A New Approach to the Entropy Power Inequality, via Rearrangements

Liyao Wang
Department of Physics
Yale University
New Haven CT, USA
Email: liyao.wang@yale.edu

Mokshay Madiman
Department of Mathematical Sciences
University of Delaware
Newark DE, USA
Email: mokshay.madiman@yale.edu

***Abstract*—A new lower bound on the entropy of the sum of independent random vectors is demonstrated in terms of rearrangements. This lower bound is better than that given by the entropy power inequality. In fact, we use it to give a new, independent, and simple proof of the entropy power inequality in the case when the summands are identically distributed. We also give a more involved but new way to recover the full entropy power inequality, without invoking Fisher information, MMSE or any differentiation of information functionals.**

I. Introduction

The entropy power inequality (EPI) is a basic and powerful tool in information theory, and also has relevance to probability theory and mathematical physics. Firstly, it has been used very effectively in proofs of converse results for various channel coding theorems; indeed this was Shannon's original motivation for introducing it in [1]. Another famous application of it to Gaussian broadcast channels was developed by Bergmans [2], and this was significantly generalized in an important (and much more recent) paper of Weingarten, Steinberg and Shamai [3] to the Gaussian MIMO setting. Secondly, the EPI plays a key role in the information-theoretic understanding of probabilistic limit theorems; not only does it imply that entropy increases monotonically along a subsequence for the normalized sums in the central limit theorem, but the search for a discrete analogue of it has motivated much interesting recent research (see [4], [5] and references therein). Thirdly, the EPI is closely related to important results in mathematical physics– it can be used to deduce the Fourier transform formulation of Heisenberg's uncertainty principle, as well as the Gaussian logarithmic Sobolev inequality (both these deductions are in Stam's beautiful paper [6]).

Numerous variants and generalizations of the EPI have been proved, discussed, and applied to various problems in the literature. However, to our knowledge, there are no known direct refinements of the EPI in the sense that a natural information-theoretic quantity is inserted between the left and right sides of the classical formulations of the inequality. Our first contribution is to provide such a refinement. The refinement is stated in terms of the notion of the spherically symmetric decreasing rearrangement of a random vector (or its probability density function); we defer the definition of this notion to Section II.

Theorem 1. *Let $f_i, i = 1, 2, \cdots, n$ be n probability densities on $\mathbb{R}^d$ and $f_i^*, i = 1, 2, \cdots, n$ be the spherically symmetric decreasing rearrangements of the corresponding densities. Then*

$$h(f_1 \star f_2 \star \cdots \star f_n) \geq h(f_1^* \star f_2^* \star \cdots \star f_n^*), \tag{1}$$

as long as both sides are well defined. Here

$$h(f) = -\int_{\mathbb{R}^d} f(x) \log(f(x)) dx,$$

is the differential entropy of a probability density f, and $\star$ is ordinary convolution for densities on $\mathbb{R}^d$.

Indeed, the classical EPI can be used to show that if g_i are independent isotropic (i.e., covariance being a multiple of identity) Gaussians with $h(f_i) = h(g_i)$, then

$$h(f_1^* \star f_2^* \star \cdots \star f_n^*) \geq h(g_1 \star g_2 \star \cdots \star g_n).$$

Since the classical EPI can be formulated as saying that

$$h(f_1 \star f_2 \star \cdots \star f_n) \geq h(g_1 \star g_2 \star \cdots \star g_n), \tag{2}$$

Theorem 1 is saying that $h(f_1^* \star f_2^* \star \cdots \star f_n^*)$ can be inserted between the two sides of (2). The details of this argument are given in Section III, but for now we simply note that this argument does *not* immediately give a new proof of the EPI, since the EPI itself was used to show that (1) is a refinement of (2).

The proof of Theorem 1 is independent of the EPI or any Fisher information, MMSE or entropy differentiation arguments, and relies on rearrangement inequalities developed by Brascamp, Lieb and Luttinger [7]. The details of the proof of Theorem 1, as well as various applications, will be contained in the forthcoming paper [8] by the authors. In this note, we focus on one particular application– providing a new proof of the EPI. This is particularly easy to do in the special case when the summands are identically distributed (which is the relevant case for the application to the plain vanilla entropic CLT) and we explain this in Section IV; the deduction of the general case from Theorem 1 is more involved and presented in Section V. Section VI contains some concluding remarks.

