Introduction to Malliavin Calculus

March 25, 2021

M. Hairer

Imperial College London

Contents

1 Introduction

One of the main tools of modern stochastic analysis is Malliavin calculus. In a nutshell, this is a theory providing a way of differentiating random variables defined on a Gaussian probability space (typically Wiener space) with respect to the underlying noise. This allows to develop an "analysis on Wiener space", an infinite-dimensional generalisation of the usual analytical concepts we are familiar with on $\mathbf{R}^n$. (Fourier analysis, Sobolev spaces, etc.)

The main goal of this course is to develop this theory with the proof of Hörmander's theorem in mind. This was actually the original motivation for the development of the theory and states the following. Consider a stochastic differential equation on $\mathbf{R}^n$ given by

$$\begin{aligned}dX_j(t) &= V_{j,0}(X(t))\,dt + \sum_{i=1}^m V_{j,i}(X(t)) \circ dW_i(t) \\ &= V_{j,0}(X)\,dt + \frac{1}{2}\sum_{i=1}^m \sum_{k=1}^d V_{k,i}(X)\partial_k V_{j,i}(X)\,dt + \sum_{i=1}^m V_{j,i}(X)\,dW_i(t)\;,\end{aligned}$$

where the $V_{j,k}$ are smooth functions, the W_i are i.i.d. Wiener processes, $\circ dW_i$ denotes Stratonovich integration, and dW_i denotes Itô integration. We also write this in the shorthand notation

$$dX_t = V_0(X_t)\,dt + \sum_{i=1}^m V_i(X_t) \circ dW_i(t)\;, \tag{1.1}$$

where the V_i are smooth vector fields on $\mathbf{R}^n$ with all derivatives bounded. One might then ask under what conditions it is the case that the law of X_t has a density with respect to Lebesgue measure for $t > 0$. One clear obstruction would be the existence of a (possibly time-dependent) submanifold of $\mathbf{R}^n$ of strictly smaller dimension (say $k < n$) which is invariant for the solution, at least locally. Indeed, n-dimensional Lebesgue measure does not charge any such submanifold, thus ruling out that transition probabilities are absolutely continuous with respect to it.

If such a submanifold exists, call it say $\mathscr{M} \subset \mathbf{R} \times \mathbf{R}^n$, then it must be the case that the vector fields $\partial_t - V_0$ and $\{V_i\}_{i=1}^m$ are all tangent to $\mathscr{M}$. This implies in particular that all Lie brackets between the V_j's (including $j = 0$) are tangent to $\mathscr{M}$, so that the vector space spanned by them is of dimension strictly less than $n + 1$. Since the vector field $\partial_t - V_0$ is the only one spanning the "time" direction, we conclude that if such a submanifold exists, then the dimension of the vector space $\mathscr{V}(x)$ spanned by $\{V_i(x)\}_{i=1}^m$ as well as all the Lie brackets between the V_j's evaluated at x, is strictly less than n for some values of x.

This suggests the following definition. Define $\mathscr{V}_0 = \{V_i\}_{i=1}^m$ and then set recursively

$$\mathscr{V}_{n+1} = \mathscr{V}_n \cup \{[V_i, V] \;:\; V \in \mathscr{V}_n,\ i \geq 0\}\;, \qquad \mathscr{V} = \bigcup_{n\geq 0} \mathscr{V}_n\;,$$

as well as $\mathscr{V}(x) = \operatorname{span}\{V(x) \;:\; V \in \mathscr{V}\}$.

Definition 1.1 Given a collection of vector fields as above, we say that it satisfies the parabolic Hörmander condition if $\dim \mathscr{V}(x) = n$ for every $x \in \mathbf{R}^n$.

Conversely, Frobenius's theorem (see for example [Law77]) is a deep theorem in Riemannian geometry which can be interpreted as stating that if $\dim \mathscr{V}(x) = k < n$

for all x in some open set $\mathcal{O}$ of $\mathbf{R}^n$, then $\mathbf{R}\times\mathcal{O}$ can be foliated into $k+1$-dimensional submanifolds with the property that $\partial_t - V_0$ and $\{V_i\}_{i=1}^m$ are all tangent to this foliation. This discussion points towards the following theorem.

Theorem 1.2 (Hörmander) *Consider (1.1), as well as the vector spaces $\mathcal{V}(x) \subset \mathbf{R}^n$ constructed as above. If the parabolic Hörmander condition is satisfied, then the transition probabilities for (1.1) have smooth densities with respect to Lebesgue measure.*

The original proof of this result goes back to [Hör67] and relied on purely analytical techniques. However, since it has a clear probabilistic interpretation, a more "pathwise" proof of Theorem 1.2 was sought for quite some time. The breakthrough came with Malliavin's seminal work [Mal78], where he laid the foundations of what is now known as the "Malliavin calculus", a differential calculus in Wiener space, and used it to give a probabilistic proof of Hörmander's theorem. This new approach proved to be extremely successful and soon a number of authors studied variants and simplifications of the original proof [Bis81b, Bis81a, KS84, KS85, KS87, Nor86]. Even now, more than three decades after Malliavin's original work, his techniques prove to be sufficiently flexible to obtain related results for a number of extensions of the original problem, including for example SDEs with jumps [Tak02, IK06, Cas09, Tak10], infinite-dimensional systems [Oco88, BT05, MP06, HM06, HM11], and SDEs driven by Gaussian processes other than Brownian motion [BH07, CF10, HP11, CHLT15].

1.1 Original references

The material for these lecture notes was taken mostly from the monographs [Nua06, Mal97], as well as from the note [Hai11]. Additional references to some of the original literature can be found at the end.

2 White noise and Wiener chaos

Let $H = L^2(\mathbf{R}_+, \mathbf{R}^m)$ (but for the purpose of much of this section, H could be any real separable Hilbert space), then *white noise* is a linear isometry $W\colon H \to L^2(\Omega, \mathbf{P})$ for some probability space $(\Omega, \mathbf{P})$, such that each $W(h)$ is a real-valued centred Gaussian random variable. In other words, for all $f, g \in H$, one has

$$\mathbf{E}W(h) = 0\;, \qquad \mathbf{E}W(h)W(g) = \langle h, g\rangle\;,$$

and each $W(h)$ is Gaussian. Such a construct can easily be shown to exist.

Indeed, it suffices to take a sequence $\{\xi_n\}_{n\ge0}$ of i.i.d. normal random variables and an orthonormal basis $\{e_n\}_{n\ge0}$ of H. For $h = \sum_{n\ge0} h_n e_n \in H$, it then suffices to set $W(h) = \sum_{n\ge0} h_n \xi_n$, with the convergence taking place in $L^2(\Omega, \mathbf{P})$.

Conversely, given a white noise, it can always be recast in this form (modulo possible modifications on sets of measure 0) by setting $\xi_n = W(e_n)$.

A white noise determines an m-dimensional Wiener process, which we call again W, in the following way. Write $\mathbf{1}^{(i)}_{[0,t)}$ for the element of H given by

$$(\mathbf{1}^{(i)}_{[0,t)})_j(s) = \begin{cases} 1 & \text{if } s \in [0,t) \text{ and } j = i, \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

and set $W_i(t) = W(\mathbf{1}^{(i)}_{[0,t)})$. It is then immediate to check that one has indeed

$$\mathbf{E} W_i(s) W_j(t) = \delta_{ij}(s \wedge t) \;,$$

so that this is a standard Wiener process. For arbitrary $h \in H$, one then has

$$W(h) = \sum_{i=1}^m \int_0^\infty h_i(s)\, dW_i(s) \;, \tag{2.2}$$

with the right hand side being given by the usual Wiener–Itô integral.

Let now H_n denote the nth Hermite polynomial. One way of defining these is to set $H_0 = 1$ and then recursively by imposing that

$$H_n'(x) = n H_{n-1}(x) \tag{2.3}$$

and that, for $n \geq 1$, $\mathbf{E} H_n(X) = 0$ for a normal Gaussian random variable X with variance 1. This determines the H_n uniquely, since the first condition determines H_n up to a constant, with the second condition determining the value of this constant uniquely. The first few Hermite polynomials are given by

$$H_1(x) = x \;, \qquad H_2(x) = x^2 - 1 \;, \qquad H_3(x) = x^3 - 3x \;.$$

Remark 2.1 Beware that the definition given here differs from the one given in [Nua06] by a factor $n!$, but coincides with the one given in most other parts of the mathematical literature, for example in [Mal97]. In the physical literature, they tend to be defined in the same way, but with X of variance $1/2$, so that they are orthogonal with respect to the measure with density $\exp(-x^2)$ rather than $\exp(-x^2/2)$.

There is an analogy between expansions in Hermite polynomials and expansion in Fourier series. In this analogy, the factor $n!$ plays the same role as the factor 2π that appears in Fourier analysis. Just like there, one can shift it around to simplify certain expressions, but one can never quite get rid of it.

An alternative characterisation of the Hermite polynomials is given by

$$H_n(x) = \exp\Big(-\frac{D^2}{2}\Big)x^n\;, \tag{2.4}$$

where D represents differentiation with respect to the variable x. (Yes, this is the *inverse* heat flow appearing here!) To show that this is the case, and since it is obvious that $DH_n = nH_{n-1}$, it suffices to verify that $\mathbf{E}H_n(X) = 0$ for $n \ge 1$ and H_n as in (2.4). Since the Fourier transform of x^n is $c_n\delta^{(n)}$ for some constant c_n and since $\exp(-x^2/2)$ is a fixed point for the Fourier transform, one has for $n > 0$

$$\int e^{-\frac{x^2}{2}} e^{-\frac{D^2}{2}} x^n\,dx = c_n \int e^{-\frac{k^2}{2}} e^{\frac{k^2}{2}} \delta^{(n)}(k)\,dk = c_n \int \delta^{(n)}(k)\,dk = 0\;,$$

as required.

A different recursive relation for the H_n's is given by

$$H_{n+1}(x) = xH_n(x) - H_n'(x)\;,\qquad n \ge 0\;. \tag{2.5}$$

To show that (2.5) holds, it suffices to note that

$$[f(D), x] = f'(D)\;,$$

so that indeed

$$\begin{aligned} H_{n+1}(x) &= \exp\Big(-\frac{D^2}{2}\Big)x\,x^n = xH_n(x) + \Big[\exp\Big(-\frac{D^2}{2}\Big), x\Big]\,x^n \\ &= xH_n(x) - D\exp\Big(-\frac{D^2}{2}\Big)\,x^n = xH_n(x) - H_n'(x)\;. \end{aligned}$$

Combining both recursive characterisations of H_n, we obtain for $n, m \ge 0$ the identity

$$\begin{aligned} \int H_n(x)H_m(x)e^{-x^2/2}\,dx &= \frac{1}{n+1}\int H_{n+1}'(x)H_m(x)e^{-x^2/2}\,dx \\ &= \frac{1}{n+1}\int H_{n+1}(x)\big(xH_m(x) - H_m'(x)\big)e^{-x^2/2}\,dx \\ &= \frac{1}{n+1}\int H_{n+1}(x)H_{m+1}e^{-x^2/2}\,dx\;. \end{aligned}$$

Combining this with the fact that $\mathbf{E}H_n(X) = 0$ for $n \ge 1$ and $\mathbf{E}H_0^2(X) = 1$, we immediately obtain the identity

$$\mathbf{E}H_n(X)H_m(X) = n!\delta_{n,m}\;.$$

Fix now an orthonormal basis $\{e_i\}_{i\in\mathbf{N}}$ of H. For every multiindex k, which we view as a function $k\colon \mathbf{N}\to\mathbf{N}$ such that all but finitely many values vanish, we then define a random variable Φ_k by

$$\Phi_k \stackrel{\text{def}}{=} \prod_{i\in\mathbf{N}} H_{k_i}(W(e_i)) \;. \tag{2.6}$$

It follows immediately from the above that

$$\mathbf{E}\Phi_k\Phi_\ell = k!\delta_{k,\ell}\;,\qquad k! = \prod_i k_i!\;. \tag{2.7}$$

Write now $H^{\otimes_s n}$ for the subspace of $H^{\otimes n}$ consisting of symmetric tensors. There is a natural projection $\Pi\colon H^{\otimes n}\to H^{\otimes_s n}$ given as follows. For any permutation σ of $\{1,\ldots,n\}$ write $\Pi_\sigma\colon H^{\otimes n}\to H^{\otimes n}$ for the linear map given by

$$\Pi_\sigma\big(h_1\otimes\ldots\otimes h_n\big) = h_{\sigma(1)}\otimes\ldots\otimes h_{\sigma(n)}\;.$$

We then set $\Pi = \frac{1}{n!}\sum_\sigma \Pi_\sigma$, where the sum runs over all permutations.

Writing $|k| = \sum_i k_i$, we set $e_k = \Pi\bigotimes_i e_i^{\otimes k_i}$, which is an element of $H^{\otimes_s |k|}$. Note that the vectors e_k are not orthonormal, but that instead one has

$$\langle e_k, e_\ell\rangle = \frac{k!}{|k|!}\delta_{k,\ell}\;.$$

Comparing this to (2.7), we conclude that the maps

$$I_n\colon e_k \mapsto \frac{1}{\sqrt{n!}}\Phi_k\;,\qquad |k| = n\;, \tag{2.8}$$

yield, for every $n\ge 0$, an isometry between $H^{\otimes_s n}$ and some closed subspace $\mathscr{H}_n$ of $L^2(\Omega,\mathbf{P})$. This space is called the nth homogeneous Wiener chaos after the terminology of the original article [Wie38] by Norbert Wiener where a construction similar to this was first introduced, but with quite a different motivation. As a matter of fact, Wiener's construction was based on the usual monomials instead of Hermite polynomials and, as a consequence, the analogues of the maps I_n in his context were not isometries. The first construction equivalent to the one presented here was given almost two decades later by Irving Segal [Seg56], motivated in part by constructive quantum field theory.

We now show that the isomorphisms I_n are canonical, i.e. they do not depend on the choice of basis $\{e_i\}$. For this, it suffices to show that for any $h\in H$ with $\|h\| = 1$ one has $I_n(h^{\otimes n}) = H_n(W(h))/\sqrt{n!}$. The main ingredient for this is the following lemma.

Lemma 2.2 *Let $x, y \in \mathbf{R}$ and a, b with $a^2 + b^2 = 1$. Then, one has the identity*

$$H_n(ax + by) = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} H_k(x) H_{n-k}(y) \;.$$

Proof. As a consequence of (2.4) we have

$$H_n(ax+by) = \exp\Big(-\frac{D_x^2}{2a^2}\Big)(ax+by)^n = \exp\Big(-\frac{D_x^2}{2}\Big)\exp\Big(-\frac{b^2 D_x^2}{2a^2}\Big)(ax+by)^n \;.$$

Noting that $G(D_x/a)(ax + by)^n = G(D_y/b)(ax + by)^n$, we conclude that

$$H_n(ax + by) = \exp\Big(-\frac{D_x^2}{2}\Big)\exp\Big(-\frac{D_y^2}{2}\Big)(ax + by)^n \;.$$

Applying the binomial theorem to $(ax + by)^n$ concludes the proof. □

Applying this lemma repeatedly and taking limits, we have

Corollary 2.3 *Let $a \in \ell^2$ with $\sum a_i^2 = 1$ and let $\{x_i\}_{i\in\mathbf{N}}$ be such that $\sum_i a_i x_i$ converges. Then, one has the identity*

$$H_n\Big(\sum_{i\in\mathbf{N}} a_i x_i\Big) = \sum_{|k|=n} \frac{n!}{k!} a^k \prod_{i\in\mathbf{N}} H_{k_i}(x_i) \;,$$

where $a^k \stackrel{\text{def}}{=} \prod_i a_i^{k_i}$. In particular, whenever $h \in H$ with $\|h\| = 1$, one has $I_n(h^{\otimes n}) = H_n(W(h))/\sqrt{n!}$, independently of the choice of basis used in the definition of I_n.

