Electron. Commun. Probab. **25** (2020), no. 27, 1–13.
https://doi.org/10.1214/20-ECP308
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Donsker's theorem in Wasserstein-1 distance

Laure Coutin* Laurent Decreusefond†

Abstract

We compute the Wassertein-1 (or Kantorovitch-Rubinstein) distance between a random walk in $\mathbf{R}^d$ and the Brownian motion. The proof is based on a new estimate of the modulus of continuity of the solution of the Stein's equation. As an application, we can evaluate the rate of convergence towards the local time at 0 of the Brownian motion and to a Brownian bridge.

Keywords: Donsker theorem; Malliavin calculus; Stein's method; Wasserstein distance.
AMS MSC 2010: Primary 60F15, Secondary 60H07; 60G15; 60G55.
Submitted to ECP on August 16, 2019, final version accepted on March 11, 2020.
Supersedes arXiv:1904.07045.
Supersedes HAL:hal-02098892.

1 Motivations

For a complete, separable metric space W, the so-called Kantorovitch-Rubinstein or Wasserstein-1 distance is defined by:

$$\operatorname{dist}_{\text{KR}}(\mu,\nu) = \sup_{f\in \operatorname{Lip}_1(W)} \left(\int_W f \, \mathrm{d}\mu - \int_W f \, \mathrm{d}\nu \right) \tag{1.1}$$

where

$$\operatorname{Lip}_1(W) = \{ f \,:\, W \to \mathbf{R}, f \text{ bounded and } |f(x)-f(y)| \le \operatorname{dist}_W(x,y), \ \forall x,y \in W \}.$$

The formulation (1.1) is well suited to evaluate distance by the Stein's method. First, consider that we are interested in the convergence towards the standard Gaussian measure μ_d on $\mathbf{R}^d$. It is well known that the semi-group $(P_t, t \ge 0)$ defined by

$$P_t f \,:\, x \in \mathbf{R}^d \longmapsto \int_{\mathbf{R}^d} f(e^{-t}x + \sqrt{1-e^{-2t}}\, y) \, \mathrm{d}\mu_d(y),$$

admits μ_d as invariant and stationary measure. Let $(X_n,\, n \ge 1)$ be a sequence of independent, centered, square integrable, and identically distributed $\mathbf{R}^d$-valued random

*Institute of Mathematics, Université Toulouse 3, Toulouse, France.
E-mail: laure.coutin@math.univ-toulouse.fr
†LTCI, Télécom Paris, Institut polytechnique de Paris, Paris, France.
E-mail: Laurent.Decreusefond@mines-telecom.fr

variables such that $\mathbf{E}\left[X_1(i)X_1(j)\right] = \mathbf{1}_{\{i=j\}}$. Let $T_n = n^{-1/2}\sum_{j=1}^n X_j$ and $T_n^{\neg j} = T_n - X_j/\sqrt{n}$. Standard computations [5] yield, for f regular enough,

$$\mathbf{E}\left[f(T_n)\right] - \int_{\mathbf{R}} f \, \mathrm{d}\mu_d = \frac{1}{n}\int_0^\infty \mathbf{E}\left[\operatorname{trace}(\nabla^{(2)}P_t f)(T_n^{\neg j}) - \operatorname{trace}(\nabla^{(2)}P_t f)(T_n)\right] \, \mathrm{d}t$$
$$-\frac{1}{n}\sum_{j=1}^n \int_0^\infty\int_0^1 \mathbf{E}\left[\langle X_j \otimes X_j,\ \left(\nabla^{(2)}(P_t f)(T_n^{\neg j} + rX_j/\sqrt{n}) - \nabla^{(2)}(P_t f)(T_n^{\neg j})\right)\rangle_{\mathbf{R}^d}\right] \, \mathrm{d}r \, \mathrm{d}t. \tag{1.2}$$

The crux of the matter is then to find the regularity of f with which we can estimate uniformly the Lipschitz modulus of $\nabla^{(2)}P_t f$. If we can prove that $\nabla^{(2)}P_t f$ is uniformly Lipschitz, we see that a factor $1/\sqrt{n}$ will pop-up in the right-hand-side of (1.2). Multiplied by the already present $1/n$, this gives a $n^{-3/2}$ multiplicative factor, which is added n times hence a rate of convergence which is proportional to $n^{-1/2}$. The estimate of the Lipschitz modulus of $\nabla^{(2)}P_t f$ for f Lipschitz continuous is easily done when $W = \mathbf{R}$ but it is only recently that the case $W = \mathbf{R}^d$ has been successfully handled (see [6, 8, 11] and references therein). Now, if we are interested in the Donsker theorem, the process under study is

$$S_n(t) = \sum_{j=1}^n X_j h_j^n(t) \text{ where } h_j^n(t) = \sqrt{n}\int_0^t \mathbf{1}_{[j/n,(j+1)/n)}(s) \, \mathrm{d}s.$$

In his seminal paper, Barbour [2] gave a convergence rate $n^{-1/2}\ln n$ when W is the Skorohod space and the test functions are not Lipschitz but three times Fréchet differentiable and satisfy some additional regularity assumptions. Recently, Kasprzak used this approach to estimate the convergence rate of the diffusion approximation of the Moran model [10]. In [3], we proved that the convergence rate for the Donsker theorem in $W = W_{\eta,2}$ (see below for the definition) for some $\eta < 1/2$ is bounded by $n^{\eta-1/2}$ for test functions which are twice Fréchet differentiable. Unfortunately, Fréchet differentiability is too stringent a condition for practical purposes. In [4], we weakened this hypothesis by considering only weak differentiability. In fact, for reasons that will be explained below, when working in a Wiener space, the analog of (1.2) involves terms like

$$\left\langle h_j^n \otimes h_j^n, \nabla^{(2)}(P_t f)(S_n) - \nabla^{(2)}(P_t f)(S_n - X_j h_j^n)\right\rangle_{I_{1,2}^{\otimes 2}} \tag{1.3}$$

where ∇ is the Malliavin derivative and $I_{1,2}$ is the Cameron-Martin space

$$I_{1,2} = \left\{ f, \exists ! \dot{f} \in L^2([0,1], \, \mathrm{d}t) \text{ with } f(t) = \int_0^t \dot{f}(s) \, \mathrm{d}s \right\} \text{ and } \|f\|_{I_{1,2}} = \|\dot{f}\|_{L^2}.$$

