The Annals of Probability
2020, Vol. 48, No. 5, 2494–2524
https://doi.org/10.1214/20-AOP1429

THE CLT IN HIGH DIMENSIONS: QUANTITATIVE BOUNDS VIA MARTINGALE EMBEDDING

BY RONEN ELDAN[1,*], DAN MIKULINCER[1,†] AND ALEX ZHAI[2]

[1]*Weizmann Institute of Science*, [*]*roneneldan@gmail.com*; [†]*danmiku@gmail.com*
[2]*Stanford University*, *azhai@stanford.edu*

We introduce a new method for obtaining quantitative convergence rates for the central limit theorem (CLT) in a high-dimensional setting. Using our method, we obtain several new bounds for convergence in transportation distance and entropy, and in particular: (a) We improve the best known bound, obtained by the third named author (*Probab. Theory Related Fields* **170** (2018) 821–845), for convergence in quadratic Wasserstein transportation distance for bounded random vectors; (b) we derive the first nonasymptotic convergence rate for the entropic CLT in arbitrary dimension, for general log-concave random vectors (this adds to (*Ann. Inst. Henri Poincaré Probab. Stat.* **55** (2019) 777–790), where a finite Fisher information is assumed); (c) we give an improved bound for convergence in transportation distance under a log-concavity assumption and improvements for both metrics under the assumption of strong log-concavity. Our method is based on martingale embeddings and specifically on the Skorokhod embedding constructed in (*Ann. Inst. Henri Poincaré Probab. Stat.* **52** (2016) 1259–1280).

1. Introduction. Let $X^{(1)}, \dots, X^{(n)}$ be i.i.d. random vectors in $\mathbb{R}^d$. By the central limit theorem, it is well known that, under mild conditions, the sum $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X^{(i)}$ converges to a Gaussian. With d fixed, there is an extensive literature showing that the distance from Gaussian under various metrics decays as $\frac{1}{\sqrt{n}}$ as $n \to \infty$, and this is optimal.

However, in high-dimensional settings, it is often the case that the dimension d is not fixed but rather grows with n. It then becomes necessary to understand how the convergence rate depends on dimension, and the optimal dependence here is not well understood. We present a new technique for proving central limit theorems in $\mathbb{R}^d$ that is suitable for establishing quantitative estimates for the convergence rate in the high-dimensional setting. The technique, which is described in more detail in Section 1.1 below, is based on pathwise analysis: we first couple the random vector with a Brownian motion via a martingale embedding. This gives rise to a coupling between the sum and a Brownian motion for which we can establish bounds on the concentration of the quadratic variation. We use a multidimensional version of a Skorokhod embedding, inspired by a construction of the first named author from [21], as a manifestation of the martingale embedding.

Using our method, we prove new bounds on quadratic *transportation* (also known as "Kantorovich" or "Wasserstein") distance in the CLT, and in the case of log-concave distributions, we also give bounds for *entropy* distance. Let $\mathcal{W}_2(A, B)$ denote the quadratic transportation distance between two d-dimensional random vectors A and B. That is,

$$\mathcal{W}_2(A, B) = \sqrt{\inf_{\substack{(X,Y) \text{ s.t.} \\ X \sim A, Y \sim B}} \mathbb{E}\left[\|X - Y\|_2^2\right]},$$

Received October 2018; revised January 2020.
MSC2020 subject classifications. 60F05, 60G57.
Key words and phrases. Central limit theorem, martingale embedding.

where the infimum is taken over all couplings of the vectors A and B. As a first demonstration of our method, we begin with an improvement to the best known convergence rate in the case of bounded random vectors.

THEOREM 1. *Let X be a random d-dimensional vector. Suppose that $\mathbb{E}[X] = 0$ and $\|X\| \leq \beta$ almost surely for some $\beta > 0$. Let $\Sigma = \mathrm{Cov}(X)$, and let $G \sim \mathcal{N}(0, \Sigma)$ be a Gaussian with covariance Σ. If $\{X^{(i)}\}_{i=1}^n$ are i.i.d. copies of X and $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X^{(i)}$, then*

$$\mathcal{W}_2(S_n, G) \leq \frac{\beta\sqrt{d}\sqrt{32 + 2\log_2(n)}}{\sqrt{n}}.$$

Theorem 1 improves a result of the third named author [46] that gives a bound of order $\frac{\beta\sqrt{d}\log n}{\sqrt{n}}$ under the same conditions. It was noted in [46] that when X is supported on a lattice $\beta\mathbb{Z}^d$, then the quantity $\mathcal{W}_2(S_n, G)$ is of order $\frac{\beta\sqrt{d}}{\sqrt{n}}$. Thus, Theorem 1 is within a $\sqrt{\log n}$ factor of optimal.

When the distribution of X is isotropic and log-concave, we can improve the bounds guaranteed by Theorem 1. In this case, however, a more general bound has already been established in [19]; see discussion below.

THEOREM 2. *Let X be a random d-dimensional vector. Suppose that the distribution of X is log-concave and isotropic. Let $G \sim \mathcal{N}(0, \mathrm{I}_d)$ be a standard Gaussian. If $\{X^{(i)}\}_{i=1}^n$ are i.i.d. copies of X and $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X^{(i)}$, then there exists a universal constant $C > 0$ such that, if $d \geq 8$,*

$$\mathcal{W}_2(S_n, G) \leq \frac{Cd^{3/4}\ln(d)\sqrt{\ln(n)}}{\sqrt{n}}.$$

REMARK 3. We actually prove the slightly stronger bound

$$\mathcal{W}_2(S_n, G) \leq \frac{C\kappa_d \ln(d)\sqrt{d\ln(n)}}{\sqrt{n}},$$

where

$$\kappa_d := \sup_{\substack{\mu \text{ isotropic,} \\ \text{log-concave}}} \left\| \int_{\mathbb{R}^d} x_1 x \otimes x \mu(dx) \right\|_{\mathrm{HS}}, \tag{1}$$

as defined in [20]. Results in [20] and [35] imply that $\kappa_d = O(d^{1/4})$, leading to the bound in Theorem 2. If the *thin-shell conjecture* (see [2], as well [15]) is true, then the bound is improved to $\kappa_d = O(\sqrt{\ln(d)})$, which yields

$$\mathcal{W}_2(S_n, G) \leq \frac{C\sqrt{d\ln(d)^3\ln(n)}}{\sqrt{n}}.$$

By considering, for example, a random vector uniformly distributed on the unit cube, one can see that the above bound is sharp up to the logarithmic factors.

REMARK 4. To compare with the previous theorem, note that if $\mathrm{Cov}(X) = \mathrm{I}_d$, then $\mathbb{E}\|X\|^2 = d$. Thus, in applying Theorem 1 we must take $\beta \geq \sqrt{d}$, and the resulting bound is then of order at least $\frac{d\sqrt{\log n}}{\sqrt{n}}$.

Next, we describe our results regarding convergence rate in entropy. If A and B are random vectors such that A has density f with respect to the law of B, then relative entropy of A with respect to B is given by

$$\mathrm{Ent}(A||B) = \mathbb{E}\big[\ln\big(f(A)\big)\big].$$

As a warm-up, we first use our method to recover the entropic CLT in any fixed dimension. In dimension one this was first established by Barron [6]. The same methods may also be applied to prove a multidimensional analogue. See [13] for a more quantitative version of the theorem.

THEOREM 5. *Suppose that* $\mathrm{Ent}(X||G) < \infty$. *Then one has*

$$\lim_{n\to\infty} \mathrm{Ent}(S_n||G) = 0.$$

The next result gives the first nonasymptotic convergence rate for the entropic CLT, again under the log-concavity assumption (other nonasymptotic results appear in previous works, notably [19], but require additional assumptions; see below).

THEOREM 6. *Let X be a random d-dimensional vector. Suppose that the distribution of X is log-concave and isotropic. Let $G \sim \mathcal{N}(0, \mathrm{I}_d)$ be a standard Gaussian. If $\{X^{(i)}\}_{i=1}^n$ are i.i.d. copies of X and $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X^{(i)}$ then*

$$\mathrm{Ent}(S_n||G) \le \frac{Cd^{10}(1 + \mathrm{Ent}(X||G))}{n}$$

for a universal constant $C > 0$.

Our method also yields a different (and typically stronger) bound if the distribution is strongly log-concave.

THEOREM 7. *Let X be a d-dimensional random vector with $\mathbb{E}[X] = 0$ and $\mathrm{Cov}(X) = \Sigma$. Suppose further that X is 1-uniformly log concave (i.e., it has a probability density $e^{-\varphi(x)}$ satisfying $\nabla^2\varphi \succeq \mathrm{I}_d$) and that $\Sigma \succeq \sigma \mathrm{I}_d$ for some $\sigma > 0$.*

Let $G \sim \mathcal{N}(0, \Sigma)$ be a Gaussian with the same covariance as X and let $\gamma \sim \mathcal{N}(0, \mathrm{I}_d)$ be a standard Gaussian. If $\{X^{(i)}\}_{i=1}^n$ are i.i.d. copies of X and $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X^{(i)}$, then

$$\mathrm{Ent}(S_n||G) \le \frac{2(d + 2\,\mathrm{Ent}(X||\gamma))}{\sigma^4 n}.$$

REMARK 8. The theorem can be applied when X is isotropic and σ-uniformly log concave for some $\sigma > 0$. In this case, a change of variables shows that $\sqrt{\sigma}X$ is 1-uniformly log concave and has $\sigma\mathrm{I}_d$ as a covariance matrix. Since relative entropy to a Gaussian is invariant under affine transformations, if $G \sim \mathcal{N}(0, \mathrm{I}_d)$ is a standard Gaussian, we get

$$\mathrm{Ent}(S_n||G) = \mathrm{Ent}(\sqrt{\sigma}S_n||\sqrt{\sigma}G) \le \frac{2(d + 2\,\mathrm{Ent}(\sqrt{\sigma}X||G))}{\sigma^4 n}.$$

1.1. *An informal description of the method.* Let B_t be a standard Brownian motion in $\mathbb{R}^d$ with an associated filtration $\mathcal{F}_t$. The following definition will be central to our method.

DEFINITION 9. Let X_t be a martingale satisfying $dX_t = \Gamma_t\, dB_t$ for some adapted process Γ_t taking values in the positive semidefinite cone and let τ be a stopping time. We say that the triplet (X_t, Γ_t, τ) is a martingale embedding of the measure μ if $X_\tau \sim \mu$.

Note that if Γ_t is deterministic, then X_t has a Gaussian law for each t. At the heart of our proof is the following simple idea: Summing up n independent copies of a martingale embedding of μ, we end up with a martingale embedding of μ^{*n} whose associated covariance process has the form $\sqrt{\sum_{i=1}^n (\Gamma_t^{(i)})^2}$. By the law of large numbers, this process is well concentrated and thus the resulting martingale is close to a Brownian motion.

This suggests that it would be useful to couple the sum process $\sum_{i=1}^n X_t^{(i)}$ with the "averaged" process whose covariance is given by $\mathbb{E}[\sqrt{\sum_{i=1}^n (\Gamma_t^{(i)})^2}]$ (this process is a Brownian motion up to deterministic time change). Controlling the error in the coupling naturally leads to a bound on transportation distance. For relative entropy, we can reformulate the discrepancies in the coupling in terms of a predictable drift and deduce bounds by a judicious application of Girsanov's theorem.

In order to derive quantitative bounds, one needs to construct a martingale embedding in a way that makes the fluctuations of the process Γ_t tractable. The specific choices of Γ_t that we consider are based on a construction introduced in [21]. This construction is also related to the entropy minimizing process used by Föllmer ([29, 30], see also Lehec [36]) and to the stochastic localization which was used in [20]. Such techniques have recently gained prominence and have been used, among other things, to improve known bounds of the KLS conjecture [20, 35], calculate large deviations of nonlinear functions [22] and study tubular neighborhoods of complex varieties [34].

The basic idea underlying the construction of the martingale is a certain measure-valued Markov process driven by a Brownian motion. This process interpolates between a given measure and a delta measure via multiplication by infinitesimal linear functions. The Doob martingale associated to the delta measure (the conditional expectation of the measure, based on the past) will be a martingale embedding for the original measure. This construction is described in detail in Section 2.3 below.

1.2. *Related work.* Multidimensional central limit theorems have been studied extensively since at least the 1940s [8] (see also [10] and references therein). In particular, the dependence of the convergence rate on the dimension was studied by Nagaev [38], Senatov [44], Götze [31], Bentkus [7] and Chen and Fang [26], among others. These works focused on convergence in probabilities of convex sets. We mention that in dimension 1, the picture is much clearer and that tight estimates are known under various metrics ([9, 11, 12, 25, 42, 43]).

More recently, dependence on dimension in the high-dimensional CLT has also been studied for Wishart matrices (Bubeck and Ganguly [17], Eldan and Mikulincer [24]), maxima of sums of independent random vectors (Chernozhukov, Chetverikov and Kato [18]), and transportation distance ([46]). As mentioned earlier, Theorem 1 is directly comparable to an earlier result of the third named author [46], improving on it by a factor of $\sqrt{\log n}$ (see also the earlier work [45]). We refer to [46] for a discussion of how convergence in transportation distance may be related to convergence in probabilities of convex sets.

As mentioned above, Theorem 2 is not new, and follows from a result of Courtade, Fathi and Pananjady [19], Theorem 4.1. Their technique employs Stein's method (see also [16], for a different approach using Stein's method) in a novel way which is also applicable to entropic CLTs (see below). In a subsequent work [27], similar bounds are derived for convergence in the pth-Wasserstein transportation metric.

Regarding entropic CLTs, it was shown by Barron [6] that convergence occurs as long as the distribution of the summand has finite relative entropy (with respect to the Gaussian). However, establishing explicit rates of convergence does not seem to be a straightforward task. Even in the restricted setting of log-concave distributions, not much is known. One of the only quantitative results is Proposition 4.3 in [19], which gives near optimal convergence, provided that the distribution has finite Fisher information. We do not know of any results prior to Theorem 6 which give entropy distance bounds of the form $\frac{\mathrm{poly}(d)}{n}$ to a sum of general log-concave vectors.

A one-dimensional result was established by Artstein, Ball, Barthe and Naor [3] and independently by Barron and Johnson [33], who showed an optimal $O(1/n)$ convergence rate in relative entropy for distributions having a spectral gap (i.e., satisfying a Poincaré inequality). This was later improved by Bobkov, Chistyakov and Götze [13, 14], who derive an Edgeworth-type expansion for the entropy distance which also applies to higher dimensions. However, although their estimates contain very precise information as $n \to \infty$, the given error term is only asymptotic in n and no explicit dependence on the measure or on the dimension is given (in fact, the dependence derived from the method seems to be exponential in the dimension d).

A related "entropy jump" bound was proved by Ball and Nguyen [5] for log-concave random vectors in arbitrary dimensions (see also [4]). Essentially, the bound states that for two i.i.d. random vectors X and Y, the relative entropy $\mathrm{Ent}(\frac{X+Y}{\sqrt{2}}||G)$ is strictly less than $\mathrm{Ent}(X||G)$, where the amount is quantified by the spectral gap for the distribution of X. Repeated application gives a bound for entropy of sums of i.i.d. log-concave vectors in any dimension, but the bound is far from optimal. It is not apparent to us whether the method of [5] can be extended to provide quantitative estimates for convergence in the entropic CLT.

