A quantitative entropic CLT for radially symmetric random vectors

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

***Abstract*—A quantitative entropic central limit theorem is established for the sum of i.i.d. radially symmetric random vectors having dimension greater than one. In contrast to recent related work, strong regularity assumptions – such as positive spectral gap or log-concavity of densities – are not needed. However, this added flexibility comes at the expense of an assumption of radial symmetry.**

I. Introduction

Let X be a random vector on $\mathbb{R}^d$ with density f. The entropy associated with X is defined by

$$h(X) = -\int_{\mathbb{R}^d} f \log f, \tag{1}$$

provided the integral exists in the Lebesgue sense. The non-Gaussianness of X, denoted by $D(X)$, is defined as

$$D(X) = h(G^X) - h(X), \tag{2}$$

where G^X denotes a Gaussian random vector with the same covariance as X. Evidently, $D(X)$ is the relative entropy of X with respect to G^X, and is therefore nonnegative, with $D(X) = 0$ if and only if X is Gaussian.

In this paper, we investigate forms of the entropic central limit theorem for radially symmetric random vectors of arbitrary dimension $d \geq 2$. To this end, a random vector with density f on $\mathbb{R}^d$ is said to be radially symmetric if $f(x) = \phi(|x|)$ for some function $\phi : [0, \infty) \to [0, \infty]$, where $|\cdot|$ denotes Euclidean length. Radial symmetry arises naturally in many physical settings (e.g., particle velocities) and data science applications (e.g., random projections), and often also appears in characterizing extremal distributions in diverse situations where various symmetrization techniques are applicable (e.g., symmetric decreasing rearrangements, for which [1] is particularly relevant). As such, radial symmetry defines an important class of probability distributions that enjoys many applications, and also has explicit structure to facilitate analysis.

Our first main result pertains to 'entropy jumps' under rescaled convolution. Informally, it may be stated as follows: Let X, X_* be i.i.d. radially symmetric random vectors on $\mathbb{R}^d$, $d \geq 2$, with sufficiently regular density f. For any $\varepsilon > 0$

$$D(X) - D\left(\tfrac{1}{\sqrt{2}}(X + X_*)\right) \geq C_\varepsilon(f) \frac{d^\varepsilon\, \mathbb{E}|X|^2}{\||X|^2\|_{2+1/\varepsilon}^{1+2\varepsilon}} D(X)^{1+\varepsilon}, \tag{3}$$

where $C_\varepsilon(f) > 0$ is an explicit function depending only on ε and the regularity of f, and $\|\cdot\|_p$ denotes the usual L^p norm for random variables. In particular, (3) produces a nontrivial bound as soon as $\mathbb{E}|X|^{4+\delta} < \infty$ for some $\delta > 0$. Moreover, (3) has desired dependence on dimension d for all $\varepsilon > 0$, as discussed in Section III.

Although a radially symmetric density f has a one-dimensional parameterization in terms of its density on shells of radius $r \geq 0$, the convolution $f * f$ is inherently a d-dimensional operation unless f is Gaussian. As such, it does not appear that (3) can be easily reduced to a one-dimensional problem.

The quantity on the LHS of (3) is often referred to as the entropy jump corresponding to X, and has been the topic of several investigations: Ball, Barthe and Naor [2] and Johnson and Barron [3] independently established a bound on the entropy jump in dimension one, under the assumption that X satisfies a Poincaré inequality. Subsequently, Ball and Nguyen [4] bounded the entropy jump in general dimension under the additional assumption of log-concavity. More recently, Courtade, Fathi and Pananjady found a stability estimate for the EPI for uniformly log-concave distributions, which includes the entropy jump as a special case [5]. Although no significant symmetry assumptions are imposed in any of these previous works, all assume that the distribution in question satisfies a Poincaré inequality. Among other things, existence of a Poincaré inequality implies finite moments of all orders. In contrast, (3) assumes much weaker regularity, essentially consisting of smoothness of the density f and $|X|$ having finite $4 + \delta$ moment for some $\delta > 0$. That said, our weaker regularity assumptions come at the price of a radial symmetry assumption which is crucial in our analysis.

