An Introduction to Malliavin Calculus

Lecture Notes
SummerTerm 2013

by Markus Kunze

Contents

CHAPTER 1

Stochastic Calculus

The general setting for Malliavin calculus is a *Gaussian probability space*, i.e. a probability space $(\Omega, \Sigma, \mathbb{P})$ along with a closed subspace $\mathscr{H}$ of $L^2(\Omega, \Sigma, \mathbb{P})$ consisting of centered Gaussian random variables. It is often convenient to assume that $\mathscr{H}$ is isometric to another Hilbert space H, typically an L^2-space over a parameter set T.

1.1. The Wiener Chaos Decomposition

We recall that a real-valued random variable X, defined on a probability space $(\Omega, \Sigma, \mathbb{P})$ is called *Gaussian* (or, more precisely, *centered Gaussian*) if its characteristic function φ_X, defined by $\varphi_X(t) := \mathbb{E}\big(\exp(itX)\big)$ is of the form $\varphi_X(t) = e^{\frac{q}{2}t^2}$ for some $q \geq 0$. It is well-known that a Gaussian random variable is either constantly zero (in which case $q = 0$) or, else, $q > 0$ and the distribution has density

$$\frac{1}{\sqrt{2\pi q}} e^{-\frac{x^2}{2q}}$$

with respect to Lebesgue measure dx. In that case, the random variable has finite moments of all orders, its mean is zero (whence it is called "centered") and its variance is q. A Gaussian random variable is called *standard* if it has variance 1.

A family $(X_i)_{i\in I}$ of real-valued random variables is called *Gaussian family* or *jointly Gaussian*, if for any $n \in \mathbb{N}$ and any choice $i_1, \ldots, i_n$ of distinct indices in I, the vector $(X_{i_1}, \ldots, X_{i_n})$ is a Gaussian vector. The latter means that for any $\alpha \in \mathbb{R}^n$, the real-valued random variable $\sum_{k=1}^n \alpha_k X_{i_k}$ is Gaussian.

We now introduce *isonormal Gaussian processes.*

DEFINITION 1.1.1. Let H be a real, separable Hilbert space with inner product $\langle \cdot\,, \cdot \rangle$. An *$H$-isonormal Gaussian process* is a family $W = \{W(h) : h \in H\}$ of real-valued random variables, defined on a common probability space $(\Omega, \Sigma, \mathbb{P})$, such that $W(h)$ is a Gaussian random variable for all $h \in H$ and, for $h, g \in H$, we have $\mathbb{E}\big(W(h)W(g)\big) = \langle h, g\rangle$.

If the Hilbert space H is clear from the context, we will also speak of *isonormal Gaussian processes.* Given an isonormal Gaussian process, the probability space on which the random variables are defined will be denoted by $(\Omega, \Sigma, \mathbb{P})$.

We now note some properties of isonormal Gaussian processes.

PROPOSITION 1.1.2. *Let H be a real, separable Hilbert space.*

(1) *There exists an H-isonormal Gaussian process W.*

Now let $W = \{W(h) : h \in H\}$ be any H-isonormal Gaussian process.

(2) *The map $h \mapsto W(h)$ is an isometry, in particular, it is linear.*
(3) *W is a Gaussian family.*

PROOF. (2) It suffices to prove that $\mathscr{W}$ is linear; in that case, it follows directly from the definition of isonormal process that $\mathscr{W}$ is an isometry. Thus, let $h, g \in H$ and $\lambda, \mu \in \mathbb{R}$. We have

$$\begin{aligned} &\mathbb{E}\Big(\big(W(\lambda h + \mu g) - \lambda W(h) - \mu W(g)\big)^2\Big) \\ = \;& \mathbb{E}W(\lambda h - \mu g)^2 - 2\lambda \mathbb{E}\big(W(\lambda h + \mu g)W(h)\big) - 2\mu \mathbb{E}\big(W(\lambda h + \mu g)W(g)\big) \end{aligned}$$

$$
\begin{aligned}
&\quad +\lambda^2\mathbb{E}W(h)^2 + 2\lambda\mu\mathbb{E}\big(W(h)W(g)\big) + \mu^2\mathbb{E}W(g)^2 \\
&= \|\lambda h + \mu g\|^2 - 2\lambda\langle \lambda h + \mu g, h\rangle - 2\mu\langle \lambda h + \mu g, g\rangle + \lambda^2\|h\|^2 + 2\lambda\mu\langle g, h\rangle + \mu^2\|g\|^2 \\
&= 0\,.
\end{aligned}
$$

This implies that $W(\lambda h + \mu g) = \lambda W(h) + \mu W(g)$ almost surely.

(3) For $n \in \mathbb{N}$, $h_1, \ldots, h_n \in H$ and $\alpha \in \mathbb{R}^n$, we have, by linearity,

$$\sum_{k=1}^n \alpha_k W(h_k) = W\Big(\sum_{k=1}^n \alpha_k h_k\Big)$$

which is Gaussian by assumption.

(1) Since H is separable, it has a countable orthonormal basis $(h_k)_{k\in\mathbb{N}}$ (if the basis is finite, the proof is similar). Let γ denote standard Gaussian measure on $\mathbb{R}$ and consider the infinite product space

$$(\Omega, \Sigma, \mathbb{P}) := \Big(\prod_{k=1}^\infty \mathbb{R}, \bigotimes_{k=1}^\infty \mathcal{B}(\mathbb{R}), \bigotimes_{k=1}^\infty \gamma\Big)$$

By construction, the random variables X_n, defined by $X_n((\omega_k)_{k\in\mathbb{N}}) := \omega_n$ are independent and Gaussian. In particular, they form an orthonormal basis of their closed linear span in $L^2(\Omega, \Sigma, \mathbb{P})$. We now define $W : H \to L^2(\Omega, \Sigma, \mathbb{P})$ to be the isometry that sends h_k to X_k. Then W is an H-isonormal Gaussian process as is easy to see. □

Let us give some examples of isonormal Gaussian processes.

Example 1.1.3. (Brownian motion)
Consider the Hilbert space $H = L^2((0,\infty), \mathscr{B}(0,\infty), \lambda)$, where λ is Lebesgue measure. By Proposition 1.1.2, there exists an H-isonormal Gaussian process W. Let $B_t := W(\mathbb{1}_{(0,t]})$. Then B_t is a centered Gaussian random variable with variance

$$\mathbb{E}B_t^2 = \mathbb{E}W(\mathbb{1}_{(0,t]})W(\mathbb{1}_{(0,t]}) = \big\langle \mathbb{1}_{(0,t]}, \mathbb{1}_{(0,t]}\big\rangle = t$$

Moreover, given $0 \le t_0 < t_1 \ldots < t_n = t < t+s$, the functions $\mathbb{1}_{(t_0,t_1]}, \ldots, \mathbb{1}_{(t_{n-1},t]}, \mathbb{1}_{(t,t+s|}$ are orthogonal in H, hence the random variables $B_{t_1} - B_{t_0} = W(\mathbb{1}_{(t_0,t_1]}), \ldots, B_t - B_{t_{n-1}}, B_{t+s} - B_t$ are orthogonal in $L^2(\Omega)$, i.e. they are uncorrelated. As these random variables are jointly Gaussian (Proposition 1.1.2), they are independent. This shows that $B_{t+s} - B_t$ is independent of the σ-algebra $\mathscr{F}_t := \sigma(B_r : r \le t)$. Thus, the process (B_t) is a Brownian motion — except for the fact that we have not proved that it has continuous paths. However, using Kolmogorov's theorem, it can be proved that B_t has a continuous modification. We refer the reader to the literature for a proof.

One often writes

$$\int_0^T f(t)\,dB_t := W(\mathbb{1}_{(0,T)}f)$$

and calls $\int_0^T f(t)\,dB_t$ the *Wiener integral* of f over $(0,T)$.

Example 1.1.4. (d-dimensional Brownian motion)
Consider the Hilbert space $H = L^2((0,,\infty), \mathscr{B}((0,\infty), \lambda; \mathbb{R}^d)$. We denote the canonical Basis of $\mathbb{R}^d$ by $(e_1, \ldots, e_n)$, i.e. e_i is the vector which has a 1 at position i and 0's at all other positions. If we put $B_t^j := W(\mathbb{1}_{(0,t]}e_j)$, then the vector $B_t := (B_t^1, \ldots, B_t^d)$ is a d-dimensional Brownian motion, i.e. B_t^j is a Brownian motion for every j and B_t^j and B_t^i are independent for $j \neq i$.

Let us also mention, without going into details, a further example which motivates the abstract setting we consider.

EXAMPLE 1.1.5. (Fractional Brownian motion)
A fractional Brownian motion is a Gaussian process with covariance function

$$c_H(t,s) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H})$$

where $H \in (0,1)$ is the so-called *Hurst parameter*. The choice $H = \frac{1}{2}$ yields $c_{\frac{1}{2}}(t,s) = \min\{t,s\}$ which is the covariance function of Brownian motion.

Let $\mathcal{E}$ denote the step functions on, say, $(0,T)$. It can be proved that there exists an inner product $\langle \cdot\,,\,\cdot \rangle_H$ on $\mathcal{E}$ such that

$$\left\langle \mathbb{1}_{(0,t]}, \mathbb{1}_{(0,s]} \right\rangle_H = c_H(t,s)\,.$$

We denote by V_H the completion of $\mathcal{E}$ with respect to the inner product $\langle \cdot\,,\,\cdot \rangle_H$.

Then an V_H-isonormal Gaussian process W gives rise, as above, to a fractional Brownian motion with Hurst parameter H.

The range of the isonormal process W is the subspace $\mathscr{H}$ that was mentioned in the introduction. We now begin to study the structure of that space. This will be done using *Hermite polynomials.*

DEFINITION 1.1.6. For $n \in \mathbb{N}_0$, the n-th Hermite polynomial H_n is defined by $H_0 \equiv 1$ and

$$H_n(x) := \frac{(-1)^n}{n!} e^{\frac{x^2}{2}} \frac{d^n}{dx^n}\left(e^{-\frac{x^2}{2}}\right)$$

for $n \geq 1$.

Consider the function $F(t,x) := \exp(tx - t^2/2)$. Then the Hermite polynomial are the coefficients in the power series expansion of F with respect to t. Indeed, we have

$$\begin{aligned} F(t,x) &= \exp\left(\frac{x^2}{2} - \frac{1}{2}(x-t)^2\right) \\ &= e^{\frac{x^2}{2}} \sum_{n=0}^\infty \frac{t^n}{n!} \frac{d^n}{dt^n} e^{-\frac{(x-t)^2}{2}}\Big|_{t=0} \\ &= \sum_{n=0}^\infty t^n \frac{(-1)^n}{n!} e^{\frac{x^2}{2}} \frac{d^n}{dz^n} e^{-\frac{z^2}{2}}\Big|_{z=x} = \sum_{n=0}^\infty t^n H_n(x)\,. \end{aligned}$$

We note that the convergence of the series is uniform on compact sets.

Now some basic properties of the Hermite polynomials follow easily:

LEMMA 1.1.7. *For $n \geq 1$, we have*
1. $H_n'(x) = H_{n-1}(x)$;
2. $(n+1)H_{n+1}(x) = xH_n(x) - H_{n-1}(x)$
3. $H_n(-x) = (-1)^n H_n(x)$.

PROOF. Throughout, we set $F(t,x) := \exp(tx - t^2/2)$.

(1) We have $\frac{\partial}{\partial x}F(t,x) = tF(t,x) = \sum_{n=0}^\infty t^{n+1}H_n(x)$. On the other hand, interchanging summation and differentiation, $\frac{\partial}{\partial x}F(t,x) = \sum_{n=0}^\infty t^n H_n'(x)$. Thus (1) follows by equating coefficients.

(2) follows similarly, observing that $\frac{\partial}{\partial t}F(t,x) = (x-t)F(t,x) = \sum_{n=0}^\infty t^n x H_n(x) - t^{n+1}H_n(x)$, while, on the other hand, $\frac{\partial}{\partial t}F(t,x) = \sum_{n=0}^\infty n t^{n-1} H_n(x)$.

(3) follows directly from observing that $F(-x,t) = F(x,-t)$. □

We next insert Gaussian random variables into polynomials. Observe that if X is a Gaussian random variable, then $X \in L^r$ for all $r \in [1,\infty)$. Thus Hölder's inequality yields that for every polynomial p the random variable $p(X)$ belongs to L^r for all $r \in [1,\infty)$. The following Lemma clarifies the basic relationship between Hermite polynomials and Gaussian random variables.

LEMMA 1.1.8. *Let X, Y be standard Gaussian random variables which are jointly Gaussian. Then for $n, m \geq 0$, we have*

$$\mathbb{E}\big(H_n(X)H_m(Y)\big) = \begin{cases} 0 & \text{if } \; n \neq m \\ \frac{1}{n!}\big(\mathbb{E}(XY)\big)^n & \text{if } \; n = m. \end{cases}$$

PROOF. Let $\rho := \mathbb{E}[XY]$. Then $\rho = \operatorname{Cov}(X,Y)$, since $\mathbb{E}X^2 = \mathbb{E}Y^2 = 1$. The moment generating function of the vector (X,Y) is given by

$$\mathbb{E}e^{sX+tY} = e^{\frac{1}{2}(t^2+s^2+2st\rho)}\,.$$

Equivalently, we have

$$\mathbb{E}\big(\exp(sX - s^2/2)\exp(tY - t^2/2)\big) = \exp(st\rho)\,.$$

Differentiating n times with respect to s and m times with respect to t and evaluating at $s = t = 0$, we obtain

$$\mathbb{E}\big(n!m!H_n(X)H_m(Y)\big) = \begin{cases} 0 & \text{if } \; n \neq m \\ n!\rho^n & \text{if } \; n = m. \end{cases} \qquad \square$$

DEFINITION 1.1.9. Let W be an H-isonormal Gaussian process. The *n-th Wiener chaos* $\mathscr{H}_n$ is the closure in $L^2(\Omega,\Sigma,\mathbb{P})$ of the linear span of the set $\{H_n(W(h)) \,:\, h \in H, \|h\| = 1\}$.

As $H_0 \equiv 1$, the 0-th Wiener chaos $\mathscr{H}_0$ is the set of all constant functions, whereas $\mathscr{H}_1 = \{W(h) \,:\, h \in H\}$, since $H_1(x) = x$ and since W is linear.

As a consequence of Lemma 1.1.8 we obtain

COROLLARY 1.1.10. *Let W be an H-isonormal Gaussian process. Then for $n \neq m$ the spaces $\mathscr{H}_n$ and $\mathscr{H}_m$ are orthogonal.*

PROOF. If $g, h \in H$ with $\|g\| = \|h\| = 1$, then $W(g)$ and $W(h)$ are standard Gaussian random variables. Moreover, they are jointly Gaussian, as W is a Gaussian family. Thus, by Lemma 1.1.8, $H_n(W(g))$ and $H_m(W(h))$ are orthogonal in $L^2(\Omega,\Sigma,\mathbb{P})$. This orthogonality relation extends to the linear hull of such elements and also to their closures $\mathscr{H}_n$ and $\mathscr{H}_m$. $\square$

It is a rather natural question, how the direct sum $\bigoplus_{n=0}^\infty \mathscr{H}_n$ looks like. Clearly, any element of the latter is measurable with respect to the σ-algebra $\Sigma_W := \sigma(W(h) \,:\, h \in H)$. As it turns out, we have

THEOREM 1.1.11. *Let W be an H-isonormal Gaussian process. Then*

$$\bigoplus_{n=0}^\infty \mathscr{H}_n = L^2(\Omega,\Sigma_W,\mathbb{P})$$

and this decomposition is orthogonal.

PROOF. In view of Corollary 1.1.10 it only remains to prove that if $X \in L^2(\Omega,\Sigma_W,\mathbb{P})$ is orthogonal to $\bigoplus_{n=0}^\infty \mathscr{H}_n$, then $X = 0$. Let an $X \in L^2(\Omega,\Sigma_W,\mathbb{P})$ which is orthogonal to $\bigoplus_{n=0}^\infty \mathscr{H}_n$ be given. Then $\mathbb{E}\big(XH_n(W(h))\big) = 0$ for all $n \in \mathbb{N}_0$ and $h \in H$ with $\|h\| = 1$. Noting that H_n is a polynomial of degree n, it follows that the Hermite polynomials are a vector space basis of the space of polynomials. By linearity, it follows that $\mathbb{E}\big(Xp(W(h))\big) = 0$ for all polynomials p and every $h \in H$. Setting $p_n(x) := \sum_{k=0}^n \frac{x^k}{k!}$, we have $p_n(W(h)) \to \exp(W(h))$ pointwise and $|p_n(W(h))| \leq \exp(|W(h)|)$ where the latter is square integrable. It follows from dominated convergence that $\mathbb{E}\big(Xe^{W(h)}\big) = 0$ for all $h \in H$.

By the linearity of W, we have

$$\mathbb{E}\big(Xe^{\sum_{j=1}^d t_jW(h_j)}\big) = 0 \qquad \forall\, t_1,\ldots,t_n \in \mathbb{R}, h_1,\ldots h_n \in H\,.$$

In fact, as the left hand side is also well-defined for $t_j \in \mathbb{C}$ and is an analytic function of its variables, by the uniqueness theorem for analytic functions implies that this relation remains valid for complex $t_1, \dots, t_n$. In particular, the characteristic function of the measure ν on $\mathbb{R}^n$, defined by

$$\nu(A) = \mathbb{E}\big(X \mathbb{1}_A(W(h_1), \dots, W(h_n))\big),$$

vanishes. By the uniqueness theorem for characteristic functions, $\nu = 0$. It thus follows that $\mathbb{E}\big(X\mathbb{1}_B\big) = 0$ for all sets B of the form $\{(W(h_1), \dots, W(h_n)) \in A\}$ for $n \in \mathbb{N}$, $h_1, \dots, h_n \in H$ and $A \in \mathscr{B}(\mathbb{R}^n)$. Since sets of this form generate Σ_W, it follows that $\mathbb{E}\big(X\mathbb{1}_B\big) = 0$ for all $B \in \Sigma_W$. Since X is Σ_W-measurable, we have $X = 0$. □

In the case where $H = \mathbb{R}$ is one-dimensional this reduces to

Corollary 1.1.12. *Let γ be standard Gaussian measure. Then $\{(n!)^{\frac{1}{2}} H_n : n \in \mathbb{N}_0\}$ is an orthonormal basis of $L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), \gamma)$.*

Proof. Let $(\Omega, \Sigma, \mathbb{P}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}), \gamma)$ and $H = \mathbb{R}$. Define $W : H \to L^2(\Omega, \Sigma, \mathbb{P})$ by $\big(W(h)\big)(x) = hx$. Then W is an isonormal Gaussian process. Indeed, under γ, x is a Gaussian random variable hence so is ever multiple $W(h)$. Moreover, $\mathbb{E}\big(W(h)W(g)\big) = hg\int_{\mathbb{R}} x^2\, d\gamma = hg$.

Further note that H has only two elements of norm one: $+1$ and -1 which correspond to the random variables x and $-x$ respectively. It follows from Lemma 1.1.7 that $H_n(x) = (-1)^n H_n(-x)$, whence $\mathscr{H}_n$ is one-dimensional. Thus, by Theorem 1.1.11 and Lemma 1.1.8, $(n!)^{\frac{1}{2}} H_n$ is an orthonormal basis of $L^2(\mathbb{R}, \Sigma_W, \mathbb{P})$. Noting that $\Sigma_W = \sigma(x) = \mathscr{B}(\mathbb{R})$, the thesis follows. □

We end this section by providing more information about the spaces $\mathscr{H}_n$.

Definition 1.1.13. Let W be an H-isonormal Gaussian process. We define $\mathscr{P}_n^0$ as the set (in fact, vector space) of all random variables of the form $p(W(h_1), \dots, W(h_k))$, where $k \in \{1, \dots, n\}$, $h_1, \dots, h_k \in H$ and p is a polynomial in k variables of degree at most n.

Moreover, $\mathscr{P}_n$ is the closure of $\mathscr{P}_n^0$ in $L^2(\Omega, \Sigma, \mathbb{P})$.

Proposition 1.1.14. *We have $\mathscr{P}_n = \mathscr{H}_0 \oplus \cdots \oplus \mathscr{H}_n$.*

Proof. Clearly, $\mathscr{H}_0 \oplus \cdots \oplus \mathscr{H}_n \subset \mathscr{P}_n$. To prove the converse inclusion, it suffices to prove that $\mathscr{P}_n \perp \mathscr{H}_m$ for all $m > n$. In fact, it suffices to prove $\mathscr{P}_n^0 \perp \mathscr{H}_m$ for $m > n$. To that end, let $X = p(W(h_1), \dots, W(h_k)) \in \mathscr{P}_n^0$ and $h \in H$ with $\|h\| = 1$ be given. We have to prove that $\mathbb{E}\big(XH_m(W(h))\big) = 0$ for $m > n$.

Extend h to an orthonormal basis $\{h, e_1, \dots, e_j\}$ of $\operatorname{span}\{h_1, \dots, h_k, h\}$. Expressing the h_i's in this basis we can write $X = q(W(e_1), \dots, W(e_j), W(h))$ for a polynomial of degree at most n. Note that the random variables $W(e_1), \dots, W(e_j), W(h)$ are jointly Gaussian and uncorrelated, hence independent. It follows that

$$\mathbb{E}\big(W(e_1)^{\alpha_1} \cdot \ldots \cdot W(e_j)^{\alpha_j} W(h)^\beta H_m(W(h))\big) = \prod_{l=1}^{j} \mathbb{E}\big(W(e_j)^{\alpha_j}\big) \cdot \mathbb{E}\big(W(h)^\beta H_m(W(h))\big).$$

If $\beta = 0$, the latter is zero, as $\mathbb{E}H_m(W(h)) = \mathbb{E}\big(H_0(W(h))H_m(W(h))\big) = 0$ by Lemma 1.1.8. If, on the other hand, $0 < \beta \le n < m$, then x^β is a linear combination of Hermite polynomials H_l with indices $l \le n < m$. Thus Lemma 1.1.8 yields that the above expectation is zero. Together, it follows that $\mathbb{E}\big(XH_m(W(h))\big) = 0$, which is what we needed to prove. □

Finally, we provide an orthonormal basis of $\mathscr{H}_n$ and thus, in view of Theorem 1.1.11, also of $L^2(\Omega, \Sigma_W, \mathbb{P})$. We suppose that H is infinite dimensional, the case where H is finite dimensional is similar, but easier. We pick an orthonormal basis $(e_n)_{n\in\mathbb{N}}$ of H.

Let Λ be the set of all sequences $\alpha = (\alpha_n)_{n\in\mathbb{N}_0}$ such that $\alpha_n \in \mathbb{N}_0$ for all n and $\alpha_n = 0$ for all but finitely many values of n. For $\alpha \in \Lambda$, we put $|\alpha| := \sum_{n=0}^\infty \alpha_n$ and $\alpha! := \prod_{n=0}^\infty \alpha_n!$.

For $\alpha \in \Lambda$, we define

$$\Phi_\alpha := \sqrt{\alpha!}\prod_{n=0}^\infty H_{\alpha_n}(W(e_n))\,.$$

Note that in the above product all but finitely many factors are 1, since $H_0 \equiv 1$.

Proposition 1.1.15. *For $n \in \mathbb{N}_0$, the family $\{\Phi_\alpha : |\alpha| = n\}$ of random variables is an orthonormal basis of $\mathscr{H}_n$.*

Proof. Let us first note that since the reandom variables $W(e_j)$ are independent, we have

$$\mathbb{E}\big(\Phi_\alpha\Phi_\beta\big) = \sqrt{\alpha!}\sqrt{\beta!}\prod_{k=0}^\infty \mathbb{E}\big(H_{\alpha_k}(W(e_k))H_{\beta_k}(W(e_k))\big) = \delta_{\alpha\beta}$$

by Lemma 1.1.8. Thus $\{\Phi_\alpha\}$ is an orthonormal system.

If $|\alpha| = n$, then clearly $\Phi_\alpha \in \mathscr{P}_n^0 \subset \mathscr{P}_n$. Conversely, if $X = p(W(h_1), \ldots, W(h_k)) \in \mathscr{P}_n^0$, then X can be approximated by polynomials in the $W(e_j)$ of degree at most n. These two facts together yield that $\overline{\text{span}}\{\Phi_\alpha : |\alpha| \leq n\} = \mathscr{P}_n$.

Together with the orthogonality relation, we see inductively that $\{\Phi_\alpha : |\alpha| = n\}$ is an orthonormal basis of $\mathscr{H}_n$ as claimed. □

1.2. The Malliavin Derivative

Throughout this section, actually for the rest of this chapter, we fix an H-isonormal Gaussian process W. The underlying probability space is denoted by $(\Omega, \Sigma, \mathbb{P})$. For simplicity, we assume that $\Sigma = \Sigma_W$.

In this section, we want to define the derivative of a random variable $X \in L^2(\Omega)$ with respect to the "chance parameter" ω. However, usual definitions of "derivative" cannot be used in our situation for the following reasons: (i) our probability space Ω is in general lacking the necessary structure to define a derivative (such as linear structure, Banach space structure, etc.), (ii) random variables are usually defined only almost everywhere.

To solve this problem, we think of "randomness" as being injected only via the isonormal process W and take derivatives with respect to the "parameter" h. Thus, in particular, the derivative should take values in H^*, the dual space of H, which, however, we shall identify with H.

Before giving the formal definition, we need some preparation. By $C_p^\infty(\mathbb{R}^d)$, we denote the space of infinitely differentiable functions on $\mathbb{R}^d$ which, together with all their partial derivatives, have polynomial growth. $C_b^\infty(\mathbb{R}^d)$ denotes the subspace of $C_p^\infty(\mathbb{R}^d)$ consisting of those functions which, together with all their partial derivatives, are bounded. Finally $C_c^\infty(\mathbb{R}^d)$ denotes the subspace of $C_b^\infty(\mathbb{R}^d)$ consisting of functions with compact support.

By $\mathscr{S}$ (resp. $\mathscr{S}_b$, resp. $\mathscr{S}_c$) we denote the class of random variables X such that there exists an $n \in \mathbb{N}$, vectors $h_1, \ldots, h_n$ and a function $f \in C_p^\infty(\mathbb{R}^n)$ (resp. $f \in C_b^\infty(\mathbb{R}^n)$, resp. $f \in C_c^\infty(\mathbb{R}^n)$) such that

$$X = f(W(h_1), \ldots, W(h_n))\,.$$

The elements of $\mathscr{S}$ are called *smooth random variables.*

Clearly, $\mathscr{P}_0 \subset \mathscr{S}$, thus both spaces are dense in $L^2(\Omega)$ by Theorem 1.1.11 and Proposition 1.1.14 (recall that we assumed $\Sigma = \Sigma_W$). Moreover, $\mathscr{S}_c \subset \mathscr{S}_b \subset \mathscr{S}$. Approximating a polynomial with C_c^∞-functions, it is easy to see that also $\mathscr{S}_c$ is dense in $L^2(\Omega)$.

Definition 1.2.1. Let $X = f(W(h_1), \ldots, W(h_n)) \in \mathscr{S}$. The *Malliavin derivative of X* is the H-valued random variable DX, defined by

$$DX := \sum_{j=1}^n \partial_j f(W(h_1), \ldots, W(h_n))h_j,$$

where $\partial_j f$ denotes the j-th partial derivative of f.

Note that for $h \in H$, we have

$$\langle DX, h\rangle = \lim_{t\to 0} \frac{f(W(h_1) + t\langle h_1, h\rangle, \dots, W(h_n) + t\langle h_n, h\rangle) - f(W(h_1), \dots, W(h_n))}{t},$$

which gives an interpretation of DX as a directional derivative.

EXAMPLE 1.2.2. Let us consider again the situation of a Brownian motion, i.e. $H = L^2(0,\infty)$. Recall, that we often write $W(h) =: \int_0^\infty h(t)\,dB_t$ where B_t is Brownian motion.

Then $X := \int_0^\infty h(t)\,dB_t \in \mathscr{S}$ and $DX = h$. In particular, $DB_t = \mathbb{1}_{(0,t]}$.

As a slightly more complicated example, consider $X = B_t^2 = (W(\mathbb{1}_{(0,t]}))^2 \in \mathscr{S}$. Then $DX = 2W(\mathbb{1}_{(0,t]})\mathbb{1}_{(0,t]} = 2B_t\mathbb{1}_{(0,t]}$.

Before proceeding, we have to check that the definition of the Malliavin derivative does not depend on the representation of the random variable X.

LEMMA 1.2.3. *Assume that $X \in \mathscr{S}$ has representations*

$$X = f(W(h_1), \dots, W(h_n)) = g(W(e_1), \dots, W(e_m))$$

where $e_1, \dots, e_m$ are an orthonormal system. Then

$$\sum_{j=1}^n \partial_j f(W(h_1), \dots, W(h_n))h_j = \sum_{i=1}^m \partial_i g(W(e_1), \dots, W(e_m))e_i$$

PROOF. We may assume without loss of generality that the span of the h_j's is the same as the span of the e_i's, for otherwise we can put $h_{n+1} = e_1, \dots, h_{n+m} = e_m$ and replace f with $\tilde f$, defined by $\tilde f(x_1, \dots, x_{n+m}) := f(x_1, \dots, x_n)$ and extend $e_1, \dots, e_m$ to an orthonormal basis $e_1, \dots, e_{m+r}$ of the span of both replace g with $\tilde g$, defined by $\tilde g(x_1, \dots, x_{m+r}) := g(x_1, \dots, x_m)$. Note that this does not change the sums in the conclusion of the statement, as $\partial_j \tilde f \equiv 0$ for $j > n$ and $\partial_i \tilde g \equiv 0$ for $i > m$.

