Electronic Journal of Probability

Electron. J. Probab. **29** (2024), article no. 78, 1–18.
ISSN: 1083-6489 https://doi.org/10.1214/24-EJP1134

Improved rates of convergence for the multivariate Central Limit Theorem in Wasserstein distance

Thomas Bonis*

Abstract

We provide new bounds for the rate of convergence of the multivariate Central Limit Theorem in Wasserstein distances of order $p \geq 2$. In particular, we obtain what we conjecture to be the asymptotically optimal rate in the identically distributed case whenever the measure of the summands admits a non-zero continuous component and has a non-zero third moment.

Keywords: Central Limit Theorem; multivariate normal approximation; p-Wasserstein distance; Stein's method.
MSC2020 subject classifications: 60F05; 62E17.
Submitted to EJP on October 5, 2023, final version accepted on April 24, 2024.

1 Introduction and main result

Let $X_1, \dots, X_n$ be i.i.d. random variables drawn from a measure μ on $\mathbb{R}^d$ and such that $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1 X_1^T] = I_d$. By the Central Limit Theorem, we know that the measure μ_n of $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ converges to the d-dimensional standard normal distribution γ. In this work, we wish to quantify this convergence for the family of Wasserstein distances of order $p \geq 2$, defined between any two measures ν and ν' on $\mathbb{R}^d$ by

$$W_p(\nu, \nu')^p = \inf_\pi \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y - x\|^p \, d\pi(x, y),$$

where π has marginals ν and ν' and $\| \cdot \|$ is the traditional Euclidean norm.

In recent years, multiple works provided non-asymptotic bounds for $W_p(\mu_n, \gamma)$. For instance as long as $\mathbb{E}[\|X_1\|^4] < \infty$, Theorem 1 [2] gives

$$W_2(\mu_n, \gamma) \leq C \sqrt{\frac{\sqrt{d} \|\mathbb{E}[X_1 X_1^T \|X_1\|^2]\|_{HS}}{n}}, \tag{1.1}$$

*Université Gustave Eiffel, Université Paris-Est Créteil, CNRS, LAMA UMR8050, F-77447 Marne-La-Vallée, France. E-mail: thomas.bonis@univ-eiffel.fr

where $C > 0$ and $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm. Similar results were also obtained for other W_p distances [2, 5]. However, this bound is not sharp with respect to the dimension. Indeed, if X_1 has i.i.d. components, (1.1) scales with $d^{3/4}$ while an optimal bound would scale with $\sqrt{d}$. Sharper bounds have been obtained under additional assumptions on the measure μ. For instance, if μ satisfies a Poincaré inequality with constant $K \geq 1$, Theorem 4.1 [3] gives

$$W_2(\mu_n, \gamma) \leq C\sqrt{\frac{(K-1)d}{n}} \tag{1.2}$$

and similar results have been obtained for any W_p distances with $p \geq 1$ in Theorem 1.2 [6] under the additional assumption that μ is log-concave. As a consequence of (1.2), if μ is log-concave then it admits a Poincaré constant $K \leq C\sqrt{\log d}$ for some $C > 0$ [8] and if the Kannan-Lovász-Simonovits isoperimetric conjecture is true then $K \leq C$. Finally, for uniformly log-concave measures, the optimal dependency on $\sqrt{d}$ is obtained in Theorem 3.4 [7] without any further assumptions.

Some insight on the conditions required to obtain this optimal dependency on the dimension in a more general case can be obtained from Proposition 1.2 [13] which states that, if X_1 takes value in the lattice $h\mathbb{Z}^d$ with $h > 0$, then

$$\liminf_{n\to\infty} \sqrt{n}W_2(\mu_n, \gamma) \geq \frac{\sqrt{d}h}{4}.$$

In particular, if h is of order $\sqrt{d}$ then $\liminf_{n\to\infty} \sqrt{n}W_2(\mu_n, \gamma) \geq Cd$. Therefore, if one wants $W_2(\mu_n, \gamma)$ to scale with the square root of the dimension, one would require h to be independent of d, or X_1 to not be lattice-distributed. Such a result does not come as surprising in the light of known asymptotic results obtained in the univariate setting. Indeed, according to Theorem 1.2 [12], if X_1 takes values in $\{a + kh \mid k \in \mathbb{Z}\}$ for some $a \in \mathbb{R}, h > 0$ and has a finite moment of order $p + 2$ with $p \in]1, 2]$, then

$$\liminf_{n\to\infty} \sqrt{n}W_p(\mu_n, \gamma) = \frac{1}{6}\|\mathbb{E}[X_1^3](Z^2 - 1) + hU\|_p, \tag{1.3}$$

where $Z \sim \gamma, U$ is a uniform random variable on $[-1/2, 1/2]$ independent of Z and $\|\cdot\|_p = \mathbb{E}[\|\cdot\|^p]^{1/p}$. On the other hand, as long as X_1 is not lattice-distributed, one has

$$\liminf_{n\to\infty} \sqrt{n}W_p(\mu_n, \gamma) = \frac{1}{6}\|\mathbb{E}[X_1^3](Z^2 - 1)\|_p. \tag{1.4}$$

Furthermore, faster rates of convergence have been obtained for all $p \geq 1$ whenever the first moments of μ and γ are equal and μ satisfies the Cramer's condition [1].

One can thus expect the rate of convergence for the central limit theorem in Wasserstein distance in a high-dimensional setting to not only be determined by the moments of X_1 but to also depend on whether the measure is lattice-distributed. In other words, along with the large-scale behaviour of μ, described by its moments, we expect a tight bound on $W_p(\mu_n, \gamma)$ to include a term corresponding to the small-scale behaviour of μ. In this work, we provide a first instance of such a result in the multidimensional setting. In particular, we obtain the following asymptotic bound.

Corollary 1.1. *Let $p \geq 2$ and $X_1, \ldots, X_n$ be i.i.d. centered random variables drawn from a measure μ on $\mathbb{R}^d$ with identity covariance matrix and finite moment of order $p + 2$. Suppose there exists $h > 0$ such that the matrix*

$$\mathbb{E}[(X_2 - X_1)(X_2 - X_1)^T 1_{\|X_2 - X_1\| \leq h}]$$

is positive-definite. Then, the measure μ_n of $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ verifies

$$\sqrt{n}W_p(\mu_n,\gamma) \leq \frac{1}{6}\|\mathbb{E}[X_1^{\otimes 3}](Z^{\otimes 2} - I_d)\|_p + Cph\sqrt{d} + C_{d,p,\mu}\mathcal{O}\left(\log(n)^{-1/p}\right),$$

where $C > 0$ is a generic constant, $C_{d,p,\mu}$ is a constant depending on d, p and μ, Z is drawn from the d-dimensional standard normal distribution γ and $\mathbb{E}[X_1^{\otimes 3}](Z^{\otimes 2} - I_d)$ is a vector whose i-th coordinate is given by

$$(\mathbb{E}[X_1^{\otimes 3}](Z^{\otimes 2} - I_d))_i \coloneqq \sum_{j,k}\mathbb{E}[(X_1)_i(X_1)_j(X_1)_k](Z_jZ_k - 1_{j=k}).$$

Furthermore, if μ has a non-zero absolutely continuous component with respect to the Lebesgues measure then,

$$\sqrt{n}W_p(\mu_n,\gamma) \leq \frac{1}{6}\|\mathbb{E}[X_1^{\otimes 3}](Z^{\otimes 2} - I_d)\|_p + C_{d,p,\mu}o\left(1\right). \tag{1.5}$$

Whenever μ admits a non-zero continuous component and has a non-zero third moment, we conjecture our result to be asymptotically optimal as it is a natural multidimensional generalization of (1.4). In particular, if X_1 has i.i.d. components we recover the correct dependency on $\sqrt{d}$ since, by Lemma 6.2,

$$\sqrt{n}W_p(\mu_n,\gamma) \leq \frac{\mathbb{E}[(X_1)_1^3]\sqrt{(p-1)d}}{3\sqrt{2}} + C_{d,p,\mu}o\left(1\right).$$

Our bound is also asymptotically sharper than known existing bounds. Indeed, using Lemma 6.2 and Lemma 6.5, we obtain

$$\|\mathbb{E}[X_1^{\otimes 3}](Z^{\otimes 2} - I_d)\|_p \leq \sqrt{2(p-1)\sqrt{d}\|\mathbb{E}[X_1X_1^T\|X_1\|^2]\|_{HS}},$$

thus recovering (1.1) in the asymptotic setting. In particular, this means that if $\|X_1\| \leq M$ almost surely then, for any $p \geq 1$,

$$\sqrt{n}W_p(\mu_n,\gamma) \leq M\sqrt{\frac{(p-1)d}{18}} + C_{d,p,\mu}o\left(1\right).$$

Remark that this bound scales with at least d as M must be of order at least $\sqrt{d}$. On the other hand, if μ admits a Stein kernel τ as defined in [9], combining Lemmas 6.2 and 6.6 gives

$$\|\mathbb{E}[X_1^{\otimes 3}](Z^{\otimes 2} - I_d)\|_p \leq 2\sqrt{2(p-1)\mathbb{E}[\|\tau - I_d\|^2]}.$$

Hence, following the work of [3], if μ admits a Poincaré constant $K \geq 1$ we can generalize (1.2) to all $p \geq 1$:

$$\sqrt{n}W_p(\mu_n,\gamma) \leq \frac{\sqrt{2(p-1)(K-1)d}}{3} + C_{d,p,\mu}o\left(1\right).$$

Let us note that, asymptotically, this bound depends only on $\sqrt{p-1}$, thus improving on the bound obtained in Theorem 1.2 [6] which scales with p^2 while lifting the requirement for μ to be log-concave.