It turns out that, as long as Ω contains no other source of randomness than generated by the white noise W, then the spaces $\mathscr{H}_n$ span all of $L^2(\Omega, \mathbf{P})$. More precisely, denote by $\mathscr{F}$ the σ-algebra generated by W, namely the smallest σ-algebra such that the random variables $W(h)$ are $\mathscr{F}$-measurable for every $h \in H$. Then, one has

Theorem 2.4 *In the above context, one has*

$$L^2(\Omega, \mathscr{F}, \mathbf{P}) = \bigoplus_{n\ge 0} \mathscr{H}_n \;.$$

Proof. Denote by $H_N \subset H$ the subspace generated by $\{e_k\}_{k\le N}$, by $\mathscr{F}_N$ the σ-algebra generated by $\{W(h) : h \in H_N\}$, and assume that one has

$$L^2(\Omega, \mathscr{F}_N, \mathbf{P}) = \bigoplus_{n\ge 0} \mathscr{H}_n^{(N)} \;, \tag{2.9}$$

where $\mathcal{H}_n^{(N)} \subset \mathcal{H}_n$ is the image of $H_N^{\otimes_s n}$ under I_n. Let now $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ and set $X_N = \mathbf{E}(X \mid \mathcal{F}_N)$. By Doob's martingale convergence theorem, one has $X_N \to X$ in L^2, thus showing by (2.9) that $\bigoplus_{n\geq 0} \mathcal{H}_n$ is dense in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ as claimed.

To show (2.9), it only remains to show that if $X \in L^2(\Omega, \mathcal{F}_N, \mathbf{P})$ satisfies $\mathbf{E}XY = 0$ for every $Y \in \mathcal{H}_n^{(N)}$ and every $n \geq 0$, then $X = 0$. We can write $X(\omega) = f(W(e_1), \ldots, W(e_N))$ (almost surely) for some function $f\colon \mathbf{R}^N \to \mathbf{R}$ that is square integrable with respect to the standard Gaussian measure μ_N on $\mathbf{R}^N$. By assumption, one then has $\int f(x)\, P(x)\, \mu_N(dx) = 0$ for every polynomial P.

Note now that the truncated Taylor expansion $T_k^{(n)}$ of $f_k \stackrel{\text{def}}{=} (x \mapsto e^{ikx})$ satisfies the bound

$$|T_k^{(n)}(x)| = \Big| \sum_{|m|\leq n} \frac{(ikx)^m}{m!} \Big| \leq e^{|kx|} ,$$

uniformly over n, and $T_k^{(n)}$ converges of course pointwise to f_k. Since μ_N has subexponential tails it integrates $e^{|kx|}$ for every $k \in \mathbf{R}^N$, and it follows from Lebesgue's dominated convergence theorem that $\lim_{n\to\infty} T_k^{(n)} = f_k$ in $L^2(\mu_N)$.

Since $\int f(x) T_k^{(n)}(x)\, \mu_N(dx) = 0$ for every $k \in \mathbf{R}^N$ and every $N > 0$, we conclude that one must have $\int f(x)\, e^{ikx}\, \mu_N(dx) = 0$, whence the claim follows. $\square$

Remark 2.5 The fact that μ integrates exponentials is crucial here and not just an artefact of our proof. If this condition is dropped, then there are counterexamples to the claim that polynomials are dense in $L^2(\mu)$.

Let us now show what a typical element of $\mathcal{H}_n$ looks like. We have already seen that $\mathcal{H}_0$ only contains constants and $\mathcal{H}_1$ contains precisely all random variables of the form (2.2). Write $\Delta_n \subset \mathbf{R}_+^n$ for the cone consisting of points s with $0 < s_1 < \cdots < s_n$. We then have

Lemma 2.6 *Let $H = L^2(\mathbf{R}_+, \mathbf{R}^m)$. For $n \geq 1$, the space $\mathcal{H}_n$ consists of all random variables of the form*

$$\tilde I_n(\tilde h) = \sum_{j_1\cdots j_n} \int_0^\infty \int_0^{s_n} \cdots \int_0^{s_2} \tilde h_{j_1,\ldots,j_n}(s_1, \ldots, s_n)\, dW_{j_1}(s_1) \cdots dW_{j_n}(s_n) ,$$

with $\tilde h \in L^2(\Delta_n, \mathbf{R}^{m^n})$.

Proof. We identify $L^2(\Delta_n, \mathbf{R}^{m^n})$ with a subspace of $H^{\otimes n}$ and define the symmetrisation $\Pi\colon H^{\otimes n} \to H^{\otimes n}$ as before. The map $\sqrt{n!}\Pi$ is then an isometry between $L^2(\Delta_n, \mathbf{R}^{m^n})$ and $H^{\otimes_s n}$. Setting $h = \sqrt{n!}\Pi \tilde h$, we claim that $\tilde I_n(\tilde h) = I_n(h)$, from which the lemma then follows.

Since linear combinations of such elements are dense in $L^2(\Delta_n)$, it suffices to hake $\tilde h$ of the form

$$\tilde h_{j_1,\ldots,j_n}(s) = g_{j_1}^{(1)}(s_1)\cdots g_{j_n}^{(n)}(s_n)\;,$$

where the functions $g^{(i)} \in H$ satisfy $\|g^{(i)}\| = 1$ and have the property that $\sup\operatorname{supp} g^{(i)} < \inf\operatorname{supp} g^{(i)}$ for $i < j$. It then follows from the properties of the supports and standard properties of Itô integration that

$$\tilde I_n(\tilde h) = \prod_{i=1}^n W(g^{(i)})\;.$$

Since the functions $g^{(i)}$ have disjoint supports and are therefore all orthogonal in H, we can view them as the first n elements of an orthonormal basis of H. The claim now follows immediately from the definition (2.8) of I_n. □

3 The Malliavin derivative and its adjoint

One of the goals of Malliavin calculus is to make precise the notion of "differentiation with respect to white noise". Let us formally write $\xi_i(t) = \frac{dW_i}{dt}$, which actually makes sense as a random distribution. Then, any random variable X measurable with respect to the filtration generated by the $W(h)$'s can be viewed as a function of the ξ_i's.

At the intuitive level, one would like to introduce operators $\mathscr{D}_t^{(i)}$ which take the derivative of a random variable with respect to $\xi_i(t)$. What would natural properties of such operators be? On the one hand, one would certainly like to have

$$\mathscr{D}_t^{(i)} W(h) = h_i(t)\;, \tag{3.1}$$

since, at least formally, one has

$$W(h) = \sum_{i=1}^m \int_0^\infty h_i(t)\xi_i(t)\,dt\;.$$

On the other hand, one would like these operators to satisfy the chain rule, since otherwise they could hardly claim to be "derivatives":

$$\mathscr{D}_t^{(i)} F(X_1,\ldots,X_n) = \sum_{k=1}^n \partial_k F(X_1,\ldots,X_n)\,\mathscr{D}_t^{(i)} X_k\;. \tag{3.2}$$

Finally, when viewed as a function of t (and of the index i), the right hand side of (3.1) belongs to H, and this property is preserved by the chain rule. It is therefore natural to ask for an operator $\mathscr{D}$ that takes as an argument a sufficiently "nice" random variable and returns an H-valued random variable, such that $\mathscr{D}W(h) = h$ and such that (3.2) holds.

Let now $\mathcal{W} \subset L^2(\Omega, \mathbf{P})$ denote the set of all random variables X such that there exists $N \geq 0$, a function $F \colon \mathbf{R}^N \to \mathbf{R}$ which, together with its derivatives, grows at most polynomially at infinity, and elements $h_i \in H$ such that

$$X = F(W(h_1), \ldots, W(h_N)) \; . \tag{3.3}$$

Given such a random variable, we define an H-valued random variable $\mathscr{D}X$ by

$$\mathscr{D}X = \sum_{k=1}^{N} \partial_k F(W(h_1), \ldots, W(h_N))\, h_k \; . \tag{3.4}$$

One can show that $\mathscr{D}X$ is well-defined, i.e. does not depend on the choice of representation (3.3). Indeed, for $h \in H$, one can characterise $\langle h, \mathscr{D}X\rangle$ as the limit in probability, as $\varepsilon \to 0$, of $\varepsilon^{-1}(\tau_{\varepsilon h}X - X)$, where the translation operator τ is given by

$$\big(\tau_h X\big)(W) = X\Big(W + \int_0^\cdot h(s)\, ds\Big) \; .$$

This in turn does not depend on the representative of X in L^2 since $\tau_{\varepsilon h}^* \mathbf{P}$ is equivalent to $\mathbf{P}$ for every $h \in H$ as a consequence of the Cameron-Martin theorem, see for example [Bog98]. Since $\mathcal{W} \cap \mathcal{H}_n$ is dense in $\mathcal{H}_n$ for every n, we conclude that $\mathcal{W}$ is dense in $L^2(\Omega, \mathbf{P})$, so that $\mathscr{D}$ is a densely defined unbounded linear operator on $L^2(\Omega, \mathbf{P})$.

One very important tool in Malliavin calculus is the following integration by parts formula.

Proposition 3.1 *For every $X \in \mathcal{W}$ and $h \in H$, one has the identity*

$$\mathbf{E}\langle \mathscr{D}X, h\rangle = \mathbf{E}\big(XW(h)\big) \; .$$

Proof. By Grahm-Schmidt, we can assume that X is of the form (3.3) with the h_i orthonormal. One then has

$$\begin{aligned}
\mathbf{E}\langle \mathscr{D}X, h\rangle &= \sum_{k=1}^{N} \mathbf{E}\partial_k F(W(h_1), \ldots, W(h_N))\, \langle h_k, h\rangle \\
&= \sum_{k=1}^{N} \frac{\langle h_k, h\rangle}{(2\pi)^{N/2}} \int_{\mathbf{R}^N} e^{-|x|^2/2} \partial_k F(x_1, \ldots x_k)\, dx \\
&= \sum_{k=1}^{N} \frac{\langle h_k, h\rangle}{(2\pi)^{N/2}} \int_{\mathbf{R}^N} e^{-|x|^2/2} F(x_1, \ldots x_k) x_k\, dx
\end{aligned}$$

$$= \sum_{k=1}^{N} \mathbf{E}\big(XW(h_k)\langle h_k, h\rangle\big) = \mathbf{E}\big(XW(h)\big) \ .$$

To obtain the last identity, we used the fact that $h = \sum_{k=1}^{\infty} \langle h_k, h\rangle h_k$ for an orthonormal basis $\{h_k\}$, together with the fact that $W(h_k)$ is of mean 0 and independent of X for every $k > N$. □

Corollary 3.2 *For every $X, Y \in \mathcal{W}$ and every $h \in H$, one has*

$$\mathbf{E}\big(Y\langle \mathscr{D}X, h\rangle\big) = \mathbf{E}\big(XYW(h) - X\langle \mathscr{D}Y, h\rangle\big) \ . \tag{3.5}$$

Proof. Note that $XY \in \mathcal{W}$ and that Leibniz's rule holds. □

As a consequence of the integration by parts formula (3.5), we can show that the operator $\mathscr{D}$ is closable, which guarantees that it is "well-behaved" from a functional-analytic point of view.

Proposition 3.3 *The operator $\mathscr{D}$ is closable. In other words if, for some sequence $X_n \in \mathcal{W}$, one has $X_n \to 0$ in $L^2(\Omega, \mathbf{P})$ and $\mathscr{D}X_n \to Y$ in $L^2(\Omega, \mathbf{P}, H)$, then $Y = 0$.*

Proof. Let X_n be as in the statement of the proposition and let $Z \in \mathcal{W}$, so that in particular both Z and $\mathscr{D}Z$ have moments of all orders. It then immediately follows from (3.5) that on has

$$\mathbf{E}\big(Z\langle Y, h\rangle\big) = \lim_{n\to\infty} \mathbf{E}\big(X_n Z W(h) - X_n\langle \mathscr{D}Z, h\rangle\big) = 0 \ .$$

If Y was non-vanishing, we could find h such that the real-valued random variable $\langle Y, h\rangle$ is not identically 0. Since $\mathcal{W}$ is dense in L^2, this would entail the existence of some $Z \in \mathcal{W}$ such that $\mathbf{E}\big(Z\langle Y, h\rangle\big) \neq 0$, yielding a contradiction. □

We henceforth denote by $\mathcal{W}^{1,2}$ the domain of the closure of $\mathscr{D}$ (namely those random variables X such that there exists $X_n \in \mathcal{W}$ with $X_n \to X$ in L^2 and such that $\mathscr{D}X_n$ converges to some limit $\mathscr{D}X$) and we do not distinguish between $\mathscr{D}$ and its closure. We also follow [Nua06] in denoting the adjoint of $\mathscr{D}$ by δ. One can of course apply the Malliavin differentiation operator repeatedly, thus yielding an unbounded closed operator $\mathscr{D}^k$ from $L^2(\Omega, \mathbf{P})$ to $L^2(\Omega, \mathbf{P}, H^{\otimes k})$. We denote the domain of this operator by $\mathcal{W}^{k,2}$.

Actually, a similar proof shows that powers of $\mathscr{D}$ are closable as unbounded operators from $L^p(\Omega, \mathbf{P})$ to $L^p(\Omega, \mathbf{P}, H^{\otimes k})$ for every $p \geq 1$. We denote the domain of these operators by $\mathcal{W}^{k,p}$. Furthermore, for any Hilbert space K, we denote by $\mathcal{W}^{k,p}(K)$ the domain of $\mathscr{D}^k$ viewed as an operator from $L^p(\Omega, \mathbf{P}, K)$ to $L^p(\Omega, \mathbf{P}, H^{\otimes k} \otimes K)$. We call a random variable belonging to $\mathcal{W}^{k,p}$ for every $k, p \geq 1$ "Malliavin smooth" and we write $\mathcal{S} = \bigcap_{k,p} \mathcal{W}^{k,p}$ as well as $\mathcal{S}(K) = \bigcap_{k,p} \mathcal{W}^{k,p}(K)$. The Malliavin smooth random variables play a role analogous to that of Schwartz test functions in finite-dimensional analysis.

Remark 3.4 As an immediate consequence of Hölder's inequality and the Leibniz rule, $\mathcal{S}$ is an algebra.

Let us now try to get some feeling for the domain of δ. Recall that, by definition of the adjoint, the domain of δ is given by those elements $u \in L^2(\Omega, \mathbf{P}, H)$ such that there exists $Y \in L^2(\Omega, \mathbf{P})$ satisfying the identity

$$\mathbf{E}\langle u, \mathscr{D}X\rangle = \mathbf{E}(YX) ,$$

for every $X \in \mathscr{W}$. One then writes $Y = \delta u$. Interestingly, it turns out that the operator δ is an extension of Itô integration! It is therefore also called the Skorokhod integration operator and, instead of just writing δu, one often writes instead

$$\int_0^\infty u(t)\, \delta W(t) .$$

We now proceed to showing that it is indeed the case that, if u is a square integrable stochastic process that is adapted to the filtration generated by the increments of the underlying Wiener process W, then u belongs to the domain of δ and δu coincides with the usual Itô integral of u against W.

To formulate this more precisely, denote by $\mathscr{F}_t$ the σ-algebra generated by the random variables $W(h)$ with $\operatorname{supp} h \subset [0, t]$. Consider then the set of elementary adapted processes, which consist of all processes of the form

$$u = \sum_{k=1}^N Y_k^{(i)} \mathbf{1}^{(i)}_{[s_k,t_k)} , \tag{3.6}$$

for some $N \geq 1$, some times s_k, t_k with $0 \leq s_k < t_k < \infty$, and some random variables $Y_k^{(i)} \in L^2(\Omega, \mathscr{F}_{s_k}, \mathbf{P})$. Summation over i is also implied. We denote by $L^2_a(\Omega, \mathbf{P}, H) \subset L^2(\Omega, \mathbf{P}, H)$ the closure of this set. Recall then that, for an elementary adapted process of the type (3.6), its Itô integral is given by

$$\int_0^\infty u(t)\, dW(t) \stackrel{\text{def}}{=} \sum_{k=1}^N Y_k^{(i)} \big(W_i(t_k) - W_i(s_k)\big) = \sum_{k=1}^N Y_k^{(i)} W\big(\mathbf{1}^{(i)}_{[s_k,t_k)}\big) . \tag{3.7}$$

Using Itô's isometry, this can then be extended to all of L^2_a.

