Notes on Malliavin Calculus

Joe Jackson

May 20, 2020

These are lecture notes for a summer 2020 mini course on Malliavin Calculus. First, we will review stochastic integration, and introduce the basic operators of Malliavin calculus. We will then take a detour to study some basic SDE theory, and see the connection between SDEs and the Cauchy problem. Finally, we will explain how Malliavin calculus can be applied to give a probabilistic proof of Hörmander's Theorem. Sections 2 and 4 of the notes borrow heavily from the book *Introduction to Malliavin Calculus* by David Nualart, and much of Section 5 on Hörmander's Theorem I learned from an expository paper by Martin Hairer, titled *On Malliavin's Proof of Hörmander's Theorem*.

Contents

1 Probabilistic Setup **2**
- 1.1 Wiener Space . . . 2
- 1.2 Definite Itô Integral . . . 2
- 1.3 Indefinite Itô Integral . . . 4
- 1.4 Stratonovich Integral . . . 4

2 The Malliavin Derivative and its Adjoint **5**
- 2.1 Definitions . . . 5
- 2.2 Interpreting D and δ . . . 7
- 2.3 Some Tools . . . 8

3 Stochastic Differential Equations **10**
- 3.1 Existence and Uniqueness . . . 11
- 3.2 The Derivative of an SDE Solution . . . 13
- 3.3 Cauchy Problem and SDEs . . . 14

4 Malliavin Calculus and Densities **15**
- 4.1 Some Results in 1-D. . . . 15
- 4.2 Main Criteria for the Existence of Smooth Densities . . . 17

5 Hörmander's Theorem **17**
- 5.1 Statement of the Theorem . . . 18
- 5.2 Intuition about Hörmander's Condition . . . 19

1 Probabilistic Setup

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $X : \Omega \to \mathbb{R}$ a random variable. We would like to talk about the "derivative" of X, but this is hopeless without some more analytical structure on Ω. Luckily, many random variables of interest are defined as functionals of some Brownian motion, in which case we might as well take $(\Omega, \mathcal{F}, \mathbb{P})$ to be the Wiener space.

1.1 Wiener Space

Definition 1.1. The **Wiener space** is the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = C(\mathbb{R}_+; \mathbb{R})$, $\mathcal{F}$ is the Borel σ-algebra there, and $\mathbb{P}$ is the unique measure such that the process $B_t(\omega) = \omega(t)$ is a Brownian Motion.

The Wiener space also comes with a natural choice of filtration, namely the augmentation of the natural filtration of B. More precisely, we take $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, where

$$\mathcal{F}_t = \sigma(\{B_s : s \leq t\} \vee \{A \in \mathcal{F} : \mathbb{P}[A] = 0\}). \tag{1}$$

For the rest of these notes, $(\Omega, \mathcal{F}, \mathbb{P})$ will be the Wiener space, $B : \Omega \times \mathbb{R}_+$ will be the Brownian motion $B_t(\omega) = \omega(t)$, and $\mathbb{F}$ will be the filtration defined in (1).

Exercise 1.2. Let $\mathcal{H}$ denote the set of random variables of the form

$$F(\omega) = f(\omega(t_1), ..., \omega(t_n)) = f(B_{t_1}(\omega), ..., B_{t_n}(\omega)$$

for some $0 \leq t_1 < ... < t_n < \infty$ and $f : \mathbb{R}^n \to \mathbb{R}$ bounded and measurable. Show that $\mathcal{H}$ is dense in $L^2(\Omega, \mathcal{F})$.

1.2 Definite Itô Integral

Now we define the integral of a process with respect to B.

Definition 1.3. A process $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is called **simple** if it takes the form $X = \sum_i H_i 1_{(t_i, t_{i+1}]}(t)$ for a some $0 = t_0 < t_1 < ... < t_n < \infty$ and $H_i \in L^2(\Omega, \mathcal{F}_{t_i})$.

The integral of a simple process is easy to define, and analogous to the Riemann integral of a step function.

Definition 1.4. If $X = \sum_i H_i 1_{(t_i, t_{i+1}]}(t)$ is simple, then $\int_0^\infty X_t dB_t := \sum_i H_i (B_{t_{i+1}} - B_{t_i}) \in L^2(\Omega)$ is the **Itô integral of X with respect to B**.

We now have a well-defined map $X \mapsto \int_0^\infty X_t dB_t$ from simple processes into $L^2(\Omega) = L^2(\Omega, \mathcal{F})$. In fact, this map is an isometry.

Proposition 1.5. *If X and Y are simple processes, then*

$$\langle \int_0^\infty X_t dB_t, \int_0^\infty Y_t dB_t \rangle_{L^2(\Omega)} = \langle X, Y \rangle_{L^2(\Omega \times \mathbb{R}_+)}$$

Exercise 1.6. Prove Proposition 1.5.

Now we want to identify the closure of the set of simple processes in $L^2(\Omega \times \mathbb{R}_+)$.

Definition 1.7. A process $X : \Omega\times\mathbb{R}_+ \to \mathbb{R}$ is **progressively measurable** (or progressive) if for all t, $X|_{\Omega\times[0,t]}$ is measurable with respect to the σ-algebra $\mathcal{F}_t\otimes\mathcal{B}([0,t])$. The **progressive σ-algebra**, denoted $\mathcal{P}$, is the σ-algebra on $\Omega \times \mathbb{R}_+$ generated by all progressive processes.

Exercise 1.8. Show that

$$\mathcal{P} = \sigma(\{A \in \mathcal{F} \otimes \mathcal{B}([0, T]) : A \cap (\Omega \times [0, t]) \in \mathcal{F}_t \otimes \mathcal{B}([0, t]) \ \forall\ t\}).$$

We will use $L^2(\mathcal{P})$ to denote the space $L^2(\Omega \times \mathbb{R}_+, \mathcal{P})$, and view it in the natural way as a closed subspace of $L^2(\Omega \times \mathbb{R}_+) = L^2(\Omega \times \mathbb{R}_+, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+))$. Then we have the following:

Proposition 1.9. *The closure of the space of simple process in $L^2(\Omega \times \mathbb{R}_+)$ is $L^2(\mathcal{P})$.*

Exercise 1.10. The proof of Proposition 1.9 takes some work, but one inclusion is easy. Which one is it, and why?

In light of Propositions 1.5 and 1.9, the Itô integral on simple processes extends uniquely to an isometry $X \mapsto \int_0^\infty X_s dB_s$ on $L^2(\mathcal{P})$, which we will also call the Itô integral. More precisely, we have:

Definition 1.11. For $X \in L^2(\mathcal{P})$, the **Itô integral of X with respect to B** is given by $\int_0^\infty X_s dB_s \coloneqq \lim_{n\to\infty} \int_0^\infty X^n_s dB_s$, where the limit is taken in $L^2(\Omega)$, and $\{X^n\}$ is any sequence of simple process approaching X in $L^2(\Omega \times \mathbb{R}_+)$.

For $T < \infty$ we define $\int_0^T X_t dB_t \coloneqq \int_0^\infty X_t 1_{[0,t]} dB_t$.

