

A high-dimensional CLT in \mathcal{W}_2 distance with near optimal convergence rate

Alex Zhai¹ 

Received: 10 June 2016 / Revised: 23 February 2017 / Published online: 24 March 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract Let X_1, \dots, X_n be i.i.d. random vectors in \mathbb{R}^d with $\|X_1\| \leq \beta$. Then, we show that

$$\frac{1}{\sqrt{n}}(X_1 + \dots + X_n)$$

converges to a Gaussian in quadratic transportation (also known as “Kantorovich” or “Wasserstein”) distance at a rate of $O\left(\frac{\sqrt{d}\beta \log n}{\sqrt{n}}\right)$, improving a result of Valiant and Valiant. The main feature of our theorem is that the rate of convergence is within $\log n$ of optimal for $n, d \rightarrow \infty$.

Mathematics Subject Classification 60F05

1 Introduction

The central limit theorem states that if X_1, X_2, \dots, X_n are independent and identically distributed random variables, then $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ is approximately Gaussian. It is well-known that by various metrics the distance from Gaussian decays at a rate of $n^{-1/2}$; for example, the celebrated Berry–Esseen bound states that $|\mathbf{P}(S \leq t) - \mathbf{P}(Z \leq t)| = O(n^{-1/2} \mathbf{E}|X_i|^3)$. Moreover, this bound is optimal to within a constant.

The same principle holds if we allow the X_i to be \mathbb{R}^d -valued, and an extensive literature was developed, tracing back at least to the 1940’s [4] (see also [5] and references therein), around establishing multivariate central limit theorems with good

✉ Alex Zhai
azhai@stanford.edu

¹ Stanford University, Stanford, CA, USA

convergence rates. One new consideration that arises in the multivariate setting is that the convergence rate is expressed in terms of not only n but also the dimension d . This dependence on d , and in particular when d is growing with n , was studied by Nagaev [17], Senatov [22], Götze [15], Bentkus [2], and Chen and Fang [13], among others. These works focus on convergence in probabilities of convex sets, which we will call convergence in convex-indicator (CI) distance.

In addition to being a natural question, obtaining good dependence on dimension has recently been of interest in various applications. Bubeck and Ganguly [12] prove a central limit theorem for Wishart matrices (relevant to random geometric graphs, see also [11]), and Chernozhukov et al. [14] prove a central limit theorem for maxima of sums of independent random vectors (with applications in high-dimensional statistical inference). Another relevant work is that of Valiant and Valiant [24],¹ who prove central limit theorems for transportation distance and generalized multinomial distributions and use them to construct lower bounds for estimating entropy.

In this paper, we prove a multivariate central limit theorem for quadratic transportation distance whose rate of convergence is within $\log n$ of optimal in both the number of summands n and the dimension d , improving the result of Valiant and Valiant [24]. To our knowledge, this is the first general multivariate central limit theorem whose convergence rate is optimal to within logarithmic factors in both n and d , albeit not for the CI metric that is most commonly studied in the literature.² Additionally, we believe that the method of proof based on Talagrand's transportation inequality, described in Sect. 1.2, is of independent interest. We also note that in certain regimes our result implies stronger bounds in CI distance than what is known in the existing literature, as elaborated in Sect. 1.1.

To state the result, recall that for two probability measures μ and ν on \mathbb{R}^d and a number $p \geq 1$, the L^p transportation distance³ $\mathcal{W}_p(\mu, \nu)$ is defined to be

$$\mathcal{W}_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \int \|x - y\|^p d\gamma(x, y) \right)^{\frac{1}{p}},$$

where $\Gamma(\mu, \nu)$ is the space of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with μ and ν as marginals. In other words, $\mathcal{W}_p(\mu, \nu)$ measures how closely μ and ν may be coupled. If X and Y are random variables with distributions μ and ν , respectively, we will also write

$$\mathcal{W}_p(X, Y) = \mathcal{W}_p(\mu, \nu).$$

Our main result is the following theorem concerning the L^2 (or “quadratic”) transportation distance.

¹ See [25] for the full version.

² It should be noted that the bounds obtained by Bubeck and Ganguly [12] are also optimal to within logarithmic factors, but they are specific to Wishart matrices. We mention also the work of Bentkus and Götze [3], which obtains optimal bounds for quadratic forms under certain somewhat specialized assumptions.

³ Other names appearing in the literature include “Monge–Kantorovich distance”, “Kantorovich distance”, and “Wasserstein distance”. We refer to [26] for a historical discussion of the concept.

Theorem 1.1 *Let X_1, \dots, X_n be independent random vectors with mean zero, covariance Σ , and $\|X_i\| \leq \beta$ almost surely for each i . Let $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$, and let Z be a Gaussian with covariance Σ . Then,*

$$\mathcal{W}_2(S_n, Z) \leq \frac{5\sqrt{d}\beta(1 + \log n)}{\sqrt{n}}.$$

This bound improves by a factor of \sqrt{d} the result of Valiant and Valiant [24], who obtain under the same assumptions a $O\left(\frac{d\beta \log n}{\sqrt{n}}\right)$ bound for \mathcal{W}_1 distance. In fact, Theorem 1.1 is within a $\log n$ factor of optimal, in the sense that one cannot have a convergence rate faster than $O(\sqrt{d}\beta/\sqrt{n})$, as shown by the following proposition.

Proposition 1.2 *Let $(X_i)_{i=1}^n, S_n, Z$, and β be as in Theorem 1.1. Suppose further that the X_i take values in the lattice $\beta\mathbb{Z}^d$. Then,*

$$\liminf_{n \rightarrow \infty} \sqrt{n}\mathcal{W}_2(S_n, Z) \geq \frac{\sqrt{d}\beta}{4}.$$

The proof is routine and is given in ‘‘Proof of Proposition 1.2’’ of Appendix section; it is based on the fact that a typical point in \mathbb{R}^d will be a distance $O(\sqrt{d}\beta)$ from the closest point in $\beta\mathbb{Z}^d$.

Several other works in the literature have studied central limit theorems for \mathcal{W}_p distance. In the multivariate setting, the recent work of Bonis [10] proves a $O(1/\sqrt{n})$ convergence rate for \mathcal{W}_2 distance under the assumption $\mathbf{E}\|X_1\|^4 < \infty$. However, Bonis’ result does not have an explicit dependence on the dimension, which is the main point of this paper.

We mention also the work of Rio (see [19,20]), who analyzed for the one-dimensional setting convergence in \mathcal{W}_p distance under various moment assumptions. For \mathcal{W}_2 , he proves a $O(1/\sqrt{n})$ convergence rate under the assumption of finite fourth moments; we refer the reader to [19] for statements about other values of p . An alternative proof of Rio’s result for \mathcal{W}_2 was given by Bobkov [7] (see also [8]). We note that Talagrand’s transportation inequality also makes an appearance in [7], but the way it is used is substantially different from the approach of this paper.

The above literature leads us to believe that Theorem 1.1 can be improved to remove the $\log n$ factor (this was also conjectured in [24]). We remark that the extra $\log n$ factor in our proof comes from a harmonic series arising from repeated applications of Lemma 1.6 below.

1.1 Comparison with convex-indicator bounds

For two measures μ and ν on \mathbb{R}^d , we define the convex-indicator (CI) distance Δ_{CI} by

$$\Delta_{CI}(\mu, \nu) = \sup_{A \subset \mathbb{R}^d \text{ convex}} |\mu(A) - \nu(A)|,$$

and as with \mathcal{W}_p distance, we will write $\Delta_{CI}(X, Y) = \Delta_{CI}(\mu, \nu)$ if X has distribution μ and Y has distribution ν . As mentioned earlier, CI distance is perhaps the most widely studied metric in the high-dimensional central limit theorem literature (see e.g. [2, 6, 15, 17, 21, 22]). The best convergence rate seems to be due to Bentkus [2]. For simplicity, we state his theorem in the i.i.d. case (the original paper contains a somewhat more general formulation).

Theorem 1.3 (Bentkus, i.i.d. case of Theorem 1.1 in [2]) *Let X_1, \dots, X_n be i.i.d. \mathbb{R}^d -valued random variables with mean zero, identity covariance, and $(\mathbf{E}\|X_1\|^3)^{\frac{1}{3}} = \beta_3$. Let $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$, and let Z be a standard Gaussian. Then, there is a constant C such that*

$$\Delta_{CI}(S_n, Z) \leq \frac{Cd^{1/4}\beta_3^3}{\sqrt{n}}.$$

Note that this recovers the Berry–Esseen bound for $d = 1$. Nagaev [17] established earlier that this bound is within $d^{1/4}$ of optimal in the sense that there exist examples which would contradict the above theorem if $d^{1/4}$ were replaced with some term going to zero as $d \rightarrow \infty$. However, the family of examples in [17] is for a specific relation between n, d , and β_3 , which, as we shall see, may not be representative of the behavior of many natural cases.

Although our result is for \mathcal{W}_2 distance, when the dimension d fixed, convergence in \mathcal{W}_2 distance to a Gaussian implies convergence in probabilities of convex sets.⁴ Specifically, we have the following proposition.

Proposition 1.4 *Let T be any \mathbb{R}^d -valued random variable, and let Z be a standard d -dimensional Gaussian. Then, for a universal constant C ,*

$$\Delta_{CI}(T, Z) \leq Cd^{\frac{1}{6}}\mathcal{W}_2(T, Z)^{\frac{2}{3}}.$$

For the short proof (involving Gaussian surface area of convex sets), see “Proof of Proposition 1.4” of Appendix section. Applying Proposition 1.4 to Theorem 1.1, we have the following corollary.

