THE CLT IN HIGH DIMENSIONS: QUANTITATIVE BOUNDS VIA MARTINGALE EMBEDDING

BY RONEN ELDAN*
Weizmann Institute of Science

BY DAN MIKULINCER†
Weizmann Institute of Science

BY ALEX ZHAI‡
Stanford University

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 [46], for convergence in quadratic Wasserstein transportation distance for bounded random vectors; (b) We derive the first non-asymptotic convergence rate for the entropic CLT in arbitrary dimension, for general log-concave random vectors (this adds to [20], 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 [22].

1. Introduction. Let \(X^{(1)}, \ldots, X^{(n)}\) 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 [22], 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 \(W_2(A, B)\) denote the quadratic transportation distance between two \(d\)-dimensional random vectors \(A\) and \(B\). That is,

\[
W_2(A, B) = \sqrt{\inf_{(X,Y) \sim \pi} \mathbb{E} \left[ \|X - Y\|_2^2 \right]},
\]

*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.
†Supported by an Azrieli Fellowship award from the Azrieli Foundation.
‡Partially supported by a Stanford Graduate Fellowship.

MSC 2010 subject classifications: 60F05, 60G57
Keywords 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 $\Sigma = \text{Cov}(X)$, and let $G \sim \mathcal{N}(0, \Sigma)$ be a Gaussian with covariance $\Sigma$. 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

$$W_2(S_n, G) \leq \frac{\beta \sqrt{d} \sqrt{32 + 2 \log(n)}}{\sqrt{n}}.$$ 

Theorem 1 improves a result of the third named author [46] that gives a bound of order $\beta \sqrt{d \log n} \sqrt{n}$ under the same conditions. It was noted in [46] that when $X$ is supported on a lattice $\beta \mathbb{Z}^d$, then the quantity $W_2(S_n, G)$ is of order $\beta \sqrt{d \log n} \sqrt{n}$. 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 [20], 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 $G \sim \mathcal{N}(0, I_d)$ 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

$$W_2(S_n, G) \leq \frac{C \sqrt{d} \ln(d) \sqrt{\ln(n)}}{\sqrt{n}}.$$ 

**Remark 3.** We actually prove the slightly stronger bound

$$W_2(S_n, G) \leq \frac{C \kappa_d \ln(d) \sqrt{d \ln(n)}}{\sqrt{n}},$$

where

$$\kappa_d := \sup_{\mu \text{ isotropic, log-concave}} \left\| \int x_1 x \otimes x \mu(dx) \right\|_{HS},$$

as defined in [21]. Results in [21] 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

$$W_2(S_n, G) \leq \frac{C \sqrt{d} \ln(d)^{3/2} \ln(n)}{\sqrt{n}}.$$ 

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 $\text{Cov}(X) = I_d$, 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 $d \sqrt{\log n} \sqrt{n}$.

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

$$\text{Ent}(A||B) = \mathbb{E} [\ln(f(A))].$$
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 \( \text{Ent}(X || G) < \infty \). Then one has
\[
\lim_{n \to \infty} \text{Ent}(S_n || G) = 0.
\]

The next result gives the first non-asymptotic convergence rate for the entropic CLT, again under the log-concavity assumption (other non-asymptotic results appear in previous works, notably [20], 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 \( G \sim \mathcal{N}(0, I_d) \) 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
\[
\text{Ent}(S_n || G) \leq \frac{C d^{10} (1 + \text{Ent}(X || G))}{n},
\]
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 \( \text{Cov}(X) = \Sigma \). Suppose further that \( X \) is 1-uniformly log concave (i.e. it has a probability density \( e^{-\varphi(x)} \) satisfying \( \nabla^2 \varphi \preceq I_d \)) and that \( \Sigma \succeq \sigma I_d \) for some \( \sigma > 0 \).

Let \( G \sim \mathcal{N}(0, \Sigma) \) be a Gaussian with the same covariance as \( X \) and let \( \gamma \sim \mathcal{N}(0, I_d) \) 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
\[
\text{Ent}(S_n || G) \leq \frac{2 (d + 2 \text{Ent}(X || \gamma))}{\sigma^4 n}.
\]

**Remark 8.** The theorem can be applied when \( X \) is isotropic and \( \sigma \)-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 \( \sigma I_d \) as a covariance matrix. Since relative entropy to a Gaussian is invariant under affine transformations, if \( G \sim \mathcal{N}(0, I_d) \) is a standard Gaussian, we get
\[
\text{Ent}(S_n || G) = \text{Ent}(\sqrt{\sigma}S_n || \sqrt{\sigma}G) \leq \frac{2 (d + 2 \text{Ent}(\sqrt{\sigma}X || G))}{\sigma^4 n}.
\]

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 \( \Gamma_t \) taking values in the positive semi-definite cone and let \( \tau \) be a stopping time. We say that the triplet \( (X_t, \Gamma_t, \tau) \) is a martingale embedding of the measure \( \mu \) if \( X_\tau \sim \mu \).

Note that if \( \Gamma_t \) 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 \( \mu \), we end up with a martingale embedding of \( \mu^\otimes n \) whose associated covariance process has the form \( \sqrt{\sum_{i=1}^n \left( \Gamma_{t_i}^{(i)} \right)^2} \).

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 \( \mathbb{E} \left[ \sqrt{\sum_{i=1}^{n} (\Gamma_{t}^{(i)})^2} \right] \) (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 \( \Gamma_{t} \) tractable. The specific choices of \( \Gamma_{t} \) that we consider are based on a construction introduced in [22]. 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 [21]. Such techniques have recently gained prominence and have been used, among other things, to improve known bounds of the KLS conjecture [21, 35], calculate large deviations of non-linear functions [23] 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 Subsection 2.3 below.

### 1.2. Related work

Multidimensional central limit theorems have been studied extensively since at least the 1940’s [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 [18], 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, 26, 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 [25]), maxima of sums of independent random vectors (Chernozhukov, Chetverikov, and Kato [19]), 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 [20, 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 \( p \)'th-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 [20], 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 \( \frac{\text{poly}(d)}{n} \) 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 $\text{Ent} \left( \frac{X+Y}{\sqrt{2}} \middle| G \right)$ is strictly less than $\text{Ent}(X \middle| G)$, 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 semi-definite symmetric matrix $A$ we denote by $\sqrt{A}$ the unique positive semi-definite 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 semi-definite matrix. By $A^\dagger$ we denote the pseudo inverse of $A$. Put succinctly, this means that in $A^\dagger$ every non-zero 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

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

when $F_t$ is adapted to the natural filtration of $B_t$.

$\mu$ will always stand for a probability measure. To avoid confusion, when integrating with respect to $\mu$, on $\mathbb{R}^d$, we will use the notation $\int \cdots \mu(dx)$. For a measure-valued stochastic process $\mu_t$, the expression $d\mu_t$ refers to the stochastic derivative of the process. A measure $\mu$ 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 $\rho$, twice differentiable almost everywhere, for which

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

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

$$-\nabla^2 \log(\rho(x)) \geq \sigma I_d \quad \text{for all } x,$$

we say that $\mu$ is $\sigma$-uniformly log-concave. The measure $\mu$ is called isotropic if it is centered and its covariance matrix is the identity, i.e.,

$$\int x \mu(dx) = 0 \quad \text{and} \quad \int x \otimes x \mu(dx) = I_d.$$

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

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.

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

Throughout this section we fix a measure $\mu$ 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 $\text{Cov}(X) = \Sigma$. 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 $\mathcal{F}_t$. 

\[ \text{Ent} \left( \frac{X+Y}{\sqrt{2}} \middle| G \right) \lessdot \text{Ent}(X \middle| G), \]
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 \((X_t, \Gamma_t, \tau)\) be a martingale embedding of \( \mu \). Set \( \Gamma_t = 0 \) for \( t > \tau \), then

\[
W_2^2(S_n, G) \leq \int_0^\infty \min \left( \frac{1}{n} \text{Tr} \left( \mathbb{E} \left[ \Gamma_t^4 \right] \mathbb{E}[\Gamma_t^2]^\dagger \right), 4 \text{Tr} \left( \mathbb{E} \left[ \Gamma_t^2 \right] \right) \right) dt.
\]

To illustrate how such a result might be used, let us assume, for simplicity, that \( \Gamma_t \preceq k I_d \) almost-surely for some \( k > 0 \) and that \( \tau \) has a sub-exponential tail, i.e., there exist positive constants \( C, c > 0 \) such that for any \( t > 0 \),

\[
\mathbb{P}(\tau > t) \leq C e^{-ct}.
\]

Under these assumptions,

\[
W_2^2(S_n, G) \leq \int_0^\infty \min \left( \frac{1}{n} \text{Tr} \left( \mathbb{E} \left[ \Gamma_t^4 \right] \mathbb{E}[\Gamma_t^2]^\dagger \right), 4k^2 d \mathbb{P}(\tau > t) \right) dt
\]

\[
\leq dk^2 \int_0^{\frac{\log(n)}{cn}} \frac{1}{n} dt + 4Cd \int_0^\infty e^{-ct} dt = \frac{d\log(n)k^2}{cn} + \frac{4Cdk^2}{n}.
\]

