Introduction to stochastic calculus

Lecture notes
Rough preliminary version

Master of mathematics
Université Paris-Dauphine
PSL Université Paris

2017–2018
Version compiled November 2, 2018
These are the lecture notes of an introduction course on stochastic calculus, given at Université Paris-Dauphine, for second year master students in mathematics\(^1\). The prerequisite is a probability theory course based on measure theory and Lebesgue integral, including conditional expectation, gaussian random vectors, and standard notions of convergence. The initial version of these lecture notes was based on a course given by Halim Doss, inspired from the book by Nobuyuki Ikeda and Sinzo Watanabe [IW89]. The current version is largely inspired from the books by Fabrice Baudoin [Bau14] and Jean-François Le Gall [LG16].

**Notations.**

- \( \mathbb{R}_+ = [0, \infty) \);
- \( e \) = exponential;
- \( \text{iff} \) = if and only if;
- \( 1_A \) is the indicator of \( A \);
- \( \mathcal{B}_E \) : Borel \( \sigma \)-algebra of \( E \);
- \( d \) is the differential element;
- \( s \wedge t = \min(s, t) \), \( s \vee t = \max(s, t) \);
- \( d, i, j, k, m, n, \ell \) are integer numbers;
- \( p, q, r, s, t, u, v, \alpha, \beta, \varepsilon \) are real numbers;
- \( f \) is increasing means \( f(y) \geq f(x) \) if \( y \geq x \);
- \( i \) can be sometimes the complex number \( (0, 1) \);
- \( X \sim \mu \) \( X \) means the random variable \( X \) has law \( \mu \);
- \( \langle x, y \rangle_\mathcal{H} \) scalar product in the Hilbert space \( \mathcal{H} \);
- \( L^p_{\mathbb{R}^d}(\Omega, \mathbb{P}) \) : r.v. \( X : \Omega \to \mathbb{R}^d \) with \( \mathbb{E}(\|X\|^p) < \infty \);
- \( x \cdot y = x_1y_1 + \cdots + x_dy_d \) for all \( x, y \in \mathbb{R}^d \);
- \( |x| = \sqrt{x_1^2 + \cdots + x_d^2} \) for all \( x \in \mathbb{R}^d \).

\(^1\)MASEF (Mathématiques pour l’économie et la finance) and MATH (Mathématiques appliquées et théoriques).
# Contents

0 Motivation .......................... 1

1 Preliminaries ........................ 3
   1.1 Independence ...................... 3
   1.2 Markov inequality, convergences, Borel–Cantelli lemma, ...................... 3
   1.3 Uniform integrability .............. 5
   1.4 Conditioning ...................... 6
   1.5 Gaussian random vectors .......... 9
   1.6 Bounded variation and Lebesgue–Stieltjes integral ...................... 9
   1.7 Monotone class theorem and Carathéodory extension theorem .......... 11

2 Processes, filtrations, stopping times, martingales .................. 13
   2.1 Measurability ...................... 13
   2.2 Stopping times ...................... 16
   2.3 Quadratic variation .................. 17
   2.4 Martingales ...................... 18

3 Brownian motion ...................... 29
   3.1 Markov property of Brownian motion ...................... 33
   3.2 A construction of Brownian motion ...................... 36
   3.3 Wiener integral ...................... 39
   3.4 Wiener measure, canonical Brownian motion, Cameron–Martin formula .............. 41

4 Itô stochastic integral .................. 45
   4.1 Intuitive mathematical experiments ...................... 45
   4.2 Stochastic integral with respect to Brownian motion ...................... 46
      4.2.1 Stochastic integral of step processes ...................... 46
      4.2.2 Extension to square integrable progressive processes .............. 48
      4.2.3 Further extension by localization ...................... 54
   4.3 Stochastic integral with respect to continuous martingales bounded in $L^2$ ...................... 55
   4.4 Stochastic integral with respect to continuous local martingales ...................... 59
   4.5 Stochastic integral with respect to semi-martingales ...................... 61

5 Itô formula and applications .................. 65
   5.1 Itô formula ...................... 65
   5.2 Lévy characterization of Brownian motion and Dubins–Schwarz theorem ...................... 67
   5.3 Girsanov theorem for Itô integrals ...................... 69
   5.4 Sub-Gaussian deviation inequality ...................... 70
   5.5 Burkholder–Davis–Gundy inequalities ...................... 70

6 Stochastic differential equations .................. 73
   6.1 Stochastic differential equations with Lipschitz coefficients ...................... 73
   6.2 Deterministic case ...................... 77
   6.3 Homogeneous case ...................... 79
   6.4 Locally Lipschitz coefficients and explosion time ...................... 85
6.5 Girsanov theorem .................................................. 88

7 Probabilistic formulation of Dirichlet problems .......................... 91

A Selected topics related to stochastic calculus ......................... 95
   A.1 Recurrence and transience of Brownian motion .................. 95
   A.2 Fokker–Planck equation ............................................. 95
   A.3 Feynman–Kac formula ............................................... 95
   A.4 Hamilton–Jacobi–Bellman equation ............................... 95
   A.5 Statistical inference of diffusion processes .................... 95
   A.6 Euler–Maruyama numerical scheme ............................... 95
   A.7 Stratonovich integral ................................................ 95
   A.8 Local time and Tanaka formula .................................... 95
   A.9 Exemples of stochastic processes ................................ 95
      A.9.1 Infinitesimal generator = simulation algorithm .......... 95
      A.9.2 Ornstein–Uhlenbeck and Bakry–Émery processes .......... 95
      A.9.3 Laguerre and Jacobi processes ............................... 95
      A.9.4 Bessel and Cox–Ingersoll–Ross processes .................. 95
      A.9.5 Dyson Brownian motion and Dyson–Ornstein–Uhlenbeck process .... 95
      A.9.6 Fisher–Wright processes ...................................... 95
      A.9.7 Diffusions with jumps and piecewise deterministic Markov processes .... 95
   A.10 Lévy area .............................................................. 95
   A.11 Additive functionals, ergodic theorem, central limit theorem .... 95
   A.12 White noise .......................................................... 95

Bibliography ................................................................. 97
Chapter 0

Motivation

In this introductory course on stochastic calculus, the goal is to define integrals of the form

\[ I_t = \int_0^t X_s dY_s, \quad t \geq 0, \]

where \((X_t)_{t \geq 0}\) and \((Y_t)_{t \geq 0}\) are stochastic processes. The result is a stochastic process \((I_t)_{t \geq 0}\). If we subdivide the interval \([0, t]\) into \([t_0, t_1] \cup \ldots \cup [t_{n-1}, t_n]\) with \(0 = t_0\) and \(t_n = t\), then the hypothetic quantity \(I_t\) is naturally approximated by the (random) Riemann sum

\[ \sum_{i=0}^{n-1} \tilde{X}_i (Y_{t_{i+1}} - Y_{t_i}), \quad \text{where} \quad \tilde{X}_i \in [X_{t_i}, X_{t_{i+1}}], \quad i = 0, \ldots, n-1. \]

Following Riemann, Stieltjes, and L. C. Young among others. The convergence of this quantity when \(n \to \infty\) with \(\max_i (t_{i+1} - t_i) \to 0\) is guaranteed when the integrator and the integrand are regular enough. Unfortunately, this does not work for instance when both \(X\) and \(Y\) are Brownian motion, due to the fact that the sample paths of this stochastic process are of infinite variation on all finite intervals. The solution found by Itô is to take advantage of the stochastic nature of Brownian motion and to consider the limit of Riemann sums in \(L^2\) or in probability. Taking \(\tilde{X}_i = X_{t_i}\) and following this idea leads to what is known as the Itô integral, for which \((I_t)_{t \geq 0}\) is typically a martingale. This approach remains valid far beyond Brownian motion, for a class of integrators known as semi-martingales. There is also a fundamental formula of calculus called the Itô formula or lemma, involving an additional quadratic term.

The Itô integral will allow us to define and compute in particular \(I_t\) when both \(X\) and \(Y\) are Brownian motion, and more generally to solve stochastic differential equations of the form

\[ X_t = X_0 + \int_0^t \sigma(X_s) dB_s + \int_0^t b(X_s) ds, \quad t \geq 0, \]

where for instance \(B = (B_t)_{t \geq 0}\) is a Brownian motion, and where \(\sigma\) and \(b\) are regular enough coefficients, typically locally Lipschitz, as in the classical Cauchy–Lipschitz theorem. Moreover the law of the solution \(X\) on the trajectories will turn out to be absolutely continuous with respect to the one of \(B\), and there is a formula for the density, known as the Girsanov formula.

Finally we will give a probabilistic representation of the solution of the Dirichlet problem, due to Kacutani, which plays an important role in the connection between probability theory and the analysis of partial differential equations. These lecture notes end with a collection of additional themes of interest involving stochastic calculus, which is enriched progressively.

---

1Taking \(X_i = \frac{1}{2} (X_{t_i} + X_{t_{i+1}})\) leads to the Strato(novich) integral, which has advantages and drawbacks.

2A smooth function is the integral of its derivative.
Chapter 1

Preliminaries

1.1 Independence

Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space.

1. We say that a family \((\mathcal{A}_i)_{i \in I}\) of sub-\(\sigma\)-algebras of \(\mathcal{A}\) is independent when for all finite \(J \subset I\) and for all \(A_i \in \mathcal{A}_i\) we have
\[
\mathbb{P}(\bigcap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i).
\]

2. We say that a family \((X_i)_{i \in I}\) of random variables is independent, \(X_i : (\Omega, \mathcal{A}) \to (E_i, \mathcal{B}_i)\), when the family of sub-\(\sigma\)-algebras \((\sigma(X_i))_{i \in I}\) is independent, where
\[
\sigma(X_i) = \{X_i^{-1}(B) : B \in \mathcal{B}_i\}
\]
is the \(\sigma\)-algebra generated by \(X_i\). Thus \((X_i)_{i \in I}\) is independent iff for all \(J \subset I\) finite,
\[
\mathbb{P}_{X_i : i \in J} = \otimes_{i \in J} \mathbb{P}_{X_i} \quad \text{on} \quad (\prod_{i \in J} E_i, \otimes_{i \in J} \mathcal{B}_i).
\]

[If \(Z\) is a random variable then we denote its law by \(\mathbb{P}_Z\).] It follows that if \(X_1, X_2, \ldots, X_n\) are real random variables integrable and independent then
\[
\prod_{i=1}^n X_i \in L^1 \quad \text{and} \quad \mathbb{E}(\prod_{i=1}^n X_i) = \prod_{i=1}^n \mathbb{E}(X_i).
\]

1.2 Markov inequality, convergences, Borel–Cantelli lemma, …

Markov inequality. If \(\varphi(X) \in L^1\) and \(\varphi \geq 0\) then for all \(r > 0\),
\[
\mathbb{P}(\varphi(X) \geq \varphi(r)) \leq \mathbb{E}(\varphi(X))/\varphi(r).
\]

Convergences. Below \((X_n)_{n \geq 1}, (Y_n)_{n \geq 1}, X, Y\) are real random variables on a probability space \((\Omega, \mathcal{A}, \mathbb{P})\), of law \(\mu_n, \nu_n, \mu, \nu\) and cumulative distribution function \(F_n, G_n, F, G\) respectively.

Almost sure convergence. We say that \(X_n \xrightarrow{a.s.} X\) when
\[
\mathbb{P}(\lim_{n \to \infty} X_n = X) = 1
\]
in other words \(\mathbb{P}(\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)) = 1\). This is the notion of convergence which appears in the strong law of large numbers (SLLN).

Convergence in probability. We say that \(X_n \xrightarrow{p} X\) when
\[
\forall \varepsilon > 0, \quad \lim_{n \to \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0
\]
which means that \(\forall \varepsilon > 0, \lim_{n \to \infty} \mathbb{P}(\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon) = 0\). This is the notion of convergence which appears in the weak law of large numbers (a consequence of the CLT).
Mean convergence. For all \( p \in [1, \infty) \), we say that \( X_n \xrightarrow{L^p} X \) when

\[
X \in L^p \quad \text{and} \quad \lim_{n \to \infty} \mathbb{E}(|X_n - X|^p) = 0,
\]

which means that \( X_n \xrightarrow{\text{a.s.}} X \) in \( L^p \). The most useful cases are \( p \in \{1, 2, 4\} \).

Convergence in law. The following properties are equivalent and we say then that \( X_n \xrightarrow{\text{law}} X \), ou encore \( X_n \xrightarrow{d} \mu \) (convergence in distribution), or \( \mu_n \xrightarrow{\text{nar.}} \mu \) (narrow convergence). This is the notion of convergence used in the central limit theorem (CLT).

1. \( \lim_{n \to \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \) for all bounded and measurable \( f : \mathbb{R} \to \mathbb{R} \);
2. \( \lim_{n \to \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \) for all \( \mathcal{C}^\infty \) and compactly supported \( f : \mathbb{R} \to \mathbb{R} \);
3. \( \lim_{n \to \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \) for all \( f = 1_{(-\infty, x]} \) with \( x \) continuity point of \( F \), in other words \( F_n(x) \to F(x) \) as soon as \( F \) is continuous at \( x \); for all \( f = e^{itx}, t \in \mathbb{R} \);
4. \( \lim_{n \to \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \) for all \( f = e^{-itx}, t \geq 0 \).
5. (on \( \mathbb{R}_+ \)) \( \lim_{n \to \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \) for all \( f \).

Contrary to the other modes of convergence, the convergence in law does not depend on the law of the couple \((X_n, X)\). It uses only marginal laws: law of \( X_n \) (for all \( n \)) and law of \( X \). The couple of last properties above are useful to transform sums of independent random variables into products, for which the expectation is the product of expectations.

Apart the convergence in law, the other modes of convergence are stable by finite linear combinations. The almost sure convergence, the convergence in probability, and the convergence in law are stable by composition with continuous functions.

The notions of convergence extend naturally to random vectors by using a distance/norm/scalar product, for instance for the characteristic function by replacing \( itX \) by \( i(t, X) \).

\[
\begin{align*}
\text{L}^p \text{ CV} & \quad \Rightarrow \quad \text{L}^1 \text{ CV} \\
\text{a.s. CV} & \quad \Rightarrow \quad \text{CV in P} & \Rightarrow \quad \text{CV in law}
\end{align*}
\]

If \( X \) is constant then the convergence in law implies the convergence in probability. The convergence in \( \text{L}^1 \) is equivalent to uniform integrability and convergence in probability.

Monotone convergence theorem. If \( (X_n)_{n \geq 1} \) is \( \geq 0 \) and \( \sup \) then

\[
\lim_{n \to \infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \to \infty} X_n)
\]

in \( [0, \infty] \), and in particular \( \mathbb{E}(\lim_{n \to \infty} X_n) < \infty \) if \( \lim_{n \to \infty} \mathbb{E}(X_n) < \infty \).

Fatou lemma. If \( (X_n)_{n \geq 1} \) is \( \geq 0 \) then

\[
\lim_{n \to \infty} \mathbb{E}(X_n) \geq \mathbb{E}(\lim_{n \to \infty} X_n).
\]

Useful for sequences which are neither monotonous nor bounded.

Dominated convergence theorem. If \( X_n \xrightarrow{\text{a.s.}} X \) and \( \sup_n |X_n| \leq Y, \mathbb{E}(Y) < \infty \), then

\[
\lim_{n \to \infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \to \infty} X_n) = \mathbb{E}(X).
\]

The dominated convergence is an easy to check criterion of uniform integrability.

---

1 Cumulative distribution function.
2 Fourier transform or characteristic function.
3 Laplace transform.
1.3 Uniform integrability

**Scheffé lemma.** If \( X_n, X \in L_1 \), \( X_n \xrightarrow{a.s.} X \) then \( X_n \xrightarrow{L_1} X \iff \mathbb{E}(\{|X_n|\} \to \mathbb{E}(\{|X|\}).

**Slutsky lemma.** If \( X_n \xrightarrow{law} X \) and \( Y_n \xrightarrow{law} Y \) and \( Y \) is constant then \( (X_n, Y_n) \xrightarrow{law} (X, Y) \). In particular 
\( X_n, Y_n \xrightarrow{law} X + Y \); \( X_n/Y_n \xrightarrow{law} X/Y \) if \( Y \neq 0 \).

**Fubini–Tonelli theorem.** Let \((\Omega_1, \mathcal{A}_1, \mu_1)\) and \((\Omega_2, \mathcal{A}_2, \mu_2)\) two measurable spaces, and let \( f : \Omega_1 \times \Omega_2 \to \mathbb{R} \) be a measurable function. If \( f \geq 0 \) or if \( f \in L^1(\mu_1 \otimes \mu_2) \) then
\[
\int f(x, y) d(\mu_1 \otimes \mu_2)(x, y) = \int \left( \int f(x, y) d\mu_1(x) \right) d\mu_2(y).
\]

**Borel–Cantelli lemma.** Let \((A_n)_n\) be events in a probability space \((\Omega, \mathcal{A}, \mathbb{P})\). We define
\[
\begin{align*}
\limsup_n A_n &= \bigcup_n \bigcap_{m \geq n} A_m = \{ \omega \in \Omega : \omega \in A_n \text{ for } n \text{ large enough} \}, \\
\liminf_n A_n &= \bigcap_n \bigcup_{m \geq n} A_m = \{ \omega \in \Omega : \omega \in A_n \text{ for infinitely many values of } n \}.
\end{align*}
\]
We have \( (\limsup_n A_n)^c = \liminf_n A_n \), and \( \limsup_n 1_{A_n} = 1 \limsup_n A_n \) and \( \liminf_n 1_{A_n} = 1 \liminf_n A_n \).

1. (Cantelli) if \( \sum_n \mathbb{P}(A_n) < \infty \) then \( \mathbb{P}(\limsup_n A_n) = 0 \);
2. (Borel zero-one law) if \( \sum_n \mathbb{P}(A_n) = \infty \) and the \((A_n)_n\) are independent then \( \mathbb{P}(\limsup_n A_n) = 1 \).

The Borel–Cantelli lemma is a great provider of almost sure convergence. Note that if \( X \geq 0 \) then \( \mathbb{E}(X) < \infty \) implies \( \mathbb{P}(X < \infty) = 1 \), and this allows to prove the Cantelli part:
\[
\sum_n \mathbb{P}(A_n) = \mathbb{E} \sum_n 1_{A_n} = \mathbb{E} \liminf_n 1_{A_n} \quad \text{ and } \quad \left\{ \sum_n 1_{A_n} = \infty \right\} = \liminf_n A_n.
\]

1.3 Uniform integrability

Let \( \Phi \) be the class of non-decreasing functions \( \varphi : \mathbb{R}_+ \to \mathbb{R}_+ \) such that \( \lim_{x \to +\infty} \varphi(x)/x = +\infty \). It contains for instance the convex functions \( x \mapsto x^p \), \( p > 1 \), and \( x \mapsto x\log(x) \). Let \( L^q \) be the set of random variables \( X \) on \((\Omega, \mathcal{F}, \mathbb{P})\) such that \( \varphi(\{|X|\}) \in L^1 = L^1(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{R}) \). We have \( L^p \subseteq L^1 \). Clearly, if \( (X_i)_{i \in I} \subseteq L^1 \) is bounded in \( L^p \) with \( \varphi \in \Phi \) then \((X_i)_{i \in I}\) is bounded in \( L^1 \).

For any family \((X_i)_{i \in I} \subseteq L^1\), the following three properties are equivalent. When one (and thus all) of these properties holds true, we say that the family \((X_i)_{i \in I}\) is **uniformly integrable** (UI). The first property can be seen as a natural definition of uniform integrability.

1. (definition of uniform integrability) \( \lim_{x \to +\infty} \sup_{i \in I} \mathbb{E}(\varphi(\{|X_i|\} \cdot 1_{|X_i| \leq x})) = 0 \);
2. (epsilon-delta criterion) the family is bounded in \( L^1 \): \( \sup_{i \in I} \mathbb{E}(\varphi(\{|X_i|\})) < \infty \), and moreover \( \forall \varepsilon > 0, \exists \delta > 0, \forall A \in \mathcal{F}, \mathbb{P}(A) \leq \delta \Rightarrow \sup_{i \in I} \mathbb{E}(\varphi(\{|X_i|\} \cdot 1_A)) \leq \varepsilon \);
3. (de la Vallée Poussin\(^4\) boundedness in \( L^p \) criterion) there exists \( \varphi \in \Phi \) such that \( (X_i)_{i \in I} \) is bounded in \( L^p \subseteq L^1 \), namely \( \sup_{i \in I} \mathbb{E}(\varphi(\{|X_i|\})) < \infty \).

Here are examples for uniformly integrable families:

- every finite subset of \( L^1 \) is uniformly integrable. In particular if \( X \in L^1 \) then there exists \( \varphi \in \Phi \) such that \( X \in L^p \) (beware that this convex function \( \varphi \) depends on \( X \));
- if \((X_i)_{i \in I}\) is bounded in \( L^p \) with \( p > 1 \) then it is uniformly integrable;
- if \( \sup_{i \in I} \varphi(\{|X_i|\}) \in L^1 \) (domination: \( \varphi(\{|X_i|\}) \leq X \in L^1 \) for all \( i \in I \)) then \((X_i)_{i \in I}\) is UI.
- if \( X_n \xrightarrow{L_1} X \) then \((X_n)_{n \geq 1}, (X_n)_{n \geq 1} \cup \{X\}, \text{ and } (X_n - X)_{n \geq 1}\) are UI.

The notion of uniform integrability leads to a stronger version of the dominated convergence theorem: for any \((X_n)_{n \geq 1}\) and \( X \in L^1 \), we have
\[
X_n \xrightarrow{L_1} X \quad \text{if and only if } \quad X_n \xrightarrow{P} X \quad \text{and } \quad (X_n)_{n \geq 1} \text{ is UI}.
\]

The dominated convergence theorem corresponds to the special case \( \sup_n \|X_n\| \in L^1 \).

\(^4\)After Charles-Jean Étienne Gustave Nicolas de la Vallée Poussin (1866 – 1962), Belgian mathematician.
1.4 Conditioning

1. Orthogonal projection in a Hilbert space. Let $H$ be a Hilbert space and $F \subset H$ be a closed sub-space. For all $x \in H$ there exists a unique $y \in F$, called the orthogonal projection of $x$ on $F$, which satisfies one (and thus all) the following equivalent properties:

(a) for all $z \in F$, $x - y \perp z$ i.e. $\langle x, z \rangle = \langle y, z \rangle$

(b) for all $z \in F$, $\| x - y \| \leq \| x - z \|$ i.e. $\| x - y \| = \inf_{z \in F} \| x - z \|$.

2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}$ be a sub-$\sigma$-algebra of $\mathcal{A}$. Let us consider the Hilbert space $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$. The set $F = L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a closed sub-space of $H$. If $X$ is a square integrable random variable i.e. an element of $H$, it is natural to consider the best (least squares) approximation of $X$ by an element of $F$, denoted $Y$. The random variable $Y$ is the orthogonal projection of $X$ on $F$, characterized by the following:

(a) $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$;

(b) for all $Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $E(|X - Y|^2) \leq E(|X - Z|^2)$.

Using the relation to scalar product, the second property is equivalent to

- for all $Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $E(XZ) = E(YZ)$

and also to

- for all $B \in \mathcal{B}$, $E(X1_B) = E(Y1_B)$.

This gives three characterizations of $Y$. We denote $Y = E(X | \mathcal{F})$ and we call it the **conditional expectation of $Y$ given $\mathcal{F}$**. It is the best approximation in $L^2$ (in a sense least squares) of $X$ by an $\mathcal{F}$-measurable square integrable random variable.

3. Suppose now that $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. We can, by extension, define the conditional expectation $Y = E(X \mid \mathcal{F})$. It is a real random variable characterized by:

(a) $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$

(b) for all $Z$ bounded and $\mathcal{F}$ measurable, we have $E(XZ) = E(YZ)$. Equivalently, for all $B \in \mathcal{F}$, we have $E(X1_B) = E(Y1_B)$.

**Proof.** Let $\mu$ be the bounded measure on $(\Omega, \mathcal{F})$ defined by $\mu(B) = E(X1_B)$, $B \in \mathcal{F}$. Let us define $\nu = \frac{\mu}{\mathbb{P}}$. For all $B \in \mathcal{F}$, if $\nu(B) = 0$ then $\mu(B) = 0$. From the Radon–Nikodym theorem, there exists a unique $Y \in L^1(\Omega, \mathcal{F}, \nu)$ such that $\int_B Y d\nu = \mu(B)$, for all $B \in \mathcal{F}$ i.e. $E(Y1_B) = E(X1_B)$, for all $B \in \mathcal{F}$. □

The expectation and the variance of square integrable random variables have a variational interpretation. Namely if $X \in L^2$ then its variance $\text{var}(X)$ is the square distance in $L^2$ of $X$ to the sub-space of constant random variables:

$$\text{var}(X) = \inf_{c \in \mathbb{R}} E((X - c)^2) = \inf_{c \in \mathbb{R}} (E(X^2) - 2cE(X) + c^2).$$

This infimum is a minimum, achieved for $c = E(X)$, which is therefore the orthogonal projection of $X$ in $L^2$ on the sub-space of constant random variables, and

$$\text{var}(X) = E((X - E(X))^2) = E(X^2) - 2E(XE(X)) + (E(X))^2 = E(X^2) - (E(X))^2.$$

The following identity is an instance of the Pythagoras theorem in $L^2$:

$$\text{var}(X) = E(X^2) - (E(X))^2 = E(X^2) - E((E(X \mid \mathcal{F}))^2) + E((E(X \mid \mathcal{F}))^2) - (E(X))^2.$$
= \mathbb{E}(\text{var}(X \mid \mathcal{F})) + \text{var}(\mathbb{E}(X \mid \mathcal{F}))

where

\text{var}(X \mid \mathcal{F}) = \mathbb{E}(X^2 \mid \mathcal{F}) - (\mathbb{E}(X \mid \mathcal{F}))^2.

Note that by definition of \( \mathbb{E} \) where positivity, normalization, monotone convergence, Jensen and Hölder inequalities, etc.

1.4 Conditioning

The expectation operator \( \mathbb{E} \) is the conditional expectation associated to the trivial sub-\( \sigma \)-algebras \( \{\emptyset, \Omega\} \), hence a particular case. Actually the conditional expectation operator have all the properties of an expectation, and more. Namely, for all sub-\( \sigma \)-algebra \( \mathcal{F} \) of \( \mathcal{A} \):

- **Linearity.** for all \( \alpha, \beta \in \mathbb{R} \) and \( X, Y \in L^1 \), \( \mathbb{E}(\alpha X + \beta Y \mid \mathcal{F}) = \alpha \mathbb{E}(X \mid \mathcal{F}) + \beta \mathbb{E}(Y \mid \mathcal{F}); \)
- **“Projection”**. If \( X \) is \( \mathcal{F} \)-measurable, \( Y \in L^1 \), \( XY \in L^1 \), then \( \mathbb{E}(XY \mid \mathcal{F}) = X \mathbb{E}(Y \mid \mathcal{F}), \) in particular \( \mathbb{E}(X \mid \mathcal{F}) = X \) if \( X \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \) which is the case when \( X \) is constant;
- **Composed “projections”**. For all sub-\( \sigma \)-algebras \( \mathcal{F}, \mathcal{G} \) with \( \mathcal{G} \subset \mathcal{F} \) and all \( X \in L^1 \),
  \[ \mathbb{E}(X \mid \mathcal{F}) \mid \mathcal{G} = \mathbb{E}(X \mid \mathcal{G} \mid \mathcal{F}) = \mathbb{E}(X \mid \mathcal{G}), \]
  and in particular for all \( X \in L^1 \)
  - we have \( \mathbb{E}(\mathbb{E}(X \mid \mathcal{F})) = \mathbb{E}(X); \)
  - if \( X \) is independent of \( \mathcal{F} \) then \( \mathbb{E}(X \mid \mathcal{F}) = \mathbb{E}(X); \)
  - if \( X \) is constant then \( \mathbb{E}(X \mid \mathcal{F}) = X; \)
- **Normalization.** \( \mathbb{E}(1_\Omega \mid \mathcal{F}) = 1_\Omega \) (follows from projection properties);
- **Positivity or monotonicity.** For all \( X, Y \in L^1 \), if \( X \leq Y \) then \( \mathbb{E}(X \mid \mathcal{F}) \leq \mathbb{E}(Y \mid \mathcal{F}), \) or equivalently for all \( X \in L^1 \) if \( X \geq 0 \) then \( \mathbb{E}(X \mid \mathcal{F}) \geq 0. \) In particular for all \( X \in L^1 \),
  \[ |\mathbb{E}(X \mid \mathcal{F})| \leq \mathbb{E}(|X| \mid \mathcal{F}); \]
- **Convexity.** Jensen inequality: for all non-negative convex \( \varphi : \mathbb{R} \rightarrow \mathbb{R} \) and all \( X \in L^1 \),
  \[ \varphi(\mathbb{E}(X \mid \mathcal{F})) \leq \mathbb{E}(\varphi(X) \mid \mathcal{F}). \]

In particular, for all \( p \in [1, \infty) \), \( |\mathbb{E}(X \mid \mathcal{F})|^p \leq \mathbb{E}(|X|^p \mid \mathcal{F}). \) Moreover for all \( X \in L^p \) and \( Y \in L^q \) with \( 1 \leq p, q < \infty, 1/p + 1/q = 1 \) \( (q = p/(p-1)), \) we have the Hölder inequality
  \[ |\mathbb{E}(XY \mid \mathcal{F})| \leq |\mathbb{E}(X|^p \mid \mathcal{F})|^{1/p} |\mathbb{E}(|Y|^q \mid \mathcal{F})|^{1/q}. \]

The Cauchy–Schwarz inequality corresponds to the special case \( p = q = 1/2; \)
- **Monotone convergence.** If \( X_n \geq 0, X_n \not\rightarrow X, X \in L^1 \), then \( \mathbb{E}(X_n \mid \mathcal{F}) \not\rightarrow \mathbb{E}(X \mid \mathcal{F}). \) This allows to define \( \mathbb{E}(X \mid \mathcal{F}) \) for all non-negative random variable \( X. \)

The moral is that the conditional expectation is an expectation, so it has the same properties: linearity, positivity, normalization, monotone convergence, Jensen and Hölder inequalities, etc.

**Theorem 1.1** (Transfer). Being measurable with respect to the \( \sigma \)-algebra generated by a random variable is equivalent to being a measurable function of that random variable. More precisely, if \( T : \Omega \rightarrow (F, \mathcal{F}) \) and \( Y : \Omega \rightarrow (\mathbb{R}, \mathcal{B}_\mathbb{R}) \) and random variables then \( Y \) is \( \sigma(T) \) measurable if and only if there exists \( g : (F, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}_\mathbb{R}) \) measurable such that \( Y = g \circ T. \)
Proof. If \( Y = 1_A \) for some \( A \in \sigma(T) \). We have \( A = T^{-1}(B) \) for some \( B \in \mathcal{F} \), and therefore \( Y = 1_B \circ T \). If \( Y = \sum_{i \in I} a_i 1_{A_i} \) with \( I \) finite and \( A_i = T^{-1}(B_i) \), then \( Y = (\sum_{i \in I} a_i 1_{B_i}) \circ T \). The property is thus satisfied when \( Y \) is a step function. Now, if \( Y \) is non-negative and \( \sigma(T) \) measurable, then there exists a sequence \((Y_n)_n\) of step functions, \( \sigma(T) \) measurable, such that \( Y_n \not\to Y \), and \( Y_n = g_n \circ T \). By setting \( g = \lim g_n \), we get \( Y = g \circ T \). Finally, if \( Y \) is just \( \sigma(T) \) measurable, then it suffices to write \( Y = Y_+ - Y_- \). \( \blacksquare \)

Let \( X \in L^1(\Omega, \mathcal{A}, \mathbb{P}) \) and let \( T : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}) \) be a random variable. The conditional expectation of \( X \) given \( T \), denoted \( \mathbb{E}(X \mid T) \), is defined by

\[
\mathbb{E}(X \mid T) = \mathbb{E}(X \mid \sigma(T)).
\]

It is characterized by the following properties:

1. There exists \( g : (F, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}) \) with \( \mathbb{E}(X \mid T) = g(T) \) and \( g(T) \in L^1 \);

2. For all \( h : (F, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}) \) measurable and bounded,

\[
\mathbb{E}(Xh(T)) = \mathbb{E}(g(T)h(T)).
\]

If \( X \in L^2 \) then, thanks to the transfer theorem, the conditional expectation \( \mathbb{E}(X \mid T) \) is the best approximation in \( L^2 \) (least squares) of \( X \) by a measurable function of \( T \).

For a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \), an event \( A \in \mathcal{F} \), and a sub-\( \sigma \)-algebra \( \mathcal{A} \subset \mathcal{F} \), the quantity \( \mathbb{P}(A \mid \mathcal{A}) = \mathbb{E}(1_A \mid \mathcal{A}) \) is a random variable taking its values in \([0, 1]\). Similarly, conditioning with respect to an event makes sense in the sense that \( \mathbb{E}(X \mid A) = \mathbb{E}(X \mid 1_A = 1) \)

\[
\mathbb{E}(X \mid 1_A) = \frac{\mathbb{E}(X1_A)}{\mathbb{P}(A)} 1_A + \frac{\mathbb{E}(X1_{A^c})}{\mathbb{P}(A^c)} 1_{A^c}
\]

Finally, when \( X \) and \( Y \) take their values in an at most countable sets then

\[
\mathbb{E}(X \mid Y) = F(Y) \quad \text{where} \quad F(y) = \mathbb{E}(X \mid Y = y) = \sum_x \mathbb{P}(X = x \mid Y = y).
\]

Remark 1.2 (The condition expectation as an averaging of residual randomness). Let \( X \) and \( Y \) be random variables defined on a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \), and let \( \mathcal{A} \) be a sub-\( \sigma \)-algebra of \( \mathcal{F} \). If \( X \) is independent of \( \mathcal{A} \) and if \( Y \) is \( \mathcal{A} \)-measurable, then, using the monotone class theorem, for all \( \mathcal{F} \)-measurable and bounded or positive \( f : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \), we get

\[
\mathbb{E}(f(X, Y) \mid \mathcal{A}) = g(Y) \quad \text{where} \quad g(y) = \mathbb{E}(f(X, y) \mid \mathcal{F}).
\]

This suggests to interpret intuitively the conditional expectation as an averaging of residual randomness, and not only as the best approximation in the sense of least squares.

Let \( X \) and \( Y \) be two random variables taking values in the measurable spaces \((E, \mathcal{E})\) and \((F, \mathcal{F})\) respectively. The conditional law of \( X \) given \( Y \) is a family transition kernel \((N(y, \cdot))_{y \in F}\) of probability measures on \((E, \mathcal{E})\) such that for all \( A \in \mathcal{E} \), the map \( y \in F \to N(y, A) \in [0, 1] \) is measurable, and for all bounded (or positive) measurable test function \( h : E \to \mathbb{R} \),

\[
\mathbb{E}(h(X) \mid Y) = \int_E h(x)N(Y, \cdot).
\]

For all \( y \in F \), we also say that \( N(y, \cdot) \) is the conditional law of \( X \) given \( Y = y \), in other words

\[
\mathbb{E}(h(X) \mid Y = y) = \int_E h(x)N(y, \cdot).
\]

In particular \( \mathbb{P}(X \in A \mid Y) = N(Y, A) \) for all \( A \in \mathcal{E} \).
The random variables $X$ and $Y$ are independent if and only if $N(y, \cdot)$ does not depend on $y$ in the sense that for almost all $y \in F$, $N(y, \cdot) = P_X$ where $P_X$ is the law of $X$.

If $(X, Y)$ has density $f_{X,Y}(x, y)$ for the Lebesgue measure then $X$ and $Y$ have respective densities $f_X : x \mapsto \int f(x, y)dy$ and $f_Y : y \mapsto \int f(x, y)dx$ and the condition law $\text{Law}(X \mid Y = y)$ has density $f_{X \mid Y = y} : x \mapsto f_{X,Y}(x, y)\big/f_Y(y)$, in such a way that

$$f_{X,Y}(x, y) = f_{X \mid Y = y}(x)f_Y(y) = f_X(x)f_{Y \mid X = x}(y).$$

### 1.5 Gaussian random vectors

**Proof.**

1. Now $F$ is of bounded variation and is differentiable with integrable derivative then $|f'|ig/|f| = 1$-var, which is bounded (the boundedness of $|f|$ plays a role here).

### 1.6 Bounded variation and Lebesgue–Stieltjes integral

**Definition 1.3** ($p$-variation of a function on a finite interval). Let $[a, b] \subset \mathbb{R}$ be a finite interval, $−\infty < a < b < +\infty$. For all $p \in [1, \infty)$, the $p$-variation of a function $f : [a, b] \to \mathbb{R}$ is defined by

$$\|f\|_{p-\text{var}} = \left(\sup_{k} \sum_{k} \left|f(t_{k+1}) - f(t_{k})\right|^p\right)^{1/p} \in [0, +\infty]$$

where the supremum runs over all finite partitions or sub-divisions of the interval $I$ namely the finite sequences $(t_k)_{0 \leq k \leq n}$ in $[a, b]$ such that $n \geq 0$ and $a = t_0 < \cdots < t_{n+1} = b$.

- $\|f\|_{1-\text{var}}$ is called sometimes the total variation of $f$;
- if $f : [a, b] \to S$ has finite 1-variation, we say that $f$ has finite variation or is of bounded variation;
- if $f : [a, b] \to \mathbb{R}$ is of bounded variation then $f$ is bounded (the boundedness of $[a, b]$ plays a role here).
- if $f : [a, b] \to \mathbb{R}$ is of bounded variation and is differentiable with integrable derivative then

$$\|f\|_{1-\text{var}} = \int_a^b |f'(t)|dt.$$

- if $f$ is continuously differentiable then $f$ has bounded variation and the latter holds true.

**Theorem 1.4** (Representation of bounded variation functions on a finite interval). Let $[a, b] \subset \mathbb{R}$ be a finite interval, $−\infty < a < b < +\infty$. For all $f : [a, b] \to \mathbb{R}$, the following properties are equivalent:

1. $f$ is of bounded variation;
2. $f$ is the difference of two positive bounded increasing functions.

Such a decomposition is not unique in general.

**Proof.** $1 \Rightarrow 2$. Let $f$ be a function of bounded variation on $[a, b]$. For all $t \in [a, b]$, let

$$F(t) = \sup_{\delta} \sum_{k=0}^{n-1} \left|f(t_{k+1}) - f(t_k)\right|$$

where the supremum runs over the set of partitions or sub-divisions $\delta : a = t_0 < \cdots < t_n = t$ of $[a, t]$, $n = n_\delta \geq 1$. Now $F$ is bounded, and increasing by definition. It suffices now to show that $G = F - f$ is increasing and bounded. But the boundedness of $G$ follows from the one of $f$ and $F$, while for all $t_1 < t_2$ in $[a, b]$,

$$G(t_2) - G(t_1) = F(t_2) - F(t_1) - f(t_2) + f(t_1) \geq 0$$
since \( F(t_1) + f(t_2) - f(t_1) \leq F(t_1) + |f(t_2) - f(t_1)| \leq F(t_2) \).

2 \( \Rightarrow \) 1. If \( f \) and \( g \) have bounded variation on \([a, b]\), then it is also the case for \( f - g \). On the other hand, if \( f \) is monotonic on \( I \) then for all sub-division \( a = t_0 < \cdots < t_n = b, n \geq 1 \), we have

\[
\sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)| = |f(b) - f(a) |
\]

and thus \( f \) is of bounded variation. \( \blacksquare \)

**Theorem 1.5** (Lebesgue–Stieltjes integral with respect to continuous finite variation integrators). Let \( [a, b] \subset \mathbb{R} \) be a finite interval, \(-\infty < a < b < +\infty\). Let \( f : [a, b] \to \mathbb{R} \) be right continuous and of bounded variation. Then there exists a unique finite signed Borel measure \( \mu_f \) on \(([a, b], \mathcal{B}_{[a, b]})\) such that

\[
\mu_f([a]) = 0, \quad \text{and for all } t \in [a, b], \quad \mu_f([a, t]) = f(t) - f(a).
\]

It is customary to denote \( d\mu_f = df \), also for all measurable \( g : [a, b] \to \mathbb{R} \), positive or in \( L^1(\mu_f) \), we have

\[
\int_a^b g(t)df(t) = \int_g d\mu_f.
\]

Moreover, for all bounded and measurable \( g : [a, b] \to \mathbb{R} \), and for all sequence \((\delta_n)_{n \geq 1}\) of partitions or sub-divisions of \([a, b]\), \( \delta_n : a = t_0^0 < \cdots < t_{m_n}^n = b, m_n \geq 1 \), with \( \lim_{n \to \infty} \max_k (t_{k+1}^n - t_k^n) = 0 \), we have

\[
\int_a^b g(t)df(t) = \lim_{n \to \infty} \sum_k g(t_k^n)(f(t_{k+1}^n) - f(t_k^n)).
\]

Furthermore, \( h : t \in [a, b] \mapsto h(t) = \int_a^t g(s)df(s) \) is continuous and of bounded variation, and \( \mu_h = \mu_f \) in other words \( dh(t) = g(t)df(t) \), in the sense that for all bounded and measurable \( k : [a, b] \to \mathbb{R} \),

\[
\int_a^b k(t)dh(t) = \int_a^b k(t)d\int_a^t g(s)df(s) = \int_a^b k(t)g(t)df(t).
\]

**Proof.** First part. Theorem 1.4 gives \( f = f_+ - f_- \) where \( f_\pm \) are positive, bounded, and increasing. This reduces the problem to the case where \( f \) is increasing and \( \mu_f \) is a positive Borel measure. In this case, the result follows from the Carathéodory extension theorem (Theorem 1.11). Note: \( \mu_f \) is unique even if \( f_\pm \) are not.

Second part. For all \( n \geq 1 \), set \( g^{(n)}(a) = g(a) \), and for all \( t \in [a, b] \), \( g^{(n)}(t) = g(t_k^n) \) if \( t \in [t_{k-1}^n, t_k^n] \) for some \( k \in \{0, \ldots, m_n - 1\} \). Then \( g^{(n)} \) is measurable and converges pointwise to \( g \). By dominated convergence,

\[
\sum_k g(t_k^n)(f(t_{k+1}^n) - f(t_k^n)) = \int g(n) d\mu_f \xrightarrow{n \to \infty} \int g d\mu_f = \int_a^b g(t)df(t).
\]

Third part. First of all, for all \( s \in [a, b] \), we have \( \mu_{f|_{[a, s]}} = \mu_f|_{[a, s]} \).

\[
\int_a^s g(t)df(t) = \int_g d\mu_f|_{[a, s]} = \int g 1_{[0, s]} d\mu_f.
\]

The continuity of \( h \) follows now by dominated convergence. For the 1-variation, we write

\[
\sum_k |h(t_{k+1}) - h(t_k)| \leq \sum_k |g(1_{[t_{k+1}, t_k]} d|\mu_f| = \int |g|d|\mu_f| < \infty.
\]

Finally, to prove the formula, it suffices to check it for \( k = 1_{[a, c]} \) for \( c \in [a, b] \). This writes \( \mu_h(c) - \mu_h(a) = \int_a^c g(t)d\mu_f(t) = h(c) - h(a) \), which is the definition of \( \mu_h \). Note that by construction we have \( h(a) = 0 \). \( \blacksquare \)
Remark 1.6 (Riemann–Stieltjes–Young integral). Following L.C. Young, it can be shown that if a couple of continuous functions \( f, g : [a, b] \to \mathbb{R} \) such that \( f \) has finite \( p \)-variation and \( g \) has finite \( q \)-variation, with \( 1/p + 1/q > 1 \), then the following Riemann–Stieltjes–Young integral is well defined:

\[
\int_a^b f(t)dg(t) = \lim_{n \to \infty} \sum_{k=0}^{m_n} f(t_{k}^{(n)})[g(t_{k+1}^{(n)}) - g(t_{k}^{(n)})],
\]

where \( (\delta_n)_{n \geq 1} \) is an arbitrary sequence of partitions of \([a, b] , \delta_n : a = t_0 < \cdots < t_{m_n} = b, m_n \geq 1\).

1.7 Monotone class theorem and Carathéodory extension theorem

The monotone class theorem is a sort of Stone–Weierstrass theorem of measure theory.

Definition 1.7 (\( \pi \)-systems and \( \lambda \)-systems).

- We say that \( \mathcal{C} \subset \mathcal{P}(\Omega) \) is a \( \pi \)-system when \( A \cap B \in \mathcal{C} \) for all \( A, B \in \mathcal{C} \);
- We say that \( \mathcal{I} \subset \mathcal{P}(\Omega) \) is a \( \lambda \)-system (or monotone class or Dynkin\(^a\) system) when
  \[ \cup_{n} A_n \in \mathcal{C} \text{ for all } (A_n)_n \text{ such that } A_n \subset A_{n+1} \text{ and } A_n \in \mathcal{C} \text{ for all } n; \]
  \[ A \setminus B \in \mathcal{C} \text{ for all } A, B \in \mathcal{C} \text{ such that } B \subset A. \]

\(^a\)Named after Eugene Dynkin (1924 – 2014), Soviet and American mathematician.

Basic examples of \( \pi \)-systems are given by the class of singletons \( \{x\} : x \in \mathbb{R} \cup \{\varnothing\} \), the class of product subsets \( \{A \times B : A, B \in \mathcal{P}(\Omega)\} \), and the class of intervals \( \{(-\infty, x] : x \in \mathbb{R}\} \).

A basic yet important example of \( \lambda \)-system is given by \( \{A \in \mathcal{A} : \mathcal{P}(A) = \mathcal{Q}(A)\} \) where \( \mathcal{P} \) and \( \mathcal{Q} \) are probability measures on \( (\Omega, \mathcal{A}) \), see Corollary 1.10 for an application.

Lemma 1.8 (\( \sigma \)-fields). A \( \lambda \)-system that contains \( \Omega \) and which is a \( \pi \)-system is a \( \sigma \)-algebra.

Note that conversely, a \( \sigma \)-algebra is always a \( \pi \)-system, but not a \( \lambda \)-system in general.

Proof. If a \( \lambda \)-system \( \mathcal{I} \subset \mathcal{P}(\Omega) \) contains \( \Omega \) and is a \( \pi \)-system then for all \( A, B \in \mathcal{I} \) we have

\[
A \cup B = \Omega \setminus ((\Omega \setminus A) \cap (\Omega \setminus B)),
\]

which means that \( \mathcal{I} \) is table by finite union. This allows to drop the non-decreasing condition in the stability of \( \mathcal{I} \) by countable union, which simply means finally that \( \mathcal{I} \) is a \( \sigma \)-algebra.

\[ \square \]

Theorem 1.9 (Dynkin \( \pi \)-\( \lambda \) Theorem). If \( \mathcal{I} \subset \mathcal{P}(\Omega) \) is a \( \lambda \)-system containing \( \Omega \) and including a \( \pi \)-system \( \mathcal{C} \), then \( \mathcal{I} \) contains also the \( \sigma \)-algebra \( \sigma(\mathcal{C}) \) generated by \( \mathcal{C} \).

Proof. The \( \lambda \)-system generated by a subset of \( \mathcal{P}(\Omega) \) is by definition the intersection of all \( \lambda \)-systems which include this subset. This makes sense, indeed this intersection is not empty since it contains \( \mathcal{P}(\Omega) \), and it is not difficult to check that it is a \( \lambda \)-system. It is the smallest (for the inclusion) \( \lambda \)-system containing the initial subset of \( \mathcal{P}(\Omega) \).