Recall that the h_j's are represented in the orthonormal basis $e_1, \dots, e_m$ by the formula $h_j = \sum_{i=1}^m \langle h_j, e_i\rangle e_i$. We let T be the linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ given by the matrix $(\langle h_j, e_i\rangle)_{1\le j\le n, 1\le i\le m}$. Then $T(W(e_1), \dots, W(e_m)) = (W(h_1), \dots, W(h_n))$, so that $(f\circ T)(W(e_1), \dots, W(e_m)) = X = g(W(e_1), \dots, W(e_m))$. This implies that $f \circ T \equiv g$. Indeed, if $f\circ T(x_0) \neq g(x_0)$, then, by continuity, $|f\circ T - g| \ge \varepsilon$ in a neighborhood of x_0. As the probability that the standard Gaussian vector $W(e_1), \dots, W(e_m)$ takes values in that neighborhood is strictly positive, this contradicts the above equality.

Thus, using the chain rule, we find

$$\begin{aligned}
\sum_{i=1}^m \partial_i g(W(e_1), \dots, W(e_m))e_i &= \sum_{i=1}^m \partial_i (f\circ T)(W(e_1), \dots, W(e_m))e_i \\
&= \sum_{i=1}^m \sum_{j=1}^n (\partial_j f \circ T)(W(e_1), \dots, W(e_m))\langle h_j, e_i\rangle e_i \\
&= \sum_{j=1}^m \partial_j f(W(h_1), \dots, h_m)h_j\,.
\end{aligned}$$

□

We next establish an important integration-by-parts formula.

LEMMA 1.2.4. *Let $X \in \mathscr{S}$ and $h \in H$. Then $\mathbb{E}\big(\langle DX, h\rangle\big) = \mathbb{E}\big(XW(h)\big)$.*

PROOF. By linearity, we may assume that $\|h\| = 1$. Furthermore, we can assume that X is given as

$$X = f(W(e_1), \ldots, W(e_n))$$

where $f \in C_p^\infty(\mathbb{R}^n)$ and $e_1, \ldots, e_n$ is an orthonormal system with $e_1 = h$. Thus, the vector $W(e_1), \ldots, W(e_n)$ has standard normal distribution on $\mathbb{R}^n$. Note that in this case $\langle DX, h\rangle = \partial_1 f(W(e_1), \ldots, W(e_n))$.

We denote by ρ the density of the standard normal distribution, i.e.

$$\rho(x) = (2\pi)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2}\sum_{j=1}^n x_j^2\Big).$$

Using integration by parts, we find

$$\begin{aligned}\mathbb{E}\big(\langle DX, h\rangle\big) &= \int_{\mathbb{R}^n} \big(\partial_1 f(x)\big)\rho(x)\, dx \\ &= -\int_{\mathbb{R}^n} f(x)\rho(x)(-x_1)\, dx \\ &= \mathbb{E}\big(XW(h)\big).\end{aligned}$$

□

COROLLARY 1.2.5. *Let $X, Y \in \mathscr{S}$ and $h \in H$. Then*

$$\mathbb{E}\big(Y\langle DX, h\rangle\big) = \mathbb{E}\big(XYW(h) - X\langle DY, h\rangle\big).$$

PROOF. Without loss of generality, we can assume that there exist $h_1, \ldots, h_n \in H$ and $f, g \in C_p^\infty(\mathbb{R}^n)$ such that

$$X = f(W(h_1), \ldots, W(h_n)) \quad \text{and} \quad Y = g(W(h_1), \ldots, W(h_n)).$$

It now follows from the product rule that $D(XY) = XDY + YDX$. The claim thus follows from Lemma 1.2.4. □

We next show that the Malliavin derivative is *closable*. Before proceeding, let us recall some terminology. Let E and F be Banach spaces. An *(unbounded) operator* from E to F is a linear map $A : D(A) \to F$, where $D(A)$ is a subspace of E. Such an operator is called *closed*, if its *graph* $\Gamma(A) := \{(x, Ax) : x \in D(A)\}$ is a closed subspace of the product $E \times F$. Equivalently, if (x_n) is a sequence in $D(A)$ and $x_n \to x$ in E and $Ax_n \to y$ in F, it follows that $x \in D(A)$ and $Ax = y$. An operator is called *closable*, if the closure of its graph is again the graph of an operator, called the *closure of* A. Note that the closure of A – if it exists – is necessarily closed. An operator A is closable if (and only if) whenever x_n is a sequence in $D(A)$ which converges to 0 and such that Ax_n converges to some $y \in F$, it follows that $y = 0$.

Let us also recall the definition of vector valued L^p spaces. A map $f : \Omega \to H$ is called *(strongly) measurable* if it is the pointwise limit of simple functions $\sum x_k \mathbb{1}_{A_k}$. A strongly measureable map f is called p-integrable, if $\int_\Omega \|f\|^p \, d\mathbb{P} < \infty$. We then write $f \in L^p(\Omega; H)$. As usual, we identify functions which are equal almost everywhere. It is easy to see that $L^p(\Omega; H)$ is a Banach space with respect to the norm

$$\|f\|_{L^p(\Omega;H)} := \Big(\int_\Omega \|f\|_H^p \, d\mathbb{P}\Big)^{\frac{1}{p}}.$$

It is easy to see that $L^2(\Omega; H)$ is a Hilbert space with respect to the inner product $\langle f, g\rangle := \int_\Omega \langle f, g\rangle_H \, d\mathbb{P}$.

PROPOSITION 1.2.6. *Let $p \in [1, \infty)$. The operator D, viewed as an operator from $L^p(\Omega)$ to $L^p(\Omega; H)$ initially defined on $\mathscr{S}$, is closable.*

PROOF. Let X_n be a sequence in $\mathscr{S}$ which converges to 0 in $L^p(\Omega)$ and such that DX_n converges to ξ in $L^p(\Omega; H)$. We need to show that $\xi = 0$.

To that end, let $h \in H$ and $Y \in \mathscr{S}_b$ be given such that $YW(h)$ is bounded. By Corollary 1.2.5, we have

$$\begin{aligned}\mathbb{E}\big(Y\langle \xi, h\rangle\big) &= \lim \mathbb{E}\big(Y\langle DX_n, h\rangle\big)\\ &= \lim \mathbb{E}\big(YX_nW(h) - X_n\langle DY, h\rangle\big) = 0\end{aligned}$$

since $X_n \to 0$ in $L^p(\Omega)$ and $YW(h)$ and $\langle DY, h\rangle$ are bounded. For density reasons it follows that $\mathbb{E}\langle \xi, h\rangle = 0$. Since $h \in H$ was arbitrary, it follows that $\xi = 0$, which finishes the proof. □

Slightly abusing notation, we shall denote the closure of D in $L^p(\Omega)$ again by D. The domain of D in $L^p(\Omega)$ is denoted by $\mathbb{D}^{1,p}$. Thus $\mathbb{D}^{1,p}$ is the closure of $\mathscr{S}$ with respect to the norm

$$\|X\|_{1,p} := \big(\mathbb{E}|X|^p + \mathbb{E}\|DX\|_H^p\big)^{\frac{1}{p}}.$$

For $p = 2$, the space $\mathbb{D}^{1,2}$ is a Hilbert space with respect to the inner product

$$\langle X, Y\rangle_{1,2} = \mathbb{E}(XY) + \mathbb{E}\langle DX, DY\rangle_H.$$

We also note that for $p \in (1, \infty)$ the space $\mathbb{D}^{1,p}$ is reflexive as it is isometrically isomorphic to a closed subspace of the reflexive space $L^p(\Omega) \times L^p(\Omega; H)$.

We next establish a *chain rule* for Malliavin derivatives.

PROPOSITION 1.2.7. *Let $p \geq 1$, $X = (X_1, \ldots, X_m)$ be a vector of random variables with $X_j \in \mathbb{D}^{1,p}$ for $j = 1, \ldots, m$ and $\varphi : \mathbb{R}^m \to \mathbb{R}$ be continuously differentiable with bounded partial derivatives. Then $\varphi(X) \in \mathbb{D}^{1,p}$ and*

$$D\varphi(X) = \sum_{j=1}^{m} \partial_j \varphi(X) DX_j. \tag{1.1}$$

PROOF. If $X_j \in \mathscr{S}$ for $j = 1, \ldots, m$ and $\varphi \in C^\infty(\mathbb{R}^m)$ with bounded partial derivatives, then equation (1.1) follows directly from the definition and the classical chain rule. For the general case, pick sequences $X_j^{(n)}$ in $\mathscr{S}$ with $X_j^{(n)} \to X_j$ in $\mathbb{D}^{1,p}$ for $j = 1, \ldots, m$ (thus we have $X_j^{(n)} \to X_j$ in $L^p(\Omega)$ and $DX_j^{(n)} \to DX_j$ in $L^p(\Omega; H)$ for all j) and a sequence $\varphi_n \in C^\infty(\mathbb{R})$ such that $\varphi_n \to \varphi$ uniformly on compact sets, $\nabla\varphi_n \to \nabla\varphi$ uniformly on compact sets and $|\varphi_n| \leq C(1 + |x|)$ for a certain constant C (the latter is possible since φ, having bounded derivatives is of linear growth).

It then follows from dominated convergence that $\varphi_n(X^{(n)}) \to \varphi(X)$ in $L^p(\Omega)$ and that $D\varphi_n(X^{(n)}) = \sum_{j=1}^m \partial_j\varphi_n(X^{(n)})DX_j^{(n)}$ converges to $\sum_{j=1}^m \partial_j\varphi(X)DX_j$ in $L^p(\Omega; H)$. By closedness of the Malliavin derivative, it follows that $\varphi(X) \in \mathbb{D}^{1,p}$ and that (1.1) holds. □

In applications, for example to prove that solutions to certain stochastic differential equations with Lipschitz coefficients belong to the domain of the Malliavin derivative, it is important to have a generalization of the chain rule for Lipschitz maps. To prove such a generalization, we use the following important

LEMMA 1.2.8. *Let $p \in (1, \infty)$ and $X_n \in \mathbb{D}^{1,p}$ be such that $X_n \to X$ in $L^p(\Omega)$ and such that*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\|DX_n\|_H^p < \infty.$$

Then $X \in \mathbb{D}^{1,p}$ and DX_n converges weakly to DX.

Proof. The sequence X_n, being convergent in L^p is bounded in L^p. It thus follows that X_n is bounded in $\mathbb{D}^{1,p}$. As bounded subsets of reflexive spaces are relatively weakly compact, we can extract a subsequence X_{n_k} which converges weakly (in $\mathbb{D}^{1,p}$) to some $Y \in \mathbb{D}^{1,p}$. In particular, X_{n_k} converges weakly in L^p to Y, whence $Y = X$. It follows that $X \in \mathbb{D}^{1,p}$. We can apply the above to any subsequence to see that every subsequence of X_n has a subsequence converging to X weakly in $\mathbb{D}^{1,p}$. It then follows that X_n converges weakly to X in $\mathbb{D}^{1,p}$. In particular, $DX_n \to DX$ weakly in $L^p(\Omega; H)$. □

Proposition 1.2.9. *Let $p \in (1, \infty)$ and $X = (X_1, \dots, X_m)$ be a vector with $X_j \in \mathbb{D}^{1,p}$ for $j = 1, \dots, m$ and let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be Lipschitz continuous with Lipschitz constant L. Then $\varphi(X) \in \mathbb{D}^{1,p}$. Moreover, there exists a random vector $Y = (Y_1, \dots, Y_m)$ with $|Y| \le L$ almost surely, such that*

$$D\varphi(X) = \sum_{j=1}^m Y_j DX_j\,.$$

Proof. Let φ_n be a sequence of continuously differentiable functions which converges pointwise to φ and such that $|\nabla \varphi_n| \le L$ (e.g. convolute with a mollifier). By Proposition 1.2.7, $\varphi_n(X) \in \mathbb{D}^{1,p}$ and $D\varphi_n(X) = \sum \partial_j \varphi_n(X) DX_j$. Clearly, $\varphi_n(X) \to \varphi(X)$ in L^p as $n \to \infty$. Moreover, the sequence $D\varphi_n(X)$ is bounded in $L^p(\Omega; H)$. Thus Lemma 1.2.8 implies that $\varphi(X) \in \mathbb{D}^{1,p}$ and that $D\varphi_n(X)$ converges weakly in $L^p(\Omega; H)$ to $D\varphi(X)$. Since the sequence $\nabla \varphi_n(X)$ is bounded by L, it has a subsequence which converges weakly in $L^p(\Omega, \mathbb{R}^m)$ to a random vector $Y = (Y_1, \dots, Y_m)$. It follows that Y is bounded by L. Moreover, $D\varphi(X) = \sum Y_j DX_j$. □

We next want to define also higher order Malliavin derivatives. Since the Malliavin derivative of a random variable is an H-valued random variable, we need to differentiate Hilbert space valued random variables. As it turns out, we need some abstract results which we establish before returning to our main line of study.

Intermezzo: Tensor Products and Hilbert-Schmidt Operators

Let H and V be Hilbert spaces and pick orthonormal bases $(h_\alpha)_{\alpha \in A}$ of H and $(v_\beta)_{\beta \in B}$ of V. We define the *(Hilbert space) tensor product* $H \otimes V$ to be the vector space of all formal series $\sum a_{\alpha,\beta} h_\alpha \otimes v_\beta$ where $a_{\alpha,\beta} \in \ell^2(A \times B)$, endowed with the inner product

$$\langle \sum a_{\alpha,\beta} h_\alpha \otimes v_\beta, \sum b_{\alpha.\beta} h_\alpha \otimes v_\beta \rangle := \sum a_{\alpha,\beta} b_{\alpha,\beta}\,.$$

Thus, $H \otimes V$, being isometrically isomorphic with $\ell^2(A \times B)$, is a Hilbert space and the system $(h_\alpha \otimes v_\beta)_{(\alpha,\beta) \in A \times B}$ is an orthonormal basis of $V \otimes H$.

Similar as for algebraic tensor products, we can characterize $H \otimes V$ by a universal property.

Consider the map $\rho : H \times V \to H \otimes V$, defined by

$$\rho\Big(\sum a_\alpha h_\alpha, \sum b_\beta v_\beta\Big) := \sum a_\alpha b_\beta h_\alpha \otimes b_\beta\,.$$

Note that if $(a_\alpha) \in \ell^2(A)$ and $(b_\beta) \in \ell^2(B)$, then $(a_\alpha b_\beta) \in \ell^2(A \times B)$ so that this definition makes sense. In fact, we have

$$\Big\| \sum a_\alpha b_\beta h_\alpha \otimes v_\beta \Big\|_{H \otimes V}^2 = \sum_{\alpha \in A, \beta \in B} a_\alpha^2 b_\beta^2 = \sum_{\alpha \in A} a_\alpha^2 \sum_{\beta \in B} b_\beta^2 = \Big\| \sum a_\alpha h_\alpha \Big\|_H^2 \Big\| \sum b_\beta v_\beta \Big\|_V^2$$

so that ρ is a continuous, bilinear form.

Lemma 1.2.10. *Let H, V be Hilbert spaces, $H \otimes V$ and ρ as above. Let also U be a Hilbert space and $\eta : H \times V \to U$ be a continuous bilinear form (recall, that η is called continuous if there is a constant C such that $\|\eta(h, v)\|_U \le C \|h\|_H \|v\|_V$). Then there exists a bounded linear operator $T_\eta : H \otimes V \to U$ such that $\eta = T_\eta \circ \rho$.*

PROOF. Given η as in the statement of the lemma, let $u_{\alpha,\beta} := \eta(h_\alpha, v_\beta)$. For the assertion to be true, we have to have $T_\eta(h_\alpha \otimes v_\beta) := u_{\alpha,\beta}$. We note that by continuity of η, the vectors $u_{\alpha,\beta}$ are uniformly bounded. Thus, given $(a_{\alpha,\beta}) \in \ell^2(A \times B)$, the series $\sum a_{\alpha,\beta} u_{\alpha,\beta}$ converges in U. This shows that there exists a unique bounded linear map T_η from $H \otimes V$ to U with $T_\eta(h_\alpha \otimes v_\beta) := u_{\alpha,\beta}$. This map clearly satisfies $\eta = T_\eta \circ \rho$. □

We can now use this universal property to identify some tensor products.

COROLLARY 1.2.11. *Let* $(\Omega_i, \Sigma_i, \mu_i)$ *for* $i = 1, 2$ *be* σ*-finite measure spaces. Then the tensor product* $L^2(\Omega_1, \Sigma_1, \mu_1) \otimes L^2(\Omega_2, \Sigma_2, \mu_2)$ *is isometrically isomorphic to the product space* $L^2(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \otimes \mu_2)$.

PROOF. For ease of notation, we suppress the σ-algebra and the measure from our notation in L^2 spaces.

Consider the bilinear form $\eta : L^2(\Omega_1) \times L^2(\Omega_2) \to L^2(\Omega_1 \times \Omega_2)$ given by $\eta(f, g) := [(x, y) \mapsto f(x)g(y))]$. By Fubini's theorem we have

$$\int_{\Omega_1 \times \Omega_2} \big(f(x)g(y)\big)^2 \, d(\mu_1 \otimes \mu_2)(x, y) = \int_{\Omega_1} f(x)^2 \, d\mu_1(x) \int_{\Omega_2} g(y)^2 \, d\mu_2(y)$$

so that $\|\eta(f, g)\| = \|f\|\|g\|$. In particular, η is continuous. By the universal property, there exists a linear operator $T_\eta : L^2(\Omega_1) \otimes L^2(\Omega_2) \to L^2(\Omega_1 \times \Omega_2)$ such that $\eta = T_\eta \circ \rho$. We thus have $T_\eta(f \otimes g) = [(x, y) \mapsto f(x)g(y)]$. In particular, the range of T_η contains indicator functions of rectangles $A \times B$ where A and B have finite measure. Using the continuity of T_η, a monotone class argument shows that the range of T_η contains the indicator of every set $S \in \Sigma_1 \otimes \Sigma_2$ with finite measure and hence, by linearity, any integrable simple function. By density and continuity, T_η is surjective.

Now let (e_α), resp. $(\tilde{e}_\beta)$ be orthonormal bases of $L^2(\Omega_1)$, resp. $L^2(\Omega_2)$. Then every element ξ of $L^2(\Omega_1) \otimes L^2(\Omega_2)$ can be written as $\xi = \sum a_{\alpha,\beta} e_\alpha \otimes \tilde{e}_\beta$. for some $(a_{\alpha,\beta}) \in \ell^2(A \times B)$. We have

$$\begin{aligned} \|T_\eta \xi\|^2_{L^2(\Omega_1, \times \Omega_2)} &= \int_{\Omega_1 \times \Omega_2} \Big(\sum a_{\alpha,\beta} e_\alpha(x) \tilde{e}_\beta(y)\Big)^2 d(\mu_1 \otimes \mu_2)(x, y) \\ &= \sum a_{\alpha,\beta} a_{\gamma,\delta} \int_{\Omega_1} e_\alpha(x) e_\gamma(x) \, d\mu_1(x) \int_{\Omega_2} \tilde{e}_\beta(y) \tilde{e}_\delta(y) \, d\mu_2(y) \\ &= \sum a^2_{\alpha,\beta} = \|\xi\|_{L^2(\Omega_1) \otimes L^2(\Omega_2)} \, . \end{aligned}$$

This proves that T is an isometry. In particular it is injective, hence, altogether, an isomorphism. □

Also a second identification for tensor products, this time even for abstract Hilbert spaces, is of importance. It arises from the bilinear form $\eta : H \times V \to \mathscr{L}(H, V)$ given by $\eta(h, v) := [g \mapsto \langle g, h \rangle_H \cdot v]$. Note that $\eta(h, v)$ if $h \neq 0$, then the range of $\eta(h, v)$ is the span of v. Linear combinations of such *"rank one operators"* are exactly the operators with finite dimensional range, the so-called *"finite rank operators"*. In the literature, one often finds the notation $h \otimes v$ for $\eta(h, v)$ which already suggests a connection with tensor products.

It should be noted, however, that there is no hope of identifying the tensor product $H \times V$ with all of $\mathscr{L}(H, V)$. Indeed, if that was the case, then every bounded operator from H to V could be approximated by finite rank operators, in particular it must be compact. For infinite dimensional Hilbert spaces, this is certainly not true. Thus to identify the tensor product $H \times V$ with a space of operators, we have to consider a suitable aubclass of operators, the so-called *Hilbert-Schmidt operators*.

DEFINITION 1.2.12. Let H, V be separable Hilbert spaces. An operator $T \in \mathscr{L}(H, V)$ is called *Hilbert-Schmidt operator* if for some (equivalently, all) orthonormal bases (e_n) of H

the sum $\sum_n \|Te_n\|_V^2$ is finite. For a Hilbert-Schmidt operator, we put

$$\|T\|_{\mathsf{HS}} := \Big(\sum_{n\in\mathbb{N}} \|Te_n\|_V^2\Big)^{\frac{1}{2}}.$$

To see that the definition of Hilbert-Schmidt operator (and of the value of $\|T\|_{\mathsf{HS}}$) is independent of the choice of the orthonormal basis, let (e_n) and (h_n) let orthonormal bases of H. Then, using Parseval's identity, we find

$$\begin{aligned}\sum_n \|Te_n\|_V^2 &= \sum_n \Big\|T\sum_m \langle e_n, h_m\rangle_H h_m\Big\|_V^2 = \sum_n\sum_m |\langle e_n, h_m\rangle_H|^2 \|Th_m\|_V^2\\ &= \sum_m \|h_m\|_H^2 \|Th_m\|_V^2 = \sum_m \|Th_m\|_V^2\,.\end{aligned}$$

Clearly, sums and scalar multiples of Hilbert-Schmidt operators are Hilbert-Schmidt operators. Thus the Hilbert-Schmidt Operators from H to V are a vector space, denoted by $\mathscr{L}_{\mathsf{HS}}(H,V)$. It is also easy to see that $\|\cdot\|_{\mathsf{HS}}$ is a norm on $\mathscr{L}_{\mathsf{HS}}(H,V)$.

Lemma 1.2.13. *Let H, V be separable Hilbert spaces. Then $H\otimes V$ is isometrically isomorphic to $\mathscr{L}_{\mathsf{HS}}(H,V)$.*

Proof. Consider the continuous bilinear form $\eta : H\times V \to \mathscr{L}(H,V)$, defined by $\eta(h,v) := \langle\cdot, h\rangle v$. By the universal property, there exists bounded linear operator $T_\eta : H\otimes V \to \mathscr{L}(H,V)$ such that $\eta = T_\eta\circ\rho$.

Now let (h_n) be an orthonormal basis of H and (v_n) be an orthonormal basis of V. Then $(h_n\otimes v_m)$ is an orthonormal basis of $H\otimes V$. If $\xi = \sum a_{nm} h_n\otimes v_m$ is an element of $H\otimes V$, then $T_\eta(\xi)\in\mathscr{L}_{\mathsf{HS}}(H,V)$. Indeed, for the orthonormal basis (h_k) we have

$$\sum_k \|T_\eta h_k\|_V^2 = \sum_k \Big\|\sum_{n,m} a_{nm}\langle h_k, h_n\rangle_H v_m\Big\|_V^2 = \sum_k\sum_m a_{km}^2 = \|\xi\|_{H\otimes V}^2\,.$$

This moreover shows that $\|T_\eta\xi\|_{\mathsf{HS}} = \|\xi\|_{H\otimes V}$. Thus T_η is an isometry. It remains to prove that the range of T_η is all of $\mathscr{L}_{\mathsf{HS}}(H,V)$. To that end, let $T\in\mathscr{L}_{\mathsf{HS}}(H,V)$ and let (h_n), resp. (v_n) be orthonormal bases of H resp. V. We can expand $Te_n = \sum_m a_{nm}v_m$, where $a_{nm} = \langle Te_n, h_m\rangle_V$. Note that

$$\sum_{nm} a_{nm}^2 = \sum_n \|Te_n\|_V^2 < \infty\,.$$

Thus $\xi := \sum a_{nm} e_n\otimes v_m$ belongs to $H\otimes V$. We then have

$$(T_\eta\xi)(e_k) = \sum_{n,m} a_{nm}\langle e_k, e_n\rangle_H v_m = \sum_m a_{km}v_m = Te_k\,.$$

Hence, by linearity and continuity, $T_\eta\xi = T$, i.e. $\mathrm{rg}T_\eta\mathscr{L}_{\mathsf{HS}}(H,V)$ as claimed. □

It is customary to drop the operator T_η and merely write $T = \sum a_{nm}h_n\otimes v_m$ in the above situation, where a_{nm} are the coefficients in the expansion of Te_n in the basis (v_m).

It follows from Lemma 1.2.13 that the Hilbert-Schmidt norm $\|\cdot\|_{\mathsf{HS}}$ is induced by an inner product. Indeed, if $T = \sum a_{nm}h_n\otimes v_m$ and $S = \sum b_{nm}h_n\otimes v_m$ belong to $\mathscr{L}_{\mathsf{HS}}(H,V)$ and we set

$$\langle T, S\rangle_{\mathsf{HS}} := \sum_{nm} a_{nm}b_{nm}\,.$$

This defines an inner product on $\mathscr{L}_{\mathsf{HS}}(H,V)$ which induces the norm $\|\cdot\|_{\mathsf{HS}}$. Note that

$$\langle T, S\rangle_{\mathsf{HS}} = \sum_n\sum_m a_{nm}b_{nm} = \sum_n \langle Te_n, Se_n\rangle_V = \sum_n \langle S^*Te_n, e_n\rangle_V =: \mathrm{tr}(S^*T)$$

The last sum is the so-called *trace* of the operator $S^*T\in\mathscr{L}(H,H)$. In fact, the trace can be define for a more general class of operators – the so-called *trace class operators* – and the above equation shows that if $S, T \in \mathscr{L}_{\mathsf{HS}}(H, V)$, then S^*T is of trace class. We do not go into details here but note that the value of $\mathrm{tr}(S^*T)$ does not depend on the choice of the orthonormal basis (e_n). Indeed, if (h_n) is an orthonormal basis of H, then expanding h_n in the basis (e_m), we find

$$\begin{aligned}\sum_n \langle Th_n, Sh_n\rangle_V &= \sum_n\sum_k\sum_k \langle h_n, e_k\rangle_H \langle h_n, e_l\rangle_H \langle Te_k, Se_l\rangle_V\\ &= \sum_k\sum_l \langle e_k, e_l\rangle_H \langle Te_k, Se_l\rangle_V\\ &= \sum_k \langle Te_k, Se_k\rangle_V \,.\end{aligned}$$

This in particular shows that the norm in the Hilbert space $H \otimes V$ does not depend on the choice of bases of H and V, as it seems a priori.

We now return to our discussion of Malliavin derivatives and extend the definition of the Malliavin derivative to Hilbert space valued random variables. Besides our H-isonormal Gaussian process W we are given a (separable) Hilbert space V. In this setting, we have the following generalization of the Wiener chaos decomposition.

We let $\mathscr{H}_n(V)$ denote the closed subspace of $L^2(\Omega; V)$ generated by random variables of the form $\sum_{j=1}^m X_j v_j$, where $X_j \in \mathscr{H}_n$ and $v_j \in V$. This space can be canonically identified with the space $\mathscr{H}_n \otimes V$.

We now have the following generalization of Theorem 1.1.11. The proof is left to the reader.

THEOREM 1.2.14. *Let W be an H-isonormal Gaussian process and V be a separable Hilbert space. Then*

$$L^2(\Omega, \Sigma_W, \mathbb{P}; V) = \bigotimes_{n=0}^{\infty} \mathscr{H}_n(V)$$

and this decomposition is orthogonal. Moreover, if $(v_k)_{k\in K}$ is an orthonormal basis of V, then $\{\Phi_\alpha \otimes v_k : |\alpha| = n, k \in K\}$ is an orthonormal basis of $\mathscr{H}_n(V)$.

In order to generalize the definition of the Malliavin derivative, we consider the space $\mathscr{S}(V)$, consisting of V-valued random vectors X of the form

$$X = \sum_{j=1}^m Y_j v_j$$

where $Y_j \in \mathscr{S}$, thus $Y_j = f_j(W(h_1^j), \dots, W(h_{n_j}^j))$ for certain $n_j \in \mathbb{N}$, $f_j \in C_p^\infty(\mathbb{R}^{n_j}))$ and $h_1^j, \dots, h_{n_j}^j \in H$, and $v_j \in V$.

DEFINITION 1.2.15. For $X \in \mathscr{S}(V)$ as above, the Malliavin derivative of X is the $H\otimes V$-valued random variable (alternatively, the $\mathscr{L}_{\mathsf{HS}}(H, V)$-valued random variable) DX, defined by

$$DX = \sum_{j=1}^m \sum_{i=1}^{n_j} \big[\partial_i f_j(W(h_1^j), \dots, W(h_{n_j}^j)\big] h_i^j \otimes v_j = \sum_{j=1}^m \big(DX_j\big) \otimes v_j \,.$$

Here the tensors in the first sum are elements in $H \otimes V$, the tensors in the second sum are elements in $L^2(\Omega; H) \otimes V$.

We note that in the case where $V = \mathbb{R}$, i.e. in the scalar situation considered so far, the tensor $H \otimes V = H \otimes \mathbb{R}$ is isometrically isomorphic to H and the vector valued definition coincides with the one given above.

Expanding the vectors v_j in an orthonormal basis and testing against the elements of this basis, it follows from Lemma 1.2.3 that the definition is independent of the representation of the random vector X. Similar as in the scalar case, one shows that the Malliavin derivative is closable as an operator from $L^p(\Omega;V)$ to $L^p(\Omega;H\otimes V)$. We leave the details to the reader. The domain of the closure is denoted by $\mathbb{D}^{1,p}(V)$.

We can now also define higher order Malliavin derivatives. Indeed, if $X\in\mathbb{D}^{1,p}$ is such that the H-valued random variable DX belongs to $\mathbb{D}^{1,p}(H)$, then we say that $X\in\mathbb{D}^{2,p}$ and $D^2X := D(DX)$. Thus, the second derivative is an $H\otimes H$-valued random variable or, alternatively, and $\mathscr{L}_{\mathsf{HS}}(H)$-valued random variable.

EXAMPLE 1.2.16. Let us consider again the situation of Brownian motion where $H = L^2((0,\infty))$ and $B_t = W(\mathbb{1}_{(0,t]})$. We have seen that $X = B_t^2$ belongs to $\mathbb{D}^{1,2}$ and $DX = 2B_t\mathbb{1}_{(0,t]}$ Note that $DX\in\mathscr{S}(H)$. We have $D^2X = (D2B_t)\otimes\mathbb{1}_{(0,t]} = 2\mathbb{1}_{(0,t]}\otimes\mathbb{1}_{(0,t]}$.

