

ON RÉNYI ENTROPY POWER
INEQUALITIES

ESHED RAM

ON RÉNYI ENTROPY POWER INEQUALITIES

RESEARCH THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
IN ELECTRICAL ENGINEERING

ESHED RAM

SUBMITTED TO THE SENATE OF THE TECHNION – ISRAEL INSTITUTE OF TECHNOLOGY

Kislev 5777

HAIFA

DECEMBER 2016

THIS RESEARCH THESIS WAS DONE UNDER THE
SUPERVISION OF ASSOCIATE PROF. IGAL SASON IN THE
DEPARTMENT OF ELECTRICAL ENGINEERING

ACKNOWLEDGEMENT

I would like to acknowledge my supervisor, associate Prof. Igal Sason, for his support, assistance and guidance throughout this research. His devotion is more than appreciated and his professionalism revealed the world of research to me in a profound and interesting way. I especially thank him for his guidance concerning academic writing.

THE GENEROUS FINANCIAL HELP OF THE TECHNION IS
GRATEFULLY ACKNOWLEDGED

This work was published in the *IEEE Transactions on Information Theory* in December 2016. It was partially presented in the *International Symposium on Information Theory*, Barcelona, July 2016.

To my wife
Noga Ram

Contents

Abstract	2
1 Introduction	3
2 Analytical Tools	7
2.1 Rényi Entropy	7
2.2 Basic Inequalities	9
2.3 Rank-One Modification of a Symmetric Eigenproblem	12
3 A New Rényi EPI	13
4 A Further tightening of the R-EPI	22
4.1 A Tightened R-EPI for $n \geq 2$	22
4.1.1 The optimization problem in (3.18)	23
4.1.2 A reduction of the optimization problem in (3.18) to $n - 1$ variables .	23
4.1.3 Proving the convexity of the optimization problem in (4.6)	24
4.1.4 Solution of the convex optimization problem in (4.6)	25
4.2 A Tightened R-EPI for $n = 2$	31
4.3 Example: The Rényi Entropy Difference Between Data and its Filtering . . .	34

Contents - Continued

5	Summary	37
5.1	Research Results and Tools	37
5.1.1	Remarks on Tightness	38
5.2	Further Research	38
5.2.1	The $\alpha < 1$ Case	38
5.2.2	Further Generalizations	41
A	Proof of (3.4)	46
B	Proof of Proposition 1	48
C	Derivation of (4.17)–(4.19) From Lagrange Duality	55
D	On the existence and uniqueness of the solution to (4.23)	58
E	On the Asymptotic Equivalence of (4.24) and (4.25)	60
	References	63

List of Figures

3.1	A plot of $c_\alpha^{(n)}$ in (3.2), as a function of α , for $n = 2, 3, 10$ and $n \rightarrow \infty$	19
4.1	A comparison of the Rényi entropy power inequalities for $n = 3$ independent random vectors according to [11] (BC), [7] (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 (4.26) does not hold, while in the latter it does).	30
4.2	A comparison of the Rényi entropy power inequalities for $n = 2$ independent random vectors according to [11] (BC), [7] (BV), Theorem 1 and the Corollary 3. The bounds refer to the two cases where $(N_\alpha(X_1), N_\alpha(X_2)) = (10, 10)$ or $(6, 14)$ (in both cases, the sum of the entries is 20).	33

Abstract

One of the well-known inequalities in information theory is the entropy power inequality (EPI) which has been introduced by Shannon in his 1948 landmark paper. The EPI has proved to be an instrumental tool in proving converse theorems for the capacity region of the Gaussian broadcast channel, the capacity region of the Gaussian wire-tap channel, the capacity region of the Gaussian interference channel, the capacity region of the Gaussian broadcast multiple-input multiple-output (MIMO) channel, and a converse theorem in multi-terminal lossy compression. Due to its importance, the EPI has been proved with information-theoretic tools in several insightful ways.

More studies on the theme include EPIs for discrete random variables and some analogies, generalized EPIs, reverse EPIs, related inequalities to the EPI in terms of rearrangements, and some refined versions of the EPI for specialized distributions.

The Rényi entropy and divergence have been introduced by Rényi in the sixties, as a generalization of the Shannon entropy and the Kullback-Leibler relative entropy, and they evidence a long track record of usefulness in information theory and its applications. It is therefore of interest to consider the generalization of the EPI to Rényi measures.

This Master thesis focuses on Rényi entropy power inequalities (R-EPI), which generalize the celebrated EPI of Shannon to the Rényi entropy power. Consider a sum of $S_n = \sum_{k=1}^n X_k$ independent continuous random vectors taking values in \mathbb{R}^d , and let $\alpha \in [1, \infty]$. A Rényi Entropy Power Inequality provides a lower bound Rényi entropy power of S_n of an arbitrary order α that, up to a multiplicative constant (which may depend in general on n , α and d), is equal to the sum of the Rényi entropy powers (of the same order α) of the n random vectors $\{X_k\}_{k=1}^n$. For $\alpha = 1$, the R-EPIs derived in the thesis by the author coincide with the entropy power inequality by Shannon.

The first R-EPI is an improvement of a recent R-EPI by Bobkov and Chistyakov (IEEE Trans. on Information Theory, Feb. 2015) which relies on the sharpened Young's inequality. A further improvement of the R-EPI relies on convex optimization and results on rank-one modification of a real-valued diagonal matrix. The latter form of the R-EPI improves both the inequality by Bobkov and Chistyakov as well as a bound by Bercher and Vignat, and it forms the tightest R-EPI known to date. The improvements are exemplified in the context of data filtering, and are shown to be asymptotically tight as $\alpha \rightarrow \infty$ for $n = 2$.

This work has been very recently published as a journal paper:

E. Ram and I. Sason, "On Rényi entropy power inequalities," *IEEE Trans. on Information Theory*, vol. 62, no. 12, pp. 6800–6815, December 2016. It was also presented in part at the ISIT 2016 conference.

List of Abbreviations and Notations

BC	Bobckov and Chistyakov
BV	Bercher and Vignat
EPI	Entropy Power Inequality
KKT	Karush-Kuhn-Tucker
R-EPI	Rényi Entropy Power Inequality
f_X	Probability density function of a random vector X
d	Dimension of all random vectors
n	Number of summands
$H(\cdot)$	Shannon entropy
$h(\cdot)$	Differential entropy
$N(\cdot)$	Differential entropy power
$h_\alpha(\cdot)$	Differential Rényi entropy of order α
$N_\alpha(\cdot)$	Differential Rényi entropy power of order α
α'	Hölder conjugate of α
$\ \cdot\ _\alpha$	L^α -norm
*	Convolution operation
$D(\cdot \cdot)$	Discrete Kullback-Leibler divergence
$d(\cdot \cdot)$	Binary Kullback-Leibler divergence
$M(\cdot)$	Essential supremum functional
\mathcal{P}^n	n -dimensional probability simplex
$L(\dots ; \dots)$	Lagrangian of an optimization problem

Chapter 1

Introduction

One of the well-known inequalities in information theory is the entropy power inequality (EPI) which has been introduced by Shannon [50, 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:

$$N\left(\sum_{k=1}^n X_k\right) \geq \sum_{k=1}^n N(X_k) \quad (1.1)$$

with equality in (1.1) 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 [8], the capacity region of the Gaussian wire-tap channel [37], the capacity region of the Gaussian interference channel [21], the capacity region of the Gaussian broadcast multiple-input multiple-output (MIMO) channel [59], and a converse theorem in multi-terminal lossy compression [43]. Due to its importance, the EPI has been proved with information-theoretic tools in several insightful ways (see, e.g., [9], [25], [29], [34, Appendix D], [45], [53], [56]); e.g., the proof in [56] relies on fundamental relations between information and estimation measures ([28], [30]), 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 [31, 32, 33, 36, 49, 51, 60], generalized EPIs [39, 40, 62], reverse EPIs [12, 13, 42, 61], related inequalities to the EPI in terms of rearrangements [57], and some refined versions of the EPI for specialized distributions [20, 22, 32, 55]. An overview on EPIs is provided in [2]; The reader is also referred to a preprint of a recent survey paper by Madiman *et al.* [42] which addresses forward and reverse EPIs with Rényi measures, and their connections with convex geometry.

The Rényi entropy and divergence have been introduced in [44], and they evidence a long track record of usefulness in information theory and its applications. These include a source coding theorem [19], guessing moments [3], generalized cutoff rates and error exponents for hypothesis testing [24, 52], encoding tasks [18], and a channel coding theorem [47]. Recent studies of the properties of these Rényi measures have been provided in [26], [27] and [52]. 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

$$h_\alpha(X) = \frac{1}{1-\alpha} \log \left(\int_{\mathbb{R}^d} f_X^\alpha(x) dx \right) \quad (1.2)$$

$$= \frac{\alpha}{1-\alpha} \log \|f_X\|_\alpha. \quad (1.3)$$

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

$$h_0(X) = \log \lambda(\text{supp}(f_X)), \quad (1.4)$$

$$h_1(X) = h(X) = - \int_{\mathbb{R}^d} f_X(x) \log f_X(x) dx, \quad (1.5)$$

$$h_\infty(X) = - \log(\text{ess sup}(f_X)) \quad (1.6)$$

where λ in (1.4) 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

$$N_\alpha(X) = \exp\left(\frac{2}{d} h_\alpha(X)\right). \quad (1.7)$$

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. This question is stated as follows:

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

$$N_\alpha\left(\sum_{k=1}^n X_k\right) \geq c_\alpha^{(n,d)} \sum_{k=1}^n N_\alpha(X_k) \quad (1.8)$$

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

In [35, 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 [57, 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 $\alpha > \frac{d}{d+2}$. In a recent preprint ([14]) a similar question is proposed and answered; the multiplicative constant $c_\alpha^{(n,d)}$ is forced to be equal to 1, but an exponent is added to each Rényi entropy power; i.e. $N_\alpha^p(\sum_{k=1}^n X_k) \geq \sum_{k=1}^n N_\alpha^p(X_k)$ for a suitable choice of p .

Question 1 has been recently addressed by Bobkov and Chistyakov [11], showing that (1.8) holds with

$$c_\alpha = \frac{1}{e} \alpha^{\frac{1}{\alpha-1}}, \quad \forall \alpha > 1 \quad (1.9)$$

independently of the values of n, d . It is the purpose of this thesis to derive some improved R-EPIs for $\alpha > 1$ (the case of $\alpha = 1$ refers to the EPI (1.1)). A study of Question 1 for $\alpha \in (0, 1)$ is currently an open problem (see [11, p. 709]) and is discussed in Chapter 5.

In view of the close relation in (1.3) between the (differential) Rényi entropy and the L_α norm, the sharpened version of Young's inequality plays a key role in [11] for the derivation of an R-EPI, as well as in this thesis for the derivation of some improved R-EPIs. The sharpened version of Young's inequality was also used by Dembo *et al.* [25] 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 (1.9) which gets the form

$$c_\alpha^{(n)} = \alpha^{\frac{1}{\alpha-1}} \left(1 - \frac{1}{n\alpha'}\right)^{n\alpha'-1} \quad (1.10)$$

independently of the dimension d . The new R-EPI with the constant in (1.10) asymptotically coincides with the tight bound by Rogozin [46] when $\alpha \rightarrow \infty$ and $n = 2$, and it also asymptotically coincides with the R-EPI in [11] when $n \rightarrow \infty$. Moreover, the R-EPI with the new constant in (1.10) 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 thesis is organized as follows: In Chapter 2, preliminary material and notations are introduced. A new R-EPI is derived in Chapter 3 for $\alpha > 1$, and special cases of this improved bound are studied. Chapter 4 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 4.3 exemplifies numerically the tightness of the new R-EPIs in comparison to some previously reported bounds. Finally, Chapter 5 summarizes this thesis, discusses the tightness of the R-EPIs in it, and suggests some future research items, including $\alpha < 1$ case and some more generalizations of EPI related results to the Rényi entropy power.

Chapter 2

Analytical Tools

This chapter includes notation and tools which are essential to the analysis in this thesis. It starts with a short introduction on the conditional Rényi entropy followed by some properties of the differential Rényi entropy and Rényi entropy power which are useful for the derivation of the improved R-EPIs. Next, basic inequalities which play a key role in this thesis are presented; this includes the sharpened Young's inequality, and a monotonicity property of the L_α norm. We finish this chapter with some results on rank-one modification of a symmetric eigenproblem [17].

2.1 Rényi Entropy

In the following, a definition of the conditional Rényi entropy and some properties of the differential Rényi entropy and Rényi entropy power defined in Definitions 1 and 2, respectively, are presented.

In contrast to the conditional (Shannon) entropy, many different definitions have been suggested for the conditional Rényi entropy in the literature (see [54]). For example, let X, Y be continuous random vectors taking values in \mathbb{R}^d , with a joint PDF $f_{X,Y}$. Recall the definition of the conditional differential entropy, $h(Y|X) = \int_{\mathbb{R}^d} f_X(x)h(Y|X = x)dx$, where f_X is the marginal PDF of the random vector X . By mimicking this definition, one can

define the conditional Rényi entropy as

$$h_\alpha(Y|X) = \int_{\mathbb{R}^d} f_X(x) h_\alpha(Y|X=x) dx, \quad \alpha \in [0, \infty]. \quad (2.1)$$

However, this definition fails to satisfy a basic property that holds for the conditional Shannon entropy, namely, that conditioning reduces entropy:

$$h_\alpha(Y|X) \leq h_\alpha(Y), \quad \forall \alpha \in [0, \infty]. \quad (2.2)$$

The monotonicity property in (2.2) will turn out to be useful in Chapter 3, hence a different definition for the conditional Rényi entropy is needed. The next definition is due to Arimoto [4], and satisfies the monotonicity property (together with a weak version of the chain rule [27, Section 4]).