Theorem 3.5 *The space $L^2_a(\Omega, \mathbf{P}, H)$ is included in the domain of δ and, on it, δ coincides with the Itô integration operator.*

Proof. Let u be an elementary adapted process of the form (3.6) with each $Y_k^{(i)}$ in $\mathscr{W}$. For $X \in \mathscr{W}$ one then has, as a consequence of (3.5),

$$\mathbf{E}\big(\langle u, \mathscr{D}X\rangle\big) = \sum_{k=1}^N \mathbf{E}\big(Y_k^{(i)} \langle \mathbf{1}^{(i)}_{[s_k,t_k)}, \mathscr{D}X\rangle\big) \tag{3.8}$$

$$= \sum_{k=1}^N \mathbf{E}\Big(Y_k^{(i)} X\, W(\mathbf{1}^{(i)}_{[s_k,t_k)}) - X\langle \mathscr{D} Y_k^{(i)}, \mathbf{1}^{(i)}_{[s_k,t_k)}\rangle\Big)\;.$$

At this stage, we note that since $\mathscr{D} Y_k^{(i)}$ is $\mathscr{F}_{s_k}$-measurable by assumption, it has a representation of the type (3.3) with each h_j satisfying $\operatorname{supp} h_j \in [0, s_k]$. In particular, one has $\langle h_j, \mathbf{1}^{(i)}_{[s_k,t_k)}\rangle = 0$ so that, by (3.4), one has $\langle \mathscr{D} Y_k^{(i)}, \mathbf{1}^{(i)}_{[s_k,t_k)}\rangle = 0$. Combining this with the above identity and (3.7), we conclude that

$$\mathbf{E}\big(\langle u, \mathscr{D} X\rangle\big) = \mathbf{E}\Big(X \int_0^\infty u(t)\, dW(t)\Big)\;.$$

Taking limits on both sides of this identity, we conclude that it holds for every $u \in L^2_a$, thus completing the proof. □

One also has the following extension of Itô's isometry.

Theorem 3.6 *The space $\mathscr{W}^{1,2}(H)$ is included in the domain of δ and, on it, the identity*

$$\mathbf{E}|\delta u|^2 = \mathbf{E}\int_0^\infty |u(t)|^2\, dt + \mathbf{E}\int_0^\infty \int_0^\infty \mathscr{D}^{(i)}_s u_j(t)\, \mathscr{D}^{(j)}_t u_i(s)\, ds\, dt$$

holds, with summation over repeated indices implied.

Proof. Consider similarly to before u to be a process of the form $u = \sum_{i=1}^N Y^{(i)} h^{(i)}$ with $Y^{(i)} \in \mathscr{W}$ and $h^{(i)} \in H$, but this time without any adaptedness condition on the Y's. It then follows from the same calculation as (3.8) that

$$\delta u = Y^{(i)} W(h^{(i)}) - \langle \mathscr{D} Y^{(i)}, h^{(i)}\rangle\;, \tag{3.9}$$

with summation over i implied, so that

$$\mathscr{D}_h \delta u = \mathscr{D}_h Y^{(i)}\, W(h^{(i)}) + Y^{(i)}\langle h, h^{(i)}\rangle - \langle \mathscr{D}^2 Y^{(i)}, h\otimes h^{(i)}\rangle = \delta \mathscr{D}_h u + \langle h, u\rangle\;. \tag{3.10}$$

(This is nothing but an instance of the "canonical commutation relations" appearing in quantum mechanics!) Integrating by parts, applying (3.10), and then integrating by parts again it follows that

$$\begin{aligned}\mathbf{E}|\delta u|^2 &= \mathbf{E}\langle u, \mathscr{D}\delta u\rangle = \mathbf{E}\langle u, u\rangle + \mathbf{E} Y^{(i)} \delta(\mathscr{D}_{h^{(i)}} Y^{(j)} h^{(j)})\\ &= \mathbf{E}\langle u, u\rangle + \mathbf{E}\mathscr{D}_{h^{(j)}} Y^{(i)} \mathscr{D}_{h^{(i)}} Y^{(j)}\;,\end{aligned}$$

with summation over i and j implied. At this point, we note that

$$\mathscr{D}_{h^{(j)}} Y^{(i)}\, \mathscr{D}_{h^{(i)}} Y^{(j)} = \int_0^\infty \int_0^\infty \big(\mathscr{D}^{(k)}_s Y^{(i)}\big)\, h^{(j)}_k(s)\, \big(\mathscr{D}^{(\ell)}_t Y^{(j)}\big)\, h^{(i)}_\ell(t)\, ds\, dt$$

$$
\begin{aligned}
&= \int_0^\infty \int_0^\infty \big(\mathscr{D}_s^{(k)} Y^{(i)}\, h_\ell^{(i)}(t)\big)\, \big(\mathscr{D}_t^{(\ell)} Y^{(j)} h_k^{(j)}(s)\big)\; ds\, dt \\
&= \int_0^\infty \int_0^\infty \mathscr{D}_s^{(k)} u_\ell(t)\, \mathscr{D}_t^{(\ell)} u_k(s)\, ds\, dt\ .
\end{aligned}
$$

It remains to use the density of the class of processes we considered to conclude the proof. □

In a similar way, we can give a very nice characterisation of the "Ornstein-Uhlenbeck operator" $\delta\mathscr{D}$:

Proposition 3.7 *The spaces $\mathscr{H}_n$ are invariant for $\Delta = \delta\mathscr{D}$ and one has $\Delta X = nX$ for every $X \in \mathscr{H}_n$.*

Proof. Fix an orthonormal basis $\{e_k\}$ of H. Then, by definition, the random variables Φ_k as in (2.6) with $|k| = n$ are dense in $\mathscr{H}_n$. Recalling that $\mathscr{D}H_k(W(h)) = kH_{k-1}(W(h))h$, one has

$$
\mathscr{D}\Phi_k = \sum_i k_i \Phi_{k-\delta_i} e_i\ ,
$$

where δ_i is given by $\delta_i(j) = \delta_{i,j}$. We now recall that, as in (3.8), one has the identity

$$
\delta(Xh) = X\,W(h) - \langle \mathscr{D}X, h\rangle\ ,
$$

for every $X \in \mathscr{W}$ and $h \in H$, so that

$$
\begin{aligned}
\Delta\Phi_k &= \sum_i k_i \Phi_{k-\delta_i} W(e_i) - \sum_{i,j} k_i (k_j - \delta_{i,j}) \Phi_{k-\delta_i-\delta_j} \langle e_i, e_j\rangle \\
&= \sum_i k_i \big(\Phi_{k-\delta_i} W(e_i) - (k_i - 1)\Phi_{k-2\delta_i}\big)\ .
\end{aligned}
$$

Recall now that, by (2.5), one has

$$
H_{k_i-1}(x)x - (k_i - 1)H_{k_i-2}(x) = H_{k_i}(x)\ ,
$$

so that one does indeed obtain $\Delta\Phi_k = \sum_i k_i \Phi_k = n\Phi_k$ as claimed. □

An important remark to keep in mind is that while δ is an extension of Itô integration it is not the only such extension, and not even the only "reasonable" one. Actually, one may argue that it is not a "reasonable" extension of Itô's integral at all since, for a generic random variable X, one has in general

$$
\int_0^\infty X\,u(t)\,\delta W(t) \neq X \int_0^\infty u(t)\,\delta W(t)\ .
$$

Also, if one considers a one-parameter family of stochastic processes $a \mapsto u(a, \cdot)$ and sets $G(a) = \int_0^\infty u(a,t)\, \delta W(t)$, then in general one has

$$G(X) \neq \int_0^\infty u(X,t)\, \delta W(t) ,$$

if X is a random variable.

It will be useful in the sequel to be able to have a formula for the Malliavin derivative of a random variable that is already given as a stochastic integral. Consider a random variable X of the type

$$X = \int_0^\infty u(t)\, dW(t) , \tag{3.11}$$

with $u \in L^2_a$ is sufficiently "nice". At a formal level, one would then expect to have the identity

$$\mathscr{D}_s^{(i)} X = u_i(s) + \int_0^\infty \mathscr{D}_s^{(i)} u(t)\, dW(t) . \tag{3.12}$$

This is indeed the case, as the following proposition shows.

Proposition 3.8 *Let $u \in L^2_a(\Omega, \mathbf{P}, H)$ be such that $u_i(t) \in \mathscr{W}^{1,2}$ for almost every t and $\int_0^\infty \mathbf{E}\|\mathscr{D}u_i(t)\|^2\, dt < \infty$. Then (3.12) holds.*

Proof. Take u of the form (3.6) with each $Y_k^{(i)}$ in $\mathscr{W}$, so that $X = \sum Y_k^{(i)} W\big(\mathbf{1}^{(i)}_{[s_k,t_k)}\big)$. It then follows from the chain rule that

$$\mathscr{D}X = \sum_{k=1}^N \big(Y_k^{(i)} \mathbf{1}^{(i)}_{[s_k,t_k)} + \mathscr{D}Y_k^{(i)} W\big(\mathbf{1}^{(i)}_{[s_k,t_k)}\big)\big) = u + \int_0^\infty \mathscr{D}u(t)\, dW(t) ,$$

and the claim follows from a simple approximation argument, combined with the fact that $\mathscr{D}$ is closed. □

Finally, we will use the important fact that the divergence operator δ maps $\mathcal{S}$ into $\mathcal{S}$. This is a consequence of the following result.

Proposition 3.9 *For every $p \geq 2$ there exist constants k and C such that, for every separable Hilbert space K and every $u \in \mathcal{S}(H \otimes K)$, one has the bound*

$$\mathbf{E}|\delta u|^p \leq C \sum_{0 \leq \ell \leq k} \big(\mathbf{E}|\mathscr{D}^\ell u|^{2p}\big)^{1/2} .$$

Proof. For $p \in [1,2]$, the bound follows immediately from Theorem 3.6 and Jensen's inequality. Take now $p > 2$. Using the definition of δ combined with the chain rule for $\mathscr{D}$, Proposition 3.8, and Young's inequality, we obtain the bound

$$\begin{aligned}\mathbf{E}|\delta u|^p &= (p-1)\mathbf{E}\big(|\delta u|^{p-2}\langle u, \mathscr{D}\delta u\rangle\big) = (p-1)\mathbf{E}|\delta u|^{p-2}\big(|u|^2 + \langle u, \delta\mathscr{D}u\rangle\big)\\ &\le \frac{1}{2}\mathbf{E}|\delta u|^p + c\mathbf{E}\big(|u|^p + |u|^{p/2}|\delta\mathscr{D}u|^{p/2}\big)\;,\end{aligned}$$

for some constant c. We now use Hölder's inequality which yields

$$\mathbf{E}\big(|u|^{p/2}|\delta\mathscr{D}u|^{p/2}\big) \le \big(\mathbf{E}|u|^{2p}\big)^{1/4}\big(\mathbf{E}|\delta\mathscr{D}u|^{2p/3}\big)^{3/4}\;.$$

Combining this with the above, we conclude that there exists a constant C such that

$$\mathbf{E}|\delta u|^p \le C\big(\mathbf{E}|\mathscr{D}^\ell u|^{2p}\big)^{1/2} + \big(\mathbf{E}|\delta\mathscr{D}u|^{2p/3}\big)^{3/2}\;.$$

The proof is concluded by a simple inductive argument. □

Corollary 3.10 *The operator δ maps $\mathcal{S}(H \otimes K)$ into $\mathcal{S}(K)$.*

Proof. In order to estimate $\mathbf{E}|\mathscr{D}^k\delta u|^p$, it suffices to first apply Proposition 3.8 k times and then Proposition 3.9. □

Remark 3.11 The above argument is very far from being sharp. Actually, it is possible to show that δ maps $\mathscr{W}^{k,p}$ into $\mathscr{W}^{k-1,p}$ for every $p \ge 2$ and every $k \ge 1$. This however requires a much more delicate argument.

4 Smooth densities

In this section, we give sufficient conditions for the law of a random variable X to have a smooth density with respect to Lebesgue measure. The main ingredient for this is the following simple lemma.

Lemma 4.1 *Let X be an $\mathbf{R}^n$-valued random variable for which there exist constants C_k such that $|\mathbf{E}D^{(k)}G(X)| \le C_k\|G\|_\infty$ for every $G \in \mathscr{C}_0^\infty$ and $k \ge 1$. Then the law of X has a smooth density with respect to Lebesgue measure.*

Proof. Denoting by μ the law of X, our assumption can be rewritten as

$$\Big|\int_{\mathbf{R}^n} D^{(k)}G(x)\,\mu(dx)\Big| \le C_k\|G\|_\infty\;. \tag{4.1}$$

Let now $s > n/2$ so that $\|G\|_\infty \lesssim \|G\|_{H^s}$ by Sobolev embedding. By duality and the density of $\mathscr{C}_0^\infty$ in H^s, the assumption then implies that every distributional derivative of μ belongs to the Sobolev space H^{-s} so that, as a distribution, μ belongs to H^ℓ for every $\ell \in \mathbf{R}$. The result then follows from the fact that $H^\ell \subset \mathscr{C}^k$ as soon as $\ell > k + \frac{n}{2}$. □

Remark 4.2 If the bound (4.1) only holds for $k = 1$, then it is still the case that the law of X has a density with respect to Lebesgue measure.

The idea now is to make repeated use of the integration by parts formula (3.5) in order to control the expectation of $D^{(k)}G(X)$. Consider first the case $k = 1$ and write $\langle u, DG\rangle$ for the directional derivative of G in the direction $u \in \mathbf{R}^n$. Ideally, for every $i \in \{1, \ldots, n\}$, we would like to find an H-valued random variable Y_i *independent of* G such that

$$\partial_i G(X) = \langle \mathscr{D}G(X), Y_i\rangle \;, \tag{4.2}$$

where the second scalar product is taken in H, so that one has

$$\mathbf{E}\partial_i G(X) = \mathbf{E}\big(G(X)\,\delta Y_i\big) \;.$$

If Y_i can be chosen in such a way that $\mathbf{E}|\delta Y_i| < \infty$ for every i, then the bound (4.1) for $k = 1$ follows. Since $\mathscr{D}G(X) = \sum_j \partial_j G(X)\,\mathscr{D}X_j$ as a consequence of the chain rule, a random variable Y_i as in (4.2) can be found only if $\mathscr{D}X$, viewed as a random linear map from H to $\mathbf{R}^n$, is almost surely surjective. This suggests that an important condition will be that of the invertibility of the *Malliavin matrix* $\mathscr{M}$ defined by

$$\mathscr{M}_{ij} = \langle \mathscr{D}X_i, \mathscr{D}X_j\rangle \;, \tag{4.3}$$

where the scalar product is taken in H. Assuming that $\mathscr{M}$ is invertible, the solution with minimal H-norm to the overdetermined system $\delta_i = \langle \mathscr{D}X, Y_i\rangle$ (where δ_i denotes the ith canonical basis vector in $\mathbf{R}^n$) is given by

$$Y_i = (\mathscr{D}X)^*\,\mathscr{M}^{-1}\delta_i \;.$$

Assuming a sufficient amount of regularity, this shows that a bound of the type appearing in the assumption of Lemma 4.1 holds for a random variable X whose Malliavin matrix $\mathscr{M}$ is invertible and whose inverse has a finite moment of sufficiently high order. The following theorem should therefore not come as a surprise.

Theorem 4.3 *Let X be a Malliavin smooth $\mathbf{R}^n$-valued random variable such that the Malliavin matrix defined in (4.3) is almost surely invertible and has inverse moments of all orders. Then the law of X has a smooth density with respect to Lebesgue measure.*

The main ingredient of the proof of this theorem is the following lemma.

Lemma 4.4 *Let X be as above and let $Z \in \mathcal{S}$. Then, there exists $\bar Z \in \mathcal{S}$ such that the identity*

$$\mathbf{E}\big(Z\partial_i G(X)\big) = \mathbf{E}\big(G(X)\bar Z\big) \;, \tag{4.4}$$

holds for every $G \in \mathscr{C}_0^\infty$.

