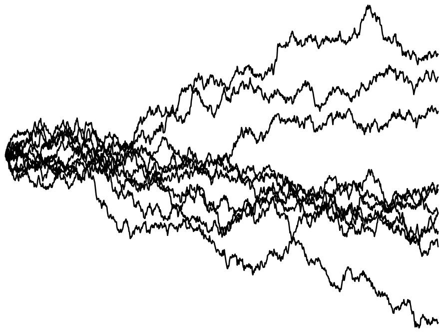
Introduction to stochastic calculus

Rough lecture notes



Master of mathematics
Université Paris-Dauphine – PSL

2020 – 2021
Corrected version
Compiled May 11, 2025

Course home page on https://djalil.chafai.net/

Dauphine UNIVERSITÉ PARIS | PSL

Suggested schedule of the lectures.

We recommend that the in class or oral lectures differ from the lecture notes, ideally they should contain less details and should be focused on the essential aspects, the structure, the culture, and the intuition.

- Lecture 1 (2 x 1.5h)
  Chapter 1 (Preliminaries)
- Lecture 2 (2 x 1.5h)
  Chapter 2 (Processes, filtrations, stopping times, martingales)
- Lecture 3 (2 x 1.5h)
  Chapter 3 (Brownian motion)
- Lecture 4 (2 x 1.5h)
  Chapter 3 (Brownian motion)
- Lecture 5 (2 x 1.5h)
  Chapter 4 (More on martingales)
- Lecture 6 (2 x 1.5h)
  Chapter 5 (Itô stochastic integral with respect to BM)
- Lecture 7 (2 x 1.5h)
  Chapter 5 (Itô stochastic integral and semi-martingales)
- Lecture 8 (2 x 1.5h)
  Chapter 6 (Itô formula and applications)
- Lecture 9 (2 x 1.5h)
  Chapter 6 (Itô formula and applications)
- Lecture 10 (2 x 1.5h)
  Chapter 7 (Stochastic differential equations)
- Lecture 11 (2 x 1.5h)
  Chapter 7 (Stochastic differential equations)
- Lecture 12 (2 x 1.5h)
  Chapter 8 (More links with partial differential equations)
- Exam

There are also separate excercises sessions (séances de travaux dirigés).

These are the lecture notes of an introduction course on stochastic calculus, given at Université Paris-Dauphine – PSL, for second year master students in mathematics[1]. The prerequisite is a probability theory course based on 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 [20]. The current version is also inspired in part from the books by Fabrice Baudoin [4] and Jean-François Le Gall [31], and by plenty of other sources. Some bits are truly original. Beware that these lecture notes are designed to constitute a rich written reference for the live course. The live course concerns only a *strict subpart focusing on intuition*, selected for being essential for understanding the concepts and techniques. At the time of writing, here are the main differences with the written lecture notes by Halim Doss before 2018:

- More on probability basics, uniform integrability, Lebesgue – Stieltjes integral
- More on martingales and local martingales
- More on examples and applications everywhere
- More on history, intuition, link with physics, programming
- Properties of Brownian motion, Dubins – Schwarz theorem, Feynman – Kac formula, Langevin processes
- More on semi-martingales, stochastic integral, and Itô formula

These lecture notes do not cover several important topics related to stochastic calculus, such as fine analysis of Brownian motion : regularity, excursions, zeros, recurrence and transcience, etc, random time change, Euler – Maruyama schemes for numerical analysis of stochastic differential equations, applications of stochastic calculus to finance, physics, biology, statistics, stochastic control, and Monte Carlo methods, Malliavin calculus, Stroock – Varadhan support theorems, local times and Tanaka formula, Schilder large deviation principle, additive functionals : law of large numbers, ergodic theorems, central limit theorems, large deviation principles, link with entropy and Poisson equation, Doob H-transforms, Friedlin – Wentzell large deviations principle for perturbation of dynamical systems, Feller branching diffusions, branching Brownian motion, Fisher – Wright diffusion, diffusions with jumps, space/time white noise, Bakry – Émery non-explosion criterion and link with Poincaré, logarithmic Sobolev, and isoperimetric functional inequalities, diffusions on manifolds, Eyrings- Kramers formula, etc. On the other hand, some topics are considered in the exams, such as Cox – Ingersoll – Ross and Bessel processes, Lévy area of planar Brownian motion, etc.

There are many other references on the subject. An accessible introduction are the books by Laurence Craig Evans [16] and by Bernt Øksendal [49]. The books by Richard Durrett [13], Philip Protter [42], and Hui-Hsiung Kuo [28] are also accessible. More advanced references include the books by Michel Métivier [36], Chris Rogers and David Williams [44, 45], Daniel Stroock and Srinivasa Varadhan [47], Ioannis Karatzas and Steven Shreve [24], Daniel Revuz and Marc Yor [43], Jean Jacod [21], Iosif Gikhman and Anatoli Skorokhod [18], and by Claude Delacherie and Paul-André Meyer [9, 10]. Finally, accessible references with exercises include the book by Francis Comets and Thierry Meyre [8] (in French) and Paolo Baldi [3] for instance.

Contributors.

- 2018 – 2021 : Djalil Chafaï
- – 2018 : Halim Doss

Glitches hunters.

- 2023 – 2024 : Aniss Fares
- 2021 – 2022 : Pauline Amrouche, Faniriana Rakoto Endor, Justin Salez
- 2020 – 2021 : Oskar Bataillon, Yi Han, Qiaoyu Luo, Gabriel Moreira-Nogueira, Diego Alejandro Murillo Taborda, Lyes Tifoun, Walid El Wahabi
- 2019 – 2020 : Oscar Cosserat, Łukasz Mądry, Alejandro Rosales Ortiz, Ziyu Zhou
- 2018 – 2019 : Clément Berenfeld

[1] MASEF (Mathématiques pour l'économie et la finance) and MATH (Mathématiques appliquées et théoriques).

Notation.

$\mathbb{R}_+$	$[0,+\infty)$
BM	Brownian motion
O, o	Landau notation
iff	if and only if
a.s.	almost surely
u.i.	uniformly integrable
w.r.t.	with respect to
$\mathbf{1}_A$	indicator of A
$x\cdot y$ or $\langle x,y\rangle$	$x_1y_1+\cdots+x_dy_d$ if $x,y\in\mathbb{R}^d$
$\lvert x\rvert$	$\sqrt{x_1^2+\cdots+x_d^2}$ if $x\in\mathbb{R}^d$
$\mathscr{B}_E$	Borel σ-algebra of E
e	exponential
d	differential element
i	the complex number $(0,1)$
d,i,j,k,m,n,ℓ	integer numbers
$p,q,r,s,t,u,v,\alpha,\beta,\varepsilon$	real numbers
$s\wedge t$ and $s\vee t$	$\min(s,t)$ and $\max(s,t)$
f is increasing	$f(y)\geq f(x)$ if $y\geq x$
$\mathrm{L}^p_{\mathbb{R}^d}(\Omega,\mathbb{P})$	$X:\Omega\to\mathbb{R}^d$ measurable with $\mathbb{E}(\\|X\\|^p)<\infty$
$\langle x,y\rangle_{\mathrm{H}}$	scalar product in the Hilbert space H
$\langle M,N\rangle$	angle bracket of local martingales M,N
$\langle M\rangle$	$\langle M,M\rangle$
$[M,N]$	square bracket of local martignales M,N
$[M]$	$[M,M]$
$X\sim\mu$	X has law μ

Some of the scientists related to Brownian motion and stochastic calculus

Life time	Scientist
(1975 –)	Martin Hairer
(1968 –)	Wendelin Werner
(1959 –)	Jean-François Le Gall
(1955 –)	Alain-Sol Sznitman
(1954 –)	Dominique Bakry
(1953 –)	Terry Lyons
(1951 –)	David Nualart
(1949 – 2014)	Marc Yor
(1947 –)	Shige Peng
(1947 –)	Étienne Pardoux
(1944 –)	Nicole El Karoui
(1944 –)	Jean Jacod
(1942 – 2004)	Catherine Doléans-Dade
(1940 –)	S. R. Srinivasa Varadhan
(1940 –)	Daniel W. Stroock
(1938 –)	Mark Iosifovich Freidlin
(1938 – 1995)	Fischer Black
(1935 –)	Shinzo Watanabe
(1934 –)	Albert Shiryaev
(1934 – 2003)	Paul-André Meyer
(1930 –)	Henry McKean
(1930 – 2011)	Anatoliy Skorokhod
(1930 – 1997)	Ruslan Stratonovich
(1927 – 2013)	Donald Burkholder
(1925 – 2010)	Paul Malliavin
(1924 – 2014)	Eugene Dynkin
(1923 – 2020)	Freeman Dyson
(1916 – 2008)	Gilbert Hunt
(1915 – 2008)	Kiyosi Itô
(1915 – 1940)	Wolfgang Doeblin
(1914 – 1984)	Mark Kac
(1911 – 2004)	Shizuo Kakutani
(1910 – 2004)	Joseph Leo Doob
(1908 – 1989)	Robert Horton Cameron
(1906 – 1970)	William Feller
(1903 – 1987)	Andrey Kolmogorov
(1900 – 1988)	George Uhlenbeck
(1896 – 1971)	Paul Lévy
(1894 – 1964)	Nobert Wiener
(1879 – 1955)	Albert Einstein
(1875 – 1941)	Henri Lebesgue
(1872 – 1946)	Paul Langevin
(1872 – 1917)	Marian Smoluchowski
(1871 – 1956)	Émile Borel
(1870 – 1942)	Jean Baptiste Perrin
(1870 – 1946)	Louis Bachelier
(1856 – 1922)	Andrey Markov
(1856 – 1894)	Thomas Joannes Stieltjes
(1773 – 1858)	Robert Brown

On ne peut non plus fixer une tangente, même de façon approchée, à aucun point de la trajectoire, et c'est un cas où il est vraiment naturel de penser à ces fonctions continues sans dérivées que les mathématiciens ont imaginées, et que l'on regarderait à tort comme de simples curiosités mathématiques, puisque la nature les suggère aussi bien que les fonctions à dérivée.

Jean Perrin (1870 – 1942), Les Atomes (1913), Chapitre 4, partie 68, [39].

Uhlenbeck's attitude to Wiener's work was brutally pragmatic and it is summarized at the end of footnote 9 in his paper (written jointly with Ming Chen Wang) "On the Theory of Brownian Motion II" (1945): the authors are aware of the fact that in the mathematical literature, especially in papers by N. Wiener, J. L Doob, and others [cf. for instance Doob (Annals of Mathematics 43, 351 1942) also for further references], the notion of a random (or stochastic) process has been defined in a much more refined way. This allows [us], for instance, to determine in certain cases the probability that the random function y(t) is of bounded variation or continuous or differentiable, etc. However it seems to us that these investigations have not helped in the solution of problems of direct physical interest and we will therefore not try to give an account of them.

Mark Kac (1914 – 1984) about George Uhlenbeck (1900 – 1988)
in *Enigmas of Chance : an autobiography* (1984).
This was before the completion of the theory of stochastic processes and stochastic calculus,
its numerical applications, and the rise of nowadays mathematical finance which is based on it.
About Brownian motion across physics and mathematics, the reader may take a look at [23, 48, 12, 38, 7].

"... *Ainsi l'intégrale et les processus d'Itô, lointains descendants de la théorie de la spéculation de Bachelier, retournent à la spéculation financière. Ils méritent à tous égards d'être intégrés dans la culture générale des mathématiciens.*"

Jean-Pierre Kahane, *Le mouvement brownien.*
Un essai sur les origines de la théorie mathématique
Société Mathématique de France, 1998.

Contents

Chapter 0

Motivation

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

$$I_t = \int_0^t X_s \mathrm{d}Y_s, \quad t \geq 0,$$

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 sub-divide the interval $[0, t]$ into $[t_0, t_1] \cup \cdots \cup [t_{n-1}, t_n]$ with $0 = t_0$ and $t_n = t$, then the hypothetic quantity I_t is naturally approximated by the (random) Riemann sum

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

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 garanteed 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[1] $\widetilde{X}_i = X_{t_i}$ 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 remarkable class of integrators called semi-martingales, which are sums of a local martingale and a finite variation process. There is also a fundamental formula of calculus[2] called the Itô formula or lemma. The set of semi-martingales is stable by stochastic integration and by composition with smooth functions.

The Itô integral allows 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

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

where for instance $B = (B_t)_{t\geq 0}$ is a Brownian motion, and where σ and b are regular enough coefficients, typically locally Lipschitz, as in the classical Cauchy – Lipschitz theorem for ordinary differential equations. In the right hand side, the first integral is an Itô stochastic integral which is a limit in probability of Riemann sums while the second integral is a Lebesgue – Stieltjes integral which is an almost sure limit of Riemann sums. Both integrals are special cases of the integral with respect to a semi-martingale. We study various properties of stochastic differential equations, including the strong Markov property, the relation to martingales and partial differential equations (Duhamel formula), and the relative asbolute continuity of the distribution of the solutions for different choices of coefficients (Girsanov formulas).

Finally we give a probabilistic interpretation of real Schrödinger operators, known as the Feynman – Kac formula, and a probabilistic representation of the solution of Dirichlet type problems, due to Kakutani.

Stochastic processes are essential in the modelling of phenomena in physics, computer science, biology, chemistry, finance, etc. Stochastic calculus is essential in the computation and estimation of distributions of stochastic objects of interests such as stopping times and solutions of stochastic differential equations. Beyond utilitarism, stochastic calculus provides also deep and aesthetic mathematics, essential for your happiness. A great thank you to Kolmogorov, Doob, Lévy, Itô, and all the others for this wonderful universe.

[1] Taking $\widetilde{X}_i = \frac{1}{2}(X_{t_i} + X_{t_{i+1}})$ leads to the Strato(novich) integral, with advantages and drawbacks.

[2] A smooth function is the integral of its derivative.

$$\left(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P}\right)$$

Unless otherwise stated, the random variables
and stochastic processes considered in this course
are defined on this enormous filtered probability space
and moreover the filtration is complete and right-continuous.

Chapter 1

Preliminaries

We refer to [14] and to [19] for the essential basic notions of probability theory (and more).

1.0 Sigma-algebras, random variables, and probabilities

A σ-field or σ-algebra – tribu in French – on a set Ω is a collection of subsets $\mathscr{A} \subset \mathscr{P}(\Omega)$ such that

- $\Omega \in \mathscr{A}$
- for all $A \in \mathscr{A}$, we have $A^c \in \mathscr{A}$
- for all at most countable family $(A_n)_n$ in $\mathscr{A}$, we have $\cap_n A_n \in \mathscr{A}$

where $A^c = \Omega \setminus A$. By combining these properties we also get $\varnothing \in \mathscr{A}$ and $\cup_n A_n \in \mathscr{A}$. We say that the couple $(\Omega, \mathscr{A})$ is a measurable space. Extreme examples of σ-algebras are $\mathscr{P}(\Omega)$ and $\{\varnothing, \Omega\}$.

- The intersection of an arbitrary family of σ-algebras is a σ-algebra.
- The σ-algebra generated by a subset of $\mathscr{P}(\Omega)$ is the $\cap$ of all the σ-algebras containing the subset.
- If Ω is equipped with a topology $\mathscr{T}$, the σ-algebra generated by $\mathscr{T}$ is called the Borel σ-algebra $\mathscr{B}$[1].

A map $f : \Omega \to E$ where $(\Omega, \mathscr{F})$ and $(E, \mathscr{E})$ are measurable is measurable when $f^{-1}(B) \in \mathscr{F}$ for all $B \in \mathscr{E}$.

A (positive) measure on a measurable space $(\Omega, \mathscr{A})$ is a map $\mu : \mathscr{A} \to [0, +\infty]$ such that

- $\mu(\varnothing) = 0$
- for all at most countable family $(A_n)_n$ of parwise disjoint elements of $\mathscr{A}$, we have $\mu(\cup_n A_n) = \sum_n \mu(A_n)$.

The triplet $(\Omega, \mathscr{A}, \mu)$ is a measured space. The measure μ is a probability measure when $\mu(\Omega) = 1$, and in this case the triplet $(\Omega, \mathscr{A}, \mu)$ is then a probability space.

A random variable X taking values in a measurable space $(E, \mathscr{E})$ is a measurable map defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$. By default we always assume that there is an underlying probability space $(\Omega, \mathscr{A}, \mathbb{P})$.

1.1 Expectation and law

If $X = \mathbf{1}_A$ then $\mathbb{E}(X) = \mathbb{P}(A)$, by linearity and monotone convergence this allows to define $\mathbb{E}(X) \in [0, +\infty]$ when X takes its values in $[0, +\infty]$. Next L^1 is the set of random variables such that $\mathbb{E}(|X|) < \infty$. If $X = X_+ - X_-$ then $|X| = X_+ + X_- \in \mathrm{L}^1$ if and only if $X_\pm \in \mathrm{L}^1$, and then we have $\mathbb{E}(|X|) = X_+ - X_-$ and $\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-)$.

The law $\mathbb{P}_X$ of a real random variable X is characterized by

- distribution: $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$ for all $B \in \mathscr{B}_{\mathbb{R}}$
- cumulative distribution function: $F_X(x) = \mathbb{P}_X((-\infty, x])) = \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$
- characteristic function: $\varphi_X(t) = \mathbb{E}_{\mathbb{P}_X}(\mathrm{e}^{\mathrm{i}t\bullet}) = \mathbb{E}(\mathrm{e}^{\mathrm{i}tX})$ for all $t \in \mathbb{R}$,

[1] Further reading: `https://djalil.chafai.net/blog/2016/03/21/integration-alpha-et-omega/`

- Laplace transform (when $X \geq 0$): $L_X(t) = \mathbb{E}_{\mathbb{P}_X}(\mathrm{e}^{-t\bullet}) = \mathbb{E}(\mathrm{e}^{-tX})$ for all $t \geq 0$.

More generally, for a random variable $X : (\Omega, \mathscr{A}) \to (E, \mathscr{B})$, the law $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ of X is a probability measure on $(E, \mathscr{B})$. This infinite dimensional dual functional object is characterized by considering its values on a sufficiently large family of test functions such as, when $(E, \mathscr{B}) = (\mathbb{R}, \mathscr{B})$, $\mathbf{1}_{(-\infty,x]}$, $x \in \mathbb{R}$, or $\mathrm{e}^{\mathrm{i}t\bullet}$, $t \in \mathbb{R}$, etc.

1.2 Independence

1. A family $(\mathscr{A}_i)_{i\in I}$ of sub-σ-algebras of $\mathscr{A}$ is independent when for all finite $J \subset I$ and all $A_i \in \mathscr{A}_i$ we have
$$\mathbb{P}(\cap_{i\in J} A_i) = \prod_{i\in J} \mathbb{P}(A_i).$$
2. We say that a family $(X_i)_{i\in I}$ of random variables is independent, $X_i : (\Omega, \mathscr{A}) \mapsto (E_i, \mathscr{B}_i)$, when the family of sub-σ-algebras $(\sigma(X_i))_{i\in I}$ is independent, where
$$\sigma(X_i) = \{X_i^{-1}(B) : B \in \mathscr{B}_i)$$
is the σ-algebra generated by X_i. Thus $(X_i)_{i\in I}$ is independent iff for all $J \subset I$ finite,
$$\mathbb{P}_{X_i : i\in J} = \otimes_{i\in J} \mathbb{P}_{X_i} \quad \text{on} \quad (\prod_{i\in J} E_i, \otimes_{i\in J} \mathscr{B}_i).$$
It follows that if $X_1, X_2, \ldots, X_n$ are real random variables integrable and independent then
$$\prod_{i=1}^n X_i \in \mathrm{L}^1 \quad \text{and} \quad \mathbb{E}\big(\prod_{i=1}^n X_i\big) = \prod_{i=1}^n \mathbb{E}(X_i).$$

1.3 Markov, Cauchy–Schwarz, Hölder, Jensen, convergence, Borel–Cantelli, LLN, LIL, CLT, ...

Markov inequality. If $U(X) \geq 0$ for a non-decreasing function U then for all $r > 0$,
$$\mathbb{P}(X \geq r) \leq \frac{\mathbb{E}(U(X))}{U(r)}.$$
This allows to control tails with moments. Conversely, we can control moments by tails via
$$\mathbb{E}(U(|X|)) = U(0) + \int_0^\infty U'(t)\mathbb{P}(|X| \geq t)\mathrm{d}t.$$
Cauchy–Schwarz inequality. In $[0, +\infty]$, with equality if and only if X and Y are colinear,
$$\mathbb{E}(XY) \leq \mathbb{E}(|X|^2)^{1/2}\mathbb{E}(|Y|^2)^{1/2}.$$
Hölder inequality. If $p \in [1, \infty]$ and $q = 1/(1-1/p) = p/(p-1)$ then, in $[0, +\infty]$,
$$\mathbb{E}(|XY|) \leq \mathbb{E}(|X|^p)^{1/p}\mathbb{E}(|Y|^q)^{1/q}.$$
Jensen inequality. If $U : \mathbb{R}^d \to \mathbb{R}$ is convex and $X \in \mathrm{L}^1$ with $U(X) \in \mathrm{L}^1$ then
$$U(\mathbb{E}(X)) \leq \mathbb{E}(U(X)),$$
moreover when U is strictly convex then equality is achieved only if X is (almost surely) constant. Useful examples include $U(x) = x^p$, $p \geq 1$, $U(x) = \mathrm{e}^{cx}$, $c \in \mathbb{R}$, $U(x) = +\infty \mathbf{1}_{x<0} + x\log(x)\mathbf{1}_{x\geq 0}$.

Convergences. Below $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$, X, Y are real random variables on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, of law μ_n, ν_n, μ, ν and cumulative distribution function F_n, G_n, F, G respectively.

Almost sure convergence. We say that $X_n \xrightarrow{\text{a.s.}} X$ 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 in the SLLN.

Convergence in probability. We say that $X_n \xrightarrow{\mathbb{P}} X$ 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 used in the weak LLN.

Mean convergence. For all $p \in [1,\infty)$, we say that $X_n \xrightarrow{\mathrm{L}^p} X$ when

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

The most useful cases are $p \in \{1,2,4\}$.

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

1. $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all bounded and continuous $f : \mathbb{R} \to \mathbb{R}$
2. $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all $\mathscr{C}^\infty$ and compactly supported $f : \mathbb{R} \to \mathbb{R}$
3. *Cumulative distribution function.* $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all $f = \mathbf{1}_{(-\infty,x]}$ with x continuity point of $\mathbb{P}(X \leq \bullet)$, in other words $F_n(x) = \mathbb{P}(X_n \leq x) \to F(x) = \mathbb{P}(X \leq x)$ as soon as F is continuous at x
4. *Fourier transform or characteristic function.* $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all $f = \mathrm{e}^{\mathrm{i}t\bullet}$, $t \in \mathbb{R}$
5. *Laplace transform.* **(on $\mathbb{R}_+$)** $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for all $f = \mathrm{e}^{-t\bullet}$, $t \geq 0$.

Contrary to the other modes of convergence, the convergence in law does not depend on the law of the couple (X_n, X) and uses only marginal laws. The Fourier and Laplace transforms convert 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, and this is referred to sometimes as the continuous mapping theorem.

The notions of convergence extend naturally to random vectors by using a distance/norm/scalar product, for instance for the characteristic function by replacing $\mathrm{i}tX$ by $\mathrm{i}\langle t, X\rangle$.

$$\begin{array}{ccccc} & & \boxed{\mathrm{L}^p \text{ CV}} & & \\ & & \Downarrow & & \\ & & \boxed{\mathrm{L}^1 \text{ CV}} & & \\ & & \Downarrow & & \\ \boxed{\text{a.s. CV}} & \Rightarrow & \boxed{\text{CV in } \mathbb{P}} & \Rightarrow & \boxed{\text{CV in law}} \end{array}$$

If X is constant then the convergence in law implies the convergence in probability. The convergence in L^1 is equivalent to uniform integrability and convergence in probability.

Monotone convergence theorem. If $(X_n)_{n\geq 1}$ takes its values in $[0,+\infty]$ and $\nearrow$ then

$$\mathbb{E}(\lim_{n\to\infty} X_n) = \lim_{n\to\infty} \mathbb{E}(X_n) \in [0,+\infty].$$

Fatou lemma. If $(X_n)_{n\geq 1}$ takes its values in $[0,+\infty]$ then

$$\mathbb{E}(\varliminf_{n\to\infty} X_n) \leq \varliminf_{n\to\infty} \mathbb{E}(X_n) \in [0,+\infty].$$

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.

Scheffé lemma. If $X_n, X \in \mathrm{L}^1$, $X_n \xrightarrow{\text{a.s.}} X$ then $X_n \xrightarrow{\mathrm{L}^1} X$ iff $\mathbb{E}(|X_n|) \to \mathbb{E}(|X|)$.

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

Fubini–Tonelli theorem. Let $(\Omega_1, \mathscr{A}_1, \mu_1)$ and $(\Omega_2, \mathscr{A}_2, \mu_2)$ two measurable spaces, and let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a measurable function. If $f \geq 0$ or if $f \in \mathrm{L}^1(\mu_1 \otimes \mu_2)$ then

$$\int f(x,y)\mathrm{d}(\mu_1 \otimes \mu_2)(x,y) = \int \Big(\int f(x,y)\mathrm{d}\mu_1(x)\Big)\mathrm{d}\mu_2(y).$$

Borel–Cantelli lemma. Let $(A_n)_n$ be events in a probability space $(\Omega, \mathscr{A}, \mathbb{P})$. We define

$$\begin{cases} \underline{\lim}_n A_n &= \cup_n \cap_{m\geq n} A_m = \{\omega \in \Omega : \omega \in A_n \text{ for } n \text{ large enough}\}, \\ \overline{\lim}_n A_n &= \cap_n \cup_{m\geq n} A_m = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many values of } n\}. \end{cases}$$

We have $(\underline{\lim}_n A_n^c)^c = \overline{\lim}_n A_n$, and $\overline{\lim}_n \mathbf{1}_{A_n} = \mathbf{1}_{\overline{\lim}_n A_n}$ and $\underline{\lim}_n \mathbf{1}_{A_n} = \mathbf{1}_{\underline{\lim}_n A_n}$.

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

The Borel–Cantelli lemma is a great provider of almost sure convergence. Note that if X takes its values in $[0, +\infty]$ then $\mathbb{E}(X) < \infty$ implies $\mathbb{P}(X < \infty) = 1$, and this allows to prove the Cantelli part:

$$\sum_n \mathbb{P}(A_n) = \sum_n \mathbb{E}\mathbf{1}_{A_n} = \mathbb{E}\sum_n \mathbf{1}_{A_n} \quad \text{and} \quad \Big\{\sum_n \mathbf{1}_{A_n} = \infty\Big\} = \overline{\lim_n}\, A_n.$$

Strong Law of Large Numbers (SLLN). If $X \in \mathrm{L}^1$ and $X_1, X_2, \ldots$ are i.i.d.[2] copies of X then, with $m = \mathbb{E}(X)$,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{a.s.}} m \quad \text{and} \quad \frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\mathrm{L}^1} m.$$

Central limit theorem (CLT). If moreover $X \in \mathrm{L}^2$, then with $\sigma^2 = \mathrm{Var}(X) = \mathbb{E}((X-m)^2) = \mathbb{E}(X^2) - m^2$,

$$\frac{\sqrt{n}}{\sigma}\Big(\frac{X_1 + \cdots + X_n}{n} - m\Big) = \frac{X_1 - m + \cdots + X_n - m}{\sqrt{n}\sigma} \xrightarrow[n\to\infty]{\text{law}} \mathscr{N}(0,1).$$

Law of iterated logarithm (LIL). Under the assumptions and with the notation of the CLT, almost surely

$$\overline{\lim_{n\to\infty}}\Big(\frac{\sqrt{n}}{\sigma\sqrt{2\log\log(n)}}\Big(\frac{X_1 + \cdots + X_n}{n} - m\Big)\Big) = \overline{\lim_{n\to\infty}}\Big(\frac{X_1 - m + \cdots + X_n - m}{\sqrt{2n\log\log(n)}\sigma}\Big) = 1$$

and

$$\underline{\lim_{n\to\infty}}\Big(\frac{\sqrt{n}}{\sigma\sqrt{2\log\log(n)}}\Big(\frac{X_1 + \cdots + X_n}{n} - m\Big)\Big) = \underline{\lim_{n\to\infty}}\Big(\frac{X_1 - m + \cdots + X_n - m}{\sqrt{2n\log\log(n)}\sigma}\Big) = -1$$

Note that the CLT gives $\frac{X_1+\cdots+X_n}{n} - m \xrightarrow[n\to\infty]{\mathbb{P}} 0$, which is a weak form of LLN.

1.4 Uniform integrability

For any family $(X_i)_{i\in I} \subset \mathrm{L}^1$, the following three properties are equivalent[3]. When one (and thus all) of these properties holds true, we say that the family $(X_i)_{i\in I}$ is uniformly integrable (u.i.) or equi-integrable[4]. The first property can be seen as a natural definition of uniform integrability.

1. *(definition of uniform integrability)* $\lim_{r\to+\infty} \sup_{i\in I} \mathbb{E}(|X_i|\mathbf{1}_{|X_i|\geq r}) = 0$

[2] Independent and identically distributed, in French “indépendantes et identiquement distribuées”.
[3] Further reading: `https://djalil.chafai.net/blog/2014/03/09/de-la-vallee-poussin-on-uniform-integrability/`
[4] The terminology comes from the fact that by dominated convergence, we have $X \in \mathrm{L}^1$ if and only if $\lim_{r\to\infty} \mathbb{E}(|X|\mathbf{1}_{|X|\geq r}) = 0$.

2. *(epsilon-delta criterion)* the family is bounded in L^1 in the sense that

$$\sup_{i\in I}\mathbb{E}(|X_i|)<\infty$$

and moreover $\forall\varepsilon>0$, $\exists\delta>0$, $\forall A\in\mathscr{F}$, $\mathbb{P}(A)\le\delta\Rightarrow\sup_{i\in I}\mathbb{E}(|X_i|\mathbf{1}_A)\le\varepsilon$

3. *(de la Vallée Poussin[5] boundedness in L^U criterion)* there exists a non-decreasing convex $U:\mathbb{R}_+\to\mathbb{R}_+$ such that $\lim_{x\to+\infty}U(x)/x=+\infty$ and such that $(U(|X_i|))_{i\in I}$ is bounded in L^1, namely

$$\sup_{i\in I}\mathbb{E}(U(|X_i|))<\infty.$$

Note that this implies boundedness in L^1, and is implied by boundedness in L^p with $p>1$.

Here are examples for uniformly integrable families:

- every finite subset of L^1 is uniformly integrable. In particular if $X\in\mathrm{L}^1$ then there exists a non-decreasing convex and super-linear U such that $U(|X|)\in\mathrm{L}^1$, but beware that this U depends on X.
- if $(X_i)_{i\in I}$ is bounded in L^p with $p>1$ then it is u.i.
- if $\sup_{i\in I}|X_i|\in\mathrm{L}^1$ (domination: $|X_i|\le X\in\mathrm{L}^1$ for all $i\in I$) then $(X_i)_{i\in I}$ is u.i.
- if $\mathscr{T}\in\{\mathbb{N},\mathbb{R}_+\}$ and $X_t\xrightarrow[t\to\infty]{\mathrm{L}^1}X\in\mathrm{L}^1$ then $(X_t)_{t\in\mathscr{T}}$, $(X_t)_{t\in\mathscr{T}}\cup\{X\}$, and $(X_t-X)_{t\in\mathscr{T}}$ are u.i.
- if $X\in\mathrm{L}^1$ and $X_i=\mathbb{E}(X\mid\mathscr{F}_i)$ for all $i\in I$ for σ-algebras $(\mathscr{F}_i)_{i\in I}$ then $(X_i)_{i\in I}$ is uniformly integrable.

The notion of uniform integrability leads to a stronger version of the dominated convergence theorem: for any $p\ge1$, and for any random variables X and $(X_t)_{t\in\mathscr{T}}$, $\mathscr{T}\in\{\mathbb{N},\mathbb{R}_+\}$, we have

$$X_t,X\in\mathrm{L}^p\text{ and }X_t\xrightarrow[t\to\infty]{\mathrm{L}^p}X\quad\text{if and only if}\quad(|X_t|^p)_{t\in\mathscr{T}}\text{ is u.i. and }X_t\xrightarrow[t\to\infty]{\mathbb{P}}X$$

In particular the convergence in probability together with u.i. implies $X\in\mathrm{L}^1$, which is remarkable!
The dominated convergence theorem corresponds to the special case $\sup_{t\in\mathscr{T}}|X_t|\in\mathrm{L}^1$.

1.5 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:
   - (orthogonality) for all $z\in F$, $x-y\perp z$ namely $\langle x,z\rangle=\langle y,z\rangle$
   - (variational: least squares) for all $z\in F$, $\|x-y\|\le\|x-z\|$ namely $\|x-y\|=\min_{z\in F}\|x-z\|$.

2. Let $(\Omega,\mathscr{A},\mathbb{P})$ be a probability space and $\mathscr{F}$ be a sub-σ-algebra of $\mathscr{A}$. Let us consider the Hilbert space $H=\mathrm{L}^2(\Omega,\mathscr{A},\mathbb{P})$. The set $F=\mathrm{L}^2(\Omega,\mathscr{F},\mathbb{P})$ is a closed sub-space of H. If $X\in H$, 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:

$$Y\in\mathrm{L}^2(\Omega,\mathscr{F},\mathbb{P})\text{ and, for all }Z\in\mathrm{L}^2(\Omega,\mathscr{F},\mathbb{P}),\ \mathbb{E}(|X-Y|^2)\le\mathbb{E}(|X-Z|^2).$$

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

   - for all $Z\in\mathrm{L}^2(\Omega,\mathscr{F},\mathbb{P})$, $\mathbb{E}(XZ)=\mathbb{E}(YZ)$, or even for all $B\in\mathscr{B}$, $\mathbb{E}(X\mathbf{1}_B)=\mathbb{E}(Y\mathbf{1}_B)$.

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

[5] After Charles-Jean Étienne Gustave Nicolas de la Vallée Poussin (1866 – 1962), Belgian mathematician.

3. If now $X \in \mathrm{L}^1(\Omega,\mathscr{A},\mathbb{P})$, we define by extension $Y = \mathbb{E}(X \mid \mathscr{F})$, a real random variable characterized by

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

   (b) for all Z bounded and $\mathscr{F}$ measurable, $\mathbb{E}(XZ) = \mathbb{E}(YZ)$, or for all $B \in \mathscr{F}$, $\mathbb{E}(X\mathbf{1}_B) = \mathbb{E}(Y\mathbf{1}_B)$.

   Proof. Let μ be the bounded measure on $(\Omega,\mathscr{F})$ defined by $\mu(B) = \mathbb{E}(X\mathbf{1}_B)$, $B \in \mathscr{F}$. Set $\nu = \mathbb{P}_{\mathscr{F}}$. For all $B \in \mathscr{F}$, if $\nu(B) = 0$ then $\mu(B) = 0$. From the Radon–Nikodym theorem, there exists a unique $Y \in \mathrm{L}^1(\Omega,\mathscr{F},\nu)$ such that $\int_B Y \mathrm{d}\nu = \mu(B)$, for all $B \in \mathscr{F}$ in other words $\mathbb{E}(Y\mathbf{1}_B) = \mathbb{E}(X\mathbf{1}_B)$, for all $B \in \mathscr{F}$. ■

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

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

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

$$\mathrm{var}(X) = \mathbb{E}((X-\mathbb{E}(X))^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + (\mathbb{E}(X)^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

which follows in fact from the Pythagoras theorem in L^2. More generally we have

$$\begin{aligned}\mathrm{var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}((\mathbb{E}(X \mid \mathscr{F}))^2) + \mathbb{E}((\mathbb{E}(X \mid \mathscr{F}))^2) - (\mathbb{E}(X))^2 \\ &= \mathbb{E}(\mathrm{var}(X \mid \mathscr{F})) + \mathrm{var}(\mathbb{E}(X \mid \mathscr{F}))\end{aligned}$$

where $\mathrm{var}(X \mid \mathscr{F}) = \mathbb{E}(X^2 \mid \mathscr{F}) - (\mathbb{E}(X \mid \mathscr{F}))^2$. Note that by definition of $\mathbb{E}(X \mid \mathscr{F})$,

$$\begin{aligned}\inf_{Y:\sigma(Y)\subset\mathscr{F}} \mathbb{E}((X-Y)^2) &= \mathbb{E}((X-\mathbb{E}(X \mid \mathscr{F}))^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X \mid \mathscr{F})) + \mathbb{E}((\mathbb{E}(X \mid \mathscr{F}))^2) \\ &= \mathbb{E}(X^2) - \mathbb{E}((\mathbb{E}(X \mid \mathscr{F}))^2) \\ &= \mathbb{E}(\mathrm{var}(X \mid \mathscr{F})).\end{aligned}$$

Note that $\mathbb{E} = \mathbb{E}(\cdot \mid \mathscr{T})$ where $\mathscr{T} = \{\varnothing,\Omega\}$. The conditional expectation generalizes the expectation and has all the properties of an expectation, and more. Namely, for all sub-σ-algebra $\mathscr{F}$ of $\mathscr{A}$:

- **Linearity.** for all $\alpha,\beta \in \mathbb{R}$ and $X, Y \in \mathrm{L}^1$, $\mathbb{E}(\alpha X + \beta Y \mid \mathscr{F}) = \alpha\mathbb{E}(X \mid \mathscr{F}) + \beta\mathbb{E}(Y \mid \mathscr{F})$

- **Independence.** If X is independent of $\mathscr{F}$ (always the case when X is constant) then $\mathbb{E}(X \mid \mathscr{F}) = \mathbb{E}(X)$

- **Factorization.** If X is $\mathscr{F}$-measurable, $Y \in \mathrm{L}^1$, $XY \in \mathrm{L}^1$, then $\mathbb{E}(XY \mid \mathscr{F}) = X\mathbb{E}(Y \mid \mathscr{F})$, in particular we recover the "projection property" $\mathbb{E}(X \mid \mathscr{F}) = X$ if $X \in \mathrm{L}^1(\Omega,\mathscr{F},\mathbb{P})$ which is the case when X is constant

- **Composed "projections" or "tower property".** For all sub-σ-algebras $\mathscr{F},\mathscr{G}$ with $\mathscr{G} \subset \mathscr{F}$ and all $X \in \mathrm{L}^1$,

  $$\mathbb{E}(\mathbb{E}(X \mid \mathscr{F}) \mid \mathscr{G}) = \mathbb{E}(\mathbb{E}(X \mid \mathscr{G}) \mid \mathscr{F}) = \mathbb{E}(X \mid \mathscr{G}),$$

  and in particular for all $X \in \mathrm{L}^1$, $\mathbb{E}(\mathbb{E}(X \mid \mathscr{F})) = \mathbb{E}(X)$, and if X is constant then $\mathbb{E}(X \mid \mathscr{F}) = X$.

- **Normalization.** $\mathbb{E}(\mathbf{1}_\Omega \mid \mathscr{F}) = \mathbf{1}_\Omega$ (follows from some of the properties above)

- **Positivity or monotonicity.** For all $X, Y \in \mathrm{L}^1$, if $X \le Y$ then $\mathbb{E}(X \mid \mathscr{F}) \le \mathbb{E}(Y \mid \mathscr{F})$, or equivalently for all $X \in \mathrm{L}^1$ if $X \ge 0$ then $\mathbb{E}(X \mid \mathscr{F}) \ge 0$. In particular for all $X \in \mathrm{L}^1$,

  $$|\mathbb{E}(X \mid \mathscr{F})| \le \mathbb{E}(|X| \mid \mathscr{F})$$

- **Convexity.** Jensen inequality: for all non-negative convex $U:\mathbb{R}^d\to\mathbb{R}$ and all $X\in\mathrm{L}^1$,

$$U(\mathbb{E}(X\mid\mathscr{F}))\le\mathbb{E}(U(X)\mid\mathscr{F}).$$

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

$$|\mathbb{E}(XY\mid\mathscr{F})|\le(\mathbb{E}(|X|^p\mid\mathscr{F})^{1/p}\mathbb{E}(|Y|^q\mid\mathscr{F})^{1/q}.$$

  The Cauchy–Schwarz inequality corresponds to the special case $p=q=1/2$

- **Monotone convergence.** If $X_n\ge0$, $X_n\nearrow X$, $X\in\mathrm{L}^1$, then $\mathbb{E}(X_n\mid\mathscr{F})\nearrow\mathbb{E}(X\mid\mathscr{F})$. This allows to define $\mathbb{E}(X\mid\mathscr{F})$ for all non-negative random variable X taking values in $[0,+\infty]$.

Theorem 1.5.1. Transfer or the meaning of being measurable.

If $T:\Omega\to(F,\mathscr{F})$ are $Y:\Omega\to(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ and random variables then Y is $\sigma(T)$ measurable if and only if there exists $g:(F,\mathscr{F})\to(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ measurable such that $Y=g\circ T$.

Proof. If $Y=\mathbf{1}_A$ for $A\in\sigma(T)$, then $A=T^{-1}(B)$ for some $B\in\mathscr{F}$, and therefore $Y=\mathbf{1}_B\circ T$. If $Y=\sum_{i\in I}a_i\mathbf{1}_{A_i}$ with I finite and $A_i=T^{-1}(B_i)$, $B_i\in\mathscr{F}$, then $Y=(\sum_{i\in I}a_i\mathbf{1}_{B_i})\circ T$. 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 $Y_n\nearrow Y$, and $Y_n=g_n\circ T$. By setting $g=\overline{\lim}\,g_n$, we get $Y=g\circ T$. Finally, if Y is just $\sigma(T)$ measurable, then it suffices to write $Y=Y_+-Y_-$. ■

Let $X\in\mathrm{L}^1(\Omega,\mathscr{A},\mathbb{P})$ and let $T:(\Omega,\mathscr{A})\to(F,\mathscr{F})$ 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 $g:(F,\mathscr{F})\to(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ with $\mathbb{E}(X\mid T)=g(T)$ and $g(T)\in\mathrm{L}^1$
2. For all $h:(F,\mathscr{F})\to(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ measurable and bounded,

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

If $X\in\mathrm{L}^2$ then, thanks to the transfer theorem (Theorem 1.5.1), 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 $(\Omega,\mathscr{F},\mathbb{P})$, an event $A\in\mathscr{F}$, and a sub-σ-algebra $\mathscr{A}\subset\mathscr{F}$, the quantity $\mathbb{P}(A\mid\mathscr{A})=\mathbb{E}(\mathbf{1}_A\mid\mathscr{A})$ is a random variable taking its values in $[0,1]$. Similarly, conditioning with respect to an event makes sense in the sense that $\mathbb{E}(X\mid A)=\mathbb{E}(X\mid\mathbf{1}_A=1)$, and

$$\begin{aligned}\mathbb{E}(X\mid\mathbf{1}_A)&=\frac{\mathbb{E}(X\mathbf{1}_A)}{\mathbb{P}(A)}\mathbf{1}_A+\frac{\mathbb{E}(X\mathbf{1}_{A^c})}{\mathbb{P}(A^c)}\mathbf{1}_{A^c}\\&=\mathbb{E}(X\mid\mathbf{1}_A=1)\mathbf{1}_A+\mathbb{E}(X\mid\mathbf{1}_A=0)\mathbf{1}_{A^c}.\end{aligned}$$

Finally, when X and Y take their values in an at most countable set then

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

Remark 1.5.2. Conditional expectation as averaging of residual randomness.

Let X and Y be random variables defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$, and let $\mathscr{A}$ be a sub-σ-algebra of $\mathscr{F}$. If X is independent of $\mathscr{A}$ and if Y is $\mathscr{A}$-measurable, then, using the monotone class theorem, for all $\mathscr{F}$-measurable and bounded or positive $f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, we get

$$\mathbb{E}(f(X,Y)\mid\mathscr{A})=\mathbb{E}(f(X,Y)\mid Y)=g(Y)\quad\text{where}\quad g(y)=\mathbb{E}(f(X,y)).$$

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 $(E,\mathscr{E})$ and $(F,\mathscr{F})$ respectively. The conditional law of X given Y is a family $(N(y,\cdot))_{y\in F}$ of probability measures on $(E,\mathscr{E})$, in other words a transition kernel, such that for all $A\in\mathscr{E}$, the map $y\in F\mapsto N(y,A)\in[0,1]$ 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 $A\in\mathscr{E}$. We sometimes speak about disintegration of measure.

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 $y\in F$, $N(y,\cdot)=\mathbb{P}_X$ where $\mathbb{P}_X$ is the law of X.

If (X,Y) has Lebesgue density $f_{X,Y}$ then X and Y have densities $f_X=\int f(\cdot,y)\mathrm{d}y$ and $f_Y=\int f(x,\cdot)\mathrm{d}x$ and the conditional law $\mathrm{Law}(X\mid Y=y)$ has density $f_{X\mid Y=y}=f_{X,Y}(x,y)/f_Y(y)$, in such a way that

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

1.6 Gaussian random vectors

A random vector $X=(X_1,\dots,X_n)$ of $\mathbb{R}^n$ is a Gaussian random vector when every linear combination of its components is Gaussian, namely for all $\alpha_1,\dots,\alpha_n\in\mathbb{R}$ the real random variable $\alpha_1X_1+\cdots+\alpha_nX_n$ is Gaussian.

Let X be a random vector with mean vector and the covariance matrix

$$m=\mathbb{E}(X)=(\mathbb{E}(X_1),\dots,\mathbb{E}(X_n))\quad\text{and}\quad\Sigma=\big(\mathbb{E}((X_j-m_j)(X_k-m_k))\big)_{1\leq j,k\leq n}$$

Then X is Gaussian iff its characteristic function is given for all $t\in\mathbb{R}^n$ by

$$\varphi_X(t)=\mathbb{E}\big(\mathrm{e}^{\mathrm{i}tX}\big)=\mathrm{e}^{\mathrm{i}tm-\frac{1}{2}\langle\Sigma t,t\rangle}.$$

We denote this law $\mathcal{N}(m,\Sigma)$. Beware that when $n=1$, we denote $\Sigma=\sigma^2$.

We say that $\mathcal{N}(0,I_d)$ is the standard Gaussian.

The law $\mathcal{N}(m,\Sigma)$ has a density iff Σ is invertible, given by

$$x\in\mathbb{R}^n\mapsto\frac{\exp\Big(-\frac{1}{2}\langle\Sigma^{-1}(x-m),x-m\rangle\Big)}{\sqrt{(2\pi)^n\det(\Sigma)}},$$

otherwise $\mathcal{N}(m,\Sigma)$ is supported by a strict sub-vector space of $\mathbb{R}^n$.

If $(X_1,\dots,X_n)$ is a Gaussian random vector, then $X_1,\dots,X_n$ are independent iif Σ is diagonal.

If $Z\sim\mathcal{N}(0,I_n)$ and $m\in\mathbb{R}^d$ and $A\in\mathcal{M}_{d,n}(\mathbb{R})$ then $AZ\sim\mathcal{N}(m,AA^\top)$ is a Gaussian random vector of $\mathbb{R}^d$.

Coding in action 1.6.1. Simulation.

Write a Python[a] or Julia[b] program for the simulation of a sample of $\mathcal{N}(m,\Sigma)$ knowing m and Σ. What is the best way to reduce to the one-dim. case? What is the best way to find A such that $AA^\top=\Sigma$?

[a] `https://en.wikipedia.org/wiki/Python_(programming_language)`
[b] `https://en.wikipedia.org/wiki/Julia_(programming_language)`

1.7 Bounded variation and Lebesgue – Stieltjes integral

Definition 1.7.1. p-variation of a function on a finite interval.

Let $[a,b] \subset \mathbb{R}$ be a finite interval. For all $p \geq 1$, the p-variation of a function $f : [a,b] \to \mathbb{R}$ is defined by

$$\|f\|_{p-\mathrm{var}} = \Big(\sup_{t_k} \sum_k |f(t_{k+1}) - f(t_k)|^p\Big)^{1/p} \in [0,+\infty]$$

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 $f : [a,b] \to \mathbb{R}$ 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 $f : [a,b] \to \mathbb{R}$ if of bounded variation and is differentiable with integrable derivative then

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

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

Theorem 1.7.2. Representation of bounded variation functions on a finite interval.

Let $[a,b] \subset \mathbb{R}$ be a finite interval. For all $f : [a,b] \mapsto \mathbb{R}$, the following properties are equivalent:

1. f is of bounded variation
2. f is the difference of two positive increasing functions $[a,b] \to \mathbb{R}$.

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

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

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 increasing (and bounded) by definition. It suffices now to show that $G = F - f$ is increasing. We observe that for all $t_1 < t_2$ in $[a,b]$, we have $F(t_1) + f(t_2) - f(t_1) \leq F(t_1) + |f(t_2) - f(t_1)| \leq F(t_2)$, and thus

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

$2 \Rightarrow 1$. 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 it is of bounded variation since for all sub-division $a = t_0 < \cdots < t_n = b$, $n \geq 1$,

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

∎

The notion of bounded variation is used for the Lebesgue – Stieltjes integral in stochastic calculus.

Theorem 1.7.3. Lebesgue – Stieltjes integral of continuous finite variation integrators.

Let $[a,b] \subset \mathbb{R}$ be a finite interval. Let $f : [a,b] \to \mathbb{R}$ be right continuous and of bounded variation.

Then there exists a unique finite signed Borel measure μ_f on $([a,b],\mathscr{B}_{[a,b]})$ 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 $\mathrm{d}\mu_f=\mathrm{d}f$, and for all measurable $g:[a,b]\to\mathbb{R}$, positive or in $\mathrm{L}^1(|\mu_f|)$,

$$\int_a^b g(t)\mathrm{d}f(t)=\int g\mathrm{d}\mu_f.$$

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

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

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

$$\int_a^b k(t)\mathrm{d}h(t)=\int_a^b k(t)\mathrm{d}\int_a^t g(s)\mathrm{d}f(s)=\int_a^b k(t)g(t)\mathrm{d}f(t).$$

In particular when $f(t)=t$ for all $t\geq 0$ then on all $[a,b]\subset[0,\infty)$, the measure μ_f is the Lebesgue measure and for all measurable $g:\mathbb{R}_+\mapsto\mathbb{R}$ which is locally bounded or positive, we have, for all $t\geq 0$,

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

Theorem 1.7.3 is used in stochastic calculus with $f(t)=V_t(\omega)$, $t\geq 0$, and for almost all fixed $\omega\in\Omega$ where $V=(V_t)_{t\geq 0}$ is a finite variation process, for instance $V=\langle M\rangle$ where M is a continuous local martingale. In particular when $M=B$ is Brownian motion then $V_t=t$ is deterministic and we recover the example above.

Theorem 3.2.1 says that Brownian motion has a.s. sample paths of infinite variation on any interval. In particular the assumptions of Theorem 1.7.3 are not satisfied when $f(t)=B_t(\omega)$, $t\in[a,b]\subset[0,+\infty)$.

Proof. First part. Theorem 1.7.2 gives $f=f_+-f_-$ where $f_\pm\geq 0$ are bounded and increasing. This reduces the problem to the case where f is increasing and μ_f is a positive Borel measure. In this case, the result follows from the Carathéodory extension theorem (Theorem 1.8.5). Note: μ_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 $t\in(t_k^{(n)},t_{k+1}^{(n)}]$ for some $k\in\{0,\ldots,m_n-1\}$. Then $g^{(n)}$ is measurable, we have $\lim_{n\to\infty}g^{(n)}(t)=g(t)$ for all $t\in[a,b]$, and moreover $\sup_n\sup_{t\in[a,b]}|g^{(n)}(t)|\leq\sup_{t\in[a,b]}|g(t)|<\infty$. By dominated convergence in $\mathrm{L}^1(|\mu|)$, we obtain

$$\sum_k g(t_k^{(n)})(f(t_{k+1}^{(n)})-f(t_k^{(n)}))=\int g^{(n)}\mathrm{d}\mu_f\underset{n\to\infty}{\longrightarrow}\int g\mathrm{d}\mu_f=\int_a^b g(t)\mathrm{d}f(t).$$

Note that if g is measurable and not continuous, then $g^{(n)}\to g$ as $n\to\infty$, almost everywhere on $[a,b]$, which is suitable for the Lebesgue measure but not necessarily for the measure $|\mu|$ which is of interest here.

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

$$\int_a^s g(t)\mathrm{d}f(t)=\int g\mathrm{d}\mu_{f\mathbf{1}_{[a,s]}}=\int g\mathbf{1}_{[a,s]}\mathrm{d}\mu_f.$$

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

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

Finally, to prove the formula, it suffices to check it for $k=\mathbf{1}_{[a,c]}$ for $c\in[a,b]$. This writes $\mu_h(c)-\mu_h(a)=\int_a^c g(t)\mathrm{d}\mu_f(t)=h(c)-h(a)$, which is the definition of μ_h. Note that by construction we have $h(a)=0$. ∎

Remark 1.7.4. Riemann – Stieltjes – Young integral.

Following L.C. Young, it can be shown that if $f, g : [a,b] \to \mathbb{R}$ are continuous with f of finite p-var. and g of finitie q-var. with $1/p + 1/q > 1$, then the Riemann – Stieltjes integral is well defined:

$$\int_a^b f(t)\mathrm{d}g(t) = \lim_{n\to\infty} \sum_{k=0}^{m_n} f(t_k^{(n)})(g(t_{k+1}^{(n)}) - g(t_k^{(n)})),$$

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

1.8 Monotone class theorem and Carathéodory extension theorem

Definition 1.8.1. π-systems and λ-systems.

- We say that $\mathscr{C} \subset \mathscr{P}(\Omega)$ is a π-system when $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$
- We say that $\mathscr{S} \subset \mathscr{P}(\Omega)$ is a λ-system (or monotone class or Dynkin[6] system) when
  - $\cup_n A_n \in \mathscr{S}$ for all $(A_n)_n$ such that $A_n \subset A_{n+1}$ and $A_n \in \mathscr{S}$ for all n
  - $A \setminus B \in \mathscr{S}$ for all $A, B \in \mathscr{S}$ such that $B \subset A$.

[6] Named after Eugene Dynkin (1924 – 2014), Soviet and American mathematician.

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

A basic yet important example of λ-system is given by $\{A \in \mathscr{A} : \mathbb{P}(A) = \mathbb{Q}(A)\}$ where $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on $(\Omega, \mathscr{A})$, see Corollary 1.8.4 for an application.

Lemma 1.8.2. σ-algebras.

A λ-system that contains Ω and which is a π-system is a σ-algebra.

Note that conversely, a σ-algebra is always a π-system, but not a λ-system in general, due to the second property of λ-systems which is not necessarily valid for a σ-algebra when $A \neq \Omega$.

Proof. If a λ-system $\mathscr{S} \subset \mathbb{P}(\Omega)$ contains Ω and is a π-system then for all $A, B \in \mathscr{S}$ we have

$$A \cup B = \Omega \setminus ((\Omega \setminus A) \cap (\Omega \setminus B)),$$

which means that $\mathscr{S}$ is table by finite union. This allows to drop the non-decreasing condition in the stability of $\mathscr{S}$ by countable union, which simply means finally that $\mathscr{S}$ is a σ-algebra. ■

Theorem 1.8.3. Dynkin π-λ Theorem.

If $\mathscr{S} \subset \mathscr{P}(\Omega)$ is a λ-system containing Ω and including a π-system $\mathscr{C}$,
then $\mathscr{S}$ contains also the σ-algebra $\sigma(\mathscr{C})$ generated by $\mathscr{C}$.

Proof. The λ-system generated by a subset of $\mathscr{P}(\Omega)$ is by definition the intersection of all λ-systems which include this subset. This intersection is not empty since it contains $\mathscr{P}(\Omega)$, and we can check that it is a λ-system. It is the smallest (for the inclusion) λ-system containing the initial subset of $\mathscr{P}(\Omega)$.

Let $\mathscr{S}'$ be the λ-system generated by $\mathscr{C}$ and Ω. It suffices to show that $\mathscr{S}'$ is a σ-algebra. For that, and thanks to lemma 1.8.2, it suffices to show that $\mathscr{S}'$ is a π-system. To do so, let us define

$$\mathscr{S}_1 = \{A \in \mathscr{S}' : A \cap B \in \mathscr{S}' \text{ for all } B \in \mathscr{C}\},$$

which is a λ-system including Ω and containing $\mathscr{C}$, hence $\mathscr{S}_1 \subset \mathscr{S}'$, and thus $\mathscr{S}_1 = \mathscr{S}'$. Now,

$$\mathscr{S}_2 = \{A \in \mathscr{S}' : A \cap B \in \mathscr{S}' \text{ for all } B \in \mathscr{S}'\}$$

is a λ-system containing Ω and including $\mathscr{S}$ and thus $\mathscr{S}_2 = \mathscr{S}'$, hence $\mathscr{S}'$ is a π-system. ■

Corollary 1.8.4. Sierpiński[a] – Dynkin (functional) monotone class theorem.

[a] Named after Wacław Sierpiński (1882 – 1969), Polish mathematician.

1. For all probability measures $\mathbb{P}$ and $\mathbb{Q}$ on a measurable space $(\Omega, \mathscr{A})$, if $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \mathscr{C}$ where $\mathscr{C}$ is a π-system such that $\sigma(\mathscr{C}) = \mathscr{A}$, then $\mathbb{P} = \mathbb{Q}$
2. Let H be a vector space of bounded measurable functions $(\Omega, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ such that
   (a) H is stable by monotone convergence namely if $(f_n)_n$ is a sequence in H such that $f_n \nearrow f$ pointwise with f bounded then $f \in H$
   (b) H contains constant functions namely $\mathbf{1}_\Omega \in H$, is stable by product namely if $f, g \in H$ then $fg \in H$, and contains all $\mathbf{1}_A$ for all A in a π-system $\mathscr{C}$ on Ω such that $\sigma(\mathscr{C}) = \mathscr{A}$

   then H contains all $\mathscr{A}$-measurable bounded functions $\Omega \to \mathbb{R}$.

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.

Proof.

1. Take $\mathscr{S} = \{A \in \mathscr{A} : \mathbb{P}(A) = \mathbb{Q}(A)\}$ and use Theorem 1.8.3.
2. Take $\mathscr{S} = \{A \in \mathscr{A} : \mathbf{1}_A \in H\}$ and use Theorem 1.8.3.

■

Theorem 1.8.5. Carathéodory extension theorem.

Let $\Omega \neq \varnothing$, $\mathscr{A} \subset \mathscr{P}(\Omega)$, and $\mu : \mathscr{A} \mapsto \mathbb{R}_+$. Let $\sigma(A)$ be the σ-algebra generated by $\mathscr{A}$. If

1. $\Omega \in \mathscr{A}$
2. (stability by complement) for all $A \in \mathscr{A}$, we have $A^c = \Omega \setminus A \in \mathscr{A}$
3. (stability by intersection) for all $A, B \in \mathscr{A}$, we have $A \cap B \in \mathscr{A}$
4. μ is σ-additive and σ-finite

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

Proof. See for instance [4]. The uniqueness can be deduced from Corollary 1.8.4. ■

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 $(\Omega, \mathscr{F}, \mathbb{P})$, and taking values in some measurable space $(G, \mathscr{B})$. By default a process takes real values. In general G is a metric space, with distance denoted d, complete, separable, and $\mathscr{B}$ is its Borel σ-algebra.

2.1 Measurability

The natural filtration of a process $(X_t)_{t\geq 0}$ is the increasing family $(\mathscr{F}_t)_{t\geq 0}$ of sub-σ-algebras of $\mathscr{F}$ defined for all $t \geq 0$ by $\mathscr{F}_t = \sigma(X_s : 0 \leq s \leq t)$. More generally, an increasing family $(\mathscr{F}_t)_{t\geq 0}$ of sub-σ-algebras of $\mathscr{F}$ is called a filtration. For a given filtration $(\mathscr{F}_t)_{t\geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$, 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 for all $t \geq 0$, X_t is $\mathscr{F}_t$ measurable
- measurable when for all $t \geq 0$, $(s, \omega) \in [0, t] \times \Omega \mapsto X_s(\omega)$ is $\mathscr{B}_{[0,t]} \otimes \mathscr{F}$ measurable
- progressive when for all $t \geq 0$, $(s, \omega) \in [0, t] \times \Omega \mapsto X_s(\omega)$ is $\mathscr{B}_{[0,t]} \otimes \mathscr{F}_t$ measurable
- right-continuous (respectively left-continuous, 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, continuous)
- square integrable when for all $t \geq 0$, $\mathbb{E}(X_t^2) < \infty$
- 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, $\sup_{t\geq 0} |X_t| \leq C$
- locally bounded when for almost all $\omega \in \Omega$ and all $t \geq 0$, $\sup_{s\in[0,t]} |X_s(\omega)| < \infty$
- of finite variation when almost surely $t \mapsto X_t$ is of bounded variation on all finite intervals of $\mathbb{R}_+$, equivalently is the difference of two positive increasing processes, see Theorem 1.7.2
- Feller continuous when $x \mapsto \mathbb{E}(f(X_t) \mid X_0 = x)$ is continuous for all $t \geq 0$ and bounded continuous f.

Theorem 2.1.1. Progressive σ-field and progressive processes.

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

2. If $X = (X_t)_{t\geq 0}$ is adapted right-continuous or left-continuous defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and taking its values in a metric space (E, d) equipped with its Borel σ-algebra, then X is progressive. In particular continuous adapted implies 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

$$X_s^n = \begin{cases} X_{kt/n}^n & \text{if } s \in [(k-1)t/n, kt/n),\, 1 \leq k \leq n, \\ X_t & \text{if } s = t. \end{cases}$$

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

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

Since $(X_t)_{t\geq 0}$ is adapted, this set belongs to $\mathscr{F}_t \otimes \mathscr{B}_{[0,t]}$. Therefore, for all $n \geq 1$, the function $(\omega, s) \in \Omega \times [0, t] \mapsto X_s^n(\omega)$ is measurable for $\mathscr{F}_t \otimes \mathscr{B}_{[0,t]}$. 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 $\mathscr{F}_t \otimes \mathscr{B}_{[0,t]}$, which means, since $t > 0$ is arbitrary, that $(X_t)_{t\geq 0}$ is progressive.

■

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" $\mathscr{P}(\mathbb{R}_+, \mathbb{R}^d)$ of functions from $\mathbb{R}_+$ to $\mathbb{R}^d$. The measurability is for free if we equip $\mathscr{P}(\mathbb{R}_+, \mathbb{R}^d)$ with the σ-algebra $\mathscr{A}_{\mathscr{P}(\mathbb{R}_+, \mathbb{R}^d)}$ generated by the cylindrical events

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

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 $\mathscr{P}(\mathbb{R}_+, \mathbb{R}^d)$ is so big that $\mathscr{A}_{\mathscr{P}(\mathbb{R}_+, \mathbb{R}^d)}$ turns out to be too small, and does not contain for instance events of interest such that $\{f \in \mathscr{P}(\mathbb{R}_+, \mathbb{R}^d) : \sup_{t\in[0,1]} f(t) < 1\}$.

We focus in this course on continuous processes. This suggests to consider $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$ and the σ-algebra $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)}$ generated by the cylindrical events $\{f \in \mathscr{C}(\mathbb{R}_+, \mathbb{R}^d) : f(t_1) \in I_1, \ldots, f(t_n) \in I_n\}$ 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 2.1.2. What a wonderful world.

On $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$, the following σ-algebras coincide:

- σ-algebra $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)}$ generated by the cylindrical events
- Borel σ-algebra $\mathscr{B}_{\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)}$ 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 $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$ equipped with the distance

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

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 $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)} \subset \mathscr{B}_{\mathscr{C}(\mathbb{R}_+, \mathbb{R})}$ since the σ-algebra $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+, \mathbb{R})}$ is generated by the cylinders

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

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

$$\{f \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}) : \max_{t\in[0,n]} |f(t) - g(t)| \leq r\} = \cap_{t\in\mathbb{Q}\cap[0,n]}\{f \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}) : |f(t) - g(t)| \leq r\}$$

belongs to $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})}$, and since these sets generate $\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$, we get $\mathscr{A}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)} = \mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})}$. ■

Theorem 2.1.3. Continuous processes as random variables on path space.

Let $X = (X_t)_{t\geq 0}$ be a continuous d-dimensional process defined on $(\Omega,\mathscr{F},\mathbb{P})$. Let $\Omega' \in \mathscr{F}$ such that $\mathbb{P}(\Omega') = 1$ and $\Omega' \subset \{X_\bullet \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)\}$. Then the map $X|_{\Omega'} : \omega \in \Omega' \to X_\bullet(\omega) \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)$ is measurable with respect to the σ-algebras $\mathscr{F}' = \{F \cap \Omega' : A \in \mathscr{A}\}$ and $\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$.

Proof. Let us consider an arbitrary cylindrical event

$$F = \{f \in \mathscr{C}(\mathbb{R}_+,\mathbb{R}^d) : f(t_1) \in I_1,\ldots, f(t_n) \in I_n\},$$

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

$$\Omega' \cap \{X_\bullet \in F\} = \Omega' \cap \{X_{t_1} \in I_1,\ldots, X_{t_n} \in I_n\} \in \mathscr{F}'.$$

Now $\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^d)}$ is generated by cylindrical events (Theorem 2.1.2). ■

Remark 2.1.4. Equality of processes, modification and indistinguishability.

Two processes $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ defined on the same probability space $(\Omega,\mathscr{F},\mathbb{P})$ are indistinguishable when for almost all $\omega \in \Omega$ the sample paths $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ coincide, namely

$$\mathbb{P}(\forall t \geq 0 : X_t = Y_t) = 1.$$

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 $\mathbb{R}^d$ with same 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 vectors $(X_{t_1},\ldots,X_{t_n})$ and $(Y_{t_1},\ldots,Y_{t_n})$ have same law in $(\mathbb{R}^d)^n$, then X and Y have same law as random variables on the path space $(\mathscr{P}(\mathbb{R}_+,\mathbb{R}),\mathscr{A}_{\mathscr{P}(\mathbb{R}_+,\mathbb{R}^d)})$. The following theorem provides a sort of converse, stated when $d = 1$ for simplicity.

Theorem 2.1.5. Kolmogorov extension theorem.

For all $n \geq 1$ and all $t \in \mathbb{R}^n$ with $0 \leq t_1 \leq \cdots \leq t_n$, let $\mu_{t_1,\ldots,t_n}$ be a probability measure on $\mathbb{R}^n$. Let us assume the following consistency condition:

- for all $n \geq 1$, $t \in \mathbb{R}^n$ with $0 \leq t_1 \leq \cdots \leq t_n$, and all $A_1,\ldots,A_{n-1} \in \mathscr{B}_\mathbb{R}$, we have

$$\mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_{n-1} \times \mathbb{R}) = \mu_{t_1,\ldots,t_{n-1}}(A_1 \times \cdots \times A_{n-1}).$$

Then there exists a unique probability measure μ on the path space $(\mathscr{P}(\mathbb{R}_+,\mathbb{R}),\mathscr{A}_{\mathscr{P}(\mathbb{R}_+,\mathbb{R})})$ such that for all $n \geq 1$, all $t \in \mathbb{R}^n$ with $0 \leq t_1 \leq \cdots \leq t_n$, and all $A_1,\ldots,A_n \in \mathscr{B}_\mathbb{R}$, we have

$$\mu(\pi_{t_1} \in A_1,\ldots,\pi_{t_n} \in A_n) = \mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n),$$

where $\pi_t(\omega) = \omega_t$, namely $\pi_t : \omega \in \mathscr{P}(\mathbb{R}_+,\mathbb{R}) \mapsto \omega_t \in \mathbb{R}$ for all $t \geq 0$.

Proof. For a cylindrical event $A_{t_1,\ldots,t_n}(B) = \{f \in \mathscr{P}(\mathbb{R}_+,\mathbb{R}) : (f(t_1),\ldots,f(t_n)) \in B\}$ where $n \geq 1$, $t \in \mathbb{R}^n$ with $0 \leq t_1 \leq \cdots \leq t_n$, and where $B \in \mathscr{B}_{\mathbb{R}^n}$, we define $\mu(B) = \mu_{t_1,\ldots,t_n}(B)$. This makes sense thanks to the consistency condition. Note that we could drop the ordering on the coordinates of t by defining $\mu_{t_1,\ldots,t_n} = \mu_{t_{(1)},\ldots,t_{(n)}}$ where $t_{(1)} \leq \cdots \leq t_{(n)}$ is the reordering. Moreover $\mu(\mathscr{P}(\mathbb{R}_+,\mathbb{R})) = 1$. Since the set of cylinders satisfies the assumptions of the Carathéodory extension theorem (Theorem 1.8.5), and generates the σ-algebra $\mathscr{A}_{\mathscr{P}(\mathbb{R}_+,\mathbb{R})}$, it remains to show that μ is a σ-finite measure, which is the difficult part of the proof. See instance [4]. ■

2.2 Completeness

Contrary to discrete processes, continuous processes lead naturally to measurability issues.

In a probability space $(\Omega,\mathscr{F},\mathbb{P})$, we say that $A \subset \Omega$ is negligible when there exists $A' \in \mathscr{F}$ with $A \subset A'$ and $\mathbb{P}(A') = 0$. We say that the $(\Omega,\mathscr{F},\mathbb{P})$ is complete when $\mathscr{F}$ contains the negligible subsets of Ω.

A filtration $(\mathscr{F}_t)_{t\geq 0}$ on $(\Omega,\mathscr{F},\mathbb{P})$ is complete when $\mathscr{F}_0$ contains the negligible subsets of $\mathscr{F}$.

Completeness emerges naturally via almost sure events which are complement of negligible subsets.

Theorem 2.2.1. Measurability of running supremum from completeness.

Let $(X_t)_{t\geq 0}$ be a continuous process defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and taking values in a topological space E equipped with its Borel σ-field $\mathscr{E}$. Let $f : E \to \mathbb{R}$ be a measurable function.

- If $(\Omega,\mathscr{F},\mathbb{P})$ is complete then $\sup_{s\in[0,t]} f(X_s)$ is measurable for all $t \geq 0$.
- If X is adapted with respect to a complete filtration $(\mathscr{F}_t)_{t\geq 0}$ then $(\sup_{s\in[0,t]} f(X_s))_{t\geq 0}$ is adapted.

Proof. Let $\Omega' \in \mathscr{F}$ be an almost sure event on which X is continuous. Set $S_t = \sup_{s\in[0,t]} f(X_s)$.

- For all $t \geq 0$ and $A \in \mathscr{E}$, we have
$$\Omega' \cap \{S_t \in A\} = \Omega' \cap \Big\{\sup_{s\in[0,t]\cap\mathbb{Q}} f(X_s) \in A\Big\} \in \mathscr{F},$$
while $(\Omega \setminus \Omega') \cap \{S_t \in A\} \subset \Omega \setminus \Omega'$ is negligible and thus belongs to $\mathscr{F}$ by completeness of $(\Omega,\mathscr{F},\mathbb{P})$.
- Same argument as before with $\mathscr{F}_t$ instead of $\mathscr{F}$.

■

The notion of completeness is relative to the probability measure $\mathbb{P}$. There is also a notion of universal completeness, see [9], that do not depend on the probability measure, but we do not use it in these notes.

2.3 Stopping times

Definition 2.3.1. Stopping time.

A map $T : \Omega \to [0,+\infty]$ is a stopping time or optional time for a filtration $(\mathscr{F}_t)_{t\geq 0}$ on $(\Omega,\mathscr{F},\mathbb{P})$ when $\{T \leq t\} \in \mathscr{F}_t$ for all $t \geq 0$. All constant non-negative random variables are stopping times.

Contrary to discrete time filtrations, the notion of stopping times for continuous time filtration leads naturally to the notions of complete filtration and right continuous filtration.

In a probability space $(\Omega,\mathscr{F},\mathbb{P})$, we say that $A \subset \Omega$ is negligible when there exists $A' \in \mathscr{F}$ with $A \subset A'$ and $\mathbb{P}(A') = 0$. A filtration $(\mathscr{F}_t)_{t\geq 0}$ on $(\Omega,\mathscr{F},\mathbb{P})$ is complete when $\mathscr{F}_0$ contains the negligible subsets of $\mathscr{F}$, in particular all almost sure events. We say then that $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq 0},\mathbb{P})$ is a complete filtered probability space.

Theorem 2.3.2. Hitting times as archetypal examples of stopping times.

Let $X = (X_t)_{t\geq 0}$ be a continuous and adapted process on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ with respect to a complete filtration $(\mathscr{F}_t)_{t\geq 0}$, and taking its values in a metric space G equipped with its Borel σ-field.

