Links between the Logarithmic Sobolev Inequality and the convolution inequalities for Entropy and Fisher Information

Thomas A. Courtade  
Department of Electrical Engineering and Computer Sciences  
University of California, Berkeley  

August 22, 2016

Abstract

Relative to the Gaussian measure on \( \mathbb{R}^d \), entropy and Fisher information are famously related via Gross’ logarithmic Sobolev inequality (LSI). These same functionals also separately satisfy convolution inequalities, as proved by Stam. We establish a dimension-free inequality that interpolates among these relations. Several interesting corollaries follow: (i) the deficit in the LSI satisfies a convolution inequality itself; (ii) the deficit in the LSI controls convergence in the entropic and Fisher information central limit theorems; and (iii) the LSI is stable with respect to HWI jumps (i.e., a jump in any of the convolution inequalities associated to the HWI functionals).

Another consequence is that the convolution inequalities for Fisher information and entropy powers are reversible in general, up to a factor depending on the Stam defect. An improved form of Nelson’s hypercontractivity estimate also follows. Finally, we speculate on the possibility of an analogous reverse Brunn-Minkowski inequality and a related upper bound on surface area associated to Minkowski sums.

1 Introduction

For a random vector \( X \) on \( \mathbb{R}^d \) with absolutely continuous density \( f \), the entropy of \( X \) is given by

\[
h(X) = - \int f \log f,
\]

provided the integral exists, and the entropy power of \( X \) is defined according to

\[
N(X) = \frac{1}{2\pi e} e^{\frac{1}{2} h(X)}.
\]

The Fisher information of \( X \) is defined by

\[
J(X) = \int f |\nabla \log f|^2,
\]

with \( J(X) = \infty \) if the integral does not exist.

The entropy and Fisher information functionals play a fundamental role in information theory and related fields, and enjoy many useful properties. Standing out among these properties is their
behavior under convolution of densities. In particular, Stam [1] and Blachman [2] proved that if $X, Y$ are independent random vectors on $\mathbb{R}^d$, then

$$N(X + Y) \geq N(X) + N(Y)$$

and

$$\frac{1}{J(X + Y)} \geq \frac{1}{J(X)} + \frac{1}{J(Y)}.$$

These inequalities may also be stated in terms of relative entropies and Fisher informations, which will be useful for our purposes. Toward this end, the entropy of $X$ relative to the standard normal $N(0, I)$ is

$$D(X) := D(f \| \phi) = \int f \log \frac{f}{\phi},$$

where $\phi(x) = (2\pi)^{-d/2}e^{-|x|^2/2}$ denotes the Gaussian density on $\mathbb{R}^d$. By Jensen’s inequality, $D(X) \geq 0$, with equality iff $X \sim N(0, I)$. Similarly, the Fisher information of $X$ relative to $N(0, I)$ is defined according to

$$I(X) := I(f \| \phi) = \int \left| \nabla \log \frac{f}{\phi} \right|^2.$$

As with $D(X)$, the quantity $I(X)$ is nonnegative, and zero only if $X \sim N(0, I)$. Completely equivalent to (4) and (5), respectively, are the inequalities

$$\theta D(X) + \bar{\theta} D(Y) \geq D(\sqrt{\theta} X + \sqrt{\bar{\theta}} Y)$$

and

$$\theta I(X) + \bar{\theta} I(Y) \geq I(\sqrt{\theta} X + \sqrt{\bar{\theta}} Y),$$

where $X, Y$ are independent zero-mean random vectors, $\theta \in [0, 1]$ and $\bar{\theta} := 1 - \theta$. Given their equivalence, we shall refer to inequalities (4) and (8) collectively as the entropy power inequality (EPI), and inequalities (5) and (9) collectively as the Fisher information inequality (FII).

Evidently, the EPI and FII apply separately to the entropy and Fisher information functionals. However, in 1975, Gross established a remarkable inequality directly relating relative Fisher information to relative entropy [3]. Namely,

$$\delta_{\text{LSI}}(X) := \frac{1}{2} I(X) - D(X) \geq 0,$$

which is known as the logarithmic Sobolev inequality (LSI) for standard Gaussian measure. The quantity $\delta_{\text{LSI}}(X)$ is the deficit in the LSI associated to $X$, and is zero iff $X$ is a translate of the standard normal [4]. The LSI has a variety of important consequences including Talagrand’s quadratic transportation cost inequality [5], the Gaussian concentration inequality for Lipschitz functions (e.g., [6]), and the Gaussian Poincaré inequality. An interesting feature of the LSI (and many of its corollaries) is the fact that it has no effective dependence on dimension. The convolution inequalities satisfied by Fisher information and relative entropy also enjoy this dimension-free property.
Although it was not recognized until the 1990s by Carlen \cite{4}, Gross’ LSI is in fact mathematically equivalent to the uncertainty principle

\[ p(X) := \frac{1}{d} N(X) J(X) \geq 1, \]  

which was observed by Stam in his 1959 proof of (4) and (5). We refer to the quantity \( p(X) \) as the \textit{Stam defect} associated to \( X \), and remark here that \( p(X) = 1 \) iff \( X \sim N(0, \sigma^2 I) \) for some \( \sigma^2 > 0 \). In fact, (11) is a direct consequence of the EPI and the so-called de Bruijn identity, which suggests a quantitative relationship between the LSI and the EPI. Unfortunately, the derivation of the LSI from the EPI does not propagate any deficit terms, so only conditions for equality are carried through. In this paper, we fill this gap by proving a general inequality that interpolates between the LSI, the EPI and the FII.

\section*{Organization}

The remainder of this paper is organized as follows. Section 2 contains our two main results, which are ultimately shown to be equivalent. In particular, Section 2.1 gives a general interpolation inequality for the LSI and EPI, followed by a brief discussion. Section gives new reverse EPI and FII, and contains proofs of all main results. Section 3 provides applications of the main results, including consequences for information-theoretic central limit theorems, stability of the LSI with respect to HWI jumps, and a sharp form of Nelson’s hypercontractive inequality. In Section 4, we conclude by speculating on the possibility of geometric analogues of our main results.

