Regularity of solutions of the Stein equation and rates in the multivariate central limit theorem

T.O. Gallouët, G. Mijoule and Y. Swan

April 24, 2017

Abstract

Consider the multivariate Stein equation $\Delta f - x \cdot \nabla f = h(x) - \mathbb{E}h(Z)$, where Z is a standard d-dimensional Gaussian random vector. We prove that when h is α-Hölder $(0 < \alpha \leqslant 1)$, all derivatives of order 2 of the solution f_h given by the generator approach are $(\alpha - \epsilon)$-Hölder for $\epsilon > 0$ arbitrarily small, improving existing regularity results on the solution of the multivariate Gaussian Stein equation. As an application, we prove a near-optimal Berry-Esseen bound of the order $\ln n/\sqrt{n}$ in the classical multivariate CLT in 1-Wasserstein distance, as long as the underlying random variables have finite moment of order 3. When it is only assumed a finite moment of order $3 - \epsilon$ $(0 < \epsilon < 1)$, we obtain the optimal rate in $\mathcal{O}(n^{\frac{\epsilon-1}{2}})$. All constants are explicit.

1 Introduction

1.1 Multivariate Stein's method

Stein's method is a powerful tool to assess the distance between probability distributions. It was first designed for a (univariate) Gaussian target, for which Stein's idea [10] is as follows. If Z is a standard Gaussian random variable, then for a large class of functions f,

$$\mathbb{E}[f'(Z) - Zf(Z)] = 0.$$

Let X be another random variable, and assume the distance between the laws of X and Z is given by

$$d(X, Z) = \sup_{h \in \mathcal{H}} \mathbb{E}[h(X) - h(Z)], \tag{1}$$

where $\mathcal{H}$ is a class of functions. Such a representation holds for many classical distances, like the Kolmogorov, total variation, and 1-Wasserstein (or Kantorovitch) distances. Let f_h be a solution to the so-called Stein equation

$$f_h'(x) - xf_h(x) = h(x) - \mathbb{E}h(Z). \tag{2}$$

Then,

$$d(X, Z) = \sup_{h \in \mathcal{H}} \mathbb{E}[f_h'(X) - Xf_h(X)].$$

Bounding last quantity can of course only be done on a case by case basis, but a common key step is to find solutions f_h that have good regularity properties – typically, one usually tries to bound the derivatives of f_h in terms of that of h. For instance, when one consider the 1-Wasserstein distance, for which $\mathcal{H}$ is the set of 1-Lipschitz functions, then one can show that there exist a solution f_h to (2) such that f_h' and f_h'' are uniformly bounded by universal constants. This can be used, for instance, to obtain Berry-Esseen bounds in the classical central limit theorem in Wasserstein distance.

Consider now a d-dimensional Gaussian target $Z \sim \mathcal{N}(0, I_d)$. In this case, one has for a large class of functions $f : \mathbb{R}^d \to \mathbb{R}$ that $\mathbb{E}[\Delta f(Z) - Z \cdot \nabla f(Z)] = 0$ (see e.g. [4, 7]). The multivariate Stein equation then reads

$$\Delta f(x) - x \cdot \nabla f(x) = h(x) - \mathbb{E}h(Z). \tag{3}$$

Note that in dimension 1, (3) is (2) applied to f_h'. Barbour [1] identified a solution of (3) to be

$$f_h(x) = \int_0^1 \frac{1}{2t}\mathbb{E}[h(\sqrt{t}x + \sqrt{1-t}Z) - h(Z)]\, dt. \tag{4}$$

This representation is indeed the most suitable to obtain bounds on the derivatives of f_h in terms of that of h. A striking contribution is due to Chatterjee et al. [4] who proved, for instance, that

$$\forall (i,j), \quad \left|\frac{\partial^2 f_h}{\partial x_i \partial x_j}\right|_\infty \leqslant |\nabla h|_\infty, \tag{5}$$

where $|\,.\,|$ holds for the supremum norm; see also [9]. Gaunt [6] later showed a generalization of this result, namely that derivatives of order k of f_h can be bounded by derivatives of order $k-1$ of h. Note that in dimension 1, when h' is bounded, one can bound one higher derivative of f_h : indeed, it holds $|f_h^{(3)}| \leqslant 2|h'|$.