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

**Lemma 1.** Let \( A, B \) be positive semi-definite matrices with \( \ker(A) \subset \ker(B) \). Then,

\[
\text{Tr} \left( (\sqrt{A} - \sqrt{B})^2 \right) \leq \text{Tr} \left( (A - B)^2 A^\dagger \right).
\]

**Proof.** Since \( A \) and \( B \) are positive semi-definite, \( \ker(\sqrt{A} + \sqrt{B}) \subset \ker(\sqrt{A} - \sqrt{B}) \). Thus, we have that

\[
\sqrt{A} - \sqrt{B} = \left( \sqrt{A} - \sqrt{B} \right) \left( \sqrt{A} + \sqrt{B} \right) \left( \sqrt{A} + \sqrt{B} \right)^\dagger
\]

\[
= (A - B + [\sqrt{A}, \sqrt{B}]) \left( \sqrt{A} + \sqrt{B} \right)^\dagger.
\]

So,

\[
\text{Tr} \left( (\sqrt{A} - \sqrt{B})^2 \right) = \text{Tr} \left( (A - B + [\sqrt{A}, \sqrt{B}]) \left( \sqrt{A} + \sqrt{B} \right)^\dagger \right)^2.
\]

Note that for any symmetric matrices \( X \) and \( Y \), by the Cauchy-Schwartz inequality,

\[
\text{Tr} \left( (XY)^2 \right) \leq \text{Tr} (XYXY) \leq \sqrt{\text{Tr}(XYXY)} \cdot \text{Tr}(YXXY) = \text{Tr} (X^2Y^2).
\]

Applying this to the above equation shows

\[
\text{Tr} \left( (\sqrt{A} - \sqrt{B})^2 \right) \leq \text{Tr} \left( (A - B + [\sqrt{A}, \sqrt{B}])^2 \left( (\sqrt{A} + \sqrt{B})^\dagger \right)^2 \right).
\]
Note that the commutator \( [\sqrt{A}, \sqrt{B}] \) is an anti-symmetric matrix, so that \((A - B) [\sqrt{A}, \sqrt{B}] + [\sqrt{A}, \sqrt{B}] (A - B)\) is anti-symmetric as well. Thus, for any symmetric matrix \( C \), we have that

\[
\text{Tr} \left( (A - B) [\sqrt{A}, \sqrt{B}] + [\sqrt{A}, \sqrt{B}] (A - B) C \right) = 0.
\]

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 \( \text{Tr} \left( [\sqrt{A}, \sqrt{B}]^2 C \right) \leq 0 \). Using these observations we obtain

\[
\text{Tr} \left( (A - B + [\sqrt{A}, \sqrt{B})^2 \left( (\sqrt{A} + \sqrt{B})^\dagger \right)^2 \right) \leq \text{Tr} \left( (A - B)^2 \left( (\sqrt{A} + \sqrt{B})^\dagger \right)^2 \right).
\]

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 \( \text{Tr} (CX) \leq \text{Tr} (CY) \). The assumption \( \ker(A) \subset \ker(B) \) implies \( \left( (\sqrt{A} + \sqrt{B})^\dagger \right)^2 \preceq A^1 \), which concludes the claim by

\[
\text{Tr} \left( (A - B)^2 \left( (\sqrt{A} + \sqrt{B})^\dagger \right)^2 \right) \leq \text{Tr} \left( (A - B)^2 A^1 \right).
\]

\[ \square \]

\textbf{Proof of Theorem 10.} Recall that \((X_t, \Gamma_t, \tau)\) is a martingale embedding of \( \mu \). Let \( X^{(i)}_t, \Gamma^{(i)}_t, \tau^{(i)}_t \) be independent copies of the embedding. We can always set \( \Gamma^{(i)}_t = 0 \) whenever \( t > \tau^{(i)} \), so that

\[
\int_0^\infty \Gamma^{(i)}_t dB^{(i)}_t \sim \mu.
\]

Define \( \tilde{\Gamma}_t = \sqrt{\frac{1}{n}} \sum_{i=1}^n \Gamma^{(i)}_t \). Our first goal is to show

\[
\mathcal{W}_2^2(G, S_n) \leq \int_0^\infty \mathbb{E} \left[ \text{Tr} \left( \left( \tilde{\Gamma}_t - \sqrt{\mathbb{E} \left[ \tilde{\Gamma}_t^2 \right]} \right)^2 \right) \right] \, dt.
\]

The theorem will then follow by deriving suitable bounds for \( \mathbb{E} \left[ \text{Tr} \left( \left( \tilde{\Gamma}_t - \sqrt{\mathbb{E} \left[ \tilde{\Gamma}_t^2 \right]} \right)^2 \right) \right] \) using Lemma 1. Consider the sum \( \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\infty \Gamma^{(i)}_t dB^{(i)}_t \), 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 \( d\tilde{B}_t := \frac{1}{\sqrt{n}} \sum_{i=1}^n \Gamma_{t}^{(i)} dB^{(i)}_t \) is a martingale whose quadratic variation matrix has derivative satisfying

\[
\frac{d}{dt} [\tilde{B}]_t = \frac{1}{n} \sum_{i=1}^n \Gamma_{t}^{(i)} \left( \Gamma_{t}^{(i)} \right)^2 \tilde{\Gamma}_t^\dagger \preceq I_d.
\]

(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

\[
S_n = \int_0^\infty \sqrt{\mathbb{E} \left[ \tilde{\Gamma}_t^2 \right]} \, d\tilde{B}_t + \int_0^\infty \left( \tilde{\Gamma}_t - \sqrt{\mathbb{E} \left[ \tilde{\Gamma}_t^2 \right]} \right) d\tilde{B}_t.
\]
Observe that \( G := \int_0^\infty \sqrt{\mathbb{E}[\Gamma_t^2]} dB_t \) 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

\[
\text{Cov}(G) = \mathbb{E} \left[ \left( \int_0^\infty \sqrt{\mathbb{E}[\Gamma_t^2]} dB_t \right)^2 \right] = \mathbb{E} \left[ \int_0^\infty \Gamma_t^2 dt \right] = \mathbb{E} \left[ \left( \int_0^\infty \Gamma_t dB_t \right)^2 \right] = \text{Cov}(X).
\]

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

\[
W_2^2(G, S_n) \leq \mathbb{E} \left[ \left\| \int_0^\infty \left( \tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]} \right) dB_t \right\|^2 \right] \overset{(5)}{=} \text{Tr} \left( \mathbb{E} \left[ \int_0^\infty \left( \tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]} \right)^2 dt \right] \right)
\]

\[
= \int_0^\infty \mathbb{E} \left[ \text{Tr} \left( \left( \tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]} \right)^2 \right) \right] dt,
\]

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

\[
(7) \quad \text{Tr} \left( \mathbb{E} \left[ \left( \tilde{\Gamma}_t - \sqrt{\mathbb{E}[\Gamma_t^2]} \right)^2 \right] \right) \leq 4 \text{Tr} \left( \mathbb{E}[\Gamma_t^2] \right).
\]

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 \( \Gamma_t \) is positive semi-definite, it is clear that \( \text{ker} \left( \mathbb{E} [\Gamma_t^2] \right) \subset \text{ker}(U_t) \). By Lemma 1,

\[
\mathbb{E} \left[ \text{Tr} \left( \left( \sqrt{U_t} - \sqrt{\mathbb{E}[\Gamma_t^2]} \right)^2 \right) \right] \leq \text{Tr} \left( \mathbb{E} \left[ (U_t - \mathbb{E} [\Gamma_t^2])^2 \right] \mathbb{E} [\Gamma_t^2]^\top \right)
\]

\[
= \frac{1}{n^2} \text{Tr} \left( \sum_{i=1}^n \mathbb{E} \left[ (\Gamma_t^{(i)})^2 - \mathbb{E} [\Gamma_t^2] \right]^2 \mathbb{E} [\Gamma_t^2]^\top \right)
\]

\[
= \frac{1}{n} \text{Tr} \left( \mathbb{E} [\Gamma_t^4] - \mathbb{E} [\Gamma_t^2]^2 \mathbb{E} [\Gamma_t^2]^\top \right)
\]

\[
\leq \frac{1}{n} \text{Tr} \left( \mathbb{E} [\Gamma_t^4] \mathbb{E} [\Gamma_t^2]^\top \right),
\]

where we have used the fact \( \mathbb{E} \left[ (\Gamma_t^{(i)})^2 \right] = \mathbb{E} [\Gamma_t^2] \) 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 \(\mu\). Assume that for every \(0 \leq t \leq 1\), 
\[ \mathbb{E} [\Gamma_t] \geq \sigma t I_d \geq 0 \] 
and that \(\Gamma_t\) is invertible a.s. for \(t < 1\). Then we have the following inequalities:

\[ \text{Ent}(S_n || G) \leq \frac{1}{n} \int_0^1 \frac{1}{(t-1)^2} \left( \int_0^1 \sigma_s^{-2} ds \right) dt, \]

and

\[ \text{Ent}(S_n || G) \leq \frac{1}{n} \int_0^1 \frac{1}{(t-1)^2} \left( \int_0^1 \sigma_s^{-2} ds \right) dt, \]

where

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

and \(\Gamma^{(i)}_t\) are independent copies of \(\Gamma_t\).

