Article

Rényi Entropy Power Inequalities via Normal Transport and Rotation

Olivier Rioul [1,2]

1   LTCI, Télécom ParisTech, Université Paris-Saclay, 75013 Paris, France; olivier.rioul@telecom-paristech.fr
     or olivier.rioul@polytechnique.edu
2   École Polytechnique, Université Paris-Saclay, 91128 Palaiseau, France

Abstract: Following a recent proof of Shannon's entropy power inequality (EPI), a comprehensive framework for deriving various EPIs for the Rényi entropy is presented that uses transport arguments from normal densities and a change of variable by rotation. 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. In particular, for log-concave densities, we obtain a simple transportation proof of a sharp varentropy bound.

Keywords: Rényi entropy; entropy power inequalities; transportation arguments; normal distributions; escort distributions; log-concave distributions

1. Introduction

The entropy power inequality (EPI) dates back to Shannon's seminal paper [1] and has a long history [2]. The link with the Rényi entropy 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 the Rényi entropy orders tend to one ([4] Theorem 17.8.3).

The Rényi entropy [5] was first defined as a generalization of Shannon's entropy for *discrete* variables, when looking for the most general definition of information measures that would preserve the additivity for independent events. It has found many applications such as source coding [6], hypothesis testing [7], channel coding [8] and guessing [9]. The (differential) Rényi entropy considered in this paper (Definition 2) generalizes the (differential) Shannon's entropy for continuous variables. It was first considered in [3] to make the transition between the entropy-power and the Brunn–Minkowski inequalities. It has also been applied to deconvolution problems [10]. A definition of the Renyi entropy-power itself appears in [11], which is essentially Definition 5 below.

Recently, there has been significant interest in Rényi entropy power inequalities for several independent variables (the survey [12] is recommended to the reader for recent developments on forward and reverse EPIs). Bobkov and Chistyakov [13] extended the classical Shannon's EPI to the Rényi entropy by incorporating a multiplicative constant that depends on the order of the Rényi entropy. Ram and Sason [14] improved the value of the constant by making it depend also on the number of variables. Bobkov and Marsiglietti [15] proved another modification of the EPI for the Rényi entropy for two independent variables, with a power exponent parameter α whose value was further improved by Li [16]. All these EPIs were found for Rényi entropies of orders >1. The α-modification of the Rényi EPI was extended to orders <1 for two independent variables having log-concave densities by Marsiglietti and Melbourne [17]. The starting point of all the above works was Young's strengthened convolutional inequality.

Recently, Shannon's original EPI was given a simple proof [18] using a transport argument from normal variables and a change of variable by rotation. In this paper, we exploit these ingredients,

described in the following lemmas, to establish all the above-mentioned Rényi EPIs and derive new ones.

Notation 1. *Throughout this article, the considered n-dimensional zero-mean random variables $X \in \mathbb{R}^n$ admit a density which is implicitly assumed continuous inside its support. We write $X \sim f$ if X has density f and write $X^* \sim \mathcal{N}(0, \mathbf{K})$ if X^* is normally distributed with $n \times n$ covariance matrix $\mathbf{K}$.*

Lemma 1 (Normal Transport). *Let $X^* \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. There exists a diffeomorphism $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $X = T(X^*) \sim f$. Moreover, T can be chosen such that its Jacobian matrix T' is (lower) triangular with positive diagonal elements.*

Lemma 1 is known in optimal transport theory as an application of the Knothe–Rosenblatt map [19,20]. Two different proofs are given in [18]. The proof is very simple for one-dimensional variables [21], where T is just an increasing function with continuous derivative $T' > 0$.

Lemma 2 (Normal Rotation [18]). *If X^*, Y^* are i.i.d. (independent and identically distributed) and normal, then for any $0 < \lambda < 1$, the rotation*

$$\begin{cases} \widetilde{X} = \sqrt{\lambda}\, X^* + \sqrt{1 - \lambda}\, Y^* \\ \widetilde{Y} = -\sqrt{1 - \lambda}\, X^* + \sqrt{\lambda}\, Y^* \end{cases} \tag{1}$$

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

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

$$\begin{aligned} X^* &= \sqrt{\lambda}\, \widetilde{X} - \sqrt{1 - \lambda}\, \widetilde{Y} \\ Y^* &= \sqrt{1 - \lambda}\, \widetilde{X} + \sqrt{\lambda}\, \widetilde{Y}. \end{aligned} \tag{2}$$

The proof of Lemma 2 is trivial considering covariance matrices. A deeper result states that this property of remaining i.i.d. by rotation characterizes the normal distribution—this is known as Bernstein's lemma (see, e.g., ([21] [Lemma 4]) and ([22] [Chapter 5])). This explains why one obtains equality in the EPI only for normal variables (see [21] for more details).

This article is a revised, full version of what was presented in part in a previous conference communication [23]. It is organized as follows. Preliminary definitions and known properties are presented in Section 2. Section 3 derives a crucial "information inequality" for Rényi entropies that enjoys a transformational invariance. The central result is in Section 4, where the first version of the Rényi EPI by Dembo, Cover and Thomas is proven using the ingredients of Lemmas 1 and 2. All previously known Rényi EPIs for finite orders—and new ones—are then derived using a simple method in Section 5. Section 6 concludes.

2. Preliminary Definitions and Properties

Throughout this article, we consider exponents $p > 0$ with $p \neq 1$. The following definition is well known and used, e.g., in Hölder's inequality.

Definition 1 (Conjugate Exponent). *The conjugate exponent of p is*

$$p' = \frac{p}{p - 1}, \tag{3}$$

that is, the number p' such that

$$\frac{1}{p} + \frac{1}{p'} = 1. \tag{4}$$

Remark 1. *There are two situations depending on whether p' is positive or negative, as summarized in the following table:*

$p > 1$	or	$0 < p < 1$
$p' > 1$		$p' < 0$

Definition 2 (Rényi Entropy). *If X has density $f \in L^p(\mathbb{R}^n)$, its Rényi entropy of order p is defined by*

$$h_p(X) = \frac{1}{1-p} \log \int_{\mathbb{R}^n} f^p(x)\, \mathrm{d}x = -p' \log \|f\|_p \tag{5}$$

where $\|f\|_p$ denotes the L^p norm of f.

It is known that the limit as $p \to 1$ is the Shannon entropy

$$h_1(X) = -\int_{\mathbb{R}^n} f(x) \log f(x)\, \mathrm{d}x. \tag{6}$$

The Rényi entropy enjoys well-known properties similar to those of the Shannon entropy, which are recalled here for completeness.

Lemma 3 (Scaling Property). *For any $a \neq 0$,*

$$h_p(aX) = h_p(X) + n \log |a|. \tag{7}$$

Proof. Making a change of variables, $h_p(aX) = \frac{1}{1-p} \log \int \left(\frac{1}{|a|^n} f(\frac{x}{a}) \right)^p \mathrm{d}x = \frac{1}{1-p} \log \int f^p(\frac{x}{a}) \frac{\mathrm{d}x}{|a|^n} + \frac{1}{1-p} \log |a|^{n(1-p)} = h_p(X) + n \log |a|$. $\square$

One recovers the usual scaling property for the Shannon entropy by letting $p \to 1$.

