Quantitative CLTs on a Gaussian space: a survey of recent developments

Giovanni Peccati

December 22, 2013

Abstract

I will provide a short survey of recent findings concerning normal approximations on a Gaussian space. The results discussed in this work involve Stein's method, Poincaré-type inequalities, as well as the use of techniques from information theory. The guiding example involves 'exploding Brownian functionals', that are used as a tool for enhancing the reader's intuition. This expository paper mirrors part of the talk that I delivered in Clermont-Ferrand in August 2012, in the occasion of the 'Journées MAS 2012'.

1   Introduction

This work is a survey of a recent and very active direction of research, concerning limit theorems and probabilistic approximations for random variables that are functionals of an infinite-dimensional Gaussian field. The main idea developed below is that, in order to deal with these problems, one can very effectively use the power and flexibility of the *Malliavin calculus* of variations (see [32, 40]). The monograph [32] provides a quite detailed introduction to the main object of the present paper, with specific emphasis on the so-called 'fourth moment theorems' (see [42]), as well as on the interaction between Malliavin calculus, Stein's method and Poincaré-type estimates. Since the publication of [32] some new techniques have been introduced (see [28, 37, 38]) — in particular related to Carbery-Wright inequalities (see [6]) and tools from information theory. One of my implicit aims is to provide a quick introduction to these recent developments. See also the paper [19].

I will start this survey (see Section 2) with the discussion of a very specific example, that is, 'exploding' quadratic functionals of a standard Brownian motion living on a compact interval. This example allows one to easily compare the new techniques based on Malliavin calculus with some classical tools, such as the *method of moments and cumulants* and the technique of *random time changes*. I was actually introduced to the latter approach by Marc Yor while writing our joint papers [47, 48], where we dealt with limit theorems for non-linear functionals of Brownian local times. It is fair to say that these two references contain the seed of the body of research described in the present survey.

It is important to mention three additional (and extremely fruitful) directions of research that are connected to Malliavin techniques and fourth moment theorems: the first one concerns limit theorems on a Poisson space, with powerful applications in stochastic geometry (see e.g. [5, 44, 49, 52]); the second one deals with limit theorems for stochastic

1

integrals in the framework of free probability (see e.g. [13, 22]); the third one deals with fourth moment theorems for random variables living in the chaos of a general Markov diffusion generator (see [2, 3, 24]). The reader can consult the constantly updated webpage

http://www.iecn.u-nancy.fr/~nourdin/steinmalliavin.htm

for many applications of Malliavin-type techniques, as well as for asymptotic results that are somehow connected with the fourth moment theorem proved in [42].

For the rest of the paper, every random object is assumed to be defined on an adequate common probability space $(\Omega, \mathscr{F}, P)$, with E indicating expectation with respect to P.

2 Introductory example: exploding Brownian functionals

As a simple illustration of the problems we are interested in, we shall present a typical situation where one can take advantage of Malliavin calculus and associated techniques, that is: the asymptotic study of the quadratic functionals of a standard Brownian motion. We will first state a general problem, and then describe two popular methods of solution (along with their drawbacks) that are not based on Malliavin calculus. The second part of the paper will demonstrate how the use of Malliavin calculus can overcome the disadvantages of both approaches. This example was also discussed in the unpublished document [43].

In what follows, the notion of *cumulant* is sometimes used. Recall that, given a random variable Y with finite moments of all orders and with characteristic function $\psi_Y(t) = E[\exp(itY)]$ $(t \in \mathbb{R})$, one defines the sequence of cumulants (sometimes known as *semi-invariants*) of Y, noted $\{\chi_n(Y) : n \geqslant 1\}$, as

$$\chi_n(Y) = (-i)^n \frac{d^n}{dt^n} \log \psi_Y(t) \mid_{t=0}, \quad n \geqslant 1. \tag{2.1}$$

For instance, $\chi_1(Y) = E(Y)$, $\chi_2(Y) = E[Y - E(Y)]^2 = \mathbf{Var}(Y)$,

$$\chi_3(Y) = E(Y^3) - 3E(Y^2)E(Y) + 2E(Y)^3,$$

and so on. In general, one deduces from (2.1) that, for every $n \geqslant 1$, the first n moments of Y can be expressed as polynomials in the first n cumulants (and vice versa). Note that (2.1) also implies that the cumulants of order $n \geqslant 3$ of a Gaussian random variable are equal to zero (recall also that the Gaussian distribution is determined by its moments, and therefore by its cumulants). We refer the reader to [45, Chapter 3] for a self-contained introduction to the basic combinatorial properties of cumulants. See [32, Chapter 8] for an approach to cumulant computations based on Malliavin operators.

2.1 Statement of the problem

Let $W = \{W_t : t \geqslant 0\}$ be a standard Brownian motion started from zero. This means that W is a centered Gaussian process such that $W_0 = 0$, W has continuous paths, and $E[W_t W_s] = t \wedge s$ for every $t, s \geqslant 0$ (see e.g. [53]). We are interested in a specific property of the paths of W (first pointed out, in a slightly different form, in [20]), namely that

$$\int_0^1 \frac{W_t^2}{t^2} dt = \infty, \quad \text{a.s.-}P. \tag{2.2}$$

As discussed e.g. in [20, 47, 48], relation (2.2) has deep connections with the theory of the (Gaussian) *initial enlargements of filtrations* in continuous-time stochastic calculus, as well as with the asymptotic analysis of non-linear functionals of Brownian local times.

Remark 2.1 Define the process $\hat{W}$ as $\hat{W}_0 = 0$ and $\hat{W}_u = uW_{1/u}$ for $u > 0$. A trivial covariance computation shows that $\hat{W}$ is also a standard Brownian motion. By using the change of variable $u = 1/t$, it now follows that property (2.2) is equivalent to the following statement:

$$\int_1^\infty \frac{W_u^2}{u^2} du = \infty, \quad \text{a.s.-}P.$$

A natural question arising from (2.2) is of course how to characterize the 'rate of explosion', as $\varepsilon \to 0$, of the quantities

$$B_\varepsilon = \int_\varepsilon^1 \frac{W_t^2}{t^2} dt, \quad \varepsilon \in (0,1). \tag{2.3}$$

One typical answer can be obtained by proving that some suitable renormalization of B_ε converges in distribution to a standard Gaussian random variable. By a direct computation, one can prove that $E[B_\varepsilon] = \log 1/\varepsilon$ and $\mathbf{Var}(B_\varepsilon) \approx 4 \log 1/\varepsilon$ *. By setting

$$\widetilde{B}_\varepsilon = \frac{B_\varepsilon - \log 1/\varepsilon}{\sqrt{4 \log 1/\varepsilon}}, \quad \varepsilon \in (0,1), \tag{2.4}$$

one can therefore meaningfully state the following problem.

Problem I. *Prove that, as $\varepsilon \to 0$,*

$$\widetilde{B}_\varepsilon \xrightarrow{\text{Law}} N \sim \mathscr{N}(0,1), \tag{2.5}$$

where, here and for the rest of the paper, $\mathscr{N}(\alpha, \beta)$ denotes a one-dimensional Gaussian distribution with mean α and variance $\beta > 0$.

We shall solve Problem I by using both the classic *method of cumulants* and a stochastic calculus technique, known as *random time-change*. We will see below that both approaches suffer of evident drawbacks, and also that these difficulties can be successfully eliminated by using tools of infinite-dimensional Gaussian analysis.

2.2 The method of cumulants

The *method of (moments and) cumulants* is a very popular approach to the proof of limit results involving non-linear functionals of Gaussian fields. Its success relies mainly on the following two facts: (i) square-integrable functionals of Gaussian fields can always be represented in terms of (possibly infinite) series of multiple Wiener-Itô integrals (see the discussion below, as well as [32]), and (ii) moments and cumulants of multiple integrals can be computed (at least formally) by means of well-established combinatorial devices, known as *diagram formulae*. See [32, 45], for a detailed introduction to these topics and for a discussion of the relevant literature.

In order to apply the method of cumulants to the proof of (2.4), one should start with the classic Itô formula $W_t^2 = 2 \int_0^t W_s dW_s + t$, $t \in [0,1]$, and then write

$$\widetilde{B}_\varepsilon = \frac{B_\varepsilon - \log 1/\varepsilon}{\sqrt{4 \log 1/\varepsilon}} = \frac{2 \int_\varepsilon^1 \left[\int_0^t W_s dW_s \right] t^{-2} dt}{\sqrt{4 \log 1/\varepsilon}}, \quad \varepsilon \in (0,1). \tag{2.6}$$

---

*In what follows, we shall write $\gamma(\varepsilon) \approx \varphi(\varepsilon)$, whenever $\frac{\gamma(\varepsilon)}{\varphi(\varepsilon)} \to 1$, as $\varepsilon \to 0$.

