Stochastic Analysis and Mathematical Finance

with applications of the Malliavin calculus to the calculation of risk numbers

Alexander Sokol

Speciale for cand.scient graden i matematik
Institut for Matematiske Fag
Københavns Universitet

Thesis for the master degree in mathematics
Department of Mathematical Sciences
University of Copenhagen

Date of Submission: September 8, 2008

Advisor: Rolf Poulsen
Co-advisors: Bo Markussen & Martin Jacobsen

Department of Mathematical Sciences
University of Copenhagen

Contents

Chapter 1

Introduction

This thesis is about stochastic integration, the Malliavin calculus and mathematical finance. The final form of the thesis has developed very gradually, and was certainly not planned in detail from the beginning.

A short history of the development of the thesis. I started out with a vague idea about developing the stochastic integral with respect to what in Øksendal (2005) is known as an Itô process, only to a higher level of rigor than seen in, say, Øksendal (2005) or Steele (2000), and in a self-contained manner, and furthermore showing that this theory rigourously could be used to obtain the fundamental results of mathematical finance. The novelty of this is that ordinary accounts of this type almost always either has some heuristic points or apply more advanced results, such as the Lévy characterisation theorem, which are left unproved in the context of the account.

I also wanted the thesis to contain some practical aspects, and I knew that the calculation of risk numbers is a very important issue for practicioners. I had heard about the results of Fournié et al. (1999), concisely introduced by Benhamou (2001), applying the Malliavin calculus to the problem of calculating risk numbers such as the delta or gamma. When investigating this, I obtained two conclusions. First, the litterature pertaining the Malliavin calculus is rather difficult, with proofs containing very little detail. Second, the numerical investigations of the applications of the Malliavin calculus to the calculation of risk numbers have mostly been constrained to the context of the Black-Scholes model. My idea was to attempt to amend these two issues, first off by developing the fundamentals of the Malliavin calculus in higher detail, and secondly by applying the results to a more advanced model than the Black-Scholes model, namely the Heston model.

In the first few months of work, I developed the hope that I could find the time and ability not only to construct the stochastic integral for Brownian motion and continuous finite variation procesess, but also consider discontinuous finite variation processes. Furthermore, I expected to be able to develop the entire fundament of the Malliavin calculus, including the Skorohod integral, the result on the Malliavin derivative of the solution to a stochastic differential equation and detailed proofs of the results of Fournié et al. (1999). I also vainly expected to find time to consider PDE methods for pricing in finance and to investigate calculation of risk numbers for path-dependent options. In retrospect, all this was very optimistic, and I have ended up restricting myself to stochastic integration for continuous processes, discussing only the fundamental results of the Malliavin derivative operator and not even going as far as to introduce the Skorohod integral. This means that the theory of the Malliavin calculus presented here is not even strong enough to be used to prove the results of Fournié et al. (1999). On the other hand, it is indisputably very rigorous and detailed, and thus constitutes a firm foundation for further theory. Finally, I also had to forget about the idea of PDE methods and path-dependent options in finance.

The final content of the thesis, then, is this: A development of the stochastic integral for Brownian motion and continuous finite variation processes, applied to set up some simple financial models and to produce sufficient conditions for the absence of arbitrage. The Malliavin deriative operator is also developed, but it is necessary to refer to the litterature for the results applicable to finance. Some Monte Carlo methods for evaluating expectations and sensitivities are reviewed, and finally we apply all of this to pricing and calculating risk numbers in the Black-Scholes and Heston models.

It can be sometimes difficult to distinguish between well-known and original work. Many of the results in this thesis are well-known, but formulated in a considerably different manner, often opting for the most pedagogical approach available. There are also many minor original results. The highlights of the original contributions in the thesis are these:

- In Theorem 3.8.3, a proof of the Girsanov theorem not invoking the Lévy characterisation theorem, building on the ideas presented in Steele (2000).

- In Theorem 4.3.5, an extension of the chain rule of the Malliavin calculus, removing the condition of boundedness of the partial derivatives.
- In Section 5.6, identification of an erroneous expression of the Malliavin estimator in Benhamou (2001) for the Heston model and correction of this.

Other significant results are of course all of the numerical results for the Black-Scholes and Heston models in Section 5.5 and Section 5.6. Also, the proof of agreement of the integrals in Theorem 3.8.4 is new. Furthermore, in Appendix C there can be found a new proof of the existence of the quadratic variation for continuous local martingales not using stochastic integration at all. Also, a version of Urysohn's Lemma is proved yielding an expression for the Lipschitz constant of the bump function.

Structure of the thesis. Overall, the thesis first develops the stochastic integral, secondly develops some of the basic results of the Malliavin calculus and finally applies all of this to mathematical finance.

In Chapter 2, we develop some preliminary results which will be essential to the rest of the thesis. We review some results on Brownian motion and the usual conditions, martingales and the like. Furthermore, we develop a general localisation concept which will be essential to the smooth development of the stochastic integral.

Chapter 3 develops the Itô integral. We begin with some results on the space of elementary processes, rigorously proving that the usual definition of the stochastic integral is independent of the representation of the elementary process. Furthermore, we prove a general density result for elementary processes. These preparations make the development of the integral for Brownian motion very easy. We then extend the integral to integrate with respect to integral proceses and continuous processes of finite variation. After having constructed the stochastic integral in the level of generality necessary, we define the quadratic variation and use it to prove Itô's formula, Itô's representation theorem and Girsanov's theorem. We also prove Kazamaki's and Novikov's conditions for an exponental martingale to be a true martingale. Finally, we discuss the pros and cons of the usual conditions.

We then proceed to Chapter 4, containing the basic results on the Malliavin calculus. We define the Malliavin derivative slightly differently than in Nualart (2006), initially using coordinates of the Brownian motion instead of stochastic integrals with deterministic integrands. We then extend the operator by the usual closedness argument.

Next, we prove the chain rule and the integration-by-parts formula. We use the generalized chain rule to show that our definition of the Malliavin derivative coincides with that in Nualart (2006). Finally, we develop the Hilbert space theory of the Malliavin derivative and use it to obtain a chain rule for Lipschitz transformations.

Chapter 5 contains the applications to mathematical finance. We first define a simple class of financial market models and develop sufficent criteria for the absence of arbitrage, and we discuss what a proper price for a contingent claim is. After this, we take a quick detour to introduce some fundamental concepts from the theory of SDEs, and we go through some Monte Carlo methods for evaulation of expectations and sensitivities. Finally, we begin the numerical experiments. We first consider the Black-Scholes model, where we evaluate call prices, call deltas and digital deltas. The purpose of this is mainly a preliminary evaluation of the methods in a simple context, allowing us to gain an idea of their effectiveness and applicability. After this, we try our luck at the Heston model. We prove absence of arbitrage building on some ideas found in Cheridito et al. (2005). We then consider some discretisation schemes for simulating from the Heston model, and apply these to calculation of the digital delta. We compare different methods and conclude that the localised Malliavin method is the superior one. Finally, we outline the difference between our Malliavin estimator and the one found in Benhamou (2001), illustrating the resulting discrepancy.

Finally, in Chapter 6, we discuss our results and opportunities for further work.

The appendices also contain several important results, with original proofs. Appendix A mostly contains well-known results, but the sections on Hermite polynomials and Hilbert spaces contain proofs of results which are hard to find in the litterature. Appendix B contains basic results from measure theory, but also has in Section B.3 an original result on convergence of normal distributions which is essential to our development of the Hilbert space theory of the Malliavin calculus. Furthermore, in Appendix C, we prove some novel results which ultimately turned out not to be necessary for the main parts of the thesis: An existence proof for the quadratic variation and a Lipschitz version of Urysohn's lemma.

Prerequisites. The thesis is written with a reader in mind who has a good grasp of the fundamentals of real analysis, measure theory, probability theory such as can be found in Carothers (2000), Hansen (2004a), Dudley (2002), Jacobsen (2003) and Rogers & Williams (2000a). Some Hilbert space theory and functional analysis is also applied, for introductions to this, see for example Hansen (2006) and Meise & Vogt

(1997).

Final words. I would like to thank by advisor Rolf Poulsen and by co-advisors Bo Markussen and Martin Jacobsen for their support in writing this thesis. Furthermore, I would like to give a special warm thanks to Martin Jacobsen and Ernst Hansen for teaching me measure theory and probability theory and spending so much time always answering mails and discussing whatever mathematical problems I have had at hand.

Chapter 2

Preliminaries

This chapter contains results which are fundamental to the theory that follows, in particular the stochastic calculus, which relies heavily on the machinery of stopping times, martingales and localisation. Many of the basic results in this chapter are assumed to be well-known, and we therefore often refer to the litterature for proofs of the results. Other results, such as the results on progressive measurability and the Brownian filtration, are more esoteric, and we give full proofs.

The first chapter of Karatzas & Shreve (1988) and the second chapter of Rogers & Williams (2000a) are excellent sources for most of the results in this section.

2.1 The Usual Conditions and Brownian motion

Let $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ be a filtered probability space. Recall that the usual conditions for a filtered probability space are the conditions:

1. The σ-algebra $\mathcal{F}$ is complete.
2. The filtration is right-continuous, $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$.
3. For $t \geq 0$, $\mathcal{F}_t$ contains all null sets in $\mathcal{F}$.

In every section of this thesis except this one, we are going to assume the usual conditions. We are also going to work with Brownian motion throughout this thesis. We therefore need to spend some time making sure that we can make our Brownian motion work together with the usual conditions. This is the subject matter of this section.

The reasons for assuming the usual conditions and ways to avoid assuming the usual conditions will be discussed in Section 3.9. Many of our results also hold without the usual conditions, but to avoid clutter, we will consistently assume the usual conditions throughout the thesis.

Our first aim is to understand how to augment a filtered probability space so that it fulfills the usual conditions. In the following, $\mathcal{N}$ denotes the null sets of $\mathcal{F}$. Before we can construct the desired augmentation of a filtered probability space, we need a few lemmas.

Lemma 2.1.1. *Letting $\mathcal{G} = \sigma(\mathcal{F}, \mathcal{N})$, it holds that*

$$\mathcal{G} = \{F \cup N | F \in \mathcal{F}, N \in \mathcal{N}\},$$

and P can be uniquely extended from $\mathcal{F}$ to a probability measure P' on $\mathcal{G}$ by putting $P'(F \cup N) = P(F)$. The space $(\Omega, \mathcal{G}, P')$ is called the completion of $(\Omega, \mathcal{F}, P)$.

Proof. We first prove the equality for $\mathcal{G}$. Define $\mathcal{H} = \{F \cup N | F \in \mathcal{F}, N \in \mathcal{N}\}$. It is clear that $\mathcal{H} \subseteq \mathcal{G}$. We need to prove the opposite inclusion. To do so, we prove directly that $\mathcal{H}$ is a σ-algebra containing $\mathcal{F}$ and $\mathcal{N}$, the inclusion follows from this. It is clear that $\Omega \in \mathcal{H}$. If $H \in \mathcal{H}$ with $H = F \cup N$, we obtain, with $B \in \mathcal{F}$ such that $P(B) = 0$ and $N \subseteq B$,

$$\begin{aligned} H^c &= (F \cup N)^c \\ &= (B^c \cap (F \cup N)^c) \cup (B \cap (F \cup N)^c) \\ &= (B \cup F \cup N)^c \cup (B \cap (F \cup N)^c) \\ &= (B \cup F)^c \cup (B \cap (F \cup N)^c), \end{aligned}$$

so since $(B \cup F)^c \in \mathcal{F}$ and $B \cap (F \cup N)^c \in \mathcal{N}$, we find $H^c \in \mathcal{H}$. If $(H_n) \subseteq \mathcal{H}$ with $H_n = F_n \cup N_n$, we find

$$\bigcup_{n=1}^{n} H_n = \left(\bigcup_{n=1}^{\infty} F_n \right) \cup \left(\bigcup_{n=1}^{\infty} N_n \right),$$

showing $\cup_{n=1}^{\infty} H_n \in \mathcal{H}$. We have now proven that $\mathcal{H}$ is a σ-algebra. Since it contains $\mathcal{F}$ and $\mathcal{N}$, we conclude $\mathcal{G} \subseteq \mathcal{H}$.

It remains to show that P can be uniquely extended from $\mathcal{F}$ to $\mathcal{G}$. We begin by proving that the proposed extension is well-defined. Let $G \in \mathcal{G}$ with two decompositions $G = F_1 \cup N_1$ and $G = F_2 \cup N_2$. We then have $G^c = F_2^c \cap N_2^c$. Thus, if $\omega \in F_2^c$ and $\omega \in G$, then $\omega \in N_2$, showing $F_2^c \subseteq G^c \cup N_2$ and therefore

$$F_1 \cap F_2^c \subseteq F_1 \cap (G^c \cup N_2) = F_1 \cap N_2 \subseteq N_2,$$

and analogously $F_2 \cap F_1^c \subseteq N_1$. We can therefore conclude $P(F_1 \cap F_2^c) = P(F_2 \cap F_1^c) = 0$ and as a consequence,

$$P(F_1) = P(F_1 \cap F_2) + P(F_1 \cap F_2^c) = P(F_2 \cap F_1) + P(F_2 \cap F_1^c) = P(F_2).$$

This means that putting $P'(G) = P(F)$ for $G = F \cup N$ is a definition independent of the representation of G, therefore well-defined. That P' is a probability measure on $\mathcal{G}$ extending P is obvious. □

In the following, we let $(\Omega, \mathcal{G}, P)$ be the completion of $(\Omega, \mathcal{F}, P)$. In particular, we use the same symbol for the measure on $\mathcal{F}$ and its completion. This will not cause any problems.

Lemma 2.1.2. *Let $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{G}$. Then*

$$\sigma(\mathcal{H}, \mathcal{N}) = \{G \in \mathcal{G} | G \Delta H \in \mathcal{N} \text{ for some } H \in \mathcal{H}\},$$

where Δ is the symmetric difference, $G \Delta H = (G \setminus H) \cup (H \setminus G)$.

Proof. Define $\mathcal{K} = \{G \in \mathcal{G} | G \Delta H \in \mathcal{N}$ for some $H \in \mathcal{H}\}$. We begin by arguing that $\mathcal{K} \subseteq \sigma(\mathcal{H}, \mathcal{N})$. Let $G \in \mathcal{K}$ and let $H \in \mathcal{H}$ be such that $G \Delta H \in \mathcal{N}$. We then find

$$\begin{aligned} G &= (H \cap G) \cup (G \setminus H) \\ &= (H \cap (H \cap G^c)^c) \cup (G \setminus H) \\ &= (H \cap (H \setminus G)^c) \cup (G \setminus H). \end{aligned}$$

Since $P(G \Delta H) = 0$, $H \setminus G$ and $G \setminus H$ are both null sets, and we conclude that $G \in \sigma(\mathcal{H}, \mathcal{N})$. Thus $\mathcal{K} \subseteq \sigma(\mathcal{H}, \mathcal{N})$.

To show the other inclusion, we will show that $\mathcal{K}$ is a σ-algebra containing $\mathcal{H}$ and $\mathcal{N}$. If $G \in \mathcal{H}$, we have $G \Delta G = \emptyset \in \mathcal{N}$ and therefore $G \in \mathcal{K}$. If $N \in \mathcal{N}$, we have

$N\Delta\emptyset = N \in \mathcal{N}$, so $N \in \mathcal{K}$. We have now shown $\mathcal{H}, \mathcal{N} \subseteq \mathcal{K}$. It remains to show that $\mathcal{K}$ is a σ-algebra.

Clearly, $\Omega \in \mathcal{K}$. Assume that $G \in \mathcal{K}$ and $H \in \mathcal{H}$ with $G\Delta H \in \mathcal{N}$. We then find

$$G^c \Delta H^c = (G^c \setminus H^c) \cup (H^c \setminus G^c) = (H \setminus G) \cup (G \setminus H) \in \mathcal{N},$$

so $G^c \in \mathcal{K}$ as well. Now assume that $(G_n) \subseteq \mathcal{K}$, and let $H_n \in \mathcal{H}$ be such that $G_n \Delta H_n \in \mathcal{N}$. Then

$$\begin{aligned}
\left(\bigcup_{n=1}^{\infty} G_n\right) \Delta \left(\bigcup_{n=1}^{\infty} H_n\right) &= \left(\bigcup_{n=1}^{\infty} G_n \setminus \bigcup_{n=1}^{\infty} H_n\right) \cup \left(\bigcup_{n=1}^{\infty} H_n \setminus \bigcup_{n=1}^{\infty} G_n\right) \\
&\subseteq \bigcup_{n=1}^{\infty} G_n \setminus H_n \cup \bigcup_{n=1}^{\infty} H_n \setminus G_n \\
&= \bigcup_{n=1}^{\infty} G_n \Delta H_n,
\end{aligned}$$

so $\cup_{n=1}^{\infty} G_n \in \mathcal{K}$. We can now conclude that $\mathcal{K}$ is a σ-algebra. Since it contains $\mathcal{H}$ and $\mathcal{N}$, $\sigma(\mathcal{H}, \mathcal{N}) \subseteq \mathcal{K}$. $\square$

Comment 2.1.3 We cannot obtain the same simple representation of $\sigma(\mathcal{H}, \mathcal{N})$ as we did for $\sigma(\mathcal{F}, \mathcal{N})$ in Lemma 2.1.1. This is because the null sets $\mathcal{N}$ under consideration here are not null sets of $\mathcal{H}$, only of $\mathcal{F}$, and therefore $\{H \cup N | H \in \mathcal{H}, N \in \mathcal{N}\}$ would not even be a σ-algebra: The choice of null sets available is too broad to retain stability under complements.

To see an example of what goes wrong, consider the set-up $(\Omega, \mathcal{G}, P) = ([0,1], \mathbb{L}[0,1], \lambda)$, where $\mathbb{L}[0,1]$ is the Lebesgue σ-algebra on $[0,1]$, the completion of $\mathcal{B}[0,1]$. λ is the Lebesgue measure on $[0,1]$. Let $\mathcal{H} = \sigma(\{[0, \frac{1}{4}]\})$ and let $\mathcal{N}$ be the null sets of $\mathbb{L}[0,1]$. We will argue that $\{H \cup N | H \in \mathcal{H}, N \in \mathcal{N}\}$ is not a σ-algebra.

To do so, consider $N = \{\frac{3}{4}\}$. If $\{H \cup N | H \in \mathcal{H}, N \in \mathcal{N}\}$ were a σ-algebra, we would have $([0, \frac{1}{4}] \cup N)^c = B \cup M$ for some $B \in \mathcal{H}$ and $M \in \mathcal{N}$. Then in particular, $\lambda([0, \frac{1}{4}]^c) = \lambda(([0, \frac{1}{4}] \cup N)^c) = \lambda(B \cup M) = \lambda(B)$. Because $\mathcal{H} = \{\emptyset, \Omega, [0, \frac{1}{4}], [0, \frac{1}{4}]^c\}$, it is clear that $B = [0, \frac{1}{4}]^c$ is necessary. Thus, we need to find $M \in \mathcal{N}$ such that $([0, \frac{1}{4}] \cup N)^c = [0, \frac{1}{4}]^c \cup M$, or, in other words

$$\left(\frac{1}{4}, 1\right] \cap N^c = \left(\frac{1}{4}, 1\right] \cup M,$$

which is clearly impossible, since $(\frac{1}{4},1]\cap N^c$ is strictly smaller than $(\frac{1}{4},1]$ and $(\frac{1}{4},1]\cup M$ is at least as large as $(\frac{1}{4},1]$. ∘

Theorem 2.1.4. *With $\mathcal{F}_{t+}=\cap_{s>t}\mathcal{F}_s$, we have*

$$\bigcap_{s>t}\sigma(\mathcal{F}_s,\mathcal{N})=\sigma(\mathcal{F}_{t+},\mathcal{N}),$$

and the filtration $\mathcal{G}_t=\sigma(\mathcal{F}_{t+},\mathcal{N})$ is the smallest filtration satisfying the usual conditions such that $\mathcal{F}_t\subseteq\mathcal{G}_t$. We call $(\mathcal{G}_t)$ the usual augmentation of $(\mathcal{F}_t)$. We call $(\Omega,\mathcal{G},P,\mathcal{G}_t)$ the usual augmentation of the filtered probability space $(\Omega,\mathcal{F},P,\mathcal{F}_t)$.

Proof. We need to prove three things. First, we need to prove the equality stated in the lemma. Second, we need to prove that $\mathcal{G}_t$ satisfies the usual conditions. And third, we need to prove that $\mathcal{G}_t$ is the smallest filtration containing $\mathcal{F}_t$ satisfying the usual conditions.

Step 1: The equality. We need to prove $\cap_{s>t}\sigma(\mathcal{F}_s,\mathcal{N})=\sigma(\mathcal{F}_{t+},\mathcal{N})$. Since we have $\mathcal{F}_{t+}\subseteq\mathcal{F}_s$ for any $s>t$, it is clear that $\sigma(\mathcal{F}_{t+},\mathcal{N})\subseteq\cap_{s>t}\sigma(\mathcal{F}_s,\mathcal{N})$. Now consider $F\in\cap_{s>t}\sigma(\mathcal{F}_s,\mathcal{N})$. By Lemma 2.1.2, for any $s>t$ there is $F_s\in\mathcal{F}_s$ such that $F\Delta F_s\in\mathcal{N}$. Put $G_n=\cup_{k\geq n}F_{t+\frac{1}{k}}$. Much like in the proof of Lemma 2.1.2, we obtain

$$F\Delta G_n\subseteq\bigcup_{k=n}^{\infty}F\Delta F_{n+\frac{1}{k}}\in\mathcal{N}.$$

Put $G=\cap_{n=1}^{\infty}G_n$. Since G_n is decreasing and $G_n\in\mathcal{F}_{t+\frac{1}{n}}$, $G\in\mathcal{F}_{t+}$. We find

$$\begin{aligned}F\Delta G &= \left(F\setminus\bigcap_{n=1}^{\infty}G_n\right)\cup\left(\bigcap_{n=1}^{\infty}G_n\setminus F\right)\\ &= \left(\bigcup_{n=1}^{\infty}F\setminus G_n\right)\cup\bigcap_{n=1}^{\infty}(G_n\setminus F)\\ &\subseteq \left(\bigcup_{n=1}^{\infty}F\setminus G_n\right)\cup\bigcup_{n=1}^{\infty}(G_n\setminus F)\\ &= \bigcup_{n=1}^{\infty}F\Delta G_n\in\mathcal{N},\end{aligned}$$

showing $F\in\sigma(\mathcal{F}_{t+},\mathcal{N})$ and thereby the inclusion $\cap_{s>t}\sigma(\mathcal{F}_s,\mathcal{N})\subseteq\sigma(\mathcal{F}_{t+},\mathcal{N})$.

Step 2: $\mathcal{G}_t$ satisfies the usual conditions. It is clear that $\mathcal{G}_t$ contains the null sets for all $t\geq 0$. To prove right-continuity of the filtration, we merely note

$$\bigcap_{s>t}\mathcal{G}_s=\bigcap_{s>t}\bigcap_{u>s}\sigma(\mathcal{F}_u,\mathcal{N})=\bigcap_{u>t}\sigma(\mathcal{F}_u,\mathcal{N})=\mathcal{G}_t.$$

Step 3: $\mathcal{G}_t$ satisfies the minimality criterion. Finally, we prove that $\mathcal{G}_t$ is the smallest filtration satisfying the usual conditions such that $\mathcal{F}_t \subseteq \mathcal{G}_t$. To do so, assume that $\mathcal{H}_t$ is another filtration satisfying the usual conditions with $\mathcal{F}_t \subseteq \mathcal{H}_t$. We need to prove $\mathcal{G}_t \subseteq \mathcal{H}_t$. To do so, merely note that since $\mathcal{H}_t$ satisfies the usual conditions, $\mathcal{F}_{t+} \subseteq \mathcal{H}_{t+} = \mathcal{H}_t$, and $\mathcal{N} \subseteq \mathcal{H}_t$. Thus, $\mathcal{G}_t = \sigma(\mathcal{F}_{t+}, \mathcal{N}) \subseteq \mathcal{H}_t$. □

Theorem 2.1.4 shows that for any filtered probability space, there exists a minimal augmentation satisfying the usual conditions, and the theorem also shows how to construct this augmentation.

Next, we consider a n-dimensional Brownian motion W on our basic, still uncompleted, probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$. We will define a criterion to ensure that the Brownian motion interacts properly with the filtration, and we will show that when augmenting the filtration induced by the Brownian motion, we still obtain a proper interaction between the Brownian motion and the filtration.

Definition 2.1.5. *An n-dimensional $\mathcal{F}_t$ Brownian motion is a continuous process W adapted to $\mathcal{F}_t$ such that for any t, the distribution of $s \mapsto W_{t+s} - W_t$ is a n-dimensional Brownian motion independent of $\mathcal{F}_t$.*

In what follows, we will assume that the filtration $\mathcal{F}_t$ is the one induced by the Brownian motion W. W is then trivially a $\mathcal{F}_t$ Brownian motion. Letting $(\Omega, \mathcal{G}, P, \mathcal{G}_t)$ be the usual augmentation of $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ as given in Theorem 2.1.4, we want to show that W is a $\mathcal{G}_t$ Brownian motion. We will do this through a few lemmas. We are led by the results of Rogers & Williams (2000a), Section II.68. By $C([0,\infty), \mathbb{R}^n)$, we denote the set of continuous mappings from $[0,\infty)$ to $\mathbb{R}^n$. By $\mathcal{C}([0,\infty), \mathbb{R}^n)$, we denote the σ-algebra on the space $C([0,\infty), \mathbb{R}^n)$ induced by the coordinate projections. By $C_b(\mathbb{R}^n)$, we denote the bounded continuous mappings from $\mathbb{R}^n$ to $\mathbb{R}$.

Lemma 2.1.6. *W is also a Brownian motion with respect to the filtration $\mathcal{F}_{t+}$.*

Proof. It is clear that W is adapted to $\mathcal{F}_{t+}$. We need to show that for any $t \geq 0$, $s \mapsto W_{t+s} - W_t$ is independent of $\mathcal{F}_{t+}$. Let $t \geq 0$ be given and define $X_s = W_{t+s} - W_t$. It will be sufficient to prove that for every bounded, real $\mathcal{C}([0,\infty), \mathbb{R}^n)$ measurable mapping f and $A \in \mathcal{F}_{t+}$,

$$E1_A f(X) = P(A) Ef(X).$$

To do so, we first prove the result in the case where $f(x) = \prod_{k=1}^n f_k(x_{t_k})$ and $f_1, \ldots, f_n \in C_b(\mathbb{R}^n)$. We will then use monotone-class arguments to obtain the general case.

Therefore, let $0 \leq t_1 \leq \cdots \leq t_n$ and $f_1, \ldots, f_n \in C_b(\mathbb{R}^n)$ be given, and let $A \in \mathcal{F}_{t+}$. Note that for any $\varepsilon > 0$ and $s \geq 0$, $W_{t+s+\varepsilon} - W_{t+\varepsilon}$ is independent of $\mathcal{F}_{t+\varepsilon}$. Therefore, $W_{t+s+\varepsilon} - W_{t+\varepsilon}$ is in particular independent of $\mathcal{F}_{t+}$. We then have, using continuity of W, f_k and dominated convergence,

$$\begin{aligned} E1_A \prod_{k=1}^n f_k(X_{t_k}) &= E1_A \prod_{k=1}^n f_k(W_{t+t_k} - W_t) \\ &= \lim_{\varepsilon \to 0^+} E1_A \prod_{k=1}^n f_k(W_{t+t_k+\varepsilon} - W_{t+\varepsilon}) \\ &= \lim_{\varepsilon \to 0^+} P(A) E \prod_{k=1}^n f_k(W_{t+t_k+\varepsilon} - W_{t+\varepsilon}) \\ &= P(A) E \prod_{k=1}^n f_k(W_{t+t_k} - W_t) \\ &= P(A) E \prod_{k=1}^n f_k(X_{t_k}). \end{aligned}$$

as desired. By the Monotone Class Theorem of Corollary B.1.4, we then conclude $E1_A f(X) = P(A)Ef(X)$ for all $f \in \mathbf{b}\mathcal{C}([0,\infty), \mathbb{R}^n)$. in particular we obtain the desired independence result. □

Comment 2.1.7 Clearly, the method of proof above can be generalized to show analogous results for all right-continuous processes with independent increments. The extension of properties from $\mathcal{F}_t$ to $\mathcal{F}_{t+}$ is a general feature of Markov processes, se for example Ethier & Kurtz (1986), Chapter 4, Jacobsen (1972) or Sokol (2007). ∘

Theorem 2.1.8. *W is a $\mathcal{G}_t$ Brownian motion.*

Proof. It is clear that W is adapted to $\mathcal{G}_t$. We therefore merely need to show that for any $t \geq 0$, $s \mapsto W_{t+s} - W_t$ is independent of $\mathcal{G}_t$. Let $t \geq 0$ and define $X_s = X_{t+s} - X_t$. With $\mathcal{N}$ the null sets of $\mathcal{F}$, we have $\mathcal{G}_t = \sigma(\mathcal{F}_{t+}, \mathcal{N})$. Thus, since both $\mathcal{F}_{t+}$ and the complements of $\mathcal{N}$ contain Ω, the sets of the form $C \cap D$, where $C \in \mathcal{F}_{t+}$ and $D^c \in \mathcal{N}$, form a generating system for $\mathcal{G}_t$, stable under intersections. It will suffice to show $E1_{C\cap D} f(X) = P(C \cap D)Ef(X)$ for any $f \in \mathbf{b}\mathcal{C}([0,\infty), \mathbb{R}^n)$. Lemma 2.1.6 yields

$$E1_{C\cap D} f(X) = E1_C f(X) = P(C)Ef(X) = P(C \cap D)Ef(X).$$

The conclusion follows. □

Theorem 2.1.8 is the result that will allow us to assume the usual conditions in what follows. Next, we prove Blumenthal's zero-one law.

Lemma 2.1.9. *It holds that $\mathcal{F}_{t+} = \sigma(\mathcal{F}_t, \mathcal{N}_{t+})$, where $\mathcal{N}_{t+}$ are the null sets of $\mathcal{F}_{t+}$.*

Proof. It is clear that $\sigma(\mathcal{F}_t, \mathcal{N}_{t+}) \subseteq \mathcal{F}_{t+}$, so it will suffice to show the other inclusion. Let $A \in \mathcal{F}_{t+}$, and put $\xi = 1_A - E(1_A|\mathcal{F}_t)$. Our first step will be to prove that ξ is almost surely zero, and to do so, we first prove $E(\xi 1_B) = 0$ for $B \in \mathcal{F}_\infty$, where $\mathcal{F}_\infty = \sigma(\cup_{t\geq 0}\mathcal{F}_t)$.

To this end, note that with $\mathcal{U}_t = \sigma(W_{t+s} - W_t)_{s\geq 0}$, $\mathcal{F}_\infty = \sigma(\mathcal{F}_t, \mathcal{U}_t)$. The sets of the form $C \cap D$, where $C \in \mathcal{F}_t$ and $D \in \mathcal{U}_t$, are then a generating system for $\mathcal{F}_\infty$, stable under intersections. Using Lemma 2.1.6, we see that $\mathcal{U}_t$ and $\mathcal{F}_{t+}$ are independent. Since ξ is $\mathcal{F}_{t+}$ measurable, we conclude, with $C \in \mathcal{F}_t$ and $D \in \mathcal{U}_t$,

$$
\begin{aligned}
E(\xi 1_{C\cap D}) &= P(D)E(\xi 1_C) \\
&= P(D)E((1_A - E(1_A|\mathcal{F}_t))1_C).
\end{aligned}
$$

Now, by the definition of conditional expectation, $EE(1_A|\mathcal{F}_t)1_C) = E1_A1_C$, and we conclude $E(\xi 1_{C\cap D}) = 0$. Then, by the Dynkin Lemma, $E(\xi 1_B) = 0$ for all $B \in \mathcal{F}_\infty$, and therefore ξ is zero almost surely. Since ξ is $\mathcal{F}_{t+}$ measurable, this means that there exists $N \in \mathcal{F}_{t+}$ with $P(N) = 0$ such that $\xi = 0$ on N^c, and therefore

$$
\begin{aligned}
1_A &= 1_A 1_N + 1_A 1_{N^c} \\
&= 1_{A\cap N} + \xi 1_{N^c} + E(1_A|\mathcal{F}_t)1_{N^c} \\
&= 1_{A\cap N} + E(1_A|\mathcal{F}_t)1_{N^c}.
\end{aligned}
$$

Because $A \cap N \in \mathcal{N}_{t+}$, the right-hand side is measurable with respect to $\sigma(\mathcal{F}_t, \mathcal{N}_{t+})$ and therefore $A \in \sigma(\mathcal{F}_t, \mathcal{N}_{t+})$. □

Lemma 2.1.10 (Blumenthal's 0-1 law). *For any $A \in \mathcal{G}_0$, $P(A)$ is either zero or one.*

Proof. Let $\mathbb{H}$ be the sets of $\mathcal{G}_0$ where the lemma holds. It is clear that $\mathbb{H}$ is a Dynkin system. Using Lemma 2.1.9, we obtain

$$\mathcal{G}_0 = \sigma(\mathcal{F}_{0+}, \mathcal{N}) = \sigma(\mathcal{F}_0, \mathcal{N}_{0+}, \mathcal{N}) = \sigma(\mathcal{F}_0, \mathcal{N}).$$

With $\mathcal{A}$ denoting the complements of the sets in $\mathcal{N}$, we then also have $\mathcal{G}_=\sigma(\mathcal{F}_0, \mathcal{A})$. Because both $\mathcal{F}_0$ and $\mathcal{A}$ contains Ω, the sets of the form $C \cap D$ where $C \in \mathcal{F}_0$ and $D \in \mathcal{A}$ form a generating system for $\mathcal{G}_0$, stable under intersections. By Dynkin's Lemma, it will suffice to show that $P(C \cap D)$ is zero or one for any $C \in \mathcal{F}_0$ and $D \in \mathcal{A}$. Obviously, $P(C \cap D) = P(C)$. Since $\mathcal{F}_0 = \sigma(W_0) = \{\Omega, \emptyset\}$, the result follows. $\square$

Before proceeding to the next section, let us review what we have obtained in this somewhat technical section. Our object was to ensure that we can assume that our filtrations in the following satisfy the usual conditions. We began by proving Theorem 2.1.4, showing that any filtered probability space admits a minimal augmentation satisfying the usual conditions.

Next, we considered a Brownian motion and its induced filtration $\mathcal{F}_t$, and in Theorem 2.1.8, we showed that with $\mathcal{G}_t$ the usual augmentation, W is a $\mathcal{G}_t$ Brownian Motion. If this was not the case, we could not assume the usual conditions in our work with stochastic integration, since we would not have any probability space satisfying the usual conditions and at the same time containing a Brownian motion with respect to the filtration. Finally, we proved Blumenthals zero-one law. This result will be of importance to us later in this chapter, when proving the Girsanov Theorem.

2.2 Stochastic processes

We will now review some basic results on stochastic processes. We work in the context of a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ satisfying the usual conditions. A stochastic process with values in a measure space $(E, \mathcal{E})$ is a family $(X_t)_{t \geq 0}$ of $(E, \mathcal{E})$-valued stochastic variables. The sample paths of the stochastic process X are the functions $X(\omega)$ for $\omega \in \Omega$.

We will in this section review some fundamental concepts pertaining stochastic processes. Many of these results were first investigated systematically in Chung & Doob (1965).

We say that two processes X and Y are versions if $P(X_t = Y_t) = 1$ for all $t \geq 0$. In this case, we say that Y is a version of X and vice versa. We say that two processes X and Y are indistinguishable if their sample paths are almost surely equal. We then say that Y is a modification of Y and vice versa. We call a process evanescent if it is indistinguishable from the zero process.

In the following, $\mathcal{B}$ denotes the Borel-σ-algebra on $\mathbb{R}$. $\mathcal{B}_k$ denotes the k-dimensional Borel-σ-algebra, and for any pseudometric space (M, d), $\mathcal{B}(M)$ denotes the Borel-σ-algebra on M. λ denotes the Lebesgue measure on $\mathbb{R}$. We have the following three measurability concepts for stochastic processes.

- A process X is adapted if $X(t)$ is $\mathcal{F}_t$-measurable for all $t \geq 0$.
- A process X is measurable if $(t, \omega) \mapsto X(t, \omega)$ is $\mathcal{B}[0, \infty) \otimes \mathcal{F}$ measurable.
- A process is progressive if $X_{|[0,t] \times \Omega}$ is $\mathcal{B}[0, t] \otimes \mathcal{F}_t$ measurable for $t \geq 0$.

Furthermore, if a process X has continuous paths in the sense that $X(\omega)$ is continuous for all $\omega \in \Omega$, we say that X is continuous. We will now show that being measurable and adapted and being progressive correspond to being measurable with respect to particular σ-algebras.

Lemma 2.2.1. *Let $\Sigma_{\mathcal{A}}$ be the family of sets $A \in \mathcal{B}[0, \infty) \otimes \mathcal{F}$ such that $A_t \in \mathcal{F}_t$ for all $t \geq 0$, where $A_t = \{\omega \in \Omega | (t, \omega) \in A\}$. Then $\Sigma_{\mathcal{A}}$ is a σ-algebra, and a process X is measurable and adapted if and only if it is $\Sigma_{\mathcal{A}}$-measurable.*

Proof. Clearly, $[0, \infty) \times \Omega \in \Sigma_{\mathcal{A}}$. To show that $\Sigma_{\mathcal{A}}$ is stable under complements, let $A \in \Sigma_{\mathcal{A}}$. Then $A^c \in \mathcal{B}[0, \infty) \otimes \mathcal{F}$ and

$$(A^c)_t = \{\omega \in \Omega | (t, \omega) \in A^c\} = \{\omega \in \Omega | (t, \omega) \in A\}^c = (A_t)^c \in \mathcal{F}_t.$$

Thus $A^c \in \Sigma_{\mathcal{A}}$. Now let $(A_n) \subseteq \Sigma_{\mathcal{A}}$. Then $\cup_{n=1}^\infty A_n \in \mathcal{B}[0, \infty) \otimes \mathcal{F}$ and

$$\begin{aligned} \left(\bigcup_{n=1}^\infty A_n \right)_t &= \left\{ \omega \in \Omega \,\middle|\, (t, \omega) \in \bigcup_{n=1}^\infty A_n \right\} \\ &= \bigcup_{n=1}^\infty \{\omega \in \Omega | (t, \omega) \in A_n\} \in \mathcal{F}_t, \end{aligned}$$

which shows that $\cup_{n=1}^\infty A_n \in \Sigma_{\mathcal{A}}$. We have now shown that $\Sigma_{\mathcal{A}}$ is a σ-algebra. That measurability and adaptedness is equivalent to $\Sigma_{\mathcal{A}}$-measurability is clear from the inclusion $\Sigma_{\mathcal{A}} \subseteq \mathcal{B}[0, \infty) \otimes \mathcal{F}$ and the equality, for any process X,

$$(X \in A)_t = \{\omega \in \Omega | X_t(\omega) \in A\}.$$

□

Lemma 2.2.2. *Let Σ_π be the familiy of sets $A \in \mathcal{B}[0,\infty) \otimes \mathcal{F}$ such that A satisfies $A \cap [0,t] \times \Omega \in \mathcal{B}[0,t] \otimes \mathcal{F}_t$ for all $t \in [0,\infty)$. Then Σ_π is a σ-algebra, and a process X is progressively measurable if and only if it is Σ_π-measurable.*

Proof. That Σ_π is a σ-algebra is clear. The equality

$$(X \in A) \cap [0,t] \times \Omega = (X_{|[0,t]\times\Omega} \in A)$$

shows that progressive measurability is equivalent to Σ_π measurability. □

Lemma 2.2.1 and Lemma 2.2.2 show that the concepts of being measurable and adapted and of being progressive correspond to conventional measurability properties with respect to the σ-algebras $\Sigma_\mathcal{A}$ and Σ_π, respectively. These results will enable us to apply the usual machinery of measure theory when dealing when these measurability concepts. The next results demonstrate some of the interplay between the regularity properties of stochastic processes. In particular, Lemma 2.2.3 shows that $\Sigma_\pi \subseteq \Sigma_\mathcal{A}$.

Lemma 2.2.3. *Let X be progressive. Then X is measurable and adapted.*

Proof. That X is measurable follows from Lemma 2.2.2. To show that X is adapted, note that when X is progressive, $X_{|[0,t]\times\Omega}$ is $\mathcal{B}[0,t] \otimes \mathcal{F}_t$-measurable, and therefore $\omega \mapsto X(t,\omega)$ is $\mathcal{F}_t$-measurable. □

Lemma 2.2.4. *Let X be adapted and continuous. Then X is progressively measurable.*

Proof. Define $X_n(t) = X(\frac{1}{2^n}[2^n t])$. We can then write

$$X_n(t) = \sum_{k=0}^{\infty} X\left(\frac{k}{2^n}\right) 1_{[\frac{k}{2^n},\frac{k+1}{2^n})}(t).$$

Since X is adapted, $X(\frac{k}{2^n})$ is $\mathcal{F}_{\frac{k}{2^n}}$ measurable. Therefore, each term in the sum is progressive. Since X_n converges pointwise to X by the continuity of X, we conclude that X is progressive as the limit of progressive processes. □

Comment 2.2.5 In the proof of the lemma, we actually implicitly employed the result from Lemma 2.2.2 that progressive measurability is equivalent to measurability with respect to a σ-algebra Σ_π, in the sense that we know that ordinary measurability is preserved by pointwise limits. Had we not known that progressive measurability is measurability with respect to Σ_π, we would have to prove manually that it is preserved by pointwise convergence. Note that the result would also hold if continuity was weakened to left-continuity or right-continuity. $\circ$

Lemma 2.2.6. *Let X and Y be measurable processes. If X and Y are versions, the set $\{(t,\omega) \in [0,\infty) \times \Omega | X_t(\omega) \neq Y_t(\omega)\}$ is a $\lambda \otimes P$ null set.*

Proof. First note that $A = \{t,\omega) \in [0,\infty) \times \Omega | X_t(\omega) \neq Y_t(\omega)\}$ is $\mathcal{B}[0,\infty) \otimes \mathcal{F}$ measurable since X and Y both are measurable. By the Tonelli Theorem, then

$$(\lambda \otimes P)(A) = \int_0^\infty P(X_t \neq Y_t)\,\mathrm{d}t = 0,$$

showing the claim. $\square$

Comment 2.2.7 Note that the other implication does not hold: We can only conclude that $P(X_t \neq Y_t) = 0$ almost surely if $\{(t,\omega) \in [0,\infty) \times \Omega | X_t \neq Y_t\}$ is a $\lambda \otimes P$ null set. Note also that it is crucial that both X and Y are measurable in the above, otherwise the set $\{(t,\omega) \in [0,\infty) \times \Omega | X_t(\omega) \neq Y_t(\omega)\}$ would not be $\mathcal{B}[0,\infty) \otimes \mathcal{F}$ measurable. $\circ$

2.3 Stopping times

This section contains some basic results on stopping times. A stopping time is a stochastic variable $\tau : \Omega \to [0,\infty]$ such that $(\tau \leq t) \in \mathcal{F}_t$ for any $t \geq 0$. We say that τ is finite if τ maps into $[0,\infty)$. We say that τ is bounded if τ maps into a bounded subset of $[0,\infty)$. If X is a stochastic process and τ is a stopping time, we denote by X^τ the process $X^\tau_t = X_{\tau \wedge t}$ and call X^τ the process stopped at τ. We denote by $X[0,\tau]$ the process $X[0,\tau]_t = X_t 1_{[0,\tau]}(t)$ and call $X[0,\tau]$ the process zero-stopped at τ.

Furthermore, we define the stopping-time σ-algebra $\mathcal{F}_\tau$ of events determined at τ by $\mathcal{F}_\tau = \{A \in \mathcal{F} | A \cap (\tau \leq t) \in \mathcal{F}_t \text{ for all } t \geq 0\}$. Clearly, $\mathcal{F}_\tau$ is a σ-algebra. Our first goal is to develop some basic results on stopping times and their interplay with stopping-time σ-algebras. By $\mathcal{B}_0$, we denote the elements $A \in \mathcal{B}$ with $0 \notin A$.

Lemma 2.3.1. *If τ and σ are stopping times, then so is $\tau \wedge \sigma$, $\tau \vee \sigma$ and $\tau + \sigma$. Furthermore, if $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.*

Proof. See Karatzas & Shreve (1988), Lemma 1.2.9 and Lemma 1.2.15. $\square$

Lemma 2.3.2. *Let τ be a stopping time. Then $(\tau < t) \in \mathcal{F}_t$ for all $t \geq 0$.*

Proof. We have $(\tau < t) = \cup_{n\geq 1} \left(\tau \leq t - \frac{1}{n}\right) \in \mathcal{F}_t$, since $(\tau \leq t - \frac{1}{n}) \in \mathcal{F}_{t-\frac{1}{n}} \subseteq \mathcal{F}_t$. $\square$

Lemma 2.3.3. *Let X be progressively measurable, and let τ be a stopping time. Then X_τ is $\mathcal{F}_\tau$ measurable and X^τ is progressively measurable.*

Proof. See Karatzas & Shreve (1988), Proposition 1.2.18. $\square$

Lemma 2.3.4. *For any stopping time τ, the process $(t, \omega) \mapsto 1_{[0,\tau(\omega)]}(t)$ is progressive. In particular, if X is Σ_π measurable, then $X[0, \tau]$ is Σ_π measurable as well.*

Proof. Let $t \geq 0$. Our fundamental observation is

$$\{(s, \omega) \in [0, t] \times \Omega | 1_{[0,\tau(\omega)]}(s) = 0\} \quad = \quad \{(s, \omega) \in [0, t] \times \Omega | \tau(\omega) < s\}.$$

We need to prove that the latter is $\mathcal{B}[0, t] \otimes \mathcal{F}_t$ measurable. Since $1_{[0,\tau]}$ only takes the values 0 and 1, this is sufficient to prove progressive measurability. To this end, define the mapping $f : [0, t] \times (\tau \leq t) \to [0, t]^2$ by $f(s, \omega) = (s, \tau(\omega))$. Let τ_t be the restriction of τ to $(\tau \leq t)$. We then have $(\tau_t \leq s) \in \mathcal{F}_s \subseteq \mathcal{F}_t$ for any $s \leq t$. Since the sets of the form $[0, s]$, $s \leq t$, form a generating family for $\mathcal{B}[0, t]$, we conclude that τ_t is $\mathcal{F}_t - \mathcal{B}[0, t]$ measurable. Therefore, f is $\mathcal{B}[0, t] \otimes \mathcal{F}_t - \mathcal{B}[0, t]^2$ measurable. With $A = \{(x, y) \in [0, t]^2 | x > y\}$, $A \in \mathcal{B}[0, t]^2$ and we obtain

$$\begin{aligned} \{(s, \omega) \in [0, t] \times \Omega | \tau(\omega) < s\} &= \{(s, \omega) \in [0, t] \times \Omega | f(s, \omega) \in A\} \\ &= f^{-1}(A), \end{aligned}$$

which is $\mathcal{B}[0, t] \otimes \mathcal{F}_t$ measurable, as desired. The remaining claims of the lemma follows immediately. $\square$

Lemma 2.3.5. *Let X be a continuous and adapted process, and let F be closed. Then the first entrance time $\tau = \inf\{t \geq 0 | X_t \in F\}$ is a stopping time.*

Proof. See Rogers & Williams (2000a), Lemma 74.2. $\square$

Lemma 2.3.6. *Let τ_k, $k \leq n$ be a family of stopping times. Then the i'th ordered value of $(\tau_k)_{k\leq n}$ is also a stopping time.*

Comment 2.3.7 Note that it does not matter how we order the stopping times in case of ties, since we are only interested in the ordered *values* of the stopping times, not the actual ordering. ◦

Proof. Let $t \geq 0$, and let $\tau_{(i)}$ be the i'th ordered value of $(\tau_k)_{k\leq n}$. Then $\tau_{(i)} \leq t$ precisely if i or more of the stopping times are less than t. That is,

$$(\tau_{(i)} \leq t) = \bigcup_{I\subseteq\{1,\ldots,n\},|I|\geq i} (\tau_k \leq t, k \in I) \cap (\tau_k > t, k \notin I) \in \mathcal{F}_t.$$

□

Lemma 2.3.8. *Let τ and σ be stopping times. Assume that $Z \in \mathcal{F}_\sigma$. Then it holds that both $Z1_{(\sigma<\tau)}$ and $Z1_{(\sigma\leq\tau)}$ are $\mathcal{F}_{\sigma\wedge\tau}$ measurable.*

Proof. We first show that $(\sigma < \tau) \in \mathcal{F}_{\sigma\wedge\tau}$. Fix $t \geq 0$, it will suffice to show that $(\sigma < \tau) \cap (\sigma \wedge \tau \leq t) \in \mathcal{F}_t$. First note

$$(\sigma < \tau) \cap (\sigma \wedge \tau \leq t) = (\sigma < \tau) \cap (\sigma \leq t).$$

Now consider some $\omega \in \Omega$ such that $\sigma(\omega) < \tau(\omega)$ and $\sigma(\omega) \leq t$. If $t < \tau(\omega)$, we have $\sigma(\omega) \leq t < \tau(\omega)$. If $\tau(\omega) \leq t$, there is some $q \in \mathbb{Q} \cap [0, t]$ such that $\sigma(\omega) \leq q < \tau(\omega)$. We thus obtain

$$(\sigma < \tau) \cap (\sigma \wedge \tau \leq t) = \bigcup_{q\in\mathbb{Q}\cap[0,t]\cup\{t\}} (\sigma \leq q) \cap (q < \tau) \cap (\sigma \leq t),$$

which is in $\mathcal{F}_t$, showing $(\sigma < \tau) \in \mathcal{F}_{\sigma\wedge\tau}$. We now show that $Z1_{(\sigma<\tau)}$ is $\mathcal{F}_{\sigma\wedge\tau}$ measurable. Let $B \in \mathcal{B}_0$. It will suffice to show that for any $t \geq 0$ it holds that $(Z1_{(\sigma<\tau)} \in B) \cap (\sigma \wedge \tau \leq t) \in \mathcal{F}_t$. To obtain this, we rewrite

$$\begin{aligned}(Z1_{(\sigma<\tau)} \in B) \cap (\sigma \wedge \tau \leq t) &= (Z \in B) \cap (\sigma < \tau) \cap (\sigma \wedge \tau \leq t) \\ &= (Z \in B) \cap (\sigma < \tau) \cap (\sigma \leq t).\end{aligned}$$

Since Z is $\mathcal{F}_\sigma$ measurable, $(Z \in B) \cap (\sigma \leq t) \in \mathcal{F}_t$. And by what we have already shown, $(\sigma < \tau) \in \mathcal{F}_\sigma$, so $(\sigma < \tau) \cap (\sigma \leq t) \in \mathcal{F}_t$. Thus, the above is in $\mathcal{F}_t$, as desired.

Next, we show that $Z1_{(\sigma\leq\tau)}$ is $\mathcal{F}_{\sigma\wedge\tau}$ measurable. Let $B \in \mathcal{B}_0$. As in the proof of the first part of the lemma, it is sufficient to show that for any $t \geq 0$ it holds that $(Z1_{(\sigma\leq\tau)} \in B) \cap (\sigma \wedge \tau \leq t) \in \mathcal{F}_t$. To obtain this, we first write

$$\begin{aligned}(Z1_{(\sigma\leq\tau)} \in B) \cap (\sigma \wedge \tau \leq t) &= (Z \in B) \cap (\sigma \leq \tau) \cap (\sigma \wedge \tau \leq t) \\ &= (Z \in B) \cap (\sigma \leq t) \cap (\sigma \leq \tau) \cap (\sigma \wedge \tau \leq t).\end{aligned}$$

Since $Z \in \mathcal{F}_\sigma$, we find $(Z \in B) \cap (\sigma \leq t) \in \mathcal{F}_t$. And since we know $(\tau < \sigma) \in \mathcal{F}_{\tau\wedge\sigma}$, $(\sigma \leq \tau) = (\tau < \sigma)^c \in \mathcal{F}_{\sigma\wedge\tau}$, so $(\sigma \leq \tau) \cap (\sigma \wedge \tau \leq t) \in \mathcal{F}_t$. This demonstrates $(Z1_{(\sigma\leq\tau)} \in B) \cap (\sigma \wedge \tau \leq t) \in \mathcal{F}_t$, as desired. □

2.4 Martingales

In this section, we review the basic theory of martingales. As before, we work in a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ satisfying the usual conditions. If M is some process, we put $M_t^* = \sup_{s\leq t} |M_s|$ and $M^* = M_\infty^*$. A process M is said to be a martingale if M_s is adapted and $E(M_t|\mathcal{F}_s) = M_s$ whenever $s \leq t$. We say that M is square-integrable if M is bounded in $\mathcal{L}^2$. As described in Appendix B.5, we say that M is uniformly integrable if

$$\lim_{x\to\infty} \sup_{t\geq 0} E|M_t|1_{(|M_t|>x)} = 0.$$

In this case, M is also bounded in $\mathcal{L}^1$ by Lemma B.5.2. By Lemma B.5.1, if M_t is bounded in $\mathcal{L}^p$ for any $p > 1$, it is uniformly integrable. We denote the space of martingales by $\mathcal{M}$. The space of martingales stating at zero is denoted $\mathcal{M}_0$.

We will mostly be concerned with continuous martingales. However, at some points we will be interested in cadlag martingales as well. The following result shows that we always can find a cadlag version of any martingale.

Theorem 2.4.1. *Let M be a martingale. Then M has a cadlag version.*

Proof. This follows from Theorem II.67.7 of Rogers & Williams (2000a). □

Next, we go on to state some of the classic results of martingale theory.

Theorem 2.4.2 (Doob's maximal inequality). *Let M be a nonnegative cadlag submartingale. For $c > 0$ and $t \geq 0$, $cP(M_t^* \geq c) \leq EM_t 1_{(M_t^*\geq c)}$.*

Proof. See Rogers & Williams (2000a), Theorem 70.1. □

Theorem 2.4.3 (Martingale Convergence Theorem). *Let M be a continuous submartingale. If M^+ is bounded in $\mathcal{L}^1$, then M is almost surely convergent to an* *integrable variable. If M is also uniformly integrable, the limit M_∞ exists in $\mathcal{L}^1$ as well, and $E(M_\infty|\mathcal{F}_t) = M_t$. In this case, we say that M_∞ closes the martingale.*

Proof. The first part follows from Theorem 1.3.15 in Karatzas & Shreve (1988), and the rest follows from Theorem II.69.2 in Rogers & Williams (2000a). □

Corollary 2.4.4. *Let M be a continuous martingale. M is closed by a variable M_∞ if and only if M is uniformly integrable, and in this case M_t converges to M_∞ almost surely and in $\mathcal{L}^1$.*

Proof. From Theorem 2.4.3, we know that if M is uniformly integrable, M is closed by some variable and we have convergence almost surely and in $\mathcal{L}^1$. It will therefore suffice to show the other implication. Assume that M is closed by a variable M_∞, we need to show that M is uniformly integrable. By Jensen's inequality, we know that $|M_t| = |E(M_\infty|\mathcal{F}_t)| \le E(|M_\infty||\mathcal{F}_t)$. In particular, M is bounded in $\mathcal{L}^1$, and therefore M_t is convergent almost surely. Therefore, M^* is almost surely finite. We obtain

$$\begin{aligned} E|M_t|1_{(|M_t|>x)} &\le EE(|M_\infty||\mathcal{F}_t)1_{(|M_t|>x)} \\ &= E|M_\infty|1_{(|M_t|>x)} \\ &\le E|M_\infty|1_{(M^*>x)}, \end{aligned}$$

and since M^* is almost surely finite, we conclude

$$\limsup_{x\to\infty} \sup_{t\ge 0} E|M_t|1_{(|M_t|>x)} \le \limsup_{x\to\infty} E|M_\infty|1_{(M^*>x)} = 0,$$

so M is uniformly integrable. □

Theorem 2.4.5 (Doob's $\mathcal{L}^p$-inequality). *Let M be a continuous martingale. For any $t \ge 0$ and $p > 1$, $\|M_t^*\|_p \le q\|M_t\|_p$, where q is the dual exponent to p.*

Proof. See Rogers & Williams (2000a), Theorem 70.2. □

Lemma 2.4.6. *If M is a martingale bounded in $\mathcal{L}^2$, then M is almost surely convergent and the limit M_∞ is square integrable. Furthermore, M_∞ closes the martingale.*

Proof. See Rogers & Williams (2000a), Theorem 70.2. □

Theorem 2.4.7 (Optional Sampling). *Let M be a continuous martingale. Let $\sigma \leq \tau$ be stopping times. Then $E(M_\tau|\mathcal{F}_\sigma) = M_\sigma$ if either M is uniformly integrable or σ and τ are bounded.*

Proof. Assume first that M is uniformly integrable. By 2.4.3, M is then convergent almost surely and in $\mathcal{L}^1$ with limit M_∞, and M_∞ closes the martingale. Then Theorem 1.3.22 of Karatzas & Shreve (1988) yields the result.

Next assume that σ and τ are bounded, say by T. Then $M_\tau = M_\tau^T$ and $M_\sigma = M_\sigma^T$. M^T is a uniformly integrable martingale by Corollary 2.4.4, so we obtain

$$E(M_\tau|\mathcal{F}_\sigma) = E(M_\tau^T|\mathcal{F}_\sigma) = M_\sigma^T = M_\sigma.$$

□

Since any square-integrable martingale is uniformly integrable, optional sampling holds for square-integrable martingales for all stopping times.

Definition 2.4.8. *By $\mathbf{c}\mathcal{M}_0^2$, we denote the space of continuous martingales bounded in $\mathcal{L}^2$. We endow $\mathbf{c}\mathcal{M}_0^2$ with the the seminorm given by $\|M\|_{\mathbf{c}\mathcal{M}_0^2} = \sqrt{EM_\infty^2}$.*

It is clear that $\|\cdot\|_{\mathbf{c}\mathcal{M}_0^2}$ is a well-defined seminorm.

Lemma 2.4.9. *Convergence in $\mathbf{c}\mathcal{M}_0^2$ is equivalent to uniform $\mathcal{L}^2$-convergence and implies almost sure uniform convergence along a subsequence. Also, $\|M-N\|_{\mathbf{c}\mathcal{M}_0^2} = 0$ if and only if M and N are indistinguishable.*

Proof. From Doob's $\mathcal{L}^2$-inequality, Lemma 2.4.5, we have for any $M, N \in \mathbf{c}\mathcal{M}_0^2$ that

$$\|M-N\|_{\mathbf{c}\mathcal{M}_0^2} \leq \|(M-N)^*\|_2 \leq 2\|M_\infty - N_\infty\|_2 = 2\|M-N\|_{\mathbf{c}\mathcal{M}_0^2}.$$

Therefore, it is immediate that convergence in $\mathbf{c}\mathcal{M}_0^2$ is equivalent to uniform $\mathcal{L}^2$-convergence. Now assume that M_n converges to M in $\mathbf{c}\mathcal{M}_0^2$. In particular, then $(M_n - M)^*$ converges to 0 in $\mathcal{L}^2$. Therefore, there is a subsequence converging almost surely, which means that $(M_{n_k} - M)^*$ converges almost surely to 0, and this is precisely the almost sure uniform convergence proposed.

To se that $\|M-N\|_{\mathbf{c}\mathcal{M}_0^2} = 0$ if and only if M and N are indistinguishable, first note that if M and N are indistinguishable, then $M_\infty = \lim M_t = \lim N_t = N_\infty$ almost surely, and therefore $\|M-N\|_{\mathbf{c}\mathcal{M}_0^2} = \|M_\infty - N_\infty\|_2 = 0$. Conversely, if $\|M-N\|_{\mathbf{c}\mathcal{M}_0^2} = 0$ then by Doob's $\mathcal{L}^2$-inequality, $\|(M-N)^*\|_2 = 0$, showing that $(M-N)^*$ is almost surely zero, and this implies indistinguishability. □

Lemma 2.4.10. *The space* $\mathbf{c}\mathcal{M}_0^2$ *is complete.*

Proof. Let (M^n) be a cauchy sequence in $\mathbf{c}\mathcal{M}_0^2$. Then (M^n_∞) is a cauchy sequence in $\mathcal{L}^2(\mathcal{F}_\infty)$, therefore convergent to some $M_\infty \in \mathcal{L}^2(\mathcal{F}_\infty)$. Now define M by putting $M_t = E(M_\infty|\mathcal{F}_t)$, it is clear from Jensen's inequality that M is a martingale bounded in $\mathcal{L}^2$. Since

$$\|(M^n - M)^*_\infty\|_2 \leq 2\|M^n_\infty - M_\infty\|_2,$$

we find that M^n converges uniformly in $\mathcal{L}^2$ to M. As in the proof of Lemma 2.4.9, we then find that there is a subsequence such that M^{n_k} converges almost surely uniformly to M. In particular, since M^{n_k} is continuous, M is almost surely continuous. Since we have assumed the usual conditions, we can pick be a modification N of M which is adapted and continuous for all sample paths. We then find that $N \in \mathbf{c}\mathcal{M}_0^2$ and M^n converges to N in $\mathbf{c}\mathcal{M}_0^2$, as desired. □

Lemma 2.4.11. *Let M be a continuous martingale and let τ be a stopping time. Then the stopped process M^τ is also a continuous martingale.*

Proof. Let $0 \leq s \leq t$. We write

$$E(M^\tau_t|\mathcal{F}_s) = 1_{(\tau \leq s)}E(M^\tau_t|\mathcal{F}_s) + 1_{(\tau > s)}E(M^\tau_t|\mathcal{F}_s)$$

and consider the two terms separately. Since $1_{(\tau\leq s)}$ is $\mathcal{F}_s$ measurable, we find

$$\begin{aligned} 1_{(\tau\leq s)}E(M^\tau_t|\mathcal{F}_s) &= E(M^\tau_t 1_{(\tau\leq s)}|\mathcal{F}_s) \\ &= E(M^\tau_s 1_{(\tau\leq s)}|\mathcal{F}_s) \\ &= M^\tau_s 1_{(\tau\leq s)}, \end{aligned}$$

since $M^\tau_t 1_{(\tau\leq s)} = M^\tau_s 1_{(\tau\leq s)}$, which is $\mathcal{F}_s$ measurable. For the second term, we find $1_{(\tau>s)}E(M^\tau_t|\mathcal{F}_s) = E(1_{(\tau>s)}M^\tau_t|\mathcal{F}_s)$, and for $F \in \mathcal{F}_s$ we have $F \cap (\tau > s) \in \mathcal{F}_{\tau\wedge s}$ by Lemma 2.3.8, yielding

$$\begin{aligned} E(1_F E(1_{(\tau>s)}M^\tau_t|\mathcal{F}_s)) &= E(1_F 1_{(\tau>s)}M^\tau_t) \\ &= E(1_F 1_{(\tau>s)}E(M^\tau_t|\mathcal{F}_{\tau\wedge s})) \\ &= E(1_F 1_{(\tau>s)}M^\tau_s), \end{aligned}$$

by optional sampling, so that $1_{(\tau>s)}E(M_t^\tau|\mathcal{F}_s) = 1_{(\tau>s)}M_s^\tau$. All in all, we conclude $E(M_t^\tau|\mathcal{F}_s) = M_s^\tau$. $\square$

The following lemma will be useful in several of the calculations we are to make when developing the stochastic integral.

Lemma 2.4.12. *Let $M \in \mathbf{c}\mathcal{M}_0^2$, and let $(t_0, \ldots, t_n)$ be a partition of $[s,t]$. Then*

$$E\left(\sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^2 \middle| \mathcal{F}_s\right) = E((M_t - M_s)^2|\mathcal{F}_s)$$

Proof. We see that

$$\begin{aligned}
& E((M_t - M_s)^2|\mathcal{F}_s) \\
= \; & E\left(\left(\sum_{k=1}^n M_{t_k} - M_{t_{k-1}}\right)^2 \middle| \mathcal{F}_s\right) \\
= \; & \sum_{k=1}^n E((M_{t_k} - M_{t_{k-1}})^2|\mathcal{F}_s) + \sum_{k\neq i}^n E((M_{t_k} - M_{t_{k-1}})(M_{t_i} - M_{t_{i-1}})|\mathcal{F}_s),
\end{aligned}$$

where, for $k < i$,

$$\begin{aligned}
& E((M_{t_k} - M_{t_{k-1}})(M_{t_i} - M_{t_{i-1}})|\mathcal{F}_s) \\
= \; & E((M_{t_k} - M_{t_{k-1}})E(M_{t_i} - M_{t_{i-1}}|\mathcal{F}_{t_{i-1}})|\mathcal{F}_s) \\
= \; & 0.
\end{aligned}$$

This shows the lemma. $\square$

We end the section with a very useful criterion for determining when a process is a martingale or a uniformly integrable martingale.

Lemma 2.4.13. *Let M be a progressive process such that the limit exists almost surely. If $EM_\tau = 0$ for any bounded stopping time τ, M is a martingale. If $EM_\tau = 0$ for any stopping time τ, M is a uniformly integrable martingale.*

Proof. The statement that M is a uniformly integrable martingale if $EM_\tau = 0$ for any stopping time τ is Theorem II.77.6 of Rogers & Williams (2000a). If we only have that $EM_\tau = 0$ for any bounded stopping time, M^t is a uniformly integrable martingale for any $t \geq 0$ and therefore, M is a martingale. $\square$

2.5 Localisation

In our construction of the stochastic integral, we will employ the technique of localisation, where we extend properties and definitions to processes where a given property only holds locally in the sense that if we somehow cut off the process from a certain point onwards, then the property holds.

We will mainly be using two localisation concepts: With $\mathbb{H}$ a family of stochastic processes, we say that a process is locally in $\mathbb{H}$ if there exists a sequence of stopping times increasing to infinity such that X^{τ_n} is in $\mathbb{H}$ and we say that a process is zero-locally in $\mathbb{H}$ if there exists a sequence of stopping times increasing to infinity such that $X[0, \tau_n]$ is in $\mathbb{H}$ for $n \geq 1$. We also write that X is $\mathfrak{L}$-locally in $\mathbb{H}$ and that X is $\mathfrak{L}_0$-locally in $\mathbb{H}$, respectively.

We will need a variety of results about these localisation procedures. To avoid having to prove the same results twice, we define an abstract concept of localisation and prove the results for this abstract type of localisation. This also has the convenient side-effect of demonstrating what is the essential properties of localisation: The ability to determine a process from its local properties, and the ability to paste together a new process from localised processes. Our definition of localisation is sufficiently strong to have these properties, but also sufficiently broad to encompass the two concepts of localisation that we will need.

In the following, we denote by $\mathbb{SP}$ the set of real stochastic processes on a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ satisfying the usual conditions.

Definition 2.5.1. *A localisation concept $\mathbb{L}$ is a family of mappings $f_\tau : \mathbb{SP} \to \mathbb{SP}$, indexed by the set of stopping times, such that*

1. *For any stopping times τ and σ, $f_\tau \circ f_\sigma = f_{\tau \wedge \sigma}$.*
2. *If $f_\tau(X) = f_\tau(Y)$, then $X_t(\omega) = Y_t(\omega)$ whenever $t \leq \tau(\omega)$.*

The interpretation of the above is the following. The mappings f_τ correspond to cutoff functions, such that $f_\tau(X)$ is the restriction of X to $\{(t, \omega) | t \leq \tau(\omega)\}$ in some sense. The first property ensures that the order of cutting off is irrelevant. The second property clarifies that $f_\tau(X)$ corresponds to cutting off X to $\{(t, \omega) | t \leq \tau(\omega)\}$. The two localisation concepts mentioned earlier is covered by our definition. Specifically, define the localisation concept of stopping $\mathfrak{L}$ as the family of mappings $f_\tau(X) = X^\tau$ and the localisation concept of zero-stopping $\mathfrak{L}_0$ as the family of mappings $f_\tau(X) = X[0,\tau]$. Then $\mathfrak{L}$ and $\mathfrak{L}_0$ are both localisation concepts according to Definition 2.5.1.

Our abstract localisation concept does not cover more esoteric localisation concepts such as the concept of pre-stopping as found in Protter (2005), Chapter IV. If, however, we exchanged $\{(t,\omega)|t \leq \tau(\omega)\}$ with $\{(t,\omega)|t < \tau(\omega)\}$, it would cover pre-stopping as well. This is unnecessary for our purposes, however, and we therefore omit this extension.

In the following, let $\mathbb{L} = (f_\tau)$ be a localisation concept. Let $\mathbb{H}$ be a subset of $\mathbb{SP}$.

Definition 2.5.2. *We say that a sequence τ_n of stopping times is determining if τ_n increases to infinity. Let X be a stochastic process. We say that X is $\mathbb{L}$-locally in $\mathbb{H}$ if there is a determining sequence τ_n such that $f_{\tau_n}(X) \in \mathbb{H}$ for all $n \geq 1$. In this case, we say that τ_n is a localising sequence for X. The set of elements $\mathbb{L}$-locally in $\mathbb{H}$ is denoted $\mathbb{L}(\mathbb{H})$.*

Note that in the above, we defined two different labels for sequences of stopping times. A sequence of stopping times (τ_n) is determining if it increases to infinity. The same sequence is localising for X to $\mathbb{H}$ if $f_{\tau_n}(X) \in \mathbb{H}$ for $n \geq 1$. The interpretation of Definition 2.5.2 is that when a process X is locally in $\mathbb{H}$, we can cut off with the localising sequence τ_n and then the result $f_{\tau_n}(X)$ will be in $\mathbb{H}$. The reason for using determining sequences is to ensure that the "localised" versions $f_{\tau_n}(X)$ actually determine the element, as the following result shows.

Lemma 2.5.3. *If τ_n is a determining sequence and $f_{\tau_n}(X) = f_{\tau_n}(Y)$ for all $n \geq 1$, then X and Y are equal.*

Proof. It follows that $X_t(\omega)$ and $Y_t(\omega)$ are equal whenever $t \leq \tau_n(\omega)$. Since τ_n tends to infinity, we obtain $X = Y$. $\square$

We will now introduce the concept of a stable subset and develop some basic results about localisation and stability. When we have done this, we continue to the main results about localisation: The ability to paste localised processes together and the ability to extend mappings $\mathbb{H}_1 \to \mathbb{H}_2$ to mappings between localised versions of $\mathbb{H}_1$ and $\mathbb{H}_2$. In order to make the contents of the following easier to understand, we will comment on the meaning of the results in the context of our staple example, that of the space of martingales. We say that M is a local martingale if M $\mathfrak{L}$-locally is a martingale.

Definition 2.5.4. *We say that $\mathbb{H}$ is $\mathbb{L}$-stable if $f_\tau(\mathbb{H}) \subseteq \mathbb{H}$ for any stopping time τ. If the localisation concept is clear from the context, we merely say that $\mathbb{H}$ is stable.*

Lemma 2.5.5. *If $\mathbb{H}$ is stable, then $\mathbb{L}(\mathbb{H})$ is also stable.*

Proof. Let $X \in \mathbb{L}(\mathbb{H})$. We need to show that $f_\tau(X) \in \mathbb{L}(\mathbb{H})$ for each stopping time τ. Let τ be given. Since $X \in \mathbb{L}(\mathbb{H})$, we know that there is a determining sequence σ_n such that $f_{\sigma_n}(X) \in \mathbb{H}$. Since $\mathbb{H}$ is stable, $f_{\sigma_n}(f_\tau(X)) = f_\tau(f_{\sigma_n}(X)) \in \mathbb{H}$. Thus, σ_n is a localising sequence for $f_\tau(X)$, so $f_\tau(X) \in \mathbb{L}(\mathbb{H})$. We conclude that $\mathbb{L}(\mathbb{H})$ is stable. □

Lemma 2.5.6. *If $\mathbb{H}$ is stable, $\mathbb{H} \subseteq \mathbb{L}(\mathbb{H})$ and any determining sequence is localising for any element of $\mathbb{H}$.*

Proof. Let $X \in \mathbb{H}$ and let τ_n be any determining sequence. Then $f_{\tau_n}(X) \in \mathbb{H}$, so $X \in \mathbb{L}(\mathbb{H})$ and τ_n is localising for X. □

Lemma 2.5.7. *Let $\mathbb{H}$ be stable and let $X \in \mathbb{L}(\mathbb{H})$ with localising sequence τ_n. Let σ be any stopping time. Then τ_n is also a localising sequence for $f_\sigma(X)$.*

Proof. Since $\mathbb{H}$ is stable, $f_{\tau_n}(f_\sigma(X)) = f_\sigma(f_{\tau_n}(X)) \in \mathbb{H}$. □

The above lemmas are fundamental workhorses for localisation. Let us see how they apply to the space of martingales. According to Lemma 2.4.11, the space of continuous martingales is stable. Therefore, Lemma 2.5.5 tells us that the space of local continous martingales is also stable. So if M is a local continuous martingale, we know that M^τ is also a local continuous martingale. Lemma 2.5.6 tells us that whenever M is a continous martingale, M is also a local continuous martingale, and any determining sequence is localising for M. This is obviously a very sensible result.

Finally, Lemma 2.5.7 tells us that when we stop a local process, we can use the same determining process for the stopped local process as for the local process.

Lemma 2.5.8. *Let $\mathbb{H}$ be stable and assume that X and Y are locally in $\mathbb{H}$ with determining sequences τ_n and σ_n, respectively. Then $\tau_n \wedge \sigma_n$ is localising for both X and Y.*

Proof. Since $\mathbb{H}$ is stable, it is clear that $f_{\tau_n \wedge \sigma_n}(X)$ and $f_{\tau_n \wedge \sigma_n}(Y)$ are both in $\mathbb{H}$. And since τ_n and σ_n both increase to infinity, so does $\tau_n \wedge \sigma_n$. □

Lemma 2.5.9. *Assume that $\mathbb{H}$ is a linear space which is stable. If each f_τ is linear, $\mathbb{L}(\mathbb{H})$ is a stable linear space.*

Proof. From Lemma 2.5.5, we know that $\mathbb{L}(\mathbb{H})$ is stable. We just have to show that $\mathbb{L}(\mathbb{H})$ is linear. To this end, let $X, Y \in \mathbb{L}(\mathbb{H})$ and let $\lambda, \mu \in \mathbb{R}$, we want to show that $\lambda X + \mu Y \in \mathbb{L}(\mathbb{H})$. By Lemma 2.5.8, there exists a common localising sequence τ_n for X and Y. Since f_τ is linear and $\mathbb{H}$ is a linear space, we have

$$f_{\tau_n}(\lambda X + \mu Y) = \lambda f_{\tau_n}(X) + \mu f_{\tau_n}(Y) \in \mathbb{H}.$$

This shows that τ_n is a localising sequence for $\lambda X + \mu Y$, so $\lambda X + \mu Y \in \mathbb{L}(\mathbb{H})$. □

Lemma 2.5.9 can be used to tell us that since the space of continous martingales obviously is linear, then the space of local continuous martingales is linear as well.

Next, we prove the important Pasting Lemma. This is the lemma which will allow us to construct new processes by local processes which fit together properly.

Lemma 2.5.10 (Pasting Lemma). *Let τ_n be a determining sequence and let X_n be a sequence of processes such that $f_{\tau_n}(X_{n+1}) = X_n$. There exists a unique process X such that $f_{\tau_n}(X) = f_{\tau_n}(X_n)$ for all $n \geq 1$.*

Proof. For any stopping time τ, we define $A_\tau = \{(t, \omega) | t \leq \tau(\omega)\}$. We first prove by induction that $f_{\tau_n}(X_{n+k}) = X_n$ for all $k \geq 1$. By assumption, the induction start holds. Assume that the statement is proven for some k. Since τ_n is increasing, $\tau_n \wedge \tau_{n+1} = \tau_n$ and therefore

$$f_{\tau_n}(X_{n+k+1}) = f_{\tau_n}(f_{\tau_{n+1}}(X_{n+k+1})) = f_{\tau_n}(X_{n+1}) = X_n.$$

This completes the induction proof. We find $f_{\tau_n}(X_{n+k}) = f_{\tau_n}(f_{\tau_n}(X_{n+k})) = f_{\tau_n}(X_n)$, so X_{n+k} and X_n are equal on A_{τ_n}. Defining X by letting X be equal to X_1 on A_{τ_1} and X_n on $A_{\tau_n} \setminus A_{\tau_{n-1}}$, $n \geq 2$, X is well-defined since τ_n is determining and we see that X has the property that X and X_n are equal on A_{τ_n}, and therefore $f_{\tau_n}(X) = f_{\tau_n}(X_n)$.

That X is unique follows from Lemma 2.5.3. □

Theorem 2.5.11. *Let $\mathbb{L}_1 = (f_\tau)$ and $\mathbb{L}_2 = (g_\tau)$ be two localisation concepts. Let $\mathbb{H}_1$ and $\mathbb{H}_2$ be spaces of stochastic processes such that $\mathbb{H}_1$ is $\mathbb{L}_1$-stable and $\mathbb{H}_2$ is $\mathbb{L}_2$-stable. Let $\alpha : \mathbb{H}_1 \to \mathbb{H}_2$ be a mapping with $g_\tau(\alpha(X)) = \alpha(f_\tau(X))$. There exists an extension $\alpha : \mathbb{L}_1(\mathbb{H}_1) \to \mathbb{L}_2(\mathbb{H}_2)$, uniquely defined by the criterion that $g_\tau(\alpha(X)) = \alpha(f_\tau(X))$.*

Proof. We proceed by first constructing a candidate mapping α_0 from $\mathbb{L}_1(\mathbb{H}_1)$ to $\mathbb{L}_2(\mathbb{H}_2)$ and then show that it is an extension of α with the desired properties.

Step 1: Construction of the mapping. Let $X \in \mathbb{L}_1(\mathbb{H}_1)$ with localising sequence τ_n. Then

$$g_{\tau_n}(\alpha(f_{\tau_{n+1}}(X))) = \alpha((f_{\tau_n} \circ f_{\tau_{n+1}})(X)) = \alpha(f_{\tau_n}(X)).$$

The Pasting Lemma 2.5.10 yields the existence of some process Y with the property that $g_{\tau_n}(Y) = \alpha(f_{\tau_n}(X))$. We want to define $\alpha_0(X)$ as this process Y. For this to be meaningful, we need to check that Y does not depend on the localising sequence used. Therefore, let τ_n^* be another localising sequence for X and let Y^* be the element such that $g_{\tau_n^*}(Y^*) = \alpha(f_{\tau_n^*}(X))$. We want to show that $Y = Y^*$. Since τ_n and τ_n^* both are localising sequences for X, by Lemma 2.5.8, so is $\tau_n^* \wedge \tau_n$, and

$$\begin{aligned}
g_{\tau_n^* \wedge \tau_n}(Y) &= (g_{\tau_n^*} \circ g_{\tau_n})(Y) \\
&= g_{\tau_n^*}(\alpha(f_{\tau_n}(X))) \\
&= \alpha((f_{\tau_n^*} \circ f_{\tau_n})(X)) \\
&= \alpha((f_{\tau_n} \circ f_{\tau_n^*})(X)) \\
&= g_{\tau_n}(\alpha(f_{\tau_n^*}(X))) \\
&= (g_{\tau_n} \circ g_{\tau_n^*})(Y^*) \\
&= g_{\tau_n^* \wedge \tau_n}(Y^*)
\end{aligned}$$

Since $\tau_n^* \wedge \tau_n$ is determining, $Y = Y^*$. Therefore, we can define $\alpha_0 : \mathbb{L}_1(\mathbb{H}_1) \to \mathbb{L}_2(\mathbb{H}_2)$ by putting $\alpha_0(X) = Y$. We need to show that α_0 is an extension of α and has the localisation property mentioned in the theorem.

Step 2: α_0 is an extension of α. To show that α_0 is an extension, let an element $X \in \mathbb{H}_1$ be given, and let τ_n be a determining sequence. Then τ_n is localising for X, and therefore $g_{\tau_n}(\alpha_0(X)) = \alpha(f_{\tau_n}(X)) = g_{\tau_n}(\alpha(X))$. Thus $\alpha_0(X) = \alpha(X)$.

Step 3: α_0 has the localisation property. We show $g_\sigma(\alpha_0(X)) = \alpha_0(f_\sigma(X))$ for any σ. Still letting τ_n be localising for X, we know by Lemma 2.5.7 that τ_n is also localising for any $f_\sigma(X)$, and therefore

$$\begin{aligned} g_{\tau_n}(g_\sigma(\alpha_0(X))) &= g_\sigma(g_{\tau_n}(\alpha_0(X))) \\ &= g_\sigma(\alpha(f_{\tau_n}(X))) \\ &= \alpha((f_\sigma \circ f_{\tau_n})(X)) \\ &= \alpha((f_{\tau_n} \circ f_\sigma)(X)) \\ &= g_{\tau_n}(\alpha_0(f_\sigma(X))). \end{aligned}$$

Since τ_n is determining, we conclude $g_\sigma(\alpha_0(X)) = \alpha_0(f_\sigma(X))$.

Step 4: Uniqueness. Finally, assume that α_1 and α_2 are two extensions of α satisfying the criterion in the theorem. We want to show that $\alpha_1 = \alpha_2$. Let an element $X \in \mathbb{L}_1(\mathbb{H}_1)$ be given and let τ_n be a localising sequence. We obtain

$$\begin{aligned} g_{\tau_n}(\alpha_1(X)) &= \alpha_1(f_{\tau_n}(X)) \\ &= \alpha(f_{\tau_n}(X)) \\ &= \alpha_2(f_{\tau_n}(X)) \\ &= g_{\tau_n}(\alpha_2(X)). \end{aligned}$$

Since τ_n is determining, $\alpha_1(X) = \alpha_2(X)$, as desired. □

Corollary 2.5.12. *Let $\mathbb{H}_1$ and $\mathbb{H}_2$ be linear spaces. If the mapping $\alpha : \mathbb{H}_1 \to \mathbb{H}_2$ in Theorem 2.5.11 is linear and each element of the localisation concepts $\mathbb{L}_1$ and $\mathbb{L}_2$ are linear, then the extension $\alpha : \mathbb{L}_1(\mathbb{H}_1) \to \mathbb{L}_2(\mathbb{H}_2)$ is linear as well.*

Proof. By Lemma 2.5.9, $\mathbb{L}_1(\mathbb{H}_1)$ and $\mathbb{L}_2(\mathbb{H}_2)$ are linear spaces, so the conclusion of the corollary is well-defined. The extension satisfies $g_\tau(\alpha(X)) = \alpha(f_\tau(x))$. Let $X, Y \in \mathbb{L}_1(\mathbb{H}_1)$ and let $\lambda, \mu \in \mathbb{R}$. Let τ_n be a localising sequence for X and Y, then $f_{\tau_n}(X), f_{\tau_n}(Y) \in \mathbb{H}_1$. Therefore, by the linearity of α on $\mathbb{H}_1$,

$$\begin{aligned} g_{\tau_n}(\alpha(\lambda X + \mu Y)) &= \alpha(f_{\tau_n}(\lambda X + \mu Y)) \\ &= \alpha(\lambda f_{\tau_n}(X) + \mu f_{\tau_n}(Y)) \\ &= \lambda\alpha(f_{\tau_n}(X)) + \mu\alpha(f_{\tau_n}(Y)) \\ &= \lambda g_{\tau_n}(\alpha(X)) + \mu g_{\tau_n}(\alpha(Y)) \\ &= g_{\tau_n}(\lambda\alpha(X) + \mu\alpha(Y)). \end{aligned}$$

Since τ_n is determining, we may conclude that α is linear. □

Lemma 2.5.13. *Assume that $\mathbb{H}$ is linear and that $f_{\tau\vee\sigma} + f_{\tau\wedge\sigma} = f_\tau + f_\sigma$. If $\mathbb{H}$ is $\mathbb{L}$-stable, then $\mathbb{L}(\mathbb{L}(\mathbb{H})) = \mathbb{L}(\mathbb{H})$.*

Proof. Since $\mathbb{H}$ is $\mathbb{L}$-stable, $\mathbb{L}(\mathbb{H})$ is $\mathbb{L}$-stable as well, and therefore $\mathbb{L}(\mathbb{H}) \subseteq \mathbb{L}(\mathbb{L}(\mathbb{H}))$. We need to show the other inclusion. Before doing so, we make an observation. Let τ and σ be stopping times, and assume that $f_\tau(X), f_\sigma(X) \in \mathbb{H}$. Since $\mathbb{H}$ is stable and linear, we find $f_{\tau\vee\sigma}(X) = f_\tau(X) + f_\sigma(X) - f_{\tau\wedge\sigma}(X) \in \mathbb{H}$.

Now let $X \in \mathbb{L}(\mathbb{L}(\mathbb{H}))$ and let τ_n be a localising sequence such that $f_{\tau_n}(X) \in \mathbb{L}(\mathbb{H})$. For any n, let σ_m^n be a localising sequence such that $f_{\sigma_m^n}(f_{\tau_n}(X) \in \mathbb{H}$. Since σ_m^n tends to infinity almost surely, $\lim_m P(\sigma_m^n \le n) = 0$ for all $n \ge 1$. For any $n \ge 1$, pick $m(n)$ such that $P(\sigma_{m(n)}^n \le n) \le \frac{1}{2^n}$. Obviously, then $\sum_{n=1}^\infty P(\sigma_{m(n)}^n \le n) < \infty$, so the Borel-Cantelli Lemma yields $\sigma_{m(n)}^n \xrightarrow{\text{a.s.}} \infty$. Since also $\tau_n \xrightarrow{\text{a.s.}} \infty$, we may conclude $\sigma_{m(n)}^n \wedge \tau_n \xrightarrow{\text{a.s.}} \infty$. Now put

$$\rho_n = \max_{k\le n} \sigma_{m(k)}^k \wedge \tau_k.$$

Then ρ is a sequence of stopping times tending almost surely to infinity, and by our earlier observation, $f_{\rho_n}(X) \in \mathbb{H}$. Therefore $X \in \mathbb{L}(\mathbb{H})$. □

The criterion in Lemma 2.5.13 is satisfied both for stopping and zero-stopping. We are now done with general results on localisation. We proceed to give a few results regarding more specific matters of localisation.

Definition 2.5.14. *Let $\mathbb{H}$ be a class of stochastic proceses. We denote the continuous elements of $\mathbb{H}$ by $\mathbf{c}\mathbb{H}$.*

Lemma 2.5.15. *It always holds that $\mathbf{c}\mathfrak{L}(\mathbb{H}) = \mathfrak{L}(\mathbf{c}\mathbb{H})$.*

Proof. Let $X \in \mathbf{c}\mathfrak{L}(\mathbb{H})$, and let τ_n be a localizing sequence for X. Since we have $X_t^{\tau_n} = X_{\tau_n\wedge t}$ and X is continuous, clearly X^{τ_n} is continuous as well and therefore $X \in \mathfrak{L}(\mathbf{c}\mathbb{H})$. Contrarily, assume that $X \in \mathfrak{L}(\mathbf{c}\mathbb{H})$, and let τ_n be a localizing sequence for X. Since X^{τ_n} is continuous, X is continuous on $[0, \tau_n]$. Because τ_n tends to infinity, it follows that X is continuous and therefore $X \in \mathbf{c}\mathfrak{L}(\mathbb{H})$. □

Lemma 2.5.16. *Let M be a local martingale. Then M is locally a uniformly integrable martingale.*

Proof. Let τ_n be a sequence such that M^{τ_n} is a martingale for $n \geq 1$. Then $M^{\tau_n \wedge n}$ is uniformly integrable, since it is a martingale closed by $M_{\tau_n \wedge n}$. The result follows. □

Lemma 2.5.17. *The following are equivalent:*

1. *M is a continuous local martingale.*
2. *M is locally a continuous martingale.*
3. *M is locally a continuous square-integrable martingale.*
4. *M is locally a continuous bounded martingale.*

In particular, $\mathbf{c}\mathfrak{L}(\mathcal{M}_0)$, $\mathfrak{L}(\mathbf{c}\mathcal{M}_0)$, $\mathfrak{L}(\mathbf{c}\mathcal{M}_0^2)$ *and* $\mathfrak{L}(\mathbf{cb}\mathcal{M}_0)$ *are equal.*

Proof. Trivially, (4) implies (3) and (3) implies (2). By Lemma 2.5.15, (2) implies (1). It will therefore suffice to show that (1) implies (4). Assume that M is a continuous local martingale. Let τ_n be a localizing sequence. Define the stopping times σ_n by $\sigma_n = \inf\{t \geq 0 || M(t)| \geq n\}$. Since M is continuous, M is bounded on bounded intervals, and therefore σ_n almost surely increases to infinity. Since martingales are stable under stopping by Lemma 2.4.11, it follows that $\tau_n \wedge \sigma_n$ is a localizing sequence, and since M is continuous, $M^{\tau_n \wedge \sigma_n}$ is bounded. □

By Lemma 2.5.17, the spaces $\mathbf{c}\mathfrak{L}(\mathcal{M}_0)$, $\mathfrak{L}(\mathbf{c}\mathcal{M}_0)$, $\mathfrak{L}(\mathbf{c}\mathcal{M}_0^2)$ and $\mathfrak{L}(\mathbf{cb}\mathcal{M}_0)$ are all equal. We will denote them by the common symbol $\mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$.

Lemma 2.5.18. *It holds that* $\mathcal{L}^2([0,\infty) \times \Omega, \Sigma_\pi, \mu)$ *is* $\mathfrak{L}_0$*-stable for any measure* μ.

Proof. Let $X \in \mathcal{L}^2([0,\infty) \times \Omega, \Sigma_\pi, \mu)$ and let τ be any stopping time. We need to argue that $X[0,\tau]$ is Σ_π measurable and integrable. By Lemma 2.3.4, $X[0,\tau]$ is Σ_π measurable. By $|X[0,\tau]| \leq |X|$, $X[0,\tau]$ is integrable. □

Lemma 2.5.19. *Let M be a continuous local martingale starting at zero. Let $p > 1$ and assume that*

$$\sup_\tau E|M_\tau|^p < \infty,$$

where the supremum is taken over all bounded stopping times. Then M is a martingale bounded in $\mathcal{L}^p$.

Proof. Let τ_n be a localising sequence for M and let τ be any bounded stopping time. Then, $\tau_n \wedge \tau$ is a bounded stopping time for each n, and $M_{\tau_n \wedge \tau}$ is bounded in $\mathcal{L}^p$, therefore uniformly integrable. Since M is continuous, $M_{\tau_n \wedge \tau}$ converges to M_τ pointwise. In particular, $M_{\tau_n \wedge \tau}$ converges in probability to M_τ. By Lemma B.5.3, $M_{\tau_n \wedge \tau}$ converges in $\mathcal{L}^1$ to M_τ. Therefore, $EM_\tau = \lim EM_{\tau_n \wedge \tau} = 0$. By Lemma 2.4.13, M is a martingale. $\square$

Chapter 3

The Itô Integral

In this chapter, we will develop the theory of integration with respect to a simple class of stochastic processes, which nevertheless will be rich enough to cover all of our purposes. Our goal will be to create a relatively self-contained and thorough presentation of the theory. We will not merely be concerned with results for use in the later development of the Malliavin calculus and the applications in finance, but will seek to convey a general understanding of the properties of the integral. Our construction will mainly follow the methods of Steele (2000), Karatzas & Shreve (1988), Øksendal (2005) and Rogers & Williams (2000b).

We will work on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with a $\mathcal{F}_t$ brownian motion W satisfying the usual conditions. Such a probability space exists according to our results from Section 2.1. Our first task will be to define a meaningful stochastic integral with respect to W. The construction will be done in several steps, first defining the integral for a simple class of integrands, and then extending the space of integrands in several steps.

After establishing the basic theory of integration with respect to one-dimensional Brownian motion, we will consider a n-dimensional Brownian motion and extend the integral to integrators of the form $M_t = \sum_{k=1}^n \int_0^t Y_s^k \, \mathrm{d}W_s^k$. The development of the integral for this type of processes will mainly be done by reusing parts of the theory for the one-dimensional brownian case. We will later see that in a special case, this class of integrands actually is equal to the class of local martingales. After defining the integral for these integrators, we extend the theory to integrators which also have a component of continuous paths with finite variation. The stochastic integral of a process Y with respect to a process X will be denoted $I_X(Y)$ or $Y \cdot X$. We also use the notation $(Y \cdot X)_t = \int_0^t Y_s \,\mathrm{d}X_s$.

After the construction of the integral, we will prove some of the fundamental theorems of the theory: the Itô formula, the Martingale Representation Theorem and Girsanov's Theorem.

Before beginning work on the stochastic integral with respect to Brownian motion, we develop some general results on the space of elementary processes that will prove very useful. This allows us a separation of concerns: We will be able to see which elements of the theory essentially only concern the nature of the space of elementary processes and which elements concern the Brownian motion specifically. We will also avoid letting the development of the integral be disturbed by the somewhat tedious results for the space of elementary processes.

The plan for the chapter is thus as follows.

1. Develop some basic results for elementary processes.
2. Construct the integral with respect to Brownian motion.
3. Construct the integral with respect to integrators of type $M_t = \sum_{k=1}^n \int_0^t Y_s^k \,\mathrm{d}W_s^k$.
4. Construct the integral with respect to integrators with finite variation.
5. Establish the basic results of the theory.

3.1 The space of elementary processes

The purpose of this section is twofold: First, we will show how to define the integral rigorously on the set $\mathbf{b}\mathcal{E}$ of elementary processes. Second, we will prove a density result for $\mathbf{b}\mathcal{E}$ and use this result to demonstrate how to extend a class of mappings from $\mathbf{b}\mathcal{E}$ to certain $\mathcal{L}^2$ spaces.

Definition 3.1.1. *Let Z be a stochastic variable and let σ and τ be stopping times. We define the process $Z(\sigma,\tau]$ by $Z(\sigma,\tau]_t = Z1_{(\sigma,\tau]}(t)$. We define the class of elementary* *processes* $\mathbf{b}\mathcal{E}$ *by*

$$\mathbf{b}\mathcal{E} = \left\{ \sum_{k=1}^{n} Z_k(\sigma_k, \tau_k] \middle| \sigma_k \le \tau_k \textit{ are bounded stopping times and } Z_k \in \mathbf{b}\mathcal{F}_{\sigma_k} \right\}.$$

Clearly, $\mathbf{b}\mathcal{E}$ is a linear space. Let M be any stochastic process, and let $H \in \mathbf{b}\mathcal{E}$ be given with $H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k]$. Intuitively, we would like to define the integral $I_M(H)_t$ of H with respect to M over $[0,t]$ by the Riemann-sum-like expression

$$I_M(H)_t = \sum_{k=1}^{n} Z_k(M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}).$$

In order for this definition to work, however, we must make sure that this definition does not depend on the representation of H. It is clear that if $H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k]$, we also find $H = \sum_{k=1}^{n-1} Z_k(\sigma_k, \tau_k] + \frac{1}{2} Z_k(\sigma_k, \tau_k] + \frac{1}{2} Z_k(\sigma_k, \tau_k]$, or, with bounded stopping times ρ_k such that $\sigma_k \le \rho_k \le \tau_k$, $H = \sum_{k=1}^{n-1} Z_k(\sigma_k, \rho_k] + Z_k(\rho_k, \tau_k]$. In other words, H can be represented on the form $\sum_{k=1}^n Z_k(\sigma_k, \tau_k]$ in many equivalent ways. We will now check that the definition of the stochastic integral does not depend on the representation.

Lemma 3.1.2. *Let $H, K \in \mathbf{b}\mathcal{E}$ with $H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k]$ and $K = \sum_{k=1}^m Z'_k(\sigma'_k, \tau'_k]$. There exists representations of H and K based on the same stopping times. One particular such representation can be obtained by putting*

$$H = \sum_{i=1}^{2n+2m-1} Y_i(U_i, U_{i+1}] \textit{ and } K = \sum_{i=1}^{2n+2m-1} Y'_i(U_i, U_{i+1}],$$

where U_i is the i'th ordered value amongst the $2n + 2m$ stopping times in the family $\{\sigma_1, \ldots, \sigma_n, \tau_1, \ldots, \tau_n, \sigma'_1, \ldots, \sigma'_m, \tau'_1, \ldots, \tau'_m\}$ and

$$Y_i = \sum_{k=1}^{n} Z_k 1_{(\sigma_k \le U_i < \tau_k)} \textit{ and } Y'_i = \sum_{k=1}^{m} Z'_k 1_{(\sigma'_k \le U_i < \tau'_k)}.$$

Proof. We need to show that the representations for H and K are on the form given for processes in $\mathbf{b}\mathcal{E}$, and we need to check that they are in fact equal to H and K, respectively. As the proofs are the same, we only prove the results for H.

By Lemma 2.3.6, the i'th ordered value U_i is a stopping time. It is clear that U_i is bounded. We want to show that $Y_i \in \mathbf{b}\mathcal{F}_{U_i}$. Boundedness is obvious. Since $Z_k \in \mathcal{F}_{\sigma_k}$, by Lemma 2.3.8, $Z_k 1_{(\sigma_k \le U_i)} \in \mathcal{F}_{\sigma_k \wedge U_i} \subseteq \mathcal{F}_{U_i}$. Therefore, we may conclude by that same lemma that $Z_k 1_{(\sigma_k \le U_i < \tau_k)} = Z_k 1_{(\sigma_k \le U_i)} 1_{(U_i < \tau_k)} \in \mathcal{F}_{U_i \wedge \tau_k}$, and then it follows that Y_i is $\mathcal{F}_{U_i}$ measurable. Thus, $\sum_{i=1}^{2n+2m-1} Y_i (U_i, U_{i+1}]$ is in $\mathbf{b}\mathcal{E}$.

It remains to show that this process is equal to H. To this end, note

$$\begin{aligned}
H_t &= \sum_{k=1}^n Z_k 1_{(\sigma_k, \tau_k]}(t) \\
&= \sum_{k=1}^n Z_k 1_{(\sigma_k, \tau_k] \cap \bigcup_{i=1}^{2n+2m-1} (U_i, U_{i+1}]}(t) \\
&= \sum_{k=1}^n Z_k 1_{(\sigma_k, \tau_k]}(t) \sum_{i=1}^{2n+2m-1} 1_{(U_i, U_{i+1}]}(t) \\
&= \sum_{i=1}^{2n+2m-1} \left(\sum_{k=1}^n Z_k 1_{(\sigma_k, \tau_k]}(t) \right) 1_{(U_i, U_{i+1}]}(t).
\end{aligned}$$

Exploiting the fact that either $(U_i, U_{i+1}] \subseteq (\sigma_k, \tau_k]$ or $(U_i, U_{i+1}] \cap (\sigma_k, \tau_k] = \emptyset$, we obtain

$$\begin{aligned}
&\sum_{i=1}^{2n+2m-1} \left(\sum_{k=1}^n Z_k 1_{(\sigma_k, \tau_k]}(t) \right) 1_{(U_i, U_{i+1}]}(t) \\
= &\sum_{i=1}^{2n+2m-1} \left(\sum_{k=1}^n Z_k 1_{((U_i, U_{i+1}] \subseteq (\sigma_k, \tau_k])}(t) \right) 1_{(U_i, U_{i+1}]}(t) \\
= &\sum_{i=1}^{2n+2m-1} \left(\sum_{k=1}^n Z_k 1_{(\sigma_k \le U_i, U_{i+1} \le \tau_k)}(t) \right) 1_{(U_i, U_{i+1}]}(t) \\
= &\sum_{i=1}^{2n+2m-1} \left(\sum_{k=1}^n Z_k 1_{(\sigma_k \le U_i < \tau_k)}(t) \right) 1_{(U_i, U_{i+1}]}(t) \\
= &\sum_{i=1}^{2n+2m-1} Y_i 1_{(U_i, U_{i+1}]}(t),
\end{aligned}$$

as desired. □

Theorem 3.1.3. *For any stochastic process M, there exists a unique mapping I_M on $\mathbf{b}\mathcal{E}$, taking its value in the space of stochastic processes, such that for H with representation $\sum_{k=1}^n Z_k (\sigma_k, \tau_k]$, it holds that*

$$I_M(H)_t = \sum_{k=1}^n Z_k (M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}).$$

Proof. Uniqueness of the mapping is clear. We need to show that if we have two different representations of an element of $\mathbf{b}\mathcal{E}$, then the stochastic integral based on these representations yield the same result. Assume that we have two different representations of H, $H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k] = \sum_{k=1}^m Z'_k(\sigma'_k, \tau'_k]$. We want to show that

$$\sum_{k=1}^n Z_k(M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}) = \sum_{k=1}^m Z'_k(M_{\tau'_k \wedge t} - M_{\sigma'_k \wedge t}).$$

By Lemma 3.1.2, we have $H = \sum_{i=1}^{2n+2m-1} Y_i(U_i, U_{i+1}] = \sum_{i=1}^{2n+2m-1} Y'_i(U_i, U_{i+1}]$, where U_i is the i'the ordered value amongst the $2n+2m$ stopping times in the family $\{\sigma_1, \ldots, \sigma_n, \tau_1, \ldots, \tau_n, \sigma'_1, \ldots, \sigma'_m, \tau'_1, \ldots, \tau'_m\}$ and

$$Y_i = \sum_{k=1}^n Z_k 1_{(\sigma_k \le U_i < \tau_k)} \text{ and } Y'_i = \sum_{k=1}^m Z'_k 1_{(\sigma'_k \le U_i < \tau'_k)}.$$

We will show the two equalities

$$\begin{aligned} \sum_{k=1}^n Z_k(M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}) &= \sum_{i=1}^{2n+2m-1} Y_i(M_{U_{i+1} \wedge t} - M_{U_i \wedge t}) \\ \sum_{k=1}^m Z'_k(M_{\tau'_k \wedge t} - M_{\sigma'_k \wedge t}) &= \sum_{i=1}^{2n+2m-1} Y'_i(M_{U_{i+1} \wedge t} - M_{U_i \wedge t}). \end{aligned}$$

The proof is basically the same for both equalities, so we only show the first one. We get

$$\begin{aligned} &\sum_{i=1}^{2n+2m-1} Y_i(M_{U_{i+1} \wedge t} - M_{U_i \wedge t}) \\ =& \sum_{i=1}^{2n+2m-1} \left(\sum_{k=1}^n Z_k 1_{(\sigma_k \le U_i < \tau_k)} \right) (M_{U_{i+1} \wedge t} - M_{U_i \wedge t}) \\ =& \sum_{k=1}^n Z_k \sum_{i=1}^{2n+2m-1} 1_{(\sigma_k \le U_i < \tau_k)} (M_{U_{i+1} \wedge t} - M_{U_i \wedge t}) \\ =& \sum_{k=1}^n Z_k(M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}), \end{aligned}$$

since the sum in the second to last equation above telescopes. To prove the theorem, then, it will suffice to show

$$\sum_{i=1}^{2n-1} Y_i(M_{U_{i+1} \wedge t} - M_{U_i \wedge t}) = \sum_{i=1}^{2n-1} Y'_i(M_{U_{i+1} \wedge t} - M_{U_i \wedge t}),$$

and to do this, it will suffice to show $Y_i(\omega) = Y_i(\omega)'$ whenever $U_i(\omega) < U_{i+1}(\omega)$. But in this case, letting s be some number between $U_i(\omega)$ and $U_{i+1}(\omega)$, $Y_i(\omega) = H(s) = Y'_i(\omega)$, as desired. $\square$

Theorem 3.1.3 ensures the existence of the stochastic integral I_M on $\mathbf{b}\mathcal{E}$, taking the values we would expect no matter what the representation of the $\mathbf{b}\mathcal{E}$-element. In the following, we will interchangeably use the notation $I_M(H)$ and $H \cdot M$ for the integral of H with respect to M over $[0,t]$. We will also write $I_M(H)_t = \int_0^t H_s \,\mathrm{d}M_s$. Note that using the notation from Chapter 2 for stopped processes, we can write the process $I_M(H)$ in compact form, with $H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k]$, by

$$I_M(H) = \sum_{k=1}^n Z_k(M^{\tau_k} - M^{\sigma_k}).$$

We also want to import meaning to the symbol $\int_s^t H(u)\,\mathrm{d}M(u)$. The following lemmas shows what the proper interpretation is.

