Beyond the entropy power inequality, via rearrangements

Liyao Wang and Mokshay Madiman, Member, IEEE

Abstract

A lower bound on the Rényi differential entropy of a sum of independent random vectors is demonstrated in terms of rearrangements. For the special case of Boltzmann-Shannon entropy, this lower bound is better than that given by the entropy power inequality. Several applications are discussed, including a new proof of the classical entropy power inequality and an entropy inequality involving symmetrization of Lévy processes.

Index Terms

Entropy power inequality; spherically symmetric rearrangement; Rényi entropy; majorization.

I. INTRODUCTION

Rearrangement is a natural and powerful notion in functional analysis, and finds use in proving many interesting inequalities. For instance, the original proofs of Young’s inequality with sharp constant (which, as is well known from [24], is a common generalization of the Brunn-Minkowski and entropy power inequalities) rely on rearrangements [4], [14]. A basic property of rearrangements is that they preserve $L^p$ norms; thus, in particular, the rearrangement of a probability density function is also a probability density function.

Our main contribution in this note is a new lower bound on the Rényi (differential) entropy of a sum of independent random vectors taking values in $\mathbb{R}^n$, for some fixed natural number $n$. Recall that for $p \in (0, 1) \cup (1, +\infty)$, the Rényi entropy of a probability density $f$ is defined as:

$$h_p(f) = \frac{1}{1-p} \log \left( \int_{\mathbb{R}^n} f^p(x) dx \right).$$

For $p = 1$, $h_1(f)$ is defined as the Shannon differential entropy

$$h(f) = -\int f(x) \log f(x) dx,$$

and for $p = 0, \infty$, it is defined in a limiting sense (see Lemma V.1 for details).

This new bound is expressed in terms of rearrangements, which we define and recall basic properties of in Section II.

Main Theorem. If $f_i$ are densities on $\mathbb{R}^n$, and $f_i^*$ are their spherically symmetric rearrangements,

$$h_p(f_1 \ast f_2 \ast \cdots \ast f_k) \geq h_p(f_1^* \ast f_2^* \ast \cdots \ast f_k^*),$$

for any $p \in [0, 1) \cup (1, \infty]$. For $p = 1$,

$$h(f_1 \ast f_2 \ast \cdots \ast f_k) \geq h(f_1^* \ast f_2^* \ast \cdots \ast f_k^*),$$

provided both sides are well defined.

If we write $X_i^*$ for a random vector drawn from the density $f_i^*$, and assume that all random vectors are drawn independently of each other, the Main Theorem says in more customary information-theoretic notation that

$$h_p(X_1 + \ldots + X_k) \geq h_p(X_1^* + \ldots + X_k^*)$$

for each $p \geq 1$.

Let us note that the special case of the Main Theorem corresponding to $p \in (0, 1)$ and $k = 2$ is implicit in [14, Proposition 9]. However, our extension includes the three most interesting values of $p$ (namely, 0, 1, and $\infty$), and also covers arbitrary positive integers $k$. Indeed, as we will discuss, the $p = 0$ case yields the Brunn-Minkowski inequality, the $p = \infty$ case yields

\[This research was supported by the U.S. National Science Foundation through the grants DMS-1409504 and CCF-1346564.

\[Portions of this paper were presented at the 2013 Information Theory and Applications Workshop in San Diego, and at the 2013 IEEE International Symposium on Information Theory in Istanbul.

L. Wang is with the Department of Physics, Yale University, 217 Prospect Street, New Haven, CT 06511, USA. Email: liyao.wang@yale.edu

M. Madiman is with the Department of Mathematical Sciences, University of Delaware, 517B Ewing Hall, Newark, DE 19716, USA. Email: madiman@udel.edu

arXiv:1307.6018v2 [cs.IT] 19 May 2014
as a corollary an inequality due to Riesz and Sobolev, and the $p = 1$ case for the Shannon entropy is new and the most interesting from an information-theoretic point of view.

In order to make the comparison with the classical Shannon-Stam entropy power inequality, we state the following standard version of it [23], focusing on real-valued random variables for simplicity.

**Theorem I.1.** [44], [46] Let $X_1$ and $X_2$ be two independent $\mathbb{R}$-valued random variables with finite differential entropies, and finite variance. Let $Z_1$ and $Z_2$ be two independent Gaussians such that

$$h(X_i) = h(Z_i), \quad i = 1, 2.$$ Then

$$h(X_1 + X_2) \geq h(Z_1 + Z_2).$$

We also need the following lemmata, which we could not find explicitly stated in the literature. (The proofs are not difficult, and given in later sections.)

**Lemma I.2.** If one of $h(X)$ and $h(X^*)$ is well defined, then so is the other one and we have

$$h(X) = h(X^*).$$

** Lemma I.3.** For any real random variable $X$,

$$\text{Var}(X^*) \leq \text{Var}(X).$$

First note that from Lemma I.2, it follows that

$$h(X^*_i) = h(X_i) = h(Z_i), \quad i = 1, 2.$$ Furthermore, if $X_1$ and $X_2$ have finite variance, then Lemma I.3 implies that $X^*_1$ and $X^*_2$ have finite variance, and therefore by the usual entropy power inequality (i.e., Theorem I.1), we have that

$$h(X^*_1 + X^*_2) \geq h(Z_1 + Z_2).$$ (3)

On the other hand, the Main Theorem gives

$$h(X_1 + X_2) \geq h(X^*_1 + X^*_2).$$ (4)

From (4) and (3), we see that we have inserted the quantity $h(X^*_1 + X^*_2)$ between the two sides of the entropy power inequality as stated in Theorem I.1. In this sense, the $p = 1$ and $k = 2$ case is a kind of strengthening of Theorem I.1.

Let us outline how this note is organized. Section II describes basic and well known facts about rearrangements in a self-contained fashion. Section III discusses a result related to the Main Theorem but for relative entropy (or more generally, Rényi divergence) rather than entropy. In Section IV, we discuss connections of our result to a recent Rényi entropy power inequality for independent random vectors due to Bobkov and Chistyakov [9].

Then we give two related proofs of the Main Theorem, both of which are based on the Rogers-Brascamp-Lieb-Luttinger inequality. The first proof based on continuity considerations for Rényi entropy in the order is described in Sections V and VI. We include the first proof mainly because along the way, it clarifies various points that may be considered folklore (in particular, the continuity of Rényi entropy in the order, which has sometimes been taken for granted in the literature leading to incomplete statements of technical assumptions).

The second proof based on majorization ideas is simpler and more general, and described in Section VII. Our approach here was inspired by slides of a 2008 talk of Bruce Hajek that we found online (after a draft of this paper was written with just the first proof). Based on comments we received after the first draft of this paper was posted online, it appears that the majorization-based approach to rearrangements is well known to experts though there does not seem to be a clear exposition of it anywhere; while its roots may be considered to lie implicitly in the famed text of Hardy, Littlewood and Polya [30], it was significantly taken forward in a difficult paper of Alvino, Trombetti and Lions [1]. As a result, a central technical result of this paper—Theorem VII.4—may not be very surprising to experts. In fact, after the first draft of this paper was circulated, it came to our attention that when $\varphi$ in Theorem VII.4 is non-negative, the $k = 2$ case is Corollary 1 in Section 3.3 of Burchard’s dissertation [16], where also the equality case is characterized (this is much more difficult than the inequality itself). For non-negative $\varphi$ and general $k$, Theorem VII.4 is proved in Corollary 3 in Section 3.4 of [16].

However, the main innovation in Theorem VII.4 is the extension to general convex functions and the streamlined development using majorization that yields at one go a unified proof of the Main Theorem for all $p$. It is pertinent to note that the most interesting case of the Main Theorem, namely for Shannon differential entropy, would not follow from the earlier results. In
the stated generality, Theorem VII.4 does not seem to have ever been written down before and its probabilistic implications—
including, in particular, the Main Theorem— are developed for the first time in this work.

Section VIII describes how various classical inequalities can be seen as special cases of the Main Theorem, while Section IX
discusses an application of the Main Theorem to bounding the entropy of the sum of two independent uniform random vectors.

One application that rearrangement inequalities have found in probability is in the area of isoperimetric inequalities for
stochastic processes. Representative works in this area include Watanabe [51] on capacities associated to Lévy processes,
Burchard and Schmuckenschläger [18] on exit times of Brownian motions on the sphere or hyperbolic space, Bañuelos and
Méndez-Hernández on exit times and more for general Lévy processes [2], and Drewitz, Sousi and Sun [26] on survival
probabilities in a field of Lévy-moving traps. Our results also have implications for stochastic processes and these are developed
in Section X.

Finally, in Section XI, we give a new proof of the classical entropy power inequality using the Main Theorem. This shares
some features in common with the proof of Szarek and Voiculescu [47], but completely bypasses the use of Fisher information,
MMSE or any differentiation of entropy functionals.

We have also obtained discrete analogues of several of the results of this paper; these analogues and their applications to
combinatorics will be presented elsewhere (see, e.g., [50]).

For convenience of the reader, we collect here some (mostly standard) notations that will be used in the rest of the paper:

1) \( \|x\| \): Euclidean norm of \( x \in \mathbb{R}^n \).
2) \( \{ \text{Condition} \} \): equals 1 if Condition is true and 0 otherwise. For example, \( \{ f(x) > 1 \} = 1 \) if \( f(x) > 1 \) and 0 otherwise.
3) \( \{ x : \text{Condition} \} \): the set of \( x \) such that Condition is true. For example, \( \{ x : f(x) > 1 \} \) is the set of all \( x \) such that \( f(x) > 1 \).
4) \( \mathbb{I}_A(x) \): indicator function of the set \( A \).
5) \( t_+ := \max(t, 0) \).
6) \( f^+(x) := \max(f(x), 0) \).
7) \( f^-(x) := (-f)^+(x) \).
8) \( f * g \): the convolution of \( f(x) \) and \( g(x) \).
9) \( \ast_{1 \leq i \leq n} f_i \): the convolution of the functions \( f_i \), namely \( f_1 * f_2 * \cdots * f_n \).
10) \( \varphi'_+(t) \): right derivative of a function \( \varphi(t) \), defined on \( \mathbb{R} \).

II. BASIC FACTS ABOUT REARRANGEMENTS

We will try to make this section as self-contained as possible. For a Borel set \( A \) with volume \( |A| \), one can define its
spherically decreasing symmetric rearrangement \( A^* \) by

\[
A^* = B(0, r),
\]

where \( B(0, r) \) stands for the open ball with radius \( r \) centered at the origin and \( r \) is determined by the condition that \( B(0, r) \)
has volume \( |A| \). Here we use the convention that if \( |A| = 0 \), then \( A^* = \emptyset \) and that if \( |A| = +\infty \), then \( A^* = \mathbb{R}^n \).

Now for a measurable non-negative function \( f \), we define its spherically decreasing symmetric rearrangement \( f^* \) by:

\[
f^*(y) = \int_0^{+\infty} \{ y \in B_r^* \} dt
\]
where \( B_t = \{ x : f(x) > t \} \).

From the definition, it is intuitively clear that \( \{ x : f(x) > t \}^* = \{ x : f^*(x) > t \} \) for all \( t \geq 0 \). The proof of this is given in the following lemma, which is listed as an exercise in [34].

**Lemma II.1.** \( \{ x : f(x) > t \}^* = \{ x : f^*(x) > t \} \) for all \( t \geq 0 \).

**Proof:** Consider the function \( g(x, t) = \mathbb{1}_{\{ y : f(y) > t \}}(x). \)

Observe that for fixed \( x \), if for some \( t_1 \) we have

\[
g(x, t_1) = 1,
\]

then for all \( t \leq t_1 \), we would have

\[
g(x, t) = 1.
\]

Because of this,

\[
\{ x : f^*(x) > t_1 \} = \bigcup_{s > t_1} \{ x : g(x, s) = 1 \} = \bigcup_{s > t_1} \{ x : f(x) > s \}^*.
\]

Notice that for \( s_1 < s_2 \),

\[
\{ x : f(x) > s_2 \}^* \subseteq \{ x : f(x) > s_1 \}^*.
\]

Hence

\[
\{ x : f^*(x) > t_1 \} = \bigcup_n \left\{ x : f(x) > t_1 + \frac{1}{n} \right\}^*.
\]

Now observe that \( \bigcup_n \{ x : f(x) > t_1 + \frac{1}{n} \}^* \) is an open ball with the same Lebesgue measure as \( \{ x : f(x) > t_1 \}^* \), which is also an open ball.

**Remark 1.** For any measurable subset \( A \) of \([0, \infty)\), a generating class argument shows easily that \( |\{ x : f(x) \in A \}| = |\{ x : f^*(x) \in A \}|. \)

**Remark 2.** A natural consequence of this is that \( f^* \) is lower semicontinuous. By the layer cake representation, another consequence is that if \( f \) is integrable, so is \( f^* \) and \( \| f \|_1 = \| f^* \|_1 \). In particular, \( f^* \) is a probability density if \( f \) is.

The second simple observation is that all Rényi entropies are preserved by rearrangements.

**Lemma II.2.** For any \( p \in [0, 1) \cup (1, \infty) \),

\[
h_p(f) = h_p(f^*).
\]

For \( p = 1 \), if one of \( h(f) \) and \( h(f^*) \) is well defined, then so is the other one and we have:

\[
h(f) = h(f^*).
\]

The preservation of \( L^p \)-norms (for \( p > 1 \)) by rearrangements is a very classical fact (see, e.g., [17, Lemma 1.4]), although somewhat surprisingly the preservation of Shannon entropy does not seem to have explicitly noted anywhere in the literature. We give a complete proof of a more general result, namely Lemma VII.2, later.

Another useful fact is that composition with non-negative increasing functions on the left commutes with taking rearrangements.

**Lemma II.3.** [34] If \( \Psi(x) \) is a non-negative real valued strictly increasing function defined on the non-negative real line, then

\[
(\Psi(f))^* = \Psi(f^*)
\]

for any non-negative measurable function \( f \).

**Proof:** It suffices to show that the following two sets are equal for every \( t \):

\[
\{ x : (\Psi(f))^* > t \} = \{ x : \Psi(f^*) > t \}.
\]

By Lemma II.1,

\[
\{ x : \Psi(f^*) > t \} = \{ x : f^* > \Psi^{-1}(t) \} = \{ x : f > \Psi^{-1}(t) \}^*.
\]

Again by Lemma II.1,

\[
\{ x : f > \Psi^{-1}(t) \}^* = \{ x : \Psi(f) > (t) \}^* = \{ x : (\Psi(f))^* > t \}.
\]
The final fact that will be useful later is that rearrangement decreases the $L_1$ distance between two functions. We refer to [17] for a proof.

Lemma II.4. [17] Let $f$ and $g$ be two integrable non-negative functions. Then

$$\|f^* - g^*\|_1 \leq \|f - g\|_1.$$ 

By construction, spherically symmetric decreasing rearrangements move the mass of functions towards the origin. A fundamental rearrangement inequality expressing this concentration is an inequality popularly known as the Brascamp-Lieb-Luttinger inequality, which we state now.

Theorem II.5. [40], [15] For any measurable functions $f_i : \mathbb{R}^n \to [0, \infty)$, with $1 \leq i \leq M$, and real numbers $a_{ij}, 1 \leq i \leq M, 1 \leq j \leq N$,

$$\int_{\mathbb{R}^n} \prod_{j=1}^{N} dx_j \prod_{i=1}^{M} f_i \left( \sum_{j=1}^{N} a_{ij} x_j \right) \leq \int_{\mathbb{R}^n} \prod_{j=1}^{N} dx_j \prod_{i=1}^{M} f_i^* \left( \sum_{j=1}^{N} a_{ij} x_j \right).$$

Remark 3. To our considerable surprise, we found while preparing this paper that Theorem II.5 was in fact discovered by C. A. Rogers [40] as far back as 1957, and using a similar proof idea as Brascamp, Lieb and Luttinger [15] rediscovered in 1974. This historical fact does not seem to be widely known, but it is the reason we call Theorem II.5 the Rogers-Brascamp-Lieb-Luttinger inequality elsewhere in this paper.

