Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2010, Vol. 46, No. 1, 45–58
DOI: 10.1214/08-AIHP308

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Multivariate normal approximation using Stein's method and Malliavin calculus

Ivan Nourdin[a], Giovanni Peccati[b] and Anthony Réveillac[c]

[a]*LPMA, Université Paris VI, Boîte courrier 188, 4 place Jussieu, 75252 Paris Cedex 05, France. E-mail:* ivan.nourdin@upmc.fr
[b]*Équipe Modal'X, Université Paris Ouest – Nanterre la Défense, 200 avenue de la République, 92000 Nanterre and LSTA, Université Paris VI, France. E-mail:* giovanni.peccati@gmail.com
[c]*Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. E-mail:* areveill@mathematik.hu-berlin.de

Received 11 April 2008; revised 20 November 2008; accepted 26 November 2008

Abstract. We combine Stein's method with Malliavin calculus in order to obtain explicit bounds in the multidimensional normal approximation (in the Wasserstein distance) of functionals of Gaussian fields. Among several examples, we provide an application to a functional version of the Breuer–Major CLT for fields subordinated to a fractional Brownian motion.

Résumé. Nous expliquons comment combiner la méthode de Stein avec les outils du calcul de Malliavin pour majorer, de manière explicite, la distance de Wasserstein entre une fonctionnelle d'un champs gaussien donnée et son approximation normale multidimensionnelle. Entre autres exemples, nous associons des bornes à la version fonctionnelle du théorème de la limite centrale de Breuer–Major, dans le cas du mouvement brownien fractionnaire.

MSC: 60F05; 60G15; 60H07

Keywords: Breuer–Major CLT; fractional Brownian motion; Gaussian processes; Malliavin calculus; Normal approximation; Stein's method; Wasserstein distance

---

1. Introduction

Let $Z \sim \mathcal{N}(0,1)$ be a standard Gaussian random variable on some probability space $(\Omega, \mathcal{F}, P)$, and let F be a real-valued functional of an infinite-dimensional Gaussian field. In the papers [11,12] it is shown that one can combine Stein's method (see e.g. [5,19] or [20]) with Malliavin calculus (see e.g. [13]), in order to deduce explicit (and, sometimes, optimal) bounds for quantities of the type $d(F, Z)$, where d stands for some distance between the law of F and the law of Z (e.g., d can be the Kolmogorov or the Wasserstein distance). The aim of this paper is to extend the results of [11,12] to the framework of the *multidimensional* Gaussian approximation in the Wasserstein distance. Once again, our techniques hinge upon the use of infinite-dimensional operators on Gaussian spaces (like the *divergence operator* or the *Ornstein–Uhlenbeck generator*) and upon an appropriate multidimensional version of Stein's method (in a form close to Chatterjee and Meckes [4], but see also Reinert and Röllin [18]). As a result, we will obtain explicit bounds, both in terms of Malliavin derivatives and contraction operators, thus providing a substantial refinement of the main findings by Nualart and Ortiz-Latorre [14] and Peccati and Tudor [17]. Note that an important part of our computations (see e.g. Lemma 3.7) are directly inspired by those contained in [14]: we shall indeed stress that this last reference contains a fundamental methodological breakthrough, showing that one can deal with (possibly multidimensional) weak convergence on a Gaussian space, by means of Malliavin-type operators and "characterizing" differential equations. See [10] for an application of these techniques to non-central limit theorems. Incidentally, observe that the paper [16], which is mainly based on martingale-type techniques, also uses distances between probability measures (such as the Prokhorov distance) to deal with multidimensional Gaussian approximations on Wiener space, but without giving explicit bounds.

The rationale behind Stein's method is better understood in dimension one. In this framework, the starting point is the following crucial result, proved, e.g., in [19].

Lemma 1.1 (Stein's lemma). *A random variable* Y *is such that* $Y \stackrel{\text{Law}}{=} Z \sim \mathcal{N}(0,1)$ *if and only if, for every continuous and piecewise continuously differentiable function* $f:\mathbb{R}\to\mathbb{R}$ *such that* $E|f'(Z)| < \infty$, *one has*

$$E\big[f'(Y) - Yf(Y)\big] = 0. \tag{1.1}$$

The fact that a random variable Y satisfying (1.1) is necessarily Gaussian can be proved by several routes: For instance, by taking f to be a complex exponential, one can show that the characteristic function of Y, say $\psi(t)$, is necessarily a solution to the differential equation $\psi'(t) + t\psi(t) = 0$, and therefore $\psi(t) = \exp(-t^2/2)$; alternatively, one can set $f(x) = x^n$, $n = 1, 2, \dots$, and observe that (1.1) implies that, for every n, one must have $E(Y^n) = E(Z^n)$, where $Z \sim \mathcal{N}(0,1)$ (note that the law of Z is determined by its moments).

Heuristically, Lemma 1.1 suggests that the distance $d(Y,Z)$, between the law of a random variable Y and that of $Z \sim \mathcal{N}(0,1)$, must be "small" whenever $E[f'(Y) - Yf(Y)] \simeq 0$, for a sufficiently large class of functions f. In the seminal works [19,20], Stein proved that this somewhat imprecise argument can be made rigorous by means of the use of differential equations. To see this, for a given function $g:\mathbb{R}\to\mathbb{R}$, define the *Stein equation* associated with g as

$$g(x) - E\big[g(Z)\big] = h'(x) - xh(x), \quad \forall x \in \mathbb{R} \tag{1.2}$$

(we recall that $Z \sim \mathcal{N}(0,1)$). A solution to (1.2) is a function h which is Lebesgue-almost everywhere differentiable, and such that there exists a version of h' satisfying (1.2) for every $x \in \mathbb{R}$. If one assumes that $g \in \mathrm{Lip}(1)$ (that is, if $\|g\|_{\mathrm{Lip}} \le 1$, where $\|\cdot\|_{\mathrm{Lip}}$ stands for the usual Lipschitz seminorm), then a standard result (see e.g. [20]) yields that (1.2) admits a solution h such that $\|h'\|_\infty \le 1$ and $\|h''\|_\infty \le 2$. Now recall that the *Wasserstein distance* between the laws of two real-valued random variables Y and X is defined as

$$d_{\mathrm{W}}(Y,X) = \sup_{g\in \mathrm{Lip}(1)} \big|E\big[g(Y)\big] - E\big[g(X)\big]\big|,$$

and introduce the notation $\mathcal{F}_{\mathrm{W}} = \{f:\ \|f'\|_\infty \le 1,\ \|f''\|_\infty \le 2\}$. By taking expectations on the two sides of (1.2), one obtains finally that, for $Z \sim \mathcal{N}(0,1)$ and for a generic random variable Y,

$$d_{\mathrm{W}}(Y,Z) \le \sup_{f\in\mathcal{F}_{\mathrm{W}}} \big|E\big[f'(Y) - Yf(Y)\big]\big|, \tag{1.3}$$

thus giving a precise meaning to the heuristic argument sketched above (note that an analogous conclusion can be obtained for other distances, such as the total variation distance or the Kolmogorov distance – see e.g. [5] for a discussion of this point). We stress that the topology induced by d_{W}, on probability measures on $\mathbb{R}$, is stronger than the topology induced by weak convergence.

The starting point of [11,12] is that a relation such as (1.3) can be very effectively combined with Malliavin calculus, whenever Y is a centered regular functional of some infinite dimensional Gaussian field. To see this, denote by DY the Malliavin derivative of Y (observe that DY is a random element with values in some adequate Hilbert space $\mathfrak{H}$), and write L to indicate the (infinite-dimensional) Ornstein–Uhlenbeck generator (see Section 2 below for precise definitions). One crucial relation proved in [11], and then further exploited in [12], is the upper bound

$$d_{\mathrm{W}}(Y,Z) \le E\big|1 - \langle DY, -DL^{-1}Y\rangle_{\mathfrak{H}}\big|. \tag{1.4}$$

As shown in [11], when specialized to the case of Y being equal to a multiple Wiener–Itô integral, relation (1.4) yields bounds that are intimately related with the CLTs proved in [14] and [15]. See [12] for a characterization of the optimality of these bounds; see again [11] for extensions to non-Gaussian approximations and for applications to the Breuer–Major CLT (stated and proved in [2]) for functionals of a fractional Brownian motion.