Then, for all closed subset $A \subset G$, the hitting time $T_A : \Omega \to [0, +\infty]$ of A, defined by

$$T_A = \inf\{t \geq 0 : X_t \in A\},$$

with convention $\inf \varnothing = +\infty$, is a stopping time.

For instance $T_n = T_{[n,\infty)} = \inf\{t \geq 0 : |X_t| \geq n\}$ when $G = \mathbb{R}^d$.

Proof. Let Ω' be the almost sure event on which X is continuous. On Ω', since X is continuous and A is closed, we have $\{t \geq 0 : X_t \in A\} = \{t \geq 0 : \mathrm{dist}(X_t, A) = 0\}$, the map $t \geq 0 \mapsto \mathrm{dist}(X_t, A)$ is continuous, and the inf in the definition of T_A is a min. Now, since X is adapted, we have, for all $t \geq 0$,

$$\Omega' \cap \{T_A \leq t\} = \Omega' \cap \bigcap_{s \in [0,t] \cap \mathbb{Q}} \{X_s \in A\} \in \mathcal{F}_t,$$

where we have also used the fact that $\Omega' \in \mathcal{F}_t$ for all $t \geq 0$ since $(\mathcal{F}_t)_{t \geq 0}$ is complete. On the other hand, $(\Omega \setminus \Omega') \cap \{T_A \leq t\} \subset \Omega \setminus \Omega'$ is negligible, and belongs then to $\mathcal{F}_t$ for all $t \geq 0$ since $(\mathcal{F}_t)_{t \geq 0}$ is complete. ■

We say that a filtration $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous when $\mathcal{F}_t = \mathcal{F}_{t^+}$ for all $t \geq 0$ where

$$\mathcal{F}_{t+} = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon} = \bigcap_{s > t} \mathcal{F}_s.$$

Theorem 2.3.3. Stopping times: alternative definition.

If $T : \Omega \to [0, +\infty]$ is a stopping time with respect to a filtration $(\mathcal{F}_t)_{t \geq 0}$ then $\{T < t\} \in \mathcal{F}_t$ for all $t \geq 0$. Conversely this property implies that T is a stopping time when the filtration is right-continuous.

Proof. If T is a stopping time then for all $t \geq 0$ we have

$$\{T < t\} = \bigcup_{n=1}^{\infty} \{T \leq t - 1/n\} \in \mathcal{F}_t,$$

(note also that $\{T = t\} = \{T \leq t\} \cap \{T < t\}^c \in \mathcal{F}_t$). Conversely $\{T \leq t\} \in \cap_{s > t} \mathcal{F}_s = \mathcal{F}_{t+}$ since for all $s > t$,

$$\{T \leq t\} = \bigcap_{n=1}^{\infty} \{T < (t + 1/n) \wedge s\} \in \mathcal{F}_s.$$

■

This can be skipped at first reading.

The following generalizes Theorem 2.3.2 to hitting times of arbitrary measurable subsets by progressive processes, at the price of assuming right continuity of the filtration in addition to completeness.

Theorem 2.3.4: Hitting times are stopping times reloaded.

Let $X = (X_t)_{t \geq 0}$ be a progressive process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a right continuous and complete filtration $(\mathcal{F}_t)_{t \geq 0}$, and taking its values in a measurable space G. Then for all measurable subset $A \subset G$, the hitting time $T_A : \Omega \to [0, +\infty]$ defined by

$$T_A = \inf\{t \geq 0 : X_t \in A\},$$

with convention $\inf \varnothing = +\infty$, is a stopping time.

Example of progressive processes include adapted right-continuous processes.

Proof. The debut D_B of any $B \in \mathscr{F} \otimes \mathscr{B}(\mathbb{R}_+)$ is defined for all $\omega \in \Omega$ by

$$D_B(\omega) = \inf\{t \geq 0 : (\omega, t) \in B\} \in [0, +\infty].$$

If B is progressive, then D_B is a stopping time (this is known as the debut theorem). Indeed, for all $t \geq 0$ the set $\{D_B < t\}$ is then the projection on Ω of $\{s \in [0,t) : (\omega, s) \in B\}$, which belongs to $\mathscr{B}(\mathbb{R}_+) \otimes \mathscr{F}_t$ since B is progressive. Since the filtration is right-continuous and complete, this projection[a] belongs to $\mathscr{F}_t$. Now $\{D_B < t\} \in \mathscr{F}_t$ for all $t \geq 0$ implies that D_B is a stopping time since the filtration is right continuous (Theorem 2.3.3). Finally it remains to note that $T_A = D_B$ with $B = \{(\omega, t) : X_t \in A\}$, which is progressive as the pre-image of $\mathbb{R}_+ \times A$ by the map $(\omega, t) \mapsto X_t(\omega)$ (recall that X is progressive). ■

[a]See [9, Th. IV.50 page 116]. This is related to a famous mistake made by the French Henri Lebesgue (1875 – 1941) on the measurability of projections of measurable sets in product spaces, that motivated the Russian Nikolai Luzin (1883 – 1950) and his student Mikhail Yakovlevich Suslin (1894 – 1919) to forge the concept of analytic set and descriptive set theory.

Remark 2.3.5. Canonical filtration.

It is customary to assume that the underlying filtration is right-continuous and complete. For a given filtration $(\mathscr{F}_t)_{t\geq 0}$, it is always possible to consider its completion $(\sigma_t)_{t\geq 0} = (\sigma(\mathscr{N} \cup \mathscr{F}_t))_{t\geq 0}$ where $\mathscr{N}$ is the collection of negligible subsets of $\mathscr{F}$. It is also customary 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 completed natural filtration.

From now on and unless otherwise stated we make the "canonical assumption":
we assume that the underlying filtration is complete and right-continuous.

Remark 2.3.6. Subtleties about righ-continuity of filtrations.

The natural filtration of a right-continuous process is not right-continuous in general, indeed a counter example is given by $X_t = tZ$ for all $t \geq 0$ where Z is a non-constant random variable, since $\sigma(X_0) = \{\varnothing, \Omega\}$ while $\sigma(X_{0+\varepsilon} : \varepsilon > 0) = \sigma(Z) \neq \sigma(X_0)$. 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.

Theorem 2.3.7. Stopping times properties.

Let S, T, and T_n, $n \geq 0$ be stopping times for some underlying filtration $(\mathscr{F}_t)_{t\geq 0}$ on an underlyning probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Then:

1. the following family is a σ-algebra called the stopping σ-algebra:
$$\mathscr{F}_T = \{A \in \mathscr{F} : \forall t \geq 0, A \cap \{T \leq t\} \in \mathscr{F}_t\}.$$
Moreover the stopping time T is $\mathscr{F}_T$-measurable
2. $X = (X_t)_{t\geq 0}$ is adapted then the stopped process $X^T = (X_{t\wedge T})_{t\geq 0}$ is also adapted. Moreover
$$(X^T)^S = X^{S\wedge T} = (X^S)^T$$
3. if $(X_t)_{t\geq 0}$ is adapted and progressive and if T is a.s. finite then $X^T = (X_{t\wedge T})_{t\geq 0}$ is progressive
4. if $X = (X_t)_{t\geq 0}$ is adapted and right-continuous then $Z = X_T \mathbf{1}_{T<\infty}$ is $\mathscr{F}_T$-mesurable
5. if $S \leq T$ then $\mathscr{F}_S \subset \mathscr{F}_T$
6. $S \wedge T$ and $S \vee T$ are stopping times and in particular $\mathscr{F}_{S\wedge T} \subset \mathscr{F}_{S\vee T}$
7. if $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous then $\underline{\lim}_n T_n$ and $\overline{\lim}_n T_n$ are stopping times and
$$\cap_n \mathscr{F}_{T_n} = \mathscr{F}_{\inf_n T_n}.$$

Proof. The proof of the first three items are left as exercises.

4. Let $B \in \mathscr{B}_{\mathbb{R}}$ and $t \geq 0$. Then we have:

$$\{Z \in B\} \cap \{T \leq t\} = \{X_{T\wedge t} \in B\} \cap \{T \leq t\}.$$

Now we consider the composition of measurable maps:

$$\omega \in (\Omega, \mathscr{F}_t) \mapsto (\sigma(\omega) \wedge t, \omega) \in ([0,t] \times \Omega, \mathscr{B}_{[0,1]} \otimes \mathscr{F}_t) \mapsto X_{\sigma(\omega)\wedge t}(\omega) \in (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$$

and we use the fact that X is progressive.

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

6. For all $t \geq 0$ we have

$$\{S \wedge T > t\} = \{S > t\} \cap \{T > t\} \in \mathscr{F}_t \quad \text{and} \quad \{S \vee T \leq t\} = \{S \leq t\} \cap \{T \leq t\} \in \mathscr{F}_t.$$

7. It suffice to show that $\sup_n T_n$ and $\inf_n T_n$ are stopping times. But

$$\{\sup_n T_n \leq t\} = \cap_n \{T_n \leq t\} \in \mathscr{F}_t \quad \text{and} \quad \{\inf_n T_n < t\} = \cup_n \{T_n < t\} \in \mathscr{F}_t$$

and therefore

$$\{\inf_n T_n \leq t\} = \cap_{\varepsilon>0}\{\inf_n T_n < t+\varepsilon\} \in \mathscr{F}_{t+} = \mathscr{F}_t.$$

Let $A \in \cap_n \mathscr{F}_{T_n}$. Then

$$A \cap \{\inf_n T_n < t\} = \cup_n A \cap \{T_n < t\} \in \mathscr{F}_t.$$

Therefore

$$A \cap \{\inf_n T_n \leq t\} \in \mathscr{F}_{t+} = \mathscr{F}_t.$$

■

Remark 2.3.8. Truncation via cutoff stopping times for continuous processes.

Truncation is an important tool in probability theory, and allows for instance to prove the strong law of large numbers for i.i.d. integrable random variables by reduction to the case of more integrable random variables. This tool is also available for stochastic processes, and its version with cutoff stopping times has the advantage of keeping the martingale structure (Doob stopping, Theorem 2.5.1). Let $X = (X_t)_{t\geq 0}$ be adapted. For all n we introduce the "truncation" or "cutoff" stopping time

$$T_n = \inf\{t \geq 0 : |X_t| \geq n\},$$

which takes its values in $[0, +\infty]$. We have $T_n \leq T_{n+1}$ for all n. If X is continuous then almost surely[a].

$$T_n \underset{n\to\infty}{\nearrow} +\infty.$$

Still if X is additionally continuous then almost surely and for all $n \geq 1$ and all $t \geq 0$,

$$|X_{t\wedge T_n}| \leq n\mathbf{1}_{|X_0|\leq n} + |X_0|\mathbf{1}_{|X_0|>n}.$$

If $X_0 = 0$ then the process $|X^{T_n}|$ is bounded by n for all $n \geq 1$. This is useful in this course[b].

[a] Indeed, almost surely, either the trajectory of X is bounded then $T_n = +\infty$ for large enough n beyond a (random) threshold, or the trajectory of X is unbounded and then by definition of being continuous and unbounded we have $T_n \nearrow +\infty$ as $n \to \infty$. Without continuity X_t could take arbitrary large values near a finite time forcing $(T_n)_n$ to be bounded.

[b] Localization is efficient for continuous processes issued from the origin. If X is discontinuous and in particular if it is a discrete time process, then, due to a possible jump at time T_n, we could have $|X_{T_n}| > n$ even if $X_0 = 0$ and n is large.

2.4 Martingales, sub-martingales, super-martingales

We restrict for simplicity to continuous martingales/sub-martingales/super-martingales. But many of the results remain actually valid for right-continuous martingales/sub-martingales/super-martingales.

The notion of martingale implements the idea of updating with a conditionally independent ingredient.

Definition 2.4.1. Martingales, sub-martingales, super-martingales.

Let $X = (X_t)_{t\geq 0}$ be a real adapted and integrable process in the sense that for all $t \geq 0$, X_t is measurable for $\mathscr{F}_t$ and $X_t \in \mathrm{L}^1$. Then, when

- $\mathbb{E}(X_t \mid \mathscr{F}_s) \geq X_s$ for all $t \geq 0$ and all $s \in [0,t]$, we say that X is a sub-martingale,
- $\mathbb{E}(X_t \mid \mathscr{F}_s) = X_s$ for all $t \geq 0$ and all $s \in [0,t]$, we say that X is a martingale
- $\mathbb{E}(X_t \mid \mathscr{F}_s) \leq X_s$ for all $t \geq 0$ and all $s \in [0,t]$, we say that X is a super-martingale.

These three notions can be seen in a sense as a probabilistic counterpart of the notions of increasing sequence, constant sequence, and decreasing sequence in basic classical analysis.

- For a sub-martingale, $t \mapsto \mathbb{E}(X_t)$ grows and in particular $\mathbb{E}(X_t) \geq \mathbb{E}(X_0)$ for all $t \geq 0$
- For a martingale, $t \mapsto \mathbb{E}(X_t)$ is constant, namely $\mathbb{E}(X_t) = \mathbb{E}(X_0)$ for all $t \geq 0$. It is a conservation law
- For a super-martingale, $t \mapsto \mathbb{E}(X_t)$ decreases and in particular $\mathbb{E}(X_t) \leq \mathbb{E}(X_0)$ for all $t \geq 0$.

The set of martingales is the intersection of the set of sub-martingales and the set of super-martingales. A super-martingale or sub-martingale is a martingale if and only if its expectation is constant along time. Being a martingale for a given filtration is a property stable by linear combinations.

If M is a martingale and if $(t_n)_{n\geq 0}$ is a strictly increasing sequence of times then the sequence of random variables $(M_{t_n})_{n\geq 0}$ is a discrete time martingale. We will try to avoid using discrete time martingales, but we will sometimes discretize time, notably to handle stopping times, which is roughly the same. The theory of discrete time martingales is similar to the theory of continuous time martingales that we develop here and comes with very similar theorems. In this course, most stochastic processes are in continuous time, and when we say "continuous process/martingale/etc", we mean that the process has continuous sample paths.

Example 2.4.2. Martingales.

1. If $Y \in \mathrm{L}^1$ then the process $(X_t)_{t\geq 0}$ defined by $X_t = \mathbb{E}(Y \mid \mathscr{F}_t)$ for all $t \geq 0$ is a martingale with respect to $(\mathscr{F}_t)_{t\geq 0}$ known as the Doob martingale or a closed martingale. It is uniformly integrable. Corollary 4.4.5 provides a sort of converse (u.i. martingales are closed)
2. If $(X_t)_{t\geq 0}$ is a martingale and if $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and such that $\varphi(X_t) \in \mathrm{L}^1$ for all $t \geq 0$, then by the Jensen inequality for conditional expectation, $(Y_t)_{t\geq 0} = (\varphi(X_t))_{t\geq 0}$ is a sub-martingale for the same filtration. In particular $(|X_t|)_{t\geq 0}$, $(X_t^2)_{t\geq 0}$, and $(\mathrm{e}^{X_t})_{t\geq 0}$ are sub-martingales
3. If $(X_t)_{t\geq 0}$ is a sub-martingale and if $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and non-decreasing such that $\varphi(X_t) \in \mathrm{L}^1$ for all $t \geq 0$, then by the Jensen inequality for condition expectation, $(Y_t)_{t\geq 0} = (\varphi(X_t))_{t\geq 0}$ is a sub-martingale for the same filtration. In particular $(\mathrm{e}^{X_t})_{t\geq 0}$ is a sub-martingale
4. A martingale $X = (X_t)_{t\geq 0}$ is also a martingale for its natural filtration $(\sigma(X_s : s \in [0,t]))_{t\geq 0}$
5. 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 $N_t = \mathrm{card}\{n \geq 1 : E_n \in [0,t]\}$. 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 $0 = t_0 \leq \cdots \leq t_n$, the random variables $N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent of law $\mathrm{Poi}(\lambda(t_1 - t_0)), \ldots, \mathrm{Poi}(\lambda(t_n - t_{n-1}))$. We say that $(N_t)_{t\geq 0}$ is the simple Poisson process of intensity λ. Now for the (natural) filtration

$(\mathscr{F}_t)_{t\geq 0}$, $\mathscr{F}_t = \sigma(N_s : 0 \leq s \leq t)$, 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 $0 \leq s \leq t$,

$$\begin{aligned}\mathbb{E}(N_t - ct \mid \mathscr{F}_s) &= \mathbb{E}(N_t - N_s - c(t-s) + N_s - cs \mid \mathscr{F}_s)\\ &= \mathbb{E}(N_t - N_s) - c(t-s) + N_s - cs\\ &= (\lambda - c)(t-s) + N_s - cs.\end{aligned}$$

This process is not continuous, but has right-continuous and left limited trajectories (càdlàg[a]).

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

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

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

7. 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 the process $(B_t)_{t\geq 0}$ is a martingale for its natural filtration, indeed, for all $0 \leq s \leq t$,

$$\mathbb{E}(B_t \mid \mathscr{F}_s) = \mathbb{E}(B_t - B_s + B_s \mid \mathscr{F}_s) = \mathbb{E}(B_t - B_s) + B_s = B_s.$$

This process has continuous trajectories. Moreover and similarly, for all $c \in \mathbb{R}$, the process $(B_t^2 - ct)_{t\geq 0}$ is a sub-martingale if $c < 1$, a martingale if $c = 1$, and a super-martingale if $c > 1$. The key is to use the decomposition $B_t = (B_t - B_s)^2 + 2B_sB_t - B_s^2$. We can also study the process $\mathrm{e}^{B_t - ct}$ and seek for a condition on c to get a martingale, and we speak about an exponential martingale. For simplicity, most of the martingales encountered in this course are continuous.

[a] *Continu à droite avec limites à gauche.*

2.5 Doob stopping theorem and maximal inequalities

Stopped martingales are martingales, and the conservation law extends to stopping times:

Theorem 2.5.1. Doob[a] stopping theorem.

[a] Named after Joseph L. Doob (1910 – 2004), American mathematician.

If M is a continuous martingale and $T : \Omega \to [0, +\infty]$ is a stopping time then $M^T = (M_{t\wedge T})_{t\geq 0}$ is a (continuous) martingale, namely for all $t \geq 0$ and $s \in [0, t]$, we have

$$M_{t\wedge T} \in \mathrm{L}^1 \quad \text{and} \quad \mathbb{E}(M_{t\wedge T} \mid \mathscr{F}_s) = M_{s\wedge T}.$$

Moreover, if T is bounded, or if T is almost surely finite and $(M_{t\wedge T})_{t\geq 0}$ is u.i.[a], then

$$M_T \in \mathrm{L}^1 \quad \text{and} \quad \mathbb{E}(M_T) = \mathbb{E}(M_0).$$

[a] For instance dominated by an integrable random variable, or even bounded by a constant.

In practice, the best is to retain that $(M_{t\wedge T})_{t\geq 0}$ is a martingale. We have $\lim_{t\to\infty} M_{T\wedge t}\mathbf{1}_{T<\infty} = M_T\mathbf{1}_{T<\infty}$ a.s. When $T < \infty$ a.s. we could use what we know on M and T to deduce by monotone or dominated convergence that this holds in L^1, giving $\mathbb{E}(M_T) = \mathbb{E}(\lim_{t\to\infty} M_{t\wedge T}) = \lim_{t\to\infty}\mathbb{E}(M_{t\wedge T}) = \mathbb{E}(M_0)$. Theorem 2.5.1 states that this is automatically the case when T is bounded or when M^T is u.i. Furthermore, if M^T is u.i. then it can be shown that M_∞ exists, giving a sense to M_T even on $\{T = \infty\}$, and then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

Proof. Let assume first that T takes a finite number of values $t_1 < \cdots < t_n$. Let us show that $M_T \in \mathrm{L}^1$ and $\mathbb{E}(M_T) = \mathbb{E}(M_0)$. We have $M_T = \sum_{k=1}^n M_{t_k} \mathbf{1}_{T=t_k} \in \mathrm{L}^1$, and moreover, using $\{T \geq t_k\} = (\cup_{i=1}^{k-1}\{T = t_i\})^c \in \mathscr{F}_{t_{k-1}}$, and the martingale property $\mathbb{E}(M_{t_k} - M_{t_{k-1}} \mid \mathscr{F}_{t_{k-1}}) = 0$, for all k, we get

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) + \mathbb{E}\Big(\sum_{k=1}^n \mathbb{E}(M_{t_k} - M_{t_{k-1}} \mid \mathscr{F}_{t_{k-1}})\mathbf{1}_{T\geq t_k}\Big) = \mathbb{E}(M_0).$$

Suppose now that T takes an infinite number of values but is bounded by some constant C. For all $n \geq 0$, we approximate T by the piecewise constant random variable (discretization of $[0, C]$)[1]

$$T_n = C\mathbf{1}_{T=C} + \sum_{k=1}^n t_k \mathbf{1}_{t_{k-1}\leq T<t_k} \quad \text{where} \quad t_k = t_{n,k} = C\frac{k}{n}.$$

This is a stopping time since it takes discrete values and for all $m \geq 0$,

$$\{T_n = m\} = \begin{cases} \varnothing \in \mathscr{F}_0 & \text{if } m \not\in \{t_k : 1 \leq k \leq n\} \\ \{T = C\} \in \mathscr{F}_C & \text{if } m = C \\ \{T < t_{k-1}\}^c \cap \{T < t_k\} \in \mathscr{F}_{t_k} & \text{if } m = t_k,\ 1 \leq k \leq n \end{cases}$$

where we used the fact that $\{T = t\} = \{T \leq t\} \cap \{T < t\}^c = \{T \leq t\} \cap \cap_{r=1}^\infty \{T > t - 1/r\} \in \mathscr{F}_t$ for all $t \geq 0$.

Since T_n takes a finite number of values, the previous step gives $\mathbb{E}(M_{T_n}) = \mathbb{E}(M_0)$. On the other hand, almost surely, $T_n \to T$ as $n \to \infty$. Since M is continuous, it follows that almost surely $M_{T_n} \to M_T$ as $n \to \infty$. Let us show now that $(M_{T_n})_{n\geq 1}$ is uniformly integrable. Since for all $n \geq 0$, T_n takes its values in a finite set $t_1 < \cdots < t_{m_n} \leq C$, the martingale property[2] and the Jensen inequality give, for all $R > 0$,

$$\begin{aligned}
\mathbb{E}(|M_{T_n}|\mathbf{1}_{|M_{T_n}|\geq R}) &= \sum_k \mathbb{E}(|M_{t_k}|\mathbf{1}_{|M_{t_k}|\geq R, T_n=t_k}) \\
&= \sum_k \mathbb{E}(|\mathbb{E}(M_C \mid \mathscr{F}_{t_k})|\mathbf{1}_{|M_{t_k}|\geq R, T_n=t_k}) \\
&\leq \sum_k \mathbb{E}(\mathbb{E}(|M_C| \mid \mathscr{F}_{t_k})\mathbf{1}_{|M_{t_k}|\geq R, T_n=t_k}) \\
&= \sum_k \mathbb{E}(|M_C|\mathbf{1}_{|M_{t_k}|\geq R, T_n=t_k}) \\
&= \mathbb{E}(|M_C|\mathbf{1}_{|M_{T_n}|\geq R}).
\end{aligned}$$

Now M is continuous and thus locally bounded, and $M_C \in \mathrm{L}^1$, thus, by dominated convergence,

$$\sup_n \mathbb{E}(|M_{T_n}|\mathbf{1}_{|M_{T_n}|>R}) \leq \mathbb{E}(|M_C|\mathbf{1}_{\sup_{s\in[0,C]}|M_s|\geq R}) \underset{R\to\infty}{\longrightarrow} 0.$$

Therefore $(M_{T_n})_{n\geq 0}$ is uniformly integrable. As a consequence

$$\overset{\text{a.s.}}{\lim_{n\to\infty}} M_{T_n} = M_T \in \mathrm{L}^1 \quad \text{and} \quad \mathbb{E}(M_T) = \lim_{n\to\infty} \mathbb{E}(M_{T_n}) = \mathbb{E}(M_0).$$

Let us suppose now that T is an arbitrary stopping time. For all $0 \leq s \leq t$ and $A \in \mathscr{F}_s$, the random variable $S = s\mathbf{1}_A + t\mathbf{1}_{A^c}$ is a (finite) stopping time, and what precedes for the finite stopping time $t \wedge T \wedge S$ gives $M_{t\wedge T\wedge S} \in \mathrm{L}^1$ and $\mathbb{E}(M_{t\wedge T\wedge S}) = \mathbb{E}(M_0)$. Now, using the definition of S, we have

$$\mathbb{E}(M_0) = \mathbb{E}(M_{t\wedge T\wedge S}) = \mathbb{E}(\mathbf{1}_A M_{s\wedge T}) + \mathbb{E}(\mathbf{1}_{A^c} M_{t\wedge T}) = \mathbb{E}(\mathbf{1}_A(M_{s\wedge T} - M_{t\wedge T})) + \mathbb{E}(M_{t\wedge T}).$$

Since $\mathbb{E}(M_{t\wedge T}) = \mathbb{E}(M_0)$, we get $\mathbb{E}((M_{t\wedge T} - M_{s\wedge T})\mathbf{1}_A) = 0$, namely the martingale property for $(M_{t\wedge T})_{t\geq 0}$.

Finally, suppose that $T < \infty$ a.s. and $(M_{t\wedge T})_{t\geq 0}$ is u.i. The random variable M_T is well defined and $\lim_{t\to\infty} M_{t\wedge T} = M_T$ a.s. because M is continuous. Furthermore, since $(M_{t\wedge T})_{t\geq 0}$ is u.i., it follows that $M_T \in \mathrm{L}^1$ and $\lim_{t\to\infty} M_{t\wedge T} = M_T$ in L^1. In particular $\mathbb{E}(M_0) \underset{\forall t}{=} \mathbb{E}(M_{t\wedge T}) = \lim_{t\to\infty}\mathbb{E}(M_{t\wedge T}) = \mathbb{E}(M_T)$. ■

[1] By using dyadics, we could define T_n in such a way that $T_n \searrow T$, giving $M_{T_n} \to M_T$ pointwise when M is right-continuous.
[2] It also works for non-negative sub-martingales.

Example 2.5.2. Example of application of Doob stopping theorem.

Let $(M_t)_{t\geq 0}$ be a continuous martingale, $a<b$, and $T=\inf\{t\geq 0 : M_t\in\{a,b\}\}$ the hitting time of the boundary of $[a,b]$. Suppose that M_0 takes its values in $[a,b]$ and that T is almost surely finite[a]. Then on the one hand, we have the equation $\mathbb{P}(M_T=a)+\mathbb{P}(M_T=b)=1$. On the other hand, by definition of T the process $(M_{t\wedge T})_{t\geq 0}$ is bounded and thus u.i. and the Doob stopping theorem (Theorem 2.5.1) gives then $x=\mathbb{E}(M_0)=\mathbb{E}(M_T)=a\mathbb{P}(M_T=a)+b\mathbb{P}(M_T=b)$. It follows by combining the equations that

$$\mathbb{P}(M_T=a)=\frac{b-x}{b-a}\quad\text{and}\quad\mathbb{P}(M_T=b)=\frac{x-a}{b-a}$$

(note that $x\in[a,b]$). This holds in particular for Brownian motion started from $x\in[a,b]$, and by using an exponential martingale, it is then even possible to compute the Laplace transform of T.

[a] Holds for BM B with $B_0=0\in(a,b)$ since $\mathbb{P}(T=\infty)\leq\mathbb{P}(T>t)\leq\mathbb{P}(B_t\in(a,b))=\mathbb{P}(\sqrt{t}Z\in(a,b))\to 0$ as $t\to\infty$.

Coding in action 2.5.3. Gambler's ruin.

Physically Brownian motion and the simple symmetric random walk are the same, it is just a matter of scale. Fix $a\leq b$ in $\mathbb{Z}$. Write a code to plot on the same graphic multiple trajectories of such a random walk started from various values of $x\in[a,b]$ and stopped when it reaches a or b. Could you verify numerically the formulas of Example 2.5.2? And mathematically?

Remark 2.5.4. Counter example with an unbounded stopping time.

If M is a continuous martingale with $M_0=0$, then, for all $a>0$, $T_a=\inf\{t>0 : M_t=a\}$ is a stopping time, but it cannot be bounded since this would give $0=\mathbb{E}(M_0)=\mathbb{E}(M_{T_a})=a>0$, a contradiction!

The following variant of the Doob stopping is useful in many applications.

Theorem 2.5.5. Doob stopping theorem for sub-martingales.

If M is a continuous sub-martingale and S and T are bounded stopping times such that $S\leq T$, $M_S\in\mathrm{L}^1$, and $M_T\in\mathrm{L}^1$, then $\mathbb{E}(M_S)\leq\mathbb{E}(M_T)$.

Proof. We proceed as in the proof of Theorem 2.5.1, by assuming first that S and T take their values in the finite set $\{t_1,\ldots,t_n\}$ where $t_1<\cdots<t_n$. In this case M_T and M_S are in L^1 automatically. The inequality $S\leq T$ gives $\mathbf{1}_{S\geq t}\leq\mathbf{1}_{T\geq t}$ for all t. Using this fact and the sub-martingale property of M, we get

$$\mathbb{E}(M_S)=\mathbb{E}(M_0)+\mathbb{E}\Big(\sum_{k=1}^n\underbrace{\mathbb{E}(M_{t_k}-M_{t_{k-1}}\mid\mathcal{F}_{t_{k-1}})}_{\geq 0}\mathbf{1}_{S\geq t_k}\Big)\leq\mathbb{E}(M_0)+\mathbb{E}\Big(\sum_{k=1}^n\mathbb{E}(M_{t_k}-M_{t_{k-1}}\mid\mathcal{F}_{t_{k-1}})\mathbf{1}_{T\geq t_k}\Big)=\mathbb{E}(M_T).$$

More generally, when S and T are arbitrary bounded stopping times satisfying $S\leq T$, and at least when M is a non-negative sub-martingale, we can proceed by approximation as in the proof of Theorem 2.5.1. ■

This can be skipped at first reading.

Theorem 2.5.6: Doob stopping theorem for non-negative super-martingales

If M is a continuous non-negative super-martingale and S and T are stopping times such that $S\leq T$, then $M_S\in\mathrm{L}^1$ and $M_T\in\mathrm{L}^1$ and $\mathbb{E}(M_S)\geq\mathbb{E}(M_T\mid\mathcal{F}_S)$, in particular $\mathbb{E}(M_S)\geq\mathbb{E}(M_T)$.

When S and T are bounded we recover Theorem 2.5.6 in the special case where $M\leq 0$.

Proof. See for instance [31, Theorem 3.25 pages 64–65]. Note that since M is a non-negative super-martingale, it is automatically bounded in L^1 since $0\leq\mathbb{E}(M_t)\leq\mathbb{E}(M_0)$ for all $t\geq 0$. ■

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 independent and identically distributed random variables.

Theorem 2.5.7. Doob maximal inequalities.

1. If M is a continuous martingale or non-negative sub-martingale then for all $p \geq 1$, $t \geq 0$, $\lambda > 0$,

$$\mathbb{P}\Big(\max_{s\in[0,t]}|M_s| \geq \lambda\Big) \leq \frac{\mathbb{E}(|M_t|^p)}{\lambda^p}.$$

2. If M is a continuous martingale then for all $p > 1$ and $t \geq 0$,

$$\mathbb{E}\Big(\max_{s\in[0,t]}|M_s|^p\Big) \leq \Big(\frac{p}{p-1}\Big)^p\mathbb{E}(|M_t|^p) \quad \text{in other words} \quad \Big\|\max_{s\in[0,t]}|M_s|\Big\|_p \leq \frac{p}{p-1}\|M_t\|_p.$$

   In particular if $M_t \in \mathrm{L}^p$ then $M_t^* = \max_{s\in[0,t]}|M_s| \in \mathrm{L}^p$.

Note that $q = 1/(1-1/p) = p/(p-1)$ is the Hölder conjugate of p in the sense that $1/p + 1/q = 1$. The Doob inequality is often used with $p = 2$, for which $(p/(p-1))^p = 4$.

Proof. We can assume that the right hand side is finite ($M_t \in \mathrm{L}^p$), otherwise the inequalities are trivial.

1. If M is a martingale, then by the Jensen inequality for the convex function $u \in \mathbb{R} \mapsto |u|^p$, the process $|M|^p$ is a sub-martingale. Similarly, If M is a non-negative sub-martingale then, since $u \in [0,+\infty) \mapsto u^p$ is convex and non-decreasing it follows that $M^p = |M|^p$ is a sub-martingale. Therefore in all cases $(|M_s|^p)_{s\in[0,t]}$ is a sub-martingale. Let us define the bounded stopping time

$$T = t \wedge \inf\{s \geq 0 : |M_s| \geq \lambda\}.$$

   Since M is continuous we have $|M_T| \leq \max(|M_0|,\lambda)$ and thus $M_T \in \mathrm{L}^1$. The Doob stopping Theorem 2.5.5 for the sub-martingale $|M|^p$ and the bounded stopping times T and t that satisfy $T \leq t$ gives

$$\mathbb{E}(|M_T|^p) \leq \mathbb{E}(|M_t|^p).$$

   On the other hand the definition of T gives

$$|M_T|^p \geq \lambda^p\mathbf{1}_{\max_{s\in[0,t]}|M_s|\geq\lambda} + |M_t|^p\mathbf{1}_{\max_{s\in[0,t]}|M_s|<\lambda} \geq \lambda^p\mathbf{1}_{\max_{s\in[0,t]}|M_s|\geq\lambda}.$$

   It remains to combine these inequalities to get the desired result

2. We first reduce to the case where M satisfies $\max_{s\in[0,t]}|M_s| \in \mathrm{L}^p$. To do so, we introduce for all $n \geq 1$ the truncation or localization stopping time[3] $T_n = \inf\{s \geq 0 : |M_s| \geq n\}$. By the Doob stopping theorem (Theorem 2.5.1), the process $(M_{s\wedge T_n})_{s\in[0,t]}$ is a martingale. Moreover, since M is continuous, we have the domination $|M_{s\wedge T_n}| \leq |M_0| \wedge n$, and since $M_t \in \mathrm{L}^p$ gives $M_0 \in \mathrm{L}^p$, we obtain $\max_{s\in[0,t]}|M_{s\wedge T_n}| \in \mathrm{L}^p$. The desired inequality for the dominated martingale $(M_{s\wedge T_n})_{s\in[0,t]}$ would give

$$\mathbb{E}(\max_{s\in[0,t]}|M_{s\wedge T_n}|^p) \leq \left(\frac{p}{p-1}\right)^p\mathbb{E}(|M_{t\wedge T_n}|^p),$$

   and the desired result for $(M_s)_{s\in[0,t]}$ would then follow by monotone convergence theorem as $n \to \infty$ since then $|M_{s\wedge T_n}| \nearrow |M_s|$ for all $s \in [0,t]$. Thus this shows that we can assume without loss of generality that $\sup_{s\in[0,t]}|M_s| \in \mathrm{L}^p$. This is our first martingale localization argument!

   By using the proof of the first item with $p = 1$ and decomposing M_t as we did for M_T, we get

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

[3] Since we are only interested by the time interval $[0,t]$, we could take $\wedge t$ which makes the stopping time bounded.

for all $\lambda > 0$, and thus

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

Now the Fubini – Tonelli theorem gives

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

and similarly (here we need $p > 1$)

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

Combining all this gives

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

But since the Hölder inequality gives

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

we obtain

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

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

■

Example 2.5.8. A consequence of Doob maximal inequality.

Let $(M_t)_{t\geq0}$ be a continuous martingale bounded in L^p, $p > 1$, namely

$$C_p = \sup_{t\geq0}\mathbb{E}(|M_t|^p) < \infty.$$

It follows that M is u.i. But the Doob maximal inequality says more. Namely, by Theorem 2.5.7, for all $t \geq 0$, $\mathbb{E}(\max_{s\in[0,t]}|M_s|^p) \leq C_p$. The monotone convergence theorem gives then

$$\mathbb{E}(\sup_{t\geq0}|M_t|^p) \leq C_p < \infty.$$

Therefore, almost surely $\sup_{t\geq0}|M_t| < \infty$. In other words M has almost surely bounded trajectories. Beware however that the bound is random and may depend on the trajectory.

The following version of Doob maximal inequality is useful for some applications.

Theorem 2.5.9. Doob maximal inequality for super-martingales.

If M is a continuous super-martingale, then for all $t \geq 0$ and $\lambda > 0$, denoting $M^- = \max(0, -M)$,

$$\mathbb{P}\Big(\max_{s\in[0,t]}|M_s| \geq \lambda\Big) \leq \frac{\mathbb{E}(M_0) + 2\mathbb{E}(M_t^-)}{\lambda}.$$

In particular when M is non-negative then $\mathbb{E}(M^-) = 0$ and the upper bound becomes $\mathbb{E}(M_0)/\lambda$.

This can be skipped at first reading.

Proof. Let us define the bounded stopping time

$$T = t \wedge \inf\{s \in [0,t] : M_s \geq \lambda\}.$$

We have $M_T \in \mathrm{L}^1$ since $|M_T| \le \max(|M_0|,|M_t|,\lambda)$. By the Doob stopping Theorem 2.5.5 with the sub-martingale $-M$ and the bounded stopping times 0 and T that satisfy $M_0 \in \mathrm{L}^1$ and $M_T \in \mathrm{L}^1$, we get

$$\mathbb{E}(M_0) \ge \mathbb{E}(M_T) \ge \lambda \mathbb{P}(\max_{s\in[0,t]} M_s \ge \lambda) + \mathbb{E}(M_t \mathbf{1}_{\max_{s\in[0,t]} M_s<\lambda})$$

hence, recalling that $M^- = \max(-M, 0)$,

$$\lambda \mathbb{P}(\max_{s\in[0,t]} M_s \ge \lambda) \le \mathbb{E}(M_0) + \mathbb{E}(M_t^-).$$

This produces the desired inequality when M is non-negative. For the general case, we observe that the Jensen inequality for the non-decreasing convex function $u \in \mathbb{R} \mapsto \max(u, 0)$ and the sub-martingale $-M$ shows that M^- is a non-negative sub-martingale. Thus, by Theorem 2.5.1,

$$\lambda \mathbb{P}(\max_{s\in[0,t]} M_s^- \ge \lambda) \le \mathbb{E}(M_t^-).$$

Finally, putting both inequalities together gives

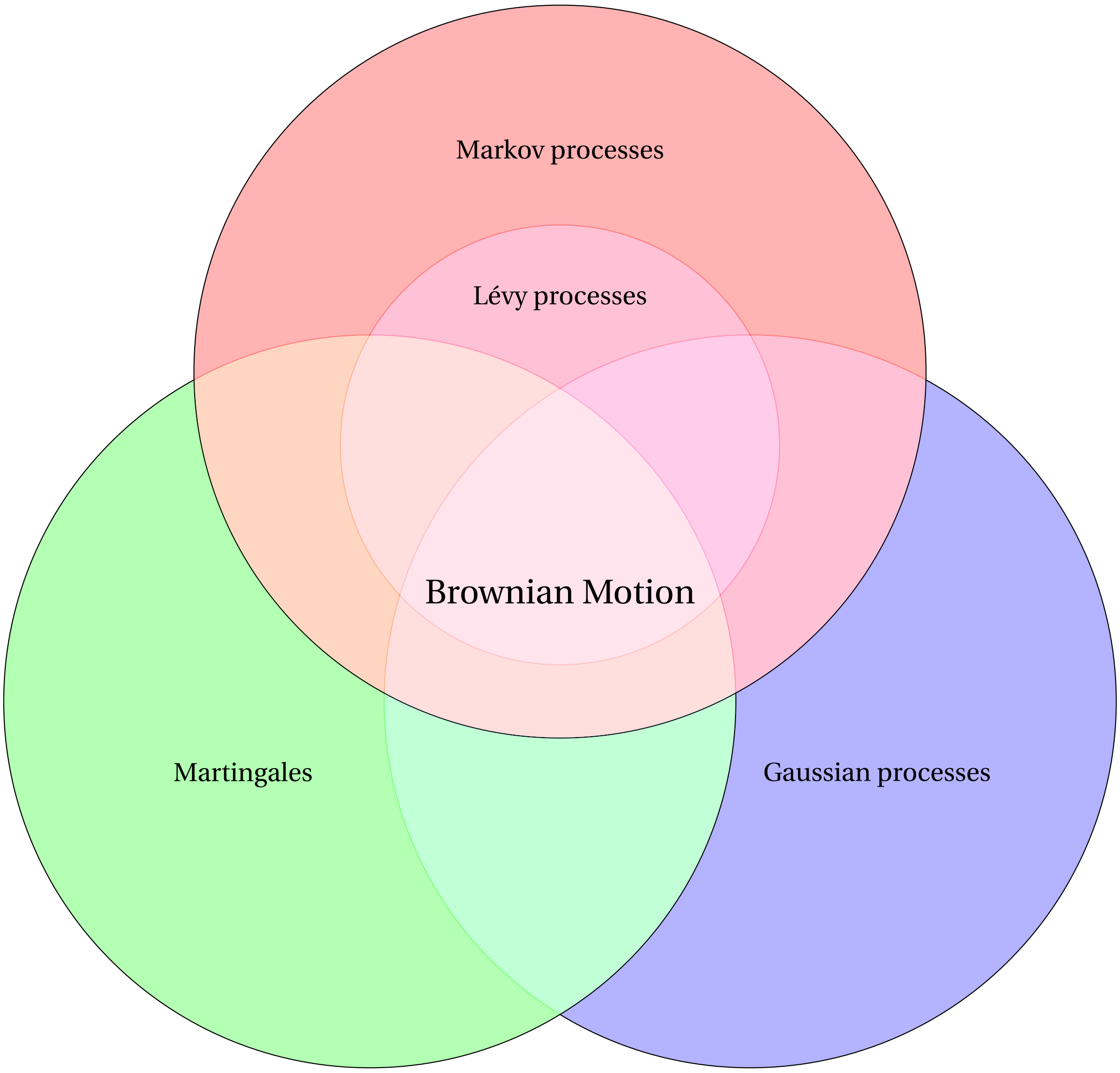
$$\lambda \mathbb{P}(\max_{s\in[0,t]} |M_s| \ge \lambda) \le \lambda \mathbb{P}(\max_{s\in[0,t]} M_s \ge \lambda) + \lambda \mathbb{P}(\max_{s\in[0,t]} M_s^- \ge \lambda) \le \mathbb{E}(M_0) + 2\mathbb{E}(M_t^-).$$

■

Chapter 3

Brownian motion

Just like the central limit theorem, Brownian motion is a physical as well as a mathematical phenomenon, see figures 3.1, 3.2, and 3.3. In this chapter, we study some properties of the mathematical Brownian motion.



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

$$x \in \mathbb{R}^d \mapsto p_t(x) = \frac{\mathrm{e}^{-\frac{|x|^2}{2t}}}{(\sqrt{2\pi t})^d} \quad \text{where} \quad |x|^2 = x_1^2 + \cdots + x_d^2.$$

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

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

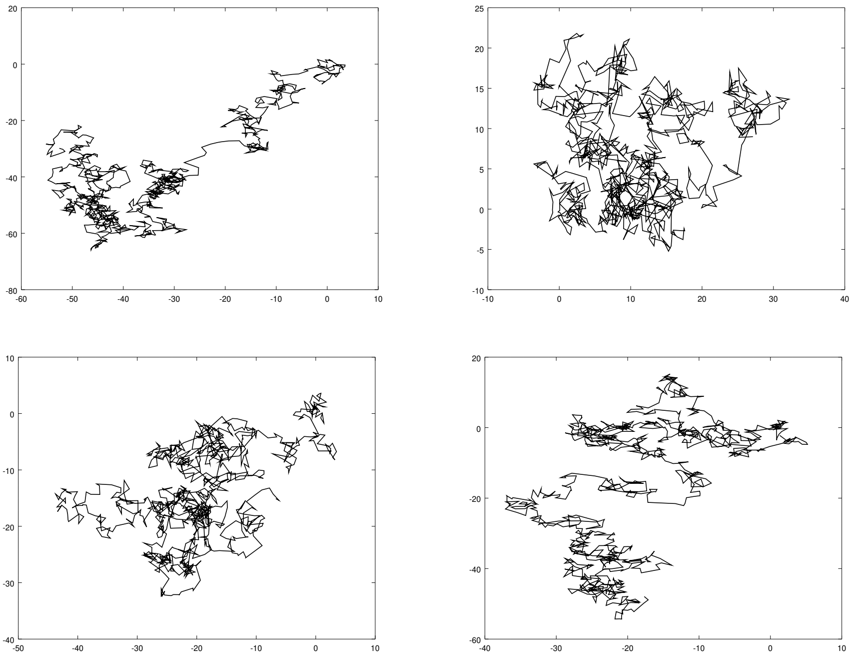


Figure 3.1: First steps of four approximated sample paths of 2-dimensional Brownian motion issued from the origin, numerically simulated with a Gaussian random walk via code `plot(cumsum(randn(2,1000)))`.

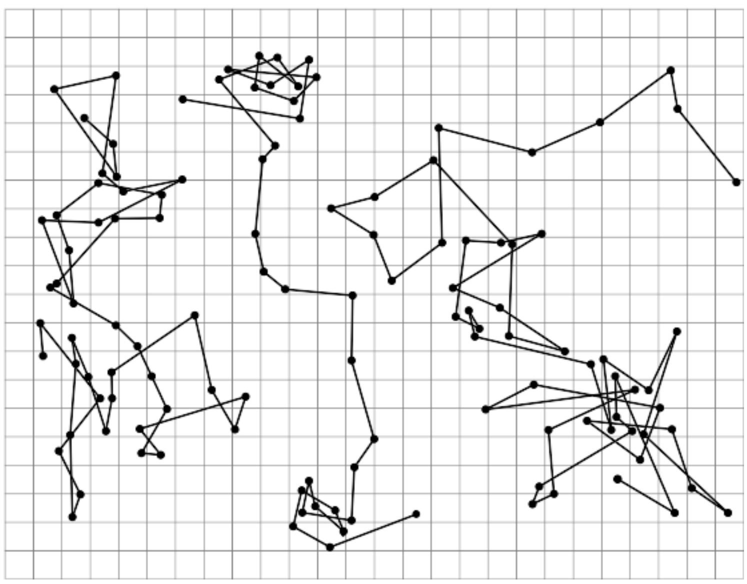


Figure 3.2: From the famous book [39] of Jean Perrin (1870 – 1942), three tracings of the motion of colloidal particles of radius 0.53 µm, as seen under the microscope are displayed. Successive positions every 30 seconds are joined by straight line segments (mesh size is 3.2 µm). These precise and systematic experiments, inspired by the historical ones by Robert Brown (1773 – 1858), allowed to test the atomistic theory of Ludwig Boltzmann (1944 – 1906), Albert Einstein (1879 – 1955), Marian Schmoluchovski (1872 – 1917), and others. “*Ainsi, la théorie moléculaire du mouvement brownien peut-être regardée comme expérimentalement établie, et, du même coup, il devient assez difficile de nier la réalité objective des molécules.*”. Louis Bachelier (1870 – 1946) identified independently a similar physical phenomenon in the behavior of stock markets.

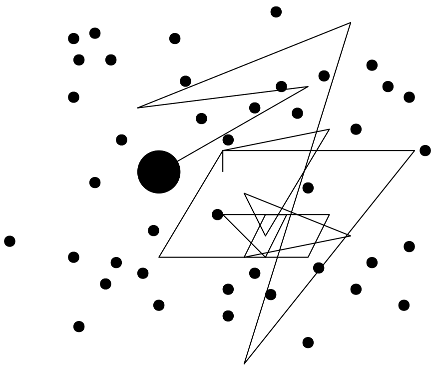


Figure 3.3: Atomistic interpretation of physical Brownian motion: a big particle of dust in a liquid is subject to a high number of collisions with the molecules of the liquid, which are much smaller and disordered by heat. This leads to the kinetic interpretation behind the Langevin equation in Example 8.2.7. In reality, the diameter ratio is high, for instance the colloidal particle observed by Perrin has diameter of 0.57 μm while a molecule of water has a diameter of 0.27 nm, which gives a diameter ratio of about 2000. Moreover in reality the molecules density is high, the distance between molecules being of 0.31 nm for water. With this atomistic interpretation, physical Brownian motion is essentially a random walk, seen at a space-time scale which makes it close to mathematical Brownian motion, its idealistic scaling limit.

Definition 3.0.1. Brownian motion[a] or Wiener[b] process.

[a] Named after Robert Brown (1773 – 1858), Scottish botanist.
[b] Named after Norbert Wiener (1894 – 1964), American mathematician.

A d-dimensional Brownian motion (BM) is a d-dimensional process $B = (B_t)_{t\geq 0}$ which has:

1. Almost surely continuous trajectories, in the sense that B is a continuous process.
2. Stationary, Gaussian, independent increments:
   - for all $0 \leq s \leq t$, $B_t - B_s \sim \mathcal{N}(0, (t-s)I_d)$
   - for all $t_0 = 0 < t_1 < \cdots < t_n$, $n \geq 0$, $B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent.

Beware that there are no conditions on B_0, and in particular $B_t = B_0 + B_t - B_0$ may not be Gaussian.

```python
# Python program generating the graphic used for the lecture notes cover
import numpy as np ; import matplotlib.pyplot as pp
for i in range(1,11):
    pp.plot(np.cumsum(np.random.randn(1,1000)[0]),'k-',linewidth=1)
pp.axis('off') ; pp.show()
```

```julia
# Julia program generating the graphic used for the lecture notes cover
using Pkg ; Pkg.add("Plots") ; using Plots
for i=1:10
plot!(cumsum(randn(1000,1),dims = 1), lw = 1, legend = false, grid = false,
 ↪  axes=([],false))
end
gui()
```

Coding in action 3.0.2. Stochastic simulation.

By using the structure of the increments, write your own program simulating and plotting approximated trajectories of BM. Can we check numerically that the mathematical object of Brownian motion exists? Let D be a closed domain of $\mathbb{R}^d$ such as a disc or a square, containing the origin 0. Let ∂D be its boundary, let B be a BM with $B_0 = 0$, and let $T = \inf\{t \geq 0 : B_t \in \partial D\}$. Write a program simulating the law of T and the law of B_T, and producing nice plots when $d = 1$, $d = 2$, $d = 3$.

Remark 3.0.3. Gaussian[a] and Lévy[b] processes.

[a]Named after Carl Friedrich Gauss (1777 – 1855), German mathematician.
[b]Named after Paul Lévy (1886 – 1971), French mathematician.

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. They form a special subclass of Markov processes. The Poisson process considered in Example 2.4.2 is also a Lévy process, for which the increments are Poisson and the trajectories right continuous with left limits.

Remark 3.0.4. Finite dimensional laws.

A d-dimensional continuous process $X = (X_t)_{t \geq 0}$ issued from $x \in \mathbb{R}^d$ is a Brownian Motion iff for all $n \geq 0$, all $0 < t_1 < \cdots < t_n$, all $A_i \in \mathcal{B}_{\mathbb{R}^d}$, $1 \leq i \leq n$, we have

$$\mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n) = \int_{A_1 \times \cdots \times A_n} p_{t_1}(x_1 - x) p_{t_2 - t_1}(x_2 - x_1) \cdots p_{t_n - t_{n-1}}(x_n - x_{n-1}) \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

Remark 3.0.5. Reduction to centered case.

From the definition, we get that if $B = (B_t)_{t \geq 0}$ is a Brownian motion issued from the origin namely $B_0 = 0$ and if H is a random variable then $(H + B_t)_{t \geq 0}$ is also a Brownian motion, issued from H.

Remark 3.0.6. Reduction to one-dimensional case.

From the definition, if $X = (X_t)_{t \geq 0}$ is d-dimensional with coordinates $X_t = (X_t^1, \ldots, X_t^d)$ in $\mathbb{R}^d$, then X is a Brownian motion issued from the origin iff the following two properties hold true:

1. for all $1 \leq i \leq d$, $(X_t^i)_{t \geq 0}$ is a Brownian motion issued from the origin
2. the processes $(X_t^1)_{t \geq 0}, \ldots, (X_t^d)_{t \geq 0}$ are independent.

3.1 Characterizations and martingales

Theorem 3.1.1. 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 $\mathbb{E}(X_t X_s) = s \wedge t$ 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 $0 \le s \le t$, we have

$$\mathbb{E}(X_s X_t) = \mathbb{E}(X_s(X_t - X_s)) + \mathbb{E}(X_s^2) = 0 + s = s = s \wedge t.$$

2. Conversely, if $X = (X_t)_{t\ge 0}$ is a Gaussian process, centered, with $\mathbb{E}(X_s X_t) = s \wedge t$ for all $s, t \ge 0$, then 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 $\mathrm{diag}(t_1, t_2 - t_1, \ldots, t_n - t_{n-1})$, which implies that $(X_t)_{t\ge 0}$ is a Brownian motion.

■

Corollary 3.1.2. Scale invariance by space-time scaling.

If $B = (B_t)_{t\ge 0}$ is a BM on $\mathbb{R}$, issued form the origin, then for all $c \in (0, +\infty)$, $\left(\frac{1}{\sqrt{c}} B_{ct}\right)_{t\ge 0}$ is a BM.

Proof. The process $\left(\frac{1}{\sqrt{c}} B_{ct}\right)_{t\ge 0}$ is continuous, Gaussian, centered, with same covariance as BM. ■

Theorem 3.1.3. Fourier and Laplace martingale characterizations of Brownian motion.

Let $X = (X_t)_{t\ge 0}$ be a d-dimensional continuous process issued from the origin.
The following properties are equivalent:

1. X is a Brownian motion
2. For all $\lambda \in \mathbb{R}^d$, $(M_t^\lambda)_{t\ge 0} = (\mathrm{e}^{\mathrm{i}\lambda\cdot X_t + \frac{|\lambda|^2 t}{2}})_{t\ge 0}$ is a martingale[a] for the natural filtration of X
3. For all $\lambda \in \mathbb{R}^d$, $(N_t^\lambda)_{t\ge 0} = (\mathrm{e}^{\lambda\cdot X_t - \frac{|\lambda|^2 t}{2}})_{t\ge 0}$ is a martingale for the natural filtration of X.

[a] The notion of martingale remains valid for complex valued processes.

Proof. Let us define $\mathcal{G}_t = \sigma(X_s : s \in [0, t])$ for all $t \ge 0$. The process X is a BM iff for all $0 \le s < t$, $X_t - X_s$ is independent of $\mathcal{G}_s$ and $X_t - X_s \sim \mathcal{N}(0, (t-s)I_d)$, in other words if and only if for all $0 \le s < t$ and $\lambda \in \mathbb{R}^d$,

$$\mathbb{E}(\mathrm{e}^{\mathrm{i}\lambda\cdot(X_t - X_s)} \mid \mathcal{G}_s) = \mathrm{e}^{-\frac{|\lambda|^2(t-s)}{2}}.$$

By multiplying both sides by $\mathrm{e}^{\mathrm{i}uZ}$ for an arbitrary bounded $\mathcal{G}_s$ measurable random variable Z and taking the expectation we get that $X_t - X_s$ is independent of $\mathcal{G}_s$ and $X_t - X_s \sim \mathcal{N}(0, (t-s)I_d)$. This shows the equivalence of the first two properties. The third property is just the Laplace (instead of Fourier) transform version. ■

Definition 3.1.4. Brownian motion with respect to a filtration.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$ be a filtered probability space. We say that a continuous d-dimensional process $X = (X_t)_{t\ge 0}$ is an $(\mathcal{F}_t)_{t\ge 0}$ Brownian motion when it is $(\mathcal{F}_t)_{t\ge 0}$ adapted and for all $t \ge 0$ and $s \in [0, t]$, the increment $X_t - X_s$ is independent of $\mathcal{F}_s$ and follows the Gaussian law $\mathcal{N}(0, (t-s)I_d)$, which is equivalent to say that for all $\lambda \in \mathbb{R}^d$, the process $(\exp(\mathrm{i}\lambda\cdot X_t + \frac{1}{2}|\lambda|^2 t))_{t\ge 0}$ is an $(\mathcal{F}_t)_{t\ge 0}$-martingale.

Remark 3.1.5. Definitions of Brownian motion (BM).

If $X = (X_t)_{t\ge 0}$ is an $(\mathcal{F}_t)_{t\ge 0}$ BM, then X is a BM in the sense of Definition 3.0.1. Conversely, a BM $(X_t)_{t\ge 0}$ in the sense of Definition 3.0.1 is an $(\mathcal{G}_t)_{t\ge 0}$ BM where $\mathcal{G}_t = \sigma(X_s : s \le t)$ for all $t \ge 0$ is the natural filtration associated to X (see Theorem 3.1.3).

Theorem 3.1.6. Martingale properties.

Let $B = (B_t)_{t\geq 0}$ be an $(\mathscr{F}_t)_{t\geq 0}$ d-dimensional Brownian motion and let $B_t = (B_t^1, \ldots, B_t^d)$ be the coordinates of the random vector B_t. Then for all $0 \leq s < t$ an $1 \leq j, k \leq d$,

$$\mathbb{E}(B_t^j - B_s^j \mid \mathscr{F}_s) = 0 \quad \text{and} \quad \mathbb{E}((B_t^j - B_s^j)(B_t^k - B_s^k) \mid \mathscr{F}_s) = (t-s)\mathbf{1}_{j=k}.$$

As a consequence, for all $1 \leq j, k \leq d$,

- $(B_t^j)_{t\geq 0}$ is a continuous $(\mathscr{F}_t)_{t\geq 0}$-martingale, provided that $B_0 \in \mathrm{L}^1$
- $(B_t^j B_t^k - \mathbf{1}_{j=k} t)_{t\geq 0}$ is a continuous $(\mathscr{F}_t)_{t\geq 0}$-martingale, provided that $B_0 \in \mathrm{L}^2$.

Actually it turns out that these properties characterize Brownian motion (see Theorem 7.2.1).

Proof. The first property follows from the fact that $(B_t^j)_{t\geq 0}$ is a BM. For the second property, we write

$$\begin{aligned}
\mathbb{E}((B_t^j - B_s^j)(B_t^k - B_s^k) \mid \mathscr{F}_s) &= \mathbb{E}((B_t^j - B_s^j)(B_t^k - B_s^k)) \\
&= \mathbb{E}((B_t^j - B_s^j))\mathbb{E}((B_t^k - B_s^k))\mathbf{1}_{j\neq k} + \mathbb{E}((B_t^j - B_s^k)^2)\mathbf{1}_{j=k} \\
&= 0 + (t-s)\mathbf{1}_{j=k}.
\end{aligned}$$

As a consequence, for all $0 \leq s \leq t$ and $1 \leq j, k \leq d$,

$$\mathbb{E}(B_t^j \mid \mathscr{F}_s) = B_s^j = \mathbb{E}(B_s^j \mid \mathscr{F}_s)$$

and

$$\mathbb{E}(B_t^j B_t^k - t\mathbf{1}_{j=k} \mid \mathscr{F}_s) = B_s^j B_s^k - s\mathbf{1}_{j=k} = \mathbb{E}(B_s^j B_s^k - s\mathbf{1}_{j=k} \mid \mathscr{F}_s).$$

■

Up to now, we study BM but it is unclear if BM exists or not! Actually an explicit construction of BM is given in Section 3.6. Other constructions are available, see for instance [28].

3.2 Variation of trajectories and quadratic variation

See Definition 1.7.1 (finite variation functions) and Definition 4.1.1 (quadratic variation of processes).

Theorem 3.2.1. Variation and quadratic variation of Brownian motion.

Let $B = (B_t)_{t\geq 0}$ be a BM issued from the origin, let $[u, v]$ be a finite interval, $0 \leq u < v$, and let δ be a partition or sub-division of $[u, v]$, $\delta : u = t_0 < \cdots < t_n = v$, $n \geq 1$. Let us consider the quantities

$$r_1(\delta) = \sum_{i=1}^{n-1} |B_{t_{i+1}} - B_{t_i}| \quad \text{and} \quad r_2(\delta) = \sum_{i=0}^{n-1} |B_{t_{i+1}} - B_{t_i}|^2.$$

Then the following properties hold true:

1. $\lim_{|\delta|\to 0} r_2(\delta) = v - u$ in L^2 and thus in $\mathbb{P}$, where $|\delta| = \sup_{0\leq i\leq n}(t_{i+1} - t_i)$. In other words, the quadratic variation of B on a finite interval is equal to the length of the interval.

2. $\sup_{\delta\in\mathscr{P}} r_1(\delta) = +\infty$ almost surely, where $\mathscr{P}$ is the set of subdivision of $[u, v]$. In other words the sample paths of B are almost surely of infinite variation on all intervals.

The second proprery implies that we cannot hope to define an integral $\int_a^b \varphi_t \mathrm{d}B_t(\omega)$ with φ continuous as in Theorem 1.7.2 because $t \mapsto B_t(\omega)$ is of infinite variation on all intervals for almost all ω. However, and following Itô, the first property will be the key to give a sort of L^2 or in $\mathbb{P}$ meaning to such stochastic integrals.

The quadratic variation of square integrable continuous martingales is considered in Theorem 4.1.4.

Proof. We could use Lemma 4.1.2 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

$$\begin{aligned}
\mathbb{E}((r_2(\delta))^2) &= \mathbb{E}\Big(\big(\sum_i |B_{t_{i+1}} - B_{t_i}|^2\big)^2\Big)\\
&= \sum_i \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|^4) + 2\sum_{i<j} \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|^2 |B_{t_{j+1}} - B_{t_j}|^2)\\
&= 3\sum_i (t_{i+1} - t_i)^2 + 2\sum_{i<j} (t_{i+1} - t_i)(t_{j+1} - t_j)\\
&= 2\sum_i (t_{i+1} - t_i)^2 + \big(\sum_i (t_{i+1} - t_i)\big)^2\\
&= 2\sum_i (t_{i+1} - t_i)^2 + (v-u)^2.
\end{aligned}$$

Moreover $\mathbb{E}(r_2(\delta)) = \sum_i (t_{i+1} - t_i) = v - u$. Thus

$$\mathbb{E}((r_2(\delta) - (v-u))^2) = 2\sum_i (t_{i+1} - t_i)^2 \le 2\max_i (t_{i+1} - t_i)(v-u) \underset{|\delta|\to 0}{\longrightarrow} 0.$$

2. From the first part, there exists a sequence of subdivisions $(\delta^k)_k$ of $[u,v]$ such that

$$\lim_{k\to\infty} r_2(\delta^k) = \lim_{k\to\infty} \sum_i |B_{t^k_{i+1}} - B_{t^k_i}|^2 = v - u \quad \text{almost surely}$$

and thus, almost surely,

$$\sup_\delta r_1(\delta) \ge r_1(\delta^k) = \sum_i |B_{t^k_{i+1}} - B_{t^k_i}| \ge \frac{\sum_i |B_{t^k_{i+1}} - B_{t^k_i}|^2}{\max_i |B_{t^k_{i+1}} - B_{t^k_i}|} \underset{k\to\infty}{\longrightarrow} +\infty,$$

where used the fact that almost surely, $\max_i |B_{t^k_{i+1}} - B_{t^k_i}| \to 0$ as $k \to \infty$ since $B_\bullet$ is continuous and hence uniformly continuous on every compact interval such as $[u,v]$ (Heine theorem).

■

3.3 Blumenthal zero-one law and its consequences on the trajectories

This can be skipped at first reading.

Theorem 3.3.1: Properties of Brownian trajectories

If $B = (B_t)_{t\ge 0}$ is a one-dimensional BM on $\mathbb{R}$ issued form the origin, and $\mathscr{F}_t = \sigma(B_s : s \in [0,t])$, then:

1. Blumenthal[a] 0-1 law. The σ-algebra $\mathscr{F}_{0^+} = \cap_{t>0} \mathscr{F}_t$ is trivial: for all $A \in \mathscr{F}_{0^+}$, $\mathbb{P}(A) \in \{0,1\}$
2. Almost surely, for all $\varepsilon > 0$, $\inf_{s\in[0,\varepsilon]} B_s < 0$ and $\sup_{s\in[0,\varepsilon]} B_s > 0$
3. For all $a \in \mathbb{R}$, almost surely[b], $T_a = \inf\{t \ge 0 : B_t = a\} < \infty$
4. Almost surely[c], $\varliminf_{t\to\infty} B_t = -\infty$ and $\varlimsup_{t\to\infty} B_t = +\infty$
5. Almost surely, the function $t \in \mathbb{R}_+ \to B_t$ is not monotone on any non singleton interval.

[a] Named after Robert McCallum Blumenthal (1931 – 2012), American mathematician.
[b] However T_a is not bounded, see Remark 2.5.4.
[c] This does not imply that a.s. $\lim_{t\to\infty} |B_t| = +\infty$.

Proof.

1. The idea is to show that $\mathscr{F}_{0^+}$ is independent of itself. For all $A \in \mathscr{F}_{0^+}$, all $k \geq 1$, all bounded continuous $f : \mathbb{R}^k \to \mathbb{R}$, and all $0 < t_1 < \cdots < t_k$, we have

$$\mathbb{E}(\mathbf{1}_A f(B_{t_1}, \ldots, B_{t_k})) = \lim_{\varepsilon \to 0^+} \mathbb{E}(\mathbf{1}_A f(B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon)).$$

Now when $0 < \varepsilon < t_1$, the random variables $B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon$ are independent of $\mathscr{F}_\varepsilon$ (structure of the increments of simple Markov property), and thus independent of $\mathscr{F}_{0^+}$. It follows that

$$\mathbb{E}(\mathbf{1}_A f(B_{t_1}, \ldots, B_{t_k})) = \lim_{\varepsilon \to 0^+} \mathbb{P}(A)\mathbb{E}(f(B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon)) = \mathbb{P}(A)\mathbb{E}(f(B_{t_1}, \ldots, B_{t_k})).$$

Hence $\mathscr{F}_{0^+}$ is independent of $\sigma(B_{t_1}, \ldots, B_{t_k})$ for all t_i's, and thus is independent of $\sigma(B_t, t > 0)$. But $\sigma(B_t, t > 0) = \sigma(B_t, t \geq 0)$ since $B_0 = 0$. It remains to note that $\mathscr{F}_{0^+} \subset \sigma(B_t, t \geq 0)$.

2. For the statement with the sup, it suffices to show that $\mathbb{P}(A) = 1$ where

$$A = \bigcap_n \Big\{ \sup_{s \in [0,1/n]} B_s > 0 \Big\}.$$

We can restrict the intersection to $n \geq N$ for an arbitrary large threshold N, therefore $A \in \mathscr{F}_{0^+}$. Next, thanks to the Blumenthal zero-one law, it suffices to show that $\mathbb{P}(A) > 0$. Now

$$\mathbb{P}\Big(\sup_{s \in [0,1/n]} B_s > 0 \Big) \underset{n \to \infty}{\searrow} \mathbb{P}(A)$$

while

$$\mathbb{P}\Big(\sup_{s \in [0,1/n]} B_s > 0 \Big) \geq \mathbb{P}(B_{1/n} > 0) = \frac{1}{2},$$

giving $\mathbb{P}(A) \geq 1/2$ and thus $\mathbb{P}(A) = 1$. The statement with inf follows by using $-B$ instead of B.

3. Thanks to the previous property,

$$\mathbb{P}\Big(\sup_{s \in [0,1]} B_s > \varepsilon \Big) \underset{\varepsilon \to 0}{\nearrow} \mathbb{P}\Big(\sup_{s \in [0,1]} B_s > 0 \Big) = 1.$$

But by the scale invariance (Corollary 3.1.2),

$$\mathbb{P}\Big(\sup_{s \in [0,1]} B_s > \varepsilon \Big) = \mathbb{P}\Big(\sup_{s \in [0,\varepsilon^{-2}]} \varepsilon^{-1} B_{\varepsilon^2 s} > 1 \Big) = \mathbb{P}\Big(\sup_{s \in [0,\varepsilon^{-2}]} B_s > 1 \Big)$$

Now, since

$$\mathbb{P}\Big(\sup_{s \in [0,\varepsilon^{-2}]} B_s > 1 \Big) \underset{\varepsilon \to 0}{\nearrow} \mathbb{P}\Big(\sup_{s \geq 0} B_s > 1 \Big),$$

we get

$$\mathbb{P}\Big(\sup_{s \geq 0} B_s > 1 \Big) = 1.$$

Again by scaling, we obtain, for all $R > 0$, and also by replacing B by $-B$,

$$\mathbb{P}\Big(\sup_{s \geq 0} B_s > R \Big) = 1 \quad \text{and} \quad \mathbb{P}\Big(\inf_{s \geq 0} B_s < -R \Big) = 1.$$

This implies that for all $a \in \mathbb{R}$, almost surely $T_a < \infty$.

4. This is implied directly by the end of the proof of the previous item.

5. From the item about the inf and sup, and the structure of increments, we have, almost surely, for all $t \in \mathbb{Q} \cap \mathbb{R}_+$ and all $\varepsilon > 0$, $\inf_{s \in [t,t+\varepsilon]} B_s < B_t$ and $\sup_{s \in [t,t+\varepsilon]} B_s > B_t$, hence the result.

■

Corollary 3.3.2: Law of hitting time via Laplace transform

Let $(B_t)_{t\geq 0}$ be a one-dimensional Brownian motion with $B_0 = 0$. For all $a > 0$, let us consider the hitting time $T_a = \inf\{t \geq 0 : B_t = a\}$, which is almost surely finite thanks to Theorem 3.3.1. Then its Laplace transform is given by $\lambda \geq 0 \mapsto \mathbb{E}(\mathrm{e}^{-\lambda T_a}) = \mathrm{e}^{-a\sqrt{2\lambda}}$, and it has density

$$t \in \mathbb{R}_+ \mapsto \frac{a}{\sqrt{2\pi t}^3}\mathrm{e}^{-\frac{a^2}{2t}}.$$

Proof. For all $c > 0$ and n, the Doob stopping (Theorem 2.5.1) with the martingale $(\mathrm{e}^{cB_t - \frac{c^2}{2}t})_{t\geq 0}$ and the bounded stopping time $T_a \wedge n$ gives $\mathbb{E}(\mathrm{e}^{cB_{T_a\wedge n} - \frac{c^2}{2}(T_a\wedge n)}) = 1$. Now, since $\mathrm{e}^{cB_{T_a\wedge n} - \frac{c^2}{2}(T_a\wedge n)} \leq \mathrm{e}^{ca}$, we get, by dominated convergence, $\mathbb{E}(\mathrm{e}^{cB_{T_a} - \frac{c^2}{2}T_a}) = 1$. Next, since B has almost surely continuous trajectories, we have $B_{T_a} = a$ almost surely, and this gives the formula for the Laplace transform. The formula for the density follows then by the inversion formula for the Laplace transform. ■

3.4 Strong law of large numbers, invariance by time inversion, law of iterated logarithm

The nature of the increments of Brownian motion leads to formulate the following theorem.

Theorem 3.4.1. Strong law of large numbers.

If $(B_t)_{t\geq 0}$ is BM on $\mathbb{R}$ with $B_0 = 0$ then $\lim_{t\to\infty} \frac{B_t}{t} = 0$ almost surely and in L^p for all $p \in [1,\infty)$.

The central limit theorem would be the trivial statement $\sqrt{t}\frac{B_t}{t} \xrightarrow[t\to\infty]{\text{law}} \mathcal{N}(0,1)$.
The a.s. remains valid for an arbitrary B_0, and the L^p convergence if $B_0 \in \mathrm{L}^p$.

