STUDIA MATHEMATICA 235 (1) (2016)

A reverse entropy power inequality for log-concave random vectors

by

KEITH BALL (Warwick), PIOTR NAYAR (Minneapolis, MN)
and TOMASZ TKOCZ (Warwick)

Abstract. We prove that the exponent of the entropy of one-dimensional projections of a log-concave random vector defines a 1/5-seminorm. We make two conjectures concerning reverse entropy power inequalities in the log-concave setting and discuss some examples.

1. Introduction. One of the most significant and mathematically intriguing quantities studied in information theory is the *entropy*. For a random variable X with density f its entropy is defined as

$$\mathcal{S}(X) = \mathcal{S}(f) = -\int_{\mathbb{R}} f \ln f \tag{1}$$

provided this integral exists (in the Lebesgue sense). Note that the entropy is translation invariant and $\mathcal{S}(bX) = \mathcal{S}(X) + \ln|b|$ for any nonzero b. If f belongs to $L_p(\mathbb{R})$ for some $p > 1$, then by the concavity of the logarithm and by Jensen's inequality, $\mathcal{S}(f) > -\infty$. If $\mathbb{E}X^2 < \infty$, then comparison with the standard Gaussian density and again Jensen's inequality yield $\mathcal{S}(X) < \infty$. In particular, the entropy of a log-concave random variable is well defined and finite. Recall that a random vector in $\mathbb{R}^n$ is called *log-concave* if it has a density of the form $e^{-\psi}$ with $\psi : \mathbb{R}^n \to (-\infty, \infty]$ being a convex function.

The *entropy power inequality* (EPI) says that

$$e^{\frac{2}{n}\mathcal{S}(X+Y)} \geq e^{\frac{2}{n}\mathcal{S}(X)} + e^{\frac{2}{n}\mathcal{S}(Y)} \tag{2}$$

for independent random vectors X and Y in $\mathbb{R}^n$ provided that all the entropies exist. Stated first by Shannon in his seminal paper [24] and first

2010 *Mathematics Subject Classification*: Primary 94A17; Secondary 52A40, 60E15.
Key words and phrases: entropy, log-concave, reverse entropy power inequality.
Received 19 October 2015; revised 18 August 2016.
Published online 23 September 2016.

DOI: 10.4064/sm8418-6-2016

rigorously proved by Stam in [25] (see also [6]), it is often referred to as the *Shannon–Stam inequality* and plays a crucial role in information theory and elsewhere (see the survey [18]). Using the AM-GM inequality, the EPI can be *linearised*: for every $\lambda \in [0,1]$ and independent random vectors X, Y we have

$$\mathcal{S}(\sqrt{\lambda}\,X + \sqrt{1-\lambda}\,Y) \geq \lambda \mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y) \tag{3}$$

provided that all the entropies exist. This formulation is in fact equivalent to (2) as first observed by Lieb in [22], where he also shows how to derive (3) from Young's inequality with sharp constants. Several other proofs of (3) are available, including versions for the Fisher information [13] and recent techniques of the minimum mean-square error [27]. There are also refinements when one variable is Gaussian [15, 17, 28].

If X and Y are independent and identically distributed random variables (or vectors), inequality (3) says that the entropy of the normalised sum

$$X_\lambda = \sqrt{\lambda}\,X + \sqrt{1-\lambda}\,Y \tag{4}$$

is at least as large as the entropy of the summands X and Y, i.e. $\mathcal{S}(X_\lambda) \geq \mathcal{S}(X)$. It is worth mentioning that this phenomenon has been quantified, first in [14], which has deep consequences in probability (see the pioneering work [4] and its sequels [1, 2], which establish the rate of convergence in the entropic central limit theorem and the "second law of probability" of the entropy growth, as well as the independent work [20], with somewhat different methods). In the context of log-concave vectors, Ball and Nguyen [5] establish dimension free lower bounds on $\mathcal{S}(X_{1/2}) - \mathcal{S}(X)$ and discuss connections between the entropy and major conjectures in convex geometry; for the latter see also [12].

In general, the EPI cannot be reversed. In [7, Proposition V.8], Bobkov and Chistyakov find a random vector X with finite entropy such that $\mathcal{S}(X+Y) = \infty$ for every random vector Y independent of X and with finite entropy. However, for log-concave vectors and, more generally, convex measures, Bobkov and Madiman have recently addressed the question of reversing the EPI (see [10, 11]). They show that for any pair X, Y of independent log-concave random vectors in $\mathbb{R}^n$, there are linear volume preserving maps $T_1, T_2 : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$e^{\frac{2}{n}\mathcal{S}(T_1(X)+T_2(Y))} \leq C\big(e^{\frac{2}{n}\mathcal{S}(X)} + e^{\frac{2}{n}\mathcal{S}(Y)}\big),$$

where C is some universal constant.

For a random variable X with finite variance its *relative entropy* $\mathcal{D}(X)$ is defined as the difference $\mathcal{S}(Z) - \mathcal{S}(X)$, where Z is a Gaussian random variable with variance $\mathrm{Var}(X)$. Relative entropy is nonnegative and provides a way to measure closeness to Gaussians. Another reverse EPI has been lately discovered in the context of the stability of Cramér's theorem (see [9]). The authors bound from below the relative entropy of the sum of independent *regularised* random variables in terms of the relative entropies of the regularised summands. The regularisation is performed by adding independent Gaussians; without it, such a lower bound does not hold in general, as shown by the same authors in [8].

The goal of this note is to investigate further, in the log-concave setting, some new forms of what could be called a reverse EPI. In the next section we present our results. The last section is devoted to their proofs.

