Introduction to stochastic calculus

Justin Salez

April 2, 2024

Contents

Disclaimer: this course is a *minimal* and *practical* introduction to the theory of stochastic calculus, with an emphasis on examples and applications rather than abstract subtleties.

Acknowledgment: Thanks are due to Josué Corujo and Damiano De Gaspari for having reported many typos in a preliminary version of these notes.

Chapter 1

Preliminaries

1.1 Stochastic processes

Stochastic processes. A stochastic process is just a collection $X = (X_t)_{t \in \mathbb{T}}$ of real-valued random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and indexed by an arbitrary set $\mathbb{T}$. Two simple choices, which should be familiar to the reader, are $\mathbb{T} = \{1, \ldots, n\}$ (random vectors) – and $\mathbb{T} = \mathbb{N}$ (random sequences). We shall here soon focus on the more involved choice $\mathbb{T} = \mathbb{R}_+$ (random functions), and interpret the parameter t as *time*: the stochastic process X may then be thought of as modeling a physical quantity which *evolves at random through time*.

Law of a process. Our collection $X = (X_t)_{t \in \mathbb{T}}$ of one-dimensional random variables can equivalently be viewed as a *single* random variable taking values in the multi-dimensional space $\mathbb{R}^{\mathbb{T}}$, equipped with the product $\sigma-$field. As any random variable, X has a well-defined law. The latter is a probability measure on $\mathbb{R}^{\mathbb{T}}$ which, by Dynkin's lemma, is uniquely characterized by the laws of the random vectors $(X_{t_1}, \ldots, X_{t_n})$ for all $n \in \mathbb{N}$ and all $(t_1, \ldots, t_n) \in \mathbb{T}^n$. In practice, specifying these finite-dimensional marginals can be very complicated, and one often restricts attention to two fundamental statistics of X: its mean and covariance.

Mean and covariance. A stochastic process X is called square-integrable if its coordinates are in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, i.e. $\mathbb{E}[X_t^2] < \infty$ for all $t \in \mathbb{T}$. By Cauchy-Schwarz, this ensures the well-definiteness of the mean $m_X \colon \mathbb{T} \to \mathbb{R}$ and covariance $\gamma_X \colon \mathbb{T}^2 \to \mathbb{R}$, given by

$$m_X(t) := \mathbb{E}[X_t], \qquad \gamma_X(s,t) := \mathrm{Cov}(X_s, X_t) = \mathbb{E}[X_s X_t] - \mathbb{E}[X_s]\mathbb{E}[X_t]. \tag{1.1}$$

Recall for future reference that the function γ_X is always symmetric in its two arguments, and positive semi-definite: for all $n \in \mathbb{N}$, all $(t_1, \ldots, t_n) \in \mathbb{T}^n$, and all $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, we have

$$\sum_{j,k=1}^{n} \sum_{k=1}^{n} \lambda_j \lambda_k \gamma_X(t_j, t_k) \;=\; \mathrm{Var}\left(\sum_{j=1}^{n} \lambda_j X_j\right) \;\geq\; 0. \tag{1.2}$$

Perhaps surprisingly, the two simple functions m_X and γ_X capture a considerable amount of structural information about the process, and play a major role in many practical aspects of signal processing and forecasting. While they are far from characterizing the law of a general square-integrable process, they do characterize it in the important case of Gaussian processes.

Gaussian processes. A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called Gaussian if every finite linear combination of its coordinates is a gaussian random variable. More explicitly, for every $n \in \mathbb{N}$, every $(t_1, \ldots, t_n) \in \mathbb{T}^n$, and every $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, the scalar random variable

$$Z \quad := \quad \lambda_1 X_{t_1} + \cdots + \lambda_n X_{t_n} \tag{1.3}$$

is a Gaussian random variable. In particular, we have

$$\mathbb{E}[e^{iZ}] \quad = \quad \exp\left\{ i\mathbb{E}[Z] - \frac{1}{2}\mathrm{Var}(Z) \right\}. \tag{1.4}$$

Note that the left-hand side is precisely the characteristic function of the random vector $(X_{t_1}, \ldots, X_{t_n})$ evaluated at the point $(\lambda_1, \ldots, \lambda_n)$: the knowledge of this quantity for every $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ suffices to determine the law of $(X_{t_1}, \ldots, X_{t_n})$. Since the right-hand side of (1.4) only depends on X through m_X and γ_X, we conclude that the law of a Gaussian process is fully determined by its mean and covariance. We will henceforth write $X \sim \mathcal{N}(m, \gamma)$ to mean that X is a Gaussian process with mean m and covariance γ. Such a process can be shown to exist for any choice of $m \colon \mathbb{T} \to \mathbb{R}$ and of the symmetric positive-definite function $\gamma \colon \mathbb{T}^2 \to \mathbb{R}$. Here are a few particularly important choices, to which we will come back a lot:

- $\mathbb{T} = \mathbb{R}_+$, $m = 0$, $\gamma(s,t) = \mathbf{1}_{(s=t)}$ (as in a white noise).
- $\mathbb{T} = \mathbb{R}_+$, $m = 0$, $\gamma(s,t) = s \wedge t$ (as in a Brownian motion).
- $\mathbb{T} = \mathbb{R}$, $m = 0$, $\gamma(s,t) = e^{-|t-s|}$ (as in an Ornstein-Uhlenbeck process).
- $\mathbb{T} = [0,1]$, $m = 0$, $\gamma(s,t) = s \wedge t - st$ (as in a Brownian bridge).

Independance. Two stochastic processes $(X_s)_{s\in\mathbb{S}}$ and $(Y_t)_{t\in\mathbb{T}}$ are independent if the random vectors $(X_{s_1}, \ldots, X_{s_n})$ and $(Y_{t_1}, \ldots, Y_{t_n})$ are independent for every $n \in \mathbb{N}$ and every choice of the indices $(s_1, \ldots, s_n) \in \mathbb{S}^n$ and $(t_1, \ldots, t_n) \in \mathbb{T}^n$. In general, this may be quite hard to check, but a huge simplification occurs when the two processes $(X_s)_{s\in\mathbb{S}}$ and $(Y_t)_{t\in\mathbb{T}}$ are jointly Gaussian, meaning that the concatenated process $((X_s)_{s\in\mathbb{S}}, (Y_t)_{t\in\mathbb{T}})$ is Gaussian. Indeed, the random vector $(X_{s_1}, \ldots, X_{s_n}, Y_{t_1}, \ldots, Y_{t_n})$ is then Gaussian, so its distribution is entirely determined by its mean and covariance. In particular, the independance between $(X_{s_1}, \ldots, X_{s_n})$ and $(Y_{t_1}, \ldots, Y_{t_n})$ reduces to the corresponding covariances being 0. Thus, two jointly Gaussian processes $X = (X_s)_{s\in\mathbb{S}}$ and $Y = (Y_t)_{t\in\mathbb{T}}$ are independent if and only if they are decorrelated in the sense that

$$\forall (s,t) \in \mathbb{S} \times \mathbb{T}, \quad \mathrm{Cov}(X_s, Y_t) = 0. \tag{1.5}$$

This simplification extends to more than two processes in the obvious way.

Indistinguishability We will say that two processes $X = (X_t)_{t\in\mathbb{T}}$ and $Y = (Y_t)_{t\in\mathbb{T}}$ are indistinguishable if the random variables X and Y coincide a.-s., i.e. if the set

$$\{\omega \in \Omega \colon \exists t \in \mathbb{T} \colon X_t \neq Y_t\}$$

is $\mathbb{P}-$negligible. In general, this is stronger than requiring that Y is a modification of X, i.e.

$$\forall t \in \mathbb{T}, \quad \mathbb{P}(X_t = Y_t) \quad = \quad 1. \tag{1.6}$$

However, the two notions coincide when $\mathbb{T}$ is countable, or when $\mathbb{T} = \mathbb{R}$ and X, Y are (right-) continuous. Note that (1.6) implies, in particular, that X and Y have the same law.

1.2 Brownian motion

We are now ready to introduce our most important stochastic process, the Brownian motion.

Definition 1.1 (Brownian motion). *A Brownian motion is a stochastic process $B = (B_t)_{t\geq 0}$ such that*

(i) B is Gaussian with mean $m_B(t) = 0$ and covariance $\gamma_B(s,t) = s \wedge t$ for all $s,t \geq 0$.

(ii) B almost-surely has continuous trajectories. More precisely, the set

$$\{\omega \in \Omega \colon \text{ the function } t \mapsto B_t(\omega) \text{ is not continuous}\} \tag{1.7}$$

is $\mathbb{P}-$negligible, which means that it lies inside an event $E \in \mathcal{F}$ such that $\mathbb{P}(E) = 0$.

The existence of this object is not obvious, and we will admit it here.

Remark 1.1 (Continuous version). *Note that one can always improve the almost-sure continuity to a pointwise continuity by re-defining B to be the zero function on the null event E above. Working with this continuous modification will be sometimes useful.*

Remark 1.2 (A subtle point). *As explained above, Condition (i) completely determines the law of B. This, however, does not determine whether Condition (ii) holds or not: there are processes satisfying (i) and (ii), and others satisfying (i) but not (ii). The reason is that the set of continuous trajectories $C^0(\mathbb{R}_+) \subseteq \mathbb{R}^{\mathbb{R}^+}$ is not in the product $\sigma-$field: in words, trajectorial continuity is too sophisticated to be expressible in terms of finite-dimensional marginals only.*

Let us now enumerate some simple properties of Brownian motion. We start with a simple description of its distribution, based on increments.

Proposition 1.1 (Increments). *Let $B = (B_t)_{t\geq 0}$ be a Brownian motion. Then,*

(i) $B_0 = 0$ almost-surely;

(ii) $B_t - B_s \sim \mathcal{N}(0, t-s)$ for all $0 \leq s \leq t$;

(iii) $B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent for any $n \in \mathbb{N}$ and any $0 \leq t_1 \leq \ldots \leq t_n$.

Conversely, any process satisfying these three properties has the law of a Brownian motion.

Proof. Since B is a Gaussian process with $m_B = 0$ and $\gamma_B(s,t) = s \wedge t$, we have $B_t \sim \mathcal{N}(0,t)$ for all $t \geq 0$. Taking $t = 0$ yields the first claim, and we now turn to the second. The fact that $B_t - B_s$ is a Gaussian random variable is clear, since B is a Gaussian process. Thus, it only remains to compute its mean and variance: by linearity of expectations and bilinearity of covariances,

$$\begin{aligned} \mathbb{E}[B_t - B_s] &= m_B(t) - m_B(s) \;=\; 0 \\ \mathrm{Var}(B_t - B_s) &= \gamma_B(t,t) + \gamma_B(s,s) - 2\gamma_B(t,s) \;=\; t - s. \end{aligned}$$

Finally, the random vector $(B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}})$ is Gaussian, because any linear combination of its coordinates is also a linear combination of coordinates of the Gaussian process B. Consequently, independance reduces to decorrelation. Now, for $1 \leq j < k \leq n$, we have

$$\begin{aligned} \mathrm{Cov}\left(B_{t_j} - B_{t_{j-1}}, B_{t_k} - B_{t_{k-1}}\right) &= \gamma_B(t_j, t_k) + \gamma_B(t_{j-1}, t_{k-1}) - \gamma_B(t_j, t_{k-1}) - \gamma_B(t_{j-1}, t_k) \\ &= t_j + t_{j-1} - t_j - t_{j-1} \\ &= 0, \end{aligned}$$

where the second line uses the fact that $t_{j-1} \leq t_j \leq t_{k-1} \leq t_k$. This establishes (iii). The converse is a good exercise, which we leave to the reader. □

We now list three elementary but important invariance properties, which confirm the robustness and canonical nature of Brownian motion.

Proposition 1.2 (Invariance). *Let $B = (B_t)_{t\geq 0}$ be a Brownian motion. Then, in each of the following cases, the process $W = (W_t)_{t\geq 0}$ is also a Brownian motion.*

(i) $W_t := B_{a+t} - B_a$ *for any fixed* $a \geq 0$ *(invariance by translation).*

(ii) $W_t := \frac{B_{at}}{\sqrt{a}}$ *for any fixed* $a > 0$ *(invariance by scaling).*

(iii) $W_t := tB_{\frac{1}{t}}\mathbf{1}_{(t>0)}$ *(invariance by time inversion).*

Proof. In each case, W is a Gaussian process because any linear combination of its coordinates is also a linear combination of coordinates of the Gaussian process B. Moreover, direct computations reveal that W has the same mean and covariance as B. From this, we can already conclude that W is distributed as a Brownian motion, but we still have to check the almost-sure continuity of $t \mapsto W_t$. The latter is clear in cases (i) and (ii), by composition of continuous functions. The same argument works in (iii), except at $t = 0$. Thus, it only remains to check that $W_t \to 0$ almost-surely as $t \to 0$. Here is a short but subtle argument: if a function $x \colon \mathbb{R}_+ \to \mathbb{R}$ is known to be continuous on $(0, \infty)$, then its convergence to 0 at $0+$ can be expressed as $x \in E$, where

$$E \quad := \quad \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{t \in [0, \frac{1}{k}] \cap \mathbb{Q}} \left\{ |x_t| \leq \frac{1}{n} \right\}.$$

Clearly, this set is in the product $\sigma-$field. Since W and B have the same law, we can safely conclude that $\mathbb{P}(W \in E) = \mathbb{P}(B \in E)$. But $\mathbb{P}(B \in E) = 1$, by the trajectorial continuity of B. □

Remark 1.3 (SLLN for the Browian motion). *The invariance (iii) as an interesting consequence: being a Brownian motion, the process* $t \mapsto tB_{\frac{1}{t}}\mathbf{1}_{(t>0)}$ *must tend to* 0 *almost-surely as* $t \to 0$*, which means that*

$$\frac{B_t}{t} \xrightarrow[t\to\infty]{a.-s.} 0. \tag{1.8}$$

This classical fact is known as the strong law of large numbers for the Brownian motion.

Finally, let us complement the invariance by translation observed above.