Proof. Following the calculation above, defining the H-valued random variable Y_i by

$$Y_i = \sum_{j=1}^{n} (\mathscr{D} X_j)\, \mathscr{M}_{ji}^{-1} \;,$$

we have the identity $\partial_i G(X) = \langle \mathscr{D} G(X), Y_i \rangle$. As a consequence, (4.4) holds with

$$\bar{Z} = \delta\big(Z\, Y_i\big) \;.$$

The claim now follows from Remark 3.4 and Proposition 3.9, as soon as we can show that $\mathscr{M}_{ji}^{-1} \in \mathcal{S}$. This however follows from the chain rule for $\mathscr{D}$ and Remark 3.4, since the former shows that $\mathscr{D}^k \mathscr{M}_{ji}^{-1}$ can be written as a polynomial in $\mathscr{M}_{ji}^{-1}$ and $\mathscr{D}^\ell X$ for $\ell \leq k$. □

Proof of Theorem 4.3. By Lemma 4.1 it suffices to show that under the assumptions of the theorem, for every multiindex k and every random variable $Y \in \mathcal{S}$, there exists a random variable $Z \in \mathcal{S}$ such that

$$\mathbf{E}\big(Y\, D^k G(X)\big) = \mathbf{E}\big(G(X) Z\big) \;. \tag{4.5}$$

We proceed by induction, the claim being trivial for $k = 0$. Assuming that (4.5) holds for some k, we then obtain as a consequence of Lemma 4.4 that

$$\mathbf{E}\big(Y D^k \partial_i G(X)\big) = \mathbf{E}\big(\partial_i G(X) Z\big) = \mathbf{E}\big(G(X) \bar{Z}\big) \;,$$

for some $\bar{Z} \in \mathcal{S}$, which is precisely the required bound (4.5), but for $k + e_i$. □

5 Malliavin Calculus for Diffusion Processes

We are now in possession of all the abstract tools required to tackle the proof of Hörmander's theorem. Before we start however, we discuss how $\mathscr{D}_s X_t$ can be computed when X_t is the solution to an SDE of the type (1.1). Recall first that, by definition, (1.1) is equivalent to the Itô stochastic differential equation

$$dX_t = \tilde{V}_0(X_t)\, dt + \sum_{i=1}^{m} V_i(X_t)\, dW_i(t) \;, \tag{5.1}$$

with $\tilde{V}_0$ given in coordinates by

$$(\tilde{V}_0)_i(x) = (V_0)_i(x) + \frac{1}{2} (\partial_k V_j)_i(x) (V_j)_k(x) \;,$$

with summation over repeated indices implied. We assume that $V_j \in \mathscr{C}_b^\infty$, the space of smooth vector fields that are bounded, together with all of their derivatives. It immediately follows that, for every initial condition $x_0 \in \mathbf{R}^n$, (5.1) can be solved by simple Picard iteration, just like ordinary differential equations, but in the space of adapted square integrable processes.

An important tool for our analysis will be the linearisation of (1.1) with respect to its initial condition. This is obtained by simply formally differentiating both sides of (1.1) with respect to the initial conditions x_0. For any $s \geq 0$, this yields the non-autonomous linear equation

$$dJ_{s,t} = D\tilde{V}_0(X_t)\, J_{s,t}\, dt + \sum_{i=1}^m DV_i(X_t)\, J_{s,t}\, dW_i(t)\;, \qquad J_{s,s} = \mathrm{id}\;, \tag{5.2}$$

where id denotes the $n \times n$ identity matrix. This in turn is equivalent to the Stratonovich equation

$$dJ_{s,t} = DV_0(X_t)\, J_{s,t}\, dt + \sum_{i=1}^m DV_i(X_t)\, J_{s,t} \circ dW_i(t)\;, \qquad J_{s,s} = \mathrm{id}\;. \tag{5.3}$$

Higher order derivatives $J_{0,t}^{(k)}$ with respect to the initial condition can be defined similarly. It is straightforward to verify that this equation admits a unique solution, and that this solution satisfies the identity

$$J_{t,u}J_{s,t} = J_{s,u}\;, \tag{5.4}$$

for any three times $s \leq t \leq u$. Under our standing assumptions for SDEs, the Jacobian has moments of all orders:

Proposition 5.1 *If $V_i \in \mathscr{C}_b^\infty$ for all i, then $\sup_{s,t\leq T} \mathbf{E}|J_{s,t}|^p < \infty$ for every $T > 0$ and every $p \geq 1$.*

Proof. We write $|A|$ for the Frobenius norm of a matrix A. A tedious application of Itô's formula shows that for even integers $p \geq 4$ one has

$$\begin{aligned} d|J_{s,t}|^p = p|J_{s,t}|^{p-2}\Big(\langle J_{s,t}, D\tilde{V}_0(X_t)\, J_{s,t}\rangle\, dt + \sum_{i=1}^m \langle J_{s,t}, DV_i(X_t)\, J_{s,t}\rangle\, dW_i(t)\Big) \\ + \frac{p}{2}|J_{s,t}|^{p-4} \sum_{i=1}^m \Big((p-2)\langle DV_i(X_t)J_{s,t}, J_{s,t}\rangle^2 + |J_{s,t}|^2 (\mathrm{tr}\, DV_i(X_t)J_{s,t})^2\Big)\, dt\;. \end{aligned}$$

Writing this in integral form, taking expectations on both sides and using the boundedness of the derivatives of the vector fields, we conclude that there exists a constant C such that

$$\mathbf{E}|J_{s,t}|^p \leq n^{p/2} + C\int_s^t \mathbf{E}|J_{s,r}|^p\, dr\;,$$

so that the claim now follows from Gronwall's lemma. (The $n^{p/2}$ comes from the initial condition, which equals $|\mathrm{id}|^p = n^{p/2}$.) □

As a consequence of (5.4), one has $J_{s,t} = J_{0,t}J_{0,s}^{-1}$. One can also verify that the inverse $J_{0,t}^{-1}$ of the Jacobian solves the SDE

$$dJ_{0,t}^{-1} = -J_{0,t}^{-1}\, DV_0(X_t)\, dt - \sum_{i=1}^m J_{0,t}^{-1}\, DV_i(X_t) \circ dW_i \;. \tag{5.5}$$

This follows from the chain rule by noting that if we denote by $\Psi(A) = A^{-1}$ the map that takes the inverse of a square matrix, then we have $D\Psi(A)H = -A^{-1}HA^{-1}$.

On the other hand, we can use (3.12) to, at least on a formal level, take the Malliavin derivative of the integral form of (5.1), which then yields for $r \le t$ the identity

$$\mathscr{D}_r^{(j)} X(t) = \int_r^t D\tilde V_0(X_s)\, \mathscr{D}_r^{(j)} X_s\, ds + \sum_{i=1}^m \int_r^t DV_i(X_s)\, \mathscr{D}_r^{(j)} X_s\, dW_i(s) + V_j(X_r) \;.$$

We see that, save for the initial condition at time $t = r$ given by $V_j(X_r)$, this equation is identical to the integral form of (5.2)! Using the variation of constants formula, it follows that for $s < t$ one has the identity

$$\mathscr{D}_s^{(j)} X_t = J_{s,t} V_j(X_s) \;. \tag{5.6}$$

Furthermore, since X_t is independent of the later increments of W, we have $\mathscr{D}_s^{(j)} X_t = 0$ for $s \ge t$. This formal manipulation can easily be justified a posteriori, thus showing that the random variables X_t belongs to $\mathscr{W}^{1,p}$ for every p. In fact, one has even more than that:

Proposition 5.2 *If the vector fields V_i belong to $\mathscr{C}_b^\infty$ and $X_0 \in \mathbf{R}^n$ is deterministic, then the solution X_t to (5.1) belongs to $\mathcal{S}$ for every $t \ge 0$.*

Before we prove this, let us recall the following bound on iterated Itô integrals.

Lemma 5.3 *Let $k \ge 1$ and let v be a stochastic process on $\mathbf{R}^k$ with $\mathbf{E}\|v\|_{L^p}^p < \infty$ for some $p \ge 2$. Then, one has the bound*

$$\mathbf{E}\Big|\int_0^t \cdots \int_0^{s_2} v(s_1,\ldots,s_k)\, dW_{i_1}(s_1)\cdots dW_{i_k}(s_k)\Big|^p \le Ct^{\frac{k(p-2)}{2}} \mathbf{E}\|v\|_{L^p}^p \;.$$

Proof. The proof goes by induction over k. For $k = 1$, it follows from the Burkholder-David-Gundy inequality followed by Hölder's inequality that

$$\mathbf{E}\Big|\int_0^t v(s)\, dW_i(s)\Big|^p \le C\mathbf{E}\Big|\int_0^t |v(s)|^2\, ds\Big|^{p/2} \le Ct^{\frac{p-2}{2}} \mathbf{E}\int_0^t |v(s)|^p\, ds \;, \tag{5.7}$$

as claimed. In the general case, we set

$$\tilde v(s_k) = \int_0^{s_k} \cdots \int_0^{s_2} v(s_1,\ldots,s_k)\,dW_{i_1}(s_1)\cdots dW_{i_{k-1}}(s_{k-1})\;.$$

Combining (5.7) with the induction hypothesis, we obtain

$$\begin{aligned}\mathbf{E}\Big|\int_0^t \tilde v(s_k)\,dW_{i_k}(s_k)\Big|^p &\le Ct^{\frac{p-2}{2}}\int_0^t \mathbf{E}|\tilde v(s_k)|^p\,ds_k\\ &\le Ct^{\frac{p-2}{2}}t^{\frac{(k-1)(p-2)}{2}}\int_0^t \mathbf{E}\Big(\int_0^{s_k}\cdots\int_0^{s_1}|v(s_1,\ldots,s_k)|^p\,ds_1\cdots ds_{k-1}\Big)\,ds_k\;,\end{aligned}$$

and the claim follows from Fubini's theorem. □

Proof of Proposition 5.2. We first derive an equation for the Malliavin derivative of the Jacobian $J_{s,t}$. Differentiating the integral form of (5.2) we obtain for $r < s$ the identity

$$\begin{aligned}\mathscr{D}_r^{(j)} J_{s,t} &= \int_s^t D\tilde V_0(X_u)\,\mathscr{D}_r^{(j)} J_{s,u}\,du + \int_s^t D^2\tilde V_0(X_u)\,\big(J_{s,u},\mathscr{D}_r^{(j)}X_u\big)\,du\\ &\quad+\sum_{i=1}^m \int_s^t DV_i(X_u)\,\mathscr{D}_r^{(j)}J_{s,u}\,dW_i(u) + \int_s^t D^2V_i(X_u)\,\big(J_{s,u},\mathscr{D}_r^{(j)}X_u\big)\,dW_i(u)\;.\end{aligned}$$

Using (5.6), we can rewrite this as

$$\begin{aligned}\mathscr{D}_r^{(j)} J_{s,t} &= \int_s^t D\tilde V_0(X_u)\,\mathscr{D}_r^{(j)} J_{s,u}\,du + \int_s^t D^2\tilde V_0(X_u)\,\big(J_{s,u},J_{r,u}V_j(X_r)\big)\,du\\ &\quad+\sum_{i=1}^m \int_s^t DV_i(X_u)\,\mathscr{D}_r^{(j)}J_{s,u}\,dW_i(u)\\ &\quad+\sum_{i=1}^m \int_s^t D^2V_i(X_u)\,\big(J_{s,u},J_{r,u}V_j(X_r)\big)\,dW_i(u)\;.\end{aligned}$$

Once again, we see that this is nothing but an inhomogeneous version of the equation for $J_{s,t}$ itself. The variation of constants formula thus yields

$$\begin{aligned}\mathscr{D}_r^{(j)} J_{s,t} &= \int_s^t J_{u,t}D^2\tilde V_0(X_u)\,\big(J_{s,u},J_{r,u}V_j(X_r)\big)\,du\\ &\quad+\sum_{i=1}^m \int_s^t J_{u,t}D^2V_i(X_u)\,\big(J_{s,u},J_{r,u}V_j(X_r)\big)\,dW_i(u)\;.\end{aligned}$$

This allows to show by induction that, for any integer k, the iterated Malliavin derivative $\mathscr{D}_{r_1}^{(j_1)}\cdots\mathscr{D}_{r_k}^{(j_k)}X(t)$ with $r_1 \le \cdots \le r_k$ can be expressed as a finite sum of terms consisting of a multiple iterated Wiener / Lebesgue integral with integrand given by a finite product of components of the type J_{s_i,s_j} with $i < j$, as well as functions in $\mathcal{C}_b^\infty$ evaluated at X_{s_i}. This has moments of all orders as a consequence of Proposition 5.1, combined with Lemma 5.3. □

Theorem 5.4 *Let $x_0 \in \mathbf{R}^n$ and let X_t be the solution to (1.1). If the vector fields $\{V_j\} \subset \mathcal{C}_b^\infty$ satisfy the parabolic Hörmander condition, then the law of X_t has a smooth density with respect to Lebesgue measure.*

Proof. Denote by $\mathcal{A}_{0,t}$ the operator $\mathcal{A}_{0,t} v = \int_0^t J_{s,t} V(X_s) v(s)\, ds$, where v is a square integrable, not necessarily adapted, $\mathbf{R}^m$-valued stochastic process and V is the $n \times m$ matrix-valued function obtained by concatenating the vector fields V_j for $j = 1, \ldots, m$. With this notation, it follows from (5.6) that the Malliavin covariance matrix $\mathcal{M}_{0,t}$ of X_t is given by

$$\mathcal{M}_{0,t} = \mathcal{A}_{0,t}\mathcal{A}_{0,t}^* = \int_0^t J_{s,t} V(X_s) V^*(X_s) J_{s,t}^*\, ds\ .$$

It follows from (5.6) that the assumptions of Theorem 4.3 are satisfied for the random variable X_t, provided that we can show that $\|\mathcal{M}_{0,t}^{-1}\|$ has bounded moments of all orders. This in turn follows by combining Lemma 6.2 with Theorem 6.3 below. □

6 Hörmander's Theorem

This section is devoted to a proof of the fact that Hörmander's condition is sufficient to guarantee the invertibility of the Malliavin matrix of a diffusion process. For purely technical reasons, it turns out to be advantageous to rewrite the Malliavin matrix as

$$\mathcal{M}_{0,t} = J_{0,t}\mathcal{C}_{0,t}J_{0,t}^*\ , \qquad \mathcal{C}_{0,t} = \int_0^t J_{0,s}^{-1} V(X_s) V^*(X_s) \big(J_{0,s}^{-1}\big)^*\, ds\ ,$$

where $\mathcal{C}_{0,t}$ is the *reduced Malliavin matrix* of our diffusion process.

Remark 6.1 The reason for considering the reduced Malliavin matrix is that the process appearing under the integral in the definition of $\mathcal{C}_{0,t}$ is adapted to the filtration generated by W_t. This allows us to use some tools from stochastic calculus that would not be available otherwise.

Since we assumed that $J_{0,t}$ has inverse moments of all orders, the invertibility of $\mathcal{M}_{0,t}$ is equivalent to that of $\mathcal{C}_{0,t}$. Note first that since $\mathcal{C}_{0,t}$ is a positive definite symmetric matrix, the norm of its inverse is given by

$$\|\mathcal{C}_{0,t}^{-1}\| = \Big(\inf_{|\eta|=1} \langle \eta, \mathcal{C}_{0,t}\eta\rangle\Big)^{-1}\ .$$

A very useful observation is then the following:

Lemma 6.2 *Let M be a symmetric positive semidefinite $n\times n$ matrix-valued random variable such that $\mathbf{E}\|M\|^p < \infty$ for every $p \geq 1$ and such that, for every $p \geq 1$ there exists C_p such that*

$$\sup_{|\eta|=1} \mathbf{P}\big(\langle \eta, M\eta\rangle < \varepsilon\big) \leq C_p \varepsilon^p\;, \tag{6.1}$$

holds for every $\varepsilon \leq 1$. Then, $\mathbf{E}\|M^{-1}\|^p < \infty$ for every $p \geq 1$.

Proof. The non-trivial part of the result is that the supremum over η is taken outside of the probability in (6.1). For $\varepsilon > 0$, let $\{\eta_k\}_{k\leq N}$ be a sequence of vectors with $|\eta_k| = 1$ such that for every η with $|\eta| \leq 1$, there exists k such that $|\eta_k - \eta| \leq \varepsilon^2$. It is clear that one can find such a set with $N \leq C\varepsilon^{2-2n}$ for some $C > 0$ independent of ε. We then have the bound

$$\begin{aligned}\langle \eta, M\eta\rangle &= \langle \eta_k, M\eta_k\rangle + \langle \eta - \eta_k, M\eta\rangle + \langle \eta - \eta_k, M\eta_k\rangle\\ &\geq \langle \eta_k, M\eta_k\rangle - 2\|M\|\varepsilon^2\;,\end{aligned}$$

so that

$$\begin{aligned}\mathbf{P}\Big(\inf_{|\eta|=1}\langle \eta, M\eta\rangle \leq \varepsilon\Big) &\leq \mathbf{P}\Big(\inf_{k\leq N}\langle \eta_k, M\eta_k\rangle \leq 4\varepsilon\Big) + \mathbf{P}\Big(\|M\| \geq \frac{1}{\varepsilon}\Big)\\ &\leq C\varepsilon^{2-2n}\sup_{|\eta|=1}\mathbf{P}\Big(\langle \eta, M\eta\rangle \leq 4\varepsilon\Big) + \mathbf{P}\Big(\|M\| \geq \frac{1}{\varepsilon}\Big)\;.\end{aligned}$$

It now suffices to use (6.1) for p large enough to bound the first term and Chebychev's inequality combined with the moment bound on $\|M\|$ to bound the second term. □

As a consequence of this, Theorem 5.4 is a corollary of:

Theorem 6.3 *Under the assumptions of Theorem 5.4, for every initial condition $x \in \mathbf{R}^n$, we have the bound*

$$\sup_{|\eta|=1} \mathbf{P}\big(\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon\big) \leq C_p \varepsilon^p\;,$$

for suitable constants C_p and all $p \geq 1$.

