Stochastic analysis

Paul Bourgade

These are lecture notes from the lessons given in the fall 2010 at Harvard University, and fall 2016 at New York University's Courant Institute.
These notes are based on distinct references. In particular, Chapter 3 is adapted from the remarkable lecture notes by Jean François Le Gall [12], in French. For Chapters 2, 4 and 5, our main references are [13], [16] and [18]. Chapter 6 is based on [23]. For Chapter 7, references are [1] and [10]. All errors are mine.

These are draft notes, so please send me any mathematical inaccuracies you will certainly find. I thank Stéphane Chrétien and Guangqu Zheng for the numerous typos they found in earlier versions of these notes.

Table of contents

Elementary probability prerequisites

These pages remind some important results of elementary probability theory that we will make use of in the stochastic analysis lectures. All the notions and results hereafter are explained in full details in *Probability Essentials*, by Jacod-Protter, for example.

Probability space

Sample space Ω	Arbitrary non-empty set.
σ-algebra $\mathcal{F}$	A set of subsets of Ω, including the empty set, stable under complements and countable union (hence countable intersection).
Probability measure $\mathbb{P}$	A function from $\mathcal{F}$ to $[0,1]$ such that $\mathbb{P}(\Omega)=1$ and $\mathbb{P}(\cup_i \mathrm{A}_i)=\sum_i \mathbb{P}(\mathrm{A}_i)$ for any disjoint elements in $\mathcal{F}$.
Probability space $(\Omega,\mathcal{F},\mathbb{P})$	A triple composed on a set Ω, a σ-algebra $\mathcal{F}\subset 2^\Omega$, and a probability measure on $\mathcal{F}$.
Random variable X	Given a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a metric space $(\mathrm{G},\mathcal{G})$, $\mathrm{X}:\Omega\to\mathrm{G}$ is measurable in the sense $\mathrm{X}^{-1}(g)\in\mathcal{F}$ for any $g\in\mathcal{G}$.
Wiener space	In these lectures, Ω can be the set W of continuous functions from $[0,1]$ to $\mathbb{R}$ (Wiener space) vanishing at 0, $\mathcal{F}=\sigma(\mathrm{W}_s, 0\leqslant s\leqslant 1)$ is the smallest σ-algebra for which all coordinates mappings $\omega\to\mathrm{W}_t(\omega)=\omega(t)$ are measurable, and $\mathrm{G}=\mathbb{R}^d$ endowed with its Borel σ-algebra.
Monotone class theorem	For $\mathrm{C}\subset 2^\Omega$, let $\sigma(\mathrm{C})$ be the smallest σ-algebra containing C (uniquely defined as an intersection of σ-algebras is a σ-algebra). Let P and Q be two probability measures on $\sigma(\mathrm{C})$. If C is stable by intersection and P, Q coincide on C, then P = Q.

Conditional expectation

Definition	On a probability space $(\Omega,\mathcal{F},\mathbb{P})$, given an integrable random variable $\mathrm{X}:\Omega\to\mathrm{G}$ and a sub-σ-algebra $\mathcal{G}\subset\mathcal{F}$, a conditional expectation of X with respect to $\mathcal{G}$ is any $\mathcal{G}$-measurable random variable $\mathbb{E}(\mathrm{X}\mid\mathcal{G}):\omega\to\mathrm{G}$ such that $\int_\mathrm{A}\mathbb{E}(\mathrm{X}\mid\mathcal{G})(\omega)\mathrm{d}\mathbb{P}(\omega)=\int_\mathrm{A}\mathrm{X}(\omega)\mathrm{d}\mathbb{P}(\omega)$ for any $\mathrm{A}\in\mathcal{G}$.
Existence	Given by the Radon-Nikodym theorem, hence an absolute continuity condition. In practice, the existence is often proved in a constructive way, i.e. by showing a random variable with the desired properties (e.g. the Brownian bridge).
Uniqueness	In the almost sure sense, i.e. two conditional expectation of X with respect to $\mathcal{G}$ only differ on a set of probability measure 0.
Property	$\mathbb{E}(\mathbb{E}(\mathrm{X}\mid\mathcal{G}))=\mathbb{E}(\mathrm{X})$ whenever $\mathcal{G}\subset\mathcal{F}$.

Functional analysis

Closed operator Let $F : \mathcal{D}(F) \subset A \to B$ be a linear operator, A, B being Banach spaces. The operator F is said to be closed if for any sequence of elements $a_n \in \mathcal{D}(F)$ converging to $a \in A$, such that $F(a_n) \to b \in B$, $a \in \mathcal{D}(F)$ and $F(a) = b$: the graph of F is closed in the direct sum $A \oplus B$.

Riesz representation Any element φ of the dual H^* of a Hilbert space H can be uniquely written $\varphi = \langle x, \cdot \rangle$ for some $x \in H$.

$L^p - L^q$ duality If $1/p + 1/q = 1$, $p, q > 0$, then $\|f\|_{L^p} = \sup_{\|g\|_{L^q} \leqslant 1} \langle f, g \rangle$.

Convergence types

Almost sure On the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, X_n converges almost surely to X if $\mathbb{P}(X_n \underset{n\to\infty}{\to} X) = 1$, where the limit is in the sense of the metric of the space G.

L^p On the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, X_n converges to X in L^p $(p > 0)$ if $\mathbb{E}(|X_n - X|^p) \underset{n\to\infty}{\to} 0$, where the distance is in the sense of the metric of the space G, and $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$.

In probability On the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, X_n converges to X in probability if for any $\varepsilon > 0$ $\mathbb{P}(|X_n - X| > \varepsilon) \underset{n\to\infty}{\to} 0$, , where the distance is in the sense of the metric of the space G.

In law On possibly distinct probability spaces, but identical image space $(G, \mathcal{G})$, X_n with law $\mathbb{P}_n$ converges in law (or in distribution, weakly) to X if for any bounded continuous function f on G, $\mathbb{E}_n(f(X_n)) \underset{n\to\infty}{\to} \mathbb{E}(f(X))$.

Portmanteau's theorem Any of the following implies the convergence in law : (i) the test function f only needs to be Lipschitz (ii) if $G = \mathbb{R}^d$ it only needs to be infinitely differentiable (iii) for any closed subset C of G, $\limsup_{n\to\infty} \mathbb{P}_n(C) \leqslant \mathbb{P}(C)$ (iv)for any open subset O of G, $\liminf_{n\to\infty} \mathbb{P}_n(O) \geqslant \mathbb{P}(O)$.

Implications

$$\begin{tikzcd} \text{Almost sure} \arrow[dr] & L^p \arrow[d] & L^q \arrow[l, "(q>p>0)"'] \\ & \text{In probability} \arrow[d] & \\ & \text{In law} & \end{tikzcd}$$

Partial reciprocal Convergence in probability implies the almost convergence along a subsequence.

Paul Lévy's theorem Take $G = \mathbb{R}^d$. If, for any $u \in \mathbb{R}^d$, $\mathbb{E}_n(e^{iu\cdot X_n}) \underset{n\to\infty}{\to} f(u) := \mathbb{E}(e^{iu\cdot X})$ and f is continuous at 0, then X_n converges in law to X.

Paul Lévy : the easy implication If X_n converges in law to X, $\mathbb{E}(e^{iu\cdot X_n})$ converges to $\mathbb{E}(e^{iu\cdot X})$ uniformly in compact subsets of $\mathbb{R}^d$.

Theorems

Strong law of large numbers	For i.i.d. random variables X_i's with finite expectation, $\frac{1}{n}\sum_1^n \mathrm{X}_i$ converges almost surely to $\mathbb{E}(\mathrm{X}_1)$.
Central limit theorem	For i.i.d. random variables X_i's in L^2, with expectation μ and variance σ^2, $\frac{1}{\sigma\sqrt{n}}\sum_1^n(\mathrm{X}_i-\mu)$ converges in law to a standard Gaussian random variable.

Useful lemmas

Borel-Cantelli	On a probability space $(\Omega,\mathcal{F},\mathbb{P})$, if $\mathrm{A}_n\in\mathcal{F}$ and $\sum_n \mathbb{P}(\mathrm{A}_n)<\infty$, then $\mathbb{P}(\cap_{n=1}^\infty \cup_{m\geqslant n}\mathrm{A}_m)=0$.
Borel-Cantelli, independent case	On a probability space $(\Omega,\mathcal{F},\mathbb{P})$, if $\mathrm{A}_n\in\mathcal{F}$, the A_n's are independent and $\sum_n \mathbb{P}(\mathrm{A}_n)=\infty$, then $\mathbb{P}(\cap_{n=1}^\infty \cup_{m\geqslant n}\mathrm{A}_m)=1$.
Fatou	On a probability space $(\Omega,\mathcal{F},\mathbb{P})$, if $\mathrm{X}_n\geqslant 0$, then $\mathbb{E}(\liminf \mathrm{X}_n)\leqslant \liminf \mathbb{E}(\mathrm{X}_n)$.

Motivations

It is very natural to think about these random functions imagined by mathematicians, and that were wrongly only seen as mathematical curiosities.

Jean Perrin
Les atomes, 1913

In the above quote, Perrin refers to his derivation of the Avogadro number[1] : previous works by Einstein, supposing that molecules move randomly, yield to a theoretical derivation of this physical constant ; Perrin's work gave further experimental support to Einsein's hypothesis on the particles motion.

These trajectories, known now as Brownian motions, appeared first in the work of Brown about pollen particles moving on the surface of water. They exhibit strong irregularities (infinite variation, nowhere differentiable, uncountable set of zeros) and are universal in many senses presented hereafter. These lecture notes aim to give a self-contained introduction to these random paths, through the important contributions of Lévy, Wiener, Kolmogorov, Doob, Itô, but also to show how they enlighten non-probabilistic questions, such as a the Dirichlet problem and concentration of measure.

A first approach towards Brownian motion consists in an asymptotic analysis of random walks. Let X_i be independent Bernoulli random variables, i.e. $\mathbb{P}(\mathrm{X}_i = 1) = \mathbb{P}(\mathrm{X}_i = -1) = 1/2$. Consider the partial sum $\mathrm{S}_n = \sum_{i=1}^n \mathrm{X}_i$. The central limit theorem states that

$$\frac{\mathrm{S}_n}{\sqrt{n}} \xrightarrow{\text{law}} \mathcal{N}, \tag{0.1}$$

a standard Gaussian random variable. Many other questions of interest about S_n include the asymptotic distribution of $\sup_{1\leqslant i\leqslant n} \mathrm{S}_i$ or $|\{i \leqslant n \mid \mathrm{S}_i \geqslant 0\}|$ for example. In the case of Bernoulli random variables, these problems can be handled by combinatorial techniques involving the Catalan numbers, but for more general variables X_i in L^2, what is required is a weak limit of $(\mathrm{S}_n, n \geqslant 0)$, the entire process.

In order to get a functional limit of S_n, (0.1) gives the required normalization. More precisely, let

$$\mathrm{S}_n(t) = \frac{\mathrm{S}_{\lfloor nt \rfloor}}{\sqrt{n}},$$

and B_n be the unique continuous piecewise affine function such that $\mathrm{B}_n(t) = \mathrm{S}_n(t)$ when $nt \in \mathbb{N}$. In the Bernoulli case, as the jumps of the function S_n are $\pm 1/\sqrt{n}$, for any $s < t$

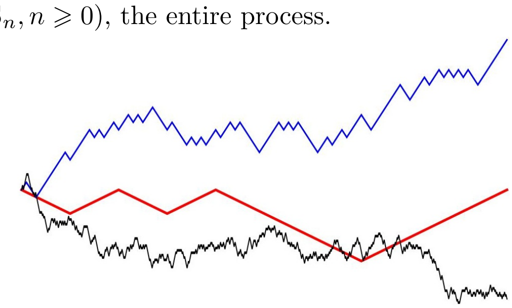

Figure 1. Samples of B_n, $0 \leqslant t \leqslant 1 : n = 10$ (red), 100 (blue) and 1000 (black).

$$\sup_{s=t_0<\dots<t_m=t} \sum |\mathrm{B}_n(t_{k+1}) - \mathrm{B}_n(t_k)| = (t-s)\sqrt{n} \underset{n\to\infty}{\longrightarrow} \infty,$$

1. Approximately 6×10^{23}, defined as the number of atoms in 12 gram of carbon-12 atoms in their ground state at rest. In his PhD thesis, Einstein derived from a Brownian motion hypothesis equations for diffusion coefficients and viscosities in which Avogadro's number appears. From experimental values of the diffusion coefficients and viscosities of sugar solutions in water, he obtained an approximation, improved by Perrin.

so if the path B_n converges in some sense to a limiting object B, this is of infinite variation on any nonempty interval.

Now, given any increasing sequence $0 = t_0 < t_1 < \dots$, the central limit theorem yields that $\mathrm{B}_n(t_{i+1}) - \mathrm{B}_n(t_i)$ converges as $n \to \infty$ to a normally distributed random variable with variance $t_{i+1} - t_i$. More precisely,

$$(\mathrm{B}_n(t_1), \mathrm{B}_n(t_2) - \mathrm{B}_n(t_1), \dots, \mathrm{B}_n(t_{k+1}) - \mathrm{B}_n(t_k)) \\ \xrightarrow{\text{law}} (\sqrt{t_1}\mathcal{N}_1, \sqrt{t_2 - t_1}\mathcal{N}_2, \dots, \sqrt{t_k - t_{k-1}}\mathcal{N}_k) \quad (0.2)$$

where $\mathcal{N}_1, \dots, \mathcal{N}_k$ are independent standard Gaussian random variables. This suggests the following definition.

Definition. *A random trajectory* $(\mathrm{B}_t, t \geqslant 0)$ *with values in* $\mathbb{R}$ *is a Brownian motion if the following four conditions are satisfied :*

(i) $\mathrm{B}_0 = 0$ *almost surely ;*

(ii) for any $n \geqslant 2$, $0 < t_1 \leqslant \dots \leqslant t_n$, $(\mathrm{B}_{t_1}, \mathrm{B}_{t_2} - \mathrm{B}_{t_1}, \dots, \mathrm{B}_{t_n} - \mathrm{B}_{t_{n-1}})$ *is a Gaussian vector with independent components ;*

(iii) for a any $0 < s < t$, $\mathrm{B}_t - \mathrm{B}_s \sim \mathcal{N}(0, t - s)$ *;*

(iv) B *is almost surely continuous.*

Note that such a definition implies strange properties of the sample trajectory. For example, for any $t > 0$, $s = t_0 < \dots < t_m = t$

$$\mathbb{E}\left(\sum_{i=1}^{m} |\mathrm{B}(t_{i+1}) - \mathrm{B}(t_i)|\right) \underset{n\to\infty}{\longrightarrow} \sum_{i=1}^{m} \sqrt{t_{i+1} - t_i}\, \mathbb{E}(|\mathcal{N}_1|) \geqslant \mathbb{E}(|\mathcal{N}_1|)\sqrt{t-s}\sqrt{m},$$

by the Cauchy-Schwarz inequality, so as expected the L^1-norm of the total variation of B on $[s, t]$ is ∞.

More interesting than the total variation is the quadratic one : from the convergence in law (0.2),

$$\mathbb{E}\left(\sum_{i=1}^{m} |\mathrm{B}_n(t_{i+1}) - \mathrm{B}_n(t_i)|^2\right) \underset{n\to\infty}{\longrightarrow} t - s,$$

suggesting that the quadratic variation of a Brownian path till any time t must be t. One can even prove from the definition of the Brownian motion that

$$\lim\left(\left(\sum_{i=1}^{m} |\mathrm{B}_n(t_{i+1}) - \mathrm{B}_n(t_i)|^2 - (t-s)\right)^2\right) = 0$$

where the limit is in the sense of the time step going to 0 for the subdivision $s = t_0 < \dots < t_m = t$. These observations can make skeptical about the existence of such paths. We will prove in Chapter 2 the following theorem, together with the first properties of Brownian motion. As a prerequisite, properties of discrete martingales, like S_n, will be studied in Chapter 1.

Theorem. *The Brownian motion exists. More precisely, there is a measure* W *on* $\mathscr{C}^o([0,1])$ *such that :*

(i) $\mathrm{W}(\{\omega(0) = 0\}) = 1$ *;*

(ii) for any $n \geqslant 2$, $0 < t_1 \leqslant \dots \leqslant t_n$, *the projection of* W *by* $\omega \to (\omega_{t_1}, \omega_{t_2} - \omega_{t_1}, \dots, \omega_{t_n} - \omega_{t_{n-1}})$ *is a Gaussian measure ;*

(iii) for a any $0 < s < t$*, the projection of* W *by* $\omega \to \omega_t - \omega_s$ *is the centered Gaussian measure with variance* $t - s$*.*

Moreover, this measure is unique and is called the Wiener measure.

Moreover, this measure is the weak limit of the random paths B_n, no matter which distribution the normalized X_i's have. This is Donsker's theorem.

Theorem. *Let* B_n *be constructed as previously from iid* X_i*'s,* $\mathbb{E}(\mathrm{X}_i) = 0$*,* $\mathbb{E}(\mathrm{X}_i^2) = 1$*. Then for any bounded continuous (for the* L^∞ *norm) functional* F *on* $\mathscr{C}^o([0,1])$*,*

$$\mathbb{E}(\mathrm{F}(\mathrm{B}_n(s), 0 \leqslant s \leqslant 1)) \underset{n\to\infty}{\longrightarrow} \mathbb{E}_{\mathrm{W}}(\mathrm{F}(\mathrm{B}(s), 0 \leqslant s \leqslant 1)).$$

We say that the process B_n *converges weakly to the Brownian motion.*

To some extent, the Brownian motion is the only *natural* random function. More precisely, take any continuous integrable random curve $(\mathrm{X}_s, s \geqslant 0)$ satisfying the martingale property

$$\mathbb{E}(\mathrm{X}_t \mid \mathcal{F}_s) = \mathrm{X}_s$$

for any $s < t$, where $\mathcal{F}_s = \sigma(\mathrm{X}_s, s \leqslant t)$. Then a theorem by Dubins and Schwartz states that there exists a nondecreasing $\mathcal{F}$-measurable random function f such that

$$(\mathrm{X}_{f(s)}, s \geqslant 0)$$

is a Wiener process. This means that, up to a change of time, any martingale (which does not presuppose normality) is a Brownian motion. In particular, this implies that any martingale must either be constant or have infinite variation on any interval : there are no smooth nontrivial martingales. A study of martingales is the purpose of Chapter 3.

Chapter 4 relates martingales and the Brownian motion through the Itô calculus. This requires the definition of a stochastic integral, as the limit in probability of

$$\mathrm{X}_t = \int_0^t a(\mathrm{B}_s)\mathrm{d}\mathrm{B}_s = \lim_{\mathrm{mesh}\to 0} \sum a(\mathrm{B}_{t_i})(\mathrm{B}_{t_{i+1}} - \mathrm{B}_{t_i})$$

under proper integrability assumptions for a. The Itô calculus gives the decomposition, as a stochastic integral, of transforms of the process X.

One application of the Itô formula will be the links between Brownian motion and harmonic functions.

More precisely, for example, consider a connected $\mathcal{D} \subset \mathbb{R}^2$ with smooth boundary $\partial\mathcal{D}$. The Dirichlet problem consists in finding a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\begin{cases} f(x) = u(x) & , x \in \partial\mathcal{D}, \\ \Delta f = 0 & , x \in \mathcal{D}^o. \end{cases}$$

Given a bidimensional Brownian motion $\mathrm{B} = (\mathrm{B}_1, \mathrm{B}_2)$ (B_1 and B_2 independent Brownian motions with values in $\mathbb{R}$), a solution to the Dirichlet problem is

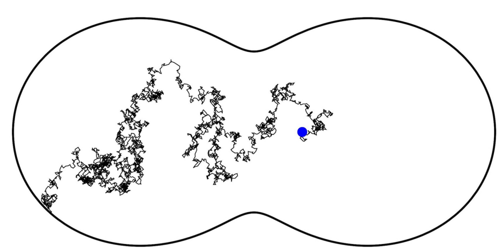

Figure 2. A bidimensional Bownian motion till hitting the peanut frontier.

$$f(x) = \mathbb{E}\left(u(\mathrm{B}_\mathrm{T}) \mid \mathrm{B}_0 = x\right)$$

where $\mathrm{T} = \inf\{t \mid \mathrm{B}_t \notin \mathrm{U}\}$. This is an example of existence results proved by probabilistic means.

Chapters 5 focuses on stochastic differential equations, here existence and uniqueness results for dynamics driven by a Brownian motion. Some analogies with ordinary differential equations appear, for example if the coefficients a and b are Lipschitzian, there is a unique path X measurable in terms of B ($\mathrm{X}_t = f(\mathrm{B}_s, s \leqslant t)$) such that

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

Surprisingly, we will see that in other situations solutions to stochastic differential equations require less smoothness assumptions than in the deterministic case. For example, for

$$\mathrm{X}_t = \mathrm{B}_t + \int_0^t b(\mathrm{X}_s)\mathrm{d}s,$$

the measurability and boundedness of b yield to existence and uniqueness of a solution. The question, important in filtering theory, whether X is a function of B in the above equation will be addressed.

Chapters 6 has a more functional analysis content. First, we deal with representations of random variables. For example, any $\mathcal{F}_1$-measurable[2] random variable X in L^2 can be written as

$$\mathrm{X} = \int_0^1 a_s \mathrm{dB}_s \tag{0.3}$$

for some adapted process a. This implies in particular such random variables can be decomposed into an orthogonal basis generated by multiple stochastic integrals, the so-called chaos decomposition. Then the Gross-Sobolev derivative is introduced. This allows to investigate questions like the infinitesimal variation of F(B) (F being $\mathcal{F}_1$-measurable, say) when B is perturbed by a deterministic element h, such that $h(t) = \int_0^t \dot{h}(s)\mathrm{d}s$, $\|h\|_{\mathrm{H}}^2 = \int_0^t |\dot{h}(s)|^2 \mathrm{d}s < \infty$: for a suitable function F, by the Riesz representation theorem, there is an element $\nabla\mathrm{F}(\mathrm{B}) \in \mathrm{H}$ such that

$$\lim_{\varepsilon} \frac{\mathrm{F}(\mathrm{B}+\varepsilon h) - \mathrm{F}(\mathrm{B})}{\varepsilon} = \langle \nabla\mathrm{F}(\mathrm{B}), h\rangle.$$

This $\nabla\mathrm{F}$ is called the Gross Sobolev derivative of F, and will make it possible to give a complete description of what the processus a is in Itô's representation theorem (0.3). This is important in control theory : it enables to get a random variable X by integrating predictably along the random path B.

Finally, we give two applications, of independent interest, of the stochastic analysis. The first one is about concentration of measure, in Chapter 8. Examples of such a concentration appear on any compact Riemannian manifold with positive Ricci curvature. For simplification here, consider the case of a n-dimensional unit sphere $\mathscr{S}^n$, with uniform probability measure μ_n. Then, for any Lipschitz function F with Lipschitz constant $\|\mathrm{F}\|_{\mathcal{L}}$, there are constants $\mathrm{C}, c > 0$ independent of n, F such that

$$\mu_n\left\{|\mathrm{F} - \mathbb{E}_\mu(\mathrm{F}) > \varepsilon|\right\} \leqslant \mathrm{C}e^{-c(n-1)\frac{\varepsilon^2}{\|\mathrm{F}\|_{\mathcal{L}}^2}}.$$

In particular, choosing

$$\mathrm{F}(x) = \min(\mathrm{dist}(x,\mathrm{M}), \varepsilon)$$

where M is the Equator, $\mathrm{M} = \{x_1 = 0\} \cap \mathscr{S}_n$, yields for some constants $\mathrm{C}, c > 0$

$$\mu_n(\mathrm{dist}(x,\mathrm{M}) < \varepsilon) \geqslant 1 - \mathrm{C}e^{-cn\varepsilon^2},$$

i.e. all the mass of the uniform measure concentrates exponentially fast around

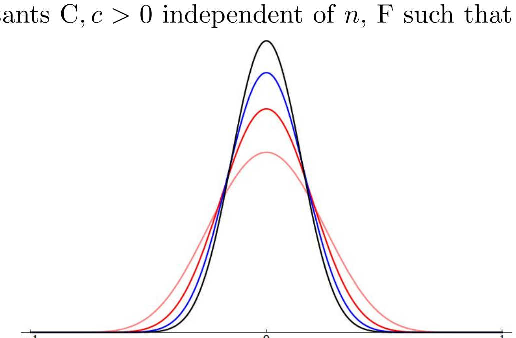

Figure 3. Density of x_1 for the uniform measure on $\mathscr{S}^n$, $n = 20$ (pink), 30 (red), 40 (blue), 50 (black).

2. $\mathcal{F}_s = \sigma(\mathrm{B}_u, u \leqslant s)$

the Equator ! How are these results related to stochastic processes ? A method to prove concentration, initiated by Bakry and Emery, roughly speaking consists in seing the uniform measure as the equilibrium measure of a Brownian motion on the manifold : $\mathrm{F}-\mathbb{E}_\mu(\mathrm{F})$ is the different between $t=0$ and $t=\infty$ of an evolution along the Browian path, whose differential can be controlled.

This shows an application of Brownian paths, and more generally dynamics, to time-independent probabilistic statements.

Chapter 1

Discrete time processes

Although these lectures focus on the continuous time setting, which involves phenomena not appearing in discrete time, this chapter aims to give the intuition and necessary tools for the following ones. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an increasing sequence $(\mathcal{F}_n)_{n\geqslant 0}$ of sub-σ-algebras of $\mathcal{F}$ is called a filtration :

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n \subset \dots$$

Intuitively, the index n is a time and a random variable is $\mathcal{F}_n$-measurable if it only depends on the past up to time n.

1. Martingales, stopping times, the martingale property for stopping times

Definition 1.1. *A sequence* $(\mathrm{X}_n)_{n\geqslant 0}$ *of random variables is called a* $((\mathcal{F}_n)_{n\geqslant 0}, \mathbb{P})$*-martingale if it satisfies the three following conditions, for any* $n \geqslant 0$ *:*

(i) $\mathbb{E}(|\mathrm{X}_n|) < \infty$ *;*

(ii) X_n *is* $\mathcal{F}_n$*-measurable ;*

(iii) $\mathbb{E}(\mathrm{X}_n \mid \mathcal{F}_m) = \mathrm{X}_m$ *for any* $m \leqslant n$*,* $\mathbb{P}$*-almost surely.*

The submartingale (resp. supermartingale) is defined in the same way, except that $\mathbb{E}(\mathrm{X}_n \mid \mathcal{F}_m) \geqslant \mathrm{X}_m$ *(resp.* $\mathbb{E}(\mathrm{X}_n \mid \mathcal{F}_m) \leqslant \mathrm{X}_m$*)*

The above conditional expectations are assumed to exist, and there is no ambiguity on the choice (necessarily made up to a set of measure 0). Note that, as a direct consequence of the above definition, a martingale X vanishing almost surely at time n vanishes almost surely on $[\![0, n]\!]$.

As an example, if the X_i's are independent Bernoulli random variables with parameter $1/2$ and $\mathcal{F}_n = \sigma(\mathrm{X}_1, \dots, \mathrm{X}_n)$,

$$\mathrm{S}_n := n\mu + \sum_{i=1}^{n} \mathrm{X}_i \qquad (1.1)$$

is a $(\mathcal{F}_n)_{n\geqslant 0}$-martingale for $\mu = 0$, a submartingale for $\mu > 0$, and a supermartingale if $\mu < 0$. This process is called the biased random walk with parameter μ.

The submartingales are stable when composing with a no-decreasing convex function, and a similar result holds for supermartingales, taking the opposite.

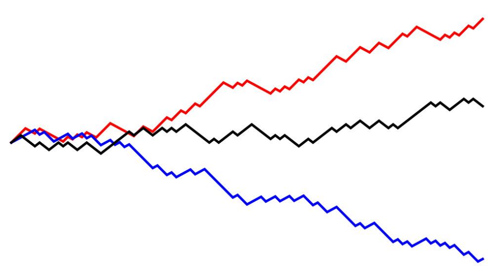

Figure 1.1. Examples of trajectories for S_n for 100 steps : martingale ($\mu = 0$, black), submartingale ($\mu > 0$, red), supermartingale ($\mu < 0$, blue).

Proposition 1.2. *Let* X *be a* $(\mathcal{F}_n)_{n\geqslant 0}$*-submartingale and* f *a Lipschitz convex non-decreasing function. Then* $(f(\mathrm{X}_n), n \geqslant 0)$ *is a* $(\mathcal{F}_n)_{n\geqslant 0}$*-submartingale.*

We will often use the following special case : if X is a submartingale, so is $\sup(\mathrm{X}, 0)$.

Proof. The Lipschitz hypothesis is just a condition ensuring that $f(\mathrm{X}_n)$ is integrable. Now, for $n \geqslant m$, by Jensen's inequality

$$\mathbb{E}\left(f(\mathrm{X}_n) \mid \mathcal{F}_m\right) \geqslant f\left(\mathbb{E}(\mathrm{X}_n \mid \mathcal{F}_m)\right).$$

As $\mathbb{E}(\mathrm{X}_n \mid \mathcal{F}_m) \geqslant \mathrm{X}_m$ and f is non-decreasing, this is greater than $f(\mathrm{X}_m)$. □

Now, consider a specific class of random times adapted to the filtration : their values are determined in view of the past.

Definition 1.3. *A function* $\mathrm{T} : \Omega \to \mathbb{N} \cup \{+\infty\}$ *is a stopping time if* $\{\mathrm{T} \leqslant n\} \in \mathcal{F}_n$ *for any* $n \geqslant 0$.

An example of stopping time is $\mathrm{T}_a = \inf\{n \geqslant 0 \mid \mathrm{S}_n > a\}$, the time when the process (1.1) reaches the level $a \in \mathbb{R}$.

The definitions prove that the set

$$\mathcal{F}_\mathrm{T} := \{\mathrm{A} \in \mathcal{F} \mid \forall n \in \mathbb{N}, \mathrm{A} \cap \{\mathrm{T} \leqslant n\} \in \mathcal{F}_n\}$$

is a σ-algebra. Intuitively, $\mathcal{F}_\mathrm{T}$ is the information available at time T[1]. Other straightforward properties are (i) if $\mathrm{S} \leqslant \mathrm{T}$, $\mathcal{F}_\mathrm{S} \subset \mathcal{F}_\mathrm{T}$ and (ii) T and $\mathrm{X}_\mathrm{T} \mathbb{1}_{\mathrm{T}<\infty}$ are $\mathcal{F}_\mathrm{T}$-measurable.

Proposition 1.4. *Let* X *be a* $(\mathcal{F}_n)_{n\geqslant 0}$*-submartingale and* $(\mathrm{H}_n, n \geqslant 0)$ *be a bounded nonnegative process, with* $\mathrm{H}_n \in \mathcal{F}_{n-1}$ *for* $n \geqslant 1$*. Then the process* Y *defined by* $\mathrm{Y}_0 = 0$ *and* $\mathrm{Y}_n = \mathrm{Y}_{n-1} + \mathrm{H}_n(\mathrm{X}_n - \mathrm{X}_{n-1})$ *is a* $(\mathcal{F}_n)_{n\geqslant 0}$*-submartingale. In particular, if* T *is a stopping time,* $(\mathrm{X}_{n\wedge \mathrm{T}}, n \geqslant 0)$ *is a submartingale.*

Proof. For the first statement, note that when $n > m$

$$\begin{aligned}\mathbb{E}(\mathrm{Y}_n \mid \mathcal{F}_m) &= \mathbb{E}(\mathrm{Y}_{n-1} + \mathrm{H}_n(\mathrm{X}_n - \mathrm{X}_{n-1}) \mid \mathcal{F}_m)\\ &= \mathbb{E}(\mathrm{Y}_{n-1} + \mathrm{H}_n\,\mathbb{E}(\mathrm{X}_n - \mathrm{X}_{n-1} \mid \mathcal{F}_{n-1}) \mid \mathcal{F}_m) \geqslant \mathbb{E}(\mathrm{Y}_{n-1} \mid \mathcal{F}_m).\end{aligned}$$

By an immediate induction, this implies $\mathbb{E}(\mathrm{Y}_n \mid \mathcal{F}_m) \geqslant \mathrm{Y}_m$. The second statement follows when choosing $\mathrm{H}_n = \mathbb{1}_{n\leqslant \mathrm{T}}$, which is $\mathcal{F}_{n-1}$-measurable by definition of a stopping time. □

Note that the above result implies that a martingale frozen at a stopping time is still a martingale.

A natural question is whether the martingale property $\mathbb{E}(\mathrm{X}_n \mid \mathcal{F}_m) = \mathrm{X}_m$ remains when m, n are random times. The answer is yes for bounded, ordered stopping times, as shown by the following theorem. To prove it, we first note that if $\mathrm{T} \leqslant c$ is a stopping time, then $\mathbb{E}(\mathrm{X}_\mathrm{T}) = \mathbb{E}(\mathrm{X}_0)$. Indeed, as $\{\mathrm{T} = n\} \in \mathcal{F}_n$,

$$\begin{aligned}\mathbb{E}(\mathrm{X}_\mathrm{T}) = \sum_{n=1}^{c} \mathbb{E}(\mathrm{X}_n \mathbb{1}_{\mathrm{T}=n}) &= \sum_{n=1}^{c} \mathbb{E}\left(\mathbb{E}(\mathrm{X}_c \mid \mathcal{F}_n) \mathbb{1}_{\mathrm{T}=n}\right)\\ &= \sum_{n=1}^{c} \mathbb{E}\left(\mathbb{E}(\mathrm{X}_c \mathbb{1}_{\mathrm{T}=n} \mid \mathcal{F}_n)\right) = \mathbb{E}(\mathrm{X}_c) = \mathbb{E}(\mathrm{X}_0).\end{aligned}$$

Theorem 1.5 (Doob's first sampling or stopping theorem)**.** *Let* $(\mathrm{X}_n)_{n\geqslant 0}$ *be a martingale, and* T, S *two stopping times such that* $\mathrm{S} \leqslant \mathrm{T} \leqslant c$ *for some constant* $c > 0$*. Then almost surely*

$$\mathbb{E}(\mathrm{X}_\mathrm{T} \mid \mathcal{F}_\mathrm{S}) = \mathrm{X}_\mathrm{S}.$$

1. Example : let T is the first sunny day of the year in Paris, A the event *my hibernation has been longer than 60 days this year*, and B the event *at some time in History, Mars is aligned with Jupiter and Saturn.* Then, assuming one does not hibernate on sunny days, $\mathrm{A} \in \mathcal{F}_\mathrm{T}$ and $\mathrm{B} \notin \mathcal{F}_\mathrm{T}$. Note that, when replacing Paris with London in the above example, remarkably $\mathrm{B} \in \mathcal{F}_\mathrm{T}$.

Proof. First, $X_T \leqslant \sum_{i=0}^{c} |X_i|$ is clearly integrable and X_S is $\mathcal{F}_S$ measurable, so from the definition of the conditional expectation we need to check that for any $\mathcal{F}_S$ measurable bounded random variable Z,

$$\mathbb{E}(X_T Z) = \mathbb{E}(X_S Z).$$

The random variable Z can be chosen of the form $\mathbb{1}_A$ where $A \in \mathcal{F}_S$ (by approximating nonnegative random variables by linear combinations of such indicators, and a substraction gives the result for any bounded Z).

Now, define U as the random time S if $\omega \in A$ and T if $\omega \notin A$. Then U is a stopping time, bounded by c, so $\mathbb{E}(X_U) = X_0 = \mathbb{E}(X_T)$. A simplification, writing $X_U = X_S \mathbb{1}_A + X_T \mathbb{1}_{A^c}$ yields the expected result $\mathbb{E}(X_T \mathbb{1}_A) = \mathbb{E}(X_S \mathbb{1}_A)$. □

This theorem has an immediate equivalent form for submartingales and supermartingales evaluated on bounded stopping times. Moreover, note that this boundedness assumption is essential : if S is the random walk (1.1) with parameter $\mu = 0$, and $T = \inf\{n \geqslant 0 \mid S_n = 1\}$, the (false) equality $\mathbb{E}(S_T \mid \mathcal{F}_0) = S_0$ yields $1 = 0$.

Finally, the above result gives easy proofs of not being a stopping time. As an example, let $m \geqslant 1$ and $T = \sup\{n \in [\![0, m]\!] \mid S_n = \sup_{[\![0,10]\!]} S_i\}$. The (false) equality $\mathbb{E}(S_T \mid \mathcal{F}_0) = S_0$ would give $\mathbb{E}(S_T) = 0$, obviously false as $\mathbb{E}(S_T) > \mathbb{P}(S_1 = 1) = 1/2$. Hence T is not a stopping time.

2. Inequalities : L^p norms in terms of final values, number of jumps.

Both theorems below are stated for nonnegative submartingale. By by replacing X by $X_+ = \sup(X, 0)$ – a submartingale from Proposition 1.2 – they also give estimates for general submartingales.

Theorem 1.6 (Doob's maximal inequality)**.** *Let* $(X_n)_{n\geqslant 0}$ *be a non-negative submartingale, and* $X_n^* = \sup_{[\![0,n]\!]} X_i$*. Then for any* $\lambda > 0$

$$\mathbb{P}(X_n^* \geqslant \lambda) \leqslant \frac{\mathbb{E}(X_n)}{\lambda}.$$

Proof. Hence for any bounded stopping time $T \leqslant n$, $\mathbb{E}(X_T) \leqslant \mathbb{E}(X_n)$.

Take $T = \inf(n, \inf\{k \mid X_k \geqslant \lambda\})$:

$$\mathbb{P}(X_n^* \geqslant \lambda) = \mathbb{P}(X_T \geqslant \lambda) \leqslant \frac{1}{\lambda}\,\mathbb{E}(X_T) \leqslant \frac{1}{\lambda}\,\mathbb{E}(X_n),$$

which is the expected result. □

Theorem 1.7 (Doob's inequality)**.** *Let* $(X_n)_{n\geqslant 0}$ *be a non-negative submartingale. Then for any* $p > 1$

$$\mathbb{E}\left((X_n^*)^p\right) \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}\left(X_n^p\right),$$

may the right member be infinite.

Proof. Note that the proof of Doob's maximal inequality yields slightly more than stated :

$$\lambda \mathbb{P}(X_n^* \geqslant \lambda) \leqslant \mathbb{E}(X_n \mathbb{1}_{X_n^* \geqslant \lambda}).$$

Indeed, still writing $T = \inf(n, \inf\{k \mid X_k \geqslant \lambda\})$, as $\mathbb{1}_{X_n^* \geqslant \lambda}$ is $\mathcal{F}_T$-measurable and $T \leqslant n$,

$$\mathbb{E}(X_n \mathbb{1}_{X_n^* \geqslant \lambda}) \geqslant \mathbb{E}(X_T \mathbb{1}_{X_n^* \geqslant \lambda}) \geqslant \mathbb{E}(\lambda \mathbb{1}_{X_T \geqslant \lambda} \mathbb{1}_{X_n^* \geqslant \lambda}) = \lambda \mathbb{P}(X_T \geqslant \lambda).$$

Hence we only need to prove that for nonnegative random variables X and Y, if for any $\lambda \geqslant 0$

$$\lambda \mathbb{P}(\mathrm{Y} \geqslant \lambda) \leqslant \mathbb{E}(\mathrm{X} \mathbb{1}_{\mathrm{Y} \geqslant \lambda}), \tag{1.2}$$

then $\mathbb{E}(\mathrm{Y}^p) \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}(\mathrm{X}^p)$. First note that, as $y^p = \int_0^y p\lambda^{p-1}\mathrm{d}\lambda$, $\mathbb{E}(\mathrm{Y}^p) = \int_0^\infty p\lambda^{p-1}\mathbb{P}(\mathrm{Y} \geqslant \lambda)\mathrm{d}\lambda$. This yields, using successively the hypothesis and Hölder's inequality (we define q by $1/p + 1/q = 1$),

$$\begin{aligned}\mathbb{E}(\mathrm{Y}^p) = \int_0^\infty p\lambda^{p-1}\mathbb{P}(\mathrm{Y} \geqslant \lambda)\mathrm{d}\lambda &\leqslant \int_0^\infty p\lambda^{p-2}\,\mathbb{E}(\mathrm{X}\mathbb{1}_{\mathrm{Y}\geqslant\lambda})\mathrm{d}\lambda \\ &= \mathbb{E}\left(\mathrm{X}\int_0^{\mathrm{Y}} p\lambda^{p-2}\mathrm{d}\lambda\right) = q\,\mathbb{E}(\mathrm{X}\mathrm{Y}^{p-1}) \leqslant q\|\mathrm{X}\|_{\mathrm{L}^p}\|\mathrm{Y}^{p-1}\|_{\mathrm{L}^q}.\end{aligned}$$

If $\mathbb{E}(\mathrm{Y}^p) < \infty$, the inequality $\mathbb{E}(\mathrm{Y}^p) \leqslant q\|\mathrm{X}\|_{\mathrm{L}^p}\|\mathrm{Y}^{p-1}\|_{\mathrm{L}^q}$ can be simplified and gives the required result. If $\mathbb{E}(\mathrm{Y}^p) = \infty$, for any $m \geqslant 0$ define $\mathrm{T}_m = \inf\{k \geqslant 0 \mid |\mathrm{X}_k| > m\} \wedge n$. Then by using the previous result for the submartingale $(\mathrm{X}_{n\wedge \mathrm{T}_m})_{n\geqslant 0}$, we obtain

$$\|\sup_{0\leqslant k\leqslant n} \mathrm{X}_{k\wedge\mathrm{T}_m}\|_{\mathrm{L}^p}^p \leqslant \left(\frac{p}{p-1}\right)^p \|\mathrm{X}_n\|_{\mathrm{L}^p}^p.$$

By monotone convergence, taking $m \to \infty$ allows to conclude. □

Remark. For $p = 1$, a similar inequality holds :

$$\mathbb{E}\left(\mathrm{X}_n^*\right) \leqslant \frac{e}{e-1}\left(1 + \mathbb{E}(|\mathrm{X}_n| \log_+ |\mathrm{X}_n|)\right),$$

Both of Doob's inequalities above will be important in the next chapters as it allows to control uniform convergence of martingales from their distribution at a given time.

Theorem 1.8 (Doob's jumps inequality). *Let $(\mathrm{X}_n)_{n\geqslant 0}$ be a submartingale and $a < b$. Let U_n be the number of jumps from a to b before time n. More precisely, define by induction $\mathrm{T}_0 = 0$, $\mathrm{S}_{j+1} = \inf\{k \geqslant \mathrm{T}_j \mid \mathrm{X}_k \leqslant a\}$ and $\mathrm{T}_{j+1} = \inf\{k > \mathrm{S}_{j+1} \mid \mathrm{X}_k \geqslant b\}$. Then $\mathrm{U}_n = \sup\{j \mid \mathrm{T}_j \leqslant n\}$.*

Then

$$\mathbb{E}(\mathrm{U}_n) \leqslant \frac{1}{b-a}\,\mathbb{E}\left((\mathrm{X}_n - a)_+\right).$$

Proof. Let $\mathrm{Y}_n := (\mathrm{X}_n - a)_+$. By definition, $\mathrm{Y}_{\mathrm{T}_i} - \mathrm{Y}_{\mathrm{S}_i} \geqslant b - a$ for the above stopping times T_i, S_i smaller than n. Consequently,

$$\begin{aligned}\mathrm{Y}_n = \mathrm{Y}_{\mathrm{S}_1\wedge n} + \sum_{i\geqslant 1}(\mathrm{Y}_{\mathrm{T}_i\wedge n} - \mathrm{Y}_{\mathrm{S}_i\wedge n}) &+ \sum_{i\geqslant 1}(\mathrm{Y}_{\mathrm{S}_{i+1}\wedge n} - \mathrm{Y}_{\mathrm{T}_i\wedge n}) \\ &\geqslant (b-a)\mathrm{U}_n + \sum_{i\geqslant 1}(\mathrm{Y}_{\mathrm{S}_{i+1}\wedge n} - \mathrm{Y}_{\mathrm{T}_i\wedge n}).\end{aligned}$$

Now, as $x \to (x-a)_+$ is convex and X is a submartingale, so is Y, from Proposition 1.2. Consequently, as $\mathrm{S}_{i+1} \wedge n \geqslant \mathrm{T}_i \wedge n$ are increasing bounded stopping times, $\mathbb{E}(\mathrm{Y}_{\mathrm{S}_{i+1}\wedge n}) \geqslant \mathbb{E}(\mathrm{Y}_{\mathrm{T}_i\wedge n})$, so taking expectations in the previous inequality yields $\mathbb{E}(\mathrm{Y}_n) \geqslant (b-a)\,\mathbb{E}(\mathrm{U}_n)$, the expected result. □

3. Convergence of martingales.

Theorem 1.9 (Convergence of submartingales). *Let X_n be a submartingale such that $\sup_n \mathbb{E}\left((\mathrm{X}_n)_+\right) < \infty$. Then X_n converges almost surely to some $\mathrm{X} \in \mathbb{R}$ and $\mathrm{X} \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof. Given any $a < b$, let $\mathrm{U}_n(a,b)$ be the number of jumps from a to b before time n, like in Theorem 1.8. By monotone convergence, $\mathbb{E}(\mathrm{U}_\infty(a,b)) = \lim_{n\to\infty} \mathbb{E}(\mathrm{U}_n(a,b))$. But $\mathbb{E}(\mathrm{U}_n(a,b))$ is uniformly bounded :

$$\mathbb{E}(\mathrm{U}_n(a,b)) \leqslant \frac{1}{b-a}\,\mathbb{E}((\mathrm{X}_n-a)_+) \leqslant \frac{1}{b-a}\left(\mathbb{E}(\mathrm{X}_n)_+ + a\right).$$

Consequently, $\mathbb{E}(\mathrm{U}_\infty(a,b)) < \infty$, so $\mathrm{U}_\infty(a,b) < \infty$ almost surely. From the inclusion of events

$$\cap_{a<b,a,b\in\mathbb{Q}}\{\mathrm{U}_\infty(a,b)<\infty\} \subset \{\mathrm{X}_n \text{ converges}\}$$

we conclude that X_n converges almost surely. The limit X eventually can be $\pm\infty$, but this is not the case. Indeed, note that $|x| = 2x_+ - x$, so

$$\mathbb{E}(|\mathrm{X}_n|) = 2\,\mathbb{E}((\mathrm{X}_n)_+) - \mathbb{E}(\mathrm{X}_n) \leqslant 2\,\mathbb{E}((\mathrm{X}_n)_+) - \mathbb{E}(\mathrm{X}_0)$$

because $\mathbb{E}(\mathrm{X}_n) \geqslant \mathbb{E}(\mathrm{X}_0)$ (X is a submartingale). Hence, from the hypothesis, $\mathbb{E}(|\mathrm{X}_n|)$ is uniformly bounded, and using Fatou's lemma

$$\mathbb{E}(|\mathrm{X}|) = \mathbb{E}(\lim_n |\mathrm{X}_n|) \leqslant \liminf \mathbb{E}(|\mathrm{X}_n|) < \infty.$$

We proved that $|\mathrm{X}|$ is integrable, so $\mathrm{X} \in \mathbb{R}$ almost surely. □

For the following result, the notion of uniformly integrable family of random variables is required : this is a set $\{\mathrm{X}_n\}_{n\geqslant 0}$ of elements in $\mathrm{L}^1(\Omega,\mathcal{F},\mathbb{P})$ such that

$$\lim_{\lambda\to\infty}\sup_n \mathbb{E}(|\mathrm{X}_n|\mathbb{1}_{|\mathrm{X}_n|\geqslant\lambda}) = 0.$$

A $(\Omega,\mathcal{F},\mathbb{P})$-martingale X is said to be uniformly integrable if $\{\mathrm{X}_n\}_{n\geqslant 0}$ is uniformly integrable. There are many criteria to prove uniform integrability. For example if, for some $p > 1$, $\sup_n \mathbb{E}(|\mathrm{X}_n|^p) < \infty$, the family is uniformly integrable. If there is an integrable Y such that, for any n, $\mathrm{X}_n \leqslant \mathrm{Y}$, then the family $\{\mathrm{X}_n\}_{n\geqslant 0}$ is integrable.

A straightforward reasoning yields that if X is integrable, $\mathrm{X}_n := \mathbb{E}(\mathrm{X} \mid \mathcal{F}_n)$ is a uniformly integrable martingale. The following result shows that the converse is true.

Theorem 1.10 (Convergence of martingales)**.** *Let* X_n *be a uniformly integrable martingale. Then* X_n *converges almost surely and in* $\mathrm{L}^1(\Omega,\mathcal{F},\mathbb{P})$ *to some integrable* $\mathrm{X}\in\mathbb{R}$ *and* $\mathrm{X}_n = \mathbb{E}\left(\mathrm{X} \mid \mathcal{F}_n\right)$ *for any* $n \geqslant 0$.

Proof. As the family is uniformly integrable, for some $\varepsilon > 0$ there is a $\lambda > 0$ such that $\sup_n \mathbb{E}(|\mathrm{X}_n|\mathbb{1}_{|\mathrm{X}_n|>\lambda}) < \varepsilon$. Hence

$$\mathbb{E}((\mathrm{X}_n)_+) \leqslant \mathbb{E}(|\mathrm{X}_n|) = \mathbb{E}(|\mathrm{X}_n|\mathbb{1}_{|\mathrm{X}_n|>\lambda}) + \mathbb{E}(|\mathrm{X}_n|\mathbb{1}_{|\mathrm{X}_n|\leqslant\lambda}) \leqslant \varepsilon + \lambda,$$

so $\sup_n \mathbb{E}((\mathrm{X}_n)_+) < \infty$ and the convergence of submartingales, Theorem 1.9, implies that $(\mathrm{X}_n)_{n\geqslant 0}$ converges almost surely to some $\mathrm{X}\in\mathrm{L}^1(\Omega,\mathcal{F},\mathbb{P})$.

Now, take any $\varepsilon > 0$ and note, for any $\lambda > 0$,

$$f_\lambda(x) = x\mathbb{1}_{|x|\leqslant\lambda} + \lambda\mathbb{1}_{x>\lambda} - \lambda\mathbb{1}_{x<-\lambda}.$$

The uniform integrability implies that there is a sufficiently large $\lambda > 0$ such that

$$\mathbb{E}(|\mathrm{X}_n - f_\lambda(\mathrm{X}_n)|) < \varepsilon. \tag{1.3}$$

Moreover, as $\mathrm{X}\in\mathrm{L}^1(\Omega,\mathcal{F},\mathbb{P})$, dominated convergence implies that, for sufficiently large λ,

$$\mathbb{E}(|\mathrm{X} - f_\lambda(\mathrm{X})|) < \varepsilon. \tag{1.4}$$

Finally, $\mathrm{X}_n \to \mathrm{X}$ a.s. so $f_\lambda(\mathrm{X}_n) \to f_\lambda(\mathrm{X})$ a.s. by continuity of f_λ. This implies by dominated convergence that given λ, for sufficiently large n

$$\mathbb{E}(|f_\lambda(\mathrm{X}) - f_\lambda(\mathrm{X}_n)|) < \varepsilon. \tag{1.5}$$

Equations (1.3), (1.4) and (1.5) together prove that X_n converges to X in L^1.

Our last task is proving that $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$. As usual, taking some $A \in \mathcal{F}_n$, we need to prove that

$$\mathbb{E}(X_n \mathbb{1}_A) = \mathbb{E}(X \mathbb{1}_A). \tag{1.6}$$

For any $m \geqslant n$, as $X_n = \mathbb{E}(X_m \mid \mathcal{F}_n)$, $\mathbb{E}(X_n \mathbb{1}_A) = \mathbb{E}(X_m \mathbb{1}_A)$. Consequently,

$$|\mathbb{E}(X_n \mathbb{1}_A) - \mathbb{E}(X \mathbb{1}_A)| \leqslant |\mathbb{E}(X_n \mathbb{1}_A) - \mathbb{E}(X_m \mathbb{1}_A)| + |\mathbb{E}(X_m \mathbb{1}_A) - \mathbb{E}(X \mathbb{1}_A)| \leqslant \mathbb{E}(|X_m - X|).$$

As $X_m \to X$ in L^1, this converges to 0 as $m \to \infty$, which proves (1.6). □

Theorem 1.11 (Doob's optional sampling or stopping theorem)**.** *Let* $(X_n)_{n \geqslant 0}$ *be a uniformly integrable martingale, and* T, S *two stopping times such that* $S \leqslant T$*. Then almost surely*

$$\mathbb{E}(X_T \mid \mathcal{F}_S) = X_S.$$

Proof. We first prove the result when $T = \infty$. We know, by Theorem 1.10, that X_n converges almost surely and in L^1 to some $X \in L^1$. Then, note that for any $n \in \mathbb{N}$,

$$\mathbb{E}(X \mid \mathcal{F}_S) \mathbb{1}_{S=n} = \mathbb{E}(X \mid \mathcal{F}_n) \mathbb{1}_{S=n}. \tag{1.7}$$

To prove the above identity, note that both terms are $\mathcal{F}_n$-measurable and that, for any $A \in \mathcal{F}_n$,

$$\int_{A \cup \{S=n\}} \mathbb{E}(X \mid \mathcal{F}_n) d\mathbb{P} = \int_{A \cup \{S=n\}} X d\mathbb{P} = \int_{A \cup \{S=n\}} \mathbb{E}(X \mid \mathcal{F}_S) d\mathbb{P}$$

because $A \cup \{S = n\}$ is both in $\mathcal{F}_n$ and $\mathcal{F}_S$. Now, using Theorem 1.10 and (1.7),

$$X_S = \sum_n X_n \mathbb{1}_{S=n} = \sum_n \mathbb{E}(X \mid \mathcal{F}_n) \mathbb{1}_{S=n} = \sum_n \mathbb{E}(X \mid \mathcal{F}_S) \mathbb{1}_{S=n} = \mathbb{E}(X \mid \mathcal{F}_S),$$

where the summation over n includes the case $n = \infty$, corresponding to $X_S = X$, in case S is ∞ with positive probability. This gives the expected result when $T = \infty$, and the conditioning

$$X_S = \mathbb{E}(X \mid \mathcal{F}_S) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_T) \mid \mathcal{F}_S) = \mathbb{E}(X_T \mid \mathcal{F}_S)$$

yields the general case. □

We end this chapter with a convergence result about processes indexed by $-\mathbb{N}$, this will be useful when extending Theorem 1.11 to the continuous setting. An inverse martingale is a sequence of random variables $(X_{-n}, n \geqslant 0)$ in L^1 such that for $m > n \geqslant 0$

$$\mathbb{E}(X_{-n} \mid \mathcal{F}_{-m}) = X_{-m}$$

almost surely, where $\mathcal{F}_{-m} \subset \mathcal{F}_{-n} : (\mathcal{F}_{-n}, n \geqslant 0)$ can be thought of as a filtration indexed by negative times.

Theorem 1.12. *Let* $(X_{-n}, n \geqslant 0)$ *be an inverse martingale for the filtration* $(\mathcal{F}_{-n}, n \geqslant 0)$*. Then, as* $n \to \infty$*,* X_n *converges a.s. and in* L^1 *to some* $X \in L^1$*.*

Proof. The proof is very similar to the one of Theorems 1.9 and 1.10. First, for any $a < b$ the number of jumps from a to b is almost surely finite from the Doob's jumps inequality Theorem 1.8 :

$$\mathbb{E}(U(a,b)) = \lim_{n \to \infty} \mathbb{E}(U_n(a,b)) \leqslant \frac{1}{b-a} \mathbb{E}((X_0 - a)_+) < \infty,$$

where $\mathrm{U}_n(a,b)$ is the number of jumps between times $-n$ and 0. As a consequence, X_{-n} converges almost surely to some X, eventually equal to $\pm\infty$. As x_+ is convex and nondecreasing, Fatou's lemma and Proposition 1.2 yield[2]

$$\mathbb{E}(\mathrm{X}_+) \leqslant \liminf_{n\to\infty} \mathbb{E}((\mathrm{X}_{-n})_+) \leqslant \mathbb{E}((\mathrm{X}_0)_+) < \infty.$$

In the same way one can consider $-\mathrm{X}$, giving $\mathrm{X} \in \mathrm{L}^1$, so in particular X is finite. The last step consists in proving the convergence in L^1. Copying the proof of Theorem 1.10, we just need to know that the family of random variables $(\mathrm{X}_{-n}, n \geqslant 0)$ is uniformly integrable. This is automatic because $\mathrm{X}_{-n} = \mathbb{E}(\mathrm{X}_0 \mid \mathcal{F}_{-n})$. □

2. The only difference with the proof of Theorem 1.9 is that here the condition $\sup_n \mathbb{E}((\mathrm{X}_{-n})_+) < \infty$ is automatically true thanks to the negative indexation

Chapter 2

Brownian motion

After reminding some elementary facts about Gaussian random variables, we will be ready for constructing one specific random trajectory, the Brownian motion, and study its basic properties. At the end of the chapter, it is shown to be universal as a continuous limit of random walks. In the next two chapters, it will be shown to be universal amongst continuous martingales.

1. Gaussian vectors

This section is a reminder about properties of Gaussian vectors useful in the following. First, a random variable X is said to be Gaussian with expectation μ and variance σ^2 ($\mathrm{X} \sim \mathcal{N}(\mu, \sigma)$) if its law has density

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

with respect to the Lebesgue measure. By convention, the Dirac measure δ_μ is Gaussian with distribution $\mathcal{N}(\mu, 0)$. Hence, if $\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2)$, $\mathrm{X} \overset{\text{law}}{=} \mu + \sigma \mathrm{Y}$ where $\mathrm{Y} \sim \mathcal{N}(0, 1)$.

Theorem 2.1. *If* $\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2)$*, for any* $t \in \mathbb{R}$

$$\mathbb{E}(e^{\mathrm{i}t\mathrm{X}}) = e^{\mathrm{i}t\mu - \frac{(\sigma t)^2}{2}}$$

If X_1 *and* X_2 *are independent Gaussian random variables* ($\mathrm{X}_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \mathrm{X}_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$)*,* $\mathrm{X}_1 + \mathrm{X}_2$ *is Gaussian with expectation* $\mu_1 + \mu_2$ *and variance* $\sigma_1^2 + \sigma_2^2$*.*

Proof. To prove the form of the characteristic function, as $\mathrm{X} \overset{\text{law}}{=} \mu + \sigma \mathrm{Y}$ where $\mathrm{Y} \sim \mathcal{N}(0, 1)$, we only need to prove that

$$f(t) := \mathbb{E}(e^{\mathrm{i}t\mathrm{Y}}) = e^{-\frac{t^2}{2}}.$$

Derivation into the integral sign is clearly allowed, and a subsequent integration by parts yields

$$f'(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{i}x e^{\mathrm{i}tx} e^{-\frac{x^2}{2}} = -t f(t).$$

As $f(0) = 1$, the unique solution is $f(t) = e^{-\frac{t^2}{2}}$.

Now, for X_1 and X_2 as in the hypothesis, by independence

$$\mathbb{E}\left(e^{\mathrm{i}t(\mathrm{X}_1 + \mathrm{X}_2)}\right) = \mathbb{E}\left(e^{\mathrm{i}t\mathrm{X}_1}\right) \mathbb{E}\left(e^{\mathrm{i}t\mathrm{X}_2}\right) = e^{\mathrm{i}t(\mu_1 + \mu_2) - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}}.$$

As the characteristic function uniquely determines the law, this means that $\mathrm{X}_1 + \mathrm{X}_2$ is Gaussian with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. □

The rigidity of the family of Gaussian measures implies that it is stable by convergence in law, and that convergence in probability implies L^p convergence.

Theorem 2.2. *Let* $\mathrm{X}_n \sim \mathcal{N}(\lambda_n, \sigma_n^2)$*.*

(i) If $X_n \xrightarrow{\text{law}} X$*, then* λ_n *(resp.* σ_n^2*) converges to some* $\lambda \in \mathbb{R}$ *(resp.* $\sigma^2 \in \mathbb{R}_+$*) and* $X \sim \mathcal{N}(\lambda, \sigma^2)$.

(ii) If $X_n \xrightarrow{P} X$ *then, for any* $p > 1$, $X_n \xrightarrow{L^p} X$.

Proof. Concerning (i), we know that the convergence in law implies the uniform convergence of characteristic functions on compact sets, i.e.

$$\mathbb{E}\left(e^{itX_n}\right) = e^{it\mu_n - \frac{\sigma_n^2 t^2}{2}} \underset{n\to\infty}{\longrightarrow} \mathbb{E}\left(e^{itX}\right).$$

Taking the modulus, σ_n^2 converges to some $\sigma^2 \in \mathbb{R}_+$ ($\sigma^2 = \infty$ is not possible as $\mathbb{E}(e^{itX})$ would be discontinuous, $\mathbb{1}_{t=0}$). If $(\mu_n, n \geqslant 0)$ is bounded, this implies that $\mu_n \to \mu \in \mathbb{R}$ because for any accumulation points μ, μ', for any t, $e^{it\mu} = e^{it\mu'}$, so $\mu = \mu'$, and by convergence of the characteristic functions X_n converges in law to Gaussian $\mathcal{N}(\mu, \sigma^2)$.

The unbounded case is impossible : if for example a subsequence $\mu_{n_k} \to \infty$, then as the variables are Gaussian and σ_n^2 is bounded $\mathbb{P}(X_{n_k} > \lambda) \to 1$ for any λ, so by weak convergence $\mathbb{P}(X > \lambda) = 1$, a contradiction when λ is large enough.

For (ii), first note that we just proved that μ_n and σ_n^2 are bounded. Hence we have $\sup_n \mathbb{E}(|X_n|^{2p}) < \infty$ for any $p > 0$. By Fatou's lemma, as X_{m_k} converges almost surely to X along a subsequence, this implies that $\mathbb{E}(|X|^{2p}) < \infty$, and therefore $\sup_n \mathbb{E}(|X_n - X|^{2p}) < \infty$. Hence the sequence $(X_n - X, n \geqslant 0)$ is bounded in L^2 and converges to 0 in probability, hence it converges to 0 in L^1, as expected. □

We now come to a multidimensional natural generalization : a random variable X with values in $\mathbb{R}^d$ is called a Gaussian vector if any linear combination of the coordinates is a Gaussian random variable : for any $u \in \mathbb{R}^d$, $u \cdot X$ is normally distributed.

For example, if $X_1, \dots, X_d$ are independent Gaussians, $X = (X_1, \dots, X_d)$ is a Gaussian vector. Under the same hypothesis, if M is a $d' \times d$, MX is still a Gaussian vector with values in $\mathbb{R}^{d'}$. Note that there are vectors with Gaussian entries which are not Gaussian vectors. For example, if $X \sim \mathcal{N}(0,1)$ and ε is an independent random variable, $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$, then $(X, \varepsilon X)$ has Gaussian entries but is not a Gaussian vector : the sum of its coordinates has probability 1/2 to be 0.

Theorem 2.3. *Let* X *be a Gaussian vector. Then its entries are independent if and only if its covariance matrix* $(\mathrm{cov}(X_i, X_j))_{1 \leqslant i,j \leqslant d}$ *is diagonal.*

Proof. If the entries are independent, the covariance matrix is obviously diagonal. Reciprocally, if the matrix is diagonal, for any $u \in \mathbb{R}^d$

$$\begin{aligned} \mathbb{E}(u \cdot X) &= \sum u_i \, \mathbb{E}(X_i) \\ \mathrm{var}(u \cdot X) &= \sum_{j,k} u_j u_k \, \mathrm{cov}(X_i, X_j) = \sum_j u_j^2 \, \mathrm{var}(X_j) \end{aligned}$$

Hence the Gaussian random variable $u \cdot X$ has characteristic function

$$\mathbb{E}(e^{iu\cdot X}) = e^{i\,\mathbb{E}(u\cdot X) - \frac{\mathrm{var}(u\cdot X)}{2}} = \prod_{k=1}^{d} e^{iu_k\,\mathbb{E}(X_k) - \frac{\mathrm{var}(X_k)}{2}} = \prod_{k=1}^{n} \mathbb{E}(e^{iu_k X_k}).$$

As a consequence of this splitting of the characteristic function, the X_k's are independent. □

Finally, in this section we want to give a useful estimate of the queuing distribution of Gaussian random variables.

