

## Chapter 6

### Stein's method, Malliavin calculus, Dirichlet forms and the fourth moment theorem

Louis H.Y. Chen\* and Guillaume Poly†

The fourth moment theorem provides error bounds of the order  $\sqrt{\mathbb{E}(F^4) - 3}$  in the central limit theorem for elements  $F$  of Wiener chaos of any order such that  $\mathbb{E}(F^2) = 1$ . It was proved by Nourdin and Peccati [31] using Stein's method and the Malliavin calculus. It was also proved by Azmoodeh, Campese and Poly [2] using Stein's method and Dirichlet forms. This paper is an exposition on the connections between Stein's method and the Malliavin calculus and between Stein's method and Dirichlet forms, and on how these connections are exploited in proving the fourth moment theorem.

#### Contents

1. Introduction . . . . .	108
2. Stein's method . . . . .	109
2.1. How it began . . . . .	109
2.2. A general framework . . . . .	110
2.3. Normal approximation . . . . .	111
3. Malliavin calculus . . . . .	113
3.1. A brief history . . . . .	113
3.2. Malliavin derivatives . . . . .	114
3.3. Wiener chaos and multiple integrals . . . . .	115
3.4. Main properties of Malliavin operators . . . . .	117
4. Connecting Stein's method with Malliavin calculus . . . . .	118
5. The Nualart-Peccati criterion of the fourth moment and Ledoux's idea . . . . .	121
5.1. Some history . . . . .	121
5.2. Overview of the proof of Nourdin and Peccati . . . . .	122
5.3. About Ledoux's generalization . . . . .	123
6. The general fourth moment Theorem for Dirichlet forms . . . . .	124
6.1. The Dirichlet structures . . . . .	124
6.2. Fourth moment theorem for Dirichlet structures with (H1) and (H2) . . . . .	125
6.3. Dirichlet structures with (H1) and (H2) . . . . .	126
7. Acknowledgments . . . . .	128
References . . . . .	128

\*Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore, matchyl@nus.edu.sg

†Mathematical research unit, University of Luxembourg, 6 rue Richard Coudenhove-Kalergi, L-1359 Luxembourg, Grand Duchy of Luxembourg, poly@crans.org

## 1. Introduction

In 2005, Nualart and Peccati [36], discovered the remarkable fact that a sequence of multiple Wiener-Itô integrals, that is, members of a Wiener chaos, converges in distribution to a Gaussian random variable if and only if their second and fourth moments converge to the corresponding moments of the limiting random variable. The proof in [36] relies on a theorem in stochastic calculus, namely the Dubins-Schwarz Theorem. Although the proof is elegant, it does not provide good bounds on the distances (for instance, the Kolmogorov distance) between the sequence and its limit.

A few years later, the result of Nualart and Peccati [36] was generalized and developed further by Peccati and Tudor [38] and also by Nourdin and Peccati [31]. Peccati and Tudor [38] considered the multivariate central limit theorem for vectors of multiple Wiener-Itô integrals and established that if the covariance matrices of the vectors of multiple Wiener-Itô integrals converge to that of a Gaussian vector, then joint convergence in distribution to Gaussian is equivalent to coordinate-wise convergence in distribution to Gaussian. As an application of this result, the central limit theorem for any sequence of random variables can be established by proving a central limit theorem for each term in the chaos expansion.

Another significant development from the work of Nualart and Peccati [36] is that of Nourdin and Peccati [31]. Developing considerably a remarkable intuition of Nualart and Ortiz-Latorre in [35], Nourdin and Peccati [31] established a fundamental connection between Stein's method and the Malliavin calculus via the use of the Malliavin operators  $(D, \delta, \mathbf{L})$ . This connection resulted in the derivation of errors bounds, often optimal, in the central limit theorems for random variables in the Wiener space. Of special interest in this paper is the proof of error bounds in the central limit theorem for multiple Wiener-Itô integrals, which we call the fourth moment theorem.

The success of such a connection relies on the fact that both Stein's method and the Malliavin calculus are built on some integration by parts techniques. In addition, the operators of the Malliavin calculus,  $D, \delta, \mathbf{L}$ , satisfy several nice integration by parts formulae which fit in perfectly with the so-called Stein equation. For a good overview of these techniques, we refer to the following website.

<https://sites.google.com/site/malliavinstein/home>

The work of Nourdin and Peccati [31] has added a new dimension to Stein's method. Their approach of combining Stein's method with the Malliavin calculus has led to improvements and refinements of many results in probability theory, such as the Breuer-Major theorem [10]. More recently, this approach has been successfully used to obtain central limit theorems in stochastic geometry, stochastic calculus, statistical physics, and for zeros of random polynomials. It has also been

extended to different settings as in non-commutative probability and Poisson chaos. Of particular interest is the connection between the Nourdin-Peccati analysis and information theory, which was recently revealed in [27, 34]. An overview of these new developments can also be found in the above website.

The approach of Nourdin and Peccati [31] entails the use of the so-called product formula for Wiener integrals. The use of this formula makes the proofs rather involved since it relies on subtle combinatorial arguments. Very recently, starting with the work of Ledoux [26], a new approach to the fourth moment theorem was developed by Azmoodeh, Campese and Poly [2] by combining Stein's method with the Dirichlet form calculus. An advantage of this new approach is that it provides a simpler proof of the theorem by avoiding completely the use of the product formula. Moreover, since a Dirichlet space is a more general concept than the Wiener space, the former contains examples of fourth moment theorems that cannot be realized on the latter.

A more algebraic flavor of this approach has enabled Azmoodeh, Malicet and Poly [3] to prove that convergence of pairs of moments other than the 2nd and 4th (for example, the 6th and 68th) also implies the central limit theorem. This new approach seems to open up new possibilities and perhaps also central limit theorems on manifolds.

This paper is an exposition on the connections between Stein's method and the Malliavin calculus and between Stein's method and Dirichlet forms, and on how these connections are exploited in proving the fourth moment theorem.

## 2. Stein's method

### 2.1. How it began

Stein's method began with Charles Stein using his own approach in the 1960's to prove the combinatorial central limit theorems of Wald and Wolfowitz [44] and of Hoeffding [24]. Motivated by permutation tests in nonparametric statistics, Wald and Wolfowitz [44] proved that under certain conditions,  $\sum_{i=1}^n a_i b_{\pi(i)}$  converges in distribution to the standard normal distribution  $\mathcal{N}(0, 1)$ , where  $\{a_i, b_j : i, j = 1, \dots, n\}$  are real numbers and  $\pi$  a random permutation of  $\{1, \dots, n\}$ . Hoeffding [24] generalized the result of Wald and Wolfowitz [44] to  $\sum_{i=1}^n c_{i\pi(i)}$ , where  $\{c_{ij} : i, j = 1, \dots, n\}$  is a square array of real numbers and  $\pi$  a random permutation of  $\{1, \dots, n\}$ .

Let  $W = \sum_{i=1}^n c_{i\pi(i)}$  and let  $\phi$  be the characteristic function of  $W$ . Assume  $c_{i \cdot} = c_{\cdot j} = 0$  where  $c_{i \cdot} = \sum_{j=1}^n c_{ij}/n$  and  $c_{\cdot j} = \sum_{i=1}^n c_{ij}/n$  (which implies  $\mathbb{E}[W] = 0$ ), and also assume  $\text{Var}(W) = 1$ . While using exchangeable pairs to show that

$$\phi'(t) \simeq -t\phi(t) \tag{2.1}$$

Stein realized that there was nothing special about the complex exponentials. Since

$\phi'(t) = i\mathbb{E}[We^{itW}]$  and  $-t\phi(t) = -t\mathbb{E}[e^{itW}] = -\frac{1}{i}\mathbb{E}[\frac{\partial}{\partial W}e^{itW}]$ , replacing the complex exponential by an arbitrary function  $f$ , (2.1) becomes

$$\mathbb{E}[Wf(W)] \simeq \mathbb{E}[f'(W)].$$

By letting  $f$  be a solution, say  $f_h$ , of the differential equation

$$f'(w) - wf(w) = h(w) - \mathbb{E}[h(Z)],$$

where  $h$  is a bounded function and  $Z \sim \mathcal{N}(0, 1)$ , one obtains

$$\mathbb{E}[h(W)] - \mathbb{E}[h(Z)] = \mathbb{E}[f'_h(W) - Wf_h(W)].$$

How close  $\mathcal{L}(W)$  is to  $\mathcal{N}(0, 1)$  can then be determined by bounding  $\mathbb{E}[f'_h(W) - Wf_h(W)]$ . There is no inversion of the characteristic function.

This story of how Stein's method began is based on personal communications with Charles Stein and Persi Diaconis and also on an interview of Stein in Leong [28].

## 2.2. A general framework

Stein's method for normal approximation was published in his seminal 1972 paper in the Proceedings of the Sixth Berkeley Symposium. Although the method was developed for normal approximation, Stein's ideas were very general and the method was modified by Chen [13] for Poisson approximation. Since then the method has been constantly developed and applied to many approximations beyond normal and Poisson and in finite as well as infinite dimensional spaces. As Stein's method works well for dependent random variables, it has been applied, and continues to be applied, to a large number of problems in many different fields. The method, together with its applications, continues to grow and remains a very active research area. See, for example, Stein [43], Arratia, Goldstein and Gordon [1], Barbour, Holst and Janson [8], Diaconis and Holmes [21], Barbour and Chen [6, 7], Chatterjee, Diaconis and Meckes [12], Chen, Goldstein and Shao [15], Ross [40], Shih [41], Nourdin and Peccati [32], and Chen and Röllin [16, 17].