1.2 Multivariate Berry-Esseen bounds

Stein's method allows, among many other applications, to obtain Berry-Esseen type bounds in distances of the type (1) where $\mathcal{H}$ is a set of smooth functions with derivatives bounded by 1 up to some order k (which are sometimes refered to as *smooth* Wasserstein distances). Of course, the smaller k, the stronger the distance; $k=1$ leads to the classical Wasserstein distance. Let us recall the result of Chatterjee and Meckes in this direction.

Let $(X_i)_{i\geqslant 1}$ be an i.i.d. sequence of centered random vectors with unit covariance matrix. Let $W = n^{-1/2}\sum_{i=1}^n X_i$. $Z \sim \mathcal{N}(0, I_d)$. In [4], it is proved (using an exchangeable pair multivariate version of Stein's method) that if X_i has finite moment of order 4, then for any smooth h,

$$\mathbb{E}[h(W) - h(Z)] \leqslant \frac{C}{\sqrt{n}}\left(|\nabla h|_\infty + \sup_{i,j}\left|\frac{\partial^2 h}{\partial x_i \partial x_j}\right|_\infty\right),$$

where the constant C is explicit and depends on the dimension and $\mathbb{E}|X_i|^4$; see [4], Theorem 3.1 for a precise statement.

Using an approach close to Stein's method, Bonis [3] showed that, when $\mathbb{E}|X_i|^{2+\alpha} < \infty$ for some $\alpha \in (0,2)$, then $\mathcal{W}_2(W,Z) = \mathcal{O}(n^{-\alpha/4})$, where $\mathcal{W}_2$ stands for the Wasserstein distance of order 2 (which is stronger than the norm used in [4]).

Zhai [12] shows that when X_i is almost surely bounded, then a near-optimal rate of convergence in $\mathcal{O}(\ln n/\sqrt{n})$ holds, again in Wasserstein distance of order 2 (this improved a former result by Valiant et al. [11]).

Recently, Courtade et al. [5] managed to obtain the optimal rate of convergence $\mathcal{O}(n^{-1/2})$, again in 2-Wasserstein distance, under the assumption that X_i satisfies a Poincaré-type inequality. This assumption, as mentioned in [5], is not directly comparable to the ones of Zhai [12].

1.3 Contribution

In this paper, we prove new regularity results on the solution of Stein's equation (4) : namely, if h is α-Hölder for some $0 < \alpha \leqslant 1$, then for all i, j, $\frac{\partial^2 f_h}{\partial x_i \partial x_j}$ is $(\alpha - \epsilon)$-Hölder for $\epsilon > 0$ arbitrarily small. A slightly stronger (and precise) statement is given in Proposition 2.1. When applied to $\alpha = 1$, this implies that when $|\nabla h| \leqslant 1$, then the derivatives of order 2 of f_h are $(1-\epsilon)$-Hölder for any $0 < \epsilon < 1$; up to now it was only known that those derivatives were bounded. Note that from Shauder's theory, one cannot hope for the second derivative of f_h to inherit the regularity of h in the multivariate case, contrary to the univariate one. In this sense, our regularity results are sharp.

In a second step, we apply those regularity results to estimate the rate of convergence in the CLT, in Wasserstein distance. Our main result is the following theorem.

Theorem 1.1. *Let $(X_i)_{i\geqslant 1}$ be an i.i.d. sequence of random vectors with unit covariance matrix. Assume that there exists $\epsilon \in (0,1)$ such that $\mathbb{E}[|X_i|^{3-\epsilon}] < \infty$. Then*

$$\mathcal{W}\left(n^{-1/2}\sum_{i=1}^n X_i, Z\right) \leqslant d\, \frac{C(d)+2\epsilon^{-1}}{n^{\frac{1-\epsilon}{2}}} \left[\mathbb{E}|X_i|^{1-\epsilon} + d\,\mathbb{E}|X_i|^{3-\epsilon}\right],$$

where $\mathcal{W}$ stands for the 1*-Wasserstein distance, and*

$$C(d) = \sqrt{2}\frac{(d+2)\,\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})}. \tag{6}$$

Note that the rate in $\mathcal{O}(n^{(\epsilon-1)/2})$ is optimal when only assuming moments of order $3-\epsilon$; see [2] or [8]. To our knowledge, those are the first optimal rates in Wasserstein distance in the multidimensional case when assuming finite moments of order $3-\epsilon$ only.

