (similarly to [9]),
4) If $n = 2$ and $\alpha \to \infty$, the constant $c_\alpha^{(n)}$ in (31) tends to $\frac{1}{2}$ which is optimal; this constant is achieved when X_1 and X_2 are independent random vectors which are uniformly distributed in the cube $[0,1]^d$.

Proof: In the first stage of this proof, we assume that

$$N_\alpha(X_k) > 0, \quad k \in \{1, \ldots, n\} \tag{32}$$

which, in view of (25), implies that $f_{X_k} \in L^\alpha(\mathbb{R}^d)$, where f_{X_k} is the density of X_k for all $k \in \{1, \ldots, n\}$. In [9, (12)] it is shown that for $\alpha > 1$,

$$N_\alpha(S_n) \geq B \prod_{k=1}^n N_\alpha^{t_k}(X_k) \tag{33}$$

with

$$B = \left(A_{\nu_1} \ldots A_{\nu_n} A_{\alpha'}\right)^{-\alpha'}, \tag{34}$$

$$\nu_k > 1, \quad \forall\, k \in \{1, \ldots, n\}, \tag{35}$$

$$\nu' = \frac{\nu}{\nu - 1}, \quad \forall\, \nu \in \mathbb{R}, \tag{36}$$

$$\sum_{k=1}^n \frac{1}{\nu_k'} = \frac{1}{\alpha'}, \tag{37}$$

$$t_k = \frac{\alpha'}{\nu_k'}, \quad \forall\, k \in \{1, \ldots, n\}. \tag{38}$$

Consequently, (35)–(38) yields

$$t_k \geq 0, \quad \forall\, k \in \{1, \ldots, n\}, \tag{39}$$

$$\sum_{k=1}^n t_k = 1. \tag{40}$$

The proof of (33), which relies on Corollaries 1 and 2, is introduced in Appendix A.

Similarly to [9, (14)], in view of the homogeneity of the entropy power functional (see (28)), it can be assumed without any loss of generality that

$$\sum_{k=1}^{n} N_\alpha(X_k) = 1. \tag{41}$$

Hence, to prove (30), it is sufficient to show that under the assumption in (41)

$$N_\alpha(S_n) \geq c_\alpha^{(n)}. \tag{42}$$

From this point, we deviate from the proof of [9, Theorem I.1]. Taking logarithms on both sides of (33) and assembling (13), (34)–(40) and (41) yield

$$\log N_\alpha(S_n) \geq f_0(\underline{t}), \tag{43}$$

where $\underline{t} = (t_1, \ldots, t_n)$, and

$$f_0(\underline{t}) = \frac{\log \alpha}{\alpha - 1} - D(\underline{t} \| \underline{N}_\alpha) + \alpha' \sum_{k=1}^{n} \left(1 - \frac{t_k}{\alpha'}\right) \log \left(1 - \frac{t_k}{\alpha'}\right), \tag{44}$$

$$\underline{N}_\alpha = \left(N_\alpha(X_1), \ldots, N_\alpha(X_n)\right), \tag{45}$$

$$D(\underline{t} \| \underline{N}_\alpha) = \sum_{k=1}^{n} t_k \log \left(\frac{t_k}{N_\alpha(X_k)}\right). \tag{46}$$

In view of (39) and (40), the bound in (43) holds for every $\underline{t} \in \mathbb{R}_+^n$ such that $\sum_{k=1}^{n} t_k = 1$. Consequently, the R-EPI in [9, Theorem I.1] can be tightened by maximizing the right side of (43), leading to the following optimization problem:

$$\begin{array}{ll} \text{maximize} & f_0(\underline{t}) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \ldots, n\}, \\ & \sum_{k=1}^{n} t_k = 1. \end{array} \tag{47}$$

Note that the convexity of the function

$$f(x) = \left(1 - \frac{x}{\alpha'}\right) \log \left(1 - \frac{x}{\alpha'}\right), \quad x \in [0, \alpha'] \tag{48}$$

yields that the third term on the right side of (44) is convex in $\underline{t}$. Since the relative entropy $D(\underline{t} \| \underline{N}_\alpha)$ is also convex in $\underline{t}$, the objective function f_0 in (44) is expressed as a difference of two convex functions in $\underline{t}$. In order to get an analytical closed-form lower bound on the solution of the optimization problem in (47), we take the sub-optimal choice $\underline{t} = \underline{N}_\alpha$ (similarly to the proof [9, Theorem I.1]) which yields that $D(\underline{t} \| \underline{N}_\alpha) = 0$; however, our proof derives an improved lower bound on the third term of $f_0(\underline{t})$ which needs to be independent of $\underline{N}_\alpha$. Let

$$\hat{t}_k = N_\alpha(X_k), \quad 1 \leq k \leq n, \tag{49}$$

then, in view of (43) and (49),

$$\begin{aligned} \log N_\alpha(S_n) &\geq f_0(\hat{\underline{t}}) &(50)\\ &= \frac{\log \alpha}{\alpha - 1} + \alpha' \sum_{k=1}^{n} \left(1 - \frac{\hat{t}_k}{\alpha'}\right) \log \left(1 - \frac{\hat{t}_k}{\alpha'}\right). &(51) \end{aligned}$$

Due to the convexity of f in (48), for all $k \in \{1, \ldots, n\}$,

$$f(\hat{t}_k) \geq f(x) + f'(x)\,(\hat{t}_k - x). \tag{52}$$

Choosing $x = \frac{1}{n}$ in the right side of (52) yields

$$\left(1 - \frac{\hat{t}_k}{\alpha'}\right) \log \left(1 - \frac{\hat{t}_k}{\alpha'}\right) \geq \log \left(1 - \frac{1}{n\alpha'}\right) + \frac{\log e}{n\alpha'} - \frac{\hat{t}_k}{\alpha'} \left[\log e + \log \left(1 - \frac{1}{n\alpha'}\right)\right] \tag{53}$$

and, in view of (41) and (49) which yields $\sum_{k=1}^{n} \hat{t}_k = 1$, summing over $k \in \{1, \dots, n\}$ on both sides of (53) implies that

$$\alpha' \sum_{k=1}^{n} \left(1 - \frac{\hat{t}_k}{\alpha'}\right) \log \left(1 - \frac{\hat{t}_k}{\alpha'}\right) \geq (n\alpha' - 1) \log \left(1 - \frac{1}{n\alpha'}\right). \tag{54}$$

Finally, assembling (50), (51) and (54) yields (42) with $c_\alpha^{(n)}$ in (31) as required.

In the sequel, we no longer assume that condition (32) holds. Define

$$\mathcal{K}_0 = \{k \in \{1, \dots, n\} \colon N_\alpha(X_k) = 0\}, \tag{55}$$

and note that

$$h_\alpha(S_n) = h_\alpha \left(\sum_{k \notin \mathcal{K}_0} X_k + \sum_{k \in \mathcal{K}_0} X_k \right) \tag{56}$$

$$\geq h_\alpha \left(\sum_{k \notin \mathcal{K}_0} X_k + \sum_{k \in \mathcal{K}_0} X_k \,\middle|\, \{X_k\}_{k \in \mathcal{K}_0} \right) \tag{57}$$

$$= h_\alpha \left(\sum_{k \notin \mathcal{K}_0} X_k \right) \tag{58}$$

where the conditional Rényi entropy is defined according to Arimoto's proposal in [2] (see also [20, Section 4]), (57) is due to the monotonicity property of the conditional Rényi entropy (see [20, Theorem 2]), and (58) is due to the independence of $X_1, \dots, X_n$. Since $N_\alpha(X_k) > 0$ for every $k \notin \mathcal{K}_0$, then from the previous analysis

$$N_\alpha \left(\sum_{k \notin \mathcal{K}_0} X_k \right) \geq c_\alpha^{(l)} \sum_{k \notin \mathcal{K}_0} N_\alpha(X_k), \tag{59}$$

where $l = n - |\mathcal{K}_0|$. In view of (31), it can be verified that $c_\alpha^{(n)}$ is monotonically decreasing in n; hence, (58), (59) and $c_\alpha^{(l)} \geq c_\alpha^{(n)}$ yield

$$N_\alpha(S_n) \geq c_\alpha^{(n)} \sum_{k=1}^{n} N_\alpha(X_k). \tag{60}$$

We now turn to prove Items 1)–4).

- To prove Item 1), note that (9) and (31) yield $c_\alpha^{(n)} > c_\alpha$ for all $\alpha > 1$ and $n \in \mathbb{N}$.
- Item 2) holds since from (31)

$$\lim_{n \to \infty} c_\alpha^{(n)} = \tfrac{1}{e}\, \alpha^{\frac{1}{1-\alpha}} \tag{61}$$

  where the right side of (61) coincides with the constant c_α in [9, (3)] (see (9)).
- Item 3) holds since $\alpha \downarrow 1$ yields $\alpha' \to \infty$, which implies that for every $n \in \mathbb{N}$

$$\lim_{\alpha \downarrow 1} c_\alpha^{(n)} = \lim_{\alpha \downarrow 1} c_\alpha = 1. \tag{62}$$

Hence, by letting α tend to 1, (30) and (62) yield the EPI in (1).

- To prove Item 4), note that from (31)

$$\lim_{\alpha\to\infty} c_\alpha^{(n)} = \left(1 - \frac{1}{n}\right)^{n-1} \tag{63}$$

which is monotonically decreasing in n for $n \geq 2$, being equal to $\frac{1}{2}$ for $n = 2$ and $\frac{1}{e}$ by letting n tend to ∞. Let X be a d-dimensional random vector with density f_X, and let

$$M(X) := \operatorname{ess\,sup}(f_X). \tag{64}$$

From (6), (7) and (64), it follows that

$$N_\infty(X) := \lim_{\alpha\to\infty} N_\alpha(X) \tag{65}$$

$$= M^{-\frac{2}{d}}(X). \tag{66}$$

By assembling (30) and (66), it follows that if $X_1, \ldots, X_n$ are independent d-dimensional random vectors with densities then

$$M^{-\frac{2}{d}}(S_n) \geq \left(1 - \frac{1}{n}\right)^{n-1} \sum_{k=1}^{n} M^{-\frac{2}{d}}(X_k). \tag{67}$$

This improves the tightness of the inequality in [8, Theorem 1] where the coefficient $\left(1 - \frac{1}{n}\right)^{n-1}$ on the right side of (67) has been loosened to $\frac{1}{e}$ (note, however, that they coincide when $n \to \infty$). For $n = 2$, the coefficient $\frac{1}{2}$ on the right side of (67) is tight, and it is achieved when X_1 and X_2 are independent random vectors which are uniformly distributed in the cube $[0,1]^d$ [8, p. 103]. ■

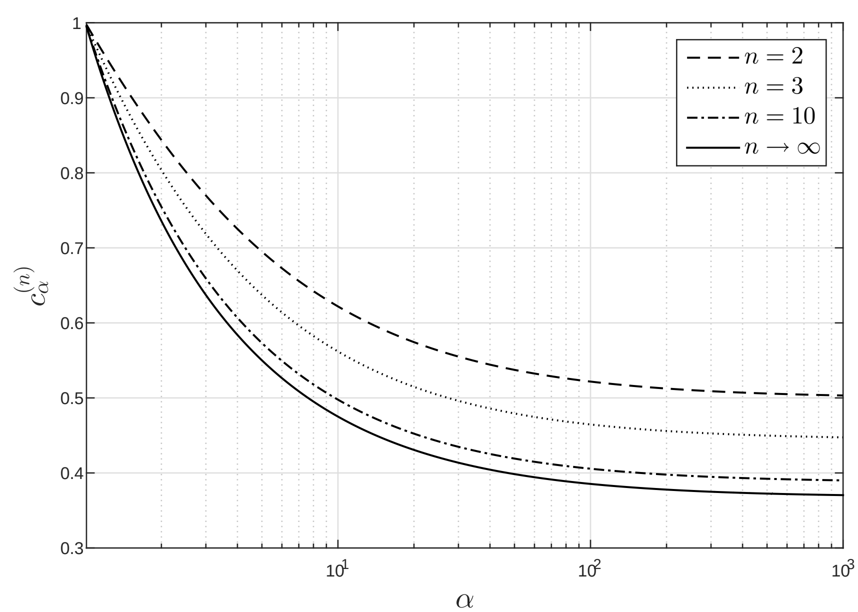

Fig. 1. A plot of $c_\alpha^{(n)}$ in (31), as a function of α, for $n = 2, 3, 10$ and $n \to \infty$.