Definition 3 (Conditional Differential Rényi Entropy). Let X, Y be random vectors taking values in \mathbb{R}^d . Assume that X and Y given X have probability density functions designated by f_X and $f_{Y|X}$, respectively. The conditional differential Rényi entropy of Y given X is defined as

$$h_\alpha(Y|X) = \frac{\alpha}{1-\alpha} \log \left(\int_{\mathbb{R}^d} f_X(x) \|f_{Y|X}(\cdot|x)\|_\alpha dx \right), \quad (2.3)$$

where

$$\|f_{Y|X}(\cdot|x)\|_\alpha = \left(\int_{\mathbb{R}^d} f_{Y|X}^\alpha(y|x) \right)^{\frac{1}{\alpha}}. \quad (2.4)$$

Next, properties of the differential Rényi entropy, the conditional Rényi entropy and the Rényi entropy power, which are relevant to this thesis, are presented:

- In view of (1.3) and (1.7), for $\alpha \in (0, 1) \cup (1, \infty)$,

$$N_\alpha(X) = (\|f_X\|_\alpha)^{-\frac{2\alpha'}{\alpha}}. \quad (2.5)$$

- The differential Rényi entropy $h_\alpha(X)$ is monotonically non-increasing in α , and so is $N_\alpha(X)$. i.e., every random vector X , if $0 \leq \beta \leq \alpha$, then

$$N_\beta(X) \geq N_\alpha(X). \quad (2.6)$$

- The conditional Rényi entropy defined in (2.3) satisfies the monotonicity property in (2.2).
- If $Y = AX + b$ where $A \in \mathbb{R}^{d \times d}$, $|A| \neq 0$, $b \in \mathbb{R}^d$, then for all $\alpha \in [0, \infty]$

$$h_\alpha(Y) = h_\alpha(X) + \log |A|, \quad (2.7)$$

$$N_\alpha(Y) = |A|^{\frac{2}{\alpha}} N_\alpha(X). \quad (2.8)$$

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

$$N_\alpha(\lambda X) = \lambda^2 N_\alpha(X), \quad \forall \lambda \in \mathbb{R}, \quad (2.9)$$

$$N_\alpha(X + b) = N_\alpha(X), \quad \forall b \in \mathbb{R}^d. \quad (2.10)$$

In view of (2.9) and (2.10), $N_\alpha(X)$ has some similar properties to the variance of X . However, if we consider a sum of independent random vectors, then $\text{Var}(\sum_{k=1}^n X_k) = \sum_{k=1}^n \text{Var}(X_k)$ 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).

2.2 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., $\frac{1}{\alpha} + \frac{1}{\alpha'} = 1$.

Note that $\alpha > 1$ if and only if $\alpha' > 0$; if $\alpha = 1$, we define $\alpha' = \infty$. This notation is known as Hölder's conjugate.

A useful consequence of (2.5) and (2.6) is the following result (a weaker version of it is given in [11, 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 $\beta \neq 1$,

$$\|f\|_\beta^{\beta'} \leq \|f\|_\alpha^{\alpha'}. \quad (2.11)$$

Proof. Let X be a random vector whose density function is f . In view of (2.5) and (2.6)

$$\begin{aligned} (\|f\|_\alpha)^{-\frac{2\alpha'}{d}} &= N_\alpha(X) \\ &\leq N_\beta(X) \\ &= (\|f\|_\beta)^{-\frac{2\beta'}{d}}. \end{aligned} \quad (2.12)$$

Increasing by the power of $-\frac{d}{2}$ yields the desired result. \square

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

$$A_t = t^{\frac{1}{t}} |t'|^{-\frac{1}{|t'|}} \quad (2.13)$$

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

$$A_{t'} = \frac{1}{A_t}. \quad (2.14)$$

The sharpened Young's inequality, first derived by Beckner [6] and re-derived with alternative proofs in, e.g., [5] and [16] is given as follows:

Fact 1 (Sharpened Young's inequality). Let $p, q, r > 0$ satisfy

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}, \quad (2.15)$$

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

$$\|f * g\|_r \leq \left(\frac{A_p A_q}{A_r} \right)^{\frac{d}{2}} \|f\|_p \|g\|_q. \quad (2.16)$$

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

$$\|f * g\|_r \geq \left(\frac{A_p A_q}{A_r} \right)^{\frac{d}{2}} \|f\|_p \|g\|_q. \quad (2.17)$$

Furthermore, (2.16) and (2.17) hold with equalities if and only if f and g are Gaussian probability density functions.

Note that the condition in (2.15) can be expressed in terms of the Hölder's conjugates as follows:

$$\frac{1}{p'} + \frac{1}{q'} = \frac{1}{r'}. \quad (2.18)$$

By using (2.18) 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

$$A = \left(\frac{1}{A_\nu} \prod_{k=1}^n A_{\nu_k} \right)^{\frac{d}{2}} \quad (2.19)$$

where the right side in (2.19) is defined by (2.13), and let $f_k \in L^{\nu_k}(\mathbb{R}^d)$ be non-negative functions.

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

$$\|f_1 * \dots * f_n\|_\nu \leq A \prod_{k=1}^n \|f_k\|_{\nu_k}. \quad (2.20)$$

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

$$\|f_1 * \dots * f_n\|_\nu \geq A \prod_{k=1}^n \|f_k\|_{\nu_k} \quad (2.21)$$

with equalities in (2.20) and (2.21) if and only if f_k are scaled versions of Gaussian probability densities for all k .

2.3 Rank-One Modification of a Symmetric Eigenproblem

This chapter is based on a paper by Bunch *et al.* [17] which addresses the eigenvectors and eigenvalues (a.k.a. eigensystem) of a rank-one modification of a real-valued diagonal matrix. We use in this thesis the following result [17]:

Fact 2. Let $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix with the eigenvalues $d_1 \leq d_2 \leq \dots \leq d_n$. Let $z \in \mathbb{R}^n$ such that $\|z\|_2 = 1$ and let $\rho \in \mathbb{R}$. Let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ 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 $i \in \{1, \dots, n\}$.
2. If $\rho > 0$, then the following interlacing property holds:

$$d_1 \leq \lambda_1 \leq d_2 \leq \lambda_2 \leq \dots \leq d_n \leq \lambda_n \quad (2.22)$$

and, if $\rho < 0$, then

$$\lambda_1 \leq d_1 \leq \lambda_2 \leq d_2 \leq \dots \leq \lambda_n \leq d_n. \quad (2.23)$$

3. If all the eigenvalues of D are different, all the entries of z are non-zero, and $\rho \neq 0$, then inequalities (2.22) and (2.23) are strict. For $i \in \{1, \dots, n\}$, the eigenvalue λ_i is a zero of

$$W(x) = 1 + \rho \sum_{j=1}^n \frac{z_j^2}{d_j - x}. \quad (2.24)$$

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

Chapter 3

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 [11, 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:

$$N_\alpha(S_n) \geq c_\alpha^{(n)} \sum_{k=1}^n N_\alpha(X_k) \quad (3.1)$$

with

$$c_\alpha^{(n)} = \alpha^{\frac{1}{\alpha-1}} \left(1 - \frac{1}{n\alpha'}\right)^{n\alpha'-1}. \quad (3.2)$$

Furthermore, the R-EPI in (3.1) has the following properties:

1. Eq. (3.1) improves the R-EPI in [11, 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 [11, Theorem I.1] as $n \rightarrow \infty$,
3. In the other limiting case where $\alpha \downarrow 1$, it coincides with the EPI (similarly to [11]),

4. If $n = 2$ and $\alpha \rightarrow \infty$, the constant $c_\alpha^{(n)}$ in (3.2) 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

$$N_\alpha(X_k) > 0, \quad k \in \{1, \dots, n\} \quad (3.3)$$

which, in view of (2.5), implies that $f_{X_k} \in L^\alpha(\mathbb{R}^d)$, where f_{X_k} is the density of X_k for all $k \in \{1, \dots, n\}$. In [11, (12)] it is shown that for $\alpha > 1$,

$$N_\alpha(S_n) \geq B \prod_{k=1}^n N_\alpha^{t_k}(X_k) \quad (3.4)$$

with

$$B = (A_{\nu_1} \dots A_{\nu_n} A_{\alpha'})^{-\alpha'}, \quad (3.5)$$

$$\nu_k > 1, \quad \forall k \in \{1, \dots, n\}, \quad (3.6)$$

$$\nu' = \frac{\nu}{\nu - 1}, \quad \forall \nu \in \mathbb{R}, \quad (3.7)$$

$$\sum_{k=1}^n \frac{1}{\nu'_k} = \frac{1}{\alpha'}, \quad (3.8)$$

$$t_k = \frac{\alpha'}{\nu'_k}, \quad \forall k \in \{1, \dots, n\}. \quad (3.9)$$

Consequently, (3.6)–(3.9) yields

$$t_k > 0, \quad \forall k \in \{1, \dots, n\}, \quad (3.10)$$

$$\sum_{k=1}^n t_k = 1. \quad (3.11)$$

The proof of (3.4), which relies on Corollaries 1 and 2, is introduced in Appendix A.

Similarly to [11, (14)], in view of the homogeneity of the entropy power functional (see (2.9)), it can be assumed without any loss of generality that

$$\sum_{k=1}^n N_\alpha(X_k) = 1. \quad (3.12)$$

Hence, to prove (3.1), it is sufficient to show that under the assumption in (3.12)

$$N_\alpha(S_n) \geq c_\alpha^{(n)}. \quad (3.13)$$

From this point, we deviate from the proof of [11, Theorem I.1]. Taking logarithms on both sides of (3.4) and assembling (2.13), (3.5)–(3.11) and (3.12) yield

$$\log N_\alpha(S_n) \geq f_0(\underline{t}), \quad (3.14)$$

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

$$f_0(\underline{t}) = \frac{\log \alpha}{\alpha - 1} - D(\underline{t} \parallel \underline{N}_\alpha) + \alpha' \sum_{k=1}^n \left(1 - \frac{t_k}{\alpha'}\right) \log \left(1 - \frac{t_k}{\alpha'}\right), \quad (3.15)$$

$$\underline{N}_\alpha = (N_\alpha(X_1), \dots, N_\alpha(X_n)), \quad (3.16)$$

$$D(\underline{t} \parallel \underline{N}_\alpha) = \sum_{k=1}^n t_k \log \left(\frac{t_k}{N_\alpha(X_k)} \right). \quad (3.17)$$

In view of (3.10) and (3.11), the bound in (3.14) holds for every $\underline{t} \in \mathbb{R}_+^n$ such that $\sum_{k=1}^n t_k = 1$. Consequently, the R-EPI in [11, Theorem I.1] can be tightened by maximizing the right side of (3.14), leading to the following optimization problem:

$$\begin{aligned} & \text{maximize} && f_0(\underline{t}) \\ & \text{subject to} && t_k \geq 0, \quad k \in \{1, \dots, n\}, \\ & && \sum_{k=1}^n t_k = 1. \end{aligned} \quad (3.18)$$

Note that the convexity of the function

$$f(x) = \left(1 - \frac{x}{\alpha'}\right) \log \left(1 - \frac{x}{\alpha'}\right), \quad x \in [0, \alpha'] \quad (3.19)$$

yields that the third term on the right side of (3.15) is convex in \underline{t} . Since the relative entropy $D(\underline{t} \parallel \underline{N}_\alpha)$ is also convex in \underline{t} , the objective function f_0 in (3.15) 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 (3.18), we take the sub-optimal choice $\underline{t} = \underline{N}_\alpha$

(similarly to the proof [11, 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}_α . Let

$$\hat{t}_k = N_\alpha(X_k), \quad 1 \leq k \leq n, \quad (3.20)$$

then, in view of (3.14) and (3.20),

$$\log N_\alpha(S_n) \geq f_0(\hat{\underline{t}}) \quad (3.21)$$

$$= \frac{\log \alpha}{\alpha - 1} + \alpha' \sum_{k=1}^n \left(1 - \frac{\hat{t}_k}{\alpha'}\right) \log \left(1 - \frac{\hat{t}_k}{\alpha'}\right). \quad (3.22)$$

Due to the convexity of f in (3.19), for all $k \in \{1, \dots, n\}$,

$$f(\hat{t}_k) \geq f(x) + f'(x)(\hat{t}_k - x). \quad (3.23)$$

Choosing $x = \frac{1}{n}$ in the right side of (3.23) yields

$$\left(1 - \frac{\hat{t}_k}{\alpha'}\right) \log \left(1 - \frac{\hat{t}_k}{\alpha'}\right) \geq \log \left(1 - \frac{1}{n\alpha'}\right) + \frac{\log e}{n\alpha'} - \frac{\hat{t}_k}{\alpha'} \left[\log e + \log \left(1 - \frac{1}{n\alpha'}\right)\right] \quad (3.24)$$

and, in view of (3.12) and (3.20) which yields $\sum_{k=1}^n \hat{t}_k = 1$, summing over $k \in \{1, \dots, n\}$ on both sides of (3.24) implies that

$$\alpha' \sum_{k=1}^n \left(1 - \frac{\hat{t}_k}{\alpha'}\right) \log \left(1 - \frac{\hat{t}_k}{\alpha'}\right) \geq (n\alpha' - 1) \log \left(1 - \frac{1}{n\alpha'}\right). \quad (3.25)$$

Finally, assembling (3.21), (3.22) and (3.25) yields (3.13) with $c_\alpha^{(n)}$ in (3.2) as required.

In the sequel, we no longer assume that condition (3.3) holds. Define

$$\mathcal{K}_0 = \{k \in \{1, \dots, n\} : N_\alpha(X_k) = 0\}, \quad (3.26)$$

and note that

$$h_\alpha(S_n) = h_\alpha \left(\sum_{k \notin \mathcal{K}_0} X_k + \sum_{k \in \mathcal{K}_0} X_k \right) \quad (3.27)$$

$$\geq h_\alpha \left(\sum_{k \notin \mathcal{K}_0} X_k + \sum_{k \in \mathcal{K}_0} X_k \mid \{X_k\}_{k \in \mathcal{K}_0} \right) \quad (3.28)$$

$$= h_\alpha \left(\sum_{k \notin \mathcal{K}_0} X_k \right) \quad (3.29)$$

where the conditional Rényi entropy is defined according to Arimoto's proposal in [4] (see also [27, Section 4]), (3.28) is due to the monotonicity property of the conditional Rényi entropy (see [27, Theorem 2]), and (3.29) 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

$$N_\alpha \left(\sum_{k \notin \mathcal{K}_0} X_k \right) \geq c_\alpha^{(l)} \sum_{k \notin \mathcal{K}_0} N_\alpha(X_k), \quad (3.30)$$

where $l = n - |\mathcal{K}_0|$. In view of (3.2), it can be verified that $c_\alpha^{(n)}$ is monotonically decreasing in n ; hence, (3.29), (3.30) and $c_\alpha^{(l)} \geq c_\alpha^{(n)}$ yield

$$N_\alpha(S_n) \geq c_\alpha^{(n)} \sum_{k=1}^n N_\alpha(X_k). \quad (3.31)$$

We now turn to prove Items 1)–4).