For lattice-distributed measures, our bound is close to matching a multidimensional equivalent of (1.3) but still requires improvements. However, obtaining the optimal rate of convergence for discrete but non lattice-distributed random variables is still an open issue. In any case, let us note that the remainder term is likely sub-optimal.

Corollary 1.1 is derived from a non-asymptotic bound obtained in Theorem 3.1 which also deals with non-identically distributed random variables. Our result is derived through refinements on a variant of Stein's method used in [2] which might be of interest in other contexts.

2 Notations

Let d be a positive integer. For any $k \in \mathbb{N}$, let $(\mathbb{R}^d)^{\otimes k}$ be the set of elements of the form $(x_j)_{j \in \{1,\dots,d\}^k} \in \mathbb{R}^{d^k}$. For $x \in \mathbb{R}^d$ and $k \in \mathbb{N}$, we denote by $x^{\otimes k}$ the element of $(\mathbb{R}^d)^{\otimes k}$ such that

$$\forall j \in \{1,\dots,d\}^k, (x^{\otimes k})_j = \prod_{i=1}^k x_{j_i}.$$

For any $x, y \in (\mathbb{R}^d)^{\otimes k}$, we denote by $< x, y >$ the Hilbert-Schmidt scalar product between x and y defined by

$$< x, y >= \sum_{i \in \{1,\dots,d\}^k} x_i y_i,$$

and, by extension, we write

$$\|x\|^2 =< x, x > .$$

Furthermore, for any $x \in (\mathbb{R}^d)^{\otimes (k+1)}$ and $y \in (\mathbb{R}^d)^{\otimes k}$, let xy be the vector defined by

$$\forall i \in \{1,\dots,d\}, (xy)_i = \sum_{j \in \{1,\dots,d\}^k} x_{i,j} y_j.$$

For any $k \in \mathbb{N}$, any function ϕ with partial derivatives of order k and any $x \in \mathbb{R}^d$, we denote by $\nabla^k \phi(x) \in (\mathbb{R}^d)^{\otimes k}$ the k-th gradient of ϕ at x:

$$\forall j \in \{1,\dots,d\}^k, (\nabla^k \phi(x))_j = \frac{\partial^k \phi}{\partial x_{j_1} \dots \partial x_{j_k}}(x).$$

For any $k \in \mathbb{N}$, let H_k be the d-dimensional Hermite polynomial, defined by

$$\forall x \in \mathbb{R}^d, H_k(x) = (-1)^k e^{\frac{\|x\|^2}{2}} \nabla^k e^{-\frac{\|x\|^2}{2}}.$$

Finally, for any random variable X on $\mathbb{R}^d$, we denote by $\|X\|_p$ the L_p-norm of X, that is

$$\|X\|_p \coloneqq \mathbb{E}[\|X\|^p]^{1/p}.$$

3 Main result

Let $n > 0$ and $W_1, \dots, W_n$ be independent centered random variables on $\mathbb{R}^d$ such that $W = \sum_{i=1}^n W_i$ has identity covariance matrix and $\max_{i \in \{1,\dots,n\}} \|\mathbb{E}[W_i^{\otimes 2}]\| < 1$. We denote by ν the measure of W. For any $i \in \{1,\dots,n\}$, let $D_i = W_i' - W_i$, where W_i' is an independent copy of W_i. Let us define a set of features describing the large-scale behaviour of the variables $(W_i)_{1 \le i \le n}$:

- $\forall i \in \{1,\dots,n\}, \xi_i = -\log(\|\mathbb{E}[W_i^{\otimes 2}]\|)$;
- $\forall q > 0, L_q = \sum_{i=1}^n \mathbb{E}[\|W_i\|^q]$;
- $\forall q > 2, N_q = \sum_{i=1}^n \frac{1}{\xi_i} \mathbb{E}[\|D_i\|^q (1_{\|D_i\|^2 \ge \xi_i \|\mathbb{E}[W_i^{\otimes 2}]\|^{2/3}} + \xi_i^{-1})^{q/2-1}]$;
- $N_4' = \sum_{i=1}^n \frac{\|\mathbb{E}[D_i^{\otimes 2} \|W_i\| \|D_i\|]\|}{\sqrt{\|\mathbb{E}[W_i^{\otimes 2}]\| \xi_i}}$.

Now, for any $\beta > 0$, let $D_{i,\beta} = D_i 1_{\|D_i\| \le \beta}$. If $\mathbb{E}\left[\sum_{i=1}^n D_{i,\beta}^{\otimes 2}\right]$ is positive-definite, we consider the following small-scale feature:

$$\forall q \ge 0, \beta_q = \sum_{i=1}^n \mathbb{E}\left[\left\| \mathbb{E}\left[\sum_{i=1}^n D_{i,\beta}^{\otimes 2}\right]^{-1} D_{i,\beta} \right\|^q\right].$$

Theorem 3.1. *Let $p \geq 2$ such that $p \leq \min_{i\in\{1,\dots,n\}} \|\mathbb{E}[W_i^{\otimes 2}]\|^{-1}$ and suppose $L_{p+2} < \infty$. Let $\epsilon \coloneqq \max_{i\in\{1,\dots,n\}} \|\mathbb{E}[W_i^{\otimes 2}]\|^{2/3}$. If there exists $0 < \beta < \sqrt{\epsilon}$ such that $\mathbb{E}\left[\sum_{i=1}^n D_{i,\beta}^{\otimes 2}\right]$ is positive-definite, then, for any $q, r > p$ such that $\frac{1}{q} + \frac{1}{r} = \frac{1}{p}$, we have*

$$W_p(\nu,\gamma) \leq \frac{\|\mathbb{E}[W^{\otimes 3}]H_2(Z)\|_p}{6} + C\left(\beta p\sqrt{d} + \frac{(2r-1)^{3/2}\|\mathbb{E}[W^{\otimes 3}]\|W_q(\nu,\gamma)}{\sqrt{(p-1)\epsilon}}\right)$$
$$+ Cp\left(\epsilon\left(\sqrt{p}(\sqrt{\beta_2}+\sqrt{d}) + p\left(\beta_p^{1/p} + L_p^{1/p}\right)\right) + \sqrt{N_4} + (pN_{p+2})^{1/p} + N_4'\right).$$

In order to prove Corollary 1.1 from this result, we take

$$\forall i \in \{1,\dots,n\}, W_i = \frac{X_i}{\sqrt{n}}$$

and $\beta = \frac{h}{\sqrt{n}}$ and let us assume n is sufficiently large so that $\beta < \frac{\|\mathbb{E}[X_1^{\otimes 2}]\|^{2/3}}{n^{2/3}} = \frac{d^{1/3}}{n^{2/3}}$. In the following, we denote by C a positive constant depending on properties of μ but independent of n. First, we have

$$\sqrt{p}(\sqrt{\beta_2}+\sqrt{d}) + p\left(\beta_p^{1/p} + L_p^{1/p}\right) \leq C$$

and, since $\epsilon = \frac{C}{n^{2/3}}$,

$$\epsilon\left(\sqrt{p}(\sqrt{\beta_2}+\sqrt{d}) + p\left(\beta_p^{1/p} + L_p^{1/p}\right)\right) \leq \frac{C}{n^{2/3}}.$$

Then,

$$N_4' = \frac{C}{\sqrt{n\log(n)}}$$

and, since we have $\lim_{n\to\infty} n\epsilon\xi_1 = +\infty$,

$$p\sqrt{N_4} + p^{1+1/p}N_{p+2}^{1/p} = o\left(\frac{1}{\sqrt{n}\log(n)^{1/p}}\right).$$

Furthermore, since X_1 has finite moment of order $p+2$, we can use Theorem 6 from [2] to obtain

$$W_{p+1/2}(\mu_n,\gamma) \leq \frac{C}{n^{1/2-1/4p}} \leq \frac{C}{n^{3/8}}.$$

Thus, since $\|\mathbb{E}[W^{\otimes 3}]\| = \frac{C}{\sqrt{n}}$,

$$\frac{\|\mathbb{E}[W^{\otimes 3}]\|W_{p+1/2}(\mu_n,\gamma)}{\sqrt{\epsilon}} \leq \frac{C}{n^{13/24}}.$$

which concludes the proof whenever μ does not admit an absolutely continuous component with respect to the Lebesgues measure. If it does, let us denote by μ_c this continuous component. For any $h > 0$, there must exist a ball $\mathcal{B}$ with radius h and non-zero mass for μ_c. Remark that

$$\int_{\mathcal{B}^2} (x'-x)^{\otimes 2} d\mu_c(x) d\mu_c(x')$$

must be positive definite. Otherwise the dimension of the support of μ_c on this ball would be lower than d which is impossible since μ_c is absolutely continuous with respect to the Lebesgues measure. Thus,

$$\forall h, q > 0, \beta_q(h) = \frac{1}{n^{q/2-1}}\mathbb{E}\left[\left\|\mathbb{E}\left[(X_i'-X_i)^{\otimes 2}1_{\|X_i'-X_i\|\leq h}\right]^{-1}(X_i'-X_i)1_{\|X_i'-X_i\|\leq h}\right\|^q\right]$$

must be finite. Therefore, for n sufficiently large, we can take h_n as

$$h_n = \inf\left\{h \geq \frac{1}{n} : \sqrt{\beta_2(h)} + \beta_p(h)^{1/p} \leq n^{1/7}\right\}.$$

Then $\lim_{n\to\infty} h_n = 0$ and

$$\epsilon\left(\sqrt{p}(\sqrt{\beta_2(h_n)} + \sqrt{d}) + p\left(\beta_p(h_n)^{1/p} + L_p^{1/p}\right)\right) = O\left(n^{1/7-2/3}\right) = o\left(\frac{1}{\sqrt{n}}\right),$$

which yields the desired result.