It is a standard result of stochastic calculus that one can interchange deterministic and stochastic integration on the right-hand-side of (2.6) as follows:

$$
\begin{aligned}
\int_\varepsilon^1 \left[\int_0^t W_s dW_s \right] \frac{dt}{t^2} &= \int_\varepsilon^1 \left[\int_0^1 \mathbf{1}_{\{s<t\}} W_s dW_s \right] \frac{dt}{t^2} \\
&= \int_0^1 \left[\int_\varepsilon^1 \mathbf{1}_{\{s<t\}} \frac{dt}{t^2} \right] W_s dW_s \\
&= \int_0^1 \left[(s \vee \varepsilon)^{-1} - 1 \right] W_s dW_s.
\end{aligned}
$$

As a consequence,

$$
\widetilde{B}_\varepsilon = \frac{2 \int_0^1 \left[(s \vee \varepsilon)^{-1} - 1 \right] W_s dW_s}{\sqrt{4 \log 1/\varepsilon}} = 2 \int_0^1 \int_0^s f_\varepsilon(s,u) \, dW_u dW_s, \tag{2.7}
$$

where f_ε is the symmetric and Lebesgue square-integrable function on $[0,1]^2$ given by

$$
f_\varepsilon(s,u) = 2 \left[(s \vee u \vee \varepsilon)^{-1} - 1 \right] \times (4 \log 1/\varepsilon)^{-1/2}. \tag{2.8}
$$

By anticipating the terminology introduced later on in the paper, formula (2.7) simply implies that each random variable $\widetilde{B}_\varepsilon$ is a member of the *second Wiener chaos* associated with W. One can now combine this fact with the results discussed e.g. in [32, Chapter 3], to deduce that, since the application $\varepsilon \mapsto \mathbf{Var}(\widetilde{B}_\varepsilon)$ is bounded, then, for every $n \geqslant 2$,

$$
\sup_{\varepsilon > 0} E|\widetilde{B}_\varepsilon|^n < \infty. \tag{2.9}
$$

Since $E(\widetilde{B}_\varepsilon) = 0$ and $\mathbf{Var}(\widetilde{B}_\varepsilon) \to 1$, relation (2.9) implies immediately that (2.5) is proved once it is shown that, as $\varepsilon \to 0$,

$$
\chi_n(\widetilde{B}_\varepsilon) \to 0, \quad \text{for every } n \geqslant 3. \tag{2.10}
$$

To prove (2.10) we make use of a well-known combinatorial result (see e.g. [45, p. 139] for a proof), stating that, for every fixed $n \geqslant 3$, the nth cumulant of $\widetilde{B}_\varepsilon$ is given by the following 'circular integral'

$$
\chi_n(\widetilde{B}_\varepsilon) = 2^{n-1} (n-1)! \int_{[0,1]^n} f_\varepsilon(t_1, t_2) f_\varepsilon(t_2, t_3) \cdots f_\varepsilon(t_{n-1}, t_n) f_\varepsilon(t_n, t_1) \, dt_1 \cdots dt_n, \tag{2.11}
$$

obtained by juxtaposing n copies of the kernel f_ε. By plugging (2.8) into (2.11), and after some lengthy (but standard) computations, one obtains that, as $\varepsilon \to 0$,

$$
\chi_n(\widetilde{B}_\varepsilon) \approx c_n \times \left(\log \frac{1}{\varepsilon} \right)^{1 - \frac{n}{2}}, \quad \text{for every } n \geqslant 3, \tag{2.12}
$$

where $c_n > 0$ is a finite constant independent of ε. This yields (2.10) and therefore (2.5). The implication (2.12) $\Rightarrow$ (2.10) $\Rightarrow$ (2.5) is a typical application of the method of cumulants to the proof of Central Limit Theorems (CLTs) for functionals of Gaussian fields.

In the following list we pinpoint some of the main disadvantages of this approach.

D1 Formulae (2.11) and (2.12) characterize the speed of convergence to zero of the cumulants of $\widetilde{B}_\varepsilon$. However, there is no way to deduce from (2.12) an estimate for quantities of the type $d\left(\widetilde{B}_\varepsilon, N\right)$, where d indicates some distance between the law of $\widetilde{B}_\varepsilon$ and the law of N. [†]

[†] This assertion is not accurate, although it is kept for dramatic effect. Indeed, we will show below that the use of Malliavin calculus <u>exactly</u> allows to deduce Berry-Esséen bounds from estimates on cumulants.

D2 Relations (2.10) and (2.11) require that, in order to prove the CLT (2.5), one verifies an *infinity* of asymptotic relations, each one involving the estimate of a multiple deterministic integral of increasing order. This task can be computationally quite demanding. Here, (2.12) is obtained by exploiting the elementary form of the kernels f_ε in (2.8).

D3 If one wants to apply the method of cumulants to elements of higher chaoses (for instance, by considering functionals involving Hermite polynomials of degree greater than 3), then one is forced to use diagram formulae that are much more involved than the neat identity (2.11). Striking examples of this situation appear e.g. when dealing with non-linear functionals of random fields defined on homogeneous spaces — see the recent monograph [25].

2.3   Random time-changes

As anticipated, this technique has been successfully used in [47, 48], and also plays an important role in the proof of the 'fourth moment theorem' given in reference [42]. Our starting point is once again formula (2.7), implying that, for each $\varepsilon \in (0,1)$, the random variable $\widetilde{B}_\varepsilon$ coincides with the value at the point $t = 1$ of the continuous Brownian martingale

$$t \mapsto M_t^\varepsilon = 2 \int_0^t \int_0^s f_\varepsilon(s,u)\, dW_u dW_s, \quad t \in [0,1]. \tag{2.13}$$

It is well-known that the martingale M_t^ε has a quadratic variation equal to

$$\langle M^\varepsilon, M^\varepsilon \rangle_t = 4 \int_0^t \left(\int_0^s f_\varepsilon(s,u)\, dW_u \right)^2 ds, \quad t \in [0,1].$$

By virtue of a classic stochastic calculus result, known as the *Dambis-Dubins-Schwarz Theorem* (DDS Theorem; see [53, Chapter V]), for every $\varepsilon \in (0,1)$ there exists (on a possibly enlarged probability space) a standard Brownian motion β^ε, initialized at zero and such that

$$M_t^\varepsilon = \beta_{\langle M^\varepsilon, M^\varepsilon \rangle_t}^\varepsilon, \quad t \in [0,1]. \tag{2.14}$$

It is important to notice that the definition of β^ε strongly depends on ε, and that β^ε is in general *not adapted* to the natural filtration of W. Moreover, one has that there exists a (continuous) filtration $\mathcal{G}_s^\varepsilon$, $s \geqslant 0$, such that β_s^ε is a $\mathcal{G}_s^\varepsilon$-Brownian motion and (for every fixed t) the positive random variable $\langle M^\varepsilon, M^\varepsilon \rangle_t$ is a $\mathcal{G}_s^\varepsilon$-*stopping time*. Formula (2.14) yields in particular that

$$\widetilde{B}_\varepsilon = M_1^\varepsilon = \beta_{\langle M^\varepsilon, M^\varepsilon \rangle_1}^\varepsilon.$$

Now consider a Lipschitz function h such that $\|h'\|_\infty \leqslant 1$, and observe that, for every $\varepsilon > 0$, $\beta_1^\varepsilon \overset{\text{Law}}{=} N \sim \mathcal{N}(0,1)$. A careful application of the Burkholder-Davis-Gundy (BDG)

inequality (in the version stated in [53, Corollary IV.4.2]) yields the following estimates:

$$
\begin{aligned}
\left| E[h(\widetilde{B}_\varepsilon)] - E\left[h\left(N\right)\right] \right| &= \left| \mathbb{E}[h(\beta^\varepsilon_{\langle M^\varepsilon, M^\varepsilon \rangle_1})] - E\left[h\left(\beta^\varepsilon_1\right)\right] \right| & (2.15) \\
&\leqslant E\left[\left| \beta^\varepsilon_{\langle M^\varepsilon, M^\varepsilon \rangle_1} - \beta^\varepsilon_1 \right| \right] \\
&\leqslant E\left[\left| \beta^\varepsilon_{\langle M^\varepsilon, M^\varepsilon \rangle_1} - \beta^\varepsilon_1 \right|^4 \right]^{\frac{1}{4}} \\
&\leqslant C \times E\left[\left| \langle M^\varepsilon, M^\varepsilon \rangle_1 - 1 \right|^2 \right]^{\frac{1}{4}} \\
&= C \times E\left[\left| 4\int_0^1 \left(\int_0^s f_\varepsilon\left(s,u\right) dW_u \right)^2 ds - 1 \right|^2 \right]^{\frac{1}{4}},
\end{aligned}
$$

where C is some universal constant independent of ε. The CLT (2.5) is now obtained from (2.15) by means of a direct computation, yielding that, as $\varepsilon \to 0$,

$$
E\left[\left| 4\int_0^1 \left(\int_0^s f_\varepsilon\left(s,u\right) dW_u \right)^2 ds - 1 \right|^2 \right] \approx \frac{\alpha}{\log 1/\varepsilon} \to 0, \tag{2.16}
$$

where $\alpha > 0$ is some constant independent of ε.

