Stephan Sturm

Calculation of the Greeks by Malliavin Calculus

Diplomarbeit zur Erlangung des Akademischen Grades
"Magister der Naturwissenschaften"

Betreuer: Walter Schachermayer
Institut für Matematik
Universität Wien

Februar 2004

Müdigkeit spürte er keine,
nur war es ihm manchmal unangenehm,
dass er nicht auf dem Kopf gehn konnte.

Georg Büchner, *Lenz*

Preface

It would not be correct to describe the work on this thesis, now having it finished successfully, as a short and easy task. On the contrary, it took me a long time and in the couse of the work I had a lot of doubts about my own mathematical abilities. It was my old friend Josef Teichmann who proposed me to write a diploma thesis on Malliavin Calculus under his guidance. It was for me quite a hard, but also very joyful path starting out as a student whith hardly any knowledge in probability theory to the heights of stochastic analysis.

The goal of my diploma thesis was a profound understanding of the calculation of the Greeks by Malliavin calculus in the n-dimensional, elliptic case as first presented by Fournié e.a. (1998) and a generalization to hypoellipticity; this generalization is the focus of current research, see e.g. the works of Malliavin and Thalmaier (2003, preprint), Gobet and Munos (2002, preprint) and Teichmann and Touzi (working paper).

Malliavin calculus, i.e. the stochastic calculus of variations which is build up on the notion of a weak derivative on the Wiener space, the Malliavin derivative, lies in the core of the intersection of stochastic analysis, functional analysis and differential geometry. It is the perfect tool for a calculation of the sensitivity of the price of an option with respect to small changes in the parameters, i.e. the Greeks. The abstract notions of functional analysis allow us to write the Greeks as the expectation of the product of the original payoff function with a specific factor, the Malliavin weight which is in fact a Skorohod integral, the adjoint operator of the Malliavin derivative and a generalization of the notion of the Itô integral.

After a short introduction to martingale theory I will give the foundations of stochastic analysis, introduce the Itô and the Stratonovich notion of the stochastic integral (with respect to a Brownian motion, but also in the more general continuous semimartingale case) and present the classical Girsanov theory of transformations of the probability measure. The proof of the unique existence of the solution of a stochastic differential equation follows an introduction to its first derivative with respect to the initial value, the first variation process.

The introduction of the Wiener chaos decomposition allows me to understand multiple Wiener-Itô integrals as iterated (classical) Itô integrals and hence to look at stochastic integration as a process of climbing up the Wiener chaos. The Malliavin derivative is introduced as the inverse climbing down and I will prove its (functional) analytic properties up to the Clark-Haussmann-Ocone formula, in the core the chain rule. The divergence operator (or Skorohod integral) is introduced as its adjoint operator and it is shown that it coincides for progressively measurable processes with the Itô integral. As last theoretical point I will show the connection between the first variation process and the Malliavin derivative which leads us to some closing remarks on the existence and smoothness of densities of random variables.

The hitherto developed mathematical theory is used to answer a specific question of mathematical finance: What is the behavior of an option if we vary the parameters a little bit? In the jargon of finance this means to calculate the Greeks. The idea behind the notion of "calculating" is here to develop a formula which is better fitted for a numerical evaluation than the simple difference quotient. The method is to express the derivative of the expectation as expectation of a product of the original payoff function and some weight.

For the n-dimensional elliptic case we follow the already classical paper by Fournié e.a. to show that we can calculate the derivatives with respect to the interest rate by classical Girsanov theory while for those with respect to the initial value and the volatility the Malliavin calculus is of great use. So we can write the weights as Skorohod integral whose integrands depend only on the underlying processes.

Dropping the ellipticity condition we will then show in an hypoelliptic setting with d Brownian motions for an n-dimensional process that we can calculate the Greeks also here.

As one concrete application of this method we calculate the Greeks in the Black-Scholes model and show the strength of the hypoelliptic formula by using it for an approximation of the Hobson-Rogers delta. In particular the obtained formulas allow simple numerical algorithms to approximate the solutions of hypoelliptic partial differential equations. This feature is applied in Hubalek, Teichmann, Tompkins (2004) to fit parameters of a model to real market data without using sophisticated PDE techniques.

The main sources for me were the book of Revuz and Yor [RY 91] for martingale theory and stochastic integration with respect to continuous semimartingales, the manuscripts of Teichmann [Tei 02] and [Tei 03] for stochastic integration, the theory of SDEs, Wiener Chaos and Malliavin Calculus, Bass [Ba 98] for the first variation process and Nualart's book on Malliavin Calculus [Nua 95]. General reference was Kallenberg [Kal 02], the calculation of the Greeks in the elliptic setting is due to Fournié e.a. [FLLLT 99], the hypoelliptic treatment was inspired by Teichmann and Touzi [TT].

First of all I have to thank Josef Teichmann who initiated me to this subject and was always there if I had to discuss some problems of my work. Sebastian Markt made some linguistical suggestion and thus helped me to master my problems with the English language. All possible faults obviously remain in my responsibility.

The Department of Financial and Actuarial Mathematics at the Technical

University of Vienna under the direction of Walter Schachermayer gave me an ideal environment for my work in a very pleasant atmosphere. I am also very grateful that the department gave my the possibility to participate in the "Berlin Workshop on Mathematical Finance for Young Researchers". Thanks also to CIMPA and Prof. S.G. Dani who gave me the possibility to spend two intensive weeks at the Summer School "Probability Measures on Groups" at the Tata Institute for Fundamental Research (TIFR), Bombay.

Last but not least I have to thank all those people who spent their days and nights with me, laughing and discussing, cooking, eating (and sometimes too much) drinking, on the mountains, in cafés or in our flats, shortly: my friends. In particular I want to mention Herwig Czech, Ulrike Girardi, Florian Huber, Sebastian Markt, Martina Punz, Christian Selinger, Martin, Susanne and Wilhelm Sturm, Josef Teichmann and Florian Wenninger.

I want to dedicate this work to three people who always looked on my formation and on my studies, but sadly could not see the fruits of all their care, love and help: My aunt Anna Kopečny (1914-2001) and my grandparents Maria Sturm (1920-2003) and Eduard Sturm (1920-2000).

Chapter 1

Preliminaries

Modern probability theory was founded by Andrei Nikolaevich Kolmogorov who, claiming that *"the theory of probability as mathematical discipline can and should be developed from axioms in exactly the same way as Geometry and Algebra"*, was the first to treat this subject from an axiomatic, measure theory based point of view. It is here not the place to go into the measure theoretic details of the foundations of modern probability theory; we will refer to the literature where needed. We give here only an introduction to stochastic process, in particular martingales, and a very short recapitulation of the basics of the theory of tensors on Hilbert spaces.

1.1 Stochastic Processes and Martingales

In this section we will give an introduction to stochastic processes, i.e. families of random variables, and martingales, stochastic processes which can be thought as "fair games".

1.1.1 Stochastic Processes

Definition 1.1.1 (Stochastic Process)
Given an index set T, a stochastic process is a family of measurable mappings X_t, $t \in T \subset \mathbb{R}_{\geq 0} \cup \{\infty\}$, from a probability space $(\Omega, \mathcal{F}, P)$ to a measurable space $(\mathcal{E}, \mathcal{G})$, the state space.

Under the path (or trajectory) of a stochastic process we understand the mapping $t \to X_t(\omega)$. A process is called continuous iff for almost all ω's the paths are continuous.

We introduce the following notions of "proximity" of stochastic processes:

Definition 1.1.2 (Modifications and Indistinguishability)
Two stochastic processes X_t, Y_t, $t \in T$ defined on the same probability space $(\Omega, \mathcal{F}, P)$ are called

(i) modifications of each other iff for each $t \in T$ we have $X_t = Y_t$ a.s.
(ii) indistinguishable iff for almost all $\omega \in \Omega$ we have $X_t = Y_t$ for all $t \in T$.

So the difference is that for a modification the null set where both processes differ is dependent on t, while for indistinguishability it is required to be independent of the concrete choice of t.

Theorem 1.1.3 (Conditions for Indistinguishability)
Given two stochastic processes X_t, Y_t, modifications of each other. If both processes have right continuous paths a.s., then they are indistinguishable.

Proof Let A, B the null sets where X_t resp. Y_t are not right continuous. Further be $N_t := \{\omega : X_t(\omega) \neq Y_t(\omega)\}$ the null set where X_t and Y_t differ and $N := \bigcup_{t \in \mathbb{Q}} N_t$. N has as rational union of null sets measure zero, so has $M := A \cup B \cup N$. For $\omega \notin M$ we can take for every t a rational sequence $t_n \in \mathbb{Q}$, $t_n \downarrow t$, so right continuity implies that $X_{t_n} = Y_{t_n}$ entails $X_t = Y_t$ for all $\omega \notin M$ independent of the choice of t. ■

The following definition will be of great use for integration theory:

Definition 1.1.4 (Total and Quadratic Variation)
Given a stochastic process X on $\mathbb{R}^d$ and a partition Δ of the interval $[0, t]$ with $0 = t_0 < t_1 < ... < t_n = t$, we consider the sums

$$\begin{aligned} S_t^\Delta(X) &:= \sum_{i=0}^{n-1} |X_{t_{i+1}} - X_{t_i}|, \\ T_t^\Delta(X) &:= \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2. \end{aligned}$$

The process X is said to be of finite total variation on $[0, t]$, iff the total variation process $S_t(X) := \sup_\Delta S_t^\Delta(X) < \infty$ and it is said to be of finite quadratic variation, iff $\langle X, X \rangle_t := \lim_{|\Delta_n| \to 0} T_t^{\Delta_n}(X) < \infty$ for a refining sequence of partitions with mesh $|\Delta_n|$ tending to zero.

The next notion we introduce is a very fundamental one, the filtration:

Definition 1.1.5 (Filtration)
A filtration of a probability space $(\Omega, \mathcal{F}, P)$ is an increasing family $(\mathcal{F}_t)$, $t \in T \subset \mathbb{R}_{\geq 0} \cup \{\infty\}$ of sub- σ-algebras of $\mathcal{F}$, i.e. $\mathcal{F}_k \subset \mathcal{F}_j$ for $k \leq j$.

For the *filtered space*, i.e. the probability space endowed with a filtration we will write $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. By a monotone class argument there exists a $\mathcal{F}_\infty := \bigcup_t \mathcal{F}_t \subset \mathcal{F}$. Iff $\mathcal{F}_\infty = \mathcal{F}$ we say that the filtration converges.

Definition 1.1.6 (Usual Conditions)
We say that a filtration $\mathcal{F}_t$, $t \in T$ satisfies the usual conditions of the Strasbourg School (the "conditions habituels de l'Ecole Strasbourgeoise") or shortly the usual conditions iff

(i) $\mathcal{F}_t$ is complete, i.e. it contains all P-null sets (meaning that for all t, iff $G \subset F \in \mathcal{F}_t$ and $P(F) = 0$, then $G \subset \mathcal{F}_t$).

(ii) $\mathcal{F}_t$ is right-continuous, i.e. $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$.

In the following all our filtrations - except when otherwise stated - will satisfy the usual conditions and converge. The connection of stochastic processes and filtrations seems obvious:

Definition 1.1.7 (Adaptedness)
A process $(X_t)_{t\in T}$ on $(\Omega, \mathcal{F}, P)$ is called adapted to the filtration $\{\mathcal{F}_t\}_{t\in T}$, iff X_t is $\mathcal{F}_t$-measurable for every t.

Every process X_t is adapted to its *natural filtration* $\mathcal{F}_t^0 := \sigma(X_s : s \leq t)$, the coarsest filtration wherefore it is adapted. The usual measurability notion of processes on filtered probability spaces is the following: we say that a process X is measurable, iff the map $t \mapsto X_t(\omega)$ is $\mathcal{B}(\mathbb{R}_{\geq 0}) \times \mathcal{F}_\infty$-measurable. But for integration theory we will need yet another notion of measurability:

Definition 1.1.8 (Progressively Measurable Processes)
A processes X_t, $t \in \mathbb{R}_{\geq 0}$ is called progressively measurable with respect to the filtration $\mathcal{F}_t$, iff its restriction to $[0,T]\times\Omega$ is $\mathcal{B}([0,T])\otimes\mathcal{F}_t$-measurable for every $t \geq 0$.

A progressively measurable process is obviously measurable and adapted. We can extend this notion to subsets of $\mathbb{R}_{\geq 0} \times \Omega$: A set $A \subset \mathbb{R}_{\geq 0} \times \Omega$ is called progressively measurable, iff its indicator function 1_A is progressively measurable. Note that progressively measurable sets form a σ-algebra.

The next objects we introduce are the so-called stopping (or optional) times:

Definition 1.1.9 (Stopping Time)
A stopping time τ relative to the filtration $\mathcal{F}_t$ is a r.v. $\tau : (\Omega, \mathcal{F}, \mathcal{F}_t, P) \to [0, \infty]$ such that $\{\tau \leq t\} := \{\omega : \tau \leq t\} \in \mathcal{F}_t$ for every $t \in T$.

For a right-continuous filtration we can formulate this differently:

Proposition 1.1.10
For a right-continuous filtration $\mathcal{F}_t$, τ is a $\mathcal{F}_t$-stopping time, iff $\{\tau < t\} \in \mathcal{F}_t$ for every $t \in T$.

Proof Given such a τ, then by right continuity

$$\{\tau \leq t\} = \bigcap_{s>t} \{\tau < s\} \in \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t,$$

and conversely for a stopping time τ

$$\{\tau < t\} = \bigcup_{0<s<t} \{\tau \leq s\} = \left(\bigcap_{s>t} \{\tau \leq s\}^c\right)^c \in \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t.$$

■

We note that for $\mathcal{F}_t$-stopping times σ, τ and a real number $\alpha > 1$ the expressions $\sigma \wedge \tau (:= inf(\sigma, \tau))$, $\sigma \vee \tau (:= sup(\sigma, \tau))$, $\sigma + \tau$ and $\alpha\tau$ are $\mathcal{F}_t$-stopping times too. To a given stopping time τ, the expression $t \wedge \tau$ is also a stopping time and we will define the *stopped process* by $X^\tau := X_{t\wedge\tau}$.

1.1.2 Martingales

One of the most central notions of modern probability theory is the martingale:

Definition 1.1.11 (Martingale)
A real-valued process X_t defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is called a $(\mathcal{F}_t)$-martingale (resp. sub- or supermartingale), iff
(i) X_t is adapted to $\mathcal{F}_t$ for all $t \in T$,
(ii) $E(|X_t|) < \infty$ for every $t \in T$,
(iii) $E(X_t|\mathcal{F}_s) = X_s$ (resp. $\geq$, $\leq$) for $s < t$.

In other words, a martingale is an adapted family of r.v.s such that for any set $A \in \mathcal{F}_s$, $s < t$ it holds that

$$\int_A X_s dP = \int_A X_t dP.$$

Given two filtrations $\mathcal{F}_t$, $\mathcal{G}_t$, $\mathcal{F}_t \subset \mathcal{G}_t$, then every $\mathcal{G}_t$-martingale is also a $\mathcal{F}_t$-martingale. In particular every $\mathcal{F}_t$-martingale X_t is a martingale with respect to its natural filtration $\sigma(X_s : s \leq t)$ too.

A martingale is called closed (on the right), iff there exists a $X_\infty \in L^1$ such that $E(X_\infty|\mathcal{F}_s) = X_s$.

Proposition 1.1.12 (Stopped Discrete Martingales)
Given a discrete $\mathcal{F}_n$-(sub-)martingale X_n, $n \geq 0$ and H_n a positive bounded stochastic process with $H_n \in \mathcal{F}_{n-1}$ for $n \geq 1$. Then the process Y given by

$$\begin{cases} Y_0 & := & X_0 \\ Y_n & := & Y_{n-1} + H_n(X_n - X_{n-1}) \end{cases}$$

is a (sub)martingale, and for a stopping time τ, the stopped process $X^\tau = X_{\tau\wedge n}$ is a (sub-)martingale too.

Proof It is enough to show that

$$\begin{aligned} E(Y_{n+1}|\mathcal{F}_n) &= E(Y_n + H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n) \\ &= Y_n + H_{n+1}E(X_{n+1} - X_n|\mathcal{F}_n) \geq Y_n \end{aligned}$$

with equality for X_n a martingale.
For the stopping we define $H_n := 1_{\{n \leq \tau\}} = 1 - 1_{\{\tau \leq n-1\}} \in \mathcal{F}_{n-1}$ which implies that $Y_n = X^\tau$ is a submartingale according to the first part. ■

Lemma 1.1.13 (Discrete Optional Sampling)
Given two bounded stopping times $\sigma \leq \tau$ (i.e. $\sigma(\omega) \leq \tau(\omega) \leq M < \infty$ for a constant M independent of ω) and a discrete (sub-)martingale X_n, then it holds that

$$X_\sigma \leq E(X_\tau|\mathcal{F}_\sigma) \quad a.s.$$

with equality for X a martingale.

Proof We define with respect to the previous proposition $H_n := 1_{\{n \leq \tau\}} - 1_{\{n \leq \sigma\}} \in \mathcal{F}_{n-1}$, then we have on the one hand side

$$Y_n - X_0 = X_\tau - X_\sigma$$

and on the other hand side

$$\begin{aligned} E(Y_n) &= E(Y_{n-1} + H_n(X_n - X_{n-1})) \\ &= E(E(Y_{n-1} + H_n(X_n - X_{n-1})|\mathcal{F}_{n-1})) \geq E(Y_{n-1}) \geq E(Y_0) = E(X_0) \end{aligned}$$

by iteration. Together this gives

$$E(X_\tau) \geq E(X_\sigma).$$

For any $B \in \mathcal{F}_\sigma$ we define the following stopping times $\sigma_B := \sigma 1_B + M 1_{B^c}$ and $\tau_B := \tau 1_B + M 1_{B^c}$ for which the result above reads

$$E(X_\sigma 1_B + X_M 1_{B^c}) \leq E(X_\tau 1_B + X_M 1_{B^c}).$$

Conditioning by $E(\cdot|\mathcal{F}_s)$ gives

$$E(X_\sigma 1_B) \leq E(E(X_\tau 1_B|\mathcal{F}_s)),$$

whence

$$X_\sigma \leq E(X_\tau|\mathcal{F}_s) \quad a.s.$$

with equality for the martingale case. ■

As a corollary we get optional stopping: $X^\sigma \leq E(X^\tau|\mathcal{F}_\sigma)$. To generalize this lemma to arbitrary stopping times and closed martingales we have yet to prove the following statement:

Lemma 1.1.14
Given a closed martingale X, then the family $\{X_\sigma\}$ is u.i. for an arbitrary stopping time σ (bounded or not).

Proof First we prove the lemma for stopping times bounded by a constant M. Then by Lemma 1.1.12

$$X_\sigma = E(X_M|\mathcal{F}_\sigma) = E(E(X_\infty|\mathcal{F}_M)|\mathcal{F}_\sigma) = E(X_\infty|\mathcal{F}_\sigma)$$

what implies for $c > 0$ that

$$\int_{\{|X_\sigma|>c\}} |X_\sigma| dP = \int_{\{|X_\sigma|>c\}} |X_\infty| dP.$$

But by Chebyshev's inequality we have

$$P(\{|X_\sigma| > c\}) \leq \frac{1}{c} E(|X_\sigma|) \leq \frac{1}{c} E(|X_\infty|)$$

implying that (for $c \to \infty$) $P(\{|X_\sigma| > c\} \to 0$ and $\int_{\{|X_\sigma|>c\}} |X_\sigma| dP \to 0$ proving that the family $\{X_\sigma\}$ is u.i.

For the generalization to unbounded stopping times we define the family $U := \{X_{\sigma\wedge M} 1_{\sigma\leq M} + X_\infty 1_{\sigma>M}\}$ for arbitrary σ, M.This family is by the first part of this proof uniformly integrable. Further we can write for an arbitrary stopping

$$X_\sigma = \lim_{M\to\infty} \left(X_{\sigma\wedge M} 1_{\sigma\leq M} + X_\infty 1_{\sigma>M} \right).$$

So X_σ is an a.s. limit of elements of U, hence it is (by Lemma 1.1.13) in $\bar{U}$, the closure of U in L^1, which is also u.i. ■

Theorem 1.1.15 (Doob's Optional Sampling)
Given a closed martingale X, then it holds for two arbitrary stopping times $\sigma \leq \tau$ that

$$X_\sigma = E(X_\tau|\mathcal{F}_\sigma) = E(X_\infty|\mathcal{F}_\sigma) \quad a.s.$$

with equality for X a martingale.

Proof For any set $B \in \mathcal{F}_\sigma$ we have by Lemma 1.1.13

$$\int_{B\cap\{\sigma\leq M\}} X_\sigma dP = \int_{B\cap\{\sigma\leq M\}} X_\infty dP$$

since $B \cap \{\sigma \leq M\} = B \cap \{\sigma \leq \sigma \wedge M\} \in \mathcal{F}_{\sigma\wedge M}$ and on the other hand side it obviously holds that

$$\int_{B\cap\{\sigma=\infty\}} X_\sigma dP = \int_{B\cap\{\sigma=\infty\}} X_\infty dP.$$

This implies for $M \to \infty$ that

$$E(X_\sigma|\mathcal{F}_\sigma) = E(X_\infty|\mathcal{F}_\sigma)$$

and hence

$$X_\sigma = E(X_\infty|\mathcal{F}_\sigma) \quad a.s.$$

since the family $\{X_\sigma\}$ is u.i. by the previous lemma. To establish the second part of the result, it is enough to observe that

$$X_\sigma = E(E(X_\infty|\mathcal{F}_\tau)|\mathcal{F}_\sigma) = E(X_\tau|\mathcal{F}_\sigma) \quad a.s.$$

■

Proposition 1.1.16
For a càdlàg (i.e. right continuous) adapted process X the following conditions are a.s. equivalent:
(i) X is a martingale.
(ii) For any bounded stopping time τ, $X_\tau \in L^1$ and $E(X_\tau) = E(X_0)$.

Proof
(i) $\Rightarrow$ (ii) This is a clear consequence of Doob's Optional Sampling Theorem.
(ii) $\Rightarrow$ (i) For $s < t$ and an arbitrary set $B \in \mathcal{F}_s$, the r.v. $\tau := t1_{B^c} + s1_B$ is a stopping time and hence

$$E(X_0) = E(X_\tau) = E(X_t 1_{B^c}) + E(X_s 1_B).$$

On the other hand side t itself is a stopping time too, so

$$E(X_0) = E(X_t) = E(X_t 1_{B^c}) + E(X_t 1_B).$$

Subtracting these two equations and conditioning to $\mathcal{F}_s$ yields $X_s = E(X_t|\mathcal{F}_s)$ a.s. ■

Corollary 1.1.17
Given a martingale X and a stopping time τ, the stopped process X^τ is a martingale with respect to the filtration $\mathcal{F}_t$.

Proof X^τ is *càdlàg*, adapted and for a bounded stopping time σ, also $\sigma \wedge \tau$ is a stopping time. Since

$$E(X^\tau_\sigma) = E(X_{\sigma\wedge\tau}) = E(X_0) = E(X^\tau_0)$$

the previous proposition implies that X^τ is a martingale. ■

As a consequence we get the optional stopping theorem $X^\sigma = E(X^\tau|\mathcal{F}_\sigma)$. The next lemmata will pave us the way to Doob's maximal inequality:

Lemma 1.1.18
Given a finite submartingale X_n, $0 \leq n \leq N$, it holds for every $\lambda > 0$ that

$$\lambda P\left(\left\{\sup_n X_n \geq \lambda\right\}\right) \leq E\left(X_n 1_{\left\{\sup\limits_n X_n \geq \lambda\right\}}\right) \leq E\left(|X_n| 1_{\left\{\sup\limits_n X_n \geq \lambda\right\}}\right).$$

Proof We define a stopping time by

$$\tau_n := \begin{cases} \inf\{n : X_n \geq \lambda\} \in \mathcal{F}_N & \text{for } \{n : X_n \geq \lambda\} \neq \emptyset \\ N \in \mathcal{F}_N & \text{for } \{n : X_n \geq \lambda\} = \emptyset. \end{cases}$$

By optional sampling (Lemma 1.1.13) we get

$$\begin{aligned} E(X_N) &\geq E(X_\tau) = E\left(X_\tau 1_{\left\{\sup\limits_n X_n \geq \lambda\right\}}\right) + E\left(X_\tau 1_{\left\{\sup\limits_n X_n < \lambda\right\}}\right) \\ &\geq \lambda P\left(\left\{\sup_n X_n \geq \lambda\right\}\right) + E\left(X_N 1_{\left\{\sup\limits_n X_n < \lambda\right\}}\right). \end{aligned}$$

since $X_\tau \geq \lambda$ on $\left\{\sup_n X_n \geq \lambda\right\}$ and by the definition of the stopping time (there is no infimum for the second term). A simple subtraction achieves the left inequality, the right one is trivial. ■

For the rest of this work we will use X^* as abbreviation for $\sup_n |X_n|$.

Lemma 1.1.19
For a finite martingale (or a finite positive submartingale) X_n, $0 \leq n \leq N$ the inequality

$$\lambda^p P(X^* \geq \lambda) \leq E(|X_N|^p)$$

holds for $\lambda > 0$ and $p \geq 1$; for $p > 1$ we have

$$E(|X_N|^p) \leq E\left(\sup_n |X_n|^p\right) \leq \left(\frac{p}{p-1}\right)^p E(|X_N|^p).$$

Proof If $E(|X_N|^p) < \infty$, Jensens's inequality implies that $|X_n|^p$ (since convex) is a submartingale

$$E(|X_n|^p|\mathcal{F}_s) \geq |E(X_n|\mathcal{F}_s)|^p \geq |X_s|^p$$

and the previous lemma entails the first inequality.
For the second part we have to concentrate on the right equation since the left one is obvious. The previous lemma for the process $|X_n|$ gives

$$\lambda P(X^* \geq \lambda) \leq E\left(|X_N| 1_{\{X^* \geq \lambda\}}\right)$$

which we will use by estimating for a fixed $k < 0$:

$$\begin{aligned}
E((X^* \wedge k)^p) &= E\left(\int_0^{X^* \wedge k} p\lambda^{p-1} d\lambda\right) \\
&= E\left(\int_0^k p\lambda^{p-1} 1_{[0,X^*]} d\lambda\right) = E\left(\int_0^k p\lambda^{p-1} 1_{[0,X^*]} d\lambda\right) \\
&= E\left(\int_0^k p\lambda^{p-1} 1_{\{X^* \geq \lambda\}} d\lambda\right) = \int_0^k p\lambda^{p-1} P(\{X^* \geq \lambda\}) d\lambda \\
&\leq \int_0^k p\lambda^{p-1} E\left(|X_N| 1_{\{X^* \geq \lambda\}}\right) d\lambda = pE\left(|X_N| \int_0^{X^* \wedge k} \lambda^{p-2} d\lambda\right) \\
&= \frac{p}{p-1} E\left(|X_N|(X^* \wedge k)^{p-1}\right).
\end{aligned}$$

Now applying Hölder's inequality we get

$$E((X^* \wedge k)^p) \leq \frac{p}{p-1}\left(E\left((X^* \wedge k)^{p-1\frac{p}{p-1}}\right)\right)^{\frac{p-1}{p}} (E(|X_N|^p))^{\frac{1}{p}}$$

and hence

$$\begin{aligned}\left(E\left((X^*\wedge k)^p\right)\right)^p &\leq \left(\frac{p}{p-1}\right)^p \left(E\left((X^*\wedge k)^{p-1\frac{p}{p-1}}\right)\right)^{p-1} E\left(|X_N|^p\right)\\ E\left((X^*\wedge k)^p\right) &\leq \left(\frac{p}{p-1}\right)^p E\left(|X_N|^p\right)\end{aligned}$$

such that for $k\to\infty$ we get the desired result. ■

Until now we remained in the domain of finite (sub-)martingales, we will extend this now to get a quite more general version of Doob's maximal inequality.

Theorem 1.1.20 (Doob's Maximal Inequality)
Let $(X_t)_{t\in T}$ be a right-continuous martingale or a right-continuous positive submartingale for an index set T which is either an interval of $\mathbb{R}\cup\{\infty\}$ or a countable subset of an interval. Then it holds for $p\geq 1$, $\lambda>0$ that

$$\lambda^p P\left(X^*\geq\lambda\right)\leq \sup_t E\left(|X_t|^p\right)$$

and for $p>1$

$$\|X^*\|_p\leq \frac{p}{p-1}\sup_t \|X_t\|_p.$$

Proof If T is an interval we choose a countable dense subset $D\subset T$. We can further choose an increasing sequence of finite subsets $D_n\subset D$ such that $\bigcup_{n=0}^{\infty} D_n = D$ which enables us to use the above lemma for the D_n. Since $E\left(|X_t|^p\right)$ increases with t, we get by passing to the limit $n\to\infty$

$$\lambda^p P\left(\sup_{t\in D}|X_t|\geq\lambda\right)\leq \sup_{t\in D} E\left(|X_t|^p\right)$$

respectively

$$E\left(\sup_{t\in D}|X_t|^p\right)\leq\left(\frac{p}{p-1}\right)^p \sup_{t\in D} E\left(|X_t|^p\right).$$

But since X_t is right continuous we have $\sup_{t\in D}|X_t| = X^*$ and get the desired result by taking the p-th root in the second inequality. ■

Note that we made no assumption of completeness or right-continuity on the filtration $\mathcal{F}_t$.

Along with semimartingales which we will encounter in the context of integration theory in Section 2.4, *local martingales* are the most important generalization of the notion of the martingale.

Definition 1.1.21 (Local Martingale)
Given a filtered probability space with a right continuous filtration. An adapted process M_t is called a local martingale, iff there exists an increasing sequence $\tau_n\uparrow\infty$ of stopping times such that $M^{\tau_n}-M_0$ is a true martingale.

Proposition 1.1.22 (Localization)
Given an increasing sequence of stopping times $\tau_n \uparrow \infty$, then the following statements are equivalent:
(i) M is a local martingale.
(ii) M^{τ_n} is a local martingale for every n.

Proof
(i) $\Rightarrow$ (ii) Since M is a local martingale, there is a localizing sequence $\sigma_n \uparrow \infty$ such that M^{σ_n} is a true martingale. This holds true as well for an arbitrary stopping time $(M^{\sigma_n})^\tau = (M^\tau)^{\sigma_n}$ by optional stopping. But this implies that M^τ is a local martingale with localizing sequence σ_n. And since τ was arbitrarily chosen it holds especially for the τ_n.

(ii) $\Rightarrow$ (i) Given the localizing sequences for the local martingales M^{τ_n} by $(\sigma_k^n) \uparrow \infty$ for $k \to \infty$. We may choose indices k_n such that

$$P\left(\{\sigma_{k_n}^n < \tau_n \wedge n\}\right) \leq 2^{-n}, n \geq 0.$$

The Borel-Cantelli Lemma (see [Kal 02], p.56) implies that for $\tau_n' := \sigma_k^n \wedge \tau_n$ we have $\tau_n' \to \infty$ a.s. for $n \to \infty$. By defining $\tau_n'' := \inf_{m \geq n} \tau_m'$ we get a sequence with monotone limit $\tau_n'' \uparrow \infty$ a.s. for $n \to \infty$. To conclude it is enough to remark that the $M^{\tau_n'}$ are clearly true martingales and hence the $M^{\tau_n''} = \left(M^{\tau_n'}\right)^{\tau_n''}$ are true martingales too, and so M is a local martingale.

■

Every local martingale can be chosen uniformly integrable since if τ_n is a localizing sequence, we we can set $\sigma_n := \tau_n \wedge \tau$ which implies that σ_n is a localizing sequence too, and $M^{\sigma_n} = M^{\tau_n \wedge \tau}$ is u.i. In the same way any continuous local martingale can - by setting $\sigma_n := \tau_n \wedge \inf\{t : |M_t| = n\}$ - be chosen bounded.

1.2 Tensor Products

Here is neither the space nor the place for a fundamental treatment of algebraic tensor theory, here we will only give the main definitions and central theorems which will be needed in the following.

Definition 1.2.1 (Tensor Product)
Given two vector spaces E, F over the same field $\mathbb{K}$, we denote the vector space of bilinear forms on $E \times F$ by $B(E, F)$. For each pair $(x, y) \in E \times F$ the mapping $u_{x,y} : B(E, F) \to \mathbb{K}$ given by $f \mapsto f(x, y)$ is an element of $B(E, F)^*$, the algebraic dual of $B(E, F)$. Now we can see that there is a unique bilinear mapping $E \times F \to B(E, F)^*$ given by $\chi : (x, y) \mapsto u_{x,y}$. The linear hull of $\chi(E \times F)$ in the dual $B(E, F)^*$ is denoted by $E \otimes F$ and called the tensor (or direct) product of E and F; the embedding map $\chi : E \times F \hookrightarrow E \otimes F$ is called the canonical bilinear map of $E \times F$ into $E \otimes F$. We will apply the same notation for the element $u_{x,y}$ which we will now write as $x \otimes y$.

This notions can easily be expanded to n-fold tensor products on multilinear forms which also arise as compositions of tensor products on bilinear forms. These are associative, so we do not have to struggle with brackets. The main properties of the tensor product are the following:

Lemma 1.2.2 (Properties of the Tensor Product)
(i) For $\lambda \in \mathbb{K}$ and x_i, y_i elements of the vector space E respective F the following rules hold:
$\lambda(x \otimes y) = (\lambda x) \otimes y = x \otimes (\lambda y)$
$(x_1 + x_2) \otimes y = x_1 \otimes y + x_2 \otimes y$ *and* $x \otimes (y_1 + y_2) = x \otimes y_1 + x \otimes y_2$
(ii) Every element $u \in E \otimes F$ can (not uniquely!) be represented as $u = \sum_{i=1}^{r} x_i \otimes y_i$ where the minimal such r is called the rank of u.
(iii) For the tensor product and the direct sum holds the distributive law. Given vector spaces E, F, G it holds that $E \otimes (F \oplus G) = (E \otimes F) \oplus (E \otimes G)$.

In the following we will be concerned only with Hilbert space tensor products over $\mathbb{R}$. The tensor product $H_1 \otimes H_2$ of two Hilbert spaces H_1 and H_2 with inner product

$$\langle h_1 \otimes h_2, g_1 \otimes g_2 \rangle_{H_1 \otimes H_2} := \langle h_1, g_1 \rangle_{H_1} \langle g_1, g_2 \rangle_{H_2}$$

for $h_i, g_i \in H_i$ is a Pre-Hilbert space. The closure of this Pre-Hilbert space is a Hilbert space which we will denote - without creating confusion - by $H_1 \otimes H_2$ too, and call it *the* tensor product of H_1 and H_2. Furthermore given Hilbert space bases $\{e_i\}_{i \geq 1}$ of H_1 and $\{f_j\}_{j \geq 1}$ of H_2, the set $\{e_i \otimes f_j\}_{i,j \geq 1}$ forms a basis of the Hilbert space $H_1 \otimes H_2$.

Looking for Hilbert spaces H_i, G_i at the bounded linear maps $\varphi_i : H_i \to G_i$ we define their tensor product as a mapping $H_1 \otimes H_2 \to G_1 \otimes G_2$ by

$$(\varphi_1 \otimes \varphi_2)(h_1 \otimes h_2) := \varphi_1(h_1) \otimes \varphi_2(h_2).$$

Lemma 1.2.3
For the tensor product of Hilbert space mappings it holds - for the Hilbert space norm $\|\cdot\|$ respectively the supremum norm for mappings - that
(i) $\|(\varphi_1 \otimes \varphi_2)(h_1 \otimes h_2)\|^2 = \|\varphi_1(h_1)\|^2 \|\varphi_2(h_2)\|^2$
(ii) $\|\varphi_1 \otimes \varphi_2\| = \|\varphi_1\| \|\varphi_2\|$

For the case of Hilbert spaces $H_i := L^2(\Omega_i, \mathcal{F}_i, \mu_i)$ with σ-finite measures μ_i we have the nice isometry

$$L^2(\Omega_1, \mathcal{F}_1, \mu_1) \otimes L^2(\Omega_2, \mathcal{F}_2, \mu_2) \simeq L^2(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$$

where $\Omega_1 \times \Omega_2$ is the topological product, $\mathcal{F}_1 \otimes \mathcal{F}_2$ the product σ-algebra and $\mu_1 \otimes \mu_2$ the product measure.

Chapter 2

Brownian Motion and Stochastic Integration

In this chapter it is our aim to develop stochastic integration. After an introduction to Gaussian processes we will give a primer on Brownian motion and then go on defining the stochastic integral with respect to Brownian motion. In the last sections we will try to generalize the notion of the stochastic integral by defining it for so called progressively measurable processes with respect to continuous semimartingales.

First we will take a look at continuity properties which we can do in a quite general setting. We will recall the definition of continuity in the sense of Hölder and then give the fundamental theorem of Kolmogorov and Čentsov which, under certain conditions, assures the existence of a modification which is Hölder continuous to a given stochastic process.

Definition 2.0.4 (Hölder Continuity)
A function $f : (S^1, \rho^1) \to (S^2, \rho^2)$ between two complete metric spaces S^1, S^2 with respective metrics ρ^1, ρ^2 is called Hölder continuous with exponent α, iff

$$\sup_{s \neq t} \left\{ \frac{\rho^2(f(s), f(t))}{\rho^1(s,t)^\alpha} : s, t \in S^1, \rho^1(s,t) < \infty \right\} < \infty.$$

It is called locally Hölder continuous iff it is Hölder continuous on every bounded set.

Theorem 2.0.5 (Kolmogorov-Čentsov Theorem)
Given a stochastic process X_t on $\mathbb{R}^d$ with values in a complete metric space (S, ρ) such that for some constants C, γ, $\varepsilon > 0$ and for all $s, t \in \mathbb{R}^d$ the following inequality holds:

$$E\left(\rho\left(X_s, X_t\right)^\gamma\right) \leq C \cdot |s-t|^{d+\varepsilon}. \tag{2.1}$$

Then there exists a continuous modification $\breve{X}_t$ with sample paths which are a.s. locally Hölder continuous with exponent α for every $\alpha \in [\,0, \frac{\varepsilon}{\gamma}\,[$.

Proof Here we will only develop the proof for the restriction of X_t to $[0,1]^d$, since for a generalized cube (where every bounded set has to be enclosed) it is the same up to a factor.

(i) First we define for positive integers m the set of points with coordinate entries which are dyadic rational numbers up to degree m:

$$\mathcal{D}_m := \{(k_1, ..., k_d) \cdot 2^{-m} : k_i \in \{0, 1, ..., 2^m\}, i \in \{1, ..., d\}, m \geq 0\}$$

and then by

$$\mathcal{D} := \bigcup_m \mathcal{D}_m,$$

the set of all points with purely dyadic coordinates in the unit cube. Furthermore we define

$$\Delta_m := \{(s,t) : s,t \in \mathcal{D}_m, |s-t| = 2^{-m}\},$$

the set of all pairs of adjoining points in $\mathcal{D}_m$. Obviously we can count exactly $|\Delta_m| = d \cdot 2^{md}$ of them.
Defining ξ_n by $\xi_n := \sup_{(s,t)\in\Delta_n} \rho(X_s, X_t)$ we can estimate it by the sum over all $(s,t) \in \Delta_n$. This fact and the given inequality (2.1) let us conclude

$$\begin{aligned} E(\xi_n^\gamma) &\leq E\left(\sum_{(s,t)\in\Delta_n} \rho(X_s,X_t)^\gamma\right) = \sum_{(s,t)\in\Delta_n} E\left(\rho(X_s,X_t)^\gamma\right) \\ &\leq \sum_{(s,t)\in\Delta_n} C \cdot |s-t|^{d+\varepsilon} \leq d2^{nd} \cdot C \cdot 2^{-n(d+\varepsilon)} \leq J2^{-n\varepsilon} \end{aligned} \tag{2.2}$$

for a constant J.

(ii) Abbreviating by $s_m := \sup\{r \in D_m : r \leq s\}$ the largest element in D_m not greater than s we can expand for $s,t \in D, |s-t| \leq 2^{-m}$

$$\rho(X_t, X_s) \leq \sum_{i=m}^{\infty} \rho\left(X_{t_{i+1}}, X_{t_i}\right) + \rho(X_{t_m}, X_{s_m}) + \sum_{i=m}^{\infty} \rho\left(X_{s_{i+1}}, X_{s_i}\right) \tag{2.3}$$

where the series are actually finite sums, whence

$$\rho(X_t, X_s) \leq \sum_{i=m+1}^{\infty} \xi_i + \xi_m + \sum_{i=m+1}^{\infty} \xi_i.$$

(iii) Setting now for $0 \leq \alpha \leq \frac{\varepsilon}{\gamma}$

$$M_\alpha := \sup\left\{\frac{\rho(X_s,X_t)}{|s-t|^\alpha} : s,t \in D, s \neq t\right\}$$

we get

$$\begin{aligned} M_\alpha &\leq \sup_{m\geq 0}\left\{\sup_{2^{-m-1}<|t-s|\leq 2^{-m}} \frac{\rho(X_s,X_t)}{|s-t|^\alpha} : s,t \in D, s \neq t\right\} \\ &\leq \sup_{m\geq 0}\left\{2^{(m+1)\alpha} \sup_{|t-s|\leq 2^{-m}} \rho(X_s,X_t) : s,t \in D, s \neq t\right\} \\ &\leq \sup_{m\geq 0}\left(2 \cdot 2^{(m+1)\alpha} \sum_{i=m}^{\infty} \xi_i\right) \leq 2^{\alpha+1} \sum_{i=0}^{\infty} 2^{i\alpha} \xi_i \end{aligned}$$

by (2.3).

(iv)To prove that $E\left(M_\alpha^\gamma\right)<\infty$ we have to discern the two cases $\gamma \geq 1$ and $\gamma < 1$:

$\gamma \geq 1$:

$$\begin{aligned}\left(E\left(M_\alpha^\gamma\right)\right)^{\frac{1}{\gamma}} &\leq \left(E\left(\left(2^{\alpha+1}\sum_{i=0}^{\infty}2^{i\alpha}\xi_i\right)^\gamma\right)\right)^{\frac{1}{\gamma}} \leq 2^{\alpha+1}\sum_{i=0}^{\infty}\left(2^{i\alpha}\left(E\left(\xi_i^\gamma\right)^{\frac{1}{\gamma}}\right)\right)\\ &\leq 2^{\alpha+1}\sum_{i=0}^{\infty}\left(2^{i\alpha}\left(J2^{-i\varepsilon}\right)^{\frac{1}{\gamma}}\right) \leq R\sum_{i=0}^{\infty}\left(2^{i(\alpha-\frac{\varepsilon}{\gamma})}\right)<\infty.\end{aligned}$$

by (2.2) for a constant R.

$\gamma < 1$:

$$\begin{aligned}\left(E\left(M_\alpha^\gamma\right)\right) &\leq \left(E\left(\left(2^{\alpha+1}\sum_{i=0}^{\infty}2^{i\alpha}\xi_i\right)^\gamma\right)\right) \leq 2^{(\alpha+1)\gamma}E\left(\sum_{i=0}^{\infty}\left(2^{i\alpha}\xi_i^\gamma\right)\right)\\ &\leq 2^{(\alpha+1)\gamma}\sum_{i=0}^{\infty}\left(2^{i\alpha\gamma}E\left(\xi_i^\gamma\right)\right) \leq 2^{(\alpha+1)\gamma}\sum_{i=0}^{\infty}\left(2^{i\alpha\gamma}J2^{-\varepsilon}\right)\\ &\leq 2^{(\alpha+1)\gamma}J\sum_{i=0}^{\infty}2^{i(\alpha\gamma-\varepsilon)}<\infty.\end{aligned}$$

(v) This enables us to show easily that the mapping $t \rightarrow X_t(\omega)$ is a.s. uniformly continuous for $\alpha < \frac{\varepsilon}{\gamma}$: We define Ω_α as the subset with uniformly continuous sample paths of order α; it is surely in $\mathcal{F}$ since

$$\Omega_\alpha = \bigcup_{n\geq 0}\bigcap_{s,t\in D}\{\rho(X_s,X_t)\leq n|t-s|^\alpha\} = \bigcup_{n\geq 0}\{M_\alpha \leq n\}.$$

It follows that $P(\Omega_\alpha)=1$ since $E\left(M_\alpha^\gamma\right)<\infty$ and $P(\Omega_0)=1$ since $\Omega_0 = \bigcap_{0\leq\alpha<\frac{\varepsilon}{\gamma}}\Omega_\alpha$ is an intersection of decreasing measurable sets.