Remark 3.2. Note that we restricted ourselves to the existence of a moment of order $p+2$ for the summands to simplify computations. Let us note that one could only consider existence of a moment of order $p+l$ with $l<2$ only in order to obtain the rate $o\left(n^{-1/2+1/p-l/2p}\log(n)^{-l/2p}\right)$ for the i.i.d. case which would slightly improve on Theorem 6 [2] in which the rate $O\left(n^{-1/2+1/p-l/2p}\right)$ was obtained. Our approach would also be able to deal with varying moment assumption where each variable W_i admits a finite moment of order $p+l_i$ for non identically distributed summands.

4 Diffusion interpolation approach

Let $p>0$ and W be a random variable drawn from a measure ν on $\mathbb{R}^d$. In the following, we assume ν admits a density h with respect to the Gaussian measure which is both bounded and with bounded gradient. These additional assumptions can later be lifted to obtain Theorem 3.1 using approximation arguments similar to those developed in Section 8 [2].

Let $t>0$ and let us consider the random variable $F_t := e^{-t}W + \sqrt{1-e^{-2t}}Z$, where Z is a random variable drawn from the d-dimensional standard Gaussian measure γ and independent of W. We denote by ν_t the measure of F_t. Due to our assumptions on h, ν_t admits a smooth density h_t with respect to γ. We can thus consider the score function of F_t defined by

$$\rho_t \coloneqq \nabla \log h_t(F_t).$$

Then, by Equation (3.8) [9], we have

$$W_p(\nu,\nu_t) \leq \int_0^t \|\rho_t\|_p \, dt$$

and, since $\lim_{t\to\infty} \nu_t = \gamma$,

$$W_p(\nu,\gamma) \leq \int_0^\infty \|\rho_t\|_p \, dt.$$

We are thus left with bounding $\|\rho_t\|_p$ for all $t \geq 0$.

One can first remark that this score function verifies the following formula (see e.g. Lemma IV.1 [10]):

$$\rho_t = e^{-t}\mathbb{E}\left[W - \frac{Z}{\sqrt{\Delta(t)}} \mid F_t\right] \text{ a.s.}, \tag{4.1}$$

where $\Delta(t) \coloneqq e^{2t}-1$. A first, somewhat trivial, bound on $\|\rho_t\|_p$ can then be obtained by applying Jensen's and the triangular inequalities:

$$\|\rho_t\|_p \leq e^{-t}\left(\|\mathbb{E}[W \mid F_t]\|_p + \frac{\|\mathbb{E}[Z \mid F_t]\|_p}{\sqrt{\Delta(t)}}\right) \leq e^{-t}\left(\|W\|_p + \frac{\|Z\|_p}{\sqrt{\Delta(t)}}\right). \tag{4.2}$$

Note that this bound can still be nearly optimal for small values of t. Indeed, whenever W takes values in $h\mathbb{Z}^d$, one has, for small enough values of $t \ll h$,

$$W_2(\nu, \nu_t) \approx \|F_t - W\|_2 = (1-e^{-t})\,\|W\|_2 + \sqrt{1-e^{-2t}}\,\|Z\|_2 = \int_0^t e^{-t}\left(\|W\|_2 + \frac{\|Z\|_2}{\sqrt{\Delta(t)}}\right) dt.$$

However, for continuous measures ν or for higher values of t, it is usually possible to obtain better bounds on $\|\rho_t\|_p$. For instance, (1.1) is obtained by combining (4.2) with another bound on $\|\rho_t\|_p$ which holds for large values of t. A similar approach was used in [5] to provide quantitative results for normal approximation in various frameworks such as Wiener chaos or homogeneous sums. In this work, we refine this approach by using three different bounds: (4.2) for small values of t, a bound for medium values of t highlighting the small-scale behaviour of the measure ν and a last bound for larger values of t which depends on the large-scale structure of ν through its moments.

5 Bounding $\|\rho_t\|_p$

5.1 Small times

Let $p \geq 2$ and let $W = \sum_{i=1}^n W_i$ where the $(W_i)_{1\leq i\leq n}$ are centered and independent random variables on $\mathbb{R}^d$ with finite moment of order p. If $\mathbb{E}[W^{\otimes 2}] = I_d$, there exists $C > 0$ such that

$$\|\rho_t\|_p \leq \Psi_1(t) \coloneqq Ce^{-t}\left(\sqrt{dp}\left(1 + \frac{1}{\sqrt{\Delta(t)}}\right) + pL_p^{1/p}\right). \tag{5.1}$$

Indeed, since the $(W_i)_{1\leq i\leq n}$ are independent and centered, we can use Rosenthal's inequality (see Lemma 6.1) to obtain

$$\|W\|_p \leq C\left(\sqrt{dp} + pL_p^{1/p}\right).$$

On the other hand, by Lemma 6.2,

$$\|Z\|_p \leq \sqrt{d(p-1)}.$$

Injecting these bounds into (4.2) then yields (5.1).

5.2 Medium times

When looking at (4.2), we can see that, for small values of t, the main contributor of this bound is $\|\mathbb{E}[Z \mid F_t]\|_p/\sqrt{\Delta(t)}$. In the previous Section, we upper bounded this quantity somewhat crudely by using Jensen's inequality. In this Section, we establish a sharper bound on $\|\rho_t\|_p$ by proving a variant of Proposition 6.1 [5] leveraging the small scale features of W. We start by covering the more general exchangeable pair framework, a standard framework for applying Stein's method, before tackling the specific Central Limit Theorem case.

5.2.1 Exchangeable pairs framework

Proposition 5.1. *Let $p \geq 2$ and (W, W') be a pair of random variables on $\mathbb{R}^d$ such that (W, W') and (W', W) follow the same law. For any $t \geq 0$, let $\eta_p(t) = \Delta(t)/(p-1)$ and $D_t = (W' - W)1_{\|W'-W\|^2 \leq \eta_p(t)}$. For any $0 < s < t$ such that $\mathbb{E}[D_s^{\otimes 2}]$ is positive-definite, we have*

$$\|\rho_t\|_p \leq e^{-t}\left(\|\mathbb{E}[\Lambda_s D_s + W \mid W]\|_p + \frac{C}{\sqrt{\eta_p(t)}}\|\mathbb{E}[\Gamma_s \mid W] - \mathbb{E}[\Gamma_s]\|_p + \frac{C\sqrt{d}\eta_p(s)}{\eta_p(t)^{3/2}}\right),$$

where $C > 0$ is a generic constant, $\Lambda_s = \mathbb{E}[D_s^{\otimes 2}]^{-1}$ and $\Gamma_s = \frac{1}{2}\Lambda_s D_s^{\otimes 2}$.

The proof of this result mostly follows the proof of Proposition 6.1 [5].

Proof. Let $0 < s < t$ and let

$$\tau_t = \left(\Lambda_s D_s + \frac{\Gamma_s Z}{\sqrt{\Delta(t)}} + \sum_{k=3}^{\infty} a_k \frac{(\Gamma_s \otimes D_s^{\otimes(k-1)}) H_k(Z)}{\Delta(t)^{k/2}} \right),$$

with $a_k = \frac{1}{k!} - \frac{1}{4(k-2)!}$. A small modification of Lemma 6.5 [4] (see also the proof of Lemma 5.4) gives

$$\mathbb{E}[\tau_t \mid F_t] = 0.$$

Therefore,

$$\rho_t = \rho_t + e^{-t}\mathbb{E}[\tau_t \mid F_t]$$

and using (4.1) along with the triangle inequality yields

$$e^t \|\rho_t\|_p \leq \|\mathbb{E}[\Lambda_s D_s + W \mid F_t]\|_p + \frac{1}{\sqrt{\Delta(t)}} \|\mathbb{E}\left[(\Gamma_s - I_d) Z \mid F_t\right]\|_p + \sum_{k=3}^{\infty} \frac{a_k}{\Delta(t)^{k/2}} \left\| \mathbb{E}[(\Gamma_s \otimes D_s^{\otimes(k-1)}) H_k(Z) \mid F_t] \right\|_p.$$