Remark 4. As noted in [15], Theorem II.5 is only nontrivial when $M > N$.

In fact, we only need the following special but important case of the Rogers-Brascamp-Lieb-Luttinger inequality in this paper.

Theorem II.6. For any non-negative measurable functions $f_i, 1 \leq i \leq k$, on $\mathbb{R}^n$, we have

$$\int f_1(y) \left[ \star_{2 \leq i \leq k} f_i(y) \right] dy \leq \int f_1^*(y) \left[ \star_{2 \leq i \leq k} f_i^*(y) \right] dy.$$

Proof: By definition, we have

$$\int f_1(y) \left[ \star_{2 \leq i \leq k} f_i(y) \right] dy = \int_{\mathbb{R}^{n(k-1)}} \prod_{j=1}^{k-1} dx_j \prod_{i=1}^{k} f_i \left( \sum_{j=1}^{k-1} a_{ij} x_j \right),$$

where $a_{1j} = \delta_{j,1}, a_{2j} = -1 + 2\delta_{j,1}$ and $a_{ij} = \delta_{i-1,j}, i > 2$. Hence we can apply Theorem II.5 with $N = k - 1, M = k$ to conclude.

Remark 5. For $k = 1$, Theorem II.6 is called the Hardy-Littlewood inequality. For $k = 2$, it is called the Riesz or Riesz-Sobolev inequality. (Riesz [39] actually proved only the one dimensional case, but it was generalized by Sobolev [45] to higher dimensions. See, for example, [17] for historical perspective on all these inequalities.) For $k > 2$, as demonstrated during the proof, Theorem II.6 is a special case of Theorem II.5.

Remark 6. Observe that when $f_2, \ldots, f_k$ are densities, we may interpret Theorem II.6 probabilistically as follows. Let $X_1, \ldots, X_M$ be random vectors with densities on $\mathbb{R}^n$. Then for any non-negative measurable function $u$ on $\mathbb{R}^n$,

$$\mathbb{E} u \left( \sum_{i=1}^{M} X_i \right) \leq \mathbb{E} u^* \left( \sum_{i=1}^{M} X_i^* \right).$$

In the following, we will see that our Main Theorem is a consequence of Theorem II.6 (and in fact, they are mathematically equivalent).
III. MOMENT AND RELATIVE ENTROPY INEQUALITIES FOR ARRANGEMENTS

In this section, we will show some moment and relative entropy inequalities, which are useful later.

Lemma III.1. Let \( g(x) \) be a non-negative increasing function defined on the non-negative real line and \( f_i, 1 \leq i \leq k \) be probability densities. Then

\[
\begin{align*}
\mathbb{E}g(\|X_1 + X_2 + \cdots + X_k\|) \\
\geq \mathbb{E}g(\|X^*_1 + X^*_2 + \cdots + X^*_k\|),
\end{align*}
\]

where all random vectors are independent, \( X_i \) is distributed according to \( f_i \), and \( X^*_i \) is distributed according to \( f^*_i \).

Proof: For any \( t > 0 \), we can apply Theorem II.6 (as interpreted in Remark 6) to obtain

\[
\mathbb{E}\left[e^{-tg(\|X_1 + X_2 + \cdots + X_k\|)}\right] \leq \mathbb{E}\left[e^{-tg(\|X^*_1 + X^*_2 + \cdots + X^*_k\|)}\right],
\]

since \( (e^{-tg(\|x\|)})^* = e^{-tg(\|x\|)} \) almost everywhere. Hence we get:

\[
\mathbb{E}\left[1 - e^{-tg(\|X_1 + X_2 + \cdots + X_k\|)}\right] \geq \mathbb{E}\left[1 - e^{-tg(\|X^*_1 + X^*_2 + \cdots + X^*_k\|)}\right].
\]

Now note that \( \frac{1 - e^{-tg(\|x\|)}}{t} \) is monotonically increasing to \( g(\|x\|) \) as \( t \) goes to zero, and that

\[
\frac{1 - e^{-tg(\|x\|)}}{t} \leq g(\|x\|)
\]

for any \( t > 0 \). Hence, as \( t \) goes to zero, we can apply monotone convergence on the right side of (5), and use the inequality (6) on the left side of (5) to obtain the claimed result.

It is easy to see that the density of \((X - c)^*\) is the same as that of \( X^* \). Hence a simple consequence of Lemma III.1 is:

Corollary III.2. For any random vector \( X \) with finite covariance matrix, \( \mathbb{E}\|X^*\|^2 \leq \mathbb{E}\|X - \mathbb{E}[X]\|^2 \). Equivalently, \( \text{tr} \text{ Cov}(X^*) \leq \text{tr} \text{ Cov}(X) \).

Lemma III.1, together with Lemma II.2, immediately imply that \( D(f^*\|g) \leq D(f\|g) \) where \( g \) is a non-degenerate isotropic Gaussian. In fact, we have the following more general fact, easiest to state in terms of the Rényi divergence (see [49] for a recent survey on this). Recall that the Rényi divergence of order \( \alpha \) between any two densities \( f \) and \( g \) is defined as

\[
D_\alpha(f\|g) = \frac{1}{\alpha - 1} \int f^\alpha g^{1-\alpha} dx
\]

for \( \alpha \in (0, 1) \cup (1, \infty) \), and that \( D_1(f\|g) = D(f\|g) \) is simply the usual relative entropy between \( f \) and \( g \).

Proposition III.3. Let \( f \) and \( g \) be two probability densities. Then

\[
D_\alpha(f^*\|g^*) \leq D_\alpha(f\|g),
\]

where \( 0 < \alpha \leq 1 \).

Proof: By Lemma II.3, we have:

\[
\int (f^*)^\alpha (g^*)^{1-\alpha} dx = \int (f^\alpha)^* (g^{1-\alpha})^* dx,
\]

where \( 0 < \alpha < 1 \). Now an easy application of Theorem II.6 will lead to:

\[
\int (f^*)^\alpha (g^*)^{1-\alpha} dx \geq \int f^\alpha g^{1-\alpha} dx,
\]

which is equivalent to \( D_\alpha(f^*\|g^*) \leq D_\alpha(f\|g) \).

It remains to prove the statement for relative entropy. We first do this under the assumptions that \( f \ll g, f^* \ll g^* \), and that the respective likelihood ratios \( r(x) \) and \( r^*_x(x) \) are uniformly bounded. By these assumptions, we can rewrite (7) in terms of likelihood ratios:

\[
\int (r^*_x)^\alpha g^* dx \geq \int r^\alpha g dx.
\]

Noting \( \int r^*_x g^*(x) dx = \int r(x) g(x) dx = 1 \), we get:

\[
\int \frac{(r^*_x)^\alpha - r^*_x}{\alpha - 1} g^* dx \leq \int \frac{r^\alpha - r}{\alpha - 1} g dx.
\]

(8)
Since we have assumed the uniform boundedness of both \( r(x) \) and \( r_\ast(x) \), as \( \alpha \) goes to 1, we can apply bounded convergence to both sides of (8) (we omit the details) to obtain:

\[
\int r_\ast \log(r_\ast)g^* \, dx \leq \int r \log(r)g \, dx,
\]

which is equivalent to

\[
D(f^* \| g^*) \leq D(f \| g).
\]

In the general case, by what has already been proved, we have:

\[
D(f^* \| (\lambda f + (1 - \lambda)g^*)) \leq D(f \| \lambda f + (1 - \lambda)g),
\]

where \( 0 < \lambda < 1 \). Here we have used the fact \((\lambda f + (1 - \lambda)g)^* \geq \lambda f^* \) since the \( * \) operation is order preserving [17]. Note that by Lemma II.4,

\[
\|g^* - [\lambda f + (1 - \lambda)g]^*\|_1 \leq \|g - [\lambda f + (1 - \lambda)g]\|_1 \leq 2\lambda.
\]

Hence as \( \lambda \) goes to 0, \((\lambda f + (1 - \lambda)g)^* \) converges to \( g^* \) in total variation distance. But it is well known that the relative entropy functional is jointly lower-semicontinuous with respect to the topology of weak convergence (see, e.g., [25]). Hence:

\[
D(f^* \| g^*) \leq \lim \inf_{\lambda \downarrow 0} D(f^* \| (\lambda f + (1 - \lambda)g)^*). \tag{9}
\]

On the other hand, the relative entropy functional is jointly convex [25]. We get:

\[
D(f \| \lambda f + (1 - \lambda)g) \leq (1 - \lambda)D(f \| g). \tag{10}
\]

Combining (9) and (10), we can conclude.

\[ \blacksquare \]

**IV. TOWARDS THE OPTIMAL RÉNYI EPI**

Very recently, Bobkov and Chistyakov [9] obtained a generalization of the entropy power inequality (henceforth EPI), due to Shannon [44] and Stam [46], to Rényi entropy of any order \( p > 1 \). First we recall that the Rényi entropy power of order \( p \) is defined for \( \mathbb{R}^n \)-valued random vectors by

\[
N_p(X) = \exp \left\{ \frac{2h_p(X)}{n} \right\}.
\]

**Theorem IV.1.** [9] If \( X_1, \ldots, X_k \) are independent random vectors taking values in \( \mathbb{R}^n \), then for any \( p \geq 1 \),

\[
N_p(X_1 + \ldots + X_k) \geq c_p \sum_{i=1}^{k} N_p(X_i),
\]

where \( c_p \) is a constant depending only on \( p \). Moreover, one may take \( c_1 = 1 \),

\[
c_p = \frac{1}{e^{p^{\frac{1}{p-1}}}} \quad \text{for} \quad p \in (1, \infty),
\]

and \( c_\infty = \frac{1}{e} \). If \( n = 1 \), we may take \( c_\infty = \frac{1}{2} \).

Note that the \( p = 1 \) case is simply the Shannon-Stam EPI, which is sharp in any dimension with equality only for Gaussians with proportional covariance matrices. In general, however, Theorem IV.1 is not sharp, and [9] does not suggest what the optimal constants or the extremizing distributions might be.

**Remark 7.** For any dimension \( n \), restricting to \( k = 2 \) and \( p = \infty \), it turns out that the sharp constant is \( \frac{1}{2} \), which is achieved for two identical uniform distributions on the unit cube; this is observed in another paper of Bobkov and Chistyakov [8]. However, note that the maximum of the density of convolution of two uniforms on the ball with unit volume (not unit ball) is 1, while the maximum of the density of each of them is also 1. Hence, some non-uniqueness of the extremal distribution arises (at least for \( p = \infty \)). Indeed, for \( k = 2 \) and \( p = \infty \), uniform distributions on any symmetric convex set \( K \) (i.e., \( K \) is convex, and \( x \in K \) if and only if \(-x \in K\)) of volume 1 will be extremal: if \( X \) and \( X' \) are independently distributed according to \( f = I_K \), then denoting the density of \( X - X' \) by \( u \), we have

\[
\|u\|_\infty = u(0) = \int f^2(x) \, dx = 1 = \|f\|_\infty,
\]

so that \( N_\infty(X + X') = N_\infty(X - X') = N_\infty(X) = \frac{1}{2}[N_\infty(X) + N_\infty(X')] \).
Our Main Theorem may be seen as refining Theorem IV.1 (in a similar way to how the $p=1$ case of it refined the classical EPI). In order to do this, however, we need to recast Theorem IV.1 in a different, more precise, conjectural form, which suggests the extremizing distributions of Theorem IV.1.

When dealing with optimization problems involving Rényi entropies, it is quite common for a certain class of generalized Gaussians to arise. A rich collection of such generalized Gaussians has been studied in the literature. The ones that are of interest to us are a one-parameter family of distributions, indexed by a parameter $-\infty < \beta \leq \frac{2}{n+2}$, of the following form: $g_0$ is the standard Gaussian density in $\mathbb{R}^n$, and for $\beta \neq 0$,

$$g_\beta(x) = A_\beta \left(1 - \frac{\beta}{2} \|x\|^2 \right)^{\frac{1}{2} - \frac{n+2}{2}};$$

where $A_\beta$ is a normalizing constant (which can be written explicitly in terms of gamma functions if needed). We call $g_\beta$ the standard generalized Gaussian of order $\beta$; any affine function of a standard generalized Gaussian yields a “generalized Gaussian”. Observe that the densities $g_\beta$ (apart from the obviously special value $\beta = 0$) are easily classified into two distinct ranges where they behave differently. First, for $\beta < 0$, the density is proportional to a negative power of $(1 + b \|x\|^2)$ for a positive constant $b$, and therefore correspond to measures with full support on $\mathbb{R}^n$ that are heavy-tailed. For $\beta > 0$, note that $(1 - b \|x\|^2)_+$ with positive $b$ is non-zero only for $\|x\| < b^{-\frac{1}{2}}$, and is concave in this region. Thus any density in the second class, corresponding to $0 < \beta \leq \frac{2}{n+2}$, is a positive power of $(1 - b \|x\|^2)_+$, and is thus a concave function supported on a centered Euclidean ball of finite radius. In particular, note that $\beta = \frac{2}{n+2}$ is the uniform distribution on the Euclidean ball of radius $\sqrt{n+2}$. It is pertinent to note that although the first class includes many distributions from what one might call the “Cauchy family”, it excludes the standard Cauchy distribution; indeed, not only do all the generalized Gaussians defined above have finite variance, but in fact the form has been chosen so that, for $Z \sim g_\beta$,

$$E[\|Z\|^2] = n$$

for any $\beta$. Incidentally, the generalized Gaussians are called by many different names in the literature; the one other nomenclature that is perhaps worth noting for its relevance in statistics is that the $\beta < 0$ class is also called the Student-$t$ class, while the $0 < \beta \leq \frac{2}{n+2}$ class is also called the Student-$r$ class.

For $p > \frac{n}{n+2}$, define $\beta_p$ by

$$1 \beta_p = \frac{1}{p-1} + \frac{n+2}{2};$$

note that $\beta_p$ ranges from $-\infty$ to $\frac{2}{n+2}$ as $p$ ranges from $\frac{n}{n+2}$ to $\infty$. Henceforth we will write $Z^{(p)}$ for a random vector drawn from $g_\beta$.

Costa, Hero and Vignat [21] showed that the maximizers of Rényi entropy under covariance constraints are the generalized Gaussians; this fact was later obtained by Lutwak, Yang and Zhang [36] in a more general setting as what they called “moment-entropy inequalities”. We find the following formulation convenient.

**Theorem IV.2.** [21], [36] If $X$ is a random vector taking values in $\mathbb{R}^n$, then for any $p > \frac{n}{n+2}$,

$$\frac{E[\|X\|^2]}{N_p(X)} \geq \frac{E[\|Z^{(p)}\|^2]}{N_p(Z^{(p)})}.$$

Clearly Theorem IV.2 implies that under a variance constraint, the Rényi entropy power of order $p$ is maximized by the generalized Gaussian $Z^{(p)}$.

This leads us to the following conjecture, refining Theorem IV.1.

**Conjecture IV.3.** Let $X_1, \ldots, X_k$ be independent random vectors taking values in $\mathbb{R}^n$, and $p > \frac{n}{n+2}$. Suppose $Z_i$ are independent random vectors, each a scaled version of $Z^{(p)}$, such that $h_p(X_i) = h_p(Z_i)$. Then

$$N_p(X_1 + \ldots + X_k) \geq N_p(Z_1 + \ldots + Z_k).$$

This conjecture is true for at least three important special cases:

1) $p = 1, \forall n, \forall k$: This is the classical EPI;
2) $p = \infty, k = 2, \forall n$: This follows from a recent result of Bobkov and Chistyakov [8] as explained in Remark 7;
3) $p = \infty, n = 1, \forall k$: This is a relatively old but nontrivial result of Rogozin [41].

