Malliavin Calculus and Applications

4th Year MMath Project

Matthew Dunlop
University of Warwick

Supervisor: Dr Xue-Mei Li

Contents

Introduction

The Malliavin calculus is a variational calculus for L^p functions on a space of paths which are not differentiable in the classical sense, for example functions on $C_0([0,T];\mathbb{R}^m)$, the space of continuous paths in $\mathbb{R}^m$ starting from the origin. The aim of this essay is to present the main elements of Malliavin calculus, and then use them to analyse the law of the Ornstein-Uhlenbeck (OU) process.

In the first chapter we introduce the Malliavin derivative operator D in terms of isonormal Gaussian processes and look at some of its properties, in particular its action on diffusion processes. We then proceed to define the divergence, which leads to an extension of the Itô integral to non-adapted integrands. This in turn leads to the Clark-Ocone theorem, which gives a more explicit form for the martingale representation theorem for differentiable L^2 functions on $C_0([0,T])$.

One of the important applications of Malliavin calculus, and the original motivation for developing the theory, is to provide a probabilistic proof of Hörmander's theorem regarding the existence and smoothness of a density with respect to the Lebesgue measure of the law of solutions to a class of hypoelliptic SDEs. In the second chapter we supply a proof of a version of this theorem following Hairer's recent paper [5], after discussing the problem of existence and smoothness of densities for the law of more general random variables.

Finally in the third chapter we analyse the laws of the one-dimensional OU process Z defined by the SDE

$$dZ_t = dB_t - Z_t dt$$

and the associated two-dimensional process (X, Y) defined by the SDE

$$\begin{cases} dX_t = V_t dt \\ dV_t = dB_t - V_t dt \end{cases}$$

More specifically, we look at the Cameron-Martin space associated with the law of Z using a classical approach and then using Malliavin calculus, before looking for its associated divergence operator in terms of the Itô integral. We then look at the problem of existence and smoothness of the density of the law of (X, V) using the tehniques covered in the second chapter.

Sufficient prerequisites would be undergraduate courses in functional analysis and measure theory (Riesz representation theorems, unbounded/closable operators, adjoints) and a first course in stochastic analysis (stochastic integrals, Itô's formula, existance and uniqueness of solutions to SDEs, Girsanov's theorem). Some knowledge of differential geometry and Sobolev spaces would be beneficial in parts. Knowledge of chaos expansions and infinite dimensional Gaussian measures is also assumed, but the relevant details can be found in the appendices.

1 Malliavin calculus

1.1 The derivative operator

In this section we define the Malliavin derivative operator and give some of its key properties.

1.1.1 Definitions

Central to our definition of the Malliavin derivative operator will be the notion of an isonormal process, which is a family of L^2 Gaussian random variables indexed by a Hilbert space. For examples and more details, see Appendix A.1.

Definition 1.1. *Let H be a real separable Hilbert space, and let $W : H \to L^2(\Omega, \mathcal{F}, \mathbb{P})$. We say that W is an* isonormal Gaussian process *if*

(i) W is a linear isometry

(ii) $W(h)$ is normally distributed with mean zero and variance $\|h\|^2$

Let $C_p^\infty(\mathbb{R}^n)$ be the space of all infinitely differentiable functions on $\mathbb{R}^n$ whose derivatives have at most polynomial growth. Let W be an isonormal Gaussian process on a real separable Hilbert space H, and define the space $\mathcal{S}$ of smooth cylindrical random variables by

$$\mathcal{S} := \{F = \psi(W(h_1), \dots, W(h_n)) \,|\, \psi \in \ C_p^\infty(\mathbb{R}^n), h_i \in H, n \in \mathbb{N}\}$$

The polynomial growth ensures existence of all moments of elements of $\mathcal{S}$, hence $\mathcal{S} \subseteq L^p(\Omega)$ for all p. We define the derivative of elements of $\mathcal{S}$ as follows.

Definition 1.2. *The (Malliavin) derivative DF of a function $F \in \mathcal{S}$ is an H-valued random element given by*

$$DF := \sum_{i=1}^n \frac{\partial \psi}{\partial x_i}(W(h_1), \dots, W(h_n))h_i$$

when F is given by $\psi(W(h_1), \dots, W(h_n))$.

It can be shown that this definition is independent of the choice of representation of F by basic linear algebra, see for example [10, p. 29].

Example 1.3. $D(W(h)) = h$. In particular, in the case where $H = L^2([0, T])$ and $W(h) = \int_0^T h_s dB_s$ is the Itô integral, we have

$$D\left(\int_0^t h_s dB_s\right) = D(W(h\mathbf{1}_{[0,t]})) = h\mathbf{1}_{[0,t]}$$

Similarly, we have that $DB_s = \mathbf{1}_{[0,s]}$, and $Df(B_s) = f'(B_s)\mathbf{1}_{[0,s]}$ for suitable smooth f.

One can easily check that D is a linear operator on $\mathcal{S}$ and that it satisfies the product rule

$$D(FG) = FDG + GDF$$

Since the derivative of a function F is valued in H, $\langle DF, h\rangle_H$ makes sense for $h \in H$, and if DF is given as in Definition 1.2, then

$$\langle DF, h\rangle_H := \sum_{i=1}^{n} \frac{\partial\psi}{\partial x_i}(W(h_1), \ldots, W(h_n))\langle h_i, h\rangle_H$$

This is the derivative of F in the direction[1] h, so the derivative operator D is comparable to the gradiant operator ∇ in the finite dimensional case. The derivative of F in direction h can be thought of in the same variational way as in the finite dimensional case. For example, suppose that $F \in \mathcal{S}$ has representation $\psi(W(k))$, then

$$\begin{aligned}\lim_{\varepsilon\to 0} \frac{\psi(W(k) + \varepsilon\langle k, h\rangle_H) - \psi(W(k)))}{\varepsilon} &= \lim_{\varepsilon'\to 0} \frac{\psi(W(k) + \varepsilon') - \psi(W(k)))}{\varepsilon'} \cdot \langle k, h\rangle_H \\ &= \psi'(W(k))\langle k, h\rangle_H \\ &= \langle DF, k\rangle_H\end{aligned}$$

A similar result holds for when F is of the more general form $\psi(W(h_1), \ldots, W(h_n))$. Note that when we are in the situation of Example 1.3, the perturbation $W(k) + \varepsilon\langle k, h\rangle_H$ corresponds to the addition of a determininstic drift,

$$W(k) + \varepsilon\langle k, h\rangle_H = \int_0^1 k_t dB_t + \varepsilon \int_0^1 k_t h_t dt$$

and so we have that

$$\begin{aligned}\left\langle D\left(\int_0^1 k_t dB_t\right), h\right\rangle_H &= \int_0^1 k_t h_t dt \\ &= \langle k, h\rangle_H\end{aligned}$$

Hence by the Riesz representation theorem we again obtain that

$$D\left(\int_0^1 k_t dB_t\right) = k$$

An important property of the derivative operator is the integration by parts formula, which is used to prove its closability:

Proposition 1.4. *Let $F \in \mathcal{S}$ and $h \in H$, then $\mathbb{E}\langle DF, h\rangle_H = \mathbb{E}(FW(h))$*

[1] The 'directions' should technically be elements of tangent spaces of H, however due to the linear structure of H we can identify its tangent spaces with H itself.

Proof. (Following [6, pp. 25-26]) By linearity and Gram-Schmidt, we lose no generality in assuming that there exist orthonormal elements $e_1, \dots e_n$ of H such that $e_1 = h$ and

$$F = f(W(e_1), \dots, W(e_n))$$

where $f \in C_p^\infty(\mathbb{R}^n)$. From the definition of W and orthogonity of the (e_i) we have that the random vector $(W(e_1), \cdots, W(e_n))$ has standard normal distribution in $\mathbb{R}^n$, and so

$$\begin{aligned}
\mathbb{E}\langle DF, h\rangle_H &= \mathbb{E}\left(\frac{\partial \psi}{\partial x_1}(W(e_1), \cdots, W(e_n))\right) \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}}^n \frac{\partial \psi}{\partial x_1}(x_1, \dots, x_n) e^{-\|x\|^2/2} dx \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}}^n x_1 \psi(x_1, \dots, x_n) e^{-\|x\|^2/2} dx \\
&= \mathbb{E}(FW(h))
\end{aligned}$$

□

Remark. We will often write $D_h F := \langle DF, h\rangle_H$. DF is sometimes viewed an an H^*-valued random variable by the identification $DF(h) := D_h F$. Viewing DF as valued in H^*, the relation between the above integration by parts formula and the formula (A.1) given in Appendix A.1 is clear when W is given by the Paley-Wiener integral. Indeed by approximating $F \in \mathcal{S}$ by BC^1 functions, a variational proof of the above proposition is obtained.

An immediate consequence of the above proposition, using the product rule, is the following:

Corollary 1.5. *Suppose $F, G \in \mathcal{S}$ and $h \in H$, then*

$$\mathbb{E}(G\langle DF, h\rangle_H) = \mathbb{E}(-F\langle DG, h\rangle_H + FGW(h))$$

We are now able to show that the derivative operator is closable. The closability of D is a useful property: it tells us that if $F_n \to F$ and $DF_n \to \eta$, then $\eta = DF$, once we have settled on notions of convergence. Since we have that $\mathcal{S} \subseteq L^p(\Omega)$ for all p, L^p convergence would seem appropriate for the F_n. For the DF_n, we use convergence in the Bochner spaces $L^p(\Omega; H)$.

Proposition 1.6. *For $p \in [1, \infty)$, the operator D is closable from $L^p(\Omega)$ to $L^p(\Omega; H)$.*

Proof. (Following [11]) By linearity, it is enough to show this for the case $F = 0$. Let $(F_n)_{n\in\mathbb{N}} \subseteq \mathcal{S}$ be a sequence of smooth random variables such that $F_n \to 0$ in $L^p(\Omega)$ and $DF_n \to \eta$ in $L^p(\Omega; H)$. We need to show that $\eta = 0$.

First note that for any $h \in H$ we have that $\langle DF_n, h\rangle_H \to \langle \eta, h\rangle_H$ in $L^p(\Omega)$, because

$$\begin{aligned}
\mathbb{E}\left[|\langle DF_n, h\rangle_H - \langle \eta, h\rangle_H|^p\right] &= \mathbb{E}\left[|\langle DF_n - \eta, h\rangle_H|^p\right] \\
&\le \mathbb{E}\left[(\|DF_n - \eta\|_H \|h\|_H)^p\right] \\
&= \|h\|_H^p \|DF_n - \eta\|_{L^p(\Omega;H)}^p \to 0
\end{aligned}$$

So given any bounded $G \in \mathcal{S}$ with $GW(h)$ also bounded, Corollary 1.5 gives

$$\begin{aligned}\mathbb{E}(\langle \eta, h\rangle_H G) &= \lim_{n\to\infty} \mathbb{E}(\langle DF_n, h\rangle_H G)\\ &= \lim_{n\to\infty} \mathbb{E}(-F_n\langle DG, h\rangle_H + F_n GW(h))\\ &= 0\end{aligned}$$

Thus for all $h \in H$, $\langle \eta, h\rangle_H = 0$ a.s. and so $\eta = 0$ $\square$

By the above proposition the operator D can be extended to a maximal closed operator with domain $\mathbb{D}^{1,p} \subseteq L^p(\Omega)$. It follows that the space $\mathbb{D}^{1,p}$ is a Banach space under the graph norm

$$\begin{aligned}\|F\|_{1,p} &:= \left[\|F\|^p_{L^p(\Omega)} + \|DF\|^p_{L^p(\Omega;H)}\right]^{1/p}\\ &= \left[\mathbb{E}|F|^p + \mathbb{E}\|DF\|^p_H\right]^{1/p}\end{aligned}$$

Let $F \in \mathcal{S}$. The kth derivative operator D^k is defined by iteration: $D^1F := DF$ and $D^kF := D(D^{k-1}F)$ for $k \geq 2$. D^kF is hence a random variable with values in $H^{\otimes k}$. Using this we can define some more spaces that D can act on. Define a norm[2] on $\mathcal{S}$ by

$$\|F\|_{k,p} := \left[\mathbb{E}|F|^p + \sum_{i=1}^k \mathbb{E}\|D^iF\|^p_{H^{\otimes i}}\right]^{1/p}$$

This makes sense as the polynomial growth of all derivatives of ψ in the definition of $\mathcal{S}$ ensures that all moments of all derivatives of elements of $\mathcal{S}$ exist. The space $\mathbb{D}^{k,p}$ is defined as the closure of $\mathcal{S}$ with respect to the norm $\|\cdot\|_{k,p}$. Note that we clearly have $\mathbb{D}^{k,p} \subseteq \mathbb{D}^{l,p}$ for $k > l$, and by Jensen's inequality we have that $\mathbb{D}^{k,p} \subseteq \mathbb{D}^{k,q}$ for $p > q$. We can hence define three final spaces:

$$\mathbb{D}^{\infty,p} := \bigcap_k \mathbb{D}^{k,p}, \quad \mathbb{D}^{k,\infty} := \bigcap_p \mathbb{D}^{k,p}, \quad \mathbb{D}^{\infty} := \bigcap_{k,p} \mathbb{D}^{k,p}$$

The spaces $\mathbb{D}^{k,p}$ can be thought of as infinite dimensional analogues to Gaussian Sobolev spaces, see for example [11]. It is worth noting that the spaces $\mathbb{D}^{k,2}$ are Hilbert spaces, with inner products

$$\langle F, G\rangle_{k,2} := \mathbb{E}(FG) + \sum_{i=1}^k \mathbb{E}\langle D^iF, D^iG\rangle_{H^{\otimes i}}$$

The space $\mathbb{D}^{1,2}$ is of particular importance, as will be seen later.

We haven't shown that D^k is well-defined on $\mathbb{D}^{k,p}$ for $k \geq 2$: the closability of D^k from $L^p(\Omega)$ to $L^p(\Omega; H^{\otimes k})$ follows from the same arguments as Proposition 1.6.

It will perhaps be helpful to see some examples of functions which lie in the above spaces, and examples of those that don't.

[2] This is technically a seminorm, but if we identify elements of $\mathcal{S}$ that agree $\mathbb{P}$-almost everywhere it is a true norm

Example 1.7. (i) Clearly $\mathcal{S} \subseteq \mathbb{D}^{k,p}$ for all $k, p \in [1, \infty]$.

(ii) An indicator function $\mathbf{1}_A$ of a measurable set A is in $\mathbb{D}^{1,p}$ if and only if $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, see Corollary 1.10.

(iii) Let $(B_t)_{t\in[0,T]}$ be a one-dimensional Brownian motion on the interval $[0, T]$ and suppose $X = (X_t)_{t\in[0,T]}$ is the process defined by the SDE

$$dX_t = a(X_t)dB_t + b(X_t)dt$$

where $a, b : \mathbb{R} \to \mathbb{R}$ are continuously differentiable and globally Lipschitz. Then $X_t \in \mathbb{D}^{1,\infty}$ for all $t \in [0, T]$, see Proposition 1.21.

The above can be extended to the case where the random variables take values in a separable Hilbert space V. We define the set $\mathcal{S}_V$ of V-valued smooth cylindrical random variables by

$$\mathcal{S}_V := \left\{ F = \sum_{j=1}^m F_j v_j \,\middle|\, F_j \in \mathcal{S}, v_j \in V \right\} \subseteq L^p(\Omega; V) \text{ for all } p$$

Now the derivative of an element $F = \sum_{j=1}^m F_j v_j$ of $\mathcal{S}_V$ is given by

$$DF := \sum_{j=1}^m v_j \otimes DF_j$$

and it can be checked similarly to the real valued case that that D defined in this way is a closable operator. Then we define the space $\mathbb{D}^{k,p}(V) \subseteq L^p(\Omega; V)$ as the closure of $\mathcal{S}_V$ with respect to the norm

$$\|F\|_{k,p,V} = \left[\mathbb{E}(\|F\|_V^p) + \sum_{i=1}^k \mathbb{E}\left(\|D^i F\|_{V\otimes H^{\otimes i}}^p\right) \right]^{1/p}$$

Thus we have that $D : \mathbb{D}^{k,p}(V) \to L^p(\Omega; V \otimes H)$, and $D^i : \mathbb{D}^{k,p}(V) \to L^p(\Omega; V \otimes H^{\otimes i})$, etc

Remark. Some authors define the derivative first on the space

$$\mathcal{S}'_V := \{F = \psi(W(h_1), \dots, W(h_n)) \,|\, \psi \in C_p^\infty(\mathbb{R}^n; V), h_i \in H, n \in \mathbb{N}\}$$

by

$$DF := \sum_{i=1}^n \frac{\partial \psi}{\partial x_i}(W(h_1), \dots, W(h_n)) \otimes h_i$$

One can check that this agrees with the above definition on $\mathcal{S}_V \subseteq \mathcal{S}'_V$, and that both spaces of smooth cylinder functions have the same closure $\mathbb{D}^{k,p}(V)$ under the $\|\cdot\|_{k,p,V}$-norm.

