Transportation Proofs of Rényi Entropy Power Inequalities

Olivier Rioul
LTCI, Télécom ParisTech, 75013, Paris, France
Institut Polytechnique de Paris
Email: olivier.rioul@telecom-paristech.fr

Abstract—A framework for deriving Rényi entropy-power inequalities (EPIs) is presented that uses linearization and an inequality of Dembo, Cover, and Thomas. Simple arguments are given to recover the previously known Rényi EPIs and derive new ones, by unifying a multiplicative form with constant c and a modification with exponent α of previous works. An information-theoretic proof of the Dembo-Cover-Thomas inequality—equivalent to Young’s convolutional inequality with optimal constants—is provided, based on properties of Rényi differential entropy and from Gaussian densities. For log-concave densities, a transportation proof of a sharp varentropy bound is presented.

I. INTRODUCTION

Throughout this paper we consider n-dimensional zero-mean random vectors $X \in \mathbb{R}^n$ having densities. If $X \sim f$ has density $f \in L^r(\mathbb{R}^n)$ where $r > 0$ and $r \neq 1$, its Rényi entropy of exponent $r$ (or $r$-entropy) is

$$h_r(X) = \frac{1}{1-r} \log \int_{\mathbb{R}^n} f^r(x) \, dx = -r' \log \| f \|_{r'}$$

where $\| f \|_r$ denotes the $L^r$ norm of $f$, and $r' = \frac{r}{r-1}$ is the conjugate exponent of $r$, such that $\frac{1}{r'} + \frac{1}{r} = 1$. Notice two distinct situations: either $r > 1$ and $r' > 1$, or $0 < r < 1$ and $r' < 0$.

It is known that the limit as $r \to 1$ is the Shannon (differential) entropy $h_1(X) = h(X) = -\int_{\mathbb{R}^n} f(x) \log f(x) \, dx$. Letting $N(X) = \exp(2h(X)/n)$ be the corresponding entropy power, the famous entropy power inequality (EPI) can be written in the form

$$N \left( \sum_{i=1}^{m} X_i \right) \geq \sum_{i=1}^{m} N(X_i)$$

for any independent random vectors $X_1, X_2, \ldots, X_m \in \mathbb{R}^n$. The EPI dates back to Shannon’s seminal paper [1] and has a long history [2]. The link with the Rényi entropy $h_r(X)$ was first made by Dembo, Cover and Thomas [3] in connection with Young’s convolutional inequality with sharp constants, where Shannon’s EPI is obtained by letting exponents $r \to 1$ [4, Thm 17.8.3].

Recently, there has been increasing interest in Rényi entropy-power inequalities [5]. The Rényi entropy-power itself was first defined in [6]. We follow a slightly different definition [7] where, as in Shannon’s original definition [1], the Rényi entropy-power $N_r(X)$ equals (up to a multiplicative constant) the average power of a white Gaussian vector having the same Rényi entropy as $X$—hence the name “entropy power”.

If $X^* \sim \mathcal{N}(0, \sigma^2 I)$ is white Gaussian, an easy calculation yields

$$h_r(X^*) = \frac{n}{2} \log(2\pi \sigma^2) + \frac{n}{2} r' \log \frac{r}{r'}.$$  \hspace{1cm} \hspace{1cm} (3)

Since equating $h_r(X^*) = h_r(X)$ gives $\sigma^2 = \frac{e^{2h_r(X)/n}}{2\pi r'^2}$, we define the Rényi entropy power as

$$N_r(X) = e^{2h_r(X)/n}. \hspace{1cm} \hspace{1cm} (4)$$

Bobkov and Chistyakov [7] extended the classical Shannon’s EPI [2] to the Rényi entropy by incorporating a multiplicative constant $c > 0$ that depends on $r$:

$$N_r \left( \sum_{i=1}^{m} X_i \right) \geq c \sum_{i=1}^{m} N_r(X_i). \hspace{1cm} \hspace{1cm} (5)$$

Ram and Sason [8] improved (increased) the value of $c$ by making it depend also on the number $m$ of independent vectors $X_1, X_2, \ldots, X_m$.

Bobkov and Marsiglietti [9] proved another modification of the EPI for the Rényi entropy:

$$N_r^\alpha \left( \sum_{i=1}^{m} X_i \right) \geq \sum_{i=1}^{m} N_r^\alpha(X_i) \hspace{1cm} \hspace{1cm} (6)$$

with a power exponent parameter $\alpha > 0$ whose value was further improved (decreased) by Li [10].

All the above EPIs were found for Rényi entropies of orders $r > 1$. Recently, the $\alpha$-modification of the Rényi EPI [4] was extended to orders $<1$ for two independent variables having log-concave densities by Marsiglietti and Melbourne [11]. The starting point of all the above works was Young’s strengthened convolutional inequality.

In this paper, we build on the results of [12] that provides a comprehensive framework with simple proofs for Rényi EPIs of the general form

$$N_r^\alpha \left( \sum_{i=1}^{m} X_i \right) \geq c \sum_{i=1}^{m} N_r^\alpha(X_i) \hspace{1cm} \hspace{1cm} (7)$$

with constant $c > 0$ and exponent $\alpha > 0$. The framework uses only basic properties of Rényi entropies and is based on a transportation argument from normal densities and a change of variable by rotation, which had been previously used to give a simple proof of Shannon’s original EPI [13].