By iterating (3), we obtain a quantitative entropic CLT for radially symmetric random vectors, to be stated in Section IV. Barron's entropic CLT ensures that the non-Gaussianness of sums of i.i.d. random variables with finite variance tends to zero, provided that the entropy is ever finite [6]. However, establishing explicit quantitative bounds on this behavior continues to be an active area of research; a brief sampling of recent results is as follows. Under moment conditions, Bobkov, Chistyakov and Götze [7], [8] establish that entropy along the CLT decays proportionally to number of convolutions, which is consistent with classical Berry-Esseen estimates for the Kolmogorov-Smirnov distance. Except for dimension one, these bounds only establish the asymptotic rate of entropy decay, and do not provide quantitative estimates. Using a different approach based on Stein kernels, Ledoux, Nourdin and Peccati [9] establish near-optimal bounds on entropy decay. However, these results only apply when a Stein kernel exists; a sufficient condition is existence of a Poincaré-type

Email: courtade@berkeley.edu. Research supported in part by NSF grants CCF-1704967, CCF-0939370 and CCF-1750430.

978-1-5386-4780-6/18/$31.00©2018 IEEE

inequality [10]. When convergence is measured under the weaker quadratic Wasserstein-Kantorovich distance, Zhai [11], Bonis [12] and Courtade, Fathi and Pananjady [10] establish various CLTs, each with their own advantages and disadvantages. See [10] for a discussion and other related results.

Our approach for establishing (3) differs from each of the previously mentioned works, but traces its roots to related problem in statistical physics. In particular, (3) most closely parallels bounds on entropy production in the Boltzmann equation (see the review [13] for an overview). Indeed, it has been observed at least since the early 90s (cf. [14], [15]) that there is a strong analogy between entropy production in the Boltzmann equation and entropy jumps associated to rescaled convolution. The details of the former are tangential to the present discussion, but a major milestone was achieved when the entropy production in the Boltzmann equation was bounded from below by an explicit function of $D(X)$ and various norms of X, where X models the velocity of a particle in a rarified gas [16], [17]. A key ingredient used to prove this bound was an earlier result by Desvillettes and Villani that controls relative Fisher information $I(X\|G)$ via entropy production in the Landau equation:

Lemma 1. *[18] Let X be a random vector on $\mathbb{R}^d$, satisfying $\mathbb{E}|X|^2 = d \geq 2$ and having density f. Then,*

$$\iint |x-x_*|^2 f(x)f(x_*)\left|\Pi(x-x_*)\left[\frac{\nabla f}{f}(x)-\frac{\nabla f}{f}(x_*)\right]\right|^2 \mathrm{d}x\,\mathrm{d}x_* \geq 2\lambda\,(d-1)I(X\|G), \tag{4}$$

where λ is the minimum eigenvalue of the covariance matrix associated to X, and $\Pi(v)$ is the orthogonal projection onto the subspace orthogonal to $v \in \mathbb{R}^d$.

In fact, our proof of (3) will follow a program similar to [16], [17]. Our starting point is to recognize that the LHS of (4) resembles dissipation of Fisher information when written in the context of L^2 projections (cf. [3, Lemma 3.1]). Using the radial symmetry assumption, we are able to bound the Fisher information dissipation from below by error terms plus entropy production in the Landau equation, which is subsequently bounded by relative Fisher information using Lemma 1. Care must be exercised in order to control error terms, but the final result (3) closely parallels that proved in [16] for the Boltzmann equation. We remark that the assumption of a non-vanishing Boltzmann collision kernel in [16] has a symmetrizing effect on the particle density functions involved; the rough analog in the present paper is the radial symmetry assumption.

The rest of this paper is organized as follows. Section II introduces notation and definitions that are used throughout. Section III is the technical core of the paper, wherein bounds on entropy jumps are established. These are used to derive a quantitative central limit theorem in Section IV.

II. Preliminaries

For a vector $v \in \mathbb{R}^d$, we let $|v| := (\sum_{i=1}^d v_i^2)^{1/2}$ denote its Euclidean norm. For a random variable X on $\mathbb{R}$ and $p \geq 1$, we write $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$ for the usual L^p-norm of X. It will be convenient to use the same notation for $0 < p < 1$, with the understanding that $\|\cdot\|_p$ is not a norm in this case.

Throughout, $G \sim N(0, \mathrm{I})$ denotes a standard Gaussian random vector on $\mathbb{R}^d$; the dimension will be clear from context. For $d \geq 2$, we denote the coordinates of a random vector X on $\mathbb{R}^d$ as $(X_1, X_2, \dots, X_d)$.