Remark 6.4 The choice $t = 1$ as the final time is of course completely arbitrary. Here and in the sequel, we will always consider functions on the time interval $[0, 1]$.

Before we turn to the proof of this result, we introduce a very useful notation which was introduced in [HM11]. Given a family $A = \{A_\varepsilon\}_{\varepsilon\in(0,1]}$ of events depending on some parameter $\varepsilon > 0$, we say that A is "almost true" if, for every $p > 0$ there exists a constant C_p such that $\mathbf{P}(A_\varepsilon) \geq 1 - C_p\varepsilon^p$ for all $\varepsilon \in (0,1]$. Similarly for "almost false". Given two such families of events A and B, we say that "A almost implies B" and we write $A \Rightarrow_\varepsilon B$ if $A \setminus B$ is almost false. It is straightforward to check that these notions behave as expected (almost implication is transitive, finite unions of almost false events are almost false, etc). Note also that these notions are unchanged under any reparametrisation of the form $\varepsilon \mapsto \varepsilon^\alpha$ for $\alpha > 0$. Given two families X and Y of real-valued random variables, we will similarly write $X \leq_\varepsilon Y$ as a shorthand for the fact that $\{X_\varepsilon \leq Y_\varepsilon\}$ is "almost true".

Before we proceed, we state the following useful result, where $\|\cdot\|_\infty$ denotes the L^∞ norm and $\|\cdot\|_\alpha$ denotes the best possible α-Hölder constant.

Lemma 6.5 *Let $f\colon [0,1] \to \mathbf{R}$ be continuously differentiable and let $\alpha \in (0,1]$. Then, the bound*

$$\|\partial_t f\|_\infty = \|f\|_1 \leq 4\|f\|_\infty \max\Big\{1, \|f\|_\infty^{-\frac{1}{1+\alpha}} \|\partial_t f\|_\alpha^{\frac{1}{1+\alpha}}\Big\}$$

holds, where $\|f\|_\alpha$ denotes the best α-Hölder constant for f.

Proof. Denote by x_0 a point such that $|\partial_t f(x_0)| = \|\partial_t f\|_\infty$. It follows from the definition of the α-Hölder constant $\|\partial_t f\|_{\mathcal{C}^\alpha}$ that $|\partial_t f(x)| \geq \frac{1}{2}\|\partial_t f\|_\infty$ for every x such that $|x - x_0| \leq \big(\|\partial_t f\|_\infty / 2\|\partial_t f\|_{\mathcal{C}^\alpha}\big)^{1/\alpha}$. The claim then follows from the fact that if f is continuously differentiable and $|\partial_t f(x)| \geq A$ over an interval I, then there exists a point x_1 in the interval such that $|f(x_1)| \geq A|I|/2$. □

With these notations at hand, we have the following statement, which is essentially a quantitative version of the Doob-Meyer decomposition theorem. Originally, it appeared in [Nor86], although some form of it was already present in earlier works. The statement and proof given here are slightly different from those in [Nor86], but are very close to them in spirit.

Lemma 6.6 *Let W be an m-dimensional Wiener process and let A and B be $\mathbf{R}$ and $\mathbf{R}^m$-valued adapted processes such that, for $\alpha = \frac{1}{3}$, one has $\mathbf{E}\big(\|A\|_\alpha + \|B\|_\alpha\big)^p < \infty$ for every p. Let Z be the process defined by*

$$Z_t = Z_0 + \int_0^t A_s\,ds + \int_0^t B_s\,dW(s)\;. \tag{6.2}$$

Then, there exists a universal constant $r \in (0,1)$ such that one has

$$\big\{\|Z\|_\infty < \varepsilon\big\} \quad \Rightarrow_\varepsilon \quad \big\{\|A\|_\infty < \varepsilon^r\big\} \;\&\; \big\{\|B\|_\infty < \varepsilon^r\big\}\;.$$

Proof. Recall the exponential martingale inequality [RY99, p. 153], stating that if M is any continuous martingale with quadratic variation process $\langle M\rangle(t)$, then

$$\mathbf{P}\Big(\sup_{t\le T}|M(t)|\ge x \quad \& \quad \langle M\rangle(T)\le y\Big)\le 2\exp\big(-x^2/2y\big)\;,$$

for every positive T, x, y. With our notations, this implies that for any $q<1$ and any adapted process F, one has the almost implication

$$\big\{\|F\|_\infty<\varepsilon\big\}\quad\Rightarrow_\varepsilon\quad\Big\{\Big\|\int_0^\cdot F_t\,dW(t)\Big\|_\infty<\varepsilon^q\Big\}\;. \tag{6.3}$$

With this bound in mind, we apply Itô's formula to Z^2, so that

$$Z_t^2=Z_0^2+2\int_0^t Z_s\,A_s\,ds+2\int_0^t Z_s\,B_s\,dW(s)+\int_0^t B_s^2\,ds\;. \tag{6.4}$$

Since $\|A\|_\infty\le_\varepsilon\varepsilon^{-1/4}$ (or any other negative exponent for that matter) by assumption and similarly for B, it follows from this and (6.3) that

$$\big\{\|Z\|_\infty<\varepsilon\big\}\quad\Rightarrow_\varepsilon\quad\Big\{\Big|\int_0^1 A_s\,Z_s\,ds\Big|\le\varepsilon^{\frac34}\Big\}\;\&\;\Big\{\Big|\int_0^1 B_s\,Z_s\,dW(s)\Big|\le\varepsilon^{\frac23}\Big\}\;.$$

Inserting these bounds back into (6.4) and applying Jensen's inequality then yields

$$\big\{\|Z\|_\infty<\varepsilon\big\}\quad\Rightarrow_\varepsilon\quad\Big\{\int_0^1 B_s^2\,ds\le\varepsilon^{\frac12}\Big\}\quad\Rightarrow\quad\Big\{\int_0^1|B_s|\,ds\le\varepsilon^{\frac14}\Big\}\;.$$

We now use the fact that $\|B\|_\alpha\le_\varepsilon\varepsilon^{-q}$ for every $q>0$ and we apply Lemma 6.5 with $\partial_t f(t)=|B_t|$ (we actually do it component by component), so that

$$\big\{\|Z\|_\infty<\varepsilon\big\}\quad\Rightarrow_\varepsilon\quad\big\{\|B\|_\infty\le\varepsilon^{\frac{1}{17}}\big\}\;,$$

say. In order to get the bound on A, note that we can again apply the exponential martingale inequality to obtain that this "almost implies" the martingale part in (6.2) is "almost bounded" in the supremum norm by $\varepsilon^{\frac{1}{18}}$, so that

$$\big\{\|Z\|_\infty<\varepsilon\big\}\quad\Rightarrow_\varepsilon\quad\Big\{\Big\|\int_0^\cdot A_s\,ds\Big\|_\infty\le\varepsilon^{\frac{1}{18}}\Big\}\;.$$

Finally applying again Lemma 6.5 with $\partial_t f(t)=A_t$, we obtain that

$$\big\{\|Z\|_\infty<\varepsilon\big\}\quad\Rightarrow_\varepsilon\quad\big\{\|A\|_\infty\le\varepsilon^{1/80}\big\}\;,$$

and the claim follows with $r=1/80$. □

Remark 6.7 By making α arbitrarily close to $1/2$, keeping track of the different norms appearing in the above argument, and then bootstrapping the argument, it is possible to show that

$$\left\{\|Z\|_\infty < \varepsilon\right\} \quad \Rightarrow_\varepsilon \quad \left\{\|A\|_\infty \le \varepsilon^p\right\} \,\&\, \left\{\|B\|_\infty \le \varepsilon^q\right\} ,$$

for p arbitrarily close to $1/5$ and q arbitrarily close to $3/10$. This seems to be a very small improvement over the exponent $1/8$ that was originally obtained in [Nor86], but is certainly not optimal either. The main reason why our result is suboptimal is that we move several times back and forth between L^1, L^2, and L^∞ norms. (Note furthermore that our result is not really comparable to that in [Nor86], since Norris used L^2 norms in the statements and his assumptions were slightly different from ours.)

We now have all the necessary tools to prove Theorem 6.3:

Proof of Theorem 6.3. We fix some initial condition $x_0 \in \mathbf{R}^n$ and some unit vector $\eta \in \mathbf{R}^n$. With the notation introduced earlier, our aim is then to show that

$$\left\{\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon\right\} \quad \Rightarrow_\varepsilon \quad \emptyset , \tag{6.5}$$

or in other words that the statement $\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon$ is "almost false". As a shorthand, we introduce for an arbitrary smooth vector field F on $\mathbf{R}^n$ the process Z_F defined by

$$Z_F(t) = \langle \eta, J_{0,t}^{-1} F(x_t)\rangle ,$$

so that

$$\langle \eta, \mathscr{C}_{0,1}\eta\rangle = \sum_{k=1}^m \int_0^1 |Z_{V_k}(t)|^2\, dt \ge \sum_{k=1}^m \Bigl(\int_0^1 |Z_{V_k}(t)|\, dt\Bigr)^2 . \tag{6.6}$$

The processes Z_F have the nice property that they solve the stochastic differential equation

$$dZ_F(t) = Z_{[F,V_0]}(t)\, dt + \sum_{i=1}^m Z_{[F,V_k]}(t) \circ dW_k(t) , \tag{6.7}$$

which can be rewritten in Itô form as

$$dZ_F(t) = \Bigl(Z_{[F,V_0]}(t) + \sum_{k=1}^m \frac{1}{2} Z_{[[F,V_k],V_k]}(t)\Bigr)\, dt + \sum_{i=1}^m Z_{[F,V_k]}(t)\, dW_k(t) . \tag{6.8}$$

Since we assumed that all derivatives of the V_j grow at most polynomially, we deduce from the Hölder regularity of Brownian motion that, provided that the derivatives of F grow at most polynomially fast, Z_F does indeed satisfy the assumptions on its Hölder norm required for the application of Norris's lemma. The idea now is to observe that, by (6.6), the left hand side of (6.5) states that Z_F is "small" for every $F \in \mathscr{V}_0$. One then argues that, by Norris's lemma, if Z_F is small for every $F \in \mathscr{V}_k$ then, by considering (6.7), it follows that Z_F is also small for every $F \in \mathscr{V}_{k+1}$. Hörmander's condition then ensures that a contradiction arises at some stage, since $Z_F(0) = \langle F(x_0), \xi\rangle$ and there exists k such that $\mathscr{V}_k(x_0)$ spans all of $\mathbf{R}^n$.

Let us make this rigorous. It follows from Norris's lemma and (6.8) that one has the almost implication

$$\big\{\|Z_F\|_\infty < \varepsilon\big\} \quad \Rightarrow_\varepsilon \quad \big\{\|Z_{[F,V_k]}\|_\infty < \varepsilon^r\big\} \;\&\; \big\{\|Z_G\|_\infty < \varepsilon^r\big\}\;,$$

for $k = 1, \ldots, m$ and for $G = [F, V_0] + \frac{1}{2}\sum_{k=1}^m [[F, V_k], V_k]$. Iterating this bound a second time, this time considering the equation for Z_G, we obtain that

$$\big\{\|Z_F\|_\infty < \varepsilon\big\} \quad \Rightarrow_\varepsilon \quad \big\{\|Z_{[[F,V_k],V_\ell]}\|_\infty < \varepsilon^{r^2}\big\}\;,$$

so that we finally obtain the implication

$$\big\{\|Z_F\|_\infty < \varepsilon\big\} \quad \Rightarrow_\varepsilon \quad \big\{\|Z_{[F,V_k]}\|_\infty < \varepsilon^{r^2}\big\}\;, \tag{6.9}$$

for $k = 0, \ldots, m$.

At this stage, we are basically done. Indeed, combining (6.6) with Lemma 6.5 as above, we see that

$$\big\{\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon\big\} \quad \Rightarrow_\varepsilon \quad \big\{\|Z_{V_k}\|_\infty < \varepsilon^{1/5}\big\}\;.$$

Applying (6.9) iteratively, we see that for every $k > 0$ there exists some $q_k > 0$ such that

$$\big\{\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon\big\} \quad \Rightarrow_\varepsilon \quad \bigcap_{V\in\mathscr{V}_k} \big\{\|Z_V\|_\infty < \varepsilon^{q_k}\big\}\;.$$

Since $Z_V(0) = \langle \eta, V(x_0)\rangle$ and since there exists some $k > 0$ such that $\mathscr{V}_k(x_0) = \mathbf{R}^n$, the right hand side of this expression is empty for some sufficiently large value of k, which is precisely the desired result. □

7 Hypercontractivity

The aim of this section is to prove the following result. Let T_t denote the semigroup generated by the Ornstein–Uhlenbeck operator Δ defined in Section 3. In other words, one sets $T_t = \exp(-\Delta t)$, which can be defined by functional calculus. Since we have an explicit eigenspace decomposition of Δ by Proposition 3.7, this is equivalent to simply setting $T_t X = e^{-nt} X$ for every $X \in \mathscr{H}_n$. The main result of this section is the following.

Theorem 7.1 *For $p, q \in (1,\infty)$ and $t \geq 0$ with $\frac{p-1}{q-1} = e^{2t}$, one has*

$$\|T_t X\|_{L^p} \leq \|X\|_{L^q}\;, \tag{7.1}$$

for every $X \in L^q(\Omega, \mathbf{P})$.

Versions of this theorem were first proved by Nelson [Nel66, Nel73] with again constructive quantum field theory as his motivation. An operator T_t which satisfies a bound of the type (7.1) for some $p > q$ is called "hypercontractive". An extremely important feature of this bound is that it holds without the appearance of any proportionality constant. As we will see in Corollary 7.3 below, this makes it stable under tensorisation, which is a very powerful property.

Let us provide a simple application of this result. An immediate corollary is that, for any $X \in \mathscr{H}_n$ and $p \geq 1$, one has the very useful "reverse Jensen's inequality"

$$\mathbf{E}X^{2p} \leq (2p-1)^{np}\big(\mathbf{E}X^2\big)^p\;. \tag{7.2}$$

Although we have made no attempt at optimising this statement, it is already remarkably precise: using Stirling's formula, one can verify from the explicit formula for the moments of a Gaussian distribution that in the case $n = 1$ (when all elements of $\mathscr{H}_n$ have Gaussian distributions) and for large p, one has the asymptotic behaviour

$$\mathbf{E}X^{2p} \leq (2p-1)^p \sqrt{2} e^{\frac{1}{2}-p}\big(\mathbf{E}X^2\big)^p\;.$$

It is however for $n \geq 2$ that the bound (7.2) reveals its full power since the possible distributions for elements of $\mathscr{H}_n$ then cannot be described by a finite-dimensional family anymore.

A crucial ingredient in the proof of Theorem 7.1 is the following tensorisation property. Consider bounded linear operators $T_i \colon L^q(\Omega_i, \mathbf{P}_i) \to L^p(\Omega_i, \mathbf{P}_i)$ for $i \in \{1,2\}$[1] and define the probability space $(\Omega, \mathbf{P})$ by

$$\Omega = \Omega_1 \times \Omega_2\;, \qquad \mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2\;.$$

Then, on random variables of the form $X(\omega) = X_1(\omega_1)X_2(\omega_2)$, one defines an operator $T = T_1 \otimes T_2$ by setting $(TX)(\omega) = (T_1X_1)(\omega_1)\,(T_2X_2)(\omega_2)$, where we used the notation $\omega = (\omega_1, \omega_2)$. Extending T by linearity, this defines T on a dense subset of $(\Omega, \mathbf{P})$. On that subset one can also write $T = \hat{T}_2\hat{T}_1$ where $\hat{T}_1$ acts on functions of Ω by

$$(\hat{T}_1 X)(\omega_1, \omega_2) = \big(T_1 X(\cdot, \omega_2)\big)(\omega_1)\;,$$

[1] We will always assume our spaces to be standard probability spaces, so that no pathologies arise when considering products and conditional expectations.

and similarly for $\hat T_2$. We claim that if each T_i satisfies a hypercontractive bound of the type (7.1), then so does T. A key ingredient for proving this is the following lemma, where we write for example $\|X\|_{L^p_1}$ as a shorthand for the function $\omega_2 \mapsto \|X(\cdot,\omega_2)\|_{L^p(\Omega_1,\mathbf{P}_1)}$.