Corollary 1.5 *Let X_1, \dots, X_n be independent random vectors in \mathbb{R}^d with mean zero, identity covariance, and $\|X_i\| \leq \beta$ almost surely for each i . Let $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$, and let Z be a standard Gaussian. Then, for a universal constant C ,*

$$\Delta_{CI}(S_n, Z) \leq \frac{Cd^{\frac{1}{2}}\beta^{\frac{2}{3}}(1 + \log n)^{\frac{2}{3}}}{n^{\frac{1}{3}}}.$$

⁴ On the other hand, convergence in probabilities of convex sets does not in general imply convergence in \mathcal{W}_2 distance, and we do not know of any easy way to derive a result similar to Theorem 1.1 from Theorem 1.3.

Before we proceed, it should be noted that a few issues arise in comparing high-dimensional central limit theorems. To start with, concepts such as “third moments” are less clear-cut. For example, for an \mathbb{R}^d -valued random variable $X = (X_1, \dots, X_d)$, both $\mathbf{E}\|X\|^3$ and $\sum_{i=1}^d \mathbf{E}|X_i|^3$ are potentially reasonable generalizations of the one-dimensional third moment. A related issue is how to normalize covariances. In the one-dimensional setting, we can always, without loss of generality, normalize X so that $\mathbf{E}X^2 = 1$. In higher dimensions, linear transformations on the covariance matrix have a more complicated effect on quantities such as the aforementioned third moments.

Corollary 1.5 has a suboptimal $n^{-1/3}$ dependence on n (compared to the correct order $n^{-1/2}$ obtained in Theorem 1.3). Nevertheless, Corollary 1.5 yields better information in some cases. Let us suppose that $\|X_1\| = \sqrt{d}$ almost surely; this includes natural examples such as when X_1 is $\pm\sqrt{d}$ times a standard basis vector, with the sign and the basis vector chosen uniformly at random. Then, we have $\beta = \beta_3 = \sqrt{d}$, so that Theorem 1.3 gives

$$\Delta_{CI}(S_n, Z) \leq \frac{Cd^{7/4}}{n^{1/2}},$$

while Corollary 1.5 gives

$$\Delta_{CI}(S_n, Z) \leq \frac{Cd^{5/6}(1 + \log n)^{2/3}}{n^{1/3}}.$$

We find that the second bound is stronger than the first whenever $d = \tilde{\Omega}(n^{2/11})$, where the tilde suppresses logarithmic factors. In particular, note that the second bound gives $\Delta_{CI}(S_n, Z) = o(1)$ (i.e. says something non-trivial) as soon as $d = \tilde{o}(n^{2/5})$, while the first bound requires $d = o(n^{2/7})$.⁵ In this sense, when $\|X_1\| = \sqrt{d}$ almost surely, Corollary 1.5 gives convergence for a larger range of d .

We mention here that in high-dimensional settings, d may indeed be as large as a power of n . For example, the earlier mentioned work of Bubeck and Ganguly [12], when applied in the context of [11], concerns $d \approx n^{2/3}$ (after converting to our notation). The work of Chernozhukov et al. [14] even considers $d \approx e^{n^c}$ for a constant c , albeit working under a much weaker notion of convergence.

1.2 Idea of the proof

The proof of Theorem 1.1 follows a Lindeberg-type strategy of gradually replacing X_i 's with Gaussians. However, instead of working with sufficiently smooth test functions, we directly compare probability densities. A major ingredient for accomplishing this is Talagrand's transportation inequality. To our knowledge, this variation of the Lindeberg strategy has not appeared before in the literature, and the idea may be of

⁵ We remark that even if the $d^{1/4}$ in Theorem 1.3 were replaced by a constant as in Nagaev's lower bound, it would only give $\Delta_{CI}(S_n, Z) = o(1)$ for $d = o(n^{1/3})$, which is still more restrictive than $d = o(n^{2/5})$. Thus, Corollary 1.5 proves that under the assumption $\|X_1\| = \sqrt{d}$, convergence in Δ_{CI} is actually faster than indicated by Nagaev's example (which does not satisfy $\|X_1\| = \sqrt{d}$).

use in other settings. Our argument rests upon the following key lemma, which bounds the error arising from replacing X_i with a Gaussian.

Lemma 1.6 *Let X be a \mathbb{R}^k -valued random variable with mean 0, covariance Σ , and $\|X\| \leq \beta$ almost surely. Let Z_t denote a Gaussian of mean 0 and covariance $t\Sigma$ independent of X . Let σ_{\min}^2 denote the smallest eigenvalue of Σ . Then, for any $n \geq \frac{5\beta^2}{\sigma_{\min}^2}$, we have*

$$\mathcal{W}_2(Z_n, Z_{n-1} + X) \leq \frac{5\sqrt{k}\beta}{n}.$$

Remark 1.1 The assumption on n implies that $n \geq 5k$, because

$$n \geq \frac{5\beta^2}{\sigma_k^2} \geq \frac{5}{\sigma_k^2} \mathbf{E}\|X\|^2 = \frac{5}{\sigma_k^2} \sum_{i=1}^k \sigma_i^2 \geq 5k.$$

Heuristically, Lemma 1.6 says that when you add an independent random variable X to a Gaussian Z_{n-1} , the resulting distribution is still nearly Gaussian. The hypothesis that n be sufficiently large is required to ensure that X is small compared to Z_{n-1} . Note that the dimension k appearing in Lemma 1.6 is not necessarily equal to d . This is a subtle but important point—we will selectively apply the estimate of Lemma 1.6 to only a subset of the coordinates depending on the variance of X in those directions.

Theorem 1.1 follows from repeated applications of Lemma 1.6. To prove Lemma 1.6, our strategy is to take advantage of the fact that we can explicitly compute the density of the Gaussian Z_n , and we also have a fairly explicit form for the density of $Z_{n-1} + X$. We can then make precise density estimates, which are conveniently translated into \mathcal{W}_2 estimates via (a variant of) Talagrand's transportation inequality.

1.3 Organization of the paper

The rest of the paper is organized as follows. In Sect. 2, we prove Theorem 1.1 assuming Lemma 1.6. In Sect. 3, we provide some background on Talagrand's transportation inequality needed to prove Lemma 1.6. In particular, whereas the inequality is usually formulated in the setting of a standard n -dimensional Gaussian, we give a version for general Gaussians. Finally, Sect. 4 gives the proof of Lemma 1.6, filling in the technical details of the strategy described above.

2 Proof of Theorem 1.1

We first show how to deduce Theorem 1.1 from Lemma 1.6. Recall however that the statement of Lemma 1.6 contains a hypothesis that $n \geq \frac{5\beta^2}{\sigma_{\min}^2}$. Thus, we will also need an *a priori* bound to estimate \mathcal{W}_2 distances for smaller n .

Luckily, a naïve bound suffices. For any mean-zero random variables X and Y , coupling them to be independent yields the inequality $\mathcal{W}_2(X, Y)^2 \leq \mathbf{E}\|X\|^2 + \mathbf{E}\|Y\|^2$.

The next lemma is a slight refinement of this observation to consider only a subset of coordinates.

Lemma 2.1 *Let $X = (X_1, \dots, X_d)$ and $Y = (Y_1, \dots, Y_d)$ be two \mathbb{R}^d -valued random variables with mean zero. Moreover, suppose that $\mathbf{E}((Y_{k+1}, \dots, Y_d) \mid Y_1, \dots, Y_k) = 0$. Then,*

$$\mathcal{W}_2(X, Y)^2 \leq \mathcal{W}_2((X_1, \dots, X_k), (Y_1, \dots, Y_k))^2 + \sum_{i=k+1}^d (\mathbf{E}X_i^2 + \mathbf{E}Y_i^2).$$

Proof For convenience, define $P_k : \mathbb{R}^d \rightarrow \mathbb{R}^d$ by $P_k(x_1, \dots, x_d) = (x_1, \dots, x_k, 0, \dots, 0)$. Let \tilde{X} and \tilde{Y} be a coupling of X and Y given by first sampling $P_k(\tilde{X})$ and $P_k(\tilde{Y})$ according to a coupling such that

$$\mathbf{E}\|P_k(\tilde{X}) - P_k(\tilde{Y})\|^2 = \mathcal{W}_2(P_k(X), P_k(Y))$$

and then sampling \tilde{X} and \tilde{Y} independently conditioned on $P_k(\tilde{X})$ and $P_k(\tilde{Y})$. Thus, $\tilde{X} - P_k(\tilde{X})$ and $\tilde{Y} - P_k(\tilde{Y})$ are independent conditioned on $P_k(\tilde{X})$ and $P_k(\tilde{Y})$. Then,

$$\begin{aligned} \mathcal{W}_2(X, Y)^2 &\leq \mathbf{E}\|\tilde{X} - \tilde{Y}\|^2 = \mathbf{E}\|\tilde{X} - \tilde{Y}\|^2 \\ &= \mathbf{E}\|(\tilde{X} - P_k(\tilde{X})) + (P_k(\tilde{X}) - P_k(\tilde{Y})) + (P_k(\tilde{Y}) - \tilde{Y})\|^2 \\ &= \mathbf{E}\|\tilde{X} - P_k(\tilde{X})\|^2 + \mathbf{E}\|P_k(\tilde{X}) - P_k(\tilde{Y})\|^2 + \mathbf{E}\|P_k(\tilde{Y}) - \tilde{Y}\|^2 \\ &= \mathcal{W}_2(P_k(X), P_k(Y))^2 + \sum_{i=k+1}^d (\mathbf{E}X_i^2 + \mathbf{E}Y_i^2). \end{aligned}$$

□

We are now ready for the main proof. The rough idea is to induct simultaneously on n and the dimension. At each step, if possible, we apply Lemma 1.6 to increase n . Otherwise, we apply Lemma 2.1 to increase the dimension.