For a random vector X with sufficiently smooth density f, we define the Fisher information

$$J(X) = 4\int \left|\nabla\sqrt{f}\right|^2 = \int_{f>0} \frac{|\nabla f|^2}{f}. \tag{5}$$

For random vectors X, Q with respective densities f, g, the relative Fisher information is defined by

$$I(X\|Q) = 4\int g\left|\nabla\sqrt{f/g}\right|^2$$

and the relative entropy is defined by

$$D(X\|Q) = \int f\log\frac{f}{g},$$

where 'log' denotes the natural logarithm throughout. Evidently, both quantities are nonnegative and, in the case of radially symmetric random vectors,

$$I(X) := I(X\|G^X) = J(X) - J(G^X)$$
$$D(X) := D(X\|G^X) = h(G^X) - h(X).$$

Finally, we recall two basic inequalities that will be taken for granted without explicit reference: for real-valued a, b we have $(a+b)^2 \leq 2a^2 + 2b^2$, and for random variables X, Y, we have Minkowski's inequality: $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$ when $p \geq 1$.

In the intermediate steps of our development, we will need to impose certain regularity conditions on densities. In particular, we need pointwise control of $|\nabla \log f(x)|$ in terms of $|x|$.

Definition 1. *A random vector X on $\mathbb{R}^d$ with smooth density f is c-regular if, for all $x \in \mathbb{R}^d$,*

$$|\nabla \log f(x)| \leq c\,(|x| + \mathbb{E}|X|)\,. \tag{6}$$

We remark that the smoothness requirement of f in the definition of c-regularity is stronger than generally required for our purposes. However, it allows us to avoid further qualifications; for instance, the identities (5) hold for any c-regular function. Moreover, since $\nabla\log f = \frac{\nabla f}{f}$ for smooth f, we have $J(X) < \infty$ for any c-regular X with $\mathbb{E}|X|^2 < \infty$. Evidently, c-regularity quantifies the smoothness of a density function. The following shows that any density can be mollified to make it c-regular.

Proposition 1. *[19] Let X and Z be independent, where $Z \sim N(0, \sigma^2\mathrm{I})$ and $\mathbb{E}|X| < \infty$. Then $Y = X + Z$ is c-regular with $c = 4/\sigma^2$.*

Observe that, in the notation of the above proposition, if X is radially symmetric then so is Y. Therefore, Proposition 1 provides a convenient means to construct radially symmetric random vectors that are c-regular. The following estimate will be needed in our development. Its proof follows similarly to [19, Proposition 2], and is omitted.

Proposition 2. *Let X and Z be independent random vectors on $\mathbb{R}^d$, where $Z \sim N(0,\mathrm{I})$ and $\mathbb{E}|X| < \infty$. Define $X_t = e^{-t}X + (1-e^{-2t})^{1/2}Z$ for $t \geq 0$. If X is c-regular, then X_t is $(5c\,e^{2t})$-regular.*

Proposition 1 ensures that X_t will be $4/(1-e^{-2t})$-regular, which yields a poor estimate for small t. However, Proposition 2 affords control over regularity in this regime.

III. Bounds on Entropy Jumps

Here, we establish quantitative estimates on the entropy jump associated with a radially symmetric random vector X in dimension $d \geq 2$. As can be expected, we begin with an inequality for Fisher information, and then obtain a corresponding entropy jump inequality by integrating and applying de Bruijn's identity. To lighten notation, we let C_ε denote a positive constant that depends only on ε, whose value may change line to line.

Our main result of this section is as follows:

Theorem 1. *Let X, X_* be i.i.d. radially symmetric random vectors on $\mathbb{R}^d$, $d \geq 2$, with $I(X) < \infty$. For any $\varepsilon > 0$*

$$D(X) - D\left(\tfrac{1}{\sqrt{2}}(X+X_*)\right) \geq C_\varepsilon \frac{d^\varepsilon\, \mathbb{E}|X|^2}{\||X|^2\|_{2+1/\varepsilon}^{1+2\varepsilon}} \frac{D(X)^{1+3\varepsilon}}{I(X)^{2\varepsilon}}. \tag{7}$$

We remark that, although the RHS of (7) appears to grow favorably with dimension d, the additional factor of d^ε actually gives appropriate dependence on dimension since $\mathbb{E}|X|^2, \||X|^2\|_{2+1/\varepsilon}, D(X)$ and $I(X)$ all typically scale linearly with d. As such, both sides of (7) will generally scale linearly with d. This is consistent with the oft-desired 'dimension-free' property of information inequalities, meaning that they are additive on product distributions. However, this usual notion of dimension-free behavior isn't directly applicable to our setting, since the only product measure that is radially symmetric is Gaussian.

The remainder of this section is dedicated to the proof of Theorem 1. We start with an analogous inequality for Fisher information, which holds under the additional assumption of c-regularity. This inequality will be integrated to obtain an entropic inequality, from which the c-regularity assumption will eventually be eliminated.