\section{Main Results}

\subsection{An Interpolation Inequality for the LSI and EPI}

\textbf{Theorem 1.} Let \( X, Y \) be independent, centered random vectors on \( \mathbb{R}^d \). For any \( \theta \in [0, 1] \)

\[ D(X) + D(Y) \leq \frac{\theta}{2} I(X) + \frac{\theta}{2} I(Y) + D(\sqrt{\theta} X + \sqrt{\theta} Y). \]  

(12)

Let us briefly discuss a few observations. If we adopt the convention that \( \delta_{\text{LSI}}(X) = \infty \) when \( I(X) = \infty \), and \( \delta_{\text{LSI}}(X) = \frac{1}{2} I(X) - D(X) \) otherwise, then (12) may be rewritten as

\[ \theta D(X) + \theta D(Y) \leq \theta \delta_{\text{LSI}}(X) + \theta \delta_{\text{LSI}}(Y) + D(\sqrt{\theta} X + \sqrt{\theta} Y), \]  

(13)

from which it is plain that (12) interpolates between the LSI for \( X \), the LSI for \( Y \) and the EPI. However, Theorem 1 also directly connects the LSI to the FII. Indeed, definitions and algebra yield the following equivalent form:

\[ \delta_{\text{LSI}}(\sqrt{\theta} X + \sqrt{\theta} Y) + \theta \frac{1}{2} I(X) + \theta \frac{1}{2} I(Y) \leq \delta_{\text{LSI}}(X) + \delta_{\text{LSI}}(Y) + \frac{1}{2} I(\sqrt{\theta} X + \sqrt{\theta} Y), \]  

(14)

which interpolates between the FII and the LSIs for \( X, Y \) and the sum \( \sqrt{\theta} X + \sqrt{\theta} Y \). Although it is a weakening of (14), we may apply the FII to find that \( \delta_{\text{LSI}} \) satisfies its own convolution inequality which remains an improvement of Gross’ LSI:
Corollary 1 (Convolution inequality for the LSI). Let $X, Y$ be independent random vectors on $\mathbb{R}^d$. For all $\theta \in [0, 1]$

$$\delta_{\text{LSI}}(\sqrt{\theta} X + \sqrt{\theta} Y) \leq \delta_{\text{LSI}}(X) + \delta_{\text{LSI}}(Y).$$  \hspace{1cm} (15)$$

Remark 1. The assumption of centered random vectors is not needed for (15) due to the fact that $\delta_{\text{LSI}}(\cdot)$ is translation invariant. In fact, the centering assumption is not needed in Theorem 1 either, and can be dealt with by incorporating the inner product $\langle \mathbb{E} X, \mathbb{E} Y \rangle$ into (12). Details are straightforward and are left to the reader.

Theorem 1 is essentially best possible for any choice of $X, Y$. Indeed, observe that definitions and the EPI imply

$$\delta_{\text{LSI}}(X) + \delta_{\text{LSI}}(Y) \leq 2 \left( \frac{1}{4} I(X) + \frac{1}{4} I(Y) - \frac{1}{2} I \left( \frac{1}{\sqrt{2}} (X + Y) \right) + \delta_{\text{LSI}} \left( \frac{1}{\sqrt{2}} (X + Y) \right) \right).$$ \hspace{1cm} (16)$$

However, Theorem 1 implies via (14) that

$$\delta_{\text{LSI}}(X) + \delta_{\text{LSI}}(Y) \geq \sup_{\theta \in [0, 1]} \left( \frac{\theta}{2} I(X) + \frac{\bar{\theta}}{2} I(Y) - \frac{1}{2} I \left( \sqrt{\theta} X + \sqrt{\bar{\theta}} Y \right) + \delta_{\text{LSI}} \left( \sqrt{\theta} X + \sqrt{\bar{\theta}} Y \right) \right) \geq \frac{1}{4} I(X) + \frac{1}{4} I(Y) - \frac{1}{2} I \left( \frac{1}{\sqrt{2}} (X + Y) \right) - \delta_{\text{LSI}} \left( \frac{1}{\sqrt{2}} (X + Y) \right),$$ \hspace{1cm} (17)$$

which differs from corresponding the upper bound (16) by precisely a factor of 2.

Loosely speaking, the conjunction of (16) and (18) suggests that $\delta_{\text{LSI}}(X) + \delta_{\text{LSI}}(Y)$ can be roughly decomposed into two nonnegative parts that depend jointly on $X, Y$: (i) the dissipation of Fisher information $\frac{1}{2} \left( I(X) + I(Y) \right) - I \left( \frac{1}{\sqrt{2}} (X + Y) \right)$; and (ii) the deficit in the LSI associated to the rescaled sum $\frac{1}{\sqrt{2}} (X + Y)$. On this note, we remark that neither of these quantities control one another in general.

To see this, let $\rho \in (0, 1)$ and consider Gaussian random vectors $X, Y$ with distributions:

$$X \sim N \left( 0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \quad Y \sim N \left( 0, \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \right).$$ \hspace{1cm} (19)$$

In this case, $\frac{1}{\sqrt{2}} (X + Y) \sim N(0, I)$ so $\delta_{\text{LSI}} \left( \frac{1}{\sqrt{2}} (X + Y) \right) = 0$. However,

$$\frac{1}{2} \left( I(X) + I(Y) \right) - I \left( \frac{1}{\sqrt{2}} (X + Y) \right) = \frac{2}{1 - \rho^2} - 2 = \frac{2\rho^2}{1 - \rho^2}. \hspace{1cm} (20)$$

On the other hand, consider $X_\ast$ to be an independent copy of $X$. In this case, $\frac{1}{\sqrt{2}} (X + X_\ast)$ is equal to $X$ in distribution, so

$$\frac{1}{2} \left( I(X) + I(X_\ast) \right) - I \left( \frac{1}{\sqrt{2}} (X + X_\ast) \right) = 0. \hspace{1cm} (21)$$

However, we may readily compute that

$$\delta_{\text{LSI}} \left( \frac{1}{\sqrt{2}} (X + X_\ast) \right) = \frac{\rho^2}{1 - \rho^2} + \frac{1}{2} \log(1 - \rho^2) > \frac{\rho^2}{2}. \hspace{1cm} (22)$$
Another simple consequence of the above discussion is that, if $X, X_*$ are independent and identically distributed, then

$$\delta_{LSI}(X) \asymp \frac{1}{2} \left( I(X) - I \left( \frac{1}{\sqrt{2}} (X + X_*) \right) \right) + \delta_{LSI} \left( \frac{1}{\sqrt{2}} (X + X_*) \right),$$

where ‘$\asymp$’ denotes equality up to an absolute constant factor. This suggests that the Fisher information jump $I(X) - I \left( \frac{1}{\sqrt{2}} (X + X_*) \right)$ can be used to quantify the stability of the LSI, a topic we will return to in Section 3.2.