Let \( \mathcal{I}' \) be the \( \lambda \)-system generated by \( \mathcal{C} \) and \( \Omega \). It suffices to show that \( \mathcal{I}' \) is a \( \sigma \)-algebra. For that, and thanks to lemma 1.8, it suffices to show that \( \mathcal{I}' \) is a \( \pi \)-system. To do so, let us define

\[
\mathcal{I}_1 = \{ A \in \mathcal{I}' : A \cap B \in \mathcal{I} \text{ for all } B \in \mathcal{C} \},
\]

which is a \( \lambda \)-system including \( \Omega \) and containing \( \mathcal{C} \), hence \( \mathcal{I}_1 \subset \mathcal{I}' \), and thus \( \mathcal{I}_1 = \mathcal{I}' \). Now,

\[
\mathcal{I}_2 = \{ A \in \mathcal{I}' : A \cap B \in \mathcal{I} \text{ for all } B \in \mathcal{C} \}
\]

is a \( \lambda \)-system containing \( \Omega \) and including \( \mathcal{I} \) and thus \( \mathcal{I}_2 = \mathcal{I}' \), hence \( \mathcal{I}' \) is a \( \pi \)-system.

\[ \square \]
Corollary 1.10 (Sierpiński—a–Dynkin monotone class theorem).

1. For all probability measures \( P \) and \( Q \) on a measurable space \((\Omega, \mathcal{A})\), if \( P(A) = Q(A) \) for all \( A \in \mathcal{C} \) where \( \mathcal{C} \) is a \( \pi \)-system such that \( \sigma(\mathcal{C}) = \mathcal{A} \), then \( P = Q \);

2. Let \( H \) be a vector space of bounded measurable functions \((\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})\) such that
   
   a. \( H \) is stable by monotone convergence: if \( f_n \in H \) \( \uparrow \) \( f \) pointwise with \( f \) bounded then \( f \in H \);
   
   b. \( H \) contains constant functions i.e. \( 1_\Omega \in H \), is stable by product i.e. if \( f, g \in H \) then \( fg \in H \), and contains all \( 1_A \) for all \( A \) in a \( \pi \)-system \( \mathcal{C} \) on \( \Omega \) such that \( \sigma(\mathcal{C}) = \mathcal{A} \);

   then \( H \) contains all \( \mathcal{A} \)-measurable bounded functions \( \Omega \to \mathbb{R} \).

\[ {}^a \text{Named after Wacław Sierpiński (1882 – 1969), Polish mathematician.} \]

Note that \( H \) is an algebra in the sense that it is a vector space stable by product.

The second statement can be seen as some sort of Stone–Weierstrass theorem of measure theory. It is useful in applications and is known under the name functional monotone class theorem.

**Proof.**

1. Take \( \mathcal{F} = \{ A \in \mathcal{A} : P(A) = Q(A) \} \) and use Theorem 1.9.

2. Take \( \mathcal{F} = \{ A \in \mathcal{A} : 1_A \in H \} \) and use Theorem 1.9.

**Exercise!**

Theorem 1.11 (Carathéodory extension theorem). Let \( \Omega \) be a non-empty set, let \( \mathcal{A} \) be a family of subsets of \( \Omega \), and let \( \mu : \mathcal{A} \to \mathbb{R}_+ \). Let \( \sigma(A) \) be the \( \sigma \)-algebra generated by \( \mathcal{A} \). If

1. \( \Omega \in \mathcal{A} \);

2. (stability by complement) for all \( A \in \mathcal{A} \), we have \( A^c = \Omega \setminus A \in \mathcal{A} \);

3. (stability by intersection) for all \( A, B \in \mathcal{A} \), we have \( A \cap B \in \mathcal{A} \);

4. \( \mu \) is \( \sigma \)-additive and \( \sigma \)-finite;

then there exists a unique \( \sigma \)-additive measure \( \mu_{\text{ext}} \) on \((\Omega, \sigma(A))\) such that \( \mu_{\text{ext}} = \mu \) on \( \mathcal{A} \).

**Proof.** FIXME:
Chapter 2

Processes, filtrations, stopping times, martingales

A stochastic process or process is a family of random variables $X = (X_t)_{t \geq 0}$, indexed by a parameter $t \in \mathbb{R}_+$ interpreted as a time, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and taking values in some measurable space $(G, \mathcal{B})$. By default a process takes real values. In general $G$ is a metric space, with distance denoted $d$, complete, separable, and $\mathcal{B}$ is its Borel $\sigma$-algebra.

2.1 Measurability

The natural filtration of a process $(X_t)_{t \geq 0}$ is the increasing family $(\mathcal{F}_t)_{t \geq 0}$ of sub-$\sigma$-algebras of $\mathcal{F}$ defined for all $t \geq 0$ by $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t)$. More generally, an increasing family $(\mathcal{F}_t)_{t \geq 0}$ of sub-$\sigma$-algebras of $\mathcal{F}$ is called a filtration. For a given filtration $(\mathcal{F}_t)_{t \geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, we say that the process $X$ is...

- real when $G = \mathbb{R}$ in other words $X$ takes real values (this is the default in this course);
- $d$-dimensional when $G = \mathbb{R}^d$ in other words $X$ takes its values in $\mathbb{R}^d$, $d \geq 1$;
- issued from the origin when $X_0 = 0$ (makes sense when $G$ is a vector space);
- adapted when $X_t$ is $\mathcal{F}_t$ measurable for all $t \geq 0$;
- measurable when $(s, \omega) \in [0, t] \times \Omega \mapsto X_s(\omega)$ is $\mathcal{B}_{[0,t]} \otimes \mathcal{F}$ measurable for all $t \geq 0$;
- progressive if for all $t > 0$, $(s, \omega) \in [0, t] \times \Omega \mapsto X_s(\omega)$ is $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$ measurable;
- right-continuous (respectively left-continuous, respectively continuous) when for almost all $\omega \in \Omega$, the sample path $t \in \mathbb{R}_+ \mapsto X_t(\omega) \in G$ is right-continuous (respectively left-continuous, respectively continuous);
- square integrable when $\mathbb{E}(X_t^2) < \infty$ for all $t \geq 0$;
- bounded in $L^p$, $p \geq 1$, when $\sup_{t \geq 0} \mathbb{E}(|X_t|^p) < \infty$;
- bounded when there exists a finite $C > 0$ such that almost surely, $\sup_{t \geq 0} |X_t| \leq C$;
- of finite variation when almost surely $t \mapsto X_t$ is of bounded variation on all finite intervals of $\mathbb{R}_+$, equivalently is difference of two positive increasing processes, see Theorem 1.4.

Theorem 2.1 (Progressive $\sigma$-field and progressive processes).

1. The family $\mathcal{P}$ of all $\Lambda \in \mathcal{F} \otimes \mathcal{B}_{\mathbb{R}_+}$ such that the process $X_t(\omega) = 1_{(\omega, t) \in \Lambda}$ is progressive is a $\sigma$-field on $\Omega \times \mathbb{R}_+$ called the progressive $\sigma$-field. For all $\Lambda \subset \Omega \times \mathbb{R}_+$, we have $\Lambda \in \mathcal{P}$ if and only if for all $t \geq 0$, $\Lambda \cap (\Omega \times [0, t]) \in \mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$. A process $X = (X_t)_{t \geq 0}$ is progressive if and only if it is, as a random variable $X : (\omega, t) \in \Omega \times \mathbb{R}_+ \mapsto X_t(\omega)$, measurable for the progressive $\sigma$-field $\mathcal{P}$ on $\Omega \times \mathbb{R}_+$;

2. If $(X_t)_{t \geq 0}$ is an adapted right-continuous (or left-continuous) process defined on a filtered proba-
bility space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) and taking its values in a metric space \((E, d)\) equipped with its Borel \(\sigma\)-algebra, then this process is progressive.

**Proof.**

1. Exercise;

2. We give the proof in the right-continuous case, the left-continuous case being entirely similar. For all \(n \geq 1, t > 0, s \in [0, t]\), we define the random variable

\[
X^n_s = \begin{cases} 
X^n_{kt/n} & \text{if } s \in [(k - 1)t/n, kt/n), 1 \leq k \leq n, \\
X_t & \text{if } s = t.
\end{cases}
\]

Since \((X_t)_{t \geq 0}\) is right-continuous, it follows that \(X_s(\omega) = \lim_{n \to \infty} X^n_s(\omega)\) for all \(t > 0, s \in [0, t]\) and all \(\omega \in \Omega\). On the other hand, for every Borel subset \(A \subseteq E\),

\[
\{ (\omega, s) \in \Omega \times [0, t] : X^n_s(\omega) \in A \} = \bigcup_{k=1}^{n} \{ (X_{kt/n} \in A) \times \{(k - 1)t/n, kt/n)\}.
\]

Since \((X_t)_{t \geq 0}\) is adapted, this set belongs to \(\mathcal{F}_t \otimes \mathcal{B}_{[0, t]}\). Therefore, for all \(n \geq 1\), the function \((\omega, s) \in \Omega \times [0, t] \mapsto X^n_s(\omega)\) is measurable for \(\mathcal{F}_t \otimes \mathcal{B}_{[0, t]}\). Now a pointwise limit of measurable functions is measurable, and therefore the function \((\omega, s) \in \Omega \times [0, t] \mapsto X_s(\omega)\) is also measurable for \(\mathcal{F}_t \otimes \mathcal{B}_{[0, t]}\), which means, since \(t > 0\) is arbitrary, that \((X_t)_{t \geq 0}\) is progressively measurable.

A process \(X = (X_t)_{t \geq 0}\) taking its values in \(\mathbb{R}^d\) can be seen as a random variable taking its values in the "path space" \(\mathcal{P}(\mathbb{R}_+, \mathbb{R}^d)\) of functions from \(\mathbb{R}_+\) to \(\mathbb{R}^d\). The measurability is for free if we equip \(\mathcal{P}(\mathbb{R}_+, \mathbb{R}^d)\) with the \(\sigma\)-algebra \(\mathcal{A}(\mathbb{R}_+, \mathbb{R}^d)\) generated by the cylindrical events

\[
\{ f \in \mathcal{P}(\mathbb{R}_+, \mathbb{R}^d) : f(t_1) \in I_1, \ldots, f(t_n) \in I_n \}
\]

where \(n \geq 1, t_1, \ldots, t_n \in \mathbb{R}_+, \) and where \(I_1, \ldots, I_n\) are products of intervals in \(\mathbb{R}^d\) of the form \(\prod_{i=1}^{d} (a_i, b_i]\). Unfortunately \(\mathcal{P}(\mathbb{R}_+, \mathbb{R}^d)\) is so big that \(\mathcal{A}(\mathbb{R}_+, \mathbb{R}^d)\) turns out to be too small, and does not contain for instance events of interest such that \(\{ f \in \mathcal{P}(\mathbb{R}_+, \mathbb{R}^d) : \sup_{t \in [0,1]} f(t) < 1 \}\).

In this course, we focus essentially, for simplicity, on continuous processes. This suggests to consider the space \(C(\mathbb{R}_+, \mathbb{R}^d)\) and the \(\sigma\)-algebra \(\mathcal{A}(C(\mathbb{R}_+, \mathbb{R}^d))\) generated by the cylindrical events \(\{ f \in C(\mathbb{R}_+, \mathbb{R}^d) : f(t_1) \in I_1, \ldots, f(t_n) \in I_n \}\) where \(n \geq 1, t_1, \ldots, t_n \in \mathbb{R}_+\), \(n \geq 1\), \(I_1, \ldots, I_n\) are products of intervals in \(\mathbb{R}^d\) of the form \(\prod_{i=1}^{d} (a_i, b_i]\). We have then the following:

**Theorem 2.2** (What a wonderful world). On \(C(\mathbb{R}_+, \mathbb{R}^d)\), the \(\sigma\)-algebra \(\mathcal{A}(C(\mathbb{R}_+, \mathbb{R}^d))\) generated by the cylindrical events coincides with the Borel \(\sigma\)-algebra \(B (C(\mathbb{R}_+, \mathbb{R}^d))\) generated by the open sets of the topology of uniform convergence on compact intervals of \(\mathbb{R}_+\).

**Proof.** Take \(d = 1\) for simplicity. It can be shown that \(C(\mathbb{R}_+, \mathbb{R})\) equipped with the distance

\[
d(f, g) = \sum_{n=1}^{\infty} 2^{-n}(1 \land \max_{t \in [0,1]} |f(t) - g(t)|)
\]

is a Polish space in other words a complete and separate metric space, and the associated topology is the one of uniform convergence on compact subsets of \(\mathbb{R}_+\). First we have the inclusion \(\mathcal{A}(C(\mathbb{R}_+, \mathbb{R})) \subset B (C(\mathbb{R}_+, \mathbb{R}))\) since the \(\sigma\)-algebra \(\mathcal{A}(C(\mathbb{R}_+, \mathbb{R}))\) is generated by the cylinders

\[
\{ f \in C(\mathbb{R}_+, \mathbb{R}) : f(t_i) < a_i, \ldots, f(t_n) < a_n \}, \quad n \geq 1, t_1, \ldots, t_n \in \mathbb{R}_+, a_1, \ldots, a_n \in \mathbb{R},
\]

which are open subsets. Conversely, for all \(g \in C(\mathbb{R}_+, \mathbb{R})\), all \(n \geq 1\), and all \(r > 0\),

\[
\{ f \in C(\mathbb{R}_+, \mathbb{R}) : \max_{t \in [0,1]} |f(t) - g(t)| \leq r \} = \bigcap_{n=0}^{\infty} \{ f \in C(\mathbb{R}_+, \mathbb{R}) : |f(t) - g(t)| \leq r \}
\]

belongs to \(\mathcal{A}(C(\mathbb{R}_+, \mathbb{R}))\), and since these sets generate \(B (C(\mathbb{R}_+, \mathbb{R}))\), we get \(\mathcal{A}(C(\mathbb{R}_+, \mathbb{R})) = B (C(\mathbb{R}_+, \mathbb{R}))\).
2.1 Measurability

**Theorem 2.3** (Continuous processes as random variables on the path space). Let \( X = (X_t)_{t \geq 0} \) be a continuous \( d \)-dimensional process defined a probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Let \( \Omega' \in \mathcal{F} \) be such that \( \mathbb{P}(\Omega') = 1 \) and \( X_* \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \) on \( \Omega' \), and let \( \sigma' = \{ A \cap \Omega' : A \in \mathcal{F} \} \). Then \( X' : \omega \in \Omega' \rightarrow X_*(\omega) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \) is measurable with respect to \( \sigma' \) and \( \mathcal{B}_c(\mathbb{R}_+, \mathbb{R}^d) \).

**Proof.** Let us consider an arbitrary cylindrical event
\[
A = \{ f \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) : f(t_1) \in I_1, \ldots, f(t_n) \in I_n \},
\]
where \( n \geq 1, t_1, \ldots, t_n \in \mathbb{R}_+, \) and \( I_1, \ldots, I_n \) are product of intervals as \( \prod_{i=1}^d (a_i, b_i] \). Then
\[
\Omega' \cap \{ X_* \in A \} = \Omega' \cap \{ X_{t_1} \in I_1, \ldots, X_{t_n} \in I_n \} \in \sigma'.
\]
Now \( \mathcal{B}_c(\mathbb{R}_+, \mathbb{R}^d) \) is generated by cylindrical events.

**Remark 2.4** (Equality of processes, modification of a process, indistinguishable processes). We say that two processes \( X = (X_t)_{t \geq 0} \) and \( Y = (Y_t)_{t \geq 0} \) defined on the same probability space \((\Omega, \mathcal{F}, \mathbb{P})\) are indistinguishable when for almost all \( \omega \in \Omega \) the sample paths \( t \rightarrow X_t(\omega) \) and \( t \rightarrow Y_t(\omega) \) coincide in other words
\[
\mathbb{P}(\forall t \geq 0 : X_t = Y_t) = 1.
\]

There is a weaker notion in which the almost sure event depends on time, namely we say that \( Y \) is a modification of \( X \) if for all \( t \geq 0 \) the event \( \Omega_t = \{ \omega \in \Omega : X_t(\omega) \neq Y_t(\omega) \} \) is negligible, in other words
\[
\forall t \geq 0 : \mathbb{P}(X_t = Y_t) = 1.
\]

If \( X \) and \( Y \) are continuous then the two notions of indistinguishable and modification coincide.

If \( X = (X_t)_{t \geq 0} \) and \( Y = (Y_t)_{t \geq 0} \) are two processes taking values in \( \mathbb{R}^d \) and if they have some finite dimensional marginal distributions, in the sense that for all \( n \geq 1 \) and all \( t_1, \ldots, t_n \in \mathbb{R}_+ \), the random variables \( (X_{t_1}, \ldots, X_{t_n}) \) and \( (Y_{t_1}, \ldots, Y_{t_n}) \) have same law in \( (\mathbb{R}^d)^n \), then the processes \( X \) and \( Y \) have same law as random variables on the path space \( \mathcal{P}(\mathbb{R}_+, \mathbb{R}), \mathcal{A}_{\mathcal{P}(\mathbb{R}_+, \mathbb{R})} \). The following theorem provides a sort of converse, stated in the case \( d = 1 \) for simplicity.

**Theorem 2.5** (Kolmogorov extension theorem). For all \( n \geq 1 \) and all \( t_1, \ldots, t_n \in \mathbb{R}_+ \), let \( \mu_{t_1, \ldots, t_n} \) be a probability measure on \( \mathbb{R}^n \). Let us assume the following consistency conditions:

1. for all \( n \geq 1, t_1, \ldots, t_n \in \mathbb{R}, \) and all permutation \( \sigma \) of \( \{1, \ldots, n\}, \) we have
\[
\mu_{t_1, \ldots, t_n} = \mu_{t_{\sigma(1)}, \ldots, t_{\sigma(n)}},
\]

2. for all \( n \geq 1, t_1, \ldots, t_n \in \mathbb{R}, A_1, \ldots, A_{n-1} \in \mathcal{B}_c, \)
\[
\mu_{t_1, \ldots, t_n}(A_1 \times \cdots \times A_{n-1} \times \mathbb{R}) = \mu_{t_1, \ldots, t_{n-1}}(A_1 \times \cdots \times A_{n-1}).
\]

Then there exists a unique probability measure \( \mu \) on the path space \( \mathcal{P}(\mathbb{R}_+, \mathbb{R}), \mathcal{A}_{\mathcal{P}(\mathbb{R}_+, \mathbb{R})} \) such that for all \( n \geq 1 \), all \( t_1, \ldots, t_n \in \mathbb{R}_+, \) and all \( A_1, \ldots, A_n \in \mathcal{B}_c, \)
\[
\mu(\pi_{t_1} \in A_1, \ldots, \pi_{t_n} \in A_n) = \mu_{t_1, \ldots, t_n}(A_1 \times \cdots \times A_n),
\]
where \( \pi_t : \omega \in \mathcal{P}(\mathbb{R}_+, \mathbb{R}) \rightarrow \omega_t \in \mathbb{R} \) for all \( t \geq 0. \)

**Proof.** For a cylindrical event \( A_{t_1, \ldots, t_n}(B) = \{ f \in \mathcal{P}(\mathbb{R}_+, \mathbb{R}) : (f(t_1), \ldots, f(t_n)) \in B \} \) where \( n \geq 1, t_1, \ldots, t_n \in \mathbb{R}^n, \) and where \( B \in \mathcal{B}_{\mathbb{R}^n}, \) we define \( \mu(B) = \mu_{t_1, \ldots, t_n}(B) \). This makes sense thanks to the consistency conditions. Moreover \( \mu(\mathcal{P}(\mathbb{R}_+, \mathbb{R})) = 1. \) Since the set of cylinders satisfies the assumptions of the Carathéodory extension theorem (Theorem 1.11), and generates the \( \sigma \)-algebra \( \mathcal{A}_{\mathcal{P}(\mathbb{R}_+, \mathbb{R})} \), it remains to show that \( \mu \) is a \( \sigma \)-finite measure, which is the difficult part of the proof. FIXME: See instance [Bau14].

15/98
2 Processes, filtrations, stopping times, martingales

2.2 Stopping times

A map $T : \Omega \to [0, +\infty]$ is a stopping time (respectively optional time) for a filtration $\left(\mathcal{F}_t\right)_{t \geq 0}$ on $(\Omega, \mathcal{F}, P)$ when for all $t \geq 0$ we have $\{T \leq t\} \in \mathcal{F}_t$ (respectively $\{T < t\} \in \mathcal{F}_t$). All stopping times are optional times, and the two notions coincide if the filtration is right-continuous:

$$\text{for all } t \geq 0, \quad \mathcal{F}_t = \mathcal{F}_t^{T}, \quad \text{where } \mathcal{F}_t^{T} = \cap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}.$$

Note that if $T : \Omega \to [0, +\infty]$ is constant then it is a stopping time.

**Theorem 2.6** (Archetypal examples of stopping times). If $X = (X_t)_{t \geq 0}$ is a continuous and adapted process taking its values in a metric space $(G, d)$, then, for all closed subset $A \subset G$, the hitting time $T_A = \inf\{t \geq 0 : X_t \in A\} : \Omega \to [0, +\infty]$ is\^a a stopping time.

\vspace{0.5cm}

**Proof.** For all $t \geq 0$, we have, denoting $A_\varepsilon = \{x \in G : d(x, A) < \varepsilon\}$,

$$\{T_A \leq t\}^c = \{T_A > t\} = \bigcup_{\varepsilon \in \mathbb{Q} \cap [0, \infty)} \bigcap_{r \in \mathbb{Q} \cap [0, t)} \{X_r \not\in A_\varepsilon\}.$$ 

\vspace{0.5cm}

More generally, it can be shown that if $\left(\mathcal{F}_t\right)_{t \geq 0}$ is right-continuous and complete in the sense that $\mathcal{F}_0$ contains all the negligible subsets of $\mathcal{F}$, and if the process $X = (X_t)_{t \geq 0}$ is adapted and \textit{right-continuous}\footnote{This remains valid more generally when $X$ is measurable as a random variable on $\Omega \times \mathbb{R}$, with respect to the $\sigma$-algebra generated by all right-continuous adapted processes. Such a process is called an \textit{optional process}.} then for all Borel set $A$, $T_A$ is a stopping time, see for instance [DM88, DM82, DM78].

**Theorem 2.7** (Stopping times properties). Let $S$, $T$, and $T_n$, $n \geq 0$ be stopping times for some underlying filtration $\left(\mathcal{F}_t\right)_{t \geq 0}$ on an underlying probability space $(\Omega, \mathcal{F}, P)$. Then:

1. the family $\mathcal{F}_T = \{A \in \mathcal{F} : \forall t \geq 0, A \cap \{T \leq t\} \in \mathcal{F}_t\}$ is a $\sigma$-algebra, called the \textit{stopping $\sigma$-algebra}, and the stopping time $T$ is $\mathcal{F}_T$-measurable;

2. $X = (X_t)_{t \geq 0}$ is adapted then the \textit{stopped process} $X^T = (X_{t\wedge T})_{t \geq 0}$ is also adapted. Moreover

$$\left(X^T\right)^S = X^{S\wedge T} = (X^S)^T;$$

3. if $(X_t)_{t \geq 0}$ is adapted and progressively measurable and if $T$ is almost surely finite then the \textit{stopped process} $X^T = (X_{t\wedge T})_{t \geq 0}$ is progressively measurable;

4. if $X = (X_t)_{t \geq 0}$ is adapted and right-continuous then $Z = X_T 1_{T < \infty}$ is $\mathcal{F}_T$-measurable;

5. if $S \leq T$ then $\mathcal{F}_S \subset \mathcal{F}_T$;

6. $S \wedge T$ and $S \vee T$ are stopping times and in particular $\mathcal{F}_{S\wedge T} \subset \mathcal{F}_{S\vee T}$;

7. if $\left(\mathcal{F}_t\right)_{t \geq 0}$ is right-continuous then $\lim_n T_n$ and $\varlimsup_n T_n$ are stopping times and

$$\cap_n \mathcal{F}_{T_n} = \mathcal{F}_{\lim_n T_n},$$

$$\cap_n \mathcal{F}_{T_n} = \mathcal{F}_{\inf_n T_n}.$$
4. Let \( B \in \mathcal{B}_\mathbb{R} \) and \( t \geq 0 \). Then we have:
\[
\{ Z \in B \} \cap \{ T \leq t \} = \{ X_{t_n} \in B \} \cap \{ T \leq t \}.
\]

Now we consider the composition of measurable maps:
\[
\omega \in (\Omega, \mathcal{F}_t) \mapsto (\sigma(\omega) \land t, \omega) \in ([0, t] \times \Omega, \mathcal{B}_{[0,1]} \otimes \mathcal{F}_t) \mapsto X_{\sigma(\omega) \land t}(\omega) \in (\mathbb{R}, \mathcal{B}_\mathbb{R})
\]
and we use the fact that \( X \) is progressively measurable.

5. If \( A \in \mathcal{F}_S \) then, for all \( t \geq 0 \), \( A \cap \{ T \leq t \} = A \cap \{ S \leq t \} \cap \{ T \leq t \} \in \mathcal{F}_t \), hence \( A \in \mathcal{F}_T \).

6. For all \( t \geq 0 \) we have
\[
\{ S \land T > t \} = \{ S > t \} \cap \{ T > t \} \in \mathcal{F}_t \quad \text{and} \quad \{ S \lor T \leq t \} = \{ S \leq t \} \cap \{ T \leq t \} \in \mathcal{F}_t.
\]

7. It suffice to show that \( \sup_n T_n \) and \( \inf_n T_n \) are stopping times. But
\[
\{ \sup_n T_n \leq t \} = \cap_n \{ T_n \leq t \} \in \mathcal{F}_t \quad \text{and} \quad \{ \inf_n T_n < t \} = \cup_n \{ T_n < t \} \in \mathcal{F}_t
\]
and therefore
\[
\{ \inf_n T_n \leq t \} = \cap_{\epsilon > 0} \{ \inf_n T_n < t + \epsilon \} \in \mathcal{F}_{t+} = \mathcal{F}_t.
\]

Remark 2.8 (Subtleties). The natural filtration of a right-continuous process is not right-continuous in general\(^a\), indeed a counter example is given by \( X_t = tZ \) for all \( t \geq 0 \) where \( Z \) is a non-constant random variable, since \( \sigma(X_0) = \{ \emptyset, \Omega \} \) while \( \sigma(X_{0+}: \epsilon > 0) = \sigma(Z) \neq \sigma(X_0) \). It is customary to assume that the underlying filtration is right-continuous and complete. For a given filtration \( (\mathcal{F}_t)_{t \geq 0} \), it is always possible to consider its completion \( (\sigma_t)_{t \geq 0} = (\sigma(\mathcal{N} \cup \mathcal{F}_t))_{t \geq 0} \) where \( \mathcal{N} \) are the negligible subsets of \( \mathcal{F} \), and to consider the right-continuous version \( (\sigma_{t+})_{t \geq 0} \), called the canonical filtration. A process is always adapted with respect to the canonical filtration constructed from its natural filtration.

\(^a\)However it can be shown that the completion of the natural filtration of a "Feller Markov process" – including all Lévy processes and in particular Brownian motion – is always right-continuous.

2.3 Quadratic variation

**Definition 2.9** (Quadratic variation). Let \( X = (X_t)_{t \geq 0} \) be a square integrable real process such that \( X_0 = 0 \). The **quadratic variation process** \( [X] = ([X]_t)_{t \geq 0} \) of \( X \) is defined for all \( t \geq 0 \) by the limit (when it exists)
\[
[X]_t = \lim_{|\delta| \to 0} \sum_k (X_{t_{k+1}} - X_{t_k})^2
\]
where the convergence takes place in probability, and where
\[
\delta : 0 = t_0 < t_1 < \cdots < t_n = t, \ n = n_\delta \geq 1,
\]
runs over all the partitions or sub-divisions of \([0, t]\), and where \(|\delta| = \max_{1 \leq k \leq n} |t_{k+1} - t_k|\) is the mesh of \( \delta \). More generally, the **quadratic covariation process** of a couple of square integrable real processes

---

17/98
2 Processes, filtrations, stopping times, martingales

\[ X = (X_t)_{t \geq 0} \text{ and } Y = (Y_t)_{t \geq 0} \text{ is defined for all } t \geq 0 \text{ by the following limit when it exists:} \]

\[ [X, Y]_t = \lim_{|\delta| \to 0} \mathbb{P} \sum_{k} (X_{t_{k+1}} - X_{t_k})(Y_{t_{k+1}} - Y_{t_k}). \]

We have \([X] = [X, X]\). The set of processes with quadratic variation is a vector space. The operator \([-\cdot -]\) is bilinear on this space and we have by polarization \([X, Y] = \frac{1}{2}([X + Y] - [X - Y])\).

Theorem 3.13 states that for Brownian motion \(B\), we have, for all \(t \geq 0\), \(|B|_t = t\). Theorem 2.15 states that for a square integrable continuous martingale \(M\) with \(M_0 = 0\), we have, for all \(t \geq 0\), \(\mathbb{E}(|M|_t) = \mathbb{E}(M^2_t)\).

**Lemma 2.10** (Continuity and finite variation implies zero quadratic variation). *If a process \(X = (X_t)_{t \geq 0}\) is continuous and has finite variation then it has zero quadratic variation.*

**Proof.** Indeed, for all \(t > 0\) and all partition \(\delta : 0 = t_0 < \cdots < t_n = t\) of \([0, t]\), \(n = n_\delta \geq 1\),

\[ \sum_{k} (X_{t_{k+1}} - X_{t_k})^2 \leq \max_{k} X_{t_{k+1}} - X_{t_k} \sum_{k} |X_{t_{k+1}} - X_{t_k}| \rightarrow 0. \]

The max part of the right hand side tends to zero since \(X\) is continuous, while the \(\sum\) part tends to the 1-variation of \(X\) on \([0, t]\) which is finite since \(X\) has finite variation.

**2.4 Martingales**

The notion of martingale implements the idea of conditionally independent updates. A stochastic real process \(X = (X_t)_{t \geq 0}\) on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) is a martingale (respectively super-martingale, respectively sub-martingale) with respect to \((\mathcal{F}_t)_{t \geq 0}\) when:

1. for all \(t \geq 0\), \(X_t\) is integrable;
2. for all \(t \geq 0\), \(X_t\) is \(\mathcal{F}_t\) measurable;
3. for all \(0 \leq s \leq t\), \(\mathbb{E}(X_t | \mathcal{F}_s) = X_s\) (respectively \(\leq\), respectively \(\geq\)).

In particular, a martingale has a conservation law: \(\mathbb{E}(X_t) = \mathbb{E}(X_0)\) for all \(t \geq 0\). Similarly, \(\mathbb{E}(X_t)\) grows with \(t\) for a sub-martingale, while \(\mathbb{E}(X_t)\) decreases with \(t\) for a super-martingale. Also these three notions can be seen in a sense as a probabilistic generalization of the notion of constant sequence, decreasing sequence, and increasing sequence in basic classical analysis.

Being a martingale for a given filtration is a property stable by linear combinations.

If \((M_t)_{t \geq 0}\) is a martingale and if \(\varphi : \mathbb{R} \to \mathbb{R}\) is convex and \(\varphi(M_t) \in L^1\) then the Jensen inequality for condition expectations implies that \((\varphi(M_t))_{t \geq 0}\) is a sub-martingale.

**Example 2.11** (Martingales).

1. If \(Y \in L^1\) then the process \((X_t)_{t \geq 0}\) defined by \(X_t = \mathbb{E}(Y | \mathcal{F}_t)\) for all \(t \geq 0\) is a martingale with respect to \((\mathcal{F}_t)_{t \geq 0}\) known as the Doob martingale or a closed martingale. It is uniformly integrable. Theorem 2.22 provides a sort of converse;
2. If \((X_t)_{t \geq 0}\) is a martingale and if \(\varphi : \mathbb{R} \to \mathbb{R}\) is convex and such that \(\varphi(X_t) \in L^1\) for all \(t \geq 0\), then \((Y_t)_{t \geq 0} = (\varphi(X_t))_{t \geq 0}\) is a sub-martingale for the same filtration. In particular \(|(X_t)_{t \geq 0}|, (X^2_t)_{t \geq 0}\) and \((e^{X_t})_{t \geq 0}\) are sub-martingales;
3. A martingale \(X = (X_t)_{t \geq 0}\) is also a martingale for its natural filtration \((\sigma(X_t))_{t \geq 0}\);
4. If \((E_n)_{n \geq 1}\) are independent and identically distributed exponential random variables of mean \(1/\lambda\), then, for all \(t \geq 0\), the number of these random variables falling in the interval \([0, t]\) is \(N_t = \text{card}(n \geq 1 : E_n \in [0, t])\). It is known that the counting process \((N_t)_{t \geq 0}\) has independent and stationary increments of Poisson law, namely for all \(n \geq 1\) and 0 = \(t_0 \leq \cdots <= t_n\), the random variables \(N_{t_0} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}\) are independent of law \(\text{Poi}(\lambda(t_1 - t_0)), \ldots, \text{Poi}(\lambda(t_n - t_{n-1}))\). We say that
(\(N_t\))_{t \geq 0} is the simple Poisson process of intensity \(\lambda\). Now for the filtration \((\mathcal{F}_t)_{t \geq 0}, \mathcal{F}_t = \sigma(N_s: 0 \leq s \leq t)\), and for all \(c \in \mathbb{R}\), the process \((N_t - ct)_{t \geq 0}\) is a sub-martingale if \(c < \lambda\), a martingale if \(c = \lambda\), and a super-martingale if \(c > \lambda\). Namely, for all \(0 \leq s \leq t\),

\[
\mathbb{E}(N_t - ct \mid \mathcal{F}_s) = \mathbb{E}(N_t - N_s - ct + N_s - cs \mid \mathcal{F}_s)
\]

\[
= \mathbb{E}(N_t - N_s - ct + N_s - cs)
\]

\[
= (\lambda - c)(t - s) + N_s - cs.
\]

This process is not continuous, but has right-continuous and left limited sample paths.

5. If \((N_t)_{t \geq 0}\) is the simple Poisson process of intensity \(\lambda\) as above, then, for all \(0 \leq s \leq t\),

\[
\mathbb{E}(e^{N_t - ct} \mid \mathcal{F}_s) = e^{N_t - cs} \mathbb{E}(e^{N_t - N_s}e^{-(t-s)}) = e^{N_s - cs}e^{\lambda(t-s)(e-1) - c(t-s)}.
\]

It follows that for the natural filtration of \((N_t)_{t \geq 0}\), the process \((e^{N_t - ct})_{t \geq 0}\) is a sub-martingale if \(c < \lambda(e - 1)\), a martingale if \(c = \lambda(e - 1)\), and a super-martingale if \(c > \lambda(e - 1)\). We often say that \((e^{N_t - ct})_{t \geq 0}\) is an exponential (sub/super-)martingale.

6. The Brownian motion \((B_t)_{t \geq 0}\) of Chapter 3 has independent and stationary Gaussian increments: for all \(n \geq 1\) and \(0 = t_0 \leq \cdots \leq t_n\) the random variables \(B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}\) are independent of law \(\mathcal{N}(0, t_1 - t_0), \ldots, \mathcal{N}(0, t_n - t_{n-1})\). Thus for all \(c \in \mathbb{R}\), \((B_t)_{t \geq 0}\) is a martingale for its natural filtration: for all \(0 \leq s \leq t\),

\[
\mathbb{E}(B_t \mid \mathcal{F}_s) = \mathbb{E}(B_t - B_s + B_s \mid \mathcal{F}_s) = \mathbb{E}(B_t - B_s) + B_s = B_s.
\]

This process has continuous sample paths. Moreover and similarly \((B^2_t - ct)_{t \geq 0}\) is a sub-martingale if \(c < 1\), a martingale if \(c = 1\), and a super-martingale if \(c > 1\). For simplicity, most of the martingales encountered in this course are continuous.

---

**Theorem 2.12** (Doob stopping theorem). Let \((M_t)_{t \geq 0}\) be a continuous \((\mathcal{F}_t)_{t \geq 0}\)-adapted process such that \(M_t \in L^1\) for all \(t \geq 0\). The following properties are equivalent:

1. the process \((M_t)_{t \geq 0}\) is a martingale;
2. \(\mathbb{E}(M_T) = \mathbb{E}(M_0)\) for all bounded stopping time \(T\) such that \(M_T \in L^1\).

Moreover, if \(M = (M_t)_{t \geq 0}\) is a continuous martingale and \(T\) is a stopping time, not necessarily bounded or finite, such that \(M\) and \(T\) belong \(L^1\) for all \(t \geq 0\), then \((M_{T \wedge t})_{t \geq 0}\) is a continuous martingale.

**Proof of Theorem 2.12.** Proof of 1 \(\Rightarrow\) 2. Let \((M_t)_{t \geq 0}\) be a continuous martingale with respect to \((\mathcal{F}_t)_{t \geq 0}\), and let \(T\) be a bounded stopping time, say \(T \leq C\) for a finite constant \(C\). If \(T\) takes a finite number of values \(t_1 < \cdots < t_n \leq C\), then by martingale and stopping time properties,

\[
\mathbb{E}(M_T) = \sum_{k=1}^n \mathbb{E}(M_{t_k}1_{T=t_k}) = \sum_{k=1}^n \mathbb{E}(M_{t_k} \mid \mathcal{F}_{t_k})1_{T=t_k} = \sum_{k=1}^n \mathbb{E}(M_{t_k}1_{T=t_k}) = \mathbb{E}(M_0).
\]

If \(T\) takes an infinite number of values, we proceed by discretization and approximation, namely

\[
T = \lim_{n \to \infty} T_n\quad \text{where} \quad T_n = C1_{T=C} + \sum_{k=1}^n \frac{k}{2n} C1_{\frac{k-1}{2n} \leq T < \frac{k}{2n} C}.
\]

and it remains to show that \(\lim_{n \to \infty} \mathbb{E}(M_{T_n}) = \mathbb{E}(M_T)\). But since \(\lim_{n \to \infty} M_{T_n} = M_T\) almost surely, it suffices to show that \((M_{T_n})_{n \geq 1}\) is uniformly integrable. Now, since \(T_n\) takes a finite number of values, the martingale property and the Jensen inequality give, for all \(R > 0\),

\[
\mathbb{E}(|M_C|1_{|M_{T_n}| \geq R}) = \sum_{k=1}^n \mathbb{E}(|M_{T_n}|1_{|M_{T_n}| \geq R}) \leq \mathbb{E}(|M_{T_n}|1_{|M_{T_n}| \geq R}).
\]
This gives, using the dominated convergence theorem,
\[
\lim_{R \to \infty} \mathbb{E}(\|M^R_t\| \mathbf{1}_{M^R_t \geq R}) \leq \lim_{R \to \infty} \mathbb{E}(\|M^R_t\| \mathbf{1}_{\sup_{s \leq 0 : t} |M_s| \geq R}) = 0,
\]
since \( M_R \in L^1 \) and since \((M_t)_{t \geq 0}\) has continuous sample paths.

**Proof of 2 \Rightarrow 1.** For all \( 0 \leq s \leq t, A \in \mathcal{F}_s \), the bounded stopping time \( T = s \mathbf{1}_A + t \mathbf{1}_{A^c} \) satisfies \( M_T = |M_s| \mathbf{1}_A + |M_t| \mathbf{1}_{A^c} \in L^1 \) and the property \( \mathbb{E}(M_T) = \mathbb{E}(M_0) \) gives
\[
\mathbb{E}((M_T - M_s) \mathbf{1}_A) = 0,
\]
which implies the martingale projection property for \((M_t)_{t \geq 0}\).

\[\Box\]

**Remark 2.13 (Counter example about stopping with an unbounded stopping time).** If \((M_t)_{t \geq 0}\) is a continuous martingale with \( M_0 = 0 \), then, for all \( a > 0 \), \( T_\alpha = \inf\{t > 0 : M_t = a\} \) is a stopping time, but it cannot be almost surely bounded since this would give \( 0 = \mathbb{E}(M_0) = \mathbb{E}(M_{T_\alpha}) = a > 0 \) which is a contradiction. In this case the process (which is a martingale) \((M_{T_\alpha})_{t \geq 0}\) is not uniformly integrable.

The following theorem allows to control the tail of the supremum of a martingale over a time interval by the moment at the terminal time. It is a continuous time martingale version of the simpler Kolmogorov maximal inequality for sums of i.i.d. random variables.

**Theorem 2.14 (Doob\textsuperscript{a} maximal inequality for non-negative sub-martingales).** Let \( M = (M_t)_{t \geq 0} \) be a continuous non-negative sub-martingale with respect to \((\mathcal{F}_t)_{t \geq 0}\). Then the following hold true:

1. For all \( t > 0 \), \( p \geq 1 \), and \( \lambda > 0 \),
\[
\mathbb{P}(\sup_{s \in [0, t]} M_s \geq \lambda) \leq \frac{\mathbb{E}(M_t^p)}{\lambda^p};
\]

2. For all \( t > 0 \) and \( p > 1 \), if \( \mathbb{E}(M_t^p) < \infty \), then \( \mathbb{E}(\sup_{s \in [0, t]} M_s^p) < \infty \) and\textsuperscript{b}
\[
\mathbb{E}(\sup_{s \in [0, t]} M_s^p) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(M_t^p) \quad \text{in other words} \quad \|\sup_{s \in [0, t]} M_s\|_p \leq \frac{p}{p-1} \|M_t\|_p.
\]

\textsuperscript{a}Named after Joseph L. Doob (1910 – 2004), American mathematician.

\textsuperscript{b}Note that \( q = p/(p-1) \) is the conjugate of \( p \) in the sense that \( 1/p + 1/q = 1 \).

There exists a stronger version of Theorem 2.14 for \(|M_t|\) where \( M \) is a surper-martingale, see for instance [LG16]. We recover the version for non-negative sub-martingales by considering the super-martingale \(-M_t\).

**Theorem 2.14** works for instance if \((M_t)_{t \geq 0} = (\varphi(X_t))_{t \geq 0}\) where \( \varphi : \mathbb{R} \to \mathbb{R}_+ \) is a non-negative convex function and where \((X_t)_{t \geq 0}\) is a continuous martingale. In particular for \((M_t)_{t \geq 0} = (|X_t|)_{t \geq 0}\).

**Proof.**

1. Let \( t > 0 \) and \( p \geq 1 \) such that \( \mathbb{E}(M_t^p) < \infty \). The Jensen inequality implies that \((M_s^p)_{0 \leq s \leq t}\) is a sub-martingale. Let \( \lambda > 0 \) and let us define the bounded stopping time
\[
T = t \wedge \inf\{s \geq 0 : M_s \geq \lambda\}.
\]
The Doob stopping Theorem 2.12 gives
\[
\mathbb{E}(M_T^p) \leq \mathbb{E}(M_t^p).
\]
On the other hand the definition of \( T \) gives
\[
M_T^p \geq \lambda^p \mathbf{1}_{\sup_{s \in [0, t]} M_s \geq \lambda} + M_t^p \mathbf{1}_{\sup_{s \in [0, t]} M_s < \lambda}.
\]
It remains to combine these inequalities to get the desired result.
2. Let $t > 0$ and $p > 1$. For all $n \geq 1$, introducing the stopping time

$$T_n = t \wedge \inf(s \geq 0 : M_s \geq n),$$

the desired inequality for the bounded sub-martingale $(M_{s \wedge T_n})_{s \in [0, t]}$ would give

$$\mathbb{E}(\sup_{s \in [0, t]} M_s^p) \leq \left( \frac{p}{p-1} \right)^p \mathbb{E}(M_t^p),$$

and the desired result for $(M_s)_{s \in [0, t]}$ would then follow using the monotone convergence theorem. Thus, this shows that we can assume without loss of generality that $(M_s)_{s \in [0, t]}$ is bounded, in particular that $\mathbb{E}(\sup_{s \in [0, t]} M_s^p) < \infty$. The previous proof gives

$$\mathbb{P}(\sup_{s \in [0, t]} M_s \geq \lambda) \leq \frac{\mathbb{E}(M_t \mathbf{1}_{\sup_{s \in [0, t]} M_s \geq \lambda})}{\lambda}$$

for all $\lambda > 0$, and thus

$$\int_0^\infty \lambda^{p-1} \mathbb{P}(\sup_{s \in [0, t]} M_s \geq \lambda) d\lambda \leq \int_0^\infty \lambda^{p-2} \mathbb{E}(M_t \mathbf{1}_{\sup_{s \in [0, t]} M_s \geq \lambda}) d\lambda.$$

Now the Fubini–Tonelli theorem gives (here we need $p > 1$)

$$\int_0^\infty \lambda^{p-1} \mathbb{P}(\sup_{s \in [0, t]} M_s \geq \lambda) d\lambda = \int_0^\infty \lambda^{p-1} \mathbb{E}(M_t \mathbf{1}_{\sup_{s \in [0, t]} M_s \geq \lambda}) d\lambda = \frac{1}{p} \mathbb{E}(\sup_{s \in [0, t]} M_s^p).$$

and similarly

$$\int_0^\infty \lambda^{p-2} \mathbb{E}(M_t \mathbf{1}_{\sup_{s \in [0, t]} M_s \geq \lambda}) d\lambda = \frac{1}{p-1} \mathbb{E}(M_t \sup_{s \in [0, t]} M_s^{p-1}).$$

Combining all this gives

$$\mathbb{E}(\sup_{s \in [0, t]} M_s^p) \leq \frac{p}{p-1} \mathbb{E}(M_t \sup_{s \in [0, t]} M_s^{p-1}).$$

But since the Hölder inequality gives

$$\mathbb{E}(M_t \sup_{s \in [0, t]} M_s^{p-1}) \leq \mathbb{E}(M_t^p)^{1/p} \mathbb{E}(\sup_{s \in [0, t]} M_s^{p-1/\tau}),$$

we obtain

$$\mathbb{E}(\sup_{s \in [0, t]} M_s^p) \leq \frac{p}{p-1} \mathbb{E}(M_t^p)^{1/p} \mathbb{E}(\sup_{s \in [0, t]} M_s^{p-1/\tau}).$$

Consequently, since $\mathbb{E}(\sup_{s \in [0, t]} M_s^p) < \infty$, we obtain the desired inequality.

\[ \square \]

**Theorem 2.15 (Increasing process).** If $M = (M_t)_{t \geq 0}$ is a square integrable continuous martingale with respect to $(\mathcal{F}_t)_{t \geq 0}$ with $M_0 = 0$ then there exists a unique continuous and increasing process denoted

$$\langle M \rangle = (\langle M_t \rangle)_{t \geq 0},$$

called the increasing process of $M$, such that

- $\langle M \rangle_0 = 0$;
- $(M_t^2 - \langle M \rangle)_t$ is a martingale with respect to $(\mathcal{F}_t)_{t \geq 0}$.

Moreover the quadratic variation of $(M_t)_{t \geq 0}$ exists and coincides with the increasing process:

$$[M]_t = \langle M \rangle_t \quad \text{for all } t \geq 0.$$
We also speak about the angle bracket (process) of \( M \). Note also in particular that for all \( t \geq 0 \),

\[
\langle M \rangle_t \geq M_0 = 0 \quad \text{and} \quad E(\langle M \rangle_t) = E(M_t^2).
\]

Moreover since \( \langle M \rangle_0 = 0 \) and \( t \mapsto \langle M \rangle_t \) is increasing, almost surely

\[
\frac{\langle M \rangle_t}{t} \xrightarrow{t \to \infty} \langle M \rangle_\infty \in [0, +\infty].
\]

Furthermore, by monotone convergence,

\[
E(M_t^2) = E(\langle M \rangle_t) / t \xrightarrow{t \to \infty} E(\langle M \rangle_\infty) \in [0, +\infty].
\]

In Theorem 2.15, \( (M_t^2)_{t \geq 0} \) is a sub-martingale, and actually Theorem 2.15 states a special case of the more general Doob–Meyer\(^2\) decomposition of sub-martingales which is beyond the scope of this course.