Acknowledgements. The authors would like to thank Assaf Naor for pointing out the Aoki–Rolewicz theorem as well as for fruitful discussions without which Theorem 1 would not have been discovered. They are also indebted to Mokshay Madiman for his help with tracking down several references.

PN was supported in part by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation as well as by NCN grant DEC-2012/05/B/ST1/00412.

2. Main results and conjectures. Suppose X is a symmetric log-concave random vector in $\mathbb{R}^n$. Then any projection of X on a certain direction $v \in \mathbb{R}^n$, that is, the random variable $\langle X, v\rangle$ is also log-concave. Here $\langle \cdot, \cdot \rangle$ denotes the standard scalar product in $\mathbb{R}^n$. If we know the entropies of projections in, say, two different directions, can we say anything about the entropy of projections in related directions? We make the following conjecture.

Conjecture 1. *Let X be a symmetric log-concave random vector in $\mathbb{R}^n$. Then the function*

$$N_X(v) = \begin{cases} e^{\mathcal{S}(\langle v, X\rangle)}, & v \neq 0, \\ 0, & v = 0, \end{cases}$$

defines a norm on $\mathbb{R}^n$.

The homogeneity of N_X is clear. To check the triangle inequality, we have to answer in fact a two-dimensional question: *is it true that for a symmetric log-concave random vector (X, Y) in $\mathbb{R}^2$ we have*

$$e^{\mathcal{S}(X+Y)} \le e^{\mathcal{S}(X)} + e^{\mathcal{S}(Y)}? \tag{5}$$

Indeed, this applied to the vector $(\langle u, X\rangle, \langle v, X\rangle)$ which is also log-concave yields $N_X(u+v) \le N_X(u) + N_X(v)$. Inequality (5) can be seen as a reverse EPI (cf. (2)). It is not too difficult to show that this inequality holds up to a multiplicative constant.

Proposition 1. *Let (X,Y) be a symmetric log-concave random vector in $\mathbb{R}^2$. Then*

$$e^{\mathcal{S}(X+Y)} \le e(e^{\mathcal{S}(X)} + e^{\mathcal{S}(Y)}).$$

Proof. The argument relies on the well-known observation that for a log-concave density $f\colon \mathbb{R} \to [0,\infty)$ its maximum and entropy are related (see for example [5] or [12]):

$$-\ln\|f\|_\infty \le \mathcal{S}(f) \le 1 - \ln\|f\|_\infty. \tag{6}$$

Suppose that w is an even log-concave density of (X,Y). The densities of X, Y and $X+Y$ equal respectively

$$f(x) = \int w(x,t)\,dt, \qquad g(x) = \int w(t,x)\,dt, \qquad h(x) = \int w(x-t,t)\,dt. \tag{7}$$

They are even and log-concave, hence attain their maximum at zero. By the result of Ball (Busemann's theorem for symmetric log-concave measures, see [3]), the function $\|x\|_w = (\int w(tx)\,dt)^{-1}$ is a norm on $\mathbb{R}^2$. In particular,

$$\begin{aligned}\frac{1}{\|h\|_\infty} &= \frac{1}{h(0)} = \frac{1}{\int w(-t,t)\,dt} = \|e_2 - e_1\|_w \le \|e_1\|_w + \|e_2\|_w \\ &= \frac{1}{\int w(t,0)\,dt} + \frac{1}{\int w(0,t)\,dt} = \frac{1}{f(0)} + \frac{1}{g(0)} = \frac{1}{\|f\|_\infty} + \frac{1}{\|g\|_\infty}.\end{aligned}$$

Using (6) twice we obtain

$$e^{\mathcal{S}(X+Y)} \le \frac{e}{\|h\|_\infty} \le e\left(\frac{1}{\|f\|_\infty} + \frac{1}{\|g\|_\infty}\right) \le e(e^{\mathcal{S}(X)} + e^{\mathcal{S}(Y)}). \ \blacksquare$$

Recall that the classical result of Aoki and Rolewicz says that a C-quasi-norm (1-homogeneous function satisfying the triangle inequality up to a multiplicative constant C) is equivalent to some κ-seminorm (κ-homogeneous function satisfying the triangle inequality) for some κ depending only on C (to be precise, it is enough to take $\kappa = \ln 2/\ln(2C)$). See for instance [21, Lemma 1.1 and Theorem 1.2]. In view of Proposition 1, for every symmetric log-concave random vector X in $\mathbb{R}^n$ the function $N_X(v)^\kappa = e^{\kappa\mathcal{S}(\langle X,v\rangle)}$ with $\kappa = \frac{\ln 2}{1+\ln 2}$ is equivalent to some nonnegative κ-seminorm. Therefore, it is natural to relax Conjecture 1 and ask whether there is a positive universal constant κ such that the function N_X^κ itself satisfies the triangle inequality for every symmetric log-concave random vector X in $\mathbb{R}^n$. Our main result answers this question positively.

Theorem 1. *There exists a universal constant $\kappa > 0$ such that for a symmetric log-concave random vector X in $\mathbb{R}^n$ and two vectors $u,v \in \mathbb{R}^n$ we have*

$$e^{\kappa\mathcal{S}(\langle u+v,X\rangle)} \le e^{\kappa\mathcal{S}(\langle u,X\rangle)} + e^{\kappa\mathcal{S}(\langle v,X\rangle)}. \tag{8}$$

Equivalently, for a symmetric log-concave random vector (X, Y) *in* $\mathbb{R}^2$ *we have*

$$e^{\kappa\mathcal{S}(X+Y))} \leq e^{\kappa\mathcal{S}(X)} + e^{\kappa\mathcal{S}(Y)}. \tag{9}$$

In fact, we can take $\kappa = 1/5$.