The difficulty is then that we do not have a $n^{-1/2}$ factor in the definition of S_n and it is easily seen that $\|h_j^n\|_{I_{1,2}} = 1$, hence no multiplicative factor will pop up in (1.3). In [4], we bypassed this difficulty by assuming that there exists $c > 0$ such that for any $g, h, \ell \in I_{1,2}$,

$$\sup_{x\in W}\left|\left\langle \nabla^{(2)}P_t f(x+\ell) - \nabla^2 P_t f(x),\ h \otimes g\right\rangle_{I_{1,2}^{\otimes(2)}}\right| \le c\|\ell\|_W \|h\|_{L^2}\|g\|_{L^2}. \tag{1.4}$$

Then, in the estimate of terms as those appearing in (1.3), it is the L^2-norm of h_j^n which appears and it turns out that $\|h_j^n\|_{L^2} \le c\, n^{-1/2}$, hence the presence of a factor n^{-1}, which saves the proof and yields a convergence rate bounded by $n^{-1/2}$. The goal of this paper is to weaken even more these hypothesis on f to be able to bound the true K-R distance between the distribution of S_n and the distribution of a Brownian motion. The space W is a Banach space we can choose arbitrarily as far as it can be equipped with the structure of an abstract Wiener space and it contains the sample paths of S_n and B. As in [4], we chose here the fractional Sobolev spaces $W_{\eta,p}$.

The main technical result of this article is Theorem 4.4 which gives a new estimate of the Lipschitz modulus of $\nabla^{(2)} P_t f$ for $t > 0$. The main idea is to introduce a hierarchy of approximations. There is a first scale induced by the time discretization coming from the definition of S_n. Then, we consider a coarser discretization onto which we project our approximations in order to benefit from the averaging effect of the ordinary CLT. It turns out that the optimal ratio is obtained when the mesh of the coarser subdivision is roughly the cubic root of the mesh of the reference partition. Moreover, after [3] and [4], we are convinced that it is simpler and as efficient to stick to finite dimension as long as possible. For, we consider the affine interpolation of the Brownian motion as an intermediary process. The distance between the Brownian sample-paths and their affine interpolation is well known. This reduces the problem to estimate the distance between S_n and the affine interpolation of B, a task which can be handled by the Stein's method. It turns out that the bottleneck is in fact the rate of convergence of the Brownian interpolation to the Brownian motion.

This paper is organized as follows. In Section 2, we show how to view fractional Sobolev spaces as Wiener spaces. In Section 3, we explain the line of thoughts we used. The main proofs are given in Section 4.

2 Preliminaries

2.1 Fractional Sobolev spaces

Let $d \geq 1$ be a fixed integer. We consider the fractional Sobolev spaces $W_{\eta,p}$ defined for $\eta \in (0,1)$ and $p \geq 1$ as the closure of $\mathcal{C}^1$ functions from $[0,1]$ to $\mathbf{R}$ with respect to the norm

$$\|f\|_{\eta,p}^p = \int_0^1 |f(t)|^p \,\mathrm{d}t + \iint_{[0,1]^2} \frac{|f(t)-f(s)|^p}{|t-s|^{1+p\eta}} \,\mathrm{d}t \,\mathrm{d}s.$$

For $\eta = 1$, $W_{1,p}$ is the completion of $\mathcal{C}^1$ for the norm:

$$\|f\|_{1,p}^p = \int_0^1 |f(t)|^p \,\mathrm{d}t + \int_0^1 |f'(t)|^p \,\mathrm{d}t.$$

They are known to be separable Banach spaces and to satisfy the Sobolev embeddings [1]:

$$W_{\eta,p} \subset \mathrm{Hol}(\eta - 1/p) \text{ for } \eta - 1/p > 0,$$

where $\mathrm{Hol}(\alpha)$ is the space of $\mathbf{R}^d$-valued, Hölder continuous functions on $[0,1]$. Recall the following lemma proved in [4]:

Lemma 2.1. *Let $0 \leq s_1 < s_2 \leq 1$ and consider*

$$h_{s_1,s_2}(t) = \int_0^t \mathbf{1}_{[s_1,s_2]}(r) \; dr.$$

For any $\eta \in [0,1/2)$, there exists $c_\eta > 0$ such that for any $p \geq 1$, for any s_1, s_2, we have

$$\|h_{s_1,s_2}\|_{\eta,p} \leq c_\eta \, |s_2 - s_1|^{1/2-\eta}. \tag{2.1}$$

We denote by $W_{0,\infty}$ the space of continuous (hence bounded) functions on $[0,1]$ equipped with the uniform norm.

2.2 Wiener spaces

In what follows, a couple (η, p) in Λ defined by:

$$\Lambda = \{(\eta, p) \in (0,1] \times [1, \infty), 0 < \eta - 1/p < 1/2\} \cup \{(0, \infty)\},$$

is fixed. Consider $Z = (Z_n,\, n \geq 1)$ a sequence of independent, standard Gaussian random variables and let $(z_n,\, n \geq 1)$ be a complete orthonormal basis of $I_{1,2}$. Then, we know from [9] that

$$\sum_{n=1}^{N} Z_n\, z_n \xrightarrow{N\to\infty} B := \sum_{n=1}^{\infty} Z_n\, z_n \text{ in } W_{\eta,p} \text{ with probability } 1, \tag{2.2}$$

where B is a Brownian motion. By taking d independent copies of the sequence Z, we can construct the Wiener measure on $W = W_{\eta,p}([0,1]; \mathbf{R}^d)$, the space of functions from $[0,1]$ into $\mathbf{R}^d$, each component of which belongs to $W_{\eta,p}$. We clearly have the diagram

$$W^* \xrightarrow{\mathfrak{e}^*} (I_{1,2}^{\otimes d})^* \simeq I_{1,2}^{\otimes d} = H \xrightarrow{\mathfrak{e}=\mathfrak{e}_{\eta,p}^{\otimes d}} W, \tag{2.3}$$

where $\mathfrak{e}_{\eta,p}$ is the embedding from $I_{1,2}$ into $W_{\eta,p}$. Note that the space H is dense in W since tensor products of polynomials do belong to H. Moreover, Eqn. (2.2) and the Parseval identity entail that for any $z \in W^*$,

$$\mathbf{E}\left[e^{i\langle z,B\rangle_{W^*,W}}\right] = \exp\left(-\frac{1}{2}\|\mathfrak{e}^*(z)\|_H^2\right). \tag{2.4}$$

We denote by μ the law of B on W. Then, the diagram (2.3) and the identity (2.4) mean that $(I_{1,2}, W, \mathfrak{e})$ is a Wiener space.

