Jan Obłój

Continuous martingales and stochastic calculus

(notes in never-ending progress)

March 6, 2025

Preface

These notes accompany my lecture on Continuous martingales and Stochastic Calculus (B8.2). However, not having the strict time-limit imposed on a lecture course, they tend to go into various (interesting!) digressions and cover additional material which is meant to provide the reader with a "larger and clearer picture". In general I try to point out the parts of the material which are additional and are not covered in the lectures. However, this is not always possible.

I should stress that the material presented in the lectures is examinable – nothing less or more. PLEASE GO TO THE LECTURES AND TAKE YOUR OWN NOTES. Then read these as a supplement.

I first gave these lectures in HT 2014. It was the very first time stochastic calculus was taught as part of the undergraduate curriculum at Oxford. It was, and remains, a rather challenging course. At the same time, it covers exciting and beautiful mathematics so, all in all, I hope you enjoy it!

These notes are work in progress and are constantly being improved. I am very grateful to all who helped me to improve them. Your comments, corrections, but also questions in lectures, are precious.

Last but not least, I want to thank Zhaoxu Hou for his help with typing up these notes.

Oxford, January 2015 *Jan Obłój*

PS. Please send all your comments and corrections to `jan.obloj@maths.ox.ac.uk`. Many thanks in advance!

Contents

1 Introduction 1
- 1.1 Stochastic process - an introduction 1
- 1.2 Monotone class theorem 3

2 An overview of Gaussian variables and processes 5
- 2.1 Gaussian variables 5
- 2.2 Gaussian vectors and spaces 7
  - 2.2.1 Gaussian spaces 9
- 2.3 Gaussian processes 10
- 2.4 Constructing distributions on $\left(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)})\right)$ 11
- 2.5 Continuity of paths and Kolmogorov's criterion 12

3 Brownian Motion and First Properties 13
- 3.1 Definition of Brownian motion 13
- 3.2 Lévy's construction of Brownian motion on $[0,1]$ 14
- 3.3 Wiener Measure 17
- 3.4 Extensions and first properties 18
- 3.5 Basic properties of Brownian sample path 18

4 Filtrations, stopping times and first applications 23
- 4.1 Filtrations 23
- 4.2 Stopping times 24
- 4.3 Strong Markov property and reflection principle for Brownian motion 26

5 (Sub/super-)Martingales in continuous time 29
- 5.1 Definition 29
- 5.2 Maximal (Doob's) inequalities 30
- 5.3 Convergence and regularisation theorems 32
- 5.4 Optional sampling (stopping) theorem(s) 34

6 Continuous semimartingales (the integrators) 39
6.1 Functions of finite variation 39
6.2 Process of finite variation 43
6.3 Continuous local martingales 45
6.4 Quadratic variation of a continuous local martingale 47
6.5 Continuous semimartingales 53

7 Stochastic Integration 55
7.1 Stochastic integral w.r.t. L^2-bounded martingales 56
7.1.1 Construction of stochastic integral through a limiting procedure 56
7.1.2 Intrinsic characterisation of stochastic integrals using the quadratic co-variation 59
7.2 Extensions: stochastic integration with respect to continuous semimartingales 62
7.3 Itô's formula and its applications 66

A A very short primer in functional analysis 73
A.1 Normed vector spaces 73
A.2 Banach spaces 75

Chapter 1
Introduction

1.1 Stochastic process - an introduction

A stochastic (from Greek word $\sigma\tau\acute{o}\chi o\varsigma$, which means "aim") process is a random phenomenon evolving in time. Our lives, science or fiction give us numerous examples of such process, from stock prices, through temperature, population size to cloud shapes or ice levels in Antarctic. In this course we will study the mathematical theory required to properly define the mathematical representations of such phenomenon. We will develop tools necessary to make sense of differential equations driven by random process. These will be studied in part C.

Our basic example will be Brownian motion. First described by an English botanist R.Brown in 1827 it represents the chaotic movement of a pollen particle in a liquid resulting from constant hitting of the pollen by the much smaller molecules of the liquid. The physical theory of this motion was set up by Einstein in 1905 and mathematics were formalised by Wiener in 1923. We will study BM in some detail, see that its paths have infinite variation, are nowhere differentiable or monotone and yet we can sensibly integrate against them. BM also yields direct representation of solutions to partial differential equations, such as

$$\begin{cases} \frac{\partial}{\partial t}v(t,x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x) = 0, \quad t \leq T, \\ v(T,x) = \phi(x), \quad x \in \mathbb{R}. \end{cases} \tag{1.1}$$

Definition 1.1.1. Let $\mathscr{T}$ be a set and $(E,\mathscr{E})$ a measurable space. A stochastic process indexed by $\mathscr{T}$ and taking values in $(E,\mathscr{E})$ is a family of random variables $\big(X_t : t \in \mathscr{T}\big)$ defined on a common probability space $(\Omega,\mathscr{F},\mathbb{P})$ and valued in $(E,\mathscr{E})$ which is called the state space.

Although sometimes it is useful to consider rather general sets $\mathscr{T}$, here we focus on the interpretation of $t \in \mathscr{T}$ as time. In past, you considered $\mathscr{T} = \mathbb{N}$, here we consider $\mathscr{T} = [0,\infty)$ or $[0,T]$.

In this perspective the mapping $t \mapsto X_t(\omega)$, for a fixed $\omega \in \Omega$, represents a given realisation of the process, called a sample path or a trajectory.

Note that it is not clear (nor true in general) that $t \mapsto X_t(\omega) : [0,\infty) \mapsto \mathbb{R}$ is a measurable function. This is why we really want to see X as a function of two variables.

Definition 1.1.2. A stochastic process indexed by $[0,\infty)$ and valued in $\mathbb{R}$ (or $\mathbb{R}^d$) is called measurable if the mapping

$$(t,\omega) \mapsto X_t(\omega) : \big([0,\infty) \times \Omega, \mathscr{B}([0,\infty)) \otimes \mathscr{F}\big) \mapsto \big(\mathbb{R}, \mathscr{B}(\mathbb{R})\big)$$

is measurable (i.e. $\forall A \in \mathscr{B}(\mathbb{R})$ $\{(t,\omega) : X_t \in A\}$ is in the product σ-algebra $\mathscr{B}([0,\infty)) \otimes \mathscr{F}$.

It follows (part of Fubini's Thm) that its trajectories are Borel-measurable. We can now talk about two ways in which two processes can be close to one another.

Definition 1.1.3. Let X, Y be two stochastic processes defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
We say that X is a modification of Y if for all $t \geq 0$ we have $X_t = Y_t$ a.s.
We say that X and Y are indistinguishable if $\mathbb{P}(X_t = Y_t \text{ for all } 0 \leq t < \infty) = 1$.

If X and Y are modifications of one another than in particular they have the same finite dimensional distributions. Note however that indistinguishability is a much stronger property and implies that X and Y are modification of each other. It also takes the trajectory as our basic object of interest and implies we could see $(X_t : t \geq 0)$ as a random variable valued in $\mathbb{R}^{[0,\infty)}$. We need to then endow it with a σ-algebra.

Definition 1.1.4. An n-dimensional cylinder set in $\mathbb{R}^{[0,\infty)}$ is a set of the form

$$C = \big\{\omega \in \mathbb{R}^{[0,\infty)} : \big(\omega(t_1), \ldots, \omega(t_n)\big) \in A\big\}$$

for some $0 \leq t_1 < t_2 \ldots < t_n$ and $A \in \mathscr{B}(\mathbb{R}^n)$.

Let $\mathscr{C}$ be the family of all finite-dimensional cylinder sets and $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ the σ-algebra it generates. This is small enough so that we will be able to build probability measures on $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ using Carathéodory's Theorem. On the other hand $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ has events which can be defined using at most countably many coordinates. It follows that $\big\{\omega \in \mathscr{B}(\mathbb{R}^{[0,\infty)}) : \omega(t) \text{ is continuous}\big\}$ is not $\mathscr{B}(\mathbb{R}^{[0,\infty)})$-measurable. This is problematic since it is natural to build process with continuous paths, or more generally to measure $\mathbb{P}\big(\{\omega \in \mathscr{B}(\mathbb{R}^{[0,\infty)}) : \omega(t) \text{ is continuous}\}\big)$. To overcome this problem we will work (hard) to show that many processes can be assumed to be continuous or right-continuous. Such samples paths are fully described by their values for $t \in \mathbb{Q}$, i.e. at countably many times. This will simultaneously simplify the study of other quantities of interest, such as

$$\sup_{0 \leq s \leq t} |X_s|, \quad \tau(\omega) = \inf\{t \geq 0 : X_t(\omega) > 0\}.$$

1.2 Monotone class theorem

We recall an extremely useful measure theoretic result, which is known under a variety of names including: π–λ systems lemma, Dynkin's lemma and monotone class theorem. It allows us to extend a property, such as equality of measures, independence of random variables, measurability of functions, from a π–system to the σ–algebra it generates. We state two versions: one in terms of sets and another is terms of functions. Both are state for a general probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Lemma 1.2.1 (π–λ systems lemma, Dynkin's lemma).
Let $\mathscr{J}$ be a collection of subsets of Ω such that

1. *$\Omega \in \mathscr{J}$*
2. *if $A, B \in \mathscr{J}$ and $A \subset B$ then $B \setminus A \in \mathscr{J}$*
3. *if $\{A_n\}$ is an increasing sequence of elements of $\mathscr{J}$ then $\cup_n A_n \in \mathscr{J}$.*

Then if $\mathscr{I} \subseteq \mathscr{J}$ is closed under finite intersections then $\sigma(\mathscr{I}) \subseteq \mathscr{J}$.

The collection $\mathscr{J}$ of sets above is often called a d–system or a λ–system. A collection $\mathscr{I}$ which is stable under intersections is called a π–system. An immediate consequence of the above lemma is that if two measures agree on a π-system then they agree on the σ-algebra it generates. In particular, if two probability distributions μ, ν on $\mathbb{R}$ have the same cumulative distribution function then $\mu = \nu$.

The functional version of the above can be phrased as follows:

Theorem 1.2.2 (Monotone Class Theorem).
Let $\mathscr{H}$ be a vector space of bounded real-valued functions on Ω such that

1. *constant functions are in $\mathscr{H}$*
2. *if $\{h_n\}$ is an increasing sequence of non-negative elements of $\mathscr{H}$ s.t. $h = \sup_n h_n$ is bounded then $h \in \mathscr{H}$.*

Then if $\mathscr{C}$ is a subset of $\mathscr{H}$ which is stable under pointwise multiplication then $\mathscr{H}$ contains all bounded $\sigma(\mathscr{C})$ -measurable functions.

Chapter 2
An overview of Gaussian variables and processes

2.1 Gaussian variables

A random variable X is called centred standard Gaussian (or normal) if its distribution admits a density

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad x \in \mathbb{R}.$$

It follows that X has finite exponential moments and its Laplace transform is easily computed as

$$\mathbb{E}[e^{\lambda X}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\lambda x - \frac{x^2}{2}} dx = e^{\frac{\lambda^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(x-\lambda)^2}{2}} dx = e^{\frac{\lambda^2}{2}}, \ \lambda \in \mathbb{R}.$$

The above also shows that $\mathbb{E}[e^{zX}]$ is well defined for all $z \in \mathbb{C}$ and is an analytic function. The above equality on $\mathbb{R}$ between two analytic functions on $\mathbb{C}$ extends to all of $\mathbb{C}$:

$$\mathbb{E}[e^{zX}] = e^{\frac{z^2}{2}}, \ z \in \mathbb{C}.$$

In particular, taking $z = i\xi$ for $\xi \in \mathbb{R}$, we find the characteristic function $\mathbb{E}[e^{i\xi X}] = e^{-\xi^2/2}$, $\xi \in \mathbb{R}$. We write $X \sim \mathcal{N}(0,1)$.

We say Y has Gaussian distribution with mean m and variance σ^2, $Y \sim \mathcal{N}(m, \sigma^2)$, if $Y = \sigma X + m$ for some $\sigma \geq 0$, $m \in \mathbb{R}$ and $X \sim \mathcal{N}(0,1)$. This is equivalent to saying that

$$\mathbb{E}[e^{i\xi Y}] = \exp\left(im\xi - \frac{\sigma^2 \xi^2}{2}\right), \quad \xi \in \mathbb{R},$$

or, if $\sigma > 0$, that the density on $\mathbb{R}$ is

$$p_Y(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Note that we consider here a constant to be a degenerate Gaussian random variable.

The space of Gaussian variables (resp. distributions) is closed under convergence in probability (resp. convergence in distribution). More precisely we have the following result.

Proposition 2.1.1. *Let (X_n) be a sequence of Gaussian random variables, $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$, which converges in distribution to a random variable X. Then*

(i) X is also Gaussian, $X \sim \mathcal{N}(m, \sigma^2)$ with $m = \lim_{n\to\infty} m_n$, $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$,
(ii)and if X_n converge to X in probability then the convergence is also in L^p for all $1 \le p < \infty$.

Proof. The convergence in distribution is equivalent to saying that the characteristic functions converge:

$$\mathbb{E}\left[e^{\mathrm{i}\xi X_n}\right] = \exp(\mathrm{i}m_n\xi - \sigma_n^2\xi^2/2) \longrightarrow \mathbb{E}\left[e^{\mathrm{i}\xi X}\right], \quad \xi \in \mathbb{R}.$$

Taking absolute values we see that the sequence $\exp(-\sigma_n^2\xi^2/2)$ converges which in turn implies that $\sigma_n^2 \to \sigma^2 \in [0,\infty)$, where we ruled out the case $\sigma_n \to \infty$ since the limit has to be the absolute value of a characteristic function and in particular has to be continuous. We deduce that

$$e^{\mathrm{i}m_n\xi} \longrightarrow e^{\frac{1}{2}\sigma^2\xi^2}\mathbb{E}\left[e^{\mathrm{i}\xi X}\right], \quad \xi \in \mathbb{R}.$$

We now argue that this implies the sequence m_n converges to some finite m. If the sequence is bounded this is clear. Indeed, suppose it had subsequences converging to m and m' then we see that $e^{\mathrm{i}\xi m} = e^{\mathrm{i}\xi m'}$, for all $\xi \in \mathbb{R}$, which gives $m = m'$. Now suppose the sequence is not bounded so that there is a subsequence $m_{n_k} \nearrow +\infty$ (the case of $-\infty$ is analogous). Then

$$\mathbb{P}(X \ge M) \ge \limsup_{k\to\infty} \mathbb{P}(X_{n_k} \ge M) \ge \frac{1}{2}, \quad \text{for any } M > 0,$$

where the first inequality follows from convergence in distribution and $[M,\infty)$ being closed and the second inequality follows since $m_{n_k} \ge M$ for any fixed M and k large enough. Clearly the above is impossible for any fixed r.v. X and gives us a desired contradiction. This completes the proof of (i).

To show (ii) observe that the convergence of σ_n and m_n implies in particular that

$$\sup_n \mathbb{E}\left[e^{\theta X_n}\right] = \sup_n \mathbb{E}^{\theta m_n + \theta^2\sigma_n^2/2} < \infty, \quad \text{or any } \theta \in \mathbb{R}.$$

Since $\exp(|x|) \le \exp(x) + \exp(-x)$, the above remains finite if we take $|X_n|$ instead of X_n. This implies that $\sup_n \mathbb{E}[|X_n|^p] < \infty$ for any $p \ge 1$ and hence also

$$\sup_n \mathbb{E}[|X_n - X|^p] < \infty, \quad \forall p \ge 1.$$

Fix $p \geq 1$. Then the sequence $|X_n - X|^p$ converges to zero in probability by assumption and is uniformly integrable by the above. It follows that it also converges in L^1 which gives the desired result. □

2.2 Gaussian vectors and spaces

We define Gaussian vectors taking values in $\mathbb{R}^d$. However it is easy to see that everything which follows also holds when we replace $\mathbb{R}^d$ with an isomorphic euclidean vector space E.

Definition 2.2.1. A random vector taking values in $\mathbb{R}^d$ is called Gaussian if and only if

$$< u, X >= u^{\mathrm{T}} X = \sum_{i=1}^{d} u_i X_i \quad \text{is Gaussian for all } u \in \mathbb{R}^d.$$

By definition, *an image of a Gaussian vector under a linear transformation remains Gaussian*: if X is a Gaussian vector taking values in $\mathbb{R}^d$ and A is an $m \times d$ matrix then AX is a Gaussian vector taking values in $\mathbb{R}^m$. We will use this key observation throughout without an explicit mention.

Lemma 2.2.2. *Let X be a Gaussian vector and define $m_X := (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$ and $\Gamma_X := \big(\mathrm{cov}(X_i, X_j)_{1 \leq i,j \leq d}\big)$, the mean vector and the covariance matrix respectively. Then $q_X(u) := u^{\mathrm{T}} \Gamma_X u$ is a non-negative quadratic form and*

$$u^{\mathrm{T}} X \sim \mathscr{N}\left(u^{\mathrm{T}} m_X, q_X(u)\right), \quad u \in \mathbb{R}^d. \tag{2.1}$$

Proof. Clearly $\mathbb{E}[u^{\mathrm{T}} X] = u^{\mathrm{T}} m_X$ and

$$\mathrm{var}(u^{\mathrm{T}} X) = \mathbb{E}\left[\left(\sum_{i=1}^{d} u_i (X_i - \mathbb{E}[X_i])\right)^2\right] = \sum_{1 \leq i,j \leq d} u_i u_j \mathrm{cov}(X_i, X_j) = u^{\mathrm{T}} \Gamma_X u = q_X(u),$$

which also shows that $q_X(u) \geq 0$. □
The identification in (2.1) is equivalent to

$$\mathbb{E}\left[e^{\mathrm{i} < u, X >}\right] = e^{\mathrm{i} < u, m_X > - \frac{1}{2} q_X(u)}, \quad u \in \mathbb{R}^d.$$

From this we derive easily the following important fact:

Proposition 2.2.3. *Let X be a Gaussian vector and Γ_X its covariance matrix. Then $X_1, \ldots, X_d$ are independent if and only if Γ_X is a diagonal matrix (i.e. the variables are pairwise uncorrelated).*

Note that it is crucial to assume X is Gaussian and not just that $X_1, \ldots, X_d$ are all Gaussian. Indeed, consider $X_1 \sim \mathscr{N}(0,1)$ and ε an independent random variable

taking values ± 1 with probability $1/2$. Let $X_2 := \varepsilon X_1$. Then $X_2 \sim \mathcal{N}(0,1)$ and $\mathrm{cov}(X_1, X_2) = 0$ while clearly X_1, X_2 are not independent.

By definition, a Gaussian X remains Gaussian if we add to it a deterministic vector $m \in \mathbb{R}^d$. Hence, without loss of generality, by considering $X - m_X$, it suffices to consider centred Gaussian vectors. We note that Γ_X is symmetric and non-negative definite, as observed above. We now show the reverse: for any such matrix Γ, a Gaussian vector X with $\Gamma_X = \Gamma$ exists and can be obtained via a linear transformation of a Gaussian vector with iid coordinates.

Theorem 2.2.4. *Let Γ be a symmetric non-negative definite $d \times d$ matrix. Let $(\varepsilon_1, \ldots, \varepsilon_d)$ be an orthonormal basis in $\mathbb{R}^d$ which diagonalises Γ: $\Gamma \varepsilon_i = \lambda_i \varepsilon_i$ for some $\lambda_1 > \lambda_2 > \ldots > \lambda_r > 0 = \lambda_{r+1} = \ldots = \lambda_d$, where $1 \le r \le d$ is the rank of Γ.*

(i) A centred Gaussian vector X with covariance matrix $\Gamma_X = \Gamma$ exists.
(ii) Further, any such vector can be represented as

$$X = \sum_{i=1}^{r} Y_i \varepsilon_i, \tag{2.2}$$

where $Y_1, \ldots, Y_r$ are independent Gaussian, $Y_i \sim \mathcal{N}(0, \lambda_i)$.
(iii) If $r = d$ then X admits a density given by

$$p_X(x) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(\Gamma)}} \exp\Big(-\frac{1}{2} x^T \Gamma^{-1} x \Big), \quad x \in \mathbb{R}^d.$$

Proof. Let A be a matrix whose columns are ε_i so that $\Gamma_X = A \Lambda A^T$ where Λ is the diagonal matrix with λ_i on the diagonal. Let $Z_1, \ldots, Z_n$ be i.i.d. standard centred Gaussian variables and $Y_i = \sqrt{\lambda_i} Z_i$. Let X be given by (2.2), i.e. $X = AY$. Then

$$< u, X > = < u, AY > = < A^T u, Y > = \sum_{i=1}^{d} \sqrt{\lambda_i} < u, \varepsilon_i > Z_i$$

is Gaussian, centred. We can compute its variance, using the formula in the proof of Lemma 2.2.2 or directly, to see that it is given by

$$\sum_{i=1}^{d} \lambda_i (u^{\mathrm{T}} \varepsilon_i)^2 = \sum_{i=1}^{d} u^{\mathrm{T}} \varepsilon_i \lambda_i \varepsilon_i^{\mathrm{T}} u = (A^T u)^T \Lambda A^T u = u^T A \Lambda A^T u = u^T \Gamma u,$$

as required for (i).
Reversely, suppose X is a centred Gaussian vector with covariance matrix Γ and let $Y = A^T X$. For $u \in \mathbb{R}^d$, $< u, Y > = < Au, X >$ is centred Gaussian with variance

$$(Au)^T \Gamma Au = u^T A^T \Gamma A u = u^T \Lambda u$$

and we conclude that Y is also a centred Gaussian vector with covariance matrix Λ. Independence between $Y_1, \ldots, Y_d$ then follows from Proposition 2.2.3. It follows that when $r = d$, Y admits density on $\mathbb{R}^d$ given by

$$p_Y(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(\Lambda)}} \exp\Big(-\frac{1}{2} y^T \Lambda^{-1} y\Big), \quad y \in \mathbb{R}^d.$$

Change of variables, together with $\det(\Lambda) = \det(\Gamma)$ and $|\det(A)| = 1$ gives the desired density for x. □

The representation in (2.2), combined with Proposition 2.2.3, allows us to compute conditional expectations within the family of random variables which are linear transformations of a Gaussian vector X. These conditional expectations essentially reduce to a suitable projection operation. Consider X a Gaussian vector in $\mathbb{R}^d$ and let K be the vector space spanned by $X_2, \ldots, X_d$. Then

$$\mathbb{E}[X|\sigma(X_2, \ldots, X_d)] = p_K(X_1) := \sum_{i=2}^{n} a_i X_i,$$

where the coefficients are obtained solving the system of equations

$$\operatorname{cov}(X_1 - p_K(X_1), X_i) = \operatorname{cov}(X_1, X_i) - \sum_{j=2}^{d} a_j \operatorname{cov}(X_j, X_i) = 0, \ 2 \le i \le d.$$

The most striking feature above is that $\mathbb{E}[X|\sigma(K)]$ is an element of the vector space K itself and not a general $\sigma(K)$–measurable random variable. In particular, thisjustifies the principle of linear regression: if (X_1, X_2, X_3) is a Gaussian vector then the best (in the L^2 sense) approximation of X_1 in terms of X_2, X_3 is in fact a *linear* function of X_1 and X_2. This extends to a more general setup of Gaussian spaces described below.

2.2.1 Gaussian spaces

Note that to a Gaussian vector X in $\mathbb{R}^d$ we can associate the vector space span by its coordinates:

$$\left\{ \sum_{i=1}^{d} u_i X_i : u_i \in \mathbb{R} \right\}$$

and by definition all elements of this space are Gaussian random variables. This is a very simple example of a *Gaussian space* and it is useful to think of such spaces in a much greater generality.

Definition 2.2.5. A closed linear subspace $H \subset \mathrm{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ is called a *Gaussian space* if all of its elements are centred Gaussian random variables.

In an analogy to Proposition 2.2.3, two elements of a Gaussian space are independent if and only if they are uncorrelated, which in turn is equivalent to being orthogonal in L^2. More generally we have the following result.

Theorem 2.2.6. *Let H_1, H_2 be two Gaussian subspaces of a Gaussian space H. Then H_1, H_2 are orthogonal if and only if $\sigma(H_1)$ and $\sigma(H_2)$ are independent.*

Proof. Use Lemma 1.2.1 and Proposition 2.2.3. □
Note that the above extends instantly to the case of an arbitrary family of Gaussian subspaces $(H_i : i \in \mathscr{I})$ of H. In particular this greatly generalises Proposition 2.2.3. As a corollary of the above property we see that the conditional expectation on such spaces is just the L^2 projection.

Corollary 2.2.7. *Let H be a Gaussian space and K its closed subspace. Let p_K denote the orthogonal projection on K. Then for $X \in H$*

$$\mathbb{E}[X|\sigma(K)] = p_K(X). \tag{2.3}$$

Proof. Let $Y = X - p_K(X)$ which, by basic properties of p_K, is orthogonal to K. It follows from Theorem 2.2.6 that Y is independent of $\sigma(K)$ and hence

$$\mathbb{E}[X|\sigma(K)] = \mathbb{E}[p_K(X)|\sigma(K)] + \mathbb{E}[Y|\sigma(K)] = p_K(X) + \mathbb{E}[Y] = p_K(X).$$

□
Recall that for an arbitrary $X \in \mathrm{L}^2$ we would have

$$\mathbb{E}[X|\sigma(K)] = p_{\mathrm{L}^2(\Omega,\sigma(K),\mathbb{P})}(X)$$

while for a Gaussian X it is enough to consider the projection on a much smaller space K, as already observed this before for a finite dimensional K. We will use this observation repeatedly in the sequel.

2.3 Gaussian processes

Definition 2.3.1. A stochastic process $(X_t : t \geq 0)$ is called a (centred) Gaussian process if any finite linear combination of its coordinates is a Gaussian variable.

Equivalently, X is a centred Gaussian process if for any n, $0 \leq t_1 < t_2 < \ldots < t_n$, $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is a (centred) Gaussian vector. It follows that the distribution of a centred Gaussian process on $\mathscr{B}(\mathbb{R}^{[0,\infty)})$, is characterised by the covariance function $\Gamma : [0,\infty)^2 \mapsto \mathbb{R}$, i.e.

$$\Gamma(s,t) := \mathrm{cov}(X_t, X_s).$$

For any fixed n-tuple $(X_{t_1}, \ldots, X_{t_n})$ the covariance matrix $(\Gamma(t_i,t_j))$ has to be symmetric and positive semi-definite. As the following result shows, the converse also holds – for any such function we may construct an associated Gaussian process.

Theorem 2.3.2. *Let $\Gamma : [0,\infty)^2 \mapsto \mathbb{R}$ be symmetric and such that for any $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 < \ldots < t_n$*

$$\sum_{1 \leq i,j, \leq n} u_i u_j \Gamma(t_i,t_j) \geq 0, \quad u \in \mathbb{R}^d.$$

Then there exists a centred Gaussian process with covariance function Γ.

This result will follow from the more general Theorem 2.4.2 below. Finally, recalling that by Proposition 2.1.1 an L^2 limit of Gaussian variables is also Gaussian, we observe that the closed linear subspace of L^2 spanned by the variables $(X_t : t \geq 0)$ is a Gaussian space.

2.4 Constructing distributions on $\left(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)})\right)$

Let $\mathbb{T}$ be the set of finite increasing sequences of non-negative numbers, i.e. $\mathbf{t} \in \mathbb{T}$ if and only if $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ for some n and $0 \leq t_1 < t_2 < \ldots, < t_n$.