In a nutshell, Stein's method can be described as follows. Let  $W$  and  $Z$  be random elements taking values in a space  $\mathcal{S}$  and let  $\mathcal{X}$  and  $\mathcal{Y}$  be some classes of real-valued functions defined on  $\mathcal{S}$ . In approximating the distribution  $\mathcal{L}(W)$  of  $W$  by the distribution  $\mathcal{L}(Z)$  of  $Z$ , we write  $\mathbb{E}[h(W)] - \mathbb{E}[h(Z)] = \mathbb{E}[Lf_h(W)]$  for a test function  $h \in \mathcal{Y}$ , where  $L$  is a linear operator (Stein operator) from  $\mathcal{X}$  into  $\mathcal{Y}$  and  $f_h \in \mathcal{X}$  a solution of the equation

$$Lf = h - \mathbb{E}[h(Z)] \quad (\text{Stein equation}). \quad (2.2)$$

The error  $\mathbb{E}[Lf_h(W)]$  can then be bounded by studying the solution  $f_h$  and exploiting the probabilistic properties of  $W$ . The operator  $L$  characterizes  $\mathcal{L}(Z)$  in the sense that  $\mathcal{L}(W) = \mathcal{L}(Z)$  if and only if for a sufficiently large class of functions  $f$  we have

$$\mathbb{E}[Lf(W)] = 0 \quad (\text{Stein identity}). \quad (2.3)$$

In normal approximation, where  $\mathcal{L}(Z)$  is the standard normal distribution, the operator used by Stein [42] is given by  $Lf(w) = f'(w) - wf(w)$  for  $w \in \mathbb{R}$ , and in Poisson approximation, where  $\mathcal{L}(Z)$  is the Poisson distribution with mean  $\lambda > 0$ , the operator  $L$  used by Chen [13] is given by  $Lf(w) = \lambda f(w+1) - wf(w)$  for  $w \in \mathbb{Z}_+$ . However the operator  $L$  is not unique even for the same approximating distribution but depends on the problem at hand. For example, for normal approximation  $L$  can also be taken to be the generator of the Ornstein-Uhlenbeck process, that is,  $Lf(w) = f''(w) - wf'(w)$ , and for Poisson approximation,  $L$  taken to be the generator of an immigration-death process, that is,  $Lf(w) = \lambda[f(w+1) - f(w)] + w[f(w-1) - f(w)]$ . This generator approach, which is due to Barbour [4], allows extensions to multivariate and process settings. Indeed, for multivariate normal approximation,  $Lf(w) = \Delta f(w) - w \cdot \nabla f(w)$ , where  $f$  is defined on the Euclidean space; see Barbour [5] and Götze [23], and also Reinert and Röllin [39] and Meckes [30].

### 2.3. Normal approximation

In many problems of normal approximation, the random variable  $W$  whose distribution is to be approximated satisfies this equation

$$\mathbb{E}[Wf(W)] = \mathbb{E}[T_1 f'(W + T_2)]. \quad (2.4)$$

where  $T_1$  and  $T_2$  are some random variables defined on the same probability space as  $W$ , and  $f$  is an absolutely continuous function for which the expectations in (2.4) exist. Examples of  $W$  satisfying this equation include sums of locally dependent random variables as considered in Chen and Shao [19] and exchangeable pairs as defined in Stein [43]. More generally, a random variable  $W$  satisfies (2.4) if there is a Stein coupling  $(W, W', G)$  where  $W, W', G$  are defined on a common probability space such that  $\mathbb{E}[Wf(W)] = \mathbb{E}[Gf(W') - Gf(W)]$  for absolutely continuous functions  $f$  for which the expectations exist (see Chen and Röllin [17]). In all cases it is assumed that  $\mathbb{E}[W] = 0$  and  $\text{Var}(W) = 1$ . Letting  $f(w) = w$ , we have  $1 = \mathbb{E}[W^2] = \mathbb{E}[T_1]$ .

As an illustration, let  $(W, W')$  be an exchangeable pair of random variables with  $\mathbb{E}(W) = 0$  and  $\text{Var}(W) = 1$  such that  $\mathbb{E}[W' - W|W] = -\lambda W$  for some  $\lambda > 0$ . Since the function  $(w, w') \mapsto (w' - w)(f(w') + f(w))$  is anti-symmetric, the exchangeability of  $(W, W')$  implies

$$E[(W' - W)(f(W') + f(W))] = 0.$$

From this we obtain

$$\begin{aligned} \mathbb{E}[Wf(W)] &= \frac{1}{2\lambda} \mathbb{E}[(W' - W)(f(W') - f(W))] \\ &= \frac{1}{2\lambda} \mathbb{E}\left[(W' - W)^2 \int_0^1 f'(W + (W' - W)t) dt\right] = \mathbb{E}[T_1 f'(W + T_2)] \end{aligned}$$

where  $T_1 = \frac{1}{2\lambda}(W' - W)^2$ ,  $T_2 = (W' - W)U$ , and  $U$  uniformly distributed on  $[0, 1]$  and independent of  $W, W', T_1$  and  $T_2$ .

Let  $f_h$  be the unique bounded solution of the Stein equation

$$f'(w) - wf(w) = h(w) - \mathbb{E}[h(Z)], \quad (2.5)$$

where  $w \in \mathbb{R}$ ,  $Z \sim \mathcal{N}(0, 1)$ , and  $h$  a bounded test function.

The following boundedness properties of  $f_h$  are useful for bounding the errors in the approximation.

If  $h$  is bounded, then

$$\|f_h\|_\infty \leq \sqrt{2\pi}\|h\|_\infty, \quad \|f'_h\|_\infty \leq 4\|h\|_\infty.$$

If  $h$  is absolutely continuous, then

$$\|f_h\|_\infty \leq 2\|h'\|_\infty, \quad \|f'_h\|_\infty \leq \sqrt{2/\pi}\|h'\|_\infty, \quad \|f''_h\|_\infty \leq 2\|h'\|_\infty.$$

If  $h = \mathbf{1}_{(-\infty, x]}$ , then for all  $w, v \in \mathbb{R}$ ,

$$0 \leq f_h(w) \leq \sqrt{2\pi}/4, \quad |wf_h(w)| \leq 1, \quad |f'_h(w)| \leq 1, \quad |f'_h(w) - f'_h(v)| \leq 1.$$

These can be found in Lemmas 2.4 and 2.5 of Chen, Goldstein and Shao [15] and in Lemma 2.2 of Chen and Shao [20]

Assume that  $\mathbb{E}[W] = 0$  and  $\text{Var}(W) = 1$ . From (2.4) and (2.5),

$$\begin{aligned} \mathbb{E}[h(W)] - \mathbb{E}[h(Z)] &= \mathbb{E}[f'_h(W) - T_1 f'_h(W + T_2)] \\ &= \mathbb{E}[T_1(f'_h(W) - f'_h(W + T_2))] + \mathbb{E}[(1 - T_1)f'_h(W)] \end{aligned} \quad (2.6)$$

Different techniques have been developed for bounding the error term on the right side of (2.6). We will consider two special cases.

Case 1. Assume that  $T_1 = 1$ . This is the case of zero-bias coupling. See Goldstein and Reinert [22], and also Chen, Goldstein and Shao [15]. Let  $h$  be absolutely continuous such that  $h'$  is bounded. From (2.6),

$$|\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]| = \left| \mathbb{E} \left[ \int_0^{T_2} f''_h(W + t) dt \right] \right| \leq \|f''_h\|_\infty \mathbb{E}[|T_2|] \leq 2\|h'\|_\infty \mathbb{E}[|T_2|].$$

From this we obtain the following bound on the Wasserstein distance between  $\mathcal{L}(W)$  and  $\mathcal{N}(0, 1)$ .

$$d_W(\mathcal{L}(W), \mathcal{N}(0, 1)) := \sup_{|h(x) - h(y)| \leq |x - y|} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]| \leq 2\mathbb{E}[|T_2|].$$

Note that  $d_W(\mathcal{L}(W), \mathcal{N}(0, 1)) = \|F - \Phi\|_1$  where  $F(x) = P(W \leq x)$  and  $\Phi(x) = P(Z \leq x)$ .

If  $W = X_1 + \dots + X_n$  where  $X_1, \dots, X_n$  are independent random variables with  $\mathbb{E}[X_i] = 0$ ,  $\text{Var}(X_i) = \sigma^2$  and  $\mathbb{E}[|X_i|^3] = \gamma_i < \infty$ , then  $T_2 = \xi_I - X_I$  where the  $X_i$ , the  $\xi_i$  and  $I$  are independent,  $\mathbb{E}[|\xi_i|] = \gamma_i/2\sigma_i^2$ , and  $P(I = i) = \sigma_i^2$ . Note that

$\sum \sigma_i^2 = \text{Var}(W) = 1$ . Then the bound on the Wasserstein distance between  $\mathcal{L}(W)$  and  $\mathcal{N}(0, 1)$  is  $2\mathbb{E}[|\xi_I - X_I|] \leq 3 \sum \gamma_i$ .

It is more difficult to obtain a bound on the Kolmogorov distance between  $\mathcal{L}(W)$  and  $\mathcal{N}(0, 1)$ , namely  $\sup_{x \in \mathbb{R}} |P(W \leq x) - \Phi(x)|$ . Such a bound can be obtained by induction or the use of a concentration inequality. For induction, see [9]. For the use of a concentration inequality, see Chen [14] and Chen and Shao [18] for sums of independent random variables, and Chen and Shao [19] for sums of locally dependent random variables. See also Chen, Goldstein and Shao [15]. For sums of independent random variables, Chen and Shao [18] obtained a bound of  $4.1 \sum \gamma_i$  on the Kolmogorov distance.