Then, since Z and W are independent, we have, by Jensen's inequality,

$$e^t \|\rho_t\|_p \leq \|\mathbb{E}[\Lambda_s D_s + W \mid W]\|_p + \frac{1}{\sqrt{\Delta(t)}} \|(\mathbb{E}\left[\Gamma_s \mid W\right] - I_d) Z\|_p + \sum_{k=3}^{\infty} \frac{a_k}{\Delta(t)^{k/2}} \left\| \mathbb{E}[\Gamma_s \otimes D_s^{\otimes(k-1)} \mid W] H_k(Z) \right\|_p$$

and, by Lemma 6.2,

$$e^t \|\rho_t\|_p \leq \|\mathbb{E}[\Lambda_s D_s + W \mid W]\|_p + \frac{1}{\sqrt{\eta_p(t)}} \|\mathbb{E}\left[\Gamma_s \mid W\right] - I_d\|_p + \sum_{k=3}^{\infty} \frac{a_k \sqrt{k!}}{\eta_p(t)^{k/2}} \left\| \mathbb{E}[\Gamma_s \otimes D_s^{\otimes(k-1)} \mid W] \right\|_p.$$

Since Γ_s is positive-definite, we have, for any $k \geq 3$,

$$\begin{aligned} \left\| \mathbb{E}[\Gamma_s \otimes D_s^{\otimes(k-1)} \mid W] \right\|_p &\leq \left\| \mathbb{E}[\Gamma_s \|D_s\|^{k-1} \mid W] \right\|_p \\ &\leq \eta_p(s)^{(k-1)/2} \|\mathbb{E}[\Gamma_s \mid W]\|_p \\ &\leq \eta_p(s)^{(k-1)/2} \left(\|\mathbb{E}[\Gamma_s \mid W] - I_d\|_p + \|I_d\| \right) \\ &\leq \eta_p(s)^{(k-1)/2} \left(\|\mathbb{E}[\Gamma_s \mid W] - I_d\|_p + \sqrt{d} \right). \end{aligned}$$

Thus, since $\sum_{k=3}^{\infty} a_k \sqrt{k!} < \infty$ and $\eta_p(s) \leq \eta_p(t)$, we obtain that there exists $C > 0$ such that

$$e^t \|\rho_t\|_p \leq \|\mathbb{E}[\Lambda_s D_s + W \mid W]\|_p + \frac{C}{\sqrt{\eta_p(t)}} \|\mathbb{E}\left[\Gamma_s \mid W\right] - I_d\|_p + \frac{C\sqrt{d}\eta_p(s)}{\eta_p(t)^{3/2}}.$$

Finally, one can remark that, by definition of Γ_s,

$$\mathbb{E}[\Gamma_s] = I_d,$$

concluding the proof. □

5.2.2 Sum of independent variables

Proposition 5.2. *Let $W = \sum_{i=1}^n W_i$ where the $(W_i)_{1\leq i\leq n}$ are independent random variables on $\mathbb{R}^d$ with finite moment of order $p \geq 2$. For any $i \in \{1,\ldots,n\}$ and $\beta > 0$, let $D_{i,\beta} = (W_i' - W_i)1_{\|D_i\|\leq\beta}$ where W_i' is an independent copy of W_i. Suppose there exists $\beta > 0$ such that*

$$\Lambda_\beta^{-1} = \sum_{i=1}^n \mathbb{E}[D_{i,\beta}^{\otimes 2}]$$

is positive-definite. Then, for any t such that $\Delta(t) \geq (p-1)\beta^2$, there exists $C > 0$ such that

$$\|\rho_t\|_p \leq \Psi_2(t) \coloneqq C\left(\sqrt{p}\left(\sqrt{\beta_2} + \sqrt{L_2}\right) + p\left(\beta_p^{1/p} + L_p^{1/p}\right) + \frac{p^{3/2}\sqrt{d}\beta^2}{\Delta(t)^{3/2}}\right),$$

where

$$\forall q \geq 0, \beta_q = \sum_{i=1}^n \mathbb{E}\left[\|\Lambda_\beta D_{i,\beta}\|^q\right]$$

and

$$\forall q \geq 0, L_q = \sum_{i=1}^n \mathbb{E}\left[\|W_i\|^q\right].$$

In the following, we denote by C a generic positive constant. Let s be such that $\Delta(s) = (p-1)\beta^2$ and let $t > s$. Let $W' = W + (W_I' - W_I)$ where I is a uniform random variable on $\{1,\ldots,n\}$. Since (W,W') and (W',W) follow the same law, we can apply Proposition 5.1 to obtain

$$e^t\|\rho_t\|_p \leq \|\mathbb{E}[\Lambda_s D_s \mid W]\|_p + \|W\|_p + \sqrt{\frac{p-1}{\Delta(t)}}\|\mathbb{E}[\Gamma_s \mid W] - \mathbb{E}[\Gamma_s]\|_p + C\frac{(p-1)^{3/2}\sqrt{d}\beta^2}{\Delta(t)^{3/2}},$$

with $\Lambda_s = n\Lambda_\beta$. First, following the proof of (5.1), we have

$$\|W\|_p \leq C\left(\sqrt{pL_2} + pL_p^{1/p}\right).$$

Then, by definition of D_s and since I is independent of W,

$$\mathbb{E}[D_s \mid W] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[D_{i,\beta} \mid W].$$

Hence,

$$\|\mathbb{E}[\Lambda_s D_s \mid W]\|_p = \left\|\Lambda_\beta \sum_{i=1}^n \mathbb{E}\left[D_{i,\beta} \mid W\right]\right\|_p$$

and, by Jensen's inequality,

$$\|\mathbb{E}[\Lambda_s D_s \mid W]\|_p \leq \left\|\sum_{i=1}^n \Lambda_\beta D_{i,\beta}\right\|_p.$$

Let $i \in \{1,\ldots,n\}$. Since W_i' and W_i are independent, we have

$$\mathbb{E}[D_{i,\beta}] = 0.$$

We can thus apply Rosenthal's inequality (see Lemma 6.1) to obtain

$$\begin{aligned}\|\mathbb{E}[\Lambda_s D_s \mid W]\|_p &\leq C\sqrt{p}\left(\sum_{i=1}^n \|\Lambda_\beta D_{i,\beta}\|_2^2\right)^{1/2} + Cp\left(\sum_{i=1}^n \|\Lambda_\beta D_{i,\beta}\|_p^p\right)^{1/p} \\ &\leq C\left(\sqrt{p\beta_2} + p\beta_p^{1/p}\right).\end{aligned}$$

Similarly,

$$\|\mathbb{E}[\Gamma_s \mid W] - \mathbb{E}[\Gamma_s]\|_p \leq C\sqrt{p}\left(\sum_{i=1}^n \left\|\Lambda_\beta D_{i,\beta}^{\otimes 2}\right\|_2^2\right)^{1/2} + Cp\left(\sum_{i=1}^n \left\|\Lambda_\beta D_{i,\beta}^{\otimes 2}\right\|_p^p\right)^{1/p}$$

and, since $\|D_{i,\beta}\| \leq \beta \leq \sqrt{\Delta(t)/(p-1)}$, we can use Cauchy-Schwarz inequality to obtain

$$\left\|\Lambda_\beta D_{i,\beta}^{\otimes 2}\right\| \leq \|\Lambda_\beta D_{i,\beta}\| \, \|D_{i,\beta}\| \leq \sqrt{\frac{\Delta(t)}{p-1}} \, \|\Lambda_\beta D_{i,\beta}\| \, .$$

Therefore,

$$\begin{aligned}\sqrt{\frac{p-1}{\Delta(t)}}\|\mathbb{E}[\Gamma_s \mid W] - \mathbb{E}[\Gamma_s]\|_p &\leq C\sqrt{p}\left(\sum_{i=1}^n \|\Lambda_\beta D_{i,\beta}\|_2^2\right)^{1/2} + Cp\left(\sum_{i=1}^n \|\Lambda_\beta D_{i,\beta}\|_p^p\right)^{1/p} \\ &\leq C\left(\sqrt{p\beta_2} + p\beta_p^{1/p}\right)\end{aligned}$$

which concludes the proof.

5.3 Large times

Finally, we are left with bounding $\|\rho_t\|_p$ for “large” values of t by using large-scale features of the $(W_i)_{1\leq i\leq n}$. In practice, we improve on Proposition 6.1 [5]. However, while this result was derived in the general exchangeable pairs framework, our improvements require dealing with sums of independent random variables and thus to the Central Limit Theorem case.

Proposition 5.3. *Suppose $W = \sum_{i=1}^n W_i$ where the $(W_i)_{1\leq i\leq n}$ are centered independent random variables on $\mathbb{R}^d$ with finite moment of order $p+2$ such that $\mathbb{E}[W^{\otimes 2}] = I_d$. There exists $C > 0$ such that for any $p < q \leq p+2$ and r verifying $\frac{1}{q} + \frac{1}{r} = \frac{1}{p}$ and any t such that $\Delta(t) > (p-1)\max_{i\in\{1,\dots,n\}} \|\mathbb{E}[W_i^{\otimes 2}]\|$, we have*

$$\begin{aligned}\|\rho_t\|_p \leq \Psi_3(t) \coloneqq \frac{e^{-3t}\|\mathbb{E}[W^{\otimes 3}]H_2(Z)\|_p}{2} &+ \frac{C\|\mathbb{E}[W^{\otimes 3}]\|W_q(\nu,\gamma)}{\eta_{2r}(t)^{3/2}} \\ &+ C\left(\sqrt{\frac{pN_4(t)}{\eta_p(t)}} + p\left(\frac{N_{p+2}(t)}{\eta_p(t)}\right)^{1/p}\right) + \frac{N_4'(t)}{\eta_p(t)},\end{aligned}$$

where

- $\eta_p(t) = \frac{\Delta(t)}{p-1}$;
- $\xi_i(t) = \log\left(\frac{\eta_p(t)}{\|\mathbb{E}[W_i^{\otimes 2}]\|}\right)$;
- $\forall q > 2, N_q(t) = \sum_{i=1}^n \frac{\mathbb{E}[\|D_i\|^q(1_{\|D_i\|^2\geq \eta_p(t)\xi_i(t)} + \xi_i(t)^{-1})^{q/2-1}]}{\xi_i(t)}$;
- $N_4'(t) = \sum_{i=1}^n \frac{\|\mathbb{E}[D_i^{\otimes 2}\|W_i\|\|D_i\|]\|}{\sqrt{\|\mathbb{E}[W_i^{\otimes 2}]\|}\xi_i(t)^{3/2}}$.