Lemma 4 (Rényi Entropy of the Normal). *If $X^* \sim \mathcal{N}(0, \mathbf{K})$ for some nonsingular covariance matrix $\mathbf{K}$, then*

$$h_p(X^*) = \frac{1}{2} \log((2\pi)^n |\mathbf{K}|) + \frac{n}{2} p' \frac{\log p}{p} \tag{8}$$

where $|\cdot|$ denotes the determinant. In particular, for $X^* \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$,

$$h_p(X^*) = \frac{n}{2} \log(2\pi\sigma^2) + \frac{n}{2} p' \frac{\log p}{p}. \tag{9}$$

Proof. By direct calculation, $h_p(X^*) = \frac{1}{1-p} \log \int \left(\frac{\exp(-\frac{1}{2} x^t \mathbf{K}^{-1} x)}{\sqrt{(2\pi)^n |\mathbf{K}|}} \right)^p \mathrm{d}x = \frac{1}{1-p} \log \frac{\sqrt{(2\pi)^n |\mathbf{K}| p^{-n}}}{\sqrt{(2\pi)^n |\mathbf{K}|^p}} = \frac{1}{2} \log((2\pi)^n |\mathbf{K}|) - \frac{n}{2} \frac{\log p}{1-p}$. $\square$

Again, one recovers the Shannon entropy of a normal variable by letting $p \to 1$ (then, $p' \frac{\log p}{p} \to \log e$). The following notion of escort distribution [24,25] is useful in the sequel.

Definition 3 (Escort Density ([25] § 2.2)). *If $f \in L^p(\mathbb{R}^n)$, its escort density of exponent p is the density defined by*

$$f_p(x) = \frac{f^p(x)}{\int_{\mathbb{R}^n} f^p(x)\, \mathrm{d}x}. \tag{10}$$

In other words, $f_p = f^p / \|f\|_p^p$, where $\|f\|_p$ denotes the L^p norm of f. We also use the notation X_p to denote the corresponding escort random variable with density f_p.

Lemma 5 (Monotonicity Property). *If $p < q$, then $h_p(X) \geq h_q(X)$ with equality if and only if X is uniformly distributed.*

Proof. Let $p \neq 1$ and assume that $f \in L^q(\mathbb{R}^n)$ for all q in a neighborhood of p so that one can freely differentiate under the integral sign:

$$
\begin{aligned}
\frac{\partial}{\partial p} h_p(X) &= \frac{1}{(1-p)^2} \log \int f^p + \frac{1}{1-p} \frac{\int f^p \log f}{\int f^p} \\
&= \frac{1}{(1-p)^2} \left(\log \int f^p + \int f_p \log f^{1-p} \right) \\
&= \frac{1}{(1-p)^2} \int f_p \log \frac{f}{f_p} \\
&= -\frac{D(f_p \| f)}{(1-p)^2} \leq 0
\end{aligned}
\tag{11}
$$

where $D(\cdot \| \cdot)$ denotes the Kullback–Leibler divergence. Equality $D(f_p \| f) = 0$ can hold only if $f = f_p$ a.e., which, since $p \neq 1$, implies that f is constant over some measurable subset $A \subset \mathbb{R}^n$, and is zero elsewhere. It follows that $h_p(X) > h_q(X)$ for any $p < q$ if X is not uniformly distributed. Conversely, if X is uniformly distributed over some measurable subset $A \subset \mathbb{R}^n$, its density can be written as $f(x) = 1/\mathrm{vol}(A)$ for all $x \in A$ and $f(x) = 0$ elsewhere. Then, $h_p(X) = \frac{1}{1-p} \log \frac{\mathrm{vol}(A)}{\mathrm{vol}(A)^p} = \log \mathrm{vol}(A)$ is independent of p. $\square$

Remark 2. *Notice the identity established in the proof:*

$$
\frac{\partial}{\partial p} h_p(X) = -\frac{D(f_p \| f)}{(1-p)^2}.
\tag{12}
$$

A similar formula for discrete variables can be found in ([26] § 5.3).

3. An Information Inequality

The Shannon entropy satisfies a fundamental "information inequality" ([4] Theorem 2.6.3) from which many classical information-theoretic inequalities can be derived. This can be written as

$$
h_1(X) \leq -\mathbb{E} \log \varphi(X)
\tag{13}
$$

for any density φ, with equality if and only if $\varphi = f$ a.e. The following Theorem 1 can be seen as the natural extension of the information inequality to Rényi entropies and is central in the following derivations of this paper. Bercher [27] pointed out to the author that it is similar to an inequality for discrete distributions established by Campbell [6] in the context of source coding (see also [24]).

Theorem 1 (Information Inequality). *For any density φ,*

$$
h_p(X) \leq -p' \log \mathbb{E}\big(\varphi^{1/p'}(X)\big)
\tag{14}
$$

with equality if and only if $\varphi = f_p$ a.e.

By letting $p \to 1$, one recovers the classical information inequality in Equation (13) for the Shannon entropy.

Proof. By definition in Equation (5),

$$
\begin{aligned}
h_p(X) &= -p' \log \|f\|_p \\
&= -p' \log \left(\int (f^p \varphi^{-1}) \varphi \right)^{1/p} \\
&\leq -p' \log \int (f^p \varphi^{-1})^{1/p} \varphi \\
&= -p' \log \int f \varphi^{1/p'}
\end{aligned}
\tag{15}
$$

where the inequality follows from Jensen's inequality applied to the function $x \mapsto x^{1/p}$, which is strictly concave if $p > 1$ (that is, $p' > 0$) and strictly convex if $p < 1$ (that is, $p' < 0$). Equality holds if and only if $f^p \varphi^{-1}$ is constant a.e., which means that φ and f^p are proportional a.e. Normalizing gives the announced condition $\varphi = f_p$ a.e.　□

Remark 3. *An alternate proof is obtained using Hölder's inequality or its reverse applied to f and $\varphi^{1/p'}$. Notice that the equality case for $\varphi = f_p$ gives*

$$
h_p(X) = -p' \log \mathbb{E}\big(f_p^{1/p'}(X)\big)
\tag{16}
$$

as can be easily checked directly.

The following conditional version of Theorem 1 involves a more complicated relation for dependent variables.

Corollary 1 (Conditional Information Inequality). *For any two random variables $X, Y \in \mathbb{R}^n$,*

$$
-p' \log \mathbb{E}_Y \exp\big(-h_p(X|Y)/p'\big) \leq -p' \log \mathbb{E}\big(\varphi^{1/p'}(X|Y)\big)
\tag{17}
$$

where $h_p(X|y)$ denotes the Rényi entropy of X knowing $Y = y$ and the expectation on the l.h.s. is taken over Y (the expectation in the r.h.s. is taken over (X, Y)).

In particular, when X and Y are independent,

$$
h_p(X) \leq -p' \log \mathbb{E}\big(\varphi^{1/p'}(X|Y)\big).
\tag{18}
$$

with equality if and only if $\varphi(x|y)$ does not depend on y and equals $f_p(x)$ a.e.

Proof. From Equation (14) for fixed y, one has $\mathbb{E}\big(\varphi^{1/p'}(X|y)\big) \leq \exp\big(-h_p(X|y)/p'\big)$ for $p' > 0$ ($p > 1$) and the opposite inequality for $p' < 0$ ($p < 1$), with equality if and only if $\varphi(x|y) = f_p(x|y)$ a.e. Taking the expectation over Y yields $\mathbb{E}\big(\varphi^{1/p'}(X|Y)\big) \leq \mathbb{E}_Y \exp\big(-h_p(X|Y)/p'\big)$ for $p' > 0$ ($p > 1$) and the opposite inequality for $p' < 0$ ($p < 1$). The result follows by taking the logarithm and multiplying by $-p'$. When X and Y are independent, equality holds if and only if $\varphi(x|y) = f_p(x)$ a.e. for all y.　□

For the Shannon entropy, the difference between the two sides of the information inequality in Equation (13) is the Kullback–Leibler divergence:

$$
D(f\|\varphi) = \mathbb{E} \log \frac{f(X)}{\varphi(X)} \geq 0
$$

which can also be noted $D(X\|Z)$ where $X \sim f$ and $Z \sim \varphi$. It is known (and easy to check) that the divergence is invariant by reversible transformations T. This means that, when $X = T(X^*)$ and $Z = T(Z^*)$, one has $D(X\|Z) = D(X^*\|Z^*)$. A natural extension to Rényi entropies can be obtained on the difference

$$- p' \log \mathbb{E}\big(\varphi^{1/p'}(X)\big) - h_p(X) \geq 0 \tag{19}$$

between the two sides of the information inequality in Equation (14).

