THE MALLIAVIN CALCULUS

Han Zhang

Supervisor: Professor Tony Dooley

School of Mathematics,
The University of New South Wales.

December 2004

Submitted in partial fulfillment of the requirements of the degree of
Bachelor of Science with Honours

Acknowledgements

I enjoyed the help and support from numerous friends and faculty members during whilst writing this thesis. My greatest debt however, goes to my supervisor, Professor Tony Dooley, who initiated me to the fantastic field of Malliavin calculus. With his incredible breadth and depth of knowledge and intuition, he guided me to structure and clarify my thought and suggested valuable insightful comments. This thesis would not have been possible to write without his help and support.

I am also indebted to express my sincere thankfulness to the honours coordinators, Dr Ian Doust and Dr Brian Jefferies, who were very kind and supportive to us during the year. Dr Ben Goldys also deserves a special mention, for the recommendation of a number of useful references, and I very much enjoyed the casual discussions I had with him.

Finally, I would like to thank everyone in the School of Mathematics, and the Department of Actuarial Studies for providing me with such a wonderful four years at UNSW.

Han Zhang, November 2004.

History and Introduction

The Malliavin calculus, also known as the stochastic calculus of variations, is an infinite dimensional differential calculus on the Wiener space. Much of the theory builds on from Itô's stochastic calculus, and aims to investigate the structure and also regularity laws of spaces of Wiener functionals. First initiated in 1974, Malliavin used it in [31] to give a probabilistic proof of the Hörmander's theorem and its importance was immediately recognized.

It has been believed up until near the end of the 19th century that a continuous function ought to be smooth at "most points". The only sort of non-differentiable incidents that were the isolated sharp corners between two pieces of smooth curves, whose behaviour is similar to the graph of $f(x) = |x|$ around $x = 0$. It was not until 1861, when the German mathematician K. Weierstrass first gave an example of a function that was continuous but nowhere differentiable on $\mathbb{R}$:

$$f(x) = \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k \cos(3^k \pi x)$$

This was a striking phenomenon at the time, as it signals that there could be a new class of continuous functions that are essentially not governed by almost all of the calculus developed at the time. For example, the integration by substitution formula for $\int g df$ breaks down completely if f is nowhere differentiable.

This type of wild sharp oscillation is not entirely abstract nonsense. In fact, a path which can be modelled by a continuous nowhere differentiable function was observed in real life 50 years before Weierstrass' example, by the English botanist Robert Brown while observing movements of pollen particles under the microscope. This was known as the Brownian motion.

In the early 20th century, many physicists including A. Einstein expressed great interest in modelling quantum particle movements with Brownian motion. Einstein's paper in 1905 was considered by many as the first breakthrough in giving a mathematical model to the Brownian motion. In 1923, an American mathematician Norbert Wiener gave a mathematically rigorous definition (in a measure theoretic sense) to Brownian motion based on the idea of independent increments. An interesting fact to note here, is that Wiener's work had appeared before Kolmogorov formalized the theory of probability which occurred in 1931.

However, a problem that remained to be unsolved for another thirty years was how one could make sense of

$$\int f dW,$$

where W is the Wiener process. Almost all of the results known at the time suggested it was impossible. Essentially, there is no hope of constructing a Lebesgue-Stieljes type integral of the form $\int f\ dg$ if g is of infinite variation - the Brownian paths have this property.

Between 1942 and 1951, a Japanese mathematician K. Itô was able to give a reasonable definition of such an integral in probabilistic terms (as opposed to pathwise). Further, he showed how change of variables were made via a lemma, now known as Itô's lemma. It essentially states that if f is a twice differentiable function in x and t and $X_t = \sigma W_t$ where W_t is a standard Wiener process, then,

$$df = \frac{\partial f}{\partial X_t} dX_t + \frac{\partial f}{\partial t} dt + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial X_t^2} dt.$$

We see that this formula serves a similar purpose to the chain rule in classical calculus, but with an extra "correction term" that can be roughly understood as something to account for the non-zero quadratic variation of Wiener paths. Itô discovered the above only as a lemma, while his ultimate goal at the time was to prove a martingale representation theorem: If M_t is a martingale with bounded quadratic variation, then there exists a square integrable adapted process f_s, such that

$$M_t = \int_0^t f_s dW_s.$$

Itô's work really opened the gate to a new world of stochastic analysis. In particular, people began to realize that there were tools available in stochastic calculus that can be used to solve problems in deterministic calculus. The Feynman-Kac formula, first appeared in 1947, and rigorously proven in 1965, was perhaps the highlight such example, where an initial problem involving partial differential equations was solved by solving a corresponding stochastic differential equation. The Itô calculus found its immediate applications in diffusion theory and quantum mechanics, and later in mathematical finance.

One question of particular interest at this stage is to ask for an explicit expression for f_s in the martingale representation. An immediate reaction at this point, taking into account of the fundamental theorem of calculus, is that the f_s term should correspond to a differentiation type of operation in the probabilistic setting. It turns out that we have $f_s = \mathbb{E}(D_t M_t | \mathcal{F}_s)$, where D_t is the Malliavin derivative. This is called Clark's representation.

Malliavin's initial intentions in developing his calculus really had very little to do with Clark's representation. He was working to give sufficient conditions to which a random variable possesses a smooth probability density. He showed that this could be done if a certain matrix involving Malliavin derivatives was invertible and its inverse is integrable in L^p for all $p \geq 1$, almost surely. Based on this, and exploiting the connections between SDEs and PDEs, he was able to give a probabilistic proof of Hörmander's theorem. There has been an extensive amount of work done to generalize Malliavin's ideas for giving regularity conditions of stochastic partial differential eqautions (SPDE). In 1982, 1984 and 1987, Stroock, Bismut and Bells respectively have demonstrated three different ways that the Malliavin calculus could be approached from.

In 1999, the Malliavin calculus found itself yet another playground in the field of mathematical finance. It is often of interest to investors to know the sensitivity of the underlying stock price with respect to various parameters. Obviously, this involves taking derivatives. These sensitivity measures are called Greeks, as they are traditionally denoted by Greek letters. They are extremely difficult to calculate even numerically. The main problem is that the derivative term needs to be approximated using the finite difference method and such approximations can become very rough. The integration by parts formula obtained from Malliavin calculus can transform a derivative into an weighted integral of random variables. This gives a much accurate and fast converging numerical solution than obtained from the classical method.

My thesis will be written in six chapters. Chapter 1 briefly refreshes the theory of functions of bounded variations, and also some basic definitions of random variables and stochastic processes.

Chapter 2 will establish the Itô integral, Itô's lemma and Itô's martingale representation theorem. This chapter aims to set a firm foundation for the development of Malliavin calculus.

Chapter 3 begins with an illustration of the chaos decomposition theorem. Then, it develops the Malliavin calculus and links it back to the chaos decomposition to establish some fascinating results. A common aim of chapters 2 and 3 is to demonstrate precisely how the classical deterministic calculus fails to extend to the infinite dimensional setting, and how the probabilistic calculus fixes these problems.

Chapter 4 provides an introduction to the first application of Malliavin calculus, we give the sufficient conditions to which the probability density of a given random variable is smooth.

Chapter 5 begins by briefly sketching through the basic theory of stochastic differential equations and stochastic flows, and their relations with partial differential equations. In particular, it demonstrates how Malliavin calculus can be mixed with these ideas to give a probabilistic proof of Hörmander's theorem.

Chapter 6 concludes the thesis by illustrating a very recent development in the area of mathematical finance, whereby Malliavin calculus is used to give stable Monte Carlo simulation algorithms.

Contents

Chapter 1

Tools From Analysis

The sole purpose of the first chapter is to introduce and revise the main ideas from analysis that are required to understand and appreciate rest of the thesis. Section 1.1 on functions of bounded variations is essentially a summary of the key concepts from [17]. Where we begin by exploring some quite general conditions for the existence of a derivative and analogues of fundamental theorem of calculus under Lebesgue's definition of integration. Section 1.2 will introduce the basics of probability and random variables from a measure theoretic point of view.

1.1 Functions of Bounded Variation

The set of functions with bounded variation are particularly nice, in the sense that most of the classical calculus operations such as differentiation are problem free. A large part of this thesis attempts to resurrect the situation in cases when we are dealing with functions or paths of unbounded variation, such as the trajectories of a Brownian motion. Perhaps it would be appropriate to first explore the case when function has bounded variation, so that we can really appreciate the efforts spent in studying the unbounded variation case.

Definition 1.1.1. Let $\mathcal{P}_n[a,b] = (x_0, x_1, ..., x_n)$ such that $a = x_0 < x_1 < ... < x_n = b$. Define δ_n to be the mesh of $\mathcal{P}_n$ by

$$\delta_n = \sup_k |x_k - x_{k-1}|.$$

Definition 1.1.2. Given a function $f : [a,b] \to \mathbb{C}$, we define the **total variation over** the interval $[a,b]$ as

$$\langle f \rangle_1^{a,b} = \lim_{n\to\infty,\delta_n\to 0} \sum_{k=1}^{n} |x(t_k) - x(t_{k-1})|.$$

Moreover, if $\langle f \rangle_1^{a,b} < \infty$, we say $f \in BV_{a,b}$, where $BV_{a,b}$ denotes the set of functions that has **bounded variation** over the interval $[a,b]$.

Heuristically, one could think of $\langle f\rangle_1^{a,b}$ as the total amount of vertical oscillation of f throughout the interval $[a, b]$. Therefore, it should be intuitive that for monotone functions f, then $\langle f\rangle_1^{a,b} = |f(a) - f(b)|$, which also implies that $f \in BV_{a,b}$. Moreover, if one could find a finite partition of $[a, b]$, such that f is monotone on each of the partitions, then $f \in BV_{a,b}$. However, perhaps against our intuition but the converse is false: consider the following example,

Example 1.1.3. Let $q_1, q_2, q_3, ...$ be an ordering of the rational numbers in $(0, 1)$, and let $0 < a < 1$. Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} a^k, & \text{if } x = q_k; \\ 0, & \text{otherwise.} \end{cases}$$

Consider any sequence of partitions $\mathcal{P}_n = (0 = x_0, x_1, ..., x_n = 1)$ of $(0, 1)$... and the contribution to total variation of each rational point is at most $2a^k$,

$$\langle f\rangle_1^{a,b} \leq \sum_{k=1}^{\infty} 2a^k = \frac{2a}{1-a} < \infty$$

Therefore, we see that $f \in BV_{0,1}$, while it's clearly impossible to partition $[0, 1]$ into subintervals I_j such that for each j, f is monotone on I_j. □

Theorem 1.1.4. ***(Jordan Decomposition)*** *Let $f \in BV_{a,b}$, then there exist non-decreasing functions g and h such that $f = g - h$.*

Proof. Define

$$\begin{aligned} g(x) &= \frac{1}{2}(\langle f\rangle_1^{a,x} + f(x)) \\ h(x) &= \frac{1}{2}(\langle f\rangle_1^{a,x} - f(x)), \end{aligned}$$

then, $f(x) = g(x) - h(x)$. Thus we need to check that g and h are increasing. Let $\alpha, \beta \in [a, b]$ with $\alpha < \beta$, then

$$\begin{aligned} g(\beta) - g(\alpha) &= \frac{1}{2}(((\langle f\rangle_1^{a,\beta} - \langle f\rangle_1^{a,\alpha}) + f(\beta) - f(\alpha)) \\ &\geq \frac{1}{2}(\langle f\rangle_1^{\alpha,\beta} - |f(\beta) - f(\alpha)|) \\ &\geq 0. \end{aligned}$$

A similar argument shows that h is also increasing, and hence we have constructed two increasing functions whereby f is their difference. □

Perhaps the most powerful consequence of Jordan's decomposition theorem is that it conveniently allows us to generalize a vast number of results that hold true for monotone functions to functions of bounded variation. The following theorems plays a vital role in the theory of differentiation.

Theorem 1.1.5. ***(Lebesgue's Theorem)*** *Let $f : [a,b] \to \mathbb{R}$ be a monotone function. Then $f'(x)$ exists and finite a.e. in (a,b).*

Theorem 1.1.6. ***(Fubini's Theorem)*** *Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of monotone functions such that $\sum_{n=1}^{\infty} f_n(x) = f(x)$. Then, $f'(x) = \sum_{n=1}^{\infty} f_n'(x)$ a.e. in (a,b).*

Corollary 1.1.7. *The statement of Theorem 1.2.4 and Theorem 1.2.5 holds true for functions of bounded variation.*

Theorem 1.1.8. ***(Stieljes Integral)*** *Let f and g be continuous functions defined on $[a,b]$, and assume that g has bounded variation. Then,*

$$\lim_{n\to\infty \delta_n \to 0} \sum_{k=0}^{n} f(x_k^*)(g(x_k) - g(x_{k-1})) = \int_a^b f dg$$

exists and agrees for all $x_k^* \in [x_{k-1}, x_k]$.

Proof. For each n, let $\mathcal{P}_n[a,b]$ be a partition of $[a,b]$. Let

$$m_k = \inf\{f(x) : x \in [x_{k-1}, x_k]\}$$

$$M_k = \inf\{f(x) : x \in [x_{k-1}, x_k]\}$$

and their corresponding sums,

$$s_n = \sum_{k=1}^{n} m_k(g(x_k) - g(x_{k-1}))$$

$$S_n = \sum_{k=1}^{n} M_k(g(x_k) - g(x_{k-1})).$$

Now consider,

$$\begin{aligned}
S - s &= \lim_{n\to\infty, \delta_n \to 0} \sum_{k=0}^{n} M_k(g(t_k) - g(t_{k-1})) - m_k(g(t_k) - g(t_{k-1})) \\
&= \lim_{n\to\infty, \delta_n \to 0} \sum_{k=0}^{n} (M_k - m_k)(g(t_k) - g(t_{k-1})) \\
&\le \sup_k |M_k - m_k| \langle g \rangle_1^{a,b}
\end{aligned}$$

By assumption, $\langle g\rangle_1^{a,b} < \infty$ and $\sup_k |M_k - m_k| \to 0$ as f is assumed to be continuous. Hence, we have shown that both sums must agree in the limit. $\square$

Definition 1.1.9. Let $f : [a,b] \to \mathbb{C}$. Suppose that for every $\varepsilon > 0$, there is a $\delta > 0$ such that,

- $\sum_{k=1}^{n} |f(d_k) - f(c_k)| < \varepsilon$
- For every finite, pairwise disjoint, family $\{(c_k, d_k)\}_{k=1}^{n}$ of open subintervals of $[a,b]$ for which $\sum_{k=1}^{n} |d_k - c_k| < \delta$.

Then, f is said to be an **absolutely continuous function** on $[a,b]$.

Theorem 1.1.10. ***(The Fundamental Theorem of Calculus)*** *Let $f : [a,b] \to \mathbb{R}$ be absolutely continuous. Then,*

1. *For every $x \in (a,b)$, there exists a function $f'(t)$ such that*

$$f(x) = f(a) + \int_a^x f'(t)dt.$$

2. *f has bounded variation, and its total variation is given by*

$$\langle f\rangle_1^{a,x} = \int_a^s |f'(t)|dt.$$

Remark 1.1.11. A key point to the concept of absolutely continuity is that there exists continuous functions f such that its derivative $f' = 0$ a.e., yet f is strictly increasing. These functions are called **singular functions** and we wish to avoid them. The first example of this kind was given by Cantor. Interested readers may consult chapter 5 of [17] and chapter 8 of [25]. $\square$

Definition 1.1.12. Let μ and ν be measures on a measurable space $(\Omega, \mathcal{F})$, such that for all $A \in \mathcal{F}$, $\mu(A) = 0 \iff \nu(A) = 0$. Then, μ is said to be an **absolutely continuous measure** with respect to ν. On the other hand, if $\mu(A) > 0 \iff \nu(A) = 0$, then we say μ is a **singular measure** with respect to ν.

Theorem 1.1.13. ***(Lebesgue Decomposition)*** *Let μ be a measure on a measurable space $(\Omega, \mathcal{F})$. Then, for a given measure ν, μ can be uniquely decomposed as*

$$\mu = \xi + \eta$$

where ξ is absolutely continuous, and η is singular with respect to ν.

1.2 Random Variables and Stochastic Processes

This section is intended to give some basic definitions to random variables and stochastic processes, so this thesis would be more self-contained. Readers can feel free to skip to the next chapter.

Definition 1.2.1. Let T be a set, $(\Omega, \mathcal{F}, \mathbb{P})$ a measure space and $(E, \mathcal{E})$ a probability space. A mapping $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E}), t \in T$ is called a **random variable** X is $(\Omega, \mathcal{F}) - (E, \mathcal{E})$ measurable.

Definition 1.2.2. Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

1. The **expectation** of X, denoted by $\mathbb{E}X$ is defined by,

$$\int_\Omega x d\mathbb{P}(x).$$

2. The **law** of X is a function $F : \mathbb{R} \to [0, 1]$ defined by $F_X(x) = \mathbb{P}(X \leq x)$. If F is an absolutely continuous function, we call $f_X(x) = \partial_x F_X(x)$ as the **probability density function** of X.

Definition 1.2.3. Let $\{X_n\}$ be a sequence of random variables. We say

1. $X_n \xrightarrow{a.s.} X \iff \mathbb{P}(\omega \in \Omega : X_n(\omega) \to X(\omega)) = 1.$
2. $X_n \xrightarrow{L^p} X \iff \mathbb{E}(|X_n - X|^p) \to 0$ for $p \geq 1$.
3. $X_n \xrightarrow{\mathbb{P}} X \iff \mathbb{P}(|X_n - X| > \varepsilon) \to 0$ for all $\varepsilon > 0$.
4. $X_n \xrightarrow{d} X \iff F_{X_n}(x) \to F_X(x)$ pointwise.

Definition 1.2.4. Let T be a set. A **stochastic process** $X_t, t \in T$, defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a family of random variables. For every $\omega \in \Omega$, the mapping $t \to X_t(\omega)$ is called the trajectory of X_t.

Definition 1.2.5. A **filtration** $(\mathcal{F}_t, t \in \mathbb{R})$ is defined to be a collection of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s < t$. Further, if $\mathcal{F}_t$ satisfies,

1. $\mathcal{F}_t = \cap_{s>t} \mathcal{F}_s$. This is called the continuity criterion.
2. $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets.

We say that $\mathcal{F}_t$ is a standard filtration.

Definition 1.2.6. A filtered probability space is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathcal{F}_t$, denoted by $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$.

Definition 1.2.7. A stochastic process X_t is said to be **adapted** to a filtration $\mathcal{F}_t$ if X_t is $\mathcal{F}_t$ measurable. It is said predictable if the map $(t, \omega) \to X_t(\omega)$ is measurable with respect to the predictable σ-algebra $\mathcal{P}$.

Definition 1.2.8. The **predictable σ-algebra** $\mathcal{P}$ is the σ-algebra on $\mathbb{R} \times \Omega$ generated by the process X_t, adapted to $\mathcal{F}_t$, with left continuous paths.

Chapter 2

Itô Calculus and Martingales

I would like to use this chapter to give an overview of Brownian motions, martingales and Itô stochastic calculus. It begins with a detailed construction of the Itô's integral with respect to an abstract martingale. We will then move onto Itô's change of variables formula and also give a number of applications, including the martingale representation theorem. It is intended to develop enough theory to talk about Malliavin calculus in chapter three. Thus, a number of closely related topics such as the reflection principle are regretfully left out. Good references to this chapter are [10], [40], [38] and [28], all of which provides a reasonably full coverage.

2.1 Brownian Motion

Between 1827 and 1829, an English botanist, Robert Brown, discovered that the movements of pollen particles under the microscope underwent extremely wild oscillations. This became later known as Brownian motion. He first hypothesised that the wild movements was related biologically to pollen particles themselves, but later he realised that other inorganic particles also exhibited the same type of motion. Today, the best accepted explanation of such a motion is caused by extremely frequent bombardments by neighbouring particles. We first look at a heuristic derivation of how such a motion may evolve in time before giving the formal definition of the Wiener process, a mathematical object that is used to model the Brownian motion.

Consider the physical movement of such a particle on $\mathbb{R}$. In every Δt units of time, a particle is bombarded from either left or from the right with probability 0.5, and the particle moves Δd units to the opposite direction after each bombardment. Let $\psi_t(x)$ be the probability distribution of the position of particle at time t. From the above physical reasoning, we have

$$\psi_t(x) = \frac{1}{2}\psi_{t-\Delta t}(x - \Delta x) + \frac{1}{2}\psi_{t-\Delta t}(x + \Delta x).$$

Subtracting $\psi_{t-\Delta t}(x)$ from both sides to give

$$\frac{1}{\Delta t}(\psi_t(x) - \psi_{t-\Delta t}(x)) = \frac{1}{\Delta t}\left(\frac{1}{2}\psi_{t-\Delta t}(x - \Delta x) - \psi_{t-\Delta t}(x) + \frac{1}{2}\psi_{t-\Delta t}(x + \Delta x)\right).$$

Since the bombardments occurs in extremely small intervals, it makes sense to consider the limit $\Delta t \to 0$ and $\Delta x \to 0$. However, this needs to be done with care, as blindly letting $\Delta t = \Delta x \to 0$ would imply that

$$\frac{\partial \psi}{\partial t} = \frac{1}{2}\left(\frac{\partial \psi}{\partial x} - \frac{\partial \psi}{\partial x}\right) = 0,$$

which means that the displacement is constant, i.e. there is no motion! However, if we let $\Delta x = \sqrt{\Delta t} \to 0$, we would obtain,

$$\frac{\partial \psi}{\partial t} = \frac{1}{2}\frac{\partial^2 \psi}{\partial x^2}$$

along with the initial condition that

$$\psi_0(x) = \delta(x)$$

where δ is the Dirac δ-function. Einstein first formulated this model in 1905, with a number of additional constant terms each with its own physical interpretation. This initial value problem is equivalent to the heat equation studied by Fourier nearly a century earlier. It can be solved via a Fourier transform over the direction of x. Let

$$\hat{\psi}_t(s) = \int_{\mathbb{R}} e^{-ixs}\psi_t(x)dx,$$

then our initial value problem becomes,

$$\frac{\partial \hat{\psi}_t(s)}{\partial t} = -\frac{1}{2}s^2\hat{\psi}_t(s)$$
$$\hat{\psi}_0(s) = 1$$

which can be solved to obtain

$$\hat{\psi}_t(s) = e^{-1/2s^2t}.$$

Hence, taking the inverse transform (or by inspection), we obtain the probability density

$$\psi_t(x) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}.$$

Thus, we see Brownian motions are characterised by independent and identical increments that are normally distributed with mean zero and variance t, where t is the amount of time elapsed. We are now in position to give some formal definitions.

Definition 2.1.1. (Wiener Process) Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space, and W_t be a stochastic process. W_t is called a Wiener process with respect to $\mathcal{F}_t$ if

1. $W_0 = 0$ a.s.
2. W_t is $\mathcal{F}_t$-measurable for every t.
3. $\mathbb{P}(\omega \in \Omega : t \to W_t(\omega)$ is a continuous function in $t) = 1$.
4. $W_t - W_s$ is independent to $\mathcal{F}_s$ for all $t > s$ and $W_t - W_s \sim N(0, t-s)$

The standard Brownian motion (starting at 0) satisfies the axioms of a Wiener process. It will be shown later this chapter that it is in fact the unique stochastic process that fulfils properties 1 - 4.

Remark 2.1.2. Strictly speaking, the Brownian motion and Wiener processes are two different things. While the former refers to the physical motions of small particles, the latter is a mathematical model of the former in an idealistic situation. Some properties of Wiener paths (as we will see) such as almost surely nowhere differentiability is in fact false for the physical Brownian motion, since it is impossible for a particle to be bombarded by its neighbours continuumly often. □

A more formal way of viewing the Wiener process is as a stochastic process taking values over the set of all possible trajectories. Let $C_\alpha[a, b]$ be the set of continuous functions f defined on $[a, b]$ with $f(a) = \alpha$. Let $\Omega = C_0[0, T]$ and $\mathcal{F} = \mathcal{B}(C_0[0, b])$, where $\mathcal{B}(C_0[0, b])$ is the σ-algebra generated by open sets (with respect to the sup metric) of $C_0[0, T]$. The filtration $\mathcal{F}_t$ in this case would be a sequence of sub-σ-algebras. In 1923, Wiener showed that there is a well defined measure μ on this measure space, known as the Wiener measure. Elements of $C_0[0, T]$ under the Wiener measure corresponds to the sample paths of the Brownian motion, and the probability space $(C_0[0, T], \mathcal{F}, \mu)$ is called the **classical Wiener space**. Readers are referred to [24] for details in the construction.

The following are some well known properties of the Brownian motion taken from [24] that are relevant to our development.

1. The Brownian paths are almost surely of unbounded variation on arbitrarily small time intervals.
2. As a consequence, the Brownian paths, with probability one (with respect to the Wiener measure), are nowhere differentiable at any point.

Proposition 2.1.3. *Let X be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} \subseteq \mathcal{F}$. Then, there exists a unique random variable Y such that*

1. *Y is $\mathcal{G}$-measurable*
2. *$\mathbb{E}(1_A Y) = \mathbb{E}(1_A X)$ for every non-empty set $A \in \mathcal{F}$.*

Proof. This is an easy consequence of the Radon-Nikodym theorem. □

Definition 2.1.4. (Conditional Expectation) The random variable Y in proposition 2.1.10 is defined to be the expectation of X conditioned on $\mathcal{G}$, denoted by $\mathbb{E}(X|\mathcal{G})$.

Properties of the conditional expectation: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ be a σ-subalgebra of $\mathcal{F}$. Then,

1. $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\forall H \in \mathcal{F}, \mathbb{E}(\mathbb{E}(X|\mathcal{G})|H) = \mathbb{E}(X|H)$.

Further properties of the conditional expectation can be obtained in chapter 1 of [10] and [40].

Example 2.1.5. Let W_t be a Wiener process, and $\mathcal{F}_t$ be the filtration generated by W_t. For $s < t$, we wish to calculate $\mathbb{E}(W_t|\mathcal{F}_s)$. Observe that W_s is $\mathcal{F}_s$-measurable, and for all $A \in \mathcal{F}_s$,

$$\mathbb{E}(1_A W_t) = \mathbb{E}(1_A W_s) + \mathbb{E}(1_A(W_t - W_s)).$$

By independence, $\mathbb{E}(1_A(W_t - W_s)) = \mathbb{E}(1_A)\mathbb{E}(W_t - W_s) = 0$, since $W_t - W_s$ is a normally distributed with 0 mean. Therefore, we have shown that $W_s = \mathbb{E}(W_t|\mathcal{F}_s)$. $\square$

We give the following definition of a martingale by means to generalize the result of the previous example.

Definition 2.1.6. Let $T \subset \mathbb{R}$ and $\mathcal{F}_t$ a filtration over Ω. A stochastic process X_t is said to be a **martingale** if

1. X_t is $\mathcal{F}_t$-measurable for every t.
2. $\mathbb{E}|X_t| < \infty$ for every t.
3. $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$.

A direct consequence of the definition is that M is a martingale iff $M_t - M_s$ is independent to $\mathcal{F}_s$. Hence,

$$\mathbb{E}(M_{s_2} - M_{s_1})(M_{t_2} - M_{t_1}) = 0$$

if $(s_1, s_2) \cap (t_1, t_2) = \emptyset$. Secondly, we observe that $\mathbb{E}M_t = \mathbb{E}M_0$ for all martingales M.

The following is a list of elementary results related to martingales that will be useful later on. Proofs of these can be found in [10] and [28].

Definition 2.1.7. Stopping Time A random variable $\tau : \Omega \to [0, \infty)$ is a **stopping time** if $\{\omega : \tau(\omega) \leq t\}$ is $\mathcal{F}_t$-measurable.

Theorem 2.1.8. *Optional Stopping Time Theorem* *Let M_t be a martingale and let ν and τ be stopping times with respect to a common filtration $\mathcal{F}_t$. Then,*

$$\mathbb{E}(M_\tau|\mathcal{F}_\nu) = M_\nu$$

Definition 2.1.9. Given a stochastic process X and a stopping time τ, we define the **stopped process** by $X^\tau = X_{t\wedge\tau}$. This is a replication of X, and frozen at time T.

Definition 2.1.10. *M is a* **local martingale** *iff there exists a sequence of stopping times $\tau_n \to \infty$, such that M^{τ_n} are martingales for all n.*

Remark 2.1.11. The concept of a local martingale is of central importance to stochastic calculus, as we will see later that the stochastic integral of a local martingale will always be another local martingale. The same cannot be said for martingales when the integral is taken over an infinite horizon. □

Proposition 2.1.12. *Every bounded local martingale is a martingale.*

Proof. Let M be a bounded local martingale, so that $M^{T_n}(\omega) \to M(\omega)$ pointwise. We may apply the dominated convergence theorem to obtain,

$$\mathbb{E}(M_t|\mathcal{F}_s) = \lim_{n\to\infty} \mathbb{E}(M_t^{T_n}|\mathcal{F}_s) = \lim_{n\to\infty} X_s^{T_n} = X_s.$$

Hence, every bounded local martingale is also a martingale. □

However, I would like to stress that local martingales are much more general than martingales. A common misconception is to believe that local martingales only need to be integrable in order to be martingales. A counter example can be constructed with the aid of so called Itô's lemma, section 5.2 of [40] has the details.

Theorem 2.1.13. *If M is a continuous local martingale with finite first and second moment and has bounded variation, then M is constant almost surely.*

Proof. Without loss of generality, we can assume the constant to be $M_0 = 0$. Since M is assumed to have bounded variation, we can apply the fundamental theorem of calculus to get

$$M_t^2 = M_0^2 + 2\int_0^t M_s dM_s.$$

We can further write the integral as a Riemann sum since M is assumed to be continuous. Hence,

$$M_t^2 = M_0^2 + \lim_{n\to\infty} 2\sum_{i=1}^n M_{s_i}\Delta M_{s_i}.$$

M was assumed to have first and second moments, applying dominated convergence theorem and exploiting independent increments of martingales, we have

$$\mathbb{E}M_t^2 = M_0^2 + \lim_{n\to\infty} 2\sum_{i=1}^n \mathbb{E}M_{s_i}\Delta M_{s_i} = M_0^2 = 0.$$

Since $M_t^2 \geq 0$, we may conclude that $M_t = 0$ almost surely. □

This is an extremely important result for our purposes, as it demonstrates that the classical integration theory does not apply when we are integrating with respect to any interesting martingales. I will give an explicit example to further demonstrate this problem in the following section, and then we will discuss the possible ways of fixing the problem.

2.2 Construction of the Itô Integral

In this section, we formulate the construction of the Itô integral with respect to an abstract martingale over $[0, T]$. Remarks will be made concerning how this formulation may generalise to obtain an integral over $[0, \infty)$.

Definition 2.2.1. Let $\mathcal{P}_n = (t_0, t_1, ..., t_n) = \{0 = t_0 < t_1 < ... < t_n = T\}$ be a partition of $[0, T]$, we define $\Delta t = t_k - t_{k-1}$ and $\Delta X_{t_k} = X_{t_k} - X_{t_{k-1}}$, and the mesh,

$$\delta_n = \sup_k |t_k - t_{k-1}|$$

for $k = 1, ..., n$ and $t_k \in \mathcal{P}_n$.