Definition 2.2 (Wiener integral)**.** *The Wiener integral, denoted as δ, is the isometric extension of the map*

$$\begin{aligned} \delta\,:\, \mathfrak{e}^*(W^*) \subset H_{1,2} &\longrightarrow L^2(\mu) \\ \mathfrak{e}^*(\eta) &\longmapsto \langle \eta,\, y\rangle_{W^*,W}\,. \end{aligned}$$

This means that if $h = \lim_{n\to\infty} \mathfrak{e}^*(\eta_n)$ in H,

$$\delta h(y) = \lim_{n\to\infty} \langle \eta_n,\, y\rangle_{W^*,W} \text{ in } L^2(\mu).$$

Definition 2.3. *Let V be a Banach space. A function $f\,:\, W \to V$ is said to be cylindrical if it is of the form*

$$f(y) = \sum_{j=1}^{k} f_j(\delta h_1(y), \cdots, \delta h_k(y))\, x_j$$

where for any $j \in \{1, \cdots, k\}$, f_j belongs to the Schwartz space on $\mathbf{R}^k$, $(h_1, \cdots, h_k)$ are elements of H and $(x_1, \cdots, x_k)$ belong to V. The set of such functions is denoted by $\mathfrak{C}(V)$. The gradient of f is then given by

$$\nabla f = \sum_{j,l=1}^{k} \partial_j f(\delta h_1(y), \cdots, \delta h_k(y))\, h_l \otimes\, x_j.$$

The space $\mathbb{D}_{1,2}$ is the closure of $\mathfrak{C}(V)$ with respect to the norm of $L^2(W; H \otimes V)$. This construction can be iterated so that we can define higher order gradients (see [14] for details).

3 Donsker's theorem in $W_{\eta,p}(\mathbf{R}^d)$

For $m \ge 1$, let $\mathcal{D}^m = \{i/m,\, i = 0, \cdots, m\}$, the regular subdivision of the interval $[0,1]$. Let

$$\mathcal{A}^m = \{1, \cdots, d\} \times \{0, \cdots, m-1\}$$

and for $a = (a_1,\, a_2) \in \mathcal{A}^m$

$$h_a^m(t) = \sqrt{m} \int_0^t \mathbf{1}_{[a_2/m,(a_2+1)/m)}(s)\, \mathrm{d}s\; e_{a_1},$$

where $(e_l,\, 1 \le l \le d)$ is the canonical basis of $\mathbf{R}^d$. Consider

$$S^m = \sum_{a \in \mathcal{A}^m} X_a\, h_a^m$$

where $(X_a,\, a \in \mathcal{A}^m)$ is a family of independent identically distributed, $\mathbf{R}^d$-valued, random variables. We denote by X a random variable which has their common distribution. Moreover, we assume that $\mathbf{E}[X] = 0$ and $\mathbf{E}\left[\|X\|^2_{\mathbf{R}^d}\right] = 1$. Remark that $(h_a^m, a \in \mathcal{A}^m)$ is an orthonormal family in H. Let $\mathcal{V}^m = \operatorname{span}(h_a^m,\, a \in \mathcal{A}^m) \subset H$. For any $m > 0$, the map π^m is the orthogonal projection from H onto $\mathcal{V}^m$. Let $0 < N < m$, for $f \in \operatorname{Lip}_1(W)$, we write

$$\begin{aligned}\mathbf{E}[f(S^m)] - \mathbf{E}[f(B)] &= \Big(\mathbf{E}[f(S^m)] - \mathbf{E}\left[f(\pi^N(S^m))\right]\Big)\\ &+ \Big(\mathbf{E}\left[f\circ\pi^N(S^m)\right] - \mathbf{E}\left[f\circ\pi^N(B^m)\right]\Big) + \Big(\mathbf{E}\left[f\circ\pi^N(B^m)\right] - \mathbf{E}[f(B)]\Big) = \sum_{i=1}^3 A_i, \quad (3.1)\end{aligned}$$

where B^m is the affine interpolation of the Brownian motion:

$$B^m(t) = \sum_{a \in \mathcal{A}^m} \sqrt{m}\, \Big(B_{a_1}(\frac{a_2+1}{m}) - B_{a_1}(\frac{a_2}{m})\Big)\, h_a^m(t).$$

The two terms A_1 and A_3 are of the same nature: We have to compare two processes which live on the same probability space. Since f is Lipschitz, we can proceed by comparison of their sample-paths. The term A_2 is different as the two processes involved live on different probability spaces. This is for this term that the Stein's method will be used. We know from [7] that

Theorem 3.1. *We have the following inequality:*

$$\sup_N N^{1/2-\eta}\, \mathbf{E}\left[\|B^N - B\|_W^p\right]^{1/p} < \infty. \quad (3.2)$$

The following upper-bound is far from being optimal and it is likely that it could be improved to obtain a factor $N^{1-\eta}$. However, in view of (3.2), it would bring no improvement to our final result.

Theorem 3.2. *There exists a constant $c > 0$ such that provided that $X \in L^p(W; \mathbf{R}^d, \mu)$,*

$$\sup_{m,N} N^{\frac{1}{2}-\eta}\, \mathbf{E}\left[\|S^m - \pi^N(S^m)\|_W^p\right]^{1/p} \le c\, \|X\|_{L^p}.$$

The main technical result is the following theorem.