As a corollary, we also obtain a near-optimal rate of order $\mathcal{O}(\ln/\sqrt{n})$ when X_i has finite moment of order 3; see Corollary 3.3. The constant we obtain behaves, when d becomes large, as $d^{7/2}$; this is to be compared with the sharpest rate that can expected, which is $\sqrt{d}$ (see [5]). Compared to [12], our assumption on the distribution of X_i is much weaker; however the distance used in [12] is stronger and the constants are sharper ([12] obtains the sharpest rate in d, which is $\mathcal{O}(\sqrt{n})$). [5] has the advantage of stronger rate of convergence (it is optimal when ours is near-optimal) and stronger distance, but the drawback of a less tractable assumption on the distribution of X_i (it should satisfy a Poincaré, or weigthed poincaré inequality).

2 Regularity of solutions of Stein's equation

Consider the d-dimensional multivariate Stein equation

$$\Delta f - x\cdot\nabla f = h - \mathbb{E}h(Z), \tag{7}$$

where $h \in \mathcal{C}_b^\infty(\mathbb{R}^d)$, is the space of smooth functions with bounded derivatives:

$$C_b^\infty(\mathbb{R}^d) = \left\{f \in C^\infty(\mathbb{R}^d) \middle| \forall k \in \mathbb{N}^d \setminus \{0\},\, |\partial_k f| < \infty\right\},$$

and $Z \sim \mathcal{N}(0, I_d)$. From the generator approach of Stein's method, we know that a solution to (7) is given by (see e.g. [4])

$$f_h(x) = \int_0^1 \frac{1}{2t}\mathbb{E}\bar{h}(\sqrt{t}x + \sqrt{1-t}Z)\, dt,$$

where $\bar{h} = h - \mathbb{E}h(Z)$. We have in particular, by Lebesgue's theorem (since the derivatives of h are bounded), that

$$\frac{\partial^2 f_h}{\partial x_i \partial x_j} = \int_0^1 \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 \bar{h}}{\partial x_j \partial x_i}(Z_{x,t})\right] dt, \tag{8}$$

where $Z_{x,t} = \sqrt{t}x + \sqrt{1-t}Z$. Performing two Gaussian integration by parts (which are valid because $\bar{h}$ and its derivatives have at most polynomial growth), we also have

$$\frac{\partial^2 f_h}{\partial x_i \partial x_j} = \int_0^1 \frac{1}{2(1-t)}\mathbb{E}[(Z_i Z_j - \delta_{ij})\bar{h}(Z_{x,t})]dt. \tag{9}$$

Define the α-Hölder semi-norm, for $\alpha \in (0,1]$, by

$$[h]_\alpha = \sup_{x\neq y}\frac{|h(x)-h(x)|}{|x-y|^\alpha}.$$

Let us state our main regularity result.

Proposition 2.1. *Let $h \in C_b^\infty(\mathbb{R}^d)$, and assume that h is α-Hölder for some $\alpha \in (0,1]$. Then the solution f_h of (7) satisfies for any $1 \leqslant i,j \leqslant d$*

$$\begin{aligned}\left|\frac{\partial^2 f_h}{\partial x_i \partial x_j}_{|x} - \frac{\partial^2 f_h}{\partial x_i \partial x_j}_{|y}\right| &\leqslant [h]_\alpha |x-y|^\alpha \left(C - 2\ln|x-y|\right), && \text{if } |x-y| \leqslant 1\\ &\leqslant C\,[h]_\alpha && \text{if } |x-y| > 1,\end{aligned}$$

where $C = 2^{\frac{\alpha}{2}} \frac{(\alpha+d+1)\,\Gamma(\frac{\alpha+d}{2})}{\alpha\,\Gamma(d/2)}$.

