Rényi Entropy Power and Normal Transport

Olivier Rioul
LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120, Palaiseau, France

***Abstract*—A framework for deriving Rényi entropy-power inequalities (REPIs) 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 conditional and relative entropies and using transportation arguments from Gaussian densities. For log-concave densities, a transportation proof of a sharp varentropy bound is presented.**

This work was partially presented at the 2019 Information Theory and Applications Workshop, San Diego, CA.

I. Introduction

We consider the r-entropy (Rényi entropy of exponent r, where $r > 0$ and $r \neq 1$) of a n-dimensional zero-mean random vector $X \in \mathbb{R}^n$ having density $f \in L^r(\mathbb{R}^n)$:

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

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 that either $r > 1$ and $r' > 1$, or $0 < r < 1$ and $r' < 0$. The limit as $r \to 1$ is the classical $h_1(X) = h(X) = -\int_{\mathbb{R}^n} f(x) \log f(x)\,\mathrm{d}x$. Letting $N(X) = \exp\bigl(2h(X)/n\bigr)$ be the corresponding entropy power [1], the famous entropy power inequality (EPI) [1], [2] writes $N\Bigl(\sum_{i=1}^m X_i\Bigr) \geq \sum_{i=1}^m N(X_i)$ for any independent random vectors $X_1, X_2, \ldots, X_m \in \mathbb{R}^n$. The link with the Rényi entropy $h_r(X)$ was first made in [3] in connection with a strengthened Young's convolutional inequality, where the EPI is obtained by letting exponents tend to 1 [4, Thm 17.8.3].

Recently, there has been increasing interest in Rényi entropy-power inequalities [5]. The Rényi entropy-power $N_r(X)$ is defined [6] as the average power of a white Gaussian vector having the same Rényi entropy as X. If $X^* \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ is white Gaussian, an easy calculation yields

$$h_r(X^*) = \tfrac{n}{2} \log(2\pi\sigma^2) + \tfrac{n}{2} r' \tfrac{\log r}{r}. \qquad (2)$$

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

Bobkov and Chistyakov [6] extended the classical EPI to the r-entropy by incorporating a r-dependent constant $c > 0$:

$$N_r\Bigl(\sum_{i=1}^m X_i\Bigr) \geq c \sum_{i=1}^m N_r(X_i). \qquad (3)$$

Ram and Sason [7] 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 [8] proved another modification of the EPI for the Rényi entropy:

$$N_r^{\alpha}\Bigl(\sum_{i=1}^m X_i\Bigr) \geq \sum_{i=1}^m N_r^{\alpha}(X_i) \qquad (4)$$

with a power exponent parameter $\alpha > 0$. Due to the non-increasing property of the α-norm, if (4) holds for α it also holds for any $\alpha' > \alpha$. The value of α was further improved (decreased) by Li [9]. All the above EPIs were found for Rényi entropies of orders $r >$1. Recently, the α-modification of the Rényi EPI (4) was extended to orders $<$1 for two independent variables having log-concave densities by Marsiglietti and Melbourne [10]. The starting point of all the above works was Young's strengthened convolutional inequality.

In this paper, we build on the results of [11] to provide simple proofs for Rényi EPIs of the general form

$$N_r^{\alpha}\Bigl(\sum_{i=1}^m X_i\Bigr) \geq c \sum_{i=1}^m N_r^{\alpha}(X_i) \qquad (5)$$

with constant $c > 0$ and exponent $\alpha > 0$. The present 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 was previously used to give a simple proof of Shannon's original EPI [12].

II. Linearization

The first step toward proving (5) is the following linearization lemma which generalizes [9, Lemma 2.1].

Lemma 1. *For independent $X_1, X_2, \ldots, X_m$, the Rényi EPI in the general form (5) is equivalent to the following inequality*

$$h_r\bigl(\sum_{i=1}^m \sqrt{\lambda_i} X_i\bigr) - \sum_{i=1}^m \lambda_i h_r(X_i) \geq \tfrac{n}{2}\bigl(\tfrac{\log c}{\alpha} + \bigl(\tfrac{1}{\alpha} - 1\bigr) H(\lambda)\bigr)$$

for any distribution $\lambda = (\lambda_1, \ldots, \lambda_m)$ of entropy $H(\lambda)$. (6)