Lemma 3.1.4. *Let $0 \leq x \leq y$ and $0 \leq t \leq s$. It holds that*

$$(x,y] \cap (s,t] = (x \wedge t \vee s, y \wedge t \vee s],$$

understanding that $x \wedge t \vee s = (x \wedge t) \vee s$ and $y \wedge t \vee s = (y \wedge t) \vee s$.

Proof. We consider the three possible cases separately.

The case $x \leq y \leq s \leq t$. We find

$$\begin{aligned} (x,y] \cap (s,y] &= \emptyset \\ (x \wedge t \vee s, y \wedge t \vee s] &= (x \vee s, y \vee s] = (s,s] = \emptyset, \end{aligned}$$

so the proposition holds in this case.

The case $x \leq s \leq y \leq t$. In this case, we see that

$$\begin{aligned} (x,y] \cap (s,y] &= (s,y] \\ (x \wedge t \vee s, y \wedge t \vee s] &= (x \vee s, y \vee s] = (s,y], \end{aligned}$$

which verifies the statement in this case also.

The case $s \leq t \leq x \leq y$. In the final case, we obtain

$$\begin{aligned} (x,y] \cap (s,y] &= \emptyset \\ (x \wedge t \vee s, y \wedge t \vee s] &= (t \vee s, t \vee s] = (t,t] = \emptyset, \end{aligned}$$

and this concludes the proof of the lemma. □

Lemma 3.1.5. *Let $H \in \mathbf{b}\mathcal{E}$ with $H = \sum_{k=1}^{n} Z_k(\sigma_k, \tau_k]$. It holds that*

$$\sum_{k=1}^{n} Z_k(M_{\tau_k \wedge t \vee s} - M_{\sigma_k \wedge t \vee s}) = I_M(H)_t - I_M(H)_s.$$

Proof. It will suffice to consider the term $Z_k(M_{\tau_k \wedge t \vee s} - M_{\sigma_k \wedge t \vee s})$ for each of the three possible configurations of the intervals $(\sigma_k, \tau_k]$ and $(s, t]$, and prove it equal to the corresponding term of $I_M(H)_t - I_M(H)_s$.

The case $s \le t \le \sigma_k \le \tau_k$. We get

$$\begin{aligned} & Z_k(M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}) - Z_k(M_{\tau_k \wedge s} - M_{\sigma_k \wedge s}) \\ = \; & Z_k(M_t - M_t) - Z_k(M_s - M_s) \\ = \; & Z_k(M_{\tau_k \wedge t \vee s} - M_{\sigma_k \wedge t \vee s}). \end{aligned}$$

showing the statement in this case.

The case $s \le \sigma_k \le t \le \tau_k$. Here, we find

$$\begin{aligned} & Z_k(M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}) - Z_k(M_{\tau_k \wedge s} - M_{\sigma_k \wedge s}) \\ = \; & Z_k(M_t - M_{\sigma_k}) - Z_k(M_s - M_s) \\ = \; & Z_k(M_{\tau_k \wedge t \vee s} - M_{\sigma_k \wedge t \vee s}). \end{aligned}$$

The case $s \le t \le \sigma_k \le \tau_k$. In the final case, we obtain

$$\begin{aligned} & Z_k(M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}) - Z_k(M_{\tau_k \wedge s} - M_{\sigma_k \wedge s}) \\ = \; & Z_k(M_{\tau_k} - M_{\sigma_k}) - Z_k(M_{\tau_k} - M_{\sigma_k}) \\ = \; & Z_k(M_{\tau_k \wedge t \vee s} - M_{\sigma_k \wedge t \vee s}), \end{aligned}$$

as desired.

Conclusion. All in all, we conclude that

$$\begin{aligned} & \sum_{k=1}^{n} Z_k(M_{\tau_k \wedge t \vee s} - M_{\sigma_k \wedge t \vee s}) \\ = \; & \sum_{k=1}^{n} Z_k(M_{\tau_k \wedge t} - M_{\sigma_k \wedge t}) - Z_k(M_{\tau_k \wedge s} - M_{\sigma_k \wedge s}) \\ = \; & I_M(H)_t - I_M(H)_s, \end{aligned}$$

and the lemma is proved. □

Lemma 3.1.5 shows that the natural definition of the integral $\int_s^t H(u)\,\mathrm{d}M(u)$ coincides with $I_M(H)_t - I_M(H)_s$, which is of course not particularly surprising. In the following, we will by $\int_s^t H(u)\,\mathrm{d}M(u)$ denote the variable $I_M(H)_t - I_M(H)_s$.

Lemma 3.1.6. *The integral mapping I_M on $\mathbf{b}\mathcal{E}$ is linear.*

Proof. Let H and K in $\mathbf{b}\mathcal{E}$ be given with representations

$$H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k] \text{ and } K = \sum_{i=1}^m Y_i(V_i, W_i]$$

and let $\lambda, \mu \in \mathbb{R}$. Then

$$\begin{aligned} & I_M\left(\lambda H + \mu K\right)_t \\ = \;& I_M\left(\sum_{k=1}^n \lambda Z_k(\sigma_k, \tau_k] + \sum_{i=1}^m \mu Y_i(V_i, W_i]\right) \\ = \;& \sum_{k=1}^n \lambda Z_k(M^{\tau_k}(t) - M^{\sigma_k}(t)) + \sum_{i=1}^m \mu Y_i(M^{W_i}(t) - M^{V_i}(t)) \\ = \;& \lambda I_M(H)_t + \mu I_M(K)_t, \end{aligned}$$

which was to be shown. □

We have now shown how to define the integral on $\mathbf{b}\mathcal{E}$ for any process M in a sensible way, and we have checked linearity of the integral. In order to extend the integral to larger spaces of integrands, we will need to require more structure on M. We will see how to do this in the following sections.

Next, we will show an extension result which will be crucial to the later development of the integral. First, we show that there is a large class of $\mathcal{L}^2$ spaces on $[0,\infty)\times\Omega$ which contain $\mathbf{b}\mathcal{E}$ as a dense subspace. Recall that Σ_π denotes the progressive σ-algebra on $[0,\infty)\times\Omega$.

Theorem 3.1.7. *Let μ be a measure on Σ_π which is absolutely continuous with respect to $\lambda \otimes P$, such that $\mu([0,T]\times\Omega) < \infty$ for all $T > 0$. Then $\mathbf{b}\mathcal{E}$ is a dense subspace of $\mathcal{L}^2(\mu)$.*

Proof. By $\|\cdot\|_\mu$, we denote the standard seminorm of $\mathcal{L}^2(\mu)$. We first show that $\mathbf{b}\mathcal{E}$ is a subspace of $\mathcal{L}^2(\mu)$. Let $H \in \mathbf{b}\mathcal{E}$ be given with $H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k]$. Using Lemma

2.3.4, we see that H is progressive. We find

$$
\begin{aligned}
\|H\|_\mu^2 &= \sum_{k=1}^n \int Z_k(\omega) 1_{(\sigma_k(\omega),\tau_k(\omega)]}(t) \,\mathrm{d}\mu(t,\omega) \\
&\leq \sum_{k=1}^n \|Z_k\|_\infty \int 1_{[0,\|\sigma_k\|_\infty + \|\tau_k\|_\infty]\times\Omega} \,\mathrm{d}\mu(t,\omega) \\
&= \sum_{k=1}^n \|Z_k\|_\infty \mu([0,\|\sigma_k\|_\infty + \|\tau_k\|_\infty]\times\Omega) < \infty,
\end{aligned}
$$

as desired. The proof of the density of $\mathbf{b}\mathcal{E}$ proceeds in four steps, each consisting of showing that a particular subset of $\mathcal{L}^2(\mu)$ is dense in $\mathcal{L}^2(\mu)$. The four subsets are:

1. Processes which are zero from a point onwards.
2. Processes which are zero from a point onwards and bounded.
3. Processes which are zero from a point onwards, bounded and continuous.
4. The set $\mathbf{b}\mathcal{E}$.

Step 1: Approximation by processes which are zero from a point onwards. Let $X \in \mathcal{L}^2(\mu)$ and define $X_n = X 1_{[0,T]\times\Omega}$. Since $[0,T]\times\Omega \in \Sigma_\pi$, X_n is progressive. Clearly $\|X_n\|_\mu \leq \|X\|_\mu < \infty$, so $X_n \in \mathcal{L}^2(\mu)$. And X_n is zero from T onwards. Finally,

$$
\lim \|X - X_n\|_\mu^2 = \lim \int 1_{(T,\infty)\times\Omega} X^2 \,\mathrm{d}\mu = 0
$$

by dominated convergence. This means that in the following, in our work to show density of $\mathbf{b}\mathcal{E}$ in $\mathcal{L}^2(\mu)$, we can assume that the process to be approximated is zero from a point onwards.

Step 2: Approximation by bounded processes. We next show that each element of $\mathcal{L}^2(\mu)$, zero from a point onwards, can be approximated by bounded processes zero from a point onwards. Therefore, let $X \in \mathcal{L}^2(\mu)$ be zero from T onwards and define $X_n(t) = X(t) 1_{[-n,n]}(X_t)$. Then X_n is progressive and bounded. And we clearly have $\|X_n\|_\mu \leq \|X\|_\mu < \infty$, so $X_n \in \mathcal{L}^2(\mu)$, and X_n is zero from T onwards. Furthermore,

$$
\lim_n \|X - X_n\|_\mu^2 = \lim \int 1_{(|X|>n)} X^2 \,\mathrm{d}\mu = 0
$$

by dominated convergence.

Step 3: Approximation by continuous processes. We now show that each element of $\mathcal{L}^2(\mu)$, bounded and zero from a point onwards, can be approximated by continuous processes of the same kind. Let $X \in \mathcal{L}^2(\mu)$ be bounded and zero from T onwards. Put

$$X_n(t) = \frac{1}{\frac{1}{n}} \int_{(t-\frac{1}{n})^+}^t X(s)\,\mathrm{d}s,$$

where $x^+ = \max\{0, x\}$. Clearly, X_n is continuous and $\|X_n\|_\infty \leq \|X\|_\infty < \infty$, so X_n is bounded. For $t \geq T + \frac{1}{n}$, $X_n(t) = 0$. We will show that the sequence X_n is in $\mathcal{L}^2(\mu)$ and approximates X.

Since X is progressive, $X_{|[0,t]\times\Omega}$ is $\mathcal{B}[0,t] \otimes \mathcal{F}_t$ measurable. Therefore, $X_n(t)$ is $\mathcal{F}_t$ measurable. Since X_n is also continuous, X_n is progressive by Lemma 2.2.4. Since X_n is bounded and zero from a certain point onwards, we conclude $X_n \in \mathcal{L}^2(\mu)$.

It remains to show convergence in $\|\cdot\|_\mu$ of X_n to X. By Theorem A.2.4, $X_n(\omega)$ converges Lebesgue almost surely to $X(\omega)$ for all $\omega \in \Omega$. This means that we have convergence $\lambda \otimes P$ everywhere. Absolute continuity yields that X_n converges μ-almost surely to X. Since X_n and X have common support $[0, T+1] \times \Omega$ of finite measure and are bounded by the common constant $\|X\|_\infty$, by dominated convergence we obtain $\lim \|X - X_n\|_\mu = 0$.

Step 4: Approximation by elementary processes. Finally, we show that for each $X \in \mathcal{L}^2(\mu)$ which is continuous, bounded and zero from some point T and onwards, we can approximate X with elements of $\mathbf{b}\mathcal{E}$.

To do this, we define

$$X_n(t) = \sum_{k=0}^{\infty} X\left(\frac{k}{2^n}\right) 1_{[\frac{k}{2^n}, \frac{k+1}{2^n})}(t).$$

Since X_n is zero from T and onwards, the sum defining X_n is zero from the term number $[2^nT] + 2$ and onwards. We therefore conclude $X_n \in \mathbf{b}\mathcal{E}$. Since $\frac{k}{2^n} = \frac{[2^nt]}{2^n}$ when $\frac{k}{2^n} \leq t < \frac{k+1}{2^n}$, we can also write $X_n(t) = X(\frac{[2^nt]}{2^n})$. In particular, X_n is zero from $T+1$ and onwards. Since $\lim \frac{[2^nt]}{2^n} = t$ for all $t \geq 0$, it follows from the continuity of X that X_n converges pointwise to X. As in the previous step, since X_n and X has common support of finite measure and are bounded by the common constant $\|X\|_\infty$, dominated convergence yields $\lim \|X - X_n\|_\mu = 0$. This shows the claim of the theorem. $\square$

Note that in the proof of Theorem 3.1.7, we actually showed a stronger result than just that $\mathbf{b}\mathcal{E}$ is dense in $\mathcal{L}^2(\mu)$. Our final stage of approximations were with processes in the set

$$\mathbf{b}\mathcal{S} = \left\{ \sum_{k=1}^n Z_k(s_k, s_k] \middle| s_k \le t_k \text{ are deterministic times and } Z_k \in \mathbf{b}\mathcal{F}_{\sigma_k} \right\}$$

so we have in fact demonstrated that $\mathbf{b}\mathcal{S}$ is dense in $\mathcal{L}^2(\mu)$, even though we will not need this stronger fact.

In our coming definition of the integral, we will have at our disposal an integral mapping $I : \mathbf{b}\mathcal{E} \to \mathbf{c}\mathcal{M}_0^2$ which is linear and isometric, when $\mathbf{b}\mathcal{E}$ is considered as a subspace of a suitable space $\mathcal{L}^2$ space on Σ_π. The following theorem shows that such mappings always can be extended from $\mathbf{b}\mathcal{E}$ to the $\mathcal{L}^2$ space.

Theorem 3.1.8. *Let μ be a measure on Σ_π, absolutely continuous with respect to $\lambda \otimes P$ and such that $\mu([0,T] \times \Omega) < \infty$ for all $T > 0$. Let $I : \mathbf{b}\mathcal{E} \to \mathbf{c}\mathcal{M}_0^2$ be linear and isometric, considering $\mathbf{b}\mathcal{E}$ as a subspace of $\mathcal{L}^2(\mu)$. There exists a linear and isometric extension of I to $\mathcal{L}^2(\mu)$. This mapping is unique up to indistinguishability.*

Comment 3.1.9 The linearity stated in the theorem is to be understood as linearity up to indistinguishability. This kind of "up to indistinguishability" qualifier will be implicit in much of the following. ∘

Proof. First, we construct the mapping. Let $X \in \mathcal{L}^2(\mu)$. By Theorem 3.1.7, $\mathbf{b}\mathcal{E}$ is dense in $\mathcal{L}^2(B)$, so there exists a sequence H_n in $\mathbf{b}\mathcal{E}$ converging to X in $\mathcal{L}^2(\mu)$. In particular, H_n is a cauchy sequence in $\mathcal{L}^2(\mu)$. By the isometry property of I on $\mathbf{b}\mathcal{E}$, $I(H_n)$ is then a cauchy sequence in $\mathbf{c}\mathcal{M}_0^2$. Since $\mathbf{c}\mathcal{M}_0^2$ is complete by Lemma 2.4.9, there exists $Y \in \mathbf{c}\mathcal{M}_0^2$ such that $I(H_n)$ converges to Y, and Y is unique up to indistinguishability. We now have a candidate for $I(X)$. However, we need to make sure that this candidate does not depend on the sequence in $\mathbf{b}\mathcal{E}$ approximating X. Assume therefore that K_n is another such sequence, and let Z be a limit of $I(K_n)$. Then, we have

$$\begin{aligned}
& \|Y - Z\|_{\mathbf{c}\mathcal{M}_0^2} \\
\le\ & \limsup \|Y - I(H_n)\|_{\mathbf{c}\mathcal{M}_0^2} + \|I(H_n) - I(K_n)\|_{\mathbf{c}\mathcal{M}_0^2} + \|I(K_n) - Z\|_{\mathbf{c}\mathcal{M}_0^2} \\
=\ & \limsup \|I_n(H_n - K_n)\|_{\mathbf{c}\mathcal{M}_0^2} \\
=\ & \limsup \|H_n - K_n\|_{\mathbf{c}\mathcal{M}_0^2},
\end{aligned}$$

and since $\limsup \|H_n - K_n\|_\mu \le \limsup \|H_n - X\|_\mu + \|X - K_n\|_\mu = 0$, we conclude that for each $X \in \mathcal{L}^2(\mu)$, there is an element Y of $\mathbf{c}\mathcal{M}_0^2$, unique up to indistinguishability, such that for any sequence (H_n) in $\mathbf{b}\mathcal{E}$ converging to X in $\mathcal{L}^2(\mu)$, it holds that $\lim I(H_n) = Y$. We define $I(X)$ as some element in the equivalence class of Y.

Having constructed the mapping I, we need to check the properties of the mapping given in the theorem. To prove the isometry property, let $X \in \mathcal{L}^2(\mu)$ and let $(H_n) \subseteq \mathbf{b}\mathcal{E}$ converge to X in $\mathcal{L}^2(\mu)$. Then $I(X) = \lim I(H_n)$ in $\mathbf{c}\mathcal{M}_0^2$, and by the isometry property on $\mathbf{b}\mathcal{E}$,

$$\|I(X)\|_{\mathbf{c}\mathcal{M}_0^2} = \lim \|I(H_n)\|_{\mathbf{c}\mathcal{M}_0^2} = \lim \|H_n\|_\mu = \lim \|X\|_\mu.$$

Next, we show linearity. We wish to prove that for $X, Y \in \mathcal{L}^2(B)$ and $\lambda, \mu \in \mathbb{R}$, it holds that $I(\lambda X + \mu Y) = \lambda I(X) + \mu I(Y)$ up to indistinguishability. We know that the claim holds for elements of $\mathbf{b}\mathcal{E}$. Letting (H_n) and (K_n) be sequences in $\mathbf{b}\mathcal{E}$ approximating X and Y, respectively, we then find

$$\begin{aligned} I(\lambda X + \mu Y) &= I(\lambda \lim H_n + \mu \lim K_n) \\ &= \lim I(\lambda H_n + \mu K_n) \\ &= \lim \lambda I(H_n) + \mu I(K_n) \\ &= \lambda I(X) + \mu I(Y), \end{aligned}$$

up to indistinguishability, where the limits are in $\mathcal{L}^2(\mu)$ and $\mathbf{c}\mathcal{M}_0^2$.

This proves the statements about I. It remains to prove uniqueness of I. Assume therefore that we have two mappings, I and J, with the properties stated in the theorem. Obviously, $I(H) = J(H)$ for $H \in \mathbf{b}\mathcal{E}$. Let $X \in \mathcal{L}^2(\mu)$ and let (H_n) be a sequence in $\mathbf{b}\mathcal{E}$ approximating H. Then,

$$\begin{aligned} \|I(X) - J(X)\|_{\mathbf{c}\mathcal{M}_0^2} &= \|\lim I(H_n) - \lim J(H_n)\|_{\mathbf{c}\mathcal{M}_0^2} \\ &= \lim \|I(H_n) - J(H_n)\|_{\mathbf{c}\mathcal{M}_0^2} \\ &= 0, \end{aligned}$$

showing that $I(X)$ and $J(X)$ are indistinguishable. □

Comment 3.1.10 The proof of the existence of the extension in Theorem 3.1.8 was made without reference to standard theorems. Another, less direct, method of proof, would be to define the integral modulo equivalence classes making the pseudometric spaces involved into metric spaces. In this case, results for continuous extensions from the theory of metric spaces can be applied directly to obtain the extension, see Theorem 8.16 of Carothers (2000). ◦

We are now ready to begin the development of the stochastic integral proper.

3.2 Integration with respect to Brownian motion

In this section, we define the stochastic integral with respect to Brownian motion. The construction takes place in three phases. Our results from Section 2.5 and Section 3.1 have paved much of the way for our work, so we will be able to concentrate on the important parts of the process, not having to deal with technical details. The three phases are:

1. Definition of the integral on $\mathbf{b}\mathcal{E}$ by a Riemann-sum definition.
2. Definition of the integral on the space $\mathcal{L}^2(W)$ by a density argument.
3. Definition of the integral on $\mathfrak{L}^2(W)$ by a localisation argument.

Let W be a one-dimensional $\mathcal{F}_t$ Brownian motion. We begin by defining and investigating the stochastic integral on $\mathbf{b}\mathcal{E}$. From Theorem 3.1.3 and Lemma 3.1.6, we know that the mapping I_W from $\mathbf{b}\mathcal{E}$ to the space of stochastic processes given by, for $H = \sum_{k=1}^{\infty} Z_k(\sigma_k, \tau_k]$,

$$I_W(H) = \sum_{k=1}^{n} Z_k(W^{\tau_k} - W^{\sigma_k})$$

is well-defined and linear. We will now show that I_W maps into the space $\mathbf{c}\mathcal{M}_0^2$ of continuous square-integrable martingales zero at zero.

Lemma 3.2.1. *W_t and $W_t^2 - t$ are $\mathcal{F}_t$ martingales.*

Comment 3.2.2 It is well known that if W_t and $W(t)^2 - t$ are martingales with respect to the filtration generated by the Brownian motion W itself. The important statement of the lemma is that in our case, where we have an $\mathcal{F}_t$ Brownian motion, the martingales can be taken with respect to $\mathcal{F}_t$. $\circ$

Proof. We first show that W is an $\mathcal{F}_t$ martingale. Let $s \leq t$ be given, it then holds that $W_t - W_s = W_{s+t-s} - W_s$ is independent of $\mathcal{F}_s$ and normally distributed. In particular, $E(W_t - W_s | \mathcal{F}_s) = E(W_t - W_s) = 0$, showing the martingale property.

To show that $W_t^2 - t$ is a $\mathcal{F}_t$ martingale, we note that for $s \leq t$, we have

$$W_t^2 - W_s^2 = (W_t - W_s)^2 + 2(W_t - W_s)W_s,$$

and therefore, since $W_t - W_s$ is independent of $\mathcal{F}_s$,

$$\begin{aligned} E(W_t^2 - W_s^2|\mathcal{F}_s) &= E((W_t - W_s)^2|\mathcal{F}_s) + 2W_s E(W_t - W_s|\mathcal{F}_s) \\ &= t - s, \end{aligned}$$

showing $E(W_t^2 - t|\mathcal{F}_s) = W_s^2 - s$. □

Theorem 3.2.3. *For any $H \in \mathbf{b}\mathcal{E}$, $I_W(H) \in \mathbf{c}\mathcal{M}_0^2$ and $EI_W(H)_\infty^2 = E\int_0^\infty H_t^2 \, dt$.*

Proof. The process $I_W(H)$ is clearly continuous, since W is continuous.

Step 1: The martingale property. We use Lemma 2.4.13. Let $H \in \mathbf{b}\mathcal{E}$ with $H = \sum_{k=1}^n Z_k[\sigma_k, \tau_k)$, we then have $\lim_{t\to\infty} I_W(H)_t = \sum_{k=1}^n Z_k(W_{\sigma_k} - W_{\tau_k})$, so $I_W(H)$ has a well-defined almost sure limit. Now let τ be any stopping time. Since τ_k and σ_k are bounded stopping times, optional sampling and Lemma 2.4.11 yields

$$\begin{aligned} E(H \cdot W)_\tau &= E\sum_{k=1}^n Z_k(W_\tau^{\tau_k} - W_\tau^{\sigma_k}) \\ &= \sum_{k=1}^n E(Z_k E(W_{\tau_k}^\tau - W_{\sigma_k}^\tau|\mathcal{F}_{\sigma_k})) \\ &= 0. \end{aligned}$$

By Lemma 2.4.13, then, $H \cdot W$ is a martingale.

Step 2: Square-integrability. To show that $I_W(H)$ is bounded in $\mathcal{L}^2$, it will suffice to show the moment equality for $I_W(H)_\infty$, since $I_W(H)^2$ is a submartingale and its second moments are therefore increasing. As before, let $H \in \mathbf{b}\mathcal{E}$ with $H = \sum_{k=1}^n Z_k[\sigma_k, \tau_k)$. By Lemma 3.1.2, we can assume that the stopping times are ordered such that $\tau_k \le \sigma_{k+1}$ for all $k < n$.

We find

$$\begin{aligned} & EI_W(H)_\infty^2 \\ =\ & E\left(\sum_{k=1}^n Z_k(W_{\tau_k} - W_{\sigma_k})\right)^2 \\ =\ & E\sum_{k=1}^n\sum_{i=1}^n Z_k Z_i (W_{\tau_k} - W_{\sigma_k})(W_{\tau_i} - W_{\sigma_i}) \\ =\ & \sum_{k=1}^n EZ_k^2(W_{\tau_k} - W_{\sigma_k})^2 + \sum_{k\neq i}^n EZ_kZ_i(W_{\tau_k} - W_{\sigma_k})(W_{\tau_i} - W_{\sigma_i}). \end{aligned}$$

Here we have, applying optional sampling for bounded stopping times and using that $W_t^2 - t$ by Lemma 3.2.1 is a $\mathcal{F}_t$-martingale,

$$\begin{aligned}
& EZ_k^2(W_{\tau_k} - W_{\sigma_k})^2 \\
= \;& EZ_k^2 E((W_{\tau_k} - W_{\sigma_k})^2|\mathcal{F}_{\sigma_k}) \\
= \;& EZ_k^2 E(W_{\tau_k}^2 - \tau_k + W_{\sigma_k}^2 + \sigma_k - 2W_{\tau_k}W_{\sigma_k}|\mathcal{F}_{\sigma_k}) + EZ_k^2 E(\tau_k - \sigma_k|\mathcal{F}_{\sigma_k}) \\
= \;& EZ_k^2 E(W_{\tau_k}^2 - \tau_k - (W_{\sigma_k}^2 - \sigma_k)|\mathcal{F}_{\sigma_k}) + EZ_k^2(\tau_k - \sigma_k) \\
= \;& EZ_k^2(\tau_k - \sigma_k).
\end{aligned}$$

Furthermore, for $i < k$ we find that, again by optional sampling, using the ordering of the stopping times,

$$\begin{aligned}
& EZ_kZ_i(W_{\tau_k} - W_{\sigma_k})(W_{\tau_i} - W_{\sigma_i}) \\
= \;& EZ_i(W_{\tau_i} - W_{\sigma_i})Z_kE((W_{\tau_k} - W_{\sigma_k})|\mathcal{F}_{\sigma_k}) \\
= \;& 0.
\end{aligned}$$

Finally, we conclude

$$EI_W(H)_\infty^2 = \sum_{k=1}^n EZ_k^2(\tau_k - \sigma_k) = E\int_0^\infty H_u^2 \, du,$$

as desired. □

We have now proven that I_W maps $\mathbf{b}\mathcal{E}$ into $\mathbf{c}\mathcal{M}_0^2$. If we can embed $\mathbf{b}\mathcal{E}$ into a $\mathcal{L}^2$-space such that I_W becomes an isometry, we can use Theorem 3.1.8 to extend I_W to this space.

Definition 3.2.4. *By $\mathcal{L}^2(W)$ we denote the set of adapted and measurable processes X such that $E\int_0^\infty X_t^2 \, dt < \infty$. $\mathcal{L}^2(W)$ is then equal to the $\mathcal{L}^2$ space of square integrable functions on $[0,\infty) \times \Omega$ with the progressive σ-algebra Σ_π under the measure $\lambda \otimes P$. We denote the $\mathcal{L}^2$ seminorm on $\mathcal{L}^2(W)$ by $\|\cdot\|_W$.*

Theorem 3.2.5. *There exists a linear isometric mapping $I_W : \mathcal{L}^2(W) \to \mathbf{c}\mathcal{M}_0^2$ such that for $H \in \mathbf{b}\mathcal{E}$, with $H = \sum_{k=1}^n Z_k[\sigma_k, \tau_k)$, $I_W(H) = \sum_{k=1}^n Z_k(W^{\tau_k} - W^{\sigma_k})$. This mapping is unique up to indistinguishability.*

Proof. Considering $\mathbf{b}\mathcal{E}$ as a subspace of $\mathcal{L}^2(W)$, we have by Theorem 3.2.3 for $H \in \mathbf{b}\mathcal{E}$ that $\|I_W(H)\|_{\mathbf{c}\mathcal{M}_0^2}^2 = \|I_W(H)_\infty\|_2^2 = E\int_0^\infty H(t)^2 \, dt = \|H\|_W^2$, so $I_W : \mathbf{b}\mathcal{E} \to \mathbf{c}\mathcal{M}_0^2$ is isometric. Since it is also linear, the conditions of Theorem 3.1.8 are satisfied, and the conclusion follows. □

Comment 3.2.6 The isometry property of I_W is known as Itô's isometry. ∘

We still need to make a final extension of the stochastic integral with respect to Brownian motion. This extension is done by the technique of localisation. It is during this step that the results of Section 2.5 will prove worth their while.

The basic idea of the localisation is that if a process is in $\mathcal{L}^2(W)$ when we cut it off and put it to zero from a certain point onwards, then we can define the integral up to this time. If the cutoff times tend to infinity, we can paste the integrals together and obtain a reasonable stochastic integral.

The cutoff procedure will, in order to obtain an appropriate level of generality, be done at a stopping time. We will need two kinds of localization, described in Section 2.5, namely stopping and zero-stopping. We review the localisation concepts here for completeness. We say that a process M is locally, or $\mathfrak{L}$-locally, in $\mathbf{c}\mathcal{M}_0^2$ if there is a sequence of stopping times τ_n increasing to infinity such that M^{τ_n}, defined by $M_t^{\tau_n} = M_{\tau_n \wedge t}$, is in $\mathbf{c}\mathcal{M}_0^2$. We say that a process X is zero-locally, or $\mathfrak{L}_0$-locally, in $\mathcal{L}^2(W)$ if there is a sequence of stopping times τ_n increasing to infinity such that $X[0,\tau_n]$, defined by $X[0,\tau_n]_t = X_t 1_{[0,\tau]}(t)$, is in $\mathcal{L}^2(W)$.

We denote the processes $\mathfrak{L}$-locally in $\mathbf{c}\mathcal{M}_0^2$ by $\mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$, and we denote the processes $\mathfrak{L}_0$-locally in $\mathcal{L}^2(W)$ by $\mathfrak{L}^2(W)$. We call $\mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$ the space of continuous local martingales. We are going to use Theorem 2.5.11 to prove that there is a unique extension of I_W from $\mathcal{L}^2(W) \to \mathbf{c}\mathcal{M}_0^2$ to $\mathfrak{L}^2(W) \to \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$ such that $I_W(X)^\tau = I_W(X[0,\tau])$ for any stopping time τ. We will therefore need to check the hypotheses of this theorem.

Lemma 3.2.7. *The spaces $\mathcal{L}^2(W)$ and $\mathfrak{L}^2(W)$ are linear and $\mathfrak{L}_0$-stable. The spaces $\mathbf{c}\mathcal{M}_0^2$ and $\mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$ are linear and $\mathfrak{L}$-stable.*

Proof. By Lemma 2.5.18, $\mathcal{L}^2(W)$ is $\mathfrak{L}_0$-stable, and it is clearly linear. By Lemma 2.5.9, $\mathfrak{L}^2(W)$ is linear and $\mathfrak{L}_0$-stable. By Lemma 2.4.11, martingales are $\mathfrak{L}$-stable. Then $\mathbf{c}\mathcal{M}_0^2$ is also $\mathfrak{L}$-stable. Since $\mathbf{c}\mathcal{M}_0^2$ is also linear, Lemma 2.5.9 yields that $\mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$ is linear. □

Theorem 3.2.8. *Let $X \in \mathcal{L}^2(W)$ and let τ be a stopping time. Then it holds that $I(X)^\tau = I(X[0,\tau])$.*

Proof. We first show that the result holds for $\mathbf{b}\mathcal{E}$, and then extend by a density argument.

Step 1: The elementary case. Let $H \in \mathbf{b}\mathcal{E}$ with $H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k]$. We have that

$$H[0,\tau](t) = \sum_{k=1}^n Z_k 1_{(\sigma_k,\tau_k]}(t) 1_{[0,\tau]}(t) = \sum_{k=1}^n Z_k 1_{(\sigma_k \le \tau)} 1_{(\sigma_k\wedge\tau, \tau_k\wedge\tau]}(t),$$

and by Lemma 2.3.8, $Z_k 1_{(\sigma_k \le \tau)} \in \mathcal{F}_{\sigma_k \wedge \tau}$. Therefore, $H[0,\tau] \in \mathbf{b}\mathcal{E}$, and we have

$$\begin{aligned} I(H[0,\tau])_t &= \sum_{k=1}^n Z_k 1_{(\sigma_k \le \tau)} (W_{\tau_k\wedge\tau\wedge t} - W_{\sigma_k\wedge\tau\wedge t}) \\ &= \sum_{k=1}^n Z_k (W_{\tau_k\wedge\tau\wedge t} - W_{\sigma_k\wedge\tau\wedge t}) \\ &= I(H)^\tau_t. \end{aligned}$$

Step 2: The general case. Now let $X \in \mathcal{L}^2(W)$ be arbitrary and let X_n be a sequence in $\mathbf{b}\mathcal{E}$ converging to X. We trivially have $\|X_n[0,\tau] - X[0,\tau]\|_W \le \|X_n - X\|_W$, so $X_n[0,\tau]$ converges to $X[0,\tau]$ in $\mathcal{L}^2(W)$. Since $I(X_n[0,\tau]) = I(X_n)^\tau$, the isometry property of the integral yields

$$\begin{aligned} & \|I(X[0,\tau]) - I(X)^\tau\|_{\mathbf{c}\mathcal{M}^2_0} \\ \le\ & \|I(X[0,\tau]) - I(X_n[0,\tau])\|_{\mathbf{c}\mathcal{M}^2_0} + \|I(X_n)^\tau - I(X)^\tau\|_{\mathbf{c}\mathcal{M}^2_0} \\ \le\ & \|I(X[0,\tau]) - I(X_n[0,\tau])\|_{\mathbf{c}\mathcal{M}^2_0} + \|I(X_n) - I(X)\|_{\mathbf{c}\mathcal{M}^2_0} \\ =\ & \|X[0,\tau] - X_n[0,\tau]\|_W + \|X_n - X\|_W \\ \le\ & 2\|X_n - X\|_W. \end{aligned}$$

Thus, $I(X^\tau) = I(X)^\tau$ up to indistinguishability, as desired. □

With Theorem 3.2.8 in hand, we are ready to extend the integral.

Theorem 3.2.9. *There exists an extension of* $I_W : \mathcal{L}^2(W) \to \mathbf{c}\mathcal{M}^2_0$ *to a mapping from* $\mathfrak{L}^2(W)$ *to* $\mathbf{c}\mathcal{M}^{\mathfrak{L}}_0$. *The extension is uniquely determined by the criterion that* $(X \cdot W)^\tau = X[0,\tau] \cdot W$ *for any* $X \in \mathfrak{L}^2(W)$. *The extension is linear.*

Proof. We apply Theorem 2.5.11 with the localisation concepts $\mathfrak{L}_0$ of zero-stopping and $\mathfrak{L}$ of stopping. From Lemma 3.2.7, we know that $\mathcal{L}^2(W)$ is $\mathfrak{L}_0$-stable and $\mathbf{c}\mathcal{M}^2_0$ is $\mathfrak{L}$-stable. And by Theorem 3.2.8, the localisation hypothesis of Theorem 2.5.11 is satisfied. Theorem 2.5.11 therefore immediately yields the desired extension, and by Corollary 2.5.12, the extension is linear. □

Theorem 3.2.9 yields the "pasting" definition of the integral of processes in $\mathfrak{L}^2(W)$ discussed earlier. For general $X \in \mathfrak{L}^2(W)$ with localising sequence τ_n, the property $I_W(X)^{\tau_n} = I_W(X[0,\tau_n])$ relates the values of the integral for general X to the values of the integral on $\mathcal{L}^2(W)$.

We have now defined the stochastic integral with respect to Brownian motion for as many processes as we will need. Our next goal is to extend the integral to other integrators than Brownian motion. Before this, we summarize our results: Our stochastic integral I_W with respect to Brownian motion is a linear mapping from $\mathfrak{L}^2(W)$ to $\mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$. It has the following properties:

1. For $X \in \mathcal{L}^2(W)$, $X \cdot W$ is a square-integrable martingale.
2. For $H \in \mathbf{b}\mathcal{E}$ with $H = \sum_{k=1}^n Z_k(\sigma_k, \tau_k]$, $H \cdot W = \sum_{k=1}^n Z_k(W^{\tau_k} - W^{\sigma_k})$.
3. For any $X \in \mathfrak{L}^2(W)$ and any stopping time τ, $(X \cdot W)^\tau = X[0,\tau] \cdot W$.

3.3 Integration with respect to integrators in $\mathcal{M}_W^{\mathfrak{L}}$

We now use the results from Section 3.2 on integration with respect to Brownian motion to define the stochastic integral for larger classes of integrators and integrands. We begin with a lemma to set the mood. We still denote by W a one-dimensional $\mathcal{F}_t$ Brownian motion.

Lemma 3.3.1. *Let $H \in \mathbf{b}\mathcal{E}$ with $H = \sum_{k=1}^\infty Z_k(\sigma_k, \tau_k]$ and let $Y \in \mathfrak{L}^2(W)$. Then $HY \in \mathfrak{L}^2(W)$ and $\sum_{k=1}^n Z_k((Y \cdot W)^{\tau_k} - (Y \cdot W)^{\sigma_k}) = HY \cdot W$.*

Proof. Since H is bounded and Σ_π-measurable, it is clear that $HY \in \mathfrak{L}^2(W)$. To prove the other statement, note that by linearity, it will suffice to consider H of the form $H = Z(\sigma, \tau]$.

Step 1: The case $Y \in \mathbf{b}\mathcal{E}$. Assume $Y = \sum_{k=1}^n X_k(\sigma_k, \tau_k]$, we then find

$$\begin{aligned} Z(\sigma,\tau]Y &= \sum_{k=1}^n Z(\sigma,\tau]X_k(\sigma_k,\tau_k] \\ &= \sum_{k=1}^n ZX_k 1_{(\sigma_k \le \tau)}(\sigma_k \wedge \tau \vee \sigma, \tau_k \wedge \tau \vee \sigma], \end{aligned}$$

by Lemma 3.1.4, where $X_k 1_{(\sigma_k \le \tau)}$ is $\mathcal{F}_{\sigma_k \wedge \tau}$ measurable by Lemma 2.3.8, so $ZX_k 1_{(\sigma_k \le \tau)}$ is $\mathcal{F}_{\sigma_k \wedge \tau \vee \sigma}$ measurable. This shows $Z(\sigma,\tau]Y \in \mathbf{b}\mathcal{E}$, and we may then conclude

$$\begin{aligned}
& Z((Y \cdot W)^\tau - (Y \cdot W)^\sigma) \\
= \;& Z\left(\sum_{k=1}^n X_k (W^{\tau_k \wedge \tau} - W^{\sigma_k \wedge \tau}) - \sum_{k=1}^n X_k (W^{\tau_k \wedge \sigma} - W^{\sigma_k \wedge \sigma})\right) \\
= \;& \sum_{k=1}^n Z X_k \left(W^{\tau_k \wedge \tau} - W^{\sigma_k \wedge \tau} - W^{\tau_k \wedge \sigma} + W^{\sigma_k \wedge \sigma}\right) \\
= \;& \sum_{k=1}^n Z X_k \left(W^{\tau_k \wedge \tau \vee \sigma} - W^{\sigma_k \wedge \tau \vee \sigma}\right) \\
= \;& \sum_{k=1}^n Z X_k 1_{(\sigma_k \le \tau)} \left(W^{\tau_k \wedge \tau \vee \sigma} - W^{\sigma_k \wedge \tau \vee \sigma}\right) \\
= \;& HY \cdot W.
\end{aligned}$$

Step 2: The case $Y \in \mathcal{L}^2(W)$. Assuming $Y \in \mathcal{L}^2(W)$, let H_n be a sequence in $\mathbf{b}\mathcal{E}$ converging towards Y. Since H is bounded, $\lim \|HY - HH_n\|_W = 0$. From the continuity of the integral, $H_n \cdot W$ converges to $Y \cdot W$. Therefore, $(H_n \cdot W)^\tau - (H_n \cdot W)^\sigma$ converges to $(Y \cdot W)^\tau - (Y \cdot W)^\sigma$. Because Z is bounded, we obtain

$$\begin{aligned}
Z((Y \cdot W)^\tau - (Y \cdot W)^\sigma) &= \lim_n Z((H_n \cdot W)^\tau - (H_n \cdot W)^\sigma) \\
&= \lim_n HH_n \cdot W \\
&= \lim_n HY \cdot W,
\end{aligned}$$

where the limits are in $\mathcal{L}^2(W)$.

Step 3: The case $Y \in \mathfrak{L}^2(W)$. Let τ_n be a localising sequence for Y. Then

$$\begin{aligned}
(Z(Y \cdot W)^\tau - (Y \cdot W)^\sigma)^\tau &= Z((Y[0,\tau_n] \cdot W)^\tau - (Y[0,\tau_n] \cdot W)^\sigma) \\
&= HY[0,\tau_n] \cdot W \\
&= (HY \cdot W)^{\tau_n},
\end{aligned}$$

and letting τ_n tend to infinity, we obtain the result. $\square$

What Lemma 3.3.1 shows is that if we consider $Y \cdot W$ as an integrator, then the integral of an element $H \in \mathbf{b}\mathcal{E}$ based on the Riemann-sum definition of Section 3.1 satisfies

$H \cdot (Y \cdot W) = HY \cdot W$. We call this property the associativity of the integral. We will use this as a criterion to generalize the stochastic integrals to integrators of the form $Y \cdot W$ where $Y \in \mathfrak{L}^2(W)$. The most obvious thing to do would be to define, given $M = Y \cdot W$ with $Y \in \mathfrak{L}^2(W)$, the integral of a process X with $XY \in \mathfrak{L}^2(W)$ with respect to M by $X \cdot M = XY \cdot W$. This is what we will do, but with a small twist. Until now, in this section and the preceeding, we have only considered a one-dimensional $\mathcal{F}_t$ Brownian motion. In our financial applications, we are going to need to work with multidimensional Brownian motion, and we are going to need integrators that depend on several coordinates of such a multidimensional Brownian motion at once. Therefore, we will now consider an n-dimensional $\mathcal{F}_t$ Brownian motion $W = (W^1, \ldots, W^n)$, as described in Definition 2.1.5. Our goal will be to define the stochastic integral with respect to integrators of the form

$$M_t = \sum_{k=1}^{n} \int_0^t Y_s^k \, \mathrm{d}W_s^k$$

by associativity, as described above. We maintain the definitions of Section 3.2 by letting $\mathcal{L}^2(W)$ denote the Σ_π measurable elements of $Y \in \mathcal{L}^2(\lambda \otimes P)$ and letting $\mathfrak{L}^2(W)$ denote the processes $\mathfrak{L}_0$-locally in $\mathcal{L}^2(W)$. The next lemma shows that each of the coordinate processes fall under the category we worked with in the last section.

Lemma 3.3.2. *If $W = (W^1, \ldots, W^n)$ is an n-dimensional $\mathcal{F}_t$ Brownian motion, then for each $k \leq n$, W^k is a one-dimensional $\mathcal{F}_t$ Brownian motion.*

Proof. Let $t \geq 0$. We know that $s \mapsto W_{t+s} - W_t$ is a n-dimensional Brownian motion which is independent of $\mathcal{F}_t$. Therefore $s \mapsto W_{t+s}^k - W_t^k$ is a one-dimensional Brownian motion independent of $\mathcal{F}_t$. $\square$

In the following, whenever we write $Y \in \mathfrak{L}^2(W)^n$, we mean that Y is an n-dimensional proces, $Y = (Y^1, \ldots, Y^n)$, such that $Y^k \in \mathfrak{L}^2(W)$ for each k. This notation is of course also extended to other spaces than $\mathfrak{L}^2(W)$.

Definition 3.3.3. *Let $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ denote the class of processes M such that there exists $Y \in \mathfrak{L}^2(W)^n$ with $M_t = \sum_{k=1}^n \int_0^t Y_s^k \, \mathrm{d}W_s^k$. In this case, we also write $M = Y \cdot W$.*

In order to define the integral with respect to processes in $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ by associativity, we need to check that the integrand processes Y^k in the representation of elements of $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ are unique. This is our next order of business.

Lemma 3.3.4. *For any $i, j \leq n$ with $i \neq j$, $W^i W^j$ is a martingale.*

Proof. Since the conditional distribution of $(W_t^i - W_s^i, W_t^j - W_s^j)$ given $\mathcal{F}_s$ is two-dimensional normal with mean zero and zero covariation, we have

$$\begin{aligned} & E(W_t^i W_t^j + W_s^i W_s^j | \mathcal{F}_s) \\ = \;& E((W_t^i - W_s^i)(W_t^j - W_s^j)|\mathcal{F}_s) + E(W_t^i W_s^j|\mathcal{F}_s) + E(W_s^i W_t^j|\mathcal{F}_s) \\ = \;& W_s^j E(W_t^i|\mathcal{F}_s) + W_s^i E(W_t^j|\mathcal{F}_s) \\ = \;& 2W_s^j W_s^i, \end{aligned}$$

so $E(W_t^i W_t^j|\mathcal{F}_s) = W_s^i W_s^j$, and $W^i W^j$ is a martingale. □

Lemma 3.3.5. *Let $i, j \leq n$ with $i \neq j$. For any $Y^i, Y^j \in \mathcal{L}^2(W)$, $(Y^i \cdot W^i)(Y^j \cdot W^j)$ is a uniformly integrable martingale.*

Proof. We use Lemma 2.4.13. Clearly, $(Y^i \cdot W^i)(Y^j \cdot W^j)$ has an almost sure limit. It will therefore suffice to show that for any stopping time τ,

$$E \int_0^\tau Y_s^i \, \mathrm{d}W_s^i \int_0^\tau Y_s^j \, \mathrm{d}W_s^j = 0.$$

Step 1: The elementary case. First consider the case where $Y^i, Y^j \in \mathbf{b}\mathcal{E}$ with

$$Y^i = \sum_{k=1}^n Z_k^i (\sigma_k, \tau_k] \text{ and } Y^j = \sum_{k=1}^n Z_k^j (\sigma_k, \tau_k].$$

We then find

$$\begin{aligned} & E \int_0^\tau Y_s^i \, \mathrm{d}W_s^i \int_0^\tau Y_s^j \, \mathrm{d}W_s^j \\ = \;& E \left(\sum_{k=1}^n Z_k^i ((W^i)_\tau^{\tau_k} - (W^i)_\tau^{\sigma_k}) \right) \left(\sum_{k=1}^n Z_k^j ((W^j)_\tau^{\tau_k} - (W^j)_\tau^{\sigma_k}) \right) \\ = \;& \sum_{k=1}^n \sum_{m=1}^n E Z_k^i Z_m^j ((W^i)_\tau^{\tau_k} - (W^i)_\tau^{\sigma_k})((W^j)_\tau^{\tau_m} - (W^j)_\tau^{\sigma_m}) \\ = \;& \sum_{k=1}^n \sum_{m=1}^n E Z_k^i Z_m^j ((W^i)_\tau^{\tau_k} - (W^i)_\tau^{\sigma_k})((W^j)_\tau^{\tau_m} - (W^j)_\tau^{\sigma_m}). \end{aligned}$$

Now, by Lemma 3.3.4, $W^i W^j$ is a martingale, and therefore $(W^i W^j)^\tau$ is also a martingale. Using the optional stopping theorem, we therefore find that the above is zero. We may then conclude $E \int_0^\tau Y_s^i \, \mathrm{d}W_s^i \int_0^\tau Y_s^j \, \mathrm{d}W_s^j = 0$.

Step 1: The general case. In the case of general $Y^i, Y^j \in \mathcal{L}^2(W)$, there exists sequences (H_n) and (K_n) in $\mathbf{b}\mathcal{E}$ converging to Y^i and Y^j, respectively. By Lemma 3.1.2, we may assume that H_n and K_n are based on the same stopping times, and therefore, by what we have just shown,

$$E\int_0^\tau H_n(s)\,\mathrm{d}W_s^i \int_0^\tau K_n(s)\,\mathrm{d}W_s^j = 0.$$

Now, since $\int_0^\tau H_n(s)\,\mathrm{d}W_s^i$ converges to $\int_0^\tau Y_s^i\,\mathrm{d}W_s^i$ and $\int_0^t K_n(s)\,\mathrm{d}W_s^j$ converges to $\int_0^\tau Y_s^j\,\mathrm{d}W_s^j$ in $\mathcal{L}^2(W)$ by the Itô isometry, Lemma B.2.1 yields that the product converges in $\mathcal{L}^1$ and therefore

$$E\int_0^\tau Y_s^i\,\mathrm{d}W_s^i \int_0^\tau Y_s^j\,\mathrm{d}W_s^j = 0 = \lim_n E\int_0^\tau H_n(s)\,\mathrm{d}W_s^i \int_0^\tau K_n(s)\,\mathrm{d}W_s^j = 0,$$

as desired. $\square$

Lemma 3.3.6. *If $Y \in \mathcal{L}^2(W)^n$, then $E(Y\cdot W)_\infty^2 = \sum_{k=1}^n \|Y^k\|_W^2$.*

Proof. By Lemma 3.3.5, the Itô isometry and optional sampling for uniformly integrable martingales,

$$\begin{aligned}
E\left(\sum_{k=1}^n \int_0^\infty Y_s^k\,\mathrm{d}W_s^k\right)^2 &= \sum_{k=1}^n\sum_{i=1}^n E\int_0^\infty Y_s^k\,\mathrm{d}W_s^k \int_0^\infty Y_s^i\,\mathrm{d}W_s^i \\
&= \sum_{k=1}^n E\left(\int_0^\infty Y_s^k\,\mathrm{d}W_s^k\right)^2 \\
&= \sum_{k=1}^n \|Y^k\|_W^2.
\end{aligned}$$

$\square$

Lemma 3.3.7. *Assume $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ with $M = \sum_{k=1}^n Y_s^k\,\mathrm{d}W_s^k$ and $M = \sum_{k=1}^n Z_s^k\,\mathrm{d}W_s^k$. Then Y^k and Z^k are equal $\lambda\otimes P$ almost surely for $k \le n$.*

Proof. First consider the case where $Y^k, Z^k \in \mathcal{L}^2(W)$ for $k \le n$. We then find, by Lemma 3.3.6,

$$\begin{aligned}
E\left(\sum_{k=1}^n \int_0^\infty Y_s^k\,\mathrm{d}W_s^k - \sum_{k=1}^n \int_0^\infty Z_s^k\,\mathrm{d}W_s^k\right)^2 &= E\left(\sum_{k=1}^n \int_0^\infty Y_s^k - Z_s^k\,\mathrm{d}W_s^k\right)^2 \\
&= \sum_{k=1}^n \|Y_s^k - Z_s^k\|_W^2.
\end{aligned}$$

This shows that Y^k and Z^k are equal except on a set of $\lambda \otimes P$ measure zero. Next, consider the general case. Let τ_n be a sequence of stopping times tending to infinity such that $(Y^k)^{\tau_n}, (Z^k)^{\tau_n} \in \mathcal{L}^2(W)$. We then know that $(Y^k)^{\tau_n}$ and $(Z^k)^{\tau_n}$ are equal $\lambda \otimes P$ almost surely, and therefore Y^k and Z^k are equal $\lambda \otimes P$ almost surely as well, as desired. □

We can now make the definition of the stochastic integral with respect to processes in $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$. As mentioned earlier, whenever we have $Y \in \mathfrak{L}^2(W)^n$, we will use the shorthand $(Y \cdot W)_t = \sum_{k=1}^n Y_s^k \,\mathrm{d}W_s^k$. Also, when X is a one-dimensional process, we will write $XY = (XY^1, \ldots, XY^n)$.

Definition 3.3.8. *Let $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ with $M = Y \cdot W$, $Y \in \mathfrak{L}^2(W)^n$. We define $\mathfrak{L}^2(M)$ as the space of Σ_π measurable processes X such that $XY \in \mathfrak{L}^2(W)^n$. For $X \in \mathfrak{L}^2(M)$, we define $I_M(X) = XY \cdot W$. I_M is then a mapping from $\mathfrak{L}^2(M)$ to $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$.*

Comment 3.3.9 Since the shorthands we use will be implicit in most of the following, let us take a moment to see how they figure in the above definition. We defined $\mathfrak{L}^2(M)$ as the space of processes X such that $XY \in \mathfrak{L}^2(W)^n$. Here, Y is a n-dimensional process such that $Y^k \in \mathfrak{L}^2(W)$. Thus, $XY = (XY^1, \ldots, XY^n)$ and the statement $XY \in \mathfrak{L}^2(W)^n$ means that $XY^k \in \mathfrak{L}^2(W)$ for all $k \leq n$.

We defined $I_M(X) = XY \cdot W$. Again XY is an n-dimensional process in $\mathfrak{L}^2(W)^n$, and W is a n-dimensional $\mathcal{F}_t$ Brownian motion. Thus, the definition really means $I_M(X)_t = \sum_{k=1}^n \int_0^t X_s Y_s^k \,\mathrm{d}W_s^k$. This is well-defined up to indistinguishability by Lemma 3.3.7, since if $M = Z \cdot W$ is another representation, Z^k and Y^k are equal $\lambda \otimes P$ almost surely, so XZ^k and ZY^k are equal $\lambda \otimes P$ almost surely, and therefore the stochastic integrals $XY^k \cdot W^k$ and $XZ^k \cdot W^k$ agree up to indistinguishability.

It is important to note that even if we are working with n-dimensional Brownian motions and n-dimensional integrand processes $Y \in \mathfrak{L}^2(W)^n$, the space $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ of integrators is still only a space of one-dimensional processes, as is the space of integrands $\mathfrak{L}^2(M)$. The multidimensionality appears because we want our integrands to be able to depend on all the coordinates of the Brownian motion, and this is done by summing integrals with respect to each coordinate. ◦

Definition 3.3.8 defines the stochastic integral for any $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ and $X \in \mathfrak{L}^2(M)$. We will spend the rest of the section investigating the properties of this integral. First, we will show linearity, that the integral can be written as a Riemann-sum for elementary processes, that it preserves stopping in an appropriate manner, and associativity. After this, we will work to understand under which conditions the integral is a square-integrable martingale. We will also develop some approximation results that will be useful in our work with the integral. In the following, let $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ with $M = Y \cdot W$, $Y \in \mathfrak{L}^2(W)^n$, unless anything else is implied.

It is clear that $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ and $\mathfrak{L}^2(M)$ are linear spaces, and by $I_M(X) = \sum_{k=1}^n I_{W^k}(XY^k)$, it is immediate that I_M is linear for any $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$.

Lemma 3.3.10 (Riemann representation). *If $H \in \mathbf{b}\mathcal{E}$ with $H = \sum_{k=1}^m Z_k(\sigma_k, \tau_k]$, $H \in \mathfrak{L}^2(M)$ and $H \cdot M = \sum_{k=1}^m Z_k(M^{\tau_k} - M^{\sigma_k})$.*

Proof. By Lemma 3.3.1, $HY^k \in \mathfrak{L}^2(W^k)$ for all $k \leq n$, so $H \in \mathfrak{L}^2(M)$, and

$$\begin{aligned}
(H \cdot M)_t &= \sum_{i=1}^n \int_0^t H_s Y_s^i \, \mathrm{d}W_s^i \\
&= \sum_{i=1}^n \sum_{k=1}^m Z_k((Y^i \cdot W^i)^{\tau_k} - (Y^i \cdot W^i)^{\sigma_k}) \\
&= \sum_{k=1}^m Z_k \sum_{i=1}^n (Y^i \cdot W^i)^{\tau_k} - (Y^i \cdot W^i)^{\sigma_k} \\
&= \sum_{k=1}^m Z_k(M^{\tau_k} - M^{\sigma_k}),
\end{aligned}$$

as desired. □

For $Y \in \mathfrak{L}^2(W)^n$, we write $Y[0,\tau] = (Y^1[0,\tau], \ldots, Y^n[0,\tau])$. From Theorem 3.2.9, it is clear that $(Y \cdot W)^\tau = Y[0,\tau] \cdot W$.

Lemma 3.3.11 (Localisation properties). *The spaces $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ and $\mathfrak{L}^2(M)$ are stable under stopping and zero-stopping, respectively, and $\mathfrak{L}^2(M) \subseteq \mathfrak{L}^2(M^\tau)$. It holds that $(X \cdot M)^\tau = X[0,\tau] \cdot M = X \cdot M^\tau$.*

Proof. $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ is $\mathfrak{L}$-stable since $(Y \cdot W)^\tau = Y[0,\tau] \cdot W$ and $Y[0,\tau] \in \mathfrak{L}^2(W)^n$ by Lemma 3.2.7. $\mathfrak{L}^2(M)$ is $\mathfrak{L}_0$-stable because, if $X \in \mathfrak{L}^2(M)$, $X[0,\tau]Y^k = (XY^k)[0,\tau]$, which is in $\mathfrak{L}^2(W)$ by Lemma 3.2.7. To show that $\mathfrak{L}^2(M) \subseteq \mathfrak{L}^2(M^\tau)$, consider $X \in \mathfrak{L}^2(M)$. Then $XY \in \mathfrak{L}^2(W)^n$, so $XY[0,\tau] \in \mathfrak{L}^2(W)^n$. Because $M^\tau = Y[0,\tau] \cdot W$, $X \in \mathfrak{L}^2(M^\tau)$

follows. We then obtain $(X \cdot M)^\tau = X[0, \tau] \cdot M$ from

$$\begin{aligned} (X \cdot M)^\tau &= (XY \cdot W)^\tau \\ &= (XY)[0, \tau] \cdot W \\ &= X[0, \tau]Y \cdot W \\ &= X[0, \tau] \cdot M. \end{aligned}$$

And the equality $(X \cdot M)^\tau = X \cdot M^\tau$ follows since

$$\begin{aligned} (X \cdot M)^\tau &= (XY \cdot W)^\tau \\ &= XY[0, \tau] \cdot W \\ &= X \cdot M^\tau. \end{aligned}$$

□

Lemma 3.3.12 (Associativity). *If $Z \in \mathfrak{L}^2(X \cdot M)$, then $ZX \in \mathfrak{L}^2(M)$ and it holds that $Z \cdot (X \cdot M) = ZX \cdot M$.*

Proof. We know that $X \cdot M = XY \cdot W$, so $Z \in \mathfrak{L}^2(X \cdot M)$ means $ZXY \in \mathfrak{L}^2(W)^n$, which is equivalent to $ZX \in \mathfrak{L}^2(M)$. We then obtain

$$\begin{aligned} Z \cdot (X \cdot M) &= Z \cdot (XY \cdot W) \\ &= ZXY \cdot W \\ &= ZX \cdot M \end{aligned}$$

as desired. □

The linearity, the localisation properties of Lemma 3.3.11 and the associativity of Lemma 3.3.12 are the basic rules for manipulating the integral. Our next goal is to identify classes of integrands such that the integral becomes a square-integrable martingale. For $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ with $M = Y \cdot W$, $Y \in \mathfrak{L}^2(W)^n$, let $\mathcal{L}^2(M)$ be the space of Σ_π measurable processes X such that $XY \in \mathcal{L}^2(W)^n$.

Lemma 3.3.13. *$\mathcal{L}^2(M)$ is $\mathfrak{L}_0$-stable, and $\mathfrak{L}_0(\mathcal{L}^2(M)) = \mathfrak{L}^2(M)$.*

Proof. We first show that $\mathcal{L}^2(M)$ is $\mathfrak{L}_0$-stable. Assume that $X \in \mathcal{L}^2(M)$, such that $XY \in \mathcal{L}^2(W)^n$, and let τ be a stopping time. Then $X[0, \tau]Y = (XY)[0, \tau] \in \mathcal{L}^2(W)$, so $X[0, \tau] \in \mathcal{L}^2(M)$. Therefore, $\mathcal{L}^2(M)$ is $\mathfrak{L}_0$-stable.

To prove the other statement of the lemma, assume $X \in \mathfrak{L}_0(\mathcal{L}^2(M))$ and let τ_n be a localising sequence, $X[0,\tau_n] \in \mathcal{L}^2(M)$. Since $X[0,\tau_n]$ is Σ_π measurable, so is X. By definition, $X[0,\tau_n]Y \in \mathcal{L}^2(W)^n$, so $(XY)[0,\tau_n] \in \mathcal{L}^2(W)^n$, and it follows that $XY \in \mathfrak{L}^2(W)^n$, showing $X \in \mathfrak{L}^2(M)$ and therefore $\mathfrak{L}_0(\mathcal{L}^2(M)) \subseteq \mathfrak{L}^2(M)$. Conversely, assume $X \in \mathfrak{L}^2(M)$. Then $XY \in \mathfrak{L}^2(W)^n$, so there is a localising sequence τ_n yielding $(XY)[0,\tau_n] \in \mathcal{L}^2(W)^n$. Then $X[0,\tau_n]Y \in \mathcal{L}^2(W)^n$ and $X[0,\tau_n] \in \mathcal{L}^2(M)$, showing $X \in \mathfrak{L}_0(\mathcal{L}^2(M))$. □

Lemma 3.3.14. *$\mathcal{L}^2(M)$ is equal to $\mathcal{L}^2([0,\infty)\times\Omega,\Sigma_\pi,\mu_M)$, where μ_M is the measure having density $\sum_{k=1}^n (Y^k)^2$ with respect to $\lambda\otimes P$. We endow $\mathcal{L}^2(M)$ with the $\mathcal{L}^2$-norm corresponding to μ_M and denote this norm by $\|\cdot\|_M$.*

Proof. Let $X \in \mathcal{L}^2(M)$. Then X is Σ_π measurable and $XY \in \mathcal{L}^2(W)^n$, so $X^2(Y^k)^2$ is integrable with respect to $\lambda \otimes P$. Thus $\mathcal{L}^2(M) \subseteq \mathcal{L}^2([0,\infty)\times\Omega,\Sigma_\pi,\mu_M)$. On the other hand, let $X \in \mathcal{L}^2([0,\infty)\times\Omega,\Sigma_\pi,\mu_M)$. Then X is obviously Σ_π measurable and

$$\|XY^k\|_W^2 = \int X^2(Y^k)^2 \,\mathrm{d}(\lambda\otimes P) \leq \int X^2 \sum_{k=1}^n (Y^k)^2 \,\mathrm{d}(\lambda\otimes P) = \int X^2 \,\mathrm{d}\mu_M < \infty,$$

so $X \in \mathcal{L}^2(M)$ and therefore $\mathcal{L}^2([0,\infty)\times\Omega,\Sigma_\pi,\mu_M) \subseteq \mathcal{L}^2(M)$. □

Comment 3.3.15 Note that since we only require $Y \in \mathfrak{L}^2(W)^n$ and not $Y \in \mathcal{L}^2(W)^n$, the measure μ may easily turn out to be unbounded. ◦

Lemma 3.3.16. *Let $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ and $X \in \mathfrak{L}^2(M)$. $X \cdot M$ is a square-integrable martingale if and only if $X \in \mathcal{L}^2(M)$.*

Proof. If $X \in \mathcal{L}^2(M)$, $XY \in \mathcal{L}^2(W)^n$ and therefore, by linearity of $\mathbf{c}\mathcal{M}_0^2$,

$$X \cdot M = XY \cdot W = \sum_{k=1}^n XY^k \cdot W^k \in \mathbf{c}\mathcal{M}_0^2.$$

To prove the other implication, assume that $X \cdot M \in \mathbf{c}\mathcal{M}_0^2$. Because $X \in \mathfrak{L}^2(M)$, we know that $XY \in \mathfrak{L}^2(W)^n$. Let τ_m be a localising sequence with $XY^k[0,\tau_m] \in \mathcal{L}^2(W)$ for $k \leq n$. Since $XY \cdot W = X \cdot M \in \mathbf{c}\mathcal{M}_0^2$, $(XY \cdot W)^{\tau_m} \in \mathbf{c}\mathcal{M}_0^2$ as well. The Itô isometry from Lemma 3.3.6 yields

$$\begin{aligned}\|(XY^k)[0,\tau_m]\|_W^2 &\leq \sum_{k=1}^n \|XY^k[0,\tau_m]\|_W^2 \\ &\leq \sum_{k=1}^n \|XY^k\|_W^2 \\ &= \|XY\cdot W\|_{\mathbf{c}\mathcal{M}_0^2}^2 \\ &= \|X\cdot M\|_{\mathbf{c}\mathcal{M}_0^2}^2,\end{aligned}$$

where the last expression is finite by assumption. Now, monotone convergence yields

$$\|XY^k\|_W = \limsup_m \|(XY^k)[0,\tau_m]\|_W \leq \|X\cdot X\|_{\mathbf{c}\mathcal{M}_0^2} < \infty,$$

so $XY^k \in \mathcal{L}^2(W)$ and thus $X \in \mathcal{L}^2(M)$. □

Lemma 3.3.17. *$I_M : \mathcal{L}^2(M) \to \mathbf{c}\mathcal{M}_0^2$ is isometric for any $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$.*

Proof. The conclusion is well-defined since $I_M(X) \in \mathbf{c}\mathcal{M}_0^2$ for $X \in \mathcal{L}^2(M)$ by Lemma 3.3.16. To prove the result, we merely note for $X \in \mathcal{L}^2(M)$ that by Lemma 3.3.6

$$\begin{aligned}\|X\cdot M\|_{\mathbf{c}\mathcal{M}_0^2}^2 &= \|XY\cdot W\|_{\mathbf{c}\mathcal{M}_0^2}^2 \\ &= \sum_{k=1}^n \|XY^k\|_W^2 \\ &= \sum_{k=1}^n \int X^2(Y^k)^2 \,\mathrm{d}(\lambda\otimes P) \\ &= \int X^2 \,\mathrm{d}\mu_M \\ &= \|X\|_M^2.\end{aligned}$$

□

Comment 3.3.18 The isometry property of I_M is, as the isometry property of I_W given in Theorem 3.2.5, also known as Itô's isometry. ∘

Lemma 3.3.16 solves the problem of determining when the stochastic integral is a square-integrable martingale, and Lemma 3.3.17 tells us that for general integrators, we still have the isometry property that we proved in the case of Brownian motion. We have yet one more property of the integral that we wish to investigate. We are interested in understanding when the integral process $X \cdot M$ can be approximated by integrals of the form $H_n \cdot M$ for $H_n \in \mathbf{b}\mathcal{E}$. To understand this, we define $\mathbf{c}\mathcal{M}^2_W$ as the subclass of processes in $\mathbf{c}\mathcal{M}^{\mathfrak{L}}_W$ of $M = Y \cdot W$ where $Y \in \mathcal{L}^2(W)^n$.

Lemma 3.3.19. *$\mathbf{c}\mathcal{M}^2_W$ is $\mathfrak{L}$-stable, and $\mathfrak{L}(\mathbf{c}\mathcal{M}^2_W) = \mathbf{c}\mathcal{M}^{\mathfrak{L}}_W$.*

Proof. That $\mathbf{c}\mathcal{M}^2_W$ is $\mathfrak{L}$-stable follows since $\mathcal{L}^2(W)$ is $\mathfrak{L}_0$-stable. To show the other statement of the lemma, assume $M \in \mathbf{c}\mathcal{M}^{\mathfrak{L}}_W$, $M = Y \cdot W$ with $Y \in \mathfrak{L}^2(W)^n$. Let τ_m be a localising sequence such that $Y[0, \tau_m] \in \mathcal{L}^2(W)^n$. Then $M^{\tau_m} = Y[0, \tau_m] \cdot W$, so M is locally in $\mathbf{c}\mathcal{M}^2_W$, showing $\mathbf{c}\mathcal{M}^{\mathfrak{L}}_W \subseteq \mathfrak{L}(\mathbf{c}\mathcal{M}^2_W)$. Conversely, assume $M \in \mathfrak{L}(\mathbf{c}\mathcal{M}^2_W)$ and let τ_n be a localising sequence. Then we have $M^{\tau_n} \in \mathbf{c}\mathcal{M}^2_W$, and there exists $Y_n \in \mathcal{L}^2(W)$ such that $M^{\tau_n} = Y_n \cdot W$. Clearly, then

$$\begin{aligned} Y_{n+1}[0, \tau_n] \cdot W &= (Y_{n+1} \cdot W)^{\tau_n} \\ &= (M^{\tau_{n+1}})^{\tau_n} \\ &= M^{\tau_n} \\ &= Y_n \cdot W. \end{aligned}$$

In detail, this means $\sum_{k=1}^n Y_{n+1}[0, \tau_n] \cdot W^k = \sum_{k=1}^n Y_n^k \cdot W^k$. By Lemma 3.3.7, $Y_{n+1}[0, \tau_n]$ and Y_n are then $\lambda \otimes P$ almost surely equal. Then the Pasting Lemma 2.5.10 yields processes Y^k such that $Y^k[0, \tau_n]$ and $Y_n^k[0, \tau_n]$ are $\lambda \otimes P$ almost surely equal for all n. In particular, $Y \in \mathfrak{L}^2(W)^n$, and clearly

$$M^{\tau_n} = Y_n[0, \tau_n] \cdot W = Y[0, \tau_n] \cdot W = (Y \cdot W)^{\tau_n},$$

so $M = Y \cdot W$ and $\mathfrak{L}(\mathbf{c}\mathcal{M}^2_W) \subseteq \mathbf{c}\mathcal{M}^{\mathfrak{L}}_W$. $\square$

Lemma 3.3.20. *If $M \in \mathbf{c}\mathcal{M}^2_W$, then $\mathbf{b}\mathcal{E}$ is dense in $\mathcal{L}^2(M)$. In particular, for any $X \in \mathcal{L}^2(M)$, $X \cdot M$ can be approximated by elements $H_n \cdot M$ with $H_n \in \mathbf{b}\mathcal{E}$.*

Proof. Since $M \in \mathbf{c}\mathcal{M}^2_W$, $M = Y \cdot W$ for some $Y \in \mathcal{L}^2(W)^n$. In particular, μ_M is a bounded measure, and by Theorem 3.1.7, $\mathbf{b}\mathcal{E}$ is dense in $\mathcal{L}^2(M)$. By Lemma 3.3.17, when $X \in \mathcal{L}^2(M)$ and $(H_n) \subseteq \mathbf{b}\mathcal{E}$ converges to X, we then obtain that $H_n \cdot M$ also converges to $X \cdot M$. $\square$

We are now done with our preliminary investigation of the stochastic integral for integrators in $\mathbf{c}\mathcal{M}^{\mathfrak{L}}_W$. Our next task will be to extend the integral to integrators with a component of finite variation. Before doing so, let is review our results so far. We have defined the stochastic integral $X \cdot M$ for integrators $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ and integrands $X \in \mathfrak{L}^2(M)$. We have proven that

- The integral is always a continuous local martingale.
- The interal is in $\mathbf{c}\mathcal{M}_0^2$ if and only if $X \in \mathcal{L}^2(M)$.
- The integral is linear in the integrand.
- I_M is isometric from $\mathcal{L}^2(M)$ to $\mathbf{c}\mathcal{M}_0^2$.
- $(X \cdot M)^\tau = X[0,\tau] \cdot M = X \cdot M^\tau$.
- $\mathbf{b}\mathcal{E} \subseteq \mathfrak{L}^2(M)$, and the integral takes its natural values on $\mathbf{b}\mathcal{E}$.
- If $M \in \mathbf{c}\mathcal{M}_W^2$, μ_M is bounded and $\mathbf{b}\mathcal{E}$ is dense in $\mathcal{L}^2(M)$.

Note that nowhere in the above properties is the Brownian motion W, which our whole construction rests upon, mentioned. This fact will enable us to manipulate the stochastic integral without having to keep in mind the underlying definition in terms of the Brownian motion.

3.4 Integration with respect to standard processes

We now come to the final extension of the stochastic integral, where we extend the integral to cover integrators of the form $A + M$, where $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ and A is adapted with continuous paths of finite variation. Since we already have defined the integral with respect to $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$, we only have to define the integral with respect to A and make sure that the integrals with respect to A and M fit properly together.

Because functions of finite variation correspond to measures, as can be seen from Appendix A.1, we can use ordinary measure theory to define the integral with respect to A in a pathwise manner. In Appendix A.1 we have listed some fundamental results regarding functions of finite variation. For any mapping $F : [0,\infty) \to \mathbb{R}$, we have the variation function,

$$V_F(t) = \sup \sum_{k=1}^{n} |F(t_k) - F(t_{k-1})|$$

where the supremum is over all partitions of $[0,t]$. V_F is finite for all t if and only if F is of finite variation. By Lemma A.1.1, if F is continuous, then so is V_F. By Theorem A.1.2, we can associate to F a measure ν_F such that $\nu_F(a,b] = F(b) - F(a)$ for $0 \leq a \leq b$, and as in Definition A.1.4, we can then define $\mathcal{L}^1(F) = \mathcal{L}^1(\nu_F)$ and consider the integral with respect to F by putting $\int_0^\infty x(s)\,\mathrm{d}F_s = \int_0^\infty x(s)\,\mathrm{d}\nu_F(s)$.

We now transfer these ideas to stochastic processes. In order to make sure that our stochastic integral with respect to processes of finite variation has sufficient measurability properties, we will need to make the appropriate measurability assumptions on our integrands. By **cFV**, we denote the class of adapted stochastic processes with continuous paths of finite variation. If $A \in \mathbf{cFV}$, we denote by $\mathcal{L}^1(A)$ the space of progressively measurable processes such that $X(\omega) \in \mathcal{L}^1(A(\omega))$ for all ω. By $\mathfrak{L}^1(A)$, we denote the space of processes locally in $\mathcal{L}^1(A)$.

We begin with a result which shows that we do not need to consider integrators which are only locally of finite variation, since this would add nothing new to the theory.

Lemma 3.4.1. **cFV** *is a linear space, stable under stopping, and* $\mathfrak{L}(\mathbf{cFV}) = \mathbf{cFV}$.

Proof. It is clear that **cFV** is a linear space. We first show that **cFV** is stable under stopping. Let $A \in \mathbf{cFV}$ be given, and let τ be any stopping time. Clearly, A^τ has continuous paths. By Lemma 2.3.3, A^τ is progressive, therefore adapted. Fix ω. Then, for any partition $(t_0, \ldots, t_n)$ of $[0,t]$,

$$\sum_{k=1}^{n} |A^\tau_{t_k}(\omega) - A^\tau_{t_{k-1}}(\omega)| = \sum_{k=1}^{n} |A_{t_k \wedge \tau}(\omega) - A_{t_{k-1} \wedge \tau}(\omega)| \leq V_{A(\omega)}(t),$$

so taking supremum over all partitions, we conclude $V_{A^\tau(\omega)}(t) \leq V_{A(\omega)}$, and therefore $A^\tau(\omega)$ has finite variation as well, thus $A^\tau \in \mathbf{cFV}$ and **cFV** is stable under stopping.

In particular, $\mathbf{cFV} \subseteq \mathfrak{L}(\mathbf{cFV})$, so to show equality it will suffice to show the other inclusion. To this end, let $A \in \mathfrak{L}(\mathbf{cFV})$, and let τ_n be a localising sequence. Fix ω, let $t > 0$ and let n be so large that $\tau_n(\omega) \geq t$. Then $V_{A(\omega)}(t) \leq V_{A(\omega)}(\tau_n) = V_{A^{\tau_n}(\omega)}(\tau_n)$, which is finite, and we conclude that $A(\omega)$ has finite variation, showing $A \in \mathbf{cFV}$. $\square$

Next, we define the stochastic integral with respect to integrators $A \in \mathbf{cFV}$ and integrands in $\mathfrak{L}^1(A)$. To show the adaptedness of the integral, we will need the concept of a Markov kernel and the associated results, see Chapter 20 of Hansen (2004a) for this.

Lemma 3.4.2. *If $X \in \mathfrak{L}^1(A)$, the integral $\int_0^t X_s(\omega)\,\mathrm{d}A_s(\omega)$ is well-defined as a path-wise integral. We write $(X \cdot A)_t(\omega) = \int_0^t X_s(\omega)\,\mathrm{d}A_s(\omega)$.*

Proof. Let τ_n be a localising sequence for X and let $\omega \in \Omega$. Let n be so large that $\tau_n(\omega) \geq t$, then

$$\int_0^t |X_s(\omega)|\,\mathrm{d}A_s(\omega) \leq \int_0^t |X[0,\tau_n]_s(\omega)|\,\mathrm{d}A_s(\omega) \leq \int_0^\infty |X[0,\tau_n]_s(\omega)|\,\mathrm{d}A_s(\omega),$$

which is finite since $X[0,\tau_n](\omega) \in \mathcal{L}^1(A(\omega))$. Thus, the integral $\int_0^t X_s(\omega)\,\mathrm{d}A_s(\omega)$ is well-defined for all $\omega \in \Omega$. □

Lemma 3.4.3. *Let $X \in \mathfrak{L}^1(A)$. The stochastic integral $X \cdot A$ is in* **cFV**.

Proof. First assume that $X \in \mathcal{L}^1(A)$. We need to show that $X \cdot A$ is adapted, continuous and of finite variation. By Lemma A.1.3, the measure corresponding to $A(\omega)$ has no atoms, and therefore $X \cdot A$ is continuous. By Lemma A.1.6, $X \cdot A$ has paths of finite variation.

To show that $X \cdot A$ is adapted, let $t > 0$ and let $\mu(t,\omega)$ be the measure on $[0,t]$ induced by $A(\omega)$ on $[0,t]$. We wish to show that $(\mu(t,\omega))_{\omega\in\Omega}$ is a $(\Omega,\mathcal{F}_t)$ Markov kernel on $([0,t],\mathcal{B}[0,t])$. We therefore need to show that $\omega \mapsto \mu(t,\omega)(B)$ is $\mathcal{F}_t$ measurable for any $B \in \mathcal{B}[0,t]$. The family of elements of $\mathcal{B}[0,t]$ for which this holds is a Dynkin class, and it will therefore suffice to show the claim for intervals in $[0,t]$. Let $0 \leq a \leq b \leq t$. We then find

$$\mu(t,\omega)[a,b] = A(b,\omega) - A(a,\omega),$$

and by the adaptedness of A, this is $\mathcal{F}_t$ measurable. Since X is progressively measurable, the extended Fubini Theorem then yields that

$$(X \cdot A)_t(\omega) = \int_0^t X_s(\omega)\,\mathrm{d}\mu(t,\omega)(s)$$

is $\mathcal{F}_t$ measurable as a function of ω, so $X \cdot A$ is adapted.

To extend this to the local case, assume $X \in \mathfrak{L}^1(A)$ and let τ_n be a localising sequence. Then, by the ordinary properties of the integral, $(X \cdot A)^{\tau_n} = (X[0,\tau_n] \cdot A)$. Therefore, $(X \cdot A)^{\tau_n}$ is adapted, continuous and of finite variation. Clearly, then $X \cdot A$ is adapted as well, and by Lemma 3.4.1 it is also continuous and of finite variation. □

Lemma 3.4.4. *$\mathfrak{L}^1(A)$ is a linear space, and the integral $I_A : \mathfrak{L}^1(A) \to$* **cFV** *is linear. Furthermore, if τ is any stopping time, $X[0,\tau] \in \mathfrak{L}^1(A)$ and $(X \cdot A)^\tau = X[0,\tau] \cdot A$.*

Proof. Since I_A is defined as a pathwise Lebesgue integral, this follows from the results of ordinary integration theory. $\square$

We now want to link the definition of the integral with respect to integrators in **cFV** together with our construction of the stochastic integral for integrators in $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$. By $\mathcal{S}$, we denote the space of processes of the form $A+M$, where $A \in \mathbf{cFV}$ and $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$. We call processes in $\mathcal{S}$ standard processes. Before making the obvious definition of integrals with respect to processes in $\mathcal{S}$, we need to make sure that the decomposition of such processes into finite variation part and local martingale part is unique.

Lemma 3.4.5. *If F has finite variation and is continuous, then V_F can be written as $V_F(t) = \sup \sum_{k=1}^n |F(t_k) - F(t_{k-1})|$, where the supremum is taken over partitions with rational timepoints.*

Proof. It will suffice to prove that for any $\varepsilon > 0$ and any partition $(t_0, \ldots, t_n)$, there exists another partition $(q_0, \ldots, q_n)$ such that

$$\left| \sum_{k=1}^n |F(t_k) - F(t_{k-1})| - \sum_{k=1}^n |F(q_k) - F(q_{k-1})| \right| < \varepsilon.$$

To this end, choose δ parrying $\frac{\varepsilon}{2n}$ for the continuity of F in $t_0, \ldots, t_n$, and let, for $k \leq n$, q_k be some rational with $|q_k - t_k| \leq \delta$. Then

$$|(F(t_k) - F(t_{k-1})) - (F(q_k) - F(q_{k-1}))| \leq \frac{\varepsilon}{n},$$

and since $|\cdot|$ is a contraction, this implies

$$||F(t_k) - F(t_{k-1})| - |F(q_k) - F(q_{k-1})|| \leq \frac{\varepsilon}{n},$$

finally yielding

$$\begin{aligned} & \left| \sum_{k=1}^n |F(t_k) - F(t_{k-1})| - \sum_{k=1}^n |F(q_k) - F(q_{k-1})| \right| \\ \leq \; & \sum_{k=1}^n ||F(t_k) - F(t_{k-1})| - |F(q_k) - F(q_{k-1})|| \\ \leq \; & \varepsilon. \end{aligned}$$

$\square$

Lemma 3.4.6. *Let $A \in$ **cFV**. Then V_A is continuous and adapted.*

Proof. From Lemma A.1.1, we know that V_A is continuous. To show that it is adapted, note that for any partition of $[0,t]$ we have $\sum_{k=1}^n |A_{t_k} - A_{t_{k-1}}| \in \mathcal{F}_t$. By Lemma 3.4.5, $V_A(t)$ can be written as the supremum of the above sums for partitions of $[0,t]$ with rational timepoints. Since there are only countably many of these, it follows that $V_A(t) \in \mathcal{F}_t$. □

Lemma 3.4.7. *Let $M \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$. If M has paths of finite variation, then M is evanescent.*

Proof. We first prove the lemma in a particularly nice setting and then use localisation to get the general result.

Step 1: Bounded variation and square-integrability. First assume that V_M is bounded and that $M \in \mathbf{c}\mathcal{M}_0^2$. Then, for any partition $(t_0, \ldots, t_n)$ of $[0,t]$ we obtain by Lemma 2.4.12,

$$
\begin{aligned}
EM_t^2 &= E\sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^2 \\
&\leq E\sup_{k\leq n} |M_{t_k} - M_{t_{k-1}}| \sum_{k=1}^n |M_{t_k} - M_{t_{k-1}}| \\
&\leq \|V_M(t)\|_\infty E\sup_{k\leq n} |M_{t_k} - M_{t_{k-1}}|.
\end{aligned}
$$

Now, by the continuity of M, $\sup_{k\leq n} |M_{t_k} - M_{t_{k-1}}|$ converges almost surely to zero as the partition grows finer. Noting that $\sup_{k\leq n} |M_{t_k} - M_{t_{k-1}}|$ is bounded by $\|V_M(t)\|_\infty$, dominated convergence yields $EM_t^2 = 0$, so M_t is almost surely zero. By continuity, M evanescent.

Step 2: Bounded variation. Now assume only that $V_M(t)$ is bounded. Since we have $M \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$, M is locally in $\mathbf{c}\mathcal{M}_0^2$. Let τ_n be a localising sequence. Then $V_{M^{\tau_n}}(t) \leq V_M(t)$, so the previous step yields that M^{τ_n} is evanescent. Letting n tend to infinity, we obtain that M is evanescent.

Step 3: The general case. Finally, merely assume that $M \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$ with finite variation. By Lemma 3.4.6, V_M is continuous and adapted. Therefore, by Lemma 2.3.5, τ_n defined by $\tau_n = \inf\{t \geq 0 | V_M(t) \geq n\}$ is a stopping time. Since V_M is bounded on compact intervals, τ_n tends to infinity. $M^{\tau_n} \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$ and $V_{M^{\tau_n}}(t)$ is bounded for any $t \geq 0$. By what was already shown, we find that M^{τ_n} is evanescent. We conclude that M is evanescent. □

Lemma 3.4.8. *Assume $X = A + M$, where A has finite variation and $M \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$. The decomposition of X into a finite variation process and a continuous local martingale is unique up to indistinguishability.*

Proof. Assume that we have another decomposition $X = B + N$. Then it holds that $A - B = N - M$. Therefore, the process $N - M$ is a continuous local martingale with paths of finite variation. By Lemma 3.4.7, $N - M$ is evanescent. Therefore, $A - B$ is also evanescent. In other words, A and B are indistinguishable, and so are M and N. $\square$

We are now finally ready to define the integral with respect to integrators in $\mathcal{S}$. Let $X \in \mathcal{S}$ with decomposition $X = A + M$. We put $L(X) = \mathfrak{L}^1(A) \cap \mathfrak{L}^2(M)$. We then simply define , for $Y \in L(X)$,

$$Y \cdot X = Y \cdot A + Y \cdot M.$$

This is well-defined up to indistinguishability, since if $X = B + N$ is another decomposition, then $Y \cdot A$ and $Y \cdot B$ are indistinguishable since A and B are indistinguishable, and $Y \cdot M$ and $Y \cdot N$ are indistinguishable since M and N are indistinguishable.

Lemma 3.4.9. *Let $X \in \mathcal{S}$ with $X = A + M$. If $Y \in L(X)$, then $Y \cdot X \in \mathcal{S}$ and has canonical decomposition given by $Y \cdot X = Y \cdot A + Y \cdot M$.*

Proof. This follows immediately from the fact that $Y \cdot A \in \mathbf{cFV}$ by Lemma 3.4.3 and the fact that $Y \cdot M \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$ by Definition 3.3.8. $\square$

Lemma 3.4.10. *$L(X)$ and $\mathcal{S}$ are linear spaces. $L(X)$ is $\mathfrak{L}_0$-stable and $\mathcal{S}$ is $\mathfrak{L}$-stable.*

Proof. By Lemma 3.4.1 and Lemma 3.3.11, $\mathbf{cFV}$ and $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ are both $\mathfrak{L}$-stable linear spaces. Therefore, $\mathcal{S}$ is also a $\mathfrak{L}$-stable linear space. Likewise, by Lemma 3.4.4 and Lemma 3.3.11, $\mathfrak{L}^1(A)$ and $\mathfrak{L}^2(M)$ are both $\mathfrak{L}_0$-stable linear spaces, and therefore $L(X)$ is a $\mathfrak{L}_0$-stable linear space. $\square$

Lemma 3.4.11. *The stochastic integral $I_X : L(X) \to \mathcal{S}$ is linear. If τ is a stopping time, $(Y \cdot X)^\tau = Y[0, \tau] \cdot X = Y \cdot X^\tau$.*

Proof. The conclusion is well-defined by Lemma 3.4.10. Since we have defined

$$Y \cdot X = Y \cdot A + Y \cdot M,$$

the linearity follows from the linearity of the integration with respect to A and M. Likewise, the stopping properties follow from the stopping properties with respect to A and M, shown in Lemma 3.4.4 and Lemma 3.3.11. $\square$

Finally, we show that the integral defined for integrators $X \in \mathcal{S}$ and integrals in $L(X)$ cannot be extended further by localisation, in other words, we show that we do not obtain larger classes of integrators and integrands by localising once more.

Lemma 3.4.12. *Let $X \in \mathcal{S}$ with decomposition $X = A + M$. Then, for the spaces of integrators it holds that $\mathfrak{L}(\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}) = \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$, $\mathfrak{L}(\mathbf{cFV}) = \mathbf{cFV}$ and $\mathfrak{L}(\mathcal{S}) = \mathcal{S}$. Also, for the spaces of integrands, we have $\mathfrak{L}_0(L(X)) = L(X)$, $\mathfrak{L}_0(\mathfrak{L}^2(M)) = \mathfrak{L}^2(M)$ and $\mathfrak{L}_0(\mathfrak{L}^1(A)) = \mathfrak{L}^1(A)$.*

Proof. This follows from Lemma 2.5.13. $\square$

This concludes our construction of the stochastic integral. We have defined the stochastic integral for any $X \in \mathcal{S}$ and $Y \in L(X)$ by putting $Y \cdot X = Y \cdot A + Y \cdot M$, where $X = A + M$ is the canonical decomposition. We have shown that the integral is linear and interacts properly with stopping times. Further properties of the integral follow by considering the components $Y \cdot A$ and $Y \cdot M$ and using respectively ordinary integration theory and the results from Section 3.3.

Before proceeding to the next section, we prove a few useful properties of the stochastic integral.

Lemma 3.4.13. *Let $X \in \mathcal{S}$. If Y if progressively measurable and locally bounded, then $Y \in L(X)$.*

Proof. Let $X = A + M$ be the canonical decomposition, and let τ_n be a localising sequence such that $M^{\tau_n} \in \mathbf{c}\mathcal{M}_W^2$. Let σ_n be a localising sequence for Y. By ordinary integration theory, $Y[0, \sigma_n] \in \mathfrak{L}^1(A)$. Since $\mu_{M^{\tau_k}}$ is bounded, $Y[0, \sigma_n] \in \mathcal{L}^2(M^{\tau_k})$. Therefore, $Y[0, \sigma_n][0, \tau_k] \in \mathcal{L}^2(M)$, so $Y[0, \sigma_n] \in \mathfrak{L}^2(M)$. In other words, Y is locally in $\mathfrak{L}^2(M)$, so by Lemma 3.4.12, $Y \in \mathfrak{L}^2(M)$. We conclude $Y \in L(X)$. $\square$

Next, we prove a result that shows that in a weak sense, the integral with respect to a standard process is consistent with the conventional concept of integration as a limit of Riemann sums. First, we need a lemma.

Lemma 3.4.14. *Let X_n be a sequence of processes and let τ_k be a sequence of stopping times tending to infinity. Let X be another process, and let $t \geq 0$ be given. Assume that X_n and X are constant from t and onwards. If for any $k \geq 1$, $X_n(\tau_k)$ converges to $X(\tau_k)$ as $n \to \infty$, either almost surely or in $\mathcal{L}^2$, then $X_n(t)$ converges to $X(t)$ in probability.*

Proof. Whether the convergence of $X_n(\tau_k)$ is almost sure or in $\mathcal{L}^2$, we have convergence in probability. To show convergence in probability of $X_n(t)$ to $X(t)$, let $\varepsilon > 0$ and $\eta > 0$ be given. Choose N so large that $P(\tau_k \leq t) \leq \eta$ for $k \geq N$. For such k, we find

$$\begin{aligned} P\left(|X_n(t) - X(t)| > \varepsilon\right) &\leq P\left(|X_n(t) - X(t)| > \varepsilon, \tau_k > t\right) + P(\tau_k \leq t) \\ &\leq P\left(|X_n(\tau_k) - X(\tau_k)| > \varepsilon\right) + \eta. \end{aligned}$$

Therefore,

$$\limsup_n P\left(|X_n(t) - X(t)| > \varepsilon\right) \leq \eta,$$

and since η was arbitrary, this shows the result. □

The usefulness of Lemma 3.4.14 may not be clear at present, but we will use it several times in the following, including the very next result to be proven. To set the scene, let $0 \leq s \leq t$, we say that $\pi = (t_0, \ldots, t_n)$ is a partition of $[s,t]$ if $s = t_0 < \cdots < t_n = t$. We define the norm of the norm $\|\pi\|$ of the partition by $\|\pi\| = \max_{k \leq n} |t_k - t_{k-1}|$. Let (X^π) be a sequence of variables indexed by the set of partitions of $[0,t]$, and let X be another variable. We say that X^π converges to X in d as the norm of the partitions tends to zero, or as the mesh tends to zero, if it holds that for any sequence of partitions π_n with norm tending to zero that X^{π_n} tends to X in d. Here, d represents some convergence concept, usually $\mathcal{L}^2$ convergence or convergence in probability. In general, we will attept to formulate our results in terms of convergence as the mesh tends to zero instead of using particular sequences of partitions, as the notation for such partitions easily becomes cumbersome. An alternative solution would be to consider nets instead of sequences. See Appendix C.1 for more on this approach.

Lemma 3.4.15. *Let $X \in \mathcal{S}$ and let Y be a bounded, adapted and continuous process. Then $Y \in L(X)$, and the Riemann sums*

$$\sum_{k=1}^{n} Y_{t_{k-1}}(X_{t_k} - X_{t_{k-1}})$$

converge in probability to $\int_0^t Y_s \, \mathrm{d}X_s$ as the mesh of the partition tends to zero.

Proof. Let $X = A + M$ be the canonical decomposition. $Y \in \mathfrak{L}^1(A)$ by Lemma 3.4.13. We need to show that for any sequence of partitions with norm tending to zero, the Riemann sums converge to the stochastic integral. Clearly, for any partition,

$$\sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} - X_{t_{k-1}}) = \sum_{k=1}^n Y_{t_{k-1}}(A_{t_k} - A_{t_{k-1}}) + \sum_{k=1}^n Y_{t_{k-1}}(M_{t_k} - M_{t_{k-1}}).$$

We will show that the first term converges to $\int_0^t Y_s \, \mathrm{d}A_s$ and that the second term converges to $\int_0^t Y_s \, \mathrm{d}M_s$. By Lemma A.1.5, the first term converges almost surely to $\int_0^t Y_s \, \mathrm{d}A_s$, therefore also in probability, when the norm of the partitions tends to zero. Next, consider the martingale part. First assume that $M \in \mathbf{c}\mathcal{M}_W^2$. Let $\pi_n = (t_0^n, \ldots, t_{m_n}^n)$ be given such that $\|\pi_n\|$ tends to zero and put

$$Y_s^n = \sum_{k=1}^{m_n} Y_{t_{k-1}^n} 1_{(t_{k-1}^n, t_k]}(s).$$

Then $Y^n \in \mathbf{b}\mathcal{E}$, and by continuity, Y^n converges pointwise to $Y[0,t]$. Since Y is bounded and μ_M is bounded, dominated convergence yields that $\lim \|Y^n - Y\|_M = 0$. By the Itô isometry, we conclude

$$\sum_{k=1}^{m_n} Y_{t_{k-1}^n}(M_{t_k^n} - M_{t_{k-1}^n}) = \int_0^t Y_s^n \, \mathrm{d}M_s \xrightarrow{\mathcal{L}^2} \int_0^t Y_s[0,t] \, \mathrm{d}M_s = \int_0^t Y_s \, \mathrm{d}M_s.$$

This shows the claim in the case where $M \in \mathbf{c}\mathcal{M}_W^2$. Now, if $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$, by Lemma 3.3.19 there exists a localising sequence τ_n such that $M^{\tau_n} \in \mathbf{c}\mathcal{M}_W^2$. We then have

$$\sum_{k=1}^n Y_{t_{k-1}}(M_{t_k}^{\tau_n} - M_{t_{k-1}}^{\tau_n}) \xrightarrow{\mathcal{L}^2} \int_0^t Y_s \, \mathrm{d}M_s^{\tau_n}$$

as the mesh tends to zero. Lemma 3.4.14 then allows us to conclude

$$\sum_{k=1}^n Y_{t_{k-1}}(M_{t_k} - M_{t_{k-1}}) \xrightarrow{P} \int_0^t Y_s \, \mathrm{d}M_s,$$

as desired. □

Comment 3.4.16 Let us take a moment to see how exactly the somewhat abstract Lemma 3.4.14 was used in the above proof, as we shall often use it again in the same manner without further details. Our situation is that we know that $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ with $M^{\tau_n} \in \mathbf{c}\mathcal{M}_W^2$, and

$$\sum_{k=1}^n Y_{t_{k-1}}(M_{t_k}^{\tau_n} - M_{t_{k-1}}^{\tau_n}) \xrightarrow{\mathcal{L}^2} \int_0^t Y_s \, \mathrm{d}M_s^{\tau_n}$$

as the mesh tends to zero. Let $\pi_n = (t_0^n, \ldots, t_{m_n}^n)$ be some sequence of partitions with norm tending to zero. We want to prove

$$\sum_{k=1}^{m_n} Y_{t_{k-1}^n}(M_{t_k^n} - M_{t_{k-1}^n}) \xrightarrow{P} \int_0^t Y_s \, \mathrm{d}M_s.$$

Define $X_n(u) = \sum_{k=1}^{m_n} Y_{t_{k-1}^n}(M^u_{t_k^n} - M^u_{t_{k-1}^n})$ and $X(u) = \int_0^t Y_s \, \mathrm{d}M_s^u$. Since all partitions are over $[0,t]$, X_n and X are constant from t onwards. And by what we already have shown, $X_n(\tau_k)$ converges in $\mathcal{L}^2$ to $X(\tau_k)$ for any $k \geq 1$. Lemma 3.4.14 may now be invoked to conclude

$$\sum_{k=1}^{m_n} Y_{t_{k-1}^n}(M_{t_k^n} - M_{t_{k-1}^n}) = X_n(t) \xrightarrow{P} X(t) = \int_0^t Y_s \, \mathrm{d}M_s,$$

as desired. Since the sequence of partitions was arbitrary, we conclude that we have convergence as the mesh tends to zero.

Obviously, then, the use of Lemma 3.4.14 in the proof of Lemma 3.4.15 included a good deal of implicit calculations, simple as they may be. It is obvious that if we had to go through all the details of taking out subsequences, applying Lemma 3.4.14, and going back to general partitions, our proofs would become very cluttered. We will therefore not go through all these diversions when using the lemma. All of the implicit details in such cases, however, are completely analogous to the ones that we have gone through above. ◦

Now that we have the definition of the integral in place, we begin the task of developing the basic results of the stochastic integral. We will prove Itô's formula, the martingale representation theorem and Girsanov's theorem. In order to do all this, we need to develop one of the basic tools of stochastic calculus, the quadratic variation process. This is the topic of the next section.

3.5 The Quadratic Variation

if F and G are two mappings from $[0,\infty)$ to $\mathbb{R}$, we say that F and G has quadratic covariation $Q : [0,\infty) \to \mathbb{R}$ if

$$\lim \sum_{k=1}^{n} (F(t_k) - F(t_{k-1}))(G(t_k) - G(t_{k-1})) = Q(t)$$

for all $t \geq 0$ as the mesh of the partitions tend to zero. The quadratic covariation of F with itself is called the quadratic variation of F. We will now investigate a concept much like this, but instead of real mappings, we will use standard processes. Thus, the basic point of this subsection is to understand the nature of sums of the form

$$\sum_{k=1}^{n}(X_{t_k} - X_{t_{k-1}})(X'_{t_k} - X'_{t_{k-1}})$$

where X and X' are standard processes, when the mesh of the partition tends to zero. Apart from being precisely the kind of variables that are necessary to consider in the proof of Itô's formula, it also turns out that the resulting *quadratic covariation process* is a tool of general applicability in the theory of stochastic calculus. We will show that if $X, X' \in \mathcal{S}$ with decompositions $X = A + M$ and $X' = A' + M'$, where $M = Y \cdot W$ and $M' = Y' \cdot W$, $Y, Y' \in \mathfrak{L}^2(W)^n$, then

$$\sum_{k=1}^{n}(X_{t_k} - X_{t_{k-1}})(X'_{t_k} - X'_{t_k}) \xrightarrow{P} \sum_{k=1}^{n} \int_0^t Y_s^k (Y'_s)^k \, \mathrm{d}s$$

as the mesh becomes finer and finer. For this reason, we will already from now on call $\sum_{k=1}^{n} \int_0^t Y_s^k (Y'_s)^k \, \mathrm{d}s$ the quadratic covariation of X and X', and we will use the notation $[X, X']_t = \sum_{k=1}^{n} \int_0^t Y_s^k (Y'_s)^k \, \mathrm{d}s$. Note that the quadratic covariation does not depend on the finite variation components of X and X' at all. In particular, $[X, X'] = [M, M']$. Also note that we always have $[X, X'] \in \mathbf{cFV}_0$.

We write $[X] = [X, X]$ and call $[X]$ the quadratic variation of X. Our strategy for developing the results of this section is first to develop all the necessary results for the quadratic variation process and then use a concept from Hilbert space theory, polarization, to obtain the corresponding results for the quadratic covariation.

Thus, at first, we will forget all about the quadratic covariation and consider only the quadratic variation. We begin by demonstrating some fundamental properties of the quadratic variation process. Our first lemma shows how to calculate the quadratic variation of a stochastic integral with respect to a martingale integrator. Note that we are here using that the quadratic variation is a standard process and can therefore by used as an integrator.

Lemma 3.5.1. *Let $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$. Then $X \in \mathcal{L}^2(M)$ if and only if $X \in \mathcal{L}^2([M])$. Likewise, $X \in \mathfrak{L}^2(M)$ if and only if $X \in \mathfrak{L}^2([M])$. In the affirmative case, the equality $[X \cdot M] = X^2 \cdot [M]$ holds.*

Proof. Let $M = Y \cdot W$, where $Y \in \mathfrak{L}^2(W)^n$. We then have

$$\begin{aligned} \sum_{k=1}^n E \int_0^\infty X_s^2 (Y_s^k)^2 \, \mathrm{d}s &= E \int_0^\infty X_s^2 \sum_{k=1}^n (Y_s^k)^2 \, \mathrm{d}s \\ &= E \int_0^\infty X_s^2 \, \mathrm{d}[M]_s. \end{aligned}$$

The first expression is finite if and only if $X \in \mathcal{L}^2(M)$, and the final expression is finite if and only if $X \in \mathcal{L}^2([M])$. Therefore, $X \in \mathcal{L}^2(M)$ if and only if $X \in \mathcal{L}^2([M])$. From localisation, it follows that $X \in \mathfrak{L}^2(M)$ if and only if $X \in \mathfrak{L}^2([M])$.

In order to show the integral equality, assume that $X \in \mathfrak{L}^2(M)$. $X \cdot M = XY \cdot W$, and therefore the definition of the quadratic variation yields

$$\begin{aligned} [X \cdot M]_t &= \sum_{k=1}^n \int_0^t (X_s Y_s^k)^2 \, \mathrm{d}s \\ &= \int_0^t X_s^2 \sum_{k=1}^n (Y_s^k)^2 \, \mathrm{d}s \\ &= \int_0^t X_s^2 \, \mathrm{d}[M]_s \\ &= (X^2 \cdot [M])_t. \end{aligned}$$

This shows the lemma. □

Our next two lemmas gives a martingale characterization of the quadratic variation and shows that the quadratic variation interacts properly with stopping.