Lemma 2.4. *Let* $X \sim \mathcal{N}(0,1)$*. Then, as* $\lambda \to \infty$,

$$\mathbb{P}(X > \lambda) \sim \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}}$$

Proof. By integration by parts,

$$\mathbb{P}(\mathrm{X} > \lambda) = \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty \frac{-1}{x}(-x)e^{-\frac{x^2}{2}} = \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} - \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty \frac{e^{-x^2/2}}{x^2} \mathrm{d}x.$$

Obviously,

$$\frac{1}{\sqrt{2\pi}} \int_\lambda^\infty \frac{e^{-x^2/2}}{x^2} \mathrm{d}x \leqslant \frac{1}{\lambda^2\sqrt{2\pi}} \int_\lambda^\infty e^{-x^2/2} \mathrm{d}x = \mathrm{o}\left(\mathbb{P}(\mathrm{X} > \lambda)\right),$$

concluding the proof. □

2. Existence of Brownian motion

Definition 2.5. $(\mathrm{X}_t, t \geqslant 0)$ *is a Gaussian process if for any* $(t_1, \dots, t_n) \in \mathbb{R}_+^n$, $(\mathrm{X}_{t_1}, \dots, \mathrm{X}_{t_n})$ *is a Gaussian vector.*

Definition 2.6. *A process* $\mathrm{B} = (\mathrm{B}_t, t \geqslant 0)$ *is called a Brownian motion if :*

(i) it is a Gaussian process ;

(ii) it is centered : for any $t \geqslant 0$, $\mathbb{E}(\mathrm{B}_t) = 0$ *;*

(iii) for any $(s, t) \in \mathbb{R}_+^2$, $\mathbb{E}(\mathrm{B}_s \mathrm{B}_t) = s \wedge t$ *;*

(iv) it is almost surely continuous.

Note that this definition implies $\mathrm{B}_0 = 0$ almost surely. One refer sometimes to this process B as a *standard* Brownian motion, in contrast to $x + \mathrm{B}$ which is called a Brownian motion starting at x. Some variants of the definition also englobe the finite horizon possibility, refering to a Brownian motion on $[0, \mathrm{T}]$.

The above definition is equivalent to another one, with independence of the increments.

Proposition 2.7. *The process* $(\mathrm{B}_t, t \geqslant 0)$ *is a Brownian motion if and only if*

(a) $\mathrm{B}_0 = 0$ *almost surely ;*

(b) for any $n \geqslant 2$, $0 \leqslant t_1 \leqslant \dots \leqslant t_n$, $(\mathrm{B}_{t_1}, \mathrm{B}_{t_2} - \mathrm{B}_{t_1}, \dots, \mathrm{B}_{t_n} - \mathrm{B}_{t_{n-1}})$ *is a Gaussian vector with independent coordinates.*

(c) for any $(s, t) \in \mathbb{R}_+^2$, $\mathrm{B}_t - \mathrm{B}_s \sim \mathcal{N}(0, |t - s|)$ *;*

(d) it is almost surely continuous.

Proof. Let us first prove that $(i), (ii), (iii)$ implies $(a), (b), (c)$. By (iii), $\mathrm{B}_0 = 0$ a.s. which is (a). Moreover, $\mathrm{X} = (\mathrm{B}_{t_1}, \mathrm{B}_{t_2} - \mathrm{B}_{t_1}, \dots, \mathrm{B}_{t_n} - \mathrm{B}_{t_{n-1}})$ is clearly a Gaussian vector, because by (i), $(\mathrm{B}_{t_1}, \mathrm{B}_{t_2}, \dots, \mathrm{B}_{t_n})$ is a Gaussian vector. To prove that the coordinates are independent, we just need to prove that the covariance matrix is diagonal. This is a direct calculation, using (iii) : for $t_1 \leqslant t_2 \leqslant t_3 \leqslant t_4$,

$$\mathrm{cov}(\mathrm{B}_{t_4} - \mathrm{B}_{t_3}, \mathrm{B}_{t_2} - \mathrm{B}_{t_1}) = \mathbb{E}((\mathrm{B}_{t_4} - \mathrm{B}_{t_3})(\mathrm{B}_{t_2} - \mathrm{B}_{t_1})) = t_4 \wedge t_2 - t_4 \wedge t_1 - t_3 \wedge t_2 + t_3 \wedge t_1 = 0.$$

By (i), $\mathrm{B}_t - \mathrm{B}_s$ is a Gaussian random variable, with expectation $\mathrm{E}(\mathrm{B}_t) - \mathbb{E}(\mathrm{B}_s) = 0$ by (ii) and variance

$$\mathbb{E}((\mathrm{B}_t - \mathrm{B}_s)^2) = \mathbb{E}(\mathrm{B}_t^2) + \mathbb{E}(\mathrm{B}_s^2) - 2\,\mathbb{E}(\mathrm{B}_t \mathrm{B}_s) = ts - 2s \wedge t = |t - s|,$$

by (iii), proving (c).

Now, assume points $(a), (b), (c)$. For any $t_1 \leqslant \dots \leqslant t_n$, as $(\mathrm{B}_{t_1}, \mathrm{B}_{t_2} - \mathrm{B}_{t_1}, \dots, \mathrm{B}_{t_n} - \mathrm{B}_{t_{n-1}})$ is Gaussian by (b), so is $(\mathrm{B}_{t_1}, \mathrm{B}_{t_2}, \dots, \mathrm{B}_{t_n})$, because $(x_1, \dots, x_n) \mapsto (x_1, x_2 -$

$x_1, \dots, x_n - x_{n-1})$ is invertible. This proves (i). Point (ii) follows from $\mathrm{B}_t \sim \mathcal{N}(0,t)$, by (a) and (c). Finally, using the independence of the increments in (b), the previously proved fact that B is centered, and the variance of B given by (c), for $t \geqslant s$

$$\mathbb{E}(\mathrm{B}_t\mathrm{B}_s) = \mathbb{E}(\mathrm{B}_s^2) + \mathbb{E}((\mathrm{B}_t - \mathrm{B}_s)\mathrm{B}_s) = \mathbb{E}(\mathrm{B}_s^2) + \mathbb{E}((\mathrm{B}_t - \mathrm{B}_s)\mathrm{B}_s) = s = s \wedge t,$$

which proves (iii). □

In the following, we give two constructions of Brownian motion. The first one, by Paul Lévy, is intuitive and proceeds by almost sure uniform convergence of properly chosen piecewise affine gaussian processes. Then a more functional analytic proof will be discussed, originating in the work of Wiener and substantially generalized by Itô and Nisio. Both proofs exhibit a Brownian motion on $[0,1]$, extending to the existence in $\mathbb{R}_+$ by simple juxtaposition of independent ones on $[0,1]$.

Theorem 2.8. *The Brownian motion exists.*

Proof. We proceed by induction to construct piecewise affine Gaussian processes converging uniformly. First, take a Gaussian random variable $\mathrm{N}_{0,0} \sim \mathcal{N}(0,1)$, and define $f_0(t) = \mathrm{N}_{0,0}t,\ 0 \leqslant t \leqslant 1$. This is a Gaussian process with the same distribution as Brownian motion at time 1, but not on $(0,1)$. To avoid this problem, define another $\mathcal{N}(0,1)$ random variable $\mathrm{N}_{1,1}$ and the Gaussian process f_1, piecewise affine, continuous, such that f_1 and f_0 coincide on 0 and 1, but $f_1(1/2) = f_0(1/2) + \frac{1}{2}\mathrm{N}_{1,1}$. Now the process f_1 has the expected covariance function on the set of points $\{0, 1/2, 1\}$. We proceed in the same way on further intervals, defining for any $n \geqslant 1$ the continuous function f_n, affine on $[(k-1)/2^n, k/2^n]$ for any $k \in [\![1, 2^n]\!]$, by

$$\begin{cases} f_n\left(\frac{2\ell}{2^n}\right) & = f_{n-1}\left(\frac{\ell}{2^{n-1}}\right) \\ f_n\left(\frac{2\ell-1}{2^n}\right) & = f_{n-1}\left(\frac{2\ell-1}{2^{n-1}}\right) + 2^{-\frac{n+1}{2}}\mathrm{N}_{\ell,n} \end{cases},$$

where the $\mathrm{N}_{n,\ell},\ n \geqslant 0,\ 1 \leqslant \ell \leqslant 2^{n-1}$ are independent standard Gaussians. Then an immediate induction proves that for any $n \geqslant 1$, $(f_n(t), 0 \leqslant t \leqslant 1)$ is a centered Gaussian process, and with covariance function

$$\mathbb{E}\left(f_n\left(\frac{j}{2^n}\right) f_n\left(\frac{k}{2^n}\right)\right) = \frac{j}{2^n} \wedge \frac{k}{2^n}, \tag{2.1}$$

the normalization $2^{-\frac{n+1}{2}}$ in the definition of f_n being chosen to get the above appropriate covariance.

We are now interested in the uniform convergence of $f_n, n \geqslant 0$. Let

$$\mathrm{A}_n = \{\sup_{[0,1]} |f_n - f_{n-1}|(t) > 2^{-n/4}\}.$$

As $|f_n - f_{n-1}|$ is continuous affine on $[(k-1)/2^n, k/2^n]$ for any $k \in [\![1, 2^n]\!]$, and 0 if k is even, its maximal value is obtained at some $t \in \{(2\ell-1)/2^n, 1 \leqslant \ell \leqslant 2^{n-1}\}$, hence

$$\begin{aligned} \mathbb{P}(\mathrm{A}_n) &= \mathbb{P}\left(\sup_{1 \leqslant \ell \leqslant 2^{n-1}} |2^{-\frac{n+1}{2}}\mathrm{N}_{\ell,n}| > 2^{-n/4}\right) \\ &\leqslant \sum_{\ell=1}^{2^{n-1}} \mathbb{P}\left(|\mathrm{N}_{\ell,n}| > 2^{-n/4+\frac{n+1}{2}}\right) \\ \mathbb{P}(\mathrm{A}_n) &\leqslant 2^{c_1 n} e^{-c_2 2^{n/2}} \end{aligned}$$

for some absolute constants $c_1, c_2 > 0$, where we used Lemma 2.4. Hence $\sum_n \mathbb{P}(\mathrm{A}_n) < \infty$, which means by the Borel-Cantelli lemma that almost surely, there is an index

$n_0(\omega)$ such that if $n \geqslant n_0(\omega)$ then $\|f_n - f_{n-1}\|_\infty \leqslant 2^{-n/4}$. This implies the almost sure uniform convergence of f_n on $[0,1]$, to a random function called f.

This process is almost surely continuous, as the a.s. uniform limit of continuous functions, and this is Gaussian process because the a.s. limit of Gaussian vectors is a Gaussian vector, by Theorem 2.2.

Our only remaining point to prove is

$$\mathbb{E}(f(s)f(t)) = s \wedge t. \tag{2.2}$$

Let $\ell_n^{(s)}$, $\ell_n^{(t)}$ be integers such that

$$\frac{\ell_n^{(s)}}{2^n} \leqslant s < \frac{\ell_n^{(s)}+1}{2^n}, \ \frac{\ell_n^{(t)}}{2^n} \leqslant t < \frac{\ell_n^{(t)}+1}{2^n}.$$

From the uniform convergence of f_n, the Gaussian vector $(f_n(\ell_n^{(s)}/2^n), f_n(\ell_n^{(t)}/2^n))$ converges almost surely to $(f(s), f(t))$, hence in L^2 by Theorem 2.2. This implies

$$\mathbb{E}(f_n(\ell_n^{(s)}/2^n) f_n(\ell_n^{(t)}/2^n)) \to \mathbb{E}(f(s)f(t)),$$

and this first term also obviously converges to $s \wedge t$ by (2.1), proving (2.2). □

We note that there is only one Brownian motion : the subsets of type

$$\{\omega : \omega(t_1) \in \mathrm{A}_1, \dots, \omega(t_m) \in \mathrm{A}_m\},$$

where the A_m's are Borel subsets of $\mathbb{R}$, have a unique possible measure given the definition of Brownian motion. As they are stable by intersection and generate the sigma algebra $\mathcal{F} = \sigma(\omega(s), 0 \leqslant s \leqslant 1)$, by the monotone class theorem there is only one measure on $\mathcal{F}$ corresponding to a Brownian motion. The Wiener measure, noted W, is the image of the Gaussian product measure (for the N_k's) on $\mathscr{C}^o([0,1])$ in the above construction.

For another construction of Brownian motion, we need the following theorem by Itô and Nisio [6], where E is a real separable Banach space, with its norm topology, E^* its dual and $\mathscr{B}$ its Borel algebra.

Theorem 2.9. *Let $(\mathrm{X}_i)_{i \geqslant 0}$ be independent* E*-valued random variables, and $\mathrm{S}_n = \sum_1^n \mathrm{X}_i$, with law μ_n. Then the following three conditions are equivalent :*

(a) S_n converges almost surely ;

(b) S_n converges in probability ;

(c) μ_n converges for the Prokhorov metric[1].

Moreover, if the X_i's have a symmetric distribution ($\mathrm{X}_i \overset{\text{law}}{=} -\mathrm{X}_i$), then each of the above conditions is equivalent to any of the following ones :

(d) μ_n is uniformly tight ;

(e) there is a E*-valued* S *such that for any $z \in \mathrm{E}^*$, $\langle z, \mathrm{S}_n\rangle \xrightarrow{\mathrm{P}} \langle z, \mathrm{S}\rangle$;*

(f) there is a measure μ on E *such that for any $z \in \mathrm{E}^*$, $\mathbb{E}\left(e^{\mathrm{i}\langle z, \mathrm{S}_n\rangle}\right) \to \int e^{\mathrm{i}\langle z, x\rangle} \mathrm{d}\mu(x)$.*

1. The Prokhoov distance between two measures μ and ν is

$$\pi(\mu,\nu) = \inf\{\varepsilon > 0 : \forall \mathrm{A} \in \mathscr{B}, \mu(\mathrm{A}) - \nu(\mathrm{A}_\varepsilon) \leqslant \varepsilon, \mu(\mathrm{A}^\varepsilon) - \nu(\mathrm{A}) \leqslant \varepsilon\},$$

where A_ε is the set of points within distance at most ε to A.

In finite dimension, the symmetry condition for points $(d), (e), (f)$ is not necessary, but it is essential in infinite dimension, as shown by the following example : if $(e_i)_{i\geqslant 1}$ is an orthonormal basis of a Hilbert space, and $\mathrm{X}_1(\omega) = e_1$, $\mathrm{X}_n(\omega) = e_n - e_{n-1}$ for $n \geqslant 2$, then $\mathrm{S}_n(\omega) = e_n$, so $\langle z, \mathrm{S}_n\rangle \to 0$ but S_n does not converge almost surely.

The application of the previous general result to Brownian motion is for $\mathrm{E} = (\mathscr{C}([0,1]), \|\|_\infty)$, the space of continuous functions on $[0,1]$ endowed with the sup norm.

Theorem 2.10. *Let $(\varphi_n)_{n\geqslant 1}$ be an orthonormal basis of $\mathrm{L}^2([0,1])$, composed of continuous functions, and $(\mathrm{N}_k)_{k\geqslant 1}$ a sequence of independent standard normal variables. Then*

$$\mathrm{S}_n(t) = \sum_{k=1}^n \mathrm{N}_k \int_0^t \varphi_k(u)\mathrm{d}u$$

converges uniformly in $t \in [0,1]$ as $n \to \infty$. The limit S *is a Brownian motion.*

Proof. Elements in E^* can be identified with bounded signed finitely additive measures on $[0,1]$ that are absolutely continuous with respect to the Lebesgue measure on [0,1], see [4].

Let $\mathrm{d}z$ be such an element : we first prove point (f) in Theorem 2.9, by writing

$$\begin{aligned}\mathbb{E}(e^{\mathrm{i}\langle z,\mathrm{S}_n\rangle}) &= \prod_{k=1}^n \mathbb{E}\left(e^{\mathrm{i}\mathrm{N}_k\langle z,\int_0^\cdot \varphi_k\rangle}\right) = \prod_{k=1}^n e^{-\frac{1}{2}|\langle z,\int_0^\cdot \varphi_k\rangle|^2}\\ &= e^{-\frac{1}{2}\sum_{k=1}^n\left(\int_0^1 \mathrm{d}z(u)\int_0^u \varphi_k(s)\mathrm{d}s\right)^2} = e^{-\frac{1}{2}\sum_{k=1}^n\left(\int_0^1 \varphi_k(s)\mathrm{d}s z([s,1])\right)^2}\\ &\underset{n\to\infty}{\longrightarrow} e^{-\frac{1}{2}\int_0^1 \mathrm{d}s z([s,1])^2} = e^{-\frac{1}{2}\int_{[0,1]^2}(s\wedge t)\mathrm{d}z(s)\mathrm{d}z(t)},\end{aligned}$$

where we used the orthonomality of the φ_k's between the second and third lines.

Now, if B is a Brownian motion, as proved by approximations by Riemann sums, $\langle z, \mathrm{B}\rangle$ is a Gaussian random variable with variance $\int_{[0,1]^2}(s\wedge t)\mathrm{d}z(s)\mathrm{d}z(t)$. Consequently, $\mathbb{E}(e^{\mathrm{i}\langle z,\mathrm{S}_n\rangle})$ converges to $\int e^{\mathrm{i}\langle z,x\rangle}\mathrm{dW}(x)$ where W is the Wiener measure. As $\mathrm{X}_k = (\mathrm{N}_k\int_0^t \varphi_k(u)\mathrm{d}u, 0 \leqslant t \leqslant 1)$ is symmetric, by Theorem 2.9 S_n converges almost surely uniformly on $[0,1]$.

The limit is a Brownian motion : all the characteristic properties of B (independent Gaussian increments) are proved by choosing $z = \sum_{k=1}^m \delta_{t_k}$ in the above calculation. □

3. Invariance properties

One of the useful features of Brownian motion is the invariance of the Wiener measure under many transformations, i.e. symmetry, time reversal, time inversion and scaling.

Theorem 2.11. *Let $(\mathrm{B}_t, t \geqslant 0)$ be a Brownian motion. Then the following processes are Brownian motions :*

(i) $\mathrm{X}_t = -\mathrm{B}_t, t \geqslant 0$;

(ii) $\mathrm{X}_t = \mathrm{B}_\mathrm{T} - \mathrm{B}_{\mathrm{T}-t}$, $0 \leqslant t \leqslant \mathrm{T}$;

(iii) $\mathrm{X}_t = t\mathrm{B}_{1/t}$ *if* $t > 0$ *and* 0 *it* $t = 0$;

(iv) $\mathrm{X}_t = \frac{1}{\sqrt{\lambda}}\mathrm{B}_{\lambda t}$, $t \geqslant 0$, *for some given parameter* $\lambda > 0$.

Proof. It is an easy task to prove that all of these processes are centered Gaussian processes with covariance function $\mathbb{E}(\mathrm{X}_s\mathrm{X}_t) = s \wedge t$. The only problem consists in proving the almost sure continuity of X at 0 in the case *(iii)*.

As $(\mathrm{B}_s, s>0)$ has the same law as $(s\mathrm{B}_{1/s}, s>0)$, for any $\varepsilon>0$

$$\mathbb{P}\left(\cap_n \cup_{t\in\mathbb{Q}\cap(0,1/n]}\left\{|t\mathrm{B}_{1/t}|>\varepsilon\right\}\right)=\mathbb{P}\left(\cap_n \cup_{t\in\mathbb{Q}\cap(0,1/n]}\left\{|\mathrm{B}_t|>\varepsilon\right\}\right).$$

This last term is 0 because almost surely $\mathrm{B}_0=0$ and B is continuous. Hence $s\mathrm{B}_{1/s}\to 0$ along positive rational numbers, hence along $\mathbb{R}_+$ by continuity. □

Note that as $\mathrm{B}_t\to 0$ almost surely as $t\to 0^+$, point *(iii)* implies that $\frac{\mathrm{B}_t}{t}\to 0$ as $t\to\infty$. One can directly show that

$$\mathrm{M}_t := e^{\mathrm{B}_t-\frac{t}{2}}$$

has constant expectation, and thanks to the above property it converges almost surely to 0 : this is an example of almost sure convergence but no convergence in L^1.

4. Regularity of trajectories

In this section, we are interested in the Hölder exponent of the trajectories of B, i.e. the values of α such that there is almost surely a constant $c_\alpha(\omega)$ such that, for any s and t in $[0,1]$,

$$|\mathrm{B}_t-\mathrm{B}_s|\leqslant c_\alpha(\omega)|t-s|^\alpha.$$

This set of possible values for α is of type $[0,\alpha_0]$ or $[0,\alpha_0)$, and in the next section we will give the precise value of α_0.

First, before giving the important criterium by Kolmogorov for Hölder-continuity, we need to introduce the following notions.

Definition 2.12. *Let $(\mathrm{X}_t)_{t\in\mathrm{I}}$ and $(\tilde{\mathrm{X}}_t)_{t\in\mathrm{I}}$ be two processes.*

(i) X *is called a version of* $\tilde{\mathrm{X}}$ *if, for any $t\in\mathrm{I}$, $\mathbb{P}(\mathrm{X}_t=\tilde{\mathrm{X}}_t)=1$;*

(ii) X *and* $\tilde{\mathrm{X}}$ *are called indistinguishable if* $\mathbb{P}(\forall t\in\mathrm{I},\mathrm{X}_t=\tilde{\mathrm{X}}_t)=1$.

Is X and $\tilde{\mathrm{X}}$ are indistinguishable, each one is a version of the other. The converse is false, as shown by $\mathrm{X}(t)=0, 0\leqslant t\leqslant 1$ and $\tilde{\mathrm{X}}_t=0$ on $[0,1]$ except on U, uniform on $[0,1]$, where $\tilde{\mathrm{X}}$ is 1. Then on a given t both processes are almost surely equal but almost surely, they differ somewhere along the trajectory.

Note that if I is countable, or if X and $\tilde{\mathrm{X}}$ are almost surely continuous, then being a version of the other implies the indistinguishability.

Theorem 2.13. *Let* I *be a compact interval and* X *a process on* I. *We suppose that there exist p, ε and c positive numbers such that for any s,t in* I

$$\mathbb{E}\left(|\mathrm{X}_s-\mathrm{X}_t|^p\right)\leqslant c|t-s|^{1+\varepsilon}.$$

Then there exists a version $\tilde{\mathrm{X}}$ *of* X *whose trajectories are Hölderian with index α for any $\alpha\in[0,\varepsilon/p)$: almost surely, there is $c_\alpha(\omega)>0$ such that for any s,t in* I

$$|\tilde{\mathrm{X}}_t-\tilde{\mathrm{X}}_s|\leqslant c_\alpha(\omega)|t-s|^\alpha.$$

Proof. If $\alpha=0$ the result is obvious by compactness of the support, so we suppose $\alpha\in(0,\varepsilon/p)$ in the following. We can suppose $\mathrm{I}=[0,1]$, and we will first show the Hölder property for X on the set $\mathscr{D}$ of dyadic numbers, i.e. of type

$$\sum_{k=1}^m \frac{\varepsilon_k}{2^k}$$

for some $m\in\mathbb{N}^*$, and $\varepsilon_k=0$ or 1. We want to prove that for $\alpha\in(0,\varepsilon/p)$,

$$\sup_{s,t\in\mathscr{D},s\neq t}\frac{|\mathrm{X}_t-\mathrm{X}_s|}{|t-s|^\alpha}<\infty.$$

Let us first prove that

$$d_\alpha = \sup_{n\geqslant 0}\sup_{1\leqslant i\leqslant 2^n}\frac{|\mathrm{X}_{\frac{i}{2^n}}-\mathrm{X}_{\frac{i-1}{2^n}}|}{|1/2^n|^\alpha}<\infty \tag{2.3}$$

almost surely. For this, let $\mathrm{A}_i^{(n)}=\{|\mathrm{X}_{\frac{i}{2^n}}-\mathrm{X}_{\frac{i-1}{2^n}}|>2^{n\alpha}\}$. As

$$\mathbb{P}(|\mathrm{X}_t-\mathrm{X}_s|>a)\leqslant a^{-p}\,\mathbb{E}(|\mathrm{X}_t-\mathrm{X}_s|^p)\leqslant ca^{-p}|t-s|^{1+\varepsilon},$$

we have $\mathbb{P}(\mathrm{A}_i^{(n)})\leqslant c2^{(p\alpha-1-\varepsilon)n}$. As $p\alpha-\varepsilon<0$, this yields

$$\sum_{n\geqslant 0}\mathbb{P}\left(\cup_{i=1}^{2^n}\mathrm{A}_i^{(n)}\right)<\infty,$$

so by the Borel Cantelli lemma $d_\alpha<\infty$ almost surely.

We now want to extend the result to dyadic numbers. For this, for given $s<t$ in $\mathscr{D}$, let q be the smallest integer such that $2^{-q}<t-s$. We can write

$$\begin{cases} s &= k2^{-q}-\sum_{i=1}^{\ell}\varepsilon_i 2^{-q-i}\\ t &= k2^{-q}+\sum_{i=1}^{m}\varepsilon_i' 2^{-q-i}\end{cases}$$

for some integers k, ℓ and m, and the ε_i's, ε_i''s being 0 or 1. Define for $j\in[\![0,\ell]\!]$ and $0\in[\![1,m]\!]$,

$$\begin{cases} s_j &= k2^{-q}-\sum_{i=1}^{j}\varepsilon_i 2^{-q-i}\\ t_k &= k2^{-q}+\sum_{i=1}^{k}\varepsilon_i' 2^{-q-i}\end{cases}$$

Then, using (2.3) between successive s_j's and t_j's

$$\begin{aligned}|\mathrm{X}_t-\mathrm{X}_s| &\leqslant \sum_{j=1}^{\ell}|\mathrm{X}_{s_j}-\mathrm{X}_{s_{j-1}}|+\sum_{k=1}^{m}|\mathrm{X}_{t_k}-\mathrm{X}_{t_{k-1}}|\\ &\leqslant d_\alpha\left(\sum_{j=1}^{\ell}2^{-(q+j)\alpha}+\sum_{k=1}^{m}2^{-(q+k)\alpha}\right)\\ &\leqslant 2d_\alpha\frac{1}{1-2^{-\alpha}}2^{-p\alpha}\\ |\mathrm{X}_t-\mathrm{X}_s| &\leqslant c_\alpha|t-s|^\alpha,\end{aligned}$$

proving that X is Hölderian with exponent α on dyadic numbers. Define $\tilde{\mathrm{X}}$ as 0 if $d_\alpha(\omega)=\infty$ (event with measure 0) and the continuous extension of X from $\mathscr{D}$ to $[0,1]$ otherwise, which exists and is unique thanks to the previous results : $\tilde{\mathrm{X}}$ is still Hölderian for the exponent α, and we just need to prove that it is a version of X. For a given $t\in[0,1]$, consider a sequence $(s_n)_{n\geqslant 0}$ of diadic numbers converging to t. From the hypothesis, X_{s_n} converges to X_t in L^p, hence in probability, hence almost surely along a subsequence. Moreover, along that subsequence, X_{s_n} converges to $\tilde{\mathrm{X}}_t$ by definition of $\tilde{\mathrm{X}}$. Consequently, almost surely both limits coincide, i.e. $\mathbb{P}(\mathrm{X}_t=\tilde{\mathrm{X}}_t)=1$. □

Corollary 2.14. *Let* $\mathrm{B}=(\mathrm{B}_t,0\leqslant t\leqslant 1)$ *be a Browian motion. Then* B *is almost surely Hölderian with exponent* α *for any* $\alpha\in[0,1/2)$.

Proof. For any s and t,, if $\mathrm{X}\sim\mathcal{N}(0,1)$,

$$\mathbb{E}\left(|\mathrm{B}_t-\mathrm{B}_s|^p\right)=\mathbb{E}\left(|\mathrm{X}|^p\right)|t-s|^{p/2}.$$

For $p>2$, if $\varepsilon=\frac{p}{2}-1$ in Theorem 2.13, we see that B has a Hölderian version with exponent α for any $\alpha\in\left[0,\frac{1}{2}-\frac{1}{p}\right)$. As B and its version are almost surely continuous, they are indistinguishable and taking $p\to\infty$ previously proves that B is almost surely Hölderian for any exponent $\alpha\in[0,1/2)$. □

Section 6 will prove that the set of values of α for which a Brownian trajectory is almost surely Hölderian is exactly $[0,1/2)$.

5. Markov properties

Let $\mathcal{F}_t = \sigma(\mathrm{B}_s, 0 \leqslant s \leqslant t)$, i.e. $\mathcal{F}_t$ is the smallest σ-algebra in Ω such that all $\mathrm{B}_s : \Omega \to \mathbb{R}$, $0 \leqslant s \leqslant t$, are measurable.

Then the weak Markov property states that, writing $\tilde{\mathrm{B}}_t = \mathrm{B}_{t+s} - \mathrm{B}_s$, $\tilde{\mathrm{B}}$ is a Brownian motion independent of $\mathcal{F}_s$. Being a Brownian motion is a straightforward consequence of the definition. To get the independence property, by the monotone class Theorem, it is sufficient to prove that for $0 \leqslant t_1 < \cdots < t_m$, $0 \leqslant s_1 < \cdots < s_m \leqslant s$, the vectors $\tilde{b} = (\tilde{\mathrm{B}}_{t_1}, \dots, \tilde{\mathrm{B}}_{t_n})$ and $b = (\mathrm{B}_{s_1}, \dots, \mathrm{B}_{s_m})$ are independent. Note that

$$\operatorname{cov}(\tilde{\mathrm{B}}_{t_j}, \mathrm{B}_{s_k}) = \mathbb{E}((\mathrm{B}_{s+t_j} - \mathrm{B}_s)\mathrm{B}_{s_k}) = \mathbb{E}(\mathrm{B}_{s+t_j} - \mathrm{B}_s)\,\mathbb{E}(\mathrm{B}_{s_k}) = 0,$$

thanks to the independence of increments of the Brownian motion and $s_k \leqslant s \leqslant s+t_j$. Hence for any $\lambda = (\lambda_1, \dots, \lambda_m)$ and $\mu = (\mu_1, \dots, \mu_n)$, the Gaussian vector $(\lambda \cdot \tilde{b}, \mu \cdot b)$ has a diagonal covariance matrix, and by Theorem 2.3 $\lambda \cdot \tilde{b}$ and $\mu \cdot b$ are independent. This being true for any λ and μ, b and $\tilde{b}$ are independent, as expected.

The above result can be extended to a σ-algebra potentially bigger than $\mathcal{F}_s$, namely

$$\mathcal{F}_s^+ = \cap_{t>s} \mathcal{F}_t.$$

Theorem 2.15 (Weak Markov property)**.** *Let* $\tilde{\mathrm{B}}_t = \mathrm{B}_{t+s} - \mathrm{B}_s$*. Then* $\tilde{\mathrm{B}}$ *is a Brownian motion independent of* $\mathcal{F}_s^+$

Proof. Proving the independence property requires some work. By the monotone class Theorem, a sufficient condition is that for any $\mathrm{A} \in \mathcal{F}_s^+$, $n \geqslant 0$, $0 \leqslant t_1 < \cdots < t_n$, and $\mathrm{F} : \mathbb{R}^n \to \mathbb{R}$ bounded and continuous,

$$\mathbb{E}\left(\mathrm{F}(\tilde{\mathrm{B}}_{t_1}, \dots, \tilde{\mathrm{B}}_{t_n})\mathbb{1}_{\mathrm{A}}\right) = \mathbb{E}\left(\mathrm{F}(\tilde{\mathrm{B}}_{t_1}, \dots, \tilde{\mathrm{B}}_{t_n})\right)\mathbb{E}\left(\mathbb{1}_{\mathrm{A}}\right). \tag{2.4}$$

For any $\varepsilon > 0$, as $(\mathrm{B}_{t+s+\varepsilon} - \mathrm{B}_{s+\varepsilon}, t \geqslant 0)$ is a Brownian motion independent of $\mathcal{F}_{s+\varepsilon}$ and $\mathrm{A} \in \mathcal{F}_{s+\varepsilon}$,

$$\begin{aligned}&\mathbb{E}\left(\mathrm{F}(\mathrm{B}_{t_1+s+\varepsilon} - \mathrm{B}_{s+\varepsilon}, \dots, \mathrm{B}_{t_n+s+\varepsilon} - \mathrm{B}_{s+\varepsilon})\mathbb{1}_{\mathrm{A}}\right)\\ &\qquad\qquad = \mathbb{E}\left(\mathrm{F}(\mathrm{B}_{t_1+s+\varepsilon} - \mathrm{B}_{s+\varepsilon}, \dots, \mathrm{B}_{t_n+s+\varepsilon} - \mathrm{B}_{s+\varepsilon})\right)\mathbb{E}\left(\mathbb{1}_{\mathrm{A}}\right).\end{aligned}$$

By dominated convergence, $\varepsilon \to 0$ in the above equation yields the result (2.4). □

As a consequence of the weak Markov property, the σ-algebra $\mathcal{F}_0^+$ is trivial : there is no information in the *germ* of the Brownian motion.

Theorem 2.16 (Blumenthal's 0-1 law)**.** *For any* $\mathrm{A} \in \mathcal{F}_0^+$, $\mathbb{P}(\mathrm{A}) = 0$ *or* $\mathbb{P}(\mathrm{A}) = 1$.

Proof. From Theorem 2.15, A is independent of $(\mathrm{B}_t - \mathrm{B}_0, t \geqslant 0)$, hence independent of $\sigma(\mathrm{B}_s, s \geqslant 0)$. Moreover, $\mathrm{A} \in \mathcal{F}_0^+ \subset \sigma(\mathrm{B}_s, s \geqslant 0)$, hence A is independent of itself :

$$\mathbb{P}(\mathrm{A}) = \mathbb{E}(\mathbb{1}_{\mathrm{A}}\mathbb{1}_{\mathrm{A}}) = \mathbb{E}(\mathbb{1}_{\mathrm{A}})\,\mathbb{E}(\mathbb{1}_{\mathrm{A}}) = \mathbb{P}(\mathrm{A})^2,$$

so $\mathbb{P}(\mathrm{A}) = 0$ or $\mathbb{P}(\mathrm{A}) = 1$. □

Corollary 2.17.

(i) Let $\tau_1 = \inf\{t > 0 : \mathrm{B}_t > 0\}$*. Then* $\tau_1 = 0$ *almost surely.*

(ii) Let $\tau_2 = \inf\{t > 0 : \mathrm{B}_t = 0\}$*. Then* $\tau_2 = 0$ *almost surely.*

(iii) Let $\tau_3 = \sup\{t > 0 : \mathrm{B}_t = 0\}$. *Then* $\tau_3 = \infty$ *almost surely.*

Proof. First note that

$$\{\tau_1 = 0\} = \cap_{n\geqslant 1}\{\sup_{[0,1/n]} \mathrm{B}_s > 0\} \in \cap_{t>0}\mathcal{F}_t = \mathcal{F}_0+,$$

so, from Theorem 2.16, $\mathbb{P}(\tau_1 = 0) = 0$ or 1. Moreover, by monotone convergence,

$$\mathbb{P}(\tau_1 = 0) = \lim_{\varepsilon\to 0} \mathbb{P}(\tau_1 \leqslant \varepsilon) \geqslant \lim_{\varepsilon\to 0} \mathbb{P}(\mathrm{B}_\varepsilon > 0) = 1/2,$$

so $\mathbb{P}(\tau_1 = 0) = 1$, proving (i). By symmetry, $\inf\{t > 0 : \mathrm{B}_t < 0\} = 0$ a.s. so point (ii) comes from the a.s. continuity of the Brownian path and the intermediate value theorem. From Theorem 2.11, $(t\mathrm{B}_{1/t}, t \geqslant 0)$ is a Brownian motion, so (ii) gives $\inf\{t > 0 : t\mathrm{B}_{1/t} = 0\} = 0$ almost surely, i.e. $\sup\{t > 0 : \mathrm{B}_t = 0\} = \infty$ almost surely, proving (iii). □

We now prove the strong Markov property, i.e. an analogue of Theorem 2.15 where the shifted Brownian motion begins from a stopping time. For this, we need to define notions analogue to what was discussed in the discrete setting, Chapter 1.

Definition 2.18. *On a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *a filtration* $(\mathcal{F}_t)_{t\geqslant 0}$ *is an increasing sequence of sub-*σ*-algebras of* $\mathcal{F}$.

A random variable $\mathrm{T} : \Omega \to \mathbb{R}_+ \cup \{\infty\}$ *is called a stopping time with respect to a filtration* $(\mathcal{F}_t)_{t\geqslant 0}$ *if, for any* $t \geqslant 0$, $\{\mathrm{T} \leqslant t\} \in \mathcal{F}_t$.

Examples of stopping times with respect to the Brownian filtration include

$$\mathrm{T}_a = \inf\{t \geqslant 0 : \mathrm{B}_t = a\}.$$

Indeed, if for example $a \geqslant 0$, $\{\mathrm{T}_a \leqslant t\} = \{\sup_{[0,t]} \mathrm{B}_s \geqslant a\} = \{\sup_{[0,t]\cap\mathbb{Q}} \mathrm{B}_s \geqslant a\}$ by continuity of the Brownian path, so $\{\mathrm{T}_a \leqslant t\}$ is a countable union of $\mathcal{F}_t$-measurable events, hence it is in $\mathcal{F}_t$. As we will see thanks to the stopping time theorems, not all $\mathcal{F}$-measurable times are stopping times. As an example, $g = \sup\{t \in [0, 1] : \mathrm{B}_t = 0\}$ is not a stopping time.

Given a stopping time T, the information available till time T is defined as

$$\mathcal{F}_\mathrm{T} = \{\mathrm{A} \in \mathcal{F} : \forall t \geqslant 0, \mathrm{A} \cap \{\mathrm{T} \leqslant t\} \in \mathcal{F}_t\}.$$

As an exercise, one can prove that this is a σ-algebra, and that $\mathrm{T}, \mathrm{B}_\mathrm{T}\mathbb{1}_{\{\mathrm{T}<\infty\}}$ are $\mathcal{F}_\mathrm{T}$-measurable.

Theorem 2.19 (Strong Markov property)**.** *Let* B *be a Brownian motion and* T *a stopping time with respect to* $(\mathcal{F}_t)_{t\geqslant 0}$, $\mathcal{F}_t = \sigma(\mathrm{B}_s, s \leqslant t)$.

Then conditionally to $\{\mathrm{T} < \infty\}$, *noting* $\tilde{\mathrm{B}}_t = \mathrm{B}_{\mathrm{T}+t} - \mathrm{B}_\mathrm{T}$, $\tilde{\mathrm{B}}$ *is a Brownian motion independent of* $\mathcal{F}_\mathrm{T}$.

Proof. First assume $\mathrm{T} < \infty$ almost surely. We need to prove that for any $\mathrm{A} \in \mathcal{F}_\mathrm{T}$, $n \geqslant 0$, $0 \leqslant t_1 < \cdots < t_n$, and $\mathrm{F} : \mathbb{R}^n \to \mathbb{R}$ bounded and continuous,

$$\mathbb{E}\left(\mathrm{F}(\tilde{\mathrm{B}}_{t_1}, \dots, \tilde{\mathrm{B}}_{t_n})\mathbb{1}_\mathrm{A}\right) = \mathbb{E}\left(\mathrm{F}(\mathrm{B}_{t_1}, \dots, \mathrm{B}_{t_n})\right)\mathbb{E}\left(\mathbb{1}_\mathrm{A}\right). \tag{2.5}$$

Indeed, the case $\mathrm{A} = \Omega$ will prove that $\tilde{\mathrm{B}}$ is a Brownian motion, and general A the independence property, by the monotone class theorem. Note that, for a given ω, as $\mathrm{T} < \infty$ a.s.

$$\mathrm{F}(\tilde{\mathrm{B}}_{t_1}, \dots, \tilde{\mathrm{B}}_{t_n})\mathbb{1}_\mathrm{A} = \lim_{m\to\infty} \sum_{i=1}^{\infty} \mathrm{F}\left(\mathrm{B}_{\frac{i}{m}+t_1} - \mathrm{B}_{\frac{i}{m}}, \dots, \mathrm{B}_{\frac{i}{m}+t_n} - \mathrm{B}_{\frac{i}{m}}\right)\mathbb{1}_\mathrm{A}\mathbb{1}_{\frac{i-1}{m}<\mathrm{T}\leqslant\frac{i}{m}}.$$

By dominated convergence, when taking expectations they can be reversed with the limit. As $\mathrm{A} \in \mathcal{F}_{\mathrm{T}}$, by definition of $\mathcal{F}_{\mathrm{T}}$ the event $\mathrm{A} \cap \{\mathrm{T} \leqslant \frac{i}{m}\}$ is $\mathcal{F}_{\frac{i}{m}}$ measurable. Hence $\mathrm{A} \cap \{\frac{i-1}{m} < \mathrm{T} \leqslant \frac{i}{m}\}$ is $\mathcal{F}_{\frac{i}{m}}$ measurable, so by Theorem 2.15

$$\begin{aligned}\mathbb{E}\left(\mathrm{F}\left(\mathrm{B}_{\frac{i}{m}+t_1} - \mathrm{B}_{\frac{i}{m}}, \dots, \mathrm{B}_{\frac{i}{m}+t_n} - \mathrm{B}_{\frac{i}{m}}\right) \mathbb{1}_{\mathrm{A}} \mathbb{1}_{\frac{i-1}{m} < \mathrm{T} \leqslant \frac{i}{m}}\right) \\ = \mathbb{E}\left(\mathrm{F}\left(\mathrm{B}_{t_1}, \dots, \mathrm{B}_{t_n}\right)\right) \mathbb{E}\left(\mathbb{1}_{\mathrm{A}} \mathbb{1}_{\frac{i-1}{m} < \mathrm{T} \leqslant \frac{i}{m}}\right).\end{aligned}$$

The proof of (2.5) now follows by summation. In the case $\mathbb{P}(\mathrm{T} = \infty) > 0$, the proof goes the same way by replacing A with $\mathrm{A} \cap \{\mathrm{T} < \infty\}$. □

One of the most famous applications of the strong Markov property is the following reflection principle. Please note that it is not just a curious example of an integrable law : the queuing distribution of the maximum of B will be useful to prove tightness in the forthcoming sections and chapters.

Theorem 2.20. *Let* $\mathrm{S}_t = \sup_{0 \leqslant u \leqslant t} \mathrm{B}_u$. *Then for any* $t \geqslant 0, a \geqslant 0$ *and* $b \leqslant a$

$$\mathbb{P}(\mathrm{S}_t \geqslant a, \mathrm{B}_t \leqslant b) = \mathbb{P}(\mathrm{B}_t \geqslant 2a - b).$$

In particular, $\mathrm{S}_t \overset{\text{law}}{=} |\mathrm{B}_t|$.

Proof. Let $\mathrm{T}_a = \inf\{t \geqslant 0 : \mathrm{B}_t = a\}$. Define the process $\tilde{\mathrm{B}}$ on $[0, t]$ by

$$\tilde{\mathrm{B}}_s = \left\{ \begin{array}{cl} \mathrm{B}_s & \text{if } s \leqslant \mathrm{T}_a \\ 2a - \mathrm{B}_s & \text{if } \mathrm{T}_a \leqslant s \leqslant t \end{array} \right. .$$

By Theorem 2.19, $(\mathrm{B}_{\mathrm{T}_a+s} - \mathrm{B}_{\mathrm{T}_a}, s \geqslant 0)$ is a Brownian motion independent of $\mathcal{F}_{\mathrm{T}_a}$, hence its reflection $(\tilde{\mathrm{B}}_{\mathrm{T}_a+s} - \tilde{\mathrm{B}}_{\mathrm{T}_a}, s \geqslant 0)$ is also a Brownian motion independent of $\mathcal{F}_{\mathrm{T}_a}$. Being the juxtaposition of the Brownian motion B till T_a with another independent Brownian motion, $\tilde{\mathrm{B}}$ is a Brownian Motion. Consequently,

$$\begin{aligned}\mathbb{P}\left(\sup_{0 \leqslant u \leqslant t} \mathrm{B}_u \geqslant a, \mathrm{B}_t \leqslant b\right) = \mathbb{P}\left(\sup_{0 \leqslant u \leqslant t} \tilde{\mathrm{B}}_u \geqslant a, \tilde{\mathrm{B}}_t \leqslant b\right) \\ = \mathbb{P}(\mathrm{T}_a \leqslant t, \mathrm{B}_t \geqslant 2a - b) = \mathbb{P}(\mathrm{B}_t \geqslant 2a - b)\end{aligned}$$

because, as $2a - b \geqslant a$, $\mathrm{B}_t \geqslant 2a - b$ implies $\mathrm{T}_a \leqslant t$ by the intermediate values theorem. To conclude, we can write

$$\begin{aligned}\mathbb{P}(\mathrm{S}_t \geqslant a) = \mathbb{P}(\mathrm{S}_t \geqslant a, \mathrm{B}_t \leqslant a) + \mathbb{P}(\mathrm{S}_t \geqslant a, \mathrm{B}_t > a) \\ = \mathbb{P}(\mathrm{B}_t \geqslant a) + \mathbb{P}(\mathrm{B}_t > a) = \mathbb{P}(|\mathrm{B}_t| \geqslant a),\end{aligned}$$

so $\mathrm{S}_t \overset{\text{law}}{=} |\mathrm{B}_t|$. □

6. Iterated logarithm law

We now turn to the exact estimate of the optimal Hölder coefficient of a Brownian trajectory : for any $\alpha \in [0, 1/2)$, we proved in Corollary 2.14 that there is a constant $c_\alpha(\omega)$ such that for any s, t in $[0, 1]$

$$|\mathrm{B}_t - \mathrm{B}_s| \leqslant c_\alpha(\omega)|t - s|^\alpha.$$

Point (*ii*) in the following theorem proves that the trajectory is not Hölderian with exponent $1/2$.

Theorem 2.21 (Iterated logarithm law)**.** *Let* B *be a Brownian motion.*

(i) Almost surely, $\limsup_{t\to\infty} \frac{\mathrm{B}_t}{\sqrt{2t\log\log t}} = 1.$

(ii) Almost surely, $\limsup_{t\to 0^+} \frac{\mathrm{B}_t}{\sqrt{2t\log(-\log t)}} = 1.$

Proof. Note that (ii) is a simple consequence of (i) by time inversion, Theorem 2.11.

To prove (i), for given $\lambda > 1$ and $c > 0$, we write

$$\mathrm{A}_k = \{\mathrm{B}_{\lambda^k} \geqslant c f(\lambda^k)\}$$

where $f(t) = \sqrt{2t\log\log t}$. Then, using Lemma 2.4, if $\mathrm{X} \sim \mathcal{N}(0,1)$ then

$$\begin{aligned}\mathbb{P}(\mathrm{A}_k) = \mathbb{P}\left(\mathrm{X} \geqslant c\sqrt{2\log\log(\lambda^k)}\right) &\underset{k\to\infty}{\sim} \frac{1}{\sqrt{2\pi}}\frac{1}{c\sqrt{2\log\log(\lambda^k)}}e^{-c^2\log\log(\lambda^k)}\\ &\underset{k\to\infty}{\sim} \frac{1}{2c\sqrt{\pi}}\left(\frac{1}{\log\lambda}\right)^{c^2}\frac{1}{\sqrt{\log k}}\left(\frac{1}{k}\right)^{c^2} \qquad (2.6)\end{aligned}$$

Hence $\sum_k \mathbb{P}(\mathrm{A}_k)$ converges (resp. diverges) if $c > 1$ (resp. $c \leqslant 1$). This proves by the Borel-Cantelli lemma that

$$\limsup \frac{\mathrm{B}_{\lambda^k}}{f(\lambda^k)} \leqslant 1 \qquad (2.7)$$

almost surely, and equality would hold if the A_k's were independent, which is not true. But there is a sufficiently small correlation between them and this problem can be handled. More precisely, noting

$$\mathrm{C}_k = \{\mathrm{B}_{\lambda^{k+1}} - \mathrm{B}_{\lambda^k} \geqslant c f(\lambda^{k+1} - \lambda^k)\},$$

a calculation similar to (2.6) yields

$$\mathbb{P}\left(\mathrm{C}_k\right) \underset{k\to\infty}{\sim} \frac{1}{2c\sqrt{\pi}}\left(\frac{1}{\log\lambda}\right)^{c^2}\frac{1}{\sqrt{\log k}}\left(\frac{1}{k}\right)^{c^2}.$$

This diverges if $c < 1$, so thanks to the independence of the increments of B, $\mathbb{P}(\cap_n \cup_{k\geqslant n} \mathrm{C}_k) = 1$: if $c < 1$, in infinitely many times

$$\frac{\mathrm{B}_{\lambda^{k+1}}}{f(\lambda^{k+1})} \geqslant c\frac{f(\lambda^{k+1}-\lambda^k)}{f(\lambda^{k+1})} + \frac{f(\lambda^k)}{f(\lambda^{k+1})}\frac{\mathrm{B}_{\lambda^k}}{f(\lambda^k)}$$

But using symmetry and (2.7), for any $\varepsilon > 0$, almost surely for sufficiently large k

$$\frac{\mathrm{B}_{\lambda^k}}{f(\lambda^k)} \geqslant -1-\varepsilon.$$

As a consequence, infinitely often

$$\frac{\mathrm{B}_{\lambda^{k+1}}}{f(\lambda^{k+1})} \geqslant c\frac{f(\lambda^{k+1}-\lambda^k)}{f(\lambda^{k+1})} - (1+\varepsilon)\frac{f(\lambda^k)}{f(\lambda^{k+1})} \underset{k\to\infty}{\sim} c\sqrt{1-\frac{1}{\lambda}} - \frac{1+\varepsilon}{\sqrt{\lambda}}$$

By choosing λ arbitrary large, we have proved that

$$\limsup_{t\to\infty} \frac{\mathrm{B}_t}{\sqrt{2t\log\log t}} \geqslant 1.$$

Our last task consists in controlling what happens between λ^k and λ^{k+1} to extend (2.7) to a lim sup with argument t. Let $t \in [\lambda^k, \lambda^{k+1}]$. Then, still for any $\lambda > 1$,

$$\frac{\mathrm{B}_t}{f(t)} = \frac{\mathrm{B}_{\lambda^k}}{f(\lambda^k)}\frac{f(\lambda^k)}{f(t)} + \frac{\mathrm{B}_t - \mathrm{B}_{\lambda^k}}{f(t)} \leqslant \frac{\mathrm{B}_{\lambda^k}}{f(\lambda^k)} + \frac{\mathrm{B}_t - \mathrm{B}_{\lambda^k}}{f(\lambda^k)}.$$

Consider the event

$$\mathrm{D}_k = \left\{ \sup_{[\lambda^k, \lambda^{k+1}]} \frac{\mathrm{B}_t - \mathrm{B}_{\lambda^k}}{f(\lambda^k)} \geqslant \alpha \right\},$$

for any given $\alpha > 0$. Using that $(\mathrm{B}_{s+\lambda^k} - \mathrm{B}_{\lambda^k}, s \geqslant 0)$ is a Brownian motion and the scaling property in Theorem 2.11,

$$\mathbb{P}(\mathrm{D}_k) = \mathbb{P}\left(\sup_{[0,1]} \mathrm{B}_u \geqslant \frac{\alpha f(\lambda^k)}{\sqrt{\lambda^{k+1} - \lambda^k}} \right) = 2\mathbb{P}\left(\mathrm{B}_1 \geqslant \frac{\alpha f(\lambda^k)}{\sqrt{\lambda^{k+1} - \lambda^k}} \right)$$

where we used Theorem 2.20. Lemma 2.4 therefore yields

$$\mathbb{P}(\mathrm{D}_k) \underset{k \to \infty}{\sim} \frac{c_1}{\sqrt{\log k}} \left(\frac{1}{k} \right)^{\frac{\alpha^2}{\lambda - 1}}$$

for some $c_1 > 0$. Hence, if $\alpha^2 > \lambda - 1$, a.s. for sufficiently large k, $t \in [\lambda^k, \lambda^{k+1}]$, then

$$\frac{\mathrm{B}_t}{f(t)} \leqslant \frac{\mathrm{B}_{\lambda^k}}{f(\lambda^k)} + \alpha.$$

This proves that $\limsup \mathrm{B}_t / f(t) \leqslant 1 + \alpha$ for any $\alpha > \sqrt{\lambda - 1}$, hence $\limsup \mathrm{B}_t / f(t) \leqslant 1$. □

Note that, by giving for any $\varepsilon > 0$ easy upper bounds to the probability of the events $\mathrm{E}_n = \{\sup_{[n,n+1]} |\mathrm{B}_s - \mathrm{B}_n| > \varepsilon\}$, the above Theorem proves

$$\limsup_{n \to \infty} \frac{\mathrm{B}_n}{\sqrt{2n \log \log n}} = 1 \tag{2.8}$$

a.s. where $\mathrm{B}_n = \sum_1^n \mathrm{X}_k$, where the X_k's defined as $\mathrm{B}_k - \mathrm{B}_{k-1}$ are independent standard Gaussians. One may wonder if this extends to the arbitrary centered reduced random variables X_k's. The answer is yes, and a possible proof makes use of the Skorokhod embedding, proved in the following section : for any X with expectation 0 and variance 1, given B a Brownian motion with natural filtration $(\mathcal{F}_t)_{t \geqslant 0}$, there is a stopping time T with expectation 1 such that

$$\mathrm{B}_\mathrm{T} \overset{\text{law}}{=} \mathrm{X}. \tag{2.9}$$

Theorem 2.22. *Let* $\mathrm{X}_1, \mathrm{X}_2, \dots$ *be iid random variables with expectation 0 and variance 1, and* $\mathrm{S}_n = \sum_{k=1}^n \mathrm{X}_k$*. Then almost surely*

$$\limsup_{n \to \infty} \frac{\mathrm{S}_n}{\sqrt{2n \log \log n}} = 1.$$

Proof. For the given Brownian motion B, define by induction the sequence of stopping times

$$\mathrm{T}_0 = 0$$
$$\mathrm{T}_{n+1} = \mathrm{T}_n + \mathrm{T}(\omega_n)$$

where T is the stopping time related to the Skorokhod embedding (2.9) and $\omega_n = (\mathrm{B}_{\mathrm{T}_n + s} - \mathrm{B}_{\mathrm{T}_n}, s \geqslant 0)$. Note that $\mathbb{E}(\mathrm{T}) = 1$, hence $\mathrm{T}_n < \infty$ almost surely, so by the strong Markov Theorem 2.19 ω_n is independent of $\mathcal{F}_{\mathrm{T}_n}$. This implies that :

(i) the random variables $\mathrm{T}_{n+1} - \mathrm{T}_n$ are independent, with expectation 1, so by the strong law of large numbers T_n / n converges to 1 almost surely ;

(ii) $(\mathrm{S}_n, n \geqslant 0) \overset{\text{law}}{=} (\mathrm{B}_{\mathrm{T}_n}, n \geqslant 0)$. Consequently, we need to prove that

$$\limsup_{n \to \infty} \frac{\mathrm{B}_{\mathrm{T}_n}}{f(n)} = 1,$$

where $f(n)=\sqrt{2n\log\log n}$, as previously. By (2.8) we only need to prove that

$$\frac{\mathrm{B}_{\mathrm{T}_n}-\mathrm{B}_n}{f(n)}\underset{n\to\infty}{\longrightarrow}0$$

almost surely. To prove this, we split the randomness coming from the Brownian motion and the stopping time by writing, for any $\varepsilon>0$ and $\delta\in(0,1)$,

$$\left\{\left|\frac{\mathrm{B}_{\mathrm{T}_n}-\mathrm{B}_n}{f(n)}\right|>\varepsilon\right\}\subset\{|\mathrm{T}_n-n|>\delta n\}\cup\left\{\sup_{s\in[(1-\delta)n,(1+\delta)n]}\left|\frac{\mathrm{B}_s-\mathrm{B}_n}{f(n)}\right|>\varepsilon\right\}.$$

As $\mathrm{T}_n/n\to 1$ a.s. the first event of the RHS doesn't occur for sufficiently large n. Concerning the second, if $\lambda:=1+\delta$ and k_n is the unique integer such that $\lambda^{k_n}\leqslant n<\lambda^{k_n+1}$, then

$$\begin{aligned}\left\{\sup_{s\in[n,(1+\delta)n]}\left|\frac{\mathrm{B}_s-\mathrm{B}_n}{f(n)}\right|>\varepsilon\right\}&\subset\left\{\sup_{s\in[\lambda^{k_n},\lambda^{k_n+2}]}\left|\frac{\mathrm{B}_s-\mathrm{B}_n}{f(\lambda^{k_n})}\right|>\varepsilon\right\}\\&\subset\left\{\sup_{s\in[\lambda^{k_n},\lambda^{k_n+2}]}\left|\frac{\mathrm{B}_s-\mathrm{B}_{\lambda^{k_n}}}{f(\lambda^{k_n})}\right|>\varepsilon/2\right\}\cup\left\{\left|\frac{\mathrm{B}_{\lambda^{k_n}}-\mathrm{B}_n}{f(\lambda^{k_n})}\right|>\varepsilon/2\right\}\\&=\left\{\sup_{s\in[\lambda^{k_n},\lambda^{k_n+2}]}\left|\frac{\mathrm{B}_s-\mathrm{B}_{\lambda^{k_n}}}{f(\lambda^{k_n})}\right|>\varepsilon/2\right\}=:\mathrm{E}_{k_n}.\end{aligned}$$

A calculation similar to the one performed in the proof of Theorem 2.21 proves that if $\varepsilon^2>\lambda^2-1$ (true by choosing δ small enough), $\sum_k\mathbb{P}(\mathrm{E}_k)<\infty$ so, for sufficiently large k, E_k does not occur almost surely. Hence, for n large enough,

$$\sup_{s\in[n,(1+\delta)n]}\left|\frac{\mathrm{B}_s-\mathrm{B}_n}{f(n)}\right|\leqslant\varepsilon.$$

The analogous result for the maximum on $[(1-\delta)n,n]$ holds similarly, concluding the proof. □

7. Skorokhod's embedding

Before making explicit the solution to (2.9), we prove the following useful lemma.

Lemma 2.23 (Wald's identities)**.** *Let* B *be a Brownian motion and* T *a stopping time such that* $\mathbb{E}(\mathrm{T})<\infty$*. Then the following identities hold :*

(i) $\mathbb{E}(\mathrm{B}_\mathrm{T})=0$ *;*

(ii) $\mathbb{E}(\mathrm{B}_\mathrm{T}^2)=\mathbb{E}(\mathrm{T})$.

Proof. To prove (i), we can bound

$$\mathrm{B}_{t\wedge\mathrm{T}}\leqslant\sum_{k=1}^{\lfloor\mathrm{T}\rfloor}\sup_{0\leqslant t\leqslant 1}|\mathrm{B}_{t+k}-\mathrm{B}_k|=:\mathrm{M},$$

and observe that M is in L^1 :

$$\begin{aligned}\mathbb{E}(\mathrm{M})&=\sum_{k=1}^{\infty}\mathbb{E}\left(\mathbb{1}_{\mathrm{T}\geqslant k}\sup_{0\leqslant t\leqslant 1}|\mathrm{B}_{t+k}-\mathrm{B}_k|\right)\\&=\sum_{k=1}^{\infty}\mathbb{P}(\mathrm{T}\geqslant k)\,\mathbb{E}\left(\sup_{0\leqslant t\leqslant 1}|\mathrm{B}_{t+k}-\mathrm{B}_k|\right)\leqslant\mathbb{E}(\mathrm{T}+1)\,\mathbb{E}\left(\sup_{0\leqslant t\leqslant 1}|\mathrm{B}_t|\right)<\infty.\end{aligned}$$

As a consequence, $(B_{t\wedge T}, t \geqslant 0)$ is uniformly integrable, and the straightforward continuous analogue[2] of Theorem 1.11 allows to apply the stopping time theorem at time T, thus $\mathbb{E}(B_T) = 0$.

To prove (*ii*), let $T_n = \inf\{t \geqslant 0 : |B_t| = n\}$. Then $(B^2_{t\wedge T_n \wedge T} - (t \wedge T_n \wedge T), t \geqslant 0)$ is a martingale, bounded by $n^2 + T$, which is in L^1, hence this martingale is uniformly integrable so

$$\mathbb{E}(B^2_{T\wedge T_n}) = \mathbb{E}(T \wedge T_n).$$

By Fatou's lemma, this yields

$$\mathbb{E}(B^2_T) = \mathbb{E}(\liminf_{n\to\infty} B^2_{T\wedge T_n}) \leqslant \liminf_{n\to\infty} \mathbb{E}(B^2_{T\wedge T_n}) = \liminf_{n\to\infty} \mathbb{E}(T \wedge T_n) = \mathbb{E}(T)$$

by monotone convergence. Conversely, note that for any stopping time $S \leqslant T$,

$$\mathbb{E}(B^2_T) = \mathbb{E}(B^2_S) + \mathbb{E}((B_T - B_S)^2) + 2\,\mathbb{E}\left(B_S\,\mathbb{E}(B_T - B_S \mid \mathcal{F}_S)\right) = \mathbb{E}(B^2_S) + \mathbb{E}((B_T - B_S)^2),$$

because $\mathbb{E}(B_T \mid \mathcal{F}_S) = B_S$, by the stopping time theorem applied to the uniformly integrable martingale $(B_{t\wedge T}, t \geqslant 0)$. As a consequence, $\mathbb{E}(B^2_T) \geqslant \mathbb{E}(B^2_S)$, which applied to $S = T \wedge T_n$ yields

$$\mathbb{E}(B^2_T) \geqslant \lim_{n\to\infty} \mathbb{E}(B^2_{T\wedge T_n}) = \lim_{n\to\infty} \mathbb{E}(T \wedge T_n) = \mathbb{E}(T)$$

by monotone convergence. This concludes the proof. □

We now come back to the embedding problem. First, note that given any random variable X (with a bounded density for the sake of simplicity here), finding a stopping time such that $B_T \sim X$ is an easy task.

1) There is a $\mathcal{F}_1$-measurable Z with $Z \sim X$: by independence of the increments of Brownian motion, there is a family of independent standard Gaussians, and from this family Z is obtained by rejection sampling.

2) The choice $T = \inf\{s \geqslant 2 : B_s = Z\}$ is almost surely finite stopping time, and obviously $B_T \sim Z$.

However, it is easy to prove[3] that $\mathbb{E}(T) = \infty$, which is not nice for applications. For example, $\mathbb{E}(T) < \infty$ is essential in the proof of Theorem 2.22.

Note that, by the Wald identities, if $\mathbb{E}(T) < \infty$, then $\mathbb{E}(B_T) = 0$ and $\mathbb{E}(B^2_T) = \mathbb{E}(T)$, so the expectation 0 and finite variance hypothesis in the following theorem are not restrictive. Moreover, the variance can be assumed to be 1, by the scaling property of Brownian motion.

Theorem 2.24 (Skorokhod's embedding)**.** *Let* X *be a centered random variable with variance 1. Then there is a stopping time* T *with expectation 1 such that*

$$B_T \overset{\text{law}}{=} X.$$

To prove this result, we follow the construction of T by Dubins. There are many other constructions, at least twenty one of them being listed, with extensions and applications, in Obłoj's survey [14].