A. Fisher Information and Rescaled Convolution

Theorem 2. *Let X, X_* be i.i.d. radially symmetric random vectors on $\mathbb{R}^d$, $d \geq 2$, with c-regular density f. For any $\varepsilon > 0$*

$$J(X) - J\left(\tfrac{1}{\sqrt{2}}(X+X_*)\right) \geq C_\varepsilon \frac{(\mathbb{E}|X|^2)^{1+\varepsilon}}{c^{2\varepsilon}\, \||X|^2\|_{2+1/\varepsilon}^{1+2\varepsilon}} I(X)^{1+\varepsilon}. \tag{8}$$

Proof of Theorem 2. Begin by observing that inequality (8) is invariant to scaling $t : X \mapsto tX$ for $t > 0$. Indeed, if X is c-regular, then a change of variables shows that tX is (c/t^2)-regular. So, using homogeneity of the norms and canceling terms, we find that RHS of (8) is homogeneous of degree -2. The same holds for the LHS, i.e., $J(tX) = t^{-2}J(X)$. Hence, there is no loss of generality in assuming that X is normalized according to $\mathbb{E}|X|^2 = d$. Also, since X is radially symmetric, $X - X_*$ is equal to $X + X_*$ in distribution, therefore we seek to lower bound the quantity

$$J(X) - J(\tfrac{1}{\sqrt{2}}(X+X_*)) = J(X) - 2J(X - X_*).$$

Toward this end, define the sum $W = X - X_*$, and denote its density by f_W. By the projection property of the score function of sums of independent random variables (e.g., [20, Lemma 3.4]):

$$2\left(J(X) - 2J(X-X_*)\right) = \mathbb{E}\left|2\rho_W(W) - (\rho(X) - \rho(X_*))\right|^2,$$

where $\rho := \nabla \log f$ is the score function of X and $\rho_W := \nabla \log f_W$ is the score function of W.

For $v \in \mathbb{R}^d$, let $\Pi(v)$ denote the orthogonal projection onto the subspace orthogonal to v. Now, we have

$$\begin{aligned}
&2J(X) - 4J(X-X_*)\\
&= \mathbb{E}\left|2\rho_W(W) - (\rho(X) - \rho(X_*))\right|^2\\
&\geq \mathbb{E}\left|2\Pi(W)\rho_W(W) - \Pi(X-X_*)\left(\rho(X) - \rho(X_*)\right)\right|^2\\
&= \mathbb{E}\left|\Pi(X-X_*)\left(\rho(X) - \rho(X_*)\right)\right|^2.
\end{aligned} \tag{9}$$

The inequality follows since $\Pi(w) = \Pi(x - x_*)$ by definition, and $|v| \geq |\Pi(w)v|$ since $\Pi(w)$ is an orthogonal projection. The last equality follows since $\Pi(w)\rho_W(w) = 0$ due to the fact that $\Pi(w)\nabla f_W(w)$ is the tangential gradient of f_W, which is identically zero due to radial symmetry of f_W.

Next, for any $R > 0$, use the inequality $1 \geq \frac{|x-x_*|^2}{R^2} - \frac{|x-x_*|^2}{R^2}\mathbf{1}_{\{|x-x_*|>R\}}$ to conclude that

$$\begin{aligned}
&2J(X) - 4J(X-X_*)\\
&\geq \frac{1}{R^2}\mathbb{E}\left[|X-X_*|^2\left|\Pi(X-X_*)\left(\rho(X)-\rho(X_*)\right)\right|^2\right] \qquad (10)\\
&- \frac{1}{R^2}\mathbb{E}\left[|X-X_*|^2\left|\Pi(X-X_*)\left(\rho(X)-\rho(X_*)\right)\right|^2\mathbf{1}_{\{|X-X_*|>R\}}\right].
\end{aligned}$$

Since $\mathbb{E}|X|^2 = d$, radial symmetry implies $\mathrm{Cov}(X) = \mathrm{I}$. Therefore, by Lemma 1, the first term in the RHS of (10) is bounded as

$$\mathbb{E}\left[|X-X_*|^2\left|\Pi(X-X_*)\left(\rho(X)-\rho(X_*)\right)\right|^2\right] \geq 2(d-1)I(X).$$

We now bound the second term in the RHS of (10). By c-regularity and the triangle inequality, we have

$$\begin{aligned}
|\Pi(x-x_*)\left(\rho(x)-\rho(x_*)\right)| &\leq |\rho(x)-\rho(x_*)|\\
&\leq c(|x| + |x_*|) + 2c\mathbb{E}|X|.
\end{aligned}$$