Theorem 2 (Transformational Invariance). *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism and suppose that*

$$X_p = T(X_p^*) \tag{20}$$
$$Z = T(Z^*) \tag{21}$$

where $Z \sim \varphi$ and $Z^* \sim \varphi^*$. Then,

$$- p' \log \mathbb{E}\big(\varphi^{1/p'}(X)\big) - h_p(X) = -p' \log \mathbb{E}\big(\varphi^{*1/p'}(X^*)\big) - h_p(X^*). \tag{22}$$

Note that, from Equation (5), this identity can be rewritten as

$$\frac{\mathbb{E}\big(\varphi^{*1/p'}(X^*)\big)}{\|f^*\|_p} = \frac{\mathbb{E}\big(\varphi^{1/p'}(X)\big)}{\|f\|_p}. \tag{23}$$

Proof. Proceed to prove Equation (23). Let f, f^* be the respective densities of X, X^* and recall that $X_p \sim f_p$ and $X_p^* \sim f_p^*$. By the transformation T, the densities are related by

$$f_p^*(x^*) = f_p(T(x^*))|T'(x^*)| \tag{24}$$
$$\varphi^*(x^*) = \varphi(T(x^*))|T'(x^*)| \tag{25}$$

where $|T'|$ denotes the Jacobian determinant of T. Using these relations and Definition 3,

$$
\begin{aligned}
\mathbb{E}\big(\varphi^{*1/p'}(X^*)\big)/\|f^*\|_p &= \mathbb{E}\big(\varphi^{1/p'}(T(X^*))|T'(X^*)|^{1/p'}\big)/\|f^*\|_p \\
&= \int \varphi^{1/p'}(T(x^*))|T'(x^*)|^{1/p'} f^*(x^*)\, dx^* /\|f^*\|_p \\
&= \int \varphi^{1/p'}(T(x^*))|T'(x^*)|^{1/p'} f_p^*(x^*)^{1/p}\, dx^* \\
&= \int \varphi^{1/p'}(T(x^*)) f_p(T(x^*))^{1/p} |T'(x^*)|\, dx^* \\
&= \int \varphi^{1/p'}(x) f_p(x)^{1/p}\, dx \\
&= \mathbb{E}\big(\varphi^{1/p'}(X)\big)/\|f\|_p. \quad \square
\end{aligned}
\tag{26}
$$

Remark 4. *φ is a density and is not used in the proof of Theorem 2. Therefore, Equation (22) holds more generally for any function φ satisfying Equation (25).*

4. First Version of the Rényi EPI

For two independent random variables X and Y, the Shannon entropy power inequality can be expressed as follows [2,3]: For any $0 < \lambda < 1$,

$$h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda h(X) + (1-\lambda)h(Y) \tag{27}$$

with equality if and only if X, Y are i.i.d. normal—more precisely (given the translation invariance of the entropy) when they are normal with identical covariance matrices; since it was assumed in Notation 1 that all considered variables have zero mean, both statements are equivalent. This means that the difference $h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y)$ is minimum (zero) for i.i.d. normal X, Y.

In this section, we study the natural generalization for Rényi entropies ([3] Theorem 12), namely that the quantity

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \tag{28}$$

is minimum for X, Y i.i.d. normal. Here, the triple (p, q, r) and its associated λ satisfy the following condition, which is used, e.g., in Young's convolutional inequality.

Definition 4 (Exponent Triple). *An exponent triple $(p, q, r)_\lambda$ has conjugates p', q', r' of the same sign and such that*

$$\frac{1}{p'} + \frac{1}{q'} = \frac{1}{r'}. \tag{29}$$

The corresponding coefficient $\lambda \in (0, 1)$ is defined by

$$\lambda = \frac{r'}{p'} = 1 - \frac{r'}{q'} \tag{30}$$

In other words, the exponents p, q, r are such that

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r} \tag{31}$$

and fulfill one the following two conditions:

$p, q, r > 1$	or	$0 < p, q, r < 1$
$p', q', r' > 1$		$p', q', r' < 0$
$r' < p', r' < q'$		$|r'| < |p'|, |r'| < |q'|$
$r > p, r > q$		$r < p, r < q$

The key argument used in this section is the following. If $X \sim f$ and $Y \sim g$, then for the escort variables, $X_p \sim f_p$ and $Y_q \sim g_q$. By normal transport (Lemma 1), one can write

$$X_p = T(X_p^*) \tag{32}$$
$$Y_q = U(Y_q^*) \tag{33}$$

for two diffeomorphims T and U, where X^*, Y^* are, say, i.i.d. standard normal $\mathcal{N}(0, \mathbf{I})$. (It follows that $X_p^* \sim \mathcal{N}(0, \mathbf{I}/p)$ and $Y_q^* \sim \mathcal{N}(0, \mathbf{I}/q)$.) We then have the following straightforward extension of Theorem 2.

Lemma 6 (Transformational Invariance for Two Independent Variables). *For a two-dimensional $\varphi(x, y)$,*

$$
\begin{aligned}
-r' \log \mathbb{E}\big(\varphi^{1/r'}(X, Y)\big) &- \lambda h_p(X) - (1-\lambda)h_q(Y) \\
&= -r' \log \mathbb{E}\big(\varphi^{*1/r'}(X^*, Y^*)\big) - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*)
\end{aligned}
\tag{34}
$$

where

$$\varphi^*(x^*, y^*) = \varphi(T(x^*), U(y^*))|T'(x^*)|^\lambda |U'(y^*)|^{1-\lambda}. \tag{35}$$

Proof. From Equation (5) and the definition of λ, Equation (34) can be rewritten as

$$\frac{\mathbb{E}\big(\varphi^{*1/r'}(X^*, Y^*)\big)}{\|f^*\|_p \|g^*\|_q} = \frac{\mathbb{E}\big(\varphi^{1/r'}(X, Y)\big)}{\|f\|_p \|g\|_q}. \tag{36}$$

By the transformations T and U, the densities of the escort variables are related by $f_p^*(x^*) = f_p(T(x^*))|T'(x^*)|$ and $g_q^*(y^*) = g_q(U(y^*))|U'(y^*)|$. Now, by the same calculation as in the proof of Theorem 2,

$$
\begin{aligned}
\frac{\mathbb{E}\big(\varphi^{*1/r'}(X^*,Y^*)\big)}{\|f^*\|_p\|g^*\|_q} &= \frac{\mathbb{E}\big(\varphi^{1/r'}(T(X^*),U(Y^*))|T'(X^*)|^{1/p'}|U'(Y^*)|^{1/q'}\big)}{\|f^*\|_p\|g^*\|_q} \\
&= \frac{\int \varphi^{1/r'}(T(x^*),U(y^*))|T'(x^*)|^{1/p'}|U'(y^*)|^{1/q'}f^*(x^*)g^*(y^*)\,\mathrm{d}x^*\,\mathrm{d}y^*}{\|f^*\|_p\|g^*\|_q} \\
&= \int \varphi^{1/r'}(T(x^*),U(y^*))|T'(x^*)|^{1/p'}|U'(y^*)|^{1/q'}f_p^*(x^*)^{1/p}g_q^*(y^*)^{1/q}\,\mathrm{d}x^*\,\mathrm{d}y^* \\
&= \int \varphi^{1/r'}(T(x^*),U(y^*))|T'(x^*)||U'(y^*)|f_p(T(x^*))^{1/p}g_q(U(y^*))^{1/q}\,\mathrm{d}x^*\,\mathrm{d}y^* \\
&= \int \varphi^{1/r'}(x,y)f_p(x)^{1/p}g_q(y)^{1/q}\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{\mathbb{E}\big(\varphi^{1/r'}(X,Y)\big)}{\|f\|_p\|g\|_q}. \quad \square
\end{aligned}
\tag{37}
$$