Proposition 1.3 (Markov property for the Brownian motion). *Let $(B_t)_{t\geq 0}$ be a Brownian motion, and let $a \geq 0$ be fixed. Then, the Brownian motion $(B_{t+a} - B_a)_{t\geq 0}$ is independent of $(B_t)_{t\in[0,a]}$.*

Proof. The processes $(B_t)_{t\in[0,a]}$ and $(B_{t+a} - B_a)_{t\geq 0}$ are jointly Gaussian, because their coordinates are linear combinations of coordinates of the same Gaussian process B. Thus, the claimed independance reduces to the decorrelation property

$$\forall (s,t) \in [0,a] \times \mathbb{R}_+, \quad \mathrm{Cov}\left(B_s, B_{t+a} - B_a\right) \quad = \quad 0, \tag{1.9}$$

which readily follows from the explicit expression of γ_B. □

1.3 Martingales

From now onward, we turn our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, by equipping it with a filtration $(\mathcal{F}_t)_{t\geq 0}$. In other words, each $\mathcal{F}_t \subseteq \mathcal{F}$ is a $\sigma-$field on Ω and $\mathcal{F}_s \subseteq \mathcal{F}_t$ for each $0 \leq s \leq t$. The intuition is that $\mathcal{F}_t$ represents the *information* that is available by time t about the various stochastic processes under consideration. For this interpretation to be valid, we shall restrict our attention to processes $X = (X_t)_{t\geq 0}$ that satisfy

$$\forall t \geq 0, \quad X_t \text{ is } \mathcal{F}_t - \text{measurable}. \tag{1.10}$$

Such processes are said to be adapted. A simple way to ensure that a given process X is adapted is to choose its natural filtration $\mathcal{F}^X = (\mathcal{F}^X_t)_{t\geq 0}$, defined by

$$\mathcal{F}^X_t \quad := \quad \sigma\left(X_s \colon s \leq t\right). \tag{1.11}$$

Of course, any larger filtration (in the coordinate-wise sense) will also work.

Definition 1.2 (Martingale). *A martingale is a stochastic process $M = (M_t)_{t\geq 0}$ such that*

(i) M is adapted, i.e. M_t is $\mathcal{F}_t-$measurable for each $t \geq 0$.

(ii) M is integrable, i.e. $\mathbb{E}[|M_t|] < \infty$ for each $t \geq 0$.

(iii) M is fair, i.e. $\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s$ for each $0 \leq s \leq t$.

Remark 1.4 (Constant mean). *In particular, the mean of a martingale is a constant function, i.e.*

$$\forall t \geq 0, \quad \mathbb{E}[M_t] \quad = \quad \mathbb{E}[M_0]. \tag{1.12}$$

Property (iii) is, however, a much deeper property: as we will see, it implies that (1.12) acutally remains valid when the deterministic time t is replaced by any "sufficiently reasonable" random time T.

Example 1.1 (Some important martingales). *Let $B = (B_t)_{t\geq 0}$ be a Brownian motion. Then, in each of the following cases, the process $(M_t)_{t\geq 0}$ is a martingale with respect to the filtration $\mathcal{F}^B$.*

(i) $M_t := B_t$

(ii) $M_t := B_t^2 - t$

(iii) $M_t := e^{\theta B_t - \frac{\theta^2 t}{2}}$, for any fixed $\theta \in \mathbb{R}$.

Proof. In each case, the process M is adapted because we have $M_t = f_t(B_t)$ for some measurable (in fact, continuous) function $f_t \colon \mathbb{R} \to \mathbb{R}$. The integrability is standard, since $B_t \sim \mathcal{N}(0, t)$. Finally, for $0 \leq s \leq t$, we may write $B_t = B_s + (B_t - B_s)$ and use the fact that B_s is $\mathcal{F}_s-$measurable while $B_t - B_s$ is independent of $\mathcal{F}_s$ (this is the Markov property for B) to obtain

$$\begin{aligned}
\mathbb{E}[B_t \mid \mathcal{F}_s] &= B_s + \mathbb{E}[B_t - B_s] \;=\; B_s \\
\mathbb{E}[B_t^2 \mid \mathcal{F}_s] &= B_s^2 + \mathbb{E}[(B_t - B_s)^2] + 2B_s\mathbb{E}[B_t - B_s] \;=\; B_s^2 + (t - s) \\
\mathbb{E}\left[e^{\theta B_t} \mid \mathcal{F}_s\right] &= e^{\theta B_s}\mathbb{E}\left[e^{\theta(B_t - B_s)}\right] \;=\; e^{\theta B_s + \frac{\theta^2}{2}(t-s)}.
\end{aligned}$$

Rearranging these identities readily gives the desired martingale property in each case. □

Remark 1.5. *The same argument works for any filtration $(\mathcal{F}_t)_{t\geq 0}$ such that B is adapted to $(\mathcal{F}_t)_{t\geq 0}$ and $B_t - B_s$ is independent of $\mathcal{F}_s$ for every $0 \leq s \leq t$. We then speak of a $(\mathcal{F}_t)_{t\geq 0}-$Brownian motion.*

Remark 1.6 (A general formula). *The above computations are special cases of a useful general formula for conditional expectations: if $\mathcal{G} \subseteq \mathcal{F}$ is any $\sigma-$field, and if X and Y are two random variables, the first being $\mathcal{G}-$measurable and the second being independent of $\mathcal{G}$, then*

$$\mathbb{E}\left[f(X,Y) \mid \mathcal{G}\right] \quad = \quad F(X), \quad \text{where} \quad F(x) \; := \; \mathbb{E}\left[f(x,Y)\right],$$

for any measurable f such that $\mathbb{E}[|f(X,Y)|] < \infty$. In particular, if B is a $(\mathcal{F}_t)_{t\geq 0}-$Brownian motion and $0 \leq s \leq t$, we may apply this to $X = B_s, Y = B_t - B_s, \mathcal{G} = \mathcal{F}_s$ and $f(x,y) = \varphi(x+y)$ to obtain

$$\mathbb{E}[\varphi(B_t) \mid \mathcal{F}_s] \quad = \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{z^2}{2}} \varphi(B_s + z\sqrt{t-s})\, \mathrm{d}z, \tag{1.13}$$

for any measurable function $\phi \colon \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[|\varphi(B_t)|] < \infty$.

As we will now see, the true strength of martingales lies in the fact that the time t in the mean conservation identity (1.12) can, under appropriate conditions, be taken to be random.

Definition 1.3 (Stopping time). *A stopping time is a $[0,\infty]-$valued random variable T such that*

$$\forall t \geq 0, \quad \{T \leq t\} \quad \in \quad \mathcal{F}_t. \tag{1.14}$$

The intuition is that, at any given time, one should be able to determine whether the random time T has already occurred or not, just by looking at the information available so far (and not in the future). For example, the first time that a $(\mathcal{F}_t)_{t\geq 0}-$Brownian motion reaches the value 1 is a stopping time, but the *last* time that a Brownian motion reaches the value 0 in the time interval $[0,1]$ is *not*. In practice, all stopping times that we shall encounter will be of the following form.

Proposition 1.4 (A useful criterion). *Suppose that $A \subseteq \mathbb{R}$ is a closed set and that $X = (X_t)_{t\geq 0}$ is an adapted, continuous process. Then, the hitting time of A by X, defined as*

$$T_A(X) \quad := \quad \inf\{t \geq 0 \colon X_t \in A\}, \tag{1.15}$$

is always a stopping time (with the usual convention $\inf \emptyset = +\infty$).

Proof. Using the continuity of X and the fact that A is closed, one can easily check that

$$\{T_A(X) \leq t\} \quad = \quad \bigcap_{k=1}^{\infty} \bigcup_{s\in[0,t]\cap\mathbb{Q}} \left\{\mathrm{dist}(X_s, A) \leq \frac{1}{k}\right\}. \tag{1.16}$$

Now, $\left\{\mathrm{dist}(X_s, A) \leq \frac{1}{k}\right\} \in \mathcal{F}_s \subseteq \mathcal{F}_t$ because X is adapted and $z \mapsto \mathrm{dist}(z, A)$ is measurable. Thus, $\{T_A(X) \leq t\} \in \mathcal{F}_t$ as a countable union and intersection of events in $\mathcal{F}_t$. □

Exercise 1.1 (Stopping times). *Show that if S, T are stopping times, then so are $S \wedge T$, $S \vee T$, $S + T$.*

We are now in position to recall the main result of martingale theory.

Theorem 1.1 (Doob's optional stopping Theorem). *If $(M_t)_{t\geq 0}$ is a continuous martingale and T a stopping time, then the stopped process $M^T := (M_{t\wedge T})_{t\geq 0}$ is a (continuous) martingale. In particular,*

$$\forall t \geq 0, \quad \mathbb{E}\left[M_{T\wedge t}\right] \quad = \quad \mathbb{E}[M_0]. \tag{1.17}$$

If $(M_{t\wedge T})_{t\geq 0}$ is uniformly integrable and $T < \infty$ a.-s., then taking $t \to \infty$ to obtain $\mathbb{E}\left[M_T\right] = \mathbb{E}[M_0]$.

Here is an example to illustrate the practical interest of Doob's optional stopping Theorem.

Example 1.2 (Exit time from an interval). *Fix two constants $a, b > 0$. How long will it take, on average, for a Brownian motion B to exit the interval $I = (-a, b)$? The variable of interest T is a stopping time, because it is the hitting time of the closed set I^c by the continuous and adapted process B. Applying Doob's optional stopping Theorem to the continuous martingale $(B_t^2 - t)_{t\geq 0}$, we deduce that*

$$\mathbb{E}[T \wedge t] \quad = \quad \mathbb{E}\left[B_{T\wedge t}^2\right],$$

for all $t \geq 0$. We now send $t \to \infty$. The left-hand side tends to $\mathbb{E}[T]$ by monotone convergence. Since the right-hand side is bounded by $(a \vee b)^2$ independently of t, we already see that $\mathbb{E}[T] \leq (a \vee b)^2$. In particular, T is a.-s. finite, and the domination $B_{T\wedge t}^2 \leq (a \vee b)^2$ now allows us to obtain the equality

$$\mathbb{E}[T] \quad = \quad \mathbb{E}[B_T^2] \quad = \quad pb^2 + (1-p)a^2,$$

where $p = \mathbb{P}(B_T = b)$. The second equality relies on the observation that B_T takes values in the two-element set $\{-a, b\}$, by continuity of B. To compute p, we now apply Doob's optional stopping theorem to the martingale $(B_t)_{t\geq 0}$. We already know that $T < \infty$ a.-s., and that $|B_{T\wedge t}| \leq a \vee b$, so we may safely conclude that $0 = \mathbb{E}[B_T] = pb - (1-p)a$, i.e. $p = a/(a+b)$. In conclusion, the answer is

$$\mathbb{E}[T] \quad = \quad \frac{ab^2}{a+b} + \frac{ba^2}{a+b} \quad = \quad ab.$$

More generally, Doob's optional stopping theorem remains true for sub-martingales or super-martingales (defined by relaxing the equality $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$ into the inequality $\mathbb{E}[M_t|\mathcal{F}_s] \geq M_s$ or $\mathbb{E}[M_t|\mathcal{F}_s] \leq M_s$, respectively). Such processes arise naturally when applying a convex or concave function to a martingale (by the conditional Jensen inequality). We end this section by mentioning a uniform refinement of Chebychev's inequality in the case of martingales.

Theorem 1.2 (Doob's maximal inequality). *If M is a square-integrable continuous martingale, then*

$$\forall a, t \geq 0, \quad \mathbb{P}\left(\sup_{s\in[0,t]} |M_s| \geq a\right) \quad \leq \quad \frac{\mathbb{E}[M_t^2]}{a^2}. \tag{1.18}$$

Here is a nice and useful application of Doob's maximal inequality.

Proposition 1.5 (Limits of continuous martingales). *Let $(M^n)_{n\geq 1}$ be continuous, square-integrable martingales, and suppose that for each $t \geq 0$ the limit $M_t := \lim_{n\to\infty} M_t^n$ exists in L^2. Then, the process $M = (M_t)_{t\geq 0}$ (has a modification which) is a continuous square-integrable martingale.*

Proof. The only real difficulty is continuity. By Doob's maximal inequality applied to the continuous square-integrable martingale $M^n - M^m$, we have for fixed $t \geq 0$ and $k \in \mathbb{N}$,

$$\mathbb{P}\left(\sup_{s\in[0,t]} |M_s^n - M_s^m| \geq \frac{1}{k^2}\right) \quad \leq \quad k^2 \mathbb{E}\left[(M_t^n - M_t^m)^2\right]. \tag{1.19}$$

Since $(M_t^n)_{n\geq 1}$ converges in L^2, the right-hand side can be made arbitrarily small by choosing $m \wedge n$ large. Consequently, there is an increasing sequence $(N_k)_{k\geq 1}$ such that

$$\forall k \in \mathbb{N}, \quad \mathbb{P}\left(\sup_{s\in[0,t]} |M_s^{N_{k+1}} - M_s^{N_k}| \geq \frac{1}{k^2}\right) \quad \leq \quad \frac{1}{k^2}. \tag{1.20}$$

By the Borel-Cantelli lemma, we deduce that almost-surely,

$$\sum_{k=1}^{\infty} \sup_{s\in[0,t]} \left|M_s^{N_{k+1}} - M_s^{N_k}\right| \quad < \quad \infty. \tag{1.21}$$

This ensures that almost-surely, the sequence $(M^{N_k})_{k\geq 1}$ is convergent in the space of continuous functions equipped with the topology of uniform convergence on every compact set. But the limit is necessarily a version of M, because for each $t \geq 0$, we have $M_t^n \to M_t$ in L^2. □

1.4 Absolute and quadratic variation

Despite being continuous, the Brownian motion – or any interesting martingale, as we will see – is extremely *rough*: it oscillates wildly. To formalize this idea, we define the absolute variation of a function $f\colon \mathbb{R}_+ \to \mathbb{R}$ on the interval $[s,t]$ as the (possibly infinite) quantity

$$V(f,s,t) \quad := \quad \sup_{(t_k)} \sum_k |f(t_k) - f(t_{k-1})|\,, \tag{1.22}$$

where the supremum is taken over all subdivisions $s = t_0 \leq t_1 \leq \ldots \leq t_n = t$ $(n \in \mathbb{N})$ of the interval $[s,t]$. Note that we have the *chain rule*

$$\forall u \in [s,t], \quad V(f,s,t) \quad = \quad V(f,s,u) + V(f,u,t).$$

The function f has finite variation if $V(f,s,t) < \infty$ for every $0 \leq s \leq t$. This is the case for most of the functions that we are used to manipulate: for example, we invite the reader to check that

(i) If f is continuously differentiable, then $V(f,s,t) = \int_s^t |f'(u)|\,\mathrm{d}u < \infty$;

(ii) If f is monotone, then $V(f,s,t) = |f(t) - f(s)| < \infty$;

(iii) $V(f+g,s,t) \leq V(f,s,t) + V(g,s,t)$.