Definition 2.2.2. Given a function $f : [a, b] \to \mathbb{R}$. We define the **quadratic variation** over the interval $[a, b]$ as

$$\langle f \rangle_2^{a,b} = \lim_{\delta_n \to 0} \sum_{k=1}^{n} |\Delta f(t_k)|^2.$$

Example 2.2.3. Let $W(\omega)$ be a trajectory of the Wiener process. Then,

$$\langle W \rangle_2^{a,b}(\omega) = \lim_{n \to \infty, \delta_n \to 0} \sum_{k=1}^{n} (\Delta W_{t_k})^2.$$

To compute the quadratic variation directly (path-wise) seems to be a difficult task! Instead, we shall take a probabilistic approach which will the type of proving technique we will be using for most of this chapter. Consider,

$$\begin{aligned}
\mathbb{E}\Big(\sum_{k=1}^{n}(\Delta W_{t_k})^2 - (b-a)\Big)^2 &= \mathbb{E}\Big(\sum_{k=1}^{n}(\Delta W_{t_k})^2 - (\Delta t_k)\Big)^2 \\
&= \sum_{k=1}^{n} \mathbb{E}(\Delta W_{t_k})^4 + 2(\Delta t_k)\mathbb{E}(\Delta W_{t_k})^2 + (\Delta t_k)^2 \\
&= \sum_{k=1}^{n} (3(\Delta t_k)^2 - 2(\Delta t_k)^2 + (\Delta t_k)^2 \\
&\leq 2t \max(\Delta t_k) \xrightarrow{\delta_n \to 0} 0,
\end{aligned}$$

where the fourth line was obtained by noting that the kurtosis of a $N(0, \Delta t)$ random variable is $3(\Delta t)^2$. Therefore, we have shown that $V_{a,b}^2 W = b - a$ in L^2. $\square$

Example 2.2.4. The following example shows that the Wiener process is not integrable in the Stieljes sense. Consider,

$$\int_a^b W dW = \lim_{n\to\infty, \delta_n \to 0} \sum_{k=0}^{n} W(t_k^*)(W(t_k) - W(t_{k-1}))$$

Next, we evaluate the above using $t_k^* = t_k$ and t_{k-1} and call the respective limits S and S'. If $\int W dW$ exists (under Stieljes), then we would expect $S - S' = 0$. However,

$$\begin{aligned} S - S' &= \lim_{n\to\infty, \delta_n \to 0} \sum_{k=0}^{n} \Big(W(t_k)(W(t_k) - W(t_{k-1})) - W(t_{k-1})(W(t_k) - W(t_{k-1})) \Big) \\ &= \lim_{n\to\infty, \delta_n \to 0} \sum_{k=0}^{n} (W(t_k) - W(t_{k-1}))^2 \\ &= V_{a,b}^2 W \\ &= b - a \end{aligned}$$

$\square$

The previous example suggests that what is stopping us from defining the stochastic integral the "usual" way more or less caused by this additional quadratic variation term that functions of bounded variation did not have. This suggests that we should study the quadratic variation in more detail before we could give the definition of a stochastic integral.

Theorem 2.2.5. *For every continuous and bounded martingale M of finite quadratic variation,*

1. $\langle M \rangle_2^{0,t} = 0$ *a.s. when* $t = 0$.
2. $\langle M \rangle_2$ *is everywhere increasing.*
3. *The process* $M^2 - \langle M \rangle_2$ *is a martingale adapted to* $\mathcal{F}_t$.

Proof. Since $\langle M \rangle_2$ is required to be increasing, by Theorem 1.1.5, $\langle M \rangle_2$ has bounded variation. Hence, if there are two valid candidates A and B for such a process, by Theorem 2.1.12, the process $A - B = 0$ a.s. Thus we have proven uniqueness.

For a given subdivision of $[0, \infty)$, $\delta = \{t_0 = 0 < t_1 < ...\}$ such that only a finite number of t_i's in each closed interval $[0, t]$, we define

$$T_t^\delta(M) = \sum_{i=0}^{k-1} (M_{t_{i+1}} - M_{t_i})^2 + (M_t - M_{t_k})^2$$

where k is such that $t_k \leq t < t_{k+1}$. Further, observe that for $t_k < t < t_{k+1}$,

$$\begin{aligned}\mathbb{E}[(M_{t_{k+1}} - M_{t_k})^2|\mathcal{F}_s] &= \mathbb{E}[((M_{t_{k+1}} - M_t) + (M_t - M_{t_k}))^2|\mathcal{F}_s] \\ &= \mathbb{E}[(M_{t_{k+1}} - M_t)^2|\mathcal{F}_s] + (M_t - M_{t_k})^2\end{aligned}$$

Thus, for $t_j < s < t_{j+1}$,

$$\begin{aligned}\mathbb{E}[T_t^\delta(M) - T_s^\delta(M)|\mathcal{F}_s] &= \mathbb{E}\bigg(\sum_{i=j+1}^{k-1}(M_{t_{i+1}} - M_{t_i})^2 + (M_t - M_{t_k})^2 + (M_{t_{j+1}} - M_s)^2|\mathcal{F}_s\bigg) \\ &= \mathbb{E}[(M_t - M_s)^2|\mathcal{F}_s] \\ &= \mathbb{E}[M_t^2 - M_s^2|\mathcal{F}_s]\end{aligned}$$

where the second to last line and the last line is obtained by exploiting the independence of increments. Hence,

$$\mathbb{E}(M_t^2 - T_t^\delta(M))|\mathcal{F}_s) = \mathbb{E}(M_t^2 - (M_t^2 - M_s^2|\mathcal{F}_s) = M_s^2$$

and thus $M_t^2 - T_t^\delta(M)$ is a continuous martingale.

Now, for any given $a > 0$ and δ_n be a sequence of subdivisions of $[0, a]$ such that $|\delta_n| \to 0$, we prove that $T_a^{\delta_n}$ converges in L^2. Let δ and δ' be two subdivisions, and $\delta\delta'$ be the subdivision obtained by taking the union of points of δ and δ'. Let $X = T^\delta(M) - T^{\delta'}(M)$, and observe that

$$\begin{aligned}\mathbb{E}(X_t|\mathcal{F}_s) &= \mathbb{E}(T_t^\delta(M) - T_t^{\delta'}(M)|\mathcal{F}_s) \\ &= \mathbb{E}(T_t^\delta(M) - T_s^\delta(M)) + (T_s^\delta(M) - T_s^{\delta'}(M)) + (T_s^{\delta'}(M) - T_t^{\delta'}(M))|\mathcal{F}_s) \\ &= \mathbb{E}(M_t^2 - M_s^2|\mathcal{F}_s) + (T_s^\delta(M) - T_s^{\delta'}(M)) - \mathbb{E}(M_t^2 - M_s^2|\mathcal{F}_s) \\ &= X_s\end{aligned}$$

and hence X is a martingale. Therefore,

$$\mathbb{E}X_a^2 = \mathbb{E}[(T_a^\delta(M) - T_a^{\delta'}(M))^2] = \mathbb{E}(T_a^{\delta\delta'}(X)).$$

Using the inequality $(x_1 + x_2 + ... + x_n)^2 \leq 2(x_1^2 + x_2^2 + ... + x_n^2)$ for real numbers $x_1, ..., x_n$, we have

$$T_a^{\delta\delta'}(X) \leq 2(T_a^{\delta\delta'}(T^\delta) + T_a^{\delta\delta'}(T^{\delta'}))$$

and thus it suffices to prove that $\mathbb{E}T_a^{\delta\delta'}(T^\delta) \to 0$ as $|\delta| + |\delta'| \to 0$.

Let $s_k \in \delta\delta'$ and t_l be the rightmost point of δ such that $t_l \leq s_k < s_{k+1} \leq t_{l+1}$; we have

$$\begin{aligned} T^{\delta}_{s_{k+1}}(M) - T^{\delta}_{s_k}(M) &= (M_{s_{k+1}} - M_{t_l})^2 - (M_{s_k} - M_{t_l})^2 \\ &= (M_{s_{k+1}} - M_{s_k})(M_{s_{k+1}} + M_{s_k} - 2M_{t_l}) \end{aligned}$$

and hence,

$$T^{\delta\delta'}_a(T^{\delta}) \leq T^{\delta\delta'}_a(M) \sup_k |M_{s_{k+1}} + M_{s_k} - 2M_{t_l}|^2.$$

Applying the Cauchy-Schwartz inequality gives,

$$\mathbb{E}T^{\delta\delta'}_a(T^{\delta}) \leq \left(\mathbb{E}(T^{\delta\delta'}_a(M))^2 \mathbb{E}\sup_k |M_{s_{k+1}} + M_{s_k} - 2M_{t_l}|^4\right)^{\frac{1}{2}}.$$

By continuity of M, the first factor would go to zero as $|\delta| + |\delta'| \to 0$. It suffices to show that the second factor is bounded by a constant independent of δ and δ'.

Let $\Delta = \delta\delta' = \{0 = t_0 < t_1 < ...\}$ and for simplicity, let $a = t_n$. Then,

$$\begin{aligned} (T^{\Delta}_a(M))^2 &= \left(\sum_{k=1}^{n}(M_{t_k} - M_{t_{k-1}})^2\right)^2 \\ &= 2\sum_{k=1}^{n}(T^{\Delta}_a(M) - T^{\Delta}_{t_k}(M))(T^{\Delta}_{t_k}(M) - T^{\Delta}_{t_{k-1}}(M)) + \sum_{k=1}^{n}(M_{t_k} - M_{t_{k-1}})^4. \end{aligned}$$

We have shown before that $\mathbb{E}[T^{\Delta}_a(M) - T^{\Delta}_{t_k}(M)|\mathcal{F}_{t_k}] = \mathbb{E}[(M_a - M_{t_k})^2|\mathcal{F}_s]$, and hence,

$$\begin{aligned} \mathbb{E}(T^{\Delta}_a(M))^2 &= 2\sum_{k=1}^{n}\mathbb{E}(M_a - M_{t_k})(T^{\Delta}_{t_k}(M) - T^{\Delta}_{t_{k-1}}(M)) + \sum_{k=1}^{n}\mathbb{E}(M_{t_k} - M_{t_{k-1}})^4 \\ &\leq \mathbb{E}\Bigg(2\sup_k |M_a - M_{t_k}| \sum_{k=1}^{n}(T^{\Delta}_{t_k}(M) - T^{\Delta}_{t_{k-1}}(M)) \\ &\quad + \sup_k |M_{t_k} - M_{t_{k-1}}|^2 \sum_{k=1}^{n}(M_{t_k} - M_{t_{k-1}})^2\Bigg) \\ &= \mathbb{E}\left((2\sup_k |M_a - M_{t_k}|^2 + \sup_k |M_{t_k} - M_{t_{k-1}}|^2)T^{\delta\delta'}_a(M)\right). \end{aligned}$$

By assumption, M is a bounded martingale, and hence there exists a constant C such that $|M| \leq C$. Since, $\mathbb{E}[T^{\Delta}_a(M) - T^{\Delta}_{t_k}(M)|\mathcal{F}_{t_k}] = \mathbb{E}[(M_a - M_{t_k})^2|\mathcal{F}_s]$, we can establish that $\mathbb{E}T^{\delta\delta'}_a(M) \leq 4C^2$. Thus,

$$\mathbb{E}(T^{\Delta}_a(M))^2 \leq 4C^2(\mathbb{E}(2\sup_k |M_a - M_{t_k}|^2 + \sup_k |M_{t_k} - M_{t_{k-1}}|^2) \leq 48C^4$$

Therefore, we have shown that $\mathbb{T}_a^{\delta_n}$ has a limit $\langle M\rangle_2^{0,a}$ as $n \to \infty$ in L^2, and hence in probability. It remains to show that we could choose $\langle M\rangle_2^{0,a}$ within the equivalence class to have the desired properties. For any sequence converging in L^2, we may extract a subsequence that converges almost surely. In particular, there exists a subsequence of partitions δ_{n_k} such that $T_t^{\delta_{n_k}}$ converges a.s. uniformly on $[0,t]$ to $\langle M\rangle_2^{0,t}$ that is continuous. Moreover, the subsequence may be chosen such that $\delta_{n_{k+1}} \subset \delta_{n_k}$, and that $\cup_k \delta_{n_k}$ is dense in $[0,t]$. Since for every $u < v$, we have $T_u^{\delta_{n_k}} \leq T_v^{\delta_{n_k}}$, and so $\langle M\rangle_2^{0,t}$ is increasing on $\cup_k \delta_{n_k}$, which is dense on $[0,t]$. By continuity, $\langle M\rangle_2$ is increasing everywhere. □

Remark 2.2.6. What theorem 2.2.5 is really telling us is that

1. The existence of the quadratic variation for a general martingale M.
2. $\langle M\rangle_2^{0,t}$ is non-decreasing, and hence a process of bounded variation. Therefore, classical Lebesgue integration theory can be applied to the quadratic variation process.

This is the key result in establishing stochastic integration with respect to abstract martingales. □

Now we would like to state a few propositions that generalises the previous theorem. In particular, we would like to extend it to martingales that are unbounded, like the Wiener process.

Proposition 2.2.7. *For every stopping time T, we have $\langle M^T\rangle_2 = \langle M\rangle_2^T$.*

Theorem 2.2.8. *For every continuous local martingale M, the result of the previous theorem applies. In particular,*

$$\sup_{s\leq t} |T_s^{\delta_n}(M) - \langle M\rangle_2^{0,s}|$$

converges to zero in probability.

Proof. Let $\{\tau_n\}$ be a sequence of stopping times defined by

$$\tau_n = \inf\{t : |M_t| \geq n\}.$$

Since M is assumed to be continuous, it cannot explode to ∞ in finite time. Then, we have $\tau_n \to \infty$ and $X_n = M^{\tau_n}$ is a bounded martingale. By theorem 2.2.5, there is, for each n, a continuous adapted and non-decreasing process A_n such that $A_0 = 0$ and $X_n^2 - A_n$ is a martingale. Furthermore, $(X_{n+1}^2 - A_{n+1})^{\tau_n}$ is a martingale and is equal to $X_n^2 - A_{n+1}^{\tau_n}$. By the uniqueness property, we have $A_{n+1}^{\tau_n} = A_n$. Hence, for each n, we can unambiguously define $\langle M\rangle_t^{\tau_n} = A_n$, as clearly we have $(M^{\tau_n})^2 - \langle M\rangle_t^{\tau_n}$ being martingales. Letting $n \to \infty$ in the definition of $\langle M\rangle_t^{\tau_n}$ will uniquely recover the quadratic variation process $\langle M\rangle_t$.

To prove the convergence property, let $\delta, \varepsilon > 0$ and t fixed. One can find a stopping time τ whereby M^τ is a bounded martingale and $\mathbb{P}(\tau \leq t) < \delta$. Since $T^\Delta(M)$ and $\langle M \rangle$ coincide with $T^\Delta(M^\tau)$ and $\langle M^\tau \rangle$ respectively on $[0, \tau]$, we have

$$\mathbb{P}\left(\sup_{s \leq t} |T_s^\Delta(M) - \langle M \rangle_s| > \varepsilon\right) < \delta + \mathbb{P}\left(\sup_{s \leq t} |T_s^\Delta(M^\tau) - \langle M^\tau \rangle_s| > \varepsilon\right)$$

and by theorem 2.2.5, the last term goes to zero as $|\Delta| \to 0$. □

Remark 2.2.9. A frequently-occurring phenomenon in this chapter is that a lot of convergence result such as above weakens from L^2 to convergence in probability when we extend the domain from $[0, T]$ to $[0, \infty)$. Their proofs are very similar to the argument carried out in the above, and will be omitted. Interested readers may consult [10] and [40] for details. □

Definition 2.2.10. (Covariation) Let M and N be continuous local martingales, we define

$$\langle M, N \rangle = \frac{1}{4}[\langle M + N \rangle_2 - \langle M - N \rangle_2].$$

In particular, we have $\langle M, M \rangle = \langle M \rangle$.

Theorem 2.2.11. *Let M and N be local martingales and let δ_n be a sequence of partitions of $[0, s]$. Define*

$$T_s^{\delta_n}(M, N) = \sum_{t_i \in \delta_n} (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}),$$

then,

1. $T_s^{\delta_n}(M, N) \xrightarrow{L^2(\Omega)} \langle M, N \rangle_s$.
2. $MN - \langle M, N \rangle$ *is a local martingale.*
3. $\langle M, N \rangle$ *is the unique continuous process with the above properties.*

Proof. Note that for fixed partitions δ_n,

$$T_s^{\delta_n}(M, N) = \frac{1}{4}(T_s^{\delta_n}(M + N) - T_s^{\delta_n}(M - N)).$$

and so $T_s^{\delta_n}(M, N) \xrightarrow{\mathbb{P}} \langle M, N \rangle$ follows by convergence of quadratic variation. Checking 2. is routine algebra, and 3. follows from the uniqueness of the quadratic variation. □

Remark 2.2.12.

1. The preceding theorem is true for $s = \infty$ as well, but convergence weakens to sense of probability. The proof is again a routine exercise with stopping times.

2. Note that $\langle M, M\rangle$ is an increasing process, and hence it has bounded variation. Thus it makes sense to talk about $\int f d\langle M, M\rangle$ or $\int f d\langle M, M\rangle$ in the Lebesgue sense.

$\square$

Carrying the idea forward from the previous remark, we have

Proposition 2.2.13. *Let $a > 0$ and $\delta_a^n = \{t_0 = 0 < t_1 < ... < t_n = a\}$ be a partition of $[0, a]$. Set $t_i^\lambda = t_i + \lambda(t_{i+1} - t_i)$, where $\lambda \in [0, 1]$. Let M and N be local martingales and H a bounded continuously adapted process. Then,*

$$\lim_{|\delta|\to 0} \sum_i H_{t_i}[(M_{t_i^\lambda} - M_{t_i})(N_{t_i^\lambda} - N_{t_i})] = \lambda \int_0^a H_s d\langle M, N\rangle_s$$

in L^2. This also hold in case when $a = \infty$, but convergence weakens to the sense of probability.

Proof. Let $\langle M, N\rangle_{t_i}^{t_i\lambda} = \langle M, N\rangle_{t_i\lambda} - \langle M, N\rangle_{t_i}$, and by the previous theorem,

$$\sum_i H_{t_i}[(M_{t_i^\lambda} - M_{t_i})(N_{t_i^\lambda} - N_{t_i}) - \langle M, N\rangle_{t_i}^{t_i\lambda}] \to 0$$

where convergence is in L^2. Since $\langle M, M\rangle$ and $\langle N, N\rangle$ are increasing processes and hence $\langle M, N\rangle$ has bounded variation. Moreover, integrals in the Lebesgue sense can be defined via a Riemann sum with respect to $\langle M, N\rangle$. Hence,

$$\begin{aligned}
\lim_{|\delta|\to 0} \sum_i H_{t_i}[(M_{t_i^\lambda} - M_{t_i})(N_{t_i^\lambda} - N_{t_i})] &= \lim_{|\delta|\to 0} \sum_i H_{t_i}\langle M, N\rangle_{t_i}^{t_i\lambda} \\
&= \lim_{|\delta|\to 0} \sum_i \lambda H_{t_i}\langle M, N\rangle_{t_i}^{t_{i+1}} \\
&= \lambda \int_0^t H_s d\langle M, N\rangle_s
\end{aligned}$$

where the second to third line is obtained by approximating $\langle M, N\rangle_{t_i}^{t_i\lambda}$ with $\lambda\langle M, N\rangle_{t_i}^{t_{i+1}}$. This error in this approximation tends to 0 uniformly in λ since $\langle M, N\rangle$ has bounded variation. (write something for case when $a = \infty$). $\square$

Corollary 2.2.14. *Let $f : \mathbb{R} \to \mathbb{R}$ be a C^1 function and f' be its formal derivative. Then,*

$$\sum_i [(f(M_{t_{i+1}}) - f(M_{t_i}))(N_{t_{i+1}} - N_{t_i}) \to \int_0^a f'(M_s) d\langle M, N\rangle_s.$$

Proof. Since $f \in C^1$ and M continuous, we have $(f(M_{t_{i+1}}) - f(M_{t_i})) \to f'(M_{t_i})(M_{t_{i+1}} - M_{t_i})$ as $|\delta| \to 0$. Hence by proposition,

$$\begin{aligned}\lim_{|\delta|\to 0} \sum_i [(f(M_{t_{i+1}}) - f(M_{t_i}))(N_{t_{i+1}} - N_{t_i}) &= \lim_{|\delta|\to 0} \sum_i f'(M_{t_i})(M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}) \\ &= \int_0^a f'(M_s) d\langle M, N\rangle_s.\end{aligned}$$

□

Next, we state a Cauchy-Schwarz type inequality that would lead us to the famous Kunita-Watanabe inequality.

Proposition 2.2.15. *For any continuous local martingales M and N, and measurable processes H and K,*

$$\int_0^t |H_s||K_s||d\langle M, N\rangle|_s \le \left(\int_0^t H_s^2 d\langle M, M\rangle_s\right)^{\frac{1}{2}} \left(\int_0^t K_s^2 d\langle N, N\rangle_s\right)^{\frac{1}{2}}$$

holds a.s. for $t \le \infty$.

Applying Hölder's inequality to the above with $\frac{1}{p} + \frac{1}{q} = 1$, we obtain

Theorem 2.2.16. ***(Kunita-Watanabe Inequality)***

$$\mathbb{E}\left(\int_0^\infty |H_s||K_s||d\langle M, N\rangle|_s\right) \le \left|\left|\int_0^\infty H_s^2 d\langle M, M\rangle_s\right|\right|_p \left|\left|\int_0^\infty K_s^2 d\langle N, N\rangle_s\right|\right|_q.$$

The main purpose of this result is to provide an upper bound for $\int .d\langle M, N\rangle_s$, which becomes a key step in setting up Itô's isometry.

We are now in position to set up the Itô integral, but we first should identity some Hilbert space type structure over the set of continuous local martingales. Let $\mathcal{H}^2$ be the space of continuous L^2-bounded martingales, that is, for each $M \in \mathcal{H}^2$, we have $\sup_t \mathbb{E}M_t^2 < \infty$. Let $\mathcal{H}_0^2$ be the subspace of $\mathcal{H}^2$ such that for every $M \in \mathcal{H}_0^2$, $M_0 = 0$. This space have a default inner product defined by

$$(M, N) = \int_0^\infty M_s N_s ds$$

We can further define an $\mathcal{H}^2$-norm to these spaces by

$$||M||_{\mathcal{H}^2} = \lim_{t\to\infty} (\mathbb{E}(M_t^2))^{1/2}.$$

Polarization of this norm gives rise to an inner product, which thus make $\mathcal{H}^2$ a Hilbert space.

For each $M \in \mathcal{H}^2$, we define $\mathcal{L}^2(M)$ as the space of martingales with the property that, if $K \in \mathcal{L}^2(M)$, then

$$||K||_M^2 = \mathbb{E}\int_0^\infty K_s^2 d\langle M, M\rangle_s < \infty.$$

Again, the norm $||K||_M$ can be made to an inner product via polarization, and hence $\mathcal{L}^2(M)$ is a Hilbert space.

Theorem 2.2.17. *[Itô's Isometry] Let $M \in \mathcal{H}^2$, for each $K \in L^2(M)$, there is a unique element of $\mathcal{H}_0^2$, denoted by $K.M$, such that*

$$\langle K.M, N\rangle = K.\langle M, N\rangle := \int_0^\infty K_s d\langle M, N\rangle_s$$

for every $N \in \mathcal{H}^2$. The map $K \to K.M$ is an isometry from $L^2(M)$ into $\mathcal{H}_0^2$.

Proof. Uniqueness is easy, since if L and L' are two martingales of $\mathcal{H}_0^2$, such that $\langle L, N\rangle = \langle L', N\rangle$, then one can establish $\langle L - L', L - L'\rangle = \langle L - L'\rangle_2 = 0$. By Theorem 2.1.13, the only martingale with zero quadratic variation in $\mathcal{H}_0^2$ is the zero process, and hence $L = L'$ a.s. It remains for us to prove the existence part.

We first work with case when $M, N \in \mathcal{H}_0^2$. By the Kunita-Watanabe's inequality,

$$\begin{aligned}\left|\mathbb{E}\left(\int_0^\infty K_s d\langle M, N\rangle\right)\right| &\leq \left|\left|\int_0^\infty K_s^2 d\langle M, M\rangle_s\right|\right|_2 \left|\left|\int_0^\infty d\langle N, N\rangle_s\right|\right|_2.\\ &\leq ||K||_M ||N||_{\mathcal{H}^2}.\end{aligned}$$

Hence, $N \to \mathbb{E}[(K.\langle M, N\rangle)_\infty]$ is a linear and continuous map on $\mathcal{H}_0^2$. By Riez representation theorem, there exists $K.M \in \mathcal{H}_0^2$, such that

$$(K.M, N)_{\mathcal{H}^2} = \mathbb{E}[(K.M)_\infty N_\infty] = \mathbb{E}[(K.\langle M, N\rangle)_\infty]$$

for every $N \in \mathcal{H}_0^2$. Since elements of $\mathcal{H}_0^2$ are L^2-bounded, it follows by an easy application of Hölder's inequality that they are also uniformly integrable. Hence, for every stopping time T, we have

$$\begin{aligned}\mathbb{E}[(K.M)_T N_T] &= \mathbb{E}[\mathbb{E}[(K.M)_\infty|\mathcal{F}_t]N_T]\\ &= \mathbb{E}[(K.M)_\infty N_\infty^T]\\ &= \mathbb{E}[(K.\langle M, N\rangle^T)_\infty]\\ &= \mathbb{E}[(K.\langle M, N\rangle)_T]\end{aligned}$$

Since the choice of T was arbitrary, it follows that $(K.M)N - K.\langle M, N\rangle$ is a martingale. Further,

$$\begin{aligned}
||K||_M^2 &= \mathbb{E}\int_0^\infty K_s^2 d\langle M, M\rangle_s \\
&= \mathbb{E}[(K^2.\langle M, M\rangle)_\infty] \\
&= \mathbb{E}[(K.\langle M, K.M\rangle)_\infty] \\
&= \mathbb{E}[(K.M)_\infty^2] \\
&= ||K.M||_{\mathcal{H}^2}^2.
\end{aligned}$$

This shows that the map $K \to K.M$ is an isometry. Now, if $M, N \in \mathcal{H}^2$ instead of $\mathcal{H}_0^2$, we still have $\langle K.M, N\rangle = K.\langle M, N\rangle$, because the covariation of any constant martingale is always zero. □

The following theorem relates the quantity $K.M$ to a Riemann sum.

Theorem 2.2.18. *Let $M \in \mathcal{H}_0^2$ and $K \in \mathcal{L}^2(M)$ as before; $\delta_n^a = \{t_0 = 0 < t_1 < ... < t_n = a\}$ be a sequence of partitions of $[0, a]$.*

$$\lim_{|\delta_n|\to 0} \sum_{i=0}^{n-1} K_{t_i}(M_{t_{i+1}} - M_{t_i}) = (K.M)_t.$$

Proof. Consider the case when K is bounded first. Let

$$T^{\delta_n} = \sum_{i=0}^{n-1} K_{t_i} 1_{(t_{i+1},t_i)} S^{\delta_n} = \sum_{i=0}^{n-1} K_{t_i}(M_{t_{i+1}} - M_{t_i}).$$

Then, one easily checks that T^{δ_n} converges to K pointwise, bounded by $||K||_\infty$, and also $S^{\delta_n} = T^{\delta_n}.M$. Thus, by uniqueness of the isometry, as $n \to \infty$, we have $T^{\delta_n} \to M$ and $S^{\delta_n} \to K.M$ boths in L^2. Finally, we relax the boundedness of K and we could achieve the same result (except now converging in probability) with an appropriate choice of stopping times. □

Definition 2.2.19. Let M be a continuous local martingale, define the space of progressively measurable processes denoted by $\mathcal{L}_{loc}^2(M)$, consisting of elements K for which there exists a sequence of stopping times $T_n \to \infty$, such that

$$\mathbb{E}\bigg(\int_0^{T_n} K_s^2 d\langle M, M\rangle_s\bigg) < \infty.$$

Theorem 2.2.20. *The previous theorem extends the choices of K to $\mathcal{L}_{loc}^2(M)$.*

Definition 2.2.21. (Itô Integral) Let M be a continuous local martingale and $K \in \mathcal{L}^2_{loc}(M)$. The Itô's stochastic integral of K with respect to M is defined by

$$\int_0^t K_s dM_s = (K.M)_t.$$

Remark 2.2.22. Many texts, such as [10] and [28] give the definition of the Itô integral as a Riemann sum at a much earlier point of the chapter. I have chosen an alternative approach, by establishing everything we need to know on quadratic variation processes first, it makes our lives a lot easier in setting up the Itô's lemma in the next section. □

Definition 2.2.23. A continuous **semi-martingale** is a process $(X_t, \mathcal{F}_t)$ which has a decomposition

$$X = X_0 + M + V$$

where M is a continuous local martingale and V a continuous process of bounded variation, both $\mathcal{F}_t$ adapted.

Remark 2.2.24. The above decomposition is unique and it is called the **Doob-Meyer decomposition**. □

Proposition 2.2.25. ***(Properties of the Itô integral)*** *Carrying forward the notation from the previous definition, let $Y \in \mathcal{L}^2(M)$. Then*

1. *The Itô integral is a continuous and linear,*

$$\int_0^t Y_s dX_s = \int_0^t Y_s dM_s + \int_0^t Y_s dV_s$$

2. *The process $Z_t = \int_0^t Y_s dM_s$ is an adapted L^2-local martingale.*

In the definition of the Itô integral, we have chosen to use the leftmost point of each interval as our sample point in constructing the Riemann sum. One may enquire what would happen if we had chosen some other point instead. It turns out that as far as convergence is concerned, the choice of the points does not really matter. The following proposition is an exercise taken from [40], it tells us exactly how other types of Riemann sums are related with the Itô integral.

Proposition 2.2.26. *Let μ be a measure on $[0, 1]$, and δ be a partition of [0,a], X a continuously adapted process and M a local martingale. We define*

$$S^\mu_\delta = \sum_i (M_{t_{i+1}} - M_{t_i}) \int_0^1 f(X_{t_i} + \lambda(X_{t_{i+1}} - X_{t_i}) d\mu(\lambda).$$

Then,

$$\lim_{|\delta|\to 0} S_\delta^\mu = \int_0^t f(X_s)dM_s + \bar{\mu}\int_0^t f'(X_s)d\langle X, M\rangle_s$$

where $\bar{\mu} = \int_0^1 s d\mu(s)$.

Proof. Since $f \in C^1$ and X continuous,

$$\begin{aligned}
S_\delta^\mu &= \sum_i (M_{t_{i+1}} - M_{t_i}) \int_0^1 f(X_{t_i} + \lambda(X_{t_{i+1}} - X_{t_i})d\mu(\lambda) \\
&= \sum_i (M_{t_{i+1}} - M_{t_i}) \int_0^1 f(X_{t_i}) + f'(X_{t_i})(\lambda(X_{t_{i+1}} - X_{t_i}))d\mu(\lambda) \\
&= \sum_i (M_{t_{i+1}} - M_{t_i}) f(X_{t_i}) + \sum_i (M_{t_{i+1}} - M_{t_i})(X_{t_{i+1}} - X_{t_i}) f'(X_{t_i}) \int_0^1 \lambda d\mu(\lambda) \\
&= \sum_i (M_{t_{i+1}} - M_{t_i}) f(X_{t_i}) + \bar{\mu}\sum_i (M_{t_{i+1}} - M_{t_i})(X_{t_{i+1}} - X_{t_i}) f'(X_{t_i})
\end{aligned}$$

By Theorem 2.2.14,

$$\sum_i (M_{t_{i+1}} - M_{t_i}) f(X_{t_i}) \xrightarrow{\mathbb{P}} \int_0^a f(X_s)dM_s$$

and by Proposition 2.2.9,

$$\sum_i (M_{t_{i+1}} - M_{t_i})(X_{t_{i+1}} - X_{t_i}) f'(X_{t_i}) \xrightarrow{\mathbb{P}} \int_0^t f'(X_s)d\langle X, M\rangle_s$$

and hence we have finished the proof. □

Corollary 2.2.27. *The Itô integral is the unique stochastic integral that is a local martingale.*

Remark 2.2.28. The case when $f(x) = x$, and μ is a probability measure has a interesting interpretation. It tells us exactly how the Riemann sum of a stochastic integral converges when we randomly choose our sample points according to a given probability distribution. One could interpret it as the law of large numbers for stochastic Riemann sums. In particular, when $\mu = \delta_0$, we recover the Itô integral; when $\mu = \delta_{1/2}$ we would get the so-called Stratonovich integral, where μ_x is the Dirac-δ measure centered at x. It turns out that each of the Itô and Stratonovich integrals has its own advantages. For example, all Itô integrals are $\mathcal{F}_t$-adapted local martingales which makes numerical calculations very easy. The Stratonovich integral, on the other hand, transforms in a much more friendlier manner under change of variables in the sense that it is follows the chain rule of ordinary calculus. □

Definition 2.2.29. Let M be a local martingale and $X \in \mathcal{L}^2(M)$. The **Stratonovich integral**, denoted by $\int . \circ dM$, is defined to be

$$\int_0^t X_s \circ dM_s = \int_0^t X_s dM_s - \frac{1}{2}\langle X, M\rangle_t.$$

2.3 Itô's Lemma and Applications

2.3.1 Itô's Lemma on $\mathbb{R}$ and $\mathbb{R}^n$

Itô's Formula, originally stated as a lemma, can be thought of as a chain rule for stochastic calculus. It is perhaps most commonly stated in most undergraduate textbooks as,

$$dF_t = \frac{\partial F}{\partial X_t} dB_t + \frac{\partial F}{\partial t} dt + \frac{1}{2}\frac{\partial^2 F}{\partial X_t^2} dt$$

where B_t is the standard Brownian motion process, and $F_t = F(t, B_t)$, for some twice differentiable function F. Here, we see that it actually looks like the chain rule, with an extra correction term involving a second derivative in it.