Theorem 3.13 states that a \( d \)-dimensional Brownian motion \( B \) issued from the origin is a continuous process with infinite variation on every interval and with quadratic variation given by \( \langle B \rangle_t = t \) for all \( t \geq 0 \). We also have \( \langle B \rangle_t = t \) for all \( t \geq 0 \). More generally Theorem 4.17 states that for all continuous local martingale \( M \) issued from the origin we have \( [M] = \langle M \rangle \).

**Proof of Theorem 2.15. Case where \( (M_t)_{t \geq 0} \) is bounded.** Let us show that \( (M_t)_{t \geq 0} \) has well defined quadratic variation, namely that if \( \delta_m : 0 = t_0^m < \cdots < t_n^m = t, n = n_m, m \geq 1 \), is a sequence of sub-divisions of \([0, t]\) such that \( \lim_{m \to \infty} |\delta_m| = \lim_{m \to \infty} \max_{1 \leq k \leq n_m} |t_k^m - t_{k-1}^m| = 0 \), then \( \lim_{m \to \infty} \sum_{k=1}^{n_m} (M_{t_k^m} - M_{t_{k-1}^m})^2 \) exists in probability. Let us in fact show that it exists in \( L^2 \).

Let us define, for all \( s \in [0, t] \), with \( t_k^m \) such that \( t_k^m \leq s \leq t_{k+1}^m \),

\[
S_{\delta_m}^\ell(M) = \sum_{i=0}^{k-1} \left( M_{t_i+1}^m - M_{t_i}^m \right)^2 + (M_s - M_{t_k^m})^2.
\]

The martingale property for \( M \) gives that the process \( (M_{t_k^m}^2 - S_{\delta_m}^\ell(M))_{s \in [0, t]} \) is also martingale.

If \( \delta \) and \( \delta' \) are two sub-divisions of the interval \([0, t]\), we denote by \( \delta \cup \delta' \) the sub-division obtained by putting together the points of both sub-divisions, removing the possible duplicates. Now if \( (\delta_m)_{m \geq 1} \) is a sequence of sub-divisions of \([0, t]\) such that \( \lim_{m \to \infty} |\delta_m| = 0 \), let us show that \( (S_{\delta_m}^\ell(M))_{m \geq 1} \) is a Cauchy sequence in \( L^2 \).

The process \( S_{\delta_m}^\ell(M) - S_{\delta_{m'}}^\ell(M) \) is a martingale as difference of two martingales, and thus

\[
E((S_{\delta_m}^\ell(M) - S_{\delta_{m'}}^\ell(M))^2) = E(S_{\delta_m}^\ell(M) - S_{\delta_{m'}}^\ell(M))
\]

\[
\leq 2E(S_{\delta_m}^\ell(M))^2 + 2E(S_{\delta_{m'}}^\ell(M))^2.
\]

Let us denote by \( s_k \)’s the points of \( \delta_m \cup \delta_{m'} \), and, for a fixed \( k \), by \( t_{\ell} \) the points of \( \delta_m \) which is the closest to \( s_k \) and such that \( t_{\ell} \leq s_k \leq t_{\ell+1} \). We have

\[
S_{\delta_{m'}}^\ell(M) - S_{\delta_m}^\ell(M) = (M_{s_{k+1}^m} - M_{s_k^m})^2 - (M_{s_k^m} - M_{t_{\ell}})^2
\]

\[
= (M_{s_{k+1}^m} - M_{s_k^m})(M_{s_{k+1}^m} + M_{s_k^m} - 2M_{t_{\ell}}).
\]

Now, from the Cauchy–Schwarz inequality,

\[
E(S_{\delta_m}^\ell(M))^2 \leq E(\sup_{k}(M_{s_{k+1}^m} + M_{s_k^m} - 2M_{t_{\ell}})^4)^{1/2}E(S_{\delta_m}^\ell(M)^2)^{1/2}.
\]

But since \( M \) is continuous, we have

\[
\lim_{m, m' \to \infty} E(\sup_{k}(M_{s_{k+1}^m} + M_{s_k^m} - 2M_{t_{\ell}})^4) = 0.
\]

\(^2\) Named after Paul-Anré Meyer (1934 – 2003), French mathematician.
On the other hand, since $\mathcal{M}$ is bounded, it follows that $\mathbb{E}(S^{\delta_m \rightarrow \delta_{m'}}_t (\mathcal{M})^2)^{1/2}$ is bounded. This proves the desired convergence in $L^2$, and therefore the quadratic variation exists:

$$ [M]_t = \lim_{m \to \infty} \sum_{k=1}^{n_m} |M^m_{t_k} - M^m_{t_{k-1}}|^2. $$

Now the process $(\mathcal{M}^2 - [M]_t)_{t \geq 0}$ is a martingale because for all $m \geq 1$ and $t \geq 0$, the process $(\mathcal{M}^2 - S^{\delta_m}_t (\mathcal{M}))_{s \in [0,t]}$ is a martingale. We can then set $\langle \mathcal{M} \rangle = [M]_t$.

The continuity of $\langle \mathcal{M} \rangle$ follows from the Doob maximal inequality of Theorem 2.14, namely for all $m, m' \geq 1$ and all $\varepsilon > 0$, we have

$$ \mathbb{P}( \sup_{s \in [0,t]} |S^{\delta_m}_s (\mathcal{M}) - S^{\delta_{m'}}_s (\mathcal{M})| \geq \varepsilon) \leq \frac{\mathbb{E}(S^{\delta_m}_t (\mathcal{M}) - S^{\delta_{m'}}_t (\mathcal{M})^2)}{\varepsilon^2}, $$

and therefore, by the Borel–Cantelli lemma, there exists a sequence $m_k$ such that the sequence of stochastic processes $(S^{\delta_{m_k}}_s)_{s \in [0,t]}$ converges almost surely to the process $(\langle \mathcal{M} \rangle_s)_{s \in [0,t]}$. This proves the existence of a continuous version of $\langle \mathcal{M} \rangle$.

To prove that $\langle \mathcal{M} \rangle$ is increasing, it suffices to consider an increasing sequence of sub-divisions whose mesh tend to zero.

To prove the uniqueness of $\langle \mathcal{M} \rangle$, if $(\mathcal{A}_t)_{t \geq 0}$ and $(\mathcal{A}'_t)_{t \geq 0}$ are two continuous increasing processes issued from the origin such that $(\mathcal{M}^2 - \mathcal{A}_t)_{t \geq 0}$ and $(\mathcal{M}^2 - \mathcal{A}'_t)_{t \geq 0}$ are continuous martingales, then their difference $(\mathcal{A}_t - \mathcal{A}'_t)_{t \geq 0}$ is a continuous finite variation martingale, and according to Lemma 2.16, it is constant. Since $A_0 = A'_0 = 0$, it follows that $A = A'$.

**Case where $\langle \mathcal{M} \rangle_{t \geq 0}$ is not bounded.** For all $N \geq 1$, we introduce the stopping time

$$ T_N = \inf\{t \geq 0 : |\mathcal{M}_t| \geq N\}. $$

From the previous part applied to the bounded martingale $(\mathcal{M}_{t \wedge T_N})_{t \geq 0}$, there exists a unique increasing process $(\mathcal{A}^N_t)_{t \geq 0}$ such that $(\mathcal{M}^2_{t \wedge T_N} - \mathcal{A}^N_t)_{t \geq 0}$ is a martingale. The uniqueness gives $A^N_{t \wedge T_N} = \mathcal{A}^N_t$, and we can define a process $(\mathcal{A}_t)_{t \geq 0}$ by setting $\mathcal{A}_t = \mathcal{A}^N_t$ on the event $\{T_N \geq t\}$. Finally the monotone and dominated converge theorems give that $(\mathcal{M}^2_t - \mathcal{A}_t)_{t \geq 0}$ is a martingale.

For the quadratic variation, it suffices to write

$$ \mathbb{P}(|A_t - \sum_{k=1}^{n} (M^m_{t_k} - M^m_{t_{k-1}})^2| \geq \varepsilon) = \mathbb{P}(T_N \leq t) + \mathbb{P}(|A^N_t - \sum_{k=1}^{n} (M^m_{t \wedge T_N} - M^m_{t_{k-1} \wedge T_N})^2| \geq \varepsilon). $$

\[\Box\]

**Lemma 2.16.** If $(\mathcal{M}_s)_{s \in [0,t]}$ is a finite variation martingale then it is constant.

**Proof of Lemma 2.16.** Let $(\mathcal{M}_s)_{s \in [0,t]}$ be a finite variation continuous martingale. We may assume without loss of generality that $M_0 = 0$. For all $N \geq 1$, we introduce the stopping time

$$ T_N = t \wedge \inf\{s \in [0,t] : |\mathcal{M}_s| \geq N, \sup_k (M^m_{t_{k+1}} - M^m_t) \geq N\} $$

where the supremum runs over all sub-divisions $0 = t^n_0 \ldots t^n_n = t$ of $[0,t]$ into $n$ parts, $n \geq 1$. The stopped process $(\mathcal{M}_{t \wedge T_N})_{t \geq 0}$ is a bounded martingale and thus, for all $s \leq t$,

$$ \mathbb{E}((\mathcal{M}_{t \wedge T_N} - \mathcal{M}_s)^2) = \mathbb{E}(\mathbb{E}((\mathcal{M}_{t \wedge T_N} - \mathcal{M}_{s \wedge T_N})^2 \mid \mathcal{F}_s)) = \mathbb{E}(\mathcal{M}^2_{t \wedge T_N} - \mathbb{E}(\mathcal{M}^2_{t \wedge T_N}). $$

This gives, using a telescoping sum, for an arbitrary sub-division as above,

$$ \mathbb{E}(\mathcal{M}^2_{T_N}) = \mathbb{E}(\mathcal{M}^2_{T_N} - \mathcal{E}(\mathcal{M}^2_{T_N})) = \sum_k (M^m_{t_{k+1} \wedge T_N} - M^m_{t_k \wedge T_N})^2$$
\[ \leq \sup_{k} |M_{t_{k+1} \wedge T_N}^{n} - M_{t_k}^{n} \wedge T_N| + \sum_{k} |M_{t_{k+1} \wedge T_N}^{n} - M_{t_k}^{n} \wedge T_N| \]
\[ \leq N \sup_{k} |M_{t_{k+1} \wedge T_N}^{n} - M_{t_k}^{n} \wedge T_N|. \]

The supremum in the right hand side is bounded uniformity in \( n, N \) since \( M \) has finite variation. As \( n \to \infty \) we get \( E(M_{T_N}^{2}) = 0 \), and therefore \( M_{T_N} = 0 \), which gives in turn \( \hat{M} = 0 \) as \( N \to \infty \) thanks to the fact that \( M \) is continuous with finite variation.

\[ \text{Corollary 2.17 (Quadratic covariation of continuous square integrable martingales).} \text{ If } (M_t)_{t \geq 0} \text{ and } N = (N_t)_{t \geq 0} \text{ are square integrable continuous martingales with respect to the same filtration } (\mathcal{F}_t)_{t \geq 0} \text{ and such that } M_0 = N_0 = 0, \text{ then there exists a unique continuous finite variation process } \langle M, N \rangle = (\langle M, N \rangle_t)_{t \geq 0} \text{ which satisfies the following properties:} \]

- \( \langle M, N \rangle_0 = 0; \)
- \( (M_t, N_t - \langle M_t, N_t \rangle)_{t \geq 0} \) is a martingale with respect to \((\mathcal{F}_t)_{t \geq 0}\).

Moreover the \textbf{quadratic covariation} of \((M, N)\) exists and \( [M, N]_t = \langle M, N \rangle_t \) for all \( t \geq 0 \).

\[ \text{Proof.} \text{ We proceed by quadratic polarization. First the processes } (M_t + N_t)_{t \geq 0} \text{ and } (M_t - N_t)_{t \geq 0} \text{ are square integrable continuous martingales with respect to } (\mathcal{F}_t)_{t \geq 0}. \text{ Next, for all } t \geq 0, \text{ denoting} \]

\[ \langle M, N \rangle_t = \frac{1}{4}((M + N)_t - \langle M - N \rangle_t), \]

we have

\[ M_t N_t - \langle M, N \rangle_t = \frac{1}{4}((M_t + N_t) - (M + N)_t - (M_t - N_t)^2 - \langle M - N \rangle_t)), \]

and thus \( (M_t + N_t - \langle M, N \rangle_t)_{t \geq 0} \) is a martingale. Moreover \( \langle M, N \rangle \) is continuous with finite variation as being the difference of two continuous and increasing processes. The uniqueness and the link with the quadratic covariation follows without difficulty. \]

A basic example is given by Theorem 3.11: if \( B = (B^1, \ldots, B^d) \) is a \( d \)-dimensional Brownian motion then

for all \( t \geq 0 \) and \( 1 \leq j, k \leq d, \quad \langle B^j, B^k \rangle_t = |B^j|^2 = t \mathbf{1}_{j=k}. \]

\[ \text{Corollary 2.18 (Angle brackets and stopping times).} \text{ For all continuous martingales } M \text{ and } N \text{ with } M_0 = N_0 = 0, \text{ and all stopping times } S \text{ and } T \text{ we have} \]

\[ \langle M^T, N^S \rangle = \langle M, N \rangle^{S \wedge T}. \]

\[ \text{Proof.} \text{ Indeed, the Doob stopping theorem (Theorem 2.24) gives that } (M^2 - \langle M \rangle)^T = (M^T)^2 - \langle M \rangle^T \text{ is a continuous martingale. Now } (\langle M^T \rangle)_0 = \langle M \rangle_0 = 0 \text{ and } \langle M \rangle^T \text{ is a continuous increasing process, and thus, by the uniqueness property of the increasing process provided by Theorem 2.15, we have } \langle M^T \rangle = \langle M \rangle^T. \text{ By polarization, } \langle M^T, N^T \rangle = \langle M, N \rangle^T, \text{ and by using this twice for } T \text{ and } S, \text{ the desired formula follows.} \]

\[ \text{Corollary 2.19 (Kunita–Watanabe inequality for continuous martingales).} \text{ For all square integrable martingales } M = (M_t)_{t \geq 0} \text{ and } N = (N_t)_{t \geq 0} \text{ and for all measurable processes } \varphi = (\varphi_t)_{t \geq 0} \text{ and } \psi = (\psi_t)_{t \geq 0}, \text{ we have, for all } t \geq 0, \text{ the following Cauchy–Schwarz type inequality:} \]

\[ \int_0^t |\varphi_s| |\psi_s| d\langle M, N \rangle_s \leq \sqrt{\int_0^t |\varphi_s|^2 d\langle M \rangle_s} \sqrt{\int_0^t |\psi_s|^2 d\langle N \rangle_s}, \]

\[ 24/98 \]
where the integrals are in the sense of bounded variation integrators (Theorem 1.5).

Proof. FIXME:

Let \( \mathcal{M}^2 \) be the vector space of continuous martingales issued from the origin and bounded in \( L^2 \). For all \( M \in \mathcal{M}^2 \), we have \( M_0 = 0 \) and \( \sup_{t \geq 0} \mathbb{E}(M_t^2) < \infty \). By Theorem 2.3, we see the elements of \( \mathcal{M}^2 \) as random variables taking values in \((\mathcal{E}(\mathbb{R}_+, \mathbb{R}), \mathfrak{B}_{\mathcal{E}}(\mathbb{R}_+, \mathbb{R}))\). In particular for all \( M,N \in \mathcal{M}^2 \), we have \( M = N \) if and only if \( M \) and \( N \) are indistinguishable in other words \( \mathbb{P}(\forall t \geq 0 : M_t = N_t) = 1 \). Also \( M = 0 \) if for all \( t \geq 0, M_t = 0 \).

\[ \text{Corollary 2.20} \text{ (Hilbert structure on } \mathcal{M}^2 \text{). The following bilinear map is a scalar product on } \mathcal{M}^2: \]
\[ \langle M, N \rangle \in \mathcal{M}^2 \rightarrow \langle M, N \rangle_{\mathcal{M}^2} = \mathbb{E}(\langle M, N \rangle_{\infty}). \]

Moreover it makes \( \mathcal{M}^2 \) a Hilbert space. Furthermore, for all \( M \in \mathcal{M}^2 \),
\[ \| M \|_{\mathcal{M}^2}^2 = \mathbb{E}(\langle M \rangle_{\infty}) = \sup_{t \geq 0} \mathbb{E}(\langle M \rangle_t) = \sup_{t \geq 0} \mathbb{E}(M_t^2). \]

Proof. The last statement follows immediately form the properties of \( \langle \cdot \rangle \). The bilinearity, symmetry, and positivity of \( \langle \cdot, \cdot \rangle_{\mathcal{M}^2} \) are also immediate. Now if \( M \in \mathcal{M}^2 \) with \( \mathbb{E}(\langle M \rangle_{\infty}) = 0 \) then we have \( \langle M \rangle_t = 0 \) for all \( t \geq 0 \), hence \( \mathbb{E}(M_t^2) = 0 \) for all \( t \geq 0 \), thus \( M_t = 0 \) for all \( t \geq 0 \). To prove completeness, let \( (M^{(n)})_{n \geq 1} \) be a Cauchy sequence in \( \mathcal{M}^2 \). Then for all \( \epsilon > 0 \), there exists \( r \geq 1 \) such that for all \( m,n \geq r \), \( \| M^{(m)} - M^{(n)} \|_{\mathcal{M}^2} \leq \epsilon \). Thus
\[ \sup_{t \geq 0} \mathbb{E}(\| M^{(m)}_t - M^{(n)}_t \|^2) \leq \epsilon^2. \]

This implies that for all \( t \geq 0 \), \( (M^{(n)}_t)_{t \geq 0} \) is a Cauchy sequence in \( L^2 \), and thus converges to an element \( M_t \in L^2 \). It follows that \( M = (M_t)_{t \geq 0} \) is a square integrable martingale, issued from the origin. It remains to prove that \( M \) is continuous. To this end, the idea is to use uniform convergence on finite time intervals. Namely, let us fix \( t > 0 \). From the \( L^2 \) convergence, there exists a sub-sequence \( (n_k)_{k \geq 1} \) such that for all \( k \geq 1 
[ \mathbb{E}(\| M^{(n_k)}_t - M^{(n_k+1)}_t \|^2) \leq 2^{-k}. \]

Now the Doob maximal inequality (Theorem 2.14) for the sub-martingale \( (M^{(n)}_t - M^{(n+1)}_t)_{t \geq 0} \) gives
\[ \mathbb{E}(\sup_{s \in [0,t]} | M^{(n)}_s - M^{(n+1)}_s |^2) \leq 4 \mathbb{E}(\| M^{(n_k)}_t - M^{(n_k+1)}_t \|^2) \leq 2^{-k+2}, \]

and thus, by monotone convergence or the Fubini–Tonelli theorem,
\[ \mathbb{E}\left( \sum_{k \geq 1} \sup_{s \in [0,t]} | M^{(n_k)}_s - M^{(n_k+1)}_s |^2 \right) = \sum_{k \geq 1} \mathbb{E}\left( \sup_{s \in [0,t]} | M^{(n_k)}_s - M^{(n_k+1)}_s |^2 \right) < \infty. \]

Therefore for all \( t > 0 \), almost surely
\[ \sum_{k \geq 1} \sup_{s \in [0,t]} | M^{(n_k)}_s - M^{(n_k+1)}_s | < \infty. \]

This implies that for all \( t > 0 \), almost surely, the sequence of continuous functions \( s \in [0,t] \rightarrow M^{(n)}_s \) converges uniformly towards a limit denoted \( (M_s)_{s \geq 0} \) which is continuous thanks the uniform convergence. This almost sure event can be chosen independent of \( t \) for instance by taking integer values for \( t \). Now for all \( t \geq 0 \), \( (M^{(n)}_t)_{n \geq 1} \) converges to \( M_t \) in \( L^2 \) and to \( M'_t \) almost surely, and therefore \( M_t = M'_t \).

\[ \text{Theorem 2.21} \text{ (Martingales bounded in } L^2 \). Let } (M_t)_{t \geq 0} \text{ be a square integrable continuous martingale with respect to } (\mathfrak{F}_t)_{t \geq 0}, \text{ bounded in } L^2, \text{ namely } \sup_{t \geq 0} \mathbb{E}(M_t^2) < \infty. \text{ Then there exists } M_\infty \in L^2 \text{ such that } \lim_{t \rightarrow \infty} M_t = M_\infty \text{ almost surely and in } L^2. \]

Note that being bounded in \( L^2 \), the martingale \( (M_t)_{t \geq 0} \) is uniformly integrable.
Proof. Let us show that \((M_t)_{t \geq 0}\) satisfies the \(L^2\) Cauchy criterion. For all \(0 \leq s \leq t\), we have
\[
\mathbb{E}((M_t - M_s)^2) = \mathbb{E}(M_t^2 - 2M_t \mathbb{E}(M_t | \mathcal{F}_s) + M_s^2) = \mathbb{E}(M_t^2 - M_s^2).
\]

But \((M_t^2)_{t \geq 0}\) is a sub-martingale, and in particular \(t \mapsto \mathbb{E}(M_t^2)\) grows, and is on the other hand bounded above by \(\sup_{t \geq 0} \mathbb{E}(M_t^2) < \infty\). Thus \(\lim_{t \to \infty} \mathbb{E}(M_t^2)\) exists. Hence \((M_t)_{t \geq 0}\) is Cauchy in \(L^2\), and therefore it converges in \(L^2\) towards some \(M_\infty \in L^2\). It remains to establish the almost sure convergence. Now, by the Chebyshev inequality, for all \(s \geq 0\) and all \(\varepsilon > 0\),
\[
\mathbb{P}(\sup_{t \geq s} |M_t - M_\infty| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbb{E}(\sup_{t \geq s} (M_t - M_\infty)^2) \leq \frac{2}{\varepsilon^2} (\mathbb{E}((M_s - M_\infty)^2) + \mathbb{E}(\sup_{t \geq s} (M_t - M_s)^2)).
\]

Now the monotone convergence theorem gives
\[
\mathbb{E}(\sup_{t \geq s} (M_t - M_s)^2) = \lim_{T \to \infty} \mathbb{E}(\sup_{t \in [s,T]} (M_t - M_s)^2).
\]

On the other hand, for all \(s \geq 0\), the process \(|(M_t - M_s)|_{t \geq s}\) is a continuous non-negative sub-martingale, for which the Doob maximal inequality of Theorem 2.14 gives
\[
\mathbb{E}(\sup_{t \geq s} (M_t - M_s)^2) \leq \lim_{T \to \infty} 4\mathbb{E}((M_T - M_s)^2) = 4\mathbb{E}((M_\infty - M_s)^2).
\]

Therefore we obtain
\[
\mathbb{P}(\sup_{t \geq s} |M_t - M_\infty| \geq \varepsilon) \leq \frac{10}{\varepsilon^2} \mathbb{E}((M_s - M_\infty)^2) \xrightarrow[\varepsilon \to 0]{} 0.
\]

Since the right hand side decreases as \(s\) grows, we get, for all \(\varepsilon > 0\),
\[
\mathbb{P}(\cap_{s \geq 0} \{\sup_{t \geq s} |M_t - M_\infty| \geq \varepsilon\}) = \lim_{s \to \infty} \mathbb{P}(\sup_{t \geq s} |M_t - M_\infty| \geq \varepsilon) = 0,
\]

Similarly, the right hand side decreases as \(\varepsilon\) grows, and then
\[
\mathbb{P}(\cup_{\varepsilon \in \mathbb{Q}, \varepsilon > 0} \cap_{s \geq 0} \{\sup_{t \geq s} |M_t - M_\infty| \geq \varepsilon\}) = \lim_{\varepsilon \to 0} \lim_{s \to \infty} \mathbb{P}(\sup_{t \geq s} |M_t - M_\infty| \geq \varepsilon) = 0,
\]

which means that \(\lim_{t \to \infty} M_t = M_\infty\) almost surely! \(\blacksquare\)

---

**Theorem 2.22** (Doob theorem on closed martingales or Doob martingale convergence theorem). Let \((M_t)_{t \geq 0}\) be a continuous martingale. The following properties are equivalent:

1. (convergence) \(M_t\) converges in \(L^1\) as \(t \to \infty\);
2. (closedness) there exists \(M_\infty \in L^1\) such that for all \(t \geq 0\), \(M_t = \mathbb{E}(M_\infty | \mathcal{F}_t)\);
3. (integrability) the family \([M_t : t \geq 0]\) is uniformly integrable.

In this case, for all \(t \geq 0\), \(M_t = \mathbb{E}(M_\infty | \mathcal{F}_t)\), and \(\lim_{t \to \infty} M_t = M_\infty\) a.s. and in \(L^1\), and \(\mathbb{E}(M_0) = \mathbb{E}(M_\infty)\).

Proof. FixMe:

---

\(^3\)Note that a very similar computation allows to show that \((M_t)_{t \geq 0}\) has orthogonal increments in \(L^2\) in the sense that for all \(0 \leq s \leq t \leq s' \leq t'\) we have \(\mathbb{E}((M_t - M_s)(M_{t'} - M_{s'})) = \mathbb{E}((M_t - M_s)(M_{t'} - M_{s'} | \mathcal{F}_{s'}))) = 0\).

---

26/98
Lemma 2.23 (Characterization of stopped martingales). If $M = (M_t)_{t \geq 0}$ is a continuous adapted process then for all stopping time $T$, the following properties are equivalent:

1. $(M_{t \wedge T})_{t \geq 0}$ is an $(\mathcal{F}_t)_{t \geq 0}$-martingale; 
2. $(M_{t \wedge T})_{t \geq 0}$ is an $(\mathcal{F}_{t\wedge T})_{t \geq 0}$-martingale.

Proof. Proof of $\Rightarrow$. For all $0 \leq s < t$ we have

$$M_{s \wedge T} = \mathbb{E}(M_{t \wedge T} \mid \mathcal{F}_s) = \mathbb{E}(\mathbb{E}(M_{t \wedge T} \mid \mathcal{F}_{s \wedge T}) \mid \mathcal{F}_s) = \mathbb{E}(M_{t \wedge T} \mid \mathcal{F}_{s \wedge T}).$$

Proof of $\Leftarrow$. For all $0 \leq s < t$ and $A \in \mathcal{F}_s$, we have, using the fact that $A \cap \{T > s\} \in \mathcal{F}_{s \wedge T}$,

$$\mathbb{E}(M_{t \wedge T} 1_A) = \mathbb{E}(M_{t \wedge T} 1_{A \cap \{T \leq s\}}) + \mathbb{E}(M_{t \wedge T} 1_{A \cap \{T > s\}})$$
$$= \mathbb{E}(M_T 1_{A \cap \{T \leq s\}}) + \mathbb{E}(\mathbb{E}(M_{t \wedge T} \mid \mathcal{F}_{s \wedge T}) 1_{A \cap \{T > s\}})$$
$$= \mathbb{E}(M_{s \wedge T} 1_{A \cap \{T < s\}}) + \mathbb{E}(M_{s \wedge T} 1_{A \cap \{T > s\}})$$
$$= \mathbb{E}(M_{s \wedge T} 1_A).$$

The following theorem generalizes the Doob stopping theorem (Theorem 2.12).

Theorem 2.24 (More Doob stopping for martingales). Let $M = (M_t)_{t \geq 0}$ be a continuous martingale and let $T$ be a stopping time (not necessarily bounded or finite). The following properties hold true:

1. $(M_{t \wedge T})_{t \geq 0}$ is a martingale;
2. if $M$ is uniformly integrable then $(M_{t \wedge T})_{t \geq 0}$ is also uniformly integrable and for all $t \geq 0$,

$$M_{t \wedge T} = \mathbb{E}(M_T \mid \mathcal{F}_t),$$

and in particular $\mathbb{E}(M_0) = \mathbb{E}(M_{t \wedge T}) = \mathbb{E}(M_T)$ for all $t \geq 0$. Note: $M_T = M_\infty$ on $\{T = \infty\}$.
3. if $M$ is uniformly integrable or if $T$ is bounded, and if $S$ is a stopping time with $S \leq T$, then

$$M_S \in \mathbb{L}^1, \quad M_T \in \mathbb{L}^1, \quad M_S = \mathbb{E}(M_T \mid \mathcal{F}_S),$$

and in particular $\mathbb{E}(M_0) = \mathbb{E}(M_S) = \mathbb{E}(M_T)$. Note: $M_S = M_\infty$ and $M_T = M_\infty$ on $\{S = \infty\}$ and $\{T = \infty\}$.

Proof. FIXME:

 FIXME: add here examples of usage of Doob stopping

Theorem 2.25 (Doob convergence theorem for non-negative sub-martingales). Let $(M_t)_{t \geq 0}$ be a non-negative sub-martingale such that $\sup_{t \geq 0} \mathbb{E}(M_t^+) < \infty$ where $M_t^+ = \max(0, M_t)$. Then there exists $M_\infty \in \mathbb{L}^1$ such that $\lim_{t \to \infty} M_t = M_\infty$ almost surely. Moreover the convergence takes place in $\mathbb{L}^1$ if and only if $(M_t)_{t \geq 0}$ is uniformly integrable.

Proof. FIXME:
Chapter 3

Brownian motion

Figure 3.1: First steps of four sample paths of 2-dimensional Brownian motion issued from the origin, numerically simulated with a programming code like `plot(cumsum(rndn(2,1000)))`.

For all $t > 0$, $d \geq 1$, the density of the Gaussian distribution $\mathcal{N}(0, tI_d)$ on $\mathbb{R}^d$ is

$$x \in \mathbb{R}^d \mapsto p_t(x) = \frac{e^{-|x|^2}}{(\sqrt{2\pi t})^d}$$

where $|x|^2 = x_1^2 + \cdots + x_d^2$.

We have, for all $s, t > 0$,

$$p_{t+s}(x) = (p_t * p_s)(x) = \int_{\mathbb{R}^d} p_t(x-z)p_s(z)dz.$$ 

**Definition 3.1** (Brownian motion). A $d$-dimensional Brownian motion is a **continuous** $d$-dimensional process $B = (B_t)_{t \geq 0}$ such that:

1. For all $0 \leq s \leq t$, the random variable $B_t - B_s$ follows the Gaussian law $\mathcal{N}(0, (t-s)I_d)$;

2. $(B_t)_{t \geq 0}$ has independent increments i.e. for all $t_0 = 0 < t_1 < \cdots < t_n$, $n \geq 0$, the random variables $B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent.
Remark 3.2 (Gaussian processes and Lévy\textsuperscript{a} processes). For all \( n \geq 1 \) and \( 0 \leq t_1 < \cdots < t_n \) the random vector \((B_{t_1}, \ldots, B_{t_n})\) is Gaussian, and we say that Brownian motion is a Gaussian process. On the other hand, for all \( n \geq 1 \) and \( 0 = t_0 < \cdots < t_n \) the increments \( B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}} \) are independent and stationary in the sense that their law depends only on the differences \( t_1 - t_0, \ldots, t_n - t_{n-1} \) between successive times. Also Brownian motion has independent and stationary increments and such processes are called Lévy processes. \textsuperscript{a}Named after Paul Lévy (1886 – 1971), French mathematician.

Remark 3.3 (Basic properties). The first two items below are easy consequences of the definition and show that the study of a \( d \)-dimensional Brownian motion reduces to the study of the one-dimensional Brownian motion issued from the origin.

1. Let \( B = (B_t)_{t \geq 0} \) be a Brownian motion issued from the origin i.e. \( B_0 = 0 \), and let \( H \) be a random variable independent of \( B \). Then according to the definition above the process \((H + B_t)_{t \geq 0}\) is also a Brownian motion;
2. Let \( X = (X_t)_{t \geq 0} \) be a \( d \)-dimensional process and let \( X_t = (X^1_t, \ldots, X^d_t) \) be the coordinates of \( X_t \) in \( \mathbb{R}^d \). Then \( X \) is a Brownian motion issued from the origin if and only if the following two properties hold true:
   \begin{enumerate}
   \item for all \( 1 \leq i \leq d \), \((X^i_t)_{t \geq 0}\) is a Brownian motion issued from the origin;
   \item the processes \((X^1_t)_{t \geq 0}, \ldots, (X^d_t)_{t \geq 0}\) are independent.
   \end{enumerate}
3. Let \( X = (X_t)_{t \geq 0} \) be continuous on \( \mathbb{R}^d \), issued from \( x \in \mathbb{R}^d \), then \( X \) is a Brownian motion if and only if for all \( n \geq 0 \), \( 0 < t_1 < \cdots < t_n \), \( A_1, \ldots, A_n \) are \( \mathcal{B}_{\mathbb{R}^d} \), \( 1 \leq i \leq n \), we have
   \[
   \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n) = \int_{A_1 \times \cdots \times A_n} p_{t_1}(x_1 - x)p_{t_2 - t_1}(x_2 - x_1) \cdots p_{t_n - t_{n-1}}(x_n - x_{n-1}) \, dx_1 \cdots dx_n.
   \]

Theorem 3.4 (Characterization of BM by Gaussianity and covariance). If \( X = (X_t)_{t \geq 0} \) is real, continuous, issued from the origin, then \( X \) is a Brownian motion if and only if \( X \) is a Gaussian process, centered, with covariance given by \( \mathbb{E}(X_tX_s) = s \wedge t \) for all \( s, t \geq 0 \).

Proof.

1. Suppose that \( X = (X_t)_{t \geq 0} \) is a Brownian motion issued from the origin, then for all \( 0 < t_1 < \cdots < t_n \) the random variables \( X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}} \) are Gaussian, centered, and independent, and \( X_0 = 0 \), and \((X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})\) and \((X_{t_1}, \ldots, X_{t_n})\) are (centered) Gaussian random vectors in the sense that all linear combinations of their coordinates are Gaussian. Moreover, for all \( 0 \leq s \leq t \), we have
   \[
   \mathbb{E}(X_sX_t) = \mathbb{E}(X_s(X_t - X_s)) + \mathbb{E}(X^2_s) = 0 + s = s.
   \]
2. Conversely, suppose that \( X = (X_t)_{t \geq 0} \) is a Gaussian process, centered, such that \( \mathbb{E}(X_sX_t) = s \wedge t \), for all \( s, t \geq 0 \). It is easy to deduce that for all \( 0 < t_1 < \cdots < t_n \), the random vector \((X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})\) is Gaussian, centered, with diagonal covariance \( \text{diag}(t_1, t_2 - t_1, \ldots, t_n - t_{n-1}) \) which implies that \((X_t)_{t \geq 0}\) is a Brownian motion.
Corollary 3.5. If \( X = (B_t)_{t \geq 0} \) is a Brownian motion on \( \mathbb{R} \), issued form the origin, then for all \( c \in \mathbb{R} \setminus \{0\} \), the process \( X^{(c)} = (\frac{1}{c} B_{ct})_{t \geq 0} \) is a Brownian motion.

Theorem 3.6 (Invariance by time change). If \( B = (B_t)_{t \geq 0} \) is a Brownian motion on \( \mathbb{R} \) issued from the origin then the process \( X = (tB_{1/t})_{t \geq 0} \), with \( X_0 = 0 \), is a Brownian motion.

Proof. The process \( X \) is Gaussian, centered, with \( \mathbb{E}(X_t, X_s) = s \wedge t \) for all \( s, t \geq 0 \). It remains to prove that \( X \) is continuous. By definition \( X \) is continuous on \( (0, \infty) \). To prove the continuity at \( t = 0 \), one can use the fact that \( (B_t)_{t \in [0,1]} \) and \( (X_t)_{t \in [0,1]} \) have same law as random variables on \( \mathcal{C}((0,1), \mathbb{R}) \). Now both processes vanish at \( t = 0 \). The continuity at \( t = 0 \) depends only on the values on \( (0,1) \) and \( B \) is continuous at \( t = 0 \), thus \( X \) is almost surely continuous at \( t = 0 \). The problem with this approach is that it requires to see \( B \) and \( X \) as random variables taking values on a functional space of trajectories with an appropriate \( \sigma \)-algebra, which is unclear.

It is possible to prove the continuity at \( t = 0 \) of \( X \) without useful the argument about the law of the sample paths of the processes. Namely, it suffices to show that almost surely,

\[
\lim_{t \to 0^+} X_t = \lim_{t \to 0^-} tB_{1/t} = \lim_{t \to \infty} \frac{B_t}{t} = 0.
\]

Thanks to the Borel–Cantelli lemma, for that is sufficient to show that for all \( \varepsilon > 0 \),

\[
\mathbb{P}\left( \lim_{t \to +\infty} \frac{|B_t|}{t} > \varepsilon \right) = 0.
\]

But

\[
\mathbb{P}\left( \lim_{t \to +\infty} \frac{|B_t|}{t} > \varepsilon \right) = \lim_{n \to \infty} \mathbb{P}\left( \sup_{t \in [n]} \frac{|B_t|}{t} > \varepsilon \right).
\]

Now

\[
\mathbb{P}\left( \sup_{t \in [n]} \frac{|B_t|}{t} > \varepsilon \right) \leq \sum_{m=n}^{\infty} \mathbb{P}\left( \max_{m \leq s < m+1} \frac{|B_s|}{s} > \varepsilon \right) \leq \sum_{m=n}^{\infty} \mathbb{P}\left( \max_{0 \leq s \leq m} |B_s| \geq \frac{\varepsilon m}{2} \right).
\]

Let us show that the series converges. For all \( t > 0 \), the symmetry of Brownian motion gives

\[
\mathbb{P}\left( \max_{0 \leq s \leq t} |B_s| \geq \frac{\varepsilon m}{2} \right) \leq \mathbb{P}\left( \max_{0 \leq s \leq t} B_s \geq \frac{\varepsilon m}{2} \right) + \mathbb{P}\left( \max_{0 \leq s \leq t} B_s \geq \frac{\varepsilon m}{2} \right) = 2\mathbb{P}\left( \max_{0 \leq s \leq t} B_s \geq \frac{\varepsilon m}{2} \right),
\]

while the Doob maximal inequality (Theorem 2.14) for the sub-martingale \( (aB_t)_{t \geq 0}, a > 0 \), gives

\[
\mathbb{P}\left( \max_{0 \leq s \leq t} B_s \geq \lambda \right) \leq \mathbb{P}\left( \max_{0 \leq s \leq t} aB_s \geq e^a \lambda \right) \leq e^{-a \lambda} \mathbb{E}(aB_t) = e^{\frac{t}{1}} - a \lambda
\]

which gives, for \( a = \lambda / t \),

\[
\mathbb{P}\left( \max_{0 \leq s \leq t} B_s \geq \lambda \right) \leq e^{-\lambda^2 / 2},
\]

and thus, with \( \lambda = m \varepsilon / 2 \) and \( t = m \),

\[
\mathbb{P}\left( \max_{0 \leq s \leq m} |B_t| \geq \frac{m \varepsilon}{2} \right) \leq 2 e^{-\varepsilon^2 / 2}.
\]
This proves the equivalence between the first two properties. For the third property, it suffices to use the Laplace (instead of Fourier) transform or an argument of analytic continuation.

Let \((\Omega, \mathcal{F}, P)\) is a probability space, equipped with a filtration \((\mathcal{F}_t)_{t \geq 0}\).

**Definition 3.9** (Brownian motion with respect to a filtration). We say that a continuous \(d\)-dimensional process \(X = (X_t)_{t \geq 0}\) is an \((\mathcal{F}_t)_{t \geq 0}\) Brownian motion when it is \((\mathcal{F}_t)_{t \geq 0}\) adapted and for all \(0 \leq s < t\), \(X_t - X_s\) is independent of \(\mathcal{F}_s\) and follows the Gaussian law \(\mathcal{N}(0, (t-s)I_d)\), which means that for all \(\lambda \in \mathbb{R}^d\), the following process is an \((\mathcal{F}_t)_{t \geq 0}\)-martingale:

\[
(e^{i\lambda \cdot X_t + \frac{|\lambda|^2}{2} t})_{t \geq 0}.
\]

**Remark 3.10** (Definitions of Brownian motion). If \(X = (X_t)_{t \geq 0}\) is an \((\mathcal{F}_t)_{t \geq 0}\) Brownian motion, then \(X\) is a Brownian motion in the sense of Definition 3.1. Conversely, a Brownian motion \((X_t)_{t \geq 0}\) in the sense of Definition 3.1 is an \((\mathcal{F}_t)_{t \geq 0}\) Brownian motion where \(\mathcal{G}_t = \sigma(X_s: s \leq t)\) for all \(t \geq 0\) is the natural filtration associated to \(X\) (see Theorem 3.8).

**Theorem 3.11** (Martingale property and quadratic covariation). Let \(B = (B_t)_{t \geq 0}\) be an \((\mathcal{F}_t)_{t \geq 0}\) \(d\)-dimensional Brownian motion and let \(B_t = (B^1_t, \ldots, B^d_t)\) be the coordinates of the random vector \(B_t\). Then for all \(0 \leq s < t\), we have, for all \(1 \leq j, k \leq d\),

1. \(E(B^j_t - B^j_s | \mathcal{F}_s) = 0;\)
2. \(E((B^j_t - B^j_s)(B^k_t - B^k_s) | \mathcal{F}_s) = (t-s)1_{j=k}.\)

In particular, if \(E(||B_0||^2) < \infty\) then, for all \(1 \leq j, k \leq d\),

- \((B^j_t)_{t \geq 0}\) is a continuous \((\mathcal{F}_t)_{t \geq 0}\)-martingale;
- \((B^j_tB^k_t - 1_{j=k})_{t \geq 0}\) is a continuous \((\mathcal{F}_t)_{t \geq 0}\)-martingale, and in particular

\[
([B^j, B^k]_t)_{t \geq 0} = ([B^j, B^k]_t)_{t \geq 0} - (1_{j=k})_{t \geq 0}.
\]

Actually Theorem 5.2 states that these properties characterize Brownian motion.
3.1 Markov property of Brownian motion

Proof. The first property is immediate since \((B^j_t)_{t \geq 0}\) is a Brownian motion. For the second property, we have, for all \(0 \leq s \leq t\) and \(1 \leq j, k \leq d\),

\[
\mathbb{E}((B^j_t - B^j_s)(B^k_t - B^k_s) \mid \mathcal{F}_s) = \mathbb{E}((B^j_t - B^j_s)(B^k_t - B^k_s)) = \mathbb{E}((B^j_t - B^j_s)B^k_t)1_{j \neq k} + \mathbb{E}((B^j_t - B^j_s)^2)1_{j = k} = 0 + (t - s)1_{j = k}.
\]

The consequences are immediate since it follows that for all \(0 \leq s \leq t\) and \(1 \leq j, k \leq d\),

\[
\mathbb{E}(B^j_t | \mathcal{F}_s) = B^j_s = \mathbb{E}(B^j_t | \mathcal{F}_s)
\]

and

\[
\mathbb{E}(B^j_t B^k_t - t1_{j = k} \mid \mathcal{F}_s) = B^j_s B^k_s - s1_{j = k} = \mathbb{E}(B^j_t B^k_t - s1_{j = k} \mid \mathcal{F}_s).
\]

Finally, since \((B^j_t B^k_t - 1_{j = k} t)_{t \geq 0}\) is a continuous martingale, Theorem 2.15 gives

\[
[B^j_t, B^k_t]_t = \langle B^j_t, B^k_t \rangle_t = t1_{j = k}.
\]

■

3.1 Markov property of Brownian motion

Let \((B_t)_{t \geq 0}\) be an \((\mathcal{F}_t)_{t \geq 0}\) \(d\)-dimensional Brownian motion. We easily check that for all fixed \(T > 0\) the process \((B_{t+T} - B_T)_{t \geq 0}\) is a Brownian motion, issued from the origin, independent of \(\mathcal{F}_T\). This is the simple Markov property, which extends to all \((\mathcal{F}_t)_{t \geq 0}\) stopping time \(T\), namely:

Theorem 3.12 (Strong Markov property). Let \(T\) be a stopping time for \((\mathcal{F}_t)_{t \geq 0}\), almost surely finite, let \(\mathcal{F}_T\) be its stopping \(\sigma\)-algebra, and let \(B = (B_t)_{t \geq 0}\) be an \((\mathcal{F}_t)_{t \geq 0}\) \(d\)-dimensional Brownian motion. Then the following properties hold true:

1. \((B_{t+T} - B_T)_{t \geq 0}\) is a Brownian motion issued from the origin, independent of \(\mathcal{F}_T\);
2. For all measurable and bounded \(f : \mathbb{R}^d \to \mathbb{R}\), we have, for all \(t > 0\),

\[
\mathbb{E}(f(B_{t+T}) \mid \mathcal{F}_T) = P_t(f)(B_T)
\]

where

\[
P_t(f)(x) = \frac{1}{(2\pi t)^{d/2}} \int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{2t}} f(y) \, dy = (p_t * f)(x).
\]

"Named after Andrey Markov (1856 – 1922), Russian mathematician.

Proof. Let us define \(B^* = (B_{T+t} - B_T)_{t \geq 0}\). For all \(n \geq 1\), let us define

\[
T_n = \sum_{k=0}^{n} \left\lfloor \frac{k+1}{2^n} \right\rfloor (T).
\]

We have that \(T_n \leq T\), and \(T_n\) takes its values in the set of dyadics \(D_n = \{k/2^n : k \geq 0\}\). We check easily that \(T_n\) is a stopping time, and that \(T_n \uparrow T\) as \(n \to \infty\). Let \(A \in \mathcal{F}_T\), \(m \geq 0\), and \(0 = t_0 < \cdots < t_m < \infty\). We have, for all continuous and bounded \(\varphi : \mathbb{R}^d \to \mathbb{R}\),

\[
\mathbb{E}(1_A \varphi(B^*_m, \ldots, B^*_0)) = \mathbb{E}(1_A \varphi(B_{T+m} - B_T, \ldots, B_{T+1} - B_T)) = \lim_{n \to \infty} \mathbb{E}(1_A \varphi(B_{T+m} - B_T, \ldots, B_{T+m+n} - B_T)).
\]

Moreover, for all \(n \geq 1\), we have \(A \in \mathcal{F}_T \subset \mathcal{F}_{T_n}\) since \(T \leq T_n\) and, using the fact that \(A \in \mathcal{F}_{T_n}\),

\[
\mathbb{E}(1_A \varphi(B_{T+m+n} - B_T, \ldots, B_{T+m+n} - B_T)) = \sum_{r \in D_m} \mathbb{E}(1_A \varphi(B_{T+m+n} - B_T, \ldots, B_{T+m+n} - B_T)).
\]

33/98
3 Brownian motion

This implies the first property since \((B_t - B_0)_{t \geq 0}\) is a Brownian motion issued from the origin. Note that this proves in the same time the fact that \(B^*\) has the law of \(B\) and is independent of \(\mathcal{F}_T\). To prove only the identity in law, we can remove \(1_A\) in other words take \(A = \Omega\).

The second property follows immediately from the first one, namely since for all \(t \geq 0\), \(B^*_{t}\) is independent of \(\mathcal{F}_T\) while \(B_{T}\) is measurable with respect to \(\mathcal{F}_T\) we get, using Remark 1.2,

\[
\mathbb{E}(f(B_t + T) | \mathcal{F}_T) = \mathbb{E}(f(B^*_t + B_t) | \mathcal{F}_T) = g_t(B_T)
\]

where

\[
g_t(x) = \mathbb{E}(f(x + B^*_t)) = \mathbb{E}(f(x + B_t)) = (p_t * f)(x).
\]

We can refer to Definition 1.3 for the notion of finite variation of a function and to Definition 2.9 for the notion of quadratic variation of a process.