Proof of Theorem 1.1 Using the notation in the statement of the theorem, we can assume without loss of generality that Σ takes the form

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{bmatrix},$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d > 0$. For each $n \geq 1$, define

$$S_n = \sum_{i=1}^n X_i,$$

and let Z_n denote a Gaussian with covariance $n\Sigma$.

Let $P_k : \mathbb{R}^d \rightarrow \mathbb{R}^k$ denote the projection onto the first k coordinates, and for $0 \leq k \leq d$, define

$$A_{n,k} = \mathcal{W}_2(P_k(S_n), P_k(Z_n)), \quad A_{n,0} = 0.$$

We will prove by induction on n and k that

$$A_{n,k} \leq 5\sqrt{k}\beta(1 + \log n) \tag{1}$$

for all $n \geq 1$ and $0 \leq k \leq d$. The theorem then follows by taking $k = d$.

Let us call (n, k) a *good pair* if (1) holds. We first prove the base cases. If $k = 0$, then (1) holds trivially. If $n = 1$, then by Lemma 2.1,

$$A_{1,k} = \mathcal{W}_2(P_k(X_1), P_k(Z_1)) \leq \sqrt{\mathbf{E}\|X_1\|^2 + \mathbf{E}\|Z_1\|^2} \leq 2\beta,$$

so again (1) holds.

For the inductive step, consider any $n > 1$ and $k > 0$. Our inductive hypothesis is that $(n - 1, k)$ and $(n, k - 1)$ are good pairs, and we will show that (n, k) is a good pair as well. If $n > \frac{5\beta^2}{\sigma_k^2}$, then we may apply Lemma 1.6 to $P_k(X_n)$, whose covariance is just the top-left $k \times k$ submatrix of Σ . This gives

$$\mathcal{W}_2(P_k(Z_{n-1} + X_n), P_k(Z_n)) \leq \frac{5\sqrt{k}\beta}{n}.$$

Consequently,

$$\begin{aligned} A_{n,k} &= \mathcal{W}_2(P_k(S_n), P_k(Z_n)) = \mathcal{W}_2(P_k(S_{n-1} + X_n), P_k(Z_n)) \\ &\leq \mathcal{W}_2(P_k(S_{n-1} + X_n), P_k(Z_{n-1} + X_n)) + \mathcal{W}_2(P_k(Z_{n-1} + X_n), P_k(Z_n)) \\ &\leq A_{n-1,k} + \frac{5\sqrt{k}\beta}{n} \leq 5\sqrt{k}\beta \left(1 + \log(n - 1) + \frac{1}{n} \right) \leq 5\sqrt{k}\beta(1 + \log n). \end{aligned}$$

Otherwise, if $n \leq \frac{5\beta^2}{\sigma_k^2}$, then by Lemma 2.1, we have

$$\begin{aligned} A_{n,k}^2 &\leq A_{n,k-1}^2 + 2n\sigma_k^2 \\ &\leq 25(k - 1)\beta^2(1 + \log n)^2 + 10\beta^2 \leq 25k\beta^2(1 + \log n)^2. \end{aligned}$$

We see in both cases that (n, k) is a good pair, completing the induction and the proof. □

3 A transportation inequality

It remains only to prove Lemma 1.6. As described earlier, the strategy we use is to translate closeness in probability densities into closeness in \mathcal{W}_2 distance. In this section, we establish the result needed for this purpose, which is based on the following inequality due to Talagrand.

Theorem 3.1 (Talagrand’s transportation inequality) *Let Z be a standard d -dimensional Gaussian with density ρ . Let μ be a probability density on \mathbb{R}^d and let $f(x) = \frac{d\mu}{d\rho}(x)$. Then,*

$$\mathcal{W}_2(\mu, \rho)^2 \leq 2\mathbf{E}f(Z) \log f(Z) = 2\mathcal{D}(\mu \parallel \rho).$$

Remark 3.1 We note that the above inequality is sharp: equality holds when Y is Gaussian with the same covariance as Z , but with a different mean. However, it can be far from optimal when the density of Y is not very “smooth”; indeed, in the extreme case where Y is not absolutely continuous with respect to Z , Theorem 3.1 says nothing at all. The need to ensure this “smoothness” explains the requirement that $n \geq \frac{5\beta^2}{\sigma_{\min}^2}$ in the statement of Lemma 1.6.

Theorem 3.1 is an example of a transportation-information inequality (also known as transportation-cost inequalities in the literature). Such inequalities were first studied by Marton [16] who showed their connection to concentration of measure phenomena (see also [9]).

In [23], Talagrand proves Theorem 3.1 using an inductive argument, following ideas of Marton [16]. The one-dimensional case is a (non-trivial!) calculus problem. Higher dimensions then follow by tensorization properties of \mathcal{W}_2 distance and relative entropy.

However, we cannot directly apply Talagrand’s transportation inequality in our case, because the covariance of our Gaussian is not the identity. Nevertheless, by modifying the proof only slightly, we can obtain a version of the inequality that applies to non-standard Gaussians, as captured in the next proposition.

Proposition 3.2 (Variant of Talagrand’s transportation inequality) *Let Z be a d -dimensional Gaussian having diagonal covariance*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{bmatrix}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d > 0$. Let $\rho : \mathbb{R}^d \rightarrow \mathbb{R}$ be the density of Z , and let Y be a \mathbb{R}^d -valued random variable with density $f(x)\rho(x)$. Then,

$$\mathcal{W}_2(Y, Z)^2 \leq 2 \sum_{i=1}^d \sigma_i^2 \left(\mathbf{E}f(Z)^2 - \mathbf{E}f_{(i)}(Z)^2 \right),$$

where $f_{(i)}$ is the “averaging” of f along the i -th coordinate defined by

$$f_{(i)}(x) = \frac{\int_{-\infty}^{\infty} f(x + te_i)\rho(x + te_i) dt}{\int_{-\infty}^{\infty} \rho(x + te_i) dt},$$

where $e_i \in \mathbb{R}^d$ are unit coordinate vectors.

The proof of Proposition 3.2 uses an elementary lemma involving conditional L^2 norms, which is proved in “Proof of Lemma 3.3” of Appendix section.

Lemma 3.3 *Let $A \in \mathcal{A}$ and $B \in \mathcal{B}$ be independent random variables and consider any function $f : \mathcal{A} \times \mathcal{B} \rightarrow \mathbb{R}$. Define*

$$\begin{aligned} f_A : \mathcal{B} &\rightarrow \mathbb{R}, & f_A(b) &= \mathbf{E}(f(A, B) \mid B = b) \\ f_B : \mathcal{A} &\rightarrow \mathbb{R}, & f_B(a) &= \mathbf{E}(f(A, B) \mid A = a). \end{aligned}$$

Then,

$$\mathbf{E}f(A, B)^2 + (\mathbf{E}f(A, B))^2 \geq \mathbf{E}f_A(B)^2 + \mathbf{E}f_B(A)^2.$$

Proof of Proposition 3.2 In fact, a slightly stronger inequality holds. In order to state it, let us define for each $0 \leq k \leq d$ the function

$$f_{[k]}(x) = \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(x + \sum_{i=k+1}^d t_i e_i\right) \rho\left(x + \sum_{i=k+1}^d t_i e_i\right) dt_{k+1} \cdots dt_d}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \rho\left(x + \sum_{i=k+1}^d t_i e_i\right) dt_{k+1} \cdots dt_d},$$

which may be thought of as the “averaging” of f over all but the first k coordinates. Note that $f_{[d]} = f$ and $f_{[0]} = 1$.

We claim that

$$\mathcal{W}_2(Y, Z)^2 \leq 2 \sum_{k=1}^d \sigma_k^2 \cdot \mathbf{E}\left(f_{[k]}(Z) \log f_{[k]}(Z) - f_{[k-1]}(Z) \log f_{[k-1]}(Z)\right). \tag{2}$$

This inequality is essentially a byproduct of the proof of Theorem 3.1 (see [23], section 3). Note that if $\sigma_k = 1$ for all k , then the sum in (2) telescopes to

$$2\mathbf{E}f(Z) \log f(Z) = 2\mathcal{D}(Y \parallel Z),$$

recovering Theorem 3.1. Although (2) is a direct consequence of the arguments in [23], for the sake of completeness we repeat the proof in “Proof of Equation (2)” of Appendix section.