Remark 1. If X and Y are independent random variables uniformly distributed on the intervals $[-t/2, t/2]$ and $[-1/2, 1/2]$ with $t < 1$, then (9) becomes $e^{\kappa t/2} \leq 1 + t^\kappa$. Letting $t \to 0$ shows that necessarily $\kappa \leq 1$. We believe that this is the extreme case and the optimal value of κ equals 1.

Remark 2. Inequality (9) with $\kappa = 1$ can be easily shown for log-concave random vectors (X, Y) in $\mathbb{R}^2$ for which one marginal has the same law as the other one rescaled, say $Y \sim tX$ for some $t > 0$. Note that the symmetry of (X, Y) is not needed here. This fact in the essential case of $t = 1$ was first observed in [16]. We recall the argument in the next section. Moreover, in that paper the converse was shown as well: given a density f, the equality

$$\max\{\mathcal{S}(X+Y) : X \sim f,\, Y \sim f\} = \mathcal{S}(2X)$$

holds if and only if f is log-concave, thus characterising log-concavity. For some bounds on $\mathcal{S}(X \pm Y)$ in higher dimensions see [23] and [11].

It will be much more convenient to prove Theorem 1 in an equivalent form, obtained by linearising inequality (9).

Theorem 2. *Let* (X, Y) *be a symmetric log-concave vector in* $\mathbb{R}^2$ *and assume that* $\mathcal{S}(X) = \mathcal{S}(Y)$. *Then for every* $\theta \in [0, 1]$ *we have*

$$\mathcal{S}(\theta X + (1-\theta)Y) \leq S(X) + \frac{1}{\kappa}\ln(\theta^\kappa + (1-\theta)^\kappa), \tag{10}$$

where $\kappa > 0$ *is a universal constant. We can take* $\kappa = 1/5$.

Remark 3. Proving Conjecture 1 is equivalent to showing Theorem 2 with $\kappa = 1$.

Notice that in the above reverse EPI we estimate the entropy of linear combinations of summands whose joint distribution is log-concave. This is different from what would be the straightforward reverse form of the EPI (3) for independent summands with weights $\sqrt{\lambda}$ and $\sqrt{1-\lambda}$ preserving variance. Suppose that the summands X, Y are independent and identically distributed, say with finite variance, and recall (4). Then, as we mentioned in the introduction, the EPI says that the function $[0, 1] \ni \lambda \mapsto \mathcal{S}(X_\lambda)$ is minimal at $\lambda = 0$ and $\lambda = 1$. Following this logic, reversing the EPI could amount to determining the λ for which the maximum of this function occurs. Our next result shows that the somewhat natural guess of $\lambda = 1/2$ is false in general.

Proposition 2. *For each positive* $\lambda_0 < \frac{1}{2(2+\sqrt{2})}$ *there is a symmetric continuous random variable* X *of finite variance for which* $\mathcal{S}(X_{\lambda_0}) > \mathcal{S}(X_{1/2})$.

Nevertheless, we believe that in the log-concave setting the function $\lambda \mapsto \mathcal{S}(X_\lambda)$ should behave nicely.

Conjecture 2. *Let* X *and* Y *be independent copies of a log-concave random variable. Then the function*

$$\lambda \mapsto \mathcal{S}(\sqrt{\lambda}\,X + \sqrt{1-\lambda}\,Y)$$

is concave on $[0,1]$.

3. Proofs

3.1. Theorems 1 and 2 are equivalent. To see that Theorem 2 implies Theorem 1, take a symmetric log-concave random vector (X,Y) in $\mathbb{R}^2$ and take θ with $\mathcal{S}(X/\theta) = \mathcal{S}(Y/(1-\theta))$, that is, $\theta = e^{\mathcal{S}(X)}/(e^{\mathcal{S}(X)} + e^{\mathcal{S}(Y)}) \in [0,1]$. Applying Theorem 2 with the vector $(X/\theta, Y/(1-\theta))$ and using the identity $\mathcal{S}(X/\theta) = \mathcal{S}(X) - \ln\theta = -\ln(e^{\mathcal{S}(X)} + e^{\mathcal{S}(Y)})$ gives

$$\mathcal{S}(X+Y) \le \mathcal{S}(X/\theta) + \frac{1}{\kappa}\ln\frac{e^{\kappa\mathcal{S}(X)} + e^{\kappa\mathcal{S}(Y)}}{(e^{\mathcal{S}(X)} + e^{\mathcal{S}(Y)})^\kappa} = \frac{1}{\kappa}\ln(e^{\kappa\mathcal{S}(X)} + e^{\kappa\mathcal{S}(Y)}),$$

so (9) follows.