Before moving on we extend the product rule given earlier to a formula for the directional derivative of the inner product of two elements of $\mathbb{D}^{1,p}(V)$:

Proposition 1.8. *Let $h \in H$ and let $F, G \in \mathbb{D}^{1,p}(V)$ be two differentiable V-valued random variables. Then*

$$D_h \langle F, G\rangle_V = \langle D_h F, G\rangle_V + \langle F, D_h G\rangle_V \tag{1.1}$$

Proof. (A similar statement is given in [1] without proof). We first prove it for $F, G \in \mathcal{S}'_V$. Let $F = \psi(W(h_1), \dots, W(h_n))$ and $G = \varphi(W(h_{n+1}), \dots, W(h_{n+m}))$ where $\psi \in C_p^\infty(\mathbb{R}^n; V)$, $\varphi \in C_p^\infty(\mathbb{R}^m; V)$ and $h_i \in H$. $F = \langle \eta, \xi\rangle_V$ is a real-valued random variable, we calculate its directional derivative:

$$\begin{aligned}
D_h\langle F, G\rangle_V &= \sum_{i=1}^{n+m} \frac{\partial}{\partial x_i} \langle \psi(x_1, \dots, x_n), \varphi(x_{n+1}, \dots, x_{n+m})\rangle_V \Big|_{x_j = W(h_j), j=1,\dots,n+m} \langle h_i, h\rangle_H \\
&= \sum_{i=1}^{n} \left\langle \frac{\partial \psi}{\partial x_i}(W(h_1), \dots, W(h_n)), \varphi(W(h_{n+1}), \dots, W(h_{n+m}))\right\rangle_V \langle h_i, h\rangle_H \\
&\quad + \sum_{i=n+1}^{n+m} \left\langle \psi(W(h_1), \dots, W(h_n)), \frac{\partial \varphi}{\partial x_i}(W(h_{n+1}), \dots, W(h_{n+m}))\right\rangle_V \langle h_i, h\rangle_H \\
&= \left\langle \sum_{i=1}^{n} \frac{\partial \psi}{\partial x_i}(W(h_1), \dots, W(h_n))\langle h_i, h\rangle_H, \varphi(W(h_{n+1}), \dots, W(h_{n+m}))\right\rangle_V \\
&\quad + \left\langle \psi(W(h_1), \dots, W(h_n)), \sum_{i=n+1}^{n+m} \frac{\partial \varphi}{\partial x_i}(W(h_{n+1}), \dots, W(h_{n+m}))\langle h_i, h\rangle_H \right\rangle_V \\
&= \langle D_h F, G\rangle_V + \langle F, D_h G\rangle_V
\end{aligned}$$

So the result is true for $F, G \in \mathcal{S}'_V$. For general $F, G \in \mathbb{D}^{1,p}$, take sequences $(F_n), (G_n) \subseteq \mathcal{S}'_V$ converging to F, G in $\mathbb{D}^{1,p}$ respectively and use standard approximation arguments to get the result. □

1.1.2 Properties

Now that most of the definitions are out of the way we can look at some of the properties of the derivative operator. The first will be a simple version of the chain rule for smooth cylindrical functions:

Proposition 1.9 (Chain rule I)**.** *Let $\varphi \in C_p^\infty(\mathbb{R}^m)$. Suppose that $F = (F^1, \dots, F^m)$ is a random vector whose components belong to the space $\mathcal{S}$. Then $\varphi(F) \in \mathcal{S}$, and*

$$D(\varphi(F)) = \sum_{i=1}^{m} \frac{\partial \varphi}{\partial x_i}(F) DF^i$$

Proof. For simplicity we assume that $m = 1$. Let $F = F^1$ have representation $F = \psi(W(h_1), \dots, W(h_n))$, where $\psi \in C_p^\infty(\mathbb{R}^n)$ and $h_i \in H$. Then $\xi = \varphi \circ \psi \in C_p^\infty(\mathbb{R}^n)$ and we have $\varphi(F) = \xi(W(h_1), \dots, W(h_n))$. So by the definition of D on $\mathcal{S}$ and the ordinary chain rule for derivatives we have

$$\begin{aligned}
D(\varphi(F)) &= \sum_{i=1}^{n} \frac{\partial \xi}{\partial x_i}(W(h_1), \dots, W(h_n))h_i \\
&= \sum_{i=1}^{n} \frac{\partial(\varphi \circ \psi)}{\partial x_i}(W(h_1), \dots, W(h_n))h_i \\
&= \sum_{i=1}^{n} \frac{\partial \varphi}{\partial x}(\psi(W(h_1), \dots, W(h_n)))\frac{\partial \psi}{\partial x_i}(W(h_1), \dots, W(h_n))h_i \\
&= \frac{\partial \varphi}{\partial x}(F)DF
\end{aligned}$$

□

Corollary 1.10. *Let $p \geq 1$ and $A \in \mathcal{F}$. Then the indicator function $\mathbf{1}_A$ belongs to $\mathbb{D}^{1,p}$ if and only if $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$*

Proof. (From [7, p. 8]) Let $\varphi \in C_p^\infty(\mathbb{R})$ be such that $\varphi(x) = x^2$ for $x \in [0,1]$. Apply the above proposition to get (using that $\mathbf{1}_A = (\mathbf{1}_A)^2$),

$$D\mathbf{1}_A = D(\mathbf{1}_A)^2 = 2\mathbf{1}_A D\mathbf{1}_A$$

Now if $\omega \in A$, the above says that $D\mathbf{1}_A(\omega) = 2D\mathbf{1}_A(\omega)$ and so $D\mathbf{1}_A(\omega) = 0$. If $\omega \notin A$ then clearly $D\mathbf{1}_A(\omega) = 0$. Thus $D\mathbf{1}_A = 0$ everywhere hence[3] $\mathbf{1}_A = \mathbb{E}\mathbf{1}_A = \mathbb{P}(A)$. □

By approximation we get a stronger version of the chain rule:

Proposition 1.11 (Chain rule II)**.** *Let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be a continously differentiable function with bounded partial derivatives, and fix $p \geq 1$. Suppose that $F = (F^1, \dots, F^m)$ is a random vector whose components belong to the space $\mathbb{D}^{1,p}$. Then $\varphi(F) \in \mathbb{D}^{1,p}$, and*

$$D(\varphi(F)) = \sum_{i=1}^{m} \frac{\partial \varphi}{\partial x_i}(F)DF^i \tag{1.2}$$

Proof. (Statement and idea of proof given in [7, p.5]) Assume first that $F^i \in \mathcal{S}$ for all i. Let η_ε be an approximation to the identity so that $\varphi * \eta_\varepsilon \in C_p^\infty(\mathbb{R}^m)$. Then $(\varphi * \eta_\varepsilon)(F)$ satisfies (1.2) for all $\varepsilon > 0$ by the previous proposition. Using that the derivative can be brought inside the convolution, we have

$$\begin{aligned}
\mathbb{E}\|D((\varphi * \eta_\varepsilon)(F)) - D(\varphi(F))\|_H^p &\leq C_p \sum_{i=1}^{m} \mathbb{E}\left\| \left(\frac{\partial \varphi}{\partial x_i} * \eta_\varepsilon\right)(F)DF^i - \frac{\partial \varphi}{\partial x_i}(F)DF^i \right\|_H^p \\
&= C_p \sum_{i=1}^{m} \mathbb{E}\left| \left(\frac{\partial \varphi}{\partial x_i} * \eta_\varepsilon\right)(F) - \frac{\partial \varphi}{\partial x_i}(F) \right|^p \|DF^i\|_H^p \\
&\leq C_p \sum_{i=1}^{m} \left\| \left(\frac{\partial \varphi}{\partial x_i} * \eta_\varepsilon\right) - \frac{\partial \varphi}{\partial x_i} \right\|_\infty^p \mathbb{E}\|DF^i\|_H^p \\
&\to 0 \ \text{ as } \varepsilon \to 0
\end{aligned}$$

[3] The fact that $DF = 0$ implies $F = \mathbb{E}F$ will be proved later. This is an expected result: if the derivative of a function vanishes, then the function is constant.

Using the closability of D gives that $\varphi(F)$ satisfies (1.2).

Now assume that $F^i \in \mathbb{D}^{1,p}$ for each i. Then there exists a sequence $(F_n^i)_{n\in\mathbb{N}} \subseteq \mathcal{S}$ such that $F_n^i \to F^i$ in $L^p(\Omega)$ and $DF_n^i \to DF^i$ in $L^p(\Omega; H)$ for each i. Now $\varphi(F_i^n) \to \varphi(F^i)$ in $L^p(\Omega)$ as the boundedness of the partial derivatives of φ mean that it's Lipschitz:

$$\begin{aligned}\mathbb{E}|\varphi(F_n) - \varphi(F)|^p &\leq K^p\,\mathbb{E}\|F_n - F\|_{\mathbb{R}^m}^p \\ &\leq C^p K^p\,\mathbb{E}\|F_n - F\|_{\ell^\infty}^p \\ &= C^p K^p\,\mathbb{E}\left(\max_i |F_n^i - F^i|^p\right) \\ &\to 0 \text{ as } n \to \infty\end{aligned}$$

where the second inequality comes from the equivalence of norms on $\mathbb{R}^m$. The convergence $D(\varphi(F_n)) \to D(\varphi(F))$ in $L^p(\Omega; H)$ also follows from the boundedness of the partial derivatives:

$$\begin{aligned}\mathbb{E}\|D(\varphi(F_n)) - D(\varphi(F))\|_H^p &\leq C_p \sum_{i=1}^m \mathbb{E}\left\|\frac{\partial\varphi}{\partial x_i}(F_n)DF_n^i - \frac{\partial\varphi}{\partial x_i}(F)DF^i\right\|_H^p \\ &\leq C_p K^p \sum_{i=1}^m \mathbb{E}\left\|DF_n^i - DF^i\right\|_H^p \\ &\to 0 \text{ as } n \to \infty\end{aligned}$$

Thus, by the closability of D, $D(\varphi(F))$ exists and satisfies (1.2). □

To extend the chain rule further to Lipschitz functions of $\mathbb{D}^{1,2}$ random variables we'll need the following result. This result is useful in its own right to decide whether a given random variable is in $\mathbb{D}^{1,2}$ or not, see [7, p. 7] for details.

Lemma 1.12. *Let $(F_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{D}^{1,2}$ such that $F_n \to F$ in $L^2(\Omega)$. If*

$$\sup_{n\in\mathbb{N}} \|DF_n\|_{L^2(\Omega;H)} < \infty$$

then $F \in \mathbb{D}^{1,2}$ and $DF_n \to DF$ in the weak topology of $L^2(\Omega; H)$

We can now give a final version of the chain rule:

Proposition 1.13 (Chain rule III). *Let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be Lipschitz, i.e. there exists a $K > 0$ such that*

$$|\varphi(x) - \varphi(y)| \leq K\|x - y\|_{\mathbb{R}^m}$$

for all $x, y \in \mathbb{R}^m$. Suppose that $F = (F^1, \ldots, F^m)$ is a random vector whose components belong to the space $\mathbb{D}^{1,2}$. Then $\varphi(F) \in \mathbb{D}^{1,2}$, and there exists a random vector $G = (G_1, \ldots, G_m)$ bounded by K such that

$$D(\varphi(F)) = \sum_{i=1}^m G_i DF^i$$

Remark. Note that this chain rule only applies to random vectors in $\mathbb{D}^{1,2}(\mathbb{R}^m)$, whereas the previous chain rule applies to random vectors in $\mathbb{D}^{1,p}(\mathbb{R}^m)$ for any $p \geq 1$.

Proof. (Following idea of [7, p. 8] and [11]) We approximate φ by smooth functions. Let η_ε be an approximation to the identity and define $\varphi_n = \varphi * \eta_{1/n}$ so that $\varphi_n \in C^\infty(\mathbb{R}^m)$, $\varphi_n \to \varphi$ uniformly and $\|\nabla\varphi_n\|_{\mathbb{R}^m} < K$ for all n. Now the uniform convergence of the φ_n gives that $\varphi_m(F) \to \varphi(F)$ in $L^2(\Omega)$. From the previous proposition we know that

$$D(\varphi_n(F)) = \sum_{i=1}^m \frac{\partial\varphi_n}{\partial x_i}(F)DF^i$$

and so the sequence $D(\varphi_n(F))$ is uniformly bounded in $L^2(\Omega; H)$. We can now apply Lemma 1.12 to get that $\varphi(F) \in \mathbb{D}^{1,2}$ and $D(\varphi_n(F)) \to D(\varphi(F))$ in the weak topology of $L^2(\Omega; H)$.

Now since the sequence $(\nabla\varphi_n(F))_{n\in\mathbb{N}}$ is bounded by K, there exists a subsequence $(\nabla\varphi_{n_k}(F))_{k\in\mathbb{N}}$ that converges to some random vector $G = (G_1, \ldots, G_m)$ in the weak topology of $L^2(\Omega; \mathbb{R}^m)$, with G bounded by K. Now we can take (weak) limits in the above expression for $D(\varphi_n(F))$ to get that

$$D(\varphi_{n_k}(F)) \to \sum_{i=1}^m G_i DF^i$$

weakly in $L^2(\Omega; H)$. By uniqueness of weak limits, we have that

$$D(\varphi(F)) = \sum_{i=1}^m G_i DF^i$$

□

Example 1.14. A nice application of Lemma 1.12 and the chain rule above is given in [8, p. 19]: let $H = L^2([0, 1])$ and let W be the Itô integral with respect to some Brownian motion B on $[0, 1]$. Define

$$M = \sup_{t\in[0,1]} B_t$$

The claim is that $M \in \mathbb{D}^{1,2}$, and $DM = \mathbf{1}_{[0,T]}$, where T is the (random) time where B attains its maximum on $[0, 1]$. The proof involves taking a countable dense subset $(t_k)_{k\in\mathbb{N}}$ of $[0, 1]$ and considering an approximation of M,

$$M_n = \max\{B_{t_1}, \ldots, B_{t_n}\}$$

Then M_n is a Lipschitz function of $\mathbb{D}^{1,2}$ random variables, and hence by the previous proposition is in $\mathbb{D}^{1,2}$ for all $n \in \mathbb{N}$. The result follows by showing that $M_n \to M$ in $L^2(\Omega)$ and that DM_n is uniformly bounded in $L^2(\Omega; H)$.

1.1.3 The white noise case

(Proofs of the statements given in this subsection are found in [6]). We now consider the case when the underlying Hilbert space is of the form $H = L^2(T, \mathcal{T}, \mu)$ for some σ-finite measure space without atoms $(T, \mathcal{T}, \mu)$. The definition of white noise is found in Appendix A.1, and that of its associated multiple integral is found in Appendix A.2.

Let $F \in \mathbb{D}^{1,2}$, then its Malliavin derivative DF is valued in $L^2(\Omega, \mathcal{F}, \mathbb{P}) \otimes L^2(T, \mathcal{T}, \mu)$. This space is naturally isomorphic to $L^2(\Omega \times T, \mathcal{F} \otimes \mathcal{T}, \mathbb{P} \otimes \mu)$, and as such we think of DF as being valued in $L^2(\Omega \times T)$. We denote by $D_t F$ the evaluation of DF at $t \in T$, so $D_t F \in L^2(\Omega)$ can be thought of as a stochastic process. Higher derivatives are treated similarly, with $D^k_{t_1,\dots,t_k} F := D^k F(\cdot, t_1, \dots, t_k) \in L^2(\Omega)$ being thought of as a random field.

Suppose now that F has chaos expansion

$$F = \sum_{n=0}^{\infty} I_n(f_n) \tag{1.3}$$

where the f_n are symmetric (and hence the decomposition is unique). It turns out that the derivative process of F has a very nice representation in terms of this expansion.

Proposition 1.15. *Let $f \in L^2(T^n)$ be symmetric. Then $I_n(f) \in \mathbb{D}^{1,2}$ and*

$$D_t I_n(f) = n I_{n-1}(f(\cdot, t)) \tag{1.4}$$

By linearity the following result is immediate:

Corollary 1.16. *Let $F \in \mathbb{D}^{1,2}$ have chaos expansion given by (1.3). Then*

$$D_t F = \sum_{n=1}^{\infty} n I_{n-1}(f_n(\cdot, t)) \tag{1.5}$$

Compare this with how the one-dimensional derivative operator acts on Taylor series!

Example 1.17. Let $H = L^2([0, 1])$ and let W be the Itô integral with respect to some Brownian motion B, se we are in the white noise case. Let $g \in H$ and consider the function $F = \exp(W(g))$. In Appendix A.2 we calculate the chaos expansion of F:

$$F = \exp\left(\int_0^1 g(s)dB_s\right) = \sum_{n=0}^{\infty} I_n\left(e^{\|g\|^2/2} \frac{g^{\otimes n}}{n!}\right)$$

We use the above corollary to calculate its derivative at time t:

$$\begin{aligned}
D_tF &= \sum_{n=1}^{\infty} nI_{n-1}\left(e^{\|g\|^2/2}\frac{g^{\otimes(n-1)}g(t)}{n!}\right)\\
&= g(t)\sum_{n=1}^{\infty} I_{n-1}\left(e^{\|g\|^2/2}\frac{g^{\otimes(n-1)}}{(n-1)!}\right)\\
&= g(t)\sum_{n=0}^{\infty} I_{n}\left(e^{\|g\|^2/2}\frac{g^{\otimes n}}{n!}\right)\\
&= g(t)\exp\left(\int_0^1 g(s)dB_s\right)
\end{aligned}$$

This is what would be expected from using the chain rule.

We now see how the derivative operator behaves on certain conditional expectations.

Proposition 1.18. *Let $F \in \mathbb{D}^{1,2}$ have chaos expansion given by (1.3) and let $A \in \mathcal{T}$. Then*

$$\mathbb{E}(F|\mathcal{F}_A) = \sum_{n=0}^{\infty} I_n(f_n \mathbf{1}_A^{\otimes n})$$

Proposition 1.19. *Let $F \in \mathbb{D}^{1,2}$ and $A \in \mathcal{T}$ with finite measure. Then $\mathbb{E}(F|\mathcal{F}_A) \in \mathbb{D}^{1,2}$ and*

$$D_t(\mathbb{E}(F|\mathcal{F}_A)) = \mathbb{E}(D_tF|\mathcal{F}_A)\mathbf{1}_A(t)$$

Corollary 1.20. *Let $F \in \mathbb{D}^{1,2}$ and suppose that F is $\mathcal{F}_A$ measurable for some $A \in \mathcal{T}$. Then $D_tF = 0$ $\mu \otimes \mathbb{P}$-a.e. on $A^c \times \Omega$*

Proof. F is $\mathcal{F}_A$ measurable, so $\mathbb{E}(F|\mathcal{F}_A) = F$. Applying the previous proposition gives that $D_tF = \mathbb{E}(D_tF|\mathcal{F}_A)\mathbf{1}_A(t) = 0$ for a.e. $t \in A^c$, i.e. $D_tF = 0$ $\mu \otimes \mathbb{P}$-a.e on $A^c \times \Omega$. □

Remark. This is called the *local property* of the derivative operator.

We can now take a moment to look about what the above results are saying in the case where $T = [0,1]$ equipped with the Lebesgue measure and W is the usual Itô integral. In this case, for symmetric $f \in L^2([0,1]^n)$ the map I_n is given by the multiple Itô integral:

$$I_n(f) = n!\int_0^1\int_0^{t_n}\cdots\int_0^{t_2} f(t_1,\ldots,t_n)dB_{t_1}\ldots dB_{t_n}$$

(The factor $n!$ accounts for the fact that we are only integrating over the simplex

$$\{(t_1,\ldots,t_n)\,|\,0 \le t_1 \le t_2 \le \ldots \le t_n \le 1\}$$

which occupies $1/n!$ of the volume of the whole cube $[0,1]^n$, see Appendix A.2). From Corollary 1.16 we then have that

$$
\begin{aligned}
D_t\left(\int_0^1\int_0^{t_n}\cdots\int_0^{t_2} f(t_1,\dots,t_n)dB_{t_1}\dots dB_{t_n}\right) &= D_t\left(\frac{1}{n!}I_n(f)\right)\\
&= \frac{1}{(n-1)!}I_{n-1}(f(\cdot,t))\\
&= \int_0^1\int_0^{t_{n-1}}\cdots\int_0^{t_2} f(t_1,\dots,t)dB_{t_1}\dots dB_{t_{n-1}}
\end{aligned}
$$

and so the outer integral is effectively stripped away.

Proposition 1.18 gives that, in the case where $A=[0,t]$ so that $\mathcal{F}_A=\mathcal{F}_t$,

$$
\mathbb{E}\left(\int_0^1\int_0^{t_n}\cdots\int_0^{t_2} f(t_1,\dots,t_n)dB_{t_1}\dots dB_{t_n}\middle|\mathcal{F}_t\right) = \int_0^t\int_0^{t_n}\cdots\int_0^{t_2} f(t_1,\dots,t_n)dB_{t_1}\dots dB_{t_n}
$$

This is indeed what would be expected from the martingale property of Itô integrals!