Figure 1 plots $c_\alpha^{(n)}$ as a function of α, for some values of n, verifying numerically Items 1)–4) in Theorem 1. In [9, Theorem I.1], $c_\alpha^{(n)}$ is independent of n, and it is equal to c_α in (8) which is the limit of $c_\alpha^{(n)}$ in (31) by letting $n \to \infty$ (the solid curve in Figure 1).

Remark 1: For independent random variables $\{X_k\}_{k=1}^n$ with densities on $\mathbb{R}$, the result in (67) with $d = 1$ can be strengthened to (see [8, p. 105] and [38])

$$\frac{1}{M^2(S_n)} \geq \tfrac{1}{2} \sum_{k=1}^{n} \frac{1}{M^2(X_k)} \tag{68}$$

where $S_n := \sum_{k=1}^n X_k$. Note that (67) and (68) coincide if $n = 2$ and $d = 1$.

Example 1: Let X and Y be d-dimensional random vectors with densities f_X and f_Y, respectively, and assume that the entries of X are i.i.d. as well as those of Y. Let X_1, X_2, Y_1, Y_2 be independent d-dimensional random vectors where X_1, X_2 are independent copies of X, and Y_1, Y_2 are independent copies of Y. Assume that

$$\begin{aligned} &\mathbb{P}[X_{1,k} = X_{2,k}] = \alpha, \\ &\mathbb{P}[Y_{1,k} = Y_{2,k}] = \beta \end{aligned} \tag{69}$$

for all $k \in \{1, \ldots, d\}$. We wish to obtain an upper bound on the probability that $X_1 + Y_1$ and $X_2 + Y_2$ are equal. From (3), (7) (with $\alpha = 2$), and (69)

$$\begin{aligned} N_2(X) &= \exp\left(\tfrac{2}{d}\, h_2(X)\right) && (70) \\ &= \left(\int_{\mathbb{R}^d} f_X^2(\underline{x})\, \mathrm{d}\underline{x}\right)^{-\frac{2}{d}} && (71) \\ &= \mathbb{P}^{-\frac{2}{d}}[X_1 = X_2] && (72) \\ &= \prod_{k=1}^{d} \mathbb{P}^{-\frac{2}{d}}[X_{1,k} = X_{2,k}] && (73) \\ &= \alpha^{-2}, && (74) \\ N_2(Y) &= \beta^{-2}, && (75) \\ N_2(X+Y) &= \mathbb{P}^{-\frac{2}{d}}[X_1 + Y_1 = X_2 + Y_2]. && (76) \end{aligned}$$

Assembling (30) with $n = \alpha = 2$, (74), (75) and (76) yield

$$\mathbb{P}[X_1 + Y_1 = X_2 + Y_2] \leq \left(\tfrac{27}{32}\left(\alpha^{-2} + \beta^{-2}\right)\right)^{-\frac{d}{2}}. \tag{77}$$

The factor $\frac{27}{32}$ on the base of the exponent on the right side of (77), instead of the looser factor $c_2 = \frac{2}{e}$ which follows from (9) with $\alpha = 2$ (see [9, Theorem I.1]), improves the exponential decay rate of the upper bound in (77) as a function of the dimension d. The optimal bound has to be with a coefficient of $\left(\alpha^{-2} + \beta^{-2}\right)$ on the base of the exponent in the right side of (77) which is less than or equal to 1; this can be verified since if X and Y are independent Gaussian random variables, then

$$N_2(X+Y) = N_2(X) + N_2(Y), \tag{78}$$

so,

$$\mathbb{P}[X_1 + Y_1 = X_2 + Y_2] = \left(\alpha^{-2} + \beta^{-2}\right)^{-\frac{d}{2}}. \tag{79}$$

This provides a reference for comparing the exponential decay which is implied by c_2 in (9), $c_2^{(2)}$ in (30), and the case where X and Y are independent Gaussian random variables:

$$\frac{2}{e} < \frac{27}{32} < 1. \tag{80}$$

IV. A Further tightening of the R-EPI

A. A Tightened R-EPI for $n \geq 2$

In the following, we wish to tighten the R-EPI in Theorem 1. It is first demonstrated that a reduction of the optimization problem in (47) to $n-1$ variables (recall that $\sum_{k=1}^n t_k = 1$) leads to a convex optimization problem. This convexity result is established by a non-trivial use of Fact 3 in Section II-B (see [14]), and it is also shown that the reduction of the optimization problem in (47) from n to $n-1$ variables is essential for its convexity. Consequently, the convex optimization problem is handled by solving the corresponding Karush-Kuhn-Tucker (KKT) equations. If $n = 2$, their solution leads to a closed-form expression which yields the R-EPI in Corollary 3. For $n > 2$, no solution is provided in closed form; nevertheless, an efficient algorithm is introduced for solving the KKT equations for an arbitrary $n > 2$, and the improvement in the tightness of the new R-EPI in this section is exemplified numerically in comparison to the bounds in [5], [9] and Theorem 1.

1) The optimization problem in (47)*:* In view of (44)–(47), the maximization problem in (47) can be expressed in the form

$$\begin{array}{ll} \text{maximize} & f_0(\underline{t}) = \sum_{k=1}^n g(t_k) + \sum_{k=1}^n t_k \log N_k + \frac{\log \alpha}{\alpha - 1} \\ \text{subject to} & \underline{t} \in \mathcal{P}^n \end{array} \tag{81}$$

where

$$g(x) = (\alpha' - x) \log\left(1 - \frac{x}{\alpha'}\right) - x \log x, \quad x \in [0,1] \tag{82}$$

$$N_k = N_\alpha(X_k), \quad k \in \{1, \ldots, n\} \tag{83}$$

(for simplicity of notation, the dependence of g and N_k in α has been suppressed in (81)), and $\mathcal{P}^n$ is the probability simplex

$$\mathcal{P}^n = \left\{ \underline{t} \in \mathbb{R}^n : t_k \geq 0, \sum_{k=1}^n t_k = 1 \right\}. \tag{84}$$

The term $\sum_{k=1}^n t_k \log N_k$ on the right side of (81) is linear in $\underline{t}$, thus the concavity of f_0 in $\underline{t}$ is only affected by the term $\sum_{k=1}^n g(t_k)$. Since $g''(x) = \frac{2x - \alpha'}{x(\alpha' - x)}$ where $x \in [0,1]$, if $\alpha' \geq 2$, then g is concave on the interval $[0,1]$. If $\alpha' \in (1,2)$ (i.e., if $\alpha \in (2, \infty)$) then g is not concave on the interval $[0,1]$; it is only concave on $[0, \frac{\alpha'}{2}]$, and it is convex on $[\frac{\alpha'}{2}, 1]$. Hence, as a maximization problem over the variables $t_1, \ldots, t_n$, the objective function f_0 in (81) is not concave if $\alpha > 2$.

2) A reduction of the optimization problem in (47) *to* $n-1$ *variables:* In view of (84), the substitution

$$t_n = 1 - \sum_{k=1}^{n-1} t_k \tag{85}$$

transforms the maximization problem in (81) to the following equivalent problem:

$$\begin{array}{ll} \text{maximize} & f(t_1, \ldots, t_{n-1}) \\ \text{subject to} & \underline{t} \in \mathcal{D}^{n-1} \end{array} \tag{86}$$

where

$$f(t_1, \ldots, t_{n-1}) = f_0\left(t_1, \ldots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k\right) \tag{87}$$

and $\mathcal{D}^{n-1}$ is the polyhedron

$$\mathcal{D}^{n-1} = \left\{ (t_1, \ldots, t_{n-1}) \colon t_k \geq 0, \, \sum_{k=1}^{n-1} t_k \leq 1 \right\}. \tag{88}$$

3) Proving the convexity of the optimization problem in (86)*:* We wish to show that the objective function f of the optimization problem in (86) is concave, i.e., it is required to assert that all the eigenvalues of the Hessian matrix $\nabla^2 f$ are non-positive.

Eqs. (81) and (87) yield

$$\begin{aligned} & f(t_1, \ldots, t_{n-1}) \\ & = \sum_{k=1}^{n-1} g(t_k) + g\left(1 - \sum_{k=1}^{n-1} t_k\right) \\ & \quad + \sum_{k=1}^{n-1} t_k \log N_k + \left(1 - \sum_{k=1}^{n-1} t_k\right) \log N_n + \frac{\log \alpha}{\alpha - 1}. \end{aligned} \tag{89}$$

Let

$$q(x) = g''(x) = \frac{2x - \alpha'}{x(\alpha' - x)}, \quad x \in [0, 1] \tag{90}$$

then, in view of (89) and (90), for all $(t_1, \ldots, t_{n-1}) \in \mathcal{D}^{n-1}$

$$\begin{aligned} \nabla^2 f(t_1, \ldots, t_{n-1}) &= \begin{pmatrix} q(t_1) & 0 & \cdots & 0 \\ 0 & q(t_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q(t_{n-1}) \end{pmatrix} + q\left(1 - \sum_{k=1}^{n-1} t_k\right) \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \\ &= D + \rho \, \underline{1} \, \underline{1}^T \end{aligned} \tag{91}$$

where

$$\begin{aligned} & D = \operatorname{diag}(q(t_1), \ldots, q(t_{n-1})), \\ & \rho = q\left(1 - \sum_{k=1}^{n-1} t_k\right). \end{aligned} \tag{92}$$

Recall that if $\alpha' \in [2, \infty)$ then $f_0(t_1, \ldots, t_n)$ is concave in $\mathcal{P}^n$, hence, so is $f(t_1, \ldots, t_{n-1})$ in D^{n-1}. We therefore need only to focus on the case where $\alpha' \in (1, 2)$ (i.e., $\alpha \in (2, \infty)$).