To prove the above Skorokhod embedding, a central tool is the convergence of a special type martingale to a random variable with law X. More precisely, a discrete martingale is said to be *binary splitting* if conditionally to the past, it only can take two values :

$$|\{X_{n+1}(\omega) : (X_0, \dots, X_n)(\omega) = (x_0, \dots, x_n)\}| \leqslant 2.$$

2. This continuous extension will be proved in Chapter 3.

3. For this, note that $\mathbb{E}(T_a) = \infty$, where $T_a = \inf\{s \geqslant 0 : B_s = a\}$ and $a \neq 0$; indeed, if not, Wald's identity would give $0 = \mathbb{E}(B_{T_a}) = a$. Then the result follows by conditioning on $\mathcal{F}_1$.

Lemma 2.25. *Let* X *be a random variable with finite variance. Then there is a binary splitting martingale* $(\mathrm{X}_n)_{n\geqslant 0}$ *such that* $\mathrm{X}_n \to \mathrm{X}$ *almost surely and in* L^2*.*

Proof. Define the sequence of random variables $(\mathrm{X}_n)_{n\geqslant 0}$ by, at rank 0,

$$\left\{\begin{array}{rcl} \mathrm{X}_0 &=& \mathbb{E}(\mathrm{X}) \\ \zeta_0 &=& 1 \text{ if } \mathrm{X} \geqslant \mathrm{X}_0, -1 \text{ otherwise} \end{array}\right.,$$

and for any $n \geqslant 1$

$$\left\{\begin{array}{rcl} \mathcal{G}_n &=& \sigma(\zeta_0, \dots, \zeta_{n-1}) \\ \mathrm{X}_n &=& \mathbb{E}(\mathrm{X} \mid \mathcal{G}_n) \\ \zeta_n &=& 1 \text{ if } \mathrm{X} \geqslant \mathrm{X}_n,\ -1 \text{ otherwise} \end{array}\right..$$

In the following, the values ± 1 for ζ_n play no role, any distinct two numbers are sufficient. The process $(\mathrm{X}_n)_{n\geqslant 0}$ is a martingale, because X is integrable and $(\mathcal{G}_n)_{n\geqslant 0}$ is a filtration. Conditionally to $(\mathrm{X}_0, \dots, \mathrm{X}_{n-1})$, X_n may be in one of the $n+1$ intervals and then one may think X_n can take $n+1$ values. But X_n is actually constrained to be in only 2 possible intervals the smallest ones surrounding X_{n-1} : a calculation shows that $\mathrm{X}_n - \mathrm{X}_k$ has the same sign as $\mathrm{X}_\ell - \mathrm{X}_k$ for any $\ell > k$. Hence X has the binary splitting property.

Moreover,

$$\begin{aligned} \mathbb{E}(\mathrm{X}^2) = \mathbb{E}(\mathrm{X}_n^2) + \mathbb{E}((\mathrm{X}-\mathrm{X}_n)^2) + 2\,\mathbb{E}\left(\mathrm{X}_n\,\mathbb{E}(\mathrm{X}-\mathrm{X}_n \mid \mathcal{G}_n)\right) \\ = \mathbb{E}(\mathrm{X}_n^2) + \mathbb{E}((\mathrm{X}-\mathrm{X}_n)^2) \geqslant \mathbb{E}(\mathrm{X}_n^2), \end{aligned}$$

as a consequence $(\mathrm{X}_n)_{n\geqslant 0}$ is a L^2-bounded martingale, hence it converges almost surely and in L^2, to a random variable noted $\overline{\mathrm{X}}$.

We still need to prove that $\mathrm{X} = \overline{\mathrm{X}}$. Note that :

- as X_n is uniformly L^2-bounded and $\mathrm{X} \in \mathrm{L}^2$, $\mathrm{Y}_n := \zeta_n(\mathrm{X}_{n+1} - \mathrm{X})$ is bounded is L^2 ;
- $\lim_{n\to\infty} \mathrm{Y}_n = |\overline{\mathrm{X}} - \mathrm{X}| =: \mathrm{Y}$ a.s. because if $\mathrm{X} > \overline{\mathrm{X}}$, for sufficiently large n $\mathrm{X}_n < \mathrm{X}$, and the same way if $\mathrm{X} < \overline{\mathrm{X}}$.

If Y_n is uniformly bounded in L^2 and converges almost surely to some $\mathrm{Y} \in \mathrm{L}^2$, then[4] $\mathbb{E}(\mathrm{Y}_n) \to \mathbb{E}(\mathrm{Y})$. In our situation, as ζ_n is $\mathcal{G}_{n+1}$-measurable, $\mathbb{E}(\mathrm{Y}_n) = 0$, so $\mathbb{E}(\mathrm{Y}) = 0$, which is the expected result : $\overline{\mathrm{X}} = \mathrm{X}$ almost surely. □

Proof of Theorem 2.24. From the previous lemma, there is a discrete martingale converging in L^2 and a.s. to X, and for any n the support of X_n is noted $\{a_n, b_n\} = f^{(n)}(\mathrm{X}_0, \dots, \mathrm{X}_{n-1})$.

Define T_n recursively by $\mathrm{T}_0 = 0$ and

$$\mathrm{T}_n = \inf\{t \geqslant \mathrm{T}_{n-1} : \mathrm{B}_t \in f^{(n)}(\mathrm{B}_{\mathrm{T}_0}, \dots, \mathrm{B}_{\mathrm{T}_{n-1}})\}.$$

Then obviously $(\mathrm{B}_{\mathrm{T}_n})_{n\geqslant 0}$ and $(\mathrm{X}_n)_{\geqslant 0}$ have he same law as, for a binary splitting martingale, there is only one possible choice in the transition probabilities.

Let $\mathrm{T} = \lim_{n\to\infty} \mathrm{T}_n$. One easily checks that this is a stopping time. The expectation of T_n is finite : there are finitely many possible a_n and b_n's, so $\mathrm{T}_n(\omega) \leqslant \mathrm{S} :=$

4. To prove it, note that for any $\alpha > 0$

$$\begin{aligned} \mathbb{E}(|\mathrm{Y}_n - \mathrm{Y}|) = \mathbb{E}(|\mathrm{Y}_n - \mathrm{Y}|\mathbb{1}_{\mathrm{Y}_n \leqslant \alpha}) + \mathbb{E}(|\mathrm{Y}_n - \mathrm{Y}|\mathbb{1}_{\mathrm{Y}_n > \alpha}) \\ \leqslant \mathbb{E}(|\mathrm{Y}_n - \mathrm{Y}|\mathbb{1}_{\mathrm{Y}_n \leqslant \alpha}) + \mathbb{E}(|\mathrm{Y}_n - \mathrm{Y}|^2)^{1/2}\mathbb{P}(\mathrm{Y}_n > \alpha)^{1/2}. \end{aligned}$$

The first term converges to 0 by dominated convergence, and in the second $\mathbb{E}(|\mathrm{Y}_n - \mathrm{Y}|^2)$ is uniformly bounded and as $\mathbb{P}(\mathrm{Y}_n > \alpha) \leqslant \mathbb{E}(\mathrm{Y}_n^2)/\alpha^2$, the second uniformly goes to 0 as $\alpha \to 0$, concluding the proof.

$\inf\{t \geqslant 0 : \mathrm{B}_t \in \{-a,a\}\}$ for sufficiently large a, and this last stopping time has a finite expectation[5]. Hence, as a consequence, by Wald's lemma, $\mathbb{E}(\mathrm{B}_{\mathrm{T}_n}^2) = \mathbb{E}(\mathrm{T}_n)$ and by monotone convergence

$$\mathbb{E}(\mathrm{T}) = \lim_{n\to\infty} \mathbb{E}(\mathrm{T}_n) = \lim_{n\to\infty} \mathbb{E}(\mathrm{B}_{\mathrm{T}_n}^2) = \mathbb{E}(\mathrm{X}^2) = 1,$$

as $(\mathrm{B}_{\mathrm{T}_n})_{n\geqslant 0} \overset{\text{law}}{=} (\mathrm{X}_n)_{\geqslant 0}$ and X_n converges to X in L^2. To conclude, finally note that $\mathrm{B}_{\mathrm{T}_n}$ converges in law to X and almost surely to B_T, hence $\mathrm{B}_\mathrm{T} \sim \mathrm{X}$. □

8. Donsker's invariance principle

Let $\mathrm{X}_1, \mathrm{X}_2, \dots$ be iid centered random variables, with variance 1, and $\mathrm{S}_n = \sum_{k=1}^n \mathrm{X}_k$. These partial sums can be extended to continuous argument by writing

$$\mathrm{S}_t = \mathrm{S}_{\lfloor t\rfloor} + (t - \lfloor t\rfloor)(\mathrm{S}_{\lfloor t\rfloor+1} - \mathrm{S}_{\lfloor t\rfloor}). \tag{2.10}$$

Consider the normalized function

$$\mathrm{S}^{(n)}(t) = \frac{\mathrm{S}_{nt}}{\sqrt{n}},\ 0 \leqslant t \leqslant 1.$$

Theorem 2.26. *On the set of continuous functions on* $[0,1]$*, as* $n\to\infty$ *the process* $\mathrm{S}^{(n)}$ *converges in law to a Browian motion* B*.*

The meaning of this convergence in law is : for any bounded $\mathrm{F} : \mathscr{C}([0,1]) \to \mathbb{R}$, continuous for the L^∞ norm,

$$\mathbb{E}(\mathrm{F}(\mathrm{S}^{(n)})) \underset{n\to\infty}{\longrightarrow} \mathbb{E}(\mathrm{F}(\mathrm{B})).$$

As an example of application, using Portmanteau's theorem and the reflection principle Theorem 2.20, if $(\mathrm{S}_n)_{n\geqslant 0}$ is a standard random walk, and $\lambda > 0$,

$$\mathbb{P}\left(\sup_{[\![1,n]\!]} \mathrm{S}_k > \lambda\sqrt{n}\right) \underset{n\to\infty}{\longrightarrow} \sqrt{\frac{2}{\pi}} \int_\lambda^\infty e^{-\frac{x^2}{2}} \mathrm{d}x.$$

Proof. Let B be a Brownian motion. From the Skorokhod embedding Theorem 2.24, there is a stopping time $\mathrm{T}(\omega)$ such that $\mathbb{E}(\mathrm{T}) = 1$ and $\mathrm{B}_\mathrm{T} \sim \mathrm{X}_1$. Define a sequence of stopping times $(\mathrm{T}_n)_{n\geqslant 0}$ by

$$\begin{aligned} \mathrm{T}_0 &= 0 \\ \mathrm{T}_{n+1} &= \mathrm{T}_n + \mathrm{T}(\omega_n) \end{aligned}$$

where $\omega_n = (\mathrm{B}_{\mathrm{T}_n+s} - \mathrm{B}_{\mathrm{T}_n}, s \geqslant 0)$. By the strong Markov Theorem 2.19,

$$(\mathrm{B}_{\mathrm{T}_n}, n \geqslant 0) \overset{\text{law}}{=} (\mathrm{S}_n, n \geqslant 0).$$

Define $\mathrm{S}_t = \mathrm{B}_{\mathrm{T}_{\lfloor t\rfloor}} + (t - \lfloor t\rfloor)(\mathrm{B}_{\mathrm{T}_{\lfloor t\rfloor+1}} - \mathrm{B}_{\mathrm{T}_{\lfloor t\rfloor}})$, which is an inoffensive redundancy with (2.10) as both functions have the same law. Imagine we can prove the tightness condition : for any $\varepsilon > 0$

$$\mathbb{P}(\mathrm{A}_n) \underset{n\to\infty}{\longrightarrow} 0,\ \mathrm{A}_n := \left\{\sup_{[0,1]} \left|\frac{\mathrm{B}_{nt}}{\sqrt{n}} - \frac{\mathrm{S}_{nt}}{\sqrt{n}}\right| > \varepsilon\right\}. \tag{2.11}$$

5. $\mathbb{E}(\mathrm{S}) = \int \mathbb{P}(\mathrm{S} > t)\mathrm{d}t \leqslant \int \mathbb{P}(\forall s \in [0,t], |\mathrm{B}_s| < a)\mathrm{d}t$, and this integrand has bounded, for $t \geqslant n$, by $\mathbb{P}(|\mathrm{B}_1 - \mathrm{B}_0| < 2a, \dots, |\mathrm{B}_k - \mathrm{B}_{k-1}| < 2a)$, hence decreases exponentially.

Then the proof of the theorem easily follows : for any compact $\mathrm{K} \subset (\mathscr{C}([0,1]), \|\|_\infty)$ with ε-neighborhood noted K^ε,

$$\mathbb{P}\left(\frac{\mathrm{S}_{n\cdot}}{\sqrt{n}} \in \mathrm{K}\right) \leqslant \mathbb{P}\left(\frac{\mathrm{B}_{n\cdot}}{\sqrt{n}} \in \mathrm{K}^\varepsilon\right) + \mathbb{P}\left(\sup_{[0,1]}\left|\frac{\mathrm{B}_{nt}}{\sqrt{n}} - \frac{\mathrm{S}_{nt}}{\sqrt{n}}\right| > \varepsilon\right),$$

so using (2.11),

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{\mathrm{S}_{n\cdot}}{\sqrt{n}} \in \mathrm{K}\right) \leqslant \mathbb{P}\left(\frac{\mathrm{B}_{n\cdot}}{\sqrt{n}} \in \mathrm{K}^\varepsilon\right).$$

As $\varepsilon \to 0$, this converges to $\mathbb{P}(\mathrm{B} \in \mathrm{K})$ by monotone convergence (K is closed). This proves the convergence in law by Portmanteau's theorem.

We are therefore led to prove (2.11). First note that, as S_{nt} is affine on any interval of type $[k/n, (k+1)/n]$, the event A_n is included in

$$\begin{aligned}\left\{\sup_{[0,1]}\left|\frac{\mathrm{B}_{nt}}{\sqrt{n}} - \frac{\mathrm{S}_{\lfloor nt\rfloor}}{\sqrt{n}}\right| > \varepsilon\right\} &\cup \left\{\sup_{[0,1]}\left|\frac{\mathrm{B}_{nt}}{\sqrt{n}} - \frac{\mathrm{S}_{\lfloor nt\rfloor+1}}{\sqrt{n}}\right| > \varepsilon\right\} \\ &= \left\{\sup_{[0,1]}\left|\frac{\mathrm{B}_{nt}}{\sqrt{n}} - \frac{\mathrm{B}_{\mathrm{T}_{\lfloor nt\rfloor}}}{\sqrt{n}}\right| > \varepsilon\right\} \cup \left\{\sup_{[0,1]}\left|\frac{\mathrm{B}_{nt}}{\sqrt{n}} - \frac{\mathrm{B}_{\mathrm{T}_{\lfloor nt\rfloor+1}+1}}{\sqrt{n}}\right| > \varepsilon\right\}.\end{aligned}$$

For any given $\delta > 0$, the previous events union is included in $\mathrm{A}_n^{(1)} \cup \mathrm{A}_n^{(2)} \cup \mathrm{A}_n^{(3)} \cup \mathrm{A}_n^{(4)}$, where

$$\begin{aligned}\mathrm{A}_n^{(1)} &= \left\{\exists (s,t) \in [0,2]^2 : |s-t| \leqslant \delta,\ \left|\frac{\mathrm{B}_{nt}}{\sqrt{n}} - \frac{\mathrm{B}_{ns}}{\sqrt{n}}\right| > \varepsilon\right\}, \\ \mathrm{A}_n^{(2)} &= \left\{\exists t \in [0,1] : \frac{1}{n}|\mathrm{T}_{\lfloor nt\rfloor} - nt| \geqslant \delta\right\}, \\ \mathrm{A}_n^{(3)} &= \left\{\exists t \in [0,1] : \frac{1}{n}|\mathrm{T}_{\lfloor nt\rfloor+1} - nt| \geqslant \delta\right\}, \\ \mathrm{A}_n^{(4)} &= \left\{\frac{\mathrm{T}_n}{n} > 2\right\}.\end{aligned}$$

From the strong law of large numbers, for sufficiently large n, $\mathrm{A}_n^{(4)}$ does not occur. Concerning $\mathrm{A}_n^{(2)}$ and $\mathrm{A}_n^{(3)}$, note that if $x_n/n \to 1$, then $\sup_{[\![1,n]\!]} |x_k - k|/n \to 0$; as $\mathrm{T}_n/n \to 1$ almost surely, there is a.s. an index $n_0(\omega)$ such that neither $\mathrm{A}_n^{(2)}$ nor $\mathrm{A}_n^{(3)}$ occur for $n \geqslant n_0(\omega)$. Finally,

$$\mathbb{P}(\mathrm{A}_n^{(1)}) = \mathbb{P}(\exists (s,t) \in [0,2]^2 : |s-t| \leqslant \delta,\ |\mathrm{B}_s - \mathrm{B}_s| > \varepsilon),$$

converges as $\delta \to 0^+$ by monotone convergence to

$$\mathbb{P}(\cap_{\delta>0}\{\exists (s,t) \in [0,2]^2 : |s-t| \leqslant \delta,\ |\mathrm{B}_s - \mathrm{B}_s| > \varepsilon\}).$$

This event is included in the non absolute continuity of Brownian motion, which is of measure 0 as B is a.s. continuous on the compact $[0,1]$. Hence $\mathbb{P}(\mathrm{A}_n^{(1)}) \to 0$, uniformly in n, as $\delta \to 0^+$, and all together allows to prove (2.11). $\square$

Chapter 3

Semimartingales

In this chapter, we aim to forget Brownian Motion to study more general stochastic processes, irrespectively to their Gaussian or Markovian structure.

For the sake of concision, all processes in this chapter have values in $\mathbb{R}$. All results are true in the $\mathbb{R}^d$ case (and normed vector spaces I guess).

1. Filtrations, processes, stopping times

Definition 3.1. *Given a probability space* $(\Omega, \mathcal{F}, \mathrm{P})$*, a filtration* $(\mathcal{F}_t, t \geqslant 0)$ *is an increasing family of sub-*σ*-algebras of* $\mathcal{F}$*.*

For instance, if $\mathrm{X} = (\mathrm{X}_t, t \geqslant 0)$ is a process (i.e. just a collection of random variables, here indexed by $\mathbb{R}^+$), $\mathcal{F}_t = \sigma(\mathrm{X}_s, s \leqslant t)$ defines a filtration[1]. Given a filtration $(\mathcal{F}_t, t \geqslant 0)$, one can define another one, $(\mathcal{F}_t^+, t \geqslant 0)$, by

$$\mathcal{F}_t^+ = \cap_{s>t} \mathcal{F}_s.$$

In general, both filtrations are not the same[2]. If case of an equality, the filtration is called right-continuous.

A given filtration $(\mathcal{F}_t, t \geqslant 0)$ is called complete if

$$\mathcal{N} = \{\mathrm{A} \in \mathcal{F} \mid \mathrm{P}(\mathrm{A}) = 0\} \subset \mathcal{F}_0.$$

From a filtration $(\mathcal{F}_t, t \geqslant 0)$, one can build its *usual augmentation* by making it complete and right continuous, adding $\mathcal{N}$ to $(\mathcal{F}_t^+, t \geqslant 0)$.

Definition 3.2. *Given a filtration* $(\mathcal{F}_t, t \geqslant 0)$*, a process* $(\mathrm{X}_t, t \geqslant 0)$ *is called adapted if, for any* $t \geqslant 0$*,* X_t *is* $\mathcal{F}_t$*-measurable.*

It is called progressively-measurable if for any $t \geqslant 0$

$$\begin{array}{ccc} [0,t] \times \Omega & \to & \mathbb{R} \\ (s, \omega) & \mapsto & \mathrm{X}_s(\omega) \end{array}$$

is measurable for the σ*-algebra* $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$*.*

The set of $\mathrm{A} \in \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}$ *such that* $\mathbb{1}_{\mathrm{A}}(s, \omega)$ *is progressively measurable is a* σ*-algebra, called the progressive* σ*-algebra.*

It is clear that a progressively-measurable process is adapted. An example of non-progressively-measurable adapted process is the following : if the X_t's are independent Gaussian standard random variables, and $\mathcal{F}_t = \sigma(\mathrm{X}_s, s \leqslant t)$, then obviously X is an adapted process. But it is not progressively-measurable : P-almost surely (hence in particular for some $\omega \in \Omega$) $\{s \leqslant t \mid \mathrm{X}_s(\omega) > 0\}$ is not a Borel subset of $[0, t]$. However, under reasonable continuity assumptions, adapted processes are progressively-measurable.

1. This is usually called the *canonical*, or *natural*, filtration of the process. More precisely, we will use this expression in the following for the *usual augmentation* of the filtration described hereafter.
2. For example, Let $\mathrm{X}_t = \varepsilon t$, where ε is a Bernoulli random variable. Then, if $\mathcal{F}_t = \sigma(\mathrm{X}_s, s \leqslant t)$, ε is $\mathcal{F}_0^+$-measurable but not $\mathcal{F}_0$-measurable

Proposition 3.3. *Let* X *be right-continuous and adapted. Then* X *is progressively-measurable. This is also true if* X *is left-continuous.*

Proof. Let $\mathcal{F}_t^\varepsilon = \mathcal{F}_{t+\varepsilon}$. Then an adapted process is progressively-measurable with respect to $(\mathcal{F}_t, t \geqslant 0)$ if and only if it for any $\varepsilon > 0$ it is progressively-measurable with respect to $(\mathcal{F}_t^\varepsilon, t \geqslant 0)$. Indeed, one implication is obvious because $\mathcal{F}_t \subset \mathcal{F}_t^\varepsilon$. Assume X is progressive with respect to any $(\mathcal{F}_t^\varepsilon, t \geqslant 0)$, $\varepsilon > 0$. This implies that, defining

$$\mathrm{X}_s^\varepsilon = \mathrm{X}_s \mathbb{1}_{s \in [0, t-\varepsilon]} + \mathrm{X}_t \mathbb{1}_{s=t},$$

the application $(s, \omega) \mapsto \mathrm{X}_s^\varepsilon(\omega)$ is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable. As X is the pointwise limit of X^ε on $[0,t]$, this implies that $(s, \omega) \mapsto \mathrm{X}_s(\omega)$, defined on $[0,t] \times \Omega$, is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable[3].

As a consequence, we only need to prove that given $\varepsilon > 0$, X is progressive with respect to $(\mathcal{F}_t^\varepsilon, t \geqslant 0)$. Define the process X^n by

$$\mathrm{X}_t^n = \sum_{k=1}^{\infty} \mathrm{X}_{\frac{k}{n}} \mathbb{1}_{t \in \left[\frac{k-1}{n}, \frac{k}{n}\right[}.$$

If X is right-continuous, X^n converges pointwise to X. Moreover, X^n is progressively-measurable with respect to $(\mathcal{F}_t^\varepsilon, t \geqslant 0)$ if $\varepsilon > 1/n$. This proves that X is progressively-measurable with respect to $(\mathcal{F}_t^\varepsilon, t \geqslant 0)$, concluding the proof. If X is left-continuous, the proof does not require the filtration $(\mathcal{F}_t^\varepsilon, t \geqslant 0)$: X is the pointwise limit of

$$\mathrm{X}_t^n = \sum_{k=1}^{\infty} \mathrm{X}_{\frac{k}{n}} \mathbb{1}_{t \in \left]\frac{k}{n}, \frac{k+1}{n}\right]},$$

and each process X^n is progressively-measurable for $(\mathcal{F}_t, t \geqslant 0)$. □

We now introduce the strict analogue of stopping times of Chapter 1, in the continuous setting. As in the discrete case, a random time is a stopping time if at any time the past allows to determine if it takes values in the past. The σ-algebra associated to a stopping time T is the set of events determined by history till T.

Definition 3.4. *Given a probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ *and a filtration* $(\mathcal{F}_t, t \geqslant 0)$*, a random variable* $\mathrm{T} : \Omega \to \mathbb{R}_+ \cup \{\infty\}$ *is* $(\mathcal{F}_t, t \geqslant 0)$*-stopping time if, for any* $t \geqslant 0$*,* $\{\mathrm{T} \leqslant t\} \in \mathcal{F}_t$*.*

To any stopping time is associated a σ*-algebra*

$$\mathcal{F}_\mathrm{T} = \{\mathrm{A} \in \mathcal{F} \mid \forall t \geqslant 0, \mathrm{A} \cap \{\mathrm{T} \leqslant t\} \in \mathcal{F}_t\}.$$

Checking that this is indeed a σ-algebra is straightforward. Other easy properties are (i) if $\mathrm{S} \leqslant \mathrm{T}$, $\mathcal{F}_\mathrm{S} \subset \mathcal{F}_\mathrm{T}$ and (ii) T is $\mathcal{F}_\mathrm{T}$-measurable. Other useful properties are listed below.

Proposition 3.5. *(i) A random variable* $\mathrm{T} : \Omega \to \mathbb{R}_+ \cup \{\infty\}$ *is a* $(\mathcal{F}_{t+}, t \geqslant 0)$ *stopping time if and only if, for any* $t \geqslant 0$*,* $\{\mathrm{T} < t\} \in \mathcal{F}_t$*.*

(ii) If S *and* T *are stopping times, so are* $\mathrm{S} \wedge \mathrm{T}$ *and* $\mathrm{S} \vee \mathrm{T}$*.*

(iii) If S_n *is an increasing sequence of stopping times, then* $\mathrm{S} = \lim_{n\to\infty} \mathrm{S}_n$ *is also a stopping time.*

(iv) If S_n *is an decreasing sequence of stopping times, then* $\mathrm{S} = \lim_{n\to\infty} \mathrm{S}_n$ *is also a stopping time for the filtration* $(\mathcal{F}_{t+}, t \geqslant 0)$*.*

3. The pointwise limit of a measurable functions is measurable, see e.g. [3]

Proof. To prove (i), assume first that T is a $(\mathcal{F}_{t+}, t \geqslant 0)$-stopping time. Then

$$\{\mathrm{T} < t\} = \cup_{s<t}\{\mathrm{T} \leqslant s\} \in \cup_{s<t}\left(\cap_{u>s}\mathcal{F}_u\right) \subset \mathcal{F}_t\,.$$

Conversely if, for any t, $\{\mathrm{T} < t\} \in \mathcal{F}_t$, then

$$\{\mathrm{T} \leqslant t\} = \cap_{s>t}\{\mathrm{T} < s\} \in \cap_{s>t}\mathcal{F}_s = \mathcal{F}_{t+}\,.$$

Point (ii) follows from

$$\begin{aligned}\{\mathrm{S} \wedge \mathrm{T} \leqslant t\} &= \{\mathrm{S} \leqslant t\} \cup \{\mathrm{T} \leqslant t\} \in \mathcal{F}_t,\\ \{\mathrm{S} \vee \mathrm{T} \leqslant t\} &= \{\mathrm{S} \leqslant t\} \cap \{\mathrm{T} \leqslant t\} \in \mathcal{F}_t,\end{aligned}$$

and (iii), relies on

$$\{\mathrm{S} \leqslant t\} = \cap\{\mathrm{S}_n \leqslant t\} \in \mathcal{F}_t\,.$$

Finally, (iv) follows from

$$\{\mathrm{S} < t\} = \cup\{\mathrm{S}_n < t\} \in \mathcal{F}_t$$

and (i). □

Most of the stopping times we will consider are hitting times of open or closed subsets of a metric space (E, d), where X takes values in E. If X is right-continuous and adapted, and O open in E, then $\mathrm{T_O} = \inf\{t \geqslant 0 \mid \mathrm{X}_t \in \mathrm{O}\}$ is a $(\mathcal{F}_{t+}, t \geqslant 0)$-stopping time. Indeed,

$$\{\mathrm{T_O} < t\} = \cup_{s\in[0,t[\cap\mathbb{Q}}\{\mathrm{X}_s \in \mathrm{O}\} \in \mathcal{F}_t\,.$$

If X is continuous, adapted, and C is closed in E, then $\mathrm{T_C} = \inf\{t \geqslant 0 \mid \mathrm{X}_t \in \mathrm{C}\}$ is a stopping time :

$$\{\mathrm{T_C} \leqslant t\} = \{\inf_{[0,t]} d(\mathrm{X}_s, \mathrm{C}) = 0\} = \{\inf_{[0,t]\cap\mathbb{Q}} d(\mathrm{X}_s, \mathrm{C}) = 0\} \in \mathcal{F}_t\,.$$

More general examples of stopping times can be obtained through the following general result.

Theorem 3.6. *Suppose that the filtration $(\mathcal{F}_t, t \geqslant 0)$ satisfies the usual conditions (right-continuity and completeness). Let $\mathrm{A} \subset \mathbb{R}_+ \times \Omega$ such that the process $(\mathbb{1}_\mathrm{A}(t,\omega), t \geqslant 0)$ is progressively-measurable, and define the beginning of* A *by*

$$\mathrm{B_A}(\omega) = \inf\{t \geqslant 0 \mid (t,\omega) \in \mathrm{A}\}\ (\inf\varnothing = \infty).$$

Then $\mathrm{B_A}$ *is a stopping time.*

Proof. The proof relies on the following difficult result of measure theory[4] : if $(\mathrm{E}, \mathcal{E})$ is a locally compact space with a countable base, endowed with its Borel σ-algebra $\mathcal{E}$, and if $(\Omega, \mathcal{F}, \mathrm{P})$ is a complete (in the sense that $\mathrm{K} \in \mathcal{F}$ and $\mathrm{P}(\mathrm{K}) = 0$ imply that if $\mathrm{L} \subset \mathrm{K}$ then $\mathrm{L} \in \mathcal{F}$) probability space, then for every $\mathrm{K} \in \mathcal{E} \otimes \mathcal{F}$, the projection

$$\pi(\mathrm{K}) = \{\omega \in \Omega \mid \exists e \in \mathrm{E}, (e,\omega) \in \mathrm{A}\}$$

of K on Ω is in $\mathcal{F}$.

One can apply this result, for a given $t \geqslant 0$, for $\mathcal{F} = \mathcal{F}_t$, $\mathrm{E} = [0,t[$ and $\mathrm{K} = \mathrm{A} \cap ([0,t[\times\Omega)$. As $(\mathbb{1}_\mathrm{A}(t,\omega), t \geqslant 0)$ is progressively-measurable, K is $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}_t$-measurable, so $\pi(\mathrm{K}) \in \mathcal{F}_t$. From the definitions we have

$$\{\mathrm{B_A} < t\} = \pi(\mathrm{K}) \in \mathcal{F}_t,$$

so $\mathrm{B_A}$ is a stopping time because the filtration is right-continuous. □

4. A proof can be found in [2]

2. Martingales

Definition 3.7. *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, and $(\mathcal{F}_t, t \geqslant 0)$ a filtration. Then a process $(\mathrm{X}_t)_{t\geqslant 0}$ adapted to $(\mathcal{F}_t, t \geqslant 0)$ is a $\mathbb{P}$-martingale if it satisfies the following conditions, for any $t \geqslant 0$:*

(i) $\mathbb{E}(|\mathrm{X}_t|) < \infty$ *;*

(ii) $\mathbb{E}(\mathrm{X}_t \mid \mathcal{F}_s) = \mathrm{X}_s$ *for any $s \leqslant t$, $\mathbb{P}$-almost surely.*

A submartingale (resp. supermartingale) is defined in the same way, except that $\mathbb{E}(\mathrm{X}_t \mid \mathcal{F}_s) \geqslant \mathrm{X}_s$ (resp. $\mathbb{E}(\mathrm{X}_t \mid \mathcal{F}_s) \leqslant \mathrm{X}_s$).

Examples of martingales are the following. Assume that, given a filtration $(\mathcal{F}_t, t \geqslant 0)$, X is an adapted process with independent increments (i.e. for any $s < t$, $\mathrm{X}_t - \mathrm{X}_s$ is independent of $\mathcal{F}_s$). Then

- if, for any t, $\mathrm{X}_t \in \mathrm{L}^1$, $(\mathrm{X}_t - \mathbb{E}(\mathrm{X}_t), t \geqslant 0)$ is a martingale ;
- if, for any t, $\mathrm{X}_t \in \mathrm{L}^2$, $(\mathrm{X}_t^2 - \mathbb{E}(\mathrm{X}_t^2), t \geqslant 0)$ is a martingale (note that there is no such statement for higher powers of X) ;
- if there is $\lambda > 0$ such that, for any t, $\mathbb{E}(e^{\lambda \mathrm{X}_t}) < \infty$, the following process is a martingale :
$$\left(\frac{e^{\lambda \mathrm{X}_t}}{\mathbb{E}(e^{\lambda \mathrm{X}_t})}, t \geqslant 0\right).$$

The properties of discrete martingales from Chapter 1 have strict counterparts in the continuous setting, given hereafter.

Proposition 3.8. *If X is a $(\mathcal{F}_t, t \geqslant 0)$-submartingale and f is a convex, Lipschitz, nondecreasing function, then $f(\mathrm{X})$ is a $(\mathcal{F}_t, t \geqslant 0)$-submartingale.*

If X is a $(\mathcal{F}_t, t \geqslant 0)$-martingale and f is a convex, Lipschitz function, then $f(\mathrm{X})$ is a $(\mathcal{F}_t, t \geqslant 0)$-submartingale.

Proof. Identical to Proposition 1.2, relying on Jensen's inequality. □

Given a process X, we note $\mathrm{X}_t^* = \sup_{[0,t]} \mathrm{X}_s$.

Theorem 3.9. *Let X be a right-continuous submartingale. Then for any $\lambda > 0$, $t > 0$,*
$$\mathbb{P}\left(\mathrm{X}_t^* \geqslant \lambda\right) \leqslant \frac{\mathbb{E}(\mathrm{X}_t \mathbb{1}_{\mathrm{X}_t^* \geqslant \lambda})}{\lambda} \leqslant \frac{\mathbb{E}(|\mathrm{X}_t|)}{\lambda}.$$
The same result holds if X is left-continuous.

Proof. From Theorem 1.6, the result is true when considering the supremum over finite subsets of $[0, t]$, and then over $[0, t] \cap \mathrm{A}$ for any countable dense subset including t. As X is right (or left) continuous, $\sup_{[0,t]\cap \mathrm{A}} \mathrm{X}_s = \sup_{[0,t]} \mathrm{X}_s$, concluding the proof. □

Theorem 3.10. *Let X be a right-continuous submartingale. Then for any $p > 1$, $t > 0$,*
$$\mathbb{E}\left(|\mathrm{X}_t^*|^p\right) \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}\left(|\mathrm{X}_t|^p\right),$$
may the right member be infinite. The same result holds if X is left-continuous.

Proof. From Theorem 1.7, the result is true when considering the supremum over finite subsets of $[0, t]$, and then over $[0, t] \cap \mathrm{A}$ for any countable dense subset including t. As X is right (or left) continuous, $\sup_{[0,t]\cap \mathrm{A}} \mathrm{X}_s = \sup_{[0,t]} \mathrm{X}_s$, concluding the proof. □

Theorem 3.11. *Let* $(\mathrm{X}_t, t \geqslant 0)$ *be a right-continuous submartingale. Assume that* $\sup_t \mathbb{E}((\mathrm{X}_t)_+) < \infty$*. Then it converges almost surely to some* $\mathrm{X} \in \mathrm{L}^1$*. The same result holds if* X *is left-continuous.*

Proof. In the same way as the proof of Theorem 1.9, using 1.8, for any $a < b$ the number of jumps from a to b along any countable subset of $\mathbb{R}_+$ is almost surely finite (this is proved fo a finite set first and for countable sets by monotone convergence). As a consequence, along $\mathbb{Q}_+$, $(\mathrm{X}_t, t \geqslant 0)$ converges to some X in L^1. The right (or left) continuity assumption allows to state the convergence along $\mathbb{R}_+$. □

Theorem 3.12. *Let* $(\mathrm{X}_t, t \geqslant 0)$ *be a right-continuous uniformly integrable martingale. Then* X_t *converges almost surely and in* $\mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ *to some integrable* $\mathrm{X} \in \mathbb{R}$*, and* $\mathrm{X}_t = \mathbb{E}(\mathrm{X} \mid \mathcal{F}_t)$ *for any* $t \geqslant 0$*. The same result holds if* X *is left-continuous.*

Proof. The proof goes exactly the same way as for Theorem 1.10, in which discretization plays no role. The right or left-continuity are required only to make use of Theorem 3.11. □

For the stopping time theorems, more care is required to adapt the discrete results in a continuous setting. In particular, we prove the bounded case after the uniformly integrable case, a fact clearly contrasting with the discrete case.

Theorem 3.13. *Let* $(\mathrm{X}_t)_{t \geqslant 0}$ *be a right-continuous uniformly integrable martingale, and* T*,* S *two stopping times such that* $\mathrm{S} \leqslant \mathrm{T}$*. Then almost surely*

$$\mathbb{E}(\mathrm{X}_\mathrm{T} \mid \mathcal{F}_\mathrm{S}) = \mathrm{X}_\mathrm{S}.$$

Proof. Let

$$\mathrm{T}_n = \inf\left\{\frac{k}{2^n} \mid k \in \mathbb{N}, \mathrm{T} < \frac{k}{2^n}\right\}.$$

Then one easily checks that the sequence $(\mathrm{T}_n, n \geqslant 0)$ is a sequence of stopping times decreasing to T. From Theorem 1.11,

$$\mathrm{X}_{\mathrm{T}_{n+1}} = \mathbb{E}(\mathrm{X}_{\mathrm{T}_{n+1}} \mid \mathcal{F}_{\mathrm{T}_n}),$$

so the sequence defined by $\mathrm{Z}_{-n} = \mathrm{X}_{\mathrm{T}_n}$ is an inverse martingale with respect to the filtration $(\mathcal{F}_{\mathrm{T}_n}, n \in \mathbb{N})$. From Theorem 1.12, this inverse martingale converges almost surely and in L^1. For $\mathrm{S} \leqslant \mathrm{T}$ another stopping time, defining in the same way

$$\mathrm{S}_n = \inf\left\{\frac{k}{2^n} \mid k \in \mathbb{N}, \mathrm{S} < \frac{k}{2^n}\right\},$$

we know from Theorem 1.12 that $\mathrm{X}_{\mathrm{S}_n} = \mathbb{E}(\mathrm{X}_{\mathrm{T}_n} \mid \mathcal{F}_{\mathrm{S}_n})$, so for any $\mathrm{A} \in \mathcal{F}_\mathrm{S} \subset \mathcal{F}_{\mathrm{S}_n}$,

$$\mathbb{E}(\mathbb{1}_\mathrm{A} \mathrm{X}_{\mathrm{S}_n}) = \mathbb{E}(\mathbb{1}_\mathrm{A} \mathrm{X}_{\mathrm{T}_n}).$$

The convergence in L^1 of $\mathrm{X}_{\mathrm{S}_n}$ (resp $\mathrm{X}_{\mathrm{T}_n}$) to X_S (resp. X_T) by right-continuity allows to conclude that for any $\mathrm{A} \in \mathcal{F}_\mathrm{S}$, $\mathbb{E}(\mathbb{1}_\mathrm{A} \mathrm{X}_\mathrm{S}) = \mathbb{E}(\mathbb{1}_\mathrm{A} \mathrm{X}_\mathrm{T})$, the expected result. □

Note that in particular, the previous proof shows that for any stopping time T, under the above hypotheses $\mathrm{X}_\mathrm{T} \in \mathrm{L}^1$.

Theorem 3.14. *Let* $(\mathrm{X}_t)_{t \geqslant 0}$ *be a right-continuous uniformly integrable martingale, and* T*,* S *two stopping times such that* $\mathrm{S} \leqslant \mathrm{T} < c$ *a.s. for some constant* c*. Then almost surely*

$$\mathbb{E}(\mathrm{X}_\mathrm{T} \mid \mathcal{F}_\mathrm{S}) = \mathrm{X}_\mathrm{S}.$$

Proof. One can apply Theorem 3.13 to $(\mathrm{X}_{t \wedge c}, t \geqslant 0)$, which is right-continuous and uniformly integrable because for $t \leqslant c$, $\mathrm{X}_t = \mathbb{E}(\mathrm{X}_c \mid \mathcal{F}_t)$. This gives the expected result. □

We finish this martingale section with the stability of (uniformly integrable) martingales when frozen after a stopping time. This natural result actually requires the stopping time theorems.

Theorem 3.15. *Let* T *be a stopping time.*

(i) If $(\mathrm{X}_t, t \geqslant 0)$ *is a uniformy integrable martingale, so is* $(\mathrm{X}_{t\wedge \mathrm{T}}, t \geqslant 0)$.

(ii) If $(\mathrm{X}_t, t \geqslant 0)$ *is a martingale, so is* $(\mathrm{X}_{t\wedge \mathrm{T}}, t \geqslant 0)$.

Proof. To prove (i), note that the equality

$$\mathrm{X}_{t\wedge \mathrm{T}} = \mathbb{E}(\mathrm{X}_\mathrm{T} \mid \mathcal{F}_t) \tag{3.1}$$

would be sufficient. Indeed, it would first imply uniform integrability, as any set of random variables of type $(\mathbb{E}(\mathrm{X}_\mathrm{T} \mid \mathcal{F}_t), t \geqslant 0)$ with $\mathrm{X}_\mathrm{T} \in \mathrm{L}^1$ is uniformly integrable (indeed, $\mathrm{X}_\mathrm{T} \in \mathrm{L}^1$ from Theorem 3.13). Moreover, the martingale property would follow from the calculation

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

To prove (3.1), note first that by Theorem 3.13,

$$\mathrm{X}_{t\wedge \mathrm{T}} = \mathbb{E}(\mathrm{X}_\mathrm{T} \mid \mathcal{F}_{t\wedge \mathrm{T}}) = \mathbb{E}(\mathrm{X}_\mathrm{T}\mathbb{1}_{\mathrm{T}\leqslant t} \mid \mathcal{F}_{t\wedge \mathrm{T}}) + \mathbb{E}(\mathrm{X}_\mathrm{T}\mathbb{1}_{\mathrm{T}>t} \mid \mathcal{F}_{t\wedge \mathrm{T}}).$$

The first term is also $\mathbb{E}(\mathrm{X}_\mathrm{T}\mathbb{1}_{\mathrm{T}\leqslant t} \mid \mathcal{F}_t)$ because $\mathrm{X}_\mathrm{T}\mathbb{1}_{\mathrm{T}\leqslant t}$ is both $\mathcal{F}_t$ and $\mathcal{F}_{t\wedge \mathrm{T}}$-measurable. The second term is also $\mathbb{E}(\mathrm{X}_\mathrm{T}\mathbb{1}_{\mathrm{T}>t} \mid \mathcal{F}_t)$ because for any $\mathrm{A} \in \mathcal{F}_t$, $\mathbb{1}_\mathrm{A}\mathbb{1}_{\mathrm{T}>t} \in \mathcal{F}_{t\wedge \mathrm{T}}$, so

$$\mathbb{E}(\mathrm{X}_\mathrm{T}\mathbb{1}_{\mathrm{T}>t}\mathbb{1}_\mathrm{A}) = \mathbb{E}(\mathbb{E}(\mathrm{X}_\mathrm{T} \mid \mathcal{F}_{\mathrm{T}\wedge t})\mathbb{1}_{\mathrm{T}>t}\mathbb{1}_\mathrm{A}) = \mathbb{E}(\mathbb{E}(\mathrm{X}_\mathrm{T}\mathbb{1}_{\mathrm{T}>t} \mid \mathcal{F}_{\mathrm{T}\wedge t})\mathbb{1}_\mathrm{A}).$$

This proves (3.1). To prove (ii), note that for any $c > 0$ the process $(\mathrm{X}_{t\wedge c}, t \geqslant 0)$ is a uniformly integrable martingale, so from (i) the process $(\mathrm{X}_{t\wedge c\wedge \mathrm{T}}, t \geqslant 0)$ is also a uniformly integrable martingale. In particular, for any $s < t$,

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

Choosing $c > s \vee t$ concludes the proof. □

As a consequence of the above stopping time theorems, hitting times by Brownian motion are well understood. In the following, B is a Brownian motion beginning at 0. For $x \in \mathbb{R}$, let $\mathrm{T}_x = \inf\{t \geqslant 0 \mid \mathrm{B}_t = x\}$.

Corollary 3.16. *For any* $a < 0 < b$,

$$\mathbb{P}(\mathrm{T}_a < \mathrm{T}_b) = \frac{b}{b-a} = 1 - \mathbb{P}(\mathrm{T}_a > \mathrm{T}_b).$$

Proof. Note first that both T_a and T_b are a.s. finite, because the Brownian motion is a.s. recurrent :

$$\mathbb{P}(\mathrm{T}_a < \mathrm{T}_b) + \mathbb{P}(\mathrm{T}_a > \mathrm{T}_b) = 1. \tag{3.2}$$

Moreover, from Proposition 3.5 and Theorem 3.15, $(\mathrm{B}_{t\wedge \mathrm{T}_a\wedge \mathrm{T}_b}, t \geqslant 0)$ is a martingale. As it is bounded, it is uniformly integrable and Theorem 3.13 applies, with the stopping time $\mathrm{T} = \infty$ (or $\mathrm{T}_a \wedge \mathrm{T}_b$) : in particular, $\mathbb{E}(\mathrm{B}_{\mathrm{T}_a\wedge \mathrm{T}_b}) = \mathrm{B}_0$, i.e. using (3.2)

$$a\mathbb{P}(\mathrm{T}_a < \mathrm{T}_b) + b\mathbb{P}(\mathrm{T}_a > \mathrm{T}_b) = 0. \tag{3.3}$$

Equations (3.2) and (3.3) together give the result. □

Corollary 3.17. *The hitting time* T_a *has the following Laplace transform : for any* $\lambda \geqslant 0$,

$$\mathbb{E}\left(e^{-\lambda \mathrm{T}_a}\right) = e^{-|a|\sqrt{2\lambda}}.$$

If $a \geqslant 0$ *and* $\mathrm{S}_a = \inf\{t \geqslant 0 \mid |\mathrm{B}_t| = a\}$, *then*

$$\mathbb{E}\left(e^{-\lambda \mathrm{S}_a}\right) = \frac{1}{\cosh(a\sqrt{2\lambda})}.$$

Proof. The process

$$\mathrm{M}_t^{(\mu)} = e^{\mu \mathrm{B}_t - \frac{\mu^2}{2}t},\ t \geqslant 0,$$

is a well-known martingale, which is bounded when stopped at T_a, and therefore uniformly bounded. The stopping time theorem applied to $(\mathrm{M}^{(\mu)}_{t\wedge \mathrm{T}_a}, t \geqslant 0)$ at time $\mathrm{T} = \infty$ (or T_a) therefore yields $\mathbb{E}(\mathrm{M}^{\mu}_{\mathrm{T}_a}) = 1$, i.e.

$$\mathbb{E}\left(e^{-\frac{\mu^2}{2}\mathrm{T}_a}\right) = e^{-\mu a}$$

because the Brownian motion is recurrent : $\mathrm{B}_{\mathrm{T}_a} = a$ almost surely. Up to a change of variables, the above equation is the expected result. Concerning S_a, look at the martingale

$$\mathrm{N}_t = \mathrm{M}_t^{(\mu)} + \mathrm{M}_t^{(-\mu)},\ t \geqslant 0.$$

Then the same reasoning yields $\mathbb{E}(\mathrm{N}_{\mathrm{S}_a}) = 2$, i.e.

$$\left(e^{\mu a} + e^{-\mu a}\right)\mathbb{E}\left(e^{-\frac{\mu^2}{2}\mathrm{S}_a}\right) = 2,$$

which gives the result by $\lambda = \mu^2/2$. □

The following application is more striking as it does not depend on any structure of the martingale M. This is a first glimpse of universality which will be proved in the next chapter through the Dubins-Schwarz theorem.

Corollary 3.18. *Let* M *be a positive continuous martingale beginning at* $\mathrm{M}_0 > 0$, *and converging to 0 almost surely*[5]*. Then*

$$\frac{\mathrm{M}_0}{\sup_{t\geqslant 0}\mathrm{M}_t}$$

is uniform on $[0,1]$*. As a consequence, for any* $\mu \neq 0$, $\sup_{t\geqslant 0}\left(\mu \mathrm{B}_t - \frac{\mu^2}{2}t\right)$ *has the same law as an exponential random variable with parameter 1.*

Proof. Take $a > \mathrm{M}_0$ and note here $\mathrm{T}_a = \inf\{t \geqslant 0 \mid \mathrm{M}_t = a\}$ ($\inf \varnothing = \infty$). Then the martingale $(\mathrm{M}_{t\wedge \mathrm{T}_a}, t \geqslant 0)$ is bounded, hence uniformly bounded, hence the stopping time theorem yields, for $\mathrm{T} = \mathrm{T}_a$ (or ∞),

$$a\mathbb{P}(\mathrm{T}_a < \infty) = \mathrm{M}_0,$$

because M is continuous and M_T equals 0 on $\{\mathrm{T} = \infty\}$. The above equation can be written

$$\mathbb{P}\left(\frac{\mathrm{M}_0}{\sup_{t\geqslant 0}\mathrm{M}_t} \leqslant \frac{\mathrm{M}_0}{a}\right) = \frac{\mathrm{M}_0}{a},$$

5. There are many examples of such martingales, like $(e^{\mathrm{B}_t - \frac{t}{2}}, t \geqslant 0)$, or any exponential martingale, cf Chapter 4.

which is the first statement. Concerning the second, note that $\left(e^{\mu \mathrm{B}_t - \frac{\mu^2}{2}t}, t \geqslant 0\right)$ is a continuous martingale beginning at 1 and converging to 0 almost surely. As a consequence, $\sup_{t\geqslant 0} e^{\mu \mathrm{B}_t - \frac{\mu^2}{2}t}$ is distributed as the inverse of a uniform random variable on $[0,1]$, noted U, so for any $\lambda > 0$

$$\mathbb{P}\left(\sup_{t\geqslant 0}\left(\mu \mathrm{B}_t - \frac{\mu^2}{2}t\right) > \lambda\right) = \mathbb{P}\left(\log\left(\frac{1}{\mathrm{U}}\right) > \lambda\right) = e^{-\lambda},$$

the expected result. □

3. Finite variation processes

The stochastic processes we will consider will have an oscillating (of martingale type) part and a finite variation component. This section reviews some basic properties of this last part, which is always assumed continuous[6].

Definition 3.19. *Given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$, a process* $\mathrm{A} : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ *is called a finite variation process if all the following conditions are satisfied :*

1. *For any ω,* A *is continuous.*
2. *For any ω,* $\mathrm{A}_0 = 0$.
3. *For any ω, there is a signed measure*[7] *μ such that for any $t \geqslant 0$*

$$\mathrm{A}_t = \mu([0,t]).$$

Note that the continuity and initial value assumptions implies in particular that μ has no atoms. Moreover, this definition and the following results could also be given in an almost sure sense. This makes no fundamental difference in the following of the course, by defining processes to be constantly 0 on the set of ω's without the required property : this will not change adaptedness as we will consider complete filtrations.

Note that the decomposition $\mu = \mu_+ - \mu_-$ as the difference between two positive measures is not unique. However, it is unique when constrained to

$$\operatorname{supp}(\mu_+) \cap \operatorname{supp}(\mu_-) = \varnothing.$$

Indeed, the uniqueness of such a decomposition follows from the necessary identity $\mu_+(\mathrm{B}) = \sup\{\mu(\mathrm{C}) \mid \mathrm{C} \subset \mathrm{B},\ \mathrm{C} \text{ Borel set}\}$. Concerning the existence, write first $\mu = \tilde{\mu}_+ - \tilde{\mu}_-$, for some positive measures $\tilde{\mu}_+$ and $\tilde{\mu}_-$. Then $\tilde{\mu}_+$ (resp. $\tilde{\mu}_-$) is absolutely continuous with respect to $\tilde{\mu} = \tilde{\mu}_+ + \tilde{\mu}_-$, so by the Radon-Nikodym theorem it has a density $\lambda_+(t)$ (resp. $\lambda_-(t)$) with respect to $\tilde{\mu}$. Then, the choice

$$\mu_+(\mathrm{d}t) = \max(\lambda_+(t) - \lambda_-(t), 0)\tilde{\mu}(\mathrm{d}t)$$
$$\mu_-(\mathrm{d}t) = \max(\lambda_-(t) - \lambda_+(t), 0)\tilde{\mu}(\mathrm{d}t)$$

gives the expected decomposition. Note S_+ (resp. S_-) the support of μ_+ (resp. μ_-), and $|\mu| = \mu_+ + \mu_-$. Then $\mathrm{d}\mu/\mathrm{d}|\mu| = \mathbb{1}_{\mathrm{S}_+} - \mathbb{1}_{\mathrm{S}_-}$.

Moreover, for a finite variation process, $\mathrm{A}(t) = mu_+([0,t]) - mu_-([0,t])$. As A is continuous, μ_+ and μ_- have no atoms (because they have disjoint supports), so A is

6. One can extend the notion of finite variation processes to the discontinuous case, but this is not our purpose. Note that a contrario, in the preceding section we proved martingale properties also in discontinuous cases, but this will be necessary when building the stochastic integral in the next chapter.
7. This is the difference of two positive measures with finite mass on any compact.

the difference of two continuous increasing functions beginning at 0. This proves that for any $t > 0$

$$\sup_{0=t_0<\dots<t_n=t} \sum_{k=1}^{n} |\mathrm{A}_{t_k} - \mathrm{A}_{t_{k-1}}| < \infty,$$

where the supremum is over all $n \geqslant 0$ and subdivisions of $[0,t]$. The above result justifies the name *finite variation process*. A less obvious fact is that the above supremum is the $|\mu|$-measure of $[0,t]$. The proof is a beautiful example of a martingale argument.

Theorem 3.20. *Let* A *be a finite variation process. Then for any* $t > 0$*,*

$$\sup_{0=t_0<\dots<t_n=t} \sum_{k=1}^{n} |\mathrm{A}_{t_k} - \mathrm{A}_{t_{k-1}}| = |\mu|([0,t]),$$

where the supremum is over all $n \geqslant 0$ *and subdivisions of* $[0,t]$.

Proof. The inequality $\sup_{0=t_0<\dots<t_n=t} \sum_{k=1}^{n} |\mathrm{A}_{t_k} - \mathrm{A}_{t_{k-1}}| \leqslant |\mu|([0,t])$ is obvious because

$$|\mathrm{A}_{t_k} - \mathrm{A}_{t_{k-1}}| = |\mu(]t_{k-1}, t_k])| \leqslant |\mu|(]t_{k-1}, t_k].$$

Consider now any sequence of refined subdivisions of $[0,t]$ with step going to 0, noted $0 = t_0^{(n)} < \dots < t_{p_n}^{(n)} = t$, and the filtration $\mathcal{F}_0 \subset \dots \subset \mathcal{F}_n \subset \dots \subset \mathcal{B}([0,t])$ defined as subsets of the Borel algebra by

$$\mathcal{F}_n = \sigma(]t_{k-1}^{(n)}, t_k^{(n)}], 1 \leqslant k \leqslant p_n).$$

This is indeed a filtration because the subdivisions are refined. Take $\Omega = [0,t]$, and the probability measure

$$\mathbb{P}(\mathrm{d}s) = \frac{|\mu|(\mathrm{d}s)}{|\mu|([0,t])}$$

on Ω. On $(\Omega, \mathcal{B}([0,t]), (\mathcal{F}_n)_{n\geqslant 0}, \mathbb{P})$, look at the random variables

$$\mathrm{X}(s) = \mathbb{1}_{\mathrm{S}_+}(s) - \mathbb{1}_{\mathrm{S}_-}(s),$$

$$\mathrm{X}_n(s) = \mathbb{E}(\mathrm{X} \mid \mathcal{F}_n)(s) = \frac{\mu(]t_{k-1}^{(n)}, t_k^{(n)}])}{|\mu|(]t_{k-1}^{(n)}, t_k^{(n)}])} = \frac{\mathrm{A}_{t_k^{(n)}} - \mathrm{A}_{t_{k-1}^{(n)}}}{|\mu|(]t_{k-1}^{(n)}, t_k^{(n)}])} \text{ when } s \in]t_{k-1}^{(n)}, t_k^{(n)}],$$

these last equalities being easily verified by definition of the conditional expectation. As $(\mathrm{X}_n, n \geqslant 0)$ is a bounded martingale, it converges almost surely and in L^1 to some $\mathrm{Y} \in \mathrm{L}^1$, and $\mathrm{X}_n = \mathbb{E}(\mathrm{Y} \mid \mathcal{F}_n)$. As a consequence, for any n, $\mathbb{E}(\mathrm{X} - \mathrm{Y} \mid \mathcal{F}_n) = 0$. As X and Y are in $\bigvee_n \mathcal{F}_n$ (concerning X, this is a consequence of the time step going to 0), this implies $\mathrm{X} = \mathrm{Y}$ almost surely. Hence $\mathrm{X}_n \to \mathrm{X}$ in L^1, so in particular $\mathbb{E}(|\mathrm{X}_n|) \to \mathbb{E}(|\mathrm{X}|)$, which means that

$$\sum_{k=1}^{p_n} |\mathrm{A}_{t_k^{(n)}} - \mathrm{A}_{t_{k-1}^{(n)}}| \underset{n\to\infty}{\longrightarrow} |\mu|([0,t]),$$

concluding the proof. □

If A is a finite variation process and $\mathrm{F} : [0,t] \to \mathbb{R}$ a process, measurable for any given ω, such that $\int_0^t |f(s)||\mu|(\mathrm{d}s)$ is finite, then we define

$$\int_0^t f(s)\mathrm{d}\mathrm{A}_s = \int_0^t f(s)\mu(\mathrm{d}s)$$

$$\int_0^t f(s)|\mathrm{d}\mathrm{A}_s| = \int_0^t f(s)|\mu|(\mathrm{d}s).$$

Then the above definitions are compatible with sums of Riemann type for continuous f for example.

Proposition 3.21. *Let* A *be a finite variation process and* $\mathrm{F} : \Omega \times [0,t] \to \mathbb{R}$ *a left-continuous process. Then for any* ω

$$\int_0^t \mathrm{F}(s)\mathrm{dA}_s = \lim_{n\to\infty} \sum_{k=1}^{p_n} \mathrm{F}(t_{k-1}^{(n)})(\mathrm{A}_{t_k^{(n)}} - \mathrm{A}_{t_{k-1}^{(n)}}), \tag{3.4}$$

$$\int_0^t \mathrm{F}(s)|\mathrm{dA}_s| = \lim_{n\to\infty} \sum_{k=1}^{p_n} \mathrm{F}(t_{k-1}^{(n)})|\mathrm{A}_{t_k^{(n)}} - \mathrm{A}_{t_{k-1}^{(n)}}|, \tag{3.5}$$

for any sequence of subdivisions of $[0,t]$, $0 = t_0^{(n)} < \cdots < t_{p_n}^{(n)} = t$, *with step going to 0. For (3.5), we require the subdivisions to be refined.*

Proof. Let F_n be the process defined as $\mathrm{F}(t_{k-1}^{(n)})$ on $]t_{k-1}^{(n)}, t_k^{(n)}]$. Then, the right hand side of (3.4) is $\int_0^t \mathrm{F}_n(s)\mu(\mathrm{d}s)$, so the result follows by dominated convergence.

Concerning (3.5),

$$\left|\sum_{k=1}^{p_n} \mathrm{F}(t_{k-1}^{(n)})|\mathrm{A}_{t_k^{(n)}} - \mathrm{A}_{t_{k-1}^{(n)}}| - \int_0^t \mathrm{F}_n(s)|\mathrm{dA}_s|\right| \\ \leqslant \|\mathrm{F}\|_{\mathrm{L}^\infty[0,1]} \left(|\mu|([0,t]) - \sum_{k=1}^{p_n} |\mathrm{A}_{t_k^{(n)}} - \mathrm{A}_{t_{k-1}^{(n)}}|\right),$$

and from the proof of Theorem 3.20 this converges to 0 along any refined sequence of subdivisions with step going to 0. Hence proving that

$$\int_0^t \mathrm{F}_n(s)|\mathrm{dA}_s| \underset{n\to\infty}{\longrightarrow} \int_0^t \mathrm{F}(s)|\mathrm{dA}_s$$

is sufficient, and true by dominated convergence. □

The refinement of the subdivisions is actually not necessary for (3.5) : exercise! Moreover, note that the result of the above theorem is true when changing the evaluation of $\mathrm{F}(t_{k-1}^{(n)})$ from $\mathrm{F}(t_k^{(n)})$. This property will be false when considering stochastic integrals in Chapter 4.

Finally, finite variation processes are stable in the following sense.

Proposition 3.22. *Given a filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$, *let* A *be a finite variation process and* F *progressively-measurable such that for any* ω, $t \geqslant 0$, $\int_0^t |\mathrm{F}_s||\mathrm{dA}_s| < \infty$. *Then*

$$\mathrm{F}\cdot\mathrm{A} : (\omega, t) \mapsto \int_0^t \mathrm{F}_s(\omega)\mathrm{dA}_s(\omega)$$

is a finite variation process.

Proof. Let μ be the signed measure associated to A. The process $\mathrm{F}\cdot\mathrm{A}$ begins at 0, is continuous and has bounded variation because

$$(\mathrm{F}\cdot\mathrm{A})_t = \tilde\mu([0,t]),\ \tilde\mu(\mathrm{d}s) = \mathrm{H}_s\mu(\mathrm{d}s),$$

with $\tilde\mu$ a signed measure with no atoms (the finite mass condition holds thanks to $\int_0^t |\mathrm{F}_s||\mathrm{dA}_s| < \infty$). Consequently, the only condition to verify carefully is the adaptedness of $\mathrm{F}\cdot\mathrm{A}$. It is true that $(\mathrm{F}\cdot\mathrm{A})_t$ is $\mathcal{F}_t$-measurable if F is of type $\mathbb{1}_{]u,v]}(s)\mathbb{1}_\mathrm{A}(\omega)$ where $u, v \leqslant t$ and $\mathrm{A} \in \mathcal{F}_t$. It is then true if $\mathrm{F} = \mathbb{1}_\mathrm{A}$ for any $\mathrm{A} \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t$, by the monotone class theorem. Finally, by taking linear combinations of such sums approximating F from below, and using dominated convergence (domination by an integrable process holds as $\int_0^t |\mathrm{F}_s||\mathrm{dA}_s| < \infty$), we get the result for general F, as the pointwise limit of measurable functions is measurable. □

4. Local martingales

Instead of considering only martingales, we focus on a more general class of processes, defined below. The main reason is that together with finite variation processes, they will have stability properties by composition (cf Chapter 4), a property not shared by martingales. In the following, a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ is given, that satisfies the usual conditions.

Definition 3.23. *A process* $(\mathrm{M}_t)_{t\geqslant 0}$ *is called a local martingale beginning at 0 if all the following conditions are satisfied.*

(i) *It is adapted.*

(ii) *For all* ω, $\mathrm{M}_0 = 0$.

(iii) *For all* ω, M *is continuous.*

(iv) *There exists a sequence of stopping times* T_n *converging to* ∞ *for any* ω *such that, for all* n, $\mathrm{M}^{\mathrm{T}_n} := (\mathrm{M}_{t\wedge \mathrm{T}_n}, t \geqslant 0)$, *is a uniformly integrable martingale.*

For such a sequence of stopping times, one says that $(\mathrm{T}_n, n \geqslant 0)$ *reduces* M. *A process* M *is called a local martingale if* $\mathrm{M}_t = \mathrm{M}_0 + \mathrm{N}_t$, *where* $\mathrm{M}_0 \in \mathcal{F}_0$ *and* N *is a local martingale beginning at 0.*

Note that one could give a similar definition of a local martingale in the discrete setting. However, for such a definition, a discrete local martingale X is a martingale if and only if, for any n, $\mathrm{X}_n \in \mathrm{L}^1$. This characterization is not true in the continuous setting, allowing local martingales to have much more exotic structures[8].