Suppose that for each $\mathbf{t} \in \mathbb{T}$ of length n we have a probability measure $\mathbb{Q}_\mathbf{t}$ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. The collection $(\mathbb{Q}_\mathbf{t} : \mathbf{t} \in \mathbb{T})$ is called a family of finite (marginal) dimensional distributions.

Definition 2.4.1. A family of finite dimensional distribution is called consistent if for any $\mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbb{T}$ and $1 \leq j \leq n$

$$\mathbb{Q}_\mathbf{t}(A_1 \times A_2 \times \ldots \times A_{j-1} \times \mathbb{R} \times A_{j+1} \times \ldots \times A_n) = \mathbb{Q}_\mathbf{s}(A_1 \times A_2 \times \ldots \times A_{j-1} \times A_{j+1} \times \ldots \times A_n)$$

where $A_i \in \mathscr{B}(\mathbb{R})$ and $\mathbf{s} := (t_1, t_2, \ldots, t_{j-1}, t_{j+1}, \ldots, t_n)$.

If we have a probability measure $\mathbb{Q}$ on $\left(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)})\right)$ then it defines a consistent family of marginals via

$$\mathbb{Q}_\mathbf{t}(A) = \mathbb{Q}(\{\omega \in \mathbb{R}^{[0,\infty)} : (\omega(t_1), \ldots, \omega(t_n)) \in A\})$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, $A \in \mathscr{B}(\mathbb{R}^n)$ and we note that the set in question is in $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ as it depends on finitely many coordinates.

We are interested in the converse, that is in existence of processes — seen as distributions on $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ – which given finite dimensional marginal distributions.

Theorem 2.4.2 (Daniell-Kolmogorov). *Let $\{\mathbb{Q}_\mathbf{t} : \mathbf{t} \in \mathbb{T}\}$ be a consistent family of finite-dimensional distributions. Then there exists a probability measure $\mathbb{P}$ on $\left(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)})\right)$ such that for any n, $\mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{T}$ and $A \in \mathscr{B}(\mathbb{R}^n)$*

$$\mathbb{Q}_\mathbf{t}(A) = \mathbb{P}(\{\omega \in \mathbb{R}^{[0,\infty)} : (\omega(t_1), \ldots, \omega(t_n)) \in A\}). \tag{2.4}$$

Remark about the proof: the relation (2.4) defines $\mathbb{P}$ on (finite dimensional) cylinders. It remains to show that so defined map is countably additive and apply Carethéodory's extension theorem. $\mathbb{P}$ is unique by $\pi - \lambda$ systems lemma.

This theorem establishes existence of a process (seen as identity mapping on $\left(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)})\right)$) with given (consistent) finite dimensional distributions. In particular (exercise) it implies Theorem 2.3.2

It does not however assert anything else and, as seen before, we can not meaningfully ask about path properties such as continuity. To remedy this we will study modifications of a given process.

2.5 Continuity of paths and Kolmogorov's criterion

Recall Definition 1.1.3 which introduced notions in which two processes can be close to one another.

Theorem 2.5.1 (Kolmogorov, Čentsov).
Suppose that a stochastic process $(X_t : t \leq T)$ *defined on* $(\Omega, \mathscr{F}, \mathbb{P})$ *satisfies*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq C|t-s|^{1+\beta}, \quad 0 \leq s,t \leq T \tag{2.5}$$

for some positive constants α, β *and* C.
Then there exists $\tilde{X}$, *a modification of* X, *whose paths are* γ*-locally Hölder continuous* $\forall \gamma \in (0, \beta/\alpha)$ *a.s. i.e. there is a positive* δ *and a constant* ε *s.t.*

$$\sup_{0 \leq t-s \leq \delta, s,t, \in [0,T]} \frac{|\tilde{X}_t - \tilde{X}_s|}{|t-s|^\gamma} \leq \varepsilon \quad \text{a.s.} \tag{2.6}$$

In particular, the sample paths of $\tilde{X}$ *are a.s. continuous.*

Chapter 3
Brownian Motion and First Properties

3.1 Definition of Brownian motion

As will become clear, we could use general results to construct a Brownian motion. We prefer however to pursue a self-contained direct construction to prove existence of Brownian motion.

Recall that Brownian motion, also called a Wiener process, is supposed to model the displacement of a pollen in a liquid. $(B_t - B_s)$ is the change of position between times s and t. It results from infinitely many tiny collisions, with molecules, each giving an identically distributed push, independent of other collisions. By Central Limit Theorem, we then expect $B_t - B_s \sim \mathcal{N}(0, \sigma^2(t-s))$ and to be independent of anything happening before time s. Lastly, the pollen travels in a continuous manner.

Definition 3.1.1 (Definition of Brownian motion).
A stochastic process $(B_t : t \geq 0)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a (standard one-dimensional) Brownian motion if

(i) $B_0 = 0$ a.s.,
(ii) for any s,t s.t. $0 \leq s \leq t$, $B_t - B_s \sim \mathcal{N}(0, t-s)$ and is independent of $\sigma(B_u : u \leq s)$,
(iii) its sample paths $t \mapsto B_t(\omega)$ are a.s. continuous.

The second requirement in the above definition is often referred to as BM having independent and stationary increments. The above definition appears to be the natural and desirable mathematical abstraction for the phenomenon we wanted to model. However the immediate question is then whether it actually makes sense, i.e. if it is non-void. To this end we need to prove that a BM exists — a task we undertake in the following section.

3.2 Lévy's construction of Brownian motion on $[0,1]$

Exercise 3.2.1. show that conditionally on $\{B_s = x, B_t = z\}$ the intermediate value $B_{\frac{t+s}{2}}$ has Gaussian distribution $\mathcal{N}(\frac{x+z}{2}, \frac{t-s}{4})$.

This suggests to build (B_t) through an interpolation:

- first draw B_1,
- then $B_{\frac{1}{2}}$ conditional on B_0 and B_1,
- then $B_{\frac{1}{4}}, B_{\frac{3}{4}}$ conditional on B_0, B_1 and $B_{\frac{1}{2}}, B_1$ respectively.

To realise this idea, take a countable collection $\{\xi_k^{(n)} : h \in I(n), n = 0,1,\dots,\}$ of independent $\mathcal{N}(0,1)$ variables, where $I(n)$ are odd integers between 0 and 2^n: $I(0) = \{1\}, I(1) = \{1\}, I(2) = \{1,3\}$, etc.

Define $(B_t^{(0)} : t \le 1)$ via $B_0^{(0)} = 0$; $B_0^{(0)} = \xi_1^{(0)}$ and a linear interpolation in between. Having defined $(B_t^{(n-1)} : t \le 1)$, let $B_{k2^{-n+1}}^{(n)} = B_{k2^{-n+1}}^{(n-1)}$ for $k = 0,1,\dots,2^{n-1}$ and $B_{k2^{-n}}^{(n)} := \frac{1}{2}\left(B_s^{(n-1)} + B_t^{(n-1)}\right) + \sqrt{\frac{t-s}{4}}\xi_k^{(n)}$, $k \in I(n)$, where $s = (k-1)2^{-n}$ and $t = (k+1)2^{-n}$, $\frac{t-s}{4} = 2^{-(n+1)}$. Then define $(B_t^{(n)} : t \le 1)$ by interpolating the values in $(k2^{-n} : 0 \le k \le 2^n)$ linearly.

We rewrite the construction using convenient orthonormal base in $\mathrm{L}^2([0,1], Leb)$. Define Haar functions $H_k^{(n)}$ by $H_1^{(0)}(t) = 1$ for $t \in [0,1]$ and

$$H_k^{(n)}(t) = \begin{cases} 2^{\frac{n-1}{2}} & \text{if } \frac{k-1}{2^n} \le t < \frac{k}{2^n} \\ -2^{\frac{n-1}{2}} & \text{if } \frac{k}{2^n} \le t < \frac{k+1}{2^n} \\ 0 & \text{otherwise} \end{cases}, \tag{3.1}$$

for $k \in I(n)$, $n \ge 1$, and Schauder functions through their integrals

$$S_k^{(n)}(t) = \int_0^t H_k^{(n)}(u)du, \quad t \le 1, n \ge 0, k \in I(n).$$

Clearly $B_t^{(0)} = \xi_1^{(0)} S_1^{(0)}(t)$ and by induction on n we obtain

$$B_t^{(n)}(\omega) = \sum_{m=0}^{n} \sum_{k \in I(m)} \xi_k^{(m)}(\omega) S_k^{(m)}(t), \quad 0 \le t \le 1; n \ge 0. \tag{3.2}$$

Lemma 3.2.2. *The sequence of functions in* (3.2) *converges uniformly in t, as* $n \to \infty$*, to a continuous function denoted* (B_t)*, a.s.*

Proof. Let $b_n = \max_{k \in I(n)} |\xi_k^{(n)}|$. Then, for $x > 0$,

$$\mathbb{P}(|\xi_k^{(n)}| > x) = \sqrt{\frac{2}{\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy \le \sqrt{\frac{2}{\pi}} \int_x^\infty \frac{y}{x} e^{-\frac{y^2}{2}} dy = \sqrt{\frac{2}{\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}. \tag{3.3}$$

In consequence,

$$\mathbb{P}(b_n > n) = \mathbb{P}\Big(\bigcup_{k\in I(n)} \{|\xi_k^{(n)}| > n\}\Big) \le \sum_{k\in I(n)} \mathbb{P}(|\xi_k^{(n)}| > n) \le \sqrt{\frac{2}{\pi}}\frac{2^n}{n} e^{-\frac{n^2}{2}},\ n \ge 1.$$

Since $\sum_{n=1}^{\infty} \frac{2^n}{n} e^{-\frac{n^2}{2}} < \infty$ the Borel Cantelli lemma implies that there exists $\tilde{\Omega} \subseteq \Omega$ with $\mathbb{P}(\tilde{\Omega}) = 1$ and on which there is an integer $n_0(\omega)$ satisfying $b_n(\omega) \le n$ for all $n \ge n_0(\omega)$. Then, for $\omega \in \tilde{\Omega}$ and $t \in [0,1]$

$$\sum_{m=n_0(\omega)}^{n} \sum_{k\in I(m)} |\xi_k^{(m)}(\omega) S_k^{(m)}(t)| \le \sum_{m=n_0(\omega)}^{n} n 2^{-\frac{n+1}{2}} < \infty$$

where we used that $S(m)_k(t)$ is non-zero for only one $k \in I(m)$. It follows that $B_t^{(n)}(\omega)$ converge uniformly in t to some $B_t(\omega)$. Continuity of $(B_t(\omega) : t \ge 1)$ follows from the uniform convergence and continuity of $B_t^{(n)}(\omega)$. $\square$

To establish that B is a Brownian motion we need to show independence of increments and that $B_t - B_s \sim \mathcal{N}(0, t-s)$. It is not hard to check that Haar functions form an orthonormal system in $\mathrm{L}^2([0,1], \mathscr{B}([0,1]), Leb)$ and hence the Parseval identity, see Proposition A.2.6, gives us that

$$\int_0^1 f(x)g(x)dx = \langle f, g\rangle = \sum_{n=0}^{\infty} \sum_{k\in I(n)} \langle f, H_k^{(n)}\rangle \langle g, H_k^{(n)}\rangle.$$

Applying it to $f = \mathbb{1}_{[0,t]}$ and $g = \mathbb{1}_{[0,s]}$ we obtain

$$s \wedge t := \min\{s,t\} = \sum_{n=0}^{\infty} \sum_{k\in I(n)} S_k^{(n)}(t) S_k^{(n)}(s); \quad 0 \le s,t \le 1. \tag{3.4}$$

Theorem 3.2.3. *The process $(B_t : t \le 1)$ defined above is a Brownian motion on $[0,1]$.*

Proof. It suffices to establish that for any $N \ge 1$ and $0 = t_0 < t_1 < \ldots < t_N \le 1$,

$$\mathbb{E}\left[\exp\left(i\sum_{j=1}^{N} \lambda_j (B_{t_j} - B_{t_{j-1}})\right)\right] = \prod_{j=1}^{N} \exp\left(-\frac{1}{2}\lambda_j^2 (t_j - t_{j-1})\right), \quad \forall \lambda_1, \ldots, \lambda_N \in \mathbb{R}^N.$$

Let $\lambda_{N+1} := 0$. Using properties of $(\xi_k^{(n)})$ and the definition of $B^{(n)}$ we observe that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-i\sum_{j=1}^{N}(\lambda_{j+1}-\lambda_j)B^{(n)}_{t_j}\right)\right]\\
=&\mathbb{E}\left[\exp\left(-i\sum_{m=0}^{n}\sum_{k\in I(m)}\xi_k^{(m)}\sum_{j=1}^{N}(\lambda_{j+1}-\lambda_j)S_k^{(m)}(t_j)\right)\right]\\
=&\prod_{m=0}^{n}\prod_{k\in I(m)}\exp\left(-\frac{1}{2}\left(\sum_{j=1}^{N}(\lambda_{j+1}-\lambda_j)S_k^{(m)}(t_j)\right)^2\right)\\
=&\exp\left(-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}(\lambda_{j+1}-\lambda_j)(\lambda_{i+1}-\lambda_i)\sum_{m=0}^{n}\sum_{k\in I(m)}S_k^{(m)}(t_i)S_k^{(m)}(t_j)\right)
\end{aligned}$$

Letting $n\to\infty$, since convergence a.s. implies convergence in distribution and hence convergence of characteristic functions, and using (3.4) we obtain

$$\mathbb{E}\left[\exp\left(-i\sum_{j=1}^{N}(\lambda_{j+1}-\lambda_j)B_{t_j}\right)\right]=\exp\left(-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}(\lambda_{j+1}-\lambda_j)(\lambda_{i+1}-\lambda_i)\min\{t_i,t_j\}\right).$$

Now, observing that the LHS equates to the quantity of interest and continuing rearranging the sums, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(i\sum_{j=1}^{N}\lambda_j(B_{t_j}-B_{t_{j-1}})\right)\right]=\mathbb{E}\left[\exp\left(-i\sum_{j=1}^{N}(\lambda_{j+1}-\lambda_j)B_{t_j}\right)\right]\\
&=\exp\left(-\sum_{j=1}^{N-1}\sum_{i=j+1}^{N}(\lambda_{j+1}-\lambda_j)(\lambda_{i+1}-\lambda_i)t_j-\frac{1}{2}\sum_{j=1}^{N}(\lambda_{j+1}-\lambda_j)^2t_j\right)\\
&=\exp\left(-\sum_{j=1}^{N-1}(\lambda_{j+1}-\lambda_j)(-\lambda_{j+1})t_j-\frac{1}{2}\sum_{j=1}^{N}(\lambda_{j+1}-\lambda_j)^2t_j\right)\\
&=\exp\left(\frac{1}{2}\sum_{j=1}^{N-1}(\lambda_{j+1}^2-\lambda_j^2)t_j-\frac{1}{2}\lambda_N^2t_N\right)\\
&=\prod_{j=1}^{N}\exp\left(-\frac{1}{2}\lambda_j^2(t_j-t_{j-1})\right),
\end{aligned}$$

as required. □

Corollary 3.2.4. *There exists a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and a stochastic process $B=(B_t:t\geq 0)$ on it which is a standard real-valued Brownian Motion.*

Proof. Using Theorem 3.2.3, take a sequence of probability spaces $(\Omega_k,\mathscr{F}_k,\mathbb{P}_k)$, each with a Brownian motion $(B_t^k:t\leq 1)$ constructed in Theorem 3.2.3. Let $\Omega:=\Omega_1\times\Omega_2\times\ldots$, $\mathscr{F}:=\mathscr{F}_1\otimes\mathscr{F}_2\otimes\ldots$ and $\mathbb{P}:=\mathbb{P}_1\times\mathbb{P}_2\times\ldots$. Define B on Ω recursively by

$$B_t = B_t^1 \quad \text{for } 0 \le t \le 1,$$
$$B_t = B_k + B_{t-k}^{k+1} \quad \text{for } k \le t < k+1.$$

It is straightforward that $(B_t : t \ge 0)$ is indeed a Brownian motion on $(\Omega, \mathscr{F}, \mathbb{P})$. □

Exercise 3.2.5. Using Kolmogorov's lemma show that Brownian motion admits a modification which is locally Hölder continuous of order γ for any $0 < \gamma < \frac{1}{2}$.

3.3 Wiener Measure

Let $C(\mathbb{R}_+, \mathbb{R})$ be the space of continuous functions from $[0,\infty)$ to $\mathbb{R}$. Given a Brownian motion $(B_t : t \ge 0)$ on $(\Omega, \mathscr{F}, \mathbb{P})$, consider

$$\Omega \mapsto C(\mathbb{R}_+, \mathbb{R}), \quad \text{given by} \quad \omega \mapsto (B_t(\omega) : t \ge 0) \tag{3.5}$$

which is measurable w.r.t. $\mathscr{B}(C(\mathbb{R}_+, \mathbb{R}))$ — the smallest σ-algebra such that the coordinate mappings (i.e. $(\omega_t : t \ge 0) \mapsto \omega(t_0)$ for a fixed t_0) are measurable. (In fact $\mathscr{B}(C(\mathbb{R}_+, \mathbb{R}))$ is also the Borel σ-algebra generated by the topology of uniform convergence on compacts).

Definition 3.3.1. The Wiener measure **W** is the image of $\mathbb{P}$ under the above mapping in (3.5), it is the probability measure on the space of continuous functions s.t. the canonical process, i.e. $(W_t(\omega) = \omega(t), t \ge 0)$, is a Brownian motion.

Corollary 3.3.2 (Wiener measure).
The Wiener measure is the unique probability measure on $(C(\mathbb{R}_+, \mathbb{R}), \mathscr{B}(C(\mathbb{R}_+, \mathbb{R})))$ *such that*

(i) $\mathbf{W}(\{\omega \in C(\mathbb{R}_+, \mathbb{R}), \omega(0) = 0\}) = 1$
(ii) for any $n \ge 1$, $\forall 0 = t_0 < t_1 < \ldots < t_n$, $A \in \mathscr{B}(\mathbb{R}^n)$

$$\mathbf{W}(\{\omega \in C(\mathbb{R}_+, \mathbb{R}) : (\omega(t_1), \ldots, \omega(t_n)) \in A\})$$
$$= \int_A \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{dy_1 \cdots dy_n}{\sqrt{t_1(t_2 - t_1)\ldots(t_n - t_{n-1})}} \exp\Big(-\sum_{i=1}^{n} \frac{(y_i - y_{i-1})^2}{2(t_i - t_{i-1})}\Big) \quad \textit{where } y_0 := 0.$$

Proof. Existence follows from construction in Section 3.2 with (ii) above following directly from (ii) in Definition 3.1.1. The uniqueness of is an immediate consequence of π–λ systems lemma since $\mathscr{B}(C(\mathbb{R}_+, \mathbb{R})))$ is generated by finitely dimensional projections.

□

3.4 Extensions and first properties

Definition 3.4.1. Let μ be a probability measure on $\mathbb{R}^d$. A d-dimensional stochastic process $(B_t : t \geq 0)$ on $(\Omega, \mathscr{F}, \mathbb{P})$ is called a d-dimensional Brownian motion with initial distribution μ if

(i) $\mathbb{P}(B_0 \in A) = \mu(A) \quad A \in \mathscr{B}(\mathbb{R}^d)$
(ii) $\forall 0 \leq s \leq t$ the increment $(B_t - B_s)$ is independent of $\sigma(B_u : u \leq s)$ and is normally distributed with mean 0 and covariance matrix $(t-s) \times \mathbf{I}_d$
(iii) B has a.s. continuous paths.

If $\mu(\{x\}) = 1$ for some $x \in \mathbb{R}^d$, we say that B starts at x.

Proposition 3.4.2. *Let B be a standard real-valued Brownian motion. Then*

(i) $-B_t$ is also a Brownian motion, (symmetry)
(ii) $\forall c \geq 0$, $cB_{\frac{t}{c^2}}$ is a Brownian motion, (scaling)
(iii) $X_0 = 0$, $X_t := tB_{\frac{1}{t}}$ is a Brownian motion, (time reversal)
(iv) $\forall s \geq 0$, $\tilde{B}_t = B_{t+s} - B_s$ is a Brownian motion independent of $\sigma(B_u : u \leq s)$, (simple Markov property).

Proof. (i) and (ii) are immediate from the definition and $\sigma X \sim \mathscr{N}(0, \sigma^2)$ where $X \sim \mathscr{N}(0,1)$. We leave (iii) as an exercise. Finally, $\tilde{B}_0 = 0$ and $\tilde{B}$ has a.s. continuous paths. Further, $\tilde{B}_t - \tilde{B}_u = B_{t+s} - B_{u+s} \sim \mathscr{N}(0, t-u)$ and is independent of $\sigma(B_r : r \leq u+s)$ which contains $\sigma(\tilde{B}_r : r \leq u)$. Finally, each $\tilde{B}_t$ is independent of $\sigma(B_u : u \leq s)$ and hence $\tilde{B}$ is independent of $\sigma(B_u : u \leq s)$. $\square$

3.5 Basic properties of Brownian sample path

From now on, when we say "Brownian motion", we mean a standard real-valued Brownian motion. We are interested in the mapping $t \mapsto B_t(\omega)$, how wiggly it is and what is its typical behaviour close to 0 or ∞. We already know that $t \mapsto B_t(\omega)$ is a.s. continuous, and in fact locally Hölder continuous of order γ, for any $\gamma < 1/2$. We now look at variation and quadratic variation of the paths. Given a subdivision Δ of $[0,t]$: $\Delta = \{0 = t_0 < t_1 < \ldots < t_n = 1\}$, we call $mesh(\Delta) = \max_{0 \leq i \leq n-1}(t_{i+1} - t_i)$.

Definition 3.5.1. A (real-valued) stochastic process $(X_t : t \geq 0)$ is said to be of finite quadratic variation if there exists a (finite) process $(\langle X, X \rangle_t, t \geq 0)$ such that for every t and every sequence (Δ_n) of subdivisions of $[0,t]$ such that $mesh(\Delta_n) \to 0$ as $n \to \infty$, we have

$$QV_t^{\Delta_n}(X) \to \langle X, X \rangle_t \quad \text{in probability, as } n \to \infty$$

where $QV_t^{\Delta_n}(X) = \sum_{i=1}^{n}(X_{t_{i+1}} - X_{t_i})^2$, $\Delta = \{0 \le t_0 < t_1 < \cdots < t_n = t\}$.

Note that it follows $\langle X,X\rangle_t - \langle X,X\rangle_s = \lim_{n\to\infty}(QV_t^{\Delta_n}(X) - QV_s^{\Delta_n}(X)) \ge 0$ in probability and hence $\langle X,X\rangle$ has a.s. non-decreasing paths.

Theorem 3.5.2. *Brownian motion is of finite quadratic variation and $\langle B,B\rangle_t = t$ a.s.*

Proof.

$$\begin{aligned}
\mathbb{E}[(QV_t^{\Delta_n}(B) - t)^2] =& \mathbb{E}\left[\left(\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_t)^2 - (t_{i+1} - t_i)\right)^2\right] \\
=& \sum_{i=0}^{n-1}\mathbb{E}\left[\left((B_{t_{i+1}} - B_t)^2 - (t_{i+1} - t_i)\right)^2\right] \\
=& \sum_{i=0}^{n-1}(t_{i+1} - t_i)^2\mathbb{E}[(\xi^2 - 1)^2] \text{ with } \xi \sim \mathcal{N}(0,1) \\
=& 2\sum_{i=0}^{n-1}(t_{i+1} - t_i)^2 \\
\le& 2\sup_{0\le i\le n-1}(t_{i+1} - t_i)\sum_{j=0}^{n-1}(t_{j+1} - t_j) \\
=& 2mesh(\Delta_t^n)t \ \to\ 0
\end{aligned}$$

where the first equality is due to independence of increments. Hence $QV_t^{\Delta_n}(B) \to t$ as $n \to \infty$ in L^2 and in particular in probability. □

Remark 3.5.3. The convergence thus is a.s along a subsequence. It can be shown that if the sequence Δ_n is nested that $QV_t^{\Delta_n}(B) \to t$ a.s. $\forall t$.

Corollary 3.5.4. *Brownian sample paths are of infinite variation on any interval almost surely.*

Corollary 3.5.5. *Brownian sample paths are almost surely nowhere locally Holder continuous of order $\gamma > \frac{1}{2}$.*

Proofs. Fix interval $[s,t]$. By Theorem 3.5.2 and the above remark, we have $\Omega_0 \subseteq \Omega$, $\mathbb{P}(\Omega_0) = 1$ s.t. $\forall \omega \in \Omega_0$

$$\lim_{n\to\infty}\sum_{t_i\in\Omega_n}(B_{t_{i+1}}(\omega) - B_{t_i}(\omega))^2 = t - s \quad \text{a.s.}$$

and $B_t(\omega)$ is continuous. Let V_s^t be the variation of $B(\omega)$ i.e.

$$V_s^t(\omega) = \sup_{\Delta\subseteq[s,t]}\sum_{t_i\in\Delta}|B_{t_{i+1}} - B_{t_i}|.$$

Then

$$\sum_{t_i\in\Omega_n}(B_{t_{i+1}}(\omega)-B_{t_i}(\omega))^2 \le \sup_{t_i\in\Omega_n}|B_{t_{i+1}}(\omega)-B_{t_i}(\omega)|V_s^t(\omega)$$

and the RHS would converge to 0 if $V_s^t(\omega)<\infty$ by continuity of $B_t(\omega)$. Hence $V_s^t(\omega)=+\infty$ on Ω_0.

Similarly, if $|B_u(\omega)-B_r(\omega)|\le K|u-r|^\gamma$, for $s<r,u<t$ and $\gamma>1/2$ then

$$\sum_{t_i\in\Omega_n}(B_{t_{i+1}}(\omega)-B_{t_i}(\omega))^2 \le K^2(t-s)\sup_{t_i\in\Omega_n}|t_{i+1}-t_i|^{2\gamma-1}\to 0 \quad \text{as } n\to\infty$$

giving a contradiction. □

To study the behaviour of Brownian motion at $0+$ it is useful to establish the following $0-1$ law. For a Brownian motion $(B_t)_{t\ge0}$ on $(\Omega,\mathscr{F},\mathbb{P})$ let $\mathscr{F}_t^B:=\sigma(B_u:u\le t)$ and $\mathscr{F}_t=\sigma(\mathscr{F}_t^B,\mathscr{N}(\mathbb{P}))$, where $\mathscr{N}(\mathbb{P})$ are $\mathbb{P}$ are negligible sets, i.e. $\mathscr{N}(\mathbb{P}):=\{A\in\Omega:\exists B\in\mathscr{F}\ A\subseteq B \text{ and } \mathbb{P}(B)=0\}$. Finally let $\mathscr{F}_{t+}:=\cap_{s>t}\mathscr{F}_s$.

Proposition 3.5.6 (Blumenthal's 0-1 law).
$\mathscr{F}_{0+}$ is trivial, that is $\mathbb{P}(A)\in\{0,1\}$ for all $A\in\mathscr{F}_{0+}$.

Proof. For $0<\varepsilon<t$ let $\mathscr{F}_{\varepsilon,t}=\sigma(B_s-B_\varepsilon:\varepsilon<s<t)\vee\sigma(\mathscr{N}(\mathbb{P}))$. Note that $A\in\mathscr{F}_{\varepsilon,t}$ if and only if $A=A'$ a.s. with $A'\in\sigma(B_s-B_\varepsilon:\varepsilon<s<t)$. By Proposition 3.4.2 (iv) the σ-algebras $\sigma(B_s-B_\varepsilon:\varepsilon<s<t)$ and $\sigma(B_u:u\le\varepsilon)$ are independent and it follows that so are $\mathscr{F}_{\varepsilon,t}$ and $\mathscr{F}_\varepsilon$. In particular, $\mathscr{F}_{\varepsilon,t}$ and $\mathscr{F}_{0+}$ are independent.