Theorem 3.3. *There exists $c > 0$ such that provided that $X \in L^p(W; \mathbf{R}^d, \mu)$ with $p \ge 3$, for any $f \in \operatorname{Lip}_1(W)$,*

$$\mathbf{E}\left[f(\pi^N(S^m))\right] - \mathbf{E}\left[f(\pi^N(B^m))\right] \le c\, \|X\|_{L^p}\, \frac{N^{1+\eta}}{\sqrt{m}} \ln(\frac{N^{1+\eta}}{\sqrt{m}}). \quad (3.3)$$

Combining these three theorems, the global upper-bound for (3.1) appears to be proportional to $N^\eta(1/\sqrt{N}+\sqrt{N/m}\ln(N^{1+\eta}/\sqrt{m}))$. See N as a function of m and note that this expression is minimal for $N\sim m^{1/3}$. Plug this into the previous expressions to obtain the main result of this paper:

Theorem 3.4. *There exists a constant $c>0$ such provided that $X\in L^p(W;\mathbf{R}^d,\mu)$ with $p\geq 3$,*

$$\sup_{f\in \operatorname{Lip}_1(W_{\eta,p})} \mathbf{E}\left[f(S^m)\right]-\mathbf{E}\left[f(B)\right]\leq c\,\|X\|_{L^p}^p\ m^{-\frac{1}{6}+\frac{\eta}{3}}\ \ln m. \tag{3.4}$$

We now give two corollaries. These results are not accessible via the standard Stein's method since we do not know any characterization of the limit distributions.

Corollary 3.5. *Let $W=W_{0,\infty}$, the space of continuous functions. Denote by L^0, the local time at 0 of the one dimensional standard Brownian motion and let $M^m(t)=\sup_{s\leq t}(S^m(s)^-)$. We have*

$$\operatorname{dist}_{\text{KR}}\Big(M^m,L^0\Big)\leq c\,m^{-1/6}\ln m.$$

Proof. The map $\Theta\ :\ W\longrightarrow W$, which sends a function f to the function $(t\longmapsto \sup_{s\leq t} f(s)^-)$ is 1-Lipschitz continuous. Moreover, L^0 has the distribution of $\Theta(B)$, hence the result. □

The number of customers in an M/GI/1 queue follows the recurrence equation

$$Z_0=0,\ Z_{n+1}=(Z_n-1)^+ + A_{n+1}$$

where $(A_n,\,n\geq 1)$ is a sequence of i.i.d. random variables which represent the number of arrivals during a service. Let $\rho=\mathbf{E}\left[A_1\right]$ and consider the sequences of processes:

$$Z^m(t)=\sum_{j=1}^m (Z_j-Z_{j-1})\ h_j^m(t) \text{ and } U^m(t)=\sum_{j=1}^m (A_j-\rho)\ h_j^m(t)+(\rho-1)\sum_{j=1}^m\ h_j^m(t).$$

Solving the Skorohod reflection problem (see [12]), we know that

$$Z^m=U^m+\Theta(U^m).$$

Corollary 3.6. *Let $\sigma^2=\text{var}(A_1)$. Then,*

$$\operatorname{dist}_{\text{KR}}(\frac{Z^m}{\sigma\sqrt{m}},\,B_\rho+\Theta(B_\rho))\leq c\,m^{-1/6}\ln m,$$

where $B_\rho(t)=B(t)+(\rho-1)/\sigma\,t$.

Proof. Remark that $\sum_{j=1}^m\ h_j^m(t)=\sqrt{m}\,t$ and apply the previous corollary. □

Corollary 3.7. *Let B^0 be the Brownian bridge, i.e. the Brownian motion conditioned to be 0 at time 1. Let $\Psi\ :\ W_{\eta,p}\longrightarrow W_{\eta,p}$ be the linear map which sends a function f to the function $(t\longmapsto f(t)-tf(1))$. Then,*

$$\operatorname{dist}_{\text{KR}}\Big(\Psi(S^m),\,B^0\Big)\leq c\|\Psi\|\,m^{-1/6+\eta/3},$$

where $\|\Psi\|$ is the norm of Ψ as a continuous linear map.

Proof. Since Ψ is continuous linear map, it is Lipschitz continuous from $W_{\eta,p}$ into itself and its Lipschitz modulus coincides with its norm as a linear map. □

4 Proofs

In what follows, c denotes a non significant constant which may vary from line to line. We borrow from the current usage in rough path theory the notation $f_{s,t} = f(t) - f(s)$. As a preparation to the proof of Theorem 3.2, we need the following lemma.

Lemma 4.1. *For all $p \geq 2$, there exists a constant c_p such that for any sequence of independent, centered, identically distributed, $\mathbf{R}$-valued random variables $(X_i, i \in \mathbf{N})$ with $X \in L^p$ and any sequence $(\alpha_i,\, i \in \mathbf{N})$.*

$$\mathbf{E}\left[\left|\sum_{i=1}^{n} \alpha_i X_i\right|^p\right] \leq c_p \Big|\{i \leq n, \alpha_i \neq 0\}\Big|^{p/2} (\sum_{i \leq n} |\alpha_i|^p)\ \mathbb{E}(|X|^p),$$

where $|A|$ is the cardinality of the set A.