Example 1.12. We can prove directly from the definitions that $\int_0^t s dB_s = tB_t - \int_0^t B_s ds$. Indeed, for a partition $\Delta = (t_0, ..., t_n)$ with $0 = t_0 < ... < t_n = t$, let $h^\Delta(t)$ be the step function

$$h^\Delta = \sum_i t_i 1_{[t_i,t_{i+1})}.$$

Then $h^\Delta(s) \to s$ in $L^2([0, t])$ as $||\Delta|| \to 0$, so by the definition of the Itô integral, $\int_0^t s dB_s$ is the L^2 limit of the random variables

$$\int_0^t h(s)dB_s = \sum_i t_i(B_{t_{i+1}} - B_{t_i})$$

as the mesh of Δ tends to zero. We have

$$\begin{aligned}\sum_i t_i(B_{t_{i+1}} - B_{t_i}) &= \sum_i t_i B_{t_{i+1}} - \sum_i t_{i+1} B_{t_{i+1}} + \sum_i t_{i+1} B_{t_{i+1}} - \sum_i t_i B_{t_i} \\ &= tB_t - \sum_i B_{t_{i+1}}(t_{i+1} - t_i).\end{aligned}$$

Thus the proof is complete if we can show that $\sum_i B_{t_{i+1}}(t_{i+1}-t_i) \to \int_0^t B_s ds$ in L^2, as $||\Delta|| \to 0$ but this follows from the dominated convergence theorem, because clearly $\sum_i B_{t_{i+1}}(t_{i+1} - t_i) \to \int_0^t B_s ds$ almost surely, and the sequence is dominated by $t \sup_{0\le s\le t} |B_s| \in L^2$.

1.3 Indefinite Itô Integral

Let $\mathcal{M}_2$ be the be the space of continuous, square integrable martingales $M = (M_t)_{t\geq 0}$ such that $\sup_t \mathbb{E}[|M_t|^2] < \infty$. The martingale convergence theorem shows that if $M \in \mathcal{M}_2$, then there exists $M_\infty \in L^2(\Omega, \mathcal{F}_\infty)$ such that $M_t \to M_\infty$ a.s. and in L^2. Furthermore, Doob's maximal inequality shows that $\mathcal{M}_2$ is a Hilbert space under the inner product $\langle M, N\rangle_{\mathcal{M}^2} = \mathbb{E}[M_\infty N_\infty]$. We will now use this Hilbert space structure to define the indefinite Itô integral.

Definition 1.13. If $X = \sum_i G_i 1_{(t_i,t_{i+1}]}$ is a simple process, then the **indefinite integral of** X **with respect to** B is the process

$$(t,\omega) \mapsto \int_0^t X_s dB_s(\omega) := \sum_i G_i(\omega)(B_{t_{i+1}\wedge t}(\omega) - B_{t_i\wedge t}(\omega)).$$

We will often denote the indefinite integral by $\int X_s dB_s$. Just as with the definite integral, we have an isometry property:

Proposition 1.14. *If X simple, then $\int X_s dB_s$ is a continuous square integrable martingale. For X, Y simple,*

$$\langle \int XdB, \int YdB\rangle_{\mathcal{M}^2} = \langle X, Y\rangle_{L^2(\Omega\times\mathbb{R}_+)}.$$

As in the definite case, this allows us to define an isometry $X \mapsto \int XdB_s$ from $L^2(\mathcal{P})$ to $\mathcal{M}_2$.

Definition 1.15. For $X \in L^2(\mathcal{P})$, $\int XdB = \lim_{n\to\infty} \int X^n dB$, where the limit is taken in $\mathcal{M}_2$ and X^n is any sequence of simple processes approaching X in $L^2(\Omega \times \mathbb{R}_+)$.

Exercise 1.16. Review Doob's L^p maximal inequality (if you need to), and use it to give a proof that $\mathcal{M}_2$ is complete.

We define $H := L^2(\mathbb{R}_+)$. If we restrict to deterministic integrands, the definite Itô integral gives an isometry $H \to L^2(\Omega)$, $h \mapsto B(h) := \int_0^\infty h(t)dB_t$. In fact, the image of the map $B : H \to L^2(\Omega)$ contains only Gaussian random variables.

Proposition 1.17. *If $h \in H$, the continuous martingale $\int hdB$ is a Gaussian process; that is, for any $0 \leq t_1, ..., t_n \leq \infty$, the random vector $(\int_0^{t_1} h(t)dB_t, ..., \int_0^{t_n} h(t)dB_t)$ is a multivariate Gaussian. In particular, $B(h)$ is a Gaussian random variable.*

The map B can be viewed as a special case of something called Gaussian white noise, and is a basic building block of Malliavin calculus.

1.4 Stratonovich Integral

For most of these notes we will use the Itô integral, but it will be helpful when stating Hörmander's theorem to also have the Stratonovich formulation of SDEs at our disposal.

This section is a very short and very informal introduction to the Stratonovich integral. One can show that for a sufficiently nice integrand $X \in L^2(\mathcal{P})$, the Itô integral of X is given by

$$\int_0^T X_t dB_t = \lim_{n\to\infty} \sum_i X_{t_i^n}(B_{t_{i+1}^n} - B_{t_i^n}),$$

where the limit is taken in L^2, $\Delta_n = \{t_1^n, ..., t_{k_n}^n\}$ is a partition of $[0, T]$ and the mesh of Δ_n tends to zero.

Interestingly, the choice to approximate X using left-endpoints matters. If instead we use the midpoint, we get the **Stratonovich integral**, which for sufficently nice X is given by

$$\int_0^T X \circ dB_t = \lim_{n\to\infty} \sum_i \Big(\frac{X_{t_i^n} + X_{t_{i+1}^n}}{2}\Big)(B_{t_{i+1}^n} - B_{t_i^n}).$$

It turns out that the Itô and Stratonovich integrals are related by the formula

$$\int_0^T X_t \circ dB_t = \int_0^T X_t dB_t + \frac{1}{2}\langle X, B\rangle_t.$$

2 The Malliavin Derivative and its Adjoint

We will now define the Malliavin derivative and its adjoint, the divergence operator. The constrution is similar to the construction of the derivative operators on Sobolev spaces; first we define the desired operations on a very nice space, and then we use functional analysis to extend.

2.1 Definitions

Let $C_p^\infty(\mathbb{R}^n)$ to be the set of smooth functions $\mathbb{R}^n \to \mathbb{R}$ all of whose derivatives grow at most polynomially. We now define $\mathcal{S}$ to be the set of random variables of the form $F = f(B(h_1), ..., B(h_n))$, where $h_i \in H$ and $f \in C_p^\infty(\mathbb{R}^n)$. Similarly, we define $\mathcal{S}_H$ to be the set of process of the form $u = \sum_{i=1}^n F_i h_i$ where $F_i \in \mathcal{S}$ and $h_i \in H$. It turns out that $\mathcal{S}$ and $\mathcal{S}_H$ give the appropriate "nice spaces" on which to initially define our differential operators.

Definition 2.1. For $F = f(B(h_1), ..., B(h_n)) \in \mathcal{S}$, the Malliavin derivative of F is the process

$$D_t F := \sum_{i=1}^n f_{x_i}(B(h_1), ..., B(h_n))h_i(t) \in L^2(\Omega \times \mathbb{R}_+).$$

Definition 2.2. For $u = \sum_{i=1}^n F_i u_i \in \mathcal{S}_H$, the **divergence** of u, denoted $\delta(u)$, is the random variable

$$\delta(u) := \sum_{i=1}^n F_i B(h_i) - \sum_i \langle DF_j, h_j\rangle_H \in L^2(\Omega).$$

It turns out that the divergence is the adjoint of the derivative.

Proposition 2.3. *If $F \in \mathcal{S}$ and $u \in \mathcal{S}_H$, then*

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

This adjointness relationship allows us to prove that the operator D is closeable.

Proposition 2.4. *The operator $D : \mathcal{S} \subset L^2(\Omega) \to L^2(\Omega \times \mathbb{R}_+)$ is closeable.*

Proof. Suppose that $\{F_n\} \subset \mathcal{S}$ with $F_n \to 0$ in $L^2(\Omega)$ and $DF_n \to u$ in $L^2(\Omega\times\mathbb{R}_+)$. To show that D is closeable, we must show that $u = 0$. Let $v \in \mathcal{S}_H$. Then by adjointness, we have

$$\mathbb{E}[\langle u, v\rangle] = \lim_{n\to\infty} \mathbb{E}[\langle DF_n, v\rangle] = \lim_{n\to\infty} \mathbb{E}[F_n\delta(v)] = 0.$$

Because $\mathcal{S}$ is dense in $L^2(\Omega \times \mathbb{R}_+)$, we conclude that $\mathbb{E}[\langle u, v\rangle] = 0$ for all $v \in L^2(\Omega \times \mathbb{R}_+)$, and thus $u = 0$ as required. $\square$

Thus there exists a closed extension of D, which we also denote by D, defined on the space

$$\mathbb{D}^{1,2} := \{F \in L^2 : \text{ there exists } F_n \in \mathcal{S}, u \in L^2(\Omega \times \mathbb{R}_+) \text{ with } F_n \to F, DF_n \to u\}.$$

Finally, we define $\delta(u)$ by extending the adjointess relationship as far as possible.