Proof. Since for all $t > 0$ and all $p > 0$, $\mathbb{E}\left(\left|\frac{B_t}{t}\right|^p\right) = \frac{\mathbb{E}(|B_1|^p)}{t^{p/2}}$ and $B_1 \sim \mathcal{N}(0,1)$, we have immediately

$$\frac{B_t}{t} \xrightarrow[t\to\infty]{\mathrm{L}^p} 0 \quad \text{and in particular} \quad \frac{B_t}{t} \xrightarrow[t\to\infty]{\mathbb{P}} 0.$$

To get the almost sure convergence, we need some tightness, a control of tails that can be done via moments. Let us prove the a.s. convergence. Let a and b be real numbers such that $0 < a < b$. We have

$$\mathbb{E}\left(\sup_{a\leq t\leq b}\left(\frac{B_t}{t}\right)^2\right) \leq \frac{1}{a^2}\mathbb{E}\left(\sup_{a\leq t\leq b} B_t^2\right).$$

The Doob maximal inequality of Theorem 2.5.7 applied to the martingale $(B_{a+t})_{t>0}$ on $[0,b-a]$ yields

$$\mathbb{E}\left(\sup_{a\leq t\leq b}\left(\frac{B_t}{t}\right)^2\right) \leq \frac{1}{a^2}4\mathbb{E}(B_b^2) = \frac{4b}{a^2}.$$

Applying this to $a = 2^n$ and $b = 2^{n+1}$ we obtain

$$\mathbb{E}\left(\sup_{2^n\leq t\leq 2^{n+1}}\left(\frac{B_t}{t}\right)^2\right) \leq \frac{8}{2^n}$$

Thus, by the Markov inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{2^n\leq t\leq 2^{n+1}}\left|\frac{B_t}{t}\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left(\sup_{2^n\leq t\leq 2^{n+1}}\left(\frac{B_t}{t}\right)^2\right) \leq \frac{8}{2^n\varepsilon^2},$$

which gives

$$\sum_{n=0}^{\infty}\mathbb{P}\left(\sup_{2^n\leq t\leq 2^{n+1}}\left|\frac{B_t}{t}\right| > \varepsilon\right) < \infty.$$

Now, according to the Borel–Cantelli lemma, there exists an almost sure event A_ε such that for all $\omega \in A_\varepsilon$, there exists a threshold n_ω such that for all $n \geq n_\omega$, $\sup_{2^n\leq t\leq 2^{n+1}}\left|\frac{B_t(\omega)}{t}\right| \leq \varepsilon$. Thus, for all $\varepsilon > 0$, there exists an a.s. event A_ε such that for all $\omega \in A_\varepsilon$, there exists t_ω, such that for all $t \geq t_\omega$,

$$\left|\frac{B_t(\omega)}{t}\right| \leq \varepsilon.$$

It remains to consider the almost sure event $A = \bigcap_{r=1}^{\infty} A_{1/r}$, on which $\lim_{t\to\infty}\frac{B_t}{t} = 0$. ■

Corollary 3.4.2. Invariance by time inversion.

If $B = (B_t)_{t\geq 0}$ is a BM on $\mathbb{R}$ with $B_0 = 0$ then $X = (tB_{1/t})_{t\geq 0}$ with the convention $X_0 = 0$ is also BM.

Proof. The process X is Gaussian, centered, with $\mathbb{E}(X_sX_t) = s \wedge t$ for all $s,t \geq 0$. It remains to prove that X is continuous. By definition X is continuous on $(0,\infty)$. It remains to prove the almost sure continuity at $t = 0$. This follows from Theorem 3.4.1, namely, almost surely, $\lim_{t\to 0^+} X_t = \lim_{t\to 0^+} tB_{1/t} = \lim_{t\to+\infty}\frac{B_t}{t} = 0$. ■

Theorem 3.4.3. Law of Iterated Logarithm.

If $(B_t)_{t\geq 0}$ is a Brownian motion on $\mathbb{R}$ then

$$\mathbb{P}\Big(\varliminf_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}=-1,\quad \varlimsup_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}=1\Big)=1$$

and

$$\mathbb{P}\Big(\varliminf_{t\to\infty}\frac{B_t}{\sqrt{2t\log(\log(t))}}=-1,\quad \varlimsup_{t\to\infty}\frac{B_t}{\sqrt{2t\log(\log(t))}}=1\Big)=1.$$

This can be skipped at first reading.

Proof. The second property follows from the first one by using invariance by time inversion (Corollary 3.4.2). Let us prove the first property. We can assume without loss of generality that $B_0=0$. Since the intersection of two almost sure events is almost sure, and since the law of B is symmetric in the sense that $-B$ and B have same law, it follows that it suffices to show that

$$\mathbb{P}\Big(\varlimsup_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}=1\Big)=1.$$

Let us define $h(t)=\sqrt{2t\log(\log(1/t))}$ for $t>1/\mathrm{e}$. For all $\alpha>0$ and $\beta>0$, the Doob maximal inequality of Theorem 2.5.7 used for the “exponential” martingale $(\mathrm{e}^{\alpha B_t-\frac{\alpha^2}{2}t})_{t\geq 0}$ gives, for all $t\geq 0$,

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

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

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

By the Borel–Cantelli lemma, we get that for almost all $\omega\in\Omega$, there exists n_ω such that for all $n\geq n_\omega$,

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

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

$$\mathbb{P}\Big(\varlimsup_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}\leq\frac{2+\delta}{2\sqrt{\theta}}\Big)=1.$$

Now we let $\theta\to 1$ and $\delta\to 0$, reducing the proof to show that

$$\mathbb{P}\Big(\varlimsup_{t\searrow 0}\frac{B_t}{\sqrt{2t\log(\log(1/t))}}\geq 1\Big)=1.$$

For that, for all $n\geq 1$ and $\theta\in(0,1)$, we define the event

$$A_n=\{B_{\theta^n}-B_{\theta^{n+1}}\geq(1-\sqrt{\theta})h(\theta^n)\}.$$

We have, denoting $a_n=(1-\sqrt{\theta})h(\theta^n)/(\theta^{n/2}\sqrt{1-\theta})$,

$$\mathbb{P}(A_n)=\frac{1}{\sqrt{2\pi}}\int_{a_n}^\infty \mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u\geq\frac{a_n}{1+a_n^2}\mathrm{e}^{-\frac{a_n^2}{2}}\geq\frac{1}{2a_n}\Big(n\log\frac{1}{\theta}\Big)^{-\frac{(1-\sqrt{\theta})^2}{1-\theta}}.$$

Thus $\sum_{n\geq 1}\mathbb{P}(A_n) = +\infty$. 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

$$B_{\theta^n} - B_{\theta^{n+1}} \geq (1-\sqrt{\theta})h(\theta^n).$$

Since B and $-B$ have same law, the first part of the proof gives that almost surely, for n large enough,

$$B_{\theta^{n+1}} > -2h(\theta^{n+1}) \geq -2\sqrt{\theta}h(\theta^n).$$

Therefore, almost surely, for an infinite number of values of n, we have

$$B_{\theta^n} > h(\theta^n)(1-3\sqrt{\theta}).$$

This gives

$$\mathbb{P}\Big(\overline{\lim_{t\searrow 0}}\frac{B_t}{\sqrt{2t\log(\log(1/t))}} \geq 1-3\sqrt{\theta}\Big) = 1.$$

It remains to send θ to 0. Note that this proof uses both sides of the Borel–Cantelli lemma. ■

Corollary 3.4.4. 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

$$\mathbb{P}\Big(\underline{\lim_{t\searrow 0}}\frac{B_{t+s}-B_s}{\sqrt{2t\log(\log(1/t))}} = -1, \quad \overline{\lim_{t\searrow 0}}\frac{B_{t+s}-B_s}{\sqrt{2t\log(\log(1/t))}} = 1\Big) = 1.$$

In particular almost surely the sample paths $t \in \mathbb{R}_+ \mapsto B_t$ of B are not $\frac{1}{2}$-Hölder[a] continuous on finite intervals and in particular are nowhere differentiable on $\mathbb{R}_+$.

[a] $f : I \to \mathbb{R}$ is γ-Hölder continuous when $(\forall \varepsilon > 0)(\exists \eta > 0)(\forall s, t \in I)(|s-t|^\gamma \leq \eta \Rightarrow |f(s)-f(t)| \leq \varepsilon)$.

Proof. Follows from Theorem 3.4.3 and the fact that $(B_{t+s}-B_s)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ have same law. ■

3.5 Strong Markov property, reflection principle, hitting time

If $(B_t)_{t\geq 0}$ is BM then we easily check that for all fixed $T > 0$, the process $(B_{t+T}-B_T)_{t\geq 0}$ is a BM, issued form the origin, independent of $\mathcal{F}_T$. This is the simple Markov property. It extends to stopping times T:

Theorem 3.5.1. Strong Markov[a] property.

[a] Named after Andrey Markov (1856 – 1922), Russian mathematician.

If $B = (B_t)_{t\geq 0}$ is a d-dimensional Brownian motion issued from the origin, then for all stopping time T such that $\mathbb{P}(T < \infty) > 0$, under the probability measure $\mathbb{P}(\cdot \mid T < \infty)$, the following properties hold:

1. $((B_{t+T}-B_T)\mathbf{1}_{\{T<\infty\}})_{t\geq 0}$ 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 $t > 0$,

$$\mathbb{E}(f(B_{t+T})\mathbf{1}_{\{T<\infty\}} \mid \mathcal{F}_T) = P_t(f)(B_T)\mathbf{1}_{\{T<\infty\}}$$

where

$$P_t(f)(x) = \mathbb{E}(f(x+B_t)) = \frac{1}{(\sqrt{2\pi t})^d}\int_{\mathbb{R}^d} \mathrm{e}^{-\frac{|x-y|^2}{2t}} f(y)\mathrm{d}y = (p_t * f)(x).$$

We say then that Brownian motion is a strong Markov process.

Proof. Suppose first that $\mathbb{P}(T<\infty)=1$. Let us define $B^*=(B_{T+t}-B_T)_{t\geq 0}$. For all $n\geq 1$, let us define

$$T_n=\sum_{k\geq 0}\frac{k+1}{2^n}\mathbf{1}_{T\in[\frac{k}{2^n},\frac{k+1}{2^n})}.$$

We have that $T\leq T_n$, 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 $T_n\searrow T$ as $n\to\infty$. Let $A\in\mathcal{F}_T$, $m\geq 0$, and $0=t_0<\cdots<t_m<\infty$. By the dominated convergence theorem, we have, for all continuous and bounded $\varphi:(\mathbb{R}^d)^m\to\mathbb{R}$,

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A\varphi(B^*_{t_1},\ldots,B^*_{t_m}))&=\mathbb{E}(\mathbf{1}_A\varphi(B_{t_1+T}-B_T,\ldots,B_{t_m+T}-B_T))\\&=\lim_{n\to\infty}\mathbb{E}(\mathbf{1}_A\varphi(B_{t_1+T_n}-B_{T_n},\ldots,B_{t_m+T_n}-B_{T_n})).\end{aligned}$$

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 $A\in\mathcal{F}_{T_n}$,

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A\varphi(B_{t_1+T_n}-B_{T_n},\ldots,B_{t_m+T_n}))&=\sum_{r\in D_n}\mathbb{E}(\mathbf{1}_{A\cap\{T_n=r\}}\varphi(B_{t_1+r}-B_r,\ldots,B_{t_m+r}-B_r))\\&=\sum_{r\in D_n}\mathbb{E}(\mathbf{1}_{A\cap\{T_n=r\}}\mathbb{E}(\varphi(B_{t_1+r}-B_r,\ldots,B_{t_m+r}-B_r)\mid\mathcal{F}_r))\\&=\sum_{r\in D_n}\mathbb{P}(A\cap\{T_n=r\})\mathbb{E}(\varphi(B_{t_1+r}-B_r,\ldots,B_{t_m+r}-B_r))\\&=\mathbb{P}(A)\mathbb{E}(\varphi(B_{t_1}-B_0,\ldots,B_{t_m}-B_0)).\end{aligned}$$

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 $\mathbf{1}_A$ 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 1.5.2,

$$\mathbb{E}(f(B_{t+T})\mid\mathcal{F}_T)=\mathbb{E}(f(B^*_t+B_T)\mid\mathcal{F}_T)=g_t(B_T)$$

where

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

Finally, for a T taking values in $[0,+\infty]$, the same argument works with A replaced by $A\cap\{T<\infty\}$. ■

This can be skipped at first reading.

Corollary 3.5.2: Reflection principle

Let B be a one-dimensional Brownian motion issued from the origin. For all $t\geq 0$, let us define $S_t=\sup_{s\in[0,t]}B_s$. Then, for all $t\geq 0$, the following properties hold:

- For all $a\geq 0$ and all $b\in(-\infty,a]$, $\mathbb{P}(S_t\geq a,B_t\leq b)=\mathbb{P}(B_t\geq 2a-b)$.
- The random variables S_t and $|B_t|$ have same law.

The reflection principle simply says that on the event $\{T_a\leq t\}$, the probability of being, at time t, below level $b=a-(a-b)$, is equal to the one of being above level $a+(a-b)$, hence the name. This is related to the fact that the process after time T_a is again BM, which has a symmetric law.

Proof.

- We know from Theorem 3.3.1 that $T_a=\inf\{t\geq 0:B_t=a\}<\infty$ almost surely. We have

$$\mathbb{P}(S_t\geq a,B_t\leq b)=\mathbb{P}(T_a\leq t,B_t\leq b)=\mathbb{P}(T_a\leq t,B'_{t-T_a}\leq b-a)$$

with $B'_t=B_{T_a+t}-B_{T_a}$, where we have used in the last step $B'_{t-T_a}=B_{T_a+t-T_a}-B_{T_a}=B_t-a$ which makes sense on $\{T_a\leq t\}$. Now by the strong Markov property (Theorem 3.5.1), B' is independent of T_a and has the same law as B. Since B' and $-B'$ have same law, it follows that (T_a, B') has the same law as $(T_a, -B')$. Also

$$\begin{aligned}\mathbb{P}(T_a \le t, B'_{t-T_a} \le b-a) &= \mathbb{P}(T_a \le t, -B'_{t-T_a} \le b-a)\\ &= \mathbb{P}(T_a \le t, -(B_t - a) \le b-a)\\ &= \mathbb{P}(T_a \le t, B_t \ge 2a-b)\\ &= \mathbb{P}(B_t \ge 2a-b)\end{aligned}$$

where we have use in the last step the fact that $\{T_a \le t\}$ contains a.s. $\{B_t \ge 2a-b\}$.

- Follows from the first identity with $b = a$, the inequality $B_t \le S_t$, and the fact that B_t and $-B_t$ have same law, which give, for all $a \ge 0$,

$$\begin{aligned}\mathbb{P}(S_t \ge a) &= \mathbb{P}(S_t \ge a, B_t \le a) + \mathbb{P}(S_t \ge a, B_t \ge a)\\ &= \mathbb{P}(B_t \ge a) + \mathbb{P}(B_t \ge a)\\ &= \mathbb{P}(B_t \ge a) + \mathbb{P}(B_t \le -a)\\ &= \mathbb{P}(|B_t| \ge a).\end{aligned}$$

■

Corollary 3.5.3: Densities

Let $(B_t)_{t\ge0}$ be a one-dimensional Brownian motion issued from the origin.

- For all $t > 0$, the law of the couple $(\sup_{s\in[0,t]} B_s, B_t)$ has density

$$(a,b) \in \mathbb{R}^2 \mapsto \frac{2(2a-b)}{\sqrt{2\pi t^3}} \mathrm{e}^{-\frac{(2a-b)^2}{2t}} \mathbf{1}_{a\ge0, b\le a}.$$

- For all $a \in \mathbb{R}$, the law of $T_a = \inf\{t \ge 0 : B_t = a\}$ is equal to the law of $\frac{a^2}{B_1^2}$ with density

$$t \in \mathbb{R} \mapsto \frac{|a|}{\sqrt{2\pi t^3}} \mathrm{e}^{-\frac{a^2}{2t}} \mathbf{1}_{t>0}.$$

The law of T_a is known as the Lévy or Bachelier distribution, special case of $T_{a,0}$ of Corollary 3.8.4. It is, up to scaling by a^2, an inverse χ^2 distribution.

Proof.

- Direct consequence of Corollary 3.5.2.
- Thanks to Corollary 3.5.2, we have, for all $t \ge 0$, denoting $S_t = \sup_{s\in[0,t]} B_s$,

$$\mathbb{P}(T_a \le t) = \mathbb{P}(S_t \ge a) = \mathbb{P}(|B_t| \ge a) = \mathbb{P}(B_t^2 \ge a^2) = \mathbb{P}(tB_1^2 \ge a^2) = \mathbb{P}(a^2/B_1^2 \le t).$$

See Corollary 3.3.2 for the law of T_a via stopped martingales instead of Markov property.

■

3.6 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 $(X_n)_{n\ge1}$ are independent and identically distributed real random variables with law $\mathcal{N}(0,1)$, then this would consist for all $n \ge 1$ to define the Gaussian process

$(X^n_t)_{t\geq 0}$ obtained by linear interpolation as

$$X^n_t = \frac{X_1+\cdots+X_{\lfloor nt\rfloor}}{\sqrt{n}} \sim \mathcal{N}\Big(0,\frac{\lfloor nt\rfloor}{n}\Big),$$

and to consider the limit in law of $(X^n_{t_1},\ldots,X^n_{t_k})$ as $n\to\infty$, for all $k\geq 1$ and $0\leq t_1\leq\cdots\leq t_k$. 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 zero mean and unit variance. Such a functional central limit phenomenon is known as the Donsker invariance principle. From this point of view, Brownian motion is just a universal Gaussian limiting object.

Beyond intuition, the mathematical existence of Brownian motion is not obvious. Historically, Norbert Wiener 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. For more information on history, see [48, 12].

The construction of Brownian motion provided below is based on another very natural idea: by seeing Brownian motion as an infinite family of orthogonal Gaussian random variables, we could start from our favorite infinite dimensional Hilbert space, such as $\mathrm{L}^2(\mathbb{R},\mathrm{d}x)$, and construct a Gaussian random variable by using a linear combination of the elements of a Hilbert basis with Gaussian i.i.d. weights. This will produce a Gaussian process with the desired covariance. It will then remain to obtain the continuity, which can be done by using a general tightness criterion on the increments due to Kolmogorov.

Theorem 3.6.1. Pre-Brownian motion or Gaussian measures.

Let us consider the Hilbert space $\mathrm{G}=\mathrm{L}^2(\mathbb{R},\mathrm{d}x)$ and

$$\langle f,g\rangle_{\mathrm{G}} = \int f(x)g(x)\mathrm{d}x,\quad f,g\in G.$$

Then there exists a centered Gaussian family $\widetilde{B}=(\widetilde{B}_g)_{g\in\mathrm{G}}$ defined on a probability space $(\Omega,\mathcal{A},\mathbb{P})$ such that $g\in\mathrm{G}\mapsto\widetilde{B}_g\in\mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$ is a linear isometry, in other words for all $f,g\in\mathrm{G}$ and $\alpha,\beta\in\mathbb{R}$,

$$\mathbb{E}(\widetilde{B}_f\widetilde{B}_g)=\langle f,g\rangle_{\mathrm{G}}\quad\text{and}\quad \widetilde{B}_{\alpha f+\beta g}=\alpha\widetilde{B}_f+\beta\widetilde{B}_g.$$

Proof. Let $(X_n)_{n\geq 0}$ be i.i.d. real random variables 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 $\mathrm{G}=\mathrm{L}^2(\mathbb{R},\mathrm{d}x)$. For all $g\in\mathrm{G}$, the series

$$\widetilde{B}_g=\sum_{n=0}^{\infty}X_n\langle g,e_n\rangle_{\mathrm{G}}$$

is well defined in $\mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$. Indeed the Cauchy criterion is satisfied:

$$\mathbb{E}\Big(\Big(\sum_{n=p}^{p+q}X_n\langle g,e_n\rangle_{\mathrm{G}}\Big)^2\Big)=\sum_{n=p}^{p+q}\langle g,e_n\rangle_{\mathrm{G}}^2\underset{p,q\to\infty}{\longrightarrow}0.$$

We see from Lemma 3.6.2 that $\widetilde{B}$ is a centered Gaussian random variable and that

$$\|\widetilde{B}_g\|^2=\mathbb{E}((\widetilde{B}_g)^2)=\langle g,g\rangle_{\mathrm{G}}=\|g\|^2_{\mathrm{G}}$$

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

$$4\mathbb{E}(\widetilde{B}_f\widetilde{B}_g)=\mathbb{E}((\widetilde{B}_f+\widetilde{B}_g)^2)-\mathbb{E}((\widetilde{B}_f-\widetilde{B}_g)^2)=\mathbb{E}(\widetilde{B}^2_{f+g})-\mathbb{E}(\widetilde{B}^2_{f-g})=\|f+g\|^2_{\mathrm{G}}-\|f-g\|^2_{\mathrm{G}}=\langle f,g\rangle_{\mathrm{G}}.$$

The sub-space $\mathrm{H}=\overline{\mathrm{span}\{\widetilde{B}_g:g\in\mathrm{G}\}}$ of $\mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$ is isomorphic via $g\mapsto\widetilde{B}_g$ to $\mathrm{G}=\mathrm{L}^2(\mathbb{R},\mathrm{d}x)$.

Lemma 3.6.2. Convergence of Gaussians.

If $(X_n)_n$ are Gaussian real random variables with $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ for a random variable X, then the convergence holds in L^p for all $p \geq 1$, $X \sim \mathcal{N}(m,\sigma^2)$, $\lim_{n\to\infty}\mathbb{E}(X_n) = m$, and $\lim_{n\to\infty}\mathbb{E}((X_n - \mathbb{E}(X_n))^2) = \sigma^2$.

Proof of Lemma 3.6.2. Let us show now that X is Gaussian. Since $X_n \to X$ in law, we get, for all $t \in \mathbb{R}$, $\varphi_{X_n}(t) = \mathrm{e}^{\mathrm{i}tm_n - \frac{t^2}{2}\sigma_n^2} \to \varphi_X(t)$. Thus, for all $\varepsilon > 0$, $|\varphi_{X_n}(\varepsilon)| = \exp(-\frac{\varepsilon^2}{2}\sigma_n^2) \to |\varphi_X(\varepsilon)|$. Since φ_X is a characteristic function, it is non vanishing in a neighborhood of the origin, and thus $\sigma_n \to \sigma_*$ for some $\sigma_* \geq 0$. It follows in turn that for all $t \in \mathbb{R}$, $\mathrm{e}^{\mathrm{i}tm_n} \to \mathrm{e}^{\mathrm{i}tm_*}$ for some m_*. Now by dominated convergence,

$$\sqrt{2\pi}\mathrm{e}^{-\frac{m_n^2}{2}} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tm_n}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tm_*}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t = \sqrt{2\pi}\mathrm{e}^{-\frac{m_*^2}{2}}$$

thus $m_n \to m_*$. Hence $X \sim \mathcal{N}(m_*,\sigma_*^2)$, and $(m_*,\sigma_*^2) = (m,\sigma^2)$. Finally, for all $p \geq 1$, since $\mathbb{E}(|X_n|^p)$ is a continuous function of m_n and σ_n, it is bounded in n, thus $(X_n)_n$ is u.i. and therefore $X_n \to X$ in L^p. ■

■

With $\widetilde{B}$ being as in Theorem 3.6.1, let us define, for all $t \geq 0$, the random variable

$$B_t = \widetilde{B}_{\mathbf{1}_{[0,t]}}.$$

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

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

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

$$\frac{B_t - B_s}{\sqrt{t-s}} \sim \mathcal{N}(0,1), \quad \text{thus}^1 \quad \mathbb{E}((B_t - B_s)^{2n}) = c_n(t-s)^n.$$

The fourth moment case $n = 2$ allows, thanks to Theorem 3.6.3 below ($p = 4$, $\varepsilon = 1$, $\gamma < \varepsilon/p = 1/4$), to construct a continuous modification B^* of B, which is a Brownian motion on $\mathbb{R}$ issued from the origin. Moreover using the higher moments for all values of n gives the optimal Hölder regularity: $\gamma < \frac{n-1}{2n} \to \frac{1}{2}$.

Theorem 3.6.3. Kolmogorov continuity criterion.

Let $X = (X_t)_{t\geq 0}$ be a process defined on a probability space $(\Omega,\mathcal{A},\mathbb{P})$ taking its values in a Banach space $\mathbb{B}$ with norm $\|\cdot\|$, and such that the following tightness of increments property holds: 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[a] of X that is a continuous process whose trajectories are, on each finite interval, γ-Hölder continuous for all $\gamma \in [0,\varepsilon/p)$, in the sense that a.s. for all $t > 0$, there exists a constant $C = C(\omega,t) > 0$ such that for all $u, v \in [0,t]$ and all η, if $|u - v| \leq \eta$ then $|X_u - X_v| \leq C\eta^\gamma$.

[a] There exists $X^* = (X^*_t)_{t\geq 0}$ such that for all $t \geq 0$, $X_t = X^*_t$ as random variables in other words almost surely.

Proof. The proof is an instance of the chaining method or technique, invented by Kolmogorov. It suffices to prove the result on a finite time interval $[0,t]$. Let us first show that X is Hölder continuous on the dyadics $\mathcal{D} = \cup_{n\geq 0}\mathcal{D}_n$ where $\mathcal{D}_n = \{tk/2^n : k \in \{0,\ldots,2^n\}\} \subset \mathcal{D}_{n+1}$. For notation simplicity we take $t = 1$. For all $n \geq 1$, all $\varepsilon > 0$, and all $\gamma > 0$, the Markov inequality gives

$$\mathbb{P}\Big(\max_{1\leq k\leq 2^n}\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\Big) \leq \sum_{k=1}^{2^n}\mathbb{P}\Big(\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\Big)$$

[1] We have $c_n = \mathbb{E}(Z^{2n}) = \frac{(2n)!}{2^n n!}$ where $Z \sim \mathcal{N}(0,1)$ but this explicit formula for c_n is useless for our purposes.

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

Now we take $\varepsilon > \gamma p$, to get

$$\sum_{n=1}^{\infty}\mathbb{P}(\max_{1\leq k\leq 2^n}\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}) < \infty.$$

Thus, the Borel–Cantelli lemma provides $A \in \mathscr{A}$ such that $\mathbb{P}(A) = 1$ and for all $\omega \in A$, there exists n_ω such that for all $n \geq n_\omega$, we have $\max_{1\leq k\leq 2^n}\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \leq 2^{-\gamma n}$. Hence there exists a random variable C such that

$$C < \infty \text{ a.s.} \quad \text{and} \quad \max_{1\leq k\leq 2^n}\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \leq C2^{-\gamma n}.$$

Let us prove that on A, the paths of X are γ-Hölder continuous on $\mathscr{D}$. Let $s,t \in \mathscr{D}$ with $s \neq t$. We have $s,t \in \mathscr{D}_n$ for large enough n, and we can take n large enough to get $2^{-n} \leq |t-s|$. Let us write $s = k_1/2^n$ et $t = k_2/2^n$, $k_1 < k_2$, which gives $(k_2 - k_1) = 2^n|t-s|$. Now we proceed by chaining to exploit the preceding estimate,

$$\|X_t - X_s\| \leq \sum_{k=k_1+1}^{k_2}\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \leq (k_2-k_1)C2^{-\gamma n} = C|t-s|2^{-n(\gamma-1)} \leq C|t-s|^\gamma.$$

Thus, on A, the sample paths of X are γ-Hölder continuous on $\mathscr{D}$. The set $\mathscr{D}$ is dense in $\mathbb{R}_+$. Now for all $\omega \in A$, let $t \mapsto X^*_t(\omega)$ be the unique continuous function[2] agreeing with $t \mapsto X_t(\omega)$ on $\mathscr{D}$.

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

Corollary 3.6.4. Existence.

One-dimensional Brownian motion exists, and thus d-dimensional Brownian motion for all $d \geq 1$. Moreover, almost surely, the trajectories of real Brownian motion are, on each finite time interval, Hölder continuous of order γ for all $\gamma \in (0,1/2)$, not more.

Proof. Theorem 3.6.3 with $p = 2n$ and $n \to \infty$ gives $\gamma \in (0,1/2)$, while Theorem 3.4.4 gives the optimality. ∎

3.7 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 3.7.1 gives a meaning to the Gaussian random variable

$$\omega \in \Omega \mapsto \int_0^\infty g(s)\mathrm{d}B_s(\omega)$$

where the integrator $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion and the integrand g is in $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+,\mathrm{d}x)$. The integrand is deterministic and square integrable, while the integrator is random and Gaussian.

Theorem 3.7.1. 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 from 0, defined on $(\Omega,\mathscr{A},\mathbb{P})$. Let G be the Gaussian sub-space of $\mathrm{L}^2(\Omega,\mathbb{P})$ generated by the real random variables $\{B^i_t :$

[2] We can use here the following general property of metric spaces: if S and T are metric spaces with S complete, if D is a dense subset of S, and if $f : D \to T$ is uniformly continuous, then there exists a unique continuous $\widetilde{f} : S \to T$ that agrees with f on D.

$t \geq 0, 1 \leq i \leq d\}$. Then there exists a unique map $I : \mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x) \to \mathrm{G}$ such that:

1. If $g = a\mathbf{1}_{(s,t]}$ with $0 \leq s \leq t$ and $a \in \mathbb{R}^d$ then $I(g) = a \cdot (B_t - B_s)$
2. I is an isometry (and thus continuous) in the sense that for all f and g in $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$, we have

$$\underbrace{\mathbb{E}(I(f)I(g))}_{\langle I(f), I(g)\rangle_{\mathrm{L}^2(\Omega,\mathbb{P})}} = \underbrace{\int_0^\infty f(s) \cdot g(s)\mathrm{d}s}_{\langle f, g\rangle_{\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+,\mathrm{d}x)}}$$

3. I is linear and bijective

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

$$I(g)(\omega) = \int_0^\infty g(s)\mathrm{d}B_s(\omega).$$

Proof. The following sub-space

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

of $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$ is dense. If $f \in S$, then $f = \sum_{\text{finite}} a_i \mathbf{1}_{(t_i, t_{i+1}]}$, 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 the map $f \mapsto I(f)$ is linear. Moreover, we remark that thanks to the properties of Brownian motion, we have,

$$\begin{aligned}
\mathbb{E}((I(f))^2) &= \sum_{i,j} \mathbb{E}((a_i \cdot (B_{t_{i+1}} - B_{t_i}))(a_j \cdot (B_{t_{j+1}} - B_{t_j}))) \\
&= \sum_i \mathbb{E}((a_i \cdot (B_{t_{i+1}} - B_{t_i}))^2) \\
&= \sum_i \mathbb{E}\Big(\Big(\sum_j a_{i,j}(B^j_{t_{i+1}} - B^j_{t_i})\Big)^2\Big) \\
&= \sum_i \sum_{j,k} a_{i,j} a_{i,k} \mathbb{E}((B^j_{t_{i+1}} - B^j_{t_i})(B^k_{t_{i+1}} - B^k_{t_i})) \\
&= \sum_i \sum_{j,k} a_{i,j} a_{i,k} (t_{i+1} - t_i)\mathbf{1}_{j=k} \\
&= \sum_i |a_i|^2 (t_{i+1} - t_i) \\
&= \int_0^\infty |f(x)|^2 \mathrm{d}x.
\end{aligned}$$

Since S is dense, I can be extended by continuity to the whole space $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$. Namely, for all $f \in \mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$, there exists a sequence $(f_n)_n$ is S such that $\|f_n - f\| \to 0$. Therefore

$$\|f_n - f_m\| = \|I(f_n) - I(f_m)\|_{\mathrm{L}^2(\Omega,\mathscr{A},\mathbb{P})} \xrightarrow[m,n\to\infty]{} 0.$$

Set $I(f) = \lim_{n\to\infty} I(f_n)$. This limit does not depend on the sequence $(f_n)_n$ 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 $f, g \in \mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$,

$$\int_0^\infty f(s)g(s)\mathrm{d}s = \frac{1}{4}\int_0^\infty ((f+g)^2 - (f-g)^2)(s)\mathrm{d}s = \frac{1}{4}\mathbb{E}((I(f+g))^2 - (I(f-g))^2) = \mathbb{E}(I(f)I(g)).$$

The map I defined this way is unique. It is injective since it is an isometry. Finally it is surjective since $F = I(\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x))$ is a closed sub-space of G while F is dense in G (taking $g = e_i \mathbf{1}_{(0,t]}$ gives $B^i_t \in F$). ■

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

Corollary 3.7.2. Properties of the Wiener integral.

1. For all $f \in \mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$, $I(f) \sim \mathcal{N}(0, \|f\|_2^2)$
2. For all $f, g \in \mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$, $I(f)$ and $I(g)$ are independent iff $\langle f, g\rangle_{\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+,\mathrm{d}x)} = \int_0^\infty f(s)g(s)\mathrm{d}s = 0$
3. For all $t \geq 0$ and $1 \leq i \leq d$ and $f \in \mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$, we have

$$\mathbb{E}\left(B_t^i \int_0^\infty f(s)\mathrm{d}B_s\right) = \int_0^t f^i(s)\mathrm{d}s$$

   where $f^i(s)$ is the i-th coordinate of $f(s) = (f^1(s), \ldots, f^d(s))$
4. For all $f \in \mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$, $(I(f\mathbf{1}_{(0,t]}))_{t\geq 0}$ is a martingale for the natural filtration of $(B_t)_{t\geq 0}$
5. If $(f_n)_{n\geq 0}$ is an orthonormal basis of $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$, then $(I(f_n))_{n\geq 0}$ are i.i.d. standard Gaussian real random variables and for all $t \geq 0$, we have the following expansion in $\mathrm{L}^2(\Omega, \mathcal{A}, \mathbb{P})$:

$$B_t = \sum_{n\geq 0} \underbrace{\left(\int_0^\infty f_n(s)\mathrm{d}B_s\right)}_{\substack{I(f_n) \\ \text{Gaussian r.r.v.}}} \underbrace{\int_0^t f_n(s)\mathrm{d}s}_{\substack{\text{deterministic} \\ \text{vector}}}.$$

The Wiener integral produces plenty of martingales from Brownian motion! These martingales are continuous, but we prove this later on for the more general concept of Itô stochastic integral (Theorem 5.2.2).

Proof. The first two items are immediate.

3. Take $g = e_i \mathbf{1}_{(0,t]}$ then by definition of I we have $I(g) = B_t^i$ and

$$\mathbb{E}\left(B_t^i \int_0^\infty f(s)\mathrm{d}B_s\right) = \mathbb{E}(I(g)I(f)) = \int_0^\infty g(s)\cdot f(s)\mathrm{d}s = \int_0^t f^i(s)\mathrm{d}s.$$

4. For all $f \in \mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$, all $t \geq 0$, and all $s \in [0,t]$, $M_{s,t} = I(f\mathbf{1}_{(s,t]})$ is the L^2 limit of linear combinations $\sum_{\text{finite}} a_i(B_{v_i} - B_{u_i})$ with $a_i \in \mathbb{R}^d$ and $u_i, v_i \in (s,t]$. In particular it is integrable, centered, measurable for $\mathcal{F}_t$, and independent of $\mathcal{F}_s$. Now $M_{0,t} = M_{0,s} + M_{s,t}$ and thus $\mathbb{E}(M_{0,t} \mid \mathcal{F}_s) = M_{0,s} + \mathbb{E}(M_{s,t}) = M_{0,s}$.

5. If $(f_n)_{n\geq 0}$ is an orthonormal basis of $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$ then $(I(f_n))_{n\geq 0}$ is an orthonormal basis of the Gaussian space G and moreover, by the previous property, $\langle B_t^i, I(f_n)\rangle_{\mathrm{G}} = \int_0^t f_n^i(s)\mathrm{d}s$. Note that $(I(f_n))_{n\geq 0}$ is orthonormal in $\mathrm{L}^2(\Omega, \mathbb{P})$ but is not a basis: the closure of its span is $G \subsetneq \mathrm{L}^2(\Omega, \mathbb{P})$.

■

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

Let $(B_t)_{t\geq 0}$ be an arbitrary d-dimensional Brownian motion issued from 0, 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 2.1.3, that we can consider $(B_t)_{t\geq 0}$ as a random variable from $(\Omega', \mathcal{A}', \mathbb{P})$ to $(\mathrm{W}, \mathcal{B}_{\mathrm{W}})$ where $\mathrm{W} = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ is equipped with the topology of uniform convergence on every compact subset of $\mathbb{R}_+$ and where $\mathcal{B}_{\mathrm{W}}$ is the associated Borel σ-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 μ. 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 $(\mathrm{W}, \mathcal{B}_W, \mu)$. Namely, on the probability space $(\mathrm{W} = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d), \mathcal{B}_w, \mu)$, where μ 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 $w \in \mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$. Under μ, the process π is a d-dimensional Brownian motion issued from the origin. It is called the canonical Brownian motion.

Theorem 3.8.1. Wiener measure.

There exists a unique probability measure μ on the canonical space $(\mathrm{W}, \mathscr{B}_{\mathrm{W}})$, called the Wiener measure, such that for all $n \geq 1$, $0 < t_1 < \cdots < t_n$, $A_1, \ldots, A_n \in \mathscr{B}_{\mathbb{R}^d}$,

$$\mu(\{w \in \mathrm{W} : w_{t_1} \in A_1, \ldots, w_{t_n} \in A_n\}) = \int_{A_1 \times \cdots \times A_n} p_{t_1 - t_0}(x_1 - x_0) \cdots p_{t_n - t_{n-1}}(x_n - x_{n-1}) \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

where $t_0 = 0$, $x_0 = 0$, p is the heat or Gaussian kernel defined for all $t > 0$ and $x \in \mathbb{R}^d$ by

$$p_t(x) = \frac{\mathrm{e}^{-\frac{|x|^2}{2t}}}{(\sqrt{2\pi t})^d}.$$

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 $\Phi : \mathrm{W} \to \mathbb{R}$,

$$\mathbb{E}(\Phi(B)) = \int_{\mathrm{W}} \Phi(w) \mu(\mathrm{d}w).$$

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

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

Finally μ is unique because it is entirely determined on the family $\mathscr{C}$ of cylindrical subsets of W, which is stable by finite intersections and generates $\mathscr{B}_{\mathrm{W}}$ (monotone class argument, Corollary 1.8.4). ■

Cameron – Martin formula

A natural motivation is the study of the solution X of the stochastic differential equation

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

where σ, b, and the Brownian motion B are given. Provided that it is well defined, we would like to know if the solution X has a density with respect to B, on the space of trajectories. The Cameron – Martin formula provides an answer in the special case in which $X_0 = 0$, $\sigma(s) = I_d$ and $b(s, x) = b(s)$, for which we have

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

This corresponds to study the density of the shifted Wiener measure with respect to the initial Wiener measure. But let us examine first the case of shifted finite dimensional Gaussian distributions. If $Z \sim \mathcal{N}(0, I_n)$ is a standard Gaussian random variable on $\mathbb{R}^n$ with density $x \mapsto \gamma_n(x) = (2\pi)^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2}|x|^2}$ with respect to the Lebesgue measure, then, for all $h \in \mathbb{R}^n$ and all bounded and measurable $\Phi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}(\Phi(Z + h)) = \int_{\mathbb{R}^n} \Phi(x + h) \gamma_n(x) \mathrm{d}x = \int_{\mathbb{R}^n} \Phi(x') \gamma_n(x' - h) \mathrm{d}x' = \int_{\mathbb{R}^n} \Phi(x') \gamma_n(x') \mathrm{e}^{x' \cdot h - \frac{|h|^2}{2}} \mathrm{d}x' = \mathbb{E}\Big(\Phi(Z) \mathrm{e}^{Z \cdot h - \frac{|h|^2}{2}}\Big).$$

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 the set of continuous functions h which are the integral of a square integrable function denoted $\dot{h}$, namely

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

This is a subspace of $\mathrm{W} = \mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$. Note that every element $h \in H$ is differentiable and its derivative is $\dot{h}$, hence the notation. Moreover the representation is unique in the sense that for all $h_1, h_2 \in H$, $h_1 = h_2$ implies $\dot{h}_1 = \dot{h}_2$. We equip H with the scalar product

$$\langle h_1, h_2 \rangle_{\mathrm{H}} = \int_0^\infty \dot{h}_1(s) \cdot \dot{h}_2(s)\mathrm{d}s.$$

This makes H a Hilbert space isomorphic to $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$. For every $h \in \mathrm{H}$ we denote

$$|h|^2_{\mathrm{H}} = \int_0^\infty |\dot{h}(s)|^2 \mathrm{d}s = \|\dot{h}\|^2_{\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)}.$$

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

$$h_n(t) = \int_0^t \dot{h}_n(s)\mathrm{d}s.$$

The sequence $(\dot{h}_n)_{n\geq 0}$ is an orthonormal basis of $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$. Let $\pi = (\pi_t)_{t\geq 0}$ be the canonical Brownian motion on $\mathbb{R}^d$ and let us define, for almost all $w \in \mathrm{W}$ and $h \in \mathrm{H}$, using the Wiener integral of Theorem 3.7.1,

$$(w, h) = \int_0^\infty \dot{h}(s)\mathrm{d}\pi_s(w) = \int_0^\infty \dot{h}(s)\mathrm{d}w_s.$$

Now Corollary 3.7.2 gives that $(w, h_n), n \geq 0$, are i.i.d. standard Gaussian real random variables, and for all fixed $t \geq 0$ we have the following expansion in $\mathrm{L}^2(\mathrm{W}, \mathscr{B}_{\mathrm{W}}, \mu)$:

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

Here t is fixed and $h_n(t)$ is a deterministic vector, while $(w, h_n)_{n\geq 0}$ are random and orthogonal in $\mathrm{L}^2(\mathrm{W}, \mathscr{B}_{\mathrm{W}}, \mu)$.

Recall that a notion of density of Wiener measure would require a notion of Lebesgue measure on Wiener space, which is missing[3]. However the shift of the Wiener measure in a direction picked in the Cameron – Martin space is absolutely continuous with respect to the Wiener measure, with explicit density. This is an infinite dimensional analogue of the formula above for finite dimensional Gaussian laws.

Theorem 3.8.2. Cameron[a] – Martin[b] formula and density of shifts.

[a]Named after Robert Horton Cameron (1908 – 1989), American mathematician.
[b]Named after William Ted Martin (1911 – 2004), American mathematician.

Let W and μ be the Wiener space and measure, $\Phi : \mathrm{W} \to \mathbb{R}$ be measurable and bounded, and h be in the Cameron – Martin space H. Then we have the Cameron – Martin formula:

$$\int \Phi(w + h)\mu(\mathrm{d}w) = \int \Phi(w) \mathrm{e}^{(w,h) - \frac{1}{2}|h|^2_{\mathrm{H}}} \mu(\mathrm{d}w).$$

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

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

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 μ.

[3]It can be shown that on an infinite-dimensional separable Banach space equipped with its Borel σ-algebra, the only locally finite and translation-invariant Borel measure is the trivial measure identically equal to zero. Equivalently, every translation-invariant measure that is not identically zero assigns infinite measure to all open subsets. See for instance `https://en.wikipedia.org/wiki/Infinite-dimensional_Lebesgue_measure` and references therein.

Proof. By a monotone class argument, we can assume without loss of generality that $\Phi(w) = f(w_{t_1}, \ldots, w_{t_n})$ where $n \geq 1$ and $0 \leq t_1 < \cdots < t_n$ and $f : \mathbb{R}^n \to \mathbb{R}$ measurable and bounded. Let $h \in \mathrm{H}$, $h \neq 0$. There exists an orthonormal basis $(h_m)_{m \geq 0}$ of H such that $h_0 = h/|h|_{\mathrm{H}}$. For all $\omega \in \mathrm{W}$ and $m \geq 1$, let us define $w^{(m)} \in \mathrm{W}$ by

$$w^{(m)} = \sum_{\ell=0}^{m} (w, h_\ell) h_\ell, \quad \mu \text{ almost surely.}$$

We have $\lim_{m\to\infty} w^{(m)} = w = \sum_{n \geq 0} (w, h_n) h_n$ in $\mathrm{L}^2(\mathrm{W})$. Since Φ 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 $(w, h) \sim \mathcal{N}(0, |h|_{\mathrm{H}}^2)$ under μ,

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

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

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

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

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

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

$$\begin{aligned}
\mathbb{E}_\mu(\Phi(w^{(m)} + h)) &= \frac{1}{(\sqrt{2\pi})^{m+1}} \int_{\mathbb{R}^{m+1}} \Phi\Big((x_0 + |h|_{\mathrm{H}}) h_0 + \sum_{\ell=0}^{m} x_\ell h_\ell\Big) \mathrm{e}^{-\frac{1}{2}\sum_{\ell=0}^{m} x_\ell^2} \mathrm{d}x_0 \cdots \mathrm{d}x_m \\
&= \frac{1}{(\sqrt{2\pi})^{m+1}} \int_{\mathbb{R}^{m+1}} \Phi\Big(x_0' h_0 + \sum_{\ell=0}^{m} x_\ell' h_\ell\Big) \mathrm{e}^{x_0'|h|_{\mathrm{H}} - \frac{1}{2}|h|_{\mathrm{H}}^2 - \frac{1}{2}\sum_{\ell=0}^{m} x_\ell'^2} \mathrm{d}x_0' \cdots \mathrm{d}x_m' \\
&= \mathbb{E}_\mu(\Phi(w^{(m)}) \mathrm{e}^{(w, h_0)|h|_{\mathrm{H}} - \frac{|h|_{\mathrm{H}}^2}{2}}) \\
&= \mathbb{E}_\mu(\Phi(w^{(m)}) \mathrm{e}^{(w, h) - \frac{|h|_{\mathrm{H}}^2}{2}}).
\end{aligned}$$

Note that we do not need a regular Φ or f to perform a linear substitution (change of variable). ■

Corollary 3.8.3. Density of Cameron – Martin shifts.

Let $B = (B_t)_{t \geq 0}$ be a d-dimensional Brownian motion issued from the origin defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, seen as a random variable of law μ taking values on the canonical space $(\mathrm{W}, \mathscr{B}_{\mathrm{W}})$. For all $h \in \mathrm{H}$, let F_h be as in Theorem 3.8.2. Then:

1. For all $h \in \mathrm{H}$, we have $\mathbb{E}(F_h(B)) = 1$ and
$$F_h(B) = \exp\Big(\int_0^\infty \dot{h}_s \mathrm{d}B_s - \frac{1}{2} \int_0^\infty |\dot{h}_s|^2 \mathrm{d}s\Big).$$
2. For all $h \in \mathrm{H}$, with respect to $(\Omega, \mathscr{A}, \mathbb{Q})$ given by $\mathrm{d}\mathbb{Q} = F_h(B)\mathrm{d}\mathbb{P}$, the shifted process
$$(B_t - h_t)_{t \geq 0}$$
is a d-dimensional Brownian motion issued from the origin.
3. For all $h \in \mathrm{H}$, the law of the shifted process
$$(B_t + h_t)_{t \geq 0}$$
is absolutely continuous with respect to the Wiener measure μ, with density F_h.

By definition of the Wiener measure μ, we have $\mathbb{E}(F_h(B)) = \mathbb{E}_\mu(F_h)$. The first $\mathbb{E}$ is on the probability space $(\Omega, \mathscr{A}, \mathbb{P})$ used to define B while the second is on the canonical Wiener space $(W, \mathscr{B}_W, \mu)$. It is customary to omit the dependency $\mathbb{E} = \mathbb{E}_\mathbb{P}$ over the underlying probability space for an abstract random variable like B.

An extension of Corollary 3.8.3 to random shifts is given by the Girsanov theorem (Theorem 7.5.1).

Proof.

1. Since $\int_0^\infty \dot{h}_s \mathrm{d}B_s \sim \mathcal{N}(0, |h|_{\mathrm{H}}^2)$, the formula $\mathbb{E}(F_h(B)) = 1$ comes from the Laplace transform formula

$$\mathbb{E}\Big(\exp\Big(\int_0^\infty \dot{h}_s \mathrm{d}B_s\Big)\Big) = \exp\Big(\frac{|h|_{\mathrm{H}}^2}{2}\Big).$$

2. By using Theorem 3.8.2 in $\star$, denoting $\Phi_h = \Phi(\cdot - h)$,

$$\mathbb{E}(\Phi(B)) = \mathbb{E}(\Phi(B+h-h)) = \mathbb{E}(\Phi_h(B+h)) \overset{\star}{=} \mathbb{E}(\Phi_h(B)F_h(B)) = \mathbb{E}(\Phi(B-h)F_h(B)) = \mathbb{E}_\mathbb{Q}(\Phi(B-h)).$$

3. Theorem 3.8.2 writes, for all measurable and bounded $\Phi : \mathrm{W} \to \mathbb{R}$,

$$\mathbb{E}(\Phi(B+h)) = \mathbb{E}\Big(\Phi(B)\exp\Big(\int_0^\infty \dot{h}_s \mathrm{d}B_s - \frac{|h|_{\mathrm{H}}^2}{2}\Big)\Big) = \mathbb{E}(\Phi(B)F_h(B)) = \int \Phi(w)F_h(w)\mu(\mathrm{d}w),$$

hence the result. Note that taking $\Phi \equiv 1$ gives also that $\mathbb{E}(F_h(B)) = 1$.

Alternatively, let $\mathbb{Q}$ be as above and $W = B - h$. Then $W + h = B$, and by definition of $\mathbb{Q}$,

$$\mathbb{E}_\mathbb{Q}(\Phi(W+h)) = \mathbb{E}_\mathbb{Q}(\Phi(B)) = \mathbb{E}(\Phi(B)N_T),$$

and on the other hand, since h is deterministic and since W has the law of BM under $\mathbb{Q}$, we have

$$\mathbb{E}_\mathbb{Q}(\Phi(W+h)) = \mathbb{E}(\Phi(B+h)).$$

■

This can be skipped at first reading.

Corollary 3.8.4: Hitting time of Brownian motion with drift

For all $a > 0$ and $c \in \mathbb{R}$, the law of $T_{a,c} = \inf\{t \geq 0 : B_t + ct = a\}$ is $\mu + p\delta_\infty$ where μ has density

$$s \in \mathbb{R} \mapsto \frac{a}{\sqrt{2\pi s^3}} \mathrm{e}^{-\frac{(a-cs)^2}{2s}} \mathbf{1}_{s>0},$$

and where

$$p = \mathbb{P}(T_{a,c} = \infty) = \begin{cases} 0 & \text{if } c \geq 0 \\ 1 - \mathrm{e}^{2ac} & \text{if } c \leq 0 \end{cases}.$$

The law μ is known as an inverse Gaussian or Wald[4] distribution.
Several other facts and formulas about Brownian motion can be found in the book [6].

Proof. Let us define $\dot{h}(s) = c\mathbf{1}_{s\leq t}$, which gives $h(s) = c(s \wedge t)$. The Cameron – Martin formula of Corollary 3.8.3 with $\Phi(w) = \mathbf{1}_{\{\max_{s\in[0,t]} w(s) \geq a\}}$ gives

$$\begin{aligned}
\mathbb{P}(T_{a,c} \leq t) &= \mathbb{E}(\Phi(B+h)) \\
&= \mathbb{E}\Big(\Phi(B)\exp\Big(\int_0^\infty \dot{h}(s)\mathrm{d}B_s - \frac{1}{2}\int_0^\infty \dot{h}(s)^2 \mathrm{d}s\Big)\Big) \\
&= \mathbb{E}\Big(\mathbf{1}_{\{T_{a,0}\leq t\}} \exp\Big(cB_t - \frac{c^2}{2}t\Big)\Big)
\end{aligned}$$

$$
\begin{aligned}
&\overset{\star}{=} \mathbb{E}\Big(\mathbf{1}_{\{T_{a,0}\le t\}} \exp\Big(cB_{t\wedge T_{a,0}} - \frac{c^2}{2}(t\wedge T_{a,0})\Big)\Big)\\
&= \mathbb{E}\Big(\mathbf{1}_{\{T_{a,0}\le t\}} \exp\Big(ca - \frac{c^2}{2}T_{a,0}\Big)\Big)\\
&\overset{\star\star}{=} \int_0^t \frac{a}{\sqrt{2\pi s^3}} \mathrm{e}^{-\frac{a^2}{2s}} \mathrm{e}^{ca-\frac{c^2}{2}s} \mathrm{d}s\\
&= \int_0^t \frac{a}{\sqrt{2\pi s^3}} \mathrm{e}^{-\frac{(a-cs)^2}{2s}} \mathrm{d}s.
\end{aligned}
$$

We have used for $\star$ the Doob stopping (Theorem 2.5.1) with the martingale $(\mathrm{e}^{cB_t - \frac{c^2}{2}t})_{t\ge 0}$ to get

$$
\mathbb{E}\Big(\mathrm{e}^{cB_t-\frac{c^2}{2}t} \mid \mathscr{F}_{t\wedge T_{a,0}}\Big) = \mathrm{e}^{cB_{t\wedge T_{a,0}} - \frac{c^2}{2}(t\wedge T_{a,0})},
$$

and for $\star\star$ the density of $T_{a,0}$ given by Corollary 3.5.3. The density of $T_{a,c}$ gives in turn the formula for $\mathbb{P}(T_{a,c} < \infty)$, which follows also from Doob stopping with the martingale $(\mathrm{e}^{-2c(B_t+ct)})_{\{t\ge 0\}}$. ■

Named after Abraham Wald (1902 – 1950), Hungarian mathematician.

Chapter 4

More on martingales

For simplicity, this chapter is about continuous processes only.

4.1 Quadratic variation, square integrable martingales, increasing process

Definition 4.1.1. Quadratric variation if square integrable processes.

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)

$$[X]_t = \overset{\mathbb{P}}{\lim_{|\delta|\to 0}} \sum_k (X_{t_{k+1}} - X_{t_k})^2$$

where the convergence takes place in probability, and where $\delta : 0 = t_0 < \cdots < t_n = t,\ n = n_\delta \geq 1$, 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 δ. More generally, the quadratic covariation process of a couple of square integrable real processes $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ is defined for all $t \geq 0$ by the following limit when it exists:

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

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

We use convergence in probability because we do not know if the process has high enough moments. Recall that for Brownian motion we have used the fourth moment for L^2 convergence of quadratic variation.

Theorem 3.2.1 states that for a BM B, we have, for all $t \geq 0$, $[B]_t = t$. Theorem 4.1.4 states that for all any square integrable continuous martingale M issued form the origin, for all $t \geq 0$, $\mathbb{E}([M]_t) = \mathbb{E}(M_t^2)$.

Lemma 4.1.2. 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. In other words, for a continuous process, non-zero quadratic variation implies infinite variation.

On the same topic, Lemma 4.1.6 states that a finite variation continuous martingale is constant.

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

$$\sum_k (X_{t_{k+1}} - X_{t_k})^2 \leq \max_k |X_{t_{k+1}} - X_{t_k}| \sum_k |X_{t_{k+1}} - X_{t_k}| \underset{|\delta|\to 0}{\longrightarrow} 0.$$

The max part of the right hand side tends to 0 since X is continuous and thus uniformly continuous (Heine), while the $\sum$ part is bounded by the 1-variation of X on $[0, t]$ which is finite since X has finite variation. ■

Coding in action 4.1.3. Quadratic variation of BM.

Could you write a code simulating approximate trajectories of one-dimensional Brownian motion and their approximate quadratic variation, and plotting both on the same graphic?

We denote by $\mathscr{M}^2$ the set of square integrable continuous martingales.
We denote by $\mathscr{M}_0^2$ the set of square integrable continuous martingales issued from the origin.
We often use the following properties for any $M \in \mathscr{M}^2$:

- Squared L^2 norm of increments: for all $0 \leq s \leq t$,

$$\mathbb{E}((M_t - M_s)^2) = \mathbb{E}(\mathbb{E}(M_t^2 - 2M_sM_t + M_s^2 \mid \mathscr{F}_s)) = \mathbb{E}(M_t^2) - \mathbb{E}(M_s^2)$$

  and thus for any subdivision $s = t_0 < \cdots < t_n = t$, by telescoping summation,

$$\sum_{i=1}^{n} \mathbb{E}((M_{t_i} - M_{t_{i-1}})^2) = \mathbb{E}(M_t^2) - \mathbb{E}(M_s^2).$$

- (Conditional) orthogonal increments in L^2: for all $0 \leq s \leq t \leq u \leq v$ we have

$$\mathbb{E}((M_t - M_s)(M_v - M_u) \mid \mathscr{F}_t) = (M_t - M_s)\underbrace{\mathbb{E}(M_v - M_u \mid \mathscr{F}_t)}_{=M_t - M_t = 0.} = 0.$$

The following theorem is a crucial result of martingale theory.

Theorem 4.1.4. Increasing process or angle bracket.

Let $M \in \mathscr{M}_0^2$.

- There exists a unique continuous and non-decreasing process denoted $\langle M\rangle = (\langle M_t\rangle)_{t\geq 0}$ such that $\langle M\rangle_0 = 0$ and $(M_t^2 - \langle M_t\rangle)_{t\geq 0}$ is a martingale. In particular $\langle M\rangle$ is adapted.
- For all $t \geq 0$, the quadratic variation $[M]_t$ exists and $[M]_t = \langle M\rangle_t$.

Uniqueness is up to indistinguishability.
The process $\langle M\rangle$ is called the increasing process or angle bracket of M, or even the *compensator* of M^2.
If $M \in \mathscr{M}^2$ with $M_0 \neq 0$ then we define $[M] = [M - M_0]$ and $\langle M\rangle = \langle M - M_0\rangle$.

If B is a Brownian motion, Theorem 3.1.6 gives that $\langle B\rangle_t = t$ for all $t \geq 0$ by showing that $(B_t^2 - t)_{t\geq 0}$ is a martingale, while Theorem 3.2.1 gives that $[B]_t = t$ for all $t \geq 0$ by computing the quadratic variation. More generally Lemma 4.2.6 states that for all continuous local martingale M issued from the origin, $[M] = \langle M\rangle$.

In Theorem 4.1.4, M^2 is a sub-martingale, and actually Theorem 4.1.4 states a special case of the more general Doob–Meyer[1] decomposition of sub-martingales which is beyond the scope of this course.

Corollary 4.1.5. Boundedness in L^2.

If $M \in \mathscr{M}_0^2$ then there exists a random variable $\langle M\rangle_\infty$ taking values in $[0, +\infty]$ such that almost surely

$$\langle M\rangle_t \underset{t\to\infty}{\nearrow} \langle M\rangle_\infty,$$

and moreover M is bounded in L^2 if and only if $\langle M\rangle_\infty \in \mathrm{L}^1$, more precisely, in $[0, +\infty]$,

$$\mathbb{E}(\langle M\rangle_\infty) = \sup_{t\geq 0} \mathbb{E}(M_t^2).$$

[1] Named after Paul-Anré Meyer (1934–2003), French mathematician.

Proof of Corollary 4.1.5. The first property follows from the monotony and positivity of $\langle M\rangle$. For the second property, since $M^2-\langle M\rangle$ is a martingale we get $\mathbb{E}(M_t^2)=\mathbb{E}(\langle M\rangle_t)$ for all $t\geq 0$, and by monotone convergence,

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

■

This can be skipped at first reading.

Proof of Theorem 4.1.4.

- Existence of $\langle M\rangle$ and $[M]$ and their equality when M is bounded. Let us fix $t>0$ and let $(\delta_n)_n$ be a sequence of partitions of $[0,t]$, $\delta_n : 0=t_0^n<\cdots<t_{r_n}^n=t$ with $|\delta_n|=\max_{1\leq k\leq r_n}(t_k^n-t_{k-1}^n)\to 0$ as $n\to\infty$. It can be checked that the process $X=(X_s)_{s\in[0,t]}$ defined by

$$X_s^n=\sum_{k=1}^{r_n}M_{t_{k-1}^n}(M_{t_k^n\wedge s}-M_{t_{k-1}^n\wedge s})$$

is a (bounded) martingale (it is crucially zero when $s\leq t_i^n$, see Lemma 6.1.3 for more), and that

$$M_{t_k^n}^2-2X_{t_k^n}^n=\sum_{i=1}^k(M_{t_i^n}-M_{t_{i-1}^n})^2.$$

Now it turns out that

$$\lim_{m,n\to\infty}\mathbb{E}((X_s^n-X_s^m)^2)=0.$$

It follows by the Doob maximal inequality (Theorem 2.5.7) that

$$\lim_{m,n\to\infty}\mathbb{E}\Big(\sup_{s\in[0,t]}(X_s^n-X_s^m)^2\Big)=0.$$

Next, for some subsequence n_k and continuous process Y, we have that almost surely $X^{n_k}\to Y$ as $k\to\infty$. Moreover Y inherits the martingale property from X. Now the process

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

is non-decreasing along t_k^n, $1\leq k\leq r_n$. Letting $n\to\infty$ gives that M^2-2Y is almost surely non-decreasing. This shows that $[M]$ exists, is equal to M^2-2Y, and that we can take $\langle M\rangle=[M]$.

- Existence of $\langle M\rangle$ and $[M]$ and their equality when M is not bounded. For all N, we introduce the stopping time $T_N=\inf\{t\geq 0:|M_t|\geq N\}$. From the bounded case applied to the bounded martingale $(M_{t\wedge T_N})_{t\geq 0}$, there exists a unique increasing process $(A_t^N)_{t\geq 0}$ such that $(M_{t\wedge T_n}^2-A_t^N)_{t\geq 0}$ is a martingale. The uniqueness gives $A_{t\wedge T_N}^N=A_t^N$, and the we can define a process $(A_t)_{t\geq 0}$ by setting $A_t=A_t^N$ on the event $\{T_N\geq t\}$. Finally by monotone and dominated converge, $(M_t^2-A_t)_{t\geq 0}$ is a martingale.

  For the quadratic variation, it suffices to write

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

  In contrast with the bounded case, here $A_t=\langle M\rangle_t$ belongs to L^1 but not necessarily to L^2, and in particular, the convergence of $S(\delta^n)=\sum_k(M_{t_k^n}-M_{t_{k-1}^n})^2$ holds in $\mathbb{P}$ but not necessarily in L^2.

- Uniqueness of $\langle M\rangle$. If $(A_t)_{t\geq 0}$ and $(A'_t)_{t\geq 0}$ are continuous, increasing, issued from 0, such that $(M_t^2-A_t)_{t\geq 0}$ and $(M_t^2-A'_t)_{t\geq 0}$ are continuous martingales, then $(A_t-A'_t)_{t\geq 0}$ is a continuous finite variation martingale, and by Lemma 4.1.6, it is constant. Since $A_0=A'_0=0$, we get $A=A'$.

Lemma 4.1.6

If $(M_s)_{s\in[0,t]}$ is a finite variation continuous martingale then it is constant.

Proof of Lemma 4.1.6. Let $(M_s)_{s\in[0,t]}$ 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

$$T_N = t\wedge\inf\{s\in[0,t]:|M_s|\geq N,\sup\sum_k|M_{t_{k+1}}-M_{t_k}|\geq N\}$$

where the supremum runs over all sub-divisions of $[0,t]$. By Theorem 2.5.1, the stopped process $(M_{s\wedge T_N})_{s\in[0,t]}$ is a bounded martingale and thus, for all $s\leq t$,

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

This gives, using a telescoping sum, for an arbitrary sub-division $\delta:0=t_0<\cdots<t_n=t$,

$$\begin{aligned}\mathbb{E}(M^2_{T_N})&=\mathbb{E}(M^2_{t\wedge T_N})-\mathbb{E}(M^2_{0\wedge T_N})\\&=\mathbb{E}\sum_k(M_{t_{k+1}\wedge T_N}-M_{t_k\wedge T_N})^2\\&\leq\mathbb{E}\sup_k|M_{t_{k+1}\wedge T_N}-M_{t_k\wedge T_N}|\sum_k|M_{t_{k+1}\wedge T_N}-M_{t_k\wedge T_N}|\\&\leq N\mathbb{E}\sup_k|M_{t_{k+1}\wedge T_N}-M_{t_k\wedge T_N}|.\end{aligned}$$

Since M is continuous, the sup in the right hand side tends a.s. to 0 as $|\delta|=\max_i(t_{i+1}-t_i)\to0$. Since it is bounded, by dominated convergence, $\mathbb{E}(M^2_{T_N})=0$. Thus $M_{T_N}=0$, which gives in turn $M_t=0$ by sending N to ∞ and using the fact that M is continuous with finite variation. ■

■

Remark 4.1.7: Stochastic integral

In the proof of Theorem 4.1.4, we have approximated $[M]_t$ as M_t^2 minus 2 times a sort of Riemann sum approximating in probability the stochastic integral $\int_0^t M_s\mathrm{d}M_s$ making this approximation and its limit a martingale. This corresponds to the following calculus formula

$$f(M_t)=f(M_0)+\int_0^t f'(M_s)\mathrm{d}M_s+\frac12\int_0^t f''(M_s)\mathrm{d}\langle M\rangle_s$$

in the special case $f(x)=x^2$. This is a remarkable special case of the Itô formula. The quadratic variation term, the second term in the right hand side, is due the roughness of M.

Corollary 4.1.8. Angle bracket, square bracket, quadratic covariation.

Let $M,N\in\mathscr{M}_0^2$.

- There exists a unique continuous finite variation process $\langle M,N\rangle=(\langle M,N\rangle_t)_{t\geq0}$ such that $\langle M,N\rangle_0=0$ and $(M_tN_t-\langle M_t,N_t\rangle)_{t\geq0}$ is a martingale. In particular $\langle M,N\rangle$ is adapted.
- The quadratic covariation of (M,N) exists and $[M,N]_t=\langle M,N\rangle_t$ for all $t\geq0$.

It is important that M and N are martingales with respect to the same filtration, the underlying $(\mathscr{F}_t)_{t\geq0}$. By Theorem 3.1.6, if B is a d-dimensional Brownian motion then for all $1\leq j,k\leq d$ and all $t\geq0$,

$$\langle B^j,B^k\rangle_t=[B^j,B^k]_t=t\mathbf{1}_{j=k}.$$

Proof. We proceed by quadratic polarization. First the processes $(M_t+N_t)_{t\geq0}$ and $(M_t-N_t)_{t\geq0}$ are square integrable continuous martingales with respect to $(\mathscr{F}_t)_{t\geq0}$. Next, for all $t\geq0$, if we define $\langle M,N\rangle_t$ as

$$\langle M,N\rangle_t=\frac14(\langle M+N\rangle_t-\langle M-N\rangle_t),$$

then $M_tN_t - \langle M,N\rangle_t = \frac{1}{4}((M_t+N_t)^2 - \langle M+N\rangle_t - ((M_t-N_t)^2 - \langle M-N\rangle_t))$, and thus $(M_tN_t - \langle M,N\rangle_t)_{t\geq 0}$ is a martingale by Theorem 4.1.4. Moreover $\langle M,N\rangle$ is continuous with finite variation as the difference of continuous and increasing processes. The uniqueness follows as in the proof of Theorem 4.1.4. The link with the quadratic covariation follows by polarization and Theorem 4.1.4. ■

Corollary 4.1.9. Stopped angle brackets.

If $M,N \in \mathscr{M}_0^2$ and S,T are stopping times then $\langle M^T, N^S\rangle = \langle M,N\rangle^{S\wedge T}$.

Proof. Theorem 2.5.1 gives that $(M^2 - \langle M\rangle)^T = (M^T)^2 - \langle M\rangle^T$ is a martingale. Now $(\langle M\rangle^T)_0 = \langle M\rangle_0 = 0$ and $\langle M\rangle^T$ is a continuous increasing process, and thus, by the uniqueness property of the increasing process provided by Theorem 4.1.4, we have $\langle M^T\rangle = \langle M\rangle^T$. By polarization we get $\langle M^T, N^T\rangle = \langle M,N\rangle^T$. Finally, $\langle M^T, N\rangle = \langle M^T, N^T\rangle$ from the equality with quadratic covariation (sum of products of increments). ■

This can be skipped at first reading.

Corollary 4.1.10: Kunita – Watanabe inequality

For all square integrable martingales M and N, the following Cauchy – Schwarz type inequality holds, for all measurable processes φ and ψ and all $t \geq 0$:

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

where the integrals are in the sense of finite variation integrators (Theorem 1.7.3).

Proof. Set $\langle M,N\rangle_s^t = \langle M,N\rangle_t - \langle M,N\rangle_s$, $s \leq t$. The Cauchy – Schwarz inequality on the sums approximating the quadratic variations with rational edges gives, via continuity, that a.s. for all $s < t$,

$$\left|\langle M,N\rangle_s^t\right| \leq \sqrt{\langle M,M\rangle_s^t}\sqrt{\langle N,N\rangle_s^t}.$$

Similarly, we can prove that almost surely

$$\int_s^t |\mathrm{d}\langle M,N\rangle_u| \leq \sqrt{\langle M,M\rangle_s^t}\sqrt{\langle N,N\rangle_s^t}.$$

By a monotone class argument, it follows that almost surely, for all bounded Borel set A,

$$\int_A |\mathrm{d}\langle M,N\rangle_u| \leq \sqrt{\int_A \mathrm{d}\langle M,M\rangle_u}\sqrt{\int_A \mathrm{d}\langle N,N\rangle_u}.$$

Now, almost surely, if $\varphi = \sum_i \lambda_i \mathbf{1}_{A_i}$ and $\psi = \sum_i \mu_i \mathbf{1}_{A_i}$ are step functions with $\lambda_i \geq 0$ and $\mu_i \geq 0$, then

$$\begin{aligned}\int \varphi(s)\psi(s)\,|\langle M,N\rangle_s| &= \sum_i \lambda_i\mu_i \int_{A_i} |\mathrm{d}\langle M,N\rangle_s| \\ &\leq \sqrt{\sum_i \lambda_i^2 \int_{A_i} \mathrm{d}\langle M,M\rangle_s}\sqrt{\sum_i \mu_i^2 \int_{A_i} \mathrm{d}\langle N,N\rangle_s} \\ &= \sqrt{\int \varphi(s)^2 \mathrm{d}\langle M,M\rangle_s}\sqrt{\int \phi(s)^2\mathrm{d}\langle N,N\rangle_s}.\end{aligned}$$

The generalization to arbitrary non-negative measurable φ, ψ follows by monotone convergence. ■

4.2 Local martingales and localization by stopping times

If $(M_t)_{t\geq 0}$ is a martingale, then the Doob stopping theorem states that for every stopping time T, the stopped process $(M_{t\wedge T})_{t\geq 0}$ is again a martingale. Stopping can be used in general to truncate the trajectories of a process with a cutoff, in order to gain more integrability or tightness. Typically if $(X_t)_{t\geq 0}$ is an adapted process, we could consider the sequence of stopping times $(T_n)_{n\geq 0}$ defined by $T_n = \inf\{t \geq 0 : |X_t| \geq n\}$, which satisfies almost surely $T_n \nearrow +\infty$ as $n \to \infty$ and for which for all n the stopped process $(X_{t\wedge T_n})_{t\geq 0}$ is bounded. We say that $(T_n)_{n\geq 0}$ is a localizing sequence. Now a local martingale is simply an adapted processes $(X_t)_{t\geq 0}$ such that for all $n \geq 0$ the stopped process $(X_{t\wedge T_n})_{t\geq 0}$ is a (bounded) martingale. Every martingale is a local martingale. However the converse is false, and strict local martingales do exist.
Local martingales popup naturally when constructing the stochastic integral (see Chapter 7).

Definition 4.2.1. Local martingale.

- A continuous process $(M_t)_{t\geq 0}$ issued from the origin is a local martingale if it is adapted and for all $n \geq 0$, introducing the stopping time $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$, the stopped process $M^{T_n} = (M_{t\wedge T_n})_{t\geq 0}$ is a martingale. It is bounded since $\sup_{t\geq 0}|M_{t\wedge T_n}| \leq |M_0| \vee n = n < \infty$.
- Since the process M is continuous, almost surely $T_n \nearrow +\infty$ as $n \to \infty$, and thus, for all $t \geq 0$, $\lim_{n\to\infty} M_{t\wedge T_n} = M_t$ almost surely. We say that the sequence $(T_n)_{n\geq 0}$ localizes or reduces M.
- If we do not have $M_0 \neq 0$, then we say that M is a local martingale when $M - M_0$ is a local martingale however we still impose that M is adapted and in particular that M_0 is $\mathscr{F}_0$ measurable.
- We denote by $\mathscr{M}^{\mathrm{loc}}$ the set of continuous local martingales w.r.t. the default filtration $(\mathscr{F}_t)_{t\geq 0}$. We denote by $\mathscr{M}_0^{\mathrm{loc}}$ the subset issued from the origin.

Remark 4.2.2. Alternative or relaxed definitions.

Equivalently we could say that a continuous adapted process $(M_t)_{t\geq 0}$ issued from the origin is a local martingale when there exists a sequence $(S_n)_{n\geq 0}$ of stopping times such that

1. almost surely $S_n \nearrow +\infty$ as $n \to \infty$
2. for all $n \geq 1$, the continuous process $M^{S_n} = (M_{t\wedge S_n})_{t\geq 0}$ is a martingale.