Now suppose that F is an $\mathcal{F}_{[s,t]}$-measurable $\mathbb{D}^{1,2}$ random variable, where $[s,t]\subseteq[0,1]$. The local property says that D_tF vanishes on $[s,t]^c$, but this is what is to be expected - perturbing the paths outside of $[s,t]$ isn't going to have any effect on the value of F, so its derivative is going to vanish there.

1.1.4 Diffusion processes

To get a feel for how the derivative operator should act on a diffusion, we first take an (unrigourous) look at a one-dimensional diffusion driven by a one-dimensional Brownian motion, following [3, pp. 25-26]. We work in the white noise case $H=L^2([0,T])$. Let B be a one-dimensional Brownian motion on $[0,T]$ with natural filtration $(\mathcal{F}_t)_{t\in[0,T]}$ and consider the SDE

$$
X_t = X_0 + \int_0^t a(X_s)dB_s + \int_0^t b(X_s)ds \tag{1.6}
$$

where $a,b:\mathbb{R}\to\mathbb{R}$ are regular enough to ensure existence and uniqueness of a strong solution $X=(X_t)_{t\in[0,T]}$. We *assume* that $X_t\in\mathbb{D}^{1,\infty}$ for all $t\in[0,T]$, and differentiate both sides of the SDE. We need to see first however how the derivative operator will act on each of the integrals.

By the linearity and closedness of D, it can be argued that D can be brought under the Lebesgue integral to give, for $u:[0,T]\times\Omega\to\mathbb{R}$ some $\mathcal{F}_\cdot$-adapted $\mathbb{D}^{1,\infty}$ process and $r\le t$,

$$
D_r\int_0^t u_s ds = \int_0^t D_r u_s ds = \int_r^t D_r u_s ds
$$

by the local property of D. The stochastic integral isn't as easy. Let $u : [0,T] \times \Omega \to \mathbb{R}$ be an elementary adapted process of the form $u_t = F\mathbf{1}_{(s_1,s_2]}$ where $s_1 < s_2$ and $F \in \mathbb{D}^{1,\infty}$ is $\mathcal{F}_{s_1}$-measurable. Then for $r \le t$,

$$\begin{aligned}
D_r \int_0^t u_s dB_s &= D_r \left(\int_{[0,r)} F\mathbf{1}_{(s_1,s_2]}(s) dB_s + \int_{[r,t]} F\mathbf{1}_{(s_1,s_2]}(s) dB_s \right) \\
&= D_r \int_0^T F\mathbf{1}_{(s_1,s_2]}(s)\mathbf{1}_{[r,s]}(s) dB_s \quad \text{(by local property of } D\text{)} \\
&= D_r(FW(\mathbf{1}_{(s_1,s_2]}\mathbf{1}_{[r,s]}) \\
&= (D_r F)W(\mathbf{1}_{(s_1,s_2]}\mathbf{1}_{[r,s]}) + F\mathbf{1}_{(s_1,s_2]}(r) \quad \text{(by the product rule)} \\
&= \int_0^T \mathbf{1}_{[r,t]}(s) D_r(F\mathbf{1}_{(s_1,s_2]}(s)) dB_s + u_r \quad \text{(by linearity of } D\text{)} \\
&= \int_r^t D_r u_s dB_s + u_r
\end{aligned}$$

By linearity this holds for all adapted $\mathbb{D}^{1,\infty}$ elementary processes, and so by approximation it can be seen to hold for general adapted $\mathbb{D}^{1,\infty}$ processes. We can apply this to the SDE (1.6) to get that

$$\begin{aligned}
D_r X_t &= a(X_r) + \int_r^t D_r a(X_s) dB_s + \int_r^t D_r b(X_s) ds \\
&= a(X_r) + \int_r^t a'(X_s) D_r X_s dB_s + \int_r^t b'(X_s) D_r X_s ds
\end{aligned}$$

by the chain rule. Thus, for fixed r the derivative process $D_r X_t$ satisfies the SDE

$$dY_t = a'(X_t) Y_t dB_t + b'(X_t) Y_t dt, \quad Y_0 = a(X_r)$$

for $t > r$.

Now we proceed to look at the general case rigourously. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the canonical probability space associated with an m-dimensional Brownian motion $B = (B^1, \ldots, B^m)$ on a finite interval $[0,T]$, thus $\Omega = C_0([0,T]; \mathbb{R}^m)$, $\mathbb{P}$ is the Wiener measure and $\mathcal{F}$ is the completion of the σ-algebra generated by B. We let $H = L^2([0,T]; \mathbb{R}^m)$ and let W be the Itô integral, i.e. for $h \in H$ we have

$$W(h) = \int_0^T \langle h_s, dB_s \rangle_{\mathbb{R}^m} = \sum_{i=1}^m \int_0^T h_s^i dB_s^i$$

In this case, the derivative DF of a random variable $F \in \mathbb{D}^{1,2}$ will be an m-dimensional process $(D_t^1 F, \ldots, D_t^m F)_{t \in [0,T]}$. We are interested in the diffusion defined by the SDE

$$X_t = X_0 + \sum_{i=1}^m \int_0^t V_i(X_s) dB_s^i + \int_0^t V_0(X_s) ds \tag{1.7}$$

where $V_i : \mathbb{R}^n \to \mathbb{R}^n$, $i = 0, \ldots, m$, are measurable vector fields. If we assert that the V_is are globally Lipschitz then we know that (1.7) has a unique strong solution $X = (X_t)_{t\in[0,T]}$. If we assume that these vetor fields are continuously differentiable also, we get the Malliavin differentiability of the solution:

Proposition 1.21. *Suppose that the vector fields V_i in the SDE (1.7) are continuously differentiable and globally Lipschitz. Then for all $t \in [0,T]$ and all $i = 1, \ldots, m$ we have $X_t^i \in \mathbb{D}^{1,\infty}$, and the derivative process $D_r^j X_t$ satisfies the following linear SDE for $r \le t$:*

$$D_r^j X_t = V_j(X_r) + \sum_{i=1}^m \int_r^t \partial V_i(X_s) D_r^j X_s dB_s^i + \int_r^t \partial V_0(X_s) D_r^j X_s ds \tag{1.8}$$

where ∂V_k denotes the Jacobian matrix of V_k, i.e. $(\partial V_k)_j^i = \partial_j V_k^i$.

The proof uses Picard iterations and can be found in [6].

Remarks. (i) If we only assume that the vector fields are Lipschitz then we still have $X_t^i \in \mathbb{D}^{1,\infty}$, and (1.8) holds with the $\partial V_i(X_s)$ processes replaced by some bounded adapted processes. The proof of this uses the final version of the chain rule proved earlier.

(ii) An analogous result holds for diffusions defined by Stratonovich integrals. If $X = (X_t)_{t\in[0,T]}$ satisfies the SDE

$$X_t = X_0 + \sum_{j=1}^m \int_0^t V_j(X_s) \circ dB_s^j + \int_0^t V_0(X_s) ds$$

with the same conditions on the vector fields V_i as in the above proposition, then $X_t^i \in \mathbb{D}^{1,\infty}$ for all $t \in [0,T]$ and all $i = 1, \ldots, m$, and for $r \le t$ the derivative process $D_r^j X_t$ satisfies the linear SDE

$$D_r^j X_t = V_j(X_r) + \sum_{i=1}^m \int_r^t \partial V_i(X_s) D_r^j X_s \circ dB_s^i + \int_r^t \partial V_0(X_s) D_r^j X_s ds$$

We aim to get a simpler expression for the derivative. We follow [7, pp. 47-48]. Let J be the $n \times n$ matrix-valued process defined by

$$J_t = I + \sum_{k=1}^m \int_0^t \partial V_k(X_s) J_s dB_s^k + \int \partial V_0(X_s) J_s ds$$

This process is the Jacobian of the solution map of the SDE (1.7), see (2.2) later for a derivation (in the Stratonovich case). We claim that J_t is invertible for all $t \in [0,T]$ and its inverse is given by the $n \times n$ matrix-valued process Z defined by

$$Z_t = I - \sum_{k=1}^m \int_0^t X_s \partial V_k(X_s) dB_s^k - \int_0^t Z_s \left(\partial V_0(X_s) - \sum_{k=1}^m \partial V_k(X_s) \partial V_k(X_s) \right) ds$$

This is checked by using Itô's formula to see that $Z_t J_t = J_t Z_t = I$. We hence write $Z_t = J_t^{-1}$.

Now let $D_r X_t$ be the $n \times m$ matrix given by $(D_r X_t)^i_j = D^j_r X^i_t$. We claim that

$$D_r X_t = J_t J_r^{-1} V(X_r) \tag{1.9}$$

where $V(X_s)$ is the $n \times m$ matrix obtained by concatenating the m vectors $V_i(X_s)$, $i = 1, \dots, m$. This is checked by verifying that the process $(J_t J_r^{-1} V(X_r))_{t>r}$ satisfies the SDE (1.7).

We use (1.9) to get an expression for the *Malliavin matrix* of the process X.

Definition 1.22. *Let $F = (F^1, \dots, F^m)$ be a random vector whose components belong to $\mathbb{D}^{1,2}$. Then the* Malliavin matrix *of F is the random symmetric non-negative definite matrix $\mathcal{M}$ given by*

$$\mathcal{M}^i_j = \langle DF^i, DF^j \rangle_H$$

So for the process X, we have for fixed $t \in [0, T]$,

$$\begin{aligned}
(\mathcal{M}^i_j)_t &= \langle DX^i_t, DX^j_t \rangle_H \\
&= \sum_{k=1}^m \int_0^T D^k_r X^i_t D^k_r X^j_t \, dr \\
&= \sum_{k=1}^m \int_0^t (D_r X_t)^i_k ((D_r X_t)^*)^k_j \, dr
\end{aligned}$$

using the local property of D, and so

$$\mathcal{M}_t = \int_0^t (D_r X_t)(D_r X_t)^* dr \tag{1.10}$$

Using (1.9) we see that

$$\begin{aligned}
\mathcal{M}_t &= \int_0^t J_t J_r^{-1} V(X_r) V^*(X_r) (J_r^{-1})^* J_t^* dr \\
&= J_t \mathcal{C}_t J_t^*
\end{aligned}$$

where

$$\mathcal{C}_t = \int_0^t J_r^{-1} V(X_r) V^*(X_r) (J_r^{-1})^* dr \tag{1.11}$$

is the *reduced Malliavin matrix.* This form can be useful since the integrand is adapted.

We later use Malliavin calculus to look at the problems of existence and regularity of a density with respect to the Lebesgue measure of the law of a diffusion process.

1.2 The divergence operator and Skorokhod integral

In Appendix A.1, in the case where the isonormal Gaussian process is given by the Paley-Wiener integral, a prototype for a 'divergence operator' is introduced for certain constant vector fields. It's characterised by the integration by parts formula, in which it is viewed as the adjoint operator of the derivative operator analogously to the Euclidean divergence operator. We use this idea to define it in the more general case, that is for possibly non-deterministic vector fields and general isonormal Gaussian processes.

1.2.1 The divergence operator

Definition 1.23. *: The divergence operator δ is an unbounded operator on $L^2(\Omega; H)$ with values in $L^2(\Omega)$ whose domain is given by*

$$\operatorname{Dom}\delta := \{u \in L^2(\Omega; H) \,|\, |\mathbb{E}\langle DF, u\rangle_H| \leq C_u \|F\|_{1,2} \text{ for all } F \in \mathbb{D}^{1,2}\}$$

If $u \in \operatorname{Dom}\delta$, then $\delta(u)$ is the element of $L^2(\Omega)$ characterised by the duality relationship

$$\mathbb{E}(F\delta(u)) = \mathbb{E}\langle DF, u\rangle_H \tag{1.12}$$

Remark. Notes that (1.12) can be rewritten as

$$\langle F, \delta(u)\rangle_{L^2(\Omega)} = \langle DF, u\rangle_{L^2(\Omega;H)}$$

which makes it clear that $\delta = D^*$, the adjoint of the derivative operator.

Proposition 1.24. *(Properties of the divergence operator)*

(i) *For any $u \in \operatorname{Dom}\delta$, $\mathbb{E}(\delta(u)) = 0$*

(ii) *δ is a linear and closed operator on* $\operatorname{Dom}\delta$

(iii) *Let $u \in \mathcal{S}_H$ be an H-valued smooth cylindrical random variable of the form $u = \sum_{j=1}^n F_j h_j$ where $F_j \in \mathcal{S}$ and $h_j \in H$. Then $u \in \operatorname{Dom}\delta$ and*

$$\delta(u) = \sum_{j=1}^n F_j W(h_j) - \sum_{j=1}^n \langle DF_j, h_j\rangle_H$$

In particular, $\delta(h) = W(h)$ for $h \in H$.

(iv) *Let $u \in \mathcal{S}_H$ and $h \in H$, then*[4]

$$D_h(\delta(u)) = \delta(D_h u) + \langle h, u\rangle_H$$

[4]Recall that we can identify $H \otimes H$ with $\mathcal{L}_{HS}(H, H)$, the space of Hilbert-Schmidt operators from H to itself. Thus, as $u = \sum_{j=1}^n F_j h_j$ is H-valued and hence $Du = \sum_{j=1}^n h_j \otimes DF_j$ is $H \otimes H$-valued, we define $D_h u$ as the H valued element $(Du)(h) := \sum_{j=1}^n DF_j(h) h_j = \sum_{j=1}^n D_h F_j \cdot h_j$

(v) *Let $u, v \in \mathcal{S}_H$ and let $(e_i)_{i\in\mathbb{N}}$ be a complete orthonormal system in H, then*

$$\begin{aligned}\mathbb{E}(\delta(u)\delta(v)) &= \mathbb{E}\langle u, v\rangle_H + \mathbb{E}\left[\sum_{i,j=1}^{\infty} D_{e_j}\langle u, e_i\rangle_H D_{e_i}\langle v, e_j\rangle_H\right] \\ &= \mathbb{E}\langle u, v, \rangle_H + \mathbb{E}\mathrm{Tr}(DuDv)\end{aligned}$$

(vi) *Let $u \in \mathcal{S}_H$ and $F \in \mathcal{S}$, then*

$$\delta(Fu) = F\delta(u) - \langle DF, u\rangle_H$$

Proofs of the above properties are found in [7] and [8].

Remark. Taking $u = v$ in part (v) of the above proposition yields

$$\|\delta(u)\|^2_{L^2(\Omega)} \leq \|u\|^2_{L^2(\Omega;H)} + \|Du\|^2_{L^2(\Omega;H\otimes H)}$$

for $u \in \mathcal{S}_H$. The right hand side of this inequality is precisely the $\mathbb{D}^{1,2}(H)$ norm of u, and so by an approximation argument we see that $\mathbb{D}^{1,2}(H) \subseteq \mathrm{Dom}\,\delta$

1.2.2 The Skorokhod integral

We now look at the divergence operator in a less abstract case: the white noise case. In this case the divegence operator is called the Skorokhod integral. Let $H = L^2(T, \mathcal{T}, \mu)$ with $(T, \mathcal{T}, \mu)$ atomless and σ-finite as usual. For $u \in \mathrm{dom}\,\delta$ we often use the notation

$$\delta(u) = \int_T u_t \delta B_t$$

which makes it clear the the divergence should be viewed as a sort of stochastic integral. We will soon see that it is in fact a strict extension of the Itô integral to non-adapted integrands, i.e. it coincides with the Itô integral when the integrand is adapted.

The results from Proposition 1.24 immediately apply in this case, for example part (v) of the proposition can now be written

$$\mathbb{E}\left(\int_T u_t\delta B_t \int_T v_t\delta B_t\right) = \mathbb{E}\left(\int_T u_t v_t d\mu(t)\right) + \mathbb{E}\left(\int_T\int_T (D_s u_t)(D_t v_s) d\mu(s)d\mu(t)\right)$$

This is know as the *Shigekawa-Nualart-Pardoux identity*, and can be seen to be an extension of the Itô isometry.

We now have chaos expansions at our disposal, and so we give the analogous result to Corollary 1.16 for the Skorokhod integral, see [11] for a proof.

Proposition 1.25. *Suppose that $u \in L^2(\Omega \times T)$ has chaos expansion*

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

for a sequence of functions $f_n \in L^2(T^{n+1})$ which are symmetric in the first n variables. Then $u \in \operatorname{dom}\delta$ if and only if $\sum_{n=0}^{\infty} I_{n+1}(\widetilde{f}_n)$ converges in $L^2(\Omega)$, and in this case

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

Example 1.26. We calculate a couple of Skorokhod integrals using the above proposition, working in the case where $T = [0,1]$ and W is the Itô integral with respect to some Brownian motion B.

(i) We again look at the random variable $F = \exp(W(g))$ for some $g \in H$. Recall this had chaos expansion

$$F = \exp\left(\int_0^1 g(t)dB_t\right) = \sum_{n=0}^{\infty} I_n\left(e^{\|g\|^2/2}\frac{g^{\otimes n}}{n!}\right)$$

Note that F is not $\mathcal{F}_t$ measurable for any $t < 1$ so its Itô integral doesn't make sense. Since there is no time dependence, we don't need to worry about the symmetrisation of the integral kernels, and we just have

$$\int_0^1 \exp\left(\int_0^1 g(t)dB_t\right)\delta B_t = \sum_{n=0}^{\infty} I_{n+1}\left(e^{\|g\|^2/2}\frac{g^{\otimes n}}{n!}\right)$$

In particular, setting $g = 1$ yields

$$\begin{aligned}\int_0^1 e^{B_1}\delta B_t &= \sum_{n=0}^{\infty} \frac{e^{1/2}}{n!} I_{n+1}(1) \\ &= \sqrt{e}\sum_{n=0}^{\infty} \frac{1}{n!} H_{n+1}(B_1)\end{aligned}$$

where H_{n+1} is the $(n+1)$th Hermite polynomial. Note that even though there is no time dependence, we can't simply bring the integrand out of the integral like a constant.

(ii) Now we try a (non-adapted) process, $(B_{1-t})_{t\in[0,1]}$. Its chaos expansion contains just one term:

$$B_{1-t} = \int_0^1 \mathbf{1}_{[0,1-t]}(s)dB_s = I_1(\mathbf{1}_{[0,1-t]})$$

Thus we have $f_1(s,t) = \mathbf{1}_{[0,1-t]}(s)$ and $f_n = 0$ for all $n \neq 1$. Symmetrising this we get

$$\widetilde{f}_1(s,t) = \frac{1}{2}\left(\mathbf{1}_{[0,1-t]}(s) + \mathbf{1}_{[0,1-s]}(t)\right)$$

and so

$$\begin{aligned}
\int_0^1 B_{1-t} dB_t &= I_2(\widetilde{f}_1) \\
&= \int_0^1 \int_0^t \mathbf{1}_{[0,1-t]}(s) dB_s dB_t + \int_0^1 \int_0^t \mathbf{1}_{[0,1-s]}(t) dB_s dB_t \\
&= 2 \int_0^1 B_{t\wedge(1-t)} dB_t \\
&= 2 \int_0^{\frac{1}{2}} B_t dB_t + 2 \int_{\frac{1}{2}}^1 B_{1-t} dB_t \\
&= B_{\frac{1}{2}}^2 - \frac{1}{2} + 2 \int_{\frac{1}{2}}^1 B_{1-t} dB_t
\end{aligned}$$

We now work in the case where $H = L^2([0,T])$ and W is given by the Itô integral with respect to some Brownian motion B with filtration $(\mathcal{F}_t)_{t\in[0,T]}$. Denote by $L_a^2([0,T])$ the space of progressively measurable[5] stochastic processes u such that $\mathbb{E}\int_0^T u_t^2 dt < \infty$. We finish this section with the following result:

Proposition 1.27. *$L_a^2([0,T]) \subseteq \operatorname{Dom}\delta$, and the Skorokhod integral of $u \in L_a^2([0,T])$ coincides with its Itô integral, i.e.*

$$\int_0^T u_t \delta B_t = \int_0^T u_t dB_t$$

for all $u \in L_a^2([0,T])$.