Definition 2.5. The domain of the divergence operator is given by

$$Dom(\delta) := \{u \in L^2(\Omega \times \mathbb{R}_+) : \text{there exists } F \in \mathbb{D}^{1,2} \text{ with } \mathbb{E}[\langle u, DG\rangle] = \mathbb{E}[FG \text{ for all } G \in \mathbb{D}^{1,2}\},$$

and for $u \in Dom(\delta)$, we define $\delta(u) = F$, where F satisfies the above condition.

For $u = \sum_i F_i h_i \in \mathcal{S}_H$, $u_t \in \mathcal{S}$ for each t, and so the two-parameter process $Du \in L^2(\Omega \times \mathbb{R}^2_+)$ given by $Du(s,t) = D_s u_t$ is well-defined, and is given explicitly by

$$D_s u_t = \sum_i D_s F_i h_i(t).$$

Thus, we have a mapping $D : \mathcal{S}_H \subset L^2(\Omega \times \mathbb{R}_+) \to L^2(\Omega \times \mathbb{R}^2_+)$. In fact, just as in the case of $D : \mathcal{S} \to L^2(\Omega \times \mathbb{R}_+)$, one can show that this operator is closeable, and so by taking its closure we get a mapping $\mathbb{D}^{1,2}(H) \to L^p(\Omega \times \mathbb{R}^2_+)$, where

$$\mathbb{D}^{1,2}(H) := \{u \in L^2(\Omega \times \mathbb{R}_+) : \text{ there exist } u_n \in \mathcal{S}_H, v \in L^2(\Omega \times \mathbb{R}^2_+) \text{ with } u_n \to u, Du_n \to v\},$$

and for $u \in \mathbb{D}^{1,2}(H)$ we have $D_s u_t = v(s,t)$ where v is as above.

For simplicity, these notes will mostly be concerned with the spaces $\mathbb{D}^{1,2}$, $Dom(\delta)$, and $\mathbb{D}^{1,2}(H)$, but there is a whole family of related spaces. If $F = f(B(h_1), ..., B(h_n)) \in \mathcal{S}$, we can define iterated derivatives

$$D^k_{t_1,\dots,t_k} F = \sum_{i_1,\dots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1},,,x_{i_k}}(B(h_1), ..., B(h_n))h_{i_1}(t)...h_{i_k}(t),$$

and another closeability argument yields an operator $D^k : \mathbb{D}^{k,2} \to L^2(\Omega \times \mathbb{R}^k_+)$. One can also view $\mathcal{S}$ as a subset of L^p, and D^k as operator from $\mathcal{S} \to L^p(\Omega \times \mathbb{R}^k_+)$. Taking the closure of D in this setting defines a space $\mathbb{D}^{k,p}$ and an operator $D : \mathbb{D}^{k,p} \to L^p(\Omega \times \mathbb{R}^k_+)$.

Finally, we define $\mathbb{D}^{k,\infty} = \cap_{p\geq 1}\mathbb{D}^{k,\infty}$, and $\mathbb{D}^{\infty} = \cap_{k\geq 1}\mathbb{D}^{k,\infty}$.

Exercise 2.6. Compute the Malliavin derivative of the following random variables:

(a) $X = \big(\int_0^T \sin(t) dB_t\big)^2$

(b) $X = B_1 B_2 B_3$

2.2 Interpreting D and δ

There are three results which I think help provide some intuition about the operators D and δ. First, we show how to interpret DF as the "gradient" of F.

Definition 2.7. The **Cameron-Martin space** is the subspace of $C(\mathbb{R}_+;\mathbb{R})$ given by $\mathcal{C} = \{g \in C(\mathbb{R}_+;\mathbb{R}) : g(t) = \int_0^t \dot{g}(s)ds, \dot{g} \in H\}$.

The Cameron-Martin space provides a set of "directions" in which we can differentiate a random variable $F \in \mathbb{D}^{1,2}$. For $F \in \mathbb{D}^{1,2}$ and $h \in H$, we define $D_h F = \langle DF, h\rangle \in L^2(\Omega)$. For $g \in \mathcal{C}$, we define $\tau_g : \Omega \to \Omega$ by $\tau_g(\omega) = \omega + g$. Since τ_g is continuous, hence measurable on Ω, $F \circ \tau_g$ is a well-defined random variable.

Theorem 2.8. *If $F \in \mathbb{D}^{1,2}$, then for any $g \in \mathcal{C}$, we have*

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon}\big(F \circ \tau_{\epsilon g} - F\big) = D_{\dot{g}} F,$$

where the limit is taken in probability.

So, to find the "directional derivative of F in the direction of g", we take the inner product of DF with $\dot{g}$. This should be compared with the usual directional derivative of a map $\mathbb{R}^n \to \mathbb{R}$.

Exercise 2.9. Prove Theorem 2.8 in the special case that $F \in \mathcal{S}$.

Analytically, δ is best understood as the adjoint of D. Since D is like the gradient, this means δ is like the classical divergence of a vector field. There is another more probabilisitc interpretation of δ; it can be viewed as an extension of the Itô integral. To prove this, we need a lemma concerning measurability and D.

Definition 2.10. For $0 \leq a < b < \infty$, we define $\mathcal{F}_a^b := \sigma(\{B_t - B_a : t \in [a,b]\})$.

Intuitively, if a $F \in \mathcal{F}_a^b$, then F depends only on the behavior of the Brownian motion between a and b. Formally, we have:

Lemma 2.11. *If $F \in \mathbb{D}^{1,2} \cap \mathcal{F}_a^b$, then $DF = 0$ on $\Omega \times [a,b]^c$, $d\mathbb{P} \otimes dt$ a.s.*

In particular, this implies that if $F \in \mathbb{D}^{1,2} \cap \mathcal{F}_t$, then $DF = 0$ on $\Omega \times [t, \infty)$, $d\mathbb{P} \otimes dt$ a.s. Here is the result stating the relationship between δ and the Itô integral.

Proposition 2.12. *We have $L^2(\mathcal{P}) \subset Dom(\delta)$, and for $u \in L^2(\mathcal{P})$, $\delta(u) = \int_0^\infty u_t dB_t$.*

Proof. Suppose that first that $u = \sum_i F_i 1_{[t_i,t_{i+1})}$ is simple. Then $u \in \mathcal{S}_H$, and by definition

$$\delta(u) = \sum_i F_i(B_{t_{i+1}} - B_{t_i}) - \sum_i \langle DF_i, 1_{[t_i,t_{i+1})}\rangle.$$

The second term is zero by lemma 2.11 because $F_i \in \mathcal{F}_t$, and thus

$$\delta(u) = \sum_i F_i(B_{t_{i+1}} - B_{t_i}) = \int_0^\infty u_t dB_t.$$

The general case follows from approximation. □

By the martingale representation theorem, any random variable $F \in L^2(\Omega)$ can be expressed as $F = \mathbb{E}[F] + \int_0^\infty Z_s dB_s$. We now prove the Clark-Ocone formula, which shows that if $F \in \mathbb{D}^{1,2}$, then in fact we can choose Z to be the optional projection of DF.