(vi) It only remains to construct our desired modification $\breve{X}_t$:

$$\breve{X}_t := \begin{cases}\lim_{s\to t, s\in D} X_s(\omega) & \text{for } \omega\in\Omega_0\\ 0 & \text{for } \omega\notin\Omega_0\end{cases}$$

By definition as limit $\breve{X}_t$ is measurable, it is a.s. uniformly Hölder continuous and it is a modification of X_t. For a sequence $t_n \rightarrow t$ we have for $\gamma \geq 1$

$$\begin{aligned}\left(E\left(\rho\left(\breve{X}_t,X_t\right)^\gamma\right)\right)^{\frac{1}{\gamma}} &\leq \left(E\left(\rho\left(\breve{X}_t,\breve{X}_{t_n}\right)^\gamma\right)\right)^{\frac{1}{\gamma}} + \left(E\left(\rho\left(X_{t_n},X_t\right)^\gamma\right)\right)^{\frac{1}{\gamma}}\\ &\leq 2C\cdot|s-t|^{\frac{d+\epsilon}{\gamma}}\end{aligned}$$

and for $\gamma < 1$ directly

$$E\left(\rho\left(\breve{X}_t,X_t\right)^\gamma\right) \leq E\left(\rho\left(\breve{X}_t,\breve{X}_{t_n}\right)^\gamma\right)+E\left(\rho\left(X_{t_n},X_t\right)^\gamma\right) \leq 2C\cdot|s-t|^{d+\epsilon} \quad (2.4)$$

since $X_{t_n} = \breve{X}_{t_n}$ by definition. Applying Fatou's Lemma (see [Kal 02], p.11) on this equation (2.4) gives as the result $X_t = \breve{X}_t$ a.s. ■

2.1 Gaussian Processes

Here we will give some general results on Gaussian processes which we will need for the construction of Brownian motion or which can help to understand and appreciate this fundamental notion of stochastic analysis. We will come back to this subject in Chapter 3 when we will give the decomposition of Wiener chaos.

Definition 2.1.1 (Gaussian Process)
A real-valued stochastic process X_t, $t \in T$ is called Gaussian, iff for any choice of $t_1, ..., t_n \in T$ and $c_1, ...c_n \in \mathbb{R}$ for any $n \geq 0$, the r.v. $\sum_{k=1}^{n} c_k X_{t_k}$ is Gaussian.

A Gaussian process is called centered, iff for all $t \in T$ $E(X_t) = 0$. Processes X^i on T_i, $i \in I$ arbitrary, are called jointly Gaussian iff the combined process $X := \{X_t^i : t \in T_i, i \in I\}$ is Gaussian. Obviously this is the case if the processes X^i are independent and Gaussian. Note that the combined process is not necessarily Gaussian if the sole processes are Gaussian, but not independent.

The following theorem motivates the use of Gaussian processes to construct Brownian motion and shows that they arise naturally from the independent increments.

Theorem 2.1.2 (Independent Increments and Gaussian Processes)
Let $(X_t)_{0 \leq t \leq T}$ be a continuous process on $\mathbb{R}^d$ with independent increments and $X_0 = 0$ a.s. Then X_t is a Gaussian process and there exist two continuous process

$$a_t : [0, T] \to \mathbb{R}^d$$
$$b_t : [0, T] \to M_d, (b_s - b_r) \geq 0 \text{ for } 0 \leq r < s \leq T$$

such that $X_s - X_r$ is $N(a_s - a_r, b_s - b_r)$.

Proof For given $r, s \in [0, T]$ we divide the interval $[r, s]$ in n subintervals of equal length. Given $u \in \mathbb{R}^d$, we denote the corresponding increments of $u(X_s - X_r)$ by $\xi_{n1}, ..., \xi_{nn}$. The continuity of X_t yields that $\sup_j |\xi_{nj}| \to 0$ for $n \to \infty$ and we can conclude by the results on Gaussian convergence (see [Kal 02], p.92ff) that $u(X_s - X_r) = \sum_{j=1}^{n} \xi_{nj}$ is a Gaussian r.v. Since the increments are independent this implies that X_t is Gaussian. We define

$$a_t := E(X_t)$$
$$b_t := cov(X_t)$$

and get

$$E(X_s - X_r) = E(X_s) - E(X_r) = a_s - a_r$$
$$0 \leq cov(X_s - X_r) = cov(X_s) - cov(X_r) = b_s - b_r$$

for $0 \leq r < s \leq T$.The continuity of X_t implies for $s \to r$ that $X_s \to X_r$ a.s. and therefore in distribution. The same holds for a_t and b_t, so both functions are continuous. ■

2.2 Brownian Motion

The term mathematical Brownian Motion traces back to the physical phenomenon of the same name which was first studied by the Scottish botanist Robert Brown in 1827. He suspended pollen of the flower *Clarkia pulchella* in water and observed it with his microscope. What he could see was a rapid oscillatory motion of the pollen grains. This phenomenon had already been observed by other researchers like Antonie Van Leeuwenhoek, but Brown was actually the first one who really studied it. With the development of modern physics, this phenomenon became explainable: The molecules of water (as the molecules of every liquid or gas) are constantly in motion, colliding with each other, bouncing back and forth which sets the pollen grains in motion. Albert Einstein introduced the Brownian Motion into mathematical theory by analyzing the main features of the physical phenomenon, but scaling down from the discrete bouncing to a continuous version. He defined the mathematical Brownian motion as Gaussian process with continuous paths and independent increments. Even five years before Einstein, Louis Bachelier came with a different motivation to the same process in his dissertation *Théorie de la Spéculation.* But it was only Norbert Wiener who in 1923 treated this subject with full mathematical rigor by constructing the Wiener measure.

Definition 2.2.1 (Wiener Process)
A Gaussian Process $(W_t)_{t\geq 0}$ is called a Wiener Process, iff it satisfies the following conditions:
(i) $E(W_t) = 0$
(ii) $cov(W_s, W_t) = E(W_s W_t) = s \wedge t$ for $s, t \geq 0$.

Theorem 2.2.2 (Existence of the Wiener Process)
To a given probability space $(\Omega, \mathcal{F}, P)$ and a sequence of independent identically N(0,1)-distributed r.v.s $(X_n)_{n\geq 1}$, there exists a process satisfying the above stated conditions.

Proof The space $L^2(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}), dx)$ of square-integrable linear functionals defined on the Borel-measurable sets of the non-negative part of the real line is a Hilbert space, and so we can pick an orthonormal basis $(e_i)_{i\geq 1}$. As we have seen above the span of the $(X_n)_{n\geq 1}$ forms a closed subspace $\mathcal{H}_1$of the Hilbert space $L^2(\Omega, \mathcal{F}, P)$. So we can for any $f, g \in L^2(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}), dx)$ define the linear Hilbert space isometry $\eta : L^2(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}), dx) \rightarrow \mathcal{H}_1 \subset L^2(\Omega, \mathcal{F}, P)$ requiring that it satisfies

$$\langle \eta(f), \eta(g) \rangle_{\mathcal{H}_1} = E\left(\eta(f)\eta(g)\right) = \int_0^\infty f(u)g(u)du = \langle f, g \rangle_{L^2(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}), dx)} .$$

Now we can define the Wiener process by

$$W_t := \eta(1_{[0,t]}).$$

By definition (the r.v. $(X_n)_{n\geq 1}$ were given as centered) the expectation $E(W_t)$ is zero. For the covariance we get by the Hilbert space isometry

$$\begin{aligned} E(W_sW_t) &= E\left(\eta(1_{[0,s]})\eta(1_{[0,t]})\right) = \int_0^\infty 1_{[0,s]}(u)1_{[0,t]}(u)du \\ &= \int_0^\infty 1_{[0,s\wedge t]}(u)du = s\wedge t \end{aligned}$$

as desired. ■

Obviously we can not hope to have something like uniqueness of the Wiener process. - In contrary, there exists a lot of them.

The following corollary of the Kolmogorov-Čentsov theorem shows us that there exists a version of the Wiener process with continuous paths.

Corollary 2.2.3
To every Wiener process $(W_t)_{t\geq 0}$ on a given probability space $(\Omega,\mathcal{F},P)$ there exists a modification $(B_t)_{t\geq 0}$ with Hölder continuous paths for any exponent $\alpha<\frac{1}{2}$.

Proof Since the Wiener process W_t is Gaussian, its characteristic function is $e^{-\frac{tu^2}{2}}$. Writing W_t as

$$W_t = W_s + (W_t - W_s), t > s$$

we get - using the independence of B_s and B_{t-s} - for the characteristic functions

$$e^{-\frac{tu^2}{2}} = e^{-\frac{su^2}{2}} E\left(e^{iu(W_t-W_s)}\right)$$

and so

$$E\left(e^{iu(W_t-W_s)}\right) = e^{-\frac{(t-s)u^2}{2}}. \tag{2.5}$$

Writing the exponential function as series, this is

$$E\left(\sum_{n=0}^{\infty}\frac{i^nu^n(W_t-W_s)^n}{n!}\right) = \sum_{n=0}^{\infty}\frac{(-1)^n\frac{(t-s)^nu^{2n}}{2^n}}{n!} = (-1)^n\sum_{n=0}^{\infty}\frac{(t-s)^nu^{2n}}{2^nn!}$$

which enables us to compare the coefficients of u:

$$E\left(\frac{i^{2k}(W_t-W_s)^{2k}}{(2k)!}\right) = (-1)^k\frac{(t-s)^k}{2^kk!},$$

whence

$$E\left((W_t-W_s)^{2k}\right) = \frac{(2k)!(t-s)^k}{2^kk!}.$$

Setting $\gamma = 2k$ and $\varepsilon = k-1$ the Kolmogorov-Čentsov theorem states that there exists a locally Hölder continuous modification of order $0\leq\alpha<\frac{k-1}{2k}$. All these modification are clearly modifications of each other and since they are continuous too, they are indistinguishable. So by $k \to \infty$ we get that there exists a modification B_t which is locally uniformly Hölder continuous for any exponent $\alpha < \frac{1}{2}$. ■

We remark here only that there exist no locally Hölder continuous modification for any $\alpha < \frac{1}{2}$.

To proceed to Brownian motion we have only to add a filtration.

Definition 2.2.4 (Brownian Motion)
The standard Brownian Motion is a $\mathbb{R}$-valued process $(B_t)_{t\geq 0}$, $B_0 = 0$ on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ satisfying the following conditions:
(i) The filtration $\mathcal{F}_t$ satisfies the usual conditions and converges to $\mathcal{F}$.
(ii) The process $(B_t)_{t\geq 0}$ is adapted to the filtration $\mathcal{F}_n$.
(iii) The increments $B_s - B_t$ of the process are independent of $\mathcal{F}_t$ for $s \geq t$.

The adjective standard refers to the variance of the process. By Theorem 2.1.2 the third condition implies that Brownian motion is a Gaussian process and by $B_s = B_t + (B_s - B_t)$, it also implies that Brownian motion is a continuous martingale. But we have yet to show that it exists:

Theorem 2.2.5 (Existence of Brownian Motion)
A Wiener Process with its natural filtration satisfies the above stated conditions.

Proof We only have to check if a Wiener process with its natural filtration satisfies the conditions stated in definition 2.2.4. That the process is adapted to its natural filtration is trivial, hence (ii) is fulfilled. The same for (iii) since the Wiener process has independent increments and the filtration is the natural one. Concerning condition one, completeness and convergence of the filtration are obvious too. It remains only to show that the filtration is right continuous, i.e. $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$.
Therefore we define a σ-algebra $\mathcal{G}_t := \sigma(B_v - B_u : v \geq u > t)$ and prove as a first step that $\mathcal{F} = \sigma(\mathcal{F}_t \cup \mathcal{G}_t)$. Clearly $\sigma(\mathcal{F}_t \cup \mathcal{G}_t) \subset \mathcal{F}$, but since $B_v - B_t = \lim_{n\to\infty} B_v - B_{t+\frac{1}{n}}$ is $\mathcal{G}_t$-measurable we can see that every $B_v = (B_v - B_t) + B_t$, $v < t$ is $\sigma(\mathcal{F}_t \cup \mathcal{G}_t)$-measurable, whence $\mathcal{F} \subset \sigma(\mathcal{F}_t \cup \mathcal{G}_t)$.
We have to show that every $A \in \mathcal{F}_{t+}$ is also $\mathcal{F}_t$-measurable or, in another formulation, that for every $A \in \mathcal{F}_t \cup \mathcal{G}_t$ the conditional expectation $E(1_A|\mathcal{F}_{t+})$ is $\mathcal{F}_t$-measurable. This is trivial for the case $A \in \mathcal{F}_t$, otherwise $A \in \mathcal{G}_t$, which leads us to construct a so called π-system for $\mathcal{G}_t$, a collection of subsets of $\mathcal{F}$ closed under finite intersections generating $\mathcal{G}_t$: The inverse images under $B_v - B_u$, $v \geq u > t$ are already a π system generating $\mathcal{G}_t$ generically. We observe that for any $t \leq t_1 < t_2 < t_3 < t_4$

$$\sigma(B_{t_3} - B_{t_1}, B_{t_4} - B_{t_2}) \subset \sigma(B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, B_{t_4} - B_{t_3})$$

which entails that we we can form this π-system by intersection of independent measurable sets. Since $\mathcal{F}_{t+}$ is independent of every $B_v - B_u$, $v \geq u > t$ we have for $A \in \mathcal{G}_t$

$$E(1_A|\mathcal{F}_{t+}) = E(1_A)$$

which is $\mathcal{F}_t$-measurable.
The conclusion follows by a monotone class argument: Monotone limits of $\mathcal{F}_t$-measurable sets remain $\mathcal{F}_t$-measurable and for the sets which are finite intersections of sets in $\mathcal{F}_t$ and $\mathcal{G}_t$ we have for $A \in \mathcal{F}_t$, $B \in \mathcal{F}_t$ by independence

$$E(1_{A\cap B}|\mathcal{F}_{t+}) = 1_A E(1_B|\mathcal{F}_{t+})$$

which is $\mathcal{F}_t$-measurable too, and remains it for monotone limits.
To recapitulate the argument: For every $A \in \mathcal{F}$, $E(1_A|\mathcal{F}_{t+})$ is $\mathcal{F}_t$-measurable, so for $A \in \mathcal{F}_{t+}$ we can conclude that $E(1_A|\mathcal{F}_{t+}) = 1_A$ which proves the right continuity of the filtration. ■

As in the case of the Wiener Process we cannot hope to have something like uniqueness of the Brownian motion. It is easy to generalize the concept of Brownian motion: We define the d-dimensional Brownian motion as d-dimensional vector whose entries are independent copies of B_t.

2.3 Itô Integration

Having introduced Brownian motion one could ask if it can be used as integrator to define integrals along the paths of Brownian motion. Since integration theory in the sense of Stieltjes requires the integrator to be locally of finite total variation we now have to look on the variation properties of Brownian motion.

Theorem 2.3.1 (Variation Properties of Brownian Motion)
For a 1-dimensional Brownian motion B_t
(i) the quadratic variation process $\langle B, B\rangle_t$ *equals a.s.* t.
(ii) the total variation process $S_T(B)$ *is a.s. infinite for* $T > 0$.

Proof
(i) We want to show that $T_t^{\Delta_n}(B) \to t$ in L^2 for every refining sequence of partitions Δ_n with mesh $|\Delta_n| \to 0$ for $n \to \infty$. For a concrete partition we calculate directly

$$\begin{aligned}
E\left(\left(T_t^{\Delta}(B) - t\right)^2\right) &= E\left(\left(\sum_{i=0}^{n-1}\left(B_{t_{i+1}} - B_{t_i}\right)^2 - t\right)^2\right) \\
&= E\left(\left(\sum_{i=0}^{n-1}\left(\left(B_{t_{i+1}} - B_{t_i}\right)^2 - (t_{i+1} - t_i)\right)\right)^2\right) \\
&= E\left(\sum_{i=0}^{n-1}\left(\left(B_{t_{i+1}} - B_{t_i}\right)^2 - (t_{i+1} - t_i)\right)^2\right) \\
&\quad +2E\left(\sum_{0=i<j}^{n-1}\left(\left(B_{t_{i+1}} - B_{t_i}\right)^2 - (t_{i+1} - t_i)\right)\left(\left(B_{t_{j+1}} - B_{t_j}\right)^2 - (t_{j+1} - t_j)\right)\right).
\end{aligned}$$

By independence of increments we get for the second term

$$2\sum_{0=i<j}^{n-1}\left(E\left(\left(B_{t_{i+1}}-B_{t_i}\right)^2-(t_{i+1}-t_i)\right)E\left(\left(B_{t_{j+1}}-B_{t_j}\right)^2-(t_{j+1}-t_j)\right)\right)$$

which is obviously zero, so

$$E\left(\left(T_t^{\Delta}(B)-t\right)^2\right)=E\left(\sum_{i=0}^{n-1}\left(\left(B_{t_{i+1}}-B_{t_i}\right)^2-(t_{i+1}-t_i)\right)^2\right).$$

Now we take a $N(0,1)$-distributed r.v. Z and observe that

$$\begin{aligned} E(B_{t+\delta}-B_t) &= 0 &= E(\sqrt{\delta}Z) \\ E\left((B_{t+\delta}-B_t)^2\right) &= \delta &= E\left((\sqrt{\delta}Z)^2\right) \end{aligned}$$

implying that $B_{t+\delta}-B_t$ and $\sqrt{\delta}Z$ have the same distribution and

$$E\left(\left((B_{t+\delta}-B_t)^2-\delta\right)^2\right)=E\left((\delta Z^2-\delta)^2\right)=\delta^2E\left((Z^2-1)^2\right)=:\delta^2C.$$

So

$$E\left(\left(\left(B_{t_{i+1}}-B_{t_i}\right)^2-(t_{i+1}-t_i)\right)^2\right)=(t_{i+1}-t_i)^2C,$$

whence

$$E\left(\left(T_t^{\Delta}(B)-t\right)^2\right)=C\sum_{i=0}^{n-1}(t_{i+1}-t_i)^2\le C\sum_{i=0}^{n-1}|t_{i+1}-t_i|^2\sup_i|t_{i+1}-t_i|\to 0$$

since $|\Delta_n|\to 0$ for $n\to\infty$ as desired.

(ii) Assuming indirectly $S_T(B)<\infty$ for an arbitrary $T>0$, then for a partition Δ_n, $0=t_0\le ...\le t_n=T$

$$\begin{aligned} &\sum_{i=0}^{n}|B_{t_{i+1}}-B_{t_i}|^2\le\sup_i\left(|B_{t_{i+1}}-B_{t_i}|\right)\sum_{i=0}^{n}|B_{t_{i+1}}-B_{t_i}|= \\ = \quad &\sup_i\left(|B_{t_{i+1}}-B_{t_i}|\right)S_T^{\Delta_n}(B)\le\sup_i\left(|B_{t_{i+1}}-B_{t_i}|\right)S_T(B). \end{aligned}$$

But taking a sequence of partitions with mesh $|\Delta_n|=\sup_i|t_{i+1}-t_i|$ tending to zero we have $\sup_i|B_{t_{i+1}}-B_{t_i}|\to 0$ by continuity of Brownian motion, implying $\sum_{i=0}^{n}|B_{t_{i+1}}-B_{t_i}|^2\to 0$. This contradicts the result of the first part of this theorem, namely that $\sum_{i=0}^{n-1}|B_{t_{i+1}}-B_{t_i}|^2\to T$. So the assumption has to be false and $S_T(B)=\infty$ a.s. ■

As it is not possible to use Stieltjes integrals, we have to search for a new notion of integral - and it helps to know that the quadratic variation is finite on every interval $[0,T]$: The idea is to define our "stochastic integral" as L^2-limit of Riemannian sums. But first we have to define our "playground":

We denote the set of all square integrable, progressively measurable processes $\Lambda : \mathbb{R}_{\geq 0} \times \Omega \to \mathbb{R}$ (i.e. progressively measurable processes satisfying $E\left(\int_0^\infty \varphi(s)^2 ds\right) = \int_\Omega \int_0^\infty \varphi(s,\omega)^2 ds dP(\omega) < \infty$) by $L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P)$, these are the processes for which we expect a notion of stochastic integral.

For the beginning we have to look for a more convenient subspace, the space $\mathcal{E}$ of predictable step processes, i.e. processes of the form

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

with $F_i \in L^2(\Omega, \mathcal{F}_{t_i}, P)$ and a partition $0 = t_0 \leq ... \leq t_n$.

Definition 2.3.2 (Itô Integral for Predictable Step Processes)
For processes $u_t \in \mathcal{E}$ we define the Itô integral (or stochastic integral) with respect to the 1-dimensional Brownian motion by

$$I(u) := \sum_{i=0}^{n-1} F_i (B_{t_{i+1}} - B_{t_i}).$$

To expand our definition to progressively measurable processes we have to prove the following theorem:

Theorem 2.3.3 (Approximation of Prog. Measurable Processes)
The vector space $\mathcal{E}$ is dense in $L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P)$.

Proof We have to give an approximation of an arbitrarily chosen element $u \in L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P)$. Since progressive measurability is defined by measurability on intervals it is sufficient to proof this for $[0,1]$ instead of $\mathbb{R}_{\geq 0}$.
In a first step we approximate the progressive measurable (and square integrable - but this goes by itself, we do not have to care about integrability) process by bounded progressive measurable processes. This goes easily by cutting down, e.g. $u' := u \wedge n$, $n \to \infty$.
The second step is to approximate a bounded progressively measurable process u' by a sequence of continuous, adapted processes. We postulate that for $t, h \geq 0$ with $h \to 0$

$$u'_h(t) := \frac{1}{h} \int_{t-h}^{t} u'(s) ds$$

is such a sequence, assuming $u'(s) = 0$ for $s < 0$ to avoid problems in the neighborhood of 0. The u'_h are clearly continuous and they are adapted, since the integral is $\mathcal{B}([0,T]) \otimes \mathcal{F}_t$-measurable. It remains to prove that the sequence converges in the sense of $L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P)$: Since the $u'_h(t)$ are non decreasing, Lebesgue's theorem on differentiation asserts that

$$u'_h(t) \to u'(t) a.e.$$

Now we can - thanks to the boundedness of u' - use Lebesgue's dominated convergence theorem to get

$$\int_0^1 |u_h'(s)(\omega) - u'(s)(\omega)|^2 ds \to 0$$

for $h \to 0$ for every $\omega \in \Omega$. The same argument combined with Fubini's theorem on multiple integrals with respect to product measures asserts

$$E\left(\int_0^1 |u_h'(s)(\omega) - u'(s)(\omega)|^2 ds\right) \to 0$$

for $h \to 0$ as desired.
As the third and last step we have to show that every continuous, adapted process can be approximated by predictable step processes, but this can be done in a very classical way using a refining sequence of partitions. ■

This enables us now to define the Itô integral in a far more general setting - for progressively measurable processes.

Definition 2.3.4 (Itô Integral)
The unique continuous extension

$$I : L^2\left(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P\right) \to L^2\left(\Omega, \mathcal{F}, P\right)$$

of the integral for predictable step processes is called the Itô integral with respect to Brownian motion and is denoted by

$$\int_0^\infty u_t dB_t := I(u)$$

.

For $t \geq 0$ we can define the definite integral by

$$\int_0^t u_s dB_s := \int_0^\infty u_s 1_{[0,t]} dB_s.$$

Theorem 2.3.5 (Itô Lemma)
The mapping $I : L^2\left(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P\right) \to L^2\left(\Omega, \mathcal{F}, P\right)$ is a well defined Pre-Hilbert space isometry; for all $u, v \in L^2\left(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P\right)$ it holds that

$$E(I(u)I(v)) = E\left(\int_0^\infty u_t v_t dt\right)$$

with expectation zero: $E(I(u)) = 0$.

Proof First we prove the theorem for predictable step processes: For the expectation we calculate for every $u \in \mathcal{E}$

$$E(I(u)) = E\left(\sum_{i=0}^{n-1} F_i(B_{t_{i+1}} - B_{t_i})\right) = \sum_{i=0}^{n-1} E\left(F_i E(B_{t_{i+1}} - B_{t_i}|\mathcal{F}_{t_i})\right) = 0$$

using elementary facts of the conditional expectation.
The same type of argument can be used for the inner product: Having $u_t := \sum_{i=0}^{n-1} F_i 1_{[r_i,r_{i+1}]}$ and $v_t := \sum_{i=0}^{n-1} F_i 1_{[s_i,s_{i+1}]}$ it is easy to see that there has to be a common partition $0 = t_0 \leq ... \leq t_n$ such one can write $u_t := \sum_{i=0}^{n-1} F_i 1_{[t_i,t_{i+1}]}$ and $u_t := \sum_{i=0}^{n-1} F_i 1_{[t_i,t_{i+1}]}$, so

$$\begin{aligned}
E(I(u)I(v)) &= E\left(\sum_{i=0}^{n-1} F_i(B_{t_{i+1}} - B_{t_i}) \sum_{i=0}^{n-1} G_j(B_{t_{j+1}} - B_{t_j})\right) \\
&= E\left(\sum_{i=0}^{n-1} F_i G_i (B_{t_{i+1}} - B_{t_i})^2\right) \\
&\quad + E\left(\sum_{i=0}^{n-1} F_i G_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right) \\
&= \sum_{i=0}^{n-1} E\left(F_i G_i E\left((B_{t_{i+1}} - B_{t_i})^2 |\mathcal{F}_{t_i}\right)\right) \\
&\quad + \sum_{i=0}^{n-1} E\left(F_i G_j (B_{t_{i+1}} - B_{t_i}) E(B_{t_{j+1}} - B_{t_j}|\mathcal{F}_{t_j})\right) \\
&= E\left(\sum_{i=0}^{n-1} F_i G_i (t_{i+1} - t_i)\right) = E\left(\int_0^\infty u_t v_t dt\right)
\end{aligned}$$

since the second term is zero and in the first term

$$E\left(B_{t_{i+1}}^2 - 2B_{t_{i+1}} B_{t_i} + B_{t_i}^2 |\mathcal{F}_{t_i}\right) = E\left(B_{t_{i+1}}^2 - B_{t_i}^2\right) = t_{i+1} - t_i.$$

Particularly we have $E(I(u)^2) = E\left(\int_0^\infty u_t^2 dt\right)$.
Passing to the limits gives the general result for progressively measurable processes. ■

Next we want to prove some important properties of this new introduced object.

Theorem 2.3.6 (Itô Integrals as Martingales)
The stochastic process $M_t := \int_0^t u_s dB_s$ *is for any* $u \in L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P)$ *a martingale with respect to* $\mathcal{F}_t$*, the natural filtration of the Brownian motion.*

Proof We prove the theorem in a first step for an $u \in \mathcal{E}$ given by

$$u_s := \sum_{i=0}^{n-1} F_i 1_{]t_i,t_{i+1}]}(s),$$

so for the process M we have

$$M_t = \int_0^\infty \left(\sum_{i=0}^{n-1} F_i 1_{]t_i,t_{i+1}]}(s) \right) 1_{[0,t]}(s) dB_s = \sum_{i=0}^{n-1} F_i (B_{t\wedge t_{i+1}} - B_{t\wedge t_i}).$$

Since we can always refine the partition, we can assume that t is one partitioning point, i.e there has to exist a $k \leq n$ with $t_k = s$. Then for any $t \geq s$

$$\begin{aligned} E(M_t|\mathcal{F}_s) &= E\left(\sum_{i=0}^{n-1} F_i (B_{t\wedge t_{i+1}} - B_{t\wedge t_i}) | \mathcal{F}_s \right) \\ &= \sum_{i=0, t_{i+1}\leq t}^{n-1} E\left(F_i (B_{t\wedge t_{i+1}} - B_{t\wedge t_i}) | \mathcal{F}_s\right) \\ &\quad + \sum_{i=0, t_{i+1}> t}^{n-1} E\left(F_i (B_{t\wedge t_{i+1}} - B_{t\wedge t_i}) | \mathcal{F}_s\right) \\ &= \sum_{i=0}^{k-1} F_i (B_{t\wedge t_{i+1}} - B_{t\wedge t_i}) = M_s \end{aligned}$$

where the second term vanishes and in the first term we add only up to $k-1$ since $t_k = s$.
The extension to square integrable progressively measurable processes follows simply by taking L^2-limits. ■

This theorem has a quite important corollary which asserts the existence of a continuous modification of a stochastic integral. So when we speak about stochastic integrals in the following, we can always think of them as processes with continuous paths.

Corollary 2.3.7 (Continuous Paths of the Itô Integral)
The process $M_t := \int_0^t u_s dB_s$ *has a modification with continuous paths.*

Proof For $u \in \mathcal{E}$ this is clear since $M_t = \sum_{i=0}^{n-1} F_i (B_{t\wedge t_{i+1}} - B_{t\wedge t_i})$ is continuous by the continuity of Brownian motion.
Otherwise we take a converging Cauchy sequence $u_n \in \mathcal{E}$ converging to u and denote the associated Martingales by M^n. By Doob's maximal inequality (Theorem 1.1.20) for $p = 2$ it follows for the process $M^n - M^m$ that

$$\begin{aligned} P\left(\sup_{t\leq T} |M_t^n - M_t^m| \geq \varepsilon \right) &\leq \frac{1}{\varepsilon^2} E\left(|M_t^n - M_t^m|^2\right) \\ &= \frac{1}{\varepsilon^2} E\left(\int_0^T |u_n(s) - u_m(s)|^2 ds \right) \end{aligned}$$

converging to zero for $n, m \to \infty$.
This allows us to find a subsequence n_k, $k \geq 0$, such that

$$P\left(|\sup_{t\leq T}\left(M_t^{n_k} - M_t^{n_{k+1}}\right)| \geq \frac{1}{2^k}\right) \leq \frac{1}{2^k}.$$

Now we can conclude by a Borel-Cantelli argument that

$$P\left(\left\{\omega : \sup_{t\leq T}\left(M_t^{n_k} - M_t^{n_{k+1}}\right)(\omega) \geq \frac{1}{2^k}\right\}\right) = 0$$

which implies that $\breve{M}_t$ defined as uniform limit of the continuous processes M^n is continuous and a modification of M_t. ■

All we learned till now about stochastic integration was more or less general abstract nonsense or - in the case of predictable step processes - material for a numerical approach. But how we can *calculate* with Itô integrals? The next theorem, the celebrated Itô formula, is the major tool for all analytical calculations, comparable to the fundamental theorem of calculus.

Theorem 2.3.8 (Itô Formula)
Given $f \in C_b^2(\mathbb{R})$, a continuous, bounded, real valued function with continuous bounded derivatives up to order 2, and a stochastic Process

$$X_t := X_0 + \int_0^t u_s dB_s + \int_0^t v_s ds$$

for $u, v \in L^2\left(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P\right)$, then it holds that

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) u_s dB_s + \int_0^t f'(X_s) v_s ds + \frac{1}{2}\int_0^t f''(X_s) u_s^2 ds.$$

Here it is the last term which looks strange from the viewpoint of classical analysis, it distinguishes the Itô formula for stochastic integrals from the fundamental theorem of calculus. In comparison: Setting $u_s \equiv 0$, $v_s \equiv 1$ and $X_0 = 0$ we get the fundamental theorem, hence the Itô integral can be seen as a generalization of the classical notion of integral. The origins of the additional term can be seen in the proof:

Proof First we prove the theorem under the assumption that $f \in C_b^\infty(\mathbb{R})$ and $u, v \in \mathcal{E}$. For the sake of simplicity we will use the notations $\Delta_i t := t_{i+1} - t_i$, $\Delta_i B := B_{t_{i+1}} - B_{t_i}$ and $\Delta_i X := X_{t_{i+1}} - X_{t_i}$.
We expand

$$f(X_t) = F(X_0) - \sum_{i=0}^{n-1} f(\Delta_i X)$$

by Taylor's formula up to order 2

$$\begin{aligned}
F(X_t) &= f(X_0) \\
&\quad + \sum_{i=0}^{n-1} f'(X_{t_i})\Delta_i X & (2.6) \\
&\quad + \frac{1}{2}\sum_{i=0}^{n-1} f''(X_{t_i})(\Delta_i X)^2 & (2.7) \\
&\quad + \frac{1}{2}\sum_{i=0}^{n-1} \int_0^1 f'''\left(X_{t_i} + s(\Delta_i X)(1-s)^2\right) ds (\Delta_i X)^3 & (2.8)
\end{aligned}$$

and now check the convergence term by term.
Since f' is bounded we have for (2.6), $n \to \infty$

$$\begin{aligned}
&\sum_{i=0}^{n-1} f'(X_{t_i})\Delta_i X \\
= &\sum_{i=0}^{n-1} f'(X_{t_i}) \left(\int_0^{t_{i+1}} u_s dB_s - \int_0^{t_i} u_s dB_s + \int_0^{t_{i+1}} v_s ds - \int_0^{t_i} v_s ds \right) \\
= &\sum_{i=0}^{n-1} f'(X_{t_i}) \left(u_{t_i}\Delta_i B + v_{t_i}\Delta_i t \right) \longrightarrow \int_0^t f'(X_s) u_s dB_s + \int_0^t f'(X_s) v_s ds
\end{aligned}$$

in L^2 along the refining sequence by definition of the respective integrals.
The term (2.7) we expand to

$$\begin{aligned}
\frac{1}{2}\sum_{i=0}^{n-1} f''(X_{t_i})(\Delta_i X)^2 &= \frac{1}{2}\sum_{i=0}^{n-1} f''(X_{t_i}) u_{t_i}^2 (\Delta_i B)^2 \\
&\quad + \sum_{i=0}^{n-1} f''(X_{t_i}) u_{t_i} v_{t_i} (\Delta_i t)(\Delta_i B) \\
&\quad + \frac{1}{2}\sum_{i=0}^{n-1} f''(X_{t_i}) v_{t_i}^2 (\Delta_i t)^2. & (2.9)
\end{aligned}$$

Since the mesh $|\sup_i \Delta_i t|$ of the refining sequence tends to zero, so does $|\Delta_i B|$ by continuity. Hence the last term converges in L^2 to zero since v_t and f'' are bounded and we can estimate for a constant M:

$$E\left(\left(\frac{1}{2}\sum_{i=0}^{n-1} f''(X_{t_i}) v_{t_i}^2 (\Delta_i t)^2 - 0 \right)^2 \right) \leq M^2 \sum_{i,j=0}^{n-1} (\Delta_i t)^2 (\Delta_j t)^2 \longrightarrow 0.$$

By the same argument we get for the second term

$$
\begin{aligned}
& E\left(\left(\sum_{i=0}^{n-1} f''(X_{t_i}) u_{t_i} v_{t_i} (\Delta_i t)(\Delta_i B) - 0\right)^2\right) \\
\leq \quad & N^2 \sum_{i=0}^{n-1} (\Delta_i t)^2 E\left((\Delta_i B)^2\right) + 2 \sum_{i,j=0}^{n-1} u_{t_i} v_{t_i} (\Delta_i t) u_{t_j} v_{t_j} (\Delta_j t) E\left((\Delta_i B)(\Delta_j B)\right) \\
\longrightarrow \quad & N^2 \sum_{i=0}^{n-1} (\Delta_i t)^3 \longrightarrow 0
\end{aligned}
$$

where the second term disappears by independence of the increments of the Brownian motion and $\lim_{n\to\infty} \sum_{i=0}^{n-1} E\left((\Delta_i B)^2\right) = t$. Concerning the first term (2.9), we write (abbreviating $\gamma_{t_i} := \frac{1}{2} f''(X_{t_i}) u_{t_i}^2$) it as

$$
\sum_{i=0}^{n-1} \gamma_{t_i} (\Delta_i B)^2 = \sum_{i=0}^{n-1} \gamma_{t_i} \left((\Delta_i B)^2 - (\Delta_i t)\right) + \sum_{i=0}^{n-1} \gamma_{t_i} (\Delta_i t) \tag{2.10}
$$

Here the first term converges to zero since

$$
\begin{aligned}
& E\left(\left(\sum_{i=0}^{n-1} \gamma_{t_i} \left((\Delta_i B)^2 - (\Delta_i t)\right) - 0\right)^2\right) \\
= \quad & E\left(\sum_{i=0}^{n-1} \gamma_{t_i} \left((\Delta_i B)^2 - (\Delta_i t)\right)^2\right) \\
& + 2E\left(\sum_{i=0}^{n-1} \gamma_{t_i} \gamma_{t_j} \left((\Delta_i B)^2 - (\Delta_i t)\right) \left((\Delta_j B)^2 - (\Delta_j t)\right)\right) \\
\leq \quad & K \sum_{i=0}^{n-1} (\Delta_i t) + 2L \sum_{i=0}^{n-1} (\Delta_i t)^2 \longrightarrow 0
\end{aligned}
$$

by analogous considerations.
The second term of (2.10) converges obviously

$$
\sum_{i=0}^{n-1} \gamma_{t_i} (\Delta_i t) = \sum_{i=0}^{n-1} \frac{1}{2} f''(X_{t_i}) u_{t_i}^2 (\Delta_i t) \longrightarrow \frac{1}{2} \int_0^t f''(X_s) u_s^2 ds.
$$

The remainder (2.8) converges to zero too, since in all the cases we have $(\Delta_i t)^{k_1} (\Delta_i B)^{k_2}$ with $k_1 + k_2 = 3$ as differences which asserts, by $\sum_{i=0}^{n-1} E\left((\Delta_i B)^2\right) \to t$ and mesh $|\sup_i \Delta_i t|$ tending to zero, the vanishing of each term in the series, whence of the whole remainder.
Having now checked all the terms it remains as not necessarily zero

$$
\int_0^t f'(X_s) u_s dB_s + \int_0^t f'(X_s) v_s ds + \frac{1}{2} \int_0^t f''(X_s) u_s^2 ds
$$

as desired.
We yet have to generalize the result by lifting the restrictions: Given $u, v \in L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P)$ we can choose by Theorem (2.3.3) sequences $u^m, v^m \in \mathcal{E}$ converging a.s. to $(u_s 1_{[0,t]})_{s\geq 0}$ resp. $(v_s 1_{[0,t]})_{s\geq 0}$. Since u^m, v^m, f' and f'' are bounded we can conclude by Lebesgue's dominated convergence theorem and the following a.s. convergences

$$\begin{aligned}
\left(f'(X_s^m)u_s^m 1_{[0,t]}(s)\right)_{s\geq 0} &\longrightarrow \left(f'(X_s)u_s 1_{[0,t]}(s)\right)_{s\geq 0}\\
\left(f'(X_s^m)v_s^m 1_{[0,t]}(s)\right)_{s\geq 0} &\longrightarrow \left(f'(X_s)v_s 1_{[0,t]}(s)\right)_{s\geq 0}\\
\left(f'(X_s^m)(u_s^m)^2 1_{[0,t]}(s)\right)_{s\geq 0} &\longrightarrow \left(f'(X_s)(u_s)^2 1_{[0,t]}(s)\right)_{s\geq 0}
\end{aligned}$$

that all limits exist. Dropping the smoothness restriction does not pose a problem since we can approximate $f \in C_b^\infty(\mathbb{R})$ uniformly by a sequence $f^m \in C_b^2(\mathbb{R})$. ■

The following notation will be of great use: Let $u, v \in L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P)$, instead of

$$X_t := X_0 + \int_0^t u_s dB_s + \int_0^t v_s ds$$

we write as shorthand the infinitesimal expression

$$dX_t = u_t dB_t + v_t ds$$

where we have to introduce the following calculation rules: $dB_t \cdot dB_t = dt$, $dB_t \cdot dt = dt \cdot dB_t = 0$ and $dt \cdot dt = 0$. For instance the Itô formula is written in this notation as

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2.$$

It is important to notice that this is only a notation to abbreviate the somehow lengthly integral notation and has nothing to do with derivatives which were not yet defined.

2.4 Generalizing the Itô Integral: Integration along Continuous Semimartingales

It is our aim in this section to generalize the Itô calculus developed in the previous one. So first we have to ask which of the many salient features of Brownian motion was the reason that we could integrate along it's paths. Going some pages back we will recognize that it was the martingale properties which enabled us to do so, or, more precisely, the fact that Brownian motion is a continuous local martingale. There even exists a theory of stochastic integration along general semimartingales but we will restrict ourselves to continuous semimartingales as most general class of integrators.

2.4.1 Martingales and Quadratic Variation

First we have to look at the quadratic variation again:

Proposition 2.4.1 (Finite Variation Martingales)
A continuous martingale M_t does not have finite total variation unless it is constant.

Proof The proof is quite similar to the proof of Theorem 2.3.1. Without loss of generality we may assume $M_0 = 0$. We define the stopping time $\sigma_n := \inf\{s : S_s \geq 0\}$ for the total variation $S_T(M)$ on $[0, T]$ entailing that the stopped martingale M^{σ_n} is of bounded total variation which allows us to restrict ourselves to martingales with bounded variation $S_T(M) < K$.
We take a refining sequence of partitions with mesh $|\Delta_n| \to 0$. For a concrete partition we have by the martingale property

$$\begin{aligned} E(M_t^2) &= E\left(\sum_{i=0}^{n-1}(M_{t_{i+1}}^2 - M_{t_i}^2)\right) = E\left(\sum_{i=0}^{n-1}(M_{t_{i+1}} - M_{t_i})^2\right) \\ &\leq E\left(\left(\sup_i |M_{t_{i+1}} - M_{t_i}|\right) S_T(M)\right) \leq KE(|\Delta_n|) \to 0 \end{aligned}$$

for $|\Delta_n| \to 0$. It follows that $M_0 = 0$ a.s. which yields the result. ∎

Theorem 2.4.2 (Quadratic Variation Process)
Given a continuous bounded martingale M, then the quadratic variation process $\langle M, M\rangle_s$ is the unique continuous increasing adapted process which vanishes at zero such that

$$M_t^2 - \langle M, M\rangle_t$$

is a martingale. In particular this implies that continuous bounded martingales are of finite quadratic variation.

Proof
(i) Uniqueness: Assume there would be two different such processes A, B. Then $A - B$ is a continuous martingale with finite total variation, whence by Proposition 2.4.1 $A - B = 0$ a.s. which implies uniqueness.
(ii) Existence: The proof of the existence of such a process is not so easy: For a given t we observe the partition Δ with $0 = t_0 < t_1 < ... < t_n = t$. For any $s < t$ there exists an $i \in \{0, ..., n-1\}$ such that $t_i \leq s < t_{i+1}$, implying

$$E\left((M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_s\right) = E\left((M_{t_{i+1}} - M_s)^2|\mathcal{F}_s\right) + (M_s - M_{t_i})^2.$$

Writing now T_s^Δ for the quadratic sum over the partition $0 = t_0 < t_1 < ... < t_i \leq s$ we get

$$\begin{aligned} & E\left((T_t^\Delta - T_s^\Delta|\mathcal{F}_s\right) \\ = \ & E\left(\sum_{j=i+1}^{n-1}(M_{j+1} - M_j)^2 + (M_{t_{i+1}} - M_{t_i})^2 - (M_s - M_{t_i})^2|\mathcal{F}_s\right) \\ = \ & E\left((M_t - M_s)^2|\mathcal{F}_s\right) = E\left(M_t^2 - M_s^2|\mathcal{F}_s\right) \end{aligned} \tag{2.11}$$

(ii)(a) We have to prove that for a sequence of partitions $\{\Delta_n\}$ with mesh $|\Delta_n|$ the sequence $\{T_t^{\Delta_n}\}$ converges in L^2 (independently of the choice of the partition). So we have to show that the difference $X_t^n(M) := T_t^{\Delta_n} - T_t^{\Delta'_n}$ for $\{\Delta_n\}$, $\{\Delta'_n\}$ two different sequences of partitions converges in L^2 to zero or, in other words, that $E\left((X_t^n(M) - 0)^2\right) = E\left((X_t^n(M))^2\right) \to 0$ for $n \to \infty$. We note that $X(M)$ is a martingale since by (2.11)

$$\begin{aligned} & E(X_t(M)|\mathcal{F}_s) = E\left(T_t^{\Delta}(M) - T_t^{\Delta'}(M)|\mathcal{F}_s\right) \\ = \; & E\left(\left(T_t^{\Delta}(M) - T_s^{\Delta}(M)\right) - \left(T_t^{\Delta'}(M) - T_s^{\Delta'}(M)\right) + \left(T_s^{\Delta}(M) - T_s^{\Delta'}(M)\right)|\mathcal{F}_s\right) \\ = \; & E\left(M_t^2 - M_s^2|\mathcal{F}_s\right) - E\left(M_t^2 - M_s^2|\mathcal{F}_s\right) + E\left(T_s^{\Delta}(M) - T_s^{\Delta'}(M)|\mathcal{F}_s\right) \\ = \; & X_s(M) \end{aligned} \tag{2.12}$$

Writing $\Delta\Delta'$ for the partition generated by taking all partitioning points of Δ and Δ' we get by considerations analogously to (2.11) for X_t

$$\begin{aligned} E\left(X_t^2(M)\right) &= E\left(\left(T_t^{\Delta}(M) - T_t^{\Delta'}(M)\right)^2\right) = E\left(T_t^{\Delta\Delta'}(X(M))\right) \\ &\leq E\left(2\left(T_t^{\Delta\Delta'}(T_t^{\Delta}(M)) + T_t^{\Delta\Delta'}(T_t^{\Delta'}(M))\right)\right) \end{aligned} \tag{2.13}$$

since

$$\begin{aligned} & T_t^{\Delta\Delta'}\left(T_t^{\Delta}(M) - T_t^{\Delta'}(M)\right) \\ = \; & \sum_{k=0}^{n-1}\left(\left(T_{k+1}^{\Delta}(M) - T_{k+1}^{\Delta'}(M)\right) - \left(T_k^{\Delta}(M) - T_k^{\Delta'}(M)\right)\right)^2 \\ = \; & \sum_{k=0}^{n-1}\left(\left(T_{k+1}^{\Delta}(M) - T_k^{\Delta}(M)\right) - \left(T_{k+1}^{\Delta'}(M) - T_k^{\Delta'}(M)\right)\right)^2 \\ \leq \; & 2\left(\sum_{k=0}^{n-1}\left(T_{k+1}^{\Delta}(M) - T_k^{\Delta}(M)\right)^2 - \left(T_{k+1}^{\Delta'}(M) - T_k^{\Delta'}(M)\right)^2\right) \\ \leq \; & 2\left(T_t^{\Delta\Delta'}(T_t^{\Delta}(M)) + T_t^{\Delta\Delta'}(T_t^{\Delta'}(M))\right) \end{aligned}$$

So it remains to show that $E\left(T_t^{\Delta\Delta'}(T_t^{\Delta}(M))\right)$ (resp. $T_t^{\Delta'}(M)$) converges to zero for $|\Delta_n| + |\Delta'_n| \to 0$.