Note that this approach is more satisfactory than the method of cumulants. Indeed, the chain of relations starting at (2.15) allows one to assess explicitly the Wasserstein distance between the law of $\widetilde{B}_\varepsilon$ and the law of N [‡] (albeit the implied rate of $(\log 1/\varepsilon)^{-1/4}$ is suboptimal). Moreover, the proof of (2.5) is now reduced to a single asymptotic relation, namely (2.16). However, at least two crucial points make this approach quite difficult to apply in general situations.

D4 The application of the DDS Theorem and of the BDG inequality requires an explicit underlying (Brownian) martingale structure. Although it is always possible to represent a given Gaussian field in terms of a Brownian motion, this operation is often quite unnatural and can render the asymptotic analysis very hard. For instance, what happens if one considers quadratic functionals of a multiparameter Gaussian process, or of a Gaussian process which is not a semimartingale (for instance, a fractional Brownian motion with Hurst parameter $H \neq 1/2$)?

D5 It is not clear whether this approach can be used in order to deal with expressions of the type (2.15), when h is not Lipschitz (for instance, when h equals the indicator of a Borel set), so that it seems difficult to use these techniques in order to assess other distances, like the *total variation distance* or the *Kolmogorov distance*.

Before dealing with the principal matter of the paper, we will now briefly describe the main objects and tools of stochastic analysis that are needed in the following.

3 Relevant elements of Gaussian analysis

One of the quickest way to introduce Malliavin type operators is to define them in terms of some countable collection of independent and identically distributed Gaussian $\mathscr{N}\left(0,1\right)$

[‡]Recall that the Wasserstein distance between the law of two variables X_1, X_2 is given by $d_W\left(X_1, X_2\right) = \sup \left| E\left[h\left(X_1\right)\right] - E\left[h\left(X_2\right)\right] \right|$, where the supremum is taken over all Lipschitz functions such that $\|h'\|_\infty \leqslant 1$. See the discussion below.

random variables, that we shall denote by

$$\mathbf{G} = \left\{ G_i : i \geqslant 1 \right\}. \tag{3.17}$$

By definition, $\mathbf{G}$ is a Gaussian family such that $E[G_i] = 0$ for every i and $E[G_i G_j] = \mathbf{1}_{\{i=j\}}$. We will write $L^2(\sigma(\mathbf{G})) := L^2(P, \sigma(\mathbf{G}))$ to indicate the class of square-integrable (real-valued) random variables that are measurable with respect to the σ-field generated by $\mathbf{G}$. Appropriate references for the material presented below are e.g. [32, 40].

3.1 Wiener chaos

The following definition is standard.

Definition 3.1 (Hermite polynomials and Wiener chaos)

1. *The sequence of Hermite polynomials $\{H_m : m \geqslant 0\}$ is defined as follows: $H_0 = 1$, and, for $m \geqslant 1$,*

$$H_m(x) = (-1)^m e^{\frac{x^2}{2}} \frac{d^m}{dx^m} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

The class $\{(m!)^{-1/2} H_m : m \geqslant 0\}$ is an orthonormal basis of $L^2(\mathbb{R}, \phi(x)dx)$, where

$$\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2), \quad x \in \mathbb{R},$$

is the standard Gaussian density.

2. *A multi-index $\alpha = \{\alpha_i : i \geqslant 1\}$ is a sequence of nonnegative integers such that $\alpha_i \neq 0$ only for a finite number of indices i. We use the symbol Λ in order to indicate the collection of all multi-indices, and we set $|\alpha| := \sum_{i \geqslant 1} \alpha_i$, for every $\alpha \in \Lambda$.*

3. *For every integer $q \geqslant 0$, the qth Wiener chaos associated with $\mathbf{G}$ is defined as follows: $C_0 = \mathbb{R}$, and, for $q \geqslant 1$, C_q is the $L^2(\sigma(\mathbf{G}))$-closed vector space generated by random variables of the type*

$$\Phi(\alpha) = \prod_{i=1}^{\infty} H_{\alpha_i}(G_i), \quad \alpha \in \Lambda \ \text{and} \ |\alpha| = q. \tag{3.18}$$

It is clear (from the orthogonality properties of Hermite polynomials, as well as from the independence assumptions) that two random variables belonging to Wiener chaoses of different orders are orthogonal in $L^2(\sigma(\mathbf{G}))$. Moreover, since linear combinations of polynomials are dense in $L^2(\sigma(\mathbf{G}))$, one has that $L^2(\sigma(\mathbf{G})) = \bigoplus_{q \geqslant 0} C_q$, that is, any square-integrable functional of $\mathbf{G}$ can be written as an infinite series, converging in L^2 and such that the qth summand is an element of C_q. This orthogonal decomposition of $L^2(\sigma(\mathbf{G}))$ is customarily called the *Wiener-Itô chaotic decomposition* of $L^2(\sigma(\mathbf{G}))$.

It is useful to encode random variables in the spaces C_q by means of increasing tensor powers of Hilbert spaces. To do this, introduce an arbitrary separable real Hilbert space $\mathfrak{H}$ having an orthonormal basis $\{e_i : i \geqslant 1\}$. For $q \geqslant 2$, denote by $\mathfrak{H}^{\otimes q}$ (resp. $\mathfrak{H}^{\odot q}$) the qth tensor power (resp. symmetric tensor power) of $\mathfrak{H}$; write moreover $\mathfrak{H}^{\otimes 0} = \mathfrak{H}^{\odot 0} = \mathbb{R}$ and $\mathfrak{H}^{\otimes 1} = \mathfrak{H}^{\odot 1} = \mathfrak{H}$. With every multi-index $\alpha \in \Lambda$, one associates the tensor $e(\alpha) \in \mathfrak{H}^{\otimes |\alpha|}$ given by

$$e(\alpha) = e_{i_1}^{\otimes \alpha_{i_1}} \otimes \cdots \otimes e_{i_k}^{\otimes \alpha_{i_k}},$$

where $\{\alpha_{i_1}, ..., \alpha_{i_k}\}$ are the non-zero elements of α. We also denote by $\tilde{e}(\alpha) \in \mathfrak{H}^{\odot |\alpha|}$ the canonical symmetrization of $e(\alpha)$. For every $q \geqslant 2$, the set $\{\tilde{e}(\alpha) : \alpha \in \Lambda, |\alpha| = q\}$ is

a complete orthogonal system in $\mathfrak{H}^{\odot q}$ (one needs to renormalise to obtain a basis). For every $q \geqslant 1$ and every $h \in \mathfrak{H}^{\odot q}$ of the form $h = \sum_{\alpha \in \Lambda, \, |\alpha|=q} c_\alpha \tilde{e}(\alpha)$, we define

$$I_q(h) = \sum_{\alpha \in \Lambda, \, |\alpha|=q} c_\alpha \Phi(\alpha), \tag{3.19}$$

where $\Phi(\alpha)$ is given in (3.18). Another classical result (see e.g. [32, 40]) is that, for every $q \geqslant 1$, the mapping $I_q : \mathfrak{H}^{\odot q} \to C_q$ (as defined in (3.19)) is surjective, and defines an isomorphism between C_q and the Hilbert space $\mathfrak{H}^{\odot q}$, endowed with the modified norm $\sqrt{q!}\| \cdot \|_{\mathfrak{H}^{\otimes q}}$. This means that, for every $h, h' \in \mathfrak{H}^{\odot q}$, $E[I_q(h)I_q(h')] = q!\langle h, h' \rangle_{\mathfrak{H}^{\otimes q}}$.