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

LEMMA 2. Let \(\Gamma_t\) 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,

\[ \text{Ent} (X_t || M_t) \leq \mathbb{E} \left[ \int_0^1 \int_0^1 \left\| F_t^{-1}(Y_t) \frac{\Gamma_s - F_s(Y_s)}{1 - s} \right\|_{HS}^2 dt ds \right]. \]

Note that if the process \(F_t\) is deterministic, i.e. it is a constant function, then \(M_t\) has a Gaussian law, so that the lemma can be used to bound the relative entropy of \(X_t\) with respect to a Gaussian.

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

\[ \tilde{\Gamma}_t := \sqrt{\frac{1}{n} \sum_{i=1}^n (\Gamma^{(i)}_t)^2} \quad \text{and} \quad d\tilde{B}_t = \frac{1}{\sqrt{n}} \tilde{\Gamma}_t^{-1} \sum_{i=1}^n \Gamma^{(i)}_t dB^{(i)}_t. \]

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

\[ M_t = \int_0^t \sqrt{\mathbb{E} |\tilde{\Gamma}_s^2|} dB_s. \]
By Itô’s isometry one has $M_1 \sim \mathcal{N}(0, \Sigma)$. Also, by Jensen’s inequality

$$\sqrt{\mathbb{E} \left[ \Gamma_t^2 \right]} \geq \mathbb{E} [\Gamma_t] \geq \sigma_t I_d.$$ 

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

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

With the use of Lemma 1 we obtain

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

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

$$M'_t := \int_0^t \mathbb{E} \left[ \tilde{\Gamma}_s \right] d\tilde{B}_s.$$ 

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

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

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

$$\mathbb{E} \left[ \Gamma_s \right] \geq \frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^{n} \Gamma_i^{(i)} \right] \geq \sigma_s I_d.$$ 

Thus,

$$\text{Ent}(S_n\|G_n) \leq \mathbb{E} \left[ \int_0^t \mathbb{E} \left[ \tilde{\Gamma}_s \right]^{-1} \left\| \tilde{\Gamma}_t - \mathbb{E} [\tilde{\Gamma}_t] \right\|^2_{HS} \right] \leq \frac{1}{\int_0^t} \mathbb{E} \left[ \left\| \tilde{\Gamma}_t - \mathbb{E} [\tilde{\Gamma}_t] \right\|^2_{HS} \right] \left( \int_0^t \sigma_s^{-2} ds \right) dt \leq \frac{1}{(1-t)^2} \mathbb{E} \left[ \text{Tr} \left( (\Gamma_t^2 - \mathbb{E} [\Gamma_t^2])^2 \right) \right] \left( \int_0^t \sigma_s^{-2} ds \right) dt.$$ 

Since $\text{Cov}(G) = \text{Cov}(S_n)$, it is now easy to verify that $\text{Ent} (S_n\|G) \leq \text{Ent} (S_n\|G_n)$, which concludes the proof. \qed
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 \( Y_t := \int_0^t F_s(Y_s)dB_s + \int_0^t u_sds. \) Then

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

**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

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

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

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

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

\[
\mathbb{E} \left[ \ln(f(Y_1)) \right] = \text{Ent}(Y_1 || M_1),
\]

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

\[
Y_1 = \int_0^1 F_t(Y_t)dB_t + \int_0^1 \frac{\Gamma_t - F_s(Y_s)}{1-s}dB_sdt = \int_0^1 F_t(Y_t)dB_t + \int_0^1 \left( \Gamma_t - F_t(Y_t) \right) dB_t = X_1, \tag{9}
\]

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
\[
\text{Ent}(X_1||M_1) \leq \int_0^1 \mathbb{E} \left[ \left\| F_{t}^{-1}(Y_t)u_t \right\|^2 \right] dt = \int_0^1 \mathbb{E} \left[ \left\| F_{t}^{-1}(Y_t) \frac{\Gamma_s - F_s(Y_s)}{1 - s} dB_s \right\|^2 \right] dt \]
\[ = \mathbb{E} \left[ \int_0^1 \int_0^t \left\| F_{t}^{-1}(Y_t) \frac{\Gamma_s - F_s(Y_s)}{1 - s} \right\|^2_{HS} ds dt \right] \]
\[ = \mathbb{E} \left[ \int_0^1 \int_s^1 \left\| F_{t}^{-1}(Y_t) \frac{\Gamma_s - F_s(Y_s)}{1 - s} \right\|^2_{HS} dt ds \right] , \]
where last equality follows from another use of Fubini’s theorem.

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 subsection are very similar to what is done in [22], 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 $\text{Sym}_d$.

**Proposition 1.** Let $\mu$ be a probability measure on $\mathbb{R}^d$ with smooth density and bounded support. For a probability measure-valued process $\mu_t$, let
\[ a_t = \int_{\mathbb{R}^d} x \mu_t(dx), \quad A_t = \int_{\mathbb{R}^d} (x - a_t)^{\otimes 2} \mu_t(dx) \]
derive its mean and covariance.

Let $C : \mathbb{R} \times \text{Sym}_d \to \text{Sym}_d$ be a continuous function. Then, we can construct $\mu_t$ 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 $\mu_t$ 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 $\mu_t$, where $\tau$ is the first time that $\mu_t$ 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 $\mu_t$ for $\epsilon \to 0$. Consider a stochastic process $\{X_s\}_{0 \leq s \leq 1}$ 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 $X_t = a_t + \sqrt{\epsilon}C_t^{-1}Z$, where $Z$ is close to a standard Gaussian. We then take $\mu_{t+\epsilon}$ to be the conditional distribution of $X_t$ given $X_\epsilon$. This immediately ensures that property 3 holds and that $a_t$ is a martingale.
It remains to see why property 2 holds. A direct calculation with conditioned Brownian bridges gives a first-order approximation

\[ \mu_{t+\epsilon}(dx) \propto e^{-\frac{1}{2}(\sqrt{C_t^{-1}}Z_\epsilon(x-a_t))}C_t^2(\sqrt{C_t^{-1}}Z_\epsilon(x-a_t))\mu_t(dx) \]

\[ \propto e^{\sqrt{\epsilon}C_t(x-a_t)+O(\epsilon)}\mu_t(dx) \]

\[ \approx (1 + \sqrt{\epsilon}(C_t Z_\epsilon x - a_t))\mu_t(dx). \]

Then, to highest order, we have

\[ a_{t+\epsilon} - a_t \approx \sqrt{\epsilon} \int_{\mathbb{R}^d} (C_t Z_\epsilon x - a_t)(x-a_t) \mu_t(dx) = \sqrt{\epsilon}A_t C_t Z_\epsilon, \]

which translates into property 2 as \( \epsilon \to 0 \).

Observe that the procedure outlined above yields measures \( \mu_t \) that have densities which are proportional to the original density \( \mu \) times a Gaussian density. (This applies at least when \( A_t \) is non-degenerate; 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 non-degenerate.

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

**Proof.** We will construct a \((\mathbb{R}^d \times \text{Sym}_d)\)-valued stochastic process \((c_t, \tilde{\Sigma}_t)\) started at \((c_0, \tilde{\Sigma}_0) = (0, I_d)\). 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 \( \frac{d\tilde{\mu}}{d\mu}(x) \propto e^{\frac{1}{2}\langle x \rangle^2} \). We will then take \( \mu_t \) 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 \( \tilde{\Sigma}_0 = I_d \), we have \( \mu_0 = \mu \).

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

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

Note that the coefficients of this SDE are continuous functions of \((c_t, \tilde{\Sigma}_t)\) so long as \( \tilde{\Sigma}_t > 0 \). 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 \( t, \tilde{\Sigma}_t \leq I_d \) and so, the matrix process is continuous on the interval \([0, T]\).

By a limiting procedure, it is easy to see that \( \mu_T \) 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 \( \mu_t \) is a martingale and \( da_t = A_tC_t dB_t \).