It follows from $\pi-\lambda$ systems lemma that $\sigma(\cup_{\varepsilon>0}\mathscr{F}_{\varepsilon,t})$ and $\mathscr{F}_{0+}$ are independent. But for any $s\in(0,t)$, $B_s=\lim_{\varepsilon\searrow0}(B_s-B_\varepsilon)$ is measurable w.r.t. $\sigma(\cup_{\varepsilon>0}\mathscr{F}_{\varepsilon,t})$ so that $\sigma(\cup_{\varepsilon>0}\mathscr{F}_{\varepsilon,t})\supseteq\mathscr{F}_t^B$ which implies that $\sigma(\cup_{\varepsilon>0}\mathscr{F}_{\varepsilon,t})=\mathscr{F}_t$. Since $\mathscr{F}_{0+}\subseteq\mathscr{F}_t$, it is independent of itself, i.e.

$$\forall A\in\mathscr{F}_{0+}\quad \mathbb{P}(A)=\mathbb{P}(A\cap A)=\mathbb{P}(A)^2 \quad \text{which gives } \mathbb{P}(A)\in\{0,1\}.$$

□

Corollary 3.5.7. *Let B be a standard real-valued Brownian motion, as above.*

1. *Then, a.s., for any $\varepsilon>0$,*

$$\sup_{0\le s\le\varepsilon}B_s>0 \quad \textit{and} \quad \inf_{0\le s\le\varepsilon}B_s<0 \quad \textit{a.s.}$$

   In particular, $\inf\{t>0:B_t=0\}=0$ *a.s.*
2. *For all $a\in\mathbb{R}$, $H_a:=\inf\{t\ge0,B_t=a\}<\infty$ a.s. (where $\inf\emptyset=\infty$) and in particular*

$$\limsup_{t\to\infty}B_t=\infty \quad \textit{and} \quad \liminf_{t\to\infty}B_t=-\infty \quad \textit{a.s.}$$

Remark 3.5.8. Note that $\sup_{0\le s\le\varepsilon}B_s$ is measurable since $\sup_{0\le s\le\varepsilon}B_s=\sup_{0\le s\le\varepsilon,s\in\mathbb{Q}}B_s$.

Proof. (i) Let $\varepsilon_n \searrow 0$ (strictly) and $A := \bigcap_{n\geq 0} A_n$, with $A_n := \{\sup_{0\leq s\leq \varepsilon_n} B_s > 0\}$. Note that A is $\mathscr{F}_{0+}$-measurable. On the other hand, by monotonicity,

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) \geq \lim_{n\to\infty} \mathbb{P}(B_{\varepsilon_n} > 0) = \frac{1}{2}.$$

Then by Blumenthal's $0-1$ law $\mathbb{P}(A) = 1$. Replacing B by $-B$ we obtain $\mathbb{P}(\inf_{0\leq s\leq \varepsilon} B_s < 0) = 1$.
(ii) We write

$$\begin{aligned}
1 = \mathbb{P}\Big(\sup_{0\leq s\leq 1} B_s > 0\Big) &= \lim_{\delta\searrow 0} \mathbb{P}\Big(\sup_{0\leq s\leq 1} B_s > \delta\Big) \\
&= \lim_{\delta\searrow 0} \mathbb{P}\Big(\sup_{0\leq s\leq 1} \delta B_{s/\delta^2} > \delta\Big) \quad \text{by Prop 3.4.2 (ii)} \\
&= \lim_{\delta\searrow 0} \mathbb{P}\Big(\sup_{0\leq u\leq 1/\delta^2} B_u > 1\Big) \\
&= \mathbb{P}\Big(\sup_{s\geq 0} B_u > 1\Big).
\end{aligned}$$

Hence for any $\lambda > 0$

$$\mathbb{P}\Big(\sup_{s\geq 0} B_s > 1/\lambda\Big) = \mathbb{P}\Big(\sup_{u\geq 0} \lambda B_{u/\lambda^2} > 1\Big) = \mathbb{P}\Big(\sup_{s\geq 0} B_s > 1\Big) = 1.$$

As above, by replacing B by $-B$, we obtain $\mathbb{P}\big(\inf_{s\geq 0} B_s < -1/\lambda\big) = 1$. Taking $\lambda = 1/|a|$, it follows that $H_a < \infty$ a.s. The last statement then follows since a continuous function $B_t(\omega)$ has range equal to $\mathbb{R}$ if and only if

$$\limsup_{t\to\infty} B_t(\omega) = \infty \text{ and } \liminf_{t\to\infty} B_t(\omega) = -\infty.$$

□

Remark 3.5.9. From (i) above and the simple Markov property in Proposition 3.4.2 (iv), we deduce that a.s. $t \mapsto B_t$ has no intervals of monotonicity. This also follows from infinite variation property.

Chapter 4
Filtrations, stopping times and first applications

4.1 Filtrations

So far we looked at stochastic processes from "end of time" perspective and we talked about processes mainly as a random variable taking values in an appropriate space of functions. We are now going to emphasise the evolution in time. We start with notions which encode "flow of information" and "strategies" and are defined with no mention of $\mathbb{P}$ (only possibly through its null sets).

Definition 4.1.1. A filtration on the measurable space $(\Omega, \mathscr{F})$ is an increasing family $(\mathscr{F}_t)_{t\geq 0}$ of σ-algebras; $\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F}\ \forall t \geq s \geq 0$. A measurable space endowed with a filtration is called a filtered space.

Definition 4.1.2. A stochastic process X on $(\Omega, \mathscr{F})$ is said to be adapted to filtration $(\mathscr{F}_t)_{t\geq 0}$ if X_t is $\mathscr{F}_t$-measurable for all $t \geq 0$.

Remark 4.1.3. A process X is always adapted to its natural filtration

$$\mathscr{F}_t^X := \sigma(X_u : 0 \leq u \leq t)$$

(and it is the smallest filtration to which X is adapted). We can think of $\mathscr{F}_t^X$ as the information set, as time t, of an agent who is observing (paths of) X.

We define $\mathscr{F}_{0-} := \mathscr{F}_0$ and

$$\mathscr{F}_{t-} := \sigma(\cup_{s<t}\mathscr{F}_s),\ t > 0,$$

the σ-algebra of events strictly prior to t. We also let

$$\mathscr{F}_{t+} := \cap_{\varepsilon>0}\mathscr{F}_{t+\varepsilon},\ t \geq 0,$$

the σ-algebra of events "with an infinitesimal peek into the future".

We say that filtration $(\mathscr{F}_t)$ is right-continuous if $\mathscr{F}_t = \mathscr{F}_{t+}\ \forall t \geq 0$ and left-continuous if $\mathscr{F}_t = \mathscr{F}_{t-}\ \forall t \geq 0$

Definition 4.1.4. A filtration $(\mathscr{F}_t)_{t\geq 0}$ (or the filtered space $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq 0},\mathbb{P})$) is said to satisfy the usual conditions (assumptions/hypothesis) if it is right-continuous and $\mathscr{F}_0$ (and hence all $\mathscr{F}_t$) contain all $\mathbb{P}$-negligible events in $\mathscr{F}$, i.e. $\mathscr{N}(\mathbb{P})\subseteq\mathscr{F}_0$ for $\mathscr{N}(\mathbb{P}):=\{A\in\Omega:\exists B\in\mathscr{F}\ A\subseteq B \text{ and } \mathbb{P}(B)=0\}$.

Given a filtered probability space on $(\Omega,\mathscr{F},(\mathscr{F}_t),\mathbb{P})$ we can always consider its natural augmentation replacing the filtration with $\mathscr{F}_{t+}\vee\sigma(\mathscr{N}(\mathbb{P}))=\sigma(\mathscr{F}_{t+},\mathscr{N})$. which satisfies the usual conditions.

Even though these may appear somewhat artificial/technical conditions, they will turn out to be very useful. They will allow us in particular to construct (modification of) martingales which are always càdlàg (RCLL = right-continuous with left limits).

We saw previously the definitions of a stochastic process and of a measurable stochastic process. Given a flow of information, the analogous notions are that of an adapted process and of a progressively measurable process.

Definition 4.1.5. A stochastic process X_t on $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq 0},\mathbb{P})$ is called progressively measurable if $\forall t\geq 0$ the mapping $(s,\omega)\mapsto X_s(\omega)$ is measurable on $([0,t]\times\Omega,\overline{\mathscr{B}([0,t])}\otimes\mathscr{F}_t)$.

Note that a progressively measurable process is in particular adapted.

Proposition 4.1.6. *An adapted process (X_t) whose paths are all right-continuous (or are all left-continuous) is progressively measurable.*

Proof. We present the argument for a right-continuous X. For $t>0$, $n\geq 1$, $k=0,1,2\ldots,2^n-1$ let $X_0^{(n)}(\omega)=X_0(\omega)$ and $X_s^{(n)}(\omega):=X_{\frac{k+1}{2^n}t}(\omega)$ for $\frac{kt}{2^n}<s\leq\frac{k+1}{2^n}t$. Clearly $(X_s^{(n)}:s\leq t)$ takes finitely many values and is $\mathscr{B}([0,t])\otimes\mathscr{F}_t$–measurable. Further $X_s(\omega)=\lim_{n\to\infty}X_s^{(n)}(\omega)$ by right-continuity and hence is also measurable as limit of measurable mappings. □

4.2 Stopping times

Definition 4.2.1. Let $(\Omega,(\mathscr{F}_t),\mathscr{P})$ be a filtered space. A random variable $\tau:\Omega\mapsto[0,+\infty]$ is called a stopping time (relative to $(\mathscr{F}_t)$) if $\{\tau\leq t\}\in\mathscr{F}_t\ \forall t\geq 0$.

τ represents a valid "stopping strategy" — at any moment t we can describe if we should stop or continue using only the information available at t i.e. $\mathscr{F}_t$. A deterministic time is a stopping time as is "the first time a certain phenomenon occurs". $\tau=\sup\{t\leq 1:B_t=0\}$ is not as we have to wait until time 1 to know when it happened.

Our fundamental example will be given by first hitting times. If (X_t) is a stochastic process and $\Gamma\in\mathscr{B}(\mathbb{R})$ we let

$$H_\Gamma(\omega)=H_\Gamma(X(\omega)):=\inf\{t\geq 0:X(\omega)\in\Gamma\}. \tag{4.1}$$

Exercise 4.2.2. Show that

- if (X_t) is adapted to $(\mathscr{F}_t)$ and has right-continuous paths then H_Γ, for Γ an open set, is a stopping time relative to $(\mathscr{F}_{t+})$.
- if (X_t) has continuous paths, then H_Γ, for Γ a closed set, is a stopping time relative to $(\mathscr{F}_t)$.

With a stopping time we can associate "the information known at time τ":

Definition 4.2.3. Given a stopping time τ relative to $(\mathscr{F}_t)$ we define

$$\begin{aligned}\mathscr{F}_\tau &:= \{A \in \mathscr{F} : A \cap \{\tau \le t\} \in \mathscr{F}_t\}\\ \mathscr{F}_{\tau+} &:= \{A \in \mathscr{F} : A \cap \{\tau < t\} \in \mathscr{F}_t\}\\ \mathscr{F}_{\tau-} &:= \sigma(\{A \cap \{\tau > t\} : t \ge 0, A \in \mathscr{F}_t\})\end{aligned}$$

which satisfy all the natural properties.

Proposition 4.2.4. *Let τ be a stopping time. Then*

(i) $\mathscr{F}_\tau$ is a σ-algebra and τ is $\mathscr{F}_\tau$-measurable.
(ii) $\mathscr{F}_{\tau-} \subseteq \mathscr{F}_\tau \subseteq \mathscr{F}_{\tau+}$ and $\mathscr{F}_\tau = \mathscr{F}_{\tau+}$ if $(\mathscr{F}_t)$ is right-continuous.
(iii) If $\tau = t$ then $\mathscr{F}_\tau = \mathscr{F}_t$, $\mathscr{F}_{\tau+} = \mathscr{F}_{t+}$.
(iv) If τ and ρ are stopping times then so are $\tau \wedge \rho$, $\tau \vee \rho$ and $\tau + \rho$ and $\{\tau \le \rho\} \in \mathscr{F}_{\tau\wedge\rho}$. Further if $\tau \le \rho$ then $\mathscr{F}_\tau \subseteq \mathscr{F}_\rho$.
(v) If τ is a stopping time and ρ is a $[0,\infty]$-valued random variable which is $\mathscr{F}_\tau$-measurable and $\rho \ge \tau$, then ρ is a stopping time. In particular

$$\tau_n := \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{1}_{\left\{\frac{k}{2^n} < \tau \le \frac{k+1}{2^n}\right\}} + \infty \mathbb{1}_{\{\tau=\infty\}}$$

is a sequence of stopping times with $\tau_n \searrow \tau$ as $n \to \infty$.

Proof.

We give proof of (v). For the rest, please see the pb sheet.
$\{\rho \le t\} = \{\rho \le t\} \wedge \{\tau \le t\} \in \mathscr{F}_t$ since ρ is $\mathscr{F}_\tau$-measurable. Hence ρ is a stopping time. We have $\tau_n \searrow \tau$ by definition, and clearly τ_n is $\mathscr{F}_\tau$-measurable since τ is $\mathscr{F}_\tau$-measurable. □

Theorem 4.2.5. *Let (X_t) be a progressively measurable process and τ a stopping time. Then $X_\tau \mathbb{1}_{\{\tau<\infty\}}$ is $\mathscr{F}_\tau$-measurable. The stopped process $X^\tau = (X_{\tau\wedge t} : t \ge 0)$ is progressively measurable (where $X_\tau \mathbb{1}_{\{\tau<\infty\}}(\omega) = X_{\tau(\omega)}(\omega)\mathbb{1}_{\{\tau(\omega)<\infty\}}$).*

Proof. We need to show that for all $A \in \mathscr{B}(\mathbb{R})$ and $t \ge 0$,

$$\{X_\tau \mathbb{1}_{\{\tau<\infty\}} \in A\} \cap \{\tau \le t\} \in \mathscr{F}_t.$$

Note that

$$\{X_\tau \mathbb{1}_{\{\tau<\infty\}} \in A\} \cap \{\tau \le t\} = \{X_\tau \mathbb{1}_{\{\tau\le t\}} \in A\} \cap \{\tau \le t\}.$$

Hence, it suffices to show that $X_\tau \mathbb{1}_{\{\tau\le t\}}$ is $\mathscr{F}_t$-measurable since $\{\tau \le t\} \in \mathscr{F}_t$. But $X_\tau \mathbb{1}_{\{\tau\le t\}} = X_{\tau\wedge t}\mathbb{1}_{\{\tau\le t\}}$ and hence it suffices to show that $X_{\tau\wedge t}$ is $\mathscr{F}_t$–measurable. This will follow if we argue that X^τ is progressively measurable. This however is immediate since $X_{\tau\wedge s}$ on $[0,t]\times\Omega$ is a composition of two maps

$$\begin{aligned}(s,\omega) &\mapsto (\tau(\omega)\wedge s, \omega),\\ ([0,t]\times\Omega, \mathscr{B}([0,t])\otimes\mathscr{F}_t) &\mapsto ([0,t]\times\Omega, \mathscr{B}([0,t])\otimes\mathscr{F}_t)\end{aligned}$$

and

$$\begin{aligned}(u,\omega) &\mapsto X_u(\omega),\\ ([0,t]\times\Omega, \mathscr{B}([0,t])\otimes\mathscr{F}_t) &\mapsto (\mathbb{R}, \mathscr{B}(\mathbb{R})),\end{aligned}$$

both of which are measurable since τ is a stopping time and X is progressively measurable. □

We end this section with an important result which we state without a proof.

Theorem 4.2.6. *For a progressively measurable process (X_t) defined on a filtered probability space satisfying the usual conditions H_Γ is a stopping time for all $\Gamma \in \mathscr{B}(\mathbb{R})$.*

The proof relies on "section theorems" which show that a projection $H \in \mathscr{B}(\mathbb{R}_+)\otimes\mathscr{F}$ on $\mathscr{F}$ is measurable.

4.3 Strong Markov property and reflection principle for Brownian motion

Theorem 4.3.1. *Let $B = (B_t : t \ge 0)$ be a standard Brownian motion on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge0}, \mathbb{P})$, i.e. $(B_t)_{t\ge0}$ is a standard Brownian motion and is $\mathscr{F}_t$-adapted and $(B_t - B_s)$ is independent of $\mathscr{F}_s$ for any $0 \le s \le t$. Let τ be a stopping time. Then, conditional on $\{\tau < \infty\}$, the process*

$$B_t^{(\tau)} := B_{\tau+t} - B_\tau \tag{4.2}$$

is a standard Brownian motion independent of $\mathscr{F}_\tau$.

Proof. Assume that $\tau < \infty$ a.s. We will show that $\forall\, A \in \mathscr{F}_\tau$, $0 \le t_1 < \ldots < t_n$ and continuous and bounded functions F on $\mathbb{R}^n$ we have

$$\mathbb{E}\big[\mathbb{1}_A F(B_{t_1}^{(\tau)}, \ldots, B_{t_1}^{(\tau)})\big] = \mathbb{P}(A)\mathbb{E}\big[F(B_{t_1}, \ldots, B_{t_n})\big]. \tag{4.3}$$

This implies that

1. (taking $A = \Omega$), B and $B^{(\tau)}$ have the same finite-dimensional distributions and hence, since $B^{(\tau)}$ has continuous paths, that $B^{(\tau)}$ is a standard Brownian motion;
2. that $(B^{(\tau)}_{t_1}, \dots, B^{(\tau)}_{t_n})$ are independent of $\mathscr{F}_\tau$ and hence by π-λ systems lemma $B^{(\tau)}$ is independent of $\mathscr{F}_\tau$.

By continuity of B and F,

$$F(B^{(\tau)}_{t_1}, \dots, B^{(\tau)}_{t_n}) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{1}_{\left\{\frac{k-1}{2^n} < \tau \le \frac{k}{2^n}\right\}} F\left(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \dots, B_{\frac{k}{2^n}+t_n} - B_{\frac{k}{2^n}+t_{n-1}}\right) \quad \text{a.s.}$$

and by dominated convergence theorem

$$\mathbb{E}\left[F(B^{(\tau)}_{t_1}, \dots, B^{(\tau)}_{t_n})\right] = \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbb{1}_{\left\{\frac{k-1}{2^n} < \tau \le \frac{k}{2^n}\right\}} F\left(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \dots, B_{\frac{k}{2^n}+t_n} - B_{\frac{k}{2^n}+t_{n-1}}\right)\right].$$

For $A \in \mathscr{F}_\tau$ the event $A \cap \{\frac{k-1}{2^n} < \tau \le \frac{k}{2^n}\} \in \mathscr{F}_{\frac{k}{2^n}}$ so is a.s. equal to an event in $\sigma(B_t : t \le \frac{k}{2^n})$. Using simple Markov property (Proposition 3.4.2 (iv)) and repeating the argument above we obtain

$$\begin{aligned} &\mathbb{E}\left[\mathbb{1}_{\left\{A\cap\{\frac{k-1}{2^n} < \tau \le \frac{k}{2^n}\}\right\}} F\left(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \dots, B_{\frac{k}{2^n}+t_n} - B_{\frac{k}{2^n}+t_{n-1}}\right)\right] \\ &= \mathbb{P}\left(A \cap \left\{\frac{k-1}{2^n} < \tau \le \frac{k}{2^n}\right\}\right) \mathbb{E}\left[F(B^{(\tau)}_{t_1}, \dots, B^{(\tau)}_{t_n})\right], \end{aligned}$$

which implies the desired result by taking sum over k. If $\mathbb{P}(\tau = \infty) > 0$ we consider $A \cap \{\tau < \infty\}$ for which the above holds and the result follows. □

We now want to apply this at $\tau = H_a = \inf\{t > 0 : B_t = a\} = \inf\{t \ge 0 : B_t \ge a\}$ for $a > 0$.

Theorem 4.3.2. *Let $S_t := \sup_{u\le t} B_u$. For $a \ge 0$ and $b \le a$ we have*

$$\mathbb{P}(S_t \ge a, B_t \le b) = \mathbb{P}(B_t \ge 2a - b) \quad \forall t \ge 0.$$

In particular S_t and $|B_t|$ have the same distribution.

Proof. We know that $H_a < \infty$ a.s.

$$\mathbb{P}(S_t \ge a, B_t \le b) = \mathbb{P}(H_a \le t, B_t \le b) = \mathbb{P}(H_a \le t, B^{(H_a)}_{t-H_a} \le b - a)$$

since $B^{(H_a)}_{t-H_a} = B_t - B_{H_a} = B_t - a$. By strong Markov Property, note that $B^{(H_a)}$ is independent of $\mathscr{F}_{H_a}$ and hence of H_a, and further that $(-B^{(H_a)}_t)_{t\ge 0}$ is also a Brownian motion hence then couples $(H_a, B^{(H_a)})$ and $(H_a, -B^{(H_a)})$ have the same distribution. Let $\Gamma = \{(s,\omega) \in \mathbb{R}_+ \times \mathscr{C}(\mathbb{R}_+, \mathbb{R}) : s \le t, \omega(t-s) \le b - a\}$ then by the above

$$\begin{aligned}\mathbb{P}(S_t \geq a, B_t \leq b) &= \mathbb{P}\big((H_a, B^{(H_a)}) \in \Gamma\big) \\ &= \mathbb{P}\big((H_a, -B^{(H_a)}) \in \Gamma\big) = \mathbb{P}\big(H_a \leq t, -B^{(H_a)}_{t-H_a} \leq b-a\big) \\ &= \mathbb{P}(H_a \leq t, B_t \geq 2a-b) = \mathbb{P}(B_t \geq 2a-b)\end{aligned}$$

as required, since $\{B_t \geq 2a-b\} \supseteq \{H_a \leq t\}$. Finally

$$\begin{aligned}\mathbb{P}(S_t \geq a) &= \mathbb{P}(S_t \geq a, B_t \geq a) + \mathbb{P}(S_t \geq a, B_t \leq a) \\ &= 2(B_t \geq a) = \mathbb{P}(|B_t| \geq a).\end{aligned}$$

□

Chapter 5
(Sub/super-)Martingales in continuous time

5.1 Definition

We assume that a filtered probability space is given $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$.

Definition 5.1.1. An adapted stochastic process $(X_t)_{t\geq 0}$ such that $X_t \in \mathrm{L}^1(\mathbb{P})$ (i.e. $\mathbb{E}[|X_t|] < \infty$) for any $t \geq 0$, is called

- a martingale if $\mathbb{E}[X_t|\mathscr{F}_s] = X_s \ \forall 0 \leq s \leq t$,
- a super-martingale if $\mathbb{E}[X_t|\mathscr{F}_s] \leq X_s \ \forall 0 \leq s \leq t$,
- a sub-martingale if $\mathbb{E}[X_t|\mathscr{F}_s] \geq X_s \ \forall 0 \leq s \leq t$,

Exercises: Suppose $(Z_t : t \geq 0)$ is an adapted process with independent increments: $\forall 0 \leq s < t$, $Z_t - Z_s$ is independent of $\mathscr{F}_s$. The following give us examples of martingales:

1. if $\forall t \geq 0 \ Z_t \in \mathrm{L}^1$ then $\tilde{Z}_t := Z_t - \mathbb{E}[Z_t]$ is a martingale,
2. if $\forall t \geq 0 \ Z_t \in \mathrm{L}^2$ then $\tilde{Z}_t^2 - \mathbb{E}[\tilde{Z}_t^2]$ is a martingale,
3. if for some $\theta \in \mathbb{R}$ and $\forall t \geq 0 \ \mathbb{E}[e^{\theta Z_t}] < \infty$ then $\frac{e^{\theta Z_t}}{\mathbb{E}[e^{\theta Z_t}]}$ is a martingale.

In particular, B_t, $B_t^2 - t$ and $e^{\theta B_t - \theta^2 t/2}$ are all martingales with respect to $(\mathscr{F}_t)_{t\geq 0}$ for which $(B_t)_{t\geq 0}$ is a Brownian motion.

Remark 5.1.2. A process is a martingale relative to a filtration. An $(\mathscr{F}_t)_{t\geq 0}$-martingale (X_t) is also a martingale with respect to $(\mathscr{F}_t^X)$ — its natural filtration. It may however fail to be martingale with respect to a larger filtration $\mathscr{G}_t \supseteq \mathscr{F}_t$. For example, recall (e.g. see Problem sheet 1) that $\mathbb{E}[B_{\frac{t+s}{2}}|B_t, B_s] = \frac{B_t+B_s}{2}$. It follows that $(B_u : s \leq u \leq t)$ is not a martingale with respect to $\mathscr{G}_u := \mathscr{F}_u^B \vee \sigma(B_t)$ since $\mathbb{E}[B_{\frac{t+s}{2}}|B_t, B_s] = \frac{B_t+B_s}{2} \neq B_s$ a.s.

Remark 5.1.3. Note that a martingale has a constant expectation $\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t|\mathscr{F}_0]] = \mathbb{E}[X_0]$. It is a mathematical model for an equitable game. A super-martingale represents a losing game and sub-martingale is a winning one.

Remark 5.1.4. A set of $(\mathscr{F}_t)_{t\geq 0}$-martingales is a vector space.

Proposition 5.1.5. *Let $(X_t)_{t\geq 0}$ be a martingale (respectively sub-martingale) and $\phi : \mathbb{R} \mapsto \mathbb{R}$ convex (respectively convex and increasing) such that $\mathbb{E}[|\phi(X_t)|] < \infty$ for any $t \geq 0$. Then $(\phi(X_t))_{t\geq 0}$ is a sub-martingale.*

Proof. Conditional Jensen inequality. □

In particular, if $(X_t)_{t\geq 0}$ is martingale with $\mathbb{E}[|X_t|^p] < \infty$, for some $p \geq 1$ and all $t \geq 0$, then $|X_t|^p$ is a sub-martingale. In consequence, $t \mapsto \mathbb{E}[|X_t|^P]$ is non-decreasing.

5.2 Maximal (Doob's) inequalities

We develop now a series of results which allow us to control the running maximum of a (sub)martingale. They will be crucial for establishing convergence theorems for (super-)martingales.