Theorem 2.13. *(Clark-Ocone Formula) If $F \in \mathbb{D}^{1,2}$, then*

$$F = \mathbb{E}[F] + \int_0^\infty \mathbb{E}[D_t|\mathcal{F}_t]dB_t.$$

Proof. We know that $F = \mathbb{E}[F] + \int_0^\infty Z_s dB_s$ for some $Z \in L^2(\mathcal{P})$, so we need only show that $(Z_t)_t$ and $(\mathbb{E}[D_t|\mathcal{F}_t])_t$ are equal as elments of $L^2(\mathcal{P})$. For $u \in L^2(\mathcal{P})$, we have

$$\mathbb{E}[\langle Z, u\rangle] = \mathbb{E}[(F - \mathbb{E}[F])\Big(\int_0^\infty u_s dB_s\Big)] = \mathbb{E}[(F - \mathbb{E}[F])\delta(u)] = \mathbb{E}[\langle DF, u\rangle]$$
$$= \int_0^\infty \mathbb{E}[D_t F u_t]dt = \int_0^\infty \mathbb{E}[\mathbb{E}[D_t F|\mathcal{F}_t]u_t]dt = \mathbb{E}[\langle \mathbb{E}[D_t F|\mathcal{F}_t], u\rangle],$$

which completes the proof. □

2.3 Some Tools

Here are some of the main tools for computing and/or estimating Malliavin derivatives. We start with the chain rule:

Proposition 2.14. *Let $\phi \in C^1(\mathbb{R})$ with bounded derivative, and $F \in \mathbb{D}^{1,2}$. Then $\phi(F) \in \mathbb{D}^{1,2}$, and $D\phi(F) = \phi'(F)DF$.*

Proof. First, suppose that $F = f(B(h_1), ..., B(h_n)) \in \mathcal{S}$. Then $\phi(F) = \phi \circ f(B(h_1), ..., B(h_n)) \in \mathcal{S}$, and we can explicitly compute

$$D_t(\phi(F)) = \sum_i (\phi \circ f)_{x_i}(B(h_1), ..., B(h_n))h_i(t)$$
$$= \sum_i \phi'(f(B(h_1), ..., B(h_n))f_{x_i}(B(h_1), ..., B(h_n)) = \phi'(F)D_t F.$$

Thus the result holds for $F \in \mathcal{S}$. For $F \in \mathbb{D}^{1,2}$, we can find a sequence $\{F_j\} \subset \mathcal{S}$, such that $F_j \to F$ in $L^2(\Omega)$, $DF_j \to DF$ in $L^2(\Omega \times \mathbb{R}_+)$. Since ϕ is Lipschitz and $F_n \to F$, $\phi(F_n) \to \phi(F)$ in $L^2(\Omega)$. Furthermore, along a subsequence we have $F_n \to F$, and $DF_n \to DF$, and so by the dominated convergence theorem $D\phi(F_n) = \phi'(F_n)DF_n \to \phi'(F)DF$. The result now follows from the closedness of D. □

A mollification argument lets us extend the chain rule to Lipschitz functions.

Proposition 2.15. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be Lipschitz, i.e. $|\phi(x) - \phi(y)| \le K|x-y|$ for some K. Let $F \in \mathbb{D}^{1,2}$. Then $F(\phi) \in \mathbb{D}^{1,2}$, and $DF = GDF$ for some random variable G with $||G||_{L^\infty} \le K$. If the law of F is absolutely continuous with respect to the Lebesgue measure, then $G = \phi^{'}(F)$, where $\phi^{'}$ is the weak derivative of ϕ.*

Exercise 2.16. Compute the derivative of $X = \sup_{0 \le t \le 1} B_t$ by approximating X by random variables of the form $X^n = \max\{B_{t_1}, ..., B_{t_n}\}$, and using the chain rule for Lipschitz functions.

Next, we show how to compute derivatives of Lebesgue integrals.

Proposition 2.17. *Let $u \in \mathbb{D}^{1,2}(H)$, and $T > 0$. Then $\int_0^T u_t dt \in \mathbb{D}^{1,2}$, and a version of $D\big(\int_0^T u_t dt\big)$ is given by*

$$D_s\Big(\int_0^T u_t dt\Big) = \int_0^T D_s u_t dt.$$

Proof. Suppose first that $u = \sum_{i=1}^n F_i h_i \in \mathcal{S}_H$. Then

$$\int_0^T u_t dt = \sum_{i=1}^n \Big(\int_0^T h_i(t) dt\Big) F_i \in \mathcal{S},$$

and so

$$D_s\Big(\int_0^T u_t dt\Big) = \sum_{i=1}^n \Big(\int_0^T h_i(t) dt\Big) D_s F_i = \int_0^T D_s\Big(\sum_i h_i(t) F_i\Big) dt = \int_0^T D_s u_t dt.$$

Now for $u \in \mathbb{D}^{1,2}(H)$, there is a sequence $u^n \in \mathcal{S}_H$ with $u^n \to u$ in $L^2(\Omega \times \mathbb{R}_+)$ and $Du^n \to Du$ in $L^2(\Omega, \mathbb{R}_+^2)$. It is easy to check that $\int_0^T u_t^n dt \to \int_0^T u_t dt$ in $L^2(\Omega)$ and $(\omega, s) \mapsto \int_0^T D_s u_t^n dt(\omega)$ converges to $(\omega, s) \mapsto \int_0^T D_s u_t dt(\omega)$ in $L^2(\Omega \times \mathbb{R}_+)$. This lets us pass to the limit in the equation

$$D_s\Big(\int_0^T u_t^n dt\Big) = \int_0^T D_s u_t dt$$

to get the result.

□

We can also differentiate integrals against Brownian motion. We need the following lemma regarding the divergence:

Lemma 2.18. *If $u, v \in \mathbb{D}^{1,2}(H)$, then*

$$\mathbb{E}[\delta(u)\delta(v)] = \mathbb{E}\Big[\int_0^\infty \int_0^\infty D_s u_t D_t v_s ds dt\Big].$$

Proposition 2.19. *Let $u \in \mathbb{D}^{1,2} \cap L^2(\mathcal{P})$. Then $\int_0^T u_t dB_t \in \mathbb{D}^{1,2}$, and a version of Du is given by*

$$D_s\Big(\int_0^T u_t dB_t\Big) = \begin{cases} u_s + \int_s^T D_s u_t dB_t & s \le T \\ 0 & s > T. \end{cases}$$

Note that it is not obvious that the process $(s, \omega) \mapsto \int_s^T u_t dB_t(\omega)$ defines an element on $L^2(\Omega \times \mathbb{R}_+)$, but this can be remedied by choosing a sufficiently nice version of Du. We will ignore any measure-theoretic difficulties, and simply assume that we can choose a verion of Du such that $D_s u_\cdot \in L^2(\mathcal{P})$ for each s (and thus $\int_s^T D_s u_t dB_T$ makes sense as a random variable), and also that these random variables can be chosen so that $(s, \omega) \mapsto \int_s^T D_s u_t dB_t(\omega)$ is measurable. Under this assumption the statement makes sense, and we can give a simple proof.