In principle, Conjecture IV.3 suggests optimal constants for Theorem IV.1; they should simply be those that arise in comparing $N_p(Z_1 + \ldots + Z_k)$ with $\sum_{i=1}^k N_p(Z_i)$ (which, by construction is simply $\sum_{i=1}^k N_p(X_i)$). The optimal constants would depend on $k$, $p$ and $n$. This explains the precise reason why the optimal constant is 1 in the classical case $p = 1$; it is because sums
of independent Gaussians are independent Gaussians. For the case \( k = 2 \), we make a more aggressive conjecture. We would like to know

\[
\inf_{a_1,a_2} N_p(a_1 Z_1^{(p)} + a_2 Z_2^{(p)}) = N_p(a_1 Z_1^{(p)}) + N_p(a_2 Z_2^{(p)}),
\]

where \( Z_1^{(p)} \) and \( Z_2^{(p)} \) are independently drawn from \( g_{\beta_p} \). Although we do not have a rigorous argument, symmetry seems to suggest that \( a_1 = a_2 \) would be optimal here, and this is borne out by some numerical tests in one dimension. One can compute (see, e.g., [35]) that, for \( p \neq 1 \),

\[
N_p(Z^{(p)}) = A_p^{-\frac{2}{p}} \left( 1 - \frac{1}{n} \beta_p \right)^{\frac{2}{p}-1}.
\]

(For \( p = 1 \), \( N_p(Z^{(p)}) \) is just the Shannon entropy power of the standard Gaussian, which is well known to be \( 2\pi e \).) It appears to be much harder to compute \( N_p \) for the self-convolution of \( Z^{(p)} \).

**Conjecture IV.4.** Let \( X_1 \) and \( X_2 \) be independent random vectors taking values in \( \mathbb{R}^n \), and \( p > \frac{n}{n+2} \). Then

\[
N_p(X_1 + X_2) \geq C_{p,n}[N_p(X_1) + N_p(X_2)],
\]

where \( C_{p,n} = \frac{1}{2} N_p(Z_1^{(p)} + Z_2^{(p)}) \) and \( Z_1^{(p)}, Z_2^{(p)} \) are independently drawn from \( g_{\beta_p} \).

This conjecture is true in two known cases: when \( p = 1 \), \( C_{1,n} = 1 \), and it is the classical EPI; when \( p = \infty \), \( C_{\infty,n} = \frac{1}{2} \), and how this follows from the work of Bobkov and Chistyakov [8] is explained in Remark 7.

We remark that in terms of importance, Conjecture IV.3 is far more important. If we can prove Conjecture IV.3 for \( k = 2 \), then proving or disproving Conjecture IV.4 reduces in principle to a calculus problem. Also, we mention that maybe the right formulation to start with is Conjecture IV.3, since \( C_{p,n} \) might not be analytically computable or may have a complicated expression. For example, when \( n = 1, p = 2 \), we computed \( C_{2,1} \) using Mathematica:

\[
C_{2,1} = \frac{166753125}{16 \left( \frac{573635\sqrt{2}}{2} - 142365\sqrt{10} \right)^2} \approx 0.956668.
\]

We mention for completeness that Johnson and Vignat [32] also demonstrated what they call an “entropy power inequality for Rényi entropy”, for any order \( p \geq 1 \). However, their inequality does not pertain to the usual convolution, but a new and somewhat complicated convolution operation (depending on \( p \)). This new operation reduces to the usual convolution for \( p = 1 \), and has the nice property that the convolution of affine transforms of independent copies of \( Z^{(p)} \) is an affine transform of \( Z^{(p)} \) (which, as observed above, fails for the usual convolution).

A variant of the classical entropy power inequality is the concavity of entropy power of a sum when one of the summands is Gaussian, due to Costa [22], Savaré and Toscani [43] recently proposed a generalization of Costa’s result to Rényi entropy power, but the notion of concavity they use based on solutions of a nonlinear heat equation does not have obvious probabilistic meaning, and their work also does not seem directly connected to the approach discussed in this section. The definition of Rényi entropy power used in [43] has a different constant in the exponent \( \frac{p}{p+1} - 1 \) as opposed to \( \frac{2}{n} \), and it is conceivable, as a reviewer suggested, that Conjecture IV.4 is true for \( p \in (1, \infty) \) only with this modified definition of \( N_p \).

**V. Preliminaries for the First Proof**

We first state a lemma, which seems to be folklore in the continuous case [24] and allows us to to obtain the cases when \( p = 0, 1, \infty \) as limiting cases.

**Lemma V.1.** (i) The following limit is well defined:

\[
h_0(f) = \lim_{p \to 0^+} h_p(f) = \log |\text{supp}(f)|,
\]

where \( \text{supp}(f) \) is the support of \( f \), defined as the set \( \{ x : f(x) \neq 0 \} \).

(ii) The following two limits are well defined:

\[
h_1^+(f) = \lim_{p \to 1^+} h_p(f), \quad h_1^-(f) = \lim_{p \to 1^-} h_p(f).
\]

If \( h_1^+(f) > -\infty \), then \( h(f) \) is well defined (possibly \(+\infty\)) and we have:

\[
h_1^+(f) = h(f).
\]
If \( h^{-1}(f) < +\infty \), then \( h(f) \) is well defined (possibly \(-\infty\)) and we have:

\[ h^{-1}(f) = h(f). \]

(iii) The following limit is well defined:

\[ h_{\infty}(f) = \lim_{p \to +\infty} h_{p}(f), \]

and we have:

\[ h_{\infty}(f) = -\log \|f\|_{\infty}, \]

where \( \|f\|_{\infty} \) is the essential supremum of \( f \).

If we suppose that \( f \) is defined everywhere and lower semicontinuous, then

\[ h_{\infty}(f) = -\log \sup_{x} f(x). \]

(iv) Suppose \( h(f) \) is well defined (possibly \( \infty \)) and let \( q \in (0, 1), p \in (1, +\infty) \). Then

\[ h_{q}(f) \geq h(f) \geq h_{p}(f). \]

Proof: We first show that all four limits are well defined. By Lyapunov’s inequality, the function \( g \), defined by

\[ g(p) = \log \int f^{p}(x)dx \]

is a convex function from \((0, +\infty)\) to \((-\infty, +\infty]\). Since \( f \) is assumed to be probability density, we have \( g(1) = 0 \).

Consider \( p, q \in (0, 1) \) with \( p < q \). Convexity will give us:

\[ \frac{g(1) - g(p)}{1 - p} \leq \frac{g(1) - g(q)}{1 - q}, \]

which is equivalent to \( h_{p}(f) \geq h_{q}(f) \). This monotonicity guarantees that the \( h_{0}(f) \) and \( h_{-1}(f) \) exist, possibly as extended real numbers.

Similarly, if we consider \( p, q \in (1, \infty) \) with \( p < q \), convexity will give us:

\[ \frac{g(p) - g(1)}{p - 1} \leq \frac{g(q) - g(1)}{q - 1}, \]

which is equivalent to \( h_{p}(f) \geq h_{q}(f) \). This will guarantee the existence of the other two limits, possibly as extended real numbers.

Now we prove the lemma.

(i) We first assume \(|\text{supp}(f)| < \infty \). It’s clear that we only need to show:

\[ \lim_{p \to 0^{+}} \int f^{p}(x)dx = |\text{supp}(f)|. \]

Note that

\[ \int f^{p}(x)dx = \int \{f(x) \neq 0\} f^{p}(x)dx \]

and

\[ \{f(x) \neq 0\} f^{p}(x) \leq \max(f(x), 1) \{f(x) \neq 0\}. \]

But \( \max(f(x), 1) \{f(x) \neq 0\} \) is integrable by our assumption that \(|\text{supp}(f)| < \infty \). Hence by dominated convergence, we obtain:

\[ \lim_{p \to 0^{+}} \int f^{p}(x)dx = \lim_{p \to 0^{+}} \int \{f(x) \neq 0\} f^{p}(x)dx = \int \{f(x) \neq 0\} dx = |\text{supp}(f)|. \]

On the other hand, if \(|\text{supp}(f)| = +\infty \), then

\[ \lim_{n} \left| \left\{ x : f(x) \geq \frac{1}{n} \right\} \right| = +\infty. \]

Note that

\[ \lim_{p \to 0^{+}} \int f^{p}(x)dx \geq \lim_{p \to 0^{+}} \frac{1}{n^{p}} \left| \left\{ x : f(x) \geq \frac{1}{n} \right\} \right| = \left| \left\{ x : f(x) \geq \frac{1}{n} \right\} \right|. \]
Hence we have
\[ \lim_{p \to 0^+} \int f^p(x)dx = +\infty = |\text{supp}(f)|. \]

(ii) Clearly, assuming \(h_1^+(f) > -\infty\) implies the existence of \(\epsilon > 0\), such that:
\[ \int f^{1+\epsilon}(x)dx < +\infty. \]

We first show that the function \(g(p)\) is continuous at \(p = 1\). Take \(1 < p < 1 + \epsilon\). Then
\[ f^p(x) \leq \{ f(x) \leq 1 \} f(x) + \{ f(x) > 1 \} f^{1+\epsilon}(x). \]

But \(\{ f(x) \leq 1 \} f(x) + \{ f(x) > 1 \} f^{1+\epsilon}(x)\) is integrable, hence we can apply dominated convergence to conclude that \(g(p)\) is continuous at \(p = 1\).

Next we show that for \(1 < p < 1 + \epsilon\), \(g(p)\) is differentiable, with a finite derivative given by:
\[ g'(p) = \int f^p \log(f)dx. \]

This basically follows from dominated convergence and the details of the argument can be found in Lemma 6.1 in [12]. We do not repeat it here.

Now, we show that the limit
\[ - \lim_{p \to 1^+} g'(p) \]
exists (possibly \(+\infty\)) and equals \(h(f)\). Note that
\[ \int f^p \log(f)dx = \int_{\{x:f(x)>1\}} f^p(x) \log^+(f(x))dx - \int_{\{x:0<f(x)<1\}} f^p(x) \log^-(f(x))dx. \]

For the first term, we can apply dominated convergence using \(\int f^{1+\epsilon}dx < +\infty\). While for the second term, we can apply monotone convergence.

Finally, we can apply a version of L’Hospital’s rule to conclude that
\[ \lim_{p \to 1^+} \frac{g(p)}{1-p} = - \lim_{p \to 1^+} g'(p) = h(f). \]

A similar analysis applies to the assumption \(h_1^-(f) < +\infty\).

(iii) The statement \(h_\infty(f) = -\log \|f\|_\infty\) follows from a classical analysis fact [42].

For the other statement, one direction is easy:
\[ h_p(f) = \frac{1}{1-p} \log \int f^p(x)dx \]
\[ \geq \frac{1}{1-p} \log \left( \sup_x f^{p-1}(x) \right) = -\log \sup_x f(x). \]

Note that the above is automatically true if \(\sup_x f(x) = +\infty\). Now fix \(y \in \text{supp}(f)\) and \(0 < \delta < f(y)\). By lower semicontinuity, we can find an open ball \(B(y, r)\), centered at \(y\) with radius \(r > 0\), such that:
\[ \inf_{x \in B(y, r)} f(x) \geq f(y) - \delta. \]

From this, we easily obtain:
\[ h_p(f) \leq \frac{\log \left( (f(y) - \delta)^p |B(y, r)| \right)}{1-p}. \]

By first sending \(p\) to infinity and then sending \(\delta\) to zero, we arrive at:
\[ h_\infty(f) \leq -\log f(y). \]

Taking infimum of the right hand side over \(y \in \text{supp}(f)\), we conclude:
\[ h_\infty(f) \leq -\log \sup_{y \in \text{supp}(f)} f(y) = -\log \sup_y f(y). \]
Corollary V.2. The fact that Lemma V.1(ii).

\[ \log \int f^p(x)dx = \log \int f^{p-1}(x)f(x)dx \geq (p-1) \int \log(f(x))f(x)dx. \]

This will give \( h(f) \leq h_p(f) \). Similarly, we can obtain the other inequality.

Remark 8. As was observed in [6], there exist densities such that \( h_p(X) = -\infty \) for every \( p > 1 \) but \( h(X) \) is well defined and finite. An example of such a density is

\[ f(x) = \frac{c}{x \log^*(1/x)}, \quad 0 < x < \frac{1}{2}, \]

where \( c \) is a normalizing constant. This shows that the continuity of Rényi entropy in the order at \( p = 1 \) is not automatic, and means that one has to be cautious about the generality in which proofs based on this continuity apply (e.g., the proof of the entropy power inequality from Young’s inequality with sharp constant in [24] needs an additional condition as in Lemma V.1(ii)).

For later reference, we also record the following corollary, which is clear from the proof of Lemma V.1(ii).

Corollary V.2. The fact that \( h_{1}^{+}(f) > -\infty \) is equivalent to the existence of \( \epsilon > 0 \), such that:

\[ \int f^{1+\epsilon}(x)dx < +\infty. \]

Similarly, the fact that \( h_{1}^{-}(f) < +\infty \) is equivalent to the existence of \( 1 > \epsilon > 0 \), such that:

\[ \int f^{1-\epsilon}(x)dx < +\infty. \]

VI. FIRST PROOF OF MAIN THEOREM

Theorem VI.1. If \( f_i \) are densities on \( \mathbb{R}^n \) and \( f_i^\star \) are their spherically symmetric rearrangements, then

\[ h_p(f_1 \ast f_2 \ast \cdots \ast f_k) \geq h_p(f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star), \]

for any \( p \in [0, 1) \cup (1, \infty] \). For \( p = 1 \), if \( h(f_1 \ast f_2 \ast \cdots \ast f_k) \) is well defined, then

\[ h(f_1 \ast f_2 \ast \cdots \ast f_k) \geq h_{1}^+(f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star); \]

if \( h(f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star) \) is well defined, then

\[ h_{1}^-(f_1 \ast f_2 \ast \cdots \ast f_k) \geq h(f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star). \]

In particular, if one of the densities, say \( f_1 \), satisfies

\[ \int f_1^{1+\epsilon}(x)dx < +\infty, \]

for some \( \epsilon > 0 \), then both \( h(f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star) \) and \( h(f_1 \ast f_2 \ast \cdots \ast f_k) \) are well defined and we have:

\[ h(f_1 \ast f_2 \ast \cdots \ast f_k) \geq h(f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star). \]

Proof: Case 1: \( p \in (1, +\infty) \).

By definition of Rényi entropy, it suffices to show

\[ \|f_1 \ast f_2 \ast \cdots \ast f_k\|_p \leq \|f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star\|_p. \]

(12)

Note that by duality,

\[ \|f_1 \ast f_2 \ast \cdots \ast f_k\|_p = \sup_{\|g\|_{\frac{1}{p}}} \int_{\mathbb{R}^n} g(x)f_1 \ast f_2 \ast \cdots \ast f_k(x)dx, \]

where \( \frac{1}{p} + \frac{1}{q} = 1 \). Hence we can apply Theorem II.6 to obtain:

\[ \int_{\mathbb{R}^n} g(x)f_1 \ast f_2 \ast \cdots \ast f_kdx \leq \int_{\mathbb{R}^n} |g|^{\ast}(x)f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star dx. \]

Due to Lemma II.2, \( \|g|^{\ast}\|_q = 1 \) and again by duality,

\[ \int_{\mathbb{R}^n} |g|^{\ast}(x)f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star dx \leq \|f_1^\star \ast f_2^\star \ast \cdots \ast f_k^\star\|_p. \]
Hence inequality (12) is shown.

**Case 2:** $p = \infty$.

This follows from Case 1 and Lemma V.1.