Let us write $H^{\otimes n}$ for the n-fold tensor product of H with itself. We define inductively

$$\mathbb{D}^{k,p} := \{X\in\mathbb{D}^{k-1,p} : D^{k-1}X\in\mathbb{D}^{1,p}(H^{\otimes(n-1)})\}\quad\text{and}\quad D^kX := D(D^{k-1}X).$$

$\mathbb{D}^{k,p}$ is a Banach space with respect to the norm

$$\|X\|_{k,p} := \Big(\mathbb{E}|X|^p + \sum_{j=1}^k \mathbb{E}\|D^jX\|^p_{H^{\otimes j}}\Big)^{\frac{1}{p}}.$$

For $p=2$ and $k\geq 1$, the space $\mathbb{D}^{k,2}$ is a Hilbert space with respect to the inner product

$$\langle X,Y\rangle_{k,2} = \mathbb{E}(XY) + \sum_{j=1}^k \mathbb{E}\big\langle D^jX, D^jY\big\rangle_{H^{\otimes j}}.$$

We next characterize the Malliavin derivative in terms of the Wiener chaos decomposition. We denote by J_n the orthogonal projection onto the n-th Wiener chaos (both in the scalar and the vector-valued situation).

PROPOSITION 1.2.17. *Let $X\in L^2(\Omega)$ have the chaos decomposition $X = \sum_{n=0}^\infty J_nX$. Then $X\in\mathbb{D}^{1,2}$ if and only if $\sum_{n=0}^\infty n\|J_nX\|_2^2<\infty$. In this case $DJ_nX = J_{n-1}DX$ and we have*

$$\mathbb{E}\|DX\|_H^2 = \sum_{n=0}^\infty n\|J_nX\|_2^2$$

PROOF. Pick an orthonormal basis (e_n) of H and consider the random variables Φ_α from Proposition 1.1.15, i.e. for $\alpha\in\Lambda$ we have

$$\Phi_\alpha = \sqrt{\alpha!}\prod_{j=1}^\infty H_{\alpha_j}(W(e_j)).$$

We have $\Phi_\alpha\in\mathscr{S}$ and using that $H_n' = H_{n-1}$, we find

$$D\Phi_\alpha = \sqrt{\alpha!}\sum_{l=1}^\infty\prod_{j\neq l,l=0}^\infty H_{\alpha_j}(W(e_n))H_{\alpha_l-1}(W(e_l))e_l$$

Note that if $|\alpha| = \sum\alpha_j = n$, then for fixed l the random variable

$$\sqrt{\alpha!}\prod_{j\neq l,l=0}^\infty H_{\alpha_j}(W(e_n))H_{\alpha_l-1}(W(e_l))$$

is equal to $\sqrt{\alpha_l}\Phi_{\beta^{(l)}}$ where $\beta_j^{(l)} = \alpha_j$ for $j\neq l$ and $\beta_l^{(l)} = \alpha_l - 1$. Therfore, $D\Phi_\alpha = \sum_{l=1}^\infty\sqrt{\alpha_l}\Phi_{\beta^{(l)}}$.

Note that $|\beta| = n-1$ whence $D\Phi_\alpha \in \mathscr{H}_{n-1}(H)$. Moreover,

$$\mathbb{E}\|D\Phi_\alpha\|_H^2 = \sum_{l=0}^\infty \alpha_l = |\alpha| = n$$

Since $\{\Phi_\alpha : |\alpha| = n\}$ is an orthonormal basis of $\mathscr{H}_n$ by Proposition 1.1.15, it follows from linearity and closedness of D that $\mathscr{H}_n \subset \mathbb{D}^{1,2}$ and $D\mathscr{H}_n \subset \mathscr{H}_{n-1}(H)$, where we set $\mathscr{H}_{-1}(H) = \{0\}$. Moreover, $\mathbb{E}\|DY\|_H^2 = n\mathbb{E}|Y|^2$ for $Y \in \mathscr{H}_n$.

Now assume that $\sum_{n=0}^\infty n\|J_nX\|_2^2 < \infty$. By the Wiener chaos decomposition, the sequence $X_N := \sum_{n=0}^N J_nX$ converges in $L^2(\Omega)$ to X. Moreover, by the above,

$$\mathbb{E}\|DX_N\|_H^2 = \sum_{n=0}^N n\|J_nX\|_2^2$$

which is bounded by assumption. By Lemma 1.2.8, $X \in \mathbb{D}^{1,2}$ and $DX = \sum_{n=0}^\infty DJ_nX$.

Conversely, let $X \in \mathbb{D}^{1,2}$. By Theorem 1.2.14, $\mathscr{O} = \{\Phi_\alpha \otimes e_k : \alpha \in \Lambda, k \in \mathbb{N}\}$ is an orthonormal basis for $L^2(\Omega; H)$. Expanding DX in this basis and X in the orthonormal basis Φ_α, the above computation shows that $DJ_nX = J_{n-1}DX$ and that $\mathbb{E}\|DX\|_H^2 = \sum_{n=0}^\infty \|J_nX\|_2^2 < \infty$. □

As a corollary, we see that a random variable whose Malliavin derivative is zero, is almost surely constant.

Corollary 1.2.18. *Let $X \in \mathbb{D}^{1,2}$ be such that $DX = 0$ a.e. Then $X = \mathbb{E}X$ a.e.*

Proof. By Proposition 1.2.17, $J_nX = 0$ for all $n \geq 1$. Thus $X = J_0X = \mathbb{E}X$ almost surely. □

Corollary 1.2.19. *Let $A \in \Sigma$. Then $\mathbb{1}_A \in \mathbb{D}^{1,2}$ if and only if $\mathbb{P}(A) \in \{0,1\}$.*

Proof. If $\mathbb{P}(A) \in \{0,1\}$, then $\mathbb{1}_A$ is almost surely constant, hence $\mathbb{1}_A \in \mathscr{H}_0 \subset \mathbb{D}^{1,2}$.

Conversely assume that $\mathbb{1}_A \in \mathbb{D}^{1,2}$. Pick a function $\varphi \in C_c^\infty(\mathbb{R})$ with $\varphi(t) = t^2$ for $t \in (-2,2)$. Then φ has bounded derivative. Noting that $\varphi(\mathbb{1}_A) = \mathbb{1}_A$ and $\varphi'(t) = 2t$ for $t \in (-2,2)$, it follows from the chain rule (Proposition 1.2.7) that

$$D\mathbb{1}_A = D\varphi\big(\mathbb{1}_A\big) = 2\mathbb{1}_A D\mathbb{1}_A\,.$$

Thus $D\mathbb{1}_A = 0$ a.e. on A^c and $D\mathbb{1}_A = 2\mathbb{1}_A$, a.e. on A. Altogether, $D\mathbb{1}_A = 0$ a.e. and hence, by Corollary 1.2.18, $\mathbb{1}_A = \mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$ almost surely, which implies $\mathbb{P}(A) \in \{0,1\}$. □

1.3. The Divergence Operator

We next define the divergence operator as the adjoint of the Malliavin derivative. As the Malliavin derivative maps $L^2(\Omega)$ to $L^2(\Omega; H)$, its adjoint will be an operator on $L^2(\Omega; H)$ taking values in $L^2(\Omega)$.

Definition 1.3.1. The *divergence operator*, sometimes also called the *Skorohod integral*, is the adjoint δ of the Malliavin derivative D on $L^2(\Omega)$. More precisely, the domain $D(\delta)$ consists of those $u \in L^2(\Omega; H)$ such that there exists $X \in L^2(\Omega)$ with

$$\mathbb{E}\langle DY, u\rangle_H = \mathbb{E}\big(Y \cdot X\big)$$

for all $Y \in \mathbb{D}^{1,2}$. Since $\mathbb{D}^{1,2}$ is dense in $L^2(\Omega)$, there is at most one such element X. We write $\delta(u) := X$.

Note that an element u of $L^2(\Omega; H)$ belongs to the domain $D(\delta)$ if and only if there exists a constant $c \geq 0$ such that

$$\big|\mathbb{E}\langle DY, u\rangle\big| \leq c\big(\mathbb{E}|Y|^2\big)^{\frac{1}{2}}$$

for all $X \in \mathbb{D}^{1,2}$. Indeed, in this case, the map $\varphi(Y) := \mathbb{E}\langle DY, h\rangle$ is a linear functional on $\mathbb{D}^{1,2}$ which is bounded with respect to the 2-norm. It thus has a unique bounded extension to all of $L^2(\Omega)$. By the Riesz-Fischer theorem, this extension is of the form $Y \mapsto \mathbb{E}(YX)$ for a certain $X \in L^2(\Omega)$.

EXAMPLE 1.3.2. Let $u \in \mathscr{S}(H)$, say $u = \sum_{j=1}^n X_j h_j$. Then $X \in D(\delta)$ and

$$\delta(u) = \sum_{j=1}^n X_j W(h_j) - \sum_{j=1}^n \langle DX_j, h_j\rangle \tag{1.2}$$

To see this, we may (by linearity) assume that $n = 1$ and write $u = Xh$ for simplicity. Then the statement reduces to equality

$$\mathbb{E}\langle DY, u\rangle = \mathbb{E}X\langle DY, h\rangle \stackrel{!}{=} \mathbb{E}Y\big(XW(h) - \langle X, h\rangle\big)$$

for all $Y \in \mathbb{D}^{1,2}$. This integration-by-parts formula was seen to be true for $Y \in \mathscr{S}$ in Corollary 1.2.5, that it is true for $Y \in \mathbb{D}^{1,2}$ follows by an approximation argument.

We next prove that a large class of H-valued processes u, namely those in $\mathbb{D}^{1,2}(H)$, belong to the domain of δ. Let us note that if $u \in \mathbb{D}^{1,2}$, then the derivative Du takes values in $H \otimes H \eqsim \mathscr{L}_{\mathsf{HS}}(H, H)$ so that the derivative is operator valued.

We first prove a "commutator relationship" between the Malliavin derivative and the divergence.

LEMMA 1.3.3. *Let $u \in \mathscr{S}(H)$ and $h \in H$. Then*

$$\langle D\delta(u), h\rangle - \delta(Du \cdot h) = \langle u, h\rangle \tag{1.3}$$

Given $h \in H$ we may define the directional derivative $D_h X$ of a random variable X as $D_h X := \langle DX, h\rangle$. For an H-valued random variable, the directional derivative can be defined as $D_h u := Du \cdot h$, i.e. the image of h under the map $Du \in \mathscr{L}_{\mathsf{HS}}$.

Doing so, equation (1.3) can be rephrased as $D_h\delta - \delta D_h = \langle \cdot, h\rangle$ which is some sort of Heisenberg commutator relationship ($AB - BA = id$) except that we have to test against an h (and that the first D_h is actually different from the second).

PROOF OF LEMMA 1.3.3. Let $u = Xg$ where $X = f(W(e_1), \ldots, W(e_n)) \in \mathscr{S}$ and $g \in H$. We assume without loss of generality that $e_1, \ldots, e_n$ is an orthonormal system with $g, h \in \operatorname{span}\{e_1, \ldots, e_n\}$, cf. the proof of Lemma 1.2.3. The general case follows from linearity. In what follows we write $X_j := (\partial_j f)(W(e_1), \ldots, W(e_n))$.

By (1.2), $\delta u = XW(g) - \langle DX, g\rangle$. We note that $\langle DX, g\rangle = \sum_{j=1}^n X_j\langle e_j, g\rangle$. Let us also note that $XW(g) = \tilde f(W(e_1), \ldots W(e_n), W(g))$ where $\tilde f(x, x_{n+1}) = x_{n+1}f(x)$ so that $D(XW(g)) = Xg + \sum_{j=1}^n W(g)X_j e_j$. It follows from the definition of D that

$$\langle D\delta(u), h\rangle = \langle u, h\rangle + \sum_{j=1}^n W(g)X_j\langle e_j, h\rangle - \sum_{j=1}^n \langle e_j, g\rangle\langle DX_j, h\rangle \tag{1.4}$$

On the other hand, $Du = \sum_{j=1}^n X_j e_j \otimes g$ so that $Du \cdot h = \sum_{j=1}^n X_j\langle h, e_j\rangle g$. Thus (1.2) yields

$$\delta(Du \cdot h) = \sum_{j=1}^n \langle h, e_j\rangle X_j W(g) - \sum_{j=1}^n \langle h, e_j\rangle\langle DX_j, g\rangle \tag{1.5}$$

Expanding h and g in the orthonormal basis $e_1, \ldots, e_n$, we find

$$\begin{aligned}\sum_{j=1}^n \langle e_j, g\rangle\langle DX_j, h\rangle &= \sum_{j=1}^n\sum_{i=1}^n \langle e_j, g\rangle\langle h, e_i\rangle\langle DX_j, e_i\rangle\\ &= \sum_{j=1}^n\sum_{i=1}^n \langle e_j, g\rangle\langle h, e_i\rangle\langle DX_i, e_j\rangle = \sum_{i=1}^n \langle h, e_i\rangle\langle DX_i, g\rangle,\end{aligned}$$

where we have used in the second equality that $\langle DX_j, e_i\rangle = \langle DX_i, e_j\rangle$ as a consequence of the Schwarz theorem on mixed partial derivatives. Combining this with (1.4) and (1.5), equation (1.3) is proved. □

We can now prove

PROPOSITION 1.3.4. *$\mathbb{D}^{1,2}(H) \subset D(\delta)$. Moreover, for $u, v \in \mathbb{D}^{1,2}(H)$ we have*

$$\mathbb{E}\big(\delta(u)\delta(v)\big) = \mathbb{E}\langle u, v\rangle + \mathbb{E}\big[\mathrm{tr}(DuDv)\big]\,. \tag{1.6}$$

PROOF. Let us first assume that $u, v \in \mathscr{S}(H) \subset D(\delta)$. We let (e_n) be an orthonormal basis of H. Using the duality between δ and D and the Parseval identity, we have

$$\mathbb{E}\big(\delta(u)\delta(v)\big) = \mathbb{E}\langle u, D\delta(v)\rangle = \sum_{k=1}^{\infty} \mathbb{E}\langle u, e_k\rangle\langle D\delta(v), e_k\rangle\,.$$

By Lemma 1.3.3 the latter equals

$$\sum_{k=1}^{\infty} \mathbb{E}\big(\langle u, e_k\rangle\big(\langle v, e_k\rangle + \delta(Dv\cdot e_k)\big)\big) = \mathbb{E}\langle u, v\rangle + \sum_{k=1}^{\infty}\mathbb{E}\langle u, e_k\rangle\delta(Dv\cdot e_k)$$

Using the duality between D and δ we find for $k \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}\big(\langle u, e_k\rangle\delta(Dv\cdot e_k)\big) &= \mathbb{E}\big\langle D\langle u, e_k\rangle, (Dv\cdot e_k)\big\rangle\\ &= \mathbb{E}\langle Du^*\cdot e_k, Dv\cdot e_k\rangle\end{aligned}$$

where we have used that $D\langle u, h\rangle = Du^*\cdot h$ which was an exercise. Thus we obtain

$$\begin{aligned}\mathbb{E}\big(\delta(u)\delta(v)\big) &= \mathbb{E}\langle u, v\rangle + \mathbb{E}\sum_{k=1}^{\infty}\langle Du^*\cdot e_k, Dv\cdot e_k\rangle\\ &= \mathbb{E}\langle u, v\rangle + \mathbb{E}\big[\mathrm{tr}(DuDv)\big]\end{aligned}$$

This establishes (1.6) for $u, v \in \mathscr{S}(H)$. For $u = v \in \mathscr{S}(H)$ we have in particular,

$$\mathbb{E}|\delta(u)|^2 = \mathbb{E}\|u\|_H^2 + \mathbb{E}\|Du\|_{H\otimes H}^2 =: \|u\|_{\mathbb{D}^{1,2}(H)}^2\,.$$

Given $u \in \mathbb{D}^{1,2}(H)$, there exists a sequence u_n in $\mathscr{S}(H)$ such that $u_n \to u$ in $L^2(\Omega; H)$ and $Du_n \to Du$ in $L^2(\Omega; H\otimes H)$. The above inequality shows that $\delta(u_n)$ is a Cauchy sequence in $L^2(\Omega)$, hence convergent to, say, X.

Now observe that for $Y \in \mathbb{D}^{1,2}$ we have

$$\mathbb{E}\langle DY, u\rangle = \lim_{n\to\infty}\langle DY, u_n\rangle = \lim_{n\to\infty}\mathbb{E}\big(Y\delta(u_n)\big) = \mathbb{E}\big(YX\big)\,.$$

This implies $u \in D(\delta)$ and $\delta(u) = X$. It also implies (1.6) for $u = v$. The general case follows by polarization. □

Our next result allows us to factor out a scalar random variable in a divergence.

PROPOSITION 1.3.5. *Let $X \in \mathbb{D}^{1,2}$ and $u \in D(\delta)$ be such that $Xu \in L^2(\Omega; H)$ and such that $X\delta(u) - \langle DX, u\rangle \in L^2(\Omega)$. Then $Xu \in D(\delta)$ and*

$$\delta(Xu) = X\delta(u) - \langle DX, u\rangle\,.$$

PROOF. First, let $X \in \mathscr{S}$. By the duality between D and δ and calculus, we find for $Y \in \mathscr{S}_b$ that

$$\begin{aligned}\mathbb{E}\langle DY, Xu\rangle = \mathbb{E}\langle u, XDY\rangle &= \mathbb{E}\langle u, D(XY) - YDX\rangle\\ &= \mathbb{E}\big(\delta(u)XY - Y\langle u, DX\rangle\big)\\ &= \mathbb{E}\big(Y\big(X\delta(u) - \langle DX, u\rangle\big)\big)\,.\end{aligned}$$

Using that $\mathscr{S}$ is dense in $\mathbb{D}^{1,2}$, it follows that the above is true for $X \in \mathbb{D}^{1,2}$, $u \in D(\delta)$ and $Y \in \mathscr{S}_b$. Using the additional assumptions on X and u, we see that it then also holds for $Y \in \mathbb{D}^{1,2}$. This yields the claim. □

We end this section by showing that both the Malliavin derivative and the Skorohod integral are *local operators*. Here we call an operator T defined on a space of random variables *local* if $X = 0$ a.e. on a set $A \in \Sigma$ implies that also $TX = 0$ a.e. on A.

PROPOSITION 1.3.6. *The divergence operator δ is local on $\mathbb{D}^{1,2}(H)$.*

PROOF. It suffices to show that $\delta(u)\mathbb{1}_{\{\|u\|=0\}} = 0$ almost everywhere.

Let $X \in \mathscr{S}_c$ and $\varphi \in C_c^\infty(\mathbb{R})$ with $\mathbb{1}_{(-1,1)} \le \varphi \le \mathbb{1}_{(-2,2)}$. Put $\varphi_n(t) := \varphi(nt)$ so that $\varphi_n(t) \to \mathbb{1}_{\{0\}}$ pointwise and $\varphi_n'(t) \to \infty \cdot \mathbb{1}_{\{0\}}$ pointwise. Note that $\sup_{t\in\mathbb{R}} |t\varphi_n'(t)| = \sup_{t\in\mathbb{R}} |nt\varphi'(nt)| \le 2\|\varphi'\|_\infty < \infty$.

It is an exercise to prove that $X\varphi_n(\|u\|^2) \in \mathbb{D}^{1,2}$ with

$$D\big(X\varphi_n(\|u\|^2)\big) = \varphi_n(\|u\|^2)DX + 2X\varphi_n'(\|u\|^2)Du \cdot u\,.$$

This together with the duality between D and δ yields

$$\mathbb{E}\big(\delta(u)X\varphi_n(\|u\|^2)\big) = \mathbb{E}\big(\varphi_n(\|u\|^2)\langle u, DX\rangle\big) + 2\mathbb{E}\big(X\varphi_n'(\|u\|^2)\langle Du\cdot u, u\rangle\big)\,.$$

Observe that $\varphi_n(\|u\|^2)\langle u, DX\rangle \to \mathbb{1}_{\{0\}}(\|u\|^2)\langle u, DX\rangle = 0$ as $n \to \infty$ and that

$$\big|\varphi_n(\|u\|^2)\langle u, DX\rangle\big| \le \|\varphi\|_\infty \|u\|_H \|DX\|_H$$

and the latter is integrable. Thus, by dominated convergence, $\mathbb{E}\big(\varphi_n(\|u\|^2)\langle u, DX\rangle\big) \to 0$. Similarly, $X\varphi_n'(\|u\|^2)\langle Du\cdot u, u\rangle \to 0$ as $n \to \infty$ and

$$\big|X\varphi_n'(\|u\|^2)\langle Du\cdot u, u\rangle\big| \le |X||\varphi_n'(\|u\|^2)\|Du\|_{H\otimes H}\|u\|^2 \le 2|X|\|\varphi'\|_\infty\|Du\|_{H\otimes H}$$

which is integrable. Altogether the dominated convergence theorem yields

$$\mathbb{E}\big(\delta(u)\mathbb{1}_{\{\|u\|=0\}}X\big) = 0$$

which finishes the proof. □

We next proof that also the Malliavin derivative is local. Note that the proof uses the divergence operator.

PROPOSITION 1.3.7. *The Malliavin derivative is a local operator on $\mathbb{D}^{1,1}$.*

PROOF. Let $X \in \mathbb{D}^{1,1}$. Replacing X with $\arctan X$ if necessary, we may and shall assume that $X \in L^\infty$ additionally. Let φ, φ_n be as in the proof of Proposition 1.3.6 and put $\psi_n(t) := \int_{-\infty}^t \varphi_n(r)\,dr$. Note that $\|\psi_n\|_\infty \le n^{-1}\|\varphi\|_1$ By the chain rule, $\psi_n(X) \in \mathbb{D}^{1,1}$ and $D\psi_n(X) = \varphi_n(X)DX$.

Now let $u \in \mathscr{S}_b(H)$. We note that $\mathbb{E}Y\delta(u) = \mathbb{E}\langle DY, u\rangle$ holds even for $Y \in \mathbb{D}^{1,1}\cap L^\infty$ and $u \in \mathscr{S}_b(H)$ as is easy to see using approximation. Hence, we find

$$\big|\mathbb{E}\varphi_n(X)\langle DX, u\rangle\big| = \big|\mathbb{E}\langle D(\psi_n(X)), u\rangle\big| = \big|\mathbb{E}\psi_n(X)\delta(u)\big| \le \frac{1}{n}\|\varphi\|_1\mathbb{E}|\delta(u)| \to 0$$

as $n \to \infty$. It follows that $\mathbb{E}\mathbb{1}_{\{|X|=0\}}\langle DX, u\rangle = 0$ and thus, since $\mathscr{S}_b(H)$ separates $L^1(\Omega; H)$, $\mathbb{1}_{\{|X|=0\}}DX = 0$ almost surely which implies the claim. □

Since δ and D are local operators, they can be defined on larger domains by localization. This is done as follows. If $\mathscr{R}$ is a vector space of (real or Hilbert space valued) random variables then we say that $\xi \in \mathscr{R}_{\text{loc}}$ if there exists a sequence $((A_n, \xi_n))_{n\in\mathbb{N}} \subset \Sigma \times \mathscr{R}$ such that

(1) $A_n \uparrow \Omega$ and
(2) $\xi = \xi_n$ on A_n.

Then, given a local operator T on $\mathscr{R}$, we can define $T\xi$ by setting $T\xi = T\xi_n$ on A_n

Proceeding in this manner, we extend the Malliavin derivative to the spaces $\mathbb{D}^{1,p}_{\text{loc}}$ and extend the divergence δ to the spaces $\mathbb{D}^{1,2}_{\text{loc}}(H)$.

1.4. The Ornstein-Uhlenbeck Semigroup

DEFINITION 1.4.1. Let J_n denote the projection on $L^2(\Omega)$ onto the n-th Wiener chaos. The *Ornstein-Uhlenbeck semigroup* is the one-parameter semigroup $(T(t))_{t\geq 0}$ defined by

$$T(t)X := \sum_{n=0}^{\infty} e^{-nt} J_n X\,.$$

From this definition, the following properties are easy to see:

LEMMA 1.4.2. *The Ornstein-Uhlenbeck semigroup satisfies the following properties:*
(1) $T(t) \in \mathscr{L}(L^2(\Omega))$ *and* $T(t)$ *is selfadjoint for all* $t \geq 0$*;*
(2) $T(0) = id_{L^2(\Omega)}$ *and* $T(t+s) = T(t)T(s)$*;*
(3) *For* $X \in L^2(\Omega)$*, the orbit* $t \mapsto T(t)X$ *is continuous.*

Lemma 1.4.2 can be summarized by saying that the Ornstein-Uhlenbeck semigroup is a strongly continuous semigroup of selfadjoint operators. We recall that the generator of a strongly continuous semigroup $(S(t))_{t\geq 0}$ on a Banach space E is the (in general unbounded) operator A, defined by

$$D(A) := \{x \in E : \lim_{t\to 0} t^{-1}(T(t)x - x) \text{ exists}\} \quad \text{and} \quad Ax := \lim_{t\to 0} t^{-1}(T(t)x - x)\,.$$

We can now characterize the generator L of the Ornstein-Uhlenbeck semigroup – the *Ornstein-Uhlenbeck operator* – in terms of the Wiener chaos decomposition.

PROPOSITION 1.4.3. *The generator* L *of the Ornstein-Uhlenbeck semigroup* $(T(t))_{t\geq 0}$ *is the operator given by* $LX := \sum_{n=0}^{\infty} -nJ_nX$ *for* $X \in D(L)$ *where*

$$D(L) := \big\{X \in L^2(\Omega) : \sum_{n=0}^{\infty} n^2 \|J_nX\|_2^2 < \infty\big\}\,.$$

PROOF. For the moment denote the generator of the Ornstein-Uhlenbeck operator by A and let L be the operator in the statement of the proposition.

First assume that $X \in D(A)$ with $AX = Y$, i.e. $\lim_{n\to\infty} t^{-1}(T(t)X - X)$ exists and equals Y. Using the continuity of the projection J_n and the fact that J_n commutes with $T(t)$ for all $t \geq 0$, we find

$$J_nY = \lim_{t\to 0} J_n \frac{T(t)X - X}{t} = \lim_{t\to 0} \frac{T(t)J_nX - J_nX}{t} = \lim_{t\to 0} \frac{e^{-nt} - 1}{t} J_nX = -nJ_nX\,.$$

Since $\sum_n \|J_nY\|^2 < \infty$ it follows that $X \in D(L)$ and $LX = Y$.

Conversely, let $X \in D(L)$. We have

$$\mathbb{E}\Big|\frac{(T(t)X - X))}{t} - LX\Big|^2 = \sum_{n=0}^{\infty} \Big(\frac{e^{-nt} - 1}{t} + n\Big)^2 \mathbb{E}|J_nX|^2$$

Each summand in the latter sum converges to 0 as $t \to 0$. Moreover, $|t^{-1}(e^{-nt} - 1)| \leq ne$ for all $t \leq 1$. Thus by dominated convergence the sum converges to 0 as $t \to 0$ implying that $X \in D(A)$ and $AX = LX$. □

The next proposition clarifies the relation between the operators L, D and δ.

PROPOSITION 1.4.4. *We have* $L = -\delta D$*, i.e.* $X \in D(L)$ *if and only if* $X \in \mathbb{D}^{1,2}$ *and* $DX \in D(\delta)$*; in that case* $\delta(DX) = -LX$*.*

PROOF. If $X \in \mathbb{D}^{1,2}$ with $DX \in D(\delta)$, then, by the duality between D and δ, we have

$$\mathbb{E}\big(Y\delta(DX)\big) = \mathbb{E}\langle DY, DX\rangle$$

for all $Y \in \mathbb{D}^{1,2}$.

By Proposition 1.2.17 and polarization $\mathbb{E}\langle J_{n-1}DX, J_{n-1}DY\rangle = n\mathbb{E}(J_nXJ_nY)$. Thus, expanding into the Wiener chaoses, we find

$$\mathbb{E}\big(Y\delta(DX)\big) = \sum_{n=0}^{\infty} n\mathbb{E}(J_nYJ_nX) = \sum_{n=0}^{\infty} n\mathbb{E}(YJ_nX)$$

It follows that $J_n(\delta(DX)) = nJ_n(DX)$. Indeed, if we pick $Y \in \mathscr{H}_m$ for $m \neq 0$, then the above implies

$$\mathbb{E}(Y(\delta(DX) - nJ_nX) = \sum_{j\neq n} n\mathbb{E}(YJ_jX) = 0\,.$$

By linearity and continuity, this remains true for $Y \in \mathscr{H}_n^\perp = \bigoplus_{m\neq n}\mathscr{H}_m$. It now follows that

$$\sum_{n=0}^{\infty} n^2\|J_nX\|^2 = \sum_{n=0}^{\infty} \|J_n\delta(DX)\|^2 < \infty$$

and hence $X \in D(L)$. Moreover, continuing the above calculation, we find

$$\mathbb{E}\big(Y\delta(DX)\big) = \sum_{n=0}^{\infty} n\mathbb{E}(J_nYJ_nX) = \sum_{n=0}^{\infty} n\mathbb{E}(YJ_nX) = \mathbb{E}Y(-LX)$$

for all $Y \in \mathbb{D}^{1,2}$. The equality $\delta(DX) = -LX$ follows from the density of $\mathbb{D}^{1,2}$ in $L^2(\Omega)$.

Conversely, let $X \in D(L)$. Then $\sum_{n=0}^\infty n^2\|J_nX\|_2^2 < \infty$ and hence $\sum_{n=0}^\infty n\|J_nX\|_2^2 < \infty$ which implies $X \in \mathbb{D}^{1,2}$ by Proposition 1.2.17. The computations above yield

$$\mathbb{E}\big[Y(-LX)\big] = \mathbb{E}\langle DY, DX\rangle$$

for all $Y \in \mathbb{D}^{1,2}$. By definition, this means that $DX \in D(\delta)$ and $\delta(DX) = -LX$. □

Remark 1.4.5. The proof of Proposition 1.4.4 shows that the Ornstein-Uhlenbeck operator can also be defined via quadratic forms. Indeed, if we define $\mathfrak{a}[X,Y] := \mathbb{E}\langle DX, DY\rangle$, then $X \in \mathbb{D}^{1,2}$ is such that there exists $Z \in L^2(\Omega)$ with $\mathbb{E}YZ = \mathfrak{a}(X,Y)$ for all $Y \in \mathbb{D}^{1,2}$, then $X \in D(L)$ and $LX = -Z$.