Using (2) and the fact that $t \log t \leq t^2 - t$, we have

$$\begin{aligned} \mathcal{W}_2(Y, Z)^2 &\leq 2 \sum_{k=1}^d \sigma_k^2 \cdot \mathbf{E} \left(f_{[k]}(Z) \log f_{[k]}(Z) - f_{[k-1]}(Z) \log f_{[k-1]}(Z) \right) \\ &= 2\sigma_d^2 \cdot \mathbf{E} \left(f_{[d]}(Z) \log f_{[d]}(Z) \right) + 2 \sum_{k=2}^d \left(\sigma_{k-1}^2 - \sigma_k^2 \right) \mathbf{E} \left(f_{[k-1]}(Z) \log f_{[k-1]}(Z) \right) \\ &\leq 2\sigma_d^2 \cdot \mathbf{E} \left(f_{[d]}(Z)^2 - f_{[d]}(Z) \right) + 2 \sum_{k=2}^d \left(\sigma_{k-1}^2 - \sigma_k^2 \right) \mathbf{E} \left(f_{[k-1]}(Z)^2 - f_{[k-1]}(Z) \right) \\ &= 2\sigma_d^2 \cdot \mathbf{E} \left(f_{[d]}(Z)^2 - 1 \right) + 2 \sum_{k=2}^d \left(\sigma_{k-1}^2 - \sigma_k^2 \right) \mathbf{E} \left(f_{[k-1]}(Z)^2 - 1 \right) \\ &= 2 \sum_{k=1}^d \sigma_k^2 \cdot \mathbf{E} \left(f_{[k]}(Z)^2 - f_{[k-1]}(Z)^2 \right) \end{aligned}$$

Finally, for each k , we claim that

$$\mathbf{E} \left(f_{[k]}(Z)^2 - f_{[k-1]}(Z)^2 \right) \leq \mathbf{E} \left(f(Z)^2 - f_{(k)}(Z)^2 \right). \tag{3}$$

Indeed, this is actually an immediate consequence of Lemma 3.3. To simplify notation, write $Z = (Z', Z'', Z''')$, where Z' denotes the first $k - 1$ coordinates of Z , Z'' denotes the k -th coordinate, and Z''' denotes the last $d - k$ coordinates. In terms of these variables, we have

$$\begin{aligned} \mathbf{E} f_{[k-1]}(Z)^2 &= \mathbf{E} \left[\mathbf{E} \left(f(Z) \mid Z' \right)^2 \right] \\ \mathbf{E} f_{[k]}(Z)^2 &= \mathbf{E} \left[\mathbf{E} \left(f(Z) \mid Z', Z'' \right)^2 \right] \\ \mathbf{E} f_{(k)}(Z)^2 &= \mathbf{E} \left[\mathbf{E} \left(f(Z) \mid Z', Z''' \right)^2 \right] \\ \mathbf{E} f(Z)^2 &= \mathbf{E} \left[\mathbf{E} \left(f(Z)^2 \mid Z' \right) \right], \end{aligned}$$

Then, applying Lemma 3.3 conditioned on Z' with $A = Z''$ and $B = Z'''$ gives us precisely (3). Thus, we conclude that

$$\mathcal{W}_2(Y, Z)^2 \leq 2 \sum_{k=1}^d \sigma_k^2 \cdot \mathbf{E} \left(f(Z)^2 - f_{(k)}(Z)^2 \right),$$

as desired. □

4 Proof of Lemma 1.6

We finally conclude by proving Lemma 1.6. Henceforth, we use the notation in the statement of Lemma 1.6 and assume without loss of generality that

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k^2 \end{bmatrix},$$

so that $\sigma_{\min} = \min_{1 \leq i \leq k} \sigma_i$. It is more convenient to work with the normalization $Y = \frac{1}{\sqrt{n}}X$, so that $\|Y\| \leq \frac{\beta}{\sqrt{n}}$. Our goal is then to prove that

$$\mathcal{W}_2(Z_1, Z_{1-1/n} + Y) \leq \frac{5\sqrt{k}\beta}{n\sqrt{n}}$$

for $n \geq \frac{5\beta^2}{\sigma_{\min}^2}$.

4.1 A density computation

The goal of this subsection is to explicitly compute the density of $Z_{1-1/n} + Y$ and its marginals needed to apply Proposition 3.2. We will want to use the approximation

$$\frac{1}{2} \log \left(1 + \frac{1}{n^2 - 1} \right) \approx \frac{1}{2(n^2 - 1)}.$$

To this end, it is convenient to define

$$r(n) = \frac{1}{2(n^2 - 1)} - \frac{1}{2} \log \left(1 + \frac{1}{n^2 - 1} \right).$$

Note that since $t - 2t^2 \leq \log(1 + t) \leq t$ for any $t \geq 0$, we have for any $n \geq 2$ that

$$0 \leq r(n) \leq \frac{1}{(n^2 - 1)^2}.$$

The following lemma gives the formula for the density of $Z_{1-1/n} + Y$.

Lemma 4.1 *Let ρ be the density of Z_1 , let τ be the density of $Z_{1-1/n} + Y$, and let $f(x) = \frac{\tau(x)}{\rho(x)}$. Then,*

$$\mathbf{E}f(Z)^2 = \mathbf{E} \left[\exp \left(\sum_{i=1}^k \frac{2n^2 Y_i Y'_i - n Y_i^2 - n (Y'_i)^2 + \sigma_i^2}{2\sigma_i^2 (n^2 - 1)} - r(n) \right) \right],$$

where Y' is an independent copy of Y .

The proof is a straightforward calculation based on the following computational lemma, proved in ‘‘Proof of Lemma 4.2’’ of Appendix section.

Lemma 4.2 *Let Z be a k -dimensional Gaussian with covariance Σ . Define $\langle u, v \rangle_{\Sigma^{-1}} = \langle u, \Sigma^{-1}v \rangle$ and $\|u\|_{\Sigma^{-1}} = \sqrt{\langle u, u \rangle_{\Sigma^{-1}}}$. Then,*

$$\mathbf{E} \left[\exp \left(a \|Z\|_{\Sigma^{-1}}^2 + b \langle Z, v \rangle_{\Sigma^{-1}} \right) \right] = \exp \left(\frac{b^2}{2 - 4a} \|v\|_{\Sigma^{-1}}^2 \right) \cdot \left(\frac{1}{1 - 2a} \right)^{k/2}.$$

Proof of Lemma 4.1 In the notation of Lemma 4.2, the formula for ρ is

$$\rho(x) = \frac{1}{\sqrt{(2\pi)^k \cdot \det \Sigma}} \exp \left(-\frac{1}{2} \|x\|_{\Sigma^{-1}}^2 \right).$$

We write can τ in terms of ρ by

$$\begin{aligned} \tau(x) &= \mathbf{E} \left[\frac{1}{(1 - 1/n)^{k/2}} \cdot \rho \left(\frac{x - Y}{\sqrt{1 - 1/n}} \right) \right] \\ &= \mathbf{E} \left[\frac{1}{(1 - 1/n)^{k/2}} \exp \left(-\frac{1}{2 - 2/n} \|x - Y\|_{\Sigma^{-1}}^2 + \frac{1}{2} \|x\|_{\Sigma^{-1}}^2 \right) \rho(x) \right] \\ &= \mathbf{E} \left[\frac{1}{(1 - 1/n)^{k/2}} \exp \left(-\frac{\|x\|_{\Sigma^{-1}}^2}{2n - 2} + \frac{n \langle x, Y \rangle_{\Sigma^{-1}}}{n - 1} - \frac{n \|Y\|_{\Sigma^{-1}}^2}{2n - 2} \right) \right] \rho(x) \end{aligned}$$

Then, we have

$$f(x) = \frac{\tau(x)}{\rho(x)} = \mathbf{E} \left[\frac{1}{(1 - 1/n)^{k/2}} \exp \left(-\frac{\|x\|_{\Sigma^{-1}}^2}{2n - 2} + \frac{n \langle x, Y \rangle_{\Sigma^{-1}}}{n - 1} - \frac{n \|Y\|_{\Sigma^{-1}}^2}{2n - 2} \right) \right].$$

It follows that

$$\begin{aligned} \mathbf{E} f(Z)^2 &= (1 - 1/n)^{-k} \cdot \mathbf{E} \left[\exp \left(-\frac{\|Z\|_{\Sigma^{-1}}^2}{2n - 2} + \frac{n \langle Z, Y \rangle_{\Sigma^{-1}}}{n - 1} - \frac{n \|Y\|_{\Sigma^{-1}}^2}{2n - 2} \right) \right]^2 \\ &= (1 - 1/n)^{-k} \cdot \mathbf{E} \left[\exp \left(-\frac{\|Z\|_{\Sigma^{-1}}^2}{n - 1} + \frac{n \langle Z, Y + Y' \rangle_{\Sigma^{-1}}}{n - 1} \right) \right] \\ &\quad \cdot \mathbf{E} \left[\exp \left(-\frac{n (\|Y\|_{\Sigma^{-1}}^2 + \|Y'\|_{\Sigma^{-1}}^2)}{2n - 2} \right) \right], \end{aligned}$$

where we have used the fact that for any function α , $(\mathbf{E} \alpha(Y))^2 = \mathbf{E} (\alpha(Y)\alpha(Y'))$.