Noting that the set $\{|x-x_*| > R\}$ is a superset of the union

$$\{|x| \geq R/2, |x_*| \leq |x|\} \cup \{|x_*| \geq R/2, |x| \leq |x_*|\},$$

we deduce the pointwise inequality

$$\begin{aligned}
&\mathbf{1}_{\{|x-x_*|>R\}}|x-x_*|^2\left|\Pi(x-x_*)\left(\rho(x)-\rho(x_*)\right)\right|^2\\
&\leq \mathbf{1}_{\{|x|\geq R/2\}}4|x|^2\left(2c|x| + 2c\mathbb{E}|X|\right)^2\\
&+ \mathbf{1}_{\{|x_*|\geq R/2\}}4|x_*|^2\left(2c|x_*| + 2c\mathbb{E}|X|\right)^2.
\end{aligned}$$

Taking expectations and using the fact that X, X_* are i.i.d. we have for any conjugate exponents $p, q \geq 1$ and $\beta > 0$,

$$\begin{aligned}
&\mathbb{E}\left[|X - X_*|^2 \,|\Pi(X - X_*)\,(\rho(X) - \rho(X_*))|^2 \,\mathbf{1}_{\{|X-X_*|>R\}}\right]\\
&\leq 16\,c^2\,\mathbb{E}\left[|X|^2\,(|X| + \mathbb{E}|X|)^2\,\mathbf{1}_{\{|X|\geq R/2\}}\right]\\
&\leq 32\,c^2\mathbb{E}\left[|X|^4\,\mathbf{1}_{\{|X|>R/2\}}\right]\\
&\quad + 32\,c^2\,(\mathbb{E}|X|)^2\,\mathbb{E}\left[|X|^2\,\mathbf{1}_{\{|X|\geq R/2\}}\right]\\
&\leq 32\,c^2\,\left\||X|^2\right\|_{2p}^2 (\Pr\{|X| \geq R/2\})^{1/q}\\
&\quad + 32\,c^2\,(\mathbb{E}|X|)^2\,\left\||X|^2\right\|_p (\Pr\{|X| \geq R/2\})^{1/q}\\
&\leq 32c^2\left(\left\||X|^2\right\|_{2p}^2 + (\mathbb{E}|X|)^2\,\left\||X|^2\right\|_p\right)\left(\frac{\mathbb{E}|X|^\beta}{(R/2)^\beta}\right)^{1/q}\\
&\leq \frac{64\cdot 2^{\beta/q}\,c^2}{R^{\beta/q}}\,\left\||X|^2\right\|_{2p}^2\,\left\||X|^2\right\|_{\beta/2}^{\beta/(2q)}.
\end{aligned}$$

Putting both bounds together, we have shown

$$\begin{aligned}
&J(X) - 2J(X - X_*)\\
&\geq \frac{d-1}{R^2} I(X) - \frac{32\cdot 2^{\beta/q}\,c^2}{R^{2+\beta/q}}\,\left\||X|^2\right\|_{2p}^2\,\left\||X|^2\right\|_{\beta/2}^{\beta/(2q)}.
\end{aligned}$$

The goal is to now balance terms by choosing R appropriately. In particular, for any $s > 0$, taking $R = \left(\frac{b(2+s)}{2a}\right)^{1/s}$ yields the identity

$$\frac{a}{R^2} - \frac{b}{R^{2+s}} = \frac{1}{1+2/s}\left(\frac{2/s}{b(1+2/s)}\right)^{2/s} a^{1+2/s}.$$

In view of this, we now set $\varepsilon = 2q/\beta$, $p = 1 + 1/(2\varepsilon)$ (which fixes $q = 1 + 2\varepsilon$ and $\beta = 4 + 2/\varepsilon$) and simplify to obtain

$$J(X) - 2J(X - X_*) \geq \frac{(\varepsilon/8)^\varepsilon\,(\mathbb{E}|X|^2)^{1+\varepsilon}}{(8(1+\varepsilon))^{1+\varepsilon} c^{2\varepsilon}\,\left\||X|^2\right\|_{2+1/\varepsilon}^{1+2\varepsilon}} I(X)^{1+\varepsilon},$$

where we have made use of the crude bound $(d-1)/d \geq 1/2$ and substituted $d = \mathbb{E}|X|^2$. □

B. Entropy Jumps under c-Regularity

As one would expect, we may 'integrate up' in Theorem 2 to obtain an analogous result in terms of entropies.