Proposition 3.24. *For a given filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$, *all the following statements hold.*

(i) *Any continuous martingale is a local martingale.*

(ii) *In Definition 3.23, condition (iv) can be equivalently replaced by : there exists a sequence of stopping times* T_n *converging to* ∞ *for any* ω *such that, for all* n, $\mathrm{M}^{\mathrm{T}_n} := (\mathrm{M}_{t\wedge \mathrm{T}_n}, t \geqslant 0)$, *is a martingale.*

(iii) *If* M *is a local martingale and* T *is a stopping time, then* $\mathrm{M}^{\mathrm{T}} = (\mathrm{M}_{t\wedge \mathrm{T}}, t \geqslant 0)$ *is a local martingale.*

(iv) *If* M *is a local martingale,* $(\mathrm{T}_n)_{n\geqslant 0}$ *reduces* M, *and* $(\mathrm{S}_n)_{n\geqslant 0}$ *are stopping times converging to* ∞, *then* $(\mathrm{S}_n \wedge \mathrm{T}_n)_{n\geqslant 0}$ *reduces* M.

(v) *The set of local martingales (e.g. with values in* $\mathbb{R}$*) is a (e.g. real) vector space.*

(vi) *If* M *is a nonnegative local martingale and* $\mathrm{M}_0 \in \mathrm{L}^1$, *then* M *is a supermartingale.*

(vii) *If* M *is a local martingale and, for all* $t \geqslant 0$, $|\mathrm{M}_t| \leqslant \mathrm{X}$ *where* $\mathrm{X} \in \mathrm{L}^1$, *then* M *is a martingale.*

(viii) *If* M *is a local martingale beginning at 0, then* $\mathrm{T}_n = \inf\{t \geqslant 0 \mid |\mathrm{M}_t| = n\}$ *reduces* M.

Proof. Point (i) follows from the possible choice $\mathrm{T}_n = n$: for any constant $c \geqslant 0$, $(\mathrm{M}_{t\wedge c})$ is uniformly integrable, as all of its values are of type $\mathbb{E}(\mathrm{M}_c \mid \mathcal{G})$ for some σ-algebra $\mathcal{G}$ and $\mathrm{M}_c \in \mathrm{L}^1$. For point (ii), note that if $\mathrm{T}_n \to \infty$ and $\mathrm{M}^{\mathrm{T}_n}$ is a martingale, then $\mathrm{M}^{\mathrm{T}_n \wedge n}$ is uniformly integrable, as shown for (i), and $\mathrm{T}_n \wedge n \to \infty$. To prove (iii) and (iv), note that by Theorem 3.15, if $\mathrm{M}^{\mathrm{T}_n}$ is a uniformly integrable martingale, so

8. See Chapter 4 for an example of a local martingale which is L^1-bounded but not a martingale.

is $M^{T_n\wedge T}$. The stability by addition mentioned in (v) is a direct consequence of (iv), by choosing $(T_n)_{n\geqslant 0}$ reducing the first martingale and $(S_n)_{n\geqslant 0}$ reducing the second. Point (vi) is a consequence of Fatou's lemma : if $M = M_0 + N$ and $(T_n)_{n\geqslant 0}$ reduces N, then

$$\mathbb{E}(M_t \mid \mathcal{F}_s) = \mathbb{E}(\lim_n M_{t\wedge T_n} \mid \mathcal{F}_s) \leqslant \liminf_n \mathbb{E}(M_{t\wedge T_n} \mid \mathcal{F}_s) = \liminf_n M_{s\wedge T_n} = M_s.$$

Note that M_t is in L^1 precisely thanks to the above equation. The result (vii) relies on dominated convergence applied to the identity

$$M_{s\wedge T_n} = \mathbb{E}(M_{t\wedge T_n} \mid \mathcal{F}_s),$$

where $(T_n)_{n\geqslant 0}$ reduces M. Finally, $(viii)$ is a direct consequence of (ii) and (vii). □

The following result states that local martingales and finite variation processes are disjoint, up to the constant processes. Some weaker statements are intuitive and easy to prove. For example, if M is $\mathscr{C}^1$ and there exist $c > 0$ such that $\sup_{[0,1]} |M'| < c$ almost surely, then M cannot be a nontrivial martingale : the martingale property gives for any $\varepsilon > 0$

$$\mathbb{E}\left(\frac{M_{t+\varepsilon} - M_t}{\varepsilon} \mid \mathcal{F}_t\right) = 0,$$

so by dominated convergence $\mathbb{E}(M'_t \mid \mathcal{F}_t) = 0$. But M'_t is $\mathcal{F}_t$-measurable (by considering increasing rates on the left now), so $M'_t = 0$ almost surely, and as M' is continuous M is constant almost surely. This result extends to finite variation processes[9].

Theorem 3.25. *Let* M *be a local martingale beginning at* 0*. If* M *is a finite variation process, then* M *is indistinguishable from 0.*

Proof. Assume M is a finite variation process, and chose

$$T_n = \inf\{t \geqslant 0 \mid \int_0^t |dM_s| \geqslant n\}.$$

Then $T_n \to \infty$ and T_n is a stopping time, for example by Theorem 3.6. The local martingale M^{T_n} is bounded by n, so it is a martingale by Proposition 3.24. As a consequence, for any subdivision $0 = t_0 < \dots < t_p = t$,

$$\begin{aligned}\mathbb{E}\left((M_t^{T_n})^2\right) &= \sum_{k=1}^p \mathbb{E}\left((M_{t_k}^{T_n})^2 - (M_{t_{k-1}}^{T_n})^2\right) = \sum_{k=1}^p \mathbb{E}\left((M_{t_k}^{T_n} - M_{t_{k-1}}^{T_n})^2\right)\\ &\leqslant \mathbb{E}\left(\max_\ell \left|M_{t_\ell}^{T_n} - M_{t_{\ell-1}}^{T_n}\right| \sum_{k=1}^p \left|M_{t_k}^{T_n} - M_{t_{k-1}}^{T_n}\right|\right) \leqslant n\,\mathbb{E}\left(\max_\ell \left|M_{t_\ell}^{T_n} - M_{t_{\ell-1}}^{T_n}\right|\right).\end{aligned}$$

As this maximum is bounded by n and M has continuous trajectories, dominated convergence allows to conclude that

$$\mathbb{E}\left((M_t^{T_n})^2\right) = 0,$$

by choosing subdivisions with time step going to 0. By Fatou's lemma, one can take the $n \to \infty$ limit to conclude $\mathbb{E}(M_t^2) = 0$, so $M_t = 0$ almost surely. As M is continuous, this is equivalent to being indistinguishable from 0. □

9. Pessimistic people will conclude that that there are no physically reasonable fair games. Optimistic people will conclude that, as fair games exist, infinite variation processes exist.

5. Bracket

Theorem 3.25 proved that nontrivial local martingales have infinite variation. But rescaling the increments, there is a process measuring the oscillations that is not trivial : the good scale consists in considering the square of the increments ; this limiting process is defined in the important theorem hereafter.

Theorem 3.26. *Let* M *be a local martingale. Then there exists a unique (up to indistinguishability) increasing continuous variation process process, noted* $\langle \mathrm{M}, \mathrm{M}\rangle$, *such that* $(\mathrm{M}_t^2 - \langle \mathrm{M}, \mathrm{M}\rangle_t, t \geqslant 0)$ *is a local martingale. Moreover, if* $(0 = t_0^{(n)} < t_1^{(n)} < \dots, n \geqslant 0)$ *is any sequence of subdivisions of* $\mathbb{R}_+$ *with step going to 0, then*

$$\langle \mathrm{M}, \mathrm{M}\rangle_t = \lim_{n\to\infty} \sum_{k\geqslant 1} \left(\mathrm{M}_{t_k^{(n)}\wedge t} - \mathrm{M}_{t_{k-1}^{(n)}\wedge t}\right)^2,$$

uniformly in the sense of convergence in probability[10]. *The process* $\langle \mathrm{M}, \mathrm{M}\rangle$, *often noted* $\langle \mathrm{M}\rangle$, *is called the bracket (or quadratic variation) of* M.

As an example, if B is a Brownian motion then $(\mathrm{B}_t^2 - t, t \geqslant 0)$ is a martingale, so $\langle \mathrm{B}\rangle_t = t$. Note that in general the bracket is not deterministic.

Proof. Uniqueness of quadratic variation is an easy consequence of Theorem 3.25

We will first prove the existence of the bracket when M is a true martingale, and $|\mathrm{M}|$ almost surely bounded by some $\mathrm{K} > 0$. For a subdivision $\delta = \{0 = t_0 < t_1 \dots\}$ and a process Y, we note (Q for quadratic)

$$\mathrm{Q}_t^{(\mathrm{Y},\delta)} = \sum_{k\geqslant 1} \left(\mathrm{Y}_{t_k^{(n)}\wedge t} - \mathrm{Y}_{t_{k-1}^{(n)}\wedge t}\right)^2,$$

By a simple calculation[11]

$$\mathrm{X}_t^{(\delta)} := \mathrm{M}_t^2 - \mathrm{Q}_t^{(\mathrm{M},\delta)} = 2\sum_{k\geqslant 1} \mathrm{M}_{t_{k-1}^{(n)}} \left(\mathrm{M}_{t_k^{(n)}\wedge t} - \mathrm{M}_{t_{k-1}^{(n)}\wedge t}\right). \tag{3.6}$$

Therefore, $(\mathrm{X}_t^{(\delta)}, t \geqslant 0)$ is a continuous martingale (just like in the case of Proposition 1.4). For a sequence $(\delta_n, n \geqslant 0)$ of subdivisions with step going to 0, we want to find a subsequence of $(\mathrm{X}^{(\delta_n)}, n \geqslant 0)$ converging uniformly on compact sets. Note

$$\Delta_t^{(n,m)} = \mathrm{X}_t^{(\delta_n)} - \mathrm{X}_t^{(\delta_m)} = \mathrm{Q}_t^{(\mathrm{M},\delta_m)} - \mathrm{Q}_t^{(\mathrm{M},\delta_n)},$$

which is a martingale, so

$$\left(\left(\Delta_t^{(n,m)}\right)^2 - \mathrm{Q}_t^{(\Delta^{(n,m)},\delta_n\cup\delta_m)}, t \geqslant 0\right)$$

is a martingale as well, by the same decomposition used to prove that $\mathrm{X}^{(\delta)}$ is a martingale. As a consequence, the expectation of $\left(\Delta_t^{(n,m)}\right)^2$ is also a discrete analogue of the quadratic variation of a finite variation process. We therefore expect this to go

10. This means that for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{s\in[0,t]} \left|\langle \mathrm{M}, \mathrm{M}\rangle_s - \sum_{k\geqslant 1} \left(\mathrm{M}_{t_k^{(n)}\wedge s} - \mathrm{M}_{t_{k-1}^{(n)}\wedge s}\right)^2\right| > \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

11. For any numbers $a_k's$, $a_n^2 - \sum_{k=1}^n (a_k - a_{k-1})^2 = 2\sum_{k=1}^n a_{k-1}(a_k - a_{k-1})$.

to 0, which would prove that the sequence of the $(\mathrm{X}_t^{(\delta_n)}, n \geqslant 0)$ is a Cauchy sequence in L^2, hence converging. Let us prove it. Note that, as $(a-b)^2 \leqslant 2(a^2+b^2)$,

$$\mathrm{Q}_t^{(\mathrm{A}-\mathrm{B},\delta)} \leqslant 2\left(\mathrm{Q}_t^{(\mathrm{A},\delta)} + \mathrm{Q}_t^{(\mathrm{B},\delta)}\right),$$

so in order to prove that $\mathbb{E}\left(\left(\Delta_t^{(n,m)}\right)^2\right)$ converges to 0, a sufficient condition is

$$\mathbb{E}\left(\mathrm{Q}_t^{(\mathrm{Q}^{(\mathrm{M},\delta_n)},\delta_n\cup\delta_m)}\right) \underset{n,m\to\infty}{\longrightarrow} 0. \tag{3.7}$$

Note ε_n the supremum of $|\mathrm{M}_u - \mathrm{M}_v|$ over all u, v, in $[0,t]$ such that $u-v$ is smaller than the time step of δ_n. Then if s_{k-1} and s_k are successive elements of $\delta_n \cup \delta_m$, then $|\mathrm{Q}_{s_k}^{(\mathrm{M},\delta_n)} - \mathrm{Q}_{s_{k-1}}^{(\mathrm{M},\delta_n)}| \leqslant \varepsilon_n |\mathrm{M}_{s_k} - \mathrm{M}_{s_{k-1}}|$, hence

$$\mathrm{Q}_t^{(\mathrm{Q}^{(\mathrm{M},\delta_n)},\delta_n\cup\delta_m)} \leqslant \varepsilon_n^2 \sum_{k\geqslant 1}\left(\mathrm{M}_{s_k\wedge t} - \mathrm{M}_{s_{k-1}\wedge t}\right)^2.$$

As the quadratic increments is uniformly bounded[12] in L^2 by $8\mathrm{K}^4$, we get by the Cauchy-Schwarz inequality

$$\mathbb{E}\left(\mathrm{Q}_t^{(\mathrm{Q}^{(\mathrm{M},\delta_n)},\delta_n\cup\delta_m)}\right) \leqslant \left(8\mathrm{K}^4\,\mathbb{E}(\varepsilon_n^4)\right)^{1/2}. \tag{3.8}$$

By dominated convergence ($\varepsilon_n \to 0$ a.s. and $\varepsilon_n \leqslant 2\mathrm{K}$), this goes to 0 as $n\to\infty$. We have therefore proved (3.7), so $\Delta_t^{(n,m)}$ converges to 0 in L^2 as $n,m\to\infty$. By Doob's inequality, this implies that

$$\mathbb{E}\left(\left(\sup_{[0,t]}(\mathrm{X}^{(\delta_n)} - \mathrm{X}^{(\delta_m)})\right)^2\right) \underset{n,m\to\infty}{\longrightarrow} 0,$$

so there is a subsequence of the $\mathrm{X}^{(\delta_n)}$'s converging almost surely, uniformly, on $[0,t]$. Let X denote this (continuous) limit. As the subsequence of the $\mathrm{X}^{(\delta_n)}$'s converges to X, in L^2, their martingale property is preserved in the limit : X is a martingale. Moreover, from the definition of $\mathrm{Q}_t^{(\mathrm{M},\delta)}$, $\mathrm{M}^2 - \mathrm{X}^{(\delta)}$ is an increasing process. This property remains for $\mathrm{M}^2 - \mathrm{X}$ by uniform convergence. For $s\in[0,t]$, define

$$\langle \mathrm{M}\rangle_s = \mathrm{M}_s^2 - \mathrm{X}_s.$$

From the previous discussion, $\langle \mathrm{M}\rangle$ satisfies all required properties of the bracket on $[0,t]$. By uniqueness of the bracket, the value $\langle \mathrm{M}\rangle_s$ is independent of the choice of the horizon $t \geqslant s$, and of the choice of the subsequence providing uniform convergence. Moreover, the above reasoning has proved that $\mathrm{X}_s^{(\delta_n)} - \mathrm{X}_s^{(\delta_m)}$ is a Cauchy sequence in L^2, and as ε_n in (3.8) can be chosen identical for any choice of $s\in[0,t]$, the convergence is in L^2 and uniform on compact sets.

12. To prove this, first check for any subdivision $\delta = \{0 = t_0 < t_1 < \dots\}$ the identity

$$\left(\mathrm{Q}_t^{(\mathrm{M},\delta)}\right)^2 = \sum_{k\geqslant 1}(\mathrm{M}_{s_k\wedge t} - \mathrm{M}_{s_{k-1}\wedge t})^4 + 2\sum_{k\geqslant 1}\left(\mathrm{Q}_{s_k\wedge t}^{(\mathrm{M},\delta)} - \mathrm{Q}_{s_{k-1}\wedge t}^{(\mathrm{M},\delta)}\right)\left(\mathrm{Q}_t^{(\mathrm{M},\delta)} - \mathrm{Q}_{s_k\wedge t}^{(\mathrm{M},\delta)}\right).$$

As $\mathrm{X}^{(\delta)}$ is a martingale, $\mathbb{E}(\mathrm{Q}_t^{(\mathrm{M},\delta)} - \mathrm{Q}_{s_k\wedge t}^{(\mathrm{M},\delta)} \mid \mathcal{F}_{s_k\wedge t}) = \mathbb{E}(\mathrm{M}_t^2 - \mathrm{M}_{s_k\wedge t}^2 \mid \mathcal{F}_{s_k\wedge t})$, so using $|\mathrm{M}| \leqslant \mathrm{K}$

$$\begin{aligned}\mathbb{E}\left(\left(\mathrm{Q}_t^{(\mathrm{M},\delta)}\right)^2\right) &\leqslant 4\mathrm{K}^2\left(\mathbb{E}\left(\mathrm{Q}_t^{(\mathrm{M},\delta)}\right) + \sum_{k\geqslant 1}\left(\mathrm{Q}_{s_k\wedge t}^{(\mathrm{M},\delta)} - \mathrm{Q}_{s_{k-1}\wedge t}^{(\mathrm{M},\delta)}\right)\right)\\ &= 8\mathrm{K}^2\,\mathbb{E}\left(\mathrm{Q}_t^{(\mathrm{M},\delta)}\right) = 8\mathrm{K}^2\,\mathbb{E}\left(\mathrm{M}_t^2\right) \leqslant 8\mathrm{K}^4.\end{aligned}$$

This bounded martingale case extends easily. First, note that if the result is true for local martingales beginning at 0, it is true for local martingales : if $\mathrm{M}_t = \mathrm{M}_0 + \mathrm{N}_t$ with $\mathrm{M}_0 \in \mathcal{F}_0$ and N a local martingale beginning at 0, as $\mathrm{M}_0\mathrm{N}_t$ is a local martingale, so is $\mathrm{M}_t^2 - \langle \mathrm{N}\rangle_t = \mathrm{N}_t^2 - \langle \mathrm{N}\rangle_t + \mathrm{M}_0^2 + 2\mathrm{M}_0\mathrm{N}_t$.

Hence we can assume $\mathrm{M}_0 = 0$. We localize M by $\mathrm{T}_n = \inf\{t \geqslant 0 \mid |\mathrm{M}_t| = n\}$. Then, by Proposition 3.24, $\mathrm{M}^{\mathrm{T}_n}$ is a local martingale, bounded by n, so we can apply the previous study : there is an increasing process, noted $\langle \mathrm{M}\rangle^{(n)}$, such that $(\mathrm{M}^{\mathrm{T}_n})^2 - \langle \mathrm{M}\rangle^{(n)}$ is a martingale. By uniqueness, for $m \leqslant n$, $(\langle \mathrm{M}\rangle^{(n)})^{\mathrm{T}_m} = \langle \mathrm{M}\rangle^{(m)}$. Thanks to this coherence property, one can define a a process $\langle \mathrm{M}\rangle$ such that for any n, $(\mathrm{M}^{\mathrm{T}_n})^2 - \langle \mathrm{M}\rangle^{\mathrm{T}_n}$ is a martingale. As $\mathrm{T}_n \to \infty$ almost surely, this means that $\mathrm{M}^2 - \langle \mathrm{M}\rangle$ is a local martingale.

Concerning the uniform convergence of quadratic increments to the bracket, this property is true for $\mathrm{M}^{\mathrm{T}_n})^2 - \langle \mathrm{M}\rangle^{\mathrm{T}_n}$ in L^2. This with the asymptotics (by dominated or monotone convergence)

$$\mathbb{P}(\mathrm{T}_n \leqslant t) \underset{n\to\infty}{\longrightarrow} 0$$

allows to conclude that, uniformly, the convergence holds in probability. □

Note that equation (3.6) can be read in the limit as $\mathrm{M}_t^2 - \langle \mathrm{M}\rangle_t = 2\int_0^t \mathrm{M}_t \mathrm{d}\mathrm{M}_t$. This is the first example of an Itô-type formula, from which we will derive more general stochastic calculus rules in Chapter 4.

As we defined the bracket of one local martingale, we can define the quadratic variation for two local martingales by polarization.

Definition 3.27. *Let* M *and* N *be two local martingales. Then*

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

Thanks to Theorem 3.26, the reader will easily check the following properties of the bracket :

(i) Up to indistinguishability, $\langle \mathrm{M}, \mathrm{N}\rangle$ is the unique finite variation process such that $\mathrm{MN} - \langle \mathrm{M}, \mathrm{N}\rangle$ is a local martingale.

(ii) The function $(\mathrm{M}, \mathrm{N}) \mapsto \langle \mathrm{M}, \mathrm{N}\rangle$ is symmetric, bilinear.

(iii) If $(0 = t_0^{(n)} < t_1^{(n)} < \dots, n \geqslant 0)$ is any sequence of subdivisions of $\mathbb{R}_+$ with step going to 0, then

$$\langle \mathrm{M}, \mathrm{N}\rangle_t = \lim_{n\to\infty} \sum_{k\geqslant 1} \left(\mathrm{M}_{t_k^{(n)}\wedge t} - \mathrm{M}_{t_{k-1}^{(n)}\wedge t}\right)\left(\mathrm{N}_{t_k^{(n)}\wedge t} - \mathrm{N}_{t_{k-1}^{(n)}\wedge t}\right), \tag{3.9}$$

in probability, uniformly for t in compact sets.

(iv) For any stopping time T, $t \geqslant 0$, $\langle \mathrm{M}, \mathrm{N}\rangle_{t\wedge \mathrm{T}} = \langle \mathrm{M}^{\mathrm{T}}, \mathrm{N}\rangle_t = \langle \mathrm{M}^{\mathrm{T}}, \mathrm{N}^{\mathrm{T}}\rangle_t$.

As an example, consider on the same probability space two independent Brownian motions B_1 and B_2. Then $\mathrm{B}_1\mathrm{B}_2$ is a local martingale, so $\langle \mathrm{B}_1, \mathrm{B}_2\rangle = 0$. If $\mathrm{B} = \mathrm{B}_1$ and $\tilde{\mathrm{B}} = \rho\mathrm{B}_1 + \sqrt{1-\rho^2}\mathrm{B}_2$ for some $\rho \in [-1, 1]$, bilinearity implies

$$\langle \mathrm{B}, \tilde{\mathrm{B}}\rangle_t = \rho t.$$

As the bracket is a finite variation process, one can consider, as in Section 3, integrals with respect to it. Then these integrals with respect to cross and diagonal brackets are related by the following inequality, by Kumita and Watanabe.

Theorem 3.28. *Let* H, K *be two progressively-measurable*[13] *processes and* M, N, *two local martingales. Then, for any* $t \in \mathbb{R}_+ \cup \{\infty\}$,

$$\int_0^t |\mathrm{H}_s \mathrm{K}_s|\, |\mathrm{d}\langle \mathrm{M}, \mathrm{N}\rangle_s| \leqslant \left(\int_0^t \mathrm{H}_s^2 \,\mathrm{d}\langle \mathrm{M}\rangle_s\right)^{1/2} \left(\int_0^t \mathrm{K}_s^2 \,\mathrm{d}\langle \mathrm{N}\rangle_s\right)^{1/2},$$

may some terms be infinite.

Proof. First note that, by the approximation (3.9) and the Cauchy-Schwarz inequality, for any given $s < t$,

$$|\langle \mathrm{M}, \mathrm{N}\rangle_t - \langle \mathrm{M}, \mathrm{N}\rangle_s| \leqslant (\langle \mathrm{M}\rangle_t - \langle \mathrm{M}\rangle_s)^{1/2} (\langle \mathrm{N}\rangle_t - \langle \mathrm{N}\rangle_s)^{1/2},$$

almost surely. As these are continuous processes, this inequality holds almost surely for any $s < t$. Using Cauchy-Schwarz again, for $s = t_0 < \dots < t_n = t$, the above inequality yields

$$\begin{aligned}
\sum_{k=1}^n \left|\langle \mathrm{M}, \mathrm{N}\rangle_{t_k} - \langle \mathrm{M}, \mathrm{N}\rangle_{t_{k-1}}\right| &\leqslant \sum_{k=1}^n \left(\langle \mathrm{M}\rangle_{t_k} - \langle \mathrm{M}\rangle_{t_{k-1}}\right)^{1/2} \left(\langle \mathrm{N}\rangle_{t_k} - \langle \mathrm{N}\rangle_{t_{k-1}}\right)^{1/2} \\
&\leqslant \left(\sum_{k=1}^n \left(\langle \mathrm{M}\rangle_{t_k} - \langle \mathrm{M}\rangle_{t_{k-1}}\right)\right)^{1/2} \left(\sum_{k=1}^n \left(\langle \mathrm{N}\rangle_{t_k} - \langle \mathrm{N}\rangle_{t_{k-1}}\right)\right)^{1/2} \\
&= (\langle \mathrm{M}\rangle_t - \langle \mathrm{M}\rangle_s)^{1/2} (\langle \mathrm{N}\rangle_t - \langle \mathrm{N}\rangle_s)^{1/2},
\end{aligned}$$

so using Theorem 3.20, we get

$$\int_s^t |\mathrm{d}\langle \mathrm{M}, \mathrm{N}\rangle_u| \leqslant \left(\int_s^t \mathrm{d}\langle \mathrm{M}\rangle_u\right)^{1/2} \left(\int_s^t \mathrm{d}\langle \mathrm{N}\rangle_u\right)^{1/2}$$

This inequality can be extended to any Borel bounded set B in $\mathbb{R}_+$, first through the finite number of intervals case (still by Cauchy-Schwarz) and then by monotone classes :

$$\int_{\mathrm{B}} |\mathrm{d}\langle \mathrm{M}, \mathrm{N}\rangle_u| \leqslant \left(\int_{\mathrm{B}} \mathrm{d}\langle \mathrm{M}\rangle_u\right)^{1/2} \left(\int_{\mathrm{B}} \mathrm{d}\langle \mathrm{N}\rangle_u\right)^{1/2}.$$

Now, for functions of type $\mathrm{H} = \sum h_\ell \mathbb{1}_{\mathrm{B}_\ell}$, $\mathrm{K} = \sum k_\ell \mathbb{1}_{\mathrm{B}_i}$, with disjoint bounded Borel sets B_i's,

$$\begin{aligned}
\int |\mathrm{H}_s \mathrm{K}_s|\, |\mathrm{d}\langle \mathrm{M}, \mathrm{N}\rangle_s| &= \sum_\ell |h_\ell k_\ell| \int_{\mathrm{B}_\ell} |\mathrm{d}\langle \mathrm{M}, \mathrm{N}\rangle_u| \\
&\leqslant \sum_\ell |h_\ell k_\ell| \left(\int_{\mathrm{B}_\ell} \mathrm{d}\langle \mathrm{M}\rangle_u\right)^{1/2} \left(\int_{\mathrm{B}_\ell} \mathrm{d}\langle \mathrm{N}\rangle_u\right)^{1/2} \\
&\leqslant \left(\sum_\ell h_\ell^2 \int_{\mathrm{B}_\ell} \mathrm{d}\langle \mathrm{M}\rangle_u\right)^{1/2} \left(\sum_\ell k_\ell^2 \int_{\mathrm{B}_\ell} \mathrm{d}\langle \mathrm{N}\rangle_u\right)^{1/2} \\
&= \left(\int_0^t \mathrm{H}_s^2 \,\mathrm{d}\langle \mathrm{M}\rangle_s\right)^{1/2} \left(\int_0^t \mathrm{K}_s^2 \,\mathrm{d}\langle \mathrm{N}\rangle_s\right)^{1/2}.
\end{aligned}$$

Approximation of progressively-measurable processes as increasing limit of such functions ends the proof. □

13. The proof actually does not require adaptedness : $(\omega, t) \mapsto \mathrm{H}_t(\omega)$ being $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}$-measurable (and the same for K) is sufficient, but in the following we will always apply it in the progressively-measurable case.

Another important property of the bracket is that it characterizes integrability of local martingales.

Theorem 3.29. *Let* M *be a local martingale,* $\mathrm{M}_0 = 0$.

(i) The process M *is a* L^2*-bounded martingale*[14] *if and only if* $\mathbb{E}(\langle \mathrm{M}\rangle_\infty) < \infty$. *In such a case,* $\mathrm{M}^2 - \langle \mathrm{M}\rangle$ *is a uniformly integrable martingale.*

(ii) The process M *is a square integrable martingale*[15] *if and only if for any* $t \geqslant 0$, $\mathbb{E}(\langle \mathrm{M}\rangle_t) < \infty$. *In such a case,* $\mathrm{M}^2 - \langle \mathrm{M}\rangle$ *is a martingale.*

Proof. For (i) assume first that M is a L^2-bounded martingale. It is therefore uniformly integrable and converges almost surely to some M_∞. Moreover, from Doob's inequality,

$$\mathbb{E}\left(\sup_{t\geqslant 0} \mathrm{M}_t^2\right) \leqslant 4 \sup_{t\geqslant 0} \mathbb{E}\left(\mathrm{M}_t^2\right) < \infty. \tag{3.10}$$

As a consequence, if we define $\mathrm{T}_n = \inf\{t \geqslant 0 \mid \langle \mathrm{M}\rangle_t \geqslant n\}$, (this is easily shown to be a stopping time because the bracket is progressively-measurable)

$$\left(\mathrm{M}^2_{t\wedge \mathrm{T}_n} - \langle \mathrm{M}\rangle_{t\wedge \mathrm{T}_n}, t \geqslant 0\right)$$

is a local martingale (by Proposition 3.24) bounded by $\left(\sup_{t\geqslant 0} \mathrm{M}_t^2\right) + n \in \mathrm{L}^1$, so it is a true martingale (still by Proposition 3.24), so

$$\mathbb{E}\left(\langle \mathrm{M}\rangle_{t\wedge \mathrm{T}_n}\right) = \mathbb{E}\left(\mathrm{M}^2_{t\wedge \mathrm{T}_n}\right).$$

Dominated convergence allows to take $t \to \infty$ on the right hand side, and monotone convergence on the left hand side :

$$\mathbb{E}\left(\langle \mathrm{M}\rangle_{\mathrm{T}_n}\right) = \mathbb{E}\left(\mathrm{M}^2_{\mathrm{T}_n}\right).$$

Now, monotone convergence on the left and dominated convergence on the right yield

$$\mathbb{E}\left(\langle \mathrm{M}\rangle_{\infty}\right) = \mathbb{E}\left(\mathrm{M}^2_{\infty}\right).$$

In particular, $\mathbb{E}\left(\langle \mathrm{M}\rangle_{\infty}\right)$ is finite. This implies that $\mathrm{M}^2 - \langle \mathrm{M}\rangle$ is bounded by an integrable random variable $(\left(\sup_{t\geqslant 0} \mathrm{M}_t^2\right) + \langle \mathrm{M}\rangle_\infty)$, so it is a uniformly integrable martingale.

Suppose now that $\mathbb{E}\left(\langle \mathrm{M}\rangle_{\infty}\right) < \infty$ and note $\tilde{\mathrm{T}}_n = \inf\{t \geqslant 0 \mid |\mathrm{M}_t| \geqslant n\}$. Then $\mathrm{M}^{\tilde{\mathrm{T}}_n}$ (bounded by n) and $(\mathrm{M}^{\tilde{\mathrm{T}}_n})^2 - \langle \mathrm{M}\rangle^{\tilde{\mathrm{T}}_n}$ (bounded by $n^2 + \langle \mathrm{M}\rangle^{\mathrm{T}_n} \in \mathrm{L}^1$) are uniformly integrable martingales. Hence, for any stopping time S, by Theorem 3.13,

$$\mathbb{E}\left(\mathrm{M}^2_{\mathrm{S}\wedge\tilde{\mathrm{T}}_n}\right) = \mathbb{E}(\langle \mathrm{M}\rangle_{\mathrm{S}\wedge\tilde{\mathrm{T}}_n}).$$

Fatou's lemma therefore yields

$$\mathbb{E}(\mathrm{M}^2_{\mathrm{S}}) \leqslant \mathbb{E}(\langle \mathrm{M}\rangle_{\mathrm{S}}) \leqslant \mathbb{E}(\langle \mathrm{M}\rangle_{\infty}) < \infty.$$

In particular, M is L^2-bounded. Moreover, it is a martingale : in the identity

$$\mathbb{E}(\mathrm{M}_{t\wedge\tilde{\mathrm{T}}_n} \mid \mathcal{F}_s) = \mathrm{M}_{s\wedge\tilde{\mathrm{T}}_n},$$

the limit $n \to \infty$ is allowed by dominated convergence. Indeed, the $\mathrm{M}_{t\wedge\tilde{\mathrm{T}}_n}$'s are bounded in L^2 ($\mathbb{E}(\sup_t \mathrm{M}_t^2) < \infty$, so $\mathbb{E}(\sup_n \mathrm{M}^2_{t\wedge\tilde{\mathrm{T}}_n}) < \infty$), hence in L^1.

For (ii), by Doob's inequality, if M is square integrable, $(\mathrm{M}_{s\wedge t}, c \geqslant 0)$ is L^2-bounded so, by (i), $\mathbb{E}(\langle \mathrm{M}\rangle_t) < \infty$. Reciprocally, if $\mathbb{E}(\langle \mathrm{M}\rangle_t) < \infty$ then (i) implies that $(\mathrm{M}_{s\wedge t}, c \geqslant 0)$ is bounded in L^2, in particular $\mathbb{E}(\mathrm{M}_t^2) < \infty$. Finally, in such a case, from (i) the process $(\mathrm{M}^2_{s\wedge t} - \langle \mathrm{M}\rangle_{s\wedge t}, s \geqslant 0)$ is a (uniformly integrable) martingale, so $\mathrm{M}^2 - \langle \mathrm{M}\rangle$ is a martingale. □

14. $\sup_{t\geqslant 0} \mathbb{E}(|\mathrm{M}_t|^2) < \infty$
15. For any $t \geqslant 0$, $\mathbb{E}(|\mathrm{M}_t|^2) < \infty$.

Part (i) of Theorem 3.29 has the following easy consequence.

Corollary 3.30. *Let* M *be a local martingale, with* $\mathrm{M}_0 = 0$ *almost surely. Then* M *is indistinguishable from 0 if and only if* $\langle \mathrm{M} \rangle$ *is identically 0.*

Proof. If M is indistinguishable from 0, it is clear that the bracket vanishes (as a limit of quadratic increments). Reciprocally, if $\langle \mathrm{M} \rangle \equiv 0$, then by (i) in Theorem 3.29 M^2 is a martingale so, for any given t, $\mathbb{E}(\mathrm{M}_t^2) = 0$, so $\mathrm{M}_t = 0$ almost surely. One can conclude by continuity that M is indistinguishable from 0. $\square$

Finally, the class of processes we will consider in the next chapter, and for which a stochastic calculus will be developed, are of the following type, gathering the types we have considered till now.

Definition 3.31. *Given a probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geqslant 0}, \mathbb{P})$*, a process* X *is called a semimartingale if it is of type*

$$\mathrm{X} = \mathrm{X}_0 + \mathrm{M} + \mathrm{A}$$

where $\mathrm{X}_0 \in \mathcal{F}_0$*,* M *is a local martingale beginning at 0 and* A *is a finite variation process.*

Note that, by Theorem 3.25, such a decomposition is unique, up to indistinguishability. As a consequence, one can define without ambiguity the bracket of $\mathrm{X} = \mathrm{X}_0 + \mathrm{M} + \mathrm{A}$ and $\tilde{\mathrm{X}} = \tilde{\mathrm{X}}_0 + \tilde{\mathrm{M}} + \tilde{\mathrm{A}}$ as

$$\langle \mathrm{X}, \tilde{\mathrm{X}} \rangle = \langle \mathrm{M}, \tilde{\mathrm{M}} \rangle.$$

In particular, the bracket of a finite variation process with any semimartingale is always 0. Then the reader can prove that the bracket is still given as a limit of increment (the finite variation part does not contribute) : if $(0 = t_0^{(n)} < t_1^{(n)} < \dots, n \geqslant 0)$ is any sequence of subdivisions of $\mathbb{R}_+$ with step going to 0, then in the sense of convergence in probability

$$\langle \mathrm{X}, \tilde{\mathrm{X}} \rangle_t = \lim_{n \to \infty} \sum_{k \geqslant 1} \left(\mathrm{X}_{t_k^{(n)} \wedge t} - \mathrm{X}_{t_{k-1}^{(n)} \wedge t} \right) \left(\tilde{\mathrm{X}}_{t_k^{(n)} \wedge t} - \tilde{\mathrm{X}}_{t_{k-1}^{(n)} \wedge t} \right). \tag{3.11}$$

After these many efforts to define semimartingales, the reader may reasonably wonder whether this is a natural class of processes to consider. One possible reason for paying attention to these processes is that they are stable when composed with smooth function (of class $\mathscr{C}^2$) : this is part of the next chapter, through the famous Itô formula. Another reason is that, up to a deterministic part, processes with independent increments are semimartingales : the proof of the following result (that we state only in the continuous case) can be found in [7].

Theorem 3.32. *If, given a probability space, the process* Y *is continuous with independent increments, then it takes the form*

$$\mathrm{Y} = \mathrm{X} + \mathrm{F},$$

where X *is a semimartingale with independent increments and* F *is a deterministic continuous function.*

Chapter 4

The Itô formula and applications

The purpose of this chapter is to give a rigorous meaning to

$$\int \mathrm{H}_s \mathrm{dM}_s$$

where H is a progressively measurable process and M is a local martingale. Once such integrals defined, calculus rules are given for composing sufficiently smooth functions with semimartingales (Itô's formula), allowing in particular to prove that semimartingales is a class of processes stable when composed with $\mathscr{C}^2$ functions. Consequences of this in terms of occupation properties of Brownian motions, partial differential equations (the Dirichlet problem) and change of measure on the Wiener space (Girsanov's theorem) are then developed.

1. The stochastic integral

Definition 4.1. *We note* H^2 *the set of (continuous)* L^2*-bounded martingales :* $\mathrm{M} \in \mathrm{H}^2$ *if* $\sup_{t \geqslant 0} \mathbb{E}((\mathrm{M}_t)^2) < \infty$.

Note that, if $\mathrm{M} \in \mathrm{H}^2$, it converges almost surely and has a finite bracket : $\mathbb{E}(\langle \mathrm{M} \rangle_\infty) < \infty$. Moreover, as a simple application of the Kunita Watanabe inequality, if M and N are in H^2,

$$\mathbb{E}(|\langle \mathrm{M}, \mathrm{N} \rangle_\infty|) \leqslant \mathbb{E}\left(\langle \mathrm{M} \rangle_\infty\right)^{1/2} \mathbb{E}\left(\langle \mathrm{N} \rangle_\infty\right)^{1/2}.$$

This means that we can define a scalar product on H^2 by

$$(\mathrm{M}, \mathrm{N})_{\mathrm{H}^2} = \mathbb{E}\left(\langle \mathrm{M}, \mathrm{N} \rangle_\infty\right),$$

and $\|\mathrm{M}\|_{\mathrm{H}^2} = (\langle \mathrm{M} \rangle_\infty)^{1/2}$ defines a norm, associated to this scalar product : we have already seen that if $\|\mathrm{M}\|_{\mathrm{H}^2} = 0$, then M is indistinguishable from 0. An important point is that the space $(\mathrm{H}^2, \|\|_{\mathrm{H}^2})$ is complete.

Proposition 4.2. *The space* $(\mathrm{H}^2, \|\|_{\mathrm{H}^2})$ *is a Hilbert space.*

Proof. We only still need to prove that this space is complete. Consider a Cauchy sequence $(\mathrm{M}^{(n)})_{n \geqslant 0}$:

$$\lim_{m,n \to \infty} \mathbb{E}\left((\mathrm{M}^{(n)}_\infty - \mathrm{M}^{(m)}_\infty)^2\right) = \lim_{m,n \to \infty} \mathbb{E}\left(\langle \mathrm{M}^{(n)} - \mathrm{M}^{(m)} \rangle_\infty\right) = 0.$$

Hence, by the Doob inequality,

$$\lim_{m,n \to \infty} \mathbb{E}\left(\sup_{t \geqslant 0} |\mathrm{M}^{(n)}_t - \mathrm{M}^{(m)}_t|^2\right) = 0, \tag{4.1}$$

so we can find an increasing sequence $(n_k)_{k \geqslant 0}$ such that

$$\mathbb{E}\left(\sum_{k=1}^{\infty} \sup_{t \geqslant 0} |\mathrm{M}^{(n_k)}_t - \mathrm{M}^{(n_{k-1})}_t|\right) \leqslant \sum_{k=1}^{\infty} \mathbb{E}\left(\sup_{t \geqslant 0} |\mathrm{M}^{(n_k)}_t - \mathrm{M}^{(n_{k-1})}_t|^2\right)^{1/2} < \infty. \tag{4.2}$$

As a consequence, $\sum_{k=1}^{\infty} \sup_{t\geqslant 0} |\mathrm{M}_t^{(n_k)} - \mathrm{M}_t^{(n_{k-1})}|$ is almost surely finite, so $\mathrm{M}^{(n_k)}$ converges uniformly to some continuous adapted (the pointwise limit of measurable functions is measurable) process M. For any given s and t, as $\mathrm{M}_t^{(n_k)}$ (resp. $\mathrm{M}_s^{(n_k)}$) converges in L^2 to M_t (resp. M_s), in the martingale property

$$\mathbb{E}(\mathrm{M}_t^{(n_k)} \mid \mathcal{F}_s) = \mathrm{M}_s^{(n_k)},$$

one can take the limits to conclude that M is a martingale. Moreover, as the $\mathrm{M}^{(n_k)}$'s satisfy (4.1), all $\mathrm{M}_t^{(n_k)}$'s are uniformly bounded in L^2 and $\mathrm{M} \in \mathrm{H}^2$. Finally, $\mathrm{M}^{(n_k)}$ converges to M in H^2, because

$$\mathbb{E}\left(\langle \mathrm{M}^{(n_k)} - \mathrm{M}\rangle_\infty\right) = \mathbb{E}\left((\mathrm{M}_\infty^{(n_k)} - \mathrm{M}_\infty)^2\right) \to 0,$$

e.g. by (4.2). This implies, by the Cauchy condition, that $\mathrm{M}^{(n)}$ converges to M in H^2 as well. □

Definition 4.3. *For* $\mathrm{M} \in \mathrm{H}^2$*, let* $\mathrm{L}^2(\mathrm{M})$ *be the space of progressively measurable processes* H *such that*

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

Note that, as an L^2 space (more precisely, one can easily check that $\mathrm{L}^2(\mathrm{M}) = \mathrm{L}^2(\mathbb{R}_+ \times \Omega, \mathcal{F}, \nu)$, with $\mathcal{F}$ is the progressive σ-algebra and $\nu(\mathrm{A}) = \mathbb{E}(\int_0^\infty \mathbb{1}_\mathrm{A}(s,\cdot)\mathrm{d}\langle \mathrm{M}\rangle_s)$ is a well-defined finite measure), this is a Hilbert space for

$$(\mathrm{H}, \mathrm{K})_{\mathrm{L}^2(\mathrm{M})} = \mathbb{E}\left(\int_0^\infty \mathrm{H}_s \mathrm{K}_s \mathrm{d}\langle \mathrm{M}\rangle_s\right)$$

in the sense that $\|\mathrm{H}\|_{\mathrm{L}^2(\mathrm{M})} = 0$ if and only if ν-almost surely $\mathrm{H} = 0$. Note also that in the above definition, H is not necessarily continuous.

Definition 4.4. *The vector subspace of* $\mathrm{L}^2(\mathrm{M})$ *consisting in step processes is noted* $\mathcal{E}$*. More precisely,* $\mathrm{H} \in \mathcal{E}$ *if there is some* $p \geqslant 1$ *and* $0 = t_0 < \cdots < t_p$ *such that*

$$\mathrm{H}_s(\omega) = \sum_{k=0}^{p-1} \mathrm{H}_k(\omega) \mathbb{1}_{]t_k, t_{k+1}]}(s),$$

where $\mathrm{H}_k \in \mathcal{F}_{t_k}$ *is bounded.*

Proposition 4.5. *For any* $\mathrm{M} \in \mathrm{H}^2$*,* $\mathcal{E}$ *is dense in* $\mathrm{L}^2(\mathrm{M})$*.*

Proof. We need to prove that if $\mathrm{K} \in \mathrm{L}^2(\mathrm{M})$ is orthogonal to $\mathcal{E}$, then $\mathrm{K} = 0$. If K is orthogonal to $\mathrm{F}\mathbb{1}_{]s,t]} \in \mathcal{E}$, where $\mathrm{F} \in \mathcal{F}_s$ is bounded, then

$$\mathbb{E}\left(\mathrm{F}\int_s^t \mathrm{K}_u \mathrm{d}\langle \mathrm{M}\rangle_u\right) = 0. \tag{4.3}$$

Let $\mathrm{X}_t = \int_0^t \mathrm{K}_u \mathrm{d}\langle \mathrm{M}\rangle_u$. Then $\mathrm{X}_t \in \mathrm{L}^1$, as a consequence of the Cauchy-Schwarz inequality and $\mathrm{M} \in \mathrm{H}^2$, $\mathrm{K} \in \mathrm{L}^2(\mathrm{M})$. Then (4.3) means that, for any bounded $\mathrm{F} \in \mathcal{F}_s$, $\mathbb{E}((\mathrm{X}_t - \mathrm{X}_s)\mathrm{F}) = 0$, so $\mathrm{X}_s = \mathbb{E}(\mathrm{X}_t \mid \mathcal{F}_s)$: X is a martingale. But this is also a finite variation process, so it is indistinguishable from 0. In other words, for any $t \geqslant 0$,

$$\int_0^t \mathrm{K}_u \mathrm{d}\langle \mathrm{M}\rangle_u = 0,$$

so $\mathrm{K} = 0$, ν-almost everywhere on $\mathbb{R}_+ \times \Omega$. □

Theorem 4.6. *Let* $\mathrm{M} \in \mathrm{H}^2$. *To* $\mathrm{H} \in \mathcal{E}$, *written* $\mathrm{H}_s(\omega) = \sum_{k=0}^{p-1} \mathrm{H}_k(\omega)\mathbb{1}_{]t_k,t_{k+1}]}(s)$, *we associate* $\mathrm{H} \cdot \mathrm{M} \in \mathrm{H}^2$ *defined by*

$$(\mathrm{H} \cdot \mathrm{M})_t = \sum_{k=1}^{p} \mathrm{H}_k \left(\mathrm{M}_{t_{k+1}\wedge t} - \mathrm{M}_{t_k \wedge t}\right).$$

Then the following results hold.

(i) *The map* $\mathrm{H} \mapsto \mathrm{H} \cdot \mathrm{M}$ *can be uniquely extended into an isometry from* $\mathrm{L}^2(\mathrm{M})$ *to* H^2.

(ii) *The process* $\mathrm{H} \cdot \mathrm{M}$ *obtained by the previous extension is characterized by*[1] [2] *: for any* $\mathrm{N} \in \mathrm{H}^2$,

$$\langle \mathrm{H} \cdot \mathrm{M}, \mathrm{N}\rangle = \mathrm{H} \cdot \langle \mathrm{M}, \mathrm{N}\rangle.$$

(iii) *If* T *is a stopping time, then*

$$(\mathbb{1}_{]0,\mathrm{T}]}\mathrm{H}) \cdot \mathrm{M} = (\mathrm{H} \cdot \mathrm{M})^{\mathrm{T}} = \mathrm{H} \cdot \mathrm{M}^{\mathrm{T}}.$$

(iv) *The so-called stochastic integral* $\mathrm{H} \cdot \mathrm{M}$ *of H with respect to M satisfies the following associativity relation : if* $\mathrm{G} \in \mathrm{L}^2(\mathrm{M})$ *anf* $\mathrm{H} \in \mathrm{L}^2(\mathrm{G} \cdot \mathrm{M})$, *then* $\mathrm{GH} \in \mathrm{L}^2(\mathrm{M})$ *and*

$$(\mathrm{GH}) \cdot \mathrm{M} = \mathrm{G} \cdot (\mathrm{H} \cdot \mathrm{M}).$$

Proof. To show (i), we first need to check that

$$\begin{array}{ccc} \mathcal{E} & \to & \mathrm{H}^2 \\ \mathrm{H} & \mapsto & \mathrm{H} \cdot \mathrm{M} \end{array}$$

is an isometry from $\mathcal{E}$ to H^2. First, one easily gets that $\mathrm{H} \cdot \mathrm{M}$ is a L^2 bounded martingale, and as $\langle \mathrm{H} \cdot \mathrm{M}\rangle_t = \sum_{k=1}^{p-1} \mathrm{H}_k^2(\langle \mathrm{M}\rangle_{t_{k+1}\wedge t} - \langle \mathrm{M}\rangle_{t_k \wedge t})$,

$$\|\mathrm{H} \cdot \mathrm{M}\|_{\mathrm{H}^2}^2 = \mathbb{E}\left(\sum_{k=1}^{p-1} \mathrm{H}_k^2(\langle \mathrm{M}\rangle_{t_{k+1}} - \langle \mathrm{M}\rangle_{t_k})\right) = \mathbb{E}\left(\int_0^\infty \mathrm{H}_s^2 \mathrm{d}\langle \mathrm{M}\rangle_s\right) = \|\mathrm{H}\|_{\mathrm{L}^2(\mathrm{M})}^2.$$

From the two preceding propositions, $\mathcal{E}$ is dense in the Hilbert space $\mathrm{L}^2(\mathrm{M})$, so this isometry can be extended in a unique way as an isometry between $(\mathrm{L}^2(\mathrm{M}), \|\|_{\mathrm{L}^2(\mathrm{M})})$ and $(\mathrm{H}^2, \|\|_{\mathrm{H}^2})$.

Concerning (ii), a calculation allows to prove the expected relation when $\mathrm{H} \in \mathcal{E}$:

1. In the following equation, the notation $\mathrm{H} \cdot \langle \mathrm{M}, \mathrm{N}\rangle$ refers to the usual Stieljes-type integral with respect to a finite variation process, in the sense of Proposition 3.22

2. This characterization also can give a definition of the stochastic integral : as

$$\begin{array}{ccc} \mathrm{H}^2 & \to & \mathbb{R} \\ \mathrm{N} & \mapsto & \mathbb{E}((\mathrm{H} \cdot \langle \mathrm{M}, \mathrm{N}\rangle)_\infty) \end{array}$$

is continuous (by Kunita-Watanabe) and linear, by the Riesz representation theorem there is a unique element $\mathrm{H} \cdot \mathrm{M} \in \mathrm{H}^2$ such that for any $\mathrm{N} \in \mathrm{H}^2$

$$\mathbb{E}((\mathrm{H} \cdot \langle \mathrm{M}, \mathrm{N}\rangle)_\infty = \mathbb{E}(\langle \mathrm{H} \cdot \mathrm{M}, \mathrm{N}\rangle_\infty)$$

$$\begin{aligned}
\langle \mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle_t &= \langle \sum_{k=1}^{p} \mathrm{H}_k\left(\mathrm{M}_{t_{k+1}\wedge\cdot}-\mathrm{M}_{t_k\wedge\cdot}\right),\mathrm{N}\rangle_t\\
&= \sum_{k=1}^{p} \mathrm{H}_k\langle \mathrm{M}_{t_{k+1}\wedge\cdot}-\mathrm{M}_{t_k\wedge\cdot},\mathrm{N}\rangle_t\\
&= \sum_{k=1}^{p} \mathrm{H}_k\left(\langle \mathrm{M},\mathrm{N}\rangle_{t_{k+1}\wedge t}-\langle \mathrm{M},\mathrm{N}\rangle_{t_k\wedge t}\right)\\
&= \int_0^t \mathrm{H}_s \mathrm{d}\langle \mathrm{M},\mathrm{N}\rangle_s\\
\langle \mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle &= \mathrm{H}\cdot\langle \mathrm{M},\mathrm{N}\rangle.
\end{aligned}$$

To prove this formula for general $\mathrm{H}\in\mathrm{L}^2(\mathrm{M})$, consider a sequence $\mathrm{H}^{(n)}$ in $\mathcal{E}$ converging to H in $\mathrm{L}^2(\mathrm{M})$. Then, by the isometry property, $\mathrm{H}^{(n)}\cdot\mathrm{M}$ converges to $\mathrm{H}\cdot\mathrm{M}$ in H^2. We will justify the steps of the following equalities afterwards :

$$\begin{aligned}
\langle \mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle_\infty &= \lim_{n\to\infty}\langle \mathrm{H}^{(n)}\cdot\mathrm{M},\mathrm{N}\rangle_\infty\\
&= \lim_{n\to\infty}(\mathrm{H}^{(n)}\cdot\langle \mathrm{M},\mathrm{N}\rangle)_\infty\\
&= (\mathrm{H}\cdot\langle \mathrm{M},\mathrm{N}\rangle)_\infty
\end{aligned}$$

The first equality is in the sense of a L^1-limit, and a consequence of the Kunita-Watanabe inequality, where we take $\mathrm{X}=\mathrm{H}^{(n)}\cdot\mathrm{M}-\mathrm{H}\cdot\mathrm{M}$:

$$\mathbb{E}(|\langle \mathrm{X},\mathrm{N}\rangle_\infty|)\leqslant\mathbb{E}(\langle \mathrm{X}\rangle_\infty)^{1/2}\,\mathbb{E}(\langle \mathrm{N}\rangle_\infty)^{1/2}=\|\mathrm{X}\|_{\mathrm{H}^2}\,\mathbb{E}(\langle \mathrm{N}\rangle_\infty)^{1/2}.$$

The second equality has just be proven and relies on $\mathrm{H}^{(n)}\in\mathcal{E}$. Finally, the third equality holds as a limit in L^1 and relies on Kunita-Watanabe as well :

$$\mathbb{E}\left|\int_0^\infty(\mathrm{H}^{(n)}_s-\mathrm{H}_s)\mathrm{d}\langle \mathrm{M},\mathrm{N}\rangle_s\right|\leqslant\mathbb{E}\left(\langle \mathrm{N}\rangle_\infty\right)\|\mathrm{H}^{(n)}-\mathrm{H}\|_{\mathrm{L}^2(\mathrm{M})}.$$

We have proven that $\mathbb{E}(|\langle \mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle_\infty-(\mathrm{H}\cdot\langle \mathrm{M},n\rangle)_\infty|)=0$ which yields that almost surely

$$\langle \mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle_\infty=(\mathrm{H}\cdot\langle \mathrm{M},\mathrm{N}\rangle)_\infty.$$

Choosing N stopped at time t in this identity proves the expected result.

To get the unique characterization of $\mathrm{H}\cdot\mathrm{M}$ by the identity in (*ii*), not that if X satisfies the same property as $\mathrm{H}\cdot\mathrm{M}$, then for any $\mathrm{N}\in\mathrm{H}^2$

$$\langle \mathrm{H}\cdot\mathrm{M}-\mathrm{X},\mathrm{N}\rangle_\infty=0.$$

The choice $\mathrm{N}=\mathrm{H}\cdot\mathrm{M}-\mathrm{X}$ proves that $\mathrm{H}\cdot\mathrm{M}$ and X are indistinguishable.

For (*iii*), we can for example first prove $(\mathbb{1}_{]0,\mathrm{T}]}\mathrm{H})\cdot\mathrm{M}=(\mathrm{H}\cdot\mathrm{M})^{\mathrm{T}}$. For this, we just need to check that for any $\mathrm{N}\in\mathrm{H}^2$

$$\langle(\mathbb{1}_{]0,\mathrm{T}]}\mathrm{H})\cdot\mathrm{M},\mathrm{N}\rangle_\infty=\langle(\mathrm{H}\cdot\mathrm{M})^{\mathrm{T}},\mathrm{N}\rangle_\infty.$$

The left hand side is also $((\mathbb{1}_{]0,\mathrm{T}]}\mathrm{H})\cdot\langle \mathrm{M},\mathrm{N}\rangle)_\infty=(\mathrm{H}\cdot\langle \mathrm{M},\mathrm{N}\rangle)_{\mathrm{T}}$, and the right hand side is $\langle \mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle_{\mathrm{T}}$. As $\langle \mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle=\mathrm{H}\cdot\langle \mathrm{M},\mathrm{N}\rangle$, this achieves the proof. Concerning the proof of $(\mathrm{H}\cdot\mathrm{M})^{\mathrm{T}}=\mathrm{H}\cdot\mathrm{M}^{\mathrm{T}}$, it proceeds the same way.

Finally, the first assertion of (*iv*) follows from definitions : as $\langle \mathrm{H}\cdot\mathrm{M}\rangle=\mathrm{H}\cdot\langle \mathrm{M},\mathrm{H}\cdot\mathrm{M}\rangle=\mathrm{H}^2\langle \mathrm{M}\rangle$,

$$\int_0^\infty(\mathrm{GH})^2\mathrm{d}\langle \mathrm{M}\rangle_s=\int_0^\infty\mathrm{H}^2\mathrm{d}\langle \mathrm{G}\cdot\mathrm{M}\rangle_s<\infty$$

because $\mathrm{H} \in \mathrm{L}^2(\mathrm{G}\cdot\mathrm{M})$, so $\mathrm{GH} \in \mathrm{L}^2(\mathrm{M})$. To prove the second assertion, we just need to test it with respect to any $\mathrm{N} \in \mathrm{H}^2$, using the associativity for integrals of Stieljes type :

$$\langle(\mathrm{GH})\cdot\mathrm{M},\mathrm{N}\rangle = (\mathrm{GH})\cdot\langle\mathrm{M},\mathrm{N}\rangle = \mathrm{G}\cdot(\mathrm{H}\cdot\langle\mathrm{M},\mathrm{N}\rangle) = \mathrm{G}\cdot\langle\mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle = \langle\mathrm{G}\cdot(\mathrm{H}\cdot\mathrm{M}),\mathrm{N}\rangle,$$

so $(\mathrm{GH})\cdot\mathrm{M} = \mathrm{G}\cdot(\mathrm{H}\cdot\mathrm{M})$. □

In the following, we will often use the notation

$$(\mathrm{H}\cdot\mathrm{M})_t = \int_0^t \mathrm{H}_s \mathrm{dM}_s.$$

In this way, property (*iv*) of the previous theorem appears as a formal evidence :

$$\int_0^t \mathrm{G}_s(\mathrm{H}_s\mathrm{dM}_s) = \int_0^t (\mathrm{G}_s\mathrm{H}_s)\mathrm{dM}_s.$$

Moreover, property (*ii*) when iterated yields the joint bracket of two stochastic integrals : as $\langle\int_0^\cdot \mathrm{H}_s\mathrm{dM}_s,\mathrm{N}\rangle_t = \int_0^t \mathrm{H}_s\mathrm{d}\langle\mathrm{M},\mathrm{N}\rangle_s$, we get for $\mathrm{H}\in\mathrm{L}^2(\mathrm{M})$ and $\mathrm{G}\in\mathrm{L}^2(\mathrm{N})$

$$\langle\int_0^\cdot \mathrm{H}_s\mathrm{dM}_s, \int_0^\cdot \mathrm{G}_s\mathrm{dM}_s\rangle_t = \int_0^t (\mathrm{G}_s\mathrm{H}_s)\mathrm{d}\langle\mathrm{M},\mathrm{N}\rangle_s. \tag{4.4}$$

In particular, the first two moment of stochastic integrals are well-known :

$$\begin{aligned}\mathbb{E}\left(\int_0^t \mathrm{H}_s\mathrm{dM}_s\right) &= 0\\ \mathbb{E}\left(\left(\int_0^t \mathrm{H}_s\mathrm{dM}_s\right)\left(\int_0^t \mathrm{G}_s\mathrm{dN}_s\right)\right) &= \mathbb{E}\left(\int_0^t (\mathrm{G}_s\mathrm{H}_s)\mathrm{d}\langle\mathrm{G},\mathrm{H}\rangle_s\right).\end{aligned}$$

We want to emphasize that the above relations are not always true when considering the following extension of stochastic integrals to local martingales. For this, we will consider integration with respect to a local integral martingale of processes H locally integrable in the following sense :

$$\mathrm{L}^2_{\mathrm{loc}}(\mathrm{M}) := \left\{\text{progressively measurable}\,\mathrm{H}\mid \text{for any } t\geqslant 0,\ \int_0^t \mathrm{H}_s^2\mathrm{d}\langle\mathrm{M}\rangle_s<\infty \text{ a.s.}\right\}.$$

Theorem 4.7. *Let* M *be a local martingale beginning at 0. For any* $\mathrm{H}\in\mathrm{L}^2_{\mathrm{loc}}(\mathrm{M})$*, there is a unique local martingale beginning at 0, noted* $\mathrm{H}\cdot\mathrm{M}$*, such that for any local martingale* N

$$\langle\mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle = \mathrm{H}\cdot\langle\mathrm{M},\mathrm{N}\rangle. \tag{4.5}$$

This definition extends the one from Theorem 4.6. Moreover, for any stopping time T*,*

$$(\mathbb{1}_{]0,\mathrm{T}]}\mathrm{H})\cdot\mathrm{M} = (\mathrm{H}\cdot\mathrm{M})^{\mathrm{T}} = \mathrm{H}\cdot\mathrm{M}^{\mathrm{T}}. \tag{4.6}$$

Proof. Note

$$\mathrm{T}_n = \inf\left\{t\geqslant 0 \mid \int_0^t (1+\mathrm{H}_s^2)\mathrm{d}\langle\mathrm{M}\rangle_s = n\right\}.$$

Then $\mathrm{M}^{\mathrm{T}_n}\in\mathrm{H}^2$ and $\mathrm{H}\in\mathrm{L}^2(\mathrm{M}^{\mathrm{T}_n})$, so we can consider the stochastic integral $\mathrm{H}\cdot\mathrm{M}^{\mathrm{T}_n}$ from Theorem 4.6. From (*iii*) of this Theorem, for $m<n$, $(\mathrm{H}\cdot\mathrm{M}^{\mathrm{T}_m})^{\mathrm{T}_n} = \mathrm{H}\cdot\mathrm{M}^{\mathrm{T}_n}$. This coherence relation proves that there is a process, noted $\mathrm{H}\cdot\mathrm{M}$ such that, for any $n\geqslant 0$, $(\mathrm{H}\cdot\mathrm{M})^{\mathrm{T}_n} = \mathrm{H}\cdot\mathrm{M}^{\mathrm{T}_n}$. Moreover, as $\mathrm{H}\in\mathrm{L}^2_{\mathrm{loc}}(\mathrm{M})$, $\mathrm{T}_n\to\infty$ almost surely with n, hence as $(\mathrm{H}\cdot\mathrm{M})^{\mathrm{T}_n}$ is a martingale, $\mathrm{H}\cdot\mathrm{M}$ is a local martingale.