**Case 3:** $p \in (0, 1)$.

As mentioned before, the $k = 2$ case was proved by [14, Proposition 9]. It is straightforward to extend the argument there to general $k$. We give the proof here for completeness. First of all, by the reverse of Hölder’s inequality that applies for $p < 1$,

$$
\int f(x)h(x)dx \geq \|f\|_p\|h\|_{p'},
$$

where $\frac{1}{p} + \frac{1}{p'} = 1$ (here $p' < 0$) and $f, h$ are non-negative. When $h(x) = \alpha f^{p-1}(x)$, with $\alpha > 0$, there is equality. Hence we have:

$$
\|f\|_p = \inf_{\|h\|_{p'} = 1, h \geq 0} \int f(x)h(x)dx.
$$

Applying this to $f_1 \ast f_2 \ast \cdots f_k(x)$, we get:

$$
\|f_1 \ast f_2 \ast \cdots f_k(x)\|_p = \inf_{\|h\|_{p'} = 1, h \geq 0} \int f_1 \ast f_2 \ast \cdots f_k(x)h(x)dx.
$$

Now define the spherically increasing symmetric rearrangement $^*h$ of $h$ by:

$$
^*h = \frac{1}{(\pi)^{\frac{1}{2}}},
$$

Then $\|^*h\|_{p'} = \|h\|_{p'}$. For $A > 0$, define:

$$
h_A(x) = \min(A, h(x)); k^A(x) = A - h_A(x).
$$

Then, as $A \to +\infty$,

$$
h_A(x) \uparrow h(x), \quad A - (k_A(x))^* \uparrow ^*h.
$$

By monotone convergence, we obtain

$$
\int f_1 \ast f_2 \ast \cdots f_k(x)h(x)dx = \lim_{A \uparrow \infty} \int f_1 \ast f_2 \ast \cdots f_k(x)h_A(x)dx
$$

$$
= \lim_{A \uparrow \infty} A - \int f_1 \ast f_2 \ast \cdots f_k(x)k_A(x)dx.
$$

Now, similar to the proof of Case 1, we can apply Theorem II.6 to obtain

$$
\int f_1 \ast f_2 \ast \cdots f_k(x)h(x)dx \geq \lim_{A \uparrow \infty} A - \int f_1^* \ast f_2^* \ast \cdots f_k^*(x)(k_A)^*(x)dx
$$

$$
= \lim_{A \uparrow \infty} \int f_1^* \ast f_2^* \ast \cdots f_k^*(x)(A - (k_A)^*)(x)dx
$$

$$
= \int f_1^* \ast f_2^* \ast \cdots f_k^*(x)[^*h(x)]dx
$$

$$
\geq \|f_1^* \ast f_2^* \ast \cdots f_k^*(x)\|_p,
$$

where the last step follows from duality. Again by duality, taking an infimum over nonnegative functions $h$ with $\|h\|_{p'} = 1$ in inequality (14) gives us

$$
\|f_1 \ast f_2 \ast \cdots f_k(x)\|_p \geq \|f_1^* \ast f_2^* \ast \cdots f_k^*(x)\|_p.
$$

**Case 4:** $p = 0$.

This follows from Case 3 and Lemma V.1.

**Case 5:** $p = 1$.

If $h(f_1 \ast f_2 \ast \cdots f_k)$ is well defined, then by Lemma V.1(iv), we have:

$$
h(f_1 \ast f_2 \ast \cdots f_k) \geq h_p(f_1 \ast f_2 \ast \cdots f_k), \forall 1 < p < +\infty.
$$

By Case 1, we have

$$
h(f_1 \ast f_2 \ast \cdots f_k) \geq h_p(f_1^* \ast f_2^* \ast \cdots f_k^*), \forall 1 < p < +\infty.
$$
Taking the limit as $p \downarrow 1$, we obtain
\[
h(f_1 \ast f_2 \ast \cdots \ast f_k) \geq h^*_1(f_1^* \ast f_2^* \ast \cdots \ast f_k^*).
\]
The other inequality is proved similarly.

Finally, if $\|f_1\|_{1+\epsilon} < +\infty$, by Lemma II.2 and Young’s inequality, we have:
\[
\|f_1^* \ast f_2^* \ast \cdots \ast f_k^*\|_{1+\epsilon} < +\infty,
\]
\[
\|f_1 \ast f_2 \ast \cdots \ast f_k\|_{1+\epsilon} < +\infty.
\]

Now we can apply Case 5, Corollary V.2 and Lemma V.1(ii) to conclude.

\[\square\]

Remark 9. In [14], the inequality (13) was claimed without proof. We sketch a proof here. Let
\[
m(x) = \frac{1}{\liminf A \uparrow \infty (A - k_M^*(x))}.
\]
To show $m(x) = (\frac{1}{h})^*$, we show that $\{x : m(x) > t\} = \{x : (\frac{1}{h})^* > t\}$. But
\[
\{x : m(x) > t\} = \bigcup_{n=1}^{\infty} \bigcap_{M=1}^{\infty} \left\{x : (M - k_M(x)) < \frac{1}{t + \frac{1}{n}} \right\}.
\]
Using Lemma II.1, we obtain:
\[
\{x : m(x) > t\} = \bigcup_{n=1}^{\infty} \bigcap_{M=1}^{\infty} \left\{x : h_M(x) < \frac{1}{t + \frac{1}{n}} \right\}^*
\]
\[
= \bigcup_{n=1}^{\infty} \bigcap_{M=1}^{\infty} \left\{x : \left(\frac{1}{h_M} \right)^* > t + \frac{1}{n} \right\}
\]
\[
= \left\{x : \left(\frac{1}{h} \right)^* > t \right\},
\]
where the last step follows since:
\[
\left(\frac{1}{h_M} \right)^* = \frac{1}{M} + \int_{\frac{1}{M}}^{\infty} \left\{\frac{1}{h(x)} > t \right\}^* dt \downarrow \left(\frac{1}{h} \right)^*,
\]
as $M$ goes to $\infty$, where the equality follows directly from definition.

We now prove a variation of Theorem VI.1 when $p = 1$. Although we will prove the best possible version of the $p = 1$ case (as stated in the Main Theorem in Section I) later, the method of proof of the following seems interesting.

Theorem VI.2. Let $f_i, 1 \leq i \leq k$ be probability densities on $\mathbb{R}^n$ and $f_i^*, 1 \leq i \leq k$, be their respective spherically symmetric decreasing rearrangements. If
\[
\int f_1 \ast f_2 \ast \cdots \ast f_k(x)\|x\|^2 dx < +\infty,
\]
then
\[
h(f_1 \ast f_2 \ast \cdots \ast f_k) \geq h(f_1^* \ast f_2^* \ast \cdots \ast f_k^*).
\]

Proof: By Lemma III.1,
\[
\int f_1^* \ast f_2^* \ast \cdots \ast f_k^*(x)\|x\|^2 dx < +\infty.
\]
Hence both $h(f_1 \ast f_2 \ast \cdots \ast f_k)$ and $h(f_1^* \ast f_2^* \ast \cdots \ast f_k^*)$ are well defined. Clearly, the Main Theorem implies that
\[
h(\sqrt{t}Z + X_1 + \cdots + X_k) \geq h(\sqrt{t}Z + X_1^* + \cdots + X_k^*),
\]
(15)
where $t > 0$, $Z$ is a $n$ dimensional standard normal, $X_i$ is distributed according to $f_i$, $X_i^*$ is distributed according to $f_i^*$ and all random vectors are independent. The rest of the argument follows by taking the limit as $t$ goes to 0. To simplify the notation, let
\[
X = X_1 + \cdots + X_k, X_\ast = X_1^* + \cdots + X_k^*.
\]
The joint lower-semicontinuity of relative entropy functional [25] gives
\[
D(X\|G) \leq \lim_{t \downarrow 0} \inf_{t} D(\sqrt{t}Z + X\|\sqrt{t}Z + G),
\]
where \( G \) is a Gaussian random vector matching the mean and covariance of \( X \), independent of \( Z \). But by a standard equality in information theory [23], we get
\[
D(X\|G) = h(G) - h(X),
\]
\[
D(\sqrt{t}Z + X\|\sqrt{t}Z + G) = h(\sqrt{t}Z + G) - h(\sqrt{t}Z + X).
\]
It’s also easy to check directly that
\[
\lim_{t\downarrow 0} h(\sqrt{t}Z + G) = h(G).
\]
Hence we obtain
\[
\lim_{t\downarrow 0} \sup_{x} h(\sqrt{t}Z + X) \leq h(X). \tag{16}
\]
On the other hand, the density \( f_{\ast} \) of \( \sqrt{t}Z + X_{\ast} \) can be expressed as
\[
\mathbb{E} f_{\ast}(x - \sqrt{t}Z),
\]
with \( f_{\ast} \) being the density of \( X_{\ast} \). If we apply Jensen’s inequality to the concave function \(-u \log(u)\), we obtain
\[
-f_{\ast}(x) \log(f_{\ast}(x)) \geq -\mathbb{E} f_{\ast}(x - \sqrt{t}Z) \log(f_{\ast}(x - \sqrt{t}Z)). \tag{17}
\]
It is easy to check the right hand side of the above is well defined due to the boundedness of the normal density and the finiteness of the second moment of \( X_{\ast} \). If \( h(X_{\ast}) = -\infty \), we trivially have
\[
h(X_{\ast} + \sqrt{t}Z) \geq h(X_{\ast}).
\]
Hence we can assume \( h(X_{\ast}) \) is finite. In this case, we can integrate both sides of (17) with respect to \( x \) and use Fubini’s Theorem to conclude that
\[
h(X_{\ast} + \sqrt{t}Z) \geq h(X_{\ast}). \tag{18}
\]
Combining the inequalities (16), (18) and (15), we can conclude the proof.

**Remark 10.** We remark that the second part of the proof can be applied to \( X \), instead of to \( X_{\ast} \), as well. Hence, under the only assumption of the finiteness of the second moment, we obtain:
\[
\lim_{t\downarrow 0} h(X + \sqrt{t}Z) = h(X).
\]
This is known implicitly in [19], with a slightly different proof, but never seems to have been explicitly noted. Of course, the continuity of \( h(X + \sqrt{t}Z) \) in \( t \) is trivial when \( t > 0 \) because then one already has as much smoothness as desired to start with.

**VII. SECOND PROOF OF MAIN THEOREM VIA MAJORIZATION**

In this section, we give a new and unified proof of the Main Theorem for all values of \( p \), using ideas from majorization theory. In particular, we will show a best possible version of the \( p = 1 \) case of the Main Theorem. A generalization of the Main Theorem is also be obtained. We first define majorization as a partial order on the set of densities.

**Definition VII.1.** For probability densities \( f \) and \( g \) on \( \mathbb{R}^n \), we say that \( f \) is majorized by \( g \) if
\[
\int_{\{x: \|x\| < r\}} f^*(x)dx \leq \int_{\{x: \|x\| < r\}} g^*(x)dx
\]
for all \( r > 0 \). In this case, we write \( f \prec g \).

We also need the following lemma, which includes Lemma II.2 as a special case.

**Lemma VII.2.** Let \( f \) be a probability density and \( \varphi(x) \) be a convex function defined on the non-negative real line such that \( \varphi(0) = 0 \) and it is continuous at 0. Then
\[
\int \varphi(f(x))dx = \int \varphi(f^*(x))dx,
\]
provided that one of these integrals is well defined (which guarantees that the other is).

**Proof:** Note that a convex function satisfying the assumed conditions is always absolutely continuous. If \( \varphi(x) \geq 0, \forall x \geq 0 \), then \( \varphi \) must be increasing and we have \( \varphi' \geq 0 \); if \( \varphi(x) \leq 0, \forall x \geq 0 \), then \( \varphi \) must be decreasing and we have \( \varphi' \leq 0 \). In both cases, straightforward applications of Tonelli’s theorem and the layer cake representation will do.
For the remaining case, we can assume that there is a unique \( \alpha > 0 \), such that \( \varphi(\alpha) = 0 \), \( \varphi(x) > 0 \), \( x > \alpha \) and \( \varphi(x) \leq 0 \), \( x < \alpha \). Noting that \( \varphi^* \) is also a continuous convex function null at zero, by what has been proved, we obtain
\[
\int \varphi^+(f(x))\,dx = \int \varphi^+(f^*(x))\,dx.
\] (19)

On the other hand, we have
\[
\varphi^-(u) + \varphi'_+(\alpha)u\{u \leq \alpha\} = \{u \leq \alpha\}\int_0^u (\varphi'_+(\alpha) - \varphi'(t))\,dt.
\]

Tonelli’s theorem gives us
\[
\int \left[ \varphi^-(f(x)) + \varphi'_+(\alpha)f(x)\{f(x) \leq \alpha\} \right] \,dx = \int \left[ \varphi^-(f^*(x)) + \varphi'_+(\alpha)f^*(x)\{f^*(x) \leq \alpha\} \right] \,dx.
\]

But it is easy to see that
\[
-\infty < \int \varphi'_+(\alpha)f(x)\{f(x) \leq \alpha\} \,dx = \int \varphi'_+(\alpha)f^*(x)\{f^*(x) \leq \alpha\} \,dx < +\infty.
\]

Hence
\[
\int \varphi^-(f(x))\,dx = \int \varphi^-(f^*(x))\,dx,
\] (20)

and we can conclude, provided the quantities in (19) and (20) are not both \(+\infty\) or both \(-\infty\).

The following is well known in majorization theory. In the discrete case, it is first proved in Hardy, Littlewood and Polya [29]. Various extensions to the continuous setting are discussed by Chong [20]. However, we are not able to find a direct reference that covers all cases of our interest. So we give the proof here.

**Proposition VII.3.** Let \( \varphi(x) \) be a convex function defined on the non-negative real line such that \( \varphi(0) = 0 \) and it is continuous at 0. If \( f \) and \( g \) are probability densities, with \( f \prec g \), then
\[
\int \varphi(f(x))\,dx \leq \int \varphi(g(x))\,dx,
\]
provided that both sides are well defined.

**Proof:** We first show that for each \( t > 0 \),
\[
\int (f(x) - t)_+\,dx \leq \int (g(x) - t)_+\,dx.
\] (21)

By Lemma VII.2, we only need to show:
\[
\int (f^*(x) - t)_+\,dx \leq \int (g^*(x) - t)_+\,dx.
\]

By Markov’s inequality, we know that the set \( \{x : f^*(x) > t\} \) is an open ball with finite radius, say \( r \). Then we have
\[
\int (f^*(x) - t)_+\,dx = \int_{\{x : \|x\| < r\}} (f^*(x) - t)\,dx
\]
\[
\leq \int_{\{x : \|x\| < r\}} (g^*(x) - t)\,dx
\]
\[
\leq \int (g^*(x) - t)_+\,dx.
\]

Next, we assume that, additionally, \( \varphi'_+(0) \) is finite. Define the second derivative measure \( \mu \) of \( \varphi \) by setting
\[
\mu((a,b]) = \varphi'_+(b) - \varphi'_+(a),
\]
\[
\mu(\{0\}) = \varphi'_+(0).
\]

Restricted to \( (0, +\infty) \), \( \mu \) is a non-negative measure. Using Tonelli’s theorem, we see:
\[
\varphi(u) = \int_0^u \varphi'(t)\,dt
\]
\[
= \int_0^u \{0 < s \leq t\} \mu(ds)\,dt + \varphi'_+(0)u
\]
Theorem

Suppose Theorem VII.4.

Remark

11

µ

respect to

real line such that

if

ϕ

If we assume that both

| |

since

But

for any density

Now we recall the following representation [17]:

provided that both sides are well defined.

By what has been proved,

Finally, if \( \varphi'_+(0) \) is not finite, then it must be \(-\infty\) and we can find a \( \alpha > 0 \) such that \( \varphi(\alpha) < 0 \). Define \( \varphi_n(x) = \varphi \left( \frac{\alpha}{n} \right) x^2 \), if \( x \leq \frac{\alpha}{n} \) and \( \varphi_n(x) = \varphi(x) \) otherwise. Then

\[
\varphi^+_n(x) = \varphi^+(x),
\]

\[
\varphi^-_n(x) = \varphi^-(x).
\]

By what has been proved,

\[
\int \varphi_n(f(x)) dx \leq \int \varphi_n(g(x)) dx,
\]

which is equivalent to,

\[
\int \varphi^+(f(x)) dx - \int \varphi^-(f(x)) dx 
\leq \int \varphi^+(g(x)) dx - \int \varphi^-(g(x)) dx.
\]

If we assume that both \( \int \varphi(f(x)) dx \) and \( \int \varphi(g(x)) dx \) are well defined, then we can use monotone convergence to conclude.