Proof. (Following [7, pp. 16-17]) We prove it first for elementary processes. Let $u \in L_a^2([0,T])$ be given by

$$u_t = \sum_{i=0}^{n-1} F_i \mathbf{1}_{(t_i,t_{i+1}]}(t)$$

where $0 = t_0 < \ldots < t_n \leq T$ and each F_i is square integrable and $\mathcal{F}_{t_i}$-measurable. Let $(F_i^n)_{n\in\mathbb{N}}$ be a sequence of smooth cylindrical $\mathcal{F}_{t_i}$-measurable random variables converging to F_i in $L^2(\Omega)$. By Proposition 1.24(vi) we have

$$\begin{aligned}
\int_0^T F_i^n \mathbf{1}_{(t_i,t_{i+1}]}(t) \delta B_t &= F_i^n W((t_i,t_{i+1}]) - \langle DF_i^n, \mathbf{1}_{(t_i,t_{i+1}]} \rangle_H \\
&= F_i^n (W_{t_{i+1}} - W_{t_i}) - \int_{t_i}^{t_{i+1}} D_t F_i^n dt \\
&= F_i^n (W_{t_{i+1}} - W_{t_i})
\end{aligned}$$

[5]We're dealing only with sample continuous processes so 'progressively measurable' and 'adapted' are interchangable

by the local property of D. Now $F_i^n \mathbf{1}_{(t_i,t_{i+1}]} \to F_i \mathbf{1}_{(t_i,t_{i+1}]}$ in $L^2(\Omega; H)$ and $F_i^n(W_{t_{i+1}} - W_{t_i}) \to F_i(W_{t_{i+1}} - W_{t_i})$ in $L^2(\Omega)$. Since δ is closed we therefore have that

$$\begin{aligned}\int_0^T u_t \delta B_t &= \sum_{i=0}^{n-1} F_i(W_{t_{i+1}} - W_{t_i}) \\ &= \int_0^T u_t dB_t\end{aligned}$$

The result is hence true for elementary processes. These processes are dense in $L_a^2([0,T])$, so for general elements of $L_a^2([0,T])$ we can approximate and use the closedness of δ to get the result. $\square$

1.3 Clark-Ocone theorem

Throughout this section we work in the white noise case $H = L^2([0,T])$. Let B be a one-dimensional Brownian motion on $[0,T]$ and let $(\mathcal{F}_t)_{t\in[0,T]}$ be the completion of the filtration generated by B. Again denote by $L_a^2([0,T])$ the space of progressively measurable stochastic processes u such that $\mathbb{E}\int_0^T u_t^2 dt < \infty$. By the martingale representation theorem, for any $\mathcal{F}_T$-measurable square integrable random variable F there is the decomposition

$$F = \mathbb{E}F + \int_0^T \varphi_t \, dB_t \tag{1.13}$$

for some almost surely unique process $\varphi \in L_a^2([0,T])$. However, we do not know what the process φ is, we just know that it exists. Fortunately, if $F \in \mathbb{D}^{1,2}$ we can in fact get an explicit formula for φ in terms of the Malliavin derivative of F:

Theorem 1.28 (Clark-Ocone)**.** *Let $F \in \mathbb{D}^{1,2}$ be $\mathcal{F}_T$-measurable. Then*

$$F = \mathbb{E}(F) + \int_0^T \mathbb{E}(D_t F | \mathcal{F}_t) dB_t \tag{1.14}$$

Remark. The above theorem also holds true for $F \in \mathbb{D}^{1,2}(V)$ for some separable Hilbert space V, where the integral is with respect to a cylindrical Brownian motion on some separable Hilbert space G. The proof is no more complicated than the one presented below, but since we haven't covered all of the requisite details we stick to the real-valued case. See [1] for details.

Proof. This may be proved using chaos expansions, see for example [9, p. 44]. We give a more direct proof, following [7, pp. 17-18]. F is square integrable and $\mathcal{F}_T$ measurable, so we know that there exists a process $\varphi \in L_a^2([0,T])$ such that (1.13) holds. Let u be any element of $L_a^2([0,T])$, so that its Itô integral coincides with its Skorokhod integral. By the Itô isometry, we have

$$\mathbb{E}(\delta(u)F) = \int_0^T \mathbb{E}(u_t \varphi_t) dt \tag{1.15}$$

since the expectation of an Itô integral vanishes. On the other hand, the duality relationship (1.12) for the Skorokhod integral gives

$$\begin{aligned}\mathbb{E}(\delta(u)F) &= \mathbb{E}\left(\int_0^T u_t D_t F dt\right) \\ &= \int_0^T \mathbb{E}(u_t \mathbb{E}(D_t F|\mathcal{F}_t))dt \end{aligned} \tag{1.16}$$

since u is adapted. Now (1.15) and (1.16) together tell us that

$$\mathbb{E}\langle u, \varphi\rangle_H = \mathbb{E}\langle u, \mathbb{E}(D_. F|\mathcal{F}_.)\rangle_H$$

for every $u \in L_a^2([0,T])$. It follows that $\varphi_t = \mathbb{E}(D_t F|\mathcal{F}_t)$ almost surely. □

Corollary 1.29. *If $F \in \mathbb{D}^{1,2}$ has $DF = 0$, then $F = \mathbb{E}F$.*

Example 1.30. If $(M_t)_{t\in[0,T]}$ is a uniformly integrable martingale, then we can apply Clark-Ocone to M_T and take conditional expectations to obtain

$$M_t = \mathbb{E}M_0 + \int_0^t \mathbb{E}(D_s M_T|\mathcal{F}_s)dB_s$$

Consider for example the exponential martingale $M_t = e^{B_t - \frac{1}{2}t^2}$. Then by the chain rule we have

$$\begin{aligned}D_t M_T &= e^{B_T - \frac{1}{2}T^2}\mathbf{1}_{[0,T]}(t) \\ &= e^{B_T - \frac{1}{2}T^2} \\ &= M_T\end{aligned}$$

Thus we have

$$M_t = 1 + \int_0^t M_s dB_s$$

which is what would be expected.

We can also use the Clark-Ocone theorem the other way around: if we don't know what $D_t F$ is, we may still be able to find $\mathbb{E}(D_t F|\mathcal{F}_t)$.

Example 1.31. (Following an exercise from [9]) Let X be the Itô diffusion given by

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt, \quad X_0 = x \in \mathbb{R}$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ and $b : \mathbb{R} \to \mathbb{R}$ are Lipschitz functions of at most linear growth. This ensures that there is a unique strong solution $X_t = X_t^x, t \in [0,T]$. Suppose further that there exists $\delta > 0$ such that $|\sigma(x)| \geq \delta$ for all $x \in \mathbb{R}$, and let $f \in C^2(\mathbb{R})$ be such that $\mathbb{E}|f(X_t^x)| < \infty$ for all x, t. Define

$$u(t,x) = P_t f(x) = \mathbb{E}f(X_t^x)$$

By a standard result in stochastic analysis, $u \in C^{1,2}(\mathbb{R}^+ \times \mathbb{R})$ and u satisfies the Kolmogorov backward equation:

$$\frac{\partial u}{\partial t} = \frac{1}{2}\sigma(x)^2 \frac{\partial^2 u}{\partial x^2} + b(x)\frac{\partial u}{\partial x} \tag{1.17}$$

We are interested in the random variable $f(X_T)$. Applying Itô's formula to the process $g(t, X_t)$ where $g(t, x) = P_{T-t}f(x)$ and using (1.17) we obtain

$$f(X_T) = P_T f(x) + \int_0^T \sigma(X_t) \frac{\partial}{\partial x}\bigg|_{x=X_t} P_{T-t}f(x)\, dB_t$$

So by the Clark-Ocone theorem, we see that

$$\mathbb{E}(D_t f(X_T)|\mathcal{F}_t) = \sigma(X_t)\frac{\partial}{\partial x}\bigg|_{x=X_t} P_{T-t}f(x)$$

This can be made more explicit in simple cases. Suppose now that X_t is a geometric Brownian motion, so that $\sigma(x) = \alpha x$ and $b(x) = \rho x$ for some constants α, ρ, and let $f(x) = x$. Then $P_t f(x) = \mathbb{E}X_t^x$ satisfies the integral equation

$$\mathbb{E}X_t^x = x + \int_0^t \rho\, \mathbb{E}X_s^x ds$$

and so $\mathbb{E}X_t^x = xe^{\rho t}$. We hence have that

$$\begin{aligned}\mathbb{E}(D_t X_T|\mathcal{F}_t) &= \alpha X_t e^{\rho(T-t)} \\ &= \alpha x \exp\left(\rho T - \frac{1}{2}\alpha^2 t + \alpha B_t\right)\end{aligned}$$

using the explicit formula for the solution X_t.

An nice application of the Clark-Ocone theorem is in the proof of Sobolev inequalities. Since these are not directly related to the project, the statements and proofs can be found in Appendix A.3.

2 Applications

2.1 Absolute continuity of probability laws

(Proofs of statements in this section can be found in [6, §2]). Malliavin calculus can be used to determine whether the law of a random variable admits a density with respect to the Lebesgue measure. This is done by analysing the Malliavin matrix $\mathcal{M}$ of the random variable, defined in the previous chapter. As usual, we let W be an isonormal Gaussian process with associated real separable Hilbert space H, defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathcal{F}$ generated by W.

We state the following theorem without proof concerning the density of the law of F.

Theorem 2.1. *Let $F = (F^1, \dots, F^m)$ be a random vector satisfying*

(i) $F^i \in \mathbb{D}^{1,2}$ for all $i = 1, \dots, m$

(ii) The Malliavin matrix $\mathcal{M}$ of F is invertible almost surely

Then the law of F is absolutely continuous with respect to the Lebesgue measure.

When $F \in \mathbb{D}^{1,2}$ is a one dimensional random variable, we may be able to get a formula for the density of the law of F:

Proposition 2.2. *Let $F \in \mathbb{D}^{1,2}$ be such that $DF/\|DF\|_H^2 \in \operatorname{Dom}\delta$. Then the law of F has a continuous and bounded density given by*

$$p(x) = \mathbb{E}\left(\mathbf{1}_{\{F>x\}}\delta\left(\frac{DF}{\|DF\|_H^2}\right)\right)$$

Theorem 2.1 gave sufficient conditions for the existence of a density, but didn't tell us anything about its regularity. If we have stronger conditions on the random vector F, we can deduce that the density will be smooth.

Theorem 2.3. *Let $F = (F^1, \dots, F^m)$ be a random vector satisfying*

(i) $F^i \in \mathbb{D}^{\infty}$ for all $i = 1, \dots, m$

(ii) The Malliavin matrix $\mathcal{M}$ of F satisfies $\mathbb{E}(\det \mathcal{M})^{-p} < \infty$ for all $p > 1$

Then the law of F admits an infinitely differentiable density with respect to the Lebesgue measure.

Remarks. (i) A random vector satisfying the hypotheses of the above theorem is said to be *non-degenerate.*

(ii) If $\mathcal{M}$ is invertible, a sufficient condition for condition (ii) to hold above is that $\mathbb{E}\|\mathcal{M}^{-1}\|^p < \infty$ for all $p > 1$. To see this, note that $\mathcal{M}$ is a symmetric positive definite matrix and hence so is its inverse. Let $(\lambda_i)_{i=1}^m$ be the eigenvalues of $\mathcal{M}$, which are all postive. Then it holds that

$$\begin{aligned}\|\mathcal{M}^{-1}\| &= \max_i \frac{1}{\lambda_i} \\ &= \left(\min_i \lambda_i\right)^{-1}\end{aligned}$$

Now since $\det \mathcal{M} = \Pi_{i=1}^m \lambda_i$, we have

$$\det \mathcal{M} \geq \left(\min_i \lambda_i\right)^m$$

and therefore

$$\mathbb{E}(\det \mathcal{M})^{-p/m} \leq \mathbb{E}\|\mathcal{M}^{-1}\|^p < \infty$$

for all $p > 1$, which gives the result. This condition will be used in the following section.

2.2 Hörmander's theorem

An important application of the Malliavin calculus is proving the existence and smoothness properties of the density of the law of solutions of certain hypoelliptic SDEs. We follow Hairer's paper [5] and fill in some details, especially those of Lemma 2.9 and Theorem 2.10. As in section 1.1.4, let $\Omega = C_0([0,1];\mathbb{R}^m)$, $\mathbb{P}$ be the Wiener measure and $\mathcal{F}$ be the completion of the σ-algebra generated by the Brownian motion $B_t(\omega) = \omega(t)$. Let $H = L^2([0,1];\mathbb{R}^m)$. We will be interested in SDEs of the form

$$dX_t = V_0(X_t)dt + \sum_{i=1}^{m} V_i(X_t) \circ dB_t^i \tag{2.1}$$

where each $V_i : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth vector field. We will want to impose further conditions on the vector fields in order to ensure the existence and smoothness of the densities. Recall that the *Lie Bracket* of two smooth vector fields $X, Y : \mathbb{R}^n \to \mathbb{R}^n$ is the vector field $[X,Y] : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$[X,Y](x) = X(Y(x)) - Y(X(x))$$

for $x \in \mathbb{R}^n$. Because of the lack of curvature in $\mathbb{R}^n$, we may rewrite this as

$$\begin{aligned}[X,Y](x) &= (\nabla_X Y)(x) - (\nabla_Y X)(x) \\ &= \partial X(x) \cdot Y(x) - \partial Y(x) \cdot X(x)\end{aligned}$$

where ∂X, ∂Y are the Jacobian matrices of X, Y respectively. With this in mind we introduce the following condition.

Definition 2.4 (Parabolic Hörmander condition)**.** *Given an SDE (2.1), define the collections of vector fields $\mathcal{V}_k$ by*

$$\mathcal{V}_0 = \{V_i \,|\, i > 0\}, \quad \mathcal{V}_{k+1} = \mathcal{V}_k \cup \{[U, V_j] \,|\, U \in \mathcal{V}_k, j \geq 0\}$$

For $x \in \mathbb{R}^n$, define also the vector spaces

$$\mathcal{V}_k(x) = \operatorname{span}\{V(x) \,|\, V \in \mathcal{V}_k\}$$

The SDE (2.1) is said to satisfy the parabolic Hörmander condition *if for every $x \in \mathbb{R}^n$,*

$$\bigcup_{k=1}^{\infty} \mathcal{V}_k(x) = \mathbb{R}^n$$

Hörmander's theorem basically states that if the parabolic Hörmander condition is satisfied by the SDE (2.1), along with some other assumptions to be defined later, then the law of the X_t has a smooth density with respect to the Lebesgue measure.

Let $\Phi_{s,t} : \mathbb{R}^n \to \mathbb{R}^n$ be the two-parameter family of solution maps of the SDE (2.1), so $X_t = \Phi_{s,t}(X_s)$ for $s \leq t$ and $\Phi_{t,u} \circ \Phi_{s,t} = \Phi_{s,u}$. Denote also $\Phi_t = \Phi_{0,t}$. Now for a given initial condition X_0 denote by $J_{s,t}$ the Jacobian of $\Phi_{s,t}$ evaluated at X_s. By the chain rule we have that $J_{s,u} = J_{t,u}J_{s,t}$. Finally, denote by $J_{s,t}^{(k)}$ the kth derivative of $\Phi_{s,t}$ evaluated at X_s.

We can obtain a SDE for $J_t := J_{0,t}$ by differentiating (2.1) with respect to X_0:

$$d\Phi_t(X_0) = V_0(\Phi_t(X_0))dt + \sum_{i=1}^{m} V_i(\Phi_t(X_0)) \circ dB_t^i, \;\; \Phi_0(X_0) = X_0$$

and so

$$\begin{aligned} dJ_t(X_0) &= \partial V_0(\Phi_t(X_0))\partial\Phi_t(X_0)dt + \sum_{i=1}^{m} \partial V_i(\Phi_t(X_0))\partial\Phi_t(X_0) \circ dB_t^i \\ &= \partial V_0(X_t)J_t dt + \sum_{i=1}^{m} \partial V_i(X_t)J_t \circ dB_t^i, \;\; J_0 = I \end{aligned} \tag{2.2}$$

where ∂F is used to denote the Jacobian of a vector field F. $J_{s,t}$ satisfies this same equation, but with initial condition $J_{s,s} = I$.

From the previous subsection, we know that

$$D_r^j X_t = V_j(X_r) + \int_r^t \partial V_0(X_s)D_r^j X_s ds + \sum_{i=1}^{m} \int_t^r \partial V_i(X_s)D_r^j X_s \circ dB_s^i$$

where $D_r^j X_t := (D_r X_t)^j$. This is just the integral form of (2.2) with a different initial condition, hence for $s < t$ we have

$$D_s^j X_t = J_{s,t}V_j(X_s) \tag{2.3}$$

Because $J_t = J_{s,t}J_s$ we get that $J_{s,t} = J_t J_s^{-1}$, where J_t^{-1} can be found by solving the SDE

$$dJ_t^{-1} = -J_t^{-1}\partial V_0(X_t)dt - \sum_{i=1}^{m} J_t^{-1}\partial V_i(X_t) \circ dB_t^i \tag{2.4}$$

As in the previous chapter, this can be checked by verifying that $J_t J_t^{-1} = J_t^{-1}J_t = I$ using Itô's formula.

From now on the following assumptions will be in place.

Assumption 2.5. *The vector fields V_i are C^∞ and all of their derivatives grow at most polynomially at infinity. Furthermore, they are such that the solutions to* (2.1)*,* (2.2) *and* (2.4) *satisfy*

$$\mathbb{E}\left(\sup_{t\le T}|X_t|^p\right) < \infty, \;\; \mathbb{E}\left(\sup_{t\le T}|J_t^{(k)}|^p\right) < \infty, \;\; \mathbb{E}\left(\sup_{t\le T}|J_t^{-1}|^p\right) < \infty$$

for every initial condition $x_0 \in \mathbb{R}^n$ and every $T, k, p > 0$

Remarks. (i) A sufficient condition for the above assumption to hold is that the V_i are bounded with bounded derivatives of all orders.