To prove relation (4.5), take N a local martingale and $\tilde{\mathrm{T}}_n = \inf\{t \geqslant 0 \mid |\mathrm{N}_t| = n\}$. Then if $\mathrm{S}_n = \tilde{\mathrm{T}}_n \wedge \mathrm{T}_n$, $\mathrm{N}^{\mathrm{S}_n} \in \mathrm{H}^2$, so we can write

$$\begin{aligned}\langle \mathrm{H}\cdot\mathrm{M},\mathrm{N}\rangle^{\mathrm{S}_n} = \langle (\mathrm{H}\cdot\mathrm{M})^{\mathrm{S}_n},\mathrm{N}^{\mathrm{S}_n}\rangle = \langle (\mathrm{H}\cdot\mathrm{M}^{\mathrm{T}_n})^{\mathrm{S}_n},\mathrm{N}^{\mathrm{S}_n}\rangle = \langle \mathrm{H}\cdot\mathrm{M}^{\mathrm{T}_n}\mathrm{N}^{\mathrm{S}_n}\rangle \\ = \mathrm{H}\cdot\langle \mathrm{M}^{\mathrm{T}_n},\mathrm{N}^{\mathrm{S}_n}\rangle = \mathrm{H}\cdot\langle \mathrm{M},\mathrm{N}\rangle^{\mathrm{S}_n} = (\mathrm{H}\cdot\langle \mathrm{M},\mathrm{N}\rangle)^{\mathrm{S}_n}.\end{aligned}$$

As $\mathrm{S}_n \to \infty$ a.s. we get the expected result (4.5). Concerning the proof of (4.6), it proceeds exactly as for (*iii*) in Theorem 4.6, relying only on (4.5). Finally, our definition coincides with the H^2 and $\mathrm{L}^2(\mathrm{M})$ case by just noting that (4.5) characterizes $\mathrm{H}\cdot\mathrm{M}$. □

To extend the notion of stochastic integral to semimartingales, we need to work with locally bounded processes : a progressively measurable process H is said to be locally bounded if for any $t \geqslant 0$, almost surely

$$\sup_{[0,t]} |\mathrm{H}_s| < \infty.$$

Note that if H is locally bounded and M is a local martingale, then $\mathrm{H} \in \mathrm{L}^2_{\mathrm{loc}}(\mathrm{M})$. Moreover, if A is a finite variation process, then for any $t \geqslant 0$ almost surely $\int_0^t |\mathrm{H}_s||\mathrm{dA}_s| < \infty$. As a consequence, for a semimartingale $\mathrm{X} = \mathrm{X}_0 + \mathrm{M} + \mathrm{A}$, and a locally bounded progressively measurable process H, the definition

$$\mathrm{H}\cdot\mathrm{X} = \mathrm{H}\cdot\mathrm{M} + \mathrm{H}\cdot\mathrm{A}$$

makes sense. We will mostly note it $\int_0^\cdot \mathrm{H}_s\mathrm{dX}_s$.

Proposition 4.8. *The stochastic integral of a locally bounded progressively measurable process with respect to a semimartingales satisfies the following properties.*

(*i*) *The application* $(\mathrm{H},\mathrm{X}) \mapsto \mathrm{H}\cdot\mathrm{X}$ *is bilinear.*

(*ii*) *If* G *and* H *are locally bounded,* $\mathrm{G}\cdot(\mathrm{H}\cdot\mathrm{X}) = (\mathrm{GH})\cdot\mathrm{X}$

(*iii*) *If* T *is a stopping time,* $(\mathbb{1}_{]0,\mathrm{T}]}\mathrm{H})\cdot\mathrm{X} = (\mathrm{H}\cdot\mathrm{X})^{\mathrm{T}} = \mathrm{H}\cdot\mathrm{X}^{\mathrm{T}}$.

(*iv*) *If* X *is a local martingale, so is* $\mathrm{H}\cdot\mathrm{X}$.

(*v*) *If* X *is a finite variation process, so is* $\mathrm{H}\cdot\mathrm{X}$.

(*vi*) *If* H *is a step process* ($\mathrm{H}_s = \sum_{k=0}^{p-1}\mathrm{H}_k\mathbb{1}_{]t_k,t_{k+1}]}(s)$)*, then*

$$(\mathrm{H}\cdot\mathrm{X})_t = \sum_{k=0}^{p-1}\mathrm{H}_k(\mathrm{X}_{t_{k+1}\wedge t} - \mathrm{X}_{t_k\wedge t}).$$

(*vii*) *If* H *is also assumed to be left-continuous, then in the sense of convergence in probability*

$$\int_0^t \mathrm{H}_s\mathrm{dX}_s = \lim_{n\to\infty}\sum_{k=0}^{p_n-1}\mathrm{H}_{t_k^{(n)}}\left(\mathrm{X}_{t_{k+1}^{(n)}} - \mathrm{X}_{t_k^{(n)}}\right),$$

where the sequence of subdivisions $0 = t_0^{(n)} < \cdots < t_{p_n}^{(n)} = t$ *has a step going to 0.*

Proof. All the above results are easy consequences of previous analogue statements concerning local martingales and finite variation processes, except (*vii*). To prove it, we can suppose that X is a local martingale, a similar statement being proved for finite variation processes in Proposition 3.21. Let $\mathrm{H}^{(n)}$ be the process equal to $\mathrm{H}_{t_k^{(n)}}$ on $]t_k^{(n)}, t_{k+1}^{(n)}]$, 0 otherwise, and

$$\mathrm{T}_m = \inf\{t \geqslant 0 \mid |\mathrm{H}_s| + \langle \mathrm{M}\rangle_s \geqslant m\}.$$

Then $\mathrm{M}^{\mathrm{T}_m}$ and $\mathrm{H}^{(n)}$ are bounded on $[0, \mathrm{T}_m]$, so the definition and properties of the stochastic integral in the $\mathrm{H}^2/\mathrm{L}^2(\mathrm{M}^{\mathrm{T}_m})$ case allow to use the isometry property :

$$\mathbb{E}\left(\left((\mathrm{H}^n \mathbb{1}_{[0,\mathrm{T}_m]} \cdot \mathrm{M}^{\mathrm{T}_m})_t - (\mathrm{H}\mathbb{1}_{[0,\mathrm{T}_m]} \cdot \mathrm{M})_t\right)^2\right) = \mathbb{E}\left(\int_0^{t\wedge \mathrm{T}_m} (\mathrm{H}_s^n - \mathrm{H}_s)^2 \mathrm{d}\langle \mathrm{M}\rangle_s\right).$$

As H is left-continuous, this last term converges to 0 as $n \to \infty$ by dominated convergence. Hence $(\mathrm{H}^n \cdot \mathrm{M})_t^{\mathrm{T}_m}$ converges in L^2 to $(\mathrm{H}\cdot\mathrm{M})_t^{\mathrm{T}_m}$. As $\mathbb{P}(\mathrm{T}_m > t) \to 1$ for $m \to \infty$, we can conclude that $(\mathrm{H}^n \cdot \mathrm{M})_t$ converges in probability to $(\mathrm{H}\cdot\mathrm{M})_t$. □

2. The Itô formula

One of the most important formulas in analysis is the change of variable formula. For example, in dimension 1, it states that if A is a continuous finite variation function and $f \in \mathscr{C}^1$, then

$$f(\mathrm{A}_t) = f(\mathrm{A}_0) + \int_0^t f'(\mathrm{A}_s)\mathrm{d}\mathrm{A}_s.$$

When A is a semimartingale, we want a similar formula. This is the cornerstone of the following of this lecture.

Theorem 4.9. *Let* $\mathrm{F} : \mathbb{R}^d \to \mathbb{R}$ *be of class* $\mathscr{C}^2$, *and* $\mathrm{X}^1, \dots, \mathrm{X}^d$ *(continuous) semimartingales. Then* $\mathrm{F}(\mathrm{X}^1, \dots, \mathrm{X}^d)$ *is a semimartingale and*

$$\mathrm{F}(\mathrm{X}_t^1, \dots, \mathrm{X}_t^d) = \mathrm{F}(\mathrm{X}_0^1, \dots, \mathrm{X}_0^d) + \sum_{k=1}^d \int_0^t \partial_{x_k}\mathrm{F}(\mathrm{X}_s^1, \dots, \mathrm{X}_s^d)\mathrm{d}\mathrm{X}_s^k + \frac{1}{2}\sum_{1\leqslant k,\ell\leqslant d} \int_0^t \partial_{x_k x_\ell}\mathrm{F}(\mathrm{X}_s^1, \dots, \mathrm{X}_s^d)\mathrm{d}\langle \mathrm{X}^k, \mathrm{X}^\ell\rangle_s.$$

Note that the the Itô formula gives the decomposition of $\mathrm{F}(\mathrm{X}^1, \dots, \mathrm{X}^d)$ as a sum of a local martingale and a finite variation process.

Proof. We first prove this formula when $f(x, y) = xy$, which is equivalent to proving a stochastic integration by parts formula. For this purpose, by polarization, we just need to prove it for $f(x) = x^2$:

$$\mathrm{X}_t^2 = \mathrm{X}_0^2 + 2\int_0^t \mathrm{X}_t \mathrm{d}\mathrm{X}_t + \langle \mathrm{X}\rangle_t.$$

We know by (vii) in Proposition 4.8 that

$$2\int_0^t \mathrm{X}_t\mathrm{d}\mathrm{X}_t = \lim_{n\to\infty} 2\sum_{k=0}^{p_n-1} \mathrm{X}_{t_k^{(n)}}\left(\mathrm{X}_{t_{k+1}^{(n)}} - \mathrm{X}_{t_k^{(n)}}\right),$$

where the sequence of subdivisions $0 = t_0^{(n)} < \dots < t_{p_n}^{(n)} = t$ has a step going to 0, and the limit is in probability. Writing this as

$$\lim_{n\to\infty}\left(\sum_{k=0}^{p_n-1}\left(\mathrm{X}_{t_{k+1}^{(n)}}^2 - \mathrm{X}_{t_k^{(n)}}^2\right) - \sum_{k=0}^{p_n-1}\left(\mathrm{X}_{t_{k+1}^{(n)}} - \mathrm{X}_{t_k^{(n)}}\right)^2\right)$$

and using (3.11) proves the Itô formula in the quadratic case, by uniqueness of the limit in probability. We proved that the formula is true for F being constant, linear of of type $x_i x_j$. As a consequence, if the result holds for $\mathrm{F}(x)$, a simple calculation proves that it is true for $\mathrm{F}(x)x_i$. Hence, iterating this reasoning, the result holds for any polynomial in the variables $x_1, \dots, x_d$ of arbitrary degree.

Moreover, by localizing our processes $\mathrm{X}_1, \dots, \mathrm{X}_d$ through their first exist time from the ball with radius n, and taking $n \to \infty$ at the end, we can assume that the processes remain in a compact K. As F is $\mathscr{C}^2$ on K, the Stone-Weierstrass theorem gives a sequence of polynomials $(\mathrm{P}_n, \geqslant 0)$ in the variables $x_1, \dots, x_n$ such that P_n, its first and second order derivatives converge uniformly to those of F on K. One can easily show by dominated convergence that if H and the H^n's are continuous uniformly bounded progressively measurable processes, H_n converging almost surely to H on $[0, t]$, then for any semimartingale X

$$\int_0^t \mathrm{H}_s^n \mathrm{dX}_s \underset{n\to\infty}{\longrightarrow} \int_0^t \mathrm{H}_s \mathrm{dX}_s$$

in the sense of convergence in probability : the result is true in the almost sure sense if X is a finite variation process by the usual dominated convergence ; if X is L^2-bounded martingale the convergence holds in L^2 by the isometry property ; if X is a local martingale, the convergence holds in probability by localization. Thus, writing the Itô formula for the polynomials P_n's and taking the limit $n \to \infty$ yields the result for F, by uniqueness of the limit in probability.

Finally, as the formula is proved, it is obvious that $\mathrm{F}(\mathrm{X}_1, \dots, \mathrm{X}_d)$ is a semimartingale, Itô's formula giving its explicit decomposition (as F is $\mathscr{C}^2$, all integrated terms are locally bounded, so the stochastic and Stieljes integrals make sense). □

In the following, the Itô formula will often be mentioned in its differential form :

$$\mathrm{dF}(\mathrm{X}_t^1, \dots, \mathrm{X}_t^d) = \sum_{k=1}^d \partial_k \mathrm{F}(\mathrm{X}_s^1, \dots, \mathrm{X}_s^d)\mathrm{dX}_s^k + \frac{1}{2} \sum_{1 \leqslant k, \ell \leqslant d} \partial_{k\ell} \mathrm{F}(\mathrm{X}_s^1, \dots, \mathrm{X}_s^d)\mathrm{d}\langle \mathrm{X}^k, \mathrm{X}^\ell \rangle_s.$$

To see how the Itô formula works in practice, the following is an important familly local martingales. Note that a complex process is called a local martingale if both its real and imaginary parts are local martingales.

Corollary 4.10. *Let* M *be a (real) local martingale and* $\lambda \in \mathbb{C}$. *Then the process*

$$\left(e^{\lambda \mathrm{M}_t - \frac{\lambda^2}{2} \langle \mathrm{M} \rangle_t}, t \geqslant 0 \right)$$

is a local martingale.

Proof. Let F be a $\mathscr{C}^2$ function from $\mathbb{R}^2$ to $\mathbb{R}$. Then, by the Itô formula, for a semimartingale M,

$$\mathrm{dF}(\mathrm{M}_t, \langle \mathrm{M} \rangle_t) = (\partial_x \mathrm{F})\mathrm{dM}_t + (\partial_y \mathrm{F})\mathrm{d}\langle \mathrm{M} \rangle_t + \frac{1}{2}(\partial_{xx}\mathrm{F})\mathrm{d}\langle \mathrm{M} \rangle_t.$$

As a consequence, if M is a local martingale and

$$\left(\partial_y + \frac{1}{2} \partial_{xx} \right) \mathrm{F} = 0, \tag{4.7}$$

then $\mathrm{F}(\mathrm{M}, \langle \mathrm{M} \rangle)$ is a stochastic integral with respect to M, hence a local martingale. In our case, both the real and imaginary parts of $f(x, y) = e^{\lambda x - \frac{\lambda^2}{2} y}$ satisfy (4.7), which yields the result. □

A fundamental application of the Itô formula is the following characterization of the Brownian motion, that we will refer to in the following of the course as *Lévy's criterion.*

Theorem 4.11. *Let* $\mathrm{M}^{(1)}, \dots, \mathrm{M}^{(d)}$ *be local martingales beginning at 0. Then the following assertions are equivalent.*

(i) The processes $\mathrm{M}^{(1)}, \dots, \mathrm{M}^{(d)}$ *are independent standard Brownian motions.*

(ii) For any $1 \leqslant k, \ell \leqslant d$ *and* $t \geqslant 0$, $\langle \mathrm{M}^{(k)}, \mathrm{M}^{(\ell)} \rangle_t = \mathbb{1}_{k=\ell} t$.

Proof. Point (i) implies (ii) because $({\mathrm{M}_t^{(k)}}^2 - t, t \geqslant 0)$ is a martingale, as well as $(\mathrm{M}_t^{(k)} \mathrm{M}_t^{(\ell)}, t \geqslant 0)$ when $k \neq \ell$.

Assume (ii), note $\mathrm{M} = (\mathrm{M}^{(1)}, \dots, \mathrm{M}^{(d)})$ and consider some $u \in \mathbb{R}^d$. Then $u \cdot \mathrm{M}$ is a local martingale with bracket $\langle u \cdot \mathrm{M} \rangle_t = |u|^2 t$. As a consequence, from Corollary 4.10,

$$\left(e^{\mathrm{i}u\cdot\mathrm{M}-\frac{1}{2}|u^2|t}, t \geqslant 0\right)$$

is a local martingale. As it is bounded, it is a martingale :

$$\mathbb{E}\left(e^{\mathrm{i}u\cdot\mathrm{M}_t-\frac{1}{2}|u^2|t} \mid \mathcal{F}_s\right) = e^{\mathrm{i}u\cdot\mathrm{M}_s-\frac{1}{2}|u^2|s}.$$

As a consequence, for any $\mathrm{A} \in \mathcal{F}_s$,

$$\mathbb{E}\left(e^{\mathrm{i}u\cdot(\mathrm{M}_t-\mathrm{M}_s)}\mathbb{1}_\mathrm{A}\right) = e^{-\frac{1}{2}|u|^2(t-s)}\mathbb{P}(\mathrm{A}).$$

The choice $\mathrm{A} = \Omega$ proves that $\mathrm{M}_t - \mathrm{M}_s$ is a Gaussian vector with covariance matrix $(t-s)\mathrm{Id}_d$, hence with independent coordinates. Moreover, the above equation also proves that $\mathrm{M}_t - \mathrm{M}_s$ is independent of $\mathcal{F}_s$. This concludes the proof. □

As an example, consider a Brownian motion B, and define the process $\tilde{\mathrm{B}}$ through

$$\tilde{\mathrm{B}}_t = \int_0^t \mathrm{sgn}(\mathrm{B}_s)\mathrm{dB}_s,$$

where $\mathrm{sgn}(x) = 1$ if $x \geqslant 0$, -1 if $x < 0$. Then $\tilde{\mathrm{B}}$ is a local martingale and from (4.4), its bracket is $\langle \tilde{\mathrm{B}} \rangle_t = t$. As a consequence, $\tilde{\mathrm{B}}$ is a Brownian motion.

Another byproduct of Lévy's criterion is that, in the multidimensional setting, the Brownian motion has a rotation invariant distribution : it does not depend on the chosen orthogonal framework (in the next section we will see that for $d = 2$ the Brownian motion satisfies the more general property of conformal invariance). More precisely, consider a d-dimensional Brownian motion $\mathrm{B} = (\mathrm{B}^1, \dots, \mathrm{B}^d)$ beginning at 0 and an orthogonal matrix $\mathrm{O} \in \mathrm{O}(d)$. Then

$$\tilde{\mathrm{B}} = \mathrm{OB}$$

is also a Brownian motion. Indeed,

$$\langle \tilde{\mathrm{B}}^k, \tilde{\mathrm{B}}^\ell \rangle_t = \langle \sum_{i=1}^d \mathrm{O}_{ki}\mathrm{B}^i, \sum_{j=1}^d \mathrm{O}_{\ell j}\mathrm{B}^j \rangle_t = \sum_{i=1}^d \mathrm{O}_{ki}\mathrm{O}_{\ell i}\langle \mathrm{B}^i \rangle_t = (\mathrm{O}_k \cdot \mathrm{O}_l)t = \mathbb{1}_{k=\ell}t,$$

which concludes the proof by Theorem 4.11.

An important consequence of the Lévy criterion is that up to a change of time, any martingale is a Brownian motion : the natural clock of martingales is the bracket. This is the Dubins-Schwarz theorem.

Theorem 4.12. *Let* M *be a local martingale such that* $\langle \mathrm{M}\rangle_\infty = \infty$ *almost surely. Note*

$$\mathrm{T}_t = \inf\{s \geqslant 0 \mid \langle \mathrm{M}\rangle_s > t\}.$$

Then for any given t *the random variable* T_t *is an almost surely finite stopping time, and the process*[3] $\mathrm{B}_t = \mathrm{M}_{\mathrm{T}_t}$ *is a Brownian motion with respect to the filtration* $(\mathcal{G}_t)_{t\geqslant 0} = (\mathcal{F}_{\mathrm{T}_t})_{t\geqslant 0}$. *Moreover*[4],

$$\mathrm{M}_t = \mathrm{B}_{\langle \mathrm{M}\rangle_t}.$$

Proof. First, as all expected results are in the almost sure sense, we can suppose that for any ω the process M begins at 0, is continuous, and $\langle \mathrm{M}\rangle_\infty = \infty$. The process $(\mathrm{T}_t)_{t\geqslant 0}$ is nondecreasing, right-continuous, and finite. Moreover,

$$\langle \mathrm{M}\rangle_t = \inf\{s \geqslant 0 \mid \mathrm{T}_s > t\}.$$

Note that the processes M and $\langle \mathrm{M}\rangle$ have the same constance intervals : if M is constant on $[\mathrm{S},\mathrm{T}]$ so is $\langle \mathrm{M}\rangle$ thanks to the characterization of the bracket as a limiting sum of squares of increments ; conversely, if $\langle \mathrm{M}\rangle$ is constant on $[\mathrm{S},\mathrm{T}]$, as a local martingale with null bracket is indistinguishable from 0, then the result follows by considering a proper shift of M after a stopping time.

To prove that B is continuous, note that this is obvious at points t where t is continuous, and if T is not continuous at t this follows from the above coincidence of constance intervals. By Lévy's criterion, we therefore just need to prove that B and $\mathrm{B}_t^2 - t$ are local martingales with respect to the filtration $(\mathcal{G}_t)_{t\geqslant 0}$.

Let X denote M or $(\mathrm{M}_t^2 - t, t \geqslant 0)$. Let $\mathrm{S}_n = \inf\{t \geqslant 0 \mid |\mathrm{X}_t| \geqslant n\}$. Then $\tilde{\mathrm{S}}_n = \langle \mathrm{M}\rangle_{\mathrm{S}_n}$ is a $(\mathcal{G}_t)_{t\geqslant 0}$-stopping time. As $\mathrm{X}^{\mathrm{S}_n}$ is a bounded $(\mathcal{F}_t)_{t\geqslant 0}$-martingale, the stopping time theorem yields

$$\mathrm{X}^{\mathrm{S}_n}_{\mathrm{T}_t} = \mathbb{E}\left(\mathrm{X}^{\mathrm{S}_n}_\infty \mid \mathcal{F}_{\mathrm{T}_t}\right).$$

This means that $\mathrm{X}_{\tau_t \wedge \tilde{\mathrm{S}}_n}$ is a $(\mathcal{G}_t)_{t\geqslant 0}$-martingale, so as $\langle \mathrm{M}\rangle_{\mathrm{S}_n} \to \infty$ we get the expected result. □

Note that one can state a similar result when $\mathbb{P}(\langle \mathrm{M}\rangle_\infty) > 0$, by constructing B on an enlarged probability space. See [16] for more details. A consequence of the Dubins-Schwarz theorem is that any local martingale M have many common properties with a Brownian motion.

- In the interior of non constant intervals, M is nowhere differentiable, has a Hölderian index $1/2$.
- If the bracket has a strictly positive right increasing rate at t, M satisfies an iterated logarithm law.
- Up to a null set, $\{\omega \mid \mathrm{M}\,\text{converges in}\ \mathbb{R}\} = \{\omega \mid \langle \mathrm{M}\rangle_\infty < \infty\}$.
- Up to a null set, $\{\omega \mid \limsup \mathrm{M} = \infty, \liminf \mathrm{M} = -\infty\} = \{\omega \mid \langle \mathrm{M}\rangle_\infty = \infty\}$

Finally, the following multidimensional extension of the Dubins-Schwarz theorem can be proved in a similar way (with the additional difficulty that independence of processes must be proven, see [16] for a proof). It will be useful in the next section.

Theorem 4.13. *Let* $\mathrm{M}^{(1)},\dots,\mathrm{M}^{(d)}$ *be continuous local martingales beginning at* 0 *such that, for any* $1 \leqslant k \leqslant d$, $\langle \mathrm{M}^{(k)}\rangle_t \to \infty$ *as* $t \to \infty$. *If, for any* $k \neq \ell$, $\langle \mathrm{M}^{(k)}, \mathrm{M}^{(\ell)}\rangle = 0$, *then there exist* $\mathrm{B}^{(1)},\dots,\mathrm{B}^{(d)}$ *independent standard Brownian motions such that, for any* $1 \leqslant k \leqslant d$, $\mathrm{M}_t^{(k)} = \mathrm{B}^{(k)}_{\langle \mathrm{M}\rangle_t}$.

3. defined up to a null set
4. still up to a null set

3. Transcience, recurrence, harmonicity

We are now interested in properties of the Brownian curve in connection with classical harmonic analysis problems. First, as we will see, transcience or recurrence properties of the Brownian motion in dimension d depend on the asymptotic behavior of harmonic functions Then, more generally, we will give a classical interpretation of solutions to the Dirichlet problem through harmonic measures defined from the Brownian motion. This point of view will be developped in the next chapter when existence and uniqueness of solutions to some stochastic differential equations will be proved.

To begin our discussion, we need first the following highly intuitive result. Note however that there exist continuous mappings from the unit interval to the unit square [15].

Theorem 4.14. *Let $d \geqslant 2$. Then for a Brownian motion B in dimension d points are polar : for any $x \neq \mathrm{B}_0$,*

$$\mathbb{P}\left(\exists t \geqslant 0 \mid \mathrm{B}_t = x\right) = 0.$$

Proof. First, by projection on a subspace of dimension 2, proving the result for $d = 2$ is sufficient. Moreover, by scaling and rotation-invariance of Brownian motion (cf the previous section), we can consider that $x = (-1, 0)$, and that the bi-dimensional Brownian motion (X, Y) begins at $(0, 0)$. Let $\mathrm{M}_t = e^{\mathrm{X}_t} \cos \mathrm{Y}_t$ and $\mathrm{N}_t = e^{\mathrm{X}_t} \sin \mathrm{Y}_t$. Then an application of Itô's formula yields

$$\begin{aligned} \mathrm{dM}_t &= \mathrm{M}_t \mathrm{dX}_t - \mathrm{N}_t \mathrm{dY}_t \\ \mathrm{dN}_t &= \mathrm{N}_t \mathrm{dX}_t + \mathrm{M}_t \mathrm{dY}_t, \end{aligned}$$

so M and N are local martingales. Moreover, $\langle \mathrm{M}, \mathrm{N} \rangle = 0$ and

$$\langle \mathrm{N} \rangle_t = \langle \mathrm{M} \rangle_t = \int_0^t e^{2\mathrm{X}_t} \mathrm{d}t.$$

The recurrence of the Brownian motion X easily implies that these brackets go to ∞ almost surely. As a consequence, one can apply Theorem 4.13 to conclude that there are two Brownian motions $\mathrm{B}^1, \mathrm{B}^2$ such that

$$\mathrm{M}_t - 1 = \mathrm{B}^1_{\langle \mathrm{M} \rangle_t},\ \mathrm{N}_t = \mathrm{B}^2_{\langle \mathrm{M} \rangle_t}.$$

As a consequence, $\langle \mathrm{M} \rangle$ being continuous and with almost sure range $[0, \infty[$, noting $\mathrm{B} = (\mathrm{B}^1, \mathrm{B}^2)$,

$$\mathbb{P}\left(\exists t \geqslant 0 \mid \mathrm{B}_t = (-1, 0)\right) = \mathbb{P}(\exists t \geqslant 0 \mid (\mathrm{M}_t, \mathrm{N}_t) = (0, 0)).$$

As $|(\mathrm{M}_t, \mathrm{N}_t)| = e^{\mathrm{X}_t}$ and almost surely X_t is finite for any $t \geqslant 0$, this last event has probability 0. □

Although the Brownian motion will not visit a given point, it will almost surely visit any of its neighborhoods when $d = 2$.

Theorem 4.15. *Let B be a Brownian motion in dimension $d = 2$, and $\mathrm{O} \subset \mathbb{R}^2$ be open. Then*

$$\mathbb{P}(\exists t \geqslant 0 \mid \mathrm{B}_t \in \mathrm{O}) = 1.$$

Proof. We can assume $\mathrm{B}_0 = a \neq 0$, and we want to prove that for any $r > 0$ the probability that for some t, B is in $\mathscr{B}(r)$ (the ball with radius r) is 1. We note here $\mathrm{X}_t = |\mathrm{B}_t|^2$.

First, note that the process $\log \mathrm{X}_t$ is a local martingale. Indeed, for a given function f of class $\mathscr{C}^2$ on $\mathbb{R}^2$,

$$\mathrm{d}f(\mathrm{B}_t^1, \mathrm{B}_t^2) = (\partial_1 f)\mathrm{dB}_t^2 + (\partial_2 f)\mathrm{dB}_t^2 + \frac{1}{2}(\Delta f)\mathrm{d}t.$$

For $f(x, y) = \log(x^2+y^2)$, $\Delta f = 0$, so $\log \mathrm{X}_t$ is a local martingale[5] Let $0 < r < |a| < \mathrm{R}$, and $\mathrm{T}_x = \inf\{t \geqslant 0 \mid |\mathrm{B}_t| = x\}$. The stopping time theorem applied to $\log(\mathrm{X}_{t\wedge \mathrm{T}_\mathrm{R}\wedge \mathrm{T}_r})$ (this is a bounded, hence uniformly integrable martingale) yields

$$\mathbb{E}(\log(\mathrm{X}_{\mathrm{T}_\mathrm{R}\wedge \mathrm{T}_r})) = \log |a|. \tag{4.8}$$

As the one-dimensional Brownian motion is recurrent, $\mathrm{T}_\mathrm{R} < \infty$ almost surely, so $\mathbb{P}(\mathrm{T}_r < \mathrm{T}_\mathrm{R}) + \mathbb{P}(\mathrm{T}_\mathrm{R} < \mathrm{T}_r) = 1$, and (4.9) means

$$(\log r)\mathbb{P}(\mathrm{T}_r < \mathrm{T}_\mathrm{R}) + (\log \mathrm{R})\mathbb{P}(\mathrm{T}_\mathrm{R} < \mathrm{T}_r) = \log |a|.$$

This implies that

$$\mathbb{P}(\mathrm{T}_r < \mathrm{T}_\mathrm{R}) = \frac{\log \mathrm{R} - \log |a|}{\log \mathrm{R} - \log r},$$

so when $\mathrm{R} \to \infty$ we get by monotone convergence $\mathbb{P}(\mathrm{T}_r < \infty) = 1$, concluding the proof. □

Note that the previous theorem can be strengthened to prove that there are arbitrary large t such that $\mathrm{B}_t \in \mathrm{O}$. Indeed, for any $n \geqslant 0$,

$$\mathbb{P}(\exists t \geqslant n \mid \mathrm{B}_t \in 0) = \mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\exists t \geqslant n \mid \mathrm{B}_t \in \mathrm{O}} \mid \mathcal{F}_n\right)\right),$$

and $\mathbb{E}\left(\mathbb{1}_{\exists t \geqslant n \mid \mathrm{B}_t \in \mathrm{O}} \mid \mathcal{F}_n\right)$ is constantly one as, from the Markov property, this is also $\mathbb{P}(\exists t \geqslant 0 \mid \tilde{\mathrm{B}}_t \in \mathrm{O}_n)$ where $\mathrm{O}_n = \mathrm{O} + \mathrm{B}_n$ and $\tilde{\mathrm{B}}$ is a Brownian motion independent of B_n.

The above property is often referred to as the *recurrence* of the Brownian motion in dimension 2. It strongly contrasts with dimension $d \geqslant 3$: for any compact set outside of the initial point, with strictly positive probability the Brownian motion will never touch it. [6]

Theorem 4.16. *Let $d \geqslant 3$ and B be a Brownian motion of dimension d. If $\mathrm{K} \subset \mathbb{R}^d$ is compact, simply connected and $\mathrm{B}_0 \notin \mathrm{K}$, then*

$$\mathbb{P}(\exists t \geqslant 0 \mid \mathrm{B}_t \in \mathrm{K}) < 1.$$

Proof. First, as K is simply connected and bounded, and $\mathrm{B}_0 \notin \mathrm{K}$, there is a path $\gamma = (x(t), 0 \leqslant t \leqslant 1)$ such that $x(0) = \mathrm{B}_0$ and $x(1)$ is strictly separate from K by an hyperplane H. As K is closed there is some $\varepsilon > 0$ such that $\gamma_\varepsilon = \{x \in \mathbb{R}^d \mid \mathrm{dist}(x, \gamma) \leqslant \varepsilon\}$ is disjoint from K. Moreover, from Corollary 4.29 in the next section, one sees that

$$\mathbb{P}(\forall t \in [0, 1], \mathrm{B}_t \in \gamma_\varepsilon) > 0.$$

As a consequence, by considering (thanks to the Markov property) $(\mathrm{B}_{t+1} - \mathrm{B}_1, t \geqslant 0)$ instead of B and embedding K in a sphere not intersecting H, we just need to prove the following : if $\mathrm{B}_0 > r$, then

$$\mathbb{P}(\exists t \geqslant 0 \mid |\mathrm{B}_t| \leqslant r) < 1.$$

5. The function log diverges at 0, so strictly speaking we cannot apply this formula directly. But the result is true, because 0 is a polar point : writing $\mathrm{S}_n = \inf\{t \geqslant 0 \mid \mathrm{X}_t < \frac{1}{n}\}$, the Itô formula yields

$$\log \mathrm{X}_{t\wedge \mathrm{S}_n} = \log \mathrm{X}_0 + 2\int_0^{t\wedge \mathrm{S}_n} \frac{\mathrm{B}_s^1 \mathrm{dB}_s^1 + \mathrm{B}_s^2 \mathrm{dB}_s^2}{\mathrm{X}_s}.$$

From Theorem 4.14, when $n \to \infty$, $\mathrm{S}_n \to \infty$ a.s. so the above formula holds when replacing $t \wedge \mathrm{S}_n$ by t.

6. An isotropic dog with independent increments will always come back to his kennel, but such an astronaut will not necessarily find back the International Space Station.

To prove this, first the process $|\mathrm{B}_t|^{2-d}$ is a local martingale. Indeed, for a given function f of class $\mathscr{C}^2$ on $\mathbb{R}^d$,

$$\mathrm{d}f(\mathrm{B}_t^1,\dots,\mathrm{B}_t^d)=\sum_{k=1}^d(\partial_k f)\mathrm{dB}_t^k+\frac{1}{2}(\Delta f)\mathrm{d}t.$$

For $f(x_1,\dots,x_d)=(x_1^2+\dots+x_d^2)^{1-\frac{d}{2}}$, a calculation proves that $\Delta f=0$, so $|\mathrm{B}_t|^{2-d}$ is a local martingale (note that, as in the proof of Theorem 4.15, the considered function is not $\mathscr{C}^2$ in $\mathbb{R}^d$, so we first need to localize the process outside of arbitrarily small neighborhoods of the origin and then use the polarity of 0). Let $0<r<|a|<\mathrm{R}$, and $\mathrm{T}_x=\inf\{t\geqslant 0\mid |\mathrm{B}_t|=x\}$. The stopping time theorem applied to $|\mathrm{X}_{t\wedge\mathrm{T}_\mathrm{R}\wedge\mathrm{T}_r}|^{2-d}$ (this is a bounded, hence uniformly integrable martingale) yields

$$\mathbb{E}(|\mathrm{B}_{\mathrm{T}_\mathrm{R}\wedge\mathrm{T}_r}|^{2-d})=|a|^{2-d}. \tag{4.9}$$

As the one-dimensional Brownian motion is recurrent, $\mathrm{T}_\mathrm{R}<\infty$ almost surely, so $\mathbb{P}(\mathrm{T}_r<\mathrm{T}_\mathrm{R})+\mathbb{P}(\mathrm{T}_\mathrm{R}<\mathrm{T}_r)=1$, and (4.9) means

$$r^{2-d}\mathbb{P}(\mathrm{T}_r<\mathrm{T}_\mathrm{R})+\mathrm{R}^{2-d}\mathbb{P}(\mathrm{T}_\mathrm{R}<\mathrm{T}_r)=|a|^{2-d}.$$

This implies that

$$\mathbb{P}(\mathrm{T}_r<\mathrm{T}_\mathrm{R})=\frac{\mathrm{R}^{2-d}-|a|^{2-d}}{\mathrm{R}^{2-d}-r^{2-d}},$$

so when $\mathrm{R}\to\infty$ we get by monotone convergence

$$\mathbb{P}(\mathrm{T}_r<\infty)=\left(\frac{r}{|a|}\right)^{d-2}<1,$$

concluding the proof. □

Note that the above proof holds for $d=2$ (resp. 1), but it gives a trivial resp. contrary) conclusion. The above result alone is not sufficient to conclude that for $d\geqslant 3$, almost surely $|\mathrm{B}_t|\to\infty$ as $t\to\infty$. We say that in dimension greater than 3 the Brownian motion is *transcient*, and this is proved hereafter.

Theorem 4.17. *Let* B *be a Brownian motion in dimension* $d\geqslant 3$*. Then, almost surely,*

$$\lim_{t\to\infty}|\mathrm{B}_t|=\infty.$$

Proof. By a projection argument, we just need to prove it for $d=3$. By shifting B, we can suppose $\mathrm{B}_0=(1,0,0)$. The proof of Theorem 4.16 involved the fact that

$$\left(\frac{1}{|\mathrm{B}_t|},t\geqslant 0\right)$$

is a local martingale. As it is positive, it is also a supermartingale and converges almost surely, to some random variable X. We want to prove that $\mathrm{X}=0$ a.s. and as $\mathrm{X}\geqslant 0$ we just need to prove $\mathbb{E}(\mathrm{X})=0$. By Fatou's lemma,

$$\mathbb{E}(\mathrm{X})\leqslant\lim_{t\to\infty}\mathbb{E}\left(\frac{1}{|\mathrm{B}_t|}\right).$$

This limit is 0, because

$$\mathbb{E}\left(\frac{1}{|\mathrm{B}_t|}\right)=\frac{1}{t}\,\mathbb{E}\left(\left(\left(\mathcal{N}_1-\frac{1}{\sqrt{t}}\right)^2+\mathcal{N}_2^2+\mathcal{N}_3^2\right)^{-1/2}\right),$$

where $\mathcal{N}_1,\mathcal{N}_2,\mathcal{N}_3$ are independent standard normal variables, and the above expectation uniformly bounded : $1/|x|$ is integrable around 0 in dimension $d\geqslant 2$. This concludes the proof. □

The interested reader can also prove Theorem 4.17 by using the estimates on probability of hitting a ball in the proof of Theorem 4.16 and a Borel-Cantelli argument. One advantage of the above proof is that it gives a first example of a strict local martingale, justifying our efforts of the previous chapter to introduce this notion : when $d=3$,

$$\left(\frac{1}{|\mathrm{B}_t|}, t \geqslant 0\right)$$

is a local martingale but not a martingale, as its expectation converges to 0.

The previous discussion highlights the links between Brownian motion and Harmonic functions. The following gives some flavor about how from the Brownian motion allows to prove some results of Harmonic analysis.

Proposition 4.18. *Let $f : \mathrm{U} \to \mathbb{R}$ be a function of class $\mathscr{C}^2$, where U is open and connected. Then the following points are equivalent.*

(i) The function f is harmonic : $\Delta f = 0$ on U.

(ii) For any $x \in \mathbb{R}^d$ and any $r > 0$ such that the ball $\mathcal{B}(x,r)$ is included in U,

$$f(x) = \int f(y) \mathrm{d}\sigma_r(y),$$

where σ_r is the uniform measure, normalized to 1, on $\partial\mathcal{B}(x,r)$.

Proof. Let B be a Brownian motion beginning at x and $\tau_r = \inf\{t \geqslant 0 \mid |\mathrm{B}_t - x| = r\}$ The Itô formula yields

$$f(\mathrm{B}_{\tau_r}) = f(x) + \int_0^{\tau_r} \nabla f(\mathrm{B}_s) \cdot \mathrm{dB}_s + \frac{1}{2}\int_0^{\tau_r} \Delta f(\mathrm{B}_s) \mathrm{d}s.$$

Consequently, assuming (i),

$$\mathbb{E}(f(\mathrm{B}_{\tau_r})) = f(x) + \mathbb{E}\left(\int_0^{\tau_r} \nabla f(\mathrm{B}_s) \cdot \mathrm{dB}_s\right)$$

As the Brownian motion is invariant by rotations (cf previous section), the left hand side is $\int f(y)\mathrm{d}\sigma_r(y)$. Moreover, as $|\nabla f|$ is uniformly bounded on the ball $\mathcal{B}(x,r)$ with center x and radius r, $\left(\int_0^{t\wedge\tau_r} \nabla f(\mathrm{B}_s) \cdot \mathrm{dB}_s, t \geqslant 0\right)$ is a martingale with the expectation of its bracket being uniformly bounded[7] by

$$\sup_{\mathcal{B}(x,r)} |\nabla f|^2 \, \mathbb{E}(\tau_r) < \infty.$$

As a consequence, it is uniformly integrable and in particular, the stopping time theorem at time τ_r implies that

$$\mathbb{E}\left(\int_0^{\tau_r} \nabla f(\mathrm{B}_s) \cdot \mathrm{dB}_s\right) = 0,$$

proving (ii). Assume now that you have the mean value condition, and suppose that, at some point $x \in \mathrm{U}$, $\Delta f(x) \neq 0$. Up to considering $-f$, assume $\Delta f(x) > 0$ for example. Then, by continuity, one can find $r > 0$ such that $\Delta f > 0$ on $\mathcal{B}(x,r) \subset \mathrm{U}$. From the previous discussion, the Itô formula and the uniformly integrable aspect of the stochastic integral yield

$$\mathbb{E}(f(\mathrm{B}_{\tau_r})) = f(x) + \frac{1}{2}\mathbb{E}\left(\int_0^{\tau_r} \Delta f(\mathrm{B}_s) \mathrm{d}s\right)$$

7. to prove that $\mathbb{E}(\tau_r) < \infty$, just note that it is sufficient to prove it when $d=1$, and it is true by Corollary 3.17.

Condition (i) and the invariance of Brownian motion by rotation therefore yield

$$\mathbb{E}\left(\int_0^{\tau_r} \Delta f(\mathrm{B}_s)\mathrm{d}s\right) = 0.$$

As Δf is strictly positive along this trajectory this is only possible if $\tau_r = 0$ almost surely, which is not the case, concluding the proof : $\Delta f(x) = 0$ necessarily. □

As an exercise, the reader can prove in the same way the maximum principle for subharmonic functions.

Consider now the following classical *Dirichlet problem* from partial differential equations. For a connected open subset $\mathrm{U} \subset \mathbb{R}^d$, bounded, with boundary $\partial\mathrm{U}$, and a continuous function $\varphi : \partial\mathrm{U} \to \mathbb{R}$, we want to find a twice continuously differentiable function $u : \mathrm{U} \to \mathbb{R}$ such that

$$\begin{cases} \Delta u(x) = 0 & \text{for } x \in \mathrm{U} \\ u(x) = \varphi(x) & \text{for } x \in \partial\mathrm{U} \end{cases} \tag{4.10}$$

The existence of a solution depends on the smoothness of the boarder $\partial\mathrm{U}$. In the following, we will say that a point $x \in \partial\mathrm{U}$ satisfies the Poincaré cone condition (or is regular) if there exists some $\delta > 0$ and a cone C with vertex x and strictly positive angle such that

$$\mathrm{C} \cap \mathcal{B}(x, \delta) \subset \mathrm{U}^{\mathrm{c}},$$

where $\mathcal{B}(x, \delta)$ is the ball with center x and radius δ, and U^{c} is the complement of U in $\mathbb{R}^d$.

Theorem 4.19. *Let* U *and* φ *satisfy the above hypotheses, and consider the Dirichlet problem* (4.10).

(i) If there is a solution u *to the problem, then it coincides with*

$$v(x) := \mathbb{E}\left(\varphi(\mathrm{B}_\tau) \mid \mathrm{B}_0 = x\right),$$

where B *is a* d*-dimensional Brownian motion and* $\tau = \inf\{t \geqslant 0 \mid \mathrm{B}_t \in \partial\mathrm{U}\}$. *In particular, there is at most one solution.*

(ii) The above v *is harmonic and continuous at any regular point of* $\partial\mathrm{U}$. *In particular, if* $\partial\mathrm{U}$ *is everywhere regular,* v *is the unique solution of the Dirichlet problem.*

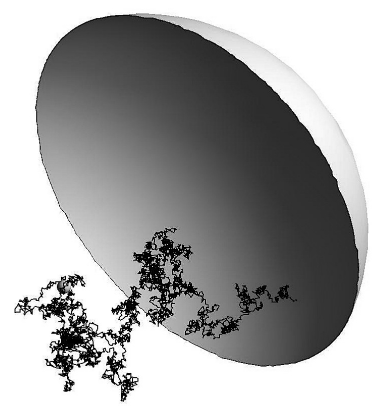

Figure 4.1. A Tridimensional Brownian motion till hitting a sphere.

Proof. To prove (i), by the Itô formula

$$u(\mathrm{B}_\tau) = u(x) + \int_0^\tau \nabla u(\mathrm{B}_s) \cdot \mathrm{dB}_s + \frac{1}{2}\int_0^\tau \Delta u(\mathrm{B}_s)\mathrm{d}s,$$

where B is a Brownian motion beginning at x. By the same argument as in the proof of Proposition 4.18, the expectation of the above stochastic integral is 0 (because $\mathbb{E}(\tau) < \infty$ and $|\nabla u|$ is bounded). As a consequence,

$$\mathbb{E}(u(\mathrm{B}_\tau)) = u(x).$$

But the left hand side is also $\mathbb{E}\left(\varphi(\mathrm{B}_\tau) \mid \mathrm{B}_0 = x\right)$, concluding the proof of (i).

Point (ii) is more subtle. We first prove that v is harmonic. Take a ball $\mathcal{B}(x,\delta) \subset \mathrm{U}$, and $\tilde{\tau} = \inf\{t \geqslant 0 \mid \mathrm{B}_t \notin \mathcal{B}(x,\delta)\}$. Then as $\tilde{\tau} \leqslant \tau$, $\mathcal{F}_{\tilde{\tau}} \subset \mathcal{F}_\tau$, so using the strong Markov property for the Brownian motion,

$$\begin{aligned} v(x) = \mathbb{E}_x\left(\varphi(\mathrm{B}_\tau)\right) &= \mathbb{E}_x\left(\mathbb{E}_x\left(\varphi(\mathrm{B}_\tau) \mid \mathcal{F}_{\tilde{\tau}}\right)\right) \\ &= \mathbb{E}_x\left(\mathbb{E}_{\mathrm{B}_{\tilde{\tau}}}\left(\varphi(\mathrm{B}_\tau)\right)\right) = \mathbb{E}_x(v(\mathrm{B}_{\tilde{\tau}})) = \int v(y)\sigma_r(\mathrm{d}y), \end{aligned}$$

where we used the invariance of B by rotation around x in the last equality, and σ_r is the uniform measure, normalized to 1, on the sphere with radius r and center x. As a consequence, using Proposition 4.18, v is harmonic in U.

We still need to prove that v is continuous at regular points. Let $x \in \partial\mathrm{U}$ be a regular point and take $\delta > 0$, a cone C with vertex x and strictly positive angle such that

$$\mathrm{C} \cap \mathcal{B}(x,\delta) \subset \mathrm{U}^{\mathrm{c}}.$$

Take a given $\varepsilon > 0$. As φ is continuous on $\partial\mathrm{U}$, we can chose the above δ such that

$$\sup_{y \in \partial\mathrm{U} \cap \mathcal{B}(x,\delta)} |\varphi(y) - \varphi(x)| < \varepsilon.$$

For a Brownian motion B, assume that there is some $0 < \theta < \delta$ such that if $|z - x| < \theta$, $z \in \mathrm{U}$,

$$\mathbb{P}_z(\tau_{\mathrm{C}} < \tau_\delta) > 1 - \varepsilon, \tag{4.11}$$

where this is the probability for a Brownian motion starting at z, τ_{C} is the hitting time of $\mathrm{C} \cap \mathcal{B}(x,\delta)$ and τ_δ is the hitting time of $\partial\mathcal{B}(x,\delta)$. Then this would imply that for $|z - x| < \theta$, $z \in \mathrm{U}$,

$$|v(z) - v(x)| = |\,\mathbb{E}_z(\varphi(\mathrm{B}_\tau) - \varphi(x))| \leqslant 2\|\varphi\|_\infty \mathbb{P}_z(\tau_\delta < \tau_{\mathrm{C}}) + \varepsilon \leqslant \varepsilon(2\|\varphi\|_\infty + 1).$$

Hence continuity will follow if we can prove (4.11). For this, by rotation invariance we can assume that B_1 corresponds to the axis of the cone, with angle α. Note that

$$\begin{aligned} \{\tau_{\mathrm{C}} < \tau_\delta\} \subset \cup_{t>0} \Bigg(&\left\{ \mathrm{B}_1(t) > \frac{1}{\tan\alpha}\left(\mathrm{B}_2(t)^2 + \cdots + \mathrm{B}_d(t)^2\right)^{1/2} \right\} \\ &\cap \cap_{k=1}^d \left\{ \sup_{[0,t]} |\mathrm{B}_k| < c \right\} \Bigg) \end{aligned}$$

for some absolute constant $c > 0$ depending only on δ and α. When $t \to 0$, these last d events have a probability converging to 1. Concerning the first event, the probability that it happens for some arbitrarily small $t > 0$ converges to 1 as well : this is an easy consequence of the iterated logarithm law ($\mathrm{B}_1(t)$ gets as big as $\sqrt{t\log(-\log t)}$ almost surely) and the independence of B_1 from the other $(\mathrm{B}_i)_{i=2}^d$. □

Concerning links with harmonic analysis, stochastic processes yield easy proofs of classical statements like the following famous Liouville's theorem.

Theorem 4.20. *Let $d \geqslant 1$. Bounded harmonic functions on $\mathbb{R}^d$ are constant.*

Proof. Let $f : \mathbb{R}^d \to [-m, m]$ be an harmonic function. As a direct consequence of the Itô formula and harmonicity, for any $t > 0$

$$\mathbb{E}_x\left(f(\mathrm{B}_t)\right) = f(x) + \mathbb{E}\left(\int_0^t \nabla f(\mathrm{B}_s) \cdot \mathrm{dB}_s\right) \tag{4.12}$$

where B is a Brownian motion (beginning at x). Chose x and y distinct points in $\mathbb{R}^d$, note $\mathscr{S}^{(n)}$ the sphere with center $\frac{x+y}{2}$ and radius $n > \|\frac{x-y}{2}\|$, and H the median hyperplane between x and y. Note

$$\mathrm{T}_n = \inf\{t \geqslant 0 \mid \mathrm{B}_t \in \mathscr{S}^{(n)}\},\ \mathrm{T}_\mathrm{H} = \inf\{t \geqslant 0 \mid \mathrm{B}_t \in \mathrm{H}\}.$$

As $|\nabla f|$ is bounded inside the sphere $\mathscr{S}^{(n)}$, at time $t \wedge \mathrm{T}_n$ the expectation on the right hand side of (4.12) is 0, so

$$f(x) = \mathbb{E}_x\left(f(\mathrm{B}_{t\wedge \mathrm{T}_n})\right) = \mathbb{E}_x\left(f(\mathrm{B}_{t\wedge \mathrm{T}_n})\mathbb{1}_{\mathrm{T}_\mathrm{H}\leqslant t\wedge \mathrm{T}_n}\right) + \mathbb{E}_x\left(f(\mathrm{B}_{t\wedge \mathrm{T}_n})\mathbb{1}_{\mathrm{T}_\mathrm{H}> t\wedge \mathrm{T}_n}\right).$$

The same formula holds concerning $f(y)$, and from the reflection principle

$$\mathbb{E}_x\left(f(\mathrm{B}_{t\wedge \mathrm{T}_n})\mathbb{1}_{\mathrm{T}_\mathrm{H}\leqslant t\wedge \mathrm{T}_n}\right) = \mathbb{E}_y\left(f(\mathrm{B}_{t\wedge \mathrm{T}_n})\mathbb{1}_{\mathrm{T}_\mathrm{H}\leqslant t\wedge \mathrm{T}_n}\right).$$

We therefore get

$$\begin{aligned}|f(x) - f(y)| = |\mathbb{E}_x\left(f(\mathrm{B}_{t\wedge \mathrm{T}_n})\mathbb{1}_{\mathrm{T}_\mathrm{H}> t\wedge \mathrm{T}_n}\right) - \mathbb{E}_y\left(f(\mathrm{B}_{t\wedge \mathrm{T}_n})\mathbb{1}_{\mathrm{T}_\mathrm{H}> t\wedge \mathrm{T}_n}\right)| \\ \leqslant 2m\mathbb{P}(\mathrm{T}_\mathrm{H} > t \wedge \mathrm{T}_n).\end{aligned}$$

When both t and n go to ∞, this last probability converges to 0, concluding the proof. □

The links between harmonic functions and the Brownian motion satisfy additional features when considering in a more specific way dimension $d = 2$. For example, in the following discussion, it will appear that up to a time change, entire functions of the Brownian motion are still Brownian motions[8]. In the following, a complex process is said to be a local martingale if both its real and imaginary parts are local martingales.

Proposition 4.21. *Let* $\mathrm{Z} = \mathrm{X} + \mathrm{i}\mathrm{Y}$ *be a complex local martingale. Then there exists a unique complex process with finite variation beginning at* 0 *such that* $\mathrm{Z}^2 - \langle \mathrm{Z}\rangle$ *is a complex local martingale. Moreover, the four following statements are equivalent.*

(i) The process Z^2 *is a local martingale.*

(ii) The process $\langle \mathrm{Z}\rangle$ *is identically 0.*

(iii) The brackets of the real and imaginary parts satisfy $\langle \mathrm{X}\rangle = \langle \mathrm{Y}\rangle$ *and* $\langle \mathrm{X}, \mathrm{Y}\rangle = 0$.

(iv) There is a Brownian motion[9] *such that* $\mathrm{Z}_t = \mathrm{B}_{\langle \mathrm{X}\rangle_t}$.

Proof. The existence of the bracket is given by

$$\langle \mathrm{Z}\rangle = \langle \mathrm{X} + \mathrm{i}\mathrm{Y}, \mathrm{X} + \mathrm{i}\mathrm{Y}\rangle = \langle \mathrm{X}\rangle - \langle \mathrm{Y}\rangle + 2\mathrm{i}\langle \mathrm{X}, \mathrm{Y}\rangle,$$

which satisfies all required properties. Uniqueness is a consequence of Theorem 3.26, and the equivalence between (*i*), (*ii*), (*iii*) is an immediate consequence of the above formula. Moreover, (*iii*) implies (*iv*) by the general Dubins-Schwarz Theorem 4.13, and (*iv*) obviously implies (*iii*). □

A local martingale is called *conformal* if any of the above properties is true. To prove that entire functions of B are conformal (this is called the *conformal invariance property* of Brownian motion), let us first discuss the translation the Itô formula in the complex analysis setting. We note

$$\partial_z = \frac{1}{2}(\partial_x - \mathrm{i}\partial_y),\ \partial_{\bar z} = \frac{1}{2}(\partial_x + \mathrm{i}\partial_y).$$

8. This property was used in the proof of Theorem 4.14

9. defined on an enlarged probability space if $\langle \mathrm{X}\rangle_\infty < \infty$ with positive probability

A function $f : \mathbb{C} \to \mathbb{C}$ is called holomorphic (resp. harmonic) of $\partial_{\overline{z}} f = 0$ (resp. $\Delta f = 4\partial_z\partial_{\overline{z}} f = 0$). For a general f of class $\mathscr{C}^2$ and Z a conformal local martingale, the Itô formula takes the form

$$f(\mathrm{Z}_t) = f(\mathrm{Z}_0) + \int_0^t \partial z f(\mathrm{Z}_s)\mathrm{d}\mathrm{Z}_s + \int_0^t \partial_{\overline{z}} f(\mathrm{Z}_s)\mathrm{d}\overline{\mathrm{Z}}_s + \frac{1}{2}\int_0^t \Delta f(\mathrm{Z}_s)\mathrm{d}\langle \mathfrak{Re}(\mathrm{Z})\rangle_s.$$

As a consequence, if f is harmonic then $f(\mathrm{Z})$ is a local martingale as well, and if f is holomorphic then the local martingale $f(\mathrm{Z})$ has the following easy decomposition

$$f(\mathrm{Z}_t) = f(\mathrm{Z}_0) + \int_0^t f'(\mathrm{Z}_s)\mathrm{d}\mathrm{Z}_s. \tag{4.13}$$

Theorem 4.22. *Let f be entire*[10] *and non constant, and* B *a standard complex Brownian motion . Then $f(\mathrm{B})$ is a time-changed Brownian motion :*

$$f(\mathrm{B}_t) = f(\mathrm{B}_0) + \tilde{\mathrm{B}}_{\langle \mathrm{X}\rangle_t}$$

where $\tilde{\mathrm{B}}$ is a standard Brownian motion, and $\langle \mathrm{X}\rangle_t := \int_0^t |f'(\mathrm{B}_s)|^2\mathrm{d}s$ is strictly increasing and converges to ∞.

Proof. As f is an entire function, so is f^2, so $f^2(\mathrm{B})$ is a local martingale, so $f(\mathrm{B})$ is a conformal local martingale. By *(iv)* of Proposition 4.21 there is a Brownian motion $\tilde{\mathrm{B}}$ such that

$$f(\mathrm{B}_t) = f(\mathrm{B}_0) + \tilde{\mathrm{B}}_{\langle \mathrm{X}\rangle_t},$$

where $\mathrm{X} = \mathfrak{Re}(f(\mathrm{B}))$. From 4.13, $\langle \mathrm{X}\rangle_t = \int_0^t |f'(\mathrm{B}_s)|^2\mathrm{d}s$. As f' is entire and not identically 0, it has a countable set of zeros so $\langle \mathrm{X}\rangle$ is strictly increasing. It converges to ∞ almost surely thanks for example to the recurrence property of the bi-dimensional Brownian motion. □

As an example the above property implies the famous d'Alembert's theorem : any non-constant complex polynomial has a complex zero. Let P be such a polynomial, and Z a complex Brownian motion. From the previous theorem,

$$\mathrm{P}(\mathrm{B}_t) = \tilde{\mathrm{B}}_{\langle \mathfrak{Re}(\mathrm{P}(\mathrm{B}_t))\rangle_t},$$

for some Brownian motion B, with the bracket going to ∞. As $\tilde{\mathrm{B}}$ is recurrent, this implies that for any $\varepsilon > 0$

$$\{z : |\mathrm{P}(z)| \leqslant \varepsilon\} \neq \varnothing,$$

concluding the proof, the intersection of nonempty embedded closed sets being nonempty. Another example of application for the conformal invariance property is about the winding number of the bi-dimensional Brownian motion around one point. This is known as *Spitzer's law.*

Theorem 4.23. *Let* B *be a planar Brownian motion beginning at $\mathrm{B}_0 \neq 0$. Define its argument at time t, θ_t continuously from $\theta_0 \in [0, 2\pi)$ at time 0. Then, as $t \to \infty$,*

$$\frac{\theta_t}{\log t} \xrightarrow{\text{law}} \frac{\mathrm{C}}{2},$$

where C *is a standard Cauchy random variable*[11].

10. i.e. everywhere holomorphic
11. This means that C has density $\frac{1}{\pi(1+x^2)}$ with respect to the Lebesgue measure.

Proof. By invariance by rotation and scaling, we can assume $B_0 = (1,0)$. From Theorem 4.22, if X and Y are independent standard real Brownian motions, then $\tilde{B}$ defined by

$$\tilde{B}_{\int_0^t e^{2X_s}ds} = e^{X_t + iY_t}$$

is a complex Brownian motion beginning at $(1,0)$. Let $a > 1$ and $S_a = \inf\{t \geqslant 0 \mid |B_t| = a\}$. We will first prove that

$$\frac{\arg B_{S_a}}{\log a} \sim C.$$

The winding number $\arg B_{S_a}$ is also the value $Y_{T_{\log a}}$ where $T_x = \inf\{t \geqslant 0 \mid X_t = x\}$. By scaling, proving that $Y_{T_1} \sim C$ is sufficient : we therefore just need to know the law of T_1 and compose it with that of Y.

- The distribution of T_1 is well-known, for example by the reflection principle 2.20 :
$$\mathbb{P}(T_1 \leqslant t) = \mathbb{P}(\sup_{[0,t]} X \geqslant 1) = \mathbb{P}(|X_t| \geqslant 1) = \mathbb{P}\left(\frac{1}{X_1^2} \leqslant t\right),$$
so $T_1 \sim 1/\mathcal{N}^2$ where $\mathcal{N}$ is a standard normal random variable.
- The density of Y_{T_1} is also that of $\sqrt{T_1}Y_1$, by conditioning on T_1. Hence it has the same law as $\mathcal{N}'/\mathcal{N}$, both random variables being independent and standard Gaussian. This is known to be a Cauchy distribution (for example by calculating the characteristic function).

From the previous discussion,

$$\frac{\arg B_{S_{\sqrt{t}}}}{\log t} \sim \frac{C}{2},$$

hence the proof will be complete if we can show that

$$\frac{\arg B_t - \arg B_{S_{\sqrt{t}}}}{\log t}$$

converges to 0 in probability. For this, consider τ the inverse of the clock, defined by

$$\int_0^{\tau_t} e^{2X_s} ds = t.$$

For some parameters $\delta \in (0,1)$ (it will go to 0) such that $\delta\sqrt{t} \geqslant 1$ and $\lambda > 0$ (it will go to ∞), consider the events

$$\mathcal{A} = \left\{S_{\delta\sqrt{t}} \leqslant t \leqslant S_{\delta^{-1}\sqrt{t}}\right\},$$
$$\mathcal{B} = \left\{\sup_{[\tau_{S_{\delta\sqrt{t}}}, \tau_{S_{\delta^{-1}\sqrt{t}}}]} |Y_s - Y_{\tau_{S_{\delta\sqrt{t}}}}| \leqslant \frac{\lambda}{2}\right\}.$$

Then $\mathcal{A} \cap \mathcal{B} \subset \left\{|\arg B_t - \arg B_{S_{\sqrt{t}}}| \leqslant \lambda\right\}$, so we just need to prove that the probability of A goes to 1 as $\delta \to 0$, and that of B goes to 1 as $\lambda \to \infty$, for any given δ.

- Concerning $\mathcal{A}$ things are easy because
$$\{|B_t| > \delta\sqrt{t}\} \cap \{\sup_{[0,t]} |B_s| < \delta^{-1}\sqrt{t}\} \subset \mathcal{A},$$
and both events have a probability going to 1 a $\delta \to 0$, independently of t by scaling.

- Concerning $\mathcal{B}$, note that

$$\tau_{\mathrm{S}_a} = \inf\{u \geqslant 0 \mid \mathrm{X}_u = \log a\},$$

so from the strong Markov property

$$\tau_{\mathrm{S}_{\delta^{-1}\sqrt{t}}} - \tau_{\mathrm{S}_{\delta\sqrt{t}}} \sim \mathrm{T}_{-2\log\delta},$$

where as previously $\mathrm{T}_x = \inf\{t \geqslant 0 \mid \mathrm{X}_t = x\}$. As Y is independent from X, this means that

$$\sup_{[\tau_{\mathrm{S}(\delta\sqrt{t})}, \tau_{\mathrm{S}(\delta^{-1}\sqrt{t})}]} |\mathrm{Y}_s - \mathrm{Y}_{\tau_{\mathrm{S}(\delta\sqrt{t})}}| \sim \sup_{[0,\Delta]} |\mathrm{Y}_s|,$$

where Δ is an almost surely finite random variable independent of Y, and with law depending on δ but not on t. As a consequence, $\mathbb{P}(\mathcal{B})$ is independent of t and goes to 1 as $\lambda \to \infty$.

This concludes the proof. □

4. The Girsanov theorem

In this section, the influence of the probability measure $\mathbb{P}$ on the semimartingale notion is discussed. As an example a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leqslant t\leqslant 1}, \mathbb{P})$, and a Brownian motion B on this space, one can define a new probability measure by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \frac{\sup_{[0,1]} \mathrm{B}}{\mathbb{E}\left(\sup_{[0,1]} \mathrm{B}\right)}.$$

The process B is certainly not a $\mathbb{Q}$-Brownian motion, as the symmetry property does not hold anymore, but is-it a $\mathbb{Q}$-semimartingale ? The main motivations for such questions are the following :

- practicing importance sampling on trajectories spaces ;
- deriving properties of the Wiener measure itself by looking at its behavior under some transformations ;
- just like changing variables allows to solve equations, changing the measure on the path space is the main tool to perform some expectations.

Before stating the general result, we need the following two propositions.

Proposition 4.24. *Assume $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{F}(= \mathcal{F}_\infty)$. For $t \in [0,\infty]$, note*

$$\mathrm{D}_t = \left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t}.$$

(i) The process D *is a uniformly integrable martingale.*

(ii) If D *is assumed right-continuous[12], then for any stopping time* T

$$\mathrm{D}_\mathrm{T} = \left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_\mathrm{T}}.$$

(iii) If D *is assumed right-continuous, and $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F}$ then, for any $t \geqslant 0$, $\mathrm{D}_t > 0$.*

12. This is not a very restrictive hypothesis, one can show that there is always a right-continuous version of the process D, see [12].