II. Brief introduction to spherically symmetric decreasing rearrangement

For a Borel set A with volume $|A|$, one can define its spherically 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 $A^* = \emptyset$.

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

$$f^*(y) = \int_0^{+\infty} I_{B_t^*}(y)dt,$$

where $B_t = \{x : f(x) > t\}$ and $I_{B_t^*}(y)$ is the indicator function of B_t^*. It follows easily from definition that $f^*(x)$ is spherically symmetric and a decreasing function of the Euclidean norm of x.

The following lemma contains some simple properties of f^*:

Lemma 2. 1) *For* $1 \le p < +\infty$*,*

$$\|f\|_p = \|f^*\|_p.$$

In particular, if f is a probability density, so is f^*.

2) *If f is bounded then so is f^*; if f is strictly positive, then so is f^*.*
3) *If f is a probability density function, then*

$$\int \|x\|^2 f^*(x)dx \le \int \|x\|^2 f(x)dx,$$

where $\|x\|$ is the Euclidean norm.

The first two statements in the above lemma are classical [9] and the last one is probably known in the literature, although we are not able to find a reference. We also need the following observation, which seems to be new but follows easily from definition.

Lemma 3. *If one of $h(f)$ and $h(f^*)$ is well defined, then so is the other one and we have:*

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

III. Theorem 1 as a strengthening of EPI

We first point out that Theorem 1 can be viewed as a strengthening of the classical EPI. Suppose that f_1 and f_2 are two densities. Then by Theorem 1, for any fixed $0 < \lambda < 1$, we have:

$$h(\sqrt{\lambda}X_1 + \sqrt{1-\lambda}X_2) \ge h(\sqrt{\lambda}X_1^* + \sqrt{1-\lambda}X_2^*), \qquad (3)$$

where X_i is distributed according to f_i, X_i^* is distributed according to f_i^* for $i = 1,2$ and all random vectors are independent.

In the literature, the EPI is typically stated under the assumption that both random vectors have finite covariance matrices. By Lemma 2 (3), we can apply EPI to X_1^* and X_2^*:

$$h(\sqrt{\lambda}X_1^* + \sqrt{1-\lambda}X_2^*) \ge \lambda h(X_1^*) + (1-\lambda)h(X_2^*). \qquad (4)$$

Now from Lemma 3, we have:

$$\lambda h(X_1^*) + (1-\lambda)h(X_2^*) = \lambda h(X_1) + (1-\lambda)h(X_2), \qquad (5)$$

From inequalities (3), (4) and (5) , we see that

$$h(\sqrt{\lambda}X_1 + \sqrt{1-\lambda}X_2) \ge \lambda h(X_1) + (1-\lambda)h(X_2),$$

which is an equivalent form of the EPI applied to X_1 and X_2.

Hence $h(\sqrt{\lambda}X_1^* + \sqrt{1-\lambda}X_2^*)$, given by Theorem 1, is a tighter lower bound of $h(\sqrt{\lambda}X_1 + \sqrt{1-\lambda}X_2)$ than $\lambda h(X_1) + (1-\lambda)h(X_2)$, given by the EPI applied to X_1 and X_2 directly. In this sense, Theorem 1 is a strengthening of the EPI.

IV. Deducing EPI from Theorem 1 assuming identical distribution

An EPI for symmetric densities 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 λ. We first work under the assumption that f_1 and f_2 are bounded and strictly positive, with finite covariance matrices. At the end of this section, we indicate how to relax this assumption to obtain the EPI under the assumption that the distribution has a finite covariance matrix.

By Lemma 2, it suffices to show the following.

Proposition 4. *Fix any $0 < \lambda < 1$. Suppose that Y_1 and Y_2 are two independent random vectors distributed according to a spherically symmetric, bounded and strictly positive density with a finite covariance matrix. Then:*

$$h(\sqrt{\lambda}Y_1 + \sqrt{1-\lambda}Y_2) \ge h(Y_1).$$

Proof: Clearly, both sides are finite under our assumptions. By independence, we have:

$$h(Y_1, Y_2) = h(Y_1) + h(Y_2) = 2h(Y_1).$$

By the scaling property for entropy,

$$h(Y_1, Y_2) = h(\sqrt{\lambda}Y_1 + \sqrt{1-\lambda}Y_2, \sqrt{1-\lambda}Y_1 - \sqrt{\lambda}Y_2).$$