To see that Theorem 1 implies Theorem 2, take a log-concave vector (X,Y) with $\mathcal{S}(X) = \mathcal{S}(Y)$ and apply (9) to the vector $(\theta X, (1-\theta)Y)$, which yields

$$\begin{aligned}\mathcal{S}(\theta X + (1-\theta)Y) &\le \frac{1}{\kappa}\ln\bigl(\theta^\kappa e^{\kappa\mathcal{S}(X)} + (1-\theta)^\kappa e^{\kappa\mathcal{S}(Y)}\bigr)\\ &= \mathcal{S}(X) + \frac{1}{\kappa}\ln(\theta^\kappa + (1-\theta)^\kappa).\end{aligned}$$

3.2. Proof of Remark 2. Let $w\colon \mathbb{R}^2 \to [0,\infty)$ be the density of such a vector and let f, g, h be the densities of $X, Y, X+Y$ as in (7). The assumption means that $f(x) = tg(tx)$. By convexity,

$$\mathcal{S}(X+Y) = \inf\Bigl\{-\int h\ln p : p \text{ is a probability density on } \mathbb{R}\Bigr\}.$$

Using Fubini's theorem and changing variables yields

$$\begin{aligned}-\int h\ln p &= -\iint w(x,y)\ln p(x+y)\,dx\,dy\\ &= -\theta(1-\theta)\iint w(\theta x, (1-\theta)y)\ln p(\theta x + (1-\theta)y)\,dx\,dy\end{aligned}$$

for every $\theta \in (0,1)$ and a probability density p. If p is log-concave we get

$$\begin{aligned}\mathcal{S}(X+Y) \le &-\theta^2(1-\theta)\iint w(\theta x,(1-\theta)y)\ln p(x)\,dx\,dy\\ &-\theta(1-\theta)^2\iint w(\theta x,(1-\theta)y)\ln p(y)\,dx\,dy\\ =&-\theta^2\int f(\theta x)\ln p(x)\,dx-(1-\theta)^2\int g\big((1-\theta)y\big)\ln p(y)\,dy.\end{aligned}$$

Set

$$p(x)=\theta f(\theta x)=t\theta g(t\theta x)$$

with θ such that $t\theta = 1-\theta$. Then the last expression becomes

$$\theta\mathcal{S}(X)+(1-\theta)\mathcal{S}(Y)-\theta\ln\theta-(1-\theta)\ln(1-\theta).$$

Since $\mathcal{S}(Y)=\mathcal{S}(X)+\ln t=\mathcal{S}(X)+\ln\frac{1-\theta}{\theta}$, we thus obtain

$$\mathcal{S}(X+Y)\le\mathcal{S}(X)-\ln\theta=\mathcal{S}(X)+\ln(1+t)=\ln(e^{\mathcal{S}(X)}+e^{\mathcal{S}(Y)}).$$

3.3. Proof of Theorem 2. The idea of our proof of Theorem 2 is very simple. For small θ we bound the quantity $\mathcal{S}(\theta X+(1-\theta)Y)$ by estimating its derivative. To bound it for large θ, we shall crudely apply Proposition 1. The exact bound based on estimating the derivative reads as follows.

PROPOSITION 3. *Let (X,Y) be a symmetric log-concave random vector on $\mathbb{R}^2$. Assume that $\mathcal{S}(X)=\mathcal{S}(Y)$ and let $0\le\theta\le\frac{1}{2(1+e)}$. Then*

$$S(\theta X+(1-\theta)Y)\le S(X)+60(1+e)\theta. \tag{11}$$

The main ingredient of the proof of the above proposition is the following lemma. We postpone its proof until the next subsection.

LEMMA 1. *Let $w:\mathbb{R}^2\to\mathbb{R}_+$ be an even log-concave function. Define $f(x)=\int w(x,y)\,dy$ and $\gamma=\int w(0,y)\,dy/\int w(x,0)\,dx$. Then*

$$\iint\frac{-f'(x)}{f(x)}yw(x,y)\,dx\,dy\le 30\gamma\int w.$$

Proof of Proposition 3. For $\theta=0$ both sides of (11) are equal. It is therefore enough to prove that $\frac{d}{d\theta}S(\theta X+(1-\theta)Y)\le 60(1+e)$ for $0\le\theta\le\frac{1}{2(1+e)}$. Let f_θ be the density of $X_\theta=\theta X+(1-\theta)Y$. Note that $f_\theta=e^{-\varphi_\theta}$, where φ_θ is convex. Let $\frac{d\varphi_\theta}{d\theta}=\Phi_\theta$ and $\frac{df_\theta}{d\theta}=F_\theta$. Then $\Phi_\theta=-F_\theta/f_\theta$. Using the chain rule we get

$$\begin{aligned}\frac{d}{d\theta}S(\theta X+(1-\theta)Y)&=-\frac{d}{d\theta}\mathbb{E}\ln f_\theta(X_\theta)=\frac{d}{d\theta}\mathbb{E}\varphi_\theta(X_\theta)\\ &=\mathbb{E}\Phi_\theta(X_\theta)+\mathbb{E}\varphi_\theta'(X_\theta)(X-Y).\end{aligned}$$

Moreover,

$$\mathbb{E}\Phi_\theta(X_\theta)=-\mathbb{E}F_\theta(X_\theta)/f_\theta(X_\theta)=-\int F_\theta(x)\,dx=-\frac{d}{d\theta}\int f_\theta(x)\,dx=0.$$

Let $Z_\theta = (X_\theta, X-Y)$ and let w_θ be the density of Z_θ. Using Lemma 1 with $w = w_\theta$ gives

$$\frac{d}{d\theta} S(\theta X + (1-\theta)Y) = -\mathbb{E}\left(\frac{f'_\theta(X_\theta)}{f_\theta(X_\theta)}(X-Y)\right)$$
$$= -\int \frac{f_\theta(x)}{f_\theta(x)} y w_\theta(x,y)\,dx\,dy \le 30\gamma_\theta,$$

where $\gamma_\theta = \int w_\theta(0,y)\,dy / \int w_\theta(x,0)\,dx$. It suffices to show that $\gamma_\theta \le 2(1+e)$ for $0 \le \theta \le \frac{1}{2(1+e)}$. Let w be the density of (X,Y). Then we have $w_\theta(x,y) = w(x+(1-\theta)y, x-\theta y)$. To finish the proof we again use the fact that $\|v\|_w = (\int w(tv)\,dt)^{-1}$ is a norm. Note that

$$\gamma_\theta = \frac{\int w_\theta(0,y)\,dy}{\int w_\theta(x,0)\,dx} = \frac{\int w((1-\theta)y, -\theta y)\,dy}{\int w(x,x)\,dx} = \frac{\|e_1+e_2\|_w}{\|(1-\theta)e_1 - \theta e_2\|_w}.$$