- To prove Item 1), note that (1.9) and (3.2) yield that $c_\alpha^{(n)} > c_\alpha$ for all $\alpha > 1$ and $n \in \mathbb{N}$.
- Item 2) holds since from (3.2)

$$\lim_{n \rightarrow \infty} c_\alpha^{(n)} = \frac{1}{e} \alpha^{\frac{1}{\alpha-1}} \quad (3.32)$$

where the right side of (3.32) coincides with the constant c_α in [11, (3)] (see (1.9)).

- Item 3) holds since $\alpha \downarrow 1$ yields $\alpha' \rightarrow \infty$, which implies that for every $n \in \mathbb{N}$

$$\lim_{\alpha \downarrow 1} c_\alpha^{(n)} = \lim_{\alpha \downarrow 1} c_\alpha = 1. \quad (3.33)$$

Hence, by letting α tend to 1, (3.1) and (3.33) yield the EPI in (1.1).

- To prove Item 4), note that from (3.2)

$$\lim_{\alpha \rightarrow \infty} c_\alpha^{(n)} = \left(1 - \frac{1}{n}\right)^{n-1} \quad (3.34)$$

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

$$M(X) := \text{ess sup}(f_X). \quad (3.35)$$

From (1.6), (1.7) and (3.35), it follows that

$$N_\infty(X) := \lim_{\alpha \rightarrow \infty} N_\alpha(X) \quad (3.36)$$

$$= M^{-\frac{2}{d}}(X). \quad (3.37)$$

By assembling (3.1) and (3.37), it follows that if X_1, \dots, X_n are d -dimensional independent random vectors with densities, then

$$M^{-\frac{2}{d}}(S_n) \geq \left(1 - \frac{1}{n}\right)^{n-1} \sum_{k=1}^n M^{-\frac{2}{d}}(X_k). \quad (3.38)$$

This tightens the inequality in [10, Theorem 1] where the coefficient $(1 - \frac{1}{n})^{n-1}$ on the right side of (3.38) has been loosened to $\frac{1}{e}$ (note, however, that they coincide when $n \rightarrow \infty$). For $n = 2$, the coefficient $\frac{1}{2}$ on the right side of (3.38) is tight, and it is achieved when X_1 and X_2 are independent random vectors which are uniformly distributed in the cube $[0, 1]^d$ [10, p. 103]. \square

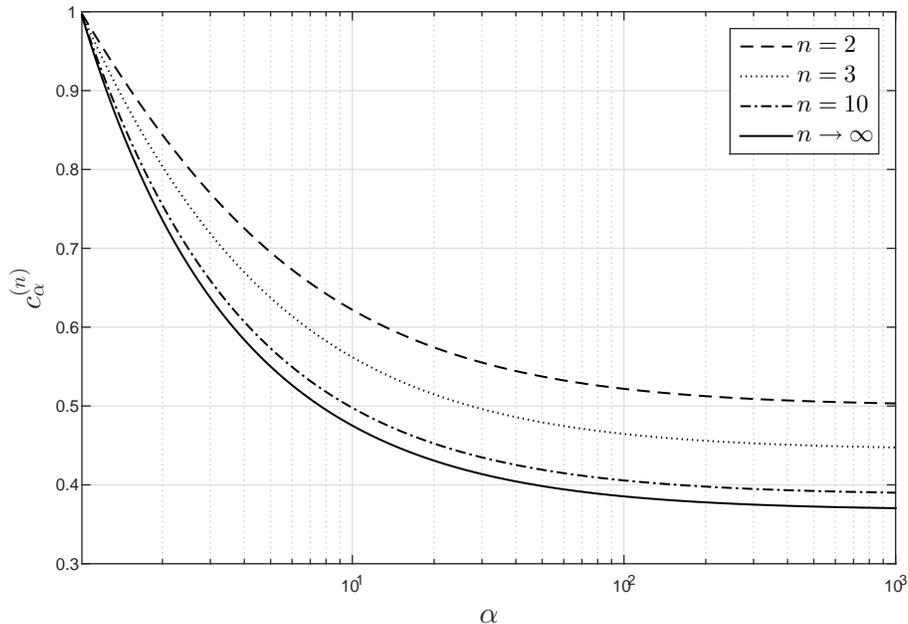


Figure 3.1: A plot of $c_\alpha^{(n)}$ in (3.2), as a function of α , for $n = 2, 3, 10$ and $n \rightarrow \infty$.

Figure 3.1 plots $c_\alpha^{(n)}$ as a function of α , for some values of n , verifying numerically Items 1)–4) in Theorem 1. In [11, Theorem I.1], $c_\alpha^{(n)}$ is independent of n , and it is equal to c_α in (1.8) which is the limit of $c_\alpha^{(n)}$ in (3.2) by letting $n \rightarrow \infty$ (the solid curve in Figure 3.1).

Remark 1. For independent random variables $\{X_k\}_{k=1}^n$ with densities on \mathbb{R} , the result in (3.38) with $d = 1$ can be strengthened to (see [10, p. 105] and [46])

$$\frac{1}{M^2(S_n)} \geq \frac{1}{2} \sum_{k=1}^n \frac{1}{M^2(X_k)} \quad (3.39)$$

where $S_n := \sum_{k=1}^n X_k$. Note that (3.38) and (3.39) 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

$$\begin{aligned}\Pr[X_{1,k} = X_{2,k}] &= \alpha, \\ \Pr[Y_{1,k} = Y_{2,k}] &= \beta\end{aligned}\tag{3.40}$$

for all $k \in \{1, \dots, n\}$. We wish to obtain an upper bound on the probability that $X_1 + Y_1$ and $X_2 + Y_2$ are equal. From (1.3), (1.7) (with $\alpha = 2$), and (3.40)

$$N_2(X) = \exp\left(\frac{2}{d} h_2(X)\right)\tag{3.41}$$

$$= \left(\int_{\mathbb{R}^d} f_X^2(\underline{x}) \, d\underline{x}\right)^{-\frac{2}{d}}\tag{3.42}$$

$$= \mathbb{P}^{-\frac{2}{d}}[X_1 = X_2]\tag{3.43}$$

$$= \prod_{k=1}^d \mathbb{P}^{-\frac{2}{d}}[X_{1,k} = X_{2,k}]\tag{3.44}$$

$$= \alpha^{-2},\tag{3.45}$$

$$N_2(Y) = \beta^{-2},\tag{3.46}$$

$$N_2(X + Y) = \mathbb{P}^{-\frac{2}{d}}[X_1 + Y_1 = X_2 + Y_2].\tag{3.47}$$

Assembling (3.1) with $n = \alpha = 2$, (3.45), (3.46) and (3.47) yield

$$\mathbb{P}(X_1 + Y_1 = X_2 + Y_2) \leq \left(\frac{27}{32} (\alpha^{-2} + \beta^{-2})\right)^{-\frac{d}{2}}.\tag{3.48}$$

The factor $\frac{27}{32}$ on the base of the exponent on the right side of (3.48), instead of the looser factor $c_2 = \frac{2}{e}$ which follows from (1.9) with $\alpha = 2$ (see [11, Theorem I.1]), improves the exponential decay rate of the upper bound in (3.48) as a function of the dimension d . The optimal bound has to be with a coefficient of $(\alpha^{-2} + \beta^{-2})$ on the base of the exponent in the right side of (3.48) which is less than or equal to 1; this can be verified since if X and Y are independent Gaussian random variables, then

$$N_2(X + Y) = N_2(X) + N_2(Y),\tag{3.49}$$

so,

$$\mathbb{P}(X_1 + Y_1 = X_2 + Y_2) = (\alpha^{-2} + \beta^{-2})^{-\frac{d}{2}}. \quad (3.50)$$

This provides a reference for comparing the exponential decay which is implied by c_2 in (1.9), $c_2^{(2)}$ in (3.1), and the case where X and Y are independent Gaussian random variables:

$$\frac{2}{e} < \frac{27}{32} < 1. \quad (3.51)$$

Chapter 4

A Further tightening of the R-EPI

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

In the following, we wish to tighten the R-EPI in Theorem 1 for an arbitrary $n \geq 2$. It is first demonstrated that a reduction of the optimization problem in (3.18) 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 2 in Section 2.3 (see [17]), and it is also shown that the reduction of the optimization problem in (3.18) 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 is exemplified numerically in comparison to the bounds in [7], [11] and Theorem 1.

4.1.1 The optimization problem in (3.18)

In view of (3.15)–(3.18), the maximization problem in (3.18) can be expressed in the form

$$\begin{aligned} & \text{maximize} && f_0(\underline{t}) = \sum_{k=1}^n g(t_k) + \sum_{k=1}^n t_k \log N_k + \frac{\log \alpha}{\alpha-1} \\ & \text{subject to} && \underline{t} \in \mathcal{P}^n \end{aligned} \tag{4.1}$$

where

$$g(x) = (\alpha' - x) \log \left(1 - \frac{x}{\alpha'} \right) - x \log x, \quad x \in [0, 1] \tag{4.2}$$

$$N_k = N_\alpha(X_k), \quad k \in \{1, \dots, n\} \tag{4.3}$$

(for simplicity of notation, the dependence of g and N_k in α has been suppressed in (4.1)), and \mathcal{P}^n is the probability simplex

$$\mathcal{P}^n = \left\{ \underline{t} \in \mathbb{R}^n : t_k \geq 0, \sum_{k=1}^n t_k = 1 \right\}. \tag{4.4}$$

The term $\sum_{k=1}^n t_k \log N_k$ on the right side of (4.1) 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, \dots, t_n , the objective function f_0 in (4.1) is not concave if $\alpha > 2$.

4.1.2 A reduction of the optimization problem in (3.18) to $n - 1$ variables

In view of (4.4), the substitution

$$t_n = 1 - \sum_{k=1}^{n-1} t_k \tag{4.5}$$

transforms the maximization problem in (4.1) to the following equivalent problem:

$$\begin{aligned} & \text{maximize} && f(t_1, \dots, t_{n-1}) \\ & \text{subject to} && \underline{t} \in \mathcal{D}^{n-1} \end{aligned} \tag{4.6}$$

where

$$f(t_1, \dots, t_{n-1}) = f_0 \left(t_1, \dots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k \right) \tag{4.7}$$

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

$$\mathcal{D}^{n-1} = \left\{ (t_1, \dots, t_{n-1}) : t_k \geq 0, \sum_{k=1}^{n-1} t_k \leq 1 \right\}. \tag{4.8}$$

4.1.3 Proving the convexity of the optimization problem in (4.6)

We wish to show that the objective function f of the optimization problem in (4.6) is concave, i.e., it is required to assert that all the eigenvalues of the Hessian matrix $\nabla^2 f$ are non-positive.

Eqs. (4.1) and (4.7) yield

$$\begin{aligned} & f(t_1, \dots, t_{n-1}) \\ &= \sum_{k=1}^{n-1} g(t_k) + g \left(1 - \sum_{k=1}^{n-1} t_k \right) \\ &+ \sum_{k=1}^{n-1} t_k \log N_k + \left(1 - \sum_{k=1}^{n-1} t_k \right) \log N_n + \frac{\log \alpha}{\alpha - 1}. \end{aligned} \tag{4.9}$$

Let

$$q(x) = g''(x) = \frac{2x - \alpha'}{x(\alpha' - x)}, \quad x \in [0, 1] \tag{4.10}$$

then, in view of (4.9) and (4.10), for all $(t_1, \dots, t_{n-1}) \in \mathcal{D}^{n-1}$

$$\begin{aligned} \nabla^2 f(t_1, \dots, t_{n-1}) &= \begin{pmatrix} q(t_1) & 0 & \cdots & 0 \\ 0 & q(t_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q(t_{n-1}) \end{pmatrix} + q \left(1 - \sum_{k=1}^{n-1} t_k \right) \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \\ &= D + \rho \underline{\mathbf{1}} \underline{\mathbf{1}}^T \end{aligned} \quad (4.11)$$

where

$$\begin{aligned} D &= \text{diag}(q(t_1), \dots, q(t_{n-1})), \\ \rho &= q \left(1 - \sum_{k=1}^{n-1} t_k \right). \end{aligned} \quad (4.12)$$

Recall that if $\alpha' \in [2, \infty)$ then $f_0(t_1, \dots, t_n)$ is concave in \mathcal{P}^n , hence, so is $f(t_1, \dots, t_{n-1})$ in \mathcal{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 $f: \mathcal{D}^{n-1} \rightarrow \mathbb{R}$ in (4.9) is concave.

Proof. See Appendix B. □

4.1.4 Solution of the convex optimization problem in (4.6)

In the following, we solve the convex optimization problem in (4.6) via the Lagrange duality and KKT conditions (see, e.g., [15, Section 5]). Since the problem is invariant to permutations of the entries of $X = (X_1, \dots, X_n)$, it can be assumed without any loss of generality that the last term of the vector \underline{N}_α in (3.16) is maximal, i.e.,

$$N_\alpha(X_k) \leq N_\alpha(X_n), \quad k \in \{1, \dots, n-1\}. \quad (4.13)$$

Moreover, it is assumed that

$$N_\alpha(X_n) > 0. \quad (4.14)$$

The possibility that $N_\alpha(X_n) = 0$ leads to a trivial bound since from (4.13), it follows that $N_\alpha(X_k) = 0$ for every $k \in \{1, \dots, n\}$; this makes the right side of (1.8) be equal to zero, while its left side is always non-negative. Let

$$c_k = \frac{N_\alpha(X_k)}{N_\alpha(X_n)}, \quad k \in \{1, \dots, n-1\}. \quad (4.15)$$

From (4.13)–(4.15), the sequence $\{c_k\}_{k=1}^{n-1}$ satisfies

$$0 \leq c_k \leq 1, \quad k \in \{1, \dots, n-1\}. \quad (4.16)$$

Let t_n be defined as in (4.5). Appendix C provides the technical details which are related to the solution of the convex optimization problem in (4.6) 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

$$t_k(\alpha' - t_k) = c_k t_n(\alpha' - t_n), \quad k \in \{1, \dots, n-1\} \quad (4.17)$$

$$\sum_{k=1}^n t_k = 1 \quad (4.18)$$

$$t_k \geq 0, \quad k \in \{1, \dots, n\} \quad (4.19)$$

with the variables \underline{t} in (4.17)–(4.19).