Case 2. Assume that  $T_2 = 0$ . This is the case if  $W$  is a functional of independent Gaussian random variables as considered by Chatterjee [11] or a functional of Gaussian random fields as considered by Nourdin and Peccati [31]. In this case, (2.6) becomes

$$\mathbb{E}[h(W)] - \mathbb{E}[h(Z)] = \mathbb{E}[(1 - T_1)f'_h(W)] = \mathbb{E}[(1 - \mathbb{E}[T_1|W])f'_h(W)].$$

Let  $h$  be such that  $|h| \leq 1$ . Then we obtain the following bound on the total variation distance between  $\mathcal{L}(W)$  and  $\mathcal{N}(0, 1)$ .

$$\begin{aligned} d_{\text{TV}}(\mathcal{L}(W), \mathcal{N}(0, 1)) &:= \frac{1}{2} \sup_{|h| \leq 1} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]| \\ &\leq \frac{1}{2} \|f'_h\|_{\infty} \mathbb{E}[|1 - \mathbb{E}[T_1|W]|] \\ &\leq 2\sqrt{\text{Var}(\mathbb{E}[T_1|W])} \end{aligned}$$

While Chatterjee [11] used a Poincaré inequality of second order to bound  $2\sqrt{\text{Var}(\mathbb{E}[T_1|W])}$ , Nourdin and Peccati [31] deployed the Malliavin calculus. In the next two sections, we will discuss how the Malliavin calculus is used to bound  $2\sqrt{\text{Var}(\mathbb{E}[T_1|W])}$ .

### 3. Malliavin calculus

#### 3.1. A brief history

The Malliavin calculus was born in 1976 in a symposium in Kyoto. Paul Malliavin presented a remarkable theory which extended the powerful calculus of variations to the framework of stochastic calculus. The initial goal of the theory was to provide a probabilistic proof of the Hörmander criterion (Hörmander [25]) of hypoellipticity by relating the smoothness of the solutions of a second order partial differential equation with the smoothness of the law of the solution of a stochastic differential equation. In order to prove that a random variable  $X$  has a smooth density, Paul Malliavin introduced the following very efficient criterion.

**Lemma 3.1.** *Assume that for each  $k \geq 1$ , there is  $C_k > 0$  such that for any  $\phi \in \mathcal{C}_C^\infty$ , the class of  $\mathcal{C}^\infty$  functions with compact support, we have*

$$\left| \mathbb{E} \left[ \phi^{(k)}(X) \right] \right| \leq C_k \|\phi\|_\infty, \quad (3.1)$$

then the distribution of  $X$  has a  $\mathcal{C}^\infty$  density.

In order to prove the inequality (3.1) for  $X$ , Malliavin showed, by integration by parts, that for some suitable weight  $H_k$ ,

$$\mathbb{E}[\phi^{(k)}(X)] = \mathbb{E}[\phi(X)H_k]. \quad (3.2)$$

Hence, (3.1) holds with  $C_k = \mathbb{E}[|H_k|]$ . In many situations of interest,  $X = F(G_1, \dots, G_n)$  where the  $\{G_i\}_{i \geq 1}$  are i.i.d. Gaussian and  $F$  is smooth. The equation (3.2) can be proved by integration by parts through the "Gaussian structure" of  $X$  together with some "non-degeneracy" assumption on  $F$ . Since its introduction, the Malliavin calculus has been extended and used in many different areas of probability theory. However, regardless of the application, the central role of the Malliavin calculus always consists in proving that some integration by parts formula holds. In the present exposition, we use integration by parts not for proving the smoothness of a density, but for establishing Stein's bounds.

### 3.2. Malliavin derivatives

Let  $\mathcal{H}$  be a real separable Hilbert space, typically  $\mathcal{H} = L^2(\mathbb{R}_+)$  but the particular choice of  $\mathcal{H}$  does not matter. We denote by  $X = \{X(h), h \in \mathcal{H}\}$  an *isonormal Gaussian process* over  $\mathcal{H}$ . That means,  $X$  is a centered Gaussian family of random variables defined in some probability space  $(\Omega, \mathcal{F}, P)$ , with covariance given by

$$\mathbb{E}[X(h)X(g)] = \langle h, g \rangle_{\mathcal{H}},$$

for any  $h, g \in \mathcal{H}$ . We also assume that  $\mathcal{F}$  is generated by  $X$ .

Let  $\mathcal{S}$  be the set of all cylindrical random variables of the form:

$$F = g(X(\phi_1), \dots, X(\phi_n)), \quad (3.3)$$

where  $n \geq 1$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  is an infinitely differentiable function such that its partial derivatives have polynomial growth, and  $\phi_i \in \mathcal{H}$ ,  $i = 1, \dots, n$ . The *Malliavin derivative* of  $F$  with respect to  $X$  is the element of  $L^2(\Omega, \mathcal{H})$  defined as

$$DF = \sum_{i=1}^n \frac{\partial g}{\partial x_i}(X(\phi_1), \dots, X(\phi_n)) \phi_i. \quad (3.4)$$

In particular,  $DX(h) = h$  for every  $h \in \mathcal{H}$ . By iteration, one can define the  $m$ -th derivative  $D^m F$ , which is an element of  $L^2(\Omega, \mathcal{H}^{\odot m})$  for every  $m \geq 2$ , where  $\mathcal{H}^{\odot m}$  commonly stands for the  $m$ -th symmetric tensor product of  $\mathcal{H}$ . Indeed, we set



$$D^m F = \sum_{i_1, \dots, i_m=1}^n \frac{\partial^m g}{\partial x_{i_1} \cdots \partial x_{i_m}} \left[ X(\phi_1), \dots, X(\phi_n) \right] \phi_{i_1} \otimes \cdots \otimes \phi_{i_m}. \quad (3.5)$$

We stress that we rather use a symmetric tensor product instead of the usual one because of the celebrated Schwarz rule  $\partial_x \partial_y = \partial_y \partial_x$ , which forces  $D^m F$  to be a *symmetric* element of  $\mathcal{H}^{\odot m}$ . For  $m \geq 1$  and  $p \geq 1$ ,  $\mathbb{D}^{m,p}$  denotes the closure of  $\mathcal{S}$  with respect to the norm  $\|\cdot\|_{m,p}$ , defined by the relation

$$\|F\|_{m,p}^p = \mathbb{E} [|F|^p] + \sum_{i=1}^m \mathbb{E} [\|D^i F\|_{\mathcal{H}^{\otimes i}}^p].$$

We often use the notation  $\mathbb{D}^\infty := \bigcap_{m \geq 1} \bigcap_{p \geq 1} \mathbb{D}^{m,p}$ . To justify properly the validity of the closure procedure of  $\mathcal{S}$  with respect to the norm  $\|\cdot\|_{m,p}$  once needs to prove that the Malliavin derivatives are *closable*. Indeed, the closability is required to ensure that the limit of  $DF_n$  does not depend on the choice of the approximating sequence  $F_n$ .

Another operator closely related to  $D$  is the Ornstein-Uhlenbeck operator. For  $F = \phi(X(h_1), \dots, X(h_n))$ , we set

$$\mathbf{L}[F] = \Delta \phi(X(h_1), \dots, X(h_n)) - \sum_{i=1}^n X(h_i) \frac{\partial \phi}{\partial x_i}(X(h_1), \dots, X(h_n)). \quad (3.6)$$

### 3.3. Wiener chaos and multiple integrals

For every  $k \geq 1$ , we denote by  $\mathcal{H}_k$  the  $k$ -th *Wiener chaos* of  $X$  defined as the closed linear subspace of  $L^2(\Omega)$  generated by the family of random variables  $\{H_k(X(h)), h \in \mathcal{H}, \|h\|_{\mathcal{H}} = 1\}$ , where  $H_k$  is the  $k$ -th Hermite polynomial given by

$$H_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} \left( e^{-\frac{x^2}{2}} \right).$$

For any  $k \geq 1$ , we denote by  $\mathcal{H}^{\otimes k}$  the  $k$ -th tensor product of  $\mathcal{H}$ . Set, for any  $h \in \mathcal{H}$  such that  $\|h\|_{\mathcal{H}} = 1$ ,

$$I_k(h^{\otimes k}) = H_k(X(h)). \quad (3.7)$$

When  $\phi = h_1^{\otimes k_1} \otimes \cdots \otimes h_p^{\otimes k_p}$  with  $(h_i)_{1 \leq i \leq p}$  an orthonormal system and  $k_1 + \cdots + k_p = k$ , we extend (3.7) by

$$I_k(\phi) = \prod_{j=1}^p H_{k_j}(X(h_j)). \quad (3.8)$$

Then  $I_k$  is a linear isometry between the symmetric tensor product  $\mathcal{H}^{\odot k}$  (equipped with the modified norm  $\sqrt{k!} \|\cdot\|_{\mathcal{H}^{\odot k}}$ ) and the  $k$ -th Wiener chaos  $\mathcal{H}_k$ . In the particular case  $\mathcal{H} = L^2(A, \mathcal{A}, \mu)$ , where  $\mu$  is a  $\sigma$ -finite measure without

atoms, then  $\mathcal{H}^{\odot k}$  coincides with the space  $L_s^2(\mu^k)$  of symmetric functions which are square integrable with respect to the product measure  $\mu^k$ , and for any  $f \in \mathcal{H}^{\odot k}$  the random variable  $I_k(f)$  is commonly denoted as the multiple stochastic integral of  $f$  with respect to the centered Gaussian measure generated by  $X$ .