Note that the symmetric form $1 + \mathfrak{a}$ is the inner product of the Hilbert space $\mathbb{D}^{1,2}$.

In the applications of Malliavin calculus that we will present, the main operators of interest are the Malliavin derivative and the Skorohod integral. However, the Ornstein-Uhlenbeck operator plays an imporatant role in theoretic considerations and we will come back to this operator in Chapter 3.

1.5. Multiple Wiener Integrals

In this section, we take a closer look at the situation where the Hilbert space H is of the form $L^2(T,\mathscr{B},\mu)$ where $(T,\mathscr{B},\mu)$ is a σ-finite measure space. This in particular includes the situation of Brownian motion from Example 1.1.3 where $(T,\mathscr{B},\mu) = ((0,\infty),\mathscr{B}((0,\infty),\lambda)$. We recall that the Brownian motion B_t was defined as $B_t := W(\mathbb{1}_{(0,t]})$. We have then remarked that one often writes

$$\int_0^T f(t)\,dB_t := W(\mathbb{1}_{(0,T]}f)$$

and calls this the *Wiener integral* of f over $(0,T)$. We note that if $f = \sum_j \alpha_j\mathbb{1}_{(a_j,b_j]}$ is a step function, then by linearity

$$\int f(t)\,dB(t) = \sum_j \alpha_jW(\mathbb{1}_{(a_j,b_j]}) = \sum_j \alpha_j\big(W(\mathbb{1}_{(0,b_j]}) - W(\mathbb{1}_{(0,a_j]})\big)\sum_j \alpha_j(B_{b_j} - B_{a_j})$$

so that the Wiener integral is indeed defined via a certain Riemannian sum. We would also like to note that $\{W(h) : h \in H\} = \mathscr{H}_1$, so that the Wiener integral yields a description of the first Wiener chaos.

In this section, we develop a theory of multiple Wiener integrals which will then give a description of all Wiener chaoses in the important case where our Hilbert space H is an L^2-space. This more detailed information about the Wiener chaos decomposition will then also yield refined results about the central operators D, δ and L.

Throughout this section, we fix a σ-finite measure space $(T, \mathscr{B}, \mu)$ and consider the Hilbert space $H = L^2(T, \mathscr{B}, \mu)$. We will assume that the measure μ has no atoms. We recall that an *atom* is a measurable set A with positive measure such that every measurable subset B of A has either measure zero or the same measure as A. It can be proved that if μ has no atoms, then for every $A \in \mathscr{B}$ with $\mu(A) > 0$ and every $r \in (0,1)$, there exists a measurable subset B of A with $\mu(B) = r\mu(A)$.

Since simple functions are dense in L^2 the isonormal Gaussian process W is uniquely determined by the random variables $\{W(A) : A \in \mathscr{B}, \mu(A) < \infty\}$ where $W(A) := W(\mathbb{1}_A)$. Sometimes this family $\{W(A) : A \in \mathscr{B}, \mu(A) < \infty\}$ is called *white noise based on* μ.

We set $\mathscr{B}_0 := \{A \in \mathscr{B} : \mu(A) < \infty\}$.

Definition 1.5.1. Let $m \in \mathbb{N}$. The *space of elementary functions* $\mathscr{E}_m$ consists of those functions $f \in L^2(T^m, \mathscr{B}^m, \mu^m)$ which are of the form

$$f(t_1, \ldots, t_m) = \sum_{i_1, \ldots, i_m = 1}^{n} a_{i_1 \ldots i_m} \mathbb{1}_{A_{i_1} \times \cdots \times A_{i_m}}(t_1, \ldots, t_m) \tag{1.7}$$

where $n \in \mathbb{N}$, $A_1, \ldots, A_n$ are pairwise disjoint sets in $\mathscr{B}_0$ and the coefficients $a_{i_1 \ldots i_m}$ are zero whenever any two of the indices $i_1, \ldots, i_m$ are equal.

Thus $\mathscr{E}_m$ is generated by the indicator functions $\mathbb{1}_B$ where $B = A_1 \times \cdots \times A_m$ is a rectangle in T^m with finite measure that does not intersect any of the diagonal subspace $\Delta_{ij} = \{t \in T^m : t_i = t_j\}$. Note that $f(t) = 0$ for every $f \in \mathscr{E}_m$ and $t \in \Delta_{ij}$. This property will play an important role in the definition of the m-fold Wiener integral. Before we introduce this integral, we note

Lemma 1.5.2. *For $m \in \mathbb{N}$, the space $\mathscr{E}_m$ is dense in $L^2(T^m, \mathscr{B}^m, \mu^m)$.*

Proof. It suffices to prove that $\mathbb{1}_B$ is contained in the closure of $\mathscr{E}_m$ whenever $B = A_1 \times \cdots \times A_m$ for certain $A_1, \ldots, A_m \in \mathscr{B}_0$. Fix such a set B. Given $\varepsilon > 0$, we find pairwise disjoint sets $E_1, \ldots, E_n$ with $\mu(E_i) \le \varepsilon$ for $1 \le i \le n$ such that each set A_j can be written as disjoint union of some of the E_i's. This is a consequence of the fact that μ has no atoms. We may assume that $\bigcup_j E_j = \bigcup_i A_i$ and denote the measure of the latter set by M.

Having the sets E_i at hand, we can now find coefficients $\varepsilon_{i_1,\ldots,i_m} \in \{0,1\}$ such that

$$\mathbb{1}_B = \sum_{i_1,\ldots,i_m=1}^{n} \varepsilon_{i_1,\ldots,i_m} \mathbb{1}_{E_{i_1} \times \ldots, \times E_{i_m}}.$$

Now let I be the set of tuples $(i_1, \ldots, i_m)$ where all indices are different and let J be the set of the remaining tuples. Setting

$$\mathbb{1}_C = \sum_{i_1,\ldots,i_m \in I} \varepsilon_{i_1,\ldots,i_m} \mathbb{1}_{E_{i_1} \times \ldots, \times E_{i_m}}.$$

Then $C \subset B$ and $\mathbb{1}_C \in \mathscr{E}_m$. Moreover, we have

$$\mathbb{1}_B - \mathbb{1}_C = \sum_{i_1,\ldots,i_m \in J} \varepsilon_{i_1,\ldots,i_m} \mathbb{1}_{E_{i_1} \times \ldots, \times E_{i_m}}$$

Let us now describe the set J in more detail. We have $\binom{m}{2}$ possibilities to pick two positions $1, \ldots, m$ onto which we put an identical index, e.e. one of the numbers $1, \ldots, n$. On the remaining positions, we put any one of these numbers.

With this information, we estimate

$$\begin{aligned}\|\mathbb{1}_B - \mathbb{1}_C\|^2_{L^2(T^m)} &= \sum_{i_1,\ldots,i_m\in J} \varepsilon_{i_1,\ldots,i_m}\mu(E_{i_1})\cdot\ldots\cdot\mu(E_{i_m})\\ &\leq \binom{m}{2}\sum_{j=1}^n \mu(E_j)^2\Big(\sum_{j=1}^n \mu(E_j)\Big)^{m-2}\\ &\leq \binom{m}{2}\varepsilon\Big(\sum_{j=1}^n \mu(E_j)\Big)^{m-1} \leq \varepsilon M^{m-1}\binom{m}{2}\end{aligned}$$

which proves that $\mathbb{1}_B$ is indeed contained in the closure of $\mathscr{E}_m$. □

Remark 1.5.3. Lemma 1.5.2 is not true for measures with atoms. Indeed, consider $T = \{0,1\}$ and $\mu(\{0\}) = \mu(\{1\}) = 1$. Then $\mathscr{E}_2$ consists of those functions f with $f(0,0) = f(1,1) = 0$. But these functions are certainly not dense in $L^2(T^2)$.

We now introduce the m-fold Wiener integral.

Definition 1.5.4. For $f \in \mathscr{E}_m$ of the form (1.7) the *m-fold Wiener integral* $I_m(f)$ is defined as

$$I_m(f) := \sum_{i_1,\ldots,i_m=1}^n a_{i_1,\ldots,i_m} W(A_{i_1})\cdot\ldots\cdot W(A_{i_m})\,.$$

We leave it to the reader to verify that the definition of $I_m(f)$ is independent of the particular representation of the function f. Clearly, the Wiener integral is linear.

Let us note some further properties of the Wiener integral.

Lemma 1.5.5. *Let $f \in \mathscr{E}_m$ and $g \in \mathscr{E}_k$.*

(1) *Let $\tilde f$ denote the symmetrization of f,*

$$\tilde f(t_1,\ldots,t_m) = \frac{1}{m!}\sum_{\sigma\in\mathfrak{S}_m} f(t_{\sigma(1)},\ldots,t_{\sigma(m)})\,.$$

Then $I_m(f) = I_m(\tilde f)$.

(2) *We have*

$$\mathbb{E}\big(I_m(f)I_k(g)\big) = \begin{cases} 0 & \text{if } m \neq k\\ m!\big\langle \tilde f, \tilde g\big\rangle_{L^2(T^m)} & \text{if } m = k\end{cases}$$

Proof. (1) By linearity, it suffices to consider $f = \mathbb{1}_{A_1\times\cdots\times A_m}$ where $A_1,\ldots,A_m$ are pairwise disjoint. Noting that

$$W(A_1)\cdot\ldots\cdot W(A_m) = W(A_{\sigma(1)})\cdot\ldots\cdot W(A_{\sigma(m)})$$

for any $\sigma \in \mathfrak{S}_m$, the thesis clearly holds in this case.

(2) We may and shall assume that f and g are based on the same partition $A_1,\ldots,A_n$. We assume that f is given by (1.7) and

$$g(t_1,\ldots,t_k) = \sum_{j_1,\ldots,j_k=1}^n b_{j_1,\ldots,j_k}\mathbb{1}_{A_{j_1}\times\ldots\times A_{j_k}}\,.$$

First assume that $m \neq k$. Note that the product $I_m(f)I_k(g)$ is a sum of terms of the form

$$a_{i_1,\ldots,i_m}b_{j_1,\ldots,j_k}W(A_{i_1})\cdot\ldots W(A_{j_m})\cdot W(A_{j_1})\cdot\ldots\cdot W(A_{j_k})\,.$$

Since $m \neq k$, at least one index in $\{1,\ldots,n\}$ appears exactly once in the above product. Since $W(A_j)$ and $W(A_i)$ are independent for $i \neq j$, it follows that the above product has expectation zero in this case.

Now assume that $m = k$. By part (1) we may and shall assume that f and g are symmetric so that $a_{i_{\sigma(1)},\ldots,i_{\sigma(m)}} = a_{i_1,\ldots,i_m}$ for all $\sigma \in \mathfrak{S}_m$ and similarly for the $b's$. We have

$$\begin{aligned}\mathbb{E}\big(I_m(f)I_m(g)\big) &= \mathbb{E}\Big[\Big(m!\sum_{i_1<i_2<\cdots<i_m} a_{i_1,\ldots,i_m}W(A_{i_1})\cdot\ldots\cdot W(A_{i_m})\Big)\\ &\qquad\times\Big(m!\sum_{j_1<j_2<\cdots<j_m} b_{j_1,\ldots,b_m}W(A_{j_1})\cdot\ldots\cdot W(A_{j_m})\Big)\Big]\\ &= (m!)^2\sum_{i_1<i_2<\cdots<i_m} a_{i_1,\ldots,i_m}b_{i_1,\ldots,i_m}\mu(A_{i_1})\cdot\ldots\cdot\mu(A_{i_m})\\ &= m!\langle f,g\rangle_{L^2(T^m)}\end{aligned}$$

□

Setting $f = g$ in part (2) of Lemma 1.5.5 and noting that $\|\tilde{f}\|_{L^2(T^m)} \leq \|f\|_{L^2(T^m)}$ it follows that I_m is a bounded linear operator from $\mathscr{E}_m$ to $L^2(\Omega)$ and thus, as a consequence of Lemma 1.5.2, has an extension to a bounded linear operator from all of $L^2(T^m)$ to $L^2(\Omega)$. We still denote this extension by I_m. We also write sometimes

$$I_m(f) =: \int_{T^m} f(t_1,\ldots,t_m)dW(t_1)\ldots dW(t_m)\,.$$

In the case where $T = (0,\infty)$ it is also customary to use the differential $dB_{t_1}\ldots dB_{t_m}$.

Our next goal is to prove that the range of I_m is exactly the m-th Wiener chaos $\mathscr{H}_m$. The proof is based on induction and uses a technical tool which we develop first.

Given $f \in L^2(T^m)$ and $g \in L^2(T^k)$ the tensor $f\otimes g$ is the function $f\otimes g \in L^2(T^{m+k})$ defined by $(f\otimes g)(t_1,\ldots,t_{m+k}) = f(t_1,\ldots,t_m)g(t_{m+1},\ldots,t_{m+k})$. We also define the contraction $f\otimes_1 g$ as the function $f\otimes_1 g \in L^2(T^{m+k-2})$ given by

$$(f\otimes_1 g)(t_1,\ldots,t_{m+k-1}) := \int_T f(t_1,\ldots,t_{m-1},s)g(t_m,\ldots,t_{m+k-2},s)\,d\mu(s)\,.$$

We have

Proposition 1.5.6. *Let $f \in L^2(T^m)$ be symmetric and $g \in L^2(T)$. Then*

$$I_m(f)I_1(g) = I_{m+1}(f\otimes g) + mI_{m-1}(f\otimes_1 g)\,.$$

Proof. First assume that $f = \tilde{\mathbb{1}}_{A_1\times\cdots\times A_m}$ and $g = \mathbb{1}_B$ where $A_1,\ldots,A_m$ are pairwise disjoint elements of $\mathscr{B}_0$. and $B \in \mathscr{B}_0$. If B is disjoint from $A_1,\ldots,A_m$, then $f\otimes g \in \mathscr{E}_{m+1}$ and $f\otimes_1 g = 0$. By the definition of the integral

$$I_m(f)I_1(g) = W(A_1)\cdot\ldots W(A_m)W(B) = I_{m+1}(f\otimes g)\,,$$

which is the thesis in this case.

If B is not disjoint from the A_j's we may and shall assume that $B = A_1$. In this case,

$$\begin{aligned}f\otimes_1 g(t_1,\ldots,t_{m-1}) &= \frac{1}{m!}\sum_{\sigma\in\mathfrak{S}_m}\mathbb{1}_{A_{\sigma(1)}}(t_1)\cdot\ldots\cdot\mathbb{1}_{A_{\sigma(m-1)}}(t_{m-1})\int_T\mathbb{1}_{A_{\sigma(m)}}(s)\mathbb{1}_{A_1}(s)\,d\mu(s)\\ &= \frac{1}{m!}\sum_{\sigma\in\mathfrak{S}_m,\sigma(m)=1}\mathbb{1}_{A_{\sigma(1)}}(t_1)\cdot\ldots\cdot\mathbb{1}_{A_{\sigma(m-1)}}(t_{m-1})\mu(A_1)\\ &= \frac{1}{m}\tilde{\mathbb{1}}_{A_2\times\cdots\times A_m}(t_1,\ldots,t_m)\mu(A_1)\end{aligned}$$

Next, given $\varepsilon > 0$, we find a measurable partition of A_1 in disjoint sets $B_1,\ldots,B_n$ with $\mu(B_i) < \varepsilon$. We can define

$$h_\varepsilon := \sum_{i\neq j}\mathbb{1}_{B_i\times B_j\times A_2\times\cdots\times A_m} \in \mathscr{E}_{m+1}\,.$$

We find

$$\begin{aligned} I_m(f)I_1(g) &= W(A_1)^2W(A_2)\cdot\ldots\cdot W(A_m) = \Big(\sum_{j=1}^n W(B_j)\Big)^2 W(A_2)\cdot\ldots\cdot W(A_m)\\ &= \mu(A_1)W(A_2)\cdot\ldots\cdot W(A_m) + \sum_{i\neq j} W(B_i)W(B_j)W(A_2)\cdot\ldots\cdot W(A_m)\\ &\qquad + \sum_{j=1}^n \big[W(B_j)^2 - \mu(B_j)\big]W(A_2)\cdot\ldots\cdot W(A_m)\\ &= mI_{m-1}(f\otimes_1 g) + I_{m+1}(h_\varepsilon) + R_\varepsilon. \end{aligned}$$

Setting $M := \mu(A_1)\cdot\ldots\cdot\mu(A_m)$, we find

$$\mathbb{E}R_\varepsilon^2 = 2\sum_{j=1}^n \mu(B_j)^2\mu(A_2)\ldots\mu(A_m) \le 2\varepsilon\sum_{j=1}^n \mu(B_j)\mu(A_2)\cdot\ldots\cdot\mu(A_m) \le 2\varepsilon M\,.$$

Here the first equality follows from the fact that $\mathbb{E}X^4 = 3\sigma^4$ for $X \sim \mathscr{N}(0,\sigma^2)$. Thus the thesis follows upon $\varepsilon \to 0$ if we prove that $h_\varepsilon \to f\otimes g$ for $\varepsilon\to 0$. Indeed we have

$$\begin{aligned} \|\tilde h_\varepsilon - f\,\tilde\otimes\, g\|_{L^2(T^{m+1})} &= \big\|\tilde h_\varepsilon - \tilde{\mathbb{1}}_{A_1\times\cdots\times A_m\times A_1}\big\|_{L^2(T^{m+1})}\\ &= \big\|\tilde h_\varepsilon - \tilde{\mathbb{1}}_{A_1\times A_1\times A_2\cdots\times A_m}\big\|_{L^2(T^{m+1})}\\ &\le \big\|h_\varepsilon - \mathbb{1}_{A_1\times A_1\times A_2\cdots\times A_m}\big\|_{L^2(T^{m+1})}\\ &= \sum_{j=1}^n \mu(B_j)^2\mu(A_2)\cdot\ldots\mu(A_m) \le \varepsilon M \end{aligned}$$

which implies the necessary convergence. □

We now obtain

Proposition 1.5.7. *Let H_m be the m-th Hermite polynomial and $h\in L^2(T)$ be of norm one. Then*

$$m!H_m(W(h)) = \int_{T^m} h(t_1)\cdot\ldots\cdot h(t_m)\,dW(t_1)\ldots dW(t_m)\,.$$

Moreover, the m-fold Wiener integral maps $L^2(T^m)$ onto $\mathscr{H}_m$. Finally, $I_m(f) = I_m(g)$ if and only if $\tilde f = \tilde g$.

Proof. Let us write $h^{\otimes m}$ for the function $h\otimes\cdots\otimes h$ with m factors.

We proof the above equation by induction on m. To include $m=0$, we set $h^{\otimes 0}\equiv 1$ and let I_0 be the identity map on the constant functions. Thus the formula trivially holds for $m=0$. For $m=1$, we clearly have $W(h) = H_1(W(h)) = \int_T h(t)\,dW(t)$. Thus assume that the formula is correct for some m and all smaller values. By Proposition 1.5.6

$$\begin{aligned} I_{m+1}(h^{\otimes(m+1)}) &= I_m(h^{\otimes m})I_1(h) - mI_{m-1}(h^{\otimes m}\otimes_1 h)\\ &= m!H_m(W(h))W(h) - mI_{m-1}(h^{\otimes(m-1)}\|h\|^2)\\ &= m!H_m(W(h))W(h) - m(m-1)!H_{m-1}(W(h))\\ &= m!(m+1)H_{m+1}(W(h)) \end{aligned}$$

since $(m+1)H_{n+1}(x) = xH_n(x) - H_{n-1}(x)$, see Lemma 1.1.7.

For the second part, let $L^2_{\mathrm{sym}}(T^m)$ denote the closed subspace of $L^2(T^m)$ consisting of symmetric functions. By Lemma 1.5.5, $\mathbb{E}I_m(f)^2 = m!\|f\|^2_{L^2(T^m)}$ for $f\in L^2_{\mathrm{sym}}(T^m)$. Thus $I_mL^2_{\mathrm{sym}}(T^m)$ is a closed subspace of $L^2(\Omega)$ which, by the above, contains $H_m(W(h))$ for $h\in H$ of norm 1. Consequently, $\mathscr{H}_m\subset I_mL^2_{\mathrm{sym}}(T^m)$. Since $I_m(f)\perp I_k(g)$ for $k\neq m$, it follows that $I_mL^2_{\mathrm{sym}}(T^m)$ is orthogonal to $\mathscr{H}_k$ for $k\neq m$, whence $I_mL^2_{\mathrm{sym}}(T^m) = \mathscr{H}_m$ □

We immediately obtain the following version of the Wiener chaos expansion, which represents an element of $L^2(\Omega)$ as a series of multiple Wiener integrals.

THEOREM 1.5.8. *Let $X \in L^2(\Omega, \Sigma, \mathbb{P})$. Then there exists unique symmetric functions $f_n \in L^2(T^n)$ with*

$$X = \sum_{n=0}^{\infty} I_n(f_n)\,.$$

Here $f_0 = \mathbb{E}X$ and I_0 is the identity on the constants.

1.6. Stochastic Calculus in the White Noise Case

We now proceed to study the special situation where $H = L^2(T, \mathscr{B}, \mu)$ for a σ-finite measure space $(T, \mathscr{B}, \mu)$ without atoms. Given a random variable $X \in \mathbb{D}^{1,2}$, the Malliavin derivative DX is an element of $L^2(\Omega; H)$ which in this special case can be identified with $L^2(T \times \Omega)$. Thus the Malliavin derivative can be viewed as a stochastic process $\{D_t X : t \in T\}$ where $D_t X$ is defined almost everywhere with respect to the measure $\mu \otimes \mathbb{P}$. Similarly, if $X \in \mathbb{D}^{k,2}$ then the k-fold Malliavin derivative $D^k X$ is an element of $L^2(\Omega; H^{\otimes k})$ which can canonically be identified with $L^2(T^k \times \Omega)$ whence the derivative $D^k X$ can be viewed as a k-parameter stochastic process $\{D^k_{t_1,\dots,t_k} X : t_j \in T\}$. Even more is true. Namely an element of X of $\mathbb{D}^{k,2}$ belongs to $\mathbb{D}^{k+1,2}$ if and only if $D^k_{t_1,\dots,t_k} X \in \mathbb{D}^{1,2}$ for $\mu^{\otimes k}$-almost every $(t_1, \dots, t_k)$ and $t \mapsto \mathbb{E}\|D_t D^k_{t_1,\dots,t_k} X\|^2_{L^2(T^k)}$ belongs to $L^2(T)$. in that case $D^{k+1}_{t,t_1,\dots,t_k} X = D_t D^k_{t_1,\dots,t_k} X$ almost surely. Indeed, consider the case $k = 1$ and let $X = f(W(h_1), \dots, W(h_n)) \in \mathscr{S}$. In this case

$$D_t X = \sum_{i=1}^{n} (\partial_i f)(W(h_1), \dots, W(h_n)) h_j(t), \quad DX = \sum_{i=1}^{n} (\partial_i f)(W(h_1), \dots, W(h_n)) h_j,$$

$$D_s D_t X = \sum_{j=1}^{n} \sum_{i=1}^{n} (\partial_j \partial_j f)(W(h_1), \dots, W(h_n)) h_i(s) h_j(t),$$

and

$$D^2 X = \sum_{j=1}^{n} \sum_{i=1}^{n} (\partial_j \partial_j f)(W(h_1), \dots, W(h_n)) h_i \otimes h_j.$$

It follows from the identifications of Tensor product spaces that

$$\mathbb{E}\big(|X|^2 + \|DX\|^2_H + \|D^2 X\|^2_{H \otimes H}\big) = \mathbb{E}\big(|X|^2 + \|t \mapsto D_t X\|^2_{L^2(T)} + \|(s,t) \mapsto D_s D_t X\|^2_{L^2(T^2)}\big)$$

for such X. Thus the claim follows from approximation.

Similar remarks also apply to the divergence operator δ. If we view the Malliavin derivative as an operator from $L^2(\Omega)$ to $L^2(T \times \Omega)$, then the adjoint δ is an operator on $L^2(T \times \Omega)$ taking values in $L^2(\Omega)$. Thus the domain of δ consists of certain stochastic processes. It is especially in this situation that one calls δ the Skorohod integral.

In this section, we study properties of the Malliavin derivative and the divergence operator in the white noise setting.

Let us start with a result about the action of D in terms of multiple Wiener integrals.

PROPOSITION 1.6.1. *Let $X \in \mathbb{D}^{1,2}$ have Wiener expansion $X = \sum_{n=0}^{\infty} I_n(f_n)$ where $f_n \in L^2_{\text{sym}}(T^n)$. Then*

$$D_t X = \sum_{n=1}^{\infty} n I_{n-1}(f_n(\cdot, t))\,. \tag{1.8}$$

Before proving Proposition 1.6.1, let us give an interpretation. If X is given as a single multiple wiener integral $X = I_n(f_n)$, say

$$X = \int_{T^n} f_n(t_1, \dots, t_n)\, dW(t_1) \dots dW(t_n),$$

then (1.8) says that the derivative of X is obtained by "removing" one of the stochastic integrals, viewing the variable which is not integrated as a parameter and multiplying with n. Note that it does not matter with respect to which variable we do not integrate, as the function f_n is assumed to be symmetric.

Proof of Proposition 1.6.1. In view of Proposition 1.2.17, it suffices to proof the result for X of the form $X = I_m(f_m)$ where f_m is symmetric. We may furthermore assume that $f_m \in \mathscr{E}_m$, say

$$f_m = \sum_{i_1, \dots i_m = 1}^{n} a_{i_1,\dots,i_m} \mathbb{1}_{A_{i_1} \times \dots \times A_{i_m}}$$

where $A_1, \dots, A_n$ are pairwise disjoint sets of finite measure. Hence

$$X = \sum_{i_1, \dots i_m = 1}^{n} a_{i_1,\dots,i_m} W(A_{i_1}) \cdot \dots \cdot W(A_{i_m})$$

Thus $X \in \mathscr{S}$ and we have, by definition,

$$DX = \sum_{i_1, \dots i_m = 1}^{n} \sum_{j=1}^{m} a_{i_1,\dots,i_m} \mathbb{1}_{A_{i_j}} \prod_{k \neq j} W(A_{j_k})$$

On the other hand,

$$\begin{aligned}
\tilde{\mathbb{1}}_{A_{i_1} \times \dots \times A_{i_m}}(t_1, t_2, \dots, t_m) &= \frac{1}{m!} \sum_{\sigma \in \mathfrak{S}_m} \prod_{k=1}^{n} \mathbb{1}_{A_{i_{\sigma(k)}}}(t_k) \\
&= \frac{1}{m!} \sum_{j=1}^{m} \mathbb{1}_{A_{i_j}}(t_m) \sum_{\sigma \in \mathfrak{S}_m, \sigma(j) = m} \prod_{k \neq j} \mathbb{1}_{A_{i_{\sigma(k)}}}(t_k) \\
&= \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{A_j}(t_m) \tilde{\mathbb{1}}_{\prod_{k \neq j} A_{i_k}}(t_1, \dots, t_{m-1})
\end{aligned}$$

This implies that

$$I_{m-1}(f_m(\cdot, t)) = \sum_{i_1, \dots i_m = 1}^{n} a_{i_1,\dots,i_m} \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{A_{i_j}}(t) \prod_{k \neq j} W(A_{i_k}),$$

finishing the proof. □

We give some applications of this result. Let us fix a set $A \in \mathscr{B}$. We put

$$\Sigma_A := \sigma(W(B) : B \in \mathscr{B}_0, B \subset A).$$

In the particular case where $T = (0, \infty)$, where we have defined $B_t := \mathbb{1}_{(0,t]}$, we can for example consider $A = (0, t]$. Then $\Sigma_{(0,t]} = \sigma(B_s : s \leq t)$. It is more customary to write $\mathscr{F}_t := \Sigma_{(0,t]}$. Note that $\mathbb{F} = (\mathscr{F}_t)_{t \geq 0}$ is the natural filtration generated by Brownian motion. We first compute the Wiener chaos representation of a conditional expectation.