Proof. Note the scaling property $h_r(aX) = h_r(X) + n \log|a|$ for any $a \neq 0$, established by a change of variable. It follows that $N_r(aX) = a^2 N_r(X)$. Now first suppose (5) holds. Then

$$\begin{aligned}
h_r\bigl(\textstyle\sum_{i=1}^m \sqrt{\lambda_i} X_i\bigr) &= \tfrac{n}{2\alpha} \log N_r^{\alpha}\bigl(\textstyle\sum_{i=1}^m \sqrt{\lambda_i} X_i\bigr) && (7)\\
&\geq \tfrac{n}{2\alpha} \log \textstyle\sum_{i=1}^m N_r^{\alpha}(\sqrt{\lambda_i} X_i) + \tfrac{n}{2\alpha} \log c\\
&= \tfrac{n}{2\alpha} \log \textstyle\sum_{i=1}^m \lambda_i^{\alpha} N_r^{\alpha}(X_i) + \tfrac{n}{2\alpha} \log c && (8)\\
&\geq \tfrac{n}{2\alpha} \textstyle\sum_{i=1}^m \lambda_i \log\bigl(\lambda_i^{\alpha-1} N_r^{\alpha}(X_i)\bigr) + \tfrac{n}{2\alpha} \log c && (9)\\
&= \textstyle\sum_{i=1}^m \lambda_i h_r(X_i) + \tfrac{n(\alpha-1)}{2\alpha} \textstyle\sum_{i=1}^m \lambda_i \log \lambda_i + \tfrac{n}{2\alpha} \log c
\end{aligned}$$

which proves (6). The scaling property is used in (8) and the concavity of the logarithm is used in (9).

Conversely, suppose that (6) is satisfied for all $\lambda_i > 0$ such that $\sum_{i=1}^m \lambda_i = 1$. Set $\lambda_i = N_r^{\alpha}(X_i) / \sum_{i=1}^m N_r^{\alpha}(X_i)$. Then

$$\begin{aligned}
N_r^{\alpha}\bigl(\textstyle\sum_{i=1}^m X_i\bigr) &= \exp \tfrac{2\alpha}{n} h_r\bigl(\textstyle\sum_{i=1}^m \sqrt{\lambda_i} \tfrac{X_i}{\sqrt{\lambda_i}}\bigr)\\
&\geq \exp \tfrac{2\alpha}{n} \textstyle\sum_{i=1}^m \lambda_i h_r\Bigl(\tfrac{X_i}{\sqrt{\lambda_i}}\Bigr) \times c \cdot e^{(1-\alpha) \sum_{i=1}^m \lambda_i \log \frac{1}{\lambda_i}}\\
&= c \prod_{i=1}^m \Bigl(N_r^{\alpha}\Bigl(\tfrac{X_i}{\sqrt{\lambda_i}}\Bigr) \lambda_i^{\alpha-1}\Bigr)^{\lambda_i} = c \prod_{i=1}^m \Bigl(N_r^{\alpha}(X_i) \lambda_i^{-1}\Bigr)^{\lambda_i}\\
&= c\bigl(\textstyle\sum_{i=1}^m N_r^{\alpha}(X_i)\bigr)^{\sum_{i=1}^m \lambda_i} = c \textstyle\sum_{i=1}^m N_r^{\alpha}(X_i)
\end{aligned}$$

which proves (5). □

III. The REPI 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 $\to 1$.

Theorem 1. *Let $r_1, \ldots, r_m, r$ be exponents those conjugates $r'_1, \ldots, r'_m, r'$ are of the same sign and satisfy $\sum_{i=1}^m \frac{1}{r'_i} = \frac{1}{r'}$ and let $\lambda_1, \ldots, \lambda_m$ be the discrete probability distribution $\lambda_i = \frac{r'}{r'_i}$. Then, for independent zero-mean $X_1, X_2, \ldots, X_m$,*

$$\begin{aligned} h_r\Bigl(\sum_{i=1}^m \sqrt{\lambda_i} X_i\Bigr) - \sum_{i=1}^m \lambda_i h_{r_i}(X_i) \\ \geq h_r\Bigl(\sum_{i=1}^m \sqrt{\lambda_i} X_i^*\Bigr) - \sum_{i=1}^m \lambda_i h_{r_i}(X_i^*) \end{aligned} \tag{10}$$

where $X_1^*, X_2^*, \ldots, X_m^*$ are i.i.d. standard Gaussian $\mathcal{N}(0, \mathbf{I})$. Equality holds if and only if the X_i are i.i.d. Gaussian.

It is easily seen from the expression (2) of the Rényi entropy of a Gaussian that (10) is equivalent to

$$h_r\Bigl(\sum_{i=1}^m \sqrt{\lambda_i} X_i\Bigr) - \sum_{i=1}^m \lambda_i h_{r_i}(X_i) \geq \tfrac{n}{2} r' \Bigl(\tfrac{\log r}{r} - \sum_{i=1}^m \tfrac{\log r_i}{r_i}\Bigr). \tag{11}$$

Note that the l.h.s. is very similar to that of (6) except that different Rényi exponents are present. This will be the crucial step toward proving (5).

Theorem 1 (for $m = 2$) was derived in [3] as a rewriting of Young's strengthened convolutional inequality with optimal constants. Section VII provides a simple transportation proof, which uses only basic properties of Rényi entropies.