1.3. *Notation.* We work in $\mathbb{R}^d$ equipped with the Euclidean norm, which we denote by $\|\cdot\|$. For a positive semidefinite symmetric matrix A we denote by $\sqrt{A}$ the unique positive semidefinite matrix B, for which the relation $B^2 = A$ holds. For symmetric matrices A and B we use $A \preceq B$ to signify that $B - A$ is a positive semidefinite matrix. By $A^\dagger$ we denote the pseudo inverse of A. Put succinctly, this means that in $A^\dagger$ every nonzero eigenvalue of A is inverted. For a random matrix A, we will write $\mathbb{E}[A]^\dagger$, for the pseudo inverse of its expectation.

If B_t is the standard Brownian motion in $\mathbb{R}^d$, then for any adapted process F_t we denote by $\int_0^t F_s\,dB_s$, the Itô stochastic integral. We refer by Itô's isometry to the fact

$$\mathbb{E}\left[\left\|\int_0^t F_s\,dB_s\right\|^2\right] = \int_0^t \mathbb{E}\left[\|F_s\|_{\mathrm{HS}}^2\right] ds$$

when F_t is adapted to the natural filtration of B_t.

μ will always stand for a probability measure. To avoid confusion, when integrating with respect to μ, on $\mathbb{R}^d$, we will use the notation $\int \ldots \mu(dx)$. For a measure-valued stochastic process μ_t, the expression $d\mu_t$ refers to the stochastic derivative of the process. A measure μ on $\mathbb{R}^d$ is said to be log-concave if it is supported on some subspace of $\mathbb{R}^d$ and, relative to the Lebesgue measure of that subspace, it has a density ρ, twice differentiable almost everywhere, for which

$$-\nabla^2 \log(\rho(x)) \succeq 0 \quad \text{for all } x,$$

where ∇^2 denotes the Hessian matrix, in the Alexandrov sense. If in addition there exists an $\sigma > 0$ such that

$$-\nabla^2 \log(\rho(x)) \succeq \sigma \mathrm{I}_d \quad \text{for all } x,$$

we say that μ is σ-uniformly log-concave. The measure μ is called *isotropic* if it is centered and its covariance matrix is the identity, that is,

$$\int_{\mathbb{R}^d} x\mu(dx) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} x \otimes x\mu(dx) = \mathrm{I}_d.$$

Finally, as a convention, we use the letters C, C', c, c' to represent positive universal constants whose values may change between different appearances.

2. Obtaining convergence rates from martingale embeddings. Suppose that we are given a measure μ and a corresponding martingale embedding (X_t, Γ_t, τ). The goal of this section is to express bounds for the corresponding CLT convergence rates (of the sum of independent copies of μ-distributed random vectors) in terms of the behavior of the process Γ_t and τ.

Throughout this section we fix a measure μ on $\mathbb{R}^d$ whose expectation is 0, a random vector $X \sim \mu$, and a corresponding Gaussian $G \sim \mathcal{N}(0, \Sigma)$, where $\mathrm{Cov}(X) = \Sigma$. Also, the sequence $\{X^{(i)}\}_{i=1}^{\infty}$ will denote independent copies of X, and we write $S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X^{(i)}$ for their normalized sum. Finally, we use B_t to denote a standard Brownian motion on $\mathbb{R}^d$ adapted to a filtration $\mathcal{F}_t$.

2.1. *A bound for Wasserstein-*2 *distance.* The following is our main bound for convergence in Wasserstein distance.

THEOREM 10. *Let S_n and G be defined as above and let (X_t, Γ_t, τ) be a martingale embedding of μ. Set $\Gamma_t = 0$ for $t > \tau$, then*

$$\mathcal{W}_2^2(S_n, G) \leq \int_0^{\infty} \min\left(\frac{1}{n} \mathrm{Tr}\big(\mathbb{E}[\Gamma_t^4]\mathbb{E}[\Gamma_t^2]^{\dagger}\big), 4\,\mathrm{Tr}\big(\mathbb{E}[\Gamma_t^2]\big)\right) dt.$$

To illustrate how such a result might be used, let us assume, for simplicity, that $\Gamma_t \prec k\mathrm{I}_d$ almost surely for some $k > 0$ and that τ has a subexponential tail, that is, there exist positive constants $C, c > 0$ such that for any $t > 0$,

$$\mathbb{P}(\tau > t) \leq Ce^{-ct}. \tag{2}$$

Under these assumptions,

$$\begin{aligned}\mathcal{W}_2^2(S_n, G) &\leq \int_0^{\infty} \min\left(\frac{1}{n} \mathrm{Tr}\big(\mathbb{E}[\Gamma_t^4]\mathbb{E}[\Gamma_t^2]^{\dagger}\big), 4k^2\, d\mathbb{P}(\tau > t)\right) dt \\ &\leq dk^2 \int_0^{\frac{\log(n)}{c}} \frac{1}{n}\, dt + 4C\, dk^2 \int_{\frac{\log(n)}{c}}^{\infty} e^{-ct}\, dt = \frac{d \log(n) k^2}{cn} + \frac{4C\, dk^2}{n}.\end{aligned}$$

Towards the proof, we will need the following technical lemma.

LEMMA 1. *Let A, B be positive semidefinite matrices with* $\ker(A) \subset \ker(B)$. *Then,*

$$\mathrm{Tr}\big((\sqrt{A} - \sqrt{B})^2\big) \leq \mathrm{Tr}\big((A - B)^2 A^{\dagger}\big).$$

PROOF. Since A and B are positive semidefinite, $\ker(\sqrt{A} + \sqrt{B}) \subset \ker(\sqrt{A} - \sqrt{B})$. Thus, we have that

$$\begin{aligned}\sqrt{A} - \sqrt{B} &= (\sqrt{A} - \sqrt{B})(\sqrt{A} + \sqrt{B})(\sqrt{A} + \sqrt{B})^{\dagger} \\ &= \big(A - B + [\sqrt{A}, \sqrt{B}]\big)(\sqrt{A} + \sqrt{B})^{\dagger}.\end{aligned} \tag{3}$$

So,

$$\mathrm{Tr}\big((\sqrt{A} - \sqrt{B})^2\big) = \mathrm{Tr}\big(\big(\big(A - B + [\sqrt{A}, \sqrt{B}]\big)(\sqrt{A} + \sqrt{B})^{\dagger}\big)^2\big).$$

Note that for any symmetric matrices X and Y, by the Cauchy–Schwarz inequality,

$$\mathrm{Tr}\big((XY)^2\big) \leq \mathrm{Tr}(XYXY) \leq \sqrt{\mathrm{Tr}(XYYX) \cdot \mathrm{Tr}(YXXY)} = \mathrm{Tr}\big(X^2 Y^2\big).$$

Applying this to the above equation shows

$$\mathrm{Tr}\big((\sqrt{A} - \sqrt{B})^2\big) \leq \mathrm{Tr}\big(\big(A - B + [\sqrt{A}, \sqrt{B}]\big)^2 \big((\sqrt{A} + \sqrt{B})^{\dagger}\big)^2\big).$$

Note that the commutator $[\sqrt{A}, \sqrt{B}]$ is an anti-symmetric matrix, so that $(A - B) \times [\sqrt{A}, \sqrt{B}] + [\sqrt{A}, \sqrt{B}](A - B)$ is anti-symmetric as well. Thus, for any symmetric matrix C, we have that

$$\mathrm{Tr}\big(\big((A - B)[\sqrt{A}, \sqrt{B}] + [\sqrt{A}, \sqrt{B}](A - B)\big)C\big) = 0.$$

Also, since all eigenvalues of anti-symmetric matrices are purely imaginary, the square of such matrices must be negative definite. And again, for any symmetric positive definite matrix C, it holds that $C^{1/2}[\sqrt{A}, \sqrt{B}]^2 C^{1/2}$ is negative definite and $\mathrm{Tr}([\sqrt{A}, \sqrt{B}]^2 C) \leq 0$. Using these observations we obtain

$$\mathrm{Tr}\big((A - B + [\sqrt{A}, \sqrt{B}])^2\big((\sqrt{A} + \sqrt{B})^\dagger\big)^2\big) \leq \mathrm{Tr}\big((A - B)^2\big((\sqrt{A} + \sqrt{B})^\dagger\big)^2\big).$$

Finally, if C, X, Y are positive definite matrices with $X \preceq Y$ then $C^{1/2}(Y - X)C^{1/2}$ is positive definite which shows $\mathrm{Tr}(CX) \leq \mathrm{Tr}(CY)$. The assumption $\ker(A) \subset \ker(B)$ implies $((\sqrt{A} + \sqrt{B})^\dagger)^2 \preceq A^\dagger$, which concludes the claim by

$$\mathrm{Tr}\big((A - B)^2\big((\sqrt{A} + \sqrt{B})^\dagger\big)^2\big) \leq \mathrm{Tr}\big((A - B)^2 A^\dagger\big) \qquad \square$$

PROOF OF THEOREM 10. Recall that (X_t, Γ_t, τ) is a martingale embedding of μ. Let $(X_t^{(i)}, \Gamma_t^{(i)}, \tau^{(i)})$ be independent copies of the embedding. We can always set $\Gamma_t^{(i)} = 0$ whenever $t > \tau^{(i)}$, so that $\int_0^\infty \Gamma_t^{(i)}\, dB_t^{(i)} \sim \mu$. Define $\tilde{\Gamma}_t = \sqrt{\frac{1}{n}\sum_{i=1}^n (\Gamma_t^{(i)})^2}$. Our first goal is to show

$$\mathcal{W}_2^2(G, S_n) \leq \int_0^\infty \mathbb{E}\big[\mathrm{Tr}\big(\big(\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]}\big)^2\big)\big]\, dt. \tag{4}$$

The theorem will then follow by deriving suitable bounds for $\mathbb{E}[\mathrm{Tr}((\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]})^2)]$ using Lemma 1. Consider the sum $\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\infty \Gamma_t^{(i)}\, dB_t^{(i)}$, which has the same law as S_n. It may be rewritten as

$$S_n = \int_0^\infty \tilde{\Gamma}_t\, d\tilde{B}_t,$$

where $d\tilde{B}_t := \frac{1}{\sqrt{n}}\tilde{\Gamma}_t^\dagger \sum_i \Gamma_t^{(i)} dB_t^{(i)}$ is a martingale whose quadratic variation matrix has derivative satisfying

$$\frac{d}{dt}[\tilde{B}]_t = \frac{1}{n}\sum_i \tilde{\Gamma}_t^\dagger (\Gamma_t^{(i)})^2 \tilde{\Gamma}_t^\dagger \preceq \mathrm{I}_d \tag{5}$$

(in fact, as long as $\mathbb{R}^d$ is spanned by the images of $\Gamma_t^{(i)}$, this process is a Brownian motion). We may now decompose S_n as

$$S_n = \int_0^\infty \sqrt{\mathbb{E}[\tilde{\Gamma}_t^2]}\, d\tilde{B}_t + \int_0^\infty \big(\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\tilde{\Gamma}_t^2]}\big)\, d\tilde{B}_t. \tag{6}$$

Observe that $G := \int_0^\infty \sqrt{\mathbb{E}[\tilde{\Gamma}_t^2]}\, d\tilde{B}_t$ has a Gaussian law and that $\mathbb{E}[\tilde{\Gamma}_t^2] = \mathbb{E}[\Gamma_t^2]$. By applying Itô's isometry, we may see that G has the "correct" covariance in the sense that

$$\mathrm{Cov}(G) = \mathbb{E}\bigg[\bigg(\int_0^\infty \sqrt{\mathbb{E}[\tilde{\Gamma}_t^2]}\, d\tilde{B}_t\bigg)^{\otimes 2}\bigg] = \mathbb{E}\bigg[\int_0^\infty \Gamma_t^2\, dt\bigg] = \mathbb{E}\bigg[\bigg(\int_0^\infty \Gamma_t\, dB_t\bigg)^{\otimes 2}\bigg] = \mathrm{Cov}(X).$$

The decomposition (6) induces a natural coupling between G and S_n, which shows, by another application of Itô's isometry, that

$$\begin{aligned}\mathcal{W}_2^2(G, S_n) &\leq \mathbb{E}\bigg[\bigg\|\int_0^\infty \big(\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]}\big)\, d\tilde{B}_t\bigg\|^2\bigg] \overset{(5)}{\leq} \mathrm{Tr}\bigg(\mathbb{E}\bigg[\int_0^\infty \big(\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]}\big)^2\, dt\bigg]\bigg) \\ &= \int_0^\infty \mathbb{E}\big[\mathrm{Tr}\big(\big(\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]}\big)^2\big)\big]\, dt,\end{aligned}$$

where the last equality is due to Fubini's theorem. Thus, (4) is established. Since $(\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]})^2 \preceq 2(\tilde{\Gamma}_t^2 + \mathbb{E}[\Gamma_t^2])$, we have

$$\mathrm{Tr}\big(\mathbb{E}\big[\big(\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]}\big)^2\big]\big) \le 4\,\mathrm{Tr}\big(\mathbb{E}[\Gamma_t^2]\big). \tag{7}$$

To finish the proof, write $U_t := \frac{1}{n}\sum_{i=1}^n (\Gamma_t^{(i)})^2$, so that $\tilde{\Gamma}_t = \sqrt{U_t}$. Since Γ_t is positive semidefinite, it is clear that $\ker(\mathbb{E}[\Gamma_t^2]) \subset \ker(U_t)$. By Lemma 1,

$$\begin{aligned}
\mathbb{E}\big[\mathrm{Tr}\big(\big(\sqrt{U_t} - \sqrt{\mathbb{E}[\Gamma_t^2]}\big)^2\big)\big] &\le \mathrm{Tr}\big(\mathbb{E}\big[\big(U_t - \mathbb{E}[\Gamma_t^2]\big)^2\big]\mathbb{E}[\Gamma_t^2]^\dagger\big) \\
&= \frac{1}{n^2}\,\mathrm{Tr}\left(\sum_{i=1}^n \mathbb{E}\big[\big((\Gamma_t^{(i)})^2 - \mathbb{E}[\Gamma_t^2]\big)^2\big]\mathbb{E}[\Gamma_t^2]^\dagger\right) \\
&= \frac{1}{n}\,\mathrm{Tr}\big(\big(\mathbb{E}[\Gamma_t^4] - \mathbb{E}[\Gamma_t^2]^2\big)\mathbb{E}[\Gamma_t^2]^\dagger\big) \\
&\le \frac{1}{n}\,\mathrm{Tr}\big(\mathbb{E}[\Gamma_t^4]\mathbb{E}[\Gamma_t^2]^\dagger\big),
\end{aligned}$$

where we have used the fact $\mathbb{E}[(\Gamma_t^{(i)})^2] = \mathbb{E}[\Gamma_t^2]$ in the second equality. Combining the last inequality with (7) and (4) produces the required result. $\square$

2.2. *A bound for the relative entropy.* As alluded to in the Introduction, in order to establish bounds on the relative entropy we will use the existence of a martingale embedding to construct an Itô process whose martingale part has a deterministic quadratic variation. This will allow us to relate the relative entropy to a Gaussian with the norm of the drift term through the use of Girsanov's theorem. As a technicality, we require the stopping time associated to the martingale embedding to be constant. Our main bound for the relative entropy reads,

THEOREM 11. *Let $(X_t, \Gamma_t, 1)$ be a martingale embedding of μ. Assume that for every $0 \le t \le 1$, $\mathbb{E}[\Gamma_t] \succeq \sigma_t \mathrm{I}_d \succneqq 0$ and that Γ_t is invertible a.s. for $t < 1$. Then we have the following inequalities*:

$$\mathrm{Ent}(S_n||G) \le \frac{1}{n}\int_0^1 \frac{\mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)]}{(1-t)^2\sigma_t^2}\left(\int_t^1 \sigma_s^{-2}\,ds\right)dt$$

and

$$\mathrm{Ent}(S_n||G) \le \int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_t^2] - \mathbb{E}[\tilde{\Gamma}_t]^2)}{(1-t)^2}\left(\int_t^1 \sigma_s^{-2}\,ds\right)dt,$$

where

$$\tilde{\Gamma}_t = \sqrt{\frac{1}{n}\sum_{i=1}^n (\Gamma_t^{(i)})^2}$$

and $\Gamma_t^{(i)}$ are independent copies of Γ_t.