In particular, for all $\alpha > \epsilon > 0$, $\frac{\partial^2 f_h}{\partial x_i \partial x_j}$ is $(\alpha-\epsilon)$-Hölder :

$$\left|\frac{\partial^2 f_h}{\partial x_i \partial x_j}_{|x} - \frac{\partial^2 f_h}{\partial x_i \partial x_j}_{|y}\right| \leqslant (C + 2\epsilon^{-1})|x-y|^{\alpha-\epsilon}[h]_\alpha. \tag{10}$$

The sharpest estimates is the $(1+\ln)$ α-Hölder regularity

$$\left|\frac{\partial^2 f_h}{\partial x_i \partial x_j}_{|x} - \frac{\partial^2 f_h}{\partial x_i \partial x_j}_{|y}\right| \leqslant C|x-y|^\alpha \left(1 + |\ln|x-y||\right)[h]_\alpha. \tag{11}$$

Proof. Recall that

$$\frac{\partial^2 f_h}{\partial x_i \partial x_j} = \int_0^1 \frac{1}{2(1-t)} \mathbb{E}[(Z_i Z_j - \delta_{ij})\bar{h}(Z_{x,t})]dt.$$

Since $\mathbb{E}[Z_i Z_j - \delta_{ij}] = 0$, we have $\mathbb{E}[(Z_i Z_j - \delta_{ij})\bar{h}(\sqrt{t}x)] = 0$, so that

$$\mathbb{E}[(Z_i Z_j - \delta_{ij})\bar{h}(Z_{x,t})] = \mathbb{E}[(Z_i Z_j - \delta_{ij})(\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}x))].$$

Since $|\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}x)| \leqslant [h]_\alpha (1-t)^{\alpha/2}||Z||^\alpha$, we have

$$|\mathbb{E}[(Z_i Z_j - \delta_{ij})\bar{h}(Z_{x,t})]| \leqslant [h]_\alpha (1-t)^{\alpha/2}\mathbb{E}[|Z_i Z_j - \delta_{ij}|\ ||Z||^\alpha].$$

For all $\beta > 0$, $\mathbb{E}||Z||^\beta = \frac{2^{\frac{\beta}{2}}\Gamma(\frac{\beta+d}{2})}{\Gamma(d/2)}$. Thus,

$$\begin{aligned}\left|\frac{1}{2(1-t)}\mathbb{E}[(Z_i Z_j - \delta_{ij})\bar{h}(Z_{x,t})]\right| &\leqslant \frac{1}{2}\frac{2^{\frac{\alpha}{2}+1}\Gamma(\frac{\alpha+d}{2}+1) + 2^{\frac{\alpha}{2}}\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)}(1-t)^{-1+\alpha/2}[h]_\alpha\\ &= 2^{\frac{\alpha}{2}-1}(\alpha+d+1)\frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)}(1-t)^{-1+\alpha/2}[h]_\alpha.\end{aligned}$$

Let $C_1 = 2^{\frac{\alpha}{2}-1}(\alpha+d+1)\frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)}$. Now we cut the integral in two parts. Let $\epsilon \in [0,1]$. We have

$$\begin{aligned}&\left|\frac{\partial^2 f_h}{\partial x_i \partial x_j}_{|x} - \frac{\partial^2 f_h}{\partial x_i \partial x_j}_{|y}\right|\\ &\leqslant \int_0^{1-\epsilon}\frac{1}{2(1-t)}\mathbb{E}[|Z_i Z_j - \delta_{ij}|\ |\bar{h}(Z_{x,t}) - \bar{h}(Z_{y,t}|]dt + \int_{1-\epsilon}^1 \frac{1}{2(1-t)}\left|\mathbb{E}[(Z_i Z_j - \delta_{ij})(\bar{h}(Z_{x,t}) - \bar{h}(Z_{x,t}))]\right|dt\\ &\leqslant [h]_\alpha |x-y|^\alpha \mathbb{E}[|Z_i Z_j - \delta_{ij}|]\int_0^{1-\epsilon}\frac{t^{\alpha/2}}{2(1-t)}dt + C_1[h]_\alpha \int_{1-\epsilon}^1 (1-t)^{\alpha/2-1}dt\\ &\leqslant [h]_\alpha\left(-|x-y|^\alpha \ln\epsilon + \frac{2C_1}{\alpha}\epsilon^{\alpha/2}\right),\end{aligned}$$