(ii)(b) For a partitioning point s_k in $\Delta\Delta'$ we denote by t_l the rightmost point of Δ such that $t_l \leq s_k < s_{k+1} \leq t_{l+1}$. Since we have

$$\begin{aligned} T_{s_{k+1}}^{\Delta}(M) - T_{s_k}t^{\Delta}(M) &= \left(M_{s_{k+1}} - M_{t_l}\right)^2 - \left(M_{s_k} - M_{t_l}\right)^2 \\ &= M_{s_{k+1}}^2 - M_{s_k}^2 - 2M_{t_l}\left(M_{s_{k+1}} - M_{s_k}\right) \\ &= \left(M_{s_{k+1}} - M_{s_k}\right)\left(M_{s_{k+1}} + M_{s_k} - 2M_{t_l}\right) \end{aligned}$$

we can conclude that

$$\begin{aligned} T_t^{\Delta\Delta'}(T_t^{\Delta}(M)) &= \sum_{k=0}^{n-1}\left(T_{s_{k+1}}^{\Delta}(M)-T_{s_k}^{\Delta}(M)\right)^2 \\ &\leq \left(\sup_k |M_{s_{k+1}}+M_{s_k}-2M_{t_l}|^2\right) T_t^{\Delta\Delta'}(M). \end{aligned}$$

Applying Cauchy's inequality to this equation we get

$$E\left(T_t^{\Delta\Delta'}(T_t^{\Delta}(M))\right) \leq \left(E\left(\sup_k |M_{s_{k+1}}+M_{s_k}-2M_{t_l}|^4\right)\right)^{\frac{1}{2}}\left(E\left(\left(T_t^{\Delta\Delta'}(M)\right)^2\right)\right)^{\frac{1}{2}}$$

where the first factor tends - by continuity of M - to zero for $|\Delta_n|+|\Delta'_n|\to 0$ and it remains only to show that the second one is bounded.

(ii)(c) Since M is bounded there exists a constant C such that $|M|\leq C$ and hence by (4.11)

$$E\left(T_t^{\Delta}(M)\right)=E\left(E\left(T_t^{\Delta}(M)-T_0^{\Delta}(M)|\mathcal{F}_0\right)\right)=E\left(E\left(M_t^2-M_0^2|\mathcal{F}_0\right)\right)\leq 2C^2.$$

For the square we note that we can write $\left(T_t^{\Delta}(M)\right)^2$ as

$$\begin{aligned} \left(T_t^{\Delta}(M)\right)^2 &= \left(\sum_{k=0}^{n-1}\left(M_{t_{k+1}}-M_{t_k}\right)^2\right)^2 \\ &= 2\sum_{0\leq k<j}^{n-1}\left(M_{t_{k+1}}-M_{t_k}\right)^2\left(M_{t_{j+1}}-M_{t_j}\right)^2+\sum_{k=0}^{n-1}\left(M_{t_{k+1}}-M_{t_k}\right)^4 \\ &= 2\sum_{0=k}^{n-1}\left(M_{t_{k+1}}-M_{t_k}\right)^2\left(T_t^{\Delta}(M)-T_{t_{k+1}}^{\Delta}(M)\right)+\sum_{k=0}^{n-1}\left(M_{t_{k+1}}-M_{t_k}\right)^4 \\ &= 2\sum_{0=k}^{n-1}\left(T_{t_{k+1}}^{\Delta}(M)-T_{t_k}^{\Delta}(M)\right)^2\left(T_t^{\Delta}(M)-T_{t_{k+1}}^{\Delta}(M)\right)+\sum_{k=0}^{n-1}\left(M_{t_{k+1}}-M_{t_k}\right)^4 \end{aligned}$$

and since by (4.11)

$$E\left(T_t^{\Delta}(M)-T_{t_{k+1}}^{\Delta}(M)|\mathcal{F}_{t_{k+1}}\right)=E\left(\left(M_t-M_{t_{k+1}}\right)^2|\mathcal{F}_{t_{k+1}}\right)$$

we get

$$\begin{aligned} &E\left(\left(T_t^{\Delta}(M)\right)^2\right) \\ &= 2\sum_{0=k}^{n-1}E\left(\left(T_{t_{k+1}}^{\Delta}(M)-T_{t_k}^{\Delta}(M)\right)\left(M_t^2-M_{t_{k+1}}^2\right)\right)+\sum_{k=0}^{n-1}E\left(\left(M_{t_{k+1}}-M_{t_k}\right)^4\right) \\ &\leq 2E\left(\left(\sup_k |M_t^2-M_{t_{k+1}}^2|^2\right)T_t^{\Delta}(M)\right)+E\left(\left(\sup_k |M_{t_{k+1}}^2-M_{t_k}^2|^2\right)T_t^{\Delta}(M)\right) \\ &\leq 2\left((2C)^2 2C^2\right)+(2C)^2 2C^2=24C^4. \end{aligned}$$

So, going backward, $E\left(\left(T_t^{\Delta}(M)\right)^2\right)$ is bounded for any partition, so particularly $E\left(\left(T_t^{\Delta\Delta'}(M)\right)^2\right)$, implying that $E\left(T_t^{\Delta\Delta'_n}\left(T_t^{\Delta_n}(M)\right)\right)$ tends to zero for $|\Delta_n| + |\Delta'_n| \to 0$ and so $E\left((X_t^n(M))^2\right) \to 0$ for $n \to \infty$ as desired.

(iii) Properties: We have now established the unique existence of the limit $\langle M, M\rangle_t$ in L^2, it remains to show that it has a modification with the required properties (i.e. to be a continuous increasing adapted process which vanishes at zero such that $M_t^2 - \langle M, M\rangle_t$ is a martingale).
Adaptedness and vanishing at zero is obvious for the limit process, so it is the martingale property since by passing to the limits in (4.11) we have

$$\begin{aligned} & E\left(M_t^2 - \langle M, M\rangle_t |\mathcal{F}_s\right) \\ = \;& E\left(\left(M_t^2 - M_s^2\right) - \left(\langle M, M\rangle_t - \langle M, M\rangle_s\right) + \left(M_s^2 - \langle M, M\rangle_s\right)|\mathcal{F}_s\right) \\ = \;& M_s^2 - \langle M, M\rangle_s . \end{aligned}$$

Since $T_t^{\Delta_n}(M) - T_t^{\Delta_m}(M)$ is a martingale by (2.12) (and it is as a step function right continuous) we have by Bob's maximal inequality for $p = 2$

$$\begin{aligned} E\left(\sup_{s\le t} |T_s^{\Delta_n}(M) - T_s^{\Delta_m}(M)|^2\right) &\le 4 \sup_{s\le t} E\left(\left(T_s^{\Delta_n}(M) - T_s^{\Delta_m}(M)\right)^2\right) \\ &= 4E\left(\left(T_t^{\Delta_n}(M) - T_t^{\Delta_m}(M)\right)^2\right) \end{aligned}$$

one can choose a subsequence $\{\Delta_{nk}\}$ such $\{T_s^{\Delta_{nk}}\}$ converges a.s. uniformly on $[0, t]$ hence $\langle M, M\rangle_s$ is continuous.
It is clear that $\langle M, M\rangle_t$ is increasing by choosing a refining sequence dense in $[0, t]$. So we have for any $r < s \in \bigcup_n \Delta_n$ the property $T_r^{\Delta_n}(M) \le T_s^{\Delta_n}(M)$ for all $n > n_0$, the smallest n such that $r, s \in \Delta_n$. The rest follows by continuity.

■

Corollary 2.4.3 (Stopped Quadratic Variation)
Given an arbitrary stopping time τ, then

$$\langle M^\tau, M^\tau\rangle = \langle M, M\rangle^\tau$$

Proof On one hand side, since $M^2 - \langle M, M\rangle$ is a martingale,

$$\left(M^2 - \langle M, M\rangle\right)^\tau = \left(M^2\right)^\tau - \langle M, M\rangle^\tau = (M^\tau)^2 - \langle M, M\rangle^\tau ,$$

but on the other hand side the previous theorem states that $\langle M^\tau, M^\tau\rangle$ is the unique increasing adapted process such that $(M^\tau)^2 - \langle M^\tau, M^\tau\rangle$ is a martingale. So they have to be equal,

$$(M^\tau)^2 - \langle M, M\rangle^\tau = (M^\tau)^2 - \langle M^\tau, M^\tau\rangle$$

and hence $\langle M^\tau, M^\tau\rangle = \langle M, M\rangle^\tau$. ■

Our next goal is to generalize these notions - into two quite distinct directions. On the one hand side we will show that the quadratic variation process exists for continuous local martingales too, and on the other hand side we will - by introduction of the bracket process - drop the symmetry.

Theorem 2.4.4 (Bracket Process)
Let M_t, N_t two continuous local martingales, then there exists a unique continuous process of finite total variation vanishing at zero such that $MN - \langle M, N\rangle$ is a continuous local martingale.

This process is called the *bracket process* of M and N or their *quadratic covariation process.*

Proof Assume first that $M = N$, then there exist stopping times $\tau_n \uparrow \infty$ such that M^{τ_n} is a bounded martingale. By Theorem 2.3.2 there exists for each n a unique continuous process $\langle M^{\tau_n}, M^{\tau_n}\rangle$ such that $(M^{\tau_n})^2 - \langle M^{\tau_n}, M^{\tau_n}\rangle$ is a martingale. Since $\langle M^{\tau_n}, M^{\tau_n}\rangle = \langle M^{\tau_{n+k}}, M^{\tau_{n+k}}\rangle^{\tau_n}$ we can define the process $\langle M, M\rangle$ by requiring it to be equal to $\langle M, M\rangle^{\tau_n}$ on $[0, \tau_n]$. Here it is unique (for every n), implying global uniqueness; and since $\left(M^2 - \langle M, M\rangle\right)^{\tau_n} = (M^{\tau_n})^2 - \langle M^{\tau_n}, M^{\tau_n}\rangle$ is a martingale this is the sought-after process.
Asymmetry we get by polarization, setting

$$\langle M, N\rangle := \frac{1}{4}\left(\langle M+N, M+N\rangle - \langle M-N, M-N\rangle\right).$$

This is a continuous process vanishing at zero and as difference of increasing processes it is of finite total variation. Furthermore it is unique by the argument 2.4.2(i). ∎

For optional stopping we get as simple consequence

Corollary 2.4.5 (Stopped Bracket)
For any stopping time τ

$$\langle M^\tau, N^\tau\rangle = \langle M, N\rangle^\tau = \langle M, N^\tau\rangle$$

Proof The left equality is clear by polarization and the right one by observing that

$$MN^\tau - \langle M, N\rangle^\tau = (M - M^\tau)\, N^\tau + (M^\tau N^\tau - \langle M, N\rangle^\tau)$$

is a local martingale which has to be unique by Theorem 2.4.4. ∎

Also the covariation process can be described as limit of a sum, taking the differences of evaluation points.

Proposition 2.4.6 (Covariance as Limit)
For any sequence $\{\Delta_n\}$ of partitions with mesh $|\Delta_n| \to 0$ we can write the covariation process as limit of sums $\hat{T}_t^\Delta(M) := \sum_{i=0}^{n-1}\left(M_{t_{i+1}} - M_{t_i}\right)\left(N_{t_{i+1}} - N_{t_i}\right)$:

$$\lim_{n\to\infty}\sup_{s\le t}|\hat{T}_s^{\Delta_n}(M) - \langle M, N\rangle_s| = 0$$

Proof We only show the proof for $M = N$, the rest is clear by polarization. For $\delta, \varepsilon > 0$ we can found a suitable stopping time such that M^σ is a bounded martingale and $P(\{\sigma \leq t\}) \leq \delta$. On $[0,\sigma]$ we have $\hat{T}_t^\Delta(M) = \hat{T}_t^\Delta(M^\sigma)$ and $\langle M, M\rangle = \langle M^\sigma, M^\sigma\rangle$, so

$$\begin{aligned} & P\left(\left\{\omega : \sup_{s\leq t} |\hat{T}_s^\Delta(M) - \langle M, M\rangle_s| > \varepsilon\right\}\right) \\ \leq \quad & \delta + P\left(\left\{\omega : \sup_{s\leq t} |\hat{T}_s^\Delta(M^\sigma) - \langle M^\sigma, M^\sigma\rangle_s| > \varepsilon\right\}\right) \end{aligned}$$

tending to zero for $|\Delta_n| \to 0$ and a sequence $\delta_n \downarrow 0$ ■

Proposition 2.4.7 (Constant Martingales)
The quadratic variation vanishes only iff (for every t) $M_t = M_0$ *a.s.*

Proof It is obvious, that for constant M the quadratic variation vanishes, the other direction of the proposition we prove for bounded M by setting in (2.11) $s = 0$ and passing to the limit

$$0 = E\left(\langle M, M\rangle_t - \langle M, M\rangle_0 |\mathcal{F}_0\right) = E\left((M_t - M_0)^2|\mathcal{F}_0\right),$$

whence $M_t = M_0$ a.s. By optional stopping and Corollary 2.4.5 this extends to continuous local martingales. ■

The class of integrators of our generalized stochastic integrals will be that of continuous semimartingales defined as follows:

Definition 2.4.8 (Continuous Semimartingales)
A stochastic process is called a continuous $(\mathcal{F}_t)$*-semimartingale, iff it can be decomposed* $X_t = M_t + A_t$ *where* M_t *is a continuous local* $(\mathcal{F}_t)$*-martingale and* A_t *is a continuous* $(\mathcal{F}_t)$*-adapted process of locally finite total variation with* $A_0 = 0$.

We determine its variation properties:

Proposition 2.4.9 (Semimartingale Properties)
Given a continuous semimartingale $X_t = M_t + A_t$*, then*
(i) it is of finite quadratic variation and $\langle X, X\rangle_t = \langle M, M\rangle_t$ *a.s.*
(ii) the decomposition $X_t = M_t + A_t$ *is a.s. unique.*

Proof (i) We take a refining sequence of partitions Δ_n of $[0,T]$ with mesh $|\Delta_n|$ tending to zero. For a concrete partition we have

$$\begin{aligned} \sum_{i=0}^{n-1}\left(X_{t_{i+1}} - X_{t_i}\right)^2 &= \sum_{i=0}^{n-1}\left(M_{t_{i+1}} - M_{t_i}\right)^2 + 2\sum_{i=0}^{n-1}\left(M_{t_{i+1}} - M_{t_i}\right)\left(A_{t_{i+1}} - A_{t_i}\right) + \\ &+ \sum_{i=0}^{n-1}\left(A_{t_{i+1}} - A_{t_i}\right)^2. \end{aligned}$$

Since by

$$\sum_{i=0}^{n-1}\left(M_{t_{i+1}}-M_{t_i}\right)\left(A_{t_{i+1}}-A_{t_i}\right)\le \sup_i |M_{t_{i+1}}-M_{t_i}|S_T(A)$$

and

$$\sum_{i=0}^{n-1}\left(A_{t_{i+1}}-A_{t_i}\right)^2=\sup_i |A_{t_{i+1}}-A_{t_i}|S_T(A)$$

we get for $|\Delta_n|\to 0$

$$\langle X,X\rangle_T=\langle M,M\rangle_T\, a.s.$$

for arbitrary T.

(ii) Given two different decompositions $X_t=M_t+A_t=N_t+B_t$, then the local martingale $M_t-N_t=B_t-A_t$ is of finite total variation a.s. and $M_0-N_0=A_0-B_0=0$, hence by Proposition 2.4.1 $M_t-N_t=0$ a.s.

■

Analogous to the local martingales we can define a bracket process $\langle X,Y\rangle_t$ by polarization which equals $\langle M,N\rangle_t$ by the above proposition.

Before developing integration theory we have yet to write something about the connection between martingale and integrability properties.

Definition 2.4.10 (L^2 Martingales)
We denote by $\mathbb{H}^2$ the space of L^2-bounded $(\mathcal{F}_t)$-martingales, i.e. martingales M such that

$$\sup_t E(M_t^2)<\infty$$

The subspace of L^2-bounded continuous martingales will be denoted by H^2 and H_0^2 will be used as abbreviation for the subspace of H^2 which consists of martingales M_t with $M_0=0$.

Proposition 2.4.11 ($\mathbb{H}^2$ as Hilbert Space)
Defining a norm by

$$\|M\|_{\mathbb{H}^2}:=\left(E(M_\infty^2)\right)^{\frac{1}{2}}=\lim_{t\to\infty}\left(E(M_t^2)\right)^{\frac{1}{2}},$$

the space $\mathbb{H}^2$ is a Hilbert space with respect to this norm and H^2 and H_0^2 are sub(-Hilbert-)spaces.

Proof The inner product can be obtained by polarization and Doob's inequality (Theorem 1.1.20) asserts that every Cauchy sequence in $\mathbb{H}^2$ converges toward an element of $\mathbb{H}^2$.
For the second statement we have to prove that the limit of H^2-martingales remains continuous which can be done by Doob's inequality too: Given a sequence $\{M_t^n\}_n\in H^2$ converging to $M_t\in\mathbb{H}^2$, then by Doob's maximal inequality

$$\|\left(M_t^n-M_t\right)^*\|_2\le\sup_t\|M_t^n-M_t\|_2\le 2\|M_t^n-M_t\|_{\mathbb{H}^2}\to 0$$

for $n \to \infty$ implies that there exists a subsequence $M_t^{n_k}$ such that

$$\sup_t |M_t^{n_k} - M_t| \to 0 \quad a.s.$$

asserting the continuity of M_t. The same argument is true for continuous L^2-bounded martingales with $M_0 = 0$. ■

The next proposition states that exactly those continuous local martingales are in H^2 which have bounded quadratic variation.

Proposition 2.4.12 (Bounded Quadratic Variation)
For continuous local martingales the following two conditions are equivalent:
(i) $M \in H^2$.
(ii) $E(\langle M,M\rangle_\infty) < \infty$ *and* $M_0 \in L^2$.

Proof
(i) ⇒ (ii) The square integrability of M_0 is obvious. To prove the boundedness of $E(\langle M,M\rangle_\infty)$ we derive from Theorem 2.4.2 the equation

$$E\left(N_t^2 - \langle N,N\rangle_t\right) = E\left(E\left(N_t^2 - \langle N,N\rangle_t |\mathcal{F}_0\right)\right) = E\left(N_0^2\right) \tag{2.14}$$

for continuous bounded martingales N. For a given localizing sequence $\tau_n \uparrow \infty$ of M we set $\sigma_n := \inf\{t : |M_t| \leq n\} \wedge \tau_n$ to get another localizing sequence such that all M^{σ_n} are martingales, so by equation (2.14) it holds that

$$E\left((M^{\sigma_n})^2\right) - E\left(\langle M,M\rangle^{\sigma_n}\right) = E\left(M_0^2\right). \tag{2.15}$$

Since $M \in H^2$ we have by definition $\sup_t E(M_t^2) < \infty$ which implies that we can pass to the limit and get

$$E\left(M_\infty^2\right) - E(\langle M,M\rangle_\infty) = E\left(M_0^2\right), \tag{2.16}$$

whence $E(\langle M,M\rangle_\infty) < \infty$.
(ii) ⇒ (i) In the other direction (2.15) yields

$$E\left((M^{\sigma_n})^2\right) \leq E\left(\langle M,M\rangle_\infty\right) + E\left(M_0^2\right) \leq K < \infty$$

for a constant K. Therefore Fatou's Lemma implies

$$E\left(M_t^2\right) = E\left(\lim_{n\to\infty} (M^{\sigma_n})^2\right) \leq \liminf_n E\left((M^{\sigma_n})^2\right) < K$$

so M_t is a L^2-bounded, continuous local martingale. ■

Corollary 2.4.13 (Bounded Quadratic Variation)
For a continuous local martingale M the following two conditions are equivalent:
(i) $\{M_s : s \leq t\}$ *is a* L^2*-bounded martingale.*
(ii) $M_0 \in L^2$ *and* $E(\langle M,M\rangle_t) < \infty$.

Proof Analogous to the previous one. ■

Corollary 2.4.14 (Norm Equality)
For $M \in H_0^2$ we have $\|M\|_{\mathbb{H}^2} = \left\| \langle M, M \rangle_\infty^{\frac{1}{2}} \right\|_2$.

Proof By definition of the $\mathbb{H}^2$-norm and equation (2.16) we have, since $M_0 = 0$,

$$\|M\|_{\mathbb{H}^2} = \left(E\left(M_\infty^2\right)\right)^{\frac{1}{2}} = \left(E\left(\langle M, M \rangle_\infty\right)\right)^{\frac{1}{2}} = \left\| \langle M, M \rangle_\infty^{\frac{1}{2}} \right\|_2$$

. ■

As the last topic of this subsection we will state some estimates for Stieltjes integrals known as the Kunita-Watanabe inequality.

Lemma 2.4.15
Given two continuous local martingales M, N and two bounded measurable processes H, K. Then it holds for $t < \infty$ that a.s.

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

Remark that the integrals here are classical Stieltjes integrals since the quadratic variation process is monotone and hence the bracket - as difference of quadratic variation processes - is of finite total variation.

Proof First we note that it is enough to prove the statement for

$$\begin{aligned} H &= H_1 1_{[t_0,t_1]} + H_2 1_{]t_1,t_2]} + ... + H_n 1_{]t_{n-1},t_n]} \\ K &= K_1 1_{[t_0,t_1]} + K_2 1_{]t_1,t_2]} + ... + K_n 1_{]t_{n-1},t_n]} \end{aligned}$$

for a specific partition Δ of $[0, t]$ and bounded measurable r.v. H_i, K_i, since the so defined processes are dense in the measurable bounded processes. So we get the general statement as limit for a refining sequence $\{\Delta_n\}$ of partitions with mesh tending to zero.
For

$$\langle M, N \rangle_s^t := \langle M, N \rangle_t - \langle M, N \rangle_s$$

we have

$$\left| \langle M, N \rangle_s^t \right| \leq \left(\langle M, M \rangle_s^t \right)^{\frac{1}{2}} \left(\langle N, N \rangle_s^t \right)^{\frac{1}{2}} \quad a.s. \tag{2.17}$$

since for a partition Δ we can require s, t to be partitioning points. So for a refining sequence $\{\Delta_n\}$ the inequality (2.17) reads as classical Cauchy inequality which is preserved for $n \to \infty$. So we estimate for our specific choice of H and

K by (2.17) and Cauchy's inequality

$$\begin{aligned}\left|\int_0^t H_sK_s d\langle M,N\rangle_s\right| &\leq \sum_{i=0}^{n-1}|H_{i+1}|\,|K_{i+1}|\left|\langle M,N\rangle_{t_i}^{t_{i+1}}\right|\\ &\leq \sum_{i=0}^{n-1}|H_{i+1}|\,|K_{i+1}|\left(\langle M,M\rangle_{t_i}^{t_{i+1}}\right)^{\frac{1}{2}}\left(\langle N,N\rangle_{t_i}^{t_{i+1}}\right)^{\frac{1}{2}}\\ &\leq \left(\sum_{i=0}^{n-1}H_{i+1}^2\langle M,M\rangle_{t_i}^{t_{i+1}}\right)^{\frac{1}{2}}\left(\sum_{i=0}^{n-1}K_{i+1}^2\langle N,N\rangle_{t_i}^{t_{i+1}}\right)^{\frac{1}{2}}\\ &\leq \left(\int_0^t H_s^2 d\langle M,M\rangle_s\right)^{\frac{1}{2}}\left(\int_0^t K_s^2 d\langle N,N\rangle_s\right)^{\frac{1}{2}}.\end{aligned}$$

. ■

Theorem 2.4.16
Given two continuous local martingales M, N and two measurable processes H_s, K_s, it holds a.s. that

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

for $t \leq \infty$.

Proof Since

$$\int_0^t |H_s|\,|K_s|\,|d\langle M,N\rangle_s| = \left|\int_0^t (H_s \operatorname{sgn}(H_sK_s)\operatorname{sgn}(d\langle M,N\rangle_s)K_s d\langle M,N\rangle_s)\right|$$

the previous lemma gives us the result for bounded H, K and $t < \infty$. But these two conditions can be dropped since the inequality is preserved under monotone limits (this is clearly true also for Lemma 2.4.15). ■

Corollary 2.4.17 (Kunita-Watanabe Inequalites)
Let M, N two continuous, local martingales and H, K two measurable processes, then for p, q satisfying $\frac{1}{p}+\frac{1}{q}=1$ it holds that

$$E\left(\int_0^\infty |H_s|\,|K_s|\,|d\langle M,N\rangle|_s\right) \leq \left\|\left(\int_0^\infty H_s^2 d\langle M,M\rangle_s\right)^{\frac{1}{2}}\right\|_p\left\|\left(\int_0^\infty K_s^2 d\langle N,N\rangle_s\right)^{\frac{1}{2}}\right\|_q$$

and

$$E\left(\left|\int_0^\infty H_sK_s d\langle M,N\rangle_s\right|\right) \leq \left\|\left(\int_0^\infty H_s^2 d\langle M,M\rangle_s\right)^{\frac{1}{2}}\right\|_p\left\|\left(\int_0^\infty K_s^2 d\langle N,N\rangle_s\right)^{\frac{1}{2}}\right\|_q$$

Proof Clearly the inequalities of the previous Theorem and Lemma 2.4.15 are preserved if we integrate both sides over Ω. Applying Hölder's inequality on the right side yields the result. ∎

2.4.2 Stochastic Integrals with respect to Continuous Semimartingales

Here we will introduce the stochastic integral from a far more abstract point of view. First we have to define the class of processes which we want to integrate:

Definition 2.4.18 (Stochastic Integrable Processes)
For a given continuous square integrable martingale $M \in H^2$ we define the space $L^2(M)$ as the equivalence class (of processes on $\mathbb{R}$ equal up to a set of measure zero) of progressively measurable processes K such that

$$\|K\|_M^2 := E\left(\int_0^\infty K_s^2 d\langle M, M\rangle_s\right) < \infty.$$

This can be seen as the space of square integrable, progressively measurable processes with respect to the measure

$$P_M(A) := E\left(\int_0^\infty 1_A(s,\omega) d\langle M, M\rangle_s(\omega)\right)$$

for a set $A \in \mathcal{B}(\mathbb{R}_{\geq 0}) \times \mathcal{F}$. The space $L^2(M)$ is obviously a Hilbert space with respect to the norm $\|\cdot\|_M$. Now we can introduce the integral:

Theorem 2.4.19
Let $M \in H^2$, then there exists for every $K \in L^2(M)$ a unique element $K \cdot M \in H_0^2$ such that

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

for every $N \in H^2$. Furthermore the mapping $K \mapsto K \cdot M$ is a Hilbert space isometry $L^2(M) \to H_0^2$.

Proof
(i) Uniqueness is proven by assuming the existence of two martingales M', $M'' \in H_0^2$ with $\langle M', N\rangle = \langle M'', N\rangle = K \cdot \langle M, N\rangle$. This implies $\langle M' - M'', N\rangle = 0$ and particularly $\langle M' - M'', M' - M''\rangle = 0$ implying by Proposition 2.4.7 that $M' - M'' = 0$ a.s. and hence, by continuity, $M' = M''$.

(ii) First we prove the existence for M, $N \in H_0^2$: The Kunita-Watanabe inequality gives for $p = q = 2$ and $H_s \equiv 0$

$$\left|E\left(\int_0^\infty K_s d\langle M, N\rangle_s\right)\right| \leq E\left(\left|\int_0^\infty K_s d\langle M, N\rangle_s\right|\right) \leq \|N\|_{\mathbb{H}^2} \|K\|_M$$

for an arbitrary N by the definition of $\|\cdot\|_M$ and Corollary 2.4.14 for $\|\cdot\|_{\mathbb{H}^2}$. So the mapping $N \to E\left((K\cdot\langle M,N\rangle)_\infty\right)$ is continuous and linear implying the existence of an element $K\cdot M$ in H_0^2 such that

$$E\left((K\cdot M)_\infty N_\infty\right) = E\left((K\cdot\langle M,N\rangle)_\infty\right). \tag{2.18}$$

For a stopping time τ we have by properties of the conditional expectation

$$\begin{aligned} E\left((K\cdot M)_\tau N_\tau\right) &= E\left(E\left((K\cdot M)_\infty|\mathcal{F}_\tau\right)N_\tau\right) = E\left((K\cdot M)_\infty N_\tau\right) \\ &= E\left((K\cdot M)_\infty N_\infty^\tau\right) = E\left((K\cdot\langle M,N^\tau\rangle)_\infty\right) \\ &= E\left((K\cdot\langle M,N\rangle^\tau)_\infty\right) = E\left((K\cdot\langle M,N\rangle)_\tau\right). \end{aligned}$$

Since all these processes are in H_0^2 we get

$$(K\cdot M)_\tau N_\tau = (K\cdot\langle M,N\rangle)_\tau \quad a.s.$$

By relaxing the (starting at zero) condition for N the result is preserved since we can write N as N_0+C for $N_0\in H_0^2$ and a constant C (the bracket $\langle X,C\rangle$ is obviously zero). Dropping the restriction for M (so that $M\in H^2$ now) requires that we define $K\cdot M := K\cdot(M-M_0)$ for this case.

(iii) We get the isometry by a norm equation; it holds that

$$\begin{aligned} \|K\cdot M\|_{\mathbb{H}^2}^2 &= E\left((K\cdot M)_\infty (K\cdot M)_\infty\right) = E\left(K\cdot\langle M,(K\cdot M)_\infty\rangle\right) \\ &= E\left(K^2\cdot\langle M,M\rangle_\infty\right) = \|K\|_M^2 \end{aligned}$$

by equation (2.18) and the respective definition of the norms. ∎

Definition 2.4.20 (Stochastic Integral)
This unique martingale $K\cdot M$ is called the stochastic integral (Itô integral) of K with respect to M and is denoted by

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

So the Itô integral is a martingale vanishing at zero. We will prove some important properties:

Proposition 2.4.21 (Chain Rule)
Given $K\in L^2(M)$ and $H\in L^2(K\cdot M)$, then $HK\in L^2(M)$ and it holds that

$$(HK)\cdot M = H\cdot(K\cdot M)$$

Proof $H\in L^2(K\cdot M)$ implies by $\langle K\cdot M, K\cdot M\rangle = K^2\cdot M$ that $HK\in L^2(M)$. By associativity of Stieltjes integrals we have

$$\begin{aligned} \langle (HK)\cdot M, N\rangle &= HK\cdot\langle M,N\rangle = H\cdot(K\cdot\langle M,N\rangle) \\ &= H\langle K\cdot M,N\rangle = \langle H\cdot(K\cdot M),N\rangle \end{aligned}$$

for an arbitrary $N\in H^2$, yielding by the uniqueness of Theorem 2.4.19 that $(HK)\cdot M = H\cdot(K\cdot M)$. ∎

The stochastic integral also has very nice stopping properties:

Proposition 2.4.22 (Optional Stopping)
For a stopping time τ it holds that

$$(K \cdot M)^{\tau} = K \cdot M^{\tau} = \left(K1_{[0,\tau]}\right) \cdot M$$

Proof We have for an arbitrary $N \in H^2$

$$\begin{aligned} &\langle (K \cdot M)^{\tau}, N\rangle = \langle K \cdot M, N^{\tau}\rangle = K \cdot \langle M, N^{\tau}\rangle \\ (1) \quad &= K \cdot \langle M^{\tau}, N\rangle = \langle K \cdot M^{\tau}, N\rangle \\ (2) \quad &= K \cdot \langle M, N\rangle^{\tau} = K1_{[0,\tau]} \cdot \langle M, N\rangle = \left\langle K1_{[0,\tau]} \cdot M, N\right\rangle \end{aligned}$$

which yields the desired result by uniqueness. ∎

It is our aim to expand the notion of the stochastic integral - in a first step to local martingales.

Definition 2.4.23
For a continuous local martingale M we define the space $L^2_{loc}(M)$ as the space of classes of progressively measurable processes K for which there exists a sequence $\tau_n \uparrow \infty$ of stopping times with

$$E\left(\int_0^{\tau_n} K_s^2 d\langle M, M\rangle_s\right) < \infty$$

Proposition 2.4.24
For any $K \in L^2_{loc}(M)$ there exists a unique local martingale $K \cdot M$ with $(K \cdot M)_0 = 0$ such that for any continuous local martingale N

$$\langle K \cdot M, N\rangle = K \cdot \langle M, N\rangle$$

Proof First we note that we can by a localizing procedure analogous to Proposition 1.1.22 choose a sequence of stopping times $\tau_n \uparrow \infty$ such that $M^{\tau_n} \in H^2$ and $K^{\tau_n} \in L^2(M^{\tau_n})$. For every n we can now define a stochastic integral $X^{(n)} := K^{\tau_n} \cdot M^{\tau_n}$ and by Proposition 2.4.22 we have $X^{(n)} \equiv X^{(n+k)}$ on $[0, \tau_n]$ for any $k \geq 0$. Now we define the unique process $K \cdot M$ by requiring equivalence with $X^{(n)}$ on $[0, \tau_n]$. This process vanishes obviously at zero, it is continuous and it is a local martingale since $(K \cdot M)^{\tau_n} = K^{\tau_n} \cdot M^{\tau_n}$ on $[0, \tau_n]$ by Proposition 2.4.22 which yields the result. By the same proposition follows the equation for a sequence of stopping times $\tau_n \uparrow \infty$:

$$\begin{aligned} \langle K \cdot M, N\rangle^{\tau_n} &= \langle (K \cdot M)^{\tau_n}, N^{\tau_n}\rangle \\ &= \left\langle (K1_{[0,\tau_n]}) \cdot M^{\tau_n}, N^{\tau_n}\right\rangle = \left(K1_{[0,\tau_n]}\right) \cdot \langle M, N\rangle^{\tau_n} \end{aligned}$$

∎

Before introducing integrals for continuous semimartingales we yet have to define locally bounded processes:

Definition 2.4.25 (Locally Bounded Process)
A progressively measurable process K is called locally bounded iff there exists a sequence of stopping times $\tau_n \uparrow \infty$ with associated constants $c_n \geq 0$ such that $|K^{\tau_n}| \leq c_n$.

Note that locally bounded processes are on the one hand side in $L^2_{loc}(M)$ for every continuous local martingale M and hence stochastic integrable and on the other hand side classical Stieltjes integrable. But we have not to fear that this class was chosen to narrow: By setting $\tau_n := \inf\{t : |K_t| \geq n\}$ all continuous adapted processes have to be locally bounded.

Definition 2.4.26 (Locally Bounded Processes)
Given a locally bounded process K and a continuous semimartingale $X = M + A$, then we define the stochastic integral $K \cdot X$ as

$$K \cdot X := K \cdot M + K \cdot A$$

where $K \cdot A$ is a Stieltjes integral of K with respect to A and $K \cdot M$ is the stochastic integral as defined in Proposition 2.4.24.

We write $\int_0^t K_s dX_s$ for $(K \cdot X)_t$ and note that iff X is a locally bounded martingale, then also $K \cdot X$, and iff X is a process with finite variation, then $K \cdot X$ is too. This asserts that the decomposition $K \cdot X := K \cdot M + K \cdot A$ is unique.

Corollary 2.4.27 (Properties of the Stochastic Integral)
For locally bounded processes H, K a continuous semimartingale X and an arbitrary stopping time τ it holds that
(i) $(HK) \cdot X = H \cdot (K \cdot X)$
(ii) $(K \cdot X)^\tau = K \cdot X^\tau = \left(K 1_{[0,\tau]}\right) \cdot X$

Proof For continuous local martingales the proofs go through as in Proposition 2.4.21. and 2.4.22, for processes of finite total variation they are trivial and the decomposition of the integral is unique. ■

We note, that iff either $K_t(\omega) = 0$ or $X_t(\omega) = X_a(\omega)$ for a.e. ω and $t \in [a, b]$, then $(K \cdot X)_t(\omega)$ is constant for a.e. ω since the quadratic and total variation are zero on this interval.

2.4.3 Itô's Formula

As for Lebesgue integrals we have a dominated convergence theorem for stochastic integrals too:

Theorem 2.4.28 (Dominated Convergence)
Given a continuous semimartingale X and a sequence of locally bounded processes $\{K^n\}_{n \geq 0}$ converging pointwise to zero. If there exists a locally bounded

process K with $|K^n| \leq K$ for every n, then $(K^n \cdot X)_{n \geq 0}$ converges in probability uniformly to zero on compact intervals, i.e. for any $T \in \mathbb{R}$

$$\lim_{n\to\infty} \sup_{t \leq T} |(K^n \cdot X)_t| = 0.$$

Proof For finite total variation processes this is clear, so we have to prove the statement only for continuous local martingales (the rest is done by a simple triangle inequality). By uniform integrability $(K^n)^\tau$ converges to zero in $L^2(X^\tau)$ for any stopping time τ of the localizing sequence of X. Hence $(K^n \cdot X)^\tau$ converges to zero in H^2 by Proposition 2.4.11 which yields the result by an argument as in Proposition 2.4.24. ■

This theorem enables us to establish the connection between this abstract approach to integration and the concrete limit of Riemannian sums:

Proposition 2.4.29 (Riemannian Sums)
Given a right-continuous, locally bounded process K and a refining sequence $\{\Delta_n\}$ of partitions $0 = t_0 < t_1 ... <_n = t$ of $[0, t]$ with mesh $|\Delta_n|$ tending to zero, then

$$(K \cdot X)_t = \int_0^t K_s dX_s = \lim_{n\to\infty} \sum_{i=0}^{n-1} K_{t_i} \left(X_{t_{i+1}} - X_{t_i} \right).$$

Proof Assume first that K is bounded, then the sums on the right hand side are stochastic integrals of the predictable step processes $\sum_{i=0}^{n-1} K_{t_i} 1_{[t_i, t_{i+1}[}$ which converge pointwise to K. But since they are bounded by $|K_{t_i}| \leq \|K\|_\infty$ we can use the previous theorem on dominated convergence for $K - \sum_{i=0}^{n-1} K_{t_i} 1_{[t_i, t_{i+1}[}$ to get the result. To generalize it we can - since K is required to be locally bounded - choose a localizing sequence $\tau_n \uparrow \infty$. ■

This proposition is not only for itself interesting but gives us the possibility to now prove an integration by parts formula for continuous semimartingales (which are obviously locally bounded).

Proposition 2.4.30 (Integration by Parts)
Given two continuous semimartigales X and Y, then it holds that

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y \rangle_t .$$

Proof Given a refining sequence $\{\Delta_n\}$ of partitions with mesh tending to zero we have for a concrete sequence

$$\begin{aligned}\sum_{i=0}^{n-1} \left(X_{t_{i+1}} - X_{t_i} \right)^2 &= \sum_{i=0}^{n-1} X_{t_{i+1}}^2 + \sum_{i=0}^{n-1} X_{t_i}^2 - \sum_{i=0}^{n-1} X_{t_{i+1}} X_{t_i} \\ &= X_t^2 + X_0^2 - 2 \sum_{i=0}^{n-1} X_{t_i} \left(X_{t_{i+1}} - X_{t_i} \right)\end{aligned}$$

and in the limit

$$\langle X, X\rangle_t = X_t^2 - X_0^2 - 2\int_0^t X_s dX_s,$$

whence

$$X_t^2 = X_0^2 + 2\int_0^t X_s dX_s + \langle X, X\rangle_t$$

From this special case for $X = Y$ we can derive the general one by polarization.

■

The integration by parts formula is a major ingredient for the Itô formula which can be derived for $X = (X^1, ..., X^d) \in \mathbb{R}^d$ where every X^i is a continuous martingale.

Theorem 2.4.31 (Itô Formula)
Let $f \in C_b^2(\mathbb{R}^d, \mathbb{R})$ and $X \in \mathbb{R}^d$ a continuous semimartingale, then $f(X)$ is a continuous semimartingale and

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x^i} f(X_s) dX_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2}{\partial x^i \partial x^j} f(X_s) d\left\langle X^i, X^j\right\rangle_s . \tag{2.19}$$

Proof We will prove the formula for the one-dimensional case to avoid that too much ink obscures the ideas behind the proof. (But for the d-dimensional case the proof is analogous since the coordinate processes are in $\mathcal{C}$ and multiplications with them too.) We denote the class of C_b^2-functions for which the equation (2.19) holds by $\mathcal{C}$. It is easy to see that it holds for constants $f(X) \equiv c$ and the identity $f(X) = X$. Next, we show that $f \in \mathcal{C}$ implies $xf(x) \in \mathcal{C}$: By the integration by parts formula we get

$$\begin{aligned}
& X_t f(X_t) - X_0 f(X_0) = (X \cdot f(X))_t + (f(X)\cdot X)_t + \langle X, f(X)\rangle \\
= & \left(X \cdot \left(f(X_0) + \int_0^{\cdot} f'(X_s)ds + \frac{1}{2}\int_0^{\cdot} f''(X_s) d\langle X, X\rangle_s\right)\right)_t + (f(X)\cdot X)_t \\
& + \left\langle X, f(X_0) + \int_0^{\cdot} f'(X_s)ds + \frac{1}{2}\int_0^{\cdot} f''(X_s)d\langle X, X\rangle_s\right\rangle \\
= & \int_0^t X_s f'(X_s)ds + \frac{1}{2}\int_0^t X_s f''(X_s) d\langle X, X\rangle_s \\
& + \int_0^t f(X_s)dX_s + \left\langle X, \int_0^t X_s f'(X_s)ds\right\rangle \\
= & \int_0^t (X_s f(X_s))' dX_s + \frac{1}{2}\int_0^t (X_s f(X_s))'' d\langle X, X\rangle_s
\end{aligned}$$

since we can expand f (in $\mathcal{C}$!) by the Itô formula and since in the last term the Stieltjes integrals have quadratic variation zero. By iteration this implies that all polynomials are in $\mathcal{C}$. Since the polynomials form a point separating subalgebra on (non-degenerated) compacts we can use a (Stone-) Weierstraß-argument to show that they are dense in $\mathcal{C}$. Choosing intervals $[-c_k, c_k]$, $c_k \uparrow \infty$ there exists a sequence p_n of polynomials such that $\lim_{n\to\infty} \sup_{|x|\le c_k} |p_n(x) - f(x)| = 0$. Lebesgue's theorem on dominated convergence (by $\|f\|_\infty$) entails that we can integrate the p's twice to get polynomials P_n, satisfying for $n \to \infty$

$$\sup_{|x|\le c_k} |P_n(x) - f(x)| \to 0$$
$$\sup_{|x|\le c_k} |P_n'(x) - f'(x)| \to 0$$
$$\sup_{|x|\le c_k} |P_n''(x) - f''(x)| \to 0.$$

Let $X = M + A$ the canonical decomposition we get by dominated convergence

$$\int_0^t P_n'(X_s)dX_s + \frac{1}{2}\int_0^t P_n''(X_s)d\langle X, X\rangle_s \to \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_0^t f''(X_s)d\langle X, X\rangle_s$$

in probability. So all C_b^2-functions on intervals $[-c_k, c_k]$ are in $\mathcal{C}$ and letting now $k \to \infty$ we get the desired result that the Itô formula holds for arbitrary C_b^2-functions on $\mathbb{R}_{\ge 0}$. Note that the approximation was only one in probability which is not necessarily uniform. ■

As in the case of Brownian motion we will find it convenient to apply the differential notation, hence the Itô formula reads

$$df(X) = \frac{1}{2}\sum_{i=1}^d \frac{\partial}{\partial x^i} f(X)dX^i + \sum_{i,j=1}^d \frac{\partial^2}{\partial x^i x^j} f(X)d\left\langle X^i, X^j\right\rangle.$$

2.4.4 The Stratonovich Integral

As the last point of this section we will present another notion of stochastic integral which was developed by Fisk and Stratonovich. Even a quite strong restriction of the integrands can not truncate two salient features of the Stratonovich integral: It gives us a substitution rule far more simple as Itô's and we get deeper insight in the convergence of the Riemannian sums.

Definition 2.4.32 (Stratonovich Integral)
Given two continuous semimartingales X and Y, the Stratonovich integral of X along Y is defined as

$$\int_0^t X_s \circ dY_s := \int_0^t X_s dY_s + \frac{1}{2}\langle X, Y\rangle_t.$$

The addition of the quadratic variation (which made the use of continuous semimartingales - also as integrands - necessary) leads to a simplification of the change of variables formula (by requiring a higher grade of differentiability of the function f!).