For any $i \in \{1,\dots,n\}$ and any $t > 0$, let $D_{i,t} = D_i 1_{\|D_i\|^2\leq \eta_p(t)\xi_i(t)}$. Let us first rewrite ρ_t with the help of the following result.

Lemma 5.4. *For any $i \in \{1,\dots,n\}$, the quantity*

$$\tau_{i,t} = \mathbb{E}\left[D_{i,t} \mid F_t\right] + \sum_{k=1}^\infty \frac{\mathbb{E}[(W_i' \otimes D_{i,t}^{\otimes k})H_k(Z) \mid F_t]}{k!\Delta(t)^{k/2}}$$

verifies

$$\mathbb{E}[\tau_{i,t} \mid F_t] = 0.$$

Proof. Let $i \in \{1, \dots, n\}$ and let ϕ be a smooth test function. Since $\Phi : x \to \mathbb{E}[\phi(e^{-t}x + \sqrt{1-e^{-2t}}Z)]$ is real analytic (see e.g. Lemma 1 [2] or Lemma 6.4 [5]), we have

$$\begin{aligned}\mathbb{E}[W_i'\phi(F_t + e^{-t}D_{i,t})1_{\|D_i\|^2 \le \eta_p(t)\xi_i(t)}] &= \sum_{k=0}^{\infty} \frac{e^{-kt}}{k!} \mathbb{E}\left[W_i' \left\langle D_{i,t}^{\otimes k}, \nabla^k \phi(F_t)\right\rangle\right] \\ &= \sum_{k=0}^{\infty} \frac{e^{-kt}}{k!} \mathbb{E}\left[(W_i' \otimes D_{i,t}^{\otimes k})\nabla^k \phi(F_t)\right].\end{aligned}$$

Thus, by performing multiple integrations by parts with respect to the Gaussian measure (see e.g. Equation (16) [2]), we obtain

$$\mathbb{E}[W_i'\phi(F_t + e^{-t}D_{i,t})1_{\|D_i\|^2 \le \eta_p(t)\xi_i(t)}] = \sum_{k=0}^{\infty} \frac{\mathbb{E}\left[(W_i' \otimes D_{i,t}^{\otimes k})H_k(Z)\phi(F_t)\right]}{k!\Delta(t)^{k/2}}.$$

Finally, since W_i and W_i' are independent and identically distributed, we have

$$\mathbb{E}[W_i'\phi(F_t + e^{-t}D_{i,t})1_{\|D_i\|^2 \le \eta_p(t)\xi_i(t)}] = \mathbb{E}[W_i\phi(F_t)1_{\|D_i\|^2 \le \eta_p(t)\xi_i(t)}]$$

concluding the proof. □

We are now ready to start the proof of Proposition 5.3. Using Lemma 5.4, we obtain

$$\rho_t = \rho_t + e^{-t}\sum_{i=1}^{n} \mathbb{E}[\tau_{i,t} \mid F_t].$$

Then, since $\sum_{i=1}^n \mathbb{E}[W_i^{\otimes 2}] = I_d$ and $\sum_{i=1}^n \mathbb{E}[W_i^{\otimes 3}] = \mathbb{E}[W^{\otimes 3}]$, we can write

$$\rho_t = \frac{e^{-t}\mathbb{E}[W^{\otimes 3}]\mathbb{E}[H_2(Z) \mid F_t]}{2\Delta(t)} + e^{-t}\sum_{i=1}^{n} \mathbb{E}\left[W_i - \frac{\mathbb{E}[W_i^{\otimes 2}]}{\sqrt{\Delta(t)}}Z - \frac{\mathbb{E}[W_i^{\otimes 3}]}{2\Delta(t)}H_2(Z) + \tau_{i,t} \mid F_t\right].$$

Thus, combining the triangle inequality, Jensen's inequality and Lemma 6.2, we obtain

$$\|\rho_t\|_p \le \frac{e^{-t}\|\|\mathbb{E}[\mathbb{E}[W^{\otimes 3}]H_2(Z) \mid F_t]\|_p}{2\Delta(t)} + e^{-t}A(t),$$

where

$$A(t) = \mathbb{E}\left[\left(\sum_{k=0}^{\infty}\left\|\sum_{i=1}^{n} A_{k,i}\right\|^2\right)^{p/2}\right]^{1/p}$$

with

- $A_{0,i} = \mathbb{E}[D_{i,t} + W_i \mid W_i]$;
- $\forall k \in \{1,2\}, A_{k,i} = \frac{\mathbb{E}[W_i' \otimes D_{i,t}^{\otimes k} | W_i] - \mathbb{E}[W_i^{\otimes(k+1)}]}{\sqrt{k!}\eta_p(t)^{k/2}}$;
- $\forall k > 2, A_{k,i} = \frac{\mathbb{E}[W_i' \otimes D_{i,t}^{\otimes k} | W_i]}{\sqrt{k!}\eta_p(t)^{k/2}}$.

First, by Lemmas 6.3 and 6.4, we have

$$\frac{e^{-t}\|\|\mathbb{E}[\mathbb{E}[W^{\otimes 3}]H_2(Z) \mid F_t]\|_p}{2\Delta(t)} \le \frac{e^{-3t}}{2}\|\mathbb{E}[W^{\otimes 3}]H_2(Z)\|_p + \frac{C\|\mathbb{E}[W^{\otimes 3}]\|W_q(\nu,\gamma)}{\eta_{2r}(t)^{3/2}}$$

where $q > p$ and r is such that

$$\frac{1}{q} + \frac{1}{r} = \frac{1}{p}.$$

Let $\bar{D}_{i,t} = D_i 1_{\|D_i\|^2 \geq \eta_p(t)\xi_i(t)}$. In order to deal with $A(t)$, let us first remark that, since $\mathbb{E}[W_i] = 0$ and since W_i' and W_i are independent, we have

$$A_{0,i} = \mathbb{E}[D_{i,t} + W_i \mid W_i] = \mathbb{E}[D_{i,t} - D_i \mid W_i] = -\mathbb{E}[\bar{D}_{i,t} \mid W_i].$$

And similarly,

$$A_{1,i} = -\frac{\mathbb{E}[W_i' \otimes \bar{D}_{i,t} \mid W_i]}{\sqrt{\eta_p(t)}}.$$

Let us also note that

$$\mathbb{E}[A_{2,i}] = -\frac{\mathbb{E}[W_i' \otimes \bar{D}_{i,t}^{\otimes 2}]}{2\eta_p(t)}.$$

Then, viewing $A(t)$ as the p-norm of an infinite-dimensional vector, we can apply Rosenthal's inequality (see Lemma 6.1) to obtain

$$\|\rho_t\|_p \leq \sum_{i=1}^{n}\left(\sum_{k=0}^{\infty}\|\mathbb{E}[A_{k,i}]\|^2\right)^{1/2} + C\sqrt{p}\left(\sum_{i=1}^{n}\mathbb{E}\left[\sum_{k=0}^{\infty}\|B_{k,i}\|^2\right]\right)^{1/2} + Cp\left(\sum_{i=1}^{n}\mathbb{E}\left[\left(\sum_{k=0}^{\infty}\|B_{k,i}\|^2\right)^{p/2}\right]\right)^{1/p},$$

where

$$\forall k \in \mathbb{N}, 1 \leq i \leq n, B_{k,i} = \begin{cases} A_{k,i} \text{ if } k \neq 2 \\ \frac{\mathbb{E}[W_i' \otimes D_{i,t}^{\otimes 2} | W_i]}{\sqrt{2}\eta_p(t)} \text{ if } k = 2. \end{cases}$$

Let us conclude the proof by bounding these quantities.