Theorem 3. *Let X, X_* be i.i.d. radially symmetric random vectors on $\mathbb{R}^d$, $d \geq 2$, with c-regular density f. For any $\varepsilon > 0$*

$$D(X) - D\left(\tfrac{1}{\sqrt{2}}(X + X_*)\right) \geq C_\varepsilon \frac{d^\varepsilon\,\mathbb{E}|X|^2}{c^{2\varepsilon}\,\left\||X|^2\right\|_{2+1/\varepsilon}^{1+2\varepsilon}} D(X)^{1+\varepsilon}. \tag{11}$$

In comparing Theorems 1 and 3, we see that the two claims are essentially identical up to constant factors, except that (7) is an upgrade of (11) in the sense that the regularity parameter c is replaced by the ratio $I(X)/D(X)$. Hence, the regularity assumption of Theorem 1 requires only finite Fisher information, which is weaker than c-regularity (see the comment following Definition 1). The ratio $I(X)/D(X)$ may be interpreted as the defectiveness of X with respect to the logarithmic Sobolev inequality. Interestingly, the defect in the logarithmic Sobolev inequality also supplies an *upper bound* on the entropy jump [21], suggesting that its appearance in both upper and lower bounds is fundamental.

Proof. Similar to before, the inequality (11) is invariant to scaling $t : X \mapsto tX$. Thus, we assume without loss of generality that X is normalized so that $\mathbb{E}|X|^2 = d$. Next, define $W = \frac{1}{\sqrt{2}}(X + X_*)$, and let X_t, W_t denote the Ornstein-Uhlenbeck evolutes of X and W, respectively. That is, for $t \geq 0$

$$\begin{aligned}
X_t &\sim e^{-t}X + (1 - e^{-2t})^{1/2} G,\\
W_t &\sim e^{-t}W + (1 - e^{-2t})^{1/2} G,
\end{aligned} \tag{12}$$

where G is independent of X, W. By Proposition 2, X_t is $(5ce^{2t})$-regular for all $t \geq 0$. An application of Theorem 2 gives

$$\begin{aligned}
I(X_t) - I(W_t) &\geq C_\varepsilon \frac{(\mathbb{E}|X_t|^2)^{1+\varepsilon}}{(5c)^{2\varepsilon}\,\left\||X_t|^2\right\|_{2+1/\varepsilon}^{1+2\varepsilon}}\, e^{-4\varepsilon t} I(X_t)^{1+\varepsilon}\\
&\geq C'_\varepsilon \frac{(\mathbb{E}|X|^2)^{1+\varepsilon}}{c^{2\varepsilon}\,\left\||X|^2\right\|_{2+1/\varepsilon}^{1+2\varepsilon}}\, e^{-4\varepsilon t} I(X_t)^{1+\varepsilon},
\end{aligned} \tag{13}$$

where C'_ε is a constant depending only on ε, and (13) holds since $\mathbb{E}|X_t|^2 = \mathbb{E}|X|^2$ and, for $p \geq 1$,

$$\begin{aligned}
\left\||X_t|^2\right\|_p &= \left(\mathbb{E}|X_t|^{2\cdot p}\right)^{1/p}\\
&\leq 2\left(\mathbb{E}(e^{-2t}|X|^2 + (1 - e^{-2t})|G|^2)^p\right)^{1/p}\\
&= 2\left\|e^{-2t}|X|^2 + (1 - e^{-2t})|G|^2\right\|_p\\
&\leq 2\left(e^{-2t}\left\||X|^2\right\|_p + (1 - e^{-2t})\left\||G|^2\right\|_p\right)\\
&\leq 2(1 + \tfrac{p}{d})\left\||X|^2\right\|_p.
\end{aligned} \tag{14}$$

The bound (14) uses the fact that $|G|^2$ is a chi-squared random variable with d degrees of freedom, and hence (using $\mathbb{E}|X|^2 = d$):

$$\begin{aligned}
\left\||G|^2\right\|_p = \left(2^p \frac{\Gamma(p + \frac{d}{2})}{\Gamma(\frac{d}{2})}\right)^{1/p} &= \mathbb{E}|X|^2\left(\frac{\Gamma(p + \frac{d}{2})}{\Gamma(\frac{d}{2})\left(\frac{d}{2}\right)^p}\right)^{1/p}\\
&\leq \mathbb{E}|X|^2\left(1 + \tfrac{p}{d}\right)\\
&\leq \left\||X|^2\right\|_p (1 + \tfrac{p}{d}).
\end{aligned}$$