Moreover in this definition we could replace martingale by the stronger conditions square integrable martingale, or u.i. martingale, or bounded in L^2 martingale, or bounded martingale. Indeed, it suffices to show that M is then localized by $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$. Indeed, since M is continuous, almost surely $T_n \nearrow +\infty$ as $n \to \infty$. Next, if $(S_n)_{n\geq 0}$ localizes M, then for all $n, k \geq 0$, by the Doob stopping theorem (Theorem 2.5.1) for the martingale M^{S_k} and the stopping time T_n, the process $(M^{S_k})^{T_n} = (M_{t\wedge S_k\wedge T_n})_{t\geq 0}$ is a martingale, thus for all $0 \leq s \leq t$, $\mathbb{E}(M_{t\wedge S_k\wedge T_n} \mid \mathscr{F}_s) = M_{s\wedge S_k\wedge T_n}$. Moreover since $M_0 = 0$ and M is continuous, by definition of T_n, we have $\sup_{t\geq 0}|M_{t\wedge S_k\wedge T_n}| \leq n$, and by dominated convergence, as $k \to \infty$, we have $\mathbb{E}(M_{t\wedge T_n} \mid \mathscr{F}_s) = M_{s\wedge T_n}$, hence $(M_{t\wedge T_n})_{t\geq 0}$ is a martingale.

- Localization is a truncation for processes by cutoff that has the advantage of preserving the continuity of the process and the martingale structure thanks to Doob stopping theorems.
- A martingale is always a local martingale: take $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$ and use Doob stopping (Theorem 2.5.1). Note that thanks to the convention $\inf\varnothing = \infty$ we have $T_n = \infty$ on $\{\sup_{t\geq 0}|M_t| < n\}$.
- If M is a local martingale, then no integrability is guaranteed for M_t for a fixed deterministic $t \geq 0$, and we may have $M_t \not\in \mathrm{L}^1$. Moreover for every stopping time T, the stopped process $M^T = (M_{t\wedge T})_{t\geq 0}$ is a local martingale but the Doob stopping theorem does not hold in general even if T is bounded.

Remark 4.2.3. Domination as a martingale criterion.

If M is a continuous local martingale dominated by an integrable random variable, in the sense that $\mathbb{E}\sup_{t\geq 0}|M_t| < \infty$, then, for all $t \geq 0$ and $s \in [0,t]$, by continuity and dominated convergence,

$$M_s = \lim_{n\to\infty} M_{s\wedge T_n} = \lim_{n\to\infty} \mathbb{E}(M_{t\wedge T_n} \mid \mathscr{F}_s) = \mathbb{E}(\lim_{n\to\infty} M_{t\wedge T_n} \mid \mathscr{F}_s) = \mathbb{E}(M_t \mid \mathscr{F}_s)$$

for any localization sequence $(T_n)_n$ for M, hence M is a u.i. martingale. However, there exists continuous local martingales which are bounded in L^2 and thus u.i. and which are not a martingale!

Remark 4.2.4. Strict local martingales.

Are there local martingales which are not martingales? Yes[a].

- If M is a martingale, for instance Brownian motion, and if U is measurable with respect to $\mathscr{F}_0$, then $(U+M_t)_{t\geq 0}$ is a local martingale, and a martingale if and only if $U \in \mathrm{L}^1$. Note that if M_0 is constant and $\mathscr{F} = \sigma(M_0) = \{\Omega, \varnothing\}$ then necessarily U is constant and we cannot have $U \notin \mathrm{L}^1$.
- Let M be a martingale such that $M_0 = 1$, such as the Doléans-Dade exponential (Theorem 7.3.1). Let U be a random variable independent of M. Then $(UM_t)_{t\geq 0}$ is a local martingale with respect to the enlarged filtration $(\sigma(\sigma(U)\cup\mathscr{F}_t))_{t\geq 0}$, localized by $T_n = \inf\{t\geq 0 : |UM_t| \geq n\}$. This is in fact an Itô stochastic integral, see Exercise 4 of the 2020-2021 exam.
- Let $(B_t)_{t\geq 0}$ be a 3-dimensional BM with $B_0 = x \neq 0$. The process $(|B_t|)_{t\geq 0}$ is a Bessel process. It can be shown that the inverse Bessel process $(|B_t|^{-1})_{t\geq 0}$ is a local martingale, localized by $T_n = \inf\{t\geq 0 : |B_t| \leq 1/(|x|+n)\}$, but is not a martingale. Moreover it is bounded in L^2 and thus u.i.! For a proof, see Exercise 3 of the 2020-2021 exam[b].

[a] Some other famous examples are listed on `https://en.wikipedia.org/wiki/Local_martingale`
[b] Or `https://djalil.chafai.net/blog/2020/10/31/back-to-basics-local-martingales/`

Remark 4.2.5. Vector spaces.

The set $\mathscr{M}^{\mathrm{loc}}$ and $\mathscr{M}_0^{\mathrm{loc}}$ are real vector spaces. Indeed if $M, M' \in \mathscr{M}_0^{\mathrm{loc}}$ are localized respectively by $(T_n)_{n\geq 1}$ and $(T'_n)_{n\geq 1}$, then by the Doob stopping theorem (Theorem 2.5.1), $(S_n)_{n\geq 0} = (T_n \wedge T'_n)_{n\geq 0}$ localizes both M and M'. For all $n \geq 0$, the process $(M+M')^{S_n} = M^{S_n} + M'^{S_n}$ is a square integrable martingale. Note that we have also the following (strict) inclusions:

$$\begin{array}{ccccc} \mathbb{M}_0^2 & \subset & \mathscr{M}_0^2 & \subset & \mathscr{M}_0^{\mathrm{loc}} \\ \cap & & \cap & & \cap \\ \mathbb{M}^2 & \subset & \mathscr{M}^2 & \subset & \mathscr{M}^{\mathrm{loc}} \end{array}$$

Lemma 4.2.6. Increasing process, angle bracket, quadratic variation, square bracket.

Let $M, N \in \mathscr{M}_0^{\mathrm{loc}}$.

1. there exists a unique continuous finite variation process denoted $(\langle M,N\rangle_t)_{t\geq 0}$ with
$$\langle M,N\rangle_0 = 0 \quad \text{and} \quad (M_tN_t - \langle M,N\rangle_t)_{t\geq 0} \in \mathscr{M}_0^{\mathrm{loc}}.$$
Moreover $\langle M,N\rangle = \frac{1}{4}(\langle M+N\rangle - \langle M-N\rangle)$ where $\langle M\rangle = \langle M,M\rangle$.
2. $\langle M\rangle$ is the unique non-decreasing process such that $M^2 - \langle M\rangle$ is a continuous local martingale
3. M is localized by $T_n = \inf\{t\geq 0 : |M_t| \geq n \text{ or } \langle M\rangle_t \geq n\}$ and for all $n \geq 0$,
$$\sup_{t\geq 0}|M_t^{T_n}| \leq n \quad \text{and} \quad \sup_{t\geq 0}\langle M^{T_n}\rangle_t \leq n.$$

4. For all $t \geq 0$ if $(\delta_n)_{n\geq 1}$ is a sequence of sub-divisions of $[0,t]$, $\delta_n : 0 = t_0^n < \cdots < t_{m_n}^n = t$, then

$$S(\delta_n) = \sum_{k=1}^{m_n} (M_{t_k^n} - M_{t_{k-1}^n})(N_{t_k^n} - N_{t_{k-1}^n}) \xrightarrow[n\to\infty]{\mathbb{P}} [M,N]_t = \langle M,N\rangle_t$$

provided that $|\delta_n| = \max_{1\leq k\leq m_n}(t_k^n - t_{k-1}^n) \to 0$ as $n \to \infty$. Furthermore

$$[M,N] = \frac{1}{4}([M+N] - [M-N]) \quad \text{where} \quad [M] = [M,M].$$

We say that $\langle M\rangle$ is the increasing process of M.
We say that $\langle M,N\rangle$ is the finite variation process or angle bracket of the couple (M,N).
We say that $[M]$ is the quadratic variation of M.
We say that $[M,N]$ is the quadratic covariation or square bracket of the couple (M,N).
As for martingales, if $M \in \mathscr{M}^{\text{loc}}$ then we set $\langle M\rangle = \langle M - M_0\rangle$ and $[M] = [M - M_0]$, in particular $\langle M\rangle = [M]$.
As for martingales, $\langle M\rangle_t$ is not necessarily in L^2, and in particular $S(\delta) \to \langle M\rangle$ may not hold in L^2.

Proof.

1. If $(S_n)_{n\geq 0}$ localizes M and $(T_n)_{n\geq 0}$ localizes N then $(U_n)_{n\geq 0} = (T_n \wedge S_n)_{n\geq 0}$ localizes both M and N. By uniqueness of the increasing process of square integrable continuous martingales (Theorem 4.1.4) used for the square integrable martingales M^{U_n} and N^{U_n}, we get that for all $0 \leq n \leq m$ and $t \geq 0$,

$$\langle M^{U_m}, N^{U_m}\rangle_{t\wedge U_n} = \langle M^{U_n}, N^{U_n}\rangle_t,$$

hence $(\langle M^{U_m}, N^{U_m}\rangle)_{t\geq 0}$ and $(\langle M^{U_n}, N^{U_n}\rangle)_{t\geq 0}$ are equal up to time U_n. We then define, for all $t \geq 0$,

$$\langle M,N\rangle_t = \lim_{n\to\infty} \langle M^{U_n}, N^{U_n}\rangle_t.$$

This is the 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 $n \geq 0$, $\langle M,N\rangle_{t\wedge U_n} = \langle M^{U_n}, N^{U_n}\rangle_t$. We then set $\langle M\rangle = \langle M,M\rangle$.

2. Take $M = N$ in the previous item.

3. It suffices to proceed as in Remark 4.2.3. Note that $\langle M^{T_n}\rangle = \langle M\rangle^{T_n}$ gives $|\langle M^{T_n}\rangle| \leq n$.

4. We reduce to $M = N$ by polarization. Next, let $(T_n)_{n\geq 0}$ be a localization sequence for M. For all $n \geq 0$, Theorem 4.1.4 used for the square integrable martingale M^{T_n} gives

$$S^{T_n}(\delta) = \sum_i (M_{t_{i+1}}^{T_n} - M_{t_i}^{T_n}) \xrightarrow[|\delta|\to 0]{\mathrm{L}^2} \langle M^{T_n}\rangle_t = \langle M\rangle_{T_n\wedge t}.$$

Now for all $\varepsilon > 0$ and all $n \geq 0$,

$$\begin{aligned}\mathbb{P}(|S(\delta) - \langle M\rangle_t| > \varepsilon) &\leq \mathbb{P}(T_n \leq t) + \mathbb{P}(|S(\delta) - \langle M\rangle_t| > \varepsilon, t < T_n)\\ &\leq \mathbb{P}(T_n \leq t) + \mathbb{P}(|S^{T_n}(\delta) - \langle M\rangle_{T_n\wedge t}| > \varepsilon),\end{aligned}$$

and therefore $\lim_{|\delta|\to 0} \mathbb{P}(|S(\delta) - \langle M\rangle_t| > \varepsilon) = 0$.

■

Lemma 4.2.7. Martingale criterion.

Let M be a continuous local martingale with $M_0 \in \mathrm{L}^2$.

1. The following properties are equivalent:

   (a) M is a martingale which is square integrable

   (b) $\mathbb{E}(\langle M\rangle_t) < \infty$ for all $t \geq 0$.

2. The following properties are equivalent:
   (a) M is a martingale which is bounded in L^2 and $\sup_{t\geq 0}\mathbb{E}(M_t^2)=\mathbb{E}(\langle M\rangle_\infty)$
   (b) $\mathbb{E}(\langle M\rangle_\infty)<\infty$

   Moreover, in this case $M^2-\langle M\rangle$ is a u.i. martingale and $\mathbb{E}(M_\infty^2)=\mathbb{E}(M_0^2)+\mathbb{E}(\langle M\rangle_\infty)$.

The proof of the lemma is rather short but uses many typical martingale ingredients!

Proof. By replacing M with $M-M_0$, we assume without loss of generality that $M_0=0$.

1. If M is a square integrable martingale then $M^2-\langle M\rangle$ is a martingale and in particular $\langle M\rangle_t\in\mathrm{L}^1$ for all $t\geq 0$. Conversely, if M is a continuous local martingale with $\langle M\rangle_t\in\mathrm{L}^1$ for all $t\geq 0$ then since $M^2-\langle M\rangle$ is a continuous local martingale, it follows that there exists a sequence $(T_n)_{n\geq 0}$ of stopping times such that almost surely $T_n\nearrow+\infty$ as $n\to\infty$ and for all $n\geq 0$ the process $(M^{T_n})^2-\langle M\rangle^{T_n}$ is a square integrable continuous martingale issued from 0. Hence, for all $t\geq 0$, using monotone convergence,
$$\mathbb{E}(M_{t\wedge T_n}^2)=\mathbb{E}(\langle M\rangle_{t\wedge T_n})\underset{n\to\infty}{\longrightarrow}\mathbb{E}(\langle M\rangle_t)<\infty.$$
   This implies that $(M_{t\wedge T_n})_{n\geq 0}$ is bounded in L^2. On the other hand, it follows by the Fatou lemma that
$$\mathbb{E}(M_t^2)=\mathbb{E}\Big(\varliminf_{n\to\infty}M_{t\wedge T_n}^2\Big)\leq\varliminf_{n\to\infty}\mathbb{E}(M_{t\wedge T_n}^2)<\infty.$$
   Finally, since for all $t\geq 0$, $(M_{t\wedge T_n})_{n\geq 0}$ is bounded in L^2, it is u.i., and thus, for all $0\leq s\leq t$, since $\lim_{n\to\infty}M_{t\wedge T_n}=M_t$ a.s., this convergence holds in L^1 and we obtain the martingale property via
$$\mathbb{E}(M_t\mid\mathscr{F}_s)=\mathbb{E}\Big(\lim_{n\to\infty}M_{t\wedge T_n}\mid\mathscr{F}_s\Big)=\lim_{n\to\infty}\mathbb{E}(M_{t\wedge T_n}\mid\mathscr{F}_s)=\lim_{n\to\infty}M_{s\wedge T_n}=M_s.$$

2. If M is a martingale bounded in L^2 then, by Corollary 4.1.5, $\langle M\rangle_\infty\in\mathrm{L}^1$. Conversely, if M is a local martingale with $\langle M\rangle_\infty\in\mathrm{L}^1$, then, by monotony and positivity of $\langle M\rangle$, $\langle M\rangle_t\in\mathrm{L}^1$ for all $t\geq 0$, next, by the first part, M is a square integrable martingale, and thus, by Corollary 4.1.5, M is bounded in L^2.

   Finally if M is a martingale bounded in L^2, then the Doob maximal inequality (Theorem 2.5.7) gives
$$\mathbb{E}\Big(\sup_{s\in[0,t]}M_s^2\Big)\leq 4\mathbb{E}(M_t^2)$$
   for all $t\geq 0$, and by sending t to ∞, we get, by monotone convergence,
$$\mathbb{E}\Big(\sup_{t\geq 0}M_t^2\Big)\leq 4\sup_{t\geq 0}\mathbb{E}(M_t^2).$$
   This gives the domination
$$\sup_{t\geq 0}|M_t^2-\langle M\rangle_t|\leq\sup_{t\geq 0}M_t^2+\langle M\rangle_\infty\in\mathrm{L}^1,$$
   which implies that $M^2-\langle M\rangle$ is uniformly integrable.

■

Remark 4.2.8. Vocabulary.

If X and A are continuous adapted processes with $X_0=A_0=0$ with A of finite variation and such that $X-A$ is a local martingale then A is unique and is called the compensator of X. For instance if X is a continuous local martingale with $X_0=0$ then the compensator of X^2 is $\langle X\rangle$.

Remark 4.2.9. Link with Brownian motion.

The Lévy characterization of Brownian motion (Theorem 7.2.1) states that among all continuous local martingales, Brownian motion is characterized by its angle bracket. On the other hand, the Dubins – Schwarz theorem (Theorem 7.4.1) states that all continuous local martingales with angle bracket tending to infinity at infinity are time changed Brownian motion by the angle bracket.

4.3 Convergence in L^2 and the Hilbert space $\mathbb{M}_0^2$

Let $\mathbb{M}_0^2$ be the set of continuous martingales issued from the origin and bounded in L^2, for some fixed underlying filtered probability space $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq 0},\mathbb{P})$.

The elements of $\mathbb{M}_0^2$ are centered: for all $M\in\mathbb{M}_0^2$ and all $t\geq 0$, $\mathbb{E}(M_t)=\mathbb{E}(M_0)=0$.

For all $M\in\mathbb{M}_0^2$, we have $M_0=0$ and $\sup_{t\geq 0}\mathbb{E}(M_t^2)<\infty$. By Theorem 2.1.3, we see the elements of $\mathbb{M}_0^2$ as random variables taking values in $(\mathscr{C}(\mathbb{R}_+,\mathbb{R}),\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})})$. In particular for all $M,N\in\mathbb{M}_0^2$, we have $M=N$ iff M and N are indistinguishable in other words $\mathbb{P}(\forall t\geq 0: M_t=N_t)=1$. Also $M=0$ iff for all $t\geq 0$, $M_t=0$.

Theorem 4.3.1. Hilbert structure on $\mathbb{M}_0^2$.

The set $\mathbb{M}_0^2$ is a Hilbert space with scalar product $\langle M,N\rangle_{\mathbb{M}_0^2}=\mathbb{E}(\langle M,N\rangle_\infty)$.
Moreover, for all $M\in\mathbb{M}_0^2$, we have $\|M\|^2_{\mathbb{M}_0^2}=\mathbb{E}(\langle M\rangle_\infty)=\sup_{t\geq 0}\mathbb{E}(\langle M\rangle_t)=\sup_{t\geq 0}\mathbb{E}(M_t^2)$.

More generally, it can be shown similarly that for all fixed $T>0$, the set $\mathbb{M}_{0,T}^2$ of square integrable continuous martingales $(M_t)_{t\in[0,T]}$ such that $M_0=0$ is a Hilbert space for the scalar product $\langle M,N\rangle_{\mathbb{M}_{0,T}^2}=\mathbb{E}(\langle M,N\rangle_T)$. In this case, for all $M\in\mathbb{M}_{0,T}^2$, we have $\|M\|^2_{\mathbb{M}_{0,T}^2}=\sup_{t\in[0,T]}\mathbb{E}(\langle M\rangle_t)=\sup_{t\in[0,T]}\mathbb{E}(M_t^2)$.

Proof. The facts that $\mathbb{M}_0^2$ is a vector space and that $\langle\cdot\rangle$ is bilinear, symmetric, and non-negative on the diagonal are almost immediate. For the positivity, if $M\in\mathbb{M}_0^2$ 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 $\mathbb{M}_0^2$. Then for all $\varepsilon>0$, there exists $r\geq 1$ such that for all $m,n\geq r$, $\|M^{(n)}-M^{(m)}\|_{\mathbb{M}_0^2}\leq\varepsilon$. Thus

$$\sup_{t\geq 0}\mathbb{E}(|M_t^{(n)}-M_t^{(m)}|^2)\leq\varepsilon^2.$$

This implies that for all $t\geq 0$, $(M_t^{(n)})_{t\geq 0}$ is a Cauchy sequence in $\mathbb{L}^2$, and thus converges to an element $M_t\in\mathrm{L}^2$. 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 $k\geq 1$,

$$\mathbb{E}(|M_t^{(n_k)}-M_t^{(n_k+1)}|^2)\leq 2^{-k}.$$

Now the Doob maximal inequality (Theorem 2.5.7) for the martingale $(M_t^{(n)}-M_t^{(n+1)})_{t\geq 0}$ gives

$$\mathbb{E}(\sup_{s\in[0,t]}|M_s^{(n_k)}-M_s^{(n_k+1)}|^2)\leq 4\mathbb{E}(|M_t^{(n_k)}-M_t^{(n_k+1)}|^2)\leq 2^{-k+2},$$

and thus, by monotone convergence or the Fubini – Tonelli theorem,

$$\mathbb{E}\Big(\sum_{k\geq 1}\sup_{s\in[0,t]}|M_s^{(n_k)}-M_s^{(n_k+1)}|^2\Big)=\sum_{k\geq 1}\mathbb{E}\Big(\sup_{s\in[0,t]}|M_s^{(n_k)}-M_s^{(n_k+1)}|^2\Big)<\infty.$$

Therefore for all $t>0$, almost surely

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

Lemma 4.3.2. Criterion.

In a Banach space if $\sum_{n=1}^{\infty}\|u_n-u_{n+1}\|<\infty$ then $(u_n)_{n\geq 1}$ converges.

The converse is false, for instance $u_n=\frac{(-1)^n}{n}\xrightarrow[n\to\infty]{}0$ but $|u_n-u_{n+1}|\underset{n\to\infty}{\sim}\frac{2}{n}$ and thus $\sum_{n=1}^{\infty}|u_n-u_{n+1}|=\infty$.

Proof of Lemma 4.3.2. The sequence $(u_n)_{n\geq 1}$ is Cauchy since for all $n\geq 1$ and $m\geq 1$ we have

$$\|u_{n+m}-u_n\|\leq\sum_{k=n}^{n+m-1}\|u_{k+1}-u_k\|\leq\sum_{k\geq n}\|u_{k+1}-u_k\|\xrightarrow[n\to\infty]{}0.$$

■

By using Lemma 4.3.2 with the Banach space $(\mathscr{C}([0,t],\mathbb{R}),\|\cdot\|=\sup_{[0,t]}|\cdot|)$, this implies that for all $t>0$, almost surely, the sequence of continuous functions $(s\in[0,t]\mapsto M_s^{(n_k)})_{k\geq 1}$ converges uniformly towards a limit denoted $(M'_s)_{s\geq[0,t]}$ 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 $t\geq 0$, $(M_t^{(n_k)})_{k\geq 1}$ converges to M_t in L^2 and to M'_t almost surely, and therefore $M_t=M'_t$. ■

Theorem 4.3.3. Convergence of martingales bounded in L^2.

Let M be a square integrable martingale bounded in L^2. Then there exists $M_\infty\in\mathrm{L}^2$ such that

$$\lim_{t\to\infty}M_t=M_\infty \text{ almost surely and in } \mathrm{L}^2.$$

Note that M is uniformly integrable because it is bounded in L^p with $p=2>1$.

Proof. Let us show that M satisfies the L^2 Cauchy criterion. Recall that for all $0\leq s\leq t$, we have

$$\mathbb{E}((M_t-M_s)^2)=\mathbb{E}(M_t^2-2M_s\mathbb{E}(M_t\mid\mathscr{F}_s)+M_s^2)=\mathbb{E}(M_t^2-M_s^2).$$

But M^2 is a sub-martingale and $t\mapsto\mathbb{E}(M_t^2)$ grows and is 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 $M_\infty\in\mathrm{L}^2$. It remains to establish the almost sure convergence. Now, by the Markov inequality, for all $s\geq 0$ and all $\varepsilon>0$,

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\geq s}|M_t-M_\infty|\geq\varepsilon\Big)&\leq\frac{1}{\varepsilon^2}\mathbb{E}\Big(\sup_{t\geq s}(M_t-M_\infty)^2\Big)\\&\leq\frac{2}{\varepsilon^2}\Big(\mathbb{E}\Big((M_s-M_\infty)^2\Big)+\mathbb{E}\Big(\sup_{t\geq s}(M_t-M_s)^2\Big)\Big).\end{aligned}$$

Now the monotone convergence theorem gives

$$\mathbb{E}\Big(\sup_{t\geq s}(M_t-M_s)^2\Big)=\lim_{T\to\infty}\mathbb{E}\Big(\sup_{t\in[s,T]}(M_t-M_s)^2\Big).$$

On the other hand, for all $s\geq 0$, the process $(|M_t-M_s|)_{t\geq s}$ is a continous non-negative sub-martingale, for which the Doob maximal inequality of Theorem 2.5.7 gives

$$\mathbb{E}\Big(\sup_{t\geq s}(M_t-M_s)^2\Big)\leq\lim_{T\to\infty}4\mathbb{E}((M_T-M_s)^2)=4\mathbb{E}((M_\infty-M_s)^2).$$

Therefore we obtain

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

Since the right hand side decreases as s grows, we get, for all $\varepsilon>0$,

$$\mathbb{P}\Big(\cap_{s\in\mathbb{Q}_+}\{\sup_{t\geq s}|M_t-M_\infty|\geq\varepsilon\}\Big)=\lim_{s\to\infty}\mathbb{P}\Big(\sup_{t\geq s}|M_t-M_\infty|\geq\varepsilon\Big)=0,$$

Similarly, the right hand side decreases as ε grows, and then

$$\mathbb{P}\Big(\cup_{\varepsilon\in\mathbb{Q}_+}\cap_{s\in\mathbb{Q}_+}\{\sup_{t\geq s}|M_t-M_\infty|\geq\varepsilon\}\Big)=\lim_{\varepsilon\to 0}\lim_{s\to\infty}\mathbb{P}\Big(\sup_{t\geq s}|M_t-M_\infty|\geq\varepsilon\Big)=0,$$

which means that $\lim_{t\to\infty}M_t=M_\infty$ almost surely! ■

4.4 Convergence in L^1, closedness, uniform integrability

As for the sum of independent and identically distributed random variables, there is, for martingales, in a way, an L^2 theory and an L^1 theory. The L^2 theory is in a sense simpler due to the Hilbert structure.

Theorem 4.4.1. Doob convergence theorem for martingales bounded in L^1.

Let M be a continuous martingale bounded in L^1. Then there exists $M_\infty \in \mathrm{L}^1$ such that

$$\lim_{n\to\infty} M_t = M_\infty \text{ almost surely.}$$

Moreover the convergence holds in L^1 if and only if M is uniformly integrable.

If M is a non-negative martingale, then it is always bounded in L^1.
If M is martingale bounded in L^1 but not u.i., then $\mathbb{E}(M_t) = \mathbb{E}(M_0)$ for all $t \geq 0$ but $\mathbb{E}(M_\infty) \neq \mathbb{E}(M_0)$.

This can be skipped at first reading.

Proof. We can assume that $M_0 = 0$, otherwise consider the martingale $M - M_0 = (M_t - M_0)_{t\geq 0}$ which is also bounded in L^1, making $M_t \to M_0 + (M - M_0)_\infty$ a.s. We proceed by truncation and reduction to the square integrable case. By the Doob maximal inequality (Theorem 2.5.7) with $p = 1$ and all $r > 0$,

$$\mathbb{P}\Big(\sup_{s\in[0,t]} |M_s| \geq r\Big) \leq \frac{\mathbb{E}(|M_t|)}{r}.$$

By monotone convergence, with $C = \sup_{t\geq 0} \mathbb{E}(|M_t|) < \infty$, for all $r > 0$,

$$\mathbb{P}\Big(\sup_{t\geq 0} |M_t| \geq r\Big) \leq \frac{C}{r}.$$

It follows that

$$\mathbb{P}\Big(\sup_{t\geq 0} |M_t| = \infty\Big) \leq \lim_{r\to\infty} \mathbb{P}\Big(\sup_{t\geq 0} |M_t| \geq r\Big) = 0,$$

in other words almost surely $(M_t)_{t\geq 0}$ is bounded: $\sup_{t\geq 0} |M_t| < \infty$. Thus, there exists an almost sure event, say Ω', on which for all $n \geq \sup_{t\geq 0} |M_t|$ (beware that this threshold on n is random),

$$T_n = \inf\{t \geq 0 : |M_t| \geq n\} = \infty.$$

Next, by Doob stopping (Theorem 2.5.1), for all $n \geq 0$, $(M_{t\wedge T_n})_{t\geq 0}$ is a martingale, bounded since $\sup_{t\geq 0} |M_{t\wedge T_n}| \leq |M_0| \vee n = n$ (M is continuous and $M_0 = 0$). Now, since $(M_{t\wedge T_n})_{t\geq 0}$ is bounded in L^2, by Theorem 4.3.3, there exists $M_\infty^{(n)} \in \mathrm{L}^2$ such that $\lim_{t\to\infty} M_{t\wedge T_n} = M_\infty^{(n)}$ almost surely (and in L^2 but this is useless). Let us denote by Ω_n the almost sure event on which this holds. Then, on the almost sure event $\Omega' \cap (\cap_n \Omega_n)$, for all $t \geq 0$ and $n \geq \sup_{s\geq 0} |M_s|$, we have $M_{t\wedge T_n} = M_t$, thus the sequence $(M_\infty^{(n)})_n$ is stationary in the sense that $M_\infty^{(n)}$ is constant when $n \geq \sup_{s\geq 0} |M_s|$, hence, if M_∞ is its limit,

$$\lim_{t\to\infty} M_t = M_\infty.$$

Contrary to $M_\infty^{(n)}$, the limit M_∞ has not reason to belong to L^2. However $M_\infty \in \mathrm{L}^1$ since from the almost sure convergence, the boundedness in L^1 of $(M_t)_{t\geq 0}$, and by using the Fatou lemma, we have

$$\mathbb{E}(|M_\infty|) = \mathbb{E}(\varliminf_{t\to\infty} |M_t|) \leq \varliminf_{t\to\infty} \mathbb{E}(|M_t|) \leq C < \infty.$$

Finally an almost sure convergence to an L^1 limit holds in L^1 if and only if the sequence is u.i. ■

The result remains valid for super-martingales.

Theorem 4.4.2: Doob convergence theorem for super-martingales bounded in L^1

Let M be a continuous super-martingale bounded in L^1. Then there exists $M_\infty \in \mathrm{L}^1$ such that

$$\lim_{n\to\infty} M_t = M_\infty \text{ almost surely.}$$

Note that a non-negative super-martingale is automatically bounded in L^1.

Proof. See for instance [31, Theorem 3.19 page 58–59] for a classical proof using oscillations. ■

Remark 4.4.3. Non-negative local martingales are super-martingales.

If $(M_t)_{t\geq 0}$ is a non-negative continuous local martingale and $M_0 \in \mathrm{L}^1$, then it is a non-negative super-martingale and by Theorem 4.4.2 it converges almost surely to an integrable random variable. Indeed, if $(T_n)_n$ is a localizing sequence then for all $t \geq 0$ and $s \in [0,t]$, by the Fatou Lemma,

$$\mathbb{E}(M_t \mid \mathscr{F}_s) = \mathbb{E}(\lim_{n\to\infty} M_{t\wedge T_n} \mid \mathscr{F}_s) \leq \lim_{n\to\infty} \mathbb{E}(M_{t\wedge T_n} \mid \mathscr{F}_s) = \lim_{n\to\infty} M_{s\wedge T_n} = M_s.$$

Note that the conditional expectations are well defined in $[0,+\infty]$ because M is non-negative.

Corollary 4.4.4. Convergence of martingales bounded in L^p, $p>1$.

If M is a continuous martingale bounded in L^p with $p>1$ then there exists $M_\infty \in \mathrm{L}^p$ such that

$$\lim_{t\to\infty} M_t = M_\infty \text{ almost surely and in } \mathrm{L}^p.$$

In particular, for $p=2$ this gives an alternative to Theorem 4.3.3.

This can be skipped at first reading.

Proof. Since M is a super-martingale bounded in L^1, Theorem 4.4.1 gives $M_\infty \in \mathrm{L}^1$ such that $\lim_{t\to\infty} M_t = M_\infty$ almost surely. But since M is bounded in L^p with $p>1$, it follows that M is uniformly integrable, and therefore $\lim_{t\to\infty} M_t = M_\infty$ in L^1. We have $M_\infty \in \mathrm{L}^p$ since by the Fatou lemma,

$$\mathbb{E}(|M_\infty|^p) = \mathbb{E}(\underline{\lim}_{t\to\infty} |M_t|^p) \leq \underline{\lim}_{t\to\infty} \mathbb{E}(|M_t|^p) < \infty.$$

On the other hand, by the Doob maximal inequality (Theorem 2.5.7), since M is bonded in L^p, for all $t\geq 0$, $\sup_{s\in[0,t]}|M_s|^p \in \mathrm{L}^1$ and $\mathbb{E}(\sup_{s\in[0,t]}|M_s|^p) \leq c_p\mathbb{E}(|M_t|^p)$. Therefore, by monotone convergence,

$$\mathbb{E}\Big(\sup_{t\geq 0}|M_t|^p\Big) \leq \sup_{t\geq 0}\mathbb{E}(|M_t|^p) < \infty.$$

Hence $\sup_{t\geq 0}|M_t|^p \in \mathrm{L}^1$, and thus, by dominated convergence, $\lim_{t\to\infty} M_t = M_\infty$ in L^p. ■

Corollary 4.4.5. Doob theorem on closed martingales or Doob martingale convergence theorem.

Let M 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 $M_\infty \in \mathrm{L}^1$ such that for all $t\geq 0$, $M_t = \mathbb{E}(M_\infty \mid \mathscr{F}_t)$
3. (integrability) the family $\{M_t : t\geq 0\}$ is uniformly integrable.

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

If M is a martingale then for all fixed $a \geq 0$, the stopped martingale $M^a = (M_{t\wedge a})_{t\geq 0}$ is closed by M_a since $\mathbb{E}(M_a \mid \mathscr{F}_t) = M_a \mathbf{1}_{a\leq t} + M_t \mathbf{1}_{a>t} = M_{t\wedge a}$. Hence M^a is uniformly integrable. Note that $\lim_{t\to\infty} M_{t\wedge a} = M_a$.

Note that in the proof below, Theorem 4.4.1 is used in every implication of the equivalence.

This can be skipped at first reading.

Proof. 1. ⇒ 2. If M converges in L^1, then it is bounded in L^1, and by Theorem 4.4.1, its converges a.s. to $M_\infty \in \mathrm{L}^1$ (the convergence holds also in L^1 but we do not use this fact now). For all $t \geq 0$ and $s \in [0,t]$ and all $A \in \mathscr{F}_s$, the martingale property for M gives $\mathbb{E}(M_t \mathbf{1}_A) = \mathbb{E}(M_s \mathbf{1}_A)$. By dominated convergence as $t \to \infty$, we get $\mathbb{E}(M_\infty \mathbf{1}_A) = \mathbb{E}(M_s \mathbf{1}_A)$ therefore $M_s = \mathbb{E}(M_\infty \mid \mathscr{F}_s)$ for all $s \geq 0$.
2. ⇒ 3. Let us assume that for some $M_\infty \in \mathrm{L}^1$ we have $M_t = \mathbb{E}(M_\infty \mid \mathscr{F}_t)$ for all $t \geq 0$. Then $\sup_{t\geq 0} \mathbb{E}(|M_t|) \leq \mathbb{E}(|M_\infty|) < \infty$ and thus, by Theorem 4.4.1, M_t converges a.s. as $t \to \infty$. It follows that almost surely $M_* = \sup_{t\geq 0} |M_t| < \infty$. Now $\lim_{R\to\infty} \mathbf{1}_{M_*\geq R} = 0$ almost surely, and for all $R \geq 0$,

$$\sup_{t\geq 0} \mathbb{E}(|M_t| \mathbf{1}_{|M_t|\geq R}) = \sup_{t\geq 0} \mathbb{E}(|\mathbb{E}(M_\infty \mid \mathscr{F}_t)| \mathbf{1}_{|M_t|\geq R}) \leq \sup_{t\geq 0} \mathbb{E}(|M_\infty| \mathbf{1}_{|M_t|\geq R}) \leq \mathbb{E}(|M_\infty| \mathbf{1}_{M_*\geq R}) \underset{R\to\infty}{\longrightarrow} 0$$

where the convergence follows by dominated convergence. Therefore M is u.i.
3. ⇒ 1. If M is u.i. then it is bounded in L^1, and from Theorem 4.4.1, there exists $M_\infty \in \mathrm{L}^1$ such that $\lim_{t\to\infty} M_t = M_\infty$ a.s. Since M is u.i., the convergence holds in L^1. ■

The following generalizes the Doob stopping theorem (Theorem 2.5.1).

Corollary 4.4.6. Doob stopping for uniformly integrable martingales.

Let M be a u.i. continuous martingale and let T be a stopping time (not necessarily bounded or finite). We set $M_T = M_\infty$ on $\{T = \infty\}$ where $M_\infty = \lim_{t\to\infty} M_t$ is as in Corollary 4.4.5. Then:

1. $(M_{t\wedge T})_{t\geq 0}$ is a uniformly integrable martingale, $M_T \in \mathrm{L}^1$, and for all $t \geq 0$, $M_{t\wedge T} = \mathbb{E}(M_T \mid \mathscr{F}_t)$.
   In particular, for all $t \geq 0$, $\mathbb{E}(M_0) = \mathbb{E}(M_{t\wedge T}) = \mathbb{E}(M_T)$.
2. Moreover if S is another stopping time with $S \leq T$ then $M_S = \mathbb{E}(M_T \mid \mathscr{F}_S)$.
   In particular, for all stopping time S, $M_S = \mathbb{E}(M_\infty \mid \mathscr{F}_S)$ and $\mathbb{E}(M_S) = \mathbb{E}(M_\infty) = \mathbb{E}(M_0)$.

This can be skipped at first reading.

Proof. We will prove the first property by using the second property.

1. For all $t \geq 0$, both $t \wedge T$ and T are stopping times. By the second property of the present theorem, $M_{t\wedge T} \in \mathrm{L}^1$ and $M_T \in \mathrm{L}^1$. Moreover $M_{t\wedge T}$ is measurable for $\mathscr{F}_{t\wedge T}$, and thus for $\mathscr{F}_t$ since $t \leq t \wedge T$. Now, in order to prove that $\mathbb{E}(M_T \mid \mathscr{F}_t) = M_{t\wedge T}$, it suffices to show that for all $A \in \mathscr{F}_t$,
   $$\mathbb{E}(\mathbf{1}_A M_T) = \mathbb{E}(\mathbf{1}_A M_{t\wedge T}).$$
   But for all $A \in \mathscr{F}_t$, we have immediately from $T = t \wedge T$ on $\{T \leq t\}$ that
   $$\mathbb{E}(\mathbf{1}_{A\cap\{T\leq t\}} M_T) = \mathbb{E}(\mathbf{1}_{A\cap\{T\leq t\}} M_{t\wedge T}).$$
   The second property of the present theorem for the stopping times $S = t \wedge T$ and T gives
   $$M_{t\wedge T} = \mathbb{E}(M_T \mid \mathscr{F}_{t\wedge T}).$$
   Now since $A \cap \{T > t\} \in \mathscr{F}_t$ and $A \cap \{T > t\} \in \mathscr{F}_T$, we get $A \cap \{T > t\} \in \mathscr{F}_t \cap \mathscr{F}_T = \mathscr{F}_{t\wedge T}$, and
   $$\mathbb{E}(\mathbf{1}_{A\cap\{T>t\}} M_T) = \mathbb{E}(\mathbf{1}_{A\cap\{T>t\}} M_{t\wedge T}).$$
   By adding this to a previous formula we get the desired result $\mathbb{E}(\mathbf{1}_A M_T) = \mathbb{E}(\mathbf{1}_A M_{t\wedge T})$.
   Finally, the fact that $M^T = (M_{t\wedge T})_{t\geq 0}$ is a martingale follows from what precedes used with the u.i. martingale $M^a = (M_{t\wedge a})_{t\geq 0}$ for all $a \geq 0$, which gives $M^a_{s\wedge T} = \mathbb{E}(M^a_T \mid \mathscr{F}_s)$ for all $s \geq 0$, in other words $M_{s\wedge a\wedge T} = \mathbb{E}(M_{a\wedge T} \mid \mathscr{F}_s)$. Taking $a = t \geq s$ gives the martingale property for M^T.

2. Following for instance [31, Theorem 3.22 page 59], we discretize as in the proof of Theorem 2.5.1 or Theorem 3.5.1. Namely, for all $n \geq 0$, let us define the stopping times

$$S_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{k2^{-n} < S \leq (k+1)2^{-n}} + \infty \mathbf{1}_{S=\infty} \quad \text{and} \quad T_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{k2^{-n} < T \leq (k+1)2^{-n}} + \infty \mathbf{1}_{T=\infty}.$$

We have $S_n \searrow S$ and $T_n \searrow T$ as $n \to \infty$, and $S_n \leq T_n$ for all $n \geq 0$. Next, for all $n \geq 0$, $2^n S_n$ and $2^n T_n$ are integer valued stopping times for the discrete time filtration $(\mathscr{F}_k^{(n)})_{k\geq 0} = (\mathscr{F}_{k2^{-n}})_{k\geq 0}$, while $M^{(n)} = (M_{k2^{-n}})_{k\geq 0}$ is a uniformly integrable discrete time martingale with respect to this filtration. By using the Doob stopping theorem for u.i. discrete time martingales, we get

$$M_{S_n} = M^{(n)}_{2^n S_n} = \mathbb{E}(M^{(n)}_{2^n T_n} \mid \mathscr{F}^{(n)}_{2^n S_n}) = \mathbb{E}(M_{T_n} \mid \mathscr{F}_{S_n}).$$

Now, for all $A \in \mathscr{F}_S \subset \mathscr{F}_{S_n}$, we have $\mathbb{E}(\mathbf{1}_A M_{S_n}) = \mathbb{E}(\mathbf{1}_A M_{T_n})$. Since M is (right) continuous, a.s.

$$M_S = \lim_{n\to\infty} M_{S_n} \quad \text{and} \quad M_T = \lim_{n\to\infty} M_{T_n}.$$

For the L^1 convergence, the Doob stopping theorem for u.i. discrete time martingales gives $M_{S_n} = \mathbb{E}(M_\infty \mid \mathscr{F}_{S_n})$ for all $n \geq 0$ and thus $(M_{S_n})_{n\geq 0}$ and $(M_{T_n})_{n\geq 0}$ are u.i. This also gives that $M_S \in \mathrm{L}^1$ and $M_T \in \mathbb{L}^1$. This also allows to pass to the limit in $\mathbb{E}(\mathbf{1}_A M_{S_n}) = \mathbb{E}(\mathbf{1}_A M_{T_n})$ to get $\mathbb{E}(\mathbf{1}_A M_S) = \mathbb{E}(\mathbf{1}_A M_T)$. This holds for all $A \in \mathscr{F}_S$, thus $M_S = \mathbb{E}(M_T \mid \mathscr{F}_S)$.

■

Chapter 5

Itô stochastic integral with respect to Brownian motion

This chapter is purely pedagogical can be completely bypassed if time is very limited.

In this chapter, we construct a stochastic integral which goes beyond the Wiener integral of Chapter 3, with an integrator $(B_t)_{t\geq 0}$ which is Brownian motion and with an integrand $(\varphi_t)_{t\geq 0}$ which can be random and square integrable. The ambition is thus to define the process

$$\Big(\int_0^t \varphi_s \mathrm{d}B_s\Big)_{t\geq 0}.$$

Following Itô, we start with a finite sum when φ 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 when the integrand φ is deterministic.

5.1 Itô versus Stratonovich stochastic integrals in a nutshell

Let $(B_t)_{t\geq 0}$ be a one-dimensional BM issued from the origin. Let us try to give a meaning to

$$\int_0^t B_s \mathrm{d}B_s.$$

This is not a Wiener integral since the integrand is random. Using an approximation with step processes, we have, rearranging terms to produce telescoping summation or already known sums,

$$\begin{aligned}\sum_i B_{t_i}(B_{t_{i+1}} - B_{t_i}) &= \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2 - B_{t_{i+1}}^2 + B_{t_i}B_{i_{i+1}}) \\ &= \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2 - B_{t_{i+1}}(B_{t_{i+1}} - B_{t_i})) \\ &= \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2) - \sum_i B_{t_i}(B_{t_{i+1}} - B_{t_i}) - \sum_i (B_{t_{i+1}} - B_{t_i})^2\end{aligned}$$

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

$$\int_0^t B_s \mathrm{d}B_s = B_t^2 - \int_0^t B_s \mathrm{d}B_s - [B]_t,$$

where $[B]_t$ is the quadratic variation of B on $[0,t]$ (Theorem 3.2.1). Since $[B]_t = t$ (Theorem 3.2.1 again),

$$\int_0^t B_s \mathrm{d}B_s = \frac{B_t^2 - t}{2}.$$

The process $(\frac{1}{2}(B_t^2 - t))_{t\geq 0}$ is a centered martingale. The term $-\frac{1}{2}t$ is the martingale correction to the differential calculus term $\frac{1}{2}B_t^2$. Taking the value at the left-end time intervals in the Riemann sum produces a centered martingale and a stochastic integral which is a centered martingale. This is the Itô[1] stochastic integral. Indeed this is confirmed by the rigorous construction in this chapter.

[1] Named after Kiyosi Itô (1915–2008), Japanese mathematician. He used the notation “Kiyosi Itô” for his name (Kunrei-shiki romanization), instead of the more standard “Kiyoshi Itō” (Hepburn romanization).

Let us examine the notion of Stratonovich[2] stochastic integral which corresponds to take the mean value of 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 to distinguish it from the Itô stochastic integral,

$$\int_0^t B_s \circ \mathrm{d}B_s = \frac{1}{2}B_t^2.$$

This time the rule of differential calculus is satisfied, but the result has no reason to be 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 $2(\mathrm{Stratonovich}_t - \frac{1}{2}\mathrm{Itô}_t) = B_t^2 + \frac{1}{2}t$.

Coding in action 5.1.1. Numerical stochastic integration.

Could you write a code approximating by simulation the sums related to the integral $\int_0^t B_s \mathrm{d}B_s$ given in Section 5.1, and plotting the resulting process?

5.2 Itô stochastic integral with respect to Brownian motion

Let $B=(B_t)_{t\geq 0}$ be a d-dimensional $(\mathscr{F}_t)_{t\geq 0}$ Brownian motion issued from the origin. In this section we construct rigorously the Itô stochastic integral with integrator B, namely, with Section 5.1 in mind,

$$\int_0^t \varphi_s \mathrm{d}B_s$$

where φ is some sort of square integrable stochastic process, that can be at least taken equal to B, say.

A d-dimensional step process $(\varphi_t)_{t\geq 0}$ is a process for which there exists $0 \leq t_0 \leq \cdots \leq t_n$, $n \geq 1$, and bounded random variables $U_0, \ldots, U_{n-1}$ which are $\mathscr{F}_{t_0}, \ldots, \mathscr{F}_{t_{n-1}}$ measurable respectively, with for all $t \geq 0$,

$$\varphi_t = U_0 \mathbf{1}_0(t) + \sum_{i=0}^{n-1} U_i \mathbf{1}_{(t_i, t_{i+1}]}(t).$$

Such a step process is progressive (measurable with respect to $\mathscr{P}$), left-continuous, and on each time interval, the random value is measurable with respect to the σ-algebra which corresponds to the left end time of the interval[3]. The vector space of step processes is denoted $\mathscr{S}_{\mathbb{R}^d}$. We have $\mathscr{S}_{\mathbb{R}^d} \subset \mathscr{L}^2_{\mathbb{R}^d}$ where

$$\mathscr{L}^2_{\mathbb{R}^d} = \mathrm{L}^2_{\mathbb{R}^d}(\Omega \times \mathbb{R}_+, \mathscr{P}, \mathbb{P} \otimes \mathrm{d}t)$$

is the Hilbert space of d-dimensional processes $(\varphi_t)_{t\geq 0}$ which are progressive with respect to $(\mathscr{F}_t)_{t\geq 0}$ and

$$\mathbb{E}\Big(\int_0^\infty |\varphi_s|^2 \mathrm{d}s\Big) = \int_0^\infty \mathbb{E}(|\varphi_s|^2)\mathrm{d}s = \|\varphi\|^2_{\mathrm{L}^2_{\mathbb{R}^d}(\Omega\times\mathbb{R}_+)} < \infty.$$

[2]Named after Ruslan Leont'evich Stratonovich (1930–1997), Russian physicist, engineer, and probabilist.
[3]Hence the name "predictable" which is used sometimes.

Lemma 5.2.1. Approximation or density.

The set $\mathscr{S}_{\mathbb{R}^d}$ is dense in $\mathscr{L}^2_{\mathbb{R}^d}$, namely for all $\varphi\in\mathscr{L}^2_{\mathbb{R}^d}$ and all $\varepsilon>0$, there exists $\psi\in\mathscr{S}_{\mathbb{R}^d}$ such that

$$\mathbb{E}\Big(\int_0^\infty|\varphi_s-\psi_s|^2\mathrm{d}s\Big)<\varepsilon.$$

Proof. We can assume that φ is bounded, since by dominated convergence,

$$\lim_{n\to\infty}\mathbb{E}\Big(\int_0^\infty|\varphi_s-\varphi_s\mathbf{1}_{\varphi_s\in[-n,n]}|^2\mathrm{d}s\Big)=0.$$

We can moreover assume that φ vanishes outside a finite time interval since

$$\lim_{n\to\infty}\mathbb{E}\Big(\int_0^\infty|\varphi_s-\varphi_s\mathbf{1}_{s\in[0,n]}|^2\mathrm{d}s\Big)=0.$$

We can assume furthermore that such a process is (left-)continuous since

$$\lim_{n\to\infty}\mathbb{E}\Big(\int_0^\infty\Big|\varphi_s-\Big(n\int_{s-\frac{1}{n}}^s\varphi_u\mathrm{d}u\Big)\mathbf{1}_{s>\frac{1}{n}}\Big|^2\mathrm{d}s\Big)=0.$$

Finally it suffices to approximate such a process with elements of $\mathscr{S}_{\mathbb{R}^d}$, namely

$$\lim_{n\to\infty}\mathbb{E}\Big(\int_0^\infty\Big|\varphi_s-\sum_{i=0}^\infty\varphi_{\frac{i}{n}}\mathbf{1}_{t\in\left(\frac{i}{n},\frac{i+1}{n}\right]}\Big|^2\mathrm{d}s\Big)=0$$

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

We denote by $\mathscr{M}^2$ be the set of continuous square integrable martingales for $(\mathscr{F}_t)_{t\geq0}$. The elements of $\mathscr{M}^2$ are seen as random variables on the Wiener space $(\mathscr{C}(\mathbb{R}_+,\mathbb{R}),\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R})})$. Two elements X and Y of $\mathscr{M}^2$ are equal iff they are indistinguishable processes: $\mathbb{P}(\forall t\geq0:X_t=Y_t)=1$.

Theorem 5.2.2. Brownian Itô stochastic integral of square integrable progressive processes.

There exists a unique linear map $I:\mathscr{L}^2\to\mathscr{M}^2$ denoted for all $\varphi\in\mathscr{L}^2_{\mathbb{R}^d}$ and all $t\geq0$ as

$$I(\varphi)_t=\int_0^t\varphi_s\mathrm{d}B_s,$$

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

1. (step processes) for all $\varphi\in\mathscr{S}_{\mathbb{R}^d}$ with decomposition $\varphi=U_0\mathbf{1}_0+\sum_{i=0}^{n-1}U_i\mathbf{1}_{(t_i,t_{i+1}]}$, for all $t\geq0$,

$$\int_0^t\varphi_s\mathrm{d}B_s=\sum_{i=0}^{n-1}U_i\cdot(B_{t_{i+1}\wedge t}-B_{t_i\wedge t})$$

2. (Itô isometry) for all $\varphi\in\mathscr{L}^2_{\mathbb{R}^d}$ and all $t\geq0$,

$$\mathbb{E}\Big(\Big(\int_0^t\varphi_s\mathrm{d}B_s\Big)^2\Big)=\mathbb{E}\Big(\int_0^t|\varphi_s|^2\mathrm{d}s\Big).$$

Moreover this linear isometry satisfies, for all $\varphi,\psi\in\mathscr{L}^2_{\mathbb{R}^d}$, $I(\varphi)_0=0$ and for all $t\geq0$,

$$\mathbb{E}\Big(\int_0^t\varphi_s\mathrm{d}B_s\Big)=0\quad\text{and}\quad\Big\langle\int_0^\bullet\varphi_s\mathrm{d}B_s,\int_0^\bullet\psi_s\mathrm{d}B_s\Big\rangle_t=\int_0^t\varphi_s\cdot\psi_s\mathrm{d}s.$$

Furthermore $I(\varphi)$ coincides with the Wiener integral of Chapter 3 when φ is not random.

Note that $\int_0^t \varphi_s \mathrm{d}B_s$ is centered even if φ_t is not. This comes from progressivity of φ and centering of B.

Theorem 5.2.2 provides plenty of continuous martingales from Brownian motion. Moreover contrary to Wiener integrals, the Itô stochastic integral may produce martingales with non-deterministic bracket.

Proof. Uniqueness. Let I and I' be two linear maps from $\mathscr{L}^2_{\mathbb{R}^d}$ to $\mathscr{M}^2$ satisfying the two properties of the theorem. Let $\varphi \in \mathscr{L}^2_{\mathbb{R}^d}$. From Lemma 5.2.1, for all $n \geq 1$, there exists $\psi^{(n)} \in \mathscr{S}_{\mathbb{R}^d}$ such that $\lim_{n\to\infty} \psi^{(n)} = \varphi$ in $\mathscr{L}^2_{\mathbb{R}^d}$. The linearity and isometry properties of I and I' and the equality of I and I' on $\mathscr{S}_{\mathbb{R}^d}$ give

$$I(\varphi)_t = \overset{\mathrm{L}^2}{\lim_{n\to\infty}} I(\psi_n)_t = \overset{\mathrm{L}^2}{\lim_{n\to\infty}} I'(\psi_n)_t = I'(\varphi)_t.$$

Therefore $I(\varphi)$ and $I'(\varphi)$ are modifications of each other: $\forall t \geq 0$, $\mathbb{P}(X_t = Y_t) = 1$, and since they are continuous, they are indistinguishable, in other words $I = I'$ in $\mathscr{M}^2$.

Properties for step processes. Let $\varphi \in \mathscr{S}_{\mathbb{R}^d}$. The process $(I(\varphi)_t)_{t\geq 0}$ satisfies $I(\varphi)_0 = 0$ and is continuous since B is continuous. Moreover, by construction $I(\varphi)_t$ is the (finite) sum of centered bounded random variables measurable with respect to $\mathscr{F}_t$. Note that here it is crucial that U_i is $\mathscr{F}_{t_i}$ measurable for all i. Furthermore, for all $0 \leq s \leq t$, if $t \in (t_k, t_{k+1}]$ and if $s \in (t_{k'}, t_{k'+1}]$, $k' \leq k$, the decomposition

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

gives, using the conditional independence of the increments of B and the measurability assumptions on U,

$$\mathbb{E}(I(\varphi)_t \mid \mathscr{F}_s) = I(\varphi)_s + 0.$$

The zero comes from the fact that for all $v \geq u \geq s$, $\mathbb{E}(U \cdot (B_v - B_u) \mid \mathscr{F}_s) = \mathbb{E}(U \cdot \mathbb{E}(B_v - B_u \mid \mathscr{F}_u) \mid \mathscr{F}_s) = 0$ when U is bounded and $\mathscr{F}_u$-measurable, since $B_v - B_u$ is conditionally independent of $\mathscr{F}_u$. Hence $X = I(\varphi) \in \mathscr{M}^2$. Let us now establish the formula for the increasing process $\langle X \rangle$. Let us show that for all $0 \leq s < t$ and $A \in \mathscr{F}_s$,

$$\mathbb{E}\Big(\mathbf{1}_A\Big(X_t^2 - X_s^2 - \int_s^t |\varphi_u|^2 \mathrm{d}u\Big)\Big) = 0.$$

Since X is a square integrable martingale, we have[4]

$$\mathbb{E}(\mathbf{1}_A(X_t^2 - X_s^2)) = \mathbb{E}(\mathbf{1}_A(X_t^2 - 2X_tX_s + X_s^2)) = \mathbb{E}(\mathbf{1}_A(X_t - X_s)^2).$$

Now, with $\varphi_t = U_0 \mathbf{1}_0(t) + \sum_i U_i \mathbf{1}_{(t_i, t_{i+1}]}(t)$, $\Delta_{u,v} = B_v - B_u$, and $t_0 = 0$, we have

$$X_t = U_0 \cdot \Delta_{t_0,t_1} + U_1 \cdot \Delta_{t_1,t_2} + \cdots + U_k \cdot \Delta_{t_k,t} \text{ if } t \in (t_k, t_{k+1}].$$

Since $s < t$, there exists $\ell \leq k$ such that $s \in (t_\ell, t_{\ell+1}]$, and thus

$$\mathbf{1}_A(X_t - X_s) = \sum_{i=\ell}^{k} \widetilde{U}_i \cdot \Delta_{s_i,s_{i+1}} \quad \text{with} \quad s_i := \begin{cases} s & \text{if } i = \ell \\ t_i & \text{if } i \in \{\ell+1, \ldots, k\} \\ t & \text{if } i = k+1 \end{cases}.$$

Now for all $i, j \in \{\ell, \ell+1, \ldots, k\}$ with $i \leq j$, since $\widetilde{U}_i$ is $\mathscr{F}_{s_i}$ measurable, since $\Delta_{s_j,s_{j+1}}$ is conditionally independent of $\mathscr{F}_{s_j}$, with mean 0 and variance $s_{j+1} - s_j$, we get

$$\mathbb{E}((\widetilde{U}_i \cdot \Delta_{s_i,s_{i+1}})(\widetilde{U}_j \cdot \Delta_{s_j,s_{j+1}})) = \begin{cases} \mathbb{E}(\mathbb{E}((\widetilde{U}_i \cdot \Delta_{s_i,s_{i+1}})^2 \mid \mathscr{F}_{s_i})) = \mathbb{E}(|\widetilde{U}_i|^2)(s_{i+1} - s_i) & \text{if } i = j \\ \mathbb{E}((\widetilde{U}_i \cdot \Delta_{s_i,s_{i+1}})(\widetilde{U}_j \cdot \mathbb{E}(\Delta_{s_j,s_{j+1}} \mid \mathscr{F}_{s_j}))) = 0 & \text{if } i < j \end{cases}.$$

It follows then that

$$\mathbb{E}(\mathbf{1}_A(X_t - X_s)^2) = \sum_{i,j=\ell}^{k} \mathbb{E}((\widetilde{U}_i \cdot \Delta_{s_i,s_{i+1}})(\widetilde{U}_j \cdot \Delta_{s_j,s_{j+1}})) = \mathbb{E}\Big(\mathbf{1}_A \sum_{i=\ell}^{k} |U_i|^2(s_{i+1} - s_i)\Big) = \mathbb{E}\Big(\mathbf{1}_A \int_s^t |\varphi_u|^2 \mathrm{d}u\Big).$$

[4] In fact Pythagoras theorem in $\mathrm{L}^2(\Omega, \mathscr{F}_s, \mathbb{P})$: $\mathbb{E}(X_t^2 \mid \mathscr{F}_s) = \mathbb{E}((X_t - X_s)^2 \mid \mathscr{F}_s) + \mathbb{E}(X_s^2 \mid \mathscr{F}_s)$.

Existence. Let $\varphi \in \mathscr{L}^2_{\mathbb{R}^d}$. From Lemma 5.2.1, for all $n \geq 1$, there exists $\psi^{(n)} \in \mathscr{S}_{\mathbb{R}^d}$ with

$$\mathbb{E}\Big(\int_0^\infty |\varphi_s - \psi_s^{(n)}|^2 \mathrm{d}s\Big) \leq \frac{1}{2^n}.$$

For all $t \geq 0$, we set $X_t^{(n)} = I(\psi^{(n)})_t$. By using the linearity and isometry of I on $\mathscr{S}_{\mathbb{R}^d}$, we get, for all n and t,

$$\mathbb{E}(|X_t^{(n)} - X_t^{(n+1)}|^2) = \mathbb{E}\Big(\int_0^t |\psi_s^{(n)} - \psi_s^{(n+1)}|^2 \mathrm{d}s\Big) \leq \frac{4}{2^n}.$$

Next, the Doob maximal inequality of Theorem 2.5.7 gives

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

Therefore,

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

and thus, almost surely,

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

By using Lemma 4.3.2 with the Banach space $(\mathscr{C}([0,t],\mathbb{R}), \|\cdot\| = \sup_{[0,t]}|\cdot|)$, it follows that almost surely, the sequence $(X^{(n)})_n$ of continuous martingales converges uniformly on every finite interval of $\mathbb{R}_+$, as $n \to \infty$, towards a continuous process $X = (X_t)_{t\geq 0}$. This process is a martingale since for all $0 \leq s < t$ and all $A \in \mathscr{F}_s$,

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

The process X depends only on φ 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 $t \geq 0$, $\lim_{n\to\infty} X_t^{(n)} = X_t$ in L^2, in particular $\mathbb{E}(X_t) = 0$ since $X_t^{(n)}$ is centered for all n, while

$$\mathbb{E}(X_t^2) = \lim_{n\to\infty}\mathbb{E}((X_t^{(n)})^2) = \lim_{n\to\infty}\mathbb{E}\Big(\int_0^t |\psi_s^{(n)}|^2\mathrm{d}s\Big) = \mathbb{E}\Big(\int_0^t |\varphi_s|^2\mathrm{d}s\Big).$$

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

Additional properties. Since almost surely the above convergence holds uniformly on $[0,t]$ and since $X_0^n = 0$ for all $n \geq 1$ it follows that $X_0 = 0$. It remains to establish the formula for $\langle X\rangle$, and for that it suffices to show that for all $0 \leq s \leq t$ and all $A \in \mathscr{F}_s$, we have $\mathbb{E}((X_t^2 - X_s^2 - \int_s^t |\varphi_u|^2\mathrm{d}u)\mathbf{1}_A) = 0$. But we have

$$\Big(X_t^2 - X_s^2 - \int_s^t |\varphi_u|^2\mathrm{d}u\Big)\mathbf{1}_A = \overset{\mathrm{L}^1}{\lim_{n\to\infty}}\Big((X_t^{(n)})^2 - (X_s^{(n)})^2 - \int_s^t |\psi_u^{(n)}|^2\mathrm{d}u\Big)\mathbf{1}_A,$$

and from what we did for step functions, the right hand side has zero mean for all $n \geq 0$.

Finally, the last formula of the theorem comes from the fact that I is a linear isometry, via the polarization $4I(\varphi)I(\psi) = I(\varphi+\psi)^2 - I(\varphi-\psi)^2$ for all $\varphi,\psi \in \mathscr{L}^2_{\mathbb{R}^d}$, giving, for all $t \geq 0$, $\langle I(\varphi), I(\psi)\rangle_t = \int_0^t(\varphi_s\cdot\psi_s)\mathrm{d}s$. ■

Example 5.2.3. $\int_0^t B_s\mathrm{d}B_s$.

In the proof of Theorem 5.2.2, the approximation by step functions remains valid for square integrable U_i's. Since $(B_s\mathbf{1}_{s\in[0,t]})_{s\geq 0} \in \mathscr{L}^2_{\mathbb{R}^d}$, this gives a meaning to the formula in Section 5.1:

$$\int_0^t B_s\mathrm{d}B_s = \int_0^t B_s\mathbf{1}_{s\in[0,t]}\mathrm{d}B_s = \overset{\mathbb{P}}{\lim_{|\delta|\to 0}}\sum_i B_{t_{i-1}}(B_{t_i} - B_{t_{i-1}}) = \frac{B_t^2 - t}{2}, \quad t \geq 0.$$

We will learn how to compute many other integrals by using the Itô formula of Chapter 7. The quantity above is clearly a centered martingale. We can also easily check the Itô isometry as

$$\mathbb{E}\Big(\Big(\int_0^t B_s \mathrm{d}B_s\Big)^2\Big) = \frac{\mathbb{E}(B_t^4 - 2tB_t^2 + t^2)}{4} = \frac{3t^2 - 2t^2 + t^2}{4} = \frac{t^2}{2} = \int_0^t \mathbb{E}(B_t^2)\mathrm{d}s.$$

Note that we have here a martingale with non-deterministic increasing process: for all $t \geq 0$,

$$\Big\langle \int_0^\bullet B_s \mathrm{d}B_s \Big\rangle_t = \int_0^t B_s^2 \mathrm{d}s \quad \text{(random and non-Gaussian).}$$

The angle bracket is a quadratic variation and is thus here some sort of χ^2.

Example 5.2.4. Itô integration by parts formula.

Let $B = (B_t)_{t\geq 0}$ and $W = (W_t)_{t\geq 0}$ be two BM on $\mathbb{R}$ for the same filtration $(\mathscr{F}_t)_{t\geq 0}$, with $B_0 = W_0 = 0$. As in Section 5.1 and Example 5.2.3, for all $t > 0$ and all subdivision $0 = t_0 < \cdots < t_n = t$, the identity

$$\begin{aligned}\sum_i W_{t_i}(B_{t_{i+1}} - B_{t_i}) = \sum_i (W_{t_{i+1}} B_{t_{i+1}} - W_{t_i} B_{t_i}) \\ - \sum_i (W_{t_{i+1}} - W_{t_i})(B_{t_{i+1}} - B_{t_i}) \\ - \sum_i B_{t_i}(W_{t_{i+1}} - W_{t_i})\end{aligned}$$

gives the integration by parts formula (will work for arbitrary continuous martingales)

$$\int_0^t W_s \mathrm{d}B_s = W_t B_t - [W,B]_t - \int_0^t B_s \mathrm{d}W_s.$$

We recover the formula of Example 5.2.3 when $B = W$. On the contrary, if B and W are independent then $[W,B]_t = 0$. Namely, with $\Delta_i^B = B_{t_{i+1}} - B_{t_i}$ and $\Delta_i^W = W_{t_{i+1}} - W_{t_i}$, we have $\mathbb{E}(\sum_i \Delta_i^W \Delta_i^B) = 0$ while

$$\mathbb{E}\big(\big(\sum_i \Delta_i^W \Delta_i^B\big)^2\big) = \sum_i \mathbb{E}((\Delta_i^B)^2)\mathbb{E}((\Delta_i^W)^2) = \sum_i (t_{i+1} - t_i)^2 \leq t \max_i (t_{i+1} - t_i) \to 0.$$

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

$$\mathbb{E}(B_t W_t \mid \mathscr{F}_s) = \mathbb{E}((B_t - B_s)(W_t - W_s) + B_t W_s + B_s W_t - B_s W_s \mid \mathscr{F}_s) = B_s W_s$$

thus BW is a martingale and thus $\langle B,W\rangle = 0$.

Remark 5.2.5. Reduction to univariate stochastic integrals.

For all $d \geq 1$, $\varphi = (\varphi^1,\ldots,\varphi^d) \in \mathscr{L}^2_{\mathbb{R}^d}$ if and only if $\varphi^i \in \mathscr{L}^2_{\mathbb{R}}$ for all $1 \leq i \leq d$, and we have, for all $t \geq 0$,

$$\int_0^t \varphi_s \mathrm{d}B_s = \sum_{i=1}^d \int_0^t \varphi_s^i \mathrm{d}B_s^i.$$

The d-dimensional integral is a linear combination of one-dimensional integrals.

5.3 Brownian semi-martingales and Itô formula

Our main objective is to solve Stochastic Differential Equations with respect to X of the form

$$X_t = X_0 + \int_0^t \sigma(s,X_s)\mathrm{d}B_s + \int_0^t b(s,X_s)\mathrm{d}s,$$

$t \geq 0$, driven by a Brownian motion B, where X_0, σ, and b are given. The first integral of the right hand side requires the notion of stochastic integral with respect to Brownian motion, defined in $\mathbb{P}$, while the second integral requires the notion of Lebesgue–Stieltjes integral ω by ω. Beyond stochastic differential equations, we say that a process $(X_t)_{t\geq 0}$ is a Brownian semi-martingale or an Itô process when it takes the form

$$X_t = X_0 + \int_0^t \varphi_s \mathrm{d}B_s + \int_0^t \psi_s \mathrm{d}s$$

where $\varphi \in \mathscr{L}^2_{\mathbb{R}}$ is a square integrable progressive process and where ψ is a locally bounded progressive process. The first integral in the right hand side is a square integrable martingale while the second integral is a finite variation process. Then Itô formula writes, for all $f \in \mathscr{C}^2(\mathbb{R}^d, \mathbb{R})$,

$$\begin{aligned} f(X_t) - f(X_0) &= \int_0^t f'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\mathrm{d}s \\ &= \int_0^t f'(X_s)\varphi_s \mathrm{d}B_s + \int_0^t f'(X_s)\psi_s \mathrm{d}s + \frac{1}{2}\int_0^t f''(X_s)\mathrm{d}s. \end{aligned}$$

The problem with this formula is that the first integral in the right hand side is not well defined because $f'(X)\varphi$ is not necessarily in $\mathscr{L}^2_{\mathbb{R}}$. Actually this stochastic integral can be defined via a limiting procedure involving truncation stopping times, producing what we call a local martingale which is not a martingale in general. Provided that we extend the stochastic integral by this way, the right hand side is a again an (extended) Itô process, showing stability of this structure by composition with $\mathscr{C}^2$ functions.

In Chapter 7 the Itô formula follows from a Taylor formula, and the last term is produced as in Section 5.1 by the quadratic variation of the Brownian integrator. In particular, if $f \in \mathscr{C}^1(\mathbb{R}, \mathbb{R})$, with primitive F, then

$$\int_0^t f(B_s)\mathrm{d}B_s = F(B_t) - \frac{1}{2}\int_0^t f'(B_s)\mathrm{d}s.$$

We recover as a special case the formula $\int_0^t B_s \mathrm{d}B_s = \frac{1}{2}B_t - \frac{1}{2}t$ identified by hand in Section 5.1.

More generally, it is proved in particular in Chapter 7 that if $(X_t)_{t\geq 0}$ is a d-dimensional process for which each coordinate process is an Itô process, then the Itô formula writes, for all $f \in \mathscr{C}^2(\mathbb{R}^d, \mathbb{R})$,

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \partial_i f(X_s)\mathrm{d}X^i_s + \frac{1}{2}\sum_{i=1}^d \int_0^t \partial^2_{i,i} f(X_s)\mathrm{d}s.$$

If needed, these Itô formulas for Itô processes are essentially sufficient for a first reading of Chapter 8 on Stochastic Differential Equations driven by Brownian motion, as well as Chapter 9 on the Feynman–Kac formula and the Kakutani representation of the solution of Dirichlet problems.

The notion of stochastic integral with respect to a general (semi-)martingale integrator beyond BM is considered in Chapter 6, and the corresponding Itô formula is proved and studied in Chapter 7. These notions are deep achievement of stochastic calculus. Beyond pragmatism, and in order to avoid being the Monsieur Jourdain of semi-martingales, the best would be to try at some point to enter chapters 6 and 7.

Chapter 6

Itô stochastic integral and semi-martingales

Our aim is to construct a stochastic integral which goes beyond the stochastic integral with respect to the Brownian motion integrator of Chapter 5, with an integrator $(M_t)_{t\geq 0}$ which can be at least a general continuous square integrable martingale $(M_t)_{t\geq 0}$ and with an integrand $(\varphi_t)_{t\geq 0}$ which can be at least random and possibly square integrable. The ambition is thus to define the process

$$\Big(\int_0^t \varphi_s \mathrm{d}M_s\Big)_{t\geq 0}.$$

- In Section 6.1, for one-dimensional processes, we construct the Itô stochastic integral when M is a continuous martingale bounded in L^2, and when φ is square integrable in a suitable sense. This is done from step processes and by using the Hilbert structure available for both φ and M.
- In Section 6.2, by using truncation (localization via stopping times cutoff), we extend the previous Itô stochastic integral 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 6.3, 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.

6.1 Stochastic integral with respect to continuous martingales bounded in L^2

In this section, the integrator is taken in $\mathbb{M}_0^2$, the set of continuous martingales bounded in L^2 and issued from the origin. This allows to benefit from the Hilbert nature of this set, see Theorem 4.3.1[1]. We would like to define, for $M\in\mathbb{M}_0^2$,

$$\int_0^t \varphi_s \mathrm{d}M_s, \quad t\geq 0,$$

for reasonable integrands φ. Since the integrator M is a one-dimensional process, it is natural to consider a one-dimensional integrand φ. Inspired by what we did before for the Brownian motion integrator in the one-dimensional case $d=1$, we denote by $\mathscr{L}^2(M)$ the set of progressive real processes $(\varphi_t)_{t\geq 0}$ such that

$$\mathbb{E}\Big(\int_0^\infty \varphi_s^2 \mathrm{d}\langle M\rangle_s\Big)<\infty.$$

In this formula, the integral on $[0,\infty)$ is understood at all fixed ω as the limit as $t\to\infty$ of the integral on $[0,t]$ with respect to the increasing and thus bounded variation function $s\in[0,t]\mapsto\langle M\rangle_s$, see Theorem 1.7.3.