IV. REPIs 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 (22) below),

$$h_{r_i}(X_i) \geq h_r(X_i) \qquad (i = 1, 2, \ldots, m). \tag{12}$$

Plugging this into (11), one obtains

$$h_r\bigl(\sum_{i=1}^m \sqrt{\lambda_i} X_i\bigr) - \sum_{i=1}^m \lambda_i h_r(X_i) \geq \tfrac{n}{2} r' \bigl(\tfrac{\log r}{r} - \textstyle\sum_{i=1}^m \tfrac{\log r_i}{r_i}\bigr) \tag{13}$$

where $\lambda_i = r'/r'_i$. From Lemma 1 it suffices to establish that the r.h.s. of this inequality exceeds that of (6) to prove (5) for appropriate constants c and α. For future reference define

$$\begin{aligned} A(\lambda) &= |r'| \bigl(\tfrac{\log r}{r} - \sum_{i=1}^m \tfrac{\log r_i}{r_i}\bigr) \\ &= |r'| \sum_{i=1}^m (1 - \tfrac{\lambda_i}{r'}) \log(1 - \tfrac{\lambda_i}{r'}) - (1 - \tfrac{1}{r'}) \log(1 - \tfrac{1}{r'}). \end{aligned} \tag{14}$$

(The absolute value $|r'|$ is needed in the next section where r' is negative.) 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 λ tends to one of the standard unit vectors $(1, 0, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 0, 1)$ and since every λ 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 [7]). *The Rényi EPI (3) holds for $r > 1$ and $c = r^{r'/r} \bigl(1 - \frac{1}{mr'}\bigr)^{mr'-1}$.*

Proof. By Lemma 1 for $\alpha = 1$ we only need to check that the r.h.s. of (13) is greater than $\frac{n}{2} \log c$ for any choice of the λ_i's, that is, for any choice of exponents r_i such that $\sum_{i=1}^m \frac{1}{r'_i} = \frac{1}{r'}$. Thus, (3) will hold for $\log c = \min_\lambda A(\lambda)$. Now, by the log-sum inequality [4, Thm 2.7.1],

$$\sum_{i=1}^m \frac{1}{r_i} \log \frac{1}{r_i} \geq \Bigl(\sum_{i=1}^m \frac{1}{r_i}\Bigr) \log \frac{\sum_{i=1}^m \frac{1}{r_i}}{m} = (m - \tfrac{1}{r'}) \log \frac{m - \frac{1}{r'}}{m} \tag{15}$$

with equality if and only if all r_i are equal, that is, the λ_i are equal to $1/m$. Thus, $\min_\lambda A(\lambda) = r' \bigl[\frac{\log r}{r} + (m - 1/r') \log \frac{m - 1/r'}{m}\bigr] = \log c$. □

Note that $\log c = r' \frac{\log r}{r} + (mr' - 1) \log\bigl(1 - \frac{1}{mr'}\bigr) < 0$ decreases (and tends to $r' \frac{\log r}{r} - 1$) as m increases. Thus, a universal constant independent of m is obtained by taking

$$c = \inf_m r^{r'/r} \bigl(1 - \frac{1}{mr'}\bigr)^{mr'-1} = \frac{r^{r'/r}}{e}, \tag{16}$$

as was established by Bobkov and Chistyakov [6].

Proposition 2 (Li [9]). *The Rényi EPI (4) holds for $r > 1$ and $\alpha = \bigl[1 + r' \frac{\log_2 r}{r} + (2r' - 1) \log_2\bigl(1 - \frac{1}{2r'}\bigr)\bigr]^{-1}$.*

Li [9] remarked that this value of α is strictly smaller (better) than the value $\alpha = \frac{r+1}{2}$ obtained previously by Bobkov and Marsiglietti [8]. In [11] it is shown that it cannot be further improved in our framework by making it depend on m.

Proof. Since the announced α 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 (13) is greater than $\frac{n}{2}(1/\alpha - 1) H(\lambda)$ for any choice of λ_is, that is, for any choice of exponents r_i such that $\sum_{i=1}^2 \frac{1}{r'_i} = \frac{1}{r'}$. Thus, (4) will hold for $\frac{1}{\alpha} - 1 = \min_\lambda \frac{A(\lambda)}{H(\lambda)}$. Li [9] showed—this is also easily proved using [10, Lemma 8]—that the minimum is obtained when $\lambda = (1/2, 1/2)$. The corresponding value of $A(\lambda)/H(\lambda)$ is $\bigl[r' \frac{\log r}{r} + (2r' - 1) \log\bigl(1 - \frac{1}{2r'}\bigr)\bigr] / \log 2 = 1/\alpha - 1$. □

The above value of α 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 (5) holds for $r > 1$, $0 < \alpha < 1$ with $c = \bigl[m\, r^{r'/r} \bigl(1 - \frac{1}{mr'}\bigr)^{mr'-1}\bigr]^\alpha / m$.*