Lemma 7.2 *If $p \ge q \ge 1$, then one has $\|\|X\|_{L^q_1}\|_{L^p} \le \|\|X\|_{L^p_2}\|_{L^q}$.*

Proof. The holds for $q = 1$ by the triangle inequality since

$$\begin{aligned}\|\|X\|_{L^1_1}\|_{L^p} &= \Big\|\int |X(\omega_1,\cdot)|\,\mathbf{P}_1(d\omega_1)\Big\|_{L^p} \le \int \Big\||X(\omega_1,\cdot)|\Big\|_{L^p}\mathbf{P}_1(d\omega_1)\\ &= \|\|X\|_{L^p_2}\|_{L^1}\;.\end{aligned}$$

Exploiting the fact that $\|X\|_{L^q} = \||X|^q\|_{L^1}^{1/q}$, the general case follows:

$$\begin{aligned}\|\|X\|_{L^q_1}\|_{L^p} &= \|\|X\|_{L^q_1}^q\|_{L^{p/q}}^{1/q} = \|\|\|X|^q\|_{L^1_1}\|_{L^{p/q}}^{1/q}\\ &\le \|\|\|X|^q\|_{L_2^{p/q}}\|_{L^1}^{1/q} = \|\|X\|_{L^p_2}^q\|_{L^1}^{1/q} = \|\|X\|_{L^p_2}\|_{L^q}\;,\end{aligned}$$

thus concluding the proof. □

Corollary 7.3 *In the above setting if, for some $p \ge q \ge 1$ one has $\|T_iX\|_{L^p} \le \|X\|_{L^q}$, then one also has $\|TX\|_{L^p} \le \|X\|_{L^q}$.*

Proof. One has

$$\begin{aligned}\|TX\|_{L^p} &= \|\|\hat T_1\hat T_2 X\|_{L^p_1}\|_{L^p} \le \|\|\hat T_2 X\|_{L^q_1}\|_{L^p}\\ &\le \|\|\hat T_2 X\|_{L^p_2}\|_{L^q} \le \|\|X\|_{L^q_2}\|_{L^q} = \|X\|_{L^q}\;,\end{aligned}$$

as claimed. □

Recall now that, in the context of a Gaussian probability space, the space $\mathscr{W}$ consisting of random variables of the type

$$X = F(\xi_1,\ldots,\xi_n)\;,\qquad F \in \mathscr{C}_b^\infty\;,\tag{7.3}$$

where $\xi_i = W(e_1)$ for an orthonormal basis $\{e_i\}$ of the Cameron-Martin space H, is dense in $L^2(\Omega,\mathbf{P})$. Similarly, one can show that it is actually dense in every $L^p(\Omega,\mathbf{P})$ for $p \in [1,\infty)$, and we will assume this in the sequel.

As a consequence, in order to prove Theorem 7.1, it suffices to show that (7.1) holds for random variables of the type (7.3) for any fixed n. Note now that, using (3.9) for the evaluation of the Skorokhod integral, one has

$$\Delta X = \delta\mathscr{D}X = \delta\sum_i (\partial_i F)(\xi_1,\ldots,\xi_n)e_i$$

$$= \sum_i \big((\partial_i F)(\xi_1, \ldots, \xi_n)\xi_i - (\partial_i^2 F)(\xi_1, \ldots, \xi_n)\big) .$$

In other words, one has $\Delta = \sum_{i=1}^n \Delta_i$, where

$$\Delta_i = -\partial_i^2 + \xi_i \, \partial_i .$$

At this stage we note that, given an ordered orthonormal basis as above, the closed subspace of $L^p(\Omega, \mathbf{P})$ given by the closure of the subspace spanned by random variables of the type (7.3) for any fixed n is canonically isomorphic (precisely via (7.3)) to $L^p(\mathbf{R}^n, \mathcal{N}(0, \mathrm{id}))$. Furthermore, T_t maps that space into itself, so let us write $T_t^{(n)}$ for the corresponding operator on $L^p(\mathbf{R}^n, \mathcal{N}(0, \mathrm{id}))$. It follows from the fact that all of the Δ_i commute that one has

$$T_t^{(n)} = T_t^{(1)} \otimes \ldots \otimes T_t^{(1)} \quad (n \text{ times}),$$

so that as a consequence of Corollary 7.3, Theorem 7.1 follows if we can show the analogous statement for $T_t^{(1)}$.

At this stage, we note that the operator $\mathcal{L} \stackrel{\text{def}}{=} \partial_x^2 - x\partial_x$ is the generator of the standard one-dimensional Ornstein-Uhlenbeck process given by the solutions to the SDE

$$dX = -X \, dt + \sqrt{2} \, dB(t) , \qquad X_0 = x , \tag{7.4}$$

where B is a standard one-dimensional Brownian motion. The Brownian motion B appearing here has nothing to do whatsoever with the white noise process W that was the start of our discussion! In other words, it follows from Itô's formula that if we define an operator P_t on $L^p(\mathbf{R}, \mathcal{N}(0, 1))$ by

$$\big(P_t\varphi\big)(x) = \mathbf{E}_x \varphi(X_t) ,$$

then $P_t\varphi$ solves the equation

$$\partial_t P_t \varphi = \mathcal{L} P_t \varphi . \tag{7.5}$$

Since the operator $\mathcal{L}$ is essentially self-adjoint on $\mathcal{C}_b^\infty$ (see for example [RS72, RS75] for more details), it follows that P_t as defined above does indeed coincide with the operator $T_t^{(1)}$, modulo the isomorphism mentioned above. By the variation of constants formula, the solution to (7.4) is given by

$$X(t) = e^{-t}x + \sqrt{2} \int_0^t e^{s-t} dB(s) .$$

In law, for any fixed t (not as a function of t!), this can be rewritten as

$$X(t) \stackrel{\text{law}}{=} e^{-t}x + \sqrt{1 - e^{-2t}}\theta , \qquad \theta \sim \mathcal{N}(0, 1) ,$$

so that

$$\big(P_t\varphi\big)(x) = \mathbf{E}\varphi\Big(e^{-t}x + \sqrt{1-e^{-2t}}\theta\Big)\;. \tag{7.6}$$

We immediately deduce from this formula the following important properties. First, by differentiating both sides in x, we see that

$$\partial_x(P_t\varphi)(x) = e^{-t}(P_t\partial_x\varphi)(x)\;. \tag{7.7}$$

Applying the Cauchy-Schwarz inequality to (7.6), we also obtain the pointwise bound

$$\big(P_t(\varphi\cdot\psi)\big)(x)^2 \le \big(P_t\varphi^2\big)(x)\,\big(P_t\psi^2\big)(x)\;. \tag{7.8}$$

We also note that if we take x random $\mathcal{N}(0,1)$, independent of θ, and interpret the expectation in the right-hand side of (7.6) as ranging over both x and θ, then it is independent of t. In other words, setting $\mu = \mathcal{N}(0,1)$, one has

$$\int \big(P_t f\big)(x)\,\mu(dx) = \int f(x)\,\mu(dx)\;,\qquad \forall t\ge 0\;. \tag{7.9}$$

Differentiating in t and setting $t=0$ thus yields for $f\in\mathcal{C}_b^\infty$

$$\int \big(\mathcal{L}f\big)(x)\,\mu(dx) = 0\;,\qquad \mu = \mathcal{N}(0,1)\;. \tag{7.10}$$

Finally, recall that integration by parts yields

$$\int f(x)\big(\mathcal{L}g\big)(x)\,\mu(dx) = -\int f'(x)g'(x)\,\mu(dx)\;. \tag{7.11}$$

(This of course also follows from how the Ornstein–Uhlenbeck operator $\delta\mathcal{D}$ was defined in the first place.)

Before we can give the proof of hypercontractivity of P_t, we need a final ingredient. Recall that Sobolev embedding guarantees that the L^p norm of a function can be bounded by its H^1 Sobolev norm, provided that $1/p \ge 1/2 - 1/n$, where n denotes the dimension of the ambient space. The problem with this embedding is twofold: the exponent p depends on the dimension of the space, as do the proportionality constants that arise. It turns out that if we weaken L^p to "$L\log L$", then a Sobolev-type embedding still holds, but this time *independently of dimension*. This was first remarked by Gross [Gro75] and has turned out to be extremely useful in a variety of context. The precise statement is as follows.

Theorem 7.4 *The measure μ satisfies the log-Sobolev inequality, namely*

$$\int f^2\log f^2\,d\mu - \int f^2\,d\mu\,\log\int f^2\,d\mu \le 2\int |\partial_x f|^2\,d\mu\;,$$

for all $f\in\mathcal{W}^{1,2}$.

Proof. (This proof is essentially taken from [Led92].) By a simple density argument, we can assume that f is smooth and bounded. One then has $\lim_{t\to\infty}(P_t f^2)(x) = \int f^2\, d\mu$, uniformly over compact sets. As a consequence, we can use the fundamental theorem of calculus to write

$$\begin{aligned}
&\int f^2 \log f^2\, d\mu - \int f^2\, d\mu \,\log \int f^2\, d\mu \\
&\quad = -\int_0^\infty \frac{d}{dt}\int P_t f^2 \,\log P_t f^2\, d\mu\, dt = -\int_0^\infty \int (\mathscr{L} P_t f^2)\,\log P_t f^2\, d\mu\, dt \\
&\quad = \int_0^\infty \int \frac{(\partial_x P_t f^2)^2}{P_t f^2}\, d\mu\, dt = 4\int_0^\infty e^{-2t}\int \frac{\big(P_t(f\partial_x f)\big)^2}{P_t f^2}\, d\mu\, dt \\
&\quad \le 4\int_0^\infty e^{-2t}\int P_t(\partial_x f)^2\, d\mu\, dt = 4\int_0^\infty e^{-2t}\int (\partial_x f)^2\, d\mu\, dt\;,
\end{aligned}$$

and the claim follows. Here, we first used (7.5), as well as the fact that $\int \mathscr{L} P_t f^2\, d\mu = 0$ by (7.10). To get the third line, we used (7.11), followed by (7.7). Finally, we used (7.8) and (7.9). □

We now have everything in place for the proof of the main theorem of this section.

Proof of Theorem 7.1. As already discussed, we only need to show (7.1) for P_t rather than T_t. Take a smooth strictly positive function $f \in \mathscr{C}_b^\infty(\mathbf{R})$ and write $\Phi(t) = \|P_t f\|_{p(t)}$, with $p(t) = 1 + (q-1)e^{2t}$, so that in particular $\dot p = 2(p-1)$. We also recall that for smooth positive functions g and h one has

$$\frac{d}{dt} g^h = g^{h-1}\big(\dot h\, g \log g + h \dot g\big)\;.$$

A simple calculation then shows that, writing $f_t = P_t f$, one has

$$\begin{aligned}
\dot\Phi(t) &= \Phi^{1-p}\Big(-\frac{2(p-1)}{p^2}\int f_t^p\, d\mu\,\log\int f_t^p\, d\mu \\
&\qquad + \frac{1}{p}\int \big(p f_t^{p-1}\,\mathscr{L} f_t + 2(p-1) f_t^p \log f_t\big)\, d\mu\Big) \\
&= -\frac{2(p-1)}{p^2}\Phi^{1-p}\Big(\int f_t^p\, d\mu\,\log\int f_t^p\, d\mu - \int f_t^p \log f_t^p\, d\mu \\
&\qquad + 2\int \big(\partial_x f^{p/2}\big)^2\, d\mu\Big)\;.
\end{aligned}$$

The log-Sobolev inequality precisely guarantees that the right-hand side is always negative, thus proving the claim. □

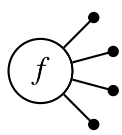
8 Graphical notations and the fourth moment theorem

In this section we show that random variables that belong to a Wiener chaos of fixed order satisfy an incredibly strong form of the central limit theorem: any sequence of random variables such that their variances converge to a finite value and their fourth cumulants converge to 0 must necessarily converge in law to a Gaussian distribution! On the other hand, we will see that if $X \in \mathscr{H}_n$ for some $n \geq 2$, then the law of X itself cannot be Gaussian. (In particular, the convergence to a Gaussian mentioned above can only ever hold in law, never in probability.) In fact, its fourth cumulant is necessarily strictly positive!

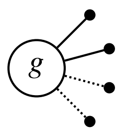
These surprising results were obtained quite recently by Nualart and Peccati in [NP05]. Their proof can make advantageous use of some graphical notation. As usual, we fix a separable Hilbert space H and we note that there is a natural definition of tensor powers $H^{\otimes S}$ for any finite set S. Elements of $H^{\otimes S}$ are linear combinations of expressions of the form $\bigotimes_{i\in S} h_i$ with the usual identifications suggested by the notation. (Of course, $H^{\otimes S}$ is isomorphic to $H^{\otimes k}$ with k the number of elements of S, but this isomorphism is not canonical since it depends on a choice of enumeration of S.) We will sometimes call elements of S "indices".

It is oftentimes natural to consider S as being endowed with a subgroup G_S of its group of permutations. There is a natural action of G_S onto $H^{\otimes S}$ and we will again write $H^{\otimes S}$ for the subspace of those elements that are invariant under that action. In all the cases we consider, one can write $S = S_1 \sqcup \ldots \sqcup S_m$ for some $m \geq 1$ and G_S is given by all the permutations leaving the S_i invariant. In such a situation, we will say that elements belonging to the same S_i are indistinguishable and $H^{\otimes S}$ is canonically isomorphic to $H^{\otimes_s S_1} \otimes \ldots \otimes H^{\otimes_s S_m}$.

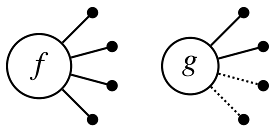
It will be convenient to use graphical notations to represent elements of spaces of the type $H^{\otimes S}$. In particular, the partition of S into subsets of indistinguishable elements will be clear from such a notation. For example, if S is a set of 4 indistinguishable elements, we may denote $f \in H^{\otimes_s S}$ by



with the four black nodes representing the elements of S. If on the other hand one has $\bar{S} = \bar{S}_1 \sqcup \bar{S}_2$ with each $\bar{S}_i$ having two indistinguishable elements, we may write $g \in H^{\otimes_s \bar{S}}$ for example as



Tensor products are then naturally denoted by juxtaposition of pictures, so $f \otimes g$ is simply denoted by

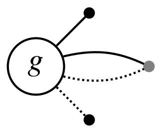


Another important operation is a "partial trace" operation. Given a set S and a pair $p = \{x, y\}$ of two distinct elements of S, we write $\mathrm{Tr}_p \colon H^{\otimes S} \to H^{\otimes(S\setminus p)}$ for the linear map such that

$$\mathrm{Tr}_p \bigotimes_{i\in S} h_i = \langle h_x, h_y\rangle \bigotimes_{i\in S\setminus p} h_i \;. \tag{8.1}$$

Note that if p and $\bar p$ are two disjoint pairs, the corresponding partial traces commute, so one has natural operators $\mathrm{Tr}_{\mathscr{P}}$ for $\mathscr{P}$ any collection of disjoint pairs. Such operations will be represented by identifying the elements in the pair p in our graphical notation and drawing them in grey.

For example, if g is as above and $p = (x, y)$ with $x \in \bar S_1$ and $y \in \bar S_2$, one would denote $\mathrm{Tr}_p\, g$ by



(8.2)

Remark 8.1 Despite (8.1) looking quite "harmless", the operator Tr_p is unbounded and not even closable! For example, if $\{e_n\}_{n\geq 1}$ is an orthonormal basis of H and we set $f = \sum_n \frac{(-1)^n}{n}(e_n \otimes e_n)$, then $f \in H \otimes H$ since $\sum 1/n^2 < \infty$, but $\mathrm{Tr}\, f$ is not defined. Furthermore, one can easily find approximations f_k such that $f_k \to f$ in $H \otimes H$ but $\mathrm{Tr}\, f_k$ converges to any given real number (or diverges).

However, we will only ever consider expressions of the form $\mathrm{Tr}_{\mathscr{P}}(f_1 \otimes \ldots \otimes f_m)$ where $f_i \in H^{\otimes S_i}$ and every pair $p = \{x, y\} \in \mathscr{P}$ is such that $x \in S_i$ and $y \in S_j$ for $i \neq j$. In this particular case, it turns out that this expression is bounded as an m-linear operator, i.e. $\|\mathrm{Tr}_{\mathscr{P}}(f_1 \otimes \ldots \otimes f_m)\| \lesssim \prod_i \|f_i\|$. In particular, expressions like (8.2) actually never appear.