Note that if $N_\alpha(X_k)$ is independent of k then, from (4.15), $c_k = 1$ for all $k \in \{1, \dots, n-1\}$. Hence, from (4.17) and (4.18), it follows that $t_1 = \dots = t_n = \frac{1}{n}$ (note that the other possibility where $t_k = \alpha' - t_n$ for some $k \in \{1, \dots, n-1\}$ contradicts (4.18) and (4.19) since in this case $\sum_{j=1}^n t_j \geq t_k + t_n = \alpha' > 1$). 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}_α 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, (4.17) yields a quadratic equation for t_k whose solutions are given by

$$t_k = \frac{1}{2} \left(\alpha' \pm \sqrt{\alpha'^2 - 4c_k t_n(\alpha' - t_n)} \right) \quad (4.20)$$

with $\alpha' = \frac{\alpha}{\alpha-1}$. The possibility of the positive sign in the right side of (4.20) is rejected since in this case $t_n + t_k \geq \alpha' > 1$, which violates (4.18). Hence, from (4.20), for all $k \in \{1, \dots, n-1\}$

$$t_k = \psi_{k,\alpha}(t_n) \quad (4.21)$$

where we define

$$\psi_{k,\alpha}(x) = \frac{1}{2} \left(\alpha' - \sqrt{\alpha'^2 - 4c_k x(\alpha' - x)} \right), \quad x \in [0, 1]. \quad (4.22)$$

In view of (4.18) and (4.21), one first calculates $t_n \in [0, 1]$ by numerically solving the equation

$$t_n + \sum_{k=1}^{n-1} \psi_{k,\alpha}(t_n) = 1. \quad (4.23)$$

The existence and uniqueness of a solution of (4.23) is proved in Appendix D. Once we compute t_n , all t_k 's for $k \in \{1, \dots, n-1\}$ are computed from (4.21). Finally, the substitution of t_1, \dots, t_n in the right side of (3.14) enables to calculate the improved R-EPI in (3.14), i.e.,

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq \exp(f_0(t_1, \dots, t_n)) \sum_{k=1}^n N_\alpha(X_k) \quad (4.24)$$

with f_0 in (3.15).

Note that due to the optimal selection of the vector $\underline{t} = (t_1, \dots, t_n)$ in (4.24), the R-EPI in this chapter 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, \dots, X_n be independent random vectors with probability densities defined on \mathbb{R}^d , let $N_\alpha(X_1), \dots, N_\alpha(X_n)$ be their respective Rényi entropy powers of order $\alpha > 1$, and let $\alpha' = \frac{\alpha}{\alpha-1}$. Let the indices of X_1, \dots, 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 (4.15);
2. $t_n \in [0, 1]$ be the unique solution of (4.23);

3. $\{t_k\}_{k=1}^{n-1}$ be given in (4.21) and (4.22).

Then, the R-EPI in (4.24) holds with f_0 in (3.15), 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 [11, Theorem 1]); if $N_\alpha(X_k)$ is independent of k , then the two R-EPIs in Theorem 1 and (4.24) coincide.
2. It improves the Bercher-Vignat (BV) bound in [7] which states that

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq \max\{N_\alpha(X_1), \dots, N_\alpha(X_n)\} \quad (4.25)$$

and the bounds in (4.24) and (4.25) asymptotically coincide as $\alpha \rightarrow \infty$ if and only if

$$\sum_{k=1}^{n-1} N_\infty(X_k) \leq N_\infty(X_n) \quad (4.26)$$

where $N_\infty(X)$ is defined in (3.37).

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 [11, Theorem 1] and Theorem 1 as $\alpha \downarrow 1$.

Proof. The proof of the R-EPI in (4.24) is provided earlier in this chapter with some additional details in Appendices B–E. In view of the 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 (3.20), whereas it is set to be optimal in the proof of Theorem 2 in (4.21)–(4.23). 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 (3.20)) reduces to $\underline{t} = (\frac{1}{n}, \dots, \frac{1}{n})$, which turns to be optimal in the sense of achieving the maximum of the objective function in (4.9).

- Item 2) holds since the selection of \underline{t} in the right side of (3.14) with $t_k = 1$ and $t_i = 0$ for all $i \neq k$ yields

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq N_\alpha(X_k) \quad (4.27)$$

which then leads to (4.25) by a maximization of the right side of (4.27) over $k \in \{1, \dots, n\}$. Appendix E proves that the bounds in (4.24) and (4.25) asymptotically coincide as $\alpha \rightarrow \infty$ if and only if the condition in (4.26) holds.

- Item 3) is proved in Section 4.2.
- Item 4) holds since the R-EPI obtained in Theorem 2 is at least as tight as the BC bound in [11, Theorem 1]; the latter coincides with the EPI as we let α tend to 1 (recall that, from (1.9), $\lim_{\alpha \downarrow 1} c_\alpha = 1$) which is known to be tight for Gaussian random vectors with proportional covariances.

□

Figure 4.1 compares the two R-EPIs in Theorems 1 and 2 with those in [11] (see (1.9)) and [7] (see (4.25)) 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 4.1 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 [7] and [11]; 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 4.1 and supported by Item 2) of Theorem 2, the bound in this

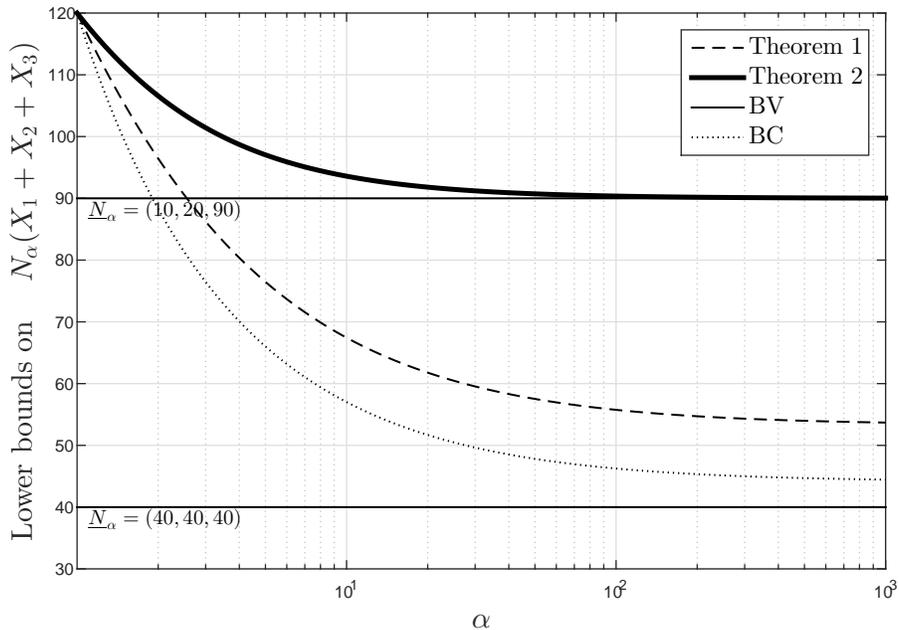


Figure 4.1: A comparison of the Rényi entropy power inequalities for $n = 3$ independent random vectors according to [11] (BC), [7] (BV), Theorem 1 and the tightest bound in Theorem 2. The bounds refer to the two cases where $(N_\alpha(X_1), N_\alpha(X_2), N_\alpha(X_3)) = (40, 40, 40)$ or $(10, 20, 90)$ (in both cases, the sum of the entries is 120; in the former case, the condition in (4.26) does not hold, while in the latter it does).

theorem asymptotically coincides with the BV bound (by letting $\alpha \rightarrow \infty$) 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 4.1 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 4.1, the R-EPIs of Theorems 1 and 2, as well as [11, Theorem 1], coincide with the EPI as we let α tend to 1 (from above).

4.2 A 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

$$d(x\|y) = x \log \left(\frac{x}{y} \right) + (1-x) \log \left(\frac{1-x}{1-y} \right). \quad (4.28)$$

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

$$\alpha' = \frac{\alpha}{\alpha - 1}, \quad (4.29)$$

$$\beta_\alpha = \frac{N_\alpha(X_1)}{N_\alpha(X_2)}, \quad (4.30)$$

$$t_\alpha = \begin{cases} \frac{\alpha'(\beta_\alpha+1)-2\beta_\alpha-\sqrt{(\alpha'(\beta_\alpha+1))^2-8\alpha'\beta_\alpha+4\beta_\alpha}}{2(1-\beta_\alpha)} & \text{if } \beta_\alpha < 1, \\ \frac{1}{2} & \text{if } \beta_\alpha = 1. \end{cases} \quad (4.31)$$

Then, the following R-EPI holds:

$$N_\alpha(X_1 + X_2) \geq c_\alpha (N_\alpha(X_1) + N_\alpha(X_2)) \quad (4.32)$$

with

$$c_\alpha = \alpha^{\frac{1}{\alpha-1}} \exp \left(-d \left(t_\alpha \parallel \frac{\beta_\alpha}{\beta_\alpha + 1} \right) \right) \left(1 - \frac{t_\alpha}{\alpha'} \right)^{\alpha' - t_\alpha} \left(1 - \frac{1 - t_\alpha}{\alpha'} \right)^{\alpha' - 1 + t_\alpha}. \quad (4.33)$$

The R-EPI in (4.32) satisfies Items 1)–4) of Theorem 2; specifically, by letting $\alpha \rightarrow \infty$, the lower bound on $N_\alpha(X_1 + X_2)$ tends to $N_\infty(X_2)$, which asymptotically coincides with the BV bound in [7].

Proof. Due the constraints in (4.1), 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 (3.12), then

$$\underline{N}_\alpha = \left(\frac{\beta_\alpha}{1+\beta_\alpha}, \frac{1}{1+\beta_\alpha} \right) \quad (4.34)$$

and, by (3.15), the maximization in (4.1) is transformed to

$$\begin{aligned} \underset{t \in [0,1]}{\text{maximize}} \left\{ \frac{\log \alpha}{\alpha - 1} - t \log((1 + \beta_\alpha)t) - (1 - t) \log\left(\frac{(1 + \beta_\alpha)(1 - t)}{\beta_\alpha}\right) \right. \\ \left. + \alpha' \left[\left(1 - \frac{t}{\alpha'}\right) \log\left(1 - \frac{t}{\alpha'}\right) + \left(1 - \frac{1-t}{\alpha'}\right) \log\left(1 - \frac{1-t}{\alpha'}\right) \right] \right\}. \end{aligned} \quad (4.35)$$

It can be verified that the objective function in (4.35) 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

$$\log\left(\frac{(1-t)\beta_\alpha}{t}\right) - \log\left(\frac{\alpha'-t}{\alpha'-1+t}\right) = 0. \quad (4.36)$$

Eq. (4.36) can be expressed as a quadratic equation whose solution is given in (4.31). Substituting the optimized value $t = t_\alpha$ in (4.31) into the objective function on the right side of (4.35) leads to the closed-form solution of the optimization problem in (4.1) for $n = 2$. Hence, under the assumption in (3.12) where $N_\alpha(X_1) + N_\alpha(X_2) = 1$, straightforward algebra yields that

$$N_\alpha(X_1 + X_2) \geq c_\alpha \quad (4.37)$$

where c_α is given in (4.33); the relaxation of this assumption requires the multiplication of the right side of (4.37) by $N_\alpha(X_1) + N_\alpha(X_2)$ (due to the homogeneity of the Rényi entropy power, see (2.9)). Note that, for $n = 2$, the condition in (4.26) becomes vacuous (since, by assumption, $N_\infty(X_1) \leq N_\infty(X_2)$) which implies that the bound in (4.32) asymptotically coincides with the BV bound when $\alpha \rightarrow \infty$. \square

Figure 4.2 compares the two R-EPIs in Theorem 1 and Corollary 3 with those in [11] (see (1.9)) and [7] (see (4.25)) for $n = 2$ independent random vectors. The four bounds are plotted in Figure 4.1 for symmetric and asymmetric cases where $(N_\alpha(X_1), N_\alpha(X_2)) = (10, 10)$ and $(6, 14)$, respectively (note that in both cases, the sum of the entries is equal to 20). Along

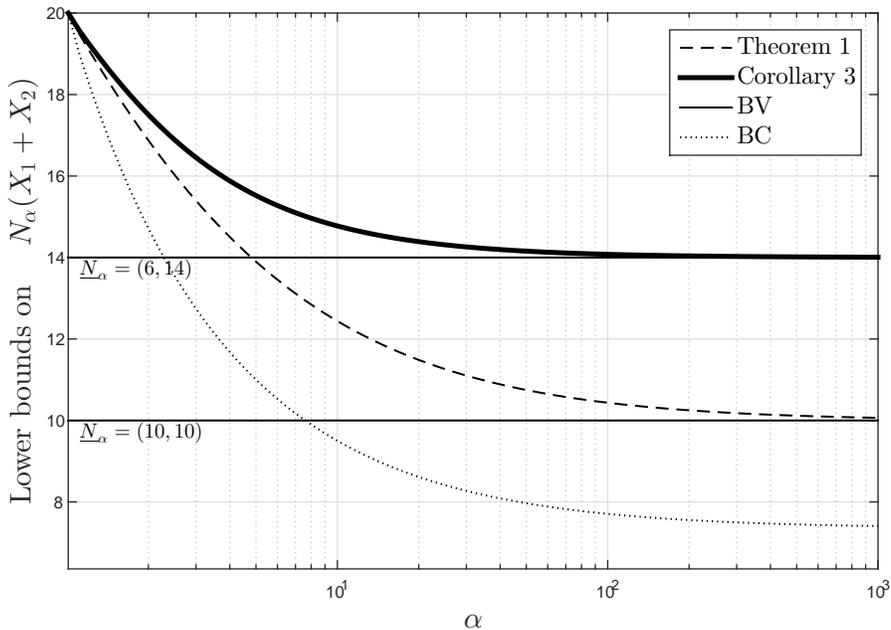


Figure 4.2: A comparison of the Rényi entropy power inequalities for $n = 2$ independent random vectors according to [11] (BC), [7] (BV), Theorem 1 and the Corollary 3. The bounds refer to the two cases where $(N_\alpha(X_1), N_\alpha(X_2)) = (10, 10)$ or $(6, 14)$ (in both cases, the sum of the entries is 20).

with the properties of the bounds as described for the general $n \geq 2$ case (and exemplified for $n = 3$ in Figure 4.1), the bound in Corollary 3 (valid for $n = 2$) asymptotically coincides with the BV bound (by letting $\alpha \rightarrow \infty$) in both the symmetric and asymmetric cases, as it is shown in Figure 4.2; indeed, the condition in (4.26) holds for $n = 2$ and for every value of $(N_\alpha(X_1), N_\alpha(X_2))$.