Proof. The Marcinkiewicz–Zygmund inequality yields

$$\mathbf{E}\left[\left|\sum_{i=1}^{n} \alpha_i X_i\right|^p\right] \leq c_p \mathbf{E}\left[\left|\sum_{i=1}^{n} \alpha_i^2 X_i^2\right|^{p/2}\right].$$

Using Jensen inequality, we obtain

$$\mathbf{E}\left[\left|\sum_{i=1}^{n} \alpha_i X_i\right|^p\right] \leq c_p \Big|\{i \leq n, \alpha_i \neq 0\}\Big|^{p/2-1}\ \mathbf{E}\left[\sum_{i=1}^{n} |\alpha_i|^p |X_i|^p\right].$$

The proof is thus complete. □

Proof of Theorem 3.2. Actually, we already proved in [4] that

$$\mathbf{E}\left[\|S^m_{s,t}\|^p\right] \leq c\|X\|_{L^p}\ \left(\sqrt{t-s} \wedge m^{-1/2}\right). \tag{4.1}$$

Assume that s and t belong to the same sub-interval: There exists $l \in \{1, ..., N\}$ such that

$$\frac{l-1}{N} \leq s < t \leq \frac{l}{N}.$$

Then we have

$$\pi^N(S^m)_{s,t} = \sqrt{N}\ \left(\sum_{k=1}^{m} X_k\, (h_k^m, h_l^N)_H\right)\ (t-s).$$

Using Lemma 4.1, there exists a constant c such that

$$\frac{\|\pi^N(S^m)_{s,t}\|_{L^p}}{\sqrt{N}\,|t-s|} \leq c\,\|X\|_{L^p}\ \left|\{k,\ (h_k^m, h_l^N)_H \neq 0\}\right|^{1/2} \sup_k \left|(h_k^m, h_l^N)_H\right|.$$

Note that $|(h_k^m, h_l^N)_H| \leq \sqrt{\frac{N}{m}}$ and there are at most $\frac{m}{N} + 2$ terms such that $(h_k^m, h_N^l)_H$ is non zero. Thus,

$$\frac{\|\pi^N(S^m)_{s,t}\|_{L^p}}{\sqrt{N}\,|t-s|} \leq c\,\|X\|_{L^p}\ \left(\frac{m}{N} + 2\right)^{1/2} \sqrt{\frac{N}{m}} \leq c\,\|X\|_{L^p},$$

as m/N tends to infinity. Since $|t-s| \leq 1/N$,

$$\|\pi^N(S^m)_{s,t}\|_{L^p} \leq c\,\|X\|_{L^p}\ \sqrt{|t-s|}. \tag{4.2}$$

For $0 \le s \le t \le 1$ let $s_+^N := \min\{l,\, s \le \frac{l}{N}\}$ and $t_-^N := \sup\{l,\, t \ge \frac{l}{N}\}$. We have

$$\pi^N(S^m)_{s,t} - S^m_{s,t} = \left(\pi^N(S^m)_{s,s_+^N} - S^m_{s,s_+^N}\right) \\ + \left(\pi^N(S^m)_{s_+^N,t_-^N} - S^m_{s_+^N,t_-^N}\right) + \left(\pi^N(S^m)_{t_-^N,t} - S^m_{t_-^N,t}\right).$$

Note that for all $f \in W$, $\pi^N(f)$ is the linear interpolation of f along the subdivision $\mathcal{D}_N$; hence, for $s,t \in \mathcal{D}_N$, $\pi^N(S^m)_{s,t} = S^m_{s,t}$. Thus the median term vanishes and we obtain

$$\mathbf{E}\left[\|\pi^N(S^m)_{s,t} - S^m_{s,t}\|^p\right] \le c\Big(\mathbf{E}\left[\|\pi^N(S^m)_{s,s_+^N}\|^p\right] + \mathbf{E}\left[\|S^m_{s,s_+^N}\|^p\right] \\ + \mathbf{E}\left[\|\pi^N(S^m)_{t_-^N,t}\|^p\right] + \mathbf{E}\left[\|S^m_{t_-^N,t}\|^p\right]\Big). \quad (4.3)$$

From (4.2), we deduce that

$$\mathbf{E}\left[\|\pi^N(S^m)_{s,s_+^N}\|^p\right]^{1/p} \le c\,\|X\|_{L^p}\,\sqrt{s_+^N - s} \le c\,\|X\|_{L^p}\,N^{-1/2}, \quad (4.4)$$

and the same holds for $\mathbf{E}\left[\|\pi^N(S^m)_{t_-^N,t}\|^p\right]$. We infer from (4.1), (4.2) and (4.4) that

$$\mathbf{E}\left[\|\pi^N(S^m)_{s,t} - S^m_{s,t}\|^p\right]^{1/p} \le c\,\|X\|_{L^p}\left(\sqrt{t-s} \wedge N^{-1/2}\right). \quad (4.5)$$

A straightforward computation shows that

$$\iint_{[0,1]^2} \frac{[|t-s| \wedge N^{-1}]^{p/2}}{|t-s|^{1+\eta p}}\,\mathrm{d}s\,\mathrm{d}t \le c\,N^{-p(1/2-\eta)}. \quad (4.6)$$

The result follows (4.5) and (4.6). □

4.1 Stein method

We wish to estimate $\mathbf{E}\left[f(\pi^N(S^m))\right] - \mathbf{E}\left[f(\pi^N(B^m))\right]$, using the Stein's method. For the sake of simplicity, we set $f_N = f \circ \pi^N$.

The Stein-Dirichlet representation formula [5] stands that, for any $\tau > 0$,

$$\mathbf{E}\left[f_N(B^m)\right] - \mathbf{E}\left[f_N(S^m)\right] = \mathbf{E}\left[\int_0^\infty \frac{d}{du} P_u f_N(S^m)\,\mathrm{d}u\right] \\ = \mathbf{E}\left[P_\tau f_N(S^m) - f_N(S^m)\right] + \mathbf{E}\left[\int_\tau^\infty LP_u f_N(S^m)\,\mathrm{d}u\right],$$

where for $g : \mathcal{V}^m \to \mathbf{R}$ regular enough, for $x \in \mathcal{V}^m \subset H$,

$$Lg(x) = -\langle x, \nabla g(x)\rangle_H + \sum_{a \in \mathcal{A}^m} \left\langle \nabla^{(2)} g(x),\, h_a^m \otimes h_a^m \right\rangle_{H^{\otimes 2}}.$$

It is straightforward (see [4, Lemma 4.1]):