Now, the claim will follow by integrating both sides. Indeed, by the integral form of de Bruijn's identity, integrating the LHS of (13) yields

$$\int_0^\infty (I(X_t) - I(W_t))\,dt = D(X) - D(W).$$

By convexity and Jensen's inequality, we may bound the integral of (13) as

$$\begin{aligned}
\int_0^\infty e^{-4\varepsilon t} I(X_t)^{1+\varepsilon} dt &\geq \frac{1}{(4\varepsilon)^\varepsilon}\left(\int_0^\infty e^{-4\varepsilon t} I(X_t) dt\right)^{1+\varepsilon}\\
&\geq \frac{1}{(4\varepsilon)^\varepsilon}\left(\int_0^\infty I(X_{t+2\varepsilon t}) dt\right)^{1+\varepsilon}\\
&= \frac{1}{(4\varepsilon)^\varepsilon (1 + 2\varepsilon)^{1+\varepsilon}} D(X)^{1+\varepsilon},
\end{aligned}$$

where we used the bound $I(X_{t+s}) \leq e^{-2s} I(X_t)$ due to the convolution inequality for Fisher information, a change of variables, and the identity $\int_0^\infty I(X_t)dt = D(X)$. This completes the proof of (11). □

C. Proof of Theorem 1

Proof. Similar to before, inequality (7) is invariant to scaling $t : X \mapsto tX$. So, once again, we will assume without loss of generality that X is normalized so that $\mathbb{E}|X|^2 = d$.

Define $W = \frac{1}{\sqrt{2}}(X + X_*)$, and as in (12) let X_t, W_t denote the Ornstein-Uhlenbeck evolutes of X and W for $t \geq 0$. Using de Bruijn's identity, i.e., $\frac{d}{dt}D(X_t) = -I(X_t)$, and the convolution inequality for Fisher information, it follows that

$$\frac{d}{dt}\left(D(X_t) - D(W_t)\right) = I(W_t) - I(X_t) \leq 0. \qquad (15)$$

Thus, $D(X) - D(W) \geq D(X_t) - D(W_t)$ for all $t \geq 0$. By Proposition 1, X_t is $4\,(1 - e^{-2t})^{-1}$-regular for all $t \geq 0$. Noting that $\mathbb{E}|X_t|^2 = \mathbb{E}|X|^2$, an application of Theorem 3 gives, for all $t \geq 0$,

$$\begin{aligned} D(X) - D(W) &\geq D(X_t) - D(W_t) \\ &\geq C_\varepsilon \frac{(1 - e^{-2t})^{2\varepsilon} d^\varepsilon\, \mathbb{E}|X|^2}{4^{2\varepsilon}\ \||X_t|^2\|_{2+1/\varepsilon}^{1+2\varepsilon}} D(X_t)^{1+\varepsilon} \\ &\geq C'_\varepsilon \frac{(1 - e^{-2t})^{2\varepsilon} d^\varepsilon\, \mathbb{E}|X|^2}{\||X|^2\|_{2+1/\varepsilon}^{1+2\varepsilon}} D(X_t)^{1+\varepsilon}, \end{aligned}$$

where the last inequality follows by the same logic as (13), with C'_ε yet another constant depending only on ε.

The map $t \mapsto D(X_t)$ is continuous and convex on $t \in [0, \infty)$ (e.g., [15]). Hence, using de Bruijn's identity, we have the inequality $D(X_t) \geq D(X) - tI(X)$, so that for all $t \geq 0$

$$D(X) - D(W) \geq \frac{C'_\varepsilon (1 - e^{-2t})^{2\varepsilon} d^\varepsilon\, \mathbb{E}|X|^2}{\||X|^2\|_{2+1/\varepsilon}^{1+2\varepsilon}} (D(X) - tI(X))^{1+\varepsilon}.$$

In principle, we may optimize over $t \geq 0$ to obtain a good lower bound. However, a simple choice suffices. Indeed, the proof is finished by putting $t = \frac{D(X)}{2I(X)}$ and noting that $1 - e^{-x} \geq \frac{1}{\sqrt{2}} x$ for $x \in [0, 1/4]$, which applies for our choice of t due to the logarithmic Sobolev inequality. ☐

IV. CLT for Radially Symmetric Random Vectors

Let $\{X^{(k)}, k \geq 1\}$ be a sequence of i.i.d. radially symmetric random vectors of dimension $d \geq 2$, each with the same distribution as X. Iterating (7) leads to quantitative estimates for convergence (to normality) of the subsequence of standardized sums

$$U_n := 2^{-n/2} \sum_{k=1}^{2^n} X^{(k)}, \qquad n \geq 0.$$

The only difficulty in this is that the norms and Fisher information in (7) change along the subsequence, meaning that the resulting bound on $D(U_n)$ would depend on the moments and Fisher information of U_k for $0 \leq k < n$. However, this can be dealt with using a simple consequence of the Minkowski and Rosenthal inequalities. In terms of the above notation,