Proof. Without loss of generality, we can assume $u_t = 0$ for $t > T$. For $v \in \mathbb{D}^{1,2}(H)$, we use Lemma 2.18 and the Clark-Ocone formula to compute

$$\begin{aligned}
\mathbb{E}[\langle D(\int_0^T u_s dB_s), v\rangle] &= \mathbb{E}[\big(\int_0^T u_s dB_s\big)\delta(v)] = \mathbb{E}[\delta(u)\delta(v)] \\
&= \mathbb{E}[\int_0^\infty u_t v_t dt] + \mathbb{E}[\int_0^\infty \int_0^\infty D_s u_t D_t v_s dt ds] \\
&= \mathbb{E}[\langle u, v\rangle] + \int_0^\infty \mathbb{E}[\big(\int_0^\infty D_s u_t dB_t\big)\big(\int_0^\infty D_t v_s dB_t\big)] ds \\
&= \mathbb{E}[\langle u, v\rangle] + \int_0^\infty \mathbb{E}[\big(\int_0^\infty D_s u_t dB_t\big) v_s ds \\
&= \mathbb{E}[\langle u, v\rangle] + \mathbb{E}[\langle \int_\cdot^T D_\cdot u_t dB_t, v\rangle],
\end{aligned}$$

from which the result follows. □

3 Stochastic Differential Equations

In this section, we fix a time horizon T, and we consider the SDE

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dB_t & 0 \le t \le T, \\ X_0 = x_0. \end{cases} \tag{2}$$

The equivalent integral formulation is

$$X_t = x_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s, \quad 0 \le t \le T. \tag{3}$$

Here B is a d-dimensional Brownian motion, X is n-dimensional, $x_0 \in \mathbb{R}^n$, and $b : \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ are both uniformly Lipschitz, i.e. $|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \le L|x - y|$. A solution to (2) is a progressive process X such that $b(X_t) \in L^{1,loc}$, $\sigma(X_t) \in L^{2,loc}$, and (3) holds for all t, almost surely.

Example 3.1. The Ornstein Uhlenbeck SDE is given by

$$\begin{cases} dX_t = bX_t dt + \sigma dB_t \\ X_0 = x_0, \end{cases}$$

and Itô's formula shows that an explicit solution is given by

$$X_t = e^{bt}x_0 + \sigma \int_0^t e^{b(t-s)} dB_s.$$

Exercise 3.2. Assume that X and B are one-dimensional, so $b, \sigma : \mathbb{R} \to \mathbb{R}$. Find $\tilde{b}$ and $\tilde{\sigma} : \mathbb{R} \to \mathbb{R}$ (in terms of b and σ) so that the Stratonovich SDE

$$\begin{cases} dX_t = \tilde{b}(X_t)dt + \tilde{\sigma}(X_t) \circ dB_t \quad 0 \leq t \leq T, \\ X_0 = x_0. \end{cases}$$

is equivalent to the Itô SDE (2).

3.1 Existence and Uniqueness

We introduce the space of processes

$$\mathbb{S}^2 := \{X : \Omega \times [0,T] \to \mathbb{R} : X \text{ is progressive }, X_\cdot \text{ is continuous } a.s., \text{ and } ||X||^2_{\mathbb{S}^2} := \mathbb{E}[\sup_{0 \leq t \leq T} |X_t|^2] < \infty\}.$$

It turns out that $\mathbb{S}^2$ is the right regularity class for solutions of (2).

We recall a couple of important inequalities, namely Gronwall's inequality (in integral form) and the Burkholder-Davis-Gundy (BDG) inequality.

Proposition 3.3. *(Gronwall's Inequality) If $u : [0,T] \to \mathbb{R}$ is a continuous function that satisfies $u(t) \leq \alpha + \int_0^t \beta(s)u(s)ds$, then*

$$u(t) \leq \alpha \exp(\int_0^t \beta(s)ds).$$

Proposition 3.4. *(BDG Inequality) There are constants c_p, C_p such that for any continuous local martingale M starting at zero, any stopping time τ and $1 \leq p < \infty$*

$$c_p \mathbb{E}[\langle M_\tau \rangle^{p/2}] \leq \mathbb{E}[\sup_{0 \leq t \leq \tau} |M_t|^p] \leq C_p \mathbb{E}[\langle M_\tau \rangle^{p/2}]$$

Exercise 3.5. Use the BDG inequality to show that if $X \in L^\infty \cap L^2(\mathcal{P})$, then $M = \int X dB$ satisfies $\sup_{0 \leq t \leq T} |M_t|^p \in L^p(\Omega)$ for all $0 \leq T < \infty$ and $1 \leq p < \infty$.

Exercise 3.6. Give examples showing that neither inequality in Proposition 3.4 holds when $p = \infty$.

Example 3.7. If B is a standard Brownian motion then $\langle B \rangle_T = T$, so the BDG inequality immediately gives that $\sup_{0 \leq t \leq T} |B_t| \in L^p$ for all $0 \leq T < \infty$ and all $1 \leq p < \infty$.

We now give an estimate which will imply uniquenes.

Proposition 3.8. *There is a constant C depending on T, L, d, and n such that if $X^i \in \mathcal{S}^2$ solves*

$$\begin{cases} dX_t^j = b(X_t^j)dt + \sigma(X_t^j) \cdot dB_t \\ X_0^j = x_j, \end{cases}$$

$j = 1, 2$, and $\Delta X = X^1 = X^2$, then $\mathbb{E}[\sup_{0 \le t \le T} |\Delta X_t|^2] \le C|x_1 - x_2|^2$.

Proof. For notational convenience, we assume that X and B are one-dimensional, so $n = d = 1$ and $b, \sigma : \mathbb{R} \to \mathbb{R}$. We allow C to vary from line to line, so long as it depends only on T and L (and implicitly on d and n). We define $\Delta X_t = X_t^1 - X_t^2$, and note that

$$\Delta X_t = x_1 - x_2 + \int_0^t b(X_s^1) - b(X_s^2)ds + \int_0^t \sigma(X_s^1) - \sigma(X_s^2, s)dB_s,$$

from which it follows that

$$|\Delta X_t|^2 \le C\big(|x_1 - x_2|^2 + \int_0^t |b(X_s^1) - b(X_s)|^2 ds + |\int_0^t \sigma(X_s^1) - \sigma(X_s^2)dB_s|^2\big).$$

Thus

$$\sup_{0 \le r \le t} |\Delta X_r|^2 \le C\big(|x_1 - x_2|^2 + \int_0^t |b(X_s^1) - b(X_s^2)|^2 ds + \sup_{0 \le r \le t} |\int_0^t \sigma(X_s^1) - \sigma(X_s^2)dB_s|^2\big).$$

Applying the Burkholder-Davis-Gundy inequality and the Lipschitz assumption on b and σ shows that

$$\begin{aligned} \mathbb{E}[\sup_{0 \le r \le t} |\Delta X_r|^2] &\le C\big(|x_1 - x_2|^2 + \int_0^t \mathbb{E}[\Delta X_s^2]ds\big) \\ &\le C\big(|x_1 - x_2|^2 + \int_0^t \mathbb{E}[\sup_{0 \le r \le s} \Delta X_r^2]ds\big). \end{aligned}$$

and finally Gronwall's inequality applied to the function $t \mapsto \mathbb{E}[\sup_{0 \le r \le t} |\Delta X_r|^2$ shows that

$$\mathbb{E}[\sup_{0 \le t \le T} |\Delta X_t|^2] \le C|x_1 - x_2|^2.$$

□

Theorem 3.9. *There exists a unique process $X \in \mathbb{S}^2$ which solves* (2)*.*

Proof. Uniqueness is given by Propositiono 3.8. We present only a sketch of the proof of existence, since it just a (slightly fancier) version of the Picard iteration proof for existence for ODEs with Lipschitz coefficients. Define processes $X^n, n \ge 0$ by

$$X_t^0 = x_0$$

$$X_t^{n+1} = x_0 + \int_0^t b(X_s^n)ds + \int_0^t \sigma(X_s^n)dB_s.$$

Then one can use the Lipschitz assumption to show that the sequence $\{X^n\}$ is Cauchy in $\mathbb{S}^2$, and hence as a limit X. Passing to the limit in (3) shows that in fact X is the desired solution. □

One can show using a the Burkholder-Davis-Gundy inequality that any solution to (2) which is square integrable (i.e. $\mathbb{E}[\int_0^T |X_s|^2 ds] < \infty$) is automatically in $\mathbb{S}^2$, and use this to give a stronger version of the uniqueness in the preceding proposition.