In particular, (ii) and (iii) imply that the difference of two non-decreasing functions has finite variation. In fact, any function of finite variation is of this form.

Proposition 1.6 (Characterization of finite variation)**.** *A function $f\colon \mathbb{R}_+ \to \mathbb{R}$ has finite variation if and only if it can be written as $f = f_1 - f_2$, where $f_1, f_2\colon \mathbb{R}_+ \to \mathbb{R}$ are non-decreasing.*

Proof. The 'if' part is trivial. Conversely, if f has finite variation, then it is immediate to check that the functions $f_1\colon t \mapsto V(f,0,t)$ and $f_2\colon t \mapsto V(f,0,t) - f(t)$ are non-decreasing. □

It is now time to give an example of a function that *fails* to have finite variation.

Example 1.3 (Variation of the Brownian motion)**.** *Let B be a Brownian motion. Fix $0 \leq s \leq t$ and consider the subdivision $(t_0,\ldots,t_n)$ of $[s,t]$ into n intervals of equal length, i.e. $t_k = s + \frac{k}{n}(t-s)$. Then,*

$$\sum_{k=1}^n |B_{t_k} - B_{t_{k-1}}| \quad \stackrel{d}{=} \quad \sqrt{\frac{t-s}{n}}\,(|\xi_1| + \cdots + |\xi_n|)\,, \tag{1.23}$$

where $(\xi_k)_{k\geq 0}$ are i.i.d. with law $\mathcal{N}(0,1)$. Now, the right-hand side diverges as $n \to \infty$ by the strong law of large numbers, implying that $\mathbb{P}(V(B,s,t) = +\infty) = 1$. By taking $s,t \in \mathbb{Q}_+$ and noting that $V(f,s,t) \leq V(f,s',t')$ whenever $[s,t] \subseteq [s',t']$, we conclude that

$$\mathbb{P}\left(\forall s,t \geq 0,\ V(B,s,t) = +\infty\right) \quad = \quad 1.$$

Thus, Brownian motion oscillates much more than the typical functions that we are used to manipulate.

The computation appearing at (1.23) strongly suggests looking at *squared* increments when measuring the variations of Brownian motion. Indeed, the quadratic version of (1.23) is

$$\sum_{k=1}^n |B_{t_k} - B_{t_{k-1}}|^2 \quad \stackrel{d}{=} \quad \frac{t-s}{n}\,(|\xi_1|^2 + \cdots + |\xi_n|^2)\,, \tag{1.24}$$

and the right-hand side now tends to $t-s$ instead of $+\infty$, by the strong law of large numbers. This idea of considering *quadratic variation* when the function of interest has infinite absolute variation turns out to work way beyond the specific example of Brownian motion, as shown in the following fundamental result.

Theorem 1.3 (Quadratic variation of squared-integrable martingales). *Let $M = (M_t)_{t\geq 0}$ be a continuous, square-integrable martingale. Then, for each $t \geq 0$, the limit*

$$\langle M\rangle_t \quad := \quad \lim_{n\to\infty} \sum_{k=1}^{n} |M_{t_k^n} - M_{t_{k-1}^n}|^2$$

exists in L^1 and does not depend on the subdivisions $(t_k^n)_{0\leq k\leq n}$ of $[0,t]$, as long as the mesh $\max_{0\leq k\leq n}|t_k^n - t_{k-1}^n|$ tends to 0 as $n\to\infty$. Moreover, the process $\langle M\rangle = (\langle M\rangle_t)_{t\geq 0}$ has the following properties:

(i) $\langle M\rangle_0 = 0$;

(ii) $t\mapsto \langle M\rangle_t$ *is non-decreasing ;*

(iii) $t\mapsto \langle M\rangle_t$ *(has a modification which) is continuous ;*

(iv) $\left(M_t^2 - \langle M\rangle_t\right)_{t\geq 0}$ *is a martingale.*

Proof. Let us here admit existence and continuity, and focus on the important martingale property (iv), whose proof is instructive (Properties (i) and (ii) are trivial). Fix $0\leq s\leq t$, and consider a subdivision (t_k^n) of $[s,t]$ with $\max_{0\leq k\leq n}|t_k^n - t_{k-1}^n| \to 0$ as $n\to\infty$. We may then write

$$\mathbb{E}\left[M_t^2 - M_s^2 \mid \mathcal{F}_s\right] \quad = \quad \sum_{k=1}^{n}\mathbb{E}\left[M_{t_k^n}^2 - M_{t_{k-1}^n}^2 \mid \mathcal{F}_s\right] \quad = \quad \sum_{k=1}^{n}\mathbb{E}\left[\left(M_{t_k^n} - M_{t_{k-1}^n}\right)^2 \mid \mathcal{F}_s\right],$$

by the conditional orthogonality of martingale increments. On the other hand, by construction,

$$\sum_{k=1}^{n}\left(M_{t_k^n} - M_{t_{k-1}^n}\right)^2 \quad \xrightarrow[n\to\infty]{L^1} \quad \langle M\rangle_t - \langle M\rangle_s.$$

Taking conditional expectation w.r.t. $\mathcal{F}_s$, we conclude from the above computation that

$$\mathbb{E}\left[M_t^2 - M_s^2 \mid \mathcal{F}_s\right] \quad = \quad \mathbb{E}\left[\langle M\rangle_t - \langle M\rangle_s|\mathcal{F}_s\right].$$

Since M_s and $\langle M\rangle_s$ are $\mathcal{F}_s$-measurable, this proves the desired martingale property. □

Example 1.4 (Brownian case). *In the case of a Brownian motion $B = (B_t)_{t\geq 0}$, the computation (1.23) shows that $\langle B\rangle_t = t$, which does indeed satisfy Properties (i)-(iv) above.*

Remark 1.7 (Quadratic covariation). *If M, N are two continuous square-integrable martingales, then we may define their quadratic covariation by the polarization formula:*

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

The above result implies that $\langle M,N\rangle$ is continuous, that $MN - \langle M,N\rangle$ is a martingale, and that

$$\sum_{k=1}^{n}\left(M_{t_k} - M_{t_{k-1}}\right)\left(N_{t_k} - N_{t_{k-1}}\right) \quad \xrightarrow[n\to\infty]{L^1} \quad \langle M,N\rangle_t,$$

for any subdivisions $(t_k^n)_{0\leq k\leq n}$ of $[0,t]$ with $\max_{0\leq k\leq n}|t_k^n - t_{k-1}^n| \to 0$ as $n\to\infty$.

Remark 1.8 (Absolute vs quadratic variation). *If f has finite variation and g is continuous, then*

$$\sum_{k=1}^{n}\left(f(t_k) - f(t_{k-1})\right)\left(g(t_k) - g(t_{k-1})\right) \quad \leq \quad V(f,0,t)\max_{u,v\in[0,t],|u-v|\leq\Delta}|g(u) - g(v)|,$$

where $\Delta = \max_{0\leq k\leq n}|t_k^n - t_{k-1}^n|$. By uniform continuity of g on compact sets (Heine's theorem), the right-hand side tends to 0 as $\Delta\to 0$, i.e. $\langle f,g\rangle = 0$. Taking $f = g$ shows that a continuous process with finite variation must have zero quadratic variation. When applied to martingales, this implies the following result, which considerably extends our observation about the roughness of Brownian motion.

Corollary 1.1 (No interesting martingale has finite variation). *If $M = (M_t)_{t\geq 0}$ is a continuous square-integrable martingale which has finite variations a.-s., then M is a.-s. constant in time:*

$$\mathbb{P}(\forall t \geq 0, \quad M_t = M_0) \quad = \quad 1. \tag{1.25}$$

Proof. Fix $t \geq 0$. By the above remark, we have $\langle M \rangle_t = 0$. On the other hand, the orthogonality of martingale increments and property (iv) above yield

$$\mathbb{E}[(M_t - M_0)^2] \quad = \quad \mathbb{E}[M_t^2] - \mathbb{E}[M_0^2] \; = \; \mathbb{E}[\langle M \rangle_t] \; = \; 0,$$

which shows that $\mathbb{P}(M_t = M_0) = 1$. This is true for any fixed $t \geq 0$, so we may take $t \in \mathbb{Q}_+$ and invoke the continuity of M to conclude that $\mathbb{P}(\forall t \geq 0, M_t = M_0) = 1$, as desired. □

Remark 1.9 (Truncation). *The result actually holds without the square-integrability assumption. Indeed, stopping preserves both the finite variation and the martingale properties, so the conclusion applies to M^{T_n}, where $T_n = \inf\{t \geq 0 \colon |M_t| \geq n\}$, and can then be transferred to M by sending $n \to \infty$.*

Remark 1.10 (Uniqueness). *The quadratic variation $\langle M \rangle = (\langle M \rangle_t)_{t\geq 0}$ defined at (1.30) is the only process satisfying the properties $(i) - (iv)$ in Theorem 1.3. Indeed, if $A = (A_t)_{t\geq 0}$ is another process with these properties, then $\langle M \rangle - A = (M^2 - A) - (M^2 - \langle M \rangle)$ is a continuous martingale (as the difference of two continuous martingales), and has finite variation a.-s. (as the difference of two non-decreasing processes). Thus, it must be a.-s. constantly equal to its initial value, which is zero.*

1.5 Levy's characterization of Brownian motion

The quadratic variation of a continuous martingale is a remarkable object. Perhaps surprisingly, it completely determines the distribution of the underlying martingale. We shall here only prove the following important special case, which constitutes a deep and celebrated result.

Theorem 1.4 (Levy's characterization of Brownian motion). *For a process $M = (M_t)_{t\geq 0}$ on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, the following two statements are equivalent*

(i) M is a continuous square-integrable martingale with $M_0 = 0$ and $\langle M \rangle_t = t$ for all $t \geq 0$;

(ii) M is a $(\mathcal{F}_t)_{t\geq 0}$−Brownian motion.

Proof. The implication $(ii) \implies (i)$ has already been established. Conversely, let us suppose that (i) holds. We will establish the following fact: for any twice-differentiable function $F \colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{C}$ whose first and second-order derivatives are all bounded, the process Z defined by

$$Z_t \quad := \quad F(t, M_t) - \int_0^t \left(\frac{\partial F}{\partial t} + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} \right)(u, M_u)\, \mathrm{d}u, \tag{1.26}$$

is a martingale. In particular, taking $F(t,x) = \exp\left(i\theta x + \frac{\theta^2 t}{2}\right)$ makes the integral vanish, so that

$$\mathbb{E}\left[e^{i\theta(M_t - M_s)} \mid \mathcal{F}_s\right] \quad = \quad e^{-\frac{\theta^2(t-s)}{2}},$$

for any $0 \leq s \leq t$ and $\theta \in \mathbb{R}$. This formula shows that $M_t - M_s$ has law $\mathcal{N}(0, t-s)$ and is independent of $\mathcal{F}_s$, as desired. To prove (1.26), we Taylor-expand F: for $0 \leq t \leq t'$ and $x, x' \in \mathbb{R}$,

$$\begin{aligned} F(t', x') - F(t, x) \quad &= \quad (t' - t)\left(\frac{\partial F}{\partial t}(t,x) + o(1)\right) \\ &\quad + (x' - x)\frac{\partial F}{\partial x}(t,x) + \frac{(x'-x)^2}{2}\left(\frac{\partial^2 F}{\partial x^2}(t,x) + o(1)\right), \end{aligned}$$

where the $o(1)$ term is uniformly bounded and can be made arbitrary small by choosing $|x' - x| + |t' - t|$ small (this uses the boundedness assumption on the derivatives of F). We now choose $(x, x') = (M_t, M_{t'})$ and take conditional expectation w.r.t. to $\mathcal{F}_t$. Using $\mathbb{E}[M_{t'} - M_t \mid \mathcal{F}_t] = 0$ and $\mathbb{E}[(M_{t'} - M_t)^2 \mid \mathcal{F}_t] = (t' - t)$, we easily obtain

$$\mathbb{E}\left[F(t', M_{t'}) - F(t, M_t) - (t' - t)\left(\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\right)(t, M_t) \mid \mathcal{F}_t\right] \quad = \quad (t' - t)o(1).$$

Finally, fix $0 \le s \le t$, set $t_k^n = \frac{(t-s)k}{n}$, and apply the above identity to $t = t_{k-1}^n$ and $t' = t_k^n$. By the tower property of conditional expectation, we may replace $\mathcal{F}_t$ by $\mathcal{F}_s$. Summing the resulting identity over $1 \le k \le n$ yields

$$\mathbb{E}\left[F(t, M_t) - F(s, M_s) - \sum_{k=1}^{n}(t_k^n - t_{k-1}^n)\left(\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\right)(t_{k-1}^n, M_{t_{k-1}^n}) \mid \mathcal{F}_s\right] \quad = \quad (t - s)o(1),$$

and taking $n \to \infty$ completes the proof. □

1.6 Local martingales

To deal with more general processes, we will need to relax the integrability requirement in the definition of a martingale. This leads to the following notion.

Definition 1.4 (Local martingale). *A stochastic process $M = (M_t)_{t \ge 0}$ is a local martingale if there exists a sequence $(T_n)_{n \ge 1}$ of stopping times (called a localizing sequence) such that*

(i) For each $n \in \mathbb{N}$, the stopped process $M^{T_n} = (M_{t \wedge T_n})_{t \ge 0}$ is a martingale.