\footnote{Due to the non-increasing property of the $\alpha$-norm, if (6) holds for $\alpha$ it also holds for any $\alpha' > \alpha$.}
II. LINEARIZATION

The first step toward proving (7) is the following linearization lemma which generalizes [10] Lemma 2.1.

**Lemma 1.** For independent \( X_1, X_2, \ldots, X_m \), the Rényi EPI in the general form (7) is equivalent to the following inequality

\[
h_r \left( \sum_{i=1}^m \sqrt{\lambda_i} X_i \right) \geq \frac{m}{2} \log c + \left( \frac{1}{\alpha} - 1 \right) H(\lambda)
\]

for any distribution \( \lambda = (\lambda_1, \ldots, \lambda_m) \) of entropy \( H(\lambda) \).

**Proof.** First note the following scaling property \( h_r(aX) = h_r(X) + n \log |a| \) for any \( a \neq 0 \), easily established by a change of variable. It follows that the Rényi entropy power enjoys the same scaling property as for the usual power: \( N_r(aX) = a^2 N_r(X) \).

Suppose (7) holds. Then

\[
h_r \left( \sum_{i=1}^m \sqrt{\lambda_i} X_i \right) = \frac{n}{2\alpha} \log N_r \left( \sum_{i=1}^m \sqrt{\lambda_i} X_i \right)
\]

\[
\geq \frac{n}{2\alpha} \log \sum_{i=1}^m N_r \left( \sqrt{\lambda_i} X_i \right) + \frac{n}{2\alpha} \log c
\]

\[
= \frac{n}{2\alpha} \log \sum_{i=1}^m \lambda_i N_r \left( \sqrt{\lambda_i} X_i \right) + \frac{n}{2\alpha} \log c
\]

\[
\geq \frac{n}{2\alpha} \sum_{i=1}^m \lambda_i \log \left( \lambda_i^{-1} N_r \left( \sqrt{\lambda_i} X_i \right) \right) + \frac{n}{2\alpha} \log c
\]

\[
= \sum_{i=1}^m \lambda_i h_r \left( X_i \right) + \frac{n(\alpha - 1)}{2\alpha} \sum_{i=1}^m \lambda_i \log \lambda_i
\]

\[
+ \frac{n}{2\alpha} \log c
\]

which proves (8). The scaling property is used in (10) and the concavity of the logarithm is used in (11).

Conversely, suppose that (8) is satisfied for all \( \lambda > 0 \) such that \( \sum_{i=1}^m \lambda_i = 1 \). Set \( \lambda_i = N_r \left( X_i \right) / \sum_{i=1}^m N_r \left( X_i \right) \). Then

\[
N_r \left( \sum_{i=1}^m X_i \right) = \exp \left( \frac{n}{m} h_r \left( \sum_{i=1}^m \sqrt{\lambda_i} X_i / \sqrt{\lambda_i} \right) \right)
\]

\[
\geq \exp \left( \frac{2\alpha}{n} \sum_{i=1}^m \sqrt{\lambda_i} h_r \left( X_i / \sqrt{\lambda_i} \right) \right)
\]

\[
\times c \cdot e^{(1-\alpha) \sum_{i=1}^m \lambda_i \log \lambda_i}
\]

\[
= c \prod_{i=1}^m \left( N_r \left( \frac{X_i}{\lambda_i} \right) \lambda_i^{-1} \right)^{\lambda_i}
\]

\[
= c \prod_{i=1}^m \left( N_r \left( X_i \lambda_i^{-1} \right) \right)^{\lambda_i}
\]

\[
= c \left( \sum_{i=1}^m N_r \left( X_i \right) \right)^{\sum_{i=1}^m \lambda_i}
\]

\[
= c \sum_{i=1}^m N_r \left( X_i \right)
\]

which proves (7).

III. THE RÉNYI EPI OF DEMBO-COVER-THOMAS

As a second ingredient we have the following result, which was essentially established by Dembo, Cover and Thomas [3]. It is this Rényi version of the EPI which led them to prove Shannon’s original EPI by letting Rényi exponents \( \rightarrow 1 \).

**Theorem 1.** Let \( r_1, \ldots, r_m, r \) be exponents those conjugates \( r_1', \ldots, r_m', r' \) of the same sign and satisfy \( \sum_{i=1}^m r_i = \frac{1}{n} \) and let \( \lambda_1, \ldots, \lambda_m \) be the discrete probability distribution \( \lambda_i = \frac{r_i'}{r'} \) (\( i = 1, 2, \ldots, m \)). Then, for independent zero-mean \( X_1, X_2, \ldots, X_m \),

\[
\frac{m}{2} \log c + \left( \frac{1}{\alpha} - 1 \right) H(\lambda)
\]

\[\geq h_r \left( \sum_{i=1}^m \sqrt{\lambda_i} X_i \right) - \sum_{i=1}^m \lambda_i h_{r_i} \left( X_i \right)
\]

(19)

where \( X_1', X_2', \ldots, X_m' \) are i.i.d. standard Gaussian \( N(0, 1) \). Equality holds if and only if the \( X_i \) are i.i.d. Gaussian.

It is easily seen from the expression (3) of the Rényi entropy of a Gaussian that (19) is equivalent to

\[
\frac{m}{2} \log c + \left( \frac{1}{\alpha} - 1 \right) H(\lambda)
\]

\[
= h_r \left( \sum_{i=1}^m \sqrt{\lambda_i} X_i \right) - \sum_{i=1}^m \lambda_i h_{r_i} \left( X_i \right)
\]

\[
\geq h_r \left( \sum_{i=1}^m \sqrt{\lambda_i} X_i' \right) - \sum_{i=1}^m \lambda_i h_{r_i} \left( X_i' \right)
\]

where \( X_1', X_2', \ldots, X_m' \) are i.i.d. standard Gaussian \( N(0, 1) \). Equality holds if and only if the \( X_i \) are i.i.d. Gaussian.