3.2 The Derivative of an SDE Solution

If the coefficients of (2) are very regular, we expect the solutions to also be very regular. The proof of this is difficult, and we won't cover it, but we can at least give a precise statement:

Proposition 3.10. *Suppose that σ and b are C^∞ with bounded derivatives of all orders, and let X be the unique solution of (2). Then for each $0 \le t \le T$, $X_t \in \mathcal{D}^\infty$.*

Assume for the moment that X and B are one-dimensional. One can show that $X \in \mathbb{D}^{1,2}(H)$, and thus the same is true of $b(X)$ and $\sigma(X)$, so we can apply Propositions 2.17 and 2.19 to compute

$$\begin{aligned} D_s X_t &= D_s\big(\int_0^t b(X_r)dr\big) + D_s\big(\int_0^t \sigma(X_r)dB_r\big) \\ &= \sigma(X_s) + \int_s^t D_s(b(X_r))dr + \int_s^t D_s(\sigma(X_r))dB_r \\ &= \sigma(X_s) + \int_s^t b^{'}(X_r)D_sX_r dr + \int_s^t \sigma^{'}(X_r)D_sX_r dB_r, \end{aligned} \tag{4}$$

for $s \le t$. Of course, we could do the same thing if n and d are not one, it just comes out a little messier. For fixed s, (4) says that the process $D_sX.$ solves a linear SDE with random coefficients on $[s,T]$. More rigorously, for any version of DX, we have that for almost every s, the process $(D_sX_t)_{s\le t\le T}$ is a version of Y, where Y solves the SDE

$$Y_t = \sigma(X_s) + \int_s^t b^{'}(X_r)Y_r dr + \int_s^t \sigma'(X_r)Y_r dB_r. \tag{5}$$

Thus a-priori estimates for the equation (5) yield estimates on the derivative of X. This idea of differentiating an SDE to get a new SDE is very useful.

Exercise 3.11. Solve in closed form the SDE

$$\begin{aligned} dX_t &= \mu X_t dt + \sigma X_t dB_t \\ X_0 &= x_0, \end{aligned}$$

where $\mu, \sigma, x_0 \in \mathbb{R}$, and verify (4) in this case.

Exercise 3.12. For fixed $t > 0$ show that $F \mapsto \mathbb{E}[F|F_t]$ is a contraction on $\mathbb{D}^{1,2}$. Conclude that if M is a square integrable martingale with $M_T \in \mathbb{D}^{1,2}$ for some T, then $M_t \in \mathbb{D}^{1,2}$ for all $0 \le t \le T$.

3.3 Cauchy Problem and SDEs

Recall the SDE (2). For convenience, here it is again:

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dB_t & 0 \le t \le T \\ X_0 = x_0. \end{cases}$$

Let $a = \sigma\sigma^T : \mathbb{R}^n \to \mathbb{R}^{n\times n}$, and consider the differential operator $\mathcal{L}$ given by

$$\mathcal{L}f(x) = \sum_{i=1}^{n} b_i(x)\partial_i f(x) + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\partial_{ij} f(x). \tag{6}$$

The basic connection between the operator $\mathcal{L}$ and the SDE (2) is that for $f \in C^2$,

$$df(X_t) = \mathcal{L}f(X_t)dt + dM_t,$$

where M is a martingale. This follows immediately from Itô 's Lemma. Consider the following Cauchy problem:

$$\begin{cases} u_t(t,x) = \mathcal{L}u(t,x) & (t,x) \in (0,\infty)\times\mathbb{R}^n \\ u(0,x) = f(x). \end{cases} \tag{7}$$

Here $u : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ is the unkown, and for simplicity we assume $f \in C_c(\mathbb{R}^n)$. For $x \in \mathbb{R}^n$ X_t^x denote the unique solution to (2) with initial condition $x_0 = x$. We now show how to represent solutions to (7) using solutions to (2). Then we have

Proposition 3.13. *If $u \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$ solves (7) and u is bounded, then $u(t,x) = \mathbb{E}[f(X_t^x)]$.*

Proof. For fixed t_0, consider the process $u(X_t, t_0 - t)$. Itô 's lemma shows that

$$du(X_t, t_0 - t) = \mathcal{L}u(X_t, t_0 - t)dt - u_t(X_t, t_0 - t)dt + dM_t = dM_t,$$

where M_t is a local martingale. Since u is bounded by assumption, $u(X_t, t_0 - t)$ is a martingale, and thus

$$u(x, t_0) = \mathbb{E}[u(X_0^x, t_0 - 0)] = \mathbb{E}[u(X_{t_0}, 0)] = \mathbb{E}[f(X_{t_0})].$$

□

Example 3.14. If $\mathcal{L} = \frac{1}{2}\Delta$, the corresponding SDE is $dX_t = dB_t$, and so the solution to (7) is given by $u(t,x) = \mathbb{E}[f(B_t + x)]$.

Exercise 3.15. Suppose that b and σ are bounded (and Lipschitz), and that X_t^x denotes the unique solution to (2) with initial condition x. Show that if $f \in C^2(\mathbb{R}^d)$ is bounded with bounded first and second order derivatives, then we have

$$\lim_{t\downarrow 0} \frac{1}{t}\big(\mathbb{E}[f(X_t^x)] - f(x)\big) = \mathcal{L}f(x).$$

For those familiar with semigroups, this shows that $\mathcal{L}$ is the generator of the semigroup T given by $T(t)f(x) = \mathbb{E}[f(X_t^x)]$.

4 Malliavin Calculus and Densities

One of the primary applications of Malliavin calculus is to the study of densities. The most powerful results about densities, which will be necessary for Hörmander's Theorem, are too involved to prove in a week long course. But I'd still like to show how Malliavin calculus can be used to prove things about densities, so I will prove some simple results in 1-D, and then state without proof the main result about random vectors which is required for Hörmander's theorem.

First, let's recall what we mean by the density of random variable. For a random vector X in $\mathbb{R}^n$, its **density** is the measure μ_X on $\mathbb{R}^n$ given by

$$\mu_X(A) = \mathbb{P}[X \in A] = \mathbb{P}[X^{-1}(A)], \quad A \in \mathcal{B}(\mathbb{R}^n).$$

Recall that X is called **continuous** if the measure μ_X is absolutely continuous with respect to the Lebesgue measure, in which case the **probability density function** (pdf) of X is the Radon Nikodym derivative $f_X = \frac{d\mu_X}{d\lambda}$, where λ is the Lebesgue measure on $\mathbb{R}^n$. When I say that Malliavin calculus can be used to study densities, what I mean is that we can use Malliavin calculus to describe conditions under which the random vector X has a nice density.

Remember that the Malliavin derivative is interpreted as a gradient. To give some intuition about how gradients relate to densities, here is a problem:

Exercise 4.1. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is smooth, and $\nabla f > 0$. Then if λ denotes the Lebesgue measure on $\mathbb{R}^n$, the push-forward measure μ defined by $\mu(A) = \lambda(f^{-1}(A))$ is atomless, i.e. $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}$. Is μ absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$? Is there an analogous condition for a map $f : \mathbb{R}^n \to \mathbb{R}^m$?

4.1 Some Results in 1-D.

We start with a lemma about indicator functions.

Lemma 4.2. *Let $A \in \mathcal{F}$. Then $1_A \in \mathbb{D}^{1,2}$ if and only if $\mathbb{P}[A]$ is zero or one.*

Proof. If $\mathbb{P}[A]$ is zero or one, then 1_A is an a.s. constant random variable, hence in $\mathbb{D}^{1,2}$. For the other direction, suppose that $1_A \in \mathbb{D}^{1,2}$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be smooth, with $\phi(x) = x^2$ for all $x \in [0, 1]$. Then $\phi(1_A) = 1_A$, and by the chain rule

$$D1_A = D\phi(1_A) = \phi'(1_A)D1_A = 21_A D1_A.$$

In particular, this implies that $D1_A = 0$, so 1_A must be (a.s.) equal to a constant. It follows that either $\mathbb{P}[A] = 1$ or $\mathbb{P}[A] = 0$. □

We now apply this to give our first result about distributions.