Let $f(x) = \int w(x,y)\,dy$ and $g(x) = \int w(y,x)\,dy$ be the densities of the log-concave random variables X and Y, respectively. Observe that by (6) we have

$$\|f\|_\infty^{-1} \le e^{\mathcal{S}(X)} \le e\|f\|_\infty^{-1}, \qquad \|g\|_\infty^{-1} \le e^{\mathcal{S}(Y)} \le e\|g\|_\infty^{-1}.$$

Since $\|f\|_\infty^{-1} = f(0)^{-1} = \|e_1\|_w$, $\|g\|_\infty^{-1} = g(0)^{-1} = \|e_2\|_w$ and $\mathcal{S}(X) = \mathcal{S}(Y)$, this gives $e^{-1} \le \|e_1\|_w / \|e_2\|_w \le e$. Thus, by the triangle inequality,

$$\gamma_\theta \le \frac{\|e_1\|_w + \|e_2\|_w}{(1-\theta)\|e_1\|_w - \theta\|e_2\|_w}$$
$$\le \frac{(1+e)\|e_1\|_w}{(1-\theta)\|e_1\|_w - \theta e\|e_1\|_w} = \frac{1+e}{1-\theta(1+e)} \le 2(1+e). \ \blacksquare$$

Proof of Theorem 2. We can assume that $\theta \in [0,1/2]$. Using Proposition 1 with the vector $(\theta X, (1-\theta)Y)$ and the fact that $\mathcal{S}(X) = \mathcal{S}(Y)$ we get $\mathcal{S}(\theta X + (1-\theta)Y) \le \mathcal{S}(X) + 1$. Thus, from Proposition 3 we deduce that it is enough to find $\kappa > 0$ such that

$$\min\{1, 60(1+e)\theta\} \le \kappa^{-1} \ln(\theta^\kappa + (1-\theta)^\kappa), \qquad \theta \in [0,1/2]$$

(if $60(1+e)\theta < 1$ then $\theta < \frac{1}{2(1+e)}$ and therefore Proposition 3 indeed can be used in this case). By the concavity and monotonicity of the right hand side it is enough to check this inequality at $\theta_0 = (60(1+e))^{-1}$, that is, we have to verify the inequality $e^\kappa \le \theta_0^\kappa + (1-\theta_0)^\kappa$. We check that it is true for $\kappa = 1/5$. $\blacksquare$

3.4. Proof of Lemma 1. We start off by establishing two simple and standard lemmas. The second one is a limiting case of the so-called Grünbaum theorem (see [19] and [26]).

Lemma 2. *Let $f:\mathbb{R}\to\mathbb{R}_+$ be an even log-concave function. For $\beta>0$ define a_β by*

$$a_\beta = \sup\{x>0 : f(x)\ge e^{-\beta}f(0)\}.$$

Then

$$2e^{-\beta}a_\beta \le \frac{1}{f(0)}\int f \le 2(1+\beta^{-1}e^{-\beta})a_\beta.$$

Proof. Since f is even and log-concave, it is maximal at zero and non-increasing on $[0,\infty)$. Consequently, the left hand inequality immediately follows from the definition of a_β. By comparing $\ln f$ with an appropriate linear function, log-concavity also guarantees that $f(x)\le f(0)e^{-\beta x/a_\beta}$ for $|x|>a_\beta$, hence

$$\int f \le 2a_\beta f(0) + 2\int_{a_\beta}^{\infty} f(0)e^{-\beta x/a_\beta}\,dx = 2a_\beta f(0) + 2f(0)\frac{a_\beta}{\beta}e^{-\beta},$$

which gives the right hand inequality. ∎

Lemma 3. *Let X be a log-concave random variable. Let a be such that $\mathbb{P}(X>a)\le e^{-1}$. Then $\mathbb{E}X\le a$.*

Proof. Without loss of generality assume that X is a continuous random variable and that $\mathbb{P}(X>a)=e^{-1}$. Moreover, the statement is translation invariant, so we can assume that $a=0$. Let $e^{-\varphi}$ be the density of X, where φ is convex. There exists a function ψ of the form

$$\psi(x)=\begin{cases} ax+b, & x\ge L,\\ \infty, & x<L,\end{cases}$$

such that $\psi(0)=\varphi(0)$ and $e^{-\psi}$ is the probability density of a random variable Y with $\mathbb{P}(Y>a)=e^{-1}$. One can check, using convexity of φ, that $\mathbb{E}X\le\mathbb{E}Y$. We have $1=\int e^{-\psi}=\frac{1}{a}e^{-(b+aL)}$ and $e^{-1}=\int_0^\infty e^{-\psi}=\frac{1}{a}e^{-b}$. It follows that $aL=-1$, and so $\mathbb{E}X\le\mathbb{E}Y=\frac{1}{a}\big(L+\frac{1}{a}\big)e^{-(b+aL)}=0$. ∎

Proof of Lemma 1. Without loss of generality assume that w is strictly log-concave and $w(0)=1$. First we derive a pointwise estimate on w which will enable us to obtain good pointwise bounds on the quantity $\int yw(x,y)\,dy$, relative to $f(x)$. To this end, set unique positive parameters a and b to be such that $w(a,0)=e^{-1}=w(0,b)$. Consider $l\in(0,a)$. We have

$$w(-l,0)=w(l,0)\ge w(a,0)^{l/a}w(0,0)^{1-l/a}=e^{-l/a}.$$

Fix $x>0$ and let $y>\frac{b}{a}x+b$. Let l be such that the line passing through the points $(0,b)$ and (x,y) intersects the x-axis at $(-l,0)$, that is $l=\frac{bx}{y-b}$.