IV. RÉNYI EPIs FOR ORDERS > 1

If \( r > 1 \), then \( r' > 0 \) and all \( r_i' \) are positive and greater than \( r' \). Therefore, all \( r_i \) are less than \( r \). Using the well-known fact that \( h_r(X) \) is non increasing in \( r \) (see also (34) below),

\[
h_r \left( X_i \right) \geq h_r \left( X_i \right) \quad (i = 1, 2, \ldots, m).
\]

(21)

Plugging this into (20), one obtains

\[
h_r \left( \sum_{i=1}^m \sqrt{\lambda_i} X_i \right) - \sum_{i=1}^m \lambda_i h_{r_i} \left( X_i \right) \geq \frac{n}{2} r' \left( \frac{\log r}{r} - \sum_{i=1}^m \log r_i \right)
\]

(22)

where \( \lambda_i = r' / r_i' \) for \( i = 1, 2, \ldots, m \). From Lemma 1 is suffices to establish that the r.h.s. of this inequality exceeds that of (8) to prove (7) for appropriate constants \( c \) and \( \alpha \).

For future reference define\( \footnote{The absolute value \( |r'| \) is needed in the next section where \( r' \) is negative.} \)

\[
A(\alpha) = |r'| \left( \frac{\log r}{r} - \sum_{i=1}^m \log r_i \right)
\]

\[
= \left| r' \right| \left( \sum_{i=1}^m \left( 1 - \frac{\lambda_i}{r} \right) \log (1 - \frac{\lambda_i}{r}) - (1 - \frac{1}{r}) \log (1 - \frac{1}{r}) \right).
\]

(23)
This function is strictly convex in $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ because $x \mapsto (1 - x/r') \log(1 - x/r')$ is strictly convex. Note that $A(\lambda)$ vanishes in the limiting cases where $\lambda$ tends to one of the standard unit vectors $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1)$ and since every $\lambda$ is a convex combination of these vectors and $A(\lambda)$ is strictly convex, one has $A(\lambda) < 0$.

Using the properties of $A(\lambda)$ it is immediate to recover known Rényi EPIs:

**Proposition 1** (Ram and Sason [8]). The Rényi EPI $\alpha$ holds for $r > 1$ and $c = r'^{\alpha}/(1 - \frac{1}{mr'})^{m-1}$. 

**Proof.** By Lemma 1 we only need to check that the r.h.s. of (22) is greater than $\frac{1}{\alpha} \log c$ for any $\alpha$ of the $\lambda_i$s, that is, for any choice of exponents $r_i$ such that $\sum_{i=1}^{m} \frac{1}{r_i} = \frac{1}{\alpha}$. Thus, (5) will hold for $c = \min_{\alpha} A(\lambda)$. Now, by the log-sum inequality [4 Thm 2.7.1],

$$\frac{1}{m} \sum_{i=1}^{m} \frac{1}{r_i} \log \frac{1}{r_i} \geq \frac{1}{m} \log \frac{1}{m} \sum_{i=1}^{m} \frac{1}{r_i} \frac{1}{r_i} = (m - 1/r') \log \left(\frac{m - 1}{m} \right)$$

(24)

with equality if and only if all $r_i$ are equal, that is, the $\lambda_i$ are equal to 1/m. Thus, $\min_{\alpha} A(\lambda) = r'^{\alpha}/(1 - \frac{1}{mr'}) + (m - 1/r') \log \left(\frac{m - 1}{m} \right) = \log c$. 

Note that $\log c = r'^{\alpha}/(1 - \frac{1}{mr'}) + (m - 1/r') \log \left(\frac{m - 1}{m} \right)$ decreases (and tends to $r'^{\alpha}/(1 - \frac{1}{mr'})$ as $m$ increases; in fact $\frac{\partial \log c}{\partial m} = r'^{\alpha}/(1 - \frac{1}{mr'}) + \frac{m - 1}{mr'} < r'^{\alpha}/(1 - \frac{1}{mr'}) + \frac{1}{m} = 0$. Thus, a universal constant independent of $m$ is obtained by taking

$$c = \inf_{m} \frac{r'^{\alpha}/(1 - \frac{1}{mr'})}{(m - 1/r') \log \left(\frac{m - 1}{m} \right)}$$

(26)

$$= \frac{r'^{\alpha}/(1 - \frac{1}{mr'})}{\inf_{m} (m - 1/r') \log \left(\frac{m - 1}{m} \right)}$$

(27)

$$= \frac{r'^{\alpha}/(1 - \frac{1}{mr'})}{e},$$

(28)

as was established by Bobkov and Chistyakov [7].

**Proposition 2** (Li [10]). The Rényi EPI $\alpha$ holds for $r > 1$ and $\alpha = \left[1 + r'^{\log \frac{1}{r'}} - (2r' - 1) \log_2(1 - \frac{1}{r'})\right]^{-1}$.

Li [10] remarked that this value of $\alpha$ is strictly smaller (better) than the value $\alpha = \frac{1}{r' - 1}$, obtained previously by Bobkov and Marsiglletti [9]. In [12] it is shown that it cannot be further improved in our framework by making it depend on $m$.