Choose $\epsilon = |x-y|^2$ if $|x-y| \leqslant 1$, $\epsilon = 1$ otherwise to get the first result. Equation (11) is a straighforward reformulation since $1 + |\ln(u)| \geqslant 1$ when $u \geqslant 1$. To get (10), simply note that for $0 < u \leqslant 1$, $-\ln u \leqslant \epsilon^{-1}u^{-\epsilon}$ and for $u \geqslant 1$, $u \leqslant |x-y|^{\alpha-\epsilon}$.

□

3 Multivariate Berry-Essen bounds in Wasserstein distance

We apply the regularity results obtained in previous section to obtain Berry-Esseen bounds in the CLT, in 1-Wasserstein distance.

Let $X_1, X_2, ...$ be an i.i.d. sequence of centered, square-integrable and isotropic random vectors; that is, $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1 X_1^T] = I_d$, the identity matrix of size d. Let $h : \mathbb{R}^d \to \mathbb{R}$ be some test function. Let $W = n^{-1/2} \sum_{i=1}^n X_i$. We are interested in $\mathcal{W}(W, Z)$ where the 1-Wasserstein distance is defined as

$$\mathcal{W}(X, Y) = \sup_{[h]_1 \leqslant 1} \mathbb{E}h(X) - \mathbb{E}h(Y).$$

We begin by a Lemma showing that one can restricts the class of functions to smooth functions with bounded derivatives.

Lemma 3.1. *We have, for any X, Y,*

$$\mathcal{W}(X, Y) = \sup_{h \in \mathcal{C}_b^\infty(\mathbb{R}^d);\ \|\nabla h\| \leqslant 1} \mathbb{E}h(X) - \mathbb{E}h(Y). \tag{12}$$

Proof. It is clear that the right-hand side of (12) is smaller than the left-hand side. To prove the converse inequality, we use a classical smoothing argument. Let h be a 1-Lipschitz function. Let ω_ϵ be a smoothing kernel, and let $h_\epsilon = h * \omega_\epsilon$, the convolution between h and ω_ϵ. It is readily checked that

$$\begin{aligned} &||h_\epsilon - h||_\infty \leqslant C\epsilon \\ &||\nabla h_\epsilon||_\infty \leqslant 1, \end{aligned}$$

where $||.||_\infty$ stands for the supremum norm and C is a constant (depending only on the dimension). It is also easily checked that all derivatives of order greater than 2 of h_ϵ are bounded. Thus, if $\tilde{\mathcal{W}}(X, Y) = \sup_{h \in \mathcal{C}_b^\infty(\mathbb{R}^d);\ |\nabla h| \leqslant 1} \mathbb{E}h(X) - \mathbb{E}h(Y)$,

$$\begin{aligned} \mathbb{E}h(X) - \mathbb{E}h(Y) =& \mathbb{E}h_\epsilon(X) - \mathbb{E}h_\epsilon(Y) + \mathbb{E}h(X) - \mathbb{E}h_\epsilon(X) \\ &+ \mathbb{E}h_\epsilon(Y) - \mathbb{E}h(Y) \\ \leqslant& \tilde{\mathcal{W}}(X, Y) + 2C\epsilon. \end{aligned}$$

Taking the supremum over h and letting $\epsilon \to 0$ finishes the proof. □

We are now in position to prove our main Theorem.