The principal contribution of the present paper (see e.g. the statement of Theorem 3.5 below) consists in showing that a relation similar to (1.4) continues to hold when Z is replaced by a d-dimensional ($d \geq 2$) Gaussian vector $F = (F_1, \dots, F_d)$ of smooth functionals of a Gaussian field, and d_W is the Wasserstein distance between probability laws on $\mathbb{R}^d$ (see Definition 3.1 below). Our results apply to Gaussian approximations by means of Gaussian vectors with arbitrary positive definite covariance matrices. The proofs rely on a multidimensional version of the Stein equation (1.2), that we combine with standard integration by parts formulae on an infinite-dimensional Gaussian space. Our approach bears some connections with the paper by Hsu [9], where the author proves an hybrid Stein/semimartingale characterization of Brownian motions on manifolds, via Malliavin-type operators.

The paper is organized as follows. In Section 2 we provide some preliminaries on Malliavin calculus. Section 3 contains our main results, concerning Gaussian approximations by means of vectors of Gaussian random variables with positive definite covariance matrices. Finally, Section 4 deals with two applications: (i) to a functional version of the Breuer–Major CLT (see [2]), and (ii) to Gaussian approximations of functionals of finite normal vectors, providing a generalization of a technical result proved by Chatterjee in [3].

2. Preliminaries and notation

In this section, we recall some basic elements of Malliavin calculus for Gaussian processes. The reader is referred to [13] for a complete discussion of this subject. Let $X = \{X(h),\ h \in \mathfrak{H}\}$ be an *isonormal Gaussian process* on a probability space $(\Omega, \mathcal{F}, P)$. This means that X is a centered Gaussian family indexed by the elements of an Hilbert space $\mathfrak{H}$, such that, for every pair $h, g \in \mathfrak{H}$ one has that $E[X(h)X(g)] = \langle h, g\rangle_{\mathfrak{H}}$.

We let $L^2(X)$ be shorthand for the space $L^2(\Omega, \sigma(X), P)$. It is well known that every random variable $F \in L^2(X)$ admits the chaotic expansion $F = E(F) + \sum_{n=1}^{\infty} I_n(f_n)$ where the deterministic kernels f_n, $n \geq 1$, belong to $\mathfrak{H}^{\odot n}$ and the convergence of the series holds in $L^2(X)$. One sometimes uses the notation $I_0(f_0) = E[F]$. In the particular case where $\mathfrak{H} := L^2(T, \mathcal{A}, \mu)$, with $(T, \mathcal{A})$ a measurable space and μ is a σ-finite measure without atoms, the random variable $I_n(f_n)$ coincides with the *multiple Wiener–Itô integral* (of order n) of f_n with respect to X (see [13], Section 1.1.2.).

Let $f \in \mathfrak{H}^{\odot p}$, $g \in \mathfrak{H}^{\odot q}$ and $0 \leq r \leq p \wedge q$. We define the rth *contraction* $f \otimes_r g$ of f and g as the element of $\mathfrak{H}^{\otimes(p+q-2r)}$ given by

$$f \otimes_r g := \sum_{i_1,\dots,i_r=1}^{\infty} \langle f, e_{i_1} \otimes \cdots \otimes e_{i_r}\rangle_{\mathfrak{H}^{\otimes r}} \otimes \langle g, e_{i_1} \otimes \cdots \otimes e_{i_r}\rangle_{\mathfrak{H}^{\otimes r}},$$

where $\{e_k,\ k \geq 1\}$ is a complete orthonormal system in $\mathfrak{H}$. Note that $f \otimes_0 g = f \otimes g$; also, if $p = q$, then $f \otimes_p g = \langle f, g\rangle_{\mathfrak{H}^{\otimes p}}$. Note that, in general, $f \otimes_r g$ is not a symmetric element of $\mathfrak{H}^{\otimes(p+q-2r)}$; the canonical symmetrization of $f \otimes_r g$ is denoted by $f \widetilde{\otimes}_r g$. We recall the product formula for multiple stochastic integrals:

$$I_p(f) I_q(g) = \sum_{r=0}^{p\wedge q} r! \binom{p}{r} \binom{p}{q} I_{p+q-2r}(f \widetilde{\otimes}_r g).$$

Now, let $\mathcal{S}$ be the set of cylindrical functionals F of the form

$$F = \varphi\bigl(X(h_1), \dots, X(h_n)\bigr), \tag{2.1}$$

where $n \geq 1$, $h_i \in \mathfrak{H}$ and the function $\varphi \in \mathcal{C}^{\infty}(\mathbb{R}^n)$ is such that its partial derivatives have polynomial growth. The *Malliavin derivative* DF of a functional F of the form (2.1) is the square integrable $\mathfrak{H}$-valued random variable defined as

$$DF = \sum_{i=1}^{n} \partial_i \varphi\bigl(X(h_1), \dots, X(h_n)\bigr) h_i, \tag{2.2}$$

where $\partial_i \varphi$ denotes the ith partial derivative of φ. In particular, one has that $DX(h) = h$ for every h in $\mathfrak{H}$. By iteration, one can define the mth derivative $D^m F$ of $F \in \mathcal{S}$, which is an element of $L^2(\Omega; \mathfrak{H}^{\odot m})$, for $m \geq 2$. As usual $\mathbb{D}^{m,2}$ denotes the closure of $\mathcal{S}$ with respect to the norm $\|\cdot\|_{m,2}$ defined by the relation $\|F\|_{m,2}^2 = E[F^2] + \sum_{i=1}^{m} E[\|D^i F\|_{\mathfrak{H}^{\otimes i}}^2]$.

Note that every finite sum of Wiener–Itô integrals always belongs to $\mathbb{D}^{m,2}$ ($\forall m \geq 1$). The Malliavin derivative D satisfies the following *chain rule formula*: if $\varphi : \mathbb{R}^n \to \mathbb{R}$ is in $\mathcal{C}_b^1$ (defined as the set of continuously differentiable functions with bounded partial derivatives) and if $(F_1, \ldots, F_n)$ is a random vector such that each component belongs to $\mathbb{D}^{1,2}$, then $\varphi(F_1, \ldots, F_n)$ is itself an element of $\mathbb{D}^{1,2}$, and moreover,

$$D\varphi(F_1, \ldots, F_n) = \sum_{i=1}^{n} \partial_i \varphi(F_1, \ldots, F_n) DF_i. \tag{2.3}$$

The *divergence operator* δ is defined as the dual operator of D. Precisely, a random element u of $L^2(\Omega; \mathfrak{H})$ belongs to the domain of δ (denoted by Domδ) if there exists a constant c_u satisfying $|E[\langle DF, u\rangle_{\mathfrak{H}}]| \leq c_u \|F\|_{L^2(\Omega)}$ for every $F \in \mathcal{S}$; in this case, the divergence of u, written $\delta(u)$, is defined by the following duality property:

$$E\big[F\delta(u)\big] = E\big[\langle DF, u\rangle_{\mathfrak{H}}\big], \quad \forall F \in \mathbb{D}^{1,2}. \tag{2.4}$$

The crucial relation (2.4) is customarily called the (Malliavin) *integration by parts formula*.

In what follows, we shall denote by $T = \{T_t : t \geq 0\}$ the *Ornstein–Uhlenbeck semigroup*. We recall that, for every $t \geq 0$ and every $F \in L^2(X)$,

$$T_t(F) = \sum_{n=0}^{\infty} \mathrm{e}^{-nt} J_n(F), \tag{2.5}$$

where, for every $n \geq 0$ and for the rest of the paper, the symbol J_n denotes the projection operator onto the nth Wiener chaos, that is onto the closed linear subspace of $L^2(X)$ generated by the random variables of the form $H_n(X(h))$ with $h \in \mathfrak{H}$ such that $\|h\|_{\mathfrak{H}} = 1$, and H_n the nth Hermite polynomial defined by (4.1). Note that T is indeed the semigroup associated with an infinite-dimensional stationary Gaussian process with values in $\mathbb{R}^{\mathfrak{H}}$, having the law of X as an invariant distribution (see e.g. [13], Section 1.4, for a more detailed discussion of the Ornstein–Uhlenbeck semigroup in the context of Malliavin calculus; see Barbour [1] for a version of Stein's method involving Ornstein–Uhlenbeck semigroups on infinite-dimensional spaces; see Götze [8] for a version of Stein's method based on multi-dimensional Ornstein–Uhlenbeck semigroups). The *infinitesimal generator of the Ornstein–Uhlenbeck semigroup* is noted L. A square integrable random variable F is in the domain of L (noted Dom$L$) if F belongs to the domain of δD (that is, if F is in $\mathbb{D}^{1,2}$ and $DF \in$ Domδ) and, in this case, $LF = -\delta DF$. One can prove that LF is such that $LF = -\sum_{n=0}^{\infty} n J_n(F)$. As an example, if $F = I_q(f_q)$, with $f_q \in \mathfrak{H}^{\odot q}$, then $LF = -qF$. Note that, for every $F \in$ Dom$L$, one has $E(LF) = 0$. The inverse L^{-1} of the operator L acts on zero-mean random variables $F \in L^2(X)$ as $L^{-1}F = -\sum_{n=1}^{\infty} \frac{1}{n} J_n(F)$. In particular, for every $q \geq 1$ and every $F = I_q(f_q)$ with $f_q \in \mathfrak{H}^{\odot q}$, one has that $L^{-1}F = -\frac{1}{q}F$.