Proposition 1: For every $\alpha' \in (1, 2)$, the function $f \colon \mathcal{D}^{n-1} \to \mathbb{R}$ in (89) is concave.

Proof: See Appendix B. ■

4) Solution of the convex optimization problem in (86)*:* In the following, we solve the convex optimization problem in (86) via the Lagrange duality and KKT conditions (see, e.g., [12, Chapter 5]). Since the problem is invariant to permutations of the entries of $X = (X_1, \ldots, X_n)$, it can be assumed without any loss of generality that the last term of the vector $\underline{N}_\alpha$ in (45) is maximal, i.e.,

$$N_\alpha(X_k) \leq N_\alpha(X_n), \quad k \in \{1, \ldots, n-1\}. \tag{93}$$

Moreover, it is assumed that

$$N_\alpha(X_n) > 0. \tag{94}$$

The possibility that $N_\alpha(X_n) = 0$ leads to a trivial bound since from (93), it follows that $N_\alpha(X_k) = 0$ for every $k \in \{1, \ldots, n\}$; this makes the right side of (8) be equal to zero, while its left side is always non-negative. Let

$$c_k = \frac{N_\alpha(X_k)}{N_\alpha(X_n)}, \qquad k \in \{1, \ldots, n-1\}. \tag{95}$$

From (93)–(95), the sequence $\{c_k\}_{k=1}^{n-1}$ satisfies

$$0 \le c_k \le 1, \quad k \in \{1, \ldots, n-1\}. \tag{96}$$

Let t_n be defined as in (85). Appendix C provides the technical details which are related to the solution of the convex optimization problem in (86) via the Lagrange duality and KKT conditions (note that strong duality holds here). The resulting simplified set of constraints which follow from the KKT conditions (see Appendix C) is given by

$$t_k(\alpha' - t_k) = c_k t_n(\alpha' - t_n), \quad k \in \{1, \ldots, n-1\} \tag{97}$$

$$\sum_{k=1}^n t_k = 1 \tag{98}$$

$$t_k \ge 0, \quad k \in \{1, \ldots, n\} \tag{99}$$

with the variables $\underline{t}$ in (97)–(99).

Note that if $N_\alpha(X_k)$ is independent of k then, from (95), $c_k = 1$ for all $k \in \{1, \ldots, n-1\}$. Hence, from (97) and (98), it follows that $t_1 = \ldots = t_n = \frac{1}{n}$ (note that the other possibility where $t_k = \alpha' - t_n$ for some $k \in \{1, \ldots, n-1\}$ contradicts (98) and (99) since in this case $\sum_{j=1}^n t_j \ge t_k + t_n = \alpha' > 1$). This implies that the selection of the t_k's in the proof of Theorem 1 is optimal when all the entries of the vector $\underline{N}_\alpha$ are equal; therefore, the R-EPI considered here improves the bound in Theorem 1 only when $N_\alpha(X_k)$ depends on the index k.

In the general case, (97) yields a quadratic equation for t_k whose solutions are given by

$$t_k = \tfrac{1}{2} \left(\alpha' \pm \sqrt{\alpha'^2 - 4c_k t_n(\alpha' - t_n)} \right) \tag{100}$$

with $\alpha' = \frac{\alpha}{\alpha-1}$. The possibility of the positive sign in the right side of (100) is rejected since in this case $t_n + t_k \ge \alpha' > 1$, which violates (98). Hence, from (100), for all $k \in \{1, \ldots, n-1\}$

$$t_k = \psi_{k,\alpha}(t_n) \tag{101}$$

where we define

$$\psi_{k,\alpha}(x) = \tfrac{1}{2} \left(\alpha' - \sqrt{\alpha'^2 - 4c_k\, x(\alpha' - x)} \right), \quad x \in [0, 1]. \tag{102}$$

In view of (98) and (101), one first calculates $t_n \in [0, 1]$ by numerically solving the equation

$$t_n + \sum_{k=1}^{n-1} \psi_{k,\alpha}(t_n) = 1. \tag{103}$$

The existence and uniqueness of a solution of (103) is proved in Appendix D. Once we compute t_n, all t_k's for $k \in \{1, \ldots, n-1\}$ are computed from (101). Finally, the substitution of $t_1, \ldots, t_n$ in the right side of (43) enables to calculate the improved R-EPI in (43), i.e.,

$$N_\alpha\left(\sum_{k=1}^n X_k\right) \ge \exp\big(f_0(t_1, \ldots, t_n)\big) \sum_{k=1}^n N_\alpha(X_k) \tag{104}$$

with f_0 in (44).

Note that due to the optimal selection of the vector $\underline{t} = (t_1, \ldots, t_n)$ in (104), the R-EPI in this section provides an improvement over the R-EPI in Theorem 1 whenever $N_\alpha(X_k)$ is not fixed as a function of the index k. This leads to the following result:

Theorem 2: Let $X_1, \ldots, X_n$ be independent random vectors with probability densities defined on $\mathbb{R}^d$, let $N_\alpha(X_1), \ldots, N_\alpha(X_n)$ be their respective Rényi entropy powers of order $\alpha > 1$, and let $\alpha' = \frac{\alpha}{\alpha-1}$. Let the indices of $X_1, \ldots, X_n$ be set such that $N_\alpha(X_n)$ is maximal, and let

1) $\{c_k\}_{k=1}^{n-1}$ be the sequence defined in (95);
2) $t_n \in [0, 1]$ be the unique solution of (103);
3) $\{t_k\}_{k=1}^{n-1}$ be given in (101) and (102).

Then, the R-EPI in (104) holds with f_0 in (44), and it satisfies the following properties:

1) It improves the R-EPI in Theorem 1 unless $N_\alpha(X_k)$ is independent of k (consequently, it also improves the R-EPI in [9, Theorem 1]); if $N_\alpha(X_k)$ is independent of k, then the two R-EPIs in Theorem 1 and (104) coincide.
2) It improves the Bercher-Vignat (BV) bound in [5] which states that
$$N_\alpha\left(\sum_{k=1}^{n} X_k\right) \geq \max\bigl\{N_\alpha(X_1), \ldots, N_\alpha(X_n)\bigr\} \tag{105}$$
and the bounds in (104) and (105) asymptotically coincide as $\alpha \to \infty$ if and only if
$$\sum_{k=1}^{n-1} N_\infty(X_k) \leq N_\infty(X_n) \tag{106}$$
where $N_\infty(X)$ is defined in (66).
3) For $n = 2$, it is expressed in a closed form (see Corollary 3).
4) It coincides with the EPI and the two R-EPIs in [9, Theorem 1] and Theorem 1 as $\alpha \downarrow 1$.

Proof: The proof of the R-EPI in (104) is provided earlier in this section with some additional details in Appendices B–E. In view of this analysis:

- Item 1) holds since the proof of the R-EPI in Theorem 1 relies in general on a sub-optimal choice of the vector $\underline{t}$ in (49), whereas it is set to be optimal in the proof of Theorem 2 in (101)–(103). Suppose, however, that $N_\alpha(X_k)$ is independent of the index k; in the latter case, the selection of the vector $\underline{t}$ in the proof of Theorem 1 (see (49)) reduces to $\underline{t} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$, which turns to be optimal in the sense of achieving the maximum of the objective function in (89).
- Item 2) holds since the selection of $\underline{t}$ in the right side of (43) with $t_k = 1$ and $t_i = 0$ for all $i \neq k$ yields
$$N_\alpha\left(\sum_{k=1}^{n} X_k\right) \geq N_\alpha(X_k) \tag{107}$$
which then leads to (105) by a maximization of the right side of (107) over $k \in \{1, \ldots, n\}$. Appendix E proves that the bounds in (104) and (105) asymptotically coincide as $\alpha \to \infty$ if and only if the condition in (106) holds.
- Item 3) is proved in Section IV-B.
- Item 4) holds since the R-EPI obtained in Theorem 2 is at least as tight as the BC bound in [9, Theorem 1]; the latter coincides with the EPI as we let α tend to 1 (recall that, from (9), $\lim_{\alpha \downarrow 1} c_\alpha = 1$) which is known to be tight for Gaussian random vectors with proportional covariances.

■

Remark 2: The R-EPI in Theorem 2 provides the tightest R-EPI known to date for $\alpha \in (1, \infty)$. Nevertheless, it is still not tight for $\alpha \in (1, \infty)$ since at least one of the inequalities involved in the derivation of (33) (see Appendix A) is loose. These include the sharpened Young's inequality in (20), and (12). The former inequality holds with equality only for Gaussians, whereas the latter inequality holds with equality only for a uniformly distributed random variable (note that in the latter case, the Rényi entropy is independent of its order). For $\alpha = \infty$ and $n = 2$, the sharpened Young's inequality (16) reduces to

$$\|f * g\|_\infty \leq \|f\|_p \, \|g\|_{p'} \tag{108}$$

where $p > 1$ and $p' = \frac{p}{p-1}$. Equality holds in (108) if f and g are scaled versions of a uniform distribution on the same convex set, which is also the same condition for tightness of (12); this is consistent with our conclusion that the R-EPIs in Theorems 1 and 2 are, however, asymptotically tight for $n = 2$ by letting $\alpha \to \infty$.

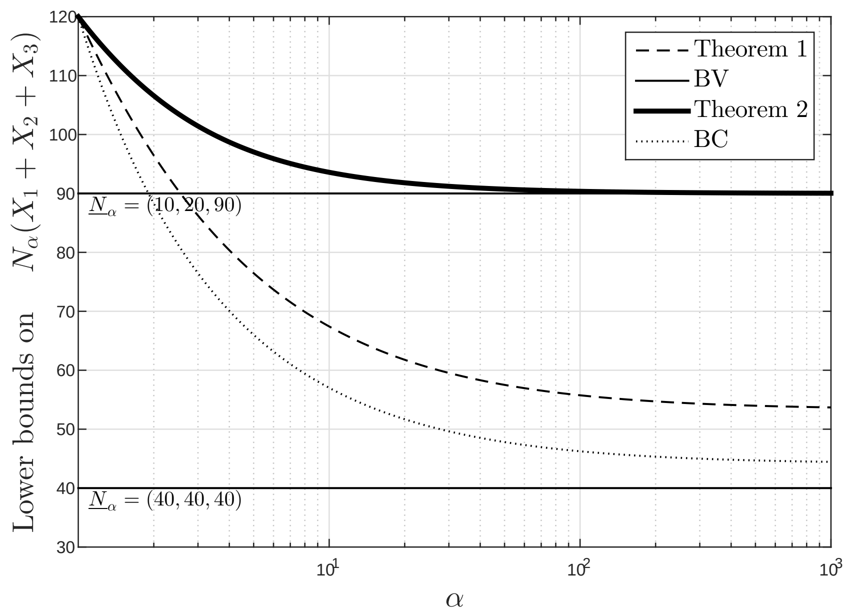

Fig. 2. A comparison of the Rényi entropy power inequalities for $n = 3$ independent random vectors according to [9] (BC), [5] (BV), Theorem 1 and the tightest bound in Theorem 2. The bounds refer to the two cases where $(N_\alpha(X_1), N_\alpha(X_2), N_\alpha(X_3)) = (40, 40, 40)$ or $(10, 20, 90)$ (in both cases, the sum of the entries is 120; in the former case, the condition in (106) does not hold, while in the latter it does).