The theorem relies on the following bound, whose proof is postponed to the end of the subsection.

LEMMA 2. *Let Γ_t be an $\mathcal{F}_t$-adapted matrix-valued processes and let $F : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be almost surely invertible and locally Lipschitz. Denote $F_t(x) := F(t,x)$ and let X_t, M_t be defined by*

$$X_t = \int_0^t \Gamma_s \, dB_s \quad \text{and} \quad M_t = \int_0^t F_s(M_s)\, dB_s.$$

Define the process Y_t by

$$Y_t = \int_0^t F_s(Y_s)\, dB_s + \int_0^t \int_0^s \frac{\Gamma_r - F_r(Y_r)}{1-r}\, dB_r \, ds.$$

Then,

$$\mathrm{Ent}(X_1 || M_1) \leq \mathbb{E}\left[\int_0^1 \int_s^1 \left\| F_t^{-1}(Y_t) \frac{\Gamma_s - F_s(Y_s)}{1-s}\right\|_{\mathrm{HS}}^2 dt\, ds\right].$$

Note that if the process F_t is deterministic, that is, it is a constant function, then M_1 has a Gaussian law, so that the lemma can be used to bound the relative entropy of X_1 with respect to a Gaussian.

PROOF OF THEOREM 11. Let $(X_t^{(i)}, \Gamma_t^{(i)}, 1)$ be independent copies of the martingale embedding. Consider the sum process $\tilde{X}_t = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_t^{(i)}$, which satisfies $\tilde{X}_t = \int_0^t \tilde{\Gamma}_s \, d\tilde{B}_s$ where we define, as in the proof of Theorem 10,

$$\tilde{\Gamma}_t := \sqrt{\frac{1}{n}\sum_{i=1}^n (\Gamma_t^{(i)})^2} \quad \text{and} \quad d\tilde{B}_t = \frac{1}{\sqrt{n}} \tilde{\Gamma}_t^{-1} \sum \Gamma_t^{(i)}\, dB_t^{(i)}.$$

By assumption $\tilde{\Gamma}_t$ is invertible, which makes $\tilde{B}_t$ a Brownian motion. In this case, $(\tilde{X}_t, \tilde{\Gamma}_t, 1)$ is a martingale embedding for the law of S_n. For the first bound, consider the process

$$M_t = \int_0^t \sqrt{\mathbb{E}[\Gamma_s^2]}\, d\tilde{B}_s.$$

By Itô's isometry, one has $M_1 \sim \mathcal{N}(0, \Sigma)$. Also, by Jensen's inequality,

$$\sqrt{\mathbb{E}[\Gamma_t^2]} \succeq \mathbb{E}[\Gamma_t] \succeq \sigma_t \mathrm{I}_d.$$

Using this observation and substituting $\sqrt{\mathbb{E}[\Gamma_t^2]}$ for a constant function F_t in Lemma 2 yields,

$$\mathrm{Ent}(S_n || G) \leq \int_0^1 \mathbb{E}\left[\left\| \frac{\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]}}{1-t}\right\|_{\mathrm{HS}}^2\right] \left(\int_t^1 \sigma_s^{-2}\, ds\right) dt. \tag{8}$$

With the use of Lemma 1, we obtain

$$\begin{aligned}
\mathbb{E}\|\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]}\|_{\mathrm{HS}}^2 &= \mathbb{E}[\mathrm{Tr}((\tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]})^2)] \\
&\leq \mathbb{E}\left[\mathrm{Tr}\left(\left(\frac{1}{n}\sum_{i=1}^n (\Gamma_t^{(i)})^2 - \mathbb{E}[\Gamma_t^2]\right)^2 \mathbb{E}[\Gamma_t^2]^{-1}\right)\right] \\
&\leq \frac{1}{n\sigma_t^2} \mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)].
\end{aligned}$$

Plugging the above into (8) shows the first bound. To see the second bound, we define a process M_t', which is similar to M_t, and is given by the equation

$$M_t' := \int_0^t \mathbb{E}[\tilde{\Gamma}_s]\, d\tilde{B}_s.$$

Let G_n denote a Gaussian which is distributed as M'_1. For any s, we now have the following Cauchy–Schwarz-type inequality:

$$n\left(\sum_{i=1}^{n}(\Gamma_s^{(i)})^2\right) \succeq \left(\sum_{i=1}^{n}\Gamma_s^{(i)}\right)^2.$$

Since the square root is monotone with respect to the order on positive definite matrices, this implies

$$\mathbb{E}[\tilde{\Gamma}_s] \succeq \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\Gamma_s^{(i)}\right] \succeq \sigma_s \mathrm{I}_d.$$

Thus,

$$\begin{aligned}
\mathrm{Ent}(S_n||G_n) &\le \mathbb{E}\left[\int_0^1\int_t^1 \left\|\mathbb{E}[\tilde{\Gamma}_s]^{-1}\frac{\tilde{\Gamma}_t-\mathbb{E}[\tilde{\Gamma}_t]}{1-t}\right\|_{\mathrm{HS}}^2 ds\,dt\right] \\
&\le \int_0^1 \mathbb{E}\left[\left\|\frac{\tilde{\Gamma}_t-\mathbb{E}[\tilde{\Gamma}_t]}{1-t}\right\|_{\mathrm{HS}}^2\right]\left(\int_t^1 \sigma_s^{-2}\,ds\right)dt \\
&= \int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_t^2]-\mathbb{E}[\tilde{\Gamma}_t]^2)}{(1-t)^2}\left(\int_t^1 \sigma_s^{-2}\,ds\right)dt.
\end{aligned}$$

Since $\mathrm{Cov}(G)=\mathrm{Cov}(S_n)$, it is now easy to verify that $\mathrm{Ent}(S_n||G)\le \mathrm{Ent}(S_n||G_n)$, which concludes the proof. □

A key component in the proof of the theorem lies in using the norm of an adapted process in order to bound the relative entropy. The following lemma embodies this idea. Its proof is based on a straightforward application of Girsanov's theorem. We provide a sketch and refer the reader to [36], where a slightly less general version of this lemma is given, for a more detailed proof.

LEMMA 3. *Let $F:\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}^{d\times d}$ be almost surely invertible and locally Lipschitz. Denote $F_t(x):=F(t,x)$ and let $M_t=\int_0^t F_s(M_s)\,dB_s$. For u_t, an adapted process, set $Y_t:=\int_0^t F_s(Y_s)\,dB_s+\int_0^t u_s\,ds$. Then*

$$\mathrm{Ent}(Y_1||M_1)\le \frac{1}{2}\int_0^1 \mathbb{E}\left[\left\|F_t^{-1}(Y_t)u_t\right\|^2\right]dt.$$

PROOF. Since M_t is an Itô diffusion, by Girsanov's theorem ([39], Theorem 8.6.5), the density of $\{Y_t\}_{t\in[0,1]}$ with respect to that of $\{M_t\}_{t\in[0,1]}$ on the space of paths is given by

$$\mathcal{E}:=\exp\left(-\int_0^1 F_t(Y_t)^{-1}u_t\,dB_t-\frac{1}{2}\int_0^1\left\|F_t(Y_t)^{-1}u_t\right\|^2dt\right).$$

If f is the density of Y_1 with respect to M_1, this implies

$$1=\mathbb{E}\left[f(Y_1)\mathcal{E}\right].$$

By Jensen's inequality,

$$0=\ln\left(\mathbb{E}\left[f(Y_1)\mathcal{E}\right]\right)\ge \mathbb{E}\left[\ln\left(f(Y_1)\mathcal{E}\right)\right]=\mathbb{E}\left[\ln\left(f(Y_1)\right)\right]+\mathbb{E}\left[\ln(\mathcal{E})\right].$$

But,

$$\mathbb{E}\left[\ln(\mathcal{E})\right]=-\frac{1}{2}\int_0^1\mathbb{E}\left[\left\|F_t^{-1}(Y_t)u_t\right\|^2\right]dt,$$

and

$$\mathbb{E}\big[\ln\big(f(Y_1)\big)\big] = \mathrm{Ent}(Y_1||M_1),$$

which concludes the proof. $\square$

The proof of Lemma 2 now amounts to invoking the above bound with a suitable construction of the drift process u_t.

PROOF OF LEMMA 2. By definition of the process Y_t, we have the following equality:

$$\begin{aligned}
Y_1 &= \int_0^1 F_t(Y_t)\,dB_t + \int_0^1 \int_0^t \frac{\Gamma_s - F_s(Y_s)}{1-s}\,dB_s\,dt \\
&= \int_0^1 F_t(Y_t)\,dB_t + \int_0^1 \big(\Gamma_t - F_t(Y_t)\big)\,dB_t = X_1,
\end{aligned} \tag{9}$$

where we have used Fubini's theorem in the penultimate equality. Now, consider the adapted process

$$u_t = \int_0^t \frac{\Gamma_s - F_s(Y_s)}{1-s}\,dB_s,$$

so that,

$$dY_t = F_t(Y_t)\,dB_t + u_t\,dt.$$

Applying Lemma 3 and using Itô's isometry, we get

$$\begin{aligned}
\mathrm{Ent}(X_1||M_1) &\le \int_0^1 \mathbb{E}\big[\|F_t^{-1}(Y_t)u_t\|^2\big]\,dt = \int_0^1 \mathbb{E}\left[\left\|\int_0^t F_t^{-1}(Y_t)\frac{\Gamma_s - F_s(Y_s)}{1-s}\,dB_s\right\|^2\right]dt \\
&= \mathbb{E}\left[\int_0^1 \int_0^t \left\|F_t^{-1}(Y_t)\frac{\Gamma_s - F_s(Y_s)}{1-s}\right\|_{\mathrm{HS}}^2 ds\,dt\right] \\
&= \mathbb{E}\left[\int_0^1 \int_s^1 \left\|F_t(Y_t)^{-1}\frac{\Gamma_s - F_s(Y_s)}{1-s}\right\|_{\mathrm{HS}}^2 dt\,ds\right],
\end{aligned}$$

where last equality follows from another use of Fubini's theorem. $\square$

2.3. *A stochastic construction.* In this section, we introduce the main construction used in our proofs, a martingale process which meets the assumptions of Theorems 10 and 11. The construction in the next proposition is based on the Skorokhod embedding described in [21]. Most of the calculations in this subsection are very similar to what is done in [21], except that we allow some inhomogeneity in the quadratic variation according to the function C_t below. In particular, C_t will be a symmetric matrix almost surely, and we will denote the space of $d \times d$ *symmetric* matrices by Sym_d.

PROPOSITION 1. *Let μ be a probability measure on $\mathbb{R}^d$ with smooth density and bounded support. For a probability measure-valued process μ_t, let*

$$a_t = \int_{\mathbb{R}^d} x\mu_t(dx), \qquad A_t = \int_{\mathbb{R}^d} (x-a_t)^{\otimes 2}\mu_t(dx)$$

denote its mean and covariance.

Let $C : \mathbb{R} \times \mathrm{Sym}_d \to \mathrm{Sym}_d$ be a continuous function. Then, we can construct μ_t so that the following properties hold:

1. $\mu_0 = \mu$,
2. *a_t is a stochastic process satisfying $da_t = A_t C(t, A_t^\dagger)\, dB_t$, where B_t is a standard Brownian motion on $\mathbb{R}^d$, and*
3. *for any continuous and bounded $\varphi : \mathbb{R}^d \to \mathbb{R}$, $\int_{\mathbb{R}^d} \varphi(x)\mu_t(dx)$ is a martingale.*

REMARK 12. We will be mainly interested in situations where μ_t converges almost surely to a point mass in finite time. In this case, we obtain a martingale embedding $(a_t, A_t C(t, A_t^\dagger), \tau)$ for μ, where τ is the first time that μ_t becomes a point mass.

In the sequel, we abbreviate $C_t := C(t, A_t^\dagger)$. We first give an informal description of how $\mu_{t+\epsilon}$ is constructed from μ_t for $\epsilon \to 0$. Consider a stochastic process $\{X_s\}_{0 \le s \le 1}$ in which we first sample $X_1 \sim \mu_t$ and then set

$$X_s = (1 - s)a_t + sX_1 + C_t^{-1} B_s,$$

where B_s is a standard Brownian bridge. We can write $X_\epsilon = a_t + \sqrt{\epsilon} C_t^{-1} Z$, where Z is close to a standard Gaussian. We then take $\mu_{t+\epsilon}$ to be the conditional distribution of X_1 given X_ϵ. This immediately ensures that property 1 holds and that a_t is a martingale.

It remains to see why property 1 holds. A direct calculation with conditioned Brownian bridges gives a first-order approximation

$$\begin{aligned}
\mu_{t+\epsilon}(dx) &\propto e^{-\frac{1}{2}(\sqrt{\epsilon} C_t^{-1} Z - \epsilon(x-a_t))^T C_t^2 (\sqrt{\epsilon} C_t^{-1} Z - \epsilon(x-a_t))} \mu_t(dx) \\
&\propto e^{\sqrt{\epsilon}\langle C_t Z, x - a_t\rangle + O(\epsilon)} \mu_t(dx) \\
&\approx \bigl(1 + \sqrt{\epsilon}\langle C_t Z, x - a_t\rangle\bigr)\mu_t(dx).
\end{aligned}$$

Then, to highest order, we have

$$a_{t+\epsilon} - a_t \approx \sqrt{\epsilon} \int_{\mathbb{R}^d} \langle C_t Z, x - a_t\rangle (x - a_t)\mu_t(dx) = \sqrt{\epsilon} A_t C_t Z,$$

which translates into property 1 as $\epsilon \to 0$.

Observe that the procedure outlined above yields measures μ_t that have densities which are proportional to the original density μ times a Gaussian density. (This applies at least when A_t is nondegenerate; something similar also holds when A_t is degenerate, as we will see shortly.) Let us now perform the construction formally. We will proceed by iterating the following preliminary construction, which handles the case when A_t remains nondegenerate.