---

1Conceptually, one can replace all instances of \( \tilde{\mu} \) with \( \mu \) 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.
By direct calculation, we have
\[
\begin{align*}
    d(\tilde{\Sigma}_t^{-1}) &= C_t^2 dt \\
    d(\tilde{\Sigma}_t^{-1} c_t) &= C_t^2 c_t dt + C_t^2 (a_t - c_t) dt + C_t dB_t \\
    &= C_t^2 a_t dt + C_t dB_t \\
    dQ_t(x) &= \left< x, \left( \frac{1}{2} C_t^2 x - C_t^2 a_t \right) dt - C_t dB_t \right> \\
    d(e^{-Q_t(x)}) &= -e^{-Q_t(x)} dQ_t(x) + \frac{1}{2} e^{-Q_t(x)} d[Q_t(x)] \\
    &= e^{-Q_t(x)} \left< x, C_t dB_t + C_t^2 a_t dt \right>
\end{align*}
\]

Integrating against \( \tilde{\mu}(dx) \), we obtain
\[
\begin{align*}
    dZ_t &= Z_t \left< a_t, C_t dB_t + C_t^2 a_t dt \right> \\
    dZ_t^{-1} &= -\frac{1}{Z_t^2} dZ_t + \frac{1}{Z_t^2} d[Z_t] = \frac{1}{Z_t} (a_t - C_t dB_t) \\
    dF_t(x) &= e^{-Q_t(x)} dZ_t^{-1} + Z_t^{-1} d(e^{-Q_t(x)}) + d[Z_t^{-1}, e^{-Q_t(x)}] \\
    &= F_t(x) \cdot \left< x - a_t, C_t dB_t \right>.
\end{align*}
\]

Thus, \( F_t(x) \) is a martingale for each fixed \( x \), and furthermore,
\[
    da_t = d \int_{\mathbb{R}^d} x \mu_t(dx) = \int_{\mathbb{R}^d} x d\mu_t(dx) = \int_{\mathbb{R}^d} x(x - a_t) C_t \mu_t(dx) dB_t = A_t C_t dB_t.
\]

**Proof of Proposition 1.** We use the process given by Lemma 4, which yields a stopping time \( T_1 \) and a measure \( \mu_{T_1} \), with a strictly lower-dimensional support. If \( \mu_T \) is a point mass, then we set \( \mu_t = \mu_T \) for all \( t \geq T \).

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

2.4. Properties of the construction. We record here various formulas pertaining to the quantities \( a_t \), \( A_t \), and \( \mu_t \) constructed in Proposition 1.

**Proposition 2.** Let \( \mu, C_t, \) and \( \mu_t \) be as in Proposition 1. Then, there is a \( \text{Sym}_d \)-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 \( \gamma_t \) on \( \mathbb{R}^d \), supported on \( L \), with covariance \( \Sigma_t \) such that \( \mu_t \) is absolutely continuous with respect to \( \gamma_t \), and
  \[
  \frac{d\mu_t}{d\gamma_t}(x) \propto \mu(x), \quad \forall x \in L.
  \]
- \( \Sigma_t \) is continuous and for almost every \( t \) obeys the differential equation
  \[
  \frac{d}{dt} \Sigma_t = -C_t^2 \Sigma_t.
  \]
- \( \lim_{t \to 0^+} \Sigma_t^{-1} = 0 \).
PROOF. For $1 \leq k \leq d$, let $T_k$ denote the first time the measure $\mu_t$ is supported in a $(d-k)$-dimensional affine subspace, and denote by $L_t$ the affine hall of the support of $\mu_t$. We will define $\Sigma_t$ inductively for each interval $[T_{k-1}, T_k]$. Recall from the proof of Proposition 1 that $\mu_t$ is constructed by iteratively applying Lemma 4 to affine subspaces of decreasing dimension $d, d-1, d-2, \ldots, 1$. 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 $L_T$. For the base case $0 < t \leq T_1$, take $\Sigma_t = (\tilde{\Sigma}_{0,t}^{-1} - I_d)^{-1}$. A straightforward calculation shows that over this time interval, $d\mu_t/du$ is proportional to the density of a Gaussian with covariance $\Sigma_t$. Note that since $\tilde{\Sigma}_{0,0}^{-1} = I_d$, we also have $\lim_{t \to 0^+} \Sigma_t^{-1} = 0$.

Now suppose that $\Sigma_t$ has been defined up until time $T_k$; we will extend it to time $T_{k+1}$. Let $L_k$ denote the affine hall of the support of $\mu_{T_k}$, so that $\dim(L_k) = d-k$ (if $\dim(L_k) < d-k$, then we simply have $T_{k+1} = T_k$). Then, for $0 \leq t \leq T_{k+1} - T_k$, we may set $\Sigma_{T_k + t} := \left(\tilde{\Sigma}_{k,t}^{-1} + \Sigma_T^{-1} - I_d\right)^{-1},$

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 $\Sigma_t$. Once again, a straightforward calculation shows that for $T_k \leq t < T_{k+1}$, $d\mu_t/du$ is proportional to the density of a Gaussian with covariance $\Sigma_t$, where we view $\mu_t$ and $\mu$ as densities on $L_k$ (for $\mu$, we take its conditional density on $L_k$).

It remains only to show that $\Sigma_t$ satisfies the required differential equation. From our construction, we see that $\Sigma_t$ always takes the form $\left(\tilde{\Sigma}_t^{-1} - H\right)^{-1}$, where $H \leq I_d$ and

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

Then, we have

$$\frac{d}{dt} \Sigma_t = -\left(\tilde{\Sigma}_t^{-1} - H\right)^{-1} \left(\frac{d}{dt} \tilde{\Sigma}_t\right) \left(\tilde{\Sigma}_t^{-1} - H\right)^{-1}$$

$$= -\Sigma_t \left(-\tilde{\Sigma}_t^{-1} \left(\frac{d}{dt} \tilde{\Sigma}_t\right) \tilde{\Sigma}_t^{-1}\right) \Sigma_t$$

$$= -\Sigma_t C_l^2 \Sigma_t,$$

as desired. $\square$

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

**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

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

Clearly the first 3 terms are local martingales, which shows, by the uniqueness of the Doob decomposi-
Also, the last 2 terms are clearly of bounded variation, which shows
\[ \int dA_t = d\int (x - a_t)\otimes 2 \mu_t(dx) = \int d((x - a_t)\otimes 2 \mu_t(dx)) \]
\[ = - \int da_t \otimes (x - a_t)\mu_t(dx) - \int (x - a_t) \otimes da_t \mu_t(dx) + \int (x - a_t)\otimes 2 d\mu_t(dx) \]
\[ - 2 \int (x - a_t) \otimes d[a_t, \mu_t(dx)] + \int d[a_t, a_t] d\mu_t(dx). \]

Note that the first 2 terms are equal to 0, since, by definition of \( a_t \),
\[ \int da_t \otimes (x - a_t)\mu_t(dx) = da_t \otimes \int (x - a_t)\mu_t(dx) = 0. \]

Also, the last 2 terms are clearly of bounded variation, which shows
\[ dM_t = \int (x - a_t)\otimes 2 d\mu_t(dx) = \int (x - a_t)\otimes 3 C_t \mu_t(dx) dB_t. \]

Define the stopping time \( \tau = \inf\{t|A_t = 0\} \). Then, at time \( \tau \), \( \mu \) is just a delta mass located at \( a_\tau \) and \( \mu_s = \mu_\tau \) for every \( s \geq \tau \). A crucial observation is

**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 \( t_0 < t < \tau \), \( \text{Tr} (A_t C_t^2 A_t) > c \),
2. \( \int_{t_0}^{\tau} \lambda_{\text{min}} (C_t^2) dt = \infty \), where \( \lambda_{\text{min}} (C_t^2) \) is the minimal eigenvalue of \( C_t^2 \),

then \( \tau \) is finite a.s. and in the second case \( \tau \leq t_0 \). Moreover, if \( \tau \) is finite a.s. then \( a_\tau \) has the law of \( \mu \).

**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 \( \text{Tr} (R_t) \) a positive local martingale; hence, a super-martingale. By the martingale convergence theorem \( \text{Tr} (R_t) \) converges to a limit almost surely. By our assumption, if \( \tau = \infty \) then
\[ \int_0^\infty \text{Tr} (A_t C_t^2 A_t) dt \geq \int_0^{t_0} \text{Tr} (A_t C_t^2 A_t) dt \geq \int_0^{t_0} c dt = \infty. \]

This would imply that \( \lim_{t \to \infty} \text{Tr} (A_t) = -\infty \) 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,
\[ \int_0^{t_0} \text{Tr} (A_t C_t^2 A_t) dt \geq \int_0^{t_0} \langle A_t v_t, C_t^2 A_t v_t \rangle dt \geq a^2 \int_0^{t_0} \lambda_{\text{min}} (C_t^2) dt = \infty, \]
which implies \( \lim_{t \to t_0} \text{Tr} (A_t) = -\infty \). Again, this cannot happen and so \( \mathbb{P}(\tau > t_0) = 0 \).
To understand the law of $a_t$, let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be any continuous bounded function. By Property 3 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 $\mu_t$ is a probability measure for every $t$, then
\[
\int_{\mathbb{R}^d} \varphi(x)\mu_t(dx) \leq \max_x |\varphi(x)|.
\]
$\tau$ 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 $\mu_\tau$ 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 $t \leq s$.

**PROOF.** To see that $\text{rank}(A_t)$ 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 $t \geq t_0$, $A_{t}v_0 = 0$. 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 $t \geq t_0$, $\langle v_0, A_t v_0 \rangle = 0$ as well.

\[
\square
\]

**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 $\mu$ on $\mathbb{R}^d$ and a random vector $X \sim \mu$ with the assumption that $\|X\| \leq \beta$ 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 $\tau$ as in Section 2.3, where we take $C_t = A_t^1$, so that $a_t = \int_0^t A_s A_s^1 dB_s$. We denote $P_t := A_t A_t^1$, and remark that since $A_t$ is symmetric, $P_t$ is a projection matrix. As such, we have that for any $t < \tau$, $\text{Tr}(P_t) \geq 1$. By Proposition 4, $a_\tau$ has the law $\mu$.