Lemma 1.6.2. *Let $X \in L^2(\Omega)$ have Wiener expansion $X = \sum_{n=0}^{\infty} I_n(f_n)$ and $A \in \mathscr{B}$. Then*

$$\mathbb{E}\big(X \big| \Sigma_A\big) = \sum_{n=0}^{\infty} I_n(f_n \mathbb{1}_A^{\otimes n}).$$

PROOF. It suffices to prove the result for $X = I_n(f_n)$ where $f_n \in \mathscr{E}_n$. Moreover, by linearity, we can assume that $f_n = \mathbb{1}_{B_1\times\cdots\times B_n}$ where $B_1, \ldots, B_n$ are pairwise disjoint sets of finite measure. Thus $X = W(B_1)\cdot\ldots\cdot W(B_n)$. We find

$$\begin{aligned}
\mathbb{E}\big(X\big|\Sigma_A\big) &= \mathbb{E}\Big(\prod_{j=1}^n \big((W(B_j\cap A) + W(B_j\cap A^c)\big)\Big|\Sigma_A\Big)\\
&= \prod_{j=1}^n \big(W(B_j\cap A) + 0\big)\\
&= I_n(\mathbb{1}_{(B_1\cap A)\times\cdots\times(B_n\cap A)}) = I_n(f_n\mathbb{1}_A^{\otimes n})\,.
\end{aligned}$$

Here we have used that the random variables $W(B_j\cap A)$ and also the random variables $W(B_j\cap A^c)$ are independent for different values of j and that the former are Σ_A measurable whereas the latter are independent from Σ_A. □

We can now prove

PROPOSITION 1.6.3. *Let $X \in \mathbb{D}^{1,2}$ and $A \in \mathscr{B}$. Then $\mathbb{E}(X|\Sigma_A) \in \mathbb{D}^{1,2}$ and*

$$D_t\mathbb{E}\big(X\big|\Sigma_A\big) = \mathbb{E}\big(D_t\big|\Sigma_A\big)\mathbb{1}_A\,.$$

PROOF. Let $X = \sum_{n=0}^\infty I_n(f_n)$. Combining Proposition 1.6.1 with Lemma 1.6.2, we have

$$\begin{aligned}
D_t\mathbb{E}(X|\Sigma_A) &= D_t\sum_{n=0}^\infty I_n(f_n\mathbb{1}_A^{\otimes n}) = \sum_{n=1}^\infty nI_{n-1}(f_n(\cdot,t)\mathbb{1}_A^{\otimes n}(\cdot,t))\\
&= \sum_{n=1}^\infty nI_{n-1}(f_n(\cdot,t)\mathbb{1}_A^{\otimes(n-1)}(\cdot))\mathbb{1}_A(t) = \mathbb{E}(D_tX|\Sigma_A)\mathbb{1}_A(t)\,.
\end{aligned}$$

In particular, the last series converges so that indeed $\mathbb{E}(X|\Sigma_A) \in \mathbb{D}^{1,2}$. □

This implies the following

COROLLARY 1.6.4. *Let $A \in \mathscr{B}$ and assume that $X \in \mathbb{D}^{1,2}$ is Σ_A-measurable. Then $D_tX = 0$ almost everywhere on $A^c \times \Omega$.*

Let us give an interpretation of this result in the case where $T = (0,\infty)$. We again put $B_t = W(\mathbb{1}_{(0,t]})$ and $\mathscr{F}_t = \sigma(B_s : s \le t)$. Then if X is $\mathscr{F}_t$ measurable, then the Malliavin derivative D_sX is supported in the interval $[0,t]$, i.e. $D_sX = 0$ for $s > t$.

Let us now address the divergence operator. As we have already mentioned above, the domain of the divergence is a subset of $L^2(T\times\Omega)$. If $u = u(t,\omega) \in D(\delta)$ we will use the notation

$$\delta(u) =: \int_T u(t)\,\delta W(t)$$

and call $\delta(u)$ the *Skorohod integral* of the "process" $(u(t))_{t\in T}$.

We have seen in Proposition 1.3.4 that if $u \in \mathbb{D}^{1,2}(H)$, then $u \in D(\delta)$. We will now, using multiple Wiener integrals, give a full description of the domain of the Skorohod integral. Given an element $u \in L^2(T\times\Omega)$ we have a Wiener expansion of the form

$$u(t) = \sum_{n=0}^\infty I_n(f_n(\cdot,t))$$

where $f_n \in L^2(T^{n+1})$ is a symmetric function in the first n variables. Using Fubini's theorem and Lemma 1.5.5 (2), we find

$$\mathbb{E}\int_T u(t)^2\,d\mu(t) = \sum_{n=0}^\infty n!\|f_n\|^2_{L^2(T^{n+1})}\,.$$

We note that f_n is not necessarily symmetric in all $n+1$ variables. The symmetrization $\tilde{f}_n$ is given by

$$\tilde{f}_n(t_1,\dots,t_n,t) = \frac{1}{n+1}\Big(f_n(t_1,\dots,t_n,t) + \sum_{j=1}^n f(t_1,\dots,t_{j-1},t,t_{j+1},\dots,t_n,t_j)\Big)\,.$$

PROPOSITION 1.6.5. *Let $u \in L^2(T\times\Omega)$ have Wiener expansion $u = \sum_{n=0}^\infty I_n(f_n(\cdot,t))$. Then $u\in D(\delta)$ if and only if $\sum_{n=0}^\infty (n+1)!\|\tilde{f}_n\|^2_{L^2(T^{n+1})} < \infty$. In that case,*

$$\delta(u) = \sum_{n=0}^\infty I_{n+1}(\tilde{f}_n)\,.$$

PROOF. Let $Y = I_n(g)$ for some symmetric g. We have

$$\begin{aligned}
\mathbb{E}\langle u, DY\rangle &= \mathbb{E}\int_T u_t D_t Y\, d\mu(t)\\
&= \mathbb{E}\int_T u_t n I_{n-1}(g(\cdot,t))\,d\mu(t)\\
&= \sum_{m=0}^\infty \int_T n\mathbb{E} I_m(f_m(\cdot,t)) I_{n-1}(g(\cdot,t))\,d\mu(t)\\
&= n\int_T \mathbb{E} I_{n-1}(f_{n-1}(\cdot,t)) I_{n-1}(g(\cdot,t))\,d\mu(t)\\
&= n(n-1)!\int_T \langle f_{n-1}(\cdot,t), g_{n-1}(\cdot,t)\rangle_{L^2(T^{n-1})}\,d\mu(t)\\
&= n!\langle f_{n-1}, g\rangle_{L^2(T^n)} = n!\Big\langle \tilde{f}_{n-1}, g\Big\rangle_{L^2(T^n)}\\
&= \mathbb{E}\big(I_n(\tilde{f}_{n-1}) I_n(g)\big) = \mathbb{E}\big(I_n(\tilde{f}_{n-1})Y\big)
\end{aligned}$$

After this preliminary computation, first assume that $u\in D(\delta)$. Then the above yields

$$\mathbb{E}\big(\delta(u)Y\big) = \mathbb{E}\langle u, DY\rangle = \mathbb{E}\big(I_n(\tilde{f}_{n-1})Y\big)$$

for all $Y\in\mathscr{H}_n$. This implies that $J_n\delta(u) = I_n(\tilde{f}_{n-1})$. Consequently, $\delta(u) = \sum_{n=0}^\infty I_{n+1}(\tilde{f}_n)$. In particular, it follows that the latter series converges. By orthogonality and Lemma 1.5.5 (2) $\mathbb{E}\delta(u)^2 = \sum_{n=0}^\infty (n+1)!\|\tilde{f}_n\|^2_{L^2(T^{n+1})} < \infty$.

Conversely, if $\sum_{n=0}^\infty (n+1)!\|\tilde{f}_n\|^2_{L^2(T^{n+1})} < \infty$, then $Z := \sum_{n=0}^\infty I_{n+1}(\tilde{f}_n)$ exists in $L^2(\Omega)$. The preceding computation yields

$$\mathbb{E}\langle u, DY\rangle = \mathbb{E}\big(ZY\big)$$

for all $Y\in\mathscr{H}_n$, hence, by linearity, for $Y\in\bigotimes_{n=0}^N \mathscr{H}_n$. By continuity, this extends even to $Y\in\mathbb{D}^{1,2}$ and we conclude that $u\in D(\delta)$ and $\delta(u) = Z$. □

Let us now compare the domain $D(\delta)$ with the space $\mathbb{D}^{1,2}(H)$ which is included in $D(\delta)$ by Proposition 1.3.4. The following Lemma gives examples of elements of $D(\delta)$ which are not included in $\mathbb{D}^{1,2}(H)$.

LEMMA 1.6.6. *Let $A\in\mathscr{B}_0$ and $X\in L^2(\Omega)$ be Σ_A^c-measurable. Then $X\mathbb{1}_A\in D(\delta)$ and*

$$\delta(X\mathbb{1}_A) = XW(A)\,.$$

PROOF. Let us first assume that $X\in\mathbb{D}^{1,2}$. By Corollary 1.6.4, $DX=0$ almost everywhere on $A\times\Omega$ so that $\langle DX,\mathbb{1}_A\rangle = \int_T D_tX\mathbb{1}_A(t)\,d\mu(t) = 0$ almost surely. Proposition 1.3.5 yields that $X\mathbb{1}_A\in D(\delta)$ and

$$\delta(X\mathbb{1}_A) = X\delta(\mathbb{1}_A) - \langle DX,\mathbb{1}_A\rangle = XW(A)\,.$$

For the general case, let X_n be a sequence in $\mathbb{D}^{1,2}$, converging to X in $L^2(\Omega)$. Replacing X_n with $\mathbb{E}(X_n|\Sigma_A)$ if necessary, we may and shall assume that X_n is Σ_A-measurable. Observe that $X_n \mathbb{1}_A$ converges to $X\mathbb{1}_A$ in $L^2(\Omega; H)$ and $\delta(X_n\mathbb{1}_A)$ converges to $XW(A)$. By closedness of δ, $X\mathbb{1}_A \in D(\delta)$ and $\delta(X\mathbb{1}_A) = XW(A)$. $\square$

By the remarks at the beginning of this section, if $\mu(A) > 0$, then $X\mathbb{1}_A \in \mathbb{D}^{1,2}(H)$ if and only if $X \in \mathbb{D}^{1,2}$. Thus there are elements in $D(\delta)$ not contained in $\mathbb{D}^{1,2}(H)$. Take, e.g., $X = \mathbb{1}_{\{W(B)>0\}}$ for some set B of positive measure disjoint from A. Then $X \notin \mathbb{D}^{1,2}$ by Corollary 1.2.19 since $\mathbb{P}(W(h) > 0) = \frac{1}{2}$.

We end this section by proving the following extension of the commutator relationship in Equation (1.3).

Proposition 1.6.7. *Let $u \in \mathbb{D}^{1,2}(H)$ be such that for almost every $t \in T$ the process $s \mapsto D_t u(s)$ belongs to $D(\delta)$ and such that there is a version of $t \mapsto \delta(D_t u(s))$ which is in $L^2(T \times \Omega)$. Then $\delta(u) \in \mathbb{D}^{1,2}$ and*

$$D_t(\delta(u)) = u(t) + \delta(D_t u)\,.$$

Proof. We assume that u has chaos expansion $u(t) = \sum_{n=0}^\infty I_n(f_n(\cdot, t))$, where f_n is symmetric in the first n variables. Since $\mathbb{D}^{1,2}(H) \subset D(\delta)$ by Proposition 1.3.4, it follows from Proposition 1.6.5 that $\sum_{n=0}^\infty (n+1)!\|\tilde f_n\|^2_{L^2(T^{n+1})} < \infty$ and $\delta(u) = \sum_{n=0}^\infty I_{n+1}(\tilde f_n)$. In particular, taking Proposition 1.5.7 into account, it follows that $J_n(\delta(u)) = I_n(\tilde f_{n-1})$. Noting that $\|J_n(\delta(u))\|_2^2 = n!\|\tilde f_{n-1}\|_{L^2(T^n)}$, it follows that $\sum n\|J_n(\delta(u))\|_2^2 < \infty$ whence $\delta(u) \in \mathbb{D}^{1,2}$ by Proposition 1.2.17. Moreover, by Proposition 1.6.1,

$$\begin{aligned} D_t(\delta(u)) &= D_t \sum_{n=0}^\infty I_{n+1}(\tilde f_n) = \sum_{n=0}^\infty (n+1) I_n(\tilde f_n(\cdot, t)) \\ &= \sum_{n=0}^\infty I_n\Big(f_n(t_1, \ldots, t_n, t) + \sum_{j=1}^n f_n(t_1, \ldots, t_{j-1}, t, t_{j+1}, \ldots, t_n, t_j\Big) \\ &= u(t) + \sum_{n=0}^\infty n I_n(\tilde g_n(\cdot, t)) \end{aligned}$$

where

$$\tilde g_n(t_1, \ldots, t_n, t) = \frac{1}{n}\sum_{j=1}^n f_n(t_1, \ldots, t_{j-1}, t, t_{j+1}, \ldots, t_n, t_j\Big)$$

Now fix $t, s \in T$. Proposition 1.2.17 yields $D_t u(s) = \sum_{n=0}^\infty n I_{n-1}(f_n(\cdot, t, s))$. For fixed t Proposition 1.6.5 implies that

$$\delta(D_t u(s)) = \sum_{n=0}^\infty n I_n(\rho_n(\cdot, t))$$

where the function ρ_n is given by

$$\begin{aligned} \rho_n(t_1, \ldots, t_n, t) &:= \frac{1}{n}\sum_{j=1}^n f_n(t_1, \ldots, t_{j-1}, t_n, t_{j+1}, \ldots, t_{n-1}, t, t_j) \\ &= \frac{1}{n}\sum_{j=1}^n f_n(t_1, \ldots, t_{j-1}, t, t_{j+1}, \ldots, t_{n-1}, t_n, t_j) = \tilde g_n(\cdot, t) \end{aligned}$$

as f_n is symmetric in the first n variables. Together with the above this implies that $D_t(\delta(u)) = u(t) + \delta(D_t u)$ as claimed. $\square$

1.7. Itô's Integral and the Clark-Ocone Formula

Let us consider once again the Skorohod integral δ in the situation of the previous section, i.e. the white noise case. The space $\mathbb{D}^{1,2}(H) =: \mathbb{L}^{1,2}$ is contained in the domain of the Skorohod integral. This space coincides with the space of processes $u \in L^2(T\times\Omega)$ such that $u(t) \in \mathbb{D}^{1,2}$ almost surely and $D_s u(t) \in L^2(T^2\times\Omega)$. $\mathbb{L}^{1,2}$ is a Hilbert space with norm

$$\|u\|^2_{\mathbb{L}^{1,2}} := \|u\|^2_{L^2(T\times\Omega)} + \|Du\|^2_{L^2(T^2\times\Omega)}\,.$$

Proposition 1.3.4 can be rephrased by saying that for $u, v \in \mathbb{L}^{1,2}$ we have

$$\mathbb{E}(\delta(u)\delta(v)) = \mathbb{E}\int_T u(t)v(t)\,d\mu(t) + \mathbb{E}\int_T\int_T D_s u(t) D_t u(s)\,d\mu(t)\,d\mu(s)\,. \tag{1.9}$$

Indeed, for Hilbert Schmidt operators S, T we have $\mathrm{tr}(ST) = \langle T, S^*\rangle_{\mathsf{HS}}$. Noting that $f\otimes g^* = g\otimes f$, we see that when identifying the tensor product $L^2(T)\otimes L^2(T)$ with the product space $L^2(T^2)$ then the adjoint corresponds to interchanging the variables.

Now we consider the special situation where $T = (0,\infty)$. We set $B_t := W(\mathbb{1}_{(0,t]})$ and consider the natural filtration $\mathbb{F} = (\mathscr{F}_t)_{t\geq 0}$ where $\mathscr{F}_t = \Sigma_{(0,t]} = \sigma(B_s : s \leq t)$. Suppose that $u = v \in \mathbb{L}^{1,2}$ is such that $u(t)$ is $\mathscr{F}_t$-measurable for $t \geq 0$, i.e. the process u is *adapted*. In this case, it follows from Corollary 1.6.4 that $D_s u(t) = 0$ for $s > t$. It follows that $D_s u(t) D_t u(s) = 0$ almost surely and equation (1.9) reduces to the following *Itô-isometry*

$$\mathbb{E}\delta(u)^2 = \mathbb{E}\int_0^\infty u(t)^2\,dt\,. \tag{1.10}$$

Let us consider a special class of stochastic processes, the so-called *elementary step-processes*. An elementary step process is a process of the form

$$u = \sum_{j=1}^n X_j \mathbb{1}_{(t_{j-1},t_j]}$$

where $0 \leq t_0 < t_1 < \cdots < t_n$ and $X_j \in L^2(\Omega, \mathscr{F}_{t_j}, \mathbb{P})$. It follows from Lemma 1.6.6 and linearity that such a process u belongs to the domain of δ and

$$\delta(u) = \sum_{j=1}^n X_j(B_{t_j} - B_{t_{j-1}})\,.$$

Now assume that u_n is a sequence of elementary step process that converges in $L^2((0,\infty)\times\Omega)$ to the process u. The Itô-isometry (1.10) yields that the Skorohod integrals $\delta(u_n)$ are Cauchy, hence convergent, in $L^2(\Omega)$. Thus, by closedness of the Skorohod integral, every process in the closure of the elementary step processes in $L^2((0,\infty)\times\Omega)$ belongs to the domain of δ. It can be proved that the elements of the closure are exactly the adapted, square integrable processes, i.e. $u \in L^2_{\mathbb{F}}((0,\infty)\times\Omega)$.

We have thus proved:

Proposition 1.7.1. *We have* $L^2_{\mathbb{F}}((0,\infty)\times\Omega) \subset D(\delta)$.

For $u \in L^2_{\mathbb{F}}((0,\infty)\times\Omega)$ is customary to write

$$\int_0^\infty u(t)\,dB_t$$

instead of $\delta(u)$ and to call this the *Itô-integral* of u.

Actually, one can prove the Itô-isometry also directly for elementary step processes. One can thus develop the Itô integral independently of Malliavin calculus. In this case, the above results show that any Itô-integrable stochastic process is in the domain of the Skorohod integral and the Skorohod integral coincides with the Itô integral.

We now have the following result about differentiability of the Itô-integral.

PROPOSITION 1.7.2. *Let* $u \in L^2_{\mathbb{F}}((0,\tau)\times\Omega)$ *and define* $X := \int_0^\tau u(s)\,dB_s$. *Then* $X \in \mathbb{D}^{1,2}$ *if and only if* $u \in \mathbb{L}^{1,2}$. *In that case,* $t \mapsto D_t u(s)$ *belongs to* $L^2_{\mathbb{F}}$ *and for* $t \in (0,\tau)$ *we have*

$$D_t X = u(t) + \int_t^\tau D_t u(s)\,dB_s$$

almost surely.

PROOF. If $u \in \mathbb{L}^{1,2}$, then, by definition, $D_t u(s)$ exists almost everywhere and is square integrable. By Corollary 1.6.4, $D_t u(s) = 0$ for $t \geq s$. Moreover, the representation in Lemma 1.6.2 together with Proposition 1.6.1 easily yields that $D_t u(s)$ is $\mathscr{F}_s$ measurable for $t \leq s$. Consequently, the map $s \mapsto D_t u(s)$ is adapted. As it is also square integrable, it belongs to the domain of δ by Proposition 1.7.1. It follows from Proposition 1.6.7 that $X \in \mathbb{D}^{1,2}$ and

$$D_t X = u(t) + \int_t^\tau D_t u(s)\,dB_s\,.$$

Conversely, assume that $X \in \mathbb{D}^{1,2}$. Let $u_n(t)$ be the orthogonal projection of $u(t)$ onto $\mathscr{P}_n = \mathscr{H}_0 \oplus \cdots \oplus \mathscr{H}_n$ so that $u_n(t) \to u(t)$ in $L^2(\Omega)$. Note that u_n is adapted and square integrable, hence Itô integrable. We put $X_n := \int_0^\tau u_n(t)\,dB_t$. Taking Proposition 1.6.5 into account, we see that X_n is the projection of X onto $\mathscr{P}_{n+1}$. Using the above formula for X_n and the Itô isometry, it follows that X_n converges to X in $\mathbb{D}^{1,2}$. In particular, its $\mathbb{D}^{1,2}$ norm is bounded.

Using the above formula, we find

$$\begin{aligned}
\mathbb{E}\int_0^\tau |D_t X_n|^2\,dt &= \mathbb{E}\int_0^\tau \Big|u(t) + \int_t^\tau D_t u_n(s)\,dB_s\Big|^2\,dt\\
&= \mathbb{E}\int_0^\tau |u(t)|^2\,dt + \mathbb{E}\int_0^\tau\int_t^\tau |D_t u_n(s)|^2\,dsdt\\
&= \mathbb{E}\int_0^\tau |u(t)|^2\,dt + \mathbb{E}\int_0^\tau\int_0^s |D_t u_n(s)|^2\,dtds\\
&\geq \mathbb{E}\int_0^\tau\int_0^s |D_t u_n(s)|^2\,dtds = \mathbb{E}\|Du_n\|^2_{L^2((0,\tau)^2)}\,.
\end{aligned}$$

Here, we have used that $\int_t^\tau D_t u_n(s)\,dB_s$ is independent of $\mathscr{F}_t$, hence of $u(t)$ and the Itô isometry in the second equality and Fubini's theorem in the third.

It thus follows that $\mathbb{E}\|Du_n\|^2_{L^2((0,\tau)^2)}$ is bounded and it now follows from Lemma 1.2.8 that $u \in \mathbb{L}^{1,2}$. ☐

Let us now consider Brownian motion with finite time horizon τ, i.e. we consider $T = [0,\tau]$. A result by Itô states that any square integrable function X can be written as

$$X = \mathbb{E}X + \int_0^\tau u(t)\,dB_t$$

for a suitable process $u \in L^2_{\mathbb{F}}((0,\tau)\times\Omega)$. It is a natural question how to compute the process u given X. For random variables $X \in \mathbb{D}^{1,2}$ we have the following result, called the *Clark-Ocone formula*.

THEOREM 1.7.3. *Let* $T = [0,\tau]$ *and set as usual* $B_t := W(\mathbb{1}_{(0,t]})$ *and* $\mathscr{F}_t = \sigma(B_s : s \leq t)$. *Then for* $X \in \mathbb{D}^{1,2}$ *we have*

$$X = \mathbb{E}X + \int_0^\tau \mathbb{E}\big(D_t X\big|\mathscr{F}_t\big)\,dB_t\,.$$

PROOF. Suppose that X has the Wiener decomposition $X = \sum_{n=0}^{\infty} I_n(f_n)$ where f_n is symmetric. Using Proposition 1.6.1 and Lemma 1.6.2, we have

$$\mathbb{E}(D_t X|\mathscr{F}_t) \quad = \quad \sum_{n=1}^{\infty} n\mathbb{E}\big(I_{n-1}(f_n(\cdot,t))\big|\mathscr{F}_t\big) = \sum_{n=1}^{\infty} nI_{n-1}(f_n(\cdot,t)\mathbb{1}_{(0,t]}^{\otimes(n-1)})$$

We set $u(t) := \mathbb{E}(D_t X|\mathscr{F}_t)$. Clearly, $u \in L^2_{\mathbb{F}}((0,\infty)\times\Omega)$ so that $u \in D(\delta)$. Moreover, Itô and Skorohod integrals coincide. We compute $\delta(u)$ using Proposition 1.6.5. To that end, let us first compute the symmetrization of $f_n \mathbb{1}_{(0,t)}^{\otimes(n-1)}$.

Using that f_n is symmetric, we find

$$\begin{aligned} & \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} f_n(t_{\sigma(1)},\dots,t_{\sigma(n)})\mathbb{1}_{(0,t_{\sigma(n)})}(t_{\sigma(1)})\cdot\ldots\cdot\mathbb{1}_{(0,t_{\sigma(n)})}(t_{\sigma(n-1)}) \\ = \quad & f_n(t_1,\dots,t_n)\frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\mathbb{1}_{\left\{t_{\sigma(n)}=\max\{t_1,\dots,t_n\}\right\}} \\ = \quad & \frac{1}{n}f_n(t_1,\dots,t_n)\,. \end{aligned}$$

Hence

$$\delta(u) = \sum_{n=1}^{\infty} nI_n(\widetilde{f_n\mathbb{1}_{(0,t)}^{\otimes(n-1)}}) = \sum_{n=1}^{\infty} I_n(f_n) = X - \mathbb{E}X \qquad \square$$

CHAPTER 2

Smoothness of Probability Laws

The basic question of this chapter is the following. Given a random variable X (or more generally, a random vector X), when has the distribution of X a density with respect to Lebesgue measure. In that case, how smooth is the density, i.e. is it continuous, differentiable, C^∞, etc.

We establish criteria for absolute continuity and smoothness in terms of the Malliavin calculus established so far. Subsequently, we apply our results to solutions of stochastic differential equations.

2.1. Absolute Continuity

Throughout this section, we fix an H-isonormal Gaussian process W and the denote the underlying probability space by $(\Omega, \Sigma, \mathbb{P})$. We assume for convenience that $\Sigma = \Sigma_W$.

The following is a simple result about the absolute continuity of the law of a random variable, based on the duality between Malliavin derivative and the Skorohod integral.

Proposition 2.1.1. *Let $X \in \mathbb{D}^{1,2}$ be such that $\|DX\|_H \neq 0$ almost surely and such that the random variable $DX/\|DX\|_H^2 \in D(\delta)$. Then the law of X is absolutely continuous with respect to Lebesgue measure. Its density is given by*

$$p(t) = \mathbb{E}\Big(\mathbb{1}_{\{X>t\}}\delta\Big(\frac{DX}{\|DX\|_H^2}\Big)\Big).$$

Proof. Let $\psi \in C_c(\mathbb{R})$ and put $\varphi(t) := \int_{-\infty}^t \psi(s)\,ds$. Then φ is continuously differentiable with bounded derivative ψ whence, by the chain rule in Proposition 1.2.7, $\varphi(X) \in \mathbb{D}^{1,2}$ and $D\varphi(X) = \psi(X)DX$. Consequently,

$$\psi(X) = \Big\langle D\varphi(X), \tfrac{DX}{\|DX\|_H^2}\Big\rangle_H$$

By the definition of the divergence,

$$\mathbb{E}\psi(X) = \mathbb{E}\Big(\varphi(X)\delta\Big(\frac{DX}{\|DX\|_H^2}\Big)\Big)$$

By approximation, the latter formula also holds for $\psi = \mathbb{1}_{(a,b)}$. Thus

$$\mathbb{P}(a < X < b) = \mathbb{E}\Big(\int_{-\infty}^X \mathbb{1}_{(a,b)}(s)\,ds \cdot \delta\Big(\frac{DX}{\|DX\|_H^2}\Big)\Big) = \int_a^b \mathbb{E}\Big(\mathbb{1}_{\{X>s\}}\delta\Big(\frac{DX}{\|DX\|_H^2}\Big)\Big)\,ds$$

by Fubini's theorem. This yields the claim. □

Example 2.1.2. Let us consider the easiest example where $X = W(h)$ for some $h \in H$ with $\|h\| = 1$, i.e. X is a standard normal random variable. In this case, $X \in \mathbb{D}^{1,2}$ with $DX = h$. Moreover, $Y = DX/\|DX\|_H^2 = h \in D(\delta)$ with $\delta(Y) = W(h) = X$. Thus in this case, the formula in Proposition 2.1.1 reduces to

$$p(t) = \mathbb{E}\big(X\mathbb{1}_{\{X>t\}}\big).$$

And indeed, using integration by parts we find

$$\mathbb{E}\big(X\mathbb{1}_{\{X>t\}}\big) = \int_t^\infty xe^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}} = \Big[-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Big|_t^\infty + 0 = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$$

Proposition 2.1.1 is a first result yielding the absolute continuity of the probability law of a random variable. Actually, one can prove absolute continuity under much weaker assumptions. This is our next goal. We need some preparation.

Intermezzo: Approximating measurable functions

In this intermezzo, we prove the following approximation result:

LEMMA 2.1.3. *Let M be a Polish space, i.e. a complete, separable metric space and μ be a finite measure on its Borel σ-algebra $\mathscr{B}(M)$. Then given a bounded, measurable function $f : M \to \mathbb{R}$, there exists a bounded sequence f_n of continuous functions on M such that $f_n \to f$ almost everywhere with respect to μ.*

Let us recall that every finite measure μ on a Polish space is *regular*, i.e. for every set $A \in \mathscr{B}(M)$ we have

$$\mu(A) = \sup\{\mu(K) : K \subset A,\, K \text{ compact}\} \quad \text{and} \quad \mu(A) = \inf\{\mu(U) : A \subset U,\, U \text{ open}\}.$$

In the proof of Lemma 2.1.3 we use the following result which is due to *Lusin*

LEMMA 2.1.4. *Let M be a Polish space, μ be a finite measure on its Borel σ-algebra and $f : M \to \mathbb{R}$ measurable. Then, given $\varepsilon > 0$ there exists a compact set $K \subset M$ with $\mu(K^c) < \varepsilon$ such that $f|_K$ is continuous.*

PROOF. Let $\mathscr{O} = \{B(q, n^{-1}) : q \in \mathbb{Q}, n \in \mathbb{N}\}$ the collection of all balls with rational centers and radii of the form n^{-1} (this is a countable basis for the topology). Let $(B_k)_{k\in\mathbb{N}}$ be an enumeration of $\mathscr{O}$. Since μ is regular, given $\varepsilon > 0$ we find compact sets K_k and open sets U_k such that

$$K_n \subset f^{-1}(B_k) \subset U_k \quad \text{and} \quad \mu(U_k \setminus K_k) < 2^{-k}\varepsilon\,.$$

The set $A := \bigcup_{k\in\mathbb{N}} U_k \setminus K_k$ is open as a union of open sets and has measure less than ε. Using again inner regularity, we find a compact set $K \subset A^c$ with $\mu(A^c \setminus K) \leq \varepsilon - \mu(A)$. It follows that $\mu(K^c) = \mu(K^c \cap A) + \mu(A^c \setminus K) \leq \mu(A) + \varepsilon - \mu(A) = \varepsilon$. Now let f_0 be the restriction of f to A^c. Then clearly, $f_0^{-1}(S) = f^{-1}(S) \cap A^c$ for any set $S \subset \mathbb{R}$. For $k \in \mathbb{N}$ we have $U_k \setminus K_k \subset A$ and thus $U_k \cap A^c = K_k \cap A^c$. It follows that

$$U_k \cap A^c = K_k \cap A^c \subset f_0^{-1}(B_k) \subset U_k \cap A^c\,.$$

This proves that the preimage of B_k under f_0 is open in A^c, thus f_0 is continuous. □

PROOF OF LEMMA 2.1.3. By Lemma 2.1.4 for $n \in \mathbb{N}$ there exists a compact subset $K_n \subset M$ such that $\mu(K_n^c) \leq n^{-1}$ and such that $f|_{K_n}$ is continuous. We assume without loss of generality that $K_n \subset K_{n+1}$. By the Tieze extension theorem, there exists a continuous function f_n such that $f_n = f$ on K_n and $\|f_n\|_\infty \leq \|f\|_\infty$. This sequence is bounded and converges pointwise to f on the set $\bigcup_n K_n$ which has full measure. □

We can now prove the following Theorem which yields absolute continuity of the law under less restrictive assumptions than Proposition 2.1.1. Note however, that in that proposition we also obtain a formula for the density, which is not the case under our weaker assumptions below.