**Theorem 3.13 (Variation of Brownian motion).** Let \(B = (B_t)_{t \geq 0}\) be a Brownian motion issued from the origin, let \([u, v]\) be a finite interval, \(0 \leq u < v\), and let \(\delta\) be a partition or subdivision of \([u, v]\), \(\delta : u = t_0 < \cdots < t_n = v\), \(n \geq 1\). Let us consider the quantities

\[
\gamma_1(\delta) = \sum_{i=1}^{n-1} |B_{t_{i+1}} - B_{t_i}| \quad \text{and} \quad \gamma_2(\delta) = \sum_{i=0}^{n-1} |B_{t_{i+1}} - B_{t_i}|^2.
\]

Then the following properties hold true:

1. \(\lim_{|\delta| \to 0} \gamma_2(\delta) = v - u \quad \text{in} \quad \mathbb{L}^2 \quad \text{and thus in} \quad \mathbb{P}, \quad \text{where} \quad |\delta| = \sup_{0 \leq i \leq n} (t_{i+1} - t_i). \quad \text{In other words the quadratic variation of} \quad B \quad \text{is given by} \quad |B|_t = t \quad \text{for all} \quad t \geq 0; \)

2. \(\sup_{\delta \in \mathcal{D}} \gamma_1(\delta) = +\infty \quad \text{almost surely, where} \quad \mathcal{D} \quad \text{is the set of subdivision of} \quad [u, v]. \quad \text{In other words the sample paths of} \quad B \quad \text{are almost surely of infinite variation on all intervals.}

**Proof.** We could use Theorem 3.11 or directly Theorem 2.15 to get that for all \(t \geq 0\), \(|B|_t\) exists and \(|B|_t = \langle B \rangle_t\) and \(\mathbb{E}(|B|_t) = \mathbb{E}(B^2_t) = t\). Moreover we could use Lemma 2.10 to get that the sample path of \(B\) have infinite variation on the time interval \([0, t]\). Let us be more precise by using the special explicit nature of Brownian motion.

1. If \(Z \sim \mathcal{N}(0, 1)\) then \(\mathbb{E}(Z^4) = 3\), hence

\[
\mathbb{E}((\gamma_2(\delta))^2) = \mathbb{E}((\sum_i |B_{t_{i+1}} - B_{t_i}|^2)^2)
\]

\[
= \sum_i \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|^4) + 2 \sum_{i < j} \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|^2 |B_{t_{j+1}} - B_{t_j}|^2)
\]

\[
= 3 \sum_i (t_{i+1} - t_i)^2 + 2 \sum_{i < j} (t_{i+1} - t_i)(t_{j+1} - t_j)
\]

\[
= 2 \sum_i (t_{i+1} - t_i)^2 + (v - u)^2.
\]

Moreover \(\mathbb{E}(\gamma_2(\delta)) = \sum_i (t_{i+1} - t_i) = v - u\). Thus

\[
\mathbb{E}((\gamma_2(\delta) - (v - u))^2) = 2 \sum_i (t_{i+1} - t_i)^2 \leq 2 \max_i (t_{i+1} - t_i)(v - u) \underset{|\delta| \to 0}{\longrightarrow} 0.
\]
2. From the first part, there exists a sequence of subdivisions $(\delta^k)_k$ of $[u, v]$ such that

$$\lim_{k \to \infty} r_2(\delta^k) = \lim_{k \to \infty} \sum_{i=1}^{\infty} |B_{t_{i+1}} - B_{t_i}|^2 = v - u \quad \text{almost surely.}$$

Moreover we see that

$$\sup_{\delta} r_1(\delta) \geq r_1(\delta^k) = \sum_{i} |B_{t_{i+1}} - B_{t_i}| \geq \frac{\sum_{i} |B_{t_{i+1}} - B_{t_i}|^2}{\max_{i} |B_{t_{i+1}} - B_{t_i}|} \to +\infty.$$  

Note that almost surely $\max_{i} |B_{t_{i+1}} - B_{t_i}| \to 0$ as $|\delta_k| \to 0$ since $t \in \mathbb{R}_+ \to B_t$ is continuous and hence uniformly continuous on every compact interval such as $[u, v]$ (Heine theorem).

\[\text{Theorem 3.14 (Law of iterated logarithm). If } (B_t)_{t \geq 0} \text{ is a Brownian motion on } \mathbb{R} \text{ then}\]

$$\mathbb{P} \left( \lim_{t \to 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} = -1, \lim_{t \to 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} = 1 \right) = 1$$

and

$$\mathbb{P} \left( \lim_{t \to \infty} \frac{B_t}{\sqrt{2t \log(\log(t))}} = -1, \lim_{t \to \infty} \frac{B_t}{\sqrt{2t \log(\log(t))}} = 1 \right) = 1.$$  

\[\text{Proof.} \quad \text{The second property follows from the first one by using Theorem 3.6. Let us prove the first property. We can assume without loss of generality that } B_0 = 0. \text{ Since the intersection of two almost sure events is always almost sure, and since the law of Brownian motion issued form the origin is symmetric } (-B \text{ and } B \text{ have same law}), \text{ it follows that it suffices to show that}\]

$$\mathbb{P} \left( \lim_{t \to 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} = 1 \right) = 1.$$  

Let us first prove that

$$\mathbb{P} \left( \lim_{t \to 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} \leq 1 \right) = 1.$$  

Let us define $h(t) = \sqrt{2t \log(\log(1/t))}$. For all $\alpha > 0$ and $\beta > 0$, the Doob maximal inequality of Theorem 2.14 used for the martingale $(e^{\alpha B_t - \frac{\alpha^2 t}{2}})_{t \geq 0}$ gives, for all $t \geq 0$,

$$\mathbb{P} \left( \max_{s \in [0, t]} (B_s - \frac{\alpha}{2} s) > \beta \right) = \mathbb{P} \left( \max_{s \in [0, t]} e^{\alpha B_s - \frac{\alpha^2 s}{2}} \geq e^{\alpha \beta} \right) \leq e^{-a \beta}.$$  

For all $\theta, \delta \in (0, 1)$ and $n \geq 1$, this inequality used with $t = \theta^n$, $\alpha = (1 + \delta) h(\theta^n)/\theta^n$ and $\beta = h(\theta^n)/2$ gives, as $n \to \infty$,

$$\mathbb{P} \left( \max_{s \in [0, \theta^n]} \left( B_s - \frac{(1 + \delta) h(\theta^n)}{2 \theta^n} s \right) > \frac{h(\theta^n)}{2} \right) = \Theta(n^{-(1+\delta)}).$$  

Now, by the Borel–Cantelli lemma, we get that for almost all $\omega \in \Omega$, there exists $N(\omega) \geq 1$ such that for all $n \geq N(\omega)$,

$$\max_{s \in [0, \theta^n]} \left( B_s - \frac{(1 + \delta) h(\theta^n)}{2 \theta^n} s \right) \leq \frac{1}{2} h(\theta^n).$$  

This inequality implies that for all $t \in [\theta^{n+1}, \theta^n]$,

$$B_t(\omega) \leq \max_{s \in [0, \theta^n]} B_s(\omega) \leq \frac{1}{2} (2 + \delta) h(\theta^n) \leq \frac{(2 + \delta) h(t)}{2 \sqrt{\theta}}.$$  

Therefore

$$\mathbb{P} \left( \lim_{t \to 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} \leq \frac{2 + \delta}{2 \sqrt{\theta}} \right) = 1.$$  

35/98
Now we let \( \theta \to 1 \) and \( \delta \to 0 \) to get
\[
P\left( \lim_{t \to 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} \leq 1 \right) = 1.
\]
It remains to prove that
\[
P\left( \lim_{t \to 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} \geq 1 \right) = 1.
\]
For that, for all \( n \geq 1 \) and \( \theta \in (0, 1) \), we define the event
\[
A_n = \{ \omega \in \Omega : B^n_\theta(\omega) - B^{\theta+1}_\theta(\omega) \geq (1 - \sqrt{\theta})h(\theta^n) \}.
\]
We have, denoting \( a_n = (1 - \sqrt{\theta})h(\theta^n)/(\theta^{n/2} \sqrt{1 - \theta}) \),
\[
P(A_n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2} du \geq \frac{a_n}{1 + a_n^2} = \Theta(n^{-\frac{\log(\log(1/t))}{\log(1)}}).
\]
Thus \( \sum_{n \geq 1} P(A_n) = +\infty \). Now the independence of the increments of \( B \) and the Borel–Cantelli lemma give that almost surely, for an infinite number of values of \( n \), we have
\[
B^n_\theta - B^{\theta+1}_\theta \geq (1 - \sqrt{\theta})h(\theta^n).
\]
But the first part of the proof gives, for almost all \( \omega \in \Omega \), that there exists \( N(\omega) \geq 1 \) such that for all \( n \geq N(\omega) \),
\[
B^{n+1}_\theta > -2h(\theta^{n+1}) \geq -2\sqrt{\theta} h(\theta^n).
\]
Therefore, almost surely, for an infinite number of values of \( n \), we have
\[
B^n_\theta > h(\theta^n)(1 - 3\sqrt{\theta}).
\]
This gives
\[
P\left( \lim_{t \to 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} \geq 1 - 3\sqrt{\theta} \right) = 1.
\]
It remains to send \( \theta \) to 0. Note that this proof uses both sides of the Borel–Cantelli lemma.

\[\square\]

**Corollary 3.15** (Regularity of Brownian motion sample paths). If \((B_t)_{t \geq 0}\) is a Brownian motion on \( \mathbb{R} \) then for all \( s \geq 0 \), we have
\[
P\left( \lim_{t \downarrow 0} \frac{B_{t+s} - B_s}{\sqrt{2t \log(\log(1/t))}} = -1, \lim_{t \downarrow 0} \frac{B_{t+s} - B_s}{\sqrt{2t \log(\log(1/t))}} = 1 \right) = 1.
\]
In particular almost surely the sample paths \( t \in \mathbb{R}_+ \rightarrow B_t \) of \( B \) are not \( \frac{1}{2} \)-Hölder\(^a\) continuous on finite intervals and in particular are nowhere differentiable on \( \mathbb{R}_+ \).

\(^a\)Recall that for all \( \gamma > 0 \), a function \( f : I \subset \mathbb{R} \) defined on a finite interval \( I \subset \mathbb{R} \) is \( \gamma \)-Hölder continuous when for all \( \varepsilon > 0 \), there exists \( \eta > 0 \) such that for all \( s, t \in I \) with \( |s - t|^\gamma \leq \eta \), we have \( |f(s) - f(t)| \leq \varepsilon \).

### 3.2 A construction of Brownian motion

A natural and intuitive idea to construct Brownian motion is to try to realize it as a scaling limit of a random walk with Gaussian increments. More precisely, if \((X_n)_{n \geq 1}\) are independent and identically distributed real random variables with law \( \mathcal{N}(0, 1) \), then this would consist for all \( n \geq 1 \) to define the Gaussian process \((X^n_t)_{t \geq 0}\) obtained by linear interpolation as
\[
X^n_t = \frac{X_1 + \cdots + X_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1}}{\sqrt{n}}, \quad t \geq 0,
\]
and to consider its limit as $n \to \infty$. Actually $X^n_t$ is a good approximation for numerical simulation. The central limit phenomenon suggests that the Brownian motion scaling limit is the same if we start from non-Gaussian ingredients: we only need finite mean and variance. This “functional” central limit phenomenon is known as the Donsker invariance principle.

Beyond intuition, the mathematical existence of Brownian motion is not obvious. Historically, Norbert Wiener\(^1\) seems to be the first scientist to give a rigorous construction, around 1923, and for this reason, Brownian motion is sometimes called the Wiener process.

**Lemma 3.16.** If $X = (X_n)_{n \geq 0}$ is a sequence of Gaussian random variables with $X_n \sim \mathcal{N}(m_n, \sigma_n)$ for all $n \geq 1$ and $\lim_{n \to \infty} X_n = X$ in probability, then $X$ is a Gaussian random variable and moreover $\lim_{n \to \infty} X_n = X$ in $L^2$.

**Proof.** Follows using characteristic functions (Fourier transform of probability measures).

---

**Theorem 3.17** (Brownian measures). Let us consider the Hilbert space $G = L^2(\mathbb{R}, dx)$ and

$$\langle f, g \rangle_G = \int f(x)g(x)dx, \quad f, g \in G.$$ 

Then there exists a centered Gaussian family $\bar{B} = (\bar{B}_g)_{g \in G}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that for all $f, g \in G$ and $\alpha, \beta \in \mathbb{R}$ the following properties hold true:

1. $\mathbb{E}(\bar{B}_f \bar{B}_g) = \langle f, g \rangle_G$;
2. $\bar{B}_{\alpha f + \beta g} = \alpha \bar{B}_f + \beta \bar{B}_g$.

The isometry $g \in G \mapsto \bar{B}_g \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ is called the Brownian measure.

Beware that the Brownian measure defined above is not a true random measure because the negligible event behind the equality in the second property (equality behind two random variables) depends on $f$ and $g$ and $G$ is not countable. Indeed, two uncountable families of random variables are not necessarily equal when they are pointwise equal as random variables!

**Proof.** Let $(X_n)_{n \geq 0}$ be a sequence of real Gaussian random variables independent and identically distributed with law $\mathcal{N}(0, 1)$, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and let $(e_n)_{n \geq 0}$ be an orthonormal sequence of the Hilbert space $G = L^2(\mathbb{R}, dx)$. For all $g \in G$, the series $\sum_{n \geq 0} X_n(\omega) \langle g, e_n \rangle_G$ converges in $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Indeed the Cauchy criterion is satisfied:

$$\mathbb{E}\left(\sum_{n=p}^{p+q} X_n(\omega) \langle g, e_n \rangle_G^2 \right) = \sum_{n=p}^{p+q} \langle g, e_n \rangle_G^2 \xrightarrow{p,q \to \infty} 0.$$ 

Let us define now, for all $g \in G$,

$$\bar{B}_g = \sum_{n \geq 0} X_n(\omega) \langle g, e_n \rangle_G.$$ 

We see from Lemma 3.16 that $\bar{B}$ is a centered Gaussian random variable and that

$$\|\bar{B}_g\|^2 = \mathbb{E}(\bar{B}_g^2) = \langle g, g \rangle_G = \|g\|_G^2$$

hence $g \mapsto \bar{B}_g$ is an isometry. Its linearity is immediate. By polarization we get, for all $f, g \in G$,

$$4\mathbb{E}(\bar{B}_f \bar{B}_g) = \mathbb{E}((\bar{B}_f + \bar{B}_g)^2) - \mathbb{E}((\bar{B}_f - \bar{B}_g)^2) = \mathbb{E}(\bar{B}_{f+g}^2) - \mathbb{E}(\bar{B}_{f-g}^2) = \|f + g\|_G^2 - \|f - g\|_G^2$$

---

\(^1\)Named after Norbert Wiener (1894 – 1964), American mathematician.
Finally, let $H \subset L^2(\Omega, \mathcal{F}, P)$ be the closed sub-space generated by the family of random variables $\{\tilde{B}_g : g \in G\}$. Then we easily check that $H$ is isomorphic, via $g \mapsto \tilde{B}_g$, to $G = L^2(\mathbb{R}, dx)$.

With $\tilde{B}$ being as in Theorem 3.17, let us define, for all $t \geq 0$, the random variable

$$B_t = \tilde{B}_{[0,t]}.$$

Now $B = (B_t)_{t \geq 0}$ is a centered Gaussian process, with covariance given for all $s, t \geq 0$ by

$$\mathbb{E}(B_s B_t) = \langle 1_{[0,s]}, 1_{[0,t]} \rangle_{L^2(\mathbb{R}, dx)} = s \wedge t.$$

However $B$ has no reason to be continuous. Let us remark however that for all $0 \leq s < t$,

$$\frac{B_t - B_s}{\sqrt{t-s}} \sim \mathcal{N}(0,1) \quad \text{and} \quad \mathbb{E}((B_t - B_s)^4) = C(t-s)^2 \text{ where } C = \int x^4 e^{-\frac{x^2}{2}} \sqrt{2\pi} dx.$$

This control of the fourth moment allows, thanks to Theorem 3.18 below, to construct a continuous modification $B^*$ of $B$, which is a Brownian motion on $\mathbb{R}$ issued from the origin.

**Theorem 3.18 (Kolmogorov continuity criterion).** Let $X = (X_t)_{t \geq 0}$ be a process defined on a probability space $(\Omega, \mathcal{F}, P)$ taking its values in a Banach space $\mathbb{B}$ with norm $\|\cdot\|$, and such that there exist $p \geq 1, \varepsilon > 0$, and $c > 0$ such that for all $s, t \geq 0$,

$$\mathbb{E}(\|X_t - X_s\|^p) \leq c|t-s|^{1+\varepsilon}.$$

Then there exists a modification* of the process $X$ that is a continuous process whose sample paths on are, on each finite interval, $\gamma$-Hölder continuous for all $\gamma \in (0, \varepsilon/p)$.

*There exists $X^* = (X^*_t)_{t \geq 0}$ such that for all $t \geq 0$, $X_t = X^*_t$ as random variables i.e. almost surely.

**Proof.** It suffices to prove the result on a finite time interval $[0, t]$. For notation simplicity we take $t = 1$. For all $n \geq 1$, and all $\gamma \in [0, \varepsilon/p)$, the Chebyshev inequality gives

$$\mathbb{P}(\max_{1 \leq k \leq n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-n}\gamma) \leq \frac{\mathbb{E}(\|X_{1/2^n} - X_{1/2^{n-1}}\| \geq 2^{-n}\gamma)}{2^{-n}\gamma \mathbb{E}(\|X_{1/2^n} - X_{1/2^{n-1}}\|)} \leq \frac{\sum_{k=1}^{2^n} \mathbb{P}(\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-n}\gamma)}{2^{-n}\gamma \mathbb{E}(\|X_{1/2^n} - X_{1/2^{n-1}}\|)} \leq \frac{\sum_{k=1}^{2^n} 2^{n\gamma p} \mathbb{P}(\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-n}\gamma)}{2^{-n}\gamma \mathbb{E}(\|X_{1/2^n} - X_{1/2^{n-1}}\|)} \leq c2^n 2^{-n(1+\varepsilon) + \gamma p} = c2^{-n(\varepsilon - \gamma p)}.$$

Now, since $\varepsilon > \gamma p$, we get $\sum_{n=1}^{\infty} \mathbb{P}(\max_{1 \leq k \leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-n}\gamma) < \infty$. Thus, from the Borel–Cantelli lemma, it follows that there exists an event $A \in \mathcal{F}$ such that $\mathbb{P}(A) = 1$ and for all $\omega \in A$, there exists $N(\omega)$ such that for all $n \geq N(\omega)$, $\max_{1 \leq k \leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \leq 2^{-n}\gamma$. In particular, there exists a random variable $C$ which is almost surely finite and such that

$$\max_{1 \leq k \leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \leq C2^{-n}\gamma.$$

Let us prove that on the event $A$ the paths of $X$ are $\gamma$-Hölder continuous on the dyadics $\mathcal{D} = \cup_{n \geq 0} \mathcal{D}_n$ where $\mathcal{D}_n = \{k/2^n : k \in \{0, \ldots, n\}\}$. For all $s, t \in \mathcal{D}$ with $s \neq t$, there exists $n \geq 0$ such that $2^{-(n+1)} \leq |s - t| \leq 2^{-n}$. Now, let $(s_k)_{k \geq 1}$ be an increasing and stationary sequence converging towards $s$ and such that $s_k \in \mathcal{D}_k$ and $|s_{k+1} - s_k| \in \{0, 2^{-(k+1)}\}$ for all $k$. Let $(t_k)_{k \geq 1}$ be a similar sequence for $t$, such that $s_k$ and $t_k$ are neighbors in $\mathcal{D}_n$ for all $k$. Then

$$\|X_s - X_t\| \leq \sum_{i=n}^{\infty} \|X_{s_i} - X_{s_{i+1}}\| + \|X_{s_n} - X_{t_n}\| + \sum_{i=n}^{\infty} \|X_{s_{i+1}} - X_{s_i}\|,$$
where the sums are actually finite since the sequences are stationary, and thus
\[
\|X_t - X_s\| \leq C2^{-\gamma n} + 2 \sum_{k=n}^{\infty} C2^{-\gamma(k+1)} \leq 2C \sum_{k=n}^{\infty} 2^{-\gamma k} \leq \frac{2C}{1 - 2^{-\gamma}}2^{-\gamma n},
\]
meaning that \(|s - t| \leq 2^{-n}\) implies \(\|X_t - X_s\| \leq C'2^{-\gamma n}\) for some random variable \(C'\), thus on the event \(A\), the sample paths of \(X\) are \(\gamma\)-Hölder continuous on \(\mathcal{D}\). Note that the set \(\mathcal{D}\) is dense in \(\mathbb{R}_+\). Now for all \(\omega \in A\), let \(t \mapsto X^*_t(\omega)\) be the unique continuous function agreeing with \(t \mapsto X_t(\omega)\) on \(\mathcal{D}\). Note that up to now, the proof extends as in to general metric spaces.

It remains to show that \(X^*\) is a modification of \(X\). By construction we know that \(X_t = X^*_t\) for all \(t \in \mathcal{D}\). Let \(t \in \mathbb{R}_+\). Since \(\mathcal{D}\) is dense in \(\mathbb{R}_+\), there exists a sequence \((t_n)\) in \(\mathcal{D}\) such that \(\lim_{n \to \infty} t_n = t\), thus \(\lim_{n \to \infty} X_{t_n} = X_t\) in \(L^p(\Omega, \mathcal{A}, \mathbb{P})\), \((\mathbb{B}, |\cdot|)\) thanks to the hypothesis. Hence there exists a subsequence \((t_{n_k})\) such that \(\lim_{k \to \infty} X_{t_{n_k}} = X_t\) almost surely. Finally, the continuity of \(X^*\) gives \(X_t = X^*_{t_k} \to X^*_t = X_t\) almost surely as \(k \to \infty\).

3.3 Wiener integral

We know that every finite and deterministic linear combination of the increments of Brownian motion is a Gaussian random variable. More generally, this phenomenon should remain valid for infinite deterministic linear combinations provided square integrability. Indeed, the Wiener integral introduced in Theorem 3.19 gives a meaning to the Gaussian random variable

\[
\omega \in \Omega \rightarrow \int_{\mathbb{R}_+} g(s)dB_s(\omega)
\]

where the integrator \((B_t)_{t \geq 0}\) is Brownian motion and the integrand \(g\) is in \(L^2_{\mathcal{B}_d}(\mathbb{R}_+, \mathbb{R})\). The integrand is deterministic and square integrable, while the integrator is random and Gaussian.

**Theorem 3.19 (Wiener integral).** Let \(B = (B_t)_{t \geq 0} = ((B^1_t, \ldots, B^d_t))_{t \geq 0}\) be a \(d\)-dimensional Brownian motion issued form the origin, defined on \((\Omega, \mathcal{A}, \mathbb{P})\). Let \(G\) be the Gaussian sub-space of \(L^2(\Omega, \mathcal{P})\) generated by the real random variables \((B^i_t : t \geq 0, 1 \leq i \leq d)\). Then there exists a unique map \(I : L^2_{\mathcal{B}_d}(\mathbb{R}_+, \mathbb{R}) \to G\) such that:

1. \(I\) is bijective, continuous, and linear;
2. If \(g = a1_{[s,t]}\) with \(0 \leq s \leq t\) and \(a \in \mathbb{R}^d\) then \(I(g) = a \cdot (B_t - B_s)\);
3. \(I\) is a Hilbertian isometry: for all \(f\) and \(g\) in \(L^2_{\mathcal{B}_d}(\mathbb{R}_+, \mathbb{R})\), we have

\[
\langle f, g \rangle_{L^2_{\mathcal{B}_d}(\mathbb{R}_+, \mathbb{R})} = \int_{\mathbb{R}_+} f(s) \cdot g(s)ds = \mathbb{E}(I(f)I(g)) = \langle I(f), I(g) \rangle_{L^2(\Omega, \mathcal{P})}.
\]

The Wiener integral of \(g\) is the random variable \(I(g)\) and we denote

\[
I(g)(\omega) = \int_{\mathbb{R}_+} g(s)dB_s(\omega).
\]

**Proof.** The following sub-space

\[
S = \left\{ f \in L^2_{\mathcal{B}_d}(\mathbb{R}_+, \mathbb{R}) : f = \sum_{i=0}^{n} a_i 1_{[t_i, t_{i+1}]} , t_0 = 0 < t_1 < \cdots < t_n, n \geq 0, a_i \in \mathbb{R}^d \right\}
\]

of \(L^2_{\mathcal{B}_d}(\mathbb{R}_+, \mathbb{R})\) is dense. If \(f \in S\), then \(f = \sum_{\text{finite}} a_i 1_{[t_i, t_{i+1}]}\), and we define

\[
I(f) = \sum_{\text{finite}} a_i \cdot (B_{t_{i+1}} - B_{t_i}).
\]
This definition does not depend on the decomposition chosen for \( f \), and that the map \( f \mapsto I(f) \) is linear. Moreover, we remark that thanks to the properties of Brownian motion, we have

\[
\mathbb{E}((I(f))^2) = \sum_{i,j} \mathbb{E}((a_i \cdot (B_{t_{i+1}} - B_{t_i}))(a_j \cdot (B_{t_{j+1}} - B_{t_j})))
\]

\[
= \sum_i \mathbb{E}((a_i \cdot (B_{t_{i+1}} - B_{t_i}))^2)
\]

\[
= \sum_i |a_i|^2 (t_{i+1} - t_i)
\]

\[
= \int_{\mathbb{R}_+} |f(x)|^2 \, dx.
\]

Since \( S \) is dense, \( I \) can be extended by continuity to the whole space \( L^2_{\text{erg}}(\mathbb{R}_+, dx) \). Namely, for all \( f \in L^2_{\text{erg}}(\mathbb{R}_+, dx) \), there exists a sequence \((f_n)\) is \( S \) such that \( \|f_n - f\| \to 0 \). Therefore \( \|f_n - f_m\| = \|I(f_n) - I(f_m)\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} \to 0 \) as \( n, m \to \infty \). Set \( I(f) = \lim_{n \to \infty} I(f_n) \). This limit does not depend on the sequence \((f_n)\) used to approximate \( f \). Moreover \( \|f\|_2^2 = \mathbb{E}((I(f))^2) \), and, by polarization, using the linearity of \( I \), we have, for all \( f, g \in L^2_{\text{erg}}(\mathbb{R}_+, dx) \),

\[
\int_{\mathbb{R}_+} f(s)g(s) \, ds = \frac{1}{4} \int_{\mathbb{R}_+} ((f + g)^2 - (f - g)^2) (s) \, ds
\]

\[
= \frac{1}{4} \mathbb{E}((I(f + g))^2 - (I(f - g))^2) = \mathbb{E}(I(f)I(g)).
\]

The map \( I \) defined this way is unique. The map \( I \) is an isometry from \( L^2_{\text{erg}}(\mathbb{R}_+, dx) \) to \( G \) since on one hand \( F = I(L^2_{\text{erg}}(\mathbb{R}_+, dx)) \) is a closed sub-space of \( G \), and on the other hand, for all \( t \geq 0 \) and \( 1 \leq i \leq d \), \( B^I_t \in F \) (take \( g = e_i \cdot (B_t - B_0) \)) and therefore \( F \) is dense in \( G \).

Note that \( L^2(\Omega) \setminus G \) is huge, and most square integrable variables are not Gaussian random variables obtained as square integrable linear combinations of increments of Brownian motion!

**Corollary 3.20** (Properties of the Wiener integral).

1. For all \( f, g \in L^2_{\text{erg}}(\mathbb{R}_+, dx) \), \( I(f) \sim \mathcal{N}(0, \|f\|_2^2) \) and the real Gaussian random variables \( I(f) \) and \( I(g) \) are independent if and only if \( \langle f, g \rangle_{L^2_{\text{erg}}(\mathbb{R}_+, dx)} = \int_{\mathbb{R}_+} f(s)g(s) \, ds = 0 \);

2. For all \( t \geq 0 \) and \( 1 \leq i \leq d \) and \( f \in L^2_{\text{erg}} \), we have

\[
\mathbb{E} \left( B^I_t \int_{\mathbb{R}_+} f(s) \, dB_s \right) = \int_0^t f^i(s) \, ds
\]

where \( f^i(s) \) is the \( i \)-th coordinate of \( f(s) = (f^1(s), ..., f^d(s)) \);

3. Let \((f_n)_{n \geq 0} \) be an orthonormal basis of \( L^2_{\text{erg}}(\mathbb{R}_+, dx) \), then \((I(f_n))_{n \geq 0} \) is a sequence of independent Gaussian real random variables with mean zero and unit variance and for all \( t \geq 0 \), we have the following expansion in \( L^2(\Omega, \mathcal{F}, \mathbb{P}) \):

\[
B_t = \sum_{n \geq 0} \left( \int_{\mathbb{R}_+} f_n(s) \, dB_s \right) \int_0^t f_n(s) ds.
\]

**Proof.**

1. Immediate;

2. Take \( g = e_i 1_{[0, t]} \) then by definition of \( I \) we have \( I(g) = B^I_t \) and

\[
\mathbb{E} \left( B^I_t \int_{\mathbb{R}_+} f(s) \, dB_s \right) = \mathbb{E}(I(g)I(f)) = \int_{\mathbb{R}_+} g(s) \cdot f(s) \, ds = \int_0^t f^i(s) ds.
\]
3. If \( (f_n)_{n \geq 0} \) is an orthonormal basis of \( L^2_{\mathbb{R}^d} \), then \( (I(f_n))_{n \geq 0} \) is an orthonormal basis of the Gaussian space \( \mathbb{G} \) and moreover \( \langle B_t, I(f_n) \rangle_G = \int_0^t f_n^i(s) \, ds \). Note that \( (I(f_n))_{n \geq 0} \) is orthonormal in \( L^2(\Omega, \mathcal{F}, \mathbb{P}) \) but is not a basis: the closure of its span is \( G \subseteq L^2(\Omega, \mathbb{P}) \).

3.4 Wiener measure, canonical Brownian motion, Cameron–Martin formula

Let \( (B_t)_{t \geq 0} \) be an arbitrary \( d \)-dimensional Brownian motion issued from the origin, and defined on a probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \). Since \( (B_t)_{t \geq 0} \) is a continuous process, we know, from Theorem 2.3, that we can consider \( (B_t)_{t \geq 0} \) as a random variable defined on \( (\Omega', \mathcal{A}', \mathbb{P}) \) and taking values in \( (W, \mathcal{B}_W) \) where \( W = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \) is equipped with the topology of uniform convergence on every compact subset of \( \mathbb{R}_+ \) and where \( \mathcal{B}_W \) is the associated Borel \( \sigma \)-algebra.

As a random variable on trajectories, Brownian motion is not unique. We can construct an infinite number of versions of it. What is unique is its law \( \mu \). This law is known as the Wiener measure. There exists however a special realization of Brownian motion as a random variable, which is called the canonical Brownian motion, defined on a canonical space \( (W, \mathcal{B}_W, \mu) \). Namely, on the probability space \( (W, \mathcal{B}_W, \mu) \), where \( \mu \) is the Wiener measure, let us consider the coordinates process \( \pi = (\pi_t)_{t \geq 0} \) defined by

\[
\pi_t(w) = w_t
\]

for all \( t \geq 0 \) and \( w \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \). Under \( \mu \), the process \( \pi \) is a \( d \)-dimensional Brownian motion issued from the origin. It is called the canonical Brownian motion.

**Theorem 3.21** (Wiener measure). There exists a unique probability measure \( \mu \) on the canonical space \( (W, \mathcal{B}_W) \), called the Wiener measure, such that for all \( n \geq 1 \) and \( 0 < t_1 < t_2 < \cdots < t_n \) and all \( A_1, A_2, \ldots, A_n \in \mathcal{B}_W \),

\[
\mu([w \in W : w_{t_1} \in A_1, w_{t_2} \in A_2, \ldots, w_{t_n} \in A_n]) = \int_{A_1 \times A_2 \times \cdots \times A_n} p_{t_n}(x_1)p_{t_{n-1}}(x_2-x_1) \cdots p_{t_1}(x_n-x_{n-1}) \, dx_1 \, dx_2 \cdots dx_n
\]

where \( p \) is the heat or Gaussian kernel defined for all \( t > 0 \) and \( x \in \mathbb{R}^d \) by

\[
p_t(x) = \frac{1}{(\sqrt{2\pi}t)^d} e^{-\frac{|x|^2}{2t}}.
\]

Moreover for all \( d \)-dimensional Brownian motion \( B = (B_t)_{t \geq 0} \) issued from the origin, we have, for all measurable and bounded or positive \( \Phi : W \to \mathbb{R} \),

\[
\mathbb{E}(\Phi(B)) = \int_W \Phi(w) \mu(dw).
\]

**Proof.** We know how to construct a \( d \)-dimensional Brownian motion \( B = (B_t)_{t \geq 0} \) issued form the origin. If \( \mu \) is the law of \( B \) seen as a random variable taking values on the canonical space \( (W, \mathcal{B}_W) \), then it is immediate to get the first desired property since

\[
\mu(B_{t_n} \in A_1, \ldots, B_{t_n} \in A_n) = \mu([w \in W : w_{t_1} \in A_1, \ldots, w_{t_n} \in A_n]).
\]

The uniqueness of \( \mu \) is a consequence of the fact that it is entirely determined on the family \( \mathcal{C} \) of cylindrical subsets of \( W \), which is stable by finite intersections and generates \( \mathcal{B}_W \), see the monotone class theorem (Corollary 1.10).

If \( X \sim \mathcal{N}(0, I_d) \) is a standard Gaussian random variable on \( \mathbb{R}^d \) with Lebesgue density \( x \mapsto \gamma_n(x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}|x|^2} \), then, for all \( h \in \mathbb{R}^n \) and all bounded and measurable \( \Phi : \mathbb{R}^n \to \mathbb{R} \),

\[
\mathbb{E}(\Phi(X+h)) = \int_{\mathbb{R}^n} \Phi(x+h) \gamma_n(x) \, dx = \int_{\mathbb{R}^n} \Phi(x') \gamma_n(x' - h) \, dx' = \mathbb{E}(\Phi(X) e^{x \cdot h - \frac{1}{2}|h|^2}.
\]
This nice “translation formula” can be interpreted in terms of Laplace transform or Gaussian integration by parts. Moreover it turns out that this formula has a counterpart for Brownian motion and the Wiener measure, which is an infinite dimensional Gaussian distribution, provided that the translation \( h \) belongs to a special space known as the Cameron–Martin space.

The Cameron–Martin space is defined by

\[
H = \left\{ h \in W : \forall t \geq 0, h(t) = \int_0^t \dot{h}(s)ds, \dot{h} \in L^2_{\mathbb{R}^d}(\mathbb{R}_+, dx) \right\}.
\]

This is a subspace of \( W = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \). We equip \( H \) with the scalar product

\[
(h_1, h_2)_H = \int_{\mathbb{R}_+} \dot{h}_1(s) \cdot \dot{h}_2(s)ds.
\]

This makes \( H \) a Hilbert space isomorphic to \( L^2_{\mathbb{R}^d}(\mathbb{R}_+, dx) \), and the canonical injection from \( H \) to \( W \) is continuous. For every \( h \in H \) we denote

\[
|h|_H^2 = \int_0^\infty |\dot{h}(s)|^2 ds = \|\dot{h}\|^2_{L^2_{\mathbb{R}^d}(\mathbb{R}_+, dx)}.
\]

Let \( (h_n)_{n \geq 0} \) be an orthonormal basis of \( H \) and for all \( n \geq 0 \) and \( t \geq 0 \),

\[
h_n(t) = \int_0^t \dot{h}_n(s)ds.
\]

The sequence \( (\dot{h}_n)_{n \geq 0} \) is an orthonormal basis of \( L^2_{\mathbb{R}^d}(\mathbb{R}_+, dx) \). Let \( \pi = (\pi_t)_{t \geq 0} \) be the canonical Brownian motion on \( \mathbb{R}^d \) and let us define, for all \( w \in W \) and \( h \in H \), using the Wiener integral,

\[
(w, h) = \int_{\mathbb{R}_+} \dot{h}(s)d\pi_s(w) = \int_{\mathbb{R}_+} \dot{h}(s)d\pi_s.
\]

Then the second property provided by Corollary 3.20 gives that the real random variables \( (w, h_n), n \geq 0 \), are independent, Gaussian, with zero mean and unit variance, and for all fixed \( t \geq 0 \) we have the following expansion in \( L^2(W, \mathcal{F}_t, \mu) \):

\[
\pi_t(w) = w_t = \sum_{n \geq 0} (w, h_n)h_n(t).
\]

Here \( t \) is fixed and \( h_n(t) \) is a real coefficient, while \( (w, h_n)_{n \geq 0} \) are orthogonal in \( L^2(W, \mathcal{F}_t, \mu) \).

**Remark 3.22** (Uniform convergence). We can prove by using a convergence theorem for vector maringales that for all \( t > 0 \) and for almost all \( w \) for \( \mu \), we have

\[
\lim_{N \to \infty} \sup_{s \in [0, t]} \left| w_s - \sum_{n=0}^N (w, h_n)h_n(s) \right| = 0.
\]

The notion of density of the Wiener measure does not make sense due to the lack of good notion of Lebesgue measure on the Wiener space. However the shift of the Wiener measure in a direction belonging to the Cameron–Martin space is absolutely continuous with respect to the Wiener measure and the density in the sense of Radon–Nikodym is explicit. This is an infinite dimensional analogue of the formula mentioned above for finite dimensional Gaussian laws.

**Theorem 3.23** (Cameron\(^a\)–Martin\(^b\) formula and density of shifts). Let \( W \) and \( \mu \) be the Wiener space and measure, \( \Phi : W \to \mathbb{R} \) be measurable and bounded, and \( h \) be in the Cameron–Martin space \( H \). Then we have the Cameron–Martin formula:

\[
\int \Phi(w + h)\mu(dw) = \int \Phi(w)e^{(w, h) - \frac{1}{2}|h|^2_0}\mu(dw).
\]

In other words, if \( B \) is canonical Brownian motion and \( F_h(w) = e^{(w, h) - \frac{1}{2}|h|^2_0} \), then

\[
\mathbb{E}(\Phi(B + h)) = \mathbb{E}(\Phi(B)F_h(B)),
\]

42/98
It particular $\mathbb{E}(F_h(B)) = 1$ and the law of $(B_t + h)_{t \geq 0}$ has density $F_h$ with respect to $\mu$.

**Proof.** By using a monotone class argument, we can assume without loss of generality that $\Phi$ is such that $\Phi(w) = f(w_n, \ldots, w_{n+1})$ where $n \geq 1$ and $0 \leq t_1 < \cdots < t_n$ and $f \in C^\infty(\mathbb{R}^n)$ with compact support. Let $h \in \mathbb{H}$, $h \neq 0$. There exists an orthonormal basis $(h_m)_{m \geq 0}$ of $\mathbb{H}$ such that $h_0 = h/|h|_H$. For all $m \geq 1$, let $w^{(m)} \in W$ be defined $\mu$ almost surely by

$$w^{(m)} = \sum_{\ell=0}^{m} (w, h_\ell) h_\ell.$$ 

We have $\lim_{m \to \infty} w_m = w = \sum_{n \geq 0} (w, h_n) h_n$ in $L^2(W)$. Since $\Phi$ is continuous and bounded, it follows that $\lim_{m \to \infty} \Phi(w_m + h) = \Phi(w + h)$ in probability, and thus, by dominated convergence,

$$\lim_{m \to \infty} \mathbb{E}_\mu(\Phi(w_m + h)) = \mathbb{E}_\mu(\Phi(w + h)).$$

Similarly we have, using the fact that $(w, h) \sim \mathcal{N}(0, |h|_H^2)$ under $\mu$,

$$\lim_{m \to \infty} \mathbb{E}_\mu\left(\Phi(w^{(m)}) \exp\left((w, h) - \frac{|h|_H^2}{2}\right)\right) = \mathbb{E}_\mu\left(\Phi(w) \exp\left((w, h) - \frac{|h|_H^2}{2}\right)\right).$$

Also, to prove the desired formula, it suffices to show that for all $m \geq 1$,

$$\mathbb{E}_\mu(\Phi(w^{(m)} + h)) = \mathbb{E}_\mu\left(\Phi(w^{(m)}) \exp\left((w, h) - \frac{|h|_H^2}{2}\right)\right).$$

This boils down to a simple computation in finite dimension. Namely, since

$$w^{(m)} + h = ((w, h_0) + |h|_H) \frac{h}{|h|_H} + \sum_{\ell=1}^{m} (w, h_\ell) h_\ell,$$

where $(w, h_\ell), \ell \geq 0$ are independent and identically distributed with law $\mathcal{N}(0, 1)$, we have

$$\mathbb{E}_\mu(\Phi(w^{(m)} + h)) = \frac{1}{(\sqrt{2\pi})^m} \int_{\mathbb{R}^m+} \Phi\left(x_0 + |h|_H h_0 + \sum_{\ell=0}^{m} x_\ell h_\ell\right) e^{-\frac{1}{2} \sum_{\ell=0}^{m} x_\ell^2} dx_0 \cdots dx_m = \frac{1}{(\sqrt{2\pi})^m} \int_{\mathbb{R}^m+} \Phi\left(x_0 h_0 + \sum_{\ell=0}^{m} x_\ell h_\ell\right) e^{x_0^2 |h|_H^2 - \frac{1}{2} \sum_{\ell=0}^{m} (x_\ell^2 |h|_H^2 - \frac{|h|_H^2}{2})} dx_0 \cdots dx_m = \mathbb{E}_\mu(\Phi(w^{(m)}) e^{w_0 h_0 |h|_H^2 - \frac{|h|_H^2}{2}}) = \mathbb{E}_\mu(\Phi(w^{(m)}) e^{(w, h) - \frac{|h|_H^2}{2}}).$$

**Corollary 3.24** (Density of Cameron–Martin shifts). If $B = (B_t)_{t \geq 0}$ is a $d$-dimensional Brownian motion issued from the origin defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then,

for all $h \in \mathbb{H}$, $\mathbb{E}(F_h(B)) = 1$, where $F_h(B) = \exp\left(\int_{\mathbb{R}^+} \dot{h}_s dB_s - \frac{1}{2} \int_{\mathbb{R}^+} |\dot{h}_s|^2 ds\right)$.

Moreover, for all $h \in \mathbb{H}$, the law of the translated process

$$\tilde{B}^{(h)} = (B_t + h_t)_{t \geq 0}$$

is absolutely continuous with respect to the law of $B$, with density given by $F_h(B)$. Furthermore the translation process $\tilde{B}^{(-h)}$ is a $d$-dimensional Brownian motion issued from the origin with respect to the probability measure $Q$ on $(\Omega, \mathcal{F})$ given by $dQ/d\mathbb{P} = F_h(B)$. 

---

\(43/98\)
Proof. Note that since \( \int_{\mathbb{R}_+} h_s dB_s \sim \mathcal{N}(0, |h|_1^2) \), we get immediately that (Laplace transform!)

\[
\mathbb{E}\left( \exp\left( \int_{\mathbb{R}_+} h_s dB_s \right) \right) = \exp\left( \frac{|h|_1^2}{2} \right),
\]

which gives \( \mathbb{E}(F_h(B)) = 1 \). But in fact, we know that up to a negligible event, we can view \( B \) as a random variable taking values in \( W \) with law \( \mu \) (under \( \mathbb{P} \)), and therefore the Cameron–Martin formula of Theorem 3.23 writes, for all measurable and bounded \( \Phi : W \to \mathbb{R} \),

\[
\mathbb{E}(\Phi(B + h)) = \mathbb{E}(\Phi(B) \exp\left( \int_{\mathbb{R}_+} h_s dB_s - \frac{|h|_1^2}{2} \right)) = \mathbb{E}(\Phi(B) F_h(B)) = \int \Phi(w) F_h(w) \mu(dw)
\]

which gives by the way that \( \mathbb{E}(F_h(B)) = 1 \) by taking \( \Phi \equiv 1 \), and which show more generally that the law of \( B^{(h)} \) has density \( F_h \) with respect to the law \( \mu \) of \( B \). Moreover

\[
\mathbb{E}(\Phi(B)) = \mathbb{E}(\Phi(B - h) F_h) = \mathbb{E}_Q(\Phi(B - h)).
\]

The law of \( Q \) under \( \mathbb{P} \) is therefore the law of \( \tilde{B}^{(-h)} = B - h \) under \( Q \). ■
Chapter 4

Itô stochastic integral

Let \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) be a filtered probability space with \((\mathcal{F}_t)_{t \geq 0}\) right-continuous and complete. Our aim is to construct a stochastic integral which goes beyond the Wiener integral of Chapter 3, with an integrator \((M_t)_{t \geq 0}\) which can be at least a Brownian motion and with an integrand \((\varphi_t)_{t \geq 0}\) which can be at least random and possibly square integrable if needed. The ambition is thus to define the process

\[
\left( \int_0^t \varphi_s \, dM_s \right)_{t \geq 0}.
\]

- In Section 4.2, we construct a stochastic integral when \(M\) is a \(d\)-dimensional Brownian motion. Following Itô, we start with a finite sum when \(\varphi\) is a \(d\)-dimensional step process. We impose a predictable structure to the step process in such a way that the resulting stochastic integral is a martingale. This corresponds to choose the value at the left-end time of intervals in the sum. This Itô stochastic integral coincides with the Wiener integral of Chapter 3 when the integrand is deterministic. The formula for the angle bracket of the martingale provided by this Itô stochastic integral leads to extend the class of integrands, by approximation with (random) step processes, in a very natural \(L^2\) sense. The final extension of this Itô stochastic integral concerns integrands that are only almost surely locally square integrable in time, and in this case, the resulting stochastic integral, obtained by cutoff (localization with stopping times), is no longer a martingale, but is a local martingale.

- In Section 4.3, for one-dimensional processes, we present a construction of the Itô stochastic integral when \(M\) is a continuous martingale bounded in \(L^2\), and when \(\varphi\) is square integrable in a suitable sense. This is done from step processes and by using the Hilbert structure available in this \(L^2\) context for both \(\varphi\) and \(M\).

- In Section 4.4, by using cutoff (localization via stopping times), we extend the previous Itô stochastic integral from the case were \(M\) is a continuous martingale bounded in \(L^2\) to the case where \(M\) is a continuous local martingale. This Itô stochastic integral coincides with the integral constructed previously with respect to Brownian motion.

- In Section 4.5, the most general integrators \(M\) that we reach for the Itô stochastic integral are sums of local martingales and bounded variation processes, called semi-martingales.

4.1 Intuitive mathematical experiments

Let us try to understand intuitively the nature of the desired object, for instance

\[
\int_0^t B_s \, dB_s
\]

where \((B_t)_{t \geq 0}\) is a one-dimensional Brownian motion issued from the origin. This is not a Wiener integral since the integrand is random. Using an approximation with step processes, we have

\[
\sum_i B_{t_i}(B_{t_{i+1}} - B_{t_i}) = \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2 - B_{t_{i+1}}^2 + B_{t_i}B_{t_{i+1}}) = \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2 - B_{t_{i+1}}(B_{t_{i+1}} - B_{t_i})
\]
Itô stochastic integral

\[ = \sum_{i} (B_{t_{i+1}}^2 - B_{t_i}^2) - \sum_{i} B_{t_{i}} (B_{t_{i+1}} - B_{t_{i}}) - \sum_{i} (B_{t_{i+1}} - B_{t_{i}})^2 \]

which gives, if we can pass to the limit in probability,

\[ \int_0^t B_s \, dB_s = B_t^2 - \int_0^t B_s \, dB_s - [B]_t, \]

and thus, using the fact that \([B]_t = t\),

\[ \int_0^t B_s \, dB_s = \frac{B_t^2 - t}{2}. \]

We check immediately that \((\frac{1}{2}(B_t^2 - t))\) is a centered martingale. The term \(-\frac{1}{2}t\) is the martingale correction to the differential calculus term \(\frac{1}{2}B_t^2\). This suggest that taking the value at the left-end time intervals in the Riemann sum produces a stochastic integral which is a centered martingale. This is the Itô stochastic integral. Indeed this is confirmed by the rigorous mathematical construction in the rest of this chapter.