Lemma 2. *For $p > 2$, $\big\||U_n|^2\big\|_p \leq dC(p)\|X_1^2\|_p$, where $C(p)$ is a constant depending only on p, and X_1 is the first coordinate of X.*

Using this together with the fact that $I(U_n) \leq I(X)$, it is an easy exercise to establish the following quantitative CLT:

Corollary 1. *Let X, U_n be as above, normalized so that $\mathbb{E}|X|^2 = d$. For any $\varepsilon > 0$*

$$D(U_n) \leq D(X) \exp\left(-K_\varepsilon(X) \sum_{k=0}^{n-1} D(U_k)^{3\varepsilon}\right),$$

where

$$K_\varepsilon(X) = \frac{C_\varepsilon}{d^\varepsilon\, (\mathbb{E}|X_1|^{4+2/\varepsilon})^\varepsilon I(X)^{2\varepsilon}},$$

and $C_\varepsilon > 0$ depends only on ε.

Clearly, this quantifies the convergence of $D(U_n)$ to zero in terms of $n, D(X), I(X)$ and moments of X_1. It also ensures rapid initial convergence when the initial distribution is far from normal (e.g., $D(X) \gg 1$).

References

[1] L. Wang and M. Madiman. Beyond the entropy power inequality, via rearrangements. *IEEE Transactions on Information Theory*, 60(9):5116–5137, 2014.

[2] K. Ball, F. Barthe, and A. Naor. Entropy jumps in the presence of a spectral gap. *Duke Math. Journal*, 119(1):41–63, 2003.

[3] O. Johnson and A. R. Barron. Fisher information inequalities and the central limit theorem. *Probability Theory and Related Fields*, 129(3):391–409, 2004.

[4] K. Ball and V. H. Nguyen. Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Mathematica*, 213(1):81–96, 2012.

[5] T. A. Courtade, M. Fathi, and A. Pananjady. Quantitative Stability of the Entropy Power Inequality. *IEEE Transactions on Information Theory*, to appear, 2018.

[6] A. R. Barron. Entropy and the central limit theorem. *The Annals of probability*, 14(1):336–342, 1986.

[7] 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*, 41(4):2479–2512, 2013.

[8] S. G. Bobkov, G. P. Chistyakov, and F. Götze. Berry-Esseen bounds in the entropic central limit theorem. *Probability Theory and Related Fields*, 159(3-4):435, 2014.

[9] M. Ledoux, I. Nourdin, and G. Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *Geometric and Functional Analysis*, 25(1):256–306, 2015.

[10] T. A. Courtade, M. Fathi, and A. Pananjady. Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Annales de l'Institut Henri Poincaré*, to appear, 2018.

[11] A. Zhai. A high-dimensional CLT in W_2 distance with near optimal convergence rate. *Prob. Theory and Rel. Fields*, 170(3):821–845, 2017.

[12] T. Bonis. Rates in the central limit theorem and diffusion approximation via Stein's method. *arXiv:1506.06966*, 2015.

[13] C. Villani. A review of mathematical topics in collisional kinetic theory. *Handbook of mathematical fluid dynamics*, 1:71–305, 2002.

[14] E. A. Carlen and M. C. Carvalho. Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. *Journal of statistical physics*, 67(3-4):575–608, 1992.

[15] E. A. Carlen and A. Soffer. Entropy production by block variable summation and central limit theorems. *Communications in mathematical physics*, 140(2):339–371, 1991.

[16] G. Toscani and C. Villani. Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. *Comms. in mathematical physics*, 203(3):667–706, 1999.

[17] C. Villani. Cercignani's conjecture is sometimes true and always almost true. *Comms. in mathematical physics*, 234(3):455–490, 2003.

[18] L. Desvillettes and C. Villani. On the spatially homogeneous Landau equation for hard potentials part II: H-theorem and applications. *Comms. in Partial Diff. Equations*, 25(1-2):261–298, 2000.

[19] Y. Polyanskiy and Y. Wu. Wasserstein continuity of entropy and outer bounds for interference channels. *IEEE Transactions on Information Theory*, 62(7):3992–4002, July 2016.

[20] O. Johnson. *Information theory and the central limit theorem*, volume 8. World Scientific, 2004.

[21] T. A. Courtade. A strong entropy power inequality. *IEEE Transactions on Information Theory*, 64(4):2173–2192. April 2018.