Proof of Theorem 1.1. Let f_h be the solution of the Stein equation defined by (4). Then,

$$\begin{aligned} \mathbb{E}[h(W) - h(Z)] &= \mathbb{E}[\Delta f_h(W) - W \cdot \nabla f_h(W)] \\ &= \frac{1}{n} \sum_{i=1}^n \left[\mathbb{E}[\Delta f_h(W) - \sqrt{n} X_i \cdot \nabla f_h(W) \right]. \end{aligned}$$

Let $W_i = W - X_i/\sqrt{n} = \frac{1}{\sqrt{n}} \sum_{j \neq i} X_j$. By Taylor's formula, we have for some uniformly distributed in $[0, 1]$ (and independent of everything else) θ

$$\mathbb{E}[X_i \cdot \nabla f_h(W)] = \frac{1}{\sqrt{n}} \mathbb{E}\left[X_i^T \nabla^2 f_h \left(W_i + \theta \frac{X_i}{\sqrt{n}} \right) X_i \right],$$

leading to

$$\mathbb{E}[h(W) - h(Z)] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[\Delta f_h(W) - X_i^T \nabla^2 f_h \left(W_i + \theta \frac{X_i}{\sqrt{n}} \right) X_i \right].$$

Let $X_{i,j}$ be the jth coordinate of X_i. Note that

$$
\begin{aligned}
&\mathbb{E}[X_i^T \nabla^2 f_h \left(W_i + \theta \frac{X_i}{\sqrt{n}} \right) X_i] \\
=& \sum_{j,k=1}^{d} \mathbb{E}\left[X_{i,j} X_{i,k} \partial^2_{jk} f_h \left(W_i + \theta \frac{X_i}{\sqrt{n}} \right) \right] \\
=& \sum_{j,k=1}^{d} \mathbb{E}\left[X_{i,j} X_{i,k} \partial^2_{jk} f_h \left(W_i \right) \right] + \sum_{j,k=1}^{d} \mathbb{E}\left[X_{i,j} X_{i,k} \left(\partial^2_{jk} f_h \left(W_i + \theta \frac{X_i}{\sqrt{n}} \right) - \partial^2_{jk} f(W_i) \right) \right] \\
=& \sum_{j=1}^{d} \mathbb{E}\left[X_{i,j}^2 \partial^2_{jj} f_h \left(W_i \right) \right] + \sum_{j,k=1}^{d} \mathbb{E}\left[X_{i,j} X_{i,k} \left(\partial^2_{jk} f_h \left(W_i + \theta \frac{X_i}{\sqrt{n}} \right) - \partial^2_{jk} f(W_i) \right) \right] \\
=& \mathbb{E}\left[\Delta f_h \left(W_i \right) \right] + \sum_{j,k=1}^{d} \mathbb{E}\left[X_{i,j} X_{i,k} \left(\partial^2_{jk} f_h \left(W_i + \theta \frac{X_i}{\sqrt{n}} \right) - \partial^2_{jk} f(W_i) \right) \right].
\end{aligned}
$$

Finally,

$$
\mathbb{E}[h(W) - h(Z)] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\Delta f_h(W) - \Delta f_h(W_i) - \sum_{j,k=1}^{d} X_{i,j} X_{i,k} \left(\partial^2_{jk} f_h \left(W_i + \theta \frac{X_i}{\sqrt{n}} \right) - \partial^2_{jk} f(W_i) \right) \right]. \tag{13}
$$

Let $h \in \mathcal{C}_b^\infty(\mathbb{R}^d)$, such that $|\nabla h| \leqslant 1$, and $W = n^{-1/2} \sum_{i=1}^n X_i$. From (10), (13) and Lemma 3.1, we have

$$
\mathbb{E}[h(W) - h(Z)] \leqslant (C(d) + 2\epsilon^{-1}) \left[d \frac{\mathbb{E}|X_i|^{1-\epsilon}}{n^{\frac{1}{2} - \frac{\epsilon}{2}}} + d^2 \frac{\mathbb{E}|X_i|^{3-\epsilon}}{n^{\frac{1}{2} - \frac{\epsilon}{2}}} \right].
$$

Rearranging, we obtain the result. □

Remark 3.2. *From Stirling's formula, $C(d) \underset{d \to +\infty}{\sim} K d^{3/2}$, with K a universal constant. Overall, the bound in last proposition behaves as $d^{7/2}$ when $d \to +\infty$. This is sub-optimal : as noted by [5], the sharpest rate one can expect is $\mathcal{O}(\sqrt{d})$.*

Corollary 3.3. *Let $(X_i)_{i \geqslant 1}$ be an i.i.d. sequence of random vectors with unit covariance matrix. Assume that $\mathbb{E}[|X_i|^3] < \infty$. Then for $n \geqslant 3$,*

$$
\mathcal{W}\left(n^{-1/2} \sum_{i=1}^{n} X_i, Z \right) \leqslant \sqrt{e}\, d(1+d)\, \frac{C(d) + 2\ln n}{\sqrt{n}}\, \mathbb{E}|X_i|^3,
$$

where $C(d)$ is given in (6).