Finally, we note that Theorem 1 allows us to easily deduce the (well-known) equality conditions for the LSI from those for the EPI. Indeed, since $\delta_{LSI}(\cdot)$ and $D(\cdot)$ are invariant to unitary transformations, it follows from Theorem 1 that $\delta_{LSI}(X) = \delta_{LSI}(Y) = 0$ only if

$$\theta D(X) + \bar{\theta} D(Y) = D(\sqrt{\theta} X + \sqrt{\bar{\theta}} U Y),$$

for all $\theta \in [0,1]$ and unitary matrices $U : \mathbb{R}^d \to \mathbb{R}^d$. From conditions for equality in the EPI, this implies $\text{Cov}(X)$ and $\text{Cov}(Y)$ are proportional to $I$; in fact, direct computation shows they must be equal to satisfy (24). Thus, evaluation of $\delta_{LSI}(X)$ for $X \sim N(\mu, \sigma^2 I)$ allows us to conclude that $\delta_{LSI}(X) = 0$ if and only if $X \sim N(\mu, I)$.

### 2.2 Reverse Entropy Power and Fisher Information Inequalities

Due to its fundamental role in information theory, there has been sustained interest in obtaining reverse forms of the entropy power inequality (4). As shown by Bobkov and Chistyakov [7], the EPI cannot be reversed in general, at least not up to a constant factor. Nevertheless, progress has been made. A notable example of a reverse EPI is due to Bobkov and Madiman [8], who show that for independent random vectors $X, Y$ with log-concave densities, there exist linear volume preserving maps $u, v$ such that

$$N(u(X) + v(Y)) \leq C(N(X) + N(Y)),$$

where $C$ is an absolute constant. Bobkov and Madiman’s result mirrors Milman’s reverse Brunn-Minkowski inequality [9], which is pleasant since the EPI itself mirrors the Brunn-Minkowski inequality. A similar statement holds for a more general class of convex measures. See also the recent survey by Madiman, Melbourne and Xu [10] for related results.

Another example of a reverse EPI is due to Ball, Nayar and Tkocz [11], who restrict attention to the class of log-concave densities. They show that, for a symmetric log-concave vector $(X, Y)$ in $\mathbb{R}^2$, there is an absolute constant $\kappa$ such that

$$N(X + Y)^\kappa \leq N(X)^\kappa + N(Y)^\kappa.$$

We show below that both the entropy power inequality (4) and the Fisher information inequality (5) can be precisely reversed, up to factors that depend only on the Stam defects associated to $X$ and $Y$. In particular, if both $X$ and $Y$ each nearly saturate Stam’s inequality (11), then the EPI and FII will also be nearly saturated. Notably, strong regularity assumptions are not imposed.

**Theorem 2.** Let $X, Y$ be independent random vectors on $\mathbb{R}^d$ with finite second moment, and choose $\lambda$ to satisfy $\lambda/(1 - \lambda) = N(Y)/N(X)$. Then

$$N(X + Y) \leq (N(X) + N(Y)) \left( \lambda p(X) + (1 - \lambda) p(Y) \right).$$

$$N(X + Y)^\kappa \leq N(X)^\kappa + N(Y)^\kappa.$$
Furthermore, if $J(X), J(Y) < \infty$, then
\[
\frac{1}{J(X+Y)} \leq \left( \frac{1}{J(X)} + \frac{1}{J(Y)} \right) p(X)p(Y). \tag{28}
\]

Since (27) does not immediately resemble either of the reverse inequalities mentioned above, let us briefly comment on how it may be understood in the context of known results. In particular, it is well known that entropy power is concave under the action of the heat semigroup [12–14]. That is, if $G \sim N(0, I)$, then
\[
\frac{d^2}{dt^2} N(X + \sqrt{t}G) \leq 0, \tag{29}
\]
which is the same as $N(X + \sqrt{t}G)$ lying below its tangents lines. By the semigroup property, it suffices to consider the tangent line at $t = 0$, so (29) is equivalent to
\[
N(X + \sqrt{t}G) \leq N(X) + t \left( \frac{d}{dt} N(X + \sqrt{t}G) \right)_{t=0}
= N(X) + t p(X), \tag{30}
\]
where the equality follows by de Bruijn’s identity (e.g., [1, 15, 16]). This coincides exactly with (27) particularized to the case where $Y = \sqrt{t}G$. Thus, we may think of (27) as a generalization of ‘concavity of entropy power’ beyond the heat semigroup.

Inequality (27) may also be viewed as a strengthening of Stam’s uncertainty principle (11). Indeed, letting $X, X^*$ be IID, (27) reduces to
\[
p(X) \geq \frac{N(\frac{1}{\sqrt{2}}(X + X^*)))}{N(X)}. \tag{32}
\]
By the EPI, the term on the RHS is strictly greater than 1 unless $X$ is Gaussian. On this note, we mention that Carlen and Soffer [16] have shown that for $X$ centered with $\text{Cov}(X) = I$, there exists a nonnegative function $\Theta$ on $[0, \infty)$, strictly increasing from 0 and depending only on certain decay and smoothness properties of $X$ such that
\[
\frac{N(\frac{1}{\sqrt{2}}(X + X^*)))}{N(X)} \geq \exp \left( \frac{2}{d} \Theta(D(X)) \right). \tag{33}
\]

Finally, we note that an equivalent version of Corollary 1 for the Stam defect follows directly from (27) and the FII:

**Corollary 2** (Convolution inequality for the Stam defect). For $X, Y$ be independent random vectors on $\mathbb{R}^d$ with finite second moment,
\[
p(X + Y) \leq p(X)p(Y). \tag{34}
\]

The proof of Theorem 2 follows rather directly from a strengthening of the entropy power inequality proved recently by the author. To state it, we first recall some notation familiar to information theorists: Let $U, V$ have joint distribution $P_{UV}$ on the space $U \times V$ and let the respective marginals be denoted by $P_U, P_V$. The mutual information $I(U; V)$ between $U$ and $V$ is given by
\[
I(U; V) := \mathbb{E} \log \left( \frac{dP_{UV}}{dP_U \times P_V} \right) = \int_{U \times V} \log \left( \frac{dP_{UV}}{dP_U \times P_V} \right) dP_{UV}. \tag{35}
\]
Theorem 3. [17] Let $X, W$ be independent random vectors on $\mathbb{R}^d$, with $W$ Gaussian. Define $Z = X + W$. For any random variable $V$ such that $X, V$ are conditionally independent given $Z$, it holds that

$$e^{-\frac{2}{d}I(X;V)}N(Z) \geq e^{-\frac{2}{d}I(Z;V)}N(X) + N(W).$$

(36)