Note that $l \in (0, a)$. Then

$$\begin{aligned} e^{-1} &= w(0,b) \geq w(x,y)^{b/y} w(-l,0)^{1-b/y} \\ &\geq w(x,y)^{b/y} e^{-(1-b/y)/a} = [w(x,y) e^{-(l/a)(y/b)(y-b)/y}]^{b/y}, \end{aligned}$$

hence

$$w(x,y) \leq e^{x/a - y/b} \qquad \text{for } x > 0 \text{ and } y > \frac{b}{a}x + b.$$

Let X be a random variable with log-concave density $y \mapsto w(x,y)/f(x)$. Set $\beta = b + b\ln(\max\{f(0), b\})$ and

$$\alpha = \frac{b}{a}x - b\ln f(x) + \beta.$$

Since f is maximal at zero (as it is an even log-concave function), we check that

$$\alpha \geq \frac{b}{a}x - b\ln f(0) + \beta \geq \frac{b}{a}x + b,$$

so we can use the pointwise estimate on w to get

$$\begin{aligned} \int_\alpha^\infty w(x,y)\,dy &\leq e^{x/a} \int_\alpha^\infty e^{-y/b}\,dy = be^{x/a - \alpha/b} \\ &= \frac{b}{\max\{f(0), b\}} e^{-1} f(x) \leq e^{-1} f(x). \end{aligned}$$

This means that $\mathbb{P}(X > \alpha) \leq e^{-1}$, which in view of Lemma 3 yields

$$\frac{1}{f(x)} \int y w(x,y)\,dy = \mathbb{E}X \leq \alpha = \frac{b}{a}x - b\ln f(x) + \beta \qquad \text{for } x > 0.$$

Having obtained this bound, we can easily estimate the quantity stated in the lemma. By the symmetry of w we have

$$\iint \frac{-f'(x)}{f(x)} y w(x,y)\,dx\,dy = 2 \int \int_{x>0} \frac{-f'(x)}{f(x)} y w(x,y)\,dx\,dy.$$

Since f decreases on $[0, \infty)$, the factor $-f'(x)$ is nonnegative for $x > 0$, thus we can further write

$$\begin{aligned} \iint \frac{-f'(x)}{f(x)} y w(x,y)\,dx\,dy &\leq 2 \int_0^\infty -f'(x) \left(\frac{b}{a}x - b\ln f(x) + \beta \right) dx \\ &= 2f(0)(-b\ln f(0) + \beta) \\ &\quad + 2 \int_0^\infty f(x) \left(\frac{b}{a} - b\frac{f'(x)}{f(x)} \right) dx \\ &= 2f(0)b\left(1 + \ln \frac{\max\{f(0), b\}}{f(0)} \right) + \frac{b}{a} \int w + 2f(0)b. \end{aligned}$$

Now we only need to put the finishing touches to this expression. By Lemma 2 applied to the functions $x \mapsto w(x,0)$ and $y \mapsto w(0,y)$ we obtain

$$\frac{b}{a} \le \frac{e}{2} 2(1+e^{-1}) \frac{\int w(0,y)\,dy}{\int w(x,0)\,dx} = (e+1)\gamma$$

and $b/f(0) \le e/2$. Estimating the logarithm yields

$$1 + \ln \frac{\max\{f(0), b\}}{f(0)} \le \frac{\max\{f(0), b\}}{f(0)} \le \frac{e}{2}.$$

Finally, by log-concavity,

$$\begin{aligned} \int w(x,y)\,dx\,dy &\ge \int \sqrt{w(2x,0)w(0,2y)}\,dx\,dy \\ &= \frac{1}{4} \int \sqrt{w(x,0)}\,dx \int \sqrt{w(0,y)}\,dy \end{aligned}$$

and

$$\int w(x,0)\,dx \le \sqrt{w(0,0)} \int \sqrt{w(x,0)}\,dx = \int \sqrt{w(x,0)}\,dx.$$

Combining these two estimates we get

$$f(0) = \int w(0,y)\,dy \le \int \sqrt{w(0,y)}\,dy \le \frac{4\int w}{\int w(x,0)\,dx},$$

and consequently

$$f(0)b \le \frac{e}{2} f(0) f(0) \le 2e f(0) \frac{\int w}{\int w(x,0)\,dx} = 2e\gamma \int w.$$

Finally,

$$\iint \frac{-f'(x)}{f(x)} y w(x,y)\,dx\,dy \le (2e^2 + 5e + 1)\gamma \int w,$$

and the assertion follows. ■

3.5. Proof of Proposition 2. For a real number s and nonnegative numbers $\alpha \le \beta$ we define the following trapezoidal function

$$T^s_{\alpha,\beta}(x) = \begin{cases} 0 & \text{if } x < s \text{ or } x > s + \alpha + \beta, \\ x - s & \text{if } s \le x \le s + \alpha, \\ \alpha & \text{if } s + \alpha \le x \le s + \beta, \\ s + \alpha + \beta - x & \text{if } s + \beta \le x \le s + \alpha + \beta. \end{cases}$$