Given a finite set S partitioned into two sets: $S = S_f \sqcup S_n$ (standing for 'free indices' and 'noise indices'), we define the linear map $\mathscr{I} \colon H^{\otimes S} \to \mathscr{H}_k(H^{\otimes S_f})$ where $k = |S_n|$ which acts on elements of the form $h = h_f \otimes h_n$ by

$$\mathscr{I}(h_f \otimes h_n) = h_f \sqrt{k!} I_k(\Pi h_n) \;, \quad k = |S_n| \;,$$

where I_k was defined in (2.8) and the symmetrisation operator Π was defined just before. Here, we identify h_n with an element of $H^{\otimes k}$ by using some enumeration of S_n. The choice of enumeration does not matter thanks to the presence of Π. Graphically, we denote $\mathscr{I}(h)$ by colouring the elements of S_n in red. This is consistent with our notation so far since, in the case when $S_n = \emptyset$, $\mathscr{I}$ is the identity under the canonical identification $\mathscr{H}_0(H^{\otimes S_f}) \simeq H^{\otimes S}$. Note that $\mathcal{J}$ is *not* an isometry in general due to the presence both of the projection Π and of the factor $\sqrt{k!}$. On the other hand, $\mathcal{J}$ is arguably more "natural" in the sense that

$$\mathcal{J}\Big(\bigotimes_{x\in S} e_{i(x)}\Big) = \Phi_k \bigotimes_{x\in S_f} e_{i(x)} , \tag{8.3}$$

where k is such that $k_j = |\{x \in S_n \,:\, i(x) = j\}|$, without any additional combinatorial factor.

The power of these notations can already be appreciated by noting that the Malliavin derivative $\mathscr{D}$ of an element depicted by such a graphical notation is obtained simply by adding up all ways of turning one red node black. Conversely, to define the divergence δ of an element of $\mathscr{H}_k(H^{\otimes S})$, one needs to specify an element $x \in S$ so that $\mathscr{H}_k(H^{\otimes S}) \simeq \mathscr{H}_k(H \otimes H^{\otimes S_x})$, where $S_x = S \setminus \{x\}$. The divergence operator $\delta\colon \mathscr{H}_k(H\otimes H^{\otimes S_x}) \to \mathscr{H}_k(H^{\otimes S_x})$ is then obtained in our graphical notations by simply colouring the node x red. The fact that $\delta\mathscr{D} = k$ on $\mathscr{H}_k$ is then immediate: an element $f \in \mathscr{H}_k$ is depicted by a graph with k red nodes and no black nodes; $\mathscr{D}f$ is obtained by summing over all k ways of turning one of these nodes black, while $\delta\mathscr{D}f$ is then obtained by turning the single black node back red again, thus yielding k times the original graph.

We have the following product formula which should be interpreted as a far-reaching generalisation of Wick's theorem for computing the moments of a Gaussian

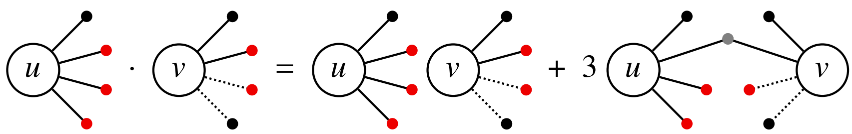
Proposition 8.2 *Let S and $\bar S$ be finite sets and let $h \in H^{\otimes S}$, $\bar h \in H^{\otimes \bar S}$. Then, one has*

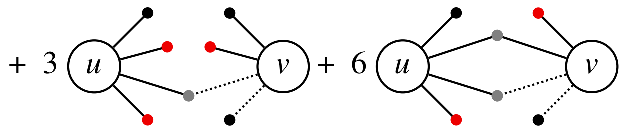
$$\mathcal{J}(h) \otimes \mathcal{J}(\bar h) = \sum_{\mathscr{P}\in P(S_n,\bar S_n)} \mathcal{J}\big(\mathrm{Tr}_{\mathscr{P}}(h\otimes \bar h)\big) , \tag{8.4}$$

where $P(S_n, \bar S_n)$ denotes all collections of disjoint pairs $\{x, y\} \subset S_n \sqcup \bar S_n$ such that $x \in S_n$, $y \in \bar S_n$.

Remark 8.3 Here, if $S = S_f \sqcup S_n$ as before and similarly for $\bar S$, we have implicitly set $(S \sqcup \bar S)_f = S_f \sqcup \bar S_f$ and similarly for the noise indices. Note that the tracing operation removes some noise indices but does not affect the free indices, so that both sides are random variables taking values in $H^{\otimes S_f} \otimes H^{\otimes \bar S_f} \simeq H^{\otimes (S_f \sqcup \bar S_f)}$.

The graphical interpretation of Proposition 8.2 is that the product of two graphs is obtained by iterating over all ways of juxtaposing them and then "contracting" any number of red nodes from the first graph with an identical number of red nodes from the second graph. For example, one has





where we made use of indistinguishability. Before we proceed to the proof of Proposition 8.2, we provide the following preliminary result on the product of Hermite polynomials.

Lemma 8.4 *For $n, m \geq 0$, one has*

$$H_n(x)H_m(x) = \sum_{p\geq 0} C(m,n;p)H_{n+m-2p}(x)\;, \quad C(m,n;p) = p!\binom{m}{p}\binom{n}{p}\;, \tag{8.5}$$

with the convention that $C(m,n;p) = 0$ whenever $p > \min\{m,n\}$.

Proof. We fix n and proceed by induction on m. The case $m = 0$ is trivial, while the case $m = 1$ reads

$$H_1 \cdot H_n = H_{n+1} + nH_{n-1}\;,$$

which is immediate from $H_1(x) = x$, combined with (2.3) and (2.5). Since n is fixed, we write $C_n^{m;p} = C(m,n;p)$ and we note that these coefficients satisfy the recursions

$$\begin{aligned} C_n^{m;p-1} &= C_n^{m+1;p}\frac{p}{(m+1)(n-p+1)}\;, \\ mC_n^{m-1;p-1} &= C_n^{m+1;p}\frac{p(m+1-p)}{(m+1)(n-p+1)}\;, \\ C_n^{m;p} &= C_n^{m+1;p}\frac{m+1-p}{m+1}\;. \end{aligned} \tag{8.6}$$

(This also holds for $p = 0$ if we use the convention that $C_n^{m;p} = 0$ for $p < 0$.) We now assume that (8.5) holds for some m (and all smaller values) and we write

$$\begin{aligned} H_nH_{m+1} &= H_n\big(H_1H_m - mH_{m-1}\big) \\ &= \sum_{p\geq 0} C_n^{m;p}H_1H_{n+m-2p} - m\sum_{p\geq 0} C_n^{m-1;p}H_{n+m-1-2p} \\ &= \sum_{p\geq 0} C_n^{m;p}H_{n+m+1-2p} + \sum_{p\geq 0}(n+m-2p)C_n^{m;p}H_{n+m-1-2p} \\ &\quad - m\sum_{p\geq 0} C_n^{m-1;p}H_{n+m-1-2p} \\ &= \sum_{p\geq 0}\big(C_n^{m;p} + (2+n+m-2p)C_n^{m;p-1} - mC_n^{m-1;p-1}\big)H_{n+m+1-2p}\;. \end{aligned}$$

Combining this with (8.6) completes the proof. □

Proof of Proposition 8.2. In view of (8.3) and the definition of $\mathrm{Tr}_{\mathscr{P}}$, we can assume without loss of generality that $S = S_n$ and similarly for $\bar{S}$. Since both sides of (8.4) are linear in h and $\bar{h}$, it suffices to show that the formula holds for h and $\bar{h}$ of the form

$$h = \bigotimes_{x\in S} e_{i(x)} \;, \qquad \bar{h} = \bigotimes_{y\in \bar{S}} e_{j(x)} \;,$$

for some fixed orthonormal basis $\{e_n\}_{n\in\mathbf{N}}$ and functions $i\colon S \to \mathbf{N}$, $j\colon \bar{S} \to \mathbf{N}$. Write $k(i)\colon \mathbf{N} \to \mathbf{N}$ for the function such that

$$k(i)_j = |\{x \in S \;:\; i(x) = j\}| \;,$$

and similarly for $k(j)$ (but with S replaced by $\bar{S}$).

Since this basis is orthonormal, the only terms that contribute to the right-hand side of (8.4) are those pairings $\mathscr{P}$ such that

$$i(x) = j(y) \;, \qquad \forall \{x, y\} \in \mathscr{P} \;.$$

Call such a pairing admissible. Every admissible pairing $\mathscr{P}$ then yields a function $k(\mathscr{P})\colon \mathbf{N} \to \mathbf{N}$ by setting $k(\mathscr{P})_\ell = |\{\{x, y\} \in \mathscr{P} \;:\; i(x) = \ell\}|$. With this notation, and in view of (8.3) and (8.1), one has

$$\mathscr{I}\big(\mathrm{Tr}_{\mathscr{P}}(h \otimes \bar{h})\big) = \Phi_{k(i)+k(j)-2k(\mathscr{P})}$$

if $\mathscr{P}$ is admissible, and 0 otherwise.

Given $p\colon \mathbf{N} \to \mathbf{N}$, we see that the number of pairings $\mathscr{P}$ such that $k(\mathscr{P}) = p$ is precisely given by $p!\binom{k(i)}{p}\binom{k(j)}{p}$. This is because it is determined by, for every $\ell \in \mathbf{N}$, first choosing p_ℓ elements among the $k(i)_\ell$ indices $x \in S$ with $i(x) = \ell$, then choosing p_ℓ elements among the $k(j)_\ell$ indices $x \in \bar{S}$ with $j(x) = \ell$, and finally choosing one of the $p_\ell!$ ways of pairing these indices up. As a consequence, the right-hand side of (8.5) is given by

$$\sum_{p\colon \mathbf{N}\to\mathbf{N}} p!\binom{k(i)}{p}\binom{k(j)}{p}\Phi_{k(i)+k(j)-2p} \;.$$

Since the left-hand side is given by $\Phi_{k(i)} \cdot \Phi_{k(j)}$ and in view of the definition (2.6) of the Φ_k, the claim immediately follows from Lemma 8.4. □

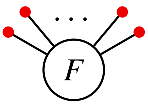
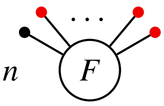
Remark 8.5 A final, but important, remark before we can proceed with the proof of the fourth moment theorem is the following. Recall that with our graphical notation, a graph Γ containing only grey and black nodes represents a deterministic element $f \in H^{\otimes S}$. The squared norm $\|f\|^2$ is then represented by the graph $|\Gamma|^2$ obtained by taking two identical copies Γ_1 and Γ_2 of Γ and pairing up each one of the black nodes of Γ_1 with the corresponding node of Γ_2. In particular, any graph that can be obtained in this way necessarily represents a positive number.

The following lemma which was obtained in [NP05] is now elementary.

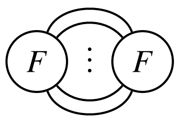
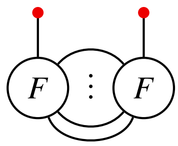
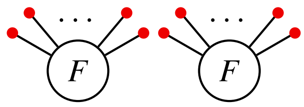
Lemma 8.6 *Let $F \in \mathscr{H}_n$ for some $n \geq 2$ with $F \neq 0$. Then, there exists $c > 0$ depending only on n such that*

$$\mathbf{E}F^4 - 3(\mathbf{E}F^2)^2 \geq c\big(\mathbf{E}\|\mathscr{D}F\|^4 - \big(\mathbf{E}\|\mathscr{D}F\|^2\big)^2\big) > 0 \;. \tag{8.7}$$

Proof. Since $F \in \mathscr{H}_n$, we can represent it and its Malliavin derivative as

$F =$  , $\mathscr{D}F = n$  .

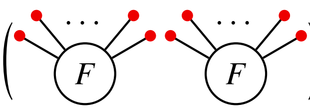
As a consequence of Proposition 8.2, we then have

$F^2 = n!$  $+\; n \cdot n!$  $+ \ldots +$  . (8.8)

(Here, each line connecting two copies of F should be thought of as having a grey node, but we don't draw these.) Since the kth term in this sum belongs to the $2k$th Wiener chaos, they are all orthogonal.

Regarding $\|\mathscr{D}F\|^2$, one also obtains a positive linear combination of the exact same terms, but with slightly different combinatorial prefactors. The main difference however is that the last term is absent since taking the norm squared corresponds to contracting the two black nodes, so one always has at least one contraction between the two copies of F. In particular, this already shows why the second inequality in (8.7) is strict when $n \geq 2$: the second term in (8.8) is of the form $\mathscr{I}(g)$ for some $g \in H^{\otimes_s 2}$ with $\operatorname{Tr} g = n\mathbf{E}F^2$. Since $\mathbf{E}F^2 > 0$, we conclude that one cannot have $g = 0$, so this term has strictly positive L^2 norm.

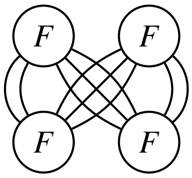
Since the first term of (8.8) (including the factor $n!$) is nothing but $\mathbf{E}F^2$, we conclude that there exists a constant $c > 0$ such that

$\mathbf{E}F^4 \geq (\mathbf{E}F^2)^2 + c\big(\mathbf{E}\|\mathscr{D}F\|^4 - \big(\mathbf{E}\|\mathscr{D}F\|^2\big)^2\big) + \mathbf{E}\Big($  $\Big)^2$.

Denoting the last term in this expression by D, it therefore remains to show that $D \geq 2(\mathbf{E}F^2)^2$. As a consequence of Proposition 8.2, we have

$$D \geq \sum_{p=0}^{n} \binom{n}{p}^2 (n!)^2 F_p^4 \;,$$

with

$F_p^4 =$  ,

where we have p “vertical connections” between any two copies of F and $n - p$ “diagonal connections”. Since each of these F_p^4 is positive by Remark 8.5 (just “untwist” the picture by flipping the two copies of F on the right), we conclude that $D \geq (n!)^2(F_0^4 + F_n^4)$, but $(n!)^2 F_0^4 = (n!)^2 F_n^4 = (\mathbf{E}F^2)^2$, thus yielding the claim. □

Note that since the last inequality in (8.7) is strict, this implies that $\mathscr{H}_n$ itself cannot contain any Gaussian random variable! We are now in a position to prove the fourth moment theorem of [NP05]:

Theorem 8.7 *Let $n \geq 2$ and let $\{F_k\}_{k\geq 0} \in \mathscr{H}_n$ be a sequence of random variables such that $\mathbf{E}F_k^2 = 1$ (say) for all k. Then, the F_k converge in law to $\mathcal{N}(0,1)$ if and only if $\lim_{k\to\infty} \mathbf{E}F_k^4 = 3$.*

Proof. We follow the exposition of [NOL08]. Since all moments of the sequence F_k are uniformly bounded by (7.2), the necessity of $\mathbf{E}F_k^4 \to 3$ is immediate. For the converse implication, we note first that by tightness we can assume that the F_k have some limit in law (modulo extraction of a subsequence which we again denote by F_k) and we write

$$\varphi(t) = \lim_{k\to\infty} \varphi_k(t) = \lim_{k\to\infty} \mathbf{E}e^{itF_k} ,$$

for its Fourier transform, so it remains to show that $\varphi(t) = e^{-t^2/2}$. Note that φ is differentiable and $\dot\varphi = \lim_{k\to\infty} \dot\varphi_k$ locally uniformly as a consequence of the boundedness of the moments of F_k. Since we know that $\varphi(0) = 1$, it is therefore sufficient to show that $\lim_{k\to\infty}\bigl(\dot\varphi_k(t) + t\varphi_k(t)\bigr) = 0$.