Finally, we notice that the Wiener-Itô chaotic decomposition of $L^2(\sigma(\mathbf{G}))$ can be rewritten as follows: every $F \in L^2(\sigma(\mathbf{G}))$ admits a unique decomposition of the type

$$F = \sum_{q=0}^{\infty} I_q(h_q), \tag{3.20}$$

where the series converges in $L^2(\sigma(\mathbf{G}))$, the symmetric kernels $h_q \in \mathfrak{H}^{\odot q}$, $q \geqslant 1$, are uniquely determined by F, and $I_0(h_0) := E[F]$. This also implies that

$$E[F^2] = E[F]^2 + \sum_{q=1}^{\infty} q!\|h_q\|_{\mathfrak{H}^{\otimes q}}^2,$$

which, by polarisation, yields an isomorphism between $L^2(\sigma(\mathbf{G}))$ and the symmetric Fock space $\bigoplus_{q \geqslant 0} \mathfrak{H}^{\odot q}$. When the functional F is sufficiently regular, one can explicitly represent the kernels h_q by means of the well-known *Stroock formula* (see e.g. [32, Corollary 2.7.8]).

3.2 Malliavin calculus

We let the previous notation and assumptions prevail: in particular, we shall fix for the rest of the section a real separable Hilbert space $\mathfrak{H}$, and represent the elements of the qth Wiener chaos of $\mathbf{G}$ in the form (3.19). In addition, we write $L^2(\mathfrak{H}) := L^2(\sigma(\mathbf{G}); \mathfrak{H})$ to indicate the space of all $\mathfrak{H}$-valued random elements u, that are measurable with respect to $\sigma(\mathbf{G})$ and verify the relation $E\big[\|u\|_{\mathfrak{H}}^2\big] < \infty$.

Let $\mathscr{S}$ be the set of all smooth cylindrical random variables of the form

$$F = g\big(I_1(h_1), \ldots, I_1(h_n)\big),$$

where $n \geqslant 1$, $g : \mathbb{R}^n \to \mathbb{R}$ is a smooth function with compact support and $h_i \in \mathfrak{H}$. The Malliavin derivative of F (with respect to $\mathbf{G}$) is the element of $L^2(\mathfrak{H})$ defined as

$$DF \;=\; \sum_{i=1}^{n} \partial_i g\big(I_1(h_1), \ldots, I_1(h_n)\big) h_i.$$

By iteration, one can define the mth derivative $D^m F$ (which is an element of $L^2(\mathfrak{H}^{\odot m})$) for every $m \geqslant 2$. For $m \geqslant 1$, $\mathbb{D}^{m,2}$ denotes the closure of $\mathscr{S}$ with respect to the norm $\| \cdot \|_{m,2}$, defined by the relation

$$\|F\|_{m,2}^2 = E[F^2] + \sum_{i=1}^{m} E\big[\|D^i F\|_{\mathfrak{H}^{\otimes i}}^2\big].$$

It is a standard result that a random variable F as in (3.20) is in $\mathbb{D}^{m,2}$ if and only if $\sum_{q \geqslant 1} q^m q!\|f_q\|_{\mathfrak{H}^{\otimes q}}^2 < \infty$, from which one deduces that $\bigoplus_{k=0}^{q} C_k \in \mathbb{D}^{m,2}$ for every

$q, m \geqslant 1$. Also, $DI_1(h) = h$ for every $h \in \mathfrak{H}$. The Malliavin derivative D satisfies the following *chain rule*: if $g : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and has bounded partial derivatives, and if $(F_1, ..., F_n)$ is a vector of elements of $\mathbb{D}^{1,2}$, then $g(F_1, \ldots, F_n) \in \mathbb{D}^{1,2}$ and

$$Dg(F_1, \ldots, F_n) = \sum_{i=1}^{n} \partial_i g(F_1, \ldots, F_n) DF_i. \tag{3.21}$$

In what follows, we denote by δ the adjoint of the operator D, also called the *divergence operator*. A random element $u \in L^2(\mathfrak{H})$ belongs to the domain of δ, written $\mathrm{Dom}\,\delta$, if and only if it satisfies

$$|E\left[\langle DF, u \rangle_{\mathfrak{H}}\right]| \leqslant c_u \sqrt{E[F^2]} \quad \text{for any } F \in \mathscr{S},$$

for some constant c_u depending only on u. If $u \in \mathrm{Dom}\,\delta$, then the random variable $\delta(u)$ is defined by the duality relationship (usually called 'integration by parts formula'):

$$E\left[F\delta(u)\right] = E\left[\langle DF, u \rangle_{\mathfrak{H}}\right], \tag{3.22}$$

which holds for every $F \in \mathbb{D}^{1,2}$. A crucial object for our discussion is the Ornstein-Uhlenbeck semigroup associated with $\mathbf{G}$.

Definition 3.2 (Ornstein-Uhlenbeck semigroup) Let $\mathbf{G}'$ be an independent copy of $\mathbf{G}$, and denote by E' the mathematical expectation with respect to $\mathbf{G}'$. For every $t \geqslant 0$ the operator $P_t : L^2(\sigma(\mathbf{G})) \to L^2(\sigma(\mathbf{G}))$ is defined as follows: for every $F(\mathbf{G}) \in L^2(\sigma(\mathbf{G}))$,

$$P_t F(\mathbf{G}) = E'[F(e^{-t}\mathbf{G} + \sqrt{1 - e^{-2t}}\mathbf{G}')],$$

in such a way that $P_0 F(\mathbf{G}) = F(\mathbf{G})$ and $P_\infty F(\mathbf{G}) = E[F(\mathbf{G})]$. The collection $\{P_t : t \geqslant 0\}$ verifies the semigroup property $P_t P_s = P_{t+s}$ and is called the *Ornstein-Uhlenbeck semigroup* associated with $\mathbf{G}$.

Some properties of the semigroup $\{P_t : t \geqslant 0\}$ are collected in the next statement.

Proposition 3.3   *1. For every $t > 0$, the eigenspaces of the operator P_t coincide with the Wiener chaoses C_q, $q = 0, 1, ...$, the eigenvalue of C_q being given by the positive constant e^{-qt}.*

2. *The infinitesimal generator of $\{P_t : t \geqslant 0\}$, denoted by L, acts on square-integrable random variables as follows: a random variable F with the form (3.20) is in the domain of L, written $\mathrm{Dom}\,L$, if and only if $\sum_{q \geqslant 1} q I_q(h_q)$ is convergent in $L^2(\sigma(\mathbf{G}))$, and in this case*

$$LF = -\sum_{q \geqslant 1} q I_q(h_q).$$

In particular, each Wiener chaos C_q is an eigenspace of L, with eigenvalue equal to $-q$.

3. *The operator L verifies the following properties: (i) $\mathrm{Dom}\,L = \mathbb{D}^{2,2}$, and (ii) a random variable F is in $\mathrm{Dom}L$ if and only if $F \in \mathrm{Dom}\,\delta D$ (i.e. $F \in \mathbb{D}^{1,2}$ and $DF \in \mathrm{Dom}\delta$), and in this case one has that $\delta(DF) = -LF$.*

It is also immediate to describe the *pseudo-inverse* of L, denoted by L^{-1}, as follows: for every mean zero random variable $F = \sum_{q \geqslant 1} I_q(h_q)$ of $L^2(\sigma(\mathbf{G}))$, one has that

$$L^{-1}F = \sum_{q \geqslant 1} -\frac{1}{q} I_q(h_q). \tag{3.23}$$

It is clear that L^{-1} is an operator with values in $\mathbb{D}^{2,2}$.

We also record the following estimate involving random variables living in a finite sum of Wiener chaoses: the proof follows from the hypercontractivity of the Ornstein-Uhlenbeck semigroup — see for instance [32, Theorem 2.7.2 and Theorem 2.8.12].

Proposition 3.4 (Hypercontractivity) *Let $q \geqslant 1$ and $1 \leqslant s < t < \infty$. Then, there exists a finite constant $c(s,t,q) < \infty$ such that, for every $F \in \bigoplus_{k=0}^{q} C_k$,*

$$E[|F|^t]^{\frac{1}{t}} \leqslant c(s,t,q)\, E[|F|^s]^{\frac{1}{s}}. \tag{3.24}$$

In particular, all L^p norms, $p \geqslant 1$, are equivalent on a finite sum of Wiener chaoses.

The following example allows one to establish an explicit connection between the content of Section 2 and our general framework.