We recall that if $(X_n)_{n\geq 0}$ is a discrete martingale (or a positive submartingale) w.r.t. some filtration $(\mathscr{F}_n)$ then for any $N \geq 0$, $p \geq 1$ and $\lambda > 0$

$$\lambda^p \mathbb{P}\Big(\sup_{n\leq N} |X_n| \geq \lambda\Big) \leq \mathbb{E}\big[|X_N|^p\big]$$

and for any $p > 1$

$$E\big[|X_N|^p\big] \leq E\left[\sup_{n\leq N} |X_n|^p\right] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_N|^p\big].$$

If X is now indexed by $t \in [0,\infty)$ we can take a dense countable set D in $[0,T]$, e.g. $D = \mathbb{Q} \cap [0,T]$, and an increasing sequence of finite subsets $D_n \subseteq D$, $\cup_{n=1}^{\infty} D_n = D$. The above inequalities then hold for X indexed with $t \in D_n \cup \{T\}$. Monotone convergence then yields

$$\begin{aligned} \lambda^p \mathbb{P}\Big(\sup_{t\in D} |X_t| \geq \lambda\Big) &\leq \mathbb{E}\big[|X_T|^p\big] \\ E\left[\sup_{t\in D} |X_t|^p\right] &\leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_T|^p\big] \end{aligned} \tag{5.1}$$

If X has regular sample paths then the supremum over a dense countable set in $[0,T]$ is equal to the supremum over the whole interval $[0,T]$. This gives:

Theorem 5.2.1 (Doob's maximal and L^p inequalities).
If $(X_t)_{t\geq 0}$ is right continuous, a martingale or a positive sub-martingale, then for any $T \geq 0$, $\lambda > 0$

$$\begin{aligned} \lambda^p \mathbb{P}\Big(\sup_{t\le T}|X_t|\ge\lambda\Big) &\le \mathbb{E}\big[|X_T|^p\big], \quad p\ge 1 \\ E\left[\sup_{t\le T}|X_t|^p\right] &\le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_T|^p\big], \quad p>1. \end{aligned} \tag{5.2}$$

As an application of Doob's maximal inequality we derive the following bound:

Proposition 5.2.2. *For any $\lambda > 0$ we have*

$$\mathbb{P}(S_t \ge \lambda t) \le e^{-\frac{\lambda^2 t}{2}}, \quad \textit{where } S_t = \sup_{u\le t} B_u.$$

Proof. Recall that $e^{\alpha B_t - \alpha^2 t/2}$, $t \ge 0$, is a non-negative martingale. It follows that, for $\alpha \ge 0$,

$$\begin{aligned} \mathbb{P}(S_t \ge \lambda t) &\le \mathbb{P}\Big(\sup_{u\le t}\big(e^{\alpha B_u - \alpha^2 t/2}\big) \ge e^{\alpha\lambda t - \alpha^2 t/2}\Big) \\ &\le \mathbb{P}\Big(\sup_{u\le t}\big(e^{\alpha B_u - \alpha^2 u/2}\big) \ge e^{\alpha\lambda t - \alpha^2 t/2}\Big) \\ &\le e^{-\alpha\lambda t + \alpha^2 t/2} \underbrace{\mathbb{E}\left[e^{\alpha B_t - \alpha^2 t/2}\right]}_{=1}. \end{aligned}$$

The bound now follows since $\min_{\alpha\ge 0} e^{-\alpha\lambda t + \alpha^2 t/2} = e^{-\lambda^2 t/2}$ (with minimum achieved when $\alpha = \lambda$). $\square$

Note that the above bound could be also easily derived from the explicit distribution of S_t obtained in Theorem 4.3.2.

Finally, we state another version of Doob's maximal inequality (5.1), this time for a supermartingale, which will be used later in the proof of Theorem 5.3.3.

Proposition 5.2.3. *Let $(X_t : t \ge 0)$ be a supermartingale. Then*

$$\mathbb{P}\Big(\sup_{t\in[0,T]\cap\mathbb{Q}}|X_t|\ge\lambda\Big) \le \frac{1}{\lambda}\left(2\mathbb{E}[|X_T|] + \mathbb{E}[|X_0|]\right), \quad \forall \lambda, T > 0. \tag{5.3}$$

In particular, $\sup_{t\in[0,T]\cap\mathbb{Q}}|X_t| < \infty$ *a.s.*

Proof. This relies on a variant of Doob's maximal inequality for discrete time supermartingales. Let us sketch its proof here. Take a sequence of rational numbers $0 = t_0 < t_1 < \ldots < t_n = T$. Applying Doob's optional stopping theorem to discrete supermartingale $(X_{t_i})_{i=1}^n$ and stopping time $\tau = \min\{t_i : X_{t_i} \ge \lambda\} \wedge T$, we obtain

$$\mathbb{E}[X_0] \ge \mathbb{E}[X_\tau] \ge \lambda\mathbb{P}\Big(\sup_{1\le i\le n} X_{t_i} \ge \lambda\Big) + \mathbb{E}[X_T \mathbf{1}_{\sup_{1\le i\le n} X_{t_i} < \lambda}].$$

This leads to

$$\lambda\mathbb{P}\Big(\sup_{1\le i\le n} X_{t_i} \ge \lambda\Big) \le \mathbb{E}[X_0] + \mathbb{E}[X_T^-].$$

On the other hand, the process $X_{t_i}^-$ is a non-negative submartingale so we may apply (5.1) directly to it giving

$$\lambda \mathbb{P}(\sup_{1\le i\le n} X_{t_i}^- \ge \lambda) \le \mathbb{E}[X_T^-].$$

Combining, using $\mathbb{E}[X_T^-] \le \mathbb{E}[|X_T|]$, and taking limit in $n \to \infty$ with nested sequences in $[0,T]\cap\mathbb{Q}$, gives the desired inequality. □

5.3 Convergence and regularisation theorems

In many instances so far we saw that processes with right-continuous paths are much easier to deal with. In this section we show that when looking at martingales there is no loss of generality to assume such regularity, provided the filtration satisfies the usual conditions.

We first recall the Doob's upcrossing lemma in discrete time:

Lemma 5.3.1. *Let $(X_t)_{t\ge 0}$ be a supermartingale and F a finite subset of $[0,T]$. If $a<b$ and $U\big([a,b],(X_t : t\in F)\big)$ is the number of upcrossings of $[a,b]$ by $(X_t : t\in\mathbb{F})$ then we have*

$$\mathbb{E}\Big[U\big([a,b],(X_t : t\in F)\big)\Big] \le \sup_{t\in F}\frac{\mathbb{E}\big[(X_t-a)^-\big]}{b-a} \le \frac{\mathbb{E}\big[(X_T-a)^-\big]}{b-a}.$$

The last inequality follows since $(X_t-a)^-$ is a submartingale. Further, taking increasing sequence F_n, $\cup_n F_n = F$, the above Lemma instantly extends to a countable $F\subset[0,T]$. This allows us to deduce two important results. First, non-negative supermartingales with regular paths converge. More generally,

Theorem 5.3.2. *If (X_t) is a right-continuous super-martingale and $\sup_t \mathbb{E}[X_t^-]<\infty$ then $\lim_{t\to\infty} X_t$ exists a.s. In particular, a non-negative right-continuous supermartingale converges a.s. as $t\to\infty$.*

Proof. This is a direct consequence of Lemma 5.3.1 and the fact that for a right-continuous process upcrossings of (a,b) over $[0,\infty)$ are the same as upcrossings of (a,b) over $t\in[0,\infty)\cap\mathbb{Q}$. □

Secondly, for a general super-martingale we can talk about its left- and right- limits.

Theorem 5.3.3. *If $(X_t : t\ge 0)$ is a supermartingale then for $\mathbb{P}$-almost every $\omega\in\Omega$,*

$$\forall t\in(0,\infty) \quad \lim_{r\nearrow t, r\in\mathbb{Q}} X_r(\omega) \text{ and } \lim_{r\searrow t, r\in\mathbb{Q}} X_r(\omega) \text{ exist and are finite.} \tag{5.4}$$

Proof. Recall that a sequence of numbers (x_n) converges (possibly to $\pm\infty$) if and only if $U([a,b],(x_n))<\infty$ for all $a<b$, $a,b\in\mathbb{Q}$. Now fix $T>0$. From Proposition

5.2.3 and Lemma 5.3.1 that there exists a $\Omega^T \subseteq \Omega$, with $\mathbb{P}(\Omega^T) = 1$, such that for any $\omega \in \Omega^T$

$$\forall_{\substack{a<b \\ a,b\in\mathbb{Q}}} U\big([a,b],(X_t(\omega) : t \in [0,T]\cap\mathbb{Q})\big) < \infty \qquad \text{and} \qquad \sup_{t\in[0,T]\cap\mathbb{Q}} |X_t(\omega)| < \infty.$$

It follows that the limits in (5.4) are well defined and finite for all $t \leq T$ and $\omega \in \Omega^T$. We may then take $\Omega := \Omega^1 \cap \Omega^2 \cap \Omega^3 \cap \ldots$ and the result follows. □

The above result shows that even if X is not right-continuous its right-continuous version is a.s. well defined. This leads to the following fundamental regularisation result.

Theorem 5.3.4. *Let X be a submartingale with respect to a right-continuous and complete filtration $(\mathscr{F}_t)$. If $t \mapsto \mathbb{E}[X_t]$ is right continuous (e.g. if X is a martingale) then X admits a modification with càdlàg (RCLL) paths which is also an $(\mathscr{F}_t)$-submartingale.*

Remark 5.3.5. If X is a martingale then its càdlàg modification above is a martingale.

Proof. By Theorem 5.3.3 and properties of the filtration, the process

$$\tilde{X}_t(\omega) = \begin{cases} \lim_{r\searrow t, r\in\mathbb{Q}} X_r(\omega) & \omega \in \Omega_0 \\ 0 & \omega \notin \Omega_0 \end{cases}$$

is well defined and adapted to $(\mathscr{F}_t)$, for some $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$. In particular, $\tilde{X}$ has càdlàg paths. Let $t_n \searrow t$ be a sequence of rationals. Then (X_{t_n}) is a backwards submartingale, $\sup_n \mathbb{E}[|X_{t_n}|] < \infty$ and hence X_{t_n} converges to $\tilde{X}_t = X_{t+}$ in L^1. This allows us to take limits in

$$X_t \leq \mathbb{E}[X_{t_n}|\mathscr{F}_t], \text{ to obtain } X_t \leq \mathbb{E}[\tilde{X}_t|\mathscr{F}_t] = \tilde{X}_t.$$

In particular, $\mathbb{E}[\tilde{X}_t - X_t] \geq 0$. But then right-continuity of $t \mapsto \mathbb{E}[X_t]$ implies $\mathbb{E}[\tilde{X}_t - X_t] = 0$ so that $X_t = X_{t+}$ a.s.

Finally, let $s < t$ and $s_n \searrow s$ rationals smaller than t. Then

$$X_{s_n} \leq \mathbb{E}[X_t|\mathscr{F}_{s_n}] = \mathbb{E}[X_{t+}|\mathscr{F}_{s_n}].$$

Taking limits as $n \to \infty$, we have $X_{s_n} \to X_{s+} = \tilde{X}_s$ a.s. and $\mathbb{E}[X_{t+}|\mathscr{F}_{s_n}] \to \mathbb{E}[X_{t+}|\mathscr{F}_{s+}] = \mathbb{E}[X_{t+}|\mathscr{F}_s]$ a.s. and we conclude that $\tilde{X}$ is an $(\mathscr{F}_s)$ sub-martingale. □

From now onwards, unless otherwise stated, we only consider right-continuous (sub-/super-) martingales (with left limits) and filtrations satisfying the usual assumptions.

5.4 Optional sampling (stopping) theorem(s)

We saw above that a (super/sub) martingale will often converge a.s. to some X_∞. We want to ask if the martingale property is preserved if we include X_∞. Further we want to understand if the martingale property is preserved under stopping.

Theorem 5.4.1. *Let $(X_t : t \geq 0)$ be a (càdlàg) martingale. TFAE:*

(i) X_t converges in L^1 as $t \to \infty$ to some X_∞
(ii) $\exists\, X_\infty$ in L^1 s.t. $X_t = \mathbb{E}[X_\infty|\mathscr{F}_t]$ $\forall t \geq 0$
(iii) the family $\{X_t : t > 0\}$ is uniformly integrable (UI).

Under these conditions, we also have $X_t \to X_\infty$ a.s.

Proof. (ii) $\Rightarrow$ (iii) Let $\Gamma_t = \{|\mathbb{E}[X_\infty|\mathscr{F}_t]| > \alpha\}$ and

$$a_t := \int_{\Gamma_t} |\mathbb{E}[X_\infty|\mathscr{F}_t]| d\mathbb{P} \leq \int_{\Gamma_t} |\mathbb{E}[|X_\infty||\mathscr{F}_t] d\mathbb{P} = \int_{\Gamma_t} |X_\infty| d\mathbb{P}.$$

On the other hand, Markov's inequality gives

$$\mathbb{P}(\Gamma_t) \leq \frac{1}{\alpha}\mathbb{E}\Big[\big|\mathbb{E}[X_\infty|\mathscr{F}_t]\big|\Big] \leq \frac{\mathbb{E}[|X_\infty|]}{\alpha}$$

and hence $\sup_t a_t \to 0$ as $\alpha \to \infty$.

(iii) $\Rightarrow$ (i) $\sup_t \mathbb{E}[|X_t|] < \infty$ and hence by Theorem 5.3.2 $X_t \to X_\infty$ a.s. for some X_∞. Therefore X_t converges to X_∞ in L^1 by uniform integrability.

(i) $\Rightarrow$ (ii) $X_t = \mathbb{E}[X_{t+s}|\mathscr{F}_t] \to \mathbb{E}[X_\infty|\mathscr{F}_t]$ as $s \to \infty$ by continuity of conditional expectation with respect to L^1 convergence.

□

Remark 5.4.2. If $\sup_t \mathbb{E}[|X_t|^p] < \infty$ for some $p > 1$ then the above conditions hold and convergence is in L^p. Note however that $p = 1$ is not enough.

Theorem 5.4.3. *If X is a UI martingale then in fact $\{X_\tau : \tau$ is a stopping time $\}$ is UI and for any two stopping times $\tau \leq \rho$*

$$X_\tau = \mathbb{E}[X_\rho|\mathscr{F}_\tau] = \mathbb{E}[X_\infty|\mathscr{F}_\tau] \quad a.s.$$

Remark 5.4.4. The UI assumption is important. If $0 < a < b$ then $H_a(B) \leq H_b(B)$ and

$$\mathbb{E}[B_{H_b}|\mathscr{F}_{H_a}] = \mathbb{E}[b|\mathscr{F}_{H_a}] = b \neq a = B_{H_a}.$$

Corollary 5.4.5. *In particular for any martingale and two bounded stopping times $\tau \leq \rho$*

$$\mathbb{E}[X_\rho|\mathscr{F}_\tau] = X_\tau \quad a.s.$$

Proof. Apply the previous result to $X_{\cdot\wedge M}$, where $\tau \leq \rho \leq M$. □

The above two results, Theorem 5.4.3 and Corollary 5.4.5, are known as the *Optional Sampling Theorem* and are an essential tool in many computations involving martingales. To prove Theorem 5.4.3, we first establish the following lemma.

Lemma 5.4.6. *Let (M_n) be a discrete time martingale w.r.t. $(\mathscr{G}_n)_{n\geq 0}$ and $\tau \leq \rho$ two bounded integer valued stopping times. Then*

$$M_\tau = \mathbb{E}[M_\rho|\mathscr{G}_\tau].$$

Proof. Recall that if φ is bounded process with φ_n being $\mathscr{G}_{n-1}$ measurable then

$$(M\cdot\varphi)_n := \sum_{k=1}^{n} \varphi_k(M_k - M_{k-1})$$

is an $\mathscr{G}_n$-martingale and hence $\mathbb{E}[(M\cdot\varphi)_n] = 0$.

Take $\varphi_n = \mathbb{1}_{\{\tau\geq n\}} - \mathbb{1}_{\{\rho\geq n\}} = \mathbb{1}_{\{\tau>n-1\}} - \mathbb{1}_{\{\rho>n-1\}}$. Then for $n \geq K \geq \rho \geq \tau$ a.s. we have

$$(M\cdot\varphi)_n - M_0 = M_\tau - M_\rho$$

so that $\mathbb{E}[M_\tau] = \mathbb{E}[M_\rho]$.

Let $\Gamma \in \mathscr{G}_\tau$ and $\tau^\Gamma := \tau\mathbb{1}_\Gamma + K\mathbb{1}_{\Gamma^\complement}$; $\rho^\Gamma := \rho\,\mathbb{1}_\Gamma + K\mathbb{1}_{\Gamma^\complement}$. Recalling item (v) in Proposition 4.2.4, we see that τ^Γ and ρ^Γ are (bounded) integer valued stopping times and by the above

$$\mathbb{E}\Big[M_\tau\mathbb{1}_\Gamma + M_K\mathbb{1}_{\Gamma^\complement}\Big] = \mathbb{E}\Big[M_\rho\,\mathbb{1}_\Gamma + M_K\mathbb{1}_{\Gamma^\complement}\Big]$$

i.e. $\mathbb{E}[M_\tau\mathbb{1}_\Gamma] = \mathbb{E}[M_\rho\,\mathbb{1}_\Gamma]$, which implies

$$\mathbb{E}[M_\rho|\mathscr{G}_\tau] = M_\tau \quad \text{a.s.}$$

□

Proof of Theorem 5.4.3. Lemma 5.4.6 applies to any discrete time martingale. It follows that if τ,ρ take values in a finite set in $[0,+\infty]$ then the statement is true.

Recall that a family U of r.v. $\mathbb{E}[X_\infty|\mathscr{G}]$ where $\mathscr{G}$ runs through all sub σ-fields of $\mathscr{F}$ is UI. Its closure in L^1, $\bar{U}$ is also UI.

Now, as in Proposition 4.2.4, for any stopping time τ we have $\tau_n \searrow \tau$ and τ_n takes finitely many values. Hence $X_{\tau_n} = \mathbb{E}[X_\infty|\mathscr{F}_{\tau_n}] \in U$ and $X_{\tau_n} \to X_\tau$ a.s. by right continuity. It follows that $Y_{-n} := X_{\tau_n}$, $n \geq 0$ is a backward martingale w.r.t. ($\mathscr{G}_{-n} := \mathscr{F}_{\tau_n}$, $n \in \mathbb{N}$) and hence that $X_{\tau_n} \to X_\tau$ a.s. and in L^1. It follows that $\{X_\tau : \tau \text{ is a stopping time }\}$ is UI. Further, for any $\Gamma \in \mathscr{F}_\tau \subseteq \mathscr{F}_{\tau_n}$

$$E[X_{\tau_n}\mathbb{1}_\Gamma] = \mathbb{E}[X_\infty\mathbb{1}_\Gamma]$$

implies $E[X_\tau\mathbb{1}_\Gamma] = \mathbb{E}[X_\infty\mathbb{1}_\Gamma]$ which further implies

$$X_\tau = \mathbb{E}[X_\infty|\mathscr{F}_\tau] \quad \text{a.s.}$$

In particular, if $\tau \leq \rho$, $X_\tau = \mathbb{E}[X_\rho|\mathscr{F}_\tau]$ □

Theorem 5.4.7. *If X is a supermartingale and $\tau \leq \rho$ bounded stopping times, then*

$$\mathbb{E}[X_\rho|\mathscr{F}_\tau] \leq X_\tau \quad \textit{a.s.}$$

Proof. Exercise. □

Theorem 5.4.8. *An adapted càdlàg process X is a martingale if and only if for every bounded stopping time τ, $X_\tau \in L^1$ and $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.*

Proof. $\Rightarrow$ by previous results
$\Leftarrow$ Let $A \in \mathscr{F}_s$ and $\tau := t\mathbb{1}_{A^\complement} + s\mathbb{1}_A$ a stopping time for $0 \leq s < t$. So

$$\mathbb{E}[X_0] = \mathbb{E}[X_\tau] = \mathbb{E}[X_t\mathbb{1}_{A^\complement}] + \mathbb{E}[X_s\mathbb{1}_A].$$

On the other hand, since t is also a stopping time, we have $\mathbb{E}[X_0] = \mathbb{E}[X_t] = \mathbb{E}[X_t\mathbb{1}_{A^\complement}] + \mathbb{E}[X_s\mathbb{1}_A]$. Combining we obtain

$$\begin{aligned} &\mathbb{E}[X_t] = \mathbb{E}[X_t\mathbb{1}_{A^\complement}] + \mathbb{E}[X_s\mathbb{1}_A] \\ \Rightarrow\ &\mathbb{E}[X_t\mathbb{1}_A] = \mathbb{E}[X_s\mathbb{1}_A] \end{aligned}$$

i.e. $X_s = \mathbb{E}[X_t|\mathscr{F}_s]$ a.s. □

Corollary 5.4.9. *If X is a martingale and τ a stopping time then $X^\tau = (X_{\tau\wedge t} : t \geq 0)$ is a martingale relative to both $(\mathscr{F}_t)$ and $(\mathscr{F}_{t\wedge\tau})$.*

Proof. If ρ is a bounded stopping time, then so is $\tau\wedge\rho$ and we can apply Theorem 5.4.8 to X^τ:

$$\mathbb{E}[X^\tau_\rho] = \mathbb{E}[X_{\tau\wedge\rho}] = \mathbb{E}[X_0] = \mathbb{E}[X^\tau_0] \implies X^\tau \text{ is } (\mathscr{F}_t)\text{-martingale.}$$

$\mathbb{E}[X_{\tau\wedge t}|\mathscr{F}_{\tau\wedge s}] = X_{\tau\wedge s}$ directly by stopping theorem, Corollary 5.4.5. □

We conclude with a simple application. Let

$$H_a = H_a(B) = \begin{cases} H_{[a,\infty)}(B) & a \geq 0 \\ H_{(-\infty,a]}(B) & a < 0, \end{cases}$$

which is a stopping time by Exercise 4.2.2.

Proposition 5.4.10. *The Laplace transforms of the distributions of H_a and $H_a \wedge H_{-a}$ are given by*

$$\mathbb{E}[e^{-\lambda H_a}] = e^{-|a|\sqrt{2\lambda}}, \quad \mathbb{E}[e^{-\lambda H_a \wedge H_{-a}}] = \frac{1}{\cosh(|a|\sqrt{2\lambda})}$$

Proof. With no loss of generality, by symmetry of BM, take $a > 0$. Let $\theta > 0$ and recall that $M_t = \exp\left(\theta B_t - \frac{\theta^2}{2}t\right)$ is a martingale. Further, $M_{t\wedge H_a}$ is bounded by $e^{\theta a}$ and hence M^{H_a} is also UI. We know that $H_a < \infty$ a.s. We get

$$\mathbb{E}[M^s_{H_a}] = \mathbb{E}[M^s_0] = 1 \quad \Rightarrow \quad \mathbb{E}\left[e^{-\frac{\theta^2}{2}H_a}\right] = e^{-\theta a}$$

and taking $\theta = \sqrt{2\lambda}$ we obtain the first desired formula. The second one is left as an exercise. □

Note that, in particular, if $\eta_a \sim H_a$ and $\eta_b \sim H_b$ are independent then $\eta_a + \eta_b \sim H_{a+b}$ which one may argue directly using the strong Markov property.

Chapter 6
Continuous semimartingales (the integrators)

We will see that there are at least two natural operations which we would expect to operate from martingales to martingales but which in fact lead to a larger class of processes called local martingales. This includes the operation $(\varphi_t) \mapsto \int \varphi dM_t$. We will later define this integral and show that it exhibits many desirable features. We can then go on to ask if such an integral may be extended to more general integrators. Naturally, we can add a finite variation term for which the Stieltjes-Lebesgue integral is well defined. This leads to

$$\text{semimartingale} \;=\; \text{local martingale} \;+\; \text{finite variation process.}$$

It then turns out this class of processes is closed under stochastic integration and when taking (C^2) functions on it, as well as under equivalent changes of measure. We shall here equip ourselves with tools to work with them and in the subsequent chapter apply these to define stochastic integrals.

Throughout the chapter, we assume that a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathbb{P})$ satisfying the usual hypothesis is given.

6.1 Functions of finite variation

We consider first functions. Throughout this section we only consider real-valued right-continuous functions on $[0,\infty)$. Our arguments will be shift invariant so, with no loss of generality, we assume that any such function a satisfies $a(0) = 0$.

Definition 6.1.1. The (total) variation of a function a over $[0,T]$ is defined as

$$V(a)_T = \sup_{\Delta} \sum_{i=0}^{n_\Delta - 1} |a_{t_{i+1}} - a_{t_i}|,$$

where $\Delta = \{0 = t_0 < t_1 < \ldots < t_{n_\Delta} = T\}$ is a partition of $[0,T]$.
We say that a is of finite variation on $[0,T]$ if $V(a)_T < \infty$. Function a is of *finite variation* if $V(a)_T < \infty$ for all $T \geq 0$ and of bounded variation if $\lim_{T\nearrow\infty} V(a)_T < \infty$.

Remark 6.1.2. Note that $t \to V(a)_t$ is non-negative, right-continuous and non-decreasing in t. This follows since any partition of $[0,s]$ may be included in a partition of $[0,t]$, $t \geq s$.

Our aim is define integration with respect to a function of finite variation. Recall that there is a one–to–one correspondence between non-decreasing functions a and measures on $(0,\infty)$ given by $a(t) = \mu((0,t])$, i.e. a is the CDF of μ. This allows to define integration with respect to a non-decreasing a. In order to extend this to a of finite variation we need to build a similar generalised equivalence. The first step is achieved with the following characterisation of functions of finite variation.

Proposition 6.1.3. *a is of finite variation if and only if it is equal to a difference of two non-decreasing functions.*

Proof.
"$\Leftarrow$": Let $a(t) = a_+(t) - a_-(t)$ for two non-decreasing functions $a_\pm$. Then

$$V(a)_t \leq V(a_+)_t + V(a_-)_t = a_+(t) - a_-(0) + a_-(t) - a_-(0) < \infty.$$

"$\Rightarrow$" Note that

$$a(t) = \frac{V(a)_t + a(t)}{2} - \frac{V(a)_t - a(t)}{2} \tag{6.1}$$

and the two functions are non-decreasing. □
This suggests that the following object on the measure side of equivalence should be suitable.

Definition 6.1.4. A *generalised (or signed) measure* on a measurable space $(E,\mathscr{E})$ is a difference of two σ–finite measures on $(E,\mathscr{E})$.

A signed measure has a unique decomposition $\mu = \mu_+ - \mu_-$ where μ_+ and μ_- are non-negative measures with disjoint supports. Indeed let $\mu = \mu_1 - \mu_2$ be some decomposition, which exists by definition, and let $\nu := \mu_1 + \mu_2$. Radon-Nikodym theorem gives

$$\mu_1(\mathrm{d}x) = h_1(x)\nu(\mathrm{d}x) \text{ and } \mu_2(\mathrm{d}x) = h_2(x)\nu(\mathrm{d}x).$$

Letting $h(x) := h_1(x) - h_2(x)$ we obtain

$$\mu(\mathrm{d}x) = h(x)\nu(\mathrm{d}x) = \underbrace{h^+(x)\nu(\mathrm{d}x)}_{=:\mu_+(\mathrm{d}x)} - \underbrace{h^-(x)\nu(\mathrm{d}x)}_{=:\mu_-(\mathrm{d}x)}$$

with μ_+ supported on $S^\mu_+ := \{x : h(x) > 0\}$ and μ_- on $S^\mu_- := \{x : h(x) < 0\}$, i.e.

$$\mu_+(\Gamma) = \mu(\Gamma \cap S^\mu_+) \quad \text{and} \quad \mu_-(\Gamma) = \mu(\Gamma \cap S^\mu_-), \quad \Gamma \in \mathscr{E}.$$

Further, the supports are disjoint: $S^\mu_+ \cap S^\mu_- = \emptyset$, as required. It follows also that $\mu_+(\Gamma) = \sup\{\mu(C) : C \subseteq \Gamma, C \in \mathscr{C}\}$ and hence μ_+ is defined in a unique way. Finally, we could have also taken $S^\mu_- = E \setminus S^\mu_+ = \{x : h(x) \leq 0\}$. The decomposition of E into $\setminus S^\mu_+$ and $\setminus S^\mu_-$ is referred to as the *Hahn decomposition* and the decomposition $\mu = \mu_+ - \mu_-$ is referred to as the (associated) *Jordan decomposition* of μ. From now on, we restrict our attention to measures on $(0,\infty)$ with its Borel σ-algebra.