We now apply this to obtain a proof of the Main Theorem (indeed, a generalization of it) under minimal assumptions.

**Theorem VII.4.** Suppose \( f_i, 1 \leq i \leq k \) are probability densities. Let \( \varphi(x) \) be a convex function defined on the non-negative real line such that \( \varphi(0) = 0 \) and \( \varphi \) is continuous at 0. Then

\[
\int \varphi(f_1 \ast f_2 \ast \cdots \ast f_k(x)) dx \leq \int \varphi(f_1^* \ast f_2^* \ast \cdots \ast f_k^*(x)) dx,
\]

provided that both sides are well defined.

**Proof:** We only need to show \( f = f_1 \ast f_2 \ast \cdots \ast f_k < g = f_1^* \ast f_2^* \ast \cdots \ast f_k^* \). Again by Theorem II.6, we have:

\[
\int r(x) f(x) dx \leq \int r^*(x) g(x) dx. \tag{22}
\]

Now we recall the following representation [17]:

\[
\int_{B(0,r)} f^*(x) dx = \sup_{|C| = |B(0,r)|} \int_C f(x) dx,
\]

for any density \( f \), where \( B(0,r) \) is the open ball with radius \( r \). By (22), we have that

\[
\int_{B(0,r)} f^*(x) dx \leq \sup_{|C| = |B(0,r)|} \int_{C^*} g(x) dx.
\]

But

\[
\sup_{|C| = |B(0,r)|} \int_{C^*} g(x) dx \leq \sup_{|A| = |B(0,r)|} \int_A g(x) dx,
\]

since \( |C^*| = |C| \). Using the representation again, we obtain:

\[
\int_{B(0,r)} f^*(x) dx \leq \int_{B(0,r)} g^*(x) dx.
\]

**Remark 11.** By taking \( \varphi(x) = x^p \) for \( p > 1 \), \( \varphi(x) = -x^p \) for \( 0 < p < 1 \) and \( \varphi(x) = x \log(x) \) for \( p = 1 \), we recover the Main Theorem.
VIII. IMPLICATIONS

In this section, we point out several implications of the Main Theorem, some of which have already been mentioned in the Introduction. The first implication is the Brunn-Minkowski inequality. One can recover it from the Main Theorem in full generality.

Corollary VIII.1. Let $A$ and $B$ be two nonempty Borel sets. Then:

$$|A + B|^\frac{1}{p} \geq |A^* + B^*|^\frac{1}{p} = |A|^\frac{1}{p} + |B|^\frac{1}{p}.$$  

Proof: We first assume both of them have non-zero and finite volume. We take $f_1 = \frac{I_A}{|A|}$ and $f_2 = \frac{I_B}{|B|}$. Then

$$f_1^* = \frac{\|A^*\|}{|A|}, f_2^* = \frac{\|B^*\|}{|B|}.$$

By the $p = 0$ case of the Main Theorem and Lemma V.1(i), we have:

$$|\text{supp}(f_1 \ast f_2)| \geq |\text{supp}(f_1^* \ast f_2^*)|.$$

As something that can be checked easily,

$$\text{supp}(f_1 \ast f_2) \subseteq A + B, \text{supp}(f_1^* \ast f_2^*) = A^* + B^*.$$

Hence we can conclude. If one of them has zero volume, say $B$, we take a point $x_0 \in B$. Then:

$$|A + B|^\frac{1}{p} = |A + B - x_0|^\frac{1}{p} \geq |A|^\frac{1}{p},$$

where the inequality follows since $A + B - x_0 \supseteq A$. If one of them has infinite volume, say $B$, we take a point $x_0 \in A$. Then:

$$|A + B|^\frac{1}{p} = |A + B - x_0|^\frac{1}{p} \geq |B|^\frac{1}{p}.$$

We next derive Theorem II.6 from the Main Theorem, thus showing that they are mathematically equivalent to each other. However, to recover Theorem II.6 for $k$ functions, we need the Main Theorem for $k$ densities, while, if we look back at the proof of the Main Theorem for $k$ densities, we need Theorem II.6 for $k + 1$ functions. To summarize the following, Theorem II.6 for $k$ functions can be seen as the $p = +\infty$ case of the Main Theorem for $k$ densities.

Corollary VIII.2. Let $A_i, 1 \leq i \leq k$ be measurable subsets of $\mathbb{R}^n$ with finite volume. Then

$$\ast_{i \in [k]} \|A_i(0) \leq \ast_{i \in [k]} \|A_i^*(0).$$

Proof: Clearly, we can assume all the sets have non-zero volume. We first prove the corollary under the assumption that one of the sets, say $A_1$, is open. But the indicator of an open set is an lower semicontinuous function. Hence by Fatou’s lemma,

$$\liminf_{m \to x} \int \|A_1(x_m - y) \ast_{2\leq i \leq k} A_i(y) dy \geq \int \liminf_{m \to x} I A_1(x_m - y) \ast_{2\leq i \leq k} A_i(y) dy$$

$$\geq \int I A_1(x - y) \ast_{2\leq i \leq k} A_i(y) dy,$$

which shows that $\ast_{i \in [k]} \|A_i(x)$ is lower-semicontinuous. Similarly, $\ast_{i \in [k]} \|A_i^*(x)$ is also lower-semicontinuous. Now the claimed result follows from the $p = \infty$ case of the Main Theorem, Lemma V.1(iii) and the classical fact that

$$\ast_{i \in [k]} \|A_i^*(0) = \sup_x \ast_{i \in [k]} \|A_i^*(x).$$

In the general case, by the regularity property of the Lebesgue measure, we can find a sequence of open set $B_m \supseteq A_1$, such that

$$\lim_{m \to \infty} |B_m| = |A_1|,$$

from which it follows that

$$\lim_{m \to \infty} \|B_m - I_{A_1} \|_1 = 0.$$  

(23)

By Lemma II.4, we also have:

$$\lim_{m \to \infty} \|B_m^* - I_{A_1^*} \|_1 = 0.$$  

(24)
The inequalities (23) and (24) imply easily that for each \( x \),
\[
\int \mathbb{I}_{B_m}(x - y) \ast 2 \leq i \leq k \mathbb{I}_{A_i}(y) dy \rightarrow \ast_{i \in [k]} \mathbb{I}_{A_i}(x),
\]
\[
\int \mathbb{I}_{B_m}(x - y) \ast 2 \leq i \leq k \mathbb{I}_{A_i}(y) dy \rightarrow \ast_{i \in [k]} \mathbb{I}_{A_i}(x).
\]
By what has been proved,
\[
\int \mathbb{I}_{B_m}(-y) \ast 2 \leq i \leq k \mathbb{I}_{A_i}(y) dy \leq \int \mathbb{I}_{B_m}(-y) \ast 2 \leq i \leq k \mathbb{I}_{A_i}(y) dy.
\]
Hence the desired result follows by taking the limit as \( m \) goes to \( \infty \).

**Remark 12.** By taking \( A_1 \) to be \( -A_1 \), we obtain
\[
\int \mathbb{I}_{A_i}(y) \ast 2 \leq i \leq k \mathbb{I}_{A_i}(y) dy \leq \int \mathbb{I}_{A_i}(y) \ast 2 \leq i \leq k \mathbb{I}_{A_i}(y) dy.
\]

For any \( k \) densities, \( f_i, 1 \leq i \leq k \), we have the following layer cake representation:
\[
\int f_i(y) \ast 2 \leq i \leq k f_i(y) dy = \int_0^\infty dt_1 \cdots \int_0^\infty dt_k \int \mathbb{I}_{A_{t_i}}(y) \ast 2 \leq i \leq k \mathbb{I}_{A_{t_i}}(y) dy,
\]
where \( A_{t_i} = \{ x : f_i(x) > t_i \} \). This, combined with Lemma II.1, will give us:
\[
\int f_i(y) \ast 2 \leq i \leq k f_i(y) dy \leq \int f_i^*(y) \ast 2 \leq i \leq k f_i^*(y) dy.
\]
This is precisely Theorem II.6.

Another corollary of the Main Theorem is a Fisher information inequality. **From now on, until the end of this section, we assume** \( n = 1 \). We first define Fisher information of a probability measure.

**Definition VIII.3.** If a density \( f \) on the real line is locally absolutely continuous, with the derivative \( f' \) (defined almost everywhere), then its Fisher information \( I(f) \) is defined by
\[
I(f) = \int_{\{x: f(x) > 0\}} \frac{f'^2(x)}{f(x)} dx.
\]
For other densities and for probability measures without densities we define \( I \) to be \( +\infty \).

We will sometimes abuse notation by writing \( I(X) \) to mean \( I(f) \), if \( X \) is distributed according to \( f \). One can show that if \( I \) is finite, then the derivative of the density \( f \) is absolutely integrable [11]. We also need some important properties of Fisher information. One is that Gaussian convolution decreases Fisher information. This is a slight extension of the argument in [3]. We give this as a lemma and give a complete proof.

**Lemma VIII.4.**
\[
I(X + G) \leq I(X),
\]
where \( X \) is any random variable with a density \( f \) and \( G \) is a non-degenerate Gaussian with density \( g \), independent of \( X \).

**Proof:** Clearly, we can assume \( I(X) < +\infty \). Let \( S = X + G \), with density \( h(x) \). It is easy to see that \( h(x) \) is strictly positive and differentiable, with
\[
h'(x) = Ef'(x - X),
\]
which can be justified by dominated convergence. By the finiteness of Fisher information, \( f' \) is absolutely integrable. Hence
\[
h'(x) = \int \int g'(x - z)[t \leq z]f'(t) dt dz = \int g(x - t)f'(t) dt,
\]
which can justified by Fubini’s Theorem. We now show that
\[
\frac{h'(S)}{h(S)} = E \left[ \frac{f'(X)\{ f(X) > 0 \}}{f(X)} \bigg| S \right].
\]
(25)

Once (25) is shown, we can apply conditional version of Jensen’s inequality to obtain:
\[
\left( \frac{h'(S)}{h(S)} \right)^2 \leq E \left[ \left( \frac{f'(X)\{ f(X) > 0 \}}{f(X)} \right)^2 \bigg| S \right],
\]
which, after taking expectations, will give us the desired result. To show (25), it suffices to show:

$$E \left[ m(S) \frac{h'(S)}{h(S)} \right] = E \left[ \frac{f'(X) \{ f(X) > 0 \}}{f(X)} m(S) \right],$$

(26)

where $m(x)$ is any bounded measurable function. The left side of inequality (26) is

$$\int m(x) h'(x) dx = \int \int m(x) f'(t) g(x-t) dx dt$$

$$= \int \int m(z + t) f'(t) \{ f(t) > 0 \} f(t) g(z) dz dt$$

$$+ \int \int m(z + t) f'(t) \{ f(t) = 0 \} g(z) dz dt,$$

where we have used Fubini’s Theorem implicitly. The first term on the right side of the above display is precisely the right side of the inequality (26), while the fact that the second term is zero follows from (30).

The other one is the lower semicontinuity of Fisher information functional with respect to weak convergence topology [11]. Lemma VIII.4 and the lower semicontinuity allow us to extend [3, Lemma 6.1] to $t = 0$, which is of some independent interest.

**Lemma VIII.5.** Assume that $X$ has finite second moment, and a density with $h(X) > -\infty$. Then

$$\frac{dh(X + \sqrt{t}Z)}{dt} \bigg|_{t=0} = \frac{1}{2} I(X),$$

where $I(X)$ might be infinity and $Z$ is a standard normal, independent of $X$.

**Proof:** For $t > 0$, Lemma 6.1 in [3] implies that

$$\frac{dh(X + \sqrt{t}Z)}{dt} = \frac{1}{2} I(X + \sqrt{t}Z).$$

Remark 10 shows that $h(X + \sqrt{t}Z)$ is continuous at $t = 0$. Hence we can apply the mean value Theorem to obtain:

$$\lim_{k \to \infty} \frac{h(X + \sqrt{t_k}Z) - h(X)}{t_k} = \frac{1}{2} \lim_{k \to \infty} I(X + \sqrt{t_k}Z),$$

where $t_k \downarrow 0, s_k \downarrow 0$. Now by Lemma VIII.4, $I(X + \sqrt{t_k}Z) \leq I(X)$. This, combined with lower-semicontinuity of Fisher information, allows us to conclude.

The final one is the continuity of the Fisher information functional after Gaussian convolution (see [33]).

**Lemma VIII.6.** Let $G$ be a non-degenerate Gaussian, independent of $X$. Then $I(X + G)$ is a continuous functional of the density of $X$, with respect to the topology of weak convergence.

**Corollary VIII.7.** For any density $f$, let $X$ be a random vector distributed according to $f$ and $X^*$ distributed according to $f^*$. Then

$$I(X) \geq I(X^*).$$

**Proof:** We first assume that $X$ has finite second moment and that $h(X) > -\infty$. The $p = 1$ case of the Main Theorem implies that:

$$h(X + \sqrt{t}Z) \geq h(X^* + \sqrt{t}Z),$$

where $Z$ is a standard normal and all random variables are independent. By Lemma II.2 and our assumption,

$$-\infty < h(X) = h(X^*) < +\infty.$$  

(27)

This further implies:

$$\frac{h(X + \sqrt{t}Z) - h(X)}{t} \geq \frac{h(X^* + \sqrt{t}Z) - h(X^*)}{t}.$$  

Now we can apply Lemma VIII.5 to conclude (note that by Lemma III.1, $X^*$ also has finite second moment).

In the general case, we use an approximation argument. Specifically, note that continuous functions with compact support is dense in the space $L_1(\mathbb{R})$. Hence we can find $f_n \in C_c(\mathbb{R})$, such that:

$$\| f_n - f \|_1 \to 0.$$  

Then we have

$$\| f_n^+ - f \|_1 \to 0.$$
and

\[ \| f_n^+ - f \|_1 \to 0. \]

Hence we conclude we can find a sequence of densities \( g_n \in C_c(\mathbb{R}) \), such that:

\[ \| g_n - f \|_1 \to 0. \]

We now show that

\[ I(X_n + \sqrt{t} Z) \geq I((X_n + \sqrt{t} Z)^*), \tag{28} \]

where \( X_n \) has density \( g_n \) and all random variables are independent. Clearly, \( X_n + \sqrt{t} Z \) has finite second moment. Since the density of \( X_n + \sqrt{t} Z \) is bounded, \( h(X_n + \sqrt{t} Z) > -\infty. \) Hence (28) follows from what has been proved.

Finally, taking the limit as \( n \) goes to infinity and applying Lemma VIII.6 to the left of (28) and lower semicontinuity to the right of (28), we see that (note that the density of \( (X_n + \sqrt{t} Z)^* \) converges to that of \( (X + \sqrt{t} Z)^* \) in total variation distance due to Lemma II.4):

\[ I(X + \sqrt{t} Z) \geq I((X + \sqrt{t} Z)^*). \tag{29} \]

We then apply Lemma VIII.4 to the left of (29) and lower semicontinuity to the right of (29) to obtain (note that the density of \( (X + \sqrt{t} Z)^* \) converges to that of \( X^* \) in total variation distance due to Lemma II.4):

\[ I(X) \geq I(X^*), \]

by taking the limit as \( t \) goes to 0.

\[ \blacksquare \]

Remark 13. Two standard facts are used implicitly about Gaussian convolution in the above:

\[ \| g_t * g_n - g_t * f \|_1 \leq \| g_n - f \|_1; \]

\[ \| g_t * f - f \|_1 \to 0, \]

as \( t \) goes to 0.