We conclude this section by recalling two important characterizations of the Ornstein–Uhlenbeck semigroup and its generator.

(i) *Mehler's formula*. Let F be an element of $L^2(X)$, so that F can be represented as an application from $\mathbb{R}^{\mathfrak{H}}$ into $\mathbb{R}$. Then, an alternative representation (due to Mehler) of the action of the Ornstein–Uhlenbeck semigroup T (as defined in (2.5)) on F, is the following:

$$T_t(F) = E\big[F\big(\mathrm{e}^{-t}a + \sqrt{1 - \mathrm{e}^{-2t}}X\big)\big]\big|_{a=X}, \quad t \geq 0, \tag{2.6}$$

where a designs a generic element of $\mathbb{R}^{\mathfrak{H}}$. See Nualart [13], Section 1.4.1, for more details on this and other characterizations of T.

(ii) *Differential characterization of* L. Let $F \in L^2(X)$ have the form $F = f(X(h_1), \ldots, X(h_d))$, where $f \in \mathcal{C}^2(\mathbb{R}^d)$ has bounded first and second derivatives, and $h_i \in \mathfrak{H}$, $i = 1, \ldots, d$. Then,

$$LF = \sum_{i,j=1}^{d} \frac{\partial^2 f}{\partial x_i \partial x_j}\big(X(h_1), \ldots, X(h_d)\big)\langle h_i, h_j\rangle_{\mathfrak{H}} - \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}\big(X(h_1), \ldots, X(h_d)\big)X(h_i). \tag{2.7}$$

See Propositions 1.4.4 and 1.4.5 in [13] for a proof and some generalizations of (2.7).

3. Stein's method and Gaussian vectors

We start by giving a definition of the Wasserstein distance, as well as by introducing some useful norms over classes of real-valued matrices.

Definition 3.1. (i) *The Wasserstein distance between the laws of two $\mathbb{R}^d$-valued random vectors X and Y, noted $d_{\mathrm{W}}(X,Y)$, is given by*

$$d_{\mathrm{W}}(X,Y) := \sup_{g\in\mathcal{H};\|g\|_{\mathrm{Lip}}\le 1} \bigl|E\bigl[g(X)\bigr]-E\bigl[g(Y)\bigr]\bigr|,$$

where $\mathcal{H}$ indicates the class of Lipschitz functions, that is, the collection of all functions $g:\mathbb{R}^d\to\mathbb{R}$ such that $\|g\|_{\mathrm{Lip}} := \sup_{x\neq y}\frac{|g(x)-g(y)|}{\|x-y\|_{\mathbb{R}^d}} < \infty$ (with $\|\cdot\|_{\mathbb{R}^d}$ the usual Euclidian norm on $\mathbb{R}^d$).

(ii) *The Hilbert–Schmidt inner product and the Hilbert–Schmidt norm on the class of $d\times d$ real matrices, denoted respectively by $\langle\cdot,\cdot\rangle_{\mathrm{H.S.}}$ and $\|\cdot\|_{\mathrm{H.S.}}$, are defined as follows: for every pair of matrices A and B, $\langle A,B\rangle_{\mathrm{H.S.}} := \mathrm{Tr}(AB^{\mathrm{T}})$ and $\|A\|_{\mathrm{H.S.}} := \sqrt{\langle A,A\rangle_{\mathrm{H.S.}}}$.*

(iii) *The operator norm of a $d\times d$ matrix A over $\mathbb{R}$ is given by $\|A\|_{\mathrm{op}} := \sup_{\|x\|_{\mathbb{R}^d}=1}\|Ax\|_{\mathbb{R}^d}$.*

Remark 3.2.

1. *For every $d\ge 1$ the topology induced by d_{W}, on the class of all probability measures on $\mathbb{R}^d$, is strictly stronger than the topology induced by weak convergence (see e.g. Dudley [6], Chapter 11).*
2. *The reason why we focus on the Wasserstein distance is nested in the statement of the forthcoming Lemma 3.3. Indeed, according to relation (3.4), in order to control the second derivatives of the solution of the Stein equation (3.3) associated with g, one must use the fact that g is Lipschitz.*
3. *According to the notation introduced in Definition 3.1(ii), relation (2.7) can be rewritten as*

$$LF = \bigl\langle C,\mathrm{Hess}\, f(Z)\bigr\rangle_{\mathrm{H.S.}} - \bigl\langle Z,\nabla f(Z)\bigr\rangle_{\mathbb{R}^d}, \tag{3.1}$$

where $Z=(X(h_1),\dots,X(h_d))$, and $C=\{C(i,j)\colon i,j=1,\dots,d\}$ is the $d\times d$ covariance matrix such that $C(i,j)=E(X(h_i)X(h_j))=\langle h_i,h_j\rangle_{\mathfrak{H}}$.

Given a $d\times d$ positive definite symmetric matrix C, we use the notation $\mathcal{N}_d(0,C)$ to indicate the law of a d-dimensional Gaussian vector with zero mean and covariance C. The following result, which is basically known (see e.g. [4] or [18]), is the d-dimensional counterpart of Stein's Lemma 1.1. In what follows, we provide a new proof which is almost exclusively based on the use of Malliavin operators.

Lemma 3.3. *Fix an integer $d\ge 2$ and let $C=\{C(i,j)\colon i,j=1,\dots,d\}$ be a $d\times d$ positive definite symmetric real matrix.*

(i) *Let Y be a random variable with values in $\mathbb{R}^d$. Then $Y\sim\mathcal{N}_d(0,C)$ if and only if, for every twice differentiable function $f:\mathbb{R}^d\to\mathbb{R}$ such that $E|\langle C,\mathrm{Hess}\, f(Y)\rangle_{\mathrm{H.S.}}| + E|\langle Y,\nabla f(Y)\rangle_{\mathbb{R}^d}|<\infty$, it holds that*

$$E\bigl[\bigl\langle Y,\nabla f(Y)\bigr\rangle_{\mathbb{R}^d} - \bigl\langle C,\mathrm{Hess}\, f(Y)\bigr\rangle_{\mathrm{H.S.}}\bigr] = 0. \tag{3.2}$$

(ii) *Consider a Gaussian random vector $Z\sim\mathcal{N}_d(0,C)$. Let $g:\mathbb{R}^d\to\mathbb{R}$ belong to $\mathcal{C}^2(\mathbb{R}^d)$ with first and second bounded derivatives. Then, the function $U_0(g)$ defined by*

$$U_0 g(x) := \int_0^1 \frac{1}{2t} E\bigl[g\bigl(\sqrt{t}x+\sqrt{1-t}Z\bigr)-g(Z)\bigr]\,\mathrm{d}t$$

is a solution to the following differential equation (with unknown function f):

$$g(x)-E\bigl[g(Z)\bigr] = \bigl\langle x,\nabla f(x)\bigr\rangle_{\mathbb{R}^d} - \bigl\langle C,\mathrm{Hess}\, f(x)\bigr\rangle_{\mathrm{H.S.}},\quad x\in\mathbb{R}^d. \tag{3.3}$$

Moreover, one has that

$$\sup_{x\in\mathbb{R}^d} \|\operatorname{Hess} U_0 g(x)\|_{\text{H.S.}} \le \|C^{-1}\|_{\text{op}} \|C\|_{\text{op}}^{1/2} \|g\|_{\text{Lip}}. \tag{3.4}$$

Remark 3.4.