Lemma 3.5.2. *Let X be a standard process, and let τ be a stopping time. Then $[X]^\tau = [X^\tau]$.*

Proof. Let $X = A + M$ with $M = Y \cdot W$, $Y \in \mathfrak{L}^2(W)^n$. X^τ then has decomposition $X^\tau = B + N$, where $B = A^\tau$ and $N = Y[0, \tau] \cdot W$, and

$$[X]_t^\tau = \sum_{k=1}^n \int_0^{t \wedge \tau} (Y_s^k)^2 \, \mathrm{d}s = \sum_{k=1}^n \int_0^t (Y^k[0, \tau]_s)^2 \, \mathrm{d}s = [X^\tau]_t,$$

as desired. □

Lemma 3.5.3. *If $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$, $[M]$ is the unique process in $\mathbf{cFV}_0$, up to indistinguishability, that makes $M_t^2 - [M]_t$ a local martingale. Also, if $M \in \mathbf{c}\mathcal{M}_W^2$, $M_t^2 - [M]_t$ is a uniformly integrable martingale.*

Proof. We first prove that $M_t^2 - [M]_t$ is a uniformly integrable martingale when we have $M \in \mathbf{c}\mathcal{M}_W^2$. Assume $M = Y \cdot W$, $Y \in \mathcal{L}^2(W)^n$. First note that M has an almost sure limit. Letting τ be any stopping time, we obtain

$$E\left(M_\tau^2 - [M]_\tau\right) = E\left(\sum_{k=1}^n \int_0^\infty Y[0,\tau]_s \, \mathrm{d}W_s\right)^2 - \sum_{k=1}^n E\int_0^\infty (Y^k[0,\tau])_s^2 \, \mathrm{d}s = 0$$

by Itô's isometry, Lemma 3.3.6. Lemma 2.4.13 now yields that $M_t^2 - [M]_t$ is a uniformly integrable martingale. In the case $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$, letting τ_n be a localising sequence, by Lemma 3.5.2, $[M]^{\tau_n} = [M^{\tau_n}]$ and therefore $(M^2 - [M])_t^{\tau_n} = (M^{\tau_n})_t^2 - [M^{\tau_n}]_t$ is a martingale, showing that $M_t^2 - [M]_t$ is a local martingale.

It remains to show uniqueness. Let A and B be two processes in $\mathbf{cFV}_0$ such that $N_t^A = M_t^2 - A_t$ and $N_t^B = M_t^2 - B_t$ both are local martingales. Then $N^A - N^B = B - A$, so $B - A$ is a continuous local martingale of finite variation. By Lemma 3.4.7, A and B are indistinguishable. □

As we will experience in the following, Lemma 3.5.3 is a very important result in the sense that the fact that $M_t^2 - [M]_t$ is a uniformly integrable martingale whenever $M \in \mathbf{c}\mathcal{M}_W^2$ is the *only* fact we will need to identify $[M]_t$ as the quadratic variation. In particular, we can in the following completely forget about the specific form of M as an integral with respect to W. Our next goal will be to show that $[X]$ really can be interpreted as a quadratic variation, we will show the result

$$\sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^2 \xrightarrow{P} [X]_t.$$

For any partition π and stochastic processes X and X' we define the quadratic covariation of X and X' over π by $Q^\pi(X, X') = \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})(X'_{t_k} - X'_{t_{k-1}})$. We write $Q^\pi(X) = Q^\pi(X, X)$. The next two lemmas are technical results that pave the road for the result on the convergence of the squared increments $Q^\pi(X)$ to the quadratic variation.

Lemma 3.5.4. *If $M \in \mathbf{c}\mathcal{M}_W^2$ and Y is a bounded and adapted process, it holds that*

$$\begin{aligned} & E\left(\sum_{k=1}^n Y_{t_{k-1}}((M_{t_k} - M_{t_{k-1}})^2 - ([M]_{t_k} - [M]_{t_{k-1}}))\right)^2 \\ = \quad & E\sum_{k=1}^n \left(Y_{t_{k-1}}((M_{t_k} - M_{t_{k-1}})^2 - ([M]_{t_k} - [M]_{t_{k-1}}))^2\right)^2, \end{aligned}$$

in other words, the cross-terms equal zero.

Proof. Consider $i < j$ with $i \leq n, j \leq n$. Let $X_i = (M_{t_i} - M_{t_{i-1}})^2 - ([M]_{t_i} - [M]_{t_{i-1}})$. Then $X_i \in \mathcal{F}_{t_i} \subseteq \mathcal{F}_{t_{j-1}}$, so we obtain

$$EY_{t_{i-1}}X_iY_{t_{j-1}}X_j = E(Y_{t_{i-1}}X_iY_{t_{j-1}}E(X_j|\mathcal{F}_{t_{j-1}}).$$

Now note

$$\begin{aligned}
E(X_j|\mathcal{F}_{t_{j-1}}) &= E((M_{t_j} - M_{t_{j-1}})^2 - ([M]_{t_j} - [M]_{t_{j-1}})|\mathcal{F}_{t_{j-1}}) \\
&= E(M_{t_j}^2 - [M]_{t_j} + M_{t_{j-1}}^2 + [M]_{t_{j-1}} - 2M_{t_j}M_{t_{j-1}}|\mathcal{F}_{t_{j-1}}) \\
&= E(M_{t_j}^2 - [M]_{t_j} - (M_{t_{j-1}}^2 - [M]_{t_{j-1}})|\mathcal{F}_{t_{j-1}}) \\
&= 0,
\end{aligned}$$

by Lemma 3.5.3. Therefore, we find

$$\begin{aligned}
& E\left(\sum_{k=1}^n Y_{t_{k-1}}((M_{t_k} - M_{t_{k-1}})^2 - ([M]_{t_k} - [M]_{t_{k-1}}))^2\right)^2 \\
= \; & E\left(\sum_{k=1}^n Y_{t_{k-1}}X_k\right)^2 \\
= \; & \sum_{k=1}^n\sum_{i=1}^n EY_{t_{k-1}}X_kY_{t_{i-1}}X_i \\
= \; & E\sum_{k=1}^n Y_{t_{k-1}}^2X_k^2 \\
= \; & E\sum_{k=1}^n \left(Y_{t_{k-1}}((M_{t_k} - M_{t_{k-1}})^2 - ([M]_{t_k} - [M]_{t_{k-1}}))^2\right)^2,
\end{aligned}$$

as desired. □

Lemma 3.5.5. *Let M be a continous bounded martingale, zero at zero. Then*

$$E\sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^4$$

tends to zero as the norm of the partition tends to zero.

Proof. We first employ the Cauchy-Schwartz inequality to obtain

$$\begin{aligned}
\sum_{k=1}^n E(M_{t_k} - M_{t_{k-1}})^4 &\leq E\sup_{k\leq n}(M_{t_k} - M_{t_{k-1}})^2\sum_{k=1}^n(M_{t_k} - M_{t_{k-1}})^2 \\
&\leq \left(E\sup_{k\leq n}(M_{t_k} - M_{t_{k-1}})^4\right)^{\frac{1}{2}}\left(EQ^\pi(M)^2\right)^{\frac{1}{2}}.
\end{aligned}$$

The first factor converges to zero by dominated convergence as $\|\pi\|$ tends to zero, since M is continuous and bounded. It is therefore sufficient for our needs to show that the second factor is bounded. We rewrite as

$$\begin{aligned}
& EQ^\pi(M)^2 \\
= \;& E\left(\sum_{k=1}^n (M_{t_k}-M_{t_{k-1}})^2\right)^2 \\
= \;& \sum_{k=1}^n E(M_{t_k}-M_{t_{k-1}})^4 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n E(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}-M_{t_{j-1}})^2.
\end{aligned}$$

The last term can be computed using Lemma 2.4.12 twice,

$$\begin{aligned}
& 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n E(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}-M_{t_{j-1}})^2 \\
= \;& 2\sum_{i=1}^{n-1} E\left((M_{t_i}-M_{t_{i-1}})^2 E\left(\left.\sum_{j=i+1}^n (M_{t_j}-M_{t_{j-1}})^2\right|\mathcal{F}_{t_i}\right)\right) \\
= \;& 2\sum_{i=1}^{n-1} E\left((M_{t_i}-M_{t_{i-1}})^2 E((M_t-M_{t_i})^2|\mathcal{F}_{t_i})\right) \\
\leq \;& 8\|M^*\|_\infty^2 \sum_{i=1}^{n-1} E(M_{t_i}-M_{t_{i-1}})^2 \\
\leq \;& 8\|M^*\|_\infty^2 EM_t^2.
\end{aligned}$$

We then find, again using Lemma 2.4.12,

$$\begin{aligned}
EQ^\pi(M)^2 &\leq \sum_{k=1}^n E(M_{t_k}-M_{t_{k-1}})^4 + 8\|M^*\|_\infty^2 EM_t^2 \\
&\leq 4\|M^*\|_\infty^2 \sum_{k=1}^n E(M_{t_k}-M_{t_{k-1}})^2 + 8\|M^*\|_\infty^2 EM_t^2 \\
&\leq 4\|M^*\|_\infty^2 EM_t^2 + 8\|M^*\|_\infty^2 EM_t^2 \\
&= 12\|M^*\|_\infty^2 EM_t^2,
\end{aligned}$$

showing the desired boundedness. □

Finally, we are ready to show that the quadratic variation can actually be interpreted as a quadratic variation. We begin with the martingale case.

Theorem 3.5.6. *Let $M \in \mathbf{c}\mathcal{M}_W^2$. Assume that M and $[M]$ are bounded. With π a partition of $[0,t]$, $Q^\pi(M)$ tends to $[M]_t$ in $\mathcal{L}^2$ as $\|\pi\|$ tends to zero.*

Proof. By Lemma 3.5.4, using $(x-y)^2 \leq 2x^2 + 2y^2$,

$$\begin{aligned}
E\left(Q^\pi(M) - [M]_t\right)^2 &= E\left(\sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^2 - ([M]_{t_k} - [M]_{t_{k-1}})\right)^2 \\
&= \sum_{k=1}^n E((M_{t_k} - M_{t_{k-1}})^2 - ([M]_{t_k} - [M]_{t_{k-1}}))^2 \\
&\leq 2\sum_{k=1}^n E(M_{t_k} - M_{t_{k-1}})^4 + 2\sum_{k=1}^n E([M]_{t_k} - [M]_{t_{k-1}})^2.
\end{aligned}$$

To show the result of the theorem, it therefore suffices to show that each of the two sums above converges to zero as $\|\pi\|$ tends to zero. Now, the first sum converges to zero by 3.5.5. Concerning the second sum, we have, since $[M]$ is increasing,

$$\begin{aligned}
\sum_{k=1}^n E\left([M]_{t_k} - [M]_{t_{k-1}}\right)^2 &\leq E \sup_{k\leq n} \left|[M]_{t_k} - [M]_{t_{k-1}}\right| \sum_{k=1}^n \left|[M]_{t_k} - [M]_{t_{k-1}}\right| \\
&\leq E \sup_{k\leq n} \left|[M]_{t_k} - [M]_{t_{k-1}}\right| \sum_{k=1}^n [M]_{t_k} - [M]_{t_{k-1}} \\
&= E[M]_t \sup_{k\leq n} \left|[M]_{t_k} - [M]_{t_{k-1}}\right|,
\end{aligned}$$

which tends to zero by dominated convergence as $\|\pi\|$ tends to zero, since $[M]$ is continuous and bounded. $\square$

Corollary 3.5.7. *Let $M \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$. Then $Q^\pi(M)$ converges in probability to $[M]_t$ as $\|\pi\|$ tends to zero.*

Proof. Let $\tau_n = \inf\{t \geq 0 | M_t > n \text{ or } [M]_t > n\}$. Then τ_n is a stopping time, and τ_n tends to infinity because M and $[M]$ are continuous. Furthermore, M^{τ_n} and $[M^{\tau_n}]$ are bounded. Thus, the hypotheses of Theorem 3.5.6 are satisfied and $Q^\pi(M^{\tau_n})$ converges in $\mathcal{L}^2$ to $\int_0^t Y[0,\tau_n]_s^2 \, \mathrm{d}s$ as $\|\pi\|$ tends to zero. Since we know $Q^\pi(M^{\tau_n}) = Q^\pi(M)^{\tau_n}$ and $[M]^{\tau_n} = [M^{\tau_n}]$, the result now follows from Lemma 3.4.14. $\square$

Theorem 3.5.8. *Let $X \in \mathcal{S}$. With π a partition of $[0,t]$, $Q^\pi(X)$ tends to $[X]_t$ in probability as $\|\pi\|$ tends to zero.*

Proof. Letting $X = A + M$, we first note

$$\begin{aligned}\sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^2 &= \sum_{k=1}^n (A_{t_k} - A_{t_{k-1}} + M_{t_k} - M_{t_{k-1}})^2 \\ &= \sum_{k=1}^n (A_{t_k} - A_{t_{k-1}})^2 \\ &+ 2\sum_{k=1}^n (A_{t_k} - A_{t_{k-1}})(M_{t_k} - M_{t_{k-1}}) \\ &+ \sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^2.\end{aligned}$$

For the first term, we obtain

$$\left|\sum_{k=1}^n (A_{t_k} - A_{t_{k-1}})^2\right| \le \max_{k\le n} |A_{t_k} - A_{t_{k-1}}| V_A(t),$$

and since A is continuous, the above tends to zero almost surely. For the second term, we find

$$\left|\sum_{k=1}^n (A_{t_k} - A_{t_{k-1}})(M_{t_k} - M_{t_{k-1}})\right| \le \max_{k\le n} |M_{t_k} - M_{t_{k-1}}| V_A(t),$$

and again conclude that this tends to zero almost surely, by the continuity of M. And from Corollary 3.5.7, we know that the last term tends to $[M]_t$, which is equal to $[X]_t$. □

Theorem 3.5.8 shows that the process $[X]$ really is a quadratic variation in the sense that it is a limit of quadratic increments. It remains to show that $[X, X']$ has the analogous interpretation of a quadratic covariation. Before this, we show a simple generalisation of Theorem 3.5.8 which will be useful to us in the proof of Itô's formula. It is the analogue to Lemma 3.4.15, using quadratic Riemann sums instead of ordinary Riemann sums.

Corollary 3.5.9. *Let $X \in \mathcal{S}$. Let Y be a bounded, adapted and continuous process. Then the quadratic Riemann sums*

$$\sum_{k=1}^n Y_{t_{k-1}} (X_{t_k} - X_{t_{k-1}})^2$$

converge in probability to $\int_0^t Y_s \, \mathrm{d}[X]_s$ as the mesh of the partition tends to zero.

Proof. Letting $X = A + M$, we have

$$\begin{aligned}
\sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} - X_{t_{k-1}})^2 &= \sum_{k=1}^n Y_{t_{k-1}}(A_{t_k} - A_{t_{k-1}} + M_{t_k} - M_{t_{k-1}})^2 \\
&= \sum_{k=1}^n Y_{t_{k-1}}(A_{t_k} - A_{t_{k-1}})^2 \\
&+ 2\sum_{k=1}^n Y_{t_{k-1}}(A_{t_k} - A_{t_{k-1}})(M_{t_k} - M_{t_{k-1}}) \\
&+ \sum_{k=1}^n Y_{t_{k-1}}(M_{t_k} - M_{t_{k-1}})^2.
\end{aligned}$$

For the first term, we obtain

$$\left|\sum_{k=1}^n Y_{t_{k-1}}(A_{t_k} - A_{t_{k-1}})^2\right| \le \|Y\|_\infty \max_{k \le n} |A_{t_k} - A_{t_{k-1}}| V_A(t),$$

and since A is continuous, the above tends to zero almost surely. For the second term, we find

$$\left|\sum_{k=1}^n Y_{t_{k-1}}(A_{t_k} - A_{t_{k-1}})(M_{t_k} - M_{t_{k-1}})\right| \le \|Y\|_\infty \max_{k \le n} |M_{t_k} - M_{t_{k-1}}| V_A(t),$$

and again conclude that this tends to zero almost surely, by the continuity of M. Therefore, it will suffice to show that the final term converges in probability to $\int_0^t Y_s \,\mathrm{d}[X]_s$. Recall here that $[X] = [M]$, so we really have to show convergence to $\int_0^t Y_s \,\mathrm{d}[M]_s$.

We first consider the case where M and $[M]$ are bounded. By Lemma 3.4.15,

$$\sum_{k=1}^n Y_s([M]_{t_k} - [M]_{t_{k-1}}) \xrightarrow{P} \int_0^t Y_s \,\mathrm{d}[M]_s,$$

so it will suffice to show that $\sum_{k=1}^n Y_s((M_{t_k} - M_{t_{k-1}})^2 - ([M]_{t_k} - [M]_{t_{k-1}}))$ tends to zero in probability. But we have, by Lemma 3.5.4,

$$\begin{aligned}
& E\left(\sum_{k=1}^{n} Y_s((M_{t_k}-M_{t_{k-1}})^2-([M]_{t_k}-[M]_{t_{k-1}}))\right)^2 \\
= \;& E\sum_{k=1}^{n}(Y_s((M_{t_k}-M_{t_{k-1}})^2-([M]_{t_k}-[M]_{t_{k-1}})))^2 \\
\leq \;& \|Y^*\|_\infty^2 E\sum_{k=1}^{n}((M_{t_k}-M_{t_{k-1}})^2-([M]_{t_k}-[M]_{t_{k-1}}))^2 \\
= \;& \|Y^*\|_\infty^2 E\left(\sum_{k=1}^{n}((M_{t_k}-M_{t_{k-1}})^2-([M]_{t_k}-[M]_{t_{k-1}}))\right)^2 \\
= \;& \|Y^*\|_\infty^2 E\left(Q^\pi(M)-[M]_t\right)^2,
\end{aligned}$$

which converges to zero by Theorem 3.5.6. Now consider the general case where $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$. Let τ_n be a localising sequence such that M^{τ_n} and $[M]^{\tau_n}$ are bounded, this can be done since $[M]$ and M both are continuous. By Lemma 3.5.2, $[M]^{\tau_n} = [M^{\tau_n}]$, and we then obtain

$$\sum_{k=1}^{n} Y_{t_{k-1}}(M_{t_k}^{\tau_n}-M_{t_{k-1}}^{\tau_n})^2 \xrightarrow{P} \int_0^{t\wedge\tau_n} Y_s \,\mathrm{d}[X]_s.$$

The result then follows from Lemma 3.4.14. $\square$

We now transfer the results on the quadratic variation to the setting of the quadratic covariation. Recall that we defined $[X,X']_t = \int_0^t Y_s Y'_s \,\mathrm{d}s$. As mentioned earlier, the method to do so is called polarization. The name originates from the polarization identity for inner product spaces,

$$\langle x,y\rangle = \frac{1}{4}\left(\|x+y\|^2-\|x-y\|^2\right),$$

see Lemma 11.2 of Meise & Vogt (1997). The polarization identity allows one to recover the inner product from the norm. Thinking of the quadratic variation as the square norm and of the quadratic covariation as the inner product, we will show that the two satisfies a relation analogous to the above, and we will use this relation to obtain results on the quadratic covariation from the quadratic variation.

Lemma 3.5.10. *Let $X, X' \in \mathcal{S}$. Then*

$$[X,X'] = \frac{1}{4}\left([X+X']-[X-X']\right).$$

Proof. Assume that $X = A+M$ and $X' = A'+M'$, where $M = Y\cdot W$ and $M' = Y'\cdot W$. Therefore,

$$\begin{aligned}\frac{1}{4}\left([X+X']_t - [X-X']_t\right) &= \frac{1}{4}\left(\int_0^t (Y_s+Y_s')^2\,\mathrm{d}s - \int_0^t (Y_s-Y_s')^2\,\mathrm{d}s\right)\\ &= \frac{1}{4}\int_0^t 4Y_sY_s'\,\mathrm{d}s\\ &= [X,X'].\end{aligned}$$

□

Lemma 3.5.11. *Let $M, N \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$. Then $[M,N]$ is the unique process in $\mathbf{cFV}_0$ such that $MN - [M,N]$ is a local martingale.*

Proof. That $MN - [M,N]$ is a local martingale follows from Lemma 3.5.3, since

$$\begin{aligned}MN - [M,N] &= \frac{1}{4}((M+N)^2 - (M-N)^2) - \frac{1}{4}([M+N]-[M-N])\\ &= \frac{1}{4}((M+N)^2 - [M+N]) - \frac{1}{4}((M-N)^2 - [M-N]).\end{aligned}$$

Uniqueness follows as in Lemma 3.5.3 from Lemma 3.4.7. □

Theorem 3.5.12. *Let $X, X' \in \mathcal{S}$ and let Y be a bounded, adapted and continuous process. The quadratic Riemann sums*

$$\sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} - X_{t_{k-1}})(X'_{t_k} - X'_{t_{k-1}})$$

tend to $\int_0^t Y_s\,\mathrm{d}[X,X']_s$ as the mesh of the partitions tend to zero.

Proof. We find that

$$\begin{aligned}&\sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} - X_{t_{k-1}})(X'_{t_k} - X'_{t_{k-1}})\\ =\;& \frac{1}{4}\sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} + X'_{t_k} - (X_{t_{k-1}} + X'_{t_{k-1}}))^2\\ -\;& \frac{1}{4}\sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} - X'_{t_k} - (X_{t_{k-1}} - X'_{t_{k-1}}))^2.\end{aligned}$$

By Corollary 3.5.9, the first term converges in probability to $\frac{1}{4}\int_0^t Y_s\,\mathrm{d}[X+X']_s$ and the second term converges in probability to $\frac{1}{4}\int_0^t Y_s\,\mathrm{d}[X-X']_s$, as the mesh of the partitions tend to zero. Therefore, $\sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} - X_{t_{k-1}})(X'_{t_k} - X'_{t_{k-1}})$ converges in probability to

$$\begin{aligned}\frac{1}{4}\int_0^t Y_s \,\mathrm{d}[X+X']_s + \frac{1}{4}\int_0^t Y_s \,\mathrm{d}[X-X']_s &= \int_0^t Y_s \,\mathrm{d}\left(\frac{1}{4}[X+X']_s + \frac{1}{4}[X-X']_s\right) \\ &= \int_0^t Y_s \,\mathrm{d}[X,X']_s\end{aligned}$$

by Lemma 3.5.10. □

The main results on the quadratic covariation of this section are Lemma 3.5.11 and Theorem 3.5.12. Putting $Y = 1$ in Theorem 3.5.12 shows that our definition of the quadratic covariation of two standard processes really can be interpreted as a quadratic covariation, although the convergence is only in probability.

Before concluding, we note an important consequence of the definition of the quadratic variation. Since the quadratic variation is a continuous and adapted process of finite variation, it is a standard process, and we can therefore integrate with respect to it. The following lemma exploits this feature.

Lemma 3.5.13. *Let $X, X' \in \mathcal{S}$ and assume that $Y \in L^1(X)$ and $Y' \in L^1(X')$. Then $YY' \in L^1([X,X'])$ and*

$$[Y \cdot X, Y' \cdot X'] = YY' \cdot [X, X'].$$

Proof. Let the canonical decompositions be $X = A + M$ and $X' = A' + M'$. Then $[Y \cdot X, Y' \cdot X'] = [Y \cdot M, Y' \cdot M']$ and $[X, X'] = YY' \cdot [M, M']$, so it will suffice to show the result in the case where the finite variation components are zero. The integrability result follows from the definition of the quadratic covariation and results from measure theory. For the second statement of the lemma, we calculate

$$[Y \cdot M, Y' \cdot M']_t = \sum_{k=1}^n \int_0^t Y_s Z_s^k Y'_s (Z'_s)^k \,\mathrm{d}s = \int_0^t Y_s Y'_s \,\mathrm{d}[M,M']_s = (YY' \cdot [M,M'])_s.$$

□

Lemma 3.5.11 provides a characterisation of the quadratic covariation only in terms of martingales. This result shows that quadratic covariation is essentially a purely martingale concept, even though we here have developed it in the context of stochastic integration. This observation is the key to extending the stochastic integral to using general continuous local martingales as integrators. The price of this generality is that in a purely martingale setting, we do not have any obvious candidate for the quadratic covariation. In our setting, the existence of the quadratic variation was a triviality once the correct idea had been identified. In general, the proof of the existence of the quadratic covariation is quite difficult. In Rogers & Williams (2000b), Theorem 30.1, a direct construction of the quadratic variation of a continuous square-integrable martingale is carried out. Alternatively, one can obtain the existence in a more abstract manner by invoking the Doob-Meyer decomposition theorem, as is done in Karatzas & Shreve (1988). Another alternative is presented in Appendix C.1, where we give a general existence result on the quadratic variation in the continuous case based only on relatively elementary martingale theory.

3.6 Itô's Formula

In this section, we will prove the multidimensional Itô formula. We say that X is an n-dimensional standard process if each of X's coordinates is a standard process. The Itô formula states that if X is a n-dimensional standard process and $f \in C^2(\mathbb{R}^n)$, then

$$f(X_t) = f(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\,\mathrm{d}[X^i, X^j]_s.$$

This important result in particular shows that a C^2 transformation of a standard process is again a standard process, and allows us to identify the canonical decomposition. To prove this result, we will first localise to a particularly nice setting, use a Taylor expansion of f and take limits. Before starting the proof, we prepare ourselves with the right version of the Taylor expansion theorem and a result that will help us when localising in the proof.

Theorem 3.6.1. *Let $f \in C^2(\mathbb{R}^n)$, and let $x, y \in \mathbb{R}^n$. There exists a point $\xi \in \mathbb{R}^n$ on the line segment between x and y such that*

$$f(y) = f(x) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x)(y_i - x_i) + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\xi)(y_i - x_i)(y_j - x_j).$$

Proof. Define $g : \mathbb{R} \to \mathbb{R}$ by $g(t) = f(x + t(y - x))$. Note that $g(1) = f(y)$ and $g(0) = f(x)$. We will prove the theorem by applying the one-dimensional Taylor formula, see Apostol (1964) Theorem 7.6, to g. Clearly, $g \in C^2(\mathbb{R})$, and we obtain

$g(1) = g(0) + g'(0) + \frac{1}{2}g''(s)$, where $0 \leq s \leq 1$. Applying the chain rule, we find

$$g'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x + t(y-x))(y_i - x_i)$$

and

$$g''(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x + t(y-x))(y_i - x_i)(y_j - x_j).$$

Substituting and writing $\xi = x + s(y-x)$, we may conclude

$$f(y) = f(x) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)(y_i - x_i) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\xi)(y_i - x_i)(y_j - x_j).$$

□

Lemma 3.6.2. *Let $f \in C^2(\mathbb{R}^n)$. For any compact set K, There exists $g \in C_c^2(\mathbb{R}^n)$ such that f and g are equal on K.*

Proof. By Urysohn's Lemma, Theorem A.3.4, there exists $\psi \in C_c^\infty(\mathbb{R}^n)$ with $K \prec \psi$. Defining $g(x) = \psi(x)f(x)$, $g \in C_c^2(\mathbb{R}^n)$ and f and g are clearly equal on K. □

Theorem 3.6.3. *Let X be a n-dimensional standard process, and let $f \in C^2(\mathbb{R}^n)$. Then*

$$f(X_t) = f(X_0) + \sum_{i=1}^{n} \int_0^t \frac{\partial f}{\partial x_i}(X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\,\mathrm{d}[X^i, X^j]_s,$$

up to indistinguishability.

Proof. First note that all of the integrands are continous, therefore locally bounded and therefore, by Lemma 3.4.13 the stochastic integrals are well-defined.

Our plan is first to reduce by localization to a particularly nice setting, apply the Taylor formula to obtain an expression for $f(X_t) - f(X_0)$ in terms of Riemann-like sums and then use Lemma 3.4.15 and Theorem 3.5.12 on convergence of Riemann and quadratic Riemann sums.

Step 1: The localized case. We first consider the case where X is bounded. The variables $f(X_t)$, $\frac{\partial f}{\partial x_i}(X_t)$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}(X_t)$ does not depend on the values of f outside the range of X. Therefore, we can by Lemma 3.6.2 assume that f has compact support. In that case, all of f, $\frac{\partial f}{\partial x_i}$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ are bounded. Consider a partition $\pi = (t_0, \ldots, t_n)$ of $[0,t]$. The Taylor Formula 3.6.1 with Lagrange's remainder term then yields

$$\begin{aligned}
& f(X_t) - f(X_0) \\
= \; & \sum_{k=1}^n f(X_{t_k}) - f(x_{t_{k-1}}) \\
= \; & \sum_{i=1}^n \sum_{k=1}^n \frac{\partial f}{\partial x_i}(X_{t_{k-1}})(X^i_{t_k} - X^i_{t_{k-1}}) \\
+ \; & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\xi_k)(X^i_{t_k} - X^i_{t_{k-1}})(X^j_{t_k} - X^j_{t_{k-1}}).
\end{aligned}$$

We will investigate the limit of the sums as the mesh of the partition tends to zero. By Lemma 3.4.15, since $\frac{\partial f}{\partial x_i}$ is bounded, $\sum_{k=1}^n \frac{\partial f}{\partial x_i}(X_{t_{k-1}})(X^i_{t_k} - X^i_{t_{k-1}})$ converges in probability to $\int_0^t \frac{\partial f}{\partial x_i}(X_s) \, \mathrm{d}X^i_s$ as the mesh of the partitions tend to zero. Therefore,

$$\sum_{i=1}^n \sum_{k=1}^n \frac{\partial f}{\partial x_i}(X_{t_{k-1}})(X^i_{t_k} - X^i_{t_{k-1}}) \xrightarrow{P} \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s) \, \mathrm{d}X^i_s.$$

It remains to consider the second-order terms. Fix $i, j \leq n$. We first note that the difference between the sums given by $\sum_{k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\xi_k)(X^i_{t_k} - X^i_{t_{k-1}})(X^j_{t_k} - X^j_{t_{k-1}})$ and $\sum_{k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_{k-1}})(X^i_{t_k} - X^i_{t_{k-1}})(X^j_{t_k} - X^j_{t_{k-1}})$ is bounded by

$$\max_{k \leq n} \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(\xi_k) - \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_{k-1}}) \right| \sum_{k=1}^n (X^i_{t_k} - X^i_{t_{k-1}})(X^j_{t_k} - X^j_{t_{k-1}}).$$

By continuity of X and f'', the maximum tends to zero almost surely. By Theorem 3.5.12, $\sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^2$ tends to $[X]_t$ in probability. Therefore, the above tends to zero in probability as the mesh tends to zero. It will therefore suffice to prove that

$$\sum_{k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_{k-1}})(X^i_{t_k} - X^i_{t_{k-1}})(X^j_{t_k} - X^j_{t_{k-1}}) \xrightarrow{P} \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \, \mathrm{d}[X^i, X^j]_s,$$

as the mesh tends to zero, but this follows immediately from Theorem 3.5.12. We may now conclude that for each $t \geq 0$,

$$f(X_t) = f(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s) \, \mathrm{d}X^i_s + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \, \mathrm{d}[X^i, X^j]_s,$$

almost surely. Since both sides are continuous, we have equality up to indistinguishability.

Step 2: The general case. Since each X^i is continuous, there exists a sequence of stopping times τ_n tending to infinity such that X^{τ_n} is bounded. By what we have already shown, we then have

$$\begin{aligned}
f(X_t)^{\tau_n} &= f(X_t^{\tau_n}) \\
&= f(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s^{\tau_n})\,\mathrm{d}(X^{\tau_n})_s^i \\
&+ \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s^{\tau_n})\,\mathrm{d}[(X^{\tau_n})^i, (X^{\tau_n})^j]_s \\
&= f(X_0) + \sum_{i=1}^n \int_0^{t\wedge\tau_n} \frac{\partial f}{\partial x_i}(X_s)\,\mathrm{d}(X)_s^i \\
&+ \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^{t\wedge\tau_n} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\,\mathrm{d}[X^i, X^j]_s.
\end{aligned}$$

The Itô formula therefore holds whenever $t \le \tau_n$. Since τ_n tends to infinity, the theorem is proven. □

We have now proven Itô's formula. This result is more or less the center of the theory of stochastic integration. We will sometimes need to apply Itô's formula to mappings which are not defined on all of $\mathbb{R}^n$. The following corollary shows that this is possible.

Corollary 3.6.4. *Let X be a n-dimensional standard process taking its values in an open set U. Let $f \in C^2(U)$. Then*

$$f(X_t) = f(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\,\mathrm{d}[X^i, X^j]_s,$$

up to indistinguishability.

Proof. Define $F_n = \{x \in \mathbb{R}^n | d(x, U^c) \ge \frac{1}{n}\}$. Our plan is to in some sense localise to F_n and prove the result there. F_n is a closed set with $F_n \subseteq U$. Let $g_n \in C^\infty(\mathbb{R}^n)$ be such that $F_n \prec g_n \prec F_{n+1}^c$, such a mapping exists by Lemma A.3.5. Define $f_n(x) = g_n(x)f(x)$ when $x \in U$ and zero otherwise. Clearly, f_n and f agree on F_n. We will argue that $f_n \in C^2(\mathbb{R}^n)$. To this end, note that since f and g_n are both C^2 on U, it is clear that f_n is C^2 on U. Since f_n is zero on F_{n+1}^c, f_n is C^2 on F_{n+1}^c. Because $\mathbb{R}^n = U^c \cup U \subseteq F_{n+1}^c \cup U$, this shows $f_n \in C^2(\mathbb{R}^n)$. Itô's Lemma then yields

$$f_n(X_t) = f_n(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f_n}{\partial x_i}(X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^t \frac{\partial^2 f_n}{\partial x_i \partial x_j}(X_s)\,\mathrm{d}[X^i, X^j]_s.$$

Now define $\tau_n = \inf\{t \geq 0 | d(X_t, U^c) \leq \frac{1}{n}\}$. τ_n is a stopping time, and since $d(X_t, U^c)$ is a positive continuous process, τ_n tends to infinity. When $t \leq \tau_n$, $d(X_t, U^c) < \frac{1}{n}$, so $X_t \in F_n$. Since f_n and f agree on F_n, we conclude

$$\begin{aligned}
& f(X_t)^{\tau_n} \\
= \quad & f_n(X_0) + \sum_{i=1}^n \int_0^{t\wedge\tau_n} \frac{\partial f_n}{\partial x_i}(X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^{t\wedge\tau_n} \frac{\partial^2 f_n}{\partial x_i \partial x_j}(X_s)\,\mathrm{d}[X^i, X^j]_s \\
= \quad & f(X_0) + \sum_{i=1}^n \int_0^{t\wedge\tau_n} \frac{\partial f}{\partial x_i}(X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^{t\wedge\tau_n} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\,\mathrm{d}[X^i, X^j]_s.
\end{aligned}$$

Letting τ_n tend to infinity, we obtain the result. □

3.7 Itô's Representation Theorem

In this section and the following, we develop some results specifically for the context where we have a n-dimensional Brownian motion W on $(\Omega, \mathcal{F}, P)$ and let the filtration $\mathcal{F}_t$ be the one induced by W, augmented in the usual manner as described in Section 2.1. We will need to work with processes on $[0, T]$ and integration on $[0, T]$. We will therefore make some notation for this situation. By $\mathcal{L}_T^2(W)$, we denote the space $\mathcal{L}^2([0, T] \times \Omega, \Sigma_\pi[0, T], \lambda \otimes P)$, understanding that $\Sigma_\pi[0, T]$ is the restriction of Σ_π to $[0, T] \times \Omega$. The norm on $\mathcal{L}_T^2(W)$ is denoted $\|\cdot\|_{W^T}$. The integral of a process $Y \in \mathcal{L}_T^2(W)^n$ is then considered a process on $[0, T]$, defined as the integral of the process $Y[0, T]$.

Our goal of this section is to prove Itô's representation theorem, which yields a representation of all cadlag $\mathcal{F}_t$ local martingales. To prove the result, we will need some preparation. We begin with a few lemmas.

Lemma 3.7.1. *Assume that $X \in \mathcal{S}$ with $X_0 = 0$ and let Y_0 be a constant. There exists precisely one solution to the equation $Y_t = Y_0 + \int_0^t Y_s\,\mathrm{d}X_s$ in Y, unique up to indistinguishability, and the solution is given by $Y_t = Y_0 \exp(X_t - \frac{1}{2}[X]_t)$.*

Proof. We first check that Y as given in the lemma in fact solves the equation. The case $Y_0 = 0$ is trivial, so we assume $Y_0 \neq 0$. We note that $Y_t = f(X_t, [X]_t)$ with $f(x, y) = Y_0 \exp(x - \frac{1}{2}y)$. Itô's formula then yields, using that $X_0 = 0$ and therefore

$\exp(X_0 - \frac{1}{2}[X]_0) = 1$,

$$\begin{aligned} Y_t &= Y_0 + \int_0^t Y_s \, \mathrm{d}X_s - \frac{1}{2}\int_0^t Y_s \, \mathrm{d}[X]_s + \frac{1}{2}\int_0^t Y_s \, \mathrm{d}[X]_s \\ &= Y_0 + \int_0^t Y_s \, \mathrm{d}X_s, \end{aligned}$$

as desired. Now consider uniqueness. Let Y be any solution of the equation, and define Y' by $Y' = \exp(-X_t + \frac{1}{2}[X]_t)$. Itô's formula yields

$$\begin{aligned} Y'_t &= Y_0 - \int_0^t Y'_s \, \mathrm{d}X_s + \frac{1}{2}\int_0^t Y'_s \, \mathrm{d}[X]_s + \frac{1}{2}\int_0^t Y'_s \, \mathrm{d}[X]_s \\ &= Y_0 - \int_0^t Y'_s \, \mathrm{d}X_s + \int_0^t Y'_s \, \mathrm{d}[X]_s. \end{aligned}$$

Therefore, again by Itô's formula, using Lemma 3.5.13,

$$\begin{aligned} Y_t Y'_t &= Y_0 + \int_0^t Y_s \, \mathrm{d}Y'_s + \int_0^t Y'_s \, \mathrm{d}Y_s + [Y, Y']_t \\ &= Y_0 - \int_0^t Y_s Y'_s \, \mathrm{d}X_s + \int_0^t Y_s Y'_s \, \mathrm{d}[X]_s + \int_0^t Y'_s Y_s \, \mathrm{d}X_s - \int_0^t Y_s Y'_s \, \mathrm{d}[X]_s \\ &= Y_0, \end{aligned}$$

up to indistinguishability. Inserting the expression for Y', we conclude

$$Y_t = Y_0 \exp\left(X_t - \frac{1}{2}[X]_t\right),$$

demonstrating uniqueness. □

The equation $Y_t = Y_0 + \int_0^t \mathrm{d}X_s$ of Lemma 3.7.1 is called a stochastic differential equation, or a SDE. This equation is one of the few SDEs where an explicit solution is available. This type of equation will play an important role in the following, and we therefore introduce some special notation for the solution. For any $X \in \mathcal{S}$ with $X_0 = 0$, we define the stochastic exponential of X by $\mathcal{E}(X)_t = \exp(X_t - \frac{1}{2}[X]_t)$. The stochastic exponential is also known as the Doléans-Dade exponential. If M is a local martingale with $M_0 = 0$, we see from $\mathcal{E}(M)_t = 1 + \int_0^t \mathcal{E}(M)_t \, \mathrm{d}M_t$ that $\mathcal{E}(M)$ also is a local martingale. We will now give a criterion to check when $\mathcal{E}(M)$ is a square-integrable martingale.

Lemma 3.7.2. *Let M be a nonnegative continuous local martingale. Then M is a supermartingale.*

Proof. Let $0 \leq s \leq t$, and let τ_n be a localising sequence. By Fatou's Lemma for conditional expectations and continuity of M,

$$\begin{aligned} E(M_t|\mathcal{F}_s) &= E(\liminf M_t^{\tau_n}|\mathcal{F}_s) \\ &\geq \liminf E(M_t^{\tau_n}|\mathcal{F}_s) \\ &= \liminf M_s^{\tau_n} \\ &= M_s. \end{aligned}$$

□

Lemma 3.7.3. *Assume that $Y \in \mathfrak{L}^2(W)^n$ and that there exists $f_k \in \mathcal{L}^2[0,\infty)$ such that $|Y_t^k| \leq f_k(t)$ for $k \leq n$. Put $M = Y \cdot W$. Then $\mathcal{E}(M) \in \mathbf{c}\mathcal{M}_0^2$.*

Proof. From Lemma 3.7.1, we have $\mathcal{E}(M)_t = 1 + \int_0^t \mathcal{E}(M)_s \,\mathrm{d}M_s$. Therefore, Itô's formula yields

$$\begin{aligned} \mathcal{E}(M)_t^2 &= 1 + 2\int_0^t \mathcal{E}(M)_s \,\mathrm{d}\mathcal{E}(M)_s + [\mathcal{E}(M)]_t \\ &= 1 + 2\int_0^t \mathcal{E}(M)_s^2 \,\mathrm{d}M_s + \int_0^t \mathcal{E}(M)_s^2 \,\mathrm{d}[M]_s \\ &= 1 + 2\int_0^t \mathcal{E}(M)_s^2 \,\mathrm{d}M_s + \sum_{k=1}^n \int_0^t \mathcal{E}(M)_s^2 (Y_s^k)^2 \,\mathrm{d}s \\ &\leq 1 + 2\int_0^t \mathcal{E}(M)_s^2 \,\mathrm{d}M_s + \int_0^t \mathcal{E}(M)_s^2 \sum_{k=1}^n f_k(s)^2 \,\mathrm{d}s. \end{aligned}$$

Define $b(s) = \sum_{k=1}^n f_k(s)^2$. Since $\int_0^t \mathcal{E}(M)_s^2 \,\mathrm{d}M_s$ is a continuous local martingale, there exists a localising sequence τ_m such that $\mathcal{E}(M)_{t\wedge\tau_m}^2$ is bounded over t and $\int_0^{t\wedge\tau_m} \mathcal{E}(M)_s^2 \,\mathrm{d}M_s$ is a bounded martingale in t. We then obtain

$$\begin{aligned} E\mathcal{E}(M)_{t\wedge\tau_m}^2 &\leq 1 + E\int_0^{t\wedge\tau_m} \mathcal{E}(M)_s^2 b(s) \,\mathrm{d}s \\ &= 1 + E\int_0^t \mathcal{E}(M)_s^2 b(s) 1_{[0,\tau_m]}(s) \,\mathrm{d}s \\ &= 1 + E\int_0^t \mathcal{E}(M)_{s\wedge\tau_m}^2 b(s) 1_{[0,\tau_m]}(s) \,\mathrm{d}s \\ &\leq 1 + E\int_0^t \mathcal{E}(M)_{s\wedge\tau_m}^2 b(s) \,\mathrm{d}s \\ &= 1 + \int_0^t E(\mathcal{E}(M)_{s\wedge\tau_m}^2) b(s) \,\mathrm{d}s, \end{aligned}$$

by the Tonelli theorem. Now consider m fixed. Gronwall's Lemma A.2.1 then yields

$$E\mathcal{E}(M)^2_{t\wedge\tau_m} \leq \exp\left(\int_0^t b(s)\,\mathrm{d}s\right).$$

Using Fatou's Lemma, we then obtain

$$E\mathcal{E}(M)^2_t = E\liminf_m \mathcal{E}(M)^2_{t\wedge\tau_m} \leq \liminf_m E\mathcal{E}(M)^2_{t\wedge\tau_m} = \exp\left(\int_0^t b(s)\,\mathrm{d}s\right).$$

Since $\mathcal{E}(M)$ is a nonnegative local martingale, it is a supermartingale by Lemma 3.7.2. Therefore, $\mathcal{E}(M)$ is almost surely convergent by the Martingale Convergence Theorem 2.4.3. For the limit we then obtain, again using Fatou's Lemma,

$$E\mathcal{E}(M)^2_\infty = E\liminf_t \mathcal{E}(M)^2_t \leq \liminf_t E\mathcal{E}(M)^2_t \leq \exp\left(\int_0^\infty b(s)\,\mathrm{d}s\right) < \infty.$$

Therefore, $\mathcal{E}(M)$ is bounded in $\mathcal{L}^2$. This shows the lemma. □

Our next lemma will require some basic results from the theory of weak convergence. An excellent standard reference to this is Billingsley (1999).

Lemma 3.7.4. *Let $h^1,\ldots,h^n \in \mathcal{L}^2[0,T]$, and let W denote a one-dimensional Brownian motion. Then the vector $(\int_0^T h^1_s\,\mathrm{d}W_s,\ldots,\int_0^T h^n_s\,\mathrm{d}W_s)$ follows a n-dimensional normal distribution with mean zero and variance matrix Σ given by $\Sigma_{ij} = \int_0^T h^i_s h^j_s\,\mathrm{d}s$.*

Proof. First consider the case where $h^1,\ldots,h^n$ are continuous. Let $\pi = (t_0,\ldots,t_m)$ be a partition of $[0,T]$ and define X^π by putting $X^\pi_i = \sum_{k=1}^m h^i(t_{k-1})(W_{t_k} - W_{t_{k-1}})$. X^π is then a stochastic variable with values in $\mathbb{R}^n$. Since X^π is a linear transformation of normal variables, X^π is normally distributed. It is clear that X^π has mean zero, and the covariance matrix Σ^π of X^π is

$$\begin{aligned}\Sigma^\pi_{ij} &= E\left(\sum_{k=1}^m h^i(t_{k-1})(W_{t_k} - W_{t_{k-1}})\right)\left(\sum_{k=1}^m h^j(t_{k-1})(W_{t_k} - W_{t_{k-1}})\right)\\ &= \sum_{k=1}^m h^i(t_{k-1})h^j(t_{k-1})(t_k - t_{k-1})^2.\end{aligned}$$

Now, as the mesh tends to zero, Σ^π_{ij} tends to $\int_0^T h^i_s h^j_s\,\mathrm{d}s$. Therefore, X^π tends to the normal distribution with mean zero and the covariance matrix Σ given in the statement of the lemma.

On the other hand, for any $a \in \mathbb{R}^n$, we have by Lemma 3.4.15,

$$\sum_{i=1}^n a_i X_i^\pi \xrightarrow{P} \sum_{i=1}^n a_i \int_0^T h_i(s)\,\mathrm{d}W_s.$$

as the mesh tends to zero. Since convergence in probability implies weak convergence, by the Cramér-Wold Device, X^π tends to $(\int_0^T h_n(s)\,\mathrm{d}W_s, \ldots, \int_0^T h_n(s)\,\mathrm{d}W_s)$ weakly. By uniqueness of limits, we conclude that the lemma holds in the case of continuous $h^1, \ldots, h^n$.

Now consider the general case, where $h^1, \ldots, h^n \in \mathcal{L}^2[0,T]$. For $i \leq n$, there exists a sequence (h_k^i) of continuous maps converging in $\mathcal{L}^2[0,T]$ to h^i. By what we already have shown, these are normally distributed with zero mean and covariance matrix $\Sigma_{ij}^k = \int_0^T h_k^i(s) h_k^j(s)\,\mathrm{d}s$. Since the integral is continuous, the Cramér-Wold Device and uniqueness of limits again yields the conclusion. $\square$

Lemma 3.7.5. *The span of the variables*

$$\exp\left(\sum_{k=1}^n \int_0^T h_k(t)\,\mathrm{d}W_t^k - \frac{1}{2}\sum_{k=1}^n \int_0^T h_k^2(t)\,\mathrm{d}t\right)$$

where $h_1, \ldots, h_n \in \mathcal{L}^2[0,T]$, is dense in $\mathcal{L}^2(\mathcal{F}_T)$.

Proof. Let $(\mathcal{G}_t)$ denote the filtration generated by W. From the construction of the usual augmentation in Theorem 2.1.4, we know that $\mathcal{F}_T = \sigma(\mathcal{G}_{T+}, \mathcal{N})$, where $\mathcal{N}$ denotes the null sets of $\mathcal{F}$. By Lemma 2.1.9, $\mathcal{G}_{T+} \subseteq \sigma(\mathcal{G}_T, \mathcal{N})$. We therefore conclude $\mathcal{F}_T \subseteq \sigma(\mathcal{G}_T, \mathcal{N})$, so it will suffice to show that the span of the variables given in the statement of the lemma is dense in $\mathcal{L}^2(\sigma(\mathcal{G}_T, \mathcal{N}))$. By Lemma B.6.9, it will then suffice to show that the variables are dense in $\mathcal{L}^2(\mathcal{G}_T)$.

To this end, first note that $\int_0^T 1_{[0,s]}(t)\,\mathrm{d}W_t^k = W_s^k$, so $\mathcal{G}_T$ is generated by the variables of the form $\sum_{k=1}^n \int_0^T h_k(t)\,\mathrm{d}W_t^k$. Also, by Lemma 3.7.4, the vector of variables $\sum_{k=1}^n \int_0^T h_k(t)\,\mathrm{d}W_t^k$ is normally distributed, in particular it has exponential moments of all orders. It then follows from linearity of the integral and Theorem B.6.2 that the variables $\exp(\sum_{k=1}^n \int_0^T h_k(t)\,\mathrm{d}W_t^k)$ are dense in $\mathcal{L}^2(\mathcal{G}_T)$. Since $\frac{1}{2}\sum_{k=1}^n \int_0^T h_k^2(t)\,\mathrm{d}t$ is constant, the conclusion of the lemma follows. $\square$

We are now more or less ready to begin the proof of the martingale representation theorem. We start out with a version of the theorem for finite time intervals.

Theorem 3.7.6. *Let $X \in \mathcal{L}^2(\mathcal{F}_T)$. There exists $Y \in \mathcal{L}^2_T(W)^n$ such that*

$$X = EX + \sum_{k=1}^{n} \int_0^T Y_s^k \,\mathrm{d}W_s^k$$

almost surely. Y^k is unique $\lambda \otimes P$ almost surely.

Proof. We first show the theorem for variables of the form given in Lemma 3.7.5 and then extend the result by a density argument.

Step 1: A special case. Let $h_1, \ldots, h_n \in \mathcal{L}^2[0,T]$ be given, and put

$$X_s = \exp\left(\sum_{k=1}^{n} \int_0^s h_k(t)\,\mathrm{d}W_t^k - \frac{1}{2}\sum_{k=1}^{n}\int_0^s h_k^2(t)\,\mathrm{d}t\right)$$

for $s \leq T$. Our goal is to prove the theorem for X_T. Note that defining M by putting $M_t = \sum_{k=1}^n \int_0^t h_k(s)\,\mathrm{d}W_s^k$, $X_t = \mathcal{E}(M)_t$ and so, by Lemma 3.7.1, $X_T = 1 + \int_0^T X_s \,\mathrm{d}M_s$. According to Lemma 3.7.3, X is a square-integrable martingale, and therefore we may in particular conclude $EX_T = 1$, showing

$$X_T = EX_T + \sum_{k=1}^{n} \int_0^T X_s h_k(s)\,\mathrm{d}W_s^k.$$

It remains to argue that $s \mapsto X_s h_k(s)$ is in $\mathcal{L}^2(W)$. Since X is a square-integrable martingale, its squared expectation is increasing. We therefore obtain

$$E\int_0^T (X_s h_k(s))^2 \,\mathrm{d}s = \int_0^T h_k(s)^2 E X_s^2 \,\mathrm{d}s \leq EX_T^2 \int_0^T h_k(s)^2 \,\mathrm{d}s,$$

which is finite. Thus, the process $s \mapsto X_s h_k(s)$ is in $\mathcal{L}^2_T(W)$, and the theorem is proven for X_T.

Step 2: The general case. In order to extend the result to the genral case, let $\mathbb{H}$ be the class of variables in $\mathcal{L}^2(\mathcal{F}_T)$ where the theorem holds. Clearly, $\mathbb{H}$ is a linear space. By the first step and Lemma 3.7.5, we have therefore shown the result for a dense subspace of $\mathcal{L}^2(\mathcal{F}_T)$. The general case will therefore follow if we can show that $\mathbb{H}$ is closed.

To do so, assume that X_n is a sequence in $\mathbb{H}$ converging towards $X \in \mathcal{L}^2(\mathcal{F}_T)$. Then $X_n = EX_n + \sum_{k=1}^n \int_0^T Y_n^k(s)\,\mathrm{d}W_s^k$ almost surely, for some processes $Y_n \in \mathcal{L}^2_T(W)^n$.

Since X_n converges in $\mathcal{L}^2$, the norms and therefore the second moments also converge. Therefore, using Itô's isometry, Lemma 3.3.6,

$$\begin{aligned}\sum_{k=1}^n \|Y_n^k - Y_m^k\|_{W^T}^2 &= E\left(\sum_{k=1}^n \int_0^T Y_n^k(s) - Y_m^k(s)\,\mathrm{d}W_s^k\right)^2 \\ &= E\left(X_n - EX_n - (X_m - EX_m)\right)^2 \\ &= \|(X_n - X_m) - (EX_n - EX_m)\|_2^2.\end{aligned}$$

Since both $\|X_n - X_m\|_2$ and $|EX_n - EX_m|$ tends to zero as n and m tends to infinity, we conclude that for each $k \le n$, (Y_n^k) is a cauchy sequence in $\mathcal{L}_T^2(W)$. By completeness, there exists Y^k such that $\|Y_n^k - Y^k\|_{W^T}$ tends to zero, and we then obtain

$$\begin{aligned}X &= \lim_m X_m \\ &= \lim_m EX_m + \sum_{k=1}^n Y_m^k(s)\,\mathrm{d}W_s^k \\ &= EX + \sum_{k=1}^n Y^k(s)\,\mathrm{d}W_s^k\end{aligned}$$

almost surely, by the continuity of the integral, where the limits are in $\mathcal{L}^2$. The existence part of the theorem is now proven.

Step 3: Uniqueness. To show uniqueness, assume that we have two representations

$$X = EX + \sum_{k=1}^n \int_0^T Y_s^k\,\mathrm{d}W_s^k = EX + \sum_{k=1}^n \int_0^T Z_s^k\,\mathrm{d}W_s^k.$$

The Itô isometry yields

$$\sum_{k=1}^n \|Y^k - Z^k\|_{W^T}^2 = E\left(\sum_{k=1}^n \int_0^T Y^k(s) - Z^k(s)\,\mathrm{d}W_s^k\right)^2 = 0,$$

so Y^k and Z^k are equal $\lambda \otimes P$ almost surely, as desired. □

Theorem 3.7.7 (The Martingale Representation Theorem). *Assume that M is a $\mathcal{F}_t$ cadlag local martingale. Then, there exists $Y \in \mathfrak{L}^2(W)^n$ such that*

$$M_t = M_0 + \sum_{k=1}^n \int_0^t Y_s^k\,\mathrm{d}W_s^k,$$

almost surely. If M is square-integrable, Y can be taken to be in $\mathcal{L}^2(W)^n$. Y^k is unique $\lambda \otimes P$ almost surely.

Proof. Uniqueness follows immediately from Lemma 3.3.7. We therefore need only consider existence. Obviously, it will suffice to consider the case where $M_0 = 0$. This is proven in several steps:

1. Square-integrable martingales.
2. Continuous martingales.
3. Cadlag uniformly integrable martingales.
4. Cadlag local martingales.

Step 1: The square-integrable case. Assume that M is square-integrable starting at zero. In particular, $EM = 0$. For $m \geq 1$, let $Y_m \in \mathcal{L}^2_m(W)^n$ be the process that exists by Theorem 3.7.6 such that

$$M_m = \sum_{k=1}^n \int_0^m Y^k_m(s)\,\mathrm{d}W^k_s$$

almost surely. We then have, by the martingale property of the stochastic integral,

$$\begin{aligned}\sum_{k=1}^n \int_0^m Y^k_{m+1}(s)\,\mathrm{d}W^k_s &= E\left(\sum_{k=1}^n \int_0^{m+1} Y^k_{m+1}(s)\,\mathrm{d}W^k_s \,\middle|\, \mathcal{F}_m\right)\\ &= E\left(M_{m+1}|\mathcal{F}_m\right)\\ &= M_m\\ &= \sum_{k=1}^n \int_0^m Y^k_m(s)\,\mathrm{d}W^k_s,\end{aligned}$$

almost surely. Then, by the uniqueness part of Theorem 3.7.6, Y^k_{m+1} and Y^k_m must be $\lambda \otimes P$ almost surely equal on $[0, m] \times \Omega$. By the Pasting Lemma 2.5.10, there exists Y^k such that Y^k is almost surely equal to Y^k_m on $[0, m] \times \Omega$. In particular, $Y[0, t] \in \mathcal{L}^2(W)^n$ for all $t \geq 0$. This implies that for $t \geq 0$,

$$\begin{aligned}M_t &= E(M_{[t]+1}|\mathcal{F}_t)\\ &= E\left(\sum_{k=1}^n \int_0^{[t]+1} Y^k_{[t]+1}(s)\,\mathrm{d}W^k_s \,\middle|\, \mathcal{F}_t\right)\\ &= E\left(\sum_{k=1}^n \int_0^{[t]+1} Y^k_s\,\mathrm{d}W^k_s \,\middle|\, \mathcal{F}_t\right)\\ &= \sum_{k=1}^n \int_0^t Y^k_s\,\mathrm{d}W^k_s,\end{aligned}$$

almost surely, as desired. By Lemma 3.3.16, $Y \in \mathcal{L}^2(W)^n$.

Step 2: The continuous case. Now let M be a continuous martingale starting at zero. Then M is locally a square-integrable martingale. Let τ_n be a localising sequence. By what we already have shown, there exists $Y_m \in \mathcal{L}^2(W)^n$ such that $M_t^{\tau_m} = \sum_{k=1}^n \int_0^t Y_m^k(s)\,\mathrm{d}W_s^k$ almost surely. We then obtain

$$\begin{aligned} \sum_{k=1}^n \int_0^t Y_{m+1}^k[0,\tau_m](s)\,\mathrm{d}W_s^k &= \left(\sum_{k=1}^n \int_0^t Y_{m+1}^k(s)\,\mathrm{d}W_s^k\right)^{\tau_m} \\ &= (M^{\tau_{m+1}})_t^{\tau_m} \\ &= M_t^{\tau_m} \\ &= \sum_{k=1}^n \int_0^t Y_m^k(s)\,\mathrm{d}W_s^k. \end{aligned}$$

Since both integrands are in $\mathcal{L}_t^2(W)^n$, we conclude $Y_{m+1}^k[0,\tau_m] = Y_m^k$ $\lambda \otimes P$ almost surely on $[0,t]$. Letting t tend to infinity, we may then conclude $Y_{m+1}^k[0,\tau_m] = Y_m^k$ $\lambda \otimes P$ almost surely on all of $[0,\infty)$. By the Pasting Lemma, there exists Y^k such that $Y^k[0,\tau_m] = Y_m^k[0,\tau_m]$, $\lambda \otimes P$ almost surely. In particular, $Y^k[0,\tau_m] \in \mathcal{L}^2(W)$, so $Y \in \mathfrak{L}^2(W)^n$, and

$$\begin{aligned} M_t^{\tau_m} &= \sum_{k=1}^n \int_0^t Y_m^k[0,\tau_m]_s\,\mathrm{d}W_s^k \\ &= \sum_{k=1}^n \int_0^t Y^k[0,\tau_m]_s\,\mathrm{d}W_s^k \\ &= \left(\sum_{k=1}^n \int_0^t Y_s^k\,\mathrm{d}W_s^k\right)^{\tau_m}, \end{aligned}$$

almost surely. Letting m tend to infinity, we get $M_t = \sum_{k=1}^n \int_0^t Y_s^k\,\mathrm{d}W_s^k$ almost surely. This shows the theorem in the case where M is a continuous martingale.

Step 3: The cadlag uniformly integrable case. Next, let M be a cadlag uniformly integrable martingale starting at zero. By Theorem 2.4.3, M is convergent almost surely and is closed by its limit, M_∞. Let M_∞^m be a sequence of bounded variables converging towards M_∞ in $\mathcal{L}^1$. Assume in particular that $\|M_\infty - M_\infty^m\|_1 \leq \frac{1}{3^n}$. M_∞^m is square-integrable, so putting $M_t^m = E(M_\infty^m|\mathcal{F}_t)$, M^m is a square-integrable martingale. By what we already have proven, there are $Y_m \in \mathcal{L}^2(W)^n$ such that $M_t^m = M_0 + \sum_{k=1}^n \int_0^t Y_m^k(s)\,\mathrm{d}W_s^k$ almost surely. In particular, there exists a version N^m of M^m in $\mathbf{c}\mathcal{M}_0^2$.

Next note that by Doob's maximal inequality, Theorem 2.4.2,

$$P\left((N^m - M)_t^* > \frac{1}{2^m}\right) \leq 2^m E|N_t^m - M_t| = 2^m E|M_t^m - M_t|$$

Letting t tend to infinity, we obtain by $\mathcal{L}^1$-convergence

$$P\left((N^m - M)^* > \frac{1}{2^m}\right) \leq 2^m E|M_\infty^m - M_\infty| \leq \frac{2^m}{3^m}.$$

By the Borel-Cantelli Lemma, then, $(N^m - M)^*$ tends to zero almost surely, so N^m converges almost surely uniformly to M. In particular, M must have a continuous version, so the conclusion in this case follows from what we proved in the previous step.

Step 4: The cadlag local martingale case. Finally, assume that M is a cadlag local martingale starting at zero. By Lemma 2.5.16, M is locally a cadlag uniformly integrable martingale. Letting τ_m be a localising sequence, we know that the theorem holds for M^{τ_m}. We can then use the pasting technique from step 2 to obtain the desired representation for M. □

The martingale representation theorem has the following important corollary.

Corollary 3.7.8. *Let M be a $\mathcal{F}_t$ local martingale. Then there is $Y \in \mathfrak{L}^2(W)^n$ such that M and $Y \cdot W$ are versions. In particular, any $\mathcal{F}_t$ local martingale has a continuous version.*

Proof. By Theorem 2.4.1, there is a cadlag version of M. We can then apply Theorem 3.7.7 to obtain the desired result. □

Corollary 3.7.8 essentially characterises all martingales with respect to the augmented Brownian filtration. It provides considerable insight into the structure of $\mathcal{F}_t$ martingales. In particular, it shows that when we are considering the filtration $\mathcal{F}_t$, our definition of the stochastic integral encompasses all local martingales as integrators. Therefore, in the following, whenever we are considering the augmented Brownian filtration, we can assume whenever convenient that we are dealing with continuous local martingales instead of the class $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ of stochastic integrals with respect to Brownian motion.

3.8 Girsanov's Theorem

In this section, we will prove Girsanov's Theorem. As in the previous section, we let $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ be a probability space with a n-dimensional Brownian motion W, where $(\Omega, \mathcal{F}, P)$ is complete and $\mathcal{F}_t$ is usual augmentation of the filtration induced by the Brownian motion. We let $\mathcal{F}_\infty = \sigma(\cup_{t\geq 0}\mathcal{F}_t)$. Girsanov's Theorem characterises the measures which are equivalent to P on $\mathcal{F}_\infty$ and shows how the distribution of W varies when changing to an equivalent measure. We will see that changing the measure corresponds to changing the drift of the Brownian motion.

The theorem thus has two parts, describing:

1. The form of the measures equivalent to P on $\mathcal{F}_\infty$.
2. The distribution of W under Q, when Q is equivalent to P on $\mathcal{F}_\infty$.

After proving both the results, we will prove Kazamakis and Novikovs conditions on the existence of measure changes corresponding to specific changes of drift.

The first result is relatively straightforward and only requires a small lemma.

Lemma 3.8.1. *Let Q be a measure on $\mathcal{F}$ such that Q is absolutely continuous with respect to P on $\mathcal{F}_\infty$. Let Q_t denote the restriction of Q to $\mathcal{F}_t$, and let P_t denote the restriction of P to $\mathcal{F}_t$. Define the likelihood process L_t by $L_t = \frac{\mathrm{d}Q_t}{\mathrm{d}P_t}$. Then L is a uniformly integrable martingale, and it is closed by $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}$.*

Proof. All the claims of the lemma will follow if we can show $E(\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}|\mathcal{F}_t) = L_t$ for any $t \geq 0$. To do so, let $t \geq 0$ and let $A \in \mathcal{F}_t$. Using Lemma B.6.8 twice yields

$$\int_A \frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}\,\mathrm{d}P = \int_A \frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}\,\mathrm{d}P_\infty = \int_A \mathrm{d}Q_\infty = \int_A \mathrm{d}Q_t.$$

Making the same calculations in the reverse direction then yields

$$\int_A \mathrm{d}Q_t = \int_A \frac{\mathrm{d}Q_t}{\mathrm{d}P_t}\,\mathrm{d}P_t = \int_A \frac{\mathrm{d}Q_t}{\mathrm{d}P_t}\,\mathrm{d}P = \int_A L_t\,\mathrm{d}P.$$

Thus, $E(1_A\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}) = E(1_A L_t)$ for any $A \in \mathcal{F}_t$, and therefore $E(\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}|\mathcal{F}_t) = L_t$. This shows that L is a martingale closed by $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}$. In particular, L is uniformly integrable and has the limit $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}$ almost surely and in $\mathcal{L}^1$. □

Theorem 3.8.2 (Girsanov's Theorem, part one). *Let Q be a measure on $\mathcal{F}$ equivalent to P on $\mathcal{F}_\infty$. There exists a process $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ such that $\mathcal{E}(M)$ is a uniformly integrable martingale with the property $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty} = \mathcal{E}(M)_\infty$.*

Proof. Let L_t be the likelihood process for $\frac{\mathrm{d}Q}{\mathrm{d}P}$. By Lemma 3.8.1, L is a uniformly integrable martingale. Since the filtration is assumed to satisfy the usual conditions, there exists a cadlag version of L. Since L_t is almost surely nonnegative for any $t \geq 0$, we can take the cadlag version to be nonnegative almost surely. By Theorem 3.7.7, there exists processes $Y \in \mathfrak{L}^2(W)^n$ such that we have $L_t = \sum_{k=1}^n \int_0^t Y_s \, \mathrm{d}W_s$. In particular, L can be taken to be continuous.

Our plan is to show that L satisfies an exponential SDE, this will yield the desired result. If L almost surely has positive paths, we can do this simply by writing

$$L_t = \sum_{k=1}^n \int_0^t L_s \frac{Y_s^k}{L_s} \, \mathrm{d}W_s^k,$$

and if $\frac{Y_s^k}{L_s} \in \mathfrak{L}^2(W)$, we have the result. Thus, our proof runs in three parts. First, we show that L almost surely has positive paths. Then we argue that $\frac{Y^k}{L} \in \mathfrak{L}^2(W)$. Finally, we prove the theorem using these facts.

Step 1: L has positive paths almost surely. To show that L almost surely has positive paths, define $\tau = \inf\{t \geq 0 | L_t = 0\}$. Let $t \geq 0$. Then $(\tau \leq t) \in \mathcal{F}_t$. Using that $\frac{\mathrm{d}Q_t}{\mathrm{d}P_t} = L_t$, we obtain

$$Q(\tau \leq t) = Q_t(\tau \leq t) = \int_{(\tau \leq t)} L_t \, \mathrm{d}P_t = \int_{(\tau \leq t)} L_t \, \mathrm{d}P.$$

Next using that $(\tau \leq t) \in \mathcal{F}_{\tau \wedge t}$ and that L is a P-martingale, we find

$$\int_{(\tau \leq t)} L_t \, \mathrm{d}P = \int_{(\tau \leq t)} E(L_t | \mathcal{F}_{\tau \wedge t}) \, \mathrm{d}P = \int_{(\tau \leq t)} L_{\tau \wedge t} \, \mathrm{d}P = \int_{(\tau \leq t)} L_\tau \, \mathrm{d}P = 0,$$

where we have used that whenever τ is finite, $L_\tau = 0$. This shows $Q(\tau \leq t) = 0$. Since P and Q are equivalent, we conclude $P(\tau \leq t) = 0$. Letting t tend to infinity, we obtain $P(\tau < \infty) = 0$. Therefore, P almost all paths of L are positive. By changing L on a set of measure zero, we can assume that all the paths of L are positive.

Step 2: Integrability. Next, we show that $\frac{Y_s^k}{L_s} \in \mathfrak{L}^2(W)$. Let τ_n be a localising sequence for Y^k. Define the stopping time σ_n by $\sigma_n = \inf\{t \geq 0 | L_t < \frac{1}{n}\}$. Since the paths of L are positive, σ_n tends to infinity. By $L^{\sigma_n} \geq \frac{1}{n}$, we obtain

$$\left(\left|\frac{Y_s^k}{L_s}\right|\right)^{\tau_n \wedge \sigma_n} \leq n(Y_s^k)^{\tau_n \wedge \sigma_n},$$

almost surely, so $(\frac{Y_s^k}{L_s})^{\tau_n \wedge \sigma_n}$ is in $\mathcal{L}^2(W)$, showing $\frac{Y_s^k}{L_s} \in \mathfrak{L}^2(W)$. Therefore, the process $M_t = \sum_{k=1}^n \int_0^t \frac{Y_s}{L_s} \, \mathrm{d}W_s^k$ is well-defined, and $L_t = \sum_{k=1}^n \int_0^t L_s \frac{Y_s}{L_s} \, \mathrm{d}W_s^k = \int_0^t L_s \, \mathrm{d}M_s$, as desired.

Step 3: Conclusion. Finally, by Lemma 3.7.1, we have $L = \mathcal{E}(M)$. Since L is uniformly integrable by Lemma 3.8.1 with limit $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty}$, we conclude $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty} = \mathcal{E}(M)_\infty$. □

We have now proved the first part of Girsanov's Theorem. Theorem 3.8.2 yields a concrete form for any measure Q which is equivalent to P on $\mathcal{F}_\infty$. The other half of Girsanov's Theorem uses this concrete form to describe the Q-distribution of W.

Theorem 3.8.3 (Girsanov's Theorem, part two). *Let Q be a measure on $\mathcal{F}$ equivalent to P on $\mathcal{F}_\infty$ with Radon-Nikodym derivative $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty} = \mathcal{E}(M)_\infty$, where M is in $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ such that $\mathcal{E}(M)$ is a uniformly integrable martingale. Let $M = Y \cdot W$, where $Y \in \mathfrak{L}^2(W)^n$. The process X in $\mathbb{R}^n$ given by $X_t^k = W_t^k - \int_0^t Y_s^k \, \mathrm{d}s$ is an n-dimensional $\mathcal{F}_t$-Brownian motion under Q.*

Proof. We first argue that the conclusion is well-defined. Let $Y \in \mathfrak{L}^2(W)^n$, and let τ_m be a localising sequence such that $Y[0, \tau_m] \in \mathcal{L}^2(W)^n$. Then $\int_0^t (Y^k[0, \tau_m]_s)^2 \, \mathrm{d}s$ is almost surely finite, and by Lemma B.2.4, $\int_0^t |Y^k[0, \tau_m]_s| \, \mathrm{d}s$ is almost surely finite. Thus, $Y^k \in \mathfrak{L}^1(t)$, and $\int_0^t Y_s^k \, \mathrm{d}s$ is well-defined. Therefore, X^k is well-defined and the conclusion is well-defined. Also note that since Y^k is progressive, X^k is adapted to $\mathcal{F}_t$. Therefore, to prove the theorem, it will suffice to show that $s \mapsto X_{t+s} - X_t$ is a Brownian motion for any $t \geq 0$.

To prove the result, we first consider the case where Y is bounded and zero from a deterministic point onwards. Afterwards, we will obtain the general case by approximation and localisation arguments.

Step 1: The dominated case, zero from a point onwards. First assume that there exists $T > 0$ such that Y_t is zero for $t \leq T$ and assume that Y is bounded. Having this condition will allow us to apply Lemma 3.7.3 in our calculations. To show that

X is an $\mathcal{F}_t$-Brownian motion under Q, it will suffice to show that for any $0 \le s \le t$,

$$E^Q \left(\exp\left(\sum_{k=1}^n \theta_k (X_t^k - X_s^k) \right) \middle| \mathcal{F}_s \right) = \exp\left(\frac{1}{2} \sum_{k=1}^n \theta_k^2 (t-s) \right).$$

In order to show this, let $f^k \in \mathcal{L}^2[0,\infty)$ and $A \in \mathcal{F}$. Put $Z_u^k = Y_u^k + f^k(u)$ and $N = Z \cdot W$, by Lemma 3.7.3 that the exponential martingale $\mathcal{E}(N)$ is a square-integrable martingale. In particular $\mathcal{E}(N)$ has an almost sure limit. We first prove

$$E^Q 1_A \exp\left(\sum_{k=1}^n \int_0^\infty f^k(u) \,\mathrm{d}X_u^k \right) = \exp\left(\frac{1}{2} \sum_{k=1}^n \int_0^\infty f^k(u)^2 \,\mathrm{d}u \right) E^P 1_A \mathcal{E}(N)_\infty,$$

and use this formula to show the conditional expectation formula above. Note that the stochastic integral above is well-defined in the sense that $X^k \in \mathcal{S}$ under P. We consider the left-hand side of the above. Note that from the condition on Y, we obtain $Y \in \mathcal{L}^2(W)^n$ and therefore $\int_0^\infty Y_u^k \,\mathrm{d}W_u^k$ and $\int_0^\infty (Y_u^k)^2 \,\mathrm{d}u$ are well-defined. We can therefore write

$$\begin{aligned}
& E^Q 1_A \exp\left(\sum_{k=1}^n \int_0^\infty f^k(u) \,\mathrm{d}X_u^k \right) \\
= \;& E^P 1_A \exp\left(\sum_{k=1}^n \int_0^\infty f^k(u) \,\mathrm{d}X_u^k \right) \mathcal{E}(M)_\infty \\
= \;& E^P 1_A \exp\left(\sum_{k=1}^n \int_0^\infty f^k(u) \,\mathrm{d}X_u^k + \sum_{k=1}^n \int_0^\infty Y_u^k \,\mathrm{d}W_u^k - \frac{1}{2} \sum_{k=1}^n \int_0^\infty (Y_u^k)^2 \,\mathrm{d}u \right).
\end{aligned}$$

We note that

$$\begin{aligned}
& \int_0^\infty f^k(u) \,\mathrm{d}X^k + \int_0^\infty Y_u^k \,\mathrm{d}W_u^k - \frac{1}{2} \int_0^t (Y_u^k)^2 \,\mathrm{d}u \\
= \;& \int_0^\infty f^k(u) \,\mathrm{d}W_u^k - \int_0^\infty f^k(u) Y_u^k \,\mathrm{d}u + \int_0^\infty Y_u^k \,\mathrm{d}W_u^k - \frac{1}{2} \int_0^t (Y_u^k)^2 \,\mathrm{d}u \\
= \;& \int_0^\infty f^k(u) + Y_u^k \,\mathrm{d}W_u^k + \frac{1}{2} \int_0^\infty f^k(u)^2 \,\mathrm{d}u - \frac{1}{2} \int_0^\infty (f^k(u) + Y_u^k)^2 \,\mathrm{d}u.
\end{aligned}$$

We then obtain

$$\begin{aligned}
& E^P 1_A \exp\left(\sum_{k=1}^n \int_0^\infty f^k(u)\,\mathrm{d}X_u^k + \sum_{k=1}^n \int_0^\infty Y_u^k\,\mathrm{d}W_u^k - \frac{1}{2}\sum_{k=1}^n \int_0^\infty (Y_u^k)^2\,\mathrm{d}u\right) \\
= \;& E^P 1_A \exp\left(\sum_{k=1}^n \int_0^\infty f^k(u) + Y_u^k\,\mathrm{d}W_u^k + \frac{1}{2}\int_0^\infty f^k(u)^2\,\mathrm{d}u - \frac{1}{2}\int_0^\infty (f^k(u)+Y_u^k)^2\,\mathrm{d}u\right) \\
= \;& E^P 1_A \exp\left(\frac{1}{2}\sum_{k=1}^n \int_0^\infty f^k(u)^2\,\mathrm{d}u\right)\mathcal{E}(N)_\infty \\
= \;& \exp\left(\frac{1}{2}\sum_{k=1}^n \int_0^\infty f^k(u)^2\,\mathrm{d}u\right) E^P 1_A \mathcal{E}(N)_\infty.
\end{aligned}$$

We have now proved that for any $f^k \in \mathcal{L}^2[0,\infty)$ and $A \in \mathcal{F}$,

$$E^Q 1_A \exp\left(\sum_{k=1}^n \int_0^\infty f^k(u)\,\mathrm{d}X_u^k\right) = \exp\left(\frac{1}{2}\sum_{k=1}^n \int_0^\infty f^k(u)^2\,\mathrm{d}u\right) E^P 1_A \mathcal{E}(N)_\infty.$$

We are ready to demonstrate that X is an $\mathcal{F}_t$-Brownian motion under Q. Let $0 \le s \le t$, $A \in \mathcal{F}_s$ and $\theta \in \mathbb{R}^n$. We can then define f^k by $f^k(u) = \theta_k 1_{(s,t]}(u)$. Since f^k is zero on $[0,s]$, $M_s = N_s$ and therefore, using that $\mathcal{E}(N)$ is a uniformly integrable martingale,

$$E^P 1_A \mathcal{E}(N)_\infty = E^P 1_A \mathcal{E}(N)_s = E^P 1_A \mathcal{E}(M)_s = Q(A).$$

We then obtain

$$\begin{aligned}
E^Q 1_A \exp\left(\sum_{k=1}^n \theta_k (X_t^k - X_s^k)\right) &= E^Q 1_A \exp\left(\sum_{k=1}^n \int_0^\infty f^k(u)\,\mathrm{d}X_u^k\right) \\
&= \exp\left(\frac{1}{2}\sum_{k=1}^n \int_0^\infty f^k(u)^2\,\mathrm{d}s\right) E^P 1_A \mathcal{E}(N)_\infty \\
&= \exp\left(\frac{1}{2}\sum_{k=1}^n \int_0^\infty f^k(u)^2\,\mathrm{d}s\right) Q(A) \\
&= \exp\left(\frac{1}{2}\sum_{k=1}^n \theta_k^2 (t-s)\right) Q(A).
\end{aligned}$$

This shows the desired formula,

$$E^Q\left(\exp\left(\sum_{k=1}^n \theta_k (X_t^k - X_s^k)\right)\middle|\,\mathcal{F}_s\right) = \exp\left(\frac{1}{2}\sum_{k=1}^n \theta_k^2 (t-s)\right).$$

Finally, we may conclude that X is an $\mathcal{F}_t$-Brownian motion under Q.

Step 2: The $\mathcal{L}^2(W)$ case, deterministic from a point onwards. Now only assume that $Y \in \mathcal{L}^2(W)^n$ and that Y is zero from T and onwards. In the previous step, we obtained our results by calculating the conditional Laplace transform. In the current case, we will instead calculate the conditional characteristic function. Our aim is to show that for any $0 \leq s \leq t$,

$$E^Q\left(\exp\left(i\sum_{k=1}^n \theta_k(X_t^k - X_s^k)\right)\middle|\mathcal{F}_s\right) = \exp\left(-\frac{1}{2}\sum_{k=1}^n \theta_k^2(t-s)\right).$$

We will use an approximation argument. To this end, define $Y_n^k = Y^k \wedge n \vee -n$. Y_n then satisfies the condition of the first step, and we may conclude that putting $M_n(t) = Y_n \cdot W$ and defining Q_n by $\frac{\mathrm{d}Q_n}{\mathrm{d}P_n} = \mathcal{E}(M_n)_\infty$, Q_n is a probability measure and X_n given by $X_n^k(t) = W_t^k - \int_0^t Y_n^k(s)\,\mathrm{d}s$ is a Brownian motion under Q_n.

Our plan is to show that

1. There is a subsequence of $\mathcal{E}(M_m)_\infty$ converging to $\mathcal{E}(M)_\infty$ in $\mathcal{L}^1$.
2. For $t \geq 0$, there is a subsequence of $X_m^k(t)$ converging to $X^k(t)$ almost surely.

We will be able to combine these facts to obtain our desired result.

To prove the first of the two convergence results above, note that Y_n^k tends to Y^k in $\mathcal{L}^2(W)$ by dominated convergence. Therefore, $\int_0^\infty Y_n^k(s)\,\mathrm{d}W_s^k$ tends to $\int_0^\infty Y^k(s)\,\mathrm{d}W_s^k$ in $\mathcal{L}^2$. By taking out a subsequence, we can assume that the convergence is almost sure. Furthermore, from Lemma B.6.5 we find that by taking a subsueqence, we can assume that $\int_0^\infty Y_n^k(s)^2\,\mathrm{d}s$ tends to $\int_0^\infty Y^k(s)^2\,\mathrm{d}s$ almost surely. We have now ensured that almost surely, $\int_0^\infty Y_n^k(s)\,\mathrm{d}W_s^k$ tends to $\int_0^\infty Y^k(s)\,\mathrm{d}W_s^k$, and $\int_0^\infty Y_n^k(s)^2\,\mathrm{d}s$ tends to $\int_0^\infty Y^k(s)^2\,\mathrm{d}s$. Therefore, $\mathcal{E}(M_n)_\infty$ converges almost surely to $\mathcal{E}(M)_\infty$. We know that $E^P\mathcal{E}(M_n)_\infty = 1$, and have assumed $E^P\mathcal{E}(M)_\infty = 1$. By nonnegativity, Scheffé's Lemma B.2.3 yields that $\mathcal{E}(M_n)_\infty$ converges in $\mathcal{L}^1$ to $\mathcal{E}(M)_\infty$, as desired.

To prove the second convergence result, let $t \geq 0$ be given. Note that since Y_n^k converges in $\mathcal{L}^2(W)$ to Y^k and the processes are zero from a deterministic point onwards, we obtain from Lemma B.2.4 that $E\int_0^\infty |Y_n^k(s) - Y^k(s)|\,\mathrm{d}s$ tends to zero. Therefore, by Lemma B.6.5, we can by taking out a subsequence assume that $\int_0^t Y_n^k(s)\,\mathrm{d}s$ converges almost surely to $\int_0^t Y^k(s)\,\mathrm{d}s$. We then obtain

$$\lim_n X_n^k(t) = \lim_n W^k(t) - \int_0^t Y_n^k(s)\,\mathrm{d}s = X^k(t)$$

as desired.

Now let $0 \leq s \leq t$. We select a subsequence such that

1. $\mathcal{E}(M_m)_\infty$ converges to $\mathcal{E}(M)_\infty$ in $\mathcal{L}^1$.
2. $X^k_m(s)$ converges almost surely to $X^k(s)$.
3. $X^k_m(t)$ converges almost surely to $X^k(t)$.

Since X_m is $\mathcal{F}_t$-Brownian motion under Q_m, we know that

$$E^{Q_m}\left(\exp\left(i\sum_{k=1}^n \theta_k(X^k_m(t) - X^k_m(s))\right)\middle| \mathcal{F}_s\right) = \exp\left(-\frac{1}{2}\sum_{k=1}^n \theta_k^2(t-s)\right).$$

We will identify the limit of the left-hand side. We know that

$$\begin{aligned} & E^{Q_m}\left(\exp\left(i\sum_{k=1}^n \theta_k(X^k_m(t) - X^k_m(s))\right)\middle| \mathcal{F}_s\right) \\ = \quad & E^P\left(\exp\left(i\sum_{k=1}^n \theta_k(X^k_m(t) - X^k_m(s))\right)\mathcal{E}(M_m)_\infty\middle| \mathcal{F}_s\right). \end{aligned}$$

Using our convergence results with Lemma B.2.2, we find that

$$\exp\left(i\sum_{k=1}^n \theta_k(X^k_m(t) - X^k_m(s))\right)\mathcal{E}(M_m)_\infty \xrightarrow{\mathcal{L}^1} \exp\left(i\sum_{k=1}^n \theta_k(X^k(t) - X^k(s))\right)\mathcal{E}(M)_\infty,$$

and therefore, by the $\mathcal{L}^1$ continuity of conditional expectations,

$$E^{Q}\left(\exp\left(i\sum_{k=1}^n \theta_k(X^k_t - X^k_s)\right)\middle| \mathcal{F}_s\right) = \exp\left(-\frac{1}{2}\sum_{k=1}^n \theta_k^2(t-s)\right),$$

as desired.

Step 3: The $\mathfrak{L}^2(W)$ case. Finally, merely assume $Y \in \mathfrak{L}^2(W)^n$. Let σ_n be a localising sequence such that $Y^k[0,\sigma_n] \in \mathcal{L}^2(W)$. Put $\tau_n = \sigma_n \wedge n$ and define $Y^k_n = Y^k[0,\tau_n]$. Then Y^k is in $\mathcal{L}^2(W)$ and is zero from a deterministic point onwards. Therefore, we can use the result from the previous step on Y_n. As in the preceeding step of the proof, we define $M_m = Y_m \cdot W$ and define Q_m by $\frac{\mathrm{d}Q_m}{\mathrm{d}P_m} = \mathcal{E}(M_m)_\infty$. Putting $X^k_m(t) = W^k(t) - \int_0^t Y^k_m(s)\,\mathrm{d}s$, X_m is then a $\mathcal{F}_t$ Brownian motion under Q_m. We will show

1. $\mathcal{E}(M_m)_\infty$ converges to $\mathcal{E}(M)_\infty$ in $\mathcal{L}^1$.

2. $X_m^k(t)$ converges to $X^k(t)$ almost surely.

Having these two facts, the conclusion will follow by precisely the same method as in the previous step. To prove the first convergence result, note that $\mathcal{E}(M)^{\tau_m} = \mathcal{E}(M_m)$. Since $\mathcal{E}(M)$ is assumed to be a uniformly integrable martingale, it has an almost surely limit and we may conclude $\mathcal{E}(M)_{\tau_m} = \mathcal{E}(M_m)_\infty$. Now, $(\mathcal{E}(M)_{\tau_m})_{m\geq 1}$ is a discrete-time martingale, and it is closed by $\mathcal{E}(M)_\infty$. Therefore, it is convergent in $\mathcal{L}^1$. Thus $\mathcal{E}(M_m)_\infty \xrightarrow{\mathcal{L}^1} \mathcal{E}(M)_\infty$.

The second result follows directly from

$$
\begin{aligned}
\lim_m X_m^k(t) &= \lim_m W_t^k - \int_0^t Y_m^k(s)\,\mathrm{d}s \\
&= W_t^k - \lim_m \int_0^t Y^k(s) 1_{[0,\tau_m]}(s)\,\mathrm{d}s \\
&= W_t^k - \int_0^t Y^k(s)\,\mathrm{d}s \\
&= X_t^k,
\end{aligned}
$$

almost surely. As in the previous step, we conclude that X is an $\mathcal{F}_t$ Brownian motion under Q. $\square$

We are now done with the proof of Girsanov's Theorem. In the first part of the theorem, Theorem 3.8.2, we have seen that any probability measure Q equivalent to P on $\mathcal{F}_\infty$ has the property that $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty} = \mathcal{E}(M)_\infty$ for some M such that $\mathcal{E}(M)$ is uniformly integrable. In the second part, Theorem 3.8.3, we have seen now the distribution of W changes under this measure change.

When we have two equivalent probability measures P and Q, we will usually be interested in stochastic integration under both of these probability measures, in the sense that a given process can be a standard process under P with respect to the original Brownian motion, and it can be a standard process under Q with respect to the new Q Brownian motion. We have defined a stochastic integral for each of these cases. We will now check that whenever applicable, the two stochastic integrals agree. We also show that being a standard process under P is the same as being a standard process under Q.

Theorem 3.8.4. *Let P and Q be equivalent on $\mathcal{F}_\infty$. If X is a standard process under P, it is a standard process under Q as well. Let $L^P(X)$ denote the integrands for X under P, and let $L^Q(X)$ denote the integrands for X under Q. If $Y \in L^P(X) \cap L^Q(X)$, then the integrals under P and Q agree up to indistinguishability.*

Comment 3.8.5 The meaning of the lemma is a bit more convoluted that it might seem at first. To be a standard process, it is necessary to work in the context of a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ with a $\mathcal{F}_t$ Brownian motion. That X is a standard process under Q thus assumes that we have some particular process in mind which is a $\mathcal{F}_t$ Brownian motion under Q. In our case, the implicitly understood Brownian motion is the one given by Girsanov's theorem. ∘

Proof. We need to prove two things: That X is a standard process under Q, and that if $\in L^P(X) \cap L^Q(X)$, the integrals agree.

Step 1: X is a standard process under Q. Let the canonical decomposition of X under P be given by $X = A + M$ with $M = Y \cdot W$. By Girsanov's Theorem, there is a continuous local martingale $N = Z \cdot W$ such that $\mathcal{E}(N)$ is a uniformly integrable martingale and $\frac{\mathrm{d}Q_\infty}{\mathrm{d}P_\infty} = \mathcal{E}(N)_\infty$, and then the process W^Q given by

$$W_k^Q(t) = W_k(t) - \int_0^t Z_k(s)\,\mathrm{d}s$$

is a $\mathcal{F}_t$ Brownian motion under Q. Note that since the null sets of P and Q are the same, the space $(\Omega, \mathcal{F}, Q, \mathcal{F}_t)$ satisfies the usual conditions. Thus, with W^Q as our Brownian motion, we have a setup precisely as described throughout the chapter. And by the properties of the stochastic integral under P,

$$\begin{aligned} X_t &= A_t + M_t \\ &= A_t + \sum_{k=1}^n \int_0^t Y_k(s)\,\mathrm{d}W_k(s) \\ &= A_t + \sum_{k=1}^n \int_0^t Z_k(s)\,\mathrm{d}s + \sum_{k=1}^n \int_0^t Y_s^k\,\mathrm{d}W_k^Q(s). \end{aligned}$$

Under the setup $(\Omega, \mathcal{F}, Q, \mathcal{F}_t)$, the process $A_t + \sum_{k=1}^n Z_k(s)\,\mathrm{d}s$ is of finite variation and the process $\sum_{k=1}^n Y_s^k\,\mathrm{d}W_k^Q(s)$ is in $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$. Therefore, X is a standard process under Q.

Step 2: Agreement if the integrals, part 1. Now assume that Y is some process with is in $L^P(X) \cap L^Q(X)$. Let $I_X^P(Y)$ be the integral of Y with respect to X under

P, and let $I_X^Q(Y)$ be the integral of Y with respect to X under Q. We want to show that $I_X^P(Y)$ and $I_X^Q(Y)$ are indistinguishable.

We first consider the case where Y is bounded and continuous. Let $t \geq 0$. We will show that $I_X^P(Y)_t = I_X^Q(Y)_t$ almost surely. By Lemma 3.4.15, the Riemann sums $\sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} - X_{t_{k-1}})$ converge in probability under P to $I_X^P(Y)_t$ and converge in probability under Q to $I_X^Q(Y)_t$ for some sequence of partitions of $[0,t]$ with meshes tending to zero. By taking out a subsequence, we can assume that the convergence is almost sure. Since P and Q are equivalent, we find that both under P and Q, the Riemann sums tend to both $I_X^P(Y)_t$ and $I_X^Q(Y)_t$. Therefore, $I_X^P(Y)_t$ and $I_X^Q(Y)_t$ are almost surely equal. Since both processes are continuous, we conclude that $I_X^P(Y)$ and $I_X^Q(Y)$ are indistinguishable.

Step 3: Agreement of the integrals, part 2. Now, as before, let $X = A + M$ be the canonical decomposition of X under P, and let $X = B + N$ be the canonical decomposition of X under Q. We consider the case where V_A and V_B are bounded, where $M, N \in \mathbf{c}\mathcal{M}_W^2$ and where Y is bounded and zero from a deterministic point onwards. As in the proof of Theorem 3.1.7, define

$$Y_n(t) = \frac{1}{\frac{1}{n}} \int_{(t-\frac{1}{n})^+}^t Y(s)\,\mathrm{d}s.$$

Then Y_n is bounded and continuous with $\|Y_n\|_\infty \leq \|Y\|_\infty$, and Y_n converges to Y $\lambda \otimes P$ everywhere. Since μ_M and μ_N are bounded, this implies that Y_n converges to Y in $\mathcal{L}^2(M)$ and $\mathcal{L}^2(N)$, and therefore $I_M^P(Y_n)$ converges to $I_M^P(Y)$ in $\mathbf{c}\mathcal{M}_0^2$ under P. Likewise, $I_N^Q(Y_n)$ converges to $I_N^Q(Y)$ under Q. Taking out subsequences, we can assume that the convergence is uniformly almost sure, showing in particular that $I_M^P(Y_n)_t$ converges to $I_M^P(Y)_t$ almost surely and $I_N^Q(Y_n)_t$ converges to $I_N^Q(Y)_t$ almost surely for any $t \geq 0$.

Next, consider the finite variation components. Since V_A and V_B are bounded we find that for almost all $t \geq 0$, $I_A^P(Y_n)_t$ converges almost surely to $I_A^P(Y)_t$. Likewise, for almost all $t \geq 0$, $I_B^Q(Y_n)_t$ converges almost surely to $I_B^Q(Y)_t$.

All in all, we may conclude that for almost all $t \geq 0$, $I_X^P(Y_n)_t$ converges to $I_X^P(Y)_t$ and $I_X^Q(Y_n)_t$ converges to $I_X^Q(Y)_t$ almost surely. Therefore, by what we already have shown

$$I_X^P(Y)_t = \lim_n I_X^P(Y_n)_t = \lim_n I_X^Q(Y_n)_t = I_X^Q(Y)_t.$$

Since any Lebesgue almost sure set is dense and the processes are continuous, we conclude that $I^P_X(Y)$ and $I^Q_X(Y)$ are indistinguishable.

Step 4: Agreement of the integrals, part 3. We now still assume that V_A and V_B are bounded and that $M,N \in \mathbf{c}\mathcal{M}^2_W$, but we reduce our requirement on Y to be that Y is in $\mathcal{L}^1(A)\cap\mathcal{L}^1(B)\cap\mathcal{L}^2(M)\cap\mathcal{L}^2(N)$. We will use the results of the previous step to show that the result holds in this case. Define

$$Y_n(t) = Y(t)1_{(|Y_t|>n)}1_{[0,n]}(t).$$

Then Y is bounded and zero from a deterministic point onwards. Furthermore, $|Y_n(t)| \le |Y(t)|$, and Y_n converges pointwise to Y. By dominated convergence, we therefore conclude that Y_n converges to Y in $\mathcal{L}^2(M)$ and $\mathcal{L}^2(N)$. This implies that $I^P_M(Y_n)$ tends to $I^P_M(Y)$ in $\mathbf{c}\mathcal{M}^2_0$ under P, and $I^Q_N(Y_n)$ tends to $I^Q_N(Y)$ in $\mathbf{c}\mathcal{M}^2_0$ under Q. For the finite variation compoments, we find that $I^P_A(Y_n)$ converges pointwise to $I^P_A(Y)$ and $I^Q_A(Y_n)$ converges pointwise to $I^Q_A(Y)$.

As in the previous step, this implies

$$I^P_X(Y)_t = \lim_n I^P_X(Y_n)_t = \lim_n I^Q_X(Y_n)_t = I^Q_X(Y)_t,$$

showing the result in this case.

Step 5: Agreement of the integrals, part 4. Finally, we consider the general case. Define the stopping times

$$\begin{aligned}\tau^A_n &= \inf\{t \ge 0 | V_A(t) \ge n\}\\ \tau^B_n &= \inf\{t \ge 0 | V_A(t) \ge n\},\end{aligned}$$

and let τ^M_n and τ^N_n be sequences localising M and N to $\mathbf{c}\mathcal{M}^2_W$, respectively. Let σ_n be such that Y^{σ_n} is in $\mathcal{L}^1(A)\cap\mathcal{L}^1(B)\cap\mathcal{L}^2(M)\cap\mathcal{L}^2(N)$. Defining the stopping time τ_n by $\tau_n = \tau^A_n \wedge \tau^B_n \wedge \tau^M_n \wedge \tau^N_n \wedge \sigma_n$. Then$V^{\tau_n}_A$ and $V^{\tau_n}_B$ are bounded, M^{τ_n} and N^{τ_n} are in $\mathbf{c}\mathcal{M}^2_W$ and Y^{τ_n} is in $\mathcal{L}^1(A^{\tau_n})\cap\mathcal{L}^1(B^{\tau_n})\cap\mathcal{L}^2(M^{\tau_n})\cap\mathcal{L}^2(N^{\tau_n})$. Furthermore, τ_n tends to infinity almost surely. By Lemma 3.4.11,

$$\begin{aligned}I^P_X(Y)^{\tau_n} &= I^P_{X^{\tau_n}}(Y)\\ I^Q_X(Y)^{\tau_n} &= I^Q_{X^{\tau_n}}(Y).\end{aligned}$$

Since P and Q are equivalent, τ_n converges to infinity both Q and P almost surely. Therefore, since we have shown the result in the localised case, the result follows for the general case as well. □

Next, we consider necessary requirements for the practical application of the Girsanov theorem. In practice, we are not given a probability measure Q and want to identify the distribution of W under Q. Rather, we desire to change the distribution of W in a certain manner, and wish to find an equivalent probability measure Q to obtain this change. If we plan to use Theorem 3.8.3 to facilitate this change, we need to identify a local martingale M zero at zero yielding the desired distribution and ensure that $\mathcal{E}(M)$ is a uniformly integrable martingale. This latter part is the most difficult part of the plan. Note that the condition $M_0 = 0$ is necessary for $\mathcal{E}(M)$ to be defined at all.

In Lemma 3.7.3, we obtained a sufficient condition to ensure that $\mathcal{E}(M)$ is uniformly integrable. This condition is too strong to be useful in practice, however. Our next objective is to find weaker conditions that ensure the uniformly integrable martingale property of $\mathcal{E}(M)$. These conditions take the form of Kazamaki's and Novikov's criteria. We are led by Protter (2005), Section III.8.

Lemma 3.8.6. *Let $M \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$. $\mathcal{E}(M)$ is a supermartingale with $E\mathcal{E}(M)_t \leq 1$ for all $t \geq 0$. $\mathcal{E}(M)$ is a martingale if and only if $E\mathcal{E}(M)_t = 1$ for all $t \geq 0$. $\mathcal{E}(M)$ is a uniformly integrable martingale if and only if $E\mathcal{E}(M)_\infty = 1$.*

Proof. Since $\mathcal{E}(M)$ is a continuous nonnegative local martingale, it is a supermartingale. In particular, $E\mathcal{E}(M)_t \leq E\mathcal{E}(M)_0 = 1$. Furthermore, $\mathcal{E}(M)_\infty$ always exists as an almost sure limit by the Martingale Convergence Theorem.

The martingale criterion. We now prove that $\mathcal{E}(M)$ is a martingale if and only if $E\mathcal{E}(M)_t = 1$ for $t \geq 0$. Clearly, if $\mathcal{E}(M)$ is a martingale, $E\mathcal{E}(M)_t = E\mathcal{E}(M)_0 = 1$ for all $t \geq 0$. Conversely, assume that $E\mathcal{E}(M)_t = 1$ for all $t \geq 0$. Consider $0 \leq s \leq t$. By the supermartingale property, we obtain $\mathcal{E}(M)_s \geq E(\mathcal{E}(M)_t|\mathcal{F}_s)$. This shows that

$$0 \leq E(\mathcal{E}(M)_s - E(\mathcal{E}(M)_t|\mathcal{F}_s)) = E\mathcal{E}(M)_s - E\mathcal{E}(M)_t = 0,$$

so $\mathcal{E}(M)_s - E(\mathcal{E}(M)_t|\mathcal{F}_s)$ is a nonnegative variable with zero mean. Thus, it is zero almost surely and $\mathcal{E}(M)_s = E(\mathcal{E}(M)_t|\mathcal{F}_s)$, showing the martingale property.

The UI martingale criterion. Next, we prove that $\mathcal{E}(M)$ is a uniformly integrable martingale if and only if $E\mathcal{E}(M)_\infty = 1$. If $E\mathcal{E}(M)_\infty = 1$, we obtain

$$1 = E\mathcal{E}(M)_\infty = E \liminf_t \mathcal{E}(M)_t \leq \liminf_t E\mathcal{E}(M)_t.$$

Now, $E\mathcal{E}(M)_t$ is decreasing and bounded by one, so we conclude $\lim_t E\mathcal{E}(M)_t = 1$ and therefore $E\mathcal{E}(M)_t = 1$ for all t. Therefore, by what we already have shown, $\mathcal{E}(M)$ is a martingale. By Lemma B.5.4, it is uniformly integrable.

On the other hand, if $\mathcal{E}(M)$ is a uniformly integrable martingale, it is convergent in $\mathcal{L}^1$ and in particular $E\mathcal{E}(M)_\infty = \lim_t E\mathcal{E}(M)_t = 1$. □

Lemma 3.8.7. *Let M be a continuous local martingale zero at zero. Let $\varepsilon > 0$. If it holds that $\sup_\tau E\exp\left(\left(\frac{1}{2}+\varepsilon\right)M_\tau\right) < \infty$, where the supremum is over all bounded stopping times, then $\mathcal{E}(M)$ is a uniformly integrable martingale.*

Proof. Let $\varepsilon > 0$ be given. Our plan is to identify an $a > 1$ such that $\sup_\tau E\mathcal{E}(M)_\tau^a$ is finite, where the supremum is over all bounded stopping times. If we can do this, Lemma 2.5.19 shows that $\mathcal{E}(M)$ is a martingale bounded in $\mathcal{L}^a$, in particular $\mathcal{E}(M)$ is a uniformly integrable martingale. To this end, let $a, r > 1$ be given. We then have

$$\mathcal{E}(M)_\tau^a = \exp\left(aM_\tau - \frac{a}{2}[M]_\tau\right) = \exp\left(\sqrt{\frac{a}{r}}M_\tau - \frac{a}{2}[M]_\tau\right)\exp\left(\left(a - \sqrt{\frac{a}{r}}\right)M_\tau\right).$$

The point of this observation is that when we raise the first factor in the right-hand side above to the r'th power, we obtain an exponential martingale. Therefore, letting s be the dual exponent of r, Hölder's inequality and Lemma 3.8.6 yields

$$\begin{aligned}
E\mathcal{E}(M)_\tau^a &\leq \left(E\exp\left(\sqrt{ar}M_\tau - \frac{ar}{2}[M]_\tau\right)\right)^{\frac{1}{r}}\left(E\exp\left(\left(a-\sqrt{\frac{a}{r}}\right)sM_\tau\right)\right)^{\frac{1}{s}} \\
&= \left(E\mathcal{E}(\sqrt{ar}M)_\tau\right)^{\frac{1}{r}}\left(E\exp\left(\left(a-\sqrt{\frac{a}{r}}\right)sM_\tau\right)\right)^{\frac{1}{s}} \\
&\leq \left(E\exp\left(\left(a-\sqrt{\frac{a}{r}}\right)sM_\tau\right)\right)^{\frac{1}{s}}.
\end{aligned}$$

Recalling that $s = \frac{r}{r-1}$, we then find

$$\begin{aligned}
\sup_\tau E\mathcal{E}(M)_\tau^a &\leq \left(\sup_\tau E\exp\left(\left(a-\sqrt{\frac{a}{r}}\right)\frac{r}{r-1}M_\tau\right)\right)^{\frac{1}{s}} \\
&= \left(\sup_\tau E\exp\left(\left(ar-\sqrt{ar}\right)\frac{1}{r-1}M_\tau\right)\right)^{\frac{1}{s}}.
\end{aligned}$$

Thus, if we can identify $a, r > 1$ such that $(ar-\sqrt{ar})\frac{1}{r-1} \leq \frac{1}{2}+\varepsilon$, the result will follow from our assumption. To this end, define $f(y,r) = (y-\sqrt{y})\frac{1}{r-1}$. Finding $a, r > 1$ such that $(ar-\sqrt{ar})\frac{1}{r-1} \leq \frac{1}{2}+\varepsilon$ is equivalent to finding $y > r > 1$ such that $f(y,r) \leq \frac{1}{2}+\varepsilon$.