Proof. By Lemma 1 we only need to check that the r.h.s. of Equation (13) is greater than $\frac{n}{2}\bigl((\log c)/\alpha + (1/\alpha - 1) H(\lambda)\bigr)$, that is, $A(\lambda) \geq (\log c)/\alpha + (1/\alpha - 1) H(\lambda)$ for any choice of λ_is, that is, for any choice of exponents r_i such that $\sum_{i=1}^m \frac{1}{r'_i} = \frac{1}{r'}$. Thus, for a given $0 < \alpha < 1$, (5) will hold for $\log c = \min_\lambda \alpha A(\lambda) - (1 - \alpha) H(\lambda)$. From the preceding proofs (since both $A(\lambda)$ and $-H(\lambda)$ are convex functions of λ), the minimum is attained when all λ_is are equal. This gives $\log c = \alpha\Bigl(r' \frac{\log r}{r} + (mr' - 1) \log\bigl(1 - \frac{1}{mr'}\bigr)\Bigr) - (1 - \alpha) \log m$. □

V. REPIs for Orders <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 (12) holds and the method of the preceding section fails. For log-concave densities, however, (12) 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 $0 < \mu < 1$,

$$f(x)^{\mu} f(y)^{1-\mu} \le f(\mu x + (1-\mu)y). \tag{17}$$

Theorem 2 (Fradelizi, Madiman and Wang [13]). *If X has a log-concave density, then $h_r(rX) - rh_r(X) = (1-r)h_r(X) + n\log r$ is concave in r.*

This concavity property is used in [13] 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 slope $\frac{n\log r + (1-r)h_r(X) - 0}{r-1}$ is nonincreasing in r. In other words, $h_r(X) + n\frac{\log r}{1-r}$ is nondecreasing. Now since all r_i are $> r$,

$$h_{r_i}(X) + n\tfrac{\log r_i}{1-r_i} \ge h_r(X) + n\tfrac{\log r}{1-r} \qquad (i=1,\ldots,m). \tag{18}$$

Plugging this into (11), one obtains

$$\begin{aligned} h_r\Big(\sum_{i=1}^m \sqrt{\lambda_i} X_i\Big) &- \sum_{i=1}^m \lambda_i h_r(X_i) \\ &\ge n\big(\tfrac{\log r}{1-r} - \sum_{i=1}^m \lambda_i \tfrac{\log r_i}{1-r_i}\big) + \tfrac{n}{2} r'\big(\tfrac{\log r}{r} - \sum_{i=1}^m \tfrac{\log r_i}{r_i}\big) \\ &= \frac{n}{2} r' \big(\sum_{i=1}^m \tfrac{\log r_i}{r_i} - \tfrac{\log r}{r}\big) \end{aligned} \tag{19}$$

where we have used that $\lambda_i = r'/r'_i$ for $i = 1, 2, \ldots, m$.

Notice that, quite surprisingly, the r.h.s. of (19) for $r < 1$ ($r' < 0$) is the opposite of that of (13) for $r > 1$ ($r' > 0$). However, since r' is now negative, the r.h.s. is exactly equal to $\frac{n}{2}A(\lambda)$ 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* (3) *for log-concave densities holds for $c = r^{-r'/r}\big(1 - \frac{1}{mr'}\big)^{1-mr'}$ 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 [10]). *The Rényi EPI* (4) *log-concave densities holds for $\alpha = \big[1 + |r'|\frac{\log_2 r}{r} + (2|r'|+1)\log_2\big(1 + \frac{1}{2|r'|}\big)\big]^{-1}$ 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 REPI* (5) *for log-concave densities holds for $c = \big[mr^{-r'/r}\big(1 - \frac{1}{mr'}\big)^{1-mr'}\big]^{\alpha}/m$ where $r < 1$, $0 < \alpha < 1$.*

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 transportations 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 [14]). *If $f \in L^r(\mathbb{R}^n)$, its escort density of exponent r is defined by*

$$f_r(x) = f^r(x) \Big/ \int_{\mathbb{R}^n} f^r(x)\,\mathrm{d}x. \tag{20}$$

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

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

$$\frac{\partial}{\partial r}\big((1-r)h_r(X)\big) = \mathbb{E}\log f(X_r) = -h(X_r\|X) \tag{21}$$

$$\frac{\partial}{\partial r} h_r(X) = -\frac{1}{(1-r)^2} D(X_r\|X) \le 0 \tag{22}$$

$$\frac{\partial^2}{\partial r^2}\big((1-r)h_r(X)\big) = \operatorname{Var}\log f(X_r). \tag{23}$$