The most intuitive way to understand of why it works, is to simply perform a Taylor series expansion as follows.

$$dF_t = \frac{\partial F}{\partial B_t} dB_t + \frac{\partial F}{\partial t} dt + \frac{1}{2}\left(\frac{\partial^2 F}{\partial B_t^2} dB_t^2 + 2\frac{\partial^2 F}{\partial B_t \partial t} dB_t dt + \frac{\partial^2 F}{\partial t^2} dt^2\right)$$

and then argue that $dt^2 = dB_t dt = 0$, while $dB_t^2 = dt$ and hence it would immediately give us the result.

We now give a more rigorous statement and proof of Itô's formula, in a more general setting.

Lemma 2.3.1. *(Itô's Lemma) Let X_t be a continuous local martingale and V_t be a process of locally bounded variation. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a C^2 function on (x,v), such that $F_t = f(X_t, V_t)$. Then, a.s. for each t, we have*

$$F_t - F_0 = \int_0^t \frac{\partial f}{\partial x}(X_s, V_s) dX_s + \int_0^t \frac{\partial f}{\partial v}(X_s, V_s) dV_s + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(X_s, V_s) d\langle X\rangle_s$$

Proof. Let $\{\delta_t^n\}$ be a sequence of partitions of $[0, t]$, such that $|\delta|_t^n \to 0$ as $n \to \infty$. Then, by the mean value theorem, we know there exists a sequence of random times

η_j and $\tau_j \in [t_j, t_{j+1}]$, such that,

$$\begin{aligned}
f(X_t, V_t) &= f(X_0, V_0) + \sum_{j=0}^{n-1} \Big(f(X_{t_{j+1}}, V_{t_{j+1}}) - f(X_{t_{j+1}}, V_{t_j}) + f(X_{t_{j+1}}, V_{t_j}) - f(X_{t_j}, V_{t_j}) \Big) \\
&= \sum_{j=0}^{n-1} \Bigg(\frac{\partial f}{\partial v}(X_{t_{j+1}}, V_{\tau_j})(V_{t_{j+1}} - V_{t_j}) + \frac{\partial f}{\partial x}(X_{t_j}, V_{t_j})(X_{t_{j+1}} - X_{t_j}) \\
&\qquad + \frac{\partial^2 f}{\partial x^2}(X_{\eta_j}, V_{t_j})(X_{t_{j+1}} - X_{t_j})^2 \Bigg) \\
&= \sum_{j=0}^{n-1} \Bigg(\Big(\frac{\partial f}{\partial v}(X_{t_j}, V_{t_j}) + \varepsilon_j^1 \Big)(V_{t_{j+1}} - V_{t_j}) \\
&\qquad + \frac{\partial f}{\partial x}(X_{t_j}, V_{t_j})(X_{t_{j+1}} - X_{t_j}) \\
&\qquad\quad + \Big(\frac{\partial^2 f}{\partial x^2}(X_{t_j}, V_{t_j}) + \varepsilon_j^2 \Big)(X_{t_{j+1}} - X_{t_j})^2 \Bigg)
\end{aligned}$$

where

$$\begin{aligned}
\varepsilon_j^1 &= \frac{\partial f}{\partial v}(X_{t_{j+1}}, V_{\tau_j}) - \frac{\partial f}{\partial v}(X_{t_j}, V_{t_j}) \\
\varepsilon_j^2 &= \frac{\partial^2 f}{\partial x^2}(X_{\eta_j}, V_{t_j}) - \frac{\partial^2 f}{\partial x^2}(X_{t_j}, V_{t_j}).
\end{aligned}$$

Since it was assumed that the partial derivatives of f were to be continuous over $[0, t]$, they must also be uniformly continuous as $[0, t]$ is compact. Therefore, as $n \to \infty$, we are forced to have $|\delta_t^n| \to 0$, and hence both $\sup_j |\varepsilon_j^1|$ and $\sup_j |\varepsilon_j^2|$ would tend towards 0.

Now it suffices to show that the three terms above converges to each of the three integrals respectively. We do this in two steps, very much like the approach we took to prove theorem 2.2.4. First, we will prove theorem 2.3.1 for bounded X and V, and note that this implies that both $\frac{\partial f}{\partial x}$ and $\frac{\partial^2 f}{\partial x^2}$ are also bounded, as both derivatives are assumed to be continuous over a compact set $[0, t]$. Having done that, we will construct a sequence of stopping times that would extend our result to the general case.

Since V was assumed to be of bounded variation,

$$\sum_{j=0}^{n-1} \frac{\partial f}{\partial v}(X_{t_{j+1}}, V_{\tau_j})(V_{t_{j+1}} - V_{t_j}) \xrightarrow{n \to \infty} \int_0^t \frac{\partial f}{\partial v}(X_s, V_s) dV_s$$

as an ordinary Lebesgue-Stieltjes integral. By theorem 2.2.14,

$$\sum_{j=0}^{n-1} \frac{\partial f}{\partial x}(X_{t_j}, V_{t_j})(X_{t_{j+1}} - X_{t_j}) \xrightarrow{L^2} \int_0^t \frac{\partial f}{\partial x}(X_s, V_s) dX_s,$$

and by proposition 2.2.9,

$$\sum_{j=0}^{n-1}\left(\frac{\partial^2 f}{\partial x^2}(X_{t_j},V_{t_j})+\varepsilon_j^2\right)(X_{t_{j+1}}-X_{t_j})^2\xrightarrow{L^2}\int_0^t\frac{\partial^2 f}{\partial x^2}(X_s,V_s)d\langle X\rangle_s$$

Thus, we have proved theorem 2.3.1 in the case when X and V are bounded.

To extend to the general case, let $\tau_n=\inf\{t\geq 0:|X_t|\vee|V_t|>n\}$ and let $X_t^n=X_{t\wedge\tau_n}1\tau_n>0$ and $V_t^n=V_{t\wedge\tau_n}$. Both X^n and V^n are bounded, therefore theorem 2.3.1 holds a.s. with $t\wedge n$ in place of t, as the probability of X never reaching infinity is one. Hence, theorem 2.3.1 holds in the general case, by letting $n\to\infty$, except convergence weakens to the sense of probability. □

Itô's lemma can be generalised to higher dimensions as follows,

Theorem 2.3.2. *Let* $\mathbf{X}=(X_t^1,X_t^2,...,X_t^n)$ *be a continuous local martingale and* $\mathbf{V}=(V_t^1,V_t^2,...,V_t^n)$ *be a process of locally bounded variation. Let* $f:\mathbb{R}^{2n}\to\mathbb{R}^n$ *be a* C^2 *function on* $(\mathbf{x},\mathbf{v})$*, such that* $\mathbf{F}_t=f(\mathbf{X}_t,\mathbf{V}_t)$*. Then, a.s. for each* t*, we have*

$$\begin{aligned}\mathbf{F}_t-\mathbf{F}_0=&\sum_{i=1}^{n}\left(\int_0^t\frac{\partial f}{\partial x^i}(\mathbf{X}_s,\mathbf{V}_s)dX_s^i+\int_0^t\frac{\partial f}{\partial v^i}(\mathbf{X}_s,\mathbf{V}_s)dV_s^i\right)\\&+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^t\frac{\partial^2 f}{\partial x^i\partial x^j}(\mathbf{X}_s,\mathbf{V}_s)d\langle X,Y\rangle_s\end{aligned}$$

Theorem 2.3.3. *(Itô's lemma for Stratonovich integrals) Carrying through the symbols and notations used in the previous theorem, we have*

$$\mathbf{F}_t-\mathbf{F}_0=\sum_{i=1}^{n}\left(\int_0^t\frac{\partial f}{\partial x^i}(\mathbf{X}_s,\mathbf{V}_s)\circ dX_s^i+\int_0^t\frac{\partial f}{\partial v^i}(\mathbf{X}_s,\mathbf{V}_s)dV_s^i\right)$$

The proof of Theorems 2.3.2 and 2.3.3 is analogously similar to that of Theorem 2.3.1 and thus is left out. It is worth commenting that the Stratonovich integral transforms in the exact same fashion as the Lebesgue integral for functions of bounded variations.

2.3.2 Representations of Martingales

We have seen from the beginning that the concept of a martingale is something that generalises the Wiener process. In this section, we will head backwards to see how these two concepts are really related. In previous section, we established that all processes of the form

$$F_t = F_0 + \int_0^t f_s dW_s$$

are martingales, where W_t is the standard Wiener process. The conclusion of this section is to show that the converse is also true. In fact, this was one of Itô's initial motivations for establishing earlier results. For rest of the thesis, we define $\mathcal{F}_t$ as the filtration generated by the Wiener process W_t unless otherwise defined.

Theorem 2.3.4. ***(Itô's Martingale Representation Theorem)*** *Let M_t be a continuous L^2 martingale of with respect to $\mathcal{F}_t$. Then, there exists a unique continuously adapted process $f_t \in L^2$, such that*

$$F_t = \mathbb{E}F_0 + \int_0^t f_s dW_s.$$

Proof. For simplicity, I will only prove the one-dimensional case as higher dimensional cases are similar. Before we tackle the problem directly, I would like to establish a number of lemmas.

Lemma 2.3.5. *Fix $t > 0$. The set of random variables*

$$\{\phi(W_{t_1}, ..., W_{t_n} : t_i \in [0, t], \phi \in C_0^\infty(\mathbb{R}^n), n = 1, 2, ...\}$$

is dense in $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$.

Proof. Let $\{t_i\}$ be a dense subset of $[0, T]$ and for each $n = 1, 2, ..$, let $H_n = \sigma(W_{t_1}, ..., W_{t_n})$. Then $H_n \subset H_{n+1}$ and $\mathcal{F}_t = \sigma(\bigcup_{n=1}^\infty H_n)$. By the towering containment property, for each $g \in L^2(\Omega, \mathcal{F}_t, \mathbb{P})$,

$$g_t = \mathbb{E}(g|\mathcal{F}_t) = \lim_{n\to\infty} \mathbb{E}(g|H_n)$$

where the limit is taken in L^2. By the Doob-Dynkin theorem (c.f. page 7 [38]), there exists Borel measurable functions g_n such that $\mathbb{E}(g|H_n) = g_n(W_{t_1}, ..., W_{t_n})$, while each Borel measurable functions can be approximated in L^2 by a member of C^∞. $\square$

Lemma 2.3.6. *The linear span of random variables of the type*

$$\exp\left(\int_0^t h(s)dW_s(\omega) - \frac{1}{2}\int_0^t h^2(s)ds\right)$$

is dense in $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$, where $h \in L^2[0,t]$ and h is independent of ω (i.e. functions of the above form form a basis of $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$). The set of processes of this form are termed as exponential martingales denoted by $\mathcal{E}$.

Proof. Suppose $g \in L^2(\Omega, \mathcal{F}_t, \mathbb{P})$ and is orthogonal to all functions of the above form with respect to $(\Omega, \mathcal{F}_t, \mathbb{P})$. Then, in particular,

$$G(\lambda) := \int_\Omega \exp(\lambda_1 W_{t_1}(\omega) + ... + \lambda_n W_{t_n}(\omega)) g(\omega) d\mathbb{P}(\omega) = 0$$

for all $\lambda = (\lambda_1, ..., \lambda_n) \in \mathbb{R}^n$, and all $t_1, ..., t_n \in [0,t]$. Since $G(\lambda)$ is real analytic in $\lambda \in \mathbb{R}^n$, it follows that it has an analytic extension to the n-dimensional complex space $\mathbb{C}^n$ given by,

$$G(z) := \int_\Omega \exp(z_1 W_{t_1}(\omega) + ... + z_n W_{t_n}(\omega)) g(\omega) d\mathbb{P}(\omega) = 0$$

for $z = (z_1, ..., z_n) \in \mathbb{C}^n$. In particular, $G = 0$ on the imaginary axis, namely, $G(i\lambda_1, ..., i\lambda_n) = 0$ for all $(\lambda_1, ..., \lambda_n) \in \mathbb{R}^n$. For $\phi \in C^\infty(\mathbb{R}^n)$, we have

$$\begin{aligned}
\int_\Omega \phi(W_{t_1}, ..., W_{t_n}) g(\omega) d\mathbb{P}(\omega) &= \int_\Omega (2\pi)^{-n/2} \left(\int_{\mathbb{R}^n} \hat{\phi}(y) e^{i(y_1 W_{t_1} + ... + y_n W_{t_n})} dy \right) g(\omega) d\mathbb{P}(\omega) \\
&= (2\pi)^{-n/2} \int_{\mathbb{R}^n} \hat{\phi}(y) \left(\int_\Omega e^{i(y_1 W_{t_1} + ... + y_n W_{t_n})} g(\omega) d\mathbb{P}(\omega) \right) dy \\
&= (2\pi)^{-n/2} \int_{\mathbb{R}^n} \hat{\phi}(y) G(iy) dy = 0
\end{aligned}$$

where

$$\hat{\phi} = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \phi(x) e^{-ix.y} dx;$$

we have used the Fourier inversion theorem in to get the first line and Fubini's theorem to obtain the second line of the calculations (Folland 1984). Hence, we have shown that g is orthogonal to a dense subset of $\mathcal{L}^2(\Omega, \mathcal{F}_t, \mathbb{P})$, and we conclude that $g = 0$, and the exponential martingales do form a basis of $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$. □

Having proved the previous two lemmas, it makes sense for us to first prove the representation theorem on the set of exponential martingales. Define

$$Y_t = \exp\left(\int_0^t h(s) dW_s - \frac{1}{2} \int_0^t h^2(s) ds \right).$$

A straight forward application of Itô's Lemma shows that

$$Y_t = Y_0 + \int_0^t Y_s h(s) dB_s,$$

and hence Y_t satisfies the martingale representation theorem. Now, we can approximate a general $F \in L^2(\Omega, \mathcal{F}_t, \mathbb{P})$, by a linear combination F^n of exponential martingales. Then for each n, we have

$$F_t^n(\omega) = \mathbb{E}F^n + \int_0^t f^n(s,\omega)dB_s(\omega)$$

where $f^n \in L^2[0,t]$ are continuous $\mathcal{F}_t$-adapted processes. Observe that, by Itô's isometry

$$\begin{aligned}
\mathbb{E}(F^n - F^m)^2 &= \mathbb{E}\Big((\mathbb{E}(F^n - F^m) + \int_0^t (f_s^n - f_s^m)dW_s)^2\Big) \\
&= (\mathbb{E}(F^n - F^m))^2 + \int_0^t \mathbb{E}(f_s^n - f_s^m)^2 ds \\
&\to 0
\end{aligned}$$

as m and n tends to infinity. Hence, $\{f^n\}$ is a Cauchy sequence in $\mathcal{L}^2([0,t] \times \Omega)$, and hence converges to some limit $f \in \mathcal{L}^2([0,t] \times \Omega)$. Moreover, there exists a subsequence f^{n_k} of f^n that converges to f almost surely on $(0,t) \times \Omega$. Therefore, $f(t,.)$ is a measurable function for almost every t. By modifying $f(t,\omega)$ on a set of Lebesgue measure zero (in the t-direction), we can obtain a new $f(t,\omega)$ that is $\mathcal{F}_t$-adapted. Hence,

$$F = \lim_{n\to\infty} F_n = \lim_{n\to\infty} \left(\mathbb{E}F^n + \int_0^t f_s^n dB_s\right) = \mathbb{E}F + \int_0^t f_s dB_s,$$

where the limit holds in the L^2 sense. Hence we have shown that the martingale representation theorem holds for all $F \in L^2(\Omega, \mathcal{F}_t, \mathbb{P})$.

To show the uniqueness of f, suppose that $F \in L^2(\Omega, \mathcal{F}_t, \mathbb{P})$ and

$$F = \mathbb{E}F + \int_0^t f^1(t,\omega)dW_t(\omega) = \mathbb{E}F + \int_0^t f^2(t,\omega)dW_t(\omega).$$

Then, by Itô's isometry,

$$0 = \mathbb{E}\bigg(\int_0^t (f_s^1 - f_s^2)dW_s\bigg)^2 = \int_0^t \mathbb{E}(f_s^1 - f_s^2)^2 ds.$$

Hence, f^1 and f^2 disagree on at most a set of measure zero, and therefore the martingale representation is unique. □

The martingale representation theorem has impact in a number of areas. One that is particularly important is that by solely developing a calculus on Wiener processes is enough to solve almost all the problems we want with calculus of martingales. This is generally highly desirable in dealing with problems in finance. In the language of mathematical finance, the existence and uniqueness of the process f_s corresponds to that of replicating hedging strategies. However, in practical situations, one would like to obtain a formula for the replicating strategy f_s, as opposed to only the knowledge of its existence. To this extent, we will see in the next chapter that f_s can actually be explicitly evaluated using the Malliavin calculus.

Chapter 3

Concepts of Malliavin Calculus

3.1 Introduction and Motivations

We aim to develop a probabilistic differential stochastic calculus over an infinite dimensional space. The standard example of such an "infinite dimensional space" is the classical Wiener space, $(C_0[0,T], \mathcal{F}, \mu)$. The theory will be developed on a more general level, along with some solid examples in more familiar spaces such as $C_0[0,T]$. A particular focus will be made on illustrating how classical deterministic calculus fails, and how the problems are fixed by the probabilistic calculus.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, and $L^2(\Omega, \mathcal{F}, \mathbb{P})$ or simply L^2 when there is no risk of confusion, denote the set of square integrable random variables on that space. Loosely speaking, the Malliavin calculus aims to talk about quantities such as $\frac{dF}{d\omega}$, where $F \in L^2$ and $\omega \in \Omega$. To define such a term over a finite dimensional subspace is relatively straight forward. Essentially the theory boils down to classical functional calculus. However, we would like to extend this theory to an infinite dimensional space like $L^2(\Omega, \mathcal{F}, \mathbb{P})$. We will do this in four stages. The first three stages involve looking at mainly L^2 spaces, where the celebrated chaos decomposition theorem plays a central role. The chaos decomposition is essentially an orthogonal basis of the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ in terms of multiple Itô integrals; which makes it possible to approach Malliavin calculus from a Fourier type perspective. Finally we will give a brief examination of Skorohod integration in L^p spaces.

Good references for the material developed in the next three chapters include [2], [3], [4], [11], [23], [19], [31],[32], [35], [37], [41] and [43]. [2] [37] gives a friendly introduction, while [19], [35] and [41] covers the theory to much greater detail. [31] is Malliavin's original paper in 1976, [32] is also written by Malliavin, but it is written for advanced audiences. [4] and [11] provides very interesting alternative approaches to the development of Malliavin calculus. [3], [23] and [43] are much more application focused, but the theory are sufficiently well treated and they provide a good insight to how Malliavin calculus connects with other areas of mathematics.

3.2 Itô-Wiener Chaos Decomposition

3.2.1 Multiple Itô Integrals

We will set up the multiple Itô-integrals that are central to the Itô-Wiener chaos decomposition. First, I would like to stress that there are some potential difficulties with defining the iterated Itô integral. We cannot simply proceed as one does in ordinary several variable calculus, since Itô integrals are processes that are adapted to a filtration $\mathcal{F}_t$. As a consequence, the ordering of the iteration must agree with the ordering of time. One approach to fix this problem is

Let $T > 0$ be fixed, $n \in \mathbb{N}$ and $S_n(T) = \{(t_1, ..., t_n) \in [0, T]^n : t_1 \leq ... \leq t_n\}$. Let W be a Wiener process adapted to the filtration $\mathcal{F}_t$. We would like to make sense of an iterated Itô integral of the following form,

$$J_n f = \int_0^T \left(\int_0^{t_n} \left(\int_0^{t_2} f(t_1, t_2, ..., t_n) dW_{t_1} \right) ... dW_{t_{n-1}} \right) dW_{t_n}.$$

Let $k \geq 2 \in \mathbb{N}$ and $t > 0$, for an arbitrary function $g : \mathbb{R}^k \to \mathbb{R}$, let $g_s(u) = g(u, s)$ where $u \in \mathbb{R}^{k-1}$ and $s \in \mathbb{R}$. Observe that

$$\mathbf{1}_{S_k(t)}(u, s) = \mathbf{1}_{S_{k-1}}(u)\mathbf{1}_{[0,t]}(s)$$

for all $(u, s) \in S_k(t)$. Hence, by Fubini's theorem, for each $g \in L^2(S_k(t))$, we have $g_s \in L^2(S_{k-1}(s))$ for all $s \in [0, t]$ and furthermore,

$$\int_{S_k(t)} g(v) d^k v = \int_0^t \left(\int_{S_{k-1}(s)} g_s(u) d^{k-1} u \right) ds.$$

The above identity and Fubini's theorem allow us to interchange $\mathbb{E}$ and $\int_0^t$ for functions in $L^2(S_k(t))$. Then we can recursively define multiple integrals over $S_k(t)$ as follows:

$$\begin{cases} Y^1 g(\omega, s) := g(s), & \text{(deterministic)}; \\ Y^k g(\omega, s) := \int_0^s Y^{k-1} g(\omega, u) dW_u, & \text{(in } L^2\text{)}. \end{cases}$$

where we can understand $Y^k : L^2(S_k(t)) \to \mathcal{L}_t^2(W)$ as the operator that performs the $k-1$-fold integral. The next proposition will show us that these operators are well defined.

Proposition 3.2.1. *Let $k, l \in \mathbb{N}$ and $t > 0$. Then, $Y^k : L^2(S_k(t)) \to \mathcal{L}_t^2(W)$ defines a linear isometry between the Hilbert spaces. Further, for $k \neq l$,*

$$(Y^k g, Y^l h)_{\mathcal{L}_t^2(W)} = 0$$

for all $g \in L^2(S_k(t))$ and $h \in L^2(S_l(t))$.

Proof. There is nothing to prove for $k=1$. For $k>1$, we proceed by induction to show that the Y^k's are linear isometries,

$$\begin{aligned}
||Y^k g||_{\mathcal{L}^2(W)} &= \mathbb{E}\bigg(\int_0^t (Y^k g)^2(s)ds\bigg)\\
&= \int_0^t \mathbb{E}((Y^k g)(s)^2)ds\\
&= \int_0^t \mathbb{E}\bigg(\int_0^s (Y^{k-1}g_s)(u)^2 du\bigg)ds\\
&= \int_0^t \bigg(\int_{S_{k-1}(s)} g_s^2(u)du\bigg)ds\\
&= \int_{S_k(t)} g^2(v)d^k v\\
&= ||g||^2_{L^2(S_k(t))} < \infty,
\end{aligned}$$

where the third to fourth line was obtained by using the inductive hypothesis. Now, we prove the orthogonality relations as follows. For each fixed $m \in \mathbb{N}$, suffice to show that $(Y^k g, Y^{k+m}h)_{\mathcal{L}_t^2(W)} = 0$ for all $k \in N$. First, consider the case $k=1$. We have,

$$(Y^1 g, Y^{1+m}h)_{\mathcal{L}_t^2(W)} = \mathbb{E}\bigg(\int_0^t g(s)(Y^{1+m}h)(s)ds\bigg) = \int_0^t g(s)\mathbb{E}((Y^{1+m}h)(s))ds = 0,$$

since $g(s)$ is deterministic and $Y^m h$ is a martingale that starts at 0. For the inductive step, observe that

$$\mathbb{E}((Y^k g)(s)(Y^{k+m}h)(s)) = (Y^k g, Y^{k+m}h)_{L^2(\Omega,\mathcal{F}_s,\mathbb{P})} = (Y^{k-1}g_s, Y^{k+m-1}h_s)_{\mathcal{L}_s^2(W)} = 0$$

for all $s \in [0,t]$, by the induction hypothesis. Integrating both sides over $[0,t]$ gives the desired result. □

Definition 3.2.2. For arbitrary $T>0$ and $f \in L^2(S_n(T))$, we now define the iterated Itô integrals recursively as follows. Let,

$$J_0 f := f$$

$$J_n f := \int_0^T (Y^n f)(s)dW_s = \int_0^T (J_{n-1}f_s)dW_s.$$

for $n>1$.

We have the following proposition.

Proposition 3.2.3. *Let $n \in N$ and $T>0$. Then, $J_n : L^2(S_n(T)) \to L^2(\Omega, \mathcal{F}_t, \mathbb{P})$ is a linear isometry. In particular, for $n \neq m$, $(J_n f, J_m g)_{L^2(\Omega,\mathcal{F}_t,\mathbb{P})} = 0$ for all $f \in L^2(S_n(T))$ and $g \in L^2(S_m(T))$.*

Proof. J_n is the composition of two linear isometries, namely,

$$\begin{aligned} L^2(S_n(T)) &\rightarrow \mathcal{L}^2_T(W) \rightarrow L^2(\Omega, \mathcal{F}_T, \mathbb{P}) \text{ and} \\ f &\rightarrow Y^n f \rightarrow \int_0^t (Y^n f)(s) dW_s. \end{aligned}$$

□

Let $\hat{L^2}[0,T]^n$ denote the closed subspace of $L^2[0,T]^n$ consisting of symmetric functions, i.e. functions satisfying

$$f(t_1, ..., t_n) = f(t_{\sigma(1)}, ..., t_{\sigma(n)})$$

for all permutations $\sigma \in \mathfrak{S}_n$. The following result from analysis, which we will not prove, will shed some light as to how should the multiple Itô integrals be extended to $\hat{L^2}[0,T]^n$

Theorem 3.2.4. *Let $n \in \mathbb{N}$ and $T > 0$. Then,*

$$||f||^2_{L^2([0,T]^n)} = n!||f|_{S_n(T)}||^2_{L^2(S_n(T))}$$

for all $f \in \hat{L^2}([0,T])$.

With this result in mind, it is reasonable to have the following definition.

Definition 3.2.5. Let $f \in \hat{L^2}([0,T]^n)$, we define the **multiple Itô integral** of f by

$$I_n f = n! J_n(f|_{S_n(T)}).$$

Since I_n is merely a scalar multiple of J_n, it follows from our previous considerations that I_n is a continuous linear operator and

$$\mathbb{E}(I_n f)^2 = ||I_n f||^2_{L^2(\Omega, \mathcal{F}_T, \mathbb{P})} = n!||f||^2_{L^2([0,T]^n)}.$$

3.2.2 Hermite Polynomials and Chaos Decomposition

We have developed a machinery that allows us to talk about multiple Itô-integrals of symmetric functions. In particular, we will consider functions in $\hat{L^2}([0,T]^n)$ of the form,

$$g^{\otimes n}(x_1, ..., x_n) = \prod_{i=1}^n g(x_i)$$

where $g \in L^2([0,T])$.

Definition 3.2.6. The n-th **Hermite polynomial** is defined by,

$$h_n(x) = (-1)^n e^{(\frac{x^2}{2})} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}.$$

One can also obtain these via the Gram-Schmidt process. The first of these polynomials are, $h_0(x) = 1, h_1(x) = x, h_2(x) = x^2 - 1$ and $h_3(x) = x^3 - 3x$. [37] provides a very thorough discussion on the construction and properties of Hermite polynomials, we will assume them well known.

Given $a > 0$, we define

$$H_n(x, a) = \sqrt{a^n} h_n\Big(\frac{x}{\sqrt{a}}\Big).$$

We have the following lemma,

Lemma 3.2.7. *Let $x, t \in \mathbb{R}$ and $a > 0$, then*

$$e^{tx - \frac{at^2}{2}} = \sum_{n=0}^{\infty} \frac{t^n}{n!} H_n(x, a).$$

Proof. Let $x \in \mathbb{R}$ be fixed, $s = t\sqrt{a}, y = \frac{x}{\sqrt{a}}$, then

$$e^{tx - \frac{at^2}{2}} = e^{sy - \frac{s^2}{2}}.$$

Therefore, without loss of generality, we may assume that $a = 1$. Let $\tau_x(t) = x - t$ and $g(x) = e^{-\frac{x^2}{2}}$, so that $e^{tx - \frac{t^2}{2}} = e^{\frac{x^2}{2}}(g \circ \tau_x)(t)$. Apply Taylor's formula to $g \circ \tau_x$ gives

$$\begin{aligned}
e^{tx - \frac{t^2}{2}} &= e^{\frac{x^2}{2}} \sum_{k=0}^{\infty} \frac{(g \circ \tau_x)^{(k)}(0)}{k!} t^k \\
&= e^{\frac{x^2}{2}} \sum_{k=0}^{\infty} (-1)^k \frac{(g^{(k)} \circ \tau_x)(0)}{k!} t^k \\
&= \sum_{k=0}^{\infty} (-1)^k e^{\frac{x^2}{2}} \frac{d^n}{dx^k} e^{-\frac{x^2}{2}} \frac{t^k}{k!} \\
&= \sum_{k=0}^{\infty} \frac{t^k}{k!} h_k(x).
\end{aligned}$$

$\square$

It can be further checked by elementary calculus that

$$\frac{\partial}{\partial x} H_n(x,a) = nH_{n-1}(x,a) \quad \text{and} \quad \Big(\frac{1}{2}\frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial a}\Big) H_n(x,a) = 0.$$

This allows us to prove the following theorem.

Proposition 3.2.8. *Let $T > 0$ and $g \in L^2([0,T])$. Then, $g^{\otimes n} \in \bar{L}^2([0,T]^n)$ for all $n \in \mathbb{N}$ and,*

$$I_n(g^{\otimes n}) = H_n(X_T, \langle X, X\rangle_T)$$

where $X_t := \int_0^t g(s)dW_s$.