Let us examine the notion of Stratonovich stochastic integral, possibly distinct from the Itô integral, which corresponds to take the mean value between the left-end and right-end times of the intervals in the Riemann sum. Namely, we have

\[ \sum_{i} \frac{B_{t_{i+1}} + B_{t_{i}}}{2} (B_{t_{i+1}} - B_{t_{i}}) = \frac{1}{2} \sum_{i} (B_{t_{i+1}}^2 - B_{t_{i}}^2) = \frac{1}{2} B_t^2, \]

which gives, provided that the convergence holds in probability, and denoting with \(\circ\) the Stratonovich stochastic integral,

\[ \int_0^t B_s \, dB_s = \frac{1}{2} B_t^2. \]

This time the rule of differential calculus is satisfied, but the result is not a martingale. By symmetry we can also define some sort of anticipative integral from the formula

\[ \sum_{i} B_{t_{i+1}} (B_{t_{i+1}} - B_{t_{i}}), \]

which will lead to

\[ 2(\text{Stratonovich}_t - \frac{1}{2} \text{Itô}_t) = B_t^2 + \frac{1}{2} t. \]

### 4.2 Stochastic integral with respect to Brownian motion

Let \(B = (B_t)_{t \geq 0}\) be a \(d\)-dimensional \((\mathcal{F}_t)_{t \geq 0}\) Brownian motion issued from the origin. In this section we construct rigorously the Itô stochastic integral with integrator \(B\).

#### 4.2.1 Stochastic integral of step processes

A step process is a \(d\)-dimensional process \((\varphi_t)_{t \geq 0}\) for which there exists \(n \geq 1\) and \(0 \leq t_0 \leq \cdots \leq t_n\) and random variables \(U_0, \ldots, U_{n-1}\) in \(L^2(\Omega, \mathcal{F}, \mathbb{P})\), \(\mathcal{F}_{t_0}, \ldots, \mathcal{F}_{t_n}\) measurable respectively, such that for all \(t \geq 0\)

\[ \varphi_t = U_0 1_{0}(t) + \sum_{i=0}^{n-1} U_{i} 1_{[t_{i}, t_{i+1})}(t). \]

Such a step process is progressive, left-continuous, and on each time interval, the random value is measurable with respect to the \(\sigma\)-algebra which corresponds to the left end time of the interval, hence the name “predictable” which is used sometimes. The vector space of step processes is denoted \(\mathcal{F}^2\). For all \(\varphi \in \mathcal{F}^2\), we define the Itô stochastic integral of the step process \(\varphi\) with respect to the Brownian motion \(B\) by for the following formula, for all \(t \geq 0\),

\[ I(\varphi)_t = \int_0^t \varphi_s \, dB_s = \sum_{i=0}^{n-1} U_{i} \cdot (B_{t_{i+1}} - B_{t_{i}}). \]

It is a square integrable real random variable. The following properties are immediate:
1. \( I(\varphi)_0 = 0; \)

2. \( I(\varphi) \) does not depend on the decomposition chosen for \( \varphi; \)

3. the map \( \varphi \mapsto I(\varphi) \) is linear;

4. When the integrand \( \varphi \) is deterministic in other words when the \( U_i \)'s are deterministic, then we recover exactly the notion of Wiener integral introduced in Chapter 3.

Here are additional important properties of the Itô integral of step processes.

**Theorem 4.1** (Brownian Itô stochastic integral of step processes). For all \( \varphi \in \mathcal{F}^2 \), the real stochastic process issued from the origin

\[
(I(\varphi)_t)_{t \geq 0} = \left( \int_0^t \varphi_s dB_s \right)_{t \geq 0}
\]

is a continuous centered and square integrable martingale, and for all \( t \geq 0 \),

\[
\langle I(\varphi) \rangle_t = \int_0^t |\varphi_s|^2 dB_s,
\]

in particular we have the Itô isometry

\[
\mathbb{E}\left( \left( \int_0^t \varphi_s dB_s \right)^2 \right) = \mathbb{E}\left( \int_0^t |\varphi_s|^2 ds \right).
\]

Note that \( I(\varphi) \) is centered even if \( \varphi \) (in other words the \( U_i \)'s) are not.

Since the map \( I \) is linear, it follows from the theorem above that for all \( \varphi, \psi \in \mathcal{F}^2 \), by using the polarization

\[
4I(\varphi)I(\psi) = I(\varphi + \psi)^2 - I(\varphi - \psi)^2,
\]

we have, for all \( t \geq 0 \),

\[
\langle I(\varphi), I(\psi) \rangle_t = \int_0^t (\varphi_s \cdot \psi_s) ds.
\]

**Proof.** Let \( \varphi \in \mathcal{F}^2 \).

1. The continuity of \( t \mapsto I(\varphi)_t \) follows by construction from the one of \( t \mapsto B_t \). Moreover, we have \( I(\varphi)_0 = 0 \). Furthermore, for all \( t > 0 \), the random variable \( I(\varphi)_t \) is the (finite) sum of \( U_0 \cdot (B_{t_1} - B_{t_0}), \ldots, U_k \cdot (B_{t_k} - B_{t_{k-1}}) \) where \( k \) is such that \( t \in (t_k, t_{k+1}]. \) These random variables are centered square integrable and \( \mathcal{F}_t \) measurable. Note that here it is crucial for the \( \mathcal{F}_t \) measurability of \( I(\varphi)_t \) that \( U_i \) is \( \mathcal{F}_t \) measurable for all \( i \) in the definition on \( \varphi \). Moreover, for all \( 0 \leq s \leq t \), if \( s \in (t_k, t_{k+1}] \), \( k' \leq k \), the decomposition

\[
I(\varphi)_t = U_0 \cdot (B_{t_1} - B_{t_0}) + \cdots + U_{k'} \cdot (B_{k'} - B_{t_{k'}}) + U_{k'} \cdot (B_{t_{k'} + 1} - B_s) + \cdots + U_k \cdot (B_t - B_{t_{k+1}})
\]

gives, using the properties of \( B \) and the measurability assumptions on \( U \), that

\[
\mathbb{E}(I(\varphi)_t | \mathcal{F}_s) = I(\varphi)_s + 0.
\]

The zero comes from the fact that for all \( v \geq u \geq s \),

\[
\mathbb{E}(U \cdot (B_v - B_u) | \mathcal{F}_s) = \mathbb{E}(U \cdot \mathbb{E}(B_v - B_u | \mathcal{F}_u) | \mathcal{F}_s) = 0
\]

when \( U \) is bounded and \( \mathcal{F}_u \)-measurable, since \( B_v - B_u \) is conditionally independent of \( \mathcal{F}_u \). Hence \( (I_t)_{t \geq 0} \) is a continuous, centered, and square integrable martingale.

2. Let us compute \( \langle X \rangle \) where \( X = I(\varphi) \). Let us show that for all \( 0 \leq s < t \) and \( A \in \mathcal{F}_s \),

\[
\mathbb{E}\left( \mathbb{1}_A \left( X_t^2 - X_s^2 - \int_s^t |\varphi_u|^2 du \right) \right) = 0.
\]
Since $X$ is a square integrable martingale, we have\(^1\)
\[
\mathbb{E}(1_A(X_t^2 - X_s^2)) = \mathbb{E}(1_A(X_t^2 - 2X_tX_s + X_s^2)) = \mathbb{E}(1_A(X_t - X_s)^2).
\]

Now, with $\varphi_t = U_01_0(t) + \sum_j U_j1_{[t_j, t_{j+1})}(t)$, $\Delta_{u,v} = B_v - B_u$, and $t_0 = 0$, we have
\[
X_t = U_0 \cdot \Delta t_0, t_1 + U_1 \cdot \Delta t_1, t_2 + \cdots + U_k \cdot \Delta t_k, t_{k+1}
\]
if $t \in (t_k, t_{k+1})$.

Since $s < t$, there exists $\ell \leq k$ such that $s \in (t_\ell, t_{\ell+1})$, and thus
\[
1_A(X_t - X_s) = \tilde{U}_\ell \cdot \Delta s, t_\ell + \tilde{U}_{\ell+1} \Delta t_\ell, t_{\ell+2} + \cdots + \tilde{U}_k \cdot \Delta t_k, t
\]
where $\tilde{U}_\ell = 1_A U_\ell$, $\ell \leq i \leq k$, is still $\mathcal{F}_t$ measurable. Let $i, i' \in (\ell, \ell + 1, \ldots, k)$. Now $\tilde{U}_i$ is $\mathcal{F}_t$ measurable and $\Delta t_i, t_i$ is conditionally independent of $\mathcal{F}_t^i$ with $\mathbb{E}(\Delta t_i, t_i \cdot \Delta t_{i', t_{i'}}) = (t_{i+1} - t_i)1_{i = j}$. If $i < i'$ then
\[
\mathbb{E}(\tilde{U}_i \cdot \Delta t_i, t_i) = \mathbb{E}(\tilde{U}_i \cdot \Delta t_{i', t_{i'}}) = \mathbb{E}(\tilde{U}_i \cdot \Delta t_{i', t_{i'}} | \mathcal{F}_t) = \cdots = 0,
\]
while if $i = i'$ then
\[
\mathbb{E}(\tilde{U}_i \cdot \Delta t_i, t_i)^2 = \mathbb{E}(\tilde{U}_i \cdot \Delta t_{i', t_{i'}})^2 | \mathcal{F}_t) = \cdots = \mathbb{E}(\tilde{U}_i)^2(t_{i+1} - t_i).
\]

Finally it remains to write
\[
\mathbb{E}(1_A(X_t - X_s)^2) = \sum_{i, i' = \ell}^{k} \mathbb{E}(\tilde{U}_i \cdot \Delta t_i, t_i)(\tilde{U}_{i'} \cdot \Delta t_{i', t_{i'}})
\]
\[
= \sum_{i = \ell}^{k} \mathbb{E}(\tilde{U}_i)^2(t_{i+1} - t_i)
\]
\[
= \mathbb{E}(1_A \sum_{i = \ell}^{k} |U_i|^2(t_{i+1} - t_i))
\]
\[
= \mathbb{E}(1_A \int_s^t |\varphi_u|^2 du).
\]

\[\blacksquare\]

### 4.2.2 Extension to square integrable progressive processes

Are there natural candidates for being integrands beyond step processes? The Itô isometry in Theorem 4.1 suggest to introduce the vector space $L^2$ of $d$-dimensional processes $(\varphi_t)_{t \geq 0}$ which are progressive for the filtration $(\mathcal{F}_t)_{t \geq 0}$, and such that for all $t \geq 0$,
\[
\mathbb{E}\left(\int_0^t |\varphi_s|^2 ds\right) = \int_0^t \mathbb{E}(|\varphi_s|^2) ds < \infty.
\]

Note that this quantity is nothing else but $\|\varphi\|_{[0,t]}^2 = \mathbb{E}(\int_{[0,t]} |\varphi_s|^2 ds)$. Note that $(B_t)_{t \geq 0} \in L^2$. We have $\mathcal{F} \subset L^2$ and the following lemma states a density on finite time intervals.

**Lemma 4.2** (Approximation). For all $\varphi \in L^2$, $t \geq 0$, $\varepsilon > 0$, there exists $\psi \in \mathcal{F}$ with
\[
\mathbb{E}\left(\int_0^t |\varphi_s - \psi_s|^2 ds\right) < \varepsilon.
\]

\[\text{In fact Pythagoras theorem in } L^2(\Omega, \mathcal{F}, \mathbb{P}): E(X_t^2 | \mathcal{F}_s) = E((X_t - X_s)^2 | \mathcal{F}_s) + E(X_s^2 | \mathcal{F}_s).
\]

---

\[48/98\]
4.2 Stochastic integral with respect to Brownian motion

Proof. We can assume that \( \varphi \) is bounded, since for all \( t \geq 0 \), by dominated convergence,

\[
\lim_{n \to \infty} \mathbb{E}\left( \int_0^t |\varphi_s - \varphi_s 1_{[n-1,n]}|^2 \, ds \right) = 0.
\]

We can moreover assume that \( \varphi \) vanishes outside a finite time interval since for all \( t \geq 0 \),

\[
\lim_{n \to \infty} \mathbb{E}\left( \int_0^t |\varphi_s - \varphi_s 1_{[0,n]}|^2 \, ds \right) = 0.
\]

We can assume furthermore that such a process is (left-)continuous since for all \( t \geq 0 \),

\[
\lim_{n \to \infty} \mathbb{E}\left( \int_0^t |\varphi_s - \delta_n \cdot \varphi_s 1_{[0,n]}|^2 \, ds \right) = 0.
\]

Finally it suffices to approximate such a process with elements of \( \mathcal{S}^2 \), namely

\[
\lim_{n \to \infty} \mathbb{E}\left( \int_0^t \left| \varphi_s - \sum_{i=0}^{\infty} \varphi_{i/n} 1_{(i/n,(i+1/n))} \right|^2 \, ds \right) = 0
\]

(makes sense since \( \varphi \) is bounded, left-continuous, and supported in a finite time interval).

We denote by \( \mathcal{M}_c^2 \) be the set of continuous square integrable martingales for \( (\mathcal{F}_t)_{t \geq 0} \). The elements of \( \mathcal{M}_c^2 \) are random variables on the Wiener space \( (\mathcal{C}([0,\infty), \mathbb{R}), \mathcal{F}, \mathbb{P}) \). Two elements \( X \) and \( Y \) of \( \mathcal{M}_c^2 \) are equal if and only if they are indistinguishable processes: \( \mathbb{P}(\forall t \geq 0 : X_t = Y_t) = 1 \).

**Theorem 4.3** (Brownian Itô stochastic integral of square integrable progressive processes). There exists a unique linear map \( I : \mathcal{L}^2 \to \mathcal{M}_c^2 \) denoted

\[
I(\varphi)_t = \int_0^t \varphi_s \, dB_s, \quad \varphi \in \mathcal{L}^2, \ t \geq 0,
\]

and called the Itô stochastic integral with respect to Brownian motion, such that

1. (step processes) for all \( \varphi \in \mathcal{S}^2 \) with decomposition \( \varphi_t = U_0 1_{(0,t]} + \sum_{i=0}^{n-1} U_i 1_{(t_{i+1},t_i)} \),

\[
I(\varphi)_t = \sum_{i=0}^{n-1} U_i \cdot (B_{t_{i+1}} - B_{t_i});
\]

2. (Itô isometry) for all \( \varphi \in \mathcal{L}^2 \),

\[
\mathbb{E}(I(\varphi)_t) = \mathbb{E}\left( \int_0^t |\varphi_s|^2 \, ds \right).
\]

Moreover this maps satisfies, for all \( \varphi \in \mathcal{L}^2 \), \( I(\varphi)_0 = 0 \) and for all \( t \geq 0 \),

\[
\mathbb{E}(I(\varphi)_t) = 0 \quad \text{and} \quad \langle I(\varphi) \rangle_t = \int_0^t |\varphi_s|^2 \, ds,
\]

which is stronger than the Itô isometry.

Proof. **Uniqueness.** Let \( I \) and \( I' \) be two linear maps from \( \mathcal{L}^2 \) to \( \mathcal{M}_c^2 \) satisfying the two properties of the theorem. Let \( \varphi \in \mathcal{L}^2 \). From Lemma 4.2, for all \( t \geq 0, n \geq 1 \), there exists \( \psi^{(n)} \in \mathcal{S}^2 \) with

\[
\mathbb{E}\left( \int_0^t |\varphi_s - \psi^{(n)}_s|^2 \, ds \right) \leq \frac{1}{2^n}.
\]

The linearity and isometry properties of \( I \) and \( I' \) and the equality of \( I \) and \( I' \) on \( \mathcal{S}^2 \) give

\[
I(\varphi)_t = \lim_{n \to \infty} I(\psi)_t = \lim_{n \to \infty} I'(\psi)_t = I'(\varphi)_t.
\]
Therefore $I(\varphi)$ and $I'(\varphi)$ are modifications of each other: $\forall t \geq 0$, $\mathbb{E}(X_t = Y_t) = 1$, and since they are continuous, they are indistinguishable, in other words $I = I'$ in $\mathcal{H}_t^\mathbb{C}$.

**Existence.** Let $\varphi \in \mathcal{L}^2$ and $\psi^{(n)} \in \mathcal{S}^2$ be as above. For all $t \geq 0$, let us set $X^{(n)}_t = I(\psi^{(n)})_t$. The linearity and isometry of $I$ on $\mathcal{S}^2$ and the definition of $\psi^{(n)}$ give, for all $n \geq t$,

$$E((X^{(n)}_t - X^{(n+1)}_t)^2) = E\left(\int_0^t |\psi^{(n)}_s - \psi^{(n+1)}_s|^2 ds\right) \leq \frac{4}{2^n}.
$$

Next, using the Doob maximal inequality of Theorem 2.14 we get

$$E\left(\sup_{s \in [0,t]} |X_s^{(n)} - X_s^{(n+1)}|^2\right) \leq 4E((X^{(n)}_t - X^{(n)}_t)^2) \leq \frac{16}{2^n}.
$$

Therefore,

$$\mathbb{E}\sum_{n \geq 0} \sup_{s \in [0,t]} |X^{(n)}_s - X^{(n+1)}_s| = \mathbb{E}\sum_{n \geq 0} \sup_{s \in [0,t]} |X^{(n)}_s - X^{(n+1)}_s| \leq \sum_{n \geq 0} \mathbb{E}\|X^{(n)}_s - X^{(n+1)}_s\|_2 < \infty
$$

and thus, almost surely,

$$\sum_{n \geq 0} \sup_{s \in [0,t]} |X^{(n)}_s - X^{(n+1)}_s| < \infty.
$$

It follows that the sequence $(X^{(n)})_n$ of continuous martingales converges almost surely and uniformly on every compact subset of $\mathbb{R}_+$, as $n \to \infty$, towards a continuous process $X = (X_t)_{t \geq 0}$. This process is an $(\mathcal{F}_t)_{t \geq 0}$-martingale since for all $0 \leq s < t$ and all $A \in \mathcal{F}_s$, we have

$$\mathbb{E}(1_A(X_t - X_s)) = \lim_{n \to \infty} \mathbb{E}(1_A(X^{(n)}_t - X^{(n)}_s)) = 0.
$$

The process $X$ depends only on $\varphi$ and does not depend on the particular sequence $(\psi^{(n)})_n$, chosen to construct it. Moreover, from the preceding estimates, it follows that for all $t \geq 0$, $\lim_{n \to \infty} X^{(n)}_t = X_t$ in $\mathbb{L}^2$, in particular $\mathbb{E}(X_t) = 0$ since $X^{(n)}_t$ is centered for all $n$, while

$$\mathbb{E}(X^2_t) = \lim_{n \to \infty} \mathbb{E}(X^{(n)}_t)^2 = \lim_{n \to \infty} \mathbb{E}\left(\int_0^t |\psi^{(n)}_s|^2 ds\right) = \mathbb{E}\left(\int_0^t |\varphi_s|^2 ds\right).
$$

We set $I(\varphi) = X$. The linearity of $I$ follows also from the construction above.

**Additional properties.** For all $t \geq 0$ we have $\lim_{n \to \infty} I(\psi^{(n)})_t = I(\varphi)_t$, in $\mathbb{L}^2$ and thus in $\mathbb{L}^1$ and since $\mathbb{E}(I(\psi^{(n)})_t) = 0$ for all $n \geq 1$, we obtain $\mathbb{E}(I(\varphi)_t) = 0$. Since almost surely the above convergence holds uniformly on $[0, t]$ and since $I(\psi^{(n)}_0) = 0$ for all $n \geq 1$ it follows that $I(\varphi)_0 = 0$. It remains to compute the increasing process $I(\varphi)$ by showing that $(X^2_t - \int_0^t |\varphi_s|^2 ds)_{t \geq 0}$ is a martingale, which is equivalent to show that for all $0 \leq s \leq t$ and all $A \in \mathcal{F}_s$, we have

$$\mathbb{E}\left(\left(X^2_t - X^2_s - \int_s^t |\varphi_u|^2 du\right)1_A\right) = 0.
$$

But

$$\left(X^2_t - X^2_s - \int_s^t |\varphi_u|^2 du\right)1_A = \lim_{n \to \infty} \left((X^{(n)}_t)^2 - (X^{(n)}_s)^2 - \int_s^t |\psi^{(n)}_u|^2 du\right)1_A,
$$

and from Theorem 4.1 the right hand side has zero mean for all $n \geq 0$.

\[\text{Remark 4.4 (Approximating a stochastic integral by finite sums). From the proof of Theorem 4.3, if } \varphi \in \mathcal{L}^2, \text{ and if } (\psi^{(n)})_{n \geq 1} \text{ is a sequence in } \mathcal{S}^2 \text{ such that for all } t \geq 0,
\]

$$\lim_{n \to \infty} \mathbb{E}\left(\int_0^t |\psi_s - \psi^{(n)}_s|^2 ds\right) = 0,
$$

then, denoting $\psi^{(n)}_t = U^{(n)}_0 1_0(t) + \sum_{i=0}^{m_n} U^{(n)}_{t^i(t)} 1_{t^i(t), t^i_{i+1}(t)}(t)$ for all $t \geq 0$,

$$\sum_{i=0}^{m_n} U^{(n)}_{t^i(t)} (B_{t^i(t)} - B_{t^i_{i+1}(t)}) = I(\psi^{(n)}_t) + \frac{1}{n-\infty} I(\varphi)_t = \int_0^t \varphi_s dB_s.
$$
4.2 Stochastic integral with respect to Brownian motion

**Theorem 4.5** (Properties of the Brownian Itô stochastic integral).

1. If \( \varphi \in \mathcal{L}^2 \) is deterministic then \( I(\varphi) \) is the Wiener integral of Chapter 3;

2. For all \( \varphi, \psi \in \mathcal{L}^2 \) we have, for all \( t \geq 0 \),
   \[
   \langle I(\varphi), I(\psi) \rangle_t = \int_0^t (\varphi_s \cdot \psi_s) \, ds;
   \]

3. For all \( \varphi \in \mathcal{L}^2 \) and all stopping time \( T \), we have \( (\varphi_t 1_{t \leq T})_{t \geq 0} \in \mathcal{L}^2 \) and
   \[
   \left( \int_0^t \varphi_s 1_{s \leq T} \, dB_s \right)_{t \geq 0} = \left( \int_0^{t \wedge T} \varphi_s \, dB_s \right)_{t \geq 0};
   \]

4. For all \( \varphi \in \mathcal{L}^2 \) which has almost surely left-continuous and locally bounded trajectories, for all \( t \geq 0 \), and all sequence \((\delta_n)_{n \geq 1}\) of subdivisions of \([0, t]\), \( \delta_n : 0 = t_0^n < \ldots < t_{m_n}^n = t \), with
   \[
   \lim_{n \to \infty} \max_{1 \leq k \leq m_n} |t_k^n - t_{k-1}^n| = 0,
   \]
   we have
   \[
   \sum_{k=0}^{m_n-1} \varphi_{t_k^n} (B_{t_{k+1}^n} - B_{t_k^n}) \xrightarrow{P} \int_0^t \varphi \, dB_s.
   \]

**Proof.**

1. The first statement follows from the coincidence of the definition of the Wiener integral of step functions and the Itô integral of step processes, together with \( L^2 \) convergence;

2. Follows from the formula \( \langle I(\varphi), I(\psi) \rangle_t = \int_0^t |\varphi_s|^2 \, ds \) for all \( \varphi \in \mathcal{L}^2 \) and \( t \geq 0 \), and from linearity of \( I \) and polarization namely \( 4I(\varphi)I(\psi) = I(\varphi + \psi) - I(\varphi - \psi) \) for all \( \varphi, \psi \in \mathcal{L}^2 \);

3. For all \( t \geq 0 \), \( \{t \leq T\} \in \mathcal{F}_t = \{T < t\}^c \) since \( T \) is a stopping time and \((\mathcal{F}_t)_{t \geq 0}\) is right-continuous, therefore the process \((1_{t \leq T})_{t \geq 0}\) is adapted, and since it is right-continuous, it follows from Theorem 2.1 that it is progressive. The product process \((\varphi_t 1_{t \leq T})_{t \geq 0}\) is thus also progressive. Moreover, for all \( t \geq 0 \),
   \[
   \mathbb{E}\left( \int_0^t |\varphi_s 1_{s \leq T}|^2 \, ds \right) \leq \mathbb{E}\left( \int_0^t |\varphi_s|^2 \, ds \right) < \infty.
   \]

Hence \( (\varphi_t 1_{t \leq T})_{t \geq 0} \in \mathcal{L}^2 \). It remains to prove the formula with \( T \). To this end, let us assume first that \( T \) takes its values in the finite set \( \{t_0, \ldots, t_0\} \) with \( 0 = t_0 < \cdots < t_0 \). Let us take \( \varphi \in \mathcal{F}^2 \). We can always refine the decomposition of the step function \( \varphi \) and assume without loss of generality that \( \varphi_t = U_0 1_{t_0}(t) + \sum_{i=0}^{n-1} U_i 1_{(t_i, t_{i+1})}(t) \). Then \( \varphi_{1_{t \leq T}} \in \mathcal{F}^2 \), since for all \( t > 0 \),
   \[
   \varphi_{1_{t \leq T}} = \sum_{k=0}^n 1_{T=t_k} \sum_{i=0}^{k-1} U_i 1_{(t_i, t_{i+1})}(t)
   = \sum_{i=0}^{n-1} \left( \sum_{k=i+1}^n 1_{T=t_k} \right) U_i 1_{(t_i, t_{i+1})}(t)
   = \sum_{i=0}^{n-1} U_i 1_{T=t_i} 1_{(t_i, t_{i+1})}(t)
   = \sum_{i=0}^{n-1} \tilde{U}_i 1_{(t_i, t_{i+1})}(t).
   \]

Moreover, by the same way,
   \[
   I(\varphi)_{t \wedge T} = \sum_{k=0}^n 1_{T=t_k} \sum_{i=0}^{n-1} U_i (B_{t_{i+1} \wedge T} - B_{t_i \wedge T})
   \]

51/98
Itô stochastic integral

Finally, if \( \phi \in \mathcal{S}^2 \), a stopping time \( T \) with values in a given finite set \( \{t_0, \ldots, t_n\} \), and we fix \( t \geq 0 \). Lemma 4.2 gives \( (\phi^{(n)})_{n \geq 1} \) in \( \mathcal{S}^2 \) with \( \lim_{n \to \infty} \mathbb{E}(\int_0^t |\phi_s - \phi^{(n)}_s|^2 \, ds) = 0 \). Then

\[
\mathbb{E}(||I(\phi)_{t \wedge T} - I(\phi^{(n)}_{t \wedge T})||^2) = \sum_{k=0}^n \mathbb{E}(||I(\phi)_{t \wedge t_k} - I(\phi^{(n)}_{t \wedge t_k})||^2 1_{T = t_k}) \\
\leq \sum_{k=0}^n \mathbb{E}(||I(\phi)_{t \wedge t_k} - I(\phi^{(n)}_{t \wedge t_k})||^2) \\
= \sum_{k=0}^n \mathbb{E}\left(\int_0^{t \wedge t_k} |\phi_s - \phi^{(n)}_s|^2 \, ds\right)_{n \to \infty} \rightarrow 0
\]

and

\[
\mathbb{E}(||I(\mathbf{1}_{1 \leq T}) - I(\mathbf{1}^{(n)}_{1 \leq T})||^2) = \mathbb{E}\left(\int_0^{T \wedge T} |\phi_s - \phi^{(n)}_s|^2 \, ds\right) \leq \mathbb{E}\left(\int_0^T |\phi(s) - \phi^{(n)}(s)|^2 \, ds\right)_{n \to \infty} \rightarrow 0.
\]

Finally, if \( \phi \in \mathcal{S}^2 \) and \( T \) is a general stopping time, we proceed as in Doob stopping theorem (Theorem 2.12), and approximate \( T \) by stopping times taking a finite number of values as

\[
T_{N,n} = N \mathbf{1}_{T > n} + \sum_{k=1}^{2^n} \frac{k-1}{2^n} N \mathbf{1}_{N \leq T < \frac{k+1}{2^n} N} \quad \text{as } N \to \infty
\]

and by dominated convergence \( \mathbb{E}(||I(\phi)_{t \wedge T_{N,n}} - I(\phi)_{t \wedge T_{N,n}}||^2) \to 0 \) while

\[
\mathbb{E}(||I(\mathbf{1}_{1 \leq T_{N,n}}) - I(\mathbf{1}^{(n)}_{1 \leq T_{N,n}})||^2) = \cdots = \mathbb{E}\left(\int_0^T |\phi_s|^2 |\mathbf{1}_{t \wedge T_{N,n}} - \mathbf{1}_{1 \leq T_{N,n}}|^2 \, ds\right)_{N,n \to \infty} \to 0.
\]

4. Set \( t \geq 0 \) and \( \delta_n \). Suppose that \( \phi \) is almost surely bounded: there exists a finite constant \( C \) such that \( \sup_{s \in [0,t]} |\phi_s| \leq C \) almost surely. Let us define \( \phi^{(n)} = \sum_{k=0}^{m_n-1} \phi^{(n)}_k \mathbf{1}_{t_k \leq t < t_{k+1}} \) for all \( n \geq 0 \). It is a good approximation of \( \phi \) in \( \mathcal{S}^2 \). Namely since \( \phi \) is left-continuous, we have \( \lim_{n \to \infty} |\phi_s - \phi^{(n)}_s| = 0 \) for almost all \( (s,\omega) \in [0,t] \times \Omega \), and by dominated convergence

\[
\lim_{n \to \infty} \mathbb{E}\left(\int_0^t |\phi_s - \phi^{(n)}_s|^2 \, ds\right) = 0.
\]

Now \( \phi^{(n)} \in \mathcal{S}^2 \) and by definition of the Itô integral,

\[
\sum_{k=0}^{m_n-1} \phi^{(n)}_k (B^n_{t_{k+1}} - B^n_{t_k}) = \int_0^t \phi^{(n)}_s \, dB_s.
\]

Hence by the linear Itô isometry we get

\[
\sum_{k=0}^{m_n-1} \phi^{(n)}_k (B^n_{t_{k+1}} - B^n_{t_k}) = \int_0^t \phi^{(n)}_s \, dB_s \xrightarrow{n \to \infty} \int_0^t \phi_s \, dB_s.
\]

If \( \phi \) is not almost surely bounded, then we localize with the stopping time

\[
T_N = \inf\{t \geq 0 : |\phi_t| \geq N\}, \quad N \geq 1.
\]
4.2 Stochastic integral with respect to Brownian motion

Now for all \( N \geq 1 \) and \( \epsilon > 0 \) we have
\[
P(\{I(\varphi^{(n)})_t - I(\varphi)_t \geq \epsilon\} \leq \mathbb{P}(T_N \leq t) + \mathbb{P}(\{I(\varphi^{(n)})_t - I(\varphi)_t \geq \epsilon; T_N \geq t\}
\]
which is in turn bounded above by
\[
P(T_N \leq t) + \mathbb{P}\left( \left| \sum_{k=0}^{m-1} \varphi_t^k \mathbf{1}_{\{|\varphi| \leq N(B_{t+1}^{n}) - B_{t+1}^n\}} - \int_0^t \varphi_s \mathbf{1}_{\{|\varphi| \leq N\}dB_s} \right| \geq \epsilon \right)
\]
where we have used on the event \( \{T_N \geq t\} \) the identity (previous item of the theorem!)
\[
\int_0^t \varphi_s dB_s = \int_0^{\tau \wedge T_N} \varphi_s dB_s = \int_0^t \varphi_s \mathbf{1}_{s \leq T_N} dB_s = \int_0^t \varphi_s \mathbf{1}_{\{|\varphi| \leq N\}dB_s}.
\]
Now, since \( \varphi \) is almost surely locally bounded, for all \( t \geq 0 \), \( M_t = \sup_{s \in [0,t]} |\varphi_s| < \infty \) almost surely, and thus, by dominated convergence,
\[
P(T_N \leq t) = \mathbb{P}\left( \sup_{s \in [0,1]} |\varphi_s| \geq N \right) = \mathbb{E}(1_{M_t \geq N}) \xrightarrow{N \to \infty} 0.
\]
On the other hand, by the Markov inequality and the first part of the proof above used with the almost surely bounded process \( \{\varphi_t \mathbf{1}_{|\varphi| \leq N}\}_{t \geq 0} \)
\[
P\left( \left| \sum_{k=0}^{m-1} \varphi_t^k \mathbf{1}_{|\varphi| \leq N(B_{t+1}^{n}) - B_{t+1}^n\} - \int_0^t \varphi_s \mathbf{1}_{|\varphi| \leq N\}dB_s} \right| \geq \epsilon \right) \leq \frac{\mathbb{E}((I(\varphi^{(n)})_t - I(\varphi)_t)^2)}{\epsilon^2} \xrightarrow{n \to \infty} 0.
\]
For fixed \( t \geq 0, \epsilon > 0 \), it remains to select first \( N \) large enough and then \( n \) large enough.

\[\square\]

**Example 4.6 (Itô stochastic integral).** We know now that the formula in Section 4.1 is licit:
\[
\int_0^t B_s dB_s = \frac{B_t^2 - t}{2}, \quad t \geq 0.
\]
We will learn how to compute many other integrals by using the Itô formula of Chapter 5. The quantity above is clearly a centered martingale. We can also easily check the Itô isometry as
\[
\mathbb{E}\left( \left( \int_0^t B_s dB_s \right)^2 \right) = \frac{\mathbb{E}(B_t^2 - 2tB_t^2 + t^2)}{4} = \frac{3t^2 - 2t^2 + t^2}{4} = \frac{t^2}{2} = \int_0^t \mathbb{E}(B_s^2)ds.
\]

**Example 4.7 (Itô integration by parts formula).** Let \( B = (B_t)_{t \geq 0} \) and \( W = (W_t)_{t \geq 0} \) be two one-dimensional Brownian motion for the same filtration \( (\mathcal{F}_t)_{t \geq 0} \), both issued from the origin. For all \( t > 0 \) and all subdivision \( 0 = t_0 < \cdots < t_n = t \), the identity
\[
\sum_{i} W_{t_i}(B_{t_{i+1}} - B_{t_i}) = \sum_{i} (W_{t_{i+1}}B_{t_{i+1}} - W_{t_i}B_{t_i}) - \sum_{i} (W_{t_{i+1}} - W_{t_i})(B_{t_{i+1}} - B_{t_i}) - \sum_{i} B_{t_i}(W_{t_{i+1}} - W_{t_i})
\]
gives the integration by parts formula (will work for arbitrary continuous martingales)
\[
\int_0^t W_s dB_s = W_t B_t - [W, B]_t - \int_0^t B_s dW_s.
\]
We recover the formula of Example 4.6 when \( B = W \). On the contrary, if \( B \) and \( W \) are independent then
4 Itô stochastic integral

\[ [W, B]_t = 0. \text{ Namely, denoting } \Delta^B_t = B_{t+1} - B_t \text{ and } \Delta^W_t = W_{t+1} - W_t, \text{ we have } \mathbb{E}(\sum_i \Delta^W_i \Delta^B_i) = 0 \text{ while} \]

\[ \mathbb{E}\left( \left( \sum_i \Delta^W_i \Delta^B_i \right)^2 \right) = \sum_i \mathbb{E}(\Delta^B_i)^2 \mathbb{E}(\Delta^W_i)^2 = \sum_i (t_{i+1} - t_i)^2 \leq t \max(t_{i+1} - t_i) \to 0. \]

Alternatively we may use \([B, W]_t = \langle B, W \rangle_t\) and note that \(\langle B, W \rangle_t = 0\) since for all \(0 \leq s \leq t,\)

\[ \mathbb{E}(B_s W_t | \mathcal{F}_s) = \mathbb{E}(B_s - B_t)(W_t - W_s) + B_t W_s + B_s W_t - B_s W_s | \mathcal{F}_s) = 0. \]

4.2.3 Further extension by localization

Let \(\mathcal{L}_d^2\) be the set of \(d\)-dimensional progressive processes \(\varphi = (\varphi_t)_{t \geq 0}\) such that

\[ \text{for all } t \geq 0, \text{ almost surely, } \int_0^t |\varphi_s|^2 ds < \infty. \]

We extend below the Itô stochastic integral with respect to Brownian motion to this larger set of integrands. The stochastic integral is no longer a martingale. It is however a local martingale issued from the origin.

Recall that a process \(M = (M_t)_{t \geq 0}\) is a continuous local martingale for \((\mathcal{F}_t)_{t \geq 0}\) if it is real, continuous, adapted, and there exists a sequence \((T_n)_{n \geq 0}\) of stopping times such that

- almost surely \(T_n \not\to +\infty \text{ as } n \to \infty;\)
- for all \(n \geq 0\) the process \((M_{T_n \wedge T})_{t \geq 0}\) is a continuous square integrable \((\mathcal{F}_t)_{t \geq 0}\)-martingale.

We denote by \(\mathcal{M}_{\text{c,loc}}\) the set of continuous local martingales. Note that \(\mathcal{M}_d^2 \subsetneq \mathcal{M}_{\text{c,loc}}\). The reduction to square integrable martingales via stopping times is called localization and can be seen as a sort of probabilistic cutoff. Note also that no integrability is guaranteed for \(M_t\) for a fixed deterministic \(t \geq 0.\)

**Lemma 4.8** (Boundedness by localization). Let \(M \in \mathcal{M}_{\text{c,loc}}\) with \(M_0 = 0\), and, for all \(n \geq 0,\)

\[ T_n = \inf\{t \geq 0 : |M_t| \geq n\} \quad \text{with as usual } \inf\emptyset = +\infty. \]

Then almost surely \(T_n \not\to +\infty \text{ as } n \to \infty,\) and \((M_{T_n \wedge T})_{t \geq 0}\) is a continuous and bounded martingale with

\[ \sup_{t \geq 0} |M_{T_n \wedge T_t}| \leq n. \]

In particular for all \(n \geq 0\) the family \((M_{T_n \wedge T})_{t \geq 0}\) is uniformly integrable, and \(M \in \mathcal{M}_{\text{c,loc}}.\)

**Proof.** For all \(n \geq 0,\) the random variable \(T_n\) is a stopping time, and almost surely \(T_n \not\to +\infty \text{ as } n \to \infty.\) Moreover, for all \(n \geq 0,\) by definition of \(T_n,\) we have \(\sup_{t \geq 0} |M_{T_n \wedge T}| \leq n.\) Since \(M \in \mathcal{M}_{\text{c,loc}},\) there exists a sequence \((S_k)_{k \geq 0}\) of stopping times such that almost surely \(S_k \not\to +\infty \text{ as } k \to \infty,\) and for all \(k \geq 0, M_{S_k} = (M_{T_n \wedge T})_{t \geq 0}\) is a continuous square integrable martingale. For all \(n, k \geq 0,\) by the Doob optional stopping (Theorem 2.24), \((M_{S_k})_{T_n} = (M_{T_n \wedge T_k})_{t \geq 0}\) is a continuous martingale, and \(\sup_{t \geq 0} |M_{T_n \wedge T_k}| \leq n.\) In particular, for all \(0 \leq s \leq t,\)

\[ \mathbb{E}(M_{T_n \wedge T_k} | \mathcal{F}_s) = M_{T_n \wedge T_k}. \]

Now for all fixed \(n \geq 0,\) by dominated convergence, as \(k \to \infty,\) we get \(\mathbb{E}(M_{T_n \wedge T_k} | \mathcal{F}_s) = M_s,\) hence \((M_{T_n \wedge T_k})_{t \geq 0}\) is a (square integrable) martingale. Finally the continuity of this martingale follows from the one of \(M.\)

**Lemma 4.9** (Vector space). The sets \(\mathcal{M}_{\text{c,loc}}\) and \(\{M \in \mathcal{M}_{\text{c,loc}} : M_0 = 0\}\) are real vector space.

**Proof.** The first property follows from the second, that we prove now. Let \(M, M' \in \mathcal{M}_{\text{c,loc}}.\) By Lemma 4.8, there exist sequences of stopping times \((T_n)_{n \geq 1}\) and \((T'_n)_{n \geq 1}\) such that almost surely \(T_n, T'_n \not\to +\infty,\) and such that for all \(n \geq 0\) the processes \((M_{T_n \wedge T_n})_{t \geq 0}\) and \((M'_{T'_n \wedge T'_n})_{t \geq 0}\) are bounded martingales. Now for all \(n \geq 0,\) the random variable \(S_n = T_n \wedge T'_n\) is a stopping time, and almost surely \(S_n \not\to +\infty.\) By the Doob stopping theorem (Theorem 2.24), for all \(n \geq 0,\) the processes \((M_{T_n \wedge S_n})_{t \geq 0} = (M_{T_n \wedge T_n})_{t \geq 0}\) and \((M'_{T'_n \wedge S_n})_{t \geq 0} = (M'_{T'_n \wedge T'_n})_{t \geq 0}\) are martingales. Moreover they inherit the continuity and boundedness from \((M_{T_n \wedge T_n})_{t \geq 0}\) and \((M'_{T'_n \wedge T'_n})_{t \geq 0}.$
4.3 Stochastic integral with respect to continuous martingales bounded in $L^2$

**Theorem 4.10** (Extension of Itô integral to $L^2_0$). There exists a unique linear map $I: L^2_0 \to \mathcal{M}_{c,loc}$, such that for all $\varphi \in L^2_0$ and for all stopping time $T$ such that

$$(\varphi_t 1_{t \leq T})_{t \geq 0} \in L^2$$

in other words

$$E \left( \int_0^T |\varphi_s|^2 \, ds \right) < \infty,$$

we have, for all $t \geq 0$,

$$I(\varphi)_{t \wedge T} = \int_0^t \varphi_s 1_{0 \leq s \leq T} \, dB_s.$$ 

For all $\varphi \in L^2_0$ and all $t \geq 0$, we still denote

$$I(\varphi)_t = \int_0^t \varphi_s \, dB_s.$$ 

**Proof.** Let us first construct $I(\varphi)$ by approximation, actually by cutoff. Let $\varphi \in L^2_0$. For all $n \geq 0$, we define

$$T_n = \inf \left\{ t \geq 0 : \int_0^t |\varphi_s|^2 \, ds \geq n \right\} \quad \text{with usual } \inf \emptyset = +\infty.$$ 

We can check easily that $(T_n)_{n \geq 0}$ is a sequence of stopping times such that almost surely $T_n \nearrow +\infty$ as $n \to \infty$. Let us define $\varphi^{(n)}_s = \varphi_s 1_{0 \leq s \leq T_n}$. By definition of $\varphi$ and $T_n$, we see that $E(\int_0^T |\varphi^{(n)}_s|^2 \, ds) \leq t \wedge n$, thus $\varphi^{(n)} \in L^2$. Hence for all $t \geq 0$ the following stochastic integral is well defined:

$$X^{(n)}_t = \int_0^t \varphi^{(n)}_s \, dB_s = \int_0^t \varphi_s 1_{0 \leq s \leq T_n} \, dB_s.$$ 

Moreover Theorem 4.5 gives, for all $0 \leq m \leq n$, almost surely, for all $t \geq 0$,

$$X^{(m)}_{t \wedge T_n} = X^{(m \wedge n)}_t = X^{(m)}_t.$$ 

Now since almost surely $T_n \to +\infty$ as $n \to \infty$, almost surely, for all $t \geq 0$, the sequence of random variables $(X^{(m)}_t)_{m \geq 0}$ is stationary. This allows us to define the process $X = (X_t)_{t \geq 0}$ outside a negligible event by setting, for all $t \geq 0$, $X_t = \lim_{m \to \infty} X^{(m)}_t$. With this definition, almost surely, $X_{t \wedge T_m} = X^{(m)}_t$ for all $m \geq 0$ and $t \geq 0$. Let $T$ be a stopping time such that $E(\int_0^T |\varphi_s|^2 \, ds) < \infty$, in other words $(\varphi_t 1_{t \leq T})_{t \geq 0} \in L^2$. This allows to define, for all $t \geq 0$, $Y_t = \int_0^t \varphi_s 1_{s \leq T} \, dB_s$. Thanks to Theorem 4.5, we have, for all $t \geq 0$ and all $m \geq 0$,

$$Y_{t \wedge T_m} = \int_0^t \varphi_s 1_{s \leq T \wedge T_m} \, dB_s = X^{(m)}_{t \wedge T} \quad \text{and thus } \quad Y_t = \lim_{m \to \infty} Y_{t \wedge T_m} = \lim_{m \to \infty} X^{(m)}_{t \wedge T} = X_{t \wedge T}.$$ 

Setting $I(\varphi) = X$ gives a map from $L^2_0 \to \mathcal{M}_{c,loc}$ with the desired property, which is linear, and unique. \hfill $\blacksquare$

### 4.3 Stochastic integral with respect to continuous martingales bounded in $L^2$

Up to know the integrator in our Itô stochastic integral is Brownian motion. In view of the generalization to an integrator which is a continuous square integrable martingale and even a continuous local martingale, we consider in this section, in a first step, the Itô stochastic integral for an integrator which is a continuous martingale bounded in $L^2$. Recall that this is not the case of Brownian motion.

Recall that the set $\mathbb{M}^2$ of continuous martingales bounded in $L^2$ and issued from the origin is a Hilbert space (Corollary 2.20). Such martingales are centered. We would like to define, for $M \in \mathbb{M}^2$, the integral

$$\int_0^t \varphi_s \, dM_s, \quad t \geq 0,$$

for reasonable integrands $\varphi$. Since the integrator $M$ is a one-dimensional process, it is natural to consider a one-dimensional integrand $\varphi$. Actually we avoid as much as possible to deal with general vector valued
martingales in this introductory course. Inspired by what we do for the Brownian motion integrator, and to simplify as much as possible, we denote by $\mathcal{L}^2(M)$ the set of progressive real processes $(\varphi_t)_{t \geq 0}$ such that

$$\mathbb{E}\left(\int_0^\infty \varphi_t^2 \, d\langle M \rangle_s\right) < \infty.$$ 

In this formula, the integral on $[0, \infty)$ is understood at all fixed $\omega$ as the limit as $t \to \infty$ of the integral on $[0, t]$ with respect to the increasing and thus bounded variation function $s \in [0, t] \to \langle M \rangle_s$, see Theorem 1.5.

Since the (random) process $\langle M \rangle$ may be constant on some intervals of time, we define the equivalence relation $\sim$ on $\mathcal{L}^2(M)$ by setting $\varphi \sim \psi$ iff $\mathbb{E}(\int_0^\infty (\varphi_s - \psi_s)^2 \, d\langle M \rangle_s) = 0$, and we consider the quotient space $\mathcal{L}^2(M) = \mathcal{L}^2(M) / \sim$, still denoted $\mathcal{L}^2(M)$ for convenience. In fact, with this definition and convention,

$$\mathcal{L}^2(M) = L^2(\Omega \times \mathbb{R}_+, \mathcal{F}, \mu),$$

where $\mathcal{F}$ is the progressive $\sigma$-field (Theorem 2.1) and where $\mu$ is the finite measure defined for all $A \in \mathcal{F}$ by

$$\mu(A) = \int_\Omega \int_0^\infty 1_A(\omega, s) \, d\mathbb{P}(\omega) \, d\langle M \rangle_s(\omega).$$

Its total mass is $\mathbb{E}(\langle M \rangle_\infty) = \|M\|^2_{L^2}$. Note that the increasing process $\langle M \rangle$ is random in general, and this makes a notable difference with the Brownian motion case studied before for which $\langle B \rangle_t = t$ it is deterministic. The scalar product and the norm of the Hilbert space $\mathcal{L}^2(M)$ are given respectively by

$$\langle \varphi, \psi \rangle_{\mathcal{L}^2(M)} = \mathbb{E}\left(\int_0^\infty (\varphi_s \psi_s) \, d\langle M \rangle_s\right)$$

and

$$\|\varphi\|^2_{\mathcal{L}^2(M)} = \mathbb{E}\left(\int_0^\infty \varphi_s^2 \, d\langle M \rangle_s\right).$$

Let $\mathcal{A}_b$ be the vector space of bounded step processes, i.e. with bounded $U_i$'s. Then $\mathcal{A}_b \subset \mathcal{L}^2(M)$.