THEOREM 2.1.5. *Let $X \in \mathbb{D}^{1,1}_{\text{loc}}$ be such that $\|DX\|_H \neq 0$ almost surely. Then the law of X is absolutely continuous with respect to Lebesgue measure.*

PROOF. By localization, we can assume that $X \in \mathbb{D}^{1,1}$. We have to show that $\mathbb{E}\mathbb{1}_E(X) = 0$ for every set $E \in \mathscr{B}(\mathbb{R})$ with Lebesgue measure zero. It actually suffices to prove this for *bounded* sets of measure zero (for then dominated convergence yields the result for all sets of measure zero). Thus, let $E \subset (-1, 1)$ be a set of Lebesgue measure zero. We denote the law of X by μ. By Lemma 2.1.3, there exists a sequence f_n of continuous functions on $[-1,1]$ which converge to $\mathbb{1}_E$ pointwise almost everywhere with respect to $\mu + dx$. We put $\varphi_n(t) = \int_{-1}^t f_n(x)\,dx$. Then φ_n is continuously differentiable. By the chain rule (Proposition 1.2.7), $\varphi_n(X) \in \mathbb{D}^{1,1}$ and $D\varphi_n(X) = f_n(X)DX$. Since $f_n \to \mathbb{1}_E$ almost everywhere with respect to Lebesgue measure, it follows that $\varphi_n(x) \to 0$ for every x. Consequently, $\varphi_n(X) \to 0$ almost surely and even in $L^1(\Omega)$. On the other hand, since $f_n \to \mathbb{1}_E$ almost everywhere with respect to μ, we have $f_n(X) \to \mathbb{1}_E(X)$ almost surely and thus $f_n(X)DX \to \mathbb{1}_E(X)DX$ almost surely. By dominated convergence, we even have $f_n(X)DX \to \mathbb{1}_E(X)DX$ in $L^1(\Omega; H)$. Since D is a closed operator, $\mathbb{1}_E(X)DX = 0$ almost surely and thus $\mathbb{1}_E(X)\|DX\|_H = 0$ almost surely. Since $\|DX\|_H \neq 0$ almost surely it follows that $\mathbb{1}_E(X) = 0$ almost surely. □

EXAMPLE 2.1.6. Consider Brownian motion $(B_t)_{t\in[0,1]}$ and let $M := \max_{t\in[0,1]} B_t$. We have seen in the exercises that $M \in \mathbb{D}^{1,2}$ and $DM = \mathbb{1}_{(0,t^*)}$ where t^* is the almost sure unique point, where $B_{t^*} = M$. Thus $\|DM\|_H = \sqrt{t^*} \neq 0$ unless $t^* = 0$. However, since $B_0 = 0$ a.s. $t^* = 0$ implies $B_t \equiv 0$ almost surely. But this has probability zero. Thus $\|DM\|_H \neq 0$ almost surely and it follows from Theorem 2.1.5 that the law of M is absolutely continuous with respect to Lebesgue measure.

Using the reflection principle (i.e. the strong Markov property) of Brownian motion, it can be shown that M has the density

$$p(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathbb{1}_{(0,\infty)}(x)$$

with respect to Lebesgue measure. This is much more information about M than we obtain from Theorem 2.1.5. However, Theorem 2.1.5 merely uses distributional information about Brownian motion.

We would also like to note that corresponding result is also true for the maximum of the so-called *Wiener sheet* which is constructed from an $L^2((0,1)^2, d\lambda_s)$-isonormal Gaussian process, see [**6**, Section 2.1.7]. In that situation the law of the maximum process is not known.

We next turn our attention to the law of a random *vector* $X = (X_1, \dots, X_d)$. Note that the requirement that the law of X has a density with respect to d-dimensional Lebesgue measure is stronger than merely requiring that the law of each X_j has a density with respect to one dimensional Lebesgue measure. Indeed, it could still happen that the vector X is concentrated on a $(d-1)$-dimensional subspace (or even a subspace of lower dimension) which is a set of measure zero with respect to d-dimensional Lebesgue measure.

As a first attempt to prove absolute continuity of the law of a vector, we try to iterate the proof of Proposition 2.1.1. Note that there we had to normalize DX by dividing by $\|DX\|_H^2$. In the multidimensional setting, we use the so-called *Malliavin matrix*.

DEFINITION 2.1.7. Let $X : \Omega \to \mathbb{R}^d$ be a random vector whose components X_j belong to $\mathbb{D}^{1,1}_{\text{loc}}$ for $j = 1, \dots, d$. The *Malliavin matrix* γ is the matrix with entries $\gamma_{ij} = \langle DX_i, DX_j\rangle_H$.

Now assume that $X : \Omega \to \mathbb{R}^d$ is a random vector whose components belong to $\mathbb{D}^{1,2}$. We further assume that the Malliavin matrix γ is almost surely invertible and the random variables $(\gamma^{-1})_{ij}DX_j$ belong to the domain of δ.

We want to repeat the proof of Proposition 2.1.1. To that end, let $\varphi \in C_b^\infty(\mathbb{R}^d)$. It follows from the chain rule, that $\varphi(X) \in \mathbb{D}^{1,2}$ and $D\varphi(X) = \sum_{j=1}^d (\partial_j\varphi)(X)DX_j$. Taking the inner product with DX_i, we find $\langle D\varphi(X), DX_i\rangle = \sum_{j=1}^d (\partial_j\varphi)(X)\langle DX_j, DX_i\rangle = \sum_{j=1}^d \gamma_{ji}(\partial_j\varphi)(X)$.

Thus, writing $x_j = (\partial_j\varphi)(X)$ and $b_i = \langle D\varphi(X), DX_i\rangle$, the above gives a system of linear equations $\gamma x = b$. Since γ is almost surely invertible, we can solve for x and obtain $x = \gamma^{-1}b$, that is

$$(\partial_j\varphi)(X) = \sum_{k=1}^{d}(\gamma^{-1})_{jk}\langle D\varphi(X), DX_k\rangle = \Big\langle D\varphi(X), \sum_{k=1}^{d}(\gamma^{-1})_{jk}DX_k\Big\rangle_H .$$

The duality between D and δ yields

$$\begin{aligned}\mathbb{E}(\partial_j\varphi)(X) &= \sum_{k=1}^{d}\mathbb{E}\big\langle D\varphi(X), (\gamma^{-1})_{jk}DX_k\big\rangle \\ &= \sum_{k=1}^{d}\mathbb{E}\Big[\varphi(X)\delta\Big((\gamma^{-1})_{jk}DX_k\Big)\Big] \\ &= \mathbb{E}\Big[\varphi(X)\delta\Big(\sum_{k=1}^{d}(\gamma^{-1})_{jk}DX_k\Big)\Big] \qquad (2.1)\end{aligned}$$

To proof the existence of a density, we would like to use the function

$$\varphi(x_1,\dots,x_d) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d}\mathbb{1}_{(a_1,b_1)\times\cdots\times(a_d,b_d)}(t_1,\dots,t_d)\,dt_1\dots dt_d\,.$$

To be more precise, we would like to find an expression for $\partial_1\dots\partial_d\varphi$, i.e. we have to iterate (2.1). The problem is that it is not clear whether the random variables $Y_j := \delta\big(\sum(\gamma_{jk}^{-1}DX_k)\big)$ are such that they belong to $\mathbb{D}^{1,2}$ and, multiplied with entries of γ^{-1}, belong to $D(\delta)$. We will tackle this problem in the next section. However, the above computations together with Proposition 2.1.12 below are enough to conclude that the distribution of X has a density with respect to Lebesgue measure. We have thus proved

PROPOSITION 2.1.8. *Let $X:\Omega\to\mathbb{R}^d$ be a random vector with components in $\mathbb{D}^{1,2}$ such that*

1. *The Malliavin matrix γ is almost surely invertible;*
2. *For every $i,j=1,\dots,d$, the random variable $(\gamma^{-1})_{ij}DX_j$ belongs to $D(\delta)$.*

Then the distribution of X has a density p with respect to d-dimensional Lebesgue measure.

Let us illustrate this before completing the proof.

EXAMPLE 2.1.9. We consider the general linear stochastic differential equation

$$\begin{cases} dX(t) &= AX(t)dt + BdW(t) \\ X(0) &= x \end{cases}$$

where X is a vector in $\mathbb{R}^m$, $A\in\mathbb{R}^{m\times m}$, $B\in\mathbb{R}^{m\times d}$ and W is an d-dimensional Brownian motion. The initial datum $x\in\mathbb{R}^d$ is nonrandom. It can be proved that the solution of the above equation is given by $X(t) = S(t)x_0 + \int_0^t S(t-s)BdW(s)$ where $S(t) := e^{tA}$ and the latter integral is a Wiener integral. Let us look at the equation

$$\begin{cases} dX(t) &= Y(t)dt \\ dY(t) &= dW(t) \end{cases}$$

which corresponds to the choice $m=2, d=1$ $A = \begin{pmatrix} 0 & 1 \\ 0 & 0\end{pmatrix}$ and $B = \begin{pmatrix} 0 \\ 1\end{pmatrix}$. Thus $S(t) = \begin{pmatrix} 1 & t \\ 0 & 1\end{pmatrix}$ and hence, by the above formula,

$$X(t) = \begin{pmatrix} x_1 + tx_2 + \int_0^t(t-s)dB_s \\ x_2 + \int_0^t dB_s \end{pmatrix}$$

Note that the components of $X(t)$ belong to $\mathbb{D}^{1,2}$, in fact, the above gives their Wiener chaos decomposition. We have

$$DX(t) = \begin{pmatrix} (t-\cdot)\mathbb{1}_{(0,t]}(\cdot) \\ \mathbb{1}_{(0,t]}(\cdot) \end{pmatrix}.$$

Thus

$$\gamma_t = \begin{pmatrix} t^3/3 & t^2/2 \\ t^2/2 & t \end{pmatrix} \quad \text{and hence} \quad \det\gamma_t = t^4/12 \neq 0$$

for $t > 0$. Consequently, γ is invertible for $t > 0$ and hence the law of the vector $X(t)$ has a density with respect to 2-dimensional Lebesgue measure.

If, in contrast, we pick $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and B as above, then the solution is $X(t) = (e^t x_1, B_t)$ and the law of this vector has no density with respect to 2-dimensional Lebesgue measure. Note that in this case the Malliavin matrix of $X(t)$ is given by $\gamma = \begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix}$ which is not invertible.

Intermezzo: Gagliardo-Nirenberg inequality

In the proof of Proposition 2.1.12, we use the following Gagliardo-Nirenberg inequality:

LEMMA 2.1.10. *Let $d > 1$. For $f \in C_c^1(\mathbb{R}^d)$ we have*

$$\|f\|_{\frac{d}{d-1}} \leq \prod_{j=1}^{d} \|\partial_j f\|_1^{\frac{1}{d}}$$

PROOF. We write $x = (x_1, \ldots, x_d)$. For $j \in \{1, \ldots, d\}$ we have

$$|f(x)| = \Big| \int_{-\infty}^{x_j} \partial_j f dx_j \Big| \leq \int_{\mathbb{R}} |\partial_j f| \, dx_j =: F_j(x).$$

Hence

$$|f(x)|^{\frac{d}{d-1}} \leq (F_1(x) \cdots F_d(x))^{\frac{1}{d-1}}.$$

Integrating with respect to x_1 over $\mathbb{R}$, noting that $F_1(x)$ is independent of x_1 and using the generalized Hölder inequality for $\sum_{j=2}^{d} \frac{1}{d-1} = 1$,

$$\begin{aligned} \int_{\mathbb{R}} |f(x)|^{\frac{d}{d-1}} \, dx_1 &\leq F_1(x)^{\frac{1}{d-1}} \int_{\mathbb{R}} F_2^{\frac{1}{d-1}} \cdots F_d^{\frac{1}{d-1}} \, dx_1 \\ &\leq F_1(x) \Big(\int_{\mathbb{R}} F_2(x) \, dx_1 \Big)^{\frac{1}{d-1}} \cdots \Big(\int_{\mathbb{R}} F_d(x) \, dx_1 \Big)^{\frac{1}{d-1}}. \end{aligned}$$

We now integrate this inequality over $\mathbb{R}$ with respect to x_2, noting that $\Big(\int_{\mathbb{R}} F_2(x) \, dx_1 \Big)^{\frac{1}{d-1}}$ is independent of x_2 and using again the generalized Hölder inequality

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x)|^{\frac{d}{d-1}} \, dx_1 \, dx_2 \leq \Big(\int_{\mathbb{R}} F_2(x) \, dx_1 \Big)^{\frac{1}{d-1}} \Big(\int_{\mathbb{R}} F_1(x) \, dx_2 \Big)^{\frac{1}{d-1}} \\ \Big(\int_{\mathbb{R}} \int_{\mathbb{R}} F_3(x) \, dx_1 \, dx_2 \Big)^{\frac{1}{d-1}} \cdots \Big(\int_{\mathbb{R}} \int_{\mathbb{R}} F_1(x) \, dx_1 \, dx_2 \Big)^{\frac{1}{d-1}} \end{aligned}$$

Iterating this until we have integrated over all variables $x_1, \ldots, x_d$, we obtain

$$\int_{\mathbb{R}^d} |f(x)|^{\frac{d}{d-1}} \, dx \leq \prod_{j=1}^{d} \Big(\int_{\mathbb{R}^d} |\partial_j f(x)| \, dx \Big)^{\frac{1}{d-1}}.$$

□

REMARK 2.1.11. Note that a corresponding inequality in dimension 1 fails, even if one is willing to allow a constant in front of the product. This is seen by considering functions f, supported in $(-n, n)$, say, which are equal to 1 on the interval $(-(n-1), n-1)$.

We now formulate and prove the following Proposition.

PROPOSITION 2.1.12. *Let μ be a finite measure on $\mathbb{R}^d$. If there exists a constant $C \geq 0$ such that for all $\varphi \in C_b^\infty(\mathbb{R}^d)$ and all $j = 1, \dots, d$ we have*

$$\Big| \int_{\mathbb{R}^d} \partial_j \varphi \, d\mu \Big| \leq C \|\varphi\|_\infty,$$

then μ is absolutely continuous with respect to Lebesgue measure. The density belongs to the space $L^{\frac{d}{d-1}}$.

PROOF. We first consider $d = 1$. Fixing $a < b$, let

$$\varphi(t) := \begin{cases} 0 & \text{if } t \leq a \\ t-a & \text{if } a < t < b \\ b-a & \text{if } t \geq b\,. \end{cases}$$

We note that φ is not infinitely differentiable. However, we can approximate this function by C_b^∞ functions such that the estimate in the proposition remains true when we use $\partial\varphi = \mathbb{1}_{[a,b]}$. We hence obtain $\mu([a, b]) \leq C(b - a)$ and this implies that μ is absolutely continuous with respect to Lebesgue measure.

Now consider the case $d > 1$. The assumption implies that the map $\varphi \mapsto \int \partial_i \varphi \, d\mu$, initially defined on $C_c^\infty(\mathbb{R})$, extends to a bounded linear functional χ_i on $C_0(\mathbb{R}^d)$. Thus, there exists a signed measure ν_i such that $\chi_i(\varphi) = \int \varphi \, d\nu_i$ for all $\varphi \in C_0(\mathbb{R}^d)$.

In order to prove that μ is absolutely continuous, we consider a standard mollifier ρ_n, say $\rho_n(x) = n^d \rho(nx)$ where $\rho(x) = c\exp(-(1 - |x|^2)^{-1})$ for $|x| < 1$ and $\rho(x) = 0$ else. Here c is chosen such that $\int \rho dx = 1$. The convolution $\rho_n * \mu$, defined by

$$\rho_n * \mu(x) := \int_{\mathbb{R}^d} \rho_n(x - y) \, d\mu(y),$$

is a bounded, infinitely differentiable function. Moreover, $\partial_\alpha(\rho_n * \mu) = (\partial_\alpha \rho_n) * \mu$. To obtain a function with compact support, we multiply with $\vartheta_m \in C_c^\infty(\mathbb{R}^d)$ which is chosen such that $\mathbb{1}_{\{|x| \leq m\}} \leq \vartheta_m \leq \mathbb{1}_{\{|x| \leq m+1\}}$. By Lemma 2.1.10

$$\|\vartheta_m(\rho_n * \mu)\|_{\frac{d}{d-1}} \leq \prod_{j=1}^d \|\partial_j \vartheta_m(\rho_n * \mu)\|_1^{\frac{1}{d}}\,.$$

Note that

$$\partial_j(\vartheta_m(\rho_n * \mu)) = (\partial_j \vartheta_m)(\rho_n * \mu) + \vartheta_n((\partial_j \rho_n) * \mu)\,.$$

It follows from the definition of ν_i that we can estimate $|(\partial_j \rho_n) * \mu| \leq |\rho_n * \nu_j|$. We thus obtain

$$\begin{aligned} \|\partial_j \vartheta_m(\rho_n * \mu)\|_1 &\leq \|\vartheta_m\|_\infty \|\rho_n * \nu_i\|_1 + \|\partial_j \vartheta_m\|_\infty \|\rho_n * \mu\| \\ &\leq \|\vartheta_m\|_\infty \|\nu_i\|_{\mathrm{TV}} + \|\partial_j \vartheta_m\|_\infty \|\mu\|_{\mathrm{TV}} =: K_j \end{aligned}$$

where K_j is a constant independent of n and m. It follows that the set $\{\vartheta_m(\rho_n * \mu) : n, m \in \mathbb{N}\}$ is a bounded subset of $L^{\frac{d}{d-1}}$ and hence, by reflexivity, relatively weakly compact.

Thus, for every $n \in \mathbb{N}$ there is a function $f_n \in L^{\frac{d}{d-1}}$ and a subsequence m_k such that $\vartheta_{m_k}(\rho_n * \mu) \rightharpoonup f_n$ as $k \to \infty$. Consequently, for $g \in C_c \subset L^d = (L^{\frac{d}{d-1}})^*$ we have

$$\int g f_n \, dx \;=\; \lim_{k\to\infty} \int g \vartheta_{m_k}(\rho_n * \mu) \, dx$$

$$= \lim_{k\to\infty} \int \rho_n * (g\vartheta_{m_k})\, d\mu = \int \rho_n * g\, d\mu.$$

For the last equality note that by dominated convergence $\rho_n * (g\vartheta_{m_k}) \to \rho_n * g$ pointwise. Moreover, $\|\rho_n * (g\vartheta_{m_k})\|_\infty \leq \|\rho_n\|_\infty \|g\|_1$. Thus the equality follows from applying dominated convergence again.

Now note that the sequence f_n is also bounded in $L^{\frac{d}{d-1}}$. Thus there is a function $f \in L^{\frac{d}{d-1}}$ and a subsequence n_k such that $f_{n_k} \rightharpoonup f$. Now let $g \in C_c$. Since g is uniformly continuous on its compact support we find, given $\varepsilon > 0$ an $n \in \mathbb{N}$ such that $|g(x) - g(y)| \leq \varepsilon$ whenever $|x - y| \leq n^{-1}$. Thus

$$|\rho_n * g(x) - g(x)| \leq \int \rho_n(y)|g(x-y) - g(x)|\, dy \leq \varepsilon \int \rho_n(y)\, dy = \varepsilon\,.$$

This shows that $\rho_n * g \to g$ uniformly. Altogether,

$$\int gf\, dx = \lim_{k\to\infty} \int gf_{n_k}\, dx = \lim_{k\to\infty} \int \rho_{n_k} * g\, d\mu = \int g\, d\mu\,.$$

This finishes the proof. □

Also in the case of random vectors, absolute continuity of the law can be proved under weaker assumptions. The proof, however, is not a generalization of the one-dimensional proof. We quote the result, due to *Bouleau* and *Hirsch* without proof.

THEOREM 2.1.13. *Let $X = (X_1, \ldots, X_d)$ be a random vector such that $X_j \in \mathbb{D}^{1,p}_{\mathrm{loc}}$ for some $p > 1$ and $j = 1, \ldots, d$ and such that the Malliavin matrix γ is almost surely invertible. Then the law of X is absolutely continuous with respect to d-dimensional Lebesgue measure.*

2.2. Smoothness of the Density

We now return to the question of finding a formula for the density of a random vector which was left open in the last section. As we have noted there, we need regularity of Skorohod integrals to iterate the one-dimensional proof to find such a formula. To that end, given a Hilbert space V, we define the space $\mathbb{D}^\infty(V)$ by

$$\mathbb{D}^\infty(V) := \bigcap_{k\geq 1} \bigcap_{p\geq 1} \mathbb{D}^{k,p}(V),$$

that is $\mathbb{D}^\infty(V)$ consists of those random elements who have Malliavin derivatives of all orders and these derivatives lie in all L^p spaces. Note that $\mathscr{S}(V) \subset \mathbb{D}^\infty$. In the case where $V = \mathbb{R}$, we merely write $\mathbb{D}^\infty$. The following is easy to see:

LEMMA 2.2.1. *Suppose that $X = (X_1, \ldots, X_n)$ has components in $\mathbb{D}^\infty$ and that $\varphi \in C_p^\infty(\mathbb{R}^n)$. Then $\varphi(X) \in \mathbb{D}^\infty$ and*

$$D\varphi(X) = \sum_{j=1}^n (\partial_j \varphi)(X) DX_j\,.$$

In particular, chosing $n = 2$ and $\varphi(x, y) = xy$, it follows that $\mathbb{D}^\infty$ is an algebra. We now obtain the following result about the inverse of a matrix with entires in $\mathbb{D}^\infty$.

LEMMA 2.2.2. *Let $M = (M_{ij})$ is an $n \times n$ matrix with entries in $\mathbb{D}^\infty$. Moreover, assume that M is almost surely invertible and that $(\det M)^{-1} \in L^p(\Omega)$ for all $p \geq 1$. Then the inverse $M^{-1} = ((M^{-1})_{ij})$ has entries in $\mathbb{D}^\infty$. Moreover, for $i, j \in \{1, \ldots, n\}$ we have*

$$D(M^{-1})_{ij} = -\sum_{k,l=1}^n (M^{-1})_{ik}(M^{-1})_{lj} DM_{kl}\,.$$

PROOF. As a consequence of Cramer's rule, the entries of $\det M \cdot M^{-1}$ are polynomials in the entries of M, hence elements of $\mathbb{D}^\infty$ by Lemma 2.2.1. Moreover, $\det M$, being itself a polynomial in the entries of M is also an element of $\mathbb{D}^\infty$. To finish the proof, it suffices to show that $(\det M)^{-1} \in \mathbb{D}^\infty$.

To that end, let us first show that $\det M$ is either almost surely positive or almost surely negative. To see this, let $\psi_n = n\mathbb{1}_{(0,1/n)}$ and $\varphi_n(t) = \int_{-\infty}^t \psi_n(s)\,ds$. Then $\phi_n(\det M) \in \mathbb{D}^{1,2}$ by the chain rule. Moreover, $D\phi_n(\det M) = n\mathbb{1}_{(0,1/n)}(\det M)D(\det M)$ (well, not exactly, but we merely need the following estimate which can be proved rigorously, approximating ψ_n). We have $\|D\varphi_n(\det M)\|_H \leq \|X^{-1}DX\|_H \leq |\det M|^{-2}\|D(\det M)\|_H^2$ which is integrable by assumption. It follows that $\|D\varphi_n(\det M)\|_H$ is uniformly bounded in $L^2(\Omega; H)$. Since $\varphi_n(\det M) \to \mathbb{1}_{(0,\infty)}(\det M)$ pointwise and also in L^2, Lemma 1.2.8 yields that $\mathbb{1}_{(0,\infty)}(\det M) \in \mathbb{D}^{1,2}$. Now Corollary 1.2.19 implies that $\mathbb{P}(\det M > 0) \in \{0, 1\}$.

We assume without loss of generality that $\det M > 0$ almost surely. The function $\varphi_n(t) := (t + n^{-1})^{-1}$, initially defined for $t > 0$ can be extended to a function in $C_p^\infty(\mathbb{R})$. Hence, by Lemma 2.2.1, $\varphi_n(\det M) \in \mathbb{D}^\infty$. Note that $\varphi_n(\det M) \to (\det M)^{-1}$ almost surely. Moreover, using that $(\det M)^{-1} \in \bigcup L^p$, we see that $\varphi_n(\det M) \to (\det M)^{-1}$ in every L^p, for $p \geq 1$. Next note that

$$D\varphi_n(\det M) = -(\det M + n^{-1})^{-2}D(\det M) \to -(\det M)^{-1}D(\det M)$$

pointwise and also in L^p as is easy to see using our assumption. By the closedness of the Malliavin derivative, $(\det M)^{-1} \in \mathbb{D}^{1,p}$ for every $p \geq 1$. In general, the k-th derivative of $\varphi_n(\det M)$ can be written as a sum of terms which are multiples products of derivatives of $\varphi_n(\det M)$ and higer order Malliavin derivatives of $\det M$. Also these converge in $L^p(\Omega; H^{\otimes k})$ so that by closedness we obtain inductively that $(\det M)^{-1} \in \mathbb{D}^{k,p}$ for every $k, p \geq 1$, i.e. $(\det M)^{-1} \in \mathbb{D}^\infty$.

Finally, the formula in the Lemma follows by differentiating the equality $M^{-1}M = I$. □

We next state a result about the regularity of Skorohod integrals that we will use. The proof will be given in Chapter 3.

THEOREM 2.2.3. *Let $u \in \mathbb{D}^\infty(H)$. Then $u \in D(\delta)$ and $\delta(u) \in \mathbb{D}^\infty$.*

We can now iterate the arguments in the previous section. We use the following notation for partial derivatives. If $f : \mathbb{R}^d \to \mathbb{R}$ is a sufficiently smooth function and $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ is a multiindex, then $\partial_\alpha f$ denotes the partial derivative

$$\partial_\alpha f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdot \ldots \cdot \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}} f$$

PROPOSITION 2.2.4. *Let $X = (X_1, \ldots, X_d)$ be such that $X_j \in \mathbb{D}^\infty$ for $j = 1, \ldots, d$ and such that the Malliavin matrix γ satisfies $(\det\gamma)^{-1} \in L^p$ for all $p \geq 1$. Then for $Y \in \mathbb{D}^\infty$ and a multiindex $\alpha \in \mathbb{N}_0^d$, there exists a random variable $Z_\alpha = Z_\alpha(X, Y)$ in $\mathbb{D}^\infty$ such that for every $\varphi \in C_p^\infty(\mathbb{R}^d)$ we have*

$$\mathbb{E}\big[(\partial_\alpha\varphi)(X)Y\big] = \mathbb{E}\big[\varphi(X)Z_\alpha\big]\,.$$

Moreover, the random variables Z_α are defined as follows. If $\alpha = e_j$, then

$$Z_\alpha = \delta\Big(\sum_{k=1}^d (Y\gamma^{-1})_{jk}DX_k\Big)\,.$$

For general α we have

$$Z_{\alpha+e_j}(X, Y) = Z_{e_j}(X, Z_\alpha(X, Y))\,.$$

Proof. If $\alpha = e_j$, then the same argument as in the last section, cf. equation (2.1), yields that $\mathbb{E}\big[(\partial_\alpha\varphi)(X)Y\big] = \mathbb{E}\big[\varphi(X)Z_{e_j}\big]$ for Z_{e_j} as in the statement of the proposition. Lemma 2.2.2 yields that $\sum_{k=1}^d (Y\gamma^{-1})_{jk}DX_k \in \mathbb{D}^\infty(H)$. Thus $Z_{e_j} \in \mathbb{D}^\infty$. This shows that the thesis is true for $\alpha = e_j$. The general case follows by induction on $|\alpha|$, repeating the above arguments. □

Definition 2.2.5. A random vector $X = (X_1, \dots, X_d)$ is called *nondegenerate* if $X_j \in \mathbb{D}^\infty$ for $j = 1, \dots, d$ and the Malliavin matrix γ is almost surely invertible with $(\det\gamma)^{-1} \in L^p$ for all $p \geq 1$.

We have now all probabilistic tools at hand to state and prove the main result of this section. We will also need some facts about the Fourier transform and Schwartzfunctions which we recall next.

Intermezzo: Schwartz Functions and the Fourier Transform

Definition 2.2.6. A function $f : \mathbb{R}^d \to \mathbb{C}$ is called *rapidly decreasing*, if

$$\lim_{|x|\to\infty} x^\alpha f(x) = 0$$

for all multiindices $\alpha \in \mathbb{N}_0^d$. Here $x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_d^{\alpha_d}$. A function is called a *Schwartz function* if it is infinitely differentiable and the function and all its partial derivatives are rapidly decreasing. The space of all Schwarz functions is denoted by $\mathscr{S}(\mathbb{R}^d)$.

Obviously, the testfunctions $C_c^\infty(\mathbb{R}^d)$ are Schwartz functions. Also $f(x) := \exp(-|x|^2)$ is a Schwartz function.

It is easy to see that a C^∞-function f is rapidly decreasing if and only if

$$\sup_{x\in\mathbb{R}^d} (1 + |x|^m)|\partial_\alpha f| < \infty$$

for all $m \in \mathbb{N}$ and $\alpha \in \mathbb{N}_0^d$. In particular, Schwarz functions belong to $L^p(\mathbb{R}^d)$ since they can be majorized by a multiple of the function $x \mapsto (1 + |x|^m)^{-1}$ which is p-integrable if m is large enough. As the testfunctions are Schwartz functions, $\mathscr{S}(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d)$.

Definition 2.2.7. For $f \in L^1(\mathbb{R}^d)$, its *Fourier transform* $\mathscr{F}f : \mathbb{R}^d \to \mathbb{C}$ is defined by

$$(\mathscr{F}f)(\xi) = \int_{\mathbb{R}^d} f(x)e^{-ix\xi}\,dx$$

where $x\xi := \sum_j x_j\xi_j$.

Clearly, for $f \in L^1$ the Fourier transform is well-defined and a bounded function on $\mathbb{C}$. We even have

Proposition 2.2.8. *(Riemann-Lebesgue Lemma)*
The Fourier transform defines a bounded linear operator from $L^1(\mathbb{R}^d)$ to $C_0(\mathbb{R}^d)$.