Theorem 2.4.33 (Stratonovich Formula)
Given $X \in \mathbb{R}^d$ a continuous semimartingale and $f \in C^3(\mathbb{R}^d, \mathbb{R})$, it holds that

$$df(X) = \sum_{i=1}^{d} \frac{\partial}{\partial x^i} f(X) \circ dX^i$$

Proof For $f(x) = x^2$ we get for instance by definition $2 \int_0^t X_s \circ dX_s = X_t^2 - X_0^2$ which implies by polarization the integration of parts formula for Stratonovich integrals:

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s \circ dY_s + \int_0^t Y_s \circ dX_s.$$

From Itô's formula for f'

$$f'(X_t) = f'(X_0) + \int_0^t f''(X_s) dX_s + \frac{1}{2} \int_0^t f'''(X_s) d\langle X, X \rangle_s$$

we derive

$$\langle f'(X), N \rangle_t = \left\langle \int_0^{\cdot} f''(X_s) dX_s, N \right\rangle_t = (f''(X_s) \cdot \langle X_s, N \rangle)_t$$

for any continuous semimartingale N since the Stieltjes integral has quadratic variation zero. Hence

$$\begin{aligned} \int_0^t f'(X_s) \circ dX_s &= \int_0^t f'(X_s) dX_s + \langle f'(X_s), X_s \rangle \\ &= \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X, X \rangle_s = f(X_t) - f(X_0) \end{aligned}$$

by Itô's formula. ■

In Proposition 2.4.29 we saw that for right continuous, locally bounded processes K the stochastic integral can be written as limit of Riemannian sums $\sum_{i=0}^{n-1} K_{t_i} \left(X_{t_{i+1}} - X_{t_i} \right)$. Here we took the initial point of the interval to evaluate the function K. But was this choice a specific one or is the stochastic integral - as the classical Riemann integral - independent of the choice of the evaluation point in the interval? We will see that the latter is not the case, since

Proposition 2.4.34 (Riemannian Sums)
Given two continuous semimartingales X, Y and a refining sequence $\{\Delta_n\}$ of partitions $0 = t_0 < ... < t_n = t$ of $[0, t]$ with mesh $|\Delta_n|$ tending to zero, then

$$\int_0^t Y_s \circ dX_s = \lim_{n\to\infty} \sum_{i=0}^{n-1} \frac{Y_{t_{i+1}} + Y_{t_i}}{2} \left(X_{t_{i+1}} - X_{t_i}\right).$$

Proof We can write the sum as

$$\sum_{i=0}^{n-1} \frac{Y_{t_{i+1}} + Y_{t_i}}{2} \left(X_{t_{i+1}} - X_{t_i}\right) = \sum_{i=0}^{n-1} \left(Y_{t_i} + \frac{1}{2}\left(Y_{t_{i+1}} - Y_{t_i}\right)\right) \left(X_{t_{i+1}} - X_{t_i}\right)$$

$$= \sum_{i=0}^{n-1} Y_{t_i} \left(X_{t_{i+1}} - X_{t_i}\right) + \frac{1}{2} \sum_{i=0}^{n-1} \left(Y_{t_{i+1}} - Y_{t_i}\right) \left(X_{t_{i+1}} - X_{t_i}\right)$$

where the first sum converges by Proposition 2.4.29 to the Itô integral and the second one by Proposition 2.4.6 to the quadratic variation which yields the result by definition of the Stratonovich integral. ■

Hence we have seen that the integral is not independent of choice of the evaluation point. For the initial point we get the Itô integral, for the mean point the Stratonovich integral. By analogous considerations we can see that we get for the final point $\int_0^t X_s dY_s + \langle X, Y \rangle_t$ and with $\lambda \in [0, 1]$ for a weighted mean point $\lambda Y_{t_i} + (1 - \lambda) Y_{t_{i+1}}$ the expression $\int_0^t X_s dY_s + (1 - \lambda) \langle X, Y \rangle_t$. But for an arbitrary choice (as we can do it for classical Riemann integrals where the evaluation point may in every interval be differently chosen!) the stochastic integral does not have to exist since the "quadratic variation" in the sense $\sup_\Delta T_t^\Delta(Y)$ may be not finite, even if our quadratic variation defined as limit of a refining sequence of partitions exists!

2.5 Transformations of the Probability Measure

The crucial question which lies at the core of this section is the following: for which transformations of the Brownian motion (with respect to the probability space $(\Omega, \mathcal{F}, P)$) exists a probability measure Q on $(\Omega, \mathcal{F})$ such that the transformed process is again a Brownian motion, this time with respect to $(\Omega, \mathcal{F}, Q)$? The answer to this question gives the Cameron-Martin-Girsanov theorem, but before this proof we have to recall the certainly most famous elementary theorem on measure transformations, the Radon-Nikodym theorem.

Theorem 2.5.1 (Lebesgue-Radon-Nikodym Theorem)
Given a signed measure ν and a positive measure μ on $(X, \mathcal{A})$, both σ-finite, then it holds that

(i) there exist unique signed measures ν_1, ν_2 such that

$$\nu_1 \perp \mu, \quad \nu_2 \ll \mu, \quad \nu_1 + \nu_2 = \nu,$$

the Lebesgue decomposition of ν*,*

(ii) there exists a μ*-measurable function* g *a.s. unique, such that* $\nu_2 = g\mu$ *(i.e.* $\nu_2(A) = \int_A g d\mu$ *for every* $A \in \mathcal{A}$*), the Radon-Nikodym derivative* $g = \frac{d\nu_2}{d\mu}$*.*

Proof By the Jordan decomposition and the σ-finiteness of the measures we can choose μ, ν finite (true) measures without loss of generality.

Uniqueness: The uniqueness of (i) is clear since $\nu = \nu_1 + \nu_2 = \lambda_1 + \lambda_2$ implies $\nu_1 - \lambda_1 = \lambda_2 - \nu_2$ with $(\nu_1 - \lambda_1) \perp \mu$ and $(\lambda_2 - \nu_2) \ll \mu$ and $\nu_1 - \lambda_1 = \lambda_2 - \nu_2 \equiv 0$. If $\nu_2 = g\mu = h\mu$, it follows on $\{h \geq g\}$ that

$$\nu_2(\{h \geq g\}) = \int_{\{h \geq g\}} h d\mu = \int_{\{h \geq g\}} g d\mu$$

and hence $\int_{\{h \geq g\}} (h - g) d\mu = 0$. Analogously we get $\int_{\{h \leq g\}} (h - g) d\mu = 0$ and hence

$$\nu_2(h - g) = \int (h - g) d\mu = \int_{\{h \geq g\}} (h - g) d\mu + \int_{\{h \leq g\}} (h - g) d\mu = 0$$

yields the uniqueness of (ii).

Existence: We define the set

$$F := \left\{ f \geq 0 : \int_A f d\mu \leq \nu(A) \quad \text{for all} \quad A \in \mathcal{A} \right\}$$

which is, since $f \equiv 0$ is therein, not empty and so there exists a supremum $\alpha := \sup_{t \in F} \int f d\mu < \infty$. We can find a monotone increasing sequence f_n with $\int f_n d\mu \to \alpha$ and the limit $g := \lim_n f_n$ lies in F since

$$\int_A g d\mu = \int_A \lim_n f_n d\mu = \lim_n \int_A f_n d\mu \leq \nu(A)$$

by monotone convergence. Now we define our measures ν_1, ν_2 by $\nu_2 := g\mu$ and $\nu_1 := \nu - \nu_2$. By $\nu_2(A) = \int g 1_A d\mu = 0$ for A with $\mu(A) = 0$ follows $\nu_2 \ll \mu$ and it remains only to show $\nu_1 \perp \mu$.

This is done by defining a sequence of finite measures $\lambda_n := \nu_1 - \frac{1}{n}\mu$ with Hahn decomposition $P_n \cup N_n$. Setting $g_n := g + \frac{1}{n} 1_{P_n}$ we have for $A \subset P_n$

$$\int_A g_n d\mu = \left(\nu_2 + \frac{1}{n}\mu \right)(A) = (\nu - \lambda_n)(A) \leq \nu(A)$$

and for $A \subset N$ clearly

$$\int_A g_n d\mu = \nu_2(A) \leq \nu(A),$$

hence $g_n \in F$. So $\alpha \geq \int g_n d\mu = \int g d\mu + \frac{1}{n}\mu(P_n)$ implies $\mu(P_n) = 0$. On the other hand side

$$\nu_1(\Omega \setminus P) = \nu_1 \left(\bigcap_n N_n \right) \leq \nu_1(N_n) \leq \frac{1}{n}\mu(N_n) \leq \frac{1}{n}\mu(\Omega),$$

so $\nu_1(\Omega \setminus P) = 0$ implying $\nu_1 \perp \mu$. ∎

That a Brownian motion with respect to a filtered probability space endowed with a measure P is usually not a Brownian motion with respect to the same probability space with another probability measure Q seems to be clear. The essence of the Cameron-Martin-Girsanov theory of changes of probability measures is the following question: For which transformations does there a new probability measure exist, absolute continuous to the original one, under which martingale (or even semimartingale) properties are preserved. We will show Girsanov's theorem on the preservation of Brownian motion under translations, but first we have to prove a lemma which we will use therefore:

Lemma 2.5.2
Given a probability space $(\Omega, \mathcal{F}, P)$ and a non-negative function $f \in L^1(\Omega, \mathcal{F}, P)$ with $E(f) = 1$, then the probability measure Q given by $dQ := fdP$ is well defined and for every sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ it holds for any $X \in L^1(\Omega, \mathcal{F}, P)$ that

$$E_Q(X|\mathcal{G}) = \frac{E(fX|\mathcal{G})}{E(f|\mathcal{G})} \qquad Q\text{-}a.s.$$

Proof For any set $G \in \mathcal{G}$ we have on the one hand side

$$\begin{aligned} \int_G E_Q(X|\mathcal{G}) f dP &= \int_G E_Q(X|\mathcal{G})\, dQ = \int_G X dQ \\ &= \int_G X f dP = \int_G E(fX|\mathcal{G}) dP \end{aligned}$$

and on the other hand side

$$\begin{aligned} \int_G E_Q(X|\mathcal{G}) f dP &= E(E_Q(X|\mathcal{G}) f 1_G) = E(E(E_Q(X|\mathcal{G}) f 1_G|\mathcal{G})) \\ &= E(E_Q(X|\mathcal{G}) 1_G E(f|\mathcal{G})) = \int_G E_Q(X|\mathcal{G}) E(f|\mathcal{G}) dP. \end{aligned}$$

Putting both equations together and taking into account that they hold for arbitrary sets $G \in \mathcal{G}$ the equality of the two rightmost terms yields the result. ∎

Theorem 2.5.3 (Girsanov's Theorem)
Given a progressively measurable function $u_t \in L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{F}_p, dt \otimes P)$ and defining

$$\xi_t := e^{\int_0^t u_s dB_s - \frac{1}{2}\int_0^t u_s^2 ds},$$

then the following statements are equivalent:

(i) For all $T \geq 0$ there exists a probability measure Q_T on $\mathcal{F}_T$ which is absolutely continuous to $P|_{\mathcal{F}_T}$ such that $\left\{ B_t - \int_0^t u_s ds \right\}_{0 \leq t \leq T}$ has the law of a Brownian motion under Q_T on $\mathcal{F}_T$,

(ii) $E(\xi_T) = 1$,
(ii') $\{\xi_t\}_{0\leq t\leq T}$ *is a continuous martingale.*

In this case we have on $\tilde{\mathcal{F}}_T$ *(the natural filtration of* $\left\{B_t - \int_0^t u_s ds\right\}_{0\leq t\leq T}$*) the Radon-Nikodym derivative given by*

$$E\left(\frac{dQ_T}{dP|_{\mathcal{F}_T}}\middle|\tilde{\mathcal{F}}_T\right) = \xi_T.$$

Proof

(ii) $\Leftrightarrow$ (ii') That the latter statement implies the previous one is clear by the martingale property $E(\xi_T) = E(\xi_0) = 1$, on the other hand side defining stopping times $\tau_k := \inf\left\{t : \int_0^t u_s^2 ds \geq k\right\}$ gives us continuous martingales $\{\xi_t^{\tau_k}\}_{0\leq t\leq T}$ implying that $\{\xi_t\}_{0\leq t\leq T}$ is a continuous local martingale. But since $E(\xi_T) = 1$, it is a continuous true martingale.

(ii) $\Rightarrow$ (i) We want to show that $\left\{B_t - \int_0^t u_s ds\right\}_{0\leq t\leq T}$ is a Brownian motion with respect to the measure Q (yet to define), that is, in the language of the characteristic functions (compare proof of Corollary 2.2.3),

$$E\left(e^{i\lambda\left(\left(B_t - \int_0^t u_r dr\right) - \left(B_s - \int_0^s u_r dr\right)\right)}\right) = E\left(e^{i\lambda\left(B_t - B_s - \int_s^t u_r dr\right)}\right) = e^{-\frac{\lambda^2}{2}(t-s)}.$$

By Lemma 2.5.2 there exists a probability measure Q with $\frac{dQ}{dP} = \xi_T$ and we get by Itô's formula (Theorem 2.3.8)

$$d\xi_r = \xi_r\left(u_r dB_r - \frac{1}{2}u_r^2 dr\right) + \frac{1}{2}\xi_r u_r^2 dr = \xi_r u_r dB_r.$$

Defining X_r as the characteristic function of the increments of the shifted Brownian motion

$$X_r := e^{i\lambda\left(B_r - B_s - \int_s^r u_v dv\right)}$$

we get

$$\begin{aligned} dX_r &= X_r i\lambda dB_r - X_r i\lambda u_r dr + \frac{1}{2}X_r i^2\lambda^2 dr \\ &= i\lambda X_r dB_r - i\lambda X_r u_r dr - \frac{1}{2}\lambda^2 X_r dr \end{aligned}$$

and for the product

$$\begin{aligned} d(X_r\xi_r) &= \xi_r dX_r + X_r d\xi_r + dX_r d\xi_r \\ &= i\lambda\xi_r X_r dB_r - i\lambda\xi_r X_r u_r dr - \frac{1}{2}\lambda^2 X_r\xi_r dr + \\ &\quad + \xi_r X_r u_r dB_r + i\lambda\xi_r X_r u_r dr \\ &= \xi_r X_r u_r dB_r + i\lambda\xi_r X_r dB_r - \frac{1}{2}\lambda^2 X_r\xi_r dr. \end{aligned}$$

So we can calculate by the concrete representation of the previous lemma for $t \geq s$

$$\begin{aligned}
E_Q(X_t|\mathcal{F}_s) &= \frac{E(X_t\xi_T|\mathcal{F}_s)}{E(\xi_T|\mathcal{F}_s)} = \frac{E(X_t\xi_t|\mathcal{F}_s)}{\xi_s} \\
&= \frac{E\left(\xi_s + \int_s^t X_r u_r \xi_r dB_r + i\lambda \int_s^t \xi_r X_r dB_r - \frac{\lambda^2}{2}\int_s^t X_r\xi_r dr \middle| \mathcal{F}_s\right)}{\xi_s} \\
&= \frac{E\left(\xi_s - \frac{\lambda^2}{2}\int_s^t X_r\xi_r dr \middle| \mathcal{F}_s\right)}{\xi_s} \\
&= 1 - \frac{\lambda^2}{2}\int_s^t \frac{E(X_r\xi_r dr|\mathcal{F}_s)}{\xi_s} dr \\
&= E_Q(X_s|\mathcal{F}_s) - \frac{\lambda^2}{2}\int_s^t E_Q(X_r|\mathcal{F}_s)dr
\end{aligned}$$

by martingale properties and the conditional version of Fubini's theorem. The equation $Y_t = Y_s + k\int_s^t Y_r dr$ is a.s. uniquely fulfilled by $Y_t = e^{kt}$, so we get

$$E_Q(X_r|\mathcal{F}_s) = e^{-\frac{\lambda^2}{2}(t-s)}$$

as desired, the transformed process is under the changed measure a Brownian motion.

(i) $\Rightarrow$ (ii) In the other direction we have the measure Q_T given and set $\psi := \frac{dQ_T}{dP|_{\mathcal{F}_T}} \in L^1(\Omega, \mathcal{F}_T, P|_{\mathcal{F}_T})$. Then for any $h \in L^2([0,T])$ the exponentials $e^{\psi(h)}$ are dense in $L^2(\Omega, \mathcal{F}_T, P|_{\mathcal{F}_T})$ and $L^2(\Omega, \mathcal{F}_T, Q_T)$, so for non-negative bounded functions it holds that

$$E\left(f\left(B_t - \int_0^{t\wedge\tau_k} u_s ds\right)\xi_T^{\tau_k}\right) = E(f(B_t)) \tag{2.20}$$

by the same methods as in the other direction and, obviously, $E(\xi_T^{\tau_k}) = 1$. But by the definition of ψ we have

$$E\left(f\left(B_t - \int_0^t u_s ds\right)\psi\right) = E(f(B_t)) \tag{2.21}$$

too, so by conditioning (2.20) and (2.21) it follows that $\xi_T^{\tau_k} = E(\psi|\tilde{\mathcal{F}}_T^{\tau_k})$ a.s. which assures us that we can pass to the limit $k \uparrow \infty$, so $E(\xi_T) = 1$. ∎

Since it is not always easy to show that $E(\xi_T) = 1$ or that $\{\xi_t\}_{0\leq t\leq T}$ is a martingale one might be interested in an easier sufficient condition (which is, indeed, far from being necessary):

Corollary 2.5.4 (Novikov's Condition)
If $E\left(e^{\frac{1}{2}\int_0^T u_s^2 ds}\right) < \infty$, *then* $\{\xi_t\}_{0\le t\le T}$ *is a true martingale, hence the absolute continuous probability measure* Q_T *of Girsanov's theorem exists.*

Proof

(i) In a first step we proof that Novikov's condition implies that $e^{\int_0^t u_s dB_s}$ is an uniformly integrable submartingale: For stopping times τ_n as above we have

$$e^{\frac{1}{2}\int_0^{T\wedge\tau_n} u_s dB_s} = e^{\frac{1}{2}\int_0^{T\wedge\tau_n} u_s dB_s - \frac{1}{4}\int_0^{T\wedge\tau_n} u_s^2 ds} e^{\frac{1}{4}\int_0^{T\wedge\tau_n} u_s^2 ds},$$

so the Cauchy-Schwarz inequality implies

$$E\left(e^{\frac{1}{2}\int_0^{T\wedge\tau_n} u_s dB_s}\right) \le E(\xi_T^{\tau_n})^{\frac{1}{2}} E\left(e^{\frac{1}{2}\int_0^{T\wedge\tau_n} u_s^2 ds}\right)^{\frac{1}{2}} < \infty$$

and since the limit $n \to \infty$ on the left hand side is by the choice of the stopping times monotone, we can conclude that

$$E\left(e^{\frac{1}{2}\int_0^T u_s dB_s}\right) < \infty.$$

Furthermore $e^{\frac{1}{2}\int_0^T u_s dB_s}$ is an u.i. submartingale since $\frac{1}{2}\int_0^T u_s dB_s$ is an u.i. martingale.

(ii) For every $a \in]0,1[$ we get

$$\begin{aligned} e^{\int_0^t a u_s dB_s - \frac{1}{2}\int_0^t (a u_s)^2 ds} &= e^{a^2\left(\int_0^t u_s dB_s - \frac{1}{2}\int_0^t u_s^2 ds\right)} e^{a\int_0^t u_s dB_s (1-a)} \\ &= (\xi_t)^{a^2} (Z_t^a)^{1-a^2} \end{aligned}$$

for the u.i. submartingale

$$Z_t^a := e^{\left(\frac{a}{1+a}\int_0^t u_s dB_s\right)};$$

the family $\{Z_\tau^a\}$, τ a stopping time, is u.i. by Lemma 1.3.14 which can be proven analogously for submartingales (instead of martingales). We can now apply Hölder's inequality for $\frac{1}{\frac{1}{a^2}} + \frac{1}{\frac{1}{1-a^2}} = 1$ to get for some $\Gamma \in \mathcal{F}$ and a stopping time τ

$$\begin{aligned} E\left(1_\Gamma e^{\int_0^{T\wedge\tau} a u_s dB_s - \frac{1}{2}\int_0^{T\wedge\tau} (a u_s)^2 ds}\right) &\le E\left(\xi_t^\tau\right)^{a^2} E\left(Z_{T\wedge\tau}\right)^{1-a^2} \\ &= E\left(Z_{T\wedge\tau}\right)^{1-a^2} \end{aligned}$$

since $E(\xi_t^\tau) = 1$. So

$$\left\{ e^{\int_0^{T\wedge\tau} au_s dB_s - \frac{1}{2}\int_0^{T\wedge\tau} (au_s)^2 ds} \right\}_\tau$$

is an u.i. family and hence

$$\left\{ e^{\int_0^t au_s dB_s - \frac{1}{2}\int_0^t (au_s)^2 ds} \right\}_{0\le t\le T}$$

an u.i. martingale. This implies

$$1 = E\left(e^{\int_0^T au_s dB_s - \frac{1}{2}\int_0^T (au_s)^2 ds} \right) \le E\left(e^{\int_0^T u_s dB_s - \frac{1}{2}\int_0^T u_s^2 ds} \right)^{a^2} E(Z_T^a)^{1-a^2} \quad (2.22)$$

and since

$$Z_T^a \le 1_{\{\int_0^T u_s dB_s < 0\}} + e^{\frac{1}{2}\int_0^T u_s^2 dB_s} 1_{\{\int_0^T u_s dB_s > 0\}}$$

we can use dominated convergence to conclude that

$$\lim_{a\uparrow 1} E(Z_T^a)^{1-a^2} = 1.$$

So letting $a \uparrow 1$ in (2.22) we get $E(\xi_T) \ge 1$ whence $E(\xi_T) = 1$ yielding the result, namely that $\{\xi_t\}_{0\le t\le T}$ is a true martingale. ■

Chapter 3

Stochastic Differential Equations

The subject of this chapter are stochastic differential equations (abbreviated SDEs), surely the most important application of the notions of stochastic calculus. As ordinary or partial differential equations have wide use for modeling deterministic phenomena, so stochastic differential equations can serve as a tool for the description of stochastic phenomena, hence phenomena whose development in time is scattered by a white noise effect.

Definition 3.0.5 (Stochastc Differential Equation)
A (strong) solution of a stochastic differential equation is an adapted stochastic process $X_t^x : (\Omega, \mathcal{F}, \mathcal{F}_t, P) \to \mathbb{R}^n$ with continuous paths and initial value $X_0^x = x \in \mathbb{R}^n$ which satisfies for vector fields $V; V^1, ...V^d \in \mathbb{R}^n \to \mathbb{R}^n$ the equation

$$X_t^x = x + \int_0^t V(X_s^x)\, ds + \sum_{i=1}^{d} \int_0^t V^i(X_s^x)\, dB_s^i$$

for $t \geq 0$.

Following the notation of the previous chapter we will write

$$dX_t^x = V(X_t^x)dt + \sum_{i=1}^{d} V^i(X_t^x)dB_t^i$$

as a short form of this SDE. In opposition to the strong solution we call a pair $(\mathcal{F}_t, X_t^x)$ to a given probability space $(\Omega, \mathcal{F}, P)$ satisfying the above conditions a weak solution of the SDE. The vector field V is called the (Itô)-drift and the vector fields V^i the volatilities.

3.1 Some Inequalities

In this chapter we will a lot of inequalities need for the proofs. Besides the well known Hölder inequality and Doob's inequality for continuous martingales we will be concerned with Gronwall's lemma and the Burkholder-Davis-Gundy inequalities. Gronwall's lemma gives us an exponential restriction for functions which can be estimated by an integral over their past.

Lemma 3.1.1 (Gronwall's Lemma)
Let $v(t) : [0,T] \to \mathbb{R}_{\geq 0}$ be a bounded and non-negative measurable function such that

$$v(t) \leq b + a \int_0^t v(s)ds$$

for $t \in [0,T]$ and some $a, b \geq 0$. Then it holds that

$$v(t) \leq be^{at}$$

for all $t \in [0,T]$.

Proof Since $v(t)$ is bounded we can define a function by $w(t) := \int_0^t v(s)ds$ which is, since $v(t)$ has non-negative values, monotone increasing. So we can, even if $w(t)$ is not differentiable, note that

$$0 \leq \limsup_{h\to 0} \frac{w(t+h) - w(t)}{h} \leq b + aw(t)$$

and hence

$$w(t) \leq \int_0^t (b + aw(s))\, ds. \tag{3.1}$$

Now we define $f(t) := e^{-at}w(t)$ for $t \in [0,T]$ and conclude by basic theorems on Stieltjes integrals and (3.1) that the following inequality holds:

$$\begin{aligned} f(t) &= \int_0^t 1df(s) = \int_0^t e^{-as}dw(s) + \int_0^t w(s)de^{-as} \\ &\leq \int_0^t e^{-as}(b + aw(s))\, ds + \int_0^t w(s)(-a)e^{-as}ds \\ &= \int_0^t be^{-as}ds = \frac{b}{a}\left(e^{at} - 1\right)e^{-at} \end{aligned}$$

By the definition of $f(t)$ we get $w(t) \leq \frac{b}{a}(e^{at} - 1)$ and hence by the initial boundedness condition of $v(t)$

$$v(t) \leq b + aw(t) \leq be^{at}.$$

■

Another important tool are the Burkholder-Davis-Gundy (or abbreviated BDG) inequalities for continuous (local) martingales which estimate the p-th mean oft the supremum process M^* by that of the root of the quadratic variation process. We prove the inequalities first in a restricted case to develop then a procedure to generalize them for arbitrary $p > 0$.

Lemma 3.1.2
For any $p \geq 2$ there exists a constant C_p such that for every continuous local martingale M vanishing at zero the following inequality holds:

$$E\left(\left(\sup_{0\leq s\leq\infty}|M_s|\right)^p\right) = E\left((M^*_\infty)^p\right) \leq C_p E\left(\langle M,M\rangle_\infty^{\frac{p}{2}}\right).$$

Proof Remark first that it is enough to prove the lemma for bounded martingales M, since for a given localizing sequence $\tau_n \uparrow \infty$ we can define by continuity stopping times $\sigma_n := \inf\{t : |M_t| = n\}$ which also form a localizing sequence and all the respective M^{σ_n} are bounded martingales. So by passing to the limit $\sigma_n \uparrow \infty$ we get the result.
Since the function $f(x) = |x|^p$ is in $C_b^2(\mathbb{R})$ we can apply Itô's formula to get for $t \to \infty$

$$|M_\infty|^p = \int_0^\infty p|M_s|^{p-1}\operatorname{sgn}(M_s)dM_s + \frac{1}{2}\int_0^\infty p(p-1)|M_s|^{p-2}d\langle M,M\rangle_s,$$

which implies (since the expectation of stochastic integrals of processes vanishing at zero is zero) by taking the expectation

$$\begin{aligned} E\left(|M_\infty|^p\right) &= \frac{p(p-1)}{2}E\left(\int_0^\infty |M_s|^{p-2}d\langle M,M\rangle_s\right) \\ &\leq \frac{p(p-1)}{2}E\left((M^*_\infty)^{p-2}\langle M,M\rangle_\infty\right) \\ &\leq \frac{p(p-1)}{2}\left\|(M^*_\infty)^{p-2}\right\|_{\frac{p}{p-2}}\left\|\langle M,M\rangle_\infty\right\|_{\frac{p}{2}} \end{aligned}$$

by a classical supremum estimate for Stieltjes integrals by Hölder's inequality for $\left(\frac{p}{p-2}\right)^{-1} + \left(\frac{p}{2}\right)^{-1} = 1$. But by Doob's maximal inequality (Theorem 1.1.20)

$$\left(E\left(|M^*_\infty|^p\right)\right)^{\frac{1}{p}} = \|M^*_\infty\|_p \leq \frac{p}{p-1}\|M_\infty\|_p = \frac{p}{p-1}\left(E\left(|M_\infty|^p\right)\right)^{\frac{1}{p}},$$

implying that

$$\begin{aligned} E\left((M^*_\infty)^p\right) &\leq \left(\frac{p}{p-1}\right)^p E\left(|M_\infty|^p\right) \\ &\leq \left(\frac{p}{p-1}\right)^p\frac{p(p-1)}{2}\left\|(M^*_\infty)^{p-2}\right\|_{\frac{p}{p-2}}\left(E\left(\langle M,M\rangle_\infty^{\frac{p}{2}}\right)\right)^{\frac{2}{p}} \\ &\leq C_p E\left(\langle M,M\rangle_\infty^{\frac{p}{2}}\right) \end{aligned}$$

for a constant C_p since we can assume that $E\left(\langle M,M\rangle_\infty^{\frac{p}{2}}\right) \geq 1$ without loss of generality (because otherwise we could could make the whole estimate for a multiple since $\langle M,M\rangle = 0$ only if M is constant by Theorem 2.4.7). ∎

Lemma 3.1.3
For any $p \geq 4$ there exists a constant c_p such that for every continuous local martingale M vanishing at zero the following inequality holds:

$$c_p E\left(\langle M,M\rangle_\infty^{\frac{p}{2}}\right) \leq E\left((M_\infty^*)^p\right).$$

Proof In this proof we let $c_p^1, \ldots$ signify different constants depending only on p. From the integration of parts formula for continuous semimartingales (Proposition 2.4.30) we derive

$$M_t^2 = 2\int_0^t M_s dM_s + \langle M,M\rangle_t$$

implying by a triangle inequality

$$|\langle M,M\rangle_t| \leq \left|2\int_0^t M_s dM_s\right| + \left|M_t^2\right|.$$

Under $E\left((\cdot)^{\frac{p}{2}}\right)$ this gives for $t \to \infty$

$$E\left((\langle M,M\rangle_t)^{\frac{p}{2}}\right) \leq c_p^1\left(E\left(\left(\left|\int_0^\infty M_s dM_s\right|\right)^{\frac{p}{2}}\right)\right) + E\left((M_\infty^*)^p\right). \tag{3.2}$$

Since $\int_0^\cdot M_s dM_s$ is by definition a continuous local martingale vanishing at zero, we can apply the previous lemma, so

$$E\left(\left(\sup_{0\leq s\leq\infty}\left|\int_0^t M_s dM_s\right|\right)^p\right) \leq C_p E\left(\left\langle \int_0^\cdot M_s dM_s, \int_0^\cdot M_s dM_s\right\rangle_\infty^{\frac{p}{2}}\right)$$

implying

$$E\left(\left|\int_0^\infty M_s dM_s\right|^p\right) \leq c_p^2 E\left(\left(\int_0^\infty M_s^2 d\langle M_s, M_s\rangle_s\right)^{\frac{p}{2}}\right).$$

Applying this estimate on the inequality (3.2) we get

$$\begin{aligned}
E\left(\langle M,M\rangle_\infty^{\frac{p}{2}}\right) &\leq c_p^3\left(E\left((M_\infty^*)^p\right) + E\left(\left(\int_0^\infty M_s^2 d\langle M,M\rangle_s\right)^{\frac{p}{4}}\right)\right) \leq \\
&\leq c_p^4\left(E\left((M_\infty^*)^p\right) + \left(E\left((M_\infty^*)^p\right) E\left(\langle M,M\rangle_\infty^{\frac{p}{2}}\right)\right)^{\frac{1}{2}}\right)
\end{aligned}$$

by a supremum estimate for the Stieltjes integral and equivalence of norms. Abbreviating

$$\begin{aligned} x &= \left(E\left(\langle M, M\rangle_\infty^{\frac{p}{2}}\right)\right)^{\frac{1}{2}} \\ y &= \left(E\left((M_\infty^*)^p\right)\right)^{\frac{1}{2}} \end{aligned}$$

this reads

$$x^2 \leq c_p^4(y^2 + xy)$$

or

$$0 \geq x^2 - c_p^4 xy - c_p^4 y^2.$$

Since the adjoint quadratic equation has by Vieta for x two real solutions of type $c_p^5 y$ this entails that we can choose for every $p \geq 4$ a constant c_p such that the inequality holds. ■

To generalize these results to all positive p's we will use a reduction process based on the notion of the dominated process.

Definition 3.1.4
A positive, adapted right-continuous process X is said to be dominated by an increasing (non-negative) process A, iff for every bounded stopping time τ it holds that $E(X_\tau) \leq E(A_\tau)$.

Lemma 3.1.5
Given a process X dominated by a continuous process A, then there exist positive real numbers x, y such that

$$P\left(\{\omega : X_\infty^* > x, A_\infty \leq y\}\right) \leq \frac{1}{x} E\left(A_\infty \wedge y\right).$$

Proof Defining stopping times $\rho := \inf\{t : A_t > y\}$ and $\sigma := \inf\{t : X_t > x\}$ we have $\{\omega : A_\infty \leq y\} = \{\omega : \rho = \infty\}$ and - for any $n \geq 0$ - $\{\omega : X_n^* > x\} \subset \{\omega : \sigma < \infty\}$, so we can estimate by Chebyshev's inequality and dominance

$$\begin{aligned} P\left(\{\omega : X_n^* > x, A_\infty \leq y\}\right) &\leq P\left(\{\omega : \sigma < \infty, \rho = \infty\}\right) \\ &\leq P\left(\{\omega : X_{\sigma\wedge\rho} > x\}\right) \\ &\leq \frac{1}{x} E\left(X_{\sigma\wedge\rho}\right) \leq \frac{1}{x} E\left(A_{\sigma\wedge\rho}\right). \end{aligned}$$

Since $A_\rho = A_t 1_{\inf\{t : A_t > y\}} = y$ by continuity of A we have $A_{\sigma\wedge\rho} \leq A_\infty \wedge y$ and conclude that for an arbitrary $n \geq 0$

$$P\left(\{\omega : X_n^* > x, A_\infty \leq y\}\right) \leq \frac{1}{x} E\left(A_\infty \wedge y\right).$$

Applying Fatou's Lemma yields the result. ■

We use this result to prove the proper reducing lemma:

Lemma 3.1.6
Given a stochastic process X dominated by a continuous process A, it holds for any $k \in]0,1[$ that

$$E\left((X_\infty^*)^k\right) \leq \frac{2-k}{1-k} E\left(A_\infty^k\right).$$

Proof Taking a continuous increasing function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $f(0) = 0$ we can reason by Fubini's theorem

$$\begin{aligned}
E\left(f(X_\infty^*)\right) &= E\left(\int_0^\infty 1_{X_\infty^*} df(x)\right) \\
&\leq \int_0^\infty \left(P\left(\{\omega : X_\infty^* > x, A_\infty \leq x\}\right) + P\left(\{\omega : A_\infty > x\}\right)\right) df(x) \\
&\leq \int_0^\infty \left(\frac{1}{x} E\left(A_\infty \wedge x\right) + P\left(\{\omega : A_\infty > x\}\right)\right) df(x) \\
&\leq \int_0^\infty \left(\frac{1}{x} E\left(A_\infty 1_{\{A_\infty \leq x\}}\right) + 2P\left(\{\omega : A_\infty > x\}\right)\right) df(x) \\
&= E\left(A_\infty \int_{A_\infty}^\infty \frac{1}{x} df(x)\right) + 2E\left(f(A_\infty)\right)
\end{aligned}$$

since $E(A_\infty \wedge x) \leq E\left(A_\infty 1_{\{A_\infty \leq x\}}\right) + xP(\{A_\infty > x\})$. Setting now $f(x)x^k$ we get

$$\begin{aligned}
E\left((X_\infty^*)^k\right) &\leq E\left(A_\infty \int_{A_\infty}^\infty \frac{1}{x} dx^k\right) + 2E\left(A_\infty^k\right) \\
&\leq \left(2 + \frac{k}{1-k}\right) E\left(A_\infty^k\right) = \frac{2-k}{1-k} E\left(A_\infty^k\right)
\end{aligned}$$

since $\int_{A_\infty}^\infty \frac{1}{x} dx^k = \frac{k}{k-1} x^{\frac{k-1}{k}} \Big|_{A_\infty^k}^\infty = \frac{k}{1-k} A_\infty^{k-1}$. ■

Theorem 3.1.7 (BDG Inequalities)
Given $p > 0$, there exist two (non-negative) universal constants c_p and C_p such that for all continuous local martingales M vanishing at zero

$$c_p E\left(\langle M, M\rangle_\infty^{\frac{p}{2}}\right) \leq E\left((M_\infty^*)^p\right) \leq C_p E\left(\langle M, M\rangle_\infty^{\frac{p}{2}}\right).$$

Proof For the cases $p \geq 4$ resp. $p \geq 2$ the inequalities are already proven. For the right inequality we set $X = (M^*)^2$ (which is obviously continuous, positive and adapted) and $A = C_2 \langle M, M\rangle$ (clearly continuous and increasing) so that the above lemma gives us

$$E\left((M_\infty^*)^{2k}\right) \leq \frac{2-k}{1-k} C_2^k E\left(\langle M, M\rangle_\infty^k\right),$$

hence filling the gap for $p \in]0,2[$: $C_p := \frac{4-p}{2-p} C_2^{\frac{p}{2}}$. By the same processes we get for the left inequality

$$E\left(\left(\langle M, M \rangle_\infty^* \right)^{2k} \right) = E\left(\langle M, M \rangle_\infty^{2k} \right) \le \frac{2-k}{1-k} \left(\frac{1}{c_4} \right)^k E\left((M_\infty^*)^{4k} \right)$$

implying the result for $p \in]0,4[$: $c_p := \frac{4-p}{8-p} c_4^{\frac{p}{4}}$. ■

By optional stopping we get as corollary:

Corollary 3.1.8 (Stopped BDG Inequalities)
For an arbitrary stopping time τ it holds under the above conditions that

$$c_p E\left(\langle M, M \rangle_\tau^{\frac{p}{2}} \right) \le E\left((M_\tau^*)^p \right) \le C_p E\left(\langle M, M \rangle_\tau^{\frac{p}{2}} \right).$$

3.2 Solutions of SDEs

Now we can proceed to the proof of the first central theorem for SDEs, the theorem of the existence and uniqueness of a solution, which is a straight forward generalization of the analogon for classical ODEs. For the sake of simplicity we will use a slightly different notation, even as it is not so obvious as the vector field notation: Here we write instead of V for the deterministic part $b(t,x)$ and instead of the volatilities V^i we will use a notation with matrix valued $\sigma(t,x) : [0,T] \to M_{n,d}(\mathbb{R})$. As norms we will use the 1-norms for vectors and matrices, hence $\|b\| = \sum_{i=1}^{n} |b_i|$ for $b \in \mathbb{R}^n$ and $\|A\| = \sum_{i,j=1}^{n} |A_{i,j}|$ for $A \in M_{n,d}(\mathbb{R})$.

Theorem 3.2.1 (Existence and Uniqueness of Solutions of SDEs)
Under the growth restriction that for all pairs $(t,x) \in [0,T] \times \mathbb{R}^n$ there exists a constant $C \in \mathbb{R}_{\ge 0}$ such that

$$\|b(t,x)\| + \|\sigma(t,x)\| \le C\left(1 + \|(x)\|\right)$$

and assuming the Lipschitz condition that a constant $D \in \mathbb{R}_{\ge 0}$ exists for (t,x), $(t,y) \in [0,T] \times \mathbb{R}^n$ such that

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \le D\|x - y\|,$$

for the given SDE

$$dX_t^x = b(t, X_t^x)dt + \sigma(t, X_t^x)dB_t$$

with initial value $Z \in (\Omega, \mathcal{F}_0, P)$ satisfying $E(|Z|^2) < \infty$ there exists a unique continuous stochastic process $X_t(\omega)$ with $Z = X_0$ on $[0,T]$ satisfying the SDE.

Proof First we will proof the uniqueness of the solution: Assume that there are two processes $(X_t)_{0 \le t \le T}$ and $(Y_t)_{0 \le t \le T}$ with initial values $X_0 = Z^X$ and $Y_0 = Z^Y$ both satisfying the SDE and the other required conditions. With respect to the stochastic part it is clear, that in this case the essential convergence is that in the L^2-sense, so we have to show that $E\left(\|X_t - Y_t\|^2\right) = 0$. For the expansion we can, abbreviating the kernel of the deterministic integral by a_s and that of the stochastic integral by γ_s, use the triangle equation and estimate the mixed terms by a multiple of the squares

$$\begin{aligned}
& E\left(\|X_t - Y_t\|^2\right) \\
= \; & E\left(\|Z^X - Z^Y + \int_0^t b(s,X_s) - b(s,Y_s)ds + \int_0^t \sigma(s,X_s) - \sigma(s,Y_s)dB_s\|^2\right) \\
\leq \; & K\left(E\left(\|Z^X - Z^Y\|^2\right) + E\left(\|\int_0^t a_s ds\|^2\right) + E\left(\|\int_0^t \gamma_s dB_s\|^2\right)\right) \\
\leq \; & KE\left(\|Z^X - Z^Y\|^2\right) + tKE\left(\int_0^t \|a_s\|^2 ds\right) + KE\left(\int_0^t tr\left(\gamma_s \gamma_s^\tau\right) ds\right) \\
\leq \; & KE\left(\|Z^X - Z^Y\|^2\right) + tD^2 KE\left(\int_0^t \|X_s - Y_s\|^2 ds\right) \\
& + KD^2 E\left(\int_0^t \|X_s - Y_s\|^2 ds\right)
\end{aligned}$$

using for the deterministic part we can now estimate by the Cauchy inequality

$$\|\int_0^t a_s ds\|^2 = \|\left\langle 1_{[0,t]}, a_s\right\rangle\|^2 \leq \left\langle 1_{[0,t]}, 1_{[0,t]}\right\rangle \langle a_s, a_s\rangle \leq t\int_0^t \|a_s\|^2 ds, \qquad (3.3)$$

for the stochastic part simply the Itô-lemma (2.3.5) and an estimate by the Lipschitz condition for both parts. This means (since $t \leq T$) nothing else as that there exist some constant $A, F \in \mathbb{R}$ such that

$$E\left(\|X_t - Y_t\|^2\right) \leq F + A\int_0^t E\left(\|X_s - Y_s\|^2\right) ds$$

which allows us to apply Gronwall's lemma to get

$$E\left(\|X_t - Y_t\|^2\right) \leq Fe^{At}.$$

But obviously both processes have to have the same initial value $Z^X = Z^Y$ and hence $F = 0$ whence

$$E\left(\|X_t - Y_t\|^2\right) = 0$$

for all $t \in [0,T]$. This and the fact that the paths of of the processes are continuous let us conclude that both solutions are indistinguishable.

To proof the existence of a solution of the SDE under the given conditions we will use more or less the same approach as in the case of classical ODEs: the Picard-Lindelöf iteration. We define a sequence $(Y_t^k)_{0\leq t\leq T}$ for $k \geq 0$ by the following recursion:

$$\begin{cases} Y_t^0 & := Z \in \mathbb{R}^n \text{ for all } t \in [0,T] \\ Y_t^{k+1} & := Z + \int_0^t b\left(s, Y_s^k\right) ds + \int_0^t \sigma\left(s, Y_s^k\right) dB_s \end{cases}$$

First we remark that the Y_t^k are all adapted and continuous since they are a sum of the initial value (which is trivially continuous and was required to be $\mathcal{F}_0$-measurable too) and finite classical and stochastic integrals and so (σ and b are bounded on compacts by the above growth restriction) fulfill the conditions trivially.
Now we have to proof that the sequence Y_t^k is converging: Since

$$Y_t^{k+1} - Y_t^k = \int_0^t b\left(s, Y_s^k\right) - b\left(s, Y_s^{k-1}\right) ds + \int_0^t \sigma\left(s, Y_s^k\right) - \sigma\left(s, Y_s^{k-1}\right) dB_s$$

we can conclude with the above argument (4.2) from the uniqueness part that there exists an $A \in \mathbb{R}$ such that for $k \geq 0$

$$\begin{aligned} & E\left(\|Y_t^{k+1} - Y_t^k\|^2\right) \\ = \; & E\left(\|\int_0^t b\left(s, Y_s^k\right) - b\left(s, Y_s^{k-1}\right) ds + \int_0^t \sigma\left(s, Y_s^k\right) - \sigma\left(s, Y_s^{k-1}\right) dB_s\|^2\right) \\ \leq \; & A \int_0^t E\left\|Y_s^k - Y_s^{k-1}\|^2\right) ds. \end{aligned}$$

For the first difference we have by the above argumentation and the growth restriction

$$\begin{aligned} E\left(\|Y_t^1 - Y_t^0\|^2\right) &= E\left(\|\int_0^t b(s,Z)ds + \int_0^t \sigma(s,Z)dB_s\|^2\right) \\ &\leq E\left(t\int_0^t \|b(s,Z)\|^2 ds + \int_0^t \|\sigma(s,Z)\|^2 ds\right) \\ &\leq E\left((1+t)\int_0^t \left(C\left(1 + \|Z\|\right)\right)^2 ds\right), \end{aligned}$$

so that we have $E\left(\|Y_t^1 - Y_t^0\|^2\right) \leq Bt^2$ for a $B \in \mathbb{R}$ and can conclude by induction that

$$E\left(\|Y_t^{k+1} - Y_t^k\|^2\right) \leq A^k B \frac{t^{k+2}}{(k+1)!}. \tag{3.4}$$

Estimating the supremum we have

$$\begin{aligned} & \sup_{0\leq t\leq T} E\left(\|Y_t^{k+1} - Y_t^k\|\right) \\ \leq \; & \int_0^T \|b\left(s, Y_s^k\right) - b\left(s, Y_s^{k-1}\right)\|ds + \sup_{0\leq t\leq T} \|\int_0^t \sigma\left(s, Y_s^k\right) - \sigma\left(s, Y_s^{k-1}\right) dB_s\| \end{aligned}$$

so that we now can look where the processes distinguish: We define the sets $G_k := \left\{ \sup\limits_{0 \le t \le T} E\left(\|Y_t^{k+1} - Y_t^k\|^2 \right) > \frac{1}{2^k} \right\}$ and estimate

$$\begin{aligned}
P\left(G_k\right) &\le P\left(\left\{ \left(\int_0^T \|b\left(s, Y_s^k\right) - b\left(s, Y_s^{k-1}\right)\|ds \right)^2 > \frac{1}{2^{2k+2}} \right\} \right) \\
&\quad + P\left(\left\{ \sup_{0 \le t \le T} \| \int_0^t \sigma\left(s, Y_s^k\right) - \sigma\left(s, Y_s^{k-1}\right) dB_s \| > \frac{1}{2^{k+1}} \right\} \right) \\
&\le 2^{2k+2} T E\left(\int_0^T \|b\left(s, Y_s^k\right) - b\left(s, Y_s^{k-1}\right)\|^2 ds \right) \\
&\quad + 2^{2k+2} E\left(\int_0^T \|\sigma\left(s, Y_s^k\right) - \sigma\left(s, Y_s^{k-1}\right)\|^2 ds \right) \\
&\le 2^{2k+2}(T+1) \int_0^T A^k B \frac{t^{k+2}}{(k+1)!} ds \le R \frac{(4At)^{k+2}}{(k+1)!}
\end{aligned}$$

Chebyshev's and Jensen's inequalities for the deterministic part, the Doob inequality (1.1.20) for the stochastic part for $p = 2$ and by (4.3) for a $R \in \mathbb{R}$ and hence $\sum\limits_{k=1}^{\infty} P\left(G_k\right) < \infty$. So we can conclude by the Borel-Cantelli Lemma that $N := \{\omega : \omega \in G_k \text{ for infinitely many } k\}$ has measure zero: $P(N) = 0$. For $\omega \notin N$ there exists a $\mathcal{K}$ such that $\omega \notin A_k$ for $k > \mathcal{K}$, i.e.

$$\sup_{0 \le t \le T} \|Y_t^{k+1} - Y_t^k\|(\omega) \le \frac{1}{2^k}$$

and since $Y_t^{k+1} = \sum\limits_{i=0}^{k} \left(Y_t^{i+1} - Y_t^i\right)(\omega) + Z$ we know that the series Y_t^k is dominated by $\frac{1}{2^k}$, whence it converges. Since the approximation is uniformly not only for the deterministic but also for the stochastic part on $[0, T]$, there exists a process X_t with continuous paths such that $\lim\limits_{k \to \infty} Y_t^k = X_t$ for almost all ω. But since

$$E\left(\int_0^T \|Y_t^m - Y_t^n\|^2 dt \right) \to 0$$

Y_t^k is a Cauchy sequence and converges hence in $L^2_{prog}(\Omega \times [0, T])$ too. Finally the limit process X_t is also in $L^2_{prog}(\Omega \times [0, T])$ since by the argument (4.2) of the existence proof and Gronwall's lemma $E\left(\|X_t\|^2\right) \le \infty$, so X_t is a solution of the given SDE. ∎

In the following we want to point out the relationship between stochastic and partial differential equations, which is only possible in a setting a bit more specialized. So we introduce the notion of C^∞-bounded vector fields:

Definition 3.2.2 (C^∞-bounded Vector Fields)
A vector field $W : \mathbb{R}^n \to \mathbb{R}^n$ is called C^∞-bounded (or smooth bounded), iff for every multiindex $\alpha = (\alpha_1, ..., \alpha_n) \in c_0$ there exists a constant $C(\alpha)$ such that

$$|\frac{\partial^\alpha W}{\partial x^\alpha}(x)| \leq C(\alpha)$$

for partial derivatives of any degree n.