Lemma 4.2. *There exists a constant $c > 0$ such that for any $\tau > 0$, for any sequence of independent, centered random vectors $(X_a,\, a \in \mathcal{A}^m)$ such that $\mathbf{E}\left[\|X\|^p\right] < \infty$, for any $g \in \mathrm{Lip}_1(W)$, we have*

$$\mathbf{E}\left[g(S^m)\right] - \mathbf{E}\left[P_\tau g(S^m)\right] \le c\,\|X\|_{L^p}\,\sqrt{1 - e^{-\tau}}.$$

We now show, that as usual, the rate of convergence in the Stein's method is related to the Lipschitz modulus of the second order derivative of the solution of the Stein's equation. Namely, we have

Lemma 4.3. *For any $f \in \mathrm{Lip}_1(W)$, we have*

$$\mathbf{E}\left[LP_\tau f_N(S^m)\right] = -\,\mathbf{E}\left[\sum_{a\in\mathcal{A}^m}\left\langle \nabla^{(2)}P_\tau f_N(S^m_{\neg a}) - \nabla^{(2)}P_\tau f_N(S^m),\, h^m_a\otimes h^m_a\right\rangle_{H^{\otimes 2}}\right]$$
$$+\ \mathbf{E}\left[\sum_{a\in\mathcal{A}} X_a^2\int_0^1 \left\langle \nabla^{(2)}P_\tau f_N(S^m_{\neg a}+rX_a h^m_a) - \nabla^{(2)}P_\tau f_N(S^m_{\neg a}),\, h^{m\otimes 2}_a\right\rangle_{H^{\otimes 2}} dr\right].$$

Proof of Lemma 4.3. Let $S^m_{\neg a} = S^m - X_a h^m_a$. Since the X_a's are independent,

$$\begin{aligned}
\mathbf{E}\left[\langle \nabla P_\tau f_N,\, S^m\rangle_H\right] &= \mathbf{E}\left[\sum_{a\in\mathcal{A}^m} X_a\ \langle \nabla P_\tau f_N(S^m), h^m_a\rangle_H\right]\\
&= \mathbf{E}\left[\sum_{a\in\mathcal{A}^m} X_a\ \langle \nabla P_\tau f_N(S^m) - \nabla P_\tau f_N(S^m_{\neg a}),\, h^m_a\rangle_H\right]\\
&= \mathbf{E}\left[\sum_{a\in\mathcal{A}^m} X_a^2\ \left\langle \nabla^{(2)}P_\tau f_N(S^m_{\neg a}),\, h^m_a\otimes h^m_a\right\rangle_{H^{\otimes 2}}\right]\\
&+\ \mathbf{E}\left[\sum_{a\in\mathcal{A}} X_a^2\int_0^1 \left\langle \nabla^{(2)}P_\tau f_N(S^m_{\neg a}+r\,X_a h^m_a) - \nabla^{(2)}P_\tau f_N(S^m_{\neg a}), h^{m\otimes 2}_a\right\rangle_{H^{\otimes 2}} \mathrm{d}r\right],
\end{aligned}$$

according to the Taylor formula. Since $\mathbf{E}\left[X_a^2\right] = 1$, we have

$$\begin{aligned}
&\mathbf{E}\left[\sum_{a\in\mathcal{A}^m} X_a^2\ \left\langle \nabla^{(2)}P_\tau f_N(S^m_{\neg a}),\, h^m_a\otimes h^m_a\right\rangle_{H^{\otimes 2}}\right]\\
&\quad= \mathbf{E}\left[\sum_{a\in\mathcal{A}^m} \left\langle \nabla^{(2)}P_\tau f_N(S^m_{\neg a}),\, h^m_a\otimes h^m_a\right\rangle_{H^{\otimes 2}}\right].
\end{aligned}$$

The result follows by difference. □

The main difficulty and then the main contribution of this paper is to find an estimate of

$$\sup_{v\in\mathcal{V}^m} \left\langle \nabla^{(2)}P_\tau f_N(v) - \nabla^{(2)}P_\tau f_N(v+\varepsilon h^m_a),\, h^m_a\otimes h^m_a\right\rangle_{H^{\otimes 2}}.$$

Theorem 4.4. *There exists a constant $c>0$ such that for any $\tau>0$, for any $\varepsilon>0$, for any $v\in\mathcal{V}^m$, for any $f\in\mathrm{Lip}(W)$,*

$$\left|\left\langle \nabla^{(2)}P^m_\tau f_N(v+\varepsilon h^m_a) - \nabla^{(2)}P^m_\tau f_N(v),\, h^m_a\otimes h^m_a\right\rangle_{H^{\otimes 2}}\right| \le c\,\frac{e^{-5\tau/2}}{\beta^2_{\tau/2}}\,\varepsilon N^{\eta-\frac12}\sqrt{\frac{N^3}{m^3}}\cdot \tag{4.7}$$

Proof of Theorem 4.4. We know from [13, 4] that we have the following representation: for any $h\in H$,

$$\left\langle \nabla^{(2)}P^m_\tau f(v),\, h\otimes h\right\rangle_{H^{\otimes 2}} = \frac{e^{-3\tau/2}}{\beta^2_{\tau/2}}\,\mathbf{E}\left[f\Big(w_\tau(v,B^m,\hat{B}^m)\Big)\,\delta h(B^m)\,\delta h(\hat{B}^m)\right] \tag{4.8}$$

where

$$w_\tau(v,y,z) = e^{-\tau/2}(e^{-\tau/2}v+\beta_{\tau/2}y)+\beta_{\tau/2}z$$

and $\hat{B}^m$ is an independent copy of B^m. Since the map v is linear with respect to its three arguments,

$$f_N\Big(w_\tau(v,\,B^m,\,\hat{B}^m)\Big) = f_N\Big(w_\tau(\pi^N v,\,\pi^N B^m,\,\pi^N\hat{B}^m)\Big).$$

Hence,

$$\left(\frac{e^{-3\tau/2}}{\beta^2_{\tau/2}}\right)^{-1}\left\langle \nabla^{(2)}P^m_\tau f_N(v),\, h\otimes h\right\rangle_{H^{\otimes 2}} \\ = \mathbf{E}\left[f_N\Big(w_\tau(\pi^N v, \pi^N B^m, \pi^N \hat{B}^m)\Big)\mathbf{E}\left[\delta h(B^m)\,|\,\pi^N B^m\right]\mathbf{E}\left[\delta h(\hat{B}^m)\,|\,\pi^N\hat{B}^m\right]\right] \quad (4.9)$$