(ii) Under the above assumption, Proposition 1.21 tells us that the solution X_t of SDE (2.1) has $X_t^i \in \mathbb{D}^{1,\infty}$ for all t and all i, and provides us with an SDE for $D_s^j X_t$. The coefficients of this SDE again satisy the hypothesis of Proposition 1.21, and so $D_s^j X_t \in \mathbb{D}^{1,\infty}$, or equivalently $X_t \in \mathbb{D}^{2,\infty}$. Proceeding inductively we see that $X_t \in \mathbb{D}^{k,\infty}$ for all k, and so $X_t \in \mathbb{D}^{\infty}$.

Before moving on, we introduce the useful notion of 'almost truth'.

Definition 2.6. *Let $X = (X_\varepsilon)_{\varepsilon\in(0,1]}$, $Y = (Y_\varepsilon)_{\varepsilon\in(0,1]}$ be families of events depending on some parameter $\varepsilon > 0$.*

(i) X is said to be almost true *if for every $p > 0$ there exists a constant C_p such that $\mathbb{P}(X_\varepsilon) \geq 1 - C_p\varepsilon^p$ for all $\varepsilon \in (0, 1]$*

(ii) X is said to be almost false *if for every $p > 0$ there exists a constant C_p such that $\mathbb{P}(X_\varepsilon) \leq C_p\varepsilon^p$ for all $\varepsilon \in (0, 1]$*

(iii) We say X almost implies *Y and write $X \Rightarrow_\varepsilon Y$ if $X \setminus Y$ is almost false.*

Lemma 2.7 (Norris)**.** *Let B be an m-dimensional Brownian motion and let K and L be $\mathbb{R}$- and $\mathbb{R}^m$-valued adapted processes respectively such that, for $\alpha = \frac{1}{3}$, one has $\mathbb{E}(\|K\|_\alpha + \|L\|_\alpha)^p < \infty$ for every p. Let Z be the process defined by*

$$Z_t = Z_0 + \int_0^t K_s ds + \sum_{k=1}^m \int_0^t L_s^k dB_s^k \tag{2.5}$$

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

$$\{\|Z\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \{\|K\|_\infty < \varepsilon^r\} \text{ and } \{\|L\|_\infty < \varepsilon^r\}$$

Lemma 2.8. *Given a smooth vector field F on $\mathbb{R}^n$ and a unit vector $\eta \in \mathbb{R}^n$, define the process Z_F by*

$$Z_F(t) = \langle \eta, J_t^{-1} F(X_t)\rangle_{\mathbb{R}^n}$$

Then Z_F satisfies the SDE

$$dZ_F(t) = Z_{[F,V_0]}(t)dt + \sum_{k=1}^m Z_{[F,V_k]}(t) \circ dB_t^k \tag{2.6}$$

In addition, we have that

$$\langle \eta, \mathcal{C}_1 \eta\rangle = \sum_{k=1}^m \int_0^1 |Z_{V_k}(t)|^2 dt \geq \sum_{k=1}^m \left(\int_0^1 |Z_{V_k}(t)| dt\right)^2 \tag{2.7}$$

where $\mathcal{C}_1$ is the reduced Malliavin matrix at time 1.

Proof. (This statement of this lemma was given without proof) Write Q_t for J_t^{-1} and let $Q^i_j(t)$ be its component processes. By (2.4) these satisfy the following SDE[6]

$$
\begin{aligned}
dQ^i_j(t) &= -Q^i_k(t)(\partial V_0(X(t)))^k_j dt - Q^i_k(t)(\partial V_\alpha(X(t)))^k_j \circ dB^\alpha(t) \\
&= -Q^i_k(t)\partial_j V^k_0(X(t))dt - Q^i_k(t)\partial_j V^k_\alpha(X(t)) \circ dB^\alpha(t)
\end{aligned}
$$

We also have

$$
dX^i(t) = V^i_0(X(t))dt + V^i_\alpha(X(t)) \circ dB^\alpha(t)
$$

Now

$$
\begin{aligned}
Z_F(t) &= \langle \eta, J_t^{-1} F(X_t)\rangle_{\mathbb{R}^n} \\
&= \sum_{i=1}^n \eta^i (Q(t)F(X(t))^i \\
&= \sum_{i=1}^n \eta^i Q^i_j(t) F^j(X(t))
\end{aligned}
$$

Applying Itô's formula gives, for each $i = 1, \dots, n$,

$$
\begin{aligned}
d(Q^i_j(t)F^j(X(t))) &= Q^i_j(t) \circ dF^j(X(t)) + F^j(X(t)) \circ dQ^i_j(t) \\
&= Q^i_j(t)\partial_k F^j(X(t)) \circ dX^k(t) \\
&\quad + F^j(X(t))(-Q^i_k(t)\partial_j V^k_0(X(t))dt - Q^i_k(t)\partial_j V^k_\alpha(X(t)) \circ dB^\alpha(t)) \\
&= Q^i_j(t)\partial_k F^j(X(t))V^k_0(X(t))dt + Q^i_j(t)\partial_k F^j(X(t))V^k_\alpha(X(t)) \circ dB^\alpha(t) \\
&\quad - Q^i_k(t)F^j(X(t))\partial_j V^k_0(X(t))dt - Q^i_k(t)F^j(X(t))\partial_j V^k_\alpha(X(t)) \circ dB^\alpha(t)) \\
&= \{Q^i_j(t)\partial_k F^j(X(t))V^k_0(X(t)) - Q^i_k(t)F^j(X(t))\partial_j V^k_0(X(t))\}dt \\
&\quad + \{Q^i_j(t)\partial_k F^j(X(t))V^k_\alpha(X(t)) - Q^i_k(t)F^j(X(t))\partial_j V^k_\alpha(X(t))\} \circ dB^\alpha(t)
\end{aligned}
$$

But, for each $\alpha = 0, \dots, m$,

$$
\begin{aligned}
Z_{[F,V_\alpha]}(t) &= \langle \eta, Q(t)(\partial F \cdot V_\alpha - \partial V_\alpha \cdot F)(X(t))\rangle_{\mathbb{R}^n} \\
&= \sum_{i=1}^n \eta^i Q^i_j(t)((\partial F)^j_k V^k_\alpha - (\partial V_\alpha)^j_k F^k)(X(t)) \\
&= \sum_{i=1}^n \eta^i (Q^i_j(t)\partial_k F^j(X(t))V^k_\alpha(X(t)) - Q^i_k(t)F^j(X(t))\partial_j V^k_\alpha(X(t)))
\end{aligned}
$$

Combining the above gives that Z_F satisfies (2.6).

[6]Einstein's summation convention is in use, and we are using that Roman indices are to be summed from 1 to n and Greek indices are to be summed to 0 to m

Now we need to prove (2.7). The inequality follows straight from Jensen's inequality after noting that the interval $[0,1]$ equipped with the Lebesgue measure is a probability space. To prove the equality, just note that both sides are equal to

$$\sum_{\alpha=1}^{m}\sum_{i,l=1}^{n}\int_0^1 \eta^i Q^i_j(s)V^j_\alpha(X_s)V^k_\alpha(X_s)Q^l_k(s)\eta^l ds$$

$\square$

Remark. With a little more work we obtain the Itô form of (2.6):

$$dZ_F(t) = \left(Z_{[F,V_0]}(t) + \frac{1}{2}\sum_{k=1}^{m} Z_{[[F,V_k],V_k]}(t)\right)dt + \sum_{k=1}^{m} Z_{[F,V_k]}(t)dB^k_t \tag{2.8}$$

Theorem 2.9. *Consider the SDE (2.1) and assume that Assumption 2.5 holds. If (2.1) satisfies the parabolic Hörmander condition, then for every initial condition $X_0 \in \mathbb{R}^n$ we have the bound*

$$\sup_{\|\eta\|=1} \mathbb{P}(\langle \eta, \mathcal{C}_1\eta\rangle < \varepsilon) \leq C_p\varepsilon^p$$

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

Proof. Fix an initial condition $X_0 \in \mathbb{R}^n$ and a unit vector $\eta \in \mathbb{R}^n$. The result is true if

$$\{\langle \eta, \mathcal{C}_1\eta\rangle < \varepsilon\} \Rightarrow_\varepsilon \emptyset \tag{2.9}$$

i.e. that the statement $\langle \eta, \mathcal{C}_1\eta\rangle < \varepsilon$ is 'almost false'. Given a smooth vector field F on $\mathbb{R}^n$, define Z_F as in the previous lemma. Then the conclusion of the lemma holds. We want to use Norris' lemma on Z_F as given by (2.8), so we need that the coefficients satisfy the hypothesis. Let $G = [F,V_0] + \frac{1}{2}\sum_{k=1}^{m}[[F,V_k],V_k]$, then we need to show that $\mathbb{E}\|Z_{[F,V_k]}\|^p_{1/3} < \infty$ and $\mathbb{E}\|Z_G\|^p_{1/3} < \infty$. This is true provided F grows at most polynomially fast, see [5]. Therefore for each $k = 1,\ldots,m$, we have the almost implication

$$\{\|Z_F\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \{\|Z_{[F,V_k]}\|_\infty < \varepsilon^r\} \text{ and } \{\|Z_G\|_\infty < \varepsilon^r\}$$

Applying Norris' lemma to Z_G tells us that for each $k,l = 0,\ldots,m$

$$\{\|Z_G\|_\infty < \varepsilon^r\} \Rightarrow_\varepsilon \{\|Z_{[[F,V_k],V_l]}\|_\infty < (\varepsilon^r)^r\}$$

and so

$$\{\|Z_F\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \{\|Z_{[[F,V_k],V_l]}\|_\infty < \varepsilon^{r^2}\}$$

Therefore we have for each $k = 0,\ldots,m$

$$\{\|Z_F\|_\infty < \varepsilon\} \Rightarrow_\varepsilon \{\|Z_{[F,V_k]}\|_\infty < \varepsilon^{r^2}\} \tag{2.10}$$

Now by (2.7) it can be shown that

$$\{\langle \eta, \mathcal{C}_1\eta\rangle < \varepsilon\} \Rightarrow_\varepsilon \{\|Z_{V_k}\| < \varepsilon^{1/5}\}$$

Combining this with (2.10) we get that for each $k = 1, \dots, m$, $l = 0, \dots, m$

$$\{\langle \eta, \mathcal{C}_1 \eta \rangle < \varepsilon\} \Rightarrow_\varepsilon \{\|Z_{[V_k, V_l]}\|_\infty < \varepsilon^{1/25}\}$$

Iterating this gives that for every $k > 0$ there exists a $q_k > 0$ such that

$$\{\langle \eta, \mathcal{C}_1 \eta \rangle < \varepsilon\} \Rightarrow_\varepsilon \bigcap_{V \in \mathcal{V}_k} \{\|Z_V\|_\infty < \varepsilon^{q_k}\}$$

where $\mathcal{V}_k$ are the collections of Hörmander vector fields defined earlier. By our assumptions on these collections, for every $x \in \mathbb{R}^n$ there exists a $k > 0$ such that $\mathcal{V}_k(x) = \mathbb{R}^n$. Since $Z_V(0) = \langle \eta, V(X_0) \rangle$ it follows that the intersection above is empty for large enough k, and so the result is proved. □

We now just need one final lemma, whose proof is found in [5].

Lemma 2.10. *Let M be a symmetric positive definite $n \times n$ matrix-valued random variable such that $\mathbb{E}\|M\|^p < \infty$ for every $p \geq 1$ and such that, for every $p \geq 1$ there exists a C_p such that*

$$\sup_{\|\eta\|=1} \mathbb{P}(\langle \eta, M\eta \rangle < \varepsilon) \leq C_p \varepsilon^p \tag{2.11}$$

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

Combining the above two results, we get that the inverse of the reduced Malliavin matrix (and hence the inverse of the Malliavin matrix itself) has bounded moments of all orders, provided that Assumption 2.5 and the parabolic Hörmander condition are satisfied. We are hence finally ready to prove a version of Hörmander's theorem:

Theorem 2.11 (Hörmander)**.** *Let $X_0 \in \mathbb{R}^n$ and let X_t be a solution to the SDE (2.1). If (2.1) satisfies the parabolic Hörmander condition and Assumption 2.5 is satisfied, then the law of X_t has a smooth density with respect to the Lebesgue measure.*

Proof. Recall (1.10) from the previous chapter: the Malliavin matrix of X_t is given by

$$\begin{aligned}\mathcal{M}_t &= \int_0^t J_t J_s^{-1} V(X_s) V^*(X_s) (J_s^{-1})^* J_t^* ds \\ &= \int_0^t J_{s,t} V(X_s) V^*(X_s) J_{s,t}^* ds\end{aligned}$$

where $V(X_s)$ is the $n \times m$ matrix obtained by concatenating the m vectors $V_i(X_s)$, $i = 1, \dots, m$. By the (2.3) coupled with the remark following Assumption 2.5, we have that $X_t \in \mathbb{D}^\infty$ for all t. From the above we know that $\mathbb{E}\|\mathcal{M}_t^{-1}\|^p < \infty$ for all $p > 1$, and so by the remark following Theorem 2.3 the theorem is proved. □

3 Law of the Ornstein-Uhlenbeck process

The aim of this chapter is to use the tools of Malliavin calculus to analyse the law of the Ornstein-Uhlenbeck (OU) process. Let $i : H \to \Omega$ be classical Wiener space with the Wiener measure $\mathbb{P}$, so that $H = L_0^{2,1}([0,1])$ and $\Omega = C_0([0,1])$. Take a Brownian motion $B : [0,1] \times \Omega \to \mathbb{R}$ on Ω, so that $B_t(\omega) = \omega(t)$ for all $\omega \in \Omega$, $t \geq 0$. The OU process $X_t : \Omega \to \mathbb{R}$ is the solution to the SDE

$$dX_t = dB_t - X_t dt \tag{3.1}$$

which is given by

$$X_t = e^{-t} X_0 + \int_0^t e^{s-t} dB_s \tag{3.2}$$

where we assume that X_0 is constant.

Before we proceed we first review some basic facts about Gaussian measures - a more detailed treatment can be found in Appendix A.1. Recall the definition of a Gaussian measure:

Definition 3.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let λ be the Lebesgue measure on $\mathbb{R}$. A (centred) Gaussian measure on $\mathbb{R}$ is a Borel measure γ on $\mathbb{R}$ such that either there exists $\alpha > 0$ such that*

$$\gamma(A) = (2\pi\alpha)^{-1/2} \int_A e^{-|x|^2/2\alpha} d\lambda(x)$$

or $\gamma = \delta_0$, the degenerate case.

If E is a separable Banach space, a Borel measure γ on E is a (centred) Gaussian measure on E if $\ell_* \gamma$ is a (centred) Gaussian measure on $\mathbb{R}$ for all $\ell \in E^*$

Example 3.2. The Wiener measure is a Gaussian measure. The law of the OU process defined above is a Gaussian measure as it is the pushforward of the Wiener measure by the linear map $X_\cdot$.

To each Gaussian measure μ on a separable Banach space E we can associate a Hilbert subspace H_μ, called the *Cameron-Martin space* or *reproducing kernel Hilbert space*. This space comprises of the directions in which we can translate μ whilst preserving null sets. It can also be thought of the directions in which we can differentiate functions on Ω.

Definition 3.3. *Let μ be a Gaussian measure on a separable Banach space E. We define the* covariance map $R : E^* \to E$ *by*

$$R(\ell) := \int_E \ell(x) x d\mu(x)$$

where the integral is in the Bochner sense. The reproducing kernel Hilbert space (RKHS) H_μ *of μ is the completion of the image of the covariance map with respect to the inner product*

$$\langle R(\ell_1), R(\ell_2) \rangle_\mu := \int_E \ell_1(x) \ell_2(x) d\mu(x)$$

Remark. It can be checked that H_μ is still a subset of E after the completion procedure. The inclusion $i : H_\mu \to E$ gives an abstract Wiener space, see Appendix A.1.

We look at how to construct the RKHS for the law of a real-valued Gaussian process. Let $\xi = (\xi_t)_{t\in[0,1]}$ be a centred real-valued sample continuous Gaussian process and let $\gamma(s,t) = \mathbb{E}\xi_s\xi_t$ be its covariance function. Let μ be the law of ξ, which is a Gaussian measure on $E = C([0,1])$. We have that

$$(R\ell)(t) = \int_E \ell(x)x(t)d\mu(x)$$

We identify E^* with the space of signed measures on $[0,1]$ by the relation

$$\ell(x) = \int_0^1 x(s)d\alpha_\ell(s)$$

for some signed measure α_ℓ. By the Riesz representation theorem for linear functionals on $C_c([0,1]) = C([0,1])$, there is a unique function of bounded variation f such that $\alpha_\ell([s,t]) = f(t) - f(s)$ for all intervals $[s,t]$. The integral can hence be interpreted as a Lebesgue-Stieltjes integral, and by abuse of notation we write

$$\ell(x) = \int_0^1 x(s)d\ell(s)$$

since the function f above is unique. We then have the integration by parts formula for Lebesgue-Stieltjes integrals[7]:

$$\int_0^1 x(s)d\ell(x) = x(1)\ell(1) - x(0)\ell(0) - \int_0^1 \ell(s)dx(s)$$

Now using this representation along with Fubini's theorem, we have

$$\begin{aligned}(R\ell)(t) &= \int_E \left(\int_0^1 x(s)d\ell(s)\right) x(t)d\mu(x) \\ &= \int_0^1 \left(\int_E x(s)x(t)d\mu(x)\right) d\ell(s) \\ &= \int_0^1 \gamma(s,t)d\ell(s) \qquad (3.3)\end{aligned}$$

Assuming that γ is almost everywhere differentiable, we can use integration by parts to get this in terms of a Lebesgue integral:

$$\begin{aligned}(R\ell)(t) &= \gamma(1,t)\ell(1) - \gamma(0,t)\ell(0) - \int_0^1 \ell(s)d\gamma(s,t) \\ &= \gamma(1,t)\ell(1) - \gamma(0,t)\ell(0) - \int_0^1 \ell(s)\frac{\partial\gamma}{\partial s}(s,t)ds \qquad (3.4)\end{aligned}$$

[7]When we write $\ell(s)$ for $\ell \in E^*$, $s \in \mathbb{R}$, we mean $f(s)$ where f is the corresponding unique bounded variation function defined as above

We use (3.3) to get an expression for the inner product on H_μ. Let $\ell_1, \ell_2 \in E^*$, then

$$\begin{aligned}
\langle R\ell_1, R\ell_2\rangle_\mu &:= \int_E \ell_1(x)\ell_2(x)d\mu(x) \\
&= \ell_2\left(\int_E \ell_1(x)xd\mu(x)\right) \quad \text{(since } \ell_2 \text{ is bounded and linear)} \\
&= \ell_2\left(\int_0^1 \gamma(s,\cdot)d\ell_1(s)\right) \\
&= \int_0^1\int_0^1 \gamma(s,t)d\ell_1(s)d\ell_2(s) \qquad (3.5)
\end{aligned}$$

Again if γ is differentiable almost everywhere we can use integration by parts to get this in terms of Lebesgue integrals.