Remark 14. The above inequality is completely equivalent to the Polya-Szego inequality for \( p = 2 [17] \). Suppose that a non-negative function \( g \), locally absolutely continuous, satisfies \( \| g \|_2 < \infty \) and \( \| g' \|_2 < \infty \) \( g \in H_2(\mathbb{R}) \). We assume for now that \( \| g \|_2 = 1. \) Then \( f = g^2 \) is a probability density and

\[ I(f) = 4 \int_{\{x: g(x) > 0\}} g^2(x) dx < +\infty. \]

By the above corollary and Lemma II.3, we know that \( f^* \) is also absolutely continuous and that

\[ f^* = (g^*)^2, \]

\[ I(f^*) = \int_{\{x: f^*(x) > 0\}} (f^{**}(x))^2 \frac{f^{**}(x)}{f^*(x)} dx \leq I(f) < +\infty. \]

We now show that \( g^* = \sqrt{f^*} \) is locally absolutely continuous. Fix \( \epsilon > 0, \) then

\[ \sqrt{f^*(b)} + \epsilon - \sqrt{f^*(a)} + \epsilon = \frac{1}{2} \int_a^b \frac{f^{**}}{\sqrt{f^* + \epsilon}} dx. \]

But for any non-negative measurable function \( f \), if \( f \) is differentiable at \( x_0 \) such that \( f(x_0) = 0 \), then we must have

\[ f'(x_0) = 0. \tag{30} \]

(This fact appears as [11, Proposition 2.1], with a complicated proof, but it is actually a simple consequence of the definitions. Indeed, non-negativity gives us that the left derivative \( \partial_- f(x_0) \leq 0 \) and the right derivative \( \partial_+ f(x_0) \geq 0 \), while differentiability tells us that \( \partial_- f(x_0) = \partial_+ f(x_0) \), so that both are 0, and hence so is \( f'(x_0) \).) Hence

\[ \sqrt{f^*(b)} + \epsilon - \sqrt{f^*(a)} + \epsilon = \frac{1}{2} \int_a^b \{ f^*(x) > 0 \} \frac{f'^*(x)}{\sqrt{f^*(x) + \epsilon}} dx. \]

By dominated convergence (using the finiteness of the Fisher information), we obtain

\[ \sqrt{f^*(b)} - \sqrt{f^*(a)} = \frac{1}{2} \int_a^b \{ f^*(x) > 0 \} \frac{f'^*(x)}{\sqrt{f^*(x)}} dx, \]
which implies that \( g^* \) is locally absolutely continuous. Hence,

\[
I(f^*) = 4 \int_{\{x: g^*(x) > 0\}} g'^*(x)^2 dx.
\]

Finally, the argument that leads to (30) shows also that

\[
I(f^*) = 4 \int g'^2 dx, \\
I(f) = 4 \int g^2 dx.
\]

Hence we obtain

\[
\int g'^2 dx \leq \int g^2 dx. \tag{31}
\]

By Lemma II.2, the assumption \( ||g||_2 = 1 \) can be removed easily. Thus we have shown that if a non-negative function \( g \in \mathcal{H}_2(\mathbb{R}) \), then (31) is true. This is precisely the Polya-Szego inequality for \( p = 2 \).

**Remark 15.** Here is another perspective on Corollary VIII.7. Entirely similar to its proof, we can deduce the following inequality from Theorem I.1 (under the same assumptions of Theorem I.1):

\[
I(f) \geq I(g) = \frac{1}{N(f)}, \tag{32}
\]

where \( g \) is a centered normal density such that \( h(g) = h(f) \) and \( N(f) = \frac{1}{\pi} e^{2h(f)} \), the entropy power of \( f \). This inequality, first proved by Stam [46] (by combining the entropy power inequality and de Bruijn’s identity) is sometimes called the “isoperimetric inequality for entropy”. Hence, just as our Main Theorem strengthens the entropy power inequality, Corollary VIII.7 can be seen as a strengthening of the isoperimetric inequality for entropy since it inserts \( I(f^*) \) in between \( I(f) \) and \( I(g) \) in the inequality (32).

**Remark 16.** It is a folklore fact that the isoperimetric inequality for entropy is related to the Gaussian logarithmic Sobolev inequality (usually attributed to Gross [27], who developed its remarkable applications to hypercontractivity and mathematical physics) under a finite variance constraint on \( f \). Indeed, if \( \tilde{g} \) is a normal density with the same mean and variance as \( f \), then

\[
h(f) = h(\tilde{g}) - D(f\|g),
\]

which means that the inequality (32) can be rewritten as

\[
N(\tilde{g}) I(f) \geq e^{2D(\|f\|g)}.
\]

Using the fact that \( N(\tilde{g}) \) is just the variance of \( \tilde{g} \) and hence the variance \( \sigma_f^2 \) of \( f \), we have

\[
D(f\|g) \leq \frac{1}{2} \log \left[ \sigma_f^2 I(f) \right].
\]

Since \( \log x \leq x - 1 \) for \( x > 0 \), we obtain

\[
D(f\|g) \leq \frac{1}{2} J(f), \tag{33}
\]

where

\[
J(f) := \sigma_f^2 I(f) - 1
\]

is the standardized Fisher information of \( f \), which is location and scale invariant. The inequality (33) is a form of the Gaussian logarithmic Sobolev inequality. (In a related fashion, Toscani [48] uses Costa’s concavity of entropy power [22] to prove Nash’s inequality with the sharp constant.)

**IX. AN APPLICATION: BOUNDING THE ENTROPY OF THE SUM OF TWO UNIFORMS**

**Proposition IX.1.** Let \( X \) and \( Y \) be independent uniform distributions on two Borel sets \( M_1 \) and \( M_2 \), both with finite, non-zero volume. Then the following is true:

\[
h(X + Y) \geq h(X^* + Y^*) = \log \left( B\left(\frac{n + 1}{2}, \frac{1}{2}\right) V_n(1)r_1^n r_2^n \right) + \int_0^{r_1 + r_2} \log \left( \frac{1}{g(r)} \right) \frac{ng(r) r^{n-1}}{r_1^n r_2^n B\left(\frac{n+1}{2}, \frac{1}{2}\right)} dr,
\]

where \( B(\cdot, \cdot) \) is the beta function, \( V_n(1) \) is the volume of the \( n \) dimensional unit ball and

\[
V_n(1) r_i^n = |M_i|,
\]

for \( i = 1, 2 \). The function \( g(r) \) is defined in the following way: if \( r > |r_1 - r_2| \),

\[
g(r) = r_1^n h\left( \arcsin\left( \frac{r_2^2 - r_1^2 + r_1^2}{2rr_1} \right) \right) + r_2^n h\left( \arcsin\left( \frac{r_2^2 - r_1^2 + r_2^2}{2rr_2} \right) \right),
\]

if \( r < |r_1 - r_2| \),

\[
g(r) = r_1^n h\left( \arcsin\left( \frac{r_2^2 - r_1^2 + r_1^2}{2rr_1} \right) \right) + r_2^n h\left( \arcsin\left( \frac{r_2^2 - r_1^2 + r_2^2}{2rr_2} \right) \right).
\]
where \( h(\theta) = \int_0^{\theta} \cos^n(x) \, dx \) and if \( r < |r_1 - r_2| \),

\[
g(r) = \min(r_1^n, r_2^n)B\left(\frac{n + 1}{2}, \frac{1}{2}\right).
\]

**Proof:** By the \( p = 1 \) case of the Main Theorem, we get

\[
h(X + Y) \geq h(X^* + Y^*),
\]

where \( X^* \) and \( Y^* \) are independent uniform distributions on the balls centered at the origin, with radius \( r_1 \) and \( r_2 \) respectively. We just need to compute explicitly the density of \( X^* + Y^* \). This is given by:

\[
\frac{1}{|M_1||M_2|} \int I_{M_1^*}(x - y)I_{M_2^*}(y)\,dy.
\]

Note that \( \int I_{M_1^*}(x - y)I_{M_2^*}(y)\,dy \) is nothing but the volume of the intersection of two balls \( |B(x, r_1) \cap B(0, r_2)| \). If \( ||x|| < |r_2 - r_1| \), this volume is that of the smaller ball; if \( ||x|| \geq r_1 + r_2 \), this volume is zero; if \( |r_2 - r_1| \leq ||x|| \leq r_1 + r_2 \), this volume is the sum of the volumes of two spherical caps. Then if \( ||x|| = r \), the spherical cap with radius \( r_2 \) will have height \( h_2 = \frac{r_2^2 - (r-r_2)^2}{2r_2} \) and the cap with radius \( r_1 \) will have height \( h_1 = \frac{r_1^2 - (r-r_1)^2}{2r_1} \). But the volume of a spherical cap with given radius and height is classical and can be computed from elementary calculus. The rest follows from simple algebra.  

**Corollary IX.2.** If \( n = 1 \) and \(|M_2| > |M_1|\), then

\[
h(X + Y) \geq h(Y) + \frac{|M_1|}{2|M_2|} \geq \log \left( |M_2| + \frac{1}{2}|M_1| \right).
\]

**Proof:** The first inequality follows from by doing all the integrations explicitly. The second follows from the simple inequality \( e^x \geq 1 + x \).

**Remark 17.** Note that since \( X + Y \) is supported in \( M_1 + M_2 \), we have the upper bound

\[
|M_1 + M_2| \geq e^{h(X + Y)}, \tag{34}
\]

since for all random vectors supported in \( M_1 + M_2 \), the uniform distribution on \( M_1 + M_2 \) maximizes the entropy [23]. On the other hand, entropy power will give us the lower bound:

\[
e^{\frac{1}{2}h(X + Y)} \geq \sqrt{|M_1|^\frac{1}{2} + |M_2|^\frac{1}{2}}. \tag{35}
\]

Combining the inequalities (34) and (35) gives an inequality weaker than the Brunn-Minkowski inequality:

\[
|M_1 + M_2|^\frac{1}{2} \geq \sqrt{|M_1|^\frac{1}{2} + |M_2|^\frac{1}{2}} = |M_2|^{\frac{1}{2}} \sqrt{1 + \lambda^2} = |M_2|^{\frac{1}{2}} (1 + \frac{1}{2} \lambda^2 + o(\lambda^2)), \tag{36}
\]

where \( \lambda = (\frac{|M_1|}{|M_2|})^{\frac{1}{2}} \). In contrast, the Brunn-Minkowski inequality gives

\[
|M_1 + M_2|^\frac{1}{2} \geq |M_2|^{\frac{1}{2}} (1 + \lambda).
\]

(37)

It is well known and easy to see that (37) implies the following isoperimetric inequality:

\[
\lim_{\epsilon \downarrow 0} \inf \frac{|eM_1 + M_2| - |M_2|}{\epsilon} \geq n|M_2| \left( \frac{|M_1|}{|M_2|} \right)^{\frac{1}{n}}, \tag{38}
\]

where the equality holds when \( M_1 \) and \( M_2 \) are homothetic convex bodies. On the other hand, (36) will only give

\[
\lim_{\epsilon \downarrow 0} \inf \frac{|eM_1 + M_2| - |M_2|}{\epsilon} \geq 0.
\]

Hence, when \( \lambda \) is small, (36) is especially poor compared to (37) and is not reflecting the correct behavior as \( \lambda \) goes to 0. But note that when \( n = 1 \), Corollary IX.2 implies the following entropic isoperimetric inequality:

\[
\lim_{\epsilon \downarrow 0} \frac{h(Y + eX) - h(Y)}{\epsilon} \geq \frac{1}{2} \frac{|M_1|}{|M_2|}, \tag{39}
\]
with equality when $M_1$ and $M_2$ are symmetric intervals. This implies
\[
\lim \inf_{\epsilon \downarrow 0} \frac{e^{h(Y + \epsilon X)} - e^{h(Y)}}{\epsilon} \geq \frac{1}{2} \frac{|M_1|}{|M_2|},
\]
Using (34) again, we get
\[
\lim \inf_{\epsilon \downarrow 0} \frac{|\epsilon M_1 + M_2| - |M_2|}{\epsilon} \geq \frac{1}{2} \frac{|M_1|}{|M_2|},
\]
which, although still weaker than (38), is reflecting the correct behavior.

Remark 18. If $X$ is a uniform distribution on a symmetric interval and if we define
\[
\tilde{I}(Y) = \lim \inf_{\epsilon \downarrow 0} \frac{h(Y + \epsilon X) - h(Y)}{\epsilon},
\]
then the inequality (39) can be rewritten as
\[
\tilde{I}(Y) \geq \tilde{I}(Y^*).
\]
This is very similar to Corollary VIII.7, but with the role of Gaussian replaced by a uniform.

Remark 19. In the case where the sets under consideration are convex, much better bounds can be given using the fact that the convolution of the uniforms yields a $\kappa$-concave measure for $\kappa > 0$. More details, including the definition of $\kappa$-concavity, can be found in [7].

X. ANOTHER APPLICATION: ENTROPY OF MARGINALS OF LÉVY PROCESSES

In this section, we develop some simple applications to Lévy processes of our results. The main result is Theorem X.3, whose method of proof is very similar to that of [2, Theorem 1.1] and [26, Lemma 3.2] (the latter deals with the notion of symmetric rearrangements around infinity instead of the usual symmetric rearrangements, though). The key is that Lévy processes are weak limits of compound Poisson processes. In fact, the inequality (41), an intermediate step in the proof of our Theorem X.3, is readily implied by [2, Theorem 1.1], although we give a full proof here for completeness.

Proposition X.1. Suppose the $n$-dimensional process \( \{X_t : t \geq 0\} \) can be represented as

\[
X_t = A^\frac{1}{2} \cdot B_t + \sum_{i=1}^{N_t} Y_i,
\]
where $A$ is an $n$ by $n$ symmetric and strictly positive definite matrix, $B_t$ is the standard $n$ dimensional Brownian motion, $N_t$ is the Poisson process with rate $\lambda \geq 0$, independent of the process $B_t$, and $Y_i$s, independent of the processes $B_t$ and $N_t$, are i.i.d. random vectors with density $f$.

We now define the rearranged process $Z_t$ to be:

\[
Z_t = \det \left( A^\frac{1}{2} \right) B_t + \sum_{i=1}^{N_t} Y_i^*,
\]
where $B_t$ and $N_t$ are as before and $Y_i^*$s, independent of the processes $B_t$ and $N_t$, are i.i.d. random vectors with density $f^*$. We then have:

\[
h_p(X_t) \geq h_p(Z_t),
\]
for $t > 0$ and $0 < p \leq +\infty$.

Proof: Let the density of $A^\frac{1}{2} \cdot B_t$ be $g_t$, the density of $X_t$ be $p_t$ and the density of $Z_t$ be $q_t$. Then

\[
p_t = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} g_t \ast f \ast \cdots \ast f,
\]
and

\[
q_t = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} g_t^* \ast f^* \ast \cdots \ast f^*.
\]

It suffices to show $p_t \prec q_t$ by Proposition VII.3. We recall the following representation [17] (which has been used several times in Section VII):

\[
\int_{B(0,r)} f^*(x)dx = \sup_{|C|=|B(0,r)|} \int_C f(x)dx,
\]

(40)
for any density \( f \), where \( B(0, r) \) is the open ball with radius \( r \). Hence
\[
\int_{B(0, r)} p_t^*(x)dx = \sup_{|C|=|B(0, r)|} \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \int_{C} g_t * \underbrace{f * \cdots * f}_{k} dx.
\]
By Theorem II.6,
\[
\int_{C} g_t * \underbrace{f * \cdots * f}_{k} dx \leq \int_{C^*} g_t * \underbrace{f * \cdots * f}_{k} dx
\]
Since \( C^* = B(0, r) \) by definition, we get
\[
\int_{B(0, r)} p_t^*(x)dx \leq \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \int_{B(0, r)} g_t * \underbrace{f * \cdots * f}_{k} dx = \int_{B(0, r)} q_t(x)dx.
\]
By (40) again,
\[
\int_{B(0, r)} q_t(x)dx \leq \int_{B(0, r)} q_t^*(x)dx.
\]
We finally get:
\[
\int_{B(0, r)} p_t^*(x)dx \leq \int_{B(0, r)} q_t^*(x)dx,
\]
which shows \( p_t \prec q_t \).