From Lemma 4.7, we know that

$$\operatorname{Var}\Big(\mathbf{E}\left[\delta h(\hat{B}^m)\,|\,\pi^N\hat{B}^m\right]\Big) \le c\,\frac{N}{m} \quad (4.10)$$

for $m > 8\,N$, and the same holds for the other conditional expectation. Use Cauchy-Schwarz inequality in (4.9) and take (4.10) into account to obtain

$$\left(\frac{e^{-3\tau/2}}{\beta^2_{\tau/2}}\right)^{-1}\left|\left\langle \nabla^{(2)}P^m_\tau f_N(v+\varepsilon h^m_a) - \nabla^{(2)}P^m_\tau f_N(v),\, h^m_a\otimes h^m_a\right\rangle_{H^{\otimes 2}}\right| \\ \le c\left(\frac{N}{m}\right)^2\left\|w_\tau(\pi^N v, \pi^N B^m, \pi^N\hat{B}^m) - w_\tau(\pi^N v + \varepsilon\pi^N h^m_a, \pi^N B^m, \pi^N \hat{B}^m)\right\|_W \\ = ce^{-\tau}\varepsilon\left(\frac{N}{m}\right)^2\left\|\pi^N h^m_a\right\|_W \quad (4.11)$$

since f_N belongs to $\operatorname{Lip}_1(W)$. Furthermore,

$$\pi^N(h^m_a) = \sum_{b\in\mathcal{A}^N}\left\langle h^m_a, h^N_b\right\rangle_H\, h^N_b.$$

We already know that

$$0 \le |\left\langle h^m_a, h^N_b\right\rangle_H| \le \sqrt{\frac{N}{m}}$$

and that at most two terms $\left\langle h^m_a, h^N_b\right\rangle_H$ are non zero. Moreover, according to Lemma 2.1

$$\|h^N_b\|_W \le c\,N^{\eta-\frac{1}{2}}.$$

Thus,

$$\|\pi^N(h^m_a)\|_W \le c\sqrt{\frac{N}{m}}\,N^{\eta-\frac{1}{2}}. \quad (4.12)$$

Plug estimation (4.12) into estimation (4.11) yields estimate (4.7). ☐

According to (4.7) and Lemma 4.3, since the cardinality of $\mathcal{A}^m$ is md, we obtain the following theorem.

Theorem 4.5. *There exists $c > 0$ such that provided that X belongs to L^p, for any $\tau > 0$,*

$$\mathbf{E}\left[\int_\tau^\infty LP_u f_N(S^m)\,du\right] \le c\,\|X\|_{L^p}\,\frac{N^{1+\eta}}{\sqrt{m}}\int_\tau^\infty \frac{e^{-5u/2}}{1-e^{-u/2}}\,du. \quad (4.13)$$

If we combine Lemma 4.2 and (4.13), we get

$$|\mathbf{E}\left[f_N(S^m)\right] - \mathbf{E}\left[f_N(B^m)\right]| \le c\|X\|_{L^p}\left(1-e^{-\tau}+\frac{N^{1+\eta}}{\sqrt{m}}\int_\tau^\infty\frac{e^{-5u/2}}{1-e^{-u/2}}\,\mathrm{d}u\right)^{1/2}.$$

Optimizing with respect to τ yields Theorem 3.3.

It remains to prove (4.10). For the sake of simplicity, we give the proof for $d = 1$. The general situation is similar but with more involved notations. We recall that

$$\pi^N(B^m) = \sum_{b=0}^{N-1} G_b^{m,N} h_b^N.$$

where

$$G_b^{m,N} = \sum_{a=0}^{m-1} \left\langle h_a^m, h_b^N \right\rangle_H \delta(h_a^m). \tag{4.14}$$

Lemma 4.6. *The covariance matrix* Γ *of the Gaussian vector* $(G_b^{m,N},\, b = 0, \cdots, N-1)$ *is invertible and satisfies*

$$\|\Gamma^{-1}\|_\infty \le 2. \tag{4.15}$$

Proof. Since the h_a^m are orthogonal in L^2, for any $b, c \in \{0, \cdots, N-1\}$,

$$\Gamma_{b,c} = \sum_{a=0}^{m-1} \left\langle h_a^m, h_b^N \right\rangle_H \left\langle h_a^m, h_c^N \right\rangle_H. \tag{4.16}$$

Since a sub-interval of $\mathcal{D}_m$ intersects at most two sub-intervals of $\mathcal{D}_N$, the matrix Γ is tridiagonal. Furthermore, we know that

$$0 \le \left\langle h_a^m, h_b^N \right\rangle_H \le \sqrt{\frac{N}{m}}, \tag{4.17}$$

and for each b, there are at least $(N/m - 3)$ terms of this kind which are equal to $(N/m)^{-1/2}$. Hence,

$$\Gamma_{b,b} \ge (\frac{m}{N} - 3)(\sqrt{\frac{N}{m}})^2 \ge \frac{3}{4}.$$

Since Γ is tridiagonal, this implies that it is invertible. Moreover, let D be the diagonal matrix extracted from Γ. We have proved that $\|D\|_\infty \ge 3/4$.