The following fact is fundamental in the theory of Gaussian spaces.

**Theorem 3.1.** *Any random variable  $F \in L^2(\Omega)$  admits an orthogonal decomposition of the form*

$$F = \mathbb{E}[F] + \sum_{k=1}^{\infty} I_k(f_k),$$

where the kernels  $f_k \in \mathcal{H}^{\odot k}$  are uniquely determined by  $F$ . In the sequel, we shall also denote  $J_k(F) = I_k(f_k)$ .

The random variables  $I_k(f_k)$  inherit some properties from the algebraic structure of the Hermite polynomials, such that the product formula (3.9) below. To state it, let us introduce a definition. Let  $\{e_i, i \geq 1\}$  be a complete orthonormal system in  $\mathcal{H}$ .

**Definition 3.1.** Given  $f \in \mathcal{H}^{\odot k}$  and  $g \in \mathcal{H}^{\odot j}$ , for every  $r = 0, \dots, k \wedge j$ , the contraction of  $f$  and  $g$  of order  $r$  is the element of  $\mathcal{H}^{\otimes(k+j-2r)}$  defined by

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

When  $\mathcal{H} = L^2(\mathbb{R}_+)$ , the latter formula simply becomes

$$\begin{aligned} f \otimes_r g(x, y) &= \int_{\mathbb{R}_+^r} f(x_1, x_2, \dots, x_{k-r}, t_1, \dots, t_r) g(y_1, y_2, \dots, y_{k-r}, t_1, \dots, t_r) dt_1 \dots dt_r. \end{aligned}$$

The contraction  $f \otimes_r g$  is not necessarily symmetric, and we denote by  $f \tilde{\otimes}_r g$  its symmetrization. We end this section by summarizing four important properties of *Wiener chaos* which will be used in the sequel. Throughout the sequel,  $\mathbf{Ker}$  denotes the kernel of a linear operator and  $\mathbf{Id}$  stands for the identity operator.

- (1)  $\mathcal{H}_k = \mathbf{Ker}(\mathbf{L} + k\mathbf{Id})$ ,
- (2)  $\mathcal{H}_k \subset \mathbb{D}^\infty$ ,
- (3) For every  $(p, q) \in \mathbb{N}^2$ , we have the following product formula

$$I_p(f)I_q(g) = \sum_{r=0}^{p \wedge q} r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}(f \tilde{\otimes}_r g). \quad (3.9)$$

- (4) For any  $f \in \mathcal{H}^{\odot k}$ , we have the following isometry:

$$\mathbb{E}[I_k(f)^2] = k! \|f\|_{\mathcal{H}^{\otimes k}}^2.$$

### 3.4. Main properties of Malliavin operators

The Malliavin derivative  $D$ , defined in (3.4), obeys the following *chain rule*. If  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable with bounded partial derivatives and if  $F = (F_1, \dots, F_n)$  is a vector of elements of  $\mathbb{D}^{1,2}$ , then  $\varphi(F) \in \mathbb{D}^{1,2}$  and

$$D \varphi(F) = \sum_{i=1}^n \frac{\partial \varphi}{\partial x_i}(F) D F_i. \quad (3.10)$$

The domain  $\mathbb{D}^{1,2}$  can be precisely related to the Wiener chaos decomposition. Indeed,

$$\mathbb{D}^{1,2} = \left\{ F \in L^2(\Omega) : \sum_{k=1}^{\infty} k \|J_k F\|_{L^2(\Omega)}^2 < \infty \right\}. \quad (3.11)$$

In the particular case where  $\mathcal{H} = L^2(A, \mathcal{A}, \mu)$  (with  $\mu$  non-atomic), then the derivative of a random variable  $F$  in  $L^2(\Omega)$  whose chaotic expansion is

$$F = \mathbb{E}[F] + \sum_{k=1}^{\infty} I_k(f_k), \text{ with } f_k \in \mathcal{H}^{\odot k},$$

can be identified with the element of  $L^2(A \times \Omega)$  given by

$$D_x F = \sum_{k=1}^{\infty} k I_{k-1}(f_k(\cdot, x)), \quad x \in A. \quad (3.12)$$

The notation  $I_{k-1}(f_k(\cdot, x))$  means that we freeze one coordinate and take the Wiener-Itô integral of order  $k-1$  with respect to the  $k-1$  remaining coordinates. One should notice that, since the  $f_k$  are taken symmetric, then the above notation do not depend on the choice of the frozen coordinate. As a matter of fact,  $D_x F$  ( $x \in A$ ) is an element of  $L^2(A \times \Omega)$ . We denote by  $\delta$  the adjoint of the operator  $D$ , also called the *divergence operator*. We recall that  $\delta$  exists since the operator  $D$  is closed. A random element  $u \in L^2(\Omega, \mathcal{H})$  belongs to the domain of  $\delta$ , noted  $\text{Dom } \delta$ , if and only if it verifies  $|\mathbb{E}[\langle DF, u \rangle_{\mathcal{H}}]| \leq c_u \|F\|_{L^2(\Omega)}$  for any  $F \in \mathbb{D}^{1,2}$ , where  $c_u$  is a constant depending only on  $u$ . If  $u \in \text{Dom } \delta$ , then the random variable  $\delta(u)$  is defined by the duality relationship (customarily called *integration by parts formula*)

$$\mathbb{E}[F \delta(u)] = \mathbb{E}[\langle DF, u \rangle_{\mathcal{H}}], \quad (3.13)$$

which holds for every  $F \in \mathbb{D}^{1,2}$ . More generally, if  $F \in \mathbb{D}^{1,2}$  and  $u \in \text{Dom } \delta$  are such that the three expectations  $\mathbb{E}[F^2 \|u\|_{\mathcal{H}}^2]$ ,  $\mathbb{E}[F^2 \delta(u)^2]$  and  $\mathbb{E}[\langle DF, u \rangle_{\mathcal{H}}^2]$  are finite, then  $Fu \in \text{Dom } \delta$  and

$$\delta(Fu) = F \delta(u) - \langle DF, u \rangle_{\mathcal{H}}. \quad (3.14)$$

The *Ornstein-Uhlenbeck* operator  $\mathbf{L}$ , defined in (3.6) satisfies the following relation

$$\mathbf{L} = \sum_{k=0}^{\infty} -k J_k, \quad (3.15)$$

expressing that  $\mathbf{L}$  is diagonalizable with spectrum  $-\mathbb{N}$  with the Wiener chaos being its eigenspaces. Besides, the domain of  $\mathbf{L}$  is

$$\text{Dom}(\mathbf{L}) = \left\{ F \in L^2(\Omega) : \sum_{k=1}^{\infty} k^2 \|J_k F\|_{L^2(\Omega)}^2 < \infty \right\} = \mathbb{D}^{2,2}. \quad (3.16)$$

There is an important relation between the operators  $D$ ,  $\delta$  and  $\mathbf{L}$ . A random variable  $F$  belongs to  $\mathbb{D}^{2,2}$  if and only if  $F \in \text{Dom}(\delta D)$  (i.e.  $F \in \mathbb{D}^{1,2}$  and  $DF \in \text{Dom} \delta$ ) and, in this case,

$$\delta DF = -\mathbf{L}F. \quad (3.17)$$

In particular, if  $F \in \mathbb{D}^{2,2}$  and  $H, G \in \mathbb{D}^{1,2}$  are such that  $HG \in \mathbb{D}^{1,2}$ , then

$$\begin{aligned} -\mathbb{E}[HG\mathbf{L}F] &= \mathbb{E}[HG\delta DF] \\ &= \mathbb{E}[\langle D(HG), DF \rangle_{\mathcal{H}}] \\ &= \mathbb{E}[H\langle DG, DF \rangle_{\mathcal{H}}] + \mathbb{E}[G\langle DH, DF \rangle_{\mathcal{H}}]. \end{aligned}$$

We end this section by introducing the operator  $\mathbf{L}^{-1}$  which is central in the next section in concern with applications of the Malliavin calculus to Stein's method. For any  $F \in L^2(\Omega)$ , we know that  $F$  can be expanded over the Wiener chaos, namely one has  $F = \mathbb{E}[F] + \sum_{k=1}^{\infty} J_k F$ .

Thus we set

$$\mathbf{L}^{-1}F = \mathbf{L}^{-1} \left[ F - \mathbb{E}[F] \right] = \sum_{k=1}^{\infty} \frac{-1}{k} J_k F.$$

One can show by using (3.16) that  $\mathbf{L}^{-1}F \in \text{Dom}(\mathbf{L})$ . Thus, relying on (3.15) we have  $\mathbf{L}\mathbf{L}^{-1}F = F - \mathbb{E}[F]$ . The operator  $\mathbf{L}^{-1}$  is called the *pseudo-inverse* of  $\mathbf{L}$ . Notice that, since  $\text{Dom}(\mathbf{L}) \subset \mathbb{D}^{1,2}$  [compare (3.11) and (3.16)], for any  $F \in \mathbb{D}^{1,2}$  the quantity  $\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}$  is well defined.