5.3.1 Bounding the first term

Let $i \in \{1, \ldots, n\}$. First, let us note that since W_i and W_i' are independent,

$$\mathbb{E}[A_{0,i}] = \mathbb{E}[\bar{D}_{i,t}] = 0.$$

On the other hand, since $\mathbb{E}[W_i] = 0$ and $\mathbb{E}[W_i^{\otimes 2}] = I_d$, we have $\mathbb{E}[W_i' \otimes \bar{D}_{i,t}] = -\mathbb{E}[W_i \otimes \bar{D}_{i,t}]$. Hence,

$$\mathbb{E}[W_i' \otimes \bar{D}_{i,t}] = \frac{\mathbb{E}[\bar{D}_{i,t}^{\otimes 2}]}{2}$$

and

$$\|\mathbb{E}[A_{1,i}]\|^2 = \frac{\|\mathbb{E}[\bar{D}_{i,t}^{\otimes 2}]\|^2}{4\eta_p(t)} \leq \frac{\|\mathbb{E}[D_i^{\otimes 2}\|D_i\|^2]\|^2}{4\eta_p(t)^3\xi_i(t)^2} \leq \frac{\|\mathbb{E}[D_i^{\otimes 2}\|D_i\|\|W_i\|]\|^2}{(\eta_p(t)\xi_i(t))^3}\xi_i(t).$$

On the other hand,

$$\|\mathbb{E}[A_{2,i}]\|^2 \leq \frac{\|\mathbb{E}[\bar{D}_{i,t}^{\otimes 2}\|W_i\|]\|^2}{\eta_p(t)^2} \leq \frac{\|\mathbb{E}[D_i^{\otimes 2}\|D_i\|\|W_i\|]\|^2}{(\eta_p(t)\xi_i(t))^3}\frac{\xi_i(t)^2}{2}.$$

Finally, for any $k \geq 3$,

$$\|\mathbb{E}[A_{k,i}]\|^2 \leq \frac{\|\mathbb{E}[D_{i,t}^{\otimes 2}\|D_{i,t}\|^{k-2}\|W_i\|]\|^2}{k!\eta_p(t)^k} \leq \frac{\|\mathbb{E}[D_i^{\otimes 2}\|D_i\|\|W_i\|]\|^2}{(\eta_p(t)\xi_i(t))^3}\frac{\xi_i(t)^k}{k!}.$$

Therefore, by definition of $\xi_i(t)$,

$$\begin{aligned}\sum_{k=0}^{\infty}\|\mathbb{E}[A_{k,i}]\|^2 &\leq \frac{\|\mathbb{E}[D_i^{\otimes 2}\|D_i\|\|W_i\|]\|^2}{(\eta_p(t)\xi_i(t))^3}\sum_{k=1}^{\infty}\frac{\xi_i(t)^k}{k!}\\ &\leq \frac{\|\mathbb{E}[D_i^{\otimes 2}\|D_i\|\|W_i\|]\|^2}{(\eta_p(t)\xi_i(t))^3}e^{\xi_i(t)}\\ &\leq \frac{\|\mathbb{E}[D_i^{\otimes 2}\|D_i\|\|W_i\|]\|^2}{\|\mathbb{E}[W_i^{\otimes 2}]\|\eta_p(t)^2\xi_i(t)^3}\end{aligned}$$

and thus

$$\sum_{i=1}^{n}\left(\sum_{k=0}^{\infty}\|\mathbb{E}[A_{k,i}]\|^2\right)^{1/2} \leq \frac{N_4'(t)}{\eta_p(t)}.$$

5.3.2 Bounding the last two terms

Let $i \in \{1,\dots,n\}, q \in [2,p]$. First, by Jensen' inequality and by definition of $\bar{D}_{i,t}$,

$$\|B_{0,i}\|^2 \leq \|\mathbb{E}[\bar{D}_{i,t} \mid W_i]\|^2 \leq \mathbb{E}[\|\bar{D}_{i,t}\|^2 \mid W_i] \leq \frac{\mathbb{E}[\|\bar{D}_{i,t}\|^{2+4/q} \mid W_i]}{(\eta_p(t)\xi_i(t))^{2/q}}.$$

Let W_i'' and $\bar{D}_{i,t}'$ be a conditionally independent copies of W_i' and $\bar{D}_{i,t}$ with respect to W_i. We have

$$\begin{aligned}\|\mathbb{E}[W_i' \otimes \bar{D}_{i,t} \mid W_i]\|^2 &= \mathbb{E}[\langle W_i' \otimes \bar{D}_{i,t}, W_i'' \otimes \bar{D}_{i,t}'\rangle \mid W_i]\\ &= \mathbb{E}[\langle W_i', W_i''\rangle \langle \bar{D}_{i,t}, \bar{D}_{i,t}'\rangle \mid W_i]\end{aligned}$$

and, by Cauchy-Schwarz's inequality,

$$\begin{aligned}\|\mathbb{E}[W_i' \otimes \bar{D}_{i,t} \mid W_i]\|^2 &\leq \mathbb{E}[\langle W_i', W_i''\rangle^2 \mid W_i]^{1/2}\mathbb{E}[\langle \bar{D}_{i,t}, \bar{D}_{i,t}'\rangle^2 \mid W_i]^{1/2}\\ &\leq \mathbb{E}[\langle W_i'^{\otimes 2}, W_i''^{\otimes 2}\rangle \mid W_i]^{1/2}\mathbb{E}[\langle \bar{D}_{i,t}^{\otimes 2}, \bar{D}_{i,t}'^{\otimes 2}\rangle \mid W_i]^{1/2}\\ &\leq \|\mathbb{E}[W_i'^{\otimes 2} \mid W_i]\|\|\mathbb{E}[\bar{D}_{i,t}^{\otimes 2} \mid W_i]\|.\end{aligned}$$

Since W_i' is independent of W, $\|\mathbb{E}[W_i'^{\otimes 2} \mid W_i]\| = \|\mathbb{E}[W_i'^{\otimes 2}]\| = \|\mathbb{E}[W_i^{\otimes 2}]\|$. Thus,

$$\begin{aligned}\|B_{1,i}\|^2 &= \frac{\|\mathbb{E}[W_i' \otimes \bar{D}_{i,t} \mid W_i]\|^2}{\eta_p(t)}\\ &\leq \frac{\|\mathbb{E}[W_i^{\otimes 2}]\|\mathbb{E}[\|\bar{D}_{i,t}\|^2 \mid W_i]}{\eta_p(t)}\\ &\leq \frac{\|\mathbb{E}[W_i^{\otimes 2}]\|\mathbb{E}[\|D_i\|^{2+4/q} \mid W_i]}{(\eta_p(t)\xi_i(t))^{1+2/q}}\xi_i(t).\end{aligned}$$

On the other hand,

$$\begin{aligned}\|B_{2,i}\|^2 &\leq \frac{\|\mathbb{E}[W_i' \otimes D_{i,t}^{\otimes 2} \mid W_i]\|^2}{2\eta_p(t)^2}\\ &\leq \frac{\|\mathbb{E}[W_i^{\otimes 2}]\|\mathbb{E}[\|D_{i,t}\|^4 \mid W_i]}{2\eta_p(t)^2}\\ &\leq \frac{\|\mathbb{E}[W_i^{\otimes 2}]\|\mathbb{E}[\|D_i\|^{2+4/q} \mid W_i]}{(\eta_p(t)\xi_i(t))^{1+2/q}}\frac{\xi_i^2(t)}{2}.\end{aligned}$$

Finally, for any $k \geq 3$,

$$\|B_{k,i}\|^2 \leq \frac{\|\mathbb{E}[W_i^{\otimes 2}]\|\mathbb{E}[\|D_{i,t}\|^{2k} \mid W_i]}{k!\eta_p(t)^k} \leq \frac{\|\mathbb{E}[W_i^{\otimes 2}]\|\mathbb{E}[\|D_i\|^{2+4/q} \mid W_i]}{(\eta_p(t)\xi_i(t))^{1+2/q}} \frac{\xi_i(t)^k}{k!}.$$

Combining both these bounds yields, by definition of $\xi_i(t)$

$$\begin{aligned}\sum_{k=1}^{\infty} \|B_{k,i}\|^2 &\leq \frac{\|\mathbb{E}[W_i^{\otimes 2}]\|\mathbb{E}[\|D_i\|^{2+4/q} \mid W_i]}{(\eta_p(t)\xi_i(t))^{1+2/q}} \sum_{k=1}^{\infty} \frac{\xi_i(t)}{k!} \\ &\leq \frac{\|\mathbb{E}[W_i^{\otimes 2}]\|\mathbb{E}[\|D_i\|^{2+4/q} \mid W_i]}{(\eta_p(t)\xi_i(t))^{1+2/q}} e^{\xi_i(t)} \\ &\leq \frac{\mathbb{E}[\|D_i\|^{2+4/q} \mid W_i]}{\xi_i(t)^{1+2/q}\eta_p(t)^{2/q}}\end{aligned}$$

Thus,

$$\sum_{k=0}^{\infty} \|B_{k,i}\|^2 \leq \frac{\mathbb{E}[\|D_i\|^{2+4/q}(1_{\|D_i\|^2 \geq \eta_p(t)\xi_i(t)} + \xi_i(t)^{-1}) \mid W_i]}{(\eta_p(t)\xi_i(t))^{2/q}}$$

and, by Jensen's inequality,

$$\mathbb{E}\left[\left(\sum_{k=0}^{\infty} \|B_{k,i}\|^2\right)^{q/2}\right] \leq \frac{\mathbb{E}[\|D_i\|^{q+2}(1_{\|D_i\|^2 \geq \eta_p(t)\xi_i(t)} + \xi_i(t)^{-1})^{q/2}]}{\eta_p(t)\xi_i(t)}.$$

Therefore,

$$\left(\sum_{i=1}^{n} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \|B_{k,i}\|^2\right)^{q/2}\right]\right)^{1/q} \leq \left(\frac{N_{q+2}(t)}{\eta_p(t)}\right)^{1/q}.$$