Since $\Delta F_k = nF_k$ we then have

$$\begin{aligned}\dot\varphi_k(t) &= i\mathbf{E}\bigl(F_k e^{itF_k}\bigr) = \frac{i}{n}\mathbf{E}\bigl((\delta\mathscr{D}F_k)e^{itF_k}\bigr) = \frac{i}{n}\mathbf{E}\langle \mathscr{D}F_k, \mathscr{D}e^{itF_k}\rangle \\ &= -\frac{t}{n}\mathbf{E}\bigl(\|\mathscr{D}F_k\|^2 e^{itF_k}\bigr) .\end{aligned}$$

Since $\mathbf{E}\|\mathscr{D}F_k\|^2 = n$, we conclude that

$$\begin{aligned}|\dot\varphi_k(t) + t\varphi_k(t)| &= \frac{t}{n}|\mathbf{E}\bigl(\bigl(\|\mathscr{D}F_k\|^2 - n\bigr)e^{itF_k}\bigr)| \\ &\leq \frac{t}{n}|\mathbf{E}\bigl(\bigl(\|\mathscr{D}F_k\|^2 - n\bigr)^2\bigr)|^{1/2} \lesssim \frac{t}{n}\sqrt{\mathbf{E}F_k^4 - 3} ,\end{aligned}$$

where we used Lemma 8.6 in the last bound, whence the claim follows at once. □

9 Construction of the Φ^4_2 field

We now sketch the argument given by Nelson in [Nel66], showing how the hypercontractive bounds of the previous section can be used to construct Φ^4_2 Euclidean field theory. The goal is to build a measure $\mathbf{P}$ on the space $\mathcal{D}'(\mathbf{T}^2)$ of distributions on the two-dimensional torus which is *formally* given by

$$\mathbf{P}(d\Phi) \propto \exp\Big(-\frac{1}{2}\int |\nabla\Phi(x)|^2\,dx - \int |\Phi(x)|^4\,dx\Big)\,\text{“}d\Phi\text{”}\;.$$

This expression is of course completely nonsensical in many respects, not least because there is no “Lebesgue measure” in infinite dimensions. However, the first part in this description is quadratic and should therefore define a Gaussian measure. Recalling that the Gaussian measure with covariance C has density $\exp(-\frac{1}{2}\langle x, C^{-1}x\rangle)$ with respect to Lebesgue measure, this suggests that we should rewrite $\mathbf{P}$ as

$$\mathbf{P}(d\Phi) \propto \exp\Big(-\int |\Phi(x)|^4\,dx\Big)\,\mathbf{Q}(d\Phi)\;, \tag{9.1}$$

where $\mathbf{Q}$ is Gaussian with covariance given by the inverse of the Laplacian. In other words, under $\mathbf{Q}$, the Fourier modes of Φ are distributed as independent Gaussian random variables (besides the reality constraint) with $\hat\Phi(k)$ having variance $1/|k|^2$. In order to simplify things, we furthermore postulate that $\hat\Phi(0) = 0$.

The measure $\mathbf{Q}$ is the law of the *free field*, which also plays a crucial role in the study of critical phenomena in two dimensions due to its remarkable invariance properties under conformal transformations. However, it turns out that (9.1) is unfortunately still nonsensical. Indeed, for this to make sense, one would like at the very least to have $\Phi \in L^4$ almost surely. It turns out that one does not even have $\Phi \in L^2$ since, at least formally, one has

$$\mathbf{E}\|\Phi\|^2_{L^2} = \sum_k \mathbf{E}|\hat\Phi(k)|^2 = \sum_{k\neq 0} \frac{1}{|k|^2} = \infty\;,$$

since we are in two dimensions.

Denote now by G the Green’s function for the Laplacian. One way of defining G is the following. Take a cut-off function $\chi\colon \mathbf{R}^2 \to \mathbf{R}$ which is smooth, positive, compactly supported, and such that $\chi(k) = 1$ for $|k| \le 1$.

$$G(x) = \lim_{N\to\infty} G_N(x)\;,\qquad G_N(x) = \sum_{k\neq 0} \varphi(k/N)|k|^{-2}e^{ikx}\;.$$

It is a standard result that this limit exists, does not depend on the choice of χ, and is such that $G(x) \sim -\frac{1}{2\pi}\log|x|$ for small values of x, and is smooth otherwise.

Furthermore, the function G_N has the property that $|G_N(x)| \lesssim \log |x|^{-1} \wedge \log N$ for all x. Finally, one has

$$|G(x) - G_N(x)| \lesssim |\log N|x|| \wedge \frac{1}{N^2|x|^2} .$$

Note now that for every N, one can find fields Φ_N and Ψ_N that are independent and with independent Gaussian Fourier coefficients such that

$$\mathbf{E}|\hat\Phi_N(k)|^2 = \varphi(k/N)|k|^{-2} , \qquad \mathbf{E}|\hat\Psi_N(k)|^2 = \bigl(1 - \varphi(k/N)\bigr)|k|^{-2} .$$

One then has $\Phi_N + \Psi_N \stackrel{\text{law}}{=} \Phi$ with Φ a free field. We can furthermore choose Φ_N to be a function of Φ by simply setting

$$\hat\Phi_N(k) = \sqrt{\varphi(k/N)}\hat\Phi(k) . \tag{9.2}$$

Furthermore, Ψ_N is "small" in some sense to be made precise and Φ_N is almost surely a smooth Gaussian field with covariance G_N.

Note however that $C_N^2 \stackrel{\text{def}}{=} \mathbf{E}|\Phi_N(x)|^2 = G_N(0) \sim \log N$ as $N \to \infty$. The idea now is to reinterpret the quantity Φ^4 appearing in (9.1) as a "Wick power" which is defined in terms of the 4th Hermite polynomial by

$$:\!\Phi_N(x)^4\!: \stackrel{\text{def}}{=} C_N^4 H_4(\Phi_N(x)/C_N) . \tag{9.3}$$

The point here is that by the defining properties of the Hermite polynomials, one has $\mathbf{E}:\!\Phi_N(x)^4\!: = 0$. Furthermore, and even more importantly, one can easily verify from a simple calculation using Wick's formula that

$$\mathbf{E}\Bigl(:\!\Phi_N(x)^4\!:\,:\!\Phi_N(y)^4\!:\Bigr) = 24 G_N^4(x-y) .$$

In particular, setting

$$X_N \stackrel{\text{def}}{=} \int_{\mathbf{T}^2} :\!\Phi_N(x)^4\!:\, dx ,$$

one has

$$\mathbf{E}X_N^2 = 24 \int_{\mathbf{T}^2}\int_{\mathbf{T}^2} G_N^4(x-y)\, dx\, dy ,$$

which is uniformly bounded as $N \to \infty$. Furthermore, by (9.3), X_N is an element of the fourth homogeneous Wiener chaos $\mathscr{H}_4$ on the Gaussian probability space generated by Φ. It is now a simple exercise to show that there exists a random variable X belonging to $\mathscr{H}_4$ such that $\lim_{N\to\infty} X_N = X$ in L^2, and therefore in every L^p by (7.2). At this stage, we would like to define our measure $\mathbf{P}$ by

$$\mathbf{P}(d\Phi) \propto \exp(-X(\Phi))\, \mathbf{Q}(d\Phi) , \tag{9.4}$$

with X given by the finite random variable we just constructed.

The problem with this is that although $|\Phi_N(x)|^4$ is obviously bounded from below, uniformly in N, $:\Phi_N(y)^4:$ is not! Indeed, the best one can do is

$$:\Phi_N(y)^4: = \big(\Phi_N(y)^2 - 3C_N^2\big)^2 - 6C_N^4 \ge -6C_N^4 \sim -c(\log N)^2\;,$$

for some constant $c > 0$ (actually $c = \frac{3}{2\pi^2}$), so that

$$X_N \ge -c(\log N)^2\;. \tag{9.5}$$

In order to make sense of (9.4) however, we would like to show that the random variable $\exp(-X)$ is integrable. This is where the optimal bound (7.2) plays a crucial role. The idea of Nelson is to exploit the decomposition $\Phi_N + \Psi_N \stackrel{\text{law}}{=} \Phi$ together with Lemma 2.2 to write

$$\begin{aligned}
X &\stackrel{\text{law}}{=} X_N + Y_N\;,\\
Y_N &\stackrel{\text{def}}{=} 4\int_{\mathbf{T}^2} :\Phi_N(x)^3::\Psi_N(x):\,dx + 6\int_{\mathbf{T}^2} :\Phi_N(x)^2::\Psi_N(x)^2:\,dx\\
&\qquad + 4\int_{\mathbf{T}^2} :\Phi_N(x)::\Psi_N(x)^3:\,dx + \int_{\mathbf{T}^2} :\Psi_N(x)^4:\,dx = Y_N^{(1)} + \ldots + Y_N^{(4)}\;.
\end{aligned}$$

Note now that, setting $\tilde G_N = G - G_N$, one has

$$\mathbf{E}|Y_N^{(1)}|^2 = 96\int_{\mathbf{T}^2}\int_{\mathbf{T}^2} G_N^3(x-y)\,\tilde G_N(x-y)\,dx\,dy \lesssim \frac{|\log N|^4}{N^2}\;.$$

Analogous bounds can be obtained for the other $Y_N^{(i)}$. Combining this with (7.2) and the fact that Y_N belongs to the chaos of order 4, we obtain the existence of finite constants c and C such that

$$\mathbf{E}|Y_N|^{2p} \le \frac{c^p p^{4p}|\log N|^{4p}}{N^{2p}} \le C\frac{p^{4p}}{N^p}\;,$$

uniformly over $N \ge C$ and $p \ge 1$. In the sequel, the values of the constants c and C are allowed to change from expression to expression. We conclude that

$$\begin{aligned}
\mathbf{P}(X < -K) &= \mathbf{P}(X_N + Y_N < -K) \le \mathbf{P}\big(Y_N \le c(\log N)^2 - K\big)\\
&\le \mathbf{P}\big(|Y_N| \ge K - c(\log N)^2\big) \le \frac{\mathbf{E}|Y_N|^{2p}}{(K - c(\log N)^2)^p}\\
&\le \frac{Cp^{4p}}{N^p(K - c(\log N)^2)^p}\;,
\end{aligned}$$

provided that $c(\log N)^2 \le K$ and $N \ge C$. We now exploit our freedom to choose both N and p. First, we choose N such that $c(\log N)^2 - K \in [1,2]$ (this is always possible if K is large enough), so that

$$\mathbf{P}(X < -K) \le \frac{Cp^{4p}}{N^p} \le C(pe^{-c\sqrt{K}})^{4p} .$$

We now choose $p = e^{\tilde c\sqrt{K}}$ for some $\tilde c < c$, so that eventually

$$\mathbf{P}(X < -K) \le C\exp\bigl(-ce^{\tilde c\sqrt{K}}\bigr) .$$

We can rewrite this as

$$\mathbf{P}(\exp(-X) > M) \le C\exp\bigl(-ce^{\tilde c\sqrt{\log M}}\bigr) .$$

In particular, the right hand side of this expression is smaller than any inverse power of M, so that $\exp(-X)$ is indeed integrable as claimed.

References

[BH07] F. BAUDOIN and M. HAIRER. A version of Hörmander's theorem for the fractional Brownian motion. *Probab. Theory Related Fields* **139**, no. 3-4, (2007), 373–395.

[Bis81a] J.-M. BISMUT. Martingales, the Malliavin calculus and Hörmander's theorem. In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, vol. 851 of *Lecture Notes in Math.*, 85–109. Springer, Berlin, 1981.

[Bis81b] J.-M. BISMUT. Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. *Z. Wahrsch. Verw. Gebiete* **56**, no. 4, (1981), 469–505.

[Bog98] V. I. BOGACHEV. *Gaussian measures*, vol. 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998. doi:10.1090/surv/062.

[BT05] F. BAUDOIN and J. TEICHMANN. Hypoellipticity in infinite dimensions and an application in interest rate theory. *Ann. Appl. Probab.* **15**, no. 3, (2005), 1765–1777.

[Cas09] T. CASS. Smooth densities for solutions to stochastic differential equations with jumps. *Stochastic Process. Appl.* **119**, no. 5, (2009), 1416–1435.

[CF10] T. CASS and P. FRIZ. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)* **171**, no. 3, (2010), 2115–2141.

[CHLT15] T. CASS, M. HAIRER, C. LITTERER, and S. TINDEL. Smoothness of the density for solutions to Gaussian rough differential equations. *Ann. Probab.* **43**, no. 1, (2015), 188–239. doi:10.1214/13-AOP896.

[Gro75] L. GROSS. Logarithmic Sobolev inequalities. *Amer. J. Math.* **97**, no. 4, (1975), 1061–1083.

[Hai11] M. HAIRER. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.* **135**, no. 6-7, (2011), 650–666. `doi:10.1016/j.bulsci.2011.07.007`.

[HM06] M. HAIRER and J. C. MATTINGLY. Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Ann. of Math. (2)* **164**, no. 3, (2006), 993–1032.

[HM11] M. HAIRER and J. C. MATTINGLY. A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electron. J. Probab.* **16**, (2011), 658–738.

[Hör67] L. HÖRMANDER. Hypoelliptic second order differential equations. *Acta Math.* **119**, (1967), 147–171.

[HP11] M. HAIRER and N. PILLAI. Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. *Ann. IHP Ser. B* (2011). To appear.

[IK06] Y. ISHIKAWA and H. KUNITA. Malliavin calculus on the Wiener-Poisson space and its application to canonical SDE with jumps. *Stochastic Process. Appl.* **116**, no. 12, (2006), 1743–1769.

[KS84] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, vol. 32 of *North-Holland Math. Library*, 271–306. North-Holland, Amsterdam, 1984.

[KS85] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **32**, no. 1, (1985), 1–76.

[KS87] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34**, no. 2, (1987), 391–442.

[Law77] H. B. LAWSON, JR. *The quantitative theory of foliations*. American Mathematical Society, Providence, R. I., 1977. Expository lectures from the CBMS Regional Conference held at Washington University, St. Louis, Mo., January 6–10, 1975, Conference Board of the Mathematical Sciences Regional Conference Series in Mathematics, No. 27.

[Led92] M. LEDOUX. On an integral criterion for hypercontractivity of diffusion semigroups and extremal functions. *J. Funct. Anal.* **105**, no. 2, (1992), 444–465. `doi:10.1016/0022-1236(92)90084-V`.

[Mal78] P. MALLIAVIN. Stochastic calculus of variation and hypoelliptic operators. In *Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976)*, 195–263. Wiley, New York-Chichester-Brisbane, 1978.

[Mal97] P. MALLIAVIN. *Stochastic analysis*, vol. 313 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1997. `doi:10.1007/978-3-642-15074-6`.

[MP06] J. C. MATTINGLY and É. PARDOUX. Malliavin calculus for the stochastic 2D Navier-Stokes equation. *Comm. Pure Appl. Math.* **59**, no. 12, (2006), 1742–1790.

[Nel66] E. NELSON. A quartic interaction in two dimensions. In *Mathematical Theory of Elementary Particles (Proc. Conf., Dedham, Mass., 1965)*, 69–73. M.I.T. Press, Cambridge, Mass., 1966.

[Nel73] E. NELSON. Construction of quantum fields from Markoff fields. *J. Functional Analysis* **12**, (1973), 97–112.

[NOL08] D. NUALART and S. ORTIZ-LATORRE. Central limit theorems for multiple stochastic integrals and Malliavin calculus. *Stochastic Process. Appl.* **118**, no. 4, (2008), 614–628. doi:10.1016/j.spa.2007.05.004.

[Nor86] J. NORRIS. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, vol. 1204 of *Lecture Notes in Math.*, 101–130. Springer, Berlin, 1986.

[NP05] D. NUALART and G. PECCATI. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33**, no. 1, (2005), 177–193. doi:10.1214/009117904000000621.

[Nua06] D. NUALART. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second ed., 2006. doi:10.1007/3-540-28329-3.

[Oco88] D. OCONE. Stochastic calculus of variations for stochastic partial differential equations. *J. Funct. Anal.* **79**, no. 2, (1988), 288–331.

[RS72] M. REED and B. SIMON. *Methods of modern mathematical physics. I. Functional analysis*. Academic Press, New York-London, 1972.

[RS75] M. REED and B. SIMON. *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

[RY99] D. REVUZ and M. YOR. *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third ed., 1999.

[Seg56] I. E. SEGAL. Tensor algebras over Hilbert spaces. I. *Trans. Amer. Math. Soc.* **81**, (1956), 106–134. doi:10.1090/S0002-9947-1956-0076317-8.

[Tak02] A. TAKEUCHI. The Malliavin calculus for SDE with jumps and the partially hypoelliptic problem. *Osaka J. Math.* **39**, no. 3, (2002), 523–559.

[Tak10] A. TAKEUCHI. Bismut–Elworthy–Li-type formulae for stochastic differential equations with jumps. *Journal of Theoretical Probability* **23**, (2010), 576–604.

[Wie38] N. WIENER. The Homogeneous Chaos. *Amer. J. Math.* **60**, no. 4, (1938), 897–936. doi:10.2307/2371268.