We apply Lemma 4.2 with $a = -\frac{1}{n-1}$, $b = \frac{n}{n-1}$, and $v = Y + Y'$. Note that $1 - 2a = 1 + \frac{2}{n-1} = \frac{n+1}{n-1}$. The above expression then becomes

$$\begin{aligned} \mathbf{E}f(Z)^2 &= (1 - 1/n)^{-k} \cdot \left(\frac{n-1}{n+1}\right)^{k/2} \mathbf{E} \left[\exp \left(\frac{n^2 \|Y + Y'\|_{\Sigma^{-1}}^2}{2(n^2 - 1)} \right) \right] \\ &\quad \cdot \mathbf{E} \left[\exp \left(-\frac{n(\|Y\|_{\Sigma^{-1}}^2 + \|Y'\|_{\Sigma^{-1}}^2)}{2(n-1)} \right) \right] \\ &= \left(\frac{n^2}{n^2 - 1}\right)^{k/2} \mathbf{E} \left[\exp \left(\frac{2n^2 \langle Y, Y' \rangle_{\Sigma^{-1}} - n\|Y\|_{\Sigma^{-1}}^2 - n\|Y'\|_{\Sigma^{-1}}^2}{2(n^2 - 1)} \right) \right] \\ &= \mathbf{E} \left[\exp \left(\frac{2n^2 \langle Y, Y' \rangle_{\Sigma^{-1}} - n\|Y\|_{\Sigma^{-1}}^2 - n\|Y'\|_{\Sigma^{-1}}^2}{2(n^2 - 1)} + \frac{k}{2} \log \left(1 + \frac{1}{n^2 - 1} \right) \right) \right] \\ &= \mathbf{E} \left[\exp \left(\sum_{i=1}^k \frac{2n^2 Y_i Y'_i - n Y_i^2 - n (Y'_i)^2}{2\sigma_i^2 (n^2 - 1)} + \frac{1}{2} \log \left(1 + \frac{1}{n^2 - 1} \right) \right) \right] \\ &= \mathbf{E} \left[\exp \left(\sum_{i=1}^k \frac{2n^2 Y_i Y'_i - n Y_i^2 - n (Y'_i)^2 + \sigma_i^2}{2\sigma_i^2 (n^2 - 1)} - r(n) \right) \right] \end{aligned}$$

□

Note that any projection of $Z_{1-1/n} + Y$ onto a subset of its coordinates still takes the form of a Gaussian plus an independent random vector. Therefore, Lemma 4.1 can also be applied to projections of Z and Y , leading to the following corollary.

Corollary 4.3 *Let Y_i denote the i -th coordinate of Y . For each $1 \leq i \leq k$, define*

$$Q_i = \frac{2n^2 Y_i Y'_i - n Y_i^2 - n (Y'_i)^2 + \sigma_i^2}{2\sigma_i^2 (n^2 - 1)} - r(n), \quad Q = \sum_{i=1}^k Q_i.$$

Then, for each i , we have

$$\mathbf{E}f_{(i)}(Z)^2 = \mathbf{E} \exp \left(\sum_{j \neq i} Q_j \right) = \mathbf{E} \exp(Q - Q_i),$$

where the notation $f_{(i)}$ follows that of Proposition 3.2.

Proof Let $P_{(i)} : \mathbb{R}^k \rightarrow \mathbb{R}^{k-1}$ denote the projection onto all but the i -th coordinate. Then, the result follows by replacing $Z_{1-1/n}$ and Y in Lemma 4.1 with $P_{(i)}(Z_{1-1/n})$ and $P_{(i)}(Y)$, respectively. □

4.2 Some computational estimates of the Q_i

Our strategy was to bound \mathcal{W}_2 distance via Proposition 3.2, which reduces the problem to estimating various densities. By Lemma 4.1 and Corollary 4.3, we have now

expressed the densities of interest in terms of the quantities Q_i , so the next step is to estimate the Q_i . In what follows, recall that we assumed $n \geq \frac{5\beta^2}{\sigma_i^2}$ for each i , and consequently, $n \geq 5k$ (see Remark 1.1). Also, recall that by assumption we have

$$\|Y\| \leq \frac{\beta}{\sqrt{n}}, \quad \mathbf{E}Y_i = 0, \quad \text{and} \quad \mathbf{E}Y_i^2 = \frac{\sigma_i^2}{n}.$$

The bounds we obtain are summarized in the next two lemmas.

Lemma 4.4 *We have*

$$|Q_i| \leq \frac{n^2|Y_i Y'_i|}{\sigma_i^2(n^2 - 1)} + \frac{1}{2n}, \quad |Q| \leq 1, \quad \text{and} \quad |Q - Q_i| \leq 1.$$

Proof To prove the first inequality, we have

$$\begin{aligned} |Q_i| &= \left| \frac{2n^2 Y_i Y'_i - n Y_i^2 - n (Y'_i)^2 + \sigma_i^2}{2\sigma_i^2(n^2 - 1)} - r(n) \right| \\ &\leq \frac{n^2 |Y_i Y'_i|}{\sigma_i^2(n^2 - 1)} + \frac{\beta^2}{\sigma_i^2(n^2 - 1)} + \frac{1}{n^2 - 1} + r(n) \\ &\leq \frac{n^2 |Y_i Y'_i|}{\sigma_i^2(n^2 - 1)} + \frac{n}{5(n^2 - 1)} + \frac{1}{n^2 - 1} + \frac{1}{(n^2 - 1)^2} \\ &\leq \frac{n^2 |Y_i Y'_i|}{\sigma_i^2(n^2 - 1)} + \frac{1}{2n}. \end{aligned}$$

Summing over all i , we obtain

$$\begin{aligned} |Q| &\leq \sum_{i=1}^k |Q_i| \leq \sum_{i=1}^k \left(\frac{n^2 |Y_i Y'_i|}{\sigma_i^2(n^2 - 1)} + \frac{1}{2n} \right) \leq \frac{n^2}{\sigma_{\min}^2(n^2 - 1)} \left(\sum_{i=1}^k |Y_i Y'_i| \right) + \frac{k}{2n} \\ &\leq \frac{n\beta^2}{\sigma_{\min}^2(n^2 - 1)} + \frac{k}{2n} \leq \frac{n^2}{5(n^2 - 1)} + \frac{1}{2} \leq 1, \end{aligned}$$

proving the second inequality. The third inequality follows by a similar argument, except that we omit one of the $|Q_i|$ terms in the sum. □

Lemma 4.5 *We have*

$$\mathbf{E}Q_i = -\frac{1}{2(n^2 - 1)} - r(n) \tag{4}$$

$$\mathbf{E}Q_i Q_j \leq \frac{n^2}{(n^2 - 1)^2} \delta_{ij} + \frac{n^2 \mathbf{E}Y_i^2 Y_j^2}{2\sigma_i^2 \sigma_j^2 (n^2 - 1)^2} + \frac{1}{2(n^2 - 1)^2} \tag{5}$$

$$\mathbf{E}Q_i^2 \leq \frac{2n^2 + n + 1}{2(n^2 - 1)^2} \tag{6}$$

$$\mathbf{E}(Q - Q_i)Q_i \leq \frac{nk}{2(n^2 - 1)^2} \tag{7}$$

$$\mathbf{E}Q^2 \leq \frac{2k}{n^2 - 1}. \tag{8}$$

Proof To show (4), we may compute

$$\mathbf{E}Q_i = \frac{-\sigma_i^2 - \sigma_i^2 + \sigma_i^2}{2\sigma_i^2(n^2 - 1)} - r(n) = -\frac{1}{2(n^2 - 1)} - r(n).$$

To show (5), we have

$$\begin{aligned} \mathbf{E}Q_i Q_j &\leq \mathbf{E} \left[\left(Q_i - \frac{1}{2(n^2 - 1)} + r(n) \right) \left(Q_j - \frac{1}{2(n^2 - 1)} + r(n) \right) \right] \\ &\leq \mathbf{E} \left[\left(\frac{2n^2 Y_i Y'_i - n Y_i^2 - n (Y'_i)^2}{2\sigma_i^2(n^2 - 1)} \right) \left(\frac{2n^2 Y_j Y'_j - n Y_j^2 - n (Y'_j)^2}{2\sigma_j^2(n^2 - 1)} \right) \right] \\ &= \frac{4n^4 \mathbf{E}Y_i Y_j Y'_i Y'_j + n^2 \mathbf{E} \left[(Y_i^2 + (Y'_i)^2) (Y_j^2 + (Y'_j)^2) \right]}{4\sigma_i^2 \sigma_j^2 (n^2 - 1)^2} \\ &= \frac{4n^2 \sigma_i^2 \sigma_j^2 \delta_{ij} + 2n^2 \mathbf{E}Y_i^2 Y_j^2 + 2\sigma_i^2 \sigma_j^2}{4\sigma_i^2 \sigma_j^2 (n^2 - 1)^2} \\ &= \frac{n^2}{(n^2 - 1)^2} \delta_{ij} + \frac{n^2 \mathbf{E}Y_i^2 Y_j^2}{2\sigma_i^2 \sigma_j^2 (n^2 - 1)^2} + \frac{1}{2(n^2 - 1)^2}. \end{aligned}$$