Proposition 6.1.5. *There is a one-to-one correspondence between right-continuous functions of finite variation and signed measures μ on $(0,\infty)$ given by $a(t) = \mu((0,t])$. We further have $a(t-) = \mu((0,t))$ and $a(t) - a(t-) = \mu(\{t\})$ for $t > 0$.*

Proof. This is straightforward given the definition of a signed measure and Proposition 6.1.3. Indeed, a is of finite variation if and only if it is a difference of two non-decreasing functions, $a = a_+ - a_-$ which amounts to saying that μ is a difference of two measures defined by $\mu_\pm([0,t]) = a_\pm(t)$. □
Note that in particular, we may take $a(t)$ which is non-decreasing and then its associated signed measure is simply a measure and a its CDF. If we take $a(t) = t$ then its associated μ is simply the Lebesgue measures on $[0,\infty)$.

In general both a and μ admit many decompositions. We saw however that a signed measure μ admits a *unique* minimal decomposition. The same is true for a function of finite variation and such a decomposition was in fact identified in the proof of Proposition 6.1.3.

Proposition 6.1.6. *Let a be a function of finite variation and μ its associated generalised measure, $a(t) = \mu((0,t])$. Let $\mu = \mu_+ - \mu_-$ be the Jordan decomposition of μ. Then*

$$V(a)_t = \mu_+((0,t]) + \mu_-((0,t]), \quad \textit{and conversely}$$
$$\mu_+((0,t]) = \frac{V(a)_t + a(t)}{2} \textit{ and } \mu_-((0,t]) = \frac{V(a)_t - a(t)}{2}.$$

Proof. TBC – see pb. sheet. □
The above allows us to develop integration with respect to a. Let $|\mu| := \mu_+ + \mu_-$ and consider a measurable $|\mu|$-integrable function $f : [0,\infty) \mapsto \mathbb{R}$, i.e.

$$\int_{(0,t]} |f(s)||\mu|(\mathrm{d}s) = \int_{(0,t]} |f(s)|\mu_+(\mathrm{d}s) + \int_{(0,t]} |f(s)|\mu_-(\mathrm{d}s) < \infty, \quad t \geq 0.$$

Given a function of finite variation a, we say that f is a–integrable if f is $|\mu|$–integrable, where μ is associated to a, which is equivalent to f being $V(a)$–integrable. For such an f, its Lebesgue–Stieltjes integral with respect to a on $(0,t]$ is defined as

$$(f \cdot a)(t) := \int_0^t f(s)\mathrm{d}a(s) = \int_{(0,t]} f(s)\mu(\mathrm{d}s) = \int_{(0,t]} f(s)\big(\mu_+(\mathrm{d}s) - \mu_-(\mathrm{d}s)\big),$$

where $\mu = \mu_+ - \mu_-$ is the Jordan decomposition of μ. Note that

$$\left|\int_0^t f(s)\mathrm{d}a(s)\right| \le \int_0^t |f(s)||\mathrm{d}a(s)|, \quad \text{where } |\mathrm{d}a(s)| = \mathrm{d}V(a)_s = |\mu|(\mathrm{d}s).$$

Proposition 6.1.7. *Let a be of finite variation and f be a–integrable. Then the function $(f\cdot a)$ is right-continuous and of finite variation.*

Proof. Let μ be the generalised measure associated to a and $\mu = \mu_+ - \mu_-$ its Jordan decomposition. We can write

$$(f\cdot a)(t) = \big((f^+\cdot\mu_+)(t) + (f^-\cdot\mu_-)(t)\big) - \big((f^-\cdot\mu_+)(t) + (f^+\cdot\mu_-)(t)\big), \quad t\ge 0,$$

which shows that $(f\cdot a)$ is the difference of two right-continuous non-decreasing functions and hence is right-continuous and, by Proposition 6.1.3, of finite variation. □

With the above decomposition, associativity of a regular integral with respect to a measure extends readily to integrals with respect to functions of finite variation.

Proposition 6.1.8 (Associativity). *Let a be of finite variation as above and f,g measurable functions, f is a-integrable and g is $(f\cdot a)$-integrable. Then gf is a-integrable and*

$$\int_0^t g(s)\mathrm{d}(f\cdot a)(s) = \int_0^t g(s)f(s)\mathrm{d}a(s).$$

We may express the above in a more concise notation as

$$g\cdot(f\cdot a) = gf\cdot a. \tag{6.2}$$

The following is a trivial property in the deterministic setting which becomes very useful in the stochastic setting.

Proposition 6.1.9 (Stopping). *Let a be of finite variation as above. Fix $t\ge 0$ and let $a^t(s) = a(t\wedge s)$. Then a^t is of finite variation and for any measurable a-integrable function f*

$$\int_0^{u\wedge t} f(s)\mathrm{d}a(s) = \int_0^u f(s)\mathrm{d}a^t(s) = \int_0^u f(s)\mathbf{1}_{[0,t]}(s)\mathrm{d}a(s), \quad u\in[0,\infty].$$

Proof. It suffices to observe that if μ is the signed measure associated with a then the measure μ^t associated with a^t is defined via $\mu^t(\Gamma) = \mu(\Gamma\cap[0,t])$, for a Borel set $\Gamma\subset[0,\infty)$. □

Proposition 6.1.10 (Integration by parts). *Let a and b be two continuous functions of finite variation, $a(0)=b(0)=0$, then for any t*

$$a(t)b(t) = \int_0^t a(s)\mathrm{d}b(s) + \int_0^t b(s)\mathrm{d}a(s).$$

Proof. Let μ and ν be their associated signed measure. Then $LHS = \mu\otimes\nu([0,t]^2)$ and the RHS can be computed as

$$\int_0^t a(s)\mathrm{d}b(s) = \mu \otimes \nu\big(\{(s,u) \in (0,t]^2 : s \le u\}\big)$$
$$\int_0^t b(s)\mathrm{d}a(s) = \mu \otimes \nu\big(\{(s,u) \in (0,t]^2 : s \ge u\}\big)$$

and the equality follows since the diagonal has zero measure by continuity of a and b (or equivalently since $\mu(\{s\}) = \nu(\{s\}) = 0\ \forall s$). □

Remark 6.1.11. A function $a \in C^2$ is of finite variation and $\int f(s)\mathrm{d}a(s) = \int f(s)a'(s)\mathrm{d}s$.

Proposition 6.1.12 (Chain-rule). *If F is a C^1 function and a is continuous of finite variation then $F(a(t))$ is also of finite variation and $F(a(t)) = F(a(0)) + \int_0^t F'(a(s))\mathrm{d}a(s)$.*

Proof. The statement is trivially true for $F(x) = x$. Now by Proposition 6.1.10, it is straightforward to check that if the statement is true for F then it is also true for $xF(x)$. Hence by induction, the statement holds for all polynomials. The proof is finished by approximating $F \in C^1$ with a sequence of polynomials. □

Proposition 6.1.13 (Change of variables). *If a is non-decreasing and right-continuous then so is*

$$c(s) := \inf\{t \ge 0 : a(t) > s\},$$

where $\inf\emptyset = +\infty$. Let $a(0) = 0$. Then, for any Borel measurable function $f \ge 0$ on $\mathbb{R}_+$, we have

$$\int_0^\infty f(u)\mathrm{d}a(u) = \int_0^{a(\infty)} f(c(s))\mathrm{d}s.$$

Proof. That c is non-decreasing follows since a is. Further, the set $\{t : a(t) > s\}$ is the union of sets $\{t : a(t) > s + \varepsilon\}$ over $\varepsilon > 0$ which shows c is right-continuous. As for the change of variables formula, for $f = \mathbf{1}_{[0,v]}$, it reads

$$a(v) = \int_0^\infty \mathbf{1}_{c(s) \le v}\mathrm{d}s = \inf\{s : c(s) > v\}$$

and the equality follows from the definition of c. Taking differences, the desired formula holds for the indicators of sets $(u,v]$ and then, by the monotone class theorem, for any function f with compact support. It remains then to take increasing limits to establish the formula in all generality. □

6.2 Process of finite variation

Recall that a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathbb{P})$ satisfying the usual hypothesis is given.

Definition 6.2.1. An adapted right-continuous process $A = (A_t : t \geq 0)$ is called a finite variation process (or a process of finite variation) if $A_0 = 0$ and $t \to A_t$ is (a function) of finite variation a.s.

Remark 6.2.2. Note that we put $A_0 = 0$, which is a convenient normalisation. Since the filtration is complete we may always modify A on a set of measure 0 and hence we may and will assume that all paths are of finite variation.

Proposition 6.2.3. *Let A be a finite variation process and K a progressively measurable process s.t.*

$$\forall t \geq 0 \, \forall \omega \in \Omega \ \int_0^t |K_s(\omega)||\mathrm{d}A_s(\omega)| < \infty.$$

Then $((K \cdot A)_t : t \geq 0)$ defined as $(K \cdot A)_t(\omega) := \int_0^t K_s(\omega)\mathrm{d}A_s(\omega)$ is a finite variation process.

Proof. Clearly $(K \cdot A)_t(\omega)$ is right-continuous and of finite variation so it suffices to check that $((K \cdot A)_t : t \geq 0)$ is adapted. Fix $t > 0$. Consider first h defined by $h(s,\omega) = \mathbb{1}_{(u,v]}(s)\mathbb{1}_\Gamma(\omega)$ for $(u,v] \subseteq [0,t]$ and $\Gamma \in \mathscr{F}_t$. Then

$$(h \cdot A)_t = \mathbb{1}_\Gamma(A_v - A_u)$$

is $\mathscr{F}_t$-measurable. By monotone class theorem, $(h \cdot A)_t$ is $\mathscr{F}_t$-measurable for any $h = \mathbb{1}_G$ with $G \in \mathscr{B}([0,t]) \otimes \mathscr{F}_t$, or more generally any bounded $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$-measurable function h. If h is a general $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$-measurable function satisfying

$$\int_0^t |h(s,\omega)||\mathrm{d}A_s(\omega)| < \infty \quad \forall \omega \in \Omega,$$

then $h = \lim h_n$ is a pointwise limit of simple functions with $|h| \geq |h_n|$ and integrals converge by the dominated convergence theorem and hence $\int_0^t h(s,\omega)\mathrm{d}A_s(\omega)$ is also $\mathscr{F}_t$-measurable. In particular, $(K \cdot A)_t(\omega)$ is $\mathscr{F}_t$-measurable since by progressive measurability, $(s,\omega) \to K_s(\omega)$ on $[0,t]$ is $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$–measurable. □

Finally, for future convenience, we restate the fact that the integral is associative and stable under stopping. This simply combines Propositions 6.1.8 and 6.1.9 above.

Proposition 6.2.4. *Let A be a finite variation process and K, F progressively measurable processes s.t.*

$$\forall t \geq 0 \, \forall \omega \in \Omega \ \int_0^t |K_s(\omega)||\mathrm{d}A_s(\omega)| < \infty, \quad \int_0^t |F_s(\omega)||\mathrm{d}(K \cdot A)_s(\omega)| < \infty.$$

Then the integral is associative:

$$(F \cdot (K \cdot A))_t = (FK \cdot A)_t, \quad t \geq 0. \tag{6.3}$$

Let τ be a stopping time and $A^\tau = (A_{\tau \wedge s} : s \geq 0)$ be the stopped process. Then

$$(K \cdot A^\tau)_t = (K \cdot A)_{\tau \wedge t} = (K\mathbf{1}_{[0,\tau]} \cdot A)_t, \quad t \geq 0. \tag{6.4}$$

6.3 Continuous local martingales

We introduce now the generalisation of martingales which are needed to make our theory of stochastic integration well behaved.

Definition 6.3.1. An adapted process $(M_t : t \geq 0)$ is called a continuous local martingale if it has continuous trajectories a.s. and if there exist a an non-decreasing sequence of stopping times (τ_n) such that $\tau_n \nearrow \infty$ a.s. and $M^{\tau_n} = (M_{t\wedge\tau_n} : t \geq 0)$ is a martingale $\forall n$. We say (τ_n) reduces M.

Remark 6.3.2. Note that in particular $M_0 \in \mathrm{L}^1$ and we may equivalently require that M^{τ_n} is a UI martingale. In fact, by continuity of paths, we may always modify τ_n so that $M^{\tau_n} - M_0$ is bounded and hence UI. It suffices to take $\tau_n \wedge \rho_n$ for $\rho_n = \inf\{t \geq 0 : |M_t - M_0| \geq n\}$. Adding an $\mathscr{F}_0$ measurable random variable M_0 which is in L^1 does not affect the uniform integrability.

Remark 6.3.3. It is sometimes useful to define a local martingale in a slightly more general way without requiring that $M_0 \in \mathrm{L}^1$ but we do not pursue this here.

Proposition 6.3.4. *If M is a continuous local martingale then*

(i) for any stopping time ρ the stopped process M^ρ is also a continuous local martingale;
(ii) if N is also a continuous local martingale then so is $M+N$;
(iii) if M is non-negative (i.e. $\forall t \geq 0$ $M_t \geq 0$ a.s.) then it is a supermartingale;
(iv) if there exists an integrable random variable ξ such that, for all $t \geq 0$, $|M_t| \leq \xi$ a.s. then M is a (true) martingale;
(v) M is a (true) martingale if and only if for any $N > 0$ the class $\{M_\tau : \tau$ is a stopping time with $\tau \leq N\}$ is uniformly integrable.

Proof.

(i) This follows from the fact that a stopped UI martingale is a UI martingale.
(ii) It suffices to observe that the sequence $\tau_n \wedge \rho_n$ reduces $M+N$, where (τ_n) reduces M and (ρ_n) reduces N.
(iii) Exercise.
(iv) This is a consequence of the dominated convergence theorem (for the conditional expectation). Indeed, for $0 \leq s \leq t$

$$M_s = \lim_{n\to\infty} M_s^{\tau_n} = \lim_{n\to\infty} \mathbb{E}[M_t^{\tau_n}|\mathscr{F}_s] \overset{\text{DCT}}{=} \mathbb{E}[\lim_{n\to\infty} M_t^{\tau_n}|\mathscr{F}_s] = \mathbb{E}[M_t|\mathscr{F}_s].$$

(v) If M is a martingale than for any $N > 0$, M^N is a UI martingale and the statement follows from Theorem 5.4.3. For the reverse, let τ be a bounded stopping time and recall that, by Theorem 5.4.8 it suffices to show that M_τ is integrable and $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$. Take N such that $\tau \leq N$ a.s. and let (τ_n) be a localising sequence for M. Then M^{τ_n} is a martingale and, again by Theorem 5.4.8,

$\mathbb{E}[M_{\tau_n\wedge\tau}] = \mathbb{E}[M_0]$. However $\{M_{\tau_n\wedge\tau} : n \geq 0\} \subseteq \{M_\rho : \rho \leq N\}$ and hence is UI by assumption. It follows that

$$M_{\tau_n\wedge\tau} \to M_\tau \text{ a.s. and in } L^1$$

so that $\mathbb{E}[M_\tau] = \lim_{n\to\infty} \mathbb{E}[M_{\tau_n\wedge\tau}] = \mathbb{E}[M_0]$ as required.
□

Theorem 6.3.5. *A continuous local martingale M, $M_0 = 0$ a.s., is a process of finite variation if and only if M is indistinguishable from zero.*

Proof. Assume M is a continuous local martingale and of finite variation. Let

$$\tau_n = \inf\{t \geq 0 : \int_0^t |\mathrm{d}M_s| \geq n\} = \inf\{t \geq 0 : V(M)_t \geq n\}$$

which are stopping times since $V(M)_t = \int_0^t |\mathrm{d}M_s|$ is continuous and adapted.
Let $N = M^{\tau_n}$ is a bounded local martingale:

$$|N_t| = |M_{t\wedge\tau_n}| \leq |\int_0^{t\wedge\tau_n} \mathrm{d}M_u| \leq \int_0^{t\wedge\tau_n} |\mathrm{d}M_u| \leq n,$$

and hence (N_t) is a martingale.
Let $t > 0$ and $\Delta = \{0 = t_0 < t_1 < t_2 < \ldots < t_m = t\}$ a subdivision of $[0,t]$. Then

$$\begin{aligned}
\mathbb{E}[N_t^2] &= \sum_{i=1}^m \mathbb{E}[N_{t_i}^2 - N_{t_{i-1}}^2] = \sum_{i=1}^m \mathbb{E}\big[(N_{t_i} - N_{t_{i-1}})^2\big] \\
&\leq \mathbb{E}\Big[\big(\sup_{1\leq i\leq m} |N_{t_i} - N_{t_{i-1}}|\big) \cdot \underbrace{\sum |N_{t_i} - N_{t_{i-1}}|}_{\leq V(N)_t = V(M)_{t\wedge\tau_n} \leq n}\Big] \\
&\leq n\mathbb{E}\big[\sup_{1\leq i\leq m} |N_{t_i} - N_{t_{i-1}}|\big] \to 0 \quad \text{as } mesh(\Delta) \to 0
\end{aligned}$$

by Dominated Convergence Theorem since $|N_{t_i} - N_{t_{i-1}}| \leq V(N)_t \leq n$.
It then follows by Fatou lemma that

$$\mathbb{E}[M_t^2] = \mathbb{E}[\lim_{n\to\infty} M_{t\wedge\tau_n}^2] \leq \lim_{n\to\infty} \mathbb{E}[M_{t\wedge\tau_n}^2] = 0$$

which implies $M_t = 0$ a.s. which further implies $\mathbb{P}(M_t = 0\ \forall t \geq 0) = 1$ by continuity of paths. □

6.4 Quadratic variation of a continuous local martingale

Recall Definition 3.5.1 of a process admitting quadratic variation and Theorem 3.5.2 where we showed that $\langle B,B\rangle_t = t$, for a Brownian motion B. We prove in this section that any continuous local martingale M is of finite quadratic variation and that the quadratic variation allows us to describe the process M rather well. Indeed, intuitively speaking, a path of a local martingale M has to wiggle a lot and, being a martingale, it has to wiggle "up and down". So to describe it all we have to do is specify its instantaneous "speed". This is precisely what quadratic variation $\langle M,M\rangle$ does. You may think of $\langle M,M\rangle_t$ as the "clock" of the process M.

Theorem 6.4.1. *Let M be a continuous local martingale. There exists a unique (up to indistinguishability) non-decreasing continuous adapted process $(\langle M,M\rangle_t : t \geq 0)$ starting in zero such that $(M_t^2 - \langle M,M\rangle_t : t \geq 0)$ is a continuous local martingale.*

Furthermore, for any $T > 0$ and any sequence of partitions $\Delta_n = \{0 = t_0 < t_1^n < \ldots < t_{m_n}^n = T\}$ with $sup_{1\leq i\leq m_n}(t_i^n - t_{i-1}^n) \to 0$ as $n \to \infty$

$$\langle M,M\rangle_T = \lim_{n\to\infty} \sum_{i=1}^{m_n} (M_{t_i^n} - M_{t_{i-1}^n})^2, \tag{6.5}$$

where the limit is in probability.

The process $\langle M,M\rangle$ is called the quadratic variation of *M, or simply* the increasing process of *M, and is often denoted $\langle M,M\rangle_t = \langle M\rangle_t$.*

Sketch of the Proof (NOT EXAMINABLE).
Uniqueness is a direct consequence of Theorem 6.3.5 since if A,A' are two such processes then $\left(M^2 - A - (M^2 - A')\right) = A - A'$ is a local martingale starting in zero and of finite variation, which implies $A = A'$ by Theorem 6.3.5.

For existence, assume first that M is bounded (and hence also a martingale). Take a nested sequence of partitions Δ_n of $[0,T]$ with $mesh(\Delta_n) \to 0$. Then, with $\Delta_n = \{0 = t_0^n < t_1^n < \ldots < t_{m_n}^n = T\}$, we check that

$$X_t^n := \sum_{i=1}^{m_n} M_{t_{i-1}^n} (M_{t_i^n \wedge t} - M_{t_{i-1}^n \wedge t})$$

is a martingale. Further (a direct but involved) computation gives

$$\lim_{n,m\to\infty} \mathbb{E}[(X_t^n - X_t^m)^2] = 0$$

and by Doob's L^2-inequality

$$\lim_{n,m\to\infty} \mathbb{E}[\sup_{t\leq T}(X_t^n - X_t^m)^2] = 0.$$

We conclude that, passing to a subsequence, $X^{n_k} \to Y$ a.s. uniformly on $[0,T]$ to a continuous process $(Y_t)_{t\leq T}$. Martingale property of X^{n_k}, $X_t^{n_k} = \mathbb{E}[X_T^{n_k}|\mathscr{F}_t]$, is inherited by Y. Finally, we have

$$M_{t_j^n}^2 - 2X_{t_j^n}^n = \sum_{i=1}^{j} (M_{t_i^n} - M_{t_{i-1}^n})^2 = QV_{t_j^n}^{\Delta_n}(M)$$

and hence is non-decreasing along $t_j^n : j \leq m_n$. Letting $n \to \infty$ we see that $M_t^2 - 2Y_t$ is a.s. non-decreasing on $[0,T]$, and we let $\langle M,M\rangle_t := M_t^2 - 2Y_t$ (we may set it equal to 0 on the null set where it may fail to be non-decreasing if we wished). We also have that $M_t^2 - \langle M,M\rangle_t = 2Y_t$ is a martingale, as required.

We repeat the above argument for $T \in \mathbb{N}$ observing that the resulting $(\langle M,M\rangle_t : t \leq T)$ for T coincides with the one for $T-1$ on $[0,T-1]$. Uniqueness shows $\langle M,M\rangle$ does not depend on the choice of (Δ_n). We also have, by the above,

$$\langle M,M\rangle_t = M_t^2 - 2Y_t = \lim_{n\to\infty} QV_t^{\Delta_n}(M)$$

and limit is in L^2.

Now let M be a general continuous local martingale. Let $N_t = M_t - M_0$ and note that

$$M_t^2 = M_0^2 + \underbrace{2M_0N_t}_{\text{local martingale}} + N_t^2 \Rightarrow \langle M,M\rangle_t = \langle N,N\rangle_t,$$

which implies that it suffices to consider the case $M_0 = 0$ a.s.

Let $\tau_N = H_N \wedge H_{-N}(M) = \inf\{t \geq 0 : |M_t| \geq N\}$. The reasoning above applies to M^{τ_N} and gives us $A^{\tau_N} = \langle M^{\tau_N}, M^{\tau_N}\rangle_t$. Uniqueness implies $\mathbb{P}(A_t^N = A_t^{N-1}\ \forall t \leq \tau_{N-1}) = 1$ so that there exists a non-decreasing process A s.t. $\mathbb{P}(A_t = A_t^N\ \forall t \leq \tau_N) = 1$. By construction then $(M_{t\wedge\tau_N}^2 - A_{t\wedge\tau_N} : t \geq 0)$ is a martingale which means that $M^2 - A$ is a local martingale as required, $\langle M,M\rangle = A$. We also have $\langle M,M\rangle_{\tau_N\wedge t} = \lim_n QV_{\tau-N\wedge t}^{\Delta_n}(M)$ (in L^2 and probability) and observing that $\forall T\ \mathbb{P}(T < \tau_n) \to 1$ we conclude. □

Remark 6.4.2. Note that, by construction, we would expect $Y_t = \int_0^t M_u dM_u$. This will indeed be the case.

Remark 6.4.3. Note that $\langle M^\tau, M^\tau\rangle_t = \langle M,M\rangle_{\tau\wedge t}$ simply by uniqueness and since $(M_{t\wedge\tau}^2 - \langle M,M\rangle_{t\wedge\tau}, t \geq 0)$ is a stopped local martingale and hence a local martingale.

Remark 6.4.4. We often write $\langle M\rangle = \langle M,M\rangle$ for simplicity.

We turn our attention now to martingales with strong integrability properties. Although we continue to talk about processes, we note that all of what follows is invariant if we change a process on a null set. Accordingly, one can think of a martingale M as being defined up to indistinguishability. This is similar to when instead of considering functions we consider their equivalence classes up to an a.e.–equality.

Definition 6.4.5. Let $\mathbb{H}^2$ be the space of L^2-bounded càdlàg martingales i.e.

$$((\mathscr{F}_t),\mathbb{P})\text{–martingales } M \text{ s.t. } \sup_{t\geq 0} \mathbb{E}[M_t^2] < \infty.$$

and H^2 the subspace consisting of continuous L^2-bounded martingales. Finally, let $\mathrm{H}^2_0 = \{M \in \mathrm{H}^2 : M_0 = 0 \text{ a.s.}\}$.

We note that the space H^2 is also sometimes denoted $\mathscr{M}^2$. It follows from Doob's L^2-inequality that

$$\mathbb{E}\left[\sup_{t\geq 0} M_t^2\right] \leq 4 \sup_{t\geq 0} \mathbb{E}[M_t^2] < +\infty, \quad M \in \mathbb{H}^2.$$

In consequence $\{M_t : t \geq 0\}$ is bounded by a square integrable random variable ($\sup_{t\geq 0}|M_t|$) and in particular is uniformly integrable. It follows from Theorem 5.4.1 and Remark 5.4.2 that $M_t = \mathbb{E}[M_\infty|\mathscr{F}_t]$ for some square integrable r.v. M_∞. Note that M_∞ is $\mathscr{F}_\infty$–measurable, where $\mathscr{F}_\infty = \sigma(\mathscr{F}_t : t \geq 0)$. Conversely, for any $Y \in \mathrm{L}^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$, the martingale $M_t := \mathbb{E}[Y|\mathscr{F}_t]$ is in $\mathbb{H}^2$ and $M_\infty = Y$. Clearly, two L^2-bounded martingales M, M' are indistinguishable if and only if $M_\infty = M'_\infty$ a.s. This shows that if we endow $\mathbb{H}$ with the norm

$$\|M\|_{\mathbb{H}^2} := \sqrt{\mathbb{E}[M_\infty^2]} = \|M_\infty\|_{\mathrm{L}^2(\Omega, \mathscr{F}_\infty, \mathbb{P})}, \quad M \in \mathbb{H}^2, \tag{6.6}$$

then $\mathbb{H}^2$ can be identified with the familiar $\mathrm{L}^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$ space.

Theorem 6.4.6. H^2 *is a closed subspace of* $\mathbb{H}^2$.

Proof. Consider a sequence of continuous martingales $M^n \in \mathrm{H}^2$ which converges, in $\|\cdot\|_{\mathbb{H}^2}$ to some $M \in \mathbb{H}^2$. By Doob's L^2-inequalty

$$\mathbb{E}\left[\sup_{t\geq 0} |M_t^n - M_t|^2\right] \leq 4\|M^n - M\|^2_{\mathbb{H}^2} \longrightarrow 0, \quad \text{as } n \to \infty.$$

Passing to a subsequence, we have $\sup_{t\geq 0}|M_t^{n_k} - M_t| \to 0$ a.s. and hence M has continuous paths a.s., $M \in \mathrm{H}^2$ which completes the proof. □

Remark 6.4.7. In functional analytic terms, the above discussion can be summarised by saying that $\mathbb{H}^2$, endowed with the norm $\|\cdot\|_{\mathbb{H}^2}$, is a Hilbert space and so is H^2 with the inherited norm $\|M\|_{\mathrm{H}^2} = \|M\|_{\mathbb{H}^2}$, $M \in \mathrm{H}^2$.

We now re-express the norm in (6.6) for continuous martingales, i.e. on H^2, using the quadratic variation whose existence we established in Theorem 6.4.1.