4.3 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

$$\underline{Y}(n) = \sum_{k=0}^{L-1} \mathbf{H}_k \underline{X}(n-k) \quad (4.38)$$

be the filtered data at the output of a finite impulse response (FIR) filter where $\mathbf{H}_0, \dots, \mathbf{H}_{L-1}$ 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 $h_\alpha(\underline{Y}(n)) - h_\alpha(\underline{X}(n))$, 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 $k \in \{0, \dots, L-1\}$ and every n , we have

$$h_\alpha(\mathbf{H}_k \underline{X}(n-k)) = h_\alpha(\underline{X}(n)) + \log|\det(\mathbf{H}_k)| \quad (4.39)$$

and

$$\begin{aligned} N_\alpha(\mathbf{H}_k \underline{X}(n-k)) &= \exp\left(\frac{2}{d} h_\alpha(\mathbf{H}_k \underline{X}(n-k))\right) \\ &= |\det(\mathbf{H}_k)|^{\frac{2}{d}} N_\alpha(\underline{X}(n)). \end{aligned} \quad (4.40)$$

Let $\alpha > 1$, and $\alpha' = \frac{\alpha}{\alpha-1}$. Similarly to Theorem 2, it is assumed without loss of generality that $|\det(\mathbf{H}_k)| \leq |\det(\mathbf{H}_{L-1})|$ for all $k \in \{0, \dots, L-2\}$; otherwise, the indices of $\mathbf{H}_0, \dots, \mathbf{H}_{L-1}$ 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 (4.15) and (4.40), for every $k \in \{0, \dots, L-2\}$,

$$c_k = \left(\frac{|\det(\mathbf{H}_k)|}{|\det(\mathbf{H}_{L-1})|} \right)^{\frac{2}{d}} \quad (4.41)$$

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 (4.23) (with a replacement of 1 and n in (4.23) by 0 and $L-1$, respectively), and then the rest of the t_k 's for $k \in \{0, \dots, L-2\}$ are being calculated via (4.21) and (4.22). In view of (4.39), (4.40), and the R-EPI of Theorem 2, it follows that for every n

$$\begin{aligned} & h_\alpha(\underline{Y}(n)) - h_\alpha(\underline{X}(n)) \\ & \geq \frac{d}{2} \left(\frac{\log \alpha}{\alpha - 1} + \sum_{k=0}^{L-1} g(t_k) \right) + \sum_{k=0}^{L-1} t_k \log |\det(\mathbf{H}_k)| \end{aligned} \quad (4.42)$$

where the function g is given in (4.2).

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:

$$\begin{aligned} & h_\alpha(\underline{Y}(n)) - h_\alpha(\underline{X}(n)) \\ & \geq \frac{d}{2} \cdot \log \left(\sum_{k=0}^{L-1} |\det(\mathbf{H}_k)|^{\frac{2}{d}} \right) \\ & \quad + \frac{d}{2} \left(\frac{\log \alpha}{\alpha - 1} + \left(\frac{L\alpha}{\alpha - 1} - 1 \right) \log \left(1 - \frac{\alpha - 1}{L\alpha} \right) \right). \end{aligned} \quad (4.43)$$

The R-EPI of [11, Theorem I.1] leads to the following loosened bound in comparison to (4.43):

$$h_\alpha(\underline{Y}(n)) - h_\alpha(\underline{X}(n)) \geq \frac{d}{2} \left[\log \left(\sum_{k=0}^{L-1} |\det(\mathbf{H}_k)|^{\frac{2}{d}} \right) + \frac{\log \alpha}{\alpha - 1} - \log e \right] \quad (4.44)$$

and, finally, the BV bound in [7] (see (4.25)) leads to the following loosening of (4.42):

$$h_\alpha(\underline{Y}(n)) - h_\alpha(\underline{X}(n)) \geq \log \left(\max_{0 \leq k \leq L-1} |\det(\mathbf{H}_k)| \right). \quad (4.45)$$

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

$$h_\alpha(\underline{X}(n)) = \frac{d \log \alpha}{2(\alpha - 1)} + \frac{1}{2} \log \left((2\pi)^d \det(\text{Cov}(\underline{X}(n))) \right) \quad (4.46)$$

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

$$h_\alpha(\underline{Y}(n)) - h_\alpha(\underline{X}(n)) = \frac{1}{2} \log \left(\det \left(\sum_{k=0}^{L-1} \mathbf{H}_k \mathbf{H}_k^T \right) \right). \quad (4.47)$$

Example 2. Let

$$Y(n) = 2X(n) - X(n-1) - X(n-2) \quad (4.48)$$

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 (4.42), (4.43), (4.44), (4.45) 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 [11] and [7] respectively). These lower bounds are compared to the achievable value in (4.47), for an i.i.d. Gaussian input, which is equal to 0.8959 nats.

Chapter 5

Summary

5.1 Research Results and Tools

This thesis provides two forms 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 (1.8), they refer to orders $\alpha \in (1, \infty]$, and they coincide with the EPI [50] by letting $\alpha \rightarrow 1$. Theorem 1 provides an R-EPI in a closed form, improving the R-EPI by Bobkov and Chistyakov in [11, Theorem 1]; furthermore, for a sum of two independent random vectors ($n = 2$), the R-EPI in Theorem 1 is asymptotically tight when $\alpha \rightarrow \infty$. The R-EPI which is introduced in Theorem 2 can be efficiently calculated via a simple and efficient numerical algorithm, it is tighter than the R-EPI in Theorem 1 and all previously reported bounds, and it is currently the best known R-EPI for $\alpha \in (1, \infty]$. While Theorem 1 does not necessarily improve the bound by Bercher and Vignat [7], Theorem 2 does. Corollary 3 provides a closed-form expression for the R-EPI in Theorem 2 for $n = 2$. 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 [11] 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 [17] 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.

5.1.1 Remarks on Tightness

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 since at least one of the inequalities involved in the derivation of (3.4) (see Appendix A) is loose. These inequalities are the sharpened Young’s inequality in (2.20), and (2.11). 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 (2.16) reduces to

$$\|f * g\|_\infty \leq \|f\|_p \|g\|_{p'}, \tag{5.1}$$

where $p > 1$ and $p' = \frac{p}{p-1}$. Equality holds in (5.1) if f and g are scaled versions of a uniform distribution on the same convex set, which is also the same condition for tightness in (2.11); this is consistent with our conclusion that the R-EPIs in Theorems 1 and 2 are, however, asymptotically tight for $n = 2$ by letting $\alpha \rightarrow \infty$.

5.2 Further Research

5.2.1 The $\alpha < 1$ Case

In view of the derivation of (3.4) (see Appendix A), which relies on the sharpened Young’s inequality and on the monotonicity of the Rényi entropy power in its order, one might think that for $\alpha < 1$, the reverse sharpened Young’s inequality in (2.21) can be used to derive

an R-EPI (or a reversed version of this inequality). Unfortunately, this is not true since in this case, (2.11) and (2.21) yield inequalities in opposite directions. However, for $\alpha = 0$, the Brunn–Minkowski inequality can be used to derive an R-EPI. For completeness, the Brunn–Minkowski inequality is presented.

Fact 3 (Brunn–Minkowski inequality). Let $A, B \subseteq \mathbb{R}^d$ be two compact sets, and denote by ‘+’ the Minkowski sum, i.e., $A + B = \{a + b : a \in A, b \in B\}$. Then,

$$\lambda^{\frac{1}{d}}(A + B) \geq \lambda^{\frac{1}{d}}(A) + \lambda^{\frac{1}{d}}(B), \quad (5.2)$$

where λ is the Lebesgue measure in \mathbb{R}^d .

Proposition 2 (R-EPI for $\alpha = 0$). Let $\{X_k\}_{k=1}^n$ be d -dimensional independent random vectors. Then,

$$N_0\left(\sum_{k=1}^n X_k\right) \geq \sum_{k=1}^n N_0(X_k). \quad (5.3)$$

Proof. From the definition of $h_0(\cdot)$ in (1.4), it follows that for a random vector X with density f_X ,

$$N_0(X) = \exp\left(\frac{2}{d}h_0(X)\right) = \lambda^{\frac{2}{d}}(\text{supp}(f_X)). \quad (5.4)$$

If there exists $k \in \{1, \dots, n\}$ such that $N_0(X_k) = \infty$, then from (5.4), the support set of f_{X_k} has an infinite Lebesgue measure, and from the independence of X_1, \dots, X_n it follows that the density of $\sum_{k=1}^n X_k$ (which equals $f_{X_1} * \dots * f_{X_n}$) has an infinite Lebesgue measure as well, so $N_0(\sum_{k=1}^n X_k) = \infty$. Hence, we may assume that for every $k \in \{1, \dots, n\}$, $N_0(X_k) < \infty$ (or equivalently $\lambda(\text{supp}(f_{X_k})) < \infty$). In view of (5.2) and (5.4), it follows that

$$\sqrt{N_0(X_1 + X_2)} \geq \sqrt{N_0(X_1)} + \sqrt{N_0(X_2)}. \quad (5.5)$$

Since the Rényi entropy power is non-negative, by squaring the both sides of (5.5) we get

$$\begin{aligned} N_0(X_1 + X_2) &\geq N_0(X_1) + 2\sqrt{N_0(X_1)N_0(X_2)} + N_0(X_2) \\ &\geq N_0(X_1) + N_0(X_2), \end{aligned} \quad (5.6)$$

which completes the proof for $n = 2$. The proof for general n follows by a mathematical induction on n . \square

Note that Proposition 2 implies that $c_0 = 1$ (i.e., the multiplicative constant of the R-EPI for $\alpha = 0$ equals 1). From the EPI in (1.1), it follows that $c_1 = 1$. Since according to our R-EPIs in Theorems 1 and 2, c_α is monotonically decreasing in α for $\alpha \in (1, \infty)$, this leads to the following conjecture:

Conjecture 1. Let $\{X_k\}_{k=1}^n$ be independent random vectors and let $\alpha \in (0, 1)$. Then,

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq \sum_{k=1}^n N_\alpha(X_k). \quad (5.7)$$

Although our bounding techniques are not extendable for $\alpha < 1$, some R-EPIs that are restricted to certain families of distributions can be stated for $\alpha < 1$.

Proposition 3. Let $\{X_k\}_{k=1}^n$ be independent uniformly distributed random vectors and let $\alpha \in [0, 1)$. Then the following R-EPI holds,

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq \sum_{k=1}^n N_\alpha(X_k). \quad (5.8)$$

Proof. Since $\alpha < 1$, and the Rényi entropy power is non-increasing in its order, it follows that

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq N_1 \left(\sum_{k=1}^n X_k \right), \quad (5.9)$$

which in view of the EPI in (1.1) implies that

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq \sum_{k=1}^n N_1(X_k). \quad (5.10)$$

For uniformly distributed random vectors, the Rényi entropy power is independent of its order α ; in other words, for every $k \in \{1, \dots, n\}$, $N_1(X_k) = N_\alpha(X_k)$, hence (5.10) yields

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq \sum_{k=1}^n N_\alpha(X_k), \quad (5.11)$$

which completes the proof. \square

Proposition 3 supports Conjecture 1, and another confirmation comes from the Gaussian family; for independent Gaussian random vectors with proportional covariances $\{X_k\}_{k=1}^n$ we have

$$N_\alpha \left(\sum_{k=1}^n X_k \right) = \sum_{k=1}^n N_\alpha(X_k), \quad \forall \alpha \in [0, \infty]. \quad (5.12)$$

Although these two families of distributions satisfy Conjecture 1, it still needs to be proved in its full generality.

5.2.2 Further Generalizations

Future research can be done by finding generalizations of results related to the EPI to the Rényi entropy power. Before presenting some examples to EPI related results, it is mentioned that trying to generalize these results to the Rényi entropy power by following their proof and replacing Rényi's entropy in every instance of the Shannon entropy doesn't work, since in every proof there is a use of some property of the Shannon entropy which does not hold for the Rényi entropy. This means that a different approach to the proof is needed. Possible generalization of EPI related results to Rényi measures are given in the following.

R-EPIs for discrete random vectors

Although the EPI in (1.1) does not hold, in general, for discrete random vectors (consider $n = 2$, $d = 1$ and two random variables X_1 and X_2 which are actually deterministic; $N(X_1 + X_2) = N(X_1) = N(X_2) = 1$), many versions of the EPI for some families of discrete random variables have been suggested [31, 32, 60]; this encourages finding discrete R-EPIs. For the following examples of discrete EPIs, all random variables are discrete, $H(\cdot)$ stands for the (discrete) Shannon entropy, and $N(\cdot)$ is the corresponding entropy power:

- 1) In [31, Theorem 1], it is stated that

$$H(X + X') \geq H(X) + g(H(X)) \quad (5.13)$$

where X, X' are integer valued i.i.d random variables, and $g: [0, \infty) \rightarrow [0, \infty)$ is a non-negative monotonically increasing function such that $g(0) = 0$ and $\lim_{x \rightarrow \infty} g(x) = \frac{1}{8} \log_2(e)$ (g is given explicitly in [31]). By defining $c_X = \frac{1}{2} \exp\{2g(H(X))\}$, (5.13) yields the desired discrete EPI,

$$N(X + X') \geq c_X (N(X) + N(X')). \quad (5.14)$$

Note that if X is deterministic, then $H(X) = 0$ and $c_X = \frac{1}{2}$.