Finally, we can deduce (6), (7), and (8) from (5). Setting $i = j$ in (5) yields

$$\begin{aligned} \mathbf{E}Q_i^2 &\leq \frac{n^2}{(n^2 - 1)^2} + \frac{n^2 \mathbf{E}Y_i^4}{2\sigma_i^4 (n^2 - 1)^2} + \frac{1}{2(n^2 - 1)^2} \\ &\leq \frac{2n^2 + 1}{2(n^2 - 1)^2} + \frac{n\beta^2 \mathbf{E}Y_i^2}{2\sigma_i^4 (n^2 - 1)^2} \\ &= \frac{2n^2 + 1 + \beta^2/\sigma_i^2}{2(n^2 - 1)^2} \leq \frac{2n^2 + n + 1}{2(n^2 - 1)^2}, \end{aligned}$$

proving (6). If instead we sum (5) over all $j \neq i$, we obtain

$$\mathbf{E}(Q - Q_i)Q_i \leq \frac{n^2}{2\sigma_i^2(n^2 - 1)^2} \mathbf{E} \left(Y_i^2 \sum_{j \neq i} \frac{1}{\sigma_j^2} Y_j^2 \right) + \frac{k - 1}{2(n^2 - 1)^2}$$

$$\begin{aligned} &\leq \frac{n\beta^2}{2\sigma_i^2(n^2 - 1)^2} \mathbf{E} \left(\sum_{j \neq i} \frac{1}{\sigma_j^2} Y_j^2 \right) + \frac{k - 1}{2(n^2 - 1)^2} \\ &= \frac{(k - 1)(\beta^2/\sigma_i^2 + 1)}{2(n^2 - 1)^2} \leq \frac{nk}{2(n^2 - 1)^2}, \end{aligned}$$

proving (7). Finally, adding (6) and (7) and summing over all i , we obtain

$$\begin{aligned} \mathbf{E}Q^2 &= \sum_{i=1}^k \left(\mathbf{E}Q_i^2 + \mathbf{E}(Q - Q_i)Q_i \right) \leq k \cdot \frac{2n^2 + n + 1 + nk}{2(n^2 - 1)^2} \\ &\leq \frac{k(3n^2 + n + 1)}{2(n^2 - 1)^2} \leq \frac{k(4n^2 - 4)}{2(n^2 - 1)^2} = \frac{2k}{n^2 - 1}, \end{aligned}$$

which proves (8). □

4.3 Completing the proof

Proving Lemma 1.6 is now a matter of assembling together all of the bounds we have established.

Proof of Lemma 1.6 By Proposition 3.2 and Corollary 4.3, we have

$$\begin{aligned} \mathcal{W}_2(Z_1, Z_{1-1/n} + Y)^2 &\leq 2 \sum_{i=1}^k \sigma_i^2 \left(\mathbf{E}f(Z)^2 - \mathbf{E}f_{(i)}(Z)^2 \right) \\ &\leq 2 \sum_{i=1}^k \sigma_i^2 \mathbf{E} \left(e^Q - e^{Q-Q_i} \right). \end{aligned}$$

Thus, it remains to estimate the terms $\sigma_i^2 \mathbf{E} \left(e^Q - e^{Q-Q_i} \right)$. We do this by Taylor expanding the exponential. Define

$$R(t) = e^t - 1 - t - \frac{1}{2}t^2,$$

so that

$$e^t = 1 + t + \frac{1}{2}t^2 + R(t).$$

By Lemma 4.5, we can estimate the first and second order terms

$$\begin{aligned} \mathbf{E}(Q - (Q - Q_i)) &= \mathbf{E}Q_i \leq -\frac{1}{2(n^2 - 1)} \\ \frac{1}{2}\mathbf{E}(Q^2 - (Q - Q_i)^2) &= \mathbf{E} \left(\frac{1}{2}Q_i^2 + (Q - Q_i)Q_i \right) \leq \frac{n^2 + n/2 + 1/2}{2(n^2 - 1)^2} + \frac{nk}{2(n^2 - 1)^2} \end{aligned}$$

$$\leq \frac{(n^2 + nk - 1) + nk}{2(n^2 - 1)^2} = \frac{1}{2(n^2 - 1)} + \frac{nk}{(n^2 - 1)^2}.$$

To estimate the remainder term $R(Q) - R(Q - Q_i)$, note that for any $a, b \in [-1, 1]$,

$$\begin{aligned} |R(a) - R(b)| &= \left| (a - b) \sum_{m=3}^{\infty} \frac{1}{m!} (a^{m-1} + ab^{m-2} + \dots + b^{m-1}) \right| \\ &\leq |a - b| \cdot \sum_{m=3}^{\infty} \frac{1}{(m-1)!} \left(\frac{a^2 + b^2}{2} \right) \leq |a - b| \cdot \left(\frac{a^2 + b^2}{2} \right) \\ &\leq |a - b| \cdot \left(\frac{a^2 + b^2 + (2a - b)^2}{2} \right) = |a - b| \cdot \left(\frac{3}{2}a^2 + (a - b)^2 \right). \end{aligned}$$

In particular, by Lemma 4.4, both Q and $Q - Q_i$ are in $[-1, 1]$, so

$$\begin{aligned} \mathbf{E}[R(Q) - R(Q - Q_i)] &\leq \mathbf{E} \left[|Q_i| \left(\frac{3}{2}Q^2 + Q_i^2 \right) \right] \\ &\leq \frac{3}{2} \mathbf{E}|Q_i|Q^2 + \mathbf{E} \left[\left(\frac{n|Y_i Y_i'|^2}{\sigma_i^2(n^2 - 1)} + \frac{1}{2n} \right) Q_i^2 \right] \\ &\leq \frac{3}{2} \mathbf{E}|Q_i|Q^2 + \left(\frac{n\beta^2}{\sigma_i^2(n^2 - 1)} + \frac{1}{2n} \right) \cdot \frac{2n^2 + n + 1}{2(n^2 - 1)^2} \\ &\leq \frac{3}{2} \mathbf{E}|Q_i|Q^2 + \frac{2\beta^2}{\sigma_i^2 n^3} + \frac{1}{n^3}. \end{aligned}$$

Thus,

$$\begin{aligned} \sigma_i^2 \mathbf{E} \left(e^Q - e^{Q - Q_i} \right) &= \sigma_i^2 \left(\mathbf{E}(Q - (Q - Q_i)) + \frac{1}{2} \mathbf{E}(Q^2 - (Q - Q_i)^2) \right. \\ &\quad \left. + \mathbf{E}(R(Q) - R(Q - Q_i)) \right) \\ &\leq \sigma_i^2 \left(-\frac{1}{2(n^2 - 1)} + \frac{1}{2(n^2 - 1)} + \frac{nk}{(n^2 - 1)^2} \right. \\ &\quad \left. + \frac{3}{2} \mathbf{E}|Q_i|Q^2 + \frac{2\beta^2}{\sigma_i^2 n^3} + \frac{1}{n^3} \right) \\ &\leq \frac{3}{2} \mathbf{E}(\sigma_i^2 |Q_i|Q^2) + \frac{2\beta^2}{n^3} + \frac{nk\sigma_i^2}{(n^2 - 1)^2} + \frac{\sigma_i^2}{n^3} \\ &\leq \frac{3}{2} \mathbf{E}(\sigma_i^2 |Q_i|Q^2) + \frac{2\beta^2}{n^3} + \frac{3k\sigma_i^2}{n^3}. \end{aligned}$$

Summing over all i , we have

$$\begin{aligned} \sum_{i=1}^k \sigma_i^2 \mathbf{E} \left(e^Q - e^{Q-Q_i} \right) &\leq \frac{3}{2} \mathbf{E} \left(Q^2 \sum_{i=1}^k \sigma_i^2 |Q_i| \right) + \frac{2k\beta^2}{n^3} + \frac{3k}{n^3} \sum_{i=1}^k \sigma_i^2 \\ &\leq \frac{3}{2} \mathbf{E} \left(Q^2 \sum_{i=1}^k \sigma_i^2 |Q_i| \right) + \frac{5k\beta^2}{n^3} \\ &\leq \frac{3}{2} \mathbf{E} \left(Q^2 \sum_{i=1}^k \sigma_i^2 \left(\frac{n^2 |Y_i Y'_i|}{\sigma_i^2 (n^2 - 1)} + \frac{1}{2n} \right) \right) + \frac{5k\beta^2}{n^3} \\ &\leq \frac{3}{2} \mathbf{E} Q^2 \left(\frac{n\beta^2}{n^2 - 1} + \frac{\beta^2}{2n} \right) + \frac{5k\beta^2}{n^3} \\ &\leq \frac{3k}{n^2 - 1} \cdot \frac{2n\beta^2}{n^2 - 1} + \frac{5k\beta^2}{n^3} \leq \frac{25k\beta^2}{2n^3}, \end{aligned}$$

and so

$$\mathcal{W}_2 \left(Z_1, Z_{1-1/n} + Y \right)^2 \leq 2 \sum_{i=1}^k \sigma_i^2 \mathbf{E} \left(e^Q - e^{Q-Q_i} \right) \leq \frac{25k\beta^2}{n^3},$$

as desired. □

Acknowledgements We are indebted to Jian Ding for suggesting the use of Talagrand’s transportation inequality and Amir Dembo for pointing out a hole in a preliminary version of the main argument as well as many helpful comments on the exposition. We also thank Sourav Chatterjee for helpful discussions about related work. Finally, we thank the anonymous reviewers for many good suggestions and for pointing out several references.