Proof. We shall prove this by induction on n. Since $h_1(x) = x$, basic algebra shows that $H_1(x,a) = x$ also. Therefore, the statement we are proving reduces to

$$I_1(g) = X_T = \int_0^T g(s)dW_s,$$

which is true by definition. Now assume the statement is true for some n, and let $\phi_{n+1} := g^{\otimes n+1}|_{S_{n+1}(T)}$. Then, for all fixed $s \in [0,T]$, $\phi_{n+1}(u,s) = (g|_{[0,s]})^{\otimes n} g(s)$, where $u \in S_n(s)$. From the definition of I_{n+1}, we have

$$\begin{aligned}
I_{n+1} g^{\otimes n+1} &= (n+1)! J_{n+1}(\phi_{n+1}) \\
&= (n+1)! \int_0^T Y^{(n+1)}(\phi_{n+1})_s dW_s \\
&= (n+1)! \int_0^T \int_0^s Y^{(n)}(\phi_{n+1})(u,r) dW_r dW_s \\
&= (n+1)! \int_0^T \int_0^s g(s) Y^{(n)}(g|_{[0,s]^{\otimes n}}) dW_r dW_s \\
&= (n+1) \int_0^T g(s) I_n(g|_{[0,s]^{\otimes n}} dW_s \\
&= (n+1) \int_0^T g(s) H_n(X_s, \langle X, X\rangle_s) dW_s
\end{aligned}$$

where the final step follows by the induction hypothesis. On the other hand, using the previous remark, together with Itô's lemma applied to $H_{n+1}(X_T, \langle X, X\rangle_T)$, we have

$$\begin{aligned}
H_{n+1}(X_T, \langle X, X\rangle_T) &= (n+1) \int_0^T H_n(X_s, \langle X, X\rangle_s) dX_s \\
&= (n+1) \int_0^T H_n(X_s, \langle X, X\rangle_s) g(s) dW_s.
\end{aligned}$$

Thus we have established that

$$I_{n+1}(g^{\otimes n+1}) = H_{n+1}(X_T, \langle X, X \rangle_T).$$

$\square$

Remark 3.2.9. The set of random variables,

$$\left\{ X_T : X_T = \int_0^T g(s)dW_t \right\}$$

for some $g \in L^2$ is called the **Cameron-Martin** subspace of $(\Omega, \mathcal{F}_T, \mathbb{P})$. In the case of $\Omega = C_0[0,T]$, the set of allowed g is typically the set of functions with square integrable derivatives, and the forms a dense subset of $C_0[0,T]$. $\square$

We are now in position to complete the proof of the theorem.

Theorem 3.2.10. ***(Itô-Wiener Chaos Decomposition)*** *Let $T > 0$ and $F \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$. Then there exists a unique sequence $f_n \in \hat{L}^2([0,T]^n)$, such that*

$$F = \mathbb{E}F + \sum_{n=1}^{\infty} I_n(f_n)$$

and

$$||F||^2_{L^2(\Omega)} = \mathbb{E}F^2 + \sum_{n=1}^{\infty} n! ||f_n||^2_{L^2([0,T])}.$$

Proof. Let $X_T = \int_0^T g(s)dW_s$, and define the stochastic exponential of X as follows,

$$\mathcal{E}(X)_T = \sum_{n=0}^{\infty} \frac{1}{n!} H_n(X_T, \langle X, X \rangle_T) = \sum_{n=1}^{\infty} \frac{1}{n!} I_n(g^{\otimes n}).$$

where the second equality comes from applying the previous proposition. Since the H_n's are bounded by a polynomial, this infinite series converges pointwise on Ω. Moreover, using elementary calculus shows that

$$e^{X_T} = \mathcal{E}(X)_T e^{\frac{1}{2}||g||^2_{L^2([0,T])}}.$$

Recall that Lemma 2.3.6 states that,

1. $e^{\int_0^T g(s)dW_s} \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$, for all $g \in L^2([0,T])$.
2. $\{e^{\int_0^T g(s)dW_s} : g \in L^2([0,T])\}$ is a dense subset of $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$,

and thus, $\mathcal{E}(X)_T = Z \in L^2(\Omega)$ as well.

Let $F \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ be given. By lemma 2.3.6, there exists a sequence ζ_n belonging to the linear span of the set $\{e^{\int_0^T g(s)dW_s} : g \in L^2([0,T])\}$, so that $||F - \zeta_n||^2_{L^2(\Omega)} \to 0$. Each ζ_n can be written as a finite sum of the type

$$\zeta_n = \sum_{k=1}^{l_n} \alpha_k e^{X^k} = \sum_{k=1}^{l_n} \alpha_k \mathcal{E}(X^k)_T e^{\frac{1}{2}||g||^2_{L^2([0,T])}}$$

with $\alpha_k \in \mathbb{R}$, $g_k \in L^2([0,T])$ and $X^k_T = \int_0^T g(s)dW_s$. By theorem 2.3.6, and previous considerations, each stochastic exponential $\mathcal{E}(X^k)_T$ can be written as

$$\mathcal{E}(X^k)_T = \sum_{m=0}^{\infty} \frac{1}{m!} I_m(g_k^{\otimes m}),$$

so that $\zeta_n = \sum_{m=0}^{\infty} J_m(\phi^n_m)$ with $\{\phi^n_m : n \in \mathbb{N}\} \subseteq L^2(S_m(T))$. Orthogonality and Itô's isometry now lead to $||Z_i - Z_j||^2_{L^2(\Omega)} = \sum_{m=0}^{\infty} ||\phi^i_m - \phi^j_m||^2_{L^2(S_m(T))}$ for all $i, j \in \mathbb{N}$. Thus, $(\phi^i_m)_{i \in \mathbb{N}}$ is a Cauchy sequence in $L^2(S_m(T))$ for every $m \in \mathbb{N}$. By completeness, there exists a limit $\phi_m \in L^2(S_m(T))$ with $||\phi_m - \phi^i_m||^2_{L^2(S_m(T))} \to 0$ as $i \to \infty$, and thus we see that $\sum_{m=0}^{\infty} ||\phi_m - \phi^i_m||^2_{L^2(S_m(T))} \to 0$ for $i \to \infty$. Now, by orthogonality and Itô's isometry again, we see that $(\zeta_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in $L^2(\Omega)$, and so there exists $\zeta := \sum_{m=0}^{\infty} J_m(\phi_m) \in L^2(\Omega)$. Uniqueness of the limit now implies that

$$F = \zeta = \sum_{m=0}^{\infty} J_m(\phi_m).$$

To finish the proof, we may extend each ϕ_m trivially to a function $\psi_m \in L^2([0,T]^m)$ and consider then the symmetrization $\hat{\psi_m}$ of ψ_m,

$$\hat{\psi_m} := \frac{1}{m!} \sum_{\sigma \in \mathfrak{S}} \psi_m \circ A_\sigma \in \hat{L}^2([0,T]^m),$$

where $A_\sigma(t_1, ..., t_m) := (t_{\sigma(1)}, ..., t_{\sigma(m)})$ for all $(t_1, ..., t_m) \in [0,T]^m$. Since $A_\sigma(S_m(T))$ has no common points with $S_m(T)$ for all $\sigma \neq 1$, the definition of ψ_m implies that $(\psi_m \circ A_\sigma)|_{S_m(T)} = 0$ for all $\sigma \neq 1$. Therefore, $\hat{\psi_m}|_{S_m(T)} = \frac{1}{m!}\phi_m$, and we obtain

$$F = \sum_{m=0}^{\infty} J_m(\phi_m) = \mathbb{E}F + \sum_{m=1}^{\infty} I_m(\hat{\psi}).$$

Moreover, theorem 3.2.4 implies that

$$||F||^2_{L^2(\Omega)} = \sum_{m=0}^{\infty} ||\phi_m||^2_{L^2(S_m(T))} = \sum_{m=0}^{\infty} m! ||\hat{\psi_m}||^2_{L^2([0,T])^m}.$$

Thus, we have established the chaos decomposition theorem. □

Remark 3.2.11.

1. The approach I took to establish this theorem is somewhat non-standard. The more popular approach, as taken by [35] and [37] for example, is to establish that
$$L^2(\Omega, \mathcal{F}_T, \mathbb{P}) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n$$
where $\mathcal{H}_n$ is the space spanned by the set $h_n(W_{t_i})$. One then show, that each of the $\mathcal{H}_n$'s can be related to the limit of a discretisation of the multiple Itô integral. The advantage of that approach is that no prior exposure of Itô's isometry is required, but this makes the proofs are somewhat longwinded. For this reason, I have chosen a more geometric type of argument which relies on to a very large extent of Itô's isometry.
2. The decomposition $L^2(\Omega, \mathcal{F}_T, \mathbb{P}) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n$ was first known by Wiener before any stochastic integration theory had appeared. The elements in $\mathcal{H}_n$ were traditionally known as Wiener chaos. In 1951, Itô showed in [22], that these Wiener chaos can in fact be recognised as multiple Itô integrals.

$\square$

3.3 The Malliavin Derivative

This section is devoted to the development of a differential calculus on an infinite dimensional measure space like $L^p(\Omega, \mathcal{F}, \mathbb{P})$. Before we begin, let us recall some ideas from functional analysis. I will first demonstrate how far these ideas can be pushed until the approach becomes problematic. I then introduce some new probabilistic concepts to fix these problems.

Definition 3.3.1. (Fréchet Derivative) Let X and Y be Banach spaces and let U be an non-empty open subset of X. A mapping $f : U \to Y$ has a **directional derivative** at $x \in U$ in the direction of $v \in X$, $||v||_X = 1$ if

$$\mathbf{D}_v f(x) = \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} = \frac{d}{d\varepsilon} f(x + \varepsilon v)|_{\varepsilon=0}$$

exists. If $\mathbf{D}_v f(x)$ indeed exists, then we call it the **directional derivative** of f (at x in direction of v). Moreover, we say that f is **Fréchet differentiable** at $x \in U$ if there exists a linear operator $A : X \to Y$ such that

$$\lim_{h \to 0} \frac{||f(x + h) - f(x) - Ah||_Y}{||h||_X} = 0.$$

If this is the case, we call A the **Fréchet derivative** of f at x.

We now try to apply the definition of a Fréchet derivative to classical Wiener space $(\Omega, \mathcal{F}, \mu)$, where $\Omega = C_0[0, T]$ and μ is the Wiener measure. Observe that under the sup-norm, $C_0[0,T]$ is a Banach space, which has a densely embedded Hilbert subspace H^1, defined by,

$$H^1 = \{h \in C_0[0,T] : h' \in L^2[0,T]\}.$$

We call H^1, the space of all continuous functions with square integrable derivatives, the **Cameron-Martin subspace**. We equip H^1 with an inner product defined by

$$(g,h)_{H^1} = \int_0^T h'(t)g'(t)dt$$

The fact that this is dense in $C_0[0,T]$ is a consequence of the Stone-Weierstrass theorem. It turns out that to obtain a theory involving derivatives in all directions is still an open problem (see chapter 4 of [37] for details). Thus, we will first restrict ourselves to defining a directional derivative of a random variable F only in the directions in the Cameron-Martin subspace. We will later see that this theory generalises quite easily to allow derivatives in directions of the so-called isonormal Gaussian processes.

Definition 3.3.2. Let $F : \Omega \to \mathbb{R}$ be a random variable. We say F has a derivative in the direction of γ, where

$$\gamma(t) = \int_0^t g(s)ds, \qquad g \in L^2([0,T]),$$

in the strong sense at ω if

$$\mathbf{D}_\gamma F(\omega) = \frac{d}{d\varepsilon}F(\omega + \varepsilon\gamma)|_{\varepsilon=0} = \lim_{\varepsilon\to 0}\frac{F(\omega+\varepsilon\gamma) - F(\omega)}{\varepsilon}$$

exists in $L^2(\Omega)$. If in addition there exists $\psi(t,\omega) \in L^2([0,T]\times\Omega)$, such that

$$\mathbf{D}_\gamma F(\omega) = \int_0^T \psi(t,\omega)g(t)dt$$

then we say F is **Fréchet differentiable** at ω, with **Fréchet derivative** defined by

$$\mathbf{D}_t F(\omega) = \psi(t,\omega)$$

and we thus $\mathbf{D}_\gamma(F) = (\mathbf{D}_t F, \gamma)$. The set of all Fréchet differentiable random variables will be denoted by $\mathbb{D}_{1,2}$.

Remark 3.3.3. We can understand

$$\int_0^T \psi(t,\omega)g(t)dt$$

as a matrix multiplication in continuous dimensions. □

Example 3.3.4. Let $\Omega = C_0[0,T]$ be a Banach space, $H = L^2([0,T])$ and $h \in H$. Suppose that

$$F(\omega) := \left(\int_0^T h(s) dW_s \right)(\omega) = \int_0^T h(s) d\omega(s).$$

If $\gamma(t) = \int_0^t g(s) dW_s$ for some $g \in H$ and $t < T$. , then

$$\begin{aligned} F(h + \varepsilon\gamma) &= \int_0^T f(s) d(\omega(s) + \varepsilon\gamma(s)) \\ &= \int_0^T f(s) d\omega(s) + \varepsilon \int_0^T f(s) g(s) ds. \end{aligned}$$

Hence,

$$\frac{F(\omega + \varepsilon\gamma) - F(\gamma)}{\varepsilon} = \int_0^T f(s) g(s) ds$$

for all $\varepsilon > 0$. Therefore, $\mathbf{D}_t F(\omega) = \psi(t, \omega) = h(t)$ for all $t \in [0,T]$ and $\omega \in \Omega$. $\square$

Following a standard procedure in analysis, we introduce a type of Sobolev norm.

Definition 3.3.5. Let $\mathcal{D}_{1,2}$ denote the set of Fréchet differentiable random variables F with the Sobolev norm,

$$||F||_{1,2} = \sqrt{||F||_{L^2}^2 + ||DF||_{L^2([0,T] \times \Omega)}^2} < \infty.$$

At this stage, we would like to do the following two things:

1. Generalise the concept of a derivative to more general measure spaces.
2. Hope that $\mathcal{D}_{1,2}$ is a Sobolev space under the norm $||.||_{1,2}$.

Unfortunately, a derivative in the sense of Fréchet will not allows us for any of these. The reason is that in general, we will be interested in random variables F that are defined $\mathbb{P}$-almost surely, while the Fréchet derivative is implicitly dependant on the continuity of F with respect to some topology. For this reason, it is necessary that our notion of derivative should not depend on any topological structure of Ω, that is, we need a derivative which acts in the weak sense.

To address the second problem, if we are working in the classical Wiener space, it is evident that the existence of a Fréchet derivative of a random variable F depends on the existence of a continuous version of F. The following example (taken from an exercise in section 1.2 of [35]) demonstrates the existence of a random variable, F that do not possess a continuous version. Moreover, there exists a sequence of Fréchet differentiable random variables $F_n \to F$ pointwise. This demonstrates that $\mathcal{D}_{1,2}$ is not complete, and hence cannot be made into a Sobolev space as we hope it would be.

Example 3.3.6. Let $W = \{W_t : t \in [0,T]\}$ be a one dimensional Wiener process defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with natural filtration $\mathcal{F}_t$, and $h \in L^2[0,T]$. Consider the random variable

$$F = \int_0^T h(t) dW_t.$$

We claim that F will not have a continuous version if there does not exist a corresponding signed measure μ on $[0,T]$ such that $h(t) = \mu((t,1])$, for all $t \in [0,1]$, Lebesgue-almost everywhere. □

Proof. Suppose F has a continuous modification, that is, there exists $G \in C_0[0,T]$ such that $G = F$, $\mathbb{P}$-almost surely. Moreover, linearity of the Itô-integral implies that $G : C_0[0,T] \to \mathbb{R}$ is a continuous linear functional. By Riez representation theorem, for each $\omega \in \Omega$, there exists $g_\omega \in C_0[0,T]$ such that,

$$G(\omega) := \left(\int_0^T h(t) dW_t \right)(\omega) = \int_0^T g_\omega(t)\omega(t) dt.$$

Application of integration by parts (corollary 2.3.4) shows that

$$\int_0^T g_\omega(t)\omega(t) dt = \left(\int_0^T h(t) dW_t \right)(\omega) = -\int_0^T \omega(t) dh(t) + K$$

for some constant K. The expression on the far left is an honest Lebesgue integral, thus forces $-\int_0^T \omega(t) dh(t)$ to be also a well defined integral. Hence there exists a signed measure μ such that $h(t) = \mu((t,T])$. Consequently if $h(t)$ does not admit a corresponding μ, then the random variable F will not have a continuous modification.

On the other hand, we know that $dh(t)$ can be made into a signed measure if and only if $h(t)$ has bounded variation on $[0,T]$. Let $h(t)$ to be any continuous function of unbounded variation so that $F \notin \mathcal{D}_{1,2}$. For such an h, we choose a sequence of differentiable functions h_n, such that $h_n(t) \to h(t)$ uniformly on [0,T]. Let

$$F_n = \int_0^T h_n(t) dW_t.$$

Then clearly $F_n \in \mathcal{D}_{1,2}$ for all n, and by example 3.3.6, we have $\mathbf{D}_t F_n = h_n(t)$. Hence, F_n is a Cauchy sequence under the $||.||_{1,2}$ norm. But, $F_n \xrightarrow{L^2} F$, and $F \notin \mathcal{D}_{1,2}$. i.e. the space $\mathcal{D}_{1,2}$ is not complete. □

In conclusion, we see that anything remotely depends on the topology of Ω will be doomed to failure, and the Fréchet derivative is not sufficient in order to extend the theory to a more general setting. To remedy this, we introduce the Malliavin derivative, a generalisation of the Fréchet derivative defined in the weak sense. The

Malliavin derivative give the solution to both of the two problems that we have encountered.

Remark 3.3.7. A close analogy of the relationship between Fréchet and Malliavin derivative for random variables F, is the relationship between Riemann and Lebesgue integration for some function f. The definitions of both Riemann and Fréchet sets up the theoretical foundation at an intuitive level, yet both approaches had a common problem of the domain being an incomplete space. One of the main purposes of the work of Lebesgue and Malliavin serves is to solve this problem. □

The set of directions in whom Malliavin differentiation is defined is the normal generalisation of the concept of a Cameron-Martin subspace, the space of so-called isonormal Gaussian processes.

Definition 3.3.8. A stochastic process $W = \{W(h), h \in H\}$ defined in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an **isonormal Gaussian process** (on H), if W is a centered Gaussian family of random variables, such that

$$(W(h), W(g))_\Omega = (h, g)_H$$

for all $h, g \in H$.

Example 3.3.9. Let $\Omega = L^2[0, T]$ and $h \in \Omega$, and if for all $T > 0$,

$$W_T(h) = \int_0^T h(s) dW_s,$$

then observe that for each $g \in \Omega$,

$$(W_T(h), W_T(g))_\Omega = \mathbb{E}\Big(\int_0^T h(s)dW_s, \int_0^T g(s)dW_s\Big) = \int_0^T f(s)g(s)ds$$

where the last equality was obtained by Itô's isometry. Hence, $\{W(h) : h \in L^2[0, T]\}$ is an isonormal Gaussian process. □

Remark 3.3.10.

1. Suppose H be a Hilbert space and for $h \in H$, let $W(h)$ be an isonormal Gaussian process. Then the map $h \to W(h)$ is a linear map. For any λ and $\mu \in \mathbb{R}$, and $g, h \in H$, we have

$$\begin{aligned}
&\mathbb{E}(W(\lambda h + \mu g) - \lambda W(h) - \mu W(g))^2 \\
&= ||\lambda h + \mu g||_H^2 + \lambda^2 ||h||_H^2 + \mu^2 ||g||_H^2 \\
&\qquad - 2\lambda(\lambda h + \mu g, h)_H - 2\mu(\lambda h + \mu g, g)_H + 2\lambda\mu(h, g)_H \\
&= \lambda^2 ||h||_H^2 + \mu^2 ||g||_H^2 - \lambda(\lambda h + \mu g, h)_H - \mu(\lambda h + \mu g, g)_H + 2\lambda\mu(h, g)_H \\
&= 0
\end{aligned}$$

The mapping $h \to W(h)$ provides a linear isometry of H onto a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, consisting of zero-mean Gaussian random variables.

2. One can always associate an abstract Wiener space to a Hilbert space H. That is, a Gaussian measure μ on a Banach space Ω, such that H is continuously injected onto Ω, with the following inclusions, $\Omega^* \subset H^* \simeq H \subset \Omega$, dense, then
$$\int_{\Omega} e^{it(x^*y)}\mu(dy) = \frac{1}{2}||x||_H^2$$
for any $x \in \Omega^*$. Readers are referred to section 1.4 of [19] and [30] for a detailed construction of Gaussian measures on general Banach spaces.
This is in fact a very popular way of generalising the notion of differentiation from classical Wiener spaces, it is pursued by [4],[11], [19] and [23]. In such cases, the probability space Ω is again endowed with a reasonably nice topology, however such topological structures are redundant as the concepts we introduce from this point are aimed to hold on general measure spaces.

□

Let $C_p^\infty(\mathbb{R}^n)$ denote the set of infinitely differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, such that f and its partial derivatives of all orders have polynomial growth. Let $\mathcal{S}$ be the set of all **smooth random variables**, such that if $F \in \mathcal{S}$, then there exists $h_1, ..., h_n \in H$ such that
$$F = f(W(h_1), ..., W(h_n)),$$
where $f \in C_p^\infty(\mathbb{R}^n)$. Let $\mathcal{P}$, $\mathcal{S}_b$ and $\mathcal{S}_0$ be the set of smooth random variables of the above form such that $f \in \mathbb{R}[x_1, ..., x_n]$, $f \in C_b^\infty(\mathbb{R}^n)$ (f and its partial derivatives of all orders are bounded) and $C_0^\infty(\mathbb{R}^n)$ (f has compact support) respectively. Note that $\mathcal{P} \subset \mathcal{S}, \mathcal{S}_0 \subset \mathcal{S}_b \subset \mathcal{S}$, and that both $\mathcal{P}$ and $\mathcal{S}_0$ are dense in $L^2(\Omega)$ (see section 1.2 of [35] for detailed proof).

Definition 3.3.11. The **Malliavin derivative** of a smooth random variable F of the above form is the stochastic process $\{D_tF, t \in T\}$ given by
$$D_tF = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(W(h_1), ..., W(h_n))h_i(t).$$
We will drop the subscript t where there is no risk of confusion.

Example 3.3.12. Consider the case when $f(x) = x$, so $F = W(h)$. Then, trivially we obtain $D_tW(h) = h(t)$. This agrees with what we obtained for the Fréchet derivative. □

In fact, we have a much stronger result.

Proposition 3.3.13. *Let F be a smooth random variable over $C_0[0,T]$ and H^1 be its canonical Cameron-Martin subspace. Then if $F \in \mathcal{D}_{1,2}$ and $\mathbf{D}_\gamma F = D_\gamma F$, for all H^1.*

Proof. Let $h, h_1, h_2, ..., h_n \in H^1$. Since $\sup_{s\in[0,T]} \mathbb{E}|W_s(g)|^N < \infty$ for all $n \in \mathbb{N}$ and $g \in H^1$, we see that

$$\begin{aligned}
(\mathbf{D}F, h)_H &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[f(W(h_1) + \varepsilon(h_1, h)_H, ..., W(h_n) + \varepsilon(h_n, h)_H) - f(W(h_1), ..., W(h_n))] \\
&\to \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(W(h_1), ..., W(h_n))\mathbf{D}_i W(h) \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(W(h_1), ..., W(h_n))h_i(t) \\
&= (DF, h)_H,
\end{aligned}$$

where the chain rule (c.f. chapter 7 of [9]) for Fréchet derivative was used to obtain the second line. □

Remark 3.3.14.

1. Since $\mathcal{S}$ is dense in $L^p(\Omega)$, intuitively we would like to define the Malliavin derivative of a general $F \in L^p(\Omega)$ by means of taking limits. However, there is still one potential problem. Suppose $\{F_n\}$ and $\{G_n\}$ are two sequences both approaching F under the L^p norm. There is no guarantee at this stage, why should DF_n and DG_n should also approach the same limit. This problem will be solved in the forthcoming theorem, where we establish the so called closability property.
2. When working with a general measure space $(\Omega, \mathcal{F}, \mathbb{P})$ that may not necessarily be endowed with a topology, perhaps the closest analogy to "continuous maps" would be the closure of the set of smooth maps. For this reason, we first defined the Malliavin derivative over the set of smooth random variables, and we will prove that such a derivative is stable under taking limits, and hence obtain its closure.

□

We would like to prove that D is closable as an operator from $L^p(\Omega)$ to $L^p(\Omega, H)$, and thus hope to define the Malliavin derivative of a general object F by means of a limit. Before this is possible, we need to introduce the idea of the product rule for differentiation and also the integration by parts formula.

Lemma 3.3.15. *Let $F, G \in \mathcal{S}$, then $D(FG) = G.DF + F.DG$.*

The proof of this lemma is a direct consequence of the definition of the Malliavin derivative, and also the product rule for ordinary calculus.

Lemma 3.3.16. *If $F \in \mathcal{S}$ and $h \in H$, we have*

$$\mathbb{E}(DF, h)_H = \mathbb{E}(FW(h)).$$

Proof. Let $\{e_n\}_{n\in N}$ be a complete orthogonal system of H. Without loss of generality, we may assume that $h = e_1$, and that F is of the form

$$F = f(W(e_1), ..., W(e_n)),$$

where $f \in C_p^\infty(\mathbb{R}^n)$. Let μ_n denote the n-fold Wiener measure, then

$$\begin{aligned}
\mathbb{E}(DF, h)_H &= \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_1}(x) d\mu_n(x) \\
&= \int_{\mathbb{R}^n} f(x) x_1 d\mu_n(x) \\
&= \mathbb{E}(FW(h)),
\end{aligned}$$

and hence the lemma follows. □

Lemma 3.3.17. *Let $F, G \in \mathcal{S}$, and $h \in H$. Then,*

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

Proof. Apply the integration by parts formula to FG to obtain,

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

and the lemma follows. □

Theorem 3.3.18. *The Malliavin derivative $D : L^p(\Omega) \to L^p([0, T] \times \Omega)$ is closable as an operator.*

Proof. It suffices to prove that if a sequence of smooth random variables $\{F_k\} \xrightarrow{L^p(\Omega)} 0$, and $DF_k \xrightarrow{L^p(L^2[0,T],\Omega)} \xi$, then $\xi = 0$ in the sense of $L^p(L^2[0, T], \Omega)$, since $\mathcal{S}$ is dense in $L^p(\Omega)$. Let $G \in \mathcal{S}_b$ and $h \in H$. Then by the previous lemma, we have for each $k \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}((\xi, h)_H G) &= \lim_{k\to\infty} \mathbb{E}(G(DF_k, h)_H) \\
&= \lim_{k\to\infty} \mathbb{E}(-F_k(DG, h)_H + F_k GW(h)) \\
&= 0.
\end{aligned}$$

The last equality holds since F_k converges to zero in L^p, and both G and DG were assumed to be bounded. Since the choice of G was arbitrary in $\mathcal{S}_b$, and that $\mathcal{S}_b$ is dense, this implies that $\xi = 0$. □

Definition 3.3.19. (Malliavin Derivative) Let $F \in L^p(\Omega)$ and $\{F_n\}$ a sequence of smooth random variables converging to F in L^p. We define, the Malliavin derivative of F to be

$$DF = \lim_{n\to\infty} DF_n.$$

Remark 3.3.20. To see the above definition is well defined, we would like to verify that if $F_n \to F$ and $G_n \to F$, it follows that $\lim_{n\to\infty} DF_n = \lim_{n\to\infty} DG_n$. To see this, consider $H_n = F_n - G_n$, so that $H_n \to 0$. The preceding theorem implies that $DH_n \to 0$, and hence DF is well defined. $\square$

Definition 3.3.21. We will denote the domain of D in $L^p(\Omega)$ by $\mathbb{D}_{1,p}$, and we equip this space with the norm

$$||F||_{1,p} = \left(||F||^p_{L^2(\Omega)} + ||DF||^p_{L^p(L^2[0,T]\times\Omega)}\right)^{\frac{1}{p}}.$$

for every $F \in \mathbb{D}_{1,p}$.

We have apparently two different derivatives at this stage, and we wish to investigate the relations between them. By lemma 3.3.13, we know that if $F \in \mathcal{S}$, then $DF = \mathbf{D}F$. Combining definition 3.3.19 and theorem 3.3.18, we have $\mathbf{D}F = DF$ for all $F \in \mathbb{D}_{1,2} \cap \mathcal{D}_{1,2}$.

Example 3.3.6 showed that $\mathbb{D}_{1,2} \setminus \mathcal{D}_{1,2} \neq \emptyset$. On the other hand, it is well known that $\mathcal{D}_{1,2} \setminus \mathbb{D}_{1,2} \neq \emptyset$ (c.f. [37]). The reason is that the Fréchet derivative is defined by a local property. But to prove the closability criteria for Malliavin derivatives, we had to assume that $F \in L^p(\Omega)$, which is a global condition. Therefore, it is no surprise that $\mathcal{D}_{1,2} \setminus \mathbb{D}_{1,2} \neq \emptyset$, as F can be locally smooth to accommodate for taking Fréchet derivatives, yet globally not integrable.

Definition 3.3.22. We can further make $\mathbb{D}_{1,2}$ into a Hilbert space by equipping it with the inner product

$$(F, G)_{1,2} = \mathbb{E}(FG) + \mathbb{E}((DF, DG)_H).$$

for $F, G \in \mathbb{D}_{1,2}$.

Furthermore, we shall define the iterated derivatives of k-times weakly differentiable random variables.

Definition 3.3.23. Let F be a smooth random variable and k a positive integer. We define

$$D^k_{t_1,\ldots,t_k} F = D_{t_1} D_{t_2} \ldots D_{t_k} F.$$

Let $\mathbb{D}_{k,p}$ denote the set of all k-times differentiable random variables, subject to

$$||F||_{k,p} = \left(||F||^p_{L^p(\Omega)} + \sum_{j=1}^{k} ||D^j F||^p_{L^p(L^2[0,T]\times[0,T_j])} \right) < \infty.$$

In particular, we define,

$$\mathbb{D}_{k,\infty} := \bigcap_{1<p<\infty} \mathbb{D}_{k,p}$$

and

$$\mathbb{D}_{\infty} := \bigcap_{k\in\mathbb{N}} \mathbb{D}_{k,\infty}.$$

Remark 3.3.24. Note that the derivative $D^k F$ is considered as a measurable function on the product space $[0,T]^k \times \Omega$. □

For higher order derivative operators D^k, we also have a closable condition analogous to that of D. See section 1.5 of [35] for proof.

Proposition 3.3.25. *Let $\{F_n, n \geq 1\}$ be a sequence of random variables in $\mathbb{D}_{k,p}$, with $k \geq 1$, and $p > 1$. Assume that $F_n \xrightarrow{L^p(\Omega)} F$ and $\sup_n ||F_n||_{k,p} < \infty$, then $F \in \mathbb{D}_{k,p}$.*

So far, we have given a rigorous but not so instructive definition of the Malliavin derivative. The following theorem will tell us how the D operator behaves in the L^2 setting, with respect to the orthogonal basis we constructed from the Wiener chaos decomposition.

Theorem 3.3.26. *Let $F \in L^2(\Omega)$, with chaos decomposition,*

$$F = \sum_{m=0}^{\infty} I_m(f_m),$$

where $f_m \in L^2(S(T)^m)$. Then $F \in \mathbb{D}_{1,2}$ if and only if

$$\sum_{m=1}^{\infty} mm! ||f_m||^2_{L^2([0,T]^m)} < \infty.$$

In case when the above hold, we have

$$DF = \sum_{m=1}^{\infty} m I_{m-1}(f_m)$$

and that

$$||DF||^2_{L^2([0,T]\times\Omega)} = \sum_{m=1}^{\infty} mm! ||f_m||^2_{L^2([0,T]^m)}.$$

Heuristically, this theorem tells us that the Malliavin derivative essentially removes the iterates of multiple Itô integrals as ordinary operators of differentiation do to polynomials. This intuition can be made rigorous via the so called "Wick product", where Itô integrals can be recognised as algebras of Wick polynomials. Its applications span from quantum field theory to fractional Brownian motion and stochastic PDEs. Readers are advised to read [37] for an introductory treatment, and [18] for a more detailed study.