Example 3.5 (Brownian motion) *Let $W = \{W_t : t \geqslant 0\}$ be a standard Brownian motion, let $\{e_j : j \geqslant 1\}$ be an orthonormal basis of $L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), dt) =: L^2(\mathbb{R}_+)$, and define $G_j = \int_0^\infty e_j(t)dW_t$. Then, $\sigma(W) = \sigma(\mathbf{G})$, where $\mathbf{G} = \{G_j : j \geqslant 1\}$. In this case, the natural choice of a Hilbert space is $\mathfrak{H} = L^2(\mathbb{R}_+)$ and one has the following explicit characterisation of the Wiener chaoses associated with W: for every $q \geqslant 1$, one has that $F \in C_q$ if and only if there exists a symmetric kernel $f \in L^2(\mathbb{R}_+^q)$ such that*

$$F = q! \int_0^\infty \int_0^{t_1} \cdots \int_0^{t_{q-1}} f(t_1,...,t_q)dW_{t_q}\cdots dW_{t_1} := q! J_q(f).$$

The random variable $J_q(f)$ is known as the iterated Wiener-Itô integral *of order q, of f with respect to W. It is a well-known fact that, if $F \in \mathbb{D}^{1,2}$ admits the chaotic expansion $F = E[F] + \sum_{q \geqslant 1} J_q(f_q)$, then DF equals the random function*

$$t \mapsto \sum_{q \geqslant 1} q J_{q-1}(f_q(t,\cdot))$$

$$= \sum_{q \geqslant 1} q! \int_0^\infty \int_0^{t_1} \cdots \int_0^{t_{q-2}} f(t,t_1...,t_{q-1})dW_{t_{q-1}}\cdots dW_{t_1}, \quad t \in \mathbb{R}_+,$$

which is a well-defined element of $L^2(\mathfrak{H})$.

3.3 Stein factors

The statement of the forthcoming Proposition 3.6 is an important application of the chain rule (3.21). A proof of the first part (focussing on continuously differentiable functions) can be found e.g. in [32, Proposition 5.1.1]. The second part (which is slightly less standard) can be proved by applying the well-known Lusin's theorem of real analysis, see [27, Corollary 4.17].

Proposition 3.6 *Let F be centered element of $\mathbb{D}^{1,2}$.*

1. *For every continuously differentiable and Lipschitz function $g : \mathbb{R} \to \mathbb{R}$*

$$\begin{aligned} E[Fg(F)] &= E[g'(F)\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}] \\ &= E[g'(F)E[\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}|F]]. \end{aligned} \tag{3.25}$$

2. *Relation (3.25) continues to hold whenever g has the form $g(x) = \int_{-\infty}^x \mathbf{1}_C(a)da$, and C is a bounded Borel set[§], in other words:*

$$\begin{aligned} E[Fg(F)] &= E[\mathbf{1}_C(F)\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}] \\ &= E[\mathbf{1}_C(F)E[\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}|F]]. \end{aligned} \tag{3.26}$$

[§]That is, there exists a finite $\alpha > 0$ such that $C \subset [-\alpha, \alpha]$.

The mapping

$$x \mapsto \tau_F(x) := E[\langle DF, -DL^{-1}F \rangle_{\mathfrak{H}} | F = x], \tag{3.27}$$

appearing in (3.25)–(3.26) is customarily called the *Stein factor* associated with F. It is easily seen that, up to P-negligible sets, the random variable $\tau_F(F)$ is the unique element of $L^1(\sigma(F), P)$ verifying the relation

$$E[Fg(F)] = E[\tau_F(F)g'(F)]$$

for every smooth test function g. A simple integration by parts reveals that, if $F \sim \mathcal{N}(0, \sigma^2)$, then necessarily $\tau_F = \sigma^2$.

The results discussed in the next section will show that the analysis of the factor τ_F is actually a crucial tool in order to measure the distance between the law of F and that of a centred Gaussian random variable. As a first step in this direction, we now present a statement (Theorem 3.7) connecting Stein factors to the shape of the density of a random variable $F \in \mathbb{D}^{1,2}$ having an absolutely continuous distribution. A proof of the first part (which is originally due to Nourdin and Viens) can be found in [32, Theorem 10.1.1], while the second part is a new powerful representation of relative entropy, that is the starting point of the paper [37].

In what follows, we shall assume that F is a centered random variable with unit variance. Recall that, if the distribution F admits a density f, denoting by ϕ the density of a standard Gaussian random variable $N \sim \mathcal{N}(0, 1)$, the *relative entropy* of F is defined as

$$D(F\|N) := \int_{\mathbb{R}} f(x) \log \frac{f(x)}{\phi(x)} dx. \tag{3.28}$$

The relevance of the quantity $D(F\|N)$ in the classical central limit theorem is beautifully demonstrated in the monograph [21]. Here, we shall recall that, in view of the Csiszar-Kullback-Pinsker inequality (see [12, 23, 50]), one has the upper bound

$$D(F\|N) \geqslant \frac{1}{2} \left(\int_{\mathbb{R}} |f(x) - \phi(x)| dx \right)^2, \tag{3.29}$$

that is: the relative entropy controls (up to a universal multiplicative factor) the square of the total variation distance between the laws of F and N. Note that many other deep inequalities involving relative entropy are available, like for instance Talagrand's transport inequality [54] — see [1, Chapter 11] or [17, 21] for an overview. Without loss of generality, we shall implicitly assume that F and N are stochastically independent, and we introduce the notation $F_t := \sqrt{t}F + \sqrt{1-t}N$, $t \in [0, 1]$.

Theorem 3.7 *Let $F \in \mathbb{D}^{1,2}$ be centered and have unit variance.*

1. *The law of F admits a density (with respect to the Lebesgue measure), say f, if and only of the random variable $\tau_F(F)$ is P-almost surely strictly positive. In this case, the support of f is given by a closed interval of $\mathbb{R}$ containing zero, and for dx-almost all x in the support of f,*

$$f(x) = \frac{E|F|}{2\tau_F(x)} \exp\left\{ -\int_0^x \frac{a\, da}{\tau_F(a)} \right\}. \tag{3.30}$$

2. *Assume that the distribution of F admits a density f. Then,*

$$D(F\|N) = \frac{1}{2} \int_0^1 \frac{t}{1-t} E\left[E\left[N(1 - \tau_F(F)) \big| F_t \right]^2 \right] dt. \tag{3.31}$$

The proof of (3.31) makes use of a rescaled integral version (due to Barron [4]) of the so-called *de Bruijn's identity* of information theory, providing a representation of the derivative (in t) of the relative entropy of the variable $F + \sqrt{t}N$ in terms of its Fisher information. A mildly interesting side remark: we could find no direct simple argument, allowing one to deduce (3.31) by plugging (3.30) into (3.28).

4 Some explicit bounds

We let the notation and assumptions of the previous section prevail. Our aim is to discuss several explicit estimates, assessing the distance between the distribution of a given $F \in \mathbb{D}^{1,2}$ and the law of a $\mathcal{N}(0,1)$ random variable. We recall that the *total variation distance* between the distribution of two $\mathbb{R}^d$-valued random variables X, Y is given by

$$d_{TV}(X,Y) = \sup_{C \in \mathscr{B}(\mathbb{R}^d)} |P[X \in C] - P[Y \in C]|.$$

Moreover, if X, Y have densities, say f_X, f_Y, one has that

$$d_{TV}(X,Y) = \frac{1}{2} \int_{\mathbb{R}^d} |f_X(\mathbf{x}) - f_Y(\mathbf{x})| \, d\mathbf{x}. \tag{4.32}$$

The *Fortet-Mourier distance* between the distributions of X and Y is given by

$$d_{FM}(X,Y) = \sup_{\|g\|_\infty, \|g\|_L \leqslant 1} |E[g(X)] - P[g(Y)]|,$$

where $\|g\|_L$ denotes the Lipschitz constant of g, that is: in the previous expression, the supremum is taken over the class of all 1-Lipschitz continuous functions that are bounded by 1. It is a well-known fact that the topology induced by d_{TV}, on the class of all probability measures on $\mathbb{R}^d$, is strictly stronger than the topology of convergence in distribution, whereas d_{FM} metrizes convergence in distribution (see e.g. [15]).

4.1 Bounds derived from Stein's method

Stein's method can be roughly described as collection of analytical techniques, allowing one to measure the distance between random objects by assessing the regularity of the solutions to some ordinary (in dimension 1) or partial (in higher dimensions) differential equations. An introduction to this important topic can be found in the monographs [9, 32]. In the case of normal approximations, one of the most powerful bounds that can be derived from Stein's method is the following estimate on the total variation distance between the distribution of an integrable random variable F and that of $N \sim \mathcal{N}(0,1)$:

$$d_{TV}(F,N) \leqslant \sup |E[Fg(F) - g'(F)]|, \tag{4.33}$$

where the supremum runs over all test functions $g : \mathbb{R} \to \mathbb{R}$ such that $\|g\|_\infty \leqslant \sqrt{\pi/2}$ and $\|g'\|_\infty \leqslant 2$. See e.g. [32, Theorem 3.3.1] for a proof.