**Proof.** Since the announced $\alpha$ does not depend on $m$, we can always assume that $m = 2$. By Lemma 1 for $c = 1$, we only need to check that the r.h.s. of (22) is greater than $\frac{n}{2} (1/\alpha - 1) H(\lambda)$ for any choice of $\lambda_i$s, that is, for any choice of exponents $r_i$ such that $\sum_{i=1}^{m} \frac{1}{r_i} = \frac{1}{\alpha}$. Thus, (5) will hold for $\frac{1}{\alpha} - 1 = \min_{\alpha} A(\lambda)$. Li [10] showed—this is also easily proved using [11] Lemma 8—that the minimum is obtained when $\lambda = (1/2, 1/2)$. The corresponding value of $A(\lambda)/H(\lambda)$ is $\left[r'^{\log \frac{1}{r'}} + (2r' - 1) \log(1 - \frac{1}{r'})\right]/\log 2 = 1/\alpha - 1$. 

The above value of $\alpha$ is > 1. However, using the same method, it is easy to obtain Rényi EPIs with exponent values $\alpha < 1$. In this way we obtain a new Rényi EPI:

**Proposition 3.** The Rényi EPI $\alpha$ holds for $r > 1$, $0 < \alpha < 1$ with $c = \left[m r'^{\alpha}/(1 - \frac{1}{mr'})^{m-1}\right]^{1/\alpha}$. 

**Proof.** By Lemma 1 we only need to check that the r.h.s. of Equation (22) is greater than $\frac{1}{\alpha} (\log c)/\alpha + (1/\alpha - 1) H(\lambda)$, that is, 

$$A(\lambda) = \left[(\log c)/\alpha + (1/\alpha - 1) H(\lambda)\right]$$

for any choice of $\lambda_i$s, that is, for any choice of exponents $r_i$ such that $\sum_{i=1}^{m} \frac{1}{r_i} = 1/\alpha$. Thus, for a given $0 < \alpha < 1$, (7) will hold for $\log c = \min_{\alpha} A(\lambda) - (1/\alpha - 1) H(\lambda)$. From the preceding proofs (since both $A(\lambda)$ and $H(\lambda)$ are convex functions of $\lambda$), the minimum is attained when all $\lambda_i$s are equal. This gives $c = \alpha^{r'^{\log \frac{1}{r'}} + (mr' - 1) \log(1 - \frac{1}{mr'})} \cdot (1 - \alpha) \log m$. 

V. RÉNYI EPIs FOR ORDERS $\alpha < 1$ AND LOG-CONCAVE DENSITIES

If $r < 1$, then $r' < 0$ and all $r'_i$ are negative and $< r'$. Therefore, all $r_i$ are $> r$. Now the opposite inequality of (21) holds and the method of the preceding section fails. For log-concave densities, however, (21) can be replaced by a similar inequality in the right direction.

A density $f$ is log-concave if $\log f$ is concave in its support, i.e., for all $0 < \mu < 1$,

$$f(x)^{\mu} f(y)^{1-\mu} \leq f(\mu x + (1-\mu)y).$$

(29)

**Theorem 2** (Fradelizi, Madiman and Wang [14]). If $X$ has a log-concave density, then $h_r(rX - rX) = (1 - r)h_r(X) + n \log r$ is concave in $r$.

This concavity property is used in [14] to derive a sharp “varentropy bound”. Section VIII provides an alternate transportation proof along the same lines as in Section VII.

By Theorem 2 since $n \log r + (1 - r)h_r(X)$ is concave and vanishes for $r = 1$, the slopes $n \log r + (1 - r) h_r(X) = 0$ are nonincreasing in $r$. In other words, $h_r(X) + n \log r$ is nondecreasing. Now since all $r_i$ are $> r$,

$$h_r(X) + n \log \left(\frac{n}{r} \right) \leq h_r(X) + \frac{n \log r}{1 - r}$$

(30)

Plugging this into (29), one obtains

$$h_r \left[ \sum_{i=1}^{m} \frac{1}{r_i} X_i \right] \leq - \sum_{i=1}^{m} \lambda_i h_r(X_i)$$

$$\geq n \left(1 - r \right) h_r(X) - \sum_{i=1}^{m} \lambda_i r_i \log \frac{r - \sum_{i=1}^{m} \lambda_i r_i}{r_i}$$

$$= \frac{n}{2} \left(1 - r \right) \sum_{i=1}^{m} \frac{r_i \log r_i}{r_i} - \frac{r \log r}{r}$$

(31)

where we have used that $\lambda_i = r'^{r'}$ for $i = 1, 2, \ldots, m$. Notice that, quite surprisingly, the r.h.s. of (31) for $r < 1$ ($r' < 0$) is the opposite of that of (22) for $r > 1$ ($r' > 0$). However, since $r'$ is now negative, the r.h.s. is exactly equal to $\frac{1}{2} A(\lambda)$ which is still convex and negative. For this reason, the
proofs of the following theorems for \( r < 1 \) are such repeats of the theorems obtained previously for \( r > 1 \).

**Proposition 4.** The Rényi EPI (5) for log-concave densities holds for \( c = r^{-r/r} (1 - \frac{1}{m})^{1-\alpha/m} \) and \( r < 1 \).

**Proof.** Identical to that of Theorem 1 except for the change \( |r'| = -r' \) in the expression of \( A(\lambda) \).

**Proposition 5.** (Marsiglietti and Melbourne [11]). The Rényi EPI (6) log-concave densities holds for \( \alpha = [1 + |r'\log_x r' + (2|r'| + 1)\log_2(1 + \frac{1}{2|r'|})]^{-1} \) and \( r < 1 \).