1. *If* $C = \sigma^2 \boldsymbol{I}_d$ *for some* $\sigma > 0$ (*that is, if* Z *is composed of i.i.d. centered Gaussian random variables with common variance equal to* σ^2), *then*

$$\|C^{-1}\|_{\text{op}} \|C\|_{\text{op}}^{1/2} = \|\sigma^{-2}\boldsymbol{I}_d\|_{\text{op}} \|\sigma^2 \boldsymbol{I}_d\|_{\text{op}}^{1/2} = \sigma^{-1}.$$

2. *Unlike formulae* (1.1) *and* (1.2) (*associated with one-dimensional Gaussian approximations*) *the relation* (3.2) *and the Stein equation* (3.3) *involve second-order differential operators. A discussion of this fact is detailed e.g. in* [4], *Theorem* 4.

Proof of Lemma 3.3. We start by proving Point (ii). First observe that, without loss of generality, we can suppose that $Z = (Z_1, \dots, Z_d) := (X(h_1), \dots, X(h_d))$, where X is an isonormal Gaussian process over $\mathfrak{H} = \mathbb{R}^d$, the kernels h_i belong to $\mathfrak{H}$ ($i = 1, \dots, d$), and $\langle h_i, h_j\rangle_{\mathfrak{H}} = E(X(h_i)X(h_j)) = E(Z_i Z_j) = C(i, j)$. By using the change of variable $2u = -\log t$, one can rewrite $U_0 g(x)$ as follows

$$U_0 g(x) = \int_0^\infty \bigl\{E\bigl[g\bigl(\mathrm{e}^{-u}x + \sqrt{1-\mathrm{e}^{-2u}}Z\bigr)\bigr] - E\bigl[g(Z)\bigr]\bigr\}\,\mathrm{d}u.$$

Now define $\widetilde{g}(Z) := g(Z) - E[g(Z)]$, and observe that $\widetilde{g}(Z)$ is by assumption a centered element of $\mathrm{L}^2(X)$. For $q \ge 0$, denote by $J_q(\widetilde{g}(Z))$ the projection of $\widetilde{g}(Z)$ on the qth Wiener chaos, so that $J_0(\widetilde{g}(Z)) = 0$. According to Mehler's formula (2.6),

$$E\bigl[g\bigl(\mathrm{e}^{-u}x + \sqrt{1-\mathrm{e}^{-2u}}Z\bigr)\bigr]\big|_{x=Z} - E\bigl[g(Z)\bigr] = E\bigl[\widetilde{g}\bigl(\mathrm{e}^{-u}x + \sqrt{1-\mathrm{e}^{-2u}}Z\bigr)\bigr]\big|_{x=Z} = T_u\widetilde{g}(Z),$$

where x denotes a generic element of $\mathbb{R}^d$. In view of (2.5), it follows that

$$U_0 g(Z) = \int_0^\infty T_u \widetilde{g}(Z)\,\mathrm{d}u = \int_0^\infty \sum_{q\ge1} \mathrm{e}^{-qu} J_q\bigl(\widetilde{g}(Z)\bigr)\,\mathrm{d}u = \sum_{q\ge1}\frac{1}{q} J_q\bigl(\widetilde{g}(Z)\bigr) = -L^{-1}\widetilde{g}(Z).$$

Since g belongs to $\mathcal{C}^2(\mathbb{R}^d)$ with bounded first and second derivatives, it is easily seen that the same holds for $U_0 g$. By exploiting the differential representation (3.1), one deduces that

$$\bigl\langle Z, \nabla U_0 g(Z)\bigr\rangle_{\mathbb{R}^d} - \bigl\langle C, \operatorname{Hess} U_0 g(Z)\bigr\rangle_{\text{H.S.}} = -LU_0 g(Z) = LL^{-1}\widetilde{g}(Z) = g(Z) - E\bigl[g(Z)\bigr]$$

Since the matrix C is positive definite, we infer that the support of the law of Z coincides with $\mathbb{R}^d$, and therefore (e.g. by a continuity argument) we obtain that

$$\bigl\langle x, \nabla U_0 g(x)\bigr\rangle_{\mathbb{R}^d} - \bigl\langle C, \operatorname{Hess} U_0 g(x)\bigr\rangle_{\text{H.S.}} = g(x) - E\bigl[g(Z)\bigr],$$

for every $x \in \mathbb{R}^d$. This yields that the function $U_0 g$ solves the Stein's equation (3.3).

To prove the estimate (3.4), we first recall that there exists a unique non-singular *symmetric* matrix A such that $A^2 = C$, and that one has that $A^{-1}Z \sim \mathcal{N}_d(0, \mathbf{I}_d)$. Now write $U_0 g(x) = h(A^{-1}x)$, where

$$h(x) = \int_0^1 \frac{1}{2t} E\bigl[g_A\bigl(\sqrt{t}x + \sqrt{1-t}A^{-1}Z\bigr) - g_A\bigl(A^{-1}Z\bigr)\bigr]\,\mathrm{d}t,$$

and $g_A(x) = g(Ax)$. Note that, since $A^{-1}Z \sim \mathscr{N}_d(0, \mathbf{I}_d)$, the function h solves the Stein's equation $\langle x, \nabla h(x)\rangle_{\mathbb{R}^d} - \Delta h(x) = g_A(x) - E[g_A(Y)]$, where $Y \sim \mathscr{N}_d(0, \mathbf{I}_d)$. We can now use the same arguments as in the proof of Lemma 3 in [4] to deduce that

$$\sup_{x\in\mathbb{R}^d} \| \operatorname{Hess} h(x)\|_{\text{H.S.}} \le \|g_A\|_{\text{Lip}} \le \|A\|_{\text{op}}\|g\|_{\text{Lip}}. \tag{3.5}$$

On the other hand, by noting $h_{A^{-1}}(x) = h(A^{-1}x)$, one obtains by standard computations (recall that A is symmetric) that $\operatorname{Hess} U_0 g(x) = \operatorname{Hess} h_{A^{-1}}(x) = A^{-1} \operatorname{Hess} h(A^{-1}x)A^{-1}$, yielding

$$\begin{aligned}
\sup_{x\in\mathbb{R}^d} \| \operatorname{Hess} U_0 g(x)\|_{\text{H.S.}} &= \sup_{x\in\mathbb{R}^d} \|A^{-1} \operatorname{Hess} h(A^{-1}x)A^{-1}\|_{\text{H.S.}} \\
&= \sup_{x\in\mathbb{R}^d} \|A^{-1} \operatorname{Hess} h(x)A^{-1}\|_{\text{H.S.}} \\
&\le \|A^{-1}\|_{\text{op}}^2 \sup_{x\in\mathbb{R}^d} \| \operatorname{Hess} h(x)\|_{\text{H.S.}} && (3.6)\\
&\le \|A^{-1}\|_{\text{op}}^2 \|A\|_{\text{op}}\|g\|_{\text{Lip}} && (3.7)\\
&\le \|C^{-1}\|_{\text{op}} \|C\|_{\text{op}}^{1/2}\|g\|_{\text{Lip}}. && (3.8)
\end{aligned}$$

The chain of inequalities appearing in formulae (3.6)–(3.8) are mainly a consequence of the usual properties of the Hilbert–Schmidt and operator norms. Indeed, to prove inequality (3.6) we used the relations

$$\|A^{-1} \operatorname{Hess} h(x)A^{-1}\|_{\text{H.S.}} \le \|A^{-1}\|_{\text{op}} \| \operatorname{Hess} h(x)A^{-1}\|_{\text{H.S.}} \le \|A^{-1}\|_{\text{op}} \| \operatorname{Hess} h(x)\|_{\text{H.S.}} \|A^{-1}\|_{\text{op}};$$

relation (3.7) is a consequence of (3.5); finally, to show the inequality (3.8), one uses the fact that

$$\|A^{-1}\|_{\text{op}} \le \sqrt{\|A^{-1}A^{-1}\|_{\text{op}}} = \sqrt{\|C^{-1}\|_{\text{op}}} \quad \text{and} \quad \|A\|_{\text{op}} \le \sqrt{\|AA\|_{\text{op}}} = \sqrt{\|C\|_{\text{op}}}.$$

We are now left with the proof of Point (i) in the statement. The fact that a vector $Y \sim \mathscr{N}_d(0, C)$ necessarily verifies (3.2) can be proved by standard integration by parts. On the other hand, suppose that Y verifies (3.2). Then, according to Point (ii), for every $g \in \mathcal{C}^2(\mathbb{R}^d)$ with bounded first and second derivatives,

$$E\big(g(Y)\big) - E\big(g(Z)\big) = E\big(\langle Y, \nabla U_0 g(Y)\rangle_{\mathbb{R}^d} - \langle C, \operatorname{Hess} U_0 g(Y)\rangle_{\text{H.S.}}\big) = 0,$$

where $Z \sim \mathscr{N}_d(0, C)$. Since the collection of all such functions g generates the Borel σ-field on $\mathbb{R}^d$, this implies that $Y \overset{\text{Law}}{=} Z$, thus yielding the desired conclusion. □

The following statement is the main result of this paper. Its proof makes a crucial use of the integration by parts formula (2.4) discussed in Section 2.