We can easily see that a C^∞-bounded vector field satisfies the growth restriction using the Taylor formula. The C^∞-boundedness gives us a constant estimate for the sum and a multiple of $\|x\|$ as estimate for the remainder. It also satisfies the Lipschitz condition since for the i-th component $W(y) - W(x) = \int_0^1 \frac{\partial}{\partial x_i} W(x + s(y-x))(x-y)_i ds$ where $\frac{\partial}{\partial x_i} W(\cdot)$ is now obviously bounded. So an SDE with C^∞-bounded vector fields as coefficients has a unique solution. We define the differential operator $\mathcal{L}$ by

$$(\mathcal{L}f)(x) := \sum_{j=1}^{n} V_j(x) \frac{\partial f}{\partial x_j}(x) + \sum_{k,l=1}^{n} \sum_{i=1}^{d} V_k^i(x) V_l^i(x) \frac{\partial^2 f}{\partial x_k \partial x_l}(x)$$

for vector fields $V; V^1, ...V^d$ in C_0^∞, hence smooth bounded vector fields with compact support for all (partial) derivatives. Using the conventions ∇ for the gradient operator $\left(\frac{\partial}{\partial x_i}\right)$ and H for the Hessian matrix $\left(\frac{\partial^2}{\partial x_i x_j}\right)$ we can write this in our b, σ notation as

$$(\mathcal{L}f)(x) = b^\tau(x) \nabla f(x) + \frac{1}{2} tr \left(\sigma(x) \sigma^\tau(x) H f(x)\right).$$

Lemma 3.2.3 (Dynkin's Formula)
If the stochastic process X_t^x is a solution of the SDE

$$X_t^x = x + \int_0^t V\left(X_s^x\right) ds + \sum_{i=0}^{d} \int_0^t V^i\left(X_s^x\right) dB_s^i$$

with initial value $X_0^x = x \in \mathbb{R}^n$, then for every smooth bounded function with compact support (so $f \in C_0^\infty$)

$$M_t^f := f(X_t^x) - f(x) - \int_0^t (\mathcal{L}f)(X_s^x) ds$$

is a martingale.

Proof The proof goes straight forward. Applying to the SDE the Itô formula for a $f \in C_0^\infty$ we get

$$df\left(X_t^x\right) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\left(X_t^x\right)\left(dX_t^x\right)_j + \frac{1}{2} \sum_{k,l=1}^{n} \frac{\partial^2 f}{\partial x_k \partial x_l}\left(X_t^x\right)\left(dX_t^x\right)_k \left(dX_t^x\right)_l$$

where by the basic rules of the Itô calculus

$$\begin{aligned}
& (dX_t^x)_k (dX_t^x)_l \\
= & \left(V_k (X_t^x) dt + \sum_{i=1}^{d} V_k^i (X_t^x) dB_t^i \right) \left(V_l (X_t^x) dt + \sum_{i=1}^{d} V_l^i (X_t^x) dB_t^i \right) \\
= & \sum_{i=1}^{d} V_k^i (X_t^x) V_l^i (X_t^x) dt.
\end{aligned}$$

Hence

$$\begin{aligned}
df (X_t^x) \quad = & \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} (X_t^x) V_j (X_t^x) dt + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} (X_t^x) \sum_{i=1}^{n} V_j^i (X_t^x) dB_t^i \\
& + \frac{1}{2} \sum_{k,l=1}^{n} \frac{\partial^2 f}{\partial x_k \partial x_l} (X_t^x) \sum_{i=1}^{d} V_k^i (X_t^x) V_l^i (X_t^x) dt.
\end{aligned}$$

Observing that the deterministic derivatives are nothing else than $(\mathcal{L}f)(X_t^x) dt$ we get

$$df (X_t^x) = (\mathcal{L}f)(X_t^x) + \sum_{i=1}^{d} V^i (X_t^x) \nabla f (X_t^x) dB_t^i.$$

Switching to the integral notation this is

$$f (X_t^x) = f(x) + \int_0^t (\mathcal{L}f)(X_s^x) ds + \sum_{i=1}^{d} \int_0^t V^i (X_s^x) \nabla f (X_s^x) dB_s^i \tag{3.5}$$

and since stochastic integrals are martingales (Theorem 2.3.6) we can conclude that

$$M_t^f = f(X_t^x) - f(x) - \int_0^t (\mathcal{L}f)(X_s^x) ds$$

is a martingale too. ∎

We only note here that also the converse is true: If M_t^f is a martingale for every $f \in C_0^\infty$, then X_t^x is a solution of the related SDE.

3.3 Flows and the First Variation

In the last sections we were interested especially in the properties of X_t with respect to the time t, i.e. path properties. We have seen that there exist solutions of SDEs with a.s. continuous paths and by Dynkin's lemma we even got differentiability of X_t with respect to the time since all terms of the right hand side of (4.4) are differentiable. Here we will ask the question if the process X_t is differentiable with respect to the *initial value*. But before we come to the derivative properties we first have to show that there exists a modification of X_t which is jointly continuous in t and x.

Proposition 3.3.1 (Joint Continuity)
Given X_t^x a solution of the SDE

$$dX_t^x = b(t, X_t^x)dt + \sigma(t, X_t^x)dB_t$$

with initial value x and satisfying the growth restriction and the Lipschitz condition for pairs $(t, x) \in [0, T] \times \mathbb{R}^n$. Then there exists a modification $\breve{X}_t^x$ of X_t^x which is jointly continuous in t and x.

To make the notion of modification clear in this context: For every pair (x, t) there exist a set N of measure zero such that X_t^x and the modification $\breve{X}_x^t$ agree outside N, i.e. the null set may depend on x and t.

Proof To proof the existence of a modification we use the Kolmogorov-Chentsov Theorem (Theorem 2.0.5). Now we profit of our quite general proof for d-dimensional time since we can use it for the dependence of the initial value. But some preparatory work is nevertheless necessary:

Looking at

$$X_t^x - X_t^y = x - y + \int_0^t b(s, X_s^x) - b(s, X_s^y)\, ds + \int_0^t \sigma(s, X_s^x) - \sigma(s, X_s^y)\, dB_s$$

we estimate for the deterministic part

$$E\left(\sup_{r \le t} \left\| \int_0^r b(s, X_s^x) - b(s, X_s^y) ds \right\|^{2p}\right) \le KE\left(\left(\int_0^t \|X_s^x - X_s^y\|^2\, ds\right)^p\right)$$

by a Cauchy argument analogously to (4.2). For the stochastic part the second BDG inequalities and Theorem 2.4.19 assert that

$$\begin{aligned} & E\left(\sup_{r \le t} \left\| \int_0^r \sigma(s, X_s^x) - \sigma(s, X_s^y) dB_s \right\|^{2p}\right) \\ \le \quad & C_{2p} E\left(\left(\int_0^r (\sigma(s, X_s^x) - \sigma(s, X_s^y))^2\, ds\right)^p\right) \\ \le \quad & C_{2p} D^{2p} E\left(\left(\int_0^t \|X_s^x - X_s^y\|^2\, ds\right)^p\right) \end{aligned}$$

by the Lipschitz condition. Hence we have by a triangle inequality constants c_1, c_2

$$\begin{aligned} E\left(\sup_{r \le t} \|X_r^x - X_r^y\|^{2p}\right) & \le \quad c_1 |x - y|^{2p} c_2 E\left(\left(\int_0^t \|X_s^x - X_s^y\|^2\, ds\right)^p\right) \\ & \le \quad c_1 \|x - y\|^{2p} c_2 E\left(\int_0^t \|X_s^x - X_s^y\|^{2p}\, ds\right) \end{aligned}$$

by Hölder's inequality.

Now we can use Gronwall's lemma which assures us the existence of a constant c_3 such

$$E\left(\sup_{s\le T}\|X_s^x - X_s^y\|^{2p}\right) \le c_3\|x-y\|^{2p}.$$

Since the supremum on the left hand side defines clearly a metric we have only to chose p large enough ($p > \frac{d}{2}$) and the Kolmogorov-Chentsov Theorem implies that $\breve{X}_s^x$ is a.s. uniformly continuous for dyadic rationals x. We only have to define $\breve{X}_t^y := \lim_{x\to y} X_t^x$, so $\breve{X}_t^x$ is jointly continuous in x and t and since $\breve{X}_t^x = X_t^x$ a.s. it is a solution of the SDE too. ■

Analogously to the notion of the *path* as collection $\{X_t\}$ depending on the time we will call the collection $\{X_t^x\}$ depending on x and t the *flow* of the SDE.

We have now proven that we can take a solution of an SDE as jointly continuous in t and x and it seems to be clear that with increasing smoothness of b and σ, also X_t^x becomes more and more smooth. But how far does this go? Can we - in the one dimensional case - in

$$\begin{aligned}\frac{X_t^{x+h}-X_t^x}{h} = \frac{(x+h)-x}{h} &+ \int_0^t \frac{b\left(s,X_s^{x+h}\right)-b\left(s,X_s^x\right)}{h}ds + \\ &+ \int_0^t \frac{\sigma\left(s,X_s^{x+h}\right)-\sigma\left(s,X_s^x\right)}{h}dB_s\end{aligned}$$

let decrease $h \downarrow 0$ to get something like

$$\frac{dX_t^x}{dx} = 1 + \int_0^t b'\left(s,X_s^x\right)\frac{dX_s^x}{dx}ds + \int_0^t \sigma'\left(s,X_s^x\right)\frac{dX_s^x}{dx}dB_s?$$

Indeed - under certain conditions - we can do so as the next theorem will prove. For the sake of commodity we will meanwhile remain in dimension one.

Theorem 3.3.2 (First Variation)
Given b, $\sigma \in C_b^2([0,T],\mathbb{R})$ so that they satisfy the growth restriction and the Lipschitz condition also for the derivatives. Then the SDE

$$dY_t^x = b'(t,X_t^x)Y_t^x dt + \sigma'(t,X_t^x)Y_t^x dB_t, \quad Y_0^x = 1$$

has a unique solution jointly continuous in x and t with absolute moments of all orders.

We will call this solution the first variation of X_t^x and denote it by $(DX)_t^x$.

Proof
(i) Uniqueness: For two solutions Y_t^x, $\hat{Y}_t^x$ and $N >$ we define the stopping

time $\tau_N := \inf\{t : |Y_t^x| \wedge |\hat{Y}_t^x| \geq N\}$. So the functions $b'(t, X_t^x)Z_{t\wedge\tau_n}^x$ and $\sigma'(t, X_t^x)Z_{t\wedge\tau_n}^x$ for $Z \in \{Y, \hat{Y}\}$ fulfill the Lipschitz condition and the growth restriction in t and z. Analogously to the uniqueness part of Theorem 3.2.1 we get

$$E\left(\sup_{s\leq t\wedge\tau_N} \|Y_s^x - \hat{Y}_s^x\|^2\right) \leq c_1 + c_2 \int_0^t E\left(\sup_{s\leq t\wedge\tau_N} \|Y_s^x - \hat{Y}_s^x\|^2\right) dt$$

for $t \leq T$. We can conclude again by Gronwall's lemma that

$$E\left(\sup_{s\leq t\wedge\tau_N} \|Y_s^x - \hat{Y}_s^x\|^2\right) = 0,$$

entailing $Y_s^x = \hat{Y}_s^x$ since N was arbitrary chosen.

(ii) The proof of the existence goes also analogously to the proof of Theorem 3.2.1 using the Picard-Lindelöf sequence

$$(DX)_t^{k+1} := 1 + \int_0^t b'(t, X_t^x)(DX)_t^k ds + \int_0^t \sigma'(t, X_t^x)(DX)_t^k dB_s.$$

(iii) Moments: We have to show that $E(|(DX)_t^x|^p)$ for all $p > 0$. Defining for $N > 0$ the stopping time $\sigma_N := \inf\{t : |(DX)_t^x| \geq N\}$ we get (since we can - at the price of some constants c_i depending on p - distribute the p-th absolute moment on the different summands)

$$\begin{aligned}
&E\left(\sup_{s\leq t\wedge\sigma_N} |(DX)_s^x|^p\right) \\
\leq\ & c_1 + c_2 E\left(\left(\int_0^{t\wedge\sigma_N} |b'(s, X_s^x)|\,|(DX)_s^x|\,ds\right)^p\right) + \\
&+ c_3 E\left(\left(\int_0^{t\wedge\sigma_N} \left(\sigma'(s, X_s^x)(DX)_s^x\right)^2 d_s\right)^{\frac{p}{2}}\right) \\
\leq\ & c_4 + c_5 E\left(\int_0^{t\wedge\sigma_N} |(DX)_s^x|^p\,ds\right),
\end{aligned}$$

using for the stochastic part the right BDG-inequality. Gronwall's lemma asserts us therefore that the left hand side is bounded by a constant; so we can let $N \to \infty$, implying $\sigma_N \uparrow \infty$ yielding the result.

(iv) Joint Continuity: To prove the existence of a jointly continuous modification we have to give an estimate

$$E\left(\|(DX)_t^x - (DX)_t^y\|^p\right) \leq k\|x - y\|^p.$$

The the Kolmogorov-Chentsov argument goes through as in Proposition 4.3.1; hence

$$
\begin{aligned}
& E\left(\|(DX)_t^x - (DX)_t^y\|^p\right) \\
\leq\ & c_1 E\left(\left\|\int_0^t b'(s, X_s^x)(DX)_s^x - b'(s, X_s^y)(DX)_s^y ds\right\|^p\right) + \qquad (3.6)\\
& + c_2 E\left(\left\|\int_0^t \sigma'(s, X_s^x)(DX)_s^x - \sigma'(s, X_s^y)(DX)_s^y dB_s\right\|^p\right). \qquad (3.7)
\end{aligned}
$$

For (3.6) we have by the triangle inequality

$$
\begin{aligned}
& E\left(\left\|\int_0^t b'(s, X_s^x)(DX)_s^x - b'(s, X_s^y)(DX)_s^y ds\right\|^p\right) \\
\leq\ & E\left(\left\|\int_0^t \left(b'(s, X_s^x) - b'(s, X_s^y)\right)(DX)_s^x ds\right\|^p\right) + \qquad (3.8)\\
& + E\left(\left\|\int_0^t b'(s, X_s^y)\left((DX)_s^x - (DX)_s^y\right) ds\right\|^p\right) \qquad (3.9)
\end{aligned}
$$

where we can estimate (3.8) by Cauchy's, Hölder's, Fubini's and again Cauchy's inequality

$$
\begin{aligned}
& E\left(\left\|\int_0^t \left(b'(s, X_s^x) - b'(s, X_s^y)\right)(DX)_s^x ds\right\|^p\right) \\
\leq\ & c_3 \int_0^t \|E\left(\left(b'(s, X_s^x) - b'(s, X_s^y)\right)(DX)_s^x\right)\|^p ds \\
\leq\ & c_3 \int_0^t \left(E\left(\|b'(s, X_s^x) - b'(s, X_s^y)\|^{2p}\right)\right)^{\frac{1}{2}} \left(E\left(\|(DX)_s^x\|^{2p}\right)\right)^{\frac{1}{2}} ds \\
\leq\ & c_4 \int_0^t \left(E\left(\|X_s^x - X_s^y\|^{2p}\right)\right)^{\frac{1}{2}} ds \\
\leq\ & c_5 \|x - y\|^p.
\end{aligned}
$$

where we used the Lipschitz condition and the boundedness of the 2p-th moment of (DX) to prepare the application of Kolmogorov-Chentsov theorem. The inequality (3.9) we simply estimate by the boundedness of b'

$$
\begin{aligned}
& E\left(\left\|\int_0^t b'(s, X_s^y)\left((DX)_s^x - (DX)_s^y\right) ds\right\|^p\right) \\
\leq\ & c_6 \int_0^t E\left(\|(DX)_s^x - (DX)_s^y\|^p\right) ds.
\end{aligned}
$$

The considerations for the stochastic part (3.7) are similar, to eliminate the stochastic integrals we have only to use the BDG inequalities; altogether we get

$$E\left(\|(DX)_t^x - (DX)_t^y\|^p\right) \le c_7\|x-y\|^p + c_8 \int_0^t E\left(\|(DX)_s^x - (DX)_s^y\|^p\right) ds$$

such that we can again use Gronwall's lemma to achieve

$$E\left(\|(DX)_t^x - (DX)_t^y\|^p\right) \le c_9\|x-y\|^p.$$

■

We have now proven that the first variation exists and it is clear that for smoother functions also higher derivatives exist. Especially for b, $\sigma \in C_b^\infty([0,T]\times\mathbb{R}_{\ge 0})$ the solution of the SDE is smooth with respect to the initial value. But we can prove even more: $(D\cdot)_t^x$ is a differential operator in the sense that it satisfies a version of the fundamental theorem of calculus.

Theorem 3.3.3 (Differentiability)
Let b, $\sigma \in C_b^\infty([0,T]\times\mathbb{R}_{\ge 0})$, then it holds for the first variation $(DX)_t^x$ that a.s.

$$X_t^x - X_t^y = \int_y^x (DX)_t^z dz.$$

Proof
(i) Since for deterministic processes $(DX)_t^x$ this is clear by Fubini's theorem, it is enough to proof the statement for purely stochastic processes $(DX)_t^x$ (i.e. $b\equiv 0$), the decomposition of both is obvious.

(ii) We define a process Z_t^x by

$$Z_t^x := X_t^x - \int_0^x (DX)_t^z dz, \tag{3.10}$$

so it is enough to show that this process is constant in x. Defining

$$\begin{aligned} F_t &:= \int_0^t (\sigma(s,X_s^x) - \sigma(s,X_s^y)) - \sigma'(s,X_s^x)(X_s^x - X_s^y) dB_s \\ G_t &:= \int_0^t \sigma'(s,X_s^x)(Z_s^x - Z_s^y) dB_s \\ H_t &:= \int_0^t \int_y^x (DX)_s^z \left(\sigma'(s,X_s^x) - \sigma'(s,X_s^z)\right) dz dB_s \end{aligned}$$

we can write by the definition of the first variation and stochastic Fubini

$$\begin{aligned}
& Z_t^x - Z_t^y \\
= \; & (x-y) + \int_0^t \sigma(s, X_s^x) - \sigma(s, X_s^y) dB_s - \int_y^x (DX)_t^z dz \\
= \; & (x-y) + \int_0^t \sigma(s, X_s^x) - \sigma(s, X_s^y) dB_s - \int_y^x 1 + \int_0^t \sigma'(s, X_s^z)(DX)_s^z dB_s dz \\
= \; & \int_0^t \sigma(s, X_s^x) - \sigma(s, X_s^y) dB_s - \int_0^t \int_y^x (DX)_s^z \sigma'(s, X_s^z) dz dB_s \\
= \; & F_t + G_t + H_t
\end{aligned}$$

since the sum of all products of $\sigma'(s, X_s^x)$ vanishes by summation; so now we estimate every summand for its own.
Developing $\sigma(s, X_s^x)$ in a Taylor series around X_y^s we have

$$\sigma(s, X_s^x) \leq \sigma(s, X_s^y) + \sigma'(s, X_s^x)(X_s^x - X_s^y) + \frac{1}{2} \|\sigma''(s, X_s^x)\|_\infty (X_s^x - X_s^y)^2$$

and can estimate

$$\begin{aligned}
E\left(\sup_{s \leq t} \|F_s\|^2\right) & \leq c_1 E\left(\sup_{s \leq t} \left\| \int_0^t (X_s^x - X_s^y)^2 dB_s \right\|^2\right) \\
& \leq c_1 E\left(\int_0^t \left\|(X_s^x - X_s^y)^2\right\|^2 ds\right) \\
& \leq c_1 E\left(\int_0^t \|X_s^x - X_s^y\|^4 ds\right) \\
& \leq c_2 \|x - y\|^4
\end{aligned}$$

by the right BDG-inequality, Hölder's inequality and a Gronwall argument as in 3.3.1.
By the Itô lemma we get for G_t

$$E\left(\sup_{s \leq t} \|G_s\|^2\right) \leq c_3 \int_0^t E\left(\|Z_s^x - Z_s^y\|^2\right) ds.$$

The case of H_t is a little bit harder; BDG inequalities, Lipschitz condition and a Cauchy argument as in (3.2) assert that

$$\begin{aligned}
&E\left(\sup_{s\leq t}\|H_s\|^2\right)\\
\leq\;& c_4\int_0^t E\left(\|\int_y^x (DX)^z_s(X^x_s-X^z_s)dz\|^2\right)ds\\
\leq\;& c_4\int_0^t E\left(\left\langle 1_{[y,x]},1_{[y,x]}\right\rangle\left\langle (DX)^z_s(X^x_s-X^z_s),(DX)^z_s(X^x_s-X^z_s)\right\rangle\right)ds\\
\leq\;& c_4\|x-y\|\int_0^t E\left(\|(DX)^z_s\|^2\,\|X^x_s-X^z_s\|^2\right)ds\\
\leq\;& c_4\|x-y\|\int_0^t\int_y^x\left(E\left(\|(DX)^z_s\|^4\right)\right)^{\frac{1}{2}}\left(E\left(\|X^x_s-X^z_s\|^4\,2\right)\right)^{\frac{1}{2}}dzds\\
\leq\;& c_5\|x-y\|\int_y^x\|x-y\|^2\,dz\\
\leq\;& c_6\|x-y\|
\end{aligned}$$

by Hölder's inequality, the boundedness of the fourth absolute moment and a Gronwall argument.
Putting all together we get

$$\begin{aligned}
E\left(\sup_{s\leq t}\|Z^x_s-Z^y_s\|^2\right) &\leq c_7E\left(\sup_{s\leq t}\|F_s\|^2+\|G_s\|^2+\|H_s\|^2\right)\\
&\leq c_8\|x-y\|^4+c_9\int_0^t E\left(\sup_{s\leq t}\|Z^x_s-Z^y_s\|^2\right)ds
\end{aligned}$$

so Gronwall's lemma implies

$$E\left(\sup_{s\leq t}\|Z^x_s-Z^y_s\|^2\right)\leq c_{10}\|x-y\|^4.$$

(iii) To show that Z^x_t is constant in x we estimate the set where it is greater then λ. For $n>0$ and $i\in\{0,...,n\}$ we define a partition $\{x_i\}$ by $x_i:=x+i\frac{y-x}{n}$, so we can conclude by Chebyshev's inequality that

$$\begin{aligned}
P\left(\{\omega : \|Z_t^x - Z_t^y\| > \lambda\}\right) &\leq P\left(\left\{\omega : \exists i, \|Z_t^{x_{i+1}} - Z_t^{x_i}\| > \frac{\lambda}{n}\right\}\right) \\
&\leq n \sup_i P\left(\left\{\omega : \|Z_t^{x_{i+1}} - Z_t^{x_i}\| > \frac{\lambda}{n}\right\}\right) \\
&\leq n \sup_i \left(\frac{n}{\lambda}\right)^2 E\left(\|Z_t^{x_{i+1}} - Z_t^{x_i}\|^2\right) \\
&\leq \frac{n^3}{\lambda} c_{10} \frac{\|x-y\|^4}{n^4} \\
&= c_{10} \frac{\|x-y\|^4}{n\lambda} \to 0
\end{aligned}$$

for $n \to \infty$. This implies $Z_t^x = Z_t^y$ a.s and hence

$$\int_y^x (DX)_t^z dz = X_t^x - X_t^y \quad a.s.$$

■

For higher dimensional SDE's the proofs go analogously (one only risks to get lost in the abundance of indices...), having the $n \times n$ matrix $(DX)_t^x$ as the first variation. It satisfies the stochastic differential equation

$$\begin{aligned}
&(DX)_t^x = ((DX)_t^x)_{ij} \\
&= I + \int_0^t \left(\frac{\partial b_i(s, X_s^x)}{\partial j}\right)_{ij} (DX)_s^x ds + \sum_{k=1}^d \int_0^t \left(\frac{\partial \sigma_{ik}(s, X_s^x)}{\partial j}\right)_{ij} (DX)_s^x dB_s^k \\
&= I + \int_0^t \left(\frac{\partial V_i(X_s^x)}{\partial j}\right)_{ij} (DX)_s^x ds + \sum_{k=1}^d \int_0^t \left(\frac{\partial V_i^k(X_s^x)}{\partial j}\right)_{ij} (DX)_s^x dB_s^k.
\end{aligned}$$

Chapter 4

Wiener Chaos and Malliavin Derivatives

In Chapter 2 we gave a classical approach to stochastic integration. Now we will look at this subject from a different point of view which concentrates more on the structural properties. We will present a decomposition of $L^2(\Omega, \mathcal{F}, P)$ by an orthogonal direct sum of Hilbert spaces, the so called Wiener chaos decomposition. Then we construct (multiple) Wiener-Itô integrals and point out the relationship to the classical Itô integrals of Chapter 2. In the third part we present the main ideas of Malliavin Calculus which introduces the weak differentiation (in the spirit of Laurent Schwartz's distribution theory) to stochastic analysis. The adjoint of this derivative operator, the so called Skorohod integral is the object of interest in Section 4 and we show that it is nothing else then a generalization of the now well known Itô integral. The chapter closes by pointing out the connection of Malliavin calculus to the first variation process of Chapter 3 and some cursory considerations on the existence of densities.

4.1 The Wiener Chaos Decomposition

Before we can present the decomposition of $L^2(\Omega, \mathcal{F}, P)$ we yet have to know some facts about Hermite polynomials.

4.1.1 Hermite Polynomials

On $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$, the space of square-integrable functions with respect to the Gaussian measure $\nu(A) := \int\limits_A e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$, we can define the operators d and δ for $\varphi \in \mathbb{R}[x]$, the ring of polynomials, by (requiring φ, $d\varphi$ and $\delta\varphi$ to be smooth functions in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$) $d\varphi := \frac{d\varphi}{dx}$ and δ as its adjoint operator. We can calculate δ directly by partial integration where the first term vanishes since φ and ψ were both required to be square-integrable and hence are dominated by $e^{-\frac{x^2}{2}}$:

$$
\begin{aligned}
\langle d\varphi, \psi \rangle &= \int_{\mathbb{R}} \frac{d\varphi(x)}{dx} \psi(x) e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} = - \int_{\mathbb{R}} \varphi(x) \frac{d}{dx} \left(\psi(x) e^{-\frac{x^2}{2}} \right) \frac{dx}{\sqrt{2\pi}} \\
&= \int_{\mathbb{R}} \varphi(x) \left(-\frac{d\psi(x)}{dx} + x\psi(x) \right) e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} = \langle \varphi, -d\psi + x\psi \rangle \\
&= \langle \varphi, \delta\psi \rangle . \qquad (4.1)
\end{aligned}
$$

Hence $\delta\varphi = -d\varphi + x\varphi$ and we can easily see that $d\delta - \delta d = id_{\mathbb{R}[x]}$, an equation which is sometimes called Heisenberg relation.

Definition 4.1.1 (Hermite Polynomials)
We define the Hermite polynomials $H_n \in \mathbb{R}[x]$ by

$$H_n(x) := (\delta)^n 1.$$

All Hermite polynomials are normed and of n-th degree. We give here the first polynomials as example: $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$,... By the following lemma we want to present some of their elementary properties:

Lemma 4.1.2 (Properties of Hermite Polynomials)
The following properties hold for all Hermite polynomials $H_n(x)$, $n \geq 1$:
(i) $dH_n = nH_{n-1}$.
(ii) $\delta dH_n = nH_n$.
(iii) $(d + \delta)H_n = xH_n$.
(iv) $H_{n+1} = xH_n - nH_{n-1}$.
(v) $H_n(-x) = (-1)^n H_n(x)$.

Proof We will proof these statements by induction (for $H_1 = x$ they are trivial) or as simple consequences of already proven ones.

(i) We use the above stated identity $d\delta - \delta d = id_{\mathbb{R}[x]}$ and conclude by induction:

$$
\begin{aligned}
dH_n &= d\delta H_{n-1} = \delta d H_{n-1} + H_{n-1} \\
&= \delta(n-1)H_{n-2} + H_{n-1} = (n-1)H_{n-1} + H_{n-1} = nH_{n-1}.
\end{aligned}
$$

(ii) By (i) we get

$$\delta dH_n = \delta n H_{n-1} = nH_n.$$

(iii) By the explicit representation of δ we get

$$(\delta + d)H_n = ((-d + x) + d)\, H_n = xH_n.$$

(iv) Using identity (iii) we get

$$H_{n+1} = \delta H_n = xH_n - dH_n = xH_n - nH_{n-1}.$$

(v) This point is also proven by an induction argument using property (iv):

$$\begin{aligned} H_n(-x) &= (-x)H_{n-1}(-x) - (n-1)H_{n-2}(-x) \\ &= (-x)(-1)^{n-1}H_{n-1}(x) - (n-1)(-1)^{n-2}H_{n-2}(x) \\ &= (-1)^n \left(H_{n-1}(x) - (n-1)H_{n-2}\right) = (-1)^n H_n(x). \end{aligned}$$

■

Proposition 4.1.3 (Hermite Polynomials as ONB)
The set $\left\{\frac{1}{\sqrt{n!}}H_n\right\}_{n\geq 0}$ *forms an orthonormal basis of* $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$.

Proof In fact this means that we have to prove three different statements: (i) that $\left\{\frac{1}{\sqrt{n!}}H_n\right\}_{n\geq 0}$ forms an orthonormal set, (ii) that its elements are linearly independent and (iii) that they are dense in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$.

(i) We have to prove that $\left\langle \frac{H_k}{\sqrt{k!}}, \frac{H_l}{\sqrt{l!}} \right\rangle = \delta_{k,l}$ (where δ here symbolizes the Kronecker delta and should not be confused with our operator adjoint to d). Take without loss of generality $k > l$:

$$\left\langle \frac{H_k}{\sqrt{k!}}, \frac{H_l}{\sqrt{l!}} \right\rangle = \left\langle \frac{1}{\sqrt{k!}}(\delta)^k 1, \frac{1}{\sqrt{l!}}(\delta)^l 1 \right\rangle = \left\langle 1, \frac{1}{\sqrt{k!l!}}(d)^k(\delta)^l 1 \right\rangle = 0.$$

Since by iteration of point (i) of the lemma above $(d)^k H_k = k!$, we get in the case of equality

$$\left\langle \frac{H_k}{\sqrt{k!}}, \frac{H_k}{\sqrt{k!}} \right\rangle = \left\langle \frac{1}{\sqrt{k!}}(\delta)^k 1, \frac{H_k}{\sqrt{k!}} \right\rangle = \left\langle 1, \frac{1}{k!}(d)^k H_k \right\rangle = \langle 1, 1 \rangle = 1$$

as desired.

(ii) Linear independence is obvious by the fact that the n-th Hermite polynomial is a normed polynomial of order n.

(iii) By (ii) it is clear that the Hermite polynomials generate obviously the ring of polynomials over $\mathbb{R}$ which lies dense in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$. This can be proven indirectly: Assume that there exists a nonzero $g \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$ such that for all $k \geq 0$ with $\langle g, x^k \rangle = 0$. Its Fourier transform is

$$\widehat{g}(t) = \int_{\mathbb{R}} g(x) e^{itx - \frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

By the Cauchy inequality we can show that the kernel of this integral is dominated (setting $t = w + iz$)

$$\int_{\mathbb{R}} |g(x)| e^{-zx - \frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \leq \left(\int_{\mathbb{R}} |g(x)|^2 e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} e^{-2zx - \frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \right)^{\frac{1}{2}},$$

hence we can calculate derivatives of $\widehat{g}(x)$ directly by differentiating the kernel, whence (since the inner product vanishes by assumption) for $t = 0$ and all $k \in \mathbb{N}$

$$\widehat{g}^{(k)}(0) = i^k \int_{\mathbb{R}} x^k g(x) e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} = 0.$$

So we can conclude that $\widehat{g} \equiv 0$ and hence $g \equiv 0$ which contradicts the assumption.

■

Since bases are usually used for representations it is quite logical to ask how to expand smooth functions. This expansion will give us an easy way to determine the generating function of the Hermite polynomials.

Proposition 4.1.4 (Hermite Expansion and Generating Function)
(i) Given a C^∞-function f with all its derivatives in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$, then it can be expanded as

$$f(x) = \sum_{n=0}^{\infty} E\left(\frac{d^n f(x)}{dx^n}\right) \frac{H_n(x)}{n!}.$$

(ii) The generating function of the Hermite polynomials is given by

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

Proof
(i) As we have seen in the previous proposition, the Hermite polynomials $H_n(x)$ form a basis of $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$, so there have to exist c_i such that $f(x) = \sum_{n=0}^{\infty} c_n H_n(x)$. Multiplying this equation with $H_k(x)$ and integrating it (term by term) we get on the right hand side (since the Hermite polynomials are orthogonal) $c_k k!$ and can conclude that $c_k k! = E\left(H_k(x) f(x)\right) = \langle H_k(x), f(x)\rangle = \langle \delta^k 1, f(x)\rangle = \langle 1, d^k f(x)\rangle = E\left(\frac{d^k f(x)}{dx^k}\right)$ whence $c_k = \frac{1}{k!} E\left(\frac{d^k f(x)}{dx^k}\right)$ which leads to the desired result.
(ii) We have $\frac{d^k}{dx^k} e^{tx - \frac{t^2}{2}} = t^k e^{tx - \frac{t^2}{2}}$ and hence

$$E\left(\frac{d^k}{dx^k} e^{tx - \frac{t^2}{2}}\right) = \int_{\mathbb{R}} t^k e^{tx - \frac{t^2}{2}} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} = t^k \int_{\mathbb{R}} e^{-\frac{(t-x)^2}{2}} \frac{dx}{\sqrt{2\pi}} = t^k$$

which we can directly insert into the expansion of point (i). ■

The generating function also pave us the way to a nice closed expression for the Hermite polynomials (which could also be used as their definition). Splitting the generating function in two terms, we get by expanding the second factor in a Taylor series around zero

$$\begin{aligned} e^{tx - \frac{t^2}{2}} &= e^{\frac{x^2}{2}} e^{-\frac{(t-x)^2}{2}} = e^{\frac{x^2}{2}} \sum_{n=0}^{\infty} \frac{t^n}{n!} \frac{d^n}{dt^n} e^{-\frac{(t-x)^2}{2}} \Bigg|_{t=0} \\ &= e^{\frac{x^2}{2}} \sum_{n=0}^{\infty} \frac{t^n}{n!} (-1)^n \frac{d^n}{dx^n} e^{-\frac{x^2}{2}} = \sum_{n=0}^{\infty} \frac{t^n}{n!} (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}} \end{aligned}$$

and by a comparison of coefficients with 4.1.4(ii)

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

4.1.2 Wiener Chaos

Now we have to establish a relationship between the Hermite polynomials and Gaussian random variables which will be done by the following proposition.

Proposition 4.1.5 (Gaussian R.V.s and Hermite Polynomials)
Let X and Y be two jointly $\mathcal{N}(0,1)$-Gaussian r.v.s. For $m,n \geq 0$ the following property holds:

$$E\left(H_m(X)H_n(Y)\right) = \delta_{m,n} \cdot n!E(XY)^n.$$

Proof We start with the equation $E\left(e^{sX+tY}\right) = e^{E\left(\frac{(sX+tY)^2}{2}\right)}$ which e.g. can be derived from the fact that both sides satisfy the partial differential equation $\frac{\partial^2 f(x,y)}{\partial x \partial y} = stf(x,y)$. By simple transformations, using the fact that the variance of both r.v.s is one, we come to the relation

$$\begin{aligned} E\left(e^{sX-\frac{s^2}{2}}e^{tY-\frac{t^2}{2}}\right) &= e^{-\frac{s^2}{2}-\frac{t^2}{2}}E\left(e^{sX+tY}\right) = e^{-\frac{s^2}{2}-\frac{t^2}{2}}e^{E\left(\frac{(sX+tY)^2}{2}\right)} = \\ &= e^{-\frac{s^2}{2}-\frac{t^2}{2}}e^{\frac{s^2}{2}E(X^2)}e^{\frac{t^2}{2}E(Y^2)}e^{stE(XY)} = e^{stE(XY)}. \end{aligned}$$

We will take on both sides the derivative by the operator $\frac{\partial^{n+m}}{\partial s^n \partial t^m}$ and evaluate it at $s = t = 0$. This gives on the left hand side by 4.1.4(ii) $E\left(H_m(X)H_n(Y)\right)$; on the right hand side the expression is for $n \neq m$ zero, for $n = m$ it gives $n!\left(E(XY)\right)^n$ which yields the required expression. ■

Coming from Hermite polynomials and having now pointed out the connection to Gaussian r.v.s it seems rather logical that we will now focus on the Gaussian properties. The notion of the Gaussian Space, introduced in the next definition, will be the central founding concept of this chapter.

Definition 4.1.6 (Gaussian Space)
A probability space $(\Omega, \mathcal{F}, P)$ is called Gaussian iff there exists a closed subspace $\mathcal{H} \subset L^2(\Omega, \mathcal{F}, P)$ of $\mathcal{N}(0,\sigma)$-distributed (therefore Gaussian) r.v.s such that $\sigma(X : X \in \mathcal{H}) = \mathcal{F}$.

This seems quite familiar to us, looking back to the proof of the existence of the Wiener process (Theorem 2.2.2). There the Hilbert space isomorphism η played a central role, so we will do the same here in the more abstract and general setting.

Definition 4.1.7 (Isonormal Gaussian Process)
Given a real separable Hilbert space H, we say that an Hilbert space isometry $\eta : H \to \mathcal{H} \subset L^2(\Omega, \mathcal{F}, P)$ with

$$E\left(\eta(h)\eta(g)\right) = \langle h, g \rangle$$

for all $g, h \in H$ is called an isonormal Gaussian process.

As a direct consequence of the definition we can remark that η is linear; for all $g, h \in H$:

$$E\left((\eta(\alpha g+\beta h)-\alpha\eta(g)-\beta\eta(h))^2\right) = \|\alpha g+\beta h\|^2+\alpha^2\|g\|^2+$$
$$+\beta^2\|h\|^2-2\alpha\langle\alpha g+\beta h,g\rangle-2\beta\langle\alpha g+\beta h,h\rangle+2\alpha\beta\langle g,h\rangle$$
$$= \|\alpha g+\beta h\|^2-\alpha^2\|g\|^2-\beta^2\|h\|^2-2\alpha\beta\langle g,h\rangle=0.$$

Isonormal Gaussian processes are something quite common since Kolmogorov's extension theorem (see [Tei 03], p.4) allows us to construct to a given Hilbert space H an isonormal process $\eta : H \to L^2(\Omega, \mathcal{F}, P)$. In order not to create useless notations in abundance we will always assume that the space structures are compatible, i.e. $\eta(H) = \mathcal{H}$, so we can also see in the process something like a coordinate chart map which pushes the structure of the Hilbert space forward to $\mathcal{H}$.

Proposition 4.1.8 (Properties of the Hilbert Space Isometry)
The exponential of the Hilbert space isometry spans a dense subset of the space of square-integrable, real-valued functions on the probability space Ω:

$$\overline{\left\langle\left\{e^{\eta(h)} : h \in H\right\}\right\rangle} = L^2(\Omega, \mathcal{F}, P).$$

Proof We assume indirectly that there exists an $X \in L^2(\Omega,\mathcal{F},P)$, $X \neq 0$ with $E\left(Xe^{\eta(h)}\right) = 0$. Taking $\{h_i\}$ as a basis of H we define $X_i := \eta(h_i)$, then for all $A \in \mathcal{F}$ there exist $t_i \in \mathbb{R}$ with $\sum_{i=1}^{\infty} t_i X_i = 1_A$. Defining $\mathcal{F}_n := \sigma(X_1, ..., X_n)$ for $n \geq 1$ we get

$$E\left(E\left(X|\mathcal{F}_n\right)e^{\sum_{i=1}^{n} X_i}\right) = 0. \tag{4.2}$$

Decomposing $\xi^n := E\left(X|\mathcal{F}_n\right)$ into its positive (ξ^n_+) and (absolutely valued) negative (ξ^n_-) part $\xi^n = \xi^n_+ - \xi^n_-$ we can - since $E(X) = E(Xe^0) = 0$ - normalize them such that $E(\xi^n_+) = E(\xi^n_-) = 1$. So we can derive from (4.2)

$$E\left(\xi^n_+ e^{\sum_{i=1}^{n} X_i}\right) = E\left(\xi^n_- e^{\sum_{i=1}^{n} X_i}\right)$$

implying the equality of the characteristic functions of the X_i with respect to the probability measures $\xi^n_+ dP$ and $\xi^n_- dP$. So for $n \to \infty$ we have $\xi^n \to X$ implying $X_+ = X_-$ and hence $X = 0$, contradicting the assumption. ■

Having now $\eta(h)$ as r.v. one surely can ask in the spirit of Proposition 4.1.5 what happens if we apply the Hermite polynomials to it. So we can define the following subspace:

Definition 4.1.9 (Wiener Chaos)
To a given Hilbert space isometry η we define for $n \geq 1$ the n-th Wiener chaos $\mathcal{H}_n$ as a closed linear subspace of $L^2(\Omega, \mathcal{F}, P)$ by

$$\mathcal{H}_n := \overline{\langle\{H_n\left(\eta(h)\right) : h \in H, \|h\| = 1\}\rangle}.$$

Additionally we define $\mathcal{H}_0$ as the set of constants.

This leads us to ask in which relation these different subspaces - now called chaoses - stand. The answer is quite simple but important: they are mutually orthogonal and their direct sum spans the whole space.

Theorem 4.1.10 (Chaos Decomposition)
Given a Gaussian probability space $(\Omega, \mathcal{F}, P)$, a Hilbert space H and an isonormal Gaussian process η, then

(i) for $n \neq m$ the respective Wiener chaos are orthogonal $\mathcal{H}_n \perp \mathcal{H}_m$,

(ii) the space of square integrable r.v.s can be decomposed as direct sum of Wiener chaoses

$$L^2(\Omega, \mathcal{F}, P) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n.$$

Proof

(i) This is a direct consequence of proposition 4.1.5. and the definition of the chaoses.