The motivation is the following convolution identity: for real numbers a, a' and nonnegative numbers h, h' such that $h \le h'$ we have

$$\mathbf{1}_{[a,a+h]} \star \mathbf{1}_{[a',a'+h']} = T^{a+a'}_{h,h'}. \tag{12}$$

It is also easy to check that

$$\int_{\mathbb{R}} T^s_{\alpha,\beta} = \alpha\beta. \tag{13}$$

We shall need one more formula: for any real number s and nonnegative numbers A,α,β with $\alpha\le\beta$ we have

$$I(A,\alpha,\beta)=\int_{\mathbb{R}} AT^s_{\alpha,\beta}\ln(AT^s_{\alpha,\beta})=A\alpha\beta\ln(A\alpha)-\tfrac{1}{2}A\alpha^2. \tag{14}$$

Fix $0<a<b=a+h$. Let X be a random variable with density

$$f(x)=\frac{1}{2h}(\mathbf{1}_{[-b,-a]}(x)+\mathbf{1}_{[a,b]}(x)).$$

We shall compute the density f_λ of X_λ. Denote $u=\sqrt{\lambda}$, $v=\sqrt{1-\lambda}$ and without loss of generality let $\lambda\le 1/2$. Clearly, $f_\lambda(x)=\frac{1}{u}f\big(\frac{\cdot}{u}\big)\star\frac{1}{v}f\big(\frac{\cdot}{v}\big)(x)$, so by (12) we have

$$\begin{aligned}f_\lambda(x)&=\big(\mathbf{1}_{u[-b,-a]}\star\mathbf{1}_{v[-b,-a]}+\mathbf{1}_{u[a,b]}\star\mathbf{1}_{v[-b,-a]}\\&\qquad+\mathbf{1}_{u[-b,-a]}\star\mathbf{1}_{v[a,b]}+\mathbf{1}_{u[a,b]}\star\mathbf{1}_{v[a,b]}\big)(x)\cdot\frac{1}{(2h)^2uv}\\&=\Big(\underbrace{T^{-(u+v)b}_{uh,vh}}_{T_1}(x)+\underbrace{T^{ua-vb}_{uh,vh}}_{T_2}(x)+\underbrace{T^{-ub+va}_{uh,vh}}_{T_3}(x)+\underbrace{T^{(u+v)a}_{uh,vh}}_{T_4}(x)\Big)\cdot\frac{1}{(2h)^2uv}.\end{aligned}$$

This symmetric density is a superposition of four trapezoid functions T_1, T_2, T_3, T_4 which are certain shifts of the same trapezoid function $T_0=T^0_{uh,vh}$. The shifts may overlap depending on the value of λ. Now we shall consider two cases.

Case 1: $\lambda=1/2$. Then $u=v=1/\sqrt{2}$. Notice that T_0 becomes a triangle looking function and $T_2=T_3$, so we obtain

$$f_{1/2}(x)=\frac{1}{2h^2}(T^{-b\sqrt{2}}_{h/\sqrt{2},h/\sqrt{2}}+2T^{-h/\sqrt{2}}_{h/\sqrt{2},h/\sqrt{2}}+T^{a\sqrt{2}}_{h/\sqrt{2},h/\sqrt{2}})(x).$$

If $h/\sqrt{2}<a\sqrt{2}$ then the supports of the summands are disjoint and with the aid of identity (14) we obtain

$$\mathcal{S}(X_{1/2})=-2I\bigg(\frac{1}{2h^2},\frac{h}{\sqrt{2}},\frac{h}{\sqrt{2}}\bigg)-I\bigg(\frac{1}{h^2},\frac{h}{\sqrt{2}},\frac{h}{\sqrt{2}}\bigg)=\ln(2h)+\frac{1}{2}.$$

Case 2: λ small. Now we choose $\lambda=\lambda_0$ so that the supports of T_1 and T_2 intersect in such a way that the down-slope of T_1 adds up to the up-slope of T_2 giving a flat piece. This happens when $-b(u+v)+vh=ua-bv$, that is,

$$\sqrt{\frac{1-\lambda_0}{\lambda_0}}=\frac{v}{u}=\frac{a+b}{h}=2\frac{a}{h}+1. \tag{15}$$

The earlier condition $a/h>1/2$ implies that $\lambda_0<1/5$. With the above choice for λ we have $T_1+T_2=T^{-b(u+v)}_{uh,2vh}$, hence by symmetry

$$f_\lambda=(T^{-b(u+v)}_{uh,2vh}+T^{-ub+va}_{uh,2vh})\cdot\frac{1}{(2h)^2uv}.$$

As long as $-ub+va>0$, the supports of these two trapezoid functions are disjoint. Given our choice for λ, this is equivalent to $v/u > b/a = 1+h/a = 1+2/(v/u-1)$, or if we set $v/u = \sqrt{1/\lambda_0 - 1}$, to $\lambda_0 < \frac{1}{2(2+\sqrt{2})}$. Then once again $\lambda_0 < 1/5$ and we get

$$\begin{aligned}\mathcal{S}(X_{\lambda_0}) &= -2I\left(\frac{1}{(2h)^2uv}, uh, 2vh\right) = \ln(4vh) + \frac{u}{4v} \\ &= \ln(4h\sqrt{1-\lambda_0}) + \frac{1}{4}\sqrt{\frac{\lambda_0}{1-\lambda_0}}.\end{aligned}$$

Furtheremore,

$$\mathcal{S}(X_{\lambda_0}) - \mathcal{S}(X_{1/2}) = \ln 2 - \frac{1}{2} + \ln\sqrt{1-\lambda_0} + \frac{1}{4}\sqrt{\frac{\lambda_0}{1-\lambda_0}}.$$

We check that the right hand side is positive for $\lambda_0 < \frac{1}{2(2+\sqrt{2})}$. Therefore, we have shown that for each such λ_0 there is a choice of the parameters a and h (given by (15)), and hence a random variable X, for which $\mathcal{S}(X_{\lambda_0}) > \mathcal{S}(X_{1/2})$.