Theorem 3.5. *Fix $d \ge 2$ and let $C = \{C(i, j)\colon i, j = 1, \dots, d\}$ be a $d \times d$ positive definite matrix. Suppose that $Z \sim \mathscr{N}_d(0, C)$ and that $F = (F_1, \dots, F_d)$ is a $\mathbb{R}^d$-valued random vector such that $E[F_i] = 0$ and $F_i \in \mathbb{D}^{1,2}$ for every $i = 1, \dots, d$. Then,*

$$\begin{aligned}
d_{\text{W}}(F, Z) &\le \|C^{-1}\|_{\text{op}} \|C\|_{\text{op}}^{1/2} \sqrt{E\|C - \Phi(DF)\|_{\text{H.S.}}^2} && (3.9)\\
&= \|C^{-1}\|_{\text{op}} \|C\|_{\text{op}}^{1/2} \sqrt{\sum_{i,j=1}^{d} E\big[\big(C(i, j) - \langle DF_i, -DL^{-1}F_j\rangle_{\mathfrak{H}}\big)^2\big]}, && (3.10)
\end{aligned}$$

where we write $\Phi(DF)$ to indicate the matrix $\Phi(DF) := \{\langle DF_i, -DL^{-1}F_j\rangle_{\mathfrak{H}}\colon 1 \le i, j \le d\}$.

Proof. We start by proving that, for every $g \in \mathcal{C}^2(\mathbb{R}^d)$ with bounded first and second derivatives,

$$\big|E\big[g(F)\big] - E\big[g(Z)\big]\big| \le \big\|C^{-1}\big\|_{\mathrm{op}} \|C\|_{\mathrm{op}}^{1/2} \|g\|_{\mathrm{Lip}} \sqrt{E\big\|C - \Phi(DF)\big\|_{\mathrm{H.S.}}^2}.$$

To prove such a claim, observe that, according to Point (ii) in Lemma 3.3, $E[g(F)] - E[g(Z)] = E[\langle F, \nabla U_0 g(F)\rangle_{\mathbb{R}^d} - \langle C, \operatorname{Hess} U_0 g(F)\rangle_{\mathrm{H.S.}}]$. Moreover,

$$\begin{aligned}
&\big|E\big[\big\langle C, \operatorname{Hess} U_0 g(F)\big\rangle_{\mathrm{H.S.}} - \big\langle F, \nabla U_0 g(F)\big\rangle_{\mathbb{R}^d}\big]\big| \\
&\quad = \left|E\left[\sum_{i,j=1}^d C(i,j)\partial_{ij}^2 U_0 g(F) - \sum_{i=1}^d F_i \partial_i U_0 g(F)\right]\right| \\
&\quad = \left|\sum_{i,j=1}^d E\big[C(i,j)\partial_{ij}^2 U_0 g(F)\big] - \sum_{i=1}^d E\big[\big(LL^{-1}F_i\big)\partial_i U_0 g(F)\big]\right| \quad \big(\text{since } E(F_i) = 0\big) \\
&\quad = \left|\sum_{i,j=1}^d E\big[C(i,j)\partial_{ij}^2 U_0 g(F)\big] + \sum_{i=1}^d E\big[\delta\big(DL^{-1}F_i\big)\partial_i U_0 g(F)\big]\right| \quad (\text{since } \delta D = -L) \\
&\quad = \left|\sum_{i,j=1}^d E\big[C(i,j)\partial_{ij}^2 U_0 g(F)\big] - \sum_{i=1}^d E\big[\big\langle D\big(\partial_i U_0 g(F)\big), -DL^{-1}F_i\big\rangle_{\mathfrak{H}}\big]\right| \quad \big(\text{by } (2.4)\big) \\
&\quad = \left|\sum_{i,j=1}^d E\big[C(i,j)\partial_{ij}^2 U_0 g(F)\big] - \sum_{i,j=1}^d E\big[\partial_{ji}^2 U_0 g(F)\big\langle DF_j, -DL^{-1}F_i\big\rangle_{\mathfrak{H}}\big]\right| \quad \big(\text{by } (2.3)\big) \\
&\quad = \left|\sum_{i,j=1}^d E\big[\partial_{ij}^2 U_0 g(F)\big(C(i,j) - \big\langle DF_i, -DL^{-1}F_j\big\rangle_{\mathfrak{H}}\big)\big]\right| \\
&\quad = \big|E\big\langle \operatorname{Hess} U_0 g(F), C - \Phi(DF)\big\rangle_{\mathrm{H.S.}}\big| \\
&\quad \le \sqrt{E\big\|\operatorname{Hess} U_0 g(F)\big\|_{\mathrm{H.S.}}^2}\sqrt{E\big\|C - \Phi(DF)\big\|_{\mathrm{H.S.}}^2} \quad (\text{by the Cauchy–Schwarz inequality}) \\
&\quad \le \big\|C^{-1}\big\|_{\mathrm{op}} \|C\|_{\mathrm{op}}^{1/2} \|g\|_{\mathrm{Lip}} \sqrt{E\big\|C - \Phi(DF)\big\|_{\mathrm{H.S.}}^2} \quad \big(\text{by } (3.4)\big).
\end{aligned}$$

To prove the Wasserstein estimate (3.9), it is sufficient to observe that, for every globally Lipschitz function g such that $\|g\|_{\mathrm{Lip}} \le 1$, there exists a family $\{g_\varepsilon\colon \varepsilon > 0\}$ such that:

(i) for each $\varepsilon > 0$, the first and second derivatives of g_ε are bounded;
(ii) for each $\varepsilon > 0$, one has that $\|g_\varepsilon\|_{\mathrm{Lip}} \le \|g\|_{\mathrm{Lip}}$;
(iii) as $\varepsilon \to 0$, $\|g_\varepsilon - g\|_\infty \downarrow 0$.

For instance, we can choose $g_\varepsilon(x) = E[g(x + \sqrt{\varepsilon}N)]$ with $N \sim \mathcal{N}_d(0, \mathbf{I}_d)$. □

Observe that Theorem 3.5 generalizes relation (1.4) (that was proved in [11], Theorem 3.1). We now aim at applying Theorem 3.5 to vectors of multiple stochastic integrals.

Corollary 3.6. *Fix $d \ge 2$ and $1 \le q_1 \le \cdots \le q_d$. Consider a vector $F := (F_1, \ldots, F_d) = (I_{q_1}(f_1), \ldots, I_{q_d}(f_d))$ with $f_i \in \mathfrak{H}^{\odot q_i}$ for any $i = 1, \ldots, d$. Let $Z \sim \mathcal{N}_d(0, C)$, with C positive definite. Then,*

$$d_{\mathrm{W}}(F, Z) \le \big\|C^{-1}\big\|_{\mathrm{op}} \|C\|_{\mathrm{op}}^{1/2} \sqrt{\sum_{1 \le i,j \le d} E\left[\left(C(i,j) - \frac{1}{q_j}\langle DF_i, DF_j\rangle_{\mathfrak{H}}\right)^2\right]}. \tag{3.11}$$

Proof. We have $-L^{-1}F_j = \frac{1}{q_j}F_j$ so that the desired conclusion follows from (3.10). □

When one applies Corollary 3.6 in concrete situations (see e.g. Section 4 below), one can use the following result in order to evaluate the right-hand side of (3.11).

Lemma 3.7. *Let $F = I_p(f)$ and $G = I_q(g)$, with $f \in \mathfrak{H}^{\odot p}$ and $g \in \mathfrak{H}^{\odot q}$ $(p, q \geq 1)$. Let a be a real constant. If $p = q$, one has the estimate*:

$$E\left[\left(a - \frac{1}{p}\langle DF, DG\rangle_{\mathfrak{H}}\right)^2\right]$$
$$\leq \left(a - p!\langle f, g\rangle_{\mathfrak{H}^{\otimes p}}\right)^2 + \frac{p^2}{2}\sum_{r=1}^{p-1}(r-1)!^2\binom{p-1}{r-1}^4(2p-2r)!\left(\|f \otimes_{p-r} f\|^2_{\mathfrak{H}^{\otimes 2r}} + \|g \otimes_{p-r} g\|^2_{\mathfrak{H}^{\otimes 2r}}\right).$$

On the other hand, if $p < q$, one has that

$$E\left[\left(a - \frac{1}{q}\langle DF, DG\rangle_{\mathfrak{H}}\right)^2\right]$$
$$\leq a^2 + p!^2\binom{q-1}{p-1}^2(q-p)!\|f\|^2_{\mathfrak{H}^{\otimes p}}\|g \otimes_{q-p} g\|_{\mathfrak{H}^{\otimes 2p}}$$
$$+ \frac{p^2}{2}\sum_{r=1}^{p-1}(r-1)!^2\binom{p-1}{r-1}^2\binom{q-1}{r-1}^2(p+q-2r)!\left(\|f \otimes_{p-r} f\|^2_{\mathfrak{H}^{\otimes 2r}} + \|g \otimes_{q-r} g\|^2_{\mathfrak{H}^{\otimes 2r}}\right).$$

Remark 3.8.