Appendix

Proof of Proposition 1.2

Proof Let $\ell_n = \frac{\beta}{\sqrt{n}}$, and consider the lattice $L = \ell_n \mathbb{Z}^d$. For any $x \in \mathbb{R}^d$, let $d_L(x)$ denote the minimum Euclidean distance from x to L . Note that S_n takes values in L . Thus, letting ρ denote the density of Z , we have

$$\mathcal{W}_2(S_n, Z) \geq \int \rho(x) d_L(x) dx.$$

To estimate the right hand side, for any $y \in L$, let $Q_n(y)$ denote the cube of side length ℓ_n centered at y (which is also the set of points in \mathbb{R}^d closer to y than to any other point in L). We find that

$$\begin{aligned} \frac{1}{\text{Vol } Q_n(y)} \int_{Q_n(y)} d_L(x) dx &= \frac{\ell_n}{2^d} \int_{[-1,1]^d} \|x\| dx = \frac{\ell_n}{2^d} \int_{[-1,1]^d} \sqrt{x_1^2 + \dots + x_d^2} dx \\ &\geq \frac{\ell_n}{2^d} \int_{[-1,1]^d} \frac{1}{\sqrt{d}} (|x_1| + \dots + |x_d|) dx = \frac{1}{2} \ell_n \sqrt{d}. \end{aligned} \tag{9}$$

Next, let M be large enough so that,

$$\int_{[-M,M]^d} \rho(x) dx \geq \frac{1}{2},$$

and let

$$r_n = \inf_{\substack{x,y \in [-2M,2M]^d \\ \|x-y\| \leq \sqrt{d}\ell_n}} \frac{\rho(x)}{\rho(y)}. \tag{10}$$

Note that since ρ is positive and continuous, we have $\lim_{n \rightarrow \infty} r_n = 1$.

Assume now that n is sufficiently large so that $\ell_n < M$. Combining (10) with (9), we have for each $y \in L \cap [-M, M]^d$ that

$$\begin{aligned} \int_{Q_n(y)} \rho(x) d_L(x) dx &\geq \frac{r_n}{\text{Vol } Q_n(y)} \int_{Q_n(y)} \rho(x) dx \cdot \int_{Q_n(y)} d_L(x) dx \\ &\geq \frac{r_n \ell_n \sqrt{d}}{2} \int_{Q_n(y)} \rho(x) dx. \end{aligned}$$

Summing over all such y yields

$$\begin{aligned} \mathcal{W}_2(S_n, Z) &\geq \int \rho(x) d_L(x) dx \geq \int_{[-2M,2M]^d} \rho(x) d_L(x) dx \\ &\geq \sum_{y \in L \cap [-M,M]^d} \int_{Q_n(y)} \rho(x) d_L(x) dx \\ &\geq \frac{r_n \ell_n \sqrt{d}}{2} \int_{[-M,M]^d} \rho(x) dx \geq \frac{r_n \beta \sqrt{d}}{4\sqrt{n}}. \end{aligned}$$

Multiplying both sides by \sqrt{n} and taking limits gives the result. □

Proof of Proposition 1.4

Proof We prove the result with $C = 5$. Let $A \subset \mathbb{R}^d$ be a given convex set. For a parameter ϵ to be specified later, define

$$A^\epsilon = \left\{ x \in \mathbb{R}^d \mid \sup_{a \in A} \|x - a\| \leq \epsilon \right\}$$

$$A_\epsilon = \left\{ x \in \mathbb{R}^d \mid \inf_{a \in \mathbb{R}^d \setminus A} \|x - a\| \geq \epsilon \right\}.$$

Ball [1] showed a $4d^{1/4}$ upper bound⁶ for the Gaussian surface area of any convex set in \mathbb{R}^d . Hence,⁷

$$\mathbf{P}(Z \in A^\epsilon \setminus A) \leq 4\epsilon d^{1/4}, \quad \text{and} \quad \mathbf{P}(Z \in A \setminus A_\epsilon) \leq 4\epsilon d^{1/4}.$$

We may regard T as being coupled to Z so that $\mathbf{E}\|T - Z\|^2 = \mathcal{W}_2(T, Z)^2$. Then,

$$\begin{aligned} \mathbf{P}(T \in A) &\leq \mathbf{P}(\|T - Z\| \leq \epsilon, T \in A) + \mathbf{P}(\|T - Z\| > \epsilon) \\ &\leq \mathbf{P}(Z \in A^\epsilon) + \epsilon^{-2}\mathcal{W}_2(T, Z)^2 \\ &\leq \mathbf{P}(Z \in A) + 4\epsilon d^{1/4} + \epsilon^{-2}\mathcal{W}_2(T, Z)^2 \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{P}(Z \in A) &\leq \mathbf{P}(Z \in A_\epsilon) + 4\epsilon d^{1/4} \\ &\leq \mathbf{P}(\|T - Z\| \leq \epsilon, Z \in A_\epsilon) + \mathbf{P}(\|T - Z\| > \epsilon) + 4\epsilon d^{1/4} \\ &\leq \mathbf{P}(T \in A) + \epsilon^{-2}\mathcal{W}_2(T, Z)^2 + 4\epsilon d^{1/4}. \end{aligned}$$

Thus,

$$|\mathbf{P}(T \in A) - \mathbf{P}(Z \in A)| \leq \epsilon^{-2}\mathcal{W}_2(T, Z)^2 + 4\epsilon d^{1/4},$$

and taking $\epsilon = d^{-1/12}\mathcal{W}_2(T, Z)^{2/3}$ gives the result. □

Proof of Lemma 3.3

Proof Let A' and B' be independent copies of A and B . Then,

$$\mathbf{E}\left(f(A, B) + f(A', B') - f(A, B') - f(A', B)\right)^2 \geq 0.$$

Expanding yields

$$\begin{aligned} 4\mathbf{E}f(A, B)^2 + 4(\mathbf{E}f(A, B))^2 &= 4\mathbf{E}f(A, B)^2 + 2\mathbf{E}f(A, B)f(A', B') \\ &\quad + 2\mathbf{E}f(A, B')f(A', B) \\ &\geq 2\mathbf{E}f(A, B)f(A, B') + 2\mathbf{E}f(A, B)f(A', B) \\ &\quad + 2\mathbf{E}f(A', B')f(A, B') + 2\mathbf{E}f(A', B')f(A', B) \end{aligned}$$

⁶ The constant was later improved to $(2\pi)^{-1/4} \approx 0.64$ by Nazarov [18], who also constructed an example with surface area of order $d^{1/4}$.

⁷ This is also given as equation (1.4) in [2].

$$\begin{aligned}
 &= 2\mathbf{E}f_B(A)^2 + 2\mathbf{E}f_A(B)^2 \\
 &\quad + 2\mathbf{E}f_A(B)^2 + 2\mathbf{E}f_B(A)^2 \\
 &= 4\mathbf{E}f_B(A)^2 + 4\mathbf{E}f_A(B)^2,
 \end{aligned}$$

as desired. □

Proof of Equation (2)

Proof We proceed by induction on the dimension d , retracing the argument of [23], section 3. The base case $d = 1$ is immediate from Theorem 3.1.

Assume now that the inequality holds in $d - 1$ dimensions. For the inductive step, we can follow the same argument used to prove Theorem 3.1 (see [23], section 3). The argument proceeds by first comparing Y to another \mathbb{R}^d -valued random variable \hat{Y} sharing the first $d - 1$ coordinates of Y , but whose last coordinate is independently drawn from $\mathcal{N}(0, \sigma_d)$.

Fix a $(d - 1)$ -dimensional vector \hat{x} , and let $T_{\hat{x}}$ denote a random variable distributed as the last coordinate of Y conditioned on the first $d - 1$ coordinates being equal to \hat{x} . Let $\hat{\rho}(\hat{x}) = \int_{-\infty}^{\infty} \rho(\hat{x}, t) dt$. Then, the density of $T_{\hat{x}}$ at t is given by

$$\frac{f(\hat{x}, t) \cdot \rho(\hat{x}, t)}{f_{(d)}(\hat{x}, 0) \cdot \hat{\rho}(\hat{x})}.$$

Noting that $\frac{\rho(\hat{x}, t)}{\hat{\rho}(\hat{x})}$ is the density of $\mathcal{N}(0, \sigma_d)$ at t , the one-dimensional case of Theorem 3.1 implies

$$\mathcal{W}_2(T_{\hat{x}}, \mathcal{N}(0, \sigma_d))^2 \leq 2\sigma_d^2 \int_{-\infty}^{\infty} \frac{f(\hat{x}, t)}{f_{(d)}(\hat{x}, t)} \log \left(\frac{f(\hat{x}, t)}{f_{(d)}(\hat{x}, t)} \right) \frac{\rho(\hat{x}, t)}{\hat{\rho}(\hat{x})} dt. \tag{11}$$

Since $T_{\hat{x}}$ and $\mathcal{N}(0, \sigma_d)$ have the same distributions as Y and \hat{Y} conditioned on \hat{x} , we may integrate (11) over \hat{x} to obtain

$$\begin{aligned}
 \mathcal{W}_2(Y, \hat{Y})^2 &\leq 2 \int_{\mathbb{R}^{d-1}} \mathcal{W}_2(T_{\hat{x}}, \mathcal{N}(0, \sigma_d))^2 \cdot f_{(d)}(\hat{x}, 0) \hat{\rho}(\hat{x}) d\hat{x} \\
 &\leq 2\sigma_d^2 \int_{\mathbb{R}^{d-1}} \int_{-\infty}^{\infty} f(\hat{x}, t) \log \left(\frac{f(\hat{x}, t)}{f_{(d)}(\hat{x}, t)} \right) \rho(\hat{x}, t) dt d\hat{x}. \\
 &= 2\sigma_d^2 \cdot \mathbf{E} \left(f(Z) \log \frac{f(Z)}{f_{(d)}(Z)} \right) \\
 &= 2\sigma_d^2 \cdot \left(\mathbf{E} (f(Z) \log f(Z)) - \mathbf{E} (f_{(d)}(Z) \log f_{(d)}(Z)) \right)
 \end{aligned}$$