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 $\frac{\partial}{\partial r}\big((1-r)h_r(X)\big) = \frac{\partial}{\partial r}\log\int f^r = \int f_r \log f$. Taking another derivative yields $\frac{\partial}{\partial r}\frac{\int f^r \log f}{\int f^r} = \int f_r(\log f)^2 - (\int f_r \log f)^2$. Since $\frac{\partial}{\partial r}\big((1-r)h_r(X)\big) = (1-r)\frac{\partial}{\partial r}h_r(X) - h_r(X)$ we have $(1-r)^2\frac{\partial}{\partial r}h_r(X) = \int f_r \log(f/f^r) + \log\int f^r = \int f_r\log(f/f_r)$. ☐

Eq. (22) 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 (23) 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 [15]). *Given $X \sim f$ and $Y \sim g$, their* relative Rényi entropy *of exponent r (relative r-entropy) is given by*

$$\Delta_r(X\|Y) = D_{\frac{1}{r}}(X_r\|Y_r)$$

where $D_r(X\|Y) = \frac{1}{r-1}\log\int f^r g^{1-r}$ is the r-divergence [16].

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 coïncide. It is easily checked from the definitions that

$$\Delta_r(X\|Y) = -r'\log\int f_r^{1/r} g_r^{1/r'} = -r'\log\mathbb{E}\big(g_r^{1/r'}(X)\big) - h_r(X) \tag{24}$$

$$h_r(X) = -r'\log\mathbb{E}\big(f_r^{1/r'}(X)\big). \tag{25}$$

Thus, just like for the case $r = 1$, the relative r-entropy (24) is the difference between the expression of the r-entropy (25) in which f is replaced by g, and the r-entropy itself.

Since the Rényi divergence $D_r(X\|Y) = \frac{1}{r-1}\int f^r g^{1-r}$ is nonnegative and vanishes if and only if the two distributions f and g coïncide, the relative entropy $\Delta_r(X\|Y)$ enjoys the same property. From (24) we have the following

Proposition 8 (Rényi-Gibbs' inequality). *If $X \sim f$,*

$$h_r(X) \le -r'\log\mathbb{E}\big(g_r^{1/r'}(X)\big) \tag{26}$$

for any density g, with equality if and only if $f = g$ a.e.

Letting $r \to 1$ one recovers the usual Gibbs' inequality.

Definition 3 (Arimoto's Conditional Rényi Entropy [18]).

$$h_r(X|Z) = -r'\log\mathbb{E}\|f(\cdot|Z)\|_r = -r'\log\mathbb{E} f_r^{1/r'}(X|Z)$$

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

$$h_r(X|Z) \le -r'\log\mathbb{E}\big(g_r^{1/r'}(X|Z)\big). \tag{27}$$

If in particular we put $g(x|z) = f(x)$ independent of z, the r.h.s. becomes equal to (25). We have thus obtained a simple proof of the following

Proposition 9 (Conditioning reduces r-entropy [18])**.**

$$h_r(X|Z) \leq h_r(X) \tag{28}$$

with equality if and only if X and Z are independent.

Another important property is the data processing inequality [16] which implies $D_r(T(X)\|T(Y)) \leq D_r(X\|Y)$ for any transformation T. The same holds for relative r-entropy when 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*

$$X_r = T(X_r^*) \quad \text{and} \quad Y_r = T(Y_r^*), \tag{29}$$

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

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

When T is invertible, inequalities in both directions hold:

Proposition 11 (Relative r-entropy preserves transport)**.** *For an invertible transport T satisfying* (29)*, $\Delta_r(X\|Y) = \Delta_r(X^*\|Y^*)$.*

From (24) the equality $\Delta_r(X\|Y) = \Delta_r(X^*\|Y^*)$ can be rewritten as the following identity:

$$-r' \log \mathbb{E}\big(g_r^{\frac{1}{r'}}(X)\big) - h_r(X) = -r' \log \mathbb{E}\big({g^*}_r^{\frac{1}{r'}}(X^*)\big) - h_r(X^*). \tag{30}$$

Assuming T is a diffeomorphism, the density g_r^* of Y_r^* is given by the change of variable formula $g_r^*(u) = g_r(T(u))|T'(u)|$ where the Jacobian $|T'(u)|$ is the absolute value of the determinant of the Jacobian matrix $T'(u)$. In this case (30) can be rewritten as

$$\begin{aligned} &-r' \log \mathbb{E}\big(g_r^{\frac{1}{r'}}(X)\big) - h_r(X) \\ &\quad = -r' \log \mathbb{E}\big(g_r^{\frac{1}{r'}}(T(X^*))|T'(X^*)|^{\frac{1}{r'}}\big) - h_r(X^*). \end{aligned} \tag{31}$$