- 2) In [32, Theorem 1.1], an EPI for binomial random variables is given: let $X_n \sim B(n, \frac{1}{2})$ be a binomial random variable with n trials and success probability of $\frac{1}{2}$. Then,

$$N(X_n + X_m) \geq N(X_n) + N(X_m). \quad (5.15)$$

- 3) In [60, Theorem 2.5], it is stated that if X and Y are uniformly distributed on some subsets of \mathbb{Z} , then

$$N(X + Y) \geq N(X) + N(Y) - 1, \quad (5.16)$$

and if in addition X and Y are not deterministic, then

$$N(X + Y) \geq N(X) + N(Y). \quad (5.17)$$

Possible generalization of these discrete versions of EPIs to discrete R-EPIs provide directions for future research and are motivated by the applications of the discrete Rényi entropy in information theory (see [19, 3, 18]).

R-EPI for linear transformed vectors

One alternative way to write the EPI in (1.1) is the following: Let $\{X_k\}_{k=1}^n$ be independent d -dimensional random vectors and let $\{\tilde{X}_k\}_{k=1}^n$ be d -dimensional independent Gaussians with proportional covariance matrices such that, for every $k \in \{1, \dots, n\}$, $h(X_k) = h(\tilde{X}_k)$. Then

$$h\left(\sum_{k=1}^n X_k\right) \geq h\left(\sum_{k=1}^n \tilde{X}_k\right). \quad (5.18)$$

In [62, Theorem 1], Zamir and Feder introduced a generalized EPI for linear transformations: Let $X = (X_1, \dots, X_d)$ be a d -dimensional random vector with independent components, and let $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_d)$ be a Gaussian random vector with independent components such that $h(X_j) = h(\tilde{X}_j)$, for every $j \in \{1, \dots, d\}$. Then, for every matrix $A \in \mathbb{R}^{m \times d}$,

$$h(AX^T) \geq h(A\tilde{X}^T). \quad (5.19)$$

Note that if $m = 1$ and $A = (1, \dots, 1)$, (5.19) boils down to the EPI in (5.18). However, for $m > 1$, one cannot use the EPI to prove (5.19), and in [62] it is proved by induction on the matrix dimensions m, d . A possible generalization of (5.19) to the Rényi entropy is of interest.

Generalized R-EPIs for arbitrary subsets of indices

Consider the set $[n] = \{1, \dots, n\}$ for $n \geq 2$, and let C be some arbitrary collection of subsets of $[n]$; i.e., $C = \{s : s \in 2^{[n]}\}$, where $2^{[n]}$ is the power set of $[n]$. For every $k \in [n]$, define $r_k = |\{s \in C : k \in s\}|$ and $r = \max_{k \in [n]} r_k$. Madiman and Barron showed in [40, Theorem 3] that for independent random variables $\{X_k\}_{k=1}^n$

$$N\left(\sum_{k=1}^n X_k\right) \geq \frac{1}{r} \sum_{s \in C} N\left(\sum_{k \in s} X_k\right). \quad (5.20)$$

For the choice of C as the set of singletons, (5.20) reduces to the EPI in (1.1) and for the choice of C as the “*leave one out*” subsets, i.e., $C = \{\{1, 2, \dots, n\} \setminus k : 1 \leq k \leq n\}$, (5.20) yields another result [1, Theorem 3]:

$$N\left(\sum_{k=1}^n X_k\right) \geq \frac{1}{n-1} \sum_{i=1}^n N\left(\sum_{k \neq i} X_k\right). \quad (5.21)$$

Since (5.20) generalizes the EPI and more related inequalities (e.g., (5.21)), a possible generalization to the Rényi entropy powers of interest.

Possible Strengthening the R-EPIs

As mentioned in Chapter 1, the EPI in (1.1) is tight and is achieved with equality for Gaussian random vectors with proportional covariances. However, it can be refined by a restriction to some families of distributions; this has been done, for example, in [20, 22, 55]. Finding some possible strengthened R-EPIs is of interest and is suggested for future research. Examples for strengthened EPIs are given in the following:

- 1) Let $X_t = X + \sqrt{t}Z$, where X is an arbitrary random vector and Z is a Gaussian random vector with i.i.d. components having mean zero and variance 1. In [20, Theorem 3], Costa proved that the function defined by $f(t) = N(X_t)$ is concave for every $t \geq 0$. This in turn implies that [20, inequality (8)],

$$N(X_t) \geq (1-t)N(X) + tN(X+Z), \quad \forall t \in [0, 1]. \quad (5.22)$$

The latter yields the EPI in (1.1); indeed, by substituting $\beta = \frac{1}{\sqrt{t}}$ and setting $Z' = \frac{1}{\beta}Z$, it follows from (5.22) and the homogeneity property of the entropy power that

$$\begin{aligned} N(X + Z') &\geq \left(1 - \frac{1}{\beta^2}\right)N(X) + \frac{1}{\beta^2}N(X + \beta Z') \\ &= \left(1 - \frac{1}{\beta^2}\right)N(X) + N\left(\frac{X}{\beta^2} + Z'\right), \quad \forall \beta \geq 1. \end{aligned} \quad (5.23)$$

In view of (5.23), by letting $\beta \rightarrow \infty$, the EPI when the additive noise is Gaussian is re-established.

- 2) In [22, Theorem 2], a strengthened and generalized EPI is given. To distinguish between random vectors and variables, random vectors are written, for this example only, in a bold font. Let \mathbf{X}, \mathbf{Z} be random vectors and Q, V be random variables and let $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. Assume that \mathbf{X} and \mathbf{Z} are independent given Q and that \mathbf{Z} is Gaussian (with an arbitrary covariance matrix) given Q . If given Q , $\mathbf{X}-\mathbf{Y}-V$ forms a Markov chain (i.e. $f_{V|X,Y,Q} = f_{V|Y,Q}$), then

$$\exp \left\{ \frac{2}{d} (h(\mathbf{Y}|Q) - I(\mathbf{X}; V|Q)) \right\} \geq \exp \left\{ \frac{2}{d} (h(\mathbf{X}|Q) - I(\mathbf{Y}; V|Q)) \right\} + \exp \left\{ \frac{2}{d} (h(\mathbf{W}|Q)) \right\}. \quad (5.24)$$

The unconditioned version of (5.24) is written as follows. If \mathbf{X} and $\mathbf{Z} \sim N(0, \Sigma)$ are independent, then for every V such that $\mathbf{X}-(\mathbf{X} + \mathbf{Y})-V$ forms a Markov chain,

$$N(\mathbf{X} + \mathbf{Z}) \exp \left\{ \frac{2}{d} (-I(\mathbf{X}; V)) \right\} \geq N(\mathbf{X}) \exp \left\{ \frac{2}{d} (-I(\mathbf{X} + \mathbf{Z}; V)) \right\} + N(\mathbf{Z}). \quad (5.25)$$

The latter implies the EPI in (1.1) if V is independent of \mathbf{X} and \mathbf{Z} .

- 3) In [55, inequality 14], it stated that if X and Y are independent random vectors with log-concave densities, then

$$N(X + Y) \geq R_{X,Y}(N(X) + N(Y)) \quad (5.26)$$

where $R_{X,Y} \geq 1$ and $R_{X,Y} = 1$ if and only if both X and Y are Gaussian random vectors ($R_{X,Y}$ is given explicitly in [55] and serves as a measure of the non-Gaussianity of the two random vectors X and Y). Possible strengthening of the R-EPIs in this thesis for log-concave random vectors is suggested as a subject for future research.

We conclude that there are many directions for future research concerning Rényi entropy power inequalities; the R-EPIs derived in this thesis may have many generalizations and extensions. It is our hope that the usefulness of the Rényi entropy and the EPI in information theory will encourage researchers to further study these possible generalizations, and to find useful applications of these generalized inequalities in information theory.

Appendix A

Proof of (3.4)

Since $\{X_k\}_{k=1}^n$ are independent, the density of $S_n = \sum_{k=1}^n X_k$ is the convolution of the densities f_{X_k} . In view of (2.20) and (2.5), for $\alpha > 1$,

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

where

$$\nu_k > 1, \quad 1 \leq k \leq n \tag{A.2}$$

$$\sum_{k=1}^n \frac{1}{\nu_k'} = \frac{1}{\alpha'} \tag{A.3}$$

and, due to (2.14) and (2.19),

$$A = \left(A_{\alpha'} \prod_{k=1}^n A_{\nu_k} \right)^{\frac{d}{2}}. \tag{A.4}$$

From (A.2) and (A.3) it follows that $\nu_k \in (1, \alpha]$ for all $k \in \{1, \dots, n\}$, hence in view of Corollary 1,

$$\|f_{X_k}\|_{\nu_k'} \leq \|f_{X_k}\|_{\alpha'}, \quad 1 \leq k \leq n. \tag{A.5}$$

Combining (A.1) and (A.5), and defining $t_k = \frac{\alpha'}{\nu'_k}$ yields

$$N_\alpha(S_n) \geq A^{-\frac{2\alpha'}{d}} \prod_{k=1}^n (\|f_{X_k}\|_\alpha)^{-\frac{2\alpha'}{d} \cdot \frac{\alpha'}{\nu'_k}} = A^{-\frac{2\alpha'}{d}} \prod_{k=1}^n N_\alpha^{t_k}(X_k) \quad (\text{A.6})$$

which by setting $B = A^{-\frac{2\alpha'}{d}}$ completes the proof of (3.4) with the constant B as given in (3.5).

Appendix B

Proof of Proposition 1

In view of (4.7), if f_0 is concave, so is f . As it is verified in Section 4.1.1, the function f_0 is concave for all $\alpha \in (1, 2)$ (i.e., $\alpha' \in (2, \infty)$), and hence also f is concave for these values of α . We therefore need to prove the concavity of f in (4.7) whenever $\alpha' \in (1, 2)$ (i.e., if $\alpha \in (2, \infty)$), although f_0 is not concave for these values of α .

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 (4.10)). In view of (4.8), 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 (4.10) 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 (4.12), $D \preceq 0$ 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 (4.10) and (4.12) that $\rho = q(1 - \sum_{j=1}^{n-1} t_j) < 0$); hence, (4.11) 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

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

Proof. In view of (4.10), the left side of (B.1) 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

$$\frac{1}{q(u)} + \frac{1}{q(1 - u - v)} - \frac{1}{q(1 - v)} > 0. \quad (\text{B.2})$$

Proof. In view of (4.10), the left side of (B.2) 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

$$\begin{aligned} t_1, \dots, t_{n-1} &> 0, \\ \sum_{k=1}^{n-1} t_k &< 1 - \frac{\alpha'}{2}, \\ t_n &= 1 - \sum_{k=1}^{n-1} t_k \end{aligned} \quad (\text{B.3})$$

then

$$\sum_{k=1}^n \frac{1}{q(t_k)} > 0. \quad (\text{B.4})$$

Proof. Lemma 3 is proved by using mathematical induction on n . In view of Lemma 1, (B.4) holds for $n = 2$. Assuming its correctness for n , we have

$$\sum_{j=1}^{n-1} \frac{1}{q(t_j)} + \frac{1}{q(\bar{t}_n)} > 0 \quad (\text{B.5})$$

where, from (B.3), $\bar{t}_n = 1 - \sum_{k=1}^{n-1} t_k$. We prove in the following that (B.4) also holds for $n + 1$ when the constraints in (B.3) are satisfied with $n + 1$, i.e.,

$$\begin{aligned} t_1, \dots, t_n &> 0, \\ \sum_{k=1}^n t_k &< 1 - \frac{\alpha'}{2}, \\ t_{n+1} &= 1 - \sum_{k=1}^n t_k. \end{aligned} \quad (\text{B.6})$$

Consequently, the left side of (B.4) is equal to

$$\begin{aligned} \sum_{k=1}^{n+1} \frac{1}{q(t_k)} &= \sum_{k=1}^{n-1} \frac{1}{q(t_k)} + \frac{1}{q(t_n)} + \frac{1}{q(t_{n+1})} \\ &> -\frac{1}{q(\bar{t}_n)} + \frac{1}{q(t_n)} + \frac{1}{q(t_{n+1})} \end{aligned} \quad (\text{B.7})$$

$$= \frac{1}{q(t_n)} + \frac{1}{q(1 - \sum_{k=1}^n t_k)} - \frac{1}{q(1 - \sum_{k=1}^{n-1} t_k)} \quad (\text{B.8})$$

$$> 0 \quad (\text{B.9})$$

where (B.7) follows from (B.5); (B.8) holds by the equality constraint in (B.6); (B.9) 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 (B.6). Hence, it follows by mathematical induction that Lemma 3 holds for every $n \geq 2$. \square

In the following, we prove the concavity of f when $q(t_k) \neq 0$ for all $k \in \{1, \dots, n - 1\}$ (recall that the case where there exists $k \in \{1, \dots, n - 1\}$ 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 $(q(t_1), \dots, q(t_{n-1}))$ is a vector whose all entries are distinct. To justify

this assumption, note that since the function q in (4.10) 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, \dots, t_{n-1}) are all distinct. Otherwise, if some of the entries of the vector (t_1, \dots, 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, \dots, 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 (4.6) is invariant to a permutation of the entries of \underline{t} , it is assumed without loss of generality that

$$q(t_1) < q(t_2) < \dots < q(t_{n-1}). \quad (\text{B.10})$$

In view of (B.10), there are only two possibilities: either

$$q(t_1) < q(t_2) < \dots < q(t_{n-2}) < q(t_{n-1}) < 0, \quad (\text{B.11})$$

or

$$q(t_1) < q(t_2) < \dots < q(t_{n-2}) < 0 < q(t_{n-1}) \quad (\text{B.12})$$

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 (4.8).