Lemma 7. *Let φ be the density of $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$. Then,*

$$
\begin{aligned}
h_r(\sqrt{\lambda}X &+ \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\
&\geq -r'\log\mathbb{E}\Big\{\Big(\varphi_r(\sqrt{\lambda}T(X^*) + \sqrt{1-\lambda}U(Y^*)) \cdot |\lambda T'(X^*) + (1-\lambda)U'(Y^*)|\Big)^{1/r'}\Big\} \\
&\quad - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*).
\end{aligned}
\tag{38}
$$

Proof. By the equality case of Theorem 1 (see (16)), one has

$$
h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) = -r'\log\mathbb{E}\varphi_r^{1/r'}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y).
\tag{39}
$$

Now, by Lemma 6 applied to $\varphi(x,y) = \varphi_r(\sqrt{\lambda}x + \sqrt{1-\lambda}y)$, we have $\varphi^*(x^*,y^*) = \varphi_r(\sqrt{\lambda}T(x^*) + \sqrt{1-\lambda}U(y^*))|T'(x^*)|^\lambda|U'(y^*)|^{1-\lambda}$, and, therefore,

$$
\begin{aligned}
h_r(\sqrt{\lambda}X &+ \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\
&= -r'\log\mathbb{E}\Big\{\Big(\varphi_r(\sqrt{\lambda}T(X^*) + \sqrt{1-\lambda}U(Y^*)) \cdot |T'(X^*)|^\lambda|U'(Y^*)|^{1-\lambda}\Big)^{1/r'}\Big\} \\
&\quad - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*).
\end{aligned}
\tag{40}
$$

Since from Lemma 1, T and U can be chosen such that T' and U' are (lower) triangular with positive diagonal elements, it follows easily from the arithmetic-geometric mean inequality that

$$
|T'(X^*)|^\lambda|U'(Y^*)|^{1-\lambda} \leq |\lambda T'(X^*) + (1-\lambda)U'(Y^*)|.
\tag{41}
$$

The result follows at once (for either positive or negative r'). $\quad\square$

We can now use the normal rotation Lemma 2 to conclude by proving the following,

Theorem 3 (Rényi EPI [3]). *For independent X, Y and exponent triple $(p, q, r)_\lambda$,*

$$
h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \geq \frac{n}{2}r'\Big(\frac{\log r}{r} - \frac{\log p}{p} - \frac{\log q}{q}\Big)
\tag{42}
$$

with equality if and only if X, Y are i.i.d. normal.

Proof. If X, Y are i.i.d. normal, then $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$ is also identically distributed as X and Y, and from Lemma 4, it is immediate to check that equality holds (irrespective of their covariances). Therefore, the inequality in Equation (42) is equivalent to

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \geq h_r(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^*) - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*) \quad (43)$$

where X^*, Y^* are, say, i.i.d. standard normal $\mathcal{N}(0, \mathbf{I})$.

To prove Equation (43), consider the normal rotation of Lemma 2 and write X^*, Y^* in terms of $\widetilde{X}, \widetilde{Y}$ using Equation (2) in the first term of the r.h.s. of Equation (38) (Lemma 7). One obtains:

$$\begin{aligned} h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) &- \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &\geq -r' \log \mathbb{E}\left\{ \left(\psi(\widetilde{X}|\widetilde{Y})\right)^{1/r'} \right\} - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*), \end{aligned} \quad (44)$$

where

$$\begin{aligned} \psi(\widetilde{x}|\widetilde{y}) =& \varphi_r(\sqrt{\lambda}T(\sqrt{\lambda}\widetilde{x} - \sqrt{1-\lambda}\widetilde{y}) + \sqrt{1-\lambda}U(\sqrt{1-\lambda}\widetilde{x} + \sqrt{\lambda}\widetilde{y})) \\ &\times |\lambda T'(\sqrt{\lambda}\widetilde{x} - \sqrt{1-\lambda}\widetilde{y}) + (1-\lambda)U'(\sqrt{1-\lambda}\widetilde{x} + \sqrt{\lambda}\widetilde{y})|. \end{aligned} \quad (45)$$

Making the change of variable $z = \sqrt{\lambda}T(\sqrt{\lambda}\widetilde{x} - \sqrt{1-\lambda}\widetilde{y}) + \sqrt{1-\lambda}U(\sqrt{1-\lambda}\widetilde{x} + \sqrt{\lambda}\widetilde{y})$, one obtains

$$\int \psi(\widetilde{x}|\widetilde{y}) \, d\widetilde{x} = \int \varphi_r(z) \, dz = 1, \quad (46)$$

since φ_r is a density. Hence, $\psi(\widetilde{x}|\widetilde{y})$ is also a density in $\widetilde{x}$ for fixed $\widetilde{y}$. Now, since by Lemma 2, $\widetilde{X}$ and $\widetilde{Y}$ are independent, by the conditional information inequality in Equation (18) of Corollary 1, one has

$$-r' \log \mathbb{E}\left\{ \left(\psi(\widetilde{X}|\widetilde{Y})\right)^{1/r'} \right\} \geq h_r(\widetilde{X}) = h_r(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^*). \quad (47)$$

Combining with Equation (44) yields the announced inequality in Equation (43).

It remains to settle the equality case in Equation (43). From the above proof, equality holds in Equation (43) if and only if both Equation (41) and Equation (47) are equalities. Equality in Equation (41) holds if and only if for all $i = 1, 2, \ldots, n$,

$$\frac{\partial T_i}{\partial x_i}(X^*) = \frac{\partial U_i}{\partial y_i}(Y^*) \text{ a.s.} \quad (48)$$

Since X^* and Y^* are independent normal variables, this implies that $\frac{\partial T}{\partial x_i}$ and $\frac{\partial U}{\partial y_i}$ are constant and equal. In particular the Jacobian $|\lambda T'(\sqrt{\lambda}\widetilde{x} - \sqrt{1-\lambda}\widetilde{y}) + (1-\lambda)U'(\sqrt{1-\lambda}\widetilde{x} + \sqrt{\lambda}\widetilde{y})|$ in (45) is constant.

From Corollary 1 equality in Equation (47) holds if and 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 $j = 1, 2, \ldots, n$,

$$-\sqrt{\lambda}\sqrt{1-\lambda}\frac{\partial T_i}{\partial x_j}(\sqrt{\lambda}\widetilde{X} - \sqrt{1-\lambda}\widetilde{Y}) + \sqrt{\lambda}\sqrt{1-\lambda}\frac{\partial U_i}{\partial x_j}(\sqrt{1-\lambda}\widetilde{X} + \sqrt{\lambda}\widetilde{Y}) = 0 \quad (49)$$

which implies

$$\frac{\partial T_i}{\partial x_j}(X^*) = \frac{\partial U_i}{\partial y_j}(Y^*) \text{ a.s.} \quad (50)$$

for all $i, j = 1, 2, \ldots, n$. Therefore, T and U are linear transformations, equal up to an additive constant (equal to 0 since all variables are assumed of zero mean). It follows that $X_p = T(X_p^*)$ and $Y_q = U(Y_q^*)$

are normal 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. normal $\mathcal{N}(0, \mathbf{K})$.　□

A straightforward generalization to several independent variables is the following.