Theorem 6.4.8. *Let M be a continuous local martingale, $M_0 = 0$ a.s. We have:*

(i) $M \in \mathrm{H}^2$ if and only if $\mathbb{E}[\langle M\rangle_\infty] < \infty$, and then $M^2 - \langle M\rangle$ is a uniformly integrable martingale.

(ii) M is a square integrable martingale, i.e. $\mathbb{E}[M_t^2] < \infty$ for all $t \geq 0$ if and only if $\mathbb{E}[\langle M\rangle_t] < \infty$ for all $t \geq 0$, and then $M^2 - \langle M\rangle$ is a martingale.

Proof. (i) Let $M \in \mathrm{H}^2$ and hence, as argued above, $M_t = \mathbb{E}[M_\infty|\mathscr{F}_t]$ for $M_\infty \in \mathrm{L}^2$. Let $\rho_n = \inf\{t \geq 0 : \langle M\rangle_t \geq n\}$ which is a first hitting time for a continuous adapted

process and hence a stopping time, recall Exercise 4.2.2. It follows that $|M^2_{t\wedge\rho_n} - \langle M\rangle_{t\wedge\rho_n}| \le n + \sup_{t\ge 0} M_t^2$ which is integrable so that by Proposition 6.3.4 $(M - \langle M\rangle)^{\rho_n}$ is a (true) martingale and $\mathbb{E}[M^2_{t\wedge\rho_n}] = \mathbb{E}[\langle M\rangle_{t\wedge\rho_n}]$.

Note that by Dominated Convergence Theorem $\mathbb{E}[M^2_{t\wedge\rho_n}] \to \mathbb{E}[M^2_\rho]$ as $t\to\infty$ and $\mathbb{E}[\langle M\rangle_{t\wedge\rho_n}] \to \mathbb{E}[\langle M\rangle_{\rho_n}]$ as $t\to\infty$ by Monotone Convergence Theorem. Hence $\mathbb{E}[M^2_{\rho_n}] = \mathbb{E}[\langle M\rangle_{\rho_n}]$. The same arguments show that letting $n\to\infty$ we have $\mathbb{E}[M^2_\infty] = \mathbb{E}[\langle M\rangle_\infty] < \infty$. Since $M^2 - \langle M\rangle$ is dominated by $\sup_{t\ge 0} M_t^2 + \langle M\rangle_\infty$, which is integrable, we conclude that $M^2 - \langle M\rangle$ is a UI martingale.

To argue the reverse implication, let M be a continuous local martingale with $\mathbb{E}[\langle M\rangle_\infty] < \infty$. Let $\tau_n = \inf\{t : |M_t| \ge n\}$, a stopping time as above, and note that M^{τ_n} is bounded. By the above argument $M^2_{t\wedge\tau_n} - \langle M\rangle_{t\wedge\tau_n}$ is a UI martingale. If ρ is an a.s. finite stopping time then Optional Sampling Theorem, see Theorem 5.4.3, gives

$$\mathbb{E}[M^2_{\rho\wedge\tau_n}] = \mathbb{E}[\langle M\rangle_{\rho\wedge\tau_n}] \le \mathbb{E}[\langle M\rangle_\infty]$$

Then by Fatou's Lemma

$$\mathbb{E}[M^2_\rho] \le \lim_{n\to\infty} \mathbb{E}[M^2_{\rho\wedge\tau_n}] \le \mathbb{E}[\langle M\rangle_\infty]$$

and so $\{M_\rho : \rho \text{ is a stopping time}\}$ is bounded in L^2 and hence UI. We can conclude using Proposition 6.3.4 (v) or else, take take L^1 limit to get

$$\begin{array}{ccc} \mathbb{E}[M_{t\wedge\tau_n}|\mathscr{F}_s] & = & M_{s\wedge\tau_n} \\ \downarrow & & \downarrow n\to\infty \\ \mathbb{E}[M_t|\mathscr{F}_s] & = & M_s. \end{array}$$

(ii) This follows by applying (i) to M^T for any $T \ge 0$. □

It follows instantly from (i) above that for a martingale M with $M_0 = 0$ we have

$$\|M\|^2_{\mathrm{H}^2} = \mathbb{E}[M^2_\infty] = \mathbb{E}[\langle M\rangle_\infty]$$

which we will use in the sequel. We can also deduce a complement to the previous characterisation of non-zero continuous local martingales.

Corollary 6.4.9. *Let M be a continuous local martingale, $M_0 = 0$. Then the following are equivalent*

(i) M is indistinguishable from zero,
(ii) $\langle M\rangle_t = 0$ for all $t \ge 0$ a.s.
(iii) M is a process of finite variation.

Proof. We saw that (i) is equivalent to (iii) in Theorem 6.3.5. To see that (iii) implies (ii) note that for a partition $\Delta = \{0 = t_0 < t_1 < \ldots < t_{m_\Delta} = t\}$ we have

$$QV^\Delta(M)_t \le \sup_{1\le i\le m_\Delta} |M_{t_i} - M_{t_{i-1}}| \cdot V(M)_t \to 0 \quad \text{as } mesh(\Delta)\to 0.$$

It remains to argue that (ii) implies (i). From (ii) we have $\langle M\rangle_\infty = \lim_{t\to\infty}\langle M\rangle_t = 0$ a.s. It follows from Theorem 6.4.8 that $M \in \mathrm{H}^2$ and $\mathbb{E}[M_\infty^2] = \mathbb{E}[\langle M\rangle_\infty] = 0$ and hence $M_t = 0$ a.s. We conclude by continuity of paths that M is indistinguishable from zero. □

We now extend the notion of quadratic variation to the co-variation between two local martingales through *polarisation*, as is usually done for an inner product.

Definition 6.4.10. The quadratic co-variation (cross-variation) between two continuous local martingales M,N is defined as

$$\langle M,N\rangle := \frac{1}{2}\left(\langle M+N,M+N\rangle - \langle M\rangle - \langle N\rangle\right). \tag{6.7}$$

It is sometimes called the (angle) bracket process of M and N.

Proposition 6.4.11. *For two continuous local martingales M,N*

(i) the process $\langle M,N\rangle$ is the unique finite variation process, zero in zero, such that $(M_tN_t - \langle M,N\rangle_t : t \geq 0)$ is a continuous local martingale.
(ii) the mapping $M,N \to \langle M,N\rangle$ is bi-linear and symmetric.
(iii) for any stopping time τ

$$\langle M^\tau,N^\tau\rangle_t = \langle M^\tau,N\rangle_t = \langle M,N^\tau\rangle_t = \langle M,N\rangle_{\tau\wedge t}, \quad t \geq 0, \text{ a.s.} \tag{6.8}$$

(iv) for any $t > 0$ and a sequence of partitions Δ_n of $[0,t]$ with mesh converging to zero

$$\sum_{t_i\in\Delta_n}(M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}) \to \langle M,N\rangle_t, \tag{6.9}$$

the convergence being in probability.

Remark 6.4.12 (Fundamental remark). It follows from (6.9) that $\langle M,N\rangle$ is invariant under changing the filtration to some other $(\mathscr{G}_t)$ and $\mathbb{P}$ to an absolutely continuous $\mathbb{Q}$ (i.e. $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$) as long as M,N remain $((\mathscr{G}_t),\mathbb{Q})$–local martingales.

Proof. (i) $(M+N)^2 - \langle M+N,M+N\rangle_t$ is a continuous local martingale and by adding and subtracting terms it is equal to

$$\underbrace{M_t^2 - \langle M,M\rangle_t}_{\text{l.mat}} + \underbrace{N^2 - \langle N,N\rangle_t}_{\text{l.mat}} + 2\underbrace{\left(M_tN_t - \frac{1}{2}(\langle M+N,M+N\rangle_t - \langle M,M\rangle_t - \langle N,N\rangle_t)\right)}_{\text{hence also a l.mat}}$$

Uniqueness follows from Theorem 6.3.5.

(iv) Note that

$$(M_t+N_t-M_s-N_s)^2 - (M_t-M_s)^2 - (N_t-N_s)^2 = 2(M_t-M_s)(N_t-N_s).$$

The asserted convergence then follows from Theorem 6.4.1

(ii) Both properties follow from (iv). Symmetry is obvious from the definition in (6.7).

(iii) Follows from (iv). □

Remark 6.4.13. It follows that if M and N are two martingales bounded in L^2, $M_0N_0 = 0$ a.s., then $(M_tN_t - \langle M,N\rangle_t, t \geq 0)$ is a UI martingale. In particular,

$$\mathbb{E}[M_\tau N_\tau] = \mathbb{E}[\langle M,N\rangle_\tau] \quad \forall \text{stopping time } \tau. \tag{6.10}$$

Combining the above result with (ii) in Proposition 6.4.11 we deduce that

Corollary 6.4.14. *The mapping*

$$\mathrm{H}^2_0 \times \mathrm{H}^2_0 \ni (M,N) \to \mathbb{E}[\langle M,N\rangle_\infty] \in \mathbb{R}$$

is an inner product on the vector space H^2_0 *(over* $\mathbb{R}$*) which induces the norm*

$$\mathbb{E}[\langle M,M\rangle_\infty] = \mathbb{E}[M^2_\infty] = \|M\|^2_{\mathrm{H}^2}.$$

In particular, the usual Cauchy–Schwarz inequality gives us

$$|\mathbb{E}[\langle M,N\rangle_\infty]| \leq \sqrt{\mathbb{E}[\langle M\rangle_\infty]}\sqrt{\mathbb{E}[\langle N\rangle_\infty]}.$$

In fact we can easily deduce an a.s. version of this inequality using (6.9). Indeed, applying Cauchy-Schwarz to (6.9) we have

$$\left|\sum_{t_i\in\Delta_n}(M_{t_{i+1}}-M_{t_i})(N_{t_{i+1}}-N_{t_i})\right| \leq \sqrt{\sum_{t_i\in\Delta_n}(M_{t_{i+1}}-M_{t_i})^2}\sqrt{\sum_{t_i\in\Delta_n}(N_{t_{i+1}}-N_{t_i})^2}.$$

Taking limits we deduce that

$$|\langle M,N\rangle_t| \leq \sqrt{\langle M\rangle_t}\sqrt{\langle N\rangle_t} \quad a.s. \tag{6.11}$$

To close this section we establish a more general version of this inequality.

Theorem 6.4.15 (Kunita-Watanabe inequality). *Let M,N be continuous local martingales and K,H two measurable processes. Then*

$$\forall_{0\leq t\leq\infty}\int_0^t |H_s||K_s||\mathrm{d}\langle M,N\rangle_s| \leq \left(\int_0^t H_s^2\mathrm{d}\langle M\rangle_s\right)^{1/2}\left(\int_0^t K_s^2\mathrm{d}\langle N\rangle_s\right)^{1/2} \ a.s. \tag{6.12}$$

In particular, for any $p \geq 1$, $1/p+1/q = 1$ and $0 \leq t \leq \infty$, we have

$$\mathbb{E}\left[\int_0^t |H_s||K_s||\mathrm{d}\langle M,N\rangle_s|\right] \leq \left(\mathbb{E}\left[\left(\int_0^t H_s^2\mathrm{d}\langle M\rangle_s\right)^{p/2}\right]\right)^{1/p}\left(\mathbb{E}\left[\left(\int_0^t K_s^2\mathrm{d}\langle N\rangle_s\right)^{q/2}\right]\right)^{1/q}.$$

Proof (NOT EXAMINABLE, SKETCH).
The second statement follows from the first by Hölder's inequality. Also, it suffices to show (6.12) for a finite t since both LHS and RHS are non-decreasing in t.

First, the same arguments as for (6.11) above but considering subdivisions of $[s,t]$ gives that

$$|\langle M,N\rangle_s^t| \le \sqrt{\langle M\rangle_s^t}\sqrt{\langle N\rangle_s^t} \quad \forall\, s \le t \text{ a.s.}$$

where $\langle M,N\rangle_s^t = \langle M,N\rangle_t - \langle M,N\rangle_s$. Then, for a subdivision of $[s,t]$,

$$\begin{aligned}\sum |\langle M,N\rangle_{t_{i-1}}^{t_t}| &\le \sum \sqrt{\langle M\rangle_{t_{i-1}}^{t_i}}\sqrt{\langle N\rangle_{t_{i-1}}^{t_i}} \\ &\le \sqrt{\sum \langle M\rangle_{t_{i-1}}^{t_i}}\sqrt{\sum \langle N\rangle_{t_{i-1}}^{t_i}} = \sqrt{\langle M\rangle_s^t}\sqrt{\langle N\rangle_s^t}\end{aligned}$$

and hence $\int_s^t |\mathrm{d}\langle M,N\rangle_u| \le \sqrt{\langle M\rangle_s^t}\sqrt{\langle N\rangle_s^t}$. More generally, considering subdivisions of a set A which is a union of intervals, this generalises to

$$\int_A |\mathrm{d}\langle M,N\rangle_u| \le \sqrt{\int_A |\mathrm{d}\langle M\rangle_u|}\sqrt{\int_A |\mathrm{d}\langle N\rangle_u|} \quad \text{a.s.} \tag{6.13}$$

which then extends to any $A \in \mathscr{B}([0,t])$ by a Monotone Class argument.

Let us fix $\omega \in \Omega$ such that (6.13) holds and evaluate processes on this ω. First, let consider simple functions: $h = \sum \lambda_i \mathbb{1}_{A_i}$, $k = \sum \mu_i \mathbb{1}_{A_i}$ with $\lambda_i, \mu_i \ge 0$, $A_i \in \mathscr{B}([0,t])$. We then have

$$\begin{aligned}\int_0^t h(s)k(s)|\mathrm{d}\langle M,N\rangle_s| &= \sum \lambda_i\mu_i \int_{A_i} |\mathrm{d}\langle M,N\rangle_u| \\ &\le \sqrt{\sum \lambda_i^2 \int_{A_i} |\mathrm{d}\langle M\rangle_u|}\sqrt{\sum \mu_i^2 \int_{A_i} |\mathrm{d}\langle M\rangle_u|} \\ &= \sqrt{\int_0^t h(s)^2 \mathrm{d}\langle M\rangle_s}\sqrt{\int_0^t k(s)^2 \mathrm{d}\langle N\rangle_s}.\end{aligned}$$

The result now follows by writing a non-negative measurable function as an increasing sequence of simple functions. □

6.5 Continuous semimartingales

Definition 6.5.1. A stochastic process $X = (X_t : t \ge 0)$ is called a continuous semimartingale if it can be written as

$$X_t = X_0 + M_t + A_t, \quad t \ge 0 \tag{6.14}$$

where M is a continuous local martingale and A is a continuous process of finite variation and $M_0 = A_0 = 0$ a.s.

Remark 6.5.2. The above decomposition is unique by Theorem 6.3.5.

Remark 6.5.3. X is a semimartingale relative to $(\mathscr{F}_t)$ and $\mathbb{P}$.

Proposition 6.5.4. *A continuous semimartingale is of finite quadratic variation and* $\langle X,X\rangle = \langle M,M\rangle$.

Proof. Fix $t \geq 0$ and consider a sequence of subdivisions of $[0,t]$, $\Delta_m = \{0 = t_0 < t_1 < \ldots < t_{n_m} = t\}$ with $mesh(\Delta_m) \to 0$ as $m \to \infty$. Then

$$\sum_{i=1}^{n_m}(X_{t_i}-X_{t_{i-1}})^2 = \underbrace{\sum_{i=1}^{n_m}(M_{t_i}-M_{t_{i-1}})^2}_{(i)} + \underbrace{\sum_{i=1}^{n_m}(A_{t_i}-A_{t_{i-1}})^2}_{(ii)} + \underbrace{2\sum_{i=1}^{n_m}(M_{t_i}-M_{t_{i-1}})(A_{t_i}-A_{t_{i-1}})}_{(iii)}.$$

It follows from the properties of M and A that, as $m \to \infty$,

$$\begin{aligned}
&(i) \to \langle M,M\rangle_t \\
&(ii) \leq \sup_{1\leq i\leq m}|A_{t_i}-A_{t_{i-1}}|\cdot V_t(A) \to 0 \quad \text{a.s.} \\
&(iii) \leq \sup_{1\leq i\leq m}|M_{t_i}-M_{t_{i-1}}|\cdot V_t(A) \to 0 \quad \text{a.s.}
\end{aligned}$$

□

If X,Y are two continuous semimartingales we may define their co-variation $\langle X,Y\rangle$ via the usual polarisation formula as in (6.7). It follows instantly that the properties (ii)–(iv) in Proposition 6.4.11 extend to co-variation between continuous semimartingales. It also follows that if $X_t = X_0 + M_t + A_t$, $Y_t = Y_0 + N_t + A'_t$ then $\langle X,Y\rangle_t = \langle M,N\rangle_t$. In particular, $\langle M,A\rangle \equiv 0$.

Chapter 7
Stochastic Integration

Differential equations are a ubiquitous tool of mathematical models. The classical calculus, going back to Newton and Leibnitz, defines differentiation and integration and links the two operations via the fundamental theorem of calculus. This link, together with chain rule or the integration by parts formula, allows us to formulate and solve differential equations laying the cornerstone for successful modelling of deterministic systems.

However in many real world applications, from finance through population dynamics to geology, it is more natural to assume that our system exhibits a random component. The quantities of interest will be driven by deterministic factors as well as a random noise. A simple equation, embodying this double nature, could be written as

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \quad t \geq 0, \tag{7.1}$$

where B is a Brownian motion. We do not really know what $\mathrm{d}B_t$ means since B_t is nowhere differentiable as a function of t. In any case, if we are to make sense of this we would like to define both this differential, as well as an integral w.r.t. B, in such a way that the "fundamental theorem of calculus" still works: $\mathrm{d}(\int_0^t K_s \mathrm{d}B_s) = K_t \mathrm{d}B_t$. We will thus take this relation as the definition and understand (7.1) purely as a shorthand notation for

$$X_t = X_0 + \int_0^t f(X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}B_s, \quad t \geq 0. \tag{7.2}$$

For a nice f, e.g. f bounded, there is no ambiguity in defining the first integral. More generally, we know how to define $\int_0^t f(X_s)\mathrm{d}A_s$, where A is of finite variation. We simply define the integral ω by ω, i.e.

$$\left(\int_0^t f(X_s)\mathrm{d}A_s\right)(\omega) = \int_0^t f(X_s(\omega))\mathrm{d}A_s(\omega),$$

where the RHS is well defined as a Stieltjes integral.

In contrast, the second integral in (7.2) makes no sense so far. For almost every $\omega \in \Omega$, the path $B_t(\omega)$ is of infinite variation so we can not integrate anything non-

trivial with respect to it. There are two natural ways to try to go forward. First, we can start with integrating trivial things and hope general integrands will be obtained through taking limits. This is done in Section 7.1.1 below. Second, we can try to find desirable properties that such a stochastic integral should satisfy and show that they uniquely determine it. This is done in Section 7.1.2 and is shown to coincide with the previous approach.

7.1 Stochastic integral w.r.t. L^2-bounded martingales

We will now develop a theory of integration w.r.t. martingales in H^2. We present two ways to define the stochastic (Itô) integral. The first one proceeds through a limiting procedure. It is quite intuitive but some details are a bit tedious, even if not difficult. The second construction will be quicker and rather elegant but possibly less intuitive.

7.1.1 Construction of stochastic integral through a limiting procedure

Recall that H^2_0 is the space of continuous martingales M, zero in zero, which are bounded in L^2. It is a Hilbert space with the inner product $< M,N >_{\mathrm{H}^2} = \mathbb{E}[M_\infty N_\infty]$ and its induced norm

$$\|M\|_{\mathrm{H}^2} = \sqrt{\mathbb{E}[M_\infty^2]} = \sqrt{\mathbb{E}[\langle M\rangle_\infty]}.$$

Define $\mathscr{E}$ to be the space of simple bounded process of the form

$$\varphi_t = \sum_{i=0}^{m} \varphi^{(i)} \mathbf{1}_{(t_i,t_{i+1}]}(t), \quad t \geq 0, \tag{7.3}$$

for some $m \in \mathbb{N}$, $0 \leq t_0 < t_1 < \ldots < t_{m+1}$ and where $\varphi^{(i)}$ are bounded $\mathscr{F}_{t_i}$-measurable random variables. Define the stochastic integral $\varphi \bullet M$ of such φ in (7.3) with respect to $M \in \mathrm{H}^2$ via

$$(\varphi \bullet M)_t := \sum_{i=0}^{m} \varphi^{(i)} (M_{t\wedge t_{i+1}} - M_{t\wedge t_i}), \quad t \geq 0. \tag{7.4}$$

If we write $M^i_t := \varphi^{(i)}(M_{t\wedge t_{i+1}} - M_{t\wedge t_i})$ then clearly $M^i \in \mathrm{H}^2$. Further, since for $i \neq j$ the intervals (t_i,t_{i+1}) and (t_j,t_{j+1}) are disjoint, we check easily that $M^i_t M^j_t$ is a martingale and hence $\langle M^i, M^j\rangle_t = 0$ for all $t \geq 0$ a.s.. Bi-linearity of cross-variation then implies that

$$\langle \varphi \bullet M\rangle_t = \sum_{i=0}^{m} \langle M^i\rangle_t = \sum_{i=0}^{m} \left(\varphi^{(i)}\right)^2 \left(\langle M\rangle_{t_{i+1}\wedge t} - \langle M\rangle_{t_i\wedge t}\right) = \int_0^t \varphi_s^2 \mathrm{d}\langle M\rangle_s, \quad t \geq 0. \tag{7.5}$$

For simplicity it is useful to have a compact notation for the integral of φ^2 with respect to $\langle M\rangle$ seen as a stochastic process. More generally, if K is a progressively measurable process and A is of finite variation we denote

$$K \cdot A \text{ the process } (K\cdot A)_t(\omega) := \int_0^t K_s(\omega)\mathrm{d}A_s(\omega),\ t \geq 0,$$

wherever K and A are such that the Stieltjes integral is well defined a.s. for all $t \geq 0$. Property (7.5) can then be written as $\langle \varphi \bullet M\rangle = \varphi^2 \cdot \langle M\rangle$. More generally we see that for $N \in \mathrm{H}^2$ we have

$$\begin{aligned}\langle \varphi \bullet M, N\rangle_t &= \sum_{i=0}^{m} \langle M^i, N\rangle_t = \sum_{i=0}^{m} \varphi^{(i)} \left(\langle M,N\rangle_{t_{i+1}\wedge t} - \langle M,N\rangle_{t_i\wedge t}\right) \\ &= \int_0^t \varphi_s \mathrm{d}\langle M,N\rangle_s = \left(\varphi \cdot \langle M,N\rangle\right)_t,\end{aligned} \tag{7.6}$$

which will be exploited Section 7.1.2 below.

Using (7.5) it is straightforward to see that

Proposition 7.1.1. *Let $M \in \mathrm{H}^2$. The mapping $\varphi \to \varphi \bullet M$ is a linear map from $\mathscr{E}$ to H_0^2. Further*

$$\|\varphi \bullet M\|_{\mathrm{H}^2}^2 = \mathbb{E}\left[\int_0^\infty \varphi_t^2 \mathrm{d}\langle M\rangle_t\right]. \tag{7.7}$$

Proof. Given the derivations above, it remains to check the linearity. Note that we can always add extra time points in the representation (7.3) and equating some consecutive $\varphi^{(i)}$. It follows that given $\varphi, \psi \in \mathscr{E}$ we may assume their representations in (7.3) use the same time points. It is then clear that $(\varphi+\psi)\bullet M = \varphi \bullet M + \psi \bullet M$ as required. □

Remark 7.1.2. In general, when considering discontinuous martingales, it is important to assume that the processes in $\mathscr{E}$ are left-continuous. Indeed, let U, ξ be two independent random variables, U is uniform on $[0,1]$ and $\mathbb{P}(\xi = -1) = \mathbb{P}(\xi = 1) = 1/2$. Then $X_t := \xi \mathbf{1}_{t\geq U}$ is a right-continuous martingale w.r.t. to its natural filtration $(\mathbb{F}_t^X)$. Clearly, X is constant except one jump and hence is of finite variation. It follows that $I_t := \int_0^t X_u \mathrm{d}X_u$ is well defined as a Stieltjes integral. Clearly, $I_1 = \xi^2 = 1$ a.s. while $I_0 = 0$ so that (I_t) can not be a martingale.

Observe that the RHS in (7.7) is an L^2 norm. We define now the appropriate spaces.

Definition 7.1.3. Given $M \in \mathrm{H}^2$ we denote $\mathrm{L}^2(M)$ the space of progressively measurable processes K such that

$$\|K\|_{\mathrm{L}^2(M)}^2 := \mathbb{E}\left[\int_0^\infty K_t^2 \mathrm{d}\langle M\rangle_t\right] < +\infty. \tag{7.8}$$

Note that if we define a measure μ_M on $\mathscr{B}([0,\infty))\otimes\mathscr{F}$ via

$$\mu_M(A) := \mathbb{E}\left[\int_0^\infty \mathbf{1}_A(t,\omega)\mathrm{d}\langle M\rangle_t(\omega)\right] = \int_\Omega\int_0^\infty \mathbf{1}_A(t,\omega)\mathrm{d}\langle M\rangle_t(\omega)\mathbb{P}(\mathrm{d}\omega)$$

and let *Prog* be the σ-algebra on $[0,\infty)\times\Omega$ generated by the progressively measurable processes then $\mathrm{L}^2(M) = \mathrm{L}^2([0,\infty)\times\Omega, Prog, \mu_M)$, where we take K to be defined up to μ_M–indistinguishability. This justifies the usual L^2 notation and, in particular, $\mathrm{L}^2(M)$ is a Hilbert space with the inner product

$$H,K \longrightarrow \mathbb{E}\left[\int_0^\infty H_t K_t \mathrm{d}\langle M\rangle_t\right] = \mathbb{E}\left[(HK\cdot\langle M\rangle)_\infty\right].$$

Naturally $\mathscr{E}\subset\mathrm{L}^2(M)$ and Proposition 7.1.1 established that the map

$$\mathscr{E}\ni\varphi\to\varphi\bullet M\in\mathrm{H}^2_0$$

is a linear isometry. The last remaining piece of the puzzle is

Proposition 7.1.4. *Let $M\in\mathrm{H}^2$. Then $\mathscr{E}$ is a dense vector subspace of $\mathrm{L}^2(M)$.*

Proof. It suffices to show that if $K\in\mathrm{L}^2(M)$ is orthogonal to all $\varphi\in\mathscr{E}$ then $K=0$ (as an element of $\mathrm{L}^2(M)$). So suppose that $<K,\varphi>_{\mathrm{L}^2(M)}=0$ for all $\varphi\in\mathscr{E}$. Let $X = K\cdot\langle M\rangle$ i.e. $X_t=\int_0^t K_u\mathrm{d}\langle M\rangle_u$, which is a well defined Stieltjes integral and in fact

$$\mathbb{E}|X_t| \le \mathbb{E}\left[\int_0^t |K_u|\mathrm{d}\langle M\rangle_u\right] \le \sqrt{\mathbb{E}\left[\int_0^t K_u^2\mathrm{d}\langle M\rangle_u\right]}\sqrt{\mathbb{E}\langle M\rangle_t} < +\infty$$

since $M\in\mathrm{H}^2$ and $K\in\mathrm{L}^2(M)$, and where the second inequality is simply the Cauchy-Schwarz inequality (A.1) for $\mathrm{L}^2(M)$ with $x=K$, $y=1$.