The continuation of this proof relies on Fact 2 by Bunch *et al.* [17] (see Section 2.3), and on Lemma 3. For the continuation of this proof, let

$$t_n = 1 - \sum_{k=1}^{n-1} t_k. \quad (\text{B.13})$$

Case 1: If (B.11) holds, then (4.12) implies that

$$D \prec 0. \quad (\text{B.14})$$

- If $q(t_n) < 0$ then $\rho = q(t_n)\mathbf{1}\mathbf{1}^T \prec 0$ which, in view of (4.11) and (B.14), implies that $\nabla^2 f(t_1, \dots, t_{n-1}) \prec 0$.
- Otherwise, if $q(t_n) > 0$ then $\rho > 0$ (see (4.12) and (B.13)); from (4.11) and the interlacing property in (2.22), the eigenvalues $\lambda_1, \dots, \lambda_{n-1}$ of $\nabla^2 f(t)$ satisfy

$$q(t_1) < \lambda_1 < q(t_2) < \dots < q(t_{n-2}) < \lambda_{n-2} < q(t_{n-1}) < \lambda_{n-1} \quad (\text{B.15})$$

where, in view of the third item of Fact 2, the inequalities in (B.15) are strict. From (B.11) and (B.15), it follows that $\lambda_1, \dots, \lambda_{n-2} < 0$. To prove that $\nabla^2 f(t_1, \dots, t_{n-1}) \prec 0$, it remains to show that also $\lambda_{n-1} < 0$. In view of the third item of Fact 2 and (4.11), the eigenvalues of $\nabla^2 f(t_1, \dots, t_{n-1})$ satisfy the equation

$$1 + q(t_n) \sum_{j=1}^{n-1} \frac{1}{q(t_j) - \lambda} = 0 \quad (\text{B.16})$$

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

$$\sum_{j=1}^{n-1} \frac{1}{\lambda_k - q(t_j)} = \frac{1}{q(t_n)}. \quad (\text{B.17})$$

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 (B.13) that

$$\sum_{k=1}^{n-1} t_k < 1 - \frac{\alpha'}{2}. \quad (\text{B.18})$$

Since $q(t_j) < 0$ for all $j \in \{1, \dots, n-1\}$, if $\lambda_{n-1} > 0$, then in view of (B.17)

$$\begin{aligned} \sum_{j=1}^{n-1} \frac{1}{-q(t_j)} &\geq \sum_{j=1}^{n-1} \frac{1}{\lambda_{n-1} - q(t_j)} \\ &= \frac{1}{q(t_n)}. \end{aligned} \quad (\text{B.19})$$

Rearrangement of terms in (B.19) yields

$$\sum_{j=1}^n \frac{1}{q(t_j)} \leq 0 \quad (\text{B.20})$$

and, in view of the interior of \mathcal{D}^{n-1} in (4.8), and (B.13) and (B.18), inequality (B.20) 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 (B.11).

Case 2: We now consider the case where (B.12) holds. Under this assumption,

$$q(t_n) < 0. \quad (\text{B.21})$$

To verify (B.21), 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 (4.10)) 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 (B.13); hence, indeed $q(t_n) < 0$. Consequently, in view of (4.11), let

$$\bar{C} = \frac{1}{q(t_n)} \nabla^2 f(t_1, \dots, t_{n-1}) \quad (\text{B.22})$$

$$= \bar{D} + \mathbf{1}\mathbf{1}^T \quad (\text{B.23})$$

where

$$\bar{D} = \text{diag} \left(\frac{q(t_1)}{q(t_n)}, \dots, \frac{q(t_{n-1})}{q(t_n)} \right). \quad (\text{B.24})$$

From (B.12) and (B.21), it follows that

$$\frac{q(t_1)}{q(t_n)} > \frac{q(t_2)}{q(t_n)} > \dots > \frac{q(t_{n-2})}{q(t_n)} > 0 > \frac{q(t_{n-1})}{q(t_n)}. \quad (\text{B.25})$$

It is shown in the following that $\bar{C} \succeq 0$ which, from (B.21) and (B.22), imply that indeed $\nabla^2 f(t_1, \dots, t_{n-1}) \preceq 0$. Let $\{\lambda_k\}_{k=1}^{n-1}$ designate the eigenvalues of \bar{C} ; in view of (B.23) and the

last two items of Fact 2, it follows that

$$\overbrace{\frac{q(t_{n-1})}{q(t_n)}}^{<0} < \lambda_1 < \overbrace{\frac{q(t_{n-2})}{q(t_n)}}^{>0} < \lambda_2 < \dots < \overbrace{\frac{q(t_2)}{q(t_n)}}^{>0} < \lambda_{n-2} < \overbrace{\frac{q(t_1)}{q(t_n)}}^{>0} < \lambda_{n-1}. \quad (\text{B.26})$$

Hence, (B.26) asserts that $\lambda_2, \dots, \lambda_{n-1} > 0$, and it only remains to prove that $\lambda_1 > 0$. From the third item of Fact 2, and from (B.22), (B.23), (B.24), the eigenvalues $\{\lambda_k\}_{k=1}^n$ of the rank-one modification \bar{C} satisfy the equality

$$1 + \sum_{j=1}^{n-1} \frac{1}{\frac{q(t_j)}{q(t_n)} - \lambda_k} = 0 \quad (\text{B.27})$$

for all $k \in \{1, \dots, n-1\}$. Assume on the contrary that $\lambda_1 \leq 0$, then from (B.27)

$$1 + \sum_{j=1}^{n-1} \frac{q(t_n)}{q(t_j)} \geq 1 + \sum_{j=1}^{n-1} \frac{1}{\frac{q(t_j)}{q(t_n)} - \lambda_1} = 0. \quad (\text{B.28})$$

Consequently, from (B.21) and (B.28), it follows that $\sum_{j=1}^n \frac{1}{q(t_j)} \leq 0$ in contradiction to Lemma 3. Hence, $\lambda_k > 0$ for every $k \in \{1, \dots, n-1\}$, which therefore implies that $\nabla^2 f(t_1, \dots, t_{n-1}) \prec 0$ for all (t_1, \dots, t_{n-1}) in the interior of \mathcal{D}^{n-1} . This completes the proof of Proposition 1.

Appendix C

Derivation of (4.17)–(4.19) From Lagrange Duality

We consider the convex optimization problem in (4.6), and solve it via the use of the Lagrange duality where strong duality holds.

The Lagrangian of the convex optimization problem in (4.6) is given by

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

where $\lambda \succeq 0$, the function g is defined in (4.2), and $N_k := N_\alpha(X_k)$ (see (4.3)).

In view of the Lagrangian in (C.1) and the function g defined in (4.2), straightforward calculations of the partial derivatives of L with respect to t_k for $k \in \{1, \dots, n-1\}$ yields

$$\begin{aligned} \frac{\partial L}{\partial t_k} &= g'(t_k) - g'(1 - t_1 - \dots - t_{n-1}) + \log\left(\frac{N_\alpha(X_k)}{N_\alpha(X_n)}\right) + \lambda_k - \lambda_n \\ &= -\log\left(t_k\left(1 - \frac{t_k}{\alpha'}\right)\right) + \log\left(t_n\left(1 - \frac{t_n}{\alpha'}\right)\right) + \log\left(\frac{N_\alpha(X_k)}{N_\alpha(X_n)}\right) + \lambda_k - \lambda_n \end{aligned} \tag{C.2}$$

where $t_n := 1 - \sum_{k=1}^{n-1} t_k$. By setting the partial derivatives in (C.2) to zero, and exponentiating both sides of the equation, we get for all $k \in \{1, \dots, n-1\}$

$$\frac{t_n(\alpha' - t_n)}{t_k(\alpha' - t_k)} = \frac{N_\alpha(X_n)}{N_\alpha(X_k)} \cdot \exp(\lambda_n - \lambda_k). \quad (\text{C.3})$$

In view of (C.3) and the definition of $\{c_k\}_{k=1}^{n-1}$ in (4.15), we obtain that for all $k \in \{1, \dots, n-1\}$

$$t_k(\alpha' - t_k) = c_k t_n(\alpha' - t_n) \exp(\lambda_k - \lambda_n). \quad (\text{C.4})$$

Consequently, (C.4), the definition of t_n , and the slackness conditions lead to the following set of constraints:

$$t_k \geq 0, \quad k \in \{1, \dots, n\} \quad (\text{C.5})$$

$$\sum_{k=1}^n t_k = 1 \quad (\text{C.6})$$

$$\lambda_k \geq 0, \quad k \in \{1, \dots, n\} \quad (\text{C.7})$$

$$\lambda_k t_k = 0, \quad k \in \{1, \dots, n\} \quad (\text{C.8})$$

$$t_k(\alpha' - t_k) = c_k t_n(\alpha' - t_n) \exp(\lambda_k - \lambda_n), \quad k \in \{1, \dots, n-1\} \quad (\text{C.9})$$

with the variables $\underline{\lambda}$ and \underline{t} in (C.5)–(C.9).

Consider first the case where

$$N_\alpha(X_k) > 0, \quad \forall k \in \{1, \dots, n-1\} \quad (\text{C.10})$$

which in view of (4.15), implies

$$c_k > 0, \quad \forall k \in \{1, \dots, n-1\}. \quad (\text{C.11})$$

Under the assumption in (C.10), we prove that

$$\lambda_k = 0, \quad \forall k \in \{1, \dots, n\}. \quad (\text{C.12})$$

Assume on the contrary that there exists an index k such that $\lambda_k \neq 0$. This would imply from (C.8) that $t_k = 0$. If $k = n$ (i.e., if $t_n = 0$) then it follows from (C.9) that also $t_k = 0$

for all $k \in \{1, \dots, n\}$ (recall that $\alpha' > 1$), which violates the equality constraint in (C.6). Otherwise, if $t_k = 0$ for some $k < n$, then it follows from (C.9) and (C.11) that $t_n = 0$ which leads to the same contradiction as above.

The substitution of (C.12) into the right side of (C.9) gives the simplified equation in (4.17). In view of (C.5) and (C.6), this leads to the simplified set of KKT constraints in (4.17)–(4.19).

Finally, if the assumption in (C.10) 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 (4.9) be equal to $-\infty$. In addition, in this case $c_k = 0$, so the simplified set of KKT constraints in (4.17)–(4.19) still yields the optimal solution \underline{t} .

Appendix D

On the existence and uniqueness of the solution to (4.23)

Define

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

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

$$\phi_\alpha(1) = 1 + \sum_{k=1}^{n-1} \psi_k(1) > 1, \quad (\text{D.2})$$

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

$$\phi'_\alpha(x) = 1 + \sum_{k=1}^{n-1} \frac{c_k(\alpha' - 2x)}{\sqrt{\alpha'^2 - 4c_k x(\alpha' - x)}}, \quad (\text{D.3})$$

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 (4.23) is unique.

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

$$\phi'_\alpha(x) > 0, \quad x \in [0, \frac{\alpha'}{2}]. \quad (\text{D.4})$$

Note that

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

thus in view of (D.3),

$$\phi'_\alpha(x) = 1 + \sum_{k=1}^{n-1} \frac{c_k}{\sqrt{c_k + \frac{\alpha'^2(1-c_k)}{4(x-\frac{\alpha'}{2})^2}}}. \quad (\text{D.5})$$

Eq. (D.5) 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 (4.23), i.e.,

$$\phi_\alpha(x_1) = \phi_\alpha(x_2) = 1. \quad (\text{D.6})$$

Eq. (D.6) implies that there exists $c \in (x_1, x_2)$ such that $\phi'_\alpha(c) = 0$ and from (D.4), $c \in (\frac{\alpha'}{2}, x_2)$. 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 (4.24) and (4.25)

If $N_\infty(X_k) = 0$ for all $k \in \{1, \dots, n\}$, the bounds in (4.24) and (4.25) obviously coincide asymptotically as $\alpha \rightarrow \infty$. In addition, in this case, the condition in (4.26) clearly holds as well. It is therefore assumed that $N_\infty(X_k)$ is strictly positive for at least one value of $k \in \{1, \dots, n\}$ which, under the assumption in (4.13), yields that

$$N_\infty(X_n) > 0. \tag{E.1}$$

Let c_k^* be defined as

$$c_k^* = \lim_{\alpha \rightarrow \infty} \frac{N_\alpha(X_k)}{N_\alpha(X_n)} = \frac{N_\infty(X_k)}{N_\infty(X_n)}. \tag{E.2}$$

In view of (E.2), the condition in (4.26) is equivalent to

$$\sum_{k=1}^{n-1} c_k^* \leq 1. \tag{E.3}$$

Hence, it remains to show that the the tightest R-EPI in (4.24) and the BV bound in (4.25) asymptotically coincide, by letting $\alpha \rightarrow \infty$, if and only if the condition in (E.3) holds.

Let $\phi_\alpha: [0, 1] \rightarrow \mathbb{R}$ be the function defined in (D.1) for $\alpha \in (1, \infty)$, and define

$$\phi_\infty(x) = \lim_{\alpha \rightarrow \infty} \phi_\alpha(x) \quad (\text{E.4})$$

for $x \in [0, 1]$. In view of (4.22), (D.1) and (E.2), the limit in (E.4) is given by

$$\phi_\infty(x) = x + \frac{1}{2} \sum_{k=1}^{n-1} \left(1 - \sqrt{1 - 4c_k^* x(1-x)} \right) \quad (\text{E.5})$$

for $x \in [0, 1]$. Recall that under the assumption in (4.13), the selection of $t_n = 1$ in (4.9) leads to the BV bound in (4.25). Hence, in view of (4.23), if $t = 1$ is the unique solution of

$$\phi_\infty(t) = 1, \quad t \in [0, 1] \quad (\text{E.6})$$

then the bounds in (4.24) and (4.25) asymptotically coincide by letting $\alpha \rightarrow \infty$. Note that,

$$\phi_\infty(0) = 0, \quad (\text{E.7})$$

$$\phi_\infty(1) = 1. \quad (\text{E.8})$$

From (E.8), $t = 1$ is a solution of (E.6) regardless of the sequence $\{c_k^*\}$. Moreover, from (E.5),

$$\phi'_\infty(x) = 1 + \sum_{k=1}^{n-1} \frac{c_k^* (1-2x)}{\sqrt{1-4c_k^* x(1-x)}}, \quad (\text{E.9})$$

so

$$\phi'_\infty(x) > 0, \quad \forall x \in (0, \frac{1}{2}), \quad (\text{E.10})$$

$$\phi'_\infty(1) = 1 - \sum_{k=1}^{n-1} c_k^*. \quad (\text{E.11})$$

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 $\alpha' \rightarrow 1$ by letting $\alpha \rightarrow \infty$. Thus, if the condition in (E.3) holds, then $\phi'_\infty(x) > 0$ for all $x \in (0, 1)$

which, in view of (E.8), yields that $t = 1$ is the unique solution of (E.6). This implies that the tightest R-EPI in (4.24) and the BV bound in (4.25) asymptotically coincide by letting $\alpha \rightarrow \infty$.