To this end, note that $\frac{d}{dy}(y-\sqrt{y}) = 1-\frac{1}{2\sqrt{y}}$, which is positive whenever $y > \frac{1}{4}$. We therefore obtain

$$\inf_{y>r>1} f(y,r) = \inf_{r>1}\inf_{y>r} f(y,r) = \inf_{r>1}\frac{r-\sqrt{r}}{r-1} = \inf_{r>1}\frac{\sqrt{r}}{\sqrt{r}+1} = \inf_{r>1}\frac{1}{1+\frac{1}{\sqrt{r}}} = \frac{1}{2}.$$

We may now conclude that for any $\varepsilon > 0$, there is $y > r > 1$ such that $f(y,r) \le \frac{1}{2}+\varepsilon$. The proof of the lemma is then complete. □

Comment 3.8.8 This lemma parallels the lemma preceeding Theorem 44 of Section III.8 in Protter (2005). However, Protter avoids the analysis of the mapping f used in the proof above by formulating his lemma differently and making some very inspired choices of exponents. He shows that if $\sup_\tau E\exp(\frac{\sqrt{p}}{2\sqrt{p}-1}M_\tau)$ is finite, then $\mathcal{E}(M)$ is a martingale bounded in $\mathcal{L}^q$, where q is the dual exponent to p. His method is based on Hölder's inequality just as our proof, and he uses $r = \frac{\sqrt{p}+1}{\sqrt{p}-1}$.

We have opted for a different proof, involving analysis of the mapping f, in order to avoid having to use the very original choice $r = \frac{\sqrt{p}+1}{\sqrt{p}-1}$ out of the blue. ◦

Theorem 3.8.9 (Kazamaki's Criterion). *Let M be a continuous local martingale zero at zero. If $\sup_\tau E\exp\left(\frac{1}{2}M_\tau\right) < \infty$, where the supremum is over all bounded stopping times, then $\mathcal{E}(M)$ is a uniformly integrable martingale.*

Proof. Our plan is to show that $\mathcal{E}(aM)$ is a uniformly integrable martingale for any $0 < a < 1$ and let a tend to one in a suitable context. Let $0 < a < 1$. Then there is $\varepsilon_a > 0$ such that $a(\frac{1}{2}+\varepsilon_a) \le \frac{1}{2}$. Therefore, with the supremum over bounded stopping times,

$$\begin{aligned}
\sup_\tau E\exp\left(\left(\frac{1}{2}+\varepsilon_a\right)aM_\tau\right) &\le \sup_\tau E\left(1_{(M_\tau<0)} + 1_{(M_\tau\ge 0)}\exp\left(\left(\frac{1}{2}+\varepsilon_a\right)aM_\tau\right)\right)\\
&\le 1+\sup_\tau E1_{(M_\tau\ge 0)}\exp\left(\frac{1}{2}M_\tau\right)\\
&\le 1+\sup_\tau E\exp\left(\frac{1}{2}M_\tau\right),
\end{aligned}$$

which is finite by assumption. Lemma 3.8.7 then yields that $\mathcal{E}(aM)$ is a uniformly integrable martingale.

Next, note that there is $C > 0$ such that $x^+ \le C + \exp(\frac{1}{2}x)$. Therefore, by our assumptions, M_t^+ is bounded in $\mathcal{L}^1$, and therefore M_t converges almost surely to its limit M_∞. Since the limit $[M]_\infty$ always exists in $[0,\infty]$, we can then write

$$\begin{aligned}\mathcal{E}(aM)_\infty &= \exp\left(aM_\infty - \frac{a^2}{2}[M]_\infty\right)\\ &= \exp\left(a^2M_\infty - \frac{a^2}{2}[M]_\infty\right)\exp\left((a-a^2)M_\infty\right)\\ &= \mathcal{E}(M)^{a^2}\exp\left((a-a^2)M_\infty\right).\end{aligned}$$

Using that $\mathcal{E}(aM)$ is a uniformly integrable martingale and applying Hölder's inequality with $\frac{1}{a^2}$ and its dual exponent $\frac{1}{1-a^2}$, we obtain by Lemma 3.8.6 that

$$\begin{aligned}1 &= E\mathcal{E}(aM)_\infty\\ &\le (E\mathcal{E}(M))^{a^2}\left(E\exp\left(\frac{a-a^2}{1-a^2}M_\infty\right)\right)^{1-a^2}.\end{aligned}$$

Now, if the last factor on the right-hand side is bounded over $a > 1$, we can let a tend to 1 from above and obtain $1 \le E\mathcal{E}(M)_\infty$, which would give us the desired conclusion. Since $\frac{a-a^2}{1-a^2} = \frac{a}{1+a} = \frac{1}{\frac{1}{a}+1} < \frac{1}{2}$, it will suffice to prove that $E\exp(\frac{a}{2}M_\infty)$ is bounded over $0 < a < 1$. This is what we now set out to do.

Define $C = \sup_\tau E\exp(\frac{1}{2}M_\tau)$, C is finite by assumption. Note that since $a < 1$, there is $p > 1$ such that $ap \le 1$ and therefore

$$\sup_{t\ge 0} E\exp\left(\frac{a}{2}M_t\right)^p = \sup_{t\ge 0} E\exp\left(\frac{ap}{2}M_t\right) \le C.$$

This shows that $\exp\left(\frac{a}{2}M_t\right)$ is bounded in $\mathcal{L}^p$ for $t \ge 0$. In particular, the family is uniformly integrable. Since M_t converges almost surely to M_∞, we have convergence in probability as well. Therefore, $\exp(\frac{a}{2}M_t)$ converges in $\mathcal{L}^1$ to $\exp(\frac{a}{2}M_\infty)$, and we may conclude

$$E\exp\left(\frac{a}{2}M_\infty\right) = \lim_t E\exp\left(\frac{a}{2}M_t\right) \le C.$$

In particular, then, we finally get $1 \le (E\mathcal{E}(M)_\infty)^{a^2} C^{1-a^2}$, and letting a tend to one yields $1 \le E\mathcal{E}(M)_\infty$. By Lemma 3.8.6, the proof is complete. □

Theorem 3.8.10 (Novikov's Criterion). *Let M be a continuous local martingale zero at zero. Assume that $E\exp(\frac{1}{2}[M]_\infty)$ is finite. Then $\mathcal{E}(M)$ is a uniformly integrable martingale.*

Proof. We will show that Novikov's criterion implies Kazamaki's criterion. Let τ be any bounded stopping time. Then

$$\mathcal{E}(M)_\tau^{\frac{1}{2}} = \exp\left(\frac{1}{2}M_\tau\right)\left(\exp\left(-\frac{1}{2}[M]_\tau\right)\right)^{\frac{1}{2}},$$

which, using the Cauchy-Schwartz inequality, shows that

$$
\begin{aligned}
E \exp\left(\frac{1}{2}M_\tau\right) &= E\mathcal{E}(M)_\tau^{\frac{1}{2}}\left(\exp\left(\frac{1}{2}[M]_\tau\right)\right)^{\frac{1}{2}} \\
&\leq (E\mathcal{E}(M)_\tau)^{\frac{1}{2}}\left(E\exp\left(\frac{1}{2}[M]_\tau\right)\right)^{\frac{1}{2}} \\
&\leq \left(E\exp\left(\frac{1}{2}[M]_\infty\right)\right)^{\frac{1}{2}},
\end{aligned}
$$

where we have used that $[M]$ is increasing. Since the final right-hand side is independent of τ and finite by assumption, Kazamaki's criterion is satisfied and the conclusion follows from Theorem 3.8.9. □

This concludes our work on the stochastic calculus. In this section, we have proven Girsanov's theorem, characterising the possible measure changes on $\mathcal{F}_\infty$ and giving the change in drift of the Brownian motion. The Kazamaki and Novikov criteria can be used to obtain the existence of measure changes corresponding to specific drifts. Of these two, the Kazamaki criterion is the strongest, but the Novikov criterion is often easier to apply.

Before proceeding to the next chapter, we will in the next section discuss our choice to assume the usual conditions.

3.9 Why The Usual Conditions?

The question of whether to assume the usual conditions is a vexed one. Most books either do not comment the problem or choose to assume the usual conditions without stating why. We have assumed the usual conditions in this text. We will now explain this choice. We will also describe how we could have avoided having to make this assumption, and we will discuss general pros and cons of the usual conditions.

A review of the usual conditions. We begin by reiterating the meaning of the usual conditions and reviewing some opinions expressed about the usual conditions in the literature. Recall that for a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$, the usual conditions are conditions states that:

1. $(\Omega, \mathcal{F}, P)$ is a complete measure space.
2. The filtration is right-continuous, $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$.
3. For $t \geq 0$, $\mathcal{F}_t$ contains all P-null sets in $\mathcal{F}$.

The books Rogers & Williams (2000a) and Protter (2005) both have sections discussing the effects and the reasonableness of assuming the usual conditions.

In Rogers & Williams (2000a), Section II.76, the right-continuity of the filtration is taken as a natural choice. The authors focus on why the completion of the filtration is to be introduced. The main argument is that this is necessary for the Début Theorem and Section Theorem of Section II.77 to hold. Of these two, the Début Theorem is the easiest to understand, simply stating that for any progressive process X and Borel set B, the variable $D_B = \inf\{t \geq 0 | X \in B\}$ is a stopping time. This is not necessarily true if the filtration is not completed. The reason for this is illustrated in the proof of Lemma 75.1, where it is explained how the process can approach B in uncountably many ways, which causes problems with measurability.

In Protter (2005), Section 1.5, the completion of the filtration is instead seen as a natural choice, and the author focuses on why the right-continuity should be included. He does this by showing that the completed filtration of any cadlag Feller process is automatically right-continuous, so when assuming completion, right-continuity comes naturally.

While enlightening, none of these discussions provide reasons for us to assume the usual filtrations. Our reasons for assuming the usual conditions are much more down-to-earth and actually motivated from mostly practical concerns. Before discussing upsides and downsides from assuming the usual conditions in our case, we make some general observations regarding the effects of the usual conditions.

Regularity of paths and the usual conditions. We take as starting point the topic of cadlag versions of martingales. We claim that

1. Without any assumptions on the filtration, no cadlag versions are guaranteed.
2. With right-continuity, almost sure cadlag paths can be obtained.
3. With right-continuity and completeness, cadlag paths for all ω can be obtained.

The first claim is basically unfounded, as we have no counterexample to show that there are martingales where no cadlag versions exists. However, there are no theorems in the literature to be found giving results on general filtrations, and from the structure of the proofs for right-continuous filtrations, it seems highly unlikely that cadlag versions should exist for the general case.

To understand what is going on in the two other cases, we follow Rogers & Williams (2000a), sections II.62 through II.67. We work in a probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ where we assume that the filtration is right-continuous. Let $x : [0, \infty) \to \mathbb{R}$ be some mapping. We define $\lim_{q\downarrow\downarrow t} x_q$ as the limit of x_q as q converges downwards to t through the rationals. Likewise, we define $\lim_{q\uparrow\uparrow t} x_q$ as the limit of x_q as q converges upwards to t through the rationals. We say that x is regularisable if $\lim_{q\downarrow\downarrow t} x_q$ exists for all $t \geq 0$ and $\lim_{q\uparrow\uparrow t} x_q$ exists for all $t > 0$. As described in Rogers & Williams (2000a), Theorem II.62.13, if x is regularisable, there exists another function y such that y is cadlag and y agrees with x on $[0, \infty) \cap \mathbb{Q}$. As shown in Sokol (2007), Lemma 2.73, being regularisable is actually both a sufficient and necessary condition for this to hold. y is explicitly given by $y_t = \lim_{q\downarrow\downarrow t} x_t$.

Now let M be a martingale and define

$$G = \{\omega \in \Omega | M(\omega) \text{ is regularisable}\}.$$

By Rogers & Williams (2000a), Theorem II.65.1, G is then $\mathcal{F}$-measurable and it holds that $P(G) = 1$. We define

$$Y_t(\omega) = \begin{cases} \lim_{q\downarrow\downarrow t} M_t(\omega) & \text{if the limit exists.} \\ 0 & \text{otherwise.} \end{cases}$$

By a modification of the argument in Rogers & Williams (2000a), Theorem II.67.7, if the filtration is right-continuous, Y is then a version of M. Note that

$$G \subseteq (Y_t = \lim_{q\downarrow\downarrow t} M_t \ \forall\, t \geq 0),$$

and therefore Y is almost surely cadlag. Since the filtration is right-continuous, it is clear that Y is adapted, so Y is an $\mathcal{F}_t$ martingale. Thus, under the assumption of right-continuity, we have been able to use Rogers & Williams (2000a) Theorem II.67.7 to obtain a version Y of M which is a martingale and almost surely cadlag.

We will now consider what can be done to make sure that *all* paths of our version of M are cadlag. An obvious idea would be to put $Y' = 1_G Y$. This corresponds to letting Y' be equal to Y on G, where Y is guaranteed to be cadlag, and zero otherwise. All paths of Y' are cadlag. However, since we have no guarantee that $G \in \mathcal{F}_t$, Y' is no longer adapted. This is what ruins our ability to obtain all paths cadlag in the uncompleted case: the set of non-cadlag paths does not interact well with the filtration.

However, if we assume that $\mathcal{F}_t$ contains all P-null sets, we can use the fact that $P(G) = 1$ to see that Y' is adapted. Thus, in this case, we can obtain a version Y' of M which is a martingale and has all paths cadlag.

We have now argued for the three claims we made earlier. A general moral of our arguments is that if X is an adapted process, then we can change X on a null set and retain adaptedness if the filtration contains all null sets.

Upsides and downsides of the usual conditions. We now discuss reasons for assuming or not assuming the usual conditions.

The upsides of the usual conditions are:

1. Path properties hold for all ω, so there are fewer null sets to worry about.
2. Proofs of adaptedness are occasionally simpler.
3. Most literature assumes the usual conditions, making it easier to refer results.

The downsides of the usual conditions are:

1. We need to spend time checking that, say, the Brownian motion can be taken to be with respect to a filtration satisfying the usual conditions.
2. Density proofs are occasionally harder.

To see how the usual conditions can make density proofs harder, look at the proof of Lemma 3.7.5. To see how proofs of adaptedness can be harder without the usual conditions, we go back to the construction seen earlier, $Y_t(\omega) = \lim_{q\downarrow\downarrow t} M_t(\omega)$ whenever the limit exists and zero otherwise. Had we assumed the usual conditions, we could just define the entire trajectory of Y on some almost sure set and putting it to zero otherwise. This phenomenon would also cause some trouble when defining the stochastic integral for integrators of finite variation.

As null sets really are a frustrating aspect of the entire theory, adding almost no real value but constantly complicating proofs, we have chosen to offer a bit of extra effort to introduce the usual conditions, in return obtaining fewer null sets to worry about. We could have developed the entire theory without the usual conditions, but then all our stochastic integrals and other processes would only be continuous almost surely.

Some would say that another downside of the usual conditions is that the filtration loses some of its interpretation as "carrier of information", in the sense that $\mathcal{F}_t$ is usually interpreted as "the information available at time t". We do not find this argument to be particularly weighty. Obviously, the null sets $\mathcal{N}$ contain sets from all of $\mathcal{F}$, in particular sets from $\mathcal{F}_t$ for any $t \geq 0$. But these are sets of probability zero. It would therefore seem that the information value in these sets is quite minimal. Since completing the filtration does not have any negative impact on actual modeling issues, it seems unreasonable to maintain that the interpretation of the filtration is reduced.

There do seem to be situations, though, where the usual conditions should not be assumed. When considering the martingale problem, Jacod & Shiryaev (1987) in Section III.1 specifically notes that they are considering filtrations that do not necessarily satisfy the usual conditions. However, Rogers & Williams (2000b) also considers the martingale problem in Section V.19, even though they have only developed the stochatic integral under the usual conditions, and seem to get through with this. Changes of measure can also present situations where the usual conditions are inappropriate, see the comments in Rogers & Williams (2000b), Section IV.38, page 82.

3.10 Notes

In this section, we will review other accounts of the theory in the literature, and we will give references to extensions of the theory. We will also compare our methods to those found in other books and discuss the merits and demerits of our methods.

Other accounts of the theory of stochastic integration. There are many excellent sources for the theory of stochastic integration. We will consider the approaches to the stochastic integral taken in Øksendal (2005), Rogers & Williams (2000b), Protter (2005) and Karatzas & Shreve (1988).

In Øksendal (2005), the integrators are resticted to processes of the type given by $\int_0^t Y_t \,\mathrm{d}t + \int_0^t Y_t \,\mathrm{d}W_t$. The integral is developed over compact intevals and the viewpoint is more variable-oriented and less process-oriented. The level of rigor is lower, but the text is easier to comprehend and gives a good overview of the theory.

In Rogers & Williams (2000b), a very systematic development of the integral is made, resulting in a large generality of integrators. As integrators, the authors first take elements of $\mathbf{c}\mathcal{M}_0^2$. They develop the integral for integrands in $\mathbf{b}\mathcal{E}$, and then extend to the space of locally bounded predictable integrands, $\mathfrak{L}_0(\mathbf{b}\mathcal{P})$. Here, the predictable σ-algebra the σ-algebra on $(0,\infty)\times\Omega$ generated by the left-continuous processes, or equivalently, generated by $\mathbf{b}\mathcal{E}$. This σ-algebra is strictly smaller than Σ_π, and thus the integrands of Rogers & Williams (2000b) are less general than ours. However, the extension method is simpler, based on the monotone class theorem. Their greater range of integrators comes at the cost of having to develop the quadratic variation in greater detail. After constructing the integral for integrators in $\mathbf{c}\mathcal{M}_0^2$, they later develop a much more general theory, where the integrators are semimartingales, meaning the sum of a local martingale and a finite variation process.

In Protter (2005), a kind of converse viewpoint to other books is taken. Instead of developing the integral for more and more general processes, the author defines a semimartingale as a process X making the integral mapping with respect to X continuous under suitable topologies. In the first part of the book, the integral is developed as a mapping from left-continuous processes to right-continuous processes. In time, it is shown that this integral leads to the same result as the conventional approach. In later parts of the book, the space of integrands are extended to predictable processes.

Finally, in Karatzas & Shreve (1988), the integral is developed in the usual straightforward manner as in Rogers & Williams (2000b), but here, only continuous integrators are considered. As in Rogers & Williams (2000b), the authors begin with processes in $\mathbf{c}\mathcal{M}_0^2$ as integrators, but in this case, the space of integrands is larger than in Rogers & Williams (2000b). The account is very detailed and highly recommended.

Extensions of the theory. As noted above, the theory we have presented can be extended considerably. Rogers & Williams (2000b), Protter (2005) and Karatzas & Shreve (1988) all consider more general integrators than we do here. A very pedagogical source for the general theory is Elliott (1982), though the number of misprints can be frustrating.

The idea of extending the stochastic integral begs the question of whether results such as Itô's representation theorem and Girsanov's theorem also can be extended to more general situations. To a large extent, the answer is yes. For extensions of the representation theorem, see Section III.4 of Jacod & Shiryaev (1987) or Section IV.3 of Protter (2005). Extensions of the Girsanov theorem can be found in Section III.3 of Jacod & Shiryaev (1987), Section III.8 of Protter (2005), and also in Rogers & Williams (2000b) and Karatzas & Shreve (1988). One immediate extension is the opportunity to prove Girsanov's theorem in a version where P and Q are not equivalent on $\mathcal{F}_\infty$, but only on $\mathcal{F}_t$ for all $t \geq 0$. Having proved this extension would have made our applications to mathematical finance easier. Alas, this realization came too late.

The search for results like Kazamaki's or Novikov's criteria guaranteeing that an exponential martingale is a uniformly integrable martingale can reasonably be said to be an active field of research. See Protter & Shimbo (2006) for recent results and further references. For an example of a situation where Kazamaki's criterion applies but Novikov's criterion does not, see Revuz & Yor (1998), page 336.

Discussion of the present account. The account of the theory developed here has the same level of ambition as, say, Øksendal (2005) or Steele (2000), but with a higher level of rigor. It is an attempt to combine the methods of Steele (2000) with the stringency of Rogers & Williams (2000b) and Karatzas & Shreve (1988).

The restriction to continuous integrators simplifies the theory immensely, more or less to a point beyond comparison. One consequence of this restriction is that we can take our integrands to be progressively measurable instead of predictably measurable.

Not only is progressive measurability less restrictive than predictable measurability, but it also removes some of the trouble with the timepoint zero. Traditionally, the role of zero has not been well-liked. In Dellacherie & Meyer (1975), Section IV.61, the authors claim that zero plays "the devil's role", and in the introduction to Chapter IV of Rogers & Williams (2000b), it is "consigned to hell". These problems very much has something to do with the predictable σ-algebra, this is most obvious in Rogers & Williams (2000b) where the predictable σ-algebra is defined as a σ-algebra on $(0, \infty) \times \Omega$ instead of $[0, \infty) \times \Omega$. Even in our case, however, there are some considerations to be made. The fundamental problem with zero is that the intuitive role of an integrator X should only depend on differences $X_t - X_s$, interpreted as the measure assigned to $(s, t]$. The existence of a value at zero is therefore in fact something of a nuisance. In our definition of the integral, we have chosen to assign any non-zero value at zero of the integrator X to the finite-variation part of X and thus deferred the problems to results from ordinary measure theory.

In fact, a good deal of our results can be extended to integrators which are not even progressively measurable. This is due to a result, the Chung-Doob-Meyer theorem, stating that for any measurable and adapted process, there exists a progressive version of that process. This would enable us to define the stochastic integral with respect to all measurable and adapted processes satisfying appropriate integrability conditions. The original proof of the Chung-Doob-Meyer theorem was quite difficult and can be found in Dellacherie & Meyer (1975), Section 30 of Chapter IV, page 99. Recently, a much simpler proof has surfaced, see Kaden & Potthoff (2004). Because we have assumed the usual conditions, we could easily have adapted our results to consider meaurable and adapted integrands. Without the usual conditions, however, the integration of measurable and adapted processes causes problems with adaptedness. In essence, progressive meaurability is the "right" measurability concept for dealing with pathwise Lebesgue integration.

Now, as stated in the introduction, part of the purpose of the account given here was to explore what it would be like to develop the simple theory of stochastic integration with respect to Brownian integrators, just as in Steele (2000), but with complete rigor and in a self-contained manner. The reason to do so is mainly to avoid the trouble of having to prove the existence of the quadratic variation process for general continuous local martingales. This could potentially allow for a very accesible presentation of the basic theory of stochastic integration, which covers much of the results necessary for, say, basic mathematical finance. However, as we have seen, the lower level of generality has its price. The main demerits are:

1. The lack of general martingales makes the theory less elegant.
2. The development of the integral through associativity in Section 3.3 is tedious. The need to consider n-dimensional Brownian motion and corresponding n-dimensional integrands makes the notation cumbersome.
3. The proof of Girsanov's theorem of Section 3.8 is very lengthy.

The first two demerits more or less speak for themselves. The third demerit deserves some comments. In Øksendal (2005), the integral is also only developed for Brownian integrators, but the proof of the Girsanov theorem given there is considerably shorter.

This is because the author takes the Lévy characterisation of Brownian motion as given, and thus his account is not self-contained.

The modern proof of the Lévy characterisation theorem uses a more general theory of stochastic integration than presented here, see Theorem 33.1 of Rogers & Williams (2000b). There is an older proof, however, not based on stochastic integration theory, but it is much more cumbersome. A one-dimensional proof is given in Doob (1953). If we were to base our proof of the Girsanov theorem on the Lévy characterisation theorem, our account would become even longer than it is at present. The proof we have given is based on the proof in Steele (2000) for the case where the drift is bounded. Steele (2000) claims on page 224 that the proof for the bounded case immediately can be extended to the general case. It is quite unclear how this should be possible, however, since it would require the result that if $\mathcal{E}(M)$ is a martingale and f is bounded and deterministic, then $\mathcal{E}(M + f \cdot W)$ is a martingale as well. Our result extends the method in Steele (2000) by a localisation argument, and it is to a large degree this localisation which encumbers the proof.

If we had the theory of stochastic integration for general continuous local martingales at our disposal, we could have used the arguments of Rogers & Williams (2000b), Section IV.33 and Section IV.38, to give short and elegant proofs of both the Lévy characterisation theorem and the Girsanov theorem.

Our conclusion is that when a high level of rigor is required, it is quite doubtful that there is any use in developing the theory only for the case of Brownian motion instead of considering general continuous local martingales. A theory at the level of Karatzas & Shreve (1988) would therefore be preferable as introductory, with works such as Øksendal (2005) and Steele (2000) being useful for obtaining a heuristic overview of the theory.

Chapter 4

The Malliavin Calculus

In this section, we will define the Malliavin derivative and prove its basic properties. The Malliavin derivative and the theory surrounding it is collectively known as the Malliavin calculus. The Malliavin derivative is an attempt to define a derivative of a stochastic variable X in the direction of the underlying probability space Ω. Obviously, this makes no sense in the classical sense of differentiability unless Ω has sufficient structure. The Malliavin derivative is therefore defined for appropriate variables on general probability spaces not by a limiting procedure as in the case of classical derivative operators, but instead by analogy with a situation where the underlying space has sufficient topological structure.

We will not be able to develop the theory sufficiently to prove the results which are applied to mathematical finance. Therefore, the exposition given here is mostly designed to give a rigorous introduction to the fundamentals of the theory.

Before beginning the development of the Malliavin calculus, we make some motivating remarks on the definition of the derivative. We are interested in defining the derivative on probability spaces $(\Omega, \mathcal{F}, P)$ with a one-dimensional Wiener process W on $[0,T]$, where the variables to be differentiated are in some suitable sense related to W. With this in mind, we first consider the case where our probability space is $C_0[0,T]$, the space of continuous real functions on $[0,T]$ starting at zero, endowed with the Wiener measure P. $C_0[0,T]$ is a Banach space under the uniform norm. Let W be the identity mapping, W is then a Wiener process on $[0,T]$. We want to identify a differentiability concept for variables of the form $\omega \mapsto W_t(\omega)$ which can be generalized to the setting of an abstract probability space with a Wiener process. To this end, consider a mapping X from $C_0[0,T]$ to $\mathbb{R}$. Immediately available differentiability concepts for the mapping are:

1. The Frechét derivative: The Frechét derivative of X at ω is the linear operator $A_\omega : C_0[0,T] \to \mathbb{R}$ such that $\lim_{h\to 0} \frac{\|X(\omega+h)-X(\omega)-A_\omega(h)\|}{\|h\|} = 0$.
2. The Gâteaux directional derivatives: The derivative in direction $h \in C_0[0,T]$ at ω is the element $D_hX(\omega) \in \mathbb{R}$ such that $\lim_{\varepsilon\to 0} \frac{X(\omega+\varepsilon h)-X(\omega)}{\varepsilon} = D_hX(\omega)$.

Here, the Frechét derivative can obviously be ruled out as a candidate for generalization because of the direct dependence of the derivative on the linear structure of $C_0[0,T]$. Considering the Gâteaux derivative, we have that for $h \in C_0[0,T]$, the Gateaux derivative of W_t in direction h is

$$D_hW_t(\omega) = \lim_{\varepsilon\to 0} \frac{W_t(\omega+\varepsilon h) - W_t(\omega)}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{\omega_t + \varepsilon h_t - \omega_t}{\varepsilon} = h_t$$

This means that this derivative is not amenable to direct generalization either, because the derivative is an element of the underlying space depending directly of its properties as a function space. Consider, however, the case where $h_t = \int_0^t g(t)\,\mathrm{d}t$ for some $g \in \mathcal{L}^1[0,T]$. Then,

$$D_hW_t(\omega) = h_t = \int_0^T 1_{[0,t]}(s)g(s)\,\mathrm{d}s.$$

In this case, then, the Gâteaux derivative for any $\omega \in C_0[0,T]$ is actually characterized by the mapping $1_{[0,t]}$ from $[0,T]$ to $\mathbb{R}$. We can therefore consider $1_{[0,t]}$ as a kind of derivative of W_t. Extending this concept to general mappings $X : C_0[0,T] \to \mathbb{R}$, we can say that if there exists a mapping $f : [0,T] \to \mathbb{R}$ such that for $h \in C_0[0,T]$ with $h(t) = \int_0^t g(s)\,\mathrm{d}s$, it holds that $D_hX(\omega) = \int_0^T f(s)g(s)\,\mathrm{d}s$, then f is the derivative of X, and we define $D_{\mathbb{F}}X = f$.

To interpret this derivative, fix $y \in [0,T]$ and consider the case $h_\varepsilon(t) = \int_0^t \psi_\varepsilon(s-y)\,\mathrm{d}s$, where (ψ_ε) is a Dirac family, see Appendix A. This means that h_ε is 0 almost all the way on $[0,y]$ and 1 almost all the way on $[y,T]$. Then,

$$D_{h_\varepsilon}X(\omega) = \int_0^T D_{\mathbb{F}}X(s)\psi_\varepsilon(s-y)\,\mathrm{d}s \approx D_{\mathbb{F}}X(y).$$

In other words, $D_{\mathbb{F}}X(y)$ can be thought of as measuring the sensitivity of X to parallel shifts of the argument ω of the form $1_{[y,T]}$.

Led by the above ideas, consider now a general probability triple $(\Omega, \mathcal{F}, P)$ with a Wiener process W on $[0, T]$. By $\mathcal{F}_T$, we denote the usual augmentation of the σ-algebra generated by W. We want to define the Malliavin derivative of W_t as the mapping $DW_t = 1_{[0,t]}$ from $[0, T]$ to $\mathbb{R}$. Since this derivative in the abstract setting is not defined by a limiting procedure, but only by analogy to the case of $C_0[0, T]$, we need to impose requirements on the D to extend it meaningfully to other variables than W_t. A suitable starting point is to require that a chain rule holds for D: If f is continuously differentiable with sufficient growth conditions, namely polynomial growth of the mapping itself and its partial derivatives, we would like that

$$Df(W_{t_1}, \ldots, W_{t_n}) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(W_{t_1}, \ldots, W_{t_n}) DW_{t_k},$$

where we put $DW_{t_k} = 1_{[0,t_k]}$. It turns out that to require this form of the derivative of $f(W_{t_1}, \ldots, W_{t_n})$ is sufficient to obtain a rich theory for the operator D. The derivative given above is a stochastic function from $[0, T]$ to $\mathbb{R}$. With sufficient growth conditions on f, it is an element of $\mathcal{L}^p([0, T] \times \Omega)$, $p \geq 1$.

Our immediate goal is to formalize the definition of the derivative given above and consider its properties. To do so, we begin by accurately specifying the immediate domain of the derivative operator and considering its properties. After this, we will rigorously define the derivative operator.

4.1 The spaces $\mathcal{S}$ and $\mathcal{L}^p(\Pi)$

We will now define the space $\mathcal{S}$ which will be the initial domain of the Malliavin derivative. We work in the context of a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ with a one-dimensional Brownian motion W, where the filtration is the augmented filtration induced by the Brownian motion. A multi-index of order n is an element $\alpha \in \mathbb{N}_0^n$. The degree of the multi-index is $|\alpha| = \sum_{k=1}^n \alpha_k$. A polynomial in n variables of degree k is a map $p : \mathbb{R}^n \to \mathbb{R}$, $p(x) = \sum_{|\alpha| \leq k} a_\alpha x^\alpha$. The sum in the above runs over all multi-indicies α with $|\alpha| \leq k$, with $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$. The space of polynomials of degree k in any number of variables is denoted $\mathfrak{P}_k$. Furthermore, we use the following notation.

- $C^1(\mathbb{R}^n)$ are the mappings $f : \mathbb{R}^n \to \mathbb{R}$ which are continuously differentiable.
- $C^\infty(\mathbb{R}^n)$ are the mappings $f : \mathbb{R}^n \to \mathbb{R}$ which are differentiable infinitely often.

- $C_p^\infty(\mathbb{R}^n)$ are the elements $f \in C^\infty(\mathbb{R}^n)$ such that f and its partial derivatives are dominated by polynomials.
- $C_c^\infty(\mathbb{R}^n)$ are the elements of $C^\infty(\mathbb{R})$ with compact support.

Clearly, then $C_c^\infty(\mathbb{R}^n) \subseteq C_p^\infty(\mathbb{R}^n) \subseteq C^\infty(\mathbb{R}^n) \subseteq C^1(\mathbb{R}^n)$. For basic results on these spaces, see appendix A. We are now ready to define the space $\mathcal{S}$.

Definition 4.1.1. *If $t \in [0,T]^n$, we write $W_t = (W_{t_1}, \ldots, W_{t_n})$. By $\mathcal{S}$, we then denote the space of variables $f(W_t)$, where $f \in C_p^\infty(\mathbb{R}^n)$ and $t \in [0,T]^n$.*

Our aim in the following regarding $\mathcal{S}$ is to develop a canonical representation for elements of $\mathcal{S}$, develop results to yield flexible representations of elements in $\mathcal{S}$ and to prove that $\mathcal{S}$ is an algebra which is dense in $\mathcal{L}^p(\mathcal{F}_T)$, $p \geq 1$. These results are somewhat tedious, but they are necessary for what will follow.

Lemma 4.1.2. *Let $t \in [0,T]^n$, $s \in [0,T]^m$ and $F \in \mathcal{S}$ with $F = f(W_t)$. Assume that $\{t_1, \ldots, t_n\} \subseteq \{s_1, \ldots, s_m\}$. There exists $g \in C_p^\infty(\mathbb{R}^m)$ such that $F = g(W_s)$.*

Comment 4.1.3 Note that this lemma not only allows us to extend the coordinates which F depends on, but also allows us to reorder the coordinates. ∘

Proof. Since $\{t_1, \ldots, t_n\} \subseteq \{s_1, \ldots, s_m\}$, we have $n \leq m$ and there exists a mapping $\sigma : \{1, \ldots, n\} \to \{1, \ldots, m\}$ such that $t_k = s_{\sigma(k)}$ for $k \leq n$. We then simply define $g(x_1, \ldots, x_m) = f(x_{\sigma_1}, \ldots, x_{\sigma_n})$. Then $g \in C_p^\infty(\mathbb{R}^m)$ and

$$\begin{aligned} g(W_s) &= g(W_{s_1}, \ldots, W_{s_m}) \\ &= f(W_{s_{\sigma_1}}, \ldots, W_{s_{\sigma_n}}) \\ &= f(W_{t_1}, \ldots, W_{t_n}) \\ &= f(W_t) \end{aligned}$$

□

Lemma 4.1.4. *Let $F \in \mathcal{S}$ with $F = f(W_t)$ for $f \in C_p^\infty(\mathbb{R}^n)$. If $t_1 = 0$, there exists $g \in C_p^\infty(\mathbb{R}^{n-1})$ such that $F = g(W_{t_2}, \ldots, W_{t_n})$.*

Proof. Define $g(x_2, \ldots, x_n) = f(0, x_2, \ldots, x_n)$. Because we have $W_0 = 0$, it holds that $F = f(W_{t_1}, \ldots, W_{t_n}) = g(W_{t_2}, \ldots, W_{t_n})$. □

Lemma 4.1.5. *Let $F \in \mathcal{S}$ with $F = f(W_t)$ for $f \in C_p^\infty(\mathbb{R}^n)$. If $t_i = t_j$ for some $i \neq j$, there is $g \in C_p^\infty(\mathbb{R}^{n-1})$ such that $F = g(W_u)$, where $u = (t_1, \ldots, t_{j-1}, t_{j+1} \ldots, t_n)$.*

Proof. Define $A : \mathbb{R}^{n-1} \to \mathbb{R}^n$ by $Ax = (x_1, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_n)$. Then A is linear. Putting $g(x) = f(Ax)$, then $g \in C_p^\infty(\mathbb{R}^n)$. Clearly, $AW_u = W_t$ and therefore $F = f(W_t) = f(AW_u) = g(W_u)$, as desired. □

Corollary 4.1.6. *Let $F \in \mathcal{S}$. There exists $t \in [0,T]^n$ such that $0 < t_1 < \cdots < t_n$ and $f \in C_p^\infty(\mathbb{R}^n)$ such that $F = f(W_t)$.*

Proof. This follows by combining Lemma 4.1.2, Lemma 4.1.4 and Lemma 4.1.5. □

Corollary 4.1.6 states that any element of $\mathcal{S}$ has a representation where we can assume that all the coordinates of the Wiener process in the element are different, positive and ordered. This observation will make our lives a good deal easier in the following. Our next result shows that any element of $\mathcal{S}$ can be written as a transformation of an n-dimensional standard normal variable.

Lemma 4.1.7. *Let $F \in \mathcal{S}$ with $F = f(W_t)$ for $f \in C_p^\infty(\mathbb{R}^n)$, where $0 < t_1 < \cdots < t_n$. Define $A : \mathbb{R}^n \to \mathbb{R}^n$ by $Ax = (\frac{x_1}{\sqrt{t_1}}, \frac{x_2 - x_1}{\sqrt{t_2 - t_1}}, \ldots, \frac{x_n - x_{n-1}}{\sqrt{t_n - t_{n-1}}})$. Then A is invertible, $f \circ A^{-1} \in C_p^\infty(\mathbb{R}^n)$ and $F = (f \circ A^{-1})(AW_t)$, where AW_t has the standard normal distribution in n dimensions.*

Proof. Put $t_0 = 0$. A is invertible with inverse given by $(A^{-1}x)_k = \sum_{i=0}^k \sqrt{t_k - t_{k-1}} x_k$. Since A^{-1} is linear, $f \circ A^{-1} \in C_p^\infty(\mathbb{R}^n)$. It is clear that AW_t is standard normal. □

We are now done with our results on representations of elements of $\mathcal{S}$. Before proceeding, we prove that $\mathcal{S}$ is an algebra which is dense in $\mathcal{L}^p(\mathcal{F}_T)$.

Lemma 4.1.8. *$\mathcal{S}$ is an algebra, and $\mathcal{S} \subseteq \mathcal{L}^p(\Omega)$ for all $p \geq 1$.*

Proof. **$\mathcal{S}$ is an algebra.** $\mathcal{S}$ is a space of real random variables. Since $\mathcal{S}$ is obviously nonempty, to prove the first claim, we need to show that $\mathcal{S}$ is stable under linear combinations and multiplication. Let $F, G \in \mathcal{S}$ with $F = f(W_t)$ and $G = g(W_s)$ where $f \in C_p^\infty(\mathbb{R}^n)$ and $g \in C_p^\infty(\mathbb{R}^m)$. Let $\lambda, \mu \in \mathbb{R}$. Then,

$$\lambda F + \mu G = (\lambda f \oplus \mu g)(W_{t_1}, \ldots, W_{t_n}, W_{s_1}, \ldots, W_{s_m}),$$

where $\lambda f \oplus \mu g \in C_p^\infty(\mathbb{R}^{n+m})$. In the same manner,

$$FG = (f \otimes g)(W_{t_1}, \ldots, W_{t_n}, W_{s_1}, \ldots, W_{s_m}),$$

where, $f \otimes g \in C_p^\infty(\mathbb{R}^{n+m})$. This shows that $\mathcal{S}$ is an algebra.

$\mathcal{S}$ is a subset of $\mathcal{L}^p(\Omega)$. Next, we will show that $\mathcal{S} \subseteq \mathcal{L}^p(\Omega)$. Let $F \in \mathcal{S}$ be such that $F = f(W_t)$. By Corollary 4.1.6, we can assume $0 < t_1 < \cdots < t_n$. By Lemma 4.1.7, there then exists $g \in C_p^\infty(\mathbb{R}^n)$ such that $F = g(X)$ where X is n-dimensionally standard normally distributed. Since $g \in C_p^\infty(\mathbb{R}^n)$, $|g|^p \in C_p^\infty(\mathbb{R}^n)$. Therefore, there is a multinomial q in n variables with $|g|^p \leq q$, say $q(x) = \sum_{|\alpha| \leq m} a_\alpha x^\alpha$. We then find

$$\begin{aligned} E|F|^p &= \int |g(x)|^p \phi_n(x)\,\mathrm{d}x \\ &\leq \int q(x)\phi_n(x)\,\mathrm{d}x \\ &\leq \int \sum_{|\alpha|\leq m} \prod_{k=1}^n a_\alpha x_k^{\alpha_k} \phi_n(x)\,\mathrm{d}x \\ &= \sum_{|\alpha|\leq m} \prod_{k=1}^n a_\alpha \int x_k^{\alpha_k} \phi(x_k)\,\mathrm{d}x_k, \end{aligned}$$

which is finite, since the normal distribution has moments of all orders. □

Lemma 4.1.9. *The variables of the form $f(W_t)$ for $f \in C_c^\infty(\mathbb{R}^n)$ are dense in $\mathcal{L}^p(\mathcal{F}_T)$ for $p \geq 1$.*

Proof. As in the proof of Lemma 3.7.5, we find that with $\mathcal{G}_T$ the σ-algebra generated by W, $\mathcal{F}_T \subseteq \sigma(\mathcal{G}_T, \mathcal{N})$, so it will suffice to show that the variables given in the statement of the lemma is dense in $\mathcal{L}^2(\mathcal{G}_T)$.

We first prove that any $X \in \mathcal{L}^p(\mathcal{G}_T)$ can be approximated by variables dependent only on finitely many coordinates of the Wiener process. Let $X \in \mathcal{L}^p(\mathcal{G}_T)$ be given and let $\{t_k\}_{k\geq 1}$ be dense in $[0,T]$ and let $\mathcal{H}_n = \sigma(W_{t_1}, \ldots, W_{t_n})$. Then $\mathcal{H}_n$ is an increasing family of σ-algebras, and $\mathcal{G}_T = \sigma(\mathcal{H}_n)_{n\geq 1}$. Therefore, $E(X|\mathcal{H}_n)$ is a discrete-time martingale with limit X, convergent in $\mathcal{L}^p$. By the Doob-Dynkin lemma B.6.7, $E(X|\mathcal{H}_n) = g(W_{t_1}, \ldots, W_{t_n})$ for some measurable $g : \mathbb{R}^n \to \mathbb{R}$.

Now consider $X \in \mathcal{L}^p(\mathcal{G}_T)$ with $X = g(W_t)$ for some measurable $g : \mathbb{R}^n \to \mathbb{R}$ and $t \in [0,T]^n$. Let μ be the distribution of W_t. Since μ is bounded, μ is a Radon measure on $\mathbb{R}^n$, and $g \in \mathcal{L}^p(\mu)$ by assumption. By Theorem A.4.5, g can be approximated in $\mathcal{L}^p(\mu)$ by mappings $(g_n) \subseteq C_c^\infty(\mathbb{R}^n)$. We then obtain

$$\lim \|g(W_t) - g_n(W_t)\|_p = \lim \|g - g_n\|_{\mathcal{L}^p(\mu)} = 0,$$

showing the claim of the lemma. □

Corollary 4.1.10. *The space $\mathcal{S}$ is dense in $\mathcal{L}^p(\mathcal{F}_T)$ for any $p \geq 1$.*

Proof. This follows immediately from Lemma 4.1.9, since $C_c^\infty(\mathbb{R}^n) \subseteq C_p^\infty(\mathbb{R}^n)$. □

This concludes our basic results on $\mathcal{S}$, which we will use in the following sections when developing the Malliavin derivative. The important points of our work is this:

- By Corollary 4.1.6, we can choose a particularly nice form of any $F \in \mathcal{S}$, namely that $F = f(W_t)$ where $0 < t_1 < \cdots < t_n$. Also, by Lemma 4.1.2 we can extend to more coordinates when we wish. When working with more than one element of $\mathcal{S}$, this will allow us to assume that they depend on the same underlying coordinates of the Wiener process.
- By combining Corollary 4.1.6 and Lemma 4.1.7, we can always write any $F \in \mathcal{S}$ as a C_p^∞ transformation of a standard normal variable.
- By Lemma 4.1.8 and Corollary 4.1.10, $\mathcal{S}$ is an algebra, dense in $\mathcal{L}^p(\mathcal{F}_T)$ for any $p \geq 1$.

Before we proceed to the next section, we show some basic results on the space $\mathcal{L}^p([0,T] \times \Omega, \mathcal{B}[0,T] \otimes \mathcal{F}_T, \lambda \otimes P)$. We will use the shorthands $\Pi = \mathcal{B}[0,T] \otimes \mathcal{F}_T$ and $\eta = \lambda \otimes P$ and write $\mathcal{L}^p(\Pi)$ for $\mathcal{L}^p([0,T] \times \Omega, \mathcal{B}[0,T] \otimes \mathcal{F}_T, \lambda \otimes P)$. When $p = 2$, we denote the inner product on $\mathcal{L}^2(\Pi)$ by $\langle \cdot, \cdot \rangle_\Pi$.

Lemma 4.1.11. *The variables of the form $f(W_t)$ for $f \in C_c^\infty(\mathbb{R}^n)$ generate $\mathcal{F}_T$.*

Proof. This follows immediately from Lemma 4.1.9, since this lemma yields that for any $A \in \mathcal{F}_T$, 1_A can be approximated pointwise almost surely with elements of the form $f(W_t)$, where $f \in C_c^\infty(\mathbb{R}^n)$. □

Lemma 4.1.12. *The functions $1_{[s,t]} \otimes f(W_u)$, where $0 \leq s \leq t \leq T$, $u \in [0,T]^n$ and $f \in C_c^\infty(\mathbb{R}^n)$ is stable under multiplication and generates Π.*

Proof. Clearly, the class is stable under multiplication. We show the statement about its generated σ-algebra. We will apply Lemma B.6.4. Let $\mathbb{E}$ be the class of mappings $f(W_u)$ where $f \in C_c^\infty(\mathbb{R}^n)$ and $u \in [0,T]^n$. Let $\mathbb{K}$ be the class of functions $1_{[s,t]}$ for $0 \leq s \leq t \leq T$.

By Lemma 4.1.11, $\sigma(\mathbb{E}) = \mathcal{F}_T$. Since $f(W_0)$ is constant for any $f \in C_c^\infty(\mathbb{R})$, $\mathbb{E}$ contains the constant mappings. Also, clearly $\sigma(\mathbb{K}) = \mathcal{B}[0,T]$, and clearly the mapping which is constant 1 on $[0,T]$ is in $\mathbb{K}$. The hypotheses of Lemma B.6.4 are then fulfilled, and the conclusion follows. □

4.2 The Malliavin Derivative on $\mathcal{S}$ and $\mathbb{D}_{1,p}$

With the results of Section 4.1, we are ready to define and develop the Malliavin derivative. Letting $F \in \mathcal{S}$ with $F = f(W_t)$, where $f \in C_p^\infty(\mathbb{R}^n)$, we want to put

$$DF = \sum_{k=1}^n \frac{\partial f}{\partial x_k}(W_t)1_{[0,t_k]}.$$

We begin by showing that this is well-defined, in the sense that if we have different representations $f(W_t) = g(W_s)$, the derivative yields the same result.

Lemma 4.2.1. *Let $F \in \mathcal{S}$, and assume that we have two representations,*

$$\begin{aligned} F &= f_1(W_t) \\ F &= f_2(W_s), \end{aligned}$$

where $f_1 \in C_p^\infty(\mathbb{R}^n)$ and $f_2 \in C_p^\infty(\mathbb{R}^m)$. It then holds that

$$\sum_{k=1}^n \frac{\partial f_1}{\partial x_k}(W_t)1_{[0,t_k]} = \sum_{k=1}^m \frac{\partial f_2}{\partial x_k}(W_s)1_{[0,s_k]}.$$

Proof. We proceed in two steps, first showing the equality for particularly simple representations of F and then proceeding to the general case.

Step 1. First assume that $F = f_1(W_t) = f_2(W_t)$ for some $f_1, f_2 \in C_p^\infty(\mathbb{R}^n)$, where $0 < t_1 < \cdots < t_n$. Define $A : \mathbb{R}^n \to \mathbb{R}^n$ by $Ax = (\frac{x_1}{\sqrt{t_1}}, \frac{x_2 - x_1}{\sqrt{t_2 - t_1}}, \ldots, \frac{x_n - x_{n-1}}{\sqrt{t_n - t_{n-1}}})$. By Lemma 4.1.7, A is invertible, and AW_t has the standard normal distribution in n dimensions. Put $g_1 = f_1 \circ A^{-1}$ and $g_2 = f_2 \circ A^{-1}$. Then $g_1(AW_t) = g_2(AW_t)$.

This implies that g_1 and g_2 are equal almost surely with respect to the n-dimensional standard normal measure. Since this measure has a positive density with respect to the Lebesgue measure, this means that g_1 and g_2 are equal Lebesgue almost surely. Since both mappings are continuous, we conclude $g_1 = g_2$. Because $f_1 = g_1 \circ A$ and $f_2 = g_2 \circ A$, this implies $f_1 = f_2$ and therefore

$$\sum_{k=1}^{n} \frac{\partial f_1}{\partial x_k}(W_t)1_{[0,t_k]} = \sum_{k=1}^{n} \frac{\partial f_2}{\partial x_k}(W_t)1_{[0,t_k]},$$

as desired.

Step 2. Now assume $F = f_1(W_t)$ and $F = f_2(W_s)$. Defining $u = (t_1, \ldots, t_n, s_1, \ldots, s_m)$ and putting

$$\begin{aligned} g_1(x_1, \ldots, x_{n+m}) &= f_1(x_1, \ldots, x_n) \\ g_2(x_1, \ldots, x_{n+m}) &= f_2(x_{n+1}, \ldots, x_m), \end{aligned}$$

we obtain $f_1(W_t) = g_1(W_u)$ and $f_2(W_s) = g_2(W_u)$. By what we already have shown,

$$\sum_{k=1}^{n+m} \frac{\partial g_1}{\partial x_k}(W_u)1_{[0,t_k]} = \sum_{k=1}^{n+m} \frac{\partial g_2}{\partial x_k}(W_u)1_{[0,t_k]}.$$

Therefore, we conclude

$$\begin{aligned} \sum_{k=1}^{n} \frac{\partial f_1}{\partial x_k}(W_t)1_{[0,t_k]} &= \sum_{k=1}^{n} \frac{\partial g_1}{\partial x_k}(W_u)1_{[0,u_k]} = \sum_{k=1}^{n+m} \frac{\partial g_1}{\partial x_k}(W_u)1_{[0,u_k]} \\ &= \sum_{k=1}^{n+m} \frac{\partial g_2}{\partial x_k}(W_u)1_{[0,u_k]} = \sum_{k=1}^{m} \frac{\partial f_2}{\partial x_k}(W_s)1_{[0,s_k]}, \end{aligned}$$

showing the desired result. □

Definition 4.2.2. *Let $F \in \mathcal{S}$, with $F = f(W_{t_1}, \ldots, W_{t_n})$. We define the Malliavin derivative of F as the stochastic process on $[0,T]$ given by*

$$\sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(W_{t_1}, \ldots, W_{t_n})1_{[0,t_k]}.$$

Comment 4.2.3 The mapping defined above is well-defined by Lemma 4.2.1. ∘

We have now defined the Malliavin derivative D on $\mathcal{S}$. Next, we will investigate its basic properties and extend it to a larger space.

Lemma 4.2.4. *The mapping D is linear and maps into $\mathcal{L}^p(\Pi)$ for all $p \geq 1$.*

Proof. Let $F, G \in \mathcal{S}$ with $F = f(W_t)$ and $G = g(W_s)$. Also, let $\lambda, \mu \in \mathbb{R}$. From Lemma 4.1.8 and its proof, we have

$$\begin{aligned}
D(\lambda F + \mu G) &= D(\lambda f \oplus \mu g)(W_{t_1}, \ldots, W_{t_n}, W_{s_1}, \ldots, W_{s_m}) \\
&= \sum_{k=1}^n \lambda \frac{\partial f}{\partial x_k}(W_t) 1_{[0,t_k]} + \sum_{k=1}^m \mu \frac{\partial g}{\partial x_k}(W_s) 1_{[0,s_k]} \\
&= \lambda DF + \mu DG.
\end{aligned}$$

This shows linearity. To show the second statement, assume $F = f(W_t)$. Then,

$$\|DF\|_p = \left\| \sum_{k=1}^n \frac{\partial f}{\partial x_k}(W_t) 1_{[0,t_k]} \right\|_p \leq T^{\frac{1}{p}} \sum_{k=1}^n \left\| \frac{\partial f}{\partial x_k}(W_t) \right\|_p .$$

Since $f \in C_p^\infty(\mathbb{R}^n)$, $\frac{\partial f}{\partial x_k}$ has polynomial growth, so $\frac{\partial f}{\partial x_k}(W_t) \in \mathcal{L}^p(\mathcal{F}_T)$, proving the lemma. □

From Corollary 4.1.10, we see that we have defined the derivative on a dense subspace of $\mathcal{L}^p(\mathcal{F}_T)$. Nonetheless, even dense sets can be quite slim, and we need to extend the operator D to a larger space before it can be of any actual use. To do so, we show that D is closable. We can then extend it by taking the closure. We first need a few lemmas.

Lemma 4.2.5. *Let $F \in \mathcal{S}$ with $F = f(W_t)$ for $f \in C_p^\infty(\mathbb{R}^n)$, where $t = (t_1, \ldots, t_n)$ and $0 = t_0 < t_1 < \cdots < t_n$. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be linear and invertible with matrix A'. Then*

$$\sum_{k=1}^n \frac{\partial}{\partial x_k} f(W_t) 1_{[0,t_k]} = \sum_{k=1}^n \sum_{i=1}^n \frac{\partial}{\partial x_i}(f \circ A^{-1})(AW_t) A'_{ik} 1_{[0,t_k]}.$$

Proof. This follows from

$$\begin{aligned}
\sum_{k=1}^n \frac{\partial}{\partial x_k} f(W_t) 1_{[0,t_k]} &= \sum_{k=1}^n \frac{\partial}{\partial x_k}(f \circ A^{-1} \circ A)(W_t) 1_{[0,t_k]} \\
&= \sum_{k=1}^n \sum_{i=1}^n \frac{\partial}{\partial x_i}(f \circ A^{-1})(AW_t) \frac{\partial}{\partial x_k} A_i(W_t) 1_{[0,t_k]} \\
&= \sum_{k=1}^n \sum_{i=1}^n \frac{\partial}{\partial x_i}(f \circ A^{-1})(AW_t) A'_{ik} 1_{[0,t_k]},
\end{aligned}$$

where we have applied the chain rule. □

Lemma 4.2.6. *Let $f \in C_p^\infty(\mathbb{R}^n)$. Let ϕ_n be the n-dimensional standard normal density. Then,*

$$\int \frac{\partial f}{\partial x_k}(x)\phi_n(x)\,\mathrm{d}x = \int f(x)x_k\phi_n(x)\,\mathrm{d}x.$$

Proof. First note that the integrals are well-defined, since ϕ_n decays faster than any polynomial. We begin by showing that it will suffice to prove the result in the case where $k=1$, this will make the proof notationally easier. To this end, note that the Lebesgue measure is invariant under orthonormal transformations. Define the mapping $U:\mathbb{R}^n \to \mathbb{R}^n$ as the linear transformation exchanging the first and k'th coordinates, leaving all other coordinates the same. U is then orthonormal, and

$$\begin{aligned}\int \frac{\partial f}{\partial x_k}(x)\phi_n(x)\,\mathrm{d}x &= \int \frac{\partial f}{\partial x_k}(Ux)\phi_n(Ux)\,\mathrm{d}x\\ &= \int \sum_{i=1}^n \frac{\partial f}{\partial x_i}(Ux)\frac{\partial U_i}{\partial x_1}(x)\phi_n(Ux)\,\mathrm{d}x\\ &= \int \frac{\partial (f\circ U)}{\partial x_1}(x)\phi_n(x)\,\mathrm{d}x.\end{aligned}$$

Likewise,

$$\int f(x)x_k\phi_n(x)\,\mathrm{d}x = \int f(Ux)(Ux)_k\phi_n(Ux)\,\mathrm{d}x = \int (f\circ U)(x)x_1\phi_n(x)\,\mathrm{d}x.$$

Therefore, it will suffice to show the result for $k=1$ for mappings of the form $f\circ U$ where $f\in C_p^\infty(\mathbb{R}^n)$. But such mappings are in $C_p^\infty(\mathbb{R}^n)$ as well, and we conclude that it will suffice to show the result in the case where $k=1$ for some $f\in C_p^\infty(\mathbb{R}^n)$. In that case, we obtain

$$\begin{aligned}&\int \frac{\partial f}{\partial x_1}(x)\phi_n(x)\,\mathrm{d}x\\ =\;& \lim_{M\to\infty}\int_{[-M,M]^n}\frac{\partial f}{\partial x_1}(x)\phi_n(x)\,\mathrm{d}x\\ =\;& \lim_{M\to\infty}\int_{[-M,M]^{n-1}}\int_{-M}^{M}\frac{\partial f}{\partial x_1}(x)\phi_n(x)\,\mathrm{d}x_1\cdots\mathrm{d}x_n\\ =\;& \lim_{M\to\infty}\int_{[-M,M]^{n-1}}[f(x)\phi_n(x)]_{x_1=-M}^{x_1=M} - \int_{-M}^{M} f(x)\frac{\partial\phi_n}{\partial x_1}(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n.\end{aligned}$$

We wish to use dominated convergence to get rid of the first term. Defining x' by

$x' = (x_2, \ldots, x_n)$, we obtain

$$\begin{aligned} & \int_{[-M,M]^{n-1}} \left| [f(x)\phi_n(x)]_{x_1=-M}^{x_1=M} \right| \,\mathrm{d}x_2 \cdots \mathrm{d}x_n \\ \leq \quad & \int \left| [f(x)\phi_n(x)]_{x_1=-M}^{x_1=M} \right| \,\mathrm{d}x' \\ \leq \quad & \int |f(M,x')|\phi_n(M,x') + |f(-M,x')|\phi_n(-M,x')\,\mathrm{d}x' \\ = \quad & \int |f(M,x')|\phi(M)\phi_{n-1}(x')\,\mathrm{d}x' + \int |f(-M,x')|\phi(-M)\phi_{n-1}(x')\,\mathrm{d}x' \end{aligned}$$

Let p with $p(x) = \sum_{k=0}^m a_k \prod_{i=1}^n x^{\alpha_i^k}$, be a polynomial such that $|f| \leq p$. Then,

$$\begin{aligned} \sup_{M\in\mathbb{R}} |f(M,x')|\phi(M) &\leq \sum_{k=0}^m |a_k| \sup_{M\in\mathbb{R}} \phi(M) M^{\alpha_1^k} \prod_{i=2}^n |x|^{\alpha_i^k} \\ &= \sum_{k=0}^m \left(|a_k| \sup_{M\in\mathbb{R}} |M|^{\alpha_1^k}\phi(M) \right) \prod_{i=2}^n |x|^{\alpha_i^k}, \end{aligned}$$

where the suprema are finite since ϕ decays faster than any exponential. Therefore, $\sup_{M\in\mathbb{R}} |f(M,x')|\phi_1(M)$ has polynomial growth. From this we conclude that $\sup_{M\in\mathbb{R}} |f(M,x')|\phi_1(M)\phi_{n-1}(x')$ is integrable, and by dominated convergence, we therefore get

$$\begin{aligned} & \lim_{M\to\infty} \int_{[-M,M]^{n-1}} [f(x)\phi_n(x)]_{x_1=-M}^{x_1=M} \,\mathrm{d}x' \\ = \quad & \int \lim_{M\to\infty} 1_{[-M,M]^{n-1}}(x')(f\phi_n)(M,x') - 1_{[-M,M]^{n-1}}(x')(f\phi_n)(-M,x')\,\mathrm{d}x' \\ = \quad & \int \lim_{M\to\infty} (f\phi_n)(M,x') - (f\phi_n)(-M,x')\,\mathrm{d}x' \\ = \quad & 0. \end{aligned}$$

We may therefore conclude

$$\begin{aligned} \int \frac{\partial f}{\partial x_1}(x)\phi_n(x)\,\mathrm{d}x &= -\lim_{M\to\infty} \int_{[-M,M]^{n-1}} \int_{-M}^{M} f(x) \frac{\partial \phi_n}{\partial x_1}(x)\,\mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_n \\ &= \lim_{M\to\infty} \int_{[-M,M]^n} f(x) x_1 \phi_n(x)\,\mathrm{d}x \\ &= \int f(x) x_1 \phi_n(x)\,\mathrm{d}x, \end{aligned}$$

as desired. $\square$

Comment 4.2.7 The result of Lemma 4.2.6 is a version of a result sometimes known as Stein's lemma in the literature, see for example Zou et al. (2007). See Stein (1981), Lemma 1 and Lemma 2 for proofs of Stein's Lemma. $\circ$

Lemma 4.2.8. *Consider $G = g(W_u) \otimes 1_{[s,t]}$ with $g \in C_c^\infty(\mathbb{R}^n)$ and $u \in [0,T]^n$. There is some constant C such that for any $F \in \mathcal{S}$, $|\int G(DF)\,\mathrm{d}\eta| \leq C\|F\|_p$.*

Proof. By Lemma 4.1.2, we can assume that $F = f(W_u)$ for some $f \in C_p^\infty(\mathbb{R}^n)$, and we can assume that $0 < u_1 < \cdots < u_n$. Define $Ax = (\frac{x_1}{\sqrt{u_1}}, \frac{x_2-x_1}{\sqrt{u_2-u_1}}, \ldots, \frac{x_n-x_{n-1}}{\sqrt{u_n-u_{n-1}}})$, by Lemma 4.2.5 we then have

$$\sum_{k=1}^n \frac{\partial}{\partial x_k} f(W_u) 1_{[0,u_k]} = \sum_{k=1}^n \sum_{i=1}^n \frac{\partial}{\partial x_i} (f \circ A^{-1})(AW_u) A'_{ik} 1_{[0,u_k]},$$

where A' is the matrix for the linear mapping A. Defining $f' = f \circ A^{-1}$ and $g' = g \circ A^{-1}$, we then obtain

$$\begin{aligned}
\int GDF\,\mathrm{d}\eta &= \int G \sum_{k=1}^n \sum_{i=1}^n \frac{\partial f'}{\partial x_i}(AW_u) A'_{ik} 1_{[0,u_k]}\,\mathrm{d}\eta \\
&= \sum_{k=1}^n \sum_{i=1}^n A'_{ik} \int g(W_u) \frac{\partial f'}{\partial x_i}(AW_u) 1_{[s,t]} 1_{[0,u_k]}\,\mathrm{d}\eta \\
&= \sum_{k=1}^n \sum_{i=1}^n A'_{ik} \left(\int 1_{[s,t]} 1_{[0,u_k]}\,\mathrm{d}\lambda \right) E\left(g'(AW_u) \frac{\partial f'}{\partial x_i}(AW_u) \right).
\end{aligned}$$

In particular, then,

$$\left| \int GDF\,\mathrm{d}\eta \right| \leq nT\|A'\|_\infty \sum_{i=1}^n \left| E\left(g'(AW_u) \frac{\partial f'}{\partial x_i}(AW_u) \right) \right|.$$

We consider the mean values in the sums. Using Lemma 4.2.6, we obtain

$$\begin{aligned}
E\left(g'(AW_u) \frac{\partial f'}{\partial x_i}(AW_u) \right) &= \int g'(x) \frac{\partial f'}{\partial x_i}(x) \phi_n(x)\,\mathrm{d}x \\
&= \int \frac{\partial g' f'}{\partial x_i}(x) \phi_n(x)\,\mathrm{d}x - \int \frac{\partial g'}{\partial x_i}(x) f'(x) \phi_n(x)\,\mathrm{d}x \\
&= \int g'(x) f'(x) x_i \phi_n(x)\,\mathrm{d}x - \int \frac{\partial g'}{\partial x_i}(x) f'(x) \phi_n(x)\,\mathrm{d}x.
\end{aligned}$$

Recalling that $F = f'(AW_u)$, we can use Hölder's inequality to obtain

$$\begin{aligned}
\left|E\left(g'(AW_u)\frac{\partial f'}{\partial x_i}(AW_u)\right)\right| &= \left|\int g'(x)f'(x)x_i\phi_n(x)\,\mathrm{d}x - \int \frac{\partial g'}{\partial x_i}(x)f'(x)\phi_n(x)\,\mathrm{d}x\right| \\
&= \left|EF\left(g'(AW_u)(AW_u)_i - \frac{\partial g'}{\partial x_i}(AW_u)\right)\right| \\
&\leq \|F\|_p \left\|g'(AW_u)(AW_u)_i - \frac{\partial g'}{\partial x_i}(AW_u)\right\|_q.
\end{aligned}$$

Since g has compact support, so does g' and $\frac{\partial g'}{\partial x_i}$. Therefore, the latter factor in the above is finite. Putting

$$C = nT\|A'\|_\infty \sum_{i=1}^n \left\|g'(AW_u)(AW_u)_i - \frac{\partial g'}{\partial x_i}(AW_u)\right\|_q,$$

we have proved $|\int GDF\,\mathrm{d}\eta| \leq C\|F\|_p$, as desired. □

Theorem 4.2.9. *The operator D is closable from $\mathcal{S}$ to $\mathcal{L}^p(\Pi)$.*

Proof. Let a sequence $(F_n) \subseteq \mathcal{S}$ be given such that F_n converges to zero in $\mathcal{L}^p(\mathcal{F}_T)$ and such that DF_n converges in $\mathcal{L}^p(\Pi)$. We need to prove that the limit of DF_n is zero. Let ξ be the limit, By Theorem A.2.5 it will suffice to prove that that for any $G \in \mathcal{L}^q(\Pi)$, $\int \xi G\,\mathrm{d}\eta = 0$.

Let $\mathbb{H}$ be the class of elements $G \in \mathcal{L}^q(\Pi)$ such that $\int \xi G\,\mathrm{d}\eta = 0$. We will prove that $\mathbb{H} = \mathcal{L}^q(\Pi)$ first by proving that $\mathbf{b}\mathbb{H}$ contains all bounded mappings in $\mathcal{L}^q(\Pi)$. Obviously, then, $\mathbb{H}$ also contains all bounded mappings in $\mathcal{L}^q(\Pi)$ and we can finish the proof by a closedness argument.

Step 1: $\mathbf{b}\mathbb{H}$ contains all bounded mappings. We will use the monotone class theorem to prove that $\mathbf{b}\mathbb{H}$ contains all bounded mappings. To this end, note that $\mathbf{b}\mathbb{H}$ by dominated convergence is a monotone vector space. Defining $\mathcal{K}$ as the class of mappings $g(W_t) \otimes 1_{[s,t]}$ where $g \in C_c^\infty(\mathbb{R}^n)$ and $t \in [0,T]^n$, we know from Lemma 4.1.12 that $\mathcal{K}$ is a multiplicative class generating Π. If we can show that $\mathcal{K} \subseteq \mathbf{b}\mathbb{H}$, it will follow from the monotone class theorem B.1.2 that $\mathbf{b}\mathbb{H}$ contains all bounded mappings in $\mathcal{L}^q(\Pi)$.

Let $G \in \mathcal{K}$ be given with $G = g(W_u) \otimes 1_{[s,t]}$. Our ambition is to show $G \in \mathbb{H}$. By Lemma 4.2.8, there exists a constant C, such that for any $F \in \mathcal{S}$, it holds that

$|\int G(DF)\,\mathrm{d}\eta| \leq C\|F\|_p$. By the continuity of $F \mapsto \int FG\,\mathrm{d}\eta$, we then obtain

$$\left|\int \xi G\,\mathrm{d}\eta\right| = \left|\int G \lim_n DF_n\,\mathrm{d}\eta\right| = \lim_n \left|\int G(DF_n)\,\mathrm{d}\eta\right| \leq \lim_n C\|F_n\|_p = 0.$$

Thus, $\int G\xi\,\mathrm{d}\eta = 0$. We conclude $G \in \mathbf{b}\mathbb{H}$ and therefore $\mathcal{K} \subseteq \mathbb{H}$.

Step 2: $\mathbb{H}$ is closed. Since the bounded elements of $\mathcal{L}^q(\Pi)$ are dense in $\mathcal{L}^q(\Pi)$, it will follow that $\mathbb{H} = \mathcal{L}^q(\Pi)$ if we can prove that $\mathbb{H}$ is closed. To do so, assume that $(G_n) \subseteq \mathbb{H}$, converging to G. Then $\lim \|\xi G_n - \xi G\|_1 \leq \lim \|\xi\|_p \|G_n - G\|_q = 0$ by Hölder's inequality, so $\int \xi G\,\mathrm{d}\eta = \lim \int \xi G_n\,\mathrm{d}\eta = 0$, and $\mathbb{H}$ is closed. It follows that $\mathbb{H} = \mathcal{L}^q(\Pi)$ and we may finally conclude that $\xi = 0$ and therefore D is closable. $\square$

We may now conclude from Theorem A.7.3 that D has a unique closed extension from $\mathcal{S}$ to the subspace $\mathbb{D}_{1,p}$ of $\mathcal{L}^p(\mathcal{F}_T)$ defined as the $F \in \mathcal{L}^p(\mathcal{F}_T)$ such that there exists a sequence (F_n) in $\mathcal{S}$ converging to F such that DF_n converges as well, and in this case the value of DF is the limit of DF_n. D maps from $\mathbb{D}_{1,p}$ into $\mathcal{L}^p(\Pi)$. Note that endowing the space $\mathbb{D}_{1,p}$ with the norm given by $\|F\|_{1,p} = \|F\|_p + \|DF\|_p$, $\mathcal{S}$ is a dense subspace of $\mathbb{D}_{1,p}$. This is another way to think of the extension of D from $\mathcal{S}$ to $\mathbb{D}_{1,p}$.

We will mainly be interested in the properties of the operator from $\mathbb{D}_{1,2} \to \mathcal{L}^2(\Pi)$. We will consider this case in detail in the next section. First, we will spend a moment reviewing our progress. We have succeeded in defining, for any $p \geq 1$, a linear operator $D : \mathbb{D}_{1,p} \to \mathcal{L}^p(\Pi)$ with the properties that

- For $F \in \mathcal{S}$ with $F = f(W_t)$, $DF = \sum_{k=1}^n \frac{\partial f}{\partial x_k}(W_t) 1_{[0,t_k]}$.
- When considering $\mathbb{D}_{1,p}$ under the norm $\|\cdot\|_p$, D is closed.

These two properties are more or less all there is to the definition. The first property shows how to calculate D on $\mathcal{S}$, and the second property shows how to infer values of D outside of $\mathcal{S}$ from the values on $\mathcal{S}$. Note that our definition of $\mathbb{D}_{1,p}$ is different from the one seen in Nualart (2006). Nualart (2006) consideres the construction of the Malliavin derivative based on isonormal gaussian processes, which can be seen as a generalization of Brownian motion. This allows certain proofs the clarity of higher abstraction, but the downside is that the theory becomes harder to relate to real problems and furthermore that the development of the derivative operator is tied up to the L^2-structure of the concept of isonormal processes. For a brownian motion, the space of smooth functionals considered in Nualart (2006) would be

$$\left\{ f\left(\int_0^T h_1(t)\,\mathrm{d}W(t),\ldots,\int_0^T h_n(t)\,\mathrm{d}W(t)\right)\middle| h_1,\ldots,h_n \in \mathcal{L}^2[0,T], f\in C_p^\infty(\mathbb{R}^n)\right\},$$

and the space $\mathbb{D}_{1,p}$ would be the completion of this space with respect to the norm

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

where DF is considered as a stochastic variable with values in $L^2[0,T]$.

4.3 The Malliavin Derivative on $\mathbb{D}_{1,2}$

We will now consider specifically the properties of the Malliavin derivative as operator $D : \mathbb{D}_{1,2} \to \mathcal{L}^2(\Pi)$. We will prove two results allowing us easy manipulation of the Malliavin derivative: the chain rule and the integration-by-parts formula. Furthermore, we will prove that $\mathbb{D}_{1,2}$ has some degree of richness by showing that it contains all stochastic integrals of deterministic functions. We will also show how the derivative of such integrals are given.

We begin by proving the chain rule. The proof of the general case is rather long, so we split it in two. We begin by proving in Theorem 4.3.1 the chain rule as it is formulated in Nualart (2006) with transformations having bounded partial derivatives. After this, we prove in Theorem 4.3.5 a generalized chain rule.

Theorem 4.3.1. *Let $F = (F_1,\ldots,F_n)$, where $F_1,\ldots,F_n \in \mathbb{D}_{1,2}$, and let $\varphi \in C^1(\mathbb{R}^n)$ with bounded partial derivatives. Then $\varphi(F)\in\mathbb{D}_{1,2}$ and*

$$D\varphi(F) = \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(F)DF_k.$$

Comment 4.3.2 In the above, $\frac{\partial\varphi}{\partial x_k}(F)$ is a mapping on Ω and DF_k is a mapping on $[0,T]\times\Omega$. The multiplication is understood to be pointwise in the sense

$$\left(\frac{\partial\varphi}{\partial x_k}(F)DF_k\right)(t,\omega) = \frac{\partial\varphi}{\partial x_k}(F)(\omega)DF_k(t,\omega).$$

∘

Proof. The conclusion is well-defined since $\frac{\partial\varphi}{\partial x_k}$ is bounded, so $\varphi(F)$ is an element of $\mathcal{L}^2(\mathcal{F}_T)$ and the right-hand side is an element of $\mathcal{L}^2(\Pi)$. We prove the theorem in three steps:

1. The case where $\varphi \in C_p^\infty(\mathbb{R}^n)$ and $F \in \mathcal{S}^n$.
2. The case where $\varphi \in C_p^\infty(\mathbb{R}^n)$ with bounded partial derivatives and $F \in \mathbb{D}_{1,2}^n$.
3. The case where $\varphi \in C^1(\mathbb{R}^n)$ with bounded partial derivatives and $F \in \mathbb{D}_{1,2}^n$.

Step 1: C_p^∞ transformations of $\mathcal{S}$. Consider the case where $\varphi \in C_p^\infty(\mathbb{R}^n)$ and $F \in \mathcal{S}^n$. By Lemma 4.1.2, we can assume that the F_k are transformations of the same coordinates of the wiener process, $F_k = f_k(W_t)$ where $t \in [0,T]^m$. Putting $f = (f_1, \ldots, f_n)$, $\varphi \circ f \in C_p^\infty(\mathbb{R}^m)$. We have $\varphi(F) = (\varphi \circ f)(W_t)$, so $\varphi(F) \in \mathcal{S}$ and therefore

$$
\begin{aligned}
D\varphi(F) &= \sum_{i=1}^m \frac{\partial(\varphi \circ f)}{\partial x_i}(W_t) 1_{[0,t_i]} \\
&= \sum_{i=1}^m \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(f(W_t)) \frac{\partial f_k}{\partial x_i}(W_t) 1_{[0,t_i]} \\
&= \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(f(W_t)) \sum_{i=1}^m \frac{\partial f_k}{\partial x_i}(W_t) 1_{[0,t_i]} \\
&= \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(F) DF_k.
\end{aligned}
$$

Step 2: Special C_p^∞ transformations of $\mathbb{D}_{1,2}$. We next consider the case where $F \in \mathbb{D}_{1,2}^n$ and $\varphi \in C_p^\infty(\mathbb{R}^n)$ with bounded partial derivatives. Since D is closed, we know that there exists $F_k^j \subseteq \mathcal{S}$ such that F_k^j converges to F_k in $\mathcal{L}^2(\mathcal{F}_T)$ and DF_k^j converges to DF_k in $\mathcal{L}^2(\Pi)$. Then, by what we have already shown, $\varphi(F^j) \in \mathcal{S}$ and $D\varphi(F^j) = \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(F^j) DF_k^j$. Our plan is to show that $\varphi(F^j)$ converges to $\varphi(F)$ in $\mathcal{L}^2(\mathcal{F}_T)$ and that $D\varphi(F^j)$ converges to $\sum_{i=1}^n \frac{\partial\varphi}{\partial x_i}(F_i) DF_i$ in $\mathcal{L}^p(\Pi)$. By the closedness of D, this will imply that $\varphi(F) \in \mathbb{D}_{1,2}$ and $D\varphi(F) = \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(F) DF_k$.

By assumption, F_k^j converges to F_k in $\mathcal{L}^2(\mathcal{F}_T)$ and DF_k^j converges in $\mathcal{L}^2(\Pi)$ to DF_k. By picking a subsequence we may assume that we also have almost sure convergence. By the mean value theorem, Lemma A.2.2,

$$
\left\|\varphi(F^j) - \varphi(F)\right\|_2 = \left\| \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(\xi)(F_k^j - F_k) \right\|_2 \le \max_{k \le n} \left\| \frac{\partial\varphi}{\partial x_k} \right\|_\infty \left\| \sum_{k=1}^n (F_k^j - F_k) \right\|_2,
$$

so $\varphi(F^j)$ converges to $\varphi(F)$ in $\mathcal{L}^2(\mathcal{F}_T)$. To show the other convergence, note that

$$\begin{aligned}\left\|D\varphi(F^j) - \sum_{i=1}^n \frac{\partial \varphi}{\partial x_i}(F)DF_i\right\|_2 &= \left\|\sum_{i=1}^n \frac{\partial \varphi}{\partial x_i}(F^j)DF_i^j - \sum_{i=1}^n \frac{\partial \varphi}{\partial x_i}(F)DF_i\right\|_2 \\ &\leq \sum_{i=1}^n \left\|\frac{\partial \varphi}{\partial x_i}(F^j)DF_i^j - \frac{\partial \varphi}{\partial x_i}(F)DF_i\right\|_2,\end{aligned}$$

where, for each $i \leq n$,

$$\begin{aligned}&\left\|\frac{\partial \varphi}{\partial x_i}(F^j)DF_i^j - \frac{\partial \varphi}{\partial x_i}(F)DF_i\right\|_2 \\ \leq\ &\left\|\frac{\partial \varphi}{\partial x_i}(F^j)(DF_i^j - DF_i)\right\|_2 + \left\|\left(\frac{\partial \varphi}{\partial x_i}(F^j) - \frac{\partial \varphi}{\partial x_i}(F)\right)DF_i\right\|_2 \\ \leq\ &\left\|\frac{\partial \varphi}{\partial x_i}\right\|_\infty \left\|DF_i^j - DF_i\right\|_2 + \left\|\left(\frac{\partial \varphi}{\partial x_i}(F^j) - \frac{\partial \varphi}{\partial x_i}(F)\right)DF_i\right\|_2.\end{aligned}$$

Here, the first term converges to zero by assumption. Regarding the second term, we know that $F^j \xrightarrow{\text{a.s.}} F$, and by continuity, $\frac{\partial \varphi}{\partial x_i}(F^j) \xrightarrow{\text{a.s.}} \frac{\partial \varphi}{\partial x_i}(F)$, bounded by a constant since the partial derivatives are bounded. Dominated convergence therefore yields that the second term in the above converges to zero as well. We conclude that $\varphi(F) \in \mathbb{D}_{1,2}$ and $D\varphi(F) = \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k$.

Step 3: C^1 transformations of $\mathbb{D}_{1,2}$. Now assume $\varphi \in C^1$ with bounded partial derivatives, and let $F \in \mathbb{D}_{1,2}$. From Lemma A.4.3, we know that there exists functions $\varphi_n \in C_p^\infty(\mathbb{R}^n)$ such that φ_n converges uniformly to φ, the partial derivatives of φ_n converges pointwise to the partial derivatives of φ and $\|\frac{\partial \varphi_n}{\partial x_k}\|_\infty \leq \|\frac{\partial \varphi}{\partial x_k}\|_\infty$. By what we already have shown, $\varphi_n(F) \in \mathbb{D}_{1,2}$ and $D\varphi_n(F) = \sum_{k=1}^n \frac{\partial \varphi_n}{\partial x_k}(F)DF_k$. We want to argue that $\varphi_n(F)$ converges to $\varphi(F)$ in $\mathcal{L}^2(\mathcal{F}_T)$ and that $D\varphi_n(F)$ converges to $\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k$ in $\mathcal{L}^2(\Pi)$. As in the previous step, by closedness of D, this will yield that $\varphi(F) \in \mathbb{D}_{1,2}$ and that the derivative is given by $D\varphi(F) = \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k$.

Clearly, since $\|\varphi(F) - \varphi_n(F)\|_2 \leq \|\varphi - \varphi_n\|_\infty$, $\varphi_n(F)$ converges to $\varphi(F)$. To show the second limit statement, we note

$$\left\|D\varphi_n(F) - \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k\right\|_2 \leq \sum_{k=1}^n \left\|\left(\frac{\partial \varphi_n}{\partial x_k}(F) - \frac{\partial \varphi}{\partial x_k}(F)\right)DF_k\right\|_2.$$

Since $\frac{\partial \varphi_n}{\partial x_k}$ converges pointwise to $\frac{\partial \varphi}{\partial x_k}$, clearly $\frac{\partial \varphi_n}{\partial x_k}(F) \xrightarrow{\text{a.s.}} \frac{\partial \varphi}{\partial x_k}(F)$, and this convergence is bounded. The dominated convergence theorem therefore yields that the above sum converges to zero. $\square$

The chain rule of Theorem 4.3.1 is sufficient in many cases, but it will be convenient in the following to have an extension which does not require the boundedness of the partial derivatives of the transformation, but only requires enough integrability to ensure that the statement of the chain rule is well-defined. This is what we now set out to prove.

Lemma 4.3.3. *Let $M > 0$ and $\varepsilon > 0$. There exists a mapping $f \in C_c^\infty(\mathbb{R}^n)$ with partial derivatives bounded by $1 + \varepsilon$ such that $[-M, M]^n \prec f \prec (-(M+1), M+1)^n$.*

Proof. Let $\varepsilon > 0$ be given, put $M_{\varepsilon-} = M - \varepsilon$ and $M_{\varepsilon+} = M + \varepsilon$. Define the sets $K_\varepsilon = [-M_{\varepsilon+}, M_{\varepsilon+}]^n$ and $V_\varepsilon = (-(M_{\varepsilon-} + 1), M_{\varepsilon-} + 1)^n$. We will begin by identifying a d_∞-Lipschitz mapping $g \in C_c(\mathbb{R}^n)$ with $K_\varepsilon \prec g \prec V_\varepsilon$.

To this end, we put $g(x) = \min\{\frac{1}{1-2\varepsilon} d_\infty(x, V_\varepsilon^c), 1\}$. Clearly, then g is continuous and maps into $[0, 1]$. Since g is zero on V_ε^c and V_ε has compact closure, g has compact support. And if $x \in K_\varepsilon$, we obtain

$$g(x) = \min\left\{\frac{1}{1-2\varepsilon} d_\infty(x, V_\varepsilon^c), 1\right\} \geq \min\left\{\frac{1}{1-2\varepsilon} d_\infty(K_\varepsilon, V_\varepsilon^c), 1\right\} = 1,$$

We conclude $K_\varepsilon \prec g \prec V_\varepsilon$. We know that $|d_\infty(x, V^c) - d_\infty(y, V^c)| \leq d_\infty(x, y)$ by the inverse triangle inequality, so $x \mapsto d_\infty(x, V^c)$ is d_∞-Lipschitz continuous with Lipschitz constant 1. Therefore, the mapping $x \mapsto \frac{1}{1-2\varepsilon} d_\infty(x, V^c)$ is d_∞-Lipschitz continuous with Lipschitz constant $\frac{1}{1-2\varepsilon}$. Since $x \mapsto \min\{x, 1\}$ is a contraction, g is d_∞-Lipschitz continuous with Lipschitz constant $\frac{1}{1-2\varepsilon}$.

Now let (ψ_ε) be a Dirac family with respect to $\|\cdot\|_\infty$ and let $f = g * \psi_\varepsilon$. We claim that f satisfies the properties in the lemma. Let $K = [-M, M]^n$ and $V = (-(M+1), M+1)^n$, we want to show that $f \in C_c^\infty(\mathbb{R}^n)$, that $K \prec f \prec V$ and that f has partial derivatives bounded by $\frac{1}{1-2\varepsilon}$. By Lemma A.3.2, f is infinitely differentiable, and it is clear that f takes values in $[0, 1]$. If $x \in V^c$, we have $x - y \in V_\varepsilon^c$ for any y with $\|y\|_\infty \leq \varepsilon$. Therefore,

$$f(x) = \int g(x-y)\psi_\varepsilon(y)\,\mathrm{d}y = \int 1_{(\|y\|_\infty \leq \varepsilon)} g(x-y)\psi_\varepsilon(y)\,\mathrm{d}y = 0.$$

Likewise, if $x \in K$, then $x - y \in K_\varepsilon$ for any y with $\|y\|_\infty \leq \varepsilon$ and thus

$$f(x) = \int g(x-y)\psi_\varepsilon(y)\,\mathrm{d}y = \int 1_{(\|y\|_\infty \leq \varepsilon)} g(x-y)\psi_\varepsilon(y)\,\mathrm{d}y = 1.$$

We have now shown $K \prec f \prec V$. Since V is compact closure, it follows in particular that f has compact support. It remains to prove the bound on the partial derivatives.

To this end, we note

$$
\begin{aligned}
|f(x) - f(y)| &= \left| \int g(x-z)\psi_\varepsilon(z)\,\mathrm{d}z - \int g(y-z)\psi_\varepsilon(z)\,\mathrm{d}z \right| \\
&\leq \int |g(x-z) - g(y-z)|\psi_\varepsilon(z)\,\mathrm{d}z \\
&\leq \frac{d_\infty(x,y)}{1-2\varepsilon}.
\end{aligned}
$$

We therefore obtain, with e_k denoting the unit vector in the k'th direction,

$$
\left|\frac{\partial f}{\partial x_k}(x)\right| = \lim_{h\to 0^+} \frac{|f(x+he_k) - f(x)|}{h} \leq \lim_{h \to 0^+} \frac{1}{1-2\varepsilon}\frac{\|he_k\|_\infty}{h} = \frac{1}{1-2\varepsilon}.
$$

We have now identified a mapping $f \in C_c^\infty(\mathbb{R}^n)$ with $K \prec f \prec V$ and partial derivatives bounded by $\frac{1}{1-2\varepsilon}$. Since $\varepsilon > 0$ was arbitrary, we conclude that there exists $f \in C_c^\infty(\mathbb{R}^n)$ with $K \prec f \prec V$ and partial derivatives bounded by $1+\varepsilon$ for any $\varepsilon > 0$. This shows the lemma. □

Comment 4.3.4 Instead of explicitly constructing the Lipschitz mapping, we could also have applied the extended Urysohn's Lemma of Theorem C.2.2. This would have yielded a weaker bound for the partial derivatives, but the nature of the argument would be more general. ∘

Theorem 4.3.5 (Chain rule). *Let $F = (F_1, \ldots, F_n)$, where $F_1, \ldots, F_n \in \mathbb{D}_{1,2}$, and let $\varphi \in C^1(\mathbb{R}^n)$. Assume $\varphi(F) \in \mathcal{L}^2(\mathcal{F}_T)$ and $\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F) DF_k \in \mathcal{L}^2(\Pi)$. Then $\varphi(F) \in \mathbb{D}_{1,2}$ and $D\varphi(F) = \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(F)DF_k$.*

Proof. We first prove the corollary in the case where φ is bounded, and then extend to general φ.

Step 1: The bounded case. Assume that $\varphi \in C^1(\mathbb{R}^n)$ is bounded and that we have $\sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(F)DF_k \in \mathcal{L}^2(\Pi)$. In this case, obviously $\varphi(F) \in \mathcal{L}^2(\mathcal{F}_T)$. We will employ the bump functions analyzed in Lemma 4.3.3. Let $\varepsilon > 0$ and define the sets K_m and V_m by $K_m = [-m, m]^n$ and $V_m = (-m-1, m+1)^n$. By Lemma 4.3.3, there exists mappings (c_m) in $C_c^\infty(\mathbb{R}^n)$ with $K_m \prec c_m \prec V_m$ such that the partial derivatives of c_m are bounded by $1+\varepsilon$.

Define φ_m by putting $\varphi_m(x) = c_m(x)\varphi(x)$. We will argue that $\varphi_m \in C^1(\mathbb{R}^n)$ with bounded partial derivatives. It is clear that φ_m is continuously differentiable, and

$$
\frac{\partial \varphi_m}{\partial x_k}(x) = \frac{\partial c_m}{\partial x_k}(x)\varphi(x) + c_m(x)\frac{\partial\varphi}{\partial x_k}(x).
$$

Since $c_m \in C_c^\infty(\mathbb{R}^n)$, both c_m and $\frac{\partial c_m}{\partial x_k}$ has compact support. It follows that $\frac{\partial \varphi_m}{\partial x_k}$ has compact support, and since it is also continuous, it is bounded. We can therefore apply the ordinary chain rule of Theorem 4.3.1 to conclude that $\varphi_m(F) \in \mathbb{D}_{1,2}$ and $D\varphi_m(F) = \sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F) DF_m$. We will use the closedness of D to extend this result to $\varphi(F)$. We will show that $\varphi_m(F)$ converges to $\varphi(F)$ in $\mathcal{L}^2(\mathcal{F}_T)$ and that $\sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F) DF_k$ converges to $\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F) DF_k$ in $\mathcal{L}^2(\Pi)$. Closedness of D will then yield $\varphi(F) \in \mathbb{D}_{1,2}$ and $D\varphi(F) = \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F) DF_k$.

Since c_m converges pointwise to 1, φ_m converges pointwise to φ. Therefore, we obtain $\varphi_m(F) \xrightarrow{\text{a.s.}} \varphi(F)$. Because $\|c_m\|_\infty \le 1$, dominated convergence yields that $\varphi_m(F)$ converges in $\mathcal{L}^2(\mathcal{F}_T)$ to $\varphi(F)$. To show the corresponding result for the derivatives, first note that $\frac{\partial c_m}{\partial x_k}$ is zero on K_m°. Since c_m is one on K_m°, we find that $\frac{\partial \varphi_m}{\partial x_k}$ and $\frac{\partial \varphi}{\partial x_k}$ are equal on K_m°. We therefore obtain

$$\begin{aligned}
&\left\| \sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F) DF_k - \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F) DF_k \right\|_2 \\
= &\left\| \sum_{k=1}^n 1_{(K_m^\circ)^c}(F) \left(\frac{\partial \varphi_m}{\partial x_k}(F) - \frac{\partial \varphi}{\partial x_k}(F) \right) DF_k \right\|_2 \\
= &\left\| \sum_{k=1}^n 1_{(K_m^\circ)^c}(F) \left(\frac{\partial c_m}{\partial x_k}(F)\varphi(F) + (c_m(F)-1) \frac{\partial \varphi}{\partial x_k}(F) \right) DF_k \right\|_2 \\
\le &\left\| \sum_{k=1}^n 1_{(K_m^\circ)^c}(F) \frac{\partial c_m}{\partial x_k}(F)\varphi(F) DF_k \right\|_2 + \left\| \sum_{k=1}^n 1_{(K_m^\circ)^c}(F)(c_m(F)-1) \frac{\partial \varphi}{\partial x_k}(F) DF_k \right\|_2 .
\end{aligned}$$

We wish to show that each of these terms tend to zero by dominated convergence. Considering the first term, by definition of c_m we obtain

$$\left| \sum_{k=1}^n 1_{(K_m^\circ)^c}(F) \frac{\partial c_m}{\partial x_k}(F)\varphi(F) DF_k \right| \le (1+\varepsilon)\|\varphi\|_\infty \sum_{k=1}^n |DF_k|,$$

which is square-integrable. Likewise, for the second term we have the square-integrable bound

$$\begin{aligned}
\left| \sum_{k=1}^n 1_{(K_m^\circ)^c}(F)(c_m(F)-1) \frac{\partial \varphi}{\partial x_k}(F) DF_k \right| &= |1_{(K_m^\circ)^c}(F)(c_m(F)-1)| \left| \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F) DF_k \right| \\
&\le \left| \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F) DF_k \right| .
\end{aligned}$$

Since K_m° increases to $\mathbb{R}^n$, by dominated convergence using the two bounds obtained above, we find that both norms tends to zero and therefore we may finally conclude that $\sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F)DF_k$ tends to $\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k$ in $\mathcal{L}^2(\Pi)$. Closedness of D then implies that $\varphi(F) \in \mathbb{D}_{1,2}$ and $D\varphi(F) = \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k$.

Step 2: The general case. Now let $\varphi \in C^1(\mathbb{R}^n)$ be arbitrary and assume that $\varphi(F) \in \mathcal{L}^2(\mathcal{F}_T)$ and $\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k \in \mathcal{L}^2(\Pi)$. We still consider some fixed $\varepsilon > 0$. Define the sets $K_m = [-m, m]$ and $V_m = [-(m+1), m+1]$ and let c_m be the Lipschitz element of $C_c^\infty(\mathbb{R})$ with $K_m \prec c_m \prec V_m$ that exists by Lemma 4.3.3 with Lipschitz constant $1+\varepsilon$. Let C_m be the antiderivative of c_m which is zero at zero, given by $C_m(x) = \int_0^x c_m(y)\,\mathrm{d}y$. Note that since c_m is bounded by one and is zero outside of V_m, $\|C_m\|_\infty \leq m+1$. In particular, C_m is bounded. Defining $\varphi_m(x) = C_m(\varphi(x))$, it is then clear that φ_m is bounded. Furthermore, $\varphi_m \in C^1(\mathbb{R}^n)$ and

$$\frac{\partial \varphi_m}{\partial x_k}(x) = c_m(\varphi(x))\frac{\partial \varphi}{\partial x_k}(x).$$

We therefore have

$$\left|\sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F)DF_k\right| = \left|c_m(\varphi(F))\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k\right| \leq \left|\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k\right|,$$

so $\sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F)DF_k \in \mathcal{L}^2(\Pi)$. Thus, φ_m is covered by the previous step of the proof, and we may conclude that $\varphi_m(F) \in \mathbb{D}_{1,2}$ and $D\varphi_m(F) = \sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F)DF_k$. As in the previous step, we will extend this result to φ by applying the closedness of D. We will therefore need to prove convergence in $\mathcal{L}^2(\mathcal{F}_T)$ of $\varphi_m(F)$ to $\varphi(F)$ and to prove convergence in $\mathcal{L}^2(\Pi)$ of $\sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F)DF_k$ to $\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k$.

To this end, note that since c_m is bounded by one, $|C_m(x)| \leq \int_0^x |c_m(y)|\,\mathrm{d}y \leq |x|$, and for $x \in [-m, m]$, $C_m(x) = \int_0^x c_m(y)\,\mathrm{d}y = x$. Therefore, $\varphi_m(x) = \varphi(x)$ on $[-m, m]$ and

$$\begin{aligned}\|\varphi_m(F) - \varphi(F)\|_2 &= \|1_{[-m,m]^c}(F)(\varphi_m(F) - \varphi(F))\|_2 \\ &= \|1_{[-m,m]^c}(F)(C_m(\varphi(F)) - \varphi(F))\|_2.\end{aligned}$$

Since $|C_m(\varphi(F)) - \varphi(F)| \leq |C_m(\varphi(F))| + |\varphi(F)| \leq 2|\varphi(F)|$ by the contraction property of C_m, we conclude by dominated convergence that the above tends to zero, so $\varphi_m(F)$ tends to $\varphi(F)$ in $\mathcal{L}^2(\mathcal{F}_T)$. Likewise, for the derivatives we obtain

$$\begin{aligned}&\left\|\sum_{k=1}^n \frac{\partial \varphi_m}{\partial x_k}(F)DF_k - \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k\right\|_2 \\ =\ &\left\|\sum_{k=1}^n c_m(\varphi(F))\frac{\partial \varphi}{\partial x_k}(F)DF_k - \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k\right\|_2 \\ =\ &\left\|(c_m(\varphi(F)) - 1)\sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F)DF_k\right\|_2,\end{aligned}$$

and since $c_m - 1$ tends to zero, bounded by the constant one, we conclude by dominated convergence that the above tends to zero. By the closedness of D, we finally obtain $\varphi(F) \in \mathbb{D}_{1,2}$ and $D\varphi(F) = \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(F) DF_k$, proving the theorem. □

Corollary 4.3.6. *Let $F, G \in \mathbb{D}_{1,2}$. If the integrability conditions $FG \in \mathcal{L}^2(\mathcal{F}_T)$ and $(DF)G + F(DG) \in \mathcal{L}^2(\Pi)$ hold, $FG \in \mathbb{D}_{1,2}$ and $D(FG) = (DF)G + F(DG)$.*

Proof. Putting $\varphi(x, y) = xy$, the chain rule of Theorem 4.3.5 yields that $FG \in \mathbb{D}_{1,2}$ and $DFG = \frac{\partial \varphi}{\partial x_1}(F, G) DF + \frac{\partial \varphi}{\partial x_2}(F, G) DG = G(DF) + F(DG)$. □

We are now done with the proof of the chain rule for the Malliavin derivative. Next, we prove that stochastic integrals with deterministic integrands are in the domain of D. This will allow us to reconcile our definition of the Malliavin derivative with that found in Nualart (2006).

Lemma 4.3.7. *Let $h \in \mathcal{L}^2[0, T]$. It then holds that $\int_0^T h(t)\,\mathrm{d}W(t) \in \mathbb{D}_{1,2}$, and $D \int_0^T h(t)\,\mathrm{d}W(t) = h$.*

Proof. First assume that h is continuous. We define $h_n(t) = \sum_{k=1}^n h(\frac{k}{n}) 1_{(\frac{k-1}{n}, \frac{k}{n}]}(t)$. Since h is continuous, h is bounded and it follows that $h_n \in \mathbf{b}\mathcal{E}$. We may therefore conclude $\int_0^T h_n(t)\,\mathrm{d}W(t) \in \mathcal{S}$ and

$$\begin{aligned} D_s \int_0^T h_n(t)\,\mathrm{d}W(t) &= D_s \sum_{k=1}^n h\left(\frac{k}{n}\right)\left(W\left(\frac{k}{n}\right) - W\left(\frac{k-1}{n}\right)\right) \\ &= \sum_{k=1}^n h\left(\frac{k}{n}\right) 1_{(\frac{k-1}{n}, \frac{k}{n}]}(s) \\ &= h_n(s). \end{aligned}$$

Next, note that h_n converges pointwise to h for all $t \in (0, T]$. By boundedness of h, h_n converges in $\mathcal{L}^2[0, T]$ to h. By continuity of the stochastic integral, $\int_0^T h_n(t)\,\mathrm{d}W(t)$ converges in $\mathcal{L}^2(\mathcal{F}_T)$ to $\int_0^T h(t)\,\mathrm{d}W(t)$. Closedness then yields $\int_0^T h(t)\,\mathrm{d}W(t) \in \mathbb{D}_{1,2}$ and $D \int_0^T h(t)\,\mathrm{d}W(t) = h$, as desired. This shows the result for continuous h.

Next, consider a general $h \in \mathcal{L}^2[0, T]$. There exists continuous h_n converging in $\mathcal{L}^2[0, T]$ to h. By the previous step, $D \int_0^T h_n(t)\,\mathrm{d}W_t = h_n$. Thus, $\int_0^T h_n(t)\,\mathrm{d}W_t$ converges to $\int_0^T h(t)\,\mathrm{d}W_t$ and $D \int_0^T h_n(t)\,\mathrm{d}W_t$ converges to h. We conclude that $\int_0^T h(t)\,\mathrm{d}W_t \in \mathbb{D}_{1,2}$ and $D \int_0^T h(t)\,\mathrm{d}W_t = h$. □

In the following, we will for brevity denote the the Brownian stochastic integral operator on $[0,T]$ by θ, that is, whenever $X \in \mathcal{L}^2(\Pi)$ is progressively measurable, we will write $\theta(X) = \theta X = \int_0^T X_s \, \mathrm{d}W_s$. Furthermore, if $X_1, \ldots, X_n \in \mathcal{L}^2(\Pi)$ and we put $X = (X_1, \ldots, X_n)$, we will write θX for $(\theta X_1, \ldots, \theta X_n)$.

Now consider $h_1, \ldots, h_n \in \mathcal{L}^2[0,T]$ and let $\varphi \in C_p^\infty(\mathbb{R}^n)$. Since θh has a normal distribution by Lemma 3.7.4, it follows that $\varphi(\theta h) \in \mathcal{L}^2(\Pi)$. Furthermore, since h is deterministic,

$$\left\| \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(\theta h) h_k \right\|_2 \le \sum_{k=1}^n \left\| \frac{\partial \varphi}{\partial x_k}(\theta h) h_k \right\|_2 = \sum_{k=1}^n \left\| \frac{\partial \varphi}{\partial x_k}(\theta h) \right\|_2 \|h_k\|_2 ,$$

which is finite. By Lemma 4.3.7 and Theorem 4.3.5, we conclude that $\varphi(\theta h) \in \mathbb{D}_{1,2}$ and $D\varphi(\theta h) = \sum_{k=1}^n \frac{\partial \varphi}{\partial x_k}(\theta h) h_k$. This result shows that our method of introducing the Malliavin derivative only for variables $\varphi(W_t)$ with $\varphi \in C_p^\infty(\mathbb{R}^n)$ yields the same result as the method in Nualart (2006), where the Malliavin derivative is introduced for variables $\varphi(\theta h)$ with $\varphi \in C_p^\infty(\mathbb{R}^n)$. The two methods are thus equivalent.

We let $\mathcal{S}_h$ denote the space of variables $\varphi(\theta h)$, where $h = (h_1, \ldots, h_n)$, $h_k \in \mathcal{L}^2[0,T]$ and $\varphi \in C_p^\infty(\mathbb{R}^n)$. Clearly, $\mathcal{S} \subseteq \mathcal{S}_h$. Our final goal of this section will be to prove a few versions of the integration-by-parts formula. This formula will be very useful in the next section, where we develop the Hilbert space theory of the Malliavin calculus. If $X \in \mathcal{L}^2(\Pi)$ and $h \in \mathcal{L}^2[0,T]$, we will write $\langle X, h \rangle_{[0,T]}(\omega) = \int X(\omega,t) h(t) \, \mathrm{d}t$.

Theorem 4.3.8 (Integration-by-parts formula). *Let $h \in \mathcal{L}^2[0,T]$ and $F \in \mathbb{D}_{1,2}$. Then $E\langle DF, h\rangle_{[0,T]} = EF(\theta h)$, where $F(\theta h)(\omega) = F(\omega)(\theta h)(\omega)$.*

Proof. We first check that the conclusion is well-defined. Considering h as an element of $\mathcal{L}^2(\Pi)$, the left-hand side is simply the inner product in $\mathcal{L}^2(\Pi)$ of two elements in $\mathcal{L}^2(\Pi)$. And since F and θh are in $\mathcal{L}^2(\mathcal{F}_T)$, $F(\theta h) \in \mathcal{L}^1(\mathcal{F}_T)$. Thus, all the integrals are well-defined.