Proof. For (i), note that for any $t > 0$ and $\mathrm{A} \in \mathcal{F}_t$

$$\mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}}\mathrm{D}_t) = \mathbb{E}_{\mathbb{Q}}(\mathbb{1}_{\mathrm{A}}) = \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}}\mathrm{D}) = \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}}\,\mathbb{E}(\mathrm{D} \mid \mathcal{F}_t)).$$

where $\mathrm{D} = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\big|_{\mathcal{F}}$. By uniqueness of the Radon-Nikodym derivative, this implies $\mathrm{D}_t = \mathbb{E}(\mathrm{D} \mid \mathcal{F}_t)$ almost surely, concluding the proof as $\mathrm{D} \in \mathrm{L}^1$.

Point (ii) relies on the stopping time Theorem 3.13, which implies that $\mathrm{D}_{\mathrm{T}} = \mathbb{E}(\mathrm{D} \mid \mathcal{F}_{\mathrm{T}})$. As a consequence, for any $\mathrm{A} \in \mathcal{F}_{\mathrm{T}}$,

$$\mathbb{E}_{\mathbb{P}}\left(\mathbb{1}_{\mathrm{A}}\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_{\mathrm{T}}}\right) = \mathbb{E}_{\mathbb{Q}}(\mathbb{1}_{\mathrm{A}}) = \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}}\mathrm{D}) = \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}}\mathrm{D}_{\mathrm{T}}),$$

so $\mathrm{D}_{\mathrm{T}} = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\big|_{\mathcal{F}_{\mathrm{T}}}$.

Finally, (iii) follows from (ii) : for $\mathrm{T} = \inf\{t \geqslant 0 \mid \mathrm{D}_t = 0\}$, by right-continuity $\mathrm{D}_{\mathrm{T}} = 0$ on $\{\mathrm{T} < \infty\}$, so

$$\mathbb{Q}(\mathrm{T} < \infty) = \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{T}<\infty}\mathrm{D}_{\mathrm{T}}) = 0.$$

As $\mathbb{Q} \gg \mathbb{P}$, this implies $\mathbb{P}(\mathrm{T} < \infty) = 0$, as expected. □

Proposition 4.25. *Let* D *be a strictly positive continuous local martingale. Then there exists a unique continuous local martingale* L *such that*

$$\mathrm{D}_t = \mathcal{E}(\mathrm{L})_t := e^{\mathrm{L}_t - \frac{1}{2}\langle \mathrm{L}\rangle_t}.$$

Proof. For uniqueness, note that if L and $\tilde{\mathrm{L}}$ are solutions, then

$$\mathrm{L} - \tilde{\mathrm{L}} = \frac{1}{2}\left(\langle \mathrm{L}\rangle - \langle \tilde{\mathrm{L}}\rangle\right)$$

is a finite variation process and also a local martingale, so it is indistinguishable from 0. For the existence, the choice

$$\mathrm{L}_t = \log \mathrm{D}_0 + \int_0^t \frac{\mathrm{d}\mathrm{D}_s}{\mathrm{D}_s}$$

makes sense as D is strictly positive and, by the Itô formula,

$$\log \mathrm{D}_t = \log \mathrm{D}_0 + \int_0^t \frac{\mathrm{d}\mathrm{D}_s}{\mathrm{D}_s} - \frac{1}{2}\int_0^t \frac{\mathrm{d}\langle \mathrm{D}\rangle_s}{\mathrm{D}_s^2} = \mathrm{L}_t - \frac{1}{2}\langle \mathrm{L}\rangle_t,$$

concluding the proof. □

Both results imply the following, known as Girsanov's theorem, or Cameron-Martin's theorem in the special case of deterministic shifts.

Theorem 4.26. *Let $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F}$. Assume that the process defined by*

$$\mathrm{D}_t = \left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t}$$

is continuous, and write[13] *it as* $\mathrm{D} = \mathcal{E}(\mathrm{L})$, *with* L *a local martingale. Then if* M *is a* $\mathbb{P}$*-local martingale the process*

$$\tilde{\mathrm{M}} := \mathrm{M} - \langle \mathrm{M}, \mathrm{L}\rangle$$

is a $\mathbb{Q}$*-local martingale.*

13. This is possible by Proposition 4.24 (iii) and Proposition 4.25

Proof. We first prove that if a process XD is a $\mathbb{P}$-local martingale, then X is a $\mathbb{Q}$-local martingale. For this, we show that for T any stopping time and X a continuous adapted process such that $(\mathrm{XD})^{\mathrm{T}}$ is a $\mathbb{P}$-martingale, then X^{T} is a $\mathbb{Q}$-martingale :

- First the integrability condition is satisfied : from Proposition 4.24 (*ii*),

$$\mathbb{E}_{\mathbb{Q}}(|\mathrm{X}_t^{\mathrm{T}}|) = \mathbb{E}_{\mathbb{P}}(|\mathrm{X}_t^{\mathrm{T}}||\mathrm{D}_t^{\mathrm{T}}|) < \infty$$

  because $(\mathrm{XD})^{\mathrm{T}}$ is a $\mathbb{P}$-martingale.

- Moreover, for $s < t$ and $\mathrm{A} \in \mathcal{F}_s$, still using Proposition 4.24 (*ii*) for the second and fourth equalities, and the martingale property for $(\mathrm{XD})^{\mathrm{T}}$ in the third one,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}(\mathbb{1}_{\mathrm{A}}\mathrm{X}_t^{\mathrm{T}}) &= \mathbb{E}_{\mathbb{Q}}(\mathbb{1}_{\mathrm{A}\cap\{\mathrm{T}\leqslant s\}}\mathrm{X}_t^{\mathrm{T}}) + \mathbb{E}_{\mathbb{Q}}(\mathbb{1}_{\mathrm{A}\cap\{\mathrm{T}>s\}}\mathrm{X}_t^{\mathrm{T}}) \\
&= \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}\cap\{\mathrm{T}\leqslant s\}}\mathrm{X}_t^{\mathrm{T}}\mathrm{D}_{t\wedge\mathrm{T}}) + \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}\cap\{\mathrm{T}>s\}}\mathrm{X}_t^{\mathrm{T}}\mathrm{D}_{t\wedge\mathrm{T}}) \\
&= \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}\cap\{\mathrm{T}\leqslant s\}}\mathrm{X}_s^{\mathrm{T}}\mathrm{D}_{t\wedge\mathrm{T}}) + \mathbb{E}_{\mathbb{P}}(\mathbb{1}_{\mathrm{A}\cap\{\mathrm{T}>s\}}\mathrm{X}_s^{\mathrm{T}}\mathrm{D}_{s\wedge\mathrm{T}}) \\
&= \mathbb{E}_{\mathbb{Q}}(\mathbb{1}_{\mathrm{A}\cap\{\mathrm{T}\leqslant s\}}\mathrm{X}_s^{\mathrm{T}}) + \mathbb{E}_{\mathbb{Q}}(\mathbb{1}_{\mathrm{A}\cap\{\mathrm{T}>s\}}\mathrm{X}_s^{\mathrm{T}}) \\
&= \mathbb{E}_{\mathbb{Q}}(\mathbb{1}_{\mathrm{A}}\mathrm{X}_s^{\mathrm{T}}).
\end{aligned}$$

  Hence $\mathrm{X}_s^{\mathrm{T}} = \mathbb{E}_{\mathbb{Q}}(\mathrm{X}_t^{\mathrm{T}} \mid \mathcal{F}_s)$.

Thanks to the previous discussion, $\tilde{\mathrm{M}}\mathrm{D}$ being a $\mathbb{P}$-local martingale is enough to prove that $\tilde{\mathrm{M}}$ is a $\mathbb{Q}$-local martingale. The Itô formula yields

$$\begin{aligned}
\mathrm{d}\left((\mathrm{M} - \langle \mathrm{M},\mathrm{L}\rangle)\mathrm{D}\right)_t &= (\mathrm{M} - \langle \mathrm{M},\mathrm{L}\rangle)_t\mathrm{dD}_t + (\mathrm{dM}_t - \mathrm{d}\langle \mathrm{M},\mathrm{L}\rangle_t)\mathrm{D}_t + \mathrm{d}\langle \mathrm{M},\mathrm{D}\rangle_t \\
&= (\mathrm{M} - \langle \mathrm{M},\mathrm{L}\rangle)_t\mathrm{dD}_t + (\mathrm{dM}_t - \mathrm{d}\langle \mathrm{M},\mathrm{L}\rangle_t)\mathrm{D}_t + \mathrm{D}_t\mathrm{d}\langle \mathrm{M},\mathrm{L}\rangle_t \\
&= (\mathrm{M} - \langle \mathrm{M},\mathrm{L}\rangle)_t\mathrm{dD}_t + \mathrm{D}_t\mathrm{dM}_t,
\end{aligned}$$

where in the second equality we used the fact that $\mathrm{dD}_t = \mathrm{D}_t\mathrm{dL}_t$. So $\tilde{\mathrm{M}}\mathrm{D}$ is a stochastic integral with respect to the $\mathbb{P}$-local martingales D and M, so it is a $\mathbb{P}$-local martingale, concluding the proof. □

One particularly interesting case of the above result is the case when M is a $\mathbb{P}$-Brownian motion.

Corollary 4.27. *One can replace local martingale by Brownian motion in the hypothesis and conclusion of Theorem 4.26.*

Proof. If M is a P-Brownian motion, then $\tilde{\mathrm{M}}$ is a $\mathbb{Q}$-local martingale (by Girsanov's theorem) and its bracket is t (by the sum of quadratic increments formula in Theorem 3.26, the bracket does not depend on the underlying probability measure), so Lévy's criterion Theorem 4.11 applies. □

The condition $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F}$ in the hypothesis of the Girsanov theorem is quite restrictive. The following lines explain why, and a way to overcome this. As an example, if $\mathrm{M} = \mathrm{B}$ is a Brownian motion, one could try to use Girsanov's theorem till infinite time to study the properties of the Brownian motion with drift $\tilde{\mathrm{M}} = (\mathrm{B}_t - \nu t, t \geqslant 0)$. For this, we want a measure $\mathbb{Q}$ on $\mathcal{F}$ such that for any $t \geqslant 0$

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = e^{\nu \mathrm{B}_t - \frac{\nu^2}{2}t}. \tag{4.14}$$

By Theorem 1.3.5 in [18], such a measure $\mathbb{Q}$ exists. But it is not absolutely continuous with respect to the Wiener measure $\mathbb{P}$: the reader could prove that if $\nu > 0$ the event $\{\liminf \mathrm{B}_t/t > 0\}$ is almost sure for $\mathbb{Q}$ and its complement is almost sure for $\mathbb{P}$. To avoid this problem, the applications of the Girsanov theorem can be performed up to a finite horizon : if $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F}_t$, then $\tilde{\mathrm{M}}$ is a $\mathbb{Q}$-local martingale when one only considers its restriction to $[0,t]$: this is true by applying the general statement with a modified filtration such that $\mathcal{F}_u = \mathcal{F}_t$ when $u > t$.

Coming back to (4.14), we get that $(\mathrm{B}_s - \nu s, 0 \leqslant s \leqslant t)$ is a $\mathbb{Q}$-Brownian motion, hence for any bounded continuous functional F,

$$\mathbb{E}_{\mathbb{Q}}\left(\mathrm{F}(\mathrm{B}_s - \nu s, 0 \leqslant s \leqslant t)\right) = \mathbb{E}_{\mathbb{P}}\left(\mathrm{F}(\mathrm{B}_s, 0 \leqslant s \leqslant t)\right).$$

When F is a cylindric function, this is just a change of variables formula for finite-dimensional Gaussian measures. If F depends only on the trajectory up to a stopping time $\mathrm{T} \leqslant t$ a.s. the above equation together with Proposition 4.24 *(ii)* yield

$$\mathbb{E}_{\mathbb{P}}\left(\mathrm{F}(\mathrm{B}_s - \nu s, 0 \leqslant s \leqslant \mathrm{T}) e^{\nu \mathrm{B}_{\mathrm{T}} - \frac{\nu^2}{2}\mathrm{T}}\right) = \mathbb{E}_{\mathbb{P}}\left(\mathrm{F}(\mathrm{B}_s, 0 \leqslant s \leqslant \mathrm{T})\right). \tag{4.15}$$

As an example of application, the above equation allows to prove the following law of hitting time of a Brownian motion with drift.

Corollary 4.28. *Let* $\mathrm{T}_a^{(\nu)} = \inf\{s \geqslant 0 \mid \mathrm{B}_s + \nu s = a\}$. *Then for any* $\lambda \geqslant 0$

$$\mathbb{E}_{\mathbb{P}}\left(\mathbb{1}_{\mathrm{T}_a^{(\nu)} < \infty} e^{-\lambda \mathrm{T}_a^{(\nu)}}\right) = e^{\nu a - |a|\sqrt{2\lambda + \nu^2}}.$$

Proof. Consider the random variable $\mathrm{F} = \mathbb{1}_{\mathrm{T}_a^{(\nu)} \leqslant t} e^{-\lambda \mathrm{T}_a^{(\nu)}}$ (which is $\mathcal{F}_{t \wedge \mathrm{T}_a}$-measurable) in (4.15). Then

$$\mathbb{E}_{\mathbb{P}}\left(\mathbb{1}_{\mathrm{T}_a^{(0)} \leqslant t} e^{-\lambda \mathrm{T}_a^{(0)}} e^{\nu \mathrm{B}_{t \wedge \mathrm{T}_a^{(0)}} - \frac{\nu^2}{2}\mathrm{T}_a^{(0)}}\right) = \mathbb{E}_{\mathbb{P}}\left(\mathbb{1}_{\mathrm{T}_a^{(\nu)} \leqslant t} e^{-\lambda \mathrm{T}_a^{(\nu)}}\right).$$

Dominated convergence as $t \to \infty$ yields

$$\mathbb{E}_{\mathbb{P}}\left(\mathbb{1}_{\mathrm{T}_a^{(0)} < \infty} e^{-\left(\lambda + \frac{\nu^2}{2}\right)\mathrm{T}_a^{(0)}} e^{\nu a}\right) = \mathbb{E}_{\mathbb{P}}\left(\mathbb{1}_{\mathrm{T}_a^{(\nu)} < \infty} e^{-\lambda \mathrm{T}_a^{(\nu)}}\right).$$

The left hand side is known thanks to Corollary 3.17, so

$$\mathbb{E}_{\mathbb{P}}\left(\mathbb{1}_{\mathrm{T}_a^{(\nu)} < \infty} e^{-\lambda \mathrm{T}_a^{(\nu)}}\right) = e^{\nu a - |a|\sqrt{2\lambda + \nu^2}},$$

the expected result. □

A simple yet sometimes useful consequence of Girsanov's theorem is that the Wiener measure gives positive mass to any open subset of continuous functions on $[0,1]$.

Corollary 4.29. *Let f be continuous on* $(0,1)$, $f(0) = 0$, *and note* $\mathrm{V}_\varepsilon(f)$ *the set of continuous functions g on* $(0,1)$, *beginning at 0, such that* $\sup_{(0,1)} |f(x) - g(x)| < \varepsilon$. *Then* $\mathbb{P}(\mathrm{V}_\varepsilon) > 0$.

Proof. First, for a given $\varepsilon > 0$, there exists a continuous f_0, beginning at 0, such that its ε-neighborhood has a strictly positive probability (otherwise the measure would be null). We want to prove it for general f By interpolation and up to choosing a smaller ε, we can suppose that $h := f - f_0$ is $\mathscr{C}^1$. In the Girsanov theorem take $\mathrm{M} = \mathrm{B}$, a Brownian motion, and $\mathrm{L}_t = \int_0^t \dot{h}_s \mathrm{dB}_s$. Then $\mathrm{M} - \langle \mathrm{M}, \mathrm{L} \rangle$ is $\mathrm{B} - h$, and

$$\mathbb{P}\left(|\mathrm{B} - f|_{\mathrm{L}^\infty(0,1)} < \varepsilon\right) = \mathbb{E}_{\mathbb{P}}\left(\mathbb{1}_{|\mathrm{B} - f_0|_{\mathrm{L}^\infty(0,1)} < \varepsilon} e^{\int_0^1 \dot{h}_s \mathrm{dB}_s - \frac{1}{2}\int_0^1 \dot{h}_s^2 \mathrm{d}s}\right).$$

This right hand side is strictly positive because $\mathbb{P}\left(|\mathrm{B} - f_0|_{\mathrm{L}^\infty(0,1)} < \varepsilon\right) > 0$ and $\int_0^1 \dot{h}_s \mathrm{dB}_s > -\infty$ almost surely. □

We finally consider the question of checking the hypotheses to apply Girsanov's theorem. In most applications, a $\mathbb{P}$-local martingale L is given, and one makes the choice

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = e^{\mathrm{L}_t - \frac{1}{2}\langle \mathrm{L}\rangle_t},$$

hoping that the expectation of the right hand side is 1. Note that $\mathcal{E}(\mathrm{L})$ is a positive local martingale, hence a supermartingale, hence it converges almost surely and

$$\mathbb{E}\left(\mathcal{E}(\mathrm{L})_\infty\right) \leqslant 1. \tag{4.16}$$

There is equality if and only if $\mathcal{E}(\mathrm{L})$ is a uniformly integrable martingale[14]. Moreover, if $\mathcal{E}(\mathrm{L})$ is uniformly integrable, then when defining

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}} = \mathcal{E}(\mathrm{L})_\infty,$$

we obviously have $\mathbb{Q} \sim \mathbb{P}$. Hence, checking the uniform integrability of an exponential local martingale is important to apply Girsanov's theorem, and a famous criterium by Novikov is given hereafter.

Theorem 4.30. *Let* L *be a local martingale beginning at* 0*. If*

$$\mathbb{E}\left(e^{\frac{1}{2}\langle \mathrm{L}\rangle_\infty}\right) < \infty, \tag{4.17}$$

then $\mathcal{E}(\mathrm{L})$ *is a uniformly integrable martingale.*

Proof. We first prove that (4.17) implies that L is uniformly integrable and that $\mathbb{E}\left(e^{\frac{1}{2}\mathrm{L}_\infty}\right) < \infty$, which in turn implies the result, as we will see after.

Obviously, by (4.17) $\mathbb{E}(\langle \mathrm{L}\rangle_\infty) < \infty$, so L is bounded in L^2 by Theorem 3.29. Hence L is a uniformly integrable martingale. Moreover, by a simple application of the Cauchy-Schwarz inequality and using (4.16),

$$\begin{aligned}
\mathbb{E}\left(e^{\frac{1}{2}\mathrm{L}_\infty}\right) &= \mathbb{E}\left(\mathcal{E}(\mathrm{L})_\infty^{\frac{1}{2}}\left(e^{\frac{1}{2}\langle \mathrm{L}\rangle_\infty}\right)^{\frac{1}{2}}\right)\\
&\leqslant \mathbb{E}\left(\mathcal{E}(\mathrm{L})_\infty\right)^{\frac{1}{2}}\mathbb{E}\left(e^{\frac{1}{2}\langle \mathrm{L}\rangle_\infty}\right)^{\frac{1}{2}}\\
&\leqslant \mathbb{E}\left(e^{\frac{1}{2}\langle \mathrm{L}\rangle_\infty}\right)^{\frac{1}{2}},
\end{aligned}$$

so $\mathbb{E}\left(e^{\frac{1}{2}\mathrm{L}_\infty}\right) < \infty$. Now, we want to prove the uniform integrability of $\mathcal{E}(\mathrm{L})$. From the discussion before Theorem 4.30, a sufficient condition is $\mathbb{E}(\mathcal{E}(\mathrm{L})_\infty) = 1$.

- First, as L is uniformly integrable, $\mathrm{L}_t = \mathbb{E}(\mathrm{L}_\infty \mid \mathcal{F}_t)$, so by convexity $e^{\frac{1}{2}\mathrm{L}_t} < \mathbb{E}\left(e^{\frac{1}{2}\mathrm{L}_\infty} \mid \mathcal{F}_t\right)$. This proves that $(e^{\frac{1}{2}\mathrm{L}_t}, t \geqslant 0)$ is uniformly integrable.

- Moreover, as the exponential is convex and increasing, $(e^{\frac{1}{2}\mathrm{L}_t}, t \geqslant 0)$ is a submartingale, uniformly integrable from the previous point. Hence, for any stopping time T, $e^{\frac{1}{2}\mathrm{L}_\mathrm{T}} < \mathbb{E}\left(e^{\frac{1}{2}\mathrm{L}_\infty} \mid \mathcal{F}_\mathrm{T}\right)$: the family

$$\{e^{\frac{1}{2}\mathrm{L}_\mathrm{T}} \mid \mathrm{T} \text{ stopping time}\}$$

14. If this is a uniformly integrable martingale, obviously we need to have $\mathbb{E}\left(\mathcal{E}(\mathrm{L})_\infty\right) = 1$ by the stopping time theorem. Reciprocally, if we have equality, as the function $t \mapsto \mathbb{E}\left(\mathcal{E}(\mathrm{L})_t\right)$ decreases, it needs to be constant. Hence, for $t > s$, $\mathbb{E}(\mathbb{E}(\mathcal{E}(\mathrm{L})_t \mid \mathcal{F}_s) - \mathcal{E}(\mathrm{L})_s) = 0$, and as the integrated term is non-positive, we have $\mathbb{E}(\mathcal{E}(\mathrm{L})_t \mid \mathcal{F}_s) = \mathcal{E}(\mathrm{L})_s$ almost surely : $\mathcal{E}(\mathrm{L})$ is a martingale. Moreover, as an easy consequence of Fatou's lemma

$$\mathcal{E}(\mathrm{L})_t \geqslant \mathbb{E}(\mathcal{E}(\mathrm{L})_\infty \mid \mathcal{F}_t),$$

and as both random variables above have the same expectation they are equal : $\mathcal{E}(\mathrm{L})_t \geqslant \mathbb{E}(\mathcal{E}(\mathrm{L})_\infty \mid \mathcal{F}_t)$, so $\mathcal{E}(\mathrm{L})$ is uniformly integrable.

is uniformly integrable.

- We would like to prove that $\{\mathcal{E}(\mathrm{L})_\mathrm{T} \mid \mathrm{T}\text{ stopping time}\}$ is uniformly integrable. This is slightly too much required. We can show that $\{\mathcal{E}(\lambda\mathrm{L})_\mathrm{T} \mid \mathrm{T}\text{ stopping time}\}$ is uniformly integrable for any $0<\lambda<1$. Take any $\mathrm{A}\in\mathcal{F}$ (typically, $\mathrm{A}=\{\mathcal{E}(\lambda\mathrm{L})_\mathrm{T}>x\}$). Then

$$\begin{aligned}\mathbb{E}\left(\mathbb{1}_\mathrm{A}\mathcal{E}(\lambda\mathrm{L})_\mathrm{T}\right) &= \mathbb{E}\left(\mathbb{1}_\mathrm{A}(\mathcal{E}(\mathrm{L})_\mathrm{T})^{\lambda^2}e^{\lambda(1-\lambda)\mathrm{L}_\mathrm{T}}\right)\\ &\leqslant \mathbb{E}\left(\mathcal{E}(\mathrm{L})_\mathrm{T}\right)^{\lambda^2}\mathbb{E}\left(\mathbb{1}_\mathrm{A}e^{\frac{\lambda}{1+\lambda}\mathrm{L}_\mathrm{T}}\right)^{1-\lambda^2}\\ &\leqslant \mathbb{E}\left(\mathbb{1}_\mathrm{A}e^{\frac{\lambda}{1+\lambda}\mathrm{L}_\mathrm{T}}\right)^{1-\lambda^2}\\ &= \mathbb{E}\left(\left(\mathbb{1}_\mathrm{A}e^{\frac{1}{2}\mathrm{L}_\mathrm{T}}\right)^{\frac{2\lambda}{1+\lambda}}\right)^{1-\lambda^2}\\ &\leqslant \mathbb{E}\left(\mathbb{1}_\mathrm{A}e^{\frac{1}{2}\mathrm{L}_\mathrm{T}}\right)^{2\lambda(1-\lambda)}\end{aligned}$$

  where we used in the second inequality that for a positive supermartingale M and any stopping time T, $\mathbb{E}(\mathrm{M}_\mathrm{T})\leqslant\mathrm{M}_0$ (this is true for bounded stopping times, and then for any of them by Fatou). The last inequality relies on the concavity of $x\mapsto x^a$ when $a<1$. The last term above being uniformly integrable (from our second point), the set $\{\mathcal{E}(\lambda\mathrm{L})_\mathrm{T} \mid \mathrm{T}\text{ stopping time}\}$ is uniformly integrable for any $0<\lambda<1$.

- This implies that $\mathcal{E}(\lambda\mathrm{L})$ is a uniformly integrable martingale.As a consequence,

$$1=\mathbb{E}\left(\mathcal{E}(\lambda\mathrm{L})_\infty\right)\leqslant\mathbb{E}\left(\mathcal{E}(\mathrm{L})_\infty\right)^{\lambda^2}\mathbb{E}\left(e^{\frac{1}{2}\mathrm{L}_\infty}\right)^{2\lambda(1-\lambda)}.$$

  Taking $\lambda\to1^-$ yields $\mathbb{E}(\mathcal{E}(\mathrm{L})_\infty)\geqslant1$.

This achieves the proof. □

Chapter 5

Stochastic differential equations

Ordinary differential equations aim to find, from local data

$$\frac{\mathrm{dX}_t}{\mathrm{d}t} = b(\mathrm{X}_t),$$

a global solution X. Some criteria on b (e.g. the Cauchy-Lipschitz theorem) yield to existence and uniqueness of solutions to such equations. They describe the evolution of a deterministic physical system. A random perturbation of such a system is pertinent in many contexts : to study stability of such trajectories (e.g. bifurcations in dynamical systems), or directly because the physics of the deterministic problem are too intricate, hence this mixing property is simplified as an independence hypothesis (e.g. pollen particles floating in water) or directly because the physical assumption includes randomness (e.g. quantum mechanics). Such a random equation can be written as

$$\mathrm{dX}_t = \sigma(\mathrm{X}_t)\mathrm{dB}_t + b(\mathrm{X}_t)\mathrm{d}t.$$

The meaning to give to this local evolution is that X is a semimartingale such that, almost surely, for any $t > 0$,

$$\mathrm{X}_t = \mathrm{X}_0 + \int_0^t (\sigma(\mathrm{X}_s)\mathrm{dB}_s + b(\mathrm{X}_s)\mathrm{d}s).$$

Note first that these random perturbations of ordinary differential equations are performed through a Gaussian noise. This is not very restrictive : by the Donsker Theorem 2.26, any independent increments in L^2 lead to a Brownian motion. A more subtle discussion is required concerning whether the solution of the stochastic differential equation is a deterministic function of the random trajectory B (strong existence and uniqueness of the solution) or if it given by a law of the trajectory (weak uniqueness of the solution). This point is the purpose of the next section. In Section 2, an analogue of the Cauchy-Lipschitz theorem for the existence of strong solutions is given, based of Picard's iteration as well. The strong Markov property associated to these solutions will allow us to discuss generalizations of the Dirichlet problem, already considered in Chapter 4, when the second order differential operator is not necessarily the Euclidean Laplacian anymore (Section 3). After that, a practical way to perform simulations of stochastic differential equations (e.g. for Monte Carlo purpose) is given, as one (of the many) application(s) of the Stroock-Varadhan martingale problems theory (Section 4). Finally, we will study in details the case $\sigma \equiv 1$, pertinent in filtering theory and revealing a wide variety of existence/uniqueness cases, in Section 5.

1. Weak and strong solutions

In the following definition, σ and b depend not only on the current point X_t but also on t and the entire trajectory till time t. Moreover, they are assumed to be multidimensional :

$$\sigma = (\sigma_{ij})_{1\leqslant i\leqslant d, 1\leqslant j\leqslant m}, \ b = (b_i)_{1\leqslant i\leqslant d},$$

each σ_{ij}, b_i, being a function from $\mathbb{R}_+ \times \mathscr{C}$ to $\mathbb{R}$, $\sigma(t,\omega), b(t,\omega) \in \mathcal{X}_t$ where $\mathcal{X}_t = \sigma(\omega_s, s \leqslant t)$.

Definition 5.1. *A solution to the equation*

$$\mathrm{dX}_t = \sigma(t, \mathrm{X})\mathrm{dB}_t + b(t, \mathrm{X})\mathrm{d}t$$

on $\mathbb{R}_+$ is the collection of :

- *a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathrm{P})$;*
- *a $(\mathcal{F}_t)$-Brownian motion $\mathrm{B} = (\mathrm{B}^{(1)}, \dots, \mathrm{B}^{(m)})$ with $\mathrm{B}^{(i)}$ a $(\mathcal{F}_t)$-martingale and $\langle \mathrm{B}^{(i)}, \mathrm{B}^{(j)}\rangle_t = \mathbb{1}_{i=j} t$;*
- *a $(\mathcal{F}_t)$-adapted process $\mathrm{X} = (\mathrm{X}^{(1)}, \dots, \mathrm{X}^{(d)})$ such that*

$$\mathrm{X}_t = \mathrm{X}_0 + \int_0^t \sigma(s, \mathrm{X})\mathrm{dB}_s + \int_0^t b(s, \mathrm{X})\mathrm{d}s.$$

This is abbreviated by saying that the process X is a solution of $\mathrm{E}(\sigma, b)$. When imposed to $\mathrm{X}_0 = x$, X is said to be a solution of $\mathrm{E}_x(\sigma, b)$.

This definition has an obvious extension to the solution of the stochastic differential equation with a finite time horizon T. The main notions of uniqueness and existence are the following.

Definition 5.2. *For the equation $\mathrm{E}(\sigma, b)$, we say that there is*

- *weak existence if for any $x \in \mathbb{R}^d$ there is a solution to $\mathrm{E}_x(\sigma, b)$;*
- *weak existence and uniqueness is for any $x \in \mathbb{R}^d$ there is a solution to $\mathrm{E}_x(\sigma, b)$ and all solutions to $\mathrm{E}_x(\sigma, b)$ have the same law ;*
- *pathwise uniqueness if, given $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathrm{P})$ and B, two solutions X and X' such that $\mathrm{X}_0 = \mathrm{X}'_0$ P-almost surely cannot be distinguished.*

Moreover, a solution X of $\mathrm{E}_x(\sigma, b)$ is said to be strong if X is $(\mathcal{B}_t)$-adapted, where $\mathcal{B}_t = \sigma(\mathrm{B}_s, s \leqslant t)$.

We now want to illustrate the above definitions by some examples. Suppose given a Brownian motion $(\beta_t)_{t\geqslant 0}$ on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathrm{P})$, for example $\mathcal{F}_t = \sigma(\beta_s, s \leqslant t)$ and P is the Wiener measure.

- *Weak existence.* Consider the stochastic differential equation

$$\mathrm{dX}_t = 3\,\mathrm{sgn}(\mathrm{X}_t)|\mathrm{X}_t|^{2/3}\mathrm{dB}_t + 3\,\mathrm{sgn}(\mathrm{X}_t)|\mathrm{X}_t|^{1/3}\mathrm{d}t,\ \mathrm{X}_0 = 0.$$

  From Itô's formula, $\mathrm{X}_t = \mathrm{B}_t^3$ is a solution.

- *Weak uniqueness.* Consider the equation

$$\mathrm{dX}_t = f(\mathrm{X}_t)\mathrm{dB}_t$$

  where $f : \mathbb{R} \to \mathbb{R}$ is any measurable function with $|f(x)| = 1$ for any $x \in \mathbb{R}$. Then any solution X of the equation is a $\mathcal{F}$-martingale with bracket $\langle \mathrm{X}\rangle_t = t$, hence it is a Brownian motion. This proves uniqueness in law.

- *Pathwise uniqueness.* The following equation is the one describing the so-called geometric Brownian motion :

$$\mathrm{dX}_t = \mathrm{X}_t\mathrm{dB}_t,\ \mathrm{X}_0 = 1.$$

  Then $\mathrm{B} = \beta$, $\mathrm{X}_t = e^{\mathrm{B}_t - \frac{t}{2}}$ is a solution. Given $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathrm{P})$ and B, if Y_t is another solution, from the Itô formula

$$\mathrm{d}\left(\frac{\mathrm{Y}_t}{\mathrm{X}_t}\right) = \frac{\mathrm{dY}_t}{\mathrm{X}_t} - \frac{\mathrm{Y}_t\mathrm{dX}_t}{\mathrm{X}_t^2} + \frac{\mathrm{Y}_t}{\mathrm{X}_t^3}\mathrm{d}\langle \mathrm{X}\rangle_t - \frac{1}{\mathrm{X}_t^2}\mathrm{d}\langle \mathrm{X}, \mathrm{Y}\rangle_t = 0.$$

  Consequently X and Y cannot be distinguished.

- *Weak existence with no weak uniqueness.* Consider the same equation as previously,

$$\mathrm{dX}_t = 3\,\mathrm{sgn}(\mathrm{X}_t)|\mathrm{X}_t|^{2/3}\mathrm{dB}_t + 3\,\mathrm{sgn}(\mathrm{X}_t)|\mathrm{X}_t|^{1/3}\mathrm{d}t,\ \mathrm{X}_0 = 0.$$

  From Itô's formula, $\mathrm{B} = \beta$ and $\mathrm{X}_t = \mathrm{B}_t^3$ are solutions, as well as $\mathrm{X} = 0$. This example shows as well that there can be various strong solutions.

- *Weak existence and weak uniqueness with no pathwise uniqueness.* From the original Brownian motion β, define another Brownian motion (Lévy's criterion) by

$$\mathrm{B}_t = \int_0^t \mathrm{sgn}(\beta_s)\mathrm{d}\beta_s,$$

  where $\mathrm{sgn}(x) = 1$ if $x \geqslant 0$, -1 otherwise. Then β and $-\beta$ are solutions of

$$\mathrm{dX}_s = \mathrm{sgn}(\mathrm{X}_s)\mathrm{dB}_s,\ \mathrm{X}_0 = 0,$$

  because $\int_0^t \mathbb{1}_{\beta_s=0}\mathrm{d}s = 0$, hence $\int_0^t \mathbb{1}_{\beta_s=0}\mathrm{dB}_s = 0$. Moreover, any solution is a Brownian motion, once again by Lévy's criterion.

Section 5 will in particular give an example of weak existence, weak uniqueness, and non existence of a strong solution. Note that one cannot have pathwise uniqueness and no weak uniqueness, or pathwise uniqueness and solutions not measurable in B's canonical filtration, this is a famous result of Yamada and Watanabe.

Theorem 5.3. *If for the equation* $\mathrm{E}(\sigma, b)$ *pathwise uniqueness holds, then there is also weak uniqueness, and any solution is strong.*

Proof. For notational convenience, we consider the case $d = m = 1$. Proving the theorem for $\mathrm{E}_x(\sigma, b)$ is sufficient, by conditioning on the initial value. Note that some care is needed for this conditioning , as the event has no positive measure. However, as $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^2)$ is a Polish space, there is a regular conditional distribution $\mathrm{P}(\omega, \cdot)$ of P, the law of (X, B), with respect to $\mathcal{B}_0$. See [16] for more details about this conditioning.

Let (B, X) and $(\mathrm{B}', \mathrm{X}')$ be two solutions to $\mathrm{E}_x(\sigma, b)$, eventually defined on distinct probability spaces. We note $\mathrm{W} = \mathscr{C}(\mathbb{R}^+, \mathbb{R})$, with the compact uniform topology, $\mathcal{B}(\mathrm{W})$ the topological Borel field on W, $\mathcal{B}_t(\mathrm{W}) = \sigma(\omega_s, s \leqslant t)$. The same objects are easily defined on $\mathrm{W} \times \mathrm{W}, \mathrm{W} \times \mathrm{W} \times \mathrm{W}$. Let $\mathrm{P}(\mathrm{d}\omega, \mathrm{d}\omega')$ (resp. P') be the law of (B, X) (resp. $(\mathrm{B}', \mathrm{X}')$) on $(\mathrm{W} \times \mathrm{W}, \mathcal{B}(\mathrm{W} \times \mathrm{W}))$. Finally, we introduce $\mathrm{P}_\omega(\mathrm{d}\omega')$ the regular conditional distribution of $\mathrm{P}(\mathrm{d}\omega\mathrm{d}\omega')$ given ω, i.e.

(i) for any $\omega \in \mathrm{W}$, this is a probability measure on $(\mathrm{W}, \mathcal{B}(\mathrm{W}))$,

(ii) for any $\mathrm{A} \in \mathcal{B}(\mathrm{W})$, $\mathrm{P}_\omega(\mathrm{A})$ is $\mathcal{B}(\mathrm{W})$-measurable as a function of ω,

(iii) for any $(\mathrm{A}, \mathrm{A}') \in \mathcal{B}(\mathrm{W} \times \mathrm{W})$,

$$\mathrm{P}(\mathrm{A} \times \mathrm{A}') = \int_{\mathrm{A}'} \mathrm{P}_\omega(\mathrm{A})\mathrm{R}(\mathrm{d}\omega)$$

where R is the law of B on $(\mathrm{W}, \mathcal{B}(\mathrm{W}))$, i.e. the Wiener measure.

For the existence of this regular conditional distribution, we refer to [9]. In the same way we define $\mathrm{P}'_\omega(\mathrm{d}\omega)$.

To make use of pathwise uniqueness in the proof, we need to consider both solutions (B, X) and $(\mathrm{B}', \mathrm{X}')$ in the same probability space. For this purpose, let Q be the probability distribution defined on $(\mathrm{W} \times \mathrm{W} \times \mathrm{W}, \mathcal{B}(\mathrm{W} \times \mathrm{W} \times \mathrm{W}))$ by

$$\mathrm{Q}(\mathrm{d}\omega_1\mathrm{d}\omega_2\mathrm{d}\omega) = \mathrm{P}_\omega(\mathrm{d}\omega_1)\mathrm{P}'_\omega(\mathrm{d}\omega_2)\mathrm{R}(\mathrm{d}\omega). \tag{5.1}$$

Note that the projection of Q on along the second coordinate is P, and P′ along the first coordinate. We want to prove that ω is a $(\mathrm{Q}, \mathcal{B}(\mathrm{W}\times\mathrm{W}\times\mathrm{W})_t)$ Brownian motion. Let $\mathrm{F}_1, \mathrm{F}_2, \mathrm{F}$ be three $\mathcal{B}_t(\mathrm{W})$-measurable functions, and $u > t$. Then

$$\mathbb{E}_{\mathrm{Q}}\left((\omega(u)-\omega(t))\mathrm{F}_1(\omega_1)\mathrm{F}_2(\omega_2)\mathrm{F}(\omega)\right)$$
$$= \int_{\mathrm{W}} (\omega(u)-\omega(t)) \left(\int_{\mathrm{W}} \mathrm{P}_\omega(\mathrm{d}\omega_1)\mathrm{F}_1(\omega_1)\right)\left(\int_{\mathrm{W}} \mathrm{P}_\omega(\mathrm{d}\omega_2)\mathrm{F}_2(\omega_2)\right)\mathrm{F}(\omega)\mathrm{R}(\mathrm{d}\omega).$$

By the following important Lemma 5.4, $\int_{\mathrm{W}} \mathrm{P}_\omega(\mathrm{d}\omega_1)\mathrm{F}_1(\omega_1)$ and $\int_{\mathrm{W}} \mathrm{P}_\omega(\mathrm{d}\omega_2)\mathrm{F}_2(\omega_2)$ are $\mathcal{B}_\sqcup(\mathrm{W})$-measurable in ω, so by conditioning on $\mathcal{B}_\sqcup(\mathrm{W})$ and using that ω is a $(\mathrm{R}, (\mathcal{B}_t))$-martingale (a Brownian motion actually), the above term vanishes. In the same way,

$$\mathbb{E}_{\mathrm{Q}}\left(\left((\omega(u)-\omega(t))^2-(u-t)^2\right)\mathrm{F}_1(\omega_1)\mathrm{F}_2(\omega_2)\mathrm{F}(\omega)\right) = 0,$$

so ω is a $(\mathrm{Q}, (\mathcal{B}_t(\mathrm{W}\times\mathrm{W}\times\mathrm{W})))$-martingale with bracket $\langle\omega\rangle_s = s$: it is a $(\mathrm{Q}, (\mathcal{B}_t(\mathrm{W}\times\mathrm{W}\times\mathrm{W})))$-Brownian motion. To sum up, we have proven that on the probability space

$$(\mathrm{W}\times\mathrm{W}\times\mathrm{W}, \mathcal{B}(\mathrm{W}\times\mathrm{W}\times\mathrm{W}), (\mathcal{B}_t(\mathrm{W}\times\mathrm{W}\times\mathrm{W})), \mathrm{Q})$$

the processes $(\mathrm{B},\mathrm{X}) := (\omega,\omega_1)$ and $(\mathrm{B}',\mathrm{X}) := (\omega,\omega_2)$ are solutions to $\mathrm{E}_x(\sigma,b)$. By pathwise uniqueness, $\omega_1 = \omega_2$ Q-almost surely. Looking back at (5.1), this implies that $\mathrm{P} = \mathrm{P}'$ and there is a $\mathcal{B}(\mathrm{W})$-measurable function F such that $\omega_1 = \mathrm{F}(\omega)$, $\omega_2 = \mathrm{F}(\omega)$. By Lemma 5.4, F is adapted, hence any solution is strong. □

Lemma 5.4. *Let* $\mathrm{P}(\mathrm{d}\omega\mathrm{d}\omega_1)$ *be the law of* (B,X)*, a solution of* $\mathrm{E}_x(\sigma,b)$*.*

For $\mathrm{A} \in \mathcal{B}_t(\mathrm{W})$*, the applications* $\omega \to \mathrm{P}_\omega(\mathrm{A})$ *and* $\omega \to \mathrm{P}'_\omega(\mathrm{A})$ *are* $\mathcal{B}_t(\mathrm{W})$*-measurable.*

Proof. If $\sigma_1, \sigma_2, \sigma$ are three σ-algebras such that $\sigma_1 \vee \sigma_2$ is independent of σ under a probability measure μ, it is a general fact that for $\mathrm{A} \in \sigma_1$, μ-almost surely

$$\mu(\mathrm{A} \mid \sigma_2) = \mu(\mathrm{A} \mid \sigma_2 \vee \sigma).$$

We apply this to $\sigma_1(t) = \sigma(\omega_1(s), s \leqslant t)$, $\sigma_2(t) = \sigma(\omega(s), s \leqslant t)$, $\sigma(t) = \sigma(\omega(u) - \omega(t), u \geqslant t)$ and the measure $\mathrm{P}(\mathrm{d}\omega\mathrm{d}\omega_1)$. As, for some filtration $(\mathcal{F}_t)$, ω is a $(\mathcal{F}_t)$-Brownian motion, and X is $(\mathcal{F}_t)$-adapted, we have $\sigma_1(t) \vee \sigma_2(t)$ independent of $\sigma(t)$. Hence the above result reads, P-almost surely

$$\mathrm{P}(\cdot,\mathrm{A}) = \mathrm{E}_{\mathrm{P}}(\mathbb{1}_{\mathrm{A}} \mid \sigma_2(\infty)) = \mathrm{P}(\cdot,\mathrm{A}) = \mathrm{E}_{\mathrm{P}}(\mathbb{1}_{\mathrm{A}} \mid \sigma_2(t) \vee \sigma(t)) = \mathrm{E}_{\mathrm{P}}(\mathbb{1}_{\mathrm{A}} \mid \sigma_2(t)),$$

which is obviously $\mathcal{B}_t(\mathrm{W})$-measurable. □

2. The Lipschitz case

Like for ordinary differential equations, we will show existence and strong uniqueness for $\mathrm{E}(\sigma,b)$ under the Lipschitz hypothesis

$$|\sigma(t,\omega)-\sigma(t,\omega')| + |b(t,\omega)-b(t,\omega')| \leqslant c \sup_{s\leqslant t}|\omega(s)-\omega'(s)|.$$

Theorem 5.5. *Under the above hypothesis, there is pathwise uniqueness for* $\mathrm{E}(\sigma,b)$*. Moreover, for any filtered probability space* $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geqslant 0},\mathrm{P})$*, any* $\mathcal{F}$*-Brownian motion* B*, and any* $x \in \mathbb{R}^d$ *there is a strong solution to* $\mathrm{E}_x(\sigma,b)$*.*

Remark. There is only one strong solution in the above theorem, by pathwise uniqueness. Moreover, the above result also proves weak existence, and by the Yamada-Watanabe theorem weak uniqueness.

Proof. We assume $m = d = 1$ to simplify notations, the general case being absolutely similar. We first prove pathwise uniqueness : given $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathrm{P})$, a $\mathcal{F}$-Brownian motion B, let X and X′ be two solutions of the related stochastic differential equation such that $\mathrm{X}_0 = \mathrm{X}'_0$ almost surely. To avoid integrability problems, consider $\mathrm{L} > 0$ and the stopping time

$$\tau = \inf\{t \geqslant 0 \mid |\mathrm{X}_t| \geqslant \mathrm{L} \text{ or } |\mathrm{X}'_t| \geqslant \mathrm{M}\}.$$

Then, for any $t \geqslant 0$,

$$\mathrm{X}_{t\wedge\tau} = \mathrm{X}_0 + \int_0^{t\wedge\tau} \sigma(s, \mathrm{X})\mathrm{dB}_s + \int_0^{t\wedge\tau} b(s, \mathrm{X})\mathrm{d}s,$$

and a similar equation holds for X′. Consequently, assuming that t lies in a bounded set $[0, \mathrm{T}]$ and using the elementary identity $(a + b)^2 \leqslant 2(a^2 + b^2)$, $f(t) = \mathbb{E}\left(\sup_{s\leqslant t}(\mathrm{X}_{s\wedge\tau} - \mathrm{X}'_{s\wedge\tau})^2\right)$ is bounded by

$$2\,\mathbb{E}\left(\sup_{s\leqslant t}\left(\int_0^{s\wedge\tau}(\sigma(u, \mathrm{X}) - \sigma(u, \mathrm{X}'))\mathrm{dB}_u\right)^2\right) + 2\,\mathbb{E}\left(\sup_{s\leqslant t}\left(\int_0^{s\wedge\tau}(b(u, \mathrm{X}) - b(u, \mathrm{X}'))\mathrm{d}u\right)^2\right).$$

By the Doob (resp. Cauchy-Schwarz) inequality for the σ (resp. b) term, this is lower than

$$8\,\mathbb{E}\left(\left(\int_0^{t\wedge\tau}(\sigma(s, \mathrm{X}) - \sigma(s, \mathrm{X}'))\mathrm{dB}_s\right)^2\right) + 2\mathrm{T}\,\mathbb{E}\left(\int_0^{t\wedge\tau}(b(s, \mathrm{X}) - b(s, \mathrm{X}'))^2\mathrm{d}s\right)$$
$$= 8\,\mathbb{E}\left(\int_0^{t\wedge\tau}(\sigma(s, \mathrm{X}) - \sigma(s, \mathrm{X}'))^2\mathrm{d}s\right) + 2\mathrm{T}\,\mathbb{E}\left(\int_0^{t\wedge\tau}(b(s, \mathrm{X}) - b(s, \mathrm{X}'))^2\mathrm{d}s\right).$$

We used that the expectation of the square of a martingale is also the expectation of its bracket (the stochastic integral is L^2-bounded thanks to the stopping time τ and the boundedness of σ by the Lispschitz hypothesis). Now, we can use the Lipschitz hypothesis on σ and b to obtain finally

$$f(t) \leqslant 2c^2(4 + \mathrm{T})\int_0^t f(s)\mathrm{d}s.$$

As $\mathrm{X}_0 = \mathrm{X}'_0$ almost surely, $f(0) = 0$ a.s. and Gronwall's lemma[1] applied to the above inequality yields $f(t) = 0$ a.s. for any $t \in [0, \mathrm{T}]$, hence for any $t > 0$. This means that almost surely, for any $t \in [0, \mathrm{T}]$, $\mathrm{X}_{t\wedge\tau} = \mathrm{X}'_{t\wedge\tau}$. By making $\mathrm{L} \to \infty$, the processes X and X′ cannot be distinguished.

Now, given a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathrm{P})$, a $\mathcal{F}$-Brownian motion B, and any initial value x, one can construct a solution to $\mathrm{E}_x(\sigma, b)$ which is adapted to the canonical filtration $(\mathcal{B}_t)_{t\geqslant 0}$ of B using Picard's approximation scheme :

$$\begin{cases} \mathrm{X}^{(0)}_t = x, \\ \mathrm{X}^{(n)}_t = \int_0^t \sigma(s, \mathrm{X}^{(n-1)})\mathrm{dB}_s + \int_0^t b(s, \mathrm{X}^{(n-1)})\mathrm{d}s. \end{cases}$$

These processes are properly defined as an immediate induction shows that for any $n \geqslant 0$, $\mathrm{X}^{(n)}$ is $\mathcal{B}$-adapted, and still by induction for any $\mathrm{T} > 0$ and any $n \geqslant 0$ there is a constant $c^{(n)}$ such that

$$\mathbb{E}\left(\sup_{t\leqslant \mathrm{T}}\left(\mathrm{X}^{(n)}_t\right)^2\right) \leqslant c^{(n)}. \tag{5.2}$$

1. This lemma states in particular that if, on $\mathbb{R}^+$, $f'(t) \leqslant af(t)$, then $f(t) \leqslant f(0)e^{at}$. Proof : $f(t)e^{-at}$ decreases, by differentiation and application of the hypothesis.

This is obvious at rank $n=0$. Assuming the property at rank $n-1$, the rank n case follows by using the Lipschitz conditions (in particular $|\sigma(t,\omega)| \leqslant c' + c\sup_{[0,\mathrm{T}]}\omega$, $|b(t,\omega)| \leqslant c'+c\sup_{[0,\mathrm{T}]}\omega$ for any $t\in[0,\mathrm{T}]$) and the identity $(a+b+c)^2 \leqslant 3(a^2+b^2+c^2)$, by convexity of $x\to x^2$:

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{t\leqslant \mathrm{T}}\left(\mathrm{X}_t^{(n)}\right)^2\right)\\
\leqslant& 3|x|^2 + 3\,\mathbb{E}\left(\sup_{t\leqslant \mathrm{T}}\left(\int_0^t \sigma(s,\mathrm{X}^{(n-1)})\mathrm{dB}_s\right)^2\right) + 3\,\mathbb{E}\left(\sup_{t\leqslant \mathrm{T}}\left(\int_0^t b(s,\mathrm{X}^{(n-1)})\mathrm{d}s\right)^2\right)\\
\leqslant& 3|x|^2 + 12\,\mathbb{E}\left(\left(\int_0^{\mathrm{T}} \sigma(s,\mathrm{X}^{(n-1)})\mathrm{dB}_s\right)^2\right) + 3\mathrm{T}\,\mathbb{E}\left(\int_0^t b(s,\mathrm{X}^{(n-1)})^2\mathrm{d}s\right)\\
=& 3|x|^2 + 12\,\mathbb{E}\left(\int_0^{\mathrm{T}} \sigma(s,\mathrm{X}^{(n-1)})^2\mathrm{d}s\right) + 3\mathrm{T}\,\mathbb{E}\left(\int_0^t b(s,\mathrm{X}^{(n-1)})^2\mathrm{d}s\right)\\
\leqslant& 3|x|^2 + 12\int_0^{\mathrm{T}}\left(c' + c\,\mathbb{E}(\sup_{u\leqslant s}\mathrm{X}^{(n-1)})\right)^2\mathrm{d}s + 3\mathrm{T}\int_0^{\mathrm{T}}\left(c' + c\,\mathbb{E}(\sup_{u\leqslant s}\mathrm{X}^{(n-1)})\right)^2\mathrm{d}s\\
\leqslant& 3(|x|^2 + 2\mathrm{T}(4+\mathrm{T})(c'^2 + c^2c^{(n-1)}))
\end{aligned}
$$

where we used in the first inequality the Doob and the Cauchy Schwarz inequality as previously in the proof. This achieves the inductive proof of (5.2) by taking the above constant for $c^{(n)}$.

The a-priori local martingale $\int_0^t \sigma(s,\mathrm{X}^{(n)})\mathrm{dB}_s$ is actually a L^2-bounded martingale from the above estimate (5.2). This allows the following upper bound for the infinite norm between $\mathrm{X}^{(n+1)}$ and $\mathrm{X}^{(n)}$ on $[0,\mathrm{T}]$, along the same steps as previously :

$$
\begin{aligned}
\mathbb{E}\left(\sup_{[0,t]}(\mathrm{X}^{(n+1)}-\mathrm{X}^{(n)})^2\right) \leqslant& 2\,\mathbb{E}\left(\sup_{s\leqslant t}\left(\int_0^s\left(\sigma(u,\mathrm{X}^{(n+1)})-\sigma(u,\mathrm{X}^{(n)})\right)\mathrm{dB}_u\right)^2\right)\\
&+2\,\mathbb{E}\left(\sup_{s\leqslant t}\left(\int_0^s\left(b(u,\mathrm{X}^{(n+1)})-b(u,\mathrm{X}^{(n)})\right)\mathrm{d}u\right)^2\right)\\
\leqslant& 8\,\mathbb{E}\left(\left(\int_0^t\left(\sigma(s,\mathrm{X}^{(n+1)}-\sigma(s,\mathrm{X}^{(n)}))\mathrm{dB}_u\right)^2\right)\right)\\
&+2\mathrm{T}\,\mathbb{E}\left(\int_0^t\left(b(s,\mathrm{X}^{(n+1)})-b(s,\mathrm{X}^{(n)})\right)^2\mathrm{d}u\right)\\
\mathbb{E}\left(\sup_{[0,t]}(\mathrm{X}^{(n+1)}-\mathrm{X}^{(n)})^2\right) \leqslant& 2(4+\mathrm{T})c^2\,\mathbb{E}\left(\int_0^t\sup_{0\leqslant v\leqslant u}(\mathrm{X}_u^{(n+1)}-\mathrm{X}_u^{(n)})^2\mathrm{d}u\right).
\end{aligned}
$$

This can be written, for $t\in[0,\mathrm{T}]$, $f^{(n+1)}(t)\leqslant c_{\mathrm{T}}\int_0^t f^{(n)}(u)\mathrm{d}u$, with $f^{(n+1)}(t) = \mathbb{E}\left(\sup_{[0,t]}(\mathrm{X}^{(n+1)}-\mathrm{X}^{(n)})^2\right)$, $c_{\mathrm{T}} = 2(4+\mathrm{T})c^2$. An immediate induction therefore yields

$$
f^{(n)}(t)\leqslant\left(\sup_{[0,\mathrm{T}]}f^{(0)}(s)\right)(c_{\mathrm{T}})^n\frac{t^n}{n!}.
$$

In particular, $\sum_{n=0}^{\infty}\sqrt{f^{(n)}(\mathrm{T})}<\infty$, i.e.

$$
\sum_n\|\sup_{[0,\mathrm{T}]}(\mathrm{X}^{(n+1)}-\mathrm{X}^{(n)})\|_{\mathrm{L}^2}<\infty,
$$

so $\sum_n\sup_{[0,\mathrm{T}]}(\mathrm{X}^{(n+1)}-\mathrm{X}^{(n)})$ is bounded in L^2, hence a.s. finite : almost surely $\mathrm{X}^{(n)}$ converges uniformly on $[0,\mathrm{T}]$ to a continuous process X. As $\mathrm{X}^{(n)}$ is adapted to the canonical filtration of B, X is adapted as well.

We now need to check that X is a solution of $\mathrm{E}_x(\sigma, b)$: as $\|\sup_{[0,\mathrm{T}]}(\mathrm{X}^{(n)} - \mathrm{X})\|_{\mathrm{L}^2} \leqslant \sum_{m\geqslant n}^{\infty} \sqrt{f^{(m)}(\mathrm{T})} \to 0$, the similar inequalities as previously imply that, in L^2

$$\int_0^t \sigma(s, \mathrm{X}^{(n)})\mathrm{dB}_s \to \int_0^t \sigma(s, \mathrm{X})\mathrm{dB}_s,$$
$$\int_0^t b(s, \mathrm{X}^{(n)})\mathrm{dB}_s \to \int_0^t b(s, \mathrm{X})\mathrm{d}s.$$

Injecting this in the induction relation between $\mathrm{X}^{(n)}$ and $\mathrm{X}^{(n-1)}$, we get that

$$\mathrm{X}_t = \int_0^t \sigma(s, \mathrm{X})\mathrm{dB}_s + \int_0^t b(s, \mathrm{X})\mathrm{d}s,$$

first in L^2 but almost surelyas well[2]. □

3. The Strong Markov property and more on the Dirichlet problem

4. The Stroock-Varadhan piecewise approximation

This section is devoted to approximations of the law of stochastic processes. How can one make a reasonable simulation of the trajectory of the strong solution of a stochastic differential equation ? More generally, the Stroock-Varadhan piecewise approximation [18] provides the convergence in law of a discrete Markov chain to the solution of a martingale problem.

The reason why we consider their approach is that it ensures convergence in law of processes under very weak conditions, provided that the associated martingale problem admits a unique solution.

4.1. Martingale problems

The Stroock-Varadhan approach allows to define Markov processes through their generator with no explicit construction of the semigroup. In the case of diffusions, two other classical representations of diffusions are the following.

- Kolmogorov and Feller expressed the transition probabilities as the solution of a forward partial differential equation.
- Itô introduced its stochastic calculus and therefore represents the diffusions as solutions of stochastic differential equations.

Consider σ a $d \times d$ matrix function, $a = \sigma\, {}^{\mathrm{t}}\sigma$ and $b : \mathbb{R}^d \to \mathbb{R}^d$, all of them being measurable.

Definition 5.6. *A process* $(\mathrm{X}_t, t \geqslant 0)$ *with values in* $\mathbb{R}^d$*, together with a filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F})_t, \mathbb{P})$*, is said to solve the martingale problem* $\mathrm{M}(a, b)$ *if for any* $1 \leqslant i, j \leqslant d$

$$\begin{cases} \mathrm{Y}^i := (\mathrm{X}_t^i - \int_0^t b_i(\mathrm{X}_s)\mathrm{d}s, t \geqslant 0), \\ (\mathrm{Y}_t^i \mathrm{Y}_t^j - \int_0^t a_{ij}(\mathrm{X}_s)\mathrm{d}s, t \geqslant 0), \end{cases}$$

are local martingales. If a solution to $\mathrm{M}(a, b)$ *has a unique possible law, the problem is said to be well posed.*

2. If Y_n converges to Y_1 in L^2 and Y_2 a.s., then $\mathrm{Y}_1 = \mathrm{Y}_2$ almost surely : Y_n converges to Y_2 in probability, hence almost surely along a subsequence ; by uniqueness o the limit, $\mathrm{Y}_1 = \mathrm{Y}_2$.

We will study the martingale approach for the following reason : Theorem 5.9 hereafter states the weak convergence of discrete time Markov processes towards solutions of sde under the uniqueness condition of the the solution of the limiting martingale problem. This is a more general condition than pathwise uniqueness for example. Very general conditions for well-posedness of $\mathrm{M}(a, b)$ are given in [18]. For example, in dimension 1, measurability, boundedness for a and b and uniform positivity for a are sufficient.

Moreover, a martingale problem is well-posed if and only if there is uniqueness in law for an associated stochastic differential equation. More precisely, if σ and b are Lipschitz, there exists a unique strong solution (i.e. a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_t, \mathbb{P})$, $\mathcal{F}_t = \sigma(\mathrm{B}_s, s \leqslant t)$ where $\mathrm{B} = (\mathrm{B}_1, \dots, \mathrm{B}_d)$ is a Brownian motion and an $\mathcal{F}$-adapted X) to the stochastic differential equation $\mathrm{E}(a, b)$:

$$\mathrm{dX}_t = \sigma(\mathrm{X}_t)\mathrm{dB}_t + b(\mathrm{X}_t)\mathrm{d}t.$$

Then X together with $(\Omega, \mathcal{F}, (\mathcal{F})_t, \mathbb{P})$ solves the martingale problem $\mathrm{M}(a, b)$, $a = \sigma\, {}^{\mathrm{t}}\sigma$. An interesting point is that a reciprocal is true, with weaker assumptions on σ, b. For technical reasons, we assume that b is continuous and $a = \sigma\, {}^{\mathrm{t}}\sigma$ is elliptic, but these hypotheses can be seriously weakened, as remarked later.

Theorem 5.7. *Assume the above hypotheses and that* X, $(\Omega, \mathcal{F}, (\mathcal{F})_t, \mathbb{P})$ *is a solution to* $\mathrm{M}(a, b)$. *Then there exists a* $\mathcal{F}$*-Brownian motion* B *in* $\mathbb{R}^d$*such that* X, B *solves* $\mathrm{E}(\sigma, b)$.

Proof. In the following, all (stochastic) integrals make sense thanks to our (strong) assumptions on a and b. As X is a solution of $\mathrm{M}(a, b)$,

$$\mathrm{Y}^i := (\mathrm{X}_t^i - \int_0^t b^i(\mathrm{X}_s)\mathrm{d}s, t \geqslant 0)$$

is a local martingale, and $\mathrm{d}\langle \mathrm{Y}^i, \mathrm{Y}^j \rangle_t = a^{ij}(t)\mathrm{d}t$. Let

$$\mathrm{B}^i := \left(\int_0^t \sum_{k=1}^d (\sigma^{-1})_{ik}(\mathrm{X}_s)\mathrm{dY}_s^k, t \geqslant 0 \right). \tag{5.3}$$

This is a local martingale and the joint brackets are

$$\langle \mathrm{B}^i, \mathrm{B}^j \rangle_t = \sum_{k,l=1}^d \int_0^t (\sigma^{-1})_{ik}(\mathrm{X}_s)(\sigma^{-1})_{jl}(\mathrm{X}_s)(\sigma\, {}^{\mathrm{t}}\sigma)_{kl}(\mathrm{X}_s)\mathrm{d}s.$$

But the definition of the inverse yields

$$\sum_{k,l=1}^d (\sigma^{-1})_{ik}(\sigma^{-1})_{jl}(\sigma\, {}^{\mathrm{t}}\sigma)_{kl} = \sum_{k,l,m=1}^d (\sigma^{-1})_{ik}(\sigma^{-1})_{jl}(\sigma)_{k,m}\, {}^{\mathrm{t}}\sigma_{ml} = \mathbb{1}_{i=j}.$$

By Lévy's criterion, the B^i's are Brownian motions, and from (5.3)

$$\mathrm{dX}_t = \sigma(\mathrm{X}_t)\mathrm{dB}_t + b(\mathrm{X}_t)\mathrm{d}t,$$

which concludes the proof. □

Remark. This result of the theorem actually holds under weaker hypotheses, in particular without the invertibility/ellipticity condition, by letting B go independently for any time s such that $\sigma(\mathrm{X}_s)$ is not invertible.