Figure 2 compares the two R-EPIs in Theorems 1 and 2 with those in [9] (see (9)) and [5] (see (105)) for $n = 3$ independent random vectors; the abbreviations 'BC' and 'BV' stand, respectively, for the latter two bounds. Recall that the four bounds are independent of the dimension d of the random vectors, and they are plotted in Figure 2 for symmetric and asymmetric cases where $(N_\alpha(X_1), N_\alpha(X_2), N_\alpha(X_3)) = (40, 40, 40)$ and $(10, 20, 90)$, respectively (note that in both cases, the sum of the entries is equal to 120). In the former case, for every $\alpha > 1$, Theorem 2 provides a lower bound on $N_\alpha(X_1 + X_2 + X_3)$ which is tighter than those in [5] and [9]; furthermore, in this special case where $N_\alpha(X_k)$ is independent of the index k, the bounds in Theorems 1 and 2 coincide. In the asymmetric case, however, where $(N_\alpha(X_1), N_\alpha(X_2), N_\alpha(X_3)) = (10, 20, 90)$, the bound in Theorem 2 suggests a significant improvement over the bound in Theorem 1 due to the sub-optimality of the choice of the vector $\underline{t}$ in the proof of Theorem 1 in comparison to its optimal choice in Theorem 2. As it is shown in Figure 2 and supported by Item 2) of Theorem 2, the bound in this theorem asymptotically coincides with the BV bound (by letting $\alpha \to \infty$) in the considered asymmetric case; however, for every $\alpha \in (1, \infty)$, the bound in Theorem 2 is advantageous over the BV bound. It is also shown in Figure 2 that in this asymmetric case, the BV bound is advantageous over our bound in Theorem 1 for sufficiently large α; this observation emphasizes the significance of the optimization of the vector $\underline{t}$ in the proof of Theorem 2, yielding the tightest R-EPI known to date for $\alpha > 1$. Finally, as it is shown in Figure 2, the R-EPIs of Theorems 1 and 2, as well as [9, Theorem 1], coincide with the EPI as we let α tend to 1 (from above).

B. A Closed-Form Expression of the Tightened R-EPI for $n = 2$

We derive in the following a closed-form expression of the R-EPI in Theorem 2 for $n = 2$ independent random vectors. In the sequel, we make use of the binary relative entropy function which is defined to be the continuous extension to $[0,1]^2$ of

$$d(x\|y) = x \log\left(\frac{x}{y}\right) + (1-x)\log\left(\frac{1-x}{1-y}\right). \tag{109}$$

Corollary 3: Let X_1 and X_2 be independent random vectors with densities defined on $\mathbb{R}^d$, let $N_\alpha(X_1)$, $N_\alpha(X_2)$ be their Rényi entropy powers of order $\alpha > 1$, and assume without any loss of generality that $N_\alpha(X_1) \leq N_\alpha(X_2)$. Let

$$\alpha' = \frac{\alpha}{\alpha - 1}, \tag{110}$$

$$\beta_\alpha = \frac{N_\alpha(X_1)}{N_\alpha(X_2)}, \tag{111}$$

$$t_\alpha = \begin{cases} \frac{\alpha'(\beta_\alpha+1) - 2\beta_\alpha - \sqrt{(\alpha'(\beta_\alpha+1))^2 - 8\alpha'\beta_\alpha + 4\beta_\alpha}}{2(1-\beta_\alpha)} & \text{if } \beta_\alpha < 1, \\ \frac{1}{2} & \text{if } \beta_\alpha = 1. \end{cases} \tag{112}$$

Then, the following R-EPI holds:

$$N_\alpha(X_1 + X_2) \geq c_\alpha \left(N_\alpha(X_1) + N_\alpha(X_2)\right) \tag{113}$$

with

$$c_\alpha = \alpha^{\frac{1}{\alpha-1}} \exp\left(-d\Big(t_\alpha \,\Big\|\, \frac{\beta_\alpha}{\beta_\alpha + 1}\Big)\right) \left(1 - \frac{t_\alpha}{\alpha'}\right)^{\alpha' - t_\alpha} \left(1 - \frac{1 - t_\alpha}{\alpha'}\right)^{\alpha' - 1 + t_\alpha}. \tag{114}$$

The R-EPI in (113) satisfies Items 1)–4) of Theorem 2; specifically, by letting $\alpha \to \infty$, the lower bound on $N_\alpha(X_1 + X_2)$ tends to $N_\infty(X_2)$, which asymptotically coincides with the BV bound in [5].

Proof: Due to the constraints in (47), the vector $\underline{t}$ can be parameterized in the form $\underline{t} = (t, 1-t)$ for $t \in [0,1]$; due to the normalization of the vector $\underline{N}_\alpha = (N_\alpha(X_1), N_\alpha(X_2))$ in (41), then

$$\underline{N}_\alpha = \left(\tfrac{\beta_\alpha}{1+\beta_\alpha}, \tfrac{1}{1+\beta_\alpha}\right) \tag{115}$$

and, by (44), the maximization in (81) is transformed to

$$\begin{aligned} &\underset{t \in [0,1]}{\text{maximize}} \left\{ \frac{\log \alpha}{\alpha - 1} - t \log\big((1+\beta_\alpha)t\big) - (1-t)\log\left(\frac{(1+\beta_\alpha)(1-t)}{\beta_\alpha}\right) \right. \\ &\qquad \left. + \alpha' \left[\left(1 - \frac{t}{\alpha'}\right) \log\left(1 - \frac{t}{\alpha'}\right) + \left(1 - \frac{1-t}{\alpha'}\right) \log\left(1 - \frac{1-t}{\alpha'}\right) \right] \right\}. \end{aligned} \tag{116}$$

It can be verified that the objective function in (116) is concave on $[0,1]$, it has a right derivative at $t=0$ which is equal to $+\infty$, and a left derivative at $t=1$ which is equal to $-\infty$. This implies that the maximization of the objective function over $[0,1]$ is attained at an interior point of this interval. The optimized value of t is obtained by setting the derivative of this objective function to zero, leading to the equation

$$\log\left(\tfrac{(1-t)\beta_\alpha}{t}\right) - \log\left(\tfrac{\alpha'-t}{\alpha'-1+t}\right) = 0. \tag{117}$$

Eq. (117) can be expressed as a quadratic equation whose solution is given in (112). Substituting the optimized value $t = t_\alpha$ in (112) into the objective function on the right side of (116) leads to the closed-form solution of the optimization problem in (81) for $n=2$. Hence, under the assumption in (41) where $N_\alpha(X_1) + N_\alpha(X_2) = 1$, straightforward algebra yields that

$$N_\alpha(X_1 + X_2) \geq c_\alpha \tag{118}$$

where c_α is given in (114); the relaxation of this assumption requires the multiplication of the right side of (118) by $N_\alpha(X_1) + N_\alpha(X_2)$ (due to the homogeneity of the Rényi entropy power, see (28)). Note that, for $n=2$, the condition in (106) becomes vacuous (since, by assumption, $N_\infty(X_1) \leq N_\infty(X_2)$) which implies that the bound in (113) asymptotically coincides with the BV bound when $\alpha \to \infty$. ∎

V. Example: The Rényi Entropy Difference Between Data and its Filtering

Let $\{\underline{X}(n)\}$ be i.i.d. d-dimensional random vectors (the entries of the vector $\underline{X}(n)$ need not be independent), with arbitrary densities on $\mathbb{R}^d$. Let

$$\underline{Y}(n) = \sum_{k=0}^{L-1} \mathbf{H}_k \, \underline{X}(n-k) \tag{119}$$

be the filtered data at the output of a finite impulse response (FIR) filter where $\mathbf{H}_0, \ldots, \mathbf{H}_{L-1}$ are fixed non-singular $d \times d$ matrices.

In the following, the tightness of several R-EPIs is exemplified by obtaining universal lower bounds on the difference $h_\alpha\big(\underline{Y}(n)\big) - h_\alpha\big(\underline{X}(n)\big)$, being also compared with the actual value of this difference when the i.i.d. inputs are d-dimensional Gaussian random vectors with i.i.d. entries.

For $k \in \{0, \ldots, L-1\}$ and every n, we have

$$h_\alpha\big(\mathbf{H}_k \, \underline{X}(n-k)\big) = h_\alpha\big(\underline{X}(n)\big) + \log\big|\det(\mathbf{H}_k)\big| \tag{120}$$

and

$$\begin{aligned} &N_\alpha\big(\mathbf{H}_k \, \underline{X}(n-k)\big) \\ &= \exp\left(\tfrac{2}{d}\, h_\alpha\big(\mathbf{H}_k \, \underline{X}(n-k)\big)\right) \\ &= \big|\det(\mathbf{H}_k)\big|^{\frac{2}{d}} \, N_\alpha\big(\underline{X}(n)\big). \end{aligned} \tag{121}$$

Let $\alpha > 1$, and $\alpha' = \frac{\alpha}{\alpha-1}$. Similarly to Theorem 2, it is assumed without loss of generality that $\big|\det(\mathbf{H}_k)\big| \leq \big|\det(\mathbf{H}_{L-1})\big|$ for all $k \in \{0, \ldots, L-2\}$; otherwise, the indices of $\mathbf{H}_0, \ldots, \mathbf{H}_{L-1}$ can be permuted without affecting the differential Rényi entropy of $\underline{Y}(n)$. In the setting of the improved R-EPI of Theorem 2, in view of (95) and (121), for every $k \in \{0, \ldots, L-2\}$,

$$c_k = \left(\frac{\big|\det(\mathbf{H}_k)\big|}{\big|\det(\mathbf{H}_{L-1})\big|}\right)^{\frac{2}{d}} \tag{122}$$

which, in view of the above assumption, implies that $c_k \in [0,1]$ for $k \in \{0,\dots,L-2\}$. Given the L matrices $\{\mathbf{H}_k\}_{k=0}^{L-1}$, the vector $(t_0,\dots,t_{l-1}) \in [0,1]^L$ is calculated according to Theorem 2; first $t_{L-1} \in [0,1]$ is numerically calculated by solving the equation in (103) (with a replacement of 1 and n in (103) by 0 and $L-1$, respectively), and then the rest of the t_k's for $k \in \{0,\dots,L-2\}$ are calculated via (101) and (102). In view of (120), (121), and the R-EPI of Theorem 2, it follows that for every n

$$\begin{aligned} & h_\alpha\big(\underline{Y}(n)\big) - h_\alpha\big(\underline{X}(n)\big) \\ & \geq \frac{d}{2}\left(\frac{\log\alpha}{\alpha-1} + \sum_{k=0}^{L-1} g(t_k)\right) + \sum_{k=0}^{L-1} t_k \, \log\big|\det(\mathbf{H}_k)\big| \end{aligned} \tag{123}$$

where the function g is given in (82).