Corollary 2 (Rényi EPI for Several Variables). *Let* $r_1, r_2, \ldots, r_m, r$ *be exponents; those conjugates* $r'_1, r'_2, \ldots, r'_m, r'$ *are of the same sign and satisfy*

$$\sum_{i=1}^{m} \frac{1}{r'_i} = \frac{1}{r'} \tag{51}$$

and let $\lambda_1, \lambda_2, \ldots, \lambda_m$ *be defined by*

$$\lambda_i = \frac{r'}{r'_i} \qquad (i = 1, 2, \ldots, m). \tag{52}$$

Then, for independent $X_1, X_2, \ldots, X_m$,

$$h_r\Big(\sum_{i=1}^{m} \sqrt{\lambda_i} X_i\Big) - \sum_{i=1}^{m} \lambda_i h_{r_i}(X_i) \geq \frac{n}{2} r'\Big(\frac{\log r}{r} - \sum_{i=1}^{m} \frac{\log r_i}{r_i}\Big) \tag{53}$$

with equality if and only if the X_i *are i.i.d. normal.*

Proof. By induction on m: The result for $m = 2$ is Theorem 3. Suppose the result satisfied at order $m - 1$ and let $Y_m = \sum_{i=1}^{m-1} \sqrt{\lambda_i} X_i / \sqrt{1 - \lambda_m}$ and s_m be such that $\frac{1}{s'_m} = \sum_{i=1}^{m-1} \frac{1}{r'_i}$. Notice that $\frac{1}{r'} = \frac{1}{r'_m} + \frac{1}{s'_m} = \frac{\lambda_m}{r'} + \frac{1}{s'_m}$, hence $r' = (1 - \lambda_m) s'_m$. By Theorem 3, $h_r\big(\sum_{i=1}^{m} \sqrt{\lambda_i} X_i\big) = h_r(\sqrt{\lambda_m} X_m + \sqrt{1 - \lambda_m} Y_m) \geq \lambda_m h_{r_m}(X_m) + (1 - \lambda_m) h_{s_m}(Y_m) + \frac{n}{2} r'\big(\frac{\log r}{r} - \frac{\log r_m}{r_m} - \frac{\log s_m}{s_m}\big)$ with equality if and only if X_m, Y_m are i.i.d. normal. Now, by the induction hypothesis, $h_{s_m}(Y_m) \geq \sum_{i=1}^{m-1} \frac{\lambda_i}{1 - \lambda_m} h_{r_i}(X_i) + \frac{n}{2} s'_m\big(\frac{\log s_m}{s_m} - \sum_{i=1}^{m-1} \frac{\log r_i}{r_i}\big)$ with equality if and only if the X_i ($i = 1, 2, \ldots, m - 1$)—and hence Y_m—are i.i.d. normal. The result at order m follows by combining the two inequalities since $(1 - \lambda_m) s'_m = r'$.　□

5. Recent Versions of the Rényi EPI

Definition 5 (Rényi Entropy Power [13]). *The Rényi entropy power of order r is defined by*

$$N_r(X) = e^{2h_r(X)/n}. \tag{54}$$

Up to a multiplicative constant, $N_r(X)$ is the (average) power of a white normal variable having the same Rényi entropy as X—hence the name "entropy power". In fact, if $X^* \sim \mathcal{N}(0, \sigma^2)$ has the same Rényi entropy $h_r(X^*) = h_r(X)$, then by Lemma 4,

$$\sigma^2 = \frac{e^{2h_r(X)/n}}{2\pi r^{r'/r}}. \tag{55}$$

The Renyi entropy power enjoys the same scaling property as for the usual power: By Lemma 3, for any $a \in \mathbb{R}$,

$$N_r(aX) = a^2 N_r(X). \tag{56}$$

For independent $X_1, X_2, \ldots, X_m$, Rényi entropy power inequalities take either the form [13,14]

$$N_r\Big(\sum_{i=1}^{m} X_i\Big) \geq c \sum_{i=1}^{m} N_r(X_i) \tag{57}$$

for some positive constant c, or the form [15–17]

$$N_r^{\alpha}\left(\sum_{i=1}^{m} X_i\right) \geq \sum_{i=1}^{m} N_r^{\alpha}(X_i) \tag{58}$$

for some positive exponent α. The constants c and α may depend on the order r, the number m of variables and the dimension n. What is desired is:

- a maximum possible value of c in Equation (57) since the inequality is automatically satisfied for all positive constants $c' < c$.
- a minimum possible value of α in Equation (58) since the inequality is automatically satisfied for all positive exponents $\alpha' > \alpha$; in fact, since Equation (58) is homogeneous by scaling the variables $X_i \mapsto aX_i$ as in Equation (56), one may suppose without loss of generality that the r.h.s. of Equation (58) is $= 1$; then, $N_r(X_i) < 1$, hence $N_r^{\alpha'}(X_i) < N_r^{\alpha}(X_i)$ for all i and $1 = \sqrt[\alpha]{\sum_{i=1}^{m} N_r^{\alpha}(X_i)} \geq \sqrt[\alpha']{\sum_{i=1}^{m} N_r^{\alpha'}(X_i)}$.

The following useful characterization, which generalizes ([16] Lemma 2.1), makes the link between the various versions (Equations (53), (57), and (58)) of the Rényi entropy power inequality.

Lemma 8. *For independent $X_1, X_2, \ldots, X_m$, the Rényi EPI in the general form*

$$N_r^{\alpha}\left(\sum_{i=1}^{m} X_i\right) \geq c \sum_{i=1}^{m} N_r^{\alpha}(X_i) \tag{59}$$

for some constant $c > 0$ and exponent $\alpha > 0$ is equivalent to the following inequality

$$h_r\left(\sum_{i=1}^{m} \sqrt{\lambda_i} X_i\right) - \sum_{i=1}^{m} \lambda_i h_r(X_i) \geq \frac{n}{2}\left(\frac{\log c}{\alpha} + \left(\frac{1}{\alpha} - 1\right) H(\lambda)\right) \tag{60}$$

for any positive $\lambda_1, \lambda_2, \ldots, \lambda_m$ such that $\sum_{i=1}^{m} \lambda_i = 1$, where $H(\lambda) > 0$ denotes the discrete entropy

$$H(\lambda) = \sum_{i=1}^{m} \lambda_i \log \frac{1}{\lambda_i} > 0. \tag{61}$$

Proof. Suppose Equation (59) holds. Then,

$$h_r\left(\sum_{i=1}^{m} \sqrt{\lambda_i} X_i\right) = \frac{n}{2\alpha} \log N_r^{\alpha}\left(\sum_{i=1}^{m} \sqrt{\lambda_i} X_i\right) \tag{62}$$

$$\geq \frac{n}{2\alpha} \log \sum_{i=1}^{m} N_r^{\alpha}(\sqrt{\lambda_i} X_i) + \frac{n}{2\alpha} \log c \tag{63}$$

$$= \frac{n}{2\alpha} \log \sum_{i=1}^{m} \lambda_i^{\alpha} N_r^{\alpha}(X_i) + \frac{n}{2\alpha} \log c \tag{64}$$

$$\geq \frac{n}{2\alpha} \sum_{i=1}^{m} \lambda_i \log\left(\lambda_i^{\alpha-1} N_r^{\alpha}(X_i)\right) + \frac{n}{2\alpha} \log c \tag{65}$$

$$= \sum_{i=1}^{m} \lambda_i h_r(X_i) + \frac{n(\alpha - 1)}{2\alpha} \sum_{i=1}^{m} \lambda_i \log \lambda_i + \frac{n}{2\alpha} \log c \tag{66}$$

where the scaling property in Equation (56) is used in Equation (64) and the concavity of the logarithm is used in Equation (65). Conversely, suppose that Equation (60) 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,