Example 3.4. We check that in the case where $(\xi_t)_{t\in[0,1]}$ is the Wiener process, we recover the classical Cameron-Martin space $L_0^{2,1}([0,1])$, where

$$L_0^{2,1}([0,1]) = \left\{\sigma : [0,T] \to \mathbb{R}^m \;\middle|\; \exists\varphi \in L^2([0,T];\mathbb{R}^m) \text{ such that } \sigma(t) = \int_0^t \varphi(s)ds\right\}$$

with inner product

$$\langle g, h\rangle_{L_0^{2,1}} = \int_0^1 \dot{g}(t)\dot{h}(t)dt$$

Let $\ell \in E^*$, and recall that the covariance of ξ is $\gamma(s,t) = s\wedge t$. Then we have

$$\gamma(1,t) = t, \quad \gamma(0,t) = 0, \quad \frac{\partial\gamma}{\partial s}(s,t) = \begin{cases} 1 & s < t \\ 0 & s > t \end{cases}$$

and so by (3.4) we have

$$\begin{aligned}
(R\ell)(t) &= t\ell(1) - \int_0^t \ell(s)ds \\
&= \int_0^t (\ell(1) - \ell(s))ds \\
&= \int_0^t \ell([s,1])ds
\end{aligned}$$

This tells us that $R\ell$ is differentiable with $(R\ell)'(t) = \ell([t,1])$. We see that the inner product is given by

$$
\begin{aligned}
\langle R\ell_1, R\ell_2\rangle_\mu &= \int_0^1 \int_0^1 \gamma(s,t) d\ell_1(s) d\ell_2(t) \\
&= \int_0^1 \int_0^1 (s \wedge t) d\ell_1(s) d\ell_2(t) \\
&= \int_0^1 \int_0^t \ell_1([s,1]) ds d\ell_2(s) \\
&= \int_0^1 \int_0^1 \ell_1([s,1]) \mathbf{1}_{[0,t]}(s) d\ell_2(t) ds \\
&= \int_0^1 \ell_1([s,1]) \left(\int_s^1 d\ell_2(t) \right) ds \\
&= \int_0^1 \ell_1([s,1]) \ell_2([s,1]) ds \\
&= \int_0^1 (R\ell_1)'(t)(R\ell_2)'(t) dt
\end{aligned}
$$

That is, for h, g in the image of R, we have

$$
\langle h, g\rangle_\mu = \int_0^1 \dot{h}(t)\dot{g}(t) dt
$$

This is reassuring! We now just need to find the completion of the image of R with respect to this inner product. Let $(h_n)_{n\in\mathbb{N}} \subseteq R(E^*)$ be a Cauchy sequence, so

$$
\int_0^1 (\dot{h}_n(t) - \dot{h}_m(t))^2 dt \to 0 \text{ as } m, n \to \infty
$$

Now each $\dot{h}_n(t) = \ell_n([t,1])$ for some $\ell_n \in E^*$, and so $(\ell_n([\cdot,1]))_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2([0,1])$. By completeness it follows that it has a limit in $L^2([0,1])$, call it φ. Define h by

$$
h(t) = \int_0^t \varphi(s) dt
$$

Then clearly $h \in L_0^{2,1}$, and $h_n \to h$ with respect to $\langle\cdot,\cdot\rangle_\mu$. Since $R(E^*) \subseteq L_0^{2,1}$ and $L_0^{2,1}$ is complete with respect to $\langle\cdot,\cdot\rangle_\mu$, it follows that the RKHS of μ is a closed subspace of $L_0^{2,1}$. Checking that the space $(\ell([\cdot,1]))_{\ell\in E^*}$ is dense in L^2 gives the result. Since by $\ell_n([t,1])$ we actually mean $\alpha_{\ell_n}([t,1]) := f_n(1) - f_n(t)$ for a function of bounded variation f_n, and each function of bounded variation can defines a measure in this way, it's then clear that $(\ell([\cdot,1]))_{\ell\in E^*} \supseteq C_c^\infty([0,1])$, the space of smooth functions with compact support on $[0,1]$. Since these are dense in $L^2([0,1])$, the result follows.

We now turn to the law μ of the Ornstein-Uhlenbeck process X, so $\mu = (X.)_*(\mathbb{P})$. This Gaussian measure is not centred, but (3.3), (3.4) and(3.5) still hold with γ this time defined by

$$
\gamma(s,t) = \mathbb{E}(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)
$$

where $\mathbb{E}$ is used to denote the integral over E with respect to μ. By standard results on stochastic integrals, for $s < t$

$$\mathbb{E}X_t = e^{-t}X_0, \quad \gamma(s,t) = \frac{1}{2}e^{-(t+s)}(e^{2(s\wedge t)} - 1) = e^{-t}\sinh(s)$$

We look for the RKHS of μ. This time we have

$$\gamma(1,t) = e^{-1}\sinh t, \quad \gamma(0,t) = 0, \quad \frac{\partial\gamma}{\partial s}(s,t) = \begin{cases} e^{-t}\cosh(s) & s < t \\ -e^{-s}\sinh(t) & s > t \end{cases}$$

and so by (3.4), for $\ell \in E^*$ we have

$$(R\ell)(t) = \ell(1)e^{-1}\sinh(t) - \int_0^t \ell(s)e^{-t}\cosh(s)ds + \int_t^1 \ell(s)e^{-s}\sinh(t)ds$$

This is differentiable with

$$\begin{aligned}(R\ell)'(t) &= \ell(1)e^{-1}\cosh(t) + \int_0^t \ell(s)e^{-t}\cosh(s)ds - \ell(t)e^{-t}\cosh(t) \\ &\quad + \int_t^1 \ell(s)e^{-s}\cosh(t)ds - \ell(t)e^{-t}\sinh(t) \\ &= \ell(1)e^{-1}\cosh(t) + \int_0^t \ell(s)e^{-t}\cosh(s)ds + \int_t^1 \ell(s)e^{-s}\cosh(t)ds - \ell(t)\end{aligned}$$

since $\sinh(t) + \cosh(t) = e^t$, and so we have

$$\begin{aligned}(R\ell)(t) + (R\ell)'(t) &= \ell(1)e^{t-1} - \ell(t) + \int_t^1 \ell(s)e^{t-s}ds \\ &= \ell(s)e^{t-s}\Big|_{s=t}^{1} + \int_t^1 \ell(s)e^{t-s}ds\end{aligned}$$

Computing the inner product should[8] yield, for $\ell_1, \ell_2 \in E^*$,

$$\langle R\ell_1, R\ell_2\rangle_\mu = \int_0^1 ((R\ell_1)(t) + (R\ell_1)'(t))((R\ell_2)(t) + (R\ell_2)'(t))dt$$

and so for g, h in the range of R we have

$$\begin{aligned}\langle g, h\rangle_\mu &= \int_0^1 (g(t) + \dot{g}(t))(h(t) + \dot{h}(t))dt \\ &= \langle g + \dot{g}, h + \dot{h}\rangle_{L^2}\end{aligned}$$

This will only define an inner product on a space of functions that vanish at 0, but from the expression for $(R\ell)(t)$ above we see that $(R\ell)(0) = 0$ for all $\ell \in E^*$. We claim that the

[8] Work in progress!

completion of $R(E^*)$ with respect to this inner product is $L_0^{2,1}([0,1])$.

First we show that $R(E^*) \subseteq L_0^{2,1}$. Let $h \in R(E^*)$ so that $h = R\ell$ for some $\ell \in E^*$. Then we have

$$\int_0^1 (h(t) + \dot{h}(t))^2 dt = \int_0^1 h(t)^2 dt + \int_0^1 \dot{h}(t)^2 dt + 2\int_0^1 h(s)\dot{h}(s) ds < \infty$$

Integrating by parts we see that the rightmost term is finite, and so we have that $h, \dot{h} \in L^2$. It follows that $h \in L_0^{2,1}$. A Cauchy sequence $(h_n)_{n\in\mathbb{N}} \subseteq R(E^*)$ therefore has a limit $h \in L_0^{2,1}$ and so we see that the completion of $R(E^*)$ is a closed subspace of $L_0^{2,1}$. Using the expression for $(R\ell)(t) + (R\ell)'(t)$ above we can see that $\{(R\ell) + (R\ell)' \mid \ell \in E^*\} \supseteq C_c^\infty([0,1])$, and so $R(E^*)$ is dense in $L_0^{2,1}$ with respect to the inner product $\langle \cdot, \cdot \rangle_\mu$. It follows that the RKHS of μ is indeed $L_0^{2,1}$.

We now consider the problem of how to compute the divergence of elements of the RKHS with respect to μ, i.e. we want to find an integration by parts formula for the measure μ. Define the operator $d : \mathbb{D}^{1,2} \to H^*$ by

$$\begin{aligned}(df)_\sigma(h) &:= \left\langle \int_0^\cdot (Df)(t)dt, h \right\rangle_H (\sigma) \\ &= \langle Df, \dot{h} \rangle_{L^2}(\sigma)\end{aligned}$$

where $f \in \mathbb{D}^{1,2}$, $h \in H$ and $\sigma \in \Omega$. Then a divergence with respect to μ, div_μ, will satisfy

$$\int_\Omega (df)_\sigma(h) d\mu(\sigma) = \int_\Omega f(\sigma)(\mathrm{div}_\mu)(h) d\mu(\sigma)$$

for all $f \in \mathbb{D}^{1,2}$ and $h \in H$. Now recall the pushforward lemma:

Lemma 3.5 (Pushforward lemma)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space, $(Y, \mathcal{B})$ a measurable space, $\theta : X \to Y$ a measurable map and $f : Y \to \mathbb{R}$ measurable. Then*

$$\int_X f \circ \theta d\mu = \int_Y f d(\theta_* \mu)$$

in the sense that if one side exists so does the other, and there is equality.

$(df)_\cdot(h)$ is a real-valued function on Ω, so we can apply this lemma to get

$$\int_\Omega (df)_\sigma(h) d\mu(\sigma) = \int_\Omega (df)_{X_\cdot(\omega)}(h) d\mathbb{P}(\omega)$$

If we can get $(df)_{X_\cdot(\omega)}(h)$ into the form $(d\widetilde{f})_\omega(g)$ for some $\widetilde{f} : \Omega \to \mathbb{R}$, $g = g^h \in H$ then we can use the integration by parts formula for the Wiener measure on the right hand side above. By the chain rule for the Malliavin derivative,

$$(d(f \circ X_\cdot))_\omega(g) = (df)_{X_\cdot(\omega)}((dX_\cdot)_\omega(g))$$

so we can set $\widetilde{f} = f \circ X_{\cdot}$ and $h_t = (dX_t)_\omega(g)$. Let us calculate h. We have that

$$\begin{aligned}(DX_t)(s) &= D(W(e^{\cdot - t}\mathbf{1}_{[0,t]}(\cdot)))(s) \\ &= e^{s-t}\mathbf{1}_{[0,t]}(s)\end{aligned}$$

Thus for $g \in H$ we have

$$\begin{aligned}(dX_t)_\omega(g) &= \langle DX_t, \dot{g}\rangle_{L^2} \\ &= \int_0^t e^{s-t}\dot{g}_s ds\end{aligned}$$

We want this to equal h_t, so we have

$$e^t h_t = \int_0^t e^s \dot{g}_s ds$$

and therefore

$$\dot{g}_t = h_t + \dot{h}_t$$

Putting this together, we have

$$\begin{aligned}\int_\Omega (df)_\sigma(h)d\mu(\sigma) &= \int_\Omega (d(f \circ X_{\cdot}))_\omega(g)d\mathbb{P}(\omega) \\ &= \int_\Omega f(X_{\cdot}(\omega))(\mathrm{div}_{\mathbb{P}} g)(\omega)d\mathbb{P}(\omega) \\ &= \int_\Omega f(\sigma)\mathbb{E}((\mathrm{div}_\mu h)(\sigma)|X_{\cdot}(\omega) = \sigma)d\mu(\sigma)\end{aligned}$$

which tells us that (almost surely)

$$\begin{aligned}\mathbb{E}((\mathrm{div}_\mu h)(\sigma)|X_{\cdot}(\omega) = \sigma) &= (\mathrm{div}_{\mathbb{P}} g)(\omega) \\ &= \int_0^1 (h_t + \dot{h}_t)dB_t(\omega)\end{aligned}$$

Now using that $dB_t = dX_t + X_t dt$,

$$\int_0^1 (h_t + \dot{h}_t)dB_t = \int_0^1 (h_t + \dot{h}_t)dX_t + \int_0^1 (h_t + \dot{h}_t)X_t dt$$

and so we claim that the divergence is given by

$$(\mathrm{div}_\mu h)(\sigma) = \int_0^1 (h_t + \dot{h}_t)(d\sigma_t + \sigma_t dt)$$

Indeed we have, by the pushforward lemma again,

$$\begin{aligned}\int_\Omega (df)_\sigma(h)d\mu(\sigma) &= \int_\Omega f(X_{\cdot}(\omega))\left(\int_0^1 (h_t + \dot{h}_t)dX_t(\omega) + \int_0^1 (h_t + \dot{h}_t)X_t(\omega)dt\right)d\mathbb{P}(\omega) \\ &= \int_\Omega f(\sigma)\left(\int_0^1 (h_t + \dot{h}_t)d\sigma_t + \int_0^1 (h_t + \dot{h}_t)\sigma_t dt\right)d\mu(\sigma) \\ &= \int_\Omega f(\sigma)\left(\int_0^1 (h_t + \dot{h}_t)(d\sigma_t + \sigma_t dt)\right)d\mu(\sigma)\end{aligned}$$

This divergence operator gives a martingale with respect to μ:

Proposition 3.6. *Let $h \in H$ and define the process $M = (M_t)_{t\in[0,1]}$ by*

$$M_t(\sigma) = \int_0^t (h_r + \dot{h}_r)(d\sigma(r) + \sigma(s)dr)$$

Then M is an martingale with respect to μ.

Proof. Let $\mathbb{E}^\mu$ denote expectation with respect to μ and let $(\mathcal{F}_t)_{t\in[0,1]}$ be the filtration generated by B. Let $f : \Omega \to \mathbb{R}$ be a bounded $\mathcal{F}_s$-measurable function with $s < t$, then

$$\begin{aligned}
\mathbb{E}^\mu(M_t f) &= \mathbb{E}^\mu(M_s f) + \mathbb{E}^\mu((M_t - M_s)f) \\
&= \mathbb{E}^\mu(M_s f) + \mathbb{E}^\mu(\mathbb{E}^\mu(M_t - M_s|\mathcal{F}_s)f) \\
&= \mathbb{E}^\mu(M_s f) + \mathbb{E}^\mu\left(\mathbb{E}\left(\int_s^t (h_r + \dot{h}_s)dB_r\middle|\mathcal{F}_s\right) f\right) \\
&= \mathbb{E}^\mu(M_s f)
\end{aligned}$$

by the martingale property of Itô integrals. Thus we have that M is a martingale with respect to μ. $\square$

Now we have $X. : (H, \langle\cdot,\cdot\rangle_{L_0^{2,1}}) \to (H, \langle\cdot,\cdot\rangle_\mu)$ where $\langle\cdot,\cdot\rangle_\mu$ is the inner product on the RKHS for μ. We can also equip H with the quotient inner product $\langle\cdot,\cdot\rangle_q$, defined by

$$\langle h_1, h_2\rangle_q := \langle (dX.)^{-1}h_1, (dX.)^{-1}h_2\rangle_{L_0^{2,1}}$$

From earlier we have that for all $\omega \in \Omega$, $h = (dX.)_\omega(g) \iff \dot{g} = h + \dot{h}$, and so $\dot{\overbrace{(dX.)^{-1}h}} = h + \dot{h}$. The quotient inner product is therefore given by

$$\langle h_1, h_2\rangle_q = \langle h_1 + \dot{h}_1, h_2 + \dot{h}_2\rangle_{L^2}$$

and so $\langle\cdot,\cdot\rangle_q = \langle\cdot,\cdot\rangle_\mu$.

3.1 Existence and smoothness of density

We now consider the two dimensional system

$$\begin{cases} dX_t = V_t dt \\ dV_t = dB_t - V_t dt \end{cases}$$

X_t and V_t describe the position and velocity respectively of a massive Brownian particle. We can use Theorem 2.1 to see if the law of the random vector (X_t, V_t) has a density. The solution for V is

$$V_t = e^{-t}X_0 + \int_0^t e^{s-t}dB_t$$

and we saw earlier that $(DV_t)(s) = e^{s-t}\mathbf{1}_{[0,t]}(s)$. Assuming that we can interchange the Malliavin derivative and Lebesgue integral (which is reasonable due to linearity and closedness of D), we have

$$\begin{aligned}(DX_t)(s) &= \int_0^t (DV_r)(s)dr \\ &= \int_s^t e^{s-r}dr \\ &= 1 - e^{s-t}\end{aligned}$$

We can now calculate the Malliavin matrix:

$$\begin{aligned}\langle DX_t, DX_t\rangle_H &= \int_0^1 (1-e^{s-t})^2 ds \\ &= 1 + \frac{1}{2}e^{-2t}(e^2-1) - 2e^{-t}(e-1) \\ \langle DX_t, DV_t\rangle_H &= \int_0^t (e^{s-t} - e^{2s-2t})ds \\ &= \frac{1}{2}e^{-2t}(1-e^t)^2 \\ \langle DV_t, DV_t\rangle_H &= \int_0^t e^{2s-2t}ds \\ &= \frac{1}{2}(1-e^{-2t})\end{aligned}$$

Note that the Malliavin matrix isn't random in this case. Calculating the determinant, we see that it doesn't vanish for any $t > 0$ and so the law of (X_t, V_t) admits a density with respect to the Lebesgue measure for all $t > 0$. In fact we can use Theorem 2.3 to deduce that this density is smooth: it follows from the above that all inverse moments of the determinant exist, and X_t, Y_t can be seen to be in $\mathbb{D}^\infty$. Thus the vector (X_t, Y_t) is non-degenerate for all $t > 0$ and therefore has a smooth density.

What if we didn't have an explicit solution for the process (X_t, V_t)? We couldn't use the analysis above, but we could instead use Hörmander's theorem. For this we need to get the system into Stratonovich form, but since

$$\int_0^t 1dB_s = \int_0^t 1 \circ dB_s$$

we don't need to alter the finite variation term. In the notation of the section on Hörmander's theorem we have $n = 2$ and $m = 1$, and the system can be written as

$$d(X_t, V_t) = V_0(X_t, V_t)dt + V_1(X_t, V_t) \circ dB_t$$

where $V_0(x, y) = (0, 1)$ and $V_1(x, y) = (y, -y)$. Clearly these are smooth, so we check the parabolic Hörmander condition. We have

$$\mathcal{V}_0 = \{V_1\}, \quad \mathcal{V}_1 = \{V_1, [V_0, V_1]\}$$

We compute the Lie bracket above:

$$\begin{aligned}[V_0, V_1] &= \partial V_0 \cdot V_1 - \partial V_1 \cdot V_0 \\ &= -\begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 \\ 1 \end{pmatrix}\end{aligned}$$

It follows that $\mathcal{V}_1(x) = \mathbb{R}^2$ for all $x \in \mathbb{R}^2$ and so Hörmander's condition is satisfied. We therefore see (again) that the process (X_t, Y_t) admits a smooth density with respect to the Lebesgue measure.