#### 4. Connecting Stein's method with Malliavin calculus

As is discussed in Section 2, the Stein operator  $L$  for normal approximation is given by  $Lf(w) = f'(w) - wf(w)$  and the equation

$$\mathbb{E}[f'(Z) - Zf(Z)] = 0 \quad (4.1)$$

holds for absolutely continuous functions  $f$  for which the expectations exist if and only if  $Z \sim \mathcal{N}(0, 1)$ . Interestingly, this equation is nothing but a simple consequence of integration by parts. Since there is the integration by parts formula of the Malliavin calculus for functionals of general Gaussian processes, there is a natural connection between Stein's method and the Malliavin calculus. Indeed, integration by parts has been used in less general situations to construct the equation

$$\mathbb{E}[Wf(W)] = E[Tf'(W)] \quad (4.2)$$

which is Case 2 of (2.4) discussed in Subsection 2.3. Let us provide two examples below.

**Example 4.1.** Assume  $\mathbb{E}[W] = 0$  and  $\text{Var}(W) = 1$ . Then we have  $\mathbb{E}[T] = 1$ . If  $W$  has a density  $\rho > 0$  with respect to the Lebesgue measure, then by integration by parts,  $W$  satisfies (4.2) with  $T = h(W)$ , where

$$h(x) = \frac{\int_x^\infty y\rho(y)dy}{\rho(x)}.$$

If  $\rho$  is the density of  $\mathcal{N}(0,1)$ , then  $h(w) = 1$  and (4.2) reduces to (4.1).

**Example 4.2.** Let  $X = (X_1, \dots, X_d)$  be a vector of independent Gaussian random variables and let  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  be an absolutely continuous function. Let  $W = g(X)$ . Chatterjee in [11] used Gaussian interpolation and integration by parts to show that  $W$  satisfies (4.2) with  $T = h(X)$  where

$$h(x) = \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E} \left[ \sum_{i=1}^d \frac{\partial g}{\partial x_i}(x) \frac{\partial g}{\partial x_i}(\sqrt{tx} + \sqrt{1-t}X) \right] dt.$$

If  $d = 1$  and  $g$  the identity function, then  $W \sim \mathcal{N}(0,1)$ ,  $h(x) = 1$ , and again (4.2) reduces to (4.1).

As the previous example shows (see Chatterjee [11] for details), it is possible to construct the function  $h$  when one deals with sufficiently smooth functionals of a Gaussian vector. This is part of a general phenomenon discovered by Nourdin and Peccati in [31]. Indeed, consider a functional  $F$  of an isonormal Gaussian process  $X = \{X(h), h \in \mathcal{H}\}$  over a real Hilbert space  $\mathcal{H}$ . Assume  $F \in \mathbb{D}^{1,2}$ ,  $\mathbb{E}[F] = 0$  and  $\text{Var}(F) = 1$ . Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a bounded  $C^1$  function having a bounded derivative. Since  $\mathbf{L}^{-1}F \in \text{Dom}(\mathbf{L})$ ,  $\mathbf{L}^{-1}F \in \mathbb{D}^{2,2}$  and  $D\mathbf{L}^{-1}F \in \text{Dom} \delta$ . By (3.17) and  $\mathbb{E}[F] = 0$ , we have

$$F = \mathbf{L}\mathbf{L}^{-1}F = \delta(-D\mathbf{L}^{-1}F).$$

Therefore

$$\mathbb{E}[Ff(F)] = \mathbb{E}[\mathbf{L}\mathbf{L}^{-1}F \times f(F)] = \mathbb{E}[\delta(-D\mathbf{L}^{-1}F)f(F)].$$

By the integration by parts formula (3.13),

$$\mathbb{E}[\delta(-D\mathbf{L}^{-1}F)f(F)] = \mathbb{E}[\langle Df(F), -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}]$$

and by the chain rule,

$$\mathbb{E}[\langle Df(F), -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}] = \mathbb{E}[f'(F)\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}].$$

Hence

$$\mathbb{E}[Ff(F)] = \mathbb{E}[f'(F)\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}]$$

and  $F$  satisfies (4.2) with  $T = \langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}$ .

Now let  $f_h$  be the unique bounded solution of the Stein equation (2.5) where  $h : \mathbb{R} \rightarrow \mathbb{R}$  is continuous and  $|h| \leq 1$ . Then  $f_h \in \mathcal{C}^1$  and  $\|f_h'\|_\infty \leq 4\|h\|_\infty \leq 4$ , and we have

$$\begin{aligned} \mathbb{E}[h(F)] - \mathbb{E}[h(Z)] &= \mathbb{E}\{f_h'(F)[1 - \langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}]\} \\ &= \mathbb{E}\{f_h'(F)[1 - \mathbb{E}(\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}|F)]\}. \end{aligned}$$

Therefore

$$\begin{aligned} \sup_{h \in \mathcal{C}, |h| \leq 1} |\mathbb{E}[h(F)] - \mathbb{E}[h(Z)]| &\leq \|f_h'\|_\infty \mathbb{E}\left[|1 - \mathbb{E}(\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}|F)|\right] \\ &\leq 4\mathbb{E}\left[|1 - \mathbb{E}(\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}|F)|\right]. \end{aligned}$$

If  $F$  has a density with respect to the Lebesgue measure, then

$$\begin{aligned} d_{\text{TV}}(\mathcal{L}(F), \mathcal{N}(0, 1)) &:= \frac{1}{2} \sup_{|h| \leq 1} |\mathbb{E}[h(F)] - \mathbb{E}[h(Z)]| \\ &= \frac{1}{2} \sup_{h \in \mathcal{C}, |h| \leq 1} |\mathbb{E}[h(F)] - \mathbb{E}[h(Z)]| \\ &\leq 2\mathbb{E}\left[|1 - \mathbb{E}(\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}|F)|\right]. \end{aligned}$$

If, in addition,  $F \in \mathbb{D}^{1,4}$ , then  $\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}$  is square-integrable and

$$\mathbb{E}\left[|1 - \mathbb{E}(\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}|F)|\right] \leq \sqrt{\text{Var}[\mathbb{E}(\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}|F)]}.$$

Thus we have the following theorem of Nourdin and Peccati (2011).

**Theorem 4.1.** *Let  $F \in \mathbb{D}^{1,2}$  such that  $\mathbb{E}[F] = 0$  and  $\text{Var}(F) = 1$ . If  $F$  has a density with respect to the Lebesgue measure, then*

$$d_{\text{TV}}(\mathcal{L}(F), \mathcal{N}(0, 1)) \leq 2\mathbb{E}\left[|1 - \mathbb{E}(\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}|F)|\right]. \quad (4.3)$$

If, in addition,  $F \in \mathbb{D}^{1,4}$ , then

$$d_{\text{TV}}(\mathcal{L}(F), \mathcal{N}(0, 1)) \leq 2\sqrt{\text{Var}[\mathbb{E}(\langle DF, -D\mathbf{L}^{-1}F \rangle_{\mathcal{H}}|F)]}. \quad (4.4)$$

The bound (4.4) is optimal for normal approximation for functionals of Gaussian processes. Many examples can be found in the literature, and the reader is referred to this website

<https://sites.google.com/site/malliavinstein/home>

for a complete overview. One can also consult the good survey Peccati [37] with an emphasis on recent developments. For the sake of completeness, we shall illustrate the optimality of the bound (4.4) using the example of the quadratic variation of a fractional Brownian motion. To do this, let  $H \in (0, 1)$  be the Hurst parameter

of a fractional Brownian motion  $\{B_t^H\}_{t>0}$ . It is well known from the Breuer-Major Theorem that if  $0 < H < \frac{3}{4}$ , then for some suitable  $\sigma_H > 0$ ,

$$F_{n,H} := \frac{1}{\sigma_H \sqrt{n}} \sum_{k=1}^n ((B_{k+1}^H - B_k^H)^2 - 1) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Similarly, if  $H = \frac{3}{4}$ , then one can prove that for some  $\sigma_{\frac{3}{4}} > 0$ , we have

$$F_{n,\frac{3}{4}} := \frac{1}{\sigma_{\frac{3}{4}} \sqrt{n \log n}} \sum_{k=1}^n \left( (B_{k+1}^{\frac{3}{4}} - B_k^{\frac{3}{4}})^2 - 1 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Using the fact that  $(B_{k+1}^H - B_k^H)^2 - 1$  is an element of the second Wiener chaos, and the equation (3.4), one can obtain explicit bounds on  $\text{Var}[\mathbb{E}(\langle DF_{n,H}, -DL^{-1}F_{n,H} \rangle_{\mathcal{H}})]$  (see Nourdin and Peccati [33]). Since

$$\text{Var}[\mathbb{E}(\langle DF, -DL^{-1}F \rangle_{\mathcal{H}} | F)] \leq \text{Var}[\mathbb{E}(\langle DF, -DL^{-1}F \rangle_{\mathcal{H}})],$$

it follows from (4.4) that

$$d_{TV}(F_{n,H}, G) \leq c_H \begin{cases} \frac{1}{\sqrt{n}} & \text{if } H \in (0, \frac{5}{8}) \\ \frac{(\log n)^{\frac{3}{2}}}{\sqrt{n}} & \text{if } H = \frac{5}{8} \\ n^{4H-3} & \text{if } H \in (\frac{5}{8}, \frac{3}{4}) \\ \frac{1}{\log n} & \text{if } H = \frac{3}{4}. \end{cases}$$

These bounds are shown to be optimal in Nourdin and Peccati [33]. We wish to mention that for  $H > \frac{3}{4}$ ,  $F_{n,H}$  does not converge to a Gaussian distribution. Instead, it converges to the so-called Hermite distribution, which belongs to the second Wiener chaos and is therefore not Gaussian.