For $|b - c| = 1$, there is at most one term of the sum (4.16) which yields a non zero scalar product, hence

$$|\Gamma_{b,c}| \le \frac{N}{m}.$$

Set $S = \Gamma - D$. The matrix $D^{-1}S$ has at most two non null entries and

$$\|D^{-1}S\|_\infty \le \frac{8}{3}\frac{N}{m} \le \frac{1}{3},$$

if $m > 8N$. By iteration, we get for any $k \ge 1$,

$$\|(D^{-1}S)^k\|_\infty \le \frac{1}{3^k}.$$

Moreover,

$$\sum_{k=0}^{\infty} (-D^{-1}S)^k = (\mathrm{Id} + D^{-1}S)^{-1} = \Gamma^{-1}D.$$

Thus,

$$\|\Gamma^{-1}\|_\infty \le \frac{4}{3}\sum_{k=0}^{\infty} \frac{1}{3^k} = 2.$$

The proof is thus complete. $\square$

Lemma 4.7. *There exists a constant c which depends only on the dimension d such that for all m, N with $m > 8N$, for any $a \in \mathcal{A}^N$*

$$\operatorname{Var}\left[\mathbf{E}\left[\delta(h_a^m) \,|\, \pi^N(B^m)\right]\right] \le c\,\frac{N}{m}\cdot$$

Proof. Using the framework of Gaussian vectors, for all $a \in \{0, \cdots, m-1\}$

$$\mathbf{E}\left[\delta(h_a^m) \,|\, \pi^N(B^m)\right] = \sum_{b \in \mathcal{A}^N} C_{a,b}^{m,N} G_b^{m,N}. \tag{4.18}$$

For any $c \in \{0, \cdots, N-1\}$, on the one hand

$$\mathbf{E}\left[\mathbf{E}\left[\delta(h_a^m) \,|\, \pi^N(B^m)\right]\ G_c\right] = \sum_{b=0}^{N-1} \sum_{\tau=0}^{m-1} C_{a,b}^{m,N} \left\langle h_\tau^m,\ h_b^N \right\rangle_H \left\langle h_\tau^m,\ h_c^N \right\rangle_H = \sum_{b=0}^{N-1} C_{a,b}^{m,N} \Gamma_{b,c}.$$

and on the other hand,

$$\mathbf{E}\left[\mathbf{E}\left[\delta(h_a^m) \,|\, \pi^N(B^m)\right]\ G_c\right] = \mathbf{E}\left[\delta(h_a^m)\, G_c\right] = \left\langle h_a^m,\ h_c^N \right\rangle_H.$$

This means that

$$\left(\left\langle h_a^m,\ h_c^N \right\rangle_H, c = 0, \cdots, N-1\right) = \left(C_{a,b}^{m,N},\, b = 0, \cdots, N-1\right)\Gamma.$$

In view of Lemma 4.6, this entails that

$$\left(C_{a,b}^{m,N},\, b = 0, \cdots, N-1\right) = \left(\left\langle h_a^m,\ h_c^N \right\rangle_H, c = 0, \cdots, N-1\right)\Gamma^{-1}.$$

Once again we invoke (4.17) and the fact that at most two of the terms $\left\langle h_a^m,\ h_c^N \right\rangle_H$ are non zero for a fixed a, to deduce that

$$\sup_{a,b} |C_{a,b}^{m,N}| \le 2\|\Gamma^{-1}\|_\infty \sqrt{\frac{N}{m}} = 4\sqrt{\frac{N}{m}}\cdot \tag{4.19}$$

Now then, according to the very definition of the conditional expectation

$$\operatorname{Var}\left[\mathbf{E}\left[\delta(h_a^m)|\pi^N(B^m)\right]\right] = \mathbf{E}\left[\delta(h_a^m)\ \mathbf{E}\left[\delta(h_a^m)|\pi^N(B^m)\right]\right] = \sum_{b=0}^{N-1} C_{a,b}^{m,N} \left\langle h_a^m, h_b^N \right\rangle_H.$$

Hence,

$$\operatorname{Var}\left[\mathbf{E}\left[\delta(h_a^m)|\pi^N(B^m)\right]\right] \le 2\sup_{a,b}|C_{a,b}^{m,N}|\sqrt{\frac{N}{m}} \le 8\,\frac{N}{m},$$

according to (4.19). The constant 8 has to be modified when $d > 1$. □

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, second ed., Pure and Applied Mathematics (Amsterdam), vol. 140, Elsevier/Academic Press, Amsterdam, 2003. MR-2424078

[2] A. D. Barbour, *Stein's method for diffusion approximations*, Probability Theory and Related Fields **84** (1990), no. 3, 297–322. MR-1035659

[3] L. Coutin and L. Decreusefond, *Stein's method for Brownian approximations*, Communications on Stochastic Analysis **7** (2013), no. 3, 349–372. MR-3167403

[4] L. Coutin and L. Decreusefond, *Convergence rate in the rough Donsker theorem*, Potential Analysis (in press, 2020).

[5] L. Decreusefond, *The Stein-Dirichlet-Malliavin method*, ESAIM: Proceedings (2015), 11. MR-3440790

[6] X. Fang, Q.-M. Shao, and L. Xu, *Multivariate approximations in Wasserstein distance by Stein's method and Bismut's formula*, Probability Theory and Related Fields **174** (2019), no. 3, 945–979. MR-3980309

[7] P. Friz and N. Victoir, *Multidimensional Stochastic Processes As Rough Paths*, Cambridge Studies in Advanced Mathematics, vol. 120, Cambridge University Press, Cambridge, 2010. MR-2604669

[8] T. Gallouët, G. Mijoule, and Y. Swan, *Regularity of solutions of the stein equation and rates in the multivariate central limit theorem*, arXiv:1805.01720.

[9] K. Itô and M. Nisio, *On the convergence of sums of independent Banach space valued random variables*, Osaka Journal of Mathematics **5** (1968), 35–48. MR-0235593

[10] M. J. Kasprzak, *Diffusion approximations via Stein's method and time changes*, arXiv:1701.07633.

[11] M. Raič, *A multivariate central limit theorem for Lipschitz and smooth test functions*, arXiv:1812.08268.

[12] P. Robert, *Stochastic networks and queues*, Applications of Mathematics, vol. 52, Springer-Verlag, Berlin, 2003. MR-1996883

[13] H.-H. Shih, *On Stein's method for infinite-dimensional Gaussian approximation in abstract Wiener spaces*, Journal of Functional Analysis **261** (2011), no. 5, 1236–1283. MR-2807099

[14] A. S. Üstünel, *An Introduction to Analysis on Wiener Space*, Lectures Notes in Mathematics, vol. 1610, Springer-Verlag, 1995. MR-1439752