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

$$\mathscr{L}^2(M)=\mathrm{L}^2(\Omega\times\mathbb{R}_+,\mathscr{P},\mu),$$

[1]Note that the one-dimensional Brownian motion is square integrable but not bounded in L^2, but we could restrict our analysis on processes on the time interval $[0,t]$, and Brownian motion is bounded in L^2 on every finite time interval.

where $\mathscr{P}$ is the progressive σ-field (Theorem 2.1.1), and μ the finite measure defined for all $A \in \mathscr{P}$ by

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

Its total mass is $\mathbb{E}(\langle M \rangle_\infty) = \|M\|^2_{\mathbb{M}_0^2}$. 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 $\langle B \rangle$ it is deterministic. The scalar product and the norm of the Hilbert space $\mathscr{L}^2(M)$ are given respectively by

$$\langle \varphi, \psi \rangle_{\mathscr{L}^2(M)} = \mathbb{E}\Big(\int_0^\infty \varphi_s \psi_s \mathrm{d}\langle M \rangle_s \Big) \quad \text{and} \quad \|\varphi\|^2_{\mathscr{L}^2(M)} = \mathbb{E}\Big(\int_0^\infty \varphi_s^2 \mathrm{d}\langle M \rangle_s \Big).$$

We denote for short by $\mathscr{S}$ the set of real valued progressive bounded step processes. We have $\mathscr{S} \subset \mathscr{L}^2(M)$.

Lemma 6.1.1. Approximation or density.

Let $M \in \mathbb{M}_0^2$. The vector space $\mathscr{S}$ of bounded step processes is dense in $\mathscr{L}^2(M)$. In other words, for all $\varphi \in \mathscr{L}^2(M)$, and all $\varepsilon > 0$, there exists $\psi \in \mathscr{S}$ with

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

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

$$\Phi_t = \int_0^t \varphi_s \mathrm{d}\langle M \rangle_s.$$

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

$$\mathbb{E}\Big(\int_0^t |\varphi_s| \mathrm{d}\langle M \rangle_s \Big) \leq \Big(\mathbb{E}\Big(\int_0^t \varphi_s^2 \mathrm{d}\langle M \rangle_s \Big) \Big)^{1/2} \Big(\mathbb{E}(\langle M \rangle_\infty) \Big)^{1/2}$$

and the right hand side is finite since $M \in \mathbb{M}_0^2$ and $\varphi \in \mathscr{L}^2(M)$. This shows that almost surely, for all $t \geq 0$,

$$\int_0^t |\varphi_s| \mathrm{d}\langle M \rangle_s < \infty.$$

The process $(\Phi_t)_{t \geq 0}$ is continuous with finite variations and $\Phi_t \in \mathrm{L}^1$ for all $t \geq 0$. Now, let $0 \leq s \leq t$ and let Z be a bounded $\mathscr{F}_s$-measurable random variable. Since $(\psi_u)_{u \geq 0} = (Z \mathbf{1}_{u \in (s,t]})_{u \geq 0} \in \mathscr{S}$, we have

$$\langle \varphi, \psi \rangle_{\mathscr{L}^2(M)} = \mathbb{E}\Big(Z \int_s^t \varphi_u \mathrm{d}\langle M \rangle_u \Big) = 0.$$

Therefore $\mathbb{E}(Z(\Phi_t - \Phi_s)) = 0$. Since Z is an arbitrary bounded $\mathscr{F}_s$-measurable random variable and since $\Phi_t \in \mathrm{L}^1$ for all $t \geq 0$, it follows that $(\Phi_t)_{t \geq 0}$ is a martingale for $(\mathscr{F}_t)_{t \geq 0}$. On the other hand Φ is continuous with finite variations, and therefore, thanks to Lemma 4.1.6, we get $\Phi = 0$. Having in mind the initial definition on Φ, this means that almost surely the signed measure with density φ_s with respect to $\mathrm{d}\langle M \rangle_s$ is zero. But this is possible only if $\varphi_s = 0$, $\mathrm{d}\langle M \rangle_s$ almost everywhere, in other words only if $\varphi = 0$ in $\mathscr{L}^2(M)$, as expected. ■

Theorem 6.1.2. Itô stochastic integral with respect to elements of $\mathbb{M}_0^2$.

Let $M \in \mathbb{M}_0^2$. There exists a unique linear map $I_M : \mathscr{L}^2(M) \to \mathbb{M}_0^2$, denoted for all $\varphi \in \mathscr{L}^2(M)$ and $t \geq 0$

$$I_M(\varphi)_t = \int_0^t \varphi_s \mathrm{d}M_s,$$

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

1. for all $\varphi \in \mathscr{S}$ 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 $\varphi \in \mathscr{L}^2(M)$ add all $t \geq 0$,

$$\underbrace{\mathbb{E}\Big(\Big(\int_0^\infty \varphi_s \mathrm{d}M_s\Big)^2\Big)}_{\|I_M(\varphi)\|^2_{\mathbb{M}_0^2}} = \underbrace{\mathbb{E}\Big(\int_0^\infty \varphi_s^2 \mathrm{d}\langle M\rangle_s\Big)}_{\|\varphi\|^2_{\mathscr{L}^2(M)}}.$$

Moreover, for all $\varphi \in \mathscr{L}^2(M)$, $I_M(\varphi)$ is the unique element of $\mathbb{M}_0^2$ such that for all $N \in \mathbb{M}_0^2$ and $t \geq 0$,

$$\langle I_M(\varphi), N\rangle_t = \int_0^t \varphi_s \mathrm{d}\langle M, N\rangle_s.$$

Furthermore, for all $\varphi \in \mathscr{L}^2(M)$ and all stopping time T and all $t \geq 0$,

$$\underbrace{\int_0^{t\wedge T} \varphi_s \mathrm{d}M}_{I_M(\varphi)_t^T} = \underbrace{\int_0^t \varphi_s \mathbf{1}_{s\leq T} \mathrm{d}M}_{I_M(\varphi \mathbf{1}_{\bullet \leq T})_t} = \underbrace{\int_0^t \varphi_s \mathrm{d}M^T}_{I_{M^T}(\varphi)_t}.$$

Proof. The proof follows the lines of the proof for the Brownian motion integrator.

First of all the definition of $I_M(\varphi)$ for $\varphi \in \mathscr{S}$ does not depend on the decomposition chosen for φ. It is immediate to check that the map I_M is linear on $\mathscr{S}$. Let us prove that it is an isometry on $\mathscr{S}$.

The following lemma generalizes an observation already made in the proof of Theorem 4.1.4.

Lemma 6.1.3. Baby stochastic integral.

Let M be a continuous martingale, let $0 \leq u \leq v$, and let U be an $\mathscr{F}_u$ measurable bounded random variable. Set $M^{[u,v]} = U(M^v - M^u) = (U(M_{v\wedge t} - M_{u\wedge t}))_{t\geq 0}$.

1. $M^{[u,v]}$ is a martingale, which is almost surely constant outside the time interval (u, v), more precisely identically zero on $[0, u]$ and constant and equal to $U(M_v - M_u)$ on $[v, +\infty)$
2. If M is square integrable, then so is $M^{[u,v]}$ and $\langle M^{[u,v]}\rangle = U^2(\langle M\rangle^v - \langle M\rangle^u)$
3. If M is square integrable and $0 \leq u \leq v \leq u' \leq v'$ then $\langle M^{[u,v]}, M^{[u',v']}\rangle = 0$

Proof. Proof of Lemma 6.1.3

1. Almost immediate, the fact that $U(M^v - M^u) = 0$ on $[0, u]$ is crucial
2. $[U(M^v - M^u)] = [UM^v] + [UM^u] - 2[UM^u, UM^v] = [UM]^v + [UM]^u - 2[UM]^{u\wedge v} = U([M]^v - [M]^u)$
3. Follows from the definition of quadratic covariation and first property (zero outside intervals).

■

For all $\varphi \in \mathscr{S}$ and all $i \in \{0, \ldots, n-1\}$, by Lemma 6.1.3,

$$M^{(i)} = U_i(M_{t_{i+1}\wedge\bullet} - M_{t_i\wedge\bullet}) \in \mathbb{M}_0^2, \quad \text{and thus} \quad I_M(\varphi) = \sum_{i=0}^{n-1} M^{(i)} \in \mathbb{M}_0^2,$$

moreover $M^{(i)}$ is constant outside the interval (t_i, t_{i+1}), and for all $i, j \in \{0, \ldots, n-1\}$,

$$\langle M^{(i)}, M^{(j)}\rangle = U_i^2(\langle M\rangle_{t_{i+1}\wedge\bullet} - \langle M\rangle_{t_i\wedge\bullet})\mathbf{1}_{i=j},$$

(note that this implies orthogonality in $\mathbb{M}_0^2$ of $M^{(i)}$ and $M^{(j)}$ when $i \neq j$). This gives

$$\langle I_M(\varphi)\rangle = \sum_{i=0}^{n-1} \langle M^{(i)}\rangle = \sum_{i=1}^{n-1} U_i^2(\langle M\rangle_{t_{i+1}\wedge\bullet} - \langle M\rangle_{t_i\wedge\bullet}) = \int_0^\bullet \varphi_s^2 \mathrm{d}\langle M\rangle_s.$$

The isometry property follows, more precisely

$$\|I_M(\varphi)\|_{\mathbb{M}_0^2}^2 = \mathbb{E}(\langle I_M(\varphi)\rangle_\infty) = \mathbb{E}\Big(\int_0^\infty \varphi_s^2 \mathrm{d}\langle M\rangle_s\Big) = \|\varphi\|_{\mathscr{L}^2(M)}^2.$$

By Lemma 6.1.1, $\mathscr{S}$ is dense in the space $\mathscr{L}^2(M)$, and the linear isometry I_M extends uniquely to $\mathbb{M}_0^2$.

Let $N \in \mathbb{M}_0^2$. For all $\varphi \in \mathscr{L}^2(M)$, the Kunita – Watanabe inequality (Corollary 4.1.10) gives

$$\mathbb{E}\Big(\int_0^\infty |\varphi|_s |\mathrm{d}\langle M, N\rangle_s|\Big) \le \|\varphi\|_{\mathscr{L}^2(M)} \|N\|_{\mathbb{M}_0^2}.$$

It follows that $\int_0^\infty \varphi_s \mathrm{d}\langle M, N\rangle_s$ is well defined and belongs to L^1. If $\varphi \in \mathscr{S}$ then

$$\langle I_M(\varphi), N\rangle = \sum_{i=0}^{n-1} \langle M^i, N\rangle$$

and $\langle M^i, N\rangle = U_i(\langle M, N\rangle_{t_{i+1}\wedge\bullet} - \langle M, N\rangle_{t_i\wedge\bullet})$, and thus

$$\langle I_M(\varphi), N\rangle = \sum_{i=0}^{n-1} \langle M^i, N\rangle = \sum_{i=0}^{n-1} U_i(\langle M, N\rangle_{t_{i+1}\wedge\bullet} - \langle M, N\rangle_{t_i\wedge\bullet}) = \int_0^\bullet \varphi_s \mathrm{d}\langle M, N\rangle_s.$$

This gives the desired formula when $\varphi \in \mathscr{S}$. But the map $X \in \mathbb{M}_0^2 \mapsto \langle X, N\rangle_\infty \in \mathrm{L}^1$ is continuous since by the Kunita – Watanabe inequality (Corollary 4.1.10), we have

$$\mathbb{E}(|\langle X, N\rangle_\infty|) \le \mathbb{E}(\langle X\rangle_\infty)^{1/2} \mathbb{E}(\langle N\rangle_\infty)^{1/2} = \|N\|_{\mathbb{M}_0^2} \|X\|_{\mathbb{M}_0^2}.$$

If now $\varphi \in \mathscr{L}^2(M)$ and $\varphi = \lim_{n\to\infty} \varphi^{(n)}$ in $\mathscr{L}^2(M)$ for a sequence $(\varphi^{(n)})_{n\ge0}$ in $\mathscr{S}$, then

$$\langle I_M(\varphi), N\rangle_\infty = \lim_{n\to\infty} \langle I_M(\varphi^{(n)}), N\rangle_\infty = \lim_{n\to\infty} \int_0^\infty \varphi_s^{(n)} \mathrm{d}\langle M, N\rangle_s = \int_0^\infty \varphi_s \mathrm{d}\langle M, N\rangle_s,$$

where we have used for the last equality the inequality

$$\mathbb{E}\Big(\Big|\int_0^\infty (\varphi_s^{(n)} - \varphi_s) \mathrm{d}\langle M, N\rangle_s\Big|\Big) \le \|\varphi^{(n)} - \varphi\|_{\mathscr{L}^2(M)} \|N\|_{\mathbb{M}_0^2},$$

which follows from the Kunita – Watanabe inequality (Corollary 4.1.10). We have obtained

$$\langle I_M(\varphi), N\rangle_\infty = \int_0^\infty \varphi_s \mathrm{d}\langle M, N\rangle_s.$$

Finally, to replace ∞ by an arbitrary $t \ge 0$, we can replace N by the stopped process $N^t = (N_s \mathbf{1}_{s\le t})_{s\ge0}$.

Moreover this formula characterizes $I_M(\varphi)$ in $\mathbb{M}_0^2$. Indeed, if $X \in \mathbb{M}_0^2$ satisfies the same formula, then, for all $N \in \mathbb{M}_0^2$, $\langle I_M(\varphi) - X, N\rangle = 0$, and taking $X = I_M(\varphi) - X$ gives $\langle I_M(\varphi) - X\rangle = 0$, thus $X = I_M(\varphi)$.

Furthermore, for all $N \in \mathbb{M}_0^2$ and $\varphi \in \mathscr{L}^2(M)$, and all stopping time T, the properties of the angle bracket (Corollary 4.1.9) of two continuous square integrable martingales give, for all $t \ge 0$,

$$\langle I_M(\varphi)^T, N\rangle_t = \langle I_M(\varphi), N\rangle_{t\wedge T} = \int_0^{t\wedge T} \varphi_s \mathrm{d}\langle M, N\rangle_s = \int_0^t \varphi_s \mathbf{1}_{s\le T} \mathrm{d}\langle M, N\rangle_s.$$

Similarly, we have, for all $t \ge 0$,

$$\langle I_{M^T}(\varphi), N\rangle_t = \int_0^t \varphi_s \mathrm{d}\langle M^T, N\rangle_s = \int_0^t \varphi_s \mathrm{d}\langle M, N\rangle_s^T = \int_0^t \varphi_s \mathbf{1}_{s\le T} \mathrm{d}\langle M, N\rangle_s.$$

These formulas, together with the preceding characterization of I_M, give the desired formulas with T. ■

Corollary 6.1.4. Angle bracket, moments, associativity.

1. For all $M \in \mathbb{M}_0^2$, all $\varphi \in \mathscr{L}^2(M)$, and all $t \geq 0$,
$$\mathbb{E}\Big(\int_0^t \varphi_s \mathrm{d}M_s\Big) = 0.$$
2. For all $M, N \in \mathbb{M}_0^2$, all $\varphi \in \mathscr{L}^2(M)$, all $\psi \in \mathscr{L}^2(N)$, and all $t \geq 0$,
$$\Big\langle \int_0^\bullet \varphi_s \mathrm{d}M_s, \int_0^\bullet \psi_s \mathrm{d}N_s \Big\rangle_t = \int_0^t \varphi_s \psi_s \mathrm{d}\langle M, N\rangle_s,$$
which gives
$$\mathbb{E}\Big(\Big(\int_0^t \varphi_s \mathrm{d}M_s\Big)\Big(\int_0^t \psi_s \mathrm{d}N_s\Big)\Big) = \mathbb{E}\Big(\int_0^t \varphi_s \psi_s \mathrm{d}\langle M, N\rangle_s\Big).$$
3. For all $M \in \mathbb{M}_0^2$, all $\varphi \in \mathscr{L}^2(M)$, and all progressive process ψ, we have
$$\psi \in \mathscr{L}^2(I_M(\varphi)) \quad \text{iff} \quad \varphi\psi \in \mathscr{L}^2(M), \quad \text{and in this case} \quad I_{I_M(\varphi)}(\psi) = I_M(\varphi\psi).$$

Proof.

1. Theorem 6.1.2 gives $I_M(\varphi) \in \mathbb{M}_0^2$, in particular $I_M(\varphi)$ is centered, as for all the elements of $\mathbb{M}_0^2$
2. By polarization, we can assume without loss of generality that $M = N$. The property follows then from the isometry property of Theorem 6.1.2 applied with $(\varphi_s \mathbf{1}_{s\leq t})_{s\geq 0}$ together with the stopping time property in Theorem 6.1.2 with the deterministic stopping time t.
3. By using Theorem 1.7.3 and then the second property of the present Corollary with $N = M$ and $\psi = \varphi$,
$$\mathbb{E}\Big(\int_0^\infty \varphi_s^2 \psi_s^2 \mathrm{d}\langle M, M\rangle_s\Big) = \mathbb{E}\Big(\int_0^\infty \psi_s^2 \mathrm{d}\int_0^s \varphi_u^2 \mathrm{d}\langle M, M\rangle_u\Big) = \mathbb{E}\Big(\int_0^\infty \psi_s^2 \mathrm{d}\langle I_M(\varphi)\rangle_s\Big),$$
which gives that $\varphi\psi \in \mathscr{L}^2(M)$ iff $\psi \in \mathscr{L}^2(I_M(\varphi))$. Next, by Theorem 6.1.2, for all $N \in \mathbb{M}_0^2$ and all $t \geq 0$,
$$\langle I_M(\varphi\psi), N\rangle_t = \int_0^t \varphi_s \psi_s \mathrm{d}\langle M, N\rangle_s = \int_0^t \psi_s \mathrm{d}\int_0^s \varphi_u \mathrm{d}\langle M, N\rangle_u = \int_0^t \psi_s \mathrm{d}\langle I_M(\varphi), N\rangle_s,$$
which implies, by the characterization via N property of Theorem 6.1.2, that $I_M(\varphi\psi) = I_{I_M(\varphi)}(\psi)$.

■

6.2 Stochastic integral with respect to continuous local martingales

Up to now we have constructed the Itô integral for an integrator which is either BM or an element of $\mathbb{M}_0^2$. Our aim now is to consider an integrator $M \in \mathscr{M}_0^{\mathrm{loc}}$. The notion of increasing process of a local martingale is natural, see Lemma 4.2.6. By analogy with what we did before, we consider…

- the set $\mathscr{L}^0(M)$ of progressive $\varphi : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}$ such that almost surely
$$\int_0^\infty \varphi_s^2 \mathrm{d}\langle M\rangle_s < \infty,$$
- the set $\mathscr{L}^2(M) \subset \mathscr{L}^0(M)$ of progressive $\varphi : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}$ such that
$$\mathbb{E}\Big(\int_0^\infty \varphi_s^2 \mathrm{d}\langle M\rangle_s\Big) < \infty,$$

both quotiented by the equivalence relation related to equality $\langle M\rangle$ almost everywhere.

Theorem 6.2.1. Itô stochastic integral with respect to continuous local martingales.

Let $M \in \mathcal{M}_0^{\mathrm{loc}}$.

- For all $\varphi \in \mathcal{L}^0(M)$, there exists a unique $I_M(\varphi) \in \mathcal{M}_0^{\mathrm{loc}}$ such that for all $N \in \mathcal{M}_0^{\mathrm{loc}}$ and $t \geq 0$,

$$\langle I_M(\varphi), N\rangle_t = \int_0^t \varphi_s \mathrm{d}\langle M, N\rangle_s.$$

- For all $\varphi \in \mathcal{L}^0(M)$ and all stopping time T, we have, for all $t \geq 0$,

$$\underbrace{\int_0^{t\wedge T} \varphi_s \mathrm{d}M_s}_{I_M(\varphi)_t^T} = \underbrace{\int_0^t \varphi_s \mathbf{1}_{s\leq T} \mathrm{d}M_s}_{I_M(\varphi \mathbf{1}_{\bullet \leq T})} = \underbrace{\int_0^t \varphi_s \mathrm{d}M_s^T}_{I_{M^T}(\varphi)_t}$$

- If $\varphi \in \mathcal{L}^0(M)$ and if ψ is a progressive process then

$$\psi \in \mathcal{L}^0(I_M(\varphi)) \quad \text{iff} \quad \varphi\psi \in \mathcal{L}^0(M), \quad \text{and in this case} \quad I_{I_M(\varphi)}(\psi) = I_M(\varphi\psi)$$

- We have a <u>defective Itô isometry</u> in the sense that for all $\varphi \in \mathcal{L}^0(M)$ and all $t \geq 0$, in $[0, +\infty]$,

$$\mathbb{E}\Big(\Big(\int_0^t \varphi_s \mathrm{d}M_s\Big)^2\Big) \leq \mathbb{E}\Big(\int_0^t \varphi_s^2 \mathrm{d}\langle M\rangle_s\Big),$$

  either the right hand side is finite and equality holds or it is infinite and the inequality is trivial.

- If $M \in \mathbb{M}_0^2$ then I_M coincides on $\mathcal{L}^2(M)$ with the Itô integral of Theorem 6.1.2

- If $M \in \mathcal{M}_0^2$ and $\varphi \in \mathcal{L}^2(M)$ then $I_M(\varphi) \in \mathcal{M}_0^2$ and the Itô isometry holds: for all $t \geq 0, \varphi \in \mathcal{L}^2(M)$,

$$\mathbb{E}(I_M(\varphi)_t^2) = \mathbb{E}\int_0^t \varphi_s^2 \mathrm{d}\langle M\rangle_s.$$

  In particular if M is BM on $\mathbb{R}$ then I_M coincides on $\mathcal{L}^2_{\mathbb{R}^d}$ with Theorem 5.2.2 with $d = 1$

Note that if $M \in \mathcal{M}_0^{\mathrm{loc}}$, 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 <u>may not be centered and the Itô isometry may not hold</u>. However, the Itô stochastic integral with respect to $M \in \mathcal{M}_0^2$ such as Brownian motion do satisfy the centering and Itô isometry for integrands in $\mathcal{L}^2(M)$.

Proof. We proceed by localization with stopping times. For all $n \geq 0$ we define the stopping time

$$T_n = \inf\Big\{t \geq 0 : \int_0^t (1 + \varphi_s^2) \mathrm{d}\langle M\rangle_s \geq n\Big\}.$$

Almost surely $T_n \nearrow +\infty$ as $n \to \infty$. The fact that it grows with n comes from the way we define T via the integral of something non-negative for a positive measure. Now for all $n \geq 0$, the Doob stopping (Theorem 2.5.1) implies that $M^{T_n} \in \mathcal{M}_0^{\mathrm{loc}}$. Moreover, thanks to the $1 + \cdots$ in T_n, for all $n \geq 0$ and all $t \geq 0$, we have

$$\langle M^{T_n}\rangle_t = \langle M\rangle_{t\wedge T_n} \leq n.$$

Therefore, by Lemma 4.2.7, $M^{T_n} \in \mathbb{M}_0^2$. Moreover we have, from the properties of the angle bracket,

$$\int_0^\infty \varphi_s^2 \mathrm{d}\langle M^{T_n}\rangle_s = \int_0^{T_n} \varphi_s^2 \mathrm{d}\langle M\rangle_s,$$

therefore $\varphi \in \mathcal{L}^2(M^{T_n})$ and from Theorem 6.1.2, the stochastic process $I_{M^{T_n}}(\varphi)$ makes sense and belongs to $\mathbb{M}_0^2$. 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

$$I_{M^{T_n}}(\varphi) = I_{M^{T_m \wedge T_n}}(\varphi) = (I_{M^{T_m}}(\varphi))^{T_n}.$$

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 $\mathcal{M}_0^{\mathrm{loc}}$ since $(I_M(\varphi))^{T_n} = I_{M^{T_n}}(\varphi) \in \mathbb{M}_0^2$ for all $n \geq 0$.

Now, let $N \in \mathcal{M}_0^{\mathrm{loc}}$. For all $n \geq 0$, let us define $T'_n = \inf\{t \geq 0 : |N_t| \geq n\}$ and $S_n = T_n \wedge T'_n$. Almost surely $S_n \nearrow +\infty$ as $n \to \infty$. We have $N^{T'_n} \in \mathbb{M}_0^2$ and, thanks to Theorem 6.1.2, for all $t \geq 0$,

$$\begin{aligned}\langle I_M(\varphi), N\rangle_t^{S_n} &= \langle I_M(\varphi)^{T_n}, N^{T'_n}\rangle_t \\ &= \int_0^t \varphi_s \mathrm{d}\langle M^{T_n}, N^{T'_n}\rangle_s \\ &= \int_0^t \varphi_s \mathrm{d}\langle M, N\rangle_s^{S_n} \\ &= \int_0^{t\wedge S_n} \varphi_s \mathrm{d}\langle M, N\rangle_s,\end{aligned}$$

which gives the desired formula as $n \to +\infty$. As in Theorem 6.1.2, this formula characterizes $I_M(\varphi)$, mainly due to the fact that if $M \in \mathcal{M}_0^{\mathrm{loc}}$ is such that $\langle M\rangle = 0$ then $M = 0$.

It is not difficult to prove the remaining properties, including the relation to previous integrals. ■

Remark 6.2.2. Brownian motion as a martingale.

Let B be a real Brownian motion issued from the origin. We have $\mathbb{E}(B_t) = 0$ and $\mathbb{E}(B_t^2) = t < \infty$ for all $t \geq 0$ and $\sup_{t\geq 0} \mathbb{E}(B_t^2) = +\infty$, therefore

$$B \in \mathcal{M}_0^2 \subset \mathcal{M}_0^{\mathrm{loc}} \quad \text{while} \quad B \not\in \mathbb{M}_0^2.$$

Moreover since $B \in \mathcal{M}_0^{\mathrm{loc}}$ we get that $B^{T_n} \in \mathbb{M}_0^2$ for all n, where $T_n = \inf\{t \geq 0 : |B_t| \geq n\}$. We could define stochastic integrals for processes on the time interval $[0,t]$ and BM is bounded in L^2 on each finite interval, and this would correspond to introduce the space $\mathbb{M}_{0,[0,t]}^2$. Another possibility would be to generalize the construction of stochastic integral that we gave for BM to integrators in $\mathcal{M}_0^2$.

6.3 Notion of semi-martingale and stochastic integration

The notion of quadratic variation of a process is considered in Definition 4.1.1. The quadratic variation of Brownian motion is considered in Theorem 3.2.1. In dimension one, for all $t \geq 0$, $[B]_t = \langle B\rangle_t = t$.

Definition 6.3.1. Semi-martingales.

A continuous semi-martingale $X = (X_t)_{t\geq 0}$ is an adapted process with decomposition of the form

$$X = X_0 + M + V$$

where M and V are adapted continuous processes issued from the origin and such that

- $M = (M_t)_{t\geq 0}$ is a continuous local martingale issued from the origin ($\in \mathcal{M}_0^{\mathrm{loc}}$)
- $V = (V_t)_{t\geq 0}$ is a continuous finite variation process issued form the origin.

Note in particular that X_0 is $\mathcal{F}_0$ measurable, and that X is adapted.

Lemma 6.3.2. Canonical decomposition.

The decomposition of a continuous semi-martingale is unique.
In particular, a finite variation continuous local martingale is almost surely constant.

Proof. If $X = X_0 + M + V = X_0 + \widetilde{M} + \widetilde{V}$. Now the process $W = \widetilde{V} - V = M - \widetilde{M}$ is continuous and has finite variation, and thus, by Lemma 4.1.2 (see also 4.1.6) it has zero quadratic variation. Therefore, by Lemma 4.2.6, $\langle M - \widetilde{M}\rangle = [M - \widetilde{M}] = 0$, and since $M_0 = \widetilde{M}_0 = V_0 = \widetilde{V}_0 = 0$, this implies $M - \widetilde{M} = 0$ and thus $V = \widetilde{V}$. ■

Let $\mathscr{L}^{\text{locb}}$ be the set of processes which are progressive and locally bounded.
By Theorem 2.1.1, all continuous adapted processes belong to $\mathscr{L}^{\text{locb}}$.
Let $\mathscr{M}_0^{\text{semi}}$ be the set of continuous semi-martingales issued from the origin.
Let $\varphi = (\varphi_t)_{t\geq 0} \in \mathscr{L}^{\text{locb}}$ and $X = M + V \in \mathscr{M}_0^{\text{semi}}$. Then, for all $t \geq 0$, almost surely,

$$\int_0^t |\varphi_s||\mathrm{d}V_s| < \infty.$$

Additionally, we have $\varphi \in \mathscr{L}^0(M)$, and therefore the stochastic integral $\int_0^\bullet \varphi_t \mathrm{d}M_t$ is well defined. It follows then that we can define the integral $I_X(\varphi)$ of φ with respect to the semi-martingale $X = M + V$ as

$$I_X(\varphi) = \int_0^t \varphi_s \mathrm{d}X_s = \int_0^t \varphi_s \mathrm{d}M_s + \int_0^t \varphi_s \mathrm{d}V_s,$$

and we can see that the result $I_X(\varphi)$ is itself a continuous semi-martingale.

Theorem 6.3.3. Properties of the integral with respect to a continuous semi-martingale.

1. the map $(\varphi, X) \mapsto I_X(\varphi)$ is bilinear, from $\mathscr{L}^{\text{locb}} \times \mathscr{M}_0^{\text{semi}}$ to $\mathscr{M}_0^{\text{semi}}$
2. for all stopping time T, $\varphi \in \mathscr{L}^{\text{locb}}$, $X \in \mathscr{M}_0^{\text{semi}}$, we have

$$\underbrace{\int_0^{t\wedge T} \varphi_s \mathrm{d}X_s}_{(I_X(\varphi))^T} = \underbrace{\int_0^t \varphi_s \mathbf{1}_{s\leq T} \mathrm{d}X_s}_{I_X(\varphi \mathbf{1}_{[0,T]})} = \underbrace{\int_0^t \varphi_s \mathrm{d}X_s^T}_{I_{X^T}(\varphi)}.$$

3. for all $\varphi, \psi \in \mathscr{L}^{\text{locb}}$ and $X \in \mathscr{M}_0^{\text{semi}}$, we have $I_{I_X(\varphi)}(\psi) = I_X(\varphi\psi)$ in other words

$$\int_0^t \psi_s \mathrm{d} \int_0^s \varphi_u \mathrm{d}X_u = \int_0^t \varphi_s \psi_s \mathrm{d}X_s$$

4. for all $X \in \mathscr{M}_0^{\text{semi}}$, if X is a local martingale (respectively a finite variation process) then for all $\varphi \in \mathscr{L}^{\text{locb}}$, the process $I_X(\varphi)$ is a local martingale (respectively a finite variation process)
5. if $\varphi$is a step process with decomposition $\varphi_t = U_0 \mathbf{1}_0(t) + \sum_{i=0}^{n-1} U_i \mathbf{1}_{(t_i,t_{i+1}]}(t)$, $0 = t_0 < t_1 < \cdots < t_n$, $n \geq 1$, with U_i $\mathscr{F}_{t_i}$-measurable for all i, then for all $X \in \mathscr{M}_0^{\text{semi}}$ and all $t \geq 0$,

$$I_X(\varphi)_t = \int_0^t \varphi_s \mathrm{d}X_s = \sum_{i=0}^{n-1} U_i (X_{t_{i+1}\wedge t} - X_{t_i \wedge t}).$$

Proof. The first four properties follow immediately from the definition or from the properties of the stochastic integral with respect to continuous local martingales and with respect to finite variation processes.

The fifth and last property does not follow immediately due to the fact that the random variables U_i's are not assumed to be bounded. It suffices to overcome this difficulty when $X = M$ is a continuous local martingale. In this case, we define, for all $k \geq 1$,

$$T_k = \inf\{t \geq 0 : |\varphi_t| \geq k\} = \inf\{t_i : |U_i| \geq k\} \in [0, +\infty].$$

It is a stopping time and almost surely $T_k \nearrow +\infty$ as $k \to \infty$ and, for all $k \geq 1$ and $s \geq 0$,

$$\varphi_s \mathbf{1}_{[0,T_k]}(s) = \sum_{i=0}^{n-1} U_i^{(k)} \mathbf{1}_{(t_i,t_{i+1}]}(s) \quad \text{with} \quad U_i^{(k)} = U_i \mathbf{1}_{T_k > t_i}.$$

Now $\varphi \mathbf{1}_{[0,T_k]} \in \mathscr{S}$, which allows to write

$$(I_M(\varphi))_{t\wedge T_k} = I_M(\varphi \mathbf{1}_{[0,T_k]})_t = \sum_{i=0}^{n-1} U_i^{(k)} (M_{t_{i+1}\wedge t} - M_{t_i\wedge t})$$

and it remains to send k to $+\infty$. ∎

Theorem 6.3.4. Dominated convergence for stochastic integrals.

Let $X = M + V \in \mathcal{M}_0^{\mathrm{semi}}$. Let φ and $\varphi^{(n)}$, $n \geq 1$, be progressive locally bounded processes ($\in \mathcal{L}^{\mathrm{locb}}$). Let ψ be a non-negative progressive locally bounded process (≥ 0 and $\in \mathcal{L}^{\mathrm{locb}}$). Let $t > 0$. Suppose that almost surely the following properties hold:

- for all $s \in [0,t]$, $\lim_{n\to\infty} \varphi_s^{(n)} = \varphi_s$
- for all $s \in [0,t]$ and all $n \geq 0$, $|\varphi_s^{(n)}| \leq \psi_s$

Then

$$\int_0^t \varphi_s^{(n)} \mathrm{d}X_s \xrightarrow[n\to\infty]{\mathbb{P}} \int_0^t \varphi_s \mathrm{d}X_s.$$

Proof. From Theorem 1.7.3, the usual dominated convergence theorem gives

$$\int_0^t \varphi_s^{(n)} \mathrm{d}V_s \xrightarrow[n\to\infty]{\text{a.s.}} \int_0^t \varphi_s \mathrm{d}V_s.$$

It remains to address the local martingale part. We do it by localization with the stopping time

$$T_k = \inf\Big\{s \in [0,t] : \int_0^s \psi_u^2 \mathrm{d}\langle M\rangle_u \geq k\Big\}.$$

We have $T_k \to +\infty$ as $k \to \infty$ almost surely. Now by using the defective Itô isometry

$$\mathbb{E}\Big(\Big(\int_0^{t\wedge T_k} \varphi_s^{(n)} \mathrm{d}M_s - \int_0^{t\wedge T_k} \varphi_s \mathrm{d}M_s\Big)^2\Big) \leq \mathbb{E}\Big(\int_0^{t\wedge T_k} (\varphi_s^{(n)} - \varphi_s)^2 \mathrm{d}\langle M\rangle_s\Big)$$

But by definition of T_k, and thanks to the assumptions on φ and ψ, we can use dominated convergence to get that for all k, the right hand side tends to 0 as $n \to \infty$. It follows in particular that for all k,

$$\int_0^{t\wedge T_k} \varphi_s^{(n)} \mathrm{d}M_s - \int_0^{t\wedge T_k} \varphi_s \mathrm{d}M_s \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Therefore, for all $\varepsilon > 0$,

$$\mathbb{P}\Big(\Big|\int_0^t \varphi_s^{(n)} \mathrm{d}M_s - \int_0^t \varphi_s \mathrm{d}M_s\Big| \geq \varepsilon\Big) \leq \mathbb{P}(T_k \leq t) + \mathbb{P}\Big(\Big|\int_0^{t\wedge T_k} \varphi_s^{(n)} \mathrm{d}M_s - \int_0^{t\wedge T_k} \varphi_s \mathrm{d}M_s\Big| \geq \varepsilon\Big).$$

It remains to select k sufficiently large, and then n sufficiently large. ■

Theorem 6.3.5. From sums to stochastic integrals.

Let $X = M + V \in \mathcal{M}_0^{\mathrm{semi}}$ and let φ be a continuous adapted process (which is in particular progressive). Then for all $t > 0$ and for all sequence $(\delta_n)_{n\geq 0}$ of sub-divisions of $[0,t]$, $\delta_n : 0 = t_0^{(n)} < \cdots < t_{m_n}^{(n)} = t$, $m_n \geq 1$, such that $|\delta_n| = \max_{0\leq k\leq m_n-1} |t_{k+1}^{(n)} - t_k^{(n)}| \to 0$ as $n \to \infty$, we have

$$\sum_{k=0}^{m_n-1} \varphi_{t_k^{(n)}} (X_{t_{k+1}^{(n)}} - X_{t_k^{(n)}}) \xrightarrow[n\to\infty]{\mathbb{P}} \int_0^t \varphi_s \mathrm{d}X_s.$$

Proof. For all $n \geq 0$, the step process $\varphi^{(n)}$ defined by $\varphi^{(n)} = \varphi_0 \mathbf{1}_0 + \sum_{k=0}^{m_n-1} \varphi_{t_k^{(n)}} \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}$ is progressive and satisfies all the assumptions of the dominated convergence Theorem 6.3.4 with the process ψ defined by $\psi_s = \sup_{u\in[0,s]} |\varphi_u|$ for all $s \in [0,t]$. This process ψ is indeed continuous and adapted and thus progressive and locally bounded. The desired result follows then from Theorem 6.3.4 since

$$\int_0^t \varphi_s^{(n)} \mathrm{d}X_s = \sum_{k=0}^{m_n-1} \varphi_{t_k^n} (X_{t_{k+1}^{(n)}} - X_{t_k^{(n)}}).$$

■

Remark 6.3.6. Multivariate case.

We encounter the multivariate case when $X = B$ is BM in $\mathbb{R}^d$ and $\varphi = (\varphi^1, \dots, \varphi^d)$ is deterministic and in $\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, x)$, for which $I_X(\varphi) = I_B(\varphi)$ is the Wiener integral

$$\int_0^\bullet \varphi_s \mathrm{d}B_s = \sum_{i=1}^d \int_0^\bullet \varphi_s^i \mathrm{d}B_t^i.$$

At the opposite side of generality, let $X = (X^1, \dots, X^d)$ be a d-dimensional process such that for all $1 \le i \le d$, $X^i = M^i + V^i \in \mathcal{M}_0^{\mathrm{semi}}$. We decide to say that such a process is by definition a d-dimensional continuous semi-martingale issued from the origin. Then for all d-dimensional process $\varphi = (\varphi^1, \dots, \varphi^d)$ such that for all $1 \le i \le d$, $\varphi^i \in \mathscr{L}^{\mathrm{locb}}$, we define

$$\int_0^\bullet \varphi_s \mathrm{d}X_s = \sum_{i=1}^d \int_0^\bullet \varphi_s^i \mathrm{d}X_s^i.$$

In particular, when $\varphi^i \in \mathscr{L}^2(M^i)$ for all $1 \le i \le d$, then, for all $t \ge 0$,

$$\mathbb{E}\Big(\Big(\int_0^t \varphi_s \mathrm{d}M\Big)^2\Big) = \sum_{i,j=1}^d \mathbb{E}\int_0^t \varphi_s^i \varphi_s^j \mathrm{d}\langle M^i, M^j\rangle_s.$$

When M is a d-dimensional BM then $\langle M^i, M^j\rangle_s = s\mathbf{1}_{i=j}$ and we recover the Itô isometry

$$\mathbb{E}\Big(\Big(\int_0^t \varphi_s \mathrm{d}B_s\Big)^2\Big) = \sum_{i=1}^d \mathbb{E}\int_0^t (\varphi_s^i)^2 \mathrm{d}s = \mathbb{E}\int_0^t |\varphi_s|^2 \mathrm{d}s.$$

6.4 Summary of stochastic integrals and involved spaces

Integrator X	**Integrand** φ	**Integral** $I_X(\varphi)$
BM_0 in $\mathbb{R}^d$	$\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$	Gaussian martingale issued from the origin (Wiener integral, Chapter 3)
BM_0 in $\mathbb{R}^d$	$\mathscr{L}^2_{\mathbb{R}^d}$	$\mathcal{M}_0^2$ (Itô integral with Brownian motion integrator, Chapter 5)
$\mathbb{M}_0^2$	$\mathscr{L}^2(M)$, $X = M$	$\mathbb{M}_0^2$
$\mathcal{M}_0^{\mathrm{loc}}$	$\mathscr{L}^0(M)$, $X = M$	$\mathcal{M}_0^{\mathrm{loc}}$
$\mathcal{M}_0^{\mathrm{semi}}$	$\mathscr{L}^{\mathrm{locb}}$	$\mathcal{M}_0^{\mathrm{semi}}$

Space	**Definition**
$\mathrm{L}^2_{\mathbb{R}^d}(\mathbb{R}_+, \mathrm{d}x)$	Deterministic square integrable $\varphi : \mathbb{R}_+ \to \mathbb{R}^d$
$\mathscr{S}_{\mathbb{R}^d}$	Progressive $\varphi : \Omega \times \mathbb{R}_+ \to \mathbb{R}^d$ step processes with bounded increments
$\mathscr{L}^2_{\mathbb{R}^d}$	Progressive $\varphi : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}^d$ such that $\mathbb{E}\int_0^\infty |\varphi_s|^2 \mathrm{d}s < \infty$
$\mathscr{S}$	Progressive $\varphi : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ step processes with bounded increments
$\mathscr{L}^2(M)$	Progressive $\varphi : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}$ such that $\mathbb{E}\int_0^\infty \varphi_s^2 \mathrm{d}\langle M\rangle_s < \infty$
$\mathscr{L}^0(M)$	Progressive $\varphi : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}$ such that a.s. $\int_0^\infty \varphi_s^2 \mathrm{d}\langle M\rangle_s < \infty$
$\mathscr{L}^{\mathrm{locb}}$	Progressive $\varphi : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}$, a.s. $\sup_{s\in[0,t]} |\varphi_s| < \infty$ (locally bounded)
$\mathcal{M}_0^2$	Continuous square integrable martingales issued from the origin
$\mathbb{M}_0^2$	Continuous martingales bounded in L^2 issued from the origin
$\mathcal{M}_0^{\mathrm{loc}}$	Continuous local martingales issued from the origin
$\mathcal{M}_0^{\mathrm{semi}}$	Continuous semi-martingales issued from the origin

Chapter 7

Itô formula and applications

7.1 Itô formula

Classical analysis comes with its fundamental formula of integral calculus expressing a regular function as the Riemann/Stieltjes/Lebesgue/Young integral of its derivative. For the Itô stochastic integral of stochastic analysis, the analogue is the Itô formula. This is also known as the "Itô lemma". A theorem here! It was discovered in 2000 – see [7] – that we might speak about Döblin[1] – Itô theorem/lemma/formula.

We have already seen in Section 5.1 that if B is a real Brownian motion issued from the origin then

$$B^2 = 2\int_0^\bullet B_s \mathrm{d}B_s + \langle B\rangle \in \mathscr{M}^{\mathrm{semi}}.$$

The Itô formula goes beyond and states more generally that the image of a semi-martingale by a $\mathscr{C}^2$ function is again a semi-martingale. It can be seen as a fundamental rule of calculus for the Itô stochastic integral.

Theorem 7.1.1. Itô or Döblin formula for d-dimensional semi-martingales.

If $X = (X_t)_{t\geq 0}$ is d-dimensional such that for all $1 \leq i \leq d$ its i-th coordinate $(X_t^i)_{t\geq 0}$ is a continuous semi-martingale with decomposition $X^i = X_0^i + M^i + V^i$ then for all $f \in \mathscr{C}^2(\mathbb{R}^d, \mathbb{R})$ and all $t \geq 0$,

$$\begin{aligned} f(X_t) &= f(X_0) + \underbrace{\sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s)\mathrm{d}X_s^i}_{\text{usual calculus term}} + \underbrace{\frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\mathrm{d}\langle M^i, M^j\rangle_s}_{\text{Itô stochastic correction}} \\ &= f(X_0) + \underbrace{\sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s)\mathrm{d}M_s^i}_{\in \mathscr{M}_0^{\mathrm{loc}}} + \underbrace{\sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s)\mathrm{d}V_s^i + \frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\mathrm{d}\langle M^i, M^j\rangle_s}_{\text{finite variation process}}. \end{aligned}$$

In particular $f(X) = (f(X_t))_{t\geq 0} \in \mathscr{M}^{\mathrm{semi}}$. For $d = 1$ the formula simply writes, for all $t \geq 0$,

$$\begin{aligned} f(X_t) &= f(X_0) + \underbrace{\int_0^t f'(X_s)\mathrm{d}X_s}_{\text{usual calculus term}} + \underbrace{\frac{1}{2}\int_0^t f''(X_s)\mathrm{d}\langle M\rangle_s}_{\text{Itô stochastic correction}} \\ &= f(X_0) + \underbrace{\int_0^t f'(X_s)\mathrm{d}M_s}_{\in \mathscr{M}_0^{\mathrm{loc}}} + \underbrace{\int_0^t f'(X_s)\mathrm{d}V_s + \frac{1}{2}\int_0^t f''(X_s)\mathrm{d}\langle M\rangle_s}_{\text{finite variation process}}. \end{aligned}$$

The example above corresponds to $d = 1$, $X_0 = 0$, $M = B$, $V = 0$, and $f = (\bullet)^2$.
In the case $f(x) = x$ for all $x \in \mathbb{R}^d$ we get the very natural formula $\int_0^t \mathrm{d}X_s = X_t - X_0$.
The Itô formula remains valid for processes defined on a finite and deterministic time interval $[0, T]$.

1. The second order term with the local martingale part of X is typical from the Itô stochastic integral. It does not appear in the similar formula for the Stratonovich stochastic integral.

[1] Named after Wolfgang Döblin or Vincent Doblin, French – German mathematician (1915 – 1940).

2. Alternatively, in a more condensed form, the formula writes, denoting $\langle X\rangle = \langle M\rangle$,
$$f(X_t) = f(X_0) + \int_0^t \nabla f(X_s)\cdot \mathrm{d}X_s + \frac{1}{2}\int_0^t \mathrm{Hess}(f)(X_s)\cdot \mathrm{d}\langle X\rangle_s.$$
The formula can also be written alternatively using differential abstract notation, namely
$$\mathrm{d}f(X_t) = \nabla f(X_t)\cdot \mathrm{d}X_t + \frac{1}{2}\mathrm{Hess}(f)(X_t)\cdot \mathrm{d}\langle X\rangle_t.$$

3. When $d=1$ and $V=0$ then this gives, for all $M\in\mathscr{M}^{\mathrm{loc}}$ (beware that we incorporate X_0 into M)
$$f(M_t) = f(M_0) + \int_0^t f'(M_s)\mathrm{d}M_s + \frac{1}{2}\int_0^t f''(M_s)\cdot \mathrm{d}\langle M\rangle_s.$$
Note that M_0 plays no role in the integrators $\mathrm{d}M$ and $\mathrm{d}\langle M\rangle$ however it plays a role in the integrands $f'(M)$ and $f''(M)$. Now if M is additionally bounded then we can take expectations, and this gives
$$\mathbb{E}(f(M_t)) = \mathbb{E}(f(M_0)) + \frac{1}{2}\mathbb{E}\int_0^t f''(M_s)\mathrm{d}\langle M\rangle_s.$$
We could always localize $M-M_0$ with a stopping time in order to get it bounded (in particular in L^2).

4. When $d=1$ and $M=0$ we recover the fundamental formula of calculus for the Lebesgue–Stieltjes integral, namely for all continuous adapted finite variation process V,
$$f(V_t) = f(V_0) + \int_0^t f'(V_s)\mathrm{d}V_s.$$

5. When $d=1$ and $f=(\cdot)^2$ we obtain the formula
$$\int_0^t X_s\mathrm{d}X_s = \frac{X_t^2 - X_0^2 - \langle M\rangle_t}{2}.$$
This generalizes the formula that we have already obtained for Brownian motion in Section 5.1. In particular when $X_0=0$ and $X=M$ in other words when $V=0$ then this states
$$2\int_0^\bullet M_s\mathrm{d}M_s = M^2 - \langle M\rangle.$$
Actually we already knew that if $M\in\mathscr{M}_0^{\mathrm{loc}}$ then $M^2 = (M^2-\langle M\rangle)+\langle M\rangle \in \mathscr{M}_0^{\mathrm{semi}}$, and more generally, in the same spirit, if $M\in\mathscr{M}^{\mathrm{loc}}$ then $M^2 = M_0^2 + \underbrace{(M-M_0)^2 - \langle M-M_0\rangle + 2M_0(M-M_0)}_{\in\mathscr{M}_0^{\mathrm{loc}}} + \langle M-M_0\rangle \in \mathscr{M}^{\mathrm{semi}}$.

6. For all $M,N\in\mathscr{M}^{\mathrm{loc}}$, the Itô formula with $f(x_1,x_2)=x_1x_2$ and $X=(M,N)$ gives, for all $t\geq 0$,
$$M_tN_t = M_0N_0 + \int_0^t M_s\mathrm{d}N_s + \int_0^t N_s\mathrm{d}M_s + \langle M,N\rangle_t$$
which is an <u>integration by parts formula</u>. In the same spirit, for all $M\in\mathscr{M}^{\mathrm{loc}}$ and for all adapted continuous finite variation process V, taking $f(x_1,x_2)=x_1x_2$ and $X=(M,V)$ gives, for all $t\geq 0$,
$$M_tV_t = M_0V_0 + \int_0^t M_s\mathrm{d}V_s + \int_0^t V_s\mathrm{d}M_s.$$

7. When $V=0$, $X_0=x$, and $M=B$ is BM with $B_0=0$ then $\langle M^i,M^j\rangle_t = t\mathbf{1}_{i=k}$ and the Itô formula becomes
$$f(B_t) = f(x) + \int_0^t \nabla f(B_s)\cdot\mathrm{d}B_s + \frac{1}{2}\int_0^t \Delta f(B_s)\mathrm{d}s.$$
In particular, when $f=|\cdot|^2$, we obtain
$$|B_t|^2 = |x|^2 + 2\int_0^t B_s\cdot\mathrm{d}B_s + td.$$
When $d=1$ and $B_0=0$ we recover the formula $\frac{1}{2}(B_t^2-t) = \int_0^t B_s\mathrm{d}B_s$ that we have already found in Chapter 6. We say that $(|B_t|)_{t\geq 0}$ is a <u>Bessel process</u> and that $(|B_t|^2)_{t\geq 0}$ a <u>squared Bessel process</u>.

8. Let us consider the case $X = (X_1, \ldots, X_{d+1}) = (B, V)$, where B is a Brownian motion in $\mathbb{R}^d$, $V_t = t$ for all $t \geq 0$, and $f(x, x_{d+1}) = \exp(\lambda \cdot x - \frac{1}{2}|\lambda|^2 x_{d+1})$, with $\lambda \in \mathbb{R}$ fixed. The process $N^\lambda = f(X)$ satisfies $N_0^\lambda = 1$. As explained in the proof of Theorem 7.2.1, by the Itô formula, N^λ is a semi-martingale that solves the following stochastic differential equation which catches intuitively its exponential nature:

$$N_t^\lambda = 1 + \int_0^t N_s^\lambda \mathrm{d}(\lambda \cdot B_s).$$

An exponential semi-martingale of the same type appears also in Theorem 7.3.1 with more general ingredients. Such "exponential semi-martingales" are known as Doléans-Dade exponential martingales. When $d = \lambda = 1$ this process is also known as geometric Brownian motion.

9. For a time dependent function $f(t, x)$, the formula with $\widetilde{X} = (t, X)$ gives

$$f(t, X_t) = f(0, X_0) + \int_0^t \partial_t f(s, X_s)\mathrm{d}s + \int_0^t \nabla_x f(s, X_s) \cdot \mathrm{d}X_s + \frac{1}{2}\int_0^t \mathrm{Hess}_x(f)(s, X_s)\mathrm{d}\langle M\rangle_s \cdot \mathrm{d}\langle M\rangle_s.$$

Note that $t \in \mathbb{R}_+ \mapsto t$ is a continuous finite variation process. It does not contribute to the last term.

10. The Itô formula extends naturally to a complex valued $f : \mathbb{R}^d \to \mathbb{C}$. Let us examine a nice and important special case. Let $X = X_0 + M + V$ be a continuous semi-martingale and let A be a finite variation process. For all $\lambda \in \mathbb{R}$, by the Itô formula for $f(x, y) = \mathrm{e}^{\mathrm{i}\lambda x + \frac{\lambda^2}{2} y}$ and $(X_t, A_t)_{t \geq 0}$,

$$N_t^\lambda = \mathrm{e}^{\mathrm{i}\lambda X_t + \frac{\lambda^2}{2} A_t} = \mathrm{e}^{\mathrm{i}\lambda X_0} + \mathrm{i}\lambda \int_0^t N_s^\lambda \mathrm{d}X_s + \frac{\lambda^2}{2}\int_0^t N_s^\lambda \mathrm{d}A_s - \frac{\lambda^2}{2}\int_0^t N_s^\lambda \mathrm{d}\langle M\rangle_s.$$

Now if additionally $X_0 = 0$ and $A = \langle M\rangle$ then we get the stochastic differential equation

$$N_t^\lambda = 1 + \mathrm{i}\lambda \int_0^t N_s^\lambda \mathrm{d}X_s.$$

This semi-martingale is used in the proof of the Lévy characterization of Brownian motion in Theorem 7.2.1. The case $N^{-\mathrm{i}\lambda}$ is a special case of the Doléans-Dade exponential of Theorem 7.3.1.

The Itô formula involves $\langle M, N\rangle$ with M, N continuous local martingales. Note that we have defined $\langle M, N\rangle$ only when both M and N are continuous local martingales. This coincides with the quadratic variation $[M, N]$. Note also that we have defined (when it exists) the quadratic variation $[M, N]$ for arbitrary processes. In particular $[M, N] = 0$ if M is a continuous process and N is a finite variation process.

Proof. We suppose first that on some event Ω', the random variables and processes X_0, M, and V are bounded in the sense that for some deterministic and finite constant C, almost surely

$$\sup_{\substack{t \geq 0 \\ 1 \leq i, j \leq d}} \left(|X_0| + |M_t^i| + |V_t^i| + |\langle M^i, M^j\rangle_t|\right) \leq C.$$

Under this assumption, we can assume without loss of generality that f is compactly supported. Next, note also that under this assumption, for all $1 \leq i \leq d$, since M^i is a bounded continuous local martingale, it is, by localization and dominated convergence, a bounded continuous martingale.

A Taylor formula for f gives, for all $x, y \in \mathbb{R}^d$,

$$\begin{aligned}
f(y) - f(x) &= \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle \mathrm{Hess}(f)(x)(y - x), y - x\rangle + r(x, y)|x - y|^2 \\
&= \sum_i \frac{\partial f}{\partial x_i}(x)(y_i - x_i) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(x)(y_i - x_i)(y_j - x_j) + r(x, y)|y - x|^2.
\end{aligned}$$

Since f is $\mathscr{C}^2$ with compact support, by Heine theorem, $x \mapsto f''(x) = \mathrm{Hess}(f)(x) = (\frac{\partial^2 f}{\partial x_i \partial x_j})_{1 \leq i, j \leq d}$ is uniformly continuous, and therefore there exists a bounded continuous non-decreasing function $g : \mathbb{R}_+ \to \mathbb{R}$ such that $\lim_{u \to 0} g(u) = 0$ and $|r(x, y)| \leq g(|x - y|)$ for all $x, y \in \mathbb{R}$.

Now we fix $t > 0$ and we consider a sequence $(\delta_n)_{n\geq 0}$ of sub-divisions of $[0, t]$,

$$\delta_n : 0 = t_0^{(n)} < \cdots < t_{m_n}^{(n)} = t, \quad m_n \geq 1$$

such that $\lim_{n\to\infty} \max_{0\leq k\leq m_n-1} |t_{k+1}^{(n)} - t_k^{(n)}| = 0$. To simplify, we drop from now on the superscript ${}^{(n)}$. We denote $\Delta Y_k = Y_{t_{k+1}} - Y_{t_k}$ for all k and all process Y. A telescopic summation and the Taylor formula give

$$\begin{aligned} f(X_t) - f(X_0) &= \sum_k (f(X_{t_{k+1}}) - f(X_{t_k})) \\ &= \underbrace{\sum_k \langle \nabla f(X_{t_k}), \Delta X_k\rangle}_{S_1} + \underbrace{\frac{1}{2}\sum_k \langle \mathrm{Hess}(f)(X_{t_k})\Delta X_k, \Delta X_k\rangle}_{S_2} + \underbrace{\sum_k r(X_{t_k}, X_{t_{k+1}})|\Delta X_k|^2}_{S_3}. \end{aligned}$$

For the term S_1, we have, using Theorem 6.3.5,

$$S_1 = \sum_k \langle \nabla f(X_{t_k}), \Delta V_k\rangle + \sum_k \langle \nabla f(X_{t_k}), \Delta M_k\rangle \xrightarrow[n\to\infty]{\mathbb{P}} \int_0^t \nabla f(X_s)\mathrm{d}V_s + \int_0^t \nabla f(X_s)\mathrm{d}M_s = \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s)\mathrm{d}X_s^i.$$

For the term S_2, we have

$$\begin{aligned} S_2 &= \frac{1}{2}\sum_k \langle \mathrm{Hess}(f)(X_{t_k})\Delta X_k, \Delta X_k\rangle \\ &= \frac{1}{2}\sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})\Delta X_k^i \Delta X_k^j \\ &= \underbrace{\frac{1}{2}\sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})\Delta M_k^i \Delta M_k^j}_{S_2'} + \underbrace{\sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})\Delta M_k^i \Delta V_k^j}_{S_2''} + \underbrace{\frac{1}{2}\sum_k \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})\Delta V_k^i \Delta V_k^j}_{S_2'''}. \end{aligned}$$

Now since V^j has finite variation and since M^i is continuous, we have

$$2|S_2''| \leq \sum_{i,j} \max_k |\Delta M_k^i| \sum_k \left|\frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})\right| |\Delta V_k^j| \xrightarrow[n\to\infty]{\text{a.s.}} \sum_{i,j} 0 \times \int_0^t \left|\frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\right| \mathrm{d}|V^j|_s = 0.$$

Similarly, using this time the continuity of V^i and the finite variation of V^j, we get

$$2|S_2'''| \leq \sum_{i,j} \max_k |\Delta V_k^i| \sum_k \left|\frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})\right| |\Delta V_k^j| \xrightarrow[n\to\infty]{\text{a.s.}} \sum_{i,j} 0 \times \int_0^t \left|\frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\right| \mathrm{d}|V^j|_s = 0.$$

For S_2', using the notation $M^{i,j}$ for "M^i, M^j" and using the formulas of Lemma 7.1.2 in $(*)$ and $(**)$,

$$\begin{aligned} \mathbb{E}\Big(\Big(\sum_k \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})(\Delta M_k^i \Delta M_k^j - \Delta\langle M^{i,j}\rangle_k)\Big)^2\Big) &\overset{*}{=} \mathbb{E}\Big(\sum_k \Big(\frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})\Big)^2 (\Delta M_k^i \Delta M_k^j - \Delta\langle M^{i,j}\rangle_k)^2\Big) \\ &\leq \|\mathrm{Hess}(f)\|_\infty^2 \mathbb{E}\Big(\Big(\sum_k (\Delta M_k^i \Delta M_k^j - \Delta\langle M^{i,j}\rangle_k)\Big)^2\Big) \\ &\overset{**}{=} \|\mathrm{Hess}(f)\|_\infty^2 \mathbb{E}\Big(\Big(\sum_k \Delta M_k^i \Delta M_k^j - \langle M^{i,j}\rangle_t\Big)^2\Big). \end{aligned}$$

We have, for all i, j, using the fact that M is a bounded continuous martingale,

$$\sum_k \Delta M_k^i \Delta M_k^j - \langle M^{i,j}\rangle_t \xrightarrow[n\to\infty]{\mathrm{L}^2} 0.$$

Now, by Theorem 4.1.4 for $\langle M^i, M^j\rangle = [M^i, M^j]$ for all i, j, the fact that f is $\mathscr{C}^2$, and Theorem 1.7.3,

$$\lim_{n\to\infty}^{\mathbb{P}} S_2' = \lim_{n\to\infty}^{\mathbb{P}} \frac{1}{2}\sum_{i,j=1}^d \sum_k \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{t_k})\langle M^{i,j}\rangle_k$$

$$= \frac{1}{2}\sum_{i,j=1}^{d}\int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\mathrm{d}\langle M^i, M^j\rangle_s.$$

Regarding S_3, we have, using the monotony of g and Theorem 4.1.4,

$$\sum_k |r(X_{t_k}, X_{t_{k+1}})||\Delta X_k|^2 \le \underbrace{g(\max_k |\Delta X_k|)}_{\xrightarrow[n\to\infty]{\mathbb{P}} 0} \sum_{i=1}^{d}\Big(\underbrace{\sum_k (\Delta M_k^i)^2}_{\xrightarrow[n\to\infty]{\mathbb{P}} \langle M^i\rangle_t} + \underbrace{\sum_k (\Delta V_k^i)^2}_{\xrightarrow[n\to\infty]{\mathbb{P}} \langle V\rangle_t = 0}\Big).$$

This achieves the proof under the assumptions of boundedness of X_0, M, and V on an event Ω'.

To prove the general case, we consider the sequence $(T_k)_{n\ge 0}$ of stopping times defined for all $k \ge 0$ by

$$T_k = \inf\Big\{t \ge 0 : |X_0| + \sum_{i,j=1}^{d} |M_t^i| + |V_t^i| + |\langle M^i, M^j\rangle_t| \ge k\Big\}$$

then $T_k \nearrow +\infty$ a.s. as $k \to \infty$, and by arguing as in Remark 4.2.3, it follows that X^{T_k} is a continuous martingale. Moreover $|X^{T_k}| \le |X_0|\mathbf{1}_{|X_0|>k} + k\mathbf{1}_{|X_0|\le k}$ a.s. Using the first part on the event $\Omega' = \Omega'_k = \{|X_0| \le k\}$, this gives, on Ω'_k, thanks to the properties of (stochastic) integrals with respect to stopping times,

$$f(X_{t\wedge T_k}) = f(X_0) + \sum_{i=1}^{d}\int_0^{t\wedge T_k} \frac{\partial f}{\partial x_i}(X_s)\mathrm{d}X_s^i + \frac{1}{2}\sum_{i,j=1}^{d}\int_0^{t\wedge T_k} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\mathrm{d}\langle M^i, M^j\rangle_s.$$

Now both sides converge in probability as $k \to \infty$ to the desired formula. Indeed we get the desired formula on the event $\Omega'_k \cap \{T_k \ge t\}$, and $\mathbb{P}(\Omega'_k \cap \{T_k \ge t\}) \xrightarrow[k\to\infty]{} 1$. ■

It could be possible to prove Theorem 7.1.1 without truncation by using dominated convergence for stochastic integrals (Theorem 6.3.4) but still the second order term will remain the difficult part, see for instance [31]. Another classical proof of Theorem 7.1.1 consists in using the linearity of the statement with respect to f and the functional monotone class theorem machinery to get it for all f, see for instance [4]. This approach requires a stability by product which amounts to prove separately as a preparatory result the integration by parts formula. Such a proof may look short at first sight but remains a bit magical and artificial. The proof that we give based on a Taylor formula is maybe more intuitive and constructive.

Lemma 7.1.2. Quadratic formulas with increments of martingales.

Let M, N be bounded martingales issued from the origin and let Z be a bounded adapted process. Let $t > 0$ and let $\delta : 0 = t_0 < \cdots < t_n = t$ be a subdivision or partition of the interval $[0, t]$. Then, denoting $\Delta X_k = X_{t_{k+1}} - X_{t_k}$ and $Z_k = Z_{t_k}$, we have

$$\mathbb{E}\Big(\Big(\sum_k Z_k(\Delta M_k \Delta N_k - \Delta\langle M, N\rangle_k)\Big)^2\Big) = \mathbb{E}\Big(\sum_k Z_k^2(\Delta M_k \Delta N_k - \Delta\langle M, N\rangle_k)^2\Big)$$

and in particular, reading this formula from right to left when $Z = 1$ gives

$$\mathbb{E}\Big(\sum_k (\Delta M_k \Delta N_k - \Delta\langle M, N\rangle_k)^2\Big) = \mathbb{E}\Big(\Big(\sum_k \Delta M_k \Delta N_k - \langle M, N\rangle_t\Big)^2\Big).$$

Proof. The intuitive idea is to rely on conditional orthogonality properties of increments of square integrable martingales. For all k, we denote for simplicity $T_k = Z_k(\Delta M_k \Delta N_k - \Delta\langle M, N\rangle_k)$. By using the martingale property for the martingales M and N in $(*)$ and for the martingale $MN - \langle M, N\rangle$ in $(**)$, we get

$$\begin{aligned}
\mathbb{E}(T_k \mid \mathcal{F}_{t_k}) &= Z_k\mathbb{E}(M_{t_{k+1}}N_{t_{k+1}} + M_{t_k}N_{t_k} - M_{t_{k+1}}N_{t_k} - M_{t_k}N_{t_{k+1}} - \Delta\langle M, N\rangle_k \mid \mathcal{F}_{t_k})\\
&\overset{*}{=} Z_k\mathbb{E}(M_{t_{k+1}}N_{t_{k+1}} - M_{t_k}N_{t_k} - \Delta\langle M, N\rangle_k \mid \mathcal{F}_{t_k})\\
&= Z_k\mathbb{E}(\Delta(MN)_k - \Delta\langle M, N\rangle_k \mid \mathcal{F}_{t_k})\\
&= Z_k\mathbb{E}(\Delta(MN - \langle M, N\rangle)_k \mid \mathcal{F}_{t_k}) \overset{**}{=} 0.
\end{aligned}$$

For all $k' < k$, $T_{k'}$ is $\mathcal{F}_{t_k}$ measurable and $\mathbb{E}(T_{k'}T_k) = \mathbb{E}(T_{k'}\mathbb{E}(T_k \mid \mathcal{F}_{t_k})) = 0$. Thus $\mathbb{E}((\sum_k T_k)^2) = \mathbb{E}(\sum_k T_k^2)$. ■

Remark 7.1.3. Extension to discontinuous semi-martingales.

Let X be a real semi-martingale, right continuous with left limits (càdlàg). A jump occurs at time t when $X_t - X_{t^-} \neq 0$. It can be proved that the quadratic variation of the càdlàg local martingale part M of X, is well defined and admits a decomposition of the form

$$[M,M]_t = \sum_{s\leq t}(M_s - M_{s^-})^2 + [M,M]_t^c$$

where the sum runs over all jumps up to time t while the super-script "c" stands for "continuous". The Doléans-Dade – Meyer generalized Itô formula states then that for all $f \in \mathscr{C}^2(\mathbb{R},\mathbb{R})$ and all $t \geq 0$,

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s^-})\mathrm{d}X_s + \sum_{0<u\leq t}(f(X_s) - f(X_{s^-}) - f'(X_{s^-})(X_s - X_{s^-})) + \frac{1}{2}\int_0^t f''(X_s)\mathrm{d}[M,M]_s^c.$$

An accessible presentation of stochastic calculus for jump processes can be found in [41].

Remark 7.1.4. Stratonovich stochastic integral and Itô formula.

If M and N are continuous semi-martingales, we define the Stratonovich stochastic integral by

$$\int_0^t M_s \circ \mathrm{d}N_s = \int_0^t M_s \mathrm{d}N_s + \frac{1}{2}\langle M,N\rangle_t.$$

The Itô formula is then simpler and resembles to the fundamental rule of calculus

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s)\circ \mathrm{d}X_s^i,$$

for $f \in \mathscr{C}^2(\mathbb{R}^d,\mathbb{R})$. For more information and applications, see for instance [4, Chapter 6].

7.2 Lévy characterization of Brownian motion

Theorem 7.2.1. Lévy.

Let $X = (X_t)_{t\geq 0}$ be a d-dimensional adapted process such that:

- for all $1 \leq k \leq d$ the k-th coordinate $(X_t^k)_{t\geq 0}$ is a continuous local martingale
- for all $1 \leq j,k \leq d$ we have $(\langle X^j, X^k\rangle_t)_{t\geq 0} = (t\mathbf{1}_{j=k})_{t\geq 0}$.

Then X is a Brownian motion (with respect to the same filtration).

In practice, when $d = 1$, this is often used after showing that $(X_t)_{t\geq 0}$ and $(X_t^2 - t)_{t\geq 0}$ are local martingales.

Proof. By replacing X by $X - X_0$ we can assume that $X_0 = 0$. By Theorem 3.1.3, in order to show that X is a Brownian motion, it suffices to show that for all $\lambda \in \mathbb{R}^d$, the process $N^\lambda = (\exp(\mathrm{i}\lambda\cdot X_t + \frac{1}{2}|\lambda|^2 t))_{t\geq 0}$ is a martingale. Now $N^\lambda = f(X,V)$ where $f(x,x_{d+1}) = \exp(\mathrm{i}\lambda\cdot x + \frac{1}{2}|\lambda|^2 x_{d+1})$ and $V_t = t$ for all $t \geq 0$. The Itô formula[2] (Theorem 7.1.1) gives that N^λ is a continuous semi-martingale. It remains to use Lemma 7.2.2.

Lemma 7.2.2. Martingale criterion.

Let $(M_t)_{t\geq 0}$ be a continuous local martingale such that $M_0 \in \mathrm{L}^1$ and such that for all $t \geq 0$, there exists a finite constant C_t such that $\sup_{s\in[0,t]}|M_s| \leq C_t$. Then M is a martingale.

Note that we can construct continuous local martingales which are bounded in L^2 but not martingales!

[2] We can alternatively use the Itô formula for the time dependent $f(t,x) = \exp(\mathrm{i}\lambda\cdot x + \frac{1}{2}|\lambda|^2 t)$ and the semi-martingale X.

Proof. We can assume that $M_0 = 0$ by considering $M - M_0$. Next, we know that M is localized by $(T_n)_{n\geq 0}$ where $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$. Now, for all $t \geq 0$, we take $n > C_t$, hence $T_n > t$ and $(M_s)_{s\in[0,t]} = (M_{s\wedge T_n})_{s\in[0,t]}$, and since M^{T_n} is a martingale, we get that $(M_s)_{s\in[0,t]}$ is a martingale for all $t \geq 0$, hence M is a martingale. ∎

∎

Let us write the Itô formula in the proof of Theorem 7.2.1. It is crucial that $N^\lambda = \exp(\mathrm{i}\lambda X_t + \frac{1}{2}|\lambda|^2 V_t)$ for a finite variation process V such that that for all $1 \leq j,k \leq d$, $\langle X^j, X^k\rangle = V\mathbf{1}_{j=k}$. We have

$$\begin{aligned}
N_t^\lambda &= 1 + \sum_{k=1}^d \int_0^t \Big(\mathrm{i}\lambda_k N_s^\lambda \mathrm{d}X_s^k + \frac{1}{2}|\lambda|^2 N_s^\lambda \mathrm{d}V_s\Big) + \frac{\mathrm{i}^2}{2}\sum_{j,k=1}^d \lambda_j\lambda_k \int_0^t N_s^\lambda \mathbf{1}_{j=k}\mathrm{d}V_s\\
&= 1 + \mathrm{i}\int_0^t N_s^\lambda \mathrm{d}(\lambda\cdot X_s) + \underbrace{\frac{1}{2}|\lambda|^2\int_0^t N_s^\lambda \mathrm{d}V_s - \frac{1}{2}|\lambda|^2\int_0^t N_s^\lambda\mathrm{d}V_s}_{=0}\\
&= 1 + \mathrm{i}\int_0^t N_s^\lambda \mathrm{d}(\lambda\cdot X_s).
\end{aligned}$$

This shows that N^λ is a semi-martingale solving the stochastic differential equation

$$N^\lambda = 1 + \mathrm{i}\int_0^\bullet N_s^\lambda \lambda\cdot\mathrm{d}X_s.$$

The same idea, with the Laplace transform instead of the Fourier transform, is at the heart of the concept of Doléans-Dade exponential of Theorem 7.3.1. Note that $N^{-\mathrm{i}\lambda}$ is a multivariate Doléans-Dade exponential.

7.3 Doléans-Dade exponential

The following theorem generalizes what we already know for the Laplace transform of Brownian motion.

Theorem 7.3.1. Doléans-Dade[3] exponential.

[3] Named after Catherine Doléans-Dade (1942 – 2004), French American mathematician.

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 non-decreasing. For all $\lambda \in \mathbb{R}$ let us define the process

$$X^\lambda = (X_t^\lambda)_{t\geq 0} = (\mathrm{e}^{\lambda M_t - \frac{\lambda^2}{2}V_t})_{t\geq 0}.$$

Then the following properties are equivalent.

1. M is a local martingale and $\langle M\rangle = V$
2. X^λ is a local martingale for all $\lambda \in \mathbb{R}$.

Moreover, in this case,

3. X^λ solves the stochastic differential equation $X_0^\lambda = 1$ and $X_t^\lambda = 1 + \lambda\int_0^t X_s^\lambda \mathrm{d}M_s$, $t \geq 0$.
4. X^λ is a super-martingale, and a martingale if and only if $\mathbb{E}X_t^\lambda = 1$ for all $t \geq 0$
5. If X^λ is a martingale and $\mathbb{E}\mathrm{e}^{\lambda M_t} < \infty$ for all $\lambda \in \mathbb{R}$ and all $t \geq 0$ then M is a martingale.