1. *Recall that*

$$E(I_p(f)I_q(g)) = \begin{cases} p!\langle f, g\rangle_{\mathfrak{H}^{\otimes p}} & \text{if } p = q,\\ 0 & \text{otherwise.}\end{cases}$$

2. *In order to estimate the right-hand side of* (3.11), *we see that it is sufficient to asses the quantity $\|f_i \otimes_r f_i\|_{\mathfrak{H}^{\otimes 2(q_i - r)}}$ for any $i \in \{1, \dots, d\}$ and $r \in \{1, \dots, q_i - 1\}$ on the one hand, and $\langle f_i, f_j\rangle_{\mathfrak{H}^{\otimes q_i}}$ for any $1 \leq i, j \leq d$ such that $q_i = q_j$ on the other hand.*

Proof of Lemma 3.7. (see also [14], Lemma 2). Without loss of generality, we can assume that $\mathfrak{H} = L^2(A, \mathcal{A}, \mu)$, where $(A, \mathcal{A})$ is a measurable space, and μ is a σ-finite and non-atomic measure. Thus, we can write

$$\begin{aligned}
\langle DF, DG\rangle_{\mathfrak{H}} &= pq\big\langle I_{p-1}(f), I_{q-1}(g)\big\rangle_{\mathfrak{H}}\\
&= pq\int_A I_{p-1}\big(f(\cdot, t)\big)I_{q-1}\big(g(\cdot, t)\big)\mu(\mathrm{d}t)\\
&= pq\int_A \sum_{r=0}^{p\wedge q-1} r!\binom{p-1}{r}\binom{q-1}{r} I_{p+q-2-2r}\big(f(\cdot, t)\widetilde{\otimes}_r g(\cdot, t)\big)\mu(\mathrm{d}t)\\
&= pq\sum_{r=0}^{p\wedge q-1} r!\binom{p-1}{r}\binom{q-1}{r} I_{p+q-2-2r}(f\widetilde{\otimes}_{r+1} g)\\
&= pq\sum_{r=1}^{p\wedge q}(r-1)!\binom{p-1}{r-1}\binom{q-1}{r-1} I_{p+q-2r}(f\widetilde{\otimes}_r g).
\end{aligned}$$

It follows that

$$E\left[\left(a-\frac{1}{q}\langle DF,DG\rangle_{\mathfrak{H}}\right)^2\right]$$
$$=\begin{cases} a^2+p^2\sum_{r=1}^{p}(r-1)!^2\binom{p-1}{r-1}^2\binom{q-1}{r-1}^2(p+q-2r)!\|f\widetilde{\otimes}_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}} & \text{if } p<q,\\ \left(a-p!\langle f,g\rangle_{\mathfrak{H}^{\otimes p}}\right)^2+p^2\sum_{r=1}^{p-1}(r-1)!^2\binom{p-1}{r-1}^4(2p-2r)!\|f\widetilde{\otimes}_r g\|^2_{\mathfrak{H}^{\otimes(2p-2r)}} & \text{if } p=q.\end{cases} \tag{3.12}$$

If $r<p\le q$ then

$$\begin{aligned}\|f\widetilde{\otimes}_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}} &\le \|f\otimes_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}}=\langle f\otimes_{p-r}f, g\otimes_{q-r}g\rangle_{\mathfrak{H}^{\otimes 2r}}\\ &\le \|f\otimes_{p-r}f\|_{\mathfrak{H}^{\otimes 2r}}\|g\otimes_{q-r}g\|_{\mathfrak{H}^{\otimes 2r}}\\ &\le \frac{1}{2}\left(\|f\otimes_{p-r}f\|^2_{\mathfrak{H}^{\otimes 2r}}+\|g\otimes_{q-r}g\|^2_{\mathfrak{H}^{\otimes 2r}}\right).\end{aligned}$$

If $r=p<q$, then

$$\|f\widetilde{\otimes}_p g\|^2_{\mathfrak{H}^{\otimes(q-p)}}\le\|f\otimes_p g\|^2_{\mathfrak{H}^{\otimes(q-p)}}\le\|f\|^2_{\mathfrak{H}^{\otimes p}}\|g\otimes_{q-p}g\|_{\mathfrak{H}^{\otimes 2p}}.$$

If $r=p=q$, then $f\widetilde{\otimes}_p g=\langle f,g\rangle_{\mathfrak{H}^{\otimes p}}$. By plugging these last expressions into (3.12), we deduce immediately the desired conclusion. □

Let us now recall the following result, which is a collection of some of the findings contained in the papers by Peccati and Tudor [17] and Nualart and Ortiz-Latorre [14].

Theorem 3.9 (See [14,17]). *Fix $d\ge 2$ and let $C=\{C(i,j)\colon i,j=1,\dots,d\}$ be a $d\times d$ positive definite matrix. Fix integers $1\le q_1\le\cdots\le q_d$. For any $n\ge 1$ and $i=1,\dots,d$, let $f_i^{(n)}$ belong to $\mathfrak{H}^{\odot q_i}$. Assume that*

$$F^{(n)}=\left(F_1^{(n)},\dots,F_d^{(n)}\right):=\left(I_{q_1}\left(f_1^{(n)}\right),\dots,I_{q_d}\left(f_d^{(n)}\right)\right),\quad n\ge 1,$$

is such that

$$\lim_{n\to\infty}E\left[F_i^{(n)}F_j^{(n)}\right]=C(i,j),\quad 1\le i,j\le d. \tag{3.13}$$

Then, as $n\to\infty$, the following four assertions are equivalent:

(i) *For every $1\le i\le d$, $F_i^{(n)}$ converges in distribution to a centered Gaussian random variable with variance $C(i,i)$.*
(ii) *For every $1\le i\le d$, $E[(F_i^{(n)})^4]\to 3C(i,i)^2$.*
(iii) *For every $1\le i\le d$ and every $1\le r\le q_i-1$, $\|f_i^{(n)}\otimes_r f_i^{(n)}\|_{\mathfrak{H}^{\otimes 2(q_i-r)}}\to 0$.*
(iv) *The vector $F^{(n)}$ converges in distribution to a d-dimensional Gaussian vector $\mathcal{N}_d(0,C)$.*

Moreover, if $C(i,j)=\delta_{ij}$, where δ_{ij} is the Kronecker symbol, then either one of conditions (i)–(iv) *above is equivalent to the following*:

(v) *For every $1\le i\le d$, $\|DF_i^{(n)}\|^2_{\mathfrak{H}}\xrightarrow{L^2}q_i$.*

We conclude this section by pointing out the remarkable fact that, for vectors of multiple Wiener–Itô integrals of arbitrary length, *the Wasserstein distance metrizes the weak convergence towards a Gaussian vector with positive definite covariance*. Note that the next statement also contains a generalization of the multidimensional results proved in [14] to the case of an arbitrary covariance.

Proposition 3.10. *Fix $d \geq 2$, let C be a positive definite $d \times d$ symmetric matrix, and let $1 \leq q_1 \leq \cdots \leq q_d$. Consider vectors*

$$F^{(n)} := \left(F_1^{(n)}, \ldots, F_d^{(n)}\right) = \left(I_{q_1}\left(f_1^{(n)}\right), \ldots, I_{q_d}\left(f_d^{(n)}\right)\right), \quad n \geq 1,$$

with $f_i^{(n)} \in \mathfrak{H}^{\odot q_i}$ for every $i = 1, \ldots, d$. Assume moreover that $F^{(n)}$ satisfies condition (3.13). Then, as $n \to \infty$, the following three conditions are equivalent:

(a) $d_W(F^{(n)}, Z) \to 0$.