LEMMA 4. *Let μ be a measure on $\mathbb{R}^d$ with smooth density and bounded support, and let $C : \mathbb{R} \times \mathrm{Sym}_d \to \mathrm{Sym}_d$ be a continuous map. Then, there is a measure-valued process μ_t and a stopping time T such that μ_t satisfies the properties in Proposition 1 for $t < T$ and the affine hull of the support of μ_T has dimension strictly less than d. Moreover, if μ_T is considered as a measure on this affine hull, it has a smooth density.*

PROOF. We will construct a $(\mathbb{R}^d \times \mathrm{Sym}_d)$-valued stochastic process $(c_t, \tilde{\Sigma}_t)$ started at $(c_0, \tilde{\Sigma}_0) = (0, \mathrm{I}_d)$. Let us write

$$Q_t(x) = \frac{1}{2}\langle x - c_t, \tilde{\Sigma}_t^{-1}(x - c_t)\rangle,$$

and let $\tilde{\mu}$ be the probability measure satisfying $\frac{d\tilde{\mu}}{d\mu}(x) \propto e^{\frac{1}{2}\|x\|^2}$. We will then take μ_t to be $\mu_t(dx) = F_t(x)\tilde{\mu}(dx)$, where

$$F_t(x) = \frac{1}{Z_t} e^{-Q_t(x)}, \quad Z_t = \int_{\mathbb{R}^d} e^{-Q_t(x)} \tilde{\mu}(dx).$$

Note that since $\tilde{\Sigma}_0 = \mathrm{I}_d$, we have $\mu_0 = \mu$.[1]

In order to specify the process, it remains to construct $(c_t, \tilde{\Sigma}_t)$. We take it to be the solution to the SDE

$$dc_t = \tilde{\Sigma}_t C_t \, dB_t + \tilde{\Sigma}_t C_t^2 (a_t - c_t)\, dt, \qquad d\tilde{\Sigma}_t = -\tilde{\Sigma}_t C_t^2 \tilde{\Sigma}_t \, dt.$$

Note that the coefficients of this SDE are continuous functions of $(c_t, \tilde{\Sigma}_t)$ so long as $\tilde{\Sigma}_t \succ 0$. By standard existence and uniqueness results, this SDE has a unique solution up to a stopping time T (possibly $T = \infty$), at which point A_t (and hence $\tilde{\Sigma}_t$) becomes degenerate. Observe that, for every t, $\tilde{\Sigma}_t \preceq \mathrm{I}_d$ and so, the matrix process is continuous on the interval $[0, T]$.

By a limiting procedure, it is easy to see that μ_T has a smooth density when considered as a measure on the affine hull of its support. (Indeed, its density is proportional to the conditional density of $\tilde{\mu}$ times a Gaussian density.) It remains to verify that μ_t is a martingale and $da_t = A_t C_t \, dB_t$.

By direct calculation, we have

$$\begin{aligned}
d(\tilde{\Sigma}_t^{-1}) &= C_t^2 \, dt, \\
d(\tilde{\Sigma}_t^{-1} c_t) &= C_t^2 c_t \, dt + C_t^2 (a_t - c_t)\, dt + C_t \, dB_t \\
&= C_t^2 a_t \, dt + C_t \, dB_t, \\
dQ_t(x) &= \left\langle x, \left(\frac{1}{2} C_t^2 x - C_t^2 a_t\right) dt - C_t \, dB_t \right\rangle, \\
d(e^{-Q_t(x)}) &= -e^{-Q_t(x)} dQ_t(x) + \frac{1}{2} e^{-Q_t(x)} d[Q_t(x)] \\
&= e^{-Q_t(x)} \langle x, C_t \, dB_t + C_t^2 a_t \, dt \rangle.
\end{aligned}$$

Integrating against $\tilde{\mu}(dx)$, we obtain

$$\begin{aligned}
dZ_t &= Z_t \langle a_t, C_t \, dB_t + C_t^2 a_t \, dt \rangle, \\
dZ_t^{-1} &= -\frac{1}{Z_t^2} dZ_t + \frac{1}{Z_t^3} d[Z_t] = \frac{1}{Z_t} \langle a_t, -C_t \, dB_t \rangle, \\
dF_t(x) &= e^{-Q_t(x)} dZ_t^{-1} + Z_t^{-1} d(e^{-Q_t(x)}) + d[Z_t^{-1}, e^{-Q_t(x)}] \\
&= F_t(x) \cdot \langle x - a_t, C_t \, dB_t \rangle.
\end{aligned}$$

Thus, $F_t(x)$ is a martingale for each fixed x, and furthermore,

$$da_t = d\int_{\mathbb{R}^d} x \mu_t(dx) = \int_{\mathbb{R}^d} x d\mu_t(dx) = \int_{\mathbb{R}^d} x(x - a_t) C_t \mu_t(dx)\, dB_t = A_t C_t \, dB_t. \qquad \square$$

PROOF OF PROPOSITION 1. We use the process given by Lemma 4, which yields a stopping time T_1 and a measure μ_{T_1} with a strictly lower-dimensional support. If μ_T is a point mass, then we set $\mu_t = \mu_T$ for all $t \geq T$.

[1] Conceptually, one can replace all instances of $\tilde{\mu}$ with μ if we think of the initial value $\tilde{\Sigma}_0$ as being an "infinite" multiple of identity. However, to avoid issues with infinities, we have expressed things in terms of $\tilde{\mu}$ instead.

Otherwise, by the smoothness properties of μ_{T_1} guaranteed by Lemma 4, we can recursively apply Lemma 4 again on μ_{T_1} conditioned on the affine hull of its support. Repeating this procedure at most d times gives us the desired process. $\square$

2.4. *Properties of the construction.* We record here various formulas pertaining to the quantities a_t, A_t and μ_t constructed in Proposition 1.

PROPOSITION 2. *Let μ, C_t and μ_t be as in Proposition 1. Then, there is a* Sym_d*-valued process $\{\Sigma_t\}_{t>0}$ satisfying the following*:

- *For all t, there is an affine subspace $L = L_t \subset \mathbb{R}^d$ and a Gaussian measure γ_t on $\mathbb{R}^d$, supported on L, with covariance Σ_t such that μ_t is absolutely continuous with respect to γ_t, and*

$$\frac{d\mu_t}{d\gamma_t}(x) \propto \mu(x) \quad \forall x \in L.$$

- *Σ_t is continuous and for almost every t obeys the differential equation*

$$\frac{d}{dt}\Sigma_t = -\Sigma_t C_t^2 \Sigma_t.$$

- $\lim_{t\to 0^+} \Sigma_t^{-1} = 0$.

PROOF. For $1 \le k \le d$, let T_k denote the first time the measure μ_t is supported in a $(d-k)$-dimensional affine subspace, and denote by L_t the affine hall of the support of μ_t. We will define Σ_t inductively for each interval $[T_{k-1}, T_k]$. Recall from the proof of Proposition 1 that μ_t is constructed by iteratively applying Lemma 4 to affine subspaces of decreasing dimension $d, d-1, d-2, \dots, 1$. Let $\tilde{\Sigma}_{k,t}$ denote the quantity $\tilde{\Sigma}_t$, from the k-th application of Lemma 4, so that $\tilde{\Sigma}_{k,t}$ is a linear operator on the subspace L_{T_k}.

For the base case $0 < t \le T_1$, take $\Sigma_t = (\tilde{\Sigma}_{0,t}^{-1} - \mathrm{I}_d)^{-1}$. A straightforward calculation shows that over this time interval, $\frac{d\mu_t}{d\mu}$ is proportional to the density of a Gaussian with covariance Σ_t. Note that since $\tilde{\Sigma}_{0,0}^{-1} = \mathrm{I}_d$, we also have $\lim_{t\to 0^+} \Sigma_t^{-1} = 0$.

Now suppose that Σ_t has been defined up until time T_k; we will extend it to time T_{k+1}. Let L_k denote the affine hull of the support of μ_{T_k}, so that $\dim(L_k) = d-k$ (if $\dim(L_k) < d-k$, then we simply have $T_{k+1} = T_k$). Then, for $0 \le t \le T_{k+1} - T_k$, we may set

$$\Sigma_{T_k+t} := (\tilde{\Sigma}_{k,t}^{-1} + \Sigma_{T_k}^{-1} - \mathrm{I}_d)^{-1},$$

where the quantities involved are matrices over the subspace parallel to L_k but may also be regarded as degenerate bilinear forms in the ambient space $\mathbb{R}^d$. First, observe that continuity of the processes $\tilde{\Sigma}_{k,t}$ implies the same for Σ_t. Once again, a straightforward calculation shows that for $T_k \le t < T_{k+1}$, $\frac{d\mu_t}{d\mu}$ is proportional to the density of a Gaussian with covariance Σ_t, where we view μ_t and μ as densities on L_k (for μ, we take its conditional density on L_k).

It remains only to show that Σ_t satisfies the required differential equation. From our construction, we see that Σ_t always takes the form $(\tilde{\Sigma}_t^{-1} - H)^{-1}$, where $H \preceq \mathrm{I}_d$ and

$$\frac{d}{dt}\tilde{\Sigma}_t = -\tilde{\Sigma}_t C_t^2 \tilde{\Sigma}_t.$$

Then, we have

$$\begin{aligned}\frac{d}{dt}\Sigma_t &= -(\tilde{\Sigma}_t^{-1} - H)^{-1}\left(\frac{d}{dt}\tilde{\Sigma}_t^{-1}\right)(\tilde{\Sigma}_t^{-1} - H)^{-1} \\ &= -\Sigma_t\left(-\tilde{\Sigma}_t^{-1}\left(\frac{d}{dt}\tilde{\Sigma}_t\right)\tilde{\Sigma}_t^{-1}\right)\Sigma_t = -\Sigma_t C_t^2 \Sigma_t,\end{aligned}$$

as desired. $\square$

PROPOSITION 3. $dA_t = \int_{\mathbb{R}^d} (x - a_t)^{\otimes 3} \mu_t(dx) C_t \, dB_t - A_t C_t^2 A_t \, dt$.

PROOF. We consider the Doob decomposition of $A_t = M_t + E_t$, where M_t is a local martingale and E_t is a process of bounded variation. By the previous two propositions and the definition of A_t, we have on one hand

$$dA_t = d \int_{\mathbb{R}^d} x^{\otimes 2} \mu_t(dx) - da_t^{\otimes 2} = d \int_{\mathbb{R}^d} x^{\otimes 2} \mu_t(dx) - a_t \otimes da_t - da_t \otimes a_t - A_t C_t^2 A_t \, dt.$$

Clearly the first 3 terms are local martingales, which shows, by the uniqueness of the Doob decomposition, $dE_t = -A_t C_t^2 A_t \, dt$. On the other hand, one may also rewrite the above as

$$\begin{aligned} dA_t &= d \int_{\mathbb{R}^d} (x - a_t)^{\otimes 2} \mu_t(dx) = \int_{\mathbb{R}^d} d\big((x - a_t)^{\otimes 2} \mu_t(dx)\big) \\ &= -\int_{\mathbb{R}^d} da_t \otimes (x - a_t) \mu_t(dx) - \int_{\mathbb{R}^d} (x - a_t) \otimes da_t \mu_t(dx) + \int_{\mathbb{R}^d} (x - a_t)^{\otimes 2} d\mu_t(dx) \\ &\quad - 2 \int_{\mathbb{R}^d} (x - a_t) \otimes d\big[a_t, \mu_t(dx)\big]_t + \int_{\mathbb{R}^d} d[a_t, a_t]_t \mu_t(dx). \end{aligned}$$

Note that the first 2 terms are equal to 0, since, by definition of a_t,

$$\int_{\mathbb{R}^d} da_t \otimes (x - a_t) \mu_t(dx) = da_t \otimes \int_{\mathbb{R}^d} (x - a_t) \mu_t(dx) = 0.$$

Also, the last 2 terms are clearly of bounded variation, which shows

$$dM_t = \int_{\mathbb{R}^d} (x - a_t)^{\otimes 2} d\mu_t(dx) = \int_{\mathbb{R}^d} (x - a_t)^{\otimes 3} C_t \mu_t(dx) \, dB_t. \qquad \square$$

Define the stopping time $\tau = \inf\{t | A_t = 0\}$. Then, at time τ, μ_τ is just a delta mass located at a_τ and $\mu_s = \mu_\tau$ for every $s \geq \tau$. A crucial is observation is the following proposition.

PROPOSITION 4. *Suppose that there exists constants $t_0 \geq 0$ and $c > 0$ such that a.s. one of the following happens*:

1. *for every $t_0 < t < \tau$, $\mathrm{Tr}(A_t C_t^2 A_t) > c$,*
2. *$\int_0^{t_0} \lambda_{\min}(C_t^2) \, dt = \infty$, where $\lambda_{\min}(C_t^2)$ is the minimal eigenvalue of C_t^2,*

then τ is finite a.s. and in the second case $\tau \leq t_0$. Moreover, if τ is finite a.s. then a_τ has the law of μ.

PROOF. Consider the process $R_t = A_t + \int_0^t A_s C_s^2 A_s ds$. For the first case, the previous proposition shows that the real-valued process $\mathrm{Tr}(R_t)$ a positive local martingale; hence, a super-martingale. By the martingale convergence theorem $\mathrm{Tr}(R_t)$ converges to a limit almost surely. By our assumption, if $\tau = \infty$ then

$$\int_0^\infty \mathrm{Tr}\big(A_t C_t^2 A_t\big) \, dt \geq \int_{t_0}^\infty \mathrm{Tr}\big(A_t C_t^2 A_t\big) \, dt \geq \int_{t_0}^\infty c \, dt = \infty.$$

This would imply that $\lim_{t \to \infty} \mathrm{Tr}(A_t) = -\infty$ which clearly cannot happen.

For the second case, under the event $\{\tau > t_0\}$, by continuity of the process A_t there exists $a > 0$ such that for every $t \in [0, t_0]$, there is a unit vector $v_t \in \mathbb{R}^d$ for which $\langle v_t, A_t v_t \rangle \geq a$. We then have,

$$\int_0^{t_0} \mathrm{Tr}\big(A_t C_t^2 A_t\big) \, dt \geq \int_0^{t_0} \big\langle A_t v_t, C_t^2 A_t v_t \big\rangle dt \geq a^2 \int_0^{t_0} \lambda_{\min}\big(C_t^2\big) \, dt = \infty,$$

which implies $\lim_{t \to t_0} \mathrm{Tr}(A_t) = -\infty$. Again, this cannot happen and so $\mathbb{P}(\tau > t_0) = 0$.

To understand the law of a_τ, let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be any continuous bounded function. By Property 1 of Proposition 1 $\int_{\mathbb{R}^d} \varphi(x)\mu_t(dx)$ is a martingale. We claim that it is bounded. Indeed, observe that since μ_t is a probability measure for every t, then

$$\int_{\mathbb{R}^d} \varphi(x)\mu_t(dx) \le \max_x |\varphi(x)|.$$

τ is finite a.s., so by the optional stopping theorem for continuous time martingales ([39] Theorem 7.2.4)

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \varphi(x)\mu_\tau(dx)\right] = \int_{\mathbb{R}^d} \varphi(x)\mu(dx).$$

Since μ_τ is a delta mass, we have that $\int_{\mathbb{R}^d} \varphi(x)\mu_\tau(dx) = \varphi(a_\tau)$ which finishes the proof. □

We finish the section with an important property of the process A_t.