Proof. Suppose that M is a local martingale and that we have $\langle M\rangle = V$. For all $\lambda \in \mathbb{R}$, by the Itô formula,

$$X^\lambda = 1 + \lambda\int_0^\bullet X_s^\lambda \mathrm{d}M_s.$$

Thus X^λ is a non-negative local martingale. Conversely, suppose first that for all $\lambda\in\mathbb{R}$, X^λ is a martingale and that M and V are bounded. Then for all $0\le s<t$ and all $A\in\mathscr{F}_s$,

$$\mathbb{E}(\mathbf{1}_A \mathrm{e}^{\lambda M_t-\frac{\lambda^2}{2}V_t})=\mathbb{E}(\mathbf{1}_A X_t^\lambda)=\mathbb{E}(\mathbf{1}_A X_s^\lambda)=\mathbb{E}(\mathbf{1}_A \mathrm{e}^{\lambda M_s-\frac{\lambda^2}{2}V_s}).$$

Taking the derivative with respect to λ, which is allowed here by dominated convergence since $X^\lambda(M-\lambda V)$ is dominated by a constant thanks to the fact that M and V are bounded, gives

$$\mathbb{E}(\mathbf{1}_A \mathrm{e}^{\lambda M_t-\frac{\lambda^2}{2}V_t}(M_t-\lambda V_t))=\mathbb{E}(\mathbf{1}_A \mathrm{e}^{\lambda M_s-\frac{\lambda^2}{2}V_s}(M_s-\lambda V_s)),$$

which shows by taking $\lambda=0$ that M is a martingale. Moreover, taking again the derivative with respect to λ, which is allowed here by dominated convergence since $X^\lambda((M-\lambda V)^2-V)$ is dominated by a constant,

$$\mathbb{E}(\mathbf{1}_A X_t^\lambda((M_t-\lambda V_t)^2-V_t))=\mathbb{E}(\mathbf{1}_A X_s^\lambda((M_s-\lambda V_s)^2-V_s)),$$

which shows by taking $\lambda=0$ that M^2-V is a martingale. In particular we get $\langle M\rangle=V$. Back to the general case, if for all $\lambda\in\mathbb{R}$, X^λ is a local martingale, then we proceed by localization via $T_n=\inf\{t\ge0:|M_t|+V_t\ge n\}$, for which $T_n\nearrow+\infty$ almost surely as $n\to\infty$ and, for all $n\ge0$, $(X^\lambda)^{T_n}$ is a bounded martingale and M^{T_n} and V^{T_n} are bounded. We use then what we did for the martingale case to get that M^{T_n} is a martingale and $\langle M^{T_n}\rangle=V^{T_n}$, for all $n\ge0$, which implies that M is a local martingale and $\langle M\rangle=V$.

Let us prove the last properties stated when M is a continuous local martingale and $V=\langle M\rangle$.

3. We already have seen that at the beginning of the proof.
4. Since X^λ is a local martingale, there exists stopping times $(T_n)_{n\ge0}$ such that $T_n\nearrow+\infty$ almost surely and such that $(X^\lambda_{t\wedge T_n})_{t\ge0}$ is a martingale for all $n\ge0$ then for all $0\le s\le t$, by the Fatou lemma,

$$X_s^\lambda=\lim_{n\to\infty}X^\lambda_{s\wedge T_n}=\lim_{n\to\infty}\mathbb{E}(X^\lambda_{t\wedge T_n}\mid\mathscr{F}_s)\ge\mathbb{E}(\underline{\lim}_{n\to\infty}X^\lambda_{t\wedge T_n}\mid\mathscr{F}_s)=\mathbb{E}(X_t^\lambda\mid\mathscr{F}_s).$$

In particular $\mathbb{E}X_t^\lambda\le\mathbb{E}X_0^\lambda=1$ and X^λ is martingale iff[4] $\mathbb{E}X_t^\lambda=1$ for all $t\ge0$.
5. We have already seen that at the beginning of the proof!

■

For a continuous local martingale M issued from the origin, there are criteria to ensure that the local martingale $\mathrm{e}^{M-\frac12\langle M\rangle}$ (Doléans-Dade exponential) is a martingale or even an u.i. martingale, namely:

- Domination: a local martingale dominated by an integrable random variable is a u.i. martingale.
- Bracket: a local martingale with integrable bracket is a martingale (see Lemma 4.2.7). Note that

$$\mathbb{E}\langle X^\lambda\rangle_t=\lambda^2\mathbb{E}\Big\langle\int_0^\bullet X_s^\lambda\mathrm{d}M_s,\int_0^\bullet X_s^\lambda\mathrm{d}M_s\Big\rangle_t=\lambda^2\mathbb{E}\int_0^t\mathrm{e}^{2\lambda M_s-\lambda^2\langle M\rangle_s}\mathrm{d}\langle M\rangle_s\le\lambda^2\mathbb{E}\int_0^t\mathrm{e}^{2\lambda M_s}\mathrm{d}\langle M\rangle_s.$$

- Novikov and Kazamaki conditions or criteria, which are specific to Doléans-Dade exponentials.

Theorem 7.3.2. Novikov[a] and Kazamaki[b] conditions.

[a]Named after Alexander Novikov, Russian-Australian mathematician.
[b]Named after Norihiko Kazamaki, Japanese mathematician.

If $M=(M_t)_{t\ge0}$ is a continuous local martingale issued from the origin then 1. ⇒ 2. ⇒ 3. where

1. Novikov condition: $\mathbb{E}\mathrm{e}^{\frac12\langle M\rangle_\infty}<\infty$
2. Kazamaki condition: M is a u.i. martingale and $\mathbb{E}\mathrm{e}^{\frac12 M_\infty}<\infty$

[4]If $A\le B$ then $A=B$ iff $\mathbb{E}(A)=\mathbb{E}(B)$. In particular, a super or sub martingale is a martingale iff its expectation is constant.

3. $e^{M-\frac{1}{2}\langle M\rangle}$ is a u.i. martingale.

This can be skipped at first reading.

Proof.

- 1. ⇒ 2. We have $\mathbb{E}\langle M\rangle_\infty \le \mathbb{E}e^{\frac{1}{2}\langle M\rangle_\infty} < \infty$, hence, by Lemma 4.2.7, M is a continuous martingale bounded in L^2 and thus u.i. in particular M_∞ exists. By proceeding as in the proof Theorem, 7.3.1 $e^{M-\frac{1}{2}\langle M\rangle}$ is a super-martingale issued from 1 and by the Fatou lemma $\mathbb{E}e^{M_t-\frac{1}{2}\langle M\rangle_t} \le 1$ for all $t \in [0,+\infty]$. Finally by the Cauchy – Schwarz inequality,

$$(\mathbb{E}e^{\frac{1}{2}M_\infty})^2 \le \mathbb{E}e^{M_\infty-\frac{1}{2}\langle M\rangle_\infty}\mathbb{E}e^{\frac{1}{2}\langle M\rangle_\infty} \le \mathbb{E}e^{\frac{1}{2}\langle M\rangle_\infty} < \infty.$$

- 2. ⇒ 3. Since M is a u.i. martingale, by Corollary 4.4.6, for an arbitrary stopping time T we have $\mathbb{E}(M_\infty \mid \mathscr{F}_T) = M_T$ and by the Jensen inequality

$$e^{\frac{1}{2}M_T} \le \mathbb{E}(e^{\frac{1}{2}M_\infty} \mid \mathscr{F}_T).$$

But since $\mathbb{E}e^{\frac{1}{2}M_\infty} < \infty$, it follows that the family $\{\mathbb{E}(e^{\frac{1}{2}M_\infty} \mid \mathscr{F}_T) : T \text{ stopping time}\}$ is u.i., and by the inequality, the family $\{e^{\frac{1}{2}M_T} : T \text{ stopping time}\}$ is u.i.

For all $a \in (0,1)$ let us define $M^{(a)} = aM/(a+1)$. We have

$$e^{aM-\frac{1}{2}\langle aM\rangle} = \left(e^{M-\frac{1}{2}\langle M\rangle}\right)^{a^2} e^{aM-a^2M} = \left(e^{M-\frac{1}{2}\langle M\rangle}\right)^{a^2} \left(e^{M^{(a)}}\right)^{1-a^2},$$

and then, for all stopping time T, by the Hölder inequality,

$$\begin{aligned}\mathbb{E}(e^{aM_T-\frac{1}{2}\langle aM\rangle_T}) &\le \left(\mathbb{E}e^{M_T-\frac{1}{2}\langle M\rangle_T}\right)^{a^2} \mathbb{E}\left(e^{M_T^{(a)}}\right)^{1-a^2}\\ &\overset{\star}{\le} \mathbb{E}(e^{M_T^{(a)}})^{1-a^2}\\ &\overset{\star\star}{\le} \mathbb{E}(e^{\frac{1}{2}M_T}\mathbf{1}_A)^{2a(1-a)}.\end{aligned}$$

We have used for $\star$ the fact that $\mathbb{E}(e^{M_T-\frac{1}{2}\langle M\rangle_T}) \le 1$ which comes from Theorem 2.5.6 with $S=0$ for the non-negative super-martingale $e^{M-\frac{1}{2}\langle M\rangle}$ with expectation ≤ 1. We have used for $\star\star$ the Jensen inequality for the concave function $u \in \mathbb{R}_+ \mapsto u^{2a/(1+a)}$ thanks to $2a/(1+a) \in (0,1)$.

It follows in particular that the family $\{e^{aM_T-\frac{1}{2}\langle M\rangle_T} : T \text{ stopping time}\}$ is u.i. Now, if $(T_n)_n$ is a localizing sequence for the local martingale $e^{aM-\frac{1}{2}\langle aM\rangle}$, then for all $0 \le s \le t$,

$$\mathbb{E}(e^{aM_{t\wedge T_n}-\frac{1}{2}\langle aM\rangle_{t\wedge T_n}} \mid \mathscr{F}_s) = e^{aM_{s\wedge T_n}-\frac{1}{2}\langle aM\rangle_{s\wedge T_n}},$$

and we can pass to the limit by the u.i. property for the family indexed by stopping times, hence $e^{aM-\frac{1}{2}\langle aM\rangle}$ is a u.i. martingale. Finally, using again the Jensen inequality in the last step,

$$1 = \mathbb{E}(e^{aM_\infty-\frac{1}{2}\langle aM\rangle_\infty}) \le \mathbb{E}(e^{M_\infty-\frac{1}{2}\langle M\rangle_\infty})^{a^2}\mathbb{E}(e^{M_\infty^{(a)}})^{1-a^2} \le \mathbb{E}(e^{M_\infty-\frac{1}{2}\langle M\rangle_\infty})^{a^2}\mathbb{E}(e^{\frac{1}{2}M_\infty})^{2a(1-a)}.$$

Now $a \to 1$ gives $\mathbb{E}(e^{M_\infty-\frac{1}{2}\langle M\rangle_\infty}) \ge 1$ hence $\mathbb{E}(e^{M_\infty-\frac{1}{2}\langle M\rangle_\infty}) = 1$, thus $e^{M-\frac{1}{2}\langle M\rangle}$ is a martingale.

■

7.4 Dubins – Schwarz theorem

The following theorem says that a continuous local martingale is a time changed Brownian motion.

Theorem 7.4.1. Dubins[a] – Schwarz[b] or Dambis[c] theorem.

[a]Named after Lester Dubins (1920 – 2010), American mathematician.
[b]Named after Gideon E. Schwarz (1933 – 2007), Israeli mathematician and statistician.
[c]Named after K.È. Dambis, Russian mathematician who apparently published a single article, in Russian, in 1965.

Let M be a continuous local martingale with $M_0 = 0$ and $\langle M\rangle_\infty = \infty$ almost surely. For all $t \geq 0$, let

$$T_t = \inf\{s \geq 0 : \langle M\rangle_s > t\} = \langle M\rangle_t^{-1}$$

be the generalized inverse of the non-decreasing process $\langle M\rangle$ issued from the origin. Then

1. $B = (M_{\langle M\rangle_t^{-1}})_{t\geq 0}$ is a Brownian motion with respect to the filtration $(\mathscr{F}_{T_t})_{t\geq 0}$
2. $(B_{\langle M\rangle_t})_{t\geq 0} = (M_t)_{t\geq 0}$.

These equalities are as random variables taking values in $\mathscr{C}(\mathbb{R}_+,\mathbb{R})$, in other words "a.s. for all $t \geq 0$".

For instance, if $M = \alpha W$ where $\alpha > 0$ is a constant and W is a Brownian motion issued from the origin, then $\langle M\rangle_t = \alpha^2 t$ and $T_t = \alpha^{-2} t$, and $B = (M_{T_t})_{t\geq 0} = (\alpha W_{\alpha^{-2}t})_{t\geq 0}$ is a BM with respect to $(\mathscr{F}_{\alpha^{-2}t})_{t\geq 0}$.

This theorem is as dangerous as the Skorokhod embedding theorem, it does not state that there exists a Brownian motion B with respect to the filtration for which M is a local martingale and for which $B_{\langle M\rangle} = M$.

Proof. Beware that since $\langle M\rangle$ can be flat on an interval, the map $t \mapsto T_t$ can be discontinuous. Regarding the process $B = (M_{T_t})_{t\geq 0}$, Lemma 7.4.2 states that the processes M and $\langle M\rangle$ are constant on the same intervals.

Lemma 7.4.2. Simultaneous flatness for M and $\langle M\rangle$.

Let M be a continuous local martingale. Then the processes M and $\langle M\rangle$ are constant on same intervals, in the sense that almost surely, for all $0 \leq a < b$,

$$\forall t \in [a,b], M_t = M_a \quad \text{if and only if} \quad \langle M\rangle_b = \langle M\rangle_a.$$

Let us postpone the proof of Lemma 7.4.2.

For all $t \geq 0$, the random variable T_t is a stopping time with respect to the filtration $(\mathscr{F}_u)_{u\geq 0}$, and $s \mapsto T_s$ is non-decreasing. It follows that for all $0 \leq s \leq t$, we have $\mathscr{F}_{T_s} \subset \mathscr{F}_{T_t}$, and thus $(\mathscr{F}_{T_u})_{u\geq 0}$ is a filtration. Moreover for all $t \geq 0$, the random variable T_t is a stopping time for the filtration $(\mathscr{F}_{T_u})_{u\geq 0}$. We have $T_t < \infty$ for all $t \geq 0$ on the almost sure event $\{\langle M\rangle_\infty = \infty\}$. By construction, the process $(T_t)_{t\geq 0}$ is right continuous, non-decreasing (and thus with left limits), and adapted with respect to the filtration $(\mathscr{F}_{T_t})_{t\geq 0}$. Since M is continuous, the process $B = (M_{T_t})_{t\geq 0}$ is right continuous with left limits. Moreover, for all $t \geq 0$,

$$B_{t^-} = \lim_{\substack{s \to t \\ <}} B_s = M_{T_{t^-}}.$$

Lemma 7.4.2 implies that almost surely $B_{t^-} = B_t$ for all $t \geq 0$, in other words that B is continuous.

Let us show that B is a Brownian motion for $(\mathscr{F}_{T_t})_{t\geq 0}$. For all $n \geq 0$, the process M^{T_n} is a continuous local martingale issued from the origin and $\langle M^{T_n}\rangle_\infty = \langle M\rangle_{T_n} = n$ a.s. By Lemma 4.2.7, we get that for all $n \geq 0$, the processes M^{T_n} and $(M^{T_n})^2 - \langle M\rangle^{T_n}$ are u.i. martingales. Now, for all $0 \leq s \leq t \leq n$, and by the Doob stopping theorem for u.i. martingales (Corollary 4.4.6) and the martingale property, using $T_s \leq T_t \leq T_n$,

$$\mathbb{E}(B_t \mid \mathscr{F}_{T_s}) = \mathbb{E}(M^{T_n}_{T_t} \mid \mathscr{F}_{T_s}) = M^{T_n}_{T_s} = M_{T_n \wedge T_s} = B_s$$

and similarly, using additionally the property $\langle M\rangle^{T_n}_{T_t} = \langle M\rangle_{T_n \wedge T_t} = \langle M\rangle_{T_t} = t$,

$$\mathbb{E}(B_t^2 - t \mid \mathscr{F}_{T_s}) = \mathbb{E}((M^{T_n}_{T_t})^2 - \langle M^{T_n}\rangle_{T_t} \mid \mathscr{F}_{T_s}) = (M^{T_n}_{T_s})^2 - \langle M^{T_n}\rangle_{T_s} = B_{T_s}.$$

Thus B and $(B_t^2 - t)_{t\geq 0}$ are martingales with respect to the filtration $(\mathscr{F}_{T_t})_{t\geq 0}$. It follows now from the Lévy characterization (Theorem 7.2.1) that B is a Brownian motion with respect to the filtration $(\mathscr{F}_{T_t})_{t\geq 0}$.

Let us show that $M = B_{\langle M\rangle}$. By definition of B, a.s., for all $t \geq 0$, $B_{\langle M\rangle_t} = M_{T_{\langle M\rangle_t}}$. Now $T_{\langle M\rangle_t^-} \leq t \leq T_{\langle M\rangle_t}$ and since $\langle M\rangle$ takes the same value at $T_{\langle M\rangle_t^-}$ and $T_{\langle M\rangle_t}$, we get $t = T_{\langle M\rangle_t}$ and Lemma 7.4.2 gives $M_t = M_{T_{\langle M\rangle_t}}$ for all $t \geq 0$ a.s. In other words, using the definition of B, this means that a.s., for all $t \geq 0$, $M_t = M_{T_{\langle M\rangle_t}} = B_{\langle M\rangle_t}$.

Proof of Lemma 7.4.2. Since M and $\langle M\rangle$ are continuous, it suffices to show that for all $0 \le a \le b$, a.s.

$$\{\forall t \in [a,b] : M_t = M_a\} = \{\langle M\rangle_b = \langle M\rangle_a\}.$$

The inclusion $\subset$ comes from the approximation of the quadratic variation $\langle M\rangle = [M]$. Let us prove the converse. To this end, we consider the continuous local martingale $(N_t)_{t\ge 0} = (M_t - M_{t\wedge a})_{t\ge 0}$. We have

$$\langle N\rangle = \langle M\rangle - 2\langle M, M^a\rangle + \langle M^a\rangle = \langle M\rangle - 2\langle M\rangle^a + \langle M\rangle^a = \langle M\rangle - \langle M\rangle^a.$$

For all $\varepsilon > 0$, we set the stopping time $T_\varepsilon = \inf\{t \ge 0 : \langle N\rangle_t > \varepsilon\}$. The continuous semi-martingale N^{T_ε} satisfies $N_0^{T_\varepsilon} = 0$ and $\langle N^{T_\varepsilon}\rangle_\infty = \langle N\rangle_{T_\varepsilon} \le \varepsilon$. By Lemma 4.2.7, N^{T_ε} is a martingale bounded in L^2, and for all $t \ge 0$,

$$\mathbb{E}(N^2_{t\wedge T_\varepsilon}) = \mathbb{E}(\langle N\rangle_{t\wedge T_\varepsilon}) \le \varepsilon.$$

Let us define the event $A = \{\langle M\rangle_b = \langle M\rangle_a\}$. Then $A \subset \{T_\varepsilon \ge b\}$ and, for all $t \in [a,b]$,

$$\mathbb{E}(\mathbf{1}_A N_t^2) = \mathbb{E}(\mathbf{1}_A N^2_{t\wedge T_\varepsilon}) \le \mathbb{E}(N^2_{t\wedge T_\varepsilon}) \le \varepsilon.$$

By sending ε to 0 we obtain $\mathbb{E}(\mathbf{1}_A N_t^2) = 0$ and thus $N_t = 0$ almost surely on A. ■

■

7.5 Girsanov theorem for Itô integrals

Here is a generalization to random shifts of the Cameron – Martin theorem (Theorem 3.8.2).

Theorem 7.5.1. Girsanov[a].

[a] Named after Igor Vladimirovich Girsanov (1934 – 1967), Russian mathematician. Beware that the “G” in “Girsanov” should not be spelled like in the English word “Girt” but rather like the “gh” in the Arabic word “Maghrib” spelled in Arabic.

Let $T > 0$ be deterministic. Let $B = (B_t)_{t\in[0,T]}$ be a d-dimensional BM with $B_0 = 0$. Let $\varphi = (\varphi_t)_{t\in[0,T]}$ be d-dimensional progressive and uniformly locally bounded in the sense that there exists a deterministic $C < \infty$ such that $\sup_{s\in[0,T]} |\varphi_s| \le C$ almost surely. Set $h = \int_0^\bullet \varphi_s \mathrm{d}s$. Then:

1. The Doléans-Dade exponential $N = (N_t)_{t\in[0,T]}$ defined for all $t \in [0,T]$ by

$$N_t = \exp\left(\int_0^t \varphi_s \mathrm{d}B_s - \frac{1}{2}\int_0^t |\varphi_s|^2 \mathrm{d}s\right)$$

   is a non-negative martingale and $\mathbb{E}N_t = 1$ for all $t \in [0,T]$.

2. On the changed probability space $(\Omega, \mathscr{F}, \mathbb{Q})$ where $\mathbb{Q}$ is defined by $\mathrm{d}\mathbb{Q} = N_T \mathrm{d}\mathbb{P}$, the process

$$B - h = \left(B_t - \int_0^t \varphi_s \mathrm{d}s\right)_{t\in[0,T]}$$

   is an $(\mathscr{F}_t)_{0\le t\le T}$ Brownian motion.

3. If a continuous semi-martingale $(X_t)_{t\in[0,T]}$ solves the stochastic differential equation

$$X = x + B + \int_0^\bullet b(s, X_s)\mathrm{d}s$$

   with $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ measurable and bounded, then the law of $(X_t)_{t\in[0,T]}$, as a random variable on the canonical space $\mathrm{W} = \mathscr{C}([0,T], \mathbb{R}^d)$, has a density with respect to the law of $x + B$ given by

$$w \in \mathrm{W} \mapsto \exp\left(\int_0^T b(s, x + w_s)\mathrm{d}w_s - \frac{1}{2}\int_0^T b(s, x + w_s)\mathrm{d}s\right).$$

Finally, if N is u.i. then all this holds with $[0,T]$, N_T, F_T replaced by $\mathbb{R}_+$, N_∞, F_∞.

We recover Theorem 3.8.2 of Cameron – Martin when φ is deterministic.

Note that for all $t \in [0,T]$, we have $N_t \geq 0$ and $\mathbb{E}(N_t) = 1$ and thus N_t is a density. Moreover, for any bounded $\mathscr{F}_t$ measurable random variable Z, typically a bounded and measurable function of the trajectories up to time t of a continuous adapted process, the martingale property for N gives

$$\mathbb{E}(ZN_T) = \mathbb{E}(Z\mathbb{E}(N_T \mid \mathscr{F}_t)) = \mathbb{E}(ZN_t).$$

In other words the density N_T satisfies a rule of compatibility when T varies, which is a sort of projection property, and which is directly related to the martingale property of N.

Proof.

1. The process $N = \mathrm{e}^{M - \frac{1}{2}\langle M\rangle}$ is a Doléans-Dade exponential where M is the local martingale

$$M_t = \int_0^t \varphi_s \mathrm{d}B_s, \quad \text{which satisfies} \quad \langle M\rangle_t = \int_0^t |\varphi_s|^2 \mathrm{d}s.$$

From Theorem 7.3.1, it follows that for all $\lambda \in \mathbb{R}$, $N^\lambda = \exp(\lambda M - \frac{\lambda^2}{2}\langle M\rangle)$ is a non-negative local martingale, and by Theorem 7.3.1 a super-martingale. In particular with $\lambda = 2$ we get that for all $s \geq 0$ we have $\mathbb{E}(\mathrm{e}^{2(M_s - \langle M\rangle_s)}) \leq \mathbb{E}(\mathrm{e}^{2(M_0 - \langle M\rangle_0)}) = 1$. For $\lambda = 1$ we recover N. Let us show that N is a martingale. As in the proof of Theorem 7.3.1, it suffices to show that the expectation of the angle bracket is finite, by using the fact that N solves the SDE $N_t = 1 + \int_0^t N_s \mathrm{d}M_s$. Indeed, for all $0 \leq t < T$ we have, denoting $C = \sup_{s\in[0,T]} |\varphi_s|^2$ (recall that φ is uniformly locally bounded),

$$\begin{aligned}
\mathbb{E}\langle N\rangle_t &= \mathbb{E}\langle N, N\rangle_t \\
&= \mathbb{E}\Big\langle \int_0^\bullet N_s \mathrm{d}M_s, \int_0^\bullet N_s \mathrm{d}M_s \Big\rangle_t \\
&= \mathbb{E}\int_0^t N_s^2 \mathrm{d}\langle M\rangle_s \\
&= \mathbb{E}\int_0^t \mathrm{e}^{2M_s - \langle M\rangle_s} |\varphi_s|^2 \mathrm{d}s \\
&\leq C\int_0^t \mathbb{E}\Big(\mathrm{e}^{2M_s - \langle M\rangle_s}\Big)\mathrm{d}s \\
&= C\int_0^t \mathbb{E}\Big(\mathrm{e}^{2M_s - 2\langle M\rangle_s}\mathrm{e}^{\langle M\rangle_s}\Big)\mathrm{d}s \\
&\leq C\mathrm{e}^{Ct}\int_0^t \mathbb{E}\mathrm{e}^{2M_s - 2\langle M\rangle_s}\mathrm{d}s \\
&\leq C\mathrm{e}^{Ct} t < \infty.
\end{aligned}$$

Therefore, N is a martingale thanks to Lemma 4.2.7.

2. In order to check that $B - h$ is a Brownian motion under $\mathbb{Q}$, we use Theorem 3.1.3 which reduces the problem to show that for all $\lambda \in \mathbb{R}^d$ and all fixed $T \geq 0$, the process

$$\Big(\mathrm{e}^{\lambda\cdot(B-h)_t - \frac{|\lambda|^2}{2}t}\Big)_{0\leq t\leq T}$$

is a martingale under $\mathbb{Q}$. Indeed, for all $0 \leq s < t \leq T$ and $A \in \mathscr{F}_s$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\Big(\mathbf{1}_A \mathrm{e}^{\lambda\cdot(B-h)_t - \frac{|\lambda|^2}{2}t}\Big) &= \mathbb{E}\Big(\mathbf{1}_A \mathrm{e}^{\lambda\cdot B_t - \lambda\cdot\int_0^t \varphi_u \mathrm{d}u - \frac{|\lambda|^2}{2}t} N_T\Big) \\
&\overset{\star}{=} \mathbb{E}\Big(\mathbf{1}_A \mathrm{e}^{\lambda\cdot B_t - \lambda\cdot\int_0^t \varphi_u \mathrm{d}u - \frac{|\lambda|^2}{2}t} N_t\Big) \\
&= \mathbb{E}\Big(\mathbf{1}_A \mathrm{e}^{\int_0^t (\lambda+\varphi_u)\cdot\mathrm{d}B_u - \frac{1}{2}\int_0^t |\lambda+\varphi_u|^2 \mathrm{d}u}\Big) \\
&\overset{\star\star}{=} \mathbb{E}\Big(\mathbf{1}_A \mathrm{e}^{\int_0^s (\lambda+\varphi_u)\cdot\mathrm{d}B_u - \frac{1}{2}\int_0^s |\lambda+\varphi_u|^2 \mathrm{d}u}\Big) \\
&= \mathbb{E}\Big(\mathbf{1}_A \mathrm{e}^{\lambda\cdot B_s - \lambda\cdot\int_0^s \varphi_u \mathrm{d}u - \frac{|\lambda|^2}{2}s} N_s\Big)
\end{aligned}$$

$$
\begin{aligned}
&\overset{\star}{=} \mathbb{E}\Big(\mathbf{1}_A \mathrm{e}^{\lambda\cdot B_s - \lambda\cdot\int_0^s \varphi_u \mathrm{d}u - \frac{|\lambda|^2}{2}s} N_T\Big)\\
&= \mathbb{E}_{\mathbb{Q}}\Big(\mathbf{1}_A \mathrm{e}^{\lambda\cdot(B-h)_s - \frac{|\lambda|^2}{2}s}\Big),
\end{aligned}
$$

where we have used in $\star$ the fact that the process N is a martingale (under $\mathbb{P}$) and in $\star\star$ the fact that the process N (with φ replaced by $\lambda+\varphi$) is a martingale (under $\mathbb{P}$).

3. We give the proof when b is continuous and bounded. We can assume without loss of generality that $x = 0$, otherwise we use $X_t - x = B_t + \int_0^t b(s, x + X_s - x)\mathrm{d}s$. We have $\varphi_t = -b(t, X_t)$. Since the shift $h = \int_0^\bullet \varphi_s \mathrm{d}s$ is random, the simple argument used in the proof of Corollary 3.8.3 (Cameron–Martin) in order to get the density of the law of $B+h$ with respect to the law of B is no longer available. Nevertheless, for all bounded measurable $\Phi : \mathrm{W} \to \mathbb{R}$,

$$\mathbb{E}(\Phi(B-h)) = \mathbb{E}(\Phi(B-h)N_T^{-1}N_T).$$

Let us show now that N_T^{-1} is a function of $X = B - h$. We have $B = X + h = X + \int_0^\bullet \varphi_s \mathrm{d}s$ and thus

$$\int_0^T \varphi_s \mathrm{d}B_s = \int_0^T \varphi_s \mathrm{d}X_s + \int_0^T |\varphi_s|^2 \mathrm{d}s,$$

hence

$$
\begin{aligned}
N_T^{-1} &= \exp\Big(-\int_0^T \varphi_s \mathrm{d}B_s + \frac{1}{2}\int_0^T |\varphi_s|^2 \mathrm{d}s\Big)\\
&= \exp\Big(-\int_0^T \varphi_s \mathrm{d}X_s - \frac{1}{2}\int_0^T |\varphi_s|^2 \mathrm{d}s\Big)\\
&= \exp\Big(\int_0^T b(X_s)\mathrm{d}X_s - \frac{1}{2}\int_0^T |b(X_s)|^2 \mathrm{d}s\Big)\\
&= \Psi(X) = \Psi(B-h).
\end{aligned}
$$

The function Ψ is bounded. We admit that it is measurable. Therefore, using the previous result in $\star$,

$$\mathbb{E}(\Phi(B-h)) = \mathbb{E}(\Phi(B-h)N_T^{-1}N_T) = \mathbb{E}(\Phi(B-h)\Psi(B-h)N_T) = \mathbb{E}_{\mathbb{Q}}(\Phi(B-h)\Psi(B-h)) \overset{\star}{=} \mathbb{E}(\Phi(B)\Psi(B)).$$

Hence the density of X on the canonical space W with respect to the Wiener measure is Ψ.

Finally, if N is u.i. then by Corollary 4.4.5 there exists $N_\infty \in \mathrm{L}^1$ such that $\lim_{t\to\infty} N_t = N_\infty$ a.s. and in L^1, and $N_t = \mathbb{E}(N_\infty \mid \mathscr{F}_t)$ for all $t \geq 0$. This implies that $(\mathrm{e}^{\lambda\cdot(B-h)_t - \frac{|\lambda|^2}{2}t})_{t\geq 0}$ is a martingale for all $\lambda \in \mathbb{R}^d$ under the probability measure $\mathbb{Q}$ on $(\Omega, \mathscr{F})$ absolutely continuous with respect to $\mathbb{P}$ and with density N_∞. ■

Coding in action 7.5.2. Usage in statistics.

Suppose that a phenomenon is modeled with an unknown function $b : \mathbb{R}^d \to \mathbb{R}$, and the observation by a process $(X_t)_{t\geq 0}$ solution of the stochastic differential equation $X_t = x + B_t + \int_0^t b(X_s)\mathrm{d}s$ where x is known and where B is an unobserved BM modeling noise. We would like to estimate b from the observation of X, or test the hypothesis that $b = 0$. The Girsanov theorem (Theorem 7.5.1) provides the likelihood of the observations! Of course life is more complicated because in practice, we observe X only at a finite number of discrete times. This type of stochastic calculus in action belongs to the field of statistical analysis of diffusion processes, see [22, 33, 34, 29]. Could you write a computer program simulating approximate trajectories for a given b and approximating numerically their likelihood?

7.6 Sub-Gaussian tail bound and exponential square integrability for local martingales

Theorem 7.6.1. Sub-Gaussian tail bound and exponential square integrability.

Let $M=(M_t)_{t\geq 0}$ be a continuous local martingale issued from the origin. Then for all $t,K,r\geq 0$,

$$\mathbb{P}\Big(\sup_{s\in[0,t]}|M_s|\geq r \text{ and } \langle M\rangle_t\leq K\Big)\leq 2\mathrm{e}^{-\frac{r^2}{2K}},$$

and in particular, if $\langle M\rangle_t\leq Kt$ then

$$\mathbb{P}\Big(\sup_{s\in[0,t]}|M_s|\geq r\Big)\leq 2\mathrm{e}^{-\frac{r^2}{2Kt}}\quad\text{and}\quad\mathbb{E}\Big(\mathrm{e}^{\alpha\sup_{s\in[0,t]}|M_s|^2}\Big)<\infty\quad\text{for all}\quad\alpha<\frac{1}{2Kt}.$$

The condition on $\langle M\rangle_t$ is a comparison to Brownian motion B for which $\langle B\rangle_t=t$.

Proof. Let us prove the first inequality. For all $\lambda,t\geq 0$, by Remark 7.3.1, the process

$$X^\lambda=\Big(\mathrm{e}^{\lambda M_t-\frac{\lambda^2}{2}\langle M\rangle_t}\Big)_{t\geq 0}.$$

is a positive super-martingale issued from 1 and $\mathbb{E}X_t^\lambda\leq 1$ for all $t,\lambda\geq 0$. Therefore, for all $t,\lambda,r,K\geq 0$,

$$\begin{aligned}\mathbb{P}\Big(\langle M\rangle_t\leq K,\ \sup_{0\leq s\leq t}M_s\geq r\Big)&\leq\mathbb{P}\Big(\langle M\rangle_t\leq K,\ \sup_{0\leq s\leq t}X_s^\lambda\geq\mathrm{e}^{\lambda r-\frac{\lambda^2}{2}K}\Big)\\&\leq\mathbb{P}\Big(\sup_{0\leq s\leq t}X_s^\lambda\geq\mathrm{e}^{\lambda r-\frac{\lambda^2}{2}K}\Big)\\&\leq\mathbb{E}(X_0^\lambda)\mathrm{e}^{-\lambda r+\frac{\lambda^2}{2}K}=\mathrm{e}^{-\lambda r+\frac{\lambda^2}{2}K}\end{aligned}$$

where the last step comes from the maximal inequality (Theorem 2.5.9) for X^λ. Taking $\lambda=r/K$ gives

$$\mathbb{P}\Big(\langle M\rangle_t\leq K,\ \sup_{0\leq s\leq t}M_s\geq r\Big)\leq\mathrm{e}^{-\frac{r^2}{2K}}.$$

The same reasoning provides (note by the way that $\langle -M\rangle=\langle M\rangle$ obviously)

$$\mathbb{P}\Big(\langle M\rangle_t\leq K,\ \sup_{0\leq s\leq t}(-M_s)\geq r\Big)\leq\mathrm{e}^{-\frac{r^2}{2K}}.$$

The desired result follows now by the union bound, hence the factor 2 in the right hand side.

Finally, the exponential square integrability comes from the usual link between tail bound and integrability, namely if $X=\sup_{s\in[0,t]}|M_s|$, $U(x)=\mathrm{e}^{\alpha x^2}$, $\alpha<\frac{1}{2Kt}$, then, by the Fubini – Tonelli theorem,

$$\mathbb{E}(U(X))=\mathbb{E}\Big(\int_0^X U'(x)\mathrm{d}x\Big)=\mathbb{E}\Big(\int_0^\infty\mathbf{1}_{x\leq X}U'(x)\mathrm{d}x\Big)=\int_0^\infty U'(x)\mathbb{P}(X\geq x)\mathrm{d}x\leq\int_0^\infty 2\alpha x\mathrm{e}^{\alpha x^2}\mathrm{e}^{-\frac{x^2}{2Kt}}\mathrm{d}x<\infty.$$

■

7.7 Burkholder – Davis – Gundy inequalities

These inequalities allow to control the moments of the sup of a local martingale via the moments of its angle bracket. This is useful for stochastic integrals, and in particular for stochastic differential equations.

Theorem 7.7.1. Burkholder[a] – David[b] – Gundy[c] inequalities.

[a]Named after Donald Burkholder (1927 – 2013), American mathematician.
[b]Named after Burgess Davis, American mathematician.
[c]Named after Richard F. Gundy, American mathematician.

For all $p>0$ there exists universal constants $c_p, C_p < \infty$ such that for all continuous local martingale $M=(M_t)_{t\geq 0}$ issued form the origin, for all fixed $T\geq 0$, the following inequalities hold in $[0,+\infty]$:

$$c_p\mathbb{E}\Big(\sup_{t\in[0,T]}|M_t|^{2p}\Big)\leq \mathbb{E}(\langle M\rangle_T^p)\leq C_p\mathbb{E}\Big(\sup_{t\in[0,T]}|M|^{2p}\Big).$$

The essential ingredients of the proof are Doob maximal inequality and Itô formula.

This can be skipped at first reading.

Proof. Let us fix $T>0$ and set $\|M\|_T=\sup_{0\leq t\leq T}|M_t|$. We have, almost surely as $n\to\infty$,

$$T_n=\inf\{t\geq 0:|M_t|\geq n \text{ or } |\langle M\rangle_t|\geq n\}\nearrow +\infty.$$

By Lemma 4.2.6, for all $n\geq 0$, M^{T_n} is a continuous martingale and

$$\sup_{t\geq 0}|M_t^{T_n}|\leq n \quad\text{and}\quad \sup_{t\geq 0}\langle M^{T_n}\rangle\leq n.$$

Moreover, almost surely,

$$\langle M^{T_n}\rangle_T=\langle M\rangle_{T\wedge T_n}\underset{n\to\infty}{\nearrow}\langle M\rangle_T \quad\text{and}\quad \|M^{T_n}\|_T=\|M\|_{T\wedge T_n}\underset{n\to\infty}{\nearrow}\|M\|_T.$$

Now if the BDG inequality is satisfied by M^{T_n} for all $n\geq 0$ then, by monotone convergence, it is also satisfied by M. Therefore, from now on, we can assume without loss of generality that both M and $\langle M\rangle$ are bounded. Note that the constants must be universal, depending on p but of course neither on M nor on T.
Note at this stage that the Doob maximal inequality of Theorem 2.5.7 gives, for all $r>1$,

$$\mathbb{E}(\|M\|_T^r)\leq\Big(\frac{r}{r-1}\Big)^r\mathbb{E}(|M_T|^r).$$

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 $M_T^2\leq\|M\|_T^2$).
Case $p>1$. We have, from the Itô formula (Theorem 7.1.1) for $f(x)=|x|^{2p}$ and $X=M$, for all $t\geq 0$,

$$|M_t|^{2p}=2p\int_0^t|M_s|^{2p-1}\mathrm{sign}(M_s)\mathrm{d}M_s+p(2p-1)\int_0^t|M_s|^{2p-2}\mathrm{d}\langle M\rangle_s.$$

Despite appearances, f is $\mathscr{C}^2$. Indeed, to clarify the situation at the origin, we have, since $2p-1>0$,

$$f'(0)=\lim_{x\to 0}\frac{f(x)-f(0)}{x-0}=\lim_{x\to 0}\frac{|x|^{2p}}{x}=\mathrm{sign}(x)|x|^{2p-1}\underset{x\to 0}{\longrightarrow}0,$$

and, since $2p-2>0$,

$$f''(0)=\lim_{x\to 0}\frac{f'(x)-f'(0)}{x-0}=\lim_{x\to 0}\frac{|x|^{2p-1}}{|x|}=|x|^{2p-2}\underset{x\to 0}{\longrightarrow}0.$$

In the first (stochastic) integral in the right hand side in the display above, the integrand $|M|^{2p-1}\mathrm{sign}(M))$ is continuous and bounded while the integrator M belongs to $\mathbb{M}_0^2$ (recall that $M_0=0$ and M is bounded). As a consequence, this stochastic integral is a martingale issued from the origin and is therefore centered. Hence, for all $0\leq t\leq T$, taking expectations and using the Hölder inequality with p and $q=1/(1-1/p)=p/(p-1)$,

$$\mathbb{E}(|M_t|^{2p})=p(2p-1)\mathbb{E}\int_0^t|M_s|^{2p-2}\mathrm{d}\langle M\rangle_s$$

$$\leq p(2p-1)\mathbb{E}(\|M\|_T^{2(p-1)}\langle M\rangle_T)$$
$$\leq p(2p-1)\mathbb{E}(\|M\|_T^{2p})^{1-1/p}(\mathbb{E}(\langle M\rangle_T^p)^{1/p}.$$

Combined with the maximal inequality above used with $r=2p$, we obtain the second BGD inequality. To prove the first BGD inequality, we use the Itô formula (Theorem 7.1.1) with $f(x_1,x_2)=x_1x_2$, $X=(M,\langle M\rangle^{(p-1)/2})$,

$$M_t\langle M\rangle_t^{(p-1)/2}=\int_0^t\langle M\rangle_s^{(p-1)/2}\mathrm{d}M_s+\int_0^t M_s\mathrm{d}(\langle M\rangle_s^{(p-1)/2})$$

(note that there is no second order term here since either $\partial^2_{i,j}f=0$ or $\langle X^i,X^j\rangle=0$). Now, if we define

$$N_t=\int_0^t\langle M\rangle_s^{(p-1)/2}\mathrm{d}M_s,$$

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 $q=1/(1-1/p)=p/(p-1)$,

$$\mathbb{E}(N_t^2)\leq 4\mathbb{E}(\|M\|_T^2\langle M\rangle_T^{p-1})\leq 4(\mathbb{E}(\|M\|_T^{2p}))^{1/p}(\mathbb{E}(\langle M\rangle_T^p)^{1-1/p}.$$

Combined with

$$\mathbb{E}(N_t^2)=\mathbb{E}\int_0^t\langle M\rangle_s^{p-1}\mathrm{d}\langle M\rangle_s=\frac{1}{p}\mathbb{E}(\langle M\rangle_t^p)$$

we obtain

$$\mathbb{E}(\langle M\rangle_T^p)\leq(4p)^p\mathbb{E}(\|M\|_T^{2p}),$$

which is the first BGD inequality.

Case $0<p<1$. Let us define $N_t=\int_0^t\langle M\rangle_s^{(p-1)/2}\mathrm{d}M_s$. We have

$$M_t=\int_0^t\mathrm{d}M_s=\int_0^t\langle M\rangle_s^{(1-p)/2}\langle M\rangle_s^{(p-1)/2}\mathrm{d}s=\int_0^t\langle M\rangle_s^{(1-p)/2}\mathrm{d}N_s$$

and

$$N_t\langle M\rangle_t^{(1-p)/2}=\int_0^t\langle M\rangle_s^{(1-p)/2}\mathrm{d}N_s+\int_0^t N_s\mathrm{d}(\langle M\rangle_s^{(1-p)/2})$$
$$=M_t+\int_0^t N_s\mathrm{d}(\langle M\rangle_s^{(1-p)/2}).$$

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

$$|M_t|\leq 2\|N\|_T\langle M\rangle_T^{(1-p)/2}\quad\text{and}\quad\|M\|_T\leq 2\|N\|_T\langle M\rangle_T^{(1-p)/2},$$

thus, using the Hölder inequality with $1/p$ and its conjugate exponent $1/(1-p)$,

$$\mathbb{E}(\|M\|_T^{2p})\leq 4^p\mathbb{E}(\|N\|_T^{2p}\langle M\rangle_T^{p(1-p)})$$
$$\leq(4^p)^2(\mathbb{E}(\|N_T\|^2))^p(\mathbb{E}(\langle M\rangle_T^p)^{1-p}$$
$$\leq(4^p)^2(\mathbb{E}(N_T^2))^p(\mathbb{E}(\langle M\rangle_T^p)^{1-p}$$
$$=16^p(p^{-1}\mathbb{E}(\langle M\rangle_T^p))^p(\mathbb{E}(\langle M\rangle_T^p))^{1-p}$$
$$=\left(\frac{16}{p}\right)^p\mathbb{E}(\langle M\rangle_T^p).$$

This proves the first BGD inequality. To prove the second BGD inequality, let $\alpha > 0$ (the reason for α is to avoid the singularity at 0 of $x \mapsto x^{p-1}$ due to $p-1<0$). Now write, using the Itô formula (Theorem 7.1.1),

$$\begin{aligned} M_t(\alpha+\|M\|_t)^{p-1} &= \int_0^t (\alpha+\|M\|_s)^{p-1}\mathrm{d}M_s + \int_0^t M_s\mathrm{d}(\alpha+\|M\|_s)^{p-1} \\ &= N_t + (p-1)\int_0^t M_s(\alpha+\|M\|_s)^{p-2}\mathrm{d}\|M\|_s \end{aligned}$$

where $N_t = \displaystyle\int_0^t (\alpha+\|M\|_s)^{p-1}\mathrm{d}M_s$. We have then (taking $\alpha \to 0$)

$$|N_t| \le \|M_t\|^p + (1-p)\int_0^t \|M\|_s^{p-1}\mathrm{d}\|M\|_s = \frac{1}{p}\|M\|_t^p$$

and thus

$$\mathbb{E}\int_0^t (\alpha+\|M\|_s)^{2(p-1)}\mathrm{d}\langle M\rangle_s = \mathbb{E}(N_t^2) \le \frac{1}{p^2}\mathbb{E}(\|M\|_t^{2p}),$$

which gives finally the inequality (recall that $2(1-p)<0$)

$$\mathbb{E}((\alpha+\|M\|_t)^{2(p-1)}\langle M\rangle_t) \le \frac{1}{p^2}\mathbb{E}(\|M\|_t^{2p}).$$

But the identity

$$\langle M\rangle_t^p = (\langle M\rangle_t^p(\alpha+\|M\|_t)^{2p(p-1)})(\alpha+\|M\|_t)^{2p(1-p)}$$

gives, using the Hölder inequality with $1/p$ and its conjugate exponent $1/(1-p)$, that

$$\begin{aligned} \mathbb{E}(\langle M\rangle_t^p) &\le (\mathbb{E}(\langle M\rangle_t(\alpha+\|M\|_t)^{2(p-1)}))^p(\mathbb{E}((\alpha+\|M\|_t)^{2p}))^{1-p} \\ &\le \left(\frac{1}{p^2}\right)^p (\mathbb{E}(\|M\|_t^{2p}))^p(\mathbb{E}((\alpha+\|M\|_t)^{2p}))^{1-p}. \end{aligned}$$

Taking the limit as $\alpha \to 0$, we obtain

$$\mathbb{E}(\langle M\rangle_t^p) \le \frac{1}{p^{2p}}\mathbb{E}(\|M\|_t^{2p})$$

which is the second BGD inequality. ■

7.8 Representation of Brownian functionals and martingales as stochastic integrals

Let $B=(B_t)_{t\ge0}$ be a d-dimensional Brownian motion, and let $\varphi=(\varphi_t)_{t\ge0}$ be a d-dimensional progressive process with respect to the completed filtration of B, such that $\mathbb{E}\int_0^\infty|\varphi_s|^2\mathrm{d}s<\infty$. Then the stochastic integral $\int_0^\infty\varphi_s\mathrm{d}B_s$ is a measurable function of B seen as a random variable with values on the Wiener space. Indeed it is the limit in probability of finite sums which are measurable functions of B. Conversely, the following theorem states that every measurable function of B is the stochastic integral of a progressive process.

Theorem 7.8.1. Representation of Brownian functionals and martingales as stochastic integrals.

Let $B=(B_t)_{t\ge0}$ be a d-dimensional Brownian motion issued from the origin. Let $(\mathscr{F}_t)_{t\ge0}$ be its completed natural filtration, and let $\mathscr{F}_\infty=\sigma(\cup_{t\ge0}\mathscr{F}_t)$.

1. For all square integrable random variable $Z\in\mathrm{L}^2(\Omega,\mathscr{F}_\infty,\mathbb{P})$, there exists a unique progressive d-dimensional process $\varphi=(\varphi_t)_{t\ge0}$ such that $\mathbb{E}\int_0^\infty|\varphi_t|^2\mathrm{d}t<\infty$ and

$$Z=\mathbb{E}(Z)+\int_0^\infty\varphi_s\mathrm{d}B_s.$$

2. If M is an $(\mathcal{F}_t)_{t\geq 0}$ martingale (continuous or not) bounded in L^2 and issued from the origin then there exists a unique progressive $\varphi = (\varphi_t)_{t\geq 0}$ with $\mathbb{E}\int_0^\infty |\varphi_s|^2 \mathrm{d}s < \infty$ and, for all $t \geq 0$,
$$M_t = \int_0^t \varphi_s \mathrm{d}B_s.$$
3. If M is an $(\mathcal{F}_t)_{t\geq 0}$ continuous local martingale issued from the origin then there exists a unique progressive $\varphi = (\varphi_t)_{t\geq 0}$ such that for all $t \geq 0$, $\int_0^t |\varphi_s|^2 \mathrm{d}s < \infty$ a.s. and
$$M_t = \int_0^t \varphi_s \mathrm{d}B_s.$$

Being measurable for $\mathcal{F}_\infty$ means being a measurable function of B, and we say Brownian functional.

In the second item, the continuous process $\int_0^\bullet \varphi_s \mathrm{d}B_s$ can be seen as a continuous modification of M. if M is continuous then M and $\int_0^\bullet \varphi_s \mathrm{d}B_s$ are equal as random variables on the canonical space.

This can be skipped at first reading.

Proof.

1. The uniqueness follows from the representation with the progressive processes φ and φ' such that $\mathbb{E}\int_0^\infty |\varphi_t|^2 \mathrm{d}t < \infty$ and $\mathbb{E}\int_0^\infty |\varphi'_t|^2 \mathrm{d}t < \infty$, then $\varphi = \varphi'$ since the Itô isometry gives
$$\mathbb{E}\int_0^\infty (\varphi_s - \varphi'_s)^2 \mathrm{d}s = \mathbb{E}\Big(\Big(\int_0^\infty \varphi_s \mathrm{d}B_s - \int_0^\infty \varphi'_s \mathrm{d}B_s\Big)^2\Big) = 0.$$
Let us prove the existence. Let us consider the sub-vector space
$$F = \Big\{Z \in \mathrm{L}^2(\Omega, \mathcal{F}_\infty, \mathbb{P}) : Z = \mathbb{E}(Z) + \int_0^\infty \varphi_s \mathrm{d}B_s \text{ for a progressive } \varphi = (\varphi_t)_{t\geq 0} \text{ with } \mathbb{E}\int_0^\infty |\varphi_s|^2 \mathrm{d}s < \infty\Big\}.$$
For all Z and Z' in F, if φ and φ' are the associated progressive processes, by the Itô isometry,
$$\mathbb{E}(|Z - Z'|^2) = |\mathbb{E}Z - \mathbb{E}Z'|^2 + \mathbb{E}\int_0^\infty |\varphi_s - \varphi'_s|^2 \mathrm{d}s.$$
Hence F is a sub-Hilbert space of $\mathrm{L}^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$. In order to show that $F = \mathrm{L}^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$, it suffices to show that $F^\perp = \{0\}$. Note that F contains all the random variables of the form
$$Z = \exp\Big(\int_0^\infty \varphi_s \mathrm{d}B_s - \frac{1}{2}\int_0^\infty |\varphi_s|^2\Big)$$
where φ is deterministic and such that $\int_0^\infty |\varphi_s|^2 \mathrm{d}s < \infty$. Indeed, if we define, for all $t \geq 0$,
$$X_t = \exp\Big(\int_0^t \varphi_s \mathrm{d}B_s - \frac{1}{2}\int_0^t |\varphi_s|^2 \mathrm{d}s\Big)$$
then $Z = X_\infty$. Since it is a Doléans-Dade exponential (see Theorem 7.3.1), for all $t \in [0, \infty]$,
$$X_t = 1 + \int_0^t X_s \mathrm{d}B_s \quad \text{and in particular} \quad \mathbb{E}(Z) = \mathbb{E}(X_\infty) = 1 + \mathbb{E}\int_0^\infty X_s \mathrm{d}B_s,$$
which means that $Z \in F$. Let $Y \in F^\perp$.
We have then, for all deterministic φ such that $\int_0^\infty |\varphi_s|^2 \mathrm{d}s < \infty$,
$$\mathbb{E}\Big(Y \exp\Big(\int_0^\infty \varphi_s \mathrm{d}B_s\Big)\Big) = \mathbb{E}\Big(Y \exp\Big(\int_0^\infty \varphi_s \mathrm{d}B_s - \frac{1}{2}\int_0^\infty |\varphi_s|^2 \mathrm{d}s\Big)\Big) \exp\Big(\frac{1}{2}\int_0^\infty |\varphi_s|^2 \mathrm{d}s\Big) = 0.$$

Used with $h = \sum_{k=1}^n \lambda_k \mathbf{1}_{(s_k,s_{k+1}]}$, for arbitrary $n \geq 1$, $\lambda_1,\dots,\lambda_n \in \mathbb{R}$, and $0 \leq s_1 < \cdots < s_n$, this gives

$$\mathbb{E}\Big(Y \mathrm{e}^{\lambda_1 B_{s_1}+\cdots+\lambda_n B_{s_n}}\Big) = 0.$$

Thus, by analytic continuation,

$$\mathbb{E}\Big(Y \mathrm{e}^{\mathrm{i}\lambda_1 B_{s_1}+\cdots+i\lambda_n B_{s_n}}\Big) = 0.$$

Also

$$\mathbb{E}\Big(\mathbb{E}\Big(Y \mid (B_{s_1},\dots,B_{s_n})\mathrm{e}^{\mathrm{i}\sum_{k=1}^n \lambda_k B_{s_k}}\Big)\Big) = 0.$$

Since this is valid for all $\lambda_1,\dots,\lambda_n \in \mathbb{R}$, we get

$$\mathbb{E}(Y \mid (B_{s_1},\dots,B_{s_n})) = 0.$$

We have then, for all $n \geq 1$, all $0 \leq s_1 < \cdots < s_n$, and all bounded measurable $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}(Y f(B_{s_1},\dots,B_{s_n})) = 0.$$

Finally it follows by the monotone class theorem that $Y = 0$.

2. Since M is bounded in L^2, it is u.i. and $M_\infty \in \mathrm{L}^2(\Omega,\mathscr{F}_\infty,\mathbb{P})$, and from the first part

$$M_\infty = \mathbb{E}(M_\infty) + \int_0^\infty \varphi_s \mathrm{d}B_s$$

where $\varphi = (\varphi_t)_{t\geq 0}$ is progressive and such that $\mathbb{E}\int_0^\infty |\varphi_s|^2 \mathrm{d}s < \infty$. Now, $\mathbb{E}(M_\infty) = \mathbb{E}(M_0) = 0$, while by the martingale property of the stochastic integral, for all $t \geq 0$,

$$M_t = \mathbb{E}(M_\infty \mid \mathscr{F}_t) = \int_0^t \varphi_s \mathrm{d}B_s.$$

The uniqueness of φ follows from its uniqueness in the decomposition of M_∞.

3. For all $n \geq 0$, let $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$. The preceding item used for the martingale $M^{T_n} = (M_{t\wedge T_n})_{t\geq 0}$ which is bounded in L^2 gives

$$M_{t\wedge T_n} = \int_0^t \varphi_s^{(n)} \mathrm{d}B_s$$

for a progressive $\varphi^{(n)}$ such that $\mathbb{E}\int_0^\infty |\varphi_s^{(n)}|^2 \mathrm{d}s < \infty$. The uniqueness of the progressive process gives, for all $m < n$, $\varphi_s^{(m)} = \mathbf{1}_{[0,T_m]}(s)\varphi_s^{(n)}$ in $\mathrm{L}^2(\Omega\times\mathbb{R}_+,\mathscr{F}_\infty\otimes\mathscr{B}_{\mathbb{R}_+},\mathbb{P}\otimes\mathrm{d}s)$. This allows to construct a (unique) process φ such that for all $t \geq 0$, almost surely $\int_0^t |\varphi_s|^2 \mathrm{d}s < \infty$, and $M_t = \int_0^t \varphi_s \mathrm{d}B_s$.

■

Corollary 7.8.2. Filtration of Brownian motion and martingale regularization.

Let $B = (B_t)_{t\geq 0}$ be a d-dimensional BM with $B_0 = 0$ and let $(\mathscr{F}_t)_{t\geq 0}$ be its completed filtration. Then:

1. The filtration $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous and left-continuous in the sense that for all $t \geq 0$,

$$\mathscr{F}_t = \mathscr{F}_{t+} = \cap_{s>t}\mathscr{F}_s \quad \text{and} \quad \mathscr{F}_t = \mathscr{F}_{t-} = \sigma(\cup_{s\in[0,t)}\mathscr{F}_s)$$

2. If M is a martingale with respect to $(\mathscr{F}_t)_{t\geq 0}$, then it admits a continuous modification.

Proof.

1. Let us prove right-continuity. Let $t \geq 0$ and let Z be a bounded and $\mathscr{F}_{t+}$ measurable random variable. By Theorem 7.8.1 used with $d = 1$, there exists a progressive process φ with respect to $(\mathscr{F}_t)_{t\geq 0}$ such that $\mathbb{E}\int_0^\infty \varphi_s^2 \mathrm{d}s < \infty$ and $Z = \mathbb{E}Z + \int_0^\infty \varphi_s \mathrm{d}B_s$. For all $\varepsilon > 0$, the random variable Z in $\mathscr{F}_{t+\varepsilon}$ measurable and by the martingale property of the stochastic integral,

$$Z = \mathbb{E}(Z \mid \mathscr{F}_{t+\varepsilon}) = \mathbb{E}(Z) + \int_0^{t+\varepsilon} \varphi_s \mathrm{d}B_s \xrightarrow[\varepsilon\to 0]{\mathrm{L}^2} \mathbb{E}(Z) + \int_0^t \varphi_s \mathrm{d}B_s.$$

   Thus Z is equal (as a random variable: a.s.) to an $\mathscr{F}_t$-measurable random variable. Since the filtration is complete, this means that Z is $\mathscr{F}_t$-measurable. A similar argument works to prove left-continuity.

2. If M is bounded in L^2, this follows from the representation in Theorem 7.8.1. The proof of the general case is not very difficult but takes a page, and we can find it for instance in [31, p. 130-131].

■

Chapter 8

Stochastic differential equations

Let $B = (B_t)_{t\geq 0}$ be a d-dimensional (column vector) $(\mathscr{F}_t)_{t\geq 0}$ Brownian motion issued from the origin. Let $\mathscr{M}_{q,d}(\mathbb{R})$ be the set of $q \times p$ matrices with entries in $\mathbb{R}$.
The Hilbert–Schmidt norm of $A \in \mathscr{M}_{q,d}(\mathbb{R})$ is $|A| = (\sum_{i=1}^q \sum_{j=1}^d |A_{i,j}|^2)^{1/2} = (\sum_{i=1}^q |A_{i,\bullet}|^2)^{1/2}$.
We have seen in Theorem 7.1.1 that for all $\lambda \in \mathbb{R}^d$, the Doléans-Dade exponential $(\mathrm{e}^{\lambda\cdot B_t - \frac{|\lambda|^2}{2}t})_{t\geq 0}$ satisfies

$$X_0 = 1 \quad \text{and} \quad X_t = 1 + \int_0^t X_s \mathrm{d}(\lambda \cdot B_s) \quad \text{for all } t \geq 0.$$

The present chapter is devoted to the study of far more general stochastic differential equations (SDE).

8.1 Stochastic differential equations with general coefficients

We seek for a q-dimensional process $X = (X_t)_{t\geq s}$ solution of the stochastic differential equation

$$X_t = \eta + \int_s^t \sigma(u, X_u)\mathrm{d}B_u + \int_s^t b(u, X_u)\mathrm{d}u \quad \text{a.s.}, \quad t \geq s. \tag{SDE}$$

Here

- $s \geq 0$ is the initial time
- η is a q-dimensional random vector playing the role of initial value or initial condition or initial data
- the function $\sigma : \mathbb{R}_+ \times \Omega \times \mathbb{R}^q \to \mathscr{M}_{q,d}(\mathbb{R})$ plays the role of a diffusion matrix
- the function $b : \mathbb{R}_+ \times \Omega \times \mathbb{R}^q \to \mathbb{R}^q$ plays the role of a drift.

For the intuition, the best is to think about X_t as the position of a particle in $\mathbb{R}^d$ at time t. This physical picture is made more precise in our study of the Langevin stochastic process (Example 8.2.7).

The Doléans-Dade exponential corresponds to $q = 1$, $\eta = 1$, $\sigma = \lambda$ (constant row vector), and $b = 0$. Another basic example is given by the Ornstein–Uhlenbeck process (Example 8.2.2).

We say that (SDE) is driven by B. We can interpret (SDE) either as a deformation of Brownian motion, or as an Ordinary Differential Equation (ODE) with noise[1]. Note that (SDE) means that for all $1 \leq j \leq q$,

$$\begin{aligned} X_t^j &= \eta_j + \int_s^t \sigma_{j\bullet}(u, X_u)\mathrm{d}B_u + \int_s^t b_j(u, X_u)\mathrm{d}u \quad \text{a.s.}, \quad t \geq s \\ &= \eta_j + \sum_{k=1}^d \int_s^t \sigma_{jk}(u, X_u)\mathrm{d}B_u^k + \int_s^t b_j(u, X_u)\mathrm{d}u \quad \text{a.s.}, \quad t \geq s. \end{aligned}$$

In other words, in differential notations,

$$X_s = \eta, \quad \mathrm{d}X_t = \sigma(t, X_t)\mathrm{d}B_t + b(t, X_t)\mathrm{d}t \quad \text{a.s.}, \quad t \geq s,$$

[1] This ODE with noise point of view leads sometimes to put the noise at the end, namely $\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}B_t$.

in other words

$$X_s^j = \eta^j, \quad \mathrm{d}X_t^j = \sigma_{j,\bullet}(t, X_t)\mathrm{d}B_t + b_j(t, X_t)\mathrm{d}t \quad \text{a.s.}, \quad t \geq s,$$
$$= \sum_{k=1}^{d} \sigma_{j,k}(t, X_t)\mathrm{d}B_t^k + b_j(t, X_t)\mathrm{d}t \quad \text{a.s.}, \quad t \geq s,$$

Note that σ and b can be random and may for instance depend on B on the canonical space. We will sometimes make explicit or not the dependency over ω, namely $\sigma(t,\omega,x)$ and $b(t,\omega,x)$ or $\sigma(t,x)$ and $b(t,x)$.

For a given law of initial condition η, we say that we have...

- existence in law (or weak existence) when (SDE) has a solution on some filtered probability space and with some BM defined on it which are not necessarily the ones for which (SDE) is stated initially
- uniqueness in law (or weak uniqueness) when additionally all solutions of (SDE) (not necessarily on the same probability space or with the same Brownian motion) have same law on $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^q)$
- pathwise uniqueness (or strong uniqueness) when two solutions of (SDE) defined on the same probability space and with the same Brownian motion are indistinguishables.

For solving (SDE), we assume that the following properties hold for σ and b. Such conditions are natural, having in mind the (stochastic) nature of the solution, and how we solve basic deterministic ODEs.

- **(Lip) Lipschitz regularity in x uniformly in ω and t.** There exists a constant $c > 0$ such that for all $(t,\omega) \in \mathbb{R}_+ \times \Omega$ and all $x, y \in \mathbb{R}^q$,

$$|\sigma(t,\omega,x) - \sigma(t,\omega,y)| \leq c|x-y| \quad \text{and} \quad |b(t,\omega,x) - b(t,\omega,y)| \leq c|x-y|$$

- **(Mes) Progressive measurability in ω and t.** For all $t > 0$ and all $x \in \mathbb{R}^q$, the following maps are measurable with respect to $\mathscr{B}_{[0,t]} \otimes \mathscr{F}_t$:

$$(u,\omega) \in [0,t] \times \Omega \mapsto \sigma(u,\omega,x) \quad \text{and} \quad (u,\omega) \in [0,t] \times \Omega \mapsto b(u,\omega,x)$$

- **(Int) Square integrability in ω and in t (locally).** For all $t > 0$ and $x \in \mathbb{R}^q$,

$$\mathbb{E}\int_0^t |\sigma|^2(u,\cdot,x)\mathrm{d}u < \infty \quad \text{and} \quad \mathbb{E}\int_0^t |b|^2(u,\cdot,x)\mathrm{d}u < \infty.$$

Note that **(Lip)** implies that if **(Int)** holds for some $x \in \mathbb{R}^q$ then it holds for all $x \in \mathbb{R}^q$.
Note that **(Lip)** implies continuity in x and thus measurability in x.
Note that we do not assume continuity of $b(t,\omega,x)$ and $\sigma(t,\omega,x)$ with respect to t.

Lemma 8.1.1. Well posedness of the stochastic differential equation.

Let $s \geq 0$, and let $X = (X_t)_{t \geq s}$ be a q-dimensional continuous adapted process such that for all $t \geq s$,

$$\mathbb{E}\int_s^t |X_u|^2 \mathrm{d}u < \infty.$$

Then for all $t \geq 0$,

$$\int_s^t |b(u,X_u)|^2 \mathrm{d}u < \infty \quad \text{almost surely} \quad \text{and} \quad \sum_{i=1}^{q} \mathbb{E}\int_s^t |\sigma_{i,\bullet}(u,X_u)|^2 \mathrm{d}u < \infty.$$

In particular, the integrals in the right hand side of (SDE) make sense. Moreover, for all $1 \leq j \leq q$,

$$\Big(\int_s^t \sigma_{j,\bullet}(u,X_u)\mathrm{d}B_u\Big)_{t \geq s} = \Big(\sum_{k=1}^{d} \int_s^t \sigma_{j,k}(u,X_u)\mathrm{d}B_u^k\Big)_{t \geq s}$$

is a square integrable martingale, and the Itô isometry gives, for all $t \geq s$,

$$\mathbb{E}\Big(\Big|\int_s^t \sigma(u,X_u)\mathrm{d}B_u\Big|^2\Big) = \mathbb{E}\Big(\int_s^t |\sigma(u,X_u)|^2\mathrm{d}u\Big).$$

Furthermore if X solves (SDE) then for all j, X^j is a continuous semi-martingale, with local martingale part $\int_0^\bullet \sigma_{j,\bullet}(u,X_u)\mathrm{d}B_u$, which is a martingale, and finite variation part $\int_0^\bullet b_j(u,X_u)\mathrm{d}u$.

Proof. Thanks to **(Lip)**, for all $(u,\omega) \in \mathbb{R}_+ \times \Omega$, the maps $x \in \mathbb{R}^q \mapsto \sigma(u,\omega,x)$ and $x \in \mathbb{R}^q \mapsto b(u,\omega,x)$ are (uniformly) continuous. Since X is adapted and continuous, it is the pointwise limit of a sequence of adapted step processes taking a finite number of values (discretization of time and space). Thanks to the continuity of σ and b with respect to x, and to **(Mes)**, it follows that the processes $(\sigma(t,X_t))_{t\geq s}$ and $(b(t,X_t))_{t\geq 0}$ are the pointwise limit of progressively measurable processes, and are thus progressively measurable.

Now for the integral involving b, we write, using **(Lip)**,

$$|b(u,X_u)|^2 \leq 2(|b(u,0)|^2 + c^2|X_u|^2),$$

and by **(Int)** for b and the Cauchy–Schwarz inequality and the square integrability assumed for X, we get

$$\mathbb{E}\Big(\Big(\int_s^t |b(u,0)|\mathrm{d}u\Big)^2\Big) \leq (t-s)\mathbb{E}\Big(\int_s^t |b(u,0)|^2\mathrm{d}u\Big) < \infty$$

and thus almost surely

$$\int_s^t |b(u,X_u)|\mathrm{d}u < \infty.$$

Let us consider the integral involving σ. Similarly to what we did for b, by **(Lip)** for σ,

$$|\sigma(u,X_u)|^2 \leq |\sigma(u,X_u)|^2 \leq 2(|\sigma(u,0)|^2 + c^2|X_u|^2),$$

and thus, using **(Int)** for σ and the square integrability assumed for X, we get

$$\mathbb{E}\Big(\int_s^t |\sigma(u,X_u)|^2\mathrm{d}u\Big) \leq \mathbb{E}\Big(\int_s^t 2\Big(|\sigma(u,0)|^2 + c^2|X_u|^2\Big)\mathrm{d}u\Big) < \infty.$$

Hence for all $1 \leq i \leq q$, the stochastic integral $\int_s^t \sigma_{i,\bullet}(u,X_u)\mathrm{d}B_u$ is well defined and is a martingale. ■

Theorem 8.1.2. Solving stochastic differential equations and pathwise uniqueness.

For all $s \geq 0$ and all $\mathcal{F}_s$-measurable square integrable random vector η 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 $t \geq s$, $\mathbb{E}\int_s^t |X_u|^2\mathrm{d}u < \infty$
2. X solves (SDE) with initial condition η
3. such a solution is unique up to indistinguishability, hence pathwise uniqueness!

When σ and b do not depend on the space variable x, then (SDE) has an immediate explicit solution

$$X_t = X_s + \int_s^t \sigma(u)\mathrm{d}B_u + \int_s^t b(u)\mathrm{d}u$$

which is known as an Itô process. It is a martingale when $b = 0$, and a finite variation process when $\sigma = 0$.

The proof below is constructive in the sense that the solution is approximated by $S^n Y$ for n large enough and an arbitrary initial process Y. The solution does not come from the usage of the axiom of choice via a general theorem such as the Hahn–Banach theorem. However, a true algorithm on a computer would require to discretize time (Euler scheme for instance) and space and to control the quality of such an approximation. This is the subject of a whole theory that we can call stochastic numerical analysis, see [27, 26].

Proof. The idea is to try to use a fixed point method just like in the Picard theorem or the Cauchy–Lipschitz theorem for ordinary differential equations, adapted to our stochastic processes context.