A Appendices

A.1 Gaussian measure theory

Since we are concerned with the calculus of Gaussian random variables and processes, we will need to know some of their properties. We will first briefly review standard results about Gaussian measures, abstract Wiener spaces and Paley-Wiener integrals, before looking at isonormal Gaussian processes and white noise, which will play an important role in Malliavin calculus. Where not stated otherwise, the content of the following subsection is drawn from [2] and [4].

A.1.1 Gaussian measures in infinite dimensions

Definition A.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let λ be the Lebesgue measure on $\mathbb{R}$. A (centred) Gaussian measure on $\mathbb{R}$ is a Borel measure γ on $\mathbb{R}$ such that either there exists $\alpha > 0$ such that*

$$\gamma(A) = (2\pi\alpha)^{-1/2} \int_A e^{-|x|^2/2\alpha} d\lambda(x)$$

or $\gamma = \delta_0$, the degenerate case.

If E is a separable Banach space, a Borel measure γ on E is a (centred) Gaussian measure on E if $\ell_*\gamma$ is a (centred) Gaussian measure on $\mathbb{R}$ for all $\ell \in E^*$

Example A.2. The standard Gaussian measure γ^n on $\mathbb{R}^n$. If λ^n is the Lebesgue measure on $\mathbb{R}^n$, then

$$\gamma^n(A) = (2\pi)^{-n/2} \int_A e^{-\|x\|^2/2} d\lambda^n(x)$$

Example A.3. We extend the above example to general finite dimensional inner product spaces. Let $(V, \langle \cdot, \cdot \rangle)$ be an n-dimensional inner product space, and let λ^n be the Lebesgue measure on $\mathbb{R}^n$. Then the inner product $\langle \cdot, \cdot \rangle$ can be used to define a 'Lebesgue measure' on V. Let $u : \mathbb{R}^n \to V$ be an isometry so that $\langle u(x), u(y) \rangle = \langle u, v \rangle_{\mathbb{R}^n}$, and $u(x) = x_1 e_1 + \ldots + x_n e_n$ for some orthonormal basis $e_1, \ldots, e_n$ of V. Define

$\lambda^{\langle\cdot,\cdot\rangle} := u_*(\lambda^n)$. It can be checked that this definition does not depend on the choice of isometry u. We use this 'Lebesgue measure' on V to define the standard Gaussian measure $\gamma^{\langle\cdot,\cdot\rangle}$ on V: for each Borel subset A of V set

$$\gamma^{\langle\cdot,\cdot\rangle}(A) := (2\pi)^{-n/2} \int_A e^{-\langle x,x\rangle/2} d\lambda^{\langle\cdot,\cdot\rangle}(x)$$

Definition A.4. *Let E be a separable Banach space and let*

$$\mathcal{A}(E) = \{T \in \mathbb{L}(E;F) \mid \dim F < \infty,\ T \text{ onto}\}$$

Write F_T for F if $T \in \mathcal{A}(E)$ and $T \in \mathbb{L}(E;F)$. A cylinder set measure *(CSM) on E is a family $(\mu_T)_{T\in\mathcal{A}(E)}$ of probability measures μ_T on F_T satisfying the consistency relation: if we have*

$$\begin{tikzcd} E \arrow[r, "T"] \arrow[dr, "S"'] & F_T \arrow[d, "\pi_{ST}"] \\ & F_S \end{tikzcd}$$

then $\mu_S = (\pi_{ST})_*(\mu_T)$.

If there exists a probability measure μ on E such that $\mu_T = T_*\mu$ on F_T for each $T \in \mathcal{A}(E)$, we say that $(\mu_T)_{T\in\mathcal{A}(E)}$ 'is' a measure. To see why this is reasonable, suppose μ is a probability measure on E and define $\mu_T = T_*(\mu)$ for each $T \in \mathcal{A}(E)$. Then if $\pi_{ST} \circ T = S$ as above,

$$\mu_S = S_*(\mu) = (\pi_{ST} \circ T)_*(\mu) = (\pi_{ST})_*(T_*(\mu)) = (\pi_{ST})_*(\mu_T)$$

Example A.5. (Not from [2] or [4]) Let $E = \mathbb{R}^{[0,1]}$, the space of functions $f : [0,1] \to R$. Then the finite dimensional subspaces of E are $\mathbb{R}^n$ for each $n \in \mathbb{N}$. Let $\underline{t} = (t_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$ be a sequence of real numbers, and define the linear map $T_{\underline{t},n} : E \to \mathbb{R}^n$ by

$$T_{\underline{t},n}(f) = (f(t_1), \ldots, f(t_n)).$$

Then if μ is a probability measure on E, $A = A_1 \times \ldots \times A_n \subseteq \mathbb{R}^n$, we have

$$\begin{aligned} \mu_{T_{\underline{t},n}}(A) &= \mu(T_{\underline{t},n}^{-1}(A)) \\ &= \mu(\{f \mid (f(t_1), \ldots, f(t_n)) \in A\}) \\ &= \mu(\{f \mid f(t_1) \in A_1, \ldots, f(t_n) \in A_n\}) \end{aligned}$$

It can be seen that the measures $\mu_{T_{\underline{t},n}}$ satisfy the consistency relation. This example makes it clear why the name 'cylinder set measure' is appropriate: $\mu_{T_{\underline{t},n}}$ coincides with μ on cylinder sets, i.e. sets of a form like $A_1 \times \ldots A_n \times \mathbb{R} \times \mathbb{R} \times \ldots$.

Example A.6 (Canonical Gaussian CSM)**.** To any real Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$ we can associate a Gaussian CSM, $(\gamma_T^H)_{T\in\mathcal{A}(H)}$. Given $T : H \to F_T$ onto with $\dim F_T < \infty$, we define the measure γ_T^H on F_T by $\gamma_T^H := \gamma^{\langle\cdot,\cdot\rangle_T}$ where $\langle\cdot,\cdot\rangle_T$ is the quotient inner product on F_T:

$$\langle u, v\rangle_T := \left\langle T|_{(\ker T)^\perp}^{-1} u, T|_{(\ker T)^\perp}^{-1} v\right\rangle_H$$

Definition A.7. *Let E, G be separable Banach spaces and let $(\mu_T)_T$ be a cylinder set measure on E. A linear map $\theta : E \to G$ is said to* radonify $(\mu_T)_T$ *if $\theta_*(\mu_\cdot)$ is a measure on G.*

Definition A.8 (Abstract Wiener space)**.** *Let H be a separable Hilbert space, E a separable Banach space and $i : H \to E$ a continuous linear injective map with dense range. If i radonifies the canonical Gaussian cylinder set measure, we say that $i : H \to E$ is an* abstract Wiener space. *The induced measure $i_*(\gamma_\cdot^H)$ on E is called the* abstract Wiener measure *of $i : H \to E$.*

One can check that an abstract Wiener measure is a Gaussian measure. In fact the structure theorem for Gaussian measures tells us that the abstract Wiener space construction is the *only* way to obtain a Gaussian measure on a separable Banach space, that is every Gaussian measure on a separable Banach space is the pushforward of the canonical Gaussian cylinder set measure on some separable Hilbert space.

Example A.9. (Classical Wiener space) Let

$$\begin{aligned} H &:= L_0^{2,1}([0,T];\mathbb{R}^m) \\ &= \left\{\sigma : [0,T] \to \mathbb{R}^m \;\middle|\; \exists \varphi \in L^2([0,T];\mathbb{R}^m) \text{ such that } \sigma(t) = \int_0^t \varphi(s)ds\right\} \end{aligned}$$

So the elements of H start at the origin and have L^2 derivative. We equip H with the inner product

$$\langle\sigma_1, \sigma_2\rangle_{L_0^{2,1}} := \int_0^T \langle\dot\sigma_1(s), \dot\sigma_2(s)\rangle_{\mathbb{R}^m} ds$$

This can be seen to be a Hilbert space by noting that the operator $\frac{d}{dt} : L_0^{2,1} \to L^2$ is an isometry. Now let

$$\begin{aligned} E &:= C_0([0,t];\mathbb{R}^m) \\ &= \{\sigma : [0,T] \to \mathbb{R}^m \mid \sigma \text{ is continuous and } \sigma(0) = 0\} \end{aligned}$$

and equip it with the supremum norm. Then the inclusion $i : H \hookrightarrow E$ is continuous, linear and injective with dense range. It can be proved that the inclusion radonifies the canonical Gaussian cylinder set measure on H.

The Hilbert space H of an abstract Wiener space $i : H \to E$ is called the *Cameron-Martin space* or *reproducing kernel Hilbert space* (RKHS). One is not always provided with an abstract Wiener space, and is instead given a Gaussian measure on a separable Banach space E. The task is then to find the RKHS H of the measure so that with the inclusion i, $i : H \to E$

is an abstract Wiener space. Fortunately there is a standard way of constructing this space.

Let μ be a centred Gaussian measure on a separable Banach space E. Following [1] and [4], we define the covariance map $R : E^* \to E$ by

$$R(\ell) := \int_E \ell(x) x d\mu(x)$$

where the integral is in the Bochner sense. This integral does indeed belong to E:

$$\begin{aligned} \|R(\ell)\|_E &\leq \int_E \|\ell(x)x\|_E d\mu(x) \\ &= \int_E |\ell(x)| \|x\|_E d\mu(x) \\ &\leq \int_E \|\ell\|_{E^*} \|x\|_E^2 d\mu(x) \end{aligned}$$

which is finite by, for example, Fernique's integrability theorem for Gaussian measures (see [1, pp. 96-97]). The map R is hence a bounded linear operator. We can therefore define the RKHS:

Definition A.10. *Let μ be a centred Gaussian measure on a separable Banach space E. The* reproducing kernel Hilbert space *(RKHS) H_μ of μ is the completion of the image of the covariance map $R : E^* \to E$ with respect to the inner product*

$$\langle R(\ell_1), R(\ell_2) \rangle_\mu := \int_E \ell_1(x) \ell_2(x) d\mu(x)$$

It can be checked that H_μ is still a subset of E after the completion procedure.

An elementary stochastic integral called the Paley-Wiener map can be constructed on an abstract Wiener space, which acts on elements of the Cameron-Martin space. Let $i : H \to E$ be an abstract Wiener space with measure γ and let $j : E^* \to H \cong H^*$ be the adjoint of i, defined by $\langle j(\ell), h \rangle_H = \ell(i(h))$ for $\ell \in E^*$ and $h \in H$. This map is injective with dense range, which allows for the following theorem:

Theorem A.11. *If $\ell \in E^*$ then $\ell \in L^2(E, \gamma; \mathbb{R})$ with $\|\ell\|_{L^2} = \|j(l)\|_H$. Consequently, there is a unique continuous linear map $I : H \to L^2(E, \gamma; \mathbb{R})$, with $I(h) := \langle h, \cdot \rangle_H^\sim$, such that*

$$\begin{tikzcd} H \arrow[rr, "I"] & & L^2(E,\gamma;\mathbb{R}) \\ & E^* \arrow[ul, "j"] \arrow[ur, "\ell \mapsto [\ell]"'] & \end{tikzcd}$$

Moreover, I is an isometry into $L^2(E, \gamma; \mathbb{R})$, and is called the Paley-Wiener map.

In the classical case $i : L_0^{2,1}([0,1]) \hookrightarrow C_0([0,1])$, we think of the Paley-Wiener map as a stochastic integral. For $g, h \in H$, we have that

$$\begin{aligned} \langle g, h \rangle_{L_0^{2,1}} &= \int_0^1 \dot{g}_t \dot{h}_t dt \\ &= \int_0^1 \dot{g}_t dh_s \end{aligned}$$

where the last integral is interpreted in the Lebesgue-Stieltjes sense. We hence often write $\langle h, \cdot \rangle^{\sim}_{L_0^{2,1}} : C_0([0,1]) \to \mathbb{R}$ as

$$\langle h, \cdot \rangle^{\sim}_{L_0^{2,1}}(\sigma) = \int_0^1 \dot{h}_s d\sigma_s = \int_0^1 \dot{h}_s dB_s(\sigma)$$

since the evaluation map on classical Wiener space defines a Brownian motion B.

We quote some properties of the Paley-Wiener map:

Proposition A.12. *Let $i : H \to E$ be an abstract Wiener space with measure γ. Then for $h \in H$, the Paley-Wiener map $\langle h, \cdot \rangle^{\sim}_H$ satisfies*

(i) $\langle g, \cdot \rangle^{\sim}_H$ *is a Gaussian random variable on* $(E, \mathcal{B}(E), \gamma)$

(ii) $\displaystyle\int_E \langle h, \cdot \rangle^{\sim}_H(x) d\mu(x) = 0$

(iii) $\displaystyle\int_E \langle h, \cdot \rangle^{\sim}_H(x)^2 d\mu(x) = \|h\|_H^2$

(iv) $\displaystyle\int_E \langle g, \cdot \rangle^{\sim}_H(x) \langle h, \cdot \rangle^{\sim}_H(x) d\mu(x) = \langle g, h \rangle_H$

We now see what happens when an abstract Wiener measure is pushed forward by a translation by an element of its Cameron-Martin space.

Theorem A.13 (Cameron-Martin formula)**.** *Let $i : H \to E$ be an abstract Wiener space with measure γ and let $T_h : E \to E$ be the map $T_h(x) = x + i(h)$ for $h \in H$. Then $(T_h)_* \gamma \approx \gamma$ with*

$$\frac{d(T_h)_*\gamma}{d\gamma} = e^{\langle h, \cdot \rangle^{\sim}_H - \frac{1}{2}\|h\|_H^2}$$

Hence if $F : E \to \mathbb{R}$ is measurable, we have

$$\int_E F(x + i(h)) d\gamma(x) = \int_E F(x) e^{\langle h, \cdot \rangle^{\sim}_H(x) - \frac{1}{2}\|h\|_H^2} d\gamma(x)$$

Suppose now that $F : E \to \mathbb{R}$ is a measurable BC^1 function, i.e. a bounded measurable function with bounded Fréchet derivative. By the Cameron-Martin formula, for $t \in \mathbb{R}$ we have

$$\int_E F(x + ti(h))d\gamma(x) = \int_E F(x)e^{t\langle h,\cdot\rangle_H^\sim(x) - \frac{1}{2}t^2\|h\|_H^2}d\gamma(x)$$

Formally differentiating this at $t = 0$ we obtain the relation

$$\int_E (DF)_x(i(h))d\gamma(x) = \int_E F(x)\langle h,\cdot\rangle_H^\sim(x)d\gamma(x) \tag{A.1}$$

The integrand on the left hand side is our prototype for the Malliavin derivative in the h-direction. The relation is known as the integration by parts formula. Compare it to the divergence theorem in $\mathbb{R}^n$:

$$\int_{\mathbb{R}^n} (Df)_x(V(x))dx = -\int_{\mathbb{R}^n} \mathrm{div}(V(x))f(x)dx$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $V : \mathbb{R}^n \to \mathbb{R}^n$ are differentiable. For the (constant) vector field $V : E \to E$ given by $V(x) = i(h)$ it could hence seem reasonable to define

$$\mathrm{div}(V(x)) = -\langle h,\cdot\rangle_H^\sim(x)$$

Can we extend this to non-constant vector fields? The answer is yes, and it is done by defining the divergence operator as the adjoint of the Malliavin derivative operator. It is discussed in detail in the first chapter.

A.1.2 Isonormal Gaussian processes

Central to our definition of the Malliavin derivative operator is the notion of an isonormal process, which is a family of L^2 Gaussian random variables indexed by a Hilbert space:

Definition A.14. *Let H be a real separable Hilbert space, and let $W : H \to L^2(\Omega, \mathcal{F}, \mathbb{P})$. We say that W is an* isonormal Gaussian process *if*

(i) W is a linear isometry

(ii) $W(h)$ is normally distributed with mean zero and variance $\|h\|^2$

Remarks. (i) As an immediate consequence of the above definition we have that $\mathbb{E}[W(g)W(h)] = \langle g, h\rangle_H$ for all $g, h \in H$

(ii) The image of H under W is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, denoted $\mathcal{H}_1$. The reason for this notation will become clear later when looking at the chaos expansion of $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

This definition may seem a little abstract, so we'll look at a few examples.

Example A.15 (Paley-Wiener integral)**.** Let $i : H \to E$ be an abstract Wiener space. Define $W : H \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ by $W(h) = \langle h,\cdot\rangle_H^\sim$. Then by results from the previous section, this defines an isonormal Gaussian process.

In particular, on the canonical Wiener space $i : L_0^{2,1}([0,1]) \hookrightarrow C_0([0,1])$ we have

$$W(h) = \int_0^1 \dot{h}_s dB_s$$

where the integral with respect to the Brownian motion B is in the Itô sense. Then the isometry property of W is precisely the Itô isometry.

Remark. From now on we will identify $L_0^{2,1}$ with L^2 via the isometry $\frac{d}{dt} : L_0^{2,1} \to L^2$. Then we define an isonormal Gaussian process on $L^2([0,1])$ by

$$W(h) = \int_0^1 h_s dB_s$$

We can then easily recover the Brownian motion from W, since indicator functions lie in $L^2([0,1])$:

$$W(\mathbf{1}_{[0,t]}) = \int_0^1 \mathbf{1}_{[0,t]} dB_s = \int_0^t dB_s = B_t$$

The isometry property gives us the covariance function for Brownian motion,

$$\begin{aligned}\mathbb{E}B_sB_t &= \mathbb{E}(W(\mathbf{1}_{[0,s]})W(\mathbf{1}_{[0,t]})) \\ &= \langle \mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]} \rangle_{L^2([0,1])} \\ &= s \wedge t\end{aligned}$$

The following examples are taken from [10, pp 11-15].

Example A.16 (Brownian sheet)**.** A Brownian sheet (or two-parameter Brownian motion) is a two-parameter stochastic process defined by the three properties

(i) $B_{s,t} = 0$ when $s = 0$ or $t = 0$

(ii) B has independent increments[9]

(iii) B is a (centred) Gaussian process with covariance

$$\mathbb{E}B_{s_1,t_1}B_{s_2,t_2} = (s_1 \wedge s_2)(t_1 \wedge t_2)$$

An Itô calculus can be defined for this process, giving the integral of two-parameter processes with respect to the Brownian sheet (see [] for details). Let $H = L^2([0,1]^2)$,

[9]For every pair of disjoint rectangles R_1 and R_2 of $[0,1]^2$, the increment of B on R_1 is independent of the increment of B on R_2. The increment of B on a rectangle $R = [s_1, s_2] \times [t_1, t_2]$, $\Delta_R B$, is defined by

$$\Delta_R B = (B_{s_2,t_2} - B_{s_1,t_2}) - (B_{s_2,t_1} - B_{s_1,t_1})$$

where $[0,1]^2$ is equipped with the two dimensional Lebesgue measure. Define the process $W : H \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ by

$$W(h) = \int_0^1 h_{s,t} dB_{s,t}$$

Then W is an isonormal Gaussian process, and we can recover $B_{s,t}$ and its covariance function from W similarly to in the one-parameter case.