From now until the end of this section, we assume that the Brownian part of the standard Lévy process \( \{X_t\} \) is non-degenerate and that the Lévy measure \( \nu \) is locally absolutely continuous. Hence
\[
\nu(C) = \int_{C} m(x)dx
\]
for any Borel set \( C \subset \mathbb{R}^n \setminus \{0\} \). Since \( \nu \) is a Lévy measure, we have [38]:
\[
\int_{\mathbb{R}^n \setminus \{0\}} \min(1, |x|^2)m(x)dx < +\infty.
\]
Note that for \( t > 0 \),
\[
\int_{\{x : m(x) > t\}} dx \leq 1 + \frac{1}{t} \int_{|x| \geq 1} m(x)dx < +\infty.
\]
Hence \( m^*(x) \) is well defined. We now show \( \nu^*(dx) = m^*(x)dx \) also defines a Lévy measure.

**Lemma X.2.**
\[
\int_{\mathbb{R}^n \setminus \{0\}} \min(1, |x|^2)m^*(x)dx < +\infty.
\]

**Proof:** Define \( m_n(x) = m(x)\{x > \frac{1}{n}\} \). Then \( \int m_n(x)dx < +\infty \). Since rearrangement preserves \( L_p \) norm (Lemma II.2),
\[
\int m_n(x)dx = \int m_n^*(x)dx.
\]
By Lemma III.1, we get:
\[
\int_{\mathbb{R}^n \setminus \{0\}} \min(1, |x|^2)m_n^*(x)dx \leq \int_{\mathbb{R}^n \setminus \{0\}} \min(1, |x|^2)m_n(x)dx \leq \int_{\mathbb{R}^n \setminus \{0\}} \min(1, |x|^2)m(x)dx
\]
Note that
\[
m_n^*(x) = \int_{0}^{\infty} \mathbf{1}_{(S_t \cap \{x : x > \frac{1}{n}\})} \cdot (x)dt,
\]
where \( S_t = \{x : m(x) > t\} \). By monotone convergence,
\[
\left| \left( S_t \cap \left\{ x : x > \frac{1}{n} \right\} \right)^* \right| = \left| S_t \cap \left\{ x : x > \frac{1}{n} \right\} \right| \uparrow |S_t| = |S_t^*|.
\]
Suppose Theorem X.3. But since Lemma X.5. Since \( (S_i \cap \{ x : x > \frac{1}{n} \})^*, n = 1, 2, \ldots \), are open balls, we must have \( \mathbb{1}_{(S_i \cap \{ x : x > \frac{1}{n} \})^*}(x) \uparrow \mathbb{1}_{S_i^*}(x) \). By monotone convergence and the definition of rearrangement again, we obtain
\[
m_n^*(x) \uparrow m^*(x).
\]
Another application of monotone convergence will give:
\[
\int_{\mathbb{R}^n \setminus \{0\}} \min(1, |x|^2)m^*(x)dx \leq \int_{\mathbb{R}^n \setminus \{0\}} \min(1, |x|^2)m(x)dx.
\]

We now recall the Lévy-Khinchine formula [38]. For any Lévy process \( \{ X_t \} \), we have:
\[
\mathbb{E}[e^{itX_t}] = e^{-t\psi(u)},
\]
where
\[
\psi(u) = \frac{1}{2}(Au, u) - i\gamma \cdot u + \int (1 - e^{iuA} + iu \cdot xI_1(|x| \leq 1))\nu(dx).
\]
We call \( (A, \gamma, \nu) \) the Lévy triple of \( X_t \). We define the rearranged process \( Z_t \) to be a Lévy process with Lévy triple \( (|A|^\frac{1}{2}I_{\mathbb{R}^n}, 0, \nu^*) \). We can show the following

**Theorem X.3.** Suppose \( A \) is non-degenerate, then
\[
h_p(X_t) \geq h_p(Z_t),
\]
where \( 0 < p \leq 1 \) and \( t > 0 \). Moreover, if \( 0 < t_1 < t_2 < \cdots < t_n \), then
\[
h(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \geq h(Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}).
\]
To show this, we need two additional lemmas:

**Lemma X.4.** For \( p \in (0, 1) \cup (1, \infty) \), \( h_p(f \ast \mu) \geq h_p(f) \), for any two density \( f \) and any probability measure \( \mu \). If \( f \) is bounded, then this is also true for \( p = 1 \).

*Proof:* This fact is well known but let us sketch the simple proof here. We will only prove the case when \( p \in (0, 1) \). The other case is entirely similar. By Jensen’s inequality,
\[
(f \ast \mu)^p(x) \geq \int f^p(x - y)\mu(dy).
\]
Integrating the above with respect to \( x \) and rearranging, we get the desired result. Finally, when \( f \) is bounded, we may apply Lemma V.1 to take the limit as \( p \) goes to 1. ■

**Lemma X.5.** Let \( g \) be a non-degenerate Gaussian density. Then for \( 0 < p \leq 1 \), \( h_p(\mu \ast g) \), as a functional (of \( \mu \)) on the space of all probability measures, is lower semi-continuous with respect to the weak convergence topology.

*Proof:* Let \( \mu_n \) be a sequence of probability measures converging weakly to \( \mu \) and assume \( g \leq C \). Then by definition of weak convergence, for each \( x \),
\[
\mu_n \ast g(x) \to \mu \ast g(x).
\]
When \( p \neq 1 \), by Fatou’s lemma,
\[
\liminf_n h_p(\mu_n \ast g) \geq h_p(\mu \ast g).
\]
When \( p = 1 \), we can apply an argument in [31] to conclude. For completeness, let us sketch the argument here. Note that it suffices to show
\[
\liminf_n - \int \log \left( \frac{\mu_n \ast g}{C} \right) \frac{\mu_n \ast g}{C} dx \geq - \int \log \left( \frac{\mu \ast g}{C} \right) \frac{\mu \ast g}{C} dx.
\]
But since \( \mu_n \ast g \leq C \) and \( \mu \ast g \leq C \), the above also follows from Fatou’s lemma. ■

Now we finish the proof of Theorem X.3.

*Proof:* By the Lévy-Itô decomposition [38], we can write
\[
X_t = X^n_t + Y^n_t,
\]
where \( X^n \) is a Lévy process with Lévy triple \( (A, \gamma, \nu_n) \), with \( \nu_n(dx) = m_n(x)dx = m(x)\{ x > \frac{1}{n} \}dx \) and the process \( Y^n \) is independent of \( X^n \). Clearly, the density of \( X^n_t \) is bounded. By Lemma X.4,
\[
h_p(X_t) \geq h_p(X^n_t).
\]
\(X_t^n\) can be written as the sum of a Brownian motion (with constant drift) and an independent compound Poisson process [38]. Hence by Proposition X.1,
\[
h_p(X_t^n) \geq h_p(Z_t^n),
\]
where \(Z_t^n\) is a Lévy process with Lévy triple \((|A|^\frac{1}{2} B_t + U_t^n, 0, m_t^n(x)dx)\). Again by the Lévy -Ito decomposition, we can write
\[
Z_t^n = |A|^\frac{1}{2} B_t + U_t^n,
\]
\[
Z_t = |A|^\frac{1}{2} B_t + U_t,
\]
where \(B_t\) is a standard Brownian motion independent of \(U_t^n\) and \(U_t\) is a Lévy process with Lévy triple \((0, 0, m_t^n(x)dx)\) and \(U\) is a Lévy process with Lévy triple \((0, 0, m^*(x)dx)\). Finally, since \(m_t^n(x) \uparrow m^*(x), U_n(t)\) converges weakly to \(U_t\) for each \(t\). Hence by Lemma X.5,
\[
\lim \inf h_p(Z_t^n) \geq h_p(Z_t).
\]
Combining, we get:
\[
h_p(X_t) \geq h_p(Z_t).
\]
Finally, by the chain rule and the markov property of Lévy process, we have for any Lévy process \(X\),
\[
h(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = \sum_{i=1}^{n} h(X_{t_i} | X_{t_{i-1}}).
\]
By the independent and stationary increment property of Lévy process and translation invariance of entropy,
\[
h(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = \sum_{i=1}^{n} h(X_{t_i - t_{i-1}}).
\]
A similar expression holds for \(Z\). Hence we can conclude.

\section{XI. Yet another Proof of the classical Entropy Power Inequality}

The goal of this section is to give a new proof the entropy power inequality (Theorem I.1) starting from the Main Theorem. We comment here that in this section we actually only need the Main Theorem for \(p = 1\) and \(k = 2\). By Remark 10, to prove Theorem I.1, we can consider the case when the two densities are bounded, strictly positive and have finite covariance matrices. These will be assumed throughout this section. For convenience, we will use the following well-known equivalent formulation of the entropy power inequality [24]:
\[
h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda h(X) + (1 - \lambda) h(Y),
\]
for all \(0 < \lambda < 1\), where \(X\) has density \(f_1\) and \(Y\) has density \(f_2\). By the Main Theorem and Lemma II.2, we can assume that \(f_1\) and \(f_2\) are spherically symmetric decreasing. Note that by Lemma II.2, Lemma VI.1 and Lemma III.1, if a density \(f\) is bounded, strictly positive and has finite covariance matrix, then so is \(f^\ast\). Hence we will assume from now that the two densities are spherically symmetric decreasing, bounded, strictly positive and have finite covariance matrices.

We first show that an EPI comes almost for free if we assume identical distribution. The case when \(\lambda = \frac{1}{2}\) seems to be folklore; we learned it from Andrew Barron several years ago. For completeness, we sketch the easy proof for all \(\lambda\).

\textbf{Proposition XI.1.} Fix any \(0 < \lambda < 1\). Suppose that \(X\) and \(Y\) have the same distribution. Then:
\[
h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq h(X),
\]
\textbf{Proof:} By independence, we have:
\[
h(X, Y) = h(X) + h(Y).
\]
By spherical symmetry (in fact, we only need central symmetry) and i.i.d. assumption, we have
\[
\sqrt{\lambda}X + \sqrt{1-\lambda}Y = \lambda^d \sqrt{1-\lambda}X + \sqrt{\lambda}Y.
\]
By the scaling property for entropy,
\[
h(X, Y) = h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y, \sqrt{1-\lambda}X + \sqrt{\lambda}Y).
\]
Now we can use subadditivity of entropy to conclude.

We now give a slightly involved proof of the full entropy power inequality starting from the Main Theorem. For notational simplicity, we assume \(n = 1\) until the end of this section. Our proof is inspired by and may be considered as an adaptation of Brascamp and Lieb’s proof of Young’s inequality with sharp constant [14].
First, we do a simple reduction. Since we assumed \( f \) and \( g \) are bounded symmetric decreasing, we can approximate these densities pointwise and monotonically from below by symmetric decreasing simple functions of the form \( f_n \) and \( g_n \):

\[
f_n = \sum_{i=1}^{k_n} c_i^n \mathbb{1}_{I_i^n},
\]

where \( \mathbb{1}_{I_i^n} \) are indicators of symmetric finite intervals with \( \mathbb{1}_{I_i^n} \leq \mathbb{1}_{I_{i+1}^n} \) and \( c_i^n > 0 \) (note that \( c_i^n > 0 \) since \( f_n \) is decreasing) and a similar expression for \( g_n \). By our assumption, we can show that for fixed \( 0 < \lambda < 1 \),

\[
h(f_n) \to h(f), \quad h(g_n) \to h(g),
\]

\[
h\left( \frac{1}{\sqrt{\lambda}} f_n \left( \frac{\cdot}{\sqrt{\lambda}} \right) \right) \ast \frac{1}{\sqrt{1-\lambda}} g_n \left( \frac{\cdot}{\sqrt{1-\lambda}} \right) \to h\left( \frac{1}{\sqrt{\lambda}} f \left( \frac{\cdot}{\sqrt{\lambda}} \right) \right) \ast \frac{1}{\sqrt{1-\lambda}} g \left( \frac{\cdot}{\sqrt{1-\lambda}} \right),
\]

where \( f_n \) and \( g_n \) are normalized versions of \( f_n \) and \( g_n \). This is because, as shown by Harremöes and Vignat [31], if a sequence of uniformly bounded densities converges pointwise to a density, and the first two moments also converge, then one has convergence of entropies. Hence, without loss of generality, we can assume that \( f \) and \( g \) are of the following form:

\[
f = \sum_{i=1}^{k_1} c_i^1 \mathbb{1}_{I_i^1}, \quad g = \sum_{i=1}^{k_2} c_i^2 \mathbb{1}_{I_i^2},
\]

where \( c_i^1 > 0, c_i^2 > 0 \).

The main trick is to use tensorization, or what physicists call the replica method. Consider \( X_1, X_2, \ldots, X_M \), which are \( M \) independent copies of \( X \), and independent of these, \( Y_1, Y_2, \ldots, Y_M \), which are \( M \) independent copies of \( Y \). Let \( X = (X_1, X_2, \ldots, X_M) \), \( Y = (Y_1, Y_2, \ldots, Y_M) \). The densities of \( X \) and \( Y \) are

\[
F(x_1, x_2, \ldots, x_M) = \prod_{i=1}^{M} f(x_i),
\]

\[
G(y_1, y_2, \ldots, y_M) = \prod_{i=1}^{M} g(y_i).
\]

Next, we show that \( F^* \) and \( G^* \) are both finite mixtures and the number of densities in the mixture grows at most polynomially in \( M \). It is easy to see \( F \) takes at most \( (M + 1)^{k_1} \) values and \( G \) takes at most \( (M + 1)^{k_2} \) values [14]. Hence just by looking at the definitions of rearrangements, one sees that \( F^* \) takes at most \( (M + 1)^{k_1} \) values and \( G^* \) takes at most \( (M + 1)^{k_2} \) values. This allows us to express \( F^* \) and \( G^* \) as, using the spherically symmetric decreasing property,

\[
F^* = \sum_{i=1}^{(M+1)^{k_1}} b_i \mathbb{1}_{I_{n_i}^1},
\]

\[
G^* = \sum_{j=1}^{(M+1)^{k_2}} b_j \mathbb{1}_{I_{n_j}^2},
\]

where \( b_i > 0, b_j > 0 \) and \( I_{n_i}^1, I_{n_j}^2 \) are indicators of \( M \) dimensional balls \( |\eta_i| \) and \( |\eta_j| \), centered at the origin and \( |\eta_i| \leq |\eta_{i+1}^1|, |\eta_j| \leq |\eta_{j+1}^2| \). Since both are probability densities, we have

\[
\sum_{i=1}^{(M+1)^{k_1}} b_i |\eta_i| = 1,
\]

\[
\sum_{j=1}^{(M+1)^{k_2}} b_j |\eta_j| = 1,
\]

\[
\sum_{i,j} b_i^1 b_j^2 |\eta_i| |\eta_j| = 1.
\]

We rewrite \( F^* \) and \( G^* \) as

\[
F^* = \sum_{i=1}^{(M+1)^{k_1}} b_i \mathbb{1}_{|\eta_i|} \mathbb{1}_{|\eta_i|},
\]

\[
G^* = \sum_{j=1}^{(M+1)^{k_2}} b_j \mathbb{1}_{|\eta_j|} \mathbb{1}_{|\eta_j|}.
\]
But $\frac{\|\eta_i\|}{|\eta_i|}$ and $\frac{1}{|\eta_i^2|}$ are exactly the uniform distributions on the balls $\eta_i^1$ and $\eta_i^2$. Hence both $F^*$ and $G^*$ are mixtures of uniform distribution on balls.

Two more ingredients, of independent interest, are needed for the proof. The first is the concavity of entropy and the following simple lemma, which may be thought of as a “reverse concavity” property of entropy when taking finite mixtures.