(i) Almost-surely, $T_n \uparrow \infty$ as $n \uparrow \infty$.

Of course, any continuous martingale is a local martingale (take $T_n = +\infty$), but the converse is far from true. In fact, a local martingale needs not even be integrable ! However, any local martingale which is uniformly dominated is a true martingale, as we now show.

Proposition 1.7 (Uniform domination). *For a local martingale M to be a martingale, it suffices that*

$$\forall t \ge 0, \quad \mathbb{E}\left[\sup_{s \in [0,t]} |M_s|\right] \quad < \quad \infty. \tag{1.27}$$

Proof. As any local martingale, M is adapted: it is the pointwise limit of the sequence of adapted processes $(M^{T_n})_{n \ge 1}$, where $(T_n)_{n \ge 1}$ is a localizing sequence. Moreover, the above domination ensures that M is integrable. Finally, fix $0 \le s \le t$. For all $n \in \mathbb{N}$, we know that

$$\mathbb{E}\left[M_{T_n \wedge t} \mid \mathcal{F}_s\right] \quad = \quad M_{T_n \wedge s}. \tag{1.28}$$

To conclude that $\mathbb{E}\left[M_t \mid \mathcal{F}_s\right] = M_s$, we now take $n \to \infty$: the random variables $M_{T_n \wedge t}$ and $M_{T_n \wedge s}$ tend to M_t and M_s a.-s., because $T_n \uparrow \infty$. Moreover, the domination $|M_{T_n \wedge t}| \le Z$ with $Z := \sup_{s \in [0,t]} |M_s| \in L^1$ allows us to safely interchange the limit and conditional expectation. □

Local martingales are easy to work with, because we can always *localize* them to obtain true martingales (for which we have a well-developed theory), and then transfer the desired conclusion by taking a limit. As a consequence, many of the results that we have mentioned about martingales extend easily to local martingales. Here are a few important examples, which we really invite the reader to prove.

Proposition 1.8 (Doob's optional stopping theorem for local martingales). *If M is a continuous local martingale and T a stopping time, then the stopped process $M^T = (M_{t\wedge T})_{t\geq 0}$ is a local martingale.*

Proof. Let $(T_n)_{n\geq 1}$ be a localizing sequence for M, and fix $n \in \mathbb{N}$. Since M^{T_n} is a (continuous) martingale and T a stopping time, the non-local version of Doob's optional stopping Theorem ensures that $M^{T_n \wedge T}$ is a martingale. Thus, $(T_n)_{n\geq 1}$ is also a localizing sequence for M^T. □

Remark 1.11 (A smart localizing sequence). *If M is a continuous local martingale with $M = 0$, then*

$$T_n \quad := \quad \inf\{t \geq 0 \colon |M_t| \geq n\}, \tag{1.29}$$

is a stopping time for any $n \in \mathbb{N}$ (hitting time of a closed set by a continuous adapted process), so the local version of Doob's optional stopping Theorem ensures that M^{T_n} is a local martingale. But M^{T_n} is $[-n, n]-$valued by construction, so the uniform domination (1.27) trivially holds, showing that M^{T_n} is in fact a martingale. Finally, $T_n \to +\infty$ a.-s. as $n \to \infty$, because $\sup_{s\in[0,t]} |M_s| < \infty$ a.-s.. In conclusion, the sequence $(T_n)_{n\geq 1}$ defined by (1.29) is always a localizing sequence for M. It has the additional advantage that the stopped martingale M^{T_n} is bounded for every $n \in \mathbb{N}$, which can be useful.

Proposition 1.9 (Addition of local martingales). *Continuous local martingales form a vector space.*

Proof. Let M and $\widetilde{M}$ be two local martingales, with localizing sequences $(T_n)_{n\geq 1}$ and $(\widetilde{T}_n)_{n\geq 1}$. Fix $n \in \mathbb{N}$. Since M^{T_n} and $\widetilde{M}^{\widetilde{T}_n}$ are continuous martingales, Doob's optional stopping Theorem ensures that the stopped processes $M^{T_n \wedge \widetilde{T}_n}$ and $\widetilde{M}^{T_n \wedge \widetilde{T}_n}$ are martingales. Thus, so is $\lambda M^{T_n \wedge \widetilde{T}_n} + \mu \widetilde{M}^{T_n \wedge \widetilde{T}_n}$, for any $\lambda, \mu \in \mathbb{R}$. But this shows that $(T_n \wedge \widetilde{T}_n)_{n\geq 1}$ is a localizing sequence for $\lambda M + \mu \widetilde{M}$, thereby completing the proof (note that $T_n \wedge \widetilde{T}_n \to \infty$ because $T_n, \widetilde{T}_n \to \infty$). □

Proposition 1.10 (No interesting local martingale has finite variation). *If M is a continuous local martingale which has finite variation a.-s., then $\mathbb{P}(\forall t \geq 0,\, M_t = M_0) = 1$.*

Proof. Assume without loss of generality that $M_0 = 0$. The smart localizing sequence (1.29) makes the stopped process M^{T_n} a square-integrable martingale. Moreover, we have $V(M^{T_n}, 0, t) = V(M, 0, t \wedge T_n) < \infty$ for all $t \geq 0$. Thus, the non-local version of the result ensures that M^{T_n} is a.-s. constant in time, and letting $n \to \infty$ shows that M is a.-s. constant in time, as desired. □

Proposition 1.11 (Quadratic variation). *Let M be a continuous local martingale. Then, the limit*

$$\langle M \rangle_t \quad := \quad \lim_{n\to\infty} \sum_{k=1}^{n} |M_{t_k^n} - M_{t_{k-1}^n}|^2$$

exists in probability for each $t \geq 0$, and does not depend on the subdivisions $(t_k^n)_{0\leq k\leq n}$ of $[0, t]$, as long as $\max_{0\leq k\leq n} |t_k^n - t_{k-1}^n| \to 0$ as $n \to \infty$. Moreover, $\langle M \rangle$ is the unique process (up to modification) so that

(i) $\langle M \rangle_0 = 0$;

(ii) $t \mapsto \langle M \rangle_t$ *is a.-s. continuous ;*

(iii) $t \mapsto \langle M \rangle_t$ *is a.-s. non-decreasing ;*

(iv) $\left(M_t^2 - \langle M \rangle_t\right)_{t\geq 0}$ *is a local martingale.*

Exercise 1.2 (Square-integrable local martingales). *Fix $t \geq 0$. Show that a continuous local martingale $M = (M_s)_{s\in[0,t]}$ is a square-integrable martingale if and only if $M_0 \in L^2$ and $\langle M \rangle_t \in L^1$.*

Exercise 1.3 (Local martingales are unbounded). *Let M be a continuous local martingale such that a.-s., $\langle M \rangle_\infty = \infty$. Prove that a.-s., $\limsup_{t\to\infty} M_t = +\infty$ and $\liminf_{t\to\infty} M_t = -\infty$.*

Chapter 2

Stochastic integration

We have now arrived to the main theoretical challenge of this introductory course: giving a proper meaning to a stochastic integral of the form

$$I_t \quad = \quad \int_0^t X_u \, \mathrm{d}Y_u, \tag{2.1}$$

where $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ are stochastic processes. A natural idea is of course to define this integral as a limit of Riemann sums, just as one would do if X and Y were deterministic:

$$I_t \quad := \quad \lim_{n \to \infty} \sum_{k=1}^{n} X_{t_{k-1}^n} \left(Y_{t_k^n} - Y_{t_{k-1}^n} \right), \tag{2.2}$$

where $(t_k^n)_{0 \leq k \leq n}$ is a subdivision of $[0, t]$ such that $\max_{1 \leq k \leq n} t_k^n - t_{k-1}^n \to 0$ as $n \to \infty$. Unfortunately, the almost-sure convergence of these Riemann sums requires the process Y to have finite variations, thereby excluding Brownian motions as well as any interesting martingale.

The solution found by Itô consists in compensating roughness by randomness: with a bit of work, it will be shown that the above limit does in fact exist when taken in the L^2 sense, for a wide class of stochastic processes X, Y which includes Brownian motion. The general construction is rather delicate, but will eventually provide us with an extremely robust theory of stochastic integration and differentiation. As a warm-up, let us first restrict our attention to the special case where X is deterministic and Y is a Brownian motion. In this very comfortable setting, the integral I_t is known as a Wiener integral, and it enjoys remarkable properties.

2.1 The Wiener isometry

Let $\mathcal{H}, \mathcal{H}'$ be Hilbert spaces. An isometry is an additive and norm-preserving map $I \colon \mathcal{H} \to \mathcal{H}'$:

$$\forall x, y \in \mathcal{H}, \quad I(x+y) \;=\; I(x) + I(y) \quad \text{and} \quad \|I(x)\|_{\mathcal{H}'} \;=\; \|x\|_{\mathcal{H}}. \tag{2.3}$$

In particular, this implies that I is linear and continuous. When I is only defined on a vector subspace $V \subseteq \mathcal{H}$ (and is linear and norm-preserving thereon), we speak of a partial isometry. The following classical result will play a fundamental role in this chapter.

Theorem 2.1 (Isometry extension)**.** *Let $I \colon V \to \mathcal{H}'$ be a partial isometry whose domain V is dense in $\mathcal{H}$. Then I admits a unique continuous extension to $\mathcal{H}$, and the latter is an isometry.*

Proof. Fix $x \in \mathcal{H} \setminus V$, and take $(x_n)_{n \geq 1}$ in V which converges to x. Clearly, any continuous extension must satisfy

$$I(x) \quad = \quad \lim_{n \to \infty} I(x_n), \tag{2.4}$$

which establishes uniqueness. To prove existence, one would like to use (2.4) as a definition, but there are two potential problems: it is not clear that the limit exists, and even if it does, it might a priori depend on the particular sequence $(x_n)_{n\geq 1}$ chosen to approximate x. Fortunately, both issues are solved by the fact that I is a partial isometry. Indeed, for all $n, m \in \mathbb{N}$, we have

$$\|I(x_n) - I(x_m)\| = \|I(x_n - x_m)\| = \|x_n - x_m\| \xrightarrow[n\wedge m\to\infty]{} 0, \tag{2.5}$$

because $(x_n)_{n\geq 1}$ is convergent. Thus, $(I(x_n))_{n\geq 1}$ is a Cauchy sequence, hence the limit (2.4) exists. Moreover, the latter does not depend on the chosen approximation $(x_n)_{n\geq 1}$. Indeed, if $(y_n)_{n\geq 1}$ is another sequence in V which converges to x, then

$$\|I(x_n) - I(y_n)\| = \|x_n - y_n\| \xrightarrow[n\to\infty]{} 0. \tag{2.6}$$

Thus, the formula (2.4) defines a continuous extension, and the latter is automatically linear and norm-preserving, because these properties depend continuously on their arguments. □

Now, given a Brownian motion $B = (B_t)_{t\geq 0}$ and a deterministic, square-integrable function $f\colon \mathbb{R}_+ \to \mathbb{R}$, our goal is to give a meaning to the random variable

$$I(f) = \int_0^\infty f(u)\,\mathrm{d}B_u. \tag{2.7}$$

Of course, one should have $I(f) = B_t$ in the basic case $f = \mathbf{1}_{(0,t]}$. Also, as any reasonable integral, $f \mapsto I(f)$ should be linear. Together, these two requirements impose that

$$f = \sum_{k=1}^{n} a_k \mathbf{1}_{(t_{k-1},t_k]} \implies I(f) = \sum_{k=0}^{n} a_k \left(B_{t_k} - B_{t_{k-1}}\right), \tag{2.8}$$

for any $n \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathbb{R}$ and $0 = t_0 \leq t_1 \leq \ldots \leq t_n$. With this definition, we have

$$\mathbb{E}\left[(I(f))^2\right] = \sum_{k=0}^{n} a_k^2 (t_k - t_{k-1}) = \int_0^t f^2(u)\,\mathrm{d}u.$$

Thus, our map I is a partial isometry on the subspace $\mathcal{E} \subseteq L^2(\mathbb{R}_+)$ of all step functions:

$$\mathcal{E} := \left\{\sum_{k=1}^{n} a_k \mathbf{1}_{(t_{k-1},t_k]} : n \in \mathbb{N}, (a_1, \ldots, a_n) \in \mathbb{R}^n, 0 = t_0 \leq t_1 \leq \ldots \leq t_n\right\}. \tag{2.9}$$

It turns out that this set is *large enough*, in the precise sense that it is dense in $L^2(\mathbb{R}_+)$.

Lemma 2.1 (Approximation by step functions)**.** *Any function $f \in L^2(\mathbb{R}_+)$ is the limit in $L^2(\mathbb{R}_+)$ of the sequence of step functions $(P_n f)_{n\geq 1}$, where*

$$P_n f := \sum_{k=1}^{n^2} \left(n \int_{\frac{k-1}{n}}^{\frac{k}{n}} f(u)\,\mathrm{d}u\right) \mathbf{1}_{\left(\frac{k}{n}, \frac{k+1}{n}\right]}. \tag{2.10}$$

Moreover, when $f \in C_0^c(\mathbb{R}_+)$, we can replace the $(\cdot)$ term by $f(k/n)$.

Thus, the isometry extension theorem applies, leading to the following result.

Theorem 2.2 (Wiener isometry). *Let $B = (B_t)_{t\geq 0}$ be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. Then, there exists a unique linear and continuous map $I\colon L^2(\mathbb{R}_+) \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ such that for all $t \geq 0$,*

$$I(\mathbf{1}_{(0,t]}) \quad = \quad B_t. \tag{2.11}$$

Moreover, I is an isometry, in the sense that for all $f \in L^2(\mathbb{R}_+)$,

$$\|I(f)\|_{L^2(\Omega)} \quad = \quad \|f\|_{L^2(\mathbb{R}_+)}. \tag{2.12}$$

The map I is called the Wiener isometry, and denoted $I(f) = \int_0^\infty f(u)\, \mathrm{d}B_u$.