It is useful to note that both one- and two-parameter Brownian motions are particular cases of white noise, which is defined as follows.

Definition A.17 (White noise)**.** *Let $(T, \mathcal{T}, \mu)$ be a σ-finite measure space without atoms and let H be the Hilbert space $L^2(T, \mathcal{T}, \mu)$. Let W be an isonormal Gaussian process on H. We define a* white noise with intensity μ *as the process*

$$\{W(\mathbf{1}_A) \,|\, A \in \mathcal{T}, \mu(A) < \infty\}$$

We define the natural filtration $(\mathcal{F}_A, A \in \mathcal{T})$ *of this process as the completion of the filtration given by*

$$\widetilde{\mathcal{F}_A} = \sigma(\{W(\mathbf{1}_B) \,|\, B \in \mathcal{T}, \mu(B) < \infty, B \subseteq A\})$$

Remarks. (i) From now on we will write $W(A) := W(\mathbf{1}_A)$ when A is a set. In this way we think of W as a (random) vector measure on $(T, \mathcal{T}, \mu)$. For $h \in H$ we can think of $W(h)$ as an integral with respect to this measure.

(ii) In the case that $(T, \mathcal{T}, \mu) = ([0,T], \mathcal{B}([0,T]), \lambda)$ and W is given by the Itô integral, we recover standard Brownian motion by defining

$$B_t := W([0,t]), \; \mathcal{F}_t := \mathcal{F}_{[0,t]}$$

Similarly we can recover two-parameter Brownian motion.

Are there any useful Gaussian processes which aren't white noise? There are, and one commonly used example is fractional Brownian motion (fBm):

Definition A.18. *A* fractional Brownian motion with Hurst parameter $\mathfrak{h} \in (0,1)$ *is a centred Gaussian process $(B_t^{\mathfrak{h}})_{t\in[0,1]}$ starting from 0 whose covariance is given by*

$$\mathbb{E} B_s^{\mathfrak{h}} B_t^{\mathfrak{h}} = \frac{1}{2}\left(s^{2\mathfrak{h}} + t^{2\mathfrak{h}} - |t-s|^{2\mathfrak{h}}\right)$$

Note that in the case $\mathfrak{h} = 1/2$ we have standard Brownian motion (a centred Gaussian process is entirely determined by its covariance function). For general $\mathfrak{h} \in (0,1)$, fBm has the following properties.

Proposition A.19 (Properties of fBm)**.** *Let $\mathfrak{h} \in (0,1)$ and let $B^{\mathfrak{h}}$ be fractional Brownian motion with Hurst parameter $\mathfrak{h}$. Then,*

(i) $B^{\mathfrak{h}}$ has Hölder continuous paths of order α for all $\alpha < \mathfrak{h}$

(ii) $B^{\mathfrak{h}}$ *is self-similar: for each* $a > 0$, $(a^{-\mathfrak{h}} B^{\mathfrak{h}}_{at})$ *has the same law as* $(B^{\mathfrak{h}}_t)$

(iii) $B^{\mathfrak{h}}$ *has stationary increments*

(iv) $B^{\mathfrak{h}}$ *has independent increments if and only if* $\mathfrak{h} = 1/2$

(v) $B^{\mathfrak{h}}$ *is a semimartingale if and only if* $\mathfrak{h} = 1/2$

(vi) $B^{\mathfrak{h}}$ *is a Markov process if and only if* $\mathfrak{h} = 1/2$

The first property shows that the sample paths of $B^{\mathfrak{h}}$ are less regular than those of B for $\mathfrak{h} < 1/2$. The reduced regularity coupled with the lack of semimartingale property means that the definition of a stochastic integral with respect to such $B^{\mathfrak{h}}$ is non-trivial. Nualart develops a stochastic calculus for fractional Brownian motion with Hurst parameter less than 1/2 using Malliavin calculus. This won't be discussed here, but details can be found in [].

We check that fBm can be viewed as an isonormal Gaussian process. Let $\mathcal{E}$ be the set of step functions on $[0, 1]$. Define the Hilbert space $H^{\mathfrak{h}}$ as the closure of $\mathcal{E}$ with respect to the inner product

$$\langle \mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]} \rangle_{H^{\mathfrak{h}}} = \frac{1}{2} \left(s^{2\mathfrak{h}} + t^{2\mathfrak{h}} - |t-s|^{2\mathfrak{h}} \right)$$

Now define the linear isometry $W^{\mathfrak{h}} : H^{\mathfrak{h}} \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ by $W^{\mathfrak{h}}(\mathbf{1}_{[0,t]}) := B^{\mathfrak{h}}_t$. One can check that this defines an isonormal Gaussian process, but that the space $H^{\mathfrak{h}}$ is not of the form $L^2(T, \mathcal{T}, \mu)$ for some measure space $(T, \mathcal{T}, \mu)$ and so this process does not define a white noise.

A.2 Chaos expansions

We give a brief overview of the relevant properties of chaos expansions. The statements in this appendix and much more detail can be found in [9, §1].

A.2.1 Iterated Itô integrals - classical case

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, and let $(B_t, t \in [0, T])$ be a stardard one-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}_t$ be the (completion of) the σ-algebra generated by $(B_s, s \in [0, t])$. Define the simplex

$$\Delta_n = \{(t_1, \dots, t_n) \,|\, 0 \le t_1 \le t_2 \le \dots \le t_n \le T\} \subseteq [0, T]^n$$

Then for $f \in L^2(\Delta_n)$, i.e. measurable deterministic f defined on Δ_n such that

$$\|f\|^2_{L^2(\Delta_n)} := \int_{\Delta_n} f^2(t_1, \dots, t_n) dt_1 \dots dt_n < \infty,$$

we define the n-fold iterated Itô integral as

$$J_n(f) = \int_0^T \int_0^{t_n} \cdots \int_0^{t_3} \int_0^{t_2} f(t_1, ..., t_n) dB_{t_1} ... dB_{t_n}$$

Note that each successive integrand is square integrable (with respect to $d\mathbb{P} \times dt_i$) and $\mathcal{F}_\cdot$ adapted due to the conditions on $t_1, \dots t_n$, so this is well-defined.

Then for symmetric $f \in L^2([0,1]^n)$ we can define $I_n(f) = n! J_n(f)$. These are related to the Hermite polynomials $(H_n)_{n\in\mathbb{N}}$,

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n}(e^{-x^2/2}),$$

by, for example,

$$I_n(g^{\otimes n}) = \|g\|^n H_n\left(\frac{\int_0^1 g(t)dB_t}{\|g\|}\right)$$

when $g \in L^2([0,1])$.

Example A.20. $H_5(x) = x^5 - 10x^3 + 15x$, so

$$5! \int_0^1 \int_0^{t_5} \cdots \int_0^{t_2} 1 dB_{t_1} ... dB_{t_5} = 1 \cdot H_5\left(\frac{B_1}{1}\right) = B_1^5 - 10B_1^3 + 15B_1$$

Let $\widetilde{L}^2([0,T]^n) := \{f \in L^2([0,T]^n) \mid f \text{ is symmetric}\}$, and for $f \in L^2([0,T]^n)$ let $\widetilde{f}$ denote the symmetrisation of f, i.e.

$$\widetilde{f} := \sum_{\sigma\in S_n} f(\sigma(x_1), \dots, \sigma(x_n))$$

where S_n is the group of bijections of $\{1, \dots, n\}$ to itself.

Proposition A.21. *The operator $I_n : L^2([0,T]^n) \to L^2(\Omega)$ has the following properties:*

(i) *I_n is linear*

(ii) $I_n(f) = I_n(\widetilde{f})$

(iii) $\mathbb{E}(I_n(f)I_m(g)) = \delta_{mn} \cdot n! \langle \widetilde{f}, \widetilde{g} \rangle_{L^2([0,T]^n)}$

We can now state the following result:

Theorem A.22. *Let $F \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F} = \sigma(\{B_s \mid s \in [0,T]\})$. Then there exists a sequence of functions $(f_n)_{n\in\mathbb{N}}$ with $f_n \in L^2([0,T]^n)$ such that*

$$F = \sum_{n=0}^{\infty} I_n(f_n)$$

If in addition the f_n are required to be symmetric, the above expansion is unique - in this case, we say that this is 'the' chaos expansion of F.

Proof follows from iterated application of the martingale representation theorem.

If X_t is a stochastic process, we can apply the chaos expansion at each time to get

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

Example A.23. What is the chaos expansion of

$$\exp\left(\int_0^1 g(t)dB_t\right)?$$

First note that

$$\exp\left(tx - \frac{t^2}{2}\right) = \sum_{n=0}^{\infty} \frac{t^n}{n!} H_n(x)$$

and so

$$\exp\left(\int_0^1 g(t)dB_t - \frac{\|g\|^2}{2}\right) = \sum_{n=0}^{\infty} \frac{\|g\|^n}{n!} H_n\left(\frac{\theta}{\|g\|}\right)$$

thus

$$\begin{aligned}\exp\left(\int_0^1 g(t)dB_t\right) &= e^{\|g\|^2/2} \sum_{n=0}^{\infty} \frac{\|g\|^n}{n!} \cdot \frac{I_n(g^{\otimes n})}{\|g\|^n} \\ &= \sum_{n=0}^{\infty} I_n\left(e^{\|g\|^2/2} \frac{g^{\otimes n}}{n!}\right)\end{aligned}$$

A.2.2 Iterated Itô integrals - white noise case

We consider now the multiple integral in the more general case when the process isn't necessarily indexed by a time interval in the real line.

Definition A.24 (White noise)**.** *Let $(T, \mathcal{T}, \mu)$ be a σ-finite measure space without atoms and let H be the Hilbert space $L^2(T, \mathcal{T}, \mu)$. Let W be an isonormal Gaussian process on H. We define a* white noise with intensity μ *as the process*

$$\{W(\mathbf{1}_A) \,|\, A \in \mathcal{T}, \mu(A) < \infty\}$$

We define the natural filtration $(\mathcal{F}_A, A \in \mathcal{T})$ *of this process as the completion of the filtration given by*

$$\widetilde{\mathcal{F}_A} = \sigma(\{W(\mathbf{1}_B) \,|\, B \in \mathcal{T}, \mu(B) < \infty, B \subseteq A\})$$

Remark. From now on we will write $W(A) := W(\mathbf{1}_A)$ when A is a set. In this way we think of W as a (random) vector measure on $(T, \mathcal{T}, \mu)$. For $h \in H$ we can think of $W(h)$ as an integral with respect to this measure.

Remark. In the case that $(T, \mathcal{T}, \mu) = ([0,T], \mathcal{B}([0,T]), \lambda)$ and W is given by the Itô integral, we recover standard Brownian motion by defining

$$B_t := W([0,t]),\ \mathcal{F}_t := \mathcal{F}_{[0,t]}$$

Let $\mathcal{E}_n$ be the set of elementary functions of the form

$$f(t_1, \ldots, t_n) = \sum_{i_1,\ldots,i_n}^{k} a_{i_1\ldots i_n} \mathbf{1}_{A_{i_1} \times \ldots \times A_{i_n}}(t_1, \ldots, t_n) \tag{A.2}$$

where $A_1, \ldots, A_k$ are pairwise disjoint sets of finite measure, and $a_{i_1\ldots i_n} = 0$ whenever $i_\alpha = i_\beta$ for some $\alpha \neq \beta$. We can define the integral of such functions:

Definition A.25 (Multiple integral)**.** *Let $f \in \mathcal{E}_n$ be an elementary function with representation (A.2). We define the multiple integral operator $I_n : \mathcal{E}_n \subseteq L^2(T^n, \mathcal{T}^{\otimes n}, \mu^{\otimes n}) \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ by*

$$I_n(f) := \sum_{i_1,\ldots,i_n}^{k} a_{i_1\ldots i_n} W(A_{i_1})\ldots W(A_{i_n}) \tag{A.3}$$

We quote some results on this operator without proof. Compare with the results for the basic case in the previous subsection.

Proposition A.26. *The operator $I_n : \mathcal{E}_n \to L^2(\Omega)$ has the following properties:*

(i) I_n *is linear*

(ii) $I_n(f) = I_n(\widetilde{f})$

(iii) $\mathbb{E}(I_n(f)I_m(g)) = \delta_{mn} \cdot n!\langle \widetilde{f}, \widetilde{g}\rangle_{L^2(T^n)}$

(iv) There exists a continuous extension $I_n : L^2(T^n) \to L^2(\Omega)$ that satisfies the three properties above and is given by (A.3) on $\mathcal{E}_n$.

Theorem A.27. *Let $F \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ where $\mathcal{G} = \sigma(\{W(h) \,|\, h \in H\})$. Then there exists a sequence of functions $(f_n)_{n\in\mathbb{N}} \subseteq L^2(T^n)$ such that*

$$F = \sum_{n=0}^{\infty} I_n(f_n)$$

If in addition the f_n are required to be symmetric, the above expansion is unique - in this case, we say that this is 'the' chaos expansion of F.

Remark. As before, if $(F_t)_{t\in T} \subseteq L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a stochastic process we write

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

A.3 Sobolev inequalities

As a nice application of the Clark-Ocone theorem we prove two Sobolev inequalities following [1, pp. 146-148]. These are analogous to the inequalities for finite dimensional Gaussian Sobolev spaces, see [11] for details.

Theorem A.28 (Sobolev inequality)**.** *For every $F \in \mathbb{D}^{1,2}$ the following inequality holds*

$$\mathbb{E}(F^2) \leq (\mathbb{E}(F))^2 + \mathbb{E}(\|DF\|_H^2) \tag{A.4}$$

Equality is achieved if F is Gaussian, i.e. $F = C + W(h)$ for some $C \in \mathbb{E}$, $h \in H$.

Proof. By Clark-Ocone we have

$$F = \mathbb{E}(F) + \int_0^T \mathbb{E}[D_t F|\mathcal{F}_t] dB_t$$

So by the Itô isometry, Jensen's inequality and the tower property of conditional expectations we have

$$\begin{aligned}
\mathbb{E}(F^2) &= (\mathbb{E}(F))^2 + \mathbb{E}\left[\int_0^T \mathbb{E}[D_t F|\mathcal{F}_t]^2 dt\right] \\
&\leq (\mathbb{E}(F)^2 + \mathbb{E}\left[\int_0^T (D_t F)^2 dt\right] \\
&= (\mathbb{E}(F))^2 + \mathbb{E}(\|DF\|_H^2)
\end{aligned}$$

□

Theorem A.29 (log-Sobolev inequality)**.** *For every $F \in \mathbb{D}^{1,2}$ the following inequality holds*

$$\mathbb{E}(F^2 \log(F^2)) \leq \mathbb{E}(F^2) \log(\mathbb{E}(F^2)) + 2\mathbb{E}(\|DF\|_H^2) \tag{A.5}$$

Equality is achieved if F is lognormal, i.e. $F = C \exp(W(h))$ for some $C \in \mathbb{R}$, $h \in H$.

Proof. Define the family of bounded random variables $G_N := F \wedge N$ for $N \in \mathbb{N}$ so that $G_N \uparrow F$. Let $(M_t^N)_{t\geq 0}$ be the positive continuous martingale given by $M_T^N = G_N^2$ and $M_t^N = \mathbb{E}[G_N^2|\mathcal{F}_t]$. By Clark-Ocone we have

$$M_t^N = \mathbb{E}(G_N^2) + \int_0^t \mathbb{E}[D_s G_N^2|\mathcal{F}_s] dB_s$$

or equivalently

$$dM_t^N = \mathbb{E}[D_t G_N^2|\mathcal{F}_t] dB_t, \; M_0^N = \mathbb{E}(G_N^2)$$

Now by Itô's formula we have

$$M_t^N \log(M_t^N) = M_0^N \log(M_0^N) + \int_0^t (1 + \log(M_s^N)) dM_s^N + \frac{1}{2}\int_0^t \frac{1}{M_s^N} d\langle M^N, M^N\rangle_s$$

Combining the above formulas gives, using the chain rule for D, Cauchy-Schwarz and the tower property of conditional expectations,

$$\begin{aligned}
\mathbb{E}(G_N^2 \log(G_N^2)) &= \mathbb{E}(M_T^N \log(M_T^N)) \\
&= \mathbb{E}(G_N^2) \log(\mathbb{E}(G_N^2)) + \frac{1}{2}\mathbb{E}\left[\int_0^T \frac{\mathbb{E}[D_s G_N^2|\mathcal{F}_s]^2}{M_s^N} ds\right] \\
&= \mathbb{E}(G_N^2) \log(\mathbb{E}(G_N^2)) + 2\mathbb{E}\left[\int_0^T \frac{\mathbb{E}[G_N D_s G_N|\mathcal{F}_s]^2}{M_s^N} ds\right] \\
&\leq \mathbb{E}(G_N^2) \log(\mathbb{E}(G_N^2)) + 2\mathbb{E}\left[\int_0^T \frac{\mathbb{E}[G_N^2|\mathcal{F}_s]\mathbb{E}[(D_s G_N)^2|\mathcal{F}_s]}{M_s^N} ds\right] \\
&= \mathbb{E}(G_N^2) \log(\mathbb{E}(G_N^2)) + 2\mathbb{E}\left[\int_0^T (D_s G_N)^2 ds\right] \\
&= \mathbb{E}(G_N^2) \log(\mathbb{E}(G_N^2)) + 2\mathbb{E}(\|DG_N\|^2)
\end{aligned}$$

Letting $N \to \infty$ and using the monotone convergence theorem gives the result. □

Remark. In particular this theorem tells us that if $F \in \mathbb{D}^{1,2}$ then $\mathbb{E}(F^2 \log(F^2)) = 2\mathbb{E}(F^2 \log(F)) < \infty$. We therefore have that $\mathbb{D}^{1,2}$ is embedded in the space

$$L^{2,\log}(\Omega) := \{F : \Omega \to \mathbb{R} \mid \mathbb{E}(F^2 \log^+(|F|)) < \infty\}$$

where $\log^+(x) = \max\{0, \log(x)\}$.

References

[1] René Carmona and Michael Tehranchi, *Interest Rate Models: an Infinite Dimensional Stochastic Analysis Perspective*, Springer, 2006.

[2] David Elworthy, *MA482 Stochastic Analysis - lecture notes*, 2006.

[3] Peter K. Friz, *An Introduction to Malliavin Calculus*, 2002.

[4] Martin Hairer, *An Introduction to Stochastic PDEs*, 2009.

[5] ______, *On Malliavin's proof of Hörmander's theorem*, (2011).

[6] David Nualart, *The Malliavin Calculus and Related Topics*, first ed., Springer, 1995.

[7] ______, *Malliavin Calculus and its Applications*, AMS, 2008.

[8] Eulalia Nualart, *Lectures on Malliavin Calculus and its Applications to Finance*, 2009.

[9] Giulia Di Nunno and Bernt Øksendal, *Malliavin Calculus for Lèvy Processes with Applications to Finance*, Springer, 2009.

[10] Anthony Réveillac, *An Introduction to Malliavin's Calculus and its Applications*, 2010.

[11] Gordan Zitkovic, *Topics in Stochastic Analysis - lecture notes*, 2011.