$$N_r{}^\alpha \Big(\sum_{i=1}^m X_i \Big) = \exp \frac{2\alpha}{n} h_r \Big(\sum_{i=1}^m \sqrt{\lambda_i}\, \frac{X_i}{\sqrt{\lambda_i}} \Big) \tag{67}$$

$$\geq \exp \frac{2\alpha}{n} \sum_{i=1}^m \lambda_i h_r \Big(\frac{X_i}{\sqrt{\lambda_i}} \Big) \cdot c \cdot \exp(1-\alpha) \sum_{i=1}^m \lambda_i \log \frac{1}{\lambda_i} \tag{68}$$

$$= c \prod_{i=1}^m \Big(N_r{}^\alpha \Big(\frac{X_i}{\sqrt{\lambda_i}} \Big) \lambda_i^{\alpha-1} \Big)^{\lambda_i} \tag{69}$$

$$= c \prod_{i=1}^m \Big(N_r{}^\alpha (X_i) \lambda_i^{-1} \Big)^{\lambda_i} \tag{70}$$

$$= c \Big(\sum_{i=1}^m N_r{}^\alpha (X_i) \Big)^{\sum_{i=1}^m \lambda_i} \tag{71}$$

$$= c \sum_{i=1}^m N_r{}^\alpha (X_i). \quad \square \tag{72}$$

5.1. Rényi Entropy Power Inequalities for Orders >1

From Lemma 8 and Corollary 2, it is easy to recover known Rényi EPIs and obtain new ones for orders $r > 1$. In fact, if $r > 1$, then $r' > 0$ and all r'_i are positive and $> r'$. Therefore, all r_i are $< r$ and by monotonicity (Lemma 5),

$$h_{r_i}(X_i) \geq h_r(X_i) \qquad (i = 1, 2, \ldots, m). \tag{73}$$

Plugging this into Equation (53), one obtains

$$h_r \Big(\sum_{i=1}^m \sqrt{\lambda_i} X_i \Big) - \sum_{i=1}^m \lambda_i h_r(X_i) \geq \frac{n}{2} r' \Big(\frac{\log r}{r} - \sum_{i=1}^m \frac{\log r_i}{r_i} \Big) \tag{74}$$

where $\lambda_i = r'/r'_i$ for $i = 1, 2, \ldots, m$. For future reference, define

$$A(\lambda) = |r'| \Big(\frac{\log r}{r} - \sum_{i=1}^m \frac{\log r_i}{r_i} \Big) \tag{75}$$

$$= |r'| \Big(\sum_{i=1}^m (1 - \frac{\lambda_i}{r'}) \log(1 - \frac{\lambda_i}{r'}) - (1 - \frac{1}{r'}) \log(1 - \frac{1}{r'}) \Big). \tag{76}$$

(The absolute value $|r'|$ is needed in the next subsection 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 $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 0, 1)$ and since every λ is a convex combination of these vectors and $A(\lambda)$ is strictly convex, one has $A(\lambda) < 0$.

Theorem 4 (Ram and Sason [14]). *The Rényi EPI (57) holds for $r > 1$ and $c = r^{r'/r} \big(1 - \frac{1}{mr'} \big)^{mr'-1}$.*

Proof. By Lemma 8 for $\alpha = 1$ we only need to check that the r.h.s. of Equation (74) 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, Equation (57) will hold for $\log c = \min_\lambda A(\lambda)$. Now, by the log-sum inequality ([4] Theorem 2.7.1),

$$\sum_{i=1}^m \frac{1}{r_i} \log \frac{1}{r_i} \geq \Big(\sum_{i=1}^m \frac{1}{r_i} \Big) \log \frac{\sum_{i=1}^m \frac{1}{r_i}}{m} = (m - 1/r') \log \frac{m - 1/r'}{m} \tag{77}$$

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

An alternate proof is to argue that $A(\lambda)$ is convex and symmetrical in $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ and is, therefore, minimized when all λ_i are equal.

Remark 5. *The above constant c is certainly not optimal since equality in Equation* (73) *holds if and only if the X_i are uniformly distributed (Lemma* 5*) while equality in Equation* (53) *holds if and only if the X_i are identically normally distributed (Corollary* 2*). Ram and Sason [14] tightened Equation* (57) *further using optimization techniques, resulting in a constant that depends on the relative values of the entropy powers themselves.*

Remark 6. *It can be noted that* $\log c = r'\frac{\log r}{r} + (mr' - 1)\log\left(1 - \frac{1}{mr'}\right) < 0$ *decreases (and tends to* $r'\frac{\log r}{r} - 1$) *as m increases; in fact* $\frac{\partial \log c}{\partial m} = r'\log\left(1 - \frac{1}{mr'}\right) + \frac{mr'}{r'm^2} < r'\left(-\frac{1}{mr'}\right) + \frac{1}{m} = 0$. *Thus, a universal constant independent of m is obtained by taking*

$$c = \inf_m r^{r'/r}\left(1 - \frac{1}{mr'}\right)^{mr' - 1} = r^{r'/r}\lim_{m \to \infty}\left(1 - \frac{1}{mr'}\right)^{mr' - 1} = \frac{r^{r'/r}}{e} \tag{78}$$

This was the constant established by Bobkov and Chistyakov [13].

Theorem 5. *The Rényi EPI* (58) *holds for $r > 1$ and* $\alpha = \left(1 + r'\frac{\log_2 r}{r} + (2r' - 1)\log_2\left(1 - \frac{1}{2r'}\right)\right)^{-1}$ *and this value of α cannot be improved using the method of this paper by making it depend on m.*

Proof. By Lemma 8, for $c = 1$, we only need to check that the r.h.s. of Equation (74) 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 $\sum_{i=1}^m \frac{1}{r_i'} = \frac{1}{r'}$. Thus, Equation (58) will hold for $\frac{1}{\alpha} - 1 = \min_\lambda \frac{A(\lambda)}{H(\lambda)}$. By the proof of the preceding theorem, the numerator is minimized when all λ_i are equal and this also maximizes the entropy $= \log m$ in the denominator. However, one cannot conclude yet since the minimum in the numerator is negative.

A stationary point is easily obtained by the Lagrangian method, which implies that $\frac{\partial}{\partial \lambda_i}\frac{A(\lambda)}{H(\lambda)}$ is constant independent of i. This gives that $\frac{A(\lambda)}{H(\lambda)}\log \lambda_i - \log(1 - \lambda_i/r')$ is constant, hence a stationary point is obtained when all λ_i are equal (to $1/m$) and the corresponding value of $A(\lambda)/H(\lambda)$ is $\left(r'\frac{\log r}{r} + (mr' - 1)\log\left(1 - \frac{1}{mr'}\right)\right)/\log m$.

However, the boundary of the domain of $A(\lambda)/H(\lambda)$ is the simplex $\{\sum_i \lambda_i = 1, \lambda_i \geq 0\}$ where on each vertex joining two standard unit vectors, $A(\lambda)/H(\lambda)$ has the same expression as for $m = 2$. Li [16] showed—this is also easily proved using [17, Lemma 8]—that, for $m = 2$, the minimum is obtained when $\lambda = (1/2, 1/2)$. The corresponding value of $A(\lambda)/H(\lambda)$ is $\left(r'\frac{\log r}{r} + (2r' - 1)\log\left(1 - \frac{1}{2r'}\right)\right)/\log 2$, which is easily seen to be less than $\left(r'\frac{\log r}{r} + (mr' - 1)\log\left(1 - \frac{1}{mr'}\right)\right)/\log m$ for any $m \geq 2$.