5.4 Combining times

We are now ready to conclude the proof of Theorem 3.1. Let ϵ_1 and ϵ_2 such that $\eta_p(\epsilon_1) := \frac{\Delta(\epsilon_1)}{p-1} = \beta^2$ and ϵ_2 be such that $\eta_p(\epsilon_2) := \frac{\Delta(\epsilon_2)}{p-1} = \epsilon := \max_{i\in\{1,\ldots,n\}} \|\mathbb{E}[W_i^{\otimes 2}]\|^{2/3}$. Remark that, by assumption, $\epsilon_1 < \epsilon_2$. In the following computations, we will rely on the fact that $\Delta(t) \geq 2t$. By (5.1) and Propositions 5.2 and 5.3 we have

$$\int_0^{\infty} \|\rho_t\|_p \, dt \leq \int_0^{\epsilon_1} \Psi_1(t) \, dt + \int_{\epsilon_1}^{\epsilon_2} \Psi_2(t) \, dt + \int_{\epsilon_2}^{\infty} \Psi_3(t) \, dt.$$

First, since $\beta^2 < \epsilon$

$$\begin{aligned}\int_0^{\epsilon_1} \Psi_1(t) \, dt &\leq C \left(p\sqrt{\eta_p(\epsilon_1)d} + p\eta_p(\epsilon_1)\left(\sqrt{pd} + pL_p^{1/p}\right)\right) \\ &\leq C\left(p\beta\sqrt{d} + p\epsilon\left(\sqrt{pd} + pL_p^{1/p}\right)\right).\end{aligned}$$

Then,

$$\begin{aligned}\int_{\epsilon_1}^{\epsilon_2} \Psi_2(t) \, dt &\leq Cp\eta_p(\epsilon_2)\left(\sqrt{p}(\sqrt{\beta_2} + \sqrt{d}) + p\left(\beta_p^{1/p} + L_p^{1/p}\right)\right) + \frac{Cp\beta^2\sqrt{d}}{\sqrt{\eta_p(\epsilon_1)}} \\ &\leq Cp\epsilon\left(\sqrt{p}(\sqrt{\beta_2} + \sqrt{d}) + p\left(\beta_p^{1/p} + L_p^{1/p}\right)\right) + Cp\beta\sqrt{d}.\end{aligned}$$

Finally, let $t \geq \epsilon_2$. Since $\eta_p(t) \geq \max_{i\in\{1,\ldots,n\}} \|\mathbb{E}[W_i^{\otimes 2}]\|^{2/3}$, we have, for any $i \in \{1,\ldots,n\}, \xi_i(t) \geq C\xi_i$. On the other hand, since $p \leq \min_{i\in\{1,\ldots,n\}} \|\mathbb{E}[W_i^{\otimes 2}]\|^{-1}$, we have

$-\log(\epsilon_2) \leq -\log(\eta_p(\epsilon_2)) - \log(p-1) \leq C\xi_i$. Hence, since $N_4(.), N_{p+2}(.)$ and $N_4'(.)$ are decreasing functions,

$$\begin{aligned}\int_{\epsilon_2}^{\infty} \Psi_3(t)\,dt \leq \frac{\|\mathbb{E}[W^{\otimes 3}]H_2(Z)\|_p}{6} + \frac{C(2r-1)^{3/2}\|\mathbb{E}[W^{\otimes 3}]\|W_q(\nu,\gamma)}{\sqrt{(p-1)\eta_p(\epsilon_2)}}\\ + Cp\left(\sqrt{N_4(\epsilon_2)} + (pN_{p+2}(\epsilon_2))^{1/p} - N_4'(\epsilon_2)\log(\epsilon_2)\right)\\ \leq \frac{\|\mathbb{E}[W^{\otimes 3}]H_2(Z)\|_p}{6} + \frac{C(2r-1)^{3/2}\|\mathbb{E}[W^{\otimes 3}]\|W_q(\nu,\gamma)}{\sqrt{(p-1)\epsilon}}\\ + Cp\left(\sqrt{N_4} + (pN_{p+2})^{1/p} + N_4'\right).\end{aligned}$$

which concludes the proof of Theorem 3.1.

6 Technical lemmas

Lemma 6.1 (Rosenthal inequality, Theorem 5.2 [11]). *There exists $C > 0$ such that, for any $p \geq 2$ and any independent random variables $(U_i)_{1\leq i\leq n}$ with finite moment of order p taking values in a Hilbert space $\mathcal{H}$, we have*

$$\left\|\sum_{i=1}^n U_i - \mathbb{E}[U_i]\right\|_{\mathcal{H},p} \leq C\sqrt{p}\left(\sum_{i=1}^n \|U_i\|_{\mathcal{H},2}^2\right)^{1/2} + Cp\left(\sum_{i=1}^n \|U_i\|_{\mathcal{H},p}^p\right)^{1/p},$$

where, for any random variable X taking values in $\mathcal{H}$ and any $q > 0$,

$$\|X\|_{\mathcal{H},q} = \mathbb{E}[\|X\|_{\mathcal{H}}^q].$$

Proof. By Theorem 5.2 [11],

$$\left\|\sum_{i=1}^n U_i - \mathbb{E}[U_i]\right\|_{\mathcal{H},p} \leq C\sqrt{p}\left(\sum_{i=1}^n \|U_i - \mathbb{E}[U_i]\|_{\mathcal{H},2}^2\right)^{1/2} + Cp\left(\sum_{i=1}^n \|U_i - \mathbb{E}[U_i]\|_{\mathcal{H},p}^p\right)^{1/p}.$$

Now for $q \in [2,p]$, combining the triangle and Jensen's inequalities yields

$$\|U_i - \mathbb{E}[U_i]\|_{\mathcal{H},q} \leq \|U_i\|_{\mathcal{H},q} + \|\mathbb{E}[U_i]\|_{\mathcal{H}} \leq 2\|U_i\|_{\mathcal{H},q},$$

concluding the proof. ☐

Lemma 6.2 (Lemma 3 [2]). *Let Z be a d-dimensional standard normal random variable. For any $p \geq 2$, $k \in \mathbb{N}$ and $M \in (\mathbb{R}^d)^{\otimes k+1}$, we have*

$$\|MH_k(Z)\|_p^2 \leq (p-1)^k k!\|M\|^2.$$

Lemma 6.3. *Let X, Y and Z be three random variables on $\mathbb{R}^d$ such that Z is drawn from the Gaussian measure γ and is independent from (X,Y). Let $q > p \geq 2$ and suppose that X and Y have finite moment of order q. Then, for any $k \geq 0$ and any $i \in \{1,\dots,d\}^k$,*

$$\|\mathbb{E}[H_i(Z) \mid X+Z] - \mathbb{E}[H_i(Z) \mid Y+Z]\|_p \leq C\sqrt{(2r-1)^{k+1}(k+1)!}\|Y-X\|_q,$$

where $C > 0$ is a generic constant, $H_i = (H_k)_i$ and r is such that $\frac{1}{r} + \frac{1}{q} = \frac{1}{p}$.

Proof. Let $\epsilon = Y - X$. We have

$$\mathbb{E}[H_i(Z) \mid X+Z] = \mathbb{E}[H_i(Z) \mid Y+Z+\epsilon]$$

and

$$\mathbb{E}[H_i(Z) \mid Y + Z] - \mathbb{E}[H_i(Z) \mid X + Z] = \int_0^1 \frac{d}{dt}\mathbb{E}[H_i(Z) \mid Y + Z + t\epsilon]\, dt.$$

Let us denote the density of Z by f_γ and the measure of (Y, ϵ) by μ. For any $t \in [0, 1]$, let

$$f(t) = \int (-1)^k \nabla_i f_\gamma \left(Z + Y - y' + t(\epsilon - \epsilon')\right) d\mu(y', \epsilon'),$$

where $\nabla_i \cdot = (\nabla^k \cdot)_i$. We then have

$$f'(t) = \int \left\langle \epsilon - \epsilon', (-1)^k \nabla\nabla_i f_\gamma \left(Z + Y - y' + t(\epsilon - \epsilon')\right)\right\rangle d\mu(y', \epsilon').$$

Similarly, letting

$$g(t) = \int f_\gamma \left(Z + Y - y' + t(\epsilon - \epsilon')\right) d\mu(y', \epsilon'),$$

we have

$$g'(t) = \int \left\langle \epsilon - \epsilon', \nabla f_\gamma \left(Z + Y - y' + t(\epsilon - \epsilon')\right)\right\rangle d\mu(y', \epsilon').$$

By definition of the conditional expectation,

- $\frac{f(t)}{g(t)} = \mathbb{E}[H_i(Z) \mid Y + Z + t\epsilon]$;
- $\frac{g'(t)}{g(t)} = \mathbb{E}[\langle \epsilon, Z\rangle \mid Y + Z + t\epsilon] - \langle \epsilon, \mathbb{E}[Z \mid Y + Z + t\epsilon]\rangle$ and
- $\frac{f'(t)}{g(t)} = \mathbb{E}[\langle \epsilon, H_{i+1}(Z)\rangle \mid Y + Z + t\epsilon] - \langle \epsilon, \mathbb{E}[H_{i+1}(Z) \mid Y + Z + t\epsilon]\rangle$,

where $H_{i+1}(x) = (-1)^{k+1} \frac{\nabla\nabla_i f_\gamma(x)}{f_\gamma(x)}$. Therefore, letting $G_t = Y + Z + t\epsilon$, we obtain