By linearity, it will suffice to prove the result in the case where h has unit norm. First consider the case where $F \in \mathcal{S}_h$ with $F = f(\theta g)$, $g \in (\mathcal{L}^2[0,T])^n$. By using a linear transformation, we can assume without loss of generality that $g_1 = h$ and that $g_1, \ldots, g_n$ are orthonormal. We then obtain

$$E\langle DF, h\rangle_{[0,T]} = E\left\langle \sum_{k=1}^n \frac{\partial f}{\partial x_k}(\theta g) g_k, g_1 \right\rangle_{[0,T]} = E \sum_{k=1}^n \frac{\partial f}{\partial x_k}(\theta g) \langle g_k, g_1\rangle_{[0,T]} = E \frac{\partial f}{\partial x_1}(\theta g).$$

Since $g_1, \ldots, g_n$ are orthonormal, θg is standard normally distributed. Therefore, we can use Lemma 4.2.6 to obtain

$$E\frac{\partial f}{\partial x_1}(\theta g) = \int \frac{\partial f}{\partial x_1}(y)\phi_n(y)\,\mathrm{d}y = \int f(y)y_1\phi_n(y)\,\mathrm{d}y = Ef(\theta g)(\theta g)_1 = EF\theta(h).$$

This proves the claim in the case where $F \in \mathcal{S}_h$. Now consider the case where $F \in \mathbb{D}_{1,2}$. Let F_n be a sequence in $\mathcal{S}_h$ such that F_n converges to F and DF_n converges to DF, this is possible by the closedness definition of D and our earlier observation that $\mathcal{S} \subseteq \mathcal{S}_h$. By continuity of the inner products, we find

$$\begin{aligned} E\langle DF, h\rangle_{[0,T]} &= E\langle \lim DF_n, h\rangle_{[0,T]} \\ &= \lim E\langle DF_n, h\rangle_{[0,T]} \\ &= \lim EF_n\theta(h) \\ &= E\lim F_n\theta(h) \\ &= EF\theta(h), \end{aligned}$$

as desired. □

Theorem 4.3.8 is the fundamental result known as the integration-by-parts formula, even though it may not seem to bear much similarity to the integration-by-parts formula of ordinary calculus. In practice, we will not really use the integration-by-parts formula in the form given in Theorem 4.3.8, rather we will be using the following two corollaries.

Corollary 4.3.9. *Let $F, G \in \mathbb{D}_{1,2}$ with $FG \in \mathcal{L}^2(\mathcal{F}_T)$ and $F(DG)+G(DF) \in \mathcal{L}^2(\Pi)$. Then*

$$E(G\langle DF, h\rangle_{[0,T]}) = E\left(FG(\theta h)\right) - EF\langle DG, h\rangle_{[0,T]}$$

Comment 4.3.10 Note that the left-hand side of the formula above also can be written as $\int\int_{[0,T]} DF(\omega,t)G(\omega)h(t)\,\mathrm{d}t\,\mathrm{d}P$. This makes this formula useful for identifying conditional expectations, which in the Hilbert space context corresponds to orthogonal projections. ∘

Proof. By Corollary 4.3.6, $FG \in \mathbb{D}_{1,2}$ and $D(FG) = F(DG) + G(DF)$. Therefore, Theorem 4.3.8 yields

$$\begin{aligned} E(FG(\theta h)) &= E\langle D(FG), h\rangle_{[0,T]} \\ &= E\langle F(DG), h\rangle_{[0,T]} + E\langle G(DF), h\rangle_{[0,T]} \\ &= EF\langle DG, h\rangle_{[0,T]} + EG\langle DF, h\rangle_{[0,T]}, \end{aligned}$$

and rearranging yields the result. □

Corollary 4.3.11. *Let $F \in \mathbb{D}_{1,2}$ and $G \in \mathcal{S}_h$. Then*

$$E(G\langle DF, h\rangle_{[0,T]}) = E\left(FG(\theta h)\right) - EF\langle DG, h\rangle_{[0,T]}$$

Comment 4.3.12 The point of considering $G \in \mathcal{S}_h$ is to avoid the integrability requirements of Corollary 4.3.9. Particularly important for our later applications is that we avoid the requirement $FG \in \mathcal{L}^2(\mathcal{F}_T)$. ∘

Proof. Let F_n be a sequence in $\mathcal{S}_h$ such that F_n converges to F and DF_n converges to DF. Then F_n and G are both in $\mathcal{S}_h$, so F_n and G are both C_p^∞ transformations of coordinates of the Wiener process. Since C_p^∞ transformations and their partial derivatives have polynomial growth, we conclude $F_nG \in \mathcal{L}^2(\mathcal{F}_T)$ and $F_n(DG)+G(DF_n) \in \mathcal{L}^2(\Pi)$. From Lemma 4.3.9, we then have

$$E(G\langle DF_n, h\rangle_{[0,T]}) = E\left(F_nG(\theta h)\right) - EF_n\langle DG, h\rangle_{[0,T]}.$$

We want to show that each of these terms converge as F_n converges to F.

First term. We begin by considering the term on the left-hand side. Because we have $G \in \mathcal{L}^2(\mathcal{F}_T)$ and $h \in \mathcal{L}^2[0,T]$, $G \otimes h \in \mathcal{L}^2(\Pi)$. DF_n converges to DF in $\mathcal{L}^2(\Pi)$, so by the Cauchy-Schwartz inequality, $DF_n(G\otimes h)$ converges to $DF(G\otimes h)$ in $\mathcal{L}^1(\Pi)$. Noting that

$$\begin{aligned} E\left|G\langle DF_n, h\rangle_{[0,T]} - G\langle DF, h\rangle_{[0,T]}\right| &= E\left|\int_0^T (G\otimes h)DF_n \,\mathrm{d}\lambda - \int_0^T (G\otimes h)DF \,\mathrm{d}\lambda\right| \\ &\leq \left\|DF_n(G\otimes h) - DF(G\otimes h)\right\|_1, \end{aligned}$$

we obtain that $G\langle DF_n, h\rangle_{[0,T]}$ converges in $\mathcal{L}^1(\mathcal{F}_T)$ to $G\langle DF, h\rangle_{[0,T]}$, and therefore $E(G\langle DF_n, h\rangle_{[0,T]})$ converges to $E(G\langle DF, h\rangle_{[0,T]})$, as desired.

Second term. Next, consider the first term on the right-hand side. Since $G(\theta h)$ is in $\mathcal{L}^2(\mathcal{F}_T)$, the Cauchy-Schwartz inequality implies that $F_nG(\theta h)$ tends to $FG(\theta h)$ in $\mathcal{L}^1(\mathcal{F}_T)$. Therefore $E(F_nG(\theta h))$ converges to $E(FG(\theta h))$.

Third term. Finally, consider the second term on the right-hand side. We proceed as for the term on the left-hand side and first note that since $F_n \in \mathcal{L}^2(\mathcal{F}_T)$ and $h \in \mathcal{L}^2[0,T]$, $F_n \otimes h \in \mathcal{L}^2(\Pi)$. We also obtain $\|F_n\otimes h - F\otimes h\|_2^2 = \|F_n - F\|_2^2\|h\|_2^2$, showing that $F_n \otimes h$ converges in $\mathcal{L}^2(\Pi)$ to $F \otimes h$. Therefore, $(F_n \otimes h)DG$ converges in $\mathcal{L}^1(\Pi)$ to $(F \otimes h)DG$. Since we have, as for the first term considered,

$$\begin{aligned} E\left|F_n\langle DG,h\rangle_{[0,T]} - F\langle DG,h\rangle_{[0,T]}\right| &= E\left|\int_0^T (F_n\otimes h)DG\,\mathrm{d}\lambda - \int_0^T (F\otimes h)DG\,\mathrm{d}\lambda\right| \\ &\leq \|(F_n\otimes h)DG - (F\otimes h)DG\|_1 \end{aligned}$$

we conclude that $F_n\langle DG,h\rangle_{[0,T]}$ tends to $F\langle DG,h\rangle_{[0,T]}$ in $\mathcal{L}^1(\mathcal{F}_T)$, and therefore $E(F_n\langle DG,h\rangle_{[0,T]})$ tends to $E(F\langle DG,h\rangle_{[0,T]})$.

Conclusion. We have now argued that

$$\begin{aligned} E(FG\theta(h)) &= \lim_n E(F_n G\theta(h)) \\ E(F\langle DG,h\rangle_{[0,T]}) &= \lim_n E(F_n\langle DG,h\rangle_{[0,T]}) \\ E(G\langle DF,h\rangle_{[0,T]}) &= \lim_n E(G\langle DF_n,h\rangle_{[0,T]}), \end{aligned}$$

and we may therefore conclude $E(G\langle DF,h\rangle_{[0,T]}) = E\left(FG\theta(h)\right) - EF\langle DG,h\rangle_{[0,T]}$ by taking the limits. This is the result we set out to prove. □

We have now come to the conclusion of this section. Before proceeding, we will review our progress. We have developed some of the fundamental results of the Malliavin derivative on $\mathbb{D}_{1,2}$. In Lemma 4.3.7, we have checked that stochastic integrals with deterministic integrands are in $\mathbb{D}_{1,2}$. This allowed us to reconcile our Malliavin derivative with that of Nualart (2006). Also, in Theorem 4.3.5 and Theorem 4.3.8, we proved the chain rule and the integration-by-parts formula, respectively. These two results will be indispensable in the coming sections. The chain rule tells us that as long as natural integrability conditions are satisfied, $\mathbb{D}_{1,2}$ is stable under C^1 transformations. The integration-by-parts formula is essential because it can transfer an expectation involving DF into an expectation involving only F. This will be of particular use in the form of Corollary 4.3.11.

Our next goal is to introduce a series of subspaces of $\mathcal{L}^2(\mathcal{F}_T)$ and $\mathcal{L}^2(\Pi)$ and investigate the special properties of D related to these subspaces. These results are very important for the general theory of the Malliavin calculus. Unfortunately we will not be able to see the full impact of them with the limited theory we detail here. However, we will take the time to obtain one corollary from the theory, namely an extension of the chain rule to Lipschitz transformations.

4.4 Orthogonal subspaces of $\mathcal{L}^2(\mathcal{F}_T)$ and $\mathcal{L}^2(\Pi)$

In this section, we will introduce orthogonal decompositions of $\mathcal{L}^2(\mathcal{F}_T)$ and $\mathcal{L}^2(\Pi)$ given by

$$\mathcal{L}^2(\mathcal{F}_T) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n \quad \text{and} \quad \mathcal{L}^2(\Pi) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n(\Pi).$$

We will see that subspaces $\mathcal{H}_n$ have orthogonal bases for which the Malliavin derivative can be explicitly computed, and the images of the subspaces are in $\mathcal{H}_n(\Pi)$. These properties will allow us a characterization of $\mathbb{D}_{1,2}$ in terms of the projections on the subspaces $\mathcal{H}_n$ and $\mathcal{H}_n(\Pi)$, which can lead to some very useful results. In order to give an idea of what is to come, we provide the following graphical overview.

$$\begin{tikzcd}
\mathbb{H}_n \arrow[r, "\text{span}"] & \mathcal{H}''_n \arrow[d, "\text{cl}"] & \mathcal{H}''_n(\Pi) \arrow[d, "\text{cl}"] & \mathbb{H}_n(\Pi) \arrow[l, "\text{span}"'] \\
\mathcal{H}'_n \arrow[r, "\text{cl}"] & \mathcal{H}_n \arrow[r, "D"] & \mathcal{H}_n(\Pi) & \mathcal{H}'_n(\Pi) \arrow[l, "\text{cl}"]
\end{tikzcd}$$

Consider the left-hand part of graph. It is meant to state that $\mathcal{H}_n$ is the closure of $\mathcal{H}'_n$ and $\mathcal{H}''_n$, and the span of $\mathbb{H}_n$ is $\mathcal{H}''_n$. On the right-hand part, $\mathcal{H}_n(\Pi)$ is the closure of $\mathcal{H}'_n(\Pi)$ and $\mathcal{H}''_n(\Pi)$, and the span of $\mathbb{H}_n(\Pi)$ is $\mathcal{H}''_n$. The connection between the left-hand and right-hand parts is the Malliavin derivative, which maps $\mathcal{H}_n$ into $\mathcal{H}_n(\Pi)$.

Our plan is first to formally introduce $\mathcal{H}_n$, $\mathcal{H}'_n$, $\mathcal{H}''_n$ and $\mathbb{H}_n$ and consider their basic properties. After this, we introduce $\mathcal{H}_n(\Pi)$, $\mathcal{H}'_n(\Pi)$, $\mathcal{H}''_n(\Pi)$ and $\mathbb{H}_n(\Pi)$. Finally, we prove that D maps $\mathcal{H}_n$ into $\mathcal{H}_n(\Pi)$ by explicitly calculating the Malliavin derivative of the elements in $\mathbb{H}_n$. We then apply our results to obtain a characterization of the elements of the space $\mathbb{D}_{1,2}$ in terms of the projections on the spaces $\mathcal{H}_n$.

The subspaces $\mathcal{H}_n$ are based on the Hermite polynomials. As described in Appendix A.5, the n'th Hermite polynomial is defined by

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

with $H_0(x) = 1$ and $H_{-1}(x) = 0$. See the appendix for elementary properties of the Hermite polynomials. This section will use a good deal of Hilbert space theory, see Appendix A.6 for a review of the results used. We now define some of the subspaces of $\mathcal{L}^2(\mathcal{F}_T)$ under consideration. Recall that we in the previous section defined θ as the Brownian stochastic integral operator on $[0, T]$.

Definition 4.4.1. *We let $\mathcal{H}'_n$ be the linear span of $H_n(\theta h)$ for $h \in \mathcal{L}^2[0,T]$ with $\|h\| = 1$ and let $\mathcal{H}_n$ be the closure of $\mathcal{H}'_n$ in $\mathcal{L}^2(\mathcal{F}_T)$.*

Our first goal is to show that the subspaces $(\mathcal{H}_n)$ are mutually orthogonal and that their orthogonal sum is $\mathcal{L}^2(\mathcal{F}_T)$. The following lemma will be fundamental to this endeavor.

Lemma 4.4.2. *It holds that for $h, g \in \mathcal{L}^2[0,T]$ with $\|h\|_2 = \|g\|_2 = 1$,*

$$EH_n(\theta h)H_m(\theta g) = \begin{cases} 0 & \text{if } \ n \neq m \\ n!\langle h, g\rangle^n & \text{if } \ n = m \end{cases}.$$

Proof. Let $n, m \geq 1$ be given. By Lemma A.5.2, we have

$$\begin{aligned} EH_n(\theta h)H_m(\theta g) &= E\frac{\partial^n}{\partial s^n}\exp\left(s\theta h - \frac{s^2}{2}\right)\Big|_{s=0}\frac{\partial^m}{\partial t^m}\exp\left(t\theta g - \frac{t^2}{2}\right)\Big|_{t=0} \\ &= E\left(\frac{\partial^n}{\partial s^n}\exp\left(s\theta h - \frac{s^2}{2}\right)\frac{\partial^m}{\partial t^m}\exp\left(t\theta g - \frac{t^2}{2}\right)\right)\Big|_{s=0,t=0} \\ &= E\left(\frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}\exp\left(s\theta h - \frac{s^2}{2}\right)\exp\left(t\theta g - \frac{t^2}{2}\right)\right)\Big|_{s=0,t=0}. \end{aligned}$$

Our plan is to interchange the differentiation and integration above and explicitly calculate and differentiate the resulting mean.

Step 1: Interchange of differentiation and integration. We first note that by a simple induction proof,

$$\frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}\exp\left(sx - \frac{s^2}{2}\right)\exp\left(ty - \frac{t^2}{2}\right) = q_n(s,x)q_m(t,y)\exp\left(sx - \frac{s^2}{2}\right)\exp\left(ty - \frac{t^2}{2}\right),$$

where q_n is some polynomial in two variables. We will show that for any $s, t \in \mathbb{R}$, it holds that

$$\begin{aligned} &E\frac{\partial}{\partial s}q_n(s,x)q_m(t,y)\exp\left(s\theta h - \frac{s^2}{2}\right)\exp\left(t\theta g - \frac{t^2}{2}\right) \\ = \; &\frac{\partial}{\partial s}Eq_n(s,x)q_m(t,y)\exp\left(s\theta h - \frac{s^2}{2}\right)\exp\left(t\theta g - \frac{t^2}{2}\right). \end{aligned}$$

By symmetry, if this is true, then the same result will of course also hold when differentiation with respect to t. Applying this result $n + m$ times will yield the desired exchange of differentiation and integration. To show the result, fix $t \in \mathbb{R}$ and consider a bounded interval $[a,b]$. Define $c = \min\{|a|,|b|\}$. We then have, with r_{n+1} some polynomial and C and C'_t some suitably large constants,

$$\begin{aligned}
& \sup_{a\le s\le b}\left|\frac{\partial}{\partial s}q_n(s,x)q_m(t,y)\exp\left(sx-\frac{s^2}{2}\right)\exp\left(ty-\frac{t^2}{2}\right)\right| \\
= & \sup_{a\le s\le b}\left|q_{n+1}(s,x)q_m(t,y)\exp\left(sx-\frac{s^2}{2}\right)\exp\left(ty-\frac{t^2}{2}\right)\right| \\
\le & \ r_{n+1}(|x|)q_m(|t|,|y|)\exp\left(bx-\frac{c^2}{2}\right)\exp\left(ty-\frac{t^2}{2}\right) \\
\le & \ (C+\exp(|x|)(C'_t+\exp(|y|))\exp\left(bx-\frac{c^2}{2}\right)\exp\left(ty-\frac{t^2}{2}\right).
\end{aligned}$$

We have exploited that exponentials grow faster than any polynomial. Now, since $E\exp(\sum_{k=1}^n \lambda_k|X_k|)$ is finite whenever $\lambda\in\mathbb{R}^n$ and X has an n-dimensional normal distribution, we find

$$E(C+\exp(|\theta h|)(C'_t+\exp(|\theta g|))\exp\left(b\theta h-\frac{c^2}{2}\right)\exp\left(t\theta g-\frac{t^2}{2}\right)<\infty.$$

Ordinary results for exchanging differentiation and integration now certify that

$$\begin{aligned}
& E\frac{\partial}{\partial s}q_n(s,x)q_m(t,y)\exp\left(s\theta h-\frac{s^2}{2}\right)\exp\left(t\theta g-\frac{t^2}{2}\right) \\
= & \frac{\partial}{\partial s}Eq_n(s,x)q_m(t,y)\exp\left(s\theta h-\frac{s^2}{2}\right)\exp\left(t\theta g-\frac{t^2}{2}\right),
\end{aligned}$$

as desired. We can then conclude

$$\begin{aligned}
& E\left(\frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}\exp\left(s\theta h-\frac{s^2}{2}\right)\exp\left(t\theta g-\frac{t^2}{2}\right)\right) \\
= & \frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}E\left(\exp\left(s\theta h-\frac{s^2}{2}\right)\exp\left(t\theta g-\frac{t^2}{2}\right)\right),
\end{aligned}$$

and finally,

$$\begin{aligned}
& E\left(\frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}\exp\left(s\theta h-\frac{s^2}{2}\right)\exp\left(t\theta g-\frac{t^2}{2}\right)\right)\Bigg|_{s=0,t=0} \\
= & \frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}E\exp\left(s\theta h-\frac{s^2}{2}\right)\exp\left(t\theta g-\frac{t^2}{2}\right)\Bigg|_{s=0,t=0}.
\end{aligned}$$

Step 2: Calculating the mean. We now set to work to calculate

$$E\exp\left(s\theta h-\frac{s^2}{2}\right)\exp\left(t\theta h-\frac{t^2}{2}\right).$$

By Lemma 3.7.4, $(\theta h, \theta g)$ is normally distributed with mean zero, $V\theta h = 1$, $V\theta g = 1$ and $\mathrm{Cov}(\theta h, \theta g) = \langle h, g\rangle$. We then conclude for any $s, t \in \mathbb{R}$, $s\theta h + t\theta g$ is normally distributed with mean zero and variance $s^2 + t^2 + 2st\langle h, g\rangle$. Using the formula for the Laplace transform of the normal distribution, we then find

$$\begin{aligned}
& E\exp\left(s\theta h - \frac{s^2}{2}\right)\exp\left(t\theta g - \frac{t^2}{2}\right) \\
= \;& \exp\left(-\frac{s^2+t^2}{2}\right)E\exp\left(s\theta h + t\theta g\right) \\
= \;& \exp\left(-\frac{s^2+t^2}{2}\right)\exp\left(\frac{1}{2}(s^2+t^2+2st\langle h, g\rangle)\right) \\
= \;& \exp\left(st\langle h, g\rangle\right).
\end{aligned}$$

Step 3: Conclusion. With $\alpha = \max\{n, m\}$, we now obtain

$$\begin{aligned}
& \frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}E\exp\left(s\theta h - \frac{s^2}{2}\right)\exp\left(t\theta g - \frac{t^2}{2}\right) \\
= \;& \frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}\exp\left(st\langle h, g\rangle\right) \\
= \;& \frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}\sum_{k=0}^{\infty}\frac{\langle h, g\rangle^k}{k!}s^k t^k \\
= \;& \sum_{k=\alpha}^{\infty}\frac{\langle h, g\rangle^k}{k!}n!m!s^{k-n}t^{k-m}.
\end{aligned}$$

Combining our results, we conclude

$$\begin{aligned}
EH_n(\theta h)H_m(\theta g) &= E\left(\frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}\exp\left(s\theta h - \frac{s^2}{2}\right)\exp\left(t\theta g - \frac{t^2}{2}\right)\right)\Bigg|_{s=t=0} \\
&= \frac{\partial^n}{\partial s^n}\frac{\partial^m}{\partial t^m}E\exp\left(s\theta h - \frac{s^2}{2}\right)\exp\left(t\theta g - \frac{t^2}{2}\right)\Bigg|_{s=0,t=0} \\
&= \sum_{k=\alpha}^{\infty}\frac{\langle h, g\rangle^k}{k!}n!m!s^{k-n}t^{k-m}\Bigg|_{s=t=0},
\end{aligned}$$

and the final expression above is zero if $n \neq m$ and $n!\langle h, g\rangle^n$ otherwise. This shows the claim of the lemma. □

Theorem 4.4.3. *The subspaces $(\mathcal{H}_n)$ are orthogonal, and $\mathcal{L}^2(\mathcal{F}_T) = \oplus_{n=0}^{\infty}\mathcal{H}_n$.*

Proof. We first prove that the subspaces are mutually orthogonal. Let $n, m \geq 1$ be given with $n \neq m$ and let $h, g \in \mathcal{L}^2[0, T]$ with $\|h\|_2 = \|g\|_2 = 1$. By Lemma 4.4.2,

$H_n(\theta h)$ and $H_m(\theta g)$ are orthogonal. Since $\mathcal{H}_n$ is the closure of the span of elements of the form $H_n(\theta h)$, $\|h\| = 1$, Lemma A.6.2 yields that $\mathcal{H}_n \perp \mathcal{H}_m$.

Next, we must show $\mathcal{L}^2(\mathcal{F}_T) = \oplus_{n=0}^\infty \mathcal{H}_n$. By Lemma A.6.11, $\oplus_{n=0}^\infty \mathcal{H}_n$ is the closure of span $\cup_{n=0}^\infty \mathcal{H}_n$. It will therefore suffice to show that the subspace span $\cup_{n=0}^\infty \mathcal{H}_n$ is dense in $\mathcal{L}^2(\mathcal{F}_T)$. To do so, it will by Lemma A.6.4 suffice to show that the orthogonal complement of span $\cup_{n=0}^\infty \mathcal{H}_n$ is $\{0\}$. And to do so, it will by Lemma A.6.2 suffice to show that the orthogonal complement of $\cup_{n=0}^\infty \mathcal{H}_n$ is $\{0\}$. This is therefore what we set out to do.

Therefore, let $F \in \mathcal{L}^2(\mathcal{F}_T)$ be given, orthogonal to $\cup_{n=1}^\infty \mathcal{H}_n$. This means that, for any $n \geq 0$ and h with $\|h\| = 1$, $EFH_n(\theta h) = 0$. Let n be given. By Lemma A.5.6, there are coefficients such that

$$x^n = \sum_{k=0}^n \lambda_k H_k(x).$$

We then obtain $EF(\theta h)^n = \sum_{k=0}^n \lambda_k EFH_k(\theta h) = 0$. By scaling, we can conclude that $EF(\theta h)^n$ for any $h \in \mathcal{L}^2[0,T]$. Therefore, in the following h will denote an arbitrary element of $\mathcal{L}^2[0,T]$, not necessarily of unit norm. We will show $EF \exp(\theta h) = 0$. To do so, note that by Lemma B.3.2,

$$\begin{aligned}
\sum_{n=0}^\infty \left\| \frac{(\theta h)^n}{n!} \right\|_2 &= \sum_{n=0}^\infty \frac{1}{n!} \left(E(\theta h)^{2n} \right)^{\frac{1}{2}} \\
&= \sum_{n=0}^\infty \frac{1}{n!} \|h\|^n \left(\frac{(2n)!}{2^n n!} \right)^{\frac{1}{2}} \\
&= \sum_{n=0}^\infty \frac{\|h\|^n}{\sqrt{n!}} \left(\frac{(2n)!}{2^n (n!)^2} \right)^{\frac{1}{2}} \\
&\leq \sum_{n=0}^\infty \frac{\|h\|^n}{\sqrt{n!}},
\end{aligned}$$

where we have used that $\frac{(2n)!}{2^n n!} \leq n!$. $\sum_{n=0}^\infty \frac{\|h\|^n}{\sqrt{n!}}$ is convergent by the quotient test, and by completeness, we can conclude that $\sum_{n=0}^\infty \frac{(\theta h)^n}{n!}$ is convergent in $\mathcal{L}^2(\mathcal{F}_T)$. Since it also converges almost surely to $\exp(\theta h)$, this is also the limit in $\mathcal{L}^2(\mathcal{F}_T)$. By the Cauchy-Schwartz inequality, we find that $F \sum_{k=0}^n \frac{(\theta h)^k}{k!}$ converges in $\mathcal{L}^1(\mathcal{F}_T)$ to $F \exp(\theta h)$, and therefore

$$EF \exp(\theta h) = \lim_n EF \sum_{k=0}^n \frac{(\theta h)^k}{k!} = 0.$$

Since the variables of the form $\exp(\theta h)$ where $h \in \mathcal{L}^2[0,T]$ form a dense subset of

$\mathcal{L}^2(\mathcal{F}_T)$ by Theorem B.6.2, their span form a dense subspace of $\mathcal{L}^2(\mathcal{F}_T)$. Since F is orthogonal to this subspace, Lemma A.6.4 yields $F = 0$, as desired. □

Next, we introduce the orthonormal basis $\mathbb{H}_n$ for the subspace $\mathcal{H}_n$. We are going to need an orthonormal basis for $\mathcal{L}^2[0,T]$. Since $\mathcal{B}[0,T]$ is countably generated, Lemma B.4.1 yields that $\mathcal{L}^2[0,T]$ is separable, and therefore any orthonormal basis must be countable. Let (e_n) be such a basis, we will hold this basis fixed for the remainder of the section.

We let $\mathbb{I}$ be the set of sequences with values in $\mathbb{N}_0$ which are zero from a point onwards. We call the elements of $\mathbb{I}$ multi-indices. If $a \in \mathbb{I}$, we put $|a| = \sum_{n=1}^\infty a_n$ and call $|a|$ the degree of the multi-index. We also define $a! = \prod_{n=1}^\infty a_n!$. We let $\mathbb{I}_n$ be the multi-indices of degree n and let $\mathbb{I}_n^m$ be the multi-indices a of degree n such that $a_k = 0$ whenever $k > m$. We say that the multi-indices in $\mathbb{I}_n^m$ have order m.

Definition 4.4.4. *For any multi-index a of order n, we define $\Phi_a = \frac{1}{\sqrt{a!}}\prod_{n=1}^\infty H_{a_n}(\theta e_n)$. We write $\mathbb{H} = \{\Phi_a | a \in \mathbb{I}\}$ and $\mathbb{H}_n = \{\Phi_a || a| = n\}$. We let $\mathcal{H}_n''$ denote the span of $\mathbb{H}_n$.*

We begin by showing that $\mathbb{H}$ is an orthonormal set. Afterwards, we will work on showing that $\mathbb{H}_n$ is an orthonormal basis for $\mathcal{H}_n$.

Lemma 4.4.5. *$\mathbb{H}$ is an orthonormal set in $\mathcal{L}^2(\mathcal{F}_T)$.*

Proof. We need to show that each Φ_a has unit norm and that the elements of the family is mutually orthogonal.

Unit norm. Let a multi-index a be given. Since (e_n) is orthonormal, the family (θe_n) are mutually independent. Recalling that a_k is zero from a point onwards, using Lemma 3.7.4 and Lemma 4.4.2, we obtain

$$\begin{aligned}
\|\Phi_a\|_2^2 &= E\left(\frac{1}{\sqrt{a!}}\prod_{n=1}^\infty H_{a_n}(\theta e_n)\right)^2 \\
&= \frac{1}{a!}\prod_{n=1}^\infty EH_{a_n}(\theta e_n)^2 \\
&= \frac{1}{a!}\prod_{n=1}^\infty a_n!\|e_n\|^2 \\
&= 1.
\end{aligned}$$

Orthogonality. Next, consider two multi-indices a and b with $a \neq b$. We then find

$$\begin{aligned}
\langle \Phi_a, \Phi_b \rangle &= E\left(\frac{1}{\sqrt{a!}} \prod_{n=1}^{\infty} H_{a_n}(\theta e_n)\right)\left(\frac{1}{\sqrt{b!}} \prod_{n=1}^{\infty} H_{b_n}(\theta e_n)\right) \\
&= \frac{1}{\sqrt{a!}\sqrt{b!}} E \prod_{n=1}^{\infty} H_{a_n}(\theta e_n) H_{b_n}(\theta e_n) \\
&= \frac{1}{\sqrt{a!}\sqrt{b!}} \prod_{n=1}^{\infty} E H_{a_n}(\theta e_n) H_{b_n}(\theta e_n).
\end{aligned}$$

Now, since there is some n such that $a_n \neq b_n$, it follows from Lemma 4.4.2 that for this n, $EH_{a_n}(\theta e_n)H_{b_n}(\theta e_n) = 0$. Therefore, the above is zero and Φ_a and Φ_b are orthogonal. □

In order to show that $\mathbb{H}_n$ is an orthonormal basis for $\mathcal{H}_n$, we will need to add a few more spaces to our growing collection of subspaces of $\mathcal{L}^2(\mathcal{F}_T)$. Recall that $\mathfrak{P}_n$ denotes the polynomials of degree less than or equal to n. That is, $\mathfrak{P}_n$ is the family of polynomials of degree less than or equal to n in k variables for any $k \geq 1$.

Definition 4.4.6. *We let $\mathcal{P}'_n$ be the linear span of $p(\theta h)$, where $p \in \mathfrak{P}_n$ is a polynomial in k variables and $h \in (\mathcal{L}^2[0,T])^k$, and let $\mathcal{P}_n$ be the closure of $\mathcal{P}'_n$.*

Lemma 4.4.7. *Consider $h \in (\mathcal{L}^2[0,T])^n$ with $h = (h_1, \ldots, h_n)$. Let (h_k) be a sequence with $h_k = (h_1^k, \ldots, h_n^k)$ converging to h. Let $f \in C_p^1(\mathbb{R}^n)$. Then $f(\theta h_k)$ converges to $f(\theta h)$ in $\mathcal{L}^2(\mathcal{F}_T)$.*

Proof. This follows from Lemma B.3.4. □

Lemma 4.4.8. *Let $\mathbb{P}_n$ be the family of variables $p(\theta e)$, where $p \in \mathfrak{P}_n$ is a polynomium of k variables and e is a vector of distinct elements from the orthonormal basis (e_n). Then the closure of the span of $\mathbb{P}_n$ is equal to $\mathcal{P}_n$.*

Proof. Obviously, $\mathbb{P}_n \subseteq \mathcal{P}'_n$, and therefore $\overline{\text{span}}\, \mathbb{P}_n \subseteq \mathcal{P}_n$. We need to prove the other inclusion. We first consider the case $F = p(\theta h)$, where $p \in \mathfrak{P}_n$ is a polynomial of degree n in k variables and $h = (h_1, \ldots, h_k)$ with coordinates in span $\{e_i\}_{i \geq 1}$. We want to prove that $F \in \overline{\text{span}}\, \mathbb{P}_n$. To do so, first observe that there exists an orthonormal subset $\{g_1, \ldots, g_m\}$ of distinct elements from $\{e_i\}_{i \geq 1}$ such that $h_j = \sum_{i=1}^m \lambda_i^j g_i$ for some coefficients λ_i^j. We then obtain

$$p(\theta h) = p\left(\sum_{i=1}^m \lambda_i^1 g_i, \ldots, \sum_{i=1}^m \lambda_i^k g_i\right) = q(g_1, \ldots, g_m),$$

where $q(x_1,\ldots,x_m) = p(\sum_{i=1}^m \lambda_i^1 x_i,\ldots,\sum_{i=1}^m \lambda_i^k x_i)$ is a polynomium of degree less than or equal to n in m variables. We conclude $p(\theta h) \in \mathbb{P}_n$. We have now proven that $p(\theta h)$ is in $\overline{\text{span}}\, \mathbb{P}_n$ when the coordinates of h are in span $\{e_i\}_{i\geq 1}$.

Next, consider $F = p(\theta h)$, where $h \in (\mathcal{L}^2[0,T])^k$, $h = (h_1,\ldots,h_k)$. We need to prove that $p(\theta h)$ is in $\overline{\text{span}}\, \mathbb{P}_n$. To this end, note that since (e_n) is an orthonormal basis, there exists sequences (h_j^i) in span $\{e_i\}_{i\geq 1}$ for $j \leq k$ such that h_j^i converges to h_j. Putting $h^i = (h_1^i,\ldots,h_k^i)$, we then find by Lemma 4.4.7 that $p(\theta h^i)$ converges in $\mathcal{L}^2$ to $p(\theta h)$. Since we have already proven that $p(\theta h^i) \in \overline{\text{span}}\, \mathbb{P}_n$, this shows that $p(\theta h) \in \overline{\text{span}}\, \mathbb{P}_n$. By linearity, we may now conclude $\mathcal{P}_n' \subseteq \overline{\text{span}}\, \mathbb{P}_n$ and therefore also $\mathcal{P}_n \subseteq \overline{\text{span}}\, \mathbb{P}_n$, as desired. □

Lemma 4.4.9. *For each $n \geq 0$, $\mathcal{P}_n = \oplus_{k=0}^n \mathcal{H}_k$.*

Proof. We first show the easier inclusion $\oplus_{k=0}^n \mathcal{H}_k \subseteq \mathcal{P}_n$. Let $k \leq n$. By Lemma A.5.6, the k'th Hermite polynomial is a polynomial of degree k, therefore an element of $\mathfrak{P}_n$. Thus, $\mathcal{H}_k' \subseteq \mathcal{P}_n' \subseteq \mathcal{P}_n$. Since $\mathcal{P}_n$ is closed, this implies $\mathcal{H}_k \subseteq \mathcal{P}_n$, and we therefore immediately obtain $\oplus_{k=0}^n \mathcal{H}_k \subseteq \mathcal{P}_n$, as desired.

Now consider the other inclusion, $\mathcal{P}_n \subseteq \oplus_{k=0}^n \mathcal{H}_k$. From Theorem 4.4.3 and Lemma A.6.16, we have the orthogonal decomposition

$$\mathcal{L}^2(\mathcal{F}_T) = \bigoplus_{k=0}^{\infty} \mathcal{H}_k = \left(\bigoplus_{k=0}^{n} \mathcal{H}_k\right) \oplus \left(\bigoplus_{k=n+1}^{\infty} \mathcal{H}_k\right).$$

By Lemma A.6.14, we therefore find that to show $\mathcal{P}_n \subseteq \oplus_{k=0}^\infty \mathcal{H}_k$, it will suffice to show that $\mathcal{P}_n$ is orthogonal to $\oplus_{k=n+1}^\infty \mathcal{H}_k$. And to do so, it will by Lemma A.6.11 and Lemma A.6.2 suffice to show that $\mathcal{P}_n$ is orthogonal to $\mathcal{H}_k'$ for any $k > n$.

Therefore, let $k > n$ be given, and let $\|h\| = 1$. We will show that $\mathcal{P}_n$ is orthogonal to $H_k(\theta h)$. To this end, it will suffice to show that $\mathcal{P}_n'$ is orthogonal to $H_k(\theta h)$. Let $F \in \mathcal{P}_n'$ with $F = p(\theta g)$, where $g = (g_1,\ldots,g_m)$ are in $\mathcal{L}^2[0,T]$. Using the same technique as in the proof of Lemma 4.4.8, we can assume that $g_1,\ldots,g_m$ are orthonormal. In fact, by the Gram-Schmidt orthonormalization procedure, we can even assume $g_1 = h$.

We now prove $EFH_k(\theta h) = 0$. By Lemma A.5.6, we have, for any $i \geq 0$,

$$x^i = \sum_{j=0}^{i} \mu_j^i H_j(x)$$

for some coefficients $\mu_j^i \in \mathbb{R}$. Recalling that $\mathbb{I}_n^m$ denotes the multi-indicies of order m with degree n, we then have, for some coefficients λ_a,

$$\begin{aligned} p(\theta g) &= \sum_{a \in \mathbb{I}_n^m} \lambda_a \prod_{i=1}^m (\theta g_i)^{a_i} \\ &= \sum_{a \in \mathbb{I}_n^m} \lambda_a \prod_{i=1}^m \sum_{j=0}^{a_i} \mu_j^{a_i} H_j(\theta g_i) \\ &= \sum_{a \in \mathbb{I}_n^m} \lambda_a \sum_{j_1=0}^{a_1} \cdots \sum_{j_m=0}^{a_m} \prod_{i=1}^m \mu_{j_i}^{a_i} H_{j_i}(\theta g_i). \end{aligned}$$

Therefore, we obtain

$$EFH_k(\theta h) = \sum_{a \in \mathbb{I}_n^m} \lambda_a \sum_{j_1=0}^{a_1} \cdots \sum_{j_m=0}^{a_m} EH_k(\theta h) \prod_{i=1}^m \mu_{j_i}^{a_i} H_{j_i}(\theta g_i).$$

Now consider the expression $EH_k(\theta h) \prod_{i=1}^m \mu_{j_i}^{a_i} H_{j_i}(\theta g_i)$, we wish to argue that this is equal to zero. Since $g_1, \ldots, g_m$ are orthonormal, $\theta g_1, \ldots, \theta g_m$ are independent. We therefore find, recalling that $g_1 = h$,

$$EH_k(\theta h) \prod_{i=1}^m \mu_{j_i}^{a_i} H_{j_i}(\theta g_i) = E(H_k(\theta h) H_{j_1}(\theta h)) \prod_{i=2}^m E\mu_{j_i}^{a_i} H_{j_i}(\theta g_i).$$

Since $|a| \le n$, $a_1 \le n$ and in particular $j_1 \le n$. Since $k > n$, we may conclude by Lemma 4.4.2 that $E(H_k(\theta h) H_{j_1}(\theta h)) = 0$, and the above expression is therefore zero. Thus, $EFH_k(\theta h) = 0$ and we conclude that $\mathcal{P}_n$ is orthogonal to $\mathcal{H}'_k$ for $k > n$. As a consequence of this, $\mathcal{P}_n$ is orthogonal to $\oplus_{k=n+1}^\infty \mathcal{H}_k$ and therefore $\mathcal{P}_n \subseteq \oplus_{k=0}^n \mathcal{H}_k$, as desired. □

Theorem 4.4.10. *For each n, $\mathbb{H}_n$ is an orthonormal basis of $\mathcal{H}_n$.*

Proof. From Lemma 4.4.5, we already know that $\mathbb{H}_n$ is an orthonormal set, so it will suffice to show that the span $\mathcal{H}''_n$ is dense in $\mathcal{H}_n$. Let $\mathcal{K}_n$ be the closure of $\mathcal{H}''_n$. We wish to show $\mathcal{K}_n = \mathcal{H}_n$. To do so, we first prove $\mathcal{P}_n = \oplus_{k=0}^n \mathcal{K}_k$. Since Φ_a for $|a| = k$ is a polynomial transformation of degree k of variables $\theta(e_i)$, it is clear that $\mathbb{H}_k \subseteq \mathcal{P}_n$ for any $k \le n$ and therefore $\mathcal{K}_k \subseteq \mathcal{P}_n$. In particular, $\oplus_{k=0}^n \mathcal{K}_k \subseteq \mathcal{P}_n$. We need to prove the other inclusion.

By Lemma 4.4.8, it will suffice to show $\mathbb{P}_n \subseteq \oplus_{k=0}^n \mathcal{K}_k$. Therefore, let $F \in \mathbb{P}_n$ with $F = p(\theta e)$, where $e = (e_{k_1}, \ldots, e_{k_n})$ are distinct and $p(x) = \sum_{a \in \mathbb{I}_n^m} \lambda_k x^a$ is a polynomial of degree n in m variables. By Lemma A.5.6, we have, for any $n \geq 0$, $x^n = \sum_{k=0}^n \mu_k^n H_k(x)$, for some coefficients $\mu_k^n \in \mathbb{R}$. We obtain

$$\begin{aligned} p(\theta(e_{k_1}), \ldots, \theta(e_{k_n})) &= \sum_{a \in I_n^m} \lambda_a \prod_{i=1}^n (\theta e_{k_i})^{a_i} \\ &= \sum_{a \in \mathbb{I}_n^m} \lambda_a \prod_{i=1}^n \sum_{j=0}^{a_i} \mu_j^{a_i} H_j(\theta e_{k_i}) \\ &= \sum_{a \in \mathbb{I}_n^m} \lambda_a \sum_{j_1=0}^{a_1} \cdots \sum_{j_n=0}^{a_n} \prod_{i=1}^n \mu_{j_i}^{a_i} H_{j_i}(\theta e_{k_i}) \\ &= \sum_{a \in \mathbb{I}_n^m} \sum_{j_1=0}^{a_1} \cdots \sum_{j_n=0}^{a_n} \left(\lambda_a \prod_{i=1}^n \sqrt{j_i!} \mu_{j_i}^{a_i} \right) \prod_{i=1}^n \frac{1}{\sqrt{j_i!}} H_{j_i}(\theta e_{k_i}). \end{aligned}$$

Now, with $|j| = \sum_{k=1}^n j_k$, the innermost product of the above is in $\mathbb{H}_{|j|}$. Since $|a| \leq n$ and $j_i \leq a_i$ for $i \leq n$, $|j| \leq n$. We may therefore conclude that the above is in the span of $\cup_{k=0}^n \mathbb{H}_k$, therefore in $\oplus_{k=0}^n \mathcal{K}_k$.

We have now shown $\mathcal{P}_n = \oplus_{k=0}^n \mathcal{K}_k$. Recall that from Lemma 4.4.9, we also have $\mathcal{P}_n = \oplus_{k=0}^n \mathcal{H}_k$. We may then conclude

$$\mathcal{P}_{n-1} \oplus \mathcal{K}_n = \bigoplus_{k=0}^n \mathcal{K}_k = \bigoplus_{k=0}^n \mathcal{H}_k = \mathcal{P}_{n-1} \oplus \mathcal{H}_n.$$

Now, since $\mathcal{P}_{n-1} = \oplus_{k=0}^{n-1} \mathcal{K}_k$ and $\mathcal{P}_{n-1} = \oplus_{k=0}^{n-1} \mathcal{H}_k$, we conclude that $\mathcal{K}_n$ and $\mathcal{P}_{n-1}$ are orthogonal and $\mathcal{P}_{n-1}$ and $\mathcal{H}_n$ are orthogonal. Therefore, by Lemma A.6.15, $\mathcal{K}_n = \mathcal{H}_n$ and therefore $\overline{\text{span}}\, \mathbb{H}_n = \text{cl}\, \mathcal{H}_n'' = \mathcal{K}_n = \mathcal{H}_n$, as desired. □

Our results so far are the following: The subspaces $\mathcal{H}_n$ are mutually orthogonal, has orthonormal bases $\mathbb{H}_n$, and $\mathcal{L}^2(\mathcal{F}_T) = \oplus_{n=0}^\infty \mathcal{H}_n$. We also have two subspaces $\mathcal{H}_n'$ and $\mathcal{H}_n''$ which are dense in $\mathcal{H}_n$.

We will now use the families $\mathbb{H}_n$, $\mathcal{H}_n'$, $\mathcal{H}_n''$ and $\mathcal{H}_n$ to define analogous families $\mathbb{H}_n(\Pi)$, $\mathcal{H}_n'(\Pi)$, $\mathcal{H}_n''(\Pi)$ and $\mathcal{H}_n(\Pi)$ in $\mathcal{L}^2(\Pi)$. These sets will have properties much alike to their counterparts in $\mathcal{L}^2(\mathcal{F}_T)$, and will eventually be linked to these through the Malliavin derivative.

Definition 4.4.11. *By $\mathcal{H}_n'(\Pi)$, we denote the span of the variables of the form $F \otimes h$, where $F \in \mathcal{H}_n$ and $h \in \mathcal{L}^2[0, T]$. By $\mathcal{H}_n(\Pi)$, we denote the closure of $\mathcal{H}_n'(\Pi)$.*

Theorem 4.4.12. *Defining $\mathbb{H}_n(\Pi)$ as the family of elements $\Phi_a \otimes e_i$, where $|a| = n$ and $i \in \mathbb{N}$, $\mathbb{H}_n(\Pi)$ is an orthonormal basis for $\mathcal{H}_n(\Pi)$. In particular, the span $\mathcal{H}''_n(\Pi)$ of $\mathbb{H}_n(\Pi)$ is dense in $\mathcal{H}_n(\Pi)$.*

Proof. It is clear that $\|\Phi_a \otimes e_i\|_2 = \|\Phi_a\|_2 \|e_i\|_2 = 1$. For any $a, b \in \mathbb{I}_n$ and $i, j \in \mathbb{N}$, the Fubini Theorem yields $\langle \Phi_a \otimes e_i, \Phi_b \otimes e_j \rangle = \langle \Phi_a, \Phi_b \rangle \langle e_i, e_j \rangle$, so if $a \neq b$ or $i \neq j$, $\Phi_a \otimes e_i$ and $\Phi_b \otimes e_j$ are orthogonal. Thus, $\mathbb{H}_n(\Pi)$ is orthonormal.

It remains to show $\overline{\text{span}}\ \mathbb{H}_n(\Pi) = \mathcal{H}_n(\Pi)$. To this end, it will suffice to show that $\mathcal{H}'_n(\Pi) \subseteq \overline{\text{span}}\ \mathbb{H}_n(\Pi)$. Therefore, consider $F \in \mathcal{H}_n$ and $h \in \mathcal{L}^2[0, T]$. We need to demonstrate $F \otimes h \in \overline{\text{span}}\ \mathbb{H}_n(\Pi)$. Now, there are $h_n \in \text{span}\ \{e_i\}_{i \geq 1}$ such that h_n tends to h, and therefore

$$\lim_n \|F \otimes h - F \otimes h_n\|_2 = \lim_n \|F\|_2 \|h - h_n\|_2 = 0.$$

Furthermore, there are $F_n \in \text{span}\ \mathbb{H}_n$ such that F_n tends to F. Fixing n, we then find

$$\lim_k \|F \otimes h_n - F_k \otimes h_n\|_2 = \lim_k \|F - F_k\|_2 \|h_n\| = 0.$$

Since $F_k \in \text{span}\ \mathbb{H}_n$ and $h_n \in \text{span}\ \{e_i\}_{i \geq 1}$, we obtain $F_k \otimes H_n \in \text{span}\ \mathbb{H}_n(\Pi)$. Now, since $F_k \otimes h_n$ tends to $F \otimes h_n$ as k tends to infinity, $F \otimes h_n \in \overline{\text{span}}\ \mathbb{H}_n(\Pi)$. And since $F \otimes H_n$ tends to $F \otimes h$ as n tends to infinity, $F \otimes h \in \overline{\text{span}}\ \mathbb{H}_n(\Pi)$. This shows the statement of the theorem. □

Theorem 4.4.13. *The spaces $\mathcal{H}_n(\Pi)$ are orthogonal and $\mathcal{L}^2(\Pi) = \oplus_{n=0}^{\infty} \mathcal{H}_n(\Pi)$.*

Proof. Since the sets $\mathbb{H}_n$ are orthogonal, it is clear that the sets $\mathbb{H}_n(\Pi)$ are orthogonal. Therefore, $\mathcal{H}_n(\Pi)$ are orthogonal as well.

In order to show that $\mathcal{L}^2(\Pi) = \oplus_{n=0}^{\infty} \mathcal{H}_n(\Pi)$, as in Theorem 4.4.3 it will suffice to show that the orthogonal complement of $\cup_{n=0}^{\infty} \mathcal{H}_n(\Pi)$ is zero. Therefore, let $X \in \mathcal{L}^2(\Pi)$, and assume that X is orthogonal to $\mathcal{H}_n(\Pi)$ for any $n \geq 0$. In particular, X is orthogonal to $F \otimes h$ for any $F \in \mathcal{H}_n$ and any $h \in \mathcal{L}^2[0, T]$. Since $\mathcal{L}^2(\mathcal{F}_T) = \oplus_{k=0}^{\infty} \mathcal{H}_k$, we conclude by continuity of the inner product that X is orthogonal to $F \otimes h$ for any $F \in \mathcal{L}^2(\mathcal{F}_T)$ and any $h \in \mathcal{L}^2[0, T]$. By Lemma B.4.2, X is zero almost surely. □

Next, we will investigate the interplay between $\mathcal{H}_n$, $\mathcal{H}_n(\Pi)$ and D. Here, our orthonormal bases $\mathbb{H}_n$ will prove essential.

Theorem 4.4.14. *$\mathbb{H} \subseteq \mathbb{D}_{1,2}$, and for any multi-index a,*

$$D\Phi_a = \sum_{k=1}^{\infty} \alpha_k^a e_k,$$

where $\alpha_k^a = \frac{a_k}{\sqrt{a!}} H_{a_k-1}(\theta e_k) \prod_{i \neq k}^{\infty} H_{a_i}(\theta e_i)$, with the convention that $H_{-1} = 0$. The elements $(D\Phi_a)$ are mutually orthogonal, and $\|D\Phi_a\|^2 = |a|$.

Comment 4.4.15 Since a_k is zero from some point onwards, the infinite product in α_k^a only has finitely many nontrivial factors, and is therefore well-defined. Similarly, we see that α_k^a is zero from some point onwards, and therefore the infinite sum in the expression for $D\Phi_a$ is well-defined, only having finitely many nontrivial terms. ◦

Proof. We need to prove four things: That $\Phi_a \in \mathbb{D}_{1,2}$, the form of $D\Phi_a$, orthogonality and the norm of $\|D\Phi_a\|$.

Computation of $D\Phi_a$. Let $a \in \mathbb{I}$ and assume that $a_k = 0$ for $k > n$. Then $\Phi_a = \frac{1}{\sqrt{a!}} \prod_{k=1}^n H_{a_n}(\theta e_n)$, so $\Phi_a = f(H_{a_1}(\theta e_1), \ldots, H_{a_n}(\theta e_n))$, where $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = \frac{1}{\sqrt{a!}} \prod_{k=1}^n x_k$. By the chain rule, $\Phi_a \in \mathbb{D}_{1,2}$ and

$$\begin{aligned}
D\Phi_a &= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k} \left(H_{a_1}(\theta e_1), \ldots, H_{a_n}(\theta e_n)\right) DH_{a_k}(\theta e_k) \\
&= \sum_{k=1}^{n} \left(\frac{1}{\sqrt{a!}} \prod_{i \neq k}^{n} H_{a_i}(\theta e_i) \right) a_k H_{a_k-1}(\theta e_k) e_k \\
&= \sum_{k=1}^{n} \frac{a_k}{\sqrt{a!}} \left(H_{a_k-1}(\theta e_k) \prod_{i \neq k}^{n} H_{a_i}(\theta e_i) \right) e_k \\
&= \sum_{k=1}^{n} \alpha_k^a e_k \\
&= \sum_{k=1}^{\infty} \alpha_k^a e_k,
\end{aligned}$$

where we have used Lemma A.5.3. Note that in the case where $a_k = 0$, $H_{a_k} = 1$, so $H'_{a_k} = 0 = H_{-1}$. The above computation is therfore valid no matter whether a_k is zero or nonzero.

Orthogonality. Let a and b be multi-indices. Recall that the inner product on $\mathcal{L}^2(\Pi)$ is denoted $\langle \cdot, \cdot \rangle_\Pi$. We have, using Lemma 3.7.4 and recalling that in the sums and products, only finitely many terms and factors are nontrivial,

$$\begin{aligned}
\langle \Phi_a, \Phi_b \rangle_\Pi &= E\langle \Phi_a, \Phi_b \rangle_{[0,T]} \\
&= E\left\langle \sum_{k=1}^\infty \alpha_k^a e_k, \sum_{k=1}^\infty \alpha_k^b e_k \right\rangle_{[0,T]} \\
&= \sum_{k=1}^\infty E\alpha_k^a \alpha_k^b \\
&= \sum_{k=1}^\infty \frac{a_k b_k}{\sqrt{a!}\sqrt{b!}} E\left(H_{a_k-1}(\theta e_k) \prod_{i \neq k}^\infty H_{a_i}(\theta e_i) \right) \left(H_{b_k-1}(\theta e_k) \prod_{i \neq k}^\infty H_{b_i}(\theta e_i) \right) \\
&= \sum_{k=1}^\infty \frac{a_k b_k}{\sqrt{a!}\sqrt{b!}} \left(EH_{a_k-1}(\theta e_k) H_{b_k-1}(\theta e_k) \right) \left(\prod_{i \neq k}^\infty EH_{a_i}(\theta e_i) H_{b_i}(\theta e_i) \right).
\end{aligned}$$

In particular, if $a \neq b$, $\langle \Phi_a, \Phi_b \rangle_\Pi = 0$ by Lemma 4.4.2, showing orthogonality of distinct Φ_a and Φ_b.

Norm of $D\Phi_a$. Finally, we find

$$\begin{aligned}
\|\Phi_a\|_\Pi^2 &= \sum_{k=1}^\infty \frac{a_k^2}{a!} EH_{a_k-1}(\theta e_k)^2 \prod_{i \neq k}^\infty EH_{a_i}(\theta e_i)^2 \\
&= \sum_{k=1}^\infty \frac{a_k^2}{a!} (a_k - 1)! \prod_{i \neq k}^\infty a_i! \\
&= \sum_{k=1}^\infty \frac{a_k^2 (a_k - 1)!}{a_k!} \\
&= |a|.
\end{aligned}$$

□

Theorem 4.4.14 has several corollaries which gives us a good deal of insight into the nature of the Malliavin derivative.

Corollary 4.4.16. *It holds that $\mathcal{H}_n \subseteq \mathbb{D}_{1,2}$, and for any $F \in \mathcal{H}_n$, $\|DF\| = \sqrt{n}\|F\|$.*

Proof. First consider $F \in \mathcal{H}_n''$ with $F = \sum_{k=1}^m \lambda_k \Phi_{a_k}$, where $a_k \in \mathbb{I}_n$ for $k \leq m$. By Lemma 4.4.14 and the Pythagoras Theorem, we obtain $F \in \mathbb{D}_{1,2}$ and

$$\|DF\|^2 = \sum_{k=1}^m \lambda_k^2 \|D\Phi_{a_k}\|^2 = \sum_{k=1}^m \lambda_k^2 |a_k| = n \sum_{k=1}^m \lambda_k^2 = n \sum_{k=1}^m \lambda_k^2 \|\Phi_{a_k}\|^2 = n\|F\|^2.$$

We have now shown that $\mathcal{H}_n'' \subseteq \mathbb{D}_{1,2}$ and that for $F \in \mathcal{H}_n''$, $\|DF\| = \sqrt{n}\|F\|$. We need to extend this to $\mathcal{H}_n$. Therefore, let $F \in \mathcal{H}_n$ be given, and let F_n be a sequence in $\mathcal{H}_n''$ converging to F. By what we already have shown, $\|DF_n - DF_m\| = \sqrt{n}\|F_n - F_m\|$. Since (F_n) is a cauchy sequence, we conclude that DF_n is a cauchy sequence as well. By completeness, there exists a limit X of DF_n. By the closedness of D, we conclude that $F \in \mathbb{D}_{1,2}$ and $DF = X$. We then have

$$\|DF\| = \|\lim DF_n\| = \lim \|DF_n\| = \sqrt{n} \lim \|F_n\| = \sqrt{n}\|F\|.$$

This shows the claims of the corollary. □

Corollary 4.4.17. *D maps $\mathcal{H}_n$ into $\mathcal{H}_{n-1}(\Pi)$.*

Proof. Let $\Phi_a \in \mathbb{H}_n$ be given. By Lemma 4.4.14,

$$D\Phi_a = \sum_{k=1}^{\infty} \alpha_k^a e_k,$$

where $\alpha_k^a = \frac{a_k}{\sqrt{a!}} H_{a_k-1}(\theta e_k) \prod_{i \neq k}^{\infty} H_{a_i}(\theta e_i)$. By inspection, whenever $\alpha_k^a \neq 0$ we have $\alpha_k^a e_k \in \mathcal{H}_{n-1}''(\Pi) \subseteq \mathcal{H}_{n-1}(\Pi)$. By Corollary 4.4.16, D is continuous on $\mathcal{H}_n$. Since $\mathcal{H}_{n-1}(\Pi)$ is a closed linear space, the result therefore extends from $\mathbb{H}_n$ to $\mathcal{H}_n$. □

Corollary 4.4.18. *The images of $\mathcal{H}_n$ and $\mathcal{H}_m$ under D are orthogonal whenever $n \neq m$.*

Proof. This follows by combining Corollary 4.4.17 and Theorem 4.4.13. □

We have now introduced all of the subspaces and orthogonal sets described in the diagram in the beginning of the section, and we have discussed their basic properties. We are now ready to begin work on a characterisation of $\mathbb{D}_{1,2}$ in terms of the projections on the subspaces $\mathcal{H}_n$. Let P_n be the orthogonal projection in $\mathcal{L}^2(\mathcal{F}_T)$ on $\mathcal{H}_n$, and let P_n^{Π} be the orthogonal projection in $\mathcal{L}^2(\Pi)$ on $\mathcal{H}_n(\Pi)$.

Lemma 4.4.19. *If $F \in \mathcal{H}_n''$ and $h \in \mathcal{L}^2[0,T]$, $F(\theta h) \in \mathcal{H}_{n-1} \oplus \mathcal{H}_{n+1}$.*

Proof. We prove the result by first considering some simpler cases and then extending using density arguments.

Step 1: The case $F \in \mathbb{H}_n$ and $h = e_p$. First consider the case where $F \in \mathbb{H}_n$ with $F = \Phi_a$ and $h = e_p$ for some $p \in \mathbb{N}$. Supposing that $a_k = 0$ for $k > m$, we can write $\Phi_a = \frac{1}{\sqrt{a!}} \prod_{i=1}^m H_{a_i}(\theta e_i)$. In the case $p \leq m$, we find, using Lemma A.5.4,

$$\begin{aligned} \Phi_a \theta(e_p) &= \frac{1}{\sqrt{a!}} H_{a_p}(\theta e_p)\theta(e_p) \prod_{i \neq p} H_{a_i}(\theta e_i) \\ &= \frac{1}{\sqrt{a!}} \left(H_{a_p+1}(\theta e_p) + a_p H_{a_p-1}(\theta e_p) \right) \prod_{i \neq p} H_{a_i}(\theta e_i) \\ &= \frac{1}{\sqrt{a!}} H_{a_p+1}(\theta e_p) \prod_{i \neq p} H_{a_i}(\theta e_i) + \frac{1}{\sqrt{a!}} a_p H_{a_p-1}(\theta e_p) \prod_{i \neq p} H_{a_i}(\theta e_i), \end{aligned}$$

which is in $\mathcal{H}_{n-1} \oplus \mathcal{H}_{n+1}$. If $p > m$, we obtain $\Phi_a \theta(e_p) = \frac{1}{\sqrt{a!}} H_1(\theta e_p) \prod_{i=1}^m H_{a_i}(\theta e_i)$, yielding $\Phi_a \theta(e_p) \in \mathcal{H}_{n+1} \subseteq \mathcal{H}_{n-1} \oplus \mathcal{H}_{n+1}$.

Step 2: The case $F \in \mathbb{H}_n$ and $h \in \mathcal{L}^2[0,T]$. By linearity, it is clear that the result extends to $F = \Phi_a$ and h in the span of $\{e_i\}_{i \geq 1}$. Consider a general $h \in \mathcal{L}^2[0,T]$, and let h_n be in the span of $\{e_i\}_{i \geq 1}$, converging to h. Now, the mapping

$$f(x) = \left(\frac{1}{\sqrt{a!}} \prod_{i=1}^m H_{a_i}(x_i) \right) x_{m+1}$$

is in $C_p^\infty(\mathbb{R}^{m+1})$, and furthermore we have the equalities $\Phi_a \theta(h) = f(\theta e_1, \ldots, \theta e_m, \theta h)$ and $\Phi_a \theta(h_n) = f(\theta e_1, \ldots, \theta e_m, \theta h_n)$. Therefore, Lemma 4.4.7 yields that $\Phi_a \theta(h_n)$ converges in $\mathcal{L}^2(\mathcal{F}_T)$ to $\Phi_a \theta(h)$. Since $\mathcal{H}_{n-1} \oplus \mathcal{H}_{n+1}$ is a closed subspace, we may conclude $\Phi_a \theta(h) \in \mathcal{H}_{n-1} \oplus \mathcal{H}_{n+1}$.

Step 3: The case $F \in \mathcal{H}_n''$ and $h \in \mathcal{L}^2[0,T]$. The remaining extension follows directly by linearity as in the previous step. □

Lemma 4.4.20. *If $X \in \mathcal{H}_n(\Pi)$ and $h \in \mathcal{L}^2[0,T]$, $\langle X, h \rangle_{[0,T]} \in \mathcal{H}_n$.*

Proof. Let $h \in \mathcal{L}^2[0,T]$. First note that for any $F \in \mathcal{H}_n$ and $h' \in \mathcal{L}^2[0,T]$, we find $\langle F \otimes h', h \rangle_{[0,T]} = F\langle h', h \rangle \in \mathcal{H}_n$, showing the result in this simple case. Since the mapping $X \mapsto \langle X, h \rangle_{[0,T]}$ is linear and continuous and $\mathcal{H}_n$ is a closed subspace, the result extends to $X \in \mathcal{H}_n(\Pi)$. □

Theorem 4.4.21. *Let $F \in \mathbb{D}_{1,2}$. Then $P_n^\Pi DF = DP_{n+1}F$.*

Proof. Our plan is to show that $\langle Z, DF \rangle = \langle Z, DP_{n+1}F \rangle$ for any $Z \in \mathcal{H}_n(\Pi)$ and then use Lemma A.6.18 to obtain the conclusion. Showing this equality in full generality will be done in two steps, first demonstrating it in a simple case and then extending by linearity and continuity.

Step 1: The equality for $G \otimes h$. Let $G \in \mathcal{H}_n''$ and let $h \in \mathcal{L}^2[0,T]$. We will show that $\langle G \otimes h, DF \rangle = \langle G \otimes h, DP_{n+1}F \rangle$. Since $\mathcal{H}_n'' \subseteq \mathcal{S}_h$, we can use Corollary 4.3.11 and obtain

$$\begin{aligned}\langle G \otimes h, DF \rangle &= E(G\langle DF, h\rangle_{[0,T]}) \\ &= EFG(\theta h) - EF\langle DG, h\rangle_{[0,T]}.\end{aligned}$$

Using the orthogonal decomposition of $\mathcal{L}^2(\mathcal{F}_T)$ proven in Theorem 4.4.3, Lemma A.6.12 yields $F = \sum_{k=0}^\infty P_kF$, where the limit is in $\mathcal{L}^2(\mathcal{F}_T)$. Using the Cauchy-Schwartz inequality, we then obtain the two results

$$\begin{aligned}&\sum_{k=0}^\infty \|G(\theta h)P_kF\|_1 \le \|G(\theta h)\|_2 \sum_{k=0}^\infty \|P_kF\|_2 < \infty \\ &\sum_{k=0}^\infty \|\langle DG, h\rangle_{[0,T]}P_kF\|_1 \le \|\langle DG, h\rangle_{[0,T]}\|_2 \sum_{k=0}^\infty \|P_kF\|_2 < \infty,\end{aligned}$$

so $\sum_{k=0}^\infty G(\theta h)P_kF$ and $\sum_{k=0}^\infty \langle DG, h\rangle_{[0,T]}P_kF$ are convergent in $\mathcal{L}^1$. Since convergence in $\mathcal{L}^p$ implies convergence in probability, we conclude by uniqueness of limits that

$$\begin{aligned}\sum_{k=0}^\infty G(\theta h)P_kF &= G(\theta h)\sum_{k=0}^\infty P_kF \\ \sum_{k=0}^\infty \langle DG, h\rangle_{[0,T]}P_kF &= \langle DG, h\rangle_{[0,T]}\sum_{k=0}^\infty P_kF,\end{aligned}$$

where the limits on the left are in $\mathcal{L}^1(\mathcal{F}_T)$ and the limits on the right are in $\mathcal{L}^2(\mathcal{F}_T)$. This implies

$$\begin{aligned}E(G\langle DF, h\rangle_{[0,T]}) &= EFG(\theta h) - EF\langle DG, h\rangle_{[0,T]} \\ &= E\left(\sum_{k=0}^\infty G(\theta h)P_kF\right) - E\left(\sum_{k=0}^\infty \langle DG, h\rangle_{[0,T]}P_kF\right) \\ &= \sum_{k=0}^\infty EG(\theta h)P_kF - \sum_{k=0}^\infty E\langle DG, h\rangle_{[0,T]}P_kF,\end{aligned}$$

where we have used the $\mathcal{L}^1$-convergence to interchange sum and integral. By assumption, $G \in \mathcal{H}_n''$, so $DG \in \mathcal{H}_{n-1}(\Pi)$ by Corollary 4.4.17, and $\langle DG, h\rangle_{[0,T]} \in \mathcal{H}_{n-1}$ by Lemma 4.4.20. Furthermore, by Lemma 4.4.19, $G\theta(h) \in \mathcal{H}_{n-1} \oplus \mathcal{H}_{n+1}$. By orthogonality of the $\mathcal{H}_n$, we then find

$$\begin{aligned} & \sum_{k=0}^{\infty} EG(\theta h)P_k F - \sum_{k=0}^{\infty} E\langle DG, h\rangle_{[0,T]} P_k F \\ = \quad & EG(\theta h)P_{n-1}F + EG(\theta h)P_{n+1}F - E\langle DG, h\rangle_{[0,T]} P_{n-1}F. \end{aligned}$$

By the integration-by-parts formula of Corollary 4.3.11,

$$EG(\theta h)P_{n-1}F - E\langle DG, h\rangle_{[0,T]} P_{n-1}F = EG\langle DP_{n-1}F, h\rangle_{[0,T]},$$

which is zero since $G \in \mathcal{H}_n''$ and $\langle DP_{n-1}F, h\rangle_{[0,T]} \in \mathcal{H}_{n-2}$ by Lemma 4.4.19. Since also $E\langle DG, h\rangle_{[0,T]} P_{n+1}F = 0$ by the same lemma, we can use Corollary 4.3.11 once again to obtain

$$\begin{aligned} & EG(\theta h)P_{n-1}F + EG(\theta h)P_{n+1}F - E\langle DG, h\rangle_{[0,T]} P_{n-1}F \\ = \quad & EG(\theta h)P_{n+1}F \\ = \quad & EG(\theta h)P_{n+1}F - E\langle DG, h\rangle_{[0,T]} P_{n+1}F \\ = \quad & EG\langle DP_{n+1}F, h\rangle_{[0,T]} \\ = \quad & \langle G \otimes h, DP_{n+1}F\rangle. \end{aligned}$$

All in all, we have proven for $G \in \mathcal{H}_n''$ that

$$\langle G \otimes h, DF\rangle = \langle G \otimes h, DP_{n+1}F\rangle.$$

Step 2: Conclusions. Noting that both the left-hand side and the right-hand side in the equality above is continuous and linear in G, we can extend it to hold not only for $G \in \mathcal{H}_n''$, but also for $G \in \mathcal{H}_n$. Since the variables of the form $G \otimes h$ for $G \in \mathcal{H}_n$ and $h \in \mathcal{L}^2[0,T]$ has dense span in $\mathcal{H}_n(\Pi)$ by definition, we conclude, again by continuity and linearity of the inner product, that $\langle Z, DF\rangle = \langle Z, DP_{n+1}F\rangle$ for all $Z \in \mathcal{H}_n(\Pi)$. Since $DP_{n+1}F \in \mathcal{H}_n(\Pi)$ by Corollary 4.4.17, Lemma A.6.18 finally yields that $DP_{n+1}F$ is the orthogonal projection of DF onto $\mathcal{H}_n(\Pi)$, that is, $P_n^{\Pi} DF = DP_{n+1}F$. □

Theorem 4.4.22. *Let $F \in \mathcal{L}^2(\mathcal{F}_T)$. F is in $\mathbb{D}_{1,2}$ if and only if $\sum_{n=1}^{\infty} n\|P_n F\|^2$ is finite. In the affirmative case, $DF = \sum_{n=0}^{\infty} DP_n F$, and in particular the norm is given by $\|DF\|^2 = \sum_{n=1}^{\infty} n\|P_n F\|^2$.*

Proof. First assume that $F \in \mathcal{L}^2(\mathcal{F}_T)$ and that $\sum_{n=1}^{\infty} n\|P_n F\|_2^2 < \infty$. We have to show that $F \in \mathbb{D}_{1,2}$ and identify DF and $\|DF\|^2$. By Lemma A.6.12, $F = \sum_{n=0}^{\infty} P_n F$.

Define $F_k = \sum_{n=0}^k P_n F$, then F_k converges to F in $\mathcal{L}^2(\mathcal{F}_T)$. We will argue that DF_k converges as well, this will imply $F \in \mathbb{D}_{1,2}$.

Obviously, $P_n F \in \mathcal{H}_n$, so by Corollary 4.4.16, $P_n F \in \mathbb{D}_{1,2}$ and $\|DP_n F\|^2 = n\|P_n F\|^2$. Therefore, we find $\sum_{n=0}^\infty \|DP_n F\|^2 = \sum_{n=0}^\infty n\|P_n F\|^2 < \infty$, showing that the series $\sum_{n=0}^\infty DP_n F$ is convergent in $\mathcal{L}^2(\mathcal{F}_T)$, meaning that the sequence DF_n is convergent. By closedness of D, we conclude $F \in \mathbb{D}_{1,2}$ and $DF = \sum_{n=0}^\infty DP_n F$. By Corollary 4.4.18, $DP_n F$ and $DP_m F$ are orthogonal whenever $n \neq m$. We may therefore also conclude $\|DF\|^2 = \sum_{n=0}^\infty \|DP_n F\|^2 = \sum_{n=1}^\infty n\|P_n F\|^2$.

It remains to show that $\sum_{n=1}^\infty n\|P_n F\|^2 < \infty$ is also necessary to obtain $F \in \mathbb{D}_{1,2}$. Assume that $F \in \mathbb{D}_{1,2}$, we need to prove that the sum is convergent. To this end, note that using the orthogonal decomposition of $\mathcal{L}^2(\Pi)$ from Theorem 4.4.13 and applying Lemma A.6.12 and Theorem 4.4.21,

$$DF = \sum_{n=0}^\infty P_n^\Pi DF = \sum_{n=0}^\infty DP_{n+1} F = \sum_{n=1}^\infty DP_n F.$$

In particular, the sum on the right is convergent, and therefore $\sum_{n=1}^\infty \|DP_n F\|^2$ is finite. By Corollary 4.4.16, $\|DP_n F\|^2 = n\|P_n F\|^2$, and we conclude that $\sum_{n=1}^\infty n\|P_n F\|^2$ is finite, as desired. □

Theorem 4.4.22 is the promised characterisation of $\mathbb{D}_{1,2}$ in terms of the subspaces $\mathcal{H}_n$. As our final result on the Malliavin calculus, we will show how the theorem can be used to obtain an extension of the chain rule for Lipschitz mappings. We are going to employ some weak convergence results, see Appendix A.6 for an overview.

Lemma 4.4.23. *Let F_n be a sequence in $\mathbb{D}_{1,2}$ converging to F and assume that DF_n is bounded in $\mathcal{L}^2(\Pi)$. Then $F \in \mathbb{D}_{1,2}$, and there is a subsequence of DF_n converging weakly to DF.*

Proof. We will use Theorem 4.4.22 to argue that $F \in \mathbb{D}_{1,2}$. To do so, we need to prove that $\sum_{m=1}^\infty m\|P_m F\|^2$ is finite. To this end, first note that since DF_n is bounded in $\mathcal{L}^2(\Pi)$, by Theorem A.6.21 there exists a subsequence DF_{n_k} converging weakly to some $\alpha \in \mathcal{L}^2(\Pi)$. Convergence of F_{n_k} to F and Corollary 4.4.16 allows us to conclude

$$\sum_{m=1}^\infty m\|P_m F\|^2 = \sum_{m=1}^\infty \lim_k m\|P_m F_{n_k}\|^2 = \sum_{m=1}^\infty \lim_k \|DP_m F_{n_k}\|^2.$$

Now note that $P_m F_{n_k}$ is a sequence in $\mathcal{H}_m$ which converges to $P_m F$ since P_m is continous, being an orthogonal projection. By Corollary 4.4.16, D is a continuous mapping on $\mathcal{H}_m$, so we find that $DP_m F_{n_k}$ converges to $DP_m F$. On the other hand, by Theorem 4.4.21, $DP_m F_{n_k} = P^{\Pi}_{m-1} DF_{n_k}$. Since DF_{n_k} converges weakly to α, $P^{\Pi}_{m-1} DF_{n_k}$ converges weakly to $P^{\Pi}_{m-1}\alpha$ by Lemma A.6.23. Since ordinary convergence implies weak convergence by Lemma A.6.20 and weak limits are unique by Lemma A.6.19, we obtain $DP_m F = P^{\Pi}_{m-1}\alpha$. All in all, we conclude that $DP_m F_{n_k}$ converges to $P^{\Pi}_{m-1}\alpha$. Therefore, the norms converge as well. Thus, using Lemma A.6.12,

$$\sum_{m=1}^{\infty} \lim_k \|DP_m F_{n_k}\|^2 = \sum_{m=1}^{\infty} \|P^{\Pi}_{m-1}\alpha\|^2 = \|\alpha\|^2,$$

which is of course finite. Theorem 4.4.22 now yields $F \in \mathbb{D}_{1,2}$. Finally, note that $P^{\Pi}_m DF = DP_{m+1} F = P^{\Pi}_m \alpha$ for any $m \geq 0$, so by Lemma A.6.12, $DF = \alpha$. This means that DF_{n_k} converges weakly to DF. □

Corollary 4.4.24. *Let F_n be a sequence in $\mathbb{D}_{1,2}$ converging to F and assume that DF_n is bounded in $\mathcal{L}^2(\Pi)$. Then $F \in \mathbb{D}_{1,2}$, and DF_n converges weakly to DF.*

Proof. By Lemma 4.4.23, $F \in \mathbb{D}_{1,2}$. We need to show that DF_n converges weakly to DF. To this end, it will suffice to show that for any subsequence DF_{n_k}, there is a further subsequence $DF_{n_{k_l}}$ converging to DF. Therefore, let a subsequence DF_{n_k} be given. Then F_{n_k} converges to F and DF_{n_k} is bounded in $\mathcal{L}^2(\Pi)$, so Lemma 4.4.23 shows that there is a subsequence $F_{n_{k_l}}$ converging to DF. This concludes the proof. □

Theorem 4.4.25 (Lipschitz chain rule). *Let $F \in \mathbb{D}^n_{1,2}$, $F = (F_1, \ldots, F_n)$, and let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz mapping with Lipschitz constant K with respect to $\|\cdot\|_\infty$. Then $\varphi(F) \in \mathbb{D}_{1,2}$, and there is a random variable $G = (G_1, \ldots, G_n)$ with values in $\mathbb{R}^n$ and $\|G_k\|_2 \leq K$ such that $D\varphi(F) = \sum_{k=1}^n G_k DF_k$.*

Proof. By Lemma A.4.2, there exists a sequence of mappings $\varphi_n \in C^\infty(\mathbb{R}^n)$ with partial derivatives bounded by K converging uniformly to φ. By the ordinary chain rule of Theorem 4.3.1, $D\varphi_n(F) = \sum_{k=1}^n \frac{\partial \varphi_n}{\partial x_k}(F) DF_k$. Now, since φ_n converges uniformly to φ, $\varphi_n(F)$ converges in $\mathcal{L}^2(\mathcal{F}_T)$ to $\varphi(F)$. On the other hand, we have

$$\left\| \sum_{k=1}^n \frac{\partial \varphi_n}{\partial x_k}(F) DF_k \right\|_2 \leq K \sum_{k=1}^n \|DF_k\|_2,$$

so the sequence $\sum_{k=1}^n \frac{\partial \varphi_n}{\partial x_k}(F) DF_k$ is bounded in $\mathcal{L}^2(\Pi)$. By Corollary 4.4.24 we obtain $\varphi(F) \in \mathbb{D}_{1,2}$ and $\sum_{k=1}^n \frac{\partial \varphi_n}{\partial x_k}(F) DF_k$ converges weakly to $D\varphi(F)$. If we can identify the weak limit as being of the form stated in the theorem, we are done. To this end, note that since the sequence $\frac{\partial \varphi_n}{\partial x_k}(F)$ is pointwisely bounded by K, it is bounded in $\mathcal{L}^2(\mathcal{F}_T)$ by K as well. Therefore, by Theorem A.6.21 there exists a subsequence such that for all $k \leq n$, $\frac{\partial \varphi_{n_m}}{\partial x_k}(F)$ converges weakly to some $G_k \in \mathcal{L}^2(\mathcal{F}_T)$. From Lemma A.6.22 we know that $\|G_k\|_2 \leq K$, but we would like a pointwise bound instead. To obtain this, let $A \in \mathcal{F}_T$. We then obtain

$$\begin{aligned} E1_A G_k^2 &= \langle 1_A G_k, G_k \rangle \\ &= \lim_m \left\langle 1_A G_k, \frac{\partial \varphi_{n_m}}{\partial x_k}(F) \right\rangle \\ &= E1_A G_k \frac{\partial \varphi_{n_m}}{\partial x_k}(F) \\ &\leq K1_A |G_k|. \end{aligned}$$

This shows in particular that $G_k^2 \leq K|G_k|$ almost surely, and therefore $|G_k| \leq K$ almost surely. Now, for any bounded element X of $\mathcal{L}^2(\Pi)$, we find

$$\begin{aligned} \left\langle X, \frac{\partial \varphi_{n_m}}{\partial x_k}(F) DF_k - G_k DF_k \right\rangle &= \left\langle X, \left(\frac{\partial \varphi_{n_m}}{\partial x_k}(F) - G_k \right) DF_k \right\rangle \\ &= E\left(\int_0^T X(t) \left(\frac{\partial \varphi_{n_m}}{\partial x_k} - G_k \right) DF_k(t) \, \mathrm{d}t \right) \\ &= E\left(\int_0^T X(t) DF_k(t) \, \mathrm{d}t \right) \left(\frac{\partial \varphi_{n_m}}{\partial x_k} - G_k \right). \end{aligned}$$

Since X is bounded and $DF_k \in \mathcal{L}^2(\Pi)$, we conclude by Jensen's inequality that $\int_0^T X(t) DF_k(t) \, \mathrm{d}t \in \mathcal{L}^2(\mathcal{F}_T)$. By the weak convergence of $\frac{\partial \varphi_{n_m}}{\partial x_k}$ to G_k, the above therefore tends to zero. We may now conclude, still letting $X \in \mathcal{L}^2(\Pi)$ be bounded,

$$\left\langle X, D\varphi(F) - \sum_{k=1}^n G_k DF_k \right\rangle = \lim_m \left\langle X, \sum_{k=1}^n \frac{\partial \varphi_{n_m}}{\partial x_k}(F) DF_k - \sum_{k=1}^n \frac{\partial \varphi_{n_m}}{\partial x_k}(F) DF_k \right\rangle = 0.$$

Since the bounded elements of $\mathcal{L}^2(\Pi)$ are dense, by Lemma A.6.2 and A.6.13, we conclude $D\varphi(F) = \sum_{k=1}^n G_k DF_k$. This proves the theorem. □

This concludes our exposition of this Malliavin calculus. In the final next section, we give a survey of the further theory of the Malliavin calculus.

4.5 Further theory

Our description of the Malliavin calculus in the preceeding sections only covers the basics of theory, and is in fact quite inadequate for applications. We will now describe some of the further results of the theory and outline the main theoretical areas where Malliavin calculus finds application. All of these areas are covered in more or less detail in Nualart (2006).

The Skorohod integral. The largest omission in our exposition is clearly that of the Skorohod integral. The Skorohod integral δ is defined as the adjoint operator of the Malliavin derivative D. This is a linear operator defined on a dense subspace $\mathbb{S}_{1,2}$ of $\mathcal{L}^2(\Pi)$ mapping into $\mathcal{L}^2(\mathcal{F}_T)$, characterised by the duality relationship

$$\langle F, \delta u\rangle_{\mathcal{F}_T} = \langle DF, u\rangle_{\Pi}$$

for any $u \in \mathbb{S}_{1,2}$. The existence of such an operator follows from the closedness of D combined with Lemma 19.2 and Proposition 19.5 of Meise & Vogt (1997). The Skorohod integral has two extremely important properties which connect it to the theory of stochastic integration. First off, if $u \in \mathcal{L}^2(\Pi)$ is progressively measurable, it holds that

$$\delta(u) = \int_0^T u_s \,\mathrm{d}W_s,$$

where the integral on the right is the ordinary Itô stochastic integral. This justifies the name "Skorohod integral" for δ, and shows that the Skorohod integral is an extension of the Itô stochastic integral. Inspired by this fact, we will use the notation $\int_0^T u_s \delta W_s$ for $\delta(u)$. The second important property is as follows. Writing $D_t F = (DF)_t$, if u is such that $u_t \in \mathbb{D}_{1,2}$ for all $t \le T$ and such that there is a measurable version of $(t,s) \mapsto D_t u_s$ and a measurable version of $\int_0^T D_t u_s \delta W_s$, then

$$D_t \int_0^T u_s \delta W_s = \int_0^T D_t u_s \delta W_s + u_t.$$

In the case where u is progressively measurable, this shows how to differentiate a general Itô stochastic integral.

The properties of the Skorohod integral are usually investigated through the Itô-Wiener expansion, which is a series expansion for variables $F \in \mathcal{L}^2(\mathcal{F}_T)$. The basic content of the expansion is, with $\Delta_n = \{t \in [0,T]^n | t_1 \le \cdots \le t_n\}$, that there exists square-integrable deterministic functions f_n on Δ_n such that

$$F = \sum_{n=0}^{\infty} n! \int_0^T \int_0^{t_n} \cdots \int_0^{t_3} \int_0^{t_2} f_n(t) \,\mathrm{d}W(t_1) \cdots \mathrm{d}W(t_n),$$

where the convergence is in $\mathcal{L}^2$. Expanding a process on $[0,T]$ in this manner, one may then identify a necessary and sufficient criterion on the mappings f_n to ensure that the process is adapted. This can be used to prove the relationship between the Skorohod integral and the Itô integral. All of this is described in Section 1.3 of Nualart (2006).

Differentiation of SDEs. Being a stochastic calculus, it seems obvious that the Malliavin calculus should yield useful results when applied to the theory of stochastic differential equations. The main result on this topic is described in Section 2.2 of Nualart (2006) and basically states that under suitable regularity conditions, when taking the Malliavin derivative of the solution to a stochastic differential equation, interchange of differentiation and integration is allowed. This result is of fundamental importance to the applications in financial mathematics.

Regularity of densities. The Malliavin calculus can also be used to give abstract criteria for when a random variable possesses a density with respect to the Lebesgue measure, and criteria for when such a density possesses some degree of smoothness. This is shown in Section 2.1 of Nualart (2006). The basic results are Theorem 2.1.1 and Theorem 2.1.4, yielding sufficient criteria in terms of the Malliavin derivatives of a vector variable for the existence and smoothness of a density with respect to the Lebesgue measure. A further important result is found as Theorem 2.3.3, giving a sufficient criterion for the solution of a stochastic differential equation to have a smooth density.

4.6 Notes

We will in this section review our results and compare them to other accounts of the same results in the theory.

Our exposition of the theory is rather limited in its reach, striving primarily for rigor and detail. The theory is based almost entirely on Section 1.1 and Section 2.1 of Nualart (2006). Nualart (2006) does not always provide many details, so the virtue of our results is mostly the added detail. The main difficulties has been to provide details of the chain rule of Theorem 4.3.1 and the relations between the orthogonal decompositions of $\mathcal{L}^2(\mathcal{F}_T)$ and $\mathcal{L}^2(\Pi)$, culminating in the proof of Theorem 4.4.21. Further work has gone into the extension of the chain rule given in Theorem 4.3.5, which hopefully will be useful in the later development of the theory.

The Malliavin calculus is a relatively young theory, originating in the 1970s, and even though the recent applications in finance such as the results in Fournié et al. (1999) seem to have motivated an increase in production of material related to the theory, there are still quite few sources for the theory.

The most well-known book on the Malliavin calculus is Nualart (2006), which as noted has been the fundamental source for all of the theory described in this chapter. Other books on the topic are Bell (1987), Bass (1998), Malliavin (1997) and Sanz-Solé (2005).

In general, these books are all, to put it bluntly, either not very rigorous or somewhat inaccessible. Nualart (2006) is clearly the best exposition of these, but is still difficult compared to the literature available for, say, stochastic integration. We will now discuss the merits and demerits of each book.

The book by Bell (1987) is very small, and in general either defers proofs to other works or only gives sketches of proofs, making it difficult to use as an introduction. The book Bass (1998) does not have the Malliavin calculus as its main topic, and only spends the final chapter exploring it. Considering this, it cannot be blamed for not developing the theory in detail. Still, it is not useful as an introductory work.

The work Sanz-Solé (2005) looks very useful at a first glance, with a nice layout and a seemingly pedagogical introduction. However, it also defers many proofs to other books and often entirely skips proofs. This work cannot be recommended as an good introduction either.

Another work on the Malliavin calculus is Malliavin (1997). This book is not only about the Malliavin calculus, but touches upon a good deal of different topics. In any case, the main problem with the book is that it is so impressively abstract that it seems almost made to be incomprehensible on purpose.

This leaves Nualart (2006). While it, as noted earlier, is not very accessible, it is clearly superior to the other books. It is cast in a modern manner, and does not make the theory unnecessarily abstract. Our account differs from that one by considering a one-dimensional Brownian motion instead of what is known as an isonormal gaussian process. The approach based on isonormal processes restricts the theory to the context of Hilbert spaces. This would make the introductory results of Section 4.2 prettier, but would also make motivation for the definitions more difficult to comprehend.

In general, what makes the book difficult is that it very often omits a great deal of detail. This scope of these omissions can be estimated by noting that the theory which have been developed in 40 pages here and at least 15 pages more in the appendix is developed in only 10 pages in Nualart (2006).

Complementing these books, there are also notes available on the internet documenting the Malliavin calculus. Some of these are Zhang (2004), Bally (2003), Friz (2002) and Øksendal (1997). Of these, only Øksendal (1997) is useful.

The notes Bally (2003) and Friz (2002) are very easy to identify as useless. Both engage in what best can be described as some sort of mathematical name-dropping, often invoking Sobolev spaces, distribution theory and other theories, without these invocations ever resulting in any actual proofs. Furthermore, proofs are in general heuristic or omitted. Zhang (2004) at a first glance looks very good. As with Sanz-Solé (2005), the layout is professional and the author quickly obtains an air of having a good overview of the theory. However, on closer inspection most of the proofs have conspicuously similar levels of detail as in Nualart (2006), thereby not really adding anything new.

The final note we discuss is Øksendal (1997). This is by far the most accessible text available on the Malliavin calculus. It develops the theory in manner very different from Nualart (2006), basing the definition of the Malliavin derivative and Skorohod integral directly on the Wiener-Itô expansion. The proofs are in general given in detail and are very readable. Unfortunately, the note is short and does not cover so much of the theory, so at some point one has to revert to other works. However, as an introduction, it is most definitely very useful.

Complementing these works is the forthcoming book Di Nunno et al. (2008). It extends the theory of the Malliavin calculus to processes other than Brownian motion, but does also cover the Brownian case in detail. Even though the final work here is based on Nualart (2006), Di Nunno et al. (2008) has been invaluable as a first introduction, both to the Malliavin calculus in general and also for the applications to finance, because of its high level of detail and readability.

Chapter 5

Mathematical Finance

In this chapter, we will develop the theory of arbitrage pricing for a very basic class of models for financial markets. Our goals is to find sufficient criteria for the markets to be free of arbitrage, to apply these criteria to some simple models and to show how to calculate prices and sensitivities in these models.

For ease of notation, we will in general use the shorthand $\mathrm{d}X_t = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$ for $X_t = X_0 + \int_0^t \mu_t\,\mathrm{d}t + \int_0^t \sigma_t\,\mathrm{d}W_t$ for some X_0.

5.1 Financial market models

We work in the context of a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ with a n-dimensional $\mathcal{F}_t$ Brownian motion W. We assume that the usual conditions hold, in particular we assume that $\mathcal{F}_t$ is the usual augmentation of the filtration generated by W. Furthermore, we assume that $\mathcal{F} = \mathcal{F}_\infty = \sigma(\cup_{t\geq 0}\mathcal{F}_t)$.

We begin by defining the fundamental building blocks of what will follow. We will assume given a one-dimensional stochastic short rate process r and define the risk-free asset price process B as the solution to $\mathrm{d}B_t = r_t B_t\,\mathrm{d}t$ with B_0 some positive constant, equivalent to putting $B_t = B_0 \exp(\int_0^t r_s\,\mathrm{d}s)$. We furthermore assume given m financial asset price processes $(S_1, \ldots, S_m)$, which we assume to be nonnegative standard processes.

Together, the short rate, the risk-free asset price process and the financial asset price processes define a financial market model $\mathcal{M}$. In this market, a portfolio strategy is a pair of locally bounded progressive processes $h = (h^0, h^S)$, where h^0 is one-dimensional and h^S is m-dimensional. The interpretation is that $h^0(t)$ denotes the number of units held at time t in the risk-free asset B, and $h^S_k(t)$ denotes the number of units held at time t in the k'th risky asset S_k. To a given portfolio strategy, we associate the value process V^h, defined by $V^h(t) = h^0(t)B(t) + \sum_{k=1}^m h^S_k(t)S_k(t)$. We say that h is admissible if V^h is nonnegative. We say that the portfolio h is self-financing if it holds that

$$\mathrm{d}V^h(t) = h^0(t)\,\mathrm{d}B(t) + \sum_{k=1}^m h^S_k(t)\,\mathrm{d}S_k(t).$$

Here, the integrals are always well-defined because h is assumed to be locally bounded and progressive. The above essentially means that the change in value in the portfolio only comes from profits and losses in the portfolio, there is no exogenous intake or outtake of wealth. That this is the correct criterion for no exogenous infusions of wealth is not as obvious as it may seem. It is important that we are using the Itô integral and not, say, the Stratonovich integral. For an argument showing that the Itô integral yields the correct interpretation, see Chapter 6 of Björk (2004) and compare with the Riemann approximations for the Itô and Stratonovich integrals of Lemma IV.47.1 and Lemma IV.47.3 in Rogers & Williams (2000b).

We are now ready to make the central definition of this section, the notion of an arbitrage opportunity.

Definition 5.1.1. *We say that an admissible and self-financing portfolio is an arbitrage opportunity at time T if the value process V^h satisfies $V^h_0 = 0$, $P(V^h_T \geq 0) = 1$ and $P(V^h_T > 0) > 0$. If a market $\mathcal{M}$ contains no arbitrage opportunities at time T, we say that it is T-arbitrage free. If a market contains no arbitrage opportunities at any time, we say that it is arbitrage free.*

Comment 5.1.2 The notion of being arbitrage free is the basic requirement for a market to be realistic. It it reasonable to state that most real-world markets are arbitrage-free most of the time. Therefore, if we are considering a market which has arbitrage, we are some distance away from anything useful for real-world modeling.

The three requirements for h to constitute an arbitrage opportunity can be put in words as:

1. It must be possible to enter the strategy without cost.
2. The strategy should never result in a loss of money.
3. The stragegy should have a positive probability of resulting in a profit.

Note that for a portfolio to be an arbitrage opportunity, it must be both self-financing and admissible. Clearly, a portfolio satisfying the three requirements above but which is not self-financing does not say anything about whether our market is realistic, anyone can generate a sure profit if exogenous money infusions are allowed. The reason we also require admissibility is to rule out possibilities for “doubling schemes” - borrowing money until a profit turns out. For more on doubling schemes, see Steele (2000), Section 14.5 or Example 1.2.3 of Karatzas & Shreve (1998). ∘

We are interested in finding a sufficient criterion for a market to be arbitrage free. To formulate our main result on this, we first need to introduce the concept of a normalized market. Given our financial instrument vector S and the risk-free asset B, we define the normalized instruments by $S'_k(t) = \frac{S_k(t)}{B(t)}$. The instruments S'_k can be thought of as the discounted versions of S_k.

Lemma 5.1.3. *S'_k is a standard process with the dynamics*

$$\mathrm{d}S'_k(t) = \frac{1}{B(t)}\,\mathrm{d}S_k(t) - S'_k(t)r(t)\,\mathrm{d}t.$$

Proof. Since B is positive, we can use Itô's lemma in the form of Corollary 3.6.4 on the $C^2(\mathbb{R}^2 \setminus \{0\})$ mapping $(x, y) \mapsto \frac{x}{y}$ and obtain

$$\begin{aligned}
\mathrm{d}S'_k(t) &= \frac{1}{B(t)}\,\mathrm{d}S_k(t) - \frac{S_k(t)}{B(t)^2}\,\mathrm{d}B(t) - \frac{1}{2B(t)^2}\,\mathrm{d}[S_k, B]_t + \frac{S_k(t)}{B(t)^3}\,\mathrm{d}[B]_t \\
&= \frac{1}{B(t)}\,\mathrm{d}S_k(t) - \frac{S_k(t)}{B(t)^2}r(t)B(t)\,\mathrm{d}t \\
&= \frac{1}{B(t)}\,\mathrm{d}S_k(t) - S'_k(t)r(t)\,\mathrm{d}t,
\end{aligned}$$

as desired. □

Lemma 5.1.3 shows that the processes $S' = (S'_1, \ldots, S'_m)$ also describe financial instruments in the sense introduced earlier in this section, the processes are nonnegative standard processes. In particular, we can consider the normalized market consisting of a trivial risk-free asset $B' = 1$ and the normalized instruments S'. Our next lemma shows that to check arbitrage for the general market, it will suffice to consider the normalized market.

Lemma 5.1.4. *The market with instruments S and risk-free asset B is T-arbitrage free if the normalized market with instruments S' and risk-free asset $B' = 1$ is T-arbitrage free.*

Proof. Assume that the normalized market is free of arbitrage at time T. Let the portfolio $h = (h^0, h^S)$ be an arbitrage opportunity at time T for the original market. We will work to obtain a contradiction. Let V_S^h be the value process when considering h as a portfolio strategy for the original market. Let $V_{S'}^h$ be the value process when considering h as a portfolio strategy for the normalized market. Our goal is to show that the portfolio h also is an arbitrage under the normalized market.

We first examine $V_{S'}^h$. Note that since $B' = 1$, we obtain

$$V_{S'}^h(t) = h^0(t) + \sum_{k=1}^m h_k^S(t) S_k'(t) = \frac{1}{B(t)}\left(h^0(t)B(t) + \sum_{k=1}^m h_k^S(t) S_k(t)\right) = \frac{V_S^h(t)}{B(t)}.$$

In particular, since B is positive, we find that $V_{S'}^h$ is nonnegative, so h is admissible under the normalized market. To show that it is self-financing, we use Itô's lemma and Lemma 5.1.3 to obtain

$$\begin{aligned}
\mathrm{d}V_{S'}^h(t) &= \frac{1}{B(t)}\,\mathrm{d}V_S^h(t) - \frac{V_S^h(t)}{B(t)^2}\,\mathrm{d}B(t) - \frac{1}{2B(t)^2}\,\mathrm{d}[V_S^h, B]_t + \frac{V_S^h(t)}{B(t)^3}\,\mathrm{d}[B]_t \\
&= \frac{1}{B(t)} h^0(t)\,\mathrm{d}B(t) + \frac{1}{B(t)} \sum_{k=1}^m h_k^S(t)\,\mathrm{d}S_k(t) - \frac{V_{S'}^h(t)}{B(t)}\,\mathrm{d}B(t) \\
&= \frac{1}{B(t)} \sum_{k=1}^m h_k^S(t)\,\mathrm{d}S_k(t) - \frac{1}{B(t)}\left(V_{S'}^h(t) - h^0(t)\right)\,\mathrm{d}B(t) \\
&= \frac{1}{B(t)} \sum_{k=1}^m h_k^S(t)\,\mathrm{d}S_k(t) - \sum_{k=1}^m h_k^S(t) S_k'(t) r(t)\,\mathrm{d}t \\
&= \sum_{k=1}^m h_k^S(t)\,\mathrm{d}S_k'(t),
\end{aligned}$$

so h is self-financing under the normalized market. It remains to show that h is an arbitrage opportunity under the normalized market. Since h is an arbitrage for the original market, $V_S^h(0) = 0$, $P(V_S^h(T) \geq 0) = 1$ and $P(V_S^h(T) > 0) > 0$. Now note that since $B(t) > 0$ for all $t \geq 0$, we obtain $V_{S'}^h(0) = 0$, $P(V_{S'}^h(T) \geq 0) = 1$

and $P(V^h_{S'}(T) > 0) > 0$. We conclude that h is an arbitrage opportunity under the normalized market. Since this market was assumed arbitrage free, we have obtained a contradiction and must conclude that the original market is arbitrage free. $\square$

Next, we prove our main results for a market to be free of arbitrage. The theorem below is the primary criterion, with its corollaries giving criteria for special cases which are easier to check in practice. We say that a proces M is a local martingale on $[0, T]$ if there exists a localising sequence τ_n such that M^{τ_n} is a martingale on $[0, T]$ for each n.

Theorem 5.1.5. *Let $T > 0$. If there exists a probability measure Q equivalent to P such that each instrument of the normalized market is a local martingale on $[0, T]$ under Q, then $\mathcal{M}$ is arbitrage-free at time T.*

Proof. By Lemma 5.1.4, it will suffice to show that the normalized market is arbitrage free at time T. Therefore, assume that h is an arbitrage opportunity at time T under the normalized market with value process $V^h_{S'}$, we will aim to obtain a contradiction. Let Q be the probability measure equivalent to P which makes S'_k a local martingale on $[0, T]$ for any $k \le m$. By definition of an arbitrage opportunity, we know

$$\begin{aligned} V^h_{S'}(0) &= 0, \\ P(V^h_{S'}(T) \ge 0) &= 1, \\ P(V^h_{S'}(T) > 0) &> 0. \end{aligned}$$

Since P and Q are equivalent, we also have $Q(V^h_{S'}(T) \ge 0) = 1$ and $Q(V^h_{S'}(T) > 0) > 0$. In particular, $E^Q V^h_{S'}(0) = 0$ and $E^Q V^h_{S'}(T) > 0$. We will argue that h cannot be an arbitrage opportunity by showing that $V^h_{S'}$ is a supermartingale on $[0, T]$ under Q, contradicting that $E^Q V^h_{S'}(0) < E^Q V^h_{S'}(T)$.

Under the normalized market, the price of the risk-free asset is constant, so we find $\mathrm{d}V^h_{S'}(t) = \sum_{k=1}^m h^S_k(t)\,\mathrm{d}S'_k(t)$. Now, since P and Q are equivalent, S'_k is a standard process under Q according to Theorem 3.8.4. Because h is locally bounded, h is integrable with respect to S'_k both under P and Q, and Lemma 3.8.4 then yields that the integrals are the same whether under P or Q.

Under Q, each S'_k is a continuous local martingale on $[0, T]$. Therefore, the Q stochastic integral process $\sum_{k=1}^n \int_0^t h^S_k(s)\,\mathrm{d}S'_k(s)$ is a continuous local martingale on $[0, T]$ under Q. Under P, the stochastic integral process $\sum_{k=1}^n \int_0^t h^S_k(s)\,\mathrm{d}S'_k(s)$ is equal to $V^h_{S'}(t)$.

Since the integrals agree, as noted above, we conclude that $V^h_{S'}$ is a continuous local martingale on $[0,T]$ under Q. Because h is admissible, $V^h_{S'}$ is nonnegative. Thus, under Q, $V^h_{S'}$ is a nonnegative continuous local martingale on $[0,T]$. Using the same argument as in the proof of Lemma 3.7.2, V^h is then a Q supermartingale on $[0,T]$. In particular, $E^Q V^h_{S'}(t)$ is decreasing on $[0,T]$, in contradiction with our earlier conclusion that $E^Q V^h_{S'}(0) < E^Q V^h_{S'}(T)$. Therefore, there can exist no arbitrage opportunities in the normalized market. The result follows. □

Comment 5.1.6 Note that it was our assumption that h is locally bounded which ensured that the integral of h both under P and Q was well-defined. This is the reason why we in our definitions have chosen only to consider locally bounded portfolio strategies. The equivalent probability measure in the theorem making each instrument of the normalized market a local martingale on $[0,T]$ is called an equivalent local martingale measure for time T, or a T-EMM. ∘

The next two corollaries provide sufficient criteria for being arbitrage-free in the speciale case of a linear financial market, which we now define.

Definition 5.1.7. *A financial market is said to be linear if each S_k has dynamics*

$$\mathrm{d}S_k(t) = \mu_k(t)S_k(t)\,\mathrm{d}t + \sum_{i=1}^n \sigma_{ki}(t)S_k(t)\,\mathrm{d}W^i(t)$$

where $\sigma_{ki} \in \mathfrak{L}^2(W)$, $\mu_k \in \mathfrak{L}^1(t)$, and $S_k(0)$ is some positive constant. Here, μ is known as the drift vector process and σ is the volatility matrix process.

By Lemma 3.7.1, the price processes in a linear market has the explicit form

$$S_k(t) = S_k(0)\exp\left(\int_0^t \mu_k(s)\,\mathrm{d}s + \sum_{i=1}^n \int_0^t \sigma_{ki}(s)\,\mathrm{d}W^i_s - \frac{1}{2}\sum_{i=1}^n \int_0^t \sigma^2_{ki}(s)\,\mathrm{d}s\right),$$

unique up to indistinguishability. In particular, S is nonnegative, so it actually defines a financial market in the sense introduced earlier.