VII. A Transportation Proof of Theorem 1

We proceed to prove (10). It is easily seen, using finite induction on m, that it suffices to prove the corresponding inequality for $m = 2$ arguments:

$$\begin{aligned} &h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &\geq h_r(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^*) - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*) \end{aligned} \tag{32}$$

with equality if and only if X, Y are i.i.d. Gaussian. Here X^* and Y^* are i.i.d. standard Gaussian $\mathcal{N}(0, \mathbf{I})$ 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 (Normal Transport)**.** *Let f be given and $X^* \sim \mathcal{N}(0, \sigma^2\mathbf{I})$. There exists a diffeomorphism $T : \mathbb{R}^n \to \mathbb{R}^n$ with log-concave Jacobian $|T'|$ such that $X = T(X^*) \sim f$.*

Thus T transports normal X^* to X. The log-concavity property is that for any such transports T, U and $\lambda \in (0, 1)$, we have

$$|T'(X^*)|^{\lambda}|U'(Y^*)|^{1-\lambda} \leq |\lambda T'(X^*) + (1-\lambda)U'(Y^*)|. \tag{33}$$

The proof of Lemma 2 is very simple for one-dimensional variables [19], where T is just an increasing function with continuous derivative $T' > 0$ and where (33) is the classical arithmetic-geometric inequality.

For dimensions $n > 1$, Lemma 2 comes into two flavors:
(i) Knöthe maps: T can be chosen such that its Jacobian matrix T' is (lower) triangular with positive diagonal elements (Knöthe–Rosenblatt map [20], [21]). Two different elementary proofs are given in [12]. Inequality (33) results from the concavity of the logarithm applied to the Jacobian matrices' diagonal elements.
(ii) Brenier maps: T can be chosen such that its Jacobian matrix T' is symmetric positive definite (Brenier map [22], [23]). In this case (33) is Ky Fan's inequality [4, § 17.9].

The key argument is now the following. Considering escort variables, by transport (Lemma 2), one can write $X_p = T(X_p^*)$ and $Y_q = U(Y_q^*)$ for two diffeomorphims T and U satisfying (33). Then by transport preservation (Proposition 11), we have $\lambda\Delta_p(X\|U) + (1-\lambda)\Delta_p(Y\|V) = \lambda\Delta_p(X^*\|U^*) + (1-\lambda)\Delta_p(Y^*\|V^*)$ for any $U \sim \varphi$ and $V \sim \psi$, which from (31) can be easily rewritten in the form

$$\begin{aligned} &-r' \log \mathbb{E}\big(\chi^{\frac{1}{r'}}(X, Y)\big) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &= -r' \log \mathbb{E}\Big(\big(\chi(T(X^*), U(Y^*))|T'(X^*)|^{\lambda}|U'(Y^*)|^{1-\lambda}\big)^{\frac{1}{r'}}\Big) \\ &\qquad - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*) \end{aligned} \tag{34}$$

where we have noted $\chi(x, y) = \varphi_p^{\lambda}(x)\psi_q^{1-\lambda}(y)$. 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 from (25) we have

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) = -r' \log \mathbb{E}\big(\theta_r^{1/r'}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y)\big)$$

where θ is the density of $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$. Therefore, the l.h.s. of (32) can be written as

$$\begin{aligned} &h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &= -r' \log \mathbb{E}\big(\theta_r^{\frac{1}{r'}}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y)\big) - \lambda h_p(X) - (1-\lambda)h_q(Y) \end{aligned} \tag{35}$$

Applying (34) to $\chi(x, y) = \theta_r(\sqrt{\lambda}x + \sqrt{1-\lambda}y)$ and using the inequality (33) gives

$$\begin{aligned} &h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &\quad \geq -r' \log \mathbb{E}\big(\varphi^{\frac{1}{r'}}(X^*, Y^*)\big) - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*) \end{aligned} \tag{36}$$

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 [12])**.** *If X^*, Y^* are i.i.d. Gaussian, then for any $0 < \lambda < 1$, the rotation*

$$\widetilde{X} = \sqrt{\lambda}\, X^* + \sqrt{1-\lambda}\, Y^*, \quad \widetilde{Y} = -\sqrt{1-\lambda}\, X^* + \sqrt{\lambda}\, Y^* \tag{37}$$

yields i.i.d. Gaussian variables $\widetilde{X}, \widetilde{Y}$.

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 [19, Lemma 4] [24, Chap. 5]).