In view of the derivation so far, it is easy to verify that the R-EPI in Theorem 1 is equivalent to the following looser bound, which is expressed in closed form:

$$\begin{aligned} & h_\alpha\big(\underline{Y}(n)\big) - h_\alpha\big(\underline{X}(n)\big) \\ & \geq \frac{d}{2} \cdot \log\left(\sum_{k=0}^{L-1} \big|\det(\mathbf{H}_k)\big|^{\frac{2}{d}}\right) \\ & \quad + \frac{d}{2}\left(\frac{\log\alpha}{\alpha-1} + \left(\frac{L\alpha}{\alpha-1} - 1\right) \log\left(1 - \frac{\alpha-1}{L\alpha}\right)\right). \end{aligned} \tag{124}$$

The R-EPI of [9, Theorem I.1] leads to the following loosened bound in comparison to (124):

$$h_\alpha\big(\underline{Y}(n)\big) - h_\alpha\big(\underline{X}(n)\big) \geq \frac{d}{2}\left[\log\left(\sum_{k=0}^{L-1} \big|\det(\mathbf{H}_k)\big|^{\frac{2}{d}}\right) + \frac{\log\alpha}{\alpha-1} - \log e\right] \tag{125}$$

and, finally, the BV bound in [5] (see (105)) leads to the following loosening of (123):

$$h_\alpha\big(\underline{Y}(n)\big) - h_\alpha\big(\underline{X}(n)\big) \geq \log\left(\max_{0 \leq k \leq L-1} \big|\det(\mathbf{H}_k)\big|\right). \tag{126}$$

The differential Rényi entropy of order $\alpha \in (0,1) \cup (1,\infty)$ for a d-dimensional multivariate Gaussian distribution is given by

$$h_\alpha\big(\underline{X}(n)\big) = \frac{d\log\alpha}{2(\alpha-1)} + \tfrac{1}{2} \log\Big((2\pi)^d \, \det\big(\mathrm{Cov}(\underline{X}(n))\big)\Big). \tag{127}$$

Hence, if the entries of the Gaussian random vector $\underline{X}(n)$ are i.i.d.

$$h_\alpha\big(\underline{Y}(n)\big) - h_\alpha\big(\underline{X}(n)\big) = \tfrac{1}{2} \log\left(\det\left(\sum_{k=0}^{L-1} \mathbf{H}_k \, \mathbf{H}_k^T\right)\right). \tag{128}$$

Example 2: Let

$$Y(n) = 2X(n) - X(n-1) - X(n-2) \tag{129}$$

for every n where $\{X(n)\}$ are i.i.d. random variables, and consider the difference $h_2(Y) - h_2(X)$ in the quadratic differential Rényi entropy. In this example $\alpha = 2$, $d = 1$, $L = 3$, and $H_0 = 2$, $H_1 = -1$, $H_2 = -1$. The lower bounds in (123), (124), (125), (126) are equal to 0.8195, 0.7866, 0.7425 and 0.6931 nats, respectively (recall that the first two lower bounds correspond to Theorems 2 and 1 respectively, and the last two bounds correspond to [9] and [5] respectively. These lower bounds are compared to the achievable value in (128), for an i.i.d. Gaussian input, which is equal to 0.8959 nats.

VI. Summary

This work is focused on the derivation of improved Rényi entropy power inequalities (R-EPI) for a sum of n independent and continuous random vectors over $\mathbb{R}^d$. These inequalities are of the form (8), they refer to orders $\alpha \in (1, \infty]$, and they also coincide with the EPI [41] by letting $\alpha \to 1$. Theorem 1 provides an R-EPI with a constant which is given in closed form in (31), improving the R-RPI by Bobkov and Chistyakov in [9, Theorem 1]; furthermore, for $n = 2$, the R-EPI in Theorem 1 is asymptotically tight when $\alpha \to \infty$. The R-EPI which is introduced in Theorem 2 can be efficiently calculated via a simple numerical algorithm, it is tighter than Theorem 1 and all previously reported bounds, and it is currently the best known R-EPI for $\alpha \in (1, \infty)$. Corollary 3 provides a closed-form expression for the R-EPI in Theorem 2 for a sum of two independent random vectors. It should be noted that the R-EPIs in Theorems 1 and 2 coincide when the Rényi entropy powers of the n independent random vectors are all equal.

Theorem 1 is obtained by tightening the recent R-EPI by Bobkov and Chistyakov [9] with the same analytical tools, namely the monotonicity of $N_\alpha(X)$ in α, and the use of the sharpened Young's inequality. Theorem 2, which improves the tightness of the R-EPI in Theorem 1, relies on the following additional analytical tools: 1) a strong Lagrange duality of an optimization problem is asserted by invoking a theorem in matrix theory [14] regarding the rank-one modification of a real-valued diagonal matrix, and 2) a solution of the Karush-Kuhn-Tucker (KKT) equations of the related optimization problem.

Appendix A
Proof of (33)

Since $\{X_k\}_{k=1}^n$ are independent random variables, the density of $S_n = \sum\limits_{k=1}^{n} X_k$ is the convolution of the densities f_{X_k}. In view of (20) and (25), for $\alpha > 1$,

$$\begin{aligned} N_\alpha(S_n) &= \left(\|f_{X_1} * \ldots * f_{X_n}\|_\alpha\right)^{-\frac{2\alpha'}{d}} \\ &\geq A^{-\frac{2\alpha'}{d}} \textstyle\prod_{k=1}^n \left(\|f_{X_k}\|_{\nu_k}\right)^{-\frac{2\alpha'}{d}} \end{aligned} \tag{130}$$

where

$$\nu_k > 1, \quad 1 \leq k \leq n \tag{131}$$

$$\sum_{k=1}^{n} \frac{1}{\nu_k'} = \frac{1}{\alpha'} \tag{132}$$

and, due to (14) and (19),

$$A = \left(A_{\alpha'} \prod_{k=1}^{n} A_{\nu_k}\right)^{\frac{d}{2}}. \tag{133}$$

From (131) and (132) it follows that $\nu_k \in (1, \alpha]$ for all $k \in \{1, \ldots, n\}$, hence in view of Corollary 1,

$$\|f_{X_k}\|_{\nu_k}^{\nu_k'} \leq \|f_{X_k}\|_\alpha^{\alpha'}, \quad 1 \leq k \leq n. \tag{134}$$

Combining (130) and (134), and defining $t_k = \frac{\alpha'}{\nu_k'}$ yields

$$N_\alpha(S_n) \geq A^{-\frac{2\alpha'}{d}} \prod_{k=1}^{n} \left(\|f_{X_k}\|_\alpha\right)^{-\frac{2\alpha'}{d} \cdot \frac{\alpha'}{\nu_k'}} = A^{-\frac{2\alpha'}{d}} \prod_{k=1}^{n} N_\alpha^{t_k}(X_k) \tag{135}$$

which by setting $B = A^{-\frac{2\alpha'}{d}}$ completes the proof of (33) with the constant B as given in (34).

APPENDIX B
PROOF OF PROPOSITION 1

Let $\alpha' \in (1,2)$. If there exists an index $k \in \{1,\dots,n-1\}$ such that $q(t_k)=0$, then $t_k = \frac{\alpha'}{2} > \frac{1}{2}$ (see (90)). In view of (88), it follows that $t_l < \frac{1}{2}$ for every other index $l \neq k$ in the set $\{1,\dots,n-1\}$, which in turn implies from (90) that $q(t_l) < 0$ for every such index l. In other words, if there exists an index $k \in \{1,\dots,n-1\}$ such that $q(t_k)=0$, then it follows that $q(t_l) \leq 0$ for all $l \in \{1,\dots,n-1\}$. In view of (92), $D \preceq 0$ and $\rho < 0$ (to verify that $\rho < 0$, note that since $0 \leq 1 - \sum_{j=1}^{n-1} t_j \leq 1 - t_k = 1 - \frac{\alpha'}{2} < \frac{1}{2} < \frac{\alpha'}{2}$ then it follows from (90) and (92) that $\rho = q\big(1 - \sum_{j=1}^{n-1} t_j\big) < 0$); hence, (91) implies that $\nabla^2 f(t_1,\dots,t_{n-1}) \prec 0$ in the interior of $\mathcal{D}^{n-1}$, so f is (strictly) concave on $\mathcal{D}^{n-1}$.

To proceed, the following lemmas will be useful.

Lemma 1: If $\alpha' \in (1,2)$ and $x \in (0, 1-\frac{\alpha'}{2})$, then

$$\frac{1}{q(x)} + \frac{1}{q(1-x)} > 0. \tag{136}$$

Proof: In view of (90), the left side of (136) is equal to

$$\frac{\overbrace{(1-\alpha')}^{<0}\ \overbrace{(2x^2-2x+\alpha')}^{>0}}{\underbrace{(2x-\alpha')}_{<0}\ \underbrace{(2-2x-\alpha')}_{>0}} > 0.$$

■

Lemma 2: If $\alpha' \in (1,2)$, $u,v > 0$ and $u+v < 1 - \frac{\alpha'}{2}$, then

$$\frac{1}{q(u)} + \frac{1}{q(1-u-v)} - \frac{1}{q(1-v)} > 0. \tag{137}$$

Proof: In view of (90), the left side of (137) is equal to

$$\frac{\overbrace{(2\alpha' u)}^{>0}\ \overbrace{(\alpha'+v-1)}^{>0}\ \overbrace{(u+v-1)}^{<0}}{\underbrace{(2u-\alpha')}_{<0}\ \underbrace{(2-2u-2v-\alpha')}_{>0}\ \underbrace{(2-2v-\alpha')}_{>0}} > 0.$$

■

Lemma 3: If $n \geq 2$, $\alpha' \in (1,2)$ and

$$\begin{aligned} & t_1,\dots,t_{n-1} > 0, \\ & \sum_{k=1}^{n-1} t_k < 1 - \frac{\alpha'}{2}, \\ & t_n = 1 - \sum_{k=1}^{n-1} t_k \end{aligned} \tag{138}$$

then

$$\sum_{k=1}^{n} \frac{1}{q(t_k)} > 0. \tag{139}$$

Proof: Lemma 3 is proved by using mathematical induction on n. In view of Lemma 1, (139) holds for $n=2$. Assuming its correctness for n, we have

$$\sum_{j=1}^{n-1} \frac{1}{q(t_j)} + \frac{1}{q(\bar{t}_n)} > 0 \tag{140}$$

where, from (138), $\bar{t}_n = 1 - \sum_{k=1}^{n-1} t_k$. We prove in the following that (139) also holds for $n+1$ when the constraints in (138) are satisfied with $n+1$, i.e.,

$$\begin{aligned} & t_1, \ldots, t_n > 0, \\ & \sum_{k=1}^{n} t_k < 1 - \frac{\alpha'}{2}, \\ & t_{n+1} = 1 - \sum_{k=1}^{n} t_k. \end{aligned} \tag{141}$$