## 5. The Nualart-Peccati criterion of the fourth moment and Ledoux's idea

### 5.1. Some history

During the year 2005, in the seminal article [36], David Nualart and Giovanni Peccati discovered the following remarkable fact. Take  $F_n = I_k(f_n)$  a sequence of elements of the  $k$ -th Wiener chaos. Then,  $F_n$  converges in law towards the Gaussian measure  $\mathcal{N}(0, 1)$  if and only if  $\mathbb{E}[F_n^2] \rightarrow 1$  and  $\mathbb{E}[F_n^4] \rightarrow 3$ . This result can be seen as a drastic simplification of the so-called *method of moments* which consists in proving that  $\mathbb{E}[F^p] \rightarrow \int_{\mathbb{R}} x^p e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$ , for each positive integer  $p$ . A bit later, in Peccati and Tudor [38], the strength of this theorem was considerably reinforced by its multivariate counterpart. Indeed, Giovanni Peccati and Ciprian Tudor proved that a random vector  $F_n = (F_{1,n}, \dots, F_{d,n})$  with chaotic components converges in distribution towards a Gaussian vector with covariance  $C$  if and only if the covariance of  $F_n$  converges to  $C$  and for each  $i \in \{1, \dots, d\}$ ,  $F_{i,n}$  converges

in distribution to  $\mathcal{N}[0, C(i, i)]$ . That is to say, for chaotic random variables, the *componentwise* convergence implies the *joint* convergence. This observation yielded to a very efficient strategy for proving central convergence in the Wiener space, by somehow *decomposing* the convergence on each Wiener chaos. This approach, which is more and more used by practitioners as an alternative to the semi-martingale approach, is at the heart of a very active line of research. See this website for details and an exhaustive overview.

<https://sites.google.com/site/malliavinstein/home>

### 5.2. Overview of the proof of Nourdin and Peccati

Whereas the original proof of the fourth moment Theorem relied on some tools of stochastic analysis, Ivan Nourdin and Giovanni Peccati produced a new proof in [31] based on a suitable combination of Malliavin calculus and Stein's method. As noticed by the Nourdin and Peccati in [31] and in conclusion of previous sections 3 and 4, one is left to show that

$$\text{Var}(\mathbb{E}[T_1|W]) \leq \text{Var}[\langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}}] \rightarrow 0,$$

under the assumptions that  $\mathbb{E}[F_n^4] \rightarrow 3$  and  $\mathbb{E}[F_n^2] \rightarrow 1$ , if  $\{F_n\}$  is a sequence of elements in the  $k$ -th Wiener chaos. In fact, following their strategy we will prove that

$$\text{Var}[\langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}}] \leq \frac{k-1}{3k} [\mathbb{E}[F_n^4] - 3\mathbb{E}[F_n^2]]. \quad (5.1)$$

Step 1: computing  $\text{Var}[\langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}}]$ .

By equation (3.12), one has  $DF_n(t) = DI_k(f_n)(t) = kI_{k-1}(f_n(\cdot, t))$ . By applying the product formula for multiple integrals (3.9), we get:

$$\begin{aligned} \frac{1}{k} \langle DF_n, DF_n \rangle_{\mathcal{H}} &= k \int_0^\infty I_{k-1}(f_n(\cdot, t))^2 dt \\ &= k \int_0^\infty \sum_{r=0}^{k-1} r! \binom{k-1}{r}^2 I_{2k-2-2r}[f_n(\cdot, t) \tilde{\otimes}_r f_n(\cdot, t)] dt \\ &= k \sum_{r=0}^{k-1} r! \binom{k-1}{r}^2 I_{2k-2-2r} \left[ \int_0^\infty f_n(\cdot, t) \tilde{\otimes}_r f_n(\cdot, t) dt \right] \\ &= k \sum_{r=1}^k (r-1)! \binom{k-1}{r-1}^2 I_{2k-2r} [f_n \tilde{\otimes}_r f_n]. \\ &= k \sum_{r=1}^{k-1} (r-1)! \binom{k-1}{r-1}^2 I_{2k-2r} [f_n \tilde{\otimes}_r f_n] + k! \|f\|_{\mathcal{H}}^2. \end{aligned}$$



Taking into account that  $\mathbb{E}[\langle DF_n, DF_n \rangle_{\mathcal{H}}] = k\mathbb{E}[F_n^2] = kk!\|f\|_{\mathcal{H}}^2$ , the orthogonality of the Wiener chaos entails that:

$$\text{Var}[\langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}}] = \sum_{r=1}^{k-1} \frac{r^2}{k^2} (r!)^2 \binom{k}{r}^4 (2k-2r)! \|f_n \tilde{\otimes}_r f_n\|_{\mathcal{H}^{\odot(2k-2r)}}^2. \quad (5.2)$$

Step 2: computing  $\mathbb{E}[F_n^4]$ .

By product formula again, we have

$$F_n^2 = \sum_{r=0}^k r! \binom{k}{r}^2 I_{2k-2r}(f_n \tilde{\otimes}_r f_n).$$

This yields to

$$\mathbb{E}[F_n^4] = \sum_{r=0}^k (r!)^2 \binom{k}{r}^4 (2k-2r)! \|f_n \tilde{\otimes}_r f_n\|_{\mathcal{H}^{\odot(2k-2r)}}^2.$$

Unfortunately the latter expression is not immediately comparable with (5.2) because the “zero contractions”  $\|f_n \tilde{\otimes}_0 f_n\|_{\mathcal{H}^{\odot(2k-2r)}}^2$  do not appear in the decomposition of  $\langle DF_n, DF_n \rangle_{\mathcal{H}}$ . To face this problem, one has to take into account an intermediary expression, namely  $\mathbb{E}[F_n^2 \langle DF_n, DF_n \rangle_{\mathcal{H}}] = \frac{k}{3} \mathbb{E}[F_n^4]$ . After few combinatorial arguments respective to the symmetrizations of the contractions of the kernels  $f_n$  appearing in the scalar product, and by a suitable comparison of the terms  $\mathbb{E}[F_n^4]$ ,  $\text{Var}(\langle DF_n, DF_n \rangle_{\mathcal{H}})$  and  $\mathbb{E}[F_n^2 \langle DF_n, DF_n \rangle_{\mathcal{H}}]$  one may show that:

$$\mathbb{E}[F_n^4] - 3\mathbb{E}[F_n^2]^2 = \frac{3}{k} \sum_{r=1}^{k-1} r (r!)^2 \binom{k}{r}^4 (2k-2r)! \|f_n \tilde{\otimes}_{2k-2r} f_n\|_{\mathcal{H}^{\odot(2k-2r)}}^2. \quad (5.3)$$

Since this part is rather technical and irrelevant for our purpose, we refer the reader to the book Nourdin Peccati [32] for more precise arguments. By comparing the equations (5.2) and (5.3), one may recover (5.1).

### 5.3. About Ledoux's generalization

As we just showed, the usual way of proving the inequality (5.1) relies heavily on the product formula for multiple Wiener integrals. In particular, this gives the impression that the various combinatorics coefficients appearing in the formulae (5.1), (5.2) and (5.3) are playing a major role in the phenomenon. In 2012, in the very insightful article [26], Michel Ledoux tried to tackle this problem by using a more “algebraic” approach based on spectral theory and  $\Gamma$ -calculus. In particular, he was able to prove the inequality (5.1) without using the product formula, under some suitable spectral conditions. Unfortunately, the provided conditions seemed rather difficult to check in practice, and no new structure (in addition to the Wiener space one) with a fourth moment phenomenon was given.

## 6. The general fourth moment Theorem for Dirichlet forms

In this section, we keep the remarkable intuition of Ledoux in [26] of exploiting the algebraic and spectral properties of the chaotic random variables rather than product formulae techniques. However, at the very beginning, we will take a different path. As we will show, the fourth moment Theorem turns out to be a direct consequence of a very simple spectral assumption. This simplification will allow us to produce various examples of new structures where the phenomenon holds. We will make a crucial use of the very powerful formalism of Dirichlet forms to achieve our goals.

### 6.1. The Dirichlet structures

Originated from potential theory and physics, the Dirichlet form theory has become a central object in analysis. Under the impulsion of Beurling, Deny, Fukushima, Meyer or Mokobodzki, just to name a few, the Dirichlet forms theory has unveiled beautiful ramifications in many area of mathematics such as geometric measure theory, partial differential equations, Markov processes, Malliavin calculus... Here, we will adopt a modern formalism, namely the notion of Dirichlet structure.