Proof. Let f_m be a sequence of square integrable functions over $[0,T]^m$, $W(g)$ an isonormal stochastic process for some g in a Hilbert space H, and let

$$F^N = \sum_{m=0}^{N} I_m(f_m).$$

By proposition 3.2.8, and using the fact $\langle W, W\rangle_T = T$, we have

$$D_t F^N = \sum_{m=1}^{N} h'_m(W(g))g(t) = \sum_{m=1}^{N} h_{m-1}(W(g))g(t) = \sum_{m=1}^{N} m I_{m-1}(f_m(u,t)).$$

for some $u \in [0,T]^{m-1}$. By Itô's isometry, the I_{m-1} terms belongs to $L^2(\Omega)$, and hence $DF^N \in L^2([0,T]\times\Omega)$ and $F^N \in \mathbb{D}_{1,2}$ for every $N \in \mathbb{N}$. Thus it remains to find conditions in order that DF^N is stable as $N \to \infty$. Let l be a square integrable symmetric function in n variables, and $L = I_n(l)$. Then, both F_N and L are smooth random variables, so we can apply lemma 3.3.15 to obtain,

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{E}[(DF^N, h)_H L] &= \lim_{N\to\infty} \mathbb{E}[-F^N(DL, h)_H + F^N L W(g)] \\
&= \mathbb{E}[-F(DL,h)_H + FLW(g)] \\
&= \mathbb{E}[(DF, h)_H L],
\end{aligned}$$

where the second line is obtained by the dominated convergence theorem and where $F^N \xrightarrow{L^2} F$.

Now it remains to show that the derivative is convergent in the $\mathbb{D}_{1,2}$ norm. For $N > n$, we have

$$\mathbb{E}[(DF^N, h)_H L] = \mathbb{E}\left[(n+1) I_n \left(\int_0^T f_{n+1}(.,t)h(t)dt\right) L\right],$$

which means the projection of $(DF, h)_H$ onto the n-th Wiener chaos is

$$I_n\left(\int_0^T f_{n+1}(.,t)h(t)dt\right).$$

Hence, if $\{e_i\}$ is an orthonormal basis of H, we obtain that,

$$\begin{aligned}\sum_{m=1}^{\infty} mm!||f_m||^2_{L^2([0,T]^m)} &= \mathbb{E}\left[\sum_{i=1}^{\infty}\sum_{m=0}^{\infty}\left(\int_0^T f_m(.,t)e_i(t)dt\right)^2\right] \\ &= \sum_{i=1}^{\infty}\mathbb{E}[(DF,e_i)^2_H] \\ &= ||DF||^2_{L^2([0,T]\times\Omega)} \\ &< \infty,\end{aligned}$$

which completes the proof. □

We still need a final ingredient, a chain rule to govern the differential operator under composition of maps; and Leibnitz rule to govern differentiation of products.

Theorem 3.3.27. ***(Chain rule)*** *Let $\phi : \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function with bounded partial derivatives. Suppose that $F = (F^1, ..., F^m)$ is a random vector, with $F^i \in \mathbb{D}_{1,p}$ for $i = 1, 2, ..., m$. Then, $\phi(F) \in \mathbb{D}_{1,p}$ and,*

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

Theorem 3.3.28. ***(Leibnitz Rule)*** *Let I be a subset of $\{t_1, ..., t_k\}$, and $|I|$ denote the cardinality of I. Then, we have*

$$D^k_{t_1,...,t_k}(FG) = \sum_{I\subset t_1,...,t_k} D^{|I|}_I(F) D^{k-|I|}_{I^c}(G).$$

The following corollary is a consequence of the chain rule and Leibnitz rule applied simultaneously.

Corollary 3.3.29. *Let $\phi \in C^\infty_p$, and $F \in \mathbb{D}_\infty$. Then, $\phi(F) \in \mathbb{D}_\infty$.*

The proof of the preceding theorems are identical to the case of ordinary calculus, as the Malliavin derivative for smooth random variables are defined via formal differentiation, and the closable property allows us to approximate the Malliavin derivative of an arbitrary L^p random variable by that of smooth random variables.

3.4 The Skorohod Integral

In this section we consider the dual operator of the Malliavin derivative D^*, and we will primarily focus on the case for D^* acting on $L^2(\Omega)$. An interesting property of the dual is that it actually coincides with the Itô integral in the sense that $D^*_t X = \int_0^t X_s dW_s$, for Itô integrable processes X. Moreover, the Itô integrable processes forms a proper subset of the domain of D^*. One could view D^* as a generalisation of the Itô integral, and hence it has been given the name the Skorodhod integral. The Skorodhod integral and the Malliavin derivative are related by the integration by parts relation, essentially a generalised statement of lemma 3.3.15. The integration by parts formulae have some quite significant impacts in many areas of applications that will be described throughout chapters 4, 5 and 6.

Let $p, q > 1$ be such that $p^{-1} + q^{-1} = 1$. The Malliavin derivative D is closed and has domain on a dense subset of $L^p(\Omega)$, so its dual D^* should also be closed but with domain contained in $L^q(\Omega)$. In this section, we give a detailed treatment of the case $p = q = 2$ via the Itô-Wiener chaos expansions. In particular, we will show how D^* coincides with the Itô integral defined in chapter 2 for processes that are adapted to the Wiener filtration. We leave the case of a general p until the next section.

Definition 3.4.1. (Skorohod Integral) We denote the adjoint of the operator D as D^*, so D^* is an unbounded operator on $L^2(T \times \Omega)$ with values in $L^2(\Omega)$, such that,

1. The domain of D^*, denoted by $\mathbb{D}^*$ is the set of processes $\xi \in L^2(T \times \Omega)$ such that
$$\left| \int_{[0,T]} D_t F \xi_t dt \right| = |\mathbb{E}(D_t F, \xi)_{L^2[0,T]}| \le c||F||_2,$$
for all $F \in \mathbb{D}_{1,2}$, and c is a constant independent of ξ.
2. If $\xi \in \mathcal{D}^*$, then $D^*(\xi) \in L^2(\Omega)$ and satisfies
$$\mathbb{E}[(D_t\phi, F)_{L^2[0,T]}] = \mathbb{E}(\phi, (D^*F)_t).$$
for any $F \in \mathcal{D}_{1,2}$.

Remark 3.4.2.

1. The operator D^* transforms square integrable processes back to random variables. Hence, D^* is the dual of D just in the T-direction.
2. The second equality mentioned above is called the **integration by parts relation**. It is the key to many applications in Malliavin calculus.

□

We now turn to the chaos expansion of L^2 random variables to study some properties of D^*.

Theorem 3.4.3. *Let $\xi \in L^2(T \times \Omega)$ with expansion as in theorem 3.2.10. Then,*
$$D^*\xi = \sum_{m=0}^{\infty} I_{m+1}(\hat{f}_m)$$

converges in $L^2(\Omega)$, and $\hat{f}_m$ is the symmetrisation of f_m in $(m+1)$-dimensions, defined by

$$\hat{f}_m(t_1, ..., t_m, t) = \frac{1}{m+1}\bigg(f_m(t_1, ..., t_m, t) + \sum_{i=1}^{m} f_m(t_1, ..., t_{i-1}, t, t_{i+1}, ..., t_m, t_i) \bigg).$$

Remark 3.4.4. Intuitively, D^* increases the level of Wiener chaos by one degree at a time. Hence, we need to replace f_m with $\hat{f_m}$ so $\hat{f} \in \mathrm{Dom}(I_{m+1})$. □

Proof. We have the following lemma.

Lemma 3.4.5. *Let $\xi \in L^2(T \times \Omega)$, by virtue of the chaos expansion theorem, there exists a family of deterministic functions $f_m(t_1, ..., t_m, t) \in L^2([0,T]^{m+1})$ such that every f_m is symmetric in the first m variables and*

$$\xi_t = \sum_{m=0}^{\infty} I_m(f_m(., t)),$$

where convergence is taken place in $L^2(T \times \Omega)$ and

$$\mathbb{E}(||\xi||^2_{L^2[0,T]}) = \mathbb{E}\bigg(\int_{[0,T]} \xi(t)^2 dt \bigg) = \sum_{m=0}^{\infty} m! ||f_m||^2_{L^2([0,T]^{m+1})}.$$

Proof. The is an immediate consequence of the chaos decomposition theorem (theorem 3.2.10). □

Now we prove the theorem. First consider $G = I_n(g)$ for some symmetric function g and $n \geq 1$. Applying Fubini's theorem and then Itô's isometry, we have that

$$\begin{aligned}
\mathbb{E}(\xi_t, D_t G)_{L^2[0,T]} &= \sum_{m=0}^{\infty} \mathbb{E}(I_m(f_m(.,t)), nI_{n-1}(g(.,t))_{L^2[0,T]} \\
&= \mathbb{E}(I_{n-1}(f_{n-1}(.,t)), nI_{n-1}(g(.,t))_{L^2[0,T]} \\
&= \mathbb{E}\left(n\int_{[0,T]} I_{n-1}(f_{n-1}(.,t))I_{n-1}(g(.,t))dt\right) \\
&= \int_{[0,T]} n\mathbb{E}[I_{n-1}(f_{n-1}(.,t))I_{n-1}(g(.,t))]dt \\
&= n(n-1)! \int_{[0,T]} (f_{n-1}(.,t), g(.,t))_{L^2[0,T]^{n-1}} dt \\
&= n!(\hat{f}_{n-1}, g)_{L^2[0,T]^n} \\
&= \mathbb{E}(I_n(\hat{f}_{n-1})I_n(g)) \\
&= \mathbb{E}(I_n(\hat{f}_{n-1}), G)_{L^2[0,T]}.
\end{aligned}$$

Hence, for every $\xi \in \mathrm{Dom}D^*$, the above computation shows that

$$\mathbb{E}(D^*(\xi)G) = \mathbb{E}(I_n(\hat{f}_{n-1})(G)$$

for every G of the form $G = I_n(g)$. Thus, $I_n(\hat{f}_{n-1})$ coincides with the projection of $D^*(\xi)$ onto the n-th Wiener chaos. Consequently, we have

$$\sum_{m=0}^{\infty} I_{m+1}(\hat{f}_m) \xrightarrow{L^2(\Omega)} D^*\xi.$$

Conversely, if the above series converges and we denote its limit by S. The preceding computation gives

$$\mathbb{E}\left(\int_{[0,T]} \xi_t D_t\left(\sum_{n=0}^{N} I_n(g_n)\right)dt\right) = \mathbb{E}\left(V\sum_{n=0}^{N} I_n(g_n)\right),$$

for all $N \geq 0$, and hence

$$\left|\mathbb{E}\int_{[0,T]} \xi_t D_t F dt\right| \leq ||V||_{L^2(\Omega)}||F||_{L^2(\Omega)},$$

for any random variable F with a finite chaos decomposition. But such a set is dense in $L^2(\Omega) \supset \mathbb{D}_{1,2}$, and hence we conclude that $\xi \in \mathrm{Dom}D^*$. $\square$

Corollary 3.4.6. *The domain of D^* coincides with the subspace of $L^2([0,T]\times\Omega)$ formed by processes that satisfies*

$$\sum_{m=0}^{\infty}(m+1)!||\hat{f}_m||^2_{L^2[0,T]^{m+1}} < \infty.$$

Corollary 3.4.7. *Let $F \in L^2(\Omega)$ (so F is constant in time), then*

$$D^*F = \int_0^T F dW_t = FW_T,$$

and consequently,

$$D^*F1_{a,b} = F(W_b - W_a)$$

for $a < b \in \mathbb{R}$.

An immediate consequence of theorem 3.5.3 is that, if $\xi(t)$ a deterministic function, then $D^*\xi(t)$ will coincide with $\int_0^T \xi(t)dW_t$. The following theorem generalises this idea in the sense that the relation holds true for all square integrable adapted processes.

Theorem 3.4.8. *Let W_t be a Wiener process and ξ_t a square integrable process adapted to the Wiener filtration. Then, for all $t \leq T$,*

$$(D^*\xi)_t = \int_0^t \xi_s dW_s.$$

From this point, we write the Skorodhod integral as D^* or $\int dW_t$ interchangably.

Proof. Suppose first that ξ is an simple process of the form,

$$\xi_t = \sum_{j=1}^{n} \xi_j 1_{(t_j,t_{j+1}]}(t)$$

where ξ_j are square integrable random variables, and $0 \leq t_1 < ... < t_{n+1} \leq t$. Since $\xi_j 1_{(t_j,t_{j+1}]}$ are piecewise constant with respect to t, by corollary 3.4.7, we have

$$D^*\xi_t = \sum_{j=1}^{n} \xi_j (W_{t_{j+1}} - W_{t_j}).$$

Moreover, for a general square integrable adapted processes ξ, we can approximate it by simple processes ξ^n. Now, Since since D^* is closable, it follows that $D^*\xi^n_t \xrightarrow{L^2}$

$D^*\xi$. On the other hand we have

$$D^*\xi_t^n = \sum_{j=1}^{n} \xi_j (W_{t_{j+1}} - W_{t_j})(t) \xrightarrow{L^2} \int_0^T \xi_t dW_t.$$

By completeness of $L^2(\Omega)$, we conclude that

$$D^*\xi_t = \int_0^T \xi_t dW_t.$$

□

We now state the Clark-Ocone representation theorem, which can be viewed as some mixture of Itô's representation theorem and the stochastic fundamental theorem of calculus.

Theorem 3.4.9. *Let W_t be a one-dimensional Wiener process with natural filtration $\mathcal{F}_t$, and $F \in \mathbb{D}_{1,2}$. Then,*

$$F = \mathbb{E}F + \int_0^1 \mathbb{E}(D_t F|\mathcal{F}_t) dW_t.$$

Remark 3.4.10. Recall that Itô's martingale representation theorem states under certain conditions that for a square integrable F, there exists an adapted process f such that

$$F = \mathbb{E}F + \int_0^1 f_t dW_t.$$

The Clark-Ocone representation tells us exactly what Itô's mysterious f should be. It is the simply the projective image of $D_t F$ under some optional stopping times.

Indeed, we would have no hope of identifying f without the Malliavin-type of machineries which we have developed. This result is very useful in applications, since it replaces many purely existential arguments which are based on Itô's representation theorem, by constructive proofs. □

Proof. We may assume that $F \in \mathbb{D}_{1,2}$ has the form $F = \sum_m I_m(f_m)$. Then,

$$\begin{aligned}
\mathbb{E}(D_t F|\mathcal{F}_t) &= \sum_{m=1}^{\infty} m\mathbb{E}(I_{m-1}(f_m(.,t))|\mathcal{F}_t) \\
&= \sum_{m=1}^{\infty} m\mathbb{E}[I_{m-1}(f_m(t_1,...,t_{m-1},t)(1_{\{t_1\vee...\vee t_{m-1}\le t\}} + 1_{\{t_1\vee...\vee t_{m-1}\ge t\}}))|\mathcal{F}_t].
\end{aligned}$$

Now, $I_{m-1}(f_m(t_1,...,t_{m-1},t)1_{\{t_1\vee...\vee t_{m-1}\leq t\}})$ is $\mathcal{F}_t$-measurable, while by Itô's isometry,

$$\mathbb{E}[I_{m-1}f_m(t_1,...,t_{m-1},t)1_{\{t_1\vee...\vee t_{m-1}\geq t\}}|\mathcal{F}_t] = 0.$$

Hence, we have

$$\mathbb{E}(D_tF|\mathcal{F}_t) = \sum_{m=1}^{\infty} mI_{m-1}(f_m(t_1,...,t_{m-1},t)1_{\{t_1\vee...\vee t_{m-1}\leq t\}})$$

Letting $f_t = \mathbb{E}(D_tF|\mathcal{F}_t)$, we calculate D^*f using the above expression and theorem 3.5.3. We obtain

$$D^*f = \sum_{m=1}^{\infty} I_m(f_m) = F - \mathbb{E}F.$$

But now D^*f coincides with the Itô integral of f, and hence the proof is finished. □

We conclude this section by a final remark that summarises some further properties of D^*. The proofs are routine in the sense that you first check the results on the Wiener chaos, and conclude for a general L^2 random variable by a limiting argument. Precise details of these can be found in section 1.3 of [35].

Remark 3.4.11.

1. Suppose u is a Skorohod integrable process. Let $F \in \mathbb{D}_{1,2}$ such that

   $$\mathbb{E}\left(F^2\int_0^T u_t^2 dt\right) < \infty.$$

   Then we have,

   $$\int_0^T (Fu_t)dW_t = F\int_0^T u_t dW_t - \int_0^T (D_fF)u_t dt.$$

   In particular, this tells us that Fu_t is Skorohod integrable if and only if the right hand side belongs to $L^2(\Omega)$.
2. Heisenberg's commutation relation: If $F \in \mathbb{D}_{2,2}([0,T]\times\Omega)$, then $D^*F \in \mathbb{D}_{1,2}(\Omega)$ and $\forall 0 \leq t \leq T$, we have

   $$D_t(D^*F) = F_t + \int_0^T D_tF_s dW_s.$$

   It resembles the Heisenberg's relation in the sense that $DD^* - D^*D = 1$.

□

3.5 Quick Remark on Ornstein-Uhlenbeck Semigroups

Another well know operator in stochastic analysis is the Ornstein-Uhlenbeck operator. We will quickly go through its properties and state its relation with Malliavin calculus. Its action on $L^2(\Omega)$ is defined by

$$T_t(F) = \sum_{n=0}^{\infty} \frac{e^{-nt}}{n!} I_n(F),$$

where $F \in L^2(\Omega)$, and it is assumed to have Wiener chaos expansion,

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

It can be shown that (c.f. [35]),

1. The set $\{T_t, t \in \mathbb{R}_+\}$ form a Markov semigroup. In particular, we have $T_t T_s = T_{t+s}$ for all $s, t \in \mathbb{R}$.
2. We define its generator $\mathcal{L}$ to be such that

$$\mathcal{L}F = \lim_{t \to 0} \frac{T_t F - F}{t}$$

   in the sense of L^2. A remarkable fact about $\mathcal{L}$ is that $\mathcal{L}F = -D^* DF$.
3. Some authors such as [4] uses the Ornstein-Uhlenbeck generator to define the Malliavin derivative.
4. There are many other nice connections between the Ornstein-Uhlenbeck generator and the Malliavin calculus. However, we will pursue in a different direction, and turn to the integration by parts relation for the rest of this thesis.

Chapter 4

Existence and Smoothness of the Density

One of the most important applications of Malliavin calculus lies in the investigation of existence, smoothness as well as many other properties of densities of random variables that can be written as Brownian functionals via the integration by parts relation introduced towards the end of the last chapter. This was in fact the motivation for P. Malliavin to have developed such a machinery when it was first introduced in 1976 (see [31]). Malliavin's initial paper, was followed by a number of alternative developments on this theory. Interested readers may consult, [4], [3], [35] and [19]. We will take the approach that makes use of the integration by parts relation. This approach was originally introduced by Bismut and Michel in 1982, and it is one of the most popular approaches today (c.f. [1], [3], [35] and [41]).

4.1 Sufficient Conditions for Existence of Density

This chapter will be devoted to establishing various properties of the density based on the Malliavin matrix. As before, we let $W = \{W(h), h \in H\}$ be an isonormal Gaussian process associated to the Hilbert space $H = L^2([0,T], \mathcal{B}, \mu)$.

Definition 4.1.1. (The Malliavin Matrix) Let $F = (F^1, ..., F^m) \in \mathbb{D}_{1,2}$. Define,

$$\sigma_{ij} = (DF_i, DF_j)_H, 1 \le i, j \le m,$$

the matrix

$$\Sigma_F(\omega) = \begin{pmatrix} \sigma_F^{11}(\omega) & \sigma_F^{12}(\omega) & \dots & \sigma_F^{1m}(\omega) \\ \sigma_F^{21}(\omega) & \sigma_F^{22}(\omega) & \dots & \sigma_F^{2m}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_F^{m1}(\omega) & \sigma_F^{m2}(\omega) & \dots & \sigma_F^{mm}(\omega) \end{pmatrix}$$

is called the **Malliavin matrix**. If $\det \Sigma(\omega) > 0$, a.s., and $\det \Sigma(\omega)^{-1} \in L^p$ for some $p < \infty$, then Σ (and F itself) is called **non-degenerate**.

We begin with the following proposition, which is essentially a one-dimensional setting of the general case.

Proposition 4.1.2. *Let $F \in \mathbb{D}_{1,2}$, and suppose that $\frac{DF}{||DF||^2} \in \mathbb{D}^*_{1,2}$. Then the law of F has a continuous and bounded density given by*

$$f(x) = \mathbb{E}\left(1_{F>x} D^* \left(\frac{DF}{||DF||_H^2}\right)\right).$$

Proof. Let $a < b$, and consider the functions $\psi(y) = 1_{[a,b]}(y)$ and $\varphi(y) = \int_{-\infty}^{y} \psi(z)dz$. Clearly, $\varphi(F) \in \mathbb{D}_{1,2}$, and by the chain rule, we have

$$(D(\varphi(F)), DF)_H = \psi(F)(DF, DF)_H = \psi(F)||DF||_H^2.$$

Now, using the integration by parts relation,

$$\begin{aligned}\mathbb{E}(\psi(F)) &= \mathbb{E}\left[\left(D(\varphi(F)), \frac{DF}{||DF||_H^2}\right)_H\right] \\ &= \mathbb{E}\left[\varphi(F) D^* \left(\frac{DF}{||DF||_H^2}\right)\right].\end{aligned}$$

Hence applying Fubini's theorem, we obtain

$$\begin{aligned}\mathbb{P}(a \le F \le b) &= \mathbb{E}\left[\int_{-\infty}^{F} \psi(x)dx D^* \left(\frac{DF}{||DF||_H^2}\right)\right] \\ &= \int_a^b \mathbb{E}\left[1_{F>x} D^* \left(\frac{DF}{||DF||_H^2}\right)\right] ds,\end{aligned}$$

which gives the desired result. □

Remark 4.1.3. The sufficient conditions for $\frac{DF}{||DF||^2} \in \mathbb{D}^*_{1,2}$ are that

1. $F \in \mathbb{D}_{2,4}$, and
2. $\mathbb{E}(||DF||^{-8}) < \infty$.

□

To generalise the above proposition to higher dimensions, we need the following result from harmonic analysis.

Proposition 4.1.4. *Let μ be a probability measure on $\mathbb{R}^m$. Assume that for all $\varphi \in C_b^\infty(\mathbb{R}^m)$, the following inequality holds,*

$$\left|\int_{\mathbb{R}^m} \partial_j \varphi d\mu\right| \le c_j ||\varphi||_\infty, \quad 1 \le i \le m,$$

where the c_j's are constants that do not depend on φ. Then, μ is absolutely continuous with respect to the Lebesgue measure.

The most popular method of proving this involves taking Fourier transforms, and hence it is considered a result of harmonic analysis. Readers are advised to see [31] for details.

Theorem 4.1.5. *Let $F = (F^1, ..., F^m)$ be a random vector satisfying the assumptions,*

1. *$F^i \in \mathbb{D}_{2,4}$ for all $i, j = 1, ..., m$.*
2. *The matrix Σ_F is invertible a.s.*

Then, the law of F is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$.

Proof. Let $\varphi \in C_b^\infty(\mathbb{R}^m)$ be a fixed test function. By the chain rule, we know that $\varphi(F) \in \mathbb{D}_{1,4}$, and that

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

Hence,

$$(D(\varphi(F)), DF^j)_H = \sum_{i=1}^{m} \partial_i \varphi(F) \Sigma_F^{ij}$$

i.e.

$$\begin{pmatrix} \partial_1 \varphi(F) \\ \partial_2 \varphi(F) \\ \vdots \\ \partial_m \varphi(F) \end{pmatrix} = \begin{pmatrix} \sigma_F^{11}(\omega) & \sigma_F^{12}(\omega) & \dots & \sigma_F^{1m}(\omega) \\ \sigma_F^{21}(\omega) & \sigma_F^{22}(\omega) & \dots & \sigma_F^{2m}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_F^{m1}(\omega) & \sigma_F^{m2}(\omega) & \dots & \sigma_F^{mm}(\omega) \end{pmatrix}^{-1} \begin{pmatrix} (D(\varphi(F)), DF^1)_H \\ (D(\varphi(F)), DF^2)_H \\ \vdots \\ (D(\varphi(F)), DF^m)_H \end{pmatrix}.$$

In order to apply proposition 4.1.4, we need to deal with a potential integrability problem of Σ_F^{-1}. To this end, we use a localising argument.

For any integer $N > 1$, we consider a function $\Psi_N \in C_0^\infty(\mathbb{R}^m \otimes \mathbb{R}^m)$ such that

$$\Psi_N(\Xi) = \begin{cases} 1, & \text{if } \Xi \in K_N; \\ 0, & \text{if } \Xi \notin K_{N+1}. \end{cases}$$

where,

$$K_N = \left\{ \Xi \in \mathbb{R}^m \otimes \mathbb{R}^m : |\Xi^{ij}| \leq N \forall i, j, \text{ and } |\det \Xi| \geq \frac{1}{N} \right\}.$$

Note that K_N is a compact subset of $GL_m \subset \mathbb{R}^m \otimes \mathbb{R}^m \simeq \mathrm{End}(\mathbb{R}^m, \mathbb{R}^m)$. Now, multiplying Ψ_N to the previous matrix equation we get for each i,

$$|\mathbb{E}(\Psi_N(\Sigma_F) \partial_i \varphi(F))| = \sum_{j=1}^{m} \mathbb{E}\left(\Psi_N(\Sigma_F)(D(\varphi(F)), DF^j)_H (\Sigma_F^{-1})^{ji}\right).$$

Now, the second assumption gives us the invertibility of Σ_F, which implies that $G = \Psi_N(\Sigma_F)(\Sigma_F^{-1})^{ji} \in \mathbb{D}_{1,2}$. Moreover, G is bounded and the first assumption gives us $(DG, DF^j)_H \in L^2([0,T])$. By property 1 of remark 3.4.11, this implies that $\Psi_N(\Sigma_F)(\Sigma_F^{-1})^{ji} DF^j \in \mathbb{D}^*$, and hence we may apply integration by parts to get,

$$\begin{aligned}\Big|\mathbb{E}(\Psi_N(\Sigma_F)\partial_i\varphi(F))| &= |\mathbb{E}\left(\varphi(F)\sum_{j=1}^m D^*(\Psi_N(\Sigma_F)(\Sigma_F^{-1})^{ji}DF^j)\right)\Big| \\ &\leq \mathbb{E}\left(\left|\sum_{j=1}^m \left|D^*(\Psi_N(\Sigma_F)(\Sigma_F^{-1})^{ji}DF^j)\right|\right|\right)||\phi||_\infty \qquad < \infty.\end{aligned}$$

Therefore, by proposition 4.1.4, the measure $(\Psi_N(\gamma_F).P)\circ F^{-1}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$. Thus for any Borel set $A \in \mathbb{R}^m$ with Lebesgue measure zero, we have

$$\int_{F^{-1}(A)} \Psi_N(\Sigma_F)d\mathbb{P} = 0.$$

Letting $N \to \infty$ and using the second assumption, we can establish that $\mathbb{P}(F^{-1}(A)) = 0$, and thereby proving that $\mathbb{P}\circ F^{-1}$ is absolutely continuous with respect to the Lebesgue measure. □

4.2 Sufficient Conditions for Smoothness of Density

We extend the argument given in the previous section to deduce a sufficient condition for smoothness of density of a $\mathbb{R}^m$ valued random variable. More specifically, we will prove the following theorem.

Theorem 4.2.1. *Let $F = (F^1, ..., F^m)$, such that $F^i \in \mathbb{D}_{1,2}$, and satisfying the following assumptions,*

1. *$F^i \in \mathbb{D}^\infty$ for all $i = 1, ..., m$.*
2. *The Malliavin matrix Σ_F satisfies*

$$(\det \Sigma_F)^{-1} \in \bigcap_{p>1} L^p(\Omega).$$

Then, F has an infinitely differentiable density.

Before we begin the proof, we need to first state a lemma. It is a generalisation to proposition 4.1.4, a standard result in harmonic analysis that we will not prove. Interested readers are directed to see [31] for details.

Lemma 4.2.2. *Let μ be a probability measure in $\mathbb{R}^m$, and fix an open set $A \subset \mathbb{R}^m$. If for all $\varphi \in C_A^\infty(\mathbb{R}^m)$, and multi-index $\alpha = (\alpha_1, ..., \alpha_k)$, there exists a constant C_α independent of φ such that,*

$$\left|\int_{\mathbb{R}^m} \partial_\alpha \varphi d\mu\right| \leq C_\alpha ||\varphi||_\infty,$$

where C_A^∞ is the set of smooth functions compactly supported on a set $K \subset A$.

Proof. We first prove that $\det \Sigma_F^{-1} \in \mathbb{D}_\infty$. Let,

$$Y_n = \left(\det \Sigma^{-1} + \frac{1}{n}\right)^{-1}$$

for $n = 1, 2, ...$ We have assumed that $(\det \Sigma_F)^{-1} \in \bigcap_{p>1} L^p(\Omega)$, and hence $Y_n \to \det \Sigma_F^{-1}$ in $L^p(\Omega)$. Clearly, $\det \Sigma_F \in \mathbb{D}_\infty$. Observe that the functions $\zeta_n(x) = \left(x + \frac{1}{n}\right)^{-1} \in C_p^\infty$ for $x > 0$. Then by corollary 3.3.30, we conclude that $\zeta_n(\det \Sigma_F) = Y_n \in \mathbb{D}_\infty$ for all n. On the other hand, the sequence Y_n converges to a limit in L^p, and the operator D^k is closed for all k. Hence, $D^k Y_n \to D^k \det \Sigma^{-1}$ for all k, and therefore $\det \Sigma^{-1} \in D_\infty$.

Now we prove the theorem. The main direction of the proof is to construct an upper bound for

$$\left|\mathbb{E}\left(\frac{\partial^k}{\partial x_{\alpha_1} ... \partial x_{\alpha_k}}(F)\right)\right|$$

so that lemma 4.2.1 can be applied.

Let $\varphi \in C^\infty(\mathbb{R}^m)$ with compact support contained in A. By the chain rule, we obtain

$$(D(\varphi(F)), DF^j)_H = \sum_{i=1}^m \partial_i \varphi(F)(DF^i, DF^j)_H = \sum_{i=1}^m \partial_i \varphi(F) \Sigma_A^{ij}.$$

Treat the above as a system of linear equations in the $\partial_i \varphi(F)$'s. Solving the system we obtain,

$$\partial_i \varphi(F) = \sum_{j=1}^m (D(\varphi(F)), DF^j)_H (\Sigma_A^{-1})^{ji}.$$

Let R be a fixed element in $\mathbb{D}_\infty$, and using integration by parts relation we get,

$$\begin{aligned} \mathbb{E}(R(\partial_i \varphi)(F)) &= \sum_{j=1}^m \mathbb{E}[R(D(\varphi(F)), (\Sigma_A^{-1})^{ji} DF^j)_H] \\ &= \mathbb{E}[\varphi(F)\Phi_i(R)] \end{aligned}$$

where

$$\Phi_i(R) = \sum_{j=1}^m D^* \left(R u_A^j (\Sigma_A^{-1})^{ji}\right).$$

We have shown in the beginning of the proof that, $(\Sigma_A^{-1})^{ji} \in \mathbb{D}_\infty$. Consequently, since R and DF^j are assumed to be in $\mathbb{D}_\infty$, it follows that $\Phi_i(R) \in \mathbb{D}_\infty$, and consequently Φ is a linear functional of R. Define the multi-index $\alpha = (\alpha_1, ..., \alpha_k)$, where $\alpha_p \in \{1, ..., m\}$ for all $p = 1, ..., k$. Recursively applying the relationship

$$\mathbb{E}(R(\partial_i \varphi)(F)) = \mathbb{E}(\varphi(F)\Phi_i(R))$$

to

$$R = \{1, \Phi_{\alpha_1}(1), \Phi_{\alpha_2}(\Phi_{\alpha_1}(1)), ..., \Phi_{\alpha_{k-1}}(\Phi_{\alpha_{k-2}}(...(1)...))\},$$

we obtain

$$\begin{aligned}\left|\mathbb{E}\left(\frac{\partial^k}{\partial x_{\alpha_1}...\partial x_{\alpha_k}}(F)\right)\right| &= |\mathbb{E}(\varphi(F)\Phi_{\alpha_{k-1}}(\Phi_{\alpha_{k-2}}(...(1)...)))| \\ &\leq ||\phi||_\infty |\mathbb{E}[\varphi(F)\Phi_{\alpha_{k-1}}(\Phi_{\alpha_{k-2}}(...(1)...))]| \\ &\leq ||\phi||_\infty C_\alpha,\end{aligned}$$

where we know $|\mathbb{E}[\varphi(F)\Phi_{\alpha_{k-1}}(\Phi_{\alpha_{k-2}}(...(1)...))]| < \infty$ as Φ was shown to be a linear functional. Finally, the theorem holds upon applying lemma 4.2.1. □

Remark 4.2.3.