Now we can use subadditivity of entropy to obtain:

$$h(Y_1, Y_2) \le h(\sqrt{\lambda}Y_1 + \sqrt{1-\lambda}Y_2) + h(\sqrt{1-\lambda}Y_1 - \sqrt{\lambda}Y_2)$$

$$= 2h(\sqrt{\lambda}Y_1 + \sqrt{1-\lambda}Y_2),$$

where the last equality follows from spherical symmetry (in fact, we only need central symmetry) and the i.i.d. assumption. ■

The inequality (3), Proposition 4 and Lemma 3 imply the EPI for bounded and strictly positive density if we assume identical distribution. Note that it is known [10] if X has a finite covariance matrix, then $h(X + \sqrt{t}Z)$ is continuous at

$t = 0$, where Z is a standard Gaussian random vector on $\mathbb{R}^d$, independent of X. But $X + \sqrt{t}Z$ has a bounded and strictly positive density. Hence we can relax our assumption at the beginning by a limiting argument. We conclude that the assumption of a finite covariance matrix is enough.

V. The full EPI via rearrangement

In this section, we will assume $d = 1$ for convenience. Let X and Y be two independent random variables. Similarly to the previous section, to prove the full EPI, we can assume that X and Y are both symmetric random variables distributed according to bounded, decreasing densities with finite second moments.

Our proof is inspired by and essentially an adaption of Brascamp and Lieb's proof of Young's inequality with sharp constant [11]. The trick is to use tensorization, or what physicists call the replica method. Consider $X_1, X_2, \cdots, X_M$, which are M independent copies of X, and independent of these, $Y_1, Y_2, \cdots, Y_M$, which are M independent copies of Y. Then by independence, we have:

$$\begin{aligned}&h(\sqrt{\lambda}X_1 + \sqrt{1-\lambda}Y_1, \sqrt{\lambda}X_2 + \sqrt{1-\lambda}Y_2, \cdots,\\ &\qquad \sqrt{\lambda}X_M + \sqrt{1-\lambda}Y_M) = Mh(\sqrt{\lambda}X_1 + \sqrt{\lambda}Y_1).\end{aligned} \tag{6}$$

Now we apply the M dimensional version of Theorem 1:

$$\begin{aligned}&h(\sqrt{\lambda}X_1 + \sqrt{1-\lambda}Y_1, \cdots, \sqrt{\lambda}X_M + \sqrt{1-\lambda}Y_M)\\ &= h(\sqrt{\lambda}\mathbf{X} + \sqrt{1-\lambda}\mathbf{Y})\\ &\geq h(\sqrt{\lambda}\mathbf{X}^* + \sqrt{1-\lambda}\mathbf{Y}^*),\end{aligned} \tag{7}$$

where $\mathbf{X} = (X_1, X_2, \cdots, X_M)$, $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_M)$ and all random variables are independent.

Since we assumed f and g are bounded symmetric decreasing, we can approximate these densities pointwisely and monotonically from below by symmetric decreasing simple functions of the form f_n and g_n:

$$f_n = \sum_{i=1}^{k_n} c_i^n I_i^n,$$

where I_i^n are indicators of symmetric finite intervals with $I_i^n \leq I_{i+1}^n$ 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 (details omitted) that for fixed $0 < \lambda < 1$,

$$\begin{gathered}h(\widetilde{f_n}) \to h(f),\\ h(\widetilde{g}_n) \to h(g),\\ h\left(\frac{1}{\sqrt{\lambda}}\widetilde{f}_n\left(\frac{\cdot}{\sqrt{\lambda}}\right) \star \frac{1}{\sqrt{1-\lambda}}\widetilde{g}_n\left(\frac{\cdot}{\sqrt{1-\lambda}}\right)\right) \to\\ h\left(\frac{1}{\sqrt{\lambda}}f\left(\frac{\cdot}{\sqrt{\lambda}}\right) \star \frac{1}{\sqrt{1-\lambda}}g\left(\frac{\cdot}{\sqrt{1-\lambda}}\right)\right),\end{gathered}$$

where $\widetilde{f}_n$ and $\widetilde{g}_n$ are normalized versions of f_n and g_n. Hence, without loss of generality, we can assume that f and g are of the following form:

$$f = \sum_{i=1}^{k_1} c_i^1 I_i^1, \quad g = \sum_{i=1}^{k_2} c_i^2 I_i^2.$$

where $c_i^1 > 0, c_i^2 > 0$. Let

$$F(x_1, x_2, \cdots, x_M) = \prod_{i=1}^{M} f(x_i),$$

$$G(y_1, y_2, \cdots, y_M) = \prod_{i=1}^{M} g(y_i)$$

be the densities of $\mathbf{X}$ and $\mathbf{Y}$. Then it's easy to see F takes at most $(M+1)^{k_1}$ values and G takes at most $(M+1)^{k_2}$ values. Hence just by looking at the definitions of rearrangements, one sees that (we omit the elementary detail), 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 spherically symmetric decreasing property,

$$F^* = \sum_{i=1}^{(M+1)^{k_1}} b_i^1 I_{\eta_i^1}, \quad G^* = \sum_{j=1}^{(M+1)^{k_2}} b_j^2 I_{\eta_j^2},$$

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

$$\sum_{i=1}^{(M+1)^{k_1}} b_i^1 |\eta_i^1| = 1, \sum_{j=1}^{(M+1)^{k_2}} b_j^2 |\eta_j^2| = 1,$$

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

By concavity of entropy as a functional of density, we get:

$$\begin{aligned}&h(\sqrt{\lambda}\mathbf{X}^* + \sqrt{1-\lambda}\mathbf{Y}^*)\\ &\geq \sum_{i,j} b_i^1 b_j^2 |\eta_i^1||\eta_j^2| h(\sqrt{\lambda}\mathbf{Z_i^1} + \sqrt{1-\lambda}\mathbf{Z_j^2}),\end{aligned} \tag{8}$$

where $\mathbf{Z_i^1}$ and $\mathbf{Z_j^2}$ are independent uniform distributions on the M dimensional balls η_i^1 and η_j^2 respectively. We now use the following simple lemma:

Lemma 5. *Let f be a mixture of densities:*

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

Then

$$\begin{aligned}h(f) &\leq \sum_i c_i h(f_i) - \sum_i c_i \log c_i\\ &\leq \sum_i c_i h(f_i) + \log n.\end{aligned}$$

Hence we have, using Lemma 5 and Lemma 3:

$$Mh(X) = h(\mathbf{X}) = h(\mathbf{X}^*)$$

$$\leq k_1 \log(M+1) + \sum_{i=1}^{(M+1)^{k_1}} b_i^1 |\eta_i^1| h(\mathbf{Z_i^1}),$$

$$Mh(Y) = h(\mathbf{Y}) = h(\mathbf{Y}^*)$$

$$\leq k_2 \log(M+1) + \sum_{j=1}^{(M+1)^{k_2}} b_j^2 |\eta_j^2| h(\mathbf{Z_j^2}). \quad (9)$$

Now let us temporarily assume that the EPI is true when both X and Y are uniform distributions on balls centered at the origin. Under this assumption,

$$h(\sqrt{\lambda}\mathbf{Z_i^1} + \sqrt{1-\lambda}\mathbf{Z_j^2}) \geq \lambda h(\mathbf{Z_i^1}) + (1-\lambda) h(\mathbf{Z_j^2}). \quad (10)$$

Then if we combine (6) to (10), we get:

$$\begin{aligned} & h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \\ & \geq -\lambda k_1 \frac{\log(M+1)}{M} - (1-\lambda) k_2 \frac{\log(M+1)}{M} \\ & + \lambda h(X) + (1-\lambda) h(Y). \end{aligned}$$

Taking a limit as M goes to infinity, we get full EPI.