(ii) It only remains to show that the direct sum really spans the whole space. So assume that there would exist a nonzero $X \in L^2(\Omega, \mathcal{F}, P)$ with $X \perp \mathcal{H}_n$ for all $n \in \mathbb{N}$. This means that for all $h \in H$ with $\|h\| = 1$ we have $E\left(XH_n(\eta(h))\right) = 0$. But since x^n can be generated uniquely by Hermite polynomials of degree less or equal than n we have also $E\left(X(\eta(h))^n\right) = 0$ and by simply summing up a series $E\left(Xe^{\eta(h)}\right) = 0$. But by proposition 4.1.8 it follows that $X \equiv 0$ which contradicts our assumption. ■

Now one can ask if there is something as a basis of $L^2(\Omega, \mathcal{F}, P)$ which is compatible with the chaos decomposition or, with other words, if there exists a general representation of the basis of $\mathcal{H}_n$. To give an affirmative answer to this question we have to introduce multiindices. For

$$c_0 := \{(x_1, x_2, ..., x_n, 0, 0, 0, ...) : 0 \leq x_i < \infty, 1 \leq n < \infty\},$$

the set of all sequences of non-negative integers with only finitely many unequal zero we define for the multiindices $a \in c_0$ a norm by $|a| := \sum\limits_{i=1}^{\infty} a_i$ and a faculty by $a! := \prod\limits_{i=1}^{\infty} a_i!$. Furthermore we can define for $x \in \mathbb{R}^{\mathbb{N}}$ a generalization of Hermite polynomials by

$$H_a(x) := \prod_{i=1}^{\infty} H_{a_i}(x_i)$$

which is, in fact, only a finite product since $H_0 \equiv 1$. As a generalization of the Gaussian measure ν to $\mathbb{R}^{\mathbb{N}}$ (with the Borel-σ-algebra $\mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right)$) we have $\nu^{\otimes \mathbb{N}}$ as an inductive limit, hence the projections $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^n$ give $(\pi_n)_* \nu^{\otimes \mathbb{N}} = \nu^{\otimes n} := \prod\limits_{i=1}^{n} e^{-\frac{x_i^2}{2}} \frac{dx_i}{\sqrt{2\pi}}$ for all $n \in \mathbb{N}$. Easily we can now generalize Proposition 4.1.5 in this setting to get

$$E\left(\prod_{i=1}^{\infty} H_{a_i}(\eta(g_i)) H_{b_i}(\eta(h_i))\right) = \delta_{a,b} \cdot a! E\left(\prod_{i=1}^{\infty} \eta(g_i)\eta(h_i)\right)^{|a|} \tag{4.3}$$

for g_i, $h_i \in H$.

We define the function Φ_a by

$$\Phi_a(h_i) := \frac{1}{\sqrt{a!}} \prod_{i=1}^{\infty} H_{a_i}(\eta(h_i))$$

to give an explicit representation for the basis of the Wiener chaos:

Theorem 4.1.11 (Basis of $\mathcal{H}_n$)
For every $n \geq 0$ the set $\{\Phi_a(e_i) : a \in c_0, |a| = n\}$, $\{e_i\}$ an orthonormal basis of H, is an orthonormal basis of $\mathcal{H}_n$.

Proof By (4.3) it is clear that $\{\Phi_a(e_i) : a \in c_0, |a| = n\}$ is an orthonormal system in $\mathcal{H}_n$, it remains to prove that it is also dense in it. Since e_i is a basis of H, all polynomials of $\eta(h_i)$ can also be written as polynomials of $\eta(e_i)$, hence the union of all Φ_a's for m's of all different orders spans the whole space,

$$\bigcup_{m=0}^{\infty} \overline{\{\Phi_a(e_i) : a \in c_0, |a| = m\}} = \bigotimes_{m=0}^{\infty} \mathcal{H}_m.$$

But since clearly on the one hand side for $|b| = k \neq n$ we have

$$\{\Phi_b(e_i) : b \in c_0, |b| = k\} \perp \mathcal{H}_n,$$

implying that there are no other elements then the Φ_a for $|a| = n$ in the union which are in $\mathcal{H}_n$, and on the other hand side $\mathcal{H}_n \subset \bigotimes_{m=0}^{\infty} \mathcal{H}_m$, the density follows. ■

As a consequence we see that $\{\Phi_a : a \in c_0\}$ forms an orthonormal basis of $L^2(\Omega, \mathcal{F}, P)$ and hence, since

$$L^2\left(\mathbb{R}^n, \mathcal{B}\left(\mathbb{R}^n\right), \nu^{\otimes n}\right) \simeq \bigotimes_{i=1}^{n} L^2\left(\mathbb{R}, \mathcal{B}\left(\mathbb{R}\right), \nu\right),$$

Proposition 4.1.3 implies

$$L^2\left(\Omega, \mathcal{F}, P\right) = \bigotimes_{i=1}^{\infty} \mathcal{H}_i \simeq L^2\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right), \nu^{\otimes \mathbb{N}}\right).$$

4.2 Multiple Wiener-Itô Integrals

In the spirit of the Wiener chaos and based on the notion of the tensor product introduced in the first chapter we will now look for a different interpretation of the stochastic integral.

4.2.1 Multiple Wiener-Itô Integrals

To the n-fold tensor product $H^{\otimes n} := H \otimes ... \otimes H$ of a given Hilbert space H we have the permutation group $\mathfrak{S}_n$ whose elements $\pi \in \mathfrak{S}_n$ act on $H^{\otimes n}$ by $\pi(e_i \otimes ... \otimes e_n) = e_{\pi(1)} \otimes ... \otimes e_{\pi(n)}$. A function f is called symmetric if $\pi(f) = f$. We define the symmetrization $\tilde{g}$ of g by $\tilde{g} := \sum_{\pi \in \mathfrak{S}_n} \frac{\pi(g)}{n!}$ and for $0 \leq r \leq (m \wedge n)$ we define the r-fold right contraction as bilinear mapping $H^{\otimes m} \times H^{\otimes n} \rightarrow H^{\otimes (m+n-2r)}$ by

$$(e_1 \otimes ... \otimes e_m) \otimes_r (f_1 \otimes ... \otimes f_n) := e_1 \otimes ... \otimes e_{m-r} \otimes f_1 \otimes ... \otimes f_{n-r} \prod_{i=1}^{r} \langle e_{m-r+i}, f_{n-r+i} \rangle$$

These definitions give us the required foundations for the multiple Wiener-Itô integrals.

Definition 4.2.1 (Multiple Wiener-Itô Integrals)
Given a real, separable Hilbert space H and an isonormal Gaussian process η to $L^2(\Omega, \mathcal{F}, P)$, we define for $n \geq 0$ the multiple Wiener-Itô integral as a continuous linear mapping $I_n : H^{\otimes n} \rightarrow L^2(\Omega, \mathcal{F}, P)$ satisfying the following recursion for $f \in H^{\otimes n}$ and $g \in H$

$$\begin{cases} I_0(\lambda) = \lambda 1_\Omega \text{ for constant } \lambda \in \mathbb{R} \\ I_1(g) = \eta(g) \\ I_{n+1}(f \otimes g) = I_n(f) I_1(g) - n I_{n-1}(\tilde{f} \otimes_1 g) \end{cases}$$

Theorem 4.2.2 (Existence of Multiple Wiener-Itô Integrals)
For given H and η, the Wiener-Itô integrals are unique for every $n \geq 0$ and it holds for an orthonormal basis $\{e_i\}_{i \geq 1}$ of H and non-negative integers n_j satisfying $\sum_{j=1}^{m} n_j = n$ for a $m < \infty$ that

$$I_n \left(\bigotimes_{j=1}^{m} e_j^{\otimes n_j} \right) = \prod_{j=1}^{m} H_{n_j} \left(\eta(e_j) \right) \tag{4.4}$$

and

$$I_n(f) = I_n(\tilde{f}). \tag{4.5}$$

Proof The main goal is to prove equation (4.4), which is done by mathematical induction:

(i) We have

$$\begin{aligned} I_0(\pi(\lambda)) &= \lambda 1_\Omega \\ I_1(\pi(e_i)) &= \eta(e_i) = H_1(\eta(e_i)) = I_1(e_i), \end{aligned}$$

so (4.4) and (4.5) hold for $n = 0$ and $n = 1$.

(ii) Assume that (4.4) is already proven for all $p < n$, i.e.

$$I_p \left(\bigotimes_{j=1}^{m} e_j^{\otimes p_j} \right) = \prod_{j=1}^{m} H_{p_j} \left(\eta(e_j) \right)$$

holds. Then we have by the commutativity of the product

$$I_p\left(\pi\left(\bigotimes_{j=1}^{m} e_j^{\otimes p_j}\right)\right) = \prod_{j=1}^{m} H_{p_j}\left(\eta(e_j)\right)$$

for e_j elements of the basis $\{e_i\}$ and hence $I_p(f) = I_p(\tilde{f})$ for any $f \in H^{\otimes p}$

(iii) Now we have to make the proper induction step. Under the assumption that (4.4) holds for all $p < n$ we want to show that it holds for n. Therefore we introduce the following notations: Let $f = f_1 \otimes ... \otimes f_n$ the n-fold tensor product of (not necessarily distinct) basis elements $f_j \in \{e_i\}_{i\geq 1}$. We introduce non-negative integers m_j satisfying $\sum_{j=1}^{m} m_j = n-1$ and define

$$\begin{cases} \tilde{m}_j & := & m_j \text{ for } e_j \neq f_n \\ \tilde{m}_j & := & m_j - 1 \text{ for } e_j = f_n \end{cases}$$

Furthermore it is clear that if k elements of $\{f_1, ...f_{n-1}\}$ are equal to f_n, then there exist $(n-2)!k$ permutations $\pi(f_1 \otimes ... \otimes f_{n-1})$ where the last element of the tensor product equals f_n. By the definition of the first right contraction we get hence

$$\begin{aligned} I_n(f) &= I_{n-1}(f_1 \otimes ... \otimes f_{n-1}) - (n-1)I_{n-2}(\widetilde{f_1 \otimes ... \otimes f_{n-1}} \otimes_1 f_n) \\ &= \left(\prod_{j=1}^{m} H_{m_j}(\eta(e_j))\right) I_1(f_n) - (n-1)I_{n-2}\left(\frac{\sum_{\pi}(f_1 \otimes ... \otimes f_{n-1})}{(n-1)!} \otimes_1 f_n\right) \\ &= \left(\prod_{j=1}^{m} H_{m_j}(\eta(e_j))\right) I_1(f_n) - \frac{(n-1)(n-2)!k}{(n-1)!} I_{n-2}\left(\bigotimes_{j=1}^{m} e_j^{\tilde{m}_j}\right) \\ &= \left(\prod_{j=1}^{m} H_{m_j}(\eta(e_j))\right) \eta(f_n) - k\left(\prod_{j=1}^{m} H_{\tilde{m}_j}(\eta(e_j))\right) \\ &= \left(\prod_{j=1, e_j \neq f_n}^{m} H_{m_j}(\eta(e_j))\right) \left(H_k(\eta(f_n))\eta(f_n) - kH_{k-1}(\eta(f_n))\right) \\ &= \left(\prod_{j=1, e_j \neq f_n}^{m} H_{m_j}(\eta(e_j))\right) \left(H_{k+1}(\eta(f_n))\right) = \prod_{j=1}^{m} H_{n_j}(\eta(e_j)) \end{aligned}$$

by property 4.1.2(iv) of Hermite polynomials which proves (4.4) and so by (ii) also (4.5).

(iv) It only remains to show that the definition of the I_n is unique and since they are given by a recursion formula it is enough to show that they are bounded in $L^2(\Omega, \mathcal{F}, P)$. This is done for basis elements by (4.3) and the isometry property of the Gaussian process:

$$E\left(\left(I_n\left(\bigotimes_{j=1}^{m} e_j^{\otimes n_j}\right)\right)^2\right) = E\left(\left(\prod_{j=1}^{m} H_{n_j}\left(\eta(e_j)\right)\right)^2\right) = n! < \infty.$$

■

The multiple Wiener-Itô integrals enjoy the following properties:

Proposition 4.2.3 (Properties of the Wiener-Itô Integral)
For the mapping $I_n : H^{\otimes n} \to L^2(\Omega, \mathcal{F}, P)$, $n \geq 0$ *it holds that*
(i) $I_n\left(H^{\otimes n}\right) = \mathcal{H}_n$
(ii) $\langle I_n(f), I_m(g)\rangle = \delta_{n,m} n! \left\langle \tilde{f}, \tilde{g} \right\rangle$ *for* $f \in H^{\otimes n}$ *and* $g \in H^{\otimes m}$.

Proof (i) Looking to the n_j of (4.4) as entries of the multiindex n we get by definition of the Φ_a for an orthonormal basis $\{e_i\}$ of H

$$I_n\left(\bigotimes_{j=1}^m e_j^{\otimes n_j}\right) = \prod_{j=1}^m H_{n_j}(\eta(e_j)) = \sqrt{n!}\Phi_n(e_i)$$

implying, since by Theorem 4.1.11 the Φ_a form an orthonormal basis of $\mathcal{H}_{|a|}$, that the multiple Wiener-Itô integrals I_n lie dense in $\mathcal{H}_n$.

(ii) It is clear that the inner product is zero for $n \neq m$ since the Wiener-Itô integrals lie in different chaoses orthogonal to each other.
For $n = m$ we have for an $f \in H^{\otimes n}$ with $\tilde{f} \not\equiv 0$ the relation $\tilde{f} = \bigotimes_{j=1}^m \left(f^{\frac{1}{n}} e_j\right)^{\otimes n_j}$ and hence

$$I_n(\tilde{f}) = \prod_{j=1}^m H_{n_j}\left(\eta\left(f^{\frac{1}{n}} e_j\right)\right) = \|f\| I_n\left(\bigotimes_{j=1}^m e_j^{\otimes n_j}\right).$$

So we can conclude by step (iv) of the proof of the previous theorem that

$$E\left((I_n(f))^2\right) = E\left(\left(I_n(\tilde{f})\right)^2\right) = E\left(\left(\|f\| I_n\left(\bigotimes_{j=1}^m e_j^{\otimes n_j}\right)\right)^2\right) = \|f\|^2 n!. \tag{4.6}$$

Polarization

$$\begin{aligned}
\langle I_n(f), I_m(g)\rangle &= \left\langle I_n(\tilde{f}), I_m(\tilde{g})\right\rangle = \frac{1}{4} E\left(\left(I_n(\tilde{f}+\tilde{g}\right)^2 - \left(I_n(\tilde{f}-\tilde{g}\right)^2\right) \\
&= \frac{1}{4} E\left(\|\tilde{f}+\tilde{g}\ \|^2 n! - \|\tilde{f}-\tilde{g}\ \|^2 n!\right) \\
&= n! \frac{\|\tilde{f}+\tilde{g}\ \|^2 - \|\tilde{f}-\tilde{g}\ \|^2}{4} = \delta_{n,m} n! \left\langle \tilde{f}, \tilde{g} \right\rangle
\end{aligned}$$

yields the result. ■

Theorem 4.2.4 (Chaos Expansion)
For $(\Omega, \mathcal{F}, P)$ *a Gaussian probability space and* H *a real, separable Hilbert space it holds that*
(i) the Wiener-Itô integrals I_n *are injective on the closed subspace of symmetric tensors* $\tilde{f}$.

(ii) every $F \in L^2(\Omega, \mathcal{F}, P)$ *has a unique decomposition* $F = \sum_{i=0}^{\infty} I_n(\tilde{f}_n)$ *for symmetric tensors* $\tilde{f}_n \in H^{\otimes n}$*, the so-called Wiener chaos expansion.*

Proof (i) Injectivity is clear since we have the following implication chain for symmetric tensors by the above proof:

$$\begin{aligned} I_n(\tilde{f}) &= I_n(\tilde{g}) \\ \Longrightarrow \quad \|\tilde{f}\| I_n\left(\bigotimes_{j=1}^{m} e_j^{\otimes n_j}\right) &= \|\tilde{g}\| I_n\left(\bigotimes_{j=1}^{m} e_j^{\otimes n_j}\right) \\ \Longrightarrow \quad \|\tilde{f}\| &= \|\tilde{g}\| \\ \Longrightarrow \quad \tilde{f} &= \tilde{g}. \end{aligned}$$

(ii) By proposition 4.2.3(i) it is clear that there exists a decomposition $F = \sum_{i=0}^{\infty} I_n(g_n)$, by (4.5) we can take $\tilde{g}_n$ instead of g_n to get a symmetric decomposition which is, by point (i) of this proof, unique. ∎

As last point we note here that if the Hilbert space H is infinite dimensional, the recursive definition of I_n yields for orthogonal $h_1, ...h_n$ - since the first right contraction vanishes -

$$I_n(h_1 \otimes ... \otimes h_n) = I_{n-1}(h_1 \otimes ... \otimes h_{n-1}) I_1(h_n)$$

whence

$$I_n\left(\bigotimes_{i=1}^{n} h_i\right) = \prod_{i=1}^{n} \eta(h_i). \tag{4.7}$$

Without proof we remark that the relation (4.6) for arbitrary orthogonal h_i suffices to define multiple Wiener-Itô integrals on infinite dimensional Hilbert spaces, so they are fully characterized by the off-diagonal elements, also for the elements on some diagonal (i.e. $h_k = ch_l$ for some $1 \leq k, l \leq n$).

4.2.2 Wiener-Itô Integrals as Stochastic Integrals

The aim of this chapter is to point out the connection between the stochastic (Itô) integrals and the multiple Wiener-Itô integrals (which is already suggested by their names). We assume here that the separable real Hilbert space H has a concrete representation $L^2(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}), \mu)$ for a σ-finite measure μ without atoms and hence $H^{\otimes n} \simeq L^2(\mathbb{R}^n_{\geq 0}, \mathcal{B}(\mathbb{R}^n_{\geq 0}), \mu)$; the process η is given by $\eta(h) := \int_0^{\infty} h(s) dB_s$.

First we prove the existence of iterated stochastic integrals:

Lemma 4.2.5 (Existence of the Iterated Integral)
Given $\tilde{f} \in H^{\otimes n}$ *a symmetric tensor and* $0 \leq t_1 \leq t_2 ... \leq t_n$*, then there exists the iterated stochastic integral*

$$\int_0^{\infty} \int_0^{t_n} ... \int_0^{t_2} \tilde{f}_n(t_1, ...t_n) dB_{t_1} ... dB_{t_{n-1}} dB_{t_n}$$

a.s. unique.

Proof For fixed n we have to prove that the processes

$$V^k(t_{k+1},...,t_n) := \int_0^{t_{k+1}} ... \int_0^{t_2} \tilde{f}_n(t_1,...t_n)dB_{t_1}...dB_{t_k}$$

exist for almost all $t_{k+1},...t_n$ and that they could be used as integrands in t_{k+2}, i.e. that they have a progressively measurable, square integrable modification.

First we note that square integrability is given by repeated use of the Itô-Lemma

$$E\left(\left(V^k(t_{k+1},...,t_n)\right)^2\right) = \int_0^{t_{k+1}} ... \int_0^{t_2} \left(\tilde{f}_n(t_1,...t_n)\right)^2 dt_1...dt_k. \tag{4.8}$$

The rest is done by induction; we note that for $k = 0$ stochastic integrability is clearly given since $V^0(t_{k+1},...,t_n) = \tilde{f}_n(t_1,...t_n)$. Suppose now that for $V^{k-1}(t_{k+1},...,t_n)$ there exists a progressively measurable modification, then the process

$$\int_0^t V^{k-1}(t_k,...,t_n)dB_{t_k}$$

has for fixed $t_{k+1},...t_n$ a modification with a.s continuous paths (by Corollary 2.3.7) which is adapted (by Theorem 2.3.6), whence progressively measurable. ∎

Theorem 4.2.6 (Wiener-Itô Integrals as Stochastic Integrals)
Under the above assumptions it holds for symmetric $\tilde{f}_n \in H^{\otimes n}$ *that*

$$I_n(\tilde{f}_n) = n! \int_0^{\infty} \int_0^{t_n} ... \int_0^{t_2} \tilde{f}_n(t_1,...t_n)dB_{t_1}...dB_{t_{n-1}}dB_{t_n} \quad a.s.$$

Proof For elementary step processes the equality holds since on the one hand side

$$E\left(\left(I_n(\tilde{f}_n)\right)^2\right) = n!\|\tilde{f}_n\|^2$$

by (4.6) and on the other hand side the mean square of the iterated integral is $\|\tilde{f}_n\|^2$ by (4.8). By a classical density argument as in Theorem 2.3.3 this extends to square integrable, progressively measurable (symmetric) functions. The uniqueness of the multiple Wiener-Itô integrals concludes the proof. ∎

Hence the n-th iterated stochastic integral lies in the n-th Wiener chaos, integrating can be seen as climbing up the chaoses. This forces the crucial question: But how climb down again?

4.3 Malliavin Derivatives

Exactly this climbing down will be the subject of this chapter. But first we have to bother a little bit with notations. For smooth functions $f : \mathbb{R}^n \to \mathbb{R}^n$ we have the following inclusion relations

$$C_0^\infty(\mathbb{R}^n) \subset C_b^\infty(\mathbb{R}^n) \subset C_p^\infty(\mathbb{R}^n)$$

of classes of compact supported functions, bounded functions with bounded derivatives and functions with derivatives of polynomial growth. Let as in the previous chapter η be an isonormal Gaussian process and H a separable real Hilbert space (with - if needed - concrete representation $L^2(T, \mathcal{B}, \mu)$, μ atomless and σ-finite). Then for random variables $F : L^2(\Omega, \mathcal{F}, P) \to \mathbb{R}^n$, $F = f(\eta(h_1), ..., \eta(h_n))$, $h_i \in H$ orthonormal, we will use analogously the notations

$$\mathcal{S}_0 \subset \mathcal{S}_b \subset \mathcal{S}_p$$

for the different classes of smooth r.v.s with $f \in C_0^\infty(\mathbb{R}^n)$ (resp. $C_b^\infty(\mathbb{R}^n)$, $C_p^\infty(\mathbb{R}^n)$), especially we will understand under a "smooth random variable" a r.v. $F \in \mathcal{S}_p$. Since we can approximate polynomials in $\eta(h_1), ..., \eta(h_n)$ uniformly on compacts by $\mathcal{S}_0$-functions it is clear that $\mathcal{S}_0$ is dense in $L^2(\Omega, \mathcal{F}, P)$.

Definition 4.3.1 (Malliavin Derivative)
Let $F = f(\eta(h_1), ..., \eta(h_n)) \in \mathcal{S}_p$, then we understand under the Malliavin derivative of F the mapping $DF \in L^2(\Omega, \mathcal{F}, P) \otimes H$ given by

$$DF := \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\eta(h_1), ..., \eta(h_n)) h_i$$

.

For the concrete representation of the Hilbert space we get by

$$L^2(\Omega, \mathcal{F}, P) \otimes L^2(T, \mathcal{B}, \mu) \simeq L^2(\Omega \times T, \mathcal{F} \otimes \mathcal{B}, P \otimes \mu)$$

the Malliavin derivative $D_t F$ parametrized by $t \in T$

$$D_t F = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\eta(h_1), ..., \eta(h_n)) h_i(t).$$

The sole example $D_t \eta(h) = h(t)$ indicates that this mapping has to be something like an inverse of our Wiener-Itô integrals I_n. Using the Hilbert space structure-we can promote the concept of the directional derivative with respect to $h \in H$ since

$$\begin{aligned} \langle DF, h \rangle_H &= \langle DF, id \otimes h \rangle_H \\ &= \left(\frac{d}{dt} f(\eta(h_1) + t \langle h_1, h \rangle_H, ..., \eta(h_n) + t \langle h_n, h \rangle_H) \right) \Big|_{t=0} \end{aligned}$$

To understand the concrete meaning of this differentiation we give the following example:

Example 4.3.2 (Malliavin Differentiation)
Let $\eta(h) = \int_0^t h(s)dB_s$*, then for*

$$F = \left(\int_0^t 1_{[0,s]} dB_s \right)^2$$

we have

$$DF = 2 \int_0^t 1_{[0,s]} dB_s 1_{[0,t]}$$

which is obviously not $\{\mathcal{F}_s\}$*-adapted (but* $\mathcal{F}_t$*-measurable).*

Proposition 4.3.3 (Independence of the Choice of the Basis)
The Malliavin derivative DF *is independent of the representation of* F*.*

Proof Let

$$\begin{aligned} F &= f(\eta(h_1), ..., \eta(h_n)) \\ &= g(\eta(k_1), ..., \eta(k_m)) \end{aligned}$$

two different representations of F. Then we can choose in $\langle h_1, ..., h_n, k_1, ..., k_m \rangle$ a basis $e_1, ...e_l$ and get

$$\begin{aligned} h &= (h_1, ..., h_n)^\tau = A(e_1, ..., e_l)^\tau \\ k &= (k_1, ..., k_m)^\tau = B(e_1, ..., e_l)^\tau \end{aligned}$$

for matrices A, B. The composite functions $(f \circ A)(\eta(e_1), ..., \eta(e_l))$ and $(g \circ B)(\eta(e_1), ..., \eta(e_l))$ are equal by linearity of η and since

$$\int_{\mathbb{R}^l} h(x)(g \circ B)(\eta(e_1), ..., \eta(e_l))\nu_l(dx) = \int_{\mathbb{R}^l} h(x)(f \circ A)(\eta(e_1), ..., \eta(e_l))\nu_l(dx)$$

for every $h \in C_0^\infty(\mathbb{R}^l)$.
For the Malliavin derivative we have then

$$\begin{aligned} DF &= \sum_{i=1}^n \frac{\partial}{\partial x_i} f(\eta(h_1), ..., \eta(h_n))h_i \\ &= \sum_{j=1}^l \frac{\partial}{\partial y_i} (f \circ A)(\eta(e_1), ..., \eta(e_l))e_j \\ &= \sum_{j=1}^l \frac{\partial}{\partial y_i} (g \circ B)(\eta(e_1), ..., \eta(e_l))e_j \\ &= \sum_{r=1}^m \frac{\partial}{\partial z_r} g(\eta(k_1), ..., \eta(k_m))k_r. \end{aligned}$$

■

Our first important result will be on partial integration:

Lemma 4.3.4
For $F \in \mathcal{S}_p$ and $h \in H$ it holds with respect to the Gaussian measure ν that

$$E\left(\langle DF, h\rangle_H\right) = E\left(F(\eta(h))\right).$$

Proof It is enough to prove the result for H with $\|h\| = 1$ since otherwise we could normalize. We can choose orthonormal elements e_i such that $e_1 = h$ and $F = f(\eta(e_1), ..., \eta(e_n))$ by a change of bases. Now we can conclude by partial integration as in (4.1)

$$\begin{aligned} E\left(\langle DF, h\rangle_H\right) &= \int_{\mathbb{R}^n} \frac{\partial}{\partial x_1} f(x)\nu_n(dx) = \int_{\mathbb{R}^n} f(x)x_1\nu_n(dx) = E\left(F(\eta(e_1))\right) \\ &= E\left(F(\eta(h))\right). \end{aligned}$$

■

As a simple consequence we get the proper integration by parts formula:

Proposition 4.3.5 (Partial Integration)
For F, $G \in \mathcal{S}_p$, $h \in H$ it holds that

$$E\left(G\langle DF, h\rangle_H\right) + E\left(F\langle DG, h\rangle_H\right) = E\left(FG(\eta(h))\right).$$

Proof Since F, $G \in \mathcal{S}_p$ and by the product rule of differentiation we have

$$\begin{aligned} E\left(FG(\eta(h))\right) &= E\left(\langle D(FG), h\rangle_H\right) = E\left(\langle (DF)G + F(DG), h\rangle_H\right) \\ &= E\left(G\langle DF, h\rangle_H\right) + E\left(F\langle DG, h\rangle_H\right) \end{aligned}$$

■

Corollary 4.3.6 (Closeability)
The operator D is closed in the L^p-sense as

$$D : L^p(\Omega, \mathcal{F}, P) \supset \mathcal{S}_p \longrightarrow L^p(\Omega, \mathcal{F}, P) \otimes H.$$

Proof Since D is a linear operator it is closed, iff $F_n \to 0$ implies $DF_n \to 0$. This is clear by the above lemma: Suppose

$$\begin{aligned} F_n \to 0 &\quad \text{in} \quad L^p(\Omega, \mathcal{F}, P) \\ DF_n \to G &\quad \text{in} \quad L^p(\Omega, \mathcal{F}, P) \otimes H \end{aligned}$$

then

$$\begin{aligned} E(F\langle G, h\rangle) &= \lim_{n\to\infty} E(F\langle DF_n, h\rangle) \\ &= \lim_{n\to\infty} E(FF_n\eta(h)) - \lim_{n\to\infty} E(F_n\langle DF, h\rangle) = 0 \end{aligned}$$

for all $F \in \mathcal{S}_p$ with $F\eta(h)$ bounded. But these are dense in $L^p(\Omega, \mathcal{F}, P)$ and so $G \equiv 0$, yielding the closeability of D. ■

This now allows us to endow our space $\mathcal{S}_p$ with a topological vector space structure (by an abuse of notation we call D the Malliavin derivative again, tampering slightly with the domain).

Definition 4.3.7 (Operator Norm)
For $F \in \mathcal{S}_p$ and $p \geq 0$ we define the operator norm on $\mathcal{S}_p$ by

$$\|F\|_{1,p} := \left(E\left(|F|^p\right) + E\left(\|DF\|_H^p\right)\right)^{\frac{1}{p}}.$$

Since D is closeable this norm defines a Banach space $\mathbb{D}^{1,p}(\mathbb{R}^n)$ which we can continuously embed

$$\mathbb{D}^{1,p} \hookrightarrow L^p(\Omega, \mathcal{F}, P),$$

the image of this embedding is the maximal domain of D in L^p. $\mathbb{D}^{1,2}$ is obviously a Hilbert space with the inner product

$$\langle F, G \rangle_{1,2} = E(FG) + E(\langle DF, DG \rangle_H).$$

To define higher derivatives we remark that $DF \in L^p(\Omega, \mathcal{F}, P) \otimes H$. Hence we have to define the Malliavin derivatives on spaces

$$\mathcal{S}_p \otimes V \subset L^p(\Omega, \mathcal{F}, P) \otimes V$$

where V is another (real, separable) Hilbert space. This is done - since an embedding in a bigger space virtually changes nothing - by setting it $D \otimes id$, or more exact: For $F = f \otimes v \in \mathcal{S}_p \otimes v$ we define

$$DF = D(f \otimes v) := (DF) \otimes V.$$

So we can define higher Malliavin derivatives by iteration: Having $D^{k-1}F$ for $F \in L^p(\Omega, \mathcal{F}, P) \otimes V$ already defined, we set $D^k F := D(D^{k-1}F)$. The operators

$$D^k : L^p(\Omega, \mathcal{F}, P) \otimes V \to L^p(\Omega, \mathcal{F}, P) \otimes V \otimes H^{\otimes k}$$

are again closeable; with the family of operator norms

$$\|F\|_{k,p} := \left(E\left(|F|^p\right) + \sum_{j=1}^{k} E\left(\|D^j F\|_{V \otimes H^{\otimes j}}^p\right)\right)^{\frac{1}{p}}$$

for $k, p \geq 1$ we get the Banach spaces $\mathbb{D}^{k,p}(V)$ which - by embedding in $L^p(\Omega, \mathcal{F}, P) \otimes V$ - define the maximal domains of D^k , for the limit we define $\mathbb{D}^\infty := \bigcap_{p \geq 1} \bigcap_{k \geq 1} \mathbb{D}^{k,p}(V)$.

The family of operator norms enjoys the following consistency properties:

(i) *Monotonicity*: From $\|f\|_p \leq \|f\|_q$ for $p \leq q$ for ordinary Banach space norms follows for any $F \in \mathcal{S}_p$, $p \leq q$, $k \leq j$ that $\|F\|_{k,p} \leq \|F\|_{j,q}$.

(ii) *Compatibility*: Given a sequence $F_n \in \mathcal{S}_p$ and arbitrary k, p, j, q. If F_n converges with respect to $\|\cdot\|_{k,p}$ to zero and is a Cauchy sequence with respect to $\|\cdot\|_{j,q}$, then it converges also with respect to $\|\cdot\|_{j,q}$ to zero. (Convergence of Cauchy sequences is in a Banach space obvious and monotonicity yields that the limit has to be equal with respect to both norms).
The monotonicity implies obviously $\mathbb{D}^{k+1,p} \subset \mathbb{D}^{k,q}$ for any $p > q$.

The above founded definitions on the structure of the underlying spaces enables us to point out the (already suspected) connection between Malliavin derivatives and Wiener-Itô integrals:

Theorem 4.3.8 (Fundamental Theorem of the Malliavin Calculus)
Given $F \in L^2(\Omega, \mathcal{F}, P)$, then we can decompose it (by Theorem 4.2.4) $F = \sum_{m=0}^{\infty} I_m(\tilde{f}_m)$, $\tilde{f}_m \in H^{\otimes n}$ symmetric. The following statements are equivalent:

(i) $F \in \mathbb{D}^{1,2}$

(ii) $\sum_{m=1}^{\infty} mm! \|\tilde{f}_m\|^2 < \infty$,

and they imply

(iii) $\langle DF, h \rangle_H = \sum_{m=1}^{\infty} m I_{m-1}(\tilde{f}_m \otimes_1 h)$.

Proof We define truncations of F

$$F_n := \sum_{m=0}^{n} I_m(\tilde{f}_m) = \sum_{|p|=m=0}^{n} \|\tilde{f}_m\| H_p(\eta(e)),$$

$e = \{e_i\}_{i \geq 1}$, $e_i \in \{h_j\}_{j \geq 1}$ a basis of H, and get - setting as convention $H_{-1} = 0$ - as Malliavin derivatives for the (generalized) Hermite polynomials

$$\begin{aligned} DH_m(\eta(g)) &= mH_{m-1}(\eta(g))g \\ DH_p(\eta(h_1), ...) &= \sum_{i=1}^{\infty} p_i H_{p_1, p_2, ..., p_i - 1, ...}(\eta(h_1), ...)h_i. \end{aligned}$$

Since we have

$$\|\tilde{f}_m \otimes_1 h\| = \|\tilde{f}_m^1 \otimes ... \otimes \tilde{f}_m^{m-1} \left\langle \tilde{f}_m^m, h \right\rangle \| = \|\tilde{f}_m\| \left\langle e_m, h \right\rangle,$$

we get for the directional derivative (for some m_j with $\sum_{j=1}^{\infty} m_j = m - 1$)

$$\begin{aligned} \langle DF_n, h \rangle &= \left\langle D \sum_{m=0}^{n} I_m(\tilde{f}_m), h \right\rangle \\ &= \sum_{|p|=m=0}^{n} \|\tilde{f}_m\| \left\langle DH_p(\eta(e)), h \right\rangle \\ &= \sum_{|p|=m=1}^{n} m\|\tilde{f}_m\| H_{p-1}(\eta(e)) \left\langle e_m, h \right\rangle \\ &= \sum_{m=1}^{n} m\|\tilde{f}_m\| \left\langle e_m, h \right\rangle I_{m-1} \left(\bigotimes_{j=1}^{m-1} e_j^{\otimes m_j} \right) \\ &= \sum_{m=1}^{n} m I_{m-1} \left(\tilde{f}_m \otimes_1 h \right). \end{aligned} \tag{4.9}$$

This result allows us to prove the theorem:

(ii) ⇒ (i) By Parseval's equation we have

$$\begin{aligned}\|DF_n\|^2 &= \sum_{i=1}^{\infty} E\left(\langle DF_n, h_i\rangle^2\right)\\ &= \sum_{m=1}^{n}\sum_{i=1}^{\infty} E\left(m^2\left(I_{m-1}(\tilde{f}_m \otimes_1 h_i)\right)^2\right)\\ &= \sum_{m=1}^{n}\sum_{i=1}^{\infty} m^2(m-1)!\|\tilde{f}_m\|^2 \langle e_m, h_i\rangle\\ &= \sum_{m=1}^{n} mm!\|\tilde{f}_m\|^2 < \infty,\end{aligned}$$

so the sequence $\|DF_n\|^2$ is bounded by $\sum_{m=1}^{\infty} mm!\|\tilde{f}_m\|^2$, hence it converges; further F is bounded with respect to the norm $\|\cdot\|_{1,2}$, so $F \in \mathbb{D}^{1,2}$.

(i) ⇒ (ii) The required limit regularity we get for DF_n by the integration by parts formula (Proposition 4.3.4) for some $G = I_n(\tilde{g}_n) \in \mathbb{D}^{1,2}$:

$$\begin{aligned}\lim_{n\to\infty} E(G\langle DF_n, h\rangle) &= -\lim_{n\to\infty} E(F_n\langle DG, h\rangle) + \lim_{n\to\infty} E(F_n G\eta(h))\\ &= -E(F\langle DG, h\rangle) + E(FG\eta(h))\\ &= E(G\langle DF, h\rangle).\end{aligned}$$

We can now simply change the direction in the above argument to yield the result.

(iii) By the above proven boundedness and limit properties it is enough to pass in (4.9) to the limit $n \to \infty$ to get

$$\langle DF, h\rangle_H = \sum_{m=1}^{\infty} mI_{m-1}(\tilde{f}_m \otimes_1 h).$$

■

For the concrete representation $H = L^2(T, \mathcal{B}, \mu)$ (iii) reads

$$D_t F = \sum_{m=1}^{\infty} mI_{m-1}(\tilde{f}_m(\eta(h_1), ..., \eta(h_{m-1}), t);$$

this theorem implies in particular that the Wiener chaoses $\mathcal{H}_n$ lie in $\mathbb{D}^{1,2}$.

One of the major tools in dealing with derivatives is usually the chain rule, so let's have a look how this presents itself in Malliavin calculus:

Proposition 4.3.9 (Chain Rule)
Given a differentiable function $\varphi \in C^1$ with bounded derivatives. Then it holds for fixed $p \geq 1$ and $F \in \mathbb{D}^{1,p}(\mathbb{R}^n)$ that $\varphi \circ F \in \mathbb{D}^{1,p}(\mathbb{R}^n)$ and

$$D(\varphi \circ F) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\varphi(F)DF^i = \langle \nabla\varphi(F), DF\rangle.$$

Proof If φ is smooth, the result is a simple consequence of the chain rule in classical analysis. Otherwise we use the mollifyer ρ_ε (defined by $\rho(x) := ce^{\frac{1}{x^2-1}}$, c chosen such that the integral over rho is one, and $\rho_\varepsilon(x) := \varepsilon^n \rho(\varepsilon x)$) to get a smooth approximation $\varphi * \rho_\varepsilon$ for φ. Taking smooth approximations F_n^i of F^i, then $(\varphi * \rho_\varepsilon) \circ F_n^i \to \varphi \circ F^i$ for $\varepsilon \wedge n \to \infty$ in L^p and the closedness of D implies

$$\begin{aligned} & \left\| D(\varphi \circ F) - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \varphi(F) DF^i \right\|_p \\ \leq \quad & \| D(\varphi \circ F) - D\left((\varphi * \rho_\varepsilon) \circ F\right) \|_p + \\ & + \left\| D\left((\varphi * \rho_\varepsilon) \circ F\right) - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} (\varphi * \rho_\varepsilon)(F) DF_n^i \right\|_p + \\ & + \left\| \sum_{i=1}^{n} \frac{\partial}{\partial x_i} (\varphi * \rho_\varepsilon)(F) DF_n^i - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \varphi(F) DF^i \right\|_p \to 0 \end{aligned}$$

■

As consequence we get for instance for φ an approximation of x^2 that $D(F^2) = 2FDF$ and by polarization $D(FG) = FDG + GDF$, F, $G \in \mathbb{D}^{1,2}$. Another direct corollary is that this proposition implies that $\mathbb{D}^\infty$ is a smooth algebra. To generalize this result we have to prove the following lemma of more technical character:

Lemma 4.3.10
Given a sequence $\{F_n\}_{n\geq 1} \in \mathbb{D}^{1,2}$ *converging to* F *in* $L^2(\Omega, \mathcal{F}, P)$ *fulfilling the boundedness condition*

$$\sup_n E\left(\|DF\|_H^2\right) < \infty,$$

then $F \in \mathbb{D}^{1,2}$ *and the sequence of derivatives* $\{DF_n\}_{n\geq 1}$ *converges weakly to* DF *in* $L^2(\Omega, \mathcal{F}, P) \otimes H$.

Proof It is a direct consequence of the Banach-Alaogu theorem on the weak*-compactness of the unit ball of a dual space that every bounded sequence in the dual of a separable topological vector space has a converging subsequence. Hence there is a subsequence DF_{n_k} of DF_n converging weakly to some $\alpha \in L^2(\Omega, \mathcal{F}, P) \otimes H$ and - projecting the inner product $\langle DF_{n_k}, h\rangle$ down to the k-th Wiener Chaos - the projections of $\langle DF_{n_k}, h\rangle$ converge weakly to $\langle \alpha, h\rangle^k$, the projection of $\langle \alpha, h\rangle$. Parseval's identity and the orthogonality of the Wiener chaos decomposition allows us to write for a basis $\{e_i\}_{i\geq 1}$ of H

$$\|\alpha\|^2 = \sum_{i=1}^{\infty} \| \langle \alpha, e_i\rangle \|^2 = \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} \| \langle \alpha, e_i\rangle^k \|^2.$$

Since we can decompose $F = \sum_{m=0}^{\infty} I_m(\tilde{f}_m)$ we have by Theorem 4.3.8

$$\| \langle \alpha, e_i\rangle^k \|^2 = (k+1)(k+1)! \|\tilde{f}_k\|^2,$$

so

$$\sum_{k=0}^{\infty}(k+1)(k+1)!\|\tilde{f}_k\|^2 < \infty$$

and hence $F \in \mathbb{D}^{1,2}$. It follows that $DF = \alpha$, the limit of the subsequence DF_{n_k}; but this argument forces the limit of any weakly converging subsequence of DF_n to be DF, hence the whole sequence DF_n converges weakly to DF. ■

Now we can generalize the chain rule of Malliavin calculus - to even not necessarily differentiable functions!

Theorem 4.3.11 (Generalized Chain Rule)
Given a global Lipschitz function φ*,*

$$|\varphi(x) - \varphi(y)| \leq K|x-y| \qquad x, y \in \mathbb{R}^n$$

and $F \in \mathbb{D}^{1,2}$*. Then* $\varphi \circ F \in \mathbb{D}^{1,2}$ *and there exists a random vector* $G \in \mathbb{R}^n$*,* $|G| < K$ *such that*

$$D(\varphi \circ F) = \sum_{i=1}^{n} G^i DF^i.$$

Proof The proof goes nearly analogous as for Proposition 4.3.9, we can take even the same mollifyer, $\varphi * \rho_\varepsilon$ converges to φ uniformly on compacts. The sequence $D\left((\varphi * \rho_\varepsilon) \circ F\right)$ is bounded in $L^2(\Omega, \mathcal{F}, P) \otimes H$ since $|\nabla\left(\varphi * \rho_\varepsilon\right)| \leq K$ for ε large enough. So we can use the previous lemma which asserts that $\varphi \circ F \in \mathbb{D}^{1,2}$ and $D((\varphi * \rho_\varepsilon) \circ F) \to D(\varphi \circ F)$ weakly. On the other hand side $\nabla\left(\varphi * \rho_{\varepsilon_k}\right) \circ F$ converges weakly to some $G \in \mathbb{R}^n$, $|G| < K$. So it is sufficient to take the weak limit in

$$D((\varphi * \rho_\varepsilon) \circ F) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i}(F) DF^i$$

to yield the result. ■

Now we will calculate Malliavin derivatives of conditional expectations. Therefore we will use the concrete representation $H = L^2(T, \mathcal{B}, \mu)$ and define the σ-algebra

$$\mathcal{F}_A := \{\eta(1_B) : B \subset A \in \mathcal{B}\}$$

which we assume - without loss of generality - to be complete.

Lemma 4.3.12 (Conditioning)
Given $F \in L^2(\Omega, \mathcal{F}, P)$ *with decomposition* $F = \sum_{m=0}^{\infty} I_m(\tilde{f}_m)$*, then it holds for every* $A \in \mathcal{B}$ *that*

$$E(F|\mathcal{F}_A) = \sum_{m=0}^{\infty} I_m\left(\tilde{f}_m 1_A^{\otimes m}\right).$$

Proof By their linearity and density it is enough to prove the lemma for functions f_m of the form $1_{B_1\times\ldots\times B_m}$, the B_m mutually disjoint and of finite measure. For $F = I_m(f_m)$ we have by (4.7)

$$\begin{aligned} E(F|\mathcal{F}_A) &= E\left(\eta\left(1_{B_1}\right)\cdot\ldots\cdot\eta\left(1_{B_m}\right)|\mathcal{F}_A\right) \\ &= E\left(\prod_{i=1}^{m}\left(\eta\left(1_{B_i}\cap 1_A\right)+\eta\left(1_{B_i}\cap 1_{A^c}\right)\right)|\mathcal{F}_A\right) \\ &= \prod_{i=1}^{m}\eta\left(1_{B_i}\cap 1_A\right) \\ &= I_m\left(1_{B_1\times\ldots\times B_m}1_A^{\otimes m}\right). \end{aligned}$$

■

Now we can formulate the theorem:

Theorem 4.3.13 (Conditional Malliavin Derivatives)
Let $F \in \mathbb{D}^{1,2}$ and $A \in \mathcal{B}$, then $E(F|\mathcal{F}_A) \in \mathbb{D}^{1,2}$ and its Malliavin derivative is in $\Omega \times T$ a.s.

$$D_t E(F|\mathcal{F}_A) = E(D_t F|\mathcal{F}_A)1_A(t).$$

Proof By the above lemma and Theorem 4.3.8 we have

$$\begin{aligned} D_t E(F|\mathcal{F}_A) &= D_t\sum_{m=0}^{\infty} I_m\left(\tilde{f}_m 1_A^{\otimes m}\right) \\ &= \sum_{m=1}^{\infty} m I_{m-1}\left(\tilde{f}_m(\eta(h_1),\ldots,\eta(h_{m-1}),t)1_A^{\otimes(m-1)}\right)1_A(t) \\ &= E\left(\sum_{m=1}^{\infty} m I_{m-1}\left(\tilde{f}_m(\eta(h_1),\ldots,\eta(h_{m-1}),t)\right|\mathcal{F}_A\right)1_A(t) \\ &= E(D_t F|\mathcal{F}_A)1_A(t). \end{aligned}$$

■

As last point in this section we prove the Clark-Ocone Haussmann formula which shows that any r.v. $F \in \mathbb{D}^{1,2}$ can be written as the sum of its expectation and a stochastic integral of conditional expectations of its Malliavin derivative.