Corollary 5.1.8. *Assume that the market is linear and that there exists a process λ in $\mathfrak{L}^2(W)^n$ such that $\mu_k(t) - r(t) = \sum_{i=1}^n \sigma_{ki}(t)\lambda_i(t)$ for any $k \le m$, $\lambda \otimes P$ almost surely. Assume further that with $M_t = -\sum_{i=1}^n \int_0^t \lambda_i(s)\,\mathrm{d}W^i_s$, $\mathcal{E}(M)$ is a martingale. Then the market is free of arbitrage, and there is a T-EMM Q such that the Q-dynamics on $[0,T]$ are given by $\mathrm{d}S_k(t) = r(t)S_k(t)\,\mathrm{d}t + \sum_{i=1}^n \sigma_{ki}(t)S_k(t)\,\mathrm{d}\overline{W}^i(t)$, where $\overline{W}$ is a $\mathcal{F}_t$ Brownian motion under Q.*

Proof. We will use Theorem 5.1.5. Let $T > 0$ be given. Since $\mathcal{E}(M)$ is a martingale and $\mathcal{E}(M^T) = \mathcal{E}(M)^T$, we obtain $E\mathcal{E}(M^T)_\infty = E\mathcal{E}(M)_T = 1$. Recalling that $\mathcal{F} = \mathcal{F}_\infty$, we can then define the measure Q on $\mathcal{F}$ by $Q(A) = E^P 1_A \mathcal{E}(M^T)_\infty$. Since $\mathcal{E}(M^T)_\infty$ is almost surely positive and has unit mean, P and Q are equivalent. Now, with $Y_t^i = -\lambda_i(t)1_{[0,T]}(t)$, we obtain $M^T = Y \cdot W$. Therefore, by the Girsanov Theorem of Theorem 3.8.3, we can define $\overline{W}$ by $\overline{W}_t^k = W_t^k + \int_0^t \lambda_k(s)1_{[0,T]}(s)\,\mathrm{d}s$ and obtain that $\overline{W}$ is an $\mathcal{F}_t$ Brownian motion under Q.

We claim that under Q, S'_k is a local martingale on $[0,T]$. To this end, we note that, using Lemma 5.1.3,

$$\begin{aligned}
&S'_k(t) - S'_k(0) \\
=\; & \int_0^t (\mu_k(s) - r(s))S'_k(s)\,\mathrm{d}s + \sum_{i=1}^n \int_0^t \sigma_{ki}(s)S'_k(s)\,\mathrm{d}W_s^i \\
=\; & \int_0^t \left((\mu_k(s) - r(s) - \sum_{i=1}^n \sigma_{ki}(s)\lambda_i(s)1_{[0,T]}(s) \right) S'_k(s)\,\mathrm{d}s + \sum_{i=1}^n \int_0^t \sigma_{ki}(s)S'_k(s)\,\mathrm{d}\overline{W}_s^i \\
=\; & \int_0^t (\mu_k(s) - r(s))1_{(T,\infty)}(s)S'_k(s)\,\mathrm{d}s + \sum_{i=1}^n \int_0^t \sigma_{ki}(s)S'_k(s)\,\mathrm{d}\overline{W}_s^i,
\end{aligned}$$

where we have used that $\sum_{i=1}^n \sigma_{ki}(s)\lambda_i(s)1_{[0,T]}(s) = (\mu_k(s) - r(s))1_{[0,T]}(s)$. Letting τ_n be a determining sequence making the latter integral above into a martingale, it is then clear that $(S'_k)^{\tau_n}$ is a martingale on $[0,T]$ under Q. Thus, S'_k is a local martingale on $[0,T]$ under Q, so Q is a T-EMM. By Theorem 5.1.5, the market is free of arbitrage. Using the proof technique of Lemma 5.1.3, we find that the Q-dynamics of S_k on $[0,T]$ are

$$\begin{aligned}
\mathrm{d}S_k(t) &= B(t)\,\mathrm{d}S'_k(t) + S_k(t)r(t)\,\mathrm{d}t \\
&= \sum_{i=1}^n \sigma_{ki}(t)S_k(t)\,\mathrm{d}\overline{W}_k^i + S_k(t)r(t)\,\mathrm{d}t,
\end{aligned}$$

as desired. □

Comment 5.1.9 In the above, our assumption that $\mathcal{F} = \mathcal{F}_\infty$ became important. Had we not known that $\mathcal{F} = \mathcal{F}_\infty$, we would only have obtained a measure on $\mathcal{F}_\infty$ equivalent to the restriction of P to $\mathcal{F}_\infty$, and not a measure on all of $\mathcal{F}$.

The equation $\mu_k(t) - r(t) = \sum_{i=1}^n \sigma_{ki}(t)\lambda_i(t)$ is called the market price of risk equations, and any solution λ is known as a market price of risk specification for the market. In the one-dimensional case, we have a unique solution $\lambda_t = \frac{\mu_t - r_t}{\sigma_t}$, showing that λ can be interpreted as the excess return per volatility. Therefore, λ can in an intuitive sense be said to measure the price of uncertainty, or risk, in the market. $\circ$

Corollary 5.1.10. *Assume that the market is linear and only has one asset. Let μ denote the drift and let σ denote the volatility. Assume that the drift and short rate are constant, and that σ is positive $\lambda \otimes P$ almost surely. If there is $k \leq n$ such that $E\exp(\frac{(\mu-r)^2}{2}\int_0^t \frac{1}{\sigma_k(s)^2}\,\mathrm{d}s)$ is finite for any $t \geq 0$, the market is free of arbitrage, and there is a T-EMM such that the Q-dynamics on $[0,T]$ are given on the form $\mathrm{d}S_k(t) = r(t)S_k(t)\,\mathrm{d}t + \sum_{i=1}^n \sigma_{ki}(t)S_k(t)\,\mathrm{d}\overline{W}^i(t)$, where $\overline{W}$ is a $\mathcal{F}_t$ Brownian motion under Q.*

Proof. We want to solve the market price of risk equations of Corollary 5.1.8. Since we only have one asset, they reduce to

$$\mu - r = \sum_{i=1}^n \sigma_i(t)\lambda_i(t).$$

Let k be such that $E\exp(\frac{(\mu-r)^2}{2}\int_0^t \frac{1}{\sigma_k(s)^2}\,\mathrm{d}s)$ is finite for any $t \geq 0$. To simplify, we define $\lambda_i = 0$ for $i \neq k$. We are then left with the equation $\mu - r = \sigma_k(t)\lambda_k(t)$. Let A denote the subset of $[0,\infty) \times \Omega$ where σ is positive. Since σ is progressive, $A \in \Sigma_\pi$. Defining $\lambda_k(t,\omega) = \frac{\mu - r}{\sigma_k(t,\omega)}$ whenever $(t,\omega) \in A$ and zero otherwise, λ satisfies the equations in the statement of Corollary 5.1.8. By that same corollary, in order for the market to be arbitrage free, it will then suffice to show that with $M_t = -\sum_{i=1}^n \int_0^t \lambda_i(s)\,\mathrm{d}W_s^i$, $\mathcal{E}(M)$ is a martingale, and to do so it will suffice to show that $\mathcal{E}(M^t)$ is a uniformly integrable martingale for any $t \geq 0$. By the Novikov criterion of Theorem 3.8.10, this is the case if $E\exp(\frac{1}{2}[M^t]_\infty)$ is finite for any $t \geq 0$. But we have

$$\begin{aligned} E\exp\left(\frac{1}{2}[M^t]_\infty\right) &= E\exp\left(\frac{1}{2}[M]_t\right) \\ &= E\exp\left(\frac{1}{2}\int_0^t \lambda_k(s)^2\,\mathrm{d}s\right) \\ &= E\exp\left(\frac{(\mu-r)^2}{2}\int_0^t \frac{1}{\sigma_k(s)^2}\,\mathrm{d}s\right), \end{aligned}$$

so that the Novikov criterion is satisfied follows directly from our assumptions. The existence of a Q measure with the given dynamics then follows from Corollary 5.1.8. $\square$

We are now done with exploring sufficient criteria for no-arbitrage. Next, we consider pricing of claims in markets without arbitrage. For $T > 0$, we define a T-claim as a

$\mathcal{F}_T$ measurable, integrable variable X. The intuition behind this is that the owner of the claim receives the amount X at time T. Our goal is to argue for a reasonable price of this contract at $t \leq T$. We will not be able to do this in full generality, but will only consider markets satisfying the criterion of Theorem 5.1.5. As discussed in the notes, this actually covers all arbitrage-free markets, but we will not be able to show this fact.

Theorem 5.1.11. *Let $T > 0$ and consider a market with a T-EMM Q. Let X be any T-claim. Assume that $X \geq 0$ and define*

$$S_{m+1}(t) = E^Q\left(\frac{B(t)}{B(T)}X \,\middle|\, \mathcal{F}_t\right) 1_{[0,T]}(t) + \frac{B(t)}{B(T)}X 1_{(T,\infty)}(t).$$

The market with instruments $(S_1, \ldots, S_{m+1})$ is free of arbitrage at time T.

Proof. We first argue that $(S_1, \ldots, S_{m+1})$ is a well-defined financial market. Under Q, S'_{m+1} is obviously a martingale. Since we are working under the augmented Brownian filtration, it has a continuous version. Since X is nonnegative, $S'_{m+1}(t)$ is almost surely nonnegative for ant $t \geq 0$, and therefore there is a version of S'_{m+1} which is nonnegative in addition to being continuous. Thus, S'_{m+1} is a nonnegative standard process under Q. Therefore, S_{m+1} is also a nonnegative standard process under Q, and so it is a standard process under P. Finally, we conclude that $(S_1, \ldots, S_{m+1})$ defines a financial market.

Under Q, each of the instruments $S'_1, \ldots, S'_m$ are local martingales on $[0, T]$. Since S'_{m+1} trivially is a martingale on $[0, T]$ under Q, Theorem 5.1.5 yields that the market with instruments $(S_1, \ldots, S_{m+1})$ is free of arbitrage at time T. □

Comment 5.1.12 The somewhat cumbersome form of S_{m+1}, splitting up into a part on $[0, T]$ and a part on (T, ∞), is an artefact of considering markets with infinite horizon. ○

What Theorem 5.1.11 shows is that for any nonnegative claim, using $E^Q(\frac{B(t)}{B(T)}X|\mathcal{F}_t)$ as the price for the claim X for $t \leq T$ is consistent with no arbitrage in the market where the claim is traded. And since $E^Q(\frac{B(T)}{B(T)}X|\mathcal{F}_T) = X$, buying the asset with price process $E^Q(\frac{B(t)}{B(T)}X|\mathcal{F}_t)$ actually corresponds to buying the claim, in the sense that at the time of expiry T, one owns an asset with value X. Assuming that prices are linear, we conclude that a reasonable price for any T-claim for $t \leq T$ is $E^Q(\frac{B(t)}{B(T)}X|\mathcal{F}_t)$, where Q is a T-EMM.

We are now done with our discussion of general financial markets. Our results are the following. We have in Theorem 5.1.5 obtained general criteria for a market to be free of arbitrage, and we have obtained specialised criteria for the case of linear markets in the corollaries. Afterwards, we have shortly argued, based on the result of Theorem 5.1.11, what a reasonable price of a contingent claim should be modeled as. Note that lack of uniqueness of T-EMMs can imply that the price of a T-claim is not uniquely determined. This will not bother us particularly, however.

We will now proceed to discuss the practical matter of calculating the prices and sensitivities associated with contingent claims. We will begin by a very short review of some basic notions from the theory of SDEs and afterwards proceeed to go through some methods for Monte Carlo evaulation of prices and sensitivities. Having done so, we will apply these techniques to the Black-Scholes and Heston models.

5.2 Stochastic Differential Equations

We will review some fundamental definitions regarding SDEs of the form

$$\mathrm{d}X_t^k = \mu_k(X_t)\,\mathrm{d}t + \sum_{i=1}^n \sigma_{ki}(X_t)\,\mathrm{d}W_t^i,$$

where X_0 is constant, $k \le m$, W is a n-dimensional Brownian motion and the mappings $\mu : \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : \mathbb{R}^m \to \mathbb{R}^{m\times n}$ are measurable. We also consider a deterministic initial condition $x \in \mathbb{R}^m$. We write the SDE in the shorthand

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,$$

the underlying idea being that the m-vector $\mu(X_t)$ is being multiplied by the scalar $\mathrm{d}t$, yielding a new m-vector, and the $m \times n$ matrix $\sigma(X_t)$ is begin multiplied by the n-vector $\mathrm{d}W_t$, yielding a m-vector, so that we formally obtain

$$\begin{aligned}
\begin{pmatrix} \mathrm{d}X_t^1 \\ \vdots \\ \mathrm{d}X_t^m \end{pmatrix} &= \begin{pmatrix} \mu_1(X_t) \\ \vdots \\ \mu_m(X_t) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sigma_{11}(X_t) & \cdots & \sigma_{1n}(X_t) \\ \vdots & & \vdots \\ \sigma_{m1}(X_t) & \cdots & \sigma_{mn}(X_t) \end{pmatrix} \begin{pmatrix} \mathrm{d}W_t^1 \\ \vdots \\ \mathrm{d}W_t^n \end{pmatrix} \\
&= \begin{pmatrix} \mu_1(X_t)\,\mathrm{d}t + \sum_{i=1}^n \sigma_{1i}(X_t)\,\mathrm{d}W_t^i \\ \vdots \\ \mu_m(X_t)\,\mathrm{d}t + \sum_{i=1}^n \sigma_{mi}(X_t)\,\mathrm{d}W_t^i \end{pmatrix},
\end{aligned}$$

in accordance with our first definition of the SDE.

Analogously with the conventions for linear markets in Section 5.1, μ is called the drift coefficient and σ is called the the volatility coefficient. We define a set-up as a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ satisfying the usual conditions and being endowed with a $\mathcal{F}_t$ Brownian motion W. We say that the SDE has a weak solution if there for any set-up exists a m-dimensional standard process X satisfying the SDE. We say that the SDE satisfies uniqueness in law if any two solutions on any two set-ups have the same distribution. We say that the SDE satisfies pathwise uniqueness if any two solutions on the same set-up are indistinguishable.

Furthermore, we say that the SDE has a strong solution if there exists a mapping $F : \mathbb{R}^m \times C([0,\infty), \mathbb{R}^n)$ such that for any set-up with Brownian motion W and any $x \in \mathbb{R}^m$, the process $X^x = F(x, W)$ solves the SDE with initial condition x. We call F a strong solution of the SDE. If the SDE satisfies pathwise uniqueness, $F(x, \cdot)$ is almost surely unique when $C([0,\infty) \times \mathbb{R}^n)$ is endowed with the Wiener measure. When the SDE has a strong solution F, we can define the flow of the solution as the mapping $\phi : \mathbb{R}^m \times C([0,\infty), \mathbb{R}^n) \times [0,\infty) \to \mathbb{R}^m$ given by $\phi(x, w, t) = F(x, w)_t$. We say that the flow is differentiable if the mapping $\phi(\cdot, w, t)$ is differentiable for all w and t.

This concludes our review of fundamental notions for SDEs.

5.3 Monte Carlo evaluation of expectations

As we saw in the first section of this chapter, an arbitrage-free price of a contingent T-claim is given as a discounted expected value under the Q-measure. The problem of evaluating prices can therefore be seen as a special case of the problem of evaluating expectations. As we shall see, the same is the case for the problem of evaluating sensitivities of prices to changes in the model parameters. For this reason, we will now review some basic Monte Carlo methods for evaluating expectations. Our basic resource is Glasserman (2003), in particular Chapter 4 of that work.

Throughout the section, consider an integrable stochastic variable X. Our goal is to find the expectation of X. The basis for the Monte Carlo methods of evalutating expectations is the strong law of large numbers, stating that if (X_n) is a sequence of independent variables with the same distribution as X, then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} EX,$$

This means that if we can draw a sequence of independent numbers sampled from the distribution of X and take their average, the result should approach the expectation of X as the number of samples tend to infinity. Furthermore, we can measure the speed of convergence by the variance,

$$V\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} VX_k = \frac{1}{n}VX.$$

In reality, we can of course not truly draw independent random numbers, we are only able to generate "pseudorandom" sequences, numbers which behave randomly in some suitable sense. We will not discuss the implications of this observation, mostly because the pseudorandom number generators available today are of sufficiently high quality to ensure that the topic is of little practical significance. For more information on the fundamentals of generating random numbers, see Chapter 2 of Glasserman (2003).

The methods which we will concentrate upon are how the basic application of the law of large numbers can be modified to speed up the convergence to the expectation, in the sense of reducing the variance calculated above. The methods we will consider are, ordered by increased potential efficiency improvement:

1. Antithetic sampling.
2. Control variates.
3. Stratified sampling.
4. Importance sampling.

Antithetic sampling. The idea behind antithetic sampling is to even out deviations from the sample mean in the i.i.d. sequence. As before, we let (X_n) be an i.i.d. sequence with the same distribution as X. Let (X'_n) be another sequence, and assume that the sequence (X_n, X'_n) is also i.i.d. Assume furthermore that $EX'_n = EX_n$. Thus, the pairs are independent, and each marginal in any pair has the mean EX, it is however possible that X_n and X'_n are dependent, in fact such dependence is the very idea of antithetic sampling. We then find

$$\frac{1}{2n}\left(\sum_{k=1}^n X_k + \sum_{k=1}^n X'_k\right) = \frac{1}{2}\left(\frac{1}{n}\sum_{k=1}^n X_k + \frac{1}{n}\sum_{k=1}^n X'_k\right) \xrightarrow{\text{a.s.}} EX.$$

Thus, sampling from each sequence and averaging yields a consistent estimator of the mean. If the samples from X'_k on average are larger than the mean when the samples from X_k are smaller than the mean and vice versa, we could hope that this would even out deviations from the mean and improve convergence. To make this concrete, let X' be such that (X, X') has the common distribution of (X_n, X'_n). We then obtain

$$\begin{aligned} V\left(\frac{1}{2n}\left(\sum_{k=1}^n X_k + \sum_{k=1}^n X'_k\right)\right) &= \frac{1}{n^2}\sum_{k=1}^n V\left(\frac{X_k + X'_k}{2}\right) \\ &= \frac{1}{n^2}\sum_{k=1}^n \frac{1}{4}\left(2VX_k + 2\mathrm{Cov}(X_k, X'_k)\right) \\ &= \frac{1}{2n}VX + \frac{1}{2n}\mathrm{Cov}(X, X'). \end{aligned}$$

The first term is the usual Monte Carlo estimator variance. Thus, we see that antithetic sampling can reduce variance if $\mathrm{Cov}(X, X')$ is negative, precisely corresponding to the values of X'_k on average being larger than the mean when the values of X_k are smaller than the mean and vice versa.

As a simple example of antithetic sampling, let us assume that we wanted to calculate the second moment of the unit uniform distribution by antithetic sampling. We would need to identify a two-dimensional distribution (X, X') such that X and X' both have the means of the the second moment of the uniform distribution and are negatively correlated. A simple possibility would be to let U be uniformly distributed and define $X = U^2$ and $X' = (1-U)^2$. X and X' are then both distributed as the squares of uniform distributions. We would intuitively expect X and X' to be negatively correlated, since large values of X corresponds to large values of U, corresponding to small values of X'. In this case, we can check that this is actually true, since

$$\mathrm{Cov}(X, X') = \mathrm{Cov}(U^2, (1-U)^2) = -\frac{1}{12}.$$

In practice, we would therefore consider a sequence (U_n) of uniform distributions and construct the expectation estimator

$$\frac{1}{2n}\left(\sum_{k=1}^n U_k^2 + \sum_{k=1}^n (1-U_k)^2\right),$$

which would be a consistent estimator of the mean, and its variance would be, since $VU^2 = \frac{4}{45}$, $\frac{1}{2n}\left(\frac{4}{45} - \frac{1}{12}\right)$. The is an improvement over the ordinary Monte Carlo estimator of a factor $\frac{4}{45}(\frac{4}{45} - \frac{1}{12})^{-1} = 16$. This means that for any n, the antithetic estimator would have a standard deviation four times smaller than the ordinary estimator. Usually, the efficiency boost of antithetic sampling is not much larger than this, actually it is usually somewhat more modest. However, whenever we are considering random variables based on distributions with some kind of symmetry, such as the uniform distribution or the normal distribution, antithetic sampling is easy to implement and almost always provides an improvement. There is therefore rarely any reason not to use antithetic sampling in such situations.

Control variates. Like antithetic sampling, the method of control variates operates by attempting to balance samples which are below or above the true expectation by correlation considerations. Instead of adding an extra sequence of variables with the same mean as the original sequence, however, the method of control variates adds an extra sequence of variables with some known mean which is not necessarily the same as the mean of X.

Therefore, let X' be some variable with known mean, possibly correlated with X. Suppose that the sequence (X_n, X'_n) is i.i.d. with the common distribution being the same as that of (X, X'). With $b \in \mathbb{R}$, We can then consider the estimator

$$\frac{1}{n}\left(\sum_{k=1}^{n} X_k - b\sum_{k=1}^{n}(X'_k - EX')\right).$$

The law of large numbers applies to show that this yields a consistent estimator. The idea behind the estimator is that if X_k and X'_k are correlated, we can choose b such that the deviations from the mean of X_k are evened out by the deviations from the mean of X'_k. In the case of positive correlation, this would mean choosing a positive b, since large values of X_k then would be offset by large values of X'_k and vice versa. Likewise, in the case of negative correlation, we would expect that choosing a negative b would yield the best results. We call X'_k the control variate.

We will now analyze the variance of the control variate estimator. We find

$$\begin{aligned} V\left(\frac{1}{n}\left(\sum_{k=1}^{n} X_k - b\sum_{k=1}^{n}(X'_k - EX')\right)\right) &= \frac{1}{n^2}\sum_{k=1}^{n} V(X_k - b(X'_k - EX')) \\ &= \frac{1}{n}(VX_k + b^2 V(X'_k) - 2b\mathrm{Cov}(X_k, X'_k)) \\ &= \frac{1}{n}VX + \frac{1}{n}\left(b^2 VX' - 2b\mathrm{Cov}(X, X')\right). \end{aligned}$$

The first term is the variance of the ordinary Monte Carlo estimator. If the control variate estimator is to be useful, the second term must be negative, corresponding to $b^2VX' - 2b\mathrm{Cov}(X,X') < 0$. To find out when we can make sure this is the case, we first identify the optimal value of b. This is the value minimizing $b^2VX' - 2b\mathrm{Cov}(X,X')$ over b. The formula for the minimum of a quadratic polynomial yields the optimal value $b^* = \frac{\mathrm{Cov}(X,X')}{VX'}$. The corresponding variance change is

$$(b^*)^2VX' - 2b^*\mathrm{Cov}(X,X') \quad = \quad -\frac{\mathrm{Cov}(X,X')^2}{VX'}.$$

Since this is always negative, we conclude that, disregarding considerations of computing time, control variates can only improve convergence, and there is an improvement whenever X and X' are correlated. The factor of improvement of the control variate estimator to the ordinary estimator is

$$\frac{VX}{VX - \frac{\mathrm{Cov}(X,X')^2}{VX'}} = \frac{1}{1-\mathrm{Corr}(X,X')^2}.$$

This means that if, for example, the correlation is $\frac{1}{2}$, the variance of the control variate estimator is a factor of $\frac{4}{3}$ larger than the variance of the ordinary Monte Carlo estimator. If the correlation is $\frac{3}{4}$, it is a factor of $\frac{16}{7}$ larger and so on. The improvement grows drastically as the correlation tends to one, with the limit corresponding to the degenerated case where X' has the distribution of X, which is of course absurd since EX then would be known, making the estimation unnecessary.

As an example, we can again consider the problem of numerically evaluating the second moment of the uniform distribution. As we have seen, to find an effective control variate, we need to identify a variable which is highly correlated with the variables we are using to estimate the second moment, and whose mean we know. Thus, let U be uniformly distributed and let $X = U^2$. Assuming that we know that the mean of the uniform distribution is $\frac{1}{2}$, we can use the control variate $X' = U$. Obviously, X and X' should be quite highly correlated. Note that unlike the method of antithetic sampling, it does not matter whether the correlation is positive or negative, this is taken care of in the choice of coefficient. We find that the correlation is

$$\mathrm{Corr}(X,X') = \frac{\mathrm{Cov}(U^2,U)}{\sqrt{VU^2}\sqrt{VU}} = \sqrt{\frac{45}{48}},$$

which is pretty close to one, leading to an reduction in variance of the estimator to the ordinary estimator of a factor $\frac{1}{1-\frac{45}{48}} = 16$. The optimal coefficient is

$$\frac{\mathrm{Cov}(X,X')}{VX'} = \frac{\mathrm{Cov}(U^2,U)}{VU} = 1.$$

The corresponding control variate estimator is, letting (U_n) be a sequence of i.i.d. uniforms, $\frac{1}{n}\left(\sum_{k=1}^n U_k^2 - \sum_{k=1}^n \left(U_k - \frac{1}{2}\right)\right)$. As expected, very large values of U_k correspond to shooting over the true mean, and this is offset by the control variate. Since the control variate method in this case does not entail much extra calculation, we can reasonably say that the associated computational time is the same as that of the ordinary Monte Carlo estimator. Therefore, we appreciate that the control variate method provides an estimator with one fourth of the standard deviation of the ordinary estimator with the same amount of computing time.

In practice, the optimal coefficient cannot be calculated explicitly. Instead, one can apply the Monte Carlo simulations to estimate it, using

$$\hat{b}^* = \frac{\sum_{k=1}^n (X_k - \overline{X})(X_k' - \overline{X}')}{\sum_{k=1}^n (X_k' - \overline{X}')^2}.$$

The extra computational effort required from this is usually more than offset by the reduction in standard deviation.

Stratified sampling. Stratified sampling is a sampling method where the sample space of the distribution being sampled is partitioned and the number of samples from each part of the partition is being controlled. In this way, the deviation from the mean cannot go completely awry.

We begin by introducing a stratification variable, Y. Let $A_1, \ldots, A_m$ be disjoint subsets of $\mathbb{R}$ such that $P(Y \in \cup_{i=1}^m A_i) = 1$ and $P(Y \in A_i) > 0$ for all i. We call these sets the strata. The idea is that instead of sampling from the distribution of X to obtain the mean of X, we will sample from the distribution of X given $Y \in A_i$.

Let X_i have the distribution of X given $Y \in A_i$. We then find, putting $p_i = P(Y \in A_i)$,

$$EX = \sum_{i=1}^m EX1_{(Y \in A_i)} = \sum_{i=1}^m p_i E\left(\frac{X1_{(Y \in A_i)}}{p_i}\right) = \sum_{i=1}^m p_i EX_i,$$

so if we can sample from each of the conditional distributions, we can make Monte Carlo estimates of each term in the sum above and thereby obtain an estimate for EX. To make this concrete, let for each i (X_{ik}) be an i.i.d. sequence with the common distribution being the same as that of X_i. All these variables are assumed independent. Fix some n. We are free to choose how many samples to use from each strata. We will restrict ourselves to what is known as proportional allocation, meaning that we choose to draw a fraction p_i of samples from the i'th stratum. Assume for simplicity that np_i is an integer, we can then put $n_i = np_i$, obtain $n = \sum_{i=1}^m n_i$ and define the stratified Monte Carlo estimator

$$\frac{1}{n}\sum_{i=1}^{m}\sum_{k=1}^{n_i} X_{ik}.$$

By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{m}\sum_{k=1}^{n_i} X_{ik} = \sum_{i=1}^{m}\frac{p_i}{n_i}\sum_{k=1}^{n_i} X_{ik} \xrightarrow{\text{a.s.}} \sum_{i=1}^{m} p_i E X_i = EX$$

so the estimator is consistent. Its variance is

$$V\left(\frac{1}{n}\sum_{i=1}^{m}\sum_{k=1}^{n_i} X_{ik}\right) = \frac{1}{n^2}\sum_{i=1}^{m}\sum_{k=1}^{n_i} VX_{ik} = \frac{1}{n}\sum_{i=1}^{m} p_i V X_i.$$

To compare this with the variance of the ordinary Monte Carlo estimator, define U by letting $U = i$ if $Y \in A_i$. Then, the distribution of Y given $U = i$ is the distribution of X_i, and for the distribution of U we find $P(U = i) = P(Y \in A_i) = p_i$. All in all, we may conclude

$$\frac{1}{n}\sum_{i=1}^{m} p_i V X_i = \frac{1}{n} EV(X|U) = \frac{1}{n} VX - \frac{1}{n} VE(X|U),$$

where the first term is the variance of the ordinary Monte Carlo estimator. Since the second term is a negative scalar times a variance, it is always negative. Therefore, disregarding questions of computation time, the convergence of the stratified estimator is always at least as fast as that of the ordinary estimator.

Let us review the components of the stratified sampling technique. We start out with X, the distribution whose mean we wish to identify. We then consider a stratification variable Y and strata $A_1, \ldots, A_m$. It is the simultaneous distribution of (X, Y) combined with the strata which determine how our stratified estimator will function. We define i.i.d. sequences X_{ik} such that X_{ik} has the distribution of X given $Y \in A_i$. The estimator $\frac{1}{n}\sum_{i=1}^m \sum_{k=1}^{n_i} X_{ik}$ is then consistent and has superior variance compared to the ordinary Monte Carlo estimator.

In practice, the critical part of this procedure is to certify that it is possible and computationally tractable to sample from the distribution of X given $Y \in A_i$. This is often accomplished by ensuring that $X = t(Y)$ for some transformation t. In that case, the distribution of X given $Y \in A_i$ is just the distribution of Y given $Y \in A_i$ transformed with t. If Y is a distribution such that it is easy to sample from Y

given $Y \in A_i$, we have obtained a simple procedure for sampling from the conditional distribution. This is the case if, say, Y is a distribution with an easily computable quantile function and A_i is an interval.

To see an example of how stratified sampling can be used, we will use it to numerically calculate the second moment of the uniform distribution. Let U be uniformly distributed, we then wish to identify the mean of U^2. Fix $m \in \mathbb{N}$, we will use U as our stratification variable with the strata $A_i = (\frac{i-1}{m}, \frac{i}{m}]$. We then find $p_i = \frac{1}{m}$ and $n_i = \frac{n}{m}$. We need to identify the distribution of U^2 given $U \in A_i$. We know that the distribution of U given $U \in A_i$ is the uniform distribution on A_i. Therefore, U^2 given $U \in A_i$ has the same distribution as $(\frac{i-1}{m} + \frac{U}{m})^2$. Letting U_{ik} be unit uniformly distributed i.i.d. sequences, we can write our stratified estimator as

$$\frac{1}{n}\sum_{i=1}^{m}\sum_{k=1}^{\frac{n}{m}}\left(\frac{i-1}{m}+\frac{U_{ik}}{m}\right)^2.$$

Its variance is

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{m}\frac{1}{m}V\left(\frac{i-1}{m}+\frac{U_i}{m}\right)^2 \\
= & \frac{1}{n}\sum_{i=1}^{m}\frac{1}{m^5}V(i-1+U_i)^2 \\
= & \frac{1}{n}\sum_{i=1}^{m}\frac{1}{m^5}V((i-1)^2+2(i-1)U_i+U_i^2) \\
= & \frac{1}{n}\sum_{i=1}^{m}\frac{1}{m^5}\left(4(i-1)^2VU_i+VU_i^2+4(i-1)\mathrm{Cov}(U_i,U_i^2)\right) \\
= & \frac{1}{n}\sum_{i=1}^{m}\frac{1}{m^5}\left(\frac{(i-1)^2}{3}+\frac{4}{45}+\frac{i-1}{3}\right) \\
= & \frac{1}{n}\frac{1}{m^5}\left(\frac{(m-1)m(2m-1)}{18}+\frac{4m}{45}+\frac{(m-1)(m-2)}{3}\right),
\end{aligned}$$

which is of order $O(m^{-2})$ when m tends to infinity. This basically shows that we in theory can obtain as small a variance as we desire by picking as many strata as possible. In reality, of course, it is insensible to pick more strata than observations. The moral of the example, however, is the correct one: by using a large number of strata, a very dramatic reduction of variance is sometimes possible. An example of how stratified sampling can improve results can be seen in Figure 5.1, where we compare sampling from squared uniforms with and without stratification. We see that the histogram on the right, using stratified sampling, is considerably closer to the true form of the distribution.

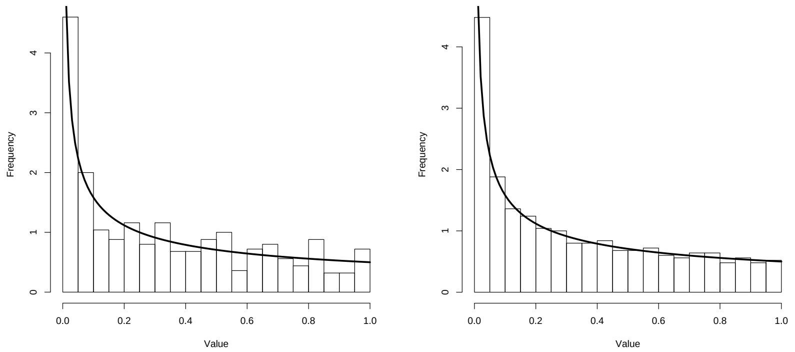

Figure 5.1: Comparison of samples from a squared uniform distribution with and without stratified sampling. On the left, 500 samples without stratification. On the right, 500 samples with 100 strata.

The ease of implementation and potentially large improvements make stratified sampling an very useful method for improving convergence.

Importance sampling. Importance sampling is arguably the method described here which has the largest potential impact. Its usefulness varies greatly from situation to situation, and if used carelessly, it can significantly reduce the effectivity of an estimator. On the other hand, in situations where no other method is applicable, it sometimes can be applied to obtain immense improvements.

The underlying idea of importance sampling is very simple. It is most naturally described in the context where $X = h(X')$ for some variable X' and some mapping h with sufficient integrability conditions. Let μ be the distribution of X'. Let Y' be some other variable with distribution ν and assume that $\mu \ll \nu$. We can then use the Radon-Nikodym theorem to write

$$Eh(X') = \int h(x)\,\mathrm{d}\mu(x) = \int h(x)\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\,\mathrm{d}\nu(x) = E\left(h(Y')\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Y')\right).$$

Letting Y'_n be a i.i.d. sequence with common distibution being the same as that of Y', we can then define the importance sampling estimator as

$$\frac{1}{n}\sum_{k=1}^{n} h(Y'_k)\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Y'_k).$$

The law of large numbers yields that this is a consistent estimator of $Eh(X')$. Its variance can be larger or smaller than that of the ordinary Monte Carlo estimator depending on the choice of ν. Results on the effectiveness of the importance sampling estimator as a function of the choice of ν can only be obtained in specialised situations. Furthermore, the objective of the importance sampling method will in fact often not be to reduce variance, but rather to change the shape of a distribution when it is somehow suboptimal for the purposes of Monte Carlo estimation. Is is difficult to describe quantitatively when this is the case, we will instead give a small example.

Let X' be the standard exponential distribution and consider the problem of estimating $P(X' > \alpha)$ for some large α. This means considering the transformation h given by $h(x) = 1_{(\alpha,\infty)}(x)$. Let $p = P(X > \alpha)$, the variance of $h(X')$ is $p(1-p)$, which is pretty small. In theory, then, the ordinary Monte Carlo estimator should yield relatively effective estimates. However, this is not actually the case.

To see why the ordinary estimator often will underestimate the true probability, we cannot rely only on moment considerations, we need to analyze the distribution of the estimator in more detail. To this end, note that $\frac{1}{n}\sum_{k=1}^n h(X'_k)$ is binomially distributed with length n, probability parameter p and scale $\frac{1}{n}$. A fundamental problem with this is that the binomial distribution is discrete, this immediately puts an upper limit to the accuracy of the estimator. Also, the estimator will often underestimate the true probability. To see how this is the case, simply note that the probability of the estimator being zero is $(1-p)^n$. When p is small compared to n, this can be very large. In the case $\alpha = 10$, for example, we obtain $p = e^{-10}$, $(1-p)^{10000} \approx 63\%$ and $(1-p)^{100000} \approx 6.3\%$. This means that with 10000 samples, our estimator will yield zero in 63% of our trials. Using ten times as many simulations yields zero only 6.3% percent of the time, but the distribution will still very often be far from the correct probability, which in this case is $e^{-10} \approx 0.0000453$.

We can use importance sampling to improve the efficiency of our estimator. We let the importance sampling measure ν be the exponential distribution with scale λ. We then have

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) = \frac{e^{-x}}{\frac{1}{\lambda}e^{-\frac{x}{\lambda}}} = \lambda \exp\left(-x\left(1-\frac{1}{\lambda}\right)\right),$$

and with (Y'_n) being an i.i.d. sequence with distribution ν, the importance sampling estimator is

$$\frac{1}{n}\sum_{k=1}^n h(Y'_k)\lambda \exp\left(-Y'_k\left(1-\frac{1}{\lambda}\right)\right).$$

In Figure 5.2, we have demonstrated the effect of using the importance sampling estimator. We use $\alpha = 10$ and compare the histograms of 1000 independent ordinary

Monte Carlo estimators with 1000 independent importance sampling estimators, letting $\lambda = 10$ and basing the estimators on 1000 samples.

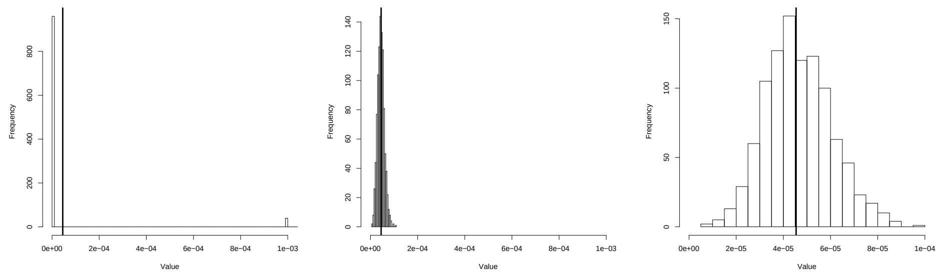

Figure 5.2: Comparison of estimation of exponential probabilities with and without importance sampling. From left to right, histogram of 1000 ordinary Monte Carlo estimators, histogram of 1000 importance sampling estimators and finally, close-up histogram of 1000 importance sampling estimators.

Obviously, the ordinary Monte Carlo estimator cannot get very close to the true value of p because of its discreteness. As expected, it hits zero most of the time, with a few samples hitting $\frac{1}{1000}$. On the other hand, the importance sampling estimator can take any value in $[0, nc)$ for some $c > 0$, removing any limits on the accuracy arising from discreteness. We see that even with only the 1000 samples which were completely useless in the ordinary Monte Carlo case, the importance sampling estimator hits the true probability very well most of the time. Furthermore, the computational overhead arising from the importance sampling technique is minimal. We conclude that with very little effort, we have improved the ordinary estimator by an extraordinary factor.

5.4 Methods for evaluating sensitivities

Next, we turn our attention to methods specifically designed to evalutate sensitivities such as the delta or gamma of a claim. We will cast our discussion in a general framework, which will then be specialised to financial cases in the following sections. We assume that we have a family of integrable variables $X(\theta)$ indexed by an open subset Θ of the real line. Our object is to understand how to estimate $\Delta_\theta = \frac{d}{d\theta}EX(\theta)$, assuming that the derivative exists. We will consider five methods, being:

1. Finite difference.
2. Pathwise differentiation.
3. Likelihood ratio.
4. Malliavin weights.
5. Localisation.

These methods in general transform the problem of estimating a derivative into estimating one or more expectations. These expectations can then be evaluated by the methods described in Section 5.3. In the following we always assume that the mapping $\theta \mapsto EX(\theta)$ is differentiable.

Finite difference. The finite difference method aims to calculate the sensitivity by approximating the derivative by a finite difference quotient and estimating this by Monte Carlo. Define $\alpha(\theta) = EX(\theta)$. Fixing θ and leting $h > 0$, our approximation is defined by

$$\hat{\Delta}_\theta = \frac{\alpha(\theta + h) - \alpha(\theta - h)}{2h}.$$

Introducing i.i.d. variables $(X_n(\theta+h), X_n(\theta-h))$ such that $X_n(\theta+h)$ and $X_n(\theta-h)$ has the distributions of $X(\theta+h)$ and $X(\theta-h)$, respectively, but possibly are correlated, we can define the finite difference estimator as

$$\frac{\overline{X}_n(\theta + h) - \overline{X}_n(\theta - h)}{2h}.$$

The law of large numbers yields $\frac{\overline{X}_n(\theta+h)-\overline{X}_n(\theta-h)}{2h} \xrightarrow{\text{a.s.}} \hat{\Delta}_\theta$, so the estimator is in general biased. To analyze the bias, note that if α is C^2, we can form second-order expansions of α at θ, yielding

$$\begin{aligned} \alpha(\theta + h) &= \alpha(\theta) + \alpha'(\theta)h + \frac{1}{2}\alpha''(\theta)h^2 + o(h^2) \\ \alpha(\theta - h) &= \alpha(\theta) - \alpha'(\theta)h + \frac{1}{2}\alpha''(\theta)h^2 + o(h^2), \end{aligned}$$

so that in this case, the bias has the form

$$\hat{\Delta}_\theta - \alpha'(\theta) = \frac{\alpha(\theta + h) - \alpha(\theta - h)}{2h} - \alpha'(\theta) = o(h).$$

If α is C^3, we can obtain the third-order expansions

$$\begin{aligned} \alpha(\theta + h) &= \alpha(\theta) + \alpha'(\theta)h + \frac{1}{2}\alpha''(\theta)h^2 + \frac{1}{6}\alpha'''(\theta)h^3 + o(h^3) \\ \alpha(\theta - h) &= \alpha(\theta) - \alpha'(\theta)h + \frac{1}{2}\alpha''(\theta)h^2 - \frac{1}{6}\alpha'''(\theta)h^3 + o(h^3), \end{aligned}$$

which yields the bias $\frac{1}{6}\alpha'''(\theta)h^2 + o(h^2)$, which is not only $o(h)$ but is $O(h^2)$.

Conditional on a level of smoothness of α, then, the finite difference estimator is at least asymptotically unbiased. Letting $(X'(\theta+h), X'(\theta-h))$ have the common distribution of $(X_n(\theta+h), X_n(\theta-h))$, the variance of the estimator is

$$\begin{aligned} V\left(\frac{\overline{X}_n(\theta+h) - \overline{X}_n(\theta-h)}{2h}\right) &= \frac{1}{4h^2n^2}V\left(\sum_{k=1}^{n} X_k(\theta+h) - X_k(\theta-h)\right) \\ &= \frac{1}{4h^2n}V(X'(\theta+h) - X'(\theta-h)). \end{aligned}$$

We see that the variance is highly dependent both on h and on the dependence structure between $X'(\theta+h)$ and $X'(\theta-h)$. Both of these are under our control, and we can therefore seek to choose them in a manner which reduces the mean square error as much as possible. This will provide a reasonable tradeoff between reducing bias and reducing variance. The context which we are working in at present is too general for us to be able to make any optimal choice of the dependence between $X'(\theta+h)$ and $X'(\theta-h)$, so we will content ourselves with making some assumptions about the asymptotic behaviour of the bias and variance and examining how best to choose h as a function of n from this.

We will assume that

$$\begin{aligned} \hat{\Delta}_\theta - \alpha'(\theta) &= O(h^2) \\ V(X'(\theta+h) - X'(\theta-h)) &= O(h^\beta), \end{aligned}$$

for some $\beta > 0$. This corresponds to situations often encountered in practice. We then define $h_n = cn^{-\gamma}$ and want to identify the rate of decrease γ yielding the asymptotically minimal mean square error. The mean square error is

$$\begin{aligned} \mathrm{MSE}(\hat{\Delta}_n) &= V(\hat{\Delta}_n) + (E\hat{\Delta}_n - \theta)^2 \\ &= \frac{1}{4h_n^2 n}V(X'(\theta+h_n) - X'(\theta-h_n)) + O(h_n^2)^2 \\ &= \frac{1}{4h_n^2 n}O(h_n^\beta) + O(h_n^4) \\ &= \frac{1}{4c^2n^{1-2\gamma}}O(n^{-\beta\gamma}) + O(n^{-4\gamma}) \\ &= O(n^{2\gamma-1})O(n^{-\beta\gamma}) + O(n^{-4\gamma}) \\ &= O(n^{\gamma(2-\beta)-1}) + O(n^{-4\gamma}) \end{aligned}$$

This is $O(n^{\delta(\gamma)})$, where $\delta(\gamma) = \max\{\gamma(2-\beta)-1, -4\gamma\}$. From this we see that with a small positive γ, we obtain $\delta(\gamma) < 0$. This implies that $n^{-\delta(\gamma)}\text{MSE}(\hat{\Delta}_n)$ is bounded as n tends to infinity. Since $n^{-\delta(\gamma)}$ tends to infinity, we conclude that in this case, $\text{MSE}(\hat{\Delta}_n)$ tends to zero. In other words, if we decrease h as n tends to zero, the mean square error tends to zero and we obtain a consistent estimator. This is a very reasonable conclusion.

However, if $\beta < 2$, choosing a γ which is too large, corresponding to a very fast decrease in the step size, can end up yielding $\delta(\gamma) > 0$, corresponding to increasing mean square error. We see that choosing γ properly is paramount to obtaining effective finite difference estimates. The condition $\beta < 2$ corresponds to the situation where the decrease in $V(X'(\theta+h) - X'(\theta-h))$ cannot keep up with the factor $\frac{1}{h^2}$ in the expression for the total variance of $\hat{\Delta}_n$, and therefore large γ can cause an explosion in variance. The optimal choice of γ is found by minimizing δ over γ, leading to the fastest decrease of the mean squared error.

To this end, note that a negative γ always yields $\delta(\gamma) > 0$, so this is never a good choice. It will therefore suffice to optimize over positive γ. We split up according to whether $\beta < 2$, $\beta = 2$ or $\beta > 2$. If $\beta < 2$, $\delta(\gamma)$ is the maximum of a line with increasing slope and a line with decreasing slope. The minimum is therefore where the two lines intersect, meaning the solution to $\gamma(2-\beta) - 1 = -4\gamma$, which is $\gamma = \frac{1}{6-\beta}$. If $\beta = 2$, the minimum is obtained for any $\gamma \geq \frac{1}{4}$. The optimal choice of γ would in this case depend on the explicit form of the mean squared error. If $\beta > 2$, δ is downwards unbounded and γ should be chosen as large as possible. This latter case, however, probably will not happen in reality, at least not in any of the practial cases we consider.

Note that our analysis is somewhat different from the one found in Section 7.1 of Glasserman (2003), and the case $\beta > 2$ does not correspond to any of the cases considered there. This is because Glasserman (2003) always assumes that the total variance of the estimator tends to zero as h tends to zero. In a sense, the parametrization of the orders of decrease is different. The results, however, are equivalent.

To apply the finite difference method, then, very little actual implementation work is needed, but it may be necessary to spend some time figuring out the best selection of h.

We now consider two examples of application of the finite difference method. First, let $U(\theta)$ have the normal distribution with unit variance and mean θ, assume that we are interested in estimating $\frac{d}{d\theta}EU(\theta)^2$. If we let $U'(\theta+h)$ and $U'(\theta-h)$ be independent normal variables with unit variance and means $\theta+h$ and $\theta-h$, respectively, we obtain

$$\begin{aligned}
V\left(U'(\theta+h)^2 - U'(\theta-h)^2\right) &= VU'(\theta+h)^2 + VU'(\theta-h)^2 \\
&= 2(1+2(\theta+h)^2) + 2(1+2(\theta-h)^2) \\
&= 4 + 4((\theta+h)^2 + (\theta-h)^2) \\
&= 4 + 8\theta^2 + 8h^2,
\end{aligned}$$

which is $O(1)$ as h tends to zero. We have used that $U'(\theta+h)^2$ has a noncentral χ^2 distribution to obtain the variances. This corresponds to a β value of zero, and we would therefore expect that the optimal choice of step size is $h_n = cn^{-\frac{1}{6}}$ for some c. This would yields a total variance of

$$\frac{4+8\theta^2+8c^2n^{-\frac{1}{3}}}{4c^2n^{1-\frac{1}{3}}},$$

which is of order $o(n)$. To find the corresponding bias of the estimator, note that putting $\alpha(\theta) = EU(\theta)^2 = 1+\theta^2$, we find $\alpha'(\theta) = 2\theta$. The bias is then

$$\frac{1+(\theta+h)^2-(1+(\theta-h)^2)}{2h} - 2\theta = 0,$$

so the estimator is unbiased. On the other hand, if we let U be a standard normal and define $U'(\theta+h) = U+\theta+h$ and $U'(\theta-h) = U+\theta-h$, we find

$$\begin{aligned}
V\left(U'(\theta+h)^2 - U'(\theta-h)^2\right) &= V\left(U^2 + 2(\theta+h)U + h^2 - (U^2 + 2(\theta-h)U + h^2)\right) \\
&= V(4hU) \\
&= 16h^2,
\end{aligned}$$

which is $O(h^2)$. Using this simulation scheme then yields $\beta = 2$ and a reasonable choice is $h_n = cn^{-\gamma}$ for any $\gamma \geq \frac{1}{4}$. The total variance becomes

$$\frac{16c^2n^{-2\gamma}}{4c^2n^{1-2\gamma}} = \frac{4}{n},$$

which, as expected, is asymptotically independent of γ (in fact, completely independent). If the estimator were biased, we could have used our freedom in choosing γ to minimize the bias.

Pathwise differentiation. In the finite difference method, we considered a finite difference approximation to $\frac{d}{d\theta}EX(\theta)$ and let the parameter increment tend to zero at the same time as we let the number of samples tend to infinity. In the method of pathwise differentiation, we instead simply exchange differentiation and mean and use ordinary Monte Carlo to estimate the result. Thus, the basic operation we desire to make is

$$\frac{d}{d\theta}EX(\theta) = E\frac{d}{d\theta}X(\theta).$$

While the left-hand side is only dependent on each of the marginal distributions $X(\theta)$, the right-hand side is dependent on the distribution of the ensemble $(X(\theta))_{\theta\in\Theta}$. For the interchange to make sense at all, we therefore assume that we are working in a context where all of the variables X_θ are defined on the same probability space. Furthermore, it is of course necessary that the variable $\frac{d}{d\theta}X(\theta)$ is almost surely well-defined and integrable, so we need to assume that X_θ is almost surely differentiable in θ for any $\theta \in \Theta$. Finally, we need to make sure that the exchange of integration and differentiation is allowed.

If all this is the case, we can let $X'_n(\theta)$ be an i.i.d. sequence with common distribution being the same as the distribution of $\frac{d}{d\theta}X(\theta)$ and define the pathwise differentiation estimator

$$\frac{1}{n}\sum_{k=1}^{n} X'_k(\theta),$$

which by the law of large numbers and our assumptions satisfies

$$\frac{1}{n}\sum_{k=1}^{n} X'_k(\theta) \xrightarrow{\text{a.s.}} E\frac{d}{d\theta}X(\theta) = \frac{d}{d\theta}EX(\theta),$$

so the pathwise differentiation estimator, when it exists, is always consistent. However, we have no guarantees that the variance of the pathwise estimator is superior to that of the finite difference estimator. Two definite benefits of the pathwise differentiation method over the finite difference method, though, are its unbiasedness and that we no longer need to worry about the optimal parameter increment in the difference quotient. Other than that, we have no guarantees for improvement. General experience states that the pathwise differentiation estimator rarely is worse than the finite difference estimator, but also rarely yields more than moderate improvements in variance.

We will now consider sufficient conditions for the existence and consistency of the pathwise differentiation estimator. For clarity, we reiterate that we at all times assume

1. The variables X_θ are all defined on the same probability space.
2. For each θ, $X(\theta)$ is almost surely differentiable.
3. The variable $\frac{d}{d\theta}X(\theta)$ is integrable.

A sufficient condition for the interchange of differentiation and integration is then that θ is if the family $(\frac{X(\theta+h)-X(\theta)}{h})$ is uniformly integrable over h in some punctured neighborhood of zero: Since the variables converge almost surely to $\frac{d}{d\theta}X(\theta)$, we also have convergence in probability, and this combined with uniform integrability yields convergence in mean and therefore convergence of the means, that is,

$$\frac{d}{d\theta}EX(\theta) = \lim_{h\to 0} E\left(\frac{X(\theta+h)-X(\theta)}{h}\right) = E\left(\lim_{h\to 0}\frac{X(\theta+h)-X(\theta)}{h}\right) = E\frac{d}{d\theta}X(\theta),$$

as desired. We now claim that the following is a sufficient condition for the interchange of differentiation and integration:

- $V(X(\theta+h)-X(\theta)) = O(h^2)$.

To see this, note that under this condition,

$$\begin{aligned}
E(X(\theta+h)-X(\theta))^2 &= V(X(\theta+h)-X(\theta)) + E^2(X(\theta+h)-X(\theta)) \\
&= O(h^2) + h^2\left(\frac{EX(\theta+h)-EX(\theta)}{h}\right)^2 \\
&= O(h^2),
\end{aligned}$$

so the family $\frac{X(\theta+h)-X(\theta)}{h}$ is bounded in $\mathcal{L}^2$ over a punctured neighborhood of zero. In particular, it is uniformly integrable, so the interchange is allowed by what we already have shown. Another sufficient condition is:

- For $\theta\in\Theta$, there is an integrable variable κ_θ such that $|X(\theta+h)-X(\theta)| \le \kappa_\theta |h|$ for all h in a punctured neighborhood of zero.

In order the realize this, merely note that in this case, κ_θ is an integrable bound for $\frac{|X(\theta+h)-X(\theta)|}{h}$, and therefore the dominated convergence theorem yields the result.

We now turn to a concrete example illustrating the use of the pathwise differentiation method. We consider the same situation as in the subsection on the finite difference method, estimating $\frac{d}{d\theta}EU(\theta)^2$ where $U(\theta)$ is normally distributed with unit variance and mean θ. To use the pathwise method, we first need to make sure that all of the variables are defined on the same probability space. This is easily obtained by letting U be some variable with the standard normal distribution and defining $U(\theta) = U+\theta$. Clearly, then, $\theta \mapsto (U+\theta)^2$ is always everywhere differentiable, and the derivative is

$\frac{d}{d\theta}(U+\theta)^2 = 2(U+\theta)$, which is integrable. We earlier found

$$V(U(\theta+h)^2 - U(\theta)^2) = 4h^2 = O(h^2),$$

so by our earlier results, the interchange of differentiation and integration is allowed, and we conclude

$$\frac{d}{d\theta}EU(\theta)^2 = E\frac{d}{d\theta}U(\theta)^2 = E2(U+\theta).$$

Now, we can of course evalutate this expectation directly, but for the sake of comparing with the finite difference method, let us consider what would happen if we used a Monte Carlo estimator. The variance based on n samples would then be $\frac{1}{n}V(2U+2\theta) = \frac{4}{n}$, which is the same as the finite difference estimator.

Likelihood ratio. The likelihood ratio method, like the pathwise differentiation method, seeks to evaluate the sensitivity by interchanging differentiation and integration. But instead of differentiating the variables themselves, the likelihood ratio method differentiates their densities. Since densities usually in most cases are relatively smooth, while the stochastic processes often, in particular in the case of option valuation, have discontinuities or breaks, this means that the likelihood ratio method in many cases applies when the pathwise differentiation method does not.

To set the stage, assume that $X(\theta) = f(X'(\theta))$ for some $X'(\theta)$ with density g_θ, where $f : \mathbb{R} \to \mathbb{R}$ is some measurable mapping. This is the formulation best suited to later applications. Our goal is then to evaluate $\frac{d}{d\theta}Ef(X'(\theta))$. The basis of the likelihood ratio method is the observation that if the interchange is allowed and the denominators are nonzero, we have

$$\begin{aligned}
\frac{d}{d\theta}Ef(X'(\theta)) &= \frac{d}{d\theta}\int f(x)g_\theta(x)\,\mathrm{d}x \\
&= \int f(x)\frac{d}{d\theta}g_\theta(x)\,\mathrm{d}x \\
&= \int f(x)\frac{g'_\theta(x)}{g_\theta(x)}g_\theta(x)\,\mathrm{d}x \\
&= Ef(X'(\theta))\frac{g'_\theta(X'(\theta))}{g_\theta(X'(\theta))}.
\end{aligned}$$

While the pathwise method rewrote the sensitivity as an expectation of a somewhat different variable, the likelihood ratio method rewrites the sensitivity as a new transformation of the old variable.

As with the pathwise method, there is no guarantee that the likelihood ratio method yields better results than the finite difference estimator. In fact, it is easy to find cases where it yields substantially worse results than finite difference, and we shall see examples of this later. As a rule of thumb, if the mapping f has discontinuities, the likelihood ratio has a good chance of improving convergence, otherwise it may easily worsen convergence. In this sense, the likelihood method complements the pathwise method: the pathwise method is usually useful in the presence of continuity, and the likelihood ratio is usually useful in the presence of discontinuities. A limitation of the likelihood ratio method is that it is necessary to know the density of the variables and their derivatives, as this is rarely the case in practice. However, the scheme can be extended to cover very general cases by approximation arguments, see Section 7.3 of Glasserman (2003).

There are no criteria in simple terms for when the interchange of differentiation and integration of the likelihood ratio method is allowed. However, general criteria can be expressed in the case where the family $X'(\theta)$ is an exponential family. See for example the comments to Lemma 1.3 of Jensen (1992), page 7.

Malliavin weights. The method of Malliavin weights, much as the likelihood method, computes sensitivities by transforming the derivative of the expectation into another expectation. The method does not work in general cases, but is specifically suited to the context where the sensitivity to be calculated is with respect to a transformation of the solution to an SDE. Specifically, consider a n-dimensional Brownian motion W and assume that the SDE

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$$

has a strong solution F, such that for each $x \in \mathbb{R}^n$, $X^x = F(x, W)$ is a solution to the above SDE with initial condition $x \in \mathbb{R}^n$. We further assume that the corresponding flow is differentiable and let Y^x be the Jacobian of X^x. Furthermore, let $T > 0$ be some constant and let $f : \mathbb{R}^n \to \mathbb{R}$ be some mapping such that $Ef(X_T^x)^2$ is finite. The basic result which we will need to derive the Malliavin estimator is that putting $u(x) = Ef(X_T^x)$, it holds under suitable regularity conditions that

$$\nabla u(x) = E\left(f(X_T^x)\frac{1}{T}\int_0^T (\sigma^{-1}(X_t^x)Y_t^x)^t\,\mathrm{d}W_t \right)$$

This result is Proposition 3.2 of Fournié et al. (1999). Here, σ^{-1} denotes matrix inverse and not the inverse mapping. We are not able to prove this with the theory that we have developed. The proof exploits several of the results from Section 4.5, among others the duality formula. It also uses the Diffeomorphism Theorem and certain criteria for uniform integrability from the theory of SDEs. See the notes for further details and references.

We will now see how this result specialises to the case of one-dimensional financial markets. First consider the case where the financial instrument is given on the form $\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma(S_t)S_t\,\mathrm{d}W_t$ for some positive measurable mapping σ. Let f be such that $Ef(S_T)^2$ is finite, we will find the delta of the T-claim $f(S_T)$. We assume that the short rate is zero, our results will then obviously immediately extend to the case of deterministic short rate. Assuming that the result from Fournié et al. (1999) can be applied, we can use the fact that $\frac{\partial}{\partial S_0}S_T = \frac{S_T}{S_0}$ to immediately conclude

$$\begin{aligned}\frac{\partial}{\partial S_0}Ef(S_T) &= E\left(f(S_T)\frac{1}{T}\int_0^T \frac{1}{\sigma(S_T)S_T}\frac{S_T}{S_0}\,\mathrm{d}W_t\right)\\ &= E\left(f(S_T)\frac{1}{TS_0}\int_0^T \frac{1}{\sigma(S_T)}\,\mathrm{d}W_t\right).\end{aligned}$$

This is the basis for the Malliavin weight estimator of the delta in the case where the volatility is of the local form. This applies, for example, to the Black-Scholes model. Next, we consider a market of the form

$$\begin{aligned}\mathrm{d}S_t &= rS_t\,\mathrm{d}t + \sigma_t S_t\sqrt{1-\rho^2}\,\mathrm{d}W_t^1 + \sigma_t S_t \rho\,\mathrm{d}W_t^2\\ \mathrm{d}\sigma_t &= \alpha(\sigma_t)\,\mathrm{d}t + \beta(\sigma_t)\,\mathrm{d}W_t^2,\end{aligned}$$

with constant initial conditions S_0 and σ_0, where we assume that α and β are differentiable and $\rho \in (-1,1)$. We further assume that β is positive. This covers for example the Heston model. As before, consider f such that $Ef(S_T)^2$ is finite, we desire to find the delta of the T-claim $f(S_T)$, that is, $\frac{\partial}{\partial S_0}Ef(S_T)$. To do so, we define

$$\mu(x) = \begin{pmatrix} rx_1 \\ \alpha(x_2)\end{pmatrix} \qquad \text{and} \qquad \sigma(x) = \begin{pmatrix} x_2x_1\sqrt{1-\rho^2} & x_2x_1\rho \\ 0 & \beta(x_2)\end{pmatrix},$$

and put $X_t = (S_t, \sigma_t)$. We then obtain

$$\begin{aligned}\mathrm{d}X^1(t) &= \mu_1(X_t)\,\mathrm{d}t + \sigma_{11}(X_t)\,\mathrm{d}W_t^1 + \sigma_{12}(X_t)\,\mathrm{d}W^2\\ \mathrm{d}X^2(t) &= \mu_2(X_t)\,\mathrm{d}t + \sigma_{21}(X_t)\,\mathrm{d}W_t^1 + \sigma_{22}(X_t)\,\mathrm{d}W_t^2\end{aligned}$$

with the initial conditions $X_0 = (S_0, \sigma_0)$. Note that $\frac{\partial}{\partial S_0}Ef(S_T) = \frac{\partial}{\partial X_1(0)}Ef(X_1(T))$. This observation leads us to believe that we may be able to use the result from Fournié et al. (1999) to obtain an expression for the sensitivity. Therefore, we assume that there is a strong solution to this SDE, yielding a differentiable flow, and we assume that the results of Fournié et al. (1999) applies to our situation. Let Y be the first variation process of X under the initial condition X_0. We find

$$\begin{aligned}
\sigma^{-1}(X_t)Y_t &= \begin{pmatrix} \frac{1}{X_t^2 X_t^1 \sqrt{1-\rho^2}} & -\frac{\rho}{\sqrt{1-\rho^2}\beta(X_t^2)} \\ 0 & \frac{1}{\beta(X_t^2)} \end{pmatrix} \begin{pmatrix} Y_t^{11} & Y_t^{12} \\ Y_t^{21} & Y_t^{22} \end{pmatrix} \\
&= \begin{pmatrix} \frac{Y_t^{11}}{X_t^2 X_t^1 \sqrt{1-\rho^2}} - \frac{\rho Y_t^{21}}{\sqrt{1-\rho^2}\beta(X_t^2)} & \frac{Y_t^{12}}{X_t^2 X_t^1 \sqrt{1-\rho^2}} - \frac{\rho Y_t^{22}}{\sqrt{1-\rho^2}\beta(X_t^2)} \\ \frac{Y_t^{21}}{\beta(X_t^2)} & \frac{Y_t^{22}}{\beta(X_t^2)} \end{pmatrix}.
\end{aligned}$$

Because X^2 does not depend on X_0^1, $Y_t^{21} = \frac{\partial}{\partial X_0^1} X_t^2 = 0$. And since X^1 is given as the solution to an exponential SDE, $Y^{11} = \frac{\partial}{\partial X_0^1} X_t^1 = \frac{1}{X_0^1} X_t^1$. Therefore, we can reduce the above to

$$\sigma^{-1}(X_t)Y_t = \begin{pmatrix} \frac{1}{X_t^2 X_0^1 \sqrt{1-\rho^2}} & \frac{Y_t^{12}}{X_t^2 X_t^1 \sqrt{1-\rho^2}} - \frac{\rho Y_t^{22}}{\sqrt{1-\rho^2}\beta(X_t^2)} \\ 0 & \frac{Y_t^{22}}{\beta(X_t^2)} \end{pmatrix}.$$

Transposing this, the result from Fournié et al. (1999) yields

$$\frac{\partial}{\partial X_0^1} Ef(X_T^1) = E\left(f(X_T^1) \frac{1}{T} \int_0^T \frac{1}{X_t^2 X_0^1 \sqrt{1-\rho^2}} \, \mathrm{d}W_t^1 \right).$$

Collecting our results and substituting S and σ for X^1 and X^2, we may conclude

$$\frac{\partial}{\partial S_0} Ef(S_T) = E\left(f(S_T) \frac{1}{S_0 T} \int_0^T \frac{1}{\sigma_t \sqrt{1-\rho^2}} \, \mathrm{d}W_t^1 \right),$$

and this is the Malliavin estimator for the delta. Note that this is *not* the same estimator as is obtained in Benhamou (2001). We will return to this point later, when considering the practical implementation of the Malliavin estimator in the Heston model.

Localisation. The method of localisation is more of a shrewd observation than an actual method. It can in principle be used in conjunction with any of the Monte Carlo methods we have described, but its primary usefulness for us will be together with the likelihood ratio and Malliavin methods. We show how the method applies to the likelihood ratio method, the application to the Malliavin method is similar. We therefore consider the same situation as in our discussion of the likelihood method, where we put $X(\theta) = f(X'(\theta))$ and desire to evaluate $\frac{d}{d\theta} EX(\theta)$. Letting g_θ be the density of $X'(\theta)$ and putting $\alpha_\theta(x) = \frac{g'_\theta(x)}{g_\theta(x)}$, the likelihood ratio method yields the equality

$$\frac{d}{d\theta} EX(\theta) = Ef(X'(\theta))\alpha_\theta(X'(\theta)).$$

Now, if α_θ has a large variability and the supports of f and α_θ have a large common support, the factor $\alpha_\theta(X'(\theta))$ could be instrumental in making the variance of estimators based upon the above rather large. Such detrimental effects are actually observed in reality, which means that the removal of the derivative in such cases have a very large price, and ordinary finite difference methods may be more effective.

One way to avoid such enlargements of variance is to make the simple observation that for any sufficiently integrable mapping h, we have

$$\begin{aligned}\frac{d}{d\theta}EX(\theta) &= \frac{d}{d\theta}Eh(X'(\theta)) + \frac{d}{d\theta}E(f-h)(X'(\theta)) \\ &= Eh(X'(\theta))\alpha_\theta(X'(\theta)) + \frac{d}{d\theta}E(f-h)(X'(\theta)).\end{aligned}$$

Here, we have split the mapping f into two parts, h and $f-h$, and have applied the likelihood ratio method only to the first part. The second term can be evaluated using other methods, such as the finite difference method or the pathwise method. If we can choose h such that the variance of $h(X'(\theta))\alpha_\theta(X'(\theta))$ is small, then we have obtained an improvement over the usual likelihood ratio estimator provided that the second term does not add too much variance. One way of obtaining this reduced variance is to make sure that the h has a small support. The choice of h, however, should be made such that the usefulness of the likelihood ratio method is preserved and such that the remained term can be calculated with other methods. In other words, we we would like the h to have small support and $f-h$ to have reasonable smoothness. One example of this is when working with the digital option, where $f-h$ can be chosen as a smoothened version of the digital payoff, and h is then a kind of small bump function with a jump. This enables one to use the likelihood ratio to take care of the discontinuity and using, say, the pathwise method, to evaluate the remainder. We shall se examples of how to do this in the next section.

5.5 Estimation in the Black-Scholes model

We now consider the Black-Scholes model, given by the dynamics

$$\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t,$$

endowed with a risk free asset B based on a constant short rate r with initial condition $B_0 = 1$. We will prove that the model is free of arbitrage and implement the Monte Carlo methods discussed in the previous sections for the Black-Scholes model. Now, our main goal of our numerical work is to analyze the performance of the Malliavin estimator and the Malliavin estimator combined with ordinary Monte Carlo methods, comparing it to other estimators. As we shall see, in the Black-Scholes case, the Malliavin method coincides with the Likelihood ratio method. Therefore, in this case, the Malliavin calculus brings nothing new at all. As a consequence, the Black-Scholes case is mostly a kind of warm-up case for us, allowing us to familiarize ourselves with the various methods in a simple setting before turning to the somewhat more cumbersome case of the Heston model.

Theorem 5.5.1. *The Black-Scholes model is free of arbitrage, and there is a T-EMM yielding the dynamics* $\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$.

Proof. Both statements of the theorem follows immediately from Corollary 5.1.10. □

Comment 5.5.2 The T-EMM is in fact in this case unique, corresponding to completeness of the Black-Scholes model. ∘

For the numerical experiments, we consider three problems:

1. The price of a call option.
2. The price of a digital option.
3. The delta of a digital option.

It is easy to check for consistency of our methods, because in all three cases, the correct value has an analytical formula, as the following lemma shows.

Lemma 5.5.3. *Define* $d_1(s) = \frac{1}{\sigma\sqrt{T}}(\log(\frac{s}{K}) + (r + \frac{1}{2}\sigma^2)T)$ *and* $d_2(s) = d_1(s) - \sigma\sqrt{T}$.

1. *The price of a strike K T-call option is* $s\Phi(d_1(S_0)) - e^{-rT}K\Phi(d_2(S_0))$.
2. *The delta of a strike K T-call option is* $\Phi(d_1(S_0))$.
3. *The delta of a strike K T-digital option is* $e^{-rT}\frac{\phi(d_2(S_0))}{\sigma\sqrt{T}S_0}$.

Proof. The first result is the Black-Scholes formula, see Proposition 7.10 in Björk (2004). The second is Proposition 9.5 in Björk (2004). To prove the last result, we first calculate the price of a digital option as

$$
\begin{aligned}
& e^{-rT}E^Q(1_{(S_T\geq K)}) \\
= \;& e^{-rT}Q(S_T \geq K) \\
= \;& e^{-rT}Q\left(\frac{1}{\sigma\sqrt{T}}\left(\log\frac{S_T}{s}-\left(r-\frac{1}{2}\sigma^2\right)T\right)\geq\frac{1}{\sigma\sqrt{T}}\left(\log\frac{K}{s}-\left(r-\frac{1}{2}\sigma^2\right)T\right)\right) \\
= \;& e^{-rT}\left(1-\Phi\left(\frac{1}{\sigma\sqrt{T}}\left(\log\frac{K}{s}-\left(r-\frac{1}{2}\sigma^2\right)T\right)\right)\right) \\
= \;& e^{-rT}\Phi(d_2(S_0)).
\end{aligned}
$$

From this we immediately obtain $\frac{d}{dS_0}e^{-rT}E^Q(1_{(S_T\geq K)}) = e^{-rT}\phi(d_2(S_0))\frac{1}{\sigma\sqrt{T}S_0}$. □

We now proceed to the numerical experiments. All three problems all yield differing degrees of irregularity, and we shall see that the methods applicable are different for all three cases.

The price of a call option. Letting W be a Brownian motion and letting S be the corresponding geometric Brownian motion with drift r and volatility σ, we want to evaluate $e^{-rT}E(S_T-K)^+$. We consider $r = 4\%$, $\sigma = 0.2$, $S_0 = 100$, $T = 1$ and $K = 104$. We consider a plain Monte Carlo estimator, an antithetic estimator, a stratified estimator, a control variate estimator and finally, an estimator combining all of these.

The plain Monte Carlo estimator is simply implemented by simulating scaled lognormal variables with scale S_0, log mean $(r-\frac{1}{2}\sigma^2)T$ and log variance σ^2T. The antithetic estimator stratifies uniform variables and transforms them with the quantile function of the lognormal distribution. The control variate estimator uses S_T as a control variate, which has the known mean $e^{rT}S_0$ and has a correlation with $(S_T-K)^+$ of around 60%. To analyze the results, we know that all the estimators are unbiased, so it will suffice to compare standard deviations. We base our standard deviation estimates on 80 Monte Carlo simulations for each method with batches from 2000 simulations to 90000 simulations with step sizes of 2000. Since some methods use more computational time per simulation than others, we compare the standard deviations as functions of computing time instead of as functions of number of simulations. The results can be seen in Figure 5.3. We see that the stratified and the combined estimators have graphs which contain no data before around 0.002 seconds of computing time. This is because we base our results on the same batches of simulations. Since these two methods take significantly longer time per simulation than the other methods, their

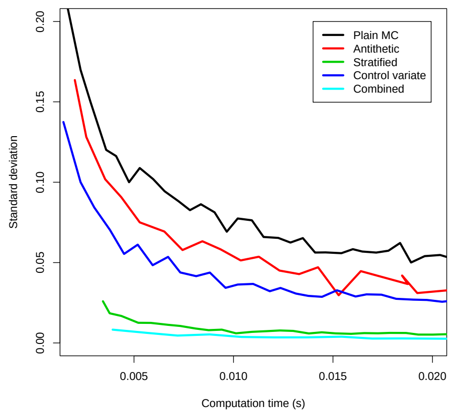

Figure 5.3: Comparison of standard deviations of Monte Carlo estimators for the price of a call option in the Black-Scholes model.

batches correspond to longer computational times. The effect is not particularly large here, but the distinction between computing time and number of simulations will become important for example when considering the digital delta, where the most time-consuming method will spend six times as much computing time per simulation compared to the simplest method. Comparing standard deviations as functions of the number of simulations would therefore have made the more advanced methods seem six times more effective than they really are.

The results show, consistently with the general rules of thumb set forth in Secion 4.7 of Glasserman (2003), that stratified sampling yields the greatest benefit, followed by the control variate method and then antithetic sampling. Combining all of these yields the best estimator. This estimator has a standard deviation which is up to 20 times smaller than the ordinary Monte Carlo estimator.

Importance sampling for OTM call options. So far, we have neglected the importance sampling method. This is because for options which are not far out of the money, it has little impact. For simplicity, we will in general only consider call and digital options which are not far out of the money. For completeness, however, we now illustrate how importance sampling can be used to facilitate computations for out of the money options.

We consider the same call as before, but change the strike from 104 to 225. Since the initial value of the instrument is 100, this makes the call quite far out of the money. This means that many of the simulated values of S_T will yield a payoff of zero, which makes the estimation of the mean difficult. We will use importance sampling to change the distribution of the underlying normal distribution such that samples giving a payoff are more likely. To this end, we note that with $d_2(x) = \frac{1}{\sigma\sqrt{T}}(\log\frac{x}{S_0} - (r - \frac{1}{2}\sigma^2)T)$ and $Z = \frac{W_T}{\sigma\sqrt{T}}$, we have $E(S_T - K)^+ = E(d_2(Z\sigma\sqrt{T}) - K)^+$. Let μ be the distribution of Z, μ is then the standard normal distribution. Let ν be the normal distribution with mean ξ and unit variance. Let f_μ and f_ν be the corresponding densities. We then have

$$\begin{aligned}
\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) &= \frac{f_\mu(x)}{f_\nu(x)} \\
&= \frac{\exp\left(-\frac{1}{2}x^2\right)}{\exp\left(-\frac{1}{2}(x-\xi)^2\right)} \\
&= \exp\left(-\frac{1}{2}\left(x^2 - (x-\xi)^2\right)\right) \\
&= \exp\left(-\xi x + \frac{1}{2}\xi^2\right).
\end{aligned}$$

Letting Y be a normal distribution with mean ξ and unit variance, we therefore obtain

$$E(S_T - K)^+ = E(d_2(Y\sigma\sqrt{T}) - K)^+ \exp\left(-\xi Y + \frac{1}{2}\xi^2\right),$$

which is the basis of the importance sampling estimator based on ν. By trial and error experimentation, we find that $\xi = 2$ is a good choice for the problem at hand. Figure 5.4 compares estimated means and standard deviations for the ordinary Monte Carlo estimator, the importance sampling estimator and the importance sampling estimator with antithetic and stratified sampling and using Y as a control variate. Not surprisingly, we see that the importance sampling estimators are clearly superior. The combined estimator has a standard deviation which is up to 130 times smaller than the ordinary estimator.

The delta of a call option. Next, we consider the delta of a call option. This situation will allow us to use some of the Monte Carlo methods designed specifically for evaluation of sensitivities. We implement the finite difference method, the pathwise differentiation method and the likelihood ratio method. As we shall see, in this case the pathwise differentiation method seems to be the optimal. We also consider an estimator combining the pathwise differentiation method with antithetic and stratified sampling. The methods for estimating sensitivities are not as straightforward to use

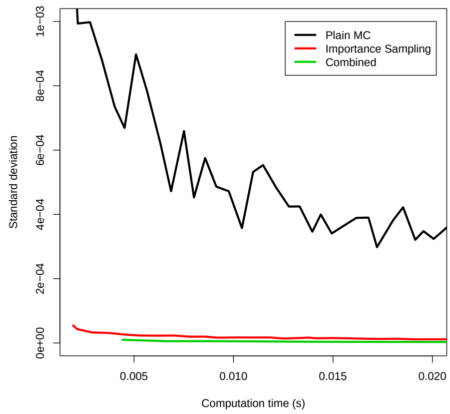

Figure 5.4: Comparison of means and standard deviations of Monte Carlo estimators for an OTM call option in the Black-Scholes model.

as those for plain Monte Carlo estimation. We will spend a short time reviewing the analysis and implementation.

As we saw in Section 5.4, to implement the finite difference method, we need to choose how fast to let the denominator in the finite difference approximation tend to zero as the number of simulations increase, and this choice depends on the variance of the numerator of the finite difference apprroximation. In practice, we also need to choose the constant factor c in $h_n = cn^{-\gamma}$. In our case, the numerator is $V(f(S_T(S_0+h)) - f(S_T(S_0-h)))$, where $S_T(S_0)$ is the solution of the SDE for the financial instrument with initial value S_0, and f is the call payoff. Our simulation method is to simulate a geometric Brownian motion with initial value one, multiply by S_0+h and S_0-h and transform with the payoff function. $f(S_T(S_0+h))$ and $f(S_T(S_0-h))$ are then somewhat correlated. This both improves convergence and saves time generating simulations. When dicsussing the finite difference method in the last section, we gave arguments for different choices according to different situations. We begin by examining how much of a difference these choices really makes. Figure 5.5 shows the results. Basically, the figure shows that as long as our choices aren't completely unreasonable, the difference is quite small, in particular for large samples. The behaviour is consistent with a β value of 2. We pick $\gamma = \frac{1}{6}$ and $c = 100$. For the pathwise differentiation method, we

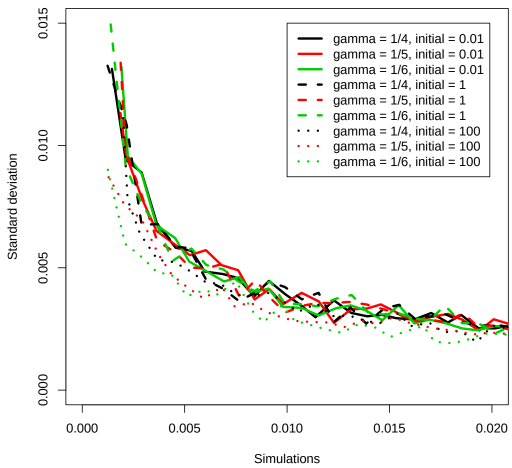

Figure 5.5: Comparison of standard deviations of Monte Carlo estimators for the delta of a call option in the Black-Scholes model.

wish to make the interchange

$$\frac{\partial}{\partial S_0}E(S_T-K)^+ = E\frac{\partial}{\partial S_0}E(S_T-K)^+.$$

We need to check the conditions in Section 5.4. Since we have the explicit formula $S_T = S_0\exp((r-\frac{1}{2}\sigma^2)T+\sigma W_T)$, the variables for varying S_0 are clearly defined on the same probability space, and S_T is always differentiable as a function of S_0. Since $x \mapsto (x-K)^+$ is everywhere differentiable except in K and $S_T = K$ with probability zero for any value of S_0, we conclude that $(S_T-K)^+$ is almost surely differentiable with respect to S_0, and since $\frac{\partial S_T}{\partial S_0} = \frac{S_T}{S_0}$, a derivative is $1_{[K,\infty)}(S_T)\frac{S_T}{S_0}$, which is integrable. To check that the interchange of differentiation and integration is allowed, we merely note that with $S_T(S_0) = S_0\exp((r-\frac{1}{2}\sigma^2)T+\sigma W_T)$,

$$|(S_T(S_0+h)-K)^+ - (S_T(S_0)-K)^+| \le |h|\exp\left(\left(r-\frac{1}{2}\sigma^2\right)T+\sigma W_T\right),$$

using that $x \mapsto (x-K)^+$ is Lipschitz with constant 1. Since the coefficient to $|h|$ is integrable, by the results of Section 5.4, the interchange of differentiation and integration is allowed, and the corresponding pathwise estimator is based on $E1_{[K,\infty)}(S_T)\frac{S_T}{S_0}$.

For the likelihood ratio method, we want to make the interchange

$$\frac{\partial}{\partial S_0} E(S_T - K)^+ = E(S_T - K)^+ \frac{\frac{\partial}{\partial S_0} g_{S_0}(S_T)}{g_{S_0}(S_T)},$$

to obtain the right-hand side as the basis for the likelihood ratio estimator, where g_{S_0} is the density of S_T under the initial condition S_0, which is given by the expression $g_{S_0}(x) = \frac{1}{x\sigma\sqrt{T}}\phi(d_2(S_0, x))$, with $d_2(S_0, x) = \frac{1}{\sigma\sqrt{T}}(\log \frac{x}{S_0} - (r - \frac{1}{2}\sigma^2)T)$. We will not try to justify the interchange, but merely conclude that numerical experiments show that it in fact yields the correct result. To obtain a simpler expression, we note

$$\frac{\frac{\partial}{\partial S_0} g_{S_0}(x)}{g_{S_0}(x)} = \frac{\frac{1}{x\sigma\sqrt{T}} \frac{1}{S_0\sigma\sqrt{T}} d_2(S_0, x)\phi(d_2(S_0, x))}{\frac{1}{x\sigma\sqrt{T}}\phi(d_2(S_0, x))} = \frac{d_2(S_0, x)}{S_0\sigma\sqrt{T}},$$

which shows, using the representation of S_T in terms of W_T,

$$\frac{\frac{\partial}{\partial S_0} g_{S_0}(S_T)}{g_{S_0}(S_T)} = \frac{d_2(S_0, S_T)}{S_0\sigma\sqrt{T}} = \frac{W_T}{S_0\sigma\sqrt{T}},$$

so that we need to calculate the mean $E(S_T - K)^+ \frac{W_T}{S_0\sigma\sqrt{T}}$. We mentioned in the introduction to this section that the likelihood method and the Malliavin weights method are equivalent for the Black-Scholes model. This is quite clear, since the Malliavin weights method yields

$$\frac{\partial}{\partial S_0} E(S_T - K)^+ = E(S_T - K)^+ \frac{1}{S_0 T} \int_0^T \frac{1}{\sigma} \, \mathrm{d}W_t = E(S_T - K)^+ \frac{W_T}{S_0\sigma\sqrt{T}},$$

which is precisely the same as the likelihood ratio method.

We know that the pathwise differentiation and likelihood ratio methods are unbiased, but the finite difference method has a bias. Our simulation results show that this bias generally is between 10^{-4} and 10^{-5}. This is small enough that we can disregard it, and to compare the estimators, it will then suffice to compare standard deviations. We compare a finite difference estimator, a pathwise estimator, a likelihood ratio estimator and finally, a pathwise estimator with antithetic and stratified sampling. As in the previous case, we base our standard deviation estimates on 80 Monte Carlo simulations for each method with batches from 2000 simulations to 90000 simulations with step sizes of 2000. The results can be seen in Figure 5.6. We see that the finite difference and the pathwise differentiation methods yield almost similar results. When combined with antithetic and stratified sampling, a considerable boost in effectivity is obtained. The likelihood ratio is in this case clearly inferior to the other methods, approximately doubling the standard deviation. This conclusion supports the common

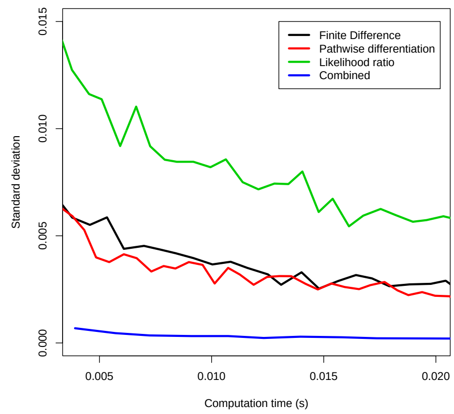

Figure 5.6: Comparison of standard deviations of Monte Carlo estimators for the delta of a call option in the Black-Scholes model.

wisdom that the likelihood ratio method mostly is useful when applied in the context of discontinuities. This conclusion is consistent with the results of Benhamou (2001) and Benhamou (2003). To see what goes wrong with the likelihood ratio method, recall that the likelihood ratio estimator is based on the formula

$$\frac{d}{dS_0}E(S_T - K)^+ = E(S_T - K)^+ \frac{W_T}{S_0\sigma\sqrt{T}}.$$

While the weight in the expectation on the right-hand side enables us to remove the derivative, it also introduces extra variance: The variable $(S_T - K)^+ W_T$ gets very large when W_T is large. We could use the localisation methods described in Section 5.4 to minimize this effect, but in reality, the likelihood ratio method simply is not suited to this particular problem. When considering the delta of a digital option, we will see a situation where the likelihood ratio and localised likelihood ratio methods provide effective results.

The delta of a digital option. Our final numerical experiment for the Black-Scholes model is the evaluation of the delta for a digital option. The methods which are a priori at our disposal are that of finite difference, pathwise differentiation and likelihood ratio. However, in this case, the pathwise differentiation method cannot be applied: Even though $1_{[K,\infty)}(S_T)$ is almost surely differentiable for any initial condition, the derivative is also almost surely zero, so the pathwise differentiation method yields a delta for the digital option of zero, which is clearly nonsense. We conclude that the interchange of differentiation and integration required for application of the pathwise differentiation method is not justified. Intuitively, this is because the payoff is not Lipschitz. We are therefore left with the finite difference method and the likelihood ratio method. We implement these, and we also implement a localised version of the likelihood ratio method. Finally, we also consider a localised likelihood ratio estimator with antithetic and stratified sampling.

The finite difference and likelihood ratio methods are implemented as for the call delta. We will discuss the details of the localisation of the likelihood method, where we proceed as described in Section 5.4. Let $f(x) = 1_{[K,\infty)}(x)$ be the digital payoff. Our goal is to make a decomposition of the form $f(x) = h_\varepsilon(x) + (f - h_\varepsilon)(x)$, where $f - h_\varepsilon$ is reasonably smooth and h_ε has small support and contains the discontinuity of the payoff. Led by Glasserman (2003), Section 7.3, we define h_ε by the relation

$$f(x) - h_\varepsilon(x) = \min\left\{1 - \frac{1}{2\varepsilon}\max\{0, x - K + \varepsilon\}\right\}$$

We have in this way ensured that $f - h_\varepsilon$ is continuous. We then obtain

$$h_\varepsilon(x) = 1_{[K,\infty)}(x)(K - \varepsilon - x)^- + 1_{(-\infty,K)}(K + \varepsilon - x)^+.$$

h_ε is a function with small support containing the discontinuity of the payoff. $f - h_\varepsilon$ is continuous. See Figure 5.7 for illustrations of the decomposition of the payoff.

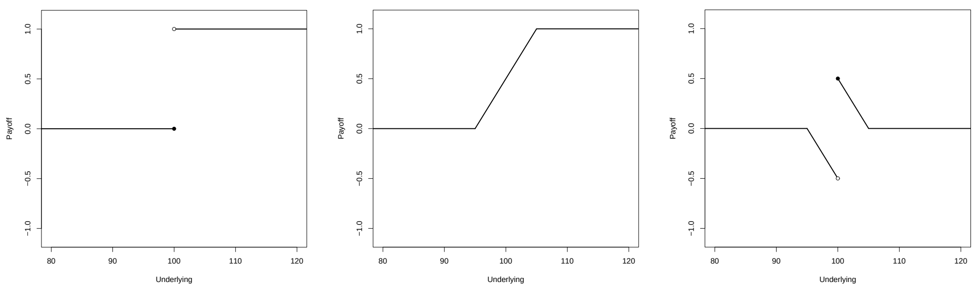

Figure 5.7: From left to right: The payoff of a digital option, the smoothed payoff $f - h_\varepsilon$ and the jump part h_ε.

The likelihood ratio method applied to the h_ε term then yields

$$\begin{aligned} \frac{\partial}{\partial S_0} E(S_T - K)^+ &= \frac{\partial}{\partial S_0} E h_\varepsilon(S_T) + \frac{\partial}{\partial S_0} E(f - h_\varepsilon)(S_T) \\ &= E h_\varepsilon(S_T)\frac{W_T}{S_0\sigma\sqrt{T}} + \frac{\partial}{\partial S_0} E(f - h_\varepsilon)(S_T). \end{aligned}$$

We will use pathwise differentiation on the second term. $f - h_\varepsilon$ is almost surely differentiable for any initial condition with $\frac{\partial}{\partial S_0}(f - h_\varepsilon)(S_T) = \frac{1}{2\varepsilon}1_{(|S_T - K| < \varepsilon)}$. Note that $f - h_\varepsilon$ is Lipschitz with constant $\frac{1}{2\varepsilon}$, and therefore

$$\begin{aligned} |(f - h_\varepsilon)(S_T(S_0 + h)) - (f - h_\varepsilon)(S_T(S_0))| &\leq \frac{1}{2\varepsilon}|S_T(S_0 + h) - S_T(S_0)| \\ &\leq \frac{1}{2\varepsilon}|h| \exp\left(\left(r - \frac{1}{2}\sigma^2\right)T + \sigma W_T\right), \end{aligned}$$

so the interchange of differentiation and integration is justified, and we finally obtain

$$\frac{\partial}{\partial S_0}E(S_T - K)^+ = Eh_\varepsilon(S_T)\frac{W_T}{S_0\sigma\sqrt{T}} + E1_{(|S_T - K| < \varepsilon)}\frac{S_T}{2\varepsilon S_0},$$

which forms the basis of the localised likelihood ratio / pathwise estimator. The standard deviations of all of the estimators are compared in Figure 5.8. The picture is very

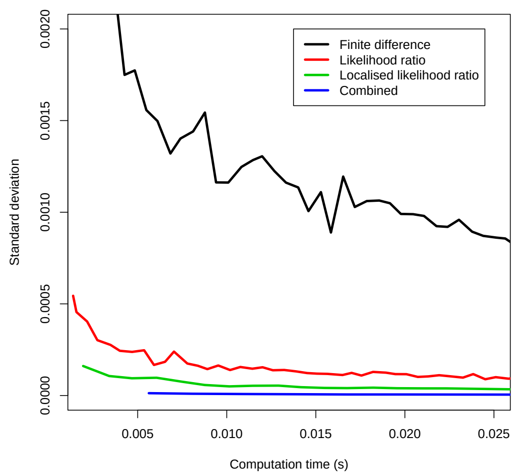

Figure 5.8: Comparison of standard deviations of Monte Carlo estimators for the delta of a digital option in the Black-Scholes model.

different from what we saw for the delta of the call. The likelihood ratio beats the finite difference method even for the non-localised case. Localisation only improves effectivity further, and the combination with antithetic and stratified sampling yields an estimator whose efficiency is a tremendous improvement over the plain finite difference estimator.

Method	Improvement factor
Call price	
Plain MC	1.0
Antithetic	1.2-1.8
Control variates	1.7-2.2
Stratified	7-12
Combined	14-22
OTM Call price	
Plain MC	1.0
Importance sampling	20-37
Combined	69-132
Call delta	
Plain finite difference	1.0
Pathwise	0.8-1.5
Likelihood ratio	0.3-0.5
Combined	9-14
Digital delta	
Plain finite difference	1.0
Likelihood ratio	5-10
Localised likelihood ratio	16-29
Combined	122-205

Table 5.1: Comparison of results for estimation of the call price, call delta and digital delta in the Black-Scholes model.

Conclusions. We have now analyzed various Monte Carlo methods for the price of a call, the delta of a call and the delta of a digital in the context of the Black-Scholes estimator. The overarching conclusion of our efforts is that when applied properly, using more than just plain Monte Carlo methods for pricing and plain finite difference for sensitivities can have a very large payoff.

Before proceeding to our numerical experiments for the Heston model, we compare the efficiency improvements for the various methods. Table 5.5 shows the factor of improvement for the standard deviations for the various methods implemented compared to the simplest methods. The factors are calculated as the minimal and maximal factors of improvement for the standard deviation for all computational times.

We see that in all cases, the addition of antithetic and stratified sampling adds considerably to the cost effectiveness of the estimators in terms of standard deviation per computing time. In particular, the combination of antithetic and stratified sampling with the localised likelihood ratio method for the digital delta yields almost unbelievable results. Our results show that this method allows one to calculate the digital delta with far more precision using the combined method with 2000 simulations than using the plain finite difference method with 90000 simulations.

5.6 Estimation in the Heston Model

The Heston model is a model for a single financial asset given by

$$\begin{aligned}
\mathrm{d}S_t &= \mu S(t)\,\mathrm{d}t + \sqrt{\sigma(t)}S(t)\sqrt{1-\rho^2}\,\mathrm{d}W^1(t) + \sqrt{\sigma(t)}S(t)\rho\,\mathrm{d}W_t^2 \\
\mathrm{d}\sigma_t &= \kappa(\theta-\sigma_t)\,\mathrm{d}t + \nu\sqrt{\sigma_t}\,\mathrm{d}W^2(t),
\end{aligned}$$

where (W^1, W^2) is a two-dimensional Brownian motion, μ, κ, θ and ν are positive parameters and $\rho \in (-1,1)$. We call this equation the Heston SDE for S and σ. We also assume given a risk-free asset B based on a constant short rate r. As in the previous section, our mission is to argue that the model is free of arbitrage and analyze different estimation methods in this model. In contrast to the previous section, it is not obvious that there exists a solution (S, σ) to the pair of SDEs describing the model. The question of arbitrage is also considerably more difficult than in the previous section. In order not to get carried away by too much theory, we will without proof apply some results from the theory of SDEs to obtain the results necessary for our purposes.

Theorem 5.6.1. *Let κ, θ and ν be positive numbers. Assume that $2\kappa\theta \geq \nu^2$. The SDE given by $\mathrm{d}\sigma_t = \kappa(\theta-\sigma_t)\,\mathrm{d}t + \nu\sqrt{\sigma_t^+}\,\mathrm{d}W_t$ has a weak solution, unique in law, and the solution is strictly positive. In particular, the solution also satisfies the SDE $\mathrm{d}\sigma_t = \kappa(\theta-\sigma_t)\,\mathrm{d}t + \nu\sqrt{\sigma_t}\,\mathrm{d}W_t$.*

Comment 5.6.2 The process solving the SDE is known as the Cox-Ingersoll-Ross process, from Cox, Ingersoll & Ross (1985). The criterion $2\kappa\theta \geq \nu^2$ ensures that the mean reversion level θ is sufficiently large in comparison to the volatility so that the variability of the solution cannot drive the process below zero. ∘

Proof. The existence of a positive weak solution is proved in Example 11.8 of Jacobsen

(1989). To prove uniqueness in law, we use the Yamada-Watanabe uniqueness theorem, Theorem V.40.1 of Rogers & Williams (2000b). This result states that the SDE is unique in law if the drift cofficient is Lipschitz and the volatility coefficient satisfies $(\sigma(x) - \sigma(y))^2 \leq \rho(|x-y|)$ for some $\rho : [0,\infty) \to [0,\infty)$ with $\int_0^\infty \frac{1}{\rho(x)}\,\mathrm{d}x$ infinite. In our case, the drift is $\mu(x) = \kappa(\theta - x)$, which is clearly Lipschitz as it has constant derivative. For the volatility, we have $\sigma(x) = \sqrt{x^+}$. We put $\rho(x) = x$ and note that for $0 \leq x \leq y$,

$$(\sigma(x) - \sigma(y))^2 = x - 2\sqrt{x}\sqrt{y} + y \leq y - x = \rho(|x-y|).$$

For $x, y \leq 0$, we clearly also have $(\sigma(x) - \sigma(y))^2 = 0 \leq \rho(|x-y|)$. Finally, in the case $x \leq 0 \leq y$ we find $(\sigma(x) - \sigma(y))^2 = y \leq \rho(|x-y|)$. We conclude that for any $x, y \in \mathbb{R}$, $(\sigma(x) - \sigma(y))^2 \leq \rho(|x-y|)$. Obviously, ρ satisfies the integrability criterion. Thus, the hypotheses of the Yamada-Watanabe theorem are satisfied, and we have uniqueness in law. □

With Theorem 5.6.1 in hand, we can conclude that the volatility process of the Heston model exists and is positive if $2\kappa\theta \geq \nu^2$. By positivity, we can also take its square root. Therefore, the SDE for the financial instrument in the Heston SDE is well-defined, and by Lemma 3.7.1, there exists a process S solving it. Because of these considerations, we shall in the following always assume $2\kappa\theta \geq \nu^2$.

Next, we consider the question of arbitrage. Because the volatility is stochastic, we cannot directly apply the criterions of Section 5.1. With a bit of work, however, we can use the results to obtain the desired conclusion.

Theorem 5.6.3. *The Heston model is free of arbitrage, and for any $T > 0$ there is a T-EMM Q yielding the dynamics on $[0, T]$ given by*

$$\begin{aligned} \mathrm{d}S_t &= rS(t)\,\mathrm{d}t + \sqrt{\sigma(t)}S(t)\sqrt{1-\rho^2}\,\mathrm{d}W^1(t) + \sqrt{\sigma(t)}S(t)\rho\,\mathrm{d}W_t^2 \\ \mathrm{d}\sigma_t &= \kappa(\theta - \sigma_t)\,\mathrm{d}t + \nu\sqrt{\sigma_t}\,\mathrm{d}W^2(t), \end{aligned}$$

that is, the dynamics for the volatility is preserved, and the drift for the instrument is changed from μ to r.

Proof. We wish to use Corollary 5.1.8 to prove freedom from arbitrage. We therefore need to analyze the equation

$$\mu - r = \sqrt{\sigma_t}\sqrt{1-\rho^2}\lambda_1(t) + \sqrt{\sigma_t}\rho\lambda_2(t).$$

To solve this, we put $\lambda_2 = 0$. The equation reduces to $\mu - r = \sigma_t\sqrt{1-\rho^2}\lambda(t)$, where we for notational simplicity have removed the subscript from the lambda. Since σ is positive, this equation has a unique solution given by

$$\lambda_t = \frac{\mu - r}{\sqrt{1-\rho^2}\sqrt{\sigma_t}}.$$

Since σ is a standard process, it is clear that λ is progressive. By continuity, λ is locally bounded and therefore in $\mathfrak{L}^2(W)$. By Corollary 5.1.8, if we can prove that $\mathcal{E}(M)$ is a martingale with $M_t = -\int_0^t \lambda_t \,\mathrm{d}W_t^1$, the market is free of arbitrage. It is then clear that the Q-dynamics of S and σ are as stated in the theorem.

To this end, it will by Lemma 3.8.6 suffice to show that $E\mathcal{E}(M)_t = 1$ for all $t \geq 0$. Let $t \geq 0$ be given, let $\tau_n = \inf\{s \geq 0 || \lambda_s| \geq n\}$. Since λ is continuous, τ_n increases to infinity and λ^{τ_n} is bounded. Then $[M^{\tau_n \wedge t}]_\infty = \int_0^t (\lambda_s^{\tau_n})^2 \,\mathrm{d}s$ is bounded, so the Novikov criterion applies to show that $\mathcal{E}(M^{\tau_n \wedge t})$ is a uniformly integrable martingale for any $t \geq 0$.

Next, note that since τ_n increases, $\tau_n \geq t$ from a point onwards and $\mathcal{E}(M)$ is non-negative, the variables $\mathcal{E}(M)_t 1_{(\tau_n \geq t)}$ converge upwards to $\mathcal{E}(M)_t$. By monotone convergence, $E\mathcal{E}(M)_t = \lim_n E\mathcal{E}(M)_t 1_{(\tau_n \geq t)}$. But $\mathcal{E}(M)_t 1_{(\tau_n \geq t)} = \mathcal{E}(M^{\tau_n \wedge t})_\infty 1_{(\tau_n \geq t)}$, so letting Q^n be the measure such that $\frac{\mathrm{d}Q^n_\infty}{\mathrm{d}P_\infty} = \mathcal{E}(M^{\tau_n \wedge t})_\infty$, we may now conclude

$$E\mathcal{E}(M)_t = \lim_n E\mathcal{E}(M)_t 1_{(\tau_n \geq t)} = \lim_n E\mathcal{E}(M^{\tau_n \wedge t})_\infty 1_{(\tau_n \geq t)} = \lim_n Q^n(\tau_n \geq t).$$

We need to show that the latter limit is equal to one. To do so, we first show that $Q^n(\tau_n \geq t) = P(\tau_n \geq t)$. Under Q^n, we know from Girsanov's theorem that the process $(W_t^1 + \int_0^t \lambda_s 1_{[0,\tau_n \wedge t]}(s)\,\mathrm{d}s, W_t^2)$ is a $\mathcal{F}_t$ Brownian motion. In particular, W^2 is a $\mathcal{F}_t$ Brownian motion under Q^n for all n. Now, we know that under P, σ satisfies

$$\sigma_t = \sigma_0 + \int_0^t \kappa(\theta - \sigma_s)\,\mathrm{d}s + \int_0^t \nu\sqrt{\sigma_s}\,\mathrm{d}W_s^2.$$

By Theorem 3.8.4, integrals under P and Q^n are the same. Since $\nu\sqrt{\sigma_s}$ is continuous, it is integrable both under P and Q_n, and therefore σ also satisfies the above equation under the set-up with the Q^n $\mathcal{F}_t$ Brownian motion W^2. Since the SDE satisfies uniqueness in law by Theorem 5.6.1, the Q_n distribution of σ must be the same as the P distribution. Since λ is a transformation of σ and τ_n is a transformation of λ, we conclude that the distribution of τ_n is the same under Q_n and P. In particular, $Q^n(\tau_n \geq t) = P(\tau_n \geq t)$. Since τ_n tends almost surely to infinity, $\lim_n P(\tau_n \geq t) = 1$. Combining this with our earlier findings, we conclude $E\mathcal{E}(M)_t = 1$. Therefore, by Lemma 3.8.6, $\mathcal{E}(M)$ is a martingale and by Corollary 5.1.8, the market is free of arbitrage. □

We will now discuss estimating the digital delta in the Heston model. Fix a maturity time T. Our first objective is to find out how to simulate values from S_T. In the Black-Scholes model, this was a trivial question as the stock price then followed a scaled lognormal distribution. For the Heston model, there is no simple way to simulate from S_T, we need to discretise the SDE. Before considering how to do this, we will rewrite the SDE to a form better suited for discretisation. Define $X_t = \log S_t$, Itô's lemma in the form of Lemma 3.6.4 yields

$$\begin{aligned}
\mathrm{d}X_t &= \frac{1}{S_t}\,\mathrm{d}S_t - \frac{1}{2S_t^2}\,\mathrm{d}[S]_t \\
&= r\,\mathrm{d}t + \sqrt{\sigma_t}\sqrt{1-\rho^2}\,\mathrm{d}W_t^1 + \sqrt{\sigma_t}\rho\,\mathrm{d}W_t^2 - \frac{1}{2}\sigma_t\,\mathrm{d}t,
\end{aligned}$$

with $X_0 = \log S_0$. In other words, in order to simulate values of S_T, it will suffice to consider the SDE

$$\begin{aligned}
\mathrm{d}X_t &= \left(r - \frac{1}{2}\sigma_t\right)\mathrm{d}t + \sqrt{\sigma_t}\sqrt{1-\rho^2}\,\mathrm{d}W_t^1 + \sqrt{\sigma_t}\rho\,\mathrm{d}W_t^2 \\
\mathrm{d}\sigma_t &= \kappa(\theta - \sigma_t)\,\mathrm{d}t + \nu\sqrt{\sigma_t}\,\mathrm{d}W^2(t),
\end{aligned}$$

find a way to simulate from X_T and then take the exponent of X_T to obtain a simulation from S_T. A discretisation of the above set of SDEs will yield a way to simulate from a distribution approximating that of X_T. Such a discretisation consists of defining discrete processes $X'_{n\Delta}$ and $\sigma'_{n\Delta}$ for $n \le N$ with $N\Delta = T$ such that the distribution of $X'(T)$ approximates that of $X(T)$. There are several ways to do this. Andersen (2008) provides a useful overview. We will consider two methods: The full truncation modified Euler scheme and a scheme with exact simulation of the volatility. For brevity, we will denote the full truncation modified Euler scheme as the FT scheme, short for full truncation, and we will call the other scheme the SE scheme, short for semi-exact.