**Lemma XI.2.** Let $f$ be a finite mixture of densities, i.e.,

$$f = \sum_{i=1}^{n} c_i f_i,$$

where $c_i$ are nonnegative constants summing to 1, and $f_i$ are densities on $\mathbb{R}^n$. Then

$$h(f) \leq \sum_{i} c_i h(f_i) - \sum_{i} c_i \log c_i.$$

*In particular, $h(f) \leq \sum_{i} c_i h(f_i) + \log n$.*

**Proof:** By definition,

$$h(f) = -\sum_{i} c_i \int f_i \log(f_i) dx.$$

Since log is increasing,

$$-\log(f) \leq -\log(c_i f_i) = -\log(c_i) - \log(f_i).$$

Hence

$$h(f) \leq -\sum_{i} \log(c_i) c_i + \sum_{i} c_i h(f_i).$$

The term $-\sum_i \log(c_i) c_i$ is exactly the discrete entropy $H(c)$ of $c = (c_1, c_2, \ldots, c_n)$. Hence

$$h(f) \leq H(c) + \sum_{i} c_i h(f_i) \leq \log(n) + \sum_{i} c_i h(f_i).$$

\[\Box\]

**Remark 20.** For discrete entropy, this lemma is well known and can be found as an exercise in [23]; the standard way to prove it is using the data processing inequality for discrete entropy. For differential entropy (which is our focus), there is no data processing inequality, and we could not find Lemma XI.2 in the literature even though it has an extremely simple alternate proof.

The last ingredient is the following simple lemma, which is the EPI for uniform distributions on balls, but with an error term.

**Lemma XI.3.** Let $Z_1$ and $Z_2$ be two independent uniforms on $M$-dimensional balls centered at the origin, then

$$h(\sqrt{\lambda} Z_1 + \sqrt{1-\lambda} Z_2) \geq \lambda h(Z_1) + (1-\lambda) h(Z_2) + o(M),$$

where the $o$ symbol is uniform with respect to all pairs of balls centered at the origin.

**Proof:** We will only sketch the argument here. Let

$$Z = \sqrt{\lambda} Z_1 + \sqrt{1-\lambda} Z_2$$

and the radii of the balls corresponding to $Z_1$ and $Z_2$ be $b_1$ and $b_2$ respectively. Suppose the densities of $Z, Z_i$ are $f, f_i, i = 1, 2$. We now define two $M$-dimensional Gaussian densities:

$$g_i(x) = \left( \frac{M}{2\pi b_i^2} \right)^{\frac{M}{2}} e^{-\frac{M|x|^2}{2b_i^2}}, i = 1, 2,$$

and let $G = \sqrt{\lambda} G_1 + \sqrt{1-\lambda} G_2$, with density $g$, where $G_1$ and $G_2$ are independent random vectors with densities $g_1$ and $g_2$. We indicate that, using Stirling’s approximation, one can show (assuming $M$ even, without loss of generality):

$$f_i \leq \sqrt{\pi M} e^{\frac{|x|^2}{2\pi}} g_i,$$

$$f \leq \pi M e^{\frac{|x|^2}{2\pi}} g.$$

Hence it is easily seen that

$$D(Z\|G) = \lambda D(Z_1\|G_1) + (1-\lambda) D(Z_2\|G_2) + O(\log(M)).$$
where $D(\cdot \| \cdot)$ is the relative entropy and the $O(\log(M))$ means it is bounded by $\log(M)$ times a universal constant. Now some easy calculations show that

$$
D(Z\|G) = -h(Z) + \frac{M^2}{M+2} - \frac{M}{2} \log \left( \frac{M}{2\pi(\lambda b_1^2 + (1-\lambda)b_2^2)} \right),
$$

$$
D(Z_i\|G_i) = -h(Z_i) + \frac{M^2}{M+2} - \frac{M}{2} \log \left( \frac{M}{2\pi b_i^2} \right).
$$

Hence we get

$$
\begin{align*}
  h(Z) - \lambda h(Z_1) - (1-\lambda)h(Z_2) &= \frac{M}{2} \left( \log(\lambda b_1^2 + (1-\lambda)b_2^2) \\
  &- \lambda \log(b_1^2) - (1-\lambda) \log(b_2^2) \right) + O(\log(M)) \\
  &\geq O(\log(M)),
\end{align*}
$$

(42)

where the last step follows from concavity of $\log$ and the meaning of the $O$ symbol is as before.

\[\text{Remark 21.}\] The above proof gives very strong information about the entropy of the sum of two independent uniform on balls, in high dimensions. The last equality is in fact equivalent to

$$
\frac{2h(Z_1 + Z_2)}{M} = \log \left( e^{\frac{2h(Z_1)}{M}} + e^{\frac{2h(Z_2)}{M}} \right) + O(\log(M)).
$$

(43)

Hence we obtain the conclusion that asymptotically, EPI becomes an equality for two independent uniform distribution on balls in high dimensions. The expression (43) implies:

$$
e^{\frac{2h(Z_1 + Z_2)}{M}} \leq c_M \left( e^{\frac{2h(Z_1)}{M}} + e^{\frac{2h(Z_2)}{M}} \right),
$$

(44)

where $c_M$, depending only on $M$, goes to 1 as $M$ goes to infinity. This is not surprising because uniform distributions on high-dimensional balls are close to Gaussians. We may also note in passing that it was recently shown in [37] (and also in [13] with a slightly worse constant) that for IID, log-concave random vectors $U_1$ and $U_2$ taking values in $\mathbb{R}^M$,

$$
e^{\frac{2h(U_1 + U_2)}{M}} \leq 2 \left( e^{\frac{2h(U_1)}{M}} + e^{\frac{2h(U_2)}{M}} \right),
$$

while the more general reverse entropy power inequality of [5], [7] gives such an inequality for arbitrary independent log-concave random vectors (thus covering two balls of different radii), but with a non-explicit constant.

We now complete the proof of the original EPI. Let $Z_1^1$ and $Z_2^2$ be independent random vectors uniformly distributed on the balls $\eta_i^1$ and $\eta_i^2$.

$$
Mh(\sqrt{\lambda}X_1 + \sqrt{1-\lambda}Y_1) \overset{(a)}{=} h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \overset{(1)}{\geq} h(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^*) \\
\overset{(2)}{\geq} \sum_{i=1}^{(M+1)^{k_1}} \sum_{j=1}^{(M+1)^{k_2}} b_i^1 b_j^2 |\eta_i^1| |\eta_j^2| h(\sqrt{\lambda}Z_i^1 + \sqrt{1-\lambda}Z_j^2) \\
\overset{(3)}{\geq} \lambda \sum_{i=1}^{(M+1)^{k_1}} b_i^1 |\eta_i^1| h(Z_i^1) + (1-\lambda) \sum_{j=1}^{(M+1)^{k_2}} b_j^2 |\eta_j^2| h(Z_j^2) - C \log(M) \\
\overset{(4)}{\geq} \lambda h(X^*) + (1-\lambda) h(Y^*) - C \log(M) - \lambda k_1 \log(M+1) - (1-\lambda) k_2 \log(M+1) \\
\overset{(b)}{=} \lambda h(X) + (1-\lambda) h(Y) + O(\log(M)) \\
\overset{(c)}{=} M \lambda h(X_1) + M(1-\lambda) h(Y_1) + O(\log(M)).
$$

Here (a) follows from independence; (1) follows from $M$ dimensional version of the Main Theorem; (2) follows from concavity of entropy and the simple fact that convolution of mixtures is a mixture of convolutions; (3) follows from Lemma XI.3; (4) follows from Lemma XI.2; (b) follows from Lemma I.2; (c) follows again from independence. We finally get:

$$
Mh(\sqrt{\lambda}X_1 + \sqrt{1-\lambda}Y_1) \geq M \lambda h(X_1) + M(1-\lambda) h(Y_1) + O(\log(M)).
$$

Dividing both sides by $M$ and taking the limit as $M$ goes to infinity, we recover the full EPI.
ACKNOWLEDGMENT

We would like to thank the Department of Statistics at Yale University, where this research was begun, and the Department of Mathematics at the Indian Institute of Science, Bangalore, where this research was completed. We are also indebted to Bruce Hajek. Professor Hajek rediscovered the Riesz rearrangement inequality while solving an optimization problem for paging and registration in cellular networks with K. Mitzel and S. Yang in [28], and suggested in a talk in 2008 (that the authors were not fortunate enough to attend) that it should be interesting to information theorists. We discovered the slides of his talk after a version of this paper had been written with just the first proof of the Main Theorem, and they led us to the second, simpler and more general, proof in this paper. We are grateful to Sergey Bobkov for sharing two preprints of his recent papers [9], [10] with G. P. Chistyakov; although we received these after the first version of this paper was written, we added the discussion in Section IV in response to their results. Finally, we thank Almut Burchard for valuable feedback on an earlier version of this paper, as well as for supplying several relevant references.

REFERENCES

[1] A. Alvino, G. Trombetti, and P.-L. Lions, “On optimization problems with prescribed rearrangements,” Nonlinear Anal., vol. 13, no. 2, pp. 185–220, 1989. [Online]. Available: http://dx.doi.org/10.1016/0362-546X(89)90043-6
[2] R. Bañuelos and P. J. Mendoza-Hernández, “Symmetrization of Lévy processes and applications,” J. Funct. Anal., vol. 258, no. 12, pp. 4026–4051, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.jfa.2010.02.013
[3] A. R. Barron, “Monotonic central limit theorem for densities,” Department of Statistics, Stanford University, California, Tech. Rep. #50, 1984. [Online]. Available: http://www.stat.yale.edu/~arb/publications/files/monotoning%20central%20limit.pdf
[4] W. Beckner, “Inequalities in Fourier analysis,” Ann. of Math. (2), vol. 102, no. 1, pp. 159–182, 1975.
[5] S. Bobkov and M. Madiman, “Dimensional behaviour of entropy and information,” C. R. Acad. Sci. Paris Sér. I Math., vol. 349, pp. 201–204, Février 2011.
[6] ——, “The entropy per coordinate of a random vector is highly constrained under convexity conditions,” IEEE Trans. Inform. Theory, vol. 57, no. 8, pp. 4940–4954, August 2011.
[7] ——, “Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures,” J. Funct. Anal., vol. 262, pp. 3309–3339, 2012. [Online]. Available: http://arxiv.org/abs/1109.5287
[8] S. G. Bobkov and G. P. Chistyakov, “Bounds on the maximum of the density for sums of independent random variables (russian),” Zapiski Nauchn. Semin. POMI, 2012.
[9] ——, “Entropy power inequality for the Rényi entropy,” Preprint, 2013.
[10] ——, “On concentration functions of random variables,” To appear in JOTP, 2013.
[11] S. G. Bobkov, G. P. Chistyakov, and F. Götze, “Fisher information and the central limit theorem,” Probab. Theory Relat. Fields, vol. 159, pp. 1–59, June 2014.
[12] S. G. Bobkov, M. Madiman, and L. Wang, “Fractional generalizations of Young and Brunn-Minkowski inequalities,” in Concentration, Functional Inequalities and Isoperimetry, ser. Contemp. Math., C. Houdré, M. Ledoux, E. Milman, and M. Milman, Eds., vol. 545. Amer. Math. Soc., 2011, pp. 35–53.
[13] S. G. Bobkov and M. M. Madiman, “On the problem of reversibility of the entropy power inequality,” in Limit Theorems in Probability, Statistics, and Number Theory (in honor of Friedrich Götze), ser. Springer Proceedings in Mathematics and Statistics, P. E. et al., Ed. Springer-Verlag, 2013, vol. 42, available online at http://arxiv.org/abs/1111.6807.
[14] H. J. Brascamp and E. H. Lieb, “Best constants in Young’s inequality, its converse, and its generalization to more than three functions,” Advances in Math., vol. 20, no. 2, pp. 151–173, 1976.
[15] H. J. Brascamp, E. H. Lieb, and J. M. Luttinger, “A general rearrangement inequality for multiple integrals,” J. Functional Analysis, vol. 17, pp. 227–237, 1974.
[16] A. Burchard, “Cases of equality in the Riesz rearrangement inequality,” Ph.D. dissertation, Georgia Institute of Technology, Atlanta, USA, 1994.
[17] —— (2009, June) A short course on rearrangement inequalities. [Online]. Available: http://www.math.utoronto.ca/almot/rearrange.pdf
[18] A. Burchard and M. Schmuckenschläger, “Comparison theorems for exit times,” Geom. Funct. Anal., vol. 11, no. 4, pp. 651–692, 2001. [Online]. Available: http://dx.doi.org/10.1007/PL00001681
[19] E. A. Carlen and A. Soffer, “Entropy production by block variable summation and central limit theorems.” Comm. Math. Phys., vol. 140, 1991.
[20] K. M. Chong, “Some extensions of a theorem of Hardy, Littlewood and Pólya and their applications,” Canad. J. Math., vol. 26, pp. 1321–1340, 1974.
[21] J. Costa, A. Hero, and C. Vignat, “On solutions to multivariate maximum alpha-entropy problems,” Lecture Notes in Computer Science, vol. 2683, no. EMMCVPR 2003, Lisbon, 7–9 July 2003, pp. 211–228, 2003.
[22] M. Costa, “A new entropy power inequality,” IEEE Trans. Inform. Theory, vol. 31, no. 6, pp. 751–760, 1985.
[23] T. M. Cover and J. A. Thomas, Elements of information theory, 2nd ed. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2006.
[24] A. Dembo, T. Cover, and J. Thomas, “Information-theoretic inequalities,” IEEE Trans. Inform. Theory, vol. 37, no. 6, pp. 1501–1518, 1991.
[25] A. Dembo and O. Zeitouni, Large Deviations Techniques And Applications, 2nd ed. New York: Springer-Verlag, 1998.
[26] A. Drewitz, P. Sousi, and R. Sun, “Symmetric rearrangements around infinity with applications to Lévy processes,” Probab. Theory Relat. Fields, 2013.
[27] L. Gross, “Logarithmic Sobolev inequalities,” Amer. J. Math., vol. 97, no. 4, pp. 1061–1083, 1975.
[28] B. Hajek, K. Mitzel, and S. Yang, “Paging and registration in cellular networks: jointly optimal policies and an iterative algorithm,” IEEE Trans. Inform. Theory, vol. 54, no. 2, pp. 608–622, 2008. [Online]. Available: http://dx.doi.org/10.1010/JIT/2007.913566
[29] G. H. Hardy, J. E. Littlewood, and G. Pólya, Some simple inequalities satisfied by convex functions,” Messenger of Mathematics, vol. 58, pp. 145–152, 1929.
[30] G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, 2nd ed., ser. Cambridge Mathematical Library. Cambridge: Cambridge University Press, 1988, reprint of the 1952 edition.
[31] P. Harremoës and C. Vignat, “A short information theoretic proof of CLT,” Unpublished, 2005.
[32] O. Johnson and C. Vignat, “Some results concerning maximum Rényi entropy distributions,” Ann. Inst. H. Poincaré Probab. Statist., vol. 43, no. 3, pp. 339–351, 2007.
[33] O. Johnson, Information theory and the central limit theorem. London: Imperial College Press, 2004.
[34] E. H. Lieb and M. Loss, Analysis, 2nd ed., ser. Graduate Studies in Mathematics. Providence, RI: American Mathematical Society, 2001, vol. 14.
[35] E. Lutwak, S. Lv, D. Yang, and G. Zhang, “Affine moments of a random vector,” IEEE Trans. Inform. Theory, vol. 59, no. 9, pp. 5592–5599, September 2013.
[36] E. Lutwak, D. Yang, and G. Zhang, “Moment-entropy inequalities for a random vector,” IEEE Trans. Inform. Theory, vol. 53, no. 4, pp. 1603–1607, 2007.
[37] M. Madiman and I. Kontoyiannis, “The Ruzsa divergence for random elements in locally compact abelian groups,” Preprint, 2014.