Proof of Theorem 2. Let $Y$ be any random vector on $\mathbb{R}^d$, independent of $X, W$, and suppose $X, Y$ have finite second moments. Then, $V = Z + Y$ is such that $X, V$ are conditionally independent given $Z$. By the definition of mutual information,

$$I(X;V) = h(X + Y + W) - h(Y + W)$$

(37)

and

$$I(Z;V) = h(X + Y + W) - h(Y).$$

(38)

Thus, rearranging exponents in (36) gives the following

$$N(X + W)N(Y + W) \geq N(X)N(Y) + N(X + Y + W)N(W).$$

(39)

Now, let $W = \sqrt{t}G$, where $G \sim N(0, I)$, in which case (39) particularizes to

$$N(X + \sqrt{t}G)N(Y + \sqrt{t}G) - N(X)N(Y) \geq N(X + Y + \sqrt{t}G) \geq N(X + Y).$$

(40)

An application of de Bruijn’s identity $\frac{d}{dt}N(X + \sqrt{t}G)\bigg|_{t=0} = p(X)$ and the chain rule for derivatives proves

$$N(X + Y) \leq N(X)p(Y) + N(Y)p(X),$$

(41)

which is the same as (27). Stam’s inequality $p(X + Y) \geq 1$ and algebra (valid when $J(X), J(Y) < \infty$) shows that (28) is a corollary of (27).

Theorem 1 now follows from Theorem 2. In fact:

Proposition 1. Theorems 1 and 2 are equivalent.

Proof. For convenience, we recall the scaling properties $N(tZ) = t^2N(Z)$ and $t^2J(tZ) = J(Z)$. Also, if $G_s \sim N(0, sI)$, the relative entropy $D(Z\|G_s)$ and Fisher information $I(Z\|G_s)$ are related to $h(Z)$ and $J(Z)$ via

$$h(Z) - \frac{d}{2} \log(2\pi e s) = -D(Z\|G_s) + \frac{1}{2s} \mathbb{E}|Z|^2 - \frac{d}{2}$$

(42)

and

$$J(Z) = I(Z\|G_s) + \frac{2}{s}d - \frac{1}{s^2} \mathbb{E}|Z|^2,$$

(43)

holding for any random vector $Z$ on $\mathbb{R}^d$ with $\mathbb{E}|Z|^2 < \infty$. Further, $N(\cdot), J(\cdot)$ are translation invariant, so we may assume without loss of generality that all random vectors are centered.

• Proof of (27) $\Rightarrow$ (12): We assume $I(X), I(Y) < \infty$, else (12) is a tautology. Now, finiteness of $I(X)$ implies $\mathbb{E}|X|^2 < \infty$, and similarly for $Y$ (see, e.g., [4, Proof of Thm. 5]). Using the scaling properties of $N(\cdot)$ and $J(\cdot)$, (27) implies

$$N\left(\sqrt{\theta}X + \sqrt{\theta}Y\right) \leq N(X)N(Y) \left(\frac{\theta J(X) + \theta J(Y)}{d}\right),$$

(44)
Now, taking logarithms, multiplying through by \( d/2 \) and recalling \( \log x \leq x - 1 \), we have:

\[
\begin{align*}
  h \left( \sqrt{\theta}X + \sqrt{\theta}Y \right) &- \frac{d}{2} \log(2\pi e) \\
  &\leq h(X) - \frac{d}{2} \log(2\pi e) + h(Y) - \frac{d}{2} \log(2\pi e) \\
  &\quad + \frac{d}{2} \log \left( \frac{\theta J(X) + \theta J(Y)}{d} \right) \\
  &\leq h(X) - \frac{d}{2} \log(2\pi e) + h(Y) - \frac{d}{2} \log(2\pi e) \\
  &\quad + \frac{1}{2} \left( \theta J(X) + \theta J(Y) \right) - \frac{d}{2} 
\end{align*}
\]

Now, \((12)\) follows from the identities \((42)\) and \((43)\) for \( s = 1 \).

- **Proof of \((12) \Rightarrow (27)\)**: We may assume \( X, Y \) have finite Fisher information and second moments. With this assumption in place, consider any \( s > 0 \) and observe via straightforward manipulation using the identities \((42)-(43)\) that \((12)\) is equivalent to

\[
D(X\|G_s) + D(Y\|G_s) \leq s \left( \frac{\theta}{2} I(X\|G_s) + \frac{\theta}{2} I(Y\|G_s) \right) + D\left( \sqrt{\theta}X + \sqrt{\theta}Y\|G_s \right). 
\]

Hence, using \((42)-(43)\) again and rearranging, we find that this is the same as

\[
\begin{align*}
  \log N \left( \sqrt{\theta}X + \sqrt{\theta}Y \right) &\leq \log N(X)N(Y) + \frac{\theta}{2} J(X) + \frac{\theta}{2} J(Y) - \log s - 1. 
\end{align*}
\]

Recalling \( 1 + \log a = \inf_{s>0} (as - \log s) \), we may minimize the RHS over \( s > 0 \) to obtain

\[
\begin{align*}
  dN \left( \sqrt{\theta}X + \sqrt{\theta}Y \right) &\leq N(X)N(Y) \left( \theta J(X) + \theta J(Y) \right) \\
  &\quad = N(\sqrt{\theta}X)N(\sqrt{\theta}Y) \left( J(\sqrt{\theta}X) + J(\sqrt{\theta}Y) \right), 
\end{align*}
\]

where the last equality follows via the scaling properties of \( N(\cdot) \) and \( J(\cdot) \). A simple rescaling recovers \((27)\) and completes the proof.

### 3 Applications

#### 3.1 Short-term convergence rates in information-theoretic CLTs

Let \( Z \) be a centered random vector on \( \mathbb{R}^d \) with \( \text{Cov}(Z) = I \), and define the normalized sums \( U_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n Z_k \), where \( Z_1, Z_2, \ldots, Z_n \) are independent copies of \( Z \). The entropic central limit theorem due to Barron [18] asserts that \( D(U_n) \to 0 \), provided \( D(U_{n_0}) < \infty \) for some \( n_0 \). Likewise, the CLT for Fisher information, due to Barron and Johnson [19], asserts that \( I(U_n) \to 0 \), provided \( I(U_{n_0}) < \infty \) for some \( n_0 \).