Proof. The continuity of $\mathscr{F}f$ for $f \in L^1$ is an easy consequence of the dominated convergence theorem. Obviously, $\|\mathscr{F}f\|_\infty \leq \|f\|_1$, whence $\mathscr{F} \in \mathscr{L}(L^1(\mathbb{R}^d); C_b(\mathbb{R}^d))$. It remains fo show that for $f \in L^1(\mathbb{R}^d)$ we have $(\mathscr{F}f)(\xi) \to 0$ as $|\xi| \to \infty$. In fact, by the boundedness of $\mathscr{F}$ and the closedness of C_0 in C_b, it suffices to show this convergence for f in a dense subspace of L^1, say $f \in C_c^\infty(\mathbb{R}^d)$.

Let such an f be given and let $|\xi| \geq R$. Then there is an index j with $|\xi_j| \geq R/\sqrt{d}$. Integration by parts yields

$$|\mathscr{F}f(\xi)| = \Big|\int_{\mathbb{R}^d} \partial_j f \frac{1}{-i\xi_j} e^{-ix\xi}\,dx\Big| \leq \frac{\sqrt{d}}{R}\max_j \|\partial_j f\|_1 \to 0$$

as $R \to \infty$. □

We want to show that $\mathscr{F}$ maps $\mathscr{S}(\mathbb{R}^d)$ bijectively onto itself. Note that if $f \in \mathscr{S}(\mathbb{R}^d)$, then for every multiindex α also the functions $\partial_\alpha f$ and $x^\alpha f : x \mapsto x^\alpha f(x)$ belong to $\mathscr{S}(\mathbb{R}^d)$, which follows immediately from the definition.

LEMMA 2.2.9. *Let $f \in \mathscr{S}(\mathbb{R}^d)$. Then $\mathscr{F} \in C^\infty(\mathbb{R}^d)$. Moreover, for every multiindex α*
(1) $\partial_\alpha \mathscr{F} f = (-i)^{|\alpha|}\mathscr{F}(x^\alpha f)$.
(2) $\mathscr{F}(\partial_\alpha f)(\xi) = i^{|\alpha|}\xi^\alpha(\mathscr{F} f)(\xi)$.

PROOF. (1) Differentiability of parameter integrals is a standard application of the dominated convergence theorem. Here, we have

$$\begin{aligned}\partial_\alpha \mathscr{F} f &= \partial_\alpha \int_{\mathbb{R}^d} f(x)e^{-ix\xi}\,dx = \int_{\mathbb{R}^d} f(x)\partial_\alpha e^{-ix\xi}\,dx \\ &= (-i)^{|\alpha|}\int_{\mathbb{R}^d} x^\alpha f(x)e^{-ix\xi}\,dx = (-i)^{|\alpha|}\mathscr{F}(x^\alpha f)\end{aligned}$$

where we can interchange differentiation and integration since $x^\alpha f(x)$ is integrable as a Schwartz function.

(2) By integration by parts (note that all boundary terms vanish as $x^\beta f \in \mathscr{S}(\mathbb{R}^d)$) we obtain

$$\mathscr{F}(\partial_\alpha f)(\xi) = \int_{\mathbb{R}^d}(\partial_\alpha f)(x)e^{-ix\xi}\,dx = (-1)^{|\alpha|}f(x)\partial_\alpha e{-}ix\xi\,dx = (-1)^{|\alpha|}(-i)^{|\alpha|}\xi^\alpha(\mathscr{F} f)(\xi)\,.$$

□

COROLLARY 2.2.10. *For $f \in \mathscr{S}(\mathbb{R}^d)$ we have $\mathscr{F} f \in \mathscr{S}(\mathbb{R}^d)$.*

PROOF. We know already that $\mathscr{F} f \in C^\infty$. Thus we need to show that $x^\alpha \partial_\beta f$ is rapidly decreasing for all multiindices α, β. As a consequence of Lemma 2.2.9,

$$\xi^\alpha \partial_\beta(\mathscr{F} f) = i^{-|\alpha|}(-i)^{|\beta|}\mathscr{F}(\partial_\alpha x^\beta f)(\xi)\,.$$

By the Riemann-Lebesgue Lemma, the latter is an element of C_0. □

We can now also compute the density of the standard normal density $\gamma(x) := c_d e^{-|x|^2/2}$ where $c_d := (2\pi)^{-d/2}$ is chosen such that $\int_{\mathbb{R}^d}\gamma(x)\,dx = 1$.

COROLLARY 2.2.11. *We have $(\mathscr{F}\gamma)(\xi) = c_d^{-1}\gamma(\xi)$.*

PROOF. First let $d = 1$. Then γ solves the ODE $y' = -xy$ with initial datum $y(0) = c_1$. By Lemma 2.2.9

$$0 = \mathscr{F}(\gamma' + x\gamma) = i\xi\mathscr{F}\gamma + i(\mathscr{F}\gamma)'$$

that is $\mathscr{F}\gamma$ solves the same ODE. Actually, $c_d\mathscr{F}\gamma$ also solves that ODE and, moreover, $c_d\mathscr{F}\gamma)(0) = c_d$. Thus, as the solution of that ODE is unique, $c_d\mathscr{F}\gamma = \gamma$.

In the case $d > 1$ we use Fubini's theorem to obtain

$$\mathscr{F}\gamma(\xi) = \prod_{j=1}^d \int_{\mathbb{R}} e^{-x_j^2/2}e^{-ix_j\xi_j}\,dx_j = \prod_{j=1}^d (2\pi)^{\frac{1}{2}}e^{-\xi_j^2/2} = c_d\gamma(\xi)\,.$$

□

LEMMA 2.2.12. *For $f \in \mathscr{S}(\mathbb{R}^d)$ we have $\mathscr{F}\mathscr{F} f(x) = c_d^2 f(-x)$.*

PROOF. First note that for Schwartz functions f also the Fourier transform $\mathscr{F} f$ is a Schwartz function. In particular, $\mathscr{F}\mathscr{F} f$ is well-defined. However, when trying to evaluate the double integral via Fubini's theorem, we end up with a divergent integral. Hence we use the following trick.

We set $g(x) := e^{-ix\xi_0}\gamma(rx)$ where $\xi_0 \in \mathbb{R}^d$ and $r > 0$ are parameters. Then we have

$$\mathscr{F} g(\xi) = \int_{\mathbb{R}^d} e^{-ix\xi_0}\gamma(rx)e^{-ix\xi}\,dx = (\mathscr{F}\gamma_r)(\xi_0 + \xi)$$

where we write $\gamma_r(x) := \gamma(rx)$. Using this and Fubini's theorem, we find

$$\begin{aligned}\int_{\mathbb{R}^d} (\mathscr{F} f)(\xi) g(\xi)\, d\xi &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x) g(\xi) e^{-ix\xi}\, dx\, d\xi \\ &= \int_{\mathbb{R}^d} f(x)(\mathscr{F} g)(x)\, dx = \int_{\mathbb{R}} f(x)(\mathscr{F}\gamma_r)(x+\xi_0) \\ &= \int_{\mathbb{R}^d} f(x)\frac{1}{r^d}(\mathscr{F}\gamma)(r^{-1}(x+\xi_0))\, dx \\ &= c_d^{-1}\int_{\mathbb{R}^d} f(ry-\xi_0)\gamma(y)\, dy\,.\end{aligned}$$

Here we have used the fact that $(\mathscr{F}(\gamma_r))(\xi) = r^{-d}\mathscr{F}\gamma(r^{-1}\xi)$, which easily follows from substituting $z = rx$ in the integral defining $\mathscr{F}$. In the last equality we have substituted $y = r^{-1}(x+\xi_0)$.

Now we let $r \to 0$. From dominated convergence, we obtain

$$\int_{\mathbb{R}^d} (\mathscr{F} f)(\xi) g(\xi)\, d\xi = \int_{\mathbb{R}} (\mathscr{F} f)(\xi) e^{-i\xi\xi_0}\gamma(r\xi)\, d\xi \to \gamma(0)\int_{\mathbb{R}^d} (\mathscr{F} f)(\xi) e^{-i\xi\xi_0}\, d\xi = c_d(\mathscr{F}\mathscr{F} f)(\xi)$$

as $n \to \infty$ and that

$$c_d\int_{\mathbb{R}^d} f(ry-\xi_0)\gamma(y)\, dy \to c_d\int_{\mathbb{R}^d} f(-\xi_0)\gamma(y)\, dy = c_d f(-\xi_0)$$

as $n \to \infty$. Now the thesis follows. □

THEOREM 2.2.13. *The Fourier transformation is a bijection from $\mathscr{S}(\mathbb{R}^d)$ onto itself.*

PROOF. By Lemma 2.2.12, $\mathscr{F}^4 = c_d^4 id_{\mathscr{S}(\mathbb{R}^d)}$. This implies that $\mathscr{F}$ is invertible with $\mathscr{F}^{-1} = c_d^{-4}\mathscr{F}^3$. □

This ends our brief digression.

The following is the main result of this Section.

THEOREM 2.2.14. *Let $X = (X_1, \ldots, X_d)$ be a nondegenerate vector. Then the law of X has a density $p \in \mathscr{S}(\mathbb{R}^d)$ with respect to d-dimensional Lebesgue measure.*

PROOF. As a consequence of Proposition 2.1.8, the law of X has a density p with respect to d-dimensional Lebesgue measure. We denote the Fourier transform of p by u. Thus

$$u(\xi) = \int_{\mathbb{R}^d} p(x) e^{-ix\xi}\, dx = \mathbb{E} e^{-iX\xi} =: \mathbb{E}\varphi_\xi(X)$$

where $\varphi_\xi(x) = \exp(-ix\xi)$. If $\Delta = \sum_{j=1}^d \partial_j^2$ denotes the Laplace operator, then $\Delta^k\varphi_\xi = |\xi|^{2k}\varphi_\xi$. It follows that

$$|\xi|^{2k} u(\xi) = \int_{\mathbb{R}^d} p(x)\Delta^k\varphi_\xi(x)\, dx = \mathbb{E}\Delta^k\varphi_\xi(X)\,.$$

As a consequence of Proposition 2.2.4, there exists a random variable $Z \in \mathbb{D}^\infty$ such that the latter equals $\mathbb{E}\big[\varphi_\xi(X) Z\big]$. It follows that $|\xi|^{2k}|u(\xi)| \le \mathbb{E}|Z|$ which implies that u is rapidly decreasing.

To take care of the derivatives, first note that $\mathbb{E}\prod |X_j|^{\alpha_j} < \infty$ since the coordinates X_j belong to any L^p. It follows that polynomials in d variables are integrable with respect to p. Now a well-known result about differentiability of parameter dependent integrals yields that u is a differentiable function of ξ and for any multiindex α we have

$$\partial_\alpha u(\xi) = \int_{\mathbb{R}^d} p(x)(-ix)^\alpha\varphi_x i(x)\, dx$$

Combining this with the above,

$$|\xi|^{2k}\partial_\alpha u(\xi) = \int_{\mathbb{R}^d} p(x)(-ix)^\alpha \Delta^k \varphi_\xi(x)\,dx = \mathbb{E}(\Delta_k\varphi_\xi)(X)(-iX)^\alpha = \mathbb{E}\varphi_\xi(X)\tilde{Z}$$

for a certain $\tilde{Z} \in \mathbb{D}^\infty$. It follows that also the partial derivatives are rapidly decreasing, hence $u \in \mathscr{S}(\mathbb{R}^d)$. By Theorem 2.2.13 it follows that $p \in \mathscr{S}(\mathbb{R}^d)$. □

2.3. Stochastic Differential Equations

The main application of the results established so far is to solutions of stochastic differential equations. Here, our underlying isonormal Gaussian process is of the form $H = L^2((0,\infty);\mathbb{R}^d)$ (or $H = L^2((0,\tau);\mathbb{R}^d)$ if we consider a finite time horizon τ) so that we can consider a d-dimensional Brownian motion $B_t := (B_t^1,\dots,B_t^d)$ on our underlying Probability space.

We will also consider the *natural filtration* $\tilde{\mathbb{F}} = \sigma(B_s^j : s \le t, j = 1,\dots,d)$. Actually, the theory of stochastic differential equations is based on the Itô integral and uses heavily the fact that Itô stochastic integrals are martingales. To avoid technical difficulties about measurability, it is customary to assume that the given filtration satisfies the *usual conditions*, i.e. $\mathscr{F}_0$ contains the $\mathbb{P}$-null sets and the filtration is right-continuous. It is well-known (see, e.g., Section 2.7 in [**4**]) that the augmentation $\mathbb{F}$ of the Brownian filtration by the null sets is right-continuous, it thus satisfies the usual conditions.

This additional assumption does not pose difficulties for the Malliavin calculus. Indeed, the basic result involving measurability, Theorem 1.1.11 about the Wiener chaos decomposition, remains valid when Σ_W is replaced with its augmentation by the $\mathbb{P}$-null sets.

Intermezzo: SDE – Existence and uniqueness of solutions

In this intermezzo, we recall some standard results about stochastic differential equations. For more information, we refer the reader to standard literature, e.g. [**4, 7**].

We will consider m-dimensional stochastic differential equations of the form

$$\begin{cases} dX(t) &= B(t,X(t))\,dt + \sum_{j=1}^d V_j(t,X(t))\,dB_t^j \\ X(0) &= x_0 \end{cases}$$

A *solution* of this equation is always understood in integrated form, i.e. a solution is an m-dimensional stochastic process X so that for every $t \ge 0$ we have

$$X(t) = x_0 + \int_0^t B(s,X(s))\,ds + \sum_{j=1}^d \int_0^t V_j(s,X(s))\,dB_s^j$$

almost surely where the latter are Itô integrals. The process X has to be such that both integrals are well-defined. As a matter of fact, we will only be interested in equations where there is a solution X which has *continuous paths* (which, together with a continuity assumption on B will yields that the deterministic integral is well-defined pathwise) and also belongs to the space $L^2_{\mathbb{F}}((0,\infty)\times\Omega;\mathbb{R}^m)$ (which together with a continuity assumption on the V_j's will yield that the integrands in the stochastic integrals belongs to $L^2_{\mathbb{F}}((0,\infty)\times\Omega;\mathbb{R}^m)$, hence to the domain of the Itô integral, see Section 1.7.

The proof of existence and uniqueness of solutions to stochastic differential equations is an application of Banach's fixed point theorem. However, it also uses additional properties of the (one-dimensional) Itô integral, namely that if $\phi \in L^2_{\mathbb{F}}((0,T)\times\Omega)$ then the integral process $M(t) := \int_0^t \phi(s)\,dB_s$ is a continuous martingale with quadratic variation process $\langle M\rangle_t = \int_0^t \|\phi(s)\|^2\,ds$. Let us recall that the quadratic variation of a continuous local martingale M is the unique adapted and increasing process $\langle M\rangle$ such that $M^2 - \langle M\rangle$ is a local martingale.

The martingale property allows us to employ the following results in the proof of existence and uniqueness.

THEOREM 2.3.1. *(Doob's maximal inequality;* [**3**, Proposition 7.16]*)*
Let $(M(t))_{t\in[0,T]}$ *be a continuous martingale on some stochastic basis and* $p \in (1,\infty)$*. Then*

$$\mathbb{E}\big(\sup_{t\in[0,T]} |M(t)|^p\big) \le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}|M(T)|^p\,.$$

THEOREM 2.3.2. *(Burkholder-Davies-Gundy inequality;* [**3**, Theorem 17.7]*)*
Let $p \in (0,\infty)$*. Then there exist constants* $c_p, C_p > 0$*such that for every continuous martingale* $(M(t))_{t\in[0,T]}$ *we have*

$$c_p\mathbb{E}\langle M\rangle_T^{\frac{p}{2}} \le \mathbb{E}\sup_{t\in[0,T]} |M(t)|^p \le C_p\mathbb{E}\langle M\rangle_T^{\frac{p}{2}}\,.$$

We now make the following assumption

HYPOTHESIS 2.3.3. *Let,* $d,m\in\mathbb{N}$*,* $\tau>0$ *and* $B:(0,\tau)\times\mathbb{R}^m\to\mathbb{R}^m$ *and, for* $j=1,\ldots,d$*,* $V_j:(0,\tau)\times\mathbb{R}^m\to\mathbb{R}^m$ *be measurable maps such* $|B(\cdot,0)|$ *and* $|V_j(\cdot,0)|$ *are bounded, say by* C *and there exists a constant* L *with*

$$|B(s,x)-B(s,y)| + \sum_{j=1}^d |V_j(s,x)-V_j(s,y)| \le L|x-y| \quad \textit{for all } s\in(0,\tau) \textit{ and } x,y\in\mathbb{R}^m.$$

THEOREM 2.3.4. *Assume Hypothesis 2.3.3. Moreover, let* $p\ge 2$ *and* $x_0\in\mathbb{R}^d$ *Then there exists a unique continuous, adapted process* $(X(t))_{t\in[0,\tau]}$ *such that for every* $t\in[0,\tau]$ *we have*

$$X(t) = x_0 + \int_0^t B(s,X(s))\,ds + \sum_{j=1}^d\int_0^t V_j(s,X(s))\,dB_s^j$$

almost surely. Moreover, there exists a constant C*, depending only on* p,T,x_0,C *and* L *such that*

$$\mathbb{E}\sup_{t\in[0,T]}|X(t)|^p \le C\,.$$

PROOF. We consider the Banach space $E = L^p_{\mathbb{F}}(\Omega; C([0,T];\mathbb{R}^m))$ of continuous, adapted $\mathbb{R}^m$-valued process which are p-integrable. We claim that the map Φ, defined by

$$\Phi(X)(t) := x_0 + \int_0^t B(s,X(s))\,ds + \sum_{j=1}^d\int_0^t V_j(s,X(s))\,dB_s^j$$

maps E to itself. Indeed, the constant x_0 belongs to E. The deterministic integral is clearly continuous and adapted. Moreover,

$$\begin{aligned}\mathbb{E}\sup_{t\in[0,T]}\Big|\int_0^t B(s,X(s))\,ds\Big|^p &\le T^{p-1}\mathbb{E}\int_0^T |B(s,X(s))|^p\,ds\\ &\le T^{p-1}\mathbb{E}\int_0^T (C+L|X(s)|)^p\,ds < \infty\,.\end{aligned}$$

As for the stochastic integral, first note that as a consquence of our assumptions, the integrand has components in $L^2_{\mathbb{F}}((0,T)\times\Omega)$ whence the stochastic integral defines a continuous martingale. As a consequence of (vector valued versions of) Doob's maximal inequality and the Burkholder-Davies-Gundy inequality,

$$\begin{aligned}\mathbb{E}\sup_{t\in[0,T]}\Big|\int_0^t V_j(s,X(s))\,dB_s^j\Big|^p &\le \Big(\frac{p}{p-1}\Big)^p\mathbb{E}\Big|\int_0^T V_j(s,X(s))\,dB_s^j\Big|^p\\ &\le C_p'\mathbb{E}\Big(\int_0^T |V_j(s,X(s))|^2\,ds\Big)^{\frac{p}{2}}\end{aligned}$$

$$\begin{aligned} &\leq C'_p T^{\frac{p}{2}-1}\mathbb{E}\int_0^T |V_j(s,X(s))|^p\,ds \\ &\leq C'_p T^{\frac{p}{2}-1}\mathbb{E}\int_0^T (C+L|X(s)|)^p\,ds < \infty. \end{aligned}$$

Altogether, this shows that Φ maps E to itself as claimed.

Arguing similarly, i.e. considering the deterministic and the stochastic integral separately where for the stochastic integral we use Doob's maximal inequality and the Burkholder-Davies-Gundy inequality, one shows that there is a constant K_p such that for $0 < r \leq T$ we have

$$\mathbb{E}\sup_{t\in[0,r]}|\Phi(X)(t)-\Phi(Y)(t)|^p \leq K_p r^{p-1}\int_0^r \mathbb{E}|X(s)-Y(s)|^p\,ds\,.$$

This implies inductively that

$$\|\Phi(X)-\Phi(Y)\|_E \leq \frac{K_p^n T^n}{n!}\|X-Y\|_E\,.$$

Now Banach's fixed point theorem implies that Φ has a unique fixed point X_*. □

This ends our brief intermezzo. We now return to the problem of proving that solutions of stochastic differential equations belong to $\mathbb{D}^{1,\infty} := \bigcap_{p\geq 1}\mathbb{D}^{1,p}$.

The basic strategy for the proof is as follows. The proof of Theorem 2.3.4 shows that the solutions of a SDE can be approximated (in L^p) by the Picard iteration. Thus to proof that the solution belongs to $\mathbb{D}^{1,\infty}$ it suffices to show that the Picard approximations belong to $\mathbb{D}^{1,p}$ for every p and that the derivatives are bounded in $L^p(\Omega;H)$. In this case Lemma 1.2.8 yields the claim.

To prove that the Picard iterates are smooth, one proceeds inductively and uses the Lipschitz condition on the coefficients and the chain rule 1.2.9 and Proposition 1.7.2 about the differentiability of Itô integrals. Note, that the latter is a Hilbert space result, i.e. it can only be used for $p = 2$. To overcome this, we will also use the following result, which will be proved in the next chapter.

PROPOSITION 2.3.5. *Let $p > 1$ and $X \in \mathbb{D}^{1,1+\varepsilon}$ for some $\varepsilon > 0$ be such that $X \in L^p(\Omega)$ and $DX \in L^p(\Omega;H)$. Then $X \in \mathbb{D}^{1,p}$.*

We now have the following result about differentiability of solutions of stochastic differential equations.

THEOREM 2.3.6. *Assume Hypothesis 2.3.3, let $x_0 \in \mathbb{R}^m$ and let $(X(t))_{t\in[0,\tau]}$ be the unique solution of the SDE*

$$X(t) = x_0 + \int_0^t B(s,X(s))\,ds + \sum_{j=1}^d \int_0^t V_j(s,X(s))\,dB_s^j$$

which exists as a consequence of Theorem 2.3.4. Then the components $X_j(t)$ belong to $\mathbb{D}^{1,\infty}$ for $j = 1,\dots,m$ and we have

$$\sup_{0\leq r\leq t}\mathbb{E}\sup_{r\leq s\leq\tau}\left|D_r^k X_j(s)\right|^p < \infty$$

for all p and $k = 1,\dots,m$.

Moreover, there exists uniformly bounded and adapted m-dimensional processes a_{kl} and b_k, for $k = 1,\dots,m$ and $l = 1,\dots,d$, such that the derivative $D^jX(t)$ satisfies the stochastic equation

$$D_r^j X(t) = V_j(r,X(r)) + \sum_{k=1}^m \int_r^t b_k(s)D_r^j X_k(s)\,ds + \sum_{k=1}^m\sum_{l=1}^d \int_r^t a_{kl}(s)D_r^j X_k(s)\,dB_s^l\,.$$

In the case where the coefficients V_j and B are continuously differentiable, the processes a_{kl} and b_k are given by

$$a_{kl}(s) := (\partial_k V_l)(s, X(s)) \quad \text{and} \quad b_k = (\partial_k B)(s, X(s))\,.$$

PROOF. We consider the Picard iterates X_n, defined inductively by $X_0 \equiv x_0$ and $X_{n+1} = \Phi(X_n)$ where Φ is as in the proof of Theorem 2.3.4.

We claim that components of $X_n(t)$ belong to $\mathbb{D}^{1,\infty}$ for all n and $t \in (0,\tau)$. Moreover, if we set $\psi_n(t) := \sup_{0\le r\le t} \mathbb{E}\sup_{s\in[r,t]} |D_r X_n(s)|^p$, we claim that $\psi_n(t) < \infty$ and there are constants c_1, c_2 such that $\psi_{n+1}(t) \le c_1 + c_2 \int_0^t \psi_n(s)\,ds$.

This is certainly true for the constant x_0. Now assume that it is true for $X_n(t)$ for all $t \in (0,\tau)$. By the chain rule for Lipschitz functions in Proposition 1.2.9, $B(s, X_n(s))$ and $V_j(s, X_n(s))$ belong to $\mathbb{D}^{1,\infty}$. Moreover, there exist random vectors $b^{n,i,k}(s)$ and $a^{n,i,j,k(s)}$ which are bounded by L such that

$$D_r\big[B^i(s, X_n(s))\big] = \sum_{k=1}^m b^{n,i,k}(s) D_r X_n^k(s) \quad \text{and} \quad D_r\big[V_j^i(s, X(s))\big] = \sum_{k=1}^m a^{n,i,j,k}(s) D_r X_n^k(s)\,.$$

The proof of Proposition 1.2.9 shows that $b^{n,i,k}$ and $a^{n,i,j,k}$ are obtained as a weak limit of $\mathscr{F}_s$-measurable random variables and are thus $\mathscr{F}_s$-measurable themselves. It follows that the processes $(D_r B(s, X_n(s)))_{r<s}$ and $(D_r V_j(s, X_n(s)))_{r<s}$ are adapted and square integrable. By Proposition 1.7.2, $\int_0^t V_j(s, X_n(s))\,dB_s^j \in \mathbb{D}^{1,2}$ and

$$D_r \int_0^t V_j(s, X_n(s))\,dB_s^j = V_j(r, X_n(r)) + \int_r^t D_r V_j(s, X_n(s))\,dB_s^j\,.$$

Now note that for $p > 1$ the right hand side belongs to $L^p(\Omega; L^2(0,\tau))$ where the second integration is with respect to r. Indeed, for the first term on the right hand side this follows from the Lipschitz assumption on the V_j together with the integrability properties of X_n. For the stochastic integral, this follows from the Burkholder-Davies-Gundy inequality, standard estimates and the boundedness of ψ_n.

It follows from Proposition 2.3.5 that $\int_0^t V_j(s, X_n(s))\,dB_s^j \in \mathbb{D}^{1,\infty}$. As for the deterministic integral, it is a consequence of the closedness of the Malliavin derivative on L^2 that integration and differentiation can be interchanged. Now a similar argument as above shows that the deterministic integral belongs to $\mathbb{D}^{1,\infty}$. By the definition of Φ, it follows that $X_{n+1}(t) \in \mathbb{D}^{1,\infty}$.

Moreover, standard estimates show that

$$\mathbb{E}\sup_{s\in[r,t]} |D_r X_{n+1}(s)|^p \le c_p\Big[\gamma_p + T^{p-1} L^p \int_r^t \mathbb{E}|D_r X_n(s)|\,ds\Big]$$

where $\gamma_p = \sup_{n,j} \mathbb{E}\sup_{s\in(0,\tau)} |V_j(s, X_n(s))|^p$. Since X_n converges in $L^p(\Omega; C([0,\tau]))$, it is bounded in that space and the Lipschitz assumption on V_j yields that γ_p is finite.

This shows that $\psi_n(t) \le c_1 + c_2 \int_0^t \psi_n(s)\,ds$. Taking into account that $\psi_0(s) \equiv 0$, it follows inductively that $\psi_n(t) \le c_1 \sum_{j=0}^n \frac{c_2^j t^j}{j!}$. In particular, the ψ_n are uniformly bounded on $(0,\tau)$ whence the sequence of derivatives is bounded in $L^p(\Omega; H)$. It follows from Lemma 1.2.8 that the components of $X(t)$ belong to $\mathbb{D}^{1,\infty}$. The stochastic equation for the Malliavin derivatives follows by differentiating the SDE, taking into account Proposition 1.7.2 and the chain rule. □

In the case where the coefficients B and V_j are C^∞ with all partial derivatives bounded, the above argument can be iterated to show that $X(t)$ has also higher order Malliavin derivatives in every L^p. The actual proof contains no new ideas but makes it necessary to keep track of rather involved notations for the higher order derivatives. We omit it here and refer the reader to [**6**, Lemma 2.2.2 and Theorem 2.2.2] for more information. However, we formulate the result:

Theorem 2.3.7. *Assume Hypothesis 2.3.3 and, additionally, assume that the functions B and V_j are infinitely differentiable functions of x with all partial derivatives bounded. Then the unique solution of the stochastic differential equation*

$$X(t) = x_0 + \int_0^t B(s, X(s))\, ds + \sum_{j=1}^d \int_0^t V_j(s, X(s))\, dB_s^j$$

has components in $\mathbb{D}^\infty$.

2.4. Hörmander's Theorem

By what was done so far, the solutions of stochastic differential equations with smooth coefficients have components in $\mathbb{D}^\infty$. Theorem 2.2.14 shows that to conclude that the law of the solution has a smooth density with respect to m-dimensional Lebesgue measure, we need to show that the Malliavin matrix γ is almost surely invertible with $(\det \gamma)^{-1} \in L^p$ for all $p \geq 1$.

To insure the latter, we need an extra assumption on the coefficients. In fact, whether or not the solution has a density with respect to m-dimensional Lebesgue measure somehow depends on "algebraic" properties of the fields B and V_j. For example, if $B = V_j \equiv 0$, then the solution $X \equiv x_0$ has degenerate distribution.

Moreoever, there is a difference between B and V_j. Indeed, if $m = d = 1$ and $B \equiv 0$ and $V_1 \equiv 1$, then the solution is Brownian motion and has a smooth density. If, on the other hand, $B \equiv 1$ and $V_1 \equiv 0$, then the solution is a deterministic function and has degenerate diestribution.

We can also take another look at Example 2.1.9. In the first example discussed there, we have $B(x, y) = (y, 0)^*$ and $V_1(x, y) = (0, 1)^*$. We have showed that in this case, the solution has a density with respect to 2-dimensional Lebesgue measure.

On the other hand, in the second example discussed, $B(x, y) = (x, 0)^*, V_1(x, y) = (0, 1)^*$ the solution fails to have a density with respect to 2-dimensional Lebesgue measure.

Note that the only difference between these to examples is in exchanging an x for a y.

The "correct" condition to ensure invertibility of the Malliavin matrix goes back to the seminal work of Hörmander [**2**] on hypoelliptic differential operators. We recall that a differential operator A is called hypoelliptic, if for a distribution u the statement $Au \in C^\infty$ implies that u is in C^∞. Hörmanders condition is formulated in the language of differential geometry.

A *(smooth) vector field* on $\mathbb{R}^m$ is a map $U : \mathbb{R}^m \to \mathbb{R}^m$ with C^∞ entries. Given two vector fields U and V, the *Lie Bracket* $[U, V]$ is the vector field defined by $[U, V](x) = DV(x) \cdot U(x) - DU(x) \cdot V(x)$, where DU is the derivative matrix with entries $(DU)_{ij} = \partial_j U_i$ and "$\cdot$" is matrix-vector multiplication.