(b) *For every $1 \leq i \leq d$, $q_i^{-1} \| DF_i^{(n)} \|_{\mathfrak{H}}^2 \xrightarrow{L^2} C(i,i)$ and, for every $1 \leq i \neq j \leq d$, $\langle DF_i^{(n)}, -DL^{-1}F_j^{(n)} \rangle_{\mathfrak{H}} = q_j^{-1} \langle DF_i^{(n)}, DF_j^{(n)} \rangle_{\mathfrak{H}} \xrightarrow{L^2} C(i,j)$.*

(c) *$F^{(n)}$ converges in distribution to $Z \sim \mathcal{N}_d(0, C)$.*

Proof. Since convergence in the Wasserstein distance implies convergence in distribution, the implication (a) → (c) is trivial. The implication (b) → (a) is a consequence of relation (3.11). Now assume that (c) is verified, that is, $F^{(n)}$ converges in law to $Z \sim \mathcal{N}_d(0, C)$ as n goes to infinity. By Theorem 3.9 we have that, for any $i \in \{1, \ldots, d\}$ and $r \in \{1, \ldots, q_i - 1\}$,

$$\left\| f_i^{(n)} \otimes_r f_i^{(n)} \right\|_{\mathfrak{H}^{\otimes 2(q_i - r)}} \underset{n \to \infty}{\longrightarrow} 0.$$

By combining Corollary 3.6 with Lemma 3.7 (see also Remark 3.8(2)), one therefore easily deduces that, since (3.13) is in order, condition (b) must necessarily be satisfied. □

4. Applications

4.1. *Convergence of marginal distributions in the functional Breuer–Major CLT*

In this section, we use our main results in order to derive an explicit bound for the celebrated *Breuer–Major CLT* for fractional Brownian motion (fBm). We recall that a fBm $B = \{B_t \colon t \geq 0\}$, with Hurst index $H \in (0, 1)$, is a centered Gaussian process, started from zero and with covariance function $E(B_s B_t) = R(s, t)$, where

$$R(s,t) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t - s|^{2H}\right), \quad s, t \geq 0.$$

For any choice of the Hurst parameter $H \in (0, 1)$, the Gaussian space generated by B can be identified with an isonormal Gaussian process of the type $X = \{X(h) \colon h \in \mathfrak{H}\}$, where the real and separable Hilbert space $\mathfrak{H}$ is defined as follows: (i) denote by $\mathcal{E}$ the set of all $\mathbb{R}$-valued step functions on $[0, \infty)$, (ii) define $\mathfrak{H}$ as the Hilbert space obtained by closing $\mathcal{E}$ with respect to the scalar product $\langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]} \rangle_{\mathfrak{H}} = R(t, s)$. In particular, with such a notation, one has that $B_t = X(\mathbf{1}_{[0,t]})$. The reader is referred e.g. to [13] for more details on fBm, including crucial connections with fractional operators. We also define $\rho(\cdot)$ to be the covariance function associated with the stationary process $x \mapsto B_{x+1} - B_x$ ($x \in \mathbb{R}$), that is

$$\rho(x) := \frac{1}{2}\left(|x+1|^{2H} + |x-1|^{2H} - 2|x|^{2H}\right) \underset{|x| \to \infty}{\sim} H|2H - 1||x|^{2H-2}.$$

Now fix an integer $q \geq 2$, assume that $H < 1 - \frac{1}{2q}$ and set

$$S_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{k=0}^{\lfloor nt \rfloor - 1} H_q(B_{k+1} - B_k), \quad t \geq 0,$$

where H_q is the qth Hermite polynomial defined as

$$H_q(x) = (-1)^q \mathrm{e}^{x^2/2} \frac{\mathrm{d}^q}{\mathrm{d}x^q} \mathrm{e}^{-x^2/2}, \quad x \in \mathbb{R}, \tag{4.1}$$

and where $\sigma = \sqrt{q! \sum_{r\in\mathbb{Z}} \rho^2(r)}$. According e.g. to the main results in [2] or [7], one has the following CLT:

$$\{S_n(t), t \ge 0\} \xrightarrow[n\to\infty]{\text{f.d.d.}} \text{standard Brownian motion},$$

where "f.d.d." indicates convergence in the sense of finite-dimensional distributions. To our knowledge, the following statement contains the first multidimensional bound for the Wasserstein distance ever proved for $\{S_n(t), t \ge 0\}$.

Theorem 4.1. *For any fixed $d \ge 1$ and $0 = t_0 < t_1 < \cdots < t_d$, there exists a constant c, (depending only on d, H and $(t_0, t_1, \ldots, t_d)$, and not on n) such that, for every $n \ge 1$:*

$$d_{\mathrm{W}}\left(\left(\frac{S_n(t_i) - S_n(t_{i-1})}{\sqrt{t_i - t_{i-1}}}\right)_{1\le i\le d}; \mathscr{N}_d(0, \mathbf{I}_d)\right) \le c \times \begin{cases} n^{-1/2} & \text{if } H \in (0, \frac{1}{2}], \\ n^{H-1} & \text{if } H \in (\frac{1}{2}, \frac{2q-3}{2q-2}], \\ n^{qH-q+1/2} & \text{if } H \in (\frac{2q-3}{2q-2}, \frac{2q-1}{2q}). \end{cases}$$

Proof. Fix $d \ge 1$ and $t_0 = 0 < t_1 < \cdots < t_d$. In the sequel, c will denote a constant independent of n, which can differ from one line to another.

First, observe that

$$\frac{S_n(t_i) - S_n(t_{i-1})}{\sqrt{t_i - t_{i-1}}} = I_q\big(f_i^{(n)}\big) \quad \text{with } f_i^{(n)} = \frac{1}{\sigma\sqrt{n}\sqrt{t_i - t_{i-1}}} \sum_{k=\lfloor nt_{i-1}\rfloor}^{\lfloor nt_i\rfloor - 1} \mathbf{1}_{[k,k+1]}^{\otimes q}.$$

In [11], proof of Theorem 4.1, it is shown that, for any $i \in \{1, \ldots, d\}$ and $r \in \{1, \ldots, q_i - 1\}$:

$$\big\| f_i^{(n)} \otimes_r f_i^{(n)} \big\|_{\mathfrak{H}^{\otimes 2(q_i - r)}} \le c \times \begin{cases} n^{-1/2} & \text{if } H \in (0, \frac{1}{2}], \\ n^{H-1} & \text{if } H \in (\frac{1}{2}, \frac{2q-3}{2q-2}], \\ n^{qH-q+1/2} & \text{if } H \in (\frac{2q-3}{2q-2}, \frac{2q-1}{2q}). \end{cases} \tag{4.2}$$

Moreover, when $1 \le i < j \le d$, we have:

$$\begin{aligned}
&\big|\langle f_i^{(n)}, f_j^{(n)}\rangle_{\mathfrak{H}^{\otimes q}}\big| \\
&\quad = \left| \frac{1}{\sigma^2 n \sqrt{t_i - t_{i-1}}\sqrt{t_j - t_{j-1}}} \sum_{k=\lfloor nt_{i-1}\rfloor}^{\lfloor nt_i\rfloor - 1} \sum_{l=\lfloor nt_{j-1}\rfloor}^{\lfloor nt_j\rfloor - 1} \rho^q(l-k) \right| \\
&\quad = \frac{c}{n} \left| \sum_{|r|=\lfloor nt_{j-1}\rfloor - \lfloor nt_i\rfloor + 1}^{\lfloor nt_j\rfloor - \lfloor nt_{i-1}\rfloor - 1} \big[\big(\lfloor nt_j\rfloor - 1 - r\big) \wedge \big(\lfloor nt_i\rfloor - 1\big) - \big(\lfloor nt_{j-1}\rfloor - r\big) \vee \big(\lfloor nt_{i-1}\rfloor\big)\big]\rho^q(r) \right| \\
&\quad \le c \frac{\lfloor nt_i\rfloor - \lfloor nt_{i-1}\rfloor - 1}{n} \sum_{|r| \ge \lfloor nt_{j-1}\rfloor - \lfloor nt_i\rfloor + 1} \big|\rho(r)\big|^q = \mathrm{O}\big(n^{2qH-2q+1}\big) \quad \text{as } n \to \infty,
\end{aligned} \tag{4.3}$$

the last equality coming from

$$\sum_{|r|\ge N} \big|\rho(r)\big|^q = \mathrm{O}\left(\sum_{|r|\ge N} |r|^{2qH-2q}\right) = \mathrm{O}\big(N^{2qH-2q+1}\big), \quad \text{as } N \to \infty.$$

Finally, by combining (4.2), (4.3), Corollary 3.6 and Lemma 3.7, we obtain the desired conclusion. □

4.2. *Vector-valued functionals of finite Gaussian sequences*

Let $Y = (Y_1, \dots, Y_n) \sim \mathscr{N}_n(0, \mathbf{I}_n)$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be an absolutely continuous function such that f and its partial derivatives have subexponential growth at infinity. The following result has been proved by Chatterjee in [3], in the context of limit theorems for linear statistics of eigenvalues of random matrices. We use the notation d_{TV} to indicate the *total variation distance* between laws of real valued random variables.