Next, define $Y_{(d)}$ and $Z_{(d)}$ to be the projections onto the first $d - 1$ coordinates of Y and Z , respectively. Note that the coupling of Y to \hat{Y} changes only d -th coordinate. Furthermore, the d -th coordinates of \hat{Y} and Z are both distributed as $\mathcal{N}(0, \sigma_d)$

independent of the first $d - 1$ coordinates. Thus, a coupling of $Y_{(d)}$ to $Z_{(d)}$ induces a coupling of \hat{Y} to Z in which the last coordinate does not change. Consequently,

$$\mathcal{W}_2(Y, Z)^2 \leq 2\sigma_d^2 \cdot \left(\mathbf{E} (f(Z) \log f(Z)) - \mathbf{E} (f_{(d)}(Z) \log f_{(d)}(Z)) \right) + \mathcal{W}_2(Y_{(d)}, Z_{(d)})^2. \tag{12}$$

Now, recall that the density of $Y_{(d)}$ at a point $\hat{x} \in \mathbb{R}^{d-1}$ is $f_{(d)}(\hat{x}, 0) \cdot \hat{\rho}(\hat{x})$, and so applying the inductive hypothesis to $\mathcal{W}_2(Y_{(d)}, Z_{(d)})^2$ yields

$$\begin{aligned} \mathcal{W}_2(Y_{(d)}, Z_{(d)})^2 &\leq 2 \sum_{k=1}^{d-1} \sigma_k^2 \cdot \mathbf{E} (f_{[k]}(Z_{(d)}) \log f_{[k]}(Z_{(d)}) - f_{[k-1]}(Z_{(d)}) \log f_{[k-1]}(Z_{(d)})) \\ &= 2 \sum_{k=1}^{d-1} \sigma_k^2 \cdot \mathbf{E} (f_{[k]}(Z) \log f_{[k]}(Z) - f_{[k-1]}(Z) \log f_{[k-1]}(Z)). \end{aligned}$$

Substituting into (12), we obtain

$$\mathcal{W}_2(Y, Z)^2 \leq 2 \sum_{k=1}^{d-1} \sigma_k^2 \cdot \mathbf{E} (f_{[k]}(Z) \log f_{[k]}(Z) - f_{[k-1]}(Z) \log f_{[k-1]}(Z)),$$

completing the induction. □

Proof of Lemma 4.2

Proof Let $C_k = (2\pi)^{-\frac{k}{2}}$. We have

$$\begin{aligned} &\mathbf{E} \left[\exp \left(a \|Z\|_{\Sigma^{-1}}^2 + b \langle Z, v \rangle_{\Sigma^{-1}} \right) \right] \\ &= \frac{C_k}{\sqrt{\det \Sigma}} \int_{\mathbb{R}^k} \exp \left(- \left(\frac{1}{2} - a \right) \|x\|_{\Sigma^{-1}}^2 + b \langle x, v \rangle_{\Sigma^{-1}} \right) dx \\ &= \frac{C_k}{\sqrt{\det \Sigma}} \int_{\mathbb{R}^k} \exp \left(- \left(\frac{1}{2} - a \right) \left\| x - \left(\frac{b}{1-2a} \right) v \right\|_{\Sigma^{-1}}^2 + \frac{b^2}{2-4a} \|v\|_{\Sigma^{-1}}^2 \right) dx \\ &= \exp \left(\frac{b^2}{2-4a} \|v\|_{\Sigma^{-1}}^2 \right) \cdot \frac{C_k}{\sqrt{\det \Sigma}} \int_{\mathbb{R}^k} \exp \left(- \frac{1-2a}{2} \left\| x - \left(\frac{b}{1-2a} \right) v \right\|_{\Sigma^{-1}}^2 \right) dx \\ &= \exp \left(\frac{b^2}{2-4a} \|v\|_{\Sigma^{-1}}^2 \right) \cdot \frac{C_k}{\sqrt{\det \Sigma}} \int_{\mathbb{R}^k} \exp \left(- \frac{1-2a}{2} \|x\|_{\Sigma^{-1}}^2 \right) dx \\ &= \exp \left(\frac{b^2}{2-4a} \|v\|_{\Sigma^{-1}}^2 \right) \cdot \frac{C_k}{\sqrt{\det \Sigma}} \int_{\mathbb{R}^k} \exp \left(- \frac{1}{2} \|y\|_{\Sigma^{-1}}^2 \right) \frac{dy}{\sqrt{(1-2a)^k}} \quad (y := \sqrt{1-2a} \cdot x) \end{aligned}$$

$$\begin{aligned}
&= \exp\left(\frac{b^2}{2-4a}\|v\|_{\Sigma^{-1}}^2\right) \cdot \left(\frac{1}{1-2a}\right)^{k/2} \cdot \frac{C_k}{\sqrt{\det \Sigma}} \int_{\mathbb{R}^k} \exp\left(-\frac{1}{2}\|y\|_{\Sigma^{-1}}^2\right) dy \\
&= \exp\left(\frac{b^2}{2-4a}\|v\|_{\Sigma^{-1}}^2\right) \cdot \left(\frac{1}{1-2a}\right)^{k/2}.
\end{aligned}$$

□

References

- Ball, K.: The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10**(4), 411–420 (1993)
- Bentkus, V.: On the dependence of the Berry–Esseen bound on dimension. *J. Stat. Plan. Inference* **113**(2), 385–402 (2003)
- Bentkus, V., Götze, F.: Optimal rates of convergence in the CLT for quadratic forms. *Ann. Probab.* **24**(1), 466–490 (1996)
- Bergström, H.: On the central limit theorem in the space R_k , $k > 1$. *Scand. Actuar. J.* **1945**(1–2), 106–127 (1945)
- Bhattacharya, R.N.: Refinements of the multidimensional central limit theorem and applications. *Ann. Probab.* **5**(1), 1–27 (1977)
- Bhattacharya, R., Holmes, S.: An exposition of Götze’s estimation of the rate of convergence in the multivariate central limit theorem. Preprint [arXiv:1003.4254](https://arxiv.org/abs/1003.4254) (2010)
- Bobkov, S.G.: Entropic approach to E. Rio’s central limit theorem for \mathcal{W}_2 transport distance. *Stat. Prob. Lett.* **83**(7), 1644–1648 (2013)
- Bobkov, S.G., Chistyakov, G., Götze, F.: Berry–Esseen bounds in the entropic central limit theorem. *Probab. Theory Relat. Fields* **159**(3–4), 435–478 (2014)
- Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163**(1), 1–28 (1999)
- Bonis, T.: Rates in the central limit theorem and diffusion approximation via Stein’s method. Preprint [arXiv:1506.06966](https://arxiv.org/abs/1506.06966) (2015)
- Bubeck, S., Ding, J., Eldan, R., Rácz, M.: Testing for high-dimensional geometry in random graphs. *Random Struct. Algorithms* **49**(3), 503–532 (2016)
- Bubeck, S., Ganguly, S.: Entropic CLT and phase transition in high-dimensional Wishart matrices. Preprint [arXiv:1509.03258](https://arxiv.org/abs/1509.03258) (2015)
- Chen, L.H.Y., Fang, X.: Multivariate normal approximation by Stein’s method: the concentration inequality approach. Preprint [arXiv:1111.4073](https://arxiv.org/abs/1111.4073) (2011)
- Chernozhukov, V., Chetverikov, D., Kato, K.: Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Stat.* **41**(6), 2786–2819 (2013)
- Götze, F.: On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**(2), 724–739 (1991)
- Marton, K.: Bounding \bar{d} -distance by informational divergence: a method to prove measure concentration. *Ann. Probab.* **24**(2), 857–866 (1996)
- Nagaev, S.V.: An estimate of the remainder term in the multidimensional CLT. In: *Proceedings of the third Japan-USSR Symposium on Probability Theory*, pp. 419–438. Springer, Berlin (1976)
- Nazarov, F.: On the maximal perimeter of a convex set in \mathbb{R}^n with respect to a Gaussian measure. In: *Geometric Aspects of Functional Analysis*, pp. 169–187. Springer, Berlin (2003)
- Rio, E.: Upper bounds for minimal distances in the central limit theorem. *Annales de l’IHP Probabilités et Statistiques* **45**(3), 802–817 (2009)
- Rio, E.: Asymptotic constants for minimal distance in the central limit theorem. *Electron. Commun. Probab.* **16**, 96–103 (2011)
- Sazanov, V.V.: On the multi-dimensional central limit theorem. *Sankhyā Indian J. Stat. Ser. A* **30**(2), 181–204 (1968)
- Senatov, V.V.: Uniform estimates of the rate of convergence in the multi-dimensional central limit theorem. *Theory Probab. Appl.* **24**(4), 745–759 (1980)
- Talagrand, M.: Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6**(3), 587–600 (1996)

24. Valiant, G., Valiant, P.: Estimating the unseen: an $n/\log(n)$ -sample estimator for entropy and support size, shown optimal via new CLTs. In: Proceedings of the Forty-Third Annual ACM Symposium on the Theory of Computing, pp. 685–694 (2011)
25. Valiant, G., Valiant, P.: A CLT and tight lower bounds for estimating entropy. <http://www.eccc.uni-trier.de/report/2010/179/> (2010)
26. Vershik, A.M.: Long history of the Monge–Kantorovich transportation problem. *Math. Intell.* **35**(4), 1–9 (2013)