Taking $\varphi=\xi\mathbf{1}_{(s,t]}\in\mathscr{E}$, with $0\le s<t$ and ξ a bounded $\mathscr{F}_s$-measurable r.v., we see that

$$0 = <K,\varphi>_{\mathrm{L}^2(M)} = \mathbb{E}\left[\xi\int_s^t K_u\mathrm{d}\langle M\rangle_u\right] = \mathbb{E}\left[\xi(X_t-X_s)\right].$$

Since this holds for any $\mathscr{F}_s$-measurable bounded ξ we conclude that $\mathbb{E}[(X_t - X_s)|\mathscr{F}_s]=0$ which means that X is a martingale. But X is also continuous and of finite variation and hence (recall Thm VI.3.3 in the lectures) $X\equiv 0$ a.s. This means that $K=0$ $\mathrm{d}\langle M\rangle-a.e.\ a.s.$ i.e. $K=0$ $\mathrm{d}\mu_M$-a.e. and hence $K=0$ in $\mathrm{L}^2(M)$. □

We can now bring all the pieces together and define a stochastic integral starting from simple processes. More precisely, the above shows that any $K\in\mathrm{L}^2(M)$ is a limit of simple processes $\varphi^n\to K$. For each φ^n we can define the stochastic integral $\varphi^n\bullet M$ via (7.4). The isometry property (7.7) then shows that $\varphi^n\bullet M$ converge in H^2 to some element which we denote $K\bullet M$ and which does not depend on the choice of approximating sequence φ^n, see the proof of Proposition A.2.7.

Theorem 7.1.5. *Let $M \in H^2$. The mapping $\varphi \to \varphi \bullet M$ from $\mathscr{E}$ to H_0^2 defined in* (7.4) *has a unique extension to a linear isometry from $L^2(M)$ to H_0^2 which we denote $K \to K \bullet M$.*

Remark 7.1.6. For $K \in L^2(M)$, the martingale $K \bullet M$ is called the *stochastic (Itô) integral of K with respect to M* and is often written as $(K \bullet M)_t = \int_0^t K_u \mathrm{d}M_u$. The isometry property may be then written as

$$\|K \bullet M\|_{H^2}^2 = \mathbb{E}\left[\left(\int_0^\infty K_t \mathrm{d}M_t\right)^2\right] = \mathbb{E}\left[\int_0^\infty K_t^2 \mathrm{d}\langle M\rangle_t\right] = \|K\|_{L^2(M)}^2. \tag{7.9}$$

It also implies that if $K \in L^2(M)$ and a sequence of processes $\varphi^n \in L^2(M)$ (e.g. simple processes $\varphi^n \in \mathscr{E}$) converges to K in $\|\cdot\|_{L^2(M)}$, i.e. $\|\varphi^n - K\|_{L^2(M)} \to 0$ as $n \to \infty$, then $\varphi^n \bullet M$ converge to $K \bullet M$ in $\|\cdot\|_{H^2}$.

Proof. This is a direct consequence of combining Propositions A.2.7, 7.1.1 and 7.1.4. □

7.1.2 Intrinsic characterisation of stochastic integrals using the quadratic co-variation

We come now to a second definition or characterisation of the Itô integral. We will show that it coincides with the definition developed in the previous section.

Theorem 7.1.7. *Let $M \in H^2$. For any $K \in L^2(M)$ there exists a unique element in H_0^2, denoted $K \bullet M$, such that*

$$\langle K \bullet M, N\rangle = K \cdot \langle M, N\rangle, \quad \forall N \in H^2. \tag{7.10}$$

Furthermore, $\|K \bullet M\|_{H^2} = \|K\|_{L^2(M)}$ and the map

$$L^2(M) \ni K \longrightarrow K \bullet M \in H_0^2$$

is a linear isometry.

Proof. We first show uniqueness. Suppose that there are two such elements, X and X'. Then (7.10) gives that

$$\langle X, N\rangle - \langle X', N\rangle = \langle X - X', N\rangle \equiv 0, \quad \forall N \in H^2.$$

Taking $N = X - X'$ we conclude, by Corollary 6.4.9, that $X = X'$.

It remains to argue existence. Consider the mapping

$$H_0^2 \ni N \longrightarrow \mathbb{E}[(K \cdot \langle M, N\rangle)_\infty] \in \mathbb{R}.$$

By linearity of expectation, of Stieltjes integral and of $\langle M,\cdot\rangle$, see Proposition 6.4.11 (ii), the above mapping is linear. Further, by Kunita-Watanabe inequality in Theorem 6.4.15

$$|\mathbb{E}[(K\cdot\langle M,N\rangle)_\infty]| \le \|N\|_{\mathrm{H}^2}\|K\|_{\mathrm{L}^2(M)}$$

and hence it is also continuous. It follows, by the representation theorem for continuous linear functionals on a Hilbert space, that there exists (a unique) element $K\bullet M\in \mathrm{H}^2_0$ such that

$$\mathbb{E}[(K\bullet M)_\infty N_\infty] = \mathbb{E}[(K\cdot\langle M,N\rangle)_\infty], \quad \forall N\in \mathrm{H}^2_0. \tag{7.11}$$

Let τ now be a stopping time. Then, using uniform integrability of $K\bullet M$, we obtain

$$\begin{aligned}\mathbb{E}[(K\bullet M)_\tau N_\tau] &= \mathbb{E}[\mathbb{E}[(K\bullet M)_\infty|\mathscr{F}_\tau]N_\tau] = \mathbb{E}[(K\bullet M)_\infty N_\tau]\\ &= \mathbb{E}[(K\bullet M)_\infty N^\tau_\infty] \overset{\text{by (7.11)}}{=} \mathbb{E}[(K\cdot\langle M,N^\tau\rangle)_\infty]\\ &= \mathbb{E}[(K\cdot\langle M,N\rangle^\tau)_\infty] = \mathbb{E}[(K\cdot\langle M,N\rangle)_\tau].\end{aligned}$$

It follows by Theorem 5.4.8 that $(K\bullet M)N - K\cdot\langle M,N\rangle$ is a martingale and hence $\langle K\bullet M,N\rangle = K\cdot\langle M,N\rangle$, by uniqueness of the co-variation process in Proposition 6.4.11 (i). This establishes (7.10) for $N\in\mathrm{H}^2_0$. For $N\in\mathrm{H}^2$ we have $\tilde N = N - N_0\in\mathrm{H}^2_0$ and it follows that

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

as required. The right-hand side in (7.11) is linear in K and hence so is $K\bullet M$. It remains to show the isometry property:

$$\begin{aligned}\|K\bullet M\|^2_{\mathrm{H}^2} &= \mathbb{E}\left[(K\bullet M)^2_\infty\right] = \mathbb{E}[(K\bullet M)_\infty(K\bullet M)_\infty] \overset{\text{by (7.11)}}{=} \mathbb{E}[(K\cdot\langle K\bullet M,M\rangle)_\infty]\\ &= \mathbb{E}\left[\left(K^2\cdot\langle M,M\rangle\right)_\infty\right] = \|K\|^2_{\mathrm{L}^2(M)},\end{aligned}$$

where the last equalities follow by associativity of Stieljtes integral (6.3) and (7.10) which in particular gives $\langle K\bullet M,M\rangle = K\cdot\langle M,M\rangle$. □

Note that it follows instantly from the fact that the initial value M_0 does not influence the quadratic (co)variation of a martingale M, that $K\bullet M = K\bullet(M-M_0)$, as one would expect. We also have:

Corollary 7.1.8. *The two definitions of stochastic integral in Theorems 7.1.5 and 7.1.7 coincide for $M\in\mathrm{H}^2$.*

Proof. We saw that the stochastic integral in (7.4) satisfies (7.6) which is precisely (7.10) for $K\in\mathscr{E}$. It follows that the two definitions of the stochastic integral coincide for simple integrands in $K\in\mathscr{E}$ and then the uniqueness of extension of the isometry in Theorem 7.1.5, and the uniqueness in Theorem 7.1.7, show that the two notions coincide for all $K\in\mathrm{L}^2(M)$. □

Defining stochastic integral as in Theorem 7.1.7 in certainly elegant – it is simple to state and pins down the precise property which characterises the stochastic

integral. Moreover, it is very useful. What (7.10) does is to characterise *stochastic integration* on the LHS through the usual well-understood *Stieltjes integration* on the RHS. This has immediate consequences.

Proposition 7.1.9 (Associativity of stochastic integration). *Let $M \in \mathrm{H}^2$, $K \in \mathrm{L}^2(M)$ and $F \in \mathrm{L}^2(K \bullet M)$. Then $FK \in \mathrm{L}^2(M)$ and*

$$FK \bullet M = F \bullet (K \bullet M). \tag{7.12}$$

Proof. We have

$$\langle K \bullet M, K \bullet M \rangle = K \cdot \langle M, K \bullet M \rangle = K^2 \cdot \langle M, M \rangle$$

so that $\int_0^\infty F_s^2 \mathrm{d}\langle K \bullet M \rangle_s = \int_0^\infty F_s^2 K_s^2 \mathrm{d}\langle M \rangle_s$, which shows that $FK \in \mathrm{L}^2(M)$. Then, for $N \in \mathrm{H}^2$, we have

$$\begin{aligned} \langle (FK) \bullet M, N \rangle &\overset{\text{Thm } 7.1.7}{=} FK \cdot \langle M, N \rangle \overset{\text{by } (6.3)}{=} F \cdot (K \cdot \langle M, N \rangle) \\ &\overset{\text{Thm } 7.1.7}{=} F \cdot \langle K \bullet M, N \rangle \\ &\overset{\text{Thm } 7.1.7}{=} \langle F \bullet (K \bullet M), N \rangle, \end{aligned}$$

which gives the desired result by the uniqueness property in Theorem 7.1.7. □

Recall that if $M \in \mathrm{H}^2$ and τ is a stopping time then $M^\tau = (M_{t \wedge \tau} : t \geq 0)$ denotes the stopped process which is itself a martingale and clearly $M^\tau \in \mathrm{H}^2$. For any $N \in \mathrm{H}^2$ we have

$$\langle M^\tau, N \rangle = \langle M, N \rangle^\tau = \mathbf{1}_{[0,\tau]} \cdot \langle M, N \rangle = \langle \mathbf{1}_{[0,\tau]} \bullet M, N \rangle,$$

where we first used property (iii) in Proposition 6.4.11, then properties of Stieltjes integral and finally (7.10). It then follows by uniqueness in Theorem 7.1.7 that $\mathbf{1}_{[0,\tau]} \bullet M = M^\tau$. In fact a more general property holds true.

Proposition 7.1.10 (Stopped stochastic integrals). *Let $M \in \mathrm{H}^2$, $K \in \mathrm{L}^2(M)$ and τ a stopping time. Then*

$$(K \bullet M)^\tau = K \bullet M^\tau = K \mathbf{1}_{[0,\tau]} \bullet M.$$

Proof. We already argued above that the result holds for $K \equiv 1$. This gives

$$K \bullet M^\tau = K \bullet \left(\mathbf{1}_{[0,\tau]} \bullet M \right) = K \mathbf{1}_{[0,\tau]} \bullet M,$$

where we used (7.12). Further, applying the same result to the martingale $K \bullet M$ we also obtain

$$(K \bullet M)^\tau = \mathbf{1}_{[0,\tau]} \bullet (K \bullet M) = \mathbf{1}_{[0,\tau]} K \bullet M$$

which gives the desired equalities. □

7.2 Extensions: stochastic integration with respect to continuous semimartingales

Definition 7.2.1. For a continuous local martingale M, we denote $\mathrm{L}^2_{\mathrm{loc}}(M)$ the space of progressively measurable processes K such that

$$\forall t \geq 0 \quad \int_0^t K_s^2 \mathrm{d}\langle M\rangle_s < +\infty \; a.s.$$

Theorem 7.2.2. *Let M be a continuous local martingale. For any $K \in \mathrm{L}^2_{loc}(M)$ there exists a unique continuous local martingale, zero in zero, denoted $K \bullet M$ and called the stochastic integral of K with respect to M, such that for any continuous local martingale N*

$$\langle K \bullet M, N\rangle = K \cdot \langle M, N\rangle. \tag{7.13}$$

Furthermore, if $M \in \mathrm{H}^2$ and $K \in \mathrm{L}^2(M)$ then this definition coincides with the previously defined stochastic integral $K \bullet M$.

Proof. To see that this definition extends the previous one, observe that (7.13) and associativity of Stieltjes integral imply that $\langle K \bullet M, K \bullet M\rangle = K^2 \cdot \langle M, M\rangle$ so that if $M \in \mathrm{H}^2$ and $K \in \mathrm{L}^2(M)$ then, by Theorem 6.4.8, $K \bullet M \in \mathrm{H}^2_0$ and hence it coincides with the integral defined in Theorem 7.1.7.

Let

$$\tau_n = \inf\left\{ t \geq 0 : \int_0^t (1 + K_s^2)\mathrm{d}\langle M\rangle_s \geq n \right\},$$

which are stopping times (recall Exercise 4.2.2) and $\tau_n \nearrow \infty$ a.s. with $n \to \infty$. Note that $\langle M^{\tau_n}\rangle_\infty = \langle M\rangle_{\tau_n} \leq n$ so that $M^{\tau_n} \in \mathrm{H}^2$ by Theorem 6.4.8. Further, $(K^2 \cdot \langle M^{\tau_n}\rangle)_\infty \leq n$ so that $K \in \mathrm{L}^2(M^{\tau_n})$. Finally, by Proposition 7.1.10, for any m, n

$$(K \bullet M^{\tau_n})_{t\wedge\rho} = (K \bullet M^{\tau_m})_{t\wedge\rho}, \quad \text{where } \rho = \tau_n \wedge \tau_m.$$

It follows that we may define a unique continuous process $K \bullet M$ by setting it to be equal to $K \bullet M^{\tau_n}$ on $[0, \tau_n]$. By definition, $K \bullet M$ is a continuous local martingale and (τ_n) is its reducing sequence.

Let N be a continuous local martingale. Without loss of generality let $N_0 = 0$. We consider its reducing sequence

$$\tilde{\tau}_n = \inf\{t \geq 0 : |N_t| \geq n\} \quad \text{and} \quad \rho_n := \tau_n \wedge \tilde{\tau}_n.$$

Then $N^{\rho_n} \in \mathrm{H}^2_0$ and hence

$$\begin{aligned}\langle K \bullet M, N\rangle^{\rho_n} &= \langle (K \bullet M)^{\rho_n}, N^{\rho_n}\rangle \overset{\tau_n \geq \rho_n}{=} \langle (K \bullet M^{\tau_n})^{\rho_n}, N^{\rho_n}\rangle \overset{(6.8)}{=} \langle K \bullet M^{\tau_n}, N^{\rho_n}\rangle \\ &\overset{\text{Thm } 7.1.7}{=} K \cdot \langle M^{\tau_n}, N^{\rho_n}\rangle \overset{(6.8)}{=} K \cdot \langle M, N\rangle^{\rho_n} = (K \cdot \langle M, N\rangle)^{\rho_n},\end{aligned}$$

so that $\langle K \bullet M, N\rangle = K \cdot \langle M, N\rangle$ as required. Uniqueness of $K \bullet M$ follows as in Theorem 7.1.7. □

Remark 7.2.3. It follows from the proofs above and the characterisation in (7.13), that the associativity of stochastic integral in (7.12) and stability under stopping in Proposition 7.1.10 instantly extend to a continuous local martingale M and $K \in \mathrm{L}^2_{\mathrm{loc}}(M)$, $F \in \mathrm{L}^2_{\mathrm{loc}}(K \bullet M)$.

Recall that a continuous semimartingale X can be represented as $X = X_0 + M + A$, where X_0 is an $\mathscr{F}_0$–measurable random variable, M is a continuous local martingale and A is of finite variation. It follows that such a decomposition is unique since if $X = X_0 + M' + A'$ then $M' - M = A - A'$ is both a continuous local martingale and a process of finite variation and hence is indistinguishable from zero by Corollary 6.4.9.

Naturally, we can define an integral with respect to a continuous semimartingale $X = X_0 + M + A$ simply by linearity as sum of integrals w.r.t. M and w.r.t. A. We just have to take integrands K which will be both in $\mathrm{L}^2_{\mathrm{loc}}(M)$ and will satisfy $\int_0^t |K_s||\mathrm{d}A_s| < \infty$ for all $t \geq 0$. A subset of such integrands, particularly convenient since it does not depend on M or A, is given by

Definition 7.2.4. We say that a process K is locally bounded if it is progressively measurable and

$$a.s.\ \forall t \geq 0 \quad \sup_{u \leq t} |K_u| < +\infty.$$

Note that a continuous process is necessarily locally bounded since a continuous function achieves its maximum and minimum on a compact interval.

Definition 7.2.5. Let $X = X_0 + M + A$ be a continuous semimartingale and K a locally bounded process. The **(Itô) stochastic integral** of K with respect to X is the continuous semimartingale $K \bullet X$ defined by

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

and we often write simply

$$(K \bullet X)_t = \int_0^t K_s \mathrm{d}X_s = \int_0^t K_s \mathrm{d}M_s + \int_0^t K_s \mathrm{d}A_s.$$

Naturally this integral inherits the common properties of the stochastic and Stieltjes integrals.

Proposition 7.2.6. *In the statements below X is a continuous semimartingale and K is a locally bounded process.*

(i) The map which to X and K associates their (stochastic) integral $K \bullet X$ is bilinear.

*(ii) If F is another locally bounded process then $F \bullet (K \bullet X) = FK \bullet X$ (*associativity*)*

(iii) If τ is a stopping time then

$$(K \bullet X)^\tau = K\mathbf{1}_{[0,\tau]} \bullet X = K \bullet X^\tau.$$

(iv) If X is a local martingale (resp. a process of finite variation) then $K \bullet M$ is a local martingale (resp. a process of finite variation).

(v) For $\varphi \in \mathscr{E}$ given in (7.3)

$$(\varphi \bullet X)_t = \sum_{i=1}^{m} \varphi^{(i)}(X_{t_{i+1}\wedge t} - X_{t_i \wedge t}).$$

Proof. This is a straightforward consequence of the definition, Proposition 6.2.4 and Remark 7.2.3. □

Remark 7.2.7 (Fundamental remark). It follows from the properties of the quadratic co-variation, recall Remark 6.4.12, and the characterisation in (7.13), that the stochastic integral in Definition 7.2.5 is invariant under an equivalent change of measure. That is if $\mathbb{Q}$ is a probability measure with $\mathbb{Q}(\Gamma) = 0$ iff $\mathbb{P}(\Gamma) = 0$ for all $\Gamma \in \mathscr{F}$ then the stochastic integrals under $\mathbb{P}$ and $\mathbb{Q}$ coincide.

This shows that our newly defined integral has all the natural properties, for a given integrand and integrator. To complete the theory it remains to understand how it behaves under limits. Of course we saw already in Theorem 7.1.5 that under suitable integrability conditions (appropriate) L^2 convergence of integrands and stochastic integrals were equivalent. In general we have a weaker result.

Proposition 7.2.8 (Stochastic dominated convergence theorem). *Let X be a continuous semimartingale and K^n a sequence of locally bounded processes $K^n_t \to 0$ as $n \to \infty$ a.s. for all t and further $|K^n_t| \le K_t$ for all n and some locally bounded process K. Then $K^n \bullet X$ converge to zero in probability and more precisely*

$$\forall t \ge 0 \quad \sup_{s \le t} \left| \int_0^s K^n_u \mathrm{d}X_u \right| \longrightarrow 0 \text{ in probability as } n \to \infty.$$

Proof. We can treat the finite variation part $X_0 + A$ and the local martingale part M separately. For the former, we first note that

$$\begin{aligned} \left| \int_0^t K^n_u \mathrm{d}A_u \right| &= \left| \int_0^t K^n_u \mathrm{d}A^+_u - \int_0^t K^n_u \mathrm{d}A^-_u \right| \\ &\le \int_0^t |K^n_u| \mathrm{d}A^+_u + \int_0^t |K^n_u| \mathrm{d}A^-_u = \int_0^t |K^n_u| |\mathrm{d}A_u|. \end{aligned}$$

The a.s. pointwise convergence of K^n to 0, together with the bound $|K^n| \le K$, allow us to apply the (usual) dominated convergence theorem to conclude that, for any $t > 0$, $\int_0^t |K^n_u||\mathrm{d}A_u|$ converges to 0 a.s. In fact, as $\int_0^t |K^n_u||\mathrm{d}A_u|$ is non-decreasing in t, the convergence is uniform on any compact interval.

For the continuous local martingale part M, let (τ_m) be a reducing sequence such that $M^{\tau_m} \in \mathrm{H}^2_0$ and $K \in \mathrm{L}^2(M^{\tau_m})$. Then, by Itô's isometry (7.9),

$$\|K^n \bullet M^{\tau_m}\|^2_{\mathrm{H}^2} = \mathbb{E}\left[\left(\int_0^{\tau_m} K^n_t \mathrm{d}M_t\right)^2\right] = \mathbb{E}\left[\int_0^\infty (K^n_t)^2 \mathbf{1}_{[0,\tau_m]}(t)\mathrm{d}\langle M\rangle_t\right] = \|K^n\|^2_{\mathrm{L}^2(M^{\tau_m})}.$$

It then follows from the dominated convergence theorem that $\|K^n\|^2_{\mathrm{L}^2(M^{\tau_m})} \to 0$ as $n \to \infty$, and hence $\|K^n \bullet M^{\tau_m}\|^2_{\mathrm{H}^2} \to 0$ as $n \to \infty$. For a fixed $t \geq 0$, and any given $\varepsilon > 0$, we may take m large enough so that $\mathbb{P}(\tau_m \leq t) \leq \varepsilon/2$. We then have

$$\begin{aligned}\mathbb{P}\left(\sup_{s\leq t}|(K^n \bullet M)_s| > \varepsilon\right) &\leq \mathbb{P}\left(\sup_{s\leq t\wedge\tau_m}|(K^n \bullet M)_s| > \varepsilon\right) + \varepsilon/2 \leq \varepsilon \\ &\leq \frac{1}{\varepsilon}\|K^n \bullet M^{\tau_m}\|^2_{\mathrm{H}^2} + \varepsilon/2 \leq \varepsilon,\end{aligned}$$

for n large enough. □

Using the above we can finally understand in what sense a stochastic integral can be seen as a limit of simple integrals in (7.4).

Proposition 7.2.9. *Let X be a continuous semimartingale and K a left-continuous locally bounded process. If Δ^n is a sequence of subdivisions on $[0,t]$ with mesh converging to zero then*

$$\sum_{t_i\in\Delta^n} K_{t_i}(X_{t_{i+1}} - X_{t_i}) \longrightarrow \int_0^t K_s \mathrm{d}X_s \text{ in probability as } n \to \infty.$$

Proof. Assume first that K is a bounded process. Let $K^n_t = K_{t_i}$ if $t \in (t_i, t_{i+1}]$ with $t_i \in \Delta^n$. Then $K^n_t \to K_t$ as $n \to \infty$ by left continuity and $|K^n - K| \leq c$ by some constant c. Now K^n is a simple integrand and

$$(K^n \bullet M)_t = \sum_{t_i\in\Delta^n} K_{t_i}(X_{t_{i+1}} - X_{t_i}).$$

Applying Proposition 7.2.8, $\big((K_n - K) \bullet M\big)_t$ converges to 0 in probability with $n \to \infty$, as required.

In the general case, we consider $\tau_m = \inf\{t \geq 0 : |K_t| \geq m\}$ and note that, by left-continuity, the stopped process is bounded: $|K^{\tau_m}| \leq m$. In particular the above reasoning shows the result holds for K^{τ_m}. We conclude since K is locally bounded and hence $\tau_m \nearrow \infty$ as $m \to \infty$. □

The sum $\sum_{t_i\in\Delta^n} K_{t_i}(X_{t_{i+1}} - X_{t_i})$ is the usual Riemann-sum approximation which uses the left end-point K_{t_i}. This ties in with our definitions in (7.3)–(7.4) and is what allows to preserve the martingale property, see (iv) in Proposition 7.2.6. However other choices are possible and lead to different notions of stochastic integrals. These will not preserve the martingale property but may have their own merits. In particular it is possible to preserve the usual simple integration by parts formula from the

Newton-Leibnitz calculus. In contrast, Itô stochastic integral has its own integration by parts which features the so-called Itô correction. We explore it in the next section.

7.3 Itô's formula and its applications

We state now the fundamental integration by parts (or chain rule) formula for stochastic integrals.

Proposition 7.3.1 (Integration by parts). *If X and Y are two continuous semimartingales then*

$$X_tY_t = X_0Y_0 + \int_0^t X_s \mathrm{d}Y_s + \int_0^t Y_s \mathrm{d}X_s + \langle X,Y\rangle_t, \quad t \geq 0 \quad a.s. \tag{7.14}$$

Proof. Fix t and let Δ^n be a sequence of subdivisions on $[0,t]$ with mesh converging to zero. Note that

$$X_tY_t - X_sY_s = X_s(Y_t - Y_s) + Y_s(X_t - X_s) + (X_t - X_s)(Y_t - Y_s)$$

so for any n

$$\begin{aligned} X_tY_t - X_0Y_0 = &\sum_{t_i \in \Delta^n} \Big(X_{t_i}(Y_{t_{i+1}} - Y_{t_i}) + Y_{t_i}(X_{t_{i+1}} - X_{t_i}) + (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}) \Big) \\ &\longrightarrow (X \bullet Y)_t + (Y \bullet X)_t + \langle X,Y\rangle_t \quad \text{as } n \to \infty \end{aligned}$$

by Proposition 6.4.11 and Proposition 7.2.9. □

Theorem 7.3.2 (Itô's formula). *Let $X^1,\dots,X^d$ be continuous semimartingales and $F : \mathbb{R}^d \to \mathbb{R}$ a C^2 function. Then $(F(X_t^1,\dots,X_t^d) : t \geq 0)$ is a continuous semimartingale and a.s. for all $t \geq 0$*

$$\begin{aligned} F(X_t^1,\dots,X_t^d) = &F(X_0^1,\dots,X_0^d) + \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x^i}(X_s^1,\dots,X_s^d)\mathrm{d}X_s^i \\ &+ \frac{1}{2}\sum_{1\leq i,j\leq d} \int_0^t \frac{\partial^2 F}{\partial x^i \partial x^j}(X_s^1,\dots,X_s^d)\mathrm{d}\langle X^i, X^j\rangle_s. \end{aligned} \tag{7.15}$$

In particular, for $d = 1$, we have

$$F(X_t) = F(X_0) + \int_0^t F'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^t F''(X_s)\mathrm{d}\langle X\rangle_s.$$

Proof. Let $X = X_0 + M + A$ be the semimartingale decomposition of X and denote V the total variation process of A. Let

$$\tau_r = \inf\{t \geq 0 : |X_t| + V_t + \langle M\rangle_t > r\}.$$

Then $(\tau_r)_{r\geq 0}$ is a family of stopping times with $\tau_r \nearrow \infty$. It is sufficient to prove (7.15) up to time τ_r. Let $\mathscr{A} \subset C^2(\mathbb{R}^d,\mathbb{R})$ be the subset of functions F for which the formula holds. We will argue below that $\mathscr{A}$ satisfies the following four properties:

(i) $\mathscr{A}$ contains the functions $F(x_1,\dots,x_d) \equiv 1$ and $F(x_1,\dots,x_d) = x_i$ for $i \leq d$,
(ii) $\mathscr{A}$ is a vector space,
(iii) if $F,G \in \mathscr{A}$, then $FG \in \mathscr{A}$,
(iv) if $F_n \in \mathscr{A}$ and $F_n \to F$ in $C^2(B_r,\mathbb{R})$ for all $r>0$ then $F \in \mathscr{A}$, where $F_n \to F$ in $C^2(B_r,\mathbb{R})$ means that

$$\Delta_{n,r} := \sup_{\mathbf{x}\in B_r} \max\left\{ |F_n(\mathbf{x}) - F(\mathbf{x})|, \sum_{i=1}^d \left| \frac{\partial F_n(\mathbf{x})}{\partial x^i} - \frac{\partial F(\mathbf{x})}{\partial x^i} \right|, \sum_{i=1}^d \sum_{j=1}^d \left| \frac{\partial^2 F_n(\mathbf{x})}{\partial x^i \partial x^j} - \frac{\partial^2 F(\mathbf{x})}{\partial x^i \partial x^j} \right| \right\}$$

converges to zero as $n \to \infty$ with $B_r = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| \leq r\}$.