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

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

**PROOF.** Let $H_t = \|a_t\|^2$. By Itô's formula and since $P_t$ is a projection matrix,
\[
dH_t = 2\langle a_t, P_t dB_t \rangle + \text{Tr}(P_t) dt = 2\langle a_t, P_t dB_t \rangle + \text{rank}(P_t) dt.
\]
So, $\frac{d}{dt}\mathbb{E}[H_t] = \mathbb{E}[\text{rank}(P_t)]$. Since $\mathbb{E}[H_\infty] \leq \beta^2$,
\[
\beta^2 \geq \mathbb{E}[H_\infty] - \mathbb{E}[H_0] = \int_0^\infty \mathbb{E}[\text{rank}(P_t)] dt \geq \int_0^\infty \mathbb{P}(\tau > t) dt = \mathbb{E}[\tau].
\]
The above claim gives bounds on the expectation of $\tau$, however in order to use Theorem 10, we need bounds for its tail behaviour in the sense of (2). To this end, we can use a bootstrap argument and invoke the above lemma with the measure $\mu_t$ in place of $\mu$, recalling that $X_\infty |F_t \sim \mu_t$ and noting that $\|X_\infty |F_t\| \leq \beta$ almost surely. Therefore, we can consider the conditioned stopping time $\tau |F_t - t$ and get that

$$E[\tau |F_t] \leq t + \beta^2.$$ 

The following lemma will make this precise.

**Lemma 6.** Suppose that, for the stopping time $\tau$, it holds that for every $t > 0$, $E[\tau |F_t] \leq t + \beta^2$ a.s., then

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

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

$$P(\tau \geq t_{i+1}) \leq P(\tau > t_i) \text{ess sup} \mu_t \left( P(\tau - t_i \geq 2\beta^2 |F_{t_i}) \right),$$

where the essential supremum is taken over all possible states of $\mu_t$. Using Markov’s inequality, we almost surely have

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

which is also true for the essential supremum. Clearly $P(\tau \geq 0) = 1$ 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 $(X_t, \Gamma_t, \tau)$ becomes a martingale embedding according to Proposition 4. In this case, we have that $\Gamma_t$ is a projection matrix almost surely. Thus,

$$\text{Tr} \left( E[\Gamma_t^4]E[\Gamma_t^2] \right) \leq d,$$

and

$$\text{Tr} \left( E[\Gamma_t^2] \right) \leq dP(\tau > t).$$

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

$$W_2^2(S_n, G) \leq \int_0^{\frac{2\beta^2 \log_2(n)}{n}} \frac{d}{n} dt + \int_0^\infty 4dP(\tau > t) dt$$

$$\leq \frac{2d\beta^2 \log_2(n)}{n} + 4d \int_0^\infty P(\tau > \left\lfloor \frac{t}{2\beta^2} \right\rfloor 2\beta^2) dt$$

$$\leq \frac{2d\beta^2 \log_2(n)}{n} + 4d \int_0^\infty \left( \frac{1}{2} \right)^{\frac{t}{2\beta^2}} dt$$

$$\leq \frac{2d\beta^2 \log_2(n)}{n} + 8d\beta^2 \sum_{j=\lfloor \log_2(n) \rfloor}^\infty \frac{1}{2^j} \leq \frac{2d\beta^2 \log_2(n)}{n} + \frac{32d\beta^2}{n}.$$
Taking square roots, we finally have
\[
W_2(S_n, G) \leq \frac{\beta \sqrt{d \sqrt{32 + 2 \log_2(n)}}}{\sqrt{n}},
\]
as required. \[\square\]

### 3.2. The case of log-concave vectors: proof of Theorem 2.

In this section we fix \(\mu\) 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 \(\tau\). To define the matrix process \(C_t\), we first define a new stopping time \(T \doteq 1 \wedge \inf \{ t | \| A_t \|_{op} \geq 2 \}\).

\(C_t\) is then defined in the following manner:
\[
C_t = \begin{cases} 
\min(A_t^\dagger, I_d) & \text{if } t \leq T \\
A_t^\dagger & \text{otherwise}
\end{cases}
\]
where, again, \(A_t^\dagger\) denotes the pseudo-inverse of \(A_t\) and \(\min(A_t^\dagger, I_d)\) 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 eigenvalue of \(A_t^\dagger\) truncated at 1. Since \(\text{Tr}(A_t A_t^\dagger) \geq 1\) whenever \(t \leq \tau\), then the conditions of Proposition 4 are clearly met for \(t_0 = 1\) and \(a_\tau\) has the law of \(\mu\).

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

**Lemma 7.** \(\tau < 1 + \frac{4}{T}\).

**Proof.** Let \(\Sigma_t\) be as in Proposition 2. As described in the proposition, \(\mu_t\) is proportional to \(\mu\) times a Gaussian of covariance \(\Sigma_t\), on an appropriate affine subspace. In this case, an application of the Brascamp-Lieb inequality (see [32] for details) shows that \(A_t = \text{Cov}(\mu_t) \preceq \Sigma_t\). In particular, this means that for \(t > T\), when restricted to the orthogonal complement of \(\ker(A_t)\), the following inequality holds,
\[
\frac{d}{dt} \Sigma_t = -\Sigma_t C_t^2 \Sigma_t \preceq -I_d.
\]
So, \(\tau \leq T + \| \Sigma_T \|_{op}\).

It remains to estimate \(\| \Sigma_T \|_{op}\). To this end, recall that for \(0 < t \leq T\), we have \(\| A_t \|_{op} \leq 2\), 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 \(f(T) = \| \Sigma_T \|_{op}\), which has solution \(f(t) = \frac{4}{t - T} \frac{4}{\| \Sigma_T \|_{op}}\). By Gronwall’s inequality, \(f(t)\) lower bounds \(\| \Sigma_t \|_{op}\) for \(0 < t \leq T\), and so, in particular, \(f(t)\) must remain finite within that interval. Consequently, we have
\[
\frac{4}{\| \Sigma_T \|_{op}} > T \implies \| \Sigma_T \|_{op} < \frac{4}{T}.
\]
We conclude that
\[
\tau \leq T + \| \Sigma_T \|_{op} < 1 + \frac{4}{T},
\]
as desired. \[\square\]
LEMMA 8. There exists universal constants \( c, C > 0 \) such that whenever \( s > C \cdot \kappa_d^2 \ln(d)^2 \) then
\[
P(\tau > s) \leq e^{-cs},
\]
where \( \kappa_d \) is the constant defined in (1).

PROOF. First, by using the previous claim, we may see that for any \( s \geq 5 \),
\[
P(\tau > s) \leq P\left( \frac{1}{T} \geq \frac{s - 1}{4} \right) \leq P\left( \frac{1}{T} \geq \frac{s}{5} \right) = P\left( 5s^{-1} \geq T \right) = P\left( \max_{0 \leq t \leq 5s^{-1}} \|A_t\|_{op} \geq 2 \right).
\]
Recall from Proposition 3,
\[
dA_t = \int_{\mathbb{R}^d} (x - a_t) \otimes (x - a_t) (C_t (x - a_t), dB_t) \mu_t(dx) - A_t C_t^2 A_t dt.
\]
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 \( t \leq T \),
\[
\int_0^t A_s C_s^2 A_s ds \leq I_d.
\]
Thus, for \( t \in [0, T] \),
\[
3I_d \geq A_t + I_d \geq \tilde{A}_t \geq A_t.
\]
Also, \( \tilde{A}_t \) can be written as,
\[
d\tilde{A}_t = \int_{\mathbb{R}^d} (x - a_t) \otimes (x - a_t) (C_t (x - a_t), dB_t) \mu_t(dx), \quad \tilde{A}_0 = I_d.
\]
The above shows
\[
P\left( \max_{0 \leq t \leq 5s^{-1}} \|A_t\|_{op} \geq 2 \right) \leq P\left( \max_{0 \leq t \leq 5s^{-1}} \|\tilde{A}_t\|_{op} \geq 2 \right).
\]
We note than whenever \( \|\tilde{A}_t\|_{op} \geq 2 \) then also \( \text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right)^{\frac{1}{4 \ln(d)}} \geq 2 \), so that
\[
P\left( \max_{0 \leq t \leq 5s^{-1}} \|\tilde{A}_t\|_{op} \geq 2 \right) \leq P\left( \max_{0 \leq t \leq 5s^{-1}} \text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right)^{\frac{1}{4 \ln(d)}} \geq 2 \right)
\]
\[
\leq P\left( \max_{0 \leq t \leq 5s^{-1}} \ln \left( \text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right) \right) \geq 2 \ln(d) \right) = P\left( \max_{0 \leq t \leq 5s^{-1}} (M_t + E_t) \geq 2 \ln(d) \right),
\]
where \( M_t \) and \( E_t \) form the Doob-decomposition of \( \ln \left( \text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right) \right) \). 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 \( v_1, ..., v_n \) be the unit eigenvectors of \( \tilde{A}_t \) and let \( \alpha_{i,j} = \langle v_i, \tilde{A}_t v_j \rangle \) with
\[
d\alpha_{i,j} = \int_{\mathbb{R}^d} \langle x, v_i \rangle \langle x, v_j \rangle (C_t x, dB_t) \mu_t(dx + a_t),
\]
which follows from (12). Also define
\[
\xi_{i,j} = \frac{1}{\sqrt{\alpha_{i,j} \alpha_{j,i}}} \int_{\mathbb{R}^d} \langle x, v_i \rangle \langle x, v_j \rangle C_t x \mu_t(dx + a_t).
\]
So that
\[ d\alpha_{i,j} = \sqrt{\alpha_{i,j} \langle \xi_{i,j}, dB_t \rangle}, \quad \frac{d}{dt}[\alpha_{i,j}]_t = \alpha_{i,j} \| \xi_{i,j} \|^2. \]