Remark 2.1 (Explicit formula). *Let us make several important remarks about this result.*

1. *By construction, for any $f \in L^2(\mathbb{R}_+)$, we have the explicit formula*

$$\int_0^\infty f(t)\, \mathrm{d}B_t \quad = \quad \lim_{n\to\infty} \sum_{k=1}^{n^2} a_{n,k}(f) \left(B_{\frac{k+1}{n}} - B_{\frac{k}{n}}\right), \tag{2.13}$$

   where $a_{n,k}(f) = n \int_{(k-1)/n}^{k/n} f(u)\, \mathrm{d}u$, and where the limit is taken in the L^2 sense. Moreover, in the particular case where $f \in C_c^0(\mathbb{R}_+)$, we can take the simpler choice $a_{n,k}(f) = f(k/n)$.

2. *As any distributional limit of a sequence of (centered) Gaussian random variables, the Wiener integral is a (centered) Gaussian random variable, with variance given by (2.12):*

$$\forall f \in L^2(\mathbb{R}_+), \quad \int_0^\infty f(u)\, \mathrm{d}B_u \quad \sim \quad \mathcal{N}\left(0, \int_0^\infty f^2(u)\, \mathrm{d}u\right). \tag{2.14}$$

3. *Thanks to the polarization identity $2\langle f, g\rangle = \langle f+g, f+g\rangle - \langle f, f\rangle - \langle g, g\rangle$, valid for any symmetric bilinear form, the isometry property (2.12) leads to the covariance formula*

$$\mathit{Cov}\left(\int_0^\infty f(u)\, \mathrm{d}B_u, \int_0^\infty g(u)\, \mathrm{d}B_u\right) \quad = \quad \int_0^\infty f(u)g(u)\, \mathrm{d}u. \tag{2.15}$$

2.2 The Wiener integral as a process

For $0 \leq s \leq t$, it is natural to introduce the notation

$$\int_s^t f(u)\, \mathrm{d}B_u \quad := \quad \int_0^\infty f(u)\mathbf{1}_{(s,t]}(u)\, \mathrm{d}B_u. \tag{2.16}$$

Note that for this definition to make sense, we only need the function $f\colon \mathbb{R}_+ \to \mathbb{R}$ to be locally square-integrable (written $f \in L^2_{\mathrm{LOC}}$), in the sense that $\int_0^t f^2(u)\, \mathrm{d}u < \infty$ for each $t \geq 0$. Also, by linearity of the Wiener integral, we have the Chasles relation

$$\int_0^t f(u)\, \mathrm{d}B_u \quad = \quad \int_0^s f(u)\, \mathrm{d}B_u + \int_s^t f(u)\, \mathrm{d}B_u \quad (0 \leq s \leq t). \tag{2.17}$$

To any $f \in L^2_{\mathrm{LOC}}$, we may now naturally associate a process $M^f = (M_t^f)_{t\geq 0}$, given by

$$M_t^f \quad := \quad \int_0^t f(u)\, \mathrm{d}B_u. \tag{2.18}$$

Clearly, M^f is a centered Gaussian process, with covariance function given by

$$\mathrm{Cov}\left(M_t^f, M_s^f\right) \quad = \quad \int_0^{s\wedge t} f^2(u)\, \mathrm{d}u. \tag{2.19}$$

But M^f has an even more remarkable property, which justifies by itself the interest of stochastic integration. In the following result, the underlying filtration $(\mathcal{F}_t)_{t\geq 0}$ can be taken to be the natural filtration of B, or any filtration for which B is a $(\mathcal{F}_t)_{t\geq 0}-$Brownian motion.

Theorem 2.3 (Wiener martingale). *M^f is a continuous square-integrable martingale, with*

$$\forall t \geq 0, \quad \langle M^f \rangle_t \;=\; \int_0^t f^2(u)\,\mathrm{d}u. \tag{2.20}$$

Proof. The square-integrability is clear, by construction of the Wiener integral. Now, for any fixed $0 \leq s \leq t$, the function $f\mathbf{1}_{(s,t]}$ is the L^2-limit of a sequence of step functions supported on $(s,t]$. In view of our construction of the Wiener integral, this implies that

$$M_t^f - M_s^f \;=\; \int_s^t f(u)\,\mathrm{d}B_u \;\in\; \overline{\mathrm{Vect}\,(B_u - B_s \colon u \in [s,t])}. \tag{2.21}$$

Since B is a $(\mathcal{F}_t)_{t\geq 0}-$Brownian motion, it follows that M_t^f is $\mathcal{F}_t-$measurable and that $M_t^f - M_s^f$ is independent of $\mathcal{F}_s$. In particular, we have

$$\begin{aligned} \mathbb{E}[M_t^f - M_s^f|\mathcal{F}_s] \;&=\; \mathbb{E}[M_t^f - M_s^f] = 0; \\ \mathbb{E}[(M_t^f)^2 - (M_s^f)^2|\mathcal{F}_s] \;=\; \mathbb{E}[(M_t^f - M_s^f)^2|\mathcal{F}_s] \;&=\; \mathbb{E}[(M_t^f - M_s^f)^2] \;=\; \int_s^t f^2(u)\,\mathrm{d}u, \end{aligned}$$

where the last identity uses the isometry property. The first line shows that M^f is a martingale, and the second that $\left\{(M^f)_t^2 - \int_0^t f^2(u)\,\mathrm{d}u\right\}_{t\geq 0}$ also is. It thus only remains to prove the continuity of M^f. With $f_n := P_n f$ as in (2.10), the L^2 convergence $f_n\mathbf{1}_{(0,t]} \to f\mathbf{1}_{(0,t]}$ implies that

$$M_t^{f_n} \xrightarrow[n\to\infty]{L^2} M_t^f, \tag{2.22}$$

for each $t \geq 0$. In view of Proposition 1.5, we thus only have to establish the continuity of M^f when f is a step function. By linearity, we may further assume that $f = \mathbf{1}_{(a,b]}$, $0 \leq a \leq b$. But then the result is trivial, since $M_t^f = B_{b\wedge t} - B_{a\wedge t}$. □

When $f = 1$, we have $M^f = B$, so we recover earlier observations. Likewise, we had seen earlier that $(e^{\theta B_t - \theta^2 t/2})_{t\geq 0}$ is a martingale, for any $\theta \in \mathbb{R}$. Here is a considerable generalization.

Proposition 2.1 (Exponential martingale). *For any $f \in L^2_{\mathrm{LOC}}$, the process $Z^f = (Z_t^f)_{t\geq 0}$ defined by*

$$Z_t^f \;:=\; \exp\left(\int_0^t f(u)\,\mathrm{d}B_u - \frac{1}{2}\int_0^t f^2(u)\,\mathrm{d}u\right), \tag{2.23}$$

is a (continuous, square-integrable) martingale.

Proof. The integrability poses no problem, because the stochastic integral is a Gaussian random variable. Now, fix $0 \leq s \leq t$. As already observed, the random variable $\int_0^s f(u)\,\mathrm{d}B_u$ is $\mathcal{F}_s-$measurable, while $\int_s^t f(u)\,\mathrm{d}B_u$ is independent of $\mathcal{F}_s$. Consequently, we have

$$\begin{aligned} \mathbb{E}\left[Z_t^f|\mathcal{F}_s\right] \;&=\; Z_s^f\,\mathbb{E}\left[\exp\left(\int_s^t f(u)\,\mathrm{d}B_u - \frac{1}{2}\int_s^t f^2(u)\,\mathrm{d}u\right)\right] \\ &=\; Z_s^f, \end{aligned}$$

because $\int_s^t f(u)\,\mathrm{d}B_u \sim \mathcal{N}\left(0, \int_s^t f^2(u)\,\mathrm{d}u\right)$. □

Exercise 2.1. *Determine the law of the process X in the following two cases*

$$\begin{aligned} X_t \;&:=\; (1-t)\int_0^t \frac{1}{1-u}\,\mathrm{d}B_u, \quad t \in (0,1); \\ X_t \;&:=\; e^{-t}\left(X_0 + \int_0^t e^u\,\mathrm{d}B_u\right), \quad t \in \mathbb{R}_+ \quad \textit{with } X_0 \sim \mathcal{N}(0,1/2) \textit{ indep. of } B. \end{aligned}$$

2.3 Progressive processes

As above, we consider a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ on which is given a $(\mathcal{F}_t)_{t\geq 0}-$Brownian motion $B = (B_t)_{t\geq 0}$. We seek to extend the Wiener integral to the case where the deterministic function f is replaced by a stochastic process $\phi = (\phi_t)_{t\geq 0}$. To do so, we will need to require that ϕ is progressive, in the sense that for each fixed $t \geq 0$, the function

$$\begin{array}{ccc} ([0,t]\times\Omega, \mathcal{B}([0,t])\otimes\mathcal{F}_t) & \to & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ (u,\omega) & \mapsto & \phi_u(\omega) \end{array}$$

is measurable. This is more than asking that $\omega \mapsto \phi_t(\omega)$ is $\mathcal{F}_t-$measurable for each $t \geq 0$, and that $t \mapsto \phi_t(\omega)$ is Borel-measurable for each $\omega \in \Omega$. It ensures, by Fubini's Theorem, that $\int_0^t \phi_u \,\mathrm{d}u$ is $\mathcal{F}_t-$measurable whenever $\int_0^t |\phi_u| \,\mathrm{d}u < \infty$ a.-s.

Remark 2.2 (Progressive $\sigma-$field)**.** *It is easy to check that the set $\mathcal{P}$ defined by*

$$\mathcal{P} \quad := \quad \bigcap_{t\geq 0} \{A \subseteq \mathbb{R}_+ \times \Omega \colon A \cap ([0,t]\times\Omega) \in \mathcal{B}([0,t])\otimes\mathcal{F}_t\},$$

is a $\sigma-$field, and that a process ϕ is progressive if and only if the map $(t,\omega) \mapsto \phi_t(\omega)$ is $\mathcal{P}-$measurable.

Proposition 2.2 (Sufficient conditions)**.** *The class of progressive processes include:*

(o) any deterministic process $\phi_t(\omega) = f(t)$, where $f\colon \mathbb{R}_+ \to \mathbb{R}$ is measurable.

(i) any process of the form $\phi_t(\omega) = X(\omega)\mathbf{1}_{(a,b]}(t)$ where $0 \leq a \leq b$, and where X is $\mathcal{F}_a-$measurable;

(ii) any process of the form $\phi_t(\omega) = \mathbf{1}_{[0,T(\omega)]}(t)$, where T is a stopping time;

(iii) any process of the form $\phi_t(\omega) = F\left(\phi_t^1(\omega), \ldots, \phi_t^n(\omega)\right)$, where $F\colon \mathbb{R}^n \to \mathbb{R}$ is a measurable function and $(\phi^1, \ldots, \phi^n)$ are progressive processes (so sums, products, etc);

(iv) any pointwise limit $\phi = \lim_{n\to\infty} \phi^n$ of a sequence $(\phi^n)_{n\geq 1}$ of progressive processes;

(v) any continuous and adapted process.

Proof. For (i), we write for any Borel set $B \in \mathcal{B}(\mathbb{R})$ which does not contain 0 (otherwise, take B^c)

$$\{(u,\omega) \in [0,t]\times\Omega \colon \phi_u(\omega) \in B\} \quad = \quad \{u \in [0,t]\cap(a,b]\} \times \{\omega \in \Omega \colon X(\omega) \in B\},$$

which is either empty (if $t \leq a$), or of the form $I \times A$ with $I \in \mathcal{B}[0,t])$ and $A \in \mathcal{F}_t$ (if $t \geq u$). For (ii), we note that ϕ is $\{0,1\}-$valued, and that

$$\{(u,\omega) \in [0,t]\times\Omega \colon \phi_u(\omega) = 0\} \quad = \quad \bigcup_{q\in[0,t]\cap\mathbb{Q}} (q,t] \times \{\omega \in \Omega \colon T(\omega) \leq q\}.$$

For (iii) and (iv), we simply use the fact that limits and compositions of measurable functions are measurable. Finally, (v) follows from (i),(iii) and (iv) once we observe that any continuous adapted process ϕ is the pointwise limit of the sequence $(\phi^n)_{n\geq 1}$, where

$$\phi_t^n(\omega) \quad := \quad X_0(\omega)\mathbf{1}_{(t=0)} + \sum_{k=0}^{n^2} X_{\frac{k}{n}}(\omega)\mathbf{1}_{(\frac{k}{n},\frac{k+1}{n}]}(t). \tag{2.24}$$

This concludes the proof. $\square$

2.4 The Itô isometry

We let $\mathbb{M}^2(\mathbb{R}_+)$ denote the space of progressive processes $\phi = (\phi_t)_{t\geq 0}$ such that

$$\mathbb{E}\left[\int_0^\infty \phi_u^2 \, \mathrm{d}u\right] \quad < \quad \infty. \tag{2.25}$$

By Remark 2.2, $\mathbb{M}^2(\mathbb{R}_+) = L^2\left(\mathbb{R}_+ \times \Omega, \mathcal{P}, \mathrm{d}t \otimes \mathbb{P}(\mathrm{d}\omega)\right)$ is a Hilbert space, with scalar product $\langle \psi, \phi \rangle_{\mathbb{M}^2} := \mathbb{E}\left[\int_0^\infty \psi_u \phi_u \, \mathrm{d}u\right]$. This space contains every elementary random step function

$$\phi_u(\omega) \quad := \quad X(\omega)\mathbf{1}_{(s,t]}(u), \tag{2.26}$$

where $0 \leq s \leq t$ and $X \in L^2(\Omega, \mathcal{F}_s, \mathbb{P})$. For such a basic process, it makes sense to define

$$\int_0^\infty \phi_u \, \mathrm{d}B_u \quad := \quad X(\omega)\left(B_t - B_s\right). \tag{2.27}$$

As in the Wiener case, this definition extends uniquely to the whole Hilbert space:

Theorem 2.4 (Itô integral). *There exists a unique continuous and linear map $I\colon \mathbb{M}^2(\mathbb{R}_+) \to L^2(\Omega)$ such that $I(\phi) = X(B_t - B_s)$ whenever ϕ is as in (2.26). Moreover, I is an isometry, i.e.*

$$\forall \psi, \phi \in \mathbb{M}^2(\mathbb{R}_+), \quad \mathbb{E}\left[I(\psi)I(\phi)\right] \quad = \quad \mathbb{E}\left[\int_0^\infty \psi_u \phi_u \, \mathrm{d}u\right]. \tag{2.28}$$

We call I the Itô integral, and write $I(\phi) = \int_0^\infty \phi_u \, \mathrm{d}B_u$.