Combining relation (4.33) with the content of Proposition 3.6, one obtains the following statement, which is crucial in understanding the interaction between Stein's method and Malliavin calculus. Note that (differently e.g. from [32, Chapter 5]) we are not assuming that the random variable has a density: this generalisation can once again be achieved by exploiting Lusin's theorem (see [27, Theorem 5.2]).

Theorem 4.1 *Let F be a centered element of $\mathbb{D}^{1,2}$, and let $N \sim \mathcal{N}(0,1)$. Then,*

$$d_{TV}(F,N) \leqslant 2E \left| 1 - \langle DF, -DL^{-1}F \rangle_{\mathfrak{H}} \right|. \tag{4.34}$$

In particular, if F belongs to the qth Wiener chaos C_q, one has that $\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}} = \frac{1}{q}\|DF\|_{\mathfrak{H}}^2$, $E[F^4] > 3$, and the following inequality holds:

$$E\left|1 - \frac{1}{q}\|DF\|_{\mathfrak{H}}^2\right| \leqslant \sqrt{\frac{q-1}{3q}(E[F^4] - 3)}. \tag{4.35}$$

Note that $3 = E[N^4]$. Using the hypercontractivity relation (3.24) one deduces immediately from (4.35) the following result, originally discovered by Nualart and Peccati in [42] (by using the random time-change technique of Section 2.3). It represents a drastic simplification of the method of moments and cumulants, as applied to random variables living inside a fixed Wiener chaos.

Theorem 4.2 (Fourth moment theorem – see [42]) *Fix $q \geqslant 2$, let $\{F_n : \geqslant 1\}$ be a sequence of random variables living on the qth Wiener chaos C_q, and assume that $E[F_n^2] \to 1$. Then, F_n converges in distribution to $N \sim \mathscr{N}(0,1)$ if and only if $E[F_n^4] \to 3$. In this case, the convergence takes place in the sense of total variation, with an upper bound on the speed of convergence obtained by combining the two estimates (4.34)–(4.35).*

Several criteria for proving that the rates of convergence implied by (4.34)–(4.35) are optimal are pointed out in [29, 33]. In particular, in [33] the following 'definitive' optimality result is proved:

Theorem 4.3 (Optimal fourth moment theorem – see [33]) *[Fix $q \geqslant 2$, let $\{F_n : \geqslant 1\}$ be a sequence of random variables living on the qth Wiener chaos C_q, and assume that $E[F_n^2] = 1$ and $E[F_n^4] \to 3$. Then, $E[F_n^3] \to 0$, and there exist two finite constants $0 < c < C$ (possibly depending on q and on the sequence $\{F_n\}$, but not on n) such that*

$$c\,\mathbf{M}(F_n) \leqslant d_{TV}(F_n, N) \leqslant C\,\mathbf{M}(F_n), \quad n \geqslant 1, \tag{4.36}$$

where $N \sim \mathscr{N}(0,1)$ and

$$\mathbf{M}(F_n) := \max\{E[F_n^3] \,;\, E[F_n^4] - 3\}.$$

Note that, differently from (4.35), the constants in (4.36) are not explicit. Extensions to the framework of functionals of Poisson measures can be found e.g. in [5, 44, 49, 52]. The paper [24] provides similar estimates for random variables living inside the chaos of a Markov operator; further important developments can be found in [2, 3]. Analogous statements in the setting of free probability are discussed in [13, 22].

Multidimensional generalisations of the previous results are derived in [32, Chapter 6, Chapter 12], as well as in [39] (providing a characterisation of asymptotic independence on a Gaussian space) and [35] (focussing on universality results for homogeneous sums). An extension to non-central limit theorems (in particular, the Gamma distribution) is developed in [31].

It is important to notice that the power of Stein's method significantly breaks down whenever one deals either (a) with an absolutely continuous limit distribution that is not Gaussian, or (b) with multidimensional limit theorems. Albeit many progresses are being made (see e.g. [14, 16]), it is indeed safe to say that Stein's method for absolutely continuous non-Gaussian distributions is quite difficult to implement (both in the one- and multi-dimensional cases), one of the main obstacles being the lack of adequate estimates on the solutions of the corresponding differential equations. For instance, the bounds proved in [31] only allow one to deal with very smooth test functions, thus making it impossible to deduce rates of convergence in the total variation distance. A similar situation also concerns multidimensional approximations, where (even in the Gaussian case) Stein's method can only be applied in the framework of distances involving sufficiently smooth

test functions, like for instance the 1-Wasserstein distance (see e.g. [32, Chapters 4 and 6] and [8]).

Some of these difficulties can be overcome by combining Malliavin calculus with different techniques, like for instance the Carbery-Wright inequalities [6] or tools from information theory. This corresponds to the content of references [28, 37, 38], on which we shall focus in Section 4.3.

In the next section, we demonstrate how to relate the above described estimates to Poincaré-type inequalities.

4.2 Second order Poincaré inequalities

Recall the classical Poincaré inequality for Gaussian vectors (a result usually attributed to Nash [26] – see e.g. [1] for a general discussion): fix $d \geqslant 1$, let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth mapping, and let $G = (G_1, ..., G_d)$ be a vector of i.i.d. $\mathcal{N}(0, 1)$ random variables. Then,

$$\mathbf{Var}(f(G)) \leqslant E\|\nabla f(G)\|_{\mathbb{R}^d}^2. \tag{4.37}$$

A proof of (4.37) can be easily derived by first expanding f and ∇f into a series of Hermite polynomials — see e.g. [10]. Such an estimate provides an explicit quantitative counterpart to the following heuristic consideration: if $E\|\nabla f(G)\|_{\mathbb{R}^d}^2$ is small, then the mapping f is close to a constant, and therefore the variance of $f(G)$ should also be close to zero.

The estimate (4.37) can be easily extended to a regular functional F of a general Gaussian sequence $\mathbf{G} = \{G_j : j \geqslant 1\}$. Indeed, applying the Cauchy-Shchwarz inequality to the right-hand side of (3.25) in the case $g(x) = x$, and using the relation $E[\|DF\|_{\mathfrak{H}}^2] \leqslant E[\|DL^{-1}F\|_{\mathfrak{H}}^2]$ (which is a consequence of (3.23)), yields the following general inequality: for every $F \in \mathbb{D}^{1,2}$,

$$\mathbf{Var}(F) \leqslant E\|DF\|_{\mathfrak{H}}^2. \tag{4.38}$$

See e.g. [18] for more details.

Now let $F \in \mathbb{D}^{2,4}$ be centered and with unit variance, and write N to indicate a Gaussian $\mathcal{N}(0, 1)$ random variable. The next result, which was originally proved in [34] by expanding ideas first introduced in [7], provides a bound on $d_{TV}(F, N)$ having the form of a 'second order' inequality of the type (4.38). In what follows, the symbol $\|D^2F\|_{op}$ stands for the operator norm of the random Hilbert-Schmidt operator associated with the second Malliavin derivative D^2F (which is a random element with values in $\mathfrak{H}^{\odot 2}$).

Theorem 4.4 (Second order Poincaré inequality – see [34]) *Assume that $F \in \mathbb{D}^{2,4}$, $E(F) = 0$ and $\mathrm{Var}(F) = 1$. Let $N \sim \mathcal{N}(\mu, \sigma^2)$. Then,*

$$d_{TV}(F, N) \quad \leqslant \quad 2E\left|1 - \langle DF, -DL^{-1}F \rangle_{\mathfrak{H}}\right| \tag{4.39}$$

$$\leqslant \quad \sqrt{10}E\left[\|D^2F\|_{op}^4\right]^{\frac{1}{4}} \times E\left[\|DF\|_{\mathfrak{H}}^4\right]^{\frac{1}{4}}. \tag{4.40}$$

A proof of the estimate (4.40) is obtained by carefully applying the first-order Poincaré inequality (4.38) to the random variable $\langle DF, -DL^{-1}F \rangle_{\mathfrak{H}}$. The requirement that $F \in \mathbb{D}^{2,4}$ ensures that the right-hand side of (4.40) is well-defined and finite. The intuition behind the previous bound is the following: if the random variable $\|D^2F\|_{op}$ is negligeable with respect to $\|DF\|_{\mathfrak{H}}$, then F is intuitively close to a linear functional of the underlying field $\mathbf{G}$, and therefore the distribution of F must be not far from Gaussian.