4.2. Convergence of Markov chains

Consider a Markov chain $(\mathrm{U}_n^{(\varepsilon)})_{n\geqslant 0}$, depending on a parameter $\varepsilon > 0$, with a transition kernel $\mathrm{P}^{(\varepsilon)}$:

$$\mathbb{P}(\mathrm{U}_{n+1}^{(\varepsilon)} \in \mathrm{B} \mid \mathrm{U}_n^{(\varepsilon)} = x) = \mathrm{P}^{(\varepsilon)}(x, \mathrm{B}).$$

In the context of the strong solution a stochastic differential equation in $\mathbb{R}^d$ (e.g. with Lipschitz coefficients),

$$\mathrm{dX}_t = \sigma(\mathrm{X}_t)\mathrm{dB}_t + b(\mathrm{X}_t)\mathrm{d}t,$$

one can think about $\mathrm{U}_{n+1}^{(\varepsilon)}$ as

$$\begin{cases} \mathrm{U}_0^{(\varepsilon)} &= \mathrm{U}_0, \\ \mathrm{U}_{n+1}^{(\varepsilon)} &= \mathrm{U}_n^{(\varepsilon)} + \sigma(\mathrm{U}_n^{(\varepsilon)})\sqrt{\varepsilon}\mathcal{N}_n + b(\mathrm{U}_n^{(\varepsilon)})\varepsilon, \end{cases}$$

where $\mathcal{N}_1, \mathcal{N}_2, \dots$ are independent standard Gaussian vectors in $\mathbb{R}^d$.

Assume now that a_{ij}, b_i, are continuous coefficients in $\mathbb{R}^d$ such that the martingale problem $\mathrm{M}(a, b)$ has a unique solution in distribution : for any $x \in \mathbb{R}^d$ there is a process X, with unique possible distribution, such that

$$\mathrm{Y}_t^i := \mathrm{X}_t^i - \int_0^t b_i(\mathrm{X}_s)\mathrm{d}s, \ \mathrm{Y}_t^i\mathrm{Y}_t^j - \int_0^t a_{ij}(\mathrm{X}_s)\mathrm{d}s$$

are local martingales. Note that, in the above context of Lipschitz coefficients, writing $a = \sigma\, {}^{\mathrm{t}}\sigma$, the martingale problem has a unique solution $\sigma(a, b)$, which is the strong solution of the differential equation.

Assume that the scaled means and variances of the jumps of $\mathrm{U}^{(\varepsilon)}$ are uniformly convergent to a and b on any compact K, as $\varepsilon \to 0$:

$$b_i^{(\varepsilon)}(x) := \frac{1}{\varepsilon}\int_{|y-x|\leqslant 1} (y_i - x_i)\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y), \qquad \lim_{\varepsilon\to 0}\sup_{\mathrm{K}} |b_i^{(\varepsilon)} - b_i| = 0, \tag{5.4}$$

$$a_{ij}^{(\varepsilon)}(x) := \frac{1}{\varepsilon}\int_{|y-x|\leqslant 1} (y_i - x_i)(y_j - x_j)\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y), \quad \lim_{\varepsilon\to 0}\sup_{\mathrm{K}} |a_{ij}^{(\varepsilon)} - a_{ij}| = 0. \tag{5.5}$$

Moreover, assume that for any $\delta > 0$ the probability of a jump greater than δ is $\mathrm{o}(\varepsilon)$, uniformly in any compact K :

$$\lim_{\varepsilon\to 0}\sup_{\mathrm{K}} \frac{1}{\varepsilon}\mathrm{P}^{(\varepsilon)}(x, \mathrm{B}(x,\delta)^c) = 0 \tag{5.6}$$

Lemma 5.8. *Conditions (5.4), (5.5) and (5.6) together are equivalent to*

$$\frac{1}{\varepsilon}\int (f(y) - f(x))\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y) \underset{\varepsilon\to 0}{\longrightarrow} \mathrm{L}f(x) \tag{5.7}$$

uniformly on compact sets, for any $f \in \mathscr{C}_c^\infty$, where

$$\mathrm{L} = \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\partial_{ij} + \sum_{i=1}^d b_i(x)\partial_i.$$

Proof. Assume first that the conditions (5.4), (5.5) and (5.6) hold, and note $\mathrm{L}^{(\varepsilon)}$ the analogue of L for the coefficients $a_{ij}^{(\varepsilon)}$, $b_i^{(\varepsilon)}$. It is clear, thanks to (5.4), (5.5), that $\mathrm{L}^{(\varepsilon)}f \to \mathrm{L}f$ uniformly on compacts, so we just need to prove that

$$\left|\frac{1}{\varepsilon}\int (f(y) - f(x))\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y) - \mathrm{L}^{(\varepsilon)}f(x)\right| \underset{\varepsilon\to 0}{\longrightarrow} 0 \tag{5.8}$$

uniformly on compacts. By Taylor's theorem, for an absolute constant c_f,

$$\left| f(y) - f(x) - \sum_{i=1}^{d}(y_i - x_i)\partial_i f(x) - \frac{1}{2}\sum_{i,j=1}^{d}(y_i - x_i)(y_j - x_j)\partial_{ij} f(x)\right| \leqslant c_f |y - x|^3,$$

so the left hand side of (5.8)is bounded by

$$c_f \int_{|y-x|\leqslant 1} |y-x|^3 \frac{\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y)}{\varepsilon} + \int_{|y-x|>1} |f(y) - f(x)| \frac{\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y)}{\varepsilon}.$$

The second term is bounded by $2\|f\|_\infty \frac{1}{\varepsilon}\mathrm{P}^{(\varepsilon)}(x, \mathrm{B}(x,1)^c)$, hence converges to 0, using (5.6). Concerning the first one, for any $0 < \delta < 1$ it is smaller than

$$c_f \left(\delta \int_{|x-y|<\delta} \frac{\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y)}{\varepsilon} |x-y|^2 + \frac{\mathrm{P}^{(\varepsilon)}(x, \mathrm{B}(x,\delta)^c)}{\varepsilon}\right).$$

This last term converges to 0 by (5.6) and the first one is uniformly bounded on a compact set K by

$$2\delta \sum_{i=1}^{d} \|a_{ii}\|_{\mathrm{L}^\infty(\mathrm{K})},$$

for sufficiently small ε, from 5.5. Hence it converges uniformly to 0 on K.

Conversely, assume that (5.7) holds. We first prove (5.6), which is equivalent to : for any $\delta > 0$, $x_\varepsilon \to x$ as $\varepsilon \to 0$,

$$\frac{1}{\varepsilon}\mathrm{P}^{(\varepsilon)}(x_\varepsilon, \mathrm{B}(x_\varepsilon, \delta)^c) \to 0. \tag{5.9}$$

Chose $\varphi \in \mathscr{C}_c^\infty$ with $\varphi \equiv 1$ on $\mathrm{B}(x, \delta/4)$, $\varphi \equiv 0$ out of $\mathrm{B}(x, \delta/2)$, $0 \leqslant \varphi \leqslant 1$. For ε sufficiently small, $\varphi \leqslant \mathbb{1}_{\mathrm{B}(x_\varepsilon,\delta)}$ and $\varphi(x_\varepsilon) = 1$, so

$$-\frac{1}{\varepsilon}\int (\varphi(y) - \varphi(x_\varepsilon))\mathrm{P}^{(\varepsilon)}(x_\varepsilon, \mathrm{d}y) = \frac{1}{\varepsilon}\int (1 - \varphi(y))\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y) \geqslant \frac{1}{\varepsilon}\mathrm{P}_\varepsilon(x_\varepsilon, \mathrm{B}(x_\varepsilon, \delta)^c).$$

Now, by (5.7) the left hand side converges to 0, as $\mathrm{L} \equiv 0$ on $\mathrm{B}(x_0, \delta/4)$. This proves (5.9), hence (5.6). Now, proving (5.4) and (5.5) is easy : we only need to prove that for any compact K, $f \in \mathscr{C}^\infty$ with support included in K,

$$\frac{1}{\varepsilon}\sup_{\tilde{\mathrm{K}}} \left| \int_{|x-y|<1} (f(y) - f(x))\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y) - \mathrm{L}f(x)\right| \underset{\varepsilon\to 0}{\longrightarrow} 0,$$

because specializing f to x_i (resp. $x_i x_j$) in K and 0 out of $\tilde{\mathrm{K}} = \{x \in \mathbb{R}^d \mid \mathrm{dist}(x, \mathrm{K}) \leqslant 1\}$ yields (5.4) (resp. (5.5)). As we have the uniform convergence hypothesis (5.7), the above result follows from

$$\sup_{\tilde{\mathrm{K}}} \frac{1}{\varepsilon}\left| \int_{|x-y|>1} (f(x) - f(y))\mathrm{P}^{(\varepsilon)}(x, \mathrm{d}y)\right| \underset{\varepsilon\to 0}{\longrightarrow} 0,$$

which is a straightforward consequence of (5.6) that we have proved. □

In the following Theorem and till the end of this section, we will talk about the weak convergence of cadlag processes (emerging from Markov chains) to continuous solutions of martingale problems or stochastic differential solutions. This is justified by using the Skorokhod topology, see [7]. The reader uncomfortable with this topology can replace the following discontinuous processes by the continuous affine extension of the Markov chains.

Theorem 5.9. *Let* $X_t^{(\varepsilon)} = U_{\lfloor t/\varepsilon\rfloor}^{(\varepsilon)}$. *Assume that* $x^{(\varepsilon)} := X_0^{(\varepsilon)} \to x_0$ *as* $\varepsilon \to 0$, *and that conditions (5.4), (5.5) and (5.6) are satisfied. Then, if* $M(a,b)$ *is well posed,* $X^{(\varepsilon)}$ *converges weakly to* X, *unique solution in distribution of the martingale problem* $M(a,b)$.

Proof. We first prove the result under the additional hypotheses

$$\lim_{\varepsilon\to 0}\sup_{\mathbb{R}^d}|b_i^{(\varepsilon)} - b_i| = 0, \tag{5.10}$$

$$\lim_{\varepsilon\to 0}\sup_{\mathbb{R}^d}|a_{ij}^{(\varepsilon)} - a_{ij}| = 0, \tag{5.11}$$

$$\lim_{\varepsilon\to 0}\sup_{\mathbb{R}^d}\frac{1}{\varepsilon}P^{(\varepsilon)}(x, B(x,\delta)^c) = 0. \tag{5.12}$$

For a given $T > 0$, let $P_{x_\varepsilon}^{(\varepsilon)}$ be the law of $(X_t^{(\varepsilon)}, 0 \leqslant t \leqslant T)$. The proof relies on two points. First the tightness[3] of these processes holds, i.e. the set $\{P_{x_\varepsilon}^{(\varepsilon)}, \varepsilon > 0\}$ is pre-compact. Here we refer to Theorem 1.4.11 from [18], which asserts that the two following criteria are sufficient to deduce tightness :

- for any smooth compactly supported f there is a constant c_f such that for any $\varepsilon > 0$
$$((f(X^{(\varepsilon)}(k\varepsilon)) + c_f k\varepsilon)_k, (\mathcal{F}_{k\varepsilon})_k, P_{x_\varepsilon}^{(\varepsilon)})$$
is a discrete submartingale. This is true as, from our assumptions (5.10), (5.11) and (5.12), there is an absolute c_f such that $\sup_{\varepsilon>0}\frac{1}{\varepsilon}\int(f(y)-f(x))P^{(\varepsilon)}(x,\mathrm{d}y) < c_f < \infty$;
- the probability of a large gap is bounded by our assumption (5.12) : for any $\delta > 0$,
$$\sum_{0\leqslant j\varepsilon\leqslant T} P_{x_\varepsilon}^{(\varepsilon)}(|X^{(\varepsilon)}((k+1)\varepsilon) - X^{(\varepsilon)}(k\varepsilon)| > \delta) \leqslant \lceil T/\varepsilon\rceil \sup_{\mathbb{R}^d} P^{(\varepsilon)}(x, B(x,\delta)^c),$$
which converges to 0 as $\varepsilon \to 0$.

Secondly if, for $\varepsilon_n \to 0$, $P_{x_{\varepsilon_n}}^{(\varepsilon_n)}$ converges weakly to a measure P, then P solves the martingale problem $M(a,b)$ starting from x_0. Indeed, to prove it, note first that necessarily $P(X_0 = x_0) = 1$. Moreover, let $s < t$, $f \in \mathscr{C}_c^\infty(\mathbb{R}^d)$ and $F : \Omega \to \mathbb{R}$ bounded, continuous and $\mathcal{F}_s$-measurable. Then, writing $k_n = \lfloor s/\varepsilon_n\rfloor + 1$, $\ell_n = \lfloor t/\varepsilon_n\rfloor + 1$,

$$\mathbb{E}_{P_{x_{\varepsilon_n}}^{(\varepsilon_n)}}\left(\left(f(X_{\ell_n\varepsilon_n}) - f(X_{k_n\varepsilon_n}) - \sum_{j=k_n}^{\ell_n-1}\int(f(y) - f(X_{j\varepsilon_n}))P^{(\varepsilon_n)}(X_{j\varepsilon_n}, \mathrm{d}y)\right)F\right) = 0,$$

because

$$\left(f(X_{k\varepsilon_n}) - \sum_{j=0}^{k-1}\int(f(y) - f(X_{j\varepsilon_n}))\,P^{(\varepsilon_n)}(X_{j\varepsilon_n}, \mathrm{d}y), k \geqslant 0\right)$$

is obviously a $(\mathcal{F}_{k\varepsilon_n})$-martingale. Moreover, using Lemma (5.8),

$$f(X(\ell_n\varepsilon_n)) - f(X_{k_n\varepsilon_n}) - \sum_{j=k_n}^{\ell_n-1}\int(f(y) - f(X_{j\varepsilon_n}))P^{(\varepsilon_n)}(X_{j\varepsilon_n}, \mathrm{d}y)$$
$$\underset{n\to\infty}{\longrightarrow} f(X_t) - f(X_s) - \int_s^t Lf(X_u)\mathrm{d}u.$$

3. A sequence of measures $(\mu_n, n \geqslant 0)$ is called tight if for any $\varepsilon > 0$ there is a compact set K such that, for any n, $\mu_n(K) \geqslant 1 - \varepsilon$.

uniformly on compact sets. As a consequence,

$$\mathbb{E}_{\mathrm{P}}\left(\left(f(\mathrm{X}_t) - f(\mathrm{X}_s) - \int_s^t \mathrm{L}f(\mathrm{X}_u)\mathrm{d}u\right)\mathrm{F}\right) = 0,$$

hence $(f(\mathrm{X}_t) - \int_0^t \mathrm{L}f(\mathrm{X}_u)\mathrm{d}u, (\mathcal{F}_t), \mathrm{P})$ is a martingale. Specializing f and localization prove therefore that X together with the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathrm{P})$ solves the martingale problem $\mathrm{M}(a, b)$.

To remove the extra assumptions (5.10), (5.11) and (5.12), we proceed by localization : for $k \geqslant 1$, let φ_k be smooth such that $\varphi_k \equiv 1$ on $\mathrm{B}(0, k)$ $\varphi_k \equiv 0$ out of $\mathrm{B}(0, k+1)$, $0 \leqslant \varphi_k \leqslant 1$,

$$\mathrm{P}_k^{(\varepsilon_n)}(x, \Gamma) := \varphi_k(x)\mathrm{P}^{(\varepsilon_n)}(x, \Gamma) + (1 - \varphi_k(x))\mathbb{1}_{x \in \Gamma}.$$

Then he coefficients $a_k^{(\varepsilon_n)}$ (resp. $b_k^{(\varepsilon_n)}$) associated to $\mathrm{P}_k^{(\varepsilon_n)}$ are uniformly bounded in $\mathbb{R}^d$ and n, and converge to $\varphi_k a$ (resp $\varphi_k b$) on compacts. Hence, from the previous discussion, the sequence of associated measures $(\mathrm{P}_k^n)_{n \geqslant 0}$ is pre-compact and all its limit points solve the martingale problem for $\varphi_k a$, $\varphi_k b$. This implies that any limit point of $(\mathrm{P}_k^n)_{n \geqslant 0}$ equals $\mathrm{P}_{\mathrm{X}_0}$on $\mathcal{F}_{\tau_k}$ where $\tau_k = \inf\{t \geqslant 0 \mid |\mathrm{X}_t| \geqslant k\}$ (see Theorem 10.1.1 in [18]). As $\mathrm{P}^{n,k} = \mathrm{P}^n$ on $\mathcal{F}_{\tau_k}$, any limit of P^n coincides with $\mathrm{P}_{\mathrm{X}_0}$ on $\mathcal{F}_{\tau_k}$. Taking $k \to \infty$ is allowed and yields the result (see Lemma 11.1.1 in [18]). $\square$

Amongst the many applications of Theorem 5.9, the following is linked to an important problem in the History of statistical physics. After Gibbs and Boltzmann, the following paradox emerged, being the source of some disputes with Poincaré and Zermelo. For a system of particles in a closed container, in the (common) ergodic case we expect that any state is recurrent, in particular the initial one. However, Boltzman's work and its H-entropy theorem, based on Newtonian dynamics, proves non-reversibility, through the monotonicity of its entropy. Smoluchowski formulated the following explanation : the states far away from equilibrium have a considerable recurrence time, so the system appears to be irreversible only for reasonable observations. Paul and Tatiana Ehrenfest formalized this idea by the following simple model, known as Ehrenfest's urn : it is exactly solvable and they proved that it has a very large recurrence time for states far from equilibrium.

Consider $2m$ particles, in 2 urns communicating through a small hole. At time n, chose uniformly one particle amongst $2m$ and change its urn : it is a model for the gaz diffusion between two urns through a small hole. Let $\mathrm{N}_n^{(m)}$ be the number of particles on the left urn at time n. Consider the process $\mathrm{X}^{(m)}$ defined as

$$\left(\frac{\mathrm{N}_{\lfloor mt \rfloor}^{(m)} - m}{\sqrt{m}}, t \geqslant 0\right).$$

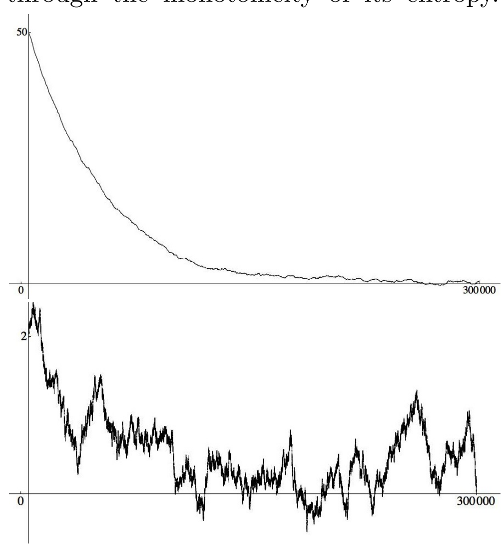

Figure 5.1. Samples of $\mathrm{X}^{(m)}$ for $m = 50000$, $0 \leqslant t \leqslant 6$ (i.e. 300000 exchanges in the urn) : for $\mathrm{X}_0^{(m)} = 50$ (upper graph), the dynamics are first irreversible, then recurrent, like in the lower graph, beginning at $\mathrm{X}_0^{(m)} = 2$, discrete approximation of an Ornstein-Uhlenbeck process.

The following corollary of Theorem 5.9 goes in the direction of Hilbert's wish in

1900 : *Boltzmann's work on the principles of mechanics suggest the problem of developing mathematically the limiting processes (...) which lead from the atomistic view to the laws of motion of continua.*

Corollary 5.10. *Assume that, as $m \to \infty$, $\mathrm{N}_0^{(m)} - m \sim \alpha\sqrt{m}$. Then, as $m \to \infty$, for any finite horizon $\mathrm{T} > 0$, the process $(\mathrm{X}_t^{(m)}, 0 \leqslant t \leqslant \mathrm{T})$ converges weakly to the unique (strong) solution of the Ornstein-Uhlenbeck stochastic differential equation*

$$\begin{cases} \mathrm{dX}_t &= -\mathrm{X}_t \mathrm{d}t + \mathrm{dB}_t, \\ \mathrm{X}_0 &= \alpha. \end{cases}$$

Note first that the above continuous limit visits all states, hanks to small fluctuations from the mean in the initial condition : $\mathrm{N}_0^{(m)} - m \sim \alpha\sqrt{m}$. If $\mathrm{N}_0^{(m)} - m \sim \alpha m$ for some $\alpha \in (0,1)$, then there is no diffusive limit, the way to equilibrium being deterministic and non-reversible, the random term is not visible compared to the drift one. On a different topic, the above stochastic differential equation clearly has a unique strong solution because its coefficients are Lipschitz, and using the Itô formula the reader can check that the explicit solution is

$$\mathrm{X}_t = e^{-t}\left(\alpha + \int_0^t e^s \mathrm{dB}_s\right).$$

Proof. The set of possible values for $\mathrm{X}_t^{(m)}$ is $\{\frac{k}{\sqrt{m}}, -m \leqslant k \leqslant m\}$, and the probability transitions are

$$\mathbb{P}^{(m)}\left(x, x + \frac{1}{\sqrt{m}}\right) = \frac{1}{2} - \frac{x}{2\sqrt{m}},\ \mathbb{P}^{(m)}\left(x, x - \frac{1}{\sqrt{m}}\right) = \frac{1}{2} + \frac{x}{2\sqrt{m}}.$$

As a consequence, with notations analogous to (5.4) and (5.5), a direct calculation yields

$$b^{(m)}(x) = -x,\ a^{(m)}(x) = 1.$$

As condition (5.6) is clearly satisfied and $\mathrm{X}_0^{(m)} \underset{m\to\infty}{\longrightarrow} \alpha$, the result follows by the general Theorem 5.9. □

5. Shifts in the Cameron-Martin space

The purpose of this section is to consider stochastic differential equations corresponding to shift of the Bownian motion along a predictable element of the Cameron-Martin space :

$$\mathrm{dX}_t = \mathrm{dB}_t + b(\mathrm{X}_s, 0 \leqslant s \leqslant t)\mathrm{d}t. \tag{5.13}$$

By the Girsanov Theorem, we know some properties of any measure weak solution to this equation : it is absolutely continuous with respect to the Wiener measure, with an explicit density.

A more subtle question concerns the existence/uniqueness of a weak/strong solution to this equation. We will show that in the Markovian case, under very weak hypotheses on b, there is a unique strong solution (Zvonkin, [26]). It was conjectured that the Markovian hypothesis on b could be relaxed, which was shown to be false in a famous counterexample due to Tsirelson ([22], see also [25]). Finally, a necessary and sufficient condition on b for existence/uniqueness of a strong solution to (5.13) is given, from the work of Ustunel [24].

5.1. The Zvonkin theorem

Consider a general Markovian differential equation

$$\mathrm{dX}_t = \sigma(\mathrm{X}_t)\mathrm{dB}_t + b(\mathrm{X}_t)\mathrm{d}t. \tag{5.14}$$

A natural idea to study the existence/uniqueness of such an equation consists in changing the variable, $Y_t = f(X_t)$, with $f \in \mathscr{C}^2$: Itô's formula yields

$$\mathrm{d}Y_t = (\sigma f')(X_t)\mathrm{d}B_t + \left(\frac{1}{2}\sigma^2 f'' + bs'\right)(X_t)\mathrm{d}t,$$

so Y is a local martingale if and only if

$$\frac{1}{2}\sigma^2 f'' + bf' \equiv 0. \tag{5.15}$$

Assume that $b\sigma^{-2}$ is locally integrable, then

$$f'(x) = e^{-\int_\cdot^x 2\frac{b(u)}{\sigma(u)^2}\mathrm{d}u}$$

is a solution to (5.15)[4]. The function f is strictly increasing, so f^{-1} is well defined, and Y solves the stochastic differential equation

$$\mathrm{d}Y_t = s(Y_t)\mathrm{d}B_t \tag{5.16}$$

with $s(x) = (\sigma f') \circ f^{-1}$. The above stochastic differential equation (and therefore general markovian shifts in the Cameron-Martin space) has a unique weak solution, as shown by the following lemma and delocalization (make $\varepsilon \to 0$ hereafter).

Lemma 5.11. *Let s be a measurable function with $s(x) > \varepsilon > 0$. Then (5.16) admits a unique weak solution, which is a local martingale.*

Proof. Let Y be a solution of (5.16). Then its bracket converges to ∞, as

$$\langle Y\rangle_t = \int_0^t g(Y_s)^2\mathrm{d}s > \varepsilon t.$$

Hence, by the Dubins-Schwarz theorem, β defined by

$$\beta_{\langle Y\rangle_t} = Y_t$$

is a Brownian motion. More explicitly,

$$\beta_t = Y_{\tau(t)}$$

where

$$\int_0^{\tau(t)} s(Y_u)^2\mathrm{d}u = t.$$

Differentiating the above equation yields

$$1 = \tau'(t)s(Y_{\tau(t)})^2 = \tau'(t)s(\beta_t)^2.$$

This yields an explicit expression of $\langle Y\rangle_t$ in terms of β :

$$\langle Y\rangle_t = \inf\{u \mid \tau(u) > t\} = \inf\{u \mid \int_0^u \tau'(u)\mathrm{d}u > t\} = \inf\{u \mid \int_0^u \frac{\mathrm{d}u}{s(\beta_u)^2} > t\}.$$

Consequently, Y is a continuous function of the Brownian motion β, hence there is a unique possible law for it. □

4. n.b. : such an f is not necessarily $\mathscr{C}^2$, but this is overcome by IV.45.9 in [17]

Much better than the above uniqueness in law, strong existence and uniqueness hold under weak assumptions : Zvonkin observed that when σ is Lipschitz, bounded away from 0, and b only bounded and measurable, the function s in (5.16) is locally Lipschitz, hence there is a unique strong solution to (5.16). As f is invertible, there is a unique strong solution to (5.14), and in particular to (5.13) where $\sigma \equiv 1$ and a bounded b depends only on the current point (Markovian case).

By scaling, the above discussion proves that there is always a strong solution to

$$\mathrm{dY}_t = \varepsilon \mathrm{dB}_t + b(\mathrm{Y}_t)\mathrm{d}t,$$

no matter how small $\varepsilon > 0$ is. This contrasts with the ordinary differential equation

$$\mathrm{dY}_t = b(\mathrm{Y}_t)\mathrm{d}t$$

which can have several solutions[5] for a bounded measurable b.

5.2. The Tsirelson example

The preceding paragraph contrasts with the following example where b depends on all the past. Note the fractional part

$$\{x\} = x - \lfloor x \rfloor$$

and consider the ounded predictable drift

$$b((\mathrm{X}_s)_{0\leqslant s\leqslant t}) = \begin{cases} \left\{\frac{\mathrm{X}_{t_k}-\mathrm{X}_{t_{k-1}}}{t_k - t_{k-1}}\right\} & \text{if} \quad t_k < t \leqslant t_{k+1} \\ 0 & \text{if} \qquad t = 0 \end{cases},$$

where $(t_k)_{k\in\mathbb{Z}}$ is an increasing sequence converging to 0 as $k \to -\infty$. Then b is a bounded measurable function, so from the Girsanov theorem for the stochastic differential equation

$$\mathrm{dX}_t = \mathrm{dB}_t + b((\mathrm{X}_s)_{0\leqslant s\leqslant t})\mathrm{d}t \tag{5.17}$$

uniqueness in law holds (here $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ is a filtered probability space, B is a $((\mathcal{F}_t), \mathbb{P})$-Brownian motion and X is $(\mathcal{F}_t)$-adapted). However, there is no strong solution : for any solution X to the above equation, surprisingly for any $t > 0$ the fractional part $\{(\Delta\mathrm{X})_t\}$ of the slope $(t_k < t \leqslant t_{k+1})$

$$(\Delta\mathrm{X})_t := \frac{\mathrm{X}_t - \mathrm{X}_{t_k}}{t - t_k}$$

is uniform on $[0,1]$ and independent of $\mathcal{B}_s = \sigma(\mathrm{B}_u, 0 \leqslant u \leqslant s)$ for any $s > 0$.

Theorem 5.12. *For any $t > 0$, $\{(\Delta\mathrm{X})_t\}$ is uniform on $[0,1]$ and independent of $\mathcal{B}_\infty$. Moreover, noting $\mathcal{X}_t = \sigma(\mathrm{X}_s, s \leqslant t)$, for any $0 < s < t$*

$$\mathcal{X}_t = \mathcal{B}_t \vee \sigma(\{(\Delta\mathrm{X})_s\}). \tag{5.18}$$

Proof. Similarly to $(\Delta\mathrm{X})_t$, let

$$(\Delta\mathrm{B})_t = \frac{\mathrm{B}_t - \mathrm{B}_{t_k}}{t - t_k},$$

for $t_k < t \leqslant t_{k+1}$. Then, from the stochastic differential equation (5.17), for $t_{k-1} < t \leqslant t_k$

$$(\Delta\mathrm{X})_t = (\Delta\mathrm{B})_t + \{(\Delta\mathrm{X})_{t_{k-1}}\}.$$

5. Example : $\mathrm{dY}_t = 2\,\mathrm{sgn}(\mathrm{Y}_t)|\mathrm{Y}_t|^{1/2}\mathrm{d}t$ has three solutions $0, \pm t^2$, due to the bifurcation choice at $t = 0$. Such a phenomena cannot exist in the stochastic case because the Brownian motion prevents from making any initial choice.

This yields, for any $p \in \mathbb{Z}$

$$\begin{aligned}\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_{t_k}\}}\right) &= \mathbb{E}\left(e^{\mathrm{i}2\pi p((\Delta\mathrm{B})_{t_k}+(\Delta\mathrm{X})_{t_{k-1}})}\right)\\ &= \mathbb{E}\left(e^{\mathrm{i}2\pi p(\Delta\mathrm{B})_{t_k}}\right)\mathbb{E}\left(e^{\mathrm{i}2\pi p(\Delta\mathrm{X})_{t_{k-1}}}\right) = e^{-\frac{2\pi^2p^2}{t_k-t_{k-1}}}\,\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_{t_{k-1}}\}}\right),\end{aligned}$$

where we have used the independence between $(\Delta\mathrm{B})_{t_k}$ and $\mathcal{F}_{t_{k-1}}$: B is a B is a $(\mathcal{F}_t)$-Brownian motion, hence $\mathrm{B}_t - \mathrm{B}_s$ is independent of $\mathcal{F}_s$ for $t > s$. An iteration of the above equality gives for any $n \geqslant 1$ the upper-bound

$$\left|\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_{t_k}\}}\right)\right| \leqslant e^{-2\pi^2p^2c_kn}\left|\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_{t_{k-n}}\}}\right)\right| \leqslant e^{-2\pi^2p^2c_kn},$$

where $c_k = \inf_{\ell\leqslant k}(t_\ell - t_{\ell-1})^{-1} > 0$. This proves that for any $p \in \mathbb{Z}^*$ and $k \in \mathbb{Z}$

$$\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_{t_k}\}}\right) = 0,$$

i.e. $\{(\Delta\mathrm{X})_{t_k}\}$ is uniform on $[0,1]$. Moreover, for any $t_k < t < t_{k+1}$, $(\Delta\mathrm{X})_t = (\Delta\mathrm{B})_t + \{(\Delta\mathrm{X})_{t_k}\}$, where the last term was just proven to be uniform on $[0,1]$, and is independent of $(\Delta\mathrm{B})_t$, hence $\{(\Delta\mathrm{X})_t\}$ is uniform on $[0,1]$ for any $t > 0$.

Now note $\mathcal{B}_v^u = \sigma(\mathrm{B}_s, u \leqslant s \leqslant v)$, $0 < u < v$. For any $v > 0$, $p \in \mathbb{Z}$,

$$\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_t\}} \mid \mathcal{B}_v\right) = \lim_{n\to-\infty}\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_t\}} \mid \mathcal{B}_v^{t_n}\right),$$

and this last term vanishes : for sufficiently large n, $t_n < t_k < t \leqslant t_{k+1}$ and supposing $t < v$,

$$\begin{aligned}\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_t\}} \mid \mathcal{B}_v^{t_n}\right) &= \mathbb{E}\left(e^{\mathrm{i}2\pi p((\Delta\mathrm{X})_{t_n}+(\Delta\mathrm{B})_{t_{n+1}}+\cdots+(\Delta\mathrm{B})_{t_k}+(\Delta\mathrm{B})_t)} \mid \mathcal{B}_v^{t_n}\right)\\ &= e^{\mathrm{i}2\pi p((\Delta\mathrm{B})_{t_{n+1}}+\cdots+(\Delta\mathrm{B})_{t_k}+\cdots+(\Delta\mathrm{B})_t)}\,\mathbb{E}\left(e^{\mathrm{i}2\pi p(\Delta\mathrm{X})_{t_n}} \mid \mathcal{B}_v^{t_n}\right)\\ &= e^{\mathrm{i}2\pi p((\Delta\mathrm{B})_{t_{n+1}}+\cdots+(\Delta\mathrm{B})_{t_k}+\cdots+(\Delta\mathrm{B})_t)}\,\mathbb{E}\left(e^{\mathrm{i}2\pi p(\Delta\mathrm{X})_{t_n}}\right)\end{aligned}$$

where the last equality relies on the independence of $\mathcal{B}_v^{t_n}$ and $\mathcal{F}_{t_n}$. We proved that $\mathbb{E}\left(e^{\mathrm{i}2\pi p(\Delta\mathrm{X})_{t_n}}\right) = 0$ for $p \in \mathbb{Z}^*$, hence if $0 < t < v$

$$\mathbb{E}\left(e^{\mathrm{i}2\pi p\{(\Delta\mathrm{X})_t\}} \mid \mathcal{B}_v\right) = 0,$$

which concludes the proof that $\{(\Delta\mathrm{X})_t\}$ is independent of $\mathcal{B}_v$ whenever $v > t$, hence independent of any $\mathcal{B}_v$, $v > 0$.

Finally, to show that $\{(\Delta\mathrm{X})_s\}$ is exactly the missing information to get X from B, i.e. (5.18), we proceed by double inclusion.

First, from the stochastic differential equation (5.17), B is a functional of X so $\mathcal{B}_t \subset \mathcal{X}_t$. Obviously $\sigma(\{(\Delta\mathrm{X})_s\}) \subset \mathcal{X}_t$ $(s < t)$, so

$$\sigma(\{(\Delta\mathrm{X})_s\}) \vee \mathcal{B}_t \subset \mathcal{X}_t.$$

Concerning the other inclusion, we make the observation that $(b(\mathrm{X}_u), 0 \leqslant u \leqslant t)$ is a function of $(\mathrm{B}_u)_{0\leqslant u\leqslant t}$ and $\{(\Delta\mathrm{X})_s\}$ for any choice of $0 < s < t$. Indeed, if $t_k < s \leqslant t_{k+1}$,

$$(\Delta\mathrm{X})_s = (\Delta\mathrm{B})_s + \{(\Delta\mathrm{X})_{t_k}\},$$

so $\{(\Delta\mathrm{X})_{t_k}\} \in \sigma(\{(\Delta\mathrm{X})_s\}) \vee \mathcal{B}_s$ (n.b. : $\{a - b\} = \{\{a\} - \{b\}\}$). Moreover,

$$(\Delta\mathrm{X})_{t_k} = (\Delta\mathrm{B})_{t_k} + \{(\Delta\mathrm{X})_{t_{k-1}}\},$$

so $\{(\Delta X)_{t_{k-1}}\}\sigma(\{(\Delta X)_{t_k}\})\vee\mathcal{B}_{t_k} \subset \sigma(\{(\Delta X)_s\})\vee\mathcal{B}_s$, and by induction any $\{(\Delta X)_{t_n}\}$, $n \leqslant k$, is $\sigma(\{(\Delta X)_s\})\vee\mathcal{B}_s$-measurable. Proving the measurability of the drifts for $t > s$ is also straightforward :

$$\{(\Delta X)_{t_{k+1}}\} = \{\{(\Delta X)_{t_k}\} + (\Delta B)_{t_{k+1}}\} \in \sigma(\{(\Delta X)_{t_k}\}) \vee \mathcal{B}_{t_{k+1}},$$

so by an immediate induction for any n with $t_n < t$, $\{(\Delta X)_{t_n} \in \sigma(\{(\Delta X)_s\}) \vee \mathcal{B}_t$. By going below and up s, we proved that, for any $u < t$, $b((X_v)_{0\leqslant v\leqslant u})$ is always $\sigma(\{(\Delta X)_s\}) \vee \mathcal{B}_t$-measurable, hence for $u < t$

$$X_u = \int_0^u b((X_y)_{0\leqslant y\leqslant v})\mathrm{d}v + B_u \in \sigma(\{(\Delta X)_s\}) \vee \mathcal{B}_t$$

which concludes the proof. □

5.3. Energy, entropy.

A simple characterization of the existence of a strong solution to

$$\mathrm{d}X_t = \mathrm{d}B_t + \dot{u}_t\mathrm{d}t, \tag{5.19}$$

where $\dot{u}_t \in \mathcal{X}_t = \sigma(X_s, s \leqslant t)$, was given in [24]. We abbreviate the above equation as $B = (\mathrm{Id} - u)X$ and wonder about P-almost sure invertibility of $\mathrm{Id} - u$. To avoid integrability problems, we assume that $\dot{u}_t$ is uniformly bounded in ω and t. Note that by the Girsanov theorem the law R of X is uniquely determined and absolutely continuous with respect to that of B. Hence the P-almost sure and R-almost sure invertibility conditions are the same, so for notational convenience we will consider the equation

$$X = (\mathrm{Id} - u)B$$

instead (for the same function u, where $\dot{u}_t \in \mathcal{B}_t = \sigma(B_s, s \leqslant t)$ now), and wonder about its $\mathbb{P}$-almost sure invertibility. From the above equation, it follows that

$$\mathcal{X}_t \subset \mathcal{B}_t \subset \mathcal{F}_t\,.$$

As previously, $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ is a filtered probability space, B is a $((\mathcal{F}_t), \mathbb{P})$-Brownian motion and X is $(\mathcal{F}_t)$-adapted. Moreover, we consider the context of a finite horizon $(0, 1)$. Let

$$L_t = \left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = \mathbb{E}_{\mathrm{P}}(L_1 \mid \mathcal{F}_t),\ \mathbb{Q} = (\mathrm{Id} - u)\mathbb{P},$$

and we will abbreviate L for L_1. By the Itô representation theorem, $L_t = \rho(-\delta y^t)$ for some $\mathcal{F}$-adapted y, $\int_0^t \dot{y}_s^2\mathrm{d}s < \infty$ a.s., with the notation

$$\rho(-\delta y^t) = e^{-\int_0^t \dot{y}_s\mathrm{d}B_s - \frac{1}{2}\int_0^t \dot{y}_s^2\mathrm{d}s}.$$

Indeed, the Itô representation yields $L_t = 1 + \int_0^t \alpha_s\mathrm{d}B_s$ for some adapted α, and as $L_s > 0$ a.s., this can be written

$$\mathrm{d}L_t = \left(\frac{\alpha_s}{L_s}\right)L_s\mathrm{d}B_s,$$

so L is an exponential martingale.

Theorem 5.13. *The application* $\mathrm{Id} - u$ *is* P*-almost surely invertible, with inverse* $\mathrm{Id} + y$*, if and only if*

$$\mathbb{E}_{\mathrm{P}}(L\log L) = \frac{1}{2}\,\mathbb{E}_{\mathrm{P}}\left(\|u\|_{\mathrm{H}}^2\right).$$

Remark. The above equation means that the relative entropy $\mathrm{H}(\mathrm{Q} \mid \mathrm{P})$ coincides with the kinetic energy of u. The following proof yields to the unconditional inequality

$$\mathrm{H}(\mathrm{Q} \mid \mathrm{P}) \leqslant \frac{1}{2}\, \mathbb{E}_{\mathrm{P}}\left(\|u\|_{\mathrm{H}}^2\right).$$

As entropy measures the number of accessible states, the theorem means that the application is invertible if and only if it has just enough energy to fulfill the accessible states.

Proof. Assume first that $\mathbb{E}_{\mathrm{P}}(\mathrm{L}\log\mathrm{L}) = \frac{1}{2}\,\mathbb{E}_{\mathrm{P}}\left(\|u\|_{\mathrm{H}}^2\right)$. By substitutions,

$$\begin{aligned}\mathbb{E}_{\mathrm{P}}(\mathrm{L}\log\mathrm{L}) = \mathbb{E}_{\mathrm{Q}}(\log\mathrm{L}) &= \mathbb{E}_{\mathrm{Q}}(-\int_0^1 \dot{y}_s \mathrm{dB}_s - \frac{1}{2}\int_0^1 \dot{y}_s^2 \mathrm{d}s) \\ &= \mathbb{E}_{\mathrm{Q}}(-\int_0^1 \dot{y}_s(\dot{y}_s \mathrm{d}s + \mathrm{dB}_s) + \frac{1}{2}\,\mathbb{E}_{\mathrm{Q}}(\int_0^t \dot{y}_s^2 \mathrm{d}s).\end{aligned}$$

From the Girsanov theorem, $(\mathrm{B}_s + \int_0^s \dot{y}_s \mathrm{d}s)_{s\geqslant 0}$ is a $(\mathcal{F}, \mathrm{Q})$-Brownian motion. As $\dot{y}_s$ is $\mathcal{F}_s$-adapted, this implies that $\mathbb{E}_{\mathrm{Q}}(-\int_0^1 \dot{y}_s(\dot{y}_s\mathrm{d}s + \mathrm{dB}_s) = 0$, so

$$\begin{aligned}\mathbb{E}_{\mathrm{P}}(\mathrm{L}\log\mathrm{L}) = \frac{1}{2}\,\mathbb{E}_{\mathrm{Q}}(\|y\|_{\mathrm{H}}^2) &= \frac{1}{2}\int \|y\|_{\mathrm{H}}^2 \frac{\mathrm{d}(\mathrm{Id}-u)\mathrm{P}}{\mathrm{dP}}\mathrm{dP} \\ &= \frac{1}{2}\int \|y(\omega)\|_{\mathrm{H}}^2 \mathrm{P}((\mathrm{Id}-u)^{-1}(\mathrm{d}\omega)) = \frac{1}{2}\,\mathbb{E}_{\mathrm{P}}(|y\circ(\mathrm{Id}-u)|_{\mathrm{H}}^2).\end{aligned}$$

Moreover, as X represents LdP, from Lemma 5.14, $\dot{y}_s \circ (\mathrm{Id}-u) = \mathbb{E}_{\mathrm{P}}(\dot{u}_s \mid \mathcal{X}_s)\ \mathrm{d}s \times \mathrm{dP}$-a.s., so

$$\mathbb{E}_{\mathrm{P}}(\mathrm{L}\log\mathrm{L}) = \frac{1}{2}\,\mathbb{E}_{\mathrm{P}}\left(\int_0^1 (\mathbb{E}_{\mathrm{P}}(\dot{u}_s \mid \mathcal{X}_s))^2 \mathrm{d}s\right),$$

and using the hypothesis

$$\mathbb{E}_{\mathrm{P}}\left(\int_0^1 \dot{u}_s^2 \mathrm{d}s\right) = \mathbb{E}_{\mathrm{P}}\left(\int_0^1 (\mathbb{E}_{\mathrm{P}}(\dot{u}_s \mid \mathcal{X}_s))^2 \mathrm{d}s\right),$$

which easily implies

$$\mathbb{E}_{\mathrm{P}}\left(\int_0^1 (\dot{u}_s - \mathbb{E}_{\mathrm{P}}(\dot{u}_s \mid \mathcal{X}_s))^2 \mathrm{d}s\right) = 0,$$

i.e. $\dot{u}_s = \mathbb{E}_{\mathrm{P}}(\dot{u}_s \mid \mathcal{X}_s\ \mathrm{d}t \times \mathrm{dP}$ almost surely. Hence

$$\dot{y}_s \circ (\mathrm{Id}-u) = \dot{u}_s$$

$\mathrm{d}t \times \mathrm{dP}$ a.s. which means that $\mathrm{Id}+y$ is the almost sure inverse of $\mathrm{Id}-u$. □

Lemma 5.14. *Keeping the previous notations, in particular*

$$\mathrm{L}_t = \frac{\mathrm{d}(\mathrm{Id}-u)\mathrm{P}}{\mathrm{dP}}\ |_{\mathcal{F}_t} = \rho(-\delta y^t) = e^{-\int_0^t \dot{y}_s \mathrm{dB}_s - \frac{1}{2}\int_0^t \dot{y}_s^2 \mathrm{d}s},$$

then $\dot{y}_s \circ (\mathrm{Id}-u) = \mathbb{E}_{\mathrm{P}}(\dot{u}_s \mid \mathcal{X}_s)\ \mathrm{d}s \times \mathrm{dP}$*-almost surely.*

Proof. An easy substitution yields

$$\mathrm{L}_t \circ (\mathrm{Id}-u) = e^{-\int_0^t (\dot{y}_s \circ (\mathrm{Id}-u))(\mathrm{dB}_s + \dot{u}_s \mathrm{d}s + \frac{1}{2}\dot{y}_s \circ (\mathrm{Id}-u)\mathrm{d}s)}. \tag{5.20}$$

On the other hand, applications of the Girsanov theorem yield the following expressions.

- First,
$$(\mathrm{L}_t \circ (\mathrm{Id} - u))\,\mathbb{E}(\rho(-\delta u^t) \mid \mathcal{X}_t) = 1. \tag{5.21}$$
Indeed, for any $\mathcal{F}_t$-measurable f, from the Girsanov theorem,
$$\mathbb{E}\left(f \circ (\mathrm{Id} - u)\mathrm{L}_t \circ (\mathrm{Id} - u)\rho(-\delta u^t)\right) = \mathbb{E}(f\mathrm{L}_t) = \mathbb{E}(f \circ (\mathrm{Id} - u)),$$
where the last equality just translates $\mathrm{L}_t = \frac{\mathrm{d}(\mathrm{Id}-u)\mathrm{P}}{\mathrm{dP}} \mid_{\mathcal{F}_t}$. Hence
$$\mathbb{E}\left(f \circ (\mathrm{Id} - u)\left(f \circ (\mathrm{Id} - u)\mathrm{L}_t \circ (\mathrm{Id} - u)\,\mathbb{E}(\rho(-\delta u^t) \mid \mathcal{X}_t) - 1\right)\right) = 0.$$
We have $\mathrm{B} := \mathbb{E}(\rho(-\delta u^t) \mid \mathcal{X}_t) \in \mathcal{X}_t$ and the above equation means $\mathbb{E}(\mathrm{AB}) = 0$ for any $\mathrm{A} \in \mathcal{X}_t$, so $\mathrm{B} = 0$ P-a.s., proving (5.21).

- Moreover, defining Z by $\mathrm{X}_t = \mathrm{Z}_t + \int_0^t \mathbb{E}(\dot{u}_s \mid \mathcal{X}_s)\mathrm{d}s$,
$$\mathbb{E}(\rho(-\delta u) \mid \mathcal{X}_t) = e^{-\int_0^t \mathbb{E}(\dot{u}_s|\mathcal{X}_s)\mathrm{dZ}_s - \frac{1}{2}\int_0^t \mathbb{E}(\dot{u}_s|\mathcal{X}_s)^2\mathrm{d}s}(=: \ell_t) \tag{5.22}$$
, by a double application of the Girsanov theorem :
$$\mathbb{E}(f \circ (\mathrm{Id} - u)\rho(-\delta_u)) = \mathbb{E}(f) = \mathbb{E}(f \circ (\mathrm{Id} - u)\ell_1),$$
the last equality relying on the fact that Z is a $\mathcal{X}$-Brownian motion (it is a $\mathcal{X}$-martingale with bracket $\langle \mathrm{Z}\rangle_t = t$). In the above equality, conditioning on $\mathcal{X}_t$ implies the expected result (5.22).

Gathering (5.20), (5.21) and (5.22) yields

$$\begin{aligned}1 &= (\mathrm{L}_t \circ (\mathrm{Id} - u))\,\mathbb{E}(\rho(-\delta u^t) \mid \mathcal{X}_t)\\ &= e^{-\int_0^t (\dot{y}_s \circ (\mathrm{Id}-u))(\mathrm{dB}_s + \dot{u}_s\mathrm{d}s + \frac{1}{2}\dot{y}_s \circ (\mathrm{Id}-u)\mathrm{d}s)} e^{-\int_0^t \mathbb{E}(\dot{u}_s|\mathcal{X}_s)\mathrm{dZ}_s - \frac{1}{2}\int_0^t \mathbb{E}(\dot{u}_s|\mathcal{X}_s)^2\mathrm{d}s}.\end{aligned}$$

Make he substitution $\mathrm{dZ}_t = \mathrm{dB}_t + (\dot{u}_s - \mathbb{E}(\dot{u}_s \mid \mathcal{X}_s))\mathrm{d}s$ in the above formula shows that a continuous semimartingale vanishes almost surely for any $t > 0$, hence its bracket is 0, which means $\dot{y}_s \circ (\mathrm{Id} - u) = \mathbb{E}_{\mathrm{P}}(\dot{u}_s \mid \mathcal{X}_s)$ $\mathrm{d}t \times \mathrm{dP}$ almost surely. □

Chapter 6

Representations

Chapter 7

Concentration of measure

Obtaining sub-gaussian tails for probability measures, like

$$\mathbb{P}(|\mathrm{X}-\mathbb{E}(\mathrm{X})| \geqslant \delta) \leqslant \mathrm{C}e^{-c\delta^2}, \tag{7.1}$$

is an important goal and tool, making it possible to quantify fluctuations of a random variable around its mean. For instance, for random variables depending on an index n (often the dimension), this allows to bound the variance and therefore to prove tightness, which is an important step towards convergence in law.

To motivate our study and understand the nature of concentration results (high dimension, small fluctuations around the mean), we consider in this introduction some examples proved in the following sections. First, let σ be the uniform probability measure on the unit sphere $\mathscr{S}^{(n)} \subset \mathbb{R}^{n+1}$, and S a subset of $\mathscr{S}^{(n)}$ such that $\mu^{(n)}(\mathrm{S}) \geqslant 1/2$. Consider the neighborhood

$$\mathrm{S}_\varepsilon = \{x \in \mathscr{S}^{(n)} \mid \mathrm{dist}(x, \mathrm{S}) \leqslant \varepsilon\}, \tag{7.2}$$

where dist is the geodesic distance on the sphere. Then

$$\sigma(\mathrm{S}_\varepsilon) \geqslant 1 - e^{-\frac{(n-1)\varepsilon^2}{2}}. \tag{7.3}$$

In particular, as $n \to \infty$ the uniform measure gets concentrated in a belt around the equator. This result (7.3) can be shown by first proving that the minimum of $\sigma(\mathrm{S}_\varepsilon)$ is obtained for an hemisphere, and then by a direct integration. The intuition behind this last step is that a unit vector of $\mathbb{R}^n$ has typical amplitude $\mathrm{O}(1/\sqrt{n})$ for the nth coordinate, so the volume of the neighborhood will be close to 1 if $n\varepsilon^2 \gg 1$.

Concerning the first step, this isoperimetry result holds for all compact manifold with strictly positive Ricci curvature, as shown by Gromov, Lévy. More precisely, let $\mathscr{M}$ be a compact manifold with Riemannian metric g and Riemannian probability measure μ. Let $\mathrm{R}(\mathscr{M})$ be the infimum of $\mathrm{Ric}(u,u)$, over all unit tangent vectors u on $\mathscr{M}$, and assume $\mathrm{R}(\mathscr{M}) > 0$. For some $r > 0$, $\mathrm{R}(\mathscr{M}) = \mathrm{R}(\mathscr{S}^n(r))$, where $\mathscr{S}^{(n)}(r)$ is a Euclidean sphere of dimension n and radius r, with uniform probability measure noted σ_r. Then, if some Borell subsets $\mathrm{M} \subset \mathscr{M}$, $\mathrm{S} \subset \mathscr{S}$, have the same measure ($\mu(\mathrm{E}) = \sigma_r(\mathrm{S})$),

$$\mu(\mathrm{M}_\varepsilon) \geqslant \sigma_r(\mathrm{S}_\varepsilon),$$

for any $\varepsilon > 0$[1]. Note that, by the Minkowski content formula, this implies that the length of the border $\partial\mathrm{M}$ is greater than $\partial\mathrm{S}$: this is an isoperimetry result.

As $\sigma_r(\mathrm{S}_\varepsilon)$ can be easily estimated, we get that for any $\varepsilon > 0$, if $\mu(\mathrm{M}) \geqslant 1/2$, for some absolute constant $c, \mathrm{C} > 0$

$$\mu(\mathrm{M}_\varepsilon) \geqslant 1 - \mathrm{C}e^{-c\mathrm{R}\varepsilon^2}. \tag{7.4}$$

We will directly prove such results for any Lipschiz[2] function F by using stochastic analysis tools, regardless isoperimetry considerations.

---

1. for a Riemannian manifold $\mathcal{M}$ with metric g, the geodesic distance allows to define a neighborhood M_ε of $\mathrm{M} \subset \mathscr{M}$ up to distance ε in the same way as (7.2)
2. in this chapter, the Lipschitz constant $\|\mathrm{F}\|_{\mathcal{L}}$ of a Lipschitz function $\mathrm{F} : \mathbb{R}^n \to \mathbb{R}$ is defined with respect to the Euclidean norm : for all x, y in $\mathbb{R}^n$, $|\mathrm{F}(x) - \mathrm{F}(y)| \leqslant \|\mathrm{F}\|_{\mathcal{L}} |x-y|_{\mathrm{L}^2(\mathbb{R}^n)}$

Theorem 7.1. *For some* $\mathrm{C}, c > 0$*, independent of the manifold* $\mathscr{M}$ *and the function* F *with Lipschitz constant (with respect to the Riemanian metric)* $\|\mathrm{F}\|_{\mathcal{L}}$

$$\mu\left\{|\mathrm{F} - \mathbb{E}_\mu(\mathrm{F}) > \varepsilon|\right\} \leqslant \mathrm{C}e^{-c\mathrm{R}(\mathscr{M})\frac{\varepsilon^2}{\|\mathrm{F}\|^2_{\mathcal{L}}}}.$$

There are two interesting regimes for these results. For $\varepsilon \to 0$, this can yield isoperimetric estimates like (7.3). For $\varepsilon > 0$, as the dimension increases, this often yields concentration, because the curvature of the sphere $\mathscr{S}^{(n)}$, with radius 1, is $\mathrm{R} = n - 1$. Note that Theorem 7.1 easily implies (7.4), by choosing $\mathrm{F}(x) = \min(\mathrm{dist}(x, \mathrm{M}), \varepsilon)$.

Bounds like (7.1) or more specifically (7.3) are called *concentration inequalities*, and many general criteria allow to prove them. Our aim consists in giving a non-exhaustive (see [10] for a review of the concentration phenomenon) description of sufficient conditions, emphasizing on how, quite surprisingly, some stochastic processes allow to prove these time-independent relations. For example, concentration yields a good understanding of the maximum of any Gaussian process, at first order.

Theorem 7.2. *Let* $g(t)_{t\in\mathrm{E}}$ *be a centered Gaussian process indexed by a countable set* E*, and such that* $\sup_{\mathrm{E}} g(t) < \infty$ *almost surely. Then* $\mathbb{E}(\sup_{\mathrm{E}} g(t)) < \infty$ *and for every* $\delta \geqslant 0$

$$\mathbb{P}\left(|\sup_{\mathrm{E}} g(t) - \mathbb{E}(\sup_{\mathrm{E}} g(t))| > \delta\right) \leqslant 2e^{-\frac{\delta^2}{2\sigma^2}},$$

where σ^2 *is the maximum of the variances*[3]*,* $\sup_{\mathrm{E}} \mathbb{E}(g(t)^2)$*.*

As we will see, this result by Borell is for example a consequence of the hypercontractivity of the Ornstein-Uhlenbeck semigroup.

Going to the discrete setting, hypercontractivity is an important tool in computer science for example. Consider the state space $\{-1, 1\}^d$ endowed with the uniform probability measure. To a function $f : \{-1, 1\}^d \to \{-1, 1\}$ (f can be thought as a decision[4], function of d individual behaviors) one associates the function, $\mathrm{T}_\rho f$ equal at point x to $\mathbb{E}(f(u))$, where u is obtained from x by independently flipping each bit with probability $\frac{1-\rho}{2}$: T_ρ is called the noise operator, and is an analogue of the Ornstein-Uhlenbeck semigroup[5]. Then this semigroup satisfies an hypercontrativity property, which is fundamental in the proof of the following famous result with from Kahn, Kanal, Linial [8] (the slightly stronger version stated here comes from Talagrand [20]).

For $x = (x_1, \ldots, x_d) \in \{-1, 1\}^d$, define the influence of individual i as the probability that changing x_i changes f. Let $-i$ for the vector of size $n - 1$ obtained from $(1, \ldots, d)$ by removing the ith coordinate, and E_{S} the expectation with respect to the uniform measures for coordinates in $\mathrm{S} \subset [\![1, d]\!]$,

$$\mathrm{infl}_i(f) = \mathbb{E}_{-i}\,\mathrm{var}_i(f).$$

Then, an easy calculation proves that

$$\sum_{i=1}^{d} \mathrm{infl}_i(f) \geqslant \mathrm{var}(f).$$

Indeed, one naturally expects that the influence of each coordinate on f will be of order $1/n$ times the variance. The remarkable result from [8] is that for some individuals, the influence goes till $\frac{\log d}{d}$ times the variance.

3. That $\sigma^2 < \infty$ is an easy consequence of $\mathbb{E}(\sup_{\mathrm{E}} g(t)) < \infty$
4. e.g. $f(x_1, \ldots, x_d) = x_1$ is called the dictatorship case, $f(x_1, \ldots, x_d) = x_1 \ldots x_d$ is the parity, and $f(x_1, \ldots, x_d) = \mathrm{sgn}(x_1 + \cdots + x_d)$ is the majority.
5. Think about $\rho = e^{-t}$, with t a time, so that $\mathrm{T}_1 f = f$ and $\mathrm{T}_0 f = \mathbb{E}(f)$ are natural

Theorem 7.3. *Under the above assumptions, thee is a constant $c > 0$ not depending on d such that for any Boolean function f*

$$\sum_{k=1}^{d} \frac{\mathrm{infl}_k(f)}{-\log \mathrm{infl}_k(f)} \geqslant c\,\mathrm{var}(f).$$

This statement allows to give a framework for sentences like 3% *of the population can decide about the ending of the vote with probability* 99%. Of course, the proof will not be constructive, so gives no clues about how to find this precious sampling amongst the coordinates. In the proof we will give, the noise operator is useful because it improves integrability (hypercontractivity) but also because $\mathrm{T}_\rho f$ very explicit in terms of f, when this function is expanded on a Fourier basis. This is very similar to the use of the Ornstein-Uhlenbeck semigroup to prove continuity of chaos in L^p, see Chapter 6.

This chapter only gives a partial view of concentration results. Another aspect I particularly like (but we won't treat in this class) is the Talagrand inequalities [21]. They allow for example to prove this visual example of concentration, given by Tao and Vu [19], which is an important tool in Random Matrix Theory.

Theorem 7.4. *Let $\mathrm{X}_1, \dots, \mathrm{X}_n$ be i.i.d. centered*[6] *Bernoulli random variables, $\mathrm{X} = (\mathrm{X}_1, \dots, \mathrm{X}_n)$, and V a given subspace of $\mathbb{R}^n$ of dimension d. Then*

$$\mathbb{P}(|\mathrm{dist}(\mathrm{X}, \mathrm{V}) - \sqrt{n-d}| \geqslant \delta) \leqslant \mathrm{C}e^{-c\delta^2},$$

for some constants C and c independent of n and d.

The above result means that the typical distance between a random element of the hypercube and a subspace of dimension d is $\sqrt{n-d}$, with small (concentrated with no normalization) deviations from the mean.

1. Hypercontractivity, logarithmic Sobolev inequalities and concentration

1.1. Diffusion semigroups

On $\mathbb{R}^d$ endowed with a measure μ, consider a family of operators $(\mathrm{P}_t)_{t\geqslant 0}$ acting on bounded measurable functions through a transition kernel

$$\mathrm{P}_t\, f(x) = \int_{\mathbb{R}^d} f(y) p_t(x, \mathrm{d}y),$$

the measures $p_t(x,\cdot)$ being bounded and positive, $t > 0$. We will only consider cases in which the domain of P_t can be extended to $\mathrm{L}^2(\mu)$: $\|\mathrm{P}_t f\|_{\mathrm{L}^2(\mu)} \leqslant c(t)\|f\|_{\mathrm{L}^2(\mu)}$. We are interested in the case of continuous semigroups, meaning :

- the semigroup property : $\mathrm{P}_t \circ \mathrm{P}_s = \mathrm{P}_{t+s},\ s,t > 0$;
- the continuity in $\mathrm{L}^2(\mu)$: for all $f \in \mathrm{L}^2(\mu)$, as $t \to 0$,

$$\mathrm{P}_t\, f \underset{\mathrm{L}^2(\mu)}{\longrightarrow} f.$$

Moreover, if $\mathrm{P}_t\, \mathbb{1} = \mathbb{1}$ for any $t > 0$ (Markovian semigroup), there exists a Markov process $(\mathrm{X}_t)_{t\geqslant 0}$ such that[7]

$$\mathrm{P}_t\, f(x) = \mathbb{E}\left(f(\mathrm{X}_t) \mid \mathrm{X}_0 = x\right).$$

6. $\mathbb{P}(\mathrm{X}_1 = 1) = 1/2,\ \mathbb{P}(\mathrm{X}_1 = -1) = 1/2$
7. Cf the Hille-Yoshida theorem

Definition 7.5. *The domain $\mathbb{D}_2(\mathrm{L}) \subset \mathrm{L}^2(\mu)$ is the set of functions $f \in \mathrm{L}^2(\mu)$ such that the limit*

$$\mathrm{L}\, f := \lim_{t\to 0} \frac{\mathrm{P}_t\, f - f}{t}$$

exists in $\mathrm{L}^2(\mu)$. Then $\mathbb{D}_2(\mathrm{L})$ is dense in $\mathrm{L}^2(\mu)$ and L characterizes the semigroup :

$$\partial_t\, \mathrm{P}_t\, f = \mathrm{L}\, \mathrm{P}_t\, f = \mathrm{P}_t\, \mathrm{L}\, f.$$

The semigroup $(\mathrm{P}_t)_{t\geqslant 0}$ is said to be a diffusion semigroup if for any $f_1, \dots, f_n$ in the domain and $\Phi : \mathbb{R}^n \to \mathbb{R}$ in $\mathscr{C}^\infty$

$$\mathrm{L}\, \Phi(f_1, \dots, f_n) = \sum_i \partial_{x_i}\Phi(f_1, \dots, f_n)\, \mathrm{L}\, f_i + \sum_{i,j} \partial_{x_i x_j}(f_1, \dots, f_n)\Gamma_1(f_i, f_j),$$

where $\Gamma_1(f, g) = \frac{1}{2}\left(\mathrm{L}(fg) - f\, \mathrm{L}\, g - g\, \mathrm{L}\, f\right)$ is the operator carré du champ.

The measure μ is said to be invariant by $(\mathrm{P}_t)_{t\geqslant 0}$ if for any $f \in \mathrm{L}^1(\mu)$, $t \geqslant 0$,

$$\int \mathrm{P}_t\, f \mathrm{d}\mu = \int f \mathrm{d}\mu.$$

The measure μ is said to be reversible by $(\mathrm{P}_t)_{t\geqslant 0}$ if for any $f, g \in \mathrm{L}^1(\mu)$, $t \geqslant 0$

$$\int (\mathrm{P}_t\, f) g \mathrm{d}\mu = \int g(\mathrm{P}_t\, f)\mathrm{d}\mu.$$

Note that, for Markovian semigroups, the reversibility implies the invariance. Moreover, for a diffusion semigroup, $\mathrm{L}(1) = 0$, so the semigroup is Markovian.