**Lemma 4.11** (Approximation or density). Let $M \in \mathbb{M}^2$. The vector space $\mathcal{A}_b$ of bounded step processes is dense in $\mathcal{L}^2(M)$. In other words, for all $\varphi \in \mathcal{L}^2(M)$, and all $\varepsilon > 0$, there exists $\psi \in \mathcal{A}_b$ with

$$\mathbb{E}\left(\int_0^\infty (\varphi_s - \psi_s)^2 \, d\langle M \rangle_s\right) < \varepsilon.$$ 

**Proof.** Since $\mathcal{L}^2(M)$ is a Hilbert space, it suffices to show that for all $\varphi \in \mathcal{L}^2(M)$, if $\langle \varphi, \psi \rangle_{\mathcal{L}^2(M)} = 0$ for all $\psi \in \mathcal{A}_b$, then $\varphi = 0$. Let $\varphi \in \mathcal{L}^2(M)$ be such an element, and set, for all $t \geq 0$,

$$\Phi_t = \int_0^t \varphi_s \, d\langle M \rangle_s.$$ 

The integral in the right hand side makes sense since by the Cauchy–Schwarz inequality,

$$\mathbb{E}\left(\int_0^t |\varphi_s| \, d\langle M \rangle_s\right) \leq \left(\mathbb{E}\left(\int_0^t \varphi_s^2 \, d\langle M \rangle_s\right)\right)^{1/2} \left(\mathbb{E}(\langle M \rangle_\infty)\right)^{1/2}$$

and the right hand side is finite since $M \in \mathbb{M}^2$ and $\varphi \in \mathcal{L}^2(M)$. This shows that almost surely, for all $t \geq 0$,

$$\int_0^t |\varphi_s| \, d\langle M \rangle_s < \infty.$$ 

The process $(\Phi_t)_{t \geq 0}$ has finite variations and $\Phi_t \in L^1$ for all $t \geq 0$. Now, let $0 \leq s \leq t$ and let $Z$ be a bounded $\mathcal{F}_s$-measurable random variable. Since $(\psi_u)_{u \geq 0} = (Z 1_{u \in (s, t)})_{u \geq 0} \in \mathcal{A}_b$, we have $\langle \varphi, \psi \rangle_{\mathcal{L}^2(M)} = 0$, namely

$$\mathbb{E}\left(Z \int_s^t \varphi_u \, d\langle M \rangle_u\right) = 0.$$ 

Therefore $\mathbb{E}(Z(\Phi_t - \Phi_s)) = 0$. Since $Z$ is an arbitrary bounded $\mathcal{F}_s$-measurable random variable and since $\Phi_t \in L^1$ for all $t \geq 0$, it follows that $(\Phi_t)_{t \geq 0}$ is a martingale for $(\mathcal{F}_t)_{t \geq 0}$. On the other hand $\Phi$ is a finite variation process, and therefore, thanks to Lemma 2.16, we get $\Phi = 0$. Having in mind the initial definition on $\Phi$, this means that almost surely the signed measure with density $\varphi_s$ with respect to $d\langle M \rangle_s$ is zero. But this is possible only if $\varphi_s = 0$, $d\langle M \rangle_s$ almost everywhere, in other words only if $K = 0$ in $\mathcal{L}^2(M)$, as expected. \[\blacksquare\]
4.3 Stochastic integral with respect to continuous martingales bounded in $L^2$

**Theorem 4.12** (Itô stochastic integral with respect to elements of $\mathcal{M}^2$). Let $M \in \mathcal{M}^2$. There exists a unique linear map $I_M : L^2(M) \to \mathcal{M}^2$, denoted for all $\varphi \in L^2(M)$ and all $t \geq 0$

$$I_M(\varphi)_t = \int_0^t \varphi_s \, dM_s,$$

and called the Itô stochastic integral with respect to elements of $\mathcal{M}^2$, such that

1. for all $\varphi \in \mathcal{S}_b$ with decomposition $\varphi_t = U_0 \mathbf{1}_0(t) + \sum_{i=0}^{n-1} U_i \mathbf{1}_{(t_i, t_{i+1})}(t)$, we have

$$I_M(\varphi)_t = \sum_{i=0}^{n-1} U_i (M_{t_{i+1} \wedge t} - M_{t_i \wedge t});$$

2. the map $I_M$ is an isometry, namely for all $\varphi \in L^2(M)$,

$$E \left( \left( \int_0^\infty \varphi_s \, dM_s \right)^2 \right) = E \left( \left\langle \int_0^\infty \varphi_s \, dM_s \right\rangle \right) = \| I_M(\varphi) \|_{L^2(\mathcal{M})}^2 = \| \varphi \|_{L^2(M)}^2 = E \left( \int_0^\infty \varphi_s^2 \, d\langle M \rangle_s \right).$$

Moreover, for all $\varphi \in L^2(M)$, $I_M(\varphi)$ is the unique element of $\mathcal{M}^2$ such that for all $N \in \mathcal{M}^2$ and $t \geq 0$,

$$\langle I_M(\varphi), N \rangle_t = \int_0^t \varphi_s \, d\langle M, N \rangle_s.$$ 

Furthermore, for all $\varphi \in L^2(M)$ and all stopping time $T$ and all $t \geq 0$,

$$\int_0^{t \wedge T} \varphi_s \, dM = \int_0^t \varphi_s \mathbf{1}_{s \leq T} \, dM = \int_0^t \varphi_s \, dM^{\wedge T}.$$

Note that by definition of $\mathcal{M}^2$ the martingale $I(\varphi)$ is centered and issued from the origin.

**Proof.** The proof follows closely the lines of the proof for the Brownian motion integrator.

First of all the definition of $I_M(\varphi)$ for $\varphi \in \mathcal{S}_b$ does not depend on the decomposition chosen for $\varphi$. It is immediate to check that the map $I_M$ is linear on $\mathcal{S}_b$. Let us prove that it is an isometry on $\mathcal{S}_b$. For all $\varphi \in \mathcal{S}_b$ and all $i \in \{0, \ldots, n-1\}$, we have $M^{(i)} = U_i (M_{t_{i+1} \wedge \bullet} - M_{t_i \wedge \bullet}) \in \mathcal{M}^2$, and thus $I_M(\varphi) = \sum_{i=0}^{n-1} M^{(i)} \in \mathcal{M}^2$. Moreover $M^{(i)}$ is constant outside the interval $(t_i, t_{i+1})$, and a computation reveals that for all $i, j \in \{0, \ldots, n-1\}$,

$$\langle M^{(i)}, M^{(j)} \rangle = H^2_t ((M)_{t_{i+1} \wedge \bullet} - (M)_{t_i \wedge \bullet}) \mathbf{1}_{i=j},$$

which gives

$$\langle I_M(\varphi) \rangle = \sum_{i=0}^{n-1} \langle M^{(i)} \rangle = \sum_{i=0}^{n-1} H^2_t ((M)_{t_{i+1} \wedge \bullet} - (M)_{t_i \wedge \bullet}) = \int_0^t \varphi_s^2 \, d\langle M \rangle_s.$$

The isometry property follows, more precisely

$$\| I_M(\varphi) \|_{\mathcal{M}^2}^2 = E(\langle I_M(\varphi) \rangle) = E \left( \int_0^\infty \varphi_s^2 \, d\langle M \rangle_s \right) = \| \varphi \|_{L^2(M)}^2.$$

By Lemma 4.11, $\mathcal{S}_b$ is dense in the Hilbert space $\mathcal{M}^2$. Thus the linear isometry $I_M$ extends uniquely to $\mathcal{M}^2$.

Let $N \in \mathcal{M}^2$. For all $\varphi \in L^2(M)$, the Kunita–Watanabe inequality (Corollary 2.19) gives

$$E \left( \int_0^\infty |\varphi_s| \, d\langle M, N \rangle_s \right) \leq \| \varphi \|_{L^2(M)} \| N \|_{\mathcal{M}^\infty}.$$ 

It follows that $\int_0^\infty \varphi_s \, d\langle M, N \rangle_s$ is well defined and belongs to $L^1$. If $\varphi \in \mathcal{S}_b$ then

$$\langle I_M(\varphi), N \rangle = \sum_{i=0}^{n-1} \langle M^{(i)}, N \rangle.$$
and \( \langle M^t, N \rangle = H_t(\langle M, N \rangle_{t \wedge \lambda}, \langle M, N \rangle_{t \wedge \lambda}) \), and thus

\[
\langle I_M(\varphi), N \rangle = \sum_{i=0}^{n-1} \langle M_i, N \rangle = \sum_{i=0}^{n-1} H_i(\langle M, N \rangle_{t_i \wedge \lambda}, \langle M, N \rangle_{t_i \wedge \lambda}) = \int_0^t \varphi_s d\langle M, N \rangle_s.
\]

This gives the desired formula when \( \varphi \in \mathcal{S}_b \). But the map \( X \in \mathbb{M}^2 \to \langle X, N \rangle_{\infty} \in \mathbb{L}^1 \) is continuous since by the Kunita–Watanabe inequality (Corollary 2.19), we have

\[
E(|\langle X, N \rangle_{\infty}|) \leq E(\langle X \rangle_{\infty})^{1/2} E(\langle N \rangle_{\infty})^{1/2} = \|N\|_{\mathbb{M}^2} \|X\|_{\mathbb{M}^2}.
\]

If now \( \varphi \in \mathcal{L}^2(M) \) and \( \varphi = \lim_{n \to \infty} \varphi^{(n)} \) in \( \mathcal{L}^2(M) \) for a sequence \( (\varphi^{(n)})_{n \geq 0} \) in \( \mathcal{S}_b \), then

\[
\langle I_M(\varphi), N \rangle_{\infty} = \lim_{n \to \infty} \langle I_M(\varphi^{(n)}), N \rangle_{\infty} = \lim_{n \to \infty} \int_0^\infty \varphi_s^{(n)} d\langle M, N \rangle_s = \int_0^\infty \varphi_s d\langle M, N \rangle_s,
\]

where we have used for the last equality the inequality

\[
E\left( \left| \int_0^\infty (\varphi_s^{(n)} - \varphi_s) d\langle M, N \rangle_s \right| \right) \leq \|\varphi^{(n)} - \varphi\|_{\mathcal{L}^2(M)} \|N\|_{\mathbb{M}^2},
\]

which follows from the Kunita–Watanabe inequality (Corollary 2.19). We have obtained

\[
\langle I_M(\varphi), N \rangle_{\infty} = \int_0^\infty \varphi_s d\langle M, N \rangle_s.
\]

Finally, to replace \( \infty \) by an arbitrary \( t \geq 0 \), we can replace \( N \) by the stopped process \( N^t = (N_s1_{s \leq t})_{s \geq 0} \).

Moreover this formula characterizes \( I_M(\varphi) \) in \( \mathbb{M}^2 \). Indeed, if \( X \in \mathbb{M}^2 \) satisfies the same formula, then, for all \( N \in \mathbb{M}^2 \), \( \langle I_M(\varphi) - X, N \rangle = 0 \), and taking \( X = I_M(\varphi) - X \) gives \( \langle I_M(\varphi) - X \rangle = 0 \), thus \( X = I_M(\varphi) \).

Furthermore, for all \( N \in \mathbb{M}^2 \) and \( \varphi \in \mathcal{L}^2(M) \), and all stopping time \( T \), the properties of the angle bracket (Corollary 2.18) of two continuous square integrable martingales give, for all \( t \geq 0 \),

\[
\langle I_M(\varphi)^T, N \rangle_T = \langle I_M(\varphi), N \rangle_{T \wedge T} = \int_0^{T \wedge T} \varphi_s d\langle M, N \rangle_s = \int_0^T \varphi_s 1_{s \leq T} d\langle M, N \rangle_s,
\]

and similarly, we have, for all \( t \geq 0 \),

\[
\langle I_M^T(\varphi), N \rangle_T = \int_0^T \varphi_s d\langle M^T, N \rangle_s = \int_0^T \varphi_s d\langle M, N \rangle_s^T = \int_0^T \varphi_s 1_{s \leq T} d\langle M, N \rangle_s.
\]

These formulas, together with the preceding characterization of \( I_M \), give the desired formulas with \( T \). ■

**Corollary 4.13** (Angle bracket, moments, associativity).

1. For all \( M \in \mathbb{M}^2 \), \( \varphi \in \mathcal{L}^2(M) \), and all \( t \geq 0 \),

\[
E\left( \int_0^t \varphi_s dM_s \right) = 0,
\]

2. For all \( M, N \in \mathbb{M}^2 \), all \( \varphi \in \mathcal{L}^2(M) \), all \( \psi \in \mathcal{L}^2(N) \), and all \( t \geq 0 \),

\[
\left\langle \int_0^t \varphi_s dM_s, \int_0^t \psi_s dN_s \right\rangle = \int_0^t (\varphi_s \psi_s) (M, N)_s,
\]

in other words

\[
E\left( \left( \int_0^t \varphi_s dM_s \right) \left( \int_0^t \psi_s dN_s \right) \right) = E\left( \int_0^t \varphi_s \psi_s d\langle M, N \rangle_s \right);
\]

3. For all \( M \in \mathbb{M}^2 \), all \( \varphi \in \mathcal{L}^2(M) \), and all progressive process \( \psi \), we have

\[
\varphi \psi \in \mathcal{L}^2(M) \quad \text{iff} \quad \psi \in \mathcal{L}^2(I_M(\varphi)), \quad \text{and in this case} \quad I_M(\varphi \psi) = I_{I_M(\varphi)}(\psi).
\]
4.4 Stochastic integral with respect to continuous local martingales

Proof.

1. According to Theorem 4.12, we have \( I_M(\varphi) \in \mathbb{M}^2 \), and it is thus in particular a martingale issued from the origin, hence \( I_M(\varphi)_0 = 0 \) and for all \( t \geq 0 \), \( \mathbb{E}(I_M(\varphi)_t) = \mathbb{E}(I_M(\varphi)_0) = 0 \);

2. By polarization, we can assume without loss of generality that \( M = N \). The property follows then from the isometry property of Theorem 4.12 applied with \( (\varphi, 1_{s \geq t})_{s \geq 0} \) together with the stopping time property in Theorem 4.12 with the deterministic stopping time \( t \).

3. From Theorem 4.12, from the second property of the theorem (with \( N = M \) and \( \psi = \varphi \)), we get,

\[
\mathbb{E}\left( \int_0^\infty (\varphi_s^2 \psi_s^2) d\langle M, M \rangle_s \right) = \mathbb{E}\left( \int_0^\infty \varphi_s^2 \psi_s^2 d\langle M, M \rangle_u \right) = \mathbb{E}\left( \int_0^\infty \psi_s^2 d\langle I_M(\varphi) \rangle_s \right),
\]

which gives that \( \varphi \psi \in \mathcal{L}^2(\mathcal{M}) \) iff \( \psi \in \mathcal{L}^2(I_M(\varphi)) \). Next, by Theorem 4.12, for all \( N \in \mathbb{M}^2 \) and all \( t \geq 0 \),

\[
\langle I_M(\varphi \psi), N \rangle_t = \int_0^t (\varphi_s^2 \psi_s^2) d\langle M, N \rangle_s = \int_0^t \varphi_s^2 d\langle M, N \rangle_u = \int_0^t \psi_s d\langle I_M(\varphi), N \rangle_s,
\]

which implies, by the characterization via \( N \) property of Theorem 4.12, that \( I_M(\varphi \psi) = I_{I_M(\varphi)}(\psi) \).

\[
\square
\]

4.4 Stochastic integral with respect to continuous local martingales

Up to now we have constructed the Itô integral for an integrator which is either Brownian motion or an arbitrary continuous martingale bounded in \( \mathbb{L}^2 \). Our aim now is to consider an integrator which is an arbitrary continuous local martingale. Let \( M \in \mathcal{M}_{c,loc} \) with \( M_0 = 0 \). The notion of increasing process of a local martingale is natural, see Lemma 4.14. By analogy with what we did before, we consider respectively the sets \( \mathcal{L}_0^2(M) \) and \( \mathcal{L}^2(M) \) of \underline{progressive real processes} \( (\varphi_t)_{t \geq 0} \) such that

for all \( t \geq 0 \), almost surely, \( \int_0^t \varphi_s^2 d\langle M \rangle_s < \infty \), and respectively \( \mathbb{E}\left( \int_0^\infty \varphi_s^2 d\langle M \rangle_s \right) < \infty \),

quotteded by the equivalence relation related to equality \( \langle M \rangle \) almost everywhere.

Lemma 4.14 (Increasing process or angle bracket of local martingales). For all \( M, N \in \mathcal{M}_{c,loc} \) with \( M_0 = N_0 = 0 \), there exists a unique continuous bounded variation process denoted \( \langle M, N \rangle_t \) such that

1. \( \langle M, N \rangle_0 = 0 \);

2. \( (M_t, N_t - \langle M, N \rangle_t)_{t \geq 0} \in \mathcal{M}_{c,loc} \).

We have \( \langle M, N \rangle = \frac{1}{2} \langle (M+N) - (M-N) \rangle \) where \( \langle M \rangle = \langle M, M \rangle \) and \( \langle M \rangle \) is an increasing process. Moreover for all \( t \geq 0 \) if \( (\delta_n)_{n \geq 1} \) is a sequence of subdivisions of \( [0, t] \) of the form \( \delta_n \colon 0 = t_0 < \cdots < t_{m_n} = t \), with

\[
\lim_{n \to \infty} \max_{1 \leq k \leq m_n} (t_k^n - t_{k-1}^n) = 0,
\]

then

\[
[M, N]_t = \lim_{n \to \infty} \sum_{k=1}^{m_n} (M_{t_k^n} - M_{t_{k-1}^n})(N_{t_k^n} - N_{t_{k-1}^n}) = \langle M, N \rangle_t.
\]

We have \( [M, N] = \frac{1}{2} ([M + N] - [M - N]) \) where \([M] = [M, M] \) and \([M] \) is the quadratic variation of \( M \).

Proof. The case \( M \neq N \) follows from the case \( M = N \) by polarization, as for the case of martingales. Let us assume that \( M = N \). By localization with a stopping time, one can reduce the problem to a bounded martingale, which is in particular square integrable, and for which the notion of increasing process is well defined (Theorem 2.15) and coincides with the notion of quadratic variation. For instance let \( M, N \in \mathcal{M}_{c,loc} \). We can
find a sequence \((T_n)_{n \geq 0}\) of stopping times such that \(T_n \nearrow +\infty\) as \(n \to \infty\) and for all \(n \geq 0\), \(M^{T_n} = (M_{t \land T_n})_{t \geq 0}\) and \(N^{T_n} = (N_{t \land T_n})_{t \geq 0}\) are square integrable continuous martingales. A consequence of the uniqueness of the increasing process in Theorem 2.15 is that for all \(0 \leq n \leq m\) and all \(t \geq 0\),

\[
\langle M^{T_m}, N^{T_n} \rangle_{t \land T_n} = \langle M^{T_n}, N^{T_n} \rangle_t.
\]

It follows that there exists a unique continuous process with finite variations and issued from the origin, denoted \(\langle M, N \rangle\) such that for all \(t \geq 0\) and all \(n \geq 0\), \(\langle M, N \rangle_{t \land T_n} = \langle M^{T_n}, N^{T_n} \rangle_t\). We then set \(\langle M \rangle = \langle M, M \rangle\). ■

**Theorem 4.15** (Itô stochastic integral with respect to continuous local martingales). Let \(M \in \mathcal{M}_{c,loc}\) such that \(M_0 = 0\). For all \(\varphi \in \mathcal{L}_0^2(M)\), there exists a unique \(I_M(\varphi) \in \mathcal{M}_{c,loc}\) issued from the origin and such that for every \(N \in \mathcal{M}_{c,loc}\) with \(N_0 = 0\) and all \(t \geq 0\),

\[
\langle I_M(\varphi), N \rangle_t = \int_0^t \varphi_s d\langle M, N \rangle_s.
\]

Moreover, for all stopping time \(T\), we have, for all \(t \geq 0\),

\[
\int_0^t \varphi_s 1_{s \leq T} dM_s = \int_0^{\inf_t T} \varphi_s dM_s = \int_0^t \varphi_s dM_s^T.
\]

Furthermore, if \(\varphi \in \mathcal{L}_0^2(M)\) and if \(\psi\) is a progressive process then

\[
\psi \in \mathcal{L}_0^2(I_M(\varphi)) \iff \varphi\psi \in \mathcal{L}_0^2(M), \quad \text{and in this case} \quad I_M(\varphi\psi) = I_{I_M(\varphi)}(\psi).
\]

Finally if \(M \in \mathcal{M}^2\) then \(I_M\) coincides with the Itô integral of Theorem 4.12, while if \(M\) is a one-dimensional Brownian motion then \(I_M\) coincides with the Brownian Itô integral of Theorem 4.10 (with \(d = 1\)).

Note that if \(M \in \mathcal{M}_{c,loc}\) with \(M_0 = 0\), then for all \(t \geq 0\), the random variable \(M_t\) may be not integrable and in particular not square integrable. In particular, the Itô stochastic integral \(I_M(\varphi)\) with respect to a local martingale may not be centered and the Itô isometry may not hold. However, the Itô stochastic integral with respect to a square integrable continuous martingale \(M\) issued from the origin such as Brownian motion do satisfy the centering and Itô isometry for integrands in \(\mathcal{L}_0^2(M)\).

**Proof.** We proceed by localization with stopping times. For all \(n \geq 0\) we define the stopping time

\[
T_n = \inf\left\{ t \geq 0 : \int_0^t (1 + \varphi_s^2) d\langle M \rangle_s \geq n \right\}.
\]

Almost surely \(T_n \nearrow +\infty\) as \(n \to \infty\). For all \(n \geq 0\) and all \(t \geq 0\), we have

\[
\langle M^{T_n} \rangle_t = \langle M \rangle_{t \land T_n} \leq n
\]

and therefore \((M^{T_n})_{t \geq 0}\) is bounded. Since it is a martingale by the Doob stopping theorem (Theorem 2.24), we obtain that \(M^{T_n} \in \mathcal{M}^2\). Moreover we have, from the properties of the angle bracket,

\[
\int_0^{\infty} \varphi_s^2 d\langle M^{T_n} \rangle_s = \int_0^{T_n} \varphi_s^2 d\langle M \rangle_s,
\]

therefore \(\varphi \in \mathcal{L}_0^2(M^{T_n})\) and from Theorem 4.12, the stochastic process \(I_{M^{T_n}}(\varphi)\) makes sense and belongs to \(\mathcal{M}^2\). The sequence of processes \((M^{T_n})_{n \geq 0}\) is stationary since for all \(m > n\), we have \(T_n \leq T_m\) and thus

\[
I_{M^{T_n}}(\varphi) = I_{M^{T_m \land T_n}}(\varphi) = (I_{M^{T_n}}(\varphi))^{T_n}.
\]

Therefore there exists a unique process \(I_M(\varphi) = \lim_{n \to \infty} I_{M^{T_n}}(\varphi)\) such that \(I_M(\varphi)^{T_n} = I_{M^{T_n}}(\varphi)\) for all \(n \geq 0\). This process is continuous, adapted, and belongs to \(\mathcal{M}_{c,loc}\) since \((I_M(\varphi))^{T_n} = I_{M^{T_n}}(\varphi) \in \mathcal{M}^2\) for all \(n \geq 0\).
4.5 Stochastic integral with respect to semi-martingales

Now, let $N \in \mathcal{M}_{c,loc}$ such that $N_0 = 0$. For all $n \geq 0$, let us define $T^n_n = \inf\{t \geq 0 : |N_t| \geq n\}$ and $S_n = T_n \wedge T^n_n$. Almost surely $S_n \nearrow +\infty$ as $n \to \infty$. We have $N^{T^n_n} \in \mathcal{M}$ and, thanks to Theorem 4.12, for all $t \geq 0$,

$$
(I_M(\varphi), N_{T^n_n})_t = \int_0^t \varphi_s d\langle M^{T^n_n}, N^{T^n_n} \rangle_s,
$$

which gives the desired formula as $n \to +\infty$. As in Theorem 4.12, this formula characterizes $I_M(\varphi)$, mainly due to the fact that if a continuous local martingale $M$ satisfies $M_0 = 0$ and $\langle M \rangle$ is a continuous local martingale, and in particular, up to localization by stopping times, which is a sort of probabilistic cutoff, it becomes a bounded continuous martingale, in particular bounded in $L^2$.

**Remark 4.16** (Brownian motion as a martingale). Brownian motion issued from the origin is a continuous square integrable martingale, but is not bounded in $L^2$ since its $L^2$ norm at time $t$ is $\sqrt{t}$. However it is a continuous local martingale, and in particular, up to localization by stopping times, which is a sort of probabilistic cutoff, it becomes a bounded continuous martingale, in particular bounded in $L^2$.

4.5 Stochastic integral with respect to semi-martingales

The notion of quadratic variation of a process is considered in Definition 2.9. The quadratic variation of Brownian motion is considered in Theorem 3.13. In dimension one, for all $t \geq 0$, $|B_t| = (B_t)_t = t$.

**Theorem 4.17** (Quadratic variation of local martingales). Let $M = (M_t)_{t \geq 0}$ and $N = (N_t)_{t \geq 0}$ be in $\mathcal{M}_{c,loc}$ with $M_0 = N_0 = 0$. For all $t > 0$ and all partition or sub-division $\delta : 0 < t_1 < \cdots < t_n = s$ of $[0, t]$, set

$$
S(\delta) = \sum_{i=0}^{n-1} (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}).
$$

Then, denoting $|\delta| = \max_{0 \leq i \leq n-1} (t_{i+1} - t_i)$, we have, for all $t > 0$,

$$
|M, N|_t = \lim_{|\delta| \to 0} S(\delta) = \langle M, N \rangle_t.
$$

In particular $|M| = |M, M| = \langle M, M \rangle = \langle M \rangle$. In other words, for continuous local martingales, the quadratic variation matches the increasing process, as for continuous martingales.

**Proof.** We can assume without loss of generality that $M = N$ by polarization. Suppose first that for all $t \geq 0$ we have $|M_t| + \langle M \rangle_t \leq C$ almost surely where $C$ is a fixed constant. We have

$$
\mathbb{E}(S(\delta)) = \sum_i \mathbb{E}((M_{t_{i+1}} - M_{t_i})^2)
\leq \sum_i \mathbb{E}(M_{t_{i+1}}^2 - M_{t_i}^2)
\leq \sum_i \mathbb{E}((M_{t_{i+1}}) - \langle M \rangle_t)
\leq \mathbb{E}((M_t)_t)
\leq C.
$$
Moreover, denoting $\Delta_l = M_{t_{l+1}} - M_{t_l}$ and $\langle \Delta \rangle_l = \langle M \rangle_{t_{l+1}} - \langle M \rangle_{t_l}$,

$$E((S(\delta) - \langle M \rangle_t)^2) = E\left(\sum_{i,j} (\Delta_i^2 - \langle \Delta \rangle_l) (\Delta_j^2 - \langle \Delta \rangle_j)\right)$$

$$= E\sum_{i,j} (\Delta_i^2 - \langle \Delta \rangle_l) (\Delta_j^2 - \langle \Delta \rangle_j)$$

$$\leq 2\sum_n E(\Delta_i^4) + E(\langle \Delta \rangle_l^2)$$

$$\leq 2E(\langle \Delta \rangle_l^2) + 2E(\langle M \rangle_t \max \langle \Delta \rangle_l)$$

$$\leq 2 \times 4 \times C \times C^2 + 2C^2.$$  

Therefore

$$\sup_{\delta}(E(S(\delta)^2))^{1/2} \leq (8C^3 + 2C^2)^{1/2} + C < \infty.$$  

Back to the $\ast$ inequality above, we have

$$E((S(\delta) - \langle M \rangle_t)^2) \leq 2\sup_{\delta} \|S(\delta)\|_2 (E(\max \Delta_i^4))^{1/2} + 2CE \max \langle \Delta \rangle_l \longrightarrow 0$$

where we used the uniform continuity of $s \in [0, t] \rightarrow M_s$ and $s \in [0, t] \rightarrow \langle M \rangle_s$ (Heine theorem).

Let us pass to the general case by localization with stopping times. For all $n \geq 0$ we define

$$T_n = \inf\{t \geq 0 : |M_t| + \langle M \rangle_t \geq n\}.$$  

Then $(T_n)_{n \geq 0}$ is a sequence of stopping times such that $T_n \leq T_{n+1}$ for all $n \geq 0$ and $T_n \nearrow +\infty$ almost surely.

We know that $\langle M^{T_n} \rangle_t = \langle M \rangle_{T_n}$ for all $n \geq 0$ and all $t \geq 0$, hence

$$\sup_t (|M^{T_n}_t| + \langle M^{T_n} \rangle_t) \leq n.$$  

From the first part of the proof, we have

$$S^{T_n}(\delta) = \sum_T (M^{T_n}_{t_{l+1}} - M^{T_n}_{t_l}) \frac{1}{\max \langle \Delta \rangle_l} \langle M \rangle_t.$$  

Now for all $\epsilon > 0$ and all $n \geq 0$,

$$\mathbb{P}(|S(\delta) - \langle M \rangle_t| > \epsilon) \leq \mathbb{P}(T_n \leq t) + \mathbb{P}(|S(\delta) - \langle M \rangle_t| > \epsilon, t < T_n)$$

$$\leq \mathbb{P}(T_n \leq t) + \mathbb{P}(|S^{T_n}(\delta) - \langle M \rangle_t| > \epsilon).$$  

This gives easily that $\lim_{\|\delta\| \rightarrow 0} \mathbb{P}(|S(\delta) - \langle M \rangle_t| > \epsilon) = 0.$

---

**Definition 4.18** (Semi-martingales). A real $(\mathcal{F}_t)_{t \geq 0}$-adapted process $X = (X_t)_{t \geq 0}$ is a **continuous semi-martingale** when it admits a decomposition of the form

$$X = X_0 + M + V$$

where $M$ and $V$ are $(\mathcal{F}_t)_{t \geq 0}$-adapted continuous processes issued from the origin such that

- $M = (M_t)_{t \geq 0}$ is a local martingale;
- $V = (V_t)_{t \geq 0}$ has finite variation on every compact subset of $\mathbb{R}_+.$

**Remark 4.19** (Uniqueness of the decomposition). The decomposition of a semi-martingale is unique. Indeed, if $X = X_0 + M + V = X_0 + \tilde{M} + \tilde{V}$ then, with $W = \tilde{V} - V = M - \tilde{M}$, Theorem 4.17 gives, for all $t > 0$, $n \geq 0$, $\delta : 0 = t_0 < t_1 < \cdots < t_n = t$,

$$\langle M - \tilde{M} \rangle_t = \lim_{\|\delta\| \rightarrow 0} \sum_k (W_{t_{k+1}} - W_{t_k})^2,$$
4.5 Stochastic integral with respect to semi-martingales

and thus, by using the uniform continuity of $s \in [0, t] \rightarrow W_s$ (Heine theorem) we get

$$\langle M - \tilde{M} \rangle_t \leq \lim_{|\delta| \to 0} \max_k |W_{t_{k+1}} - W_{t_k}|(|V|_t + |\tilde{V}|_t) = 0.$$ 

Therefore $\langle M - \tilde{M} \rangle = 0$, which implies $M - \tilde{M} = 0$ and thus $V = \tilde{V}$.

FIXME: stochastic integral with respect to semi-martingale, nature of the integrand and nature of the integral.
Chapter 5

Itô formula and applications

Let \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) be a filtered probability space, with \((\mathcal{F}_t)_{t \geq 0}\) complete and right continuous.

5.1 Itô formula

Classical integral calculus has a fundamental formula expressing a regular function as the integral of its derivative. For the stochastic integral, the analogue is the Itô formula. The lack of regularity of the process used for integration produces an additional term of second order.

**Theorem 5.1** (Itô\(^a\) formula). If \(X = (X_t)_{t \geq 0}\) is a \(d\)-dimensional continuous process such that for all \(1 \leq i \leq d\) its \(i\)-th coordinate \((X^i_t)_{t \geq 0}\) is a semi-martingale with decomposition \(X^i_t = X^i_0 + M^i_t + V^i_t\) then for all \(f : \mathbb{R}^d \to \mathbb{R}\) of class \(C^2\) and all \(t \geq 0\),

\[
 f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dM^i_s + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dV^i_s + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d\langle M^i_t, M^j_t \rangle_s,
\]

(equality holds as random variables, in other words almost surely).

\(^a\)Named after Kiyoshi Itô (1915 – 2008), Japanese mathematician. He used the notation “Kiyosi Itô” for his name (Kunreishiki romanization), instead of the more standard “Kiyoshi Itô” (Hepburn romanization).

**Proof.** We suppose first that there exists a constant \(C > 0\) such that

\[
 \sup_{t \geq 0} \left( |X_0| + |M^i_t| + |V^i_t| + |(M^i_t, M^j_t)_s| \right) \leq C
\]

and that \(f\) has compact support. The Taylor formula gives, for all \(x, y \in \mathbb{R}^d\),

\[
 f(y) - f(x) = \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \text{Hess}(f)(x)(y - x), y - x \rangle + r(x, y)|y - x|^2
\]

\[
 = \sum_i \frac{\partial f}{\partial x_i}(a)(y_i - x_i) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(a)(y_i - x_i)(y_j - x_j) + r(x, y)|y - x|^2.
\]

Since \(f\) is \(C^2\) with compact support, by Heine theorem, \(x \mapsto f''(x) = \text{Hess}(f)(x) = (\frac{\partial^2 f}{\partial x_i \partial x_j})_{1 \leq i, j \leq d}\) is uniformly continuous, and therefore there exists a bounded continuous non-decreasing function \(g : \mathbb{R}^d \to \mathbb{R}\) such that \(\lim_{x \to 0} g(x) = 0\) and \(|r(x, y)| \leq g(|x - y|)\) for all \(x, y \in \mathbb{R}\).

Now, for all \(n \geq 0\) and \(t > 0\) and all sub-division \(\delta : 0 = t_0 < t_1 < \cdots < t_{n+1} = t\) of \([0, t]\), denoting \(\Delta X_k = X_{t_{k+1}} - X_{t_k}\) for all \(0 \leq k \leq n\), we have

\[
 f(X_t) - f(X_0) = \sum_k (f(X_{t_{k+1}}) - f(X_{t_k}))
\]

\[
 = \sum_k \langle \nabla f(X_{t_k}), \Delta X_k \rangle + \frac{1}{2} \sum_k \langle \text{Hess}(f)(X_{t_k})\Delta X_k, \Delta X_k \rangle + \sum_k r(X_{t_k}, X_{t_{k+1}})|\Delta X_k|^2
\]

\[
 = S_1 + S_2 + S_3.
\]
Denoting $\Delta V_k = V_{t_{k+1}} - V_{t_k}$, $\Delta M_k = M_{t_{k+1}} - M_{t_k}$, and $|\delta| = \max_{0 \leq k \leq n} (t_{k+1} - t_k)$, we have

$$S_1 = \sum_k (\nabla f(X_{t_k}) \Delta V_k) + \sum_k (\nabla f(X_{t_k}) \Delta M_k)$$

$$\quad = \frac{L^2}{|\delta|} \int_0^t \nabla f(X_s) dV_s + \int_0^t \nabla f(X_s) dM_s$$

$$\quad = \sum_{i=1}^{d} \int_0^t \frac{\partial f}{\partial x_i}(X_s)(dV_s^i + dM_s^i).$$

Next, denoting $\Delta X_k^i = X_{t_{k+1}}^i - X_{t_k}^i$, $\Delta M_k^i = M_{t_{k+1}}^i - M_{t_k}^i$, $\Delta V_k^i = V_{t_{k+1}}^i - V_{t_k}^i$,

$$S_2 = \frac{1}{2} \sum_k (\text{Hess}(f)(X_{t_k}) \Delta X_k, \Delta X_k)$$

$$\quad = \frac{1}{2} \sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) \Delta X_k^i \Delta X_k^j$$

$$\quad = \frac{1}{2} \sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) \Delta M_k^i \Delta M_k^j$$

$$\quad + \sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) \Delta M_k^i \Delta V_k^j$$

$$\quad = \frac{1}{2} \sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) \Delta V_k^i \Delta V_k^j$$

$$\quad = S_2'' + S_2'''.$$

Now we have

$$|S_2''| \leq \sum_{i,j} \max_k |\Delta M_k^i| \sum_k \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) \right| |\Delta V_k^j| \int_{|\delta|}^0 \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \right| |dV_s^j| = 0,$$

and similarly

$$|S_2'''| \leq \sum_{i,j} \max_k |\Delta V_k^i| \sum_k \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) \right| |\Delta M_k^j| \int_{|\delta|}^0 \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \right| |dM_s^j| = 0.$$

For $S_2$, denoting $\langle \Delta M^{i,j} \rangle_k = \langle M^i, M^j \rangle_{t_{k+1}} - \langle M^i, M^j \rangle_{t_k}$, we have, using Theorem 4.17,

$$\mathbb{E}\left( \left( \sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) (\Delta M_k^i \Delta M_k^j - \langle \Delta M^{i,j} \rangle_k) \right)^2 \right)$$

$$\quad = \sum_k \mathbb{E}\left( \left( \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) (\Delta M_k^i \Delta M_k^j - \langle \Delta M^{i,j} \rangle_k) \right)^2 \right)$$

$$\quad \leq C \mathbb{E}\left( \sum_k \left( \Delta M_k^i \Delta M_k^j - \langle \Delta M^{i,j} \rangle_k \right)^2 \right) \int_{|\delta|}^0 0.$$

It follows that in $L^2$,

$$\lim_{|\delta| \to 0} S_2' = \lim_{|\delta| \to 0} \frac{1}{2} \sum_{i,j=1}^{d} \sum_{k=0}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k}) (\langle M^i, M^j \rangle_{t_{k+1}} - \langle M^i, M^j \rangle_{t_k})$$

$$\quad = \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d\langle M^i, M^j \rangle_s.$$
\[
\sum_{k=0}^{n} |r(X_{t_k}, X_{t_{k+1}})| |X_{t_{k+1}} - X_{t_k}|^2 
\leq 2 g \left( \max_{0 \leq k \leq n} |X_{t_{k+1}} - X_{t_k}| \right) \sum_{i=1}^{d} \left( \sum_{k=0}^{n} (M^i_{t_{k+1}} - M^i_{t_k})^2 + \sum_{k=0}^{n} (V^i_{t_{k+1}} - V^i_{t_k})^2 \right).
\]

This achieves the proof under the assumptions of boundedness of \(M\) and \(V\) and compactness of the support of \(f\). To prove the general case, we consider the sequence \((T_n)_{n \geq 0}\) of stopping times defined for all \(n \geq 0\) by

\[
T_n = \inf \left\{ t \geq 0 : \sum_{i,j=1}^{d} |X_0| + |M^i_t| + |V^i_t| + |\langle M^i, M^j \rangle_t| \geq n \right\}
\]

Then \(T_n \nearrow +\infty\) almost surely and from the first part of the proof

\[
f(X_{t \wedge T_n}) = f(X_0) + \sum_{i=1}^{d} \int_0^{t \wedge T_n} \frac{\partial f}{\partial x_i}(X_s) (dM^i_s + dV^i_s) + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^{t \wedge T_n} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d\langle M^i, M^j \rangle_s.
\]

It suffices then to let \(n \to \infty\).

An immediate consequence of the Itô formula of Theorem 5.1 is that for all continuous local martingales \(M, N \in \mathcal{M}_{c, loc}\) and for all continuous process \(V\) issued from the origin and of bounded variation on every compact subset of \(\mathbb{R}_+\), we have, for all \(t \geq 0\),

\[
M_t N_t = M_0 N_0 + \int_0^t M_s dN_s + \int_0^t N_s dM_s + \langle M, N \rangle_t
\]

and

\[
M_t V_t = M_0 V_0 + \int_0^t M_s dV_s + \int_0^t V_s dM_s,
\]

obtained with \(f(x_1, x_2) = x_1 x_2\) and \(X = (M, N)\) for the first and \(X = (M, V)\) for the second.

### 5.2 Lévy characterization of Brownian motion and Dubins–Schwarz theorem

**Theorem 5.2** (Lévy\(^a\) characterization of Brownian motion). If \(M = (M_t)_{t \geq 0}\) is a \(d\)-dimensional \((\mathcal{F}_t)_{t \geq 0}\)-adapted process such that for \(1 \leq i \leq d\) its \(i\)-th coordinate \((M^i_t)_{t \geq 0}\) is a continuous local martingale issued from the origin and if \(\langle M^i, M^j \rangle_t = 1_{i = j t}\) for all \(1 \leq i, j \leq d\) and all \(t \geq 0\), then \(M\) is an \((\mathcal{F}_t)_{t \geq 0}\) Brownian motion.

\(^a\)See footnote a page 30.

**Proof.** Let \(\lambda \in \mathbb{R}^d\) and \(N^\lambda_t = e^{i\lambda \cdot M_t + \frac{1}{2} |\lambda|^2 t}\) and let us reserve the notation \(i\) for the complex number \((0, 1)\) in this proof. Thanks to Theorem 3.8, it suffices to show that \((N^\lambda_t)_{t \geq 0}\) is an \((\mathcal{F}_t)_{t \geq 0}\)-martingale. From the Itô formula of Theorem 5.1 with \(f(x) = e^{i\lambda \cdot x + \frac{1}{2} |x|^2 t}\), we have

\[
N^\lambda_t = 1 + \int_0^t N^\lambda_s \left( i d(\lambda \cdot M_s) + \frac{1}{2} |\lambda|^2 ds \right) + \frac{1}{2} \sum_{j,k=1}^{d} \lambda_j \lambda_k \int_0^t N^\lambda_s 1_{j=k} ds
\]

\[
= 1 + i \int_0^t N^\lambda_s d(\lambda \cdot M_s) + \frac{1}{2} |\lambda|^2 \int_0^t N^\lambda_s ds - \frac{1}{2} |\lambda|^2 \int_0^t N^\lambda_s ds
\]

\[
= 1 + i \int_0^t N^\lambda_s d(\lambda \cdot M_s).
\]

The stochastic integral \(\int_0^t N^\lambda_s d(\lambda \cdot M_s)\) is an \((\mathcal{F}_t)_{t \geq 0}\)-martingale since for all \(t \geq 0\),

\[
\mathbb{E} \int_0^t |N^\lambda_s|^2 d(\lambda \cdot M, \lambda \cdot M)_s = \int_0^t e^{\frac{1}{2} |\lambda|^2 s} |\lambda|^2 ds < \infty.
\]

We used the criterion for being a martingale given by Remark 5.3.

\[67/98\]
Remark 5.3 (Martingale criterion). For all $X \in \mathcal{M}_{\text{loc}}$, a necessary (and sufficient!) condition for $X$ to be a square integrable martingale is that $\mathbb{E}(\langle X \rangle_1) < \infty$ for all $t \geq 0$. Proof: exercise!

Corollary 5.4 (Dubins–Schwarz theorem). If $M = (M_t)_{t \geq 0}$ is a continuous local martingale issued from the origin, with $\langle M \rangle_\infty = \infty$, and if we define, for all $t \geq 0$, the stopping time

$$T_t = \inf\{s \geq 0 : \langle M \rangle_s > t\},$$

then $M_{T_t} = M_{\langle M \rangle}$ for all $t \geq 0$ and the process $(M_{\langle M \rangle})_{t \geq 0}$ is an $(\mathcal{F}_{T_t})_{t \geq 0}$-Brownian motion.

Proof. FIXME:

Theorem 5.5 (Martingale criterion). Let $M = (M_t)_{t \geq 0}$ and $V = (V_t)_{t \geq 0}$ be continuous adapted real processes issued from the origin, with $V$ increasing. For all $\lambda \in \mathbb{R}$ let us define the process

$$X^\lambda = (X^\lambda_t)_{t \geq 0} = (e^{\lambda M_t - {\lambda^2 \over 2} V_t})_{t \geq 0}.$$

Then the following properties are equivalent.

1. $M$ is a local martingale and $\langle M \rangle = V$;
2. $X^\lambda$ is a local martingale for all $\lambda \in \mathbb{R}$.

In this case $X^\lambda$ is a super-martingale, and is a martingale if and only if $\mathbb{E}X^\lambda_t = 1$ for all $t \geq 0$.

Moreover if $\mathbb{E}\int_0^t e^{2\lambda M_s} \, dV_s < \infty$ for all $t \geq 0$ then $X^\lambda$ is a martingale. Furthermore if $X^\lambda$ is a martingale for all $\lambda \in \mathbb{R}$ and if $\mathbb{E}e^{\lambda M_t} < \infty$ for all $t \geq 0$ then $M$ is a martingale.

Proof. Suppose that $M$ is a local martingale and that we have $\langle M \rangle = V$. The Itô formula gives

$$X^\lambda_t = 1 + \lambda \cdot \int_0^t X^\lambda_s \, dM_s$$

for all $\lambda \in \mathbb{R}$ and all $t \geq 0$, and therefore $X^\lambda$ is a non-negative local martingale. On the other hand, the Fatou lemma gives that $X^\lambda$ is a non-negative local super-martingale: indeed since $X^\lambda$ is a local martingale, there exists a sequence of stopping times $(T_n)_{n \geq 0}$ such that $T_n \not\to +\infty$ almost surely and such that $(X^\lambda_{T_n})_{t \geq 0}$ is a martingale for all $n \geq 0$ then for all $0 \leq s \leq t$,

$$X^\lambda_s = \lim_{n \to \infty} X_{s \wedge T_n}^\lambda = \lim_{n \to \infty} \mathbb{E}(X_{s \wedge T_n}^\lambda | \mathcal{F}_s) \geq \mathbb{E}(\lim_{n \to \infty} X_{s \wedge T_n}^\lambda | \mathcal{F}_s) = \mathbb{E}(X_s^\lambda | \mathcal{F}_s).$$

Also $\mathbb{E}X^\lambda_s \leq \mathbb{E}X^\lambda_0 = 1$ and $X^\lambda$ is martingale if and only if $\mathbb{E}X^\lambda_t = 1$ for all $t \geq 0$.

Moreover, for all $t \geq 0$,

$$\mathbb{E}(X^\lambda)_t = \lambda^2 \mathbb{E}\int_0^t e^{2\lambda M_s - \lambda^2 V_s} \, dV_s \leq \lambda^2 \mathbb{E}\int_0^t e^{2\lambda M_s} \, dV_s.$$

If this last term is finite, then thanks to Remark 5.3, the process $X^\lambda$ is a martingale.

Conversely, suppose first that $X^\lambda$ is a martingale and that $\mathbb{E}e^{\lambda M_t} < \infty$ for all $\lambda \in \mathbb{R}$. Then for all $0 \leq s < t$ and all $A \in \mathcal{F}_s$,

$$\mathbb{E}(1_A e^{\lambda M_t - \lambda^2 V_t}) = \mathbb{E}(1_A X^\lambda_t) = \mathbb{E}(1_A X^\lambda_t) = \mathbb{E}(1_A e^{\lambda M_s - \lambda^2 V_s}).$$

Taking the derivative with respect to $\lambda$, which is allowed here by dominated convergence, gives

$$\mathbb{E}(1_A e^{\lambda M_t - \lambda^2 V_t} (M_t - \lambda V_t)) = \mathbb{E}(1_A e^{\lambda M_s - \lambda^2 V_s} (M_s - \lambda V_s)).$$
and additionally by taking the derivative with respect to $\lambda$ again,

$$
\mathbb{E}(1_A X^\lambda_t ((M_t - AV_j)^2 - V_j)) = \mathbb{E}(1_A X^\lambda_t ((M_2 - \lambda V_j)^2 - V_j)).
$$

Taking $\lambda = 0$ gives that $M$ and $(M_t^2 - V_t)_{t \geq 0}$ are martingales, and in particular $(M_t)_{t \geq 0}$ is a martingale under $Q$. Finally, to address the general case, we consider the sequence $(T_n)_{n \geq 0}$ of stopping times defined by $T_n = \inf \{t \geq 0 : |M_t| \geq n \}$ for all $n \geq 0$, then $(X^\lambda_{t \wedge T_n})_{t \geq 0}$ is a bounded local martingale, and thus a martingale, and we are in the conditions above.