Since the starred variables can be expressed in terms of the tilde variables by the inverse rotation $X^* = \sqrt{\lambda}\, \widetilde{X} - \sqrt{1-\lambda}\, \widetilde{Y}$, $Y^* = \sqrt{1-\lambda}\, \widetilde{X} + \sqrt{\lambda}\, \widetilde{Y}$, inequality (36) can be written as

$$\begin{aligned} &h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &\geq -r' \log \mathbb{E}\big(\psi^{1/r'}(\widetilde{X}|\widetilde{Y})\big) - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*), \end{aligned} \tag{38}$$

where $\psi(\widetilde{x}|\widetilde{y}) = \theta_r(\sqrt{\lambda}T(\sqrt{\lambda}\widetilde{x}-\sqrt{1-\lambda}\widetilde{y})+\sqrt{1-\lambda}U(\sqrt{1-\lambda}\widetilde{x}+\sqrt{\lambda}\widetilde{y}))\cdot|\lambda T'(\sqrt{\lambda}\widetilde{x}-\sqrt{1-\lambda}\widetilde{y})+(1-\lambda)U'(\sqrt{1-\lambda}\widetilde{x}+\sqrt{\lambda}\widetilde{y})|$. Making the change of variable $z = \sqrt{\lambda}T(\sqrt{\lambda}\widetilde{x} - \sqrt{1-\lambda}\widetilde{y}) + \sqrt{1-\lambda}U(\sqrt{1-\lambda}\widetilde{x} + \sqrt{\lambda}\widetilde{y})$, we check that $\int\psi(\widetilde{x}|\widetilde{y})\,\mathrm{d}\widetilde{x} = \int\theta_r(z)\,\mathrm{d}z = 1$ since θ_r is a density. Hence, $\psi(\widetilde{x}|\widetilde{y})$ is a conditional density, and by (27),

$$-r'\log\mathbb{E}\big(\psi^{1/r'}(\widetilde{X}|\widetilde{Y})\big) \geq h_r(\widetilde{X}|\widetilde{Y}) \tag{39}$$

where $h_r(\widetilde{X}|\widetilde{Y}) = h_r(\widetilde{X}) = h_r(\sqrt{\lambda}\,X^* + \sqrt{1-\lambda}\,Y^*)$ since $\widetilde{X}$ and $\widetilde{Y}$ are independent. Combining with (38) yields the announced inequality (32).

It remains to settle the equality case in (32). From the above proof, equality holds in (32) if and only if both (33) and (39) are equalities. The rest of the argument depends on whether Knöthe or Brenier maps are used:

(i) Knöthe maps: In the case of Knöthe maps, Jacobian matrices are triangular and equality in (33) holds if and only if for all i, $\frac{\partial T_i}{\partial x_i}(X^*) = \frac{\partial U_i}{\partial y_i}(Y^*)$ a.s. Since X^* and Y^* are independent Gaussian, this implies that $\frac{\partial T}{\partial x_i}$ and $\frac{\partial U}{\partial y_i}$ are constant and equal. In particular the Jacobian $|\lambda T'(\sqrt{\lambda}\widetilde{x}-\sqrt{1-\lambda}\widetilde{y})+(1-\lambda)U'(\sqrt{1-\lambda}\widetilde{x}+\sqrt{\lambda}\widetilde{y})|$ is constant. Now since $h_r(\widetilde{X}|\widetilde{Y}) = h_r(\widetilde{X})$ equality in (39) holds only if $\psi(\widetilde{x}|\widetilde{y})$ does not depend on $\widetilde{y}$, which implies that $\sqrt{\lambda}T(\sqrt{\lambda}\widetilde{x} - \sqrt{1-\lambda}\widetilde{y}) + \sqrt{1-\lambda}U(\sqrt{1-\lambda}\widetilde{x} + \sqrt{\lambda}\widetilde{y})$ does not depend on the value of $\widetilde{y}$. Taking derivatives with respect to y_j for all j, we have $\frac{\partial T_i}{\partial x_j}(X^*) = \frac{\partial U_i}{\partial y_j}(Y^*)$ a.s. for all i,j. In other words, $T'(X^*) = U'(Y^*)$ a.s.

(ii) 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 (33) 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 $X_p \sim \mathcal{N}(0,\mathbf{K}/p)$ and $Y_q \sim \mathcal{N}(0,\mathbf{K}/q)$. Hence, X and Y are i.i.d. Gaussian $\mathcal{N}(0,\mathbf{K})$. This ends the proof of Theorem 1. $\square$

We note that this section has provided an information-theoretic proof the strengthened Young's convolutional inequality (with optimal constants), since (32) 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 $(1-r)h_r(X) + n\log r \geq \lambda\big((1-p)h_p(X) + n\log p\big) + (1-\lambda)\big((1-q)h_q(X) + n\log q\big)$.