References

[1] S. Artstein, K. M. Ball, F. Barthe and A. Naor, *On the rate of convergence in the entropic central limit theorem*, Probab. Theory Related Fields 129 (2004), 381–390.

[2] S. Artstein, K. M. Ball, F. Barthe and A. Naor, *Solution of Shannon's problem on the monotonicity of entropy*, J. Amer. Math. Soc. 17 (2004), 975–982.

[3] K. Ball, *Logarithmically concave functions and sections of convex sets in* $\mathbb{R}^n$, Studia Math. 88 (1988), 69–84.

[4] K. Ball, F. Barthe and A. Naor, *Entropy jumps in the presence of a spectral gap*, Duke Math. J. 119 (2003), 41–63.

[5] K. Ball and V. H. Nguyen, *Entropy jumps for isotropic log-concave random vectors and spectral gap*, Studia Math. 213 (2012), 81–96.

[6] N. M. Blachman, *The convolution inequality for entropy powers*, IEEE Trans. Inform. Theory 11 (1965), 267–271.

[7] S. G. Bobkov and G. P. Chistyakov, *Entropy power inequality for the Rényi entropy*, IEEE Trans. Inform. Theory 61 (2015), 708–714.

[8] S. G. Bobkov, G. P. Chistyakov and F. Götze, *Stability problems in Cramér-type characterization in case of i.i.d. summands*, Theory Probab. Appl. 57 (2013), 568–588.

[9] S. G. Bobkov, G. P. Chistyakov and F. Götze, *Stability of Cramér's characterization of normal laws in information distances*, arXiv:1512.03571 (2015).

[10] S. G. Bobkov and M. Madiman, *Reverse Brunn–Minkowski and reverse entropy power inequalities for convex measures*, J. Funct. Anal. 262 (2012), 3309–3339.

[11] S. G. Bobkov and M. Madiman, *On the problem of reversibility of the entropy power inequality*, in: Limit Theorems in Probability, Statistics and Number Theory, Springer Proc. Math. Statist. 42, Springer, Heidelberg, 2013, 61–74.

[12] S. G. Bobkov and M. Madiman, *The entropy per coordinate of a random vector is highly constrained under convexity conditions*, IEEE Trans. Inform. Theory 57 (2011), 4940–4954.

[13] E. A. Carlen, *Superadditivity of Fisher's information and logarithmic Sobolev inequalities*, J. Funct. Anal. 101 (1991), 194–211.

[14] E. A. Carlen and A. Soffer, *Entropy production by block variable summation and central limit theorems*, Comm. Math. Phys. 140 (1991), 339–371.

[15] M. H. M. Costa, *A new entropy power inequality*, IEEE Trans. Inform. Theory 31 (1985), 751–760.

[16] T. M. Cover and Z. Zhang, *On the maximum entropy of the sum of two dependent random variables*, IEEE Trans. Inform. Theory 40 (1994), 1244–1246.

[17] A. Dembo, *Simple proof of the concavity of the entropy power with respect to added Gaussian noise*, IEEE Trans. Inform. Theory 35 (1989), 887–888.

[18] A. Dembo, T. M. Cover and J. A. Thomas, *Information-theoretic inequalities*, IEEE Trans. Inform. Theory 37 (1991), 1501–1518.

[19] B. Grünbaum, *Partitions of mass-distributions and of convex bodies by hyperplanes*, Pacific J. Math. 10 (1960), 1257–1261.

[20] O. Johnson and A. Barron, *Fisher information inequalities and the central limit theorem*, Probab. Theory Related Fields 129 (2004), 391–409.

[21] N. J. Kalton, N. T. Peck and J. W. Roberts, *An F-space Sampler*, London Math. Soc. Lecture Note Ser. 89, Cambridge Univ. Press, Cambridge, 1984.

[22] E. H. Lieb, *Proof of an entropy conjecture of Wehrl*, Comm. Math. Phys. 62 (1978), 35–41.

[23] M. Madiman and I. Kontoyiannis, *The Ruzsa divergence for random elements in locally compact abelian groups*, arXiv:1508.04089v1 (2015).

[24] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. 27 (1948), 379–423, 623–656.

[25] A. J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Inform. Control 2 (1959), 101–112.

[26] S. Szarek, *On measures of symmetry and floating bodies*, arXiv:1302.2076 (2013).

[27] S. Verdú and D. Guo, *A simple proof of the entropy-power inequality*, IEEE Trans. Inform. Theory 52 (2006), 2165–2166.

[28] C. Villani, *A short proof of the "concavity of entropy power"*, IEEE Trans. Inform. Theory 46 (2000), 1695–1696.

Keith Ball, Tomasz Tkocz
Mathematics Institute
University of Warwick
Coventry CV4 7AL, UK
E-mail: k.m.ball@warwick.ac.uk
t.tkocz@warwick.ac.uk

Piotr Nayar
Institute of Mathematics & Applications
Minneapolis, MN 55455, U.S.A.
E-mail: nayar@mimuw.edu.pl