Consequently, the left side of (139) is equal to

$$\begin{aligned} \sum_{k=1}^{n+1} \frac{1}{q(t_k)} &= \sum_{k=1}^{n-1} \frac{1}{q(t_k)} + \frac{1}{q(t_n)} + \frac{1}{q(t_{n+1})} \\ &> -\frac{1}{q(\bar{t}_n)} + \frac{1}{q(t_n)} + \frac{1}{q(t_{n+1})} \qquad (142) \\ &= \frac{1}{q(t_n)} + \frac{1}{q\left(1 - \sum_{k=1}^{n} t_k\right)} - \frac{1}{q(1 - \sum_{k=1}^{n-1} t_k)} \qquad (143) \\ &> 0 \qquad (144) \end{aligned}$$

where (142) follows from (140); (143) holds by the equality constraint in (141); (144) follows from Lemma 2 by setting $u = t_n$, $v = \sum_{k=1}^{n-1} t_k$ which satisfy $u + v < 1 - \frac{\alpha'}{2}$ in view of (141). Hence, it follows by mathematical induction that Lemma 3 holds for every $n \geq 2$. ■

In the following, we prove the concavity of f when $q(t_k) \neq 0$ for all $k \in \{1, \ldots, n-1\}$ (recall that the case where there exits $k \in \{1, \ldots, n-1\}$ such that $q(t_k) = 0$ was addressed in the paragraph before Lemma 1). Without loss of generality, we prove that $\nabla^2 f(\underline{t}) \preceq 0$ when $\big(q(t_1), \ldots, q(t_{n-1})\big)$ is a vector whose all entries are distinct. To justify this assumption, note that since the function q in (90) is monotonically increasing ($q'(t) = \frac{1}{t^2} + \frac{1}{(\alpha'-t)^2} > 0$), we actually restrict ourselves under the latter assumption to the case where the entries of the vector $(t_1, \ldots, t_{n-1})$ are all distinct. Otherwise, if some of the entries of the vector $(t_1, \ldots, t_{n-1})$ are equal, then the proof that the Hessian matrix is non-positive definite continues to hold by relying on the satisfiability of this property when all the entries of $(t_1, \ldots, t_{n-1})$ are distinct, and from the continuity in $\underline{t}$ of the eigenvalues of the Hessian matrix $\nabla^2 f(\underline{t})$.

Since the optimization problem in (86) is invariant to a permutation of the entries of $\underline{t}$, it is assumed without loss of generality that

$$q(t_1) < q(t_2) < \ldots < q(t_{n-1}). \tag{145}$$

In view of (145), there are only two possibilities: either

$$q(t_1) < q(t_2) < \ldots < q(t_{n-2}) < q(t_{n-1}) < 0, \tag{146}$$

or

$$q(t_1) < q(t_2) < \ldots < q(t_{n-2}) < 0 < q(t_{n-1}) \tag{147}$$

as if it was possible that $q(t_{n-2}) \geq 0$, it would have implied that $q(t_{n-1}) > q(t_{n-2}) \geq 0$ which in turn yields that $t_{n-1} > t_{n-2} \geq \frac{\alpha'}{2}$. This, however, cannot be true since otherwise

$$\sum_{k=1}^{n-1} t_k \geq t_{n-2} + t_{n-1} > \alpha' > 1$$

which violates the inequality constraint $\sum_{k=1}^{n-1} t_k \leq 1$ in (88).

The continuation of this proof relies on Fact 3 by Bunch *et al.* [14] (see Section II-B), and on Lemma 3. Let

$$t_n = 1 - \sum_{k=1}^{n-1} t_k. \tag{148}$$

Case 1: If (146) holds, then (92) implies that

$$D \prec 0. \tag{149}$$

- If $q(t_n) < 0$ then $\rho = q(t_n)\mathbf{1}\mathbf{1}^T \prec 0$ which, in view of (91) and (149), implies that $\nabla^2 f(t_1, \ldots, t_{n-1}) \prec 0$.
- Otherwise, if $q(t_n) > 0$ then $\rho > 0$ (see (92) and (148)); from (91) and the interlacing property in (22), the eigenvalues $\lambda_1, \ldots, \lambda_{n-1}$ of $\nabla^2 f(\underline{t})$ satisfy

$$q(t_1) < \lambda_1 < q(t_2) < \ldots < q(t_{n-2}) < \lambda_{n-2} < q(t_{n-1}) < \lambda_{n-1} \tag{150}$$

where, in view of the third item of Fact 3, the inequalities in (150) are strict. From (146) and (150), it follows that $\lambda_1, \ldots, \lambda_{n-2} < 0$. To prove that $\nabla^2 f(t_1, \ldots, t_{n-1}) \prec 0$, it remains to show that also $\lambda_{n-1} < 0$. In view of the third item of Fact 3 and (91), the eigenvalues of $\nabla^2 f(t_1, \ldots, t_{n-1})$ satisfy the equation

$$1 + q(t_n) \sum_{j=1}^{n-1} \frac{1}{q(t_j) - \lambda} = 0 \tag{151}$$

which therefore implies that, for all $k \in \{1, \ldots, n-1\}$,

$$\sum_{j=1}^{n-1} \frac{1}{\lambda_k - q(t_j)} = \frac{1}{q(t_n)}. \tag{152}$$

Let us assume on the contrary that $\lambda_{n-1} > 0$. Since it is assumed here that $q(t_n) > 0$ then $t_n > \frac{\alpha'}{2}$, and it follows from (148) that

$$\sum_{k=1}^{n-1} t_k < 1 - \frac{\alpha'}{2}. \tag{153}$$

Since $q(t_j) < 0$ for all $j \in \{1, \ldots, n-1\}$, if $\lambda_{n-1} > 0$, then in view of (152)

$$\begin{aligned} \sum_{j=1}^{n-1} \frac{1}{-q(t_j)} &\geq \sum_{j=1}^{n-1} \frac{1}{\lambda_{n-1} - q(t_j)} \\ &= \frac{1}{q(t_n)}. \end{aligned} \tag{154}$$

Rearrangement of terms in (154) yields

$$\sum_{j=1}^{n} \frac{1}{q(t_j)} \leq 0 \tag{155}$$

and, in view of the interior of $\mathcal{D}^{n-1}$ in (88), and (148) and (153), inequality (155) contradicts the result in Lemma 3. This therefore proves by contradiction that $\lambda_{n-1} < 0$, so all the $n-1$ eigenvalues of the Hessian are negative, and therefore f is strictly concave under the assumption in (146).

Case 2: We now consider the case where (147) holds. Under this assumption,

$$q(t_n) < 0. \tag{156}$$

To verify (156), note that $q(t_{n-1}) > 0$ yields that $t_{n-1} > \frac{\alpha'}{2}$; assume by contradiction that $q(t_n) \geq 0$, then $t_n \geq \frac{\alpha'}{2}$ (see (90)) which implies that $\sum_{j=1}^{n} t_j \geq t_n + t_{n-1} > \alpha' > 1$ in contradiction to the equality $\sum_{j=1}^{n} t_j = 1$ in (148); hence, indeed $q(t_n) < 0$. Consequently, in view of (91), let

$$\overline{C} = \frac{1}{q(t_n)} \, \nabla^2 f(t_1, \ldots, t_{n-1}) \tag{157}$$

$$= \overline{D} + \mathbf{1}\mathbf{1}^T \tag{158}$$

where

$$\overline{D} = \mathrm{diag}\left(\frac{q(t_1)}{q(t_n)}, \ldots, \frac{q(t_{n-1})}{q(t_n)} \right). \tag{159}$$

From (147) and (156), it follows that

$$\frac{q(t_1)}{q(t_n)} > \frac{q(t_2)}{q(t_n)} > \ldots > \frac{q(t_{n-2})}{q(t_n)} > 0 > \frac{q(t_{n-1})}{q(t_n)}. \tag{160}$$

It is shown in the following that $\overline{C} \succeq 0$ which, from (156) and (157), imply that indeed $\nabla^2 f(t_1, \ldots, t_{n-1}) \preceq 0$. Let $\{\lambda_k\}_{k=1}^{n-1}$ designate the eigenvalues of $\overline{C}$; in view of (158) and the last two items of Fact 3, it follows that

$$\overbrace{\frac{q(t_{n-1})}{q(t_n)}}^{<0} < \lambda_1 < \overbrace{\frac{q(t_{n-2})}{q(t_n)}}^{>0} < \lambda_2 < \ldots < \overbrace{\frac{q(t_2)}{q(t_n)}}^{>0} < \lambda_{n-2} < \overbrace{\frac{q(t_1)}{q(t_n)}}^{>0} < \lambda_{n-1}. \tag{161}$$

Hence, (161) asserts that $\lambda_2, \ldots, \lambda_{n-1} > 0$, and it only remains to prove that $\lambda_1 > 0$. From the third item of Fact 3, and from (157), (158), (159), the eigenvalues $\{\lambda_k\}_{k=1}^{n}$ of the rank-one modification $\overline{C}$ satisfy the equality

$$1 + \sum_{j=1}^{n-1} \frac{1}{\frac{q(t_j)}{q(t_n)} - \lambda_k} = 0 \tag{162}$$

for all $k \in \{1, \ldots, n-1\}$. Assume on the contrary that $\lambda_1 \leq 0$, then from (162)

$$1 + \sum_{j=1}^{n-1} \frac{q(t_n)}{q(t_j)} \geq 1 + \sum_{j=1}^{n-1} \frac{1}{\frac{q(t_j)}{q(t_n)} - \lambda_1} = 0. \tag{163}$$

Consequently, from (156) and (163), it follows that $\sum_{j=1}^{n} \frac{1}{q(t_j)} \leq 0$ in contradiction to Lemma 3. Hence, all $\lambda_k > 0$ for $k \in \{1, \ldots, n-1\}$, which therefore implies that $\nabla^2 f(t_1, \ldots, t_{n-1}) \prec 0$ for all $(t_1, \ldots, t_{n-1})$ in the interior of $\mathcal{D}^{n-1}$. This completes the proof of Proposition 1.

APPENDIX C
DERIVATION OF (97)–(99) FROM LAGRANGE DUALITY

We consider the convex optimization problem in (86), and solve it via the use of the Lagrange duality where strong duality holds.

The Lagrangian of the convex optimization problem in (86) is given by

$$\begin{aligned}
&L(t_1, \ldots, t_{n-1}; \lambda_1, \ldots, \lambda_n) \\
&= \sum_{k=1}^{n-1} g(t_k) + g\left(1 - \sum_{k=1}^{n-1} t_k\right) + \sum_{k=1}^{n-1} t_k \log N_k \\
&\quad + \left(1 - \sum_{k=1}^{n-1} t_k\right) \log N_n + \sum_{k=1}^{n-1} \lambda_k t_k + \lambda_n \left(1 - \sum_{k=1}^{n-1} t_k\right)
\end{aligned} \tag{164}$$

where $\lambda \succeq 0$, the function g is defined in (82), and $N_k := N_\alpha(X_k)$ (see (83)).