In this section we will consider coefficients $B, V_1, \ldots, V_d$ as in Hypothesis 2.3.3 but *independent of time* and infinitely differentiable with bounded partial derivatives. Given such coefficients, we put

$$V_0 := B - \frac{1}{2} \sum_{j=1}^d DV_j \cdot V_j\, .$$

This vector fields appear when rewriting our stochastic differential equation in *Stratonovich form*. We now introduce Hörmander's condition.

Definition 2.4.1. Define $S_0 := \{V_1, \ldots, V_d\}$ and, inductively,

$$S_{k+1} := S_k \cup \{[U, V_j] : U \in S_k, j = 0, 1, \ldots, d\}.$$

Moreover, we put $\mathscr{V}_k(x) := \operatorname{span}\{U(x) : U \in S_k\}$. We say that $B, V_1, \ldots, V_d$ *satisfies Hörmander's condition* if $\bigcup_{k\geq 0} \mathscr{V}_k(x) = \mathbb{R}^m$ for all $x \in \mathbb{R}^m$.

EXAMPLE 2.4.2. Let us again consider the situation of Example 2.1.9 where $B(x,y) = (y,0)^*$ and $V_1(x,y) = (0,1)^*$. Note that in this case $DV_1 = 0$, so that $V_0 = B$. Then we have $S_0 = \{(0,1)^*\}$ so that $\mathscr{V}_1 = \mathbb{R}(0,1)^*$. As for the commutator $[V_1, V_0]$, we find

$$[V_1, V_0](x,y) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} y \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

so that $\mathscr{V}_2(x,y) = \operatorname{span}\{(0,1)^*, (1,0)^*\} = \mathbb{R}^2$ so Hörmander's condition is satisfied.

On the other hand, if $B(x,y) = (x,0)^*$ and $V_1(x,y) = (0,1)^*$, we have

$$[V_1, V_0](x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

so that $\mathscr{V}_2(x,y) = \mathscr{V}_1(x,y) = \mathbb{R}(0,1)^*$. Note that also further "bracketing" yields does yield a larger space (we either bracket $[V_1, V_0] = 0$ or $[V_1, V_1] = 0$ trivially). Thus in this situation, Hörmander's condition is not satisfied.

We can now formulate the main result of this section, which can be though of as a probabilistic version of Hörmander's theorem.

THEOREM 2.4.3. *Let $m, d \in \mathbb{N}$, $x_0 \in \mathbb{R}^m$ and $B, V_1, \ldots, V_d : \mathbb{R}^m \to \mathbb{R}^m$ be Lipschitz continuous with C^∞ components such that all partial derivatives are bounded. If Hörmanders condition is satisfied, then the law of the solution $X(t)$ to the stochastic differential equation*

$$X(t) = x_0 + \int_0^t B(X(s))\, ds + \sum_{j=1}^d \int_0^t V_j(X(s))\, dB_s^j \tag{2.2}$$

has a density $p_t(x) \in \mathscr{S}(\mathbb{R}^m)$ with respect to m-dimensional Lebesgue measure.

In view of Theorem 2.3.6 and Theorem 2.2.14, to prove Theorem 2.4.3, it suffices to prove that the Malliavin matrix γ is almost surely invertible with $(\det\gamma)^{-1} \in \bigcap_{p\geq 1} L^p(\Omega)$. To that end, we will use the following

LEMMA 2.4.4. *Let M be a random, symmetric, positive semidefinite matrix with entries in $\bigcap_{p\geq 1} L^p(\Omega)$. Assume that for $p \geq 2$ there exists a constant C_p and an $\varepsilon_p > 0$ such that for $0 < \varepsilon < \varepsilon_p$ we have*

$$\sup_{|x|=1} \mathbb{P}(x^* M x \leq \varepsilon) \leq C_p \varepsilon^p .$$

Then $(\det M)^{-1} \in \bigcap_{p\geq 1} L^p(\Omega)$.

PROOF. Let $\lambda := \inf_{|x|=1} x^* M x$ be the smallest eigenvalue of M. Then $\lambda^m \leq \det M$. Consequently, it suffices to prove that $\mathbb{E}\lambda^{-q} < \infty$ for all $q \geq 2$ as this entails the assertion. Noting that

$$\mathbb{E}\lambda^{-q} = \int_0^\infty q t^{q-1} \mathbb{P}(\lambda^{-1} > t)\, dt) = \int_0^\infty q t^{q-1} \mathbb{P}(\lambda < t^{-1})\, dt$$

we see that it suffices to show that fpr $p \geq 2$ there exists a constant C such that $\mathbb{P}(\lambda < \varepsilon) \leq C\varepsilon^p$ for all ε small enough.

Given $\varepsilon > 0$ we can cover the unit sphere $\{|x| = 1\}$ by finitely many balls of radius ε^2. We thus find $x_1, \ldots, x_n$ such that for every x with $|x| = 1$, we find a $k \in \{1, \ldots, n\}$ with $|x - x_k| \leq \varepsilon^2$. Cleary, n can be bounded by a constant times ε^{-2m}.

Now let x with $|x| = 1$ be given and pick k with $|x - x_k| \leq \varepsilon^2$. Then

$$\langle x, Mx\rangle = \langle x_k, Mx_k\rangle + \langle x - x_k, Mx\rangle + \langle x - x_k, Mx_k\rangle \geq \langle x_k, Mx_k\rangle - 2\|M\|\varepsilon^2 .$$

Consequently, $\{\|M\| < \varepsilon^{-1}\} \cap \bigcap_{k=1}^n \{x_k^* M x_k > 3\varepsilon\} \subset \{x^* M x > \varepsilon\ \forall\, |x| = 1\}$. Thus

$$\mathbb{P}(\lambda \leq \varepsilon) \;=\; \mathbb{P}\big(\{\exists\, |x| = 1 : x^* M x \leq \varepsilon\}\big)$$

$$\begin{aligned} &\leq \mathbb{P}(\|M\| \geq \varepsilon^{-1}) + \sum_{k=1}^{n} \mathbb{P}(x_k^* M x_k \leq 3\varepsilon) \\ &\leq \varepsilon^p \mathbb{E}\|M\|^p + n \sup_{|x|=1} \mathbb{P}(x^* M x \leq 3\varepsilon)\,. \end{aligned}$$

Note that $\mathbb{E}\|M\|^p$ is finite, since M has entries which belong to all L^p's. And our assumption imlies that the last sum is bounded by a constant times $n\varepsilon^{p+2m}$ for $\varepsilon < \varepsilon_{p+2m}$. Noting that $n\varepsilon^{2m}$ is bounded, we are done. □

Intermezzo: Itô's Formula

Before proceeding, let us briefly recall Itô's formula. It will be convienient to formulate Itô's formula basis free, i.e. we consider processes taking values in a finite dimensional vector space E. Moreover, we assume that $B^1, \ldots, B^d$ are independent Brownian motions and we are given an E-valued stochastic X process which has the representation

$$X(t) = x_0 + \int_0^t \varphi(s)\, ds + \sum_{j=1}^{d} \int_0^t \Phi_j(s)\, dB_s^j\,.$$

Here, $\varphi, \Phi_j \in L^2_{\mathbb{F}}((0,\infty) \times \Omega; E)$. Finally, we are given a twice continuously differentiable function $f : E \to F$ to another finite dimensional vector space F. We recall that the derivative f' takes values in $\mathscr{L}(E,F)$, the (bounded) linear maps from E to F. Picking bases, f' can be identified with the Jacobian matrix whose entries are the partial derivatives of the components of f. Applying the linear operator $f'(x)$ to the vector $h \in E$ is expressed by mere concatenation: $f'(x_0)h$. The second derivative f'' takes values in $\mathscr{L}(E, \mathscr{L}(E,F))$ which can canonically be identified with the bilinear maps from $E \times E$ to F. We write $f''(x_0)[h_s, h_2]$ for $(f''(x_0)h_1)h_2$.

With these notation at hand, Itô's formula asserts that the stochastic process $f(X)$ can be represented in the form

$$\begin{aligned} f(X(t)) =& f(x_0) + \int_0^t f'(X(s))\varphi(s)\, ds + \sum_{j=1}^{d} \int_0^t f'(X(s))\Phi_j(s)\, dB_s^j \\ &+ \frac{1}{2} \sum_{j=1}^{d} \int_0^t f''(X(s))[\Phi_j(s), \Phi_j(s)]\, ds\,. \end{aligned}$$

After this brief excursion, let us now look at the Malliavin derivative itself. If we put $\gamma_t = (\gamma_{ij}(t))$ then, by definition,

$$\gamma_{ij}(t) = \langle DX_i(t), DX_j(t) \rangle_{L^2((0,\tau);\mathbb{R}^d)} = \sum_{k=1}^{d} \int_0^t D_r^k X_i(t) D_r^k X_j(t)\, dr\,.$$

As a consequence of Theorem 2.3.6, we have stochastic equations for the Malliavin derivatives $D_r^k X_i(t)$. However, inserting these into the equation above yields integrands which are *not adapted.* Our next task will be to establish a different representation of the Malliavin derivative, involving an integral over an adapted process, thus enabling us to use some results about Itô stochastic integrals.

To that end, let $Y(t)$ be the matrix valued process which solves the stochastic equation

$$Y(t) = I + \int_0^t B'(X(s))Y(s)\, ds + \sum_{j=1}^{d} \int_0^t V_j'(X(s))Y(s)\, dB_s^j\,.$$

Formally, Y can be thought of as the derivative of the solution $X(t) = X(t, x_0)$ with respect to the initial datum x_0 (differentiate the equation solved by X with respect to x_0 not worrying about interchanging differentiation and integration!) This actually can be made precise, however we can also take the above equation as the definition of Y.

We claim that $Y(t)$ is almost surely invertible. If that was the case, then we can (formally) obtain a stochastic differential equation for $Y(t)^{-1}$ from Itô's formula. Indeed if we define f on the invertible matrices by $f(A) = A^{-1}$, then f is twice continuously differentiable with $f'(A)H = -A^{-1}HA^{-1}$ and $f''(A)[H_1, H_2] = 2A^{-1}H_1A^{-1}H_sA^{-1}$. Thus Itô's formula yields, formally,

$$\begin{aligned} Y(t)^{-1} &= I^{-1} - \int_0^t Y(s)^{-1}B'(X(s))Y(s)^{-1}\,ds - \sum_{j=1}^d \int_0^t Y(s)^{-1}V_j'(X(s))Y(s)Y(s)^{-1}\,dB_s^j \\ &\quad + \sum_{j=1}^d \int_0^t Y(s)^{-1}V_j'(X(s))Y(s)Y(s)^{-1}V_j(X(s))Y(s)Y(s)1^{-1}\,ds\,. \end{aligned}$$

Performing some calculations, we would thus expect that $Y(t)^{-1}$ is the solution Z of the following equation:

$$Z(t) = I - \int_0^t Z(s)B'(X(s)) - \sum_{j=1}^d Z(s)V_j'(X(s))^2\,ds - \sum_{j=1}^d Z(s)V_j'(X(s))\,dB_s^j\,. \tag{2.3}$$

As this equation has coefficients of linear growth, it can easily be proved that this equation does, in fact, have a unique solution $Z(t)$. By Itô's formula (applied to the map $f(A, B) = AB$) we find

$$\begin{aligned} Z(t)Y(t) &= I + \int_0^t Z(s)B'(X(s))Y(s)\,ds + \sum_{j=1}^d \int_0^t Z(s)V_j'(X(s))Y(s)\,dB_s^j \\ &\quad - \int_0^t Z(s)B'(X(s))Y(s)\,ds + \sum_{j=1}^d \int_0^t Z(s)V_j'(X(s))^2Y(s)\,ds \\ &\quad - \sum_{j=1}^d \int_0^t Z(s)V_j'(X(s))Y(s)\,dB_s^j - \sum_{j=1}^d \int_0^t Z(s)V_j'(X(s))V_j'(X(s))Y(s)\,ds \\ &= I\,. \end{aligned}$$

By linear algebra, $Y(t)$ is almost surely invertible and $Y(t)^{-1} = Z(t)$.

We claim that $D_r^jX(t) = Y(t)Y(r)^{-1}V_j(X(s))$. Indeed, we have

$$\begin{aligned} & V_j(X(r)) + \int_r^t B'(X(s))Y(s)Y(r)^{-1}V_j(X(r))\,ds \\ & \quad + \sum_{j=1}^d \int_0^t V_j'(X(s))Y(s)Y(r)^{-1}V_j(X(r))\,dB_s^j \\ &= V_j(X(r)) + \Big[\int_r^t B'(X(s))Y(s)\,ds + \sum_{j=1}^d V_j'(X(s))Y(s)\,dB_s^j\Big]Y(r)^{-1}V_j(X(r)) \\ &= V_j(X(r)) + \big[Y(t) - Y(r)\big]Y(r)^{-1}V_j(X(r)) = Y(t)Y(r)^{-1}V_j(X(r))\,. \end{aligned}$$

Thus $Y(t)Y(r)^{-1}V_j(X(s))$ solves the same stochastic equation as, according to Theorem 2.3.6, does $D_r^j(X(t))$. Thus the two are equal as claimed.

It follows that

$$\gamma_t = \sum_{j=1}^d \int_0^t D_r^jX(t))(D_r^jX(t))^*\,dr$$

$$
\begin{aligned}
&= \sum_{j=1}^{d} \int_0^t Y(t)Y(r)^{-1}V_j(X(r))V_j(X(r))^*(Y(r)^{-1})^*Y(t)^*\,dr \\
&= Y(t)\sum_{j=1}^{d} \int_0^t Y(r)^{-1}V_j(X(r))V_j(X(r))^*(Y(r)^{-1})^*\,drY(t)^* \\
&=: Y(t)C(t)Y(t)^*\,.
\end{aligned}
$$

We have now achieved our goal. The matrix $C(t)$ is a sum of integrals over adapted processes. Moreover, the matrix $Y(t)$ is almost surely invertible. Finally, we have a stochastic differential equation for its inverse which shows that (cf. Theorem 2.3.4) that the entries of the inverse have finite moments of all orders. In particular, it follows that $(\det \gamma_t)^{-1}$ belongs to all L^p's if and only if $(\det C(t))^{-1}$ belongs to all L^p's.

By the observations made so far, Theorem 2.4.3 follows at once from the following Theorem.

THEOREM 2.4.5. *Under the assumptions of Theorem 2.4.3, for every $p \geq 2$ there exists a constant C_p and an $\varepsilon_p > 0$ such that for $0 < \varepsilon < \varepsilon_p$ we have*

$$
\sup_{|x|=1} \mathbb{P}(x^*C(t)x \leq \varepsilon) \leq C_p \varepsilon^p\,.
$$

We next explain how the brackets $[V_k, V_l]$ come into play. Fixing x with $|x| = 1$ and given a smooth vector field U, let us write $Z_U(t) := \big\langle x, Y(t)^{-1}U(X(s))\big\rangle$, where X solves (2.2) and Y^{-1} solves (2.3). Then we have

$$
\begin{aligned}
\langle x, C(t)x\rangle &= \sum_{j=1}^{d} \int_0^t \big\langle x, Y(r)^{-1}V_j(X(r))V_j(X(r))^*(Y(r)^{-1})^*x\big\rangle\,dr \\
&= \sum_{j=1}^{d} \int_0^t |Z_{V_j}(r)|^2\,dr\,.
\end{aligned}
\tag{2.4}
$$

Moreover, using equations (2.2), (2.3) and Itô's formula, we can compute Z_U explicitly:

$$
\begin{aligned}
Y(t)^{-1}U(X(t)) \;=\; & U(x_0) + \sum_{j=1}^{d} \int_0^t Y(s)^{-1}\big[U'V_j - V_j'U\big]dB_s^j \\
&+ \int_0^t Y(s)^{-1}\big[U'B - B'U\big](X(s))\,ds + \sum_{j=1}^{d} \int_0^t Y(s)^{-1}\big[V_j'V_j'U\big](X(s))\,ds \\
&+ \frac{1}{2}\int_0^t Y(s)^{-1}\sum_{j=1}^{d} U''(X(s))[V_j(X(s)), V_j(X(s))]\,ds \\
&- \sum_{j=1}^{d} \int_0^t Y(s)^{-1}\big[V_j'UV_j\big](X(s))\,ds\,.
\end{aligned}
$$

Observe that $U'V_j - V_jU = [V_j, U]$ and $U'B - B'U = [B, U]$. Moreover, a somewhat tedious but straightforward computation shows

$$
[V_0, U] + \frac{1}{2}\sum_{j=1}^{d} \big[V_j, [V_j, U]\big] - [B, U] = \sum_{j=1}^{d}(V_jV_jU + \frac{1}{2}U''[V_j, V_j] - V_j'UV_j)
$$

so that alltogether, we have

$$Y(t)^{-1}U(X(t)) = U(x_0) + \int_0^t Y(s)^{-1}\big([V_0,U] + \frac{1}{2}\sum_{j=1}^d [V_j,[V_j,U]]\big)(X(s))\,ds + \sum_{j=1}^d Y(s)^{-1}[V_j,U](X(s))\,dB_s^j\,.$$

Taking inner products with x, it follows that Z_U satisfies the following stochastic equation

$$Z_U(t) = \langle x, U(x_0)\rangle + \int_0^t Z_{[V_0,U]+\frac{1}{2}\sum_{j=1}^d [V_j,[V_j,U]]}(s)\,ds + \sum_{j=1}^d \int_0^t Z_{[V_j,U]}(s)\,dB_s^j\,. \tag{2.5}$$

Motivated by the above equation for Z_U, we introduce the sets S'_k and the vector spaces $\mathscr{V}'_K(x)$ as follows:

We put $S'_0 := S_0 = \{V_1,\dots,V_d\}$ and then, inductively,

$$S'_{k+1} := \Big\{[V_j,U], j=1,\dots,d, U\in S'_k; [V_0,U] + \frac{1}{2}\sum_{j=1}^d [V_j,]V_j,U]], U\in S'_k\Big\}\,.$$

Moreover, $\mathscr{V}'_k(x) := \operatorname{span}\{U(x), U\in S_k\}$. It is better to use the vector spaces $\mathscr{V}'_k(x)$ rather than $\mathscr{V}_k(x)$ due to the above equation for S_U. This equation, in turn, is a consequence of our using Itô integrals rather than Stratonovich integrals.

We can now sketch the strategy for the proof of Theorem 2.4.5. We need to proof that $x^*C(t)x$ gets small only with small probability. At to get the idea of the proof, let us assume that we would like to prove the (deterministic) statement that $x^*C(t)x \neq 0$. Aiming for a contradiction, assume that $x^*C(t)x = 0$ for all x with $|x| = 1$. Then equation (2.4) implies that $Z_{V_j} \equiv 0$ for $j = 1,\dots,d$. Equation (2.5) gives the semimartingale decomposition for Z_{V_j}. As this decomposition is unique, it follows that both the martingale part and the bounded variation part are zero. Consequently, $Z_{[V_0,V_k]+\frac{1}{2}\sum_{j=1}^d [V_j,[V_j,V_k]]} \equiv 0$ for $k = 1,\dots,d$ and $Z_{[V_j,V_k]} \equiv 0$ for $j,k = 1,\dots,d$. Stated differently, if $Z_U \equiv 0$ for $U \in S'_0$, then (2.5) implies that $Z_U \equiv 0$ for all $U \in S'_1$. Inductively, $S_U \equiv 0$ for $U \in S'_k$ for all $k \geq 0$. It would thus follow that $0 = S_U(0) = \langle x, U(x_0)\rangle$ for all $U \in \bigcup_{k\geq 0}\mathscr{V}'_k(x_0)$. As the latter is all of $\mathbb{R}^m$ by Hörmander's condition, it follows that $x = 0$ – a contradiction.

For the actual proof of the theorem, we need a "qualitative version" of the above argument and thus a "qualitative version" of the semimartingale decomposition. This result is due to Norris [**5**] and somtimes referred to as *Norris Lemma*. In order to highlight the above strategy, we follow Hairer [**1**] for the proof. In particular, we make use of the notion of "almost truth" and "almost implication" introduced there to streamline the proof.

Given a family $A = (A_\varepsilon)_{\varepsilon\in(0,1]}$ of events we say that A is *almost false* if, for every $p \geq 1$ there exists a constant C_p such that $\mathbb{P}(A_\varepsilon) \leq C_p\varepsilon^p$ for all $\varepsilon \in (0,1]$. We say that it is *almost true* if $A^c = (A^c_\varepsilon)_{\varepsilon\in(0,1]}$ is almost false. Given two such families A and B, we say that A *almost implies* B and write $A \Rightarrow_\varepsilon B$ if $A\setminus B = (A_\varepsilon \setminus B_\varepsilon)$ is almost false.

EXAMPLE 2.4.6. If $X \in \bigcap_{p\geq 1} L^p$, then $(\{|X| \leq \varepsilon^{-r}\})$ is almost true for every $r \in (0,1)$. Indeed, by Chebyshev's inequality $\mathbb{P}(|X| > \varepsilon^{-r}) \leq \varepsilon^{rq}\mathbb{E}|X|^q$ so that for $p \geq 1 > r$ we can choose $C_p = \mathbb{E}|X|^{\frac{p}{r}}$.

We will use this notions to formulate our version of Norris Lemma 2.4.8 below. In the proof, we use the following estimate, in which $\|f\|_\alpha$ refers to the best possible α-Hölder constant of f, i.e. $\|f\|_\alpha = \sup_{t\neq s}|t-s|^{-\alpha}|f(t)-f(s)|$.

LEMMA 2.4.7. *Let $f : [0,1] \to \mathbb{R}$ be continuously differentiable, $\alpha \in (0,1]$ and assume that f' is α-Hölder continuous. Then*

$$\|f'\|_\infty = \|f\|_1 \le 4\|f\|_\infty^{\frac{\alpha}{1+\alpha}} \|f'\|_\alpha^{\frac{1}{1+\alpha}} .$$

PROOF. Let us first note that if $|f'| \ge r$ on some closed interval $[a,b]$, then at some point $t_1 \in [a,b]$, we have $|f(t_1)| \ge r(b-a)/2$. To see this, first note that by continuity, f' does not change its sign on $[a,b]$ so that we can assume without loss of generality that $f' > 0$ on $[a,b]$. If $|f(a)| \ge r(b-a)/2$, then we are done. Otherwise, $f(a) > -r(b-a)/2$. In this case

$$f(b) = f(a) + \int_a^b f'(s)\,ds > -r(b-a)/2 + \int_a^b r\,ds = r(b-a)/2 .$$

Now let t_0 be such that $|f'(t_0)| = \|f'\|_\infty$. By the definition of Hölder continuity, if $|t| \le (2^{-1}\|f'\|_\alpha^{-1}\|f'\|_\infty)^{\frac{1}{\alpha}}$, then $|f'(t_0+t) - f'(t_0)| \le \|f'\|_\infty/2$. Thus $|f'(t_0+t)| \ge \|f'\|_\infty/2$ for such t. Applying the above estimate with $r = \|f'\|_\infty/2$ and $[a,b] = \{t_0 + t : |t| \le (2^{-1}\|f'\|_\alpha^{-1}\|f'\|_\infty)^{\frac{1}{\alpha}}\}$ it follows that

$$\frac{1}{2}\|f'\|_\infty \Big(\frac{\|f'\|_\infty}{2\|f'\|_\alpha}\Big)^{\frac{1}{\alpha}} \le |f(t_1)| \le \|f\|_\infty .$$

Now the thesis follows with elementary computations. ☐

LEMMA 2.4.8. *Let B be a d-dimensional Wiener process, a and b be $\mathbb{R}$ respectively $\mathbb{R}^d$-valued adapted processes such that, for $\alpha = 1/3$, we have $\|a\|_\alpha, \|b\|_\alpha \in \bigcap_p L^p$. Moreover, let Z be defined by*

$$Z(t) = Z_0 + \int_0^t a(s)\,ds + \sum_{j=1}^d \int_0^t b_j(s)\,dB_s^j .$$

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

$$\{\|Z\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \{\|a\|_\infty < \varepsilon^r\} \& \{\|b\|_\infty < \varepsilon^r\} .$$

PROOF. In the proof, we use the following "exponential martingale inequality" see [**7**, Exercise 3.16]. If M is a continuous martingale with quadratic variation $\langle M\rangle$, then

$$\mathbb{P}\big(\sup_{t\le T}|M(t)| \ge x, \langle M\rangle_T \le y\big) \le 2e^{-\frac{x^2}{2y}} .$$

Applying this to the martingale $M(t) := \sum_{j=1}^d \int_0^t b_j(s)dB_s^j =: \int_0^t b(s)dB_s$, whose quadratic variation is $\langle M\rangle_t = \int_0^t \|b(s)\|^2\,ds$ and noting that $\|b\|_\infty < \varepsilon$ implies that $\langle M\rangle_T \le T\varepsilon^2$, we see that

$$\mathbb{P}\Big(\sup_{t\le T}\Big\|\int_0^t b(s)\,dB_s\Big\| \ge \varepsilon^q, \|b\|_\infty \le \varepsilon\Big) \le 2e^{-\frac{\varepsilon^{2q}}{2\varepsilon^2 T}} .$$

For $q \in (0,1)$, the latter converges exponentially to zero as $\varepsilon \downarrow 0$. Consequently, we have proves that

$$\{\|b\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \Big\{\Big\|\int_0^\cdot b(s)\,dB_s\Big\|_\infty < \varepsilon^q\Big\} \tag{2.6}$$

for $q \in (0,1)$.

We now apply Itô's formula to Z^2, obtaining

$$Z(t)^2 = Z_0^2 + 2\int_0^t Z(s)a(s)\,ds + 2\int_0^t Z(s)b(s)\,dB_s + \int_0^t |b(s)|^2\,ds .$$

As $\|a\|_\infty \le \varepsilon^{-1/4}$ is almost true by the example above, we see that

$$\{\|Z\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \Big\{\Big\|\int_0^\cdot Z(s)a(s)\,ds\Big\|_\infty \le \varepsilon^{3/4}\Big\}$$

Taking equation (2.6) into account, it follows similarly that

$$\{\|Z\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \Big\{\Big\| \int_0^{\cdot} Z(s)b(s)\,ds\Big\|_\infty \le \varepsilon^{2/3}\Big\}$$

Now note that for a random variable X we have $\{|X| < \varepsilon^r\} \Rightarrow_\varepsilon \{|X| < \varepsilon^s\}$ for $s < r$, since $\{|X| < \varepsilon^r, |X| \ge \varepsilon^s\} \subset \{\varepsilon^s < \varepsilon^r\} = \emptyset$ for $\varepsilon \in (0,1)$. We can thus insert the above estimates for the deterministic and the stochastic integral back into the equation for Z^2 and obtain

$$\{\|Z\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \Big\{\int_0^T |b(s)|^2\,ds \le \varepsilon^{\frac{1}{2}}\Big\} \Rightarrow \Big\{\int_0^T |b(s)|\,ds \le \varepsilon^{\frac{1}{4}}\Big\}$$

as $\int_0^T |b(s)|\,ds \le \sqrt{T}\int_0^T |b(s)|^2\,ds$. We now use Lemma 2.4.7. We find

$$\|b\|_\infty \le 4\Big(\int_0^T |b(s)|\,ds\Big)^{\frac{1}{4}} \|b\|_\alpha^{\frac{3}{4}}\,.$$

By the above, $\|Z\|_\infty < \varepsilon$ almost implies $\int_0^T |b(s)|\,ds \le \varepsilon^{\frac{1}{4}}$. Moreover it follows from our assumption that $\|b\|_\alpha \le \varepsilon^{-q}$ is almost true, for every $q \in (0,1)$. Together, this implies that

$$\{\|Z\|_\infty \le \varepsilon\} \Rightarrow_\varepsilon \{\|b\|_\infty \le \varepsilon^s\}$$

for any $s < \frac{1}{16}$, say $s = \frac{1}{17}$. By equation (2.6) this almost implies that $\|\int b(s)dB_s\|_\infty \le \varepsilon^{q\frac{1}{17}}$ for $q < 1$, thus it almost implies that $\|\int b(s)dB_s\|_\infty \le \varepsilon^{\frac{1}{18}}$. This, in turn, together with the equation for Z implies that $\{\|Z\|_\infty < \varepsilon\}$ almost implies $\{\|\int a(s)ds\|_\infty \le \varepsilon^{1/18}\}$. Employing Lemma 2.4.7 as above, we see that

$$\{\|Z\|_\infty \le \varepsilon\} \Rightarrow_\varepsilon \{\|a\|_\infty \le \varepsilon^{\frac{1}{80}}\}\,.$$

Altogether, the claim is proved with $r = 1/80$. □

We have now all tools at hand to prove Hörmander's theorem.

Proof of Theorem 2.4.5. Fix x with $|x| = 1$. It follows from (2.4), using Lemma 2.4.7 as above that

$$\{x^*C(T)x \le \varepsilon\} \Rightarrow_\varepsilon \{\|Z_{V_k}\|_\infty \le \varepsilon^{\frac{1}{5}}\}\,.$$

Using Equation (2.5) and Lemma 2.4.8, it follows inductively that

$$\{x^*C(T)x \le \varepsilon\} \Rightarrow_\varepsilon \bigcap_{V\in\mathscr{V}_k'} \{\|Z_V\|_\infty \le \varepsilon^{r_k}\}$$

for suitable $r_k > 0$. Now observe that $Z_V(0) = \langle x, U(x_0)\rangle$. By Hörmander's condition, $\mathscr{V}_k'(x_0) = \mathbb{R}^m$ for k large enough. However, if $\mathscr{V}_k(x_0') = \mathbb{R}^m$, we can pick $V \in \mathscr{V}_k'(x_0)$ such that $Z_V(0) = 1$, so that the right-hand-side in the above equation is the empty set. We have thus proved $\{x^tC(T)x \le \varepsilon\} \Rightarrow_\varepsilon \emptyset$, which is exactly the thesis. □

Bibliography

[1] M. Hairer, *On Malliavin's proof of Hörmander's theorem.* arXiv:1103.1998, 2011.

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

[3] O. Kallenberg, *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York, second ed., 2002.

[4] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, vol. 113 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1991.

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

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

[7] 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.