Let $\mathscr{D}$ be the set of continuous adapted q-dimensional processes $(Y_t)_{t\geq s}$ with, for all $t\geq 0$,

$$\|Y\|_t^2 = \mathbb{E}\Big(\sup_{s\leq u\leq t}|Y_u|^2\Big)<\infty.$$

For all $Y\in\mathscr{D}$, thanks to Lemma 8.1.1, we can define, for all $t\geq s$,

$$SY(t)=\eta+\int_s^t\sigma(u,Y_u)\mathrm{d}B_u+\int_s^t b(u,Y_u)\mathrm{d}u.$$

It is unclear for now if SY belongs to $\mathscr{D}$ or not. For all Y^1 and Y^2 in $\mathscr{D}$ we have, for all $t\geq s$,

$$(SY^1)_t-(SY^2)_t=\int_s^t(\sigma(u,Y_u^1)-\sigma(u,Y_u^2))\mathrm{d}B_u+\int_s^t(b(u,Y_u^1)-b(u,Y_u^2))\mathrm{d}u,$$

and then, by using the Cauchy–Schwarz inequality twice,

$$|(SY^1)_t-(SY^2)_t|^2\leq 2\left|\int_s^t(\sigma(u,Y_u^1)-\sigma(u,Y_u^2))\mathrm{d}B_u\right|^2+2(t-s)\int_s^t|b(u,Y_u^1)-b(u,Y_u^2)|^2\mathrm{d}u.$$

Note that there is not hope to use Itô isometry because the norm in $\mathscr{D}$ is the expectation of a supremum not the converse, but this reminds Doob maximal inequality! Using **(Lip)** for b, we get, for all $t\geq s$,

$$\|SY^1-SY^2\|_t^2\leq 2\mathbb{E}\sup_{s\leq u\leq t}\left|\int_s^u(\sigma(v,Y_v^1)-\sigma(v,Y_v^2))\mathrm{d}B_v\right|^2+2c^2(t-s)\int_s^t\mathbb{E}(|Y_u^1-Y_u^2|^2)\mathrm{d}u.$$

Next, by Lemma 8.1.1, the Doob maximal inequality (Theorem 2.5.7), the Itô isometry, and **(Lip)** for σ,

$$\begin{aligned}\|SY^1-SY^2\|_t^2&\leq 8\mathbb{E}\int_s^t|\sigma(u,Y_u^1)-\sigma(u,Y_u^2)|^2\mathrm{d}u+2c^2(t-s)\int_s^t\mathbb{E}(|Y_u^1-Y_u^2|^2)\mathrm{d}u\\&\leq 2c^2(4+(t-s))\int_s^t\mathbb{E}(|Y_u^1-Y_u^2|^2)\mathrm{d}u\\&=C_t\int_s^t\mathbb{E}(|Y_u^1-Y_u^2|^2)\mathrm{d}u.\end{aligned}$$

Taking $Y^2\equiv 0$, this shows that $SY\in\mathscr{D}$ when $Y\in\mathscr{D}$ (beware that $S0\neq 0$). This gives also the inequality

$$\|SY^1-SY^2\|_t^2\leq C_t\|Y^1-Y^2\|_t^2.$$

So S is Lipschitz, but S is not necessarily a contraction because C_t can be arbitrarily large. To circumvent the problem, we bootstrap the estimate by plugin the same estimate into itself recursively. Namely, if we define

$$\varphi(u)=\mathbb{E}(|Y_u^1-Y_u^2|^2),$$

and if we denote, for all $n\geq 1$, by $S^n=S\circ\cdots\circ S$ the n-th iteration of S, we get

$$\begin{aligned}\|S^nY^1-S^nY^2\|_t^2&\leq (C_t)^2\int_s^t\mathrm{d}u\int_s^u\mathbb{E}(|(S^{n-2}Y^1)_v-(S^{n-2}Y^2)_v|^2)\mathrm{d}v \qquad (\star)\\&\ \ \vdots\\&\leq (C_t)^n\int\mathbf{1}_{t\geq u_1\geq\cdots\geq u_n\geq s}\varphi(u_n)\mathrm{d}u_1\ldots\mathrm{d}u_n\\&\leq (C_t)^n\|Y^1-Y^2\|_t^2\frac{(t-s)^n}{n!},\end{aligned}$$

where we used the basic estimate (also used in the study of order statistics and simple Poisson process)

$$1=\int_{[0,1]^n}\mathrm{d}u_1\cdots\mathrm{d}u_n=\sum_{\sigma\in\Sigma_n}\int_{[0,1]^n}\mathbf{1}_{u_{\sigma(1)}\geq\cdots\geq u_{\sigma(n)}}\mathrm{d}u_1\cdots\mathrm{d}u_n=n!\int_{[0,1]^n}\mathbf{1}_{u_1\geq\cdots\geq u_n}\mathrm{d}u_1\cdots\mathrm{d}u_n.$$

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

$$\mathbb{E}\Big(\sup_{s\leq u\leq t} |X^n_u - X^{n+1}_u|^2\Big) \leq \frac{(C_t(t-s))^n}{n!}\|Y - SY\|_t^2. \qquad (\star\star)$$

It follows that

$$\mathbb{E}\sum_{n\geq 0}\sup_{s\leq u\leq t} |X^n_u - X^{n+1}_u|^2 \leq \sum_{n\geq 0}\frac{(C_t(t-s))^n}{n!}\|Y - SY\|_t^2 \leq \|Y-SY\|_t^2 \mathrm{e}^{C_t(t-s)} < \infty.$$

Thus, for all $t > s$, almost surely

$$\sum_{n\geq 0}\sup_{s\leq u\leq t} |X^n_u - X^{n+1}_u|^2 < \infty.$$

By using Lemma 4.3.2 in the Banach space $\mathscr{C}([s,t],\mathbb{R}^q)$ for an arbitrary say integer $t \geq s$, it follows that almost surely, the sequence of continuous functions $(u \geq s \mapsto X^n_u)_{n\geq 0}$ converges uniformly on every compact subset of $[s,\infty)$ towards the trajectory of a continuous adapted process denoted $X = (X_u)_{u\geq 0}$ and from $(\star\star)$ we get

$$\Big(\mathbb{E}\Big(\sup_{s\leq u\leq t}|X^n_u - X_u|^2\Big)\Big)^{1/2} \leq \sum_{m\geq n}\|X^m - X^{m+1}\|_t \underset{n\to\infty}{\longrightarrow} 0.$$

It follows that $X \in \mathcal{D}$, that $X^n \to X$ in $\mathcal{D}$, and that (recall that $X^{n+1} = S^{n+1}X = SX^n$)

$$\|X - SX\|_t \leq \|X - X^{n+1}\|_t + \|SX^n - SX\|_t \underset{n\to\infty}{\longrightarrow} 0.$$

It follows that $X = SX$ in other words X is a fixed point of S. Finally, if X and $\widetilde{X}$ are two fixed points of S, then for all $n \geq 0$, $X - \widetilde{X} = S^n X - S^n\widetilde{X}$, and from $(\star)$, for all $t \geq 0$,

$$\|X - \widetilde{X}\|_t^2 \leq \frac{(C_t(t-s))^n}{n!}\|X - \widetilde{X}\|_t^2 \underset{n\to\infty}{\longrightarrow} 0$$

and therefore $X = \widetilde{X}$, hence the uniqueness up to indistinguishability. ∎

Theorem 8.1.3. Dependency over initial condition.

For all $s \geq 0$, for all $\mathscr{F}_s$-measurable square integrable random vectors η and $\widetilde{\eta}$ of $\mathbb{R}^q$, if X and $\widetilde{X}$ are solutions on the same space and for the same B, σ, b of

$$X_t = \eta + \int_s^t \sigma(u, X_u)\mathrm{d}B_u + \int_s^t b(u, X_u)\mathrm{d}u \quad \text{a.s.,} \quad t \geq s,$$

and

$$\widetilde{X}_t = \widetilde{\eta} + \int_s^t \sigma(u, \widetilde{X}_u)\mathrm{d}B_u + \int_s^t b(u, \widetilde{X}_u)\mathrm{d}u \quad \text{a.s.,} \quad t \geq s$$

respectively, then, for all $t \geq s$, there exists a constant $C_t > 0$ such that

$$\mathbb{E}\Big(\sup_{s\leq u\leq t}|X_u - \widetilde{X}_u|^2\Big) \leq C_t\mathbb{E}(|\eta - \widetilde{\eta}|^2).$$

Proof. This is a byproduct of the proof of Theorem 8.1.2. Let us give a direct proof via Lemma 8.1.4. We have

$$X_t - \widetilde{X}_t = \eta - \widetilde{\eta} + \int_s^t (\sigma(u, X_u) - \sigma(u, \widetilde{X}_u))\mathrm{d}B_u + \int_s^t (b(u, X_u) - b(u, \widetilde{X}_u))\mathrm{d}u.$$

For all $t \geq s$, setting

$$f(t) = \mathbb{E}\Big(\sup_{s\leq u\leq t}|X_u - \widetilde{X}_u|^2\Big),$$

we get, by Lemma 8.1.1, the Doob maximal inequality (Theorem 2.5.7), the Itô isometry, and **(Lip)**,

$$\begin{aligned} f(t) &\leq 3\mathbb{E}(|\eta-\widetilde{\eta}|^2) + 12\mathbb{E}\int_s^t |\sigma(u,X_u) - \sigma(u,\widetilde{X}_u)|^2\mathrm{d}u + 3(t-s)\mathbb{E}\int_s^t |b(u,X_u) - b(u,\widetilde{X}_u)|^2\mathrm{d}u \\ &\leq 3\mathbb{E}(|\eta-\widetilde{\eta}|^2) + c^2(12 + 3(t-s))\int_s^t f(u)\mathrm{d}u. \end{aligned}$$

It remains to use the Grönwall lemma (Lemma 8.1.4) with $a = 3\mathbb{E}(|\eta - \widetilde{\eta}|^2)$ and $b = c^2(12 + 3(t-s))$.

Lemma 8.1.4. Grönwall[a] lemma[b].

[a]Named after Thomas Hakon Grönwall (1877 – 1932), Swedish mathematician.
[b]Original version published in 1919 by Grönwall. There are plenty of versions, differential, integral, with variable coefficients, etc, including a non-linear version (Bihari – LaSalle inequality). Such lemmas are essential for ODEs and SDEs.

Let $s < u$ and $f : [s,u] \to \mathbb{R}$ bounded measurable. If for constants $a \in \mathbb{R}$ and $b \geq 0$, and all $t \in [s,u]$,

$$f(t) \leq a + b\int_s^t f(v)\mathrm{d}v, \quad \text{then for all } t \in [s,u], f(t) \leq a\mathrm{e}^{b(t-s)}.$$

Proof. By iterating the condition we obtain, by induction on n, for all $n \geq 0$ and all $t \in [s,u]$, with $t_0 = t$,

$$f(t) \leq a + a(b(t-s)) + \cdots + a\frac{(b(t-s))^n}{n!} + b^{n+1}\int_s^{t_0}\cdots\int_s^{t_n} f(t_{n+1})\mathbf{1}_{t_0\geq\cdots\geq t_{n+1}}\mathrm{d}t_1\cdots\mathrm{d}t_{n+1}.$$

Now the integral term is bounded above by $\|f\|_\infty \frac{(b(t-s))^{n+1}}{(n+1)!}$ which tends[2] to 0 as $n \to \infty$. ■

■

Theorem 8.1.5. Regular solution of the stochastic differential equation.

For all $s \geq 0$, there exists a family $(X_t^s(x) : x \in \mathbb{R}^q, t \geq s)$ of random variables such that:

1. for all $t \geq s$, the map $(x,\omega) \in \mathbb{R}^q \times \Omega \mapsto X_t^s(x,\omega) \in \mathbb{R}^q$ is measurable with respect to $\mathscr{B}_{\mathbb{R}^q} \otimes \mathscr{F}_t$
2. for all square integrable random vector η of $\mathbb{R}^q$ measurable with respect to $\mathscr{F}_s$, the stochastic process $(Y_t)_{t\geq 0}$ defined by $Y_t(\omega) = X_t^s(\eta(\omega),\omega)$ solves the stochastic differential equation

$$Y_t = \eta + \int_s^t \sigma(u, Y_u)\mathrm{d}B_u + \int_s^t b(u, Y_u)\mathrm{d}u \quad \text{a.s.,} \quad t \geq s. \qquad (\star)$$

Proof. In order to construct a solution measurable with respect to the initial condition, we discretize the space using an at most countable mesh and we rely on the regularity of the solution with respect to the initial condition. Namely, for all $n \geq 0$, let $(A_{n,k})_{k\geq 0}$ be an at most countable partition of $\mathbb{R}^q$ such that for all $k \geq 0$, $\mathrm{diam}(A_{n,k}) \leq 2^{-n}$. For each $k \geq 0$, we select $z_{n,k} \in A_{n,k}$, and we define, for all $x \in \mathbb{R}^q$,

$$g_n(x) = z_{n,k} \text{ where } k \text{ is such that } x \in A_{n,k}.$$

Let $z \in \mathbb{R}^q$. We consider the solution $\widetilde{X}_t(z,\omega)$ of

$$\widetilde{X}_t = z + \int_s^t \sigma(u, \widetilde{X}_u)\mathrm{d}B_s + \int_s^t b(u, \widetilde{X}_u)\mathrm{d}u,$$

for all $t \geq s$ and all $\omega \not\in N_z$ where N_z is a negligible set. Let us define

$$N_n = \bigcup_k N_{z_{n,k}} \quad \text{and} \quad X_t^n(x,\omega) = \widetilde{X}_t(g_n(x),\omega)\mathbf{1}_{\Omega\setminus N_n}(\omega).$$

The map $(x,\omega) \mapsto X_t^n(x,\omega)$ is measurable with respect to $\mathscr{B}_{\mathbb{R}^q} \otimes \mathscr{F}_t$, and by Theorem 8.1.3, for all $x \in \mathbb{R}^q$,

$$\mathbb{E}\Big(\sup_{s\leq u\leq t} |X_u^n(x) - \widetilde{X}_u(x)|^2\Big) \leq C_t|x - g_n(x)|^2 \leq C_t\Big(\frac{1}{2^n}\Big)^2.$$

Thus, for all $x \in \mathbb{R}^q$,

$$\mathbb{E}\sum_{n\geq 0}\sup_{s\leq u\leq t} |X_u^n(x) - \widetilde{X}_u(x)| < \infty$$

[2]The Stirling formula is $n! \sim_{n\to\infty} \sqrt{2\pi}n^{n+\frac{1}{2}}\mathrm{e}^{-n}$.

and therefore, for all $x \in \mathbb{R}^q$, for all $t \geq s$, almost surely,

$$\sup_{s \leq u \leq t} |X_u^n(x) - \widetilde{X}_u(x)| \underset{n \to \infty}{\longrightarrow} 0.$$

Now we define (the limit is taken component by component)

$$X_t^s(x, \omega) = \lim_{n \to \infty} X_t^n(x, \omega).$$

This limit is measurable as a pointwise limit of measurable functions. Now, let η be a square integrable $\mathscr{F}_s$-measurable random vector η of $\mathbb{R}^q$. One can check easily that $Y_t^n(\omega) = X_t^n(\eta(\omega), \omega)$ solves

$$Y_t^n = g_n(\eta) + \int_s^t \sigma(u, Y_u^n) \mathrm{d}B_u + \int_s^t b(u, Y_u^n) \mathrm{d}u, \quad t \geq s,$$

indeed, for all k, almost surely,

$$\mathbf{1}_{\eta \in A_k} \int_s^t \sigma(u, X_u^n(z_k)) \mathrm{d}B_u = \int_s^t \mathbf{1}_{\eta \in A_k} \sigma(u, X_u^n(z_k)) \mathrm{d}B_u = \int_s^t \mathbf{1}_{\eta \in A_k} \sigma(u, Y_u^n) \mathrm{d}B_u.$$

Finally, one can check easily using Theorem 8.1.3 and Lemma 4.3.2 that for all $t \geq s$, almost surely,

$$\sup_{s \leq u \leq t} |Y_u^n - Y_u| \underset{n \to \infty}{\longrightarrow} 0,$$

where $Y = (Y_t)_{t \geq 0}$ is the solution of $(\star)$. It follows that for all $t \geq s$, almost surely,

$$X_t^s(\eta(\omega), \omega) = Y_t(\omega).$$

■

Corollary 8.1.6. Composition.

For all $0 \leq s \leq t \leq u$, $x \in \mathbb{R}^q$, with the notions of Theorem 8.1.5, a.s.

$$X_u^s(x, \omega) = X_u^t(X_t^s(x, \omega), \omega)).$$

Proof. We have

$$\begin{aligned} X_u^s(x) &= x + \int_s^u \sigma(v, X_v^s) \mathrm{d}B_v + \int_s^u b(v, X_v^s) \mathrm{d}v \\ &= x + \int_s^t \sigma(v, X_v^s) \mathrm{d}B_v + \int_s^t b(v, X_v^s) \mathrm{d}v + \int_t^u \sigma(v, X_v^s) \mathrm{d}B_v + \int_t^u b(v, X_v^s) \mathrm{d}v \\ &= X_t^s(x) + \int_t^u \sigma(v, X_v^s) \mathrm{d}B_v + \int_t^u b(v, X_v^s) \mathrm{d}v \end{aligned}$$

where the last equality holds almost surely. Therefore $(X_u^s)_{u \geq t}$ solves the SDE started from X_t^s at time t. Now by the representation of Theorem 8.1.5 and the pathwise uniqueness of Theorem 8.1.2, we get, a.s.

$$X_u^s(x, \omega) = X_u^t(X_t^s(x, \omega), \omega).$$

■

8.2 Ornstein – Uhlenbeck, Bessel, and Langevin processes

Example 8.2.1. Shifted Brownian motion.

If $q = d$, $\eta = 0$, $\sigma(t, \omega, x) = I_d$ (constant), and $b(t, \omega, x) = b(t, \omega)$ (possibly random and time varying but constant in x), then (SDE) gives, for all $t \geq s$,

$$X_t = B_t + \int_s^t b(u)\mathrm{d}u.$$

If $b = 0$ then $X = B$. If b is deterministic, then Theorem 3.8.2 (Cameron – Martin) gives the density of the law of X with respect to Wiener measure. See also Theorem 7.5.1 and Theorem 8.4.11.

Example 8.2.2. Ornstein[a] – Uhlenbeck[b] process.

[a]Named after Leonard Ornstein (1880 – 1941), Deutch physicist.
[b]Named after George Eugene Uhlenbeck (1900 – 1988), Dutch-American theoretical physicist.

For simplicity, let $X_0 \in \mathrm{L}^2_{\mathbb{R}^d}$ independent of $(B_t)_{t\geq 0}$. The Ornstein – Uhlenbeck process starting from $X_0 \in \mathbb{R}^d$ solves the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \sigma \mathrm{d}B_t - \mu X_t \mathrm{d}t, \ t \geq 0,$$

where $\sigma \geq 0$ and $\mu \in \mathbb{R}$ are constants (the standard O.-U. is with $\sigma = \sqrt{2}$ and $\mu = 1$), in other words

$$X_t = X_0 + \sigma \int_0^t B_s \mathrm{d}s + \mu \int_0^t X_s \mathrm{d}s = X_0 + \sigma B_t - \mu \int_0^t X_s \mathrm{d}s.$$

This corresponds to (SDE) with $q = d$, $\sigma(u,x) = \sigma I_d$ (constant), and $b(u,x) = -\mu x$. Let us identify the solution of this SDE. We use an apriori estimate. More precisely, if X exists and is a semi-martingale, then, by the Itô formula with $f(x,y) = xy$ and the process $(\mathrm{e}^{\mu t}, X_t)$, and by using the SDE, we get

$$\mathrm{d}(\mathrm{e}^{\mu t} X_t) = \mathrm{e}^{\mu t}\mathrm{d}X_t + \mathrm{e}^{\mu t}\mu X_t \mathrm{d}t = \mathrm{e}^{\mu t}\sigma \mathrm{d}B_t \quad \text{and thus} \quad \mathrm{e}^{\mu t}X_t - \mathrm{e}^0 X_0 = \sigma \int_0^t \mathrm{e}^{\mu s}\mathrm{d}B_s,$$

which gives[a]

$$X_t = \mathrm{e}^{-\mu t}X_0 + \sigma \int_0^t \mathrm{e}^{\mu(s-t)}\mathrm{d}B_s.$$

This gives uniqueness, and we check from this formula that this process solves the SDE. The integral in the right hand side is a Wiener integral. For all $t \geq 0$, since $\int_0^t (\sigma \mathrm{e}^{\mu(s-t)})^2 \mathrm{d}s = \frac{\sigma^2}{2}\frac{1-\mathrm{e}^{-2\mu t}}{\mu}$, we get

$$\mathrm{Law}(X_t \mid X_0 = x) = \mathcal{N}\Big(x\mathrm{e}^{-\mu t}, \frac{\sigma^2}{2}\frac{1-\mathrm{e}^{-2\mu t}}{\mu} I_d\Big) \quad \text{with convention } \frac{1-\mathrm{e}^{-2\mu t}}{\mu} = 2t \text{ if } \mu = 0.$$

If $X_0 = x = (x_1, \ldots, x_d)$, then the d coordinates of X are independent one-dimensional O.-U. processes started from $x_1, \ldots, x_d$. By the isometry property for Wiener – Itô integrals, for all $s, t \geq 0$, $1 \leq i, j \leq d$,

$$\begin{aligned}
\mathrm{Cov}(X_s^i, X_t^j \mid X_0 = x) &= \sigma^2 \mathbb{E}\Big(\int_0^s \mathrm{e}^{\mu(u-s)}\mathrm{d}B_u^i, \int_0^t \mathrm{e}^{\mu(u-t)}\mathrm{d}B_u^j\Big)\\
&= \sigma^2 \mathbf{1}_{i=j}\mathrm{e}^{-\mu(t+s)}\int_0^{s\wedge t}\mathrm{e}^{2\mu u}\mathrm{d}u\\
&= \sigma^2 \mathbf{1}_{i=j}\Big(\frac{\mathrm{e}^{-\mu|t-s|} - \mathrm{e}^{-\mu(s+t)}}{2\mu}\mathbf{1}_{\mu\neq 0} + (s\wedge t)\mathbf{1}_{\mu=0}\Big).
\end{aligned}$$

When $\mu > 0$ then this quantity is small when both $s + t$, and $|t - s|$ are large. For all $s, t \geq 0$,

$$X_{t+s} = \mathrm{e}^{-\mu s}X_t + \mathrm{e}^{-\mu s}\sigma \int_t^{t+s} \mathrm{e}^{u-t}\mathrm{d}B_u = \mathrm{e}^{-\mu s}X_t + \mathrm{e}^{-\mu s}\sigma \int_0^s \mathrm{e}^{-\mu u}\mathrm{d}B_{t+u} = F_s(X_t, (B_u)_{u\in(s,t]}).$$

The process X is a continuous auto-regressive Gaussian process, constructed by incorporating along the time a new independent input, nevertheless it does not have independent increments. The SDE allows simulation as well as a dynamical interpretation of the trajectories, see Figure 8.1.

[a]This is not the canonical decomposition of the semi-martingale X since $\int_0^t \mathrm{e}^{\mu(s-t)}\mathrm{d}B_s$ is not a local martingale even if $\int_0^\bullet \mathrm{e}^{\mu s}\mathrm{d}B_s$ is a martingale. The canonical decomposition is given by the SDE. See also Exercise 1 of the 2020-2021 exam.

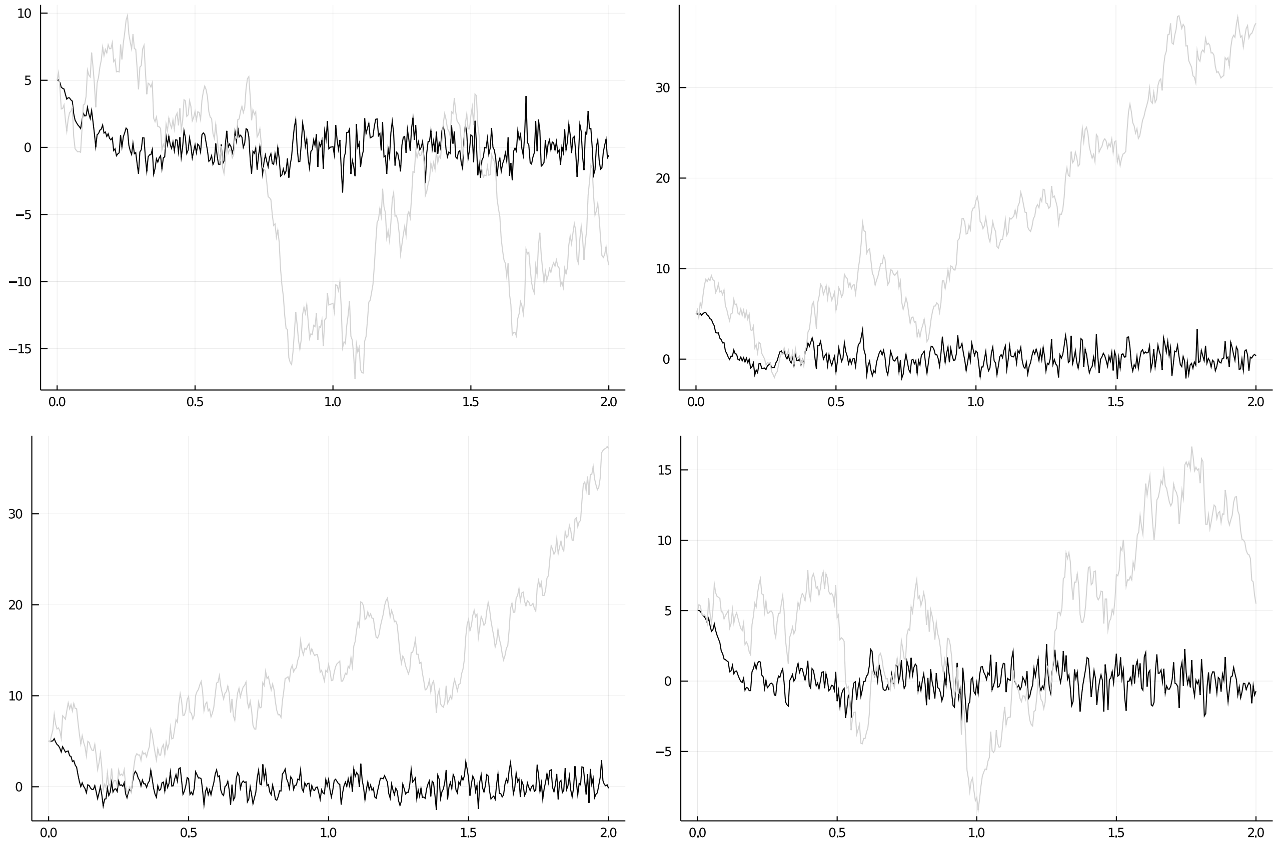


Figure 8.1: In black, four trajectories of the Ornstein – Uhlenbeck process with $X_0 = 5$, $\sigma = 1$, $\mu = 1$. The mean converges to 0 while the variance converges to $\frac{\sigma^2}{2\mu} = \frac{1}{2}$. At the beginning, the drift term in the SDE is stronger than the diffusion term and this has the effect to drive exponentially fast the process to a neighborhood of the origin. The process fluctuates then in this neighborhood forever, the drift term and the diffusion term remaining at the same order. The drift term $-\mu X_t \mathrm{d}t$ has the effect of a spring force (*force de rappel* in French) since $\mu > 0$. In light gray, the trajectories of the underlying driving Brownian motion of the SDE, the mean remains at 5 for all times while the variance grows linearly in time (they also corresponds to $\mu = 0$).

Remark 8.2.3. Quantitative exponential long time behavior of O. – U. via coupling.

Let $X = (X_t)_{t\geq 0}$ be an Ornstein – Uhlenbeck process solving the SDE

$$\mathrm{d}X_t = \sigma \mathrm{d}B_t - \mu X_t \mathrm{d}t,$$

with $\mu > 0$. Let $X' = (X'_t)_{t\geq 0}$ be another Ornstein – Uhlenbeck process in $\mathbb{R}^d$ solving the same SDE (same Brownian motion) but with an initial condition X'_0 possibly distinct from X_0. This is a way to construct the couple (X, X'). The law of (X, X') is a coupling of the laws of X and X'. Now

$$\mathrm{d}(X_t - X'_t) = -\mu(X_t - X'_t)\mathrm{d}t \quad \text{and thus} \quad X_t - X'_t = (X_0 - X'_0)\mathrm{e}^{-\mu t}.$$

Let $\mathscr{P}_2$ be the set of probability measures on $\mathbb{R}^d$ integrating $|\cdot|^2$. The Wasserstein[a] – Kantorovich[b] – [c]Fréchet – Monge[d] coupling distance W_2 on $\mathscr{P}_2$ is defined for all $\mu, \nu \in \mathscr{P}_2$ by

$$\mathrm{W}_2(\mu,\nu)^2 = \inf_{\pi} \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \pi(\mathrm{d}x,\mathrm{d}y) = \inf_{(U,V)} \mathbb{E}(|U-V|^2)$$

where the first infimum runs over all the probability measures π on the product space with marginal distribution μ and ν, and the second infimum over all couple of random variables with marginal laws

μ and ν. From now on, let us assume that the laws of X_0 and X'_0 are in $\mathscr{P}_2$. It can be shown then that the laws of X_t and X'_t are also in $\mathscr{P}_2$ for all $t \geq 0$. We have, from our previous estimates,

$$\mathrm{W}_2(\mathrm{Law}(X_t), \mathrm{Law}(X'_t)) \leq |X_0 - X'_0| \mathrm{e}^{-\mu t}.$$

By taking in turn the infimum over all couplings of X_0 and X'_0 we get

$$\mathrm{W}_2(\mathrm{Law}(X_t), \mathrm{Law}(X'_t)) \leq \mathrm{W}_2(\mathrm{Law}(X_0), \mathrm{Law}(X'_0)) \mathrm{e}^{-\mu t}.$$

Finally, recall that we already know that $\gamma = \mathcal{N}(0, \frac{\sigma^2}{2\mu} I_d)$ is an invariant law of our Ornstein–Uhlenbeck process, in the sense that $X'_0 \sim \gamma$ implies $X'_t \sim \gamma$ for all $t \geq 0$. It follows that

$$\mathrm{W}_2(\mathrm{Law}(X_t), \gamma) \leq \mathrm{W}_2(\mathrm{Law}(X_0), \gamma) \mathrm{e}^{-\mu t}.$$

This is a quantitative version of the long time exponential behavior of the process.

[a] Named after Leonid Vaseršteĭn, Russian-American mathematician.
[b] Named after Leonid Vitaliyevich Kantorovich (1912 – 1986), Soviet mathematician and economist.
[c] Maurice René Fréchet (1878 – 1973), French mathematician.
[d] Gaspard Monge, Comte de Péluse (1746 – 1818), French mathematician.

Coding in action 8.2.4. Simulation.

Write a code to simulate approximate trajectories of the Ornstein – Uhlenbeck process in dimension $d = 1$ and plot them on the same graphics. Hint: use the structure of the increments. What is the effect of changing μ and σ on the trajectories? Consider in particular the case $\sigma = 0$ versus $\sigma > 0$, and the cases $\mu < 0$, $\mu = 0$, and $\mu > 0$. Could you check numerically the exponential convergence in law to the standard Gaussian as time tends to infinity when $\sigma, \mu > 0$? See Figure 8.1 for an example of plots.

Example 8.2.5. Bessel[a] processes.

[a] Named after Friedrich Bessel (1784 - 1846), German astronomer, mathematician, physicist and geodesist.

For all $x \in \mathbb{R}^d$, we define the process $X = (X_t)_{t\geq 0}$ by

$$X_t = |x + B_t|^2.$$

Let $r = |x|^2 = X_0$. We say that X is a squared Bessel process issued from r. The Itô formula (Theorem 7.1.1) with $f(x) = |x|^2$ which is $\mathscr{C}^2(\mathbb{R}^d, \mathbb{R})$ gives, via $\nabla f(x) = 2x$ and $\Delta f(x) = 2d$,

$$|x + B_t|^2 = |x|^2 + 2\int_0^t (x + B_s) \mathrm{d}B_s + td, \quad t \geq 0.$$

Thus on the canonical space for B, the process X solves the stochastic differential equation

$$X_t = r + \int_0^t \sigma(s, X_s) \mathrm{d}B_s + \int_0^t b(s, X_s) \mathrm{d}s$$

with a random $\sigma(s, x, \omega) = 2(x + \omega)$ and $b(s, x, \omega) = d$. They are constant in s and Lipschitz in x and we can use Theorem 8.1.2 to get the existence and pathwise uniqueness of the solution. Alternatively, the Lévy characterization of Brownian motion (Theorem 7.2.1) shows that the continuous martingale

$$W = \Big(\int_0^t \frac{x + B_s}{|x + B_s|} \mathrm{d}B_s\Big)_{t\geq 0} = \Big(\int_0^t \frac{x + B_s}{\sqrt{X_s}} \mathrm{d}B_s\Big)_{t\geq 0}$$

(with convention $0/|0| = 1$) issued from the origin is a Brownian motion on $\mathbb{R}$ since for all $t \geq 0$,

$$\langle W \rangle_t = \Big\langle \int_0^\bullet \frac{x + B_s}{|x + B_s|} \mathrm{d}B_s \Big\rangle_t = \int_0^t \frac{(x + B_s)\cdot(x + B_s)}{|x + B_s|^2} \mathrm{d}\langle B \rangle_s = \int_0^t \mathrm{d}s = t.$$

Now by writing $x+B=\sqrt{X}\frac{x+B}{\sqrt{X}}$ we see that X solves the stochastic differential equation

$$X_t = r + 2\int_0^t \sqrt{X_s}\mathrm{d}W_s + td.$$

Note that $\sigma(x)=2\sqrt{x}$ is not Lipschitz at $x=0$ and Theorem 8.1.2 does not apply. However a theorem due to Yamada–Watanabe states existence and pathwise uniqueness for the stochastic differential equation $\mathrm{d}X_t=\sigma(X_t)\mathrm{d}B_t+b(X_t)\mathrm{d}t$ as soon as $\sigma:\mathbb{R}\to\mathbb{R}$ and $b:\mathbb{R}\to\mathbb{R}$ satisfy

$$|\sigma(x)-\sigma(y)|\leq C\sqrt{|x-y|} \quad \text{and} \quad |b(x)-b(y)|\leq C|x-y|$$

for some $C>0$ and all $x,y\in\mathbb{R}$. See [31, Exercise 8.14 pages 231–232].
The local martingale $\int_0^\bullet \sqrt{X_s}\mathrm{d}W_s$ is a martingale since

$$\mathbb{E}\Big(\int_0^t (\sqrt{X_s})^2\mathrm{d}s\Big) = \int_0^t \mathbb{E}(|x+B_s|^2)\mathrm{d}s<\infty.$$

Therefore $(X_t-td)_{t\geq 0}$ is a martingale. This is not a surprise since it is the sum of the martingales $((B_t^i)^2-t)_{t\geq 0}$. More generally, a squared Bessel process of dimension $\alpha>0$ solves

$$X_t = r + 2\int_0^t \sqrt{X_s}\mathrm{d}W_s + t\alpha.$$

It is not obvious that such a process stays non-negative. The process $Y=\sqrt{X}$ in known as a Bessel process. It can be shown that when $r>0, \alpha>1$, Y solves the following SDE with singular drift

$$\mathrm{d}Y_t = \mathrm{d}W_t + \frac{\alpha-1}{2}\frac{\mathrm{d}t}{Y_t},$$

see [31, Ex. 8.13]. See also [31, Ex. 5.31 & Sec. 8.4.3] and [43, Ch. XI] for more on Bessel processes.

Example 8.2.6. Time change.

Let $X=(X_t(x))_{t\geq 0}$ be the solution of the stochastic differential equation

$$X_t = x + \int_0^t \sigma(u,X_u)\mathrm{d}B_u + \int_0^t b(u,X_u)\mathrm{d}u.$$

Let $\alpha>0$. Then the time changed process $Y=(X_{\alpha t}(x))_{t\geq 0}$ solves the stochastic differential equation

$$Y_t = x + \int_0^{\alpha t} \sigma(u,X_u)\mathrm{d}B_u + \int_0^{\alpha t} b(u,X_u)\mathrm{d}u.$$

Now, denoting $\widetilde{B}=(\frac{1}{\sqrt{\alpha}}B_{\alpha t})_{t\geq 0}$, we get, with the substitution $u=\alpha v$ that Y is a solution of

$$Y_t = x + \int_0^t \sqrt{\alpha}\sigma(\alpha v,Y_v)\mathrm{d}\widetilde{B}_v + \int_0^t \alpha b(\alpha v,Y_v)\mathrm{d}v.$$

Since $\widetilde{B}$ and B have the same law, it follows that Y is a weak solution of

$$Y_t = x + \int_0^t \sqrt{\alpha}\sigma(\alpha u,Y_u)\mathrm{d}B_u + \int_0^t \alpha b(\alpha u,Y_u)\mathrm{d}u.$$

For example, if X is an Ornstein–Uhlenbeck process solution of $\mathrm{d}X_t=\sigma\mathrm{d}B_t-\mu X_t\mathrm{d}t$, with $\sigma\geq 0$ and $\mu\in\mathbb{R}^d$, then for all $\alpha>0$ the process $Y=(X_{\alpha t})_{t\geq 0}$ is a weak solution of the SDE $\mathrm{d}Y_t=\sqrt{\alpha}\sigma\mathrm{d}B_t-\alpha\mu Y_t\mathrm{d}t$. We speed up (respectively slow down) the process when $\alpha>1$ (respectively $\alpha<1$).

Example 8.2.7. Langevin[a] process.

[a]Named after Paul Langevin (1872 – 1946), French physicist.

A unit mass particle in $\mathbb{R}^d$, with position Y and velocity Z, feels a force that depends on its position via a potential $U : \mathbb{R}^d \mapsto \mathbb{R}$, a friction force that depends on its velocity via a potential $V : \mathbb{R}^d \mapsto \mathbb{R}$, and a random Brownian force of variance $\sigma^2 > 0$ due to the medium. The Langevin stochastic differential equation follows from Newton fundamental relation of dynamics[a]:

$$\begin{cases} \mathrm{d}Y_t &= Z_t \mathrm{d}t \\ \mathrm{d}Z_t &= \sqrt{\gamma}\sigma \mathrm{d}B_t - \gamma \nabla V(Z_t)\mathrm{d}t - \nabla U(Y_t)\mathrm{d}t. \end{cases}$$

Here $\gamma \geq 0$ is the "friction" parameter. This is our (SDE) with $q = 2d$, $X = (Y, Z)$, and coefficients

$$\sigma_{i,j} = \begin{cases} \sqrt{\gamma}\sigma & \text{if } 1 \leq j \leq d \text{ and } i = d + j \\ 0 & \text{otherwise} \end{cases}, \quad \text{and} \quad b_i(y,z) = \begin{cases} z & \text{if } 1 \leq i \leq d \\ -\gamma \nabla V(z) - \nabla U(y) & \text{if } d+1 \leq i \leq 2d. \end{cases}$$

They are deterministic, constant in time, and Lipschitz iff ∇U and ∇V are Lipschitz. When $U(y) = \frac{1}{2}|y|^2$ and $V(z) = \frac{1}{2}|z|^2$, the Langevin process $X = (Y, Z)$ is known as a kinetic Ornstein – Uhlenbeck process. When $\gamma = 0$ then there is no randomness and we speak about a Hamiltonian equation.
The position Y_t and the velocity Z_t are coupled in the second equation above via the drift term $-\nabla U(Y_t)\mathrm{d}t$. It turns out that they decouple in the limit of a time – friction scaling. Namely, we can dilate time with a factor $\alpha > 0$, giving the equation (we keep same notations for processes)

$$\begin{cases} \mathrm{d}Y_t &= \alpha Z_t \mathrm{d}t \\ \mathrm{d}Z_t &= \sqrt{\alpha\gamma}\sigma \mathrm{d}B_t - \alpha\gamma \nabla V(Z_t)\mathrm{d}t - \alpha \nabla U(Y_t)\mathrm{d}t. \end{cases}$$

If $\alpha \to 0$ (slow down the process) and $\gamma \to \infty$ (high friction) while keeping $\alpha\gamma = 1$, we get $\mathrm{d}Y_t = 0$ and

$$\mathrm{d}Z_t = \sigma \mathrm{d}B_t - \nabla V(Z_t)\mathrm{d}t.$$

This is known as an overdamped Langevin equation, as a generalized Ornstein – Uhlenbeck equation, and also as a Kolmogorov equation in [46]. We recover Ornstein – Uhlenbeck when $V(z) = |z|^2$. The initial Langevin equation is called sometimes the underdamped or kinetic Langevin equation. We refer to [12] for a presentation of the physical aspects of Brownian motion and the Langevin equation, from the historical roots to nowadays physics, see also [17]. Beyond its physical signifiance, the underdamped Langevin process is a key ingredient in the Hamiltonian or Hybrid Monte Carlo (HMC) computational algorithms for the simulation of probability measures, see for instance [32].

[a]The first equation expresses the fact that the derivative of position with respect to time is the velocity, while the second the fact that mass × acceleration = $\frac{\mathrm{d}Z_t}{\mathrm{d}t}$ = sum of forces = $\sqrt{\gamma}\sigma\frac{\mathrm{d}B_t}{\mathrm{d}t} - \gamma\nabla V(Z_t) - \nabla U(Y_t)$. The term $\frac{\mathrm{d}B_t}{\mathrm{d}t}$ is a white noise.

Example 8.2.8. Geometric Brownian Motion and Black[a] – Scholes[b] process.

[a]Fisher Black, American economist (1938 – 1995).
[b]Myron Scholes, Canadian-American financial economist (1941 –).

The Black – Scholes process solves the SDE

$$\mathrm{d}S_t = S_t(\sigma_t \mathrm{d}B_t + \mu_t \mathrm{d}t), \quad S_0 > 0.$$

By using the Itô formula for $\log(S)$, we obtain

$$S_t = S_0 \exp\left(\int_0^t \sigma_s \mathrm{d}B_s - \frac{1}{2}\int_0^t \sigma_s^2 \mathrm{d}s + \int_0^t \mu_s \mathrm{d}s\right).$$

In the original model, σ and μ are constants and we find the geometric BM

$$S_t = S_0 e^{\sigma B_t - \frac{\sigma^2}{2}t + \mu t}.$$

This was used historically in the study of European options pricing, see [30, 25]. The usage of stochastic calculus in mathematical finance is widely developed in specialized courses of the Master MASEF.

Coding in action 8.2.9. Simulation.

Write a code to simulate approximate trajectories of the Bessel processes with various parameters, and for the Black–Scholes process. How to deal with the singularity at the origin of the Bessel SDE for non integer parameter? Do the same for the overdamped Langevin process with potential $V = |\cdot|^4$ by using an Euler scheme for the SDE.

8.3 Markov property, Markov semi-group, weak uniqueness

In this section, we assume that σ and b are deterministic (do not depend on ω) and that:

- there exists a constant $C > 0$ such that for all $x, y \in \mathbb{R}^q$ and all $u \in \mathbb{R}_+$,

$$|\sigma(u,x) - \sigma(u,y)| + |b(u,x) - b(u,y)| \leq C|x-y|$$

- σ and b are measurable maps from $\mathbb{R}_+ \times \mathbb{R}^q$ to $\mathscr{M}_{q,d}(\mathbb{R})$ and $\mathbb{R}^q$ respectively
- for all $t > 0$ and all $x \in \mathbb{R}^q$,

$$\int_0^t (|\sigma(u,x)|^2 + |b(u,x)|^2)\mathrm{d}u < \infty.$$

With these simplified assumptions, the initial assumptions **(Lip)**, **(Mes)**, and **(Int)** are satisfied. For all $s \geq 0$, we denote by $(X_t^s(x,\omega))_{x\in\mathbb{R}^q, t\geq s}$ the regular solution of (SDE) provided by Theorem 8.1.5.

$$X_s^s(x) = x, \quad \mathrm{d}X_t^s(x) = \sigma(t, X_t^s(x))\mathrm{d}B_s + b(t, X_t^s(x))\mathrm{d}t, t \geq s.$$

Theorem 8.3.1. Weak Markov property.

For all $s \geq 0$ and $x \in \mathbb{R}^q$, let $(X_t^s(x))_{t\geq s}$ be the solution of (SDE) with $\eta = x$. Then for all bounded and measurable $f : \mathbb{R}^q \to \mathbb{R}$, and for all $u \geq t \geq s$, almost surely

$$\mathbb{E}(f(X_u^s(x)) \mid \mathscr{F}_t) = \mathbb{E}(f(X_u^s(x)) \mid X_t^s(x)) = \Pi_{t,u}(f)(X_t^s(x)),$$

where for all $z \in \mathbb{R}^q$,

$$\Pi_{t,u}(f)(z) = \mathbb{E}(f(X_u^t(z))).$$

A Markov process has no memory, it the sense that for any time t interpreted as the present, the conditioning of its future with respect to its present and past is equal to the conditioning with respect to the present. This is equivalent to conditional independence of future and past given the present. On the other hand, a process with long memory can always be seen as a Markov process by seeing the whole trajectory as a state, simple examples are provided by ARMA time series and high order Markov chains for instance.

Proof. A key observation is that the stochastic integral $\int_t^u \sigma(v, X_v^t)\mathrm{d}B_v$ involves the increments of B after time t and is thus independent of the portion of B before time t. For all $z \in \mathbb{R}^q$, almost surely,

$$\begin{aligned} X_u^t(z) &= z + \int_t^u \sigma(v, X_v^t(z))\mathrm{d}B_v + \int_t^u b(v, X_v^t(z))\mathrm{d}v \\ &= z + \int_0^{u-t} \sigma(t+v, X_{t+v}^t(z))\mathrm{d}B_v^t + \int_0^{u-t} b(t+v), X_{t+v}^t(z))\mathrm{d}v, \end{aligned}$$

where $B^t = (B^t_v)_{v\geq 0} = (B_{t+v} - B_t)_{v\geq 0}$ is a translated Brownian motion, independent of $\mathscr{F}_t$. Corollary 8.1.6 gives $X^s_u = X^t_u(X^s_t(x)) = F_{t,u}(X^s_t(x), B^t)$ where $F_{t,u}$ is measurable. Since $X^s_t(x)$ depends only on $(B_v)_{s\leq v\leq t}$, it is $\mathscr{F}_t$-measurable and independent of B^t. Hence, for all bounded measurable $f:\mathbb{R}^q\to\mathbb{R}$, by Remark 1.5.2,

$$\mathbb{E}(f(X^s_u)\mid\mathscr{F}_t) = \mathbb{E}(f(F_{t,u}(X^s_t,B^t))\mid\mathscr{F}_t) = \mathbb{E}(f(F_{t,u}(X^s_t,B^t))\mid X^s_t) = \Pi_{t,u}(f)(X^s_t)$$

where for all $z\in\mathbb{R}^q$,

$$\Pi_{t,u}(f)(z) = \mathbb{E}(f(F_{t,u}(z,B^t))) = \mathbb{E}(f(X^t_u(z))).$$

An explicit construction of $F_{t,u}$ can be done on the Wiener space. ■

Remark 8.3.2. Markov transition kernel and Markov semi-group.

For all $0\leq s\leq t$, let $\Pi_{s,t}(x,\mathrm{d}y)$ be the Markov transition kernel on $\mathbb{R}^q$ given for $x\in\mathbb{R}^q$ and $A\in\mathscr{B}_{\mathbb{R}^q}$ by

$$\Pi_{s,t}(x,A) = \mathbb{P}(X^s_t(x)\in A).$$

It acts on bounded measurable $f:\mathbb{R}^q\to\mathbb{R}$ as

$$\Pi_{s,t}(f)(x) = \int_{\mathbb{R}^q} f(y)\Pi_{s,t}(x,\mathrm{d}y),\quad x\in\mathbb{R}^q.$$

Theorem 8.3.1 gives, for $u\geq t$, $\Pi_{s,u} = \Pi_{s,t}\circ\Pi_{t,u}$ in the sense that for all f and x we have

$$\Pi_{s,u}(f)(x) = \mathbb{E}(f(X^s_u(x))) = \mathbb{E}(\mathbb{E}(f(X^s_u(x))\mid\mathscr{F}_t)) = \mathbb{E}(\Pi_{t,u}(f)(X^s_t(x))) = \Pi_{s,t}(\Pi_{t,u}(f))(x),$$

and Theorem 8.3.1 with f replaced by $\Pi_{t,u}(f)$ gives that the process

$$(\Pi_{t,u}(f)(X^s_t(x)))_{t\in[s,u]}$$

is an $(\mathscr{F}_t)_{t\in[s,u]}$ martingale. This gives the (non-homogeneous) Markov semi-group property:

$$\Pi_{s,u}(x,\mathrm{d}y) = \int_{\mathbb{R}^q}\Pi_{s,t}(x,\mathrm{d}z)\Pi_{t,u}(z,\mathrm{d}y),\quad u\geq t\geq s\geq 0,\quad x\in\mathbb{R}^q.$$

Conversely, the Markov semi-group $(\Pi_{s,t}(x,\mathrm{d}y))_{0\leq s\leq t}$ fully determines the law of $(X^s_t(x))_{t\geq s}$. Indeed, for all $n\geq 1$, $0\leq s\leq t_1\leq t_2\leq\cdots\leq t_n$, and bounded and measurable $f_1,\ldots,f_n$ from $\mathbb{R}^q$ to $\mathbb{R}$, we have

$$\begin{aligned}\mathbb{E}(f_1(X^s_{t_1}(x))\cdots f_n(X^s_{t_n}(x))) &= \mathbb{E}(f_1(X^s_{t_1}(x))\cdots f_{n-1}(X^s_{t_{n-1}})\Pi_{t_{n-1},t_n}(f_n)(X^s_{t_{n-1}}(x)))\\ &= \int_{\mathbb{R}^q}\Pi_{s,t_1}(x,\mathrm{d}y_1)\Pi_{t_1,t_2}(y_1,\mathrm{d}y_2)\cdots\Pi_{t_{n-1},t_n}(y_{n-1},\mathrm{d}y_n)f_1(y_1)\cdots f_n(y_n).\end{aligned}$$

Theorem 8.3.3. Uniqueness in law or weak uniqueness.

Let $(\widetilde{\Omega},\widetilde{\mathscr{F}},(\widetilde{\mathscr{F}}_t)_{t\geq 0},\widetilde{\mathbb{P}})$ be another filtered probability space on which is defined a d-dimensional $(\widetilde{\mathscr{F}}_t)_{t\geq 0}$ Brownian motion $\widetilde{B} = (\widetilde{B}_t)_{t\geq 0}$ issued from the origin. Let $x\in\mathbb{R}^q$, and let $X = (X_t(x,\omega))_{t\geq 0}$ and $\widetilde{X} = (\widetilde{X}_t(x,\omega)_{t\geq 0}$ be the solutions of the respective stochastic differential equations:

$$X_t(x) = x + \int_0^t\sigma(u,X_u(x))\mathrm{d}B_u + \int_0^t b(u,X_u(x))\mathrm{d}u\quad a.s.\quad t\geq 0,$$

and

$$\widetilde{X}_t(x) = x + \int_0^t\sigma(u,\widetilde{X}_u(x))\mathrm{d}\widetilde{B}_u + \int_0^t b(u,\widetilde{X}_u(x))\mathrm{d}u\quad a.s.\quad t\geq 0.$$

Then these processes X and $\widetilde{X}$ have the same law on $(\mathscr{C}(\mathbb{R}_+,\mathbb{R}^q),\mathscr{B}_{\mathscr{C}(\mathbb{R}_+,\mathbb{R}^q)})$.

Proof. Since σ and b do not depend on the randomness, regarding weak solutions, we can play with the probability space. We consider the canonical Brownian motion $\pi = (\pi_t(\omega))_{t\geq 0}$ defined on the Wiener space

$$(W = \mathscr{C}(\mathbb{R}_+,\mathbb{R}^d),\mathscr{B}_W,(\mathscr{F}_t)_{t\geq 0},\mu)$$

where μ is the Wiener measure. Let $(Y_t(x,w))_{t\geq 0, w\in W}$ be the regular solution provided by Theorem 8.1.5 of the stochastic differential equation

$$Y_t(x) = x + \int_0^t \sigma(u, Y_u(x))\mathrm{d}\pi_u + \int_0^t b(u, Y_u(x))\mathrm{d}u \quad \mu \text{ almost surely.}$$

We can check easily that the processes $Z_t(x,\omega) = Y_t(x, B(\omega))$, $t \geq 0$, $\omega \in \Omega$, and $\widetilde{Z}_t(x,\widetilde{\omega}) = Y_t(x, \widetilde{B}(\widetilde{\omega}))$, $t \geq 0$, $\widetilde{\omega} \in \widetilde{\Omega}$ are respectively solutions of the SDE satisfied by X and $\widetilde{X}$. The pathwise uniqueness of these solutions provided by Theorem 8.1.2 gives that $(Y_t(x, B(\omega))_{t\geq 0} = (X_t(x,\omega))_{t\geq 0}$ $\mathbb{P}$-a.s. and $(Y_t(x, \widetilde{B}(\widetilde{\omega}))_{t\geq 0} = (\widetilde{X}_t(x,\widetilde{\omega}))_{t\geq 0}$ $\widetilde{\mathbb{P}}$-a.s. But the Brownian motions B and $\widetilde{B}$ have same law on $W = \mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$, which is the Wiener measure μ, and therefore the processes X and $\widetilde{X}$ and $(Y_t)_{t\geq 0}$ have the same law on $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^q)$. ■

8.4 Martingale, generator, Kolmogorov equations, strong Markov property, Girsanov theorem

In this section, we consider the deterministic case and we assume furthermore that $\sigma(t,x)$ and $b(t,x)$ do not depend on the time variable t, in other words σ and b are two deterministic maps from $\mathbb{R}^q$ to $\mathscr{M}_{q,d}(\mathbb{R})$ and $\mathbb{R}^q$ respectively. This case is also referred to as the deterministic and time homogeneous case.

We denote by $(X_t(x))_{t\geq 0} = (X_t^0(x))_{t\geq 0}$ the regular solution of the SDE provided by Theorem 8.1.5:

$$X_t(x) = x + \int_0^t \sigma(X_u(x))\mathrm{d}B_u + \int_0^t b(X_u(x))\mathrm{d}u \quad \text{a.s.,} \quad t \geq 0,\ x \in \mathbb{R}^q.$$

Theorem 8.4.1. Simple Markov property.

For all $u \geq t \geq 0$ and all measurable and bounded $f : \mathbb{R}^q \mapsto \mathbb{R}$,

$$\mathbb{E}(f(X_u(x)) \mid \mathscr{F}_t) = \mathbb{E}(f(X_u(x)) \mid X_t(x)) = \Pi_{u-t}(f)(X_t(x)) \quad \text{a.s.}$$

where for all $s \geq 0$ and $x \in \mathbb{R}^q$,

$$\Pi_s(f)(x) = \mathbb{E}(f(X_s(x))).$$

In the case of the Ornstein – Uhlenbeck process of Example 8.2.2, we have the “Mehler formula”

$$\Pi_t(f)(x) = \mathbb{E}\left(f\left(x\mathrm{e}^{-\mu t} + \sigma\sqrt{\frac{1-\mathrm{e}^{-2\mu t}}{2\mu}}Z\right)\right) \quad \text{where} \quad Z \sim \mathcal{N}(0, I_d).$$

Note that with $\sigma = 1$ and $\mu \to 0$ we recover the heat kernel formula for Brownian motion, namely

$$\mathbb{E}(f(x + B_t)) = \mathbb{E}(x + \sigma\sqrt{t}Z).$$

Proof. Thanks to Theorem 8.3.1 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

$$X_u^t(x) = x + \int_0^{u-t} \sigma(X_{t+s}^t(x))\mathrm{d}B_s^t + \int_0^{u-t} b(X_{t+s}^t)\mathrm{d}s,$$

where $B_s^t = B_{t+s} - B_t$ for all $s \geq 0$, in other words, setting $Y_s(x) = X_{t+s}^t(x)$,

$$Y_s(x) = x + \int_0^s \sigma(Y_u(x))\mathrm{d}B_u^t + \int_0^s b(Y_u(x))\mathrm{d}u \quad \text{a.s.,} \quad s \geq 0,$$

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 the weak uniqueness property (Theorem 8.3.3), it follows that the processes $X(x)$ and $Y(x)$ have same law, and thus, for all $s \geq 0$,

$$\mathbb{E}(f(X_{t+s}^t(x))) = \mathbb{E}(f(X_s^0(x))).$$

■

For all $t \geq 0$, let $\Pi_t(\cdot, \mathrm{d}y)$ be the Markov transition kernel on $\mathbb{R}^q$ defined by

$$\Pi_t(x, A) = \mathbb{P}(X_t(x) \in A), \quad x \in \mathbb{R}^q, \ A \in \mathscr{B}_{\mathbb{R}^q}.$$

It acts on a bounded or positive measurable function $f : \mathbb{R}^q \to \mathbb{R}$ as

$$\Pi_t(f)(x) = \int_{\mathbb{R}^q} f(y)\Pi(x, \mathrm{d}y) = \mathbb{E}(f(X_t(x))), \quad x \in \mathbb{R}^q.$$

On bounded measurable functions, it defines a homogeneous Markov semi-group $(\Pi_t(x, \mathrm{d}y))_{t\geq 0}$,

$$\Pi_0 = \mathrm{Id}, \quad \Pi_s \circ \Pi_t = \Pi_{t+s}, \ s, t \geq 0.$$

In other words for all $s, t \geq 0$ and all $x \in \mathbb{R}^q$,

$$\Pi_{s+t}(x, \mathrm{d}y) = \int_{\mathbb{R}^q} \Pi_t(x, \mathrm{d}z)\Pi_s(z, \mathrm{d}y) = (\Pi_t\Pi_s)(x, \mathrm{d}y)).$$

Theorem 8.4.2. Markov semi-group properties.

For all $t \geq 0$ the operator Π_t preserves globally

1. the set $\mathscr{M}(\mathbb{R}^q, \mathbb{R})$ of bounded and measurable functions $\mathbb{R}^q \to \mathbb{R}$
2. the set $\mathscr{C}_b(\mathbb{R}^q, \mathbb{R})$ of bounded and continuous functions $\mathbb{R}^q \to \mathbb{R}$
3. the set $\mathscr{C}_0(\mathbb{R}^q, \mathbb{R})$ of bounded and continuous functions $\mathbb{R}^q \to \mathbb{R}$ vanishing at infinity provided however that the coefficients σ and b are bounded.

The stability of bounded continuous functions is known as the Feller[3] continuity.

The condition for the last property to hold is very simple but too restrictive, for instance the Ornstein–Uhlenbeck process satisfies the property as one can check using the Mehler formula and dominated convergence, while the drift is not bounded.

Proof.

1. Immediate for a Markov transition kernel
2. We need to establish preservation of continuity. Let $t \geq 0$, $f \in \mathscr{C}_b(\mathbb{R}^q, \mathbb{R})$, $x = \lim_{n\to\infty} x_n \in \mathbb{R}^q$. We have,
$$|\Pi_t(f)(x_n) - \Pi_t(f)(x)| = |\mathbb{E}(f(X_t(x_n))) - \mathbb{E}(f(X_t(x)))|.$$
Since by Theorem 8.1.3, $\mathbb{E}(|X_t(x_n) - X_t(x)|^2) \leq C_t|x_n - x|^2$, it follows that $\lim_{n\to\infty} X_t(x_n) = X_t(x)$ in L^2, and thus in law, and therefore $\lim_{n\to\infty} \Pi_t(f)(x_n) = \Pi_t(f)(x)$, which implies $\Pi_t(f) \in \mathscr{C}_b(\mathbb{R}^q, \mathbb{R})$.
3. It suffices to establish the preservation of nullity at infinity. Let $f \in \mathscr{C}_0(\mathbb{R}^q, \mathbb{R})$ and $\varepsilon > 0$. There exists $A > 0$ such that for all $y \in \mathbb{R}^q$ such that $|y| > A$, we have $|f(y)| < \varepsilon$. Let $(X_t(x))_{t\geq 0}$ be the solution of the stochastic differential equation associated to the semi-group, namely
$$X_t(x) = x + \int_0^t \sigma(X_s(x))\mathrm{d}B_s + \int_0^t b(X_s(x))\mathrm{d}s.$$
We have, for all $x \in \mathbb{R}^q$ such that $|x| > B > A$, using the Markov inequality and the Itô isometry,
$$\begin{aligned}|\mathbb{E}(f(X_t(x))| &\leq \mathbb{E}(|f(X_t(x))|\mathbf{1}_{|X_t(x)|>A}) + \|f\|_\infty \mathbb{P}(|X_t(x)| \leq A)\\ &\leq \varepsilon + \|f\|_\infty \mathbb{P}\Big(\Big|\int_0^t \sigma(X_s(x))\mathrm{d}B_s + \int_0^t b(X_s(x))\mathrm{d}s\Big| \geq B - A\Big)\\ &\leq \varepsilon + \frac{\|f\|_\infty}{(B-A)^2}\mathbb{E}\Big(\Big|\int_0^t \sigma(X_s(x))\mathrm{d}B_s + \int_0^t b(X_s(x))\mathrm{d}s\Big|^2\Big)\\ &\leq \varepsilon + 2\frac{\|f\|_\infty}{(B-A)^2}(\||\sigma|\|_\infty^2 t + (\||b|\|_\infty t)^2)\\ &\leq 2\varepsilon \text{ for } B \text{ sufficiently large.}\end{aligned}$$

■

[3] Named after William Feller (1906–1970), Croatian-American mathematician specializing in probability theory.

Remark 8.4.3. Brownian motion and the Laplacian.

If $f:\mathbb{R}^q\to\mathbb{R}$ is $\mathscr{C}^2$ then the Itô formula gives, for all $x\in\mathbb{R}^q$ and $t>0$,

$$f(x+B_t)=f(x)+\int_0^t\nabla f(x+B_s)\mathrm{d}B_s+\frac{1}{2}\int_0^t(\Delta f)(x+B_s)\mathrm{d}s.$$

If f has bounded first and second order derivatives then in particular the stochastic integral term is a centered martingale and the second integral is integrable, and by taking the expectation we get

$$\mathbb{E}(f(x+B_t))=f(x)+\int_0^t\mathbb{E}\Big(\frac{1}{2}(\Delta f)(x+B_s)\Big)\mathrm{d}s.$$

If we denote $\Pi_t(f)(x)=\mathbb{E}(f(x+B_t))$, this implies

$$\partial_{t=0}\Pi_t(f)(x)=\frac{1}{2}\Delta f(x).$$

Alternatively, if we do not know the Itô formula and if $f:\mathbb{R}^q\to\mathbb{R}$ has third order derivative with bounded second order derivative (Hessian), then a Taylor formula gives,

$$f(x+h)=f(x)+\sum_{i=1}^q\frac{\partial f}{\partial x_i}(x)h_i+\frac{1}{2}\sum_{i,j=1}^q\frac{\partial^2 f}{\partial x_i\partial x_j}(x)h_ih_j+\sum_{\substack{k_1\geq0,\ldots,k_q\geq0\\k_1+\cdots+k_q=3}}r_{x,k}(h)h_1^{k_1}\cdots h_q^{k_q}$$

where $r_{x,k}$ is continuous with $\lim_{h\to0}r_{x,k}(h)=0$ and $\sup_{h\in\mathbb{R}^q}|r_{x,k}(h)|<\infty$. Taking $h=B_t$, $t\geq0$, and the expectation, using the independence of the components $B_t^1,\ldots,B_t^q$ of B_t, their mean 0, variance t, and third absolute moment $t^{3/2}=o(t)$, and dominated convergence for the remainder term,

$$\mathbb{E}(f(x+B_t))=f(x)+t\frac{1}{2}\Delta f(x)+o(t).$$

In other words, here again, the partial derivative of $\Pi_t(f)(x)=\mathbb{E}(f(x+B_t))$ at $t=0$ is equal to $\frac{1}{2}\Delta f(x)$. We will see that more generally, we can associate to the solution of (SDE) a second order partial differential operator, involved in formulas related to the time derivative of the Markov semigroup.

Let $\mathscr{C}^2(\mathbb{R}^q,\mathbb{R})$ be the space of functions $\mathbb{R}^q\to\mathbb{R}$ of class $\mathscr{C}^2$ in other words twice differentiable with continuous second derivative (Hessian). We define the second order linear partial differential operator without constant term $L:\mathscr{C}^2(\mathbb{R}^q,\mathbb{R})\to\mathscr{C}(\mathbb{R}^q,\mathbb{R})$, by, for all $f\in\mathscr{C}^2(\mathbb{R}^q,\mathbb{R})$ and all $x\in\mathbb{R}^q$,

$$L(f)(x)=\frac{1}{2}\sum_{i,j=1}^q a_{i,j}(x)\frac{\partial^2 f}{\partial x_i\partial x_j}(x)+\sum_{i=1}^q b_i(x)\frac{\partial f}{\partial x_i}(x),\tag{L}$$

where $b(x)=(b_1(x),\ldots,b_q(x))$ 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 $x\in\mathbb{R}^q$, the matrix $a(x)$ is symmetric, and positive[4] since for all $y\in\mathbb{R}^q$,

$$\langle a(x)y,y\rangle=|\sigma(x)^\top y|^2\geq0.$$

We say that L is elliptic when the inequality is >0 for all x and all $y\neq0$, and sub-elliptic otherwise. There are also other notions such as uniformly elliptic and hypo-elliptic, which are outside the scope of this course.

If L is elliptic in the sense that $a(x)$ has full rank d for all $x\in\mathbb{R}^q$, then there exists linearly independent vector fields $V_0,V_1,\ldots,V_d$ on $\mathbb{R}^q$ such that $L=V_1^2+\cdots+V_d^2+V_0$, see for instance [4, Proposition 6.32].

[4]When the inequality is strict whenever $y\neq0$ it is customary in matrix analysis to say that $a(x)$ is positive definite.

Example 8.4.4. Langevin, Ornstein – Uhlenbeck, Brownian motion.

From Example 8.2.2 & 8.2.7, we get

Process	SDE	Operator
Brownian motion	$\mathrm{d}X_t = \mathrm{d}B_t$	$L(f)(x) = \frac{1}{2}(\Delta f)(x)$
Ornstein – Uhlenbeck	$\mathrm{d}X_t = \sigma \mathrm{d}B_t - \mu X_t \mathrm{d}t$	$L(f)(x) = \frac{\sigma^2}{2}(\Delta f)(x) - \mu x \cdot \nabla f(x)$.
Overdamped Langevin	$\mathrm{d}Z_t = \sigma \mathrm{d}B_t - \nabla V(Z_t)\mathrm{d}t$	$L(f)(x) = \frac{\sigma^2}{2}(\Delta f)(x) - \nabla V(x) \cdot \nabla f$.

For the underdamped Lagevin, we find $L(f)(y,z) = \gamma\frac{\sigma^2}{2}(\Delta_z f) - \gamma \nabla V(z) \cdot \nabla_z f - \nabla U(y) \cdot \nabla_z f + \nabla V(z) \cdot \nabla_y f$.

Theorem 8.4.5. Martingale, generator, Duhamel[a] formula, and Kolmogorov equation.

[a] Named after Jean-Marie Duhamel (1797 – 1872), French mathematician.

Let $x \in \mathbb{R}^q$ and let $(X_t(x))_{t\geq 0}$ be the regular solution of the SDE as in Theorem 8.4.1.

- For all $f \in \mathscr{C}^2(\mathbb{R}^q, \mathbb{R})$, the following process is an $(\mathscr{F}_t)_{t\geq 0}$ local martingale issued from the origin:

$$M^f = (M^f_t)_{t\geq 0} = \left(f(X_t(x)) - f(x) - \int_0^t Lf(X_s(x))\mathrm{d}s\right)_{t\geq 0},$$

  where L is the differential operator defined in (L) and where

$$\langle M^f \rangle = \int_0^\bullet \Gamma(f)(X_s(x))\mathrm{d}s \quad \text{where} \quad \Gamma(f)(x) = |\sigma(x)^\top \nabla f(x)|^2 = a(x)\nabla f(x) \cdot \nabla f(x).$$

- For all $f \in \mathscr{C}^2_b(\mathbb{R}^q, \mathbb{R})$ in other words has bounded first and second order derivatives then M^f is a martingale with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$ and we have the Duhamel formula:

$$\Pi_t(f)(x) = f(x) + \int_0^t \Pi_s(L(f))(x)\mathrm{d}s, \quad t \geq 0,$$

  in particular we obtain the Kolmogorov equation

$$\partial_t \Pi_t(f) = \Pi_t(Lf).$$

If $Lf = 0$, we say that f is harmonic with respect to L, and $M^f = (f(X_t) - f(x))_{t\geq 0}$ is a local martingale.
The functional quadratic form Γ is known as the carré du champ operator of the Markov process X. The formula $\Gamma(f) = \frac{1}{2}L(f^2) - fLf$ is at the heart of [2]. When σ is constant and equal to I_q then $\Gamma f = |\nabla f|^2$.
We say that X solves a Stroock – Varadhan martingale problem with respect to the differential operator L, see [47, 21, 15, 24].

Proof. Let $X^i(x)$, $1 \leq i \leq q$, be the coordinates processes of the process $X(x)$. Let $f \in \mathscr{C}^2(\mathbb{R}^q, \mathbb{R})$. From the SDE, $X^i(x)$ is a continuous semi-martingale with martingale part

$$M^i = \int_0^\bullet \sigma_{i,\bullet}(X_s)\mathrm{d}B_s = \sum_{k=1}^d \int_0^\bullet \sigma_{i,k}(X_s(x))\mathrm{d}B^k_s$$

and finite variation part $\int_0^\bullet b_i(X_s(x))\mathrm{d}s$. Now the idea is to use the Itô formula for $f(X(x))$ and to collect all the non-martingale parts into an operator. The Itô formula of Theorem 7.1.1 applied to $f(X_t(x))$ gives

$$\begin{aligned} &f(X_t(x)) \\ &= f(x) + \sum_{i=1}^q \int_0^t \frac{\partial f}{\partial x_i}(X_s(x))\mathrm{d}M^i_s + \sum_{i=1}^q \int_0^t \frac{\partial f}{\partial x_i}(X_s(x))b_i(X_s(x))\mathrm{d}s + \frac{1}{2}\sum_{i,j=1}^q \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s(x))\mathrm{d}\langle M^i, M^j\rangle_s. \end{aligned}$$

But now, since we have

$$\langle M^i, M^j\rangle = \sum_{k,\ell=1}^{d}\int_0^t \sigma_{i,k}(X_s(x))\sigma_{j,\ell}(X_s(x))\mathrm{d}\langle B^k, B^\ell\rangle_s = \sum_{k=1}^{d}\int_0^t \sigma_{i,k}(X_s(x))\sigma_{j,k}(X_s(x))\mathrm{d}s = \int_0^t a_{i,j}(X_s(x))\mathrm{d}s,$$

we get that

$$M_t^f = f(X_t(x)) - f(x) - \int_0^t L(f)(X_s(x))\mathrm{d}s = \sum_{i=1}^{q}\int_0^t \frac{\partial f}{\partial x_i}(X_s(x))\mathrm{d}M_s^i = \sum_{i=1}^{q}\sum_{k=1}^{d}\int_0^t \frac{\partial f}{\partial x_i}(X_s(x))\sigma_{i,k}(X_s(x))\mathrm{d}B_s^k$$

is an $(\mathscr{F}_t)_{t\geq 0}$ local martingale. Note that in particular, for all $1\leq i\leq q$,

$$\langle M^i\rangle = \langle M^i, M^i\rangle = \int_0^\bullet a_{i,i}(X_s(x))\mathrm{d}s = \int_0^\bullet \sum_{k=1}^{d}\sigma_{i,k}^2(X_s(x))\mathrm{d}s.$$

Moreover, using the fact that $\langle B^k, B^{k'}\rangle_s = s\mathbf{1}_{k=k'}$,

$$\langle M^f\rangle_t = \int_0^t \sum_{k=1}^{d}\sum_{i,j=1}^{q}\frac{\partial f}{\partial x_i}(X_s(x))\frac{\partial f}{\partial x_j}(X_s(x))\sigma_{i,k}(X_s(x))\sigma_{j,k}(X_s(x))\mathrm{d}s.$$

Furthermore, if now $f\in\mathscr{C}_b^2(\mathbb{R}^q,\mathbb{R})$, then for all $1\leq i\leq q$, by using the boundedness of $\partial_i f$, the Fubini–Tonelli theorem, and the square integrability of $\sigma_{i,k}(X_s(x))$ which comes from **(Lip)** and Theorem 8.1.2,

$$\mathbb{E}\int_0^t\Big(\frac{\partial f}{\partial x_i}(X_s(x))\Big)^2\mathrm{d}\langle M^i\rangle_s = \mathbb{E}\int_0^t\Big(\frac{\partial f}{\partial x_i}(X_s(x))\Big)^2\sum_{k=1}^{d}\sigma_{i,k}^2(X_s(x))\mathrm{d}s<\infty,$$

and thus M^f is a martingale as a finite sum of martingales. Finally, when M^f is a martingale, then the initial condition $M_0^f=0$ gives $\mathbb{E}(M_t^f)=0$ for all $t\geq 0$, and the Duhamel formula follows from the expression of M^f by taking expectations and using the Fubini–Tonelli theorem. ■

Corollary 8.4.6. Infinitesimal generator of Markov semi-group.

The following properties hold true:

1. **Continuity.** For all $f\in\mathscr{C}_0(\mathbb{R}^q,\mathbb{R})$, $\lim_{t\to 0^+}\|\Pi_t(f)-f\|_\infty = 0$
2. **Differentiability.** For all $f\in\mathscr{C}_c^2(\mathbb{R}^q,\mathbb{R})$ in other words $\mathscr{C}^2(\mathbb{R}^q,\mathbb{R})$ with compact support,

$$\lim_{t\to 0^+}\Big\|\frac{\Pi_t(f)-f}{t}-Lf\Big\|_\infty = 0.$$

We say that L is the infinitesimal generator of the semigroup $(\Pi_t)_{t\geq 0}$, and formally $\Pi_t = \mathrm{e}^{tL}$.

Proof.