Proof. If $\phi = (\phi_t)_{t\geq 0}$ is a random step function of the form

$$\phi_t(\omega) \quad = \quad \sum_{k=0}^{n-1} X_k(\omega)\mathbf{1}_{(t_k, t_{k+1}]}, \tag{2.29}$$

with $n \in \mathbb{N}$, $0 \leq t_0 \leq \ldots \leq t_n$, and $X_k \in L^2(\Omega, \mathcal{F}_{t_k}, \mathbb{P})$ for each $0 \leq k < n$, we are forced to set

$$I(\phi) \quad := \quad \sum_{k=0}^{n-1} X_k\left(B_{t_{k+1}} - B_{t_k}\right).$$

Note that $I(\phi) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, for $0 \leq j < k < n$, we have

$$\mathbb{E}\left[X_j(B_{t_{j+1}} - B_{t_j})X_k(B_{t_{k+1}} - B_{t_k})\right] \quad = \quad \mathbb{E}\left[X_j(B_{t_{j+1}} - B_{t_j})X_k\right]\mathbb{E}\left[B_{t_{k+1}} - B_{t_k}\right] \;=\; 0,$$

because $X_j, X_k, (B_{t_{j+1}} - B_{t_j})$ are $\mathcal{F}_{t_k}-$measurable, while $B_{t_{k+1}} - B_{t_k}$ is independent of $\mathcal{F}_{t_k}$. Thus,

$$\mathbb{E}\left[|I(\phi)|^2\right] \quad = \quad \sum_{k=0}^{n-1} \mathbb{E}\left[X_k^2\left(B_{t_{k+1}} - B_{t_k}\right)^2\right] \;=\; \sum_{k=0}^{n-1} \mathbb{E}\left[X_k^2\right](t_{k+1} - t_k) \;=\; \mathbb{E}\left[\int_0^\infty \phi_t^2 \, \mathrm{d}t\right].$$

This identity shows that our linear map I – so far defined on random step functions – is an isometry. To conclude, it thus only remains to show that random step functions are dense in $\mathbb{M}^2(\mathbb{R}_+)$. For this, we again use the approximation operators $(P_n)_{n\geq 1}$ from Lemma 2.1, i.e.

$$(P_n\phi)_t \quad = \quad \sum_{k=1}^{n^2}\left(n\int_{\frac{k-1}{n}}^{\frac{k}{n}} \phi_u \, \mathrm{d}u\right)\mathbf{1}_{\left(\frac{k}{n}, \frac{k+1}{n}\right]}(t). \tag{2.30}$$

Note that $P_n\phi$ is a random step function for any $\phi \in \mathbb{M}^2(\mathbb{R}_+)$ and $n \in \mathbb{N}$, because the random variable $n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \phi_s \, \mathrm{d}s$ is $\mathcal{F}_{k/n}-$measurable (the progressivity of ϕ is used here) with

$$\mathbb{E}\left[\left(n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \phi_u \, \mathrm{d}u\right)^2\right] \quad \leq \quad \mathbb{E}\left[n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \phi_u^2 \, \mathrm{d}u\right] \quad < \quad +\infty \tag{2.31}$$

($\|\phi\|_{\mathbb{M}^2} < \infty$ is used here). Now, to prove that $P_n\phi \to \phi$ in M^2, we estimate

$$\|P_n\phi - \phi\|_{\mathbb{M}^2}^2 \quad = \quad \mathbb{E}\left[\|P_n\phi - \phi\|_{L^2(\mathbb{R}_+)}^2\right]. \tag{2.32}$$

Lemma 2.1 ensures that the term inside the expectation tends a.-s. to 0 as $n \to \infty$ (the random function $u \mapsto \phi_u$ is a.-s. in $L^2(\mathbb{R}_+)$, because $\|\phi\|_{\mathbb{M}^2} < \infty$). Moreover, we have the domination

$$\|P_n\phi - \phi\|_{L^2(\mathbb{R}_+)}^2 \quad \leq \quad \left(\|P_n\phi\|_{L^2(\mathbb{R}_+)} + \|\phi\|_{L^2(\mathbb{R}_+)}\right)^2 \ \leq \ 4\|\phi\|_{L^2(\mathbb{R}_+)}^2, \tag{2.33}$$

and the right-hand side has expectation $4\|\phi\|_{\mathbb{M}^2}^2 < \infty$. □

Remark 2.3 (Important comments). *Here are a few elementary but important observations.*

1. *For any $\phi \in \mathbb{M}^2(\mathbb{R}_+)$, we have, in the L^2 sense,*

$$\int_0^\infty \phi_u \, \mathrm{d}B_u \quad = \quad \lim_{n\to\infty} \sum_{k=1}^{n^2} \left(n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \phi_u \, \mathrm{d}u\right) \left(B_{\frac{k+1}{n}} - B_{\frac{k}{n}}\right). \tag{2.34}$$

2. *In the deterministic case $\phi_t(\omega) = f(t)$ with $f \in L^2(\mathbb{R}_+)$, we recover the Wiener integral.*

3. *Regardless of whether the process $\phi \in \mathbb{M}^2(\mathbb{R}_+)$ is centered or not, we always have*

$$\mathbb{E}\left[\int_0^\infty \phi_u \, \mathrm{d}B_u\right] \quad = \quad 0. \tag{2.35}$$

4. *For any $\phi, \psi \in \mathbb{M}^2(\mathbb{R}_+)$, the isometry formula also reads (by polarization)*

$$Cov\left(\int_0^\infty \phi_u \, \mathrm{d}B_u, \int_0^\infty \psi_u \, \mathrm{d}B_u\right) \quad = \quad \mathbb{E}\left[\int_0^\infty \phi_u\psi_u \, \mathrm{d}u\right]. \tag{2.36}$$

5. *Even in the elementary case (2.26), the random variable $\int_0^\infty \phi_u \, \mathrm{d}B_u$ has no reason to be Gaussian!*

2.5 The Itô integral as a process

As in the Wiener case, we adopt the natural notation

$$\int_s^t \phi_u \, \mathrm{d}B_u \quad := \quad \int_0^\infty \phi_u \mathbf{1}_{(s,t]}(u) \, \mathrm{d}B_u, \tag{2.37}$$

for all $0 \leq s \leq t$. Note that the right-hand side makes sense as soon as ϕ is progressive with

$$\forall t \geq 0, \quad \int_0^t \mathbb{E}\left[\phi_u^2\right] \, \mathrm{d}u \quad < \quad \infty. \tag{2.38}$$

The space of such processes is quite larger than $\mathbb{M}^2(\mathbb{R}_+)$, and will be denoted by $\mathbb{M}^2$. The interest of stochastic integration is essentially contained in the following fundamental result.

Theorem 2.5 (Itô martingale). *For any $\phi \in \mathbb{M}^2$, the process $M^\phi = (M_t^\phi)_{t\geq 0}$ defined by*

$$M_t^\phi := \int_0^t \phi_u \, \mathrm{d}B_u \tag{2.39}$$

is a continuous square-integrable martingale, with quadratic variation

$$\langle M^\phi \rangle_t = \int_0^t \phi_u^2 \, \mathrm{d}u. \tag{2.40}$$

Proof. Let us first consider the case of an elementary random step function $\phi_t(\omega) = X(\omega)\mathbf{1}_{(a,b]}(t)$, with $0 \leq a \leq b$ and $X \in L^2(\Omega, \mathcal{F}_a, \mathbb{P})$. By definition, we then have for $t \geq 0$,

$$M_t^\phi = X\left(B_{b\wedge t} - B_{a\wedge t}\right) = \begin{cases} X(B_{b\wedge t} - B_a) & \text{if } a \leq t \\ 0 & \text{else.} \end{cases} \tag{2.41}$$

The continuity of M^ϕ is clear from the first expression, and the adaptedness and square-integrability easily follow from the second expression. Moreover, for $0 \leq s \leq t$, we have

$$M_t^\phi - M_s^\phi = \begin{cases} X(B_{b\wedge t} - B_{s\vee a}) & \text{if } s \leq b \text{ and } a \leq t \\ 0 & \text{else.} \end{cases} \tag{2.42}$$

In either case, we easily find

$$\mathbb{E}\left[M_t^\phi - M_s^\phi \mid \mathcal{F}_{s\vee a}\right] = 0 \tag{2.43}$$

$$\mathbb{E}\left[\left(M_t^\phi - M_s^\phi\right)^2 \mid \mathcal{F}_{s\vee a}\right] = X^2 \int_s^t \mathbf{1}_{(a,b]}(u) \, \mathrm{d}u = \int_s^t \phi_u^2 \, \mathrm{d}u. \tag{2.44}$$

By the tower property of conditional expectation, this implies $\mathbb{E}[M_t^\phi - M_s^\phi \mid \mathcal{F}_s] = 0$ and $\mathbb{E}[(M_t^\phi - M_s^\phi)^2 - \int_s^t \phi_u^2 \, \mathrm{d}u \mid \mathcal{F}_s] = 0$. Thus, M^ϕ is a martingale, and $\langle M^\phi \rangle_t = \int_0^t \phi_u^2 \, \mathrm{d}u$. Now, if $\widetilde{\phi}_t(\omega) = \widetilde{X}\mathbf{1}_{(\widetilde{a},\widetilde{b}]}$ is another step function with $b \leq \widetilde{a}$, then a similar reasoning as above yields

$$\mathbb{E}\left[\left(M_t^\phi - M_s^\phi\right)\left(M_t^{\widetilde{\phi}} - M_s^{\widetilde{\phi}}\right) \mid \mathcal{F}_{s\vee\widetilde{a}}\right] = 0 = \int_s^t \phi_u \widetilde{\phi}_u \, \mathrm{d}u. \tag{2.45}$$

This remains true if $\mathcal{F}_{s\vee\widetilde{a}}$ is replaced by $\mathcal{F}_s$, and we conclude that the formula

$$\langle M^\phi, M^{\widetilde{\phi}} \rangle_t = \int_0^t \phi_u \widetilde{\phi}_u \, \mathrm{d}u$$

holds whenever the elementary random step functions $\phi, \widetilde{\phi}$ are equal or have disjoint support. Now, if ϕ an arbitrary random step functions, then ϕ is a linear combination of elementary random step functions with disjoint supports, so the above computations show that M^ϕ is a square-integrable martingale with $\langle M^\phi \rangle_t = \int_0^t \phi_u^2 \, \mathrm{d}u$. Finally, this extends to any $\phi \in \mathbb{M}^2$ by Proposition 1.5, once we have observed that

$$\forall t \geq 0, \quad M_t^{\phi^n} \xrightarrow[n\to\infty]{L^2} M_t^\phi, \tag{2.46}$$

with $\phi^n = P_n\phi$, because $(\omega, s) \mapsto \phi_s^n(\omega)\mathbf{1}_{(0,t]}(s)$ converges to $(\omega, s) \mapsto \phi_s(\omega)\mathbf{1}_{(0,t]}(s)$ in $\mathbb{M}^2(\mathbb{R}_+)$. □

Remark 2.4 (Quadratic covariation). *By polarization, we have for all $\phi, \psi \in \mathbb{M}^2$ and all $t \geq 0$,*

$$\langle M^\phi, M^\psi \rangle_t = \int_0^t \phi_u \psi_u \, \mathrm{d}u. \tag{2.47}$$

Remark 2.5 (Two differences with the Wiener integral). *Unlike the Wiener case, the random variable $\int_s^t \phi_u \, \mathrm{d}B_u$ has, in general, no reason to be Gaussian, and no reason to be independent of $\mathcal{F}_s$!*

2.6 Generalized Itô integral

We now extend the Itô integral to the class $\mathbb{M}^2_{\mathrm{LOC}}$ of all progressive process $\phi = (\phi_t)_{t\geq 0}$ satisfying

$$\forall t \geq 0 \quad \int_0^t \phi_u^2 \, \mathrm{d}u \quad < \quad \infty, \tag{2.48}$$

almost-surely. Note that this new space is much larger than $\mathbb{M}^2$: in particular, it contains every continuous adapted process ! Now, fix $\phi \in \mathbb{M}^2_{\mathrm{LOC}}$ and $n \in \mathbb{N}$, and consider the stopping time

$$T_n \quad := \quad \inf\left\{ t \geq 0 \colon \int_0^t \phi_u^2 \, \mathrm{d}u \geq n \right\}, \tag{2.49}$$

which is the hitting time of the closed set $[n, \infty)$ by the continuous adapted process $t \mapsto \int_0^t \phi_u^2 \, \mathrm{d}u$. Thanks to Proposition 2.2, the truncated process

$$\phi_t^n(\omega) \quad := \quad \phi_t(\omega) \mathbf{1}_{[0,T_n(\omega)]}(t)$$

is progressive, and it is even in $\mathbb{M}^2(\mathbb{R}_+)$ because

$$\int_0^\infty (\phi_u^n)^2 \, \mathrm{d}u \quad = \quad \int_0^{T_n} \phi_u^2 \, \mathrm{d}u \; \leq \; n,$$

by definition of T_n. Consequently, the process M^n given by

$$M_t^n \quad := \quad \int_0^t \phi_u^n \, \mathrm{d}B_u \; = \; \int_0^\infty \phi_u \mathbf{1}_{[0,T_n \wedge t]}(u) \, \mathrm{d}B_u, \tag{2.50}$$

is a perfectly well-defined continuous (square-integrable) martingale, for each $n \in \mathbb{N}$. Now, fix $t \geq 0$. On the event $\{T_n \geq t\}$, we have $\phi^m \mathbf{1}_{[0,t]} = \phi^n \mathbf{1}_{[0,t]}$ for all $m \geq n$. By virtue of (2.34), this implies that $M_t^m = M_t^n$ for all $m \geq n$. Thus the sequence $(M_t^m)_{m\geq 1}$ converges a.-s. on the event $\{T_n \geq t\}$. Since $n \in \mathbb{N}$ is arbitrary, it follows that $(M_t^m)_{m\geq 1}$ converges a.-s. on the event $\bigcup_{n\geq 1}\{T_n \geq t\}$. But the latter has probability 1, thanks to (2.48). Thus, $(M_t^m)_{m\geq 1}$ converges a.-s., to a limit M_t, and we may safely set

$$\int_0^t \phi_u \, \mathrm{d}B_u \quad := \quad M_t \; = \; \lim_{n\to\infty} \int_0^t \phi_u \mathbf{1}_{[0,T_n]}(u) \, \mathrm{d}B_u. \tag{2.51}$$

The fact that $M_{t\wedge T_n} = M_t^n$ shows that M is a continuous local martingale. Let us sum this up.