Proposition 4.3. *If $X \in \mathbb{D}^{1,2}$, then supp(μ_X) is a closed interval.*

Proof. Since $\mathrm{supp}(\mu_X)$ is always a closed subset of $\mathbb{R}$, it suffices to show that $\mathrm{supp}(\mu_X)$ is connected. Suppose towards a contradiction that $\mathrm{supp}(\mu_X)$ is not connected. Then we can find $a < b$ such that $\mathbb{P}[X < a] > 0$, $\mathbb{P}[X > b] > 0$, and $\mathbb{P}[a \leq X \leq b] = 0$. Let $\phi \in C_c^\infty(\mathbb{R})$ be such that $\phi = 0$ on $(-\infty, a]$, and $\phi = 1$ on $[b, \infty)$. Then since $X \in \mathbb{D}^{1,2}$, so is $\phi(X)$. But $\phi(X) = 1_{X \geq b}$, and $0 < \mathbb{P}[X \geq b] < 1$. This contradicts Lemma 4.2, and completes the proof. □

Now we state a simple criterion for the existence of a density.

Proposition 4.4. *Let $X \in \mathbb{D}^{1,2}$, and suppose that $||DX||_H > 0$ a.s. Then μ_X is absolutely continuous with respect to the Lebesgue measure.*

In fact, with a little more information about DX, we can get an explicit formula for the density.

Proposition 4.5. *Let $X \in \mathbb{D}^{1,2}$, and suppose that $||DX||_H > 0$ a.s., and also $\frac{DX}{||DX||_H^2} \in$ Dom(δ). Then X has a continuous and bounded density f_X given by*

$$f_X(x) = \mathbb{E}[1_{\{X>x\}}\delta\Big(\frac{DX}{||DX||_H^2}\Big)]. \tag{8}$$

Proof. It suffices to show that for $a < b$, we have $\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(x)dx$. We use adjointness and the chain rule to compute

$$\begin{aligned}
\int_a^b f_X(x)dx &= \int_a^b \mathbb{E}[1_{\{X>x\}}\delta\Big(\frac{DX}{||DX||_H^2}\Big)]dx \\
&= \mathbb{E}[\Big(\int_a^b \mathbb{E}[1_{\{X>x\}}]dx\Big)\delta\Big(\frac{DX}{||DX||_H^2}\Big)] \\
&= \mathbb{E}[(a \vee X \wedge b - a)\delta\Big(\frac{DX}{||DX||_H^2}\Big)] \\
&= \mathbb{E}[\langle D(a \vee X \wedge b - a)\frac{DX}{||DX||_H^2}\rangle_H] \\
&= \mathbb{E}[1_{[a,b]}(X)\langle DX, \frac{DX}{||DX||_H^2}\rangle_H] \\
&= \mathbb{E}[1_{[a,b]}(X)] = \mathbb{P}[a \leq X \leq b].
\end{aligned}$$

□

Example 4.6. If $X = B_1$, then $DX = 1_{[0,1]}$, and so applying Proposition 4.5 shows that the density f of B_1 satisfies $f(x) = \mathbb{E}[1_{B_1>x}B_1]$.

Exercise 4.7. Verify (8) in the case that $X = B(h)$ for some $h \in H$.

4.2 Main Criteria for the Existence of Smooth Densities

Let $X = (X_1, ..., X_n)$ be a random vector in $(\mathbb{D}^{1,2})^n$. The **Malliavin matrix** of F is the random $n \times n$ matrix γ_X given by

$$(\gamma_X)_{ij} = \langle DX_i, DX_j \rangle_H.$$

Roughly speaking, the random variable $det(\gamma_X)$ plays the same role for random vectors as $||DX||_H$ does for random variables.

Definition 4.8. A random vector $X = (X_1, ..., X_n)$ is called **non-degenerate** if $X_i \in \mathbb{D}^{1,2}$, and

$$\mathbb{E}[(\det(\gamma_X))^{-p}] < \infty]$$

for all $p \geq 2$.

The main result concerning existence and smoothness of densities is the following.

Theorem 4.9. *Let $X = (X_1, ..., X_n)$ be a non-degenerate random vector such that $X_i \in \mathbb{D}^{\infty}$ for each i. Then X has an infinitely differentiable density.*

5 Hörmander's Theorem

Recall the Cauchy problem (7) from section 3.3. When $\mathcal{L} = \Delta$, (7) reduces to the heat equation

$$\begin{cases} \partial_t u(t,x) = \Delta u(t,x) & (t,x) \in (0,\infty) \times \mathbb{R}^n \\ u(0,x) = f(x), \end{cases}$$

and it is well know that the unique solution is given by

$$u(t,x) = \int_{\mathbb{R}^n} p(t,x,y) f(y) dy,$$

where p is the fundamental solution

$$p(t,x,y) = (4\pi t)^{-n/2} e^{-|x-y|^2/4t}.$$

In fact, it is a classical result that the same thing works when $\mathcal{L}$ is uniformly elliptic. That is, if $\mathcal{L}$ is uniformly elliptic, we can represent the unique solution to (7) as $u(t,x) = \int_{\mathbb{R}^n} p(t,x,y) f(y) dy$ for some fundamental solution p such that $p(t,\cdot,\cdot) : \mathbb{R}^{2n} \to \mathbb{R}$ is smooth for each t. Hörmander's Theorem says that this is possible for a much larger class of operators.

5.1 Statement of the Theorem

First, we formulate Hörmander's bracket condition. Let σ_i be a vector field on $\mathbb{R}^n$, for $0 \le i \le d$. We define

$$V_0 = \{\sigma_i : 1 \le i \le d\},$$
$$V_k = \{[v, \sigma_i] : v \in V_{k-1}, 0 \le i \le d\} \text{ for } 1 \le k \le d.$$

Finally we define $V = \cup_{k=0}^d V_k$. We say that $\{\sigma_i\}_{i=1}^d$ satisfies **Hörmander's condition** if at each point $x \in \mathbb{R}^n$, $\{v(x) : v \in V\}$ spans $\mathbb{R}^n$.

Exercise 5.1. If $\sigma_0, \sigma_1 : \mathbb{R} \to \mathbb{R}$ are viewed as one-dimensional vector fields, what does it mean for $\{\sigma_0, \sigma_1\}$ to satisfy Hörmander's condition?

Next, for vector fields b, σ_j, $1 \le j \le d$, we define the operator

$$\mathcal{L} = \sum_{i=1}^n b_i \partial_i + \frac{1}{2} \sum_{i,j=1}^n (\sigma\sigma^T)_{ij} \partial_{ij}.$$

As we saw in Setion 3.3, the operator $\mathcal{L}$ is related to the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t. \tag{9}$$

Next, we define σ_0 to be the unique vector field such that (9) is equivalent to the Stratonovich SDE

$$dX_t = \sigma_0(X_t)dt + \sigma(X_t) \circ dB_t. \tag{10}$$

It is a good exercise to find a formula for σ_0 in terms of b and σ, but not important for our understanding of the theorem.

To summarize, we started with vector fields b, σ_j, $1 \le j \le d$, and defined from them an operator $\mathcal{L}$, as well as a new vector field σ_0. These objects are related by the connection between $\mathcal{L}$ and the Itô SDE (9), and the equivalence between the Itô SDE (9) and the Stratonovich SDE (10). With all this in place, we can now state Hörmander's theorem.

Theorem 5.2. *(Hörmander's Theorem) Suppose that b and σ_j, $1 \le j \le d$ are smooth vector fields with bounded derivatives of all orders, and that $\{\sigma_j\}_{0 \le j \le d}$ satifies Hörmander's condition. Then the Cauchy problem (7) has a fundamental soultion $p = p(t, x, y)$ such that $p(t, \cdot, \cdot)$ is smooth for each t.*

There is also a probabilistic statement of Hörmander's Theorem.