1. In the finite-dimensional setting, one could formally express the density of F by $f(x) = \mathbb{E}(\delta_x \circ F)$. S.Watanabe gave a rigorous interpretation of the above statement in an infinite dimensional setting (via Malliavin derivatives), and he was able to deduce an identical result as the preceding theorem. This approach was illustrated in detail in section 2.4 of [19].
2. As we shall see in the next chapter, a particular interest of studying stochastic differential equations is to determine the behaviour of the density of the underlying solution. The results developed in this chapter serves as powerful tools in dealing with such classes of problems.

□

Chapter 5

Stochastic Differential Equations and Stochastic Flows

5.1 Introduction

5.1.1 Formal Definitions

Stochastic differential equations (SDEs) arise naturally in many problems of practice ranging from quantum mechanics to mathematical finance. Philosophically speaking, whenever we have imperfect information we can expect randomness of some degree that perturbs our observations. To include such random behaviour in our model, intuitively, the differential equation that governs the motion of these things would take the form

$$\frac{dX_t}{dt} = \mu(t, X_t) + \sigma(t, X_t)\frac{dW_t}{dt}.$$

But of course, $\frac{dW_t}{dt}$ is undefined with probability one. An alternative approach would be to re-write the above equation in an integral form, where

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$$

where the latter integral is in the sense of Itô.

Definition 5.1.1. Let μ and σ be Borel-measurable functions, with values in $\mathbb{R}^m$ and $\mathbb{R}^m \otimes \mathbb{R}^d$ respectively. A **solution to the stochastic differential equation** is a pair (X, W) of adapted processes defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, such that

1. W is a standard $\mathcal{F}_t$-Wiener process in $\mathbb{R}^d$.
2. X_t satisfies

$$\begin{cases} X_t = X_0 + \int_0^t \mu(s, X_s)dW_s + \int_0^t \sigma(s, X_s)ds \quad \text{(SDE)} \\ X_0 = x, \end{cases}$$

The above equation is sometimes written in the differential form:

$$dX_t = \mu(s, X_s)dW_s + \sigma(s, X_s)ds.$$

We say that the function μ is the **drift coefficient** and σ the **diffusion coefficient**, for historical reasons that the original motivation of studying SDEs was to model physical diffusions. The process X is sometimes also termed as a **diffusion** driven by W. When there is no risk of confusion, we simply say X is the solution to (SDE) instead of the pair (X, W).

Definition 5.1.2. A solution X of (SDE) on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ is said to be a **strong solution** if X is adapted to the filtration $\mathcal{F}_t^W$. A solution which is not strong will be termed a **weak solution**.

Example 5.1.3. (Ornstein-Uhlenbeck process) The Ornstein-Uhlenbeck process X_t is defined by the following SDE:

$$dX_t = aX_t dt + \sigma dW_t$$

$$X_0 = x.$$

We wish to find an explicit formula for X_t that depends only on W and t. The first equation can be written as

$$dX_t - aX_t dt = \sigma dW_t.$$

We multiply through by the integrating factor e^{-at} to get

$$\begin{aligned} e^{-at}\sigma dW_t &= e^{-at}(dX_t - aX_t dt) \\ &= d(e^{-at}X_t) \end{aligned}$$

Therefore,

$$e^{-at}X_t = X_0 + \int_0^t e^{as}\sigma dW_s$$

and so

$$X_t = e^{at}x + \int_0^t e^{a(t-s)}\sigma dW_s.$$

This gives the martingale representation of the Ornstein-Uhlenbeck process. □

Remark 5.1.4. Most SDEs we encounter are unlikely to have closed form solutions like the Ornstein-Uhlenbeck process. In fact, most often diffusion processes are defined by the SDE which it satisfies, rather than an explicit formula. For interest of the reader, Section 6.1 of [36] has a section discussing various types of SDEs with explicit form solutions. □

Theorem 5.1.5. ***(Existence and Uniqueness of Solution)*** *If μ and σ satisfies the* ***Lipschitz condition****, that is if*

$$||\mu(x_1, y_1) - \mu(x_2, y_2)|| + ||\sigma(x_1, y_1) - \sigma(x_2, y_2)|| \leq C(||x_1 - x_2|| + ||y_1 - y_2||),$$

$\forall x_1, y_1 \in \mathbb{R}$ and $\forall x_2, y_2 \in \mathbb{R}^m$. Then, (SDE) has a unique solution X_t adapted to the filtration $\mathcal{F}_t = \sigma(W_t)$, where uniqueness is in the sense of L^2.

Proof. Let E be the set of square integrable adapted processes, such that

$$||X||_E = \mathbb{E}\left(\left|\int_0^t X_s ds\right|\right) + \left[\mathbb{E}\left(\int_0^t X_s dW_s\right)^2\right]^{\frac{1}{2}} < \infty.$$

Then, it is easily verified that $||.||_E$ is a well defined norm, and hence $(E, ||.||_E)$ is a normed linear space. Let $X_t^0 = x$, and for each $n \in \mathbb{N}$, we carry out the Picard iterations as follows. Define,

$$C(X_t^n) = X_0^n + \int_0^t \mu(s, X_s^n) ds + \int_0^t \sigma(s, X_s^n) ds$$

and $X^{n+1} = C(X^n)$. Then it can be shown that under Lipschitz conditions, C is a contraction mapping. The proof can be found in many texts such as [10], [38] and [40]. Hence, by the contraction mapping theorem, there exists a unique point X, such that $X^n \to X$ in the norm $||.||_E$. □

Corollary 5.1.6. *(Markov Property)* *Let X_t be a solution of (SDE) with μ and σ being $\mathcal{F}_t$ adapted Lipchitz functions. Then, X_t satisfies the Markov property, that*

$$\mathbb{E}(\phi(X_t)|\mathcal{F}_s) = \mathbb{E}(\phi(X_t)|\sigma(W_s)).$$

for all functions ϕ such that the above expectation is well defined. This is an easy consequence of uniqueness of solutions.

Remark 5.1.7. The Lipchitz condition is sometimes considered to be too restrictive. Some very innocent looking SDE's like

$$dX_t = X_t^2 dt + X_t^3 dW_t$$

has solution

$$X_t = \frac{1}{1 - W_t}$$

which means the behaviour of X will become unstable in finite time with probability one, as $\mathbb{P}(W_t = 1 \text{ in finite time }) = 1$. □

5.1.2 Connections with Partial Differential Equations

In this section, I shall introduce a surprising connection between SDEs and PDEs. Quite often the stochastic method actually provides an easier route than solving the PDE directly.

Consider the time-homogeneous m-dimensional SDE, driven by a d-dimensional Wiener process defined as follows,

$$\begin{cases} dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad \text{(SDE1)} \\ X_0 = x, \end{cases}$$

where $\sigma = (\sigma_{ij})$ is a $m \times d$ matrix. Applying Itô's lemma to $f(X_t)$, for some $f \in C^2$ to get

$$\begin{aligned} f(X_t) - f(x) &= \int_0^t \sum_{j=1}^m \mu^j(X_s) \frac{\partial f}{\partial x^j}(X_s) dX_s + \frac{1}{2} \int_0^t \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s) d\langle X^i, X^j \rangle_s \\ &= \int_0^t \left(\sum_{j=1}^m \mu^j(X_s) \frac{\partial f}{\partial x^j}(X_s) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^d \sigma_{ik}\sigma_{kj} \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s) \right) d\langle W, W \rangle_s \\ &+ \int_0^t \sum_{j=1}^m \sigma_{ij}(X_s) \frac{\partial f}{\partial x^j}(X_s) dW_s \\ &= \int_0^t Af(X_s) ds + \int_0^t \sum_{j=1}^m \sigma_{ij}(X_s) \frac{\partial f}{\partial x^j}(X_s) dW_s \end{aligned}$$

where A is called the infinitesimal generator associated with the SDE $e(\mu(X_t), \sigma(X_t))$ defined by,

$$A = \sum_{j=1}^m \mu^j(X_s) \frac{\partial}{\partial x^j}(X_s) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m a_{ij} \frac{\partial^2}{\partial x^i \partial x^j}(X_s)$$

where

$$a_{ij} = \sum_{k=1}^d \sigma_{ik}\sigma_{kj}.$$

Now, taking expectations on both sides and differentiate with respect to t, we get

$$\frac{\partial}{\partial t} \mathbb{E} f(X_t) = A \mathbb{E} f(X_t).$$

Hence we deduce that if we let $u(x,t) = \mathbb{E}(f(X_t)|X_0 = x)$, then u satisfies the Cauchy problem,

$$\begin{cases} \frac{\partial u}{\partial t} = Au \\ u(x,0) = f(x). \end{cases}$$

The above approach can be generalised to solve the Schröndinger's equation, a wave equation that governs quantum mechanical motion:

$$-\frac{\hbar}{2m} \nabla^2 \psi(x,t) + V(x)\psi(x,t) = i\hbar \frac{\partial}{\partial t} \psi(x,t),$$

where $x = (x_1, x_2, x_3) \in \mathbb{R}^3$, $\bar{h}$ is the normalised Plank's constant and m is the mass of the particle. In 1947, Richard Feynman introduced a path integral approach to express solutions to the above problem at an intuitive level, it was not until 1965 when Kac had made this mathematically rigorous. We can assure that physical measurements are made to that the constants are all one, and so I shall ignore all constants that appear in the equation. We will be solving the problem,

$$\begin{cases} \frac{\partial u}{\partial t} = Lu, & \text{on } \mathbb{R}^d; \\ u(0,x) = f(x), & \text{on } \partial\mathbb{R}^d. \end{cases}$$

where

$$L = A + v(X_t) = \frac{1}{2}\nabla^2 + v(X_t).$$

Theorem 5.1.8. *(Feynman-Kac Representation) Let $u \in C^{1,2}(\mathbb{R}\times\mathbb{R}^d)$ be a solution of the above initial value problem, and W_t^x be a translated Wiener process on $\mathbb{R}^d$, so that $W_0^x = x$. Then,*

$$u(t,x) = \mathbb{E}^x\Big(f(X_t)\exp\Big(\int_0^t v(W_s)ds\Big)\Big).$$

where X satisfies the SDE $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$.

The proof of this resembles very similar ideas to the case of the Cauchy problem. Readers are referred to chapter 7 of [38] for a detailed argument.

5.2 Stochastic Flows and Malliavin Calculus

Let us remind ourselves of SDE1 defined in section 5.1.1,

$$\begin{cases} X_t = x + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dW_s, & \text{(SDE1)} \\ X_0 = x, \end{cases}$$

where $\sigma(X_s)$ is an $m \times d$ matrix, X_t and $\mu(X_s)$ are m dimensional vectors, and W_s is a d-dimensional Wiener process. The study of stochastic flows is about studying the map $\phi : (x,t,\omega) \to \mathbb{R}^m$, where $\phi(x,t,\omega) = X_t(\omega)$, where X is the process that solves the above SDE. In particular, we are interested in looking at how ϕ behaves under differentiation. Obviously,

$$\frac{\partial}{\partial t}\phi(x,t,\omega) = \frac{d}{dt}X_t$$

is undefined $\mathbb{P}$ a.s., one of the first properties that was known about solutions to SDE's. However, it turns out that both $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial \omega}$ turns out to be well defined quantities. If we assume both μ and σ are C^∞ functions, with bounded first partial derivatives, it can be shown that the map $\phi(.,t,\omega) : \mathbb{R}^m \to \mathbb{R}^m$ is a diffeomorphism for every fixed t and ω. On the other hand, $\frac{\partial}{\partial \omega}$ corresponds to, in the weak sense, of the Malliavin derivative D_sX_t. If we assume again that $\mu, \sigma \in C^\infty$, then it can be shown the solution to the SDE, $X \in \mathbb{D}_\infty$, confirming the existence of D_sX_t.

We shall devote this section in proving these results under the assumption that $\mu, \sigma \in C^\infty$. [35] give a more general treatment to higher order derivatives, [39] and [29] illustrates analogous results for the case when μ and σ are only assumed to be Lipschitz.

Theorem 5.2.1. *Let $\phi : (x,t,\omega) \to X_t(\omega)$ where X is a process satisfying (SDE1) with $X_0 = x$. Then, for almost every (t,ω), the function $\phi(.,t,\omega)$ is a C^∞-homeomorphism from $\mathbb{R}^m$ to $\mathbb{R}^m$.*

Theorem 5.2.2. *Let X satisfies (SDE1), and $\phi = \phi(x,t,\omega)$ as before. For $p \geq 2, T > 0, k \in \mathbb{N}$ and $R > 0$, $\exists C = C(p,T,k,R)$ such that*

$$\sup_{|x|\leq R} \mathbb{E}\left(\sup_{0\leq t\leq T} |\partial_\alpha \phi(x,t,\omega)|^p\right) \leq C,$$

where $\alpha = (\alpha_1, ..., \alpha_m), |\alpha| = \sum_1^m \alpha_m, \partial_\alpha = \partial_1^{\alpha_1}...\partial_n^{\alpha_m}$ are partial derivatives with respect to x. Moreover, for $t \geq 0$, let

$$J_t := J(x,t,\omega) = (\partial_j X^i(x,t,\omega))_{1\leq i,j,\leq m}$$

be the Jacobian of X with respect to x. Then, J_t and J_t^{-1} respectively satisfies the following SDE's,

$$J_t = I + \int_0^t A_0^{(1)}(X_s) J_s ds + \sum_{k=1}^m \int_0^t A_k^{(1)}(X_s) J_s dW_s^k,$$

$$J_t^{-1} = I - \int_0^t J_s^{-1}\left(A_0^{(1)}(X_s) - \sum_{k=1}^m (A_k^{(1)}(X_s))^2\right) ds - \sum_{k=1}^d \int_0^t J_s^{-1} A_k^{(1)}(X_s) dW_s^k$$

where I is the $m \times m$ identity matrix, $A_0^{(1)} := (\partial_j b^i(x))_{1\leq i,j\leq m}$ and $A_k^{(1)}(x) = (\partial_j \sigma_k^i(x))_{1\leq i,j\leq m}, k = 1,2,...,m$.

The preceding two theorems are regarded as well known and their proofs are available in [21], [29] (chapter 4) and [39] (chapter 7). Now we consider the derivative of X with respect to the "sample paths", ω; this correspond to the weak derivatives in the sense of Malliavin.

Theorem 5.2.3. *If $\mu, \sigma \in C^\infty$ in (SDE1), with bounded partial derivatives of all orders, then its unique solution $X = X(x,t,\omega) \in \mathbb{D}_\infty(\mathbb{R}^m), \forall x \in \mathbb{R}^m, t > 0$, and its Malliavin matrix $\Sigma_t := \Sigma(x,t,\omega)$ is given by*

$$\Sigma_t = J_t \left[\int_0^t J_s^{-1} a(X_s)(J_s^{-1})^* ds\right] J_t^*$$

where $a = \sigma\sigma^*$, and J_t is the Jacobian of X_t with respect to the initial value x.

Proof. By the Heisenberg's commutation relation,

$$\begin{aligned} D_s X_t &= \int_s^t D_s b(X_r)dr + \sigma(X_s) + \int_{k=1}^d \int_s^t D_s \sigma_k(X_r) dW_r^k \\ &= \int_s^t A_0^{(1)}(X_r) D_s X_r dr + \sigma(X_s) + \sum_{k=1}^d \int_s^t A_k^{(1)}(X_r) D_s X_r dW_r^k, \end{aligned}$$

where σ_k is the k-th column of the matrix σ, for $k = 1, ..., d$. On the other hand, by theorem 5.3.1 and orthogonality of stochastic integrals, we get

$$J_t J_s^{-1} = I + \int_s^t A_0^{(1)}(X_r) J_t J_s^{-1} dr + \sigma(X_s) + \sum_{k=1}^d \int_s^t A_k^{(1)}(X_r) J_t J_s^{-1} dW_r^k,$$

and hence, multiplying through by $\sigma(X_s)$, we get

$$J_t J_s^{-1}\sigma(X_s) = \sigma(X_s) + \int_s^t A_0^{(1)}(X_r) J_t J_s^{-1}\sigma(X_s) dr + \sigma(X_s) + \sum_{k=1}^d \int_s^t A_k^{(1)}(X_r) J_t J_s^{-1}\sigma(X_s) dW_r^k.$$

Observe that $J_t J_s^{-1}\sigma(X_s)$ and $D_s X_t$ are satisfied by the same SDE and initial conditions. Hence, by uniqueness of solution, we conclude that

$$D_s X_t = J_t J_s^{-1} \sigma(X_s) 1_{[0,t]}(s), a.s.$$

Therefore,

$$\begin{aligned} \Sigma_t &= \int_0^t (D_s X_t)(D_s X_t)^* ds \\ &= J_t \left[\int_0^t J_s^{-1} a(X_s)(J_s^{-1})^* ds \right] J_t^*. \end{aligned}$$

Moreover, $||D_s X_t||^2_{\Omega \times \mathbb{R}^m} = \mathrm{Tr}\Sigma_t \in L^p$ for some $p < \infty$. Recursively repeating this procedure, we can obtain higher order derivatives and show that $\forall k \in \mathbb{N}, 0 \leq t \leq T, ||D^k X_t|| \in L^p$ for some $p < \infty$, and consequently, $X_t \in \mathbb{D}_\infty(\mathbb{R}^m)$. □

Remark 5.2.4. In the proof of the preceding theorem, we have deduced that

$$D_s X_t = J_t J_s^{-1} \sigma(X_s) 1_{[0,t]}(s), a.s.$$

This formula is especially useful since it tells us in general how a diffusion driven by an SDE behaves under the Malliavin derivative operator. □

5.3 Hypoellipticity and the Hörmander's Theorem

In this section, I intend to discuss the first significant application of Malliavin calculus, a probabilistic proof of the Hörmander theorem. Let,

$$L = \frac{1}{2}\sum_{i,j=1}^{m} a^{ij}(.)\partial_i\partial_j + \sum_{i=1}^{m} b^i(.)\partial_i$$

and consider the Cauchy problem for heat equation,

$$\begin{cases} \partial_t u(t,x) = Lu(t,x), & t>0, x \in \mathbb{R}^m; \quad \text{(PDE)} \\ u(0,x) = f(x), & . \end{cases}$$

A question of particular interest in PDE theory is to obtain the fundamental solution of a given problem, that is a smooth function $p(t,.)$ on $\mathbb{R}^{2m}$ so that the solution to (PDE) is given by,

$$u(t,x) = \int_{\mathbb{R}^m} p(t,x,y)f(y)dy = \mathbb{E}[f(\phi(x,t,.))]$$

is the solution to (PDE), where $\phi(x,t,.) = X_t(.)$ and X_t is the solution to a suitable SDE with initial condition $X_0 = x$. In this case, the fundamental solution of (PDE) is precisely given by the transition density of the process X_t.

By theorem 4.2.1, we know that if the Malliavin matrix for X_t, Σ_t satisfies

$$(\det \Sigma_t)^{-1} \in L^p$$

for all $p < \infty$, then the transition probability density $p(t,x,y)$ exists and it is smooth. Theorem 5.2.3 allows us to calculate Σ_t for a reasonably general class of diffusion processes. Thus, we will develop a sufficient condition for the existence of a fundamental solution following this path.

Traditionally, it is known that if the matrix $a(x)$ is uniformly elliptic (c.f. [3]), then a smooth fundamental solution exists. In 1967, L. Hörmander obtained a much weaker condition for hypoellipticity of differential operators, namely the well known Hörmander's condition. To state this condition, we write L in the form of vector fields, and we shall adopt Einstein's summation convention for the remainder of this chapter. Let,

$$A_k(.) := \sigma_k^i(.)\partial_i, k = 1, ..., d,$$

$$A_0(.) := \left(b^i(.) - \frac{1}{2}\sum_{k=1}^{d} \sigma_k^j(.)\partial_j\sigma_k^i(.) \right) \partial_i.$$

Observe that, $A_0, A_1, ..., A_k$ are C^∞ vector fields on $\mathbb{R}^m$; and

$$\sum_{k=1}^{d} A_k^2 = a^{ij}\partial_i\partial_j + \sum_{k=1}^{d} \sigma_k^j[\partial_j\sigma_k^i]\partial_i.$$

Hence, we have,

$$L = \frac{1}{2}\sum_{k=1}^{d} A_k^2 + A_0.$$

Furthermore, we the Lie bracket between the vector fields are given by

$$[A_j, A_k] = A_jA_k - A_kA_j.$$

Theorem 5.3.1. ***(Hörmander's Theorem)*** *If the Lie algebra generated by vector fields $\{A_k, [A_0, A_k], k = 1, ..., d\}$ is m dimensional at any $x \in \mathbb{R}^m$, then the fundamental solution to (PDE) exists and is unique.*

Remark 5.3.2. The condition introduced in the preceding theorem is called Hörmander's condition. □

Proof. The proof will be roughly broken into three parts. First of all, we translate our (PDE) into the probabilistic setting. The second part, I will state to establish an upper bound, setting up for applying theorem 4.2.1. Finally, we use theorem 5.3.3 to calculate the Malliavin matrix, and combining with the upper bound derived in the second part to conclude that its inverse is in L^p. It is then a consequence of theorem 4.2.1 that the fundamental solution (or the transition density) exists and it is unique. I shall be mainly concentrating on explaining how the Lie algebras come into play, essentially as a consequence of Itô's lemma; and also the role of Malliavin calculus in the proof. For a more thorough treatment, readers are advised to see section 2.3 of [35].

For simplicity of transformation, we will work with stochastic differential equations in the sense of Itô and Statonovich interchangeably. Let

$$\tilde{b} := b - \frac{1}{2}\sum_{k=1}^{d} A_k^{(1)}\sigma_k$$

where $A_k^{(1)} := (\partial_j\sigma_k^i)_{1\leq i,j\leq m}$, then $A_0(.) = \tilde{b}^i(.)\partial_i$. Observe that the Itô equation (SDE) can be transformed to the following Stratonovich equation,

$$dX_t = A_0(X_t)dt + A_k(X_t) \circ dW_t^k,$$

where $\int . \circ dW_t^k$ is the Stratonovich integral defined in section 2.2. By theorem 2.2.3, Itô's lemma under Stratonovich integration boils down to ordinary chain rule. Hence, $\forall f \in C_b^\infty(\mathbb{R}^m)$, we have

$$df(X_t) = (A_0 f)(X_t)dt + (A_k f)(X_t) \circ dW_t^k.$$

In the sequel, for $V \in C^\infty(\mathbb{R}^m, \mathbb{R}^m)$, V is also understood as a C^∞ vector field: $V(.) = V^i(.)\partial_i$. Note also that the Itô equations that the Jacobian process and its inverse satisfies in theorem 5.2.2, are transformed to

$$dJ_t = \tilde{A}_0^{(1)}(X_t)J_t dt + A_k^{(1)}(X_t)J_t \circ dW_t^k$$

and

$$dJ_t^{-1} = -\tilde{A}_0^{(1)}(X_t)J_t^{-1}dt - J_t^{-1}A_k^{(1)}(X_t) \circ dW_t^k$$

respectively as Stratonovich equations, where $\tilde{A}_0^{(1)} = (\partial_j \tilde{b}^i(x))_{1\leq i,j\leq m}$. Applying Itô's lemma in the Stratonovich setting (see theorem 2,¿), we have

$$\begin{aligned}
d[J_t^{-1}V(X_t)] &= (dJ_t^{-1}) \circ V(X_t) + J_t^{-1} \circ dV(X_t) \\
&= -J_t^{-1}\tilde{A}_0^{(1)}(X_t)V(X_t)dt - J_t^{-1}A_k^{(1)}(X_t)V(X_t) \circ dW_t^k \\
&\quad + J_t^{-1}(A_0 V)(X_t)dt + J_t^{-1}(A_k V)(X_t) \circ dW_t^k.
\end{aligned}$$

Now observe that,

$$[\tilde{A}_0^{(1)}(x)V(x)]^i = V^j(x)\partial_j\tilde{b}^i(x) = (V\tilde{b}^i)(x),$$

and by notations of vector fields, we have $\tilde{A}_0^{(1)}(x)V(x) = (VA_0)(x)$ and similarly, $\tilde{A}_k^{(1)}(x)V(x) = (VA_k)(x)$. Hence,

$$\begin{aligned}
d[J_t^{-1}V(X_t)] &= J_t^{-1}(A_0 V - VA_0)(X_t)dt + J_t \in (A_k V - VA_k)(X_t) \circ dW_t^k \\
&= J_t^{-1}[A_0, V](X_t)dt + J_t^{-1}[A_k, V](X_t) \circ dW_t^k.
\end{aligned}$$

Now, let $R_t := (X_t, J_t)$, so R_t is an $\mathbb{R}^m \times (\mathbb{R}^m \otimes \mathbb{R}^m)$ valued stochastic process with $R_0 = (x, I)$. For any vector field V, define $\xi_V : \mathbb{R}^m \times (\mathbb{R}^m \otimes \mathbb{R}^m) \to \mathbb{R}^m$ by $\xi_V(r) = J^{-1}V(x)$ for $r = (x, J)$. Hence, the preceding equation takes form,

$$\begin{cases} d\xi_V(R_t) = \xi_{[A_0,V]}(R_t)dt + \xi_{[A_k,V]}(R_t) \circ dW_t^k & ; \\ \xi_V(R_0) = V(x). \end{cases}$$

In order to use theorem 5.2.3 to compute the Malliavin matrix, we need to translate the above back to Itô equations. Since,

$$\xi_{[A_k,V]}(R_t) \circ dW_t^k = \xi_{[A_k,V]}(R_t)dW_t^k + \frac{1}{2}\sum_{k=1}^d \xi_{[A_k,[A_k,V]]}(R_t)dt,$$

it follows that our Stratonovich equation becomes,

$$\begin{cases} d\xi_V(R_t) = \xi_{\{A_0,V\}}(R_t)dt + \xi_{\{A_k,V\}}(R_t)dW_t^k \quad ; \quad \text{SDE1} \\ \xi_V(R_0) = V(x). \end{cases}$$

where we define the stochastic Lie brackets by,

$$\{A_k, V\} = [A_k, V] \quad k = 1, ..., d$$

$$\{A_0, V\} = [A_0, V] + \frac{1}{2}\sum_{k=1}^{d}[A_k, [A_k, V]].$$

Let $\hat{\mathcal{V}}_n$ and $\mathcal{V}_n$ be the following sets of vector fields,

$$\begin{aligned} &\hat{\mathcal{V}}_0 := \{A_1, ..., A_d\} \\ &\hat{\mathcal{V}}_n := \{\{A_0, V\}, \{A_k, V\}, V \in \hat{\mathcal{V}}_{n-1}, k = 1, ..., d\}, \quad n \geq 1, \\ &\mathcal{V}_n := \bigcup_{m=0}^{n} \hat{\mathcal{V}}_m, \quad n = 0, 1, 2, ... \end{aligned}$$

Now, we translate Hörmander's condition to one that accommodates us to give a bound on $|\det \Sigma_t|^{-1}$. An alternative way to state Hörmander's condition is, $\forall x \in \mathbb{R}^m, \exists N \geq 0$ such that $V_1, ..., V_m \in \mathcal{V}_N$, such that $V_1(x), ..., V_m(x)$ are linearly independent. Yet, this condition is equivalent to

(H): $\forall x \in \mathbb{R}^m, \exists N \geq 0$ such that

$$\inf_{l \in S} \max_{V \in \mathcal{V}_N} (l, V(x))^2_{\mathbb{R}^m} > 0.$$

where $S = \{x \in \mathbb{R}^m : |x| = 1\}$ is the unit sphere in $\mathbb{R}^m$.

The reason is that since there are only a finite number of vector fields in each of the $\mathcal{V}_n$'s, we can arrange them as a matrix. If the matrix is not full rank, then its rows are linearly independent, and hence $\exists l \in S$ such that

$$\inf_{l \in S} \max_{V \in \mathcal{V}_N} (l, V(x))^2_{\mathbb{R}^m} = 0$$

. Conversely, if the matrix is full rank, then its rows are linearly independent, and hence for any $l \in S$, $\inf_{l \in S} \max_{V \in \mathcal{V}_N} (l, V(x))^2_{\mathbb{R}^m}$ is strictly positive.

Let $p \geq 2$, all there is left is to check that under (H), $\forall t > 0$, the covariance matrix Σ_t satisfies $(\det \Sigma_t)^{-1} \in L^p$. Note that $(\det J_t)^{-1} \in L^p$ and hence it suffices to prove the non-degeneracy condition for

$$\Xi_t := \int_0^t J_s^{-1} a(X_s)(J_s^{-1})^* ds,$$

as by theorem 5.2.3, $\Sigma_t = J_t \Xi_t J_t^*$. Fix $t > 0$ and $c > 0$, define,

$$\tau_c := \int \{s \geq 0 : |X_s - x| \vee ||J_s^{-1} - I|| \geq c^{-1}\} \wedge t.$$

Then τ_c is a stopping time, and for $\varepsilon \in (0, t)$, we have

$$\{\tau_c \leq \varepsilon\} = \left\{ \sup_{s \leq \varepsilon} |X_s - x| \vee ||J_s^{-1} - I|| \geq c^{-1} \right\}.$$

By estimating X_t and J_t from their defining SDE's, it can be shown that, for all $p > 1$,

$$\mathbb{E}\left(\sup_{s \leq \varepsilon} |X_s - x|^p \vee ||J_s^{-1} - I||^p \right) = o(\varepsilon^{p/2}),$$

and therefore, $\tau_c^{-1} \in L^p$. Assuming (H) holds, and taking into account of the continuous dependence of (SDE1) with respect to the initial value, we see that $\forall l_0 \in S, \exists N \in \mathbb{N}_0, V \in \mathcal{V}_N$ and some neighbourhood S_0 of l_0, for sufficiently large c and small $\delta > 0$, we have

$$\int_{l \in S_0} \sup_{s \leq \tau_c} (l, \xi_V(R_s))^2_{\mathbb{R}^m} \geq \delta.$$

Hence, $\forall p > 1$,

$$\sup_{l \in S_0} \mathbb{P}\left(\int_0^\tau (l, \xi_V(R_s))^2_{\mathbb{R}^m} ds < \varepsilon \right) \leq \mathbb{P}(\delta\tau < \varepsilon) = o(\varepsilon^p).$$

Suppose that $V = \{A_{k_j}, \{A_{k_{j-1}}, ..., \{A_{k_1}, A_{k_0}\}...\}\}$, where $0 \leq j \leq N, 1 \leq k_0 \leq d, 0 \leq k_1, ..., k_j \leq d$. For such a V, define $V_0 = A_{k_0}$ and $V_i = \{A_{k_i}, V_{i-1}\}$ for $i = 1, ..., j$. We shall prove by induction that for $i = j, j-1, ..., 0$, we have

$$\sup_{l \in S_0} \mathbb{P}\left(\int_0^\tau (l, \xi_{V_i}(R_s))^2_{\mathbb{R}^m} ds < \varepsilon \right) = o(\varepsilon^p).$$

We have already shown the case when $i = j$, so assume that the above holds for i, and we need to show it also holds for $i - 1$. To this end, we need the following lemma, whose proof can be found on section 2.3 of [35].