**Proof.** Identical to that of Theorem 2 except for the change \( |r'| = -r' \) in the expression of \( A(\lambda) \).

**Proposition 6.** The Rényi EPI (7) for log-concave densities holds for \( c = [mr^{-r/r} (1 - \frac{1}{m})]^{\alpha/m} \) where \( r < 1 \) and \( 0 < \alpha < 1 \).

**Proof.** It is identical to that of Theorem 3 except for the change \( |r'| = -r' \) in the expression of \( A(\lambda) \).

**VI. RELATIVE AND CONDITIONAL RÉNYI ENTROPIES**

Before turning to transportation proofs of Theorems 1 and 2 it is convenient to review some definitions and properties. The following notions were previously used for discrete variables, but can be easily adapted to variables with densities.

**Definition 1 (Escort Variable [15]).** If \( f \in L^r(\mathbb{R}^n) \), its escort density of exponent \( r \) is defined by

\[
\tilde{f}_r(x) = \frac{f^r(x)}{\int_{\mathbb{R}^n} f^r(x) \, dx}.
\]

Let \( X_r \sim f_r \) denote the corresponding escort random variable.

We mention, in passing, the following identities.

**Proposition 7.** Let \( r \neq 1 \) and assume that \( X \sim f \in L^r(\mathbb{R}^n) \) for all \( s \) in a neighborhood of \( r \). Then

\[
\frac{\partial}{\partial r} ((1-r)h_r(X)) = \mathbb{E} \log f(X_r) = -h(X_r|X)
\]

\[
\frac{\partial}{\partial r} h_r(X) = -\frac{1}{(1-r)^2} D(X_r|X)
\]

\[
\frac{\partial^2}{\partial r^2} ((1-r)h_r(X)) = \text{Var} \log f(X_r).
\]

where \( h(X|Y) = \int f \log(1/g) \) denotes cross-entropy and \( D(X||Y) = \int f \log(f/g) \) is the Kullback-Leibler divergence.

**Proof.** By the hypothesis, one can differentiate under the integral sign. It is easily seen that

\[
\frac{\partial}{\partial r} ((1-r)h_r(X)) = f_r \log f = f_r \log f.
\]

Taking another derivative yields

\[
\frac{\partial}{\partial r} f_r \log f = f_r \log(f^r) - (\int f_r \log(f))^2.\]

Since \( \frac{\partial}{\partial r} ((1-r)h_r(X)) = (1-r) \frac{\partial}{\partial r} h_r(X) - h_r(X) \) we have \( (1-r)^2 \frac{\partial}{\partial r} h_r(X) = \int f_r \log(f/f^r) + \log f^r = \int f_r \log(f/f_r) \).

Notice that the derivative is \( \leq 0 \) in (34), which gives a new proof that \( h_r(X) \) is nonincreasing in \( r \). It is strictly decreasing if \( X_r \) is not distributed as \( X \), that is, if \( X \) is not uniformly distributed. Equation (35) shows that \( (1-r)h_r(X) \) is convex in \( r \), that is, \( \int f^r \) is log-convex in \( r \) (which is essentially equivalent to Hölder’s inequality).

**Definition 2 (Relative Rényi Entropy [16]).** Given \( X \sim f \) and \( Y \sim g \), their relative Rényi entropy of exponent \( r \) (relative \( r \)-entropy) is given by

\[
\Delta_r(X||Y) = D_r(X||Y)_r
\]

where \( D_r(X||Y) = \frac{1}{1-r} \int f^r g^{1-r} \) denotes the Rényi \( r \)-divergence [17].

When \( r \to 1 \) both the relative \( r \)-entropy and the \( r \)-divergence tend to the Kullback-Leibler divergence \( D(X||Y) = \Delta(X||Y) \) (also known as the relative entropy). For \( r \neq 1 \) the two notions do not coincide. It is easily checked from the definitions that

\[
\Delta_r(X||Y) = -r' \log \int f_r' g^{1-r'} = -r' \log \mathbb{E}(g_r^{1/r'}(X)) - h_r(X) \quad (37)
\]

and

\[
h_r(X) = -r' \log \mathbb{E}(f_{r'}/r')(X) \quad (38)
\]

Thus, just like for the case \( r = 1 \), the relative \( r \)-entropy (37) is the difference between the expression of the \( r \)-entropy (38) in which \( f \) is replaced by \( g \), and the \( r \)-entropy itself.

Since the Rényi divergence \( D_r(X||Y) = \frac{1}{1-r} \int f^r g^{1-r} \) is nonnegative and vanishes if and only if the two distributions \( f \) and \( g \) coincide, the relative entropy \( \Delta_r(X||Y) \) enjoys the same property. From (37) we have the following

**Proposition 8 (Rényi-Gibbs’ inequality).** If \( X \sim f \) and \( g \) is any density,

\[
h_r(X) \leq -r' \log \mathbb{E}(g_r^{1/r'}(X))
\]

with equality if and only if \( f = g \) a.e.

Letting \( r \to 1 \) one recovers the classical Gibbs’ inequality.