PROPOSITION 5. *The rank of A_t is monotonic decreasing in t, and* $\ker(A_t) \subset \ker(A_s)$ *for $t \le s$.*

PROOF. To see that $\operatorname{rank}(A_t)$ is indeed monotonic decreasing, let v_0 be such that $A_{t_0} v_0 = 0$ for some $t_0 > 0$, we will show that for any $t \ge t_0$, $A_t v_0 = 0$. In a similar fashion to Proposition 4, we define the process $\langle v_0, A_t v_0\rangle + \int_0^t \langle v_0, A_s C_s^2 A_s v_0\rangle\, ds$, which is, using Proposition 3, a positive local martingale and so a super-martingale. This then implies that $\langle v_0, A_t v_0\rangle$ is itself a positive super-martingale. Since $\langle v_0, A_{t_0} v_0\rangle = 0$, we have that for any $t \ge t_0$, $\langle v_0, A_t v_0\rangle = 0$ as well. □

3. Convergence rates in transportation distance.

3.1. *The case of bounded random vectors*: *Proof of Theorem* 1. In this subsection, we fix a measure μ on $\mathbb{R}^d$ and a random vector $X \sim \mu$ with the assumption that $\|X\| \le \beta$ almost surely for some $\beta > 0$. We also assume that $\mathbb{E}[X] = 0$.

We define the martingale process a_t along with the stopping time τ as in Section 2.3, where we take $C_t = A_t^\dagger$, so that $a_t = \int_0^t A_s A_s^\dagger\, dB_s$. We denote $P_t := A_t A_t^\dagger$, and remark that since A_t is symmetric, P_t is a projection matrix. As such, we have that for any $t < \tau$, $\operatorname{Tr}(P_t) \ge 1$. By Proposition 4, a_τ has the law μ.

In light of the remark following Theorem 10, our first objective is to understand the expectation of τ.

LEMMA 5. *Under the boundedness assumption $\|X\| \le \beta$, we have $\mathbb{E}[\tau] \le \beta^2$.*

PROOF. Let $H_t = \|a_t\|^2$. By Itô's formula and since P_t is a projection matrix,

$$dH_t = 2\langle a_t, P_t\, dB_t\rangle + \operatorname{Tr}(P_t)\, dt = 2\langle a_t, P_t\, dB_t\rangle + \operatorname{rank}(P_t)\, dt.$$

So, $\frac{d}{dt}\mathbb{E}[H_t] = \mathbb{E}[\operatorname{rank}(P_t)]$. Since $\mathbb{E}[H_\infty] \le \beta^2$,

$$\beta^2 \ge \mathbb{E}[H_\infty] - \mathbb{E}[H_0] = \int_0^\infty \mathbb{E}\big[\operatorname{rank}(P_t)\big]\, dt \ge \int_0^\infty \mathbb{P}(\tau > t)\, dt = \mathbb{E}[\tau].$$ □

The above claim gives bounds on the expectation of τ, however in order to use Theorem 10, we need bounds for its tail behavior in the sense of (2). To this end, we can use a bootstrap argument and invoke the above lemma with the measure μ_t in place of μ, recalling that $X_\infty|\mathcal{F}_t \sim \mu_t$ and noting that $\|X_\infty|\mathcal{F}_t\| \le \beta$ almost surely. Therefore, we can consider the conditioned stopping time $\tau|\mathcal{F}_t - t$ and get that

$$\mathbb{E}[\tau|\mathcal{F}_t] \le t + \beta^2.$$

The following lemma will make this precise.

LEMMA 6. *Suppose that, for the stopping time τ, it holds that for every $t > 0$, $\mathbb{E}[\tau|\mathcal{F}_t] \leq t + \beta^2$ a.s., then*

$$\forall i \in \mathbb{N}, \quad \mathbb{P}(\tau \geq i \cdot 2\beta^2) \leq \frac{1}{2^i}. \tag{10}$$

PROOF. Denote $t_i = i \cdot 2\beta^2$. Since μ_t is Markovian, and by the law of total probability, for any $i \in \mathbb{N}$ we have the relation

$$\mathbb{P}(\tau \geq t_{i+1}) \leq \mathbb{P}(\tau > t_i) \operatorname*{ess\,sup}_{\mu_{t_i}} \big(\mathbb{P}(\tau - t_i \geq 2\beta^2|\mathcal{F}_{t_i})\big),$$

where the essential supremum is taken over all possible states of μ_{t_i}. Using Markov's inequality, we almost surely have

$$\mathbb{P}(\tau - t_i \geq 2\beta^2|\mathcal{F}_{t_i}) \leq \frac{\mathbb{E}[\tau - t_i|\mathcal{F}_{t_i}]}{2\beta^2} \leq \frac{1}{2},$$

which is also true for the essential supremum. Clearly $\mathbb{P}(\tau \geq 0) = 1$ which finishes the proof. □

PROOF OF THEOREM 1. Our objective is to apply Theorem 10, defining $X_t = a_t$ and $\Gamma_t = P_t$ so that (X_t, Γ_t, τ) becomes a martingale embedding according to Proposition 4. In this case, we have that Γ_t is a projection matrix almost surely. Thus,

$$\operatorname{Tr}\big(\mathbb{E}[\Gamma_t^4]\mathbb{E}[\Gamma_t^2]^\dagger\big) \leq d$$

and

$$\operatorname{Tr}\big(\mathbb{E}[\Gamma_t^2]\big) \leq d\mathbb{P}(\tau > t).$$

Therefore, if G and S_n are defined as in Theorem 10, then

$$\begin{aligned}
\mathcal{W}_2^2(S_n, G) &\leq \int_0^{2\beta^2 \log_2(n)} \frac{d}{n}\, dt + \int_{2\beta^2 \log_2(n)}^\infty 4\, d\mathbb{P}(\tau > t)\, dt \\
&\leq \frac{2\, d\beta^2 \log_2(n)}{n} + 4d \int_{2\beta^2 \log_2(n)}^\infty \mathbb{P}\left(\tau > \left\lfloor \frac{t}{2\beta^2} \right\rfloor 2\beta^2\right) dt \\
&\stackrel{(10)}{\leq} \frac{2\, d\beta^2 \log_2(n)}{n} + 4d \int_{2\beta^2 \log_2(n)}^\infty \left(\frac{1}{2}\right)^{\lfloor \frac{t}{2\beta^2} \rfloor} dt \\
&\leq \frac{2\, d\beta^2 \log_2(n)}{n} + 8\, d\beta^2 \sum_{j=\lfloor \log_2(n) \rfloor}^\infty \frac{1}{2^j} \leq \frac{2\, d\beta^2 \log_2(n)}{n} + \frac{32\, d\beta^2}{n}.
\end{aligned}$$

Taking square roots, we finally have

$$\mathcal{W}_2(S_n, G) \leq \frac{\beta\sqrt{d}\sqrt{32 + 2\log_2(n)}}{\sqrt{n}},$$

as required. □

3.2. *The case of log-concave vectors*: *Proof of Theorem* 2. In this section, we fix μ to be an isotropic log concave measure. The processes $a_t = a_t^\mu$, $A_t = A_t^\mu$ are defined as in Section 2.3 along with the stopping time τ. To define the matrix process C_t, we first define a new stopping time

$$T := 1 \wedge \inf\{t \,|\, \|A_t\|_{\text{op}} \geq 2\}.$$

C_t is then defined in the following manner:

$$C_t = \begin{cases} \min(A_t^\dagger, \mathrm{I}_d) & \text{if } t \le T, \\ A_t^\dagger & \text{otherwise,} \end{cases}$$

where, again, $A_t^\dagger$ denotes the pseudo-inverse of A_t and $\min(A_t^\dagger, \mathrm{I}_d)$ is the unique matrix which is diagonalizable with respect to the same basis as $A_t^\dagger$ and such that each of its eigenvalues corresponds to an an eigenvalue of $A_t^\dagger$ truncated at 1. Since $\mathrm{Tr}(A_t A_t^\dagger) \ge 1$ whenever $t \le \tau$, then the conditions of Proposition 4 are clearly met for $t_0 = 1$ and a_τ has the law of μ.

In order to use Theorem 10, we will also need to demonstrate that τ has subexponential tails in the sense of (2). For this, we first relate τ to the stopping time T.

LEMMA 7. $\tau < 1 + \frac{4}{T}$.

PROOF. Let Σ_t be as in Proposition 2. As described in the proposition, μ_t is proportional to μ times a Gaussian of covariance Σ_t, on an appropriate affine subspace. In this case, an application of the Brascamp-Lieb inequality (see [32] for details) shows that $A_t = \mathrm{Cov}(\mu_t) \preceq \Sigma_t$. In particular, this means that for $t > T$, when restricted to the orthogonal complement of $\ker(A_t)$, the following inequality holds:

$$\frac{d}{dt}\Sigma_t = -\Sigma_t C_t^2 \Sigma_t \preceq -\mathrm{I}_d.$$

So, $\tau \le T + \|\Sigma_T\|_{\mathrm{op}}$.

It remains to estimate $\|\Sigma_T\|_{\mathrm{op}}$. To this end, recall that for $0 < t \le T$, we have $\|A_t\|_{\mathrm{op}} \le 2$, which implies

$$\frac{d}{dt}\Sigma_t = -\Sigma_t C_t^2 \Sigma_t \preceq -\frac{1}{4}\Sigma_t^2.$$

Now, consider the differential equation $f'(t) = -\frac{1}{4} f(t)^2$ with $f(T) = \|\Sigma_T\|_{\mathrm{op}}$, which has solution $f(t) = \frac{4}{t - T + \frac{4}{\|\Sigma_T\|_{\mathrm{op}}}}$. By Gronwall's inequality, $f(t)$ lower bounds $\|\Sigma_t\|_{\mathrm{op}}$ for $0 < t \le T$, and so, in particular, $f(t)$ must remain finite within that interval. Consequently, we have

$$\frac{4}{\|\Sigma_T\|_{\mathrm{op}}} > T \quad \Longrightarrow \quad \|\Sigma_T\|_{\mathrm{op}} < \frac{4}{T}.$$

We conclude that

$$\tau \le T + \|\Sigma_T\|_{\mathrm{op}} < 1 + \frac{4}{T},$$

as desired. □

LEMMA 8. *There exist universal constants $c, C > 0$ such that if $s > C \cdot \kappa_d^2 \ln(d)^2$ and $d \ge 8$ then*

$$\mathbb{P}(\tau > s) \le e^{-cs},$$

where κ_d is the constant defined in (1).

PROOF. First, by using the previous claim, we may see that for any $s \ge 5$,

$$\mathbb{P}(\tau > s) \le \mathbb{P}\left(\frac{1}{T} \ge \frac{s-1}{4}\right) \le \mathbb{P}\left(\frac{1}{T} \ge \frac{s}{5}\right) = \mathbb{P}(5s^{-1} \ge T) = \mathbb{P}\left(\max_{0 \le t \le 5s^{-1}} \|A_t\|_{\mathrm{op}} \ge 2\right).$$

Recall from Proposition 3,

$$dA_t = \int_{\mathbb{R}^d} (x - a_t) \otimes (x - a_t) \langle C_t(x - a_t), dB_t \rangle \mu_t(dx) - A_t C_t^2 A_t \, dt.$$

Since we are trying to bound the operator norm of A_t, we might as well just consider the matrix $\tilde{A}_t = A_t + \int_0^t A_s C_s^2 A_s ds$. Note that, by definition of T, for any $t \le T$,

$$\int_0^t A_s C_s^2 A_s ds \preceq \mathrm{I}_d.$$

Thus, for $t \in [0, T]$,

$$3\mathrm{I}_d \succeq A_t + \mathrm{I}_d \succeq \tilde{A}_t \succeq A_t. \tag{11}$$

Also, $\tilde{A}_t$ can be written as,

$$d\tilde{A}_t = \int_{\mathbb{R}^d} (x - a_t) \otimes (x - a_t) \langle C_t(x - a_t), dB_t \rangle \mu_t(dx), \quad \tilde{A}_0 = \mathrm{I}_d. \tag{12}$$

The above shows

$$\mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} \|A_t\|_{\mathrm{op}} \ge 2\Big) \le \mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} \|\tilde{A}_t\|_{\mathrm{op}} \ge 2\Big).$$

We note than whenever $\|\tilde{A}_t\|_{\mathrm{op}} \ge 2$ then also $\mathrm{Tr}(\tilde{A}_t^{4\ln(d)})^{\frac{1}{4\ln(d)}} \ge 2$, so that

$$\begin{aligned}
\mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} \|\tilde{A}_t\|_{\mathrm{op}} \ge 2\Big) &\le \mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} \mathrm{Tr}\big(\tilde{A}_t^{4\ln(d)}\big)^{\frac{1}{4\ln(d)}} \ge 2\Big) \\
&\le \mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} \ln\big(\mathrm{Tr}\big(\tilde{A}_t^{4\ln(d)}\big)\big) \ge 2\ln(d)\Big) \\
&= \mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} (M_t + E_t) \ge 2\ln(d)\Big),
\end{aligned} \tag{13}$$

where M_t and E_t form the Doob-decomposition of $\ln(\mathrm{Tr}(\tilde{A}_t^{4\ln(d)}))$. That is, M_t is a local martingale and E_t is a process of bounded variation. To calculate the differential of the Doob-decomposition, fix t, let $v_1, \dots, v_n$ be the unit eigenvectors of $\tilde{A}_t$ and let $\alpha_{i,j} = \langle v_i, \tilde{A}_t v_j \rangle$ with

$$d\alpha_{i,j} = \int_{\mathbb{R}^d} \langle x, v_i \rangle \langle x, v_j \rangle \langle C_t x, dB_t \rangle \mu_t(dx + a_t),$$

which follows from (12). Also define

$$\xi_{i,j} = \frac{1}{\sqrt{\alpha_{i,i}\alpha_{j,j}}} \int_{\mathbb{R}^d} \langle x, v_i \rangle \langle x, v_j \rangle C_t x \mu_t(dx + a_t).$$

So that

$$d\alpha_{i,j} = \sqrt{\alpha_{i,i}\alpha_{j,j}} \langle \xi_{i,j}, dB_t \rangle, \qquad \frac{d}{dt}[\alpha_{i,j}]_t = \alpha_{i,i}\alpha_{j,j} \|\xi_{i,j}\|^2.$$

Now, since v_i is an eigenvector corresponding to the eigenvalue $\alpha_{i,i}$, we have

$$\xi_{i,j} = \int_{\mathbb{R}^d} \langle \tilde{A}_t^{-1/2} x, v_i \rangle \langle \tilde{A}_t^{-1/2} x, v_j \rangle C_t x \mu_t(dx + a_t).$$

If we define the measure $\tilde{\mu}_t(dx) = \det(\tilde{A}_t)^{1/2} \mu_t(\tilde{A}_t^{1/2} dx + a_t)$, then $\tilde{\mu}_t$ has the law of a centered log-concave random vector with covariance $\tilde{A}_t^{-1/2} A_t \tilde{A}_t^{-1/2} \preceq \mathrm{I}_d$. By making the substitution $y = \tilde{A}_t^{-1/2} x$, the above expression becomes

$$\xi_{i,j} = \int_{\mathbb{R}^d} \langle y, v_i \rangle \langle y, v_j \rangle C_t \tilde{A}_t^{1/2} y \tilde{\mu}_t(dy).$$

By (11) and the definition of T, C_t, for any $t \le T$, $\tilde{A}_t^{1/2} \preceq 2\mathrm{I}_d$ and $C_t \preceq \mathrm{I}_d$. So, $\|C_t \tilde{A}_t^{1/2}\|_{\mathrm{op}} \le 2$. Under similar conditions, it was shown in [20], Lemma 3.2, that there exists a universal constant $C > 0$ for which:

- for any $1 \le i \le d$, $\|\xi_{i,i}\|^2 \le C$,
- for any $1 \le i \le d$, $\sum_{j=1}^{d} \|\xi_{i,j}\|^2 \le C\kappa_d^2$.