Note that (i)-(iii) imply that $\mathscr{A}$ contains all polynomials. By Weierstrass's approximation theorem, the polynomials are dense in $C^2(B_r,\mathbb{R})$ and so (iv) then implies $\mathscr{A}$ contains $C^2(B_r,\mathbb{R})$ for all r and hence $\mathscr{A} = C^2(\mathbb{R}^d,\mathbb{R})$.

Properties (i) and (ii) above are immediate. Let us argue (iii). Suppose $F,G \in \mathscr{A}$ and set $F_t = F(X_t)$, $G_t = G(X_t)$. Since the formula holds for F and G, (F_t) and (G_t) are continuous semimartingales. Integration by parts (Proposition 7.3.1) yields

$$F_t G_t - F_0 G_0 = \int_0^t F_s \mathrm{d}G_s + \int_0^t G_s \mathrm{d}F_s + \langle F,G\rangle_t. \tag{7.16}$$

By associativity of stochastic integration, see Proposition 7.2.6, we have

$$\int_0^t F_s \mathrm{d}G_s = \sum_{i=1}^d \int_0^t F(X_s) \frac{\partial G_s}{\partial x^i} \mathrm{d}X_s^i + \frac{1}{2} \sum_{1\leq i,j\leq d} \int_0^t F(X_s) \frac{\partial^2 G_s}{\partial x^i \partial x^j} \mathrm{d}\langle X^i, X^j\rangle_s,$$

with a similar expression for $\int_0^t G_s \mathrm{d}F_s$. By Theorem 7.2.2 we have

$$\langle F,G\rangle_t = \sum_{i=1}^d \sum_{j=1}^d \int_0^t \frac{\partial F_s}{\partial x^i} \frac{\partial G_s}{\partial x^i} \mathrm{d}\langle X^i, X^j\rangle_s.$$

Substituting these into (7.16), we obtain Itô's formula for FG.

It remains to argue (iv). Let $F_n \in \mathscr{A}$ such that $F_n \to F$ in $C^2(B_r,\mathbb{R})$. Then

$$\begin{aligned}\sum_{i=1}^d \int_0^{t\wedge\tau_r} |\frac{\partial F_n}{\partial x^i}(\mathbf{X}_s) - \frac{\partial F}{\partial x^i}(\mathbf{X}_s)| \mathrm{d}V_s + \frac{1}{2} \sum_{1\leq i,j\leq d} \int_0^t \left| \frac{\partial^2 F_n}{\partial x^i \partial x^j}(\mathbf{X}_s) - \frac{\partial^2 F}{\partial x^i \partial x^j}(\mathbf{X}_s) \right| \mathrm{d}\langle M^i, M^j\rangle_s \\ \leq \Delta_{n,r}(V_{t\wedge\tau_r} + \frac{1}{2}\langle M\rangle_{t\wedge\tau_r}) \leq \Delta_{n,r} r \to 0 \quad \text{as } n\to\infty,\end{aligned}$$

where $\mathbf{X}_s = (X_s^1,\dots,X_s^n)$. It follows that

$$\sum_{i=1}^{d}\int_0^{t\wedge\tau_r}\frac{\partial F_n}{\partial x^i}(\mathbf{X}_s)\mathrm{d}A_s+\frac{1}{2}\sum_{1\le i,j\le d}\int_0^t\frac{\partial^2 F_n}{\partial x^i\partial x^j}(\mathbf{X}_s)\mathrm{d}\langle M^i,M^j\rangle_s$$
$$\longrightarrow\sum_{i=1}^{d}\int_0^{t\wedge\tau_r}\frac{\partial F}{\partial x^i}(\mathbf{X}_s)\mathrm{d}A_s+\frac{1}{2}\sum_{1\le i,j\le d}\int_0^t\frac{\partial^2 F}{\partial x^i\partial x^j}(\mathbf{X}_s)\mathrm{d}\langle M^i,M^j\rangle_s\ a.s.$$

Moreover, since F_n and F are uniformly bounded on B_r, by Proposition 7.2.8,

$$\sum_{i=1}^{d}\int_0^{t\wedge T_t}\frac{\partial F_n}{\partial x^i}(\mathbf{X}_s)\mathrm{d}M_s^i\longrightarrow\sum_{i=1}^{d}\int_0^{t\wedge\tau_r}\frac{\partial F}{\partial x^i}(\mathbf{X}_s)\mathrm{d}M_s^i\quad\text{ in probability as } n\to\infty.$$

Putting the above together and taking the limit in $n\to\infty$, we obtain

$$F(\mathbf{X}_{t\wedge\tau_r})=F(\mathbf{X}_0)+\sum_{i=1}^{d}\int_0^{t\wedge\tau_r}\frac{\partial F}{\partial x^i}(\mathbf{X}_s)\mathrm{d}X_s^i+\frac{1}{2}\sum_{1\le i,j\le d}\int_0^{t\wedge\tau_r}\frac{\partial^2 F}{\partial x^i\partial x^j}(\mathbf{X}_s)\mathrm{d}\langle X^i,X^j\rangle_s.$$

□

The last statement for $d=1$ is the canonical form of the Itô's formula and it highlights its most striking feature – the presence of the last term which arises due to non-trivial quadratic variation. Note that the formula holds for any semimartingale and in particular for X of bounded variation – in this case the last term is zero and it reduces to the "fundamental theorem of calculus."

The multidimensional statement in (7.15) is not given just for generality – it is in fact extremely useful. In the first instance we shall apply it for the case $d=2$ with $X^1=M$ a continuous local martingale and $X^2=A$ a process of bounded variation. We know that $\langle M,A\rangle\equiv 0$ so (7.15) then reduces to

$$F(M_t,A_t)=F(M_0,A_0)+\int_0^t\frac{\partial F}{\partial m}(M_s,A_s)\mathrm{d}M_s+\int_0^t\frac{\partial F}{\partial a}(M_s,A_s)\mathrm{d}A_s+\frac{1}{2}\int_0^t\frac{\partial^2 F}{\partial m^2}(M_s,A_s)\mathrm{d}\langle M\rangle_s.$$

Typical applications include choice of $A_t=t$, $A_t=\langle M\rangle_t$ or $A_t=\sup_{s\le t}M_s$. In particular, taking $F(x,y)=\exp(\lambda x-\frac{\lambda^2}{2}y)$ we derive the following useful result:

Proposition 7.3.3. *Let M be a continuous local martingale and $\lambda\in\mathbb{R}$. Then*

$$\mathscr{E}^{\lambda}(M)_t:=\exp\left(\lambda M_t-\frac{\lambda^2}{2}\langle M\rangle_t\right),\quad t\ge 0,\tag{7.17}$$

is a continuous local martingale. In fact the same holds true for any $\lambda\in\mathbb{C}$ with the real and imaginary parts being local martingales.

Proof. Note that $\langle M\rangle$ is of finite variation (in fact is non-decreasing) and in particular $\langle M,\langle M\rangle\rangle\equiv 0$ and also $\langle\langle M\rangle,\langle M\rangle\rangle\equiv 0$. Let $F(x,y)=\exp\left(\lambda x-\frac{\lambda^2}{2}y\right)$. $F\in C^2(\mathbb{R}^2,\mathbb{C})$ so we may apply Itô's formula to $\mathscr{E}^{\lambda}(M)_t=F(M_t,\langle M\rangle_t)$. Computing the partial derivatives and simplifying gives:

$$\mathscr{E}^{\lambda}(M)_t = \mathscr{E}^{\lambda}(M)_0 + \int_0^t \frac{\partial}{\partial x} F^{\lambda}(M_s, \langle M\rangle_s)\mathrm{d}M_s,$$

which shows that $\mathscr{E}^{\lambda}(M)$ is a continuous local martingale. □
Note that we have $\frac{\partial}{\partial x}F(x,y) = \lambda F(x,y)$ so that

$$\mathscr{E}^{\lambda}(M)_t = \mathscr{E}^{\lambda}(M)_0 + \lambda \int_0^t \mathscr{E}^{\lambda}(M)_s \mathrm{d}M_s$$

which we can write in a simplified "differential form" as

$$\mathrm{d}\mathscr{E}^{\lambda}(M)_t = \lambda \mathscr{E}^{\lambda}(M)_t \mathrm{d}M_t$$

which shows $\mathscr{E}^{\lambda}(M)$ solves the stochastic exponential differential equation driven by M: $\mathrm{d}Y_t = \lambda Y_t \mathrm{d}M_t$. We already saw that exponential martingales can be very useful, e.g. to characterise distributions of hitting times (via their Laplace transforms). Here we see another beautiful application.

Theorem 7.3.4 (Lévy's characterisation of Brownian motion). *Let M be a continuous local martingale starting in zero. Then M is a standard Brownian motion if and only if $\langle M\rangle_t = t$ a.s. for all $t \geq 0$.*

Proof. We know that the quadratic variation of a Brownian motion B is given by $\langle B\rangle_t = t$ so we just have to show the reverse statement. Suppose M is a continuous local martingale starting in zero with $\langle M\rangle_t = t$ a.s. for all $t \geq 0$. Then, by Proposition 7.3.3,

$$\exp\left(\mathrm{i}\xi M_t + \frac{\xi^2}{2}t\right), \quad t \geq 0$$

is a local martingale for any $\xi \in \mathbb{R}$ and, since it is bounded, it is a martingale. Let $0 \leq s < t$. We have

$$\mathbb{E}\left[\exp\left(\mathrm{i}\xi M_t + \frac{\xi^2}{2}t\right)\middle|\mathscr{F}_s\right] = \exp\left(\mathrm{i}\xi M_s + \frac{\xi^2}{2}s\right)$$

which we can rewrite as

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi(M_t - M_s)}\middle|\mathscr{F}_s\right] = \mathrm{e}^{-\frac{\xi^2}{2}(t-s)}.$$

Taking expectations is follows that $M_t - M_s$ has a centred Gaussian distribution with variance $(t-s)$. Using this, together with the definition of the conditional expectation, we have

$$\mathbb{E}\left[\mathbf{1}_A \mathrm{e}^{\mathrm{i}\xi(M_t - M_s)}\right] = \mathbb{P}(A)\mathrm{e}^{-\frac{\xi^2}{2}(t-s)} = \mathbb{P}(A)\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi(M_t - M_s)}\right], \quad A \in \mathscr{F}_s, \xi \in \mathbb{R}.$$

Fix $A \in \mathscr{F}_s$ with $\mathbb{P}(A) > 0$ and let $\mathbb{P}_A := \mathbb{P}(\cdot \cap A)/\mathbb{P}(A)$, which a probability measure on $\mathscr{F}$. The above shows that $M_t - M_s$ has the same distribution under $\mathbb{P}$ and $\mathbb{P}_A$ so that

$$\forall f \geq 0 \quad \mathbb{E}_A[f(M_t - M_s)] = \mathbb{E}[f(M_t - M_s)]$$
$$\text{i.e.} \quad \mathbb{E}[\mathbf{1}_A f(M_t - M_s)] = \mathbb{P}(A)\mathbb{E}[f(M_t - M_s)]$$

which now holds for any $A \in \mathscr{F}_s$ and shows that $(M_t - M_s)$ is independent of $\mathscr{F}_s$. It follows that M is an $(\mathscr{F}_t)$–Brownian motion. □

The above result shows that the quadratic variation captures, in a way, all the relevant information. We know that a continuous martingale has to wiggle a lot, the martingale property dictates that is has to be equally likely to wiggle up and down, and the only remaining information is the precise (instantaneous) speed at which the process wiggles which is captured by the (change of) quadratic variation.

The above has a natural intriguing consequence: if I take an arbitrary continuous local martingale M, can I change the time so that the process becomes a Brownian motion? That is to say, can I keep adjusting, accelerating or decelerating, the flow of time, in such a way that the relative "speed at which M wiggles" is constant and hence M looks like a Brownian motion? The answer is Yes!

Theorem 7.3.5 (Dambis, Dubins and Schwarz). *Let M be an $((\mathscr{F}_t),\mathbb{P})$–continuous local martingale, $M_0 = 0$ and $\langle M\rangle_\infty = \infty$ a.s. Let $\tau_s := \inf\{t \geq 0 : \langle M\rangle_t > s\}$. Then the process B, $B_s := M_{\tau_s}$, is an $((\mathscr{F}_{\tau_s}),\mathbb{P})$–Brownian motion and $M_t = B_{\langle M\rangle_t}$, $\forall t \geq 0$ a.s.*

Proof. Note that τ_s is the first hitting time of an open set (s,∞) for an adapted process $\langle M\rangle$ with continuous sample paths and hence τ_s is a stopping time (recall that $(\mathscr{F}_t)$ is right-continuous). Further, $\langle M\rangle_\infty = \infty$ a.s. implies that $\tau_s < \infty$ a.s. The process $(\tau_s : s \geq 0)$ is non-decreasing and right-continuous (in fact $s \to \tau_s$ is the right-continuous inverse of $t \to \langle M\rangle_t$). Let $\mathscr{G}_s := \mathscr{F}_{\tau_s}$. Note that it satisfies the usual assumptions. The process B is right continuous by continuity of M and right-continuity of τ. We have

$$\lim_{u \nearrow s} B_u = \lim_{u \nearrow s} M_{\tau_u} = M_{\tau_{s-}}.$$

But $[\tau_{s-}, \tau_s]$ is either a point or an interval of constancy of $\langle M\rangle$. The latter are known (exercise) to coincide a.s. with the intervals of constancy of M and hence $M_{\tau_{s-}} = M_{\tau_s} = B_s$ so that B has a.s. continuous paths. To conclude that B is a $(\mathscr{G}_s)$–Brownian motion, by Lévy's theorem, it remains to see that (B_s) and $(B_s^2 - s)$ are $(\mathscr{G}_s)$–local martingales.

Note that M^{τ_n} and $(M^{\tau_n})^2 - \langle M\rangle^{\tau_n}$ are uniformly integrable martingales. Taking $0 \leq u < s < n$ and applying the optional stopping theorem we obtain

$$\mathbb{E}[B_s|\mathscr{G}_u] = \mathbb{E}[M^{\tau_n}_{\tau_s}|\mathscr{F}_{\tau_u}] = M^{\tau_n}_{\tau_u} = M_{\tau_u} = B_u$$

and

$$\mathbb{E}[B_s^2 - s|\mathscr{G}_u] = \mathbb{E}\left[(M^{\tau_n}_{\tau_s})^2 - \langle M\rangle^{\tau_n}_{\tau_s}|\mathscr{F}_{\tau_u}\right] = (M^{\tau_n}_{\tau_u})^2 - \langle M\rangle^{\tau_n}_{\tau_u} = (M_{\tau_u})^2 - \langle M\rangle_{\tau_u} = B_u^2 - u,$$

where we used continuity of $\langle M\rangle$ to write $\langle M\rangle_{\tau_u} = u$. It follows that B is indeed a $(\mathscr{G}_s)$–Brownian motion.

Finally, $B_{\langle M\rangle_t} = M_{\tau_{\langle M\rangle_t}} = M_t$ again since the intervals of constancy of M and of $\langle M\rangle$ coincide a.s. so that $s \to \tau_s$ is constant on $[t, \tau_{\langle M\rangle_t}]$.

□

Appendix A
A very short primer in functional analysis

We start with a brief recall of basic notions of functional analysis leading to Hilbert spaces and identification of their dual mentioned above.

A.1 Normed vector spaces

We start with basics. A vector space V over $\mathbb{R}$ is a set endowed with two binary operations: addition and multiplication by a scalar, which satisfy the natural axioms. We focus the discussion on real scalars as this is relevant for us but most of what follows applies to spaces over complex numbers (or more general fields).

Definition A.1.1. A norm $\|\cdot\|$ on a vector space V is a mapping from V to $[0,\infty)$ such that

(i) for any $a \in \mathbb{R}$, $v \in V$, $\|av\| = |a|\|v\|$ (absolute homogeneity);
(ii) for any $x, y \in V$, $\|x+y\| \leq \|x\| + \|y\|$ (triangle inequality);
(iii) $\|v\| = 0$ if and only if v is the zero vector in V (separates points).

Note that a norm induces a metric on V through $d(x,y) = \|x-y\|$ and hence a topology on V. A norm is then a continuous function from V to $\mathbb{R}$. The space of continuous linear functions plays a special role:

Definition A.1.2. Given a normed vector space over reals $(V, \|\cdot\|_V)$, its dual V' is the space of all continuous linear maps (functionals) from V to $\mathbb{R}$. V' itself is a vector space over $\mathbb{R}$ equipped with a norm

$$\|\phi\|_{V'} := \sup_{v \in V, \|v\|_V \leq 1} |\phi(v)|.$$

The classical examples of spaces to consider are spaces of sequences or of functions. Let $(\mathscr{S}, \mathbb{F}, \mu)$ be a measurable space endowed with a σ-finite measure. Then for a real valued measurable function f on $\mathscr{S}$ we can consider

$$\|f\|_p := \left(\int_{\mathscr{S}} |f(x)|^p \mu(\mathrm{d}x) \right)^{1/p}$$

and let $\mathscr{L}^p(\mathscr{S}, \mathbb{F}, \mu)$ be the space of such functions for which $\|f\|_p < \infty$. Observe that $\|\cdot\|_p$ is not yet a norm on $\mathscr{L}^p$ – indeed it fails to satisfy (*iii*) in Definition A.1.1 since if $f = 0$ μ-a.e. but is not zero, e.g. $f = \mathbf{1}_A$ for a measurable $A \in \mathbb{F}$ with $\mu(A) = 0$, then still $\|f\|_p = 0$. We then say that $\|\cdot\|_p$ is a semi-norm on $\mathscr{L}^p$.

To remedy this, we consider the space $\mathrm{L}^p(\mathscr{S}, \mathbb{F}, \mu)$ which is the quotient of $\mathscr{L}^p$ with respect to the equivalence relation $f \sim g$ iff $f = g$ μ-a.e. Put differently, L^p is the space of equivalence classes of functions equal μ-a.e. and which are integrable with p^{th} power. Then $(\mathrm{L}^p(\mathscr{S}, \mathbb{F}, \mu), \|\cdot\|_p)$ is a normed vector space for $p \geq 1$. The triangle inequality for $\|\cdot\|_p$ is simply the *Minkowski inequality*.

A more geometric notion of measuring the relation between vectors is given by an inner product.

Definition A.1.3. Given a vector space V over $\mathbb{R}$, a mapping $< \cdot, \cdot >: V \times V \to \mathbb{R}$ is called an *inner product* if

(i) it is bilinear and symmetric: $< ax + bz, y >= a < x, y > + b < z, y >$ and $< x, y >=< y, x >$ for $a, b \in \mathbb{R}$, $x, y, z \in V$;
(ii) for any $x \in V$, $< x, x >\geq 0$;
(iii) $< x, x >= 0$ if and only if x is the zero vector in V.

This notion is all too well known on $V = \mathbb{R}^n$ where an inner product is given by

$$< x, y >= x^{\mathrm{T}} y = \sum_{i=1}^{n} x_i y_i.$$

An inner product satisfies the Cauchy-Schwartz inequality

Proposition A.1.4. *An inner product on a vector space satisfies*

$$| < x, y > | \leq \sqrt{< x, x >} \sqrt{< y, y >}, \quad x, y \in V. \tag{A.1}$$

Proof. Let $\|x\| := \sqrt{< x, x >}$, $x \in V$. Fix $x, y \in V$ and define a quadratic function $Q : \mathbb{R} \to \mathbb{R}$ by

$$Q(r) = \|x + ry\|^2 = \|y\|^2 r^2 + 2 < x, y > r + \|x\|^2, \quad r \in \mathbb{R}$$

which clearly is non-negative and hence its discriminant has to be non-positive i.e.

$$4| < x, y > |^2 - 4\|x\|^2 \|y\|^2 \leq 0 \text{ that is } | < x, y > | \leq \|x\| \|y\|,$$

as required. We note also that equality holds if and only if the vectors x, y are linearly dependent i.e. $x = ry$ for some $r \in \mathbb{R}$. □

The above implies that $\|x\| := \sqrt{< x, x >}$ is a norm on V. We say that the norm is induced by an inner product. Among spaces L^p defined above only L^2 has norm which is induced by an inner product, namely by

$$< f,g >= \int_{\mathscr{S}} f(x)g(x)\mu(\mathrm{d}x). \tag{A.2}$$

A.2 Banach spaces

We first define Cauchy sequences which embody the idea of a converging sequence when we do not know the limiting element.

Definition A.2.1. A sequence (x_n) of elements in a normed vector space $(X, \|\cdot\|)$ is called a Cauchy sequence if for any $\varepsilon > 0$ there exists $N \geq 1$ such that for all $n, m \geq N$ we have $\|x_n - x_m\| \leq \varepsilon$.

Definition A.2.2. A normed vector space $(X, \|\cdot\|)$ is complete if every Cauchy sequence converges to an element $x \in X$. It is then called a *Banach space*. Further, if the norm is induces by an inner product then $(X, \|\cdot\|)$ is called a *Hilbert space*.

Naturally, the Euclidean space $\mathbb{R}^d$ is a Banach space (and in fact a Hilbert space) with the norm $\|x\| = \sqrt{\sum_{i=1}^d x_i^2}$. This implies (reasoning for $d = 1$) that

Proposition A.2.3. *If $(X, \|\cdot\|_X)$ is a normed vector space over $\mathbb{R}$ then its dual $(X', \|\cdot\|_{X'})$ in Definition A.1.2 is a Banach space.*

In many cases it is interesting to build linear functionals satisfying certain additional properties. This is often done using the Hahn-Banach theorem. It states in particular that a continuous linear functional defined on a linear subspace Y of X can be continuously extended to the whole of X without increasing its norm. Its version is also known as the separating hyperplane theorem since it allows to separate two convex sets (one open) using an affine hyperplane.

An important step in studying continuous linear functionals on X is achieved by describing the structure of X'. We have

Proposition A.2.4. *Let $(\mathscr{S}, \mathbb{F}, \mu)$ be a measurable space with a σ-finite measure. Then for any $p \geq 1$, $\mathrm{L}^p(\mathscr{S}, \mathbb{F}, \mu)$ is a Banach space and for $p > 1$ its dual is equivalent to (isometric to) the space $\mathrm{L}^q(\mathscr{S}, \mathbb{F}, \mu)$, where $1/p + 1/q = 1$.*

In particular we see that L^2 is its own dual. This means that any continuous linear functional on L^2 can be identified with an element in L^2. This property remains true for any Hilbert space:

Proposition A.2.5. *Let $(X, \|\cdot\|)$ be a Hilbert space with the norm induced by an inner product, $\|x\| = \sqrt{< x,x >}$. If $\phi : X \to \mathbb{R}$ is a continuous linear map then there exists an element $x_\phi \in X$ such that*

$$\phi(y) =< x,y >, \quad \forall y \in X.$$

In particular, if $\phi : \mathrm{L}^2(\mathscr{S}, \mathbb{F}, \mu) \to \mathbb{R}$ is a continuous linear map then there exists an element $f_\phi \in \mathrm{L}^2(\mathscr{S}, \mathbb{F}, \mu)$ such that

$$\phi(g) = \int_{\mathscr{S}} g(x) f_\phi(x) \mu(\mathrm{d}x), \quad \forall g \in \mathrm{L}^2(\mathscr{S}, \mathbb{F}, \mu).$$

Note that the inner product $< x, y >$, or the integral $\int_{\mathscr{S}} g(x) f_\phi(x) \mu(\mathrm{d}x)$, in the above statement are well defined by (A.1)–(A.2).

On a (separable) Hilbert space, we can also state an infinite dimensional analogue of the Pythagorean theorem. Recall that $\mathbb{R}^n$ is a Hilbert space with inner product $< x, y >= \sum_{i=1}^n x_i y_i$. This uses the canonical basis in $\mathbb{R}^n$ but if we take any orthonormal basis in $\mathbb{R}^n$, say $(\varepsilon_1, \ldots, \varepsilon_n)$, then

$$x = \sum_{i=1}^n < x, \varepsilon_i > \varepsilon_i \quad \text{and hence } \|x\|^2 = \sum_{i=1}^n < x, \varepsilon_i >^2,$$

which is the Pythagorean theorem. The same reasoning gives $< x, y >= \sum_{i=1}^n < x, \varepsilon_i >< y, \varepsilon_i >$. The infinite dimensional version is known as the

Proposition A.2.6 (Parseval's identity). *Let $(X, \|\cdot\|)$ be a separable Hilbert space with the norm induced by an inner product $(x, y) \to < x, y >$ and let $(\varepsilon_n : n \geq 1)$ be an orthonormal basis of X. Then for any $x, y \in X$*

$$< x, y >= \sum_{n \geq 1} < x, \varepsilon_n >< y, \varepsilon_n >, \quad \textit{and in particular } \|x\|^2 = \sum_{n \geq 1} < x, \varepsilon_n >^2 .$$

Finally we state one more result, which is crucial for the construction of the stochastic integral in the Section 7.1.1.

Proposition A.2.7. *Suppose $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are two Banach spaces, $\mathscr{E} \subset X$ is a dense vector subspace in X and $I : \mathscr{E} \to Y$ is a linear isometry, i.e. a linear map which preserves the norm, $\|I(x)\|_Y = \|x\|_X$ for all $x \in X$. Then I may be extended in an unique way to a linear isometry from X to Y.*

Proof. Take $x \in X$ and $x_n \to x$ with $x_n \in \mathscr{E}$. Then, by the isometry property, $(I(x_n))$ is Cauchy in Y since (x_n) is Cauchy in X. It follows that it converges to some element which we denote $I(x)$. Further, if we have two sequences in $\mathscr{E}$, (x_n) and (y_n), both converging to $x \in X$ and giving raise to potentially two elements $I(x)$ and $I(x)'$ then we can build a third sequence $z_{2n} = x_n$, $z_{2n+1} = y_n$ which also converges to x and we see that $I(z_n)$ has to converge and the limit has to agree with both $I(x)$ and $I(x)'$, so that $I(x) = I(x)'$ is unique. It follows that we defined $I(x) \in Y$ uniquely for all $x \in X$. Further,

$$\|I(x)\|_Y = \lim_n \|I(x_n)\|_Y = \lim_n \|x_n\|_X = \|x\|_X$$

so that I is norm-preserving. Finally, if $x, y \in X$ then we can write them as limits of sequences of elements in $\mathscr{E}$, say (x_n) and (y_n) respectively. For any $a, b \in \mathbb{R}$, $ax_n + by_n \in \mathscr{E}$ since $\mathscr{E}$ is a vector space, and then by the above and linearity of I on $\mathscr{E}$ we have

$$I(ax + by) = \lim_n I(ax_n + by_n) = \lim_n \Big(aI(x_n) + bI(y_n) \Big) = aI(x) + bI(y),$$

so that I is linear on X as required. $\square$