1.2. Logarithmic Sobolev inequalities

The Sobolev inequality bounds the fluctuations of a function in terms of its derivative, and therefore can somehow yield to some concentration. For example, any compactly supported smooth function $f : \mathbb{R}^n \to \mathbb{R}$, if $f \in \mathrm{L}_p(\mathbb{R}^n, \mathrm{d}x)$ and $\partial_{x_i} f \in \mathrm{L}_p(\mathbb{R}^n, \mathrm{d}x)$, $1 \leqslant i \leqslant n$, then $f \in \mathrm{L}_{p^*}(\mathbb{R}^n, \mathrm{d}x)$ and

$$\|f\|_{p^*} \leqslant c(n, p)\|\nabla f\|_p, \tag{7.5}$$

where $1 \leqslant p < n$, $p^* = \frac{np}{n-p} > p$: a Sobolev inequality provides an improved integrability on f in terms of that of its derivative. In probabilistic terms, this allows for example to bound a variance in terms of the Lipschitz constant. One may want to quantify the fluctuations of the function in terms of its gradient in a dimensionless inequality, which cannot be achieved in (7.5) (e.g. $p^* \to p$ as $n \to \infty$, hence it doesn't give additional integrability on f).

General concentration estimates, of type (7.1), can be obtained as a corollary of the following logarithmic analogue of (7.5).

Definition 7.6. *A measure μ on some space*[8] X *is said to satisfy a logarithmic Sobolev inequality with constant $c > 0$ if for any non-negative f*

$$\mathbb{E}_\mu\left(f^2 \log(f^2)\right) - \mathbb{E}_\mu\left(f^2\right) \log \mathbb{E}_\mu\left(f^2\right) \leqslant c\, \mathbb{E}_\mu\left(|\nabla f|^2\right).$$

Here, the definition of ∇f depends on the context and its modulus is considered as ∞ in the non-differentiable case.

We will see later general very efficient criteria on μ for satisfying this general inequality. The above-stated logarithmic Sobolev inequality, corresponding to $p = 2$, implies analogues for any $p > 2$.

8. In this course X will be either a Riemannian manifold or $\{-1, 1\}^n$.

Lemma 7.7. *Let* $(\mathrm{P}_t)_{t\geqslant 0}$ *be a diffusion semigroup with invariant measure* μ*. Assume that* μ *satisfies a logarithmic Sobolev inequality with constant* c*.*

Then for any $p \geqslant 2$ *and any positive function* $f \in \mathbb{D}$

$$\mathbb{E}_\mu\left(f^p \log(f^p)\right) - \mathbb{E}_\mu\left(f^p\right)\log \mathbb{E}_\mu\left(f^p\right) \leqslant -c\frac{p^2}{4(p-1)}\,\mathbb{E}_\mu\left(f^{p-1}\,\mathrm{L}\,f\right).$$

Proof. If $(\mathrm{P}_t)_{t\geqslant 0}$ is a diffusion semigroup[9], then

$$\begin{aligned}
\frac{p^2}{4(p-1)}\,\mathbb{E}_\mu\left(f^{p-1}\,\mathrm{L}\,f\right) &= -\frac{p^2}{4}\,\mathbb{E}_\mu\left(f^{p-2}\Gamma_1(f)\right)\\
&= -\frac{p}{2}\,\mathbb{E}_\mu\left(f^{\frac{p}{2}-1}\Gamma_1(f, f^{p/2})\right)\\
&= -\frac{p}{2}\,\mathbb{E}_\mu\left(\Gamma_1(f^{\frac{p}{2}})\right)\\
\frac{p^2}{4(p-1)}\,\mathbb{E}_\mu\left(f^{p-1}\,\mathrm{L}\,f\right) &= \mathbb{E}_\mu\left(f^{\frac{p}{2}}\,\mathrm{L}\,f^{\frac{p}{2}}\right),
\end{aligned}$$

hence the expected result is the logarithmic Sobolev inequality applied to the function $f^{\frac{p}{2}}$. □

An interesting feature of logarihmic Sobolev inequalities is that they tensorize, in the following way. This will be useful in Section 4. To state this property, given a function f on $\mathrm{X} = \mathrm{X}_1 \times \cdots \times \mathrm{X}_d$, note

$$f_i(x_i) = f(x_1, \ldots, x_n),$$

the function f_i depending on the frozen coordinates $x_1, \ldots, x_{i-1}, x_{i+1}, dots, x_d$.

Proposition 7.8. *Assume that for any* $1 \leqslant k \leqslant d$ *the measure* μ_k *on* X_k *satisfies a logarithmic Sobolev inequality*[10] *with constant* c_k *: for any non-negative function* h

$$\mathrm{Ent}_{\mu_k}(h^2) \leqslant c_k \int_{\mathrm{X}_k} |\nabla_k h|^2 \mathrm{d}\mu_k.$$

Then $\mu = \mu_1 \otimes \cdots \otimes \mu_k$ *satisfies a logarithmic Sobolev inequality with constant* $c = \max_{1\leqslant k\leqslant d} c_k$ *: for any non-negative function* f *on* X

$$\mathrm{Ent}_\mu(f^2) \leqslant c \int_{\mathrm{X}} |\nabla f|^2 \mathrm{d}\mu,$$

where $|\nabla f|^2$ *is defined as* $\sum_{k=1}^d |\nabla_k f_k|^2$.

Proof. The statement easily follows from the inequality

$$\mathrm{Ent}_\mu(f) \leqslant \int \sum_{k=1}^{d} \mathrm{Ent}_{\mu_k}(f_k)\mathrm{d}\mu, \tag{7.6}$$

that we will prove by an extremal characterization of the entropy :

$$\mathrm{Ent}_\mu(f) = \sup\{\int f h \mathrm{d}\mu \mid \int e^h \mathrm{d}\mu \leqslant 1\}. \tag{7.7}$$

9. we know that for diffusion semigroups

$$\Gamma_1(\mathrm{F}(f), g) = \mathrm{F}'(f)\Gamma_1(f, g),\ \mathbb{E}_\mu(\mathrm{F}'(f)\,\mathrm{L}\,f) + \mathbb{E}_\mu(\mathrm{F}''(f)\Gamma_1(f)) = 0.$$

10. Here, as in Definition 7.6, the definition of the gradient is arbitrary/depending on the context and has no influence on the conclusion of the proposition

Indeed, first note that we can suppose that $\int f\mathrm{d}\mu = 1$ by homogeneity (if $\lambda > 0$, $\mathrm{Ent}(\lambda f) = \lambda\,\mathrm{Ent}(f)$). Then, Young's inequality (f non-negative and g real) $fh \leqslant f\log f - f + e^h$ yields, once integrated,

$$\int fh\mathrm{d}\mu \leqslant \mathrm{Ent}_\mu(f) - 1 + \int e^h\mathrm{d}\mu.$$

This proves that the left hand side of (7.7) is greater than the right side. The reverse inequality is obtained by choosing $h = \log f$. As a consequence of (7.7), a sufficient condition to prove (7.6) is that if $\int e^h\mathrm{d}\mu \leqslant 1$ then

$$\int fh\mathrm{d}\mu \leqslant \int \sum_{k=1}^{d} \mathrm{Ent}_{\mu_k}(f_k)\mathrm{d}\mu.$$

This is true by interpolating between 1 and h in the following way. For $1 \leqslant k \leqslant n$, let

$$h^{(k)}(x_k, \dots, x_n) = \log\left(\int e^h\mathrm{d}\mu_1 \dots \mathrm{d}\mu_{k-1}\right) - \log\left(\int e^h\mathrm{d}\mu_1 \dots \mathrm{d}\mu_k\right).$$

Then, as $\int e^h\mathrm{d}\mu \leqslant 1$, $\sum_{k=1}^d h^{(k)} \geqslant h$. Moreover, obviously $\int e^{h_k^{(k)}}\mathrm{d}\mu_k = 1$. Hence

$$\int fh\mathrm{d}\mu \leqslant \sum_{k=1}^{d}\int fh^{(k)}\mathrm{d}\mu = \sum_{k=1}^{d}\int\left(\int f_k h_k^{(k)}\mathrm{d}\mu_k\right)\mathrm{d}\mu \leqslant \int\sum_{k=1}^{d}\mathrm{Ent}_{\mu_k}(f_k)\mathrm{d}\mu,$$

as expected. □

1.3. Herbst's lemma

The link between Logarithmic Sobolev inequalities and concentration lies in the following Lemma, by Herbst.

Lemma 7.9. *Assume that μ satisfies a logarithmic Sobolev inequality on $\mathbb{R}^n$ with constant c. Let $\mathrm{F} : \mathbb{R}^n \to \mathbb{R}$ be a Lipschiz function, with Lipschitz constant $\|\mathrm{F}\|_{\mathcal{L}}$. Then, for any $\lambda \in \mathbb{R}$,*

$$\mathbb{E}_\mu\left(e^{\lambda(\mathrm{F}-\mathbb{E}_\mu(\mathrm{F}))}\right) \leqslant e^{c\lambda^2\|\mathrm{F}\|_{\mathcal{L}}^2/4}. \tag{7.8}$$

In particular, for any $\delta > 0$,

$$\mu\left\{|\mathrm{F} - \mathbb{E}_\mu(\mathrm{F})| > \delta\right\} \leqslant 2e^{-\frac{\delta^2}{c\|\mathrm{F}\|_{\mathcal{L}}^2}}. \tag{7.9}$$

Proof. First note that (7.9) is a consequence of (7.8) : for any $\lambda > 0$, by the Bienaymé-Chebyshev inequality,

$$\begin{aligned}\mu\left\{|\mathrm{F} - \mathbb{E}_\mu(\mathrm{F})| > \delta\right\} &\leqslant e^{-\lambda\delta}\,\mathbb{E}_\mu\left(e^{\lambda|\mathrm{F}-\mathbb{E}_\mu(\mathrm{F})|}\right)\\ &\leqslant e^{-\lambda\delta}\left(\mathbb{E}_\mu\left(e^{\lambda(\mathrm{F}-\mathbb{E}_\mu(\mathrm{F}))}\right) + \mathbb{E}_\mu\left(e^{-\lambda(\mathrm{F}-\mathbb{E}_\mu(\mathrm{F}))}\right)\right) \leqslant 2e^{-\lambda\delta}e^{c\lambda^2\|\mathrm{F}\|_{\mathcal{L}}^2/4}.\end{aligned}$$

The optimal choice $\lambda = \frac{2\delta}{c\|\mathrm{F}\|_{\mathcal{L}}^2}$ yields (7.9).

To prove (7.8), assume first that F is differentiable, with uniformly bounded derivatives :

$$\|\mathrm{F}\|_{\mathcal{L}} = \sup_{x\in\mathbb{R}^n}\|\nabla\mathrm{F}\|_{\mathrm{L}^2}(x) < \infty.$$

Moreover, we can suppose $\mathbb{E}_\mu(\mathrm{F}) = 0$. Let $f(\lambda) = \log\|e^{2\mathrm{F}}\|_{\mathrm{L}^\lambda}$, $\lambda > 0$. Then a direct calculation yields

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}f(\lambda) = \frac{1}{\lambda^2}\frac{\mathbb{E}_\mu(2\lambda\mathrm{F}e^{2\lambda\mathrm{F}}) - \mathbb{E}_\mu(e^{2\lambda\mathrm{F}})\log\mathbb{E}_\mu(e^{2\lambda\mathrm{F}})}{\mathbb{E}_\mu(e^{2\lambda\mathrm{F}})},$$

which is bounded by the logarithmic Sobolev inequality by

$$\frac{c\,\mathbb{E}_\mu(\|\nabla e^{\lambda \mathrm{F}}\|^2_{\mathrm{L}^2})}{\lambda^2\,\mathbb{E}_\mu(e^{2\lambda\mathrm{F}})} \leqslant c\sup_{\mathbb{R}^n}\|\nabla \mathrm{F}\|^2_{\mathrm{L}^2}.$$

Moreover, as $\mathbb{E}_\mu(\mathrm{F}) = 0$, $\lim_{\lambda\to 0^+} f(\lambda) = 0$, hence $f(\lambda) \leqslant \lambda c \sup_{\mathbb{R}^n}\|\nabla \mathrm{F}\|^2_{\mathrm{L}^2}$ for all $\lambda > 0$, which is the expected result. If $\lambda < 0$, the same result holds by considering $-\mathrm{F}$.

If F is not differentiable with uniformly bounded derivatives, then we proceed by approximation :

- We note, for $\varepsilon > 0$, $\overline{\mathrm{F}}_\varepsilon(x) = \mathrm{F}(x)$ if $-1/\varepsilon < \mathrm{F}(x) < 1/\varepsilon$, $-1/\varepsilon$ if $\mathrm{F}(x) < \varepsilon$ and $1/\varepsilon$ if $\mathrm{F}(x) > -1/\varepsilon$. Noting p_ε the Gaussian density $p_\varepsilon(x) = e^{-\frac{|x|^2}{2\varepsilon}}/(2\pi\varepsilon)^{n/2}$, and the convolution $\mathrm{F}_\varepsilon(x) = p_\varepsilon \star \overline{\mathrm{F}}$, then $\mathrm{F}_\varepsilon(x)$ converges simply to F ($|\mathrm{F}_\varepsilon(x) - \mathrm{F}(x)| \leqslant \|\mathrm{F}\|_{\mathcal{L}}\sqrt{\varepsilon n}$), it is continuously differentiable and we will be able to apply the above discussion to it.

- Indeed, the gradient of $\mathrm{F}_\varepsilon(x)$ is uniformly bounded : it is an easy exercise that $\sup_{\mathbb{R}^n}\|\nabla \mathrm{F}_\varepsilon\|_{\mathrm{L}^2} < \|\mathrm{F}\|_{\mathcal{L}}$, thus from the differentiable case, we know that

$$\mathbb{E}_\mu\left(e^{\lambda \mathrm{F}_\varepsilon}\right) \leqslant e^{\lambda\,\mathbb{E}_\mu(\mathrm{F}_\varepsilon)} e^{c\lambda^2\|\mathrm{F}\|^2_{\mathcal{L}}/4}. \tag{7.10}$$

- Thanks to the simple convergence we can use Fatou's lemma, and get $\mathbb{E}_\mu(e^{\lambda\mathrm{F}}) \leqslant e^{\liminf_{\varepsilon\to 0}\lambda\,\mathbb{E}_\mu(\mathrm{G}_\varepsilon)} e^{c\lambda^2\|\mathrm{F}\|^2_{\mathcal{L}}/2}$. Hence the proof will be complete if $\lim_{\varepsilon\to 0}\mathbb{E}_\mu\,\mathrm{F}_\varepsilon = \mathbb{E}_\mu\,\mathrm{F}$.

- From the concentration property of F_ε (7.10),

$$\mu\{|\mathrm{F}_\varepsilon - \mathbb{E}_\mu(\mathrm{F}_\varepsilon)| > \delta\} \leqslant \mathrm{C}e^{-c\delta^2},$$

  with C and c independent of ε. Hence $\mathrm{F}_\varepsilon - \mathbb{E}_\mu(\mathrm{F}_\varepsilon)$ is uniformly integrable. If we can show that $\mathbb{E}_\mu(\mathrm{F}_\varepsilon)$ is bounded, it will give the uniform integrability of F_ε, hence the expected convergence $\lim_{\varepsilon\to 0}\mathbb{E}_\mu\,\mathrm{F}_\varepsilon = \mathbb{E}_\mu\,\mathrm{F}$.

- As F_ε converges simply to F, by dominated convergence we can find $\mathrm{K}, \varepsilon_0$ such that, for $\varepsilon < \varepsilon_0$, $\mu\{|\mathrm{F}_\varepsilon| < \mathrm{K}\} > 3/4$.

  Moreover, from (7.10), for r sufficiently large, $\mu\{|\mathrm{F}_\varepsilon - \mathbb{E}_\mu(\mathrm{F}_\varepsilon)| < r\} > 3/4$. Hence, for $\varepsilon < \varepsilon_0$, with probability at least 1/2, both $|\mathrm{F}_\varepsilon| < \mathrm{K}$ and $|\mathrm{F}_\varepsilon - \mathbb{E}_\mu(\mathrm{F}_\varepsilon)| < r$, hence $\mathbb{E}_\mu(\mathrm{F}_\varepsilon) < \mathrm{K} + r$: this event, has measure 0 or 1, hence it is almost sure.

□

1.4. Gross's Theorem

Definition 7.10. *Given a strictly increasing function $q : \mathbb{R}^+ \to [q(0), \infty)$, a semi-group $(\mathrm{P}_t)_{t\geqslant 0}$ is said to be hypercontractive with contraction function q if, for any $f \in \mathscr{D}$ and any $t \geqslant 0$,*

$$\|\mathrm{P}_t\, f\|_{q(t)} \leqslant \|f\|_{q(0)}.$$

Theorem 7.11 (Gross). *Let $(\mathrm{P}_t)_{t\geqslant 0}$ be a Markov semigroup with invariant measure μ such that either μ is reversible or $(\mathrm{P}_t)_{t\geqslant 0}$ is a diffusion semigroup.*

If $(\mathrm{P}_t)_{t\geqslant 0}$ is hypercontractive with constant $q(t) = 1 + e^{\frac{4t}{c}}$, $c > 0$, then μ satisfies a logarithmic Sobolev inequality with constant c.

If μ satisfies a logarithmic Sobolev inequality with a constant $c > 0$, then for any $q(0) > 1$ and $q(t) = 1 + (q(0) - 1)e^{\frac{4t}{c}}$, $(\mathrm{P}_t)_{t\geqslant 0}$ is hypercontractive with contraction function q.

Proof. The proof relies on the differentiation with respect to t,

$$\frac{\mathrm{d}}{\mathrm{d}t}\log\|\mathrm{P}_t f\|_{q(t)} = \frac{q'(t)}{q(t)^2\,\mathbb{E}_\mu\left((\mathrm{P}_t f)^{q(t)}\right)}\left(\int (\mathrm{P}_t f)^{q(t)}\log((\mathrm{P}_t f)^{q(t)})\mathrm{d}\mu\right.$$
$$\left. - \int (\mathrm{P}_t f)^{q(t)}\mathrm{d}\mu \log\left(\int (\mathrm{P}_t f)^{q(t)}\mathrm{d}\mu\right) + \frac{q(t)^2}{q'(t)}\,\mathbb{E}_\mu\left((\mathrm{P}_t f)^{q(t)-1}\,\mathrm{L}\,\mathrm{P}_t f\right)\right). \quad (7.11)$$

Suppose first that $(\mathrm{P}_t)_{t\geqslant 0}$ is hypercontractive with contraction function $q(t) = 1 + e^{\frac{4t}{c}}$:

$$\|\mathrm{P}_t f\|_{q(t)} \leqslant \|\mathrm{P}_0 f\|_{q(0)},$$

hence the derivative of $t \mapsto \log\|\mathrm{P}_t f\|_{q(t)}$ is negative at 0. From the above differentiation, and as $q(0)^2/q'(0) = c$, this means

$$\int f^2\log(f^2)\mathrm{d}\mu - \int f^2\mathrm{d}\mu\log\left(\int f^2\mathrm{d}\mu\right) \leqslant -c\int f\,\mathrm{L}\,f\mathrm{d}\mu,$$

which is the expected logarithmic Sobolev inequality.

If μ satisfies a logarithmic Sobolev inequality, with constant $c > 0$, then Lemma 7.7 applied to $\mathrm{P}_t f$ yields

$$\int (\mathrm{P}_t f)^{q(t)}\log((\mathrm{P}_t f)^{q(t)})\mathrm{d}\mu - \int (\mathrm{P}_t f)^{q(t)}\mathrm{d}\mu\log\left(\int (\mathrm{P}_t f)^{q(t)}\mathrm{d}\mu\right)$$
$$\leqslant -c\frac{q(t)^2}{4(q(t)-1)}\int (\mathrm{P}_t f)^{q(t)-1}\,\mathrm{L}\,\mathrm{P}_t f\mathrm{d}\mu.$$

For our choice $q(t) = 1 + (q(0)-1)e^{4t/c}$,

$$c\frac{q(t)^2}{4(q(t)-1)} = \frac{q(t)^2}{q'(t)},$$

hence from the differentiation (7.11) the function $t \mapsto \log\|\mathrm{P}_t f\|_{q(t)}$ is decreasing, which means the expected hypercontractivity. □

2. Concentration on Euclidean spaces

2.1. Concentration for the extremum of Gaussian processes.

We are now in position to prove Borell's inequality.

Theorem 7.12. *Let $g(t)_{t\in\mathrm{E}}$ be a centered Gaussian process indexed by a countable set* E*, and such that* $\sup_{\mathrm{E}} g(t) < \infty$ *almost surely. Then* $\mathbb{E}(\sup_{\mathrm{E}} g(t)) < \infty$ *and for every* $\delta \geqslant 0$

$$\mathbb{P}\left(|\sup_{\mathrm{E}} g(t) - \mathbb{E}(\sup_{\mathrm{E}} g(t))| > \delta\right) \leqslant 2e^{-\frac{\delta^2}{2\sigma^2}},$$

where σ^2 is the maximum of the variances[11], $\sup_{\mathrm{E}}\mathbb{E}(g(t)^2)$.

Proof. Consider the Gaussian probability measure

$$\mathrm{d}\mu = \frac{1}{\mathrm{Z}_n}e^{-\mathrm{V}(x)}\mathrm{d}x = \frac{1}{(2\pi)^{n/2}}e^{-\sum_{k=1}^n\frac{x_k^2}{2}}\mathrm{d}x_1\ldots\mathrm{d}x_n$$

on $\mathbb{R}^n$. Obviously, $\mathrm{Hess}(\mathrm{V}) \geqslant \mathrm{Id}_n$, so μ satisfies a Logarithmic Sobolev inequality with constant 2, by the Bakry-Emery Theorem 7.13. Now, consider a finite set of points

11. That $\sigma^2 < \infty$ is an easy consequence of $\mathbb{E}(\sup_{\mathrm{E}} g(t)) < \infty$

$t_1,\dots,t_n$ in E, and the Gaussian vector $g=(g(t_1),\dots,g(t_n))$, with covariance ${}^{\mathrm{t}}\mathrm{VV}$: $g\overset{\mathrm{law}}{=}\mathrm{V}x$ where x is a vector of standard centered Gaussians.

Consider $x\to\sup_{1\leqslant i\leqslant n}(\mathrm{V}x)_i$. This is a Lipschiz function with constant σ :

$$\begin{aligned}|\mathrm{F}(x)-\mathrm{F}(y)|&\leqslant\sup_{1\leqslant i\leqslant n}|(\mathrm{V}x)_i-(\mathrm{V}y)_i|=|\sum_{j=1}^n\mathrm{V}_{ij}(x_j-y_j)|\\&\leqslant\sqrt{\sum_{j=1}^n\mathrm{V}_{i,j}^2}|x-y|_{\mathrm{L}^2}=\sqrt{\mathrm{var}(g(t_i))}|x-y|_{\mathrm{L}^2}\leqslant\sigma|x-y|_{\mathrm{L}^2}.\end{aligned}$$

Consequently, we can apply Herbst's Lemma 7.9 and get

$$\mathbb{P}\left(\left|\sup_{1\leqslant k\leqslant n}g(t_k)-\mathbb{E}(\sup_{1\leqslant k\leqslant n}g(t_k))\right|\geqslant\delta\right)\leqslant e^{-\frac{\delta^2}{2\sigma^2}}.\tag{7.12}$$

Note that this upper bound does not depend on n, which makes us confident in extending the result to a countable set T. We can achieve this in the following way.

First, $\mathbb{E}(\sup_{\mathrm{T}}g(t))<\infty$. Indeed, let T_n be a finite subset of T. Then for any $a,b>0$,

$$\{\mathbb{E}(\sup_{\mathrm{T}_n}g(t))<a+b\}\supset\{\sup_{\mathrm{T}_n}g(t)>\mathbb{E}(\sup_{\mathrm{T}_n}g(t))-a\}\cap\{\sup_{\mathrm{T}}g(t)<b\}.$$

From (7.12), for a large enough (and not depending on n), the first event has probability 3/4, and the second has probability at least 3/4 for b large enough, as $\sup_{\mathrm{T}}g(t)<\infty$ almost surely. $\mathbb{E}(\sup_{\mathrm{T}_n}g(t))$ is shorter than $a+b$ with probability 1/2, thus with probability 1. Now, taking increasing subsets $\mathrm{T}_n\to\mathrm{T}$ (possible, as T is countable), by monotone convergence $\mathbb{E}(\sup_{\mathrm{T}}g(t))<\infty$.

To conclude, by dominated convergence, for an increasing sequence of subsets $\mathrm{T}_n\to\mathrm{T}$ and $r>0$, by (7.12)

$$\mathbb{P}(\sup_{\mathrm{T}}g(t)-\mathbb{E}(\sup_{\mathrm{T}}g(t))>r)=\lim_{n\to\infty}\mathbb{P}(\sup_{\mathrm{T}_n}g(t)-\mathbb{E}(\sup_{\mathrm{T}}g(t))>r)\leqslant e^{-\frac{\delta^2}{2\sigma^2}}$$

where $\delta=r+|\,\mathbb{E}(\sup_{\mathrm{T}}g(t))-\mathbb{E}(\sup_{\mathrm{T}_n}g(t))|$. For $n\to\infty$, $\delta\to r$ as previously seen, by monotone convergence. This concludes the proof, applying the same method for the lower bound. □

Theorem 7.12 can be applied to Gaussian processes indexed by an interval under the continuity assumption (continuity obtained e.g. by Kolmogorov's regularity criterion). Note that for the Brownian motion on $[0,1]$, $\sigma=1$ and $\sup g(t)\overset{\mathrm{law}}{=}|\mathrm{X}|$, $\mathrm{X}\sim\mathcal{N}(0,1)$, so

$$\log\mathbb{P}\left(|\sup g(t)-\mathbb{E}(\sup g(t))|>\delta\right)\underset{\delta\to\infty}{\sim}-\frac{\delta^2}{2},$$

coherent with the Theorem 7.12. However, for more general covariance structures, this theorem is one of the very few general bounds on the law of extremes. It yields the exact first order queuing probability for the maximum :

$$\log\mathbb{P}\left(\sup g(t)>\delta\right)\underset{\delta\to\infty}{\sim}-\frac{\delta^2}{2\sigma^2}.$$

To prove the above estimate, of large deviations type, the upper bound is a straightforward consequence of Theorem 7.12, and the lower bound follows from

$$\log\mathbb{P}(\sup g(t)>\delta)\geqslant\log\mathbb{P}(g(s)>\delta)\underset{\delta\to\infty}{\sim}-\frac{\delta^2}{2\sigma_s^2},$$

noting $\sigma_s=\mathrm{var}(g(s))$, and considering the supremum over the σ_s's.

2.2. Convex potentials

Theorem 7.13 (Bakry, Emery). *Assume that the potential* $\mathrm{V}:\mathbb{R}^n\to\mathbb{R}$ *is strictly convex, in the sense that for some* $c>0$

$$\mathrm{Hess}(\mathrm{V})\geqslant\frac{1}{c}\mathrm{Id},\tag{7.13}$$

for the partial order of symmetric operators. Then the probability measure

$$\mu=\frac{1}{\mathrm{Z}_n}e^{-\mathrm{V}(x_1,\dots,x_n)}\mathrm{d}x_1\dots\mathrm{d}x_n$$

satisfies the logarithmic Sobolev inequality with constant $2c$*, i.e. for any differentiable function* $f:\mathbb{R}^n\to\mathbb{R}_+^*$

$$\int f\log f\mathrm{d}\mu-\left(\int f\mathrm{d}\mu\right)\left(\log\int f\mathrm{d}\mu\right)\leqslant 2c\int|\nabla\sqrt{f}|^2\mathrm{d}\mu.$$

Proof. The idea consists in integrating along the semigroup $(\mathrm{P}_t)_{t\geqslant 0}$ associated with the invariant measure $\frac{e^{-\mathrm{V}}}{\mathrm{Z}_n}$:

$$\int f\log f\mathrm{d}\mu-\int f\mathrm{d}\mu\log\int f\mathrm{d}\mu=\mathrm{F}(0)-\mathrm{F}(\infty)$$

where $\mathrm{F}(t)=\int\mathrm{P}_t f\log\mathrm{P}_t f\mathrm{d}\mu$. That $\mathrm{F}(\infty)=\int f\mathrm{d}\mu\log\int f\mathrm{d}\mu$ follows from the ergodicity result Lemma 7.14 (we can assume that f has bounded derivatives and that its image is included in a compact of $\mathbb{R}_+^*$ and conclude by dominated convergence).

We need to control the derivative

$$\mathrm{F}'(t)=\int\left(\left(\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}_t f\right)\log\mathrm{P}_t f+\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}_t f\right)\mathrm{d}\mu.$$

The second term makes no contribution because $\int\int_0^\infty\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}_t f\mathrm{d}\mu\mathrm{d}t=\int(f-\int f\mathrm{d}\mu)\mathrm{d}\mu=0$, still by ergodicity. The first term is

$$\int(\mathrm{L}\,\mathrm{P}_t f)\log\mathrm{P}_t f\mathrm{d}\mu=-\int\Gamma_1(\mathrm{P}_t f,\log\mathrm{P}_t f)\mathrm{d}\mu=-\int\frac{\Gamma_1(\mathrm{P}_t f)}{\mathrm{P}_t f}\mathrm{d}\mu,$$

where we used $\int f(\mathrm{L}\,g)=-\int\Gamma_1(f,g)$ in the first equality. From Lemma 7.15, this is greater than

$$-e^{-\frac{2t}{c}}\int\frac{(\mathrm{P}_t\sqrt{\Gamma_1(f)})^2}{\mathrm{P}_t f}\mathrm{d}\mu.$$

Now, by the Cauchy-Schwarz inequality, for any positive functions a and b,

$$\frac{(\mathrm{P}_t a)^2}{\mathrm{P}_t b}=\frac{\left(\mathrm{P}_t\left(\frac{a}{\sqrt{b}}\sqrt{b}\right)\right)^2}{\mathrm{P}_t b}\leqslant\mathrm{P}_t\left(\frac{a^2}{b}\right),$$

so finally

$$\mathrm{F}'(t)\geqslant -e^{-\frac{2t}{c}}\int\mathrm{P}_t\left(\frac{\Gamma_1(f)}{f}\right)\mathrm{d}\mu=-e^{-\frac{2t}{c}}\int\frac{\Gamma_1(f)}{f}\mathrm{d}\mu.$$

Integrating yields $\mathrm{F}(0)-\mathrm{F}(\infty)\leqslant 2c\int|\nabla\sqrt{f}|^2\mathrm{d}\mu$, as expected. □

Lemma 7.14. *If the curvature condition (7.13) holds, then the semigroup* $(\mathrm{P}_t)_{t\geqslant 0}$ *is ergodic, in the sense that for any bounded Lipschitz* $f\in\mathscr{D}$*, almost surely (for the Lebesgue measure)*

$$\mathrm{P}_t f\underset{t\to\infty}{\longrightarrow}\int f\mathrm{d}\mu.$$

Proof. For given initial data and $t > 0$, let

$$\varphi(s) = \mathrm{P}_s(\Gamma_1(\mathrm{P}_{t-s} f)).$$

Then

$$\varphi'(s) = \mathrm{P}_s(\mathrm{L}\,\Gamma_1(\mathrm{P}_{t-s} f)) + 2\,\mathrm{P}_s(\Gamma_1(\mathrm{P}_{t-s} f, \mathrm{L}\,\mathrm{P}_{t-s} f)) = 2\,\mathrm{P}_s\,\Gamma_2(\mathrm{P}_{t-s} f). \qquad (7.14)$$

By the curvature condition (7.13), this is bounded by $\frac{2}{c}\,\mathrm{P}_s\,\Gamma_1(\mathrm{P}_{t-s} f)$, hence $\varphi'(s) \leqslant \frac{2}{c}\varphi(s)$, so by Grönwall's lemma $\varphi(t) \geqslant \varphi(0)e^{\frac{2}{c}t}$, i.e.

$$\Gamma_1(\mathrm{P}_t f) \leqslant e^{-\frac{2}{c}t}\,\mathrm{P}_t\,\Gamma_1(f). \qquad (7.15)$$

As f is Lipschitz, $\Gamma_1(f)$ is uniformly bounded in space, so $\mathrm{P}_t\,\Gamma_1(f)$ is uniformly bounded in space and time, so $\Gamma_1(\mathrm{P}_t f)$ converges uniformly to 0 in $\mathbb{R}^n$. Hence $\mathrm{P}_t f$ converges to a constant, uniformly on compact sets ($|\mathrm{P}_t f(x) - \mathrm{P}_t f(y)| \leqslant |\sqrt{\Gamma_1(\mathrm{P}_t f)}|_\infty |x-y|$). Moreover, for any $\varepsilon > 0$ and $x \in \mathbb{R}^n$ there is a compact $\mathrm{K} \subset \mathbb{R}^n$ such that $x \in \mathrm{K}$, $\mu(\mathrm{K}) > 1 - \varepsilon$, so

$$|\mathrm{P}_t f(x) - \int f \mathrm{d}\mu| \leqslant \int |\mathrm{P}_t f(x) - \mathrm{P}_t f(y)| \mathrm{d}\mu(y) \leqslant 2\varepsilon|f|_\infty + \int_{\mathrm{K}} |\mathrm{P}_t f(x) - \mathrm{P}_t f(y)| \mathrm{d}\mu(y),$$

and the last integral goes to 0 thanks to the uniform convergence on compacts. This concludes the proof. □

Lemma 7.15. *If the curvature condition (7.13) holds, then for all $f \in \mathbb{D}$*

$$\sqrt{\Gamma_1(\mathrm{P}_t f)} \leqslant e^{-\frac{t}{c}}\,\mathrm{P}_t\left(\sqrt{\Gamma_1(f)}\right).$$

Proof. Let

$$\psi(s) = e^{-\frac{s}{c}}\,\mathrm{P}_s\sqrt{\mathrm{P}_{t-s} f}.$$

Then, noting $g = \mathrm{P}_{t-s} f$,

$$\psi'(s) = -\frac{1}{c}\psi(s) + e^{-\frac{s}{c}}\,\mathrm{P}_s\left(\mathrm{L}\sqrt{\Gamma_1(g)}\right) - e^{-\frac{s}{c}}\,\mathrm{P}_s\left(\frac{\Gamma_1(g, \mathrm{L}\,g)}{\sqrt{\Gamma_1(g)}}\right).$$

Moreover, for $\mathrm{F} : \mathbb{R}^n \to \mathbb{R}$ and $f = (f_1, \dots, f_n)$,

$$\mathrm{L}\,\mathrm{F}(f) = \sum(\partial_i \mathrm{F}) f_i + \sum_{i,j}(\partial_{ij}\mathrm{F})(f)\Gamma_1(f_i, f_j),$$

so

$$\mathrm{L}\sqrt{\Gamma_1(g)} = \frac{\mathrm{L}\,\Gamma_1(g)}{2\sqrt{\Gamma_1(g)}} - \frac{\Gamma_1(\Gamma_1(g))}{4\Gamma_1(g)^{3/2}}.$$

This yields

$$\begin{aligned}\psi'(s) &= e^{-\frac{s}{c}}\,\mathrm{P}_s\left(\frac{\mathrm{L}\,\Gamma_1(g) - 2\Gamma_1(g, \mathrm{L}\,g)}{2\sqrt{\Gamma_1(g)}} - \frac{\Gamma_1(\Gamma_1(g))}{4\Gamma_1(g)^{3/2}} - \frac{1}{c}\sqrt{\Gamma_1(g)}\right) \\ &= e^{-\frac{s}{c}}\,\mathrm{P}_s\left(\frac{4\Gamma_1(g)\left(\Gamma_2(g) - \frac{1}{c}\Gamma_1(g)\right) - \Gamma_1(\Gamma(1)(g))}{4\Gamma_1(g)^{3/2}}\right).\end{aligned}$$

This is positive thanks to Lemma 7.16, so $\psi(t) \geqslant \psi(0)$, as expected. □

Lemma 7.16. *If the curvature condition (7.13) holds, for any $f \in \mathbb{D}$,*

$$4\Gamma_1(f)\left(\Gamma_2(f) - \frac{1}{c}\Gamma_1(f)\right) \geqslant \Gamma_1(\Gamma_1(f)).$$

Proof. An elementary (and long) differentiation yields, for $\mathrm{F} : \mathbb{R}^n \to \mathbb{R}$, $f = (f_1, \dots, f_n)$,

$$\begin{aligned}\Gamma_2(\mathrm{F}(f)) = \sum_{i,j} \partial_i \mathrm{F}(f) \partial_j \mathrm{F}(f) \Gamma_2(f_i, f_j) + 2 \sum_{i,j,k} \partial_i \mathrm{F}(f) \partial_{jk} \mathrm{F}(f) \mathrm{H}(f_i)(f_j, f_k) \\ + \sum_{i,j,k,l} \partial_{ij} \mathrm{F}(f) \partial_{kl} \mathrm{F}(f) \Gamma_1(f_i, f_k) \Gamma_1(f_j, f_l)\end{aligned}$$

where $\mathrm{H}(f)(g,h) = \frac{1}{2}\left(\Gamma_1(g, \Gamma_1(f,h)) + \Gamma_1(h, \Gamma_1(f,g)) - \Gamma_1(f, \Gamma_1(g,h))\right)$. Applying this to F such that, at point $(f,g)(x_0)$,

$$\nabla \mathrm{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathrm{Hess}\, \mathrm{F} = \begin{pmatrix} 0 & x \\ x & 0 \end{pmatrix},$$

we get

$$\begin{aligned}\Gamma_2(\mathrm{F}(f,g)) = \Gamma_2(f) + 4x \mathrm{H}(f)(f,g) + 2x^2 (\Gamma_1(f,g)^2 - \Gamma_1(f)\Gamma_1(g)) \\ \geqslant \frac{1}{c} \Gamma_1(\mathrm{F}(f,g)) = \frac{1}{c} \Gamma_1(f).\end{aligned}$$

The discriminant of this positive binomial needs to be negative, which implies, (by using $\mathrm{H}(f)(f,g) = \Gamma_1(\Gamma_1(f), g)$ and $\Gamma_1(f,g)^2 \leqslant \Gamma_1(f)\Gamma_1(g)$)

$$\Gamma_1(g, \Gamma_1(f))^2 \leqslant 4 \left(\Gamma_2(f) - \frac{1}{c} \Gamma_1(f)\right) \Gamma_1(f) \Gamma_1(g),$$

and the expected result for $g = \Gamma_1(f)$. □

3. Concentration on curved spaces

Theorem 7.13 admits a Riemannian analogue, based on the same ideas, and which only requires some basis on differential geometry.

Consider $\mathscr{M}$ a compact manifold of dimension n with a Riemannian metric g. Then g_x is a positive definite bilinear map on the tangent space of $\mathscr{M}$ at point x, $\mathbb{T}_x(\mathscr{M})$. The natural underlying measure, μ, associated with $(\mathscr{M}, g)$ is the locally $\mathrm{d}\mu(x) = \sqrt{\det g_x} \mathrm{d}x$. We are interested in whether smooth modifications of this volume measure satisfy a logarithmic Sobolev inequality. Let

$$\mu_{\mathrm{V}}(\mathrm{d}x) = \frac{1}{\mathrm{Z}} e^{-\mathrm{V}(x)} \mathrm{d}\mu(x)$$

with V a smooth function on $\mathscr{M}$. A stochastic process with invariant measure μ_{V} is given through its semigroup

$$\mathrm{P}_t = e^{t\, \mathrm{L}_{\mathrm{V}}},\ \mathrm{L}_{\mathrm{V}}(f) = \Delta f - g(\nabla \mathrm{V}, \nabla f),$$

where Δ is the Laplace Beltrami operator on $(\mathscr{M}, g)$, i.e. $\Delta f = \mathrm{div} \nabla f$ with $\mathrm{div} \mathrm{X} = \sum_i g(\mathrm{L}_i, [\mathrm{L}_i, \mathrm{X}])$ in local coordinates[12]. Equivalently,

$$\int g(\nabla f, \mathrm{X}) \mathrm{d}\mu = -\int f \mathrm{div} \mathrm{X} \mathrm{d}\mu.$$

Finally, the Riemannian analogues of Γ_1 and Γ_2 are naturally given by

$$\begin{aligned}\Gamma_1(f,h) &= g(\nabla f, \nabla h), \\ \Gamma_n(f,h) &= \frac{1}{2}\left(\mathrm{L}_{\mathrm{V}}\, \Gamma_{n-1}(f,h) - \Gamma_{n-1}(\mathrm{L}_{\mathrm{V}}\, f, h) - \Gamma_{n-1}(f, \mathrm{L}_{\mathrm{V}}\, h)\right).\end{aligned}$$

Additionally to the Hessian of V, a Bakry-Emery type criterium needs to take into account the shape, curvature, of $\mathscr{M}$. This is expressed through the Ricci tensor :

12. $(\mathrm{L}_i)_{1 \leqslant i \leqslant n}$ is an orthogonal frame, for the metric g, at point x, and the Lie bracket of two vector fields, noted $[\mathrm{X}, \mathrm{Y}] = \mathrm{XY} - \mathrm{YX}$, is defined as $[\mathrm{X}, \mathrm{Y}]_i = \sum_{j=1}^n (\mathrm{X}_j \partial_j \mathrm{Y}_i - \mathrm{Y}_j \partial_j \mathrm{X}_i)$

- A connection is a bilinear operation associating to vector fields X and Y a third one, $\nabla_{\mathrm{X}}\mathrm{Y}$, such that for any smooth f
$$\nabla_{f\mathrm{X}}\mathrm{Y} = f\nabla_{\mathrm{X}}\mathrm{Y},\ \nabla_{\mathrm{X}}(f\mathrm{Y}) = f\nabla_{\mathrm{X}}\mathrm{Y} + (\mathrm{X}f)\mathrm{Y}.$$
The Levi-Civita connection is the unique torsion-free connection ($\nabla_{\mathrm{X}}\mathrm{Y}-\nabla_{\mathrm{Y}}\mathrm{X} = [\mathrm{X},\mathrm{Y}]$) such that $\mathrm{X}g(\mathrm{Y},\mathrm{Z}) = g(\nabla_{\mathrm{X}}\mathrm{Y},\mathrm{Z}) + g(\mathrm{Y},\nabla_{\mathrm{X}}\mathrm{Z})$.
- The Riemann curvature tensor $\mathrm{R}(\cdot,\cdot)$ associates to vector fields X, Y an operator on vector fields $\mathrm{R}(\mathrm{X},\mathrm{Y})$ defined by (∇ is the Levi-Civita connection)
$$\mathrm{R}(\mathrm{X},\mathrm{Y})\mathrm{Z} = \nabla_{\mathrm{X}}(\nabla_{\mathrm{Y}}\mathrm{Z}) - \nabla_{\mathrm{Y}}(\nabla_{\mathrm{X}}\mathrm{Z}) - \nabla_{[\mathrm{X},\mathrm{Y}]}\mathrm{Z}.$$
- The Ricci curvature tensor associates to vector fields X and Y the function $\mathrm{Ric}(\mathrm{X},\mathrm{Y})$, smooth on $\mathscr{M}$, such that for any orthogonal local frame (L_i)
$$\mathrm{Ric}(\mathrm{X},\mathrm{Y})(x) = \sum_{i=1}^{n} g_x(\mathrm{R}(\mathrm{X},\mathrm{L}_i)\mathrm{L}_i,\mathrm{Y}).$$
In more intuitive words, the Ricci cirvature measures the first non-trivial order of difference between the Riemannian metric and its Eucliden approximation : in loal geodesic coordinates,
$$\mathrm{d}\mu = \left(1 - \frac{1}{6}\mathrm{R}_{jk}x^j x^k\right)\mathrm{d}\mu_{\mathrm{Euclidean}}.$$

The iterated carré du champ operator Γ_2 encodes both the Hessian of V and the curvature, as shown by the Bochner formula, which is a key step in the next theorem :
$$\Gamma_2(f,f) = \langle \mathrm{Hess}\, f, \mathrm{Hess}\, f\rangle + (\mathrm{Ric} + \mathrm{Hess}\,\mathrm{V})(\nabla f,\nabla f),$$
where $\langle \mathrm{Hess}\, f, \mathrm{Hess}\, h\rangle = \sum_{i,j}(\mathrm{Hess}\, f)(\mathrm{L}_i,\mathrm{L}_j)(\mathrm{Hess}\, h)(\mathrm{L}_i,\mathrm{L}_j)$.

Proof of Theorem 7.1. We closely follow the dynamical method of Ledoux [11], in the same vein as the proof of the Bakry-Emery Thorem 7.13.

First, it is sufficient to show that, for some $\mathrm{C}, c > 0$, for any F with mean $\mathbb{E}_\mu(\mathrm{F}) = 0$,
$$\mathbb{E}_\mu\left(e^{\lambda\mathrm{F}}\right) \leqslant \mathrm{C}e^{c\frac{\lambda^2}{\mathrm{R}\|\mathrm{F}\|_{\mathcal{L}}^2}}. \tag{7.16}$$
Indeed, like in the proof of the Herst Lemma 7.9, using the Bienaymé-Chebyshev inequality,
$$\mu\left\{|\mathrm{F} - \mathbb{E}_\mu(\mathrm{F}) > u|\right\} \leqslant \mathrm{C}e^{-\lambda u + c\frac{\lambda^2}{\mathrm{R}\|\mathrm{F}\|_{\mathcal{L}}^2}},$$
and minimizing over λ (take $\lambda = \frac{\mathrm{R}u}{2c\|\mathrm{F}\|_{\mathcal{L}}^2}$) yields the result.

Now, to prove (7.16), we use the Bochner formula, which relates the Ricci curvature to the Laplace-Beltrami operator, and can be written as
$$\frac{1}{2}\Delta(|\nabla\mathrm{F}|^2) - \nabla\mathrm{F}\cdot\nabla(\Delta\mathrm{F}) = \mathrm{Ric}(\nabla\mathrm{F},\nabla\mathrm{F}) + \|\mathrm{Hess}\,\mathrm{F}\|_{\mathrm{HS}};$$
Consider the Brownian motion on $\mathscr{M}$, with associated semigroup $\mathrm{P}_t = e^{t\Delta}, t \geqslant 0$.

Note $\varphi(s) = \mathrm{P}_s(|\nabla(\mathrm{P}_{t-s}f)|^2)$. Then, in the same way as (7.14),
$$\varphi'(s) = \mathrm{P}_s(\Delta\Gamma_1(\mathrm{P}_{t-s}\,\mathrm{F})) - 2\,\mathrm{P}_s(\Gamma_1(\mathrm{P}_{t-s}\,\mathrm{F},\Delta\,\mathrm{P}_{t-s}\,\mathrm{F})) \geqslant 2\mathrm{R}\varphi(s),$$
from the Bochner formula. Hence
$$|\nabla(\mathrm{P}_t\,\mathrm{F})|^2 \leqslant e^{-2\mathrm{R}t}\,\mathrm{P}_t(|\nabla\mathrm{F}|^2). \tag{7.17}$$

This estimate will be important in the end of the proof. Let $\psi(t)=\mathbb{E}_\mu(e^{\lambda\,\mathrm{P}_t\,\mathrm{F}})$. We want to show that $\Psi(0)\leqslant \mathrm{C}e^{\mathrm{C}\lambda^2/\mathrm{R}}$. Note that

$$\psi'(t)=\lambda\,\mathbb{E}_\mu(\Delta(\mathrm{P}_t\,\mathrm{F})e^{\lambda\,\mathrm{P}_t\,\mathrm{F}})=-\lambda^2\,\mathbb{E}_\mu(|\nabla(\mathrm{P}_t\mathrm{F})|^2e^{\lambda\,\mathrm{P}_t\,\mathrm{F}})\geqslant-\lambda^2e^{-2\mathrm{R}t}\psi(t).$$

This easily implies

$$\log\psi(0)-\log\psi(\infty)\leqslant\int_0^\infty\lambda^2e^{2\mathrm{R}t}\mathrm{d}t=\frac{\lambda^2}{2\mathrm{R}},$$

which concludes the proof, noting that by ergodicity $\psi(\infty)=1$. □

Theorem 7.17. *Suppose that for any $p\in\mathrm{M}$*

$$\mathrm{Ric}_p+(\mathrm{Hess}\,\mathrm{V})_p\geqslant\frac{1}{cg_p},$$

in the sense of partial order of positive operators. Then μ_V satisfies a logarithmic Sobolev inequality with constant $2c$. In particular, for any Lipschitz function F (with respect to the geodesic distance on $\mathscr{M}$),

$$\mathbb{P}_{\mu_\mathrm{V}}\left(|\mathrm{F}-\mathbb{E}_\mu\,\mathrm{F}|>\delta\right)\leqslant 2e^{-\frac{\delta^2}{c\|\mathrm{F}\|_{\mathcal{L}}^2}}$$

4. Concentration and Boolean functions

This section aims to give an idea about why hypercontractivity is an important tool to quantify influences in discrete complex systems. This applies for example to theoretical computer science and percolation theory (see [5] for a much more complete analysis).

Let μ be the uniform measure on $\mathrm{E}=\{-1,1\}^d$. For $\mathrm{S}\subset[\![1,d]\!]$ and $x\in\mathrm{E}$, let $\chi_\mathrm{S}(x)=\prod_{k\in\mathrm{E}}x_k$. These functions are a Fourier basis on $\mathrm{L}^2(\mathrm{E},\mu)$: for any $f:\mathrm{E}\to\mathbb{R}$, $f(x)=\sum_{\mathrm{S}\subset[\![1,d]\!]}\hat f(\mathrm{S})\chi_\mathrm{S}(x)$, where $\hat f(\mathrm{S})=\mathbb{E}(f(x)\chi_\mathrm{S}(x))$.

For the Gaussian measure, we first proved hypercontractivity and then the logarithmic Sobolev inequality ; the steps will be reversed in the following, concerning the uniform measure on the hypercube.

Given a function $f:\mathrm{E}\to\mathbb{R}$, let $f_{-k}(x)=f(x_1,\dots,x_{k-1},-x_k,x_{k+1},\dots,x_d)$. We define the differential operator as

$$\mathrm{D}_kf(x)=\frac{f(x)-f_{-k}(x)}{2}.$$

Theorem 7.18. *The measure μ satisfies the logarithmic Sobolev inequality with constant 2, in the sense that for any $f:\mathrm{E}\to\mathbb{R}$*

$$\mathrm{Ent}_\mu(f^2)\leqslant 2\int_\mathrm{E}\sum_{k=1}^d|\mathrm{D}_kf|^2\mathrm{d}\mu.$$

Proof. We first prove it for $d=1$, which means that for any real a and b

$$\frac12\left(a^2\log(a^2)+b^2\log(b^2)\right)-\frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2}\leqslant\frac{(a-b)^2}{2}.$$

Assume $a>b$. Writing $c=(a+b)/2$ and $x=(a-b)/2$. We need to prove that

$$\varphi(x)=(c^2+x^2)\log(c^2+x^2)+2x^2-(c+x)^2\log(c+x)-(c-x)^2\log(c-x)$$

is positive. As $\varphi(0)=0$, this will follow if $\varphi'\geqslant0$ on $[0,c)$. A calculation yields

$$\frac12\varphi'(x)=2x+2x\log(c^2+x^2)-(c+x)\log(c+x)-(x-c)\log(c-x),$$

which is 0 at $x = 0$, so we just need $\varphi'' \geqslant 0$ on $[0, c)$. Annyother calculation gives

$$\frac{1}{2}\varphi''(x) = \frac{2x^2}{c^2 + x^2} + \log\left(1 + \frac{2x^2}{c^2 - x^2}\right) \geqslant 0,$$

concluding the proof when $d = 1$. The general case follows from the tensorization property, Proposition 7.8. □

The above result implies hypercontractivity for the following semigroup. Given a function $f : \mathrm{E} \to \mathbb{R}$, let $\mathrm{T}_\rho f(x)$ be the expectation of $f(y)$ where y is obtained from x by flipping each entry of x with probability $\frac{1-\rho}{2}$:

$$\mathrm{T}_\rho f(x) = \sum_{y\in\mathrm{E}} f(y) \prod_{k=1}^{d} \left(\frac{1+\rho}{2}\mathbb{1}_{y_k=x_k} + \frac{1-\rho}{2}\mathbb{1}_{y_k=-x_k}\right) = \mathbb{E}(f(xu)\prod_{k=1}^{d}(1+\rho u_k)).$$

In particular,

$$\mathrm{T}_\rho \chi_{\mathrm{S}}(x) = \mathbb{E}(\chi_{\mathrm{S}}(xu)\prod_{k=1}^{d}(1+\rho u_k)) = \chi_{\mathrm{S}}(x)\,\mathbb{E}(\chi_{\mathrm{S}}(u)\prod_{k=1}^{d}(1+\rho u_k)) = \rho^{|\mathrm{S}|}\chi_{\mathrm{S}}(x).$$

As the χ_{S}'s are a basis of $\mathrm{L}^2(\mathrm{E}, \mu)$, this proves that $(\mathrm{T}_\rho, 0 \leqslant \rho \leqslant 1)$ is a semigroup (note that here $\mathrm{T}_1 = \mathrm{Id}$: ρ needs to be thought of as e^{-t} where t is a time) : $\mathrm{T}_{\rho_1} \star \mathrm{T}_{\rho_1} = \mathrm{T}_{\rho_1\rho_2}$.

The semigroup T is contractive in L^p for $p \geqslant 1$, as an easy consequence of Jensen's inequality. Like the Ornstein-Uhlenbeck semigroup, thanks to its smoothing property it is even hypercontractive.

Theorem 7.19. *For any $\rho \in [0, 1]$ and $1 \leqslant p \leqslant q$ satisfying $q \leqslant 1 + \rho^{-2}(p-1)$,*

$$\|\mathrm{T}_\rho f\|_{\mathrm{L}^q} \leqslant \|f\|_{\mathrm{L}^p}.$$

Proof. We proceed like in Gross's Theorem 7.11, showing that the logarithmic Sobolev inequality Theorem 7.18 implies hypercontractivity for $\mathrm{P}_t = \mathrm{T}_{e^{-t}}$. For this, we need to identify the generator of the semigroup P_t. This is easy when looking at the base functions :

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}_t\chi_s = \frac{\mathrm{d}}{\mathrm{d}t}e^{-t|\mathrm{S}|}\chi_{\mathrm{S}} = -|\mathrm{S}|\mathrm{P}_t\chi_s,$$

so $\mathrm{L}f = \mathrm{L}(\sum \hat{f}_{\mathrm{S}}\chi_{\mathrm{S}}) = -\sum|\mathrm{S}|\hat{f}_{\mathrm{S}}\chi_{\mathrm{S}}$. In particular,

$$-\mathbb{E}(f\mathrm{L}f) = \mathbb{E}\left(\left(\sum \hat{f}_{\mathrm{S}}\chi_{\mathrm{S}}\right)\left(\sum|\mathrm{S}|\hat{f}_{\mathrm{S}}\chi_{\mathrm{S}}\right)\right) = \sum|\mathrm{S}|\hat{f}_{\mathrm{S}}^2.$$

We need to prove that this coincides with $\mathbb{E}\left(\sum_k|\mathrm{D}_k f|^2\right)$. For this, note that

$$\mathrm{D}_k\chi_{\mathrm{S}} = \mathbb{1}_{k\in\mathrm{S}}\chi_{\mathrm{S}},$$

so

$$\mathbb{E}\left(\sum_k |\mathrm{D}_k f|^2\right) = \sum_k\sum_{\mathrm{S}}\mathbb{1}_{k\in\mathrm{S}}\hat{f}_{\mathrm{S}}^2 = \sum|\mathrm{S}|\hat{f}_{\mathrm{S}}^2,$$

as expected. □

We are now in a position to justify Theorem 7.3. The function f now takes values in $\{-1, 1\}$, and the influence of the kth individual is, for any $p > 0$,

$$\mathrm{infl}_k(f) = \mathbb{P}(\mathrm{D}_k f \neq 0) = \mathbb{E}(|\mathrm{D}_k f|^p). \tag{7.18}$$

In particular, the Dirichlet form is also

$$\sum_k \operatorname{infl}_k(f) = \sum_{\mathrm{S}} |\mathrm{S}|\hat{f}_{\mathrm{S}}^2 \geqslant \sum_{\mathrm{S}\neq\varnothing} \hat{f}_{\mathrm{S}}^2 = \operatorname{var}(f) \tag{7.19}$$

hence there are influences of order $\operatorname{var}(f)/d$. We prove the following slightly weaker version of Theorem 7.3, from [8]. The fact that there are influences of greater order ($\operatorname{var}(f)\log d/d$) is a consequence of both following incompatible heuristics.

- If all influences do not exceed this $\operatorname{var}(f)/d$ estimate, all $|\hat{f}_{\mathrm{S}}|$ need to be small for large $|\mathrm{S}|$: the Fourier coefficients are concentrated on low frequencies.
- If the Fourier coefficients of f are all concentrated on low frequencies the operator T_ρ does not change much f (remember that $\mathrm{T}_\rho\chi_{\mathrm{S}} = \rho^{|\mathrm{S}|}\chi_{\mathrm{S}}$), hence $\mathrm{D}_k f$. But it still improves integrability by hypercontractivity :

  $$\operatorname{infl}_k(f) = \mathbb{E}(|\mathrm{D}_k f|^2) \approx \mathbb{E}(|\mathrm{T}_\rho \mathrm{D}_k f|^2) \leqslant \operatorname{infl}_k(f)^{2/p}$$

  up to compatibility between p and ρ. This is not possible if $1 < p < 2$ as the influence is in $[0,1]$.

This explanation, is particular the $\approx$ step, is made rigorous hereafter.

Theorem 7.20. *There exists a universal constant $c > 0$ such that for any d and $f : \mathrm{E} \to \{-1,1\}$ there exists $1 \leqslant k \leqslant d$ with*

$$\operatorname{infl}_k(f) \geqslant c\frac{\log d}{d}.$$

Proof. We know that

$$\sum_{\mathrm{S}} \hat{f}_{\mathrm{S}}^2 = 1.$$

In the above sum, if we make the distinction between summands with $|\mathrm{S}| \leqslant m$ and $|\mathrm{S}| > m$ for some $1 \leqslant m \leqslant d$, we can first bound

$$\sum_{|\mathrm{S}|\geqslant m} \hat{f}_{\mathrm{S}}^2 \leqslant \frac{1}{m}\sum_k \operatorname{infl}_k(f)$$

thanks to (7.19). For the other sum, we make use of the hypercontractivity property, Theorem 7.19 in the case $1 \leqslant p \leqslant 2$, $\rho \leqslant \sqrt{p-1}$,

$$\operatorname{infl}_k(f) = \mathbb{E}(|\mathrm{D}_k f|^p) \geqslant \mathbb{E}(|\mathrm{T}_\rho \mathrm{D}_k f|^2)^{p/2}.$$

As $\mathrm{T}_\rho \mathrm{D}_k \chi_{\mathrm{S}} = \mathrm{T}_\rho(\mathbb{1}_{k\in\mathrm{S}}\chi_{\mathrm{S}}) = \mathbb{1}_{k\in\mathrm{S}}\rho^{|\mathrm{S}|}\chi_{\mathrm{S}}$, $\mathbb{E}(|\mathrm{T}_\rho \mathrm{D}_k f|^2) = \sum_{\mathrm{S}} \rho^{2|\mathrm{S}|}\hat{f}_{\mathrm{S}}^2 \mathbb{1}_{k\in\mathrm{S}}$, so

$$\sum_k \operatorname{infl}_k(f)^{2/p} \geqslant \rho^{2m} \sum_{\mathrm{S}|0<|\mathrm{S}|\leqslant m} \hat{f}(\mathrm{S})^2.$$

We finally obtain

$$\operatorname{var}(f) \leqslant \frac{1}{m}\sum_k \operatorname{infl}_k(f) + \rho^{-2m}\sum_k \operatorname{infl}_k(f)^{2/p}.$$

If, for any k, $\operatorname{infl}(f) \leqslant c\frac{\log d}{d}\operatorname{var} f$, then choosing $m = c'\log d$ yields

$$1 \leqslant \frac{c}{c'} + c^{2/p}\rho^{-2d} d\left(c\frac{\log d}{d}\right)^{2/p}.$$

Choosing $\rho = \sqrt{p-1}$, this will be impossible for large d if $c' > c$ and $-c'\log(p-1) - 2/p + 1 < 0$. Without caring about optimal coefficients, $p = 3/4$, $c' = 10^{-4}$ is appropriate. □

Note that the spectral gap inequality only yields that there are indexes k with $\operatorname{infl}_k(f) = \Omega(1/d)$. We really need the (stronger) logarithmic Sobolev inequality to prove the above theorem.

Bibliography

[1] D. Bakry, Diffusions hypercontractives, Séminaire de Probabilités XIX, 1983-84, Lecture Notes in Mathematics, 1985, Volume 1123-1985, 177-206, Springer.

[2] C. Dellacherie, P. A. Meyer, Probabilités et Potentiel, Vol I, Hermann, 1975.

[3] R. Dudley, Real analysis and probability, Cambridge University Press, 2002.

[4] N. Dunford, J. Schwartz, Linear operators, volume I, Wiley-Interscience, 1958.

[5] C. Garban, J. Steif, Noise Sensitivity of Boolean Functions and Percolation, Cambridge University Press, IMS Textbooks series, 2014.

[6] K. Itô, M. Nisio, On the convergence of sums of independent Banach space valued random variables, Osaka J. Math. Volume 5, Number 1 (1968), 35-48.

[7] J. Jacod, N. Shiryaev, Limit theorems for stochastic processes, Springer, 2002.

[8] J. Kahn, G. Kalai, N. Linial, Proceeding SFCS 1988 Proceedings of the 29th Annual Symposium on Foundations of Computer Science.

[9] O. Kallenberg, Random Measures , 4th edition. Academic Press, New York, London, Akademie-Verlag, Berlin (1986).

[10] M. Ledoux, The concentration of measure phenomenon, Mathematical Surveys and Monographs 89, American Mathematical Society, 2001.

[11] M. Ledoux, A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold, Geometric and Functional Analysis, Volume 2, Number 2, 221-224, 1992.

[12] J. F. Le Gall, Calcul stochastique et processus de Markov, Master 2 lecture notes, available on `www.math.ens.fr/~legall`.

[13] P. Morters, Y. Peres, Cambridge Series in Statistical and Probabilistic Mathematics, Volume 30.

[14] Jan Obloj, The Skorokhod embedding problem and its offspring, Probab. Surveys Volume 1 (2004), 321-392.

[15] G. Peano, Sur une courbe qui remplit toute une aire plane, Mathematische Annalen 36 (1) : 157-160, 1890.

[16] D. Revuz, M. Yor, Continuous martingales and Brownian motion, Springer, 1998.

[17] L. C. G. Rogers, D. Williams, Diffusions, Markov Processes and Martingales, Itô Calculus. Cambridge University Press, 1987.

[18] D. W. Stroock, S. R. S. Varadhan, Multidimensional diffusion processes, Springer, 1979.

[19] T. Tao, V. Vu, M. Krishnapur (appendix), Annals of Probability 38, no. 5, 2023-2065, 2010.

[20] M. Talagrand, On Russo's Approximate Zero-One Law, Ann. Probab. Volume 22, Number 3 (1994), 1576-1587.

[21] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Inst. Hautes Études Sci. Publ. Math. No. 81 (1995), 73-205.

[22] B. Tsirelson, An example of a stochastic differential equation having no strong solution, Theory Probab. Appl. 20 :2, 416-418, 1975.

[23] A. S. Ustunel, Analysis on Wiener Space and Applications, lecture notes available on the Arxiv.

[24] A. S. Ustunel, Entropy, Invertibility and Variational Calculus of the Adapted Shifts on Wiener space, Journal of Functional Analysis Volume 257, Issue 11, 1 December 2009, Pages 3655-368.

[25] M. Yor, Tsirelson's equation in discrete time, Probability Theory and Related Fields, 91(2) :135-152, 1992.

[26] A. K. Zvonkin, A transformation of the phase space of a diffusion process that will remove the drift, Mat. Sb., 152, 1974.