Therefore, the minimum of $A(\lambda)/H(\lambda)$ is attained at the boundary when all λ_i are zero except two of them equal to $1/2$. This gives $\frac{1}{\alpha} - 1 = \left(r'\frac{\log r}{r} + (2r' - 1)\log\left(1 - \frac{1}{2r'}\right)\right)/\log 2$. $\square$

Remark 7. *The case $m = 2$ yields* $\alpha = \frac{r - 1}{(r+1)\log_2(r+1) - r\log_2 r - 2}$ *which was found by Li [16] who remarked that this value of α is strictly smaller (better) than the value* $\alpha = \frac{r+1}{2}$ *obtained by Bobkov and Marsiglietti [15].*

Interestingly, for $m > 2$, the exponent of Theorem 5 *cannot be improved by this method. In fact, in the above proof, it is easily seen that* $\left(r'\frac{\log r}{r} + (mr' - 1)\log\left(1 - \frac{1}{mr'}\right)\right)/\log m$ *is negative and increases toward 0 as m increases. Therefore, the exponent α cannot be decreased (improved) as m increases.*

The above value of α is > 1. However, using the same method, it is easy to obtain Rényi EPIs with exponent values $\alpha < 1$. This is given by the following Theorem 6.

Theorem 6. *The Rényi EPI* (59) *holds for* $r > 1, 0 < \alpha < 1$ *with* $c = \left(m \, r^{r'/r} \left(1 - \frac{1}{mr'} \right)^{mr'-1} \right)^{\alpha} / m$.

Proof. By Lemma 8 we only need to check that the r.h.s. of Equation (74) is greater than $\frac{n}{2} \left((\log c)/\alpha + (1/\alpha - 1)H(\lambda) \right)$, 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 $0 < \alpha < 1$, Equation (59) 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 $\log c = \alpha \left(r' \frac{\log r}{r} + (mr' - 1) \log \left(1 - \frac{1}{mr'} \right) \right) - (1 - \alpha) \log m$. □

5.2. Rényi Entropy Power Inequalities 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 $> r$ and by monotonicity (Lemma 5), the opposite inequality of Equation (73) holds and the method of the preceding subsection fails. For log-concave densities, however, Equation (73) can be replaced by a similar inequality in the right direction.

Definition 6 (Log-Concave Density). *A density f is log-concave if $\log f$ is concave in it support, i.e., for all $0 < \mu < 1$,*

$$f(x)^{\mu} f(y)^{1-\mu} \leq f(\mu x + (1 - \mu)y). \tag{79}$$

Lemma 9. *If X has a log-concave density, then $h_p(pX) - ph_p(X) = n \log p + (1 - p)h_p(X)$ is concave in p.*

As noted below in Corollary 3, this is essentially a result obtained by Fradelizi, Madiman and Wang [28]. The following alternate proof uses the transport properties seen in Section 4.

Proof. Define $r = \lambda p + (1 - \lambda)q$ where $0 < \lambda < 1$. By Lemma 1 there exists two diffeomorphisms T, U such that one can write $pX_p = T(X^*)$ and $qX_q = U(X^*)$. Then, X^* has density

$$\frac{1}{p^n} f_p \left(\frac{T(x^*)}{p} \right) |T'(x^*)| = \frac{1}{q^n} f_q \left(\frac{U(x^*)}{q} \right) |U'(x^*)| = \frac{1}{p^{\lambda n} q^{(1-\lambda)n}} f_p \left(\frac{T(x^*)}{p} \right)^{\lambda} f_q \left(\frac{U(x^*)}{q} \right)^{1-\lambda} |T'(x^*)|^{\lambda} |U'(x^*)|^{1-\lambda}. \tag{80}$$

Now, by log-concavity, Equation (79) with $\mu = \lambda p / r$,

$$f_p \left(\frac{T(x^*)}{p} \right)^{\lambda} f_q \left(\frac{U(x^*)}{q} \right)^{1-\lambda} = \frac{1}{\|f\|_p^{\lambda p} \|f\|_q^{(1-\lambda)q}} f \left(\frac{T(x^*)}{p} \right)^{\lambda p} f \left(\frac{U(x^*)}{q} \right)^{(1-\lambda)q} \tag{81}$$

$$\leq \frac{1}{\|f\|_p^{\lambda p} \|f\|_q^{(1-\lambda)q}} f \left(\frac{\lambda T(x^*) + (1-\lambda)U(x^*)}{r} \right)^{r} \tag{82}$$

$$= \frac{\|f\|_r^r}{\|f\|_p^{\lambda p} \|f\|_q^{(1-\lambda)q}} f_r \left(\frac{\lambda T(x^*) + (1-\lambda)U(x^*)}{r} \right). \tag{83}$$

Using the arithmetic-geometric mean inequality in Equation (41) and integrating the density in Equation (80) over $x^* \in \mathbb{R}^n$, one obtains

$$(p^{\lambda} q^{1-\lambda})^n \|f\|_p^{\lambda p} \|f\|_q^{(1-\lambda)q} \leq r^n \|f\|_r^r. \tag{84}$$

Taking the logarithm yields the announced concavity. □

As a side result, it is interesting to note that we have obtained a simple transportation proof of the following varentropy bound:

Corollary 3 (Varentropy Bound [28]). *One has $\mathrm{Var} \log f(X_p) \leq n/p^2$, that is, $\mathrm{Var} \log f_p(X_p) \leq n$.*

Proof. Since $n \log p + (1-p)h_p(X)$ is concave, one has $\frac{\partial^2}{\partial p^2}\big(n \log p + (1-p)h_p(X)\big) \leq 0$, that is,

$$\frac{\partial^2}{\partial p^2}\big((1-p)h_p(X)\big) \leq \frac{n}{p^2}. \tag{85}$$

Differentiating twice using Leibniz's product rule and plugging the identity in Equation (12), the l.h.s. of this inequality becomes

$$-\frac{1}{1-p}\frac{\partial D(f_p\|f)}{\partial p} = \frac{\partial}{\partial p}\int f_p \log f = \int f_p \log^2 f - \Big(\int f_p \log f\Big)^2. \quad \square \tag{86}$$

As another easy consequence of Lemma 9, since $n \log p + (1-p)h_p(X)$ is concave and vanishes for $p = 1$, the slopes $\frac{n \log p + (1-p)h_p(X) - 0}{p - 1}$ are non-increasing in p. In other words, $h_p(X) + n\frac{\log p}{1-p}$ is nondecreasing. Therefore:

Corollary 4 (Marsiglietti and Melbourne [17])**.** *If* $p < q$ *then for any* X *with log-concave density,* $h_p(X) + n\frac{\log p}{1-p} \leq h_q(X) + n\frac{\log q}{1-q}$.

We can now use Lemma 8 and Corollary 2 to obtain Rényi EPIs for orders $r < 1$. Since all r_i are $> r$, by Corollary 4,

$$h_{r_i}(X) + n\frac{\log r_i}{1 - r_i} \geq h_r(X) + n\frac{\log r}{1 - r} \qquad (i = 1, 2, \ldots, m). \tag{87}$$

Plugging this into Equation (53), one obtains

$$h_r\Big(\sum_{i=1}^m \sqrt{\lambda_i}X_i\Big) - \sum_{i=1}^m \lambda_i h_r(X_i) \geq n\Big(\frac{\log r}{1-r} - \sum_{i=1}^m \lambda_i \frac{\log r_i}{1-r_i}\Big) + \frac{n}{2}r'\Big(\frac{\log r}{r} - \sum_{i=1}^m \frac{\log r_i}{r_i}\Big) \tag{88}$$

$$= \frac{n}{2}r'\Big(\sum_{i=1}^m \frac{\log r_i}{r_i} - \frac{\log r}{r}\Big) \tag{89}$$

where we have used that $\lambda_i = r'/r_i'$ for $i = 1, 2, \ldots, m$. Notice that the r.h.s. of Equation (89) for $r < 1$ ($r' < 0$) is the opposite of that of Equation (74) 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 $r > 1$.