In view of the Lagrangian in (164) and the function g defined in (82), straightforward calculations of the partial derivatives of L with respect to t_k for $k \in \{1, \ldots, n-1\}$ yields

$$\begin{aligned}
\frac{\partial L}{\partial t_k} &= g'(t_k) - g'(1 - t_1 - \ldots - t_{n-1}) + \log\left(\frac{N_\alpha(X_k)}{N_\alpha(X_n)}\right) + \lambda_k - \lambda_n \\
&= -\log\left(t_k\Big(1 - \frac{t_k}{\alpha'}\Big)\right) + \log\left(t_n\Big(1 - \frac{t_n}{\alpha'}\Big)\right) + \log\left(\frac{N_\alpha(X_k)}{N_\alpha(X_n)}\right) + \lambda_k - \lambda_n
\end{aligned} \tag{165}$$

where $t_n := 1 - \sum_{k=1}^{n-1} t_k$. By setting the partial derivatives in (165) to zero, and exponentiating both sides of the equation, we get for all $k \in \{1, \ldots, n-1\}$

$$\frac{t_n(\alpha' - t_n)}{t_k(\alpha' - t_k)} = \frac{N_\alpha(X_n)}{N_\alpha(X_k)} \cdot \exp(\lambda_n - \lambda_k). \tag{166}$$

In view of (166) and the definition of $\{c_k\}_{k=1}^{n-1}$ in (95), we obtain that for all $k \in \{1, \ldots, n-1\}$

$$t_k(\alpha' - t_k) = c_k\, t_n(\alpha' - t_n)\, \exp(\lambda_k - \lambda_n). \tag{167}$$

Consequently, (167), the definition of t_n, and the slackness conditions lead to the following set of constraints:

$$t_k \geq 0, \quad k \in \{1, \ldots, n\} \tag{168}$$

$$\sum_{k=1}^{n} t_k = 1 \tag{169}$$

$$\lambda_k \geq 0, \quad k \in \{1, \ldots, n\} \tag{170}$$

$$\lambda_k t_k = 0, \qquad k \in \{1, \ldots, n\} \tag{171}$$

$$t_k(\alpha' - t_k) = c_k t_n(\alpha' - t_n)\, \exp(\lambda_k - \lambda_n), \quad k \in \{1, \ldots, n-1\} \tag{172}$$

with the variables $\underline{\lambda}$ and $\underline{t}$ in (168)–(172).

Consider first the case where

$$N_\alpha(X_k) > 0, \quad \forall\, k \in \{1, \ldots, n-1\} \tag{173}$$

which in view of (95), implies

$$c_k > 0, \quad \forall\, k \in \{1, \ldots, n-1\}. \tag{174}$$

Under the assumption in (173), we prove that

$$\lambda_k = 0, \quad \forall\, k \in \{1, \dots, n\}. \tag{175}$$

Assume on the contrary that there exists an index k such that $\lambda_k \neq 0$. This would imply from (171) that $t_k = 0$. If $k = n$ (i.e., if $t_n = 0$) then it follows from (172) that also $t_k = 0$ for all $k \in \{1, \dots, n\}$ (recall that $\alpha' > 1$), which violates the equality constraint in (169). Otherwise, if $t_k = 0$ for some $k < n$, then it follows from (172) and (174) that $t_n = 0$ which leads to the same contradiction as above.

The substitution of (175) into the right side of (172) gives the simplified equation in (97). In view of (168) and (169), this leads to the simplified set of KKT constraints in (97)–(99).

Finally, if the assumption in (173) does not hold, i.e., $N_\alpha(X_k) = 0$ for some $k \in \{1, \dots, n-1\}$, then the optimal solution satisfies $t_k = 0$ (with the convention that $0 \cdot \log 0 = 0$) since any other assignment makes the objective function in (89) be equal to $-\infty$. In addition, in this case $c_k = 0$, so the simplified set of KKT constraints in (97)–(99) still yields the optimal solution $\underline{t}$.

Appendix D
On the existence and uniqueness of the solution in (103)

Define

$$\phi_\alpha(x) = x + \sum_{k=1}^{n-1} \psi_{\alpha,k}(x), \quad x \in [0, 1], \tag{176}$$

and note that we need to show that there exists a unique solution of the equation $\phi_\alpha(x) = 1$ where $x \in [0, 1]$. From the continuity of $\phi_\alpha(\cdot)$ and since $\phi_\alpha(0) = 0$ and

$$\phi_\alpha(1) = 1 + \sum_{k=1}^{n-1} \psi_{\alpha,k}(1) > 1, \tag{177}$$

the existence of such a solution is assured. To prove uniqueness, consider two cases: $\alpha' \geq 2$ and $1 < \alpha' < 2$.

The derivative of $\phi_\alpha(x)$ is given by

$$\phi'_\alpha(x) = 1 + \sum_{k=1}^{n-1} \frac{c_k(\alpha' - 2x)}{\sqrt{\alpha'^2 - 4c_k x(\alpha' - x)}}, \tag{178}$$

so if $\alpha' \geq 2$, then $\phi_\alpha(x)$ is monotonically increasing in $[0, 1]$, hence the solution $t_n \in [0, 1]$ of the equation (103) is unique.

If $\alpha' \in (1, 2)$, then

$$\phi'_\alpha(x) > 0, \qquad x \in [0, \tfrac{\alpha'}{2}]. \tag{179}$$

Note that

$$\alpha'^2 - 4c_k x(\alpha' - x) = \alpha'^2(1 - c_k) + c_k(2x - \alpha')^2,$$

thus in view of (178),

$$\phi'_\alpha(x) = 1 + \sum_{k=1}^{n-1} \frac{c_k}{\sqrt{c_k + \frac{\alpha'^2(1-c_k)}{4(x - \frac{\alpha'}{2})^2}}}. \tag{180}$$

Eq. (180) implies that $\phi'_\alpha(\cdot)$ is monotonically decreasing in $(\frac{\alpha'}{2}, 1]$; in other words, $\phi_\alpha(\cdot)$ is concave in the interval $(\frac{\alpha'}{2}, 1])$.

Assume on the contrary that there are two solutions, $0 < x_1 < x_2 < 1$ to (103), i.e.,

$$\phi_\alpha(x_1) = \phi_\alpha(x_2) = 1. \tag{181}$$

Eq. (181) implies that there exists $c \in (x_1, x_2)$ such that $\phi'_\alpha(c) = 0$ and from (179), $c \in (\frac{\alpha'}{2}, x_2)$. Since $\phi'_\alpha(\cdot)$ is monotonically decreasing in $(\frac{\alpha'}{2}, 1]$, it follows that $\phi'_\alpha(x) < 0$ for all $x \in (c, 1)$. Hence, $\phi_\alpha(\cdot)$ is monotonically decreasing in $(x_2, 1)$, which leads to the contradiction

$$1 < \phi_\alpha(1) < \phi_\alpha(x_2) = 1.$$

This therefore demonstrates the uniqueness of the solution in both cases.

Appendix E
On the Asymptotic Equivalence of (104) and (105)

If $N_\infty(X_k) = 0$ for all $k \in \{1, \ldots, n\}$, the bounds in (104) and (105) obviously coincide asymptotically as $\alpha \to \infty$. In addition, in this case, the condition in (106) clearly holds as well. It is therefore assumed that $N_\infty(X_k)$ is strictly positive for at least one value of $k \in \{1, \ldots, n\}$ which, under the assumption in (93), yields that

$$N_\infty(X_n) > 0. \tag{182}$$

Let $c_k^\star$ be defined as

$$c_k^\star = \lim_{\alpha \to \infty} \frac{N_\alpha(X_k)}{N_\alpha(X_n)} = \frac{N_\infty(X_k)}{N_\infty(X_n)}. \tag{183}$$

In view of (183), the condition in (106) is equivalent to

$$\sum_{k=1}^{n-1} c_k^\star \leq 1. \tag{184}$$

Hence, it remains to show that the tightest R-EPI in (104) and the BV bound in (105) asymptotically coincide, by letting $\alpha \to \infty$, if and only if the condition in (184) holds.

Let $\phi_\alpha \colon [0,1] \to \mathbb{R}$ be the function defined in (176) for $\alpha \in (1, \infty)$, and define

$$\phi_\infty(x) = \lim_{\alpha \to \infty} \phi_\alpha(x) \tag{185}$$

for $x \in [0,1]$. In view of (102), (176) and (183), the limit in (185) is given by

$$\phi_\infty(x) = x + \tfrac{1}{2} \sum_{k=1}^{n-1} \left(1 - \sqrt{1 - 4c_k^\star \, x(1-x)}\right) \tag{186}$$

for $x \in [0,1]$. Recall that under the assumption in (93), the selection of $t_n = 1$ in (89) leads to the BV bound in (105). Hence, in view of (103), if $t = 1$ is the unique solution of

$$\phi_\infty(t) = 1, \quad t \in [0,1] \tag{187}$$

then the bounds in (104) and (105) asymptotically coincide by letting $\alpha \to \infty$. Note that,

$$\phi_\infty(0) = 0, \tag{188}$$

$$\phi_\infty(1) = 1. \tag{189}$$

From (189), $t = 1$ is a solution of (187) regardless of the sequence $\{c_k^\star\}$. Moreover, from (186),

$$\phi'_\infty(x) = 1 + \sum_{k=1}^{n-1} \frac{c_k^\star \, (1-2x)}{\sqrt{1 - 4c_k^\star \, x(1-x)}}, \tag{190}$$

so

$$\phi'_\infty(x) > 0, \quad \forall \, x \in (0, \tfrac{1}{2}), \tag{191}$$

$$\phi'_\infty(1) = 1 - \sum_{k=1}^{n-1} c_k^\star. \tag{192}$$

The function $\phi'_\infty(\cdot)$ is monotonically decreasing in the interval $[\frac{1}{2}, 1]$; this concavity property of ϕ_∞ can be justified by Appendix D since the function $\phi_\alpha(\cdot)$ is concave in $[\frac{\alpha'}{2}, 1]$ and $\alpha' \to 1$ by letting $\alpha \to \infty$. Thus, if the condition in (184) holds, then $\phi'_\infty(x) > 0$ for all $x \in (0,1)$ which, in view of (189), yields that $t = 1$ is the unique solution of (187). This implies that the tightest R-EPI in (104) and the BV bound in (105) asymptotically coincide by letting $\alpha \to \infty$.