Reference [34] contains multidimensional extensions of Theorem 4.4, as well as several applications to the asymptotic analysis of non-linear functionals of continuous-time Gaussian processes. As already recalled, analogous estimates in a finite-dimensional setting

can be found in [7], where a large number of applications in random matrix theory is developed. It is important to notice that, when applied to random variables inside a finite sum of Wiener chaoses, the bound (4.40) yields rates of convergence that are suboptimal with respect to the ones that can be derived from a direct estimation of (4.39). However, relation (4.40) can be crucial for dealing with random variables having an infinite chaotic expansion.

4.3 Results related to the Carbery-Wright inequalities

We now recall a version of the inequality proved by Carbery and Wright in [6], yielding a control on the small ball probabilities associated with polynomials transformations of a random vector. Such a result is at the core of the forthcoming Theorem 4.5, as well as of its multidimensional extensions [28]. Recall that, in a form that is adapted to our framework, the main finding of [6] reads as follows: there is a universal constant $c > 0$ such that, for any polynomial $Q : \mathbb{R}^n \to \mathbb{R}$ of degree at most d and any $\alpha > 0$ we have

$$E[Q(G_1, \ldots, G_n)^2]^{\frac{1}{2d}} \ P(|Q(G_1, \ldots, G_n)| \leqslant \alpha) \leqslant c d \alpha^{\frac{1}{d}}, \tag{4.41}$$

where $G_1, \ldots, G_n$ are independent random variables with common distribution $\mathcal{N}(0, 1)$. Note that, as proved in [6], the estimate (4.41) extends to the general framework of log-concave density functions.

The next statement, first proved in [38], yields the following striking result: inside a finite sum of Wiener chaoses, convergence in distribution towards a non-degenerate random variable is equivalent to convergence in total variation. Moreover, explicit (albeit largely suboptimal) rates of convergence can be deduced by means of the estimate (4.42). As already discussed, this finding is outside the scope of Stein's method.

We recall that, according to a result by Shikegawa, non-zero random variables living inside a finite sum of Wiener chaoses have necessarily an absolutely continuous distribution (see e.g. [32, Section 2.10] for a proof).

Theorem 4.5 (See [38]) *Fix an integer $M \geqslant 1$. Let $\{F_n : n \geqslant 1\} \subset \bigoplus_{q=0}^{M} C_q$ be a sequence living in the sum of the first M chaoses associated with $\mathbf{G}$, and assume that, as $n \to \infty$, F_n converges in distribution towards some random variable F_∞ such that $\mathbf{Var}(F_\infty) > 0$. Then, one has the following:*

1. *The random variable F_∞ has necessarily an absolutely continuous distribution.*

2. *There exists a finite constant $c > 0$ (independent of n, but possibly depending on the sequence $\{F_n\}$) such that, for every n,*

$$d_{TV}(F_n, F_\infty) \leqslant c \, d_{FM}(F_n, F_\infty)^{\frac{1}{1+2M}}, \tag{4.42}$$

so that, in particular, F_n converges in distribution to F_∞ in the sense of the total variation distance.

We observe that, whenever $F_\infty \sim \mathcal{N}(0, 1)$ and F_n lives in the qth Wiener chaos of $\mathbf{G}$, one can prove that

$$d_{FM}(F_n, F_\infty) \leqslant \sqrt{\frac{q-1}{3q}(E[F_n^4] - 3)},$$

so that (4.42) also yields a weak version of the estimates achievable by means of Theorem 4.1.

A multidimensional extension of Theorem 4.5 is proved in [28], where the authors also provide a neat criterion (based on the use of Malliavin matrices) in order for a vector

living inside a finite sum of Wiener chaoses to have a density. As a by-product of the results proved in [28] (see Theorem 5.2 therein), one also obtains that a vector of multiple integrals converging in distribution towards a Gaussian vector with positive covariance matrix always converges in the sense of total variation – thus providing a strengthened version of a result by Peccati and Tudor [46]. One should notice, however, that (similarly to equation (4.42)) the rates of convergence in the total variation distance implied by [28] heavily depend on the order of the chaoses – whereas it is natural to conjecture that they should be the same as those deduced in [32, Chapter 6] for the multidimensional Wasserstein distance. This difficulty has been partially solved in the paper [37], where the main estimates proved in [28] are combined with tools from information theory.

4.4   Multidimensional entropic bounds

Fix $d \geqslant 1$ and let $F = (F_1, ..., F_d)$ be a centered d-dimensional vector with covariance matrix $C > 0$. Assume that the distribution of F admits a density on $\mathbb{R}^d$, say f, and denote by ϕ_C the density of a Gaussian random vector $N = (N_1, ..., N_d)$ with zero mean and covariance matrix C. Then, analogously to the one-dimensional case, the relative entropy of F is defined as

$$\frac{1}{2} D(F \| N) := \int_{\mathbb{R}^d} f(\mathbf{x}) \log \frac{f(\mathbf{x})}{\phi_C(\mathbf{x})} d\mathbf{x}. \tag{4.43}$$

Since the Csiszar-Kullback-Pinsker inequality continues to hold, one has the bound

$$\frac{1}{2} D(F \| N) \geqslant \frac{1}{4} \left(\int_{\mathbb{R}^d} |f(\mathbf{x}) - \phi(\mathbf{x})| d\mathbf{x} \right)^2 = d_{TV}(F, N)^2, \tag{4.44}$$

where we have used (4.32).

The following result has been proved in [37], by using a multidimensional version of the generalised de Bruijn identity (3.31).

Theorem 4.6 (Entropic CLTs on Wiener chaos – see [37]) *Let $d \geqslant 1$ and*

$$q_1, \ldots, q_d \geqslant 1$$

be fixed integers. Consider vectors

$$F_n = (F_{1,n}, \ldots, F_{d,n}) = (I_{q_1}(h_{1,n}), \ldots, I_{q_d}(h_{d,n})), \quad n \geqslant 1,$$

with $h_{i,n} \in \mathfrak{H}^{\odot q_i}$. Let C_n denote the covariance matrix of F_n and let $Z_n \sim \mathscr{N}_d(0, C_n)$ be a centered Gaussian random vector in $\mathbb{R}^d$ with the same covariance matrix as F_n. Write

$$\Delta_n = E[\|F_n\|_{\mathbb{R}^d}^4] - E[\|Z_n\|_{\mathbb{R}^d}^4], \quad n \geqslant 1.$$

Then, $\Delta_n > 0$ for every n. Moreover, if $C_n \to C > 0$ and $\Delta_n \to 0$, as $n \to \infty$, then, the random vector F_n admits a density for n large enough, and

$$D(F_n \| Z_n) = O(1) \Delta_n |\log \Delta_n| \quad as \ n \to \infty, \tag{4.45}$$

where $O(1)$ indicates a bounded numerical sequence depending on $d, q_1, ..., q_d$, as well as on the sequence $\{F_n\}$.

The additional logarithmic factor in the bound (4.45) is a consequence of the fact that formula (3.31) (as well as its multidimensional generalisations) involves the mapping $t \mapsto t/(1 - t)$, which is not integrable on $[0, 1]$.

The next subsection contains a short discussion of the implication of the previous results for the class of Brownian quadratic functionals introduced in Section 2.

4.5 Back to quadratic functionals

We conclude this section with a statement providing a complete solution to Problem I, as appearing in Section 2.1. The proof of the upper bounds follows immediately from Theorem 4.5 and Theorem 4.6, as well as from the fact that, according to (2.7), the random variable $\widetilde{B}_\varepsilon$ is an element of the second chaos of W, with a fourth cumulant evolving according to (2.12) (in the case $n = 4$). The proof of the lower bound follows from an application of the main findings of [29].

Proposition 4.7 *Let the notation of Section 2.1 prevail. There exist two finite constants $0 < k < K$, independent of ε and such that*

$$\frac{k}{\sqrt{\log \frac{1}{\varepsilon}}} \leqslant d_{Kol}(\widetilde{B}_\varepsilon, N) \leqslant \frac{K}{\sqrt{\log \frac{1}{\varepsilon}}}.$$

Moreover, for ε sufficiently small,

$$D(\widetilde{B}_\varepsilon \| N) \leqslant K \times \frac{\log\log \frac{1}{\varepsilon}}{\log \frac{1}{\varepsilon}}.$$

Note how the content of Proposition 4.7 is largely outside the scope of the techniques based on the method of moments or on random time-changes.