By Lemma 2 there exists two diffeomorphisms T, U such that one can write $pX_p = T(X^*)$ and $qX_q = U(X^*)$. Then, by these changes of variables X^* has density

$$\tfrac{1}{p^n}f_p\big(\tfrac{T(x^*)}{p}\big)|T'(x^*)| = \tfrac{1}{q^n}f_q\big(\tfrac{U(x^*)}{q}\big)|U'(x^*)| \tag{40}$$

which can be written

$$\frac{f^p\big(\frac{T(x^*)}{p}\big)|T'(x^*)|}{\exp\big((1-p)h_p(X)+n\log p\big)} = \frac{f^q\big(\frac{U(x^*)}{q}\big)|U'(x^*)|}{\exp\big((1-q)h_q(X)+n\log q\big)}.$$

Taking the geometric mean, integrating over x^* and taking the logarithm gives the representation

$$\begin{aligned}&\lambda\big((1-p)h_p(X)+n\log p\big)+(1-\lambda)\big((1-q)h_q(X)+n\log q\big)\\ &=\log\int f^{\lambda p}\big(\tfrac{T(x^*)}{p}\big)f^{(1-\lambda)q}\big(\tfrac{U(x^*)}{q}\big)|T'(x^*)|^{\lambda}|U'(x^*)|^{1-\lambda}\,\mathrm{d}x^*.\end{aligned}$$

Now, by log-concavity (17) (with $\mu = \lambda p/r$) and (33),

$$\begin{aligned}&\lambda\big((1-p)h_p(X) + n\log p\big) + (1-\lambda)\big((1-q)h_q(X) + n\log q\big)\\ &\leq \log\int f^r\big(\tfrac{\lambda T(x^*)+(1-\lambda)U(x^*)}{r}\big)|\lambda T'(x^*)+(1-\lambda)U'(x^*)|\,\mathrm{d}x^*\\ &= \log\big(r^n\int f^r\big) = (1-r)h_r(X) + n\log r.\end{aligned}$$

This ends the proof of Theorem 2. $\square$

This theorem asserts that the second derivative $\frac{\partial^2}{\partial r^2}\big((1-r)h_r(X)+n\log r\big) \leq 0$. From (23) this gives $\operatorname{Var}\log f(X_r) \leq n/r^2$, that is, $\operatorname{Var}\log f_r(X_r) \leq n$. Setting $r=1$, this is the varentropy bound $\operatorname{Var}\log f(X) \leq n$ of [13].

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 623–656, Oct. 1948.

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

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

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[5] M. Madiman, J. Melbourne, and P. Xu, "Forward and reverse entropy power inequalities in convex geometry," in *Convexity and Concentration*, ser. IMA Volumes in Mathematics and its Applications, E. Carlen, M. Madiman, & E. Werner, Eds. Springer, 2017, vol. 161, pp. 427–485.

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

[7] E. Ram and I. Sason, "On Rényi entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6800–6815, Dec. 2016.

[8] S. G. Bobkov and A. Marsiglietti, "Variants of the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 63, no. 12, pp. 7747–7752.

[9] J. Li, "Rényi entropy power inequality and a reverse," *Studia Mathematica*, vol. 242, pp. 303–319, Feb. 2018.

[10] A. Marsiglietti and J. Melbourne, "On the entropy power inequality for the Rényi entropy of order $[0,1]$," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1387–1396, Mar. 2019.

[11] O. Rioul, "Rényi entropy power inequalities via normal transport and rotation," *Entropy*, vol. 20, no. 9, p. 641, Sep. 2018.

[12] ——, "Yet another proof of the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3595–3599, Jun. 2017.

[13] M. Fradelizi, M. Madiman, and L. Wang, "Optimal concentration of information content for log-concave densities," in *High Dimensional Probability VII: The Cargèse Volume*, Basel: Birkhäuser, 2016.

[14] J.-F. Bercher, "Source coding with escort distributions and Rényi entropy bounds," *Physics Letters A*, vol. 373, no. 36, pp. 3235–3238.

[15] A. Lapidoth and C. Pfister, "Two measures of dependence," in *IEEE Int. Conf. Science Electrical Engineering (ICSEE 2016)*, 2016.

[16] T. van Erven and P. Harremoës, "Rényi and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.

[17] S. Verdú, "α-mutual information," in *Information Theory and Applications Workshop (ITA 2015)*, Feb. 2015.

[18] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6801–6810, Nov. 2014.

[19] O. Rioul, "Optimal transportation to the entropy-power inequality," in *IEEE Inf. Theory Applications Workshop (ITA 2017)*, Feb. 2017.

[20] M. Rosenblatt, "Remarks on a multivariate transformation," *Ann. Math. Stat.*, vol. 23, no. 3, pp. 470–472, 1952.

[21] H. Knöthe, "Contributions to the theory of convex bodies," *Michigan Math. J.*, vol. 4, pp. 39–52, 1957.

[22] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Comm. Pure Applied Math.*, vol. 44, no. 4, pp. 375–417, Jun. 1991.

[23] R. J. McCann, "Existence and uniqueness of monotone measure-preserving maps," *Duke Math. J.*, vol. 80 pp. 309–324, Nov. 1995.

[24] W. Bryc, *The Normal Distribution - Characterizations with Applications*, ser. Lecture Notes in Statistics. Springer, 1995, vol. 100.