In 2004, Artstein, Ball, Barthe and Naor established that each of these limit theorems enjoy monotone convergence [20]. However, optimal estimates on the convergence rate remained open until recently. On this front, Bobkov, Chistyakov, Gennadiy and Götze [21, 22] have settled a longstanding conjecture and shown that under moment conditions

\[
D(U_n) = O(1/n),
\]

(51)
which is consistent with the convergence rates predicted by the Berry-Esseen theorem. Although explicit constants are given for the $O(1/n)$ term as a function of the moments and $D(Z)$, the proof invokes local limit theorems for Edgeworth expansions, so the $o(1/n)$ terms are not explicitly quantified for finite $n$. Hence, although (51) provides good long-term estimates on convergence in the entropic CLT, it does not immediately provide any information about the short-term behavior of $D(U_n)$.

The next result partially addresses this issue by establishing a lower bound on $D(U_n)$ in terms of $\delta_{\text{LSI}}(Z)$ and $n$. Roughly speaking, if $\delta_{\text{LSI}}(Z) \ll D(Z)$, then $D(U_n)$ is assured to decay slowly on short time scales. A similar result holds for Fisher information. That is, if $\delta_{\text{LSI}}(Z) \ll I(Z)$, then $I(U_n)$ will decay slowly on short time scales. More precisely, each of these quantities decay at most linearly in $n$, with slope $\delta_{\text{LSI}}(Z)$.

**Theorem 4.** Let $Z$ be a centered random vector on $\mathbb{R}^d$ and define the normalized sums $U_n = \sqrt{n} \sum_{k=1}^n Z_k$, where $Z_1, Z_2, \ldots, Z_n$ are independent copies of $Z$. The sequence $\{\delta_{\text{LSI}}(U_n), n \geq 1\}$ is subadditive. Moreover, the following hold for all $n \geq 1$:

\[
D(U_n) \geq D(Z) - (n-1)\delta_{\text{LSI}}(Z) \tag{52}
\]

\[
\frac{1}{2} I(U_n) \geq \frac{1}{2} I(Z) - n \delta_{\text{LSI}}(Z) + \delta_{\text{LSI}}(U_n). \tag{53}
\]

**Proof.** We apply Theorem 1 with $\theta = \frac{n}{n+m}$, $X = U_n$ and $Y = U_m$. In this case, $\sqrt{\theta}X + \sqrt{\bar{\theta}}Y$ is equal to $U_{n+m}$ in distribution, so we obtain the inequality

\[
D(U_{n+m}) \geq D(U_n) + D(U_m) - \frac{m}{2(m+n)} I(U_n) - \frac{n}{2(m+n)} I(U_m), \tag{54}
\]

or, equivalently

\[
\delta_{\text{LSI}}(U_m) + \delta_{\text{LSI}}(U_n) \geq \delta_{\text{LSI}}(U_{m+n}) + \frac{n}{2(m+n)} I(U_n) + \frac{m}{2(m+n)} I(U_m) - \frac{1}{2} I(U_{m+n}). \tag{55}
\]

Subadditivity of $\delta_{\text{LSI}}(U_n)$ follows from applying (9).

Now, the proof of (52) and (53) will follow by induction on $n+m$. We first prove (52). The base case for $n+m = 2$ is immediate from (54) with $n = m = 1$. So, by induction, we have

\[
D(U_{n+m}) \geq D(Z) - (n-1)\delta_{\text{LSI}}(Z) + D(Z) - (m-1)\delta_{\text{LSI}}(Z) \tag{56}
\]

\[
- \frac{1}{2} \left( \frac{m}{m+n} I(U_n) + \frac{n}{m+n} I(U_m) \right)
\]

\[
= D(Z) - (n+m-1)\delta_{\text{LSI}}(Z) + \frac{1}{2} I(Z) - \frac{1}{2} \left( \frac{m}{m+n} I(U_n) + \frac{n}{m+n} I(U_m) \right), \tag{57}
\]

\[
\geq D(Z) - (n+m-1)\delta_{\text{LSI}}(Z), \tag{58}
\]

where the final inequality is due to $I(Z) \geq I(U_m)$ for $m \geq 1$, a consequence of (9).

Now, we aim to prove (53). The base case $n+m = 2$ is immediate from (54) with $n = m = 1$. 

9
Thus, by the inductive hypothesis, we have

\[ (m + n)\delta_{\text{LSI}}(Z) \]
\[ \geq \delta_{\text{LSI}}(U_m) + \frac{1}{2} (I(Z|G) - I(U_m|G)) \]
\[ + \delta_{\text{LSI}}(U_n) + \frac{1}{2} (I(Z|G) - I(U_n|G)) \]
\[ \geq \delta_{\text{LSI}}(U_{m+n}) + I(Z|G) - \left( \frac{m}{2(m+n)} I(U_n|G) + \frac{n}{2(m+n)} I(U_m|G) + \frac{1}{2} I(U_{m+n}|G) \right) \]
\[ \geq \delta_{\text{LSI}}(U_{m+n}) + \frac{1}{2} (I(Z|G) - I(U_{m+n}|G)) , \]

where (60) is (55) and, as before, the final inequality follows due to \( I(Z) \geq I(U_m) \) for \( m \geq 1 \).

3.2 The LSI is stable with respect to HWI jumps

Recall the quadratic Wasserstein distance between random vectors \( X, Y \) is defined according to

\[ W_2^2(X, Y) = \inf_{Q_{XY} \in \pi(X,Y)} \mathbb{E}|X - Y|^2, \]

where the infimum is over all couplings between \( X, Y \) that preserve their given marginals \( X \sim P_X, Y \sim P_Y \). In the case where \( G \sim N(0, I) \), Talagrand’s quadratic transportation cost inequality asserts that

\[ W_2^2(X) := W_2^2(X, G) \leq 2D(X). \]

Talagrand’s inequality is closely related to Gross’ LSI. Indeed, Otto and Villani [23] proved a remarkable inequality that interpolates between (63) and (10):

\[ D(X) \leq \sqrt{I(X)W_2(X)} - \frac{1}{2}W_2^2(X). \]

This inequality is referred to as the HWI inequality, since it simultaneously relates the relative entropy (H), Wasserstein distance (W), and Fisher information (I) functionals.