We will say that a Dirichlet structure is a set  $(E, \mathcal{F}, \mu, \mathbb{D}, \Gamma)$  such that:

- (1)  $(E, \mathcal{F}, \mu)$  is a probability space,
- (2)  $\mathbb{D}$  is a vector space *dense* in  $L^2(\mu)$ ,
- (3)  $\Gamma : \mathbb{D} \times \mathbb{D} \rightarrow L^1(\mu)$  is a *bilinear, symmetric* and *non-negative* operator,
- (4)  $\mathbb{D}$  endowed with the norm  $\|X\|_{\mathbb{D}} = \sqrt{\mathbb{E}[X^2] + \mathbb{E}[\Gamma[X, X]]}$  is *complete*.
- (5) For any vectors  $X = (X_1, \dots, X_p) \in \mathbb{D}^p$  and  $Y = (Y_1, \dots, Y_q) \in \mathbb{D}^q$ , for any functions  $F \in \mathcal{C}^1(\mathbb{R}^p, \mathbb{R}) \cap \text{Lip}$  and  $G \in \mathcal{C}^1(\mathbb{R}^q, \mathbb{R}) \cap \text{Lip}$  we have  $F(X) \in \mathbb{D}$  and  $G(Y) \in \mathbb{D}$ . Besides, one has the following *functional calculus*

$$\Gamma[F(X), G(Y)] = \sum_{i=1}^p \sum_{j=1}^q \partial_i F(X) \partial_j G(Y) \Gamma[X_i, Y_j]. \quad (6.1)$$

The mapping  $(X, Y) \rightarrow \mathcal{E}[X, Y] = \mathbb{E}[\Gamma[X, Y]]$  is customarily called the *Dirichlet form* associated with the *carré-du-champ* operator  $\Gamma$ . For any of these structures, one may associate another operator  $\mathbf{L}$  defined on some domain  $\text{dom}(\mathbf{L})$  dense in  $\mathbb{D}$  such that the next integration by parts holds for any  $(X, Y) \in \text{dom}(\mathbf{L}) \times \text{dom}(\mathbf{L})$ .

$$\mathbb{E}[\Gamma[X, Y]] = -\mathbb{E}[X\mathbf{L}[Y]] = -\mathbb{E}[Y\mathbf{L}[X]]. \quad (6.2)$$

Besides, one may derive from (1-5) the next relation between  $\Gamma$  and  $\mathbf{L}$ .

$$2\Gamma[X, Y] = \mathbf{L}[XY] - Y\mathbf{L}[X] - X\mathbf{L}[Y] \quad (6.3)$$

In addition to assumptions (1-5) which characterize a general Dirichlet structure, we assume further that

- (H1)  $-\mathbf{L}$  is diagonalizable with spectrum  $\{0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots < \lambda_p < \cdots\}$ ,  
(H2) For each  $X \in \mathbf{Ker}(\mathbf{L} + \lambda_k \mathbf{Id})$ ,

$$X^2 \in \bigoplus_{\alpha \leq 2\lambda_k} \mathbf{Ker}(\mathbf{L} + \alpha \mathbf{Id}).$$

Before stating and proving our result, we stress that the previous framework covers the case of the Wiener structure. Indeed, for the Wiener structure,  $\Gamma[F, G] = \langle DF, DG \rangle_{\mathcal{H}}$  defined on  $\mathbb{D}^{1,2}$  and the corresponding  $\mathbf{L}$  is the Ornstein-Uhlenbeck operator. Properties (1-3) are straightforward, (4) proceeds from the closability of  $\mathbb{D}^{1,2}$  see subsection (3.2), and (5) is a consequence of the chain rule (3.10). (H1) comes from (3.15) and (H2) is a trivial consequence of the product formula (3.9).

## 6.2. Fourth moment theorem for Dirichlet structures with (H1) and (H2)

Let  $X$  be an eigenfunction of  $-\mathbf{L}$  associated with eigenvalue  $\lambda_k$  with  $\mathbb{E}[X^2] = 1$ . We will show that

$$\text{Var}(\Gamma[X, X]) \leq \frac{\lambda_k^2}{3} (\mathbb{E}[X^4] - 3) \quad (6.4)$$

Suppose that the inequality (6.4) is true and take  $\phi$  any test function. On the one hand we have by the chain rule (5) that  $\Gamma[\phi(X), X] = \phi'(X)\Gamma[X, X]$ . On the other hand, by integrations by parts (6.2) we also have  $\mathbb{E}[\phi'(X)\Gamma[X, X]] = \lambda_k \mathbb{E}[X\phi(X)]$ . Thus we are in the setting of the case 2 of the section 2.3 and one has the inequality

$$d_{\text{TV}}(\mathcal{L}(X), \mathcal{N}(0, 1)) \leq \mathbb{E} \left| \frac{\Gamma[X, X]}{\lambda_k} - 1 \right| \leq \frac{1}{\lambda_k} \sqrt{\text{Var}(\Gamma[X, X])}. \quad (6.5)$$

We used above the fact that  $\mathbb{E}[\Gamma[X, X]] = -\mathbb{E}[X\mathbf{L}[X]] = \lambda_k$ . Now, using (6.3) and taking into account that  $X$  is actually an eigenfunction of  $\mathbf{L}$ , we derive  $\Gamma[X, X] = \frac{1}{2}[\mathbf{L} + 2\lambda_k \mathbf{Id}][X^2]$ . Hence, we get  $[\Gamma[X, X] - \lambda_k] = \frac{1}{2}[\mathbf{L} + 2\lambda_k \mathbf{Id}][X^2 - 1]$ . The rest of the proof is as follows. We use (H2) to say that  $X^2$  and hence  $X^2 - 1$  are expanded in finitely many eigenspaces of  $-\mathbf{L}$  with maximum eigenvalue being  $2\lambda_k$ .

$$\begin{aligned} \text{Var}(\Gamma[X, X]) &= \mathbb{E} \left[ \frac{1}{2}[\mathbf{L} + 2\lambda_k \mathbf{Id}][X^2 - 1] \times \frac{1}{2}[\mathbf{L} + 2\lambda_k \mathbf{Id}][X^2 - 1] \right] \\ &= \frac{1}{2} \mathbb{E} [\mathbf{L}[X^2 - 1] \times \frac{1}{2}[\mathbf{L} + 2\lambda_k \mathbf{Id}][X^2 - 1]] \\ &+ \lambda_k \mathbb{E} [(X^2 - 1) \times \frac{1}{2}[\mathbf{L} + 2\lambda_k \mathbf{Id}][X^2 - 1]] \\ &\stackrel{(H2)}{=} \frac{1}{4} \sum_{\lambda_i \leq 2\lambda_k} (-\lambda_i)(2\lambda_k - \lambda_i) \mathbb{E}[J_k[X^2 - 1]^2] \\ &+ \lambda_k \mathbb{E} [(X^2 - 1)[\Gamma[X, X] - \lambda_k]] \\ &\stackrel{(H2)}{\leq} \lambda_k \mathbb{E} [(X^2 - 1)\Gamma[X, X]] - \lambda_k^2 \mathbb{E}[X^2 - 1] \end{aligned}$$

$$\begin{aligned} & \stackrel{(5)}{=} \lambda_k \mathbb{E} \left[ \Gamma \left[ \frac{X^3}{3} - X, X \right] \right] \\ & \stackrel{(6.2)}{=} \lambda_k^2 \left[ \frac{\mathbb{E}[X^4]}{3} - 1 \right]. \end{aligned}$$

So, the proof of (6.4) is done and one has shown (by dividing by  $\lambda_k$ ) that

$$\text{Var} \left( \Gamma \left[ X, \frac{X}{\lambda_k} \right] \right) \leq \left[ \frac{\mathbb{E}[X^4]}{3} - 1 \right]. \quad (6.6)$$

Relying on the equations (6.1) and (6.2) it is easy to see that for any function  $f$  bounded with bounded derivatives

$$\mathbb{E} \left[ f'(X) \Gamma \left[ X, \frac{X}{\lambda_k} \right] \right] = \mathbb{E} [f(X)X].$$

Since  $\Gamma \left[ X, \frac{X}{\lambda_k} \right]$  plays the same role as the term “ $\langle DF, -DL^{-1}F \rangle_{\mathcal{H}}$ ” of Section 4, Theorem 4.1 applies and can be combined with (6.6). Now take  $X_n$  as a sequence of eigenfunctions of  $\mathbf{L}$  with same eigenvalue, such that  $\mathbb{E}[X_n^2] = 1$ . We deduce that

$$d_{\text{TV}}(\mathcal{L}(X_n), \mathcal{N}(0, 1)) \leq \sqrt{\frac{\mathbb{E}[X_n^4]}{3} - 1}.$$

In particular, if  $\mathbb{E}[X_n^4] \rightarrow 3$ , then we have convergence to  $\mathcal{N}(0, 1)$  in total variation.

### 6.3. Dirichlet structures with (H1) and (H2)

Here we give two examples of Dirichlet structures which satisfy (H1) and (H2). The first and most important example is provided by the Wiener space. First from equation 3.15, one can see that the so-called Ornstein-Uhlenbeck operator is diagonalizable with spectrum  $\mathbb{N}$  and the Wiener chaos are its eigenspaces. This fact shows that the assumption (H1) is fulfilled. On the other hand, take  $k \in \mathbb{N}$  and  $X = I_k(f)$  some element of the  $k$ -th Wiener chaos. By using the product formula (3.9),

$$X^2 = \sum_{r=0}^k k! \binom{k}{r}^2 I_{2k-2r}(f \tilde{\otimes}_r f).$$

From which one can deduce that

$$X^2 \in \bigoplus_{r \leq 2k} \text{Ker}(\mathbf{L} + r\text{Id}).$$

Hence (H1) and (H2) hold true for the Wiener structure. One should notice, that the precise combinatorial coefficients arising in the product formula play no role in the assumption (H2). In some sense, the approach by Dirichlet forms being less

combinatoric is more direct than the usual proof in Nualart Peccati [36].