To prove the 'only if' part, one needs to show that if the condition in (184) does not hold then the bounds in (104) and (105) do not coincide asymptotically in the limit where $\alpha \to \infty$; in the latter case, we prove that our bound in (104) is tighter than (105). If (184) does not hold, then (192) implies that

$$\phi'_\infty(1) < 0. \tag{193}$$

Hence, from (189), there exists $x_0 \in (0,1)$ such that $\phi_\infty(x_0) > 1$ which, in view of (188) and the continuity of $\phi_\infty(\cdot)$, implies that there exists $t \in (0, x_0)$ which is a solution of (187). This implies that there are two different solutions of (187) in the interval $[0,1]$. Let $t^{(1)} \in (0,1)$ and $t^{(2)} = 1$ denote such solutions, i.e.,

$$t^{(1)} < t^{(2)} = 1. \tag{194}$$

Note that there are no solutions of the equation $\phi_\infty(t) = 1$ in $[0,1]$, except for $t^{(1)}$ and $t^{(2)} = 1$ since $\phi_\infty(\cdot)$ is monotonically increasing in $[0, \frac{1}{2}]$ and it is concave in $[\frac{1}{2}, 1]$ with $\phi_\infty(1) = 1$.

We need to show that $t^{(1)}$ leads to an R-EPI which is tighter than the R-EPI in (105); the bound in (105) corresponds to $t^{(2)} = 1$ under the assumption in (93). For every $\alpha > 1$, let $t(\alpha)$ be the unique solution of (103) (see Appendix D). It follows that the limit of any convergent subsequence $\{t(\alpha_n)\}$, as $\alpha_n \to \infty$, is either $t^{(1)} \in (0,1)$ or $t^{(2)} = 1$. In the sequel, if the condition in (106) is not satisfied, we show that every such subsequence tends to $t^{(1)} \in (0,1)$, which therefore implies that

$$\lim_{\alpha \to \infty} t(\alpha) = t^{(1)} < 1. \tag{195}$$

From (193) and the continuity of $\phi_\infty(\cdot)$, it follows that there exists $\delta > 0$ such that

$$\phi_\infty(x) > 1, \; \forall \, x \in (1-\delta, 1). \tag{196}$$

In addition, since $\phi_\alpha(\cdot)$ is continuous in α for every $x \in [0,1]$, it follows from (196) that there exists $\alpha_0 > 1$ such that $\phi_\alpha(x) > 1$ for all $\alpha > \alpha_0$ and $x \in (1-\delta, 1]$ (note that the rightmost point is included in this interval in view of (177)). Hence, since by definition $\phi_\alpha\bigl(t(\alpha)\bigr) = 1$ for all $\alpha \in (1, \infty)$ then $t(\alpha) \leq 1 - \delta$ for all $\alpha > \alpha_0$. This therefore proves that every subsequence $\{t(\alpha_n)\}$ tends to $t^{(1)}$ as $\alpha_n \to \infty$ (since it cannot converge to $t^{(2)} = 1$), which yields (195). Hence, the R-EPI in Theorem 2 asymptotically yields a tighter bound than (105) when $\alpha \to \infty$; this therefore proves the 'only if' part of our claim.

Acknowledgement

This work has been supported by the Israeli Science Foundation (ISF) under Grant 12/12. We would like to thank the Associate Editor, and the anonymous reviewers for their valuable feedback which helped to improve the presentation in this paper. A discussion with Sergio Verdú is acknowledged.

REFERENCES

[1] V. Anantharam, "Entropy power inequalities: results and speculation," *Workshop on Coding and Information Theory*, Institute of Mathmetical Research, the University of Hong Kong, December 2013.

[2] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Information Theory - 2nd Colloquium*, Keszthely, Hungary, 1975, Colloquia Mathematica Societatis Janós Bolyai (I. Csiszár and P. Elias editors), Amsterdam, the Netherlands: North Holland, vol. 16, pp. 41-52, 1977.

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

[4] W. Beckner, "Inequalities in Fourier analysis on $\mathbb{R}^n$," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 72, no. 2, pp. 638-641, February 1975.

[5] J. F. Bercher and C. Vignat, "A Rényi entropy convolution inequality with application," *Proceedings of the 11th European Signal Processing Conference*, Toulouse, France, pp. 1–4, September 2002.

[6] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. on Information Theory*, vol. 20, no. 2, pp. 279–280, March 1974.

[7] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. on Information Theory*, vol. 11, no. 2, pp. 267–271, April 1965.

[8] S. G. Bobkov and G. P. Chistyakov, "Bounds on the maximum of the density for sums of independent random variables," *Journal of Mathematical Sciences*, vol. 199, no. 2, pp. 100–106, May 2014.

[9] S. G. Bobkov and G. P. Chistyakov, "Entropy power inequality for the Rényi entropy," *IEEE Trans. on Information Theory*, vol. 61, no. 2, pp. 708–714, February 2015.

[10] S. Bobkov and M. Madiman, "Dimensional behaviour of entropy and information," *Comptes Rendus Mathematique*, vol. 349, no. 3–4, pp. 201–204, February 2011.

[11] S. Bobkov and M. Madiman, "On the problem of reversibility of the entropy power inequality," *Limit Theorems in Probability, Statistics and Number Theory*, in honor of F. Götze's 60th birthday, P. Eichelsbacher et al. editors, Springer Proceedings in Mathematics and Statistics, vol. 42, pp. 61-74, Springer-Verlag, 2013.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[13] H. J. Brascamp and E. H. Lieb, "Best constants in Young's inequality, its converse, and its generalization to more than three functions," *Advances in Mathematics*, vol. 20, no. 2, pp. 151-173, May 1976.

[14] J. R. Bunch, C. P. Nielsen and D. C. Sorensen, "Rank-one modification of the symmetric eigenproblem," *Numerische Mathematik*, Springer, vol. 31, no. 1, pp. 31–48, March 1978.

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

[16] T. A. Courtade, "Strengthening the entropy power inequality," *Proceedings of the 2016 IEEE International Symposium on Information Theory*, pp. 2294–2298, Barcelona, Spain, July 2016.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, second edition, 2006.

[18] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. on Information Theory*, vol. 37, no. 6, pp. 1501–1518, November 1991.

[19] T. van Erven and P. Harremöes, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.

[20] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Trans. on Information Theory*, vol. 60, no. 11, pp. 6801–6810, November 2014.

[21] D. Guo, S. Shamai (Shitz) and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. on Information Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[22] D. Guo, S. Shamai (Shitz) and S. Verdú, "Proof of entropy power inequalities via MMSE," *IEEE 2006 International Symposium on Information Theory*, Seattle, Washington, USA, pp. 1011–1015, July 2006.

[23] D. Guo, S. Shamai (Shitz) and S. Verdú, "The interplay between information and estimation measures," *Foundations and Trends in Signal Processing*, vol. 6, no. 4, pp. 243–429, 2012.

[24] S. Haghighatshoar, E. Abbe and I. E. Telatar, "A new entropy power inequality for integer-valued random variables," *IEEE Trans. on Information Theory*, vol. 60, no. 7, pp. 3787–3796, July 2014.

[25] P. Harremoës and C. Vignat, "An entropy power inequality for the binomial family," *Journal of Inequalities in Pure and Applied Mathematics*, vol. 4, no. 5, article 93, 2003.

[26] V. Jog and V. Anantharam, "The entropy power inequality and Mrs. Gerber's lemma for groups of order 2^n," *IEEE Trans. on Information Theory*, vol. 60, no. 7, pp. 3773–3786, July 2014.

[27] O. Johnson, *Information Theory and the Central Limit Theorem*, Imperial College Press, 2004.

[28] O. Johnson and C. Vignat, "Some results concerning maximum Rényi entropy distributions," *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, vol. 43, no. 3, pp. 339–351, May/June 2007.

[29] O. Johnson and Y. Yu, "Monotonicity, thinning, and discrete versions of the entropy power inequality," *IEEE Trans. on Information Theory*, vol. 56, no. 11, pp. 5387–5395, November 2010.

[30] S. K. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. on Information Theory*, vol. 24, no. 4, pp. 451–456, July 1978.

[31] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. on Information Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.

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

[33] M. Madiman, "On the entropy of sums," *Proceedings of the 2008 IEEE International Symposium on Information Theory*, pp. 303–307, Porto, Portugal, May 2008.

[34] M. Madiman, J. Melbourne and P. Xu, "Forward and reverse entropy power inequalities in convex geometry," preprint, April 2016. [Online]. Available at http://arxiv.org/abs/1604.04225.

[35] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. on Information Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[36] A. Rényi, "On measures of entropy and information," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 547–561, 1961.

[37] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. on Information Theory*, vol. 57, no. 1, pp. 33–55, January 2011.

[38] B. A. Rogozin, "Estimation of the maximum of a convolution of bounded densities," (translated from Russian), *Theory of Probability and Its Applications*, vol. 32, no. 1, pp. 48–56, 1988.

[39] G. Savaré and G. Toscani, "The concavity of Rényi entropy power," *IEEE Trans. on Information Theory*, vol. 60, no. 5, pp 2687–2693, May 2014.

[40] S. Shamai and A. D. Wyner, "A binary analog to the entropy-power inequality," *IEEE Trans. on Information Theory*, vol. 36, no. 6, pp. 1428–1430, November 1990.

[41] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and pp. 623–656, July and October 1948.

[42] N. Sharma, S. Das and S. Muthukrishnan, "Entropy power inequality for a family of discrete random variables," *Proceedings of the 2011 IEEE International Symposium on Information Theory*, pp. 1871–1875, Saint Petersburg, Russia, August 2011.

[43] O. Shayevitz, "On Rényi measures and hypothesis testing," *Proceedings of the 2011 IEEE International Symposium on Information Theory*, pp. 894–898, Saint Petersburg, Russia, July 2011.

[44] 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, June 1959.

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

[46] S. Verdú and D. Guo, "A simple proof of the entropy power inequality," *IEEE Trans. on Information Theory*, vol. 52, no. 5, pp. 2165–2166, May 2006.

[47] L. Wang and M. Madiman, "Beyond the entropy power inequality, via rearrangements," *IEEE Trans. on Information Theory*, vol. 60, no. 9, pp. 5116–5137, September 2014.

[48] L. Wang, J. O. Woo and M. Madiman, "A lower bound on the Rényi entropy of convolutions in the integers," *Proceedings 2014 IEEE International Symposium on Information Theory*, pp. 2829–2833, Honolulu, Hawaii, USA, July 2014.

[49] H. Weingarten, Y. Steinberg and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Information Theory*, vol. 52, no. 9, pp. 3936–3964, September 2006.

[50] J. O. Woo and M. Madiman, "A discrete entropy power inequality for uniform distributions," *Proceedings 2015 IEEE International Symposium on Information Theory*, pp. 1625–1629, Hong Kong, China, June 2015.

[51] P. Xu, J. Melbourne and M. Mokshay, “Reverse entropy power inequalities for s-concave densities,” *Proceedings of the 2016 IEEE International Symposium on Information Theory*, pp. 2284–2288, Barcelona, Spain, July 2016.

[52] R. Zamir and M. Feder, “A generalization of the entropy power inequality with applications,” *IEEE Trans. on Information Theory*, vol. 39, no. 5, pp. 1723–1728, September 1993.