Furthermore, in the proof of Proposition 3.1 in the same paper it was shown

$$d\,\mathrm{Tr}\big(\tilde{A}_t^{4\ln(d)}\big) \le 4\ln(d) \sum_{i=1}^{d} \alpha_{i,i}^{4\ln(d)} \langle \xi_{i,i}, dB_t\rangle + 16C\kappa_d^2 \ln(d)^2 \,\mathrm{Tr}\big(\tilde{A}_t^{4\ln(d)}\big)\,dt.$$

So, using Itô's formula with the function $\ln(x)$ we can calculate the differential of the Doob decomposition (13). Specifically, we use the fact that the second derivative of $\ln(x)$ is negative and get

$$dE_t \le 16C\kappa_d^2 \ln(d)^2 \frac{\mathrm{Tr}(\tilde{A}_t^{4\ln(d)})}{\mathrm{Tr}(\tilde{A}_t^{4\ln(d)})} = 16C\kappa_d^2 \ln(d)^2, \quad E_0 = \ln(d),$$

and

$$\frac{d}{dt}[M]_t \le 16C^2 \ln(d)^2 \left(\frac{\mathrm{Tr}(\tilde{A}_t^{4\ln(d)})}{\mathrm{Tr}(\tilde{A}_t^{4\ln(d)})}\right)^2 = 16C^2 \ln(d)^2. \tag{14}$$

Hence, $E_t \le t \cdot 16C\kappa_n^2 \ln(d)^2 + \ln(d)$, which together with (13) gives

$$\mathbb{P}(\tau > s) \le \mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} M_t \ge 2\ln(d) - \ln(d) - 80s^{-1} C\kappa_d^2 \ln(d)^2\Big) \quad \forall s \ge 5.$$

Under the assumption $s > 80C\kappa_d^2 \ln(d)^2$, and since $d \ge 8$, the above can simplify to

$$\mathbb{P}(\tau > s) \le \mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} M_t \ge \frac{1}{2}\ln(d)\Big). \tag{15}$$

To bound this last expression, we will apply the Dubins–Schwarz theorem to write

$$M_t = W_{[M]_t},$$

where W_t is some Brownian motion. Combining this with (15) gives

$$\mathbb{P}(\tau > s) \le \mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} W_{[M]_t} \ge \frac{\ln(d)}{2}\Big).$$

An application of Doob's maximal inequality ([41] Proposition I.1.8) shows that for any $t', K > 0$,

$$\mathbb{P}\Big(\max_{0 \le t \le t'} W_t \ge K\Big) \le \exp\Big(-\frac{K^2}{2t'}\Big).$$

We now integrate (14) and use the above inequality to obtain

$$\mathbb{P}\Big(\max_{0 \le t \le 5s^{-1}} W_{[M]_t} \ge \frac{\ln(d)}{2}\Big) \le e^{-cs},$$

where $c > 0$ is some universal constant. □

PROOF OF THEOREM 2. By definition of T and C_t, we have that for any $t \le T$, $A_t C_t \preceq 2\mathrm{I}_d$ and for any $t > T$, $A_t C_t = A_t A_t^{\dagger} \preceq \mathrm{I}_d$. We now invoke Theorem 10, with $\Gamma_t = A_t C_t$, for which

$$\mathrm{Tr}\big(\mathbb{E}[\Gamma_t^4]\mathbb{E}[\Gamma_t^2]^{\dagger}\big) \le 4d$$

and, by Lemma 8

$$\mathrm{Tr}\big(\mathbb{E}[\Gamma_t^2]\big) \le 4d\mathbb{P}(\tau > t) \le 4de^{-ct} \quad \forall t > C \cdot \kappa_d^2 \ln(d)^2.$$

If G is the standard d-dimensional Gaussian, then the theorem yields

$$\begin{aligned}\mathcal{W}_2^2(S_n, G) &\le \int_0^{C\cdot\kappa_d^2 \ln(d)^2 \ln(n)} 4\frac{d}{n}\,dt + \int_{C\cdot\kappa_d^2 \ln(d)^2 \ln(n)}^{\infty} 16d\mathbb{P}(\tau > t) \\ &\le 4\frac{dC\cdot\kappa_d^2 \ln(d)^2 \ln(n)}{n} + 16d\int_{C\cdot\kappa_d^2 \ln(d)^2 \ln(n)}^{\infty} e^{-ct}\,dt \\ &\le C'\frac{d\cdot\kappa_d^2 \ln(d)^2 \ln(n)}{n}.\end{aligned}$$

Thus

$$\mathcal{W}_2(S_n, G) \le \frac{C\kappa_d \ln(d)\sqrt{d\ln(n)}}{\sqrt{n}}, \qquad \square$$

4. Convergence rates in entropy. Throughout this section, we fix a centered measure μ on $\mathbb{R}^d$ with an invertible covariance matrix Σ and $G \sim \mathcal{N}(0, \Sigma)$. Let $\{X^{(i)}\}$ be independent copies of $X \sim \mu$ and $S_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n X^{(i)}$.

Our goal is to study the quantity $\mathrm{Ent}(S_n||G)$. In light of Theorem 11, we aim to construct a martingale embedding $(X_t, \Gamma_t, 1)$ such that $X_1 \sim \mu$ and which satisfies appropriate bounds on the matrix Γ_t. Our construction uses the process a_t from Proposition 1 with the choice $C_t := \frac{1}{1-t}\mathrm{I}_d$. Property 1 in Proposition 1 gives

$$a_t = \int_0^t \frac{A_s}{1-s}\,dB_s.$$

Thus, we denote

$$\Gamma_t := \frac{A_t}{1-t}.$$

Since $\int_0^1 \lambda_{\min}(C_t^2) = \infty$, Proposition 4 shows that the triplet $(a_t, \Gamma_t, 1)$ is a martingale embedding of μ. As above, the sequence $\Gamma_t^{(i)}$ will denote independent copies of Γ_t and we define $\tilde{\Gamma}_t := \sqrt{\sum_{i=1}^n (\Gamma_t^{(i)})^2}$.

4.1. *Properties of the embedding.* The martingale embedding has several useful properties which we record in this section. First, we give an alternative description of the process which will be of use for us. Define the random process

$$v := \arg\min_u \frac{1}{2}\int_0^1 \mathbb{E}[\|u_t\|^2],$$

where u varies over all $\mathcal{F}_t$-adapted drifts such that $B_1 + \int_0^1 u_t\,dt \sim \mu$. Denote

$$Y_t := B_t + \int_0^t v_s ds.$$

In [23] (Section 2.2) it was shown that the density of the measure $Y_1|\mathcal{F}_t$ has the same dynamics as the density of μ_t. Thus, almost surely $Y_1|\mathcal{F}_t \sim \mu_t$ and since a_t is the expectation of μ_t, we have the identity

$$a_t = \mathbb{E}[Y_1|\mathcal{F}_t], \tag{16}$$

and in particular we have $a_1 = Y_1$. Moreover, the same reasoning implies that $A_t = \mathrm{Cov}(Y_1|\mathcal{F}_t)$ and

$$\Gamma_t = \frac{\mathrm{Cov}(Y_1|\mathcal{F}_t)}{1-t}. \tag{17}$$

The process Y_t goes back at least to the works of Föllmer [29, 30]. In a later work, by Lehec [36], it is shown that v_t is a martingale and that

$$\mathrm{Ent}(Y_1||\gamma) = \frac{1}{2}\int_0^1 \mathbb{E}\big[\|v_t\|^2\big]\,dt, \tag{18}$$

where γ denotes the standard Gaussian.

LEMMA 9. *It holds that* $\frac{d}{dt}\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] = -\mathbb{E}[\Gamma_t^2]$.

PROOF. From (16), we have

$$\mathrm{Cov}(Y_1|\mathcal{F}_t) = \mathbb{E}\big[Y_1^{\otimes 2}|\mathcal{F}_t\big] - \mathbb{E}[Y_1|\mathcal{F}_t]^{\otimes 2} = \mathbb{E}\big[Y_1^{\otimes 2}|\mathcal{F}_t\big] - a_t^{\otimes 2}.$$

a_t is a martingale, hence

$$\frac{d}{dt}\mathbb{E}\big[\mathrm{Cov}(Y_1|\mathcal{F}_t)\big] = -\frac{d}{dt}\mathbb{E}\big[[a]_t\big] = -\mathbb{E}\big[\Gamma_t^2\big]. \tag{19}$$

□

Our next goal is to recover v_t from the martingale a_t.

LEMMA 10. *The drift v_t satisfies that identity $v_t = \int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s}\,dB_s$. Furthermore,*

$$\mathbb{E}\big[\|v_t\|^2\big] = \int_0^t \frac{\mathrm{Tr}(\mathbb{E}[(\Gamma_s - \mathrm{I}_d)^2])}{(1-s)^2}\,ds. \tag{20}$$

PROOF. We begin by writing

$$da_t = dB_t + (\Gamma_t - \mathrm{I}_d)\,dB_t.$$

Using Fubini's theorem then yields

$$\int_0^1 (\Gamma_s - \mathrm{I}_d)\,dB_s = \int_0^1\int_s^1 \frac{\Gamma_s - \mathrm{I}_d}{1-s}\,dt\,dB_s = \int_0^1\int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s}\,dB_s\,dt.$$

Therefore, defining $\tilde{v}_t = \int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s}\,dB_s$ we have that $\tilde{v}_t$ is a martingale, and that $B_1 + \int_0^1 \tilde{v}_t\,dt = a_1$. It follows that $v_t - \tilde{v}_t$ is a martingale and that $\int_0^1 (v_t - \tilde{v}_t)\,dt = 0$. We will now show that if a martingale Q_t satisfies $Q_0 = 0$ and $\int_0^1 Q_t\,dt = 0$ a.s., then $Q_t = 0$ for every $t \in [0, 1]$. From this, it will follow that $v_t = \tilde{v}_t$. Indeed, write $Q_t = \int_0^t Q'_s\,dB_s$, for some adapted process Q'_t. Using Fubini's theorem, a calculation similar to the one above gives the identity

$$0 = \int_0^1 Q_t\,dt = \int_0^1 (1-t)Q'_t\,dB_t.$$

Considering the martingale $\int_0^{\cdot} (1-t)Q'_t\,dB_t$, we now have, for any $s \in [0, 1)$

$$0 = \mathbb{E}\Big[\int_0^1 (1-t)Q'_t\,dB_t\Big|\mathcal{F}_s\Big] = \int_0^s (1-t)Q'_t\,dB_t.$$

Thus, $Q' = 0$ almost surely, which implies, for every $t \in [0, 1]$, $Q_t = Q_0 = 0$. Therefore $v_t = \tilde{v}_t$, or in other words

$$v_t = \int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s}\,dB_s.$$

Finally, equation (20) follows from a direct application of Itô's isometry. □

A combination of equations (18) and (20) gives the useful identity,

$$\mathrm{Ent}(Y_1||\gamma) = \frac{1}{2}\int_0^1 \int_0^t \frac{\mathrm{Tr}(\mathbb{E}[(\Gamma_s - \mathrm{I}_d)^2])}{(1-s)^2}\,ds\,dt = \frac{1}{2}\int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[(\Gamma_t - \mathrm{I}_d)^2])}{1-t}\,dt. \tag{21}$$

The above lemma also affords a representation of $\mathbb{E}[\mathrm{Tr}(\Gamma_t)]$ in terms of $\mathbb{E}[\|v_t\|^2]$.

LEMMA 11. *It holds that*

$$\mathbb{E}\big[\mathrm{Tr}(\Gamma_t)\big] = d - (1-t)\big(d - \mathrm{Tr}(\Sigma) + \mathbb{E}\big[\|v_t\|^2\big]\big).$$

PROOF. The identity can be obtained through integration by parts. By Lemma 10,

$$\begin{aligned}\mathbb{E}\big[\|v_t\|^2\big] &\overset{(20)}{=} \int_0^t \frac{\mathrm{Tr}(\mathbb{E}[(\Gamma_s - \mathrm{I}_d)^2])}{(1-s)^2}\,ds \\ &= \int_0^t \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_s^2])}{(1-s)^2}\,ds - 2\int_0^t \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_s])}{(1-s)^2}\,ds + \int_0^t \frac{\mathrm{Tr}(\mathrm{I}_d)}{(1-s)^2}\,ds.\end{aligned}$$

Since, by Lemma 9, $\frac{d}{dt}\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] = -\mathbb{E}[\Gamma_t^2]$ integration by parts shows

$$\begin{aligned}\int_0^t \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_s^2])}{(1-s)^2}\,ds &= -\frac{\mathrm{Tr}(\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_s)])}{(1-s)^2}\bigg|_0^t + 2\int_0^t \frac{\mathrm{Tr}(\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_s)])}{(1-s)^3}\,ds \\ &= \mathrm{Tr}(\Sigma) - \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_t])}{1-t} + 2\int_0^t \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_s])}{(1-s)^2}\,ds,\end{aligned}$$

where we have used (17) and the fact $\mathrm{Cov}(Y_1|\mathcal{F}_0) = \mathrm{Cov}(Y_1) = \Sigma$. Plugging this into the previous equation shows

$$\mathbb{E}\big[\|v_t\|^2\big] = \mathrm{Tr}(\Sigma) - \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_t])}{1-t} + \frac{d}{1-t} - d,$$

or equivalently

$$\mathbb{E}\big[\mathrm{Tr}(\Gamma_t)\big] = d - (1-t)\big(d - \mathrm{Tr}(\Sigma) + \mathbb{E}\big[\|v_t\|^2\big]\big). \qquad \square$$

Next, as in Theorem 11, we define σ_t to be the minimal eigenvalue of $\mathbb{E}[\Gamma_t]$, so that

$$\mathbb{E}[\Gamma_t] \succeq \sigma_t \mathrm{I}_d.$$

Note that by Jensen's inequality we also have

$$\mathbb{E}\big[\Gamma_t^2\big] \succeq \sigma_t^2 \mathrm{I}_d. \tag{22}$$

LEMMA 12. *Assume that* $\mathrm{Ent}(Y_1||\gamma) < \infty$. *Then* Γ_t *is almost surely invertible for all* $t \in [0, 1)$ *and, moreover, there exists a constant* $m = m_\mu > 0$ *for which*

$$\sigma_t \geq m \quad \forall t \in [0, 1).$$

PROOF. We will show that for every $0 \leq t < 1$, $\sigma_t > 0$ and that there exists $c > 0$ such that $\sigma_t > \frac{1}{8}$ whenever $t > 1 - c$. The claim will then follow by continuity of σ_t. The key to showing this is identity (21), due to which,

$$\mathrm{Ent}(Y_1||\gamma) = \frac{1}{2}\int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[(\Gamma_t - \mathrm{I}_d)^2])}{1-t}\,dt.$$

Recall that, by Equation (17), $\Gamma_t = \frac{\mathrm{Cov}(Y_1|\mathcal{F}_t)}{1-t}$ and observe that, by Proposition 5, if $\mathrm{Cov}(Y_1|\mathcal{F}_s)$ is not invertible for some $0 \leq s < 1$ then $\mathrm{Cov}(Y_1|\mathcal{F}_t)$ is also not invertible for any $t > s$. Under this event, we would have that $\int_s^1 \frac{\mathrm{Tr}((\Gamma_t - \mathrm{I}_d)^2)}{1-t}\,dt = \infty$ which, using the above display, implies that the probability of this event must be zero. Therefore, Γ_t is almost surely invertible and $\sigma_t > 0$ for all $t \in [0, 1)$.