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 \( \tilde{A}_t^{-1/2} A_t \tilde{A}_t^{-1/2} \leq I_d \). 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 \( t \leq T, \tilde{A}_t^{1/2} \leq 2I_d \) and \( C_t \leq I_d \). So, \( \| C_t \tilde{A}_t^{1/2} \|_{op} \leq 2 \). Under similar conditions, it was shown in [21], Lemma 3.2, that there exists a universal constant \( C > 0 \) for which

- for any \( 1 \leq i \leq d, \| \xi_{i,i} \|^2 \leq C. \)
- for any \( 1 \leq i \leq d, \sum_j^{d} \| \xi_{i,j} \|^2 \leq C \kappa_d^2. \)

Furthermore, in the proof of Proposition 3.1 in the same paper it was shown
\[ d \text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right) \leq 4 \ln(d) \sum_{i=1}^{d} \alpha_{i,i}^{4 \ln(d)} \langle \xi_{i,i}, dB_t \rangle + 16 C \kappa_d^2 \ln(d)^2 \text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right) dt. \]

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
\[ dE_t \leq 16 C \kappa_d^2 \ln(d)^2 \frac{\text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right)}{\text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right)} = 16 C \kappa_d^2 \ln(d)^2, \quad E_0 = \ln(d), \]
and
\[ \frac{d}{dt}[M]_t \leq 16 C^2 \ln(d)^2 \left( \frac{\text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right)}{\text{Tr} \left( \tilde{A}_t^{4 \ln(d)} \right)} \right)^2 = 16 C^2 \ln(d)^2. \]

Hence, \( E_t \leq t \cdot 16 C \kappa_d^2 \ln(d)^2 + \ln(d) \), which together with (13) gives
\[ \mathbb{P}(\tau > s) \leq \mathbb{P} \left( \max_{0 \leq t \leq 5s^{-1}} M_t \geq 2 \ln(d) - \ln(d) - 80s^{-1} C \kappa_d^2 \ln(d)^2 \right) \forall s \geq 5. \]

Under the assumption \( s > 80 C \kappa_d^2 \ln(d)^2 \), the above can simplify to
\[ \mathbb{P}(\tau > s) \leq \mathbb{P} \left( \max_{0 \leq t \leq 5s^{-1}} M_t \geq \frac{1}{2} \ln(d) \right). \]

To bound this last expression, we will apply the Dubins-Schwartz theorem to write
\[ M_t = W_{[M]_t}, \]
where $W_t$ is some Brownian motion. Combining this with (15) gives

$$
\mathbb{P}(\tau > s) \leq \mathbb{P}\left(\max_{0 \leq t \leq 5s^{-1}} W_t \geq \frac{\ln(d)}{2}\right).
$$

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

$$
\mathbb{P}\left(\max_{0 \leq t \leq t'} W_t \geq K\right) \leq \exp\left(-\frac{K^2}{2t'}\right).
$$

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

$$
\mathbb{P}\left(\max_{0 \leq t \leq 5s^{-1}} W_t \geq \frac{\ln(d)}{2}\right) \leq e^{-cs},
$$

where $c > 0$ is some universal constant.

**Proof of Theorem 2.** By definition of $T$ and $C_t$, we have that for any $t \leq T$, $A_tC_t \preceq 2I_d$ and for any $t > T$, $A_tC_t = A_tA_t^\dagger \preceq I_d$. We now invoke Theorem 10, with $\Gamma_t = A_tC_t$, for which

$$
\text{Tr}\left(\mathbb{E}[\Gamma_t^4]\mathbb{E}[\Gamma_t^2]\right) \leq 4d,
$$

and, by Lemma 8

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

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

$$
\mathcal{W}_2^2(S_n, G) \leq \int_0^n \frac{4d}{n} dt + \int_0^\infty 16d\mathbb{P}(\tau > t)\int_0^\infty e^{-ct} dt
$$

$$
\leq \frac{4dC \cdot \kappa_d^2 \ln(d)^2 \ln(n)}{n} + 16d\int_0^\infty \frac{e^{-ct} dt}{C \cdot \kappa_d^2 \ln(d)^2 \ln(n)}
$$

$$
\leq \frac{Ct \cdot \kappa_d^2 \ln(d)^2 \ln(n)}{n}.
$$

Thus

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

**4. Convergence rates in entropy.** Throughout this section, we fix a centered measure $\mu$ on $\mathbb{R}^d$ with an invertible covariance matrix $\Sigma$ 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 $\text{Ent}(S_n||G)$. 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 $\Gamma_t$. Our construction uses the process $a_t$ from Proposition 1 with the choice $C_t := \frac{1}{1-t}I_d$. Property 2 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 \(\mu\).

As above, the sequence \(\Gamma^{(i)}_t\) will denote independent copies of \(\Gamma_t\) and we define \(\tilde{\Gamma}_t := \sqrt{\sum_{i=1}^n (\Gamma^{(i)}_t)^2}\).

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 [24] (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 \(\mu_t\). Thus, almost surely \(Y_1|\mathcal{F}_t \sim \mu_t\) and since \(a_t\) is the expectation of \(\mu_t\), we have the identity
\[(16) \quad a_t = \mathbb{E}[Y_1|\mathcal{F}_t],\]
and in particular we have \(a_1 = Y_1\). Moreover, the same reasoning implies that \(A_t = \text{Cov}(Y_1|\mathcal{F}_t)\) and
\[(17) \quad \Gamma_t = \frac{\text{Cov}(Y_1|\mathcal{F}_t)}{1 - t}.\]

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
\[(18) \quad \text{Ent}(Y_1||\gamma) = \frac{1}{2} \int_0^1 \mathbb{E}\left[\|v_t\|^2\right] dt,\]
where \(\gamma\) denotes the standard Gaussian.

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

**Proof.** From (16), we have
\[\text{Cov}(Y_1|\mathcal{F}_t) = \mathbb{E}\left[Y_1^\odot 2|\mathcal{F}_t\right] - \mathbb{E}[Y_1|\mathcal{F}_t]^\odot 2 = \mathbb{E}\left[Y_1|\mathcal{F}_t\right]^\odot 2 - a_t^\odot 2.\]

\(a_t\) is a martingale, hence
\[(19) \quad \frac{d}{dt}\mathbb{E}[\text{Cov}(Y_1|\mathcal{F}_t)] = -\frac{d}{dt}\mathbb{E}[a_t] = -\mathbb{E}[\Gamma_t^2].\]
Our next goal is to recover $v_t$ from the martingale $a_t$.

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

$$
\mathbb{E} \left[ \|v_t\|^2 \right] = \int_0^t \text{Tr} \left( \mathbb{E} \left[ \left( \Gamma_s - \mathbb{I}_d \right)^2 \right] \right) ds.
$$

**Proof.** We begin by writing

$$
da_t = dB_t + \left( \Gamma_t - \mathbb{I}_d \right) dB_t.
$$

Using Fubini’s theorem then yields

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

Therefore, defining $\tilde{v}_t = \int_0^t \frac{\Gamma_s - \mathbb{I}_d}{1-s} dB_s$ 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 $\int_0^1 (1-t)Q'_t dB_t$, we now have, for any $s \in [0,1]$

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

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

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

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

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

$$
\text{Ent} (Y_1 | \gamma) = \frac{1}{2} \int_0^1 \frac{1}{2} \text{Tr} \left( \mathbb{E} \left[ \left( \Gamma_t - \mathbb{I}_d \right)^2 \right] \right) ds dt = \frac{1}{2} \int_0^1 \frac{1}{2} \text{Tr} \left( \mathbb{E} \left[ \left( \Gamma_t - \mathbb{I}_d \right)^2 \right] \right) dt.
$$

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

Lemma 11. It holds that
\[
\mathbb{E} \left[ \text{Tr}(\Gamma_t) \right] = d - (1 - t) \left( d - \text{Tr}(\Sigma) + \mathbb{E} \left[ \|v_t\|^2 \right] \right).
\]

Proof. The identity can be obtained through integration by parts. By Lemma 10,
\[
\mathbb{E}[\|v_t\|^2] = \frac{\int_0^t \text{Tr} \left( \mathbb{E} \left[ (\Gamma_s - I_d)^2 \right] \right)}{(1 - s)^2} ds
= \frac{t}{(1 - s)^2} \text{Tr} \left( \mathbb{E} \left[ \Gamma_s^2 \right] \right) ds - 2 \int_0^t \frac{\text{Tr} \left( \mathbb{E} \left[ \Gamma_s \right] \right)}{(1 - s)^2} ds + \int_0^t \frac{\text{Tr} \left( I_d \right)}{(1 - s)^2} ds.
\]