This is already interesting. Because, assuming the validity of Theorem 1 when $n = 2$ (but for any d), some simple reductions reduce our task to proving EPI ONLY for uniform distributions on balls! In principle, this is a calculus problem and an explicit expression for the entropy of the sum of two independent uniforms on balls will be given in the full paper. But, to show the full entropy power, we actually only need the following lemma:

Lemma 6. *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 (details in the full paper), using Stirling's approximation, one can show (assuming M even,w.l.o.g):

$$\begin{aligned} f_i &\leq \sqrt{\pi M} e^{\frac{1}{12M}} g_i, \\ f &\leq \pi M e^{\frac{1}{6M}} g. \end{aligned}$$

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's bounded by $\log(M)$ times a universal constant. Now some easy calculations will show:

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

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

Hence we get:

$$\begin{aligned} & h(Z) - \lambda h(Z_1) - (1-\lambda) h(Z_2) \\ & = \frac{M}{2}\Bigg(\log(\lambda b_1^2 + (1-\lambda) b_2^2) \\ & - \lambda \log(b_1^2) - (1-\lambda)\log(b_2^2) \Bigg) + O(\log(M)) \quad (11) \\ & \geq O(\log(M)), \end{aligned}$$

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

The above proof gives very strong information about the entropy of the sum of two independent uniforms on balls, in high dimensions. Equality (11) is in fact equivalent to:

$$e^{\frac{2h(Z_1+Z_2)}{M}} = e^{\frac{O(\log(M))}{M}} \left(e^{\frac{2h(Z_1)}{M}} + e^{\frac{2h(Z_2)}{M}}\right), \quad (12)$$

where the meaning of the O symbol is as before. This is not surprising because uniform distributions on high-dimensional balls are close to Gaussians.

Another implication is the following. Suppose we consider two independent uniforms X and Y on Borel sets in $\mathbb{R}^M$ with finite volumes. Then Theorem 1 gives:

$$h(X+Y) \geq h(X^* + Y^*).$$

But from (12) and Lemma 3, we see that:

$$\frac{2h(X^*+Y^*)}{M} = \log(e^{\frac{2h(X)}{M}} + e^{\frac{2h(Y)}{M}}) + \frac{O(\log(M))}{M}.$$

Hence, although our Theorem 1 is a strengthening of EPI, in high dimensions, we do not gain much.

However, for a fixed dimension, our bound can give significant improvements over EPI. In fact, a simple calculation shows that when $d = 1$ and X, Y are uniforms on Borel sets A and B, Theorem 1 becomes:

$$h(X+Y) \geq \log(|A| e^{\frac{|B|}{2|A|}}) = \log(|A|) + \frac{|B|}{2|A|},$$

assuming $|A| \geq |B|$. In contrast, the EPI will give:

$$h(X+Y) \geq \log(|A|) + \frac{1}{2}\log\left(1 + \left(\frac{|B|}{|A|}\right)^2\right).$$

Our bound is more precise than that given by EPI especially when $\frac{|B|}{|A|}$ is small.

Remark 7. There have been many proofs of the EPI. The ones that are information-theoretic in nature include the early proof by Stam [6] using Fisher information and interpolation (with simplifications by Blachman [12] and Dembo, Cover and Thomas [13]), Verdú and Guo's proof [14] using a similar interpolation but with MMSE instead of Fisher information, and Rioul's proof [15] sidestepping the explicit interpolation arguments common to the previously mentioned proofs. There have also been more analytical proofs, such as Lieb's proof [16] starting from Young's convolution inequality with sharp constant (which in turn has several proofs), Szarek and Voiculescu's proof [17] based on a restricted Brunn-Minkowski inequality, and Lehec's proof [18] based on a Brownian motion representation of relative entropy to Gaussianity. We have given yet another proof of the EPI here, starting with Theorem 1, which reduces the task of proving the EPI for general independent summands to proving the EPI for summands with spherically symmetric, decreasing densities. Our proof is in some sense related to the Szarek–Voiculescu proof since the restricted Brunn-Minkowski inequality they use as their main stepping stone to the EPI is actually proved using rearrangements, but that inequality does not have a simple information-theoretic statement like Theorem 1 does.

Remark 8. Although Young's inequality with sharp constant implies the EPI, the early proofs of Young's inequality with sharp constant [19], [11], [20] are very different from direct proofs of the EPI. Even in the recent work by Cordero-Erausquin and Ledoux [21], where they give a proof of a special case of the EPI by similar entropy duality methods to their proof of Young's inequality, there are difficulties in treating the EPI as discussed by them. Our proof provides a unification of sorts in that both our proof and the proof by Brascamp and Lieb [11] of Young's inequality have two key ingredients (the replica trick and a rearrangement inequality).

VI. CONCLUSION

We indicate some extensions and possible applications of Theorem 1, the details of which will be contained in the full paper.