Theorem 4.3.14 (Clark-Ocone-Haussmann Formula)
Given $F \in \mathbb{D}^{1,2}$ and a one-dimensional Brownian motion B_t, $\eta(h) = \int_0^\infty h(t)dB_t$, then it holds that

$$F = E(F) + \int_0^\infty E\left(D_t F|\mathcal{F}_t\right)dB_t.$$

Proof We use the representation

$$F = \sum_{m=0}^{\infty} I_m \left(f_m \left(\eta(h_1), ..., \eta(h_m) \right) \right) = \sum_{m=0}^{\infty} I_m \left(f_m(t_1, ..., t_m) \right)$$

and can prove the theorem directly by the fundamental theorem of Malliavin calculus (Theorem 4.3.8), Lemma 4.3.12 and the representation of I_n as iterated stochastic integral (Theorem 4.2.6):

$$\begin{aligned}
& \int_0^\infty E\left(D_t F | \mathcal{F}_t\right) dB_t \\
= & \int_0^\infty E\left(D_t \left(\sum_{m=0}^{\infty} I_m \left(f_m(t_1, ..., t_m) \right) \right) \middle| \mathcal{F}_t \right) dB_t \\
= & \int_0^\infty \sum_{m=1}^{\infty} m E\left(I_{m-1}\left(f_m(t_1, ..., t_{m-1}, t\right) | \mathcal{F}_t\right) dB_t \\
= & \int_0^\infty \sum_{m=1}^{\infty} m I_{m-1}\left(f_m(t_1, ..., t_{m-1}, t) 1_{[0,t]} \right) dB_t \\
= & \int_0^\infty \sum_{m=1}^{\infty} \left(m! \int_0^\infty \int_0^{t_{m-1}} \cdots \int_0^{t_2} f_m(t_1, ..., t_{m-1}, t) 1_{[0,t]} dB_{t_1} \cdots dB_{t_{m-1}} \right) dB_t \\
= & \sum_{m=1}^{\infty} m! \int_0^\infty \int_0^t \int_0^{t_{m-1}} \cdots \int_0^{t_2} f_m(t_1, ..., t_{m-1}, t) dB_{t_1} \cdots dB_{t_{m-1}} dB_t \\
= & \sum_{m=1}^{\infty} I_m(f_m) = \sum_{m=0}^{\infty} I_m(f_m) - I_0(f_0) \\
= & F - E(F).
\end{aligned}$$

■

4.4 The Skorohod Integral

Let's critical re-examine the core of the introduction to the last section. To a Gaussian probability space $(\Omega, \mathcal{F}, P)$ and an isonormal Gaussian process $\eta : H \to L^2(\Omega, \mathcal{F}, P)$ (where H was a real and separable Hilbert space) we introduced the Malliavin derivative as mapping

$$\begin{aligned}
D : L^2(\Omega, \mathcal{F}, P) \supset \mathbb{D}^{1,2} & \to & L^2(\Omega, \mathcal{F}, P) \otimes H \\
F & \mapsto & DF.
\end{aligned}$$

The operator D is here unbounded, densely defined and (as proven in Corollary 4.3.6) closed.

So, from a functional analytic viewpoint, all requirements for the existence of an adjoint operator are fulfilled and one can ask if its introduction makes sense in this setting. Indeed there exists an operator

$$\begin{aligned} \delta = D^* : L^2(\Omega, \mathcal{F}, P) \otimes H \supset \mathcal{D}^{1,2} &\rightarrow L^2(\Omega, \mathcal{F}, P) \\ u &\mapsto \delta(u), \end{aligned}$$

the so-called Skorohod integral (or divergence operator). Before going into details we want to give an exact definition.

Definition 4.4.1 (Skorohod Integral)
On the domain

$$\begin{aligned} \mathcal{D}^{1,2} := &\big\{ u \in L^2(\Omega, \mathcal{F}, P) \otimes H : E\left(\langle DF, u\rangle_H\right) \leq c(u)\|F\|_2 \\ &\text{for all } F \in \mathbb{D}^{1,2} \text{ and a constant } c \text{ depending only on } u \big\} \end{aligned}$$

the Skorohod integral $\delta : \mathcal{D}^{1,2} \rightarrow L^2(\Omega, \mathcal{F}, P)$ *is defined as operator fulfilling the equation*

$$E\left(\langle DF, u\rangle_H\right) = E\left(F\delta(u)\right)$$

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

We see that δ is an unbounded, densely defined operator which is (as adjoint) obviously closed. Clearly the rest of this section will be devoted to the study of the Skorohod integral; without much work we can directly see that it coincides with the isonormal Gaussian process in the deterministic case since by Lemma 4.3.4

$$E\left(F\delta(1 \otimes h)\right) = E\left(\langle DF, 1 \otimes h\rangle\right) = E\left(F\eta(h)\right)$$

for arbitrary $h \in H$.

Lemma 4.4.2
Given $u \in L^2(\Omega, \mathcal{F}, P) \otimes H$*, then we can write it as*

$$u = \sum_{m=0}^{\infty} \left(I_m \otimes id\right)(f_m) \tag{4.10}$$

for $f_m \in H^{\otimes(m+1)}$ *symmetric the first* m *variables and it holds that*

$$\|u\|^2_{L^2(\Omega,\mathcal{F},P)\otimes H} = \sum_{m=0}^{\infty} m!\|f_m\|^2_{H^{\otimes(m+1)}}.$$

Proof The existence of the f_m is clear by the chaos expansion (Theorem 4.2.4) and the isometry

$$L^2(\Omega, \mathcal{F}, P) \otimes H \simeq \left(\bigoplus_{m=0}^{\infty} \mathcal{H}_m\right) \otimes H \simeq \bigoplus_{m=0}^{\infty} \left(\mathcal{H}_m \otimes H\right)$$

which follows from the chaos decomposition (Theorem 4.1.10) and the elementary tensor product property from Lemma 1.2.2(iii).
For the norm equality we first note that we can - by orthogonality - draw the

norm into the sum, the rest is only technical tensor analysis for $h \in \{e_i\}_{i\geq 1}$, the basis of H,

$$\begin{aligned}\|u\|^2_{L^2(\Omega,\mathcal{F},P)\otimes H} &= \|\sum_{m=0}^{\infty}(I_m\otimes id)(f_m)\|^2_{L^2\otimes H} = \sum_{m=0}^{\infty}\|(I_m\otimes id)(f_m)\|^2_{L^2\otimes H}\\ &= \sum_{m=0}^{\infty}\|I_m(f_m\otimes_1 h)\|^2_{L^2\otimes H}\|id(h)\|^2_{L^2\otimes H}\\ &= \sum_{m=0}^{\infty} m!\|f_m\|^2_{H^{\otimes(m+1)}}\|h\|^2_H\\ &= \sum_{m=0}^{\infty} m!\|f_m\|^2_{H^{\otimes(m+1)}}\end{aligned}$$

■

Proposition 4.4.3
Given $u \in L^2(\Omega,\mathcal{F},P)\otimes H$ with the representation (4.10), then the following conditions are equivalent

(i) $\sum_{m=0}^{\infty}(m+1)!\|f_m\|^2_{H^{\otimes(m+1)}} < \infty$,

(ii) $\sum_{m=0}^{\infty} I_{m+1}(f_m)$ *converges in* L^2,

(iii) $u \in \mathcal{D}^{1,2}$,

and we have the concrete representation of the Skorohod integral

(iv) $\delta(u) = \sum_{m=0}^{\infty} I_{m+1}(f_m)$.

Proof

(i) ⇔ (ii) This is clear by

$$E\left((I_{m+1}(f_m))^2\right) = (m+1)!\|f_m\|^2_{H^{\otimes(m+1)}}.$$

(i) ⇒ (iii) Assume first that $G = I_n(g)$ for some symmetric $g \in H^{\otimes n}$, then by orthogonality and since $\|\tilde{f}_n\| = \|f_n\|$

$$\begin{aligned}E\left(\langle DG,u\rangle_H\right) &= E\left(\langle DI_n(g),u\rangle_H\right) = E\left(\langle n(I_{n-1}\otimes id)(g),u\rangle_H\right)\\ &= E\left(\langle n(I_{n-1}\otimes id)(g),(n-1)(I_{n-1}\otimes id)(f_{n-1})\rangle_H\right)\\ &= n(n-1)!\,\langle g,f_{n-1}\rangle_{H^{\otimes n}} = n!\left\langle g,\tilde{f}_{n-1}\right\rangle_{H^{\otimes n}}\\ &= E\left(I_n(g)I_n(\tilde{f}_{n-1})\right) = E\left(GI_n(\tilde{f}_{n-1})\right). \qquad (4.11)\end{aligned}$$

So under the convergence assumption $\sum_{m=0}^{\infty} I_{m+1}(\tilde{f}_m) = V$ we have for partial sums

$$\begin{aligned}E\left(\left\langle D\sum_{n=0}^{k} I_n(g_n),u\right\rangle_H\right) &= E\left(\left(\sum_{n=0}^{k} I_n(g_n)\right)\left(\sum_{n=0}^{k} I_n(\tilde{f}_{n-1})\right)\right)\\ &\leq E\left(\left(\sum_{n=0}^{k} I_n(g_n)\right)V\right),\end{aligned}$$

hence by Hölder
$$E\left(\langle DF,u\rangle_H\right) \leq \|V\|_2\|F\|_2.$$
This holds for all F with finite chaos decomposition $u \in \mathcal{D}^{1,2}$, but since they are dense, the result follows.

(iii) $\Rightarrow$ (ii) Calculation (4.11) shows that

$$E\left(\langle DF,u\rangle_H\right) = E\left(\sum_{m=0}^{\infty} I_m(\tilde{f}_{m-1})F\right) < \infty,$$

so convergence follows.

(iv) From

$$E\left(G\delta(u)\right) = E\left(\langle DG,u\rangle_H\right) = E\left(GI_n(\tilde{f}_{n-1})\right)$$

follows that the projections of $\delta(u)$ and $\sum_{m=0}^{\infty} I_{m+1}(\tilde{f}_m)$ on the n-th Wiener chaos coincide for every n, so the functions have to be equal. ■

This representation implies that δ is a linear operator on $\mathcal{D}^{1,2}$ and

$$E\left(\delta(u)\right) = \left\langle \sum_{m=0}^{\infty} I_{m+1}(f_m), I_0(1)\right\rangle_H = 0$$

for $u \in \mathcal{D}^{1,2}$.

Our next theorem concerns the representation of the processes u:

Proposition 4.4.4 (Representation)
Every stochastic process $u \in L^2(\Omega,\mathcal{F},P) \otimes H$ has a unique orthogonal decomposition
$$u = DF + u^0$$
with $F \in \mathbb{D}^{1,2}$ and $E\left(\left\langle DG,u^0\right\rangle_H\right) = 0$ for every $G \in \mathbb{D}^{1,2}$.

Proof First we ask which processes u can be represented as DF: This is clearly the case for all functions which are symmetric in all variables (and not only all but the last). DF is symmetric in all variables as a sum of Wiener-Itô integrals (see Theorem 4.3.8), on the other hand side we can set for $u = \sum_{m=0}^{\infty}(I_m \otimes id)(f_m)$

$$F = \sum_{m=0}^{\infty} \frac{1}{m+1} I_{m+1}(f_m)$$

which is a series converging $\mathbb{D}^{1,2}$ with

$$\langle DF,h\rangle_H = \sum_{m=0}^{\infty} \frac{1}{m+1}\langle DI_{m+1}(f_m),h\rangle_H = \sum_{m=0}^{\infty} I_{m+1}(f_m \otimes_1 h) = \langle u,h\rangle_H$$

for any $h \in H$. Hence the processes u representable as DF, $F \in \mathbb{D}^{1,2}$ form a closed subspace of $L^2(\Omega,\mathcal{F},P)\otimes H$ and for $G \in \mathbb{D}^{1,2}$ we have $\left\langle DG,u^0\right\rangle_H = 0$. ■

Furthermore $E\left(\langle DG, u^0\rangle_H\right) = 0$, hence u^0 is Skorohod integrable by Definition 4.4.1 and $\delta(u^0) = 0$.

The next results will only apply to special classes of Skorohod integrable functions. First we will focus on so-called smooth elementary processes $u \in \mathcal{S}_p \otimes H$, hence processes which we can write $u = \sum\limits_{j=1}^{n} F_j \otimes h_j$ for $F_j \in \mathcal{S}_p$ and $h_j \in H$.

Proposition 4.4.5 (Smooth Elementary Processes)
Smooth elementary processes are Skorohod integrable ($\mathcal{S}_p \otimes H \subset \mathcal{D}^{1,2}$) and for u as above the divergence is

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

Proof For an arbitrary $F \in \mathbb{D}^{1,2}$ we have

$$\begin{aligned} E\left(G\delta(u)\right) &= E\left(\langle DG, u\rangle_H\right) = \sum_{j=1}^{n} E\left(F_j \langle DG, h_j\rangle_H\right) \\ &= \sum_{j=1}^{n} E\left(GF_j\eta(h_j)\right) - \sum_{j=1}^{n} E\left(G\langle DF_j, h_j\rangle_H\right) \\ &\leq c(u)\|G\|_2 \end{aligned}$$

by the integration of parts formula (Proposition 4.3.5). ■

To point out the commutativity property of the Skorohod integral and the Malliavin derivative we have to restrict ourselves to an even smaller class of functions.

Definition 4.4.6 ($\mathbb{L}^{1,2}$)
A stochastic process $u \in \mathcal{D}^{1,2}$ is said to be in $\mathbb{L}^{1,2}$, iff $u(t) \in \mathbb{D}^{1,2} \otimes H$.

$\mathbb{L}^{1,2}$ is a Hilbert space with respect to the norm

$$\|u\|_{1,2} := \left(\|u\|^2_{L^2(\Omega,\mathcal{F},P)\otimes H} + \|Du\|^2_{L^2(\Omega,\mathcal{F},P)\otimes H^2}\right)^{\frac{1}{2}}$$

Proposition 4.4.7 (Commutation)
For a process $u \in \mathbb{L}^{1,2}$ it holds that

$$\left[\langle D(\cdot), h\rangle_H, \delta(\cdot)\right]_u = \langle u, h\rangle_H .$$

Proof We can represent u as in (4.10) and get, since the Wiener-Itô integrals are invariant under symmetrization (Theorem 4.2.2),

$$\begin{aligned}
\langle D\left(\delta(u)\right), h\rangle_H &= \left\langle D\sum_{m=0}^{\infty} I_{m+1}(\tilde{f}_m), h\right\rangle_H = \sum_{m=0}^{\infty}(m+1)I_m(\tilde{f}_m \otimes_1 h) \\
&= \sum_{m=0}^{\infty} I_m(\tilde{f}_m \otimes_1 h) + \sum_{m=1}^{\infty} mI_m(\tilde{f}_m \otimes_1 h) \\
&= \sum_{m=0}^{\infty} I_m(f_m \otimes_1 h) + \sum_{m=1}^{\infty} mI_m(\widetilde{f_m \otimes_1} h) \\
&= \sum_{m=0}^{\infty} \langle (I_m \otimes id)(f_m), h\rangle + \delta\left(\sum_{m=1}^{\infty} m\left(I_{m-1} \otimes id\right)(f_m \otimes_1 h)\right) \\
&= \langle u, h\rangle_H + \delta\left(\left\langle D\left(\sum_{m=0}^{\infty}\left(I_m \otimes id\right)(f_m)\right), h\right\rangle_H\right) \\
&= \langle u, h\rangle_H + \delta\left(\langle Du, h\rangle_H\right).
\end{aligned}$$

■

As a direct consequence of this commutativity property we can calculate the covariance of the Skorohod integral for $\mathbb{L}^{1,2}$-processes.

Corollary 4.4.8 (Covariance)
The covariance of the Skorohod integral is for two processes u, $v \in \mathbb{L}^{1,2}$ given by

$$E\left(\delta(u)\delta(v)\right) = E\left(\langle u, v\rangle_H\right) + E\left(\langle Du, Dv\rangle_H\right)$$

Proof By the definition of the Skorohod integral (Definition 4.4.1) and the commutation property (Proposition 4.4.7) it holds for any $u \in \mathbb{L}^{1,2}$ with finite Wiener expansion that $\delta(u) \in \mathbb{D}^{1,2}$ and

$$\begin{aligned}
E\left(\delta(u)\delta(v)\right) &= E\left(\langle D(\delta(u)), v\rangle_H\right) \\
&= E\left(\langle u, v\rangle_H\right) + E\left(\langle \delta(Du), v\rangle_H\right) \\
&= E\left(\langle u, v\rangle_H\right) + E\left(\langle Du, Dv\rangle_H\right).
\end{aligned}$$

In the last step it is implicit that the inner product can - for a concrete representation - be written as integral and hence by Fubini commutes with the expectation. But the processes with finite Wiener expansion are dense in $\mathbb{L}^{1,2}$ and the corollary follows. ■

We are now coming back to a more general class of Skorohod kernels to calculate the Skorohod integral of the product of a r.v. with a process.

Proposition 4.4.9 (Multipication)
Given $u \in \mathcal{D}^{1,2}$ and $F \in \mathbb{D}^{1,2}$ such that $E\left(F^2\left(\delta(u)\right)^2\right) < \infty$, then it holds that

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

Proof Also this proposition we can prove by algebraic means: For $G \in \mathcal{S}_0$ we have by the chain rule for Malliavin derivatives (Proposition 4.3.9)

$$\begin{aligned} E(GF\delta(u)) &= E(\langle D(GF), u\rangle_H) \\ &= E(\langle GDF, u\rangle_H + \langle FDG, u\rangle_H) \\ &= E(G\langle DF, u\rangle_H + \langle DG, Fu\rangle_H) \\ &= E(G\langle DF, u\rangle_H) + E(G\delta(Fu)) \end{aligned}$$

which yields the result since $\mathcal{S}_0$ is dense in $\mathcal{S}_p$. ∎

By the definition of the Skorohod integral as counterpart of the Malliavin derivative it seems quite logical to ask for the connection to multiple Wiener-Itô integrals (the climbing up) and Itô integrals (here the connection was established by Theorem 4.2.6). In fact, we will show that the stochastic Itô integral is nothing else than a special case of the much more general Skorohod integral. But first we have to proof the following Lemma, therefore we have to return to our concrete representation $L^2(T, \mathcal{B}, \mu)$ of H, recalling that μ was a σ-finite and atomless measure.

Lemma 4.4.10
Given $F \in L^2(\Omega, \mathcal{F}, P)$, measurable with respect to the σ-algebra

$$\mathcal{F}_{A^c} := \{\eta(1_B) : B \subset A^c \in \mathcal{B}\},$$

then $F1_A$ is Skorohod integrable and

$$\delta(F1_A) = F\eta(1_A).$$

Proof Assume first that $F \in \mathbb{D}^{1,2}$, then it follows by Proposition 4.4.9 on the Skorohod integral of a process multiplied with a r.v. that

$$\begin{aligned} \delta(F1_A) &= F\delta(1_A) - \langle D_t F, 1_A\rangle_{L^2(T)} \\ &= F\eta(1_A) - \langle D_t(E(F|\mathcal{F}_{A^c})), 1_A\rangle_{L^2(T)} \\ &= F\eta(1_A) - \langle E(D_t F|\mathcal{F}_{A^c})\, 1_{A^c}(t), 1_A\rangle_{L^2(T)} \\ &= F\eta(1_A) \end{aligned}$$

by the $\mathcal{F}_{A^c}$-measurability of F and Theorem 4.3.13. Since $\mathbb{D}^{1,2}$ is dense in $L^2(\Omega, \mathcal{F}, P)$ and since δ is a closed operator we can pass to the limit which yields the general result. ∎

Now we can state our theorem on the connection of Itô and Skorohod integral; we do this only in the one dimensional case, but the generalization is not difficult.

Theorem 4.4.11 (Itô and Skorohod Integrals)
The domain of the Skorohod integral contains that of the Itô integral and for $\eta(h) = \int_0^\infty h(t)dB_t$,

$$u \in L^2(\mathbb{R}_{\geq 0} \times \Omega, \mathcal{B}(\mathbb{R}_{\geq 0}) \otimes \mathcal{F}_t, dt \otimes P) \subset \mathcal{D}^{1,2}$$

it holds for a one-dimensional Brownian motion B_t that

$$\delta(u) = \int_0^\infty u_t dB_t.$$

Proof Assume first that u is an elementary step process $u_t = \sum_{j=1}^{n} F_j 1_{]t_j,t_{j+1}]}(t)$ with $F_j \in L^2(\Omega, \mathcal{F}, \mathcal{F}_{[0,t_j]}, P)$ adapted. Then the previous lemma states that $u \in \mathcal{D}^{1,2}$ and

$$\begin{aligned} \delta(u) &= \delta\left(\sum_{j=1}^{n} F_j 1_{]t_j,t_{j+1}]}(t)\right) = \sum_{j=1}^{n} F_j \eta\left(1_{]t_j,t_{j+1}]}(t)\right) \\ &= \sum_{j=1}^{n} F_j \int_0^\infty 1_{]t_j,t_{j+1}]}(t) dB_t = \sum_{j=1}^{n} F_j \left(B_{t_{j+1}} - B_{t_j}\right) \\ &= \int_0^\infty u_t dB_t \end{aligned}$$

by Theorem 4.2.6. By the approximation (Theorem 2.3.3) we can pass to the limit - since δ is a closed operator - to get the result. ■

4.5 Malliavin Derivative and First Variation

Till now we developed two different concepts of the derivative of a stochastic process: on the one hand side we introduced to a process X satisfying the SDE

$$dX_t^x = b(t, X_t^x)dt + \sigma(t, X_t^x)dB_t \tag{4.12}$$

the first variation process $Y = (DX)$ which satisfies the "derived" SDE

$$dY_t^x = b'(t, X_t^x)Y_t^x dt + \sigma'(t, X_t^x)Y_t^x dB_t;$$

in Chapter 3 and now we presented the Malliavin derivative operator D (or concretely D_t). Not only notational similarities suggest a connection between these two notions... Exactly this connection is what we want to study in this section, here we can have a first glance on the real goal of the present thesis: The use of Malliavin calculus to gain information on SDEs and the behavior of their solutions. Since SDEs are highly concrete objects we do not wonder that we have to work on our concrete representation $L^2(T, \mathcal{B}, \mu)$, μ atomless and σ-finite, of the Hilbert space.

Proposition 4.5.1 (Malliavin Derivative as SDE Solution)
Under the usual conditions on b and σ (as in Chapter 3, Prop 3.3.2) and $\eta(h) = \int_0^\infty h(t)dB_t$ we have $X_t^x \in \mathbb{D}^{1,2}$ and the Malliavin derivative $D_r X_t^x$ satisfies the

SDE

$$D_r X_t^x = \sigma(r, X_r^x) + \int_r^t b'(s, X_s^x) D_r X_s^x ds + \int_r^t \sigma'(s, X_s^x) D_r X_s^x dB_s. \tag{4.13}$$

Proof First we note that $X_t^x \in \mathbb{D}^{1,2}$ implies $D_t X_s^x = 0$ for $t > s$ since

$$D_t X_s^x = D_t \left(E(X_s^x | \mathcal{F}_s) \right) = E \left(D_t X_s^x | \mathcal{F}_s \right) 1_{[0,s]}(t)$$

by Theorem 4.3.13 and hence

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

for processes $u \in \mathbb{D}^{1,2}$ by approximation with Riemann sums of mollified elementary step processes. Further we have by the commutation property of the Skorohod integral

$$D_r \int_0^t u_s dB_s = \int_r^t D_r u_s dB_s + u_r$$

so that we can try what happens if we "differentiate" directly in the SDE

$$X_t^x = x + \int_0^t b\left(s, X_s^x\right) ds + \int_0^t \sigma\left(s, X_s^x\right) dB_s.$$

We get by the chain rule

$$\begin{aligned} D_r X_t^x &= D_r \left(\int_0^t b\left(s, X_s^x\right) ds \right) + D_r \left(\int_0^t \sigma\left(s, X_s^x\right) dB_s \right) \\ &= \int_r^t b'\left(s, X_s^x\right) D_r X_s^x ds + \int_r^t \sigma'\left(s, X_s^x\right) D_r X_s^x dB_s. \end{aligned}$$

This is in no way a result since we know nothing about the solvability yet. But by exactly the same means as for the first variation (Theorem 3.3.2) we can show that $D_r X_t^x$ is the unique solution of the SDE and is jointly continuous in t and x (with absolute moments of all orders). Furthermore X_t^x is by the Picard-Lindelöf iteration really in $\mathbb{D}^{1,2}$. ∎

Now, having an even very strong similarity of the SDEs satisfied by Malliavin derivative and first variation process we will try to establish a direct connection:

Theorem 4.5.2 (Malliavin Derivative and First Variation)
Given X_t^x the solution of the SDE (4.12), then it holds for its first variation Y_t^x that

$$D_r X_t^x = Y_t^x (Y_r^x)^{-1} \sigma(r, X_r^x) 1_{[0,t]}(r).$$

Proof Proposition 4.5.1 implies that $D_r X_t^x$ suffices an (in t) inhomogeneous equation and it is, by the assumptions on b and σ, a diffeomorphism. So we can write it in the form

$$D_r X_t^x = B^x(r,t)A \tag{4.14}$$

where $B^x(r,t)$ is a flow and A a matrix derived from the inhomogeneity. We conclude by (4.13) immediately

$$A = \sigma(r, X_r^x) 1_{\{r \leq t\}}$$

and since B is a flow it holds that

$$B^x(0,t) = B^x(r,t) \circ B^x(0,r)$$

and hence by invertibility

$$B^x(r,t) = B^x(0,t) \circ \left(B^x(0,r)\right)^{-1}.$$

Therefore equation (4.13) reads by (4.14)

$$B^x(r,t)A = A + \int_r^t b'(s, X_s^x) B^x(r,s) A ds + \int_r^t \sigma'(s, X_s^x) B^x(r,s) A dB_s.$$

Setting $r = 0$ we have

$$B^x(0,t) = 1 + \int_0^t b'(s, X_s^x) B^x(0,s) ds + \int_0^t \sigma'(s, X_s^x) B^x(0,s) dB_s$$

and - from the uniqueness of the first variation process stated in Theorem 3.3.2 - can derive that $B^x(0,t) = Y_t^x$ a.s. which yields the result. ■

Corollary 4.5.3
Under the above conditions it holds a.s. for $\psi \in C_b^1(\mathbb{R})$ and $r \leq t \leq T$ that

$$D_r \int_0^t \psi(X_s^x) ds = \int_0^t \nabla\psi(X_s^x) Y_s^x (Y_r^x)^{-1} \sigma(r, X_r^x) ds.$$

Proof Applying the chain rule to the above theorem yields

$$D_r \psi(X_t^x) = \nabla\psi(X_t^x) D_r X_t^x = \nabla\psi(X_t^x) Y_t^x (Y_r^x)^{-1} \sigma(r, X_r^x) 1_{[0,t]}(r)$$

implying the result. ■

In this context yet another question rises: Under which conditions does a random variable have a density? And how smooth is it? Also for this question Malliavin calculus is the perfect tool to treat it; more, it was developed to prove Hörmanders celebrated "Sums of the Squares" theorem which is directly connected to this question. But all this is beyond the scope of the present work and it is not required for the practical applications in the next chapter. But since this is a very good example to see Malliavin calculus at work we will prove the existence of the density in the one-dimensional case and state the general result without proof.

Proposition 4.5.4 (Existence of Densities)
Given $F \in \mathbb{D}^{1,2}$ and assume that $\frac{DF}{\|DF\|_H^2}$ is Skorohod integrable. Then the law of F has a continuous and bounded density f given by

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

Proof For some reals $a < b$ we define the functions

$$\psi(y) := 1_{[a,b]}(y) \text{ and } \varphi(y) = \int_{-\infty}^{y} \psi(z)dz.$$

Obviously $\varphi \circ F \in \mathbb{D}^{1,2}$ and by the chain rule for Malliavin derivatives (Proposition 4.3.9) we get

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

By the definition of the Skorohod integral this leads to

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

by Fubini's theorem. ■

In the general case we get:

Theorem 4.5.5 (General Existence and Smoothness of Densities)
Given a random vector F, then it has an infinitely differentiable density if the following conditions hold:

(i) $F_i \in \mathbb{D}^\infty$ for all i,

(ii) The Malliavin matrix $\gamma_F := (\langle DF_i, DF_j\rangle)_{ij}$ satisfies $(\det \gamma_F)^{-1} \in \bigcap_{p>1} L^p(\Omega, \mathcal{F}, P)$.

Proof See [Nua 95], p.91. ■

Chapter 5

Calculating the Greeks

In this chapter we will be concerned with applications of the Malliavin calculus to finance. We want to determine the sensitivity of the price of an option with respect to small perturbations of the dynamics of the underlying asset, this is, in the language of finance, calculate the *Greeks*. But first we have to give an (indeed very sketchy) introduction in the theory of pricing options.

5.1 Pricing Options: A very rough Primer

The general background is the question of pricing and hedging derivative securities. In opposition to the *underlying stocks* as shares, bonds, currencies or commodities which give the holder the direct possession of a good, *derivatives* give the holder the right to buy or sell a specified quantity of a stock until a specified date, the *maturity* of the derivative for a price specified in advance, the *strike price*. (In fact everything can get more complicated if we allow also derivatives relying on other derivatives...)

While forward contracts or futures give the holder the right and obligation to buy/sell the underlying stock at a specified future date to an in advance fixed price, options give the holder only the right to buy/sell, but she has no obligation to do so (and will obviously not exercise it if the actual price is lower/higher as the strike price). Options which give the right to buy are named *call options*, those which give the right to sell *put options*. Furthermore options are discerned with respect to their dependence of the payoff: For instance *European options* give the holder the right (but not the obligation) to buy/sell the underlying for the strike price (only) at maturity while *American options* allow this at any date until maturity. But apart from that a whole bunch of different "exotic" options with rather unusual payoff functions exists...

The practical use of options can be quite diverse: On the one hand side one can use options as insurance against price hikes, on the other hand side they can also be used for speculations on increasing or decreasing stock prices. But the crucial question is: How to determine their actual value? How do we know which price we are ready to pay for them? The baffling answer is that the fair price of an option - under some assumptions - does *not* depend on the general expectations of the behavior of the prices!

The two major assumptions are the following:
(i) *No Arbitrage*: Any opportunity to get a risk-free profit is called an arbitrage opportunity. We assume that there are no such possibilities.
(ii) We require that every possible derivative can be hedged, i.e. that we can construct a portfolio consisting of a risk-free cash bond and underlying stocks which has exactly the value of the derivative at maturity.

Assuming that the underlying $X_t^x \in \mathbb{R}^n$ satisfies a SDE

$$dX_t^x = b(X_t^x)dt + \sigma(X_t^x)dB_t = b(X_t^x)dt + \sum_{i=1}^{d} (\sigma(X_t^x))^i \, dB_t,$$

the fair price is given by the expectation of the payoff function $\varphi(\cdot)$,

$$u_t^x = E(\varphi(X_t^x)).$$

Then the no-arbitrage condition is equal to requiring the existence of a probability measure Q under which

$$X_0^x = cE_Q(X_T^x),$$

the risk-neutral probability measure (or equivalent martingale measure). The connection to the Girsanov theorem (Theorem 2.5.3) is obvious, the Girsanov transform deletes the drift and this way forces the independence of the expected market movement such that the price is hence only dependent on th volatility σ. The market is complete if the equivalent martingale measure is unique.

5.2 The Greeks

The *Greeks* delta, gamma, theta, rho and the "pseudo-Greek" vega are given by

$$\Delta_t^x := \frac{\partial u_t^x}{\partial x}, \quad \Gamma_t^x := \frac{\partial^2 u_t^x}{\partial x^2}, \quad \Theta_t^x := \frac{\partial u_t^x}{\partial t}, \quad \rho_t^x := \frac{\partial u_t^x}{\partial b}, \quad \nu_t^x := \frac{\partial u_t^x}{\partial \sigma},$$

they denote the sensitivity of the price for small perturbations of the initial value (respectively the time, the drift and the volatility coefficient). At a first glance we can see that these usual definitions are for the rho and the vega not satisfactory in our framework; for constant b and σ (as in the Black-Scholes framework) this is enough, but in our case they are dependent of X_t^x. So we have to define the derivatives in the directions $\tilde{b}$ (respectively $\tilde{\sigma}$ by setting the drift (volatility) coefficient $b(X_t^x) + \varepsilon\tilde{b}(X_t^x)$ $(\sigma(X_t^x) + \varepsilon\tilde{\sigma}(X_t^x))$ and evaluate the derivatives with respect to ε at $\varepsilon = 0$.

An explicit calculation of the Greeks is not possible in the case of more difficult payoff functions, so one uses numerical approaches relying on Monte-Carlo estimates (their theory is not in the scope of the present work) for the prices u_t^x. The easiest methods like finite difference approaches, i.e. for instance for the delta calculating

$$\frac{u_t^{x+\varepsilon} - u_t^x}{\varepsilon}$$

for small ε, have the disadvantage that they have very poor convergence, in particular for not so smooth payoff functions φ.

One idea of now to overcome this problem is the following which we want to present for the delta: Assume that the function $\varphi : \mathbb{R}^n \to \mathbb{R}$ has a density $p(t, x, \cdot)$ which is dominated by a L^1-function and differentiable, then we can calculate

$$\begin{aligned}
\nabla E\left(\varphi(X_t^x)\right) &= \nabla \int\limits_{\mathbb{R}^n} \varphi(y) p(t,x,y) dy \\
&= \int\limits_{\mathbb{R}^n} \varphi(y) \nabla p(t,x,y) dy \\
&= \int\limits_{\mathbb{R}^n} \varphi(y) p(t,x,y) \frac{\nabla p(t,x,y)}{p(t,x,y)} dy \\
&= E\left(\varphi(X_t^x)\pi\right)
\end{aligned}$$

where π is the logarithmic derivative

$$\pi = \nabla \ln p(t,x,y) = \frac{\nabla p(t,x,y)}{p(t,x,y)}$$

and is called the respective Malliavin weight (for delta). We clearly see that Malliavin weights are far from being unique since for any π_0 with $\varphi(X_t^x) \perp \pi_0$ the expression $\pi + \pi_0$ is also a Malliavin weight.

Since the formula should hold for $\varphi(X_t^x) = 1_{[0,T]}(t)$ too, it follows directly by $\varphi'(X_t^x) = 0$ that $E(\pi) = 0$ since

$$E(\pi) = E\left(1_{[0,T]}(t)\pi\right) = E\left(\varphi(X_t^x)\pi\right) = E\left(\varphi'(X_t^x) Y_t^x\right) = 0,$$

so π is a stochastic process with zero expectation. One can guess that it should be something like a stochastic integral

$$\sum_{i=1}^{d} \int\limits_0^T (\psi(X_t^x))^i dB_t^i$$

for a process ψ. What we gain by this approach is the boundedness

$$|\nabla u_t^x| \leq C\|\varphi\|,$$

which depends only on φ (and not on φ'!).

Next we want calculate the Malliavin weights (and so the Greeks) in the elliptic case, following the paper by Fournié, Lasry, Lebuchoux, Lions and Touzi [FLLLT 99].

5.3 The Elliptic Case

First we have to lay out the general setting for this purpose. The stochastic process $\{X_t\}_{0\leq t\leq T} \in \mathbb{R}^n$ is driven by the SDE (with respect to an n-dimensional Brownian motion)

$$dX_t^x = b(X_t^x)dt + \sigma(X_t^x)dB_t$$

where the vector fields b, $\sigma^i : [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$ (σ^i the i-th column of σ, $1 \leq i \leq n$) are C^∞-bounded what assures us e.g. the unique existence of the solution and the first variation. Furthermore we endow σ with a yet stronger regularity property, uniform ellipticity:

Definition 5.3.1 (Uniform Ellipticity)
We say that σ satisfies the uniform ellipticity condition, iff there is an $\varepsilon > 0$ such that it holds that

$$(\sigma(x)\zeta)^\tau (\sigma(x)\zeta) \geq \varepsilon\|\zeta\|^2$$

for all $t \in [0,T]$ and x, $\zeta \in \mathbb{R}^n$.

This implies that the matrix σ is positive definite and all eigenvalues are greater then ε, hence σ is invertible and the inverse σ^{-1} is bounded. In our case this means that for any bounded function $\gamma : [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$, $\sigma^{-1}\gamma$ is bounded and the process $\left\{(\sigma^{-1}\gamma)(X_t^x)\right\}_{0\leq t\leq T}$ lies in $L^2([0,T]\times\Omega,\mathcal{F}_p, dt\otimes P)$.

5.3.1 The Rho

The no-arbitrage condition tells us exactly that the drift part vanishes under the equivalent martingale measure. So we do not have to wonder that we can determine the rho in a quite general case using Girsanov's theorem: Given a payoff function $\varphi : C([0,T]) \to \mathbb{R}$ satisfying $E\left(\varphi(X_t^x)^2\right) < \infty$. We define the perturbed process $\{\hat{X}_t^x\}_{0\leq t\leq T}$ by

$$d\hat{X}_t^x = \left(b(\hat{X}_t^x) + \varepsilon\tilde{b}(\hat{X}_t^x)\right)dt + \sigma(\hat{X}_t^x)dB_t$$

for small ε and $\tilde{b} : [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$ bounded and further

$$\xi_t^x := e^{-\varepsilon\sum\limits_{i=1}^n\int\limits_0^t\left((\sigma^{-1}\tilde{b})(X_s^x)\right)^i dB_s^i - \frac{\varepsilon^2}{2}\sum\limits_{i=1}^n\int\limits_0^t\left(\left((\sigma^{-1}\tilde{b})(X_s^x)\right)^i\right)^2 ds}$$

which is a true martingale by Novikov's condition (Corollary 2.5.4) since $(\sigma^{-1}\tilde{b})$ is bounded.

Our aim is to derive the expectation

$$\hat{u}_t^x := E\left(\varphi(\hat{X}_t^x)\right)$$

and evaluate it at $\varepsilon = 0$, this gives the price sensitivity with respect to the interest rate (which corresponds with the drift part), the ρ.

Proposition 5.3.2 (Rho)
The function $\varepsilon \mapsto \hat{u}_t^x$ *is differentiable in* $\varepsilon = 0$ *and it holds that*

$$\rho_t^x = \left.\frac{\partial \hat{u}_t^x}{\partial \varepsilon}\right|_{\varepsilon=0} = E\left(\varphi(X_t^x)\sum_{i=1}^n \int_0^T \left((\sigma^{-1}\tilde{b})(X_t^x)\right)^i dB_t^i\right).$$

Proof (i) Since $E(\xi_T^x) = 1$ (by Novikov's condition) we can define a probability measure $\hat{Q}$ with $\frac{d\hat{Q}}{dP} = (\xi_T^x)^{-1}$ which is, in fact, even equivalent to P since $\xi_T^x > 0$ a.s. So we have for

$$\begin{aligned} \hat{B}_t &:= B_t + \varepsilon \int_0^t (\sigma^{-1}\tilde{b})(X_s^x) ds \\ \hat{\xi}_t^x &:= e^{-\varepsilon \sum_{i=1}^n \int_0^t \left((\sigma^{-1}\tilde{b})(\hat{X}_s^x)\right)^i d\hat{B}_t^i - \frac{\varepsilon^2}{2} \sum_{i=1}^n \int_0^t \left(\left((\sigma^{-1}\tilde{b})(\hat{X}_s^x)\right)^i\right)^2 ds} \end{aligned}$$

by the Radon-Nikodym theorem (Theorem 2.5.1)

$$\begin{aligned} \hat{u}_t^x &= E\left(\varphi(\hat{X}_t^x)\right) \\ &= E_{\hat{Q}}\left(\varphi(\hat{X}_t^x)\hat{\xi}_T^x\right) \\ &= E\left(\varphi(X_t^x)\xi_T^x\right) \end{aligned}$$

since the joint distributions of $(\hat{X}_t^x, \hat{B}_t)$ under $\hat{Q}$ and of (X_t^x, B_t) under P coincide by definition of $\hat{X}_t^x$ and $\hat{B}_t$.

(ii) Since ξ_t^x is a stochastic exponential we have

$$d\xi_t^x = \varepsilon \sum_{i=1}^n \xi_t^x \left((\sigma^{-1}\tilde{b})(X_t^x)\right)^i dB_t$$

and since $\xi_t^x = 1$ for $\varepsilon = 0$

$$\frac{1}{\varepsilon}(\xi_T^x - 1) = \sum_{i=1}^n \int_0^T \xi_t^x \left((\sigma^{-1}\tilde{b})(X_t^x)\right)^i dB_t^i.$$

For $\varepsilon \to 0$ this yields the L^2-convergence

$$\frac{1}{\varepsilon}(\xi_T^x - 1) \to \sum_{i=1}^n \int_0^T \left((\sigma^{-1}\tilde{b})(X_t^x)\right)^i dB_t^i.$$

(iii) Since $E\left(\varphi(X_t^x)^2\right) < \infty$ we can conclude by Cauchy's inequality

$$\begin{aligned}
&\left|\frac{\hat{u}_t^x - u_t^x}{\varepsilon} - E\left(\varphi(X_t^x)\sum_{i=1}^n \int_0^T \left((\sigma^{-1}\tilde{b})(X_t^x)\right)^i dB_t^i\right)\right| \\
= \quad & \left|E\left(\varphi(X_t^x)\frac{(\xi_T^x - 1)}{\varepsilon}\right) - E\left(\varphi(X_t^x)\sum_{i=1}^n \int_0^T \left((\sigma^{-1}\tilde{b})(X_t^x)\right)^i dB_t^i\right)\right| \\
= \quad & \left|E\left(\varphi(X_t^x)\left(\frac{(\xi_T^x - 1)}{\varepsilon} - \sum_{i=1}^n \int_0^T \left((\sigma^{-1}\tilde{b})(X_t^x)\right)^i dB_t^i\right)\right)\right| \\
\leq \quad & \left(E\left(\varphi(X_t^x)^2\right)\right)^{\frac{1}{2}} \left(E\left(\left(\frac{(\xi_T^x - 1)}{\varepsilon} - \sum_{i=1}^n \int_0^T \left((\sigma^{-1}\tilde{b})(X_t^x)\right)^i dB_t^i\right)^2\right)\right)^{\frac{1}{2}} \\
\longrightarrow \quad & 0
\end{aligned}$$

for $\varepsilon \to 0$ which yields by $\lim\limits_{\varepsilon\to 0} \frac{\hat{u}_t^x - u_t^x}{\varepsilon} = \frac{\partial \hat{u}_t^x}{\partial \varepsilon}\Big|_{\varepsilon=0}$ the result. ■

That we could calculate the rho in a Girsanov way was a direct consequence of the no-arbitrage assumption (and hence the existence of an equivalent martingale measure which annihilates the drift), in the case of the delta and the vega we have to find other methods. Here Malliavin calculus enters the scene and will show that it is the perfect tool to calculate the Greeks.

5.3.2 The Delta

For the delta we consider the case of a payoff function $\varphi : \mathbb{R}^{n\times m} \to \mathbb{R}$ depending on m points in time, $0 < t_1 \leq ... \leq t_m < T$, requiring obviously

$$E\left(\varphi\left(X_{t_1}^x, ..., X_{t_m}^x\right)^2\right) < \infty.$$

Then the price of the derivative is given by

$$u_t^x := E\left(\varphi\left(X_{t_1}^x, ..., X_{t_m}^x\right)\right)$$

and for the set Γ_m^Δ of square integrable functions with integral up to the times t_i equal one, i.e.

$$\Gamma_m^\Delta := \left\{a \in L^2([0,T]) : \int_0^{t_i} a_t dt = 1,\, i \in \{1, ..., m\}\right\}$$

we get the following result for the Delta:

Proposition 5.3.3 (Delta)
Assuming σ uniformly elliptic it holds for any $a \in \Gamma_m^\Delta$ that

$$\Delta_t^x = \nabla u_t^x = E\left(\varphi(X_{t_1}^x, ...X_{t_m}^x)\sum_{i=1}^n \int_0^T a_t \left(\left(\sigma^{-1}(X_t^x)\right)^i Y_t^x\right)^\tau dB_t^i\right).$$

Proof In a first step we show that the result for the case that $\varphi \in C^1(\mathbb{R}^{n\times m}, \mathbb{R})$ and $\|\nabla\varphi\| < \infty$ where ∇_j will denote the partial derivative with respect to the j-th argument.