Since, by Lemma 9, \( \frac{d}{dt} \mathbb{E} \left[ \text{Cov} \left( Y_1 | F_t \right) \right] = -\mathbb{E} \left[ \Gamma_t^2 \right] \) integration by parts shows
\[
\int_0^t \frac{\text{Tr} \left( \mathbb{E} \left[ \Gamma_s^2 \right] \right)}{(1 - s)^2} ds = \text{Tr} \left( \mathbb{E} \left[ \Gamma_t \right] \right) - \frac{t}{1 - t} \text{Tr} \left( \mathbb{E} \left[ \Gamma_t \right] \right) + 2 \int_0^t \frac{\text{Tr} \left( \mathbb{E} \left[ \text{Cov} \left( Y_1 | F_s \right) \right] \right)}{(1 - s)^3} ds,
\]
where we have used (17) and the fact \( \text{Cov} \left( Y_1 | F_0 \right) = \text{Cov} \left( Y_1 \right) = \Sigma \). Plugging this into the previous equation shows
\[
\mathbb{E}[\|v_t\|^2] = \text{Tr} \left( \Sigma \right) - \frac{t}{1 - t} \text{Tr} \left( \mathbb{E} \left[ \Gamma_t \right] \right) + \frac{d}{1 - t} - d.
\]
or equivalently
\[
\mathbb{E} \left[ \text{Tr}(\Gamma_t) \right] = d - (1 - t) \left( d - \text{Tr}(\Sigma) + \mathbb{E} \left[ \|v_t\|^2 \right] \right).
\]

Lemma 12. Assume that \( \text{Ent} \left( Y_1 || \gamma \right) < \infty \). Then \( \Gamma_t \) is almost surely invertible for all \( t \in [0, 1) \) and, moreover, there exists a constant \( m = m_\mu > 0 \) for which
\[
\sigma_t \geq m, \ \forall t \in [0, 1).
\]

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 \( \sigma_t \). The key to showing this is identity (21), due to which,
\[
\text{Ent} \left( Y_1 || \gamma \right) = \frac{1}{2} \int_0^1 \text{Tr} \left( \mathbb{E} \left[ (\Gamma_t - I_d)^2 \right] \right) dt.
\]

Recall that, by Equation (17), \( \Gamma_t = \frac{\text{Cov}(Y_1|F_t)}{1 - t} \) and observe that, by Proposition 5, if \( \text{Cov} \left( Y_1 | F_s \right) \) is not invertible for some \( 0 \leq s < 1 \) then \( \text{Cov} \left( Y_1 | F_s \right) \) is also not invertible for any \( t > s \). Under this event,
we would have that \( \int_0^1 \frac{\text{Tr}((\Gamma_t - I_d)^2)}{1-t} dt = \infty \) which, using the above display, implies that the probability of this event must be zero. Therefore, \( \Gamma_t \) 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

\[
\text{Tr} \left( \mathbb{E} \left[ (\Gamma_t - I_d)^2 \right] \right) \geq \text{Tr} \left( \mathbb{E} \left[ (\Gamma_{t'} - I_d)^2 \right] \right) \geq (1 - \sigma_{t'})^2 \geq 1 - 2\sigma_{t'}.
\]

Since, by Lemma 9, \( \mathbb{E} [\text{Cov} (Y_1 | F_t)] \) is non-increasing, 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 \( \text{Ent}(Y_1 | \gamma) < \infty \) combined with Equation (21) and with the last two displays finally gives

\[
\int_0^1 \frac{\text{Tr} \left( \mathbb{E} \left[ (\Gamma_t - I_d)^2 \right] \right)}{1-t} dt \geq \int_0^1 \frac{1-2\sigma_t}{1-t} dt \geq \sum_{i=1}^\infty \int_{t_i}^{t_{i+1}} \frac{1}{2(1-t)} dt \geq \log 2 \sum_{i=1}^\infty \frac{1}{2^i},
\]

which leads to a contradiction and completes the proof.

\[\Box\]

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

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

The same lemma also shows that for some \( m > 0 \) one has

\[
\int_0^1 \frac{\sigma_s^2}{t} dt \leq \frac{1-t}{m^2}.
\]

Therefore, we attain that

\[
\text{Ent}(S_n || G) \leq \frac{1}{m^2} \int_0^1 \frac{\text{Tr} \left( \mathbb{E} \left[ \Gamma_t^2 \right] - \mathbb{E} \left[ \tilde{\Gamma}_t \right]^2 \right)}{1-t} dt.
\]

(23)

Next, observe that, by Itô’s isometry,

\[
\text{Cov}(X) = \int_0^1 \mathbb{E} \left[ \Gamma_t^2 \right] dt.
\]

Hence, as long as \( \text{Cov}(X) \) is finite, \( \mathbb{E} \left[ \Gamma_t^2 \right] \) 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

\[
\lim_{n \to \infty} \text{Tr} \left( \mathbb{E} \left[ \Gamma_t^2 \right] - \mathbb{E} \left[ \tilde{\Gamma}_t \right]^2 \right) = 0, \ \forall t \in A.
\]

(24)
Indeed, by the law of large numbers, \( \hat{\Gamma}_t \) almost surely converges to \( \sqrt{E[\Gamma_t^2]} \). 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

\[
\hat{\Gamma}_t \leq \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\Gamma_t(i))^2} + I_d \leq \frac{1}{n} \sum_{i=1}^{n} (\Gamma_t(i))^2 + I_d,
\]

to deduce that \( \hat{\Gamma}_t \) is uniformly integrable as well. An application of Vitali’s convergence theorem (see [28], for example) 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

\[
\frac{\text{Tr} \left( E[\Gamma_t^2] - E[\hat{\Gamma}_t]^2 \right)}{1 - t}
\]

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

\[
\lim_{n \to \infty} \text{Ent}(S_n || G) = 0,
\]

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

\[
\hat{\Gamma}_t \geq \frac{1}{n} \sum_{i=1}^{n} \Gamma_t(i).
\]

It follows that

\[
\text{Tr} \left( E[\Gamma_t^2] - E[\hat{\Gamma}_t]^2 \right) \leq \text{Tr} \left( E[\Gamma_t^2] - E[\Gamma_t]^2 \right) \leq \text{Tr} \left( E[(\Gamma_t - I_d)^2] \right).
\]

So we have

\[
\frac{1}{m^2} \int_0^{1} \frac{\text{Tr} \left( E[\Gamma_t^2] - E[\hat{\Gamma}_t]^2 \right)}{1 - t} dt \leq \frac{1}{m^2} \int_0^{1} \frac{\text{Tr} \left( E[(\Gamma_t - I_d)^2] \right)}{1 - t} dt \overset{(21)}{=} \frac{2}{m^2} \text{Ent}(Y_t || \gamma) < \infty.
\]

This completes the proof.

4.3. **Quantitative bounds for log concave random vectors.** In this section, we make the additional assumption that the measure \( \mu \) 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 \( \Gamma_t \). We begin by recording several useful facts concerning this process.

**Lemma 13.** The process \( \Gamma_t \) has the following properties:

1. If \( \mu \) is log concave, then for every \( t \in [0, 1] \), \( \Gamma_t \preceq \frac{1}{n} I_d \) almost surely.
2. If \( \mu \) is also 1-uniformly log concave, then for every \( t \in [0, 1] \), \( \Gamma_t \leq I_d \) almost surely.
PROOF. Denote by \( \rho_t \) the density of \( Y_1 | \mathcal{F}_t \) with respect to the Lebesgue measure with \( \rho := \rho_0 \) being the density of \( \mu \). By Proposition 2 with \( C_t = \frac{1}{1-t} \), we can calculate the ratio between \( \rho_t \) and \( \rho \). In particular, we have

\[
\frac{d}{dt} \Sigma_t^{-1} = - \Sigma_t^{-1} \left( \frac{d}{dt} \Sigma_t \right) \Sigma_t^{-1} = \frac{1}{(1-t)^2} I_d.
\]

Solving this differential equation with the initial condition \( \Sigma_0^{-1} = 0 \), we find that

\[
\Sigma_t^{-1} = \frac{t}{1-t} I_d.
\]

Since the ratio between \( \rho_t \) and \( \rho \) is proportional to the density of a Gaussian with covariance \( \Sigma_t \), we thus have

\[
- \nabla^2 \log(\rho_t) = - \nabla^2 \log(\rho) + \frac{t}{1-t} I_d.
\]

Now, if \( \mu \) 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

\[
\text{Cov} \left( Y_1 | \mathcal{F}_t \right) \preceq \frac{t}{1-t} I_d \quad \text{and, using (17),}
\]

\[
\Gamma_t \preceq \frac{t}{1-t} I_d.
\]

If \( \mu \) is also \( 1 \)-uniformly log-concave then \( - \nabla^2 \log(\rho) \succeq I_d \) and almost surely

\[
- \nabla^2 \log(\rho_t) \succeq \frac{t}{1-t} I_d.
\]

By the same argument this implies

\[
\Gamma_t \preceq I_d.
\]

\( \square \)

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

\[
\text{(25)} \quad \text{Ent}(S_n || G) \leq \frac{1}{n} \int_0^1 \frac{\text{Tr} \left( \frac{\left( (\Gamma_t^2 - \mathbb{E}[\Gamma_t^2]) \right)^2}{(1-t)^2 \sigma_t^2} \right)}{\sigma_t^{-2}} ds dt.
\]

Attaining an upper bound on the right hand side amounts to a concentration estimate for the process \( \Gamma_t^2 \) and a lower bound on \( \sigma_t \). These two tasks are the objective of the following two lemmas.