Proof. The bound of last Proposition holds for any $0 < \epsilon < 1$. By Hölder's and the Cauchy-Schwarz inequalities, $\mathbb{E}|X_i|^{1-\epsilon} \leqslant \left(\mathbb{E}|X_i|^3 \right)^{(1-\epsilon)/3}$. But by Jensen's inequality, $\mathbb{E}|X_i|^3 \geqslant (\mathbb{E}|X_i|^2)^{3/2} = d^{3/2} \geqslant 1$, so that, since $(1-\epsilon)/3 < 1$, $\left(\mathbb{E}|X_i|^3 \right)^{(1-\epsilon)/3} \leqslant \mathbb{E}|X_i|^3$.

Similarly, $\mathbb{E}|X_i|^{3-\epsilon} \leqslant \left(\mathbb{E}|X_i|^3 \right)^{1-\epsilon/3} \leqslant \mathbb{E}|X_i|^3$. Thus we have for all $0 < \epsilon < 1$,

$$
\mathcal{W}\left(n^{-1/2} \sum_{i=1}^{n} X_i, Z \right) \leqslant d(1+d)\, \frac{C(d) + 2\epsilon^{-1}}{n^{\frac{1}{2} - \frac{\epsilon}{2}}}\, \mathbb{E}|X_i|^3.
$$

Choosing $\epsilon = 1/\ln n$ achieves the proof since $n^{-\frac{1}{2\ln n}} = 1/\sqrt{e}$. □

Remark that Lemma 3.1 can be readily extented to $\mathcal{W}_\alpha$ spaces defined by

$$
\mathcal{W}_\alpha(X, Y) = \sup_{[h]_\alpha \leqslant 1} \mathbb{E}h(X) - \mathbb{E}h(Y).
$$

Acknowledgements

We would like to thank Guillaume Carlier who suggested this problem to us.

References

[1] A. D. Barbour. Stein's method for diffusion approximations. *Probability theory and related fields*, 84(3):297–322, 1990.

[2] A. Bikjalis. Estimates of the remainder term in the central limit theorem. *Litovsk. Mat. Sb*, 6:323–346, 1966.

[3] T. Bonis. Rates in the central limit theorem and diffusion approximation via stein's method. 06 2015.

[4] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchange-able pairs. *Alea*, 4:257–283, 2008.

[5] T. A. Courtade, M. Fathi, and A. Pananjady. Existence of stein kernels under a spectral gap, and discrepancy bound. *arXiv preprint arXiv:1703.07707*, 2017.

[6] R. E. Gaunt. Rates of convergence in normal approximation under moment conditions via new bounds on solutions of the stein equation. *Journal of Theoretical Probability*, 29(1):231–247, 2016.

[7] L. Goldstein and Y. Rinott. Multivariate normal approximations by stein's method and size bias couplings. *Journal of Applied Probability*, 33(01):1–17, 1996.

[8] Y. S. Nefedova and I. G. Shevtsova. On nonuniform convergence rate estimates in the central limit theorem. *Theory of Probability & Its Applications*, 57(1):28–59, 2013.

[9] G. Reinert and A. Röllin. Multivariate normal approximation with stein's method of exchangeable pairs under a general linearity condition. *The Annals of Probability*, 37(6):2150–2173, 2009.

[10] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. The Regents of the University of California, 1972.

[11] G. Valiant and P. Valiant. A clt and tight lower bounds for estimating entropy. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 17, pages 1–1, 2010.

[12] A. Zhai. A multivariate clt in wasserstein distance with near optimal convergence rate. *arXiv preprint arXiv:1602.05565*, 2016.