5 Conclusion

Starting from the asymptotic analysis of quadratic Brownian functionals, I have provided an overview of recent developments in the theory of normal approximations for functionals of Gaussian fields. Albeit many problems have been settled, a number of interesting questions are open for future research. Here is a short list:

– One of the next important steps should be the effective use of Malliavin techniques in order to prove limit theorems in infinite dimension, like for instance theorems of the Donsker type. Some remarkable progress in this direction has been recently made in [11].

– It is not clear how to extend the information theoretical approach initiated in [37] to non-Gaussian limit distributions. The main difficulty consists in finding an effective equivalent of the generalised de Bruijn's identity (3.31).

– More generally, we need new ideas for dealing with non-Gaussian distributions for which a version of Stein's method is not available. For instance, we still do not now how to deal with general limits of the type $P(N)$, where N is a standard Gaussian random variable and P is a polynomial of degree strictly greater than 2.

– Albeit the results from [28, 38] provide some important new insights, it is still an open problem to characterise all the possible limit laws for sequence of random variables living inside a Wiener chaos of order $\geqslant 3$

– It is not clear how one could use stochastic analysis in order to systematically deal with Edgeworth expansions for sequences of Gaussian subordinated functionals. Some first partial answers (only for Edgeworth expansions of order 1) can be found in [29].

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer (2000). *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris. With a preface by Dominique Bakry and Michel Ledoux.

[2] E. Azmoodeh, S. Campese and G. Poly (2013). Fourth moment theorems for Markov diffusion generators. Preprint.

[3] E. Azmoodeh, D. Malicet and G. Poly (2013). Generalization of the Nualart-Peccati criterion.

[4] A. R. Barron (1986). Entropy and the central limit theorem. *Ann. Probab.*, 14(1), 336–342.

[5] S. Bourguin and G. Peccati (2012). Portmanteau inequalities on the Poisson space: mixed limits and multidimensional clustering. Preprint.

[6] A. Carbery and J. Wright (2001). Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Math. Res. Lett.*, 8(3), 233–248.

[7] S. Chatterjee (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1-2), 1–40.

[8] S. Chatterjee and E. Meckes (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4, 257–283.

[9] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao (2011). *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg.

[10] H. Chernoff (1981). A note on an inequality involving the normal distribution. *Ann. Probab.* **9**(3), 533–535.

[11] L. Coutin and L. Decreusefond (2012). Stein's method for Brownian approximations. Preprint.

[12] I. Csiszár. Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen (1962). *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 7, 137–158.

[13] A. Deya, S. Noreddine, and I. Nourdin (2012). Fourth moment theorem and q-Brownian chaos. *Comm. Math. Phys., to appear*.

[14] C. Döbler (2012). Stein's method of exchangeable pairs for absolutely continuous, univariate distributions with applications to the Pólya urn model. Preprint.

[15] R. M. Dudley (1989). *Real analysis and probability*. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

[16] L. Goldstein and G. Reinert (2012). Stein's method for the Beta distribution and the Pólya-Eggenberger Urn. To appear in: *Advances in Applied Probability*

[17] N. Gozlan and C. Léonard (2010). Transport inequalities - A survey. *Markov Processes and Related Fields* 16, 635-736.

[18] C. Houdré and V. Pérez-Abreu (1995). Covariance identities and inequalities for functionals on Wiener and Poisson spaces. *Ann. Probab.* 23, 400–419.

[19] Y. Hu, F. Lu and D. Nualart (2013). Convergence of densities of some functionals of Gaussian processes. Preprint.

[20] T. Jeulin and M. Yor (1979). Inégalité de Hardy, semi-martingales et faux-amis. In: *Séminaire de Probabilités XIII*, LNM **721**, Springer-Verlag, Berlin Heidelberg New York, 332–359.

[21] O. Johnson (2004). *Information theory and the central limit theorem.* Imperial College Press, London.

[22] T. Kemp, I. Nourdin, G. Peccati, and R. Speicher (2012). Wigner chaos and the fourth moment. *Ann. Probab.*, 40(4), 1577–1635.

[23] S. Kullback (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Info. Theory*, 4.

[24] M. Ledoux (2012). Chaos of a Markov operator and the fourth moment condition. *Ann. Probab.*, 40(6), 2439–2459.

[25] D. Marinucci and G. Peccati (2011). *Random fields on the sphere : Representation, limit theorems and cosmological applications*, volume 389 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge.

[26] J. Nash (1956). Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.*, 80, 931–954.

[27] I. Nourdin (2013). Lectures on Gaussian approximations with Malliavin calculus. To appear in: *Séminaire de probabilités.*

[28] I. Nourdin, D. Nualart, and G. Poly (2012). Absolute continuity and convergence of densities for random vectors on Wiener chaos. *Electron. J. Probab.*, 18(22), 1–19.

[29] I. Nourdin and G. Peccati (2010). Stein's method and exact Berry-Esseen asymptotics for functionals of Gaussian fields. *Ann. Probab.* 37(6), 2231–2261.

[30] I. Nourdin and G. Peccati (2010). Universal Gaussian fluctuations of non-Hermitian matrix ensembles: from weak convergence to almost sure CLTs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 7, 341–375.

[31] I. Nourdin and G. Peccati (2009). Noncentral convergence of multiple integrals. *Ann. Probab.*, 37, 1412–1426.

[32] I. Nourdin and G. Peccati (2012). *Normal approximations with Malliavin calculus : from Stein's method to universality.* Cambridge Tracts in Mathematics. Cambridge University Press.

[33] I. Nourdin and G. Peccati (2013). The optimal fourth moment theorem. Preprint.

[34] I. Nourdin, G. Peccati and G. Reinert (2009). Second order Poincaré inequalities and CLTs on Wiener space. *Journal of Functional Analysis* **257**, 593-609.

[35] I. Nourdin, G. Peccati, and G. Reinert (2010). Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos. *Ann. Probab.*, 38(5), 1947–1985.

[36] I. Nourdin, G. Peccati, and A. Réveillac (2010). Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1), 45–58.

[37] I. Nourdin, G. Peccati and Y. Swan (2013). Entropy and the fourth moment phenomenon. Preprint.

[38] I. Nourdin and G. Poly (2013). Convergence in total variation on Wiener chaos. *Stochastic Process. Appl.*, 123, 651–674.

[39] I. Nourdin and J. Rosiński (2013). Asymptotic independence of multiple Wiener-Itô integrals and the resulting limit laws. *Ann. Probab.*, to appear.

[40] D. Nualart (2006). *The Malliavin calculus and related topics.* Probability and its Applications (New York). Springer-Verlag, Berlin, second edition.

[41] D. Nualart and S. Ortiz-Latorre (2008). Central limit theorems for multiple stochastic integrals and Malliavin calculus. *Stochastic Process. Appl.*, 118(4), 614–628.

[42] D. Nualart and G. Peccati (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1), 177–193.

[43] G. Peccati (2009). Stein's method, Malliavin calculus and infinite-dimensional Gaussian analysis. Unpublished lecture notes.

[44] G. Peccati, G.; J.L. Solé, M.S. Taqqu and F. Utzet (2010). Stein's method and normal approximation of Poisson functionals, *Ann. Probab.* **38**, 443–478.

[45] G. Peccati, G. and M.S. Taqqu (2011). *Wiener Chaos: Moments, Cumulants and Diagrams.* Bocconi University Press and Springer, Milan.

[46] G. Peccati and C.A. Tudor (2005). Gaussian limits for vector-valued multiple stochastic integrals. *Séminaire de Probabilités XXXVIII*, LNM **1857**, 247-262.

[47] G. Peccati and M. Yor (2004). Hardy's inequality in $L^2([0,1])$ and principal values of Brownian local times. *Asymptotic Methods in Stochastics*, AMS, Fields Institute Communications Series, 49–74.

[48] G. Peccati and M. Yor (2004). Four limit theorems for quadratic functionals of Brownian motion and Brownian bridge. *Asymptotic Methods in Stochastics*, AMS, Fields Institute Communication Series, 75–87.

[49] G. Peccati and C. Zheng (2010). Multi-dimensional Gaussian fluctuations on the Poisson space, *Electron. J. Probab.*, 15, 1487–1527.

[50] M. S. Pinsker (1964). *Information and information stability of random variables and processes.* Translated and edited by Amiel Feinstein. Holden-Day Inc., San Francisco, Calif..

[51] G. Reinert and A. Roellin (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.*, 37(6), 2150–2173.

[52] M. Reitzner and M. Schulte (2012). Central limit theorems for U-statistics of Poisson point processes. *Ann. Probab.*, to appear.

[53] D. Revuz and M. Yor (1999). *Continuous martingales and Brownian motion.* Springer-Verlag, Berlin Heidelberg New York.

[54] M. Talagrand (1986). Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3), 587–600, 1996.