**Lemma 14.** If \( \mu \) is log concave and isotropic then for any \( t \in [0, 1) \),

\[
\text{Tr} \left( \mathbb{E} \left[ (\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2 \right] \right) \leq \frac{1-t}{t^2} \left( \frac{d(1+t)}{t^2} + 2\mathbb{E} \left[ ||v_t||^2 \right] \right),
\]

and

\[
\text{Tr} \left( \mathbb{E} \left[ (\Gamma_t^2 - \mathbb{E}[\Gamma_t^2])^2 \right] \right) \leq C \frac{d^4}{(1-t)^4}
\]

for a universal constant \( C > 0 \).

**Proof.** The isotropicity of \( \mu \), used in conjunction with the formula given in Lemma 11, yields

\[
\text{Tr} \left( \mathbb{E} \left[ \Gamma_t^2 \right] \right) \geq \frac{1}{d} \text{Tr} (\mathbb{E}[\Gamma_t] )^2 \geq d - 2(1-t)\mathbb{E} \left[ ||v_t||^2 \right],
\]
where the first inequality follows by convexity. Since $\mu$ is log concave, Lemma 13 ensures that, almost surely, $\Gamma_t \preceq \frac{1}{t^2} I_d$. Therefore,

$$
\text{Tr} \left( \mathbb{E} \left[ (\Gamma_t^2 - \mathbb{E} [\Gamma_t^2])^2 \right] \right) \leq \frac{1}{t^4} \text{Tr} \left( \mathbb{E} \left[ (I_d - t^2 \Gamma_t^2)^2 \right] \right) \\
\leq \frac{1}{t^4} \text{Tr} \left( \mathbb{E} \left[ I_d - t^2 \Gamma_t^2 \right] \right) \\
\leq \frac{1}{t^4} \left( \frac{d(1 + t)}{t^2} + 2 \mathbb{E} [\|v_t\|^2] \right).
$$

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

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

So,

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

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

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

\[\square\]

**Lemma 15.** Suppose that $\mu$ 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, $t \in \left[ \frac{c}{d^2}, 1 \right]$, $\sigma_t \geq \frac{c}{d^2}$.

**Proof.** By Lemma 9, we have

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

Moreover, by convexity,

$$
\mathbb{E} \left[ \text{Cov} (Y_1 | \mathcal{F}_t)^2 \right] \preceq \mathbb{E} \left[ \text{Cov} (Y_1^\otimes 2 | \mathcal{F}_t)^2 \right] \preceq \mathbb{E} \left[ \|Y_1\|^4 \right] I_d.
$$

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

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

Consequently, $\frac{d}{dt} \mathbb{E} \left[ \text{Cov} (Y_1 | \mathcal{F}_t) \right] \geq - \frac{Cd^2}{(1 - t)^2} I_d$, and since $\text{Cov} (Y_1 | \mathcal{F}_0) = I_d$,

$$
\mathbb{E} \left[ \text{Cov} (Y_1 | \mathcal{F}_t) \right] \geq \left( 1 - Cd^2 \int_0^t \frac{1}{(1 - s)^2} ds \right) I_d = \left( 1 - \frac{Cd^4}{1 - t} \right) I_d.
$$
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

$$
\mathbb{E} \left[ \text{Cov}(Y_t \mid F_t) \right] \geq \frac{1}{2} I_d,
$$

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

$$
(26) \quad \frac{d}{dt} \mathbb{E} [\Gamma_t] = \frac{d}{dt} \mathbb{E} \left[ \text{Cov}(Y_t \mid F_t) \right] (\text{Lemma } 9) \leq \frac{\mathbb{E} [\text{Cov}(Y_t \mid F_t)] - (1-t)\mathbb{E} [\Gamma_t^2]}{(1-t)^2} = \frac{\mathbb{E} [\Gamma_t] - \mathbb{E} [\Gamma_t^2]}{1-t}.
$$

Since, by Lemma 13, $\Gamma_t \leq \frac{1}{2} I_d$, we have the bound

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

Now, consider the differential equation $f''(t) = -\frac{f(0)}{t}, f'(\frac{1}{6Cd^2}) = \frac{1}{2}$. Its unique solution is $f(t) = \frac{1}{6Cd^2}$. Thus, Gromwall’s inequality shows that $\sigma_t \geq \frac{1}{6Cd^2}$, which concludes the proof.

**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 $t \in [0, 1)$,

$$
\int_0^1 \sigma_s^{-2} ds \leq Cd^4 (1-t),
$$

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

$$
\int_0^1 \mathbb{E} \left[ \text{Tr} \left( \left( \Gamma_t^2 - \mathbb{E} [\Gamma_t^2] \right)^2 \right) \right] \left( \int_0^1 \sigma_s^{-2} ds \right) dt \leq Cd^4 \int_0^1 \mathbb{E} \left[ \text{Tr} \left( \left( \Gamma_t^2 - \mathbb{E} [\Gamma_t^2] \right)^2 \right) \right] dt.
$$

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

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

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

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

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

$$
Cd^4 \int_{d^{-1}}^1 \mathbb{E} \left[ \text{Tr} \left( \left( \Gamma_t^2 - \mathbb{E} [\Gamma_t^2] \right)^2 \right) \right] \frac{1}{(1-t)\sigma_t^2} dt \leq C'd^8 \int_{d^{-1}}^1 \frac{t^2 \mathbb{E} \left[ \text{Tr} \left( \left( \Gamma_t^2 - \mathbb{E} [\Gamma_t^2] \right)^2 \right) \right]}{1-t} dt
$$

$$
\leq 2C'd^9 \int_{d^{-1}}^1 \left( \frac{1}{t^2} + \mathbb{E} \left[ \|v_t\|^2 \right] \right) dt
$$

$$
\leq 4C'd^{10} (1 + \text{Ent}(Y_t \| G)),
$$

which concludes the proof.
where the first inequality uses Lemma 15 and the second one uses Lemma 14. This establishes
\[ \text{Ent}(S_n||G) \leq \frac{Cd^{10}(1 + \text{Ent}(Y_1||G))}{n}. \]

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

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

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

**PROOF.** By Lemma 13, we have that \( \Gamma_t \preceq I_d \) almost surely. Using this together with the identity given by Lemma 11, and proceeding in similar fashion to Lemma 14 we obtain
\[ \text{Tr} \left( E \left[ \Gamma_t^2 \right] \right) \geq \frac{1}{d} \text{Tr} (E[I_t]) \geq d - 2(1 - t) \left( d - \text{Tr}(\Sigma) + E[\|v_t\|^2] \right), \]
and
\[ \text{Tr} \left( E \left[ (\Gamma_t^2 - E[\Gamma_t^2])^2 \right] \right) \leq \text{Tr} \left( E \left[ (\Gamma_t^2 - I_d)^2 \right] \right) \leq \text{Tr} \left( E \left[ I_d - \Gamma_t^2 \right] \right) \leq 2(1 - t) \left( d - \text{Tr}(\Sigma) + E[\|v_t\|^2] \right). \]

Also, recalling (26) and since \( \Gamma_t \preceq I_d \) we get
\[ \frac{d}{dt} E[\Gamma_t] = \frac{E[\Gamma_t] - E[\Gamma_t^2]}{1 - t} \geq 0, \]
which shows that \( \sigma_t \) is bounded from below by a non-decreasing function and so \( \sigma_t \geq \sigma_0 \) which is the minimal eigenvalue of \( \Sigma. \)

**PROOF OF THEOREM 7.** Plugging the bounds given in Lemma 16 into Equation (25) yields
\[
\text{Ent}(S_n||G) \leq \frac{1}{n} \int_0^1 \frac{1}{(1 - t)^2} \text{Tr} \left( (\Gamma_t^2 - E[\Gamma_t^2])^2 \right) \left( \int_1^t \sigma_s^{-2} ds \right) dt
\]
\[
\leq \frac{2 \left( d + \frac{1}{0} E[\|v_t\|^2] dt \right)}{\sigma_0^4 n} \overset{(18)}{=} \frac{2 \left( d + 2 \text{Ent}(X||\gamma) \right)}{\sigma_0^4 n},
\]
which completes the proof.

**References.**


with applications. *Entropy* **20** Paper No. 185, 24. MR3782876


R. Eldan  
D. Mikulincer  
Weizmann Institute of Science  
234 Herzl St.  
PO Box 26  
Rehovot 7610001  
Israel  
E-mail: ronenelden@gmail.com  
danimiku@gmail.com

A. Zhai  
Stanford University  
450 Serra Mall  
Building 380  
Stanford  
CA 94305  
E-mail: azhai@stanford.edu