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[H_i(Z) \mid G_t] = \left(\frac{f}{g}\right)'(t) &= \langle \epsilon, \mathbb{E}[Z \mid G_t]\mathbb{E}[H_i(Z) \mid G_t] - \mathbb{E}[H_{i+1}(Z) \mid G_t]\rangle \\ &\quad - \left(\mathbb{E}[\langle \epsilon, Z\rangle \mid G_t]\mathbb{E}[H_i(Z) \mid G_t] - \mathbb{E}[\langle \epsilon, H_{i+1}(Z)\rangle \mid G_t]\right).\end{aligned}$$

Applying the triangle inequality along with Cauchy-Schwarz, Holder's and Jensen's inequalities then yields

$$\left\| \frac{d}{dt}\mathbb{E}[H_i(Z) \mid G_t] \right\|_p \leq C\|\epsilon\|_q (\|H_{i+1}(Z)\|_r + \|Z\|_{2r}\|H_i(Z)\|_{2r}),$$

where

$$\frac{1}{q} + \frac{1}{r} = \frac{1}{p}.$$

Finally applying Lemma 6.2 yields

$$\left\| \frac{d}{dt}\mathbb{E}[H_i(Z) \mid Y + Z + t\epsilon] \right\|_p \leq C\sqrt{(2r-1)^{(k+1)}(k+1)!}\|\epsilon\|_q,$$

concluding the proof. □

Lemma 6.4. *Let Y and Z be two independent standard normal random variables on $\mathbb{R}^d$. Then, for any $k \geq 1$ and any $\alpha > 0$, we have*

$$\mathbb{E}[H_k(Z) \mid \alpha Y + \sqrt{1 - \alpha^2} Z] = (1 - \alpha^2)^{k/2} H_k(\alpha Y + \sqrt{1 - \alpha^2} Z).$$

Proof. Let ϕ be a smooth function with compact support. By performing multiple integrations by parts with respect to the Gaussian measure (see e.g. Equation (16) [2]), we obtain

$$\begin{aligned}\mathbb{E}\big[\mathbb{E}[H_k(Z) \mid \alpha Y + \sqrt{1-\alpha^2} Z]\phi(\alpha Y + \sqrt{1-\alpha^2} Z)\big]\\ &= \mathbb{E}[H_k(Z)\phi(\alpha Y + \sqrt{1-\alpha^2} Z)]\\ &= (1-\alpha^2)^{k/2}\mathbb{E}[\nabla^k \phi(\alpha Y + \sqrt{1-\alpha^2} Z)]\\ &= (1-\alpha^2)^{k/2}\mathbb{E}[H_k(\alpha Y + \sqrt{1-\alpha^2} Z)\phi(\alpha Y + \sqrt{1-\alpha^2} Z)],\end{aligned}$$

concluding the proof. □

Lemma 6.5. *Let X be a random variable on $\mathbb{R}^d$ with identity covariance matrix. Then,*

$$\|\mathbb{E}[X^{\otimes 3}]\|^2 \leq \sqrt{d}\|\mathbb{E}[X^{\otimes 2}\|X\|^2]\|.$$

Proof. Let X' be an independent copy of X. By Cauchy-Schwarz's inequality, we have

$$\begin{aligned}\|\mathbb{E}[X^{\otimes 3}]\|^2 &= \mathbb{E}\left[\left\langle X^{\otimes 3}, X'^{\otimes 3}\right\rangle\right]\\ &= \mathbb{E}\left[\left\langle X, X'\right\rangle^3\right]\\ &\leq \mathbb{E}\left[\left\langle X, X'\right\rangle^2\right]^{1/2}\mathbb{E}\left[\left\langle X, X'\right\rangle^4\right]^{1/2}\\ &\leq \mathbb{E}\left[\left\langle X^{\otimes 2}, X'^{\otimes 2}\right\rangle\right]^{1/2}\mathbb{E}\left[\left\langle X^{\otimes 2}, X'^{\otimes 2}\right\rangle\|X\|^2\|X'\|^2\right]^{1/2}\\ &\leq \|\mathbb{E}[X^{\otimes 2}]\|\|\mathbb{E}[X^{\otimes 2}\|X\|^2]\|\\ &\leq \|\mathbb{E}[I_d]\|\|\mathbb{E}[X^{\otimes 2}\|X\|^2]\|\\ &\leq \sqrt{d}\|\mathbb{E}[X^{\otimes 2}\|X\|^2]\|.\end{aligned}$$

□

Lemma 6.6. *Let X be a centered random variable on $\mathbb{R}^d$ with identity covariance matrix and suppose there exists $\tau : \mathbb{R}^d \to (\mathbb{R}^d)^{\otimes 2}$ such that, for any smooth test function $\phi : \mathbb{R}^d \to \mathbb{R}^d$,*

$$\mathbb{E}[\langle X, \phi(X)\rangle] = \mathbb{E}[\langle \tau(X), \nabla\phi(X)\rangle]. \tag{6.1}$$

Then

$$\|\mathbb{E}[X^{\otimes 3}]\| \leq 2\mathbb{E}[\|\tau(X) - I_d\|^2]^{1/2}.$$

Proof. The proof follows the proof of Equation (3.14) [8], except we use (6.1) in place of the Poincaré inequality.

Let $B = \mathbb{E}[X^{\otimes 3}]$. We have

$$\|B\|^2 = \mathbb{E}\left[\left\langle X, BX^{\otimes 2}\right\rangle\right].$$

By definition of τ, we obtain

$$\begin{aligned}\|B\|^2 &= \mathbb{E}\left[\left\langle \tau(X), \nabla(BX^{\otimes 2})\right\rangle\right]\\ &= 2\mathbb{E}\left[\langle \tau(X), BX\rangle\right],\end{aligned}$$

where

$$(BX)_{i,j} = \sum_{k=1}^d B_{i,j,k}X_k.$$

Since X is centered, we have $\mathbb{E}[BX] = 0$ and

$$\|B\|^2 = 2\mathbb{E}\left[\langle \tau(X) - I_d, BX\rangle\right].$$

Finally, by Cauchy-Schwarz's inequality and since $\mathbb{E}[X^{\otimes 2}] = I_d$,

$$\begin{aligned}\|B\|^2 &\leq \mathbb{E}[\|\tau(X) - I_d\|^2]^{1/2}\mathbb{E}[\|BX\|^2]^{1/2} \\ &\leq 2\|B\|\mathbb{E}[\|\tau(X) - I_d\|^2]^{1/2},\end{aligned}$$

concluding the proof. □

References

[1] Sergey G. Bobkov, *Berry–esseen bounds and edgeworth expansions in the central limit theorem for transport distances*, Probability Theory and Related Fields **170** (2018), no. 1, 229–262. MR3748324

[2] Thomas Bonis, *Stein's method for normal approximation in wasserstein distances with application to the multivariate central limit theorem*, Probability Theory and Related Fields **178** (2020), no. 3, 827–860. MR4168389

[3] Thomas A. Courtade, Max Fathi, and Ashwin Pananjady, *Existence of Stein kernels under a spectral gap, and discrepancy bounds*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **55** (2019), no. 2, 777 – 790. MR3949953

[4] Xiao Fang and Yuta Koike, *New error bounds in multivariate normal approximations via exchangeable pairs with applications to Wishart matrices and fourth moment theorems*, The Annals of Applied Probability **32** (2022), no. 1, 602 – 631. MR4386537

[5] Xiao Fang and Yuta Koike, *From p-Wasserstein bounds to moderate deviations*, Electronic Journal of Probability **28** (2023), no. none, 1 – 52. MR4609449

[6] Xiao Fang and Yuta Koike, *Sharp high-dimensional central limit theorems for log-concave distributions*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques (forthcoming) (2023).

[7] Max Fathi, *Stein kernels and moment maps*, The Annals of Probability **47** (2019), no. 4, 2172 – 2185. MR3980918

[8] Bo'az Klartag, *Logarithmic bounds for isoperimetry and slices of convex sets*, Ars Inveniendi Analytica **Paper No. 4** (2023), 17 pp. MR4603941

[9] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati, *Stein's method, logarithmic sobolev and transport inequalities*, Geometric and Functional Analysis **25** (2015), no. 1, 256–306. MR3320893

[10] Ivan Nourdin, Giovanni Peccati, and Yvik Swan, *Integration by parts and representation of information functionals*, 2014 IEEE International Symposium on Information Theory, 2014, pp. 2217–2221.

[11] Iosif Pinelis, *Optimum Bounds for the Distributions of Martingales in Banach Spaces*, The Annals of Probability **22** (1994), no. 4, 1679 – 1706. MR1331198

[12] Emmanuel Rio, *Asymptotic constants for minimal distance in the central limit theorem*, Electronic Communications in Probability **16** (2011), 96 – 103. MR2772388

[13] Alex Zhai, *A high-dimensional clt in $\mathcal{W}_2$ distance with near optimal convergence rate*, Probability Theory and Related Fields **170** (2018), no. 3, 821–845. MR3773801

Acknowledgments. I would like to thank the anonymous referee for their many comments which helped improve this paper. I am also grateful to Olivier Guédon for pointing out [8] to me.