(i) We note that by Theorem 4.5.2 follows immediately for $a \in \Gamma_m^{\Delta}$

$$\begin{aligned}\int_0^T (D_t X_{t_i}^x) a_t \sigma^{-1}(X_t^x) Y_t^x dt &= \int_0^{t_i} Y_{t_i}^x (Y_t^x)^{-1} \sigma(X_t^x) a_t \sigma^{-1}(X_t^x) Y_t^x dt \\ &= \int_0^{t_i} Y_{t_i}^x a_t dt = Y_{t_i}^x \int_0^{t_i} a_t dt = Y_{t_i}^x. \end{aligned} \tag{5.1}$$

(ii) Our aim is to express the gradient of u_t^x by differentiating inside the expectation: Since φ is continuously differentiable we have for $h \in \mathbb{R}^n$, $|h| \to 0$ the a.s. convergence for the directional derivative

$$\frac{\varphi(X_{t_1}^{x+h}, ...X_{t_m}^{x+h}) - \varphi(X_{t_1}^x, ...X_{t_m}^x)}{\|h\|} - \frac{\left(\sum_{i=1}^m \left(\nabla_i^\tau \varphi(X_{t_1}^x, ...X_{t_m}^x) Y_{t_i}^x\right)\right) h}{\|h\|} \to 0 \tag{5.2}$$

Both terms are uniformly integrable; for the second one this is clear by the assumption of the bounded gradient, the first can be estimated by

$$\left\| \frac{\varphi(X_{t_1}^{x+h}, ...X_{t_m}^{x+h}) - \varphi(X_{t_1}^x, ...X_{t_m}^x)}{\|h\|} \right\| \le \|\nabla\varphi\| \sum_{k=1}^m \frac{\left\|X_{t_k}^{x+h} - X_{t_k}^x\right\|}{\|h\|} < \infty$$

where the right term is clearly u.i. by the boundedness of the gradient and the joint continuity of the first variation (Theorem 3.3.2, 3.3.3). So (5.2) converges (by Lebesgue's dominated convergence) also in the L^1-sense and we get the result.

(iii) Thereon we can now perform our Malliavin-theoretic calculations:

$$\begin{aligned}\nabla^\tau u_t^x &= E\left(\sum_{j=1}^m \nabla_j^\tau \varphi(X_{t_1}^x, ...X_{t_m}^x) Y_{t_j}^x\right) \\ &= E\left(\int_0^T \sum_{j=1}^m \nabla_j^\tau \varphi(X_{t_1}^x, ...X_{t_m}^x)(D_t X_{t_j}^x) a_t \sigma^{-1}(X_t^x) Y_t^x dt\right) \\ &= E\left(\int_0^T D_t \varphi(X_{t_1}^x, ...X_{t_m}^x) a_t \sigma^{-1}(X_t^x) Y_t^x dt\right) \\ &= E\left(\varphi(X_{t_1}^x, ...X_{t_m}^x)\delta\left(a_t \sigma^{-1}(X_t^x) Y_t^x\right)\right) \\ &= E\left(\varphi(X_{t_1}^x, ...X_{t_m}^x) \sum_{i=1}^n \int_0^T a_t \left(\sigma^{-1}(X_t^x)\right)^i Y_t^x dB_t^i\right)\end{aligned}$$

by (5.1), the chain rule of Malliavin calculus (Proposition 4.3.9) and the definition of the Skorohod integral, which could in fact - since for any concrete function a_t the process $a_t \sigma^{-1}(X_t^x) Y_t^x$ is progressively measurable - be written as classical Itô integral.

(iv) We have to generalize the result to arbitrary square-integrable φ's: Since C_0^∞ is dense in $L^2(\Omega, \mathcal{F}, P)$ (see begin of Section 4.3) we can choose a sequence φ_k therein converging to φ and define

$$u_k := E\left(\varphi_n(X_{t_1}^x, ..., X_{t_m}^x)\right).$$

Obviously we have for any initial value x

$$u_k \longrightarrow u,$$

on the other hand side for the u_k the proposition is already proved and we can conclude by the Cauchy-Schwarz inequality - using the abbreviation $\tilde{\varphi}^x := \varphi_k(X_{t_1}^x, ...X_{t_m}^x) - \varphi(X_{t_1}^x, ...X_{t_m}^x)$ - that

$$\begin{aligned} & \left| \nabla u_k - E\left(\varphi(X_{t_1}^x, ...X_{t_m}^x) \sum_{i=1}^n \int_0^T a_t \left(\left(\sigma^{-1}(X_t^x) \right)^i Y_t^x \right)^\tau dB_t^i \right) \right| \\ \leq & \left| E\left(\tilde{\varphi}^x \sum_{i=1}^n \int_0^T a_t \left(\left(\sigma^{-1}(X_t^x) \right)^i Y_t^x \right)^\tau dB_t^i \right) \right| \\ \leq & \left\| E\left(\tilde{\varphi}^x\right)^2 \right\|^{\frac{1}{2}} \left\| E\left(\sum_{i=1}^n \int_0^T a_t \left(\left(\sigma^{-1}(X_t^x) \right)^i Y_t^x \right)^\tau dB_t^i \right)^2 \right\|^{\frac{1}{2}} \\ \to & \ 0 \end{aligned}$$

uniformly on compacts for $k \to \infty$, so

$$\begin{aligned} u_k & \longrightarrow u \\ \nabla u_k & \longrightarrow E\left(\varphi(X_{t_1}^x, ...X_{t_m}^x) \sum_{i=1}^n \int_0^T a_t \left(\left(\sigma^{-1}(X_t^x) \right)^i Y_t^x \right)^\tau dB_t^i \right), \end{aligned}$$

hence u is continuously differentiable and the proposition holds in the general case as well. ■

By the same arguments we can obviously derive formulae for higher derivatives, for instance we can get the gamma, the second derivative with respect to the initial value.

5.3.3 The Vega

As for the previous Greeks we start by laying out the concrete framework, nearly the same as for delta. Additionally we define the set Γ_m^ν by

$$\Gamma_m^\nu := \left\{ a \in L^2([0,T]) : \int_{t_{i-1}}^{t_i} a_t dt = 1, \, i \in \{1, ..., m\}, \, t_0 := 0 \right\}$$

and the (volatility) perturbed process $\hat{X}_t^x$ by

$$d\hat{X}_t^x := b(\hat{X}_t^x)dt + \left(\sigma(\hat{X}_t^x) + \varepsilon\tilde{\sigma}(\hat{X}_t^x)\right) dB_t$$

where we have to require that the perturbed diffusion matrix $\sigma+\varepsilon\tilde{\sigma}$ also satisfies the uniform ellipticity condition:

$$\left(\left(\sigma(x) + \varepsilon\tilde{\sigma}(x)\right)\zeta\right)^\tau \left(\left(\sigma(x) + \varepsilon\tilde{\sigma}(x)\right)\zeta\right) \geq \eta|\zeta|^2.$$

Besides the first variation with respect to the initial value we define that with respect to ε by

$$d\hat{Z}_t^x := b'(\hat{X}_t^x)\hat{Z}_t^x dt + \left(\tilde{\sigma}(\hat{X}_t^x)\right)^i dB_t^i + \sum_{i=1}^n \left(\sigma'(\hat{X}_t^x) + \varepsilon\tilde{\sigma}(\hat{X}_t^x)\right)^i \hat{Z}_t^x dB_t^i$$

with initial value $\hat{Z}_0^x = 0$, the null column vector. $\hat{Z}_t^x$ exists uniquely due to Picard-Lindelöf and we will write Z_t^x for $\hat{Z}_t^x\Big|_{\varepsilon=0}$. Furthermore we define the processes β_t and $\tilde{\beta}_t^a$ by

$$\begin{aligned}
\beta_t &:= (Y_t^x)^{-1} Z_t^x \\
\tilde{\beta}_t^a &:= \sum_{i=1}^m a_t(\beta_{t_i} - \beta_{t_{i-1}})1_{[t_{i-1},t_i]}.
\end{aligned}$$

Proposition 5.3.4 (Vega)
Under the above assumptions the function $\varepsilon \to \hat{u}_t^x$ is differentiable in $\varepsilon = 0$ and it holds that

$$\nu_t^x = \frac{\partial \hat{u}_t^x}{\partial \varepsilon}\bigg|_{\varepsilon=0} = E\left(\varphi(X_{t_1}^x, ...X_{t_m}^x)\delta\left(\sigma^{-1}(X_t^x)Y_t^x\tilde{\beta}_T^a\right)\right).$$

Proof We note that it is enough to prove the proposition for $\varphi \in C_b^1(\mathbb{R}^{n\times m}, \mathbb{R})$ since the generalization is completely analogous to that of delta.

(i) We want to prove that $\beta_t \in \mathbb{D}^{1,2}$ (and hence also $\tilde{\beta}_T^a$): The inverse $(Y_t^x)^{-1}$ of the first variation process is given by the SDE

$$dW_t^x = W_t^x \left(-b'(X_t^x) + \sum_{i=1}^n \left((\sigma'(X_t^x))^i\right)^2\right) dt - W_t^x \sum_{i=1}^n (\sigma'(X_t^x))^i dB_t^i \quad (5.3)$$

with initial value $W_0^x = I_n$, since

$$\begin{aligned}
d\langle Y^x, W^x\rangle_t &= \left\langle \sum_{i=1}^n (\sigma'(X_t^x))^i Y_t^x dB_t^i, -W_t^x \sum_{i=1}^n (\sigma'(X_t^x))^i dB_t^i \right\rangle \\
&= -\sum_{i=1}^n (\sigma'(X_t^x))^i Y_t^x W_t^x (\sigma'(X_t^x))^i dt
\end{aligned}$$

and hence by partial integration (Proposition 2.4.30)

$$\begin{aligned}
dY_t^x W_t^x &= dY_0^x W_0^x + Y_t^x dW_t^x + W_t^x dY_t^x + d\langle Y^x, W^x\rangle_t \\
&= Y_t^x W_t^x \left(\sum_{i=1}^n (\sigma'(X_t^x))^i\right)^2 dt - \sum_{i=1}^n (\sigma'(X_t^x))^i Y_t^x W_t^x (\sigma'(X_t^x))^i dt
\end{aligned}$$

which implies that $W_t^x = (Y_t^x)^{-1}$ is a solution of (5.3) which is unique due to a Picard-Lindelöf argument. By the C^∞-boundedness of the vector fields are $(Y_t^x)^{-1}$ and Z_t^x in $\mathbb{D}^{1,2}$ and the Cauchy inequality

$$\left|(Y_t^x)^{-1} Z_t^x\right|^2 \le E\left(((Y_t^x)^{-1})^2\right) E\left((Z_t^x)^2\right)$$

yields - by the closedness of D (Corollary 4.3.6) - the result.

(ii) By Theorem 4.5.2 and the definition of $\tilde{\beta}_t^a$ it follows for any $a \in \Gamma_m^\nu$ that

$$\begin{aligned}\int_0^T D_t X_{t_i}^x \sigma^{-1}(X_t^x) Y_t^x \tilde{\beta}_T^a dt &= \int_0^{t_i} Y_{t_i}^x \tilde{\beta}_T^a dt = Y_{t_i}^x \sum_{k=1}^{i} \int_{t_{k-1}}^{t_k} a_t \left(\beta_{t_k} - \beta_{t_{k-1}}\right) dt \\ &= Y_{t_i}^x \sum_{k=1}^{i} \left(\beta_{t_k} - \beta_{t_{k-1}}\right) \int_{t_{k-1}}^{t_k} a_t dt = Y_{t_i}^x \beta_{t_i} = Z_{t_i}^x.\end{aligned}$$

(iii) Analogously to point (ii) of the proof of the delta formula we can - since $\hat{X}_t^x$ is as flow a continuously differentiable function (with respect to ε) - conclude in the L^1-sense

$$\left.\frac{\partial \hat{u}_t^x}{\partial \varepsilon}\right|_{\varepsilon=0} = E\left(\sum_{i=1}^{m} \nabla_i^\tau \varphi(X_{t_1}^x, ... X_{t_m}^x) Z_{t_i}^x\right).$$

(iv) To finish we will naturally use Malliavin calculus...

$$\begin{aligned}\left.\frac{\partial \hat{u}_t^x}{\partial \varepsilon}\right|_{\varepsilon=0} &= E\left(\sum_{i=1}^{m} \nabla_i^\tau \varphi(X_{t_1}^x, ... X_{t_m}^x) Z_{t_i}^x\right) \\ &= E\left(\int_0^T \sum_{i=1}^{m} \nabla_i^\tau \varphi(X_{t_1}^x, ... X_{t_m}^x) D_t X_{t_i}^x \sigma^{-1}(X_t^x) Y_t^x \tilde{\beta}_T^a dt\right) \\ &= E\left(\int_0^T D_t \varphi(X_{t_1}^x, ... X_{t_m}^x) \sigma^{-1}(X_t^x) Y_t^x \tilde{\beta}_T^a dt\right) \\ &= E\left(\varphi(X_{t_1}^x, ... X_{t_m}^x) \delta\left(\sigma^{-1}(X_t^x) Y_t^x \tilde{\beta}_T^a\right)\right)\end{aligned}$$

by (ii), the chain rule (Proposition 4.3.9) and since $\tilde{\beta}_T^a \in \mathbb{D}^{1,2}$ (by (i)). Furthermore $\sigma^{-1}(X_t^x) Y_t^x \in \mathbb{D}^{1,2}$ and

$$E\left(\left(\tilde{\beta}_T^a\right)^2 \left(\delta\left(\sigma^{-1}(X_t^x) Y_t^x\right)\right)^2\right) \le \left(E\left(\tilde{\beta}_T^a\right)\right)^2 \left(E\left(\delta\left(\sigma^{-1}(X_t^x) Y_t^x\right)\right)\right)^2 < \infty$$

assures by 4.4.9 the Skorohod-integrability of the kernel. ■

The Skorohod integral is not an Itô integral since $\tilde{\beta}_T^a$ is $\{\mathcal{F}_T\}$-measurable though not $\{\mathcal{F}_t\}$-adapted. But the Proposition 4.4.9 even assures us a representation of the Skorohod integral in terms of classical Itô-calculus, we have

$$\begin{aligned}
&\delta\left(\sigma^{-1}(X_t^x)Y_t^x\tilde{\beta}_T^a\right)\\
= \quad &\left(\tilde{\beta}_T^a\right)^\tau \sum_{i=1}^n \int_0^T \left(\sigma^{-1}(X_t^x)\right)^i Y_t^x dB_t^i - \sum_{i=1}^n\sum_{j=1}^n \int_0^T \left(D_t\tilde{\beta}_T^a\right)^{ij} \left(\sigma^{-1}(X_t^x)Y_t^x\right)^{ji} dt\\
= \quad &\left(\tilde{\beta}_T^a\right)^\tau \sum_{i=1}^n \int_0^T \left(\sigma^{-1}(X_t^x)\right)^i Y_t^x dB_t^i - \int_0^T \operatorname{tr}\left(D_t\tilde{\beta}_T^a\sigma^{-1}(X_t^x)Y_t^x\right) dt.
\end{aligned}$$

We remark here that we calculated the delta and the vega for φ depending on finitely many points. That one can not generalize these proofs straight forward to continuous dependencies is clear by looking at sequences $t_{1_k} \to 0$ for delta and $t_{i_k} \to t_i$ for vega, $k \to \infty$.

5.4 Beyond Ellipticity

In this chapter we will drop the ellipticity assumption what means not only to give up the uniform ellipticity condition - which assured us the necessary invertibility of the matrices - but also to admit arbitrary (hence d) Brownian motions. (Remark that in the last chapter the number of Brownian motions was exactly the dimension of the process to produce (invertible) square ($n \times n$) matrices).

For commodity we will here use Stratonovich calculus, hence the SDE is given by

$$dX_t^x = b(X_t^x)dt + \sum_{i=1}^d \left(\sigma(X_t^x)\right)^i \circ dB_t^i,$$

b and σ^i again C^∞-bounded. Further we define the sets

$$\begin{aligned}
\mathcal{V}^0 \quad &:= \quad \left\{\sigma^1, ..., \sigma^d\right\}\\
\mathcal{V}^k \quad &:= \quad \left\{[V^i, V^j] : V^i \in \mathcal{V}^r, V^j \in \mathcal{V}^s, 0 \le r \le s < k\right\}\\
\mathcal{V}^\infty \quad &:= \quad \bigcup_{k=0}^\infty \mathcal{V}^k
\end{aligned}$$

and assume that the span of $\mathcal{V}^\infty$ has constant rank

$$\operatorname{rk}\left(\langle\mathcal{V}^\infty\rangle\right) = R \le n$$

and $b \in \langle\mathcal{V}^\infty\rangle$.

So the first variation as $n \times n$ matrix exists by a Picard-Lindelöf argument and it is given by

$$dY_t^x = b'(X_t^x)Y_t^x dt + \sum_{i=1}^d \left(\sigma'(X_t^x)\right)^i Y_t^x \circ dB_t^i;$$

by the same considerations as in part (i) of the proof of Proposition 5.3.4 the inverse $(Y_t^x)^{-1}$ exists a.s. In this context the Malliavin derivative $D_s X_t^x$ is a $n \times d$ matrix which satisfies

$$D_s X_t^x = Y_t^x (Y_s^x)^{-1} \sigma(X_s^x) 1_{[0,t]}(s) \tag{5.4}$$

and conversely the Skorohod integral of a $d \times n$ (Skorohod integrable) matrix is an n-dimensional column vector.

We remember the Malliavin matrix γ of Theorem 4.5.5,

$$\gamma = \gamma_t := \left(\left\langle D_s (X_t^x)^i, D_s (X_t^x)^j \right\rangle\right)_{ij} = \int_0^t (D_s X_t^x)(D_s X_t^x)^\tau ds,$$

which is for $R = n$ by [Nua] p.116, Lemma 2.3.1 invertible, so is the matrix C_t defined

$$C_t := \int_0^t \left((Y_s^x)^{-1} \sigma(X_s^x)\right) \left((Y_s^x)^{-1} \sigma(X_s^x)\right)^\tau ds$$

which is via invertible matrices related to γ by (5.4)

$$\gamma_t = Y_t^x C_t (Y_t^x)^\tau. \tag{5.5}$$

As the last point we introduce for fixed t the set $\mathbb{A}_t^x$ by

$$\begin{aligned} \mathbb{A}_t^x &:= \left\{ a \in \mathcal{D}^{1,2} : \int_0^t (Y_s^x)^{-1} \sigma(X_s^x) a_s ds = I_n \right\} \\ &= \left\{ a \in \mathcal{D}^{1,2} : \int_0^t (D_s X_t^x) a_s ds = Y_t^x \right\} \end{aligned} \tag{5.6}$$

again by (5.4) which consists of $d \times n$ matrix valued processes satisfying the integral condition. Now we can present the central theorem:

Theorem 5.4.1 (Hypoellliptic Delta)
If $R = n$, then $\mathbb{A}_t^x \neq \emptyset$ for all $x \in \mathbb{R}^n$, $t > 0$ and it holds for $\varphi : \mathbb{R}^n \to \mathbb{R}$ that

$$\Delta_t^x = E\left(\varphi(X_t^x)\left(\delta\left(\left((Y_s^x)^{-1}\sigma(X_s^x)\right)^\tau C_t^{-1}\right)\right)^\tau\right).$$

Proof
(i) The non-emptiness of $\mathbb{A}_t^x$ we show directly by stating an element whose existence is clear:

$$a_s := (D_s X_t^x)^\tau \gamma_t^{-1} Y_t^x.$$

This is an element of $\mathbb{A}_t^x$ according to (5.4) since

$$\begin{aligned} \int_0^t (D_s X_t^x)(D_s X_t^x)^\tau \gamma_t^{-1} Y_t^x ds &= \int_0^t (D_s X_t^x)(D_s X_t^x)^\tau ds \gamma_t^{-1} Y_t^x \\ &= \gamma_t \gamma_t^{-1} Y_t^x = Y_t^x. \end{aligned}$$

(ii) Using (5.3) and (5.4) we can write

$$\begin{aligned} a_s &= (D_s X_t^x)^\tau \gamma_t^{-1} Y_t^x \\ &= (Y_t^x (Y_s^x)^{-1} \sigma(X_s^x))^\tau \left(Y_t^x C_t (Y_t^x)^\tau\right)^{-1} Y_t^x \\ &= \left((Y_s^x)^{-1} \sigma(X_s^x)\right)^\tau C_t^{-1}. \end{aligned}$$

(iii) Since there are no differences to the elliptic case of the delta (Proposition 5.2.3(ii)) we can interchange gradient and expectation and hence

$$\begin{aligned} \nabla^\tau u_t^x &= E\left(\nabla^\tau \varphi(X_t^x) Y_t^x\right) \\ &= E\left(\nabla^\tau \varphi(X_t^x) \int_0^t (D_s X_t^x) a_s ds\right) \\ &= E\left(\int_0^t \nabla^\tau \varphi(X_t^x)(D_s X_t^x) a_s ds\right) \\ &= E\left(\int_0^t (D_s \varphi(X_t^x))\, a_s ds\right) \\ &= E\left(\varphi(X_t^x) \delta(a_s)\right) \end{aligned}$$

and so by (ii)

$$\Delta_t^x = E\left(\varphi(X_t^x)\,(\delta(a_s))^\tau\right) = E\left(\varphi(X_t^x)\left(\delta\left(\left((Y_s^x)^{-1}\sigma(X_s^x)\right)^\tau C_t^{-1}\right)\right)^\tau\right)$$

■

Solving the Skorohod integral we get

$$\begin{aligned} & \delta\left(\left((Y_s^x)^{-1}\sigma(X_s^x)\right)^\tau C_t^{-1}\right) \\ = \; & \left(C_t^{-1}\right)^\tau \int_0^t \left((Y_s^x)^{-1}\sigma(X_s^x)\right)^\tau dB_s^i \\ & - \sum_{i=1}^n \sum_{j=1}^d \int_0^t \left(D_s (C_t^{-1})^\tau\right)^{kij} \left(\left((Y_s^x)^{-1}\sigma(X_s^x)\right)^\tau\right)^{ji} ds \\ = \; & \left(C_t^{-1}\right)^\tau \int_0^t \left((Y_s^x)^{-1}\sigma(X_s^x)\right)^\tau dB_s^i \\ & - \int_0^t \operatorname{tr}\left(\left(D_s (C_t^{-1})^\tau\right)\left((Y_s^x)^{-1}\sigma(X_s^x)\right)^\tau\right) ds. \end{aligned}$$

5.5 Examples

This last section will be devoted to concrete examples of the calculation of the Greeks.

Example 5.5.1 (Black-Scholes Delta)
In the Black Scholes framework the underlying is described by the SDE

$$dX_t^x = r(t)X_t^x dt + \sigma X_t^x dB_t$$

and in the case of an European call option with strike price K the function φ is given by

$$\varphi(X_t^x) := e^{-\int_0^T r(t)dt}(X_t^x - K)_+.$$

Calculate the delta of this option.

Calculation The first variation process is described by the SDE

$$dY_t^x = r(t)Y_t^x dt + \sigma Y_t^x dB_t,$$

which is identical with that of the X_t^x, only the initial value is different and hence $Y_t^x = \frac{1}{x}X_t^x$. So we can calculate by Proposition 5.3.3

$$\begin{aligned}
\Delta_T^x &= E\left(\varphi(X_T^x)\int_0^T a_t\sigma^{-1}(X_t^x)Y_t^x dB_t\right) \\
&= E\left(e^{-\int_0^T r(t)dt}(X_T^x - K)_+\int_0^T \frac{a_t}{\sigma x}dB_t\right) \\
&= E\left(e^{-\int_0^T r(t)dt}\frac{B_T}{\sigma T x}(X_T^x - K)_+\right)
\end{aligned}$$

for the (quite simple) choice $a_t = \frac{1}{T}$. ∎

Example 5.5.2 (Black-Scholes Vega)
Calculate in the above setting the vega.

Calculation As above we set $a_t = \frac{1}{T}$ and have $Y_t^x = \frac{1}{x}X_t^x$; for the first variation with respect to ε we get the inhomogeneous SDE

$$dZ_t^x = r(t)Z_t^x dt + \tilde{\sigma}X_t^x dB_t + \sigma' Z_t^x dB_t$$

which we can solve by variation of the constant to get $Z_t^x = \tilde{\sigma}X_t^x B_t$. So we get

$$\begin{aligned}
\beta_t &= (Y_t^x)^{-1} Z_t^x = x\tilde{\sigma}B_t \\
\tilde{\beta}_t^{\frac{1}{T}} &= \frac{x\tilde{\sigma}}{T}B_t 1_{[0,T]}.
\end{aligned}$$

So the vega is given by

$$\begin{aligned}\nu_t^x &= E\left(\varphi(X_t^x)\delta\left(\sigma^{-1}(X_t^x)Y_t^x\tilde{\beta}_T^a\right)\right)\\ &= E\left(e^{-\int\limits_0^T r(t)dt}(X_T^x-K)_+\delta\left(\frac{1}{\sigma X_t^x}\frac{1}{x}X_t^x\frac{x\tilde{\sigma}}{T}B_t 1_{[0,T]}\right)\right)\\ &= E\left(e^{-\int\limits_0^T r(t)dt}(X_T^x-K)_+\delta\left(\frac{\tilde{\sigma}}{\sigma T}B_t 1_{[0,T]}\right)\right)\end{aligned}$$

and for the explicit calculation of the Skorohod integral we get $D\tilde{\beta}_t^{\frac{1}{T}} = \frac{x\tilde{\sigma}}{T}1_{[0,T]}$, so

$$\begin{aligned}\delta\left(\frac{\tilde{\sigma}}{\sigma T}B_t 1_{[0,T]}\right) &= \frac{x\tilde{\sigma}}{T}1_{[0,T]}\int\limits_0^T \frac{1}{\sigma X_t^x}\frac{1}{x}X_t^x dB_t - \int\limits_0^T \frac{x\tilde{\sigma}}{T}\frac{1}{\sigma X_t^x}\frac{1}{x}X_t^x dt\\ &= \tilde{\sigma}\left(\frac{B_t^2}{\sigma T}-\frac{1}{\sigma}\right),\end{aligned}$$

whence

$$\nu_t^x = E\left(e^{-\int\limits_0^T r(t)dt}(X_T^x-K)_+\tilde{\sigma}\left(\frac{B_t^2}{\sigma T}-\frac{1}{\sigma}\right)\right).$$

■

As last application of the Malliavin-based calculation of the Greeks we give an Hobson-Rogers type example. We denote by $\gamma(Z_t^{z,\varepsilon})$ the Malliavin matrix of the process $Z_t^{z,\varepsilon}$ and assume that $\gamma^{-1}(Z_t^{z,\varepsilon}) \in \mathbb{D}^\infty$ exists. Further we define the set

$$\mathbb{A}_t^z := \left\{ a \in \mathcal{D}_{1,2} : \int\limits_0^t D_u Z_t^{z,\varepsilon} D_u Z_t^{z,\varepsilon}\gamma^{-1}(Z_t^{z,\varepsilon})a_u du = 1 \right\}$$

which is not empty since $a_u = 1_{[0,t]}(u) \in \mathbb{A}_t^z$ by the definition of the Malliavin matrix. For the Hobson-Rogers calculation we yet need a general representation theorem:

Theorem 5.5.3 *Given* $(X_t^{x,\varepsilon})_{t\geq 0}$ *a real valued process with (also with respect to* ε*)* C^∞*-bounded coefficients, such that the respective iterated Skorohod integrals of* $G^{(i)}D_u X_t^{x,0}\gamma^{-1}(X_t^{x,0})a_u$ *which appear in the proof exist. Then for all* $\varphi \in C_0^\infty(\mathbb{R})$ *we have*

$$\frac{\partial^n}{\partial\varepsilon^n}E(\varphi(X_t^{x,\varepsilon}))|_{\varepsilon=0} = E(\varphi(X_t^{x,0})\pi^n).$$

Proof By the Faa di Bruno formula (see [Mi 97], p.15f) we can expand

$$\begin{aligned}
&\frac{\partial^n}{\partial \varepsilon^n} E\left(\varphi\left(X_t^{x,\varepsilon}\right)\right)\Big|_{\varepsilon=0} \\
= & \sum_{i=0}^n E\left(\varphi^{(i)}\left(X_t^{x,0}\right) G^{(i)}\right) \\
= & \sum_{i=0}^n E\left(\varphi^{(i)}\left(X_t^{x,0}\right) G^{(i)} \int_0^t D_u X_t^{x,0} D_u X_t^{x,0} \gamma^{-1}(X_t^{x,0}) a_u du\right) \\
= & \sum_{i=1}^n E\left(\int_0^t D_u \varphi^{(i-1)}\left(X_t^{x,0}\right) G^{(i)} D_u X_t^{x,0} \gamma^{-1}(X_t^{x,0}) a_u du\right) \\
& + E\left(\varphi\left(X_t^{x,0}\right) G^{(0)}\right) \\
= & \sum_{i=1}^n E\left(\varphi^{(i-1)}\left(X_t^{x,0}\right) \delta\left(G^{(i)} D_u X_t^{x,0} \gamma^{-1}(X_t^{x,0}) a_u\right)\right) + E\left(\varphi\left(X_t^{x,0}\right) G^{(0)}\right)
\end{aligned}$$

for some $G^{(i)}$ and the usual Malliavin calculation. Iteration yields the result. ■

Corollary 5.5.4 *Given $(X_t^{x,\varepsilon})_{t\geq 0}$ a real valued process with (also with respect to ε) real analytic coefficients, such that the respective Skorohod integrals exist as above. Take a bounded real analytic function with bounded derivatives $\varphi : \mathbb{R} \to \mathbb{R}$, then*

$$E(\varphi(X_t^{x,\varepsilon})) = \sum_{n\geq 0} \frac{\varepsilon^n}{n!} E(\varphi(X_t^{x,0})\pi^n)$$

for small $\varepsilon > 0$ (the seize of the neighborhood might depend on t and x).

Proof By the Cauchy-Kowalewsky Theorem (see [Hö 83], p.348f) we know the the associated parabolic initial value problem has a real analytic solution, which coincides a fortiori with $E(\varphi(X_t^{x,\varepsilon}))$. Hence the Taylor series converges locally and the above representation holds. ■

If we want to prove the convergence of the series for more general payoff functions ϕ we can apply the following sufficient conditions:

Theorem 5.5.5 *Given $(X_t^{x,\varepsilon})_{t\geq 0}$ a real valued process with (also with respect to ε) real analytic coefficients, such that the respective Skorohod integrals exist as above. Assume furthermore that the (universally calculated) weights π^n satisfy*

$$\sum_{n\geq 0} \frac{\varepsilon_0^n}{n!} E((\pi^n)^2)^{\frac{1}{2}} < \infty$$

for some $\varepsilon_0 > 0$, then for all bounded measurable $\varphi : \mathbb{R} \to \mathbb{R}$ we obtain

$$E(\varphi(X_t^{x,\varepsilon})) = \sum_{n\geq 0} \frac{\varepsilon^n}{n!} E(\varphi(X_t^{x,0})\pi^n)$$

for $\varepsilon < \varepsilon_0$.

Proof Take a sequence of bounded real analytic ϕ_k with bounded derivatives such that $E((\varphi_k - \varphi)^2(X_t^{x,\varepsilon})) \to 0$ as $k \to \infty$ for small ε, then by the Cauchy-Schwarz inequality

$$\begin{aligned}|E(\varphi(X_t^{x,\varepsilon}) - \varphi_k(X_t^{x,\varepsilon}))| &\leq \sum_{n\geq 0} \frac{\varepsilon^n}{n!} |E((\varphi(X_t^{x,0}) - \varphi_k(X_t^{x,\varepsilon}))\pi^n)| \\ &\leq E((\varphi_k - \varphi)^2(X_t^{x,\varepsilon}))^{\frac{1}{2}} \sum_{n\geq 0} \frac{\varepsilon_0^n}{n!} E((\pi^n)^2)^{\frac{1}{2}} \to 0\end{aligned}$$

as $n \to \infty$. ■

A particular feature of the above considerations is that we do not need the integrability assumptions on the Malliavin covariance matrix off 0. If we are able to calculate the π^n (which involve terms at $\epsilon = 0$) and prove a regularity assumption on $\sum_{n\geq 0} \frac{\varepsilon_0^n}{n!} E((\pi^n)^2)^{\frac{1}{2}}$, we are able to show approximation results of the above type for $E(\varphi(X_t^{x,\varepsilon}))$.

Similar reasonings can be applied for the calculation of the Greeks: here we consider precisely the same setting, only in a two-dimensional framework, since we consider the process $(X_t^{x,\varepsilon}, \frac{d}{dx} X_t^{x,\varepsilon})_{t\geq 0}$. By the Cauchy-Kowalewsky Theorem we are able to conclude that for real analytic φ with bounded derivatives the expansion

$$\frac{d}{dx} E(\varphi(X_t^{x,\varepsilon})) = \sum_{n\geq 0} \frac{\varepsilon^n}{n!} E(\varphi(X_t^{x,0})\rho^n)$$

converges, where the ρ^n are again given by Skorohod integrals which can be calculated by the convergence in Corollary 5.54. Finally also the considerations with respect to bounded measurable functions apply.

Remark 5.5.6 *If we know from semigroup considerations (analyticity with respect to parameters) that for all bounded measurable φ the series*

$$E(\varphi(X_t^{x,\varepsilon})) = \sum_{n\geq 0} \frac{\varepsilon^n}{n!} \frac{\partial^n}{\partial \varepsilon^n} E(\varphi(X_t^{x,\varepsilon}))|_{\varepsilon=0}$$

converges for small ε, then we can conclude directly

$$\frac{\partial^n}{\partial \varepsilon^n} E(\varphi(X_t^{x,\varepsilon}))|_{\varepsilon=0} = E(\varphi(X_t^{x,0})\pi^n),$$

by the Cauchy-Schwarz inequality and the uniform convergence of

$$\frac{\partial^n}{\partial \varepsilon^n} E(\varphi_k(X_t^{x,\varepsilon}))$$

for $\{\varphi_k\}_{k\geq 1}$ real analytic with bounded derivatives,

$$E\left((\varphi_k - \varphi)^2(X_t^{x,\varepsilon})\right) \longrightarrow 0,$$

$k \to \infty$, for x, t fixed in a small interval in ε.

This fact we can use in the following example:

Example 5.5.7 (Delta for Hobson Rogers Approximation)
Calculate the delta of Z_t^x in an approximated Hobson-Rogers framework, i.e.

$$\begin{pmatrix} dZ_t^z \\ dS_t^s \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}\left(\sigma(S_t^s)\right)^2 \\ -\left(\frac{1}{2}\eta^2 + \lambda S_t^s\right) \end{pmatrix} dt + \begin{pmatrix} \sigma(S_t^s) \\ \eta \end{pmatrix} dB_t$$

with $\sigma(S_t^s) = \eta\left(1 + \frac{1}{2}\varepsilon(S_t^s)^2\right) e^{-\frac{\varepsilon(S_t^s)^2}{k}}$ for large k describes the underlying, for differentiable φ.

Calculation
The SDE reads

$$\begin{pmatrix} dZ_t^{z,\varepsilon} \\ dS_t^s \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}\eta^2\left(1 + \varepsilon(S_t^s)^2 + \frac{1}{4}\varepsilon^2(S_t^s)^4\right) e^{-\frac{2\varepsilon(S_t^s)^2}{k}} \\ -\frac{1}{2}\eta^2 - \lambda S_t^s \end{pmatrix} dt + \begin{pmatrix} \eta\left(1 + \frac{1}{2}\varepsilon(S_t^s)^2\right) e^{-\frac{2\varepsilon(S_t^s)^2}{k}} \\ \eta \end{pmatrix} dB_t, \tag{5.7}$$

this system is globally solvable since the second SDE is an inhomogeneous SDE with nice coefficients and for the first one hence the integrals exist. Since $\frac{\partial}{\partial z} Z_t^{z,\varepsilon} = 1$ and the real analytic coefficients allow us to use the above theorem, we can hence calculate directly the delta:

$$\begin{aligned} \Delta_t^{z,\varepsilon} &= \frac{\partial}{\partial z} E\left(\varphi(Z_t^{z,\varepsilon})\right) \\ &= E\left(\varphi'(Z_t^{z,\varepsilon}) \frac{\partial}{\partial z} Z_t^{z,\varepsilon}\right) \\ &= E\left(\varphi'(Z_t^{z,\varepsilon}) \int_0^t D_u Z_t^{z,\varepsilon} D_u Z_t^{z,\varepsilon} \gamma^{-1}(Z_t^{z,\varepsilon}) a_u du\right) \\ &= E\left(\int_0^t D_u \varphi(Z_t^{z,\varepsilon}) D_u Z_t^{z,\varepsilon} \gamma^{-1}(Z_t^{z,\varepsilon}) a_u du\right) \\ &= E\left(\varphi(Z_t^{z,\varepsilon}) \delta\left(D_u Z_t^{z,\varepsilon} \gamma^{-1}(Z_t^{z,\varepsilon}) a_u\right)\right). \end{aligned} \tag{5.8}$$

Now we have to show that we can calculate the ingredients of the Skorohod integral, which means essentially $D_u Z_t^{z,\varepsilon}$ (which is also contained in $\gamma(Z_t^{z,\varepsilon})$). From

$$\begin{aligned} Z_t^{z,\varepsilon} &= z - \frac{1}{2}\eta^2 \int_0^t e^{-\frac{2\varepsilon(S_r^s)^2}{k}} dr - \frac{1}{2}\eta^2\varepsilon \int_0^t (S_r^s)^2 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} dr \\ &\quad - \frac{1}{8}\eta^2\varepsilon^2 \int_0^t (S_r^s)^4 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} dr + \eta \int_0^t e^{-\frac{2\varepsilon(S_r^s)^2}{k}} dB_r \\ &\quad + \frac{1}{2}\eta\varepsilon \int_0^t (S_r^s)^2 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} dB_r \end{aligned}$$

follows directly by Malliavin differentiation that

$$\begin{aligned} D_u Z_t^{z,\varepsilon} &= \frac{2\eta^2\varepsilon}{k}\int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dr - \eta^2\varepsilon \int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dr \\ &+\frac{2\eta^2\varepsilon^2}{k}\int_u^t (S_r^s)^3 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dr - \frac{\eta^2\varepsilon^2}{2}\int_u^t (S_r^s)^3 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dr \\ &+ \frac{\eta^2\varepsilon^2}{2k}\int_u^t (S_r^s)^5 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dr - \frac{4\eta\varepsilon}{k}\int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dB_r \\ &+ \eta e^{-\frac{2\varepsilon(S_r^s)^2}{k}} 1_{[0,t]}(u) + \eta\varepsilon \int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dB_r \\ &+\frac{2\eta\varepsilon^2}{k}\int_u^t (S_r^s)^3 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dB_r + \frac{\eta\varepsilon}{2}(S_u^s)^2 e^{-\frac{2\varepsilon(S_u^s)^2}{k}} 1_{[0,t]}, \end{aligned} \tag{5.9}$$

wherefore we have

$$S_t^s = e^{-\lambda t}s - \int_0^t e^{-\lambda(t-r)}\frac{\eta^2}{2}dr + \int_0^t e^{-\lambda(t-r)}\eta dB_r$$

as solution of an inhomogeneous SDE and so

$$D_u S_t^s = \eta e^{-\lambda(t-u)} 1_{[0,t]}(u).$$

As last point it remains to give an approach for a direct calculation of the stochastic integrals of (5.9): By the second SDE we have the substitution

$$dB_t = \frac{1}{\eta}dS_t^s + \frac{\eta}{2}dt + \frac{\lambda}{\eta}S_t^s dt,$$

so

$$\begin{aligned} \frac{4\eta\varepsilon}{k}\int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dB_r &= \frac{2\eta^2\varepsilon}{k}\int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dr \\ &+\frac{4\lambda\varepsilon}{k}\int_u^t (S_r^s)^2 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dr \\ &+\frac{4\varepsilon}{k}\int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dS_r^s, \end{aligned}$$

where the last integral can be calculated by the Itô formula

$$\begin{aligned}
\frac{4\varepsilon}{k}\int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} D_u S_r^s dS_r^s &= \frac{4\eta\varepsilon}{k}\int_u^t S_r^s e^{-\frac{2\varepsilon(S_r^s)^2}{k}} e^{-\lambda(t-u)} dS_r^s \\
&= -\eta e^{-\frac{2\varepsilon(S_t^s)^2}{k}} e^{-\lambda(t-u)} + \eta e^{-\frac{2\varepsilon(S_0^s)^2}{k}} e^{-\lambda(t-u)} \\
&\quad + \frac{16\eta^3\varepsilon^2}{k^2}\int_u^t (S_r^s)^2 e^{-\frac{2\varepsilon(S_r^s)^2}{k}} e^{-\lambda(t-u)} dr,
\end{aligned}$$

since $d\langle S_r^s, S_r^s\rangle = \eta^2 dr$. Analogous considerations can be applied for the other two stochastic integrals of (5.9). ■

Bibliography

[Ba 95] BASS, Richard F. Probabilistic Techniques in Analysis. New York (Springer) 1995

[Ba 98] BASS, Richard F. Diffusions and Elliptic Operators. New York (Springer) 1998

[Bor 95] BORKAR, Vivek S. Probability Theory. An Advanced Course. New York (Springer) 1995

[ChW 83] CHUNG, Kai Lai and WILLIAMS, Ruth J. Introduction to Stochastic Integration. Boston (Birkhäuser) 1983

[DM 78] DELLACHERIE, Claude and MEYER, Paul-André. Probabilities and Potential (Chapter 0 - IV). Paris (Hermann) and Amsterdam (North Holland) 1978

[DM 80] DELLACHERIE, Claude and MEYER, Paul-André. Probabilités et Potentiel (Chapitre V - VIII). Paris (Hermann) 1980

[Eth 02] ETHERIDGE, Alison. Financial Calculus. Cambridge (Cambridge University Press) 2002

[FLLLT 99] FOURNIÉ, Eric, LASRY, Jean-Michel, LEBUCHOUX, Jérôme, LIONS, Pierre-Louis and TOUZI, Nizar. Application of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics, 3* (1999), 391-412

[FLLL 01] FOURNIÉ, Eric, LASRY, Jean-Michel, LEBUCHOUX, Jérôme and LIONS, Pierre-Louis. Application of Malliavin calculus to Monte Carlo methods in finance. II. *Finance and Stochastics, 5* (2001), 201-236

[GM 02] GOBET, Emmanuel and MUNOS, Rémi. Sensitivity Analysis using Itô-Malliavin Calculus and Martingales. Application to Stochastic Optimal Control. 2002 (preprint)

[Hö 83] HÖRMANDER, Lars. The Analysis of Partial Differential Operators I. Distribution Theory and Fourier Analysis. Berlin (Springer) 1983

[HTT] HUBALEK, Friedrich, TEICHMANN, Josef and TOMPKINS, Robert. Flexible Complete Models with Stochastic Volatility Generalising Hobson & Rogers. 2003 (working paper)

[Im 97] IMKELLER, Peter. Der stochastische Variationskalkül und Anwendungen in der multiplikativen Ergodentheorie. Manuscript. 1997
(http://wws.mathematik.hu-berlin.de/~imkeller/teaching lectures/
Supplement/buckow.pdf)

[Kal 02] KALLENBERG, Olav. Foundations of Modern Probability. New York (Springer) ^{2}2002

[**KS 88**] KARATZAS, Ioannis and SHREVE, Steven E. Brownian Motion and Stochastic Calculus. New York (Springer) 1988

[**Mal 95**] MALLIAVIN, Paul e.a. Integration and Probability. New York (Springer) 1995

[**Mi 97**] MICHOR, Peter W. Transformation Groups. Lecture Notes of a course in Vienna. 1997 (http://www.mat.univie.ac.at/~michor/tgbook.ps)

[**MT 03**] MALLIAVIN, Paul and THALMAIER, Anton. Stochastic Calculus of Variations in Mathematical Finance. Preliminary draft of the first chapters. Berlin (Springer) 2003 (preprint)

[**Nua 95**] NUALART, David. The Malliavin Calculus and Related Topics. New York (Springer) 1995

[**Øks 85**] ØKSENDAL, Bernt Karsten. Stochastic Differential Equations. An Introduction with Applications. Berlin (Springer) 1985

[**Øks 97**] ØKSENDAL, Bernt. An Introduction to Malliavin Calculus with Applications to Economics. Manuscript. Oslo 1997 (http://www.nhh.no/for/dp/1996/wp0396.pdf)

[**Rud 87**] RUDIN, Walter. Real and Complex Analysis. New York (McGraw-Hill) ^{3}1987

[**Rud 91**] RUDIN, Walter. Functional Analysis. New York (McGraw-Hill) ^{2}1991

[**RY 91**] REVUZ, Daniel and YOR, Marc. Continuous Martingales and Brownian Motion. Berlin (Springer) 1991

[**Sch 99**] SCHAEFER, H.H. with WOLFF, M.P. Topological Vector Spaces. New York (Springer) ^{2}1999

[**Tei 02**] TEICHMANN, Josef. The Malliavin Calculus. Manuscript. Vienna 2002 (http://www.fam.tuwien.ac.at/~jteichma/lecture.html)

[**Tei 03**] TEICHMANN, Josef. Stochastic Analysis with Applications to Financial and Actuarial Mathematics. Manuscript. Vienna 2003 (http://www.fam.tuwien.ac.at/~jteichma/lecture.html)

[**TT**] TEICHMANN, Josef and TOUZI, Nizar. Calculating the Greeks for hypoelliptic Problems. (working paper)

[**Wat 84**] WATANABE, S. Stochastic Differential Equations and Malliavin Calculus. Bombay (Tata Institute of Fundamental Research) and Berlin (Springer) 1984

Contents