As with relative entropy and Fisher information, \( W_2^2 \) satisfies a convolution inequality (e.g., [24]). That is, if \( X, Y \) are independent, centered random vectors on \( \mathbb{R}^d \)

\[ W_2^2(\sqrt{\theta}X + \sqrt{\bar{\theta}}Y) \leq \theta W_2^2(X) + \bar{\theta}W_2^2(Y). \]

So, if \( X, X^* \) are centered i.i.d. random vectors on \( \mathbb{R}^d \), then we have the following inequalities:

\[ D(\frac{1}{\sqrt{\theta}}(X + X^*)) \leq D(X) \]
\[ W_2(\frac{1}{\sqrt{\theta}}(X + X^*)) \leq W_2(X) \]
\[ I(\frac{1}{\sqrt{\theta}}(X + X^*)) \leq I(X). \]

We shall use the term HWI jump to refer to a deficit in any of the three inequalities above. An immediate consequence of Theorem 1 and (64) is that the LSI is stable with respect to HWI jumps. That is, if \( X \) exhibits a jump under convolution with respect to any of the HWI functionals, then we can plainly lower bound the deficit in the LSI in terms of \( I(X) \).
Theorem 5. Let \(X, X^*\) be centered i.i.d. random vectors on \(\mathbb{R}^d\). If any of the following hold:

(i) \(D(\frac{1}{\sqrt{2}}(X + X_*)) \leq (1 - \varepsilon)D(X),\) or

(ii) \(W_2(\frac{1}{\sqrt{2}}(X + X_*)) \leq (1 - \varepsilon^{1/2})W_2(X),\) or

(iii) \(I(\frac{1}{\sqrt{2}}(X + X_*)) \leq (1 - \varepsilon)I(X),\)

then \(\delta_{\text{LSI}}(X) \geq \frac{\varepsilon}{4}I(X).\)

Before proving Theorem 5, we remark that there have been a number of recent attempts to give quantitative stability estimates for the LSI. However, such stability estimates are generally dimension dependent \([25]\) or impose strong regularity conditions such as presence of a spectral gap \([26, 27]\). In contrast, Theorem 5 shows that HWI jumps give a dimension-free estimate of the deficit in the LSI that holds without stringent regularity assumptions. However, HWI jumps do not directly bound the distance of \(X\) from normal except under regularity conditions such as presence of a spectral gap \((d = 1)\) \([28]\) and log-concave density \((d \geq 2)\) \([29]\), or a radial symmetry assumption \((d \geq 2)\) \([30]\). Of course, one cannot hope that \(\delta_{\text{LSI}}(X) \geq \varepsilon I(X)\) in general for some absolute constant \(\varepsilon\), else it would contradict optimality of the constant in Gross’ LSI.

Proof of Theorem 5. The proof is an immediate consequence of Theorem 1 and the HWI inequality, but we include it for completeness. In all cases, we shall apply Theorem 1 with \(Y = X_*\) and \(\theta = 1/2\). In view of this, if (i) holds, then Theorem 1 implies

\[
D(X) \leq \frac{1}{2(1 + \varepsilon)}I(X) = \frac{1}{2(1 + \varepsilon)}I(X) \leq \frac{1}{2}I(X) - \frac{\varepsilon}{4}I(X). \tag{69}
\]

Next, if (ii) holds, then using the HWI inequality and the convolution inequality for Fisher information, we have:

\[
2D(X) \leq \frac{1}{2}I(X) + D\left(\frac{1}{\sqrt{2}}(X + X_*)\right) \tag{70}
\]

\[
\leq \frac{1}{2}I(X) + \sqrt{I(\frac{1}{\sqrt{2}}(X + X_*))}W_2(\frac{1}{\sqrt{2}}(X + X_*)) - \frac{1}{2}W_2^2(\frac{1}{\sqrt{2}}(X + X_*)) \tag{71}
\]

\[
\leq \frac{1}{2}I(X) + \sqrt{I(X)}W_2(\frac{1}{\sqrt{2}}(X + X_*)) - \frac{1}{2}W_2^2(\frac{1}{\sqrt{2}}(X + X_*)). \tag{72}
\]

Equivalently,

\[
\frac{1}{2}I(X) \leq 2\delta_{\text{LSI}}(X) + \sqrt{I(X)}W_2(\frac{1}{\sqrt{2}}(X + X_*)) - \frac{1}{2}W_2^2(\frac{1}{\sqrt{2}}(X + X_*)). \tag{73}
\]

By the conjunction of Talagrand’s inequality and the LSI, (ii) and the quadratic formula, we can conclude:

\[
(1 - \varepsilon^{1/2})^2I(X) \geq (1 - \varepsilon^{1/2})^2\theta W_2^2(X) \tag{74}
\]

\[
\geq W_2^2(\frac{1}{\sqrt{2}}(X + X_*)) \tag{75}
\]

\[
\geq \left(\sqrt{I(X)} - 2\sqrt{\delta_{\text{LSI}}(X)}\right)^2. \tag{76}
\]

Taking square roots, rearranging and squaring again, we obtain the desired inequality.

Finally, if (iii) holds, then the claim is immediate from (14).
3.3 A sharp form of Nelson’s hypercontractivity estimate

It is well known that Gross’ LSI is equivalent to Nelson’s hypercontractivity estimate for the Ornstein-Uhlenbeck semigroup [3]. To state Nelson’s result, let us first introduce the Ornstein-Uhlenbeck semigroup $(P_t)_{t \geq 0}$ defined on functions $f : \mathbb{R}^d \to \mathbb{R}$ as follows:

$$P_t f(x) = \int_{\mathbb{R}^d} f(e^{-t}x + (1 - e^{-2t})^{1/2} y) \, d\gamma(y),$$

(77)

where $\gamma$ denotes the standard Gaussian measure on $\mathbb{R}^d$. Nelson’s result is as follows:

**Theorem 6.** [31] For $f \in L^p(\gamma)$

$$\|P_t f\|_{L^q(\gamma)} \leq \|f\|_{L^p(\gamma)}$$

(78)

for all $q \geq p > 1$ such that $q \leq 1 + (p - 1)e^{2t}$.