**Definition 3 (Arimoto’s Conditional Rényi Entropy [19]).**

\[
h_r(X|Z) = -r' \log \mathbb{E}[f(\cdot|Z)]_r = -r' \log \mathbb{E}f_r^{r'/r}(X|Z)
\]

**Proposition 8** applied to \( f(x|z) \) and \( g(x|z) \) gives the inequality \( h_r(X|Z = z) \leq -r' \log \mathbb{E}(g_r^{1/r'}(X|Z = z)) \) which, averaged over \( Z \), yields the following conditional Rényi-Gibbs’ inequality

\[
h_r(X|Z) \leq -r' \log \mathbb{E}(g_r^{1/r'}(X|Z)) \quad (41)
\]

If in particular we put \( g(x|z) = f(x) \) independent of \( z \), the r.h.s. becomes equal to (38). We have thus obtained a simple proof of the following

**Proposition 9** (Conditioning reduces \( r \)-entropy [19]).

\[
h_r(X|Z) \leq h_r(X) \quad (42)
\]

with equality if and only if \( X \) and \( Z \) are independent.

Another important property is the data processing inequality for Rényi divergence [17] which implies \( D_r(T(X)||T(Y)) \leq D_r(X||Y) \) for any transformation \( T \). The same holds for
relative $r$-entropy except that the transformation is applied to escort variables:

**Proposition 10** (Data processing inequality for relative $r$-entropy). If $X^*, Y^*, X, Y$ are random vectors such that

\[ X_r = T(X^*_r) \text{ and } Y_r = T(Y^*_r), \]

then $D(X\|Y) \leq D(X^*\|Y^*)$.

**Proof.** $D(X\|Y) = D_{\frac{1}{r}}(X_r\|Y_r) = D_{\frac{1}{r}}(T(X^*_r)||T(Y^*_r)) \leq D_{\frac{1}{r}}(X^*_r||Y^*_r) = D(X^*\|Y^*)$. \(\square\)

When $T$ is invertible, inequalities in both directions hold:

**Proposition 11** (Relative $r$-entropy preserves transport). Let $T$ be an (invertible) transport satisfying (43). Then $D(X\|Y) = D(X^*\|Y^*)$.

From (37) the equality $D(X\|Y) = D(X^*\|Y^*)$ can be rewritten as the following identity:

\[ -r' \log \mathbb{E}(\frac{1}{r'}(X)) - h_r(X) = -r' \log \mathbb{E}(\frac{1}{r'}(X^*)) - h_r(X^*). \]

Assuming $T$ is a diffeomorphism, the density $g_{T^*}$ of $X_r^*$ is given by the change of variable formula $g_{T^*}(u) = g_r(T(u))|T'(u)|$ where the Jacobian $|T'(u)|$ is the absolute value of the determinant of the Jacobian matrix $T'(u)$. In this case (44) can be rewritten as

\[ -r' \log \mathbb{E}(\frac{1}{r'}(X)) - h_r(X) = -r' \log \mathbb{E}(\frac{1}{r'}(T(X^*)))|T'(X^*)|^{\frac{1}{r'}} - h_r(X^*). \]

which is valid for any $g_r$.

**VII. A TRANSPORTATION PROOF OF THEOREM 1**

We proceed to prove (19). It is easily seen, using finite induction on $m$, that it suffices to prove the corresponding inequality for $m = 2$ arguments:

\[ h_r(\sqrt{X+\sqrt{1-\lambda}Y}) - \lambda h_p(X) - (1-\lambda)h_q(Y) \geq h_r(\sqrt{X} + \sqrt{1-\lambda}Y^*) - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*) \]

with equality if and only if $X, Y$ are i.i.d. Gaussian. Here $X^*$ and $Y^*$ are i.i.d. standard Gaussian $\mathcal{N}(0, 1)$ and the triple $(p, q, r)$ and its associated $\lambda \in (0, 1)$ satisfy the following conditions: $p, q, r$ have conjugates $p', q', r'$ of the same sign which satisfy

\[ \frac{1}{p'} + \frac{1}{q'} = \frac{1}{r'} \]

(that is, $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$) and

\[ \lambda = \frac{r'}{p'} = 1 - \frac{r'}{q'}. \]

**Lemma 2** (Transport from Gaussian). Let $f$ be given and $X^* \sim \mathcal{N}(0, \sigma^2 I)$. There exists a diffeomorphism $T : \mathbb{R}^n \to \mathbb{R}^n$ with log-concave Jacobian $|T'|$ such that $X = T(X^*) \sim f$ in words $T$ transports $X^*$ to $X$. The log-concavity property states that for any two such transports $T, U$ and $\lambda \in (0, 1)$, we have the inequality

\[ |T'(X^*)|^{\lambda} |U'(Y^*)|^{1-\lambda} \leq |\lambda T'(X^*) + (1-\lambda)U'(Y^*)|. \]

The proof of Lemma 2 is very simple for one-dimensional variables [20], where $T$ is just an increasing function with continuous derivative $T' > 0$ and where (49) is the classical inequality of arithmetic and geometric means.

For dimensions $n > 1$, Lemma 2 comes into two flavours:

- $T$ can be chosen such that its Jacobian matrix $T'$ is (lower) triangular with positive diagonal elements. This is known in optimal transport theory as the Knöthe–Rosenblatt map [21], [22]. Two different elementary proofs are given in [13]. Inequality (49) results from the concavity of the logarithm applied to the diagonal elements of the Jacobian matrices.

- $T$ can be chosen such that its Jacobian matrix $T'$ is symmetric positive definite. This is known in optimal transport theory as the Brenier-McCann map [23], [24]. In this case (49) is Ky Fan’s inequality [4] § 7.9.1.