The FT scheme. We first describe the modified Euler scheme. In Lord et al. (2006), several ways of discretising the CIR process are considered, and the FT scheme is found to be the one introducing the least bias. Let Z^1 and Z^2 be discrete processes on $\{0, \Delta, \ldots, N\Delta\}$ of independent standard normal variables, the FT scheme is the discretisation defined by putting $X'(0) = X_0$, $\sigma'(0) = \sigma_0$ and

$$\begin{aligned}
X'_{(n+1)\Delta} &= X'_{n\Delta} + \left(r - \frac{1}{2}\sigma'_{n\Delta}\right)\Delta + \sqrt{\sigma'^{+}_{n\Delta}}\sqrt{1-\rho^2}\sqrt{\Delta}Z^1_{n\Delta} + \sqrt{\sigma'^{+}_{n\Delta}}\rho\sqrt{\Delta}Z^1_{n\Delta} \\
\sigma'_{(n+1)\Delta} &= \sigma'_{n\Delta} + \kappa(\theta - \sigma'^{+}_{n\Delta}) + \nu\sqrt{\sigma'^{+}_{n\Delta}}\sqrt{\Delta}Z^2_{n\Delta}.
\end{aligned}$$

The discretisation means that the volatility process σ' can turn negative. In this case, there is an upwards drift of rate $\kappa\theta$. Taking positive parts ensure that the square roots are well-defined.

The SE scheme. The SE scheme takes advantage of the fact that the transition probability of the CIR process is a scaled noncentral χ^2 distribution. More explicitly, the distribution of $\sigma_{t+\Delta}$ given $\sigma_\Delta = v$ is a scaled noncentral χ^2 distribution with scale $\frac{\nu^2}{4\kappa}(1-e^{-\kappa\Delta})$, non-centrality parameter $\frac{4v\kappa e^{-\kappa\Delta}}{\nu^2(1-e^{-\kappa\Delta})}$ and degrees of freedom $\frac{4\theta\kappa}{\nu^2}$. That this is the case is a folklore result, although it is quite unclear where to find a proof of the result. In Feller (1951), Lemma 9, the Fokker-Planck equation for the SDE is solved, yielding the density of the noncentral χ^2 distribution. It would then be reasonable to expect that one might obtain a complete proof by using some results on when the solution to the Fokker-Planck equation in fact is the transition density of the solution. In any case, the result allows exact simulation of values of the CIR process. Thus, we can define σ' by letting the conditional distribution of $\sigma'_{(n+1)\Delta}$ given $\sigma'_{n\Delta}$ be the noncentral χ^2 distribution defined above. This defines a discrete approximation of σ. We will use this to obtain a discrete approximation of the price process. To do so, we first note that

$$X_{t+\Delta} - X_t = r\Delta - \frac{1}{2}\int_t^{t+\Delta} \sigma_u \,\mathrm{d}u + \sqrt{1-\rho^2}\int_t^{t+\Delta} \sqrt{\sigma_u}\,\mathrm{d}W_u^1 + \rho\int_t^{t+\Delta} \sqrt{\sigma_u}\,\mathrm{d}W_u^2.$$

The process W^2 is correlated with σ. Since we desire to simulate directly from the transition probabilities for σ, our discretisation scheme cannot depend on this correlation. To remove the $\mathrm{d}W^2$ integral, we note that

$$\sigma_{t+\Delta} - \sigma_t = \int_t^{t+\Delta} \kappa(\theta - \sigma_u)\,\mathrm{d}u + \nu\int_t^{t+\Delta} \sqrt{\sigma_u}\,\mathrm{d}W_u^2.$$

Isolating the last integral and substituting, we find

$$\begin{aligned}
& X_{t+\Delta} - X_t \\
= \quad & r\Delta - \int_t^{t+\Delta} \frac{\sigma_u}{2} + \frac{\rho\kappa}{\nu}(\theta - \sigma_u)\,\mathrm{d}t + \sqrt{1-\rho^2}\int_t^{t+\Delta} \sqrt{\sigma_u}\,\mathrm{d}W_u^1 + \frac{\rho(\sigma_{t+\Delta} - \sigma_t)}{\nu} \\
= \quad & r\Delta - \frac{\rho\kappa\theta\Delta}{\nu} + \frac{\rho(\sigma_{t+\Delta} - \sigma_t)}{\nu} + \left(\frac{\rho\kappa}{\nu} - \frac{1}{2}\right)\int_t^{t+\Delta} \sigma_u\,\mathrm{d}u + \sqrt{1-\rho^2}\int_t^{t+\Delta} \sqrt{\sigma_u}\,\mathrm{d}W_u^1.
\end{aligned}$$

The right-hand side depends only on the joint distributions of σ_u and W^1. Since σ_u and W^1 are independent, this is an important simplification compared to our earlier expression, which included an integral with respect to W^2. Because of the independence, the conditional distribution of $\int_t^{t+\Delta} \sqrt{\sigma_u}\,\mathrm{d}W_u^1$ given σ is normal with mean zero and variance $\int_t^{t+\Delta} \sigma_u \,\mathrm{d}u$. We now make the approximations

$$
\begin{aligned}
\int_t^{t+\Delta} \sigma_u \,\mathrm{d}u &\approx \sigma_{t+\Delta}\frac{\Delta}{2} + \sigma_t \frac{\Delta}{2} \\
\int_t^{t+\Delta} \sqrt{\sigma_u}\,\mathrm{d}W_u^1 &\approx \sqrt{\sigma_{t+\Delta}\frac{\Delta}{2} + \sigma_t \frac{\Delta}{2}}\frac{1}{\sqrt{\Delta}}(W_{t+\Delta}^1 - W_t^1).
\end{aligned}
$$

Inserting these in our equation for $X_{t+\Delta}$, we obtain

$$
\begin{aligned}
X_{t+\Delta} &\approx X_t + r\Delta - \frac{\rho\kappa\theta\Delta}{\nu} + \frac{\rho(\sigma_{t+\Delta} - \sigma_t)}{\nu} \\
&+ \left(\frac{\rho\kappa}{\nu} - \frac{1}{2}\right)\frac{\Delta(\sigma_{t+\Delta} + \sigma_t)}{2} + \sqrt{1-\rho^2}\sqrt{\frac{\sigma_{t+\Delta} + \sigma_t}{2}}(W_{t+\Delta}^1 - W_t^1) \\
&= X_t + \left(r - \frac{\rho\kappa\theta}{\nu}\right)\Delta + \left(\frac{\Delta}{2}\left(\frac{\rho\kappa}{\nu} - \frac{1}{2}\right) - \frac{\rho}{\nu}\right)\sigma_t \\
&+ \left(\frac{\Delta}{2}\left(\frac{\rho\kappa}{\nu} - \frac{1}{2}\right) + \frac{\rho}{\nu}\right)\sigma_{t+\Delta} + \sqrt{\frac{(1-\rho^2)(\sigma_{t+\Delta} + \sigma_t)}{2}}(W_{t+\Delta}^1 - W_t^1).
\end{aligned}
$$

We are thus lead to defining the discretisation scheme X' for X by

$$
X'_{(n+1)\Delta} = X'_{n\Delta} + K_0 + K_1\sigma'_{n\Delta} + K_2\sigma'_{(n+1)\Delta} + \sqrt{K_3(\sigma'_{(n+1)\Delta} + \sigma'_{n\Delta})}Z_n,
$$

where Z_n is a process of independent standard normal variables on $\{0, \Delta, \ldots, N\Delta\}$ independent of σ' and

$$
\begin{aligned}
K_0 &= \left(r - \frac{\rho\kappa\theta}{\nu}\right)\Delta \\
K_1 &= \frac{\Delta}{2}\left(\frac{\rho\kappa}{\nu} - \frac{1}{2}\right) - \frac{\rho}{\nu} \\
K_2 &= \frac{\Delta}{2}\left(\frac{\rho\kappa}{\nu} - \frac{1}{2}\right) + \frac{\rho}{\nu} \\
K_3 &= \frac{\Delta(1-\rho^2)}{2}.
\end{aligned}
$$

Modulo a short rate and some weights in the discretisation of the integrals, this is the same as obtained in Andersen (2008), page 20. To sum up, the SE scheme defines the approximation σ' to σ by directly defining the conditional distributions of the process using the transition probabilities. The approximation X' to X is then obtained by

$$
X'_{(n+1)\Delta} = X'_{n\Delta} + K_0 + K_1\sigma'_{n\Delta} + K_2\sigma'_{(n+1)\Delta} + \sqrt{K_3(\sigma'_{(n+1)\Delta} + \sigma'_{n\Delta})}Z_n,
$$

where Z_n is a process of independent standard normal variables on $\{0, \Delta, \ldots, N\Delta\}$, independent of σ'. Note that while the FT scheme simulates a discretisation of the two-dimensional process (W^1, W^2), the SE scheme simulates only a discretisation of

W^1. This will be an important point when comparing our Malliavin estimator with the one obtained in Benhamou (2001). Also note that because of the exact simulation of the root volatility in the SE scheme, σ' can never reach zero, but in the FT scheme, it is quite probable that σ' hits zero. This will be important for the implementation of the Malliavin estimator.

Performance of the schemes. We will now see how well each of these two schemes can approximate the true distribution of S_T. As we have no way to obtain exact simulations from the distribution, we must contend ourselves with checking convergence and compare results from each scheme. We will use $S_0 = 100$, $\sigma_0 = 0.04$, $r = 0.04$, $\kappa = 2$, $\theta = 0.04$, $\nu = 0.4$, $\rho = 0.3$ and $T = 1$ and these parameter values will be held constant throughout all of this section unless stated otherwise. Figure 5.9 shows density estimates of S_T for each of the two schemes, using 2, 5, 50 and 100 steps. The figure shows that both schemes seem to converge to some very similar distributions. Most of the error in the discretisations seem to come from the top shape of the hump, with the SE scheme converging quicker than the FT scheme. This is not surprising, considering that the SE scheme uses the exact distribution for the volatility, while the FT scheme only has an approximation. Both the schemes seem to have achieved stability at least at the 50-step discretisation.

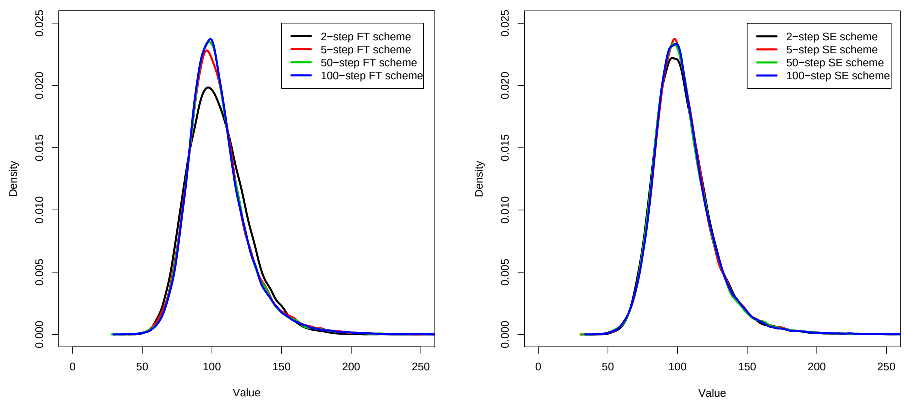

Figure 5.9: Density estimates for the FT and SE schemes based on 80000 simulations.

Next, we compare call pricing under each of the schemes. In Heston (1993), an analytical expression for the Fourier transform of the survival function of S_T is obtained, and it is explained how this can be inverted and used to obtain call prices in the Heston model. These results greatly contribute to the popularity of the Heston model and its offspring. We use the pricer at http://kluge.in-chemnitz.de/tools/pricer to obtain accurate values for call prices. This pricer seems to be reliable and is also used for the numerical experiments in Alòs & Ewald (2008). Using strike 104 yields a call price of 7.71175.

We want to see how the call prices obtained using Monte Carlo with each of the two schemes compare to this. The purpose of this is both to check how much bias is introduced by the discretisation and to check that our fundamental simulation routines work as they are supposed to. We will at the same time also check the functionality of sampling from each of the schemes using latin hypercube sampling and antithetic sampling. We have explained earlier how antithetic sampling works, but we have not mentioned latin hypercube sampling. We will not explain this method in detail, suffice it to say that it is can be thought of as a generalization of stratified sampling for high-dimensional problems, see Glasserman (2003), Section 4.4.

In the left graph of Figure 5.10, we compare the relative errors of call prices for each of the two schemes with and without latin hypercube and antithetic sampling, and we compare the standard deviations. We use discretisations steps from 10 to 100 and consider 300 replications of the estimator, each estimate based on 20000 samples. We see that as expected, each of the methods yield prices which fit well with the true value. It seems that less than 50 steps yields some inaccuracy in excess of the expected variability, with the 10-step results giving bias which is significantly larger than those of the observations for higher levels of discretisation. However, the relative error never exceeds 1%. In reality, models are inaccurate anyway, so this bias would be unimportant in practical settings. In other words, for practical applications, it would seem that 20 steps would by sufficient for the contract under investigation.

In the right graph, we compare standard deviation times computation time. This is a general measure of efficiency, with low values corresponding to high efficiency. We see that the efficiency decreases as the number of discretisation steps increases. This is not particularly surprising, but an important point nonetheless: The higher level of discretisation means that each sample from S_T takes longer to generate, but since the number of samples is held constant and the change in the distribution is minuscule, the standard deviation does not decrease. Thus, the standard deviation times computational time will increase as the number of discretisation steps increases, lowering performance. Therefore, the choice of number of discretisation steps boils down to choosing between optimising bias and optimising standard deviation times computational time.

Note that the right graph cannot be used to measure the efficiency of latin hypercube and antithetic sampling, as the computational time is not held constant. However, experimental results show an improvement factor of around 1.4. This is very moderate compared to the improvement factors found for antithetic and stratified sampling in the previous section. The reason for this is probably that the transformation from the driving normal distributions to the final S_T sample is very complex, and the latin hypercube sampling basically just stratifies in a very simple manner. To obtain the same type of efficiency gains for the Heston model as in the Black-Scholes model, it will probably be necessary to implement a more sophisticated form of stratified sampling, such as terminal stratification or one of the other methods discussed in Subsection 4.3.2 of Glasserman (2003).

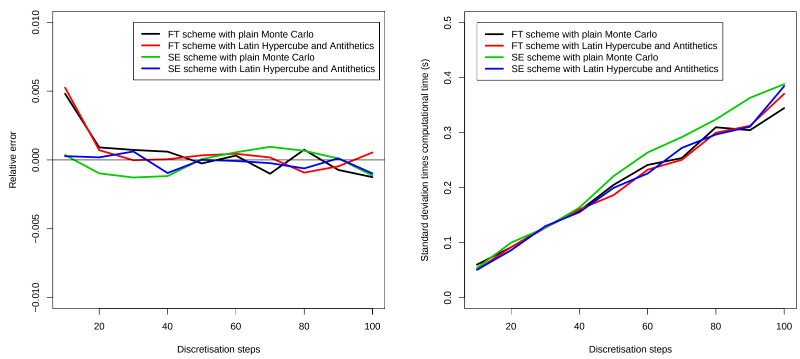

Figure 5.10: Left: Relative errors for call prices calculated using the FT and SE schemes with and without latin hypercube sampling and antithetic sampling. Right: Standard deviations times computational time.

Comparison of methods for calculating the digital delta. Being done with the introductory investigations of the Heston model, we now proceed to the main point of this section, which is the comparison of the Malliavin and localised Malliavin methods with the finite difference method. Note that we cannot directly compare these methods with the likelihood ratio and pathwise differentiation methods as in the previous section, as the likelihood ratio method does not apply directly to the Heston model (see, however, the arguments outlined in Glasserman (2003), page 414 or Chen & Glasserman (2006), for application of the likelihood ratio method to the discretisation of the Heston model) and the pathwise differentiation method does not apply to the digital delta. In the remainder of this section, we will always sample from our two schemes using latin hypercube sampling and antithetic sampling.

We consider a digital option with strike 104. Let f denote the corresponding payoff function. As described above, we have six methods to consider: The finite difference method, the Malliavin method and the localised Mallivin method, each combined with one of the two simulation schemes. In Section 5.4, we found the basis of the Malliavin method through the identity

$$\frac{\partial}{\partial S_0} Ef(S_T) = E\left(f(S_T) \frac{1}{S_0 T} \int_0^T \frac{1}{\sqrt{\sigma_t}\sqrt{1-\rho^2}} \, \mathrm{d}W_t^1 \right),$$

where we are using $\sqrt{\sigma_t}$ instead of σ_t because in our model formulation, σ_t is the root volatility and not the ordinary volatility. The localised Malliavin method is then obtained by decomposing the payoff in two parts and applying the Malliavin method to one part and the pathwise differentiation method to the other, as was also done in Section 5.5. One problem with this is that we do not know how to sample from the stochastic integral in the expectation above. Instead, we have to make to with an approximation. Based on the theory of Chapter 3, we would expect that a reasonable approximation of an integral of the form $\int_0^T H_s \, \mathrm{d}W_s^1$ is

$$\int_0^T H_s \, \mathrm{d}W_s^1 = \sum_{k=1}^n H_{t_{k-1}}(W_{t_k} - W_{t_{k-1}}),$$

where it is essential that we are using forward increments and not backwards increments in the Riemann sum. With X' and σ' denoting the discretised versions of X and σ and Z^1 denoting the driving Brownian motion for the asset price process, it is then natural to consider the approximation

$$\frac{\partial}{\partial S_0} Ef(S_T) \approx E\left(f\left(e^{X'_T}\right) \frac{1}{e^{X'_0} T\sqrt{1-\rho^2}} \sum_{k=1}^N \frac{\sqrt{\Delta} Z_{k\Delta}}{\sqrt{\sigma'_{(k-1)\Delta}}} \right).$$

This works well when using the SE scheme, where we know that σ' is positive. However, if we are using the FT scheme, we have no such guarantee, and the above is not well-defined. In this case, then, we resort to substituting $\max\{\sigma'_{(k-1)\Delta}, \varepsilon\}$ for $\sigma'_{(k-1)\Delta}$ in the sum, where ε is some small positive number. Our experimental results show that using an ε which is too small such as 0.00000001 can lead to considerable numerical instability. We use 0.0001, this leads to stable results and reasonable efficiency.

We are now ready for the numerical experiments. We begin by investigating the bias of our estimators. We expect two sources of bias. First, there is an inherent bias from the sampling of S_T since we are only sampling from a distribution approximating that of S_T. Second, each of the methods involve certain approximations - the finite difference method approximates the derivative by a finite difference, and the Malliavin estimators approximate the Itô integral by a sum.

In Figure 5.11, we have plotted the means for each of the six methods. We use 300 replications of estimators based on samples of size 1000 to 29500 with increments of 1500. We repeat the experiment for a 15-step and a 100-step discretisation. We see that in the 15-step experiment, there is a considerable variation in the means, but this variation is almost gone in the 100-step experiment. Expecting that the finite difference method produces the least bias, we obtain from the experiment an approximate value for the digital delta of 0.02132436, we will use this value as a benchmark. Note that we could also have used the semi-analytical results from Heston (1993) to obtain benchmark values, however, the results from the finite difference method will be sufficient for our needs. Using this benchmark, we find that the Malliavin methods in the 15-step experiment yields average relative errors of -6.8% and 7.7% percent for the FT and SE schemes, respectively. The localised methods have average relative errors of -3.3% and 3.7%, while the finite difference relative errors are less than 0.1%. In the 100-step experiment, the Malliavin methods have average relative errors of -0.8% and 1.9% for the FT and SE schemes, respectively, while the localised Malliavin methods have average relative errors of -0.3% and 0.9%, respectively. We conclude that it would seem that all of our estimators are asymptotically unbiased.

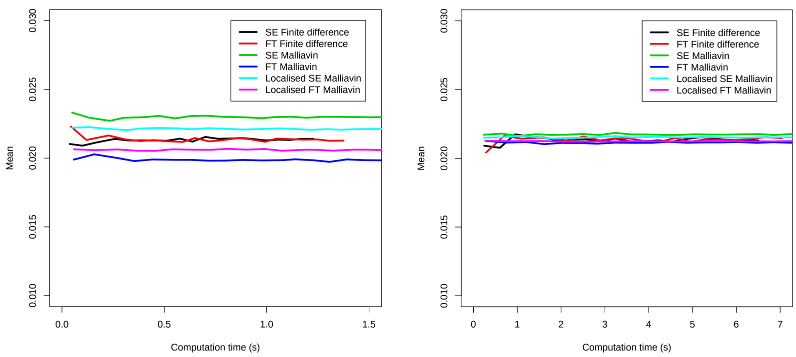

Figure 5.11: Estimated estimator means for estimating the digital delta. 15-step discretisation on the left, 100-step discretisation on the right.

Next, we consider standard deviation. The results can be found in Figure 5.12. For both the 15-step and 100-step experiments and both the FT and SE schemes, we see that the Malliavin method is superior to the finite difference method and the localised Malliavin method is superior to the ordinary Malliavin method.

For the 15-step experiment, the SE scheme yields the better performance, while the situation is reversed for the 100-step experiment. The explanation for this probably has something to do with the FT scheme being faster than the SE scheme since it is faster to generate normal distributions than noncentral χ^2 distributions. Thus, for fine discretisations, the resulting distributions of S_T are nearly the same, but the FT scheme is computationally faster.

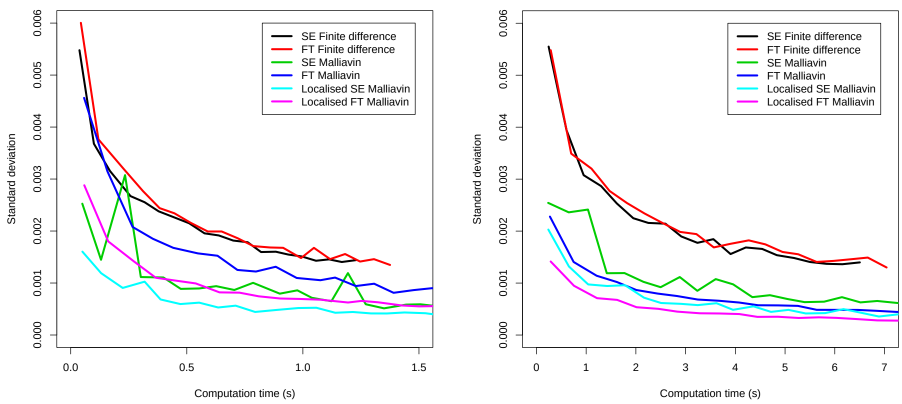

Figure 5.12: Estimated estimator standard deviations for estimating the digital delta. 15-step discretisation on the left, 100-step discretisation on the right.

The difference between our results and those of Benhamou (2001). We mentioned in Section 5.4 that the expression we have obtained for the Malliavin method is not the same as is obtained in Benhamou (2001). We will now explain what exactly the difference is and investigate how this difference manifests itself in practice.

We will first detail the model description and corresponding Malliavin estimator used by ourselves and by Benhamou (2001). Our description of the Heston model and the corresponding Malliavin estimator is

$$\begin{aligned}
\mathrm{d}S_t &= \mu S(t)\,\mathrm{d}t + \sqrt{\sigma(t)}S(t)\sqrt{1-\rho^2}\,\mathrm{d}W^1(t) + \sqrt{\sigma(t)}S(t)\rho\,\mathrm{d}W_t^2 \\
\mathrm{d}\sigma_t &= \kappa(\theta-\sigma_t)\,\mathrm{d}t + \nu\sqrt{\sigma_t}\,\mathrm{d}W^2(t) \\
\frac{\partial}{\partial S_0}Ef(S_T) &= E\left(f(S_T)\frac{1}{S_0T}\int_0^T \frac{1}{\sqrt{\sigma_t}\sqrt{1-\rho^2}}\,\mathrm{d}W_t^1\right),
\end{aligned}$$

where (W^1, W^2) is a two-dimensional Brownian motion. In contrast to this, Benhamou (2001) list the model description and Malliavin estimator as

$$\begin{aligned}
\mathrm{d}S_t &= \mu S(t)\,\mathrm{d}t + \sigma(t)S(t)\,\mathrm{d}B^1(t) \\
\mathrm{d}\sigma_t^2 &= \kappa(\theta-\sigma_t^2)\,\mathrm{d}t + \nu\sigma_t\,\mathrm{d}B^2(t) \\
\frac{\partial}{\partial S_0}Ef(S_T) &= E\left(f(S_T)\frac{1}{S_0T}\int_0^T \frac{1}{\sigma_t}\,\mathrm{d}B_t^1\right),
\end{aligned}$$

where (B^1, B^2) is a two-dimensional Brownian motion with correlation ρ. Rewriting this with our model formulation, we obtain

$$\begin{aligned}
\mathrm{d}S_t &= \mu S(t)\,\mathrm{d}t + \sqrt{\sigma(t)}S(t)\sqrt{1-\rho^2}\,\mathrm{d}W^1(t) + \sqrt{\sigma(t)}S(t)\rho\,\mathrm{d}W_t^2 \\
\mathrm{d}\sigma_t &= \kappa(\theta-\sigma_t)\,\mathrm{d}t + \nu\sqrt{\sigma_t}\,\mathrm{d}W^2(t) \\
\frac{\partial}{\partial S_0}Ef(S_T) &= E\left(f(S_T)\frac{1}{S_0T}\left(\sqrt{1-\rho^2}\int_0^T \frac{1}{\sqrt{\sigma_t}}\,\mathrm{d}W_t^1 + \rho\int_0^T \frac{1}{\sqrt{\sigma_t}}\,\mathrm{d}W_t^2\right)\right).
\end{aligned}$$

Obviously, there seems to be some difference here. The two expressions for the Malliavin estimator are equal in the case of zero correlation, but different whenever there is nonzero correlation. We will now see how this difference manifests itself when calculating the digital delta. We will compare the results of the estimator of Benhamou (2001) with our estimators and the finite difference estimator. Since the estimator of Benhamou (2001) involves an integral with respect to both W^1 and W^2, we cannot use the SE scheme with this method. We therefore restrict ourselves to comparing the Malliavin estimator of Benhamou (2001) under the FT scheme with our Malliavin estimator under both the FT and SE schemes. Finally, we also consider the finite difference method under the SE scheme. Since we have already seen that the finite difference method is quite insensitive to the choice of discretisation scheme, it should be unnecessary to consider also the finite difference method under the FT scheme.

We compare the digital delta estimators when the correlation varies between -0.95 and 0.95. We use 300 replications of estimators based on 20000 simulations each. The results can be seen in the left graph of Figure 5.13. As expected, the results from

Benhamou (2001) fits with our results in the case of zero correlation. However, it is quite obvious that when the correlation is nonzero, in particular when there is very strong positive or negative correlation, the estimator from Benhamou (2001) does not fit with the otherwise very reliable finite difference estimator, whereas our Malliavin estimator works well.

To do a more formal check that there is an actual difference, we have in the right graph of Figure 5.13 plotted density estimates for the finite difference estimator and the Benhamou estimator for the case of correlation -0.95. The thin lines are moment-matched normal densities. We see that the estimators are approximately normal, with the finite difference estimator having some minor fluctuations in the right tail. It would therefore be reasonable to use the Welch t-test for testing equality of means for normal distributions with unequal variances to test whether the two estimators have different means. The p-value of the test comes out less than 10^{-15}. For comparison, when using the same test to compare our FT scheme based Malliavin estimator with the finite difference estimator, we obtain a p-value of 71%. This strongly suggests that even when taking the bias from the approximation to the Itô integral in the Benhamou estimator into account, the Benhamou estimator does not yield the correct value of the digital delta.

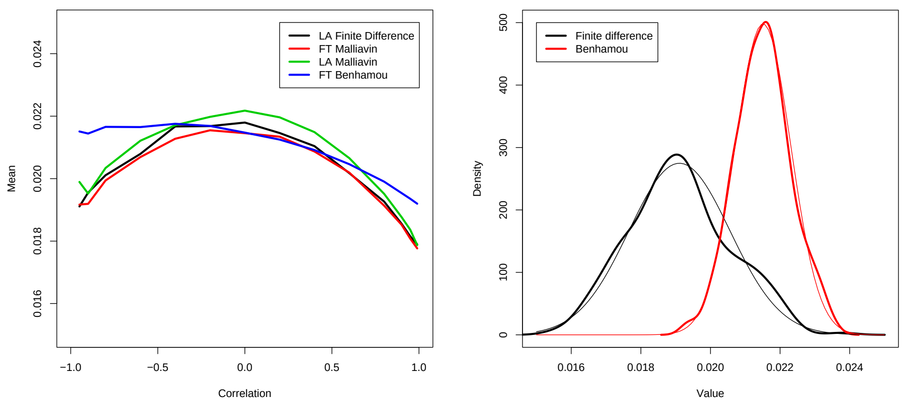

Figure 5.13: Left: Comparison of the estimator from Benhamou (2001) with our Malliavin estimators and a finite difference estimator. Right: Density plots for the finite difference and Benhamou estimators for correlation -0.95. Thick lines are density estimates, thin lines are normal approximations.

Method	Improvement factor
15-step digital delta	
FT scheme, finite difference	1.0
SE scheme, finite difference	0.9-1.1
FT scheme, Malliavin	1.0-1.4
SE scheme, Malliavin	1.0-2.5
FT scheme, localised Malliavin	1.6-2.4
SE scheme, localised Malliavin	2.6-3.8
100-step digital delta	
FT scheme, finite difference	1.0
SE scheme, finite difference	0.9-1.1
FT scheme, Malliavin	2.3-3.1
SE scheme, Malliavin	1.3-2.4
FT scheme, localised Malliavin	3.4-5.1
SE scheme, localised Malliavin	2.7-3.8

Table 5.2: Comparison of results for estimation of the digital delta in the Heston model.

Conclusions. We have considered the digital delta for three types of estimators, each combined with two different discretisation schemes. As in the previous section, we conclude that methods more exotic than just plain Monte Carlo with finite difference can make a considerable difference. Table 5.5 shows the factor of improvement for the standard deviations for the various methods implemented compared to the simplest methods.

We see that the Malliavin and localised Malliavin methods indeed yield an improvement, but the improvement is somewhat disappointing compared to the results we obtained for the Black-Scholes model, where the likelihood ratio and localised likelihood ratio methods yielded improvement factors for the digital delta of respectively 5 to 10 and 16 to 29. It is difficult to say why the improvement factors are so small in this case. The Malliavin method requires no extra simulations, only the calculation of the approximation to the Itô integral, which is merely a simple sum.

We also again note that while the SE scheme is most effective for the 15-step case, the FT scheme is most effective for the 100-step scheme. Now, the localised Malliavin method for the 15-step case yielded a bias of up to 3.7%, which may be a little much. One might get more satisfactory results using, say, a 30-step discretisation. In this case, the SE scheme is probably still the most effective. We conclude that for practical purposes, the SE scheme should be preferred above the FT scheme.

5.7 Notes

The main references for the material in Section 5.1 are Björk (2004) and Øksendal (2005), Chapter 12. Other useful resources are Karatzas & Shreve (1998) and Steele (2000). Our results are not as elegant as they could be because of the limited theory available to us. In general, the result of Theorem 5.1.5 on the existence of EMMs and freedom from arbitrage goes both ways, in the sense that under certain regularity conditions and appropriate concepts of freedom from arbitrage, a market is free from arbitrage if and only if there exists an EMM. For a precise result, see for example Delbaen & Schachermayer (1997), Theorem 3. The theory associated with the necessity of the existence of EMMs in order to preclude arbitrage is rather difficult and is presented in detail in Delbaen & Schachermayer (2008). The simplest introduction to the field seems to be Levental & Skorohod (1995).

All of the Monte Carlo methods introduced in Section 5.3 and Section 5.4 except the results on the Malliavin method are described in the excellent Glasserman (2003). Rigorous resources on the Malliavin calculus applied to calculation of risk numbers are hard to come by. Even today, the paper Fournié et al. (1999) seems to be the best, because of its very clear statement of results. Di Nunno et al. (2008) is also refreshingly explicit. Other papers on the applications of the Malliavin calculus for calculations of risk numbers are Fournié et al. (2001), which extends the investigations started in Fournié et al. (1999), and also Benhamou (2001) and Benhamou (2003). A very interesting paper is Chen & Glasserman (2006), which proves, using results from the theory of weak convergence of SDEs, that the Malliavin method in a reasonably large class of cases is equivalent to the limit of an average of likehood ratio estimators for Euler discretisations of the SDEs under consideration. Montero & Kohatsu-Higa (2002) provides an informal overview of the role of Malliavin calculus in the calculation of risk numbers. Nualart (2006), Chapter 6, also provides some resources.

Other applications of the Malliavin calculus in finance such as the calculation of conditional expectations, insider trading and explicit hedging strategies are discussed in Schröter (2007), Di Nunno et al. (2008), Nualart (2006) and Malliavin & Thalmaier (2005). Here, Schröter (2007) is very informal, and Malliavin & Thalmaier (2005) is extremely difficult to read.

Much of the SDE theory used in the derivation of the Malliavin estimator in Section 5.4 and Fournié et al. (1999) is described in Karatzas & Shreve (1988) and Rogers & Williams (2000b). Readable results on differentiation of flows is hard to find. The classical resource is Kunita (1990). A seemingly more readable proof of the main result can be found in Ikeda & Watanabe (1989), Proposition 2.2 of Chapter V.

Several of the numerical results of Section 5.5 can be compared with those of Fournié et al. (1999) and Benhamou (2001). Glasserman (2003) is, as usual, a useful general guide to the effectivity of Monte Carlo methods.

The proof in Section 5.6 that the Heston model does not allow arbitrage is based on the techniques of Cheridito et al. (2005), specifically the proof of Theorem 1 in that paper. Numerical results on the performance of Malliavin methods for the calculation of risk numbers in the Heston model are, puzzlingly, very difficult to find, actually seemlingly impossible to find. Also, the only paper giving an explicit form of the Malliavin weight for the Heston model seems to be Benhamou (2001), so it is unclear whether it is generally understood, as we concluded, that the form given is erroneous. There are, however, several other papers on applications of the Malliavin calculus to the Heston model, but these applications are not on the calculation of risk numbers. See for example Alòs & Ewald (2008).

Pertaining discretisation of the Heston model, Andersen (2008) seems to be the best resource available, and contains references to many other papers on the subject. For an curious paper giving a semi-analytical formula for the density of the log of the asset price in the Heston model, see Dragulescu & Yakovenko (2002). This could perhaps be useful as a benchmark for the true density.

The methods for calculating risk numbers considered here all have one major advantage over the finite difference method, an advantage which is not seen in our context. In our case, we only consider risk numbers with respect to changes in the initial value S_0 of the the underlying asset. Since S_T is linear in S_0, we can in the finite difference method reuse our simulations for the estimation of both means by scaling our simulations. Therefore, in spite of the finite difference method requiring the calculation of two means instead of just one, as in the other methods we consider, the computational time spent on simulations is not much different from the other methods. For risk numbers with respect to general parameters, the connection between S_T and the risk number variable is not as straightforward, and in such cases we would have to simulate twice as many variables in the finite difference method as for the other methods for calculating risk numbers. When having to calculate many different risk numbers, which is often the case in practice, this observation can have a major impact on computational time.

All of the numerical work in this chapter is done in R. We also considered using the language OCaml instead, hoping for improvements in both elegance and speed. However, in the end, the mediocrity of the IDE available for OCaml and the strong and efficient statistical library available for R made R the better choice.

Chapter 6

Conclusions

In this chapter, we will discuss the results which we have obtained in all three main parts of the thesis, and we will consider opportunities for extending this work.

6.1 Discussion of results

We wil discuss the results obtained for stochastic integrals, the Malliavin calculus and the applications to mathematical finance.

Stochastic integrals. In Chapter 3, we successfully managed to develop the theory of stochastic integration for integrators of the form $A_t + \sum_{k=1}^n \int_0^t Y_t^k \, dW_t^k$, where A is a continuous process of finite variation and $Y^k \in \mathfrak{L}^2(W)$, and we proved the basic results of the theory without reference to any external results. However, the increase in rigor compared to Øksendal (2005) or Steele (2000) came at some price, resulting in some honestly very tedious work. Furthermore, while we managed to prove Girsanov's theorem without reference to external results, the resulting proof was rather long compared to the modern proofs based on stochastic integration for general continuous local martingales.

We must therefore conclude that the ideal presentation of the theory would be based on general continuous local martingales. And with the proof of the existence of the quadratic variation of Section C.1, the biggest difficulty of that theory could be conveniently separated from the actual development of the integral.

The Malliavin calculus. We presented in Chapter 3 the basic results on the Malliavin derivative. While we only managed to cover a very small fraction of the results in Nualart (2006), we have managed to develop the proofs in much higher detail, including the development of some results used in the proofs such as Lemma 4.4.19 which curiously are completely absent from other accounts of the theory. The value of this work is considerable, clarifying the fundamentals of the theory and therefore paving the way for higher levels of detail in further proofs as well.

Also, the extension of the chain rule proved in Theorem 4.3.5 is an important new result. Considering how fundamental the chain rule is to the theory, this extension should make the Malliavin derivative somewhat easier to work with.

Mathematical finance. Chapter 5 contained our work on mathematical finance. We managed to use the theory developed Chapter 3 to set up a basic class of financial market models and proved sufficient conditions for absence of arbitrage, demonstrating that the theory of stochastic integration developed in Chapter 3 was sufficiently rich to obtain reasonable results. The use of the Theorem 3.8.4 in the proof of Theorem 5.1.5 and the Itô representation theorem in the proof of Theorem 5.1.11 were crucial, details which are often skipped in other accounts.

We also made an in-depth investigation of different estimation methods for prices and sensitivities in the Black-Scholes model, documenting the high effectivity of more advanced Monte Carlo methods and confirming results found in the literature. We applied the Malliavin methods to the Heston, obtaining new results. We saw that the efficiency gains seen in the Black-Scholes case could not be sustained in the Heston model. We also noted a problem with the Malliavin estimator given Benhamou (2001) and corrected the error.

6.2 Opportunities for further work

All three main parts of the thesis presents considerable oppportunities for further work. We will now detail some of these.

Stochastic integration for discontinuous FV processes. It is generally accepted that while the theory of stochastic integration for continuous semimartingales is somewhat manageable, the corresponding theory for general semimartingales is very difficult, as can be seen for example in Chapter VI of Rogers & Williams (2000b) or in Protter (2005). Nonetheless, financial models with jumps are becoming increasingly popular, see for example Cont & Tankov (2004). It would therefore be convenient to have a theory of stochastic integration at hand which covers the necessary results. To this end, we observe that the vast majority of financial models with jumps in use only uses the addition of a compound poisson process as a driving term in, say, the asset price. Such a process is of finite variation. Therefore, if it were possible to develop a theory of stochastic integration for integrators of the form $A+M$ where A is a possibly discontinuous process of finite variation and M is a continuous local martingale, one could potentially obtain a simplification of the general theory which would still be versatile enough to cover most financial applications. A theory alike to this is developed in Chapter 11 of Shreve (2004).

Measurability results. There are several problems relating to measurability properties in the further theory of the Malliavin calculus. As outlined in Section 4.5, if u is in the domain of the Skorohod integral with $u_t \in \mathbb{D}_{1,2}$ for all $t \leq T$ and there are measurable versions of $(t,s) \mapsto D_t u_s$ and $\int_0^T D_t u_s \delta W_s$, then

$$D_t \int_0^t u_s \delta W_s = \int_0^T D_t u_s \delta W_s + u_t.$$

It would be useful for the theory if there were general criteria to ensure that all of the measurability properties necessary for this to hold were satisfied. Another problem raises its head when considering the Itô-Wiener expansion. Also described in Section 4.5, this is a series expansion for variables $F \in \mathcal{L}^2(\mathcal{F}_T)$, stating that there exists square-integrable deterministic functions f_n on Δ_n such that we have the $\mathcal{L}^2$-summation formula

$$F = \sum_{n=0}^{\infty} n! \int_0^T \int_0^{t_n} \cdots \int_0^{t_3} \int_0^{t_2} f_n(t)\, \mathrm{d}W(t_1) \cdots \mathrm{d}W(t_n).$$

Here, it is not trivial to realize that the iterated integrals are well-defined. To see the nature of the problem, consider defining the single iterated integral

$$\int_0^T \int_0^{t_2} f(t_1,t_2)\, \mathrm{d}W(t_1)\, \mathrm{d}W(t_2).$$

Fix $t_2 \leq T$. The process $\int_0^t f(t_1,t_2)\, \mathrm{d}W(t_1)$ is well-defined if $t_1 \mapsto f(t_1,t_2)$ is progressively measurable. However, for varying t_2, the processes $\int_0^t f(t_1,t_2)\, \mathrm{d}W(t_1)$ are based on different integrands, and there is therefore a priori no guarantee that these can be combined in such a way as to make $t_2 \mapsto \int_0^{t_2} f(t_1,t_2)\,\mathrm{d}W(t_1)$ progressively measurable, which is necessary to be able to integrate this process with respect to the Brownian motion. Thus, it is not clear at all how to make the statement of the Itô-Wiener expansion rigorous. A further problem is that in the proof of the expansion based on the Itô representation theorem, it is necessary to know that the integrands in the representations can be chosen in certain measurable ways.

The simplest problems are those involving the Wiener-Itô expansion. As usual, putting $\theta(f) = \int_0^T f(t)\,\mathrm{d}W_t$, θ is a continuous mapping from $\mathcal{L}^2[0,T]$ to $\mathcal{L}^2(\mathcal{F}_T)$. Furthermore, if we let $\mathcal{R}(X)$ denote the Itô representation of a variable $X \in \mathcal{L}^2(\mathcal{F}_T)$ in the sense that

$$X = \int_0^T \mathcal{R}(X)_t\,\mathrm{d}W_t,$$

then $\mathcal{R}$ is a continuous linear mapping from $\mathcal{L}^2(\mathcal{F}_T)$ to $\mathcal{L}^2(\Sigma_\pi[0,T])$. Thus, the problems related to the Wiener-Itô expansion are both in some sense about iterating continuous linear operators over $\mathcal{L}^2$ spaces. In contrast to this, the problems related to the Malliavin derivative are in some sense about selecting measurable versions of closed linear operators over subspaces of $\mathcal{L}^2$ spaces.

It is not clear how to solve these problems in general. However, a first step would be to show a result which for example would allow us to, given a process X indexed by $[0,T]^2$ and being $\mathcal{B}[0,T]^2 \otimes \Sigma_\pi[0,T]$ measurable, select a version of $(s,t) \mapsto \int_0^t X(s,t)\,\mathrm{d}W(t)$ which is $\mathcal{B}[0,T]^2 \otimes \Sigma_\pi[0,T]$ measurable. Applying this several times in an appropriate manner could help defining the iterated integrals, and it is also a result of this type necessary in the Malliavin calculus.

In order to cover all of our cases, let us consider a mapping $A : \mathcal{L}^2(E,\mathbb{E}) \to \mathcal{L}^2(K,\mathbb{K})$ between two $\mathcal{L}^2$ spaces. We endow these $\mathcal{L}^2$ spaces with their Borel-σ-algebras $\mathcal{B}(\mathbb{E})$ and $\mathcal{B}(\mathbb{K})$ and assume that A is $\mathcal{B}(\mathbb{E})$-$\mathcal{B}(\mathbb{K})$ measurable. This obviously covers the cases from the Wiener-Itô expansion, since continuous mappings are measurable. And by Corollary, 4.4.16, the Malliavin derivative is continous on $\mathcal{H}_n$ for each $n \geq 0$, where $\mathcal{L}^2(\mathcal{F}_T) = \oplus_{n=0}^\infty \mathcal{H}_n$, leading us to suspect that the Malliavin derivative indeed is measurable as well on its domain.

Let $(D,\mathbb{D})$ be another measurable space. We will outline how to prove the following result: If $X : D \times E \to \mathbb{R}$ is $\mathbb{D} \otimes \mathbb{E}$ measurable such that for each $x \in D$ it holds that $X(x,\cdot) \in \mathcal{L}^2(E,\mathbb{E})$, there exists a $\mathbb{D} \otimes \mathbb{K}$ measurable version of $(x,y) \mapsto A(X(x,\cdot))(y)$.

This can be directly applied to obtain measurable versions of stochastic integrands where only one coordinate is integrated.

Let $X : D \times E \to \mathbb{R}$ be given with the properties stated above. We first define $Y : D \to \mathcal{L}^2(E, \mathbb{E})$ by putting $Y(x)(y) = X(x, y)$. By a monotone class argument, we prove that Y is $\mathbb{D}$-$\mathcal{B}(\mathbb{E})$ measurable. We can then use that A is measurable to obtain that $x \mapsto A(Y(x))$ is $\mathbb{D}$-$\mathcal{B}(\mathbb{K})$ measurable. Defining $Z(x, z) = A(Y(x))(z)$, Z is a mapping from $D \times K$ to $\mathbb{R}$. By definition, $Z(x, \cdot) = A(Y(x)) = A(X(x, \cdot))$. We are done if we can argue that there exists a measurable version of Z. This is the most difficult part of the argument. It is easy in the case where $x \mapsto A(Y(x))$ is simple. For the general case, we use that $\mathcal{L}^2$ convergence implies almost sure convergence for a subsequence. Selecting a subsequence for each x in a measurable manner, we can obtain a measurable version. The argument is something like that found in Protter (2005), Theorem IV.63. However, the technique of that proof depends on the linearity of the integral mapping. We need to utilize measurability properties directly, making the proof more cumbersome.

Risk numbers. In our investigation of the Malliavin method for the Heston model, we only considered the digital delta. Furthermore, we only applied a few extra Monte Carlo methods to our routines, namely latin hypercube sampling and antithetic sampling. There are several other methods available, such as those of Fries & Yoshi (2008) or those detailed in Glasserman (2003), which would be interesting to attempt to apply to the Heston model.

Furthermore, it would be interesting to investigate the efficiency gains possible for path-dependent options, both of european, bermudan or american type, and it would be interesting to extend the results to other models. Here, the extension to models with jumps poses an extra challenge, as the Malliavin calculus we have developed only works in the case of Brownian motion. For Malliavin calculus for processes with jumps such as Lévy processes, see Di Nunno et al. (2008).

Appendix A

Analysis

In this appendix, we develop the results from pure analysis that we will need. We will give some main theorems from analysis, and we will consider the basic results on spaces $\mathcal{L}^p(\mu)$, $C_c^\infty(\mathbb{R}^n)$ and $C^\infty(\mathbb{R}^n)$. We also consider Hermite polynomials and Hilbert spaces. The appendix is more or less a smorgasbord, containing a large variety of different results. For some results, references are given to proofs. For other results, full proofs are given. The reason for using so much energy on these proofs is basically that there does not seem to be any simple set of sources giving all the necessary results. Our main sources for the real analysis are Carothers (2000), Berg & Madsen (2001), Rudin (1987) and Cohn (1980). For Hilbert space theory, we use Rudin (1987), Hansen (2006) and Meise & Vogt (1997).

A.1 Functions of finite variation

We say that a function $F : [0, \infty) \to \mathbb{R}$ has finite variation if

$$V_F(t) = \sup \sum_{k=1}^{n} |F(t_k) - F(t_{k-1})|$$

is finite for all $t > 0$, where the supremum is taken over all partitions of $[0, t]$. In this case, we call V_F the variation of F over $[0, t]$. We will give the basic properties of functions of finite variation. This will be used when defining the stochastic integral for integrators which have paths of finite variation.

Lemma A.1.1. *V_F is continuous if and only of F is continuous.*

Proof. See Carothers (2000), Theorem 13.9. □

Theorem A.1.2. *There is a unique measure μ_F such that $\mu_F(a,b] = F(b) - F(a)$ for any $0 \leq a \leq b < \infty$.*

Proof. See Theorem 3.2.6 of Dudley (2002) for the case where F is monotone. The general case follows easily by applying the decomposition of F into a difference of monotone functions. □

Lemma A.1.3. *If F is continuous, then μ_F has no atoms.*

Proof. This follows directly from the definition of μ_F given in Theorem A.1.2. □

Since functions of finite variation corresponds to measures by Theorem A.1.2, we can define the integral with respect to a function of finite variation.

Definition A.1.4. *We put $\mathcal{L}^1(F) = \mathcal{L}^1(\mu_F)$ and define, for any function $x \in \mathcal{L}^1(F)$, $\int_0^t x_u \, \mathrm{d}F_u = \int_0^t x_u \, \mathrm{d}\mu_F(u)$.*

Lemma A.1.5. *Let F be continuous and of finite variation. If g is continuous, then g is integrable with respect to F over $[0,t]$, and the Riemann sums*

$$\sum_{k=1}^{n} g(t_{k-1})(F(t_k) - F(t_{k-1}))$$

converge to $\int_0^t g(s) \, \mathrm{d}F(s)$ as the mesh of the partition tends to zero.

Proof. We have

$$\sum_{k=1}^{n} g(t_{k-1})(F(t_k) - F(t_{k-1})) = \sum_{k=1}^{n} g(t_{k-1})\mu_F((t_{k-1}, t_k]) = \int_0^t s_n \, \mathrm{d}\mu_F,$$

where $s_n = \sum_{k=1}^n g(t_{k-1}) 1_{(t_{k-1}, t_k]}$. As the mesh tends to zero, s_n tends to g by the uniform continuity of g on $[0,t]$. Since g is continous, it is bounded on $[0,t]$. Therefore, by dominated convergence, $\int_0^t s_n \, \mathrm{d}\mu_F$ tends to $\int_0^t g(s) \, \mathrm{d}\mu_F(s)$, as desired. □

Lemma A.1.6. *If $x \in \mathcal{L}^1(F)$, then $t \mapsto \int_0^t x_u \,\mathrm{d}F(u)$ has finite variation.*

Proof. This follows by decomposing F into a difference between monotone functions and applying the linearity of the integral in the integrator. $\square$

A.2 Some classical results

In this section, we collect a few assorted results. These will be used in basically every part of the thesis. Gronwall's lemma is used in the proof of Itô's representation theorem, the multidimensional mean value theorem is used in various approximation arguments, the Lebesgue differentiation theorem, is used in density arguments for the Itô integral and the duality result on L^p and L^q is used when extending the Malliavin derivative.

Lemma A.2.1 (Gronwall). *Let $g, b : [0,\infty) \to [0,\infty)$ be two measurable mappings and let a be a nonnegative constant. Let $T > 0$ and assume $\int_0^T b(s)\,\mathrm{d}s < \infty$. If $g(t) \le a + \int_0^t g(s)b(s)\,\mathrm{d}s$ for $t \le T$, then $g(t) \le a\exp(\int_0^t b(s)\,\mathrm{d}s)$ for $t \le T$.*

Proof. Define $B(t) = \int_0^t b(s)\,\mathrm{d}s$ and $G(t) = \int_0^t g(s)b(s)\,\mathrm{d}s$. Then

$$\frac{d}{dt}e^{-B(t)}G(t) = e^{-B(t)}b(t)g(t) - e^{-B(t)}G(t)b(t) = e^{-B(t)}b(t)(g(t) - G(t)) \le ae^{-B(t)}b(t)$$

for $t \le T$. Integrating this equation from 0 to T, we find

$$e^{-B(t)}G(t) \le \int_0^t ae^{-B(s)}b(s)\,\mathrm{d}s = a\left(1 - e^{-B(t)}\right).$$

This shows $G(t) \le a(e^{B(t)} - 1)$ and therefore

$$g(t) \le a + \int_0^t g(s)b(s)\,\mathrm{d}s = a + G(t) \le ae^{B(t)},$$

as was to be proved. $\square$

Lemma A.2.2. *Let $\varphi \in C^1(\mathbb{R})$. For any distinct $x, y \in \mathbb{R}^n$, there exists ξ on the line between x and y such that*

$$\varphi(y) - \varphi(x) = \sum_{k=1}^n \frac{\partial\varphi}{\partial x_k}(\xi_k)(y_k - x_k).$$

Proof. Define $g : [0,1] \to \mathbb{R}^n$ by $g(t) = ty + (1-t)x$. Then, by the ordinary mean value theorem and the chain rule, for some $\theta \in [0,1]$,

$$\begin{aligned} f(y) - f(x) &= (f \circ g)(1) - (f \circ g)(0) \\ &= (f \circ g)'(\theta) \\ &= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(g(\theta)) g_k'(\theta) \\ &= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\theta y + (1-\theta)x)(y_k - x_k), \end{aligned}$$

as desired. □

Theorem A.2.3 (Lebesgue's differentiation theorem). *Let $f \in \mathcal{L}^1(\mathbb{R})$ and put $F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$. Then F is almost everywhere differentiable with derivative f.*

Proof. See Rudin (1987), Theorem 7.11 and Theorem 7.7. □

Corollary A.2.4. *Let $f : \mathbb{R} \to \mathbb{R}$ be bounded. Then it holds that*

$$\lim \frac{1}{\frac{1}{n}} \int_{t-\frac{1}{n}}^{t} f(x)\,\mathrm{d}x = f(t)$$

almost everywhere for $t \in \mathbb{R}$.

Proof. Consider any bounded interval (a,b), and put $g_{(a,b)} = f(x)1_{(a,b)}(x)$. Then $g \in \mathcal{L}^1(\mathbb{R})$, and theorem A.2.3 yields for almost all $t \in (a,b)$ that

$$\lim \frac{1}{\frac{1}{n}} \int_{t-\frac{1}{n}}^{t} f(x)\,\mathrm{d}x = \lim \frac{1}{\frac{1}{n}} \int_{t-\frac{1}{n}}^{t} g(x)\,\mathrm{d}x = g(t) = f(t).$$

Since a countable union of null sets is again a null set, the conclusion of the lemma follows. □

Theorem A.2.5. *Let $p, q \geq 1$ be dual exponents and let $(E, \mathcal{E}, \mu)$ be a measure space. The dual of $L^p(E, \mathcal{E}, \mu)$ is isometrically isomorphic to $L^q(E, \mathcal{E}, \mu)$, and an isometric isomorphism $\varphi : L^q(E, \mathcal{E}, \mu) \to L^p(E, \mathcal{E}, \mu)'$ is given by*

$$\varphi(G)(F) = \int G(x)F(x)\,\mathrm{d}\mu(x).$$

In particular, an element $F \in \mathcal{L}^p(E, \mathcal{E}, \mu)$ is almost surely zero if and only if it holds that $\int G(x)F(x)\,\mathrm{d}\mu(x) = 0$ for all $G \in \mathcal{L}^q(E, \mathcal{E}, \mu)$.

Proof. For the duality statement and the isometric isomorphism, see Proposition 13.13 of Meise & Vogt (1997). The last statement of the theorem follows from Proposition 6.10 of Meise & Vogt (1997). □

A.3 Dirac families and Urysohn's lemma

In this section, we wil develop a version of Urysohn's Lemma for smooth mappings. We first state the original Urysohn's Lemma.

Theorem A.3.1 (Urysohn's Lemma). *Let K be a compact set in $\mathbb{R}^n$, and let V be an open set. Assume $K \subseteq V$. There exists $f \in C_c(\mathbb{R}^n)$ such that $K \prec f \prec V$.*

Proof. This is proven in Rudin (1987), Theorem 2.12. □

The mappings which exist by Theorem A.3.1 are called bump functions. Our goal is to extend Theorem A.3.1 to show the existence of bump functions which are not only continuous, but are smooth. To do so, recall that for two Lebesgue integrable mappings f and g, the convolution $f * g$ is defined by $(f * g)(x) = \int f(x-y)g(y)\,\mathrm{d}y$. Fundamental results on the convolution operation can be found in Chapter 8 of Rudin (1987). The convolution of f with g has an averaging effect. The following two results will allow us to use this averaging effect to smooth out nonsmooth functions, allowing us to prove the smooth version of Urysohn's lemma.

Lemma A.3.2. *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue integrable and $g \in C^\infty(\mathbb{R}^n)$, then the convolution $f * g$ is in $C^\infty(\mathbb{R}^n)$.*

Proof. By definition, $(f * g)(x) = \int f(y)g(x-y)\,\mathrm{d}y$. Now, since g is differentiable, g is bounded on bounded sets and therefore, by Theorem 8.14 in Hansen (2004b), $(f*g)(x)$ is differentiable in the k'th direction with $\frac{\partial}{\partial x_k}(f*g)(x) = \int f(y)\frac{\partial g}{\partial x_k}(x-y)\,\mathrm{d}y$. Here $\frac{\partial g}{\partial x_k}$ is in $C^\infty(\mathbb{R}^n)$, so it follows inductively that $f * g \in C^\infty(\mathbb{R}^n)$. □

Theorem A.3.3 (Dirac family). *Let μ be a Radon measure on $\mathbb{R}^n$ and let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. There exists a family of mappings $\psi_\varepsilon \in C^\infty(\mathbb{R}^n)$ such that $\psi_\varepsilon \geq 0$, $\mathrm{supp}\psi_\varepsilon \subseteq B_\varepsilon(0)$ and $\int \psi_\varepsilon(x)\,\mathrm{d}\mu(x) = 1$. Here, $B_\varepsilon(0)$ is the ε-ball in $\|\cdot\|$. We call (ψ_ε) a Dirac family with respect to $\|\cdot\|$ and μ.*

Proof. Since all norms on $\mathbb{R}^n$ are equivalent, it will suffice to prove the result for $\|\cdot\|_2$. Therefore, $B_\varepsilon(0)$ will in the following denoted the centered ε-ball in $\|\cdot\|_2$. As seen in Berg & Madsen (2001), page 190, there exists $\chi \in C^\infty(\mathbb{R}^n)$ such that $0 \le \chi \le 1$ and supp$\chi \subseteq B_1(0)$. Then χ_ε given by $\chi_\varepsilon(x) = \chi(\frac{x}{\varepsilon})$ is also in $C^\infty(\mathbb{R}^n)$, $0 \le \chi_\varepsilon \le 1$ and supp$\chi_\varepsilon \subseteq B_\varepsilon(0)$. Since μ is Radon, any open ball in $\mathbb{R}^n$ has finite μ-measure, and therefore $\chi_\varepsilon \in \mathcal{L}^1(\mu)$. Putting $\psi_\varepsilon = \frac{1}{\|\chi_\varepsilon\|_1}\chi_\varepsilon$, ψ_ε then has the desired properties. □

Theorem A.3.4 (Smooth Urysohn's Lemma). *Let K be a compact set in $\mathbb{R}^n$, and let V be an open set. Assume $K \subseteq V$. There exists $f \in C_c^\infty(\mathbb{R}^n)$ such that $K \prec f \prec V$.*

Proof. By Theorem A.3.1, there exists $f \in C_c(\mathbb{R}^n)$ such that $K \prec f \prec V$. Let $\varepsilon \in (0, \frac{1}{2})$ and choose δ parrying ε for the uniform continuity of f. Define the sets $K' = f^{-1}([1-\varepsilon, 1])$ and $V' = f^{-1}(\varepsilon, \infty)$.

Since $K \prec f \prec V$, clearly $K \subseteq K' \subseteq V' \subseteq V$. Furthermore, K is compact and V is open. If x is such that $\|y - x\| \le \delta$ for some $y \in K$, then $|f(x) - 1| = |f(x) - f(y)| \le \varepsilon$ and therefore $f(x) \ge 1 - \varepsilon$, so $x \in K'$. On the other hand, if $x \in V'$ then $f(x) > \varepsilon$, so if $\|y - x\| \le \delta$, $|f(y) - f(x)| \le \varepsilon$ and therefore $f(y) > 0$, showing $y \in V$. In other words, $B_\delta(x) \subseteq V$.

Now let $g \in C_c(\mathbb{R}^n)$ with $K' \prec g \prec V'$ and let $h = g * \psi_\delta$. We claim that h satisfies the properties in the theorem. By Lemma A.3.2, $h \in C_c^\infty(\mathbb{R}^n)$. We will show that $h \prec V$. Let $x \in V$, we need to prove that $h(x) = 0$. We have $h(x) = \int g(y)\psi_\delta(x-y)\,\mathrm{d}y$. Now, whenever $\|x - y\| \le \delta$, the integrand is zero because $\psi_\delta(x-y)$ is zero. On the other hand, if $\|x - y\| > \delta$, we must have $y \notin V'$, because if $y \in V'$, $B_\delta(y) \subseteq V$. Therefore, $g(y)$ is zero and we again find that the integrand is zero. Thus, $h(x)$ is zero and $h \prec V$.

Next, we show that $K \prec h$. Let $x \in K$. Whenever $\|y - x\| \le \delta$, we have $y \in K'$ and therefore $g(y)$ is one. On the other hand, whenever $\|y - x\| > \delta$, ψ_δ is zero. Thus, $h(x) = \int g(y)\psi_\delta(x-y)\,\mathrm{d}y = 1$ and $K \prec h$. We conclude $K \prec h \prec V$, as desired. □

We are also going to need a version of Urysohn's lemma for closed sets instead of compact sets.

Lemma A.3.5. *Let F be a closed set in $\mathbb{R}^n$, and let V be an open set. Assume $F \subseteq V$. There exists $f \in C^\infty(\mathbb{R}^n)$ such that $F \prec f \prec V$.*

Proof. Defining $f(x) = \frac{d(x,V^c)}{d(x,F)+d(x,V^c)}$, we find that f is continuous and $F \prec f \prec V$.

Using the same technique as in the proof of Theorem A.3.4, we extend this to provide smoothness of the bump function. □

A.4 Approximation and density results

In this section, we will review an assortment of approximation and density results which will be used throughout the thesis. In particular the Malliavin calculus makes good use of various approximation results.

Whenever our approximation results involve smooth functions, we will almost always apply a Dirac family as a mollifier and take convolutions. The next three results are applications of this technique, showing how to approximate respectively continuous, Lipschitz and a class of differentiable mappings.

Lemma A.4.1. *Let $f \in C(\mathbb{R}^n)$, and let ψ_ε be a Dirac family. Then $f * \psi_\varepsilon$ converges uniformly to f on compacts.*

Proof. Let $M > 0$. For any x with $\|x\|_2 \le M$, we have

$$\begin{aligned}
|f(x) - (f * \psi_\varepsilon)(x)| &= \left| \int f(x)\psi_\varepsilon(x-y)\,\mathrm{d}y - \int f(y)\psi_\varepsilon(x-y)\,\mathrm{d}y \right| \\
&\le \int |f(x) - f(y)|\psi_\varepsilon(x-y)\,\mathrm{d}y \\
&\le \sup_{y \in B_\varepsilon(x)} |f(x) - f(y)|.
\end{aligned}$$

We may therefore conclude

$$\sup_{\|x\|_2 \le M} |f(x) - (f * \psi_\varepsilon)(x)| \le \sup_{\|x\|_2 \le M} \sup_{y \in B_\varepsilon(x)} |f(x) - f(y)|.$$

Since f is continuous, f is uniformly continuous on compact sets. In particular, f is uniformly continuous on $B_{M+\delta}(0)$ for any $\delta > 0$, and therefore the above tends to zero. □

Lemma A.4.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz mapping with respect to $\|\cdot\|_\infty$ such that $|f(x) - f(y)| \le K\|x - y\|_\infty$. There exists a sequence of mappings (g_ε) in $C^\infty(\mathbb{R}^n)$ with partial derivatives bounded by K such that g_ε converges uniformly to f.*

Proof. Let (ψ_ε) be a Dirac family on $\mathbb{R}^n$ with respect to $\|\cdot\|_\infty$ and define $g_\varepsilon = f * \psi_\varepsilon$. Then $g_\varepsilon \in C^\infty(\mathbb{R}^n)$, and we clearly have, with e_k the k'th unit vector,

$$\begin{aligned}
|g_\varepsilon(x+he_k) - g_\varepsilon(x)| &= \left| \int f(x+he_k-y)\psi_\varepsilon(y)\,\mathrm{d}y - \int f(x-y)\psi_\varepsilon(y)\,\mathrm{d}y \right| \\
&\le \int |f(x+he_k-y) - f(x)|\psi_\varepsilon(y)\,\mathrm{d}y \\
&\le Kh.
\end{aligned}$$

It follows that g_ε has partial derivatives bounded by K. It remains to show uniform convergence of g_ε to f, but this follows since

$$\begin{aligned}
|f(x) - g_\varepsilon(x)| &= \left| \int f(x)\psi_\varepsilon(x-y)\,\mathrm{d}y - \int f(y)\psi_\varepsilon(x-y)\,\mathrm{d}y \right| \\
&\le \int |f(x) - f(y)|\psi_\varepsilon(x-y)\,\mathrm{d}y \\
&\le K \int \|x-y\|_\infty \psi_\varepsilon(x-y)\,\mathrm{d}y \\
&\le \varepsilon K.
\end{aligned}$$

The proof is finished. ☐

Lemma A.4.3. *Let $f \in C^1(\mathbb{R}^n)$, and assume that f has bounded partial derivatives. The exists a sequence $(g_\varepsilon) \subseteq C_p^\infty(\mathbb{R}^n)$ with the following properties:*

1. *g_ε converges uniformly to f.*
2. *$\partial_{x_k} g_\varepsilon$ converges pointwise to $\partial_{x_k} f$ for $k \le n$.*
3. *$\|\partial_{x_k} g\|_\infty \le \|\partial_{x_k} f\|_\infty$ for $k \le n$.*

Proof. Let (ψ_ε) be a Dirac family on $\mathbb{R}^n$. Define $g_\varepsilon = f * \psi_\varepsilon$. We then have $g_\varepsilon \in C^\infty(\mathbb{R}^n)$. We claim that g_ε satisfies the properties of the lemma.

Step 1: Uniform convergence. We find

$$\begin{aligned}
|f(x) - g_\varepsilon(x)| &= \left| \int f(x)\psi_\varepsilon(x-y)\,\mathrm{d}y - \int f(y)\psi_\varepsilon(x-y)\,\mathrm{d}y \right| \\
&\leq \int |f(x) - f(y)|\psi_\varepsilon(x-y)\,\mathrm{d}y \\
&\leq \int \left| \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\xi_k)(x_k - y_k) \right| \psi_\varepsilon(x-y)\,\mathrm{d}y \\
&\leq \max_{k\leq n} \left\| \frac{\partial f}{\partial x_k}(\xi_k) \right\|_\infty \int \|x-y\|_1 \psi_\varepsilon(x-y)\,\mathrm{d}y \\
&\leq \sqrt{n} \max_{k\leq n} \left\| \frac{\partial f}{\partial x_k}(\xi_k) \right\|_\infty \int \|x-y\|_2 \psi_\varepsilon(x-y)\,\mathrm{d}y \\
&\leq \varepsilon\sqrt{n} \max_{k\leq n} \left\| \frac{\partial f}{\partial x_k}(\xi_k) \right\|_\infty,
\end{aligned}$$

so g_ε converges uniformly to f.

Step 2: Pointwise convergence of partial derivatives. We have

$$\frac{\partial g_\varepsilon}{\partial x_k}(x) = \frac{\partial}{\partial x_k} \int f(x-y)\psi_\varepsilon(y)\,\mathrm{d}y = \int \frac{\partial}{\partial x_k} f(x-y)\psi_\varepsilon(y)\,\mathrm{d}y = (\partial_{x_k} f) * \psi_\varepsilon.$$

By the same calculations as in the first step, we therefore find

$$\begin{aligned}
|\partial_{x_k} g_\varepsilon(x) - \partial_{x_k} f(x)| &\leq \int |\partial_{x_k} f(x) - \partial_{x_k} f(y)|\psi_\varepsilon(x-y)\,\mathrm{d}y \\
&\leq \sup_{y\in B_\varepsilon(x)} |\partial_{x_k} f(x) - \partial_{x_k} f(y)|,
\end{aligned}$$

where $B_\varepsilon(x)$ denotes the ε-ball around x in $\|\cdot\|_2$. Continuity of $\partial_{x_k} f$ shows the desired pointwise convergence.

Step 3: Boundedness of partial derivatives. To show the boundedness condition on the partial derivatives of g_ε, we calculate

$$\begin{aligned}
\left| \frac{\partial g_\varepsilon}{\partial x_k}(x) \right| &= \left| \frac{\partial}{\partial x_k} \int f(x-y)\psi_\varepsilon(y)\,\mathrm{d}y \right| \\
&\leq \int \left| \frac{\partial f}{\partial x_k}(x-y) \right| \psi_\varepsilon(y)\,\mathrm{d}y \\
&\leq \left\| \frac{\partial f}{\partial x_k} \right\|_\infty,
\end{aligned}$$

showing the final claim of the lemma. This also shows $g_\varepsilon \in C_p^\infty(\mathbb{R}^n)$. $\square$

We now prove that for any Radon measure μ on $\mathbb{R}^n$ and any $p \geq 1$, the space $C_c^\infty(\mathbb{R}^n)$ is dense in $\mathcal{L}^p(\mu)$.

Lemma A.4.4. *$C_c^\infty(\mathbb{R}^n)$ is uniformly dense in $C_c(\mathbb{R}^n)$. Furthermore, for any element $f \in C_c(\mathbb{R}^n)$, the approximating sequence in $C_c^\infty(\mathbb{R}^n)$ can be taken to have a common compact support.*

Proof. Let $f \in C_c(\mathbb{R}^n)$, and let ψ_ε be a Dirac family on $\mathbb{R}^n$. Let $M > 0$ be such that $\operatorname{supp} f \subseteq B_M(0)$. Put $f_\varepsilon = f * \psi_\varepsilon$. By Lemma A.3.2, $f_\varepsilon \in C^\infty(\mathbb{R}^n)$. And since $\operatorname{supp}\psi_\varepsilon \subseteq B_\varepsilon(0)$, it holds that for x with $\|x\|_2 \geq M + \varepsilon$,

$$f_\varepsilon(x) = \int f(y)\psi_\varepsilon(x-y)\,\mathrm{d}y = \int 1_{B_M(0)}(y)1_{B_\varepsilon(x)}(y)(x-y)\,\mathrm{d}y = 0,$$

so f_ε has compact support. Therefore $f_\varepsilon \in C_c^\infty(\mathbb{R}^n)$. By Lemma A.4.1, f_ε converges uniformly to f on compacts. Since $\operatorname{supp} f_\varepsilon \subseteq B_{M+\varepsilon}(0)$, our approximating sequence has a common compact support. In particular, the uniform convergence on compacts is actually true uniform convergence. This shows the claims of the lemma. □

Theorem A.4.5. *Let μ be a Radon measure on $\mathbb{R}^n$. Then $C_c^\infty(\mathbb{R}^n)$ is dense in $\mathcal{L}^p(\mu)$ for $p \geq 1$.*

Proof. First note that by combining Theorem 2.14, Theorem 2.18 and Theorem 3.14 of Rudin (1987), we can conclude that $C_c(\mathbb{R}^n)$ is dense in $\mathcal{L}^p(\mu)$. It will therefore suffice to show that we can approximate elements of $C_c(\mathbb{R}^n)$ in $\mathcal{L}^p(\mu)$ with elements of $C_c^\infty(\mathbb{R}^n)$. Let $f \in C_c(\mathbb{R}^n)$. By Lemma A.4.4, there exists $f_n \in C_c^\infty(\mathbb{R}^n)$ converging uniformly to f_n, and by the same lemma, we can assume that all the functions f_n and f have the same compact support K. Then,

$$\lim_n \int |f(x) - f_n(x)|^p \,\mathrm{d}\mu(x) \leq \lim_n \|f - f_n\|_\infty \mu(K) = 0.$$

□

A.5 Hermite polynomials

In this section, we present and prove some basic results on Hermite polynomials. A source for results such as these is Szegö (1939), Chapter V, where a good deal of results on Hermite polynomials are gathered. Most results are presented without proof, however. Also note that what is called Hermite polynomials in Szegö (1939) is a little bit different than what is called Hermite polynomials here.

Definition A.5.1. *The n'th Hermite polynomial is defined by*

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

We put $H_0(x) = 1$ *and* $H_{-1}(x) = 0$.

It is not completely obvious from Definition A.5.1 that the Hermite polynomials are polynomials at all. Before showing that this is actually the case, we will obtain a recursion relation for the Hermite polynomials.

Lemma A.5.2. *It holds that*

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

so that for any x, $H_n(x)$ *is the n'th coefficient of the Taylor expansion of* $\exp(tx - \frac{t^2}{2})$ *as a function of* t. *In particular,*

$$H_n(x) = \frac{\partial^n}{\partial t^n} \exp\left(tx - \frac{t^2}{2}\right)\bigg|_{t=0}.$$

Proof. The mapping $t \mapsto \exp(-\frac{1}{2}(x-t)^2)$ is obviously entire when considered as a complex mapping, and therefore given by its Taylor expansion. We then obtain

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

□

Lemma A.5.3. *It holds that for* $n \geq 0$, $H_n'(x) = nH_{n-1}(x)$.

Proof. We see that

$$\begin{aligned}
\sum_{n=0}^{\infty}\frac{H_n'(x)}{n!}t^n &= \frac{d}{dx}\sum_{n=0}^{\infty}\frac{H_n(x)}{n!}t^n \\
&= \frac{d}{dx}\exp\left(tx-\frac{t^2}{2}\right) \\
&= t\exp\left(tx-\frac{t^2}{2}\right) \\
&= \sum_{n=0}^{\infty}\frac{H_n(x)}{n!}t^{n+1} \\
&= \sum_{n=1}^{\infty}\frac{nH_{n-1}(x)}{n!}t^n.
\end{aligned}$$

Recalling that $H_{-1}(x)=0$, uniqueness of coefficients yields the result. □

Lemma A.5.4. *It holds that for $n\geq 0$, $H_{n+1}(x)=xH_n(x)-nH_{n-1}(x)$.*

Proof. We find, using the product rule and Lemma A.5.3,

$$\begin{aligned}
H_{n+1}(x) &= (-1)^{n+1}e^{\frac{x^2}{2}}\frac{d^{n+1}}{dx^{n+1}}e^{-\frac{x^2}{2}} \\
&= (-1)^{n+1}e^{\frac{x^2}{2}}\frac{d}{dx}\frac{d^n}{dx^n}e^{-\frac{x^2}{2}} \\
&= (-1)^{n+1}\left(\frac{d}{dx}\left(e^{\frac{x^2}{2}}\frac{d^n}{dx^n}e^{-\frac{x^2}{2}}\right)-\left(\frac{d}{dx}e^{\frac{x^2}{2}}\right)\frac{d^n}{dx^n}e^{-\frac{x^2}{2}}\right) \\
&= -H_n'(x)+(-1)^n xe^{\frac{x^2}{2}}\frac{d^n}{dx^n}e^{-\frac{x^2}{2}} \\
&= -nH_{n-1}(x)+xH_n(x),
\end{aligned}$$

as was to be proved. □

Lemma A.5.2, Lemma A.5.3 and Lemma A.5.4 along with the definition are the basic workhorses for producing results about the Hermite polynomials. In particular, we can now show that the Hermite polynomials actually are polynomials.

Lemma A.5.5. *The n'th Hermite polynomial H_n is a polynomial of degree n.*

Proof. We use complete induction. The result is clearly true for $n=0$. Assume that it is true for all $k\leq n$. We know that $H_{n+1}(x)=xH_n(x)-nH_{n-1}(x)$, by Lemma A.5.4. Here, $nH_{n-1}(x)$ is a polynomial of degree $n-1$ by assumption, and H_n is a polynomial of degree n. Therefore, $xH_n(x)$ is a polynomial of degree $n+1$, and the result follows. □

Lemma A.5.6. *The span of the n first Hermite polynomials $H_0, \ldots, H_{n-1}$ is the same as the span of the n first monomials $1, x, x^2, \ldots, x^{n-1}$.*

Proof. From Lemma A.5.5, it is clear that the span of the n first monomials includes the n first Hermite polynomials. To show the reverse inclusion, we use induction. Since $H_0(x) = 1$, the result is trivially true for $n = 1$. Assume that it has been proven for n. Let $\mathcal{H}_{n+1}$ be the span of the first $n+1$ Hermite polynomials. We need to prove that $x^k \in \mathcal{H}_{n+1}$ for $k \leq n$. By our induction hypothesis, $x^k \in \mathcal{H}_{n+1}$ for $k \leq n-1$. To prove that $x^n \in \mathcal{H}_{n+1}$, note that by Lemma A.5.5, we know that H_n is a polynomial of degree n, that is,

$$H_n(x) = \sum_{k=0}^{n} a_k x^k,$$

where $a_n \neq 0$. Since $x^k \in \mathcal{H}_{n+1}$ for $k \leq n-1$, we obtain $a_n x^n \in \mathcal{H}_{n+1}$. Since $a_n \neq 0$, this implies $x^n \in \mathcal{H}_{n+1}$, as desired. □

A.6 Hilbert Spaces

In this section, we review some results from the theory of Hilbert spaces. Our main sources are Hansen (2006) and Meise & Vogt (1997).

Definition A.6.1. *A Hilbert Space H is a complete normed linear space over $\mathbb{R}$ such that the norm of H is induced by an inner product.*

In the remainer of the section, H will denote a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We say that two elements $x, y \in H$ are orthogonal and write $x \perp y$ if $\langle x, y \rangle = 0$. We say that two subsets M and N of H are orthogonal and write $M \perp N$ if it holds that for each $x \in M$ and $y \in N$, $x \perp y$. The elements orthogonal to M are denoted $M^\perp$. We write span M for the linear span of M, and we write $\overline{\text{span}}\, M$ for the closure of the span of M.

Lemma A.6.2. *Let M and N be subsets of H. If M and N are orthogonal, then so are $\overline{\text{span}}\, M$ and $\overline{\text{span}}\, N$.*

Proof. By the linearity of the inner product, span M and span N are orthogonal. By the continuity of the inner product, $\overline{\text{span}}\, M$ and $\overline{\text{span}}\, N$ are orthogonal as well. □

Lemma A.6.3. *If U is a subspace of H, then $U^\perp$ is a closed subspace of H.*

Proof. See Rudin (1987), p. 79. □

Lemma A.6.4. *Let M is a subspace of H. M is dense in H if and only if $M^\perp = \{0\}$.*

Proof. This follows from Corollary 11.8 of Meise & Vogt (1997). □

Theorem A.6.5 (Pythagoras' Theorem). *Let $x_1, \ldots, x_n$ be orthogonal elements of H. Then*

$$\left\| \sum_{k=1}^n x_k \right\|^2 = \sum_{k=1}^n \|x_k\|^2.$$

Proof. See Hansen (2006), Theorem 3.1.14. □

Theorem A.6.6. *Let U be a closed subspace of H. Each element $x \in H$ has a unique decomposition $x = Px + x - Px$, where $Px \in U$ and $x - Px \in U^\perp$. The mapping $P : H \to U$ is a linear contraction and is called the orthogonal projection onto U.*

Proof. The existence of the decomposition and the linearity of P is proved in Rudin (1987), Theorem 4.11. That P is a contraction follows by applying Theorem A.6.5 to obtain

$$\|Px\|^2 \leq \|Px\|^2 + \|x - Px\|^2 = \|Px + x - Px\|^2 = \|x\|^2.$$

□

Lemma A.6.7. *Let U be a closed subspace of H. The orthogonal projection P onto U satisfies $P^2 = P$, $R(P) = U$ and $N(P) = U^\perp$. Also, $\|x - Px\| = d(x, U)$, where d denotes the metric induced by $\|\cdot\|$.*

Proof. Let $x \in H$. Then $Px \in U$, and therefore $P^2x = Px$, showing the first claim. Since P maps into U and $Px = x$ for $x \in U$, it is clear that range satisfies $R(P) = U$. To show the claim about the null space, is is clear that $N^\perp \subseteq N(P)$. To show the other inclusion, let $x \in N(P)$. Then $x = Px + x - Px = x - Px \in U^\perp$, as desired.

Finally, the formula $\|x - Px\| = d(x, U)$ follows by Lemma 11.5 of Meise & Vogt (1997). $\square$

If M and N are orthogonal subspaces of H, we define the orthogonal sum $M \oplus N$ by

$$M \oplus N = \{x + y | x \in M, y \in N\}.$$

Lemma A.6.8. *If M and N are closed orthogonal subspaces of H, then $M \oplus N$ is a closed subspace of H as well. Letting P and Q be the orthogonal projections onto M and N, the orthogonal projection onto $M \oplus N$ is $P + Q$.*

Proof. That $M \oplus N$ is closed is proved in Lemma 3.5.8 of Hansen (2006). Let an element $x \in H$ be given. We write $x = (P+Q)x + x - (P+Q)x$, and if we can argue that $x - (P+Q)x \in (M \oplus N)^\perp$, we are done, since $(P+Q)x$ obviously is in $M \oplus N$.

To this end, first note that we have $x = Px + (x - Px)$, where $x - Px \in M^\perp$ by the definition of the orthogonal projection P. Using the definition of the orthogonal projection Q, we also have $x - Px = Q(x - Px) + (x - Px - Q(x - Px))$, where $x - Px - Q(x - Px) \in N^\perp$. Since $R(Q) = N$, $Q(x - Px) \in M^\perp$. We conclude that $x - Px - Q(x - Px)$ is orthogonal to both N and M. In particular, then, it is orthogonal to $M \oplus N$.

Finally, by Lemma A.6.7, $N(Q) = N^\perp \supseteq M = R(P)$. Thus, $QP = 0$ and we obtain

$$x - (P+Q)x = x - Px - Q(x - Px) \in (M \oplus N)^\perp,$$

as desired. $\square$

Lemma A.6.9. *Letting M and N be orthogonal subspaces of H, it holds that $M \oplus N$ is the span of $M \cup N$.*

Proof. Clearly, $M \oplus N \subseteq \text{span}\, M \cup N$. On the other hand, if $x \in \text{span}\, M \cup N$, we have $x = \sum_{k=1}^n \lambda_k x_k$ where $x_k \in M \cup N$. Splitting the sum into parts in M and parts in N, we obtain the reverse inclusion. $\square$

Lemma A.6.10. *Let x_n be an orthogonal sequence in H. $\sum_{n=1}^\infty x_n$ is convergent if and only if $\sum_{n=1}^\infty \|x_n\|^2$ is finite.*

Proof. Both directions of the equivalence follows from the relation, with $n > m$,

$$\left\|\sum_{k=1}^{n} x_k - \sum_{k=1}^{m} x_k\right\|^2 = \sum_{k=m+1}^{n} \|x_k\|^2$$

combined with the completeness of H. $\square$

We are also going to need infinite sums of orthogonal subspaces. If (U_n) is a sequence of orthogonal subspaces of H, we define the orthogonal sum $\oplus_{n=1}^{\infty} U_n$ of the subspaces by

$$\bigoplus_{n=1}^{\infty} U_n = \left\{ \sum_{n=1}^{\infty} x_n \,\middle|\, x_n \in U_n \text{ and } \sum_{n=1}^{\infty} \|x_n\|^2 < \infty \right\}$$

This is well-defined, since we know from Lemma A.6.10 that $\sum_{n=1}^{\infty} x_n$ is convergent whenever $\sum_{n=1}^{\infty} \|x_n\|^2$ is finite.

Lemma A.6.11. *$\oplus_{n=1}^{\infty} U_n$ is a closed subspace of H, and $\oplus_{n=1}^{\infty} U_n = \overline{\operatorname{span}}\, \cup_{n=1}^{\infty} U_n$.*

Proof. We first prove that $\oplus_{n=1}^{\infty} U_n$ is closed. From the second assertion of the lemma, it will follow that $\oplus_{n=1}^{\infty} U_n$ is a subspace.

Let $x_n \in \oplus_{n=1}^{\infty} U_n$ with $x_n = \sum_{k=1}^{\infty} x_n^k$, and assume that x_n converges to x. We need to prove $x \in \oplus_{n=1}^{\infty} U_n$. From Lemma A.6.10, $\sum_{k=1}^{\infty} \|x_n^k\|^2$ is finite for any n. In particular, by the triangle inequality and the relation $(x+y)^2 \leq 2x^2 + 2y^2$, $\sum_{k=1}^{\infty} \|x_n^k - x_m^k\|^2$ is finite for any n, m. This implies that $\sum_{k=1}^{\infty} x_n^k - x_m^k$ is convergent, and we obtain

$$\|x_n - x_m\|^2 = \left\|\sum_{k=1}^{\infty} x_n^k - x_m^k\right\|^2 = \sum_{k=1}^{\infty} \left\|x_n^k - x_m^k\right\|^2.$$

Now, since x_n is convergent, it is in particular a cauchy sequence. From the above equality, we then obtain that x_n^k is a cauchy sequence for every k, convergent to a limit x^k. Since U_k is closed, $x^k \in U_k$. We want to prove $x = \sum_{k=1}^{\infty} x^k$. To this end, define $\alpha_n = (\|x_n^k\|)_{k \geq 1}$. α_n is then a member of ℓ^2, and

$$\|\alpha_n - \alpha_m\|_{\ell^2}^2 = \sum_{k=1}^{\infty} (\|x_n^k\| - \|x_m^k\|)^2 \leq \sum_{k=1}^{\infty} \|x_n^k - x_m^k\|^2 = \|x_n - x_m\|^2,$$

so α_n is cauchy in ℓ^2, therefore convergent to a limit α. In particular, α_n^k converges to α^k. But $\alpha_n^k = \|x_n^k\|$, and $\|x_n^k\|$ converges to $\|x^k\|$. Therefore, $\alpha^k = \|x^k\|$. We may now conclude

$$\lim_n \sum_{k=1}^{\infty} (\|x_n^k\| - \|x^k\|)^2 = \lim_n \|\alpha_n - \alpha\|_{\ell^2}^2 = 0,$$

in particular $\lim_n \sum_{k=1}^\infty \|x_n^k\|^2 = \sum_{k=1}^\infty \|x^k\|^2$. We then obtain by orthogonality,

$$\begin{aligned}\left\|x - \sum_{k=1}^n x_k\right\|^2 &= \lim_m \left\|\sum_{k=1}^\infty x_m^k - \sum_{k=1}^n x^k\right\|^2 \\ &= \lim_m \sum_{k=1}^n \|x_m^k - x^k\|^2 + \sum_{k=n+1}^\infty \|x_m^k\|^2 \\ &= \lim_m \sum_{k=n+1}^\infty \|x_m^k\|^2 \\ &= \sum_{k=n+1}^\infty \|x^k\|^2,\end{aligned}$$

and since $\sum_{k=1}^\infty \|x^k\|^2$ is finite, it follows that $x = \sum_{k=1}^\infty x^k$. Then $x \in \oplus_{n=1}^\infty U_n$, and we may finally conclude that $\oplus_{n=1}^\infty U_n$ is closed.

Next, we need to prove the equality $\oplus_{n=1}^\infty U_n = \overline{\text{span}}\ \cup_{n=1}^\infty U_n$. As in the proof of Lemma A.6.9, it is clear that $\oplus_{n=1}^\infty U_n \subseteq \overline{\text{span}}\ \cup_{n=1}^\infty U_n$. To prove the other inclusion, note that we clearly have span $\cup_{n=1}^\infty U_n \subseteq \oplus_{n=1}^\infty U_n$. Since we have seen that $\oplus_{n=1}^\infty U_n$ is closed, it follows that $\overline{\text{span}}\ \cup_{n=1}^\infty U_n \subseteq \oplus_{n=1}^\infty U_n$. □

Lemma A.6.12. *Let (U_n) be a sequence of closed orthogonal subspaces and let P_n be the orthogonal projection onto U_n. For any $x \in \oplus_{n=1}^\infty U_n$, it holds that $x = \sum_{n=1}^\infty P_n x$.*

Proof. Since $x \in \oplus_{n=1}^\infty U_n$, we know that $x = \sum_{n=1}^\infty x_n$ for some $x_n \in U_n$. Since P_k is continuous and the subspaces are orthogonal, we find

$$P_k x = \sum_{n=1}^\infty P_k x_n = x_n,$$

as desired. □

Lemma A.6.13. *For any subset A, it holds that $A \cap A^\perp = \{0\}$.*

Proof. Let $x \in A \cap A^\perp$. We find $\|x\|^2 = \langle x, x\rangle = 0$, showing $x = 0$. □

Lemma A.6.14. *Let M and N be orthogonal closed subspaces. Assume $x \in M \oplus N$. If $x \in M^\perp$, then $x \in N$.*

Proof. Since N and M are orthogonal, $N \subseteq M^\perp$. Let $x = y + z$ with $y \in M$ and $z \in N$. Then $y = x - z \in M^\perp$. Since also $y \in M$, Lemma A.6.13 yields $y = 0$ and therefore $x = z \in N$, as desired. □

Lemma A.6.15. *Let U, M and N be closed subspaces. Assume that U and M are orthogonal, and assume that U and N are orthogonal. If $U \oplus M = U \oplus N$, then $M = N$.*

Proof. By symmetry, it will suffice to show $M \subseteq N$. Assume that $x \in M$. Then $x \in U \oplus M$, and therefore $x \in U \oplus N$. Now, since $M \subseteq U^\perp$, we have $x \in U^\perp$. By Lemma A.6.14, $x \in N$. □

Lemma A.6.16. *Let (U_n) be a sequence of closed orthogonal subspaces. It then holds for any $n \geq 1$ that*

$$\bigoplus_{k=1}^{\infty} U_k = \left(\bigoplus_{k=1}^{n} U_k \right) \oplus \left(\bigoplus_{k=n+1}^{\infty} U_k \right).$$

Proof. It is clear that $(\oplus_{k=1}^n U_k) \oplus (\oplus_{k=n+1}^\infty U_k) \subseteq \oplus_{k=1}^\infty U_k$. To show the other inclusion, let $x \in \oplus_{k=1}^\infty U_k$. We then have $x = \sum_{k=1}^\infty x_k$ for some $x_k \in U_k$, where $\sum_{k=1}^\infty \|x_k\|^2$. Therefore, in particular $\sum_{k=n+1}^\infty \|x_k\|^2$. Thus, by the decomposition

$$x = \left(\sum_{k=1}^{n} x_k \right) + \left(\sum_{k=n+1}^{\infty} x_k \right),$$

we obtain the desired result. □

Lemma A.6.17. *Let $x, x' \in H$ and assume that $\langle y, x \rangle = \langle y, x' \rangle$ for all $y \in H$. Then $x = x'$.*

Proof. It follows that $x - x'$ is orthogonal to H. From Lemma A.6.13, $x - x' = 0$ and therefore $x = x'$. □

Lemma A.6.18. *Let H be a Hilbert space and let U be a closed subspace. Let P denote the orthogonal projection onto U. Let $x \in H$ and let $x' \in U$. If it holds for all $y \in U$ that $\langle y, x \rangle = \langle y, x' \rangle$, then $x' = Px$.*

Proof. Let Q be the orthogonal projection onto $U^\perp$. We then have, for any $y \in U$,

$$\begin{aligned}\langle y, Px\rangle &= \langle y, Qx + Px\rangle \\ &= \langle y, x\rangle \\ &= \langle y, x'\rangle.\end{aligned}$$

Since Px and x' are both elements of U, and U is a Hilbert space since it is a closed subspace of H, we conclude by Lemma A.6.17 that $Px = x'$. □

We end with a few results on weak convegence. We say that a sequence x_n is weakly convergent to x if $\langle y, x_n\rangle$ tends to $\langle y, x\rangle$ for all $y \in H$.

Lemma A.6.19. *If x_n has a weak limit, it is unique.*

Proof. See Lemma 3.6.3 of Hansen (2006). □

Lemma A.6.20. *If x_n converges to x, it also converges weakly to x.*

Proof. This follows immediately by the continuity properties of the inner product. □

Theorem A.6.21. *Let H be a Hilbert space and let (x_n) be a bounded sequence in H. Then (x_n) has a weakly convergent subsequence.*

Proof. See Schechter (2002), Theorem 8.16, where the theorem is proved for reflexive Banach spaces. Since every Hilbert space is reflexive according to Corollary 11.10 of Meise & Vogt (1997), this proves the claim. □

Lemma A.6.22. *Assume that (x_n) converges weakly to x. Then $\|x\| \leq \liminf \|x_n\|$.*

Proof. See Theorem 3.4.11 of Ash (1972). □

Lemma A.6.23. *Let x_n be a sequence in H converging weakly to x. Let $A : H \to H$ be linear and continuous. Then Ax_n converges weakly as well.*

Proof. This follows from Theorem 3.4.11 of Ash (1972). □

A.7 Closable operators

In this section, we consider the concept of closable and closed operators. This will be used when extending the Malliavin derivative. Let $\mathbb{X}$ and $\mathbb{Y}$ be Banach spaces, that is, complete normed linear spaces. Consider a subspace $D(A)$ of $\mathbb{X}$ and an operator $A : D(A) \to \mathbb{Y}$. We say that A is closable if it holds that whenever (x_n) is a sequence in $D(A)$ converging to zero and Ax_n is convergent, then the limit of Ax_n is zero. We say that A is closed if it holds that whenever (x_n) is a sequence in $D(A)$ such that both x_n and Ax_n are convergent, then $\lim x_n \in D(A)$ and $A(\lim x_n) = \lim Ax_n$.

Lemma A.7.1. *Let A be some operator with domain $D(A)$. A is closed if and only if the graph of A is closed.*

Proof. This follows since the graph of A is closed precisely if it holds for any convergent sequence (x_n, Ax_n) where $x_n \in D(A)$ that $\lim x_n \in D(A)$ and $A(\lim x_n) = \lim Ax_n$. □

Lemma A.7.2. *Let A be a closable operator. We define the domain $D(\overline{A})$ of the closure of A as the set of $x \in \mathbb{X}$ such that there exists x_n in $D(A)$ converging to x and such that Ax_n is convergent. The limit of Ax_n is independent of the sequence x_n and $D(\overline{A})$ is a linear space.*

Proof. We first prove the result on the uniqueness of the limit. Assume that $x \in D(\overline{A})$ and consider two different sequences x_n and y_n in $D(A)$ converging to x such that Ax_n and Ay_n are both convergent. We need to prove that the limits are the same. But $x_n - y_n$ converges to zero and $A(x_n - y_n)$ is convergent, so since A is closable, it follows that $A(x_n - y_n)$ converges to zero and thus $\lim_n Ax_n = \lim_n Ay_n$, as desired.

Next, we prove that $D(\overline{A})$ is a linear space. Let $x, y \in D(\overline{A})$ and let $\lambda, \mu \in \mathbb{R}$. Let x_n and y_n be sequences in $D(A)$ converging to x and y, respectively. Then $\lambda x_n + \mu y_n$ is a sequence in $D(A)$ converging to $\lambda x + \mu y$, and let $A(\lambda x_n + \mu y_n)$ is convergent since A is linear. Thus, $\lambda x + \mu y \in D(\overline{A})$ and $D(\overline{A})$ is a linear space. □

Theorem A.7.3. *Assume that A is closable. There is a unique extension $\overline{A}$ of A as a linear operator from $D(A)$ to $D(\overline{A})$, and the extension is closed.*

Proof. Let $x \in D(\overline{A})$. From Lemma A.7.2, we know that there is x_n in $D(A)$ converging to x and such that Ax_n is convergent, and the limit of Ax_n is independent of x_n.

We can then define $\overline{A}x$ as the limit of Ax_n. We need to prove that this extension is linear and closed.

To prove linearity, let $x, y \in D(\overline{A})$ and let $\lambda, \mu \in \mathbb{R}$. Let x_n and y_n be sequences in $D(A)$ converging to x and y, respectively, such that Ax_n and Ay_n converges to $\overline{A}x$ and $\overline{A}y$, respectively. Then $\lambda x_n + \mu y_n$ converges to $\lambda x + \mu y$ and $A(\lambda x_n + \mu y_n)$ is convergent. By definition, the limit is $\overline{A}(\lambda x + \mu y)$, and we conclude

$$\overline{A}(\lambda x + \mu y) = \lim_n A(\lambda x_n + \mu y_n) = \lim_n \lambda A x_n + \mu A y_n = \lambda \overline{A} x + \mu \overline{A} y,$$

as desired. To show that the extension is closed, we show that the graph of $\overline{A}$ is the closure of the graph of A. Assume that (x, y) is a limit point of the graph of A, then there exists x_n in $D(A)$ such that $x = \lim x_n$ and $y = Ax_n$. By definition, we obtain $x \in D(\overline{A})$ and $Ax = y$. Therefore, (x, y) is in the graph of $\overline{A}$. On the other hand, if (x, y) is in the graph of $\overline{A}$, in particular $x \in D(\overline{A})$ and $\overline{A}x = y$. Then, there exists x_n in $D(A)$ converging to x such that Ax_n is convergent as well, and the limit is $\overline{A}x$. Thus, (x_n, Ax_n) converges to $(x, \overline{A}x)$, so (x, y) is a limit point of the graph of A. We conclude that the graph of $\overline{A}$ is the closure of the graph of A, and by Lemma A.7.1, $\overline{A}$ is closed. $\square$

Appendix B

Measure theory

This appendix contains some results from abstract measure theory and probability theory which we will need. Our main sources are Hansen (2004a), Berg & Madsen (2001), Jacobsen (2003) and Rogers & Williams (2000a).

B.1 The Monotone Class Theorem

In this section, we state the monotone class theorem and develop an important corollary, used when analysing the Brownian motion under the usual conditions.

Definition B.1.1. *A monotone vector space $\mathbb{H}$ on Ω is a subspace of the vector space $B(\Omega)$ of bounded real functions on Ω, such that $\mathbb{H}$ contains all constant functions and such that if (f_n) is a sequence of nonnegative functions increasing pointwise towards a bounded function f, then $f \in \mathbb{H}$.*

Theorem B.1.2 (Monotone Class Theorem). *Let $\mathcal{K} \subseteq B(\Omega)$ be stable under multiplication, and let $\mathbb{H} \subseteq B(\Omega)$ be a monotone vector space. If $\mathcal{K} \subseteq \mathbb{H}$, $\mathbb{H}$ contains all $\sigma(\mathcal{K})$-measurable bounded functions.*

Proof. The theorem is stated and proved as Theorem I.21 in Dellacherie & Meyer (1975), under the assumption that $\mathbb{H}$ is closed under uniform convergence. Since

Sharpe (1988) proves that any monotone vector space is closed under uniform convergence, the claim follows. $\square$

By $C([0,\infty),\mathbb{R}^n)$, we denote the continuous mappings from $[0,\infty)$ to $\mathbb{R}^n$. We endow this space with the σ-algebra $\mathcal{C}([0,\infty),\mathbb{R}^n)$ induced by the coordinate projections X°_t, $X^\circ_t(f) = f(t)$ for any $f \in C([0,\infty),\mathbb{R}^n)$.

Lemma B.1.3. *Let F be closed in $\mathbb{R}^n$. There exists $(f_n) \subseteq C^\infty_c(\mathbb{R}^n)$ such that f_n converges pointwise to 1_F.*

Proof. By Cohn (1980), Proposition 7.1.4, there exists a sequence G_n of open sets such that $F \subseteq G_n$ and $F = \cap_{n=1}^\infty G_n$. Let K_n be an increasing sequence of compact sets with $\cup_{n=1}^\infty K_n$. Then $F \cap K_n$ is compact, so by Theorem A.3.4, there is f_n with $F \cap K_n \prec f_n \prec G_n$. We claim that this sequence of functions fulfills the conditions in the lemma. Let $x \in F$. From a certain point onwards, $x \in F \cap K_n$, and then $f_n(x) = 1$. Assume next $x \notin F$. Then, $x \notin G_n$ from a certain point onwards, and then $f_n(x) = 0$. We conclude $\lim f_n = 1_F$. $\square$

Corollary B.1.4. *Let $\mathbb{H} \subseteq B(C([0,\infty,\mathbb{R}^n))$ be a monotone vector space. If $\mathbb{H}$ contains all functions of the form $\prod_{k=1}^n f_k(X^\circ_{t_k})$ with $f_k \in C_c(\mathbb{R}^n)$, then $\mathbb{H}$ contains all $\mathcal{C}([0,\infty),\mathbb{R}^n)$ measurable functions.*

Proof. Since the compact sets are closed under finite intersections, it is clear that $C_c(\mathbb{R}^n)$ is an algebra. Therefore, putting

$$\mathcal{K} = \left\{ \prod_{k=1}^n f_k(X^\circ_{t_k}) \middle| 0 \le t_1 \le \cdots \le t_n \text{ and } f_1, \ldots, f_n \in C_c(\mathbb{R}) \right\},$$

it is clear that $\mathcal{K}$ is an algebra. By Theorem B.1.2, $\mathbb{H}$ contains all $\sigma(\mathcal{K})$ meausrable mappings. We wish to argue that $\mathcal{K} = \mathcal{C}([0,\infty),\mathbb{R}^n)$. It is clear that $\mathcal{K} \subseteq \mathcal{C}([0,\infty),\mathbb{R}^n)$, so it will suffice to show the other inclusion. To do so, we only need to show that all coordinate projections are $\mathcal{K}$ meausrable. Let $t \ge 0$ and let F be closed in $\mathbb{R}^n$. From Lemma B.1.3, there exists a sequence $(f_n) \subseteq C^\infty_c(\mathbb{R}^n)$ converging pointwise to 1_F. This means that $1_F(X^\circ_t)$ is the pointwise limit of $f_n(X^\circ_t)$, so $1_F(X^\circ_t)$ is $\mathcal{K}$-measurable. This shows that $(X^\circ_t \in F) \in \mathcal{K}$. Therefore X°_t is $\mathcal{K}$ measurable, as desired. $\square$

B.2 $\mathcal{L}^p$ and almost sure convergence

This section contains some of the classical results on $\mathcal{L}^p$ and almost sure convergence. Most of these results are surely well-known, we repeat them here for completeness.