Theorem 7. *The Rényi EPI (57) for log-concave densities holds for* $c = r^{-r'/r}\big(1 - \frac{1}{mr'}\big)^{1-mr'}$ *and* $r < 1$.

Proof. It is identical to that of Theorem 4 except for the change $|r'| = -r'$ in the expression of $A(\lambda)$. $\quad \square$

Theorem 8. *The Rényi EPI (58) for log-concave densities holds for* $r < 1$ *and* $\alpha = \Big(1 + |r'|\frac{\log_2 r}{r} + (2|r'|+1)\log_2\big(1 + \frac{1}{2|r'|}\big)\Big)^{-1}$ *and this value of* α *cannot be improved using the method of this paper by making it depend on* m.

Proof. It is identical to that of Theorem 5 except for the change $|r'| = -r'$ in the expression of $A(\lambda)$. $\quad \square$

Remark 8. *The case* $m = 2$ *yields* $\alpha = \frac{1-r}{(r+1)\log_2(r+1) - r\log_2 r - 2r}$ *which was found by Marsiglietti and Melbourne [17]. Again, the exponent of the theorem is not improved for* $m > 2$.

Theorem 9. *The Rényi EPI in Equation (59) for log-concave densities holds for* $c = \Big(mr^{-r'/r}\big(1 - \frac{1}{mr'}\big)^{1-mr'}\Big)^\alpha/m$ *where* $r < 1$ *and* $0 < \alpha < 1$.

Proof. It is identical to that of Theorem 6 except for the change $|r'| = -r'$ in the expression of $A(\lambda)$. □

6. Conclusions

This article provides a comprehensive framework to derive known Rényi entropy power inequalities (with shorter proofs), and prove new ones. The framework is based on a transport argument from normal densities and a change of variable by rotation. Only basic properties of Rényi entropies are used in the proofs.

In particular, the α-modification of the EPI is recovered for two or more independent variables for Rényi entropy orders >1 as well as for orders <1, and the Rényi EPI with multiplicative constant c is extended to Rényi entropy orders <1. In addition, a more general formulation with both exponent α and constant c is obtained for all orders. In passing, a simple proof using normal transport of a recent sharp varentropy bound was obtained for log-concave densities.

As a perspective, the methods developed in this paper can perhaps be generalized to obtain reverse Rényi entropy power inequalities (see, e.g., the discussion in [16]).

Conflicts of Interest: The author declares no conflict of interest.

References

1. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [CrossRef]
2. Rioul, O. Information Theoretic Proofs of Entropy Power Inequalities. *IEEE Trans. Inf. Theory* **2011**, *57*, 33–55. [CrossRef]
3. Dembo, A.; Cover, T.M.; Thomas, J.A. Information Theoretic Inequalities. *IEEE Trans. Inf. Theory* **1991**, *37*, 1501–1518. [CrossRef]
4. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
5. Rényi, A. On Measures of Information and Entropy. In Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1960; University of California Press: Berkeley, CA, USA, 1960; Volume 1, pp. 547–561.
6. Campbell, L.L. A Coding Theorem and Rényi's Entropy. *Inf. Control* **1965**, *8*, 423–429. [CrossRef]
7. Ben-Bassat, M.; Raviv, J. Rényi's Entropy and the Probability of Error. *IEEE Trans. Inf. Theory* **1978**, *24*, 324–331. [CrossRef]
8. Arimoto, S. Information Measures and Capacity of Order α for Discrete Memoryless Channels. In *Topics in Information Theory, Colloquia mathematica Societatis János Bolyai 16*; Csiszár, I., Elias, P., Eds.; North Holland: Amsterdam, The Netherlands, 1977; pp. 41–52.
9. Arikan, E. An Inequality on Guessing and its Application to Sequential Decoding. *IEEE Trans. Inf. Theory* **1996**, *42*, 99–105. [CrossRef]
10. Erdogmus, D.; Hild, K.E.; Principe, J.C.; Lazaro, M.; Santamaria, I. Adaptive Blind Deconvolution of Linear Channels Using Rényi's Entropy with Parzen Window Estimation. *IEEE Trans. Signal Process.* **2004**, *52*, 1489–1498. [CrossRef]
11. Savaré, G.; Toscani, G. The Concavity of Rényi Entropy Power. *IEEE Trans. Inf. Theory* **2014**, *60*, 2687–2693. [CrossRef]
12. Madiman, M.; Melbourne, J.; Xu, P. Forward and Reverse Entropy Power Inequalities in Convex Geometry. In *Convexity and Concentration*; Carlen, E., Madiman, M., Werner, E., Eds.; IMA Volumes in Mathematics and its Applications; Springer: Berlin, Germany, 2017; Volume 161, pp. 427–485.
13. Bobkov, S.G.; Chistyakov, G.P. Entropy Power Inequality for the Rényi Entropy. *IEEE Trans. Inf. Theory* **2015**, *61*, 708–714. [CrossRef]
14. Ram, E.; Sason, I. On Rényi Entropy Power Inequalities. *IEEE Trans. Inf. Theory* **2016**, *62*, 6800–6815. [CrossRef]
15. Bobkov, S.G.; Marsiglietti, A. Variants of the Entropy Power Inequality. *IEEE Trans. Inf. Theory* **2017**, *63*, 7747–7752. [CrossRef]
16. Li, J. Rényi Entropy Power Inequality and a Reverse. *Stud. Math.* **2018**, *242*, 303–319. [CrossRef]

17. Marsiglietti, A.; Melbourne, J. On the Entropy Power Inequality for the Rényi Entropy of Order $[0, 1]$. *arxiv* **2018**, arXiv:1710.00800.

18. Rioul, O. Yet Another Proof of the Entropy Power Inequality. *IEEE Trans. Inf. Theory* **2017**, *63*, 3595–3599. [CrossRef]

19. Rosenblatt, M. Remarks on a Multivariate Transformation. *Ann. Math. Stat.* **1952**, *23*, 470–472. [CrossRef]

20. Knothe, H. Contributions to the theory of convex bodies. *Mich. Math. J.* **1957**, *4*, 39–52. [CrossRef]

21. Rioul, O. Optimal Transportation to the Entropy-Power Inequality. In Proceedings of the IEEE Information Theory and Applications Workshop (ITA 2017), San Diego, CA, USA, 12–17 February 2017.

22. Bryc, W. *The Normal Distribution—Characterizations with Applications*; Lecture Notes in Statistics; Springer: Berlin, Germany, 1995.

23. Rioul, O. Optimal Transport to Rényi Entropies. In Proceedings of the 3rd Conference on Geometric Science of Information (GSI 2017), Paris, France, 7–9 November 2017; Lecture Notes in Computer Science; Springer: Berlin, Germany, 2017.

24. Bercher, J.F. Source Coding with Escort Distributions and Rényi Entropy Bounds. *Phys. Lett. A* **2009**, *373*, 3235–3238. [CrossRef]

25. Bercher, J.F. On Generalized Cramér-Rao Inequalities, Generalized Fisher Information and Characterizations of Generalized q-Gaussian Distributions. *J. Phys. A Math. Theor.* **2012**, *45*, 255303. [CrossRef]

26. Beck, C.; Schlögl, F. *Thermodynamics of Chaotic Systems: An Introduction*; Cambridge University Press: Cambridge, UK, 1993.

27. Bercher, J.F. (ESIEE Paris, Marne-la-Vallée, France). Private communication, 2018.

28. Fradelizi, M.; Madiman, M.; Wang, L. Optimal Concentration of Information Content for Log-Concave Densities. In *High Dimensional Probability VII: The Cargèse Volume*; Houdré, C., Mason, D.M., Reynaud-Bouret, P., Rosiński, J., Eds.; Birkhäuser: Basel, Switzerland, 2016.