Not only the Dirichlet forms approach simplifies the proof of the Nualart-Peccati criterion, but also enables one to give new examples of Dirichlet structures where a fourth moment phenomenon holds. In order to avoid technicalities, we restrict the exposition to the finite dimensional case. So, let us introduce the Laguerre Dirichlet structure. Let  $\nu \geq -1$ , and  $\pi_{1,\nu}(dx) = x^{\nu-1} \frac{e^{-x}}{\Gamma(\nu)} \mathbf{1}_{(0,\infty)}(x)dx$  be the Gamma distribution with parameter  $\nu$  on  $\mathbb{R}_+$ . The associated Laguerre generator is defined for any test function  $\phi$  (in dimension one) by:

$$\mathbf{L}_{1,\nu}\phi(x) = x\phi''(x) + (\nu + 1 - x)\phi'(x). \quad (6.7)$$

By a classical tensorization procedure, we obtain the Laguerre generator in dimension  $d$  associated to the measure  $\pi_{d,\nu}(dx) = \pi_{1,\nu}(dx_1)\pi_{1,\nu}(dx_2) \cdots \pi_{1,\nu}(dx_d)$ , where  $x = (x_1, x_2, \dots, x_d)$ .

$$\mathbf{L}_{d,\nu}\phi(x) = \sum_{i=1}^d \left( x_i \partial_{i,i} \phi + (\nu + 1 - x_i) \partial_i \phi \right) \quad (6.8)$$

It is well known that (see for example Ledoux [26]) that the spectrum of  $\mathbf{L}_{d,\nu}$  is given by  $-\mathbb{N}$  and moreover that

$$\mathbf{Ker}(\mathbf{L}_{d,\nu} + p\mathbf{Id}) = \left\{ \sum_{i_1+i_2+\dots+i_d=p} \alpha(i_1, \dots, i_d) \prod_{j=1}^d L_{i_j}^{(\nu)}(x_j) \right\}, \quad (6.9)$$

where  $L_n^{(\nu)}$  stands for the Laguerre polynomial of order  $n$  with parameter  $\nu$  which is defined by

$$L_n^{(\nu)}(x) = \frac{x^{-\nu} e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^{n+\nu}).$$

Again, we have the following decomposition:

$$L^2(\mathbb{R}^d, \pi_{d,\nu}) = \bigoplus_{p=0}^{\infty} \mathbf{Ker}(\mathbf{L}_{d,\nu} + p\mathbf{Id}) \quad (6.10)$$

As a matter of fact, assumption (H1) follows. Let us check assumption (H2). Assume now that  $X$  is an eigenfunction of  $\mathbf{L}_{d,\nu}$  with eigenvalue  $-\lambda_p = -p$ . In particular,  $X$  is a multivariate polynomial of degree  $p$ . Therefore,  $X^2$  is a multivariate polynomial of degree  $2p$ . Note that by expanding  $X^2$  over the basis of multivariate Laguerre polynomials  $\prod_{j=1}^d L_{i_j}^{(\nu)}(x_j), i_j \geq 0$ , we get that  $X^2$  has a finite expansion over the first eigenspaces of the generator  $\mathbf{L}_{d,\nu}$ , i.e.

$$X^2 \in \bigoplus_{p=0}^M \mathbf{Ker}(\mathbf{L}_{d,\nu} + p\mathbf{Id}).$$

Finally, by taking care of degree reasons we infer that  $M = 2p$  and thus the assumption (H2) is valid in this structure.

## 7. Acknowledgments

We would like to thank Giovanni Peccati for his valuable comments which have helped to improve the exposition of this paper. This work is partially supported by AFR grant 4897114 at the University of Luxembourg and by Grant C-146-000-034-001 and Grant R-146-000-182-112 at the National University of Singapore.

## References

- [1] Arratia, R., Goldstein, L and Gordon, L. (1990). Poisson approximation and the Chen-Stein method. *Statist. Sci.* **5**, 403–434. With comments and a rejoinder by the authors.
- [2] Azmoodeh, E., Campese, S. and Poly, G. (2014). Fourth moment theorems for Markov diffusion generators. *J. Functional Anal.* **266**, no 4, 2341-2359.
- [3] Azmoodeh, E., Malicet, D. and Poly, G. (2013). Generalization of the Nualart Peccati criterion. *Preprint*
- [4] Barbour, A. D. (1988). Stein’s method and Poisson process convergence. *J. Appl. Probab.* **25A**, 175-184.
- [5] Barbour, A. D. (1990). Stein’s method for diffusion approximations. *Probab. Theory Related Fields* **84**, 297-322.
- [6] Barbour, A. D. and Chen, L. H. Y. editors (2005a). *An Introduction to Stein’s Method*, Lecture Notes Series **No. 4**, Institute for Mathematical Sciences, National University of Singapore, Singapore University Press and World Scientific Publishing.
- [7] Barbour, A. D. and Chen, L. H. Y. editors (2005b). *Stein’s Method and Applications*, Lecture Notes Series **No. 5**, Institute for Mathematical Sciences, National University of Singapore, Singapore University Press and World Scientific Publishing.
- [8] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*, Oxford Studies in Probability **No. 2**, Oxford University Press.
- [9] Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete.* **66**. 379-386.
- [10] Breuer, P., Major, P. (1983). Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.*, **13**, no. 3, 425-441.
- [11] Chatterjee, S (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143**, 1-40.
- [12] Chatterjee, S., Diaconis, P. and Meckes, E. (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2**, 64–106.
- [13] Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534-545.
- [14] Chen, L. H. Y. (1998). Stein’s method: some perspectives with applications. *Probability Towards 2000* (L. Accardi and C. C. Heyde, eds.), Lecture Notes in Statistics **No. 128**, Springer Verlag, 97-122.
- [15] Chen, L. H. Y., Goldstein, L. and Shao, Q. M. (2011). *Normal Approximation by Stein’s Method*. Probability and its Applications, Springer.
- [16] Chen, L. H. Y. and Röllin, A. (2013). Approximating dependent rare events. *Bernoulli* **19**, 1243-1267.
- [17] Chen, L. H. Y. and Röllin, A. (2013). Stein couplings for normal approximation. *Preprint*.
- [18] Chen, L. H. Y. and Shao, Q.M. (2001). A non-uniform Berry-Esseen bound via Stein’s method. *Prob. Theo. Rel. Fields* **120**, no 3, 236-254.

- [19] Chen, L. H. Y. and Shao, Q.M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32**, no 3, 1727-2303.
- [20] Chen, L. H. Y. and Shao, Q. M. (2005). Stein's method for normal approximation. *An Introduction to Stein's Method* (A.D. Barbour and L. H. Y. Chen, eds), Lecture Notes Series **No. 4**, Institute for Mathematical Sciences, National University of Singapore, Singapore University Press and World Scientific, 1-59.
- [21] Diaconis, P. and Holmes, S. editors (2004). *Stein's Method: Expository Lectures and Applications*, IMS Lecture Notes Monogr. Ser. **46**, Inst. Math. Statist., Beachwood, OH.
- [22] Goldsetin, L. and Reinert, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, no 4, 837-1139.
- [23] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724-739.
- [24] Hoeffding, W. (1951). A combinatorial central limit theorem. *Ann. Math. Statist.* **22**, 558-566.
- [25] Hörmander, L. (1967). Hypoelliptic second order differential equations. *Acta Math.* **119**, 147-171.
- [26] Ledoux, M. (2012). Chaos of a Markov operator and the fourth moment condition. *Ann. Probab.* **40**, no. 6, 2439-2459.
- [27] Ledoux, M., Nourdin, I. and Peccati, G. (2014). Stein's method, logarithmic Sobolev and transport inequalities. *arXiv:1403.5855*.
- [28] Leong, Y. K. (2010). Charles Stein: The Invariant, the Direct and the "Pretentious". In *Creative Minds, Charmed Lives: Interviews at Institute for Mathematical Sciences, National University of Singapore*, World Scientific Publishing, Singapore.
- [29] Malliavin, P. (1978). Stochastic calculus of variations and hypoelliptic operators. *Proc. Int. Symp. on Stoch. Diff. Equations*, Kyoto 1976, Wiley, 195-263.
- [30] Meckes, E. S. (2009). On Stein's method for multivariate normal approximation. *High dimensional probability V: the Luminy volume*, 153-178, Inst. Math. Stat. Collect., 5, Inst. Math. Statist., Beachwood, OH.
- [31] Nourdin, I. and Peccati, G. (2009). Stein's method on Wiener chaos. *Probab. Theory and Related Fields.* **145**, no. 1-2, 75-118.
- [32] Nourdin, I. and Peccati, G. (2012). *Normal Approximation with Malliavin Calculus: From Stein's Method to Universality*, Volume 192 of Cambridge Tracts in Mathematics, Cambridge University Press.
- [33] Nourdin, I. and Peccati, G. (2013). Optimal fourth moment Theorem. *Proc. Amer. Math. Soc.*, to appear.
- [34] Nourdin, I., Peccati, G. and Swan, Y. (2013). Entropy and the fourth moment phenomenon. *J. Funct. Anal.* **266**, 3170-3207.
- [35] Nualart, D. and Ortiz-Latorre, S. (2008). Central limit theorems for multiple stochastic integrals and Malliavin calculus. *Stochastic Process. Appl.* **118**, no 4, 614-628.
- [36] Nualart, D. and Peccati, G. (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, **33**, 177-193.
- [37] Peccati, G. (2014). Quantitative clts on a gaussian space: a survey of recent developments. *ESAIM Proc. and Surv.* **44**, 61-78.
- [38] Peccati, G. and Tudor, C.A. (2004). Gaussian limits for vector-valued multiple stochastic integrals. *Séminaire de Probabilités XXXVIII*, 247-262.
- [39] Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37**, 2150-2173.
- [40] Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8**, 210-293.

- [41] Shih, H. H. (2011). On Stein's method for infinite-dimensional Gaussian approximation in abstract Wiener spaces. *J. Funct. Anal.* **261**, 1236–1283.
- [42] Stein, C. (1972). 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, Univ. California Press, Berkeley, Calif.
- [43] Stein, C. (1986). *Approximate Computation of Expectations*, IMS Lecture Notes Monogr. Ser. **7**, Inst. Math. Statist., Hayward, CA.
- [44] Wald, A. and Wolfowitz, J. (1944). Statistical tests based on permutations of the observations. *Ann. Math. Statist.* **15**, 358–372.