1. The Duhamel formula of Theorem 8.4.5 gives, for all $g\in\mathscr{C}_b^2(\mathbb{R}^q,\mathbb{R})$,

$$\Pi_t(g)(x) - g(x) = \int_0^t \Pi_s(L(g))(x)\mathrm{d}s = \mathbb{E}\int_0^t L(g)(X_s(x))\mathrm{d}s$$

and thus $\|\Pi(g)-g\|_\infty\leq t\|Lg\|_\infty\to 0$ as $t\to 0^+$. Now if $f\in\mathscr{C}_0(\mathbb{R}^q,\mathbb{R})$, then, for all $\varepsilon>0$, we there exists $g\in\mathscr{C}_b^2(\mathbb{R}^q,\mathbb{R})$ such that $\|f-g\|_\infty\leq\varepsilon$, and if follows then that for $t>0$ small enough,

$$\|\Pi_t(f)-f\|_\infty \leq \|\Pi_t(f-g)\|_\infty + \|\Pi_t(g)-g\|_\infty + \|g-f\|_\infty \leq 2\varepsilon + \|\Pi_t(g)-g\|_\infty \leq 3\varepsilon.$$

Let us detail an approximation argument to construct g. Since $f\in\mathscr{C}_0(\mathbb{R}^q,\mathbb{R})$, by the Heine theorem, f is uniformly continuous and thus there exists $\eta>0$ such that for all $x,y\in\mathbb{R}^q$, if $|x-y|\leq\eta$ then $|f(x)-f(y)|\leq\varepsilon$. Next, let $\rho\in\mathscr{C}^\infty(\mathbb{R}^q,\mathbb{R}_+)$ be a compactly supported probability density function with support included in the ball $\{z\in\mathbb{R}^q : |z|\leq\eta\}$. We have $g=f*\rho\in\mathscr{C}_b^\infty(\mathbb{R}^q,\mathbb{R})$, and, for all $x\in\mathbb{R}^q$,

$$|f(x)-g(x)|\leq\int_{\mathbb{R}^q}|f(x)-f(y)|\rho(x-y)\mathrm{d}y = \int_{|x-y|\leq\eta}|f(x)-f(y)|\rho(x-y)\mathrm{d}y\leq\varepsilon.$$

2. For all $f \in \mathscr{C}_c^2(\mathbb{R}^q, \mathbb{R})$ and all $t > 0$, we have, from Duhamel formula, using the first part for the last step,
$$\left\| \frac{\Pi_t(f) - f}{t} - Lf \right\|_\infty = \left\| \frac{1}{t} \int_0^t (\Pi_s(L(f)) - L(f)) \mathrm{d}s \right\|_\infty \le \sup_{s \in [0,t]} \|\Pi_s(L(f)) - L(f)\|_\infty \underset{t \to 0^+}{\longrightarrow} 0.$$
(we have used the fact that $f \in \mathscr{C}_c^2$ gives $Lf \in \mathscr{C}_c \subset \mathscr{C}_0$).

■

Theorem 8.4.7. Strong Markov property.

Let $x \in \mathbb{R}^q$ and $(X_t(x))_{t \ge 0}$ be the regular solution of the SDE as in Theorem 8.4.1. Let T be an $(\mathscr{F}_t)_{t \ge 0}$ stopping time and let $\mathscr{F}_T$ be its stopping σ-algebra.

1. For all bounded measurable $f : \mathbb{R}^q \to \mathbb{R}$, and all $t \ge 0$,
$$\mathbb{E}(f(X_{T+t}(x)) \mathbf{1}_{T<\infty} \mid \mathscr{F}_T) = \Pi_t(f)(X_T(x)) \mathbf{1}_{T<\infty}.$$
2. For all bounded measurable $\Phi : \mathscr{C}(, \mathbb{R}^q, \mathbb{R}_+) \to \mathbb{R}$,
$$\mathbb{E}(\Phi((X_{T+s}(x))_{s \ge 0}) \mathbf{1}_{T<\infty} \mid \mathscr{F}_T) = \Psi(X_T(x)) \mathbf{1}_{T<\infty}$$
where the measurable function $\Psi : \mathbb{R}^q \to \mathbb{R}$ is defined for all $y \in \mathbb{R}^q$ by
$$\Psi(y) = \mathbb{E}\Phi((X_s(y))_{s \ge 0}).$$

When $T = s$ (deterministic) then we recover the weak Markov property (Theorem 8.3.1).
If $d = q$, $\sigma = I_d$, and $b = 0$, we recover the strong Markov property for Brownian motion (Theorem 3.5.1).

This can be skipped at first reading.

Proof.

1. Suppose first that T takes its values in an at most countable set $\mathscr{T} \subset [0, \infty]$. We have to show that for all $A \in \mathscr{F}_T$ and for all $t \ge 0$,
$$\mathbb{E}(f(X_{T+t}) \mathbf{1}_{A \cap \{T<\infty\}}) = \mathbb{E}(\Pi_t(f)(X_T(x)) \mathbf{1}_{A \cap \{T<\infty\}}).$$
Indeed, using the simple Markov property of Theorem 8.4.1, the left hand side is equal to
$$\sum_{r \in \mathscr{T} \setminus \{\infty\}} \mathbb{E}(f(X_{r+t}(x)) \mathbf{1}_{A \cap \{T=r\}}) = \sum_{r \in \mathscr{T} \setminus \{\infty\}} \mathbb{E}(\Pi_t(f)(X_r(x)) \mathbf{1}_{A \cap \{T=r\}}) = \mathbb{E}(\Pi_t(f)(X_T(x)) \mathbf{1}_{A \cap \{T<\infty\}}).$$
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 $n \ge 0$, the discretized stopping time
$$T_n = \sum_{k \ge 0} \frac{k+1}{2^n} \mathbf{1}_{[k/2^n, (k+1)/2^n)}(T) + \infty \mathbf{1}_{T=\infty}.$$
We have $T_n \searrow T$. For all $n \ge 0$ and all $A \in \mathscr{F}_{T_n}$, we get, from the first part of the proof,
$$\mathbb{E}(f(X_{T_n+t}(x)) \mathbf{1}_{A \cap \{T_n<\infty\}}) = \mathbb{E}(\Pi_t(f)(X_{T_n}(x)) \mathbf{1}_{A \cap \{T_n<\infty\}}).$$
By letting $n \to \infty$ and using the right-continuity of X and dominated convergence, we obtain,
$$\mathbb{E}(f(X_{T+t}(x)) \mathbf{1}_{A \cap \{T<\infty\}}) = \mathbb{E}(\Pi_t(f)(X_T(x)) \mathbf{1}_{A \cap \{T<\infty\}}),$$
where we also used Theorem 8.4.2 about the continuity of Π_t to get
$$\Pi_t(f)(X_{T_n}(x)) \mathbf{1}_{T_n<\infty} \underset{n \to \infty}{\longrightarrow} \Pi_t(f)(X_T(x)) \mathbf{1}_{T<\infty} \quad \text{a.s.}$$

2. Suppose that Φ 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 $f_1, \ldots, f_n : \mathbb{R}^q \mapsto \mathbb{R}$, we have, for all $w \in W$,

$$\Phi(w) = f_1(w_{s_1}) \cdots f_n(w_{s_n}).$$

We have in this case to show that

$$\mathbb{E}(f_1(X_{T+s_1}(x)) \cdots f_n(X_{T+s_n}(x))\mathbf{1}_{T<\infty} \mid \mathscr{F}_T) = \Psi(X_T(x))\mathbf{1}_{T<\infty} \quad \text{a.s.}$$

where $\Psi : \mathbb{R}^q \to \mathbb{R}$ is the function defined for all $y \in \mathbb{R}^q$ by

$$\Psi(y) = \mathbb{E}(f_1(X_{s_1}(y)) \cdots f_n(X_{s_n}(y))).$$

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 $Y_i = f_i(X_{T+s_i}(x))$,

$$\begin{aligned}\mathbb{E}(Y_1 \cdots Y_n Y_{n+1}\mathbf{1}_{T<\infty} \mid \mathscr{F}_T) &= \mathbb{E}(Y_1 \cdots Y_n \mathbb{E}(Y_{n+1}\mathbf{1}_{T<\infty} \mid \mathscr{F}_{s_n+T}) \mid \mathscr{F}_T)\\ &= \mathbb{E}(Y_1 \cdots Y_n \Pi_{s_{n+1}-s_n}(f_{n+1})(X_{T+s_n}(y))\mathbf{1}_{T<\infty} \mid \mathscr{F}_T)\\ &= \Psi(X_T(x))\mathbf{1}_{T<\infty}\end{aligned}$$

where

$$\Psi(y) = \mathbb{E}(f_1(X_{s_1}(y)) \cdots f_n(X_{s_n}(y))\Pi_{s_{n+1}-s_n}(f_{n+1})(X_{s_n}(y))).$$

But using the induction hypothesis and the simple Markov property of Theorem 8.4.1,

$$\begin{aligned}\Psi(y) &= \mathbb{E}(f_1(X_{s_1}(y)) \cdots f_n(X_{s_n}(y))\mathbb{E}(f_{n+1}(X_{s_{n+1}}(y)) \mid \mathscr{F}_{s_n}))\\ &= \mathbb{E}(f_1(X_{s_1}(y)) \cdots f_n(X_{s_n}(y)) f_{n+1}(X_{s_{n+1}}(y))).\end{aligned}$$

This gives the result for Φ cylindrical. Finally we can use monotone classes (Section 1.8).

■

Theorem 8.4.8. Heat-type equation and Kolmogorov equation.

Assume that σ and b are moreover $\mathscr{C}^2_b$, in other words $\mathscr{C}^2$ with bounded first and second derivatives. Let L be the differential operator defined in (L). Let $f \in \mathscr{C}^2_b(\mathbb{R}^q, \mathbb{R})$. Then:

- There exists a unique $\Psi = (\Psi(t,x))_{t\geq 0, x\in\mathbb{R}^q}$ solution of the following problem:
  - $(t,x) \mapsto \Psi(t,x)$ is $\mathscr{C}^1$ in t and $\mathscr{C}^2_b$ in x
  - for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^q$,

  $$\frac{\partial \Psi}{\partial t}(t,x) = L(\Psi(t,\cdot))(x) = \frac{1}{2}\sum_{i,j=1}^q a_{i,j}(x)\frac{\partial^2 \Psi}{\partial x_i \partial x_j}(t,x) + \sum_{i=1}^q b_i(x)\frac{\partial \Psi}{\partial x_i}(t,x).$$

  - for all $x \in \mathbb{R}^q$, $\Psi(0,x) = f(x)$.
- For all $x \in \mathbb{R}^q$ and $t \geq 0$, denoting $(X_t(x))_{t\geq 0}$ the solution of the SDE as in Theorem 8.4.1,

  $$\Psi(t,x) = \mathbb{E}(f(X_t(x))) = \Pi_t(f)(x).$$

  In particular the infinitesimal generator L determines the Markov semi-group $(\Pi_t)_{t\geq 0}$ which characterizes the law of the Markov diffusion process $(X_t(x))_{t\geq 0}$.
- In other words, we have the Kolmogorov equation

  $$\partial_t \Pi_t(f) = L\Pi_t(f).$$

Remark 8.4.9. Kolmogorov equations and Markov processes.

The combination of the Kolmogorov equations provided by Theorem 8.4.5 and Theorem 8.4.8 provides a commutation between semi-group and generator, namely

$$\partial_t \Pi_t = L\Pi_t = \Pi_t L.$$

The special case $\partial_{t=0}\Pi_t(f) = \Pi_0(Lf) = L\Pi_0(f) = f$ is already provided by Corollary 8.4.6. The Kolmogorov equation is an essential feature of Markov processes. It expresses the remarkable fact that a Markov process is a deterministic evolution in time of distributions at time t. The determinism is at the level of the distribution of the stochastic process instead of being at the level of the trajectories of the stochastic process. In the case of diffusion processes solutions of (SDE), the deterministic evolution is described by a linear partial differential operator (L) of second order without constant term. The identification of the deterministic behavior of distributions of stochastic phenomena was a great scientific discovery, which can be traced back to Laplace[a], Quetelet[b], Boltzmann[c], and Maxwell[d], among others, connected in a way to the mechanical view of nature developed by Darwin[e].

[a] Pierre-Simon Laplace (1749 – 1827), French scholar and polymath.
[b] Adolphe Quetelet (1796 – 1874), Belgian astronomer, mathematician, statistician and sociologist.
[c] Ludwig Boltzmann (1844 – 1906), Austrian physicist and philosopher.
[d] James Clerk Maxwell (1831 – 1879), Scottish scientist in the field of mathematical physics.
[e] Charles Darwin (1809 – 1882), English naturalist, geologist and biologist.

Remark 8.4.10. Kolmogorov equations and Fokker[a] – Planck[b] equation.

[a] Named after Adriaan Fokker (1887 – 1972), Dutch physicist and musician.
[b] Named after Max Plack (1858 – 1947), German theoretical physicist.

If we denote by μ_t the law of X_t, then the Kolmogorov equation provided by Theorem 8.4.5 writes $\partial_t \int f \mathrm{d}\mu_t = \int Lf \mathrm{d}\mu_t$. This is, in the sense of Schwartz distributions, the Fokker – Planck equation,

$$\partial_t \mu_t = L\mu_t.$$

If μ_t has density p_t then, denoting L^* the adjoint of L,

$$L^* f(x) = \frac{1}{2}\sum_{i,j=1}^{q} \partial^2_{i,j}(a_{i,j}(x)f(x)) + \sum_{i=1}^{q}\partial_i(b_i(x)f(x))$$

and the Fokker – Planck equation becomes

$$\partial_t p_t = L^* p_t.$$

This is also known as the forward Kolmogorov equation, while the one provided by Theorem 8.4.5 or Theorem 8.4.8 is known as the backward Kolmogorov equation. The terms forward and backward can also be understood from the formula $\Pi_{t-s} = \Pi_{s,t}$, which allows to take the derivative with respect to t (forward in time) or with respect to s (backward in time).

This can be skipped at first reading.

Proof. We admit the following result, which relies on the assumptions made on σ and b, see [18]: for all $f \in \mathscr{C}^2_b(\mathbb{R}^q,\mathbb{R})$, the quantity $\Pi_t(f)(x)$ is $\mathscr{C}^2_b$ in x.
The fact that $\Pi_t(f)(x)$ is $\mathscr{C}^1$ in t can be checked on the Duhamel formula of Theorem 8.4.5.
Let $u \ge t > 0$. The Itô formula (Theorem 7.1.1) for function $\Pi_{u-t}(f)$ and semi-martingale $X(x)$ gives, proceeding as in the proof of Theorem 8.4.5,

$$\Pi_{u-t}(f)(X_t(x)) = \Pi_u(f)(x) + N_t + \int_0^t (L-\partial_u)(\Pi_{u-s}(f))(X_s(x))\mathrm{d}s$$

where $(N_t)_{t\geq 0}$ is a continuous local martingale. But as observed in Remark 8.3.2, the process $(\Pi_{u-t}(f)(X_t(x)))_{t\in[0,u]}$ is a continuous martingale. It follows then that the finite variation process

$$\Big(\int_0^t (L-\partial_u)(\Pi_{u-s}(f))(X_s(x))\mathrm{d}s\Big)_{t\geq 0}$$

is a continuous local martingale, issued from zero, and thus identically equal to zero. Thus

$$(L-\partial_u)(\Pi_u(f))(x) = \lim_{t\to 0}\frac{1}{t}\int_0^t \Pi_s((L-\partial_u)(\Pi_{u-s}(f))(x)\mathrm{d}s = 0.$$

Therefore the formula for Ψ in the statement of the theorem provides a solution to the problem (heat equation) considered in 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 for $(\Psi(u-t,X_t(x)))_{t\in[0,u]}$ gives

$$\begin{aligned}\Psi(0,X_u(x)) &= f(X_u(x))\\ &= \Psi(u,x)+\widetilde{N}_u+\int_0^u\Big[-(\partial_u\Psi)(u-t,X_t(x))+L(\Psi(u-t,\cdot))(X_t(x))\Big]\mathrm{d}t\\ &= \Psi(u,x)+\widetilde{N}_u\end{aligned}$$

where $(\widetilde{N}_u)_{u\geq 0}$ is a stochastic integral with zero expectation, and therefore

$$\Psi(u,x) = \mathbb{E}(f(X_u(x))) = \Pi_u(f)(x).$$

■

Theorem 8.4.11. Girsanov theorem for overdamped Langevin process.

Let $T>0$ be a fixed number. Set $s=0$, $d=q$, $\sigma=I_d$, $b=-\nabla V$ for a "potential" $V\in\mathscr{C}_b^2(\mathbb{R}^d,\mathbb{R})$. Then

$$X_t(x) = x+B_t-\int_0^t \nabla V(X_s(x))\mathrm{d}s,$$

and the law of $(X_t(x))_{t\in[0,T]}$, seen as a random variable on the canonical space $\mathscr{C}([0,T],\mathbb{R})$, is absolutely continuous with respect to the Wiener measure, with density $w\in\mathscr{C}([0,T],\mathbb{R}^d)\mapsto \mathrm{e}^{-H(w)}$ where

$$H(w) = V(x+w_T)-V(x)+\frac{1}{2}\int_0^T(|\nabla V|^2-\Delta V)(x+w_s)\mathrm{d}s.$$

Proof. Thanks to the proof of the Girsanov theorem (Theorem 7.5.1), $X=(X_t(x))_{t\in[0,T]}$ has density

$$\exp\Big(-\int_0^T\nabla V(X_s)\mathrm{d}X_s-\frac{1}{2}\int_0^T|\nabla V(X_s)|^2\mathrm{d}s\Big)$$

with respect to $\mathbb{Q}$. Now the Itô formula (Theorem 7.1.1) gives

$$V(X_T) = V(x)+\int_0^t\nabla V(X_s)\mathrm{d}X_s+\frac{1}{2}\int_0^t(\Delta V)(X_s)\mathrm{d}s,$$

Therefore

$$-\int_0^T\nabla V(X_s)\mathrm{d}X_s = V(x)-V(X_T(x))+\frac{1}{2}\int_0^T\Delta V(X_s)\mathrm{d}s,$$

hence the formula, since according to Theorem 7.5.1, X with respect to $\mathbb{Q}$ has the law of $x+B$. ■

There are other instances of the Girsanov theorem, for instance with a general L, or between the solutions of two SDE driven by the same BM with the same diffusion coefficient σ but with distinct drifts b.

This can be skipped at first reading.

8.5 Locally Lipschitz coefficients and explosion time

We assume in this section that $\sigma(t,\omega,x)$ and $b(t,\omega,x)$ depend neither on the randomness ω nor on the time t. They are defined on $\mathbb{R}^q$ and take values in $\mathcal{M}_{q,d}(\mathbb{R})$ and $\mathbb{R}^q$. We also assume that they are locally Lipschitz[a]: for all bounded $K \subset \mathbb{R}^q$, there exists a constant $C_K > 0$ such that for all $x, y \in K$,

$$|\sigma(x) - \sigma(y)| + |b(x) - b(y)| \leq C_K |x - y|.$$

Beware that these assumptions are not a specialization of the general assumptions made at the beginning of the chapter: deterministic is less general than random, but locally Lipschitz is more general than Lipschitz. The main problem with these assumptions on σ and b is that the SDE

$$X_t(x) = x + \int_0^t \sigma(X_s(x))\mathrm{d}B_s + \int_0^t b(X_s(x))\mathrm{d}s$$

may not have a solution $X_t(x)$ for all time $t \geq 0$, and an explosion may occur in finite (random) time. A way to define a solution for all time is to use a localization procedure in order to define the process before explosion, and then to stick the process to an extra point at infinity after explosion. We use the Alexandroff[b] compactification $\mathbb{R}^q \cup \{\infty\}$ of $\mathbb{R}^q$ obtained by adding to $\mathbb{R}^q$ a point at infinity denoted ∞. The neighborhoods of ∞ in $\mathbb{R}^q \cup \{\infty\}$ are the complements of the closed proper subsets of $\mathbb{R}^q$.

Theorem 8.5.1: Solving SDE with locally Lipschitz coefficients

For all $x \in \mathbb{R}^q$, there exists a unique couple (X^x, ξ^x) where ξ^x is a stopping time taking values in $(0,\infty]$ called the explosion time and where $X^x = (X_t(x))_{t\geq 0}$ is an adapted process with values in $\mathbb{R}^q \cup \{\infty\}$ such that the following properties hold:

1. a.s. the path $t \mapsto X_t(x)$ is continuous from $[0,\xi^x)$ to $\mathbb{R}^q$ and $X_t(x) = \infty$ for all $t \geq \xi^x$
2. almost surely, on the event $\{\xi^x < \infty\}$,
$$\lim_{t \overset{<}{\to} \xi^x} |X_t(x)| = +\infty$$
3. for all stopping time T such that $\{T < \xi^x\}$ almost surely on $\{\xi^x < \infty\}$,
$$X_{t\wedge T}(x) = x + \int_0^t \mathbf{1}_{s\leq T}\sigma(X_s(x))\mathrm{d}B_s + \int_0^t \mathbf{1}_{s\leq T} b(X_s(x))\mathrm{d}s \quad \text{a.s.,} \quad t \geq 0.$$
(this SDE has random coefficients, it involves only the values of X at times before T).

Before giving the proof of Theorem 8.5.1, let us prepare some ingredients.
From the assumption on σ and b, for all $n \geq 1$, there exist maps

$$\sigma_n : \mathbb{R}^q \to \mathcal{M}_{q,d}(\mathbb{R}) \quad \text{and} \quad b_n : \mathbb{R}^q \to \mathbb{R}^q$$

such that the following properties hold true:

- for all $x \in \mathbb{R}^q$ with $|x| \leq n$, we have $\sigma_n(x) = \sigma(x)$ and $b_n(x) = b(x)$
- there exists a constant $c_n \in \mathbb{R}_+$, such that for all $x, y \in \mathbb{R}^q$,
$$|\sigma_n(x) - \sigma(x)| + |b_n(x) - b(x)| \leq c_n |x - y|.$$

These extended maps can be simply obtained by using cutoff and regularization by convolution.
For all $n \geq 1$ and $x \in \mathbb{R}^q$, let $(X_t^n(x))_{t\geq 0}$ be the solution of the SDE (provided by Theorem 8.1.5)

$$X_t^n(x) = x + \int_0^t \sigma_n(X_s^n(x))\mathrm{d}B_s + \int_0^t b_n(X_s^n(x))\mathrm{d}s \quad \text{a.s.,} \quad t \geq 0.$$

For all $m \geq 1$, let us define the stopping time

$$T_n^m(x) = \inf\{t \geq 0 : |X_t^n(x)| \geq m\}.$$

Lemma 8.5.2: Cutoff and stationarity

For all $m \geq n \geq 1$ and $x \in \mathbb{R}^q$, we have

$$T_n^n(x) = T_m^n(x) \leq T_m^m(x) \quad \text{a.s.,}$$

and, for all $t \in [0, T_n^n(x)]$, we have

$$X_t^n(x) = X_t^m(x) \quad \text{a.s.}$$

Proof of Lemma 8.5.2. Let us define $T = T_n^n(x) \wedge T_m^n(x)$. We have, for all $t \geq 0$,

$$X_{t\wedge T}^n = x + \int_0^t \mathbf{1}_{s\leq T}\sigma_n(X_{s\wedge T}^n)\mathrm{d}B_s + \int_0^t \mathbf{1}_{s\leq T} b_n(X_{s\wedge T}^n)\mathrm{d}s$$

and

$$X_{t\wedge T}^m = x + \int_0^t \mathbf{1}_{s\leq T}\sigma_m(X_{s\wedge T}^m)\mathrm{d}B_s + \int_0^t \mathbf{1}_{s\leq T} b_m(X_{s\wedge T}^m)\mathrm{d}s.$$

By definition of T, the processes $(X_{t\wedge T}^n)_{t\geq 0}$ and $(X_{t\wedge T}^m)_{t\geq 0}$ solve the same SDE

$$Z_t = x + \int_0^t \mathbf{1}_{s\leq T}\sigma(Z_s)\mathrm{d}B_s + \int_0^t \mathbf{1}_{s\leq T} b(Z_s)\mathrm{d}s,$$

and thus, using the pathwise uniqueness (Theorem 8.1.2), we have $X_{t\wedge T}^n(x) = X_{t\wedge T}^m(x)$ a.s. for all $t \geq 0$. Moreover, on the event $\{0 < T < \infty\}$, for all $t \in [0, T)$,

$$|X_t^m(x)| = |X_t^n(x)| < n \quad \text{and} \quad [X_T^m(x)| = |X_T^n(x)| = n.$$

It follows that $T = T_n^n(x) = T_m^n(x)$. Furthermore, if $T = 0$, then $|x| \geq n$ and $T = T_n^n(x) = T_m^n(x) = 0$, while if $T = \infty$, then $T_n^n(x) = T_m^n(x) = \infty$. Finally the continuity of $X^m(x)$ gives $T_m^n(x) \leq T_m^m(x)$. ∎

Proof of Theorem 8.5.1. Existence. We set $\xi^x = \sup_{n\geq 0} T_n(x)$ where $T_n(x) = T_n^n(x)$. If $|x| < n$ then $T_n(x) > 0$ and thus $\xi^x > 0$. Let $t \in [0, \xi^x)$. By definition of ξ^x, there exists n such that $T_n(x) > t$ and for all $m \geq n$, we have $X_t^m(x) = X_t^n(x)$ a.s. from Lemma 8.5.2. We can then define

$$X_t(x) = \begin{cases} \lim_{n\to\infty} X_t^n(x) & \text{if } t \in [0, \xi^x) \\ \infty & \text{if } t \in [\xi^x, \infty). \end{cases}$$

This process $(X_t(x))_{t\geq 0}$ verifies the first property stated by the Theorem. Moreover, on $\{\xi^x < \infty\}$, 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 $\{\xi^x < \infty\}$,

$$\overline{\lim}_{t \xrightarrow{\leq} \xi^x} |X_t(x)| = +\infty.$$

Now let us proceed by contradiction and suppose that

$$\mathbb{P}\Big(\underline{\lim}_{t \xrightarrow{\leq} \xi^x} |X_t(x)| < +\infty,\ \xi^x < \infty\Big) > 0.$$

Then we can find real numbers r and R such that $0 < r < R < \infty$ and

$$\mathbb{P}\Big(\underline{\lim}_{t \xrightarrow{\leq} \xi^x} |X_t(x)| < r,\ \overline{\lim}_{t \xrightarrow{\leq} \xi^x} |X_t(x)| > R,\ \xi^x < \infty\Big) > 0. \qquad (\star)$$

Let $f \in \mathscr{C}_c^2(\mathbb{R}^q, \mathbb{R})$ be such that $f(x) = 0$ if $|x| = r$ and $f(x) = 1$ if $|x| = R$. Now, by Theorem 8.4.5,

$$\Big(f(X_t^n(x)) - \int_0^t L^n(f)(X_s^n(x))\mathrm{d}s\Big)_{t\geq 0}$$

is a martingale, where L^n is the infinitesimal generator of X^n. The differential operator L^n coincides with the infinitesimal generator L of $X(x)$ on $\{x \in \mathbb{R}^q : |x| \leq n\}$, and since f is compactly supported, if follows that $L^n f = Lf$ for n sufficiently large, and we can replace L^n by L. Next, it follows by the Doob stopping theorem (Theorem 2.5.1) that for all $m \geq n$, the process

$$\Big(f(X_{t\wedge T_m}^n(x)) - \int_0^{t\wedge T_m} L(f)(X_s^n(x))\mathrm{d}s\Big)_{t\geq 0}$$

is a continuous martingale. By letting $n \to \infty$, by dominated convergence, we get that

$$\Big(f(X_{t\wedge T_m}(x)) - \int_0^{t\wedge T_m} L(f)(X_s(x))\mathrm{d}s\Big)_{t\geq 0}$$

is a continuous martingale. Now by letting $m \to \infty$, we see similarly that

$$\Big(f(X_t(x))\mathbf{1}_{t<\xi^x} - \int_0^{t\wedge \xi^x} L(f)(X_s(x))\mathrm{d}s\Big)_{t\geq 0}$$

is a continuous martingale. Note that if $t < \xi^x$ then $X_{t\wedge T_m(x)}(x) \to X_t(x)$ as $m \to \infty$ while if $t \geq \xi^x$ then $f(X_{t\wedge T_m(x)}(x)) = f(X_{T_m(x)}(x)) \to 0$ as $m \to \infty$. It follows then that the process $((f(X_t(x))(x))\mathbf{1}_{t<\xi^x})_{t\geq 0}$ is a continuous martingale. Now the left continuity of this process at $t = \xi$ contradicts $(\star)$ and the definition of f. It follows that the second property stated in the theorem holds true: a.s., on $\{\xi^x < \infty\}$,

$$\lim_{t \overset{<}{\to} \xi^x} |X_t(x)| = +\infty.$$

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

$$X_{t\wedge T\wedge T_n}^n = x + \int_0^t \mathbf{1}_{s\leq T\wedge T_n}\sigma(X_s^n)\mathrm{d}B_s + \int_0^t \mathbf{1}_{s\leq T\wedge T_n} b(X_s^n)\mathrm{d}s,$$

thus,

$$X_{t\wedge T\wedge T_n} = x + \int_0^t \mathbf{1}_{s\leq T\wedge T_n}\sigma(X_s)\mathrm{d}B_s + \int_0^t \mathbf{1}_{s\leq T\wedge T_n} b(X_s)\mathrm{d}s,$$

and the desired result follows by letting $n \to \infty$ and using dominated convergence.

Uniqueness. Let (X, ξ) and (X', ξ') be two solutions with same initial condition x satisfying to the theorem properties. Then they are both solutions of the SDE with random coefficients given by the third property, with $T = \xi \wedge \xi'$. Now, if we force them to zero after T, then, by proceeding as in the proof of Theorem 8.1.3, we get $X = X'$ on $[0, T]$. Let us prove that $\xi = \xi'$. On $\{\xi < \xi'\}$, we have $T = \xi$, and by definition of ξ, $\lim_{t\nearrow\xi} |X_t| = +\infty$, while by the definition of ξ', $\lim_{t\nearrow\xi} |X'_t| = |X'_\xi|$ since X' is continuous on $[0, \xi')$ and $\xi \in [0, \xi')$. Contradiction. Thus $\{\xi < \xi'\} = \varnothing$, and by symmetry, $\xi = \xi'$. ■

[a]This is the case for instance when σ and b are $\mathscr{C}^1$ on $\mathbb{R}^q$.
[b]Named after Pavel Alexandrov (1896 – 1982), Russian mathematician.

Chapter 9

More links with partial differential equations

9.1 Feynman – Kac formula

Let $\sigma : \mathbb{R}^q \mapsto \mathcal{M}_{q,d}(\mathbb{R})$ and $b : \mathbb{R}^q \mapsto \mathbb{R}^q$ be Lipschitz. For all $x \in \mathbb{R}^q$, let $(X^x_t)_{t\geq 0}$ be the solution of the SDE

$$X^x_t = x + \int_0^t \sigma(X^x_s)\mathrm{d}B_s + \int_0^t b(X^x_s)\mathrm{d}s, \quad t \geq 0.$$

Its infinitesimal generator is the differential operator (linear, second order, without constant term)

$$L = \frac{1}{2}\sum_{i,j=1}^q (\sigma(x)\sigma^\top(x))_{i,j}\partial^2_{i,j} + \sum_{i=1}^q b_i(x)\partial_i.$$

We have seen that L is the infinitesimal generator of a Markov semigroup $(\Pi_t)_{t\geq 0}$ associated to X. But what can be done for more general linear partial differential operators? The simplest generalization beyond L is obtained by adding to L a zero order term, say $U \in \mathscr{C}^2(\mathbb{R}^q, \mathbb{R})$, which gives the linear operator

$$L_U = L + U$$

in the sense that for all $f \in \mathscr{C}^2(\mathbb{R}^q, \mathbb{R})$ and $x \in \mathbb{R}^q$,

$$L_U(f)(x) = \frac{1}{2}\sum_{i,j=1}^q (\sigma(x)\sigma^\top(x))_{i,j}\partial^2_{i,j}f(x) + \sum_{i=1}^q b_i(x)\partial_i f(x) + U(x)f(x).$$

The following theorem states that there exists a semi-group associated to L_U.

Theorem 9.1.1. Feynman[a] – Kac[b] formula and semi-group.

[a]Named after Richard Phillips Feynman (1918 – 1988), American theoretical physicist.
[b]Named after Marc Kac (1914 – 1984), Polish American mathematician.

Let $U \in \mathscr{C}^2(\mathbb{R}^q, (-\infty, c])$ be upper bounded, for some finite constant c. For all $t \geq 0$ we define the operator Q_t acting on bounded and measurable functions $f : \mathbb{R}^q \to \mathbb{R}$ by the Feynman – Kac formula

$$x \in \mathbb{R}^q \mapsto Q_t(f)(x) = \mathbb{E}(f(X_t(x))\mathrm{e}^{\int_0^t U(X_s(x))\mathrm{d}s}).$$

1. The family $(Q_t)_{t\geq 0}$ is a semi-group (known as the Feynman – Kac semi-group).
2. For all $f \in \mathscr{C}^2_b(\mathbb{R}^q, \mathbb{R})$, all $t \geq 0$, and all $x \in \mathbb{R}^q$, there is a Duhamel formula

   $$Q_t(f)(x) = f(x) + \int_0^t Q_s(L_U(f))\mathrm{d}s$$

   which involves the real Schrödinger operator (diffusion operator + multiplicative potential)

   $$L_U(f)(x) = L(f)(x) + U(x)f(x).$$

   In particular we have the Kolmogorov equation

   $$\partial_t Q_t(f)(x) = Q_t(L_U(f))(x).$$

3. The semi-group $(Q_t)_{t\geq 0}$ is continuous in the sense that for all $f \in \mathscr{C}_0(\mathbb{R}^q, \mathbb{R})$,
$$\lim_{t\to 0^+} \|Q_t(f) - f\|_\infty = 0.$$
4. The operator L_U is the infinitesimal generator of the semi-group $(Q_t)_{t\geq 0}$, namely, for all $f \in \mathscr{C}_c^2$,
$$\lim_{t\to 0^+} \left\| \frac{Q_t(f) - f}{t} - L_U f \right\|_\infty = 0.$$

We could also prove a Kolmogorov equation similar to the one of Theorem 8.4.8, namely $\partial_t Q_t = L_U Q_t(f)$, in other words $(\partial_t - L_U)Q_t(f) = 0$, which includes the version at $t = 0$, $\partial_{t=0^+} Q_t(f) = L_U Q_0(f) = L_U f$.

In the formula defining $Q_t(f)(x)$, the exponential weight involves the accumulated value of U along the trajectory $s \in [0,t] \mapsto X_s^x$. When $U \geq 0$ this is an amplification while when $U \leq 0$ this is an attenuation. The quantity $Q_t(\mathbf{1}_A)$ can be interpreted as a quantity of matter or particles in the set A at time t, which is subject to amplification or attenuation according to the Feynman–Kac evolution equation.

We have $\partial_t Q_t = Q_t(L_U)$. We say that L_U is a linear second order differential operator with a multiplicative potential U. This potential models a medium with amplification or attenuation depending on its sign.

The semi-group $(Q_t)_{t\geq 0}$ is positive in the sense that for all $t \geq 0$, $f \geq 0$ implies $Q_t(f) \geq 0$. However, it is not a Markov semi-group in the sense that if we denote by 1 the constant function equal to 1 then $Q_t(1) = \mathbb{E}(\mathrm{e}^{\int_0^t U(X_s)\mathrm{d}s})$ is not equal to 1 in general when $t > 0$.

Proof.

1. For all bounded and measurable $f : \mathbb{R}^q \to \mathbb{R}$, all $x \in \mathbb{R}^q$ and $s, t \geq 0$, the Markov property for X^x gives
$$\begin{aligned}
Q_{t+s}(f)(x) &= \mathbb{E}(\mathrm{e}^{\int_0^t U(X_v^x)\mathrm{d}v}\mathbb{E}(f(X_{t+s}^x)\mathrm{e}^{\int_t^{t+s} U(X_v^x)\mathrm{d}v} \mid \mathscr{F}_t))\\
&= \mathbb{E}(\mathrm{e}^{\int_0^t U(X_v^x)\mathrm{d}v} Q_s(f)(X_t^x))\\
&= Q_t(Q_s(f))(x).
\end{aligned}$$
2. For all $f \in \mathscr{C}^2(\mathbb{R}^q, \mathbb{R})$, all $x \in \mathbb{R}^q$, and all $t \geq 0$, we have, from Theorem 8.4.5,
$$f(X_t^x) = f(x) + M_t^f + \int_0^t (Lf)(X_s^x)\mathrm{d}s.$$
Note that M^f is a martingale and not only a local martingale because f is of class $\mathscr{C}_b^2$ and not only $\mathscr{C}^2$. Now, from Itô formula (Theorem 7.1.1) for $F(y,z) = y\mathrm{e}^z$ and semi-martingale $(Y_t, Z_t) = (f(X_t), \int_0^t U(X_s)\mathrm{d}s)$,
$$\begin{aligned}
F(Y_t, Z_t) &= F(Y_0, Z_0) + \int_0^t \mathrm{e}^{Z_s}\mathrm{d}Y_s + \int_0^t Y_s \mathrm{e}^{Z_s}\mathrm{d}Z_s\\
&= f(x) + \int_0^t \mathrm{e}^{Z_s}(\mathrm{d}M_s^f + (Lf)(X_s^x)\mathrm{d}s) + \int_0^t Y_s \mathrm{e}^{Z_s} U(X_s)\mathrm{d}s.
\end{aligned}$$
Note that we do not have a second order part in the Itô formula since the local martingale contributions for (Y,Z) comes from Y and on the other hand $\partial_y^2 F(y,z) = 0$.
By taking expectations, the martingale term disappears, and we obtain
$$\begin{aligned}
Q_t(f)(x) &= f(x) + \int_0^t \mathbb{E}((Lf)(X_s^x)\mathrm{e}^{Z_s})\mathrm{d}s + \int_0^t \mathbb{E}(Y_s U(X_s)\mathrm{e}^{Z_s})\mathrm{d}s\\
&= f(x) + \int_0^t (Q_s(Lf)(x) + Q_s(fU))\mathrm{d}s\\
&= f(x) + \int_0^t Q_s(L_U(f))\mathrm{d}s.
\end{aligned}$$
We have used the Fubini–Tonelli theorem to swap $\mathbb{E}$ and $\int_0^t \mathrm{d}s$, which is allowed since $L_U(f)(X_s)\mathrm{e}^{\int_0^s U(X_v)\mathrm{d}v}$ is integrable for $\mathbb{P} \otimes \mathbf{1}_{[0,t]}\mathrm{d}s$ since U is upper bounded and $f \in \mathscr{C}_b^2$.

3. Proceeding as in Corollary 8.4.6, we first show that the result holds for $f \in \mathscr{C}_b^2$ via the Duhamel formula for $(Q_t)_{t\geq 0}$ and the boundedness of Lf, and then we generalize to $f \in \mathscr{C}_0$ by density w.r.t. $\|\cdot\|_\infty$.
4. Proceeding as in Corollary 8.4.6, it suffices to use again the Duhamel formula for $(Q_t)_{t\geq 0}$ and the previous item for $L_U f$, which belongs to $\mathscr{C}_0$ since $f \in \mathscr{C}_c^2$ and $U \in \mathscr{C}^2$.

■

Remark 9.1.2. Killing.

We take Theorem 9.1.1 settings with $U \in \mathscr{C}^2(\mathbb{R}^q, (-\infty, 0])$ bounded. We think about X_t as the position of a particle in a medium at time t, $-U(x)$ as a quantity of poison at position x, and $-\int_0^t U(X_s)\mathrm{d}s$ as the total quantity of poison accumulated by X along its trajectory on the time interval $[0, t]$.
Let $E \sim \mathrm{Exp}(1)$ be an exponential random variable of unit mean, independent of X. Let us define

$$T = \inf\Big\{t \geq 0 : \int_0^t -U(X_s)\mathrm{d}s = E\Big\}.$$

We extend $\mathbb{R}^q$ by a *cemetery state* $\infty \notin \mathbb{R}^q$, and we define the process $\widetilde{X}$ with state space $\mathbb{R}^q \cup \{\infty\}$ by

$$\widetilde{X}_t = \begin{cases} X_t & \text{if } t < T, \\ \infty & \text{if } t \geq T. \end{cases}$$

In other words, starting from $x \in \mathbb{R}^q$, the process $\widetilde{X}$ evolves like X in $\mathbb{R}^q$ and after an exponential time it jumps to ∞ and stays there forever, while if it starts from ∞ then it never moves. It can be checked that $\widetilde{X}$ is Markov, with infinitesimal generator $\widetilde{L}$ given for $f : \mathbb{R}^q \cup \{\infty\} \to \mathbb{R}$ which is $\mathscr{C}^2$ on $\mathbb{R}^q$ by

$$\widetilde{L}(f)(x) = \begin{cases} 0 & \text{if } x = \infty \\ L(f)(x) - U(x)(f(\infty) - f(x)) & \text{if } x \in \mathbb{R}^q. \end{cases}$$

In particular, if $f(\infty) = 0$ then we recover the Feynman – Kac operator

$$\widetilde{L}(f) = L_U(f),$$

and from this point of view, when U is negative, the Feynman – Kac operator and semi-group can be interpreted as the description of the killed Markov diffusion process $\widetilde{X}$. Let us check that indeed, $\widetilde{L}$ is the generator of $\widetilde{X}$. If we define $\lambda_X(t) = -\int_0^t U(X_s)\mathrm{d}s$, then we have $T = \lambda_X^{-1}(E)$, and

$$\{t < T\} = \{t < \lambda_X^{-1}(E)\} = \{\lambda_X(t) < E\}.$$

In particular, using the boundedness of U we get (beware that $t \to 0^+$ and not $t \to +\infty$)

$$\mathbb{P}(t < T \mid \mathscr{F}_t) = \mathbb{P}(\lambda_X(t) < E \mid \mathscr{F}_t) = \mathrm{e}^{\int_0^t U(X_s)\mathrm{d}s} = 1 + o_{t\to 0^+}(1)$$

and

$$\mathbb{P}(t \geq T \mid \mathscr{F}_t) = 1 - \mathrm{e}^{\int_0^t U(X_s)\mathrm{d}s} = -tU(x) + o_{t\to 0^+}(1)$$

where the $o(1)$ are uniform in ω. Now we can write

$$\begin{aligned}
\mathbb{E}(f(\widetilde{X}_t)) - f(x) &= \mathbb{E}(\mathbb{E}(f(\widetilde{X}_t) - f(x) \mid \mathscr{F}_t)) \\
&= \mathbb{E}((f(X_t) - f(x))\mathbb{P}(t < T \mid \mathscr{F}_t)) + (f(\infty) - f(x))\mathbb{E}(\mathbb{P}(t \geq T \mid \mathscr{F}_t)) \\
&= \mathbb{E}((f(X_t) - f(x))(1 + o(1))) + (f(\infty) - f(x))(-tU(x) + o(1)) \\
&= \mathbb{E}((f(X_t) - f(x))(1 + o(1))) + (f(\infty) - f(x))(-tU(x) + o(1)) \\
&= L(f)(x) + o(1) - U(x)(f(\infty) - f(x)) + o(1) \\
&= \widetilde{L}(f)(x) + o(1).
\end{aligned}$$

This suggests to interpret L_U as the generator of a Markov process with killing. When U is constant and equal to $-\lambda$ with $\lambda > 0$ then $T \sim \mathrm{Exp}(\lambda)$ is independent of X and, as soon as $f(\infty) = 0$,

$$Q_t(f)(x) = \mathbb{E}(f(X_t))\mathrm{e}^{-\lambda t} = \mathbb{E}(f(X_t))\mathbb{P}(t < T) = \mathbb{E}(f(X_t)\mathbf{1}_{t<T}) = \mathbb{E}(f(\widetilde{X}_t)).$$

9.2 Kakutani probabilistic formulation of Dirichlet problems

Let us give first an informal presentation of the Dirichlet problem in the case of the Laplacian, and its probabilistic representation using Brownian motion. Let D be an open subset of $\mathbb{R}^d$, with boundary $\partial D = \overline{D} \setminus D$. The Dirichlet problem consists, for some prescribed $g : \partial D \to \mathbb{R}$, to find $u : \overline{D} \to \mathbb{R}$ such that u is harmonic on D in the sense that $\Delta u = 0$ in D, while $u = g$ on ∂D. Following Hilbert, the Dirichlet problem can be solved by using functional spaces and a variational formulation (it can also be solved using pure probabilistic arguments). If we assume that the solution u exists, it is also possible, following Kakutani, to obtain a probabilistic representation for u. Namely, let B be a d-dimensional Brownian motion issued from the origin. Assume that for all $x \in D$, $T_{\partial D} = \inf\{t \geq 0 : x + B_t \in \partial D\}$ satisfied $\mathbb{P}(T_{\partial D}) < \infty$. The Itô formula gives

$$u(x + B_{t\wedge T_{\partial D}}) = u(x) + M^u_{t\wedge T_{\partial D}} + \int_0^{t\wedge T_{\partial D}} \underbrace{\Delta u(x + B_s)}_{=0, s<T_{\partial D}} \mathrm{d}s,$$

for all $x \in D$ and all $t \geq 0$. The integral in the right-hand side is zero since $\Delta u = 0$ on D and $x + B_s \in D$ if $s < T_{\partial D}$. Moreover if g is bounded then $M^u_{\bullet\wedge T_{\partial D}}$ is a martingale issued from 0, and taking expectations gives

$$\mathbb{E}(u(x + B_{t\wedge T_{\partial D}})) = u(x).$$

Now, if u is continuous and bounded on $\overline{D}$, we obtain, by dominated convergence and since $u = g$ on ∂D,

$$\mathbb{E}(g(x + B_T)) = u(x).$$

This is the Kakutani probabilistic representation of the solution of the Dirichlet problem. It shows by the way the uniqueness of the solution. It allows Monte–Carlo methods. Note that when $d = 1$ and $D = (a, b)$, then $u(x) = \alpha x + \beta$ for all $x \in [a, b]$ with $\alpha = (g(b) - g(a))/(b - a)$, and $\beta = (bg(a) - ag(b))/(b - a)$.

We can define more generally the Dirichlet problem for a linear partial differential operator of second order, possibly with constant term, and with a prescribed arbitrary value inside the domain. To study this generalization, we take the following ingredients.

- $D \subset \mathbb{R}^q$ is open and bounded. Its boundary is $\partial D = \overline{D} \setminus D$. We have $D \cap \partial D = \varnothing$ and $\overline{D} = D \cup \partial D$.
- $x \in \mathbb{R}^q \mapsto a(x) = (a_{i,j}(x))_{1\leq i,j\leq d} = \sigma(x)\sigma^\top(x)$ where $\sigma(x)$ is a $d \times q$ matrix, Lipschitz in x.
- $x \in \mathbb{R}^q \mapsto b(x) = (b_i(x))_{1\leq i\leq d}$ is a vector field, Lipschitz in x
- $f : D \mapsto \mathbb{R}$ is continuous and bounded.
- $g : \partial D \mapsto \mathbb{R}$ is continuous and bounded.
- $c : D \mapsto (-\infty, 0]$ is continuous and negative.
- differential operator $\mathrm{L} = \frac{1}{2}\sum_{i,j=1}^q a_{i,j}(x)\partial^2_{x_i,x_j} + \sum_{i=1}^q b_i(x)\partial_{x_i}$.

The Dirichlet problem[1] on D consists in seeking for $u \in \mathscr{C}^2(D, \mathbb{R}) \cap \mathscr{C}^0(\overline{D}, \mathbb{R})$ such that

$$\begin{cases} \underbrace{\mathrm{L}u(x) + c(x)u(x)}_{L_c(u)(x)} = f(x) & \text{for all } x \in D, \\ \lim\limits_{\substack{x\to x_0 \\ x\in D}} u(x) = g(x_0) & \text{for all } x_0 \in \partial D, \end{cases} \tag{DirP}$$

When $c = 0$ and $f = 0$ then $Lu = 0$ on D and we say that f is harmonic for L on D. Let us consider the solution $(X^x_t)_{t\geq 0}$ provided by Theorem 8.1.5 of the stochastic differential equation

$$X^x_t = x + \int_0^t \sigma(X^x_s)\mathrm{d}B_s + \int_0^t b(X^x_s)\mathrm{d}s, \quad t \geq 0,\ x \in \mathbb{R}^q, \tag{SDE}$$

[1] Named after Peter Gustav Lejeune Dirichlet (1805 – 1859), German mathematician.

where $B=(B_s)_{s\geq 0}$ is a d-dimensional BM. For all $x\in D$, we define the stopping time

$$T_{\partial D}=\inf\{t\geq 0: X_t^x\in\partial D\}\in[0,+\infty].$$

In the sequel, we abridge X^x into X and we denote by $\mathbb{P}_x$ and $\mathbb{E}_x$ to indicate the starting point of X.

The Dirichlet problem by itself is a static problem (does not involve time), but the Kakutani probabilistic representation of its solution is dynamic, it involves a stochastic process, an evolution equation, time.

Theorem 9.2.1. Kakutani[a] probabilistic representation of the solution of the Dirichlet problem.

[a]Named after Shizuo Kakutani (1911–2004), Japanese mathematician.

Suppose that there exists a solution u to (DirP).

- If $f=c=0$, then, for all $x\in D$, if $\mathbb{P}_x(T_{\partial D}<\infty)=1$, then

$$u(x)=\mathbb{E}_x(g(X_{T_{\partial D}})).$$

- More generally, for all $x\in D$, if $\mathbb{E}_x(T_{\partial D})<\infty$ then

$$u(x)=\mathbb{E}_x\Big(g(X_{T_{\partial D}})\mathrm{e}^{\int_0^{T_{\partial D}}c(X_s)\mathrm{d}s}-\int_0^{T_{\partial D}}f(X_s)\mathrm{e}^{\int_0^s c(X_v)\mathrm{d}v}\mathrm{d}s\Big).$$

  Moreover if $f=0$ then $\mathbb{E}_x(T_{\partial D})<\infty$ can be replaced by the weaker condition $\mathbb{P}_x(T_{\partial D}<\infty)=1$.

In particular the solution in unique.

Remark 9.2.2. Existence of solution with probabilistic arguments.

Following [31, Section 7.2], it can be shown, say in the simple case $c=f=0$ and $L=\frac{1}{2}\Delta$, that we have $\mathbb{E}_x(T_{\partial D})<\infty$, and that if $g\in\mathscr{C}^0(\partial D,\mathbb{R})$ then $x\in\overline{D}\mapsto u(x)=\mathbb{E}_x(g(X_{T_{\partial D}}))$ is $\mathscr{C}^2(D,\mathbb{R})$ and harmonic on D in the sense that $\Delta u=0$ on D. This can be done by showing that u has the mean value property and this can be proved by using the strong Markov property for X. We have then also $u=g$ on ∂D. However there are extra regularity assumptions on D to get $u\in\mathscr{C}^0(\overline{D},\mathbb{R})$.

Proof. We have already produced a proof for $q=d$, $\sigma=I_d$, $b=0$, $f=0$, $c=0$. We now follow the same scheme in the general case. We suppose first that u can be extended to $\mathbb{R}^q$ as a $\mathscr{C}_b^2$ function. The Itô formula (Theorem 7.1.1) for $(x,y)\mapsto u(x)\mathrm{e}^y$ and $(X_t,Y_t=\int_0^t c(X_s)\mathrm{d}s)$ gives, for all $t\geq 0$,

$$u(X_t)\mathrm{e}^{Y_t}-u(x)=\int_0^t\langle\nabla u(X_s)\mathrm{e}^{Y_s},\sigma(X_s)\mathrm{d}B_s\rangle+\int_0^t f(X_s)\mathrm{e}^{Y_s}\mathrm{d}s+\int_0^t(\mathrm{L}u+cu-f)(X_s)\mathrm{e}^{Y_s}\mathrm{d}s.$$

If we replace t by $t\wedge T_{\partial D}$, the third and last integral in the right-hand side vanishes due to the fact that $X_s\in D$ when $s<T$ and u solves (DirP) namely $Lu+cu-f=0$ on D. Next, since u is bounded on D, the first integral in the right hand side above is a continuous martingale issued from the origin, and thus, by the Doob stopping theorem (Theorem 2.5.1), its expectation is zero. This gives finally that

$$u(x)=\mathbb{E}_x\Big(u(X_{t\wedge T_{\partial D}})\mathrm{e}^{Y_{t\wedge T_{\partial D}}}\Big)-\mathbb{E}_x\Big(\int_0^{t\wedge T_{\partial D}}f(X_s)\mathrm{e}^{Y_s}\mathrm{d}s\Big).$$

Since $\mathbb{P}_x(T_{\partial D}<\infty)=1$, the desired formula follows then by letting $t\to\infty$ and using dominated convergence. The condition $\mathbb{E}_x(T_{\partial D})<\infty$ is used to handle the integral involving f via the uniform bound $|\int_0^{T_{\partial D}}f(X_s)\mathrm{e}^{Y_s}\mathrm{d}s|\leq\|f\|_\infty T_{\partial D}$. When $f=0$, we could replace $\mathbb{E}_x(T_{\partial D})<\infty$ by $\mathbb{P}_x(T_{\partial D}<\infty)=1$.

Actually what we did is similar to the proof of the Duhamel formula for the Feynman–Kac operator L_c in the proof of Theorem 9.1.1 except that before taking expectation we stopped the process and we used the definition of the stopping time to in order to use the fact that u is a solution of the Dirichlet problem.

The general case on u can be addressed as follows. We have to take into account the fact that ∇u and Δu may blow up when approaching ∂u, because we only know that $u\in\mathscr{C}^2(D)\cap\mathscr{C}^0(\overline{D})$. Thus we restrict the domain D to gain regularity. For all $n \geq 1$, let $D_n = \{x \in D : \mathrm{dist}(x, \partial D) > 1/n\}$. For all $x \in D$, we have $x \in D_n$ for n large enough. For all $n \geq 1$, $\overline{D_n} \subset D_{n+1} \subset D$, and $\cup_n D_n = D$. For all $n \geq 1$, let $u_n \in \mathscr{C}^2_b(\mathbb{R}^q, \mathbb{R})$ be such that $u_n = u$ on D_n. This can be constructed by proceeding as in Lemma 9.2.3 below. The function u_n solves the Dirichlet problem on D_n with boundary condition $u_n|_{\partial D_n}$. Then, for all $x \in D_n$,

$$u_n(x) = u(x) = \mathbb{E}_x\Big(u(X_{T_{\partial D_n}})\mathrm{e}^{Y_{T_{\partial D_n}}}\Big) - \mathbb{E}_x\Big(\int_0^{T_{\partial D_n}} f(X_s)\mathrm{e}^{Y_s}\mathrm{d}s\Big).$$

We have $T_{\partial D_n} \leq T_{\partial D_{n+1}} \nearrow T$ as $n \to \infty$, and it remains to let $n \to \infty$ and to use dominated convergence. ■

Lemma 9.2.3. Itô formula on a domain.

The Itô formula of Theorem 7.1.1 remains valid for all $f \in \mathscr{C}^2(D, \mathbb{R})$ where $D \subset \mathbb{R}^d$ is open, provided that $X_0 = x \in D$, and, almost surely, for all $t \geq 0$, $X_t \in D$.

Proof. Since f may blow up at the boundary of D, the idea is first to reduce the domain and then to extend the restricted function, which is bounded, to the whole space. Namely, for all $n \geq 1$, let

$$D_n = \{x \in D : \mathrm{dist}(x, \partial D) > 1/n\}.$$

We have $\cup_n D_n = D$, and for all $n \geq 1$, $D_n \subset \overline{D_n} \subset D_{n+1}$. Let $\varphi_n \in \mathscr{C}^\infty(\mathbb{R}^d, [0,1])$ such that $\varphi_n = 1$ on D_n and $\varphi_n = 0$ on $(D_{n+1})^c$. Let us define $\widetilde{f}$ by $\widetilde{f} = f$ on D and $\widetilde{f} = 0$ on D^c. Let us define

$$f_n = \widetilde{f}\varphi_n + (1 - \varphi_n)(\widetilde{f} * \varphi_n).$$

Now $f \in \mathscr{C}^0(\overline{D_{n+1}}, \mathbb{R})$ gives $f_n \in \mathscr{C}^\infty_b(\mathbb{R}^d, \mathbb{R})$, and $f_n = f$ on D_n. The Itô formula (Theorem 7.1.1) for f_n gives the canonical decomposition of the semi-martingale $(f_n(X_{t \wedge T_{\partial D_n}}))_{t \geq 0}$, which depends only on f. Finally, since $f_n \to f$ as $n \to \infty$ on D, it remains to use dominated convergence as $n \to 0$ (Theorem 6.3.4). ■

Remark 9.2.4. Brownian motion.

Let us consider the simple situation where $q = d$, $\sigma = I_d$, $b = 0$, $f = 0$, $c = 0$, for which $X = x + B$. The fact that $\mathbb{P}(T_{\partial D} < \infty) = 1$ for all $x \in D$ follows from the fact that the coordinates of B are one-dimensional BM and that with probability one they exit any given finite interval, see example 2.5.2. Let us show that for all $x \in D$, $\mathbb{E}_x(T_{\partial D}) < \infty$. Since B is continuous, it suffices to show that for all $R > 0$,

$$\mathbb{E}(\tau) < \infty \quad \text{where} \quad \tau = \inf\{t \geq 0 : |B_t| = R\} = \inf\{t \geq 0 : |B_t|^2 = R^2\}$$

(recall that $B_0 = 0$). The process $|B|^2$ is a squared Bessel process issued from the origin and $(|B_t|^2 - td)_{t \geq 0}$ is a martingale. The Doob stopping theorem (Theorem 2.5.1) gives, for all $t \geq 0$,

$$\mathbb{E}(|B_{t \wedge \tau}|^2) = d\mathbb{E}(t \wedge \tau).$$

By definition of τ, the left-hand side is bounded by R^2, and by monotone convergence for the right hand side, we get $\mathbb{E}(\tau) \leq R^2/d$, which implies $\mathbb{P}(\tau < \infty) = 1$, which gives by dominated convergence

$$\mathbb{E}(\tau) = \frac{R^2}{d} < \infty.$$

We may also compute the exponential moments of τ with an exponential martingale.

Remark 9.2.5. Assumptions.

For a complete probabilistic analysis of Dirichlet type problems, we refer to [40].

1. If c takes its values in $(-\infty, -a]$, $a > 0$, then the assumption $\mathbb{E}_x(T_{\partial D}) < \infty$ is useless, and we get

from the proof by letting $t \to \infty$ in the display with $t \wedge T_{\partial D}$ that

$$u(x) = \mathbb{E}_x\Big(u(X_{T_{\partial D}})\mathrm{e}^{\int_0^{T_{\partial D}} c(X_s)\mathrm{d}s}\mathbf{1}_{T_{\partial D}<\infty}\Big) - \mathbb{E}_x\Big(\int_0^{T_{\partial D}} f(X_s)\mathrm{e}^{Y_s}\mathrm{d}s\Big).$$

2. If for some $a > 0$ and all $x \in D$, $\mathbb{E}_x(\mathrm{e}^{aT_{\partial D}}) < \infty$, then Theorem 9.2.1 remains valid for c taking values in $(-\infty, a]$ with $a > 0$. This is the case for instance if for all $x \in D$, $T_{\partial D}$ is bounded.

3. In the proof of Theorem 9.2.1, we can simply assume that (SDE) admits a (weak) solution for all $x \in D$, the coefficients σ and b being supposed mesurable (not necessarily Lipschitz!). The hypothesis of continuity for f, c, and g are superfluous as well, and one can assume that they are just mesurable, f and g being bounded and non-negative. This remark is useful for certain problems in stochastic control theory.

4. The probabilistic representation provided by Theorem 9.2.1 shows that it suffices to give g on a subset $\partial_R D$ of ∂D such that $\mathbb{P}_x(X_{T_{\partial D}} \in \partial_R D) = 1$ for all $x \in D$. The elements of such as subset are called regular points of the boundary. This same representation shows also that the Dirichlet problem is ill posed if g is arbitrary outside $\partial_R D$.

5. The Itô formula of Theorem 7.1.1 can be generalized to functions which are not $\mathscr{C}^2$ but are differentiable in a weak sense, and it follows that the probabilistic representation provided by Theorem 9.2.1 remains valid in this general case.

Remark 9.2.6. Harmonic measure and Poisson kernel.

The exit law $\Pi(x, \mathrm{d}y) = \mathbb{P}_x(X_{T_{\partial D}} \in \mathrm{d}y)$ is the harmonic measure. When the ingredients $\partial D, L$ are regular enough, then $\sup_{x\in D}\mathbb{E}_x(T_{\partial D}) < \infty$, and (DirP) has a solution u (say $f = 0$, $c = 0$), the assumptions of Theorem 9.2.1 are satisfied, and the harmonic measure is absolutely continuous. Its density $\Pi(x, \mathrm{d}y) = \Pi(x, y)\mathrm{d}y$, $x \in D$, $y \in \partial D$, is the Poisson kernel:

$$u(x) = \mathbb{E}_x(g(X_{T_{\partial D}})) = \int_{\partial D} g(y)\Pi(x, \mathrm{d}y).$$

The function $(x, y) \mapsto \Pi(x, y)$ is strictly positive, $\mathscr{C}^2$ in $x \in D$, and

$$\begin{cases} \lim\limits_{\substack{x\to y_0\\ x\in D}} \Pi(x, y) = 0 & \text{for all } y, y_0 \in \partial D, y \neq y_0,\\ \lim\limits_{\substack{x\to y_0\\ x\in D}} |\Pi(x, y_0)| = +\infty & \text{for all } y_0 \in \partial D.\end{cases}$$

We refer to [37] and [40] for more details on these aspects.

Remark 9.2.7. Discrete Dirichlet problem.

Let $(X_n)_{n\in\{0,1,2,\ldots\}}$ be a simple random walk on $\mathbb{Z}^d$ defined by $X_0 \in \mathbb{Z}^d$ and

$$X_{n+1} = X_n + \varepsilon_{n+1} = X_0 + \varepsilon_1 + \cdots + \varepsilon_{n+1}$$

where $(\varepsilon_n)_{n\geq 0}$ are independent and identically distributed "increments" uniformly distributed on

$$\{\pm e_1, \ldots, \pm e_d\}$$

where $e_1, \ldots, e_d$ is the canonical basis of $\mathbb{R}^d$. This is the discrete time and space analogue of Brownian motion. Its physics is essentially the same, since it is a matter of space-time scale. Mathematically, in discrete settings, we do not have the difficulties of regularity and infinite dimensional analysis, however we do not have the advantages of continuity and the chain rule for differentials. The simulation of Brownian motion on a computer always reduces to some sort of discrete time and space random walk, and the link between the two is the subject of stochastic numerical analysis in relation with the central limit phenomenon. The Feynman – Kac formula and the Dirichlet problem have fully discrete analogues. For instance, for the Dirichlet problem, if $D \subset \mathbb{Z}^d$, we could define

$$\partial D = \{x \not\in D : |x-y| = 1\} \quad \text{and} \quad \overline{D} = D \cup \partial D$$

where $|x-y| = 1$ is equivalent to say that $x - y \in \{\pm e_1, \dots, \pm e_d\}$, and

$$T = \inf\{n \geq 0 : X_n \in \partial D\}.$$

Now if D is not empty and bounded, then, for all $g : \partial D \to \mathbb{R}$, there exists a unique

$$u : \overline{D} \to \mathbb{R}$$

such that $u = g$ on ∂D and $\Delta u = 0$ on D where for all $x \in \mathbb{Z}^d$,

$$\Delta u(x) = \frac{1}{2d} \sum_{y:|x-y|=1} u(y) - u(x).$$

This discrete Laplacian leads to the continuous Laplacian via a second order Taylor formula and the approximation $\varepsilon\mathbb{Z}^d \to \mathbb{R}^d$ as $\varepsilon \to 0$. The discrete Kakutani probabilistic representation simply reads

$$u(x) = \mathbb{E}_x(g(X_{T_{\partial D}})).$$

Remark 9.2.8. Dynkin formula.

For all $x \in \mathbb{R}^q$, all stopping time T with $\mathbb{E}_x(T) < \infty$, and all $f \in \mathscr{C}^2(\mathbb{R}^q, \mathbb{R})$ with compact support,

$$\mathbb{E}_x(f(X_T)) = f(x) + \mathbb{E}_x \int_0^T Lf(X_s)\mathrm{d}s.$$

If T is the exist time of a bounded set then we can drop the restriction of compactness of support.

Coding in action 9.2.9. Simulation.

Write a code for the simulation of the law of $T_{\partial D}$ and $X_{T_{\partial D}}$ for various choices of D, d, when X is BM. Same question when X is an Ornstein – Uhlenbeck process. Hint: start with the discrete Dirichlet problem, and then think about the discretization of the continuous case.

Bibliography

[1] G. W. ANDERSON, A. GUIONNET & O. ZEITOUNI – *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, CUP, 2010.

[2] D. BAKRY, I. GENTIL & M. LEDOUX – *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, 2014. 124

[3] P. BALDI – *Stochastic calculus*, Universitext, Springer, Cham, 2017, An introduction through theory and exercises. iii

[4] F. BAUDOIN – *Diffusion processes and stochastic calculus*, EMS Textbooks in Mathematics, EMS, 2014. iii, 14, 18, 91, 92, 123

[5] R. M. BLUMENTHAL & R. K. GETOOR – *Markov processes and potential theory*, Pure and Applied Mathematics, Vol. 29, Academic Press, New York-London, 1968.

[6] A. N. BORODIN & P. SALMINEN – *Handbook of Brownian motion—facts and formulae*, second ed., Probability and its Applications, Birkhäuser Verlag, Basel, 2002. 51

[7] B. BRU & M. YOR – "Comments on the life and mathematical legacy of W. Doeblin", *Finance Stoch.* **6** (2002), no. 1, p. 3–47. vi, 87

[8] F. COMETS & T. MEYRE – *Calcul stochastique et modèles de diffusions : Cours et exercices corrigés*, Dunod, 2006, Coll. Sciences sup. iii

[9] C. DELLACHERIE & P.-A. MEYER – *Probabilities and potential*, North-Holland Mathematics Studies, vol. 29, North-Holland, 1978. iii, 18, 20

[10] — , *Probabilities and potential. B*, North-Holland Mathematics Studies, vol. 72, North-Holland Publishing, 1982, Theory of martingales, Translated from the French by J. P. Wilson. iii

[11] — , *Probabilities and potential. C*, North-Holland Mathematics Studies, vol. 151, North-Holland, 1988, Potential theory for discrete and continuous semigroups, Translated from the French by J. Norris.

[12] B. DUPLANTIER – "Brownian motion, "diverse and undulating"", in *Einstein, 1905–2005*, Prog. Math. Phys., vol. 47, Birkhäuser, 2006, Translated from the French by Emily Parks, p. 201–293. vi, 43, 118

[13] R. DURRETT – *Stochastic calculus*, Probability and Stochastics Series, CRC Press, 1996, A practical introduction. iii

[14] R. DURRETT – *Probability—theory and examples*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 49, Cambridge University Press, Cambridge, 2019, Fifth edition of [MR1068527]. 3

[15] S. N. ETHIER & T. G. KURTZ – *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986, Characterization and convergence. 124

[16] L. C. EVANS – *An introduction to stochastic differential equations*, American Mathematical Society, Providence, RI, 2013. iii

[17] R. F. FOX – "Stochastic calculus in physics", *J. Statist. Phys.* **46** (1987), no. 5-6, p. 1145–1157. 118

[18] I. I. GIKHMAN & A. V. SKOROKHOD – *The theory of stochastic processes. III*, Springer, 1979, Translated from Russian by Samuel Kotz, With corrections to Volumes I and II, Grundlehren der Mathematischen Wissenschaften, 232. iii, 128

[19] A. GUT – *Probability: a graduate course*, second ed., Springer Texts in Statistics, Springer, New York, 2013. 3

[20] N. IKEDA & S. WATANABE – *Stochastic differential equations and diffusion processes*, second ed., North-Holland Mathematical Library, vol. 24, North-Holland, 1989. iii

[21] J. JACOD – *Calcul stochastique et problèmes de martingales*, Lecture Notes in Mathematics, vol. 714, Springer, Berlin, 1979. iii, 124

[22] J. JACOD & A. N. SHIRYAEV – *Limit theorems for stochastic processes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 1987. 99

[23] J.-P. KAHANE – "Le mouvement brownien: un essai sur les origines de la théorie mathématique", in *Matériaux pour l'histoire des mathématiques au XX^e siècle (Nice, 1996)*, Sémin. Congr., vol. 3, Soc. Math. France, Paris, 1998, p. 123–155. vi

[24] I. KARATZAS & S. E. SHREVE – *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer, 1991. iii, 124

[25] — , *Methods of mathematical finance*, Applications of Mathematics (New York), vol. 39, Springer-Verlag, New York, 1998. 119

[26] P. E. KLOEDEN & E. PLATEN – *Numerical solution of stochastic differential equations*, Applications of Mathematics (New York), vol. 23, Springer-Verlag, Berlin, 1992. 109

[27] P. E. KLOEDEN, E. PLATEN & H. SCHURZ – *Numerical solution of SDE through computer experiments*, Universitext, Springer-Verlag, Berlin, 1994, With 1 IBM-PC floppy disk (3.5 inch; HD). 109

[28] H.-H. KUO – *Introduction to stochastic integration*, Universitext, Springer, New York, 2006. iii, 34

[29] Y. A. KUTOYANTS – *Statistical inference for ergodic diffusion processes*, Springer Series in Statistics, Springer-Verlag London, Ltd., London, 2004. 99

[30] D. LAMBERTON & B. LAPEYRE – *Introduction to stochastic calculus applied to finance*, second ed., Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, 2008. 119

[31] J.-F. LE GALL – *Brownian motion, martingales, and stochastic calculus*, french ed., Graduate Texts in Mathematics, vol. 274, Springer, 2016. iii, 25, 65, 67, 91, 106, 117, 137

[32] T. LELIÈVRE, M. ROUSSET & G. STOLTZ – *Free energy computations*, Imperial College Press, London, 2010, A mathematical perspective. 118

[33] R. S. LIPTSER & A. N. SHIRYAEV – *Statistics of random processes. I*, expanded ed., Applications of Mathematics (New York), vol. 5, Springer-Verlag, Berlin, 2001, General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability. 99

[34] — , *Statistics of random processes. II*, expanded ed., Applications of Mathematics (New York), vol. 6, Springer-Verlag, Berlin, 2001, Applications, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability. 99

[35] H. P. MCKEAN – *Stochastic integrals*, AMS Chelsea Publishing, Providence, RI, 2005, Reprint of the 1969 edition, with errata.

[36] M. MÉTIVIER – *Semimartingales*, de Gruyter Studies in Mathematics, vol. 2, Walter de Gruyter & Co., Berlin-New York, 1982, A course on stochastic processes. iii

[37] C. MIRANDA – *Partial differential equations of elliptic type*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 2, Springer, 1970, Second revised edition. Translated from the Italian by Zane C. Motteler. 139

[38] E. NELSON – *Dynamical theories of Brownian motion*, Princeton University Press, 1967. vi

[39] J. PERRIN – *Les atomes (1913)*, CNRS Éditions, 2014. vi, 30

[40] S. C. PORT & C. J. STONE – *Brownian motion and classical potential theory*, Academic Press, 1978, Probability and Mathematical Statistics. 138, 139

[41] N. PRIVAULT – *Stochastic finance*, Chapman & Hall/CRC Financial Mathematics Series, CRC Press, Boca Raton, FL, 2014, An introduction with market examples. 92

[42] P. E. PROTTER – *Stochastic integration and differential equations*, Stochastic Modelling and Applied Probability, vol. 21, Springer, 2005, Second edition. Version 2.1, Corrected third printing. iii

[43] D. REVUZ & M. YOR – *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer, 1999. iii, 117

[44] L. C. G. ROGERS & D. WILLIAMS – *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Foundations, Reprint of the second (1994) edition. iii

[45] — , *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Itô calculus, Reprint of the second (1994) edition. iii

[46] G. ROYER – *An initiation to logarithmic Sobolev inequalities*, SMF/AMS Texts and Monographs, vol. 14, AMS-SMF, 2007, Translated from the 1999 French original by Donald Babbitt. 118

[47] D. W. STROOCK & S. R. S. VARADHAN – *Multidimensional diffusion processes*, Classics in Mathematics, Springer, 2006, Reprint of the 1997 edition. iii, 124

[48] M. S. TAQQU – "Bachelier and his times: a conversation with Bernard Bru", *Finance Stoch.* **5** (2001), no. 1, p. 3–32. vi, 43

[49] B. ØKSENDAL – *Stochastic differential equations*, sixth ed., Universitext, Springer, 2003, An introduction with applications. iii