Lemma B.2.1. *Let $(E, \mathcal{E}, \mu)$ be a measure space. Assume that f_n converges to f and g_n converges to g in $\mathcal{L}^2(\mathcal{E})$. Then $f_n g_n$ converges to fg in $\mathcal{L}^1(\mathcal{E})$.*

Proof. By the Cauchy-Schwartz inequality,

$$\begin{aligned} \|f_n g_n - fg\|_1 &\leq \|f_n g_n - f_n g\|_1 + \|f_n g - fg\|_1 \\ &\leq \|f_n\|_2 \|g_n - g\|_2 + \|f_n - f\|_2 \|g\|_2, \end{aligned}$$

so since $\|f_n\|_2$ is bounded, the conclusion follows. □

Lemma B.2.2. *Let $(E, \mathcal{E}, \mu)$ be a measure space. Assume that f_n converges in $\mathcal{L}^1$ to f, and assume that g_n converges almost surely to g. If the sequence g_n is dominated by a constant, then $f_n g_n$ converges in $\mathcal{L}^1$ to fg.*

Proof. We have

$$\begin{aligned} \|f_n g_n - fg\|_1 &\leq \|f_n g_n - f g_n\|_1 + \|f g_n - fg\| \\ &\leq M\|f_n - f\|_1 + \|f g_n - fg\|_1. \end{aligned}$$

Here, the first term trivially tends to zero. By dominated convergence with bound $2Mf$, the second term also tends to zero. □

Lemma B.2.3 (Scheffé). *Let $(E, \mathcal{E}, \mu)$ be a measure space and let f_n be a sequence of probability densities with respect to μ. Let f be another probability density with respect to μ. If f_n converges μ-almost surely to f, then f_n converges to f in $\mathcal{L}^1$.*

Proof. This is Lemma 22.4 of Hansen (2004a). □

Lemma B.2.4. *Let $(E, \mathcal{E}, \mu)$ be a finite measure space. If $1 \leq p \leq r$, it holds that $\mathcal{L}^r(\mathcal{E}) \subseteq \mathcal{L}^p(\mathcal{E})$ and $\mathcal{L}^r$ convergence implies $\mathcal{L}^p$ convergence.*

Proof. See Theorem 7.11 of Berg & Madsen (2001). □

Lemma B.2.5. *Let $(E, \mathcal{E}, \mu)$ be a measure space. If f_n converges to f in $\mathcal{L}^p$, there is a subsequence f_{n_k} converging μ almost surely to f, dominated by a mapping in $\mathcal{L}^p$.*

Proof. This is proved as Corollary 7.20 in Berg & Madsen (2001). □

B.3 Convergence of normal distributions

In this section, we prove a result on convergence of normal distributions which will be essential to our development of the Malliavin calculus. We begin with a few lemmas.

Lemma B.3.1 (Hölder's inequality). *Let $(E, \mathcal{E}, \mu)$ be a measure space and let $f_1, \ldots, f_n$ be real measurable mappings for some $n \geq 2$. Let $p_1, \ldots, p_n \in (1, \infty)$ with $\sum_{k=1}^n \frac{1}{p_k} = 1$. It then holds that $\| \prod_{k=1}^n f_k \|_1 \leq \prod_{k=1}^n \|f_k\|_{p_n}$. We say that $p_1, \ldots, p_n$ are conjugate exponents.*

Proof. We use induction. The case $n = 2$ is just the ordinary Hölder's inequality, proving the induction start. Assume that the results holds in the case of n, we will prove it for $n+1$. Let $q = \frac{p_{n+1}}{p_{n+1}-1}$, then q is the conjugate exponent to p_{n+1}, so the ordinary Hölder's inequality yields $\| \prod_{k=1}^{n+1} f_k \| \leq \| \prod_{k=1}^n f_k \|_q \|f_{n+1}\|_{p_{n+1}}$. Next, note that $\sum_{k=1}^n \frac{1}{p_k} = 1 - \frac{1}{p_{n+1}} = \frac{p_{n+1}-1}{p_{n+1}} = \frac{1}{q}$, so applying the induction hypothesis with the conjugate exponents $\frac{p_1}{q}, \ldots, \frac{p_n}{q}$, we obtain

$$
\begin{aligned}
\left\| \prod_{k=1}^n f_k \right\|_q &= \left(\left\| \prod_{k=1}^n |f_k|^q \right\|_1 \right)^{\frac{1}{q}} \\
&\leq \left(\prod_{k=1}^n \| |f_k|^q \|_{\frac{p_1}{q}} \right)^{\frac{1}{q}} \\
&= \left(\prod_{k=1}^n \left(\int |f_k|^{p_1} \, d\mu \right)^{\frac{q}{p_1}} \right)^{\frac{1}{q}} \\
&= \prod_{k=1}^n \|f_k\|_{p_k},
\end{aligned}
$$

which, combined with our earlier finding, yields the desired result. □

Lemma B.3.2. *Let X be a normal distribution with mean μ and variance σ^2. Let $n \in \mathbb{N}_0$. Then we have*

$$EX^n = \sum_{k=0}^{[n/2]} \frac{(2k)!}{2^k k!} \sigma^{2k} \mu^{n-2k}.$$

Proof. Since the odd central moments of X are zero, we obtain

$$\begin{aligned} EX^n &= E(X - \mu + \mu)^n \\ &= \sum_{k=0}^{n} \binom{k}{n} E(X-\mu)^k \mu^{n-k} \\ &= \sum_{k=0}^{[n/2]} \binom{2k}{n} E(X-\mu)^{2k} \mu^{n-2k} \\ &= \sum_{k=0}^{[n/2]} \frac{(2k)!}{2^k k!} \sigma^{2k} \mu^{n-2k}, \end{aligned}$$

as desired. □

Lemma B.3.3. *Let X_k be a sequence of n-dimensional normally distributed variables. Assume that for any $i \le n$, X_k^i converges weakly to X. Let f $\mathbb{R}^n$ to $\mathbb{R}$ have polynomial growth. Then $f(X_k)$ is bounded over k in $\mathcal{L}^p$ for any $p \ge 1$.*

Proof. If f has polynomial growth, then so does $|f|^p$ for any $p \ge 1$. Therefore, it will suffice to show that $f(X_k)$ is bounded in $\mathcal{L}^1$ for any f with polynomial growth. Let p be a polynomial in n variables dominating f. Assume that p has degree m and $p(x) = \sum_{|a| \le m} \lambda_a \prod_{i=1}^n x_i^{a_i}$, using usual multi-index notation. Obviously, we then obtain

$$\begin{aligned} |f(X_k)| &\le |p(X_k)| \\ &\le \sum_{|a| \le m} |\lambda_a| \prod_{i=1}^n |(X_k^i)^{a_i}| \\ &\le \sum_{|a| \le m} |\lambda_a| \prod_{i=1}^n (1 + (X_k^i)^2)^{a_i}). \end{aligned}$$

Letting $q(x) = \sum_{|a| \le m} |\lambda_a| \prod_{i=1}^n (1 + x_i^2)^{a_i}$, q is a polynomial of degree $2m$ in n variables, and we have $E|f(X_k)| \le E|q(X_k)|$. Furthermore, the exponents in q are all even. Thus, it will suffice to show the result in the case f is a polynomial with even exponents.

Therefore, assume that f is a polynomial in n variables of degree m. Assume that we have $f(x) = \sum_{a \in \mathbb{I}_m^n} \lambda_a \prod_{i=1}^n x_i^{a_i}$, where λ_a is zero whenever a has an odd element. The generalized Hölder's inequality of Lemma B.3.1 then yields

$$\begin{aligned} E|f(X_k)| &\leq \sum_{|a|\leq m} |\lambda_a| E\left(\prod_{i=1}^n (X_k^i)^{a_i}\right) \\ &\leq \sum_{|a|\leq m} |\lambda_a| \prod_{i=1}^n \left(E(X_k^i)^{na_i}\right)^{\frac{1}{n}}. \end{aligned}$$

Let μ_k^i denote the mean of X_k^i and σ_k^i denote the standard deviation of X_k^i. Since X_k^i is normally distributed and converges weakly, μ_k^i and σ_k^i are convergent, therefore bounded. Now, by Lemma B.3.2, $E(X_k^i)^n$ is a continuous function of μ_k^i and σ_k^i for fixed i and n. Therefore, $k \mapsto E(X_k^i)^n$ is bounded for any i and any n. This shows that $E|f(X_k)|$ is bounded over k, as was to be proved. $\square$

Theorem B.3.4. *Let X_k be a sequence of n-dimensional variables converging in probability to X. Assume that (X_k, X) is normally distributed for any k. Let $f \in C_p^1(\mathbb{R}^n)$. Then $f(X_k)$ converges to $f(X)$ in $\mathcal{L}^p$, $p \geq 1$.*

Proof. Let $f \in C_p^1(\mathbb{R}^n)$ be given. Let p_i be a polynomial in n variables dominating $\frac{\partial f}{\partial x_i}$. The mean value theorem yields, for some ξ_k on the line segment between X_k and X,

$$\begin{aligned} |f(X_k) - f(X)| &= \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(\xi_k)(X_k^i - X^i) \right| \\ &\leq \sum_{i=1}^n |p_i(\xi_k)||X_k^i - X^i|. \end{aligned}$$

Now, defining $Y_k^i = 2 + (X_k^i)^2 + (X^i)^2$, we have $|\xi_k^i| \leq |X_k^i| + |X^i| \leq Y_k^i$ and therefore $|p_i(\xi_k)| \leq q_i(Y_k)$ for some other polynomial q_i. Using Jensen's inequality and the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned} E|f(X_k) - f(X)|^p &\leq E\left(\sum_{k=1}^n q_i(Y_k)|X_k^i - X^i|\right)^p \\ &\leq En^p \sum_{i=1}^n \left(q_i(Y_k)|X_k^i - X^i|\right)^p \\ &= n^p \sum_{i=1}^n Eq_i(Y_k)^p |X_k^i - X^i|^p \\ &\leq n^p \sum_{i=1}^n \|q_i(Y_k)^p\|_2 \||X_k^i - X^i|^p\|_2. \end{aligned}$$

Now, $q_i(Y_k)$ is a polynomial transformation of the $2n$-dimensional normally distributed variable (X_k, X). Since X_k converges in probability to X_k, we also have weak convergence. Thus, (X_k, X) converges coordinatewisely weakly to (X, X). By Lemma B.3.3, we may then conclude that $\|q_i(Y_k)^p\|_2$ is bounded over k for any $i \le n$. Let C be a bound, we can then conclude

$$\begin{aligned} E|f(X_k) - f(X)|^p &\le Cn^p \sum_{i=1}^n \| |X_k^i - X^i|^p \|_2 \\ &= Cn^p \sum_{i=1}^n \left(E|X_k^i - X^i|^{2p} \right)^{\frac{1}{2}}. \end{aligned}$$

To prove the desired result, it will therefore suffice to prove that X_k^i tends to X^i in $\mathcal{L}^p$ for any $p \ge 1$. It will be enough to prove that we have $\mathcal{L}^{2n}$ convergence for $n \in \mathbb{N}$. Let $n \in \mathbb{N}$ be given. Since X_k^i converges in probability to X^i, $X_k^i - X^i$ converges in probability to zero, so $X_k^i - X^i$ converges in distribution to the point measure at zero. Because we have asumed that (X_k, X) is normally distribued, $X_k^i - X^i$ is normally distributed. Let ξ_k^i be the mean and let ν_k^i be the standard deviation. We then concluce that ξ_k^i and ν_k^i both tend to zero as k tends to infinity. By Lemma B.3.2, we then conclude that $E(X_k^i - X^i)^{2n}$ tends to zero, showing the desired convergence in $\mathcal{L}^{2n}$ and implying the result of the lemma. □

B.4 Separability and bases for $\mathcal{L}^2$-spaces

This section yields some results on $\mathcal{L}^2$ spaces which will be necessary for the development of the Hilbert space results of the Malliavin calculus.

Lemma B.4.1. *If $(E, \mathcal{E}, \mu)$ is a finite measure space which is countably generated, then $\mathcal{L}^2(\mathcal{E})$ is separable as a pseudometric space.*

Proof. Let $\mathbb{D}$ be a countable generating family for $\mathcal{E}$. We can assume without loss of generality that $E \in \mathbb{D}$. Let $\mathbb{H}$ denote the family of sets $\cap_{k=1}^n A_k$ where $A_k \in \mathbb{D}$ for $k \le n$. Then $\mathbb{H}$ is countable, $\mathbb{D} \subseteq \mathbb{H}$, and $\mathbb{H}$ is stable under intersections. Put $\mathcal{S} = \{1_A | A \in \mathcal{E}\}$ and $\mathcal{S}_{\mathbb{H}} = \{1_A | A \in \mathbb{H}\}$.

Step 1: span $\mathcal{S}_{\mathbb{H}}$ **is dense in** $\mathcal{L}^2(E, \mathcal{E}, \mu)$**.** We will begin by showing that span $\mathcal{S}_{\mathbb{H}}$ is dense in $\mathcal{L}^2(\mathcal{E})$. After doing so, we will identify a dense subset of span $\mathcal{S}_{\mathbb{H}}$. To show that span $\mathcal{S}_{\mathbb{H}}$ is dense, first note that from Theorem 7.27 of Berg & Madsen (2001), span $\mathcal{S}$ is dense in $\mathcal{L}^2(\mathcal{E})$. Therefore, to show that span $\mathcal{S}_{\mathbb{H}}$ is dense in $\mathcal{L}^2(\mathcal{E})$, it will suffice to show that span $\mathcal{S}_{\mathbb{H}}$ is dense in span $\mathcal{S}$. And to do so, it will suffice to show that $\mathcal{S} \subseteq \overline{\text{span}}\ \mathcal{S}_{\mathbb{H}}$.

Define $\mathbb{K} = \{A \in \mathcal{E} | A \in \overline{\text{span}}\ \mathcal{S}_{\mathbb{H}}\}$. To show that $\mathcal{S} \subseteq \overline{\text{span}}\ \mathcal{S}_{\mathbb{H}}$, we must show that $\mathbb{K} = \mathcal{E}$. Clearly, $\mathbb{H} \subseteq \mathbb{K}$. Since $\mathbb{H}$ is a generating family for $\mathcal{E}$ stable under intersections, it will suffice to show that $\mathbb{K}$ is a Dynkin class.

Clearly, $E \in \mathbb{K}$. Assume that $A, B \in \mathbb{K}$ with $B \subseteq A$. Then $1_{A \setminus B} = 1_A - 1_B$. Since each of these is in $\overline{\text{span}}\ \mathcal{S}_{\mathbb{H}}$, so is $1_A - 1_B$ and therefore $A \setminus B \in \mathbb{K}$. Finally, assume that A_n is an increasing sequence in $\mathbb{K}$ and put $A = \cup_{n=1}^{\infty} A_n$. Since μ is finite, 1_{A_n} tends to 1_A in $\mathcal{L}^2(\mathcal{E})$ by dominated convergence. Since $1_{A_n} \in \overline{\text{span}}\ \mathcal{S}_{\mathbb{H}}$ for all n and $\overline{\text{span}}\ \mathcal{S}_{\mathbb{H}}$ is closed, $1_A \in \overline{\text{span}}\ \mathcal{S}_{\mathbb{H}}$ and therefore $A \in \mathbb{K}$. Thus, $\mathbb{K}$ is a Dynkin class. We conclude that $\mathbb{K} = \mathcal{E}$ and therefore $\mathcal{S} \subseteq \overline{\text{span}}\ \mathcal{S}_{\mathbb{H}}$, from which we deduce that span $\mathcal{S}_{\mathbb{H}}$ is dense in $\mathcal{L}^2(\mathcal{E})$.

Step 2: Identification of a dense countable set. In order to show separability of $\mathcal{L}^2(\mathcal{E})$, it will suffice to show that there exists a dense countable subset of span $\mathcal{S}_{\mathbb{H}}$. Define

$$V = \left\{ \sum_{k=1}^{n} \lambda_k 1_{A_k} \,\middle|\, \lambda_k \in \mathbb{Q}, A_k \in \mathbb{H}, k \leq n \right\}.$$

Since $\mathbb{H}$ is countable, V is countable. We will show that V is dense in span $\mathcal{S}_{\mathbb{H}}$. Let $f \in$ span $\mathcal{S}_{\mathbb{H}}$ be given, $f = \sum_{k=1}^{n} \lambda_k 1_{A_k}$. Let λ_k^j tend to λ_k through $\mathbb{Q}$ and define $f_j = \sum_{k=1}^{n} \lambda_k^j 1_{A_k}$. Then $f_j \in V$ and

$$\|f - f_j\|_2 \leq \sum_{k=1}^{n} |\lambda_k - \lambda_k^j| \|1_{A_k}\|_2,$$

showing the desired result. □

Lemma B.4.2. *Let $(E, \mathcal{E}, \mu)$ and $(K, \mathcal{K}, \nu)$ be two finite measure spaces which are countably generated. Let $([f_i])$ and $([g_j])$ be countable orthonormal bases for $L^2(\mathcal{E})$ and $L^2(\mathcal{K})$, respectively. Then the family $([f_i \otimes g_j])$ is an orthonormal basis for $L^2(\mathcal{E} \otimes \mathcal{K})$.*

Proof. We first show that the family is orthonormal. By the Tonelli theorem, we find $\|f_i \otimes g_j\|_2 = \|f_i\|_2 \|g_j\|_2 = 1$. To show that the family is orthogonal, consider $f_i \otimes g_j$ and $f_k \otimes g_n$. We then obtain by the Fubini theorem,

$$\begin{aligned} \langle f_i \otimes g_j, f_k \otimes g_n \rangle &= \int f_i(s) g_j(t) f_k(s) g_n(t) \, \mathrm{d}(\mu \otimes \nu)(s,t) \\ &= \langle f_i, f_k \rangle \langle g_j, g_n \rangle, \end{aligned}$$

showing that if $i \neq k$ or $j \neq n$, $f_i \otimes g_j$ is orthogonal to $f_k \otimes g_n$, as desired. It remains to show that the family is complete. To this end, consider $\psi \in L^2(\mathcal{E} \otimes \mathcal{K})$ and assume that ψ is orthogonal to $f_i \otimes g_j$ for any i and j. We need to show that ψ is almost surely zero. To this end, we first calculate

$$\begin{aligned} \langle \psi, f_i \otimes g_j \rangle &= \int \psi(s,t) f_i(s) g_j(t) \, \mathrm{d}(\mu \otimes \nu)(s,t) \\ &= \int \int \psi(s,t) f_i(s) \, \mathrm{d}\mu(s) g_j(t) \, \mathrm{d}\nu(t) \\ &= \int \langle \psi(\cdot,t), f_i \rangle g_j(t) \, \mathrm{d}\nu(t). \end{aligned}$$

Put $\phi_i(t) = \langle \psi(\cdot,t), f_i \rangle$. By the Cauchy-Schwartz inequality and Fubini's theorem, ψ_i is in $L^2(\mathcal{K})$, and the above shows that $\langle \phi_i, g_j \rangle = \langle \psi, f_i \otimes g_j \rangle = 0$, so since $([g_j])$ is an orthonormal basis of $L^2(\mathcal{K})$, we conclude that ϕ_i is ν almost surely zero. Since a countable union of null sets again is a null set, we find that ν almost surely, $\phi_i(t) = 0$ for all i. For any such t, $\psi(\cdot,t)$ is orthogonal to all f_i, showing that $\psi(\cdot,t)$ is zero μ almost surely. All in all, we can then conclude that ψ is zero $\mu \otimes \nu$ almost surely, as desired. □

B.5 Uniform integrability

Let $(X_i)_{i \in I}$ be a family of stochastic variables. We say that X_i is uniformly integrable if

$$\lim_{x \to \infty} \sup_{i \in I} E|X_i| 1_{(|X_i| > x)} = 0.$$

We will review some basic results about uniform integrability. We refer the results mainly for discrete sequences of variables, but the result extend to sequences indexed by $[0,\infty)$ as well.

Lemma B.5.1. *Let $(X_i)_{i \in I}$ be a family of stochastic variables. If (X_i) is bounded in $\mathcal{L}^p$ for some $p > 1$, then (X_i) is uniformly integrable.*

Proof. See Lemma 20.5 of Rogers & Williams (2000a). □

Lemma B.5.2. *Let $(X_i)_{i \in I}$ be uniformly integrable. Then (X_i) is bounded in $\mathcal{L}^1$.*

Proof. This follows from Lemma 20.7 of Rogers & Williams (2000a). □

Lemma B.5.3. *Let X_n be a sequence of stochastic variables, and let X be another variable. X_n converges in $\mathcal{L}^1$ to X if and only if X_n is uniformly integrable and converges in probability to X.*

Proof. This is Theorem 21.2 of Rogers & Williams (2000a). □

Lemma B.5.4. *Let X_n be a sequence of nonnegative integrable variables, and let X be another integrable variable. Assume that X_n converges almost surely to X. (X_n) is uniformly integrable if and only if EX_n converges to EX.*

Proof. First assume that (X_n) is uniformly integrable. Since X_n also converges in probability to X, it follows from Lemma B.5.3 that X_n converges in $\mathcal{L}^1$ to X, and therefore the means converge as well.

Conversely, assume that EX_n converges to EX. We have

$$\begin{aligned}|X_n - X| &= (X_n - X)^+ + (X_n - X)^- \\ &= X_n - X + (X_n - X)^- + (X_n - X)^- \\ &= X_n - X + 2(X_n - X)^-.\end{aligned}$$

Now, since X_n converges to X almost surely, $X_n - X$ converges to 0 almost surely. Because $x \mapsto x^-$ is continuous, $(X_n - X)^-$ converges to 0 almost surely. Clearly, X must be nonnegative. Thus, both X_n and X are nonnegative. Since $x \mapsto x^-$ is decreasing, this yields $(X_n - X)^- \leq (-X)^- = X$. By dominated convergence, we then obtain

$$\lim_n E|X_n - X| = \lim_n EX_n - EX + 2\lim_n E(X_n - X)^- = 2E\lim_n (X_n - X)^- = 0.$$

□

B.6 Miscellaneous

This final section of the appendix contains a variety of results which do not really fit anywhere else. We begin with some results on density of a certain class of stochastic variables in $\mathcal{L}^2$ spaces.

Lemma B.6.1. *Let $(E, \mathcal{E}, \mu)$ be a measure space and let $f_1, \ldots, f_n$ be real mappings on E. If $f_1, \ldots, f_n$ has exponential moments of all orders, then so does $\sum_{k=1}^n \lambda_k f_k$, where $\lambda \in \mathbb{R}^n$.*

Proof. We proceed by induction. For $n = 1$, there is nothing to prove. Assume that we have proved the claim for n. We find for $\lambda \in \mathbb{R}$ by the Cauchy-Schwartz inequality that

$$\begin{aligned} & \int \exp\left(\lambda \sum_{k=1}^{n+1} \lambda_k f_k\right) \mathrm{d}\mu \\ = & \int \exp\left(\sum_{k=1}^{n} \lambda\lambda_k f_k\right) \exp\left(\lambda\lambda_{n+1} f_{n+1}\right) \mathrm{d}\mu \\ \leq & \left(\int \exp\left(2\lambda \sum_{k=1}^{n} \lambda_k f_k\right) \mathrm{d}\mu\right)^{\frac{1}{2}} \left(\int \exp\left(2\lambda\lambda_{n+1} f_{n+1}\right) \mathrm{d}\mu\right)^{\frac{1}{2}}, \end{aligned}$$

which is finite by the induction hypothesis, proving the claim. □

Theorem B.6.2. *Let $(E, \mathcal{E}, \mu)$ be a measure space, where μ is a finite measure. Assume that $\mathcal{E}$ is generated by a family of real mappings $\mathcal{K}$, where all mappings in $\mathcal{K}$ have exponential moments of all orders. Then the variables*

$$\exp\left(\sum_{k=1}^{n} \lambda_k f_k\right),$$

where $f_1, \ldots, f_n \in \mathcal{K}$ and $\lambda \in \mathbb{R}^n$, are dense in $\mathcal{L}^2(\mathcal{E})$.

Proof. Let $\mathcal{H}$ denote the family of variables of the form $\exp\left(\sum_{k=1}^n \lambda_k f_k\right)$. Note that by Lemma B.6.1, all elements of $\mathcal{H}$ have moments of all orders, so $\mathcal{H}$ is a subset of $\mathcal{L}^2(\mathcal{E})$ and the conclusion of the theorem is well-defined.

Now, let $g \in \mathcal{L}^2(\mathcal{E})$ be given, orthogonal to $\mathcal{H}$. We wish to show that g is μ almost surely equal to zero. Consider $f_1, \ldots, f_n \in \mathcal{K}$. Since $g \in \mathcal{L}^2(\mathcal{E})$ and μ is finite, $g \in \mathcal{L}^1(\mathcal{E})$, so $g \cdot \mu$ is a well-defined signed measure. Define $\nu = (f_1, \ldots, f_n)(g \cdot \mu)$, then ν is also a signed measure, and for any $A \in \mathcal{B}_k$, we then have $\nu(A) = \int 1_A(f_1, \ldots, f_n) g \, \mathrm{d}\mu$. The Laplace transform of ν is

$$\begin{aligned} \int \exp\left(\sum_{k=1}^{n} \lambda_k x_k\right) \mathrm{d}\nu(x) &= \int \exp\left(\sum_{k=1}^{n} \lambda_k f_k\right) \mathrm{d}(g \cdot \mu) \\ &= \int \exp\left(\sum_{k=1}^{n} \lambda_k f_k\right) g \, \mathrm{d}\mu, \end{aligned}$$

and the last expression is equal to zero by assumption. Thus, ν has Laplace transform identically equal to zero, so by Lemma B.6.6, ν is the zero measure. That is, $\int 1_A(f_1, \ldots, f_n) g \, \mathrm{d}\mu$ for any $A \in \mathcal{B}_k$. Since the sets $((f_1, \ldots, f_n) \in A)$ is stable under intersections and generates $\mathcal{E}$, the Dynkin Lemma yields that $\int 1_G g \, \mathrm{d}\mu$ for any $G \in \mathcal{E}$. This shows that g is μ almost surely zero. □

Next, we review some basic results on measurability on product spaces.

Lemma B.6.3. *Let $(E, \mathcal{E})$ and $(K, \mathcal{K})$ be measurable spaces. Let $\mathbb{E}$ and $\mathbb{K}$ be generating systems for $\mathcal{E}$ and $\mathcal{K}$, respectively, closed under intersections. If $\mathbb{E}$ contains E and $\mathbb{K}$ contains K, then $\mathcal{E} \otimes \mathcal{K}$ is generated by the sets of the form $A \times B$, where $A \in \mathbb{E}$ and $B \in \mathbb{K}$.*

Proof. Let $\mathbb{H}$ be the σ-algebra generated by the sets of the form $A \times B$, where $A \in \mathbb{E}$ and $B \in \mathbb{K}$. We need to prove $\mathcal{E} \otimes \mathcal{K} = \mathbb{H}$. Clearly, $\mathbb{H} \subseteq \mathcal{E} \otimes \mathcal{K}$, we need to prove the other inclusion. To do so, note that letting $\mathbb{D}$ be the family of sets $A \times B$ where $A \in \mathcal{E}$ and $B \in \mathcal{K}$, $\mathcal{E} \otimes \mathcal{K}$ is generated by $\mathbb{D}$. It will therefore suffice to prove $\mathbb{D} \subseteq \mathbb{H}$.

To this end, let $\mathbb{F}$ be the family of sets $B \in \mathcal{K}$ such that $E \times B \in \mathbb{H}$. Then $\mathbb{F}$ is stable under complements and increasing unions, and since $E \in \mathbb{E}$, $\mathbb{K} \subseteq \mathbb{F}$. In particular, $K \in \mathbb{F}$. We conclude that $\mathbb{F}$ is a Dynkin class containing $\mathbb{K}$, therefore $\mathcal{K} \subseteq \mathbb{F}$. This shows that $E \times B \in \mathbb{H}$ for any $B \in \mathcal{K}$. Analogously, we can prove that $A \times K \in \mathbb{H}$ for any $A \in \mathcal{E}$. Letting $A \in \mathcal{E}$ and $B \in \mathcal{K}$, we then obtain $A \times B = (A \times K) \cap (E \times B) \in \mathbb{H}$, as desired. We conclude $\mathbb{D} \subseteq \mathbb{H}$ and therefore $\mathbb{H} = \mathcal{E} \otimes \mathcal{K}$. □

Lemma B.6.4. *Let $(E, \mathcal{E})$ and $(K, \mathcal{K})$ be measurable spaces. Let $\mathbb{E}$ and $\mathbb{K}$ be classes of real functions on E and K, respectively. Assume that there exists sequences of functions in $\mathbb{E}$ and $\mathbb{K}$, respectively, converging towards a nonzero constant. Assume that $\mathcal{E} = \sigma(\mathbb{E})$ and $\mathcal{K} = \sigma(\mathbb{K})$. The σ-algebra $\mathcal{E} \otimes \mathcal{K}$ is generated by the functions on the form $f \otimes g$, where $f \in \mathbb{E}$ and $g \in \mathbb{K}$.*

Proof. Let $\mathcal{H}$ be the σ-algebra generated by the functions $f \otimes g$, where $f \in \mathbb{E}$ and $g \in \mathbb{K}$. Clearly, $\mathcal{H} \subseteq \mathcal{E} \otimes \mathcal{K}$. We need to show the other inclusion. By definition, $f \otimes g$ is $\mathcal{H}$-measurable whenever $f \in \mathbb{E}$ and $g \in \mathbb{K}$. Now, let $f \in \mathbb{E}$ be fixed. By assumption, there is a sequence g_n in $\mathbb{K}$ converging towards a constant, say c. Then we find that

$$cf(x) = c(f(x) \lim g_n(y)) = \lim c(f \otimes g_n)(x, y),$$

showing that the function $(x, y) \mapsto f(x)$ is $\mathcal{H}$-measurable for any $f \in \mathbb{E}$. This means that $f^{-1}(A) \times K \in \mathcal{H}$ for any $f \in \mathbb{E}$ and $A \in \mathcal{B}$. Now, the sets $\{B \in \mathcal{E} | B \times K \in \mathcal{H}\}$ form a σ-algebra, and by what we have just shown, this σ-algebra contains the family $\{f^{-1}(A) | f \in \mathbb{E}, A \in \mathcal{B}\}$, which generates $\mathcal{E}$. We can therefore conclude that $B \times K \in \mathcal{H}$ for any $B \in \mathcal{E}$. In the samme manner, we can conclude that $E \times C \in \mathcal{H}$ for any $C \in \mathcal{K}$. Therefore, for $B \in \mathcal{E}$ and $C \in \mathcal{K}$, we find

$$B \times C = (B \times K) \cap (E \times C) \in \mathcal{H}.$$

These sets generate $\mathcal{E} \otimes \mathcal{K}$, and we may therefore conclude that $\mathcal{E} \otimes \mathcal{K} \subseteq \mathcal{H}$. □

Lemma B.6.5. *Let $(E, \mathcal{E}, \mu)$ and $(K, \mathcal{K}, \nu)$ be two measure spaces. Let f_n be a sequence of real mappings on $E \times K$, measurable with respect to $\mathcal{E} \otimes \mathcal{K}$. Let $p \geq 1$ and assume that f_n tends to f in $\mathcal{L}^p(\mathcal{E} \otimes \mathcal{K})$. There is a subsequence such that for ν almost all y, $\int |f_{n_k}(x, y)|^p \, \mathrm{d}\mu(x)$ tends to $\int |f(x, y)|^p \, \mathrm{d}\mu(x)$.*

Proof. Since $\int |f_n - f|^p \, \mathrm{d}(\mu \otimes \nu)$ tends to zero, the mapping in y on K given by $\int |f_n(x, y) - f(x, y)|^p \, \mathrm{d}\mu(x)$ tends to zero in $\mathcal{L}^1(\mathcal{K})$. By Lemma B.2.5, there is a subsequence converging ν almost surely. Let y be an element of K such that we have convergence. By the inverse triangle inequality of the norm in $\mathcal{L}^p(\mathcal{E})$, we obtain

$$\left| \left(\int |f_{n_k}(x, y)|^p \, \mathrm{d}\mu(x) \right)^{\frac{1}{p}} - \left(\int |f(x, y)|^p \, \mathrm{d}\mu(x) \right)^{\frac{1}{p}} \right| \leq \int |f_{n_k}(x, y) - f(x, y)|^p \, \mathrm{d}\mu(x),$$

showing that $\int |f_{n_k}(x, y)|^p \, \mathrm{d}\mu(x)$ tends to $\int |f(x, y)|^p \, \mathrm{d}\mu(x)$, as desired. □

Finally, a few assorted results.

Lemma B.6.6. *Let μ and ν be two bounded signed measures on $\mathbb{R}^k$ with Laplace transforms ψ and φ, respectively. If $\psi = \varphi$, then $\mu = \nu$.*

Proof. See Jensen (1992), Theorem C.1. □

Lemma B.6.7 (Doob-Dynkin lemma). *Let X be a variable on $(\Omega, \mathcal{F}, P)$ taking values in a polish space S. Let Y be another variable on the same probability space taking its values in a measurable space $(E, \mathcal{E})$. X is $\sigma(Y)$ measurable if and only if there exists a measurable mapping $\psi : E \to S$ such that $X = \psi(Y)$.*

Proof. See Dellacherie & Meyer (1975), Theorem 18 of Section 12.1. The original proof can be found in Doob (1953), p. 603. □

Lemma B.6.8. *Let $(E, \mathcal{E}, \mu)$ be a measure space and let $\mathcal{K}$ be a sub-σ-algebra of $\mathcal{E}$. Let f be $\mathcal{K}$ measurable. With $\mu_{|\mathcal{K}}$ denoting the restriction of μ to $\mathcal{K}$, we have*

$$\int f \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu_{|\mathcal{K}}.$$

Proof. If $A \in \mathcal{K}$, we clearly have

$$\int 1_A \, \mathrm{d}\mu = \mu(A) = \mu_{|\mathcal{K}}(A) = \int 1_A \, \mathrm{d}\mu_{|\mathcal{K}}.$$

By linearity, the result extends to simple $\mathcal{K}$ measurable mappings. By monotone convergence, it extends to nonnegative $\mathcal{K}$ measurable mappings. By splitting into positive and negative parts, the proof is finished. □

Lemma B.6.9. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathcal{N}$ be the null sets of the space and define $\mathcal{G} = \sigma(\mathcal{F}, \mathcal{N})$. Any $\mathcal{G}$ measurable real variable is almost surely equal to some $\mathcal{F}$ measurable variable.*

Proof. Let X be $\mathcal{G}$ measurable. Let ξ be a version of $X - E(X|\mathcal{F})$. If we can prove that ξ is almost surely zero, we are done. Note that with $\mathcal{A}$ denoting the almost sure sets of $\mathcal{F}$, $\mathcal{G} = \sigma(\mathcal{F}, \mathcal{A})$. Since both $\mathcal{F}$ and $\mathcal{A}$ contains Ω, the sets of the form $A \cap B$ where $A \in \mathcal{F}$ and $B \in \mathcal{A}$ form a generating system for $\mathcal{G}$ stable under intersections. By the Dynkin lemma, to prove that ξ is almost surely zero, it will suffice to prove that $E 1_{A \cap B} \xi$ for $A \in \mathcal{F}$ and $B \in \mathcal{N}$. To this end, we simply rewrite

$$E 1_{A \cap B} \xi) E 1_A \xi) = E 1_A (X - E(X|\mathcal{F})) = E 1_A X - E E(1_A X|\mathcal{F}) = 0.$$

The lemma is proved. □

Appendix C

Auxiliary results

This appendix contains some results which were developed in connection with the main parts of the thesis, but which eventually turned out not to fit in or became replaced by simpler arguments. Since the results have independent interest, their proofs are presented here.

C.1 A proof for the existence of $[M]$

In this section, we will discuss the existence of the quadratic variation for a continuous bounded martingale. This existence is proved in several different ways in the litteratur. Protter (2005) defines the quadratic variation directly through an integration-by-parts formula and derives the properties of the quadratic variation afterwards. Karatzas & Shreve (1988) uses the Doob-Meyer Theorem to obtain the quadratic variation as the compensator of the squared martingale. The most direct proof seems to be the one found in Rogers & Williams (2000b). This proof, however, still requires a substantial background, including preliminary results on stochastic integration with respect to elementary processes. Furthermore, even though the proof may *seem* short, a good deal of details are omitted, and the full proof is actually very lengthy and complicated, as may be seen in, say, Jacobsen (1989) or Sokol (2005), including a good deal of cumbersome calculations.

We will here present an alternative proof. The proof is a bit longer, but its structure is simpler, using quite elementary theory and requiring few prerequisites. In particular, no development of the stochastic integral is needed for the proof. The virtue of this is that it allows the development of the quadratic variation before the development of the stochastic integral. This yields a separation of concerns and therefore a more transparent structure of the theory. Before beginning the proof, we will outline the main ideas and the progression of the section.

In the remainder of the section, we work in the context of a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ and a *continuous bounded martingale M*.

Consider a partition π of $[0,t]$ and let $s \leq t$. By π_s, we denote the partition of $[0,s]$ given by $\pi_s = \{s\} \cup \pi \cap [0,s]$. Our plan is to show that the limit of the process given as $s \mapsto Q^{\pi_s}(M)$ converges uniformly in $\mathcal{L}^2$ on $[0,t]$ as the partition grows finer. We will then *define* the limit as the quadratic variation of M. Demonstrating this convergence will require two main lemmas, the relatively easy Lemma C.1.3 and the somewhat difficult Lemma C.1.4. After having defined the quadratic variation in this manner, we will argue that it is continuous and increasing and show that it can alternatively be characterised as the unique increasing process such that $M^2 - [M]$ is a uniformly integrable martingale. We will show how to extend the definition of the quadratic variation to continuous local martingales and show how to characterise the quadratic variation when M is a square-integrable martingale and when M is a continuous local martingale.

We begin with a few definitions concerning convergence of sequences indexed by partitions. Let $t \geq 0$, and let (x_π) be a family of elements of a pseudometric space (S,d), indexed by the partitions of $[0,t]$. We call (x_π) a net, inspired by the usual definition of nets as given in, say, Munkres (2000).

We say that x_π is convergent to x if it holds that for any $\varepsilon > 0$, there exists a partition π such that whenever $\varpi \supseteq \pi$ is another partition, $d(x_\varpi, x) < \varepsilon$. We say that x_π is cauchy if it holds that for any $\varepsilon > 0$, there exists a partition π such that whenever $\varpi, \vartheta \supseteq \pi$ are two other partitions, $d(x_\varpi, x_\vartheta) < \varepsilon$. Clearly, to show that a net is cauchy, it will suffice to show that for any $\varepsilon > 0$, there exists a partition such that whenever $\varpi \supseteq \pi$, $d(x_\varpi, x_\pi) \leq \varepsilon$.

Lemma C.1.1. *Let x_π be a net in (S,d). If x_π has a limit, it is unique. If (S,d) is complete and x_π is cauchy, then x_π has a limit.*

Proof. Assume that x_π has limits x and y. For any partition π, it clearly holds that $d(x,y) \leq d(x,x_\pi) + d(x_\pi,y)$. From this, it immediately follows that limits are unique.

Now assume that (S,d) is a complete pseudometric space and that x_π is cauchy. For each n, let π_n be the partition such that whenever $\varpi, \vartheta \supseteq \pi_n$, $d(x_\varpi, x_\vartheta) \leq \frac{1}{2^n}$. By expanding the partitions, we can assume without loss of generality that π_n is increasing. We then obtain $d(x_{\pi_n}, x_{\pi_{n+1}}) \leq \frac{1}{2^n}$, and in particular, for any $n \geq m$,

$$d(x_{\pi_n}, x_{\pi_m}) \leq \sum_{k=n+1}^{m} d(x_{\pi_{k-1}}, x_{\pi_k}) \leq \frac{1}{2^n}.$$

Therefore, x_{π_n} is a cauchy sequence, therefore convergent. Let x be the limit. We wish to show that x is the limit of the net x_π. By continuity of the metric, we obtain $d(x_{\pi_n}, x) \leq \frac{1}{2^n}$. Let $\varepsilon > 0$ be given. Let n be such that $\frac{1}{2^n} < \varepsilon$. Then we have, for any $\varpi \supseteq \pi_n$,

$$d(x_\varpi, x) \leq d(x_\varpi, x_{\pi_n}) + d(x_{\pi_n}, x) \leq \frac{2}{2^n} \leq 2\varepsilon.$$

Since ε was arbitrary, we have convergence. □

Lemma C.1.2. *Let x_π be convergent with limit x. Let ε_n be a positive sequence tending to zero. For each n, let π_n be a partition such that whenever $\varpi \supseteq \pi_n$, $d(x_\varpi, x) \leq \varepsilon_n$. Then x_{π_n} is a sequence converging to x.*

Proof. This is immediate. □

The following two concepts will play important roles in the coming arguments. We define the oscillation $\omega_M[s,t]$ of M over $[s,t]$ by $\omega_M[s,t] = \sup_{s \leq u,r \leq t} |M_u - M_r|$. With $\pi = (t_0, \ldots, t_n)$ a partition, by $\delta^\pi(M)$ we denote the modulus of continuity over π, defined as $\delta^\pi(M) = \max_{0 \leq i < j \leq n} \omega_M[t_i, t_j]$.

We are now ready to begin the work proper on the existence of the quadratic variation. In the following, we will generically write $\pi = (t_0, \ldots, t_n)$ and $\varpi = (s_0, \ldots, s_m)$. These will be our generic partitions.

We begin by proving a few lemmas. With these lemmas, we will be able to prove the existence of the quadratic variation. Recall that with π a partition of $[s,t]$, we define $\pi_u = \{u\} \cup \pi \cap [s,u]$ for $s \leq u \leq t$. π_u is then a partition of $[s,u]$. Note that

$$Q^{\pi_u}(M) = \sum_{k=1}^{n} (M_{t_k \wedge u} - M_{t_{k-1} \wedge u})^2.$$

Lemma C.1.3. *Let π be a partition of $[s,t]$. It holds that*

$$EQ^\pi(M)^2 \leq 3E\omega_M^2[s,t]Q^\pi(M).$$

We also have the weaker bound $EQ^\pi(M)^2 \leq 12\|M^*\|_\infty^2 E(M_t - M_s)^2$.

Proof. We first expand the square to obtain

$$\begin{aligned}
& EQ^\pi(M)^2 \\
= \; & E\left(\sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^2\right)^2 \\
= \; & E\sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^4 + 2E\sum_{i=1}^{n-1}\sum_{j=i+1}^n (M_{t_i} - M_{t_{i-1}})^2 (M_{t_j} - M_{t_{j-1}})^2.
\end{aligned}$$

The last term can be computed using Lemma 2.4.12,

$$\begin{aligned}
& 2E\sum_{i=1}^{n-1}\sum_{j=i+1}^n (M_{t_i} - M_{t_{i-1}})^2 (M_{t_j} - M_{t_{j-1}})^2 \\
= \; & 2\sum_{i=1}^{n-1} E\left((M_{t_i} - M_{t_{i-1}})^2 E\left(\left.\sum_{j=i+1}^n (M_{t_j} - M_{t_{j-1}})^2 \right| \mathcal{F}_{t_i}\right)\right) \\
= \; & 2\sum_{i=1}^{n-1} E\left((M_{t_i} - M_{t_{i-1}})^2 E((M_t - M_{t_i})^2 | \mathcal{F}_{t_i})\right) \\
= \; & 2\sum_{i=1}^{n-1} E(M_{t_i} - M_{t_{i-1}})^2 (M_t - M_{t_i})^2 \\
\leq \; & 2E\omega_M^2[s,t]\sum_{i=1}^{n-1} (M_{t_i} - M_{t_{i-1}})^2 \\
= \; & 2E\omega_M^2[s,t]Q^\pi(M).
\end{aligned}$$

We then find,

$$\begin{aligned}
EQ^\pi(M)^2 &\leq E\sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^4 + 2E\omega_M^2[s,t]Q^\pi(M) \\
&\leq E\omega_M^2[s,t]\sum_{k=1}^n (M_{t_k} - M_{t_{k-1}})^2 + 2E\omega_M^2[s,t]Q^\pi(M) \\
&\leq 3E\omega_M^2[s,t]Q^\pi(M),
\end{aligned}$$

as desired. Since $\omega_M^2[s,t] \leq 2\|M^*\|_\infty$, we also get

$$EQ^\pi(M)^2 \leq 12\|M^*\|_\infty^2 EQ^\pi(M) = 12\|M^*\|_\infty^2 E(M_t - M_s)^2.$$

□

Lemma C.1.4. *Let π be a partition of $[s,t]$. Then*

$$E \sup_{s \le u \le t} \left((M_u - M_s)^2 - Q^{\pi_u}(M)\right)^2 \le 17 E\omega_M^2[s,t] \left((M_t - M_s)^2 + Q^{\pi}(M)\right).$$

Proof. Considering $((M_u - M_s)^2 - Q^{\pi_u}(M))^2$, we note that both terms in the square are nonnegative. Therefore, the mixed term in the expansion of the square is nonpositive and we obtain the bound

$$E \sup_{s \le u \le t} \left((M_u - M_s)^2 - Q^{\pi_u}(M)\right)^2 \le E \sup_{s \le u \le t} (M_u - M_s)^4 + E \sup_{s \le u \le t} Q^{\pi_u}(M)^2.$$

The proof of the lemma will proceed by first making an estimate that allows us to get rid of the suprema above. Then, we will evaluate the resulting expression to obtain the result of the lemma. Specifically, we will remove the suprema by proving

$$\begin{aligned} & E \sup_{s \le u \le t} (M_u - M_s)^4 + E \sup_{s \le u \le t} Q^{\pi_u}(M)^2 \\ \le \quad & 5E(M_t - M_s)^4 + 2E\omega_M^2[s,t]Q^{\pi}(M) + 5EQ^{\pi}(M)^2. \end{aligned}$$

This is a somewhat intricate task. We will split it into two parts, yielding the following total structure of the proof: We first estimate the term $E \sup_{s \le u \le t}(M_u - M_s)^4$, then estimate the term $E \sup_{s \le u \le t} Q^{\pi_u}(M)^2$, and ultimately, we finalize our estimates and conclude.

Step 1: Removing the suprema, first term. We begin by considering the term given by $E \sup_{s \le u \le t}(M_u - M_s)^4$. Note that whenever $0 \le u \le r$, we have the equality $E(M_{s+r} - M_s | \mathcal{F}_{s+u}) = M_{s+u} - M_s$. Therefore, $M_{s+u} - M_s$ is a martingale in u with respect to the filtration $(\mathcal{F}_{s+u})_{u \ge 0}$. Since the dual exponent of 4 is $\frac{4}{3}$, the Doob $\mathcal{L}^p$ inequality yields

$$E \sup_{s \le u \le t} (M_u - M_s)^4 \le \left(\frac{4}{3}\right)^4 E(M_t - M_s)^4 \le 5E(M_t - M_s)^4,$$

removing the supremum.

Step 2: Removing the suprema, second term. Turning our attention to the second term, $E\sup_{s\le r\le t} Q^{\pi_r}(M)^2$, let $s \le u \le t$. We know that

$$
\begin{aligned}
& Q^{\pi_u}(M)^2 \\
= & \left(\sum_{k=1}^n (M_{t_k\wedge u} - M_{t_{k-1}\wedge u})^2\right)^2 \\
= & \sum_{k=1}^n (M_{t_k\wedge u} - M_{t_{k-1}\wedge u})^4 + \sum_{i\ne j}^n (M_{t_i\wedge u} - M_{t_{i-1}\wedge u})^2 (M_{t_j\wedge u} - M_{t_{j-1}\wedge u})^2.
\end{aligned}
$$

Therefore, we also have

$$
\begin{aligned}
& E\sup_{s\le u\le t} Q^{\pi_u}(M)^2 \\
\le & E\sup_{s\le u\le t} \sum_{k=1}^n (M_{t_k\wedge u} - M_{t_{k-1}\wedge u})^4 \\
+ & E\sup_{s\le u\le t} \sum_{i\ne j}^n (M_{t_i\wedge u} - M_{t_{i-1}\wedge u})^2 (M_{t_j\wedge u} - M_{t_{j-1}\wedge u})^2.
\end{aligned}
$$

We need to bound each of the two terms above. We begin by developing a bound for the first term. Applying Doob's $\mathcal{L}^p$ inequality used with the martingale $M^{t_k} - M^{t_{k-1}}$ yields

$$
\begin{aligned}
E\sup_{s\le u\le t} \sum_{k=1}^n (M_{t_k\wedge u} - M_{t_{k-1}\wedge u})^4 & \le \sum_{k=1}^n E\sup_{s\le u\le t} (M^{t_k}_u - M^{t_{k-1}}_u)^4 \\
& \le \sum_{k=1}^n \frac{256}{81} E(M^{t_k}_t - M^{t_{k-1}}_t)^4 \\
& \le 5\sum_{k=1}^n E(M_{t_k} - M_{t_{k-1}})^4.
\end{aligned}
$$

Next, we consider the second sum, that is,

$$
E\sup_{s\le u\le t} \sum_{i\ne j}^n (M_{t_i\wedge u} - M_{t_{i-1}\wedge u})^2 (M_{t_j\wedge u} - M_{t_{j-1}\wedge u})^2.
$$

We start out by fixing $s \le u \le t$. Assume that $u < t$, the case $u = t$ will trivially satisfy the bound we are about to prove. Letting c be the index such that $t_{c-1} \le u < t_c$, we can separate the terms where one of the indicies is equal to c and obtain three types of terms: Those where i is distinct from c and j is equal to c, those where i is equal to c and j is distinct from c and those where i and j are distinct and both distinct from c as well. Note that since we are only considering off-diagonal terms, we do not need to consider the case where i and j are both equal to c. We obtain

$$\begin{aligned}
&\sum_{i\neq j}^{n}(M_{t_i\wedge u}-M_{t_{i-1}\wedge u})^2(M_{t_j\wedge u}-M_{t_{j-1}\wedge u})^2\\
\leq\ &\sum_{i\neq c}^{n}(M_{t_i\wedge u}-M_{t_{i-1}\wedge u})^2(M_u-M_{t_{c-1}})^2\\
+\ &\sum_{j\neq c}^{n}(M_u-M_{t_{c-1}})^2(M_{t_j\wedge u}-M_{t_{j-1}\wedge u})^2\\
+\ &\sum_{i\neq j,i\neq c,j\neq c}^{n}(M_{t_i\wedge u}-M_{t_{i-1}\wedge u})^2(M_{t_j\wedge u}-M_{t_{j-1}\wedge u})^2.
\end{aligned}$$

Now note that when $i<c$, $t_i\wedge u=t_i$ and $t_{i-1}\wedge u=t_{i-1}$. And when $i>c$, $t_i\wedge u=u$ and $t_{i-1}\wedge u=u$. Therefore,

$$\begin{aligned}
\sum_{i\neq c}^{n}(M_{t_i\wedge u}-M_{t_{i-1}\wedge u})^2(M_u-M_{t_{c-1}})^2 &= \sum_{i<c}^{n}(M_{t_i}-M_{t_{i-1}})^2(M_u-M_{t_{c-1}})^2\\
&\leq \omega_M^2[s,t]\sum_{i<c}^{n}(M_{t_i}-M_{t_{i-1}})^2\\
&\leq \omega_M^2[s,t]Q^\pi(M),
\end{aligned}$$

and analogously for the second term. Likewise,

$$\begin{aligned}
&\sum_{i\neq j,i\neq c,j\neq c}^{n}(M_{t_i\wedge u}-M_{t_{i-1}\wedge u})^2(M_{t_j\wedge u}-M_{t_{j-1}\wedge u})^2\\
=\ &\sum_{i\neq j,i<c,j<c}^{n}(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}-M_{t_{j-1}})^2\\
\leq\ &\sum_{i\neq j}^{n}(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}-M_{t_{j-1}})^2.
\end{aligned}$$

All in all, inserting the bounds just obtained for each of the three terms in the sum, we obtain the pointwise bound

$$\begin{aligned}
&\sum_{i\neq j}^{n}(M_{t_i\wedge u}-M_{t_{i-1}\wedge u})^2(M_{t_j\wedge u}-M_{t_{j-1}\wedge u})^2\\
\leq\ &2\omega_M^2[s,t]Q^\pi(M)+\sum_{i\neq j}^{n}(M_{t_i}-M_{t_{i-1}})^2(M_{t_j}-M_{t_{j-1}})^2.
\end{aligned}$$

Since this bound holds for all $s \leq u \leq t$, we finally obtain

$$\begin{aligned}
& E \sup_{s \leq u \leq t} \sum_{i \neq j}^{n} (M_{t_i \wedge u} - M_{t_{i-1} \wedge u})^2 (M_{t_j \wedge u} - M_{t_{j-1} \wedge u})^2 \\
\leq \quad & 2E\omega_M^2[s,t] Q^\pi(M) + E \sum_{i \neq j}^{n} (M_{t_i} - M_{t_{i-1}})^2 (M_{t_j} - M_{t_{j-1}})^2.
\end{aligned}$$

Combining our results, we finally obtain

$$\begin{aligned}
& E \sup_{s \leq u \leq t} Q^{\pi_u}(M)^2 \\
\leq \quad & 5 \sum_{k=1}^{n} E(M_{t_k} - M_{t_{k-1}})^4 + 2E\omega_M^2[s,t]Q^\pi(M) + E \sum_{i \neq j}^{n} (M_{t_i} - M_{t_{i-1}})^2 (M_{t_j} - M_{t_{j-1}})^2 \\
\leq \quad & 2E\omega_M^2[s,t]Q^\pi(M) + 5EQ^\pi(M)^2.
\end{aligned}$$

We have now developed the bounds announced earlier, and may all in all conclude

$$\begin{aligned}
& E \sup_{s \leq r \leq t} (M_r - M_s)^4 + E \sup_{s \leq r \leq t} Q^{\pi_r}(M)^2 \\
\leq \quad & 5E(M_t - M_s)^4 + 2E\omega_M^2[s,t]Q^\pi(M) + 5EQ^\pi(M)^2,
\end{aligned}$$

as promised.

Step 3: Final estimates and conclusion. We will now make an estimate for the first and last terms in the expression above, uncovering the result of the lemma. For the first term, we easily obtain

$$5E(M_t - M_s)^4 \leq 5E\omega_M^2[s,t](M_t - M_s)^2.$$

For the other term, Lemma C.1.3 yields

$$5EQ^\pi(M)^2 \leq 15E\omega_M^2[s,t]Q^\pi(M).$$

Combining our findings, we conclude that

$$\begin{aligned}
& E \sup_{s \leq u \leq t} \left((M_u - M_s)^2 - Q^{\pi_u}(M) \right)^2 \\
\leq \quad & 5E(M_t - M_s)^4 + 2E\omega_M^2[s,t]Q^\pi(M) + 5EQ^\pi(M)^2 \\
\leq \quad & 5E\omega_M^2[s,t](M_t - M_s)^2 + 2E\omega_M^2[s,t]Q^\pi(M) + 15E\omega_M^2[s,t]Q^\pi(M) \\
\leq \quad & 17E\omega_M^2[s,t] \left((M_t - M_s)^2 + Q^\pi(M) \right),
\end{aligned}$$

which was precisely the proclaimed bound. $\square$

Lemma C.1.5. *Consider the space of processes on $[0,t]$ endowed with the norm given by $|||\cdot||| = \|\|\cdot\|_\infty\|_2$. We denote convergence in $|||\cdot|||$ as uniform $\mathcal{L}^2$ convergence. This space is complete.*

Proof. Assume that X_n is a cauchy sequence under $|||\cdot|||$. Since a cauchy sequence with a convergent subsequence is convergent, it will suffice to show that X_{n_k} has a convergent subsequence.

We know that for any $\varepsilon > 0$, there exists N such that for $n, m \geq N$, it holds that $\|\|X_n - X_m\|_\infty\|_2 \leq \varepsilon$. In particular, using $\frac{1}{8^k}$ as our ε, there exists n_k such that for any $n, m \geq n_k$,

$$P\left(\|X_n - X_m\|_\infty > \frac{1}{2^k}\right) \leq 4^k \|\|X_n - X_m\|_\infty\|_2 \leq \frac{1}{2^k}.$$

We can assume without loss of generality that n_k is increasing. We then obtain that $P(\|X_{n_k} - X_{n_{k+1}}\|_\infty > \frac{1}{2^k}) \leq \frac{1}{2^k}$, so $\|X_{n_k} - X_{n_{k+1}}\|_\infty \leq \frac{1}{2^k}$ from a point onwards almost surely, by the Borel-Cantelli Lemma. Then we also find for $i > j$

$$\|X_{n_i} - X_{n_j}\|_\infty \leq \sum_{k=i+1}^{j} \|X_{n_{k-1}} - X_{n_k}\|_\infty \leq \frac{1}{2^k},$$

so X_{n_k} is almost surely uniformly cauchy. Therefore, it is almost surely uniformly convergent. Let X be the limit. We wish to prove that X is the limit of X_{n_k} in $|||\cdot|||$. Fixing k, we have

$$\begin{aligned}
|||X_{n_k} - X||| &= \|\|X_{n_k} - X\|_\infty\|_2 \\
&= \|\lim_m \|X_{n_k} - X_{n_m}\|_\infty\|_2 \\
&= \left(E\left(\liminf_m \|X_{n_k} - X_{n_m}\|_\infty\right)^2\right)^{\frac{1}{2}} \\
&= \left(E \liminf_m (\|X_{n_k} - X_{n_m}\|_\infty)^2\right)^{\frac{1}{2}} \\
&= \liminf_m \|\|X_{n_k} - X_{n_m}\|_\infty\|_2.
\end{aligned}$$

Now note that X_{n_k} also is a cauchy sequence. Let $\varepsilon > 0$ be given and select N in accordance with the cauchy property of X_{n_k}. Let $k \geq N$, we then obtain

$$\liminf_m \|\|X_{n_k} - X_{n_m}\|_\infty\|_2 = \sup_{i \geq N} \inf_{m \geq i} \|\|X_{n_k} - X_{n_m}\|_\infty\|_2 \leq \varepsilon.$$

We have now shown that for $k \geq N$, $|||X_{n_k} - X||| \leq \varepsilon$. Thus, X_{n_k} converges to X, and therefore X_n converges to X as well. $\square$

We are now ready to prove the convergence result needed to argue the existence of the quadratic variation. Letting π be a partition of $[s,t]$, we earlier defined π_u by $\pi_u = \{u\} \cup \pi \cap [s,u]$ for $s \leq u \leq t$. We now extend this definition by defining π_u be empty for $u < s$ and letting $\pi_u = \pi$ for $u > t$. We use the convention that $Q^{\emptyset}(M) = 0$.

Theorem C.1.6. *Let $t \geq 0$, and let π denote a generic partition of $[0,t]$. The mapping $s \mapsto Q^{\pi_s}(M)$ is uniformly convergent on $[0,t]$ in $\mathcal{L}^2$.*

Proof. Since uniform $\mathcal{L}^2$ convergence is a complete mode of convergence by Lemma C.1.5, it will suffice to show that $s \mapsto Q^{\pi_s}(M)$ is a cauchy net with respect to uniform $\mathcal{L}^2$ convergence. We therefore consider the distance between elements $u \mapsto Q^{\pi_u}(M)$ and $u \mapsto Q^{\varpi_u}(M)$, where $\varpi \supseteq \pi$. As usual, $\pi = (t_0, \ldots, t_n)$. We put $\varpi = (s_0, \ldots, s_m)$ and let j_k denote the unique element of ϖ such that $s_{j_k} = t_k$. We define the partition ϖ^k of $[t_{k-1}, t_k]$ by $\varpi_k = (s_{j_{k-1}}, \ldots, s_{j_k - 1})$ and obtain

$$\begin{aligned} Q^{\varpi_u}(M) - Q^{\pi_u}(M) &= \sum_{k=1}^{m} (M^u_{s_k} - M^u_{s_{k-1}})^2 - \sum_{k=1}^{n} (M^u_{t_k} - M^u_{t_{k-1}})^2 \\ &= \sum_{k=1}^{n} \sum_{i=j_{k-1}+1}^{j_k} (M^u_{s_i} - M^u_{s_{i-1}})^2 - \sum_{k=1}^{n} (M^u_{t_k} - M^u_{t_{k-1}})^2 \\ &= \sum_{k=1}^{n} Q^{\varpi^k_u}(M) - \sum_{k=1}^{n} (M^u_{t_k} - M^u_{t_{k-1}})^2. \end{aligned}$$

This implies that

$$\begin{aligned} &\| \sup_{u \leq t} |Q^{\varpi_u}(M) - Q^{\pi_u}(M)| \|_2 \\ = &\left\| \sup_{u \leq t} \left| \sum_{k=1}^{n} Q^{\varpi^k_u}(M) - \sum_{k=1}^{n} (M^u_{t_k} - M^u_{t_{k-1}})^2 \right| \right\|_2 \\ \leq &\left\| \sum_{k=1}^{n} \sup_{u \leq t} \left| Q^{\varpi^k_u}(M) - (M^u_{t_k} - M^u_{t_{k-1}})^2 \right| \right\|_2 \\ \leq &\sum_{k=1}^{n} \left\| \sup_{u \leq t} \left| Q^{\varpi^k_u}(M) - (M^u_{t_k} - M^u_{t_{k-1}})^2 \right| \right\|_2 \\ \leq &\sum_{k=1}^{n} \left\| \sup_{t_{k-1} \leq u \leq t_k} \left| Q^{\varpi^k_u}(M) - (M^u_{t_k} - M^u_{t_{k-1}})^2 \right| \right\|_2. \end{aligned}$$

We have been able to reduce the supremum to $[t_{k-1}, t_k]$ since the expression in the supremum is constant for $u \leq t_{k-1}$ and for $u \geq t_k$, respectively. Now, using Lemma

C.1.4, we conclude

$$\begin{aligned}
&\sum_{k=1}^{n}\left\|\sup_{t_{k-1}\le u\le t_k}\left|Q^{\varpi_u^k}(M)-(M_{t_k}^u-M_{t_{k-1}}^u)^2\right|\right\|_2\\
\le\ &\sum_{k=1}^{n}17E\omega_M^2[t_{k-1},t_k]\left((M_{t_k}-M_{t_{k-1}})^2+Q^{\varpi_k}(M)\right)\\
\le\ &\sum_{k=1}^{n}17E\delta_M^2[s,t]\left((M_{t_k}-M_{t_{k-1}})^2+Q^{\varpi_k}(M)\right).
\end{aligned}$$

Since the union of the ϖ_k is ϖ, we obtain, applying the Cauchy-Schwartz inequality and Lemma C.1.3,

$$\begin{aligned}
&\sum_{k=1}^{n}17E\delta_M^2[s,t]\left((M_{t_k}-M_{t_{k-1}})^2+Q^{\varpi_k}(M)\right)\\
=\ &17E\delta_M^2[s,t](Q^{\pi}(M)+Q^{\varpi}(M))\\
\le\ &17\|\delta_M^2[s,t]\|_2\|Q^{\pi}(M)+Q^{\varpi}(M)\|_2\\
\le\ &17\|\delta_M^2[s,t]\|_2(\|Q^{\pi}(M)\|_2+\|Q^{\varpi}(M)\|_2)\\
\le\ &34\|\delta_M^2[s,t]\|_2\left(12\|M^*\|_\infty^2E(M_t-M_s)^2\right)^{\frac{1}{2}}\\
=\ &\left(34\sqrt{12}\|M^*\|_\infty\|M_t-M_s\|_2\right)\|\delta_M^2[s,t]\|_2.
\end{aligned}$$

As the mesh tends to zero, the continuity of M yields that $\delta^\pi(M)$ tends to zero. By boundedness of M, $E\delta^\pi(M)^4$ tends to zero as well. Therefore, if π is chosen sufficiently fine, the above can be made as small as desired. Therefore, we conclude that $s\mapsto Q^{\pi_s}(M)$ is uniformly $\mathcal{L}^2$ cauchy, and so it is convergent. □

Theorem C.1.7. *There exists a unique continuous, increasing and adapted process $[M]$ such that for any $t\ge 0$, $[M]_t$ is the uniform $\mathcal{L}^2$ limit in π of the processes $s\mapsto Q^{\pi_s}(M)$ on $[0,t]$.*

Proof. Letting $n\ge 1$ and letting π be a generic partition of $[0,n]$, by Lemma C.1.6, $u\mapsto Q^{\pi_u}(M)$ is uniformly $\mathcal{L}^2$ convergent. Let X^n be the limit. X^n is then a process on $[0,n]$. By Lemma C.1.2, there is a sequence of partitions π^n such that $u\mapsto Q^{\pi_u^n}(M)$ converges uniformly in $\mathcal{L}^2$ to X^n. Therefore, there is a subsequence converging almost surely uniformly. Since $u\mapsto Q^{\pi_u^n}(M)$ is continuous and adapted, X^n is continuous and adapted as well.

Now, X^n is also the uniform $\mathcal{L}^2$ limit of $u\mapsto Q^{\pi_u}(M)$, $u\le t$, where π is a generic partition of $[0,n+1]$. By uniqueness of limits, X^n and X^{n+1} are indistinguishable on $[0,n]$. By the Pasting Lemma, there exists a process $[M]$ such that $[M]^n = X^n$. Since X^n is continuous and adapted, $[M]$ is continuous and adapted as well. Since $[M]_t$ is the limit of larger partitions than $[M]_s$ whenver $s \le t$, $[M]_t \ge [M]_s$ and $[M]$ is increasing. And $[M]$ has the property stated in the lemma of being the uniform $\mathcal{L}^2$ limit on compacts of quadratic variations over partitions. □

Lemma C.1.8. *The process $M_t^2 - [M]_t$ is a uniformly integrable martingale.*

Proof. We proceed in two steps. We first show that $M_t^2 - [M]_t$ is a martingale, and then proceed to show that it is uniformly integrable.

Step 1: The martingale property. Let $0 \le s \le t$. Let π be a partition of $[0,t]$. Define $\varpi = \{s\} \cup \pi \cap [s,t]$. We then obtain

$$\begin{aligned} E(Q^\pi(M)|\mathcal{F}_s) &= E(Q^{\pi_s}(M) + Q^\varpi(M)|\mathcal{F}_s) \\ &= Q^{\pi_s}(M) + E((M_t - M_s)^2|\mathcal{F}_s) \\ &= Q^{\pi_s}(M) + E(M_t^2|\mathcal{F}_s) - M_s^2. \end{aligned}$$

This shows that

$$E(M_t^2 - Q^{\pi_t}(M)|\mathcal{F}_s) = E(M_t^2 - Q^\pi(M)|\mathcal{F}_s) = M_s^2 - Q^{\pi_s}(M),$$

so the process $M_s^2 - Q^{\pi_s}(M)$ is a martingale on $[s,t]$. Since $Q^{\pi_s}(M)$ converges in $\mathcal{L}^2$ to $[M]_s$ as π becomes finer, and conditional expectations are $\mathcal{L}^2$ continuous, we conclude

$$\begin{aligned} E(M_t^2 - [M]_t|\mathcal{F}_s) &= \lim_\pi E(M_t^2 - Q^{\pi_t}(M)|\mathcal{F}_s) \\ &= \lim_\pi M_s^2 - Q^{\pi_s}(M) \\ &= M_s^2 - [M]_s, \end{aligned}$$

almost surely, showing the martingale property.

Step 2: Uniform integrability. By the continuity of the $\mathcal{L}^2$ norm, we can use Lemma C.1.3 to obtain

$$E[M]_t^2 = \lim_\pi E(Q^\pi(M)^2) \le 12\|M^*\|_\infty^2 E M_t^2 \le 12\|M^*\|_\infty^4,$$

so $[M]$ is bounded in $\mathcal{L}^2$, therefore uniformly integrable. Since M^2 is bounded, it is clearly uniformly integrable. We can therefore conclude that $M_t^2 - [M]_t$ is a uniformly integrable martingale. □

Our final task is localise our results from continuous bounded martingales to continuous local martingales and investigate how to characterize the quadratic variation in the local cases.

Lemma C.1.9. *Let π be a partition of $[s,t]$. Let τ be a stopping time with values in $[s,\infty)$. Then*

$$Q^{\pi_\tau}(M) = Q^\pi(M^\tau).$$

Proof. This follows immediately from

$$\begin{aligned} Q^{\pi_\tau}(M) &= \sum_{k=1}^n (M^2_{t_k\wedge\tau} - M^2_{t_{k-1}\wedge\tau}) \\ &= \sum_{k=1}^n ((M^\tau_{t_k})^2 - (M^\tau_{t_{k-1}})^2) \\ &= Q^\pi(M^\tau). \end{aligned}$$

□

Lemma C.1.10. *Let M be a continuous bounded martingale, and let τ be a stopping time. Then $[M]^\tau = [M^\tau]$.*

Proof. Let $t \geq 0$. As usual, letting π be a generic partition of $[0,t]$, we know that $[M]$ is the uniform $\mathcal{L}^2$ limit on $[0,t]$ of $Q^{\pi_s}(M)$.

$$\begin{aligned} E\left([M]^\tau_t - Q^\pi(M^\tau)\right)^2 &= E\left([M]^\tau_t - Q^\pi(M)^\tau\right)^2 \\ &= E\left([M]^\tau_t - Q^{\pi_\tau}(M)\right)^2 \\ &\leq E\left(\sup_{s\leq t}|[M]_s - Q^{\pi_s}(M)|\right)^2, \end{aligned}$$

which tends to zero. Thus, $[M]^\tau_t$ is the limit of $Q^\pi(M^\tau)$. Since $[M^\tau]_t$ is also the limit of $Q^\pi(M^\tau)$, we must have $[M]^\tau_t = [M^\tau]_t$. Since the processes are continuous, we conclude that $[M]^\tau = [M^\tau]$ up to indistinguishability. □

Corollary C.1.11. *Let M be a continous local martingale. There exists a unique continuous, increasing and adapted process $[M]$ such that if τ_n is any localising sequence such that M^{τ_n} is a continuous bounded martingale, then $[M]^{\tau_n} = [M^{\tau_n}]$.*

Proof. This follows from Theorem 2.5.11. □

Lemma C.1.12. *Let A be any increasing process with $A_0 = 0$. Then A is uniformly integrable if and only if $A_\infty \in \mathcal{L}^1$.*

Proof. First note that the criterion in the lemma is well-defined, since A is convergent and therefore has a well-defined limit in $[0, \infty]$. We show each implication separately. Assume $A_\infty \in \mathcal{L}^1$. We then obtain, since A is increasing,

$$\lim_{x \to \infty} \sup_{t \geq 0} E|A_t| 1_{(|A_t| > x)} \leq \lim_{x \to \infty} \sup_{t \geq 0} E|A_\infty| 1_{(|A_\infty| > x)} = 0,$$

by dominated convergence. Therefore, A is uniformly integrable. On the other hand, if A is uniformly integrable, we find by monotone convergence

$$E|A_\infty| = \lim_t E|A_t| \leq \sup_{t \geq 0} E|A_t|,$$

which is finite, since A is uniformly integrable and therefore bounded in $\mathcal{L}^1$. □

Theorem C.1.13. *Let $M \in \mathbf{c}\mathcal{M}_0^2$. The quadratic variation is the unique process $[M]$ such that $M^2 - [M]$ is a uniformly integrable martingale.*

Proof. Let τ_n be a localising sequence for M such that M^{τ_n} is a continuous bounded martingale. We will use Lemma 2.4.13 to prove the claim. Therefore, we first prove that $M^2 - [M]$ has an almost surely finite limit. We already know that M^2 has a limit, it will therefore suffice to show that this is the case for $[M]$. This follows since

$$\begin{aligned} E[M]_\infty &= E \lim_n [M]_{\tau_n} \\ &= \lim_n E[M]_{\tau_n} \\ &= \lim_n E[M^{\tau_n}]_\infty \\ &= \lim_n E(M^{\tau_n})^2_\infty \\ &\leq \lim EM^2_\infty, \end{aligned}$$

which is finite. Now let a stopping time τ be given. Using that $M \in \mathbf{c}\mathcal{M}_0^2$ and applying monotone convergence, we obtain

$$\begin{aligned} E(M_\tau^2 - [M]_\tau) &= EM_\tau^2 - E[M]_\tau \\ &= E \lim_n M^2_{\tau_n \wedge \tau} - E \lim_n [M]_{\tau_n \wedge \tau} \\ &= \lim_n EM^2_{\tau_n \wedge \tau} - \lim_n E[M]_{\tau_n \wedge \tau} \\ &= -\lim_n E(M^{\tau_n})^2_\tau - [M^{\tau_n}]_\tau \\ &= 0. \end{aligned}$$

By Lemma 2.4.13, $M^2 - [M]$ is a uniformly integrable martingale. $\square$

Corollary C.1.14. *Let $M \in \mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$. The quadratic variation is the unique increasing process $[M]$ such that $M^2 - [M]$ is a continuous local martingale.*

C.2 A Lipschitz version of Urysohn's Lemma

In this section, we prove an version of Urysohn's Lemma giving a Lipschitz condition on the mapping supplied. This result could be used in the proof of Lemma 4.3.3, but the proof presented there seems simpler. However, since the extension of Urysohn's Lemma has independent interest, we present it here.

Lemma C.2.1. *Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$ and let d be the induced metric. Let A be a closed set and let $x \in A^c$. Then $d(x, A) = d(x, y)$ for some $y \in \partial A$.*

Proof. First consider the case where A is bounded. Then A is compact. The mapping $y \mapsto d(x, y)$ is continuous, and therefore attains its infimum over A. Thus, there exists $y \in A$ such that $d(x, A) = d(x, y)$. Assume that $y \in A^\circ$, we will seek a contradiction. Let $\varepsilon > 0$ be such that $B_\varepsilon(y) \subseteq A$. There exists $0 < \lambda \leq 1$ with $\lambda x + (1-\lambda)y \in B_\varepsilon(y)$. We then find

$$\|\lambda x + (1-\lambda)y - x\| \leq (1-\lambda)\|x - y\| < \|x - y\| = d(x, y),$$

and therefore $d(x, \lambda x + (1 - \lambda y)) < d(x, y)$, which is the desired contradiction. Thus, $y \notin A^\circ$, and since A is closed, we obtain $y \in \partial A$.

Next, consider the case where A is unbounded. Then, there exists a sequence z_n in A with norms converging to infinity. We define

$$K_n = \{y \in A | \|y - x\| \leq \|z_n - x\|\}.$$

Since the norm is continuous, K_n is closed. And any $y \in K_n$ satisfies the relation $\|y\| \leq \|y - x\| + \|x\| \leq \|z - x\| + \|x\|$, so K_n is bounded. For any $y \in A \setminus K_n$, we have $d(y, x) > d(x, z_n) \geq d(x, A)$ and therefore $d(x, A) = d(x, K_n)$.

From what we already have shown, there is $y_n \in \partial K_n$ such that $d(x, A) = d(x, y_n)$. We need to prove that $y_n \in \partial A$ for some n. Note that with $\alpha_n = \|z_n - x\|$, we have

$$\partial K_n = \partial(A \cap B_{\alpha_n}(x)) \subseteq \partial A \cup \partial B_{\alpha_n}(x) = \partial A \cup \{y \in \mathbb{R}^n | \|y - x\| = \alpha_n\}.$$

Since $\|y_n - x\| = d(y_n, x) = d(x, A)$, the sequence $\|y_n - x\|$ is constant, in particular bounded. But α_n tends to infinity, since $\|z_n\|$ tends to infinity. Thus, there is n such that $\|y_n - x\| \neq \alpha_n$, and therefore $y_n \in \partial A$, as desired. $\square$

Theorem C.2.2 (Urysohn's Lemma with Lipschitz constant). *Let K and V be compact, respectively open subsets of $\mathbb{R}^n$, with $K \subseteq V$. Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$, and let d be the induced metric. Assume that V has compact closure. There exists $f \in C_c(\mathbb{R}^n)$ such that $K \prec f \prec V$ and such that f is d-Lipschitz continuous and d-Lipschitz constant*

$$\frac{4 \sup_{x \in \partial V, y \in \partial K} d(x, y)}{(\inf_{x \in \overline{V} \setminus K^\circ} d(x, K) + d(x, V^c))^2}.$$

Proof. Clearly, we can assume $K \neq \emptyset$. We define

$$f(x) = \frac{d(x, V^c)}{d(x, K) + d(x, V^c)},$$

and claim that f satisfies the properties given in the lemma. We first argue that f is well-defined. We need to show that the denominator in the expression for f is finite and nonzero. Since $K \neq \emptyset$, $d(x, K)$ is always finite. Since V has compact closure, $V \neq \mathbb{R}^n$, so $d(x, V^c)$ is also always finite. Thus, the denominator is finite. It is also nonnegative, and if $d(x, K) + d(x, V^c) = 0$, we would have by the closedness of K and V^c that $x \in K \cap V^c$, an impossibility. This shows that f is well-defined.

It remains to prove that $f \in C_c(\mathbb{R}^n)$, that $K \prec f \prec V$ and that f has the desired Lipschitz properties. We will prove these claims in four steps.

Step 1: $f \in C_c(\mathbb{R}^n)$ with $K \prec f \prec V$. It is clear that f is continuous. If $x \in K$, $d(x, K) = 0$ and therefore $f(x) = 1$. If $x \in V^c$, $d(x, V^c) = 0$ and $f(x) = 0$. Thus, $K \prec f \prec V$. Also, since f is zero outside of V, the support of f is included in $\overline{V}$. Since V has compact closure, f has compact support.

Step 2: Lipschitz continuity of f on $\overline{V} \setminus K^\circ$. We begin by showing that f is Lipschitz continuous on the compact set $\overline{V} \setminus K^\circ$. Letting $x, y \in \overline{V} \setminus K^\circ$, we find

$$\begin{aligned} f(x) - f(y) &= \frac{d(x, V^c)}{d(x, K) + d(x, V^c)} - \frac{d(y, V^c)}{d(y, K) + d(y, V^c)} \\ &= \frac{d(x, V^c)(d(y, K) + d(y, V^c)) - d(y, V^c)(d(x, K) + d(x, V^c))}{(d(x, K) + d(x, V^c))(d(y, K) + d(y, V^c))} \\ &= \frac{d(x, V^c)d(y, K) - d(y, V^c)d(x, K)}{(d(x, K) + d(x, V^c))(d(y, K) + d(y, V^c))}. \end{aligned}$$

Since the mapping $x \mapsto d(x,K) + d(x,V^c)$ is positive and continuous in x and y and $\overline{V} \setminus K^\circ$ is compact, we conclude that the infimum of $d(x,K) + d(x,V^c)$ over x in $\overline{V} \setminus K^\circ$ is positive. Let C denote this infimum. Furthermore, note that the mappings $x \mapsto d(x,V^c)$ and $x \mapsto d(x,K)$ is continuous, and therefore their suprema over x in $\overline{V} \setminus K^\circ$ are finite. Let c_1 and c_2 denote these suprema. Note that for any $x, y, a, b \in \mathbb{R}$,

$$\begin{aligned} xa - yb &= x\left(\frac{a+b}{2} + \frac{a-b}{2}\right) - y\left(\frac{a+b}{2} - \frac{a-b}{2}\right) \\ &= (x-y)\frac{a+b}{2} + (a-b)\frac{x+y}{2}, \end{aligned}$$

so by symmetry, $|xa - yb| \le |x-y|\frac{|a+b|}{2} + |a-b|\frac{|x+y|}{2}$. This shows that

$$\begin{aligned} |f(x) - f(y)| &= \frac{|d(x,V^c)d(y,K) - d(y,V^c)d(x,K)|}{(d(x,K) + d(x,V^c))(d(y,K) + d(y,V^c))} \\ &\le \frac{1}{C^2}|d(x,V^c)d(y,K) - d(y,V^c)d(x,K)| \\ &\le \frac{1}{C^2}\left(c_2|d(x,V^c) - d(y,V^c)| + c_1|d(x,K) - d(y,K)|\right) \\ &\le \frac{1}{C^2}\left(c_2 d(x,y) + c_1 d(x,y)\right) \\ &\le \frac{c_1 + c_2}{C^2} d(x,y). \end{aligned}$$

Now, since , We have now shown the Lipschitz property of f for $x, y \in \overline{V}$, obtaining the Lipschitz constant $c = \frac{c_1+c_2}{C^2}$. We will now extend the Lipschitz property to all of $\mathbb{R}^n$.

Step 3: Lipschitz property of f on $\mathbb{R}^n$. We first extend the Lipschitz property to $\mathbb{R}^n \setminus K^\circ$. Let $x, y \in \mathbb{R}^n \setminus K^\circ$ be arbitrary. Since we already have the Lipschitz property when $x, y \in \overline{V} \setminus K^\circ$, we only need to consider the case where one of x and y are in $\overline{V}^c$. If x and y both are in $\overline{V}^c$, they are also both in V^c, so $f(x) - f(y) = 0$ and the Lipschitz property trivially holds here. It remains to consider the case where, say, $x \in \overline{V}^c$ and $y \in \overline{V} \setminus K^\circ$. Now, since $\overline{V}$ is compact, by Lemma C.2.1 there exists $z \in \partial\overline{V}$ such that $d(x, \overline{V}) = d(x,z)$. Now, since V is open, $\partial\overline{V} = \partial V = \partial V^c \subseteq V^c$. Since f is zero on V^c, we conclude $f(z) = 0$. Since also $f(x) = 0$, we obtain

$$|f(x) - f(y)| \le |f(x) - f(z)| + |f(z) - f(y)| = |f(y) - f(z)| \le cd(y,z).$$

Now, z is the element of $\overline{V}$ which minimizes the distance to x. Since $y \in \overline{V}$, it follows that $d(x,z) \le d(x,y)$ and therefore $d(y,z) \le d(y,x) + d(x,z) \le 2d(x,y)$. We conclude

$$|f(x) - f(y)| \le 2cd(y,z).$$

We have now proven that f is Lipschitz on $\mathbb{R}^n \setminus K^\circ$ with Lipschitz constant $2c$.

Finally, we extend the Lipschitz property to all of $\mathbb{R}^n$. Let $x, y \in \mathbb{R}^n$. As before, it will suffice to consider the case where either $x \in K^\circ$ or $y \in K^\circ$. In the case $x, y \in K^\circ$, $f(x) - f(y) = 0$ and the Lipschitz property is trivial.

We therefore consider the case where $x \in K^\circ$ and $y \in \mathbb{R}^n \setminus K^\circ$. By Lemma C.2.1, there is z on the boundary of $\mathbb{R}^n \setminus K^\circ$ such that $d(x, \overline{V} \setminus K^\circ) = d(x, z)$. Now, the boundary of $\mathbb{R}^n \setminus K^\circ$ is the same as ∂K, so $z \in \partial K$. K is closed and therefore $\partial K \subseteq K$, so $f(z) = 1$. Since also $x \in K^\circ \subseteq K$, $f(x) = 1$ and we find

$$|f(x) - f(y)| \leq |f(x) - f(z)| + |f(z) - f(y)| = |f(y) - f(z)|.$$

We know that f is Lipschitz on $\mathbb{R}^n \setminus K^\circ$ and that $y \in \mathbb{R}^n \setminus K^\circ$. Since $z \in \partial(\mathbb{R}^n \setminus K^\circ)$, there is a sequence $z_n \in \mathbb{R}^n \setminus K^\circ$ such that z_n converges to z. This yields

$$|f(y) - f(z)| = \lim_n |f(y) - f(z_n)| \leq \lim_n cd(y, z_n) = cd(y, z),$$

and we thus recover $|f(x) - f(y)| \leq cd(y, z)$. As in the previous step, $d(x, z) \leq d(x, y)$ and so $d(y, z) \leq d(y, x) + d(x, z) \leq 2d(x, y)$, yielding $|f(x) - f(y)| \leq 2cd(y, z)$. This shows the Lipschitz property.

Step 4: Estimating the Lipschitz constant. We now know that f is Lipschitz continuous with Lipschitz constant

$$2c = \frac{2c_1 + 2c_2}{C^2} = \frac{2\sup_{x \in \overline{V} \setminus K^\circ} d(x, K) + 2\sup_{x \in \overline{V} \setminus K^\circ} d(x, V^c)}{(\inf_{x \in \overline{V} \setminus K^\circ} d(x, K) + d(x, V^c))^2}.$$

This expression is somewhat cumbersome. By estimaing it upwards, we obtain a weaker result, but more manageable. We will leave the denominator untouched and only consider the numerator. Our goal is to show

$$\sup_{x \in \overline{V} \setminus K^\circ} d(x, K) \leq \sup_{x \in \partial V, y \in \partial K} d(x, y)$$
$$\sup_{x \in \overline{V} \setminus K^\circ} d(x, V^c) \leq \sup_{x \in \partial V, y \in \partial K} d(x, y).$$

Consider the first equation. Obviously, to prove the inequality, we can assume that $\sup_{x \in \overline{V} \setminus K^\circ} d(x, K)$ is nonzero. First note that since the mapping $x \mapsto d(x, V^c)$ is continuous and $\overline{V} \setminus K^\circ$ is compact, there is $z \in \overline{V} \setminus K^\circ$ such that

$$\sup_{x \in \overline{V} \setminus K^\circ} d(x, V^c) = d(z, V^c).$$

And by Lemma C.2.1, there is $w \in \partial V^c = \partial V$ such that $d(z, V^c) = d(z, w)$. We want to prove $z \in \partial K$. First assume that z is in the interior of $\overline{V} \setminus K^\circ$, we seek a contradiction. Let $\varepsilon > 0$ be such that $B_\varepsilon(z) \subseteq \overline{V} \setminus K^\circ$. There is $0 < \lambda \leq 1$ such that $\lambda w + (1-\lambda)z \in B_\varepsilon(z)$, and we then find

$$d(\lambda w + (1-\lambda)z, w) = \|(1-\lambda)(z-w)\| = (1-\lambda)\|z-w\| < d(z,w).$$

But z was the infimum of $d(y,w)$ over $y \in \overline{V} \setminus K^\circ$. Thus, we have obtained a contradiction and conclude that z is not in the interior of $\overline{V} \setminus K^\circ$. Since $\overline{V} \setminus K^\circ$ is closed, we conclude $z \in \partial(\overline{V} \setminus K^\circ) \subseteq \partial V \cup \partial K$. If $z \in \partial V$, we obtain $d(z, V^c) = 0$, inconsistent with our assumption that $\sup_{x \in \overline{V} \setminus K^\circ} d(x, K)$ is nonzero. Therefore, $z \in \partial K$. We may now conclude

$$\sup_{x \in \overline{V} \setminus K^\circ} d(x, K) = d(z, w) \leq \sup_{x \in \partial V, y \in \partial K} d(x, y).$$

We can use exacly the same arguments to prove the other inequality. We can therefore all in all conclude that f is Lipschitz with Lipschitz constant

$$\frac{4 \sup_{x \in \partial V, y \in \partial K} d(x, y)}{(\inf_{x \in \overline{V} \setminus K^\circ} d(x, K) + d(x, V^c))^2},$$

as desired. $\square$

Comment C.2.3 The Lipschitz constant could possibly be improved by instead considering the mapping

$$f(x) = \min\left\{\frac{d(x, V^c)}{d(K, V^c)}, 1\right\}.$$

$\circ$

Appendix D

List of Symbols

General probability

- $(\Omega, \mathcal{F}, \mathcal{F}_T, P)$ - the background probability space.
- $\mathcal{F}_\infty$ - The σ-algebra induced by $\cup_{t\geq 0}\mathcal{F}_t$.
- W - Generic Brownian motion.
- τ, σ - Generic stopping times.
- $\Sigma_\mathcal{A}$ - The σ-algebra of measurability and adaptedness.
- Σ_π - The progressive σ-algebra.
- M - Generic martingale.
- M_t^* - The maximum of M over $[0, t]$.
- M^* - The maximum of M over $[0, \infty)$.
- $\mathbb{SP}$ - The space of real stochastic processes.
- $\mathcal{M}$ - The space of martingales.
- $\mathcal{M}_0$ - The space of martingales starting at zero.
- $\mathbf{c}\mathcal{M}_0^2$ - The space of continuous square-integrable martingales starting at zero.

- $\mathbf{c}\mathcal{M}_0^{\mathfrak{L}}$- The space of continuous local martingales starting at zero.
- $\xrightarrow{P}$ - Convergence in probability.
- $\xrightarrow{\mathcal{L}^p}$ - Convergence in $\mathcal{L}^p$.
- $\xrightarrow{\text{a.s.}}$ - Almost sure convergence.
- ϕ - The standard normal density.
- ϕ_n - The n-dimensional standard normal density.

Stochastic integration

- $\mathbf{b}\mathcal{E}$ - The space of elementary processes.
- $\mathcal{L}^2(W)$ - The space of progressive processes X with $E\int_0^\infty X_t^2 \, \mathrm{d}t$ finite.
- $\mathfrak{L}^2(W)$ - The space of processes $\mathfrak{L}_0$-locally in $\mathcal{L}^2(W)$.
- $\mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$ - The processes $\sum_{k=1}^n \int_0^t Y_k \, \mathrm{d}W_t^k$, $Y \in \mathfrak{L}^2(W)^n$.
- μ_M- The measure induced by $M \in \mathbf{c}\mathcal{M}_W^{\mathfrak{L}}$.
- $\mathcal{L}^2(M)$ - The $\mathcal{L}^2$ space induced by μ_M.
- $\mathfrak{L}^2(M)$ - The space of processes $\mathfrak{L}_0$-locally in $\mathcal{L}^2(M)$.
- **cFV** - The space of continuous finite variation processes.
- $\mathbf{cFV}_0$ - The space of continuous finite variation processes starting at zero.
- $\mathcal{S}$ - The space of standard processes.
- $L(X)$ - The space of integrands for $X \in \mathcal{S}$.
- $[X]$ - The quadratic variation of a process $X \in \mathcal{S}$.
- $Q^\pi(M)$ - The quadratic variation of M over the partition π.
- $\mathcal{E}(X)$ - The Doléans-Dade exponential of $X \in \mathcal{S}$.
- X, Y - Generic elements of $\mathcal{S}$.
- H, K - Generic elements of $\mathbf{b}\mathcal{E}$.

The Malliavin calculus

- D - The Malliavin derivative.
- $\mathcal{S}$ - The space of smooth variables, $C_p^\infty(\mathbb{R}^n)$ transformations of W coordinates.
- $\mathcal{L}^p(\Pi)$ - The space $\mathcal{L}^p([0,T]\times\Omega, \mathcal{B}[0,T]\otimes\mathcal{F}_T, \lambda\otimes P)$.
- $\mathbb{D}_{1,p}$ - The domain of the extended Malliavin derivative.
- F - Generic element of $\mathbb{D}_{1,p}$.
- X - Generic element of $\mathcal{L}^2(\Pi)$.
- θ - The Brownian stochastic integral operator on $[0,T]$.
- (e_n) - Generic countable orthonormal basis of $\mathcal{L}^2[0,T]$.
- $\langle\cdot,\cdot\rangle_{[0,T]}$ - The inner product on $\mathcal{L}^2[0,T]$.
- $\langle\cdot,\cdot\rangle_\Pi$ - The inner product on $\mathcal{L}^2(\Pi)$.
- $\mathcal{H}_n$ - The subspace of $\mathcal{L}^2(\mathcal{F}_T)$ based on the n'th Hermite polynomial.
- $\mathcal{H}_n'$,$\mathcal{H}_n''$ - Dense subspaces of $\mathcal{H}_n$.
- $\mathbb{H}_n$ - Orthonormal basis of $\mathcal{H}_n$.
- $\mathcal{H}_n(\Pi)$ - The subspace of $\mathcal{L}^2(\Pi)$ based on $\mathcal{H}_n$.
- $\mathcal{H}_n'(\Pi)$, $\mathcal{H}_n''(\Pi)$ - Dense subspaces of $\mathcal{H}_n(\Pi)$.
- $\mathbb{H}_n(\Pi)$ - Orthonormal basis for $\mathcal{H}_n(\Pi)$.
- P_n - Orthogonal projection onto $\mathcal{H}_n$.
- P_n^Π - Orthogonal projection onto $\mathcal{H}_n(\Pi)$.
- Φ_a - Generic element of $\mathbb{H}_n$.
- $\mathfrak{P}_n$ - Polynomials of degree n in any number of variables.
- $\mathcal{P}_n$ - Subspace of $\mathcal{L}^2(\mathcal{F}_T)$ based on $\mathfrak{P}_n$.
- $\mathcal{P}_n'$ - Dense subspace of $\mathcal{P}_n$.
- $\mathbb{P}_n$ - Total subset of $\mathcal{P}_n$.

- δ - The Skorohod integral.

Mathematical finance

- Q - A generic T-EMM.
- S - Generic asset price process.
- r - Generic short rate process.
- B - Generic risk-free asset price process.
- S' - Generic normalized asset price process.
- B' - Generic normalized risk-free asset price process.
- $\mathcal{M}$ - Generic financial market model.
- (h^0, h^S) - Generic portfolio strategy.
- V^h - Value process of the portfolio strategy h.
- μ - Generic drift vector process.
- σ - Generic volatility matrix process.
- ρ - Generic correlation.

Measure theory

- $\mathcal{B}$ - The Borel σ-algebra on $\mathbb{R}$.
- $\mathcal{B}_k$ - The Borel σ-algebra on $\mathbb{R}^k$.
- $\mathcal{B}(M)$ - The Borel σ-algebra on a pseudoemetric space (M, d).
- λ - The Lebesgue measure.
- $[f]$ - The equivalence class corresponding to f.
- $C([0,\infty), \mathbb{R}^n)$ - The space of continuous mappings from $[0,\infty)$ to $\mathbb{R}^n$.
- X_t° - The t'th projection on $C([0,\infty), \mathbb{R}^n)$.

- $\mathcal{C}([0,\infty),\mathbb{R}^n)$ - The σ-algebra on $C([0,\infty),\mathbb{R}^n)$ induced by the projections.

Analysis

- $C^1(\mathbb{R}^n)$ - The space of continuous differentiably mappings from $\mathbb{R}^n$ to $\mathbb{R}$.
- $C^\infty(\mathbb{R}^n)$ - The infinitely differentiable elements of $C^1(\mathbb{R}^n)$.
- $C_p^\infty(\mathbb{R}^n)$ - The elements of $C^\infty(\mathbb{R}^n)$ with polynomial growth for the mapping and its partial derivatives.
- $C_c^\infty(\mathbb{R}^n)$ - The elements of $C^\infty(\mathbb{R}^n)$ with compact support.
- $C^2(U)$ - The space of twice continuously differentiable mappings from U to $\mathbb{R}^n$.
- $\|\cdot\|_\infty$ - The uniform norm.
- H_n - The n'th Hermite polynomial.
- $M \oplus N$ - orthonormal sum of two closed subspaces.
- V_F - Variation of F.
- μ_F - Signed measure induced by a finite variation mapping F.
- $f * g$ - Convolution of f and g.
- (ψ_ε) - Dirac family.
- K - Generic compact set.
- U, V - Generic open sets.

Bibliography

L. Andersen: Simple and efficient simulation of the Heston stochastic volatility model, Journal of Computational Finance, Vol. 11, No. 3, Spring 2008.

E. Alòs & C. Ewald: Malliavin Differentiability of the Heston Volatility and Applications to Option Pricing, Advances in Applied Probability 40, p. 144-162.

R. B. Ash: Measure, Integration and Functional Analysis, Academic Press, 1972.

T. M. Apostol: Calculus, Volume 1, Blaisdell Publishing Company, 1964.

V. Bally: An elementary introduction to Malliavin calculus, INRIA, 2003, http://perso-math.univ-mlv.fr/users/bally.vlad/RR-4718.pdf.

R. F. Bass: Diffusions and Elliptic Operators, Springer-Verlag, 1998.

D. R. Bell: The Malliavin Calculus, Longman Scientific & Technical, 1987.

C. Berg & T. G. Madsen: Mål- og integralteori, Matematisk afdeling, KU, 2001.

P. Billingsley: Convergence of Probability Measures, 2nd edition, 1999.

E. Benhamou: Smart Monte Carlo: Various tricks using Malliavin calculus, 2001, http://www.iop.org/EJ/article/1469-7688/2/5/301/qf2501.pdf.

E. Benhamou: Optimal Malliavin Weighting function for the computation of the greeks, Mathematical Finance, Vol. 13, No. 1, 37-53.

T. Björk: Arbitrage Theory in Continuous Time, Oxford University Press, 2nd edition, 2004

N. L. Carothers: Real Analysis, Cambridge University Press, 2000.

N. Chen & P. Glasserman: Malliavin Greeks without Malliavin Calculus, Columbia University, 2006, http://www.cfe.columbia.edu/pdf-files/Glasserman_03.pdf.

P. Cheridito, D. Filipovic & R. L. Kimmel: Market Price of Risk Specifications for Affine Models: Theory and Evidence, 2005, http://ssrn.com/abstract=524724.

K. L. Chung & J. L. Doob: Fields, Optionality and measurability, American Journal of Mathematics, 87, 397-424, 1965.

D. L. Cohn: Measure Theory, Birkhäuser, 1980.

R. Cont, P. Tankov: Financial Modeling with Jump Processes, Chapman & Hall, 2004.

J. C. Cox, J. E. Ingersoll & S. A. Ross: A Theory of the Term Structure of Interest Rates, Econometrica, Vol. 53, No. 2, p. 385-407.

C. Dellacherie & P. A. Meyer: Probabilities and Potential A, North-Holland, 1975.

R. M. Dudley: Real Analysis and Probability, Cambridge University Press, 2002.

F. Delbaen & W. Schachermayer: Non-Arbitrage and the Fundamental Theorem of Asset Pricing: Summary of Main Results, Proceedings of Symposia in Applied Mathematics, 1997, http://www.fam.tuwien.ac.at/ wschach/pubs/preprnts/prpr0094.ps.

F. Delbaen & W. Schachermayer: The Mathematics of Arbitrage, Springer, 2008.

J. L. Doob: Stochastic Processes, Wiley, 1953.

A. A. Dragulescu & V. M. Yakovenko: Probability Distribution of returns in the Heston model with stochastic volatility, Quantitative Finance 2002, Vol. 2, pp. 443-453.

R. J. Elliott: Stochastic Calculus and Applications, Springer-Verlag, 1982.

S. N. Ethier & T. G. Kurtz: Markov Processes: Characterization and Convergence, Wiley, 1986.

W. Feller: Two Singular Diffusion Problems, The Annals of Mathematics, Second Series, Vol. 54, No. 1, pp. 173-182.

E. Fournié et al.: Applications of Malliavin Calculus to Monte Carlo methods in finance, Finance and Stochastics, 1999, no.3. p. 391-412.

E. Fournié et al.: Applications of Malliavin Calculus to Monte Carlo methods in finance II, Finance and Stochastics, 2002, no.5. p. 201-236.

C. P. Fries & M. S. Yoshi: Partial Proxy simulation schemes for generic and robust Monte Carlo Greeks, Journal of Computational Finance, Vol. 11, No. 3, 2008.

P. K. Friz: An Introduction to Malliavin Calculus, Courant Institute of Mathematical Sciences, New York University, 2002, http://www.statslab.cam.ac.uk/ peter/malliavin/Malliavin2005/mall.pdf.

P. Glasserman: Monte Carlo Methods in Financial Engineering, Springer-Verlag, 2003.

E. Hansen: Sandsynlighedsregning på Målteoretisk Grundlag, Afdeling for anvendt matematik og statistik, 2004.

E. Hansen: Measure Theory, Afdeling for anvendt matematik og statistik, 2004.

V. L. Hansen: Functional Analysis: Entering Hilbert Space, World Scientific, 2006.

S. L. Heston: A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, The Review of Financial Studies, Vol. 6, No. 2, pp. 327-343.

N. Ikeda & S. Watanabe: Stochastic Differential Equations and Diffusion Processes, 2nd edition, North-Holland, 1989.

M. Jacobsen: Forelæsningsnoter om stokastiske integraler, 1989.

M. Jacobsen: Markov Processer, 1972.

M. Jacobsen: Videregående Sandsynlighedsregning, Universitetsbogladen København, 3. udgave, 2003.

J. Jacod, A. N. Shiryaev: Limit Theorems for Stochastic Processes, Springer-Verlag, 1987.

S. T. Jensen: Eksponentielle Familier, 2. udgave, Institut for Matematisk Statistik, 1992.

I. Karatzas, S. E. Shreve: Brownian Motion and Stochastic Calculus, Springer-Verlag, 1988.

I. Karatzas, S. E. Shreve: Methods of Mathematical Finance, Springer-Verlag, 1988.

S. Kaden, J. Potthoff: Progressive Stochastic Processes and an Application to the Itô integral, Stoch. Anal. and Appl, 2004, Vol 22, number 4, pages 843-866.

H. Kunita: Stochastic flows and stochastic differential equations, Cambridge University Press, 1990.

S. Levental & A. V. Skorohod: A Necessary and Sufficient Condition for Absence of Arbitrage with Tame Portfolios, The Annals of Applied Probability, Vol. 5, No. 4, pp. 906-925.

R. Lord, R. Koekkoek & D. van Dijk: A Comparison of Biased Simulation Schemes for Stochastic Volatility Models, Tinbergen Institute Discussion Paper, 2006.

P. Malliavin: Stochastic Analysis, Springer-Verlag, 1997.

P. Malliavin & A. Thalmaier: Stochastic Calculus of Variations in Mathematical Finance, Springer, 2005.

R. Meise & D. Vogt: Introduction to Functional Analysis, Clarendon Press, Oxford, 1997.

J. R. Munkres: Topology, 2nd edition, Prentice Hall, 2000.

M. Montero & A. Kohatsu-Higa: Malliavin Calculus applied to Finance, 2002, http://www.recercat.net/bitstream/2072/817/1/672.pdf.

G. Di Nunno, B. Øksendal & F. Proske: Malliavin Calculus for Lévy Processes with Applications to Finance, Universitext, Springer 2008 (to appear).

D. Nualart: The Malliavin Calculus and Related Topics, 2nd edition, Springer-Verlag, 2006.

P. Protter: Stochastic integration and differential equations, 2nd edition, Springer-Verlag, 2005.

P. Protter & K. Shimbo: No Arbitrage and General Semimartingales, 2006, http://people.orie.cornell.edu/ protter/WebPapers/na-girsanov8.pdf.

D. Revuz & M. Yor: Continuous Martingales and Brownian Motion, 3rd edition, Springer-Verlag, 1998.

L. C. G. Rogers & D. Williams: Diffusions, Markov Processes and Martingales, Volume 1: Foundations, 2nd edition, Cambridge University Press, 2000.

L. C. G. Rogers & D. Williams: Diffusions, Markov Processes and Martingales, Volume 2: Itô Calculus, 2nd edition, Cambridge University Press, 2000.

W. Rudin: Real and Complex Analysis, McGraw-Hill, 3rd edition, 1987.

M. Sanz-Solé: Malliavin Calculus with applications to stochastic partial differential equations, EPFL Press, 2005.

M. Schechter: Principles of Functional Analysis, 2nd edition, American Mathematical Society, 2002.

T. Schröter: Malliavin Calculus in Finance, Trinity College Special Topic Essay 2007, http://people.maths.ox.ac.uk/schroter/mall.pdf.

M. Sharpe: General Theory of Markov Processes, Academic Press, 1988.

S. E. Shreve: Stochastic Calculus for Finance II: Continuous-Time Models, Springer, 2004.

A. Sokol: Stokastiske integraler og Differentialligninger, 2005.

A. Sokol: Markov Processer, 2007.

J. M. Steele: Stochastic Calculus and Financial Applications, Springer-Verlag, 2000.

C. M. Stein: Estimation of the Mean of a Multivariate Normal Distribution, The Annals of Statistics, 1981, Vol. 9, No. 6, pp. 1135-1151.

G. Szegö: Orthogonal Polynomials, American Mathematical Society, 1939.

H. Zhang: The Malliavin Calculus, 2004, http://janroman.dhis.org/finance/Malliavin Calculus/Malliavin Calculus.pdf.

H. Zou, T. Hastie & R. Tibshirani: On the "Degrees of Freedom" of the Lasso, The Annals of Statistics, 2007, Vol. 35, No. 5, pp. 2173-2192.

B. Øksendal: Stochastic Differential Equations, 6th edition, Springer-Verlag, 2005.

B. Øksendal: An introduction to Malliavin calculus with applications to economics, 1997.