The essential idea behind Gross’ proof of (78) from the LSI is as follows: Let $q(t) = 1 + (p - 1)e^{2t}$. If $f \geq 0$ is a smooth function, then $\|P_t f\|_{L^{q(t)}(\gamma)}$ is differentiable on $t \in [0, \infty)$ with derivative

$$\frac{d}{dt} \log (\|P_t f\|_{L^{q(t)}(\gamma)}) = \frac{q'(t)}{q^2(t)} D(X_t) - \frac{2q(t) - 1}{q^2(t)} \frac{1}{2} I(X_t),$$

(79)

where $X_t$ is the random variable having density (with respect to $\gamma$) proportional to $(P_t f)^{q(t)}$. On account of the fact that $q'(t) = 2q(t) - 1$, Gross’ LSI implies

$$\frac{d}{dt} \|P_t f\|_{L^{q(t)}(\gamma)} \leq 0.$$  

(80)

Since $\left.\|P_t f\|_{L^{q(t)}(\gamma)}\right|_{t=0} = \|f\|_{L^p(\gamma)}$, Nelson’s result follows for smooth $f$. The extension to $f \in L^p(\gamma)$ follows by a density argument. The reverse implication is also apparent. That is, in order for (78) to hold for smooth $f \geq 0$, we must have $\frac{d}{dt} \|P_t f\|_{L^{q(t)}(\gamma)} \leq 0$ at $t = 0$, which is precisely Gross’ LSI for the random variable $X$ having density (with respect to $\gamma$) proportional to $|f|^p$. Indeed, by the semigroup property, Nelson’s inequality is completely characterized by the behavior of $\|P_t f\|_{L^{q(t)}(\gamma)}$ in a neighborhood of $t = 0$.

Extending this to the setting of Theorem 1 is straightforward. Before we state the extension, let us introduce some notation. For $f, g \in L^p(\gamma)$, let $X$ be the random vector having density (with respect to $\gamma$) proportional to $|f|^p$, and let $Y$ be the random vector having density (with respect to $\gamma$) proportional to $|g|^p$. Further, define their centered counterparts $\tilde{X} = X - \mathbb{E}X$ and $\tilde{Y} = Y - \mathbb{E}Y$, and the associated entropy production functional:

$$E_{p,\theta}(f, g) := \theta D(\tilde{X}) + \bar{\theta} D(\tilde{Y}) - D(\sqrt{\theta} \tilde{X} + \sqrt{\bar{\theta}} \tilde{Y}).$$

(81)

We have the following improvement to Nelson’s result, which interpolates between the hypercontractive estimates for two functions $f, g \in L^p(\gamma)$:

**Theorem 7.** Let $p'$ denote the Hölder conjugate of $p$. For smooth functions $f, g \in L^p(\gamma)$,

$$\|P_t f\|_{L^{q(t)}(\gamma)} \leq \|f\|_{L^p(\gamma)}^\theta E_{p,\theta}(f, g) + o(t)$$

(82)

for all $q \geq p > 1$ and $\theta \geq 0$ such that $q \leq 1 + (p - 1)e^{2t}$.
Proof. We proceed in a manner identical to Gross’ argument outlined above. In particular, we may assume \( f, g \geq 0 \). Then, using the definitions of \( X, Y \) implicit in the definition of \( E_{p, \theta}(f, g) \), we have

\[
\frac{d}{dt} \log \left( \| P_t f \|^\theta_{L^q(t)(\gamma)} \| P_t g \|^\theta_{L^q(t)(\gamma)} \right) \bigg|_{t=0} = -\theta \frac{q'(0)}{q^2(0)} \delta_{LSI}(X) - \theta \frac{q'(0)}{q^2(0)} \delta_{LSI}(Y) \\
\leq - \frac{2(p-1)}{p^2} (\theta D(\hat{X}) + \bar{\theta} D(\hat{Y}) - D(\sqrt{\theta} X + \sqrt{\theta} Y)) \tag{84}
\]

\[
= - \frac{2(p-1)}{p^2} E_{p, \theta}(f, g). \tag{85}
\]

where the inequality follows from Theorem 1, translation invariance of \( \delta_{LSI}(\cdot) \) and definition of \( q(t) \). Thus, since \( \| P_t f \|^\theta_{L^q(t)(\gamma)} \| P_t g \|^\theta_{L^q(t)(\gamma)} \) is differentiable in \( t \in [0, \infty) \) by the smoothness assumption, it follows that

\[
\log \left( \| P_t f \|^\theta_{L^q(t)(\gamma)} \| P_t g \|^\theta_{L^q(t)(\gamma)} \right) \leq \log \left( \| f \|^\theta_{L^p(\gamma)} \| g \|^\theta_{L^p(\gamma)} \right) - t \frac{2(p-1)}{p^2} E_{p, \theta}(f, g) + o(t) \tag{86}
\]

as desired. \( \square \)

Remark 2. The smoothness assumption on \( f, g \) cannot be immediately eliminated in Theorem 7 by a density argument. Indeed, smooth functions may be dense in \( L^p(\gamma) \), but the functional \( E_{p, \theta}(f, g) \) is not necessarily (semi-)continuous in its arguments with respect to the \( L^p(\gamma) \)-norm, since relative entropy is only weakly lower semicontinuous in general.

4 Concluding Remarks

To close, we briefly speculate on the potential for geometric analogues of Theorem 2. The EPI is often compared to the Brunn-Minkowski inequality, which states that for two nonempty compact subsets \( A, B \) of \( \mathbb{R}^d \),

\[
\text{Vol}(A + B)^{1/d} \geq \text{Vol}(A)^{1/d} + \text{Vol}(B)^{1/d}, \tag{87}
\]

where \( A + B \) denotes the Minkowski sum \( A + B := \{ a + b \in \mathbb{R}^d \mid a \in A, \ b \in B \} \). Indeed, if \( X \) is uniformly distributed on \( A \), then \( 2\pi e N(X) = \text{Vol}(A)^{2/d} \). As already seen in Section 2.2, the reverse EPI due to Bobkov and Madiman compares similarly with Milman’s reverse Brunn–Minkowski inequality, which holds when \( A, B \) are convex.

The classical isoperimetric inequality states that, for a sufficiently regular subset \( A \) of \( \mathbb{R}^d \), the surface area \( |\partial A| \) exceeds that of \( B_A \), a ball in \( \mathbb{R}^d \) with the same volume as \( A \). Notably, the isoperimetric inequality can be derived from the Brunn-Minkowski inequality in a manner very similar to how Stam’s inequality (11) follows from the EPI (e.g., [32]). Thus, there is a strong analogy between these information-theoretic inequalities and their geometric counterparts.