### 5.3 Girsanov theorem for Itô integrals

The following result is a generalization to random translations of the result due to Cameron–Martin about the density of translated processes (Theorem 3.24).

**Theorem 5.6** (Girsanov). Let $B = (B_t)_{t \geq 0}$ be an $(\mathcal{F}_t)_{t \geq 0}$ $d$-dimensional Brownian motion issued from the origin and $\varphi = (\varphi_t)_{t \geq 0}$ be a $d$-dimensional $(\mathcal{F}_t)_{t \geq 0}$-adapted and measurable process, bounded on each compact subset of $\mathbb{R}_+$. Then the process $M = (M_t)_{t \geq 0}$ defined for all $t \geq 0$ by

$$
M_t = \exp \left( \int_0^t \varphi_s dB_s - \frac{1}{2} \int_0^t |\varphi_s|^2 ds \right)
$$

is a martingale and $\mathbb{E}M_t = 1$ for all $t \geq 0$. Moreover for all fixed $T > 0$ the law of the process

$$
\tilde{B} = \left( B_t - \int_0^t \varphi_s dB_s \right)_{0 \leq s \leq T}
$$

is absolutely continuous with respect to the law of $(B_t)_{0 \leq s \leq T}$ with density $M_T$. In other words $\tilde{B}$ is an $(\mathcal{F}_t)_{0 \leq t \leq T}$ Brownian motion on $(\Omega, \mathcal{F}, Q)$ where $Q$ is defined by $dQ = M_T dP$.

We recover Theorem 3.24 of Cameron–Martin when $\varphi$ is deterministic.

**Proof.** From Theorem 5.5, the process $M$ is a non-negative local martingale and thus a non-negative supermartingale. For all $t \geq 0$, let us consider

$$
N_t = \int_0^t \varphi_s dB_s \quad \text{and} \quad \langle N \rangle_t = \int_0^t |\varphi_s|^2 ds.
$$

For all $0 \leq t < T$ we have, denoting $C = \sup_{s \in [0, T]} |\varphi_s|^2$,

$$
\mathbb{E}(\langle M \rangle_t) = \mathbb{E} \left( \int_0^t e^{2N_s - \langle N \rangle_s} |\varphi_s|^2 ds \right)
\leq C \int_0^t \mathbb{E}(e^{2N_s - 2\langle N \rangle_s}) ds
\leq C e^{Ct} \int_0^t e^{2\langle N \rangle_s} ds
\leq C e^{Ct} t < \infty.
$$

Therefore, $M$ is a martingale thanks to the criterion given by Remark 5.3.

In order to check that $\tilde{B}$ is a Brownian motion under $Q$, we use Theorem 3.8 which reduces the problem to show that for all $\lambda \in \mathbb{R}^d$ and all fixed $T \geq 0$, the process

$$
\left( e^{\lambda \cdot \tilde{B}_t - \frac{1}{2} \lambda^2 t} \right)_{0 \leq t \leq T}
$$

is a martingale under $Q$. Indeed, for all $0 \leq s < t \leq T$ and $A \in \mathcal{F}_s$,

$$
\mathbb{E}_Q \left( 1_A e^{\lambda \cdot \tilde{B}_s - \frac{1}{2} \lambda^2 s} M_t \right) = \mathbb{E} \left( 1_A e^{\lambda \cdot B_s - \frac{1}{2} \lambda^2 s} e^{\lambda \cdot (B_t - \tilde{B}_t) - \frac{1}{2} \lambda^2 (t-s)} M_t \right)
$$

69/98
\[ E \left( 1_A e^{\int_0^t (\lambda + \varphi_s) \, dB_s - \frac{1}{2} \int_0^t (\lambda + \varphi_s)^2 \, ds} \right) \]

\[ = E \left( 1_A e^{\int_0^t (\lambda + \varphi_u) \, dB_u - \frac{1}{2} \int_0^t (\lambda + \varphi_u)^2 \, du} \right) \]

\[ = E_Q \left( 1_A e^{\tilde{B}_s - \frac{\lambda s}{2}} \right) \]

where we used in \( \star \) the fact that \( M \) is a martingale with \( \lambda + \varphi \) instead of \( \varphi \).

\section*{5.4 Sub-Gaussian deviation inequality}

**Theorem 5.7** (Sub-Gaussian deviation inequality). For all continuous local martingale \( M = (M_t)_{t \geq 0} \) issued from the origin and for all \( t > 0, K \geq 0, r \geq 0 \), we have

\[ P\left( \langle M \rangle_t \leq K, \sup_{0 \leq s \leq t} |M_s| \geq r \right) \leq 2e^{-\frac{r^2}{2K}}. \]

In particular, if \( \langle M \rangle_s \leq K s \) for all \( 0 \leq s \leq t \) then

\[ P\left( \sup_{0 \leq s \leq t} |M_s| \geq r \right) \leq 2e^{-\frac{r^2}{2Kt}}. \]

The condition on \( \langle M \rangle_s \) is a comparison to Brownian motion, for which \( s \mapsto \langle B \rangle_s \) is linear.

**Proof.** For all \( \lambda, t \geq 0 \), by Theorem 5.5, the process

\[ X^\lambda = \left( e^{\lambda M_t - \frac{\lambda^2}{2} \langle M \rangle_t} \right)_{t \geq 0}, \]

is a non-negative super-martingale and \( E X^\lambda_t \leq 1 \) for all \( t, \lambda \geq 0 \). Therefore, for all \( t, \lambda, r, K \geq 0 \),

\[ P\left( \langle M \rangle_t \leq K, \sup_{0 \leq s \leq t} M_s \geq r \right) \leq P\left( \langle M \rangle_t \leq K, \sup_{0 \leq s \leq t} X^\lambda_s \geq e^{\frac{r - \frac{\lambda^2}{2} K}{t}} \right) \]

\[ \leq P\left( \sup_{0 \leq s \leq t} X^\lambda_s \geq e^{\frac{r - \frac{\lambda^2}{2} K}{t}} \right) \]

\[ \leq e^{-\frac{r}{t} + \frac{\lambda^2}{2} K} \]

where the last step comes from the maximal inequality (Theorem 2.14). Taking \( \lambda = r/K \) gives

\[ P\left( \langle M \rangle_t \leq K, \sup_{0 \leq s \leq t} M_s \geq r \right) \leq e^{-\frac{r^2}{2K}}. \]

The same reasoning provides

\[ P\left( \langle M \rangle_t \leq K, \sup_{0 \leq s \leq t} (-M_s) \geq r \right) \leq e^{-\frac{r^2}{2K}}. \]

The desired result follows now by the union bound, hence the factor 2 in the right hand side.

**Exercise 5.8** (Sub-Gaussian exponential integrability). Let \( M \) be as in Theorem 5.7 and for all \( t \geq 0 \). Show that if \( \langle M \rangle_t \leq K t \) for all \( t \geq 0 \) then for all \( \alpha < 1/(2Kt) \),

\[ E e^{\alpha \|M\|_t^2} < \infty \quad \text{where} \quad \|M\|_t = \sup_{0 \leq s \leq t} |M_s|. \]

\section*{5.5 Burkholder–Davis–Gundy inequalities}
5.5 Burkholder–Davis–Gundy inequalities

**Theorem 5.9 (Burkholder–David–Gundy inequalities).** For all \( p \in (0, +\infty) \) there exists universal constants \( c_p > 0 \) and \( C_p > 0 \) such that for all continuous local martingale \( M = (M_t)_{t \geq 0} \) issued from the origin, we have, for all fixed \( T \geq 0 \), denoting \( \|M_T\| = \sup_{0 \leq t \leq T} |M_t| \),

\[
c_p \mathbb{E}(\|M_T\|^{2p}) \leq \mathbb{E}(\langle M \rangle_T^{p}) \leq C_p \mathbb{E}(\|M_T\|^{2p}).
\]

**Proof.** Let \((T_n)_{n \geq 0}\) be the sequence of stopping times defined for all \( n \geq 0 \) by

\[
T_n = \inf\{t \geq 0 : |M_t| \geq n \text{ or } \langle M \rangle_t \geq n\}.
\]

Then if the desired inequalities are satisfied for \( M^{T_n} = (M_{T_nT})_{t \geq 0} \) with constants \( c_p > 0 \) and \( C_p > 0 \), then they will also be satisfied by \( M \) by letting \( n \to \infty \). This localization method allows in fact to assume without loss of generality that \( M \) is a bounded continuous martingale.

Let us fix \( T > 0 \). The maximal inequality of Theorem 2.14 writes, for all \( r \in (1, +\infty) \),

\[
\mathbb{E}(\|M\|_T^r) \leq \left( \frac{r}{r-1} \right)^{r} \mathbb{E}(\langle M \rangle_T)
\]

**Case \( p = 1 \).** In this case \( \mathbb{E}(\langle M \rangle_T) = \mathbb{E}(M_T^2) \) and the desired BGD inequality is verified with \( c_1 = 1/4 \) (maximal inequality with \( r = 2 \)) and \( C_1 = 1 \) (monotony of expectation).

**Case \( p > 1 \).** We have, from the Itô formula of Theorem 5.1, for all \( t \geq 0 \),

\[
|M_t|^{2p} = 2p \int_0^t |M_s|^{2p-1} \text{sign}(M_s) \, dM_s + p(2p-1) \int_0^t |M_s|^{2p-2} \, d\langle M \rangle_s
\]

and thus, for all \( 0 \leq t \leq T \), using the Hölder inequality with \( p \) and \( q = 1/(1-1/p) = p/(p-1) \),

\[
\mathbb{E}(|M_t|^{2p}) = p(2p-1)\mathbb{E} \left[ \int_0^t |M_s|^{2p-2} \, d\langle M \rangle_s \right]
\leq p(2p-1)\mathbb{E}(\|M^2\|^{p-1}_T \langle M \rangle_T)
\leq p(2p-1)\mathbb{E}(\|M^2\|^{p-1}_T)^{1/p} \mathbb{E}(\langle M \rangle_T^{p})^1/p.
\]

Combined with the maximal inequality above used with \( r = 2p \), we obtain the second BGD inequality. To prove the first BGD inequality, we write, using the Itô formula of Theorem 5.1,

\[
M_t \langle M \rangle_t^{(p-1)/2} = \int_0^t \langle M \rangle_s^{(p-1)/2} \, dM_s + \int_0^t M_s \, d\langle M \rangle_s^{(p-1)/2}.
\]

If we set \( N_t = \int_0^t \langle M \rangle_s^{(p-1)/2} \, dM_s \), we have, for all \( t \in [0, T] \),

\[
|N_t| \leq 2\|M\|_T \langle M \rangle_T^{(p-1)/2},
\]

which gives, using the Hölder inequality with \( p \) and \( q = 1/(1-1/p) = p/(p-1) \),

\[
\mathbb{E}(N_t^2) \leq 4\mathbb{E}(\|M\|^{2p}_T \langle M \rangle_T^{p-1}) \leq 4(\mathbb{E}(\|M\|^{2p}_T))^1/p \mathbb{E}(\langle M \rangle_T)^{1-1/p}.
\]

Combined with

\[
\mathbb{E}(N_t^2) = \mathbb{E} \left[ \int_0^t \langle M \rangle_s^{p-1} \, d\langle M \rangle_s \right] = \frac{1}{p} \mathbb{E}(\langle M \rangle_t^{p})
\]

we obtain

\[
\mathbb{E}(\langle M \rangle_T^{p}) \leq (4p)^p \mathbb{E}(\|M\|^{2p}_T),
\]

which is the first BGD inequality.

**Case \( 0 < p < 1 \).** Let us define \( N_t = \int_0^t \langle M \rangle_s^{(p-1)/2} \, dM_s \). We have

\[
M_t = \int_0^t \langle M \rangle_s^{(1-p)/2} \, dN_s
\]

71/98
and
\[
N_t(M)^{(1-p)/2}_t = \int_0^t (M)^{(1-p)/2}_s dN_s + \int_0^t N_s d((M)^{(1-p)/2})_s
\]
\[= M_t + \int_0^t N_s d((M)^{(1-p)/2})_s.
\]
Therefore, for all \(t \in [0, T]\),
\[
|M_t| \leq 2\|N\|_T(M)^{(1-p)/2}_T \quad \text{and} \quad \|M\|_T \leq 2\|N\|_T(M)^{(1-p)/2}_T,
\]
thus, using the Hölder inequality with \(1/p\) and its conjugate exponent \(1/(1-p)\),
\[
\mathbb{E}(\|M\|^{2p}_T) \leq 4^p \mathbb{E}(\|N\|^{2p}_T(M)^{p(1-p)}_T)
\]
\[\leq (4^p)^2 (\mathbb{E}(\|N_T\|^2))^{p-1} (\mathbb{E}(\|M_T\|^{p})^{1-p}
\]
\[\leq (4^p)^2 (\mathbb{E}(\|N_T\|^2))^{p-1} (\mathbb{E}(\|M_T\|^{p})^{1-p}
\]
\[= 16^p (p-1)^2 \mathbb{E}(\|M_T\|^{p})^{1-p}
\]
\[= \left(\frac{16}{p}\right)^p \mathbb{E}(\|M_T\|^{p}).
\]
This proves the first BGD inequality. To prove the second BGD inequality, let \(\alpha > 0\). The reason for \(\alpha > 0\) is to avoid the singularity at 0 of \(x \rightarrow x^{p-1}\) due to the fact that \(p - 1 < 0\). Now write, using the Itô formula of Theorem 5.1,
\[
M_t(\alpha + \|M\|_t)^{p-1} = \int_0^t (\alpha + \|M\|_s)^{p-1} dM_s + \int_0^t M_s d(\alpha + \|M\|_s)^{p-1}
\]
\[= N_t + (p-1) \int_0^t M_s(\alpha + \|M\|_s)^{p-2} d\|M\|_s
\]
where \(N_t = \int_0^t (\alpha + \|M\|_s)^{p-1} dM_s\). We have then (taking \(\alpha \rightarrow 0\))
\[
|N_t| \leq \|M_t\|^p + (1-p) \int_0^t \|M\|^{p-1} d\|M\|_s = \frac{1}{p} \|M\|^p_t
\]
and thus
\[
\mathbb{E} \int_0^t (\alpha + \|M\|_s)^{2(p-1)} d(M)^2 = \mathbb{E}(N_t^2) \leq \frac{1}{p^2} \mathbb{E}(\|M\|^{2p}_t),
\]
which gives finally the inequality (recall that \(2(1-p) < 0\))
\[
\mathbb{E}(\|M\|^{2p}_t) \leq \frac{1}{p^2} \mathbb{E}(\|M\|^{2p}_t).
\]
But the identity
\[
\langle M \rangle^p_t = (\langle M \rangle^p_t(\alpha + \|M\|_t)^{2(p-1)})(\alpha + \|M\|_t)^{2p(1-p)}
\]
gives, using the Hölder inequality with \(1/p\) and its conjugate exponent \(1/(1-p)\), that
\[
\mathbb{E}(\langle M \rangle^p_t) \leq (\mathbb{E}(\langle M \rangle_t(\alpha + \|M\|_t)^{2(p-1)}))^{p-1} (\mathbb{E}(\|M\|^{2p}_t))^{1-p}
\]
\[\leq \left(\frac{1}{p^2}\right)^p \mathbb{E}(\|M\|^{2p}_t) (\mathbb{E}(\alpha + \|M\|_t)^{2p})^{1-p}.
\]
Taking the limit as \(\alpha \rightarrow 0\), we obtain
\[
\mathbb{E}(\langle M \rangle^p_t) \leq \frac{1}{p^2} \mathbb{E}(\|M\|^{2p}_t)
\]
which is the second BGD inequality.
Chapter 6

Stochastic differential equations

Let \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)\) be a filtered probability space, with \((\mathcal{F}_t)_{t \geq 0}\) complete and right continuous. Let \(B = (B_t)_{t \geq 0}\) be a \(d\)-dimensional \((\mathcal{F}_t)_{t \geq 0}\) Brownian motion issued from the origin.

6.1 Stochastic differential equations with Lipschitz coefficients

Let us denote \(|M| = (\sum_{j,k} M_{jk}^2)^{1/2}\) for all matrix \(M \in \mathcal{M}_{q,d}(\mathbb{R})\) with \(q\) rows and \(d\) columns.

Let us consider two maps \(\sigma : \mathbb{R}_+ \times \Omega \times \mathbb{R}^q \rightarrow \mathcal{M}_{q,d}(\mathbb{R})\) and \(b : \mathbb{R}_+ \times \Omega \times \mathbb{R}^q \rightarrow \mathbb{R}^q\) such that the following properties hold true:

1. the maps \(\sigma, b\) are Lipschitz in the space variable, namely there exists a constant \(c > 0\) such that for all \((u, \omega) \in \mathbb{R}_+ \times \Omega\) and all \(x, y \in \mathbb{R}^q\),
   \[|\sigma(u, \omega, x) - \sigma(u, \omega, y)| \leq c|x - y|\] and \[|b(u, \omega, x) - b(u, \omega, y)| \leq c|x - y|;\]

2. the maps \(\sigma, b\) are measurable for the time/random variables, namely for all \(t > 0, x \in \mathbb{R}^q\),
   \((u, \omega) \in [0, t] \times \Omega \mapsto \sigma(u, \omega, x)\) and \((u, \omega) \in [0, t] \times \Omega \mapsto b(u, \omega, x)\)
   are \((\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t)\)-measurable;

3. the maps \(\sigma, b\) are locally square integrable, namely for all \(t > 0\) and \(x \in \mathbb{R}^q\),
   \[\mathbb{E} \int_0^t |\sigma|^2(u, \cdot, x)du < \infty\] and \[\mathbb{E} \int_0^t |b|^2(u, \cdot, x)du < \infty.\]

Note that thanks to the first property (Lipschitz regularity), if this square integrability property is satisfied for some \(x \in \mathbb{R}^q\) then it will be satisfied for all \(x \in \mathbb{R}^q\).

**Theorem 6.1** (Solving stochastic differential equations). For all \(s \geq 0\) and all \(\mathcal{F}_s\)-measurable and square integrable random vector \(\eta\) of \(\mathbb{R}^q\), there exists an adapted and continuous \(q\)-dimensional process \(X = (X_t)_{t \geq s}\) such that the following properties hold true:

1. for all \(t \geq s\), \(\mathbb{E} \int_s^t |X_u|^2 du < \infty;\)

2. \(X\) solves the stochastic differential equation (SDE)
   \[X_t = \eta + \int_s^t \sigma(u, X_u)dB_u + \int_s^t b(u, X_u)du \quad \text{a.s.,} \quad t \geq s,\]
   in other words for all \(1 \leq j \leq q,\)
   \[X^j_t = \eta_j + \sum_{k=1}^d \int_s^t \sigma_{jk}(u, X_u)dB^k_u + \int_s^t b_j(u, X_u)du \quad \text{a.s.,} \quad t \geq s.\]
Moreover this process $X$ is unique up to indistinguishability.

Note that the first property ensures that the second property has a meaning since
\[
|\sigma(u,X_u)|^2 \leq 2(|\sigma(u,0)|^2 + c^2|X_u|^2).
\]

The process $X$ solves the stochastic differential equation
\[
X_s = \eta, \quad dX_t = \sigma(u,X_t)dB_t + b(t,X_t)dt \quad \text{a.s.,} \quad t \geq s.
\]

Proof. Let $\mathcal{D}$ be the set of continuous adapted $q$-dimensional processes $(Y_t)_{t \geq s}$ with, for all $t \geq 0$,
\[
\|Y\|_t^2 = \mathbb{E}(\sup_{s \leq u \leq t} |Y_u|^2) < \infty.
\]

For all $Y \in \mathcal{D}$, we define, for all $t \geq s$,
\[
SY(t) = \eta + \int_s^t \sigma(u,Y_u)dB_u + \int_s^t b(u,Y_u)du.
\]

For all $Y^1$ and $Y^2$ in $\mathcal{D}$ we have, for all $t \geq s$,
\[
SY^1(t) - SY^2(t) = \int_s^t (\sigma(u,Y_u^1) - \sigma(u,Y_u^2))dB_u + \int_s^t (b(u,Y_u^1) - b(u,Y_u^2))du,
\]

and then
\[
\sup_{s \leq u \leq t} |SY^1(u) - SY^2(u)|^2 \leq 2 \sup_{s \leq u \leq t} \left( |\sigma(v,Y_u^1) - \sigma(v,Y_u^2)|^2 |dB_v|^2 \right) + 2(t-s) \sup_{s \leq u \leq t} |b(u,Y_u^1) - b(u,Y_u^2)|^2 |du|.
\]

By using the Doob maximal inequality of Theorem 2.14 with $p = 2$, we get
\[
\|SY^1 - SY^2\|_t^2 \leq 8 \int_s^t \mathbb{E}(|\sigma(u,Y_u^1) - \sigma(u,Y_u^2)|^2) |du| + 2(t-s) \int_s^t \mathbb{E}(|b(u,Y_u^1) - b(u,Y_u^2)|^2) |du|.
\]

Taking $Y^2 \equiv 0$, this shows that $SY \in \mathcal{D}$ when $Y \in \mathcal{D}$. If we set $C_t = 2c^2(4 + (t-s))$ and $\varphi(u) = \mathbb{E}(|Y_u^1 - Y_u^2|^2)$, we get, for all $n \geq 1$, denoting $S^n = S \circ \cdots \circ S$ the $n$-th iteration of $S$,
\[
\|S^n Y^1 - S^n Y^2\|_t^2 \leq (C_t)^2 \int_s^t \mathbb{E}(|S^n Y^1 - S^n Y^2(\nu)|^2) d\nu\]
\[
\vdots
\]
\[
\leq (C_t)^n \int_{t 
\leq (C_t)^n \|Y^1 - Y^2\|_t^2 \frac{(t-s)^n}{n!}.
\]

Let us show now that $S$ admits a fixed point. We start from an arbitrary $Y \in \mathcal{D}$, and we set $X^0 = Y$, and $X^n = S^n Y$ for all $n \geq 1$. Then we have
\[
\mathbb{E}\left(\sup_{s \leq u \leq t} |X_u^n - X_u^{n+1}|^2\right) \leq \frac{(C_t(t-s))}{n!} \|Y - SY\|_t^2.
\]

It follows that
\[
\mathbb{E} \sum_{n \geq 0} \sup_{s \leq u \leq t} |X_u^n - X_u^{n+1}| \leq \sum_{n \geq 0} \frac{(C_t(t-s))^n}{n!} \|Y - SY\|_t^{1/2} < \infty.
\]
Thus, for all $t > s$, almost surely

$$
\sum_{n \geq 0} \sup_{s \leq u \leq t} |X_u^n - X_u^{n+1}| < \infty.
$$

Therefore, the sequence of continuous processes $(X_u^n)_{n \geq 0}$ converges almost uniformly on every compact subset of $[s, \infty)$ towards a continuous adapted process denoted $X = (X_u)_{u \geq 0}$ and from the inequality $(\star \star)$ we get

$$
\left( \mathbb{E} \left( \sup_{s \leq u \leq t} |X_u^n - X_u| \right)^{2} \right)^{1/2} \leq \sum_{n \geq 0} \|X_u^n - X_u^{n+1}\|_t \xrightarrow{n \to \infty} 0.
$$

It follows that $X \in \mathcal{D}$, that $X_u^n \to X$ in $\mathcal{D}$, and that

$$
\|X - SX\|_t \leq \|X - X_u^{n+1}\|_t + \|SX_u^n - SX\|_t \xrightarrow{n \to \infty} 0.
$$

It follows that $X = SX$. Now let $X$ and $\tilde{X}$ two fixed points of $S$. We have, for all $n \geq 0$, $X - \tilde{X} = S^n X - S^n \tilde{X}$ and from $(\star)$, for all $t \geq 0$,

$$
\|X - \tilde{X}\|_t^2 \leq \frac{(C_t (t-s))^n}{n!} \|X - \tilde{X}\|_t \xrightarrow{n \to \infty} 0
$$

and therefore $X = \tilde{X}$, hence the uniqueness. \hfill \Box

**Lemma 6.2** (Grönwall lemma). If $f : [s, +\infty) \to \mathbb{R}$ is measurable and $h : [s, \infty) \to \mathbb{R}$ is continuous and increasing and $a > 0$ and $b > 0$ are constants such that for all $t \geq s$,

$$
f(t) \leq a + b \int_{s}^{t} f(u) dh_u.
$$

Then, for all $t \geq s$,

$$
f(t) \leq ae^{b(h_t - h_s)}.
$$

**Proof.** Let $\varepsilon > 0$ and $g(t) = (a + \varepsilon)e^{b(h_t - h_s)}$. We have, for all $t \geq s$,

$$
g(t) = (a + \varepsilon) + b \int_{s}^{t} g(u) dh_u,
$$

and therefore, for all $t \geq s$,

$$
f(t) - g(t) \leq -\varepsilon + b \int_{s}^{t} (f(u) - g(u)) dh_u.
$$

Set $T = \inf \{ t \geq s : f(t) - g(t) \geq 0 \} \in [0, +\infty]$. If $T < +\infty$ then, for some $T_n \downarrow T$,

$$
0 \leq f(T_n) - g(T_n) \leq -\varepsilon + b \int_{s}^{T_n} (f(u) - g(u)) dh_u,
$$

thus, using the definition of $T$,

$$
0 \leq \lim_{n \to \infty} (f(T_n) - g(T_n)) \leq -\varepsilon + b \int_{s}^{T} (f(u) - g(u)) dh_u \leq -\varepsilon
$$

which is a contradiction. Therefore $T = +\infty$ and thus $f(t) \leq g(t)$ for all $t \geq s$. \hfill \Box

**Theorem 6.3** (Dependency over initial condition). For all $s \geq 0$, for all $\mathcal{F}_s$-measurable square integrable random vectors $\eta$ and $\bar{\eta}$ of $\mathbb{R}^d$, if $X$ and $\tilde{X}$ are the solutions of

$$
X_t = \eta + \int_{s}^{t} \sigma(u, X_u) dB_u + \int_{s}^{t} b(u, X_u) du \text{ a.s., } t \geq s,
$$

and

$$
\tilde{X}_t = \bar{\eta} + \int_{s}^{t} \sigma(u, \tilde{X}_u) dB_u + \int_{s}^{t} b(u, \tilde{X}_u) du \text{ a.s., } t \geq s,
$$
then, for all \( t \geq s \), there exists a constant \( C_t > 0 \) such that
\[
\mathbb{E}\left( \sup_{s \leq u \leq t} |X_u - \bar{X}_u|^2 \right) \leq C_t \mathbb{E}(|\eta - \bar{\eta}|^2).
\]

**Proof.** We have
\[
X_t - \bar{X}_t = \eta - \bar{\eta} + \int_s^t (\sigma(u, X_u) - \sigma(u, \bar{X}_u)) \, dB_u + \int_s^t (b(u, X_u) - b(u, \bar{X}_u)) \, du.
\]

Setting \( f(t) = \mathbb{E}\left( \sup_{s \leq u \leq t} |X_u - \bar{X}_u|^2 \right) \), by the maximal inequality (Theorem 2.14), for all \( t \geq s \),
\[
f(t) \leq 3\mathbb{E}(|\eta - \bar{\eta}|^2) + 12\int_s^t |\sigma(u, X_u) - \sigma(u, \bar{X}_u)|^2 \, du + 3(t-s) \int_s^t \mathbb{E}(|b(u, X_u) - b(u, \bar{X}_u)|^2) \, du
\]
\[
\leq 3\mathbb{E}(|\eta - \bar{\eta}|^2) + c^2(12(3(t-s)) \int_s^t f(u) \, du.
\]

The desired result follows now from the Grönwall lemma (Lemma 6.2).

---

**Theorem 6.4** (Regular solution of the stochastic differential equation). For all \( s \geq 0 \), there exists a family \((X_t(x, \omega) : x \in \mathbb{R}^q, \omega \in \Omega, s \leq t)\) of random variables such that:

1. for all \( t \geq s \), the map \((x, \omega) \in \mathbb{R}^q \times \Omega \rightarrow X_t(x, \omega) \in \mathbb{R}^q\) is \((\mathcal{B}_{\mathbb{R}^q} \otimes \mathcal{F}_t)\)-measurable;

2. for all square integrable \( \mathcal{F}_s \)-measurable random vector \( \eta \) of \( \mathbb{R}^q \), the random variable \( Y_t(\omega) = X_t(x(X_0(\omega), \omega), \omega) \) solves the stochastic differential equation
\[
Y_t = \eta + \int_s^t \sigma(u, Y_u) \, dB_u + \int_s^t b(u, Y_u) \, du \quad \text{a.s.,} \quad t \geq s.
\]

**Proof.** For all \( n \geq 0 \), let \((T_k)_{k \geq 0}\) be an at most countable partition of \( \mathbb{R}^q\) such that for all \( k \geq 0 \), \( \text{diam}(T_k) \leq 2^{-n} \). For each \( k \geq 0 \), we select \( z_k \in T_k \), and we define, for all \( x \in \mathbb{R}^q \),
\[
g_n(x) = \sum_k z_k 1_{T_k}(x).
\]

Let \( z \in \mathbb{R}^q \). We consider the solution \( \bar{X}_t(z, \omega) \) of
\[
\bar{X}_t = z + \int_s^t \sigma(u, \bar{X}_u) \, dB_u + \int_s^t b(u, \bar{X}_u) \, du,
\]
for all \( t \geq s \) and all \( \omega \notin N_z \) where \( N_z \) is a negligible set. Let us define
\[
N_n = \cup_k N_{z_k} \text{ and } X^n_t(x, \omega) = \bar{X}_t(g_n(x), \omega) 1_{\Omega \setminus N_n}(\omega).
\]

The map \( (x, \omega) \rightarrow X^n_t(x, \omega) \) is \((\mathcal{B}_{\mathbb{R}^q} \otimes \mathcal{F}_t)\)-measurable and
\[
\mathbb{E}\left( \sup_{s \leq u \leq t} |X^n_u(x, \omega)| \right) \leq C_t |x - g_n(x)|^2 \leq C_t \left( \frac{1}{2^n} \right)^2.
\]

Thus
\[
\mathbb{E} \sum_{n \geq 0} \sup_{s \leq u \leq t} |X^n_u(x, \omega)| < \infty
\]
and therefore, for all \( t \geq s \), almost surely,
\[
\sup_{s \leq u \leq t} |X^n_u(x, \omega)| \longrightarrow 0.
\]

Now we define
\[
X^n_t(x, \omega) = \lim_{n \rightarrow \infty} X^n_t(x, \omega).
\]
Let $\eta$ be a square integrable $\mathcal{F}_s$-measurable random vector $\eta$ of $\mathbb{R}^q$. We have that the random variable $Y^n_t(\omega) = X^n_t(\eta(\omega), \omega)$ solves

$$Y^n_t = g_t(\eta) + \int_s^t \sigma(u, Y^n_u) dB_u + \int_s^t b(u, Y^n_u) du, \quad t \geq s.$$ 

This follows from the fact that, for all $k$, almost surely,

$$1_{\eta \in T_k} \int_s^t \sigma(u, X^n_t(z_k)) dB_u = \int_s^t 1_{\eta \in T_k} \sigma(u, X^n_u) dB_u.$$ 

Finally, for all $t \geq s$, almost surely,

$$\sup_{s \leq u \leq t} |Y^n_u - Y_u| \to 0, \quad n \to \infty,$$

where $Y = (Y_t)_{t \geq 0}$ is the solution of $(\ast)$. It follows that for all $t \geq s$, almost surely,

$$X^n_t(\eta(\omega), \omega) = Y_t(\omega).$$

\[\square\]

**Corollary 6.5.** For all $0 \leq s \leq t \leq u$, $x \in \mathbb{R}^d$, with the notions of Theorem 6.4, almost surely,

$$X^n_u(x, \omega) = X^n_t(X^n_t(s, \omega), \omega)).$$

**Proof.** We have

$$X^n_t(x) = x + \int_s^t \sigma(u, X^n_u) dB_u + \int_s^t b(v, X^n_u) dv$$

$$= x + \int_s^t \sigma(u, X^n_u) dB_u + \int_s^t b(v, X^n_u) dv + \int_s^t \sigma(u, X^n_u) dB_u + \int_s^t b(v, X^n_u) dv$$

$$= X^n_t(x) + \int_s^t \sigma(u, X^n_u) dB_u + \int_s^t b(v, X^n_u) dv$$

where the last equality holds almost surely, therefore, almost surely, from Theorem 6.4,

$$X^n_u(x, \omega) = X^n_t(X^n_t(x, \omega), \omega).$$

\[\square\]

### 6.2 Deterministic case

**Important remark about the assumptions**

In this section, we assume that $\sigma$ and $b$ do not depend on $\omega$ in other words that $\sigma$ and $b$ are measurable maps from $\mathbb{R}_+ \times \mathbb{R}^q$ to $\mathcal{M}_{a,d}(\mathbb{R})$ and $\mathbb{R}^q$ respectively, and, for simplicity, that

1. there exists a constant $C > 0$ such that for all $x, y \in \mathbb{R}^d$ and all $u \in \mathbb{R}_+$,

   $$|\sigma(u, x) - \sigma(u, y)| + |b(u, x) - b(u, y)| \leq C|x - y|;$$

2. for all $t > 0$ and all $x \in \mathbb{R}^q$,

   $$\int_0^t (|\sigma(u, x)|^2 + |b|^2(u, x)) du < \infty.$$

We denote by $(X^n_t(x, \omega))_{0 \leq t \leq 1}$ the regular solution, given by Theorem 6.4, of

$$X^n_t(x) = x, \quad X^n_t(x) = x + \int_s^t \sigma(u, X^n_u(x)) dB_u + \int_s^t b(u, X^n_u(x)) du, \quad t \geq s, \quad (\text{SDE})$$

in other words

$$X^n_t(x) = x, \quad dX^n_t(x) = \sigma(t, X^n_t(x)) dB_t + b(t, X^n_t(x)) dt, \quad t \geq s.$$
Theorem 6.6 (Markov property). For all \( s \geq 0 \) and \( x \in \mathbb{R}^d \), let \( (X^x_t(x))_{t \geq s} \) be the solution of \((SDE)\). Then for all bounded and measurable \( f : \mathbb{R}^d \rightarrow \mathbb{R} \), and for all \( u \geq t \geq s \),
\[
\mathbb{E}(f(X^x_{t+u}(x))) = \mathbb{E}(f(X^x_u(x)) | F_t) = \Pi_{t,u}(f)(X^x_t(x))
\]
almost surely, where for all \( z \in \mathbb{R}^d \),
\[
\Pi_{t,u}(f)(z) = \mathbb{E}(f(X^x_t(z))).
\]

Proof. Let \( z \in \mathbb{R}^d \) and \( 0 \leq t \leq u \). We have, almost surely,
\[
X^x_t(z) = z + \int_t^u \sigma(v, X^x_v(z)) dB_v + \int_t^u b(v, X^x_v(z)) dv
\]
where \( B^t = (B^t_s)_{s \geq 0} = (B_{t+s} - B_t)_{s \geq 0} \) is a translated Brownian motion, independent of \( F_t \).

Now, let \((\mathcal{F}_u)_{u \geq t}\) be the complete filtration given by \( \mathcal{F}_u = \sigma(B_{t+\varrho} - B_t : 0 \leq \varrho \leq u-t) \), for all \( u \geq t \). Thanks to Theorem 6.4, the following map is \((\mathcal{B}_Z \otimes \mathcal{F}_u)\)-measurable:
\[
(z, \omega) \in \mathbb{R}^d \times \Omega \rightarrow X^x_t(z, \omega) \in \mathbb{R}^d.
\]
By considering the process \( B^t \) as a random variables taking values in \( \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \), we see that there exists a measurable map \( \Theta^t_u : \mathbb{R}^d \times \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \rightarrow \mathbb{R}^d \) such that \( X^x_t(z, \omega) = \Theta^t_u(z, B^t(\omega)) \) for all \( (z, \omega) \in \mathbb{R}^d \times \Omega \). Let \( \mu \) be the Wiener measure on \( \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \), the law of \( B^t \). We have
\[
\mathbb{E}(f(X^x_t(z))) = \int_{\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)} f(\Theta^t_u(z, \omega)) \mu(d\omega) = \Pi_{t,u}(f)(z).
\]
Moreover, for all \( u \geq t \geq s \geq 0 \), from Corollary 6.5,
\[
\mathbb{E}(f(X^x_u) | \mathcal{F}_t) = \mathbb{E}(f(X^x_t(X^x_t(x)))) | \mathcal{F}_t)
\]
\[
= \mathbb{E}(f(\Theta^t_u(X^x_t(x), B^t))) | \mathcal{F}_t)
\]
\[
= \int_{\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)} f(\Theta^t_u(X^x_t(x), \omega)) \mu(d\omega)
\]
\[
= \Pi_{t,u}(f)(X^x_t(x))
\]
almost surely since \( X^x_t(x) \) is \( \mathcal{F}_t \)-measurable and since \( B^t \) is independent of \( \mathcal{F}_t \).

Remark 6.7 (Markov transition kernel and Markov semi-group). For all \( 0 \leq s \leq t \), let \( \Pi_{s,t}(x, dy) \) be the Markov transition kernel on \( \mathbb{R}^d \) defined for all \( x \in \mathbb{R}^d \) and \( A \in \mathcal{B}_{\mathbb{R}^d} \) by
\[
\Pi_{s,t}(x, A) = \mathbb{P}(X^x_t(x) \in A).
\]
It acts on bounded measurable \( f : \mathbb{R}^d \rightarrow \mathbb{R} \) as
\[
\Pi_{s,t}(f)(x) = \int_{\mathbb{R}^d} f(y) \Pi_{s,t}(x, dy), \quad x \in \mathbb{R}^d.
\]

Theorem 6.6 gives, of \( u \geq t \),
\[
\mathbb{E}(f(X^x_t(x))) = \mathbb{E}(\Pi_{t,u}(f)(X^x_t(x))) = \Pi_{s,u}(f)(x).
\]
This gives the (non-homogeneous) Markov semi-group property:
\[
\Pi_{s,u}(x, dy) = \int_{\mathbb{R}^d} \Pi_{s,t}(x, dz) \Pi_{t,u}(z, dy), \quad u \geq t \geq s \geq 0, \quad x \in \mathbb{R}^d.
\]
Moreover Theorem 6.6 shows that the following process is an \((\mathcal{F}_t)_{t\geq 0}\) martingale:

\[
(\Pi_{t,u}(f)(X^t_s(x)))_{s\leq t}\leq u.
\]

Conversely, the Markov semi-group \((\Pi_{t,u}(x,dy))_{0\leq s\leq t}\) determines entirely the law of process \((X^t_s(x))_{s\geq t}\).

Indeed, for all \(n\geq 1\), all \(0\leq s\leq t_1\leq t_2\leq \cdots \leq t_n\), and all bounded and measurable \(f_1, \ldots, f_n\) from \(\mathbb{R}^q\) to \(\mathbb{R}\), we have

\[
\mathbb{E}(f_1(X^t_{t_1}(x)) \cdots f_n(X^t_{t_n}(x)))
= \mathbb{E}(f_1(X^t_{t_1}(x)) \cdots f_{n-1}(X^t_{t_{n-1}}(x)) \Pi_{t_{n-1},t_n}(f_n)(X^t_{t_n}(x)))
= \int_{\mathbb{R}^q} \Pi_{t_{n_1},t_2}(x,dy_1) \Pi_{t_1,t_2}(y_1,dy_2) \cdots \Pi_{t_{n-1},t_n}(y_n,dy_n) f_1(y_1) \cdots f_n(y_n).
\]

**Theorem 6.8** (Uniqueness in law). Let \((\Omega, \tilde{\mathcal{F}}, (\tilde{F}_t)_{t\geq 0}, \tilde{P})\) be another filtered probability space with a complete and right continuous filtration, on which is defined a \(d\)-dimensional \((\tilde{F}_t)_{t\geq 0}\) Brownian motion \(\tilde{B} = (\tilde{B}_t)_{t\geq 0}\) issued from the origin. Let \(x \in \mathbb{R}^q\), and let \(X = (X_t(x,\omega))_{t\geq 0}\) and \(\tilde{X} = (\tilde{X}_t(x,\omega))_{t\geq 0}\) be the solutions of the respective stochastic differential equations:

\[
X_t(x) = x + \int_0^t \sigma(x, X_u(x)) dB_u + \int_0^t b(u, X_u(x)) du \quad \text{a.s.} \quad t \geq 0,
\]

and

\[
\tilde{X}_t(x) = x + \int_0^t \sigma(x, \tilde{X}_u(x)) d\tilde{B}_u + \int_0^t b(u, \tilde{X}_u(x)) du \quad \text{a.s.} \quad t \geq 0.
\]

Then these processes \(X\) and \(\tilde{X}\) have the same law on \(\mathcal{C}(\mathbb{R}_+, \mathbb{R}^q)\).

**Proof.** We consider the canonical Brownian motion \(\pi = (\pi_t(\omega))_{t\geq 0}\) defined on the Wiener space \((W = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d), \mathcal{B}_W, (\mathcal{F}_t))\) where \(\mu\) is the Wiener measure. Let \((Y_t(x, w))_{t\geq 0, w\in W}\) be the regular solution provided by Theorem 6.4 of the stochastic differential equation

\[
Y_t(x) = x + \int_0^t \sigma(x, Y_u(x)) d\pi_u + \int_0^t b(u, Y_u(x)) du
\]

\(\mu\) almost-surely. We can check easily that the processes \(Z_t(x,\omega) = Y_t(x, B(\omega)), t \geq 0, \omega \in \Omega\), and \(\tilde{Z}_t(x, \tilde{\omega}) = Y_t(x, \tilde{B}(\tilde{\omega})), t \geq 0, \tilde{\omega} \in \tilde{\Omega}\) are respectively solutions of the stochastic differential equations satisfied by \(X\) and \(\tilde{X}\). The sample path uniqueness of these solutions give that \((Y_t(x, B(\omega))_{t\geq 0} = (X_t(x,\omega))_{t\geq 0}\) \(\mathbb{P}\)-a.s. and \((Y_t(x, \tilde{B}(\tilde{\omega}))_{t\geq 0} = (\tilde{X}_t(x,\tilde{\omega}))_{t\geq 0}\) \(\tilde{\mathbb{P}}\)-a.s. But the Brownian motions \(B\) and \(\tilde{B}\) have same law on \(W = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)\), which is the Wiener measure \(\mu\), and therefore the processes \(X\) and \(\tilde{X}\) and \((Y_t)_{t\geq 0}\) have the same law on \(\mathcal{C}(\mathbb{R}_+, \mathbb{R}^q)\). 

6.3 Homogeneous case

In this section, we assume that \(\sigma\) and \(b\) do not depend on the randomness \(\omega\) and on time \(u\), in other words \(\sigma\) and \(b\) are two maps from \(\mathbb{R}^d\) to \(\mathcal{M}_{q,d}(\mathbb{R})\) and \(\mathbb{R}^q\) respectively. We also assume for simplicity that there exists a constant \(C > 0\) such that for all \(x, y \in \mathbb{R}^q\),

\[
|\sigma(x) - \sigma(y)| + |b(x) - b(y)| \leq C|x - y|.
\]

**Theorem 6.9** (Simple Markov property). Let \((X_t(x))_{t\geq 0}\) be the regular solution of the SDE

\[
X_t(x) = x + \int_0^t \sigma(X_u(x)) dB_u + \int_0^t b(X_u(x)) du \quad \text{a.s.,} \quad t \geq 0, x \in \mathbb{R}^q,
\]
provided by Theorem 6.4. Then for all \( t \geq u \geq 0 \) and all measurable and bounded \( f : \mathbb{R}^q \to \mathbb{R} \),
\[
\mathbb{E}(f(X_u(x)) | \mathcal{F}_t) = \mathbb{E}(f(X_u(x)) | X_t(x)) = \Pi_{u-t}(f)(X_t(x)) \quad \text{a.s.}
\]
where for all \( s \geq 0 \) and \( x \in \mathbb{R}^q \),
\[
\Pi_s(f)(x) = \mathbb{E}(f(X_s(x))).
\]

**Proof.** Thanks to Theorem 6.6 with \( s = 0 \) it suffices to show that for all \( u \geq t \geq 0 \),
\[
\mathbb{E}(f(X_u^t(x))) = \mathbb{E}(f(X_{u-t}^0(x))).
\]
But
\[
X^t_u(x) = x + \int_0^{u-t} \sigma(\tau_{t+s}^f(x))dB_s^t + \int_0^{u-t} b(\tau_{t+s}^f(x))ds,
\]
where \( B^t_s = B_{t+s} - B_t \) for all \( s \geq 0 \), in other words, setting \( Y_s(x) = X^t_{t+s}(x) \),
\[
Y_s(x) = x + \int_0^s \sigma(Y_u(x))dB^t_u + \int_0^s b(Y_u(x))du \quad \text{a.s.,} \quad s \geq 0,
\]
Thus the process \( Y(x) \) solves a stochastic differential equation similar to the one solves by \( X(x) \), obtained by replacing the Brownian motion \( B \) by the translation Brownian motion \( B^t \). From Theorem 6.8, it follows that the processes \( X(x) \) and \( Y(x) \) have same law, and thus, for all \( s \geq 0 \),
\[
\mathbb{E}(f(X^t_{t+s}(x))) = \mathbb{E}(f(X^0_s(x))).
\]
\[\square\]

For all \( t \geq 0 \), let \( \Pi_t(\cdot, dy) \) be the Markov transition kernel on \( \mathbb{R}^q \) defined by
\[
\Pi_t(x, A) = \mathbb{P}(X_t(x) \in A), \quad x \in \mathbb{R}^q, \quad A \in \mathcal{B}_{\mathbb{R}^q}.
\]
It acts on bounded measurable functions \( f : \mathbb{R}^q \to \mathbb{R} \) as
\[
\Pi_t(f)(x) = \int_{\mathbb{R}^q} f(y)\Pi(x, dy) = \mathbb{E}(f(X_t(x))), \quad x \in \mathbb{R}^q.
\]
It defines a homogeneous Markov semi-group \( (\Pi_t(x, dy))_{t \geq 0} \),
\[
\Pi_0 = \text{Id}, \quad \Pi_s \circ \Pi_t = \Pi_{t+s}, \quad s, t \geq 0.
\]
In other words for all \( s, t \geq 0 \) and all \( x \in \mathbb{R}^q \),
\[
\Pi_{s+t}(x, dy) = \int_{\mathbb{R}^q} \Pi_t(x, dz)\Pi_s(z, dy) = (\Pi_t \Pi_s)(x, dy).
\]

**Theorem 6.10** (Markov semi-group properties). For all \( t \geq 0 \) the operator \( \Pi_t \) preserves globally

1. the set \( \mathcal{M}(\mathbb{R}^q, \mathbb{R}) \) of bounded and measurable functions \( \mathbb{R}^q \to \mathbb{R} \);
2. the set \( \mathcal{C}_b(\mathbb{R}^q, \mathbb{R}) \) of bounded and continuous functions \( \mathbb{R}^q \to \mathbb{R} \);
3. the set \( \mathcal{C}_0(\mathbb{R}^q, \mathbb{R}) \) of bounded and continuous functions \( \mathbb{R}^q \to \mathbb{R} \) tending to zero at infinity when the coefficients \( \sigma \) and \( b \) are bounded.