Theorem 2.6 (Generalized Itô integral). *For any $\phi \in \mathbb{M}^2_{\mathrm{LOC}}$, the process $M^\phi = (M_t^\phi)_{t\geq 0}$ defined by*

$$\forall t \geq 0, \quad M_t^\phi \quad := \quad \int_0^t \phi_u \, \mathrm{d}B_u,$$

is a continuous local martingale with quadratic variation

$$\forall t \geq 0, \quad \langle M^\phi \rangle_t \quad = \quad \int_0^t \phi_u^2 \, \mathrm{d}u$$

Here is a useful stochastic analogue of Lebesgue's dominated convergence Theorem.

Proposition 2.3 (Stochastic dominated convergence). *Fix $t \geq 0$. In order to ensure the convergence*

$$\int_0^t \phi_u^n \, \mathrm{d}B_u \; \xrightarrow[n\to\infty]{\mathbb{P}} \; \int_0^t \phi_u \, \mathrm{d}B_u, \tag{2.52}$$

it suffices that the progressive processes $\phi, \phi^1, \phi^2 \ldots$ satisfy:

(i) *(simple convergence): for almost-every* $u \in [0,t]$, $\phi_u^n \to \phi_u$ *in probability as* $n \to \infty$.

(ii) *(domination): for all* $u \in [0,t]$ *and* $n \in \mathbb{N}$, $|\phi_u^n| \le \Psi_u$ *a.-s., with* $\Psi \in \mathbb{M}^2_{\mathrm{LOC}}$.

Proof. For $k \in \mathbb{N}$, let $T_k := \inf\left\{t \ge 0\colon \int_0^t \psi_u^2 \,\mathrm{d}u \ge k\right\}$. Then the isometry formula in $\mathbb{M}^2$ yields

$$\mathbb{E}\left[\left(\int_0^{T_k\wedge t} \phi_u^n \,\mathrm{d}B_u - \int_0^{T_k\wedge t} \phi_u \,\mathrm{d}B_u\right)^2\right] \;=\; \mathbb{E}\left[\int_0^{T_k\wedge t} (\phi_u^n - \phi_u)^2 \,\mathrm{d}u\right] \xrightarrow[n\to\infty]{} 0, \tag{2.53}$$

by dominated convergence. To conclude from this, we simply write for $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\int_0^t \phi_u^n \,\mathrm{d}B_u - \int_0^t \phi_u \,\mathrm{d}B_u\right| \ge \varepsilon\right) \;\le\; \mathbb{P}(T_k \le t) + \mathbb{P}\left(\left|\int_0^{t\wedge T_k} \phi_u^n \,\mathrm{d}B_u - \int_0^{t\wedge T_k} \phi_u \,\mathrm{d}B_u\right| \ge \varepsilon\right).$$

The first term can be made arbitrarily small by choosing k large enough, because $T_k \uparrow \infty$ a.-s.. The second term can then be made arbitrarily small by choosing n large enough, by (2.53). □

Corollary 2.1. *(Approximation of the generalized Itô integral). If* ϕ *is continuous and adapted, then*

$$\sum_{k=0}^{n-1} \phi_{t_k^n}\left(B_{t_{k+1}^n} - B_{t_k^n}\right) \;\xrightarrow[n\to\infty]{\mathbb{P}}\; \int_0^t \phi_u \,\mathrm{d}B_u,$$

for every $t \ge 0$ *and any subdivision* $(t_n^k)_{0\le k\le n}$ *of* $[0,t]$ *with* $\max_{0\le k<n} |t_{k+1}^n - t_k^n| \to 0$ *as* $n \to \infty$.

Proof. Apply the above theorem with $\phi_t^n = \sum_{k=0}^{n-1} \phi_{t_k^n} \mathbf{1}_{(t_k,t_{k+1}]}(t)$ and $\Psi_t = \sup_{u\in[0,t]} |\phi_u|$. □

In the next chapter, we will compute stochastic integrals explicitly. Here is an example.

Example 2.1 (Brownian against brownian). *For all* $t \ge 0$, *we have*

$$\int_0^t B_u \,\mathrm{d}B_u \;=\; \frac{1}{2}\left(B_t^2 - t\right). \tag{2.54}$$

Proof. Fix $t \ge 0$. Recall that for any continuous adapted process ϕ, we have

$$\sum_{k=0}^{n-1} \phi_{t_k^n}\left(B_{t_{k+1}^n} - B_{t_k^n}\right) \;\xrightarrow[n\to\infty]{\mathbb{P}}\; \int_0^t \phi_u \,\mathrm{d}B_u,$$

where $t_k^n = kt/n$. Taking $\phi = 2B$, we deduce that

$$\sum_{k=0}^{n-1} 2B_{t_k^n}\left(B_{t_{k+1}^n} - B_{t_k^n}\right) \;\xrightarrow[n\to\infty]{\mathbb{P}}\; 2\int_0^t B_u \,\mathrm{d}B_u.$$

On the other hand, we have seen at (1.24) that

$$\sum_{k=0}^{n-1} \left(B_{t_{k+1}^n} - B_{t_k^n}\right)^2 \;\xrightarrow[n\to\infty]{\mathbb{P}}\; t.$$

Adding up those two lines, and observing that $2a(b-a) + (b-a)^2 = b^2 - a^2$, we arrive at

$$\sum_{k=0}^{n-1} \left(B_{t_{k+1}^n}^2 - B_{t_k^n}^2\right) \;\xrightarrow[n\to\infty]{\mathbb{P}}\; 2\int_0^t B_u \,\mathrm{d}B_u + t.$$

But the left-hand side is a telescopic sum, which equals B_t^2 independently of n. Thus,

$$B_t^2 \;=\; 2\int_0^t B_u \,\mathrm{d}B_u + t,$$

almost-surely, as desired. □

Chapter 3

Stochastic differentiation

3.1 Itô processes

In the previous chapter, we have learnt how to integrate a stochastic process $\phi = (\phi_t)_{t\geq 0}$ against our Brownian motion B, resulting in the generalized Itô integral

$$t \quad \mapsto \quad \int_0^t \phi_u \, \mathrm{d}B_u. \tag{3.1}$$

This process is well-defined as soon as $\phi \in \mathbb{M}^2_{\mathrm{LOC}}$, and it is always a continuous local martingale. On the other hand, we of course also know how to integrate a stochastic process $\psi = (\psi_t)_{t\geq 0}$ against the Lebesgue measure, resulting in the classical integral

$$t \quad \mapsto \quad \int_0^t \psi_u \, \mathrm{d}u. \tag{3.2}$$

This process is well-defined, adapted, and continuous as soon as ψ belongs to the space $\mathbb{M}^1_{\mathrm{LOC}}$ of progressive processes satisfying almost-surely

$$\forall t \geq 0 \quad \int_0^t |\psi_u| \, \mathrm{d}u \quad < \quad \infty. \tag{3.3}$$

Note that $\mathbb{M}^2_{\mathrm{LOC}} \subseteq \mathbb{M}^1_{\mathrm{LOC}}$, (Cauchy-Schwartz) and that both spaces contain, in particular, every continuous and adapted process. Also keep in mind that those two integrals are very different: the process (3.1) is always a local martingale, while (3.2) has a.-s. finite variation ! We will now combine those two kinds of processes to construct the main object of stochastic calculus.

Definition 3.1 (Itô process). *An Itô process is a stochastic process $X = (X_t)_{t\geq 0}$ of the form*

$$\forall t \geq 0, \quad X_t \quad = \quad X_0 + \int_0^t \phi_u \, \mathrm{d}B_u + \int_0^t \psi_u \, \mathrm{d}u, \tag{3.4}$$

with $X_0 \in \mathcal{F}_0$, $\psi \in \mathbb{M}^1_{\mathrm{LOC}}$ and $\phi \in \mathbb{M}^2_{\mathrm{LOC}}$. The two integrals are called the martingale term and the drift term, respectively. Instead of (3.4), we will often use the more convenient differential notation

$$\mathrm{d}X_t \quad = \quad \phi_t \, \mathrm{d}B_t + \psi_t \, \mathrm{d}t.$$

Remark 3.1 (Linearity). *Itô processes form a vector space: if X, Y are Itô processes, and $\lambda, \mu \in \mathbb{R}$, then $Z = \lambda X + \mu Y$ is of course an Itô process, and the martingale terms and drift terms behave linearly:*

$$\mathrm{d}Z_t \quad = \quad \lambda \, \mathrm{d}X_t + \mu \, \mathrm{d}Y_t. \tag{3.5}$$

Note that an Itô process is always continuous, and adapted. Giving names to the two parts ϕ and ψ of the decomposition (3.4) suggests that they are unique. This is indeed the case.

Proposition 3.1 (Uniqueness of the drift and martingale terms)**.** *If X simultaneously satisfies*

$$\mathrm{d}X_t \;=\; \phi_t \,\mathrm{d}B_t + \psi_t \,\mathrm{d}t \quad \textit{and} \quad \mathrm{d}X_t \;=\; \widetilde{\phi}_t \,\mathrm{d}B_t + \widetilde{\psi}_t \,\mathrm{d}t,$$

for some $\phi, \widetilde{\phi} \in \mathbb{M}^2_{\mathrm{LOC}}$ and $\psi, \widetilde{\psi} \in \mathbb{M}^1_{\mathrm{LOC}}$, then $\phi, \widetilde{\phi}$ are indistinguishable, and so are $\psi, \widetilde{\psi}$.

Proof. By assumptions, we have almost-surely,

$$\forall t \geq 0, \quad \int_0^t (\phi_u - \widetilde{\phi}_u) \,\mathrm{d}B_u \;=\; \int_0^t (\psi_u - \widetilde{\psi}_u) \,\mathrm{d}u.$$

Now, the left-hand side is a continuous local martingale, while the right-hand side has finite variation almost-surely. Thus, both sides are null a.-s. In particular, the nullity of the left-hand side implies that of its quadratic variation, i.e. a.-s.,

$$\forall t \geq 0, \; \int_0^t (\phi_u - \widetilde{\phi}_u)^2 \,\mathrm{d}u \;=\; 0. \tag{3.6}$$

Letting $t \to \infty$ yields the indistinguishability of ϕ and $\widetilde{\phi}$. On the other hand, we have a.-s.,

$$\forall t \geq 0, \quad \int_0^t (\psi_u - \widetilde{\psi}_u) \,\mathrm{d}u \;=\; 0.$$

It is a classical exercise on Lebesgue integrals that this forces the integrand to be null a.-e.. □

Remark 3.2 (Itô martingales)**.** *If X is as in (3.4), then it follows from the previous chapters that*

1. *X is a local martingale if and only if $X_0 \in L^1$ and $\psi \equiv 0$.*
2. *X is a square-integrable martingale if and only if $X_0 \in L^2$, $\psi \equiv 0$ and $\phi \in \mathbb{M}^2$.*

For this reason, determining the martingale term ϕ and the drift term ψ of an Itô process is essential.

Remark 3.3 (Integral against an Itô process)**.** *Let X be as in (3.4), and let Y be a continuous and adapted process. Then, clearly, $Y\phi \in \mathbb{M}^2_{\mathrm{LOC}}$ and $Y\psi \in \mathbb{M}^1_{\mathrm{LOC}}$, so it makes sense to define for $t \geq 0$,*

$$\int_0^t Y_u \,\mathrm{d}X_u \;:=\; \int_0^t Y_u \phi_u \,\mathrm{d}B_u + \int_0^t Y_u \psi_u \,\mathrm{d}u.$$

By the dominated convergence Theorem (and its stochastic version), we then have

$$\sum_{k=0}^{n-1} Y_{t_k^n}(X_{t_{k+1}^n} - X_{t_k^n}) \;\xrightarrow[n\to\infty]{\mathbb{P}}\; \int_0^t Y_u \,\mathrm{d}X_u, \tag{3.7}$$

along any subdivisions $(t_k^n)_{0\leq k\leq n}$ of $[0,t]$ with $\Delta_n := \max_{0\leq k<n}(t_{k+1}^n - t_k^n) \to 0$ as $n \to \infty$.

Example 3.1 (Squared Brownian motion)**.** *Our Brownian motion B is of course an Itô process (take $\phi = 1$ and $\psi = 0$). A less trivial example is B^2, for which the computation in Example 2.1 shows that*

$$\mathrm{d}B_t^2 \;=\; 2B_t \,\mathrm{d}B_t + \mathrm{d}t.$$

Note the presence of the quadratic variation term $\mathrm{d}t$, compared to the classical formula $\mathrm{d}X_t^2 = 2X_t \,\mathrm{d}X_t$ that one would have in the case of a continuously differentiable process $t \mapsto X_t$. We will come back to it!