In light of this, we are moved to speculate that a geometric analogue to Theorem 1 may hold. For example, it seems reasonable to posit the following for sufficiently regular \( A, B \subset \mathbb{R}^d \):

\[
\text{Vol}(A + B)^{1/d} \leq \left( \text{Vol}(A)^{1/d} + \text{Vol}(B)^{1/d} \right) \left( \lambda \frac{|\partial A|}{|\partial B_A|} + \lambda \frac{|\partial B|}{|\partial B_B|} \right), \tag{88}
\]

\[
13
\]
where λ is chosen to satisfy \( \lambda/(1-\lambda) = \frac{\text{Vol}(B)^{1/d}}{\text{Vol}(A)^{1/d}} \). As of now, we have made no attempt to prove or disprove this inequality.

Finally, analogous to (34), we might expect to find that, for sufficiently regular \( A, B \subset \mathbb{R}^d \)

\[
\frac{|\partial (A + B)|}{|\partial B_{A+B}|} \leq \frac{|\partial A|}{|\partial B_A|} \cdot \frac{|\partial B|}{|\partial B_B|}.
\]

(89)

Again, we have made no attempt to prove or disprove this inequality.

Acknowledgment

This work was supported in part by NSF grants CCF-1528132 and CCF-0939370 (Center for Science of Information).

References

[1] A. Stam, “Some inequalities satisfied by the quantities of information of Fisher and Shannon,” Information and Control, vol. 2, no. 2, pp. 101–112, 1959.
[2] N. M. Blachman, “The convolution inequality for entropy powers,” Information Theory, IEEE Transactions on, vol. 11, no. 2, pp. 267–271, 1965.
[3] L. Gross, “Logarithmic Sobolev inequalities,” American Journal of Mathematics, vol. 97, no. 4, pp. 1061–1083, 1975.
[4] E. A. Carlen, “Superadditivity of Fisher’s information and logarithmic Sobolev inequalities,” Journal of Functional Analysis, vol. 101, no. 1, pp. 194–211, 1991.
[5] M. Talagrand, “Transportation cost for Gaussian and other product measures,” Geometric & Functional Analysis GAFA, vol. 6, no. 3, pp. 587–600, 1996.
[6] M. Ledoux, The concentration of measure phenomenon. American Mathematical Soc., 2005, no. 89.
[7] S. G. Bobkov and G. P. Chistyakov, “Entropy power inequality for the Rényi entropy,” IEEE Transactions on Information Theory, vol. 61, no. 2, pp. 708–714, 2015.
[8] S. Bobkov and M. Madiman, “Reverse Brunn–Minkowski and reverse entropy power inequalities for convex measures,” Journal of Functional Analysis, vol. 262, no. 7, pp. 3309–3339, 2012.
[9] V. D. Milman, “Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espaces normés,” CR Acad. Sci. Paris, vol. 302, no. 1, pp. 25–28, 1986.
[10] M. Madiman, J. Melbourne, and P. Xu, “Forward and reverse entropy power inequalities in convex geometry,” arXiv preprint arXiv:1604.04225, 2016.
[11] K. Ball, P. Nayar, and T. Tkocz, “A reverse entropy power inequality for log-concave random vectors,” arXiv preprint arXiv:1509.05926, 2015.
[12] M. Costa, “A new entropy power inequality,” *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 751–760, 1985.

[13] A. Dembo, “Simple proof of the concavity of the entropy power with respect to added gaussian noise,” *IEEE Transactions on Information Theory*, vol. 35, no. 4, pp. 887–888, 1989.

[14] C. Villani, “A short proof of the ?concavity of entropy power?” *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1695–1696, 2000.

[15] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*. Springer Science & Business Media, 2013, vol. 348.

[16] E. A. Carlen and A. Soffer, “Entropy production by block variable summation and central limit theorems,” *Communications in mathematical physics*, vol. 140, no. 2, pp. 339–371, 1991.

[17] T. A. Courtade, “Strengthening the entropy power inequality,” *arXiv preprint arXiv:1602.03033*, 2016.

[18] A. R. Barron, “Entropy and the central limit theorem,” *The Annals of probability*, pp. 336–342, 1986.

[19] O. Johnson and A. Barron, “Fisher information inequalities and the central limit theorem,” *Probability Theory and Related Fields*, vol. 129, no. 3, pp. 391–409, 2004.

[20] S. Artstein, K. Ball, F. Barthe, and A. Naor, “Solution of shannon?s problem on the monotonicity of entropy,” *Journal of the American Mathematical Society*, vol. 17, no. 4, pp. 975–982, 2004.

[21] S. G. Bobkov, G. P. Chistyakov, and F. Götze, “Rate of convergence and edgeworth-type expansion in the entropic central limit theorem,” *The Annals of Probability*, vol. 41, no. 4, pp. 2479–2512, 2013.

[22] ——, “Berry-esseen bounds in the entropic central limit theorem,” *Probability Theory and Related Fields*, vol. 159, no. 3-4, p. 435, 2014.

[23] F. Otto and C. Villani, “Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality,” *Journal of Functional Analysis*, vol. 173, no. 2, pp. 361–400, 2000.

[24] C. Villani, *Topics in optimal transportation*. American Mathematical Soc., 2003, no. 58.

[25] S. G. Bobkov, N. Gozlan, C. Roberto, and P.-M. Samson, “Bounds on the deficit in the logarithmic Sobolev inequality,” *Journal of Functional Analysis*, vol. 267, no. 11, pp. 4110–4138, 2014.

[26] M. Fathi, E. Indrei, and M. Ledoux, “Quantitative logarithmic Sobolev inequalities and stability estimates,” *arXiv preprint arXiv:1410.6922*, 2014.

[27] E. Indrei and D. Marcon, “A quantitative log-Sobolev inequality for a two parameter family of functions,” *International Mathematics Research Notices*, p. rnt138, 2013.
[28] K. Ball, F. Barthe, and A. Naor, “Entropy jumps in the presence of a spectral gap,” Duke Mathematical Journal, vol. 119, no. 1, pp. 41–63, 2003.

[29] K. Ball and V. H. Nguyen, “Entropy jumps for isotropic log-concave random vectors and spectral gap,” Studia Mathematica, vol. 213, no. 1, pp. 81–96, 2012.

[30] T. A. Courtade, “Entropy jumps for radially symmetric random vectors,” preprint, 2016.

[31] E. Nelson, “The free Markoff field,” Journal of Functional Analysis, vol. 12, no. 2, pp. 211–227, 1973.

[32] M. Costa and T. Cover, “On the similarity of the entropy power inequality and the Brunn-Minkowski inequality (corresp.),” IEEE Transactions on Information Theory, vol. 30, no. 6, pp. 837–839, 1984.