Lemma 5.3.3. *Let X be a one-dimensional Itô process, satisfying*

$$X_t = x + \int_0^t Y_s^0 ds + \sum_{k=1}^d \int_0^t Y_s^k dW_s^k, \quad t \geq 0,$$

where Y^0 is also a one dimensional Itô process given by,

$$Y_t^0 = y + \int_0^t Z_s^0 ds + \sum_{k=1}^d \int_0^t Z_s^k dW_s^k, \quad t \geq 0,$$

where $x, y \in \mathbb{R}, Y = (Y^1, ..., Y^d)$ and $Z = (Z^1, ..., Z^d)$ are d-dimensional adapted processes. If $\exists K > 0$ and a bounded stopping time $\tau > 0$, such that

$$\sup_{0 \leq t \leq \tau} \{|Y_t^0| + |Z_t^0| + |Y_t| + |Y_t|\} \leq K$$

then $\forall q > 8, \nu < \frac{q-8}{9}$, and sufficiently small $\varepsilon > 0$, $\exists c > 0$ such that

$$\mathbb{P}\left(\int_0^\tau X_t^2 dt < \varepsilon^q, \int_0^\tau (|Y_t^0|^2 + |Y_t|^2) dt \geq \varepsilon\right) \leq ce^{-\varepsilon^{-\nu}}.$$

Observe that for $l \in S$, and any C^∞ vector field V, we have

$$\begin{cases} d(l, \xi_V(R_t))_{\mathbb{R}^m} = (l, \xi_{\{A_0,V\}}(R_t))_{\mathbb{R}^m} dt + (l, \xi_{\{A_k,V\}}(R_t))_{\mathbb{R}^m} dW_t^k \quad ; \\ (l, \xi_V(R_0))_{\mathbb{R}^m} = (l, V(x))_{\mathbb{R}^m}. \end{cases}$$

By lemma 5.3.3, for $q > 8$ and sufficiently small ε, we have

$$\mathbb{P}\left(\int_0^\tau (l, \xi_{V_{i-1}}(R_s))^2_{\mathbb{R}^m} ds < \varepsilon^q, \int_0^\tau \sum_{k=0}^d (l, \xi_{\{A_k,V_{i-1}\}}(R_s))^2_{\mathbb{R}^m} ds \geq \varepsilon\right) \leq o(\varepsilon^p), 1 < p < \infty.$$

By the inductive assumption, we know that

$$\sup_{l \in S_0} \mathbb{P}\left(\int_0^\tau \sum_{k=0}^d (l, \xi_{\{A_k,V_{i-1}\}}(R_s))^2_{\mathbb{R}^m} ds < \varepsilon\right) \leq o(\varepsilon^p).$$

Therefore,

$$\sup_{l \in S_0} \mathbb{P}\left(\int_0^\tau (l, \xi_{V_{i-1}}(R_s))^2_{\mathbb{R}^m} ds < \varepsilon^q\right) \leq o(\varepsilon^p),$$

which finishes the inductive step. In particular, for $i = 0$, we obtain that there exists $k \in [1, d]$, so that

$$\sup_{l \in S_0} \mathbb{P}\left(\int_0^\tau (l, \xi_{A_k}(R_s))^2_{\mathbb{R}^m} ds < \varepsilon\right) \leq o(\varepsilon^p).$$

Since S is compact, we may choose a finite number of neighbourhoods to cover S, and hence,

$$\mathbb{P}\left(\inf_{l \in S} \int_0^\tau \sum_{k=1}^d (l, \xi_{A_k}(R_s))^2_{\mathbb{R}^m} ds < \varepsilon\right) \leq o(\varepsilon^p).$$

Since $\tau \leq t$, the above inequality obviously holds when τ is replaced by t. On the other hand,

$$\begin{aligned}
\inf_{l \in S} \int_0^t \sum_{k=1}^d (l, \xi_{A_k}(R_s))^2_{\mathbb{R}^m} ds &= \inf_{l \in S} \int_0^t |J_s^{-1}\sigma(X_s)^* l|^2 ds \\
&= \inf_{l \in S} (l, \Xi_t l) \\
&= \lambda_{\min}
\end{aligned}$$

where $\lambda_{\min}$ is the minimum eigenvalue of Ξ_t. It thus follows that $\lambda_{\min}^{-1} \in L^p$ for all $1 < p < \infty$, which means that $|\det \Xi_t|^{-1} \in L^p$, $\forall 1 < p < \infty$. □

Remark 5.3.4.

1. The original probabilistic proof to Hörmander's theorem was given by Malliavin in [31] in 1976. The version presented above was based on the idea of Stroock and Norris in [35].
2. Using a very similar approach to the above, Shigekawa proved in 1980 that if F is a L^2 random variable with a finite Wiener chaos expansion, then the density of F is absolutely continuous. However, it is still an open problem to give an explicit form of the densities to these Wiener chaos.

□

Chapter 6

Applications to Finance

In the final chapter of my thesis, I would like to illustrate some applications of Malliavin calculus to the industry of mathematical finance. A basic knowledge of mathematical finance is assumed, otherwise a good introductory reference for this material is [42] and the first three chapters of [34]. The later chapters of [34] takes the theory to a fairly advanced level, which might also be of interest to enthusiastic readers. We begin this chapter by briefly examine the work of Black and Scholes (1973), and Harrison and Pliska (1981). Then, we introduce some difficulties this theory faces when one tries to extend it to a more general setting, and how the Malliavin analysis of stochastic flows might give a solution to the addressed problem. This approach was initiated by [12] and [13]. [6] provides a friendly introduction to this area, while [5] focuses on looking specifically at Asian options.

6.1 Classical Theory

Typically in mathematical finance, we work with a market that has one risk free assets that admits a discount rate r, and n risky assets. The price dynamics of the risky assets is governed by the stochastic process $X = \{X_t : 0 \leq t \leq T\}$, which is quite typically defined via a **time homogeneous** stochastic differential equation,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

driven by a Wiener process W_t. We say the process is homogeneous in time if the coefficients μ and σ are independent of t. Hence the filtration $\mathcal{F}_t$ generated by W_t will be assumed as the default filtration in the market; or in plain English, it is simply the public information. For simplicity, we assume in the thesis that there will be no dividend or tax payments, traders make profits/losses only through capital gains.

Definition 6.1.1. An process $\alpha = \{\alpha_t \in \mathbb{R}^n, 0 \leq t \leq T\}$ is called a **strategy** if

1. α_t is adapted to $\mathcal{F}_t$.
2. $\int_0^T |\alpha_t| dt < \infty$.

where $|\alpha_t| = |\alpha_t^1| + ... + |\alpha_t^n|$.

A strategy is in essence a way of allocating different proportion of wealth into different risky assets at every point in time. The first condition is there to ensure a trader's strategy cannot be dependant on future events, while the second condition says nobody has access to infinite amount of wealth.

Given a simple strategy α_t, that is an adapted process whose values changes only a countable number of times at t_i, $i = 1, 2, ...$; the capital gain S of the trader is simply,

$$\sum_i a_{t_i}(X_{t_{i+1}} - X_{t_i}).$$

From the developments in chapter 2, we see that by letting $\Delta = \sup |t_{i+1} - t_i|$ tending to 0, the capital gain can be expressed as,

$$S(\alpha) = \int_0^T a_t dX_t = \int_0^T a_t I_n \mu(X_t)dt + \int_0^T a_t \sigma(X_t) dW_t$$

where I_n is the $n \times n$ identity matrix.

Definition 6.1.2. We say a strategy admits an **arbitrage** opportunity if

1. $\mathbb{P}(S(\alpha) > 0) > 0$, and
2. $\mathbb{P}(S(\alpha) < 0) = 0$.

We say a market price is **arbitrage free** if under such circumstances, there exists no strategy α that admits to arbitrage opportunities.

In a mathematical model, any presence of arbitrage opportunities is clearly undesirable, as it would mean that investors could be making instantaneous riskless profits.

Definition 6.1.3. A **contingent claim** is simply a map $\phi : X \to \mathbb{R}$, i.e. an recept or payment that depends on the asset dynamics X. We make no further restrictions of ϕ at this stage.

Example 6.1.4.

1. A **European (call) option** is a contract that gives the holder the right (but no obligation) to purchase a certain asset at a future time T for an agreed **strike price** K. In such cases we have, $\phi(X) = (X_T - K)_+$, where T is the exercising date and K the strike price.
2. An Asian option is when the payoff $\phi(X) = \phi\left(\int_0^T X_t dt, X_T\right)$.
3. An American (call) option is like an European option, except the holder can exercise the option at any time before a future time T.

□

One popular method of pricing these contingent claims is by finding a (not necessarily, but often unique) price that do not allow arbitrage. It was first shown in [15] that the price takes the form

$$P = \mathbb{E}_{\mathbb{Q}}(\phi(X)|X_0 = x),$$

where $\mathbb{E}_{\mathbb{Q}}$ here means taking expectation under the risk neutral measure $\mathbb{Q}$. Readers who have not exposed to risk neutral measures can regard it as an ordinary expectation for purposes of appreciating the ideas introduced in this chapter. I will only briefly introduce some basic definitions and state an very elementary version of Girsanov's theorem in my thesis. Interested readers are advised to consult [15], [34] and [42] for an introductory reading on the transformations of measures, and [43] provides a much deeper study.

Definition 6.1.5. Let $(\Omega, \mathcal{F})$ be a measurable space. We say $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures if $\mathbb{P}(A) = 0 \iff \mathbb{Q}(A) = 0$ for all $A \in \mathcal{F}$.

Before we state Girsanov's theorem, we first state some related facts.

1. If $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures (on $(\Omega, \mathcal{F})$), and X_t is an $\mathcal{F}_t$-adapted process. Then $\mathbb{E}_{\mathbb{Q}}(X_t) = \mathbb{E}_{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} X_t\right)$.
2. Let h be an adapted process on $[0, T]$ and consider the set of processes of the form
$$M_t = \exp\left(\int_0^t h_s dW_s - \frac{1}{2}\int_0^t h_s^2 ds\right).$$
Then, M_t is a martingales if
$$\mathbb{E}\exp\left(\frac{1}{2}\int_0^t h_s^2 ds\right) < \infty.$$
These are called exponential martingales, and they are dense over the space of L^2 martingales (lemma 2.3.6). The condition stated above is called the **Novikov condition**.

Theorem 6.1.6. ***(Girsanov's Theorem)*** *Consider M_t as above with Novikov condition satisfied. Let $\mathbb{Q}$ be a measure on $(\Omega, \mathcal{F})$ such that for all $A \in \mathcal{F}$. Then*
$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}(M_T 1_A)$$
defines a new probability measure on $(\Omega, \mathcal{F})$, and
$$\tilde{W}_t = W_t - \int_0^t h_s ds$$
is a Wiener process under $\mathbb{Q}$.

In 1981, Harrison and Pliska in [15] pursued this path and obtained the classical Black and Scholes formula as a conditional expectation under $\mathbb{Q}$-measure using the so-called risk neutral martingales. The Black and Scholes formula is a closed form solution to the price of an European option, under the assumption that the price dynamics was governed by

$$\frac{dX_t}{X_t} = \mu dt + \sigma dW_t.$$

Then, the price P is given by

$$P = x\Phi\left(\frac{\log(x/K) + rT + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right) - Ke^{rT}\Phi\left(\frac{\log(x/K) + rT - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right),$$

where r is the risk free interest rate and it is assumed to be constant over $[0, T]$; Φ is the cumulative probability distribution of a $N(0, 1)$.

The Black and Scholes option pricing formula was initially published in 1973 using an approach from PDE theory. However, the new approach taken by [15], is believed to have many advantages. For example, one can immediately deduce the Black and Scholes price is in fact arbitrage free (see [34] for details) using the martingale set up in Harrison and Pliska's method, but this property can be difficult to prove directly in the PDE approach. Secondly, the new approach can be easily generalised to give prices of more complicated contingent claims. [16] provides a very thorough discussion between the two methods.

It is often of interest for investors to look at how sensitive the price of a financial derivative is with respect to different parameters. These sensitivities coefficients are traditionally represented by Greek letters, whose definition is summarised as follows:

Greek	Sensitivity
Δ (Delta)	∂_x
Γ (Gamma)	∂_x^2
ρ (Rho)	∂_μ
$\mathcal{V}$ (Vega)	∂_σ.

When the underlying price has a closed form like the case for Black and Scholes, we can calculate the Greeks analytically - it is just a matter of taking derivatives. However we may have to resort to numerical techniques when the prices do not have a closed form. The next section address some of the challenges we face in numerical evaluation of Greeks, and also suggests a possible solution that uses the Malliavin integration by parts formula.

6.2 Monte Carlo Methods in Finance

6.2.1 Some Difficulties

We saw in the previous sections that the analytical approach gave quite promising results in terms pricing an European option, and also calculating the related sensitivities. In real life however, there are many other types of financial derivatives that are of interest which are more complicated than European option, in the sense that the option price might depend on the entire path the underlying asset might take. In such cases, we need to resort to numerical methods, namely use Monte Carlo and simulate the paths.

Let $X = \{X_t, 0 \leq t \leq T\}$ be a stochastic process that determines the price of a risky asset at time t and ϕ be a contingent claim of the form $\phi = \phi(X)$. We are interested in simulating quantities like,

$$u = \mathbb{E}_{\mathbb{Q}}(\phi(X)|X_0 = x),$$

which gives the fair price of the contingent claim, and also its partial derivatives: the Greeks, that tells us how sensitive this price is with respect to its parameters. For simplicity, we will drop the $\mathbb{Q}$ in the expectation, and simply write $\mathbb{E}$ instead for the rest of this chapter.

When using finite difference approximation for the Greeks, bumping the price and taking the sensitivity, one makes two errors: one on the numerical computation of the expectation via the Monte Carlo as for any simulations, and another one on the approximation of the derivative function by means of its finite difference. For example, when applying finite differences to the gamma, one approximates the second order derivative of the payoff function by

$$u"(x) = \frac{u(x+\varepsilon) - 2u(x) + u(x-\varepsilon)}{2\varepsilon}$$

This is obviously very inefficient for non-smooth or discontinuous payoffs, which is a common occurrence in pricing options. Figure 1 of [6] provides a good example showing how finite difference can break down. To overcome this inefficiency, [8] suggested using the likelihood ratio method. If we are interested in the sensitivity of the option price with respect to some parameter θ, and if we know explicitly the density function of the underlying variable, $p(x;\theta)$, we can compute the Greek by,

$$\frac{\partial}{\partial\theta}\mathbb{E}u = \frac{\partial}{\partial\theta}\int u(x)dp(x,\theta)dx = \int u(x)\frac{\frac{\partial}{\partial\theta}p(x,\theta)}{p(x,\theta)}p(x,\theta)dx = \mathbb{E}\left(u(x)\frac{\partial}{\partial\theta}\log p(x,\theta)\right).$$

The interest of this approach was to avoid the differentiation of the payoff function in the simulation process. However, this method was quite restrictive since one needs to have knowledge of the density function explicitly. This is precisely where Malliavin calculus comes into play, more or less in the same way in which it dealt with densities in chapters 4 and 5.

6.2.2 Simulating Greeks via Malliavin Weights

In the finance industry, we are particularly interested in computing the sensitivity of the price of a derivative u with respect to its parameters. We consider a financial market in which two types of financial securities are available, a risk free bond and

n time homogeneous risky assets whose vector of price dynamics, X_t are described by the SDE,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t,$$

where W_t is a Wiener process in $\mathbb{R}^n$ adapted to $\mathcal{F}_t$. The coefficients of μ and σ are assumed to be Lipchitz to ensure the above SDE to have a unique solution. Let J_t be the Jacobian process associated to X_t, for $0 \leq t \leq T$, defined by the stochastic differential equation,

$$\begin{cases} dJ_t = \mu'(X_t)J_t dt + \sum_{i=1}^{n} \sigma_i'(X_t)J_t dW_t^i, \\ J_0 = I_n. \end{cases}$$

where I_n is the $n \times n$ identity matrix, σ_i is the i-th column of the covariance matrix σ.

Remark 6.2.1. In the finance literature (e.g. [6]), J is also commonly termed as the **tangent process** of X. □

It is necessary at this point to assume that the covariance matrix σ satisfies the uniform ellipticity condition. That is, $\exists \varepsilon > 0$, such that

$$\xi^* \sigma^*(x)\sigma(x)\xi \geq \sigma|\xi|^2$$

for any $\xi, x \in \mathbb{R}^n$. The reason for making such an assumption is that since μ' and σ' are assumed to be Lipschitz and bounded, the Jacobian process $J_t \in L^2(\Omega, [0,T])$, (see e.g. Theorem 2.9 of [28]); hence our assumption insures that the process $\{\sigma^{-1}(X_t)J_t\} \in L^2(\Omega \times [0,T])$. Moreover, for any bounded function γ, then $\sigma^{-1}\gamma(X_t) \in L^2(\Omega \times [0,T])$ and $\sigma^{-1}\gamma$ is a bounded function.

Consider a contingent claim $\phi(X)$, with ϕ satisfying some technical conditions that will be described later; we wish to compute the Greeks of its price $u(x) = \mathbb{E}(\phi(X)|X_0 = x)$. That is, we need to take derivatives of u with respect to some parameter λ, a quantities such as drift, the initial conditions and volatility. Our aim is to express each of them in the form of

$$\frac{\partial u}{\partial \lambda} = \mathbb{E}(\phi(X) \times weight)$$

for some λ. This would allow us to avoid the trouble of using finite difference approach in our simulation procedure. The weight function appeared in the previous equation is called the Malliavin weight, as it is generally obtained from the integration by parts relation for Malliavin calculus.

The problem can be approached by looking at perturbed processes, and the limit as the "amount" of perturbation goes to 0. We first look at sensitivity of price with respect to drift. Consider a payoff function $\phi : C[0,T] \to \mathbb{R}$ with finite second moment. The perturbed process X_t^ε defined by,

$$dX_t^\varepsilon = (\mu(X_t^\varepsilon) + \varepsilon\gamma(X_t^\varepsilon))dt + \sigma(X_t^\varepsilon)dW_t,$$

with a corresponding

$$u^\varepsilon(x) = \mathbb{E}(\phi(X^\varepsilon)|X_0^\varepsilon = x),$$

and we still denote the non-perturbed process corresponding to $\varepsilon = 0$, by X_t. The following theorem gives the sensitivity of u with respect to drift.

Theorem 6.2.2. *The function $\varepsilon \to u^\varepsilon(x)$ is differentiable at $\varepsilon = 0$ for any $x \in \mathbb{R}^n$, and the derivative can be written as,*

$$\frac{\partial}{\partial\varepsilon}u^\varepsilon(x)\bigg|_{\varepsilon=0} = \mathbb{E}\left(\phi(X)\int_0^T(\sigma^{-1}\gamma(X_t), dW_t)_{\mathbb{R}^n}\bigg|X_0 = x\right)$$

Proof. We introduce the random variable

$$Z_T^\varepsilon = \exp\left(-\varepsilon\int_0^T(\sigma^{-1}\gamma(X_t), dW_t)_{\mathbb{R}^n} - \frac{\varepsilon^2}{2}\int_0^T||\sigma^{-1}\gamma(X_t)||_{\mathbb{R}^n}^2 dt\right).$$

The Novikov condition is trivially satisfied as $\sigma^{-1}\gamma$ is bounded, and hence we have $\mathbb{E}Z_T^\varepsilon = 1$ for every $\varepsilon > 0$. It then follows that the probability measure $\mathbb{Q}^\varepsilon$ defined by the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}^\varepsilon}{d\mathbb{P}} = Z_T^\varepsilon$$

is equivalent to $\mathbb{P}$, and

$$u^\varepsilon(x) = \mathbb{E}_{\mathbb{Q}^\varepsilon}(\hat{Z}^\varepsilon(T)\phi(X^\varepsilon)|X_0^\varepsilon = x),$$

where

$$\hat{Z}_T^\varepsilon = \exp\left(-\varepsilon\int_0^T(\sigma^{-1}\gamma(X_t), dW_t^\varepsilon)_{\mathbb{R}^n} - \frac{\varepsilon^2}{2}\int_0^T||\sigma^{-1}\gamma(X_t)||_{\mathbb{R}^n}^2 dt\right).$$

and $\{W_t^\varepsilon, 0 \le t \le T\}$ is defined as

$$W_t^\varepsilon = W_t - \varepsilon\int_0^t \sigma^{-1}\gamma(X_s^\varepsilon)ds.$$

By Girsanov's theorem, this is a Wiener process under $\mathbb{Q}^\varepsilon$. By considering the underlying stochastic differential equations, we observe that the joint distribution of $(X^\varepsilon, W^\varepsilon)$ under $\mathbb{Q}^\varepsilon$ coincides with that of (X, W) under $\mathbb{P}$. Hence we obtain,

$$u^\varepsilon(x) = \mathbb{E}(Z_T^\varepsilon \phi(X)|X_0 = x).$$

On the other hand, by directly calculation we have

$$\frac{1}{\varepsilon}(Z_T^\varepsilon - 1) = \int_0^T Z_t^\varepsilon (\sigma^{-1}\gamma(X_t), W_t)_{\mathbb{R}^n},$$

and hence

$$\frac{1}{\varepsilon}(Z_T^\varepsilon - 1) \xrightarrow{L^2} \int_0^T (\sigma^{-1}\gamma(X_t), W_t)_{\mathbb{R}^n},$$

by the dominated convergence theorem. Since $\mathbb{E}(\phi(X)^2)$ was assumed to be finite, we can apply the Cauchy Schwartz inequality to get,

$$\begin{aligned}
&\left| \frac{1}{\varepsilon}(u^\varepsilon(x) - u(x)) - \mathbb{E}\left(\phi(X)\int_0^T (\sigma^{-1}\gamma(X_t), dW_t)_{\mathbb{R}^n}\right) \right| \\
\leq\ & \mathbb{E}\left|\phi(X)\right|\left|\frac{1}{\varepsilon}(Z_T^\varepsilon - 1) - \int_0^T (\sigma^{-1}\gamma(X_t), dW_t)_{\mathbb{R}^n}\right| \\
\leq\ & K\mathbb{E}\left|\frac{1}{\varepsilon}(Z_T^\varepsilon - 1) - \int_0^T (\sigma^{-1}\gamma(X_t), dW_t)_{\mathbb{R}^n}\right|
\end{aligned}$$

for some constant K independent of ε. Therefore, letting $\varepsilon \to 0$ we have

$$\frac{\partial}{\partial\varepsilon}u^\varepsilon(x)\bigg|_{\varepsilon=0} = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(u^\varepsilon(x) - u(x)) = \mathbb{E}\left(\phi(X)\int_0^T (\sigma^{-1}\gamma(X_t), dW_t)_{\mathbb{R}^n}\bigg| X_0 = x\right).$$

□

We now look at sensitivities in the initial condition. Again, we hope to express the derivative as a weighted expectation of the same functional. For this case, We only consider square integrable payoff functions of the form $\phi = \phi(X_{t_1}, ..., X_{t_m})$, i.e. ϕ is only dependant on the asset price over a finite number of points in time. The price of such a contingent claim is typically given by

$$u(x) = \mathbb{E}(\phi(X_{t_1}, ..., X_{t_m})|X_0 = x).$$

We denote ∂_i as the partial derivative with respect to the i-th argument, and $\nabla = \sum_{i=1}^m \partial_i$. Define the set

$$\Gamma_m = \left\{a \in L^2([0,T]) | \int_0^{t_i} a(t)dt = 1, \forall i = 1, ..., m\right\}.$$

We have the following theorem that gives the sensitivity of u with respect to the initial conditions.

Theorem 6.2.3. *Under the assumption that the diffusion matrix σ satisfies the uniform ellipticity condition, for any $x \in \mathbb{R}^n$ and $a \in \Gamma_m$, we have,*

$$\nabla u(x) = \mathbb{E}\left(\phi(X_{t_1}, ..., X_{t_m}) \int_0^T a(t)(\sigma^{-1}(X_t)J_t)^* dW_t \middle| X_0 = x\right)$$

Proof. We assume that ϕ is continuously differentiable with bounded gradient, and we need to first justify the derivative of u with respect to x can be passed through the expectation operator. Since ϕ is continuously differentiable by assumption, we have,

$$\psi_h = \frac{1}{||h||}(\phi(X_{t_1}^x, ..., X_{t_m}^x) - \phi(X_{t_1}^{x+h}, ..., X_{t_m}^{x+h})) - \frac{1}{||h||}\left(\sum_{i=1}^m \partial_i^* \phi(X_{t_1}, ..., X_{t_m})J_{t_i}, h\right)$$

converging to zero almost surely as $h \to 0$. Since ϕ was assumed to have bounded gradient, it follows that the second term of the sum is uniformly integrable. Moreover, we can give an upper bound of the first term by,

$$\frac{1}{||h||}(\phi(X_{t_1}^x, ..., X_{t_m}^x) - \phi(X_{t_1}^{x+h}, ..., X_{t_m}^{x+h})) \leq M \sum_{j=1}^k \frac{||X_{t_j}^x - X_{t_j}^{x+h}||}{||h||}$$

where M is a uniform upper bound of the partial derivatives of ϕ. The uniform integrability of this upper bound follows from general theory of stochastic flows (see for example Theorem 37 of [39]), as the X was assumed to be governed by a stochastic differential equation with Lipschitz coefficients. Hence by dominated convergence, we apply the expectation operator through limits to obtain

$$\nabla^* u(x) = \mathbb{E}\left(\sum_{i=1}^m \partial_i^* \phi(X_{t_i}, ..., X_{t_m})J_{t_i} \middle| X_0 = x\right).$$

Now, since the drift and covariance coefficients has bounded continuous derivatives, by remark 5.2.4, $X \in \mathbb{D}_{1,2}$. Applying the Malliavin derivative, one writes $D_t X_{t_i} = J_{t_j} J_t^{-1} \sigma(t) 1_{t \leq t_i}$ for all $i = 1, ..., m$ and $t \in [0, T]$. Rearranging the terms and taking a weighted average gives,

$$J_{t_i} = \int_0^T D_t X_{t_i} a(t) \sigma^{-1} J_t dt \quad \forall a \in \Gamma_m.$$

Substituting this expression in the equation for $\nabla^* u(x)$ gives,

$$\begin{aligned}\nabla^* u(x) &= \mathbb{E}\left(\int_0^T \sum_{i=1}^m \partial_i^* \phi(X_{t_i}, ..., X_{t_m}) a(t)\sigma^{-1}(t) J_t dt \middle| X_0 = x\right) \\ &= \mathbb{E}\left(\int_0^T D_t \phi(X_{t_i}, ..., X_{t_m}) a(t)\sigma^{-1}(t) J_t dt \middle| X_0 = x\right),\end{aligned}$$

where we have applied the chain rule for the Malliavin derivative to obtain the last line. Finally, since $a(t)\sigma^{-1}(t)J_t \in L^2(\Omega \times [0, T])$ and adapted, we may apply the integration by parts formula to obtain,

$$\begin{aligned}\nabla u(x) &= \mathbb{E}\left(\phi(X_{t_1}, ..., X_{t_m}) D^*(a(t)(\sigma^{-1}(X_t)J_t)^*) \middle| X_0 = x\right) \\ &= \mathbb{E}\left(\phi(X_{t_1}, ..., X_{t_m}) \int_0^T a(t)(\sigma^{-1}(X_t)J_t)^* dW_t \middle| X_0 = x\right),\end{aligned}$$

as the D^* operator coincides with the Itô integral for arguments which are adapted processes; and thus establishing the result for ϕ with continuous and bounded gradient.

Now consider the general case for $\phi \in L^2$. Since the set C_K^∞ of infinitely differentiable functions with compact support is dense in L^2, there exists a sequence $\phi_n \in \mathbb{C}_K^\infty$ converging to ϕ in L^2. Let $u_n(x) = \mathbb{E}(\phi_n(X_{t_1}, ..., X_{t_n})|X_0 = x)$ and

$$\varepsilon_n(x) = \mathbb{E}((\phi_n(X_{t_1}, ..., X_{t_m}) - \phi)^2(X_{t_1}, ..., X_{t_m})|X_0 = x).$$

It is clear that $u_n(x) \to u(x)$ for all $x \in \mathbb{R}^n$, we only need to verify this convergence is indeed uniformly. Let

$$g(x) = \mathbb{E}\left(\phi(X_{t_1}, ..., X_{t_m}) \int_0^T a(t)(\sigma^{-1}(X_t)J_t)^* dW_t \middle| X_0 = x\right).$$

Applying the theorem to the ϕ_n's gives,

$$\begin{aligned}|\nabla u_n(x) - g(x)| &\leq \left|\mathbb{E}\left((\phi_n(X_{t_1}, ..., X_{t_m}) - \phi(X_{t_1}, ..., X_{t_m})) \int_0^T a(t)(\sigma^{-1}(X_t)J_t)^* dW_t \middle| X_0 = x\right)\right| \\ &\leq \varepsilon_n(x)\mathbb{E}\left(\int_0^T a(t)(\sigma^{-1}(X_t)J_t)^* dW_t \middle| X_0 = x\right)^2 \\ &\leq \varepsilon_n(x)\varphi(x).\end{aligned}$$

By the continuity of the expectation operator, this implies that

$$\sup_{x \in K} |\nabla u_n(x) - g(x)| \leq \varepsilon_n(\hat{x})\varphi(\hat{x}) \text{ for some } \hat{x} \in K,$$

where K is some arbitrary compact subset of $\mathbb{R}^n$. This means, $\nabla u_n(x) \to g(x)$ uniformly on compact subsets of $\mathbb{R}^n$, and hence we may conclude that u is continuous differentiable and that $\nabla u = g$. $\square$

Finally, we look at sensitivity with respect to volatility. As in the previous part, we assume the payoff $\phi = \phi(X_{t_1}, ..., X_{t_m})$ with finite second moment. Before it is possible to state the next theorem, we need to introduce some definitions. Let,

$$\hat{\Gamma}_m = \left\{ a \in L^2([0,T]) | \int_{t_{i-1}}^{t_i} a(t)dt = 1, \forall i = 1, ..., m \right\},$$

and let $\hat{\sigma} : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ be a continuously differentiable map with bounded derivatives. We assume the covariance matrix $\sigma + \varepsilon\hat{\sigma}$ satisfies the uniform ellipticity condition. That is for every $\varepsilon, \exists \eta > 0$, such that

$$\xi^*(\sigma + \varepsilon\hat{\sigma})^*(x)(\sigma + \varepsilon\hat{\sigma})(x)\xi \geq \eta|\xi|^2$$

for any $\xi, x \in \mathbb{R}^n$. In order to evaluate the functional derivative with respect to σ, we again take the perturbed process in a similar approach we took with the case of drift. Define the process $X^\varepsilon = \{X_t^\varepsilon : 0 \leq t \leq T\}$ by,

$$\begin{cases} dX^\varepsilon = \mu(X_t^\varepsilon)dt + (\sigma(X_t^\varepsilon) + \varepsilon\hat{\sigma}(X_t^\varepsilon))dW_t, \\ X_0^\varepsilon = x. \end{cases}$$

We introduce the tangent process of X^ε by the following SDE,

$$\begin{cases} dZ_t^\varepsilon = \mu'(X_t^\varepsilon)Z_t^\varepsilon dt + \hat{\sigma}(X_t^\varepsilon)dW_t + \sum_{i=1}^n (\sigma_i + \varepsilon\hat{\sigma}_i)'(X_t^\varepsilon)Z_t^\varepsilon dW_t^i, \\ X_0^\varepsilon = 0_n, \end{cases}$$

where 0_n is the zero vector in $\mathbb{R}^n$. As before, we will denote X, J and Z for X^ε, J^ε, and Z^ε when $\varepsilon = 0$. Now consider the process

$$\beta_t = Z_t J_t^{-1}, \quad 0 \leq t \leq T \ a.s.$$

Then we claim that $\beta_t \in \mathbb{D}_{1,2}$ for $0 \leq t \leq T$. This is true since we can express J^{-1} as the solution of the SDE,

$$\begin{cases} dJ_t^{-1} = J_t^{-1} \left(-\mu'(X_t) + \sum_{i=1}^n (\sigma_i'(X_t))^2 \right) dt - J_t^{-1} \sum_{i=1}^n \sigma_i'(X_t)dW_t^i, \\ J_0^{-1} = I_n. \end{cases}$$

In particular, the drift and volatility coefficients of the SDE has continuous and bounded derivatives. Hence by remark 5.2.4, the process $J_t^{-1} \in \mathbb{D}_{1,2}$, and the same argument also shows that $Z \in \mathbb{D}_{1,2}$. Therefore, the Cauchy Schwartz inequality gives $\beta_t = Z_t J_t^{-1} \in \mathbb{D}_{1,2}$.