Proposition 4.2 (Lemma 5.3 in [3]). *Assume that the random variable $W = f(Y)$ has zero mean and unit variance, and denote by $Z \sim \mathscr{N}(0,1)$ a standard Gaussian random variable. Then $d_{\mathrm{TV}}(W, Z) \le 2\operatorname{Var}(T(Y))^{1/2}$, where the function $T(\cdot)$ is defined as*

$$T(y) = \int_0^1 \frac{1}{2\sqrt{t}} \sum_{i=1}^n E\left[\frac{\partial f}{\partial y_i}(y)\frac{\partial f}{\partial y_i}\big(\sqrt{t}y + \sqrt{1-t}Y\big)\right] \mathrm{d}t.$$

In what follows, we shall use Theorem 3.5 in order to deduce a multidimensional generalization of Proposition 4.2 (with the Wasserstein distance replacing total variation).

Proposition 4.3. *Let $Y \sim \mathscr{N}_n(0, K)$, where $K = \{K(i,l)\colon i, l = 1, \dots, n\}$ is a $n \times n$ positive definite matrix. Consider absolutely continuous functions $f_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, d$. Assume that each random variable $f_j(Y)$ has zero mean, and also that each function f_j and its partial derivatives have subexponential growth at infinity. Denote by $Z \sim \mathscr{N}_d(0, C)$ a Gaussian vector with values in $\mathbb{R}^d$ and with positive definite covariance matrix $C = \{C(a,b)\colon a, b = 1, \dots, d\}$. Finally, write $W = (W_1, \dots, W_d) = (f_1(Y), \dots, f_d(Y))$. Then,*

$$d_{\mathrm{W}}(W, Z) \le \|C^{-1}\|_{\mathrm{op}} \|C\|_{\mathrm{op}}^{1/2} \sqrt{\sum_{a,b=1}^d E\big[\big(C(a,b) - T_{ab}(Y)\big)^2\big]},$$

where the functions $T_{ab}(\cdot)$ are defined as

$$T_{ab}(y) = \int_0^1 \frac{1}{2\sqrt{t}} \sum_{i,j=1}^n K(i,j) E\left[\frac{\partial f_a}{\partial y_i}(y)\frac{\partial f_b}{\partial y_j}\big(\sqrt{t}y + \sqrt{1-t}Y\big)\right] \mathrm{d}t.$$

Proof. Without loss of generality, we can assume that $Y = (Y_1, \dots, Y_n) = (X(h_1), \dots, X(h_n))$, where X is an isonormal Gaussian process over some Hilbert space $\mathfrak{H}$, and $\langle h_i, h_l\rangle_{\mathfrak{H}} = E(X(h_i)X(h_l)) = K(i,l)$. According to Theorem 3.5, it is therefore sufficient to show that, for every $a, b = 1, \dots, d$, $T_{ab}(Y) = \langle DW_a, -DL^{-1}W_b\rangle_{\mathfrak{H}}$. To prove this last claim, introduce the two $\mathfrak{H}$-valued functions $\Theta_a(y)$ and $\Theta_b(y)$, defined for $y \in \mathbb{R}^d$ as follows:

$$\Theta_a(y) = \sum_{i=1}^n \frac{\partial f_a}{\partial y_i}(y) h_i \quad \text{and} \quad \Theta_b(y) = \int_0^1 \frac{1}{2\sqrt{t}} \sum_{j=1}^n \left\{E\left[\frac{\partial f_b}{\partial y_j}\big(\sqrt{t}y + \sqrt{1-t}Y\big)\right] h_j\right\} \mathrm{d}t.$$

By using (2.2), it is easily seen that $\Theta_a(Y) = DW_a$. Moreover, by using e.g. formula (3.46) in [11], one deduces that $\Theta_b(Y) = -DL^{-1}W_b$. Since $T_{ab}(Y) = \langle \Theta_a(Y), \Theta_b(Y)\rangle_{\mathfrak{H}}$, the conclusion is immediately obtained. □

By specializing the previous statement to the case $n = d$ and $f_j(y) = y_j$, $j = 1, \dots, d$, one obtains the following simple bound on the Wasserstein distance between Gaussian vectors of the same dimension (the proof is straightforward and omitted).

Corollary 4.4. *Let $Y \sim \mathscr{N}_d(0, K)$ and $Z \sim \mathscr{N}_d(0, C)$, where K and C are two positive definite covariance matrices. Then $d_{\mathrm{W}}(Y, Z) \le Q(C, K) \times \|C - K\|_{\mathrm{H.S.}}$, where*

$$Q(C, K) := \min\big\{\|C^{-1}\|_{\mathrm{op}} \|C\|_{\mathrm{op}}^{1/2}, \|K^{-1}\|_{\mathrm{op}} \|K\|_{\mathrm{op}}^{1/2}\big\}.$$

Acknowledgments

We thank an anonymous referee for insightful remarks. Part of this paper has been written while I. Nourdin and G. Peccati were visiting the Institute of Mathematical Sciences of the National University of Singapore, in the occasion of the "Workshop on Stein's Method," from March 31 until April 4, 2008. These authors heartily thank A. Barbour, L. Chen, K.-P. Choi and A. Xia for their kind hospitality and support.

References

[1] A. D. Barbour. Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** (1990) 297–322. MR1035659
[2] P. Breuer and P. Major. Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.* **13** (1983) 425–441. MR0716933
[3] S. Chatterjee. Fluctuation of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** (2009) 1–40. MR2449121
[4] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA* **4** (2008) 257–283.
[5] L. Chen and Q.-M. Shao. Stein's method for normal approximation. In *An Introduction to Stein's Method* 1–59. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4**. Singapore Univ. Press, Singapore, 2005. MR2235448
[6] R. M. Dudley. *Real Analysis and Probability*, 2nd edition. Cambridge Univ. Press, Cambridge, 2003. MR1932358
[7] L. Giraitis and D. Surgailis. CLT and other limit theorems for functionals of Gaussian processes. *Z. Wahrsch. Verw. Gebiete* **70** (1985) 191–212. MR0799146
[8] F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** (1991) 724–739. MR1106283
[9] E. P. Hsu. Characterization of Brownian motion on manifolds through integration by parts. In *Stein's Method and Applications* 195–208. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5**. Singapore Univ. Press, Singapore, 2005. MR2205337
[10] I. Nourdin and G. Peccati. Non-central convergence of multiple integrals. *Ann. Probab.* To appear.
[11] I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*. To appear.
[12] I. Nourdin and G. Peccati. Stein's method and exact Berry–Esséen asymptotics for functionals of Gaussian fields. Preprint, 2008.
[13] D. Nualart. *The Malliavin Calculus and Related Topics*, 2nd edition. Springer, Berlin, 2006. MR2200233
[14] D. Nualart and S. Ortiz-Latorre. Central limit theorem for multiple stochastic integrals and Malliavin calculus. *Stochastic Process. Appl.* **118** (2008) 614–628. MR2394845
[15] D. Nualart and G. Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** (2005) 177–193. MR2118863
[16] G. Peccati. Gaussian approximations of multiple integrals. *Electron. Comm. Probab.* **12** (2007) 350–364 (electronic). MR2350573
[17] G. Peccati and C. A. Tudor. Gaussian limits for vector-valued multiple stochastic integrals. In *Séminaire de Probabilités XXXVIII* 247–262. *Lecture Notes in Math.* **1857**. Springer, Berlin, 2005. MR2126978
[18] G. Reinert and A. Röllin. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. Preprint, 2007.
[19] Ch. 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* (*Univ. California, Berkeley, Calif., 1970/1971*)*, Vol. II: Probability Theory* 583–602. California Univ. Press, Berkeley, CA, 1972. MR0402873
[20] Ch. Stein. *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes, Monograph Series* **7**. Inst. Math. Statist., Hayward, CA, 1986. MR0882007