To prove the 'only if' part, one needs to show that if the condition in (E.3) does not hold then the bounds in (4.24) and (4.25) do not coincide asymptotically in the limit where $\alpha \rightarrow \infty$; in the latter case, we prove that our bound in (4.24) is tighter than (4.25). If (E.3) does not hold, then (E.11) implies that

$$\phi'_\infty(1) < 0. \quad (\text{E.12})$$

Hence, from (E.8), there exists $x_0 \in (0, 1)$ such that $\phi_\infty(x_0) > 1$ which, in view of (E.7) and the continuity of $\phi_\infty(\cdot)$, implies that there exists $t \in (0, x_0)$ which is a solution of (E.6). This implies that there are two different solutions of (E.6) in the interval $[0, 1]$. Let $t^{(1)} \in (0, 1)$ and $t^{(2)} = 1$ denote such solutions, i.e.,

$$t^{(1)} < t^{(2)} = 1. \quad (\text{E.13})$$

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 (4.25); the bound in (4.25) corresponds to $t^{(2)} = 1$ under the assumption in (4.13). For every $\alpha > 1$, let $t(\alpha)$ be the unique solution of (4.23) (see Appendix D). It follows that the limit of any convergent subsequence $\{t(\alpha_n)\}$, as $\alpha_n \rightarrow \infty$, is either $t^{(1)} \in (0, 1)$ or $t^{(2)} = 1$. In the sequel, if the condition in (4.26) is not satisfied, we show that every such subsequence tends to $t^{(1)} \in (0, 1)$, which therefore implies that

$$\lim_{\alpha \rightarrow \infty} t(\alpha) = t^{(1)} < 1. \quad (\text{E.14})$$

From (E.12) and the continuity of $\phi_\infty(\cdot)$, it follows that there exists $\delta > 0$ such that

$$\phi_\infty(x) > 1, \quad \forall x \in (1 - \delta, 1). \quad (\text{E.15})$$

In addition, since $\phi_\alpha(\cdot)$ is continuous in α for every $x \in [0, 1]$, it follows from (E.15) 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 (D.2)). Hence, since by definition $\phi_\alpha(t(\alpha)) = 1$ 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 $\alpha_n \rightarrow \infty$ (since it cannot converge to $t^{(2)} = 1$), which yields (E.14). Hence, the R-EPI in Theorem 2 asymptotically yields a tighter bound than (4.25) when $\alpha \rightarrow \infty$; this therefore proves the 'only if' part of our claim.

Bibliography

- [1] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, “*Journal of the American Mathematical Society*”, vol. 17, num. 4, pp. 975–982, May, 2004.
- [2] V. Anantharam, “Entropy power inequalities: results and speculation,” *Workshop on Coding and Information Theory*, Institute of Mathematical Research, the University of Hong Kong, December 2013.
- [3] E. Arikan, “An inequality on guessing and its application to sequential decoding,” *IEEE Trans. on Information Theory*, vol. 42, no. 1, pp. 99–105, January 1996.
- [4] 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.
- [5] 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.
- [6] 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.

- [7] 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.
- [8] 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.
- [9] N. M. Blachman, “The convolution inequality for entropy powers,” *IEEE Trans. on Information Theory*, vol. 11, no. 2, pp. 267–271, April 1965.
- [10] 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.
- [11] 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.
- [12] S. Bobkov and M. Madiman, “Dimensional behaviour of entropy and information,” *Comptes Rendus Mathématique*, vol. 349, no. 3–4, pp. 201–204, February 2011.
- [13] 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.
- [14] S. Bobkov and A. Marsiglietti, “Variants of the Entropy Power Inequality”, preprint, September 2016. [Online]. Available at <http://arxiv.org/abs/1609.04897>.
- [15] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [16] 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.

- [17] 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.
- [18] C. Bunte and A. Lapidoth, “Encoding Tasks and Rényi Entropy”, *IEEE Trans. on Information Theory*, vol. 60, no. 9, pp. 5065–5076, September 2014.
- [19] L. L. Campbell, “A coding theorem and Rényi’s entropy,” *Inform. Control*, vol. 8, no. 4, pp. 423–429, August 1965.
- [20] M. H. M. Costa, “A new entropy power inequality,” *IEEE Trans. on Information Theory*, vol. 31, no. 6, pp. 751–760, November 1985.
- [21] M. H. M. Costa, “On the Gaussian interference channel,” *IEEE Trans. on Information Theory*, vol. 31, no. 5, pp. 607–615, September 1985.
- [22] 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.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, second edition, 2006.
- [24] I. Csiszár, “Generalized cutoff rates and Rényi information measures,” *IEEE Trans. on Information Theory*, vol. 41, no. 1, pp. 2634, January 1995.
- [25] 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.
- [26] 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.
- [27] S. Fehr and S. Berens, “On the conditional Rényi entropy,” *IEEE Trans. on Information Theory*, vol. 60, no. 11, pp. 6801–6810, November 2014.

- [28] 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.
- [29] 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.
- [30] 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.
- [31] S. Haghghatshoar, 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.
- [32] 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.
- [33] 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.
- [34] O. Johnson, *Information Theory and the Central Limit Theorem*, Imperial College Press, 2004.
- [35] 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.
- [36] 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.

- [37] 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.
- [38] L. Leindler, “On a certain converse of Hölder’s inequality,” *Linear Operators and Approximation*, Proceedings of the Conference held at the Oberwolfach Mathematical Research Institute, Black Forest, August, 1971.
- [39] 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.
- [40] 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.
- [41] M. Madiman, “On the entropy of sums,” *Proceedings of the 2008 IEEE International Symposium on Information Theory*, pp. 303–307, Porto, Portugal, May 2008.
- [42] 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>.
- [43] 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.
- [44] 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.
- [45] O. Rioul, “Information theoretic proofs of entropy power inequalities,” *IEEE Trans. on Information Theory*, vol. 57, no. 1, pp. 33–55, January 2011.
- [46] 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.

- [47] I. Sason, “On the Rényi Divergence, Joint Range of Relative Entropies, and a Channel Coding Theorem”, *IEEE Trans. on Information Theory*, vol. 62, no. 1, pp 23–34, January 2016.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] 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.
- [53] 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.
- [54] A. Teixeira, A. Matos, and L. Antunes, “Conditional Rényi Entropies,” *IEEE Trans. on Information Theory*, vol. 58, no. 7, pp. 4273–4277, July 2012.
- [55] 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.
- [56] 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.

- [57] 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.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] 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.

**אי שוויונות מוכללים מסוג
חזקת האנטרופיה של רייני**

אשד רם

אי שוויונות מוכללים מסוג חזקת האנטרופיה של רייני

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת תואר
מגיסטר למדעים בהנדסת חשמל

אשד רם

הוגש לסנט הטכניון – מכון טכנולוגי לישראל

דצמבר 2016

חיפה

כסליו ה'תשע"ז

המחקר נעשה בהנחיית פרופסור חבר יגאל ששון

בפקולטה להנדסת חשמל

הכרת תודה

ברצוני להודות למנחה שלי, פרופסור חבר יגאל ששון, על התמיכה, העזרה וההדרכה לאורך כל המחקר הזה. המסירות שלו יותר ממוערכת והמקצועיות שלו חשפה בפני את עולם המחקר בצורה יסודית ומעניינת. בפרט, אני מודה לו על שלימד אותי את אומנות הכתיבה האקדמית.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי

העבודה הזו פורסמה בכתב העת, "IEEE Transactions on Information Theory", בדצמבר 2016. העבודה הוצגה בכנס "International Symposium on Information Theory", בברצלונה, יולי 2016.

לאשתי, נגה רם.

תקציר

אחד מאי השוויונים הידועים בתורת האינפורמציה הוא ה-Entropy Power Inequality (EPI) אשר הוצג במאמרו פורץ הדרך של שאנון משנת 1948. לאי שוויון זה ישנם שימושים רבים בתורת האינפורמציה שעיקרם משפטים הפוכים (converse theorems) עבור מספר בעיות שמשלבות בתוכן ווקטור אקראי גאوسي: ערוץ ההפצה (broadcast) הגאوسي, ערוץ ההפרעה (interference) הגאوسي, ערוץ ההאזנה (wire-tap) הגאوسي, פונקציית קצב-עיוות (rate-distortion) בתורת האינפורמציה מרובת משתמשים וערוץ ההפצה הגאوسي מרובה כניסות ומרובה יציאות (MIMO). בגלל החשיבות הרבה שיש ל-EPI בתורת האינפורמציה, מספר הוכחות הוצעו עבורו לאורך השנים. עוד תוצאות קשורות כוללות גרסאות של ה-EPI עבור וקטורים אקראיים בדידים, EPI מוכללים, EPI הפוכים, וחיזוקים של ה-EPI עבור משפחות מסוימות של וקטורים אקראיים.

האנטרופיה של Rényi והדיברגנץ של Rényi הוצגו לראשונה ע"י אלפרד רני (Alfred Rényi) בתחילת שנות ה-60 כהכללה לאנטרופיה של שאנון ולדיברגנץ (Kullback-Liebler Divergence), ויש להם שימושים רבים בתורת האינפורמציה, החל מקידוד מקור ועד לקידוד ערוץ.

תזת המאסטר הזו מתמקדת בהכללת ה-EPI עבור הממד המתאים לו במידות Rényi, כלומר למצוא

Rényi Entropy Power Inequality (R-EPI). יהי $S_n = \sum_{k=1}^n X_k$ סכום של n וקטורים אקראיים

רציפים בלתי תלויים סטטיסטית המקבלים ערכים ב- \mathbb{R}^d , ויהי $\alpha \in [1, \infty]$. ה-R-EPI מציג חסם תחתון על ה-Rényi Entropy Power של S_n מסדר α , אשר עד כדי קבוע כפלי (אשר יכול להיות תלוי בסדר α , במימד d ובמספר הווקטורים האקראיים n) שווה לסכום של ה-Rényi Entropy Power מסדר α של כל אחד מ- n הווקטורים האקראיים $\{X_k\}_{k=1}^n$. עבור $\alpha = 1$, ה-R-EPIs שפותחו בתזה זו מתלכדים עם ה-EPI הקלאסי של שאנון.

עקב הדמיון הרב בין ה-Rényi Entropy Power לנורמת L_α , ישנו שימוש בהגדרות ותוצאות מאנליזה פונקציונלית. תוצאות אלו כוללות אי שוויונים של Hölder, Minkowski ו-Young.

התזה מאורגנת באופן הבא: בפרק 2 מוצגים הכלים האנליטיים אשר חשובים להבנה מלאה של תהליך פיתוח החסם: אי שוויונים, תכונות חשובות של מידות Rényi, ותוצאה מתורת המטריצות אשר מקשרת בין הערכים העצמיים של מטריצה אלכסונית ממשית לאילו של מטריצה אחרת שמהווה שינוי בדרגה אחת שלה (The Rank-One Modification Theorem).

ה-R-EPI הראשון מוצג בפרק 3 והוא מהווה שיפור של R-EPI קודם שהוצג לאחרונה על-ידי Bobkov and Chistyakov (IEEE Trans. on information Theory, Feb. 2015) אשר מתבסס על הגרסה ההדוקה של אי השוויון של Young. ה-R-EPI הראשון הינו הדוק אסימפטוטית עבור $n = 2$ כאשר $\alpha \rightarrow \infty$, והוא מתלכד במקרה זה עם תוצאה ישנה של Rogozin.

שיפור נוסף של ה-R-EPI נעשה בפרק 4, והוא מתבסס על אופטימיזציה קמורה ופתרון תנאי Karush-Kuhn-Tucker (KKT). הוכחת קמירות בעיית האופטימיזציה מתבססת על תוצאות הקשורות לתורת המטריצות. ה-R-EPI בפרק 4, הינו ההדוק ביותר שידוע. באופן כללי, החסם המשופר אינו נתון בצורה סגורה וניתן לחשבו בעזרת אלגוריתם נומרי פשוט ויעיל. אם זאת, במקרים מסוימים ניתן להציגו בצורה סגורה, למשל עבור $n = 2$. ה-R-EPI המשופרים מושווים לתוצאות קודמות של Bobkov and Chistyakov וBercher and Vignati בהקשר של סינון וקטורים אקראיים בלתי תלויים ושווי התפלגות במסנן בעל תגובה להלם סופית (FIR). אל מול כל החסמים הנ"ל, חישוב מדויק של ה-Rényi The Entropy Power מסדר α של פלט המסנן נעשה עבור וקטורים גאוסיים; זה מאפשר לנו לראות עד כמה החסם שלנו טוב (קרוב לתוצאה המדויקת). בפרק 5, ישנו סיכום של תוצאות המחקר והכלים האנליטיים שבהם נעשה שימוש. דיון בנוגע להדיקות של החסמים שניתנים בתזה מוצג והצעות להמשך מחקר הקשורות בהכללת תוצאות נוספות ל-Rényi Entropy Power ובהרחבת ה-R-EPI לתחומים של $\alpha \in [0,1)$ ניתנות גם כן.

עבודה זו פורסמה לאחרונה כמאמר תחת הכותרת:

E. Ram and I. Sason, "On Rényi entropy power inequalities," IEEE Trans. on Information Theory, vol. 62, no. 12, pp. 6800-6815, December 2016,

והיא הוצגה חלקית בכנס ISIT16, בברצלונה, ספרד.