We are now in position to state the theorem that allows us to express the sensitivity with respect to volatility in the desired form for purposes of Monte Carlo simulation.

Theorem 6.2.4. *For any $a \in \hat{\Gamma}_m$ and $\sigma + \varepsilon\hat{\sigma}$ satisfying the uniform ellipticity condition, we have*

$$\frac{\partial}{\partial \varepsilon} u^\varepsilon(x)\bigg|_{\varepsilon=0} = \mathbb{E}(\phi(X_{t_1}, ..., X_{t_m}) D^*(\sigma^{-1}(X) J \hat{\beta}_a(T)) | X_0 = x)$$

where

$$\hat{\beta}_a(t) = \sum_{i=1}^m a(t)(\beta_{t_i} - \beta_{t_{i-1}}) 1_{t_{i-1} \le t \le t_i}.$$

Remark 6.2.5. The operator D^* in this case cannot be written as an Itô integral, since a part of the argument $\hat{\beta}_a(T)$ is clearly non-adaptive. □

Proof. We consider only the case when ϕ has continuous and bounded derivatives, as the general case can be extended via a dense subset argument in a similar fashion as the previous case. We can also establish in a similar way to the previous part the validity of differentiation inside the expectation. Namely, we have

$$\frac{\partial}{\partial \varepsilon} u^\varepsilon(x)\bigg|_{\varepsilon=0} = \mathbb{E}\left(\sum_{i=1}^m \partial_i^* \phi(X_{t_1}, ..., X_{t_m}) Z_{t_i} \bigg| X_0 = x\right).$$

in the sense of L^1. By remark 5.2.4, we may take the Malliavin derivative to obtain $D_t X_{t_i} = J_{t_j} J_t^{-1} \sigma(t) 1_{t \le t_i}$ for any $i = 1, ..., m$ and $t \in [0, T]$. Hence, we get

$$\begin{aligned}
\int_0^T D_t X_{t_i} \sigma^{-1}(t) J_t \hat{\beta}_a(T) dt &= \int_0^{t_i} J_{t_i} \hat{\beta}_a(T) dt \\
&= J_{t_i} \sum_{k=1}^i \int_{k=1}^i \int_{t_{k-1}}^{t_k} a(t)(\beta_{t_k} - \beta_{t_{k-1}}) dt \\
&= J_{t_i} \beta_{t_i} \\
&= Z_{t_i},
\end{aligned}$$

the second to last line holds as $a \in \hat{\Gamma}_m$. Now substitution gives,

$$\begin{aligned}
\frac{\partial}{\partial \varepsilon} u^\varepsilon(x)\bigg|_{\varepsilon=0} &= \mathbb{E}\left(\int_0^T \sum_{i=1}^m \partial_i^* \phi(X_{t_1}, ..., X_{t_m}) D_t X_{t_i} \sigma^{-1}(X_t) J_t \hat{\beta}_a(T) dt \bigg| X_0 = x\right) \\
&= \mathbb{E}\left(\int_0^T D_t \phi(X_{t_1}, ..., X_{t_m}) \sigma^{-1}(X_t) J_t \hat{\beta}_a(T) dt \bigg| X_0 = x\right) \\
&= \mathbb{E}(\phi(X_{t_1}, ..., X_{t_m}) D^*(\sigma^{-1}(X) J \hat{\beta}_a(T)) | X_0 = x).
\end{aligned}$$

To justify the use of integration by parts in getting the last step, we note that $\sigma^{-1}(X_t)J_t \in L^2(\Omega \times [0,T])$ and also $\mathcal{F}_t$-adapted. Moreover, we have shown already that $\hat{\beta}_a(T) \in \mathbb{D}_{1,2}$ and it is $\mathcal{F}_T$-measurable. By Cauchy Schwartz, the product process is also in $\mathbb{D}_{1,2}$ and hence belongs to the domain of $\mathbb{D}^*$. In fact, we have

$$D^*(\sigma^{-1}(X)J\hat{\beta}_a(T)) = \hat{\beta}_a(T)\int_0^T (\sigma^{-1}(X)J\hat{\beta}_a(T))^* dW_t - \int_0^T D_t\hat{\beta}_a(T)\sigma^{-1}(X_t)J_t dt.$$

□

Remark 6.2.6.

1. The result in theorem 5.2.1 does not require the Markov property of the process X_t. The only requirement for the argument to flow is the adaptiveness of b, σ and γ.
2. The same kind of argument as in the proof of preceding three theorems generalises in an obvious way to higher order derivatives of u with respect to ε at $\varepsilon = 0$ in the sense that we could also express them in the form $\mathbb{E}(\phi \times \ weight|X_0 = x)$.

□

We now give some concrete examples. Consider the famous Black and Scholes model, where we only have one stock S and one risk free asset whose dynamics is described by

$$\begin{cases} \frac{dS_t}{S_t} = r_t dt + \sigma dW_t, \\ S_0 = x. \end{cases}$$

The tangent process J of this process is the solution to

$$\begin{cases} dJ_t = r_t J_t dt + \sigma J_t dW_t, \\ J_0 = 1, \end{cases}$$

and so we have a.s. $xJ_t = S_t$. Let ϕ be a square integrable functional that describes the payoff of a contingent claim. We denote price of such a contingent claim by $u(x)$, typically we have

$$u(x) = \mathbb{E}\left(e^{-\int_0^T r_t dt}\phi(S_T)|S_0 = x\right),$$

and we wish to simulate values for $\frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}, \frac{\partial u}{\partial r}$ and $\frac{\partial u}{\partial \sigma}$.

First we can calculate an extended ρ, the directional derivative of u for a perturbation $\hat{r}$ on the drift r. By theorem 6.2.2, we have

$$\rho_{\hat{r}} = \mathbb{E}\left(e^{-\int_0^T r_t dt}\phi(S_T)\int_0^T \frac{\hat{r}}{\sigma S_t} dW_t|S_0 = x\right).$$

For the delta, the derivative with respect to the initial condition x, we use theorem 6.2.3. It then boils down to calculating the integral $\int_0^T a(t)\frac{J_t}{\sigma S_t}dW_t$, where $a(t)$ satisfies $\int_0^T a(t)dt = 1$. A trivial choice for such a function is $a(t) = \frac{1}{T}$. Then we obtain,

$$\begin{aligned}\int_0^T a(t)\frac{J_t}{\sigma S_t}dW_t &= \frac{1}{T}\int_0^T \frac{J_t}{\sigma S_t}dW_t \\ &= \frac{1}{T}\int_0^T \frac{1}{x\sigma}dW_t \\ &= \frac{W_T}{x\sigma T},\end{aligned}$$

and hence,

$$\frac{\partial u}{\partial x}(x) = \mathbb{E}\left(e^{-\int_0^T r(t)dt}\phi(S_T)\frac{W_T}{x\sigma T}\right).$$

Applying theorem 6.2.3 again to the above expression, we may obtain an expression for the gamma,

$$\frac{\partial^2 u}{\partial x^2}(x) = \mathbb{E}\left[e^{-\int_0^T r(t)dt}\phi(S_T)\frac{1}{x^2\sigma T}\left(\frac{W_T^2}{\sigma T} - W_T - \frac{1}{\sigma}\right)\right].$$

Finally for vega, we need to apply theorem 6.2.4, and again with $a(t) = \frac{1}{T}$, we obtain,

$$\frac{\partial u}{\partial \sigma} = \mathbb{E}\left[e^{-\int_0^T r(t)dt}\phi(S_T)\left(\frac{W_T^2}{\sigma T} - W_T - \frac{1}{\sigma}\right)\right].$$

Remark 6.2.7. Of course, the Greeks and the option price of any European option in the Black and Scholes set up can be calculated analytically. However, the above analysis clearly generalises to a much more general framework. Indeed, all that's is required is an square integrable payoff contingent on a process whose SDE representation has Lipschitz drift and volatility. For example, readers are advised to see [6] for the case where the asset price follows Heston's model (which is considered as a generalisation of Black and Scholes). □

Another example we look at is when the payoff is of the form $\phi(X) = \phi\left(\int_0^T X_s ds\right)$. Derivatives of this form are called **Asian options** and its price is given by

$$u(x) = \mathbb{E}\left[\phi\left(\int_0^T X_s ds\right)\Big| X_0 = x\right].$$

It was claimed in [12] that

Proposition 6.2.8. *Let u be as above, and $\phi \in L^2[0,T]$. Then,*

$$u'(x) = \mathbb{E}\left[\phi\left(\int_0^T X_s ds\right) D^*\left(\frac{2J_t^2}{\sigma(X_t)}\left(\int_0^T J_s ds\right)^{-1}\right)\Big| X_0 = x\right].$$

Remark 6.2.9. We note again that the term $\left(\int_0^T J_s ds\right)^{-1}$ is not $\mathcal{F}_t$ adapted for $t < T$, and hence the D^* cannot be converted to an Itô integral in this case. $\square$

Proof. We consider only the case that $\phi \in C^\infty$ with compact support, as the general L^2 case can be done using a dense subsets type of argument. We also assume that there exists a process a_t satisfying,

$$\int_0^T J_s^{-1}\sigma(s, X_s)1_{s<t}a_s ds = 1.$$

In such cases, by the dominated convergence theorem, we can differentiate inside the integral. We obtain,

$$\begin{aligned}
u'(x) &= \frac{\partial}{\partial x}\mathbb{E}\left[f\left(\int_0^T X_t dt\right)\middle| X_0 = 0\right] \\
&= \mathbb{E}\left[\phi'\left(\int_0^T X_t dt\right)\int_0^T J_t dt\middle| X_0 = 0\right] \\
&= \mathbb{E}\left[\phi'\left(\int_0^T X_t dt\right)\int_0^T\int_0^T J_t J_s^{-1}\sigma(s, X_s)1_{s<t}a_s ds dt\middle| X_0 = 0\right] \\
&= \mathbb{E}\left[\phi'\left(\int_0^T X_t dt\right)\int_0^T\int_0^T D_s X_t a_s ds dt\middle| X_0 = 0\right] \\
&= \mathbb{E}\left[\int_0^T\left(\phi'\left(\int_0^T X_t dt\right)\int_0^T D_s X_t a_s dt\right) ds\middle| X_0 = 0\right] \\
&= \mathbb{E}\left[\int_0^T D_s\left(\phi\left(\int_0^T X_t dt\right) J_t\left(\int_0^T J_s ds\right) a_s\right) ds\middle| X_0 = 0\right] \\
&= \mathbb{E}\left[\phi\left(\int_0^T X_t dt\right) D^* a_s\middle| X_0 = 0\right].
\end{aligned}$$

We are justified to exchang the orders of D^* and $\int_0^T$ in the third to last line as trajectories of $D_s X_t$ and X_t are continuous over $[0, T]$, and hence dominated convergence theorem applies. It is then easy to verify that the process

$$\frac{2J_t^2}{\sigma(X_t)}\left(\int_0^T J_s ds\right)^{-1}$$

is a valid candidate for a_t. $\square$

Remark 6.2.10. Although we have used Malliavin calculus techniques to avoid the usage of finite difference method in simulating a derivative, there are still two potential problems that needs to be discussed.

1. While the finite difference method can perform poorly at places where the payoff function ϕ is non-smooth or discontinuous, its rate of convergence is

reasonably satisfactory when ϕ is smooth. In such circumstances, the extra noise that is obtained by the global effect of the Malliavin integration by parts that may bring us more trouble than benefit. The next section will precisely illustrate what exactly is meant and will suggest a way of resolving this problem via localisation. Figure 2 of [6]

2. Recall that in calculating the Malliavin weights for sensitivity with respect to the initial condition for example, there was one stage where we had the freedom to pick an arbitrary function a_t, that satisfies $\int_0^T a_t dt = 1$. We immediately took the short cut by picking the most obvious a in our preceding examples. However, what really ought be done here is to choose such a subject to certain optimality conditions, such as minimisation of variance for example. Currently, this is an area of very active research. [13] discusses some elementary treatments, and links it with the Euler-Lagrange equations. [7] is a more recent paper on this topic, and gives a more detailed treatment.

□

6.2.3 Localisation of Malliavin Weights

As mentioned in the previous section, the method of finite difference method is reasonably good for ϕ smooth, while the technique using Malliavin weights obtained from integration by parts has its advantages in when ϕ is non-smooth or discontinuous. This section is devoted to develop a technique which combines the two, namely to only apply the Malliavin integration by parts around any singularities of ϕ. We illustrate this idea with the delta of a call option in the Black and Scholes.

$$\begin{aligned}\frac{\partial}{\partial x}\mathbb{E}\left(e^{-\int_0^T r(t)dt}(S_T-K)_+\right) &= \mathbb{E}\left(e^{-\int_0^T r(t)dt}1_{S_T>K}J_T\right)\\ &= \mathbb{E}\left(e^{-\int_0^T r(t)dt}(S_T-K)_+\frac{W_T}{x\sigma T}\right).\end{aligned}$$

Now, $(S_T-K)_+W_T$ is likely to be very large if T is large, and obviously it also has a large variance. The idea is then to introduce a localisation around the singularity at K. For $\delta>0$, let

$$H_\delta(s)=\begin{cases}0, & \text{if } s\le K-\delta;\\ \frac{s-(K-\delta)}{2\delta}, & \text{if } K-\delta\le s\le K+\delta;\\ 1, & \text{if } s\ge K+\delta.\end{cases}$$

Let $G_\delta(t)=\int_{-\infty}^t H_\delta(s)ds$, $F_\delta(t)=(t-K)_+-G_\delta(t)$. Then,

$$\begin{aligned}\frac{\partial}{\partial x}\mathbb{E}\left(e^{-\int_0^T r(t)dt}(S_T-K)_+\right) &= \frac{\partial}{\partial x}\mathbb{E}\left(e^{-\int_0^T r(t)dt}G_\delta(S_T)\right)+\frac{\partial}{\partial x}\mathbb{E}\left(e^{-\int_0^T r(t)dt}F_\delta(S_T)\right)\\ &= \mathbb{E}\left(e^{-\int_0^T r(t)dt}H_\delta(S_T)J_T\right)+\mathbb{E}\left(e^{-\int_0^T r(t)dt}F_\delta(S_T)\frac{W_T}{x\sigma T}\right).\end{aligned}$$

The advantage of writing in this form is that F_δ vanishes for $s \geq |K-\delta|$, and thus $F_\delta(S_T)W_T$ vanishes when W_T is large. A similar idea can in fact be used for all Greeks, see for example [6] and [12] for details on other kinds of Greeks and/or financial derivatives.

6.2.4 American Options and Conditional Expectations

It was mentioned in section 6.1 that prices to contingent claims ϕ can generally be expressed as $\mathbb{E}_{\mathbb{Q}}(\phi(X_T)|X_0 = x)$. However, there also other types of contingent claims, like the American option (c.f. [14], [1] and section 5.1 of [34]) whose option value takes the form, $\mathbb{E}(\phi(X_T)|X_t = x)$ say for some $t < T$, and t is not necessarily zero.

It was known for a long time that a general conditional expectations of the form $\mathbb{E}(\phi(X_T)|X_t = x)$ creates computational challenge when one applies Monte Carlo techniques. The reason is that often we have $\mathbb{P}(X_t = x) \approx 0$, then essentially almost all simulated paths will not end up hitting $\{X_t = x\}$, and hence are redundant for purposes of computing the conditional expectation. The goal of this section is to transform

$$\mathbb{E}(\phi(X_T)|X_t = x) \to \mathbb{E}(\phi(X_T) \times \ weight)$$

to obtain a more numerically friendly expression for simulation.

Let δ_x denote the Dirac delta centered at x. Then, one may express the conditional expectation as,

$$\mathbb{E}(\phi(X_T)|X_t = x) = \frac{\mathbb{E}(\phi(X_T)\delta_x(X_t))}{\mathbb{E}(\delta_x(X_t))}.$$

With the aid of the joint distribution of X_T and X_t, which we shall denote by $p(x,y)$, under certain regularity conditions, one computes,

$$\begin{aligned}
\mathbb{E}(\phi(X_T)\delta_x(X_t)) &= \int\int \phi(x)\delta_x(y)p(x,y)dxdy \\
&= -\int\int \phi(x)H(y)\frac{\partial p}{\partial y}(x,y)dxdy \\
&= \int\int \phi(x)H(y)p(x,y)\pi(x,y)dxdy \\
&= \mathbb{E}(\phi(F_T)H(G)\pi),
\end{aligned}$$

where $\pi_x(x,y) = -\frac{\partial}{\partial y}\log p$, and $H(y) = 1_{\{y \geq x\}} + c$ so that $\frac{dH}{dy} = \delta_x(y)$.

This simple calculation again reveals a similar problem to the one we faced with the computation of Greeks. Namely, we require some knowledge of the underlying joint density, which is often not available in practice. However, it does explain the existence of certain weights, whose computable form will be derived by applying the integration by parts relation. We assume from now on that $D_sX_T, D_sX_t \in L^2([0,T]\times\Omega)$. We also assume there exists a smooth process $u_s \in H^1$ satisfying,

$$\mathbb{E}\left(\int_0^T D_sX_tu_sds \middle| \sigma(X_T,X_t)\right) = 1.$$

A trivial choice of u_t under certain regularity conditions is simply $u_s = \frac{1}{TD_sX_t}$. The following theorem expresses the conditional expectation in the desired form for numerical computation.

Theorem 6.2.11. *Let ϕ be a Lipschitz function, and $H(y) = 1_{\{y\geq x\}} + c$ for some $c \in \mathbb{R}$, then we have*

$$\mathbb{E}(\phi(X_T)|X_t = t) = \frac{\phi(X_T)H(X_t)D^*(u) - \phi'(X_T)H(X_t)\int_0^T D_sX_Tu_sds}{\mathbb{E}(H(X_t)D^*u)}.$$

Proof. By definition of a conditional expectation, we have

$$\mathbb{E}(\phi(X_T)|X_t = x) = \lim_{\varepsilon\to 0}\frac{\mathbb{E}(\phi(X_T)1_{(-\varepsilon,\varepsilon)}(X_t))}{\mathbb{E}(1_{(-\varepsilon,\varepsilon)}(X_t))}.$$

Now we use the integration by parts relation to get

$$\begin{aligned}
\mathbb{E}(\phi(X_T)1_{(-\varepsilon,\varepsilon)}(X_t)) &= \mathbb{E}\left(\int_0^T D_s(\phi(X_T)H_\varepsilon(X_t))u_sds\right) - \mathbb{E}\left(\phi'(X_T)H_\varepsilon(X_t)\int_0^T D_sX_Tu_sds\right)\\
&= \mathbb{E}\left(\phi(X_T)H_\varepsilon(X_t)D^*u - \phi'(X_T)H_\varepsilon(X_t)\int_0^T D_sX_Tu_sds\right)
\end{aligned}$$

where

$$H_\varepsilon(y) = \begin{cases} c, & \text{if } y \leq -\varepsilon;\\ y+\varepsilon+c, & \text{if } -\varepsilon \leq y \leq \varepsilon;\\ 2\varepsilon + c, & \text{if } y \geq \varepsilon.\end{cases}$$

On the other hand, we have

$$\begin{aligned}
\mathbb{E}(H_\varepsilon(X_t)D^*u) &= \mathbb{E}\left(\int_0^T D_sH_\varepsilon(X_t)uds\right)\\
&= \mathbb{E}\left(1_{(-\varepsilon,\varepsilon)}(X_t)\int_0^T D_sX_tuds\right)\\
&= 1_{(-\varepsilon,\varepsilon)}(X_t)
\end{aligned}$$

The proof is then finished when we let $\varepsilon \to 0$, since $\frac{1}{\varepsilon}H_\varepsilon(X_t)$ converges to $2H(X_t)$, as $\mathbb{P}(G = 0) = 0$. $\square$

Remark 6.2.12.

1. If there exists u_s that also satisfies
$$\mathbb{E}\left(\int_0^T D_s u_s ds \middle| \sigma(X_T, X_t)\right) = 0,$$
then we have
$$\mathbb{E}(\phi(X_T)|X_t = x) = \frac{\phi(X_T)H(X_t)D^*(u)}{\mathbb{E}(H(X_t)D^*u)}.$$
2. The result in theorem 6.2.5 also works for a general conditional expectation of the form $\mathbb{E}(\phi(F)|A)$ for Borel measurable functions ϕ, with at most linear growth at infinity, and A is any measurable set. Please consult [13] and [6] for details.
3. The existence of u_s in the preceding corollary really depends on that $D_s X_T$ is not proportional to $D_s X_t$. If the two derivatives are in fact proportional, it will be shown in section 6.3 that this implies there is some function φ, such that $X_T = \varphi(X_t)$. In such cases, $\mathbb{E}(X_T|X_t = x) = \varphi(0)$.

$\square$

6.3 Other Applications in Finance

As at today (2004), there are two main types applications of Malliavin calculus in finance that are known. The first of its kind dates back to 1991, it involves application Clark's theorem whose key ideas are illustrated in [26] and [27]. It can be viewed as an extension to the classical theory introduced by [15]. Where [15] uses the Itô's martingale representation to argue for the existence of a hedging strategy, [27] will use Clark's martingale representation to give an explicit form of it.

This idea was followed on by [20], who applied it to the study of inside traders. Inside traders by definition are ones whose strategies are $\mathcal{G}_t$ adapted, where the public information $\mathcal{F}_t \subset \mathcal{G}_t$. Traditionally, it is known that under certain conditions, the insider will possess arbitrage opportunities, but the proof was again an existential one. In [20], Malliavin calculus was used via a Clark-type of argument, and obtained an explicit arbitrage strategies for the insider.

The second type of application in finance is centered around the integration by parts formula, it was first introduced in 1999. With hindsight of the materials covered in chapters 4, 5 and 6, the real power of integration by parts is the ability to deal with probability densities. Traditional applications of probability theory relied very much on the knowledge of the density function, yet the density function for solutions to many important stochastic differential equations do not have an explicit form. In chapters 4, 5 and 6, we have already seen some treatments provided by integration by parts formula, and currently this remains to be an area of very active research.

A book written by P. Malliavin [32] is scheduled to be released in July 2005. It will be the first book that aims to systematically cover the ideas of Malliavin calculus applied to mathematical finance.

Another interesting application of Malliavin calculus is the ability, to some extent, describe nonlinear functional dependencies of random variables. It is well known that the covariance or correlation was traditionally used as a popular tool to determine any linear relationships between two random variables F and G. When the functional dependency is nonlinear however, we could somewhat "linearise" such a relation by looking at the Malliavin derivatives. More precisely, let F and G be $\mathcal{F}_T$-measurable and smooth in the sense that D_tF and D_tG exists for $0 \leq t \leq T$. Then if $F = \phi(G)$ for some say Lipschitz ϕ, we would then have $D_tF = \phi'(G)D_tG$ a.s., and thus D_tF and D_tG are proportional as functions of t. This leads us to consider, the Malliavin correlation defined by

$$C(F,G)^2 = \sup_{\omega} \text{ess} \left\{ \frac{\left| \int_0^T D_tFD_tGdt \right|^2}{\left(\int_0^T |D_tF|^2 dt \right) \left(\int_0^T |D_tG|^2 dt \right)} \right\}$$

and in case of D_tF or D_tG are identically zero on $[0,T]$, we then define that $C(F,G) = 1$. Two easy observations we make by just staring at the definition is that, suppose ϕ and φ are Lipschitz functions, then $F = \phi(G) \iff C(F,G) = 1$ and $C(F,G) = C(\phi(F), \phi(G))$. Since Lipschitz functions are dense in the set of measurable functions, we can then extend the previous observation to say that $C(F,G)$ is constant on $\sigma(F) \times \sigma(G)$.

Let us also mention the case when $C(F,G) = 0$, obviously this means some form of L^2-orthogonality. A question raised by Üstunel (final remarks of [13]) is to ask to what extent does this Malliavin type of correlation actually leads us to determine whether two arbitrary L^2 random variables are independent?

One should observe that if we let $X \in L^2(\Omega)$ and

$$F = \begin{cases} X, & \text{if } X \geq 0, \\ 0, & \text{otherwise;} \end{cases} \quad \text{and } G =, \begin{cases} 0, & \text{if } X \geq 0, \\ -X, & \text{otherwise;} \end{cases}$$

then F and G has disjoint support and hence $C(F,G)^2$ will always be zero, yet F and G are by no means independent. At this stage, I am hoping to define a class of "analytic random variables" as an analogue of analytic functions in the sense that some form of analytic continuation is available. Work with these class of random variables, let

$$C_k(F,G)^2 = \sup_{\omega} \text{ess} \left\{ \frac{\left| \int_0^T D_t^kFD_t^kGdt \right|^2}{\left(\int_0^T |D_t^kF|^2 dt \right) \left(\int_0^T |D_t^kG|^2 dt \right)} \right\},$$

and I hope to in the future prove something like if $C_k(F, G) = 0$ for all $k = 0, 1, 2, ...,$ then F and G are independent random variables.

References

[1] Bally, V., Caramellino, L., Zanette, A., Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach, (working paper) INRIA, 2003.

[2] Bally, V., An elementary introduction to Malliavin calculus, *Rapport de recherche no. 4718*, INRIA, 2003.

[3] Bass, R. F., *Diffusions and Elliptic Operators*, Springer-Verlag, 1997.

[4] Bell, D. R., *The Malliavin Calculus*, Longman, Harlow, Essex, 1987.

[5] Benhamou, E., Optimal Malliavin Weighting Function for the Computation of the Greeks, *Proceedings of the Monte Carlo Congress*, Monte Carlo, 2000.

[6] Benhamou, E., Smart Monte Carlo: Various tricks using Malliavin calculus, Goldman Sachs Working Paper, 2001.

[7] Benhamou, E., Optimal Malliavin Weighting Function for the Computation of the Greeks, *Mathematical Finance*, **13** (2003), 37 – 53.

[8] Broadie, M., Glasserman, P., Estimating security price derivatives using simulation, *Management Science*, **42** (1996) 269 C 285.

[9] Brown, A. L., Page, A., *Elements of Functional Analysis*, van Nostrand Reinhold, 1970.

[10] Chung, K. L., Williams, R.J., *Introduction to Stochastic Integration*, 2nd ed. Birkhauser, 1990.

[11] Ewald, C. O., *A Survey on Applications of Malliavin Calculus in Mathematical Finance*, Univ. of Kaiserslautern (Lecture Notes), 2003.

[12] Fournié, E., Lasry, J. M., Lebuchoux, J., Lions, P.L., Applications of Malliavin calculus to Monte-Carlo methods in finance, *Finance and Stochastics*, **3** (1999), 391 – 412.

[13] Fournié, E., Lasry, J. M., Lebuchoux, J., Lions, P.L., Applications of Malliavin calculus to Monte-Carlo methods in finance II, *Finance and Stochastics*, **5** (2001), 201 – 236.

[14] Fujiwara, H., Kijima, M., Pricing of Path-Dependent American Options by Monte Carlo Simulations, (working paper), Kyoto University (2004).

[15] Harrison, M. J., Pliska, S. R., Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes and Their Applications*, **11** (1981), 215 – 260.

[16] Heath, D., Schweizer, M., Martingale versus PDEs in finance: an equivalence result with examples, *J. Appl. Prob.*, **37** (2000) 947 – 957.

[17] Hewitt, E., Stromberg, K., *Real and Abstract Analysis*, Springer-Verlag, 1965.

[18] Holden, H., Oksendal, B., Uboe, J., Zhang, T., *Stochastic Partial Differential Equations: A Modelling, White Noise Functional Approach*, Birkhauser, 1996.

[19] Huang, Z. Y., Yan, J., *Introduction to Infinite Dimensional Stochastic Analysis*, Science Press, 2000.

[20] Imkeller, P.,

[21] Ikeda, N., Watanabe, S., An introduction to Malliavin calculus, *Stochastic Analysis, Proc. Taniguchi Inter. Symp. on Stochastic Analysis.,*, Katana and Kyoto, 1982, ed. Itô, K., North-Holland Amsterdam, 1 – 52.

[22] Itô, K., Multiple Wiener integral, *Journal of Math. Soc. Japan*, **3** (1951), 157 – 169.

[23] S. Janson, *Gaussian Hilbert Spaces*, Cambridge University Press, 1997.

[24] Johnson, G. W., Lapidus, M. L., *The Feynman Integral and Feynman's Operational Calculus*, Oxford Univ. Press, 2000.

[25] Kannan, R., Krueger, C. K., *Advanced Analysis on the Real Line*, Springer, 1996.

[26] Karatzas, I., Ocone, D. L., An extension of Clark's formula, *Stochastics and Stochastics Reports*, **17** (1991), 127 – 131.

[27] Karatzas, I., Ocone, D. L., A generalised Clark representation formula with applications to optimal portfolios, *Stochastics and Stochastics Reports*, **34** (1991), 187 – 220.

[28] Karatzas, I., Shreve, S.E., *Brownian Motion and Stochastic Calculus*, Springer, 2000.

[29] Kunita, H., *Stochastic flows and Stochastic Differential Equations*, Cambridge Press, 1990.

[30] Kuo, H. H., *Gaussian Measures in Banach Spaces*, Lect. Notes in Math. **463**, Springer-Verlag, 1975.

[31] Malliavin, P., Stochastic calculus of variations and hypoelliptic operators. *Proc. Int. Symp. on Stochastic Differential Equations (Kyoto, 1976)* Wiley, New York, 195 – 263.

[32] Malliavin, P., *Stochastic Analysis*, Berlin: Springer, 1997.

[33] Malliavin, P., Thalmaier, A., *Stochastic Calculus of Variations in Mathematical Finance*, Springer Finance, 2005.

[34] Musiela, M., Rutkowski, M., *Martingale Methods in Financial Modelling*, Springer-Verlag, 2004.

[35] Nualart, D., *Malliavin Calculus and Related Topics*, Berlin: Springer, 1995.

[36] Nualart, D., *Stochastic Processes*, (Lecture notes), 1997.

[37] Oksendal, B., *An Introduction to Malliavin Calculus with Applications in Economics*, (Lecture notes), 1997.

[38] Oksendal, B., *Stochastic Differential Equations: An Introduction with Applications*, 5th ed., Springer-Verlag, 1998.

[39] Protter, P., *Stochastic Integral and Differential Equations: A New Approach*, 5th ed., Springer-Verlag, 1990.

[40] Revuz, R., Yor, M., *Continuous Martingales and Brownian Motion*, Springer-Verlag, 1991.

[41] Sans-Sole, M., *A Course on Malliavin Calculus with Applications to Stochastic Partial Differential Equations*, Universitat de Barcelona (Lecture Notes), 2004.

[42] Shreve, S., *Stochastic Calculus and Finance*, (Lecture Notes), 1997.

[43] Üstünel, A. S., Zakai, M., *Transformation of Measure on Wiener Space*, Springer-Verlag, 2000.