Suppose now that for some $t' \in [0, 1]$, $\sigma_{t'} \leq \frac{1}{8}$. By Jensen's inequality, we have

$$\mathrm{Tr}\big(\mathbb{E}\big[(\Gamma_t - \mathrm{I}_d)^2\big]\big) \geq \mathrm{Tr}\big(\mathbb{E}[\Gamma_t - \mathrm{I}_d]^2\big) \geq (1-\sigma_t)^2 \geq 1 - 2\sigma_t.$$

Since, by Lemma 9, $\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)]$ is nonincreasing, for any $t' \leq t \leq t' + \frac{1-t'}{2}$,

$$\sigma_t \leq \frac{\sigma_{t'}(1-t')}{1-t} \leq \frac{1-t'}{8(1-t'-\frac{1-t'}{2})} = \frac{1}{4}.$$

Now, assume by contradiction that there exists a sequence $t_i \in (0, 1)$ such that $\sigma_{t_i} \leq \frac{1}{8}$ and $\lim_{i\to\infty} t_i = 1$. By passing to a subsequence we may assume that $t_{i+1} - t_i \geq \frac{1-t_i}{2}$ for all i. The assumption $\mathrm{Ent}(Y_1||\gamma) < \infty$ combined with Equation (21) and with the last two displays finally gives

$$\infty > \int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[(\Gamma_t - \mathrm{I}_d)^2])}{1-t}\,dt \geq \int_0^1 \frac{1-2\sigma_t}{1-t}\,dt \geq \sum_{i=1}^{\infty} \int_{t_i}^{t_i + \frac{1-t_i}{2}} \frac{1}{2}\frac{1}{1-t}\,dt \geq \log 2 \sum_{i=1}^{\infty} \frac{1}{2},$$

which leads to a contradiction and completes the proof. $\square$

4.2. *Proof of Theorem* 5. Thanks to the assumption $\mathrm{Ent}(Y_1||G) < \infty$, an application of Lemma 12 gives that Γ_t is invertible almost surely, so we may invoke the second bound in Theorem 11 to obtain

$$\mathrm{Ent}(S_n||G) \leq \int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_t^2] - \mathbb{E}[\tilde{\Gamma}_t]^2)}{(1-t)^2}\left(\int_t^1 \sigma_s^{-2}\,ds\right)dt.$$

The same lemma also shows that for some $m > 0$ one has

$$\int_t^1 \sigma_s^{-2}\,ds \leq \frac{1-t}{m^2}.$$

Therefore, we attain that

$$\mathrm{Ent}(S_n||G) \leq \frac{1}{m^2}\int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_t^2] - \mathbb{E}[\tilde{\Gamma}_t]^2)}{1-t}\,dt. \tag{23}$$

Next, observe that, by Itô's isometry,

$$\mathrm{Cov}(X) = \int_0^1 \mathbb{E}[\Gamma_t^2]\,dt.$$

Hence, as long as $\mathrm{Cov}(X)$ is finite, $\mathbb{E}[\Gamma_t^2]$ is also finite for all $t \in A$ where $[0, 1] \setminus A$ is a set of measure 0. We will use this fact to show that

$$\lim_{n\to\infty} \mathrm{Tr}\big(\mathbb{E}[\Gamma_t^2] - \mathbb{E}[\tilde{\Gamma}_t]^2\big) = 0 \quad \forall t \in A. \tag{24}$$

Indeed, by the law of large numbers, $\tilde{\Gamma}_t$ almost surely converges to $\sqrt{\mathbb{E}[\Gamma_t^2]}$. Since $(\Gamma_t^{(i)})^2$ are integrable, we get that the sequence $\frac{1}{n}\sum_{i=1}^n (\Gamma_t^{(i)})^2$ is uniformly integrable. We now use the inequality

$$\tilde{\Gamma}_t \preceq \sqrt{\frac{1}{n}\sum_{i=1}^n (\Gamma_t^{(i)})^2 + \mathrm{I}_d} \preceq \frac{1}{n}\sum_{i=1}^n (\Gamma_t^{(i)})^2 + \mathrm{I}_d,$$

to deduce that $\tilde{\Gamma}_t$ is uniformly integrable as well. An application of Vitali's convergence theorem (see, e.g., [28]) implies (24).

We now know that the integrand in the right hand side of (23) convergence to zero for almost every t. It remains to show that the expression converges as an integral, for which we intend to apply the dominated convergence theorem. It thus remains to show that the expression

$$\frac{\mathrm{Tr}(\mathbb{E}[\Gamma_t^2] - \mathbb{E}[\tilde{\Gamma}_t]^2)}{1-t}$$

is bounded by an integrable function, uniformly in n, which would imply that

$$\lim_{n\to\infty} \mathrm{Ent}(S_n||G) = 0,$$

and the proof would be complete. To that end, recall that the square root function is concave on positive definite matrices (see, e.g., [1]), thus

$$\tilde{\Gamma}_t \succeq \frac{1}{n}\sum_{i=1}^{n} \Gamma_t^{(i)}.$$

It follows that

$$\mathrm{Tr}\big(\mathbb{E}[\Gamma_t^2] - \mathbb{E}[\tilde{\Gamma}_t]^2\big) \le \mathrm{Tr}\big(\mathbb{E}[\Gamma_t^2] - \mathbb{E}[\Gamma_t]^2\big) \le \mathrm{Tr}\big(\mathbb{E}\big[(\Gamma_t - \mathrm{I}_d)^2\big]\big).$$

So we have

$$\begin{aligned}\frac{1}{m^2}\int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[\Gamma_t^2] - \mathbb{E}[\tilde{\Gamma}_t]^2)}{1-t}\,dt &\le \frac{1}{m^2}\int_0^1 \frac{\mathrm{Tr}(\mathbb{E}[(\Gamma_t - \mathrm{I}_d)^2])}{1-t}\,dt \\ &\overset{(21)}{=} \frac{2}{m^2}\mathrm{Ent}(Y_1||\gamma) < \infty.\end{aligned}$$

This completes the proof.

4.3. *Quantitative bounds for log concave random vectors.* In this section, we make the additional assumption that the measure μ is log concave. Under this assumption, we show how one can obtain explicit convergence rates in the central limit theorem. Our aim is to use the bound in Theorem 11 for which we are required to obtain bounds on the process Γ_t. We begin by recording several useful facts concerning this process.

LEMMA 13. *The process Γ_t has the following properties*:

1. *If μ is log concave, then for every $t \in [0, 1]$, $\Gamma_t \preceq \frac{1}{t}\mathrm{I}_d$, almost surely.*
2. *If μ is also 1-uniformly log concave, then for every $t \in [0, 1]$, $\Gamma_t \preceq \mathrm{I}_d$ almost surely.*

PROOF. Denote by ρ_t the density of $Y_1|\mathcal{F}_t$ with respect to the Lebesgue measure with $\rho := \rho_0$ being the density of μ. By Proposition 2 with $C_t = \frac{\mathrm{I}_d}{1-t}$, we can calculate the ratio between ρ_t and ρ. In particular, we have

$$\frac{d}{dt}\Sigma_t^{-1} = -\Sigma_t^{-1}\left(\frac{d}{dt}\Sigma_t\right)\Sigma_t^{-1} = \frac{1}{(1-t)^2}\mathrm{I}_d.$$

Solving this differential equation with the initial condition $\Sigma_0^{-1} = 0$, we find that $\Sigma_t^{-1} = \frac{t}{1-t}\mathrm{I}_d$.

Since the ratio between ρ_t and ρ is proportional to the density of a Gaussian with covariance Σ_t, we thus have

$$-\nabla^2\log(\rho_t) = -\nabla^2\log(\rho) + \frac{t}{1-t}\mathrm{I}_d.$$

Now, if μ is log concave then $Y_1|\mathcal{F}_t$ is almost surely $\frac{t}{1-t}$-uniformly log-concave. By the Brascamp-Lieb inequality (as in [32]) we get $\mathrm{Cov}(Y_1|\mathcal{F}_t) \preceq \frac{1-t}{t}\mathrm{I}_d$ and, using (17),

$$\Gamma_t \preceq \frac{1}{t}\mathrm{I}_d.$$

If μ is also 1-uniformly log-concave then $-\nabla^2 \log(\rho) \succeq \mathrm{I}_d$ and almost surely

$$-\nabla^2 \log(\rho_t) \succeq \frac{1}{1-t}\mathrm{I}_d.$$

By the same argument this implies

$$\Gamma_t \preceq \mathrm{I}_d. \qquad \square$$

The relative entropy to the Gaussian of a log concave measure with nondegenerate covariance structure is finite (it is even universally bounded, see [37]). Thus, by Lemma 12, it follows that Γ_t is invertible almost surely. This allows us to invoke the first bound of Theorem 11,

$$\mathrm{Ent}(S_n||G) \le \frac{1}{n}\int_0^1 \frac{\mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)]}{(1-t)^2\sigma_t^2}\left(\int_t^1 \sigma_s^{-2}\,ds\right)dt. \tag{25}$$

Attaining an upper bound on the right-hand side amounts to a concentration estimate for the process Γ_t^2 and a lower bound on σ_t. These two tasks are the objective of the following two lemmas.

LEMMA 14. *If μ is log concave and isotropic, then for any $t \in [0, 1)$,*

$$\mathrm{Tr}\big(\mathbb{E}\big[(\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2\big]\big) \le \frac{1-t}{t^2}\left(\frac{d(1+t)}{t^2} + 2\mathbb{E}\big[\|v_t\|^2\big]\right)$$

and

$$\mathrm{Tr}\big(\mathbb{E}\big[(\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2\big]\big) \le C\frac{d^4}{(1-t)^4}$$

for a universal constant $C > 0$.

PROOF. The isotropicity of μ, used in conjunction with the formula given in Lemma 11, yields

$$\mathrm{Tr}\big(\mathbb{E}[\Gamma_t^2]\big) \ge \frac{1}{d}\mathrm{Tr}\big(\mathbb{E}[\Gamma_t]\big)^2 \ge d - 2(1-t)\mathbb{E}\big[\|v_t\|^2\big],$$

where the first inequality follows by convexity. Since μ is log concave, Lemma 13 ensures that, almost surely, $\Gamma_t \preceq \frac{1}{t}\mathrm{I}_d$. Therefore,

$$\begin{aligned}
\mathrm{Tr}\big(\mathbb{E}\big[(\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2\big]\big) &\le \mathrm{Tr}\left(\mathbb{E}\left[\left(\Gamma_t^2 - \frac{1}{t^2}\mathrm{I}_d\right)^2\right]\right)\\
&= \frac{1}{t^4}\mathrm{Tr}\big(\mathbb{E}\big[(\mathrm{I}_d - t^2\Gamma_t^2)^2\big]\big)\\
&\le \frac{1}{t^4}\mathrm{Tr}\big(\mathbb{E}\big[\mathrm{I}_d - t^2\Gamma_t^2\big]\big)\\
&\le \frac{1-t}{t^2}\left(\frac{d(1+t)}{t^2} + 2\mathbb{E}\big[\|v_t\|^2\big]\right).
\end{aligned}$$

Which proves the first bound. Towards the second bound, we use (17) to write

$$\Gamma_t^2 \preceq \frac{1}{(1-t)^2}\mathbb{E}[Y_1^{\otimes 2}|\mathcal{F}_t]^2.$$

So,

$$\mathbb{E}[\|\Gamma_t^2\|_{\mathrm{HS}}^2] \leq \frac{1}{(1-t)^4}\mathbb{E}[\|\|Y_1\|^2 Y_1^{\otimes 2}\|_{\mathrm{HS}}^2] \leq \frac{1}{(1-t)^4}\mathbb{E}[\|Y_1\|^8].$$

For an isotropic log concave measure, the expression $\mathbb{E}[\|Y_1\|^8]$ is bounded from above by Cd^4 for a universal constant $C > 0$ (see [40]). Thus,

$$\mathrm{Tr}(\mathbb{E}[(\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2]) = \mathbb{E}[\|\Gamma_t^2 - \mathbb{E}[\Gamma_t^2]\|_{\mathrm{HS}}^2] \leq 2\mathbb{E}[\|\Gamma_t^2\|_{\mathrm{HS}}^2] \leq C\frac{d^4}{(1-t)^4}. \qquad \square$$

LEMMA 15. *Suppose that μ is log concave and isotropic, then there exists a universal constant $1 > c > 0$ such that*:

1. *For any*, $t \in [0, \frac{c}{d^2}]$, $\sigma_t \geq \frac{1}{2}$.
2. *For any*, $t \in [\frac{c}{d^2}, 1]$, $\sigma_t \geq \frac{c}{td^2}$.