Theorem 5.3. *(Probabilistic Hörmander's Theorem) Suppose that $\sigma_0, ..., \sigma_d$ are smooth vector fields with bounded derivatives of all orders, and that $\{\sigma_j\}_{0 \le j \le d}$ satisfy Hörmander's condition. Let X_t^x be the unique solution to the Stratonovich SDE (10). Then for each t and x, X_t^x has a smooth density*

Let X_t^x denote the unique solution to (9) with $X_0 = x$. Note that by Proposition 3.13, if X_t^x has a density $g_{x,t}$, then the solution u to (7) is given by

$$u(x,t) = \mathbb{E}[f(X_t^x)] = \int_{\mathbb{R}^n} f(y) g_{x,t}(y) dy.$$

Thus, if a smooth fundamental solution exists, we expect X_t^x to have a smooth density for each x and t, and $p(t,x,y) = g_{x,t}(y)$, where $g_{x,t}$ is the density of X_t^x. Thus it is not surprising that Theorem 5.1 and 5.2 are in fact equivalent. It is Theorem 5.2 that can be approached via Malliavin calculus. In fact, by Proposition 3.10 and Theorem 4.9, it suffices to show that under Hörmander's condition, the random vector X_t is non-degenerate for each t. We won't get to a proof of this fact, but hopefully it is at least clear how Malliavin calculus can help. In the next (and final) section, we will explain why Hörmander's condition is natural.

5.2 Intuition about Hörmander's Condition

In this section, we try to build some intuition about Hörmander's Condition. At this point, you are probably wondering - why the Lie brackets? Often when Lie brackets come up, the Frobenius integrability theorem is involved. We recall the notion of a smooth distribution (on $\mathbb{R}^n$).

Definition 5.4. A **distribution** on $\mathbb{R}^n$ of order k is a smooth map V, written $x \mapsto V_x$, which assigns to each $x \in \mathbb{R}^n$ a dimension k linear subspace of $\mathbb{R}^n$.

The right notion of smoothness is a bit subtle, but hopefully the idea is clear. Here are a few more definitions about distributions:

Definition 5.5. A distribution V is **involutive** if whenever $u, v : \mathbb{R}^n \to \mathbb{R}^n$ are vector fields with $u(x), v(x) \in V_x$ for all x, we have $[u, v](x) \in V_x$ for all x.

Definition 5.6. A distribution V is **integrable** if in a neighborhood of each point, there are coordinates $y_1, ..., y_n$ such that $V_x = span(\frac{\partial}{\partial y_1}, ..., \frac{\partial}{\partial y_k})$.

Now we state (a restricted version of) the Frobenius Integrability Theorem.

Theorem 5.7. *(Frobenius Integrability Theorem) A distibution V is integrable if and only if it is involutive.*

This theorem shows one way in which the Lie bracket (which at first seems to be a purely algebraic gadget) encodes geometric information.

Next, we state the Stroock-Varadhan Support Theorem, which helps explain why Hörmander's condition is more naturally stated in terms of Stratonovich SDEs. Let X denote the unique solution to the Stratonovich SDE

$$\begin{cases} dX_t = \sigma_0(X_t)dt + \sigma(X_t) \circ dB_t, \\ X_0 = x_0. \end{cases} \tag{11}$$

We can view X as a measurable map $X : \Omega \to C([0,T]; \mathbb{R}^n)$. The Stroock-Varadhan Support Theorem gives a precise and geometrically natural characterization of the support of the law of X. We define the (multi-dimensional) Cameron-Martin space $\mathcal{C} = \{g \in C(\mathbb{R}_+; \mathbb{R}^d) : g(t) = \int_0^t \dot{g}(s)ds, \dot{g} \in H^d\}$.

Theorem 5.8. *(Stroock-Varadhan Support Theorem) Let X be the solution to* (11)*. For $g \in \mathcal{C}$, let $S(g) = x \in C([0,T];\mathbb{R}^n)$, where x is the solution to the ODE*

$$\begin{cases} dx(t) = \sigma_0(x(t)) + \sigma(x(t))\dot{g}(t)dt \\ x(0) = x_0 \end{cases}$$

Then for any $0 \leq \alpha < \frac{1}{2}$, the support of the measure $\mathbb{P} \circ X^{-1}$ is the closure in C^α of $\{S(g) : g \in \mathcal{C}\}$.

So, we replace the Brownian motion in (11) with certain deterministic paths, solve all the corresponding ODEs, and then take the closure in an appropriate norm. In particular, this can be used to show that if there is a submanifold M of $\mathbb{R}^n$ such that $x_0 \in M$ and $span(\{\sigma_i(x) : 0 \leq i \leq d\}) \subset T_x M$ for each $x \in M$, then we must have $X_t \in M$ a.s., for each t.

We are now ready to understand why Hörmander's condition is natural geometrically. Let $\{\sigma_i\}$ be as before, and define vector fields on $\mathbb{R}^{n+1}$ by

$$\tilde{\sigma}_0(x,t) = \begin{bmatrix} \sigma_0(x) \\ 1 \end{bmatrix}, \tilde{\sigma}_i(x,t) = \begin{bmatrix} \sigma_i(x) \\ 0 \end{bmatrix}, \quad 0 < i \leq d.$$

Then let

$$\tilde{V}_0 = \{\sigma_i : 0 \leq i \leq d\},$$
$$\tilde{V}_k = \{[v, \sigma_i] : v \in V_{k-1}, 0 \leq i \leq d\} \text{ for } 1 \leq k \leq d,$$

and define $\tilde{V} = \cup_k \tilde{V}_k$. So the $\tilde{V}_k$'s are formed from the $\tilde{\sigma}_i$'s in the same way that the V_k's are formed from the σ_k's, except that $\sigma_0 \in V_0$. Then one can show that Hörmander's condition is equivalent to having $\{v(x,t) : v \in \tilde{V}\}$ span $\mathbb{R}^{n+1}$ for all (x,t). Suppose this condition fails. In fact, for simplicity, suppose that it fails everywhere, and that the map $(x,t) \mapsto \tilde{V}_{(x,t)} := span(\{v(x,t) : v \in \tilde{V})$ defines a dimension k distribution for some $k < n+1$. One can show that $(x,t) \mapsto \tilde{V}_{(x,t)}$ has to be involutive, and hence integrable by Frobenius.

Thus there is a dimension k submanifold of $\mathbb{R}^{n+1}$ containing $(0, x_0)$, and such that $\tilde{V}_{(x,t)} = T_{(x,t)}\tilde{M}$. In particular, $\tilde{\sigma}_i(x,t) \in T_{(x,t)}\tilde{M}$ for all i and for all $(x,t) \in M$. But the process is $(X_t, t)_t$ is the unique solution to the SDE

$$d\tilde{X}_t = \tilde{\sigma}_0(\tilde{X}_t)dt + \tilde{\sigma_0}(\tilde{X}_t) \circ dB_t$$
$$X_0 = (x_0, 0)$$

Thus by the Strook-Varadhan support theorem, we see that $(X_t, t) \in \tilde{M}$ a.s., at least until (X_t, t) exits some neighborhood U of $(x_0, 0)$. Choosing some small t_0, a transversality argument shows that $M := \tilde{M} \cap \{t_0\}$ is a dimension $k-1$ submanifold of $\mathbb{R}^n$, and clearly we must have $X_{t_0} \in M$ a.s. on the event that $(X_{t_0}, t_0) \in U$. In particular, as long as t_0 is small enough, we have that $\mathbb{P}[X_{t_0} \in M] > 0$. But as a submanifold of positive co-dimension, M has Lebesgue measure 0 in $\mathbb{R}^n$, so this shows that X_{t_0} is not a.c. with respcet to the Lebesgue measure. Hopefully this is a convincing heuristic argument that Hörmander's condition is a geometrically natural generalization of ellipticity.