First, it turns out that Theorem 1 has a natural extension to Rényi entropies of all orders. Specifically, we have:

Theorem 9. *Let $f_i, i = 1, 2, \cdots, n$ be n probability densities on $\mathbb{R}^d$. Then for $p \in [0, +\infty]$,*

$$h_p(f_1 \star f_2 \star \cdots \star f_n) \geq h_p(f_1^* \star f_2^* \star \cdots \star f_n^*),$$

as long as both sides are well defined. Here

$$h_p(f) = \frac{1}{1-p} \log\left(\int f^p(x)dx\right)$$

for $p \in (0, 1) \cup (1, +\infty)$, $h_1(f) = h(f)$ and $h_0(f), h_\infty(f)$ are defined in a limiting sense.

We will give a unified proof for all p in the full paper, using some ideas from majorization theory.

Second, some other functional inequalities can be obtained as corollaries to our results. For example, the Pólya-Szegö inequality [9] can be obtained as a corollary to Theorem 1 using standard information theoretical tools. As another example, the Brunn-Minkowski inequality [9] is implied by the $p = 0$ case of Theorem 9.

Finally, we point out that our results might be useful in reducing some optimization problems involving entropy of sums to the spherically symmetric decreasing case, for which the symmetry can be used.

ACKNOWLEDGMENT

The authors would like to thank the U.S. National Science Foundation for supporting their research through the CAREER grant DMS-1056996 and CCF-1065494, and the Department of Mathematics at the Indian Institute of Science, Bangalore, for hosting them when this research was completed.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[2] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Information Theory*, vol. IT-20, pp. 279–280, 1974.

[3] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, 2006. [Online]. Available: http://dx.doi.org/10.1109/TIT.2006.880064

[4] S. Haghighatshoar, E. Abbe, and E. Telatar, "A new entropy power inequality for integer-valued random variables," *Preprint,* `arXiv:1301:4185`, 2013.

[5] V. Jog and V. Anantharam, "The entropy power inequality and mrs gerber's lemma for abelian groups of order 2^n," *Preprint,* `arXiv:1207:6335`, 2012.

[6] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, pp. 101–112, 1959.

[7] 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.

[8] L. Wang and M. Madiman, "Beyond the entropy power inequality, via rearrangements," *Preprint*, 2013.

[9] A. Burchard. (2009, June) A short course on rearrangement inequalities. [Online]. Available: http://www.math.utoronto.ca/almut/rearrange.pdf

[10] E. A. Carlen and A. Soffer, "Entropy production by block variable summation and central limit theorems." *Comm. Math. Phys.*, vol. 140, 1991.

[11] 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.

[12] N. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Information Theory*, vol. IT-11, pp. 267–271, 1965.

[13] A. Dembo, T. Cover, and J. Thomas, "Information-theoretic inequalities," *IEEE Trans. Inform. Theory*, vol. 37, no. 6, pp. 1501–1518, 1991.

[14] S. Verdú and D. Guo, "A simple proof of the entropy-power inequality," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2165–6, May 2006.

[15] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 33–55, 2011. [Online]. Available: http://dx.doi.org/10.1109/TIT.2010.2090193

[16] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," *Comm. Math. Phys.*, vol. 62, no. 1, pp. 35–41, 1978.

[17] S. J. Szarek and D. Voiculescu, "Shannon's entropy power inequality via restricted Minkowski sums," in *Geometric aspects of functional analysis*, ser. Lecture Notes in Math. Berlin: Springer, 2000, vol. 1745, pp. 257–262. [Online]. Available: http://dx.doi.org/10.1007/BFb0107219

[18] J. Lehec, "Representation formula for the entropy and functional inequalities," *Preprint,* `arXiv:1006:3028`, 2011.

[19] W. Beckner, "Inequalities in Fourier analysis," *Ann. of Math. (2)*, vol. 102, no. 1, pp. 159–182, 1975.

[20] F. Barthe, "Optimal Young's inequality and its converse: a simple proof," *Geom. Funct. Anal.*, vol. 8, no. 2, pp. 234–242, 1998.

[21] D. Cordero-Erausquin and M. Ledoux, "The geometry of Euclidean convolution inequalities and entropy," *Proc. Amer. Math. Soc.*, vol. 138, no. 8, pp. 2755–2769, 2010.