The key argument is now the following. Considering escort variables, by transport (Lemma 2), one can write

\[ X_p = T(X^*_p) \]

\[ Y_q = U(Y^*_q) \]

for two diffeomorphisms $T$ and $U$ satisfying (49). Then by transport preservation (Proposition 11), we have $\lambda D_p(X\|U) + (1-\lambda)D_p(Y\|Y) = \lambda D_p(X^*\|U^*) + (1-\lambda)D_p(Y^*\|Y^*)$ for any $U \sim \varphi$ and $V \sim \psi$, which from (45) can be easily rewritten in the form

\[ -r' \log \mathbb{E}(\chi(\varphi(X, Y))) - \lambda h_p(X) - (1-\lambda)h_q(Y) \]

\[ = -r' \log \mathbb{E}(\chi(T(X^*, U(Y^*)))|T'(X^*)|^{\lambda} |U'(Y^*)|^{1-\lambda}) \]

\[ \lambda h_p(X^*) - (1-\lambda)h_q(Y^*) \]

where we have noted $\chi(x, y) = \varphi_x^\lambda(x) \psi_y^{1-\lambda}(y)$. Such an identity holds, by the change of variable $x = T(x^*), y = U(y^*)$, for any function $\chi(x, y)$ of $x$ and $y$.

Now we have $h_r(\sqrt{X+\sqrt{1-\lambda}Y}) - \lambda h_p(X) - (1-\lambda)h_q(Y)$.

\[ = -r' \log \mathbb{E}(\chi(T(X^*, U(Y^*)))|T'(X^*)|^{\lambda} |U'(Y^*)|^{1-\lambda}) \]

\[ \lambda h_p(X^*) - (1-\lambda)h_q(Y^*) \]

where we have noted $\chi(x, y) = \varphi_x^\lambda(x) \psi_y^{1-\lambda}(y)$. Such an identity holds, by the change of variable $x = T(x^*), y = U(y^*)$, for any function $\chi(x, y)$ of $x$ and $y$.
where \( \varphi(x^*, y^*) = \theta_r(\sqrt{\lambda} T(x^*) + \sqrt{1 - \lambda} U(y^*)) \cdot |\lambda T'(x^*) + (1 - \lambda) U'(y^*)| \).

To conclude we need the following

**Lemma 3** (Normal Rotation [13]). If \( X^*, Y^* \) are i.i.d. Gaussian, then for any \( 0 < \lambda < 1 \), the rotation

\[
\begin{align*}
\tilde{X} &= \sqrt{\lambda} X^* + \sqrt{1 - \lambda} Y^* \\
\tilde{Y} &= -\sqrt{1 - \lambda} X^* + \sqrt{\lambda} Y^*
\end{align*}
\]

(56)
yields i.i.d. Gaussian variables \( \tilde{X}, \tilde{Y} \).

Lemma 3 is easy proved considering covariance matrices. A deeper result (Bernstein’s lemma, not used here) states that this property of remaining i.i.d. by rotation characterizes the Gaussian distribution [20] Lemma 4, [25] Chap. 5).

Since the starred variables can be expressed in terms of the tilde variables by the inverse rotation

\[
\begin{align*}
X^* &= \sqrt{\lambda} \tilde{X} - \sqrt{1 - \lambda} \tilde{Y} \\
Y^* &= \sqrt{1 - \lambda} \tilde{X} + \sqrt{\lambda} \tilde{Y},
\end{align*}
\]

(57)
the inequality (55) can be written as

\[
\begin{align*}
h_r(\sqrt{\lambda} X + \sqrt{1 - \lambda} Y) - \lambda h_p(X) - (1 - \lambda) h_q(Y) \\
&\geq -r' \log \mathbb{E}(\psi^{1/r'}(\tilde{X}|\tilde{Y})) - \lambda h_p(X^*) - (1 - \lambda) h_q(Y^*),
\end{align*}
\]

where \( \psi(x|y) = \theta_r(\sqrt{\lambda} x - \sqrt{1 - \lambda} y) + \sqrt{1 - \lambda} U(\sqrt{1 - \lambda} x + \sqrt{\lambda} y) \).

Making the change of variable \( z = \sqrt{\lambda} x - \sqrt{1 - \lambda} y + \sqrt{1 - \lambda} U(\sqrt{1 - \lambda} x + \sqrt{\lambda} y) \), \( \mathbb{E}(\psi^{1/r'}(\tilde{X}|\tilde{Y})) \) is easily checked that \( \int \psi(x|y) d\tilde{X} = \int \psi(x|y) d\tilde{X} \), since \( \tilde{X} \) is a conditional density, and by the conditional Rényi-Gibbs’ inequality [41],

\[
- r' \log \mathbb{E}(\psi^{1/r'}(\tilde{X}|\tilde{Y})) \geq h(\tilde{X}|\tilde{Y})
\]

(59)
where \( h_r(\tilde{X}|\tilde{Y}) = h_r(\tilde{X}) = h_r(\sqrt{\lambda} X^* + \sqrt{1 - \lambda} Y^*) \) since \( \tilde{X} \) and \( \tilde{Y} \) are independent. Combining with (58) yields the announced inequality (46).