**Proof.**

1. Immediate for a Markov transition kernel;
of the SDE as in Theorem 6.9. Then for all \( f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}) \) and all \( x \in \mathbb{R}^d \),

\[
L(f)(x) = \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^d b_i(x) \frac{\partial f}{\partial x_i}(x),
\]

where \( b(x) = (b_1(x), \ldots, b_d(x)) \) and \( a(x) = \sigma(x) \sigma(x)^\top \) in other words

\[
a_{i,j}(x) = \sum_{k=1}^d \sigma_{i,k}(x) \sigma_{j,k}(x).
\]

For all \( x \in \mathbb{R}^d \), the matrix \( a(x) \) is symmetric positive semi-definite, i.e. for all \( y \in \mathbb{R}^d \),

\[
a(x) y \cdot y = |(\sigma(x)^\top) y|^2 \geq 0.
\]

**Theorem 6.11 (Martingale and Duhamel\(^a\) formula).** Let \( x \in \mathbb{R}^d \) and let \((X_t(x))_{t \geq 0}\) be the regular solution of the SDE as in Theorem 6.9. Then for all \( f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}) \), the following process is an \((\mathcal{F}_t)_{t \geq 0}\) local martingale issued from the origin:

\[
M^f = (M^f_t)_{t \geq 0} = \left( f(X_t(x)) - f(x) - \int_0^t Lf(X_s(x)) \, ds \right)_{t \geq 0}.
\]

Moreover if \( f \in \mathcal{C}^2_{\text{loc}}(\mathbb{R}^d, \mathbb{R}) \) i.e. has bounded first and second order derivatives then \( M^f \) is an \((\mathcal{F}_t)_{t \geq 0}\) martingale and we have the Duhamel formula:

\[
\Pi_t(f)(x) = f(x) + \int_0^t \Pi_s(L(f))(x) \, ds, \quad t \geq 0.
\]

\(^a\)Named after Jean-Marie Duhamel (1797 – 1872), French mathematician.
Proof. Let \( f \in \mathcal{C}^2(\mathbb{R}^q, \mathbb{R}) \). The Itô formula of Theorem 5.1 applied to \( f(X_t(x)) \) gives
\[
f(X_t(x)) = f(x) + \sum_{i=1}^{d} \int_{0}^{t} \frac{\partial f}{\partial x_i}(X_s(x))dM_s^i + \sum_{i=1}^{d} \int_{0}^{t} \frac{\partial f}{\partial x_i}(X_s(x))b_i(X_s(x))ds + \frac{1}{2} \sum_{i,j=1}^{d} \int_{0}^{t} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s(x))d\langle M^i, M^j \rangle_s,
\]
where
\[
M^i_t = \sum_{k=1}^{d} \int_{0}^{t} \sigma_{i,k}(X_s(x))dB_s^k.
\]
We have
\[
\langle M^i, M^j \rangle = \sum_{k,l=1}^{d} \int_{0}^{t} \sigma_{i,k}(X_s(x))\sigma_{j,l}(X_s(x))d\langle B^k, B^l \rangle_s
= \sum_{k=1}^{d} \int_{0}^{t} \sigma_{i,k}(X_s(x))\sigma_{j,k}(X_s(x))ds
= \int_{0}^{t} a_{i,j}(X_s(x))ds.
\]
Therefore
\[
M^i_t = f(X_t(x)) - f(x) - \int_{0}^{t} L(f)(X_s(x))ds
= \sum_{i=1}^{d} \int_{0}^{t} \frac{\partial f}{\partial x_i}(X_s(x))dM_s^i
\]
is an \((\mathcal{F}_t)_{t \geq 0}\) local martingale. Moreover if \( f \in \mathcal{C}^2_0(\mathbb{R}^q, \mathbb{R}) \) then we can easily check that
\[
\mathbb{E}\int_{0}^{t} \left( \frac{\partial f}{\partial x_i}(X_s(x)) \right)^2 d\langle M^i \rangle_s < \infty, \quad 1 \leq i \leq q,
\]
and then \( M^i_t \) is a martingale. The Duhamel formula is immediate. \( \blacksquare \)

**Corollary 6.12** (Infinitesimal generator of Markov semi-group). Let us equip the space \( \mathcal{C}_0(\mathbb{R}^q, \mathbb{R}) \) with the uniform norm \( \| f \| = \| f \|_\infty = \sup_{x \in \mathbb{R}^q}|f(x)| \). The following properties hold true:

1. **Continuity.** For all \( f \in \mathcal{C}_0(\mathbb{R}^q, \mathbb{R}) \), \( \lim_{t \to 0^-} \| \Pi_t(f) - f \| = 0; \)
2. **Differentiability.** For all \( f \in \mathcal{C}^2_0(\mathbb{R}^q, \mathbb{R}) \) i.e. \( \mathcal{C}^2(\mathbb{R}^q, \mathbb{R}) \) with compact support, \( \lim_{t \to 0^-} \frac{\| \Pi_t(f) - f \|}{t} \leq Lf \).

**Proof.**

1. Suppose first that \( f \in \mathcal{C}_c(\mathbb{R}^q, \mathbb{R}) \). We have then
\[
\Pi_t(f)(x) - f(x) = \int_{0}^{t} \Pi_s(L(f))(x)ds = \mathbb{E}\int_{0}^{t} L(f)(X_s(x))ds
\]
and thus \( \| \Pi(t) - f \| \leq t \| Lf \| \to 0 \) as \( t \to \infty \). Now if \( f \in \mathcal{C}_0(\mathbb{R}^q, \mathbb{R}) \), then, for all \( \varepsilon > 0 \) and \( g \in \mathcal{C}_c(\mathbb{R}^q, \mathbb{R}) \) such that \( \| f - g \| \leq \varepsilon \), we have, for sufficiently small \( t > 0 \),
\[
\| \Pi_t(f) - f \| \leq \| \Pi_t(f - g) \| + \| \Pi_t(g) \| + \| g - f \| \leq 2\varepsilon + \| \Pi_t(g) - g \| \leq 3\varepsilon.
\]

82/98
2. For all $f \in \mathcal{C}^2_c(\mathbb{R}^q, \mathbb{R})$ and all $t > 0$, we have, using the first part for the last step,

$$\left\| \Pi_t(f) - \frac{1}{t} - Lf \right\| = \left\| \frac{1}{t} \int_0^t (\Pi_s(L(f)) - L(f))ds \right\| \leq \sup_{s \in [0, t]} \left\| \Pi_s(L(f)) - L(f) \right\| \xrightarrow{t \to 0} 0.$$

\[\square\]

**Theorem 6.13** (Strong Markov property). Let $x \in \mathbb{R}^q$ and $(X_t(x))_{t \geq 0}$ be the regular solution of the SDE as in Theorem 6.9. Let $T$ be an $(\mathcal{F}_t)_{t \geq 0}$ stopping time and let $\mathcal{F}_T$ be its stopping $\sigma$-algebra. The following properties hold true:

1. for all bounded measurable $f : \mathbb{R}^q \to \mathbb{R}$, and all $t \geq 0$,

$$\mathbb{E}(f(X_{T+t}(x)) | \mathcal{F}_t) = \Pi_t(f)(X_T(x)) 1_{T < \infty} \text{ a.s.}$$

2. for all bounded measurable $\Phi : \mathcal{C}(\mathbb{R}_+, \mathbb{R}^q) \to \mathbb{R}$,

$$\mathbb{E}(\Phi(X_{T+s}(x)) \mathbb{1}_{s \geq 0} | \mathcal{F}_T) = \Psi(X_T(x)) \mathbb{1}_{T < \infty} \text{ a.s.}$$

where the measurable function $\Psi : \mathbb{R}^q \to \mathbb{R}$ is defined for all $y \in \mathbb{R}^q$ by

$$\Psi(y) = \mathbb{E}\Phi(X_s(y))_{s \geq 0}.$$ 

**Proof.**

1. Suppose first that $T$ takes its values in an at most countable set $\mathbb{3}(T) \subset [0, \infty]$. We have to show that for all $A \in \mathcal{F}_T$ and for all $t \geq 0$,

$$\mathbb{E}(f(X_{T+t}) \mathbb{1}_{A \cap \{T < \infty\}}) = \mathbb{E}(\Pi_t(f)(X_T(x)) \mathbb{1}_{A \cap \{T < \infty\}}).$$

Indeed, using the simple Markov property of Theorem 6.9, the left hand side is equal to

$$\sum_{r \in \mathbb{3}(T) \setminus \infty} \mathbb{E}(f(X_{T+r}(x)) \mathbb{1}_{A \cap \{T = r\}}) = \sum_{r \in \mathbb{3}(T) \setminus \infty} \mathbb{E}(\Pi_t(f)(X_T(x)) \mathbb{1}_{A \cap \{T = r\}})$$

$$= \mathbb{E}(\Pi_t(f)(X_T(x)) \mathbb{1}_{A \cap \{T < \infty\}}).$$

Suppose now that $T$ takes arbitrary values in $[0, \infty)$. It suffices to prove the desired property for all bounded continuous $f$. Let us define, for all $n \geq 0$, the stopping time

$$T_n = \sum_{k \geq 0} \frac{k + 1}{2^n} 1_{[k/2^n, (k+1)/2^n]}(T) + \infty 1_{T = \infty}.$$

We have $T_n \wedge T$. For all $n \geq 0$ and all $A \in \mathcal{F}_{T_n}$, we get, from the first part of the proof,

$$\mathbb{E}(f(X_{T_n+t}(x)) \mathbb{1}_{A \cap \{T_n < \infty\}}) = \mathbb{E}(\Pi_t(f)(X_{T_n}(x)) \mathbb{1}_{A \cap \{T_n < \infty\}}).$$

By letting $n \to \infty$ and using the Lebesgue dominated convergence theorem, we obtain,

$$\mathbb{E}(f(X_{T+t}(x)) \mathbb{1}_{A \cap \{T < \infty\}}) = \mathbb{E}(\Pi_t(f)(X_T(x)) \mathbb{1}_{A \cap \{T < \infty\}}),$$

where we used Theorem 6.10 to get

$$\Pi_t(f)(X_{T_n}(x)) \mathbb{1}_{T_n < \infty} \xrightarrow{n \to \infty} \Pi_t(f)(X_T(x)) \mathbb{1}_{T < \infty} \text{ a.s.}$$

2. Suppose that $\Phi$ is cylindrical, in the sense that for some $n \geq 1$, some $s_n \geq \cdots \geq s_1 \geq 0$, and some bounded measurable $f_1, \ldots, f_n : \mathbb{R}^q \to \mathbb{R}$, we have, for all $w \in W$,

$$\Phi(w) = f_1(w_{s_n}) \cdots f_n(w_{s_1}).$$
We have in this case to show that
\[ E(f_1(X_{T+s_1}(x)) \cdots f_n(X_{T+s_n}(x)) 1_{T<\infty} | \mathcal{F}_T) = \Psi(X_T(x)) 1_{T<\infty} \quad \text{a.s.} \]
where \( \Psi : \mathbb{R}^q \to \mathbb{R} \) is the function defined for all \( y \in \mathbb{R}^q \) by
\[ \Psi(y) = E(f_1(X_{s_1}(y)) \cdots f_n(X_{s_n}(y))). \]

Indeed, for \( n = 1 \), this is the first property of the Theorem that we have already proved. We then proceed by induction on \( n \), and suppose that it is already proved for some \( n \geq 1 \). Let us prove it for \( n + 1 \). We have, denoting for short \( Y_i = f_i(X_{T+s_i}(x)) \),
\[
E(Y_1 \cdots Y_n Y_{n+1} 1_{T<\infty} | \mathcal{F}_T) = E(Y_1 \cdots Y_n E(Y_{n+1} 1_{T<\infty} | \mathcal{F}_{s_{n+1}+T}) | \mathcal{F}_T)
\]
\[
= E(Y_1 \cdots Y_n \Pi_{s_{n+1}-s_n}(f_{n+1})(X_{T+s_n}(y)) 1_{T<\infty} | \mathcal{F}_T)
\]
\[
= \Psi(X_T(x)) 1_{T<\infty}
\]
where
\[ \Psi(y) = E(f_1(X_{s_1}(y)) \cdots f_n(X_{s_n}(y)) \Pi_{s_{n+1}-s_n}(f_{n+1})(X_{s_n}(y))). \]

But using the induction hypothesis and the simple Markov property of Theorem 6.9,
\[ \Psi(y) = E(f_1(X_{s_1}(y)) \cdots f_n(X_{s_n}(y)) E(f_{n+1}(X_{s_n+1}(y)) | \mathcal{F}_{s_n})) \]
\[ = E(f_1(X_{s_1}(y)) \cdots f_n(X_{s_n}(y)) f_{n+1}(X_{s_n+1}(y))). \]

This proves the property for all cylindrical \( \Phi \). It remains to use a monotone class argument (see Section 1.7) to address the case of a general \( \Phi \).

\[ \blacksquare \]

**Theorem 6.14** (Heat equation). Assume that \( \sigma \) and \( b \) are moreover \( \mathcal{C}^2_b \), in other words \( \mathcal{C}^2 \) with bounded first and second derivatives. Let \( L \) be the differential operator (L). Then, for all \( f \in \mathcal{C}^2_b(\mathbb{R}^q, \mathbb{R}) \), there exists a unique \( \Psi = (\Psi(t, x))_{(t, x) \in \mathbb{R}_+ \times \mathbb{R}^q} \) solution of the following problem:

1. \((t, x) \mapsto \Psi(t, x)\) is \( \mathcal{C}^1 \) in \( t \) and \( \mathcal{C}^2_b \) in \( x \);

2. for all \((t, x) \in \mathbb{R}_+ \times \mathbb{R}^q,\)
\[ \frac{\partial \Psi}{\partial t}(x) = L(\Psi(t, x))(x) = \frac{1}{2} \sum_{i,j=1}^q a_{i,j}(x) \frac{\partial^2 \Psi}{\partial x_i \partial x_j}(t, x) + \sum_{i=1}^q b_i(x) \frac{\partial \Psi}{\partial x_i}(t, x); \]

3. for all \( x \in \mathbb{R}^q, \Psi(0, x) = f(x) \).

Moreover \( \Psi(x, t) \) is given by the following formula:
\[ \Psi(t, x) = E(f(X_t(x))) = \Pi_t(f)(x) \]
where \((X_t(x))_{t \geq 0}\) is the solution of the stochastic differential equation as in Theorem 6.9.

Theorem 6.14 states that the infinitesimal generator \( L \) in (L) determines the Markov semi-group \((\Pi_t)_{t \geq 0}\) which characterizes the law of the Markov diffusion process \((X_t(x))_{t \geq 0}\).

**Proof.** We admit the following result, which follows from the assumptions made on \( \sigma \) and \( b \), see [GS79]: for all \( f \in \mathcal{C}^2_b(\mathbb{R}^q, \mathbb{R}) \), the quantity \( \Pi_t(f)(x) \) is \( \mathcal{C}^2_b \) in \( x \), while \( \Pi_t(f)(x) \) is \( \mathcal{C}^1 \) in \( t \) as we can check on the Duhamel formula of Theorem 6.11:
\[ \Pi_t(f)(x) = f(x) + \int_0^t \Pi_s(L(f))(x) ds. \]
Let $t \geq u > 0$. The Itô formula of Theorem 5.1 gives

$$
\mathbb{E}(f(X_u(x)) \mid \mathcal{F}_t) = \Pi_{u-t}(f)(X_t(x))
$$

$$
= \Pi_u(f)(x) + \mathcal{N}_t + \int_0^t \left( -\frac{\partial}{\partial u}\Pi_{u-s}(f)(X_s(x)) + L(\Pi_{u-s}(f))(X_s(x)) \right) ds
$$

where $(\mathcal{N}_t)_{t \geq 0}$ is a continuous local martingale. It follows that

$$
\left( \int_0^t \left( -\frac{\partial}{\partial u}\Pi_{u-s}(f)(X_s(x)) + L(\Pi_{u-s}(f))(X_s(x)) \right) ds \right)_{t \geq 0}
$$

is a continuous local martingale, with finite variation, issued from the origin, and therefore identically equal to zero, and thus for all $s \in [0, u)$,

$$
-\frac{\partial}{\partial u}\Pi_{u-s}(f)(X_s(x)) + L(\Pi_{u-s}(f))(X_s(x)) = 0.
$$

In particular, for $s = 0$, we get

$$
\frac{\partial}{\partial u}\Pi_u(f)(x) = L(\Pi_u(f))(x).
$$

Therefore the formula for $\Psi$ which appears at the end of the statement of the Theorem provides a solution to the problem (heat equation) considered in the statement of the Theorem, since $\Pi_0(f)(x) = f(x)$. Conversely, if $\Psi(t, x)$ is a solution to this problem then for all $u > 0$ the Itô formula of Theorem 5.1 for the process $(\Psi(u-t, X_t(x)))_{t \in [0, u]}$ gives

$$
\Psi(0, X_u(x)) = f(X_u(x))
$$

$$
= \Psi(u, x) + \mathcal{N}_u + \int_0^u \left[ -\frac{\partial}{\partial u}\Psi(u-t, X_t(x)) + L(\Psi(u-t, \cdot))(X_t(x)) \right] dt
$$

$$
= \Psi(u, x) + \mathcal{N}_u
$$

where $(\mathcal{N}_u)_{u \geq 0}$ is a stochastic integral with zero expectation, and therefore

$$
\Psi(u, x) = \mathbb{E}(f(X_u(x))) = \Pi_u(f)(x).
$$

\[\blacksquare\]

6.4 Locally Lipschitz coefficients and explosion time

We assume in this section that $\sigma$ and $b$ are non-random and constant in time, defined on $\mathbb{R}^q$ and taking values in $\mathcal{M}_{q,d}(\mathbb{R})$ and $\mathbb{R}^q$ respectively, and locally Lipschitz in the sense that for all bounded subset $K$ of $\mathbb{R}^q$ there exists a constant $C_K > 0$ such that for all $x, y \in K$,

$$
|\sigma(x) - \sigma(y)| + |b(x) - b(y)| \leq C_K|x - y|.
$$

This is the case for instance when $\sigma$ and $b$ are $\mathcal{C}^1$ on $\mathbb{R}^q$.

Under this regularity assumption, for all $n \geq 1$, we can find two applications $\sigma_n$ and $b_n$ from $\mathbb{R}^q$ to $\mathcal{M}_{q,d}(\mathbb{R})$ and $m\mathbb{R}^q$ respectively such that the following properties hold true:

1. for all $x \in \mathbb{R}^q$ with $|x| \leq n$, we have $\sigma_n(x) = \sigma(x)$ and $b_n(x) = b(x)$;
2. there exists a constant $D_n > 0$ such that for all $x, y \in \mathbb{R}^q$,

$$
|\sigma_n(x) - \sigma(x)| + |b_n(x) - b(x)| \leq D_n|x - y|.
$$

Beware that these assumptions are not a specialization of the general assumptions made at the beginning of the chapter, since locally Lipschitz is strictly more general than Lipschitz. The main problem with these assumptions on $\sigma$ and $b$ is that the stochastic differential equation

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

85/98
may not have a solution $X_t(x)$ for all time $t$, and an explosion may occur in finite (random) time. In this case a simple way to still define a solution for all time is to use a localization procedure in order to define the solution process before explosion, and then to stick the process to an extra point at infinity after explosion. This suggests to consider the Alexandroff\(^1\) compactification $\overline{\mathbb{R}^d} = \mathbb{R}^d \cup \{\infty\}$ of $\mathbb{R}^d$ obtained by adding to $\mathbb{R}^d$ a point at infinity denoted $\infty$. The neighborhoods of $\infty$ in $\overline{\mathbb{R}^d}$ are the complements of the closed proper subsets of $\mathbb{R}^d$.

**Theorem 6.15 (Solving SDE with locally Lipschitz coefficients).** For all $x \in \mathbb{R}^d$, there exists a unique couple $(X^x, \xi^x)$ where $\xi^x$ is a stopping time taking values in $[0, \infty]$ called explosion time and where $X^x = (X_t(x))_{t \geq 0}$ is an adapted process such that the following properties hold true:

1. a.s. the path $t \mapsto X_t(x)$ is continuous from $[0, \xi^x)$ to $\mathbb{R}^d$ and $X_t(x) = \infty$ for all $t \geq \xi^x$;
2. almost surely, on the event $\{\xi^x < \infty\}$,
   $$\lim_{t \to \xi^x} |X_t(x)| = +\infty;$$
3. for all stopping time $T$ such that $T < \xi^x$, almost surely on $\{\xi^x < \infty\}$,
   $$X_{t \wedge T}(x) = x + \int_0^t 1_{s \leq T} \sigma(X_s(x)) dB_s + \int_0^t 1_{s \leq T} b(X_s(x)) ds \quad \text{a.s.,} \quad t \geq 0.$$

Before giving the proof of Theorem 6.15, let us prepare some ingredients. For all $m \geq 1$ and $x \in \mathbb{R}^d$, let $(X^n_t(x))_{t \geq 0}$ be the solution of the stochastic differential equation

$$X^n_t(x) = x + \int_0^t \sigma_n(X^n_s(x)) dB_s + \int_0^t b_n(X^n_s(x)) ds \quad \text{a.s.,} \quad t \geq 0.$$

For all $m \geq 1$, let us define the stopping time

$$T^n_m = \inf\{t \geq 0 : |X^n_t(x)| \geq m\}.$$

**Lemma 6.16.** For all $m \geq n \geq 1$ and $x \in \mathbb{R}^d$, we have

$$T^n_n(x) = T^n_m(x) \leq T^n_m(x) \quad \text{a.s.,}$$

and, for all $t \in [0, T^n_m(x)]$, we have

$$X^n_t(x) = X^n_{T^n_m(x)}(x) \quad \text{a.s.}$$

**Proof of Lemma 6.16.** Let us define $T = T^n_n(x) \wedge T^n_m(x)$. We have

$$X^n_{t \wedge T}(x) = x + \int_0^t 1_{s \leq T} \sigma_n(X^n_{s \wedge T}(x)) dB_s + \int_0^t 1_{s \leq T} b_n(X^n_{s \wedge T}(x)) ds.$$

and

$$X^n_{m \wedge T}(x) = x + \int_0^t 1_{s \leq T} \sigma_m(X^n_{m \wedge T}(x)) dB_s + \int_0^t 1_{s \leq T} b_m(X^n_{m \wedge T}(x)) ds.$$

By definition of $T$, the processes $(X^n_{t \wedge T})_{t \geq 0}$ and $(X^n_{m \wedge T})_{t \geq 0}$ solve the same SDE

$$Z_t = x + \int_0^t 1_{s \leq T} \sigma(Z_s) dB_s + \int_0^t 1_{s \leq T} b(Z_s) ds,$$

and thus $X^n_{t \wedge T}(x) = X^n_{m \wedge T}(x)$ a.s. for all $t \geq 0$. On the event $[0 < T < \infty)$, then for all $t \in [0, T)$,

$$|X^n_t(x)| = |X^n_T(x)| < n \quad \text{and} \quad |X^n_{T^n_m(x)}(x)| = n$$

therefore $T = T^n_n(x) = T^n_m(x)$. Now, if $T = 0$, then $|x| \geq n$ and $T = T^n_n(x) = T^n_m(x) = 0$, while if $T = \infty$, then $T^n_n(x) = T^n_m(x) = \infty.$

---

\(^1\)Named after Pavel Alexandrov (1896 – 1982), Russian mathematician.
**Proof of Theorem 6.15.** Proof of existence. We set \( \xi^x = \sup_{n \geq 0} T_n(x) \) where \( T_n(x) = T_n^x(x) \). We check immediately that if \( |x| < n \) then \( T_n(x) > 0 \) and thus \( \xi^x > 0 \). Let \( t \in [0, \xi^x) \). There exists \( n \) such that \( T_n(x) > t \) and for all \( m \geq n \), we have \( X_t^m(x) = X_t^n(x) \) almost surely. From Lemma 6.16, we can then define

\[
X_t(x) = \lim_{n \to \infty} X_t^n(x) 1_{[0,\xi^x]}(t) + \infty 1_{[\xi^x,\infty)}(t).
\]

This process \( (X_t(x))_{t \geq 0} \) verifies the first property stated by the Theorem. Moreover, on the event \( \{\xi^x < \infty\} \), we have \( T_n(x) < T_{n+1}(x) < \cdots < \xi^x \) and \( |X_{T_n(x)}| = n \), and therefore, almost surely, on the event \( \{\xi^x < \infty\}, \)

\[
\lim_{t \to \xi^x} |X_t(x)| = +\infty.
\]

Suppose that

\[
P(\lim_{t \to \xi^x} |X_t(x)| < \infty; \xi^x < \infty) > 0,
\]

then we can find real numbers \( r \) and \( R \) such that \( 0 < r < R < \infty \) and

\[
P(\lim_{t \to \xi^x} |X_t(x)| < r; \xi^x < \infty) > 0 \quad \text{and} \quad P(\lim_{t \to \xi^x} |X_t(x)| > R; \xi^x < \infty) > 0.
\]

Let \( f \in \mathcal{E}_c^2(\mathbb{R}^q, \mathbb{R}) \) be such that \( f(x) = 0 \) if \( |x| = r \) and \( f(x) = 1 \) if \( |x| = R \). If \( L \) is the differential operator \( (L) \) then, for \( n \) sufficiently large,

\[
\left( f(X_t^n(x)) - \int_0^t L(f)(X_s^n(x))ds \right)_{t \geq 0}
\]

is a martingale, and thus

\[
\left( f(X_{T_m}(x)) - \int_0^{T_m} L(f)(X_s^n(x))ds \right)_{t \geq 0}
\]

is a martingale. By letting \( n \to \infty \), the Lebesgue dominated convergence theorem gives that

\[
\left( f(X_{T_m}(x)) - \int_0^{T_m} L(f)(X_s(x))ds \right)_{t \geq 0}
\]

is a martingale. Now by letting \( m \to \infty \), we see similarly that

\[
\left( f(X_t(x)) 1_{t < \xi^x} - \int_0^{T_m} L(f)(X_s(x))ds \right)_{t \geq 0}
\]

is a martingale since \( t < \xi^x \) implies \( X_{T_m}(x) \to X_t(x) \) as \( m \to \infty \) while \( t \geq \xi^x \) implies \( f(X_{T_m}(x)) = f(X_{T_m}(x)) \to 0 \) as \( m \to \infty \). We can check (exercise) that the above martingale is continuous, and it follows then that the process \( (f(X_t(x)) 1_{t < \xi^x})_{t \geq 0} \) is continuous martingale, which contradicts \((*)\) and the definition of \( f \). It follows that the second property stated in the Theorem holds true: almost surely, on \( \{\xi^x < \infty\}, \)

\[
\lim_{t \to \xi^x} |X_t(x)| = +\infty.
\]

Furthermore, the third and last property stated in the Theorem can be deduced as follows: if \( T \) is a stopping time such that \( T < \xi^x \) almost surely on \( \{\xi^x < \infty\} \) then for all \( n \geq 1 \) we have

\[
X_{T_m}(x) = x + \int_0^T 1_{s \leq T_m} \sigma(X_s^n)dB_s + \int_0^T 1_{s \leq T_m} b(X_s^n)ds,
\]

thus,

\[
X_{T_m}(x) = x + \int_0^T 1_{s \leq T_m} \sigma(X_s)dB_s + \int_0^T 1_{s \leq T_m} b(X_s)ds,
\]

and the desired result follows by letting \( n \to \infty \).

**Proof of uniqueness.** Exercise.
6.5 Girsanov theorem

This section provides an analogue of Theorem 5.6 for stochastic differential equations. More precisely, it gives the mutual density of the law of the solution of a stochastic differential equation for different drifts and same diffusion coefficient, in particular it gives the density of the law of the solution with respect to the law of the driving Brownian motion.

Let \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) be a filtered probability space, with \((\mathcal{F}_t)_{t \geq 0}\) complete and right continuous. Let \(B = (B_t)_{t \geq 0}\) be a \(d\)-dimensional \((\mathcal{F}_t)_{t \geq 0}\) Brownian motion issued from the origin. Let us consider three Lipschitz maps

\[
\sigma : \mathbb{R}^d \to \mathcal{M}_{d,d}(\mathbb{R}), \quad b : \mathbb{R}^d \to \mathbb{R}^d, \quad \tilde{b} : \mathbb{R}^d \to \mathbb{R}^d.
\]

For all \(x \in \mathbb{R}^d\), let \((X_t(x))_{t \geq 0}\) and \((\tilde{X}_t(x))_{t \geq 0}\) be the solutions of the SDE

\[
X_t(x) = x + \int_0^t \sigma(X_s(x))dB_s + \int_0^t b(X_s(x))ds
\]

and

\[
\tilde{X}_t(x) = x + \int_0^t \sigma(\tilde{X}_s(x))dB_s + \int_0^t \tilde{b}(\tilde{X}_s(x))ds.
\]

Let \(T > 0\) be a fixed real number. Let \(Q_{x,a,b}\) and \(Q_{x,a,\tilde{b}}\) be the respective laws of the processes \((X_t)_{t \in [0,T]}\) and \((\tilde{X}_t)_{t \in [0,T]}\) on the canonical space \((W = \mathcal{C}([0,T], \mathbb{R}^d), \mathcal{B}_T, (\pi_t)_{t \in [0,T]}\) where \(\mathcal{B}_T = \sigma(\pi_s : s \in [0,T])\) and \(\pi_s(w) = w_s\) for all \(w \in W\) and all \(s \in [0,T]\).

**Theorem 6.17** (Girsanov theorem for SDE solutions). If for all \(x \in \mathbb{R}^d\), \(b(x) - \tilde{b}(x) = \sigma(x)\varphi(x)\) where \(\varphi : \mathbb{R}^d \to \mathbb{R}^d\) is measurable and bounded, then the probability measures \(Q_{x,a,b}\) and \(Q_{x,a,\tilde{b}}\) are equivalent (i.e. mutually absolutely continuous) on the measurable space \((\Omega, \mathcal{B}_T), and

\[
dQ_{x,a,b} = DdQ_{x,a,\tilde{b}}
\]

where the density \(D\) is given for all \(w \in W\) by

\[
D(w) = \exp\left(\int_0^T (a(w_s))^{-1}(b(w_s) - \tilde{b}(w_s)) \cdot dw_s - \frac{1}{2} \int_0^T (a(w_s))^{-1}(b(w_s) - \tilde{b}(w_s)) \cdot (b(w_s) - \tilde{b}(w_s)) dw_s\right) \quad \text{a.s.}
\]

where \(a(x) = \sigma(x)(\sigma(x))^\top \in \mathcal{M}_{d,d}(\mathbb{R})\) for all \(x \in \mathbb{R}^d\).

**Proof.** Note first that the formula for \(D\) makes sense since under \(Q_{x,a,b}\) or \(Q_{x,a,\tilde{b}}\) the canonical process \((\pi_s(w))_{s \geq 0} = (w_s)_{s \geq 0}\) is a semi-martingale. Note also that \(a(x)|_{\text{image}(a(x))}\) is an isomorphism on \(\text{image}(\sigma(x))\) and thus \(a^{-1}(x)(b(x) - \tilde{b}(x))\) is well defined for all \(x \in \mathbb{R}^d\).

We can write

\[
\tilde{X}_t = x + \int_0^t \sigma(\tilde{X}_s(x))dB_s + \int_0^t b(\tilde{X}_s(x))ds - \int_0^t (b(\tilde{X}_s(x)) - \tilde{b}(\til{X}_s(x)))ds
\]

\[
= x + \int_0^t \sigma(\tilde{X}_s(x))(dB_s - \varphi(\tilde{X}_s(x))ds) + \int_0^t b(\tilde{X}_s(x))ds.
\]

Let \(\tilde{P}\) be the probability measure on \((\Omega, \mathcal{B}_T)\) given by

\[
\frac{d\tilde{P}}{d\mathbb{P}} = \exp\left(\int_0^T \varphi(\tilde{X}_s(x))dB_s - \frac{1}{2} \int_0^T |\varphi(\tilde{X}_s(x))|^2 ds\right) = Z.
\]

Then, under \(\tilde{P}\), the process \((B_t - \int_0^t \varphi(\tilde{X}_s(x))ds)_{t \in [0,T]}\) is an \((\mathcal{F}_t)_{t \in [0,T]}\) Brownian motion. Therefore, from the uniqueness in law provided by Theorem 6.8, the law of \(\tilde{X}\) under \(\tilde{P}\) is equal to the law of \(X\) under \(\mathbb{P}\) which is \(Q_{x,a,b}\). Moreover the law of \(\tilde{X}\) under \(\tilde{P}\) is \(Q_{x,a,b}\). For all bounded measurable \(\Psi : W \to \mathbb{R}\), we have

\[
E_{\tilde{P}}(\Psi(\tilde{X})) = E_{\mathbb{P}}(\Psi(\tilde{X})Z) = \int_W \Psi(w)dQ_{x,a,b}(w).
\]
It remains to “compute” $Z$. Suppose first that $d = q$ and that $\sigma(x)$ is a invertible matrix. Then $b(x) - \tilde{b}(x) = \sigma(x)\varphi(x)$ implies $\varphi(x) = (\sigma(x))^{-1}(b(x) - \tilde{b}(x))$ and

$$|\varphi(x)|^2 = |(\sigma(x))^{-1}(b(x) - \tilde{b}(x))|^2 = (a(x))^{-1}(b(x) - \tilde{b}(x)) \cdot (b(x) - \tilde{b}(x)),$$

and moreover

$$d\tilde{X}_t = \sigma(\tilde{X}_t)dB_t + \tilde{b}(\tilde{X}_t)dt$$

gives

$$dB_t = (\sigma(\tilde{X}_t))^{-1}(d\tilde{X}_t - \tilde{b}(\tilde{X}_t)dt)$$

and

$$\varphi(\tilde{X}_t)dB_t = (\sigma^{-1}(\tilde{X}_t)(b(\tilde{X}_t) - \tilde{b}(\tilde{X}_t)) \cdot (\sigma(\tilde{X}_t))^{-1}(d\tilde{X}_t - \tilde{b}(\tilde{X}_t)dt)

= ((a(\tilde{X}_t))^{-1}(b(\tilde{X}_t) - \tilde{b}(\tilde{X}_t)) \cdot d\tilde{X}_t \cdot (a(\tilde{X}_t))^{-1}(b - \tilde{b})(\tilde{X}_t) \cdot \tilde{b}(\tilde{X}_t),$$

hence

$$Z = \exp\left(\int_0^T (a(\tilde{X}_t))^{-1}(b(\tilde{X}_t) - \tilde{b}(\tilde{X}_t)) \cdot d\tilde{X}_t - \frac{1}{2} \int_0^T (a(\tilde{X}_t))^{-1}(b(\tilde{X}_t) - \tilde{b}(\tilde{X}_t)) \cdot (b(\tilde{X}_t) - \tilde{b}(\tilde{X}_t))ds\right)$$

which gives the desired formula by considering the image laws, namely

$$\int_W \Psi(w) dQ_{x,\sigma,b} = \int_W \Psi(w) \exp\left(\int_0^T (a(w_s))^{-1}(b(w_s) - \tilde{b}(w_s)) \cdot dw_s - \frac{1}{2} \int_0^T (a(w_s))^{-1}(b(w_s) - \tilde{b}(w_s)) \cdot (b(w_s) - \tilde{b}(w_s))ds\right) dQ_{x,\sigma,\tilde{b}}(w).$$

To address the general case on $\sigma$, we can assume that $\varphi(x) \in \mathrm{kernel}((\sigma(x))^\top) = \mathrm{image}((\sigma(x))^\top)$ and thus $\varphi(x) = (\sigma(x))^\top \tilde{\varphi}(x)$ where $\tilde{\varphi}(x) \in \mathrm{kernel}((\sigma(x))^\top)^\perp$. This gives

$$b(x) - \tilde{b}(x) = \sigma(x)(\sigma(x))^\top \tilde{\varphi}(x) = a(x)\tilde{\varphi}(x)$$

and

$$|\varphi(x)|^2 = |(\sigma(x))^\top \tilde{\varphi}(x)|^2 = a(x)\tilde{\varphi}(x) \cdot \tilde{\varphi}(x) = (b(x) - \tilde{b}(x)) \cdot (a(x))^{-1}(b(x) - \tilde{b}(x),$$

and

$$\varphi(x) \cdot dB_t = (\sigma(x))^\top \tilde{\varphi}(x) \cdot dB_t = \tilde{\varphi}(x) \cdot \sigma(x)dB_t = (a(x))^{-1}(b(x) - \tilde{b}(x)) \cdot (d\tilde{X}_t - \tilde{b}(x)dt).$$
Chapter 7

Probabilistic formulation of Dirichlet problems

The simplest instance of the Dirichlet\(^1\) problem on an open domain \(D \subset \mathbb{R}^d\) consists in giving a function \(\Psi : \partial D \to \mathbb{R}\) at the boundary \(\partial D\) of \(D\) and seeking for a function \(f : \overline{D} = D \cup \partial D \to \mathbb{R}\) such that \(\Delta f = 0\) on \(D\) and \(f = \Psi\) on \(\partial D\).

More generally, the Dirichlet problem on an open domain \(D \subset \mathbb{R}^d\) with boundary \(\partial D\) writes

\[
\begin{align*}
L u(x) - c(x) u(x) &= f(x) \quad \text{for all } x \in D, \\
\lim_{x \to x_0} u(x) &= \Psi(x) \quad \text{for all } x_0 \in \partial D,
\end{align*}
\]

\((\text{DirP})\)

where

\[
L u(x) = \frac{1}{2} \sum_{i,j=1}^{d} a_{i,j}(x) \frac{\partial^2 u}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{d} b_i(x) \frac{\partial u}{\partial x_i}(x),
\]

where for all \(x \in \mathbb{R}^d\),

- \(a(x) = (a_{i,j}(x))_{1 \leq i,j \leq d} = \sigma(x) \sigma^*(x)\) where \(\sigma(x)\) is a \(d \times q\) matrix, Lipschitz in \(x\);
- \(b(x) = (b_i(x))_{1 \leq i \leq d}\) is a vector field, Lipschitz in \(x\);
- \(f : D \to \mathbb{R}\) and \(\Psi : \partial D \to \mathbb{R}\) are continuous and bounded;
- \(c : D \to \mathbb{R}\) is continuous and non-negative.

Let us consider the stochastic differential equation

\[
X_t^x = x + \int_0^t \sigma(X_s^x) dB_s + \int_0^t b(X_s^x) ds, \quad t \geq 0, x \in \mathbb{R}^d
\]

\((\text{SDE})\)

where \(B = (B_s)_{s \geq 0}\) is an \((\mathcal{F}_s)_{s \geq 0}\) \(d\)-dimensional Brownian motion defined on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_s)_{s \geq 0}, \mathbb{P})\) satisfying the “usual” assumptions. We define the stopping time

\[
T = T_D^x = \inf\{t \geq 0 : X_t^x \in \partial D\},
\]

with the usual convention \(\inf \emptyset = +\infty\).

**Theorem 7.1** (Kakutani\(^a\) probabilistic representation of the Dirichlet problem). If \(\mathbb{E}(T_D^X) < \infty\) for all \(x \in D\) and if \(x \in D \to u(x)\) is \(\mathcal{C}^2_b(D)\) solution of \((\text{DirP})\), then, for all \(x \in D\),

\[
u(x) = \mathbb{E}\left( \Psi(X_{T_D^x}^x) \exp\left( \int_0^{T_D^x} c(X_s^x) ds \right) \right) - \mathbb{E}\left( \int_0^{T_D^x} f(X_s^x) \exp\left(-\int_0^s c(X_u^x) du\right) ds \right).
\]

\(^a\)Named after Shizuo Kakutani (1911 – 2004), Japanese mathematician.

\(^1\)Named after Peter Gustav Lejeune Dirichlet (1805 – 1859), German mathematician.
Proof. Let us suppose first that \( u \) can be extended to the whole space \( \mathbb{R}^d \) as a \( \mathcal{C}^2_b \) function. Let us consider the process \( u(X_t^x) \exp(Y_t^x) \) where \( (X_t^x)_{t \geq 0} \) is a solution of (SDE) and where \( Y_t^x = - \int_0^t c(X_s^x) \, ds \). The Itô formula of Theorem 5.1 gives, for all \( t \geq 0 \), almost surely,

\[
 u(X_t^x) \exp(Y_t^x) - u(x) = \int_0^t \langle \nabla u(X_s^x) \exp(Y_s^x), \sigma(X_s^x) \rangle dB_s + \int_0^t f(X_s^x) \exp(Y_s^x) \, ds + \int_0^t (Lu(X_s^x) - f(X_s^x) - c(X_s^x) u(X_s^x)) \exp(Y_s^x) \, ds.
\]

This equality remains valid if one replaces \( t \) by the stopping time \( t \wedge T_D^x \), which gives, thanks to the assumptions,

\[
 \mathbb{E} \left[ u(X_{t \wedge T_D^x}^x) \exp \left( - \int_0^{t \wedge T_D^x} c(X_s^x) \, ds \right) \right] = u(x) + \mathbb{E} \left[ \int_0^{t \wedge T_D^x} f(X_s^x) \exp(Y_s^x) \, ds \right].
\]

The desired formula follows then by letting \( t \to \infty \) and using dominated convergence.

Then general case on \( u \) can be addressed as follows. We consider an increasing sequence \( (D_n)_{n \geq 0} \) of open domains with smooth boundary \( \partial D_n \) such that \( \overline{D_n} \subset D_{n+1} \subset \overline{D} \) and \( \cup_n D_n = D \). The solution \( u_n \) of the Dirichlet problem on \( D_n \) can be extended on the whole \( \mathbb{R}^d \) into a \( \mathcal{C}^2_b \) function, and \( u_n = u|_{D_n} \). We have then, for all \( x \in D_n \),

\[
 u_n(x) = u(x) = \mathbb{E} \left[ u(X_{t \wedge T_D^x}^x) \exp(Y_{t \wedge T_D^x}^x) \right] - \mathbb{E} \left[ \int_0^{T_D^x} f(X_t^x) \exp(Y_t^x) \, ds \right]
\]

where \( T_n^x = \inf \{ t \geq 0 : X_t^x \in \partial D_n \} = \inf \{ t \geq 0 : X_t^x \notin D_n \} \). We have \( T_n^x \leq T_{n+1}^x \wedge T^x < \infty \), and it suffices to let \( n \to \infty \) to get the desired result.

---

**Remark 7.2.**

1. If there exists \( a > 0 \) such that \( c(x) > a > 0 \) for all \( x \in D \) then the assumption \( \mathbb{E}(T_D^x) < \infty \) for all \( x \in D \) is useless. Namely, by letting \( t \to \infty \) in (7.1), we get

\[
 u(x) = \mathbb{E} \left[ u(X_{t \wedge T_D^x}^x) \exp \left( - \int_0^{t \wedge T_D^x} c(X_t^x) \, ds \right) \right] - \mathbb{E} \left[ \int_0^{t \wedge T_D^x} f(X_t^x) \exp(Y_t^x) \, ds \right].
\]

2. If \( f = 0 \) then the condition \( \mathbb{E}(T_D^x) < \infty \) for all \( x \in D \) can be replaced by \( \mathbb{P}(T_D^x < \infty) = 1 \) for all \( x \in D \), and, from (7.1),

\[
 u(x) = \mathbb{E} \left[ \Psi(X_{T_D^x}^x) \exp \left( - \int_0^{T_D^x} c(X_t^x) \, ds \right) \right].
\]

3. If, for some \( a > 0 \), we have, for all \( x \in D \), \( \mathbb{E}(\exp(a T_D^x)) < \infty \), then the probabilistic representation provided by Theorem 7.1 remains valid for all coefficient \( c \) such that \( \inf_x c(x) \geq -a \). In particular, if, for all \( x \in D \), \( \mathbb{P}(T_D^x \leq T(x)) = 1 \) where \( T(x) \) is deterministic, then it remains valid for all coefficients bounded below by \( c \).

4. Suppose that \( D \) is a open, bounded, with regular boundary and that \( L \) is non degenerate, then \( \mathbb{E}(T_D^x) \leq c < \infty \) for all \( x \in D \). For all continuous function \( \Psi \) on \( \partial D \), the solution \( u \) of (DirP) exists with the properties required by Theorem 7.1 provided for example that \( c \) and \( f \) are Hölder and \( c \geq 0 \). In particular, for \( f = c = 0 \), the solution writes

\[
 u(x) = \mathbb{E} \left( \Psi(X_{T_D^x}^x) \right) = \int_{\partial D} \Psi(y) \Pi(x, dy)
\]

where \( \Pi(x, y) \), \( x \in D \), \( y \in \partial D \) is the Poisson kernel of \( L \), see [Mir70]. It follows that the exit law \( \Pi(x, dy) = \mathbb{P}(X_{T_D^x}^x \in dy) \), also known as the harmonic measure, has density \( \Pi(x, dy) = \Pi(x, y) \, dy \).
The function \((x, y) \mapsto \Pi(x, y)\) is strictly positive, \(\mathcal{C}^2\) in \(x \in D\), and
\[
\begin{cases}
\lim_{x \to y_0} \Pi(x, y) = 0 & \text{for all } y, y_0 \in \partial D, y \neq y_0, \\
\lim_{x \to y_0} |\Pi(x, y_0)| = +\infty & \text{for all } y_0 \in \partial D.
\end{cases}
\]

5. In the proof of Theorem 7.1, we can simply assume that (SDE) admits a (weak) solution for all \(x \in D\), the coefficients \(\sigma\) and \(b\) being supposed measurable (not necessarily Lipschitz!). The hypothesis of continuity for \(f\), \(c\), and \(\Psi\) are superfluous as well, and one can assume that they are just measurable, \(f\) and \(\Psi\) being bounded and non-negative. This remark is useful for certain problems in stochastic control theory.

6. The probabilistic representation provided by Theorem 7.1 shows that it suffices to give \(\Psi\) on a subset \(\partial_R D\) of \(\partial D\) such that \(\mathbb{P}(X_{T_D}^x \in \partial_R D) = 1\) for all \(x \in D\). The elements of such a subset are called regular points of the boundary. This same representation shows also that the Dirichlet problem is ill posed if \(\Psi\) is arbitrary outside \(\partial R D\).

7. The Itô formula of Theorem 5.1 can be generalized to functions which are not \(\mathcal{C}^2\) but are differentiable in a weak sense, and it follows that the probabilistic representation provided by Theorem 7.1 remains valid in this general case.
Appendix A

Selected topics related to stochastic calculus

FIXME:TODO

A.1 Recurrence and transience of Brownian motion
A.2 Fokker–Planck equation
A.3 Feynman–Kac formula
A.4 Hamilton–Jacobi–Bellman equation
A.5 Statistical inference of diffusion processes
A.6 Euler–Maruyama numerical scheme
A.7 Stratonovich integral
A.8 Local time and Tanaka formula
A.9 Exemples of stochastic processes
  A.9.1 Infinitesimal generator = simulation algorithm
  A.9.2 Ornstein–Uhlenbeck and Bakry–Émery processes
  A.9.3 Laguerre and Jacobi processes
  A.9.4 Bessel and Cox–Ingersoll–Ross processes
  A.9.5 Dyson Brownian motion and Dyson–Ornstein–Uhlenbeck process
  A.9.6 Fisher–Wright processes
  A.9.7 Diffusions with jumps and piecewise deterministic Markov processes
A.10 Lévy area
A.11 Additive functionals, ergodic theorem, central limit theorem
A.12 White noise
Bibliography