PROOF. By Lemma 9, we have

$$\frac{d}{dt}\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] = -\mathbb{E}[\Gamma_t^2] \stackrel{(17)}{=} -\frac{\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)^2]}{(1-t)^2}.$$

Moreover, by convexity,

$$\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)^2] \preceq \mathbb{E}[\mathbb{E}[Y_1^{\otimes 2}|\mathcal{F}_t]^2] \preceq \mathbb{E}[\|Y_1\|^4]\mathrm{I}_d.$$

It is known (see [40]) then when μ is log concave and isotropic there exists a universal constant $C > 0$ such that

$$\mathbb{E}[\|Y_1\|^4] \leq Cd^2.$$

Consequently, $\frac{d}{dt}\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] \succeq -\frac{Cd^2}{(1-t)^2}\mathrm{I}_d$, and since $\mathrm{Cov}(Y_1|\mathcal{F}_0) = \mathrm{I}_d$,

$$\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] \succeq \left(1 - Cd^2\int_0^t \frac{1}{(1-s)^2}\,ds\right)\mathrm{I}_d = \left(1 - \frac{Cd^2 t}{1-t}\right)\mathrm{I}_d.$$

By increasing the value of C, we may legitimately assume that $\frac{1}{Cd^2} \leq 1$, thus for any $t \in [0, \frac{1}{3Cd^2}]$ we get that

$$\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] \succeq \frac{1}{2}\mathrm{I}_d,$$

which implies $\sigma_t \geq \frac{1}{2}$ and completes the first part of the lemma. In order to prove the second part, we first write

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[\Gamma_t] &= \frac{d}{dt}\frac{\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)]}{1-t} \stackrel{(\text{Lemma } 9)}{=} \frac{\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] - (1-t)\mathbb{E}[\Gamma_t^2]}{(1-t)^2} \\ &= \frac{\mathbb{E}[\Gamma_t] - \mathbb{E}[\Gamma_t^2]}{1-t}.\end{aligned} \tag{26}$$

Since, by Lemma 13, $\Gamma_t \preceq \frac{1}{t}\mathrm{I}_d$, we have the bound

$$\frac{\mathbb{E}[\Gamma_t] - \mathbb{E}[\Gamma_t^2]}{1-t} \succeq \frac{1-\frac{1}{t}}{1-t}\mathbb{E}[\Gamma_t] = -\frac{1}{t}\mathbb{E}[\Gamma_t].$$

Now, consider the differential equation $f'(t) = \frac{-f(t)}{t}$, $f(\frac{1}{3Cd^2}) = \frac{1}{2}$. Its unique solution is $f(t) = \frac{1}{6Cd^2 t}$. Thus, Gromwall's inequality shows that $\sigma_t \geq \frac{1}{6Cd^2 t}$, which concludes the proof. $\square$

PROOF OF THEOREM 6. Our objective is to bound from above the right-hand side of Equation (25). As a consequence of Lemma 15, we have that, for any $t \in [0, 1)$,

$$\int_t^1 \sigma_s^{-2} \, ds \leq Cd^4(1-t),$$

for some universal constant $C > 0$. It follows that the integral in (25) admits the bound

$$\int_0^1 \frac{\mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)]}{(1-t)^2 \sigma_t^2} \left(\int_t^1 \sigma_s^{-2} \, ds \right) dt \leq Cd^4 \int_0^1 \frac{\mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)]}{(1-t)\sigma_t^2} \, dt.$$

Next, there exists a universal constant $C' > 0$ such that

$$Cd^4 \int_0^{cd^{-2}} \frac{\mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)]}{(1-t)\sigma_t^2} \, dt \leq C' \int_0^{cd^{-2}} \frac{d^8}{(1-t)^5} \, dt \leq C' d^8,$$

where we have used the second bound of Lemma 14 and the first bound of Lemma 15. Also, by applying the second bound of Lemma 15 when $t \in [cd^{-2}, d^{-1}]$ we get

$$Cd^4 \int_{cd^{-2}}^{d^{-1}} \frac{\mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)]}{(1-t)\sigma_t^2} \, dt \leq C' \int_{cd^{-2}}^{d^{-1}} \frac{d^{12} t^2}{(1-t)^5} \, dt \leq C' d^9.$$

Finally, when $t > d^{-1}$, we have

$$\begin{aligned}
Cd^4 \int_{d^{-1}}^1 \frac{\mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)]}{(1-t)\sigma_t^2} \, dt &\leq C' d^8 \int_{d^{-1}}^1 \frac{t^2 \mathbb{E}[\mathrm{Tr}((\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2)]}{1-t} \, dt \\
&\leq 2C' d^9 \int_{d^{-1}}^1 \left(\frac{1}{t^2} + \mathbb{E}[\|v_t\|^2] \right) dt \\
&\overset{(18)}{\leq} 4C' d^{10} (1 + \mathrm{Ent}(Y_1 || G)),
\end{aligned}$$

where the first inequality uses Lemma 15 and the second one uses Lemma 14. This establishes

$$\mathrm{Ent}(S_n || G) \leq \frac{Cd^{10}(1 + \mathrm{Ent}(Y_1 || G))}{n}. \qquad \square$$

Finally, we derive an improved bound for the case of 1-uniformly log concave measures, based on the following estimates.

LEMMA 16. *Suppose that μ is* 1*-uniformly log concave, then for every* $t \in [0, 1)$:

1. $\mathrm{Tr}(\mathbb{E}[(\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2]) \leq 2(1-t)(d - \mathrm{Tr}(\Sigma) + \mathbb{E}[\|v_t\|^2])$.
2. $\sigma_t \geq \sigma_0$.

PROOF. By Lemma 13, we have that $\Gamma_t \preceq \mathrm{I}_d$ almost surely. Using this together with the identity given by Lemma 11, and proceeding in similar fashion to Lemma 14 we obtain

$$\mathrm{Tr}(\mathbb{E}[\Gamma_t^2]) \geq \frac{1}{d} \mathrm{Tr}(\mathbb{E}[\Gamma_t])^2 \geq d - 2(1-t)(d - \mathrm{Tr}(\Sigma) + \mathbb{E}[\|v_t\|^2])$$

and

$$\begin{aligned}\mathrm{Tr}\big(\mathbb{E}\big[\big(\Gamma_t^2-\mathbb{E}\big[\Gamma_t^2\big]\big)^2\big]\big) &\le \mathrm{Tr}\big(\mathbb{E}\big[\big(\Gamma_t^2-\mathrm{I}_d\big)^2\big]\big) \le \mathrm{Tr}\big(\mathbb{E}\big[\mathrm{I}_d-\Gamma_t^2\big]\big)\\ &\le 2(1-t)\big(d-\mathrm{Tr}(\Sigma)+\mathbb{E}\big[\|v_t\|^2\big]\big).\end{aligned}$$

Also, recalling (26) and since $\Gamma_t \preceq \mathrm{I}_d$ we get

$$\frac{d}{dt}\mathbb{E}[\Gamma_t]=\frac{\mathbb{E}[\Gamma_t]-\mathbb{E}[\Gamma_t^2]}{1-t}\ge 0,$$

which shows that σ_t is bounded from below by a nondecreasing function and so $\sigma_t \ge \sigma_0$ which is the minimal eigenvalue of Σ. □

PROOF OF THEOREM 7. Plugging the bounds given in Lemma 16 into equation (25) yields

$$\begin{aligned}\mathrm{Ent}(S_n\|G) &\le \frac{1}{n}\int_0^1 \frac{\mathbb{E}[\mathrm{Tr}((\Gamma_t^2-\mathbb{E}[\Gamma_t^2])^2)]}{(1-t)^2\sigma_t^2}\left(\int_t^1 \sigma_s^{-2}\,ds\right)dt\\ &\le \frac{2(d+\int_0^1\mathbb{E}[\|v_t\|^2]\,dt)}{\sigma_0^4 n} \overset{(18)}{=} \frac{2(d+2\,\mathrm{Ent}(X\|\gamma))}{\sigma_0^4 n},\end{aligned}$$

which completes the proof. □

Acknowledgments. We are extremely grateful to the anonymous referee for his/her careful reading of this manuscript. His/her efforts have greatly improved the presentation and overall readability.

R. Eldan is Incumbent of the Elaine Blond career development chair. Supported by a European Research Council Starting Grant (ERC StG) and by an Israel Science Foundation Grant 715/16.

D. Mikulincer is supported by an Azrieli Fellowship award from the Azrieli Foundation.

A. Zhai is supported in part by a Stanford Graduate Fellowship.

REFERENCES

[1] ANDO, T. (1979). Concavity of certain maps on positive definite matrices and applications to Hadamard products. *Linear Algebra Appl*. **26** 203–241. MR0535686 https://doi.org/10.1016/0024-3795(79)90179-4

[2] ANTTILA, M., BALL, K. and PERISSINAKI, I. (2003). The central limit problem for convex bodies. *Trans. Amer. Math. Soc*. **355** 4723–4735. MR1997580 https://doi.org/10.1090/S0002-9947-03-03085-X

[3] ARTSTEIN, S., BALL, K. M., BARTHE, F. and NAOR, A. (2004). On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields* **129** 381–390. MR2128238 https://doi.org/10.1007/s00440-003-0329-4

[4] BALL, K., BARTHE, F. and NAOR, A. (2003). Entropy jumps in the presence of a spectral gap. *Duke Math. J*. **119** 41–63. MR1991646 https://doi.org/10.1215/S0012-7094-03-11912-2

[5] BALL, K. and NGUYEN, V. H. (2012). Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Math*. **213** 81–96. MR3024048 https://doi.org/10.4064/sm213-1-6

[6] BARRON, A. R. (1986). Entropy and the central limit theorem. *Ann. Probab*. **14** 336–342. MR0815975

[7] BENTKUS, V. (2004). A Lyapunov type bound in $\mathbf{R}^d$. *Teor. Veroyatn. Primen*. **49** 400–410. MR2144310 https://doi.org/10.1137/S0040585X97981123

[8] BERGSTRÖM, H. (1945). On the central limit theorem in the space $R_k, k > 1$. *Skand. Aktuarietidskr.* **28** 106–127. MR0015704 https://doi.org/10.1080/03461238.1945.10404921

[9] BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc*. **49** 122–136. MR0003498 https://doi.org/10.2307/1990053

[10] BHATTACHARYA, R. N. (1977). Refinements of the multidimensional central limit theorem and applications. *Ann. Probab*. **5** 1–27. MR0436273 https://doi.org/10.1214/aop/1176995887

[11] BOBKOV, S. G. (2013). Entropic approach to E. Rio's central limit theorem for W_2 transport distance. *Statist*. *Probab*. *Lett*. **83** 1644–1648. MR3062276 https://doi.org/10.1016/j.spl.2013.03.020

[12] BOBKOV, S. G. (2018). Berry–Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances. *Probab. Theory Related Fields* **170** 229–262. MR3748324 https://doi.org/10.1007/s00440-017-0756-2

[13] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2013). Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann*. *Probab*. **41** 2479–2512. MR3112923 https://doi.org/10.1214/12-AOP780

[14] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2014). Berry–Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields* **159** 435–478. MR3230000 https://doi.org/10.1007/s00440-013-0510-3

[15] BOBKOV, S. G. and KOLDOBSKY, A. (2003). On the central limit property of convex bodies. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math*. **1807** 44–52. Springer, Berlin. MR2083387 https://doi.org/10.1007/978-3-540-36428-3_5

[16] BONIS, T. (2019). Stein's method for normal approximation in Wasserstein distances with application to the multivariate central limit theorem. Preprint. Available at arXiv:1905.13615.

[17] BUBECK, S. and GANGULY, S. (2018). Entropic CLT and phase transition in high-dimensional Wishart matrices. *Int*. *Math*. *Res*. *Not*. *IMRN* **2018** 588–606. MR3801440 https://doi.org/10.1093/imrn/rnw243

[18] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann*. *Statist*. **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

[19] COURTADE, T. A., FATHI, M. and PANANJADY, A. (2019). Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **55** 777–790. MR3949953 https://doi.org/10.1214/18-aihp898

[20] ELDAN, R. (2013). Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geom*. *Funct*. *Anal*. **23** 532–569. MR3053755 https://doi.org/10.1007/s00039-013-0214-y

[21] ELDAN, R. (2016). Skorokhod embeddings via stochastic flows on the space of Gaussian measures. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **52** 1259–1280. MR3531709 https://doi.org/10.1214/15-AIHP682

[22] ELDAN, R. (2018). Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom*. *Funct*. *Anal*. **28** 1548–1596. MR3881829 https://doi.org/10.1007/s00039-018-0461-z

[23] ELDAN, R. and LEE, J. R. (2018). Regularization under diffusion and anticoncentration of the information content. *Duke Math*. *J*. **167** 969–993. MR3782065 https://doi.org/10.1215/00127094-2017-0048

[24] ELDAN, R. and MIKULINCER, D. (2016). Information and dimensionality of anisotropic random geometric graphs. Preprint. Available at arXiv:1609.02490.

[25] ESSEEN, C.-G. (1942). On the Liapounoff limit of error in the theory of probability. *Ark*. *Mat*. *Astron*. *Fys*. **28A** 19. MR0011909

[26] CHEN, L. H. and FANG, X. (2011). Multivariate normal approximation by Stein's method: The concentration inequality approach. Preprint. Available at arXiv:1111.4073.

[27] FATHI, M. (2019). Stein kernels and moment maps. *Ann*. *Probab*. **47** 2172–2185. MR3980918 https://doi.org/10.1214/18-AOP1305

[28] FOLLAND, G. B. (1999). *Real Analysis*: *Modern Techniques and Their Applications*, 2nd ed. *Pure and Applied Mathematics* (*New York*). Wiley, New York. MR1681462

[29] FÖLLMER, H. (1985). An entropy approach to the time reversal of diffusion processes. In *Stochastic Differential Systems* (*Marseille-Luminy*, 1984). *Lect*. *Notes Control Inf*. *Sci*. **69** 156–163. Springer, Berlin. MR0798318 https://doi.org/10.1007/BFb0005070

[30] FÖLLMER, H. (1986). Time reversal on Wiener space. In *Stochastic Processes—Mathematics and Physics* (*Bielefeld*, 1984). *Lecture Notes in Math*. **1158** 119–129. Springer, Berlin. MR0838561 https://doi.org/10.1007/BFb0080212

[31] GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann*. *Probab*. **19** 724–739. MR1106283

[32] HARGÉ, G. (2004). A convex/log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces. *Probab. Theory Related Fields* **130** 415–440. MR2095937 https://doi.org/10.1007/s00440-004-0365-8

[33] JOHNSON, O. and BARRON, A. (2004). Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields* **129** 391–409. MR2128239 https://doi.org/10.1007/s00440-004-0344-0

[34] KLARTAG, B. (2018). Eldan's stochastic localization and tubular neighborhoods of complex-analytic sets. *J*. *Geom*. *Anal*. **28** 2008–2027. MR3833784 https://doi.org/10.1007/s12220-017-9894-0

[35] LEE, Y. T. and VEMPALA, S. S. (2017). Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017* 998–1007. IEEE Computer Soc., Los Alamitos, CA. MR3734299 https://doi.org/10.1109/FOCS.2017.96

[36] LEHEC, J. (2013). Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 885–899. MR3112438 https://doi.org/10.1214/11-aihp464

[37] MARSIGLIETTI, A. and KOSTINA, V. (2018). A lower bound on the differential entropy of log-concave random vectors with applications. *Entropy* **20** Art. ID 185, 24. MR3782876 https://doi.org/10.3390/e20030185

[38] NAGAEV, S. V. (1976). An estimate of the remainder term in the multidimensional central limit theorem. In *Proceedings of the Third Japan–USSR Symposium on Probability Theory* (*Tashkent*, 1975). *Lecture Notes in Math.* **550** 419–438. MR0443043

[39] ØKSENDAL, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*, 6th ed. *Universitext*. Springer, Berlin. MR2001996 https://doi.org/10.1007/978-3-642-14394-6

[40] PAOURIS, G. (2006). Concentration of mass on convex bodies. *Geom. Funct. Anal.* **16** 1021–1049. MR2276533 https://doi.org/10.1007/s00039-006-0584-5

[41] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357 https://doi.org/10.1007/978-3-662-06400-9

[42] RIO, E. (2009). Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 802–817. MR2548505 https://doi.org/10.1214/08-AIHP187

[43] RIO, E. (2011). Asymptotic constants for minimal distance in the central limit theorem. *Electron. Commun. Probab.* **16** 96–103. MR2772388 https://doi.org/10.1214/ECP.v16-1609

[44] SENATOV, V. V. (1980). Some uniform estimates of the convergence rate in the multidimensional central limit theorem. *Teor. Veroyatn. Primen.* **25** 757–770. MR0595137

[45] VALIANT, G. and VALIANT, P. (2011). Estimating the unseen: An $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. In *STOC'11—Proceedings of the 43rd ACM Symposium on Theory of Computing* 685–694. ACM, New York. MR2932019 https://doi.org/10.1145/1993636.1993727

[46] ZHAI, A. (2018). A high-dimensional CLT in $\mathcal{W}_2$ distance with near optimal convergence rate. *Probab. Theory Related Fields* **170** 821–845. MR3773801 https://doi.org/10.1007/s00440-017-0771-3