It remains to settle the equality case in (46). From the above proof, equality holds in (46) if and only if both (49) and (59) are equalities. The rest of the argument depends on whether Knöthe or Brenier maps are used.

a) Knöthe maps: In the case of Knöthe maps, Jacobian matrices are triangular and equality in (49) holds if and only if for all \( i = 1, 2, \ldots, n \),

\[
\frac{\partial T_i}{\partial x_i}(x^*) = \frac{\partial U_i}{\partial y_i}(Y^*) \quad \text{a.s.}
\]

(60)
Since \( X^* \) and \( Y^* \) are independent Gaussian variables, this implies that \( \frac{\partial T_i}{\partial x_i} \) and \( \frac{\partial U_i}{\partial y_i} \) are constant and equal. In particular the Jacobian \( |\lambda T'(\sqrt{\lambda} x - \sqrt{1 - \lambda} y) + (1 - \lambda) U'(\sqrt{1 - \lambda} x + \sqrt{\lambda} y)| \) is constant. Now since \( h_r(\tilde{X}|\tilde{Y}) \) equality in (59) holds only if \( \psi(x|y) \) does not depend on \( \tilde{y} \), which implies that \( \sqrt{\lambda} T'(\sqrt{\lambda} x - \sqrt{1 - \lambda} y) + \sqrt{1 - \lambda} U'(\sqrt{1 - \lambda} x + \sqrt{\lambda} y) \) does not depend on the value of \( y \). Taking derivatives with respect to \( y_j \) for all \( j = 1, 2, \ldots, n \), we have \( -\sqrt{\lambda} \sqrt{1 - \lambda} \frac{\partial T}{\partial x_j}(\sqrt{\lambda} X - \sqrt{1 - \lambda} Y) + \sqrt{\lambda} \sqrt{1 - \lambda} \frac{\partial U}{\partial y_j}(\sqrt{\lambda} X + \sqrt{1 - \lambda} Y) = 0 \) which implies \( \frac{\partial T_i}{\partial x_i}(x^*) = \frac{\partial U_i}{\partial y_i}(Y^*) \) a.s. for all \( i, j = 1, 2, \ldots, n \). In other words, \( T'(x^*) = U'(Y^*) \) a.s.

b) Brenier maps: In the case of Brenier maps the argument is simpler. Jacobian matrices are symmetric positive definite and by strict concavity, Ky Fan’s inequality (49) is an equality only if \( T'(x^*) = U'(Y^*) \) a.s.

In both cases, since \( X^* \) and \( Y^* \) are independent, this implies that \( T'(x^*) = U'(Y^*) \) is constant. Therefore, \( T \) and \( U \) are linear transformations, equal up to an additive constant (= 0 since the random vectors are assumed of zero mean). It follows that \( X_p = T(X^*_p) \) and \( Y_q = U(Y^*_q) \) are Gaussian with respective distributions \( X_p \sim N'(0, K/p) \) and \( Y_q \sim N'(0, K/q) \). Hence, \( X \) and \( Y \) are i.i.d. Gaussian \( N'(0, K) \).

This ends the proof of Theorem 2.

We note that this section has provided an information-theoretic proof the strengthened Young’s convolutional inequality (with optimal constants), since (46) is a rewriting of this convolutional inequality [3].

**VIII. A Transportation Proof of Theorem 2**

Define \( r = \lambda p + (1 - \lambda) q \) where \( 0 < \lambda < 1 \). It is required to show that \( (1 - r) h_p(X) + n \log r \geq \lambda (1 - p) h_p(X) + n \log p + (1 - \lambda)((1 - q) h_q(X) + n \log q) \).

By Lemma 2 there exists two diffeomorphisms \( T, U \) such that one can write \( p X_p = T(X^*_p) \) and \( q X_q = U(X^*_q) \). Then, by these changes of variables \( X^* \) has density

\[
\frac{1}{p} f_p \left( \frac{T(x^*)}{p} \right) |T'(x^*)| = \frac{1}{p} f_q \left( \frac{U(x^*)}{q} \right) |U'(x^*)|
\]

(61)
which can be written

\[
\frac{f_p \left( \frac{T(x^*)}{p} \right) |T'(x^*)|}{\exp((1 - p) h_p(X) + n \log p)} = \frac{f_q \left( \frac{U(x^*)}{q} \right) |U'(x^*)|}{\exp((1 - q) h_q(X) + n \log q)}
\]

(62)
Taking the geometric mean, integrating over \( x^* \) and taking the logarithm gives the representation

\[
\lambda (1 - p) h_p(X) + n \log p + (1 - \lambda)((1 - q) h_q(X) + n \log q) \geq \log \int f_p \left( \frac{T(x^*)}{p} \right) f_q \left( \frac{U(x^*)}{q} \right) |T'(x^*)| |U'(x^*)| dx^*.
\]

(63)
Now, by log-concavity [29] (with \( \mu = \lambda p/r \)) and (49),

\[
\lambda (1 - p) h_p(X) + n \log p + (1 - \lambda)((1 - q) h_q(X) + n \log q) \leq \log \left( \int r^n f^r \left( \frac{T(x^*) + (1 - \lambda) U(x^*)}{r} \right) |T'(x^*)| + (1 - \lambda) U'(x^*)| dx^* \right)
\]

(64)
This ends the proof of Theorem 2.

This theorem asserts that the second derivative \( \frac{\partial^2}{\partial x_i^2} \) of \( (1 - r) h_p(X) + n \log r \) \( \leq 0 \). From (35) this gives Var \( \log f(X_r) \leq n/r^2 \), that is, Var \( \log f(X_r) \leq n \). Setting \( r = 1 \), this is the varentropy bound Var \( \log f(X) \leq n \) of [14].
REFERENCES