3.2 Quadratic variation of an Itô process

Lemma 3.1 (Quadratic variation of an Itô process). *Let X be an Itô process with stochastic differential*

$$\mathrm{d}X_t = \phi_t \,\mathrm{d}B_t + \psi_t \,\mathrm{d}t. \tag{3.8}$$

Then for any subdivision $(t_k^n)_{0\le k\le n}$ of $[0,t]$ with $\Delta_n := \max_{0\le k<n}(t_{k+1}^n - t_k^n) \to 0$ as $n\to\infty$, we have

$$\sum_{k=0}^{n-1}\left(X_{t_{k+1}^n} - X_{t_k^n}\right)^2 \xrightarrow[n\to\infty]{\mathbb{P}} \int_0^t \phi_u^2 \,\mathrm{d}u. \tag{3.9}$$

We will naturally denote the right-hand side by $\langle X\rangle_t$, and call $t\mapsto\langle X\rangle_t$ the quadratic variation of X. More generally, if $\widetilde{X}$ is another Itô process, with $\mathrm{d}\widetilde{X}_t = \widetilde{\phi}_t\,\mathrm{d}B_t + \widetilde{\psi}_t\,\mathrm{d}t$, then

$$\sum_{k=0}^{n-1}\left(X_{t_{k+1}^n} - X_{t_k^n}\right)\left(\widetilde{X}_{t_{k+1}^n} - \widetilde{X}_{t_k^n}\right) \xrightarrow[n\to\infty]{\mathbb{P}} \langle X,\widetilde{X}\rangle_t = \int_0^t \phi_u\widetilde{\phi}_u \,\mathrm{d}u.$$

We call $t\mapsto\langle X,\widetilde{X}\rangle_t$ the quadratic covariation of X and $\widetilde{X}$, and write $\mathrm{d}\langle X,\widetilde{X}\rangle_t = \phi_t\widetilde{\phi}_t\,\mathrm{d}t$.

Proof. We only have to prove the first claim, since the second follows by polarization. Now, when $\psi=0$, X is a continuous local martingale with quadratic variation $t\mapsto\int_0^t\phi_u^2\,\mathrm{d}u$, so the claim is Proposition 1.11. The general case then easily follows from the observation that

$$\sum_{k=0}^{n-1}(Y_{t_{k+1}^n} - Y_{t_k^n})(Z_{t_{k+1}^n} - Z_{t_k^n}) \le V(Y,0,t) \sup_{u,v\in[0,t],|u-v|\le\Delta_n} |Z_u - Z_v| \xrightarrow[n\to\infty]{a.-s.} 0,$$

whenever Y has finite variation and Z is continuous (almost-surely). □

Proposition 3.2 (Stochastic integration by parts). *If X,Y are Itô processes, then so is $(X_tY_t)_{t\ge0}$, and*

$$\mathrm{d}(X_tY_t) = X_t\,\mathrm{d}Y_t + Y_t\,\mathrm{d}X_t + \mathrm{d}\langle X,Y\rangle_t.$$

Proof. Fix $t\ge0$. Consider subdivisions $(t_k^n)_{0\le k\le n}$ of $[0,t]$ with $\max|t_{k+1}^n - t_k^n|\to0$. We have

$$\begin{aligned} X_tY_t - X_0Y_0 &= \sum_{k=0}^{n-1}\left(X_{t_{k+1}^n}Y_{t_{k+1}^n} - X_{t_k^n}Y_{t_k^n}\right) \\ &= \sum_{k=0}^{n-1} X_{t_k^n}\left(Y_{t_{k+1}^n} - Y_{t_k^n}\right) + \sum_{k=0}^{n-1} Y_{t_k^n}\left(X_{t_{k+1}^n} - X_{t_k^n}\right) + \sum_{k=0}^{n-1}(X_{t_{k+1}^n} - X_{t_k^n})(Y_{t_{k+1}^n} - Y_{t_k^n}), \end{aligned}$$

thanks to the identity $x'y' - xy = x(y'-y) + y(x'-x) + (x'-x)(y'-y)$. Letting $n\to\infty$ yields

$$X_tY_t - X_0Y_0 = \int_0^t X_u\,\mathrm{d}Y_u + \int_0^t Y_u\,\mathrm{d}X_u + \langle X,Y\rangle_t, \tag{3.10}$$

by the above Lemma and Remark 3.7. □

Remark 3.4 (Itô term). *Here again, note the extra covariation term $\mathrm{d}\langle X,Y\rangle_t$, compared to the classical integration-by-parts formula $\mathrm{d}(X_tY_t) = X_t\,\mathrm{d}Y_t + Y_t\,\mathrm{d}X_t$ for continuously differentiable trajectories.*

Remark 3.5 (Squaring). *In particular, if X is an Itô process, then so is X^2 and*

$$\mathrm{d}X_t^2 = 2X_t\,\mathrm{d}X_t + \mathrm{d}\langle X\rangle_t, \tag{3.11}$$

thereby generalizing the Brownian case studied in Example 2.1.

3.3 Itô's Formula

We have seen how to differentiate the square of an Itô process. In fact, the square may be replaced with any smooth function, thanks to the following fundamental formula, which is the stochastic analogue of the classical rule $\mathrm{d}F(X_t) = F'(X_t)\,\mathrm{d}X_t$ for differentiating a composed function. The stochastic version contains an extra term, due to the quadratic variation of X.

Theorem 3.1 (Itô's Formula)**.** *Consider an Itô process X, and a function $F \in \mathcal{C}^2(\mathbb{R})$. Then, the process $(F(X_t))_{t\geq 0}$ is again an Itô process, with stochastic differential*

$$\mathrm{d}F(X_t) \quad = \quad F'(X_t)\,\mathrm{d}X_t + \frac{1}{2}F''(X_t)\,\mathrm{d}\langle X\rangle_t. \tag{3.12}$$

Proof. As above, we fix $t \geq 0$ and consider subdivisions $(t_k^n)_{0\leq k\leq n}$ of $[0,t]$ with $\max|t_{k+1}^n - t_k^n| \to 0$. Since $F \in \mathcal{C}^2(\mathbb{R})$, we have the second-order Taylor expansion,

$$\begin{aligned} F(X_t) - F(X_0) &= \sum_{k=0}^{n-1} F(X_{t_{k+1}^n}) - F(X_{t_k^n}) \\ &= \sum_{k=0}^{n-1} F'(X_{t_k^n})\left(X_{t_{k+1}^n} - X_{t_k^n}\right) + \frac{1}{2}\sum_{k=0}^{n-1} F''(X_{U_k^n})\left(X_{t_{k+1}^n} - X_{t_k^n}\right)^2, \end{aligned}$$

for some $U_n^k \in [t_k^n, t_{k+1}^n]$. By Remark 3.7, we know that

$$\sum_{k=0}^{n-1} F'(X_{t_k^n})\left(X_{t_{k+1}^n} - X_{t_k^n}\right) \xrightarrow[n\to\infty]{\mathbb{P}} \int_0^t F'(X_u)\,\mathrm{d}X_u.$$

Thus, it only remains to show that

$$\sum_{k=0}^{n-1} F''(X_{U_k^n})\left(X_{t_{k+1}^n} - X_{t_k^n}\right)^2 \xrightarrow[n\to\infty]{\mathbb{P}} \int_0^t F''(X_u)\,\mathrm{d}\langle X\rangle_u. \tag{3.13}$$

By Lemma 3.1, we already know that

$$\sum_{k=0}^{n-1} Y_{t_k^n}\left(X_{t_{k+1}^n} - X_{t_k^n}\right)^2 \xrightarrow[n\to\infty]{\mathbb{P}} \int_0^t Y_u\,\mathrm{d}\langle X\rangle_u,$$

in the elementary case where $Y_u = 1_{(0,s]}(u)$, for any $s \geq 0$. By linearity, this immediately extends to the case where Y is a random step function. By density, it further extends to the case where Y is any continuous and adapted process. In particular, we may take $Y_u = F''(X_u)$, and this suffices to yield (3.13), since $\max_{0\leq k<n}|F''(X_{U_k^n}) - F''(X_{t_k^n})| \to 0$ a.-s. (by uniform continuity). □

The Itô formula admits a multivariate extension, allowing to combine several Itô processes.

Theorem 3.2 (Multivariate extension)**.** *Let $F \in \mathcal{C}^2(\mathbb{R}^d)$, and let $X^1,\ldots,X^d$ be Itô processes. Then, the process $t \mapsto F(X_t^1,\ldots,X_t^d)$ is again an Itô process, with*

$$\mathrm{d}F(X_t^1,\ldots,X_t^d) \quad = \quad \sum_{i=1}^d \frac{\partial F}{\partial x_i}(X_t^1,\ldots,X_t^d)\,\mathrm{d}X_t^i + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \frac{\partial^2 F}{\partial x_i x_j}(X_t^1,\ldots,X_t^d)\,\mathrm{d}\langle X^i, X^j\rangle_t.$$

Proof. The argument is the same as above, with the multivariate version of Taylor expansion:

$$F(y) \quad = \quad F(x) + \sum_{i=1}^d \frac{\partial F}{\partial x_i}(x)(y_i - x_i) + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \frac{\partial^2 F}{\partial x_i x_j}(z)(y_i - x_i)(y_j - x_j),$$

valid for any $x, y \in \mathbb{R}^d$ and some $z \in \mathrm{Conv}(x,y)$. More precisely, we here take $x = (X_{t_k^n}^1,\ldots,X_{t_k^n}^d)$ and $y = (X_{t_{k+1}^n}^1,\ldots,X_{t_{k+1}^n}^d)$, and then sum over $0 \leq k \leq n-1$, and finally let $n \to \infty$. □

Remark 3.6 (Special cases). *Here are a few special cases of interest.*

1. *One recovers the integration-by-parts formula by taking $F(x,y) = xy$.*

2. *One can add a time dependency by letting one of the Itô processes be $t \mapsto t$. For example,*

$$\mathrm{d}F(t,X_t) \quad = \quad \frac{\partial F}{\partial x}(t,X_t)\,\mathrm{d}X_t + \frac{\partial F}{\partial t}(t,X_t)\,\mathrm{d}t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t,X_t)\,\mathrm{d}\langle X\rangle_t,$$

   for any $F \in \mathcal{C}^2(\mathbb{R}^2)$. Note that $t \mapsto t$ does not contribute to the last term (finite variation).

In view of Remark 3.2, Itô's Formula is extremely useful for finding martingales. Here is a typical exercise to familiarize with this powerful technique.

Exercise 3.1 (Practicing with Itô's Formula). *In each of the following cases, compute the stochastic differential of the process $M = (M_t)_{t\geq 0}$, and deduce that it is a martingale.*

1. $M_t := B_t^2 - t$.

2. $M_t := B_t^3 - 3tB_t$.

3. $M_t := B_t^4 - 6tB_t^2 + 3t^2$.

4. $M_t := B_t^5 - 10tB_t^3 + 15t^2B_t$.

5. $M_t := \exp\left(\theta B_t - \frac{\theta^2}{2}t\right)$, *with $\theta \in \mathbb{R}$.*

6. $M_t := \cos(\theta B_t)e^{\frac{\theta^2}{2}t}$, *with $\theta \in \mathbb{R}$.*

7. $M_t := f(t,B_t)$, *where $f \in \mathcal{C}^2(\mathbb{R}^2)$ satisfies an appropriate condition.*

8. $M_t := B_t^n - \binom{n}{2}\int_0^t B_u^{n-2}\,\mathrm{d}u$, *where $n \geq 2$ is any integer.*

Exercise 3.2 (A typical exam problem). *Let B be a Brownian motion and let F denote the cumulative distribution function of a standard Gaussian random variable. Consider the process*

$$M_t \quad := \quad F\left(\frac{B_t}{\sqrt{1-t}}\right), \qquad t \in [0,1). \tag{3.14}$$

1. *Compute the stochastic differential of M.*

2. *Show that $M_1 := \lim_{t\to 1} M_t$ exists a.-s., and compute it.*

3. *Prove that $(M_t)_{t\in[0,1]}$ is a martingale.*

4. *Deduce the probability that B intersects the graph of $t \mapsto \sqrt{1-t}$, $t \in [0,1]$.*

5. *How did we guess that the Gaussian cumulative distribution function was a good choice for F?*

3.4 Exponential martingales

An important property of the Wiener integral was that the process $(e^{\int_0^t f(u)\,\mathrm{d}B_u - \frac{1}{2}f^2(u)\,\mathrm{d}u})_{t\geq 0}$ is a martingale, for any $f \in L^2_{\mathrm{LOC}}$. The argument used to prove this relied on a specific Gaussian computation, which no longer applies in the Itô case. However, an elementary application of Itô's formula yields the following important result.

Lemma 3.2 (Doléans-Dade exponential)**.** *For any $\phi \in \mathbb{M}^2_{\mathrm{LOC}}$, the process $Z^\phi = (Z^\phi_t)_{t\geq 0}$ defined by*

$$Z^\phi_t \quad := \quad \exp\left(\int_0^t \phi_u \,\mathrm{d}B_u - \frac{1}{2}\int_0^t \phi_u^2 \,\mathrm{d}u\right), \tag{3.15}$$

is a local martingale.

Proof. Applying Itô's formula with $F = \exp$ and $X_t = \int_0^t \phi_u \,\mathrm{d}B_u - \frac{1}{2}\int_0^t \phi_u^2\,\mathrm{d}u$ yields

$$\begin{aligned}
\mathrm{d}Z^\phi_t &= e^{X_t}\,\mathrm{d}X_t + \frac{1}{2}e^{X_t}\,\mathrm{d}\langle X\rangle_t \\
&= e^{X_t}\left(\phi_t\,\mathrm{d}B_t - \frac{1}{2}\phi_t^2\,\mathrm{d}t\right) + \frac{1}{2}e^{X_t}\phi_t^2\,\mathrm{d}t \\
&= e^{X_t}\phi_t\,\mathrm{d}B_t.
\end{aligned}$$

Since $Z^\phi_0 = 1$, we obtain

$$\forall t \geq 0, \quad Z^\phi_t \quad = \quad 1 + \int_0^t Z^\phi_u \phi_u \,\mathrm{d}B_u,$$

and the result follows from the general properties of the Itô integral. □

For reasons that will become clear in the next section, it is very important to ensure that Z^ϕ is really a martingale, and not just a local martingale. This holds, for example, when ϕ is deterministic: the result was proved in the section on Wiener's integral, and can be recovered by checking that the process $(Z^\phi_u\phi_u)_{u\in[0,t]}$ appearing in the stochastic differential of Z^ϕ is in $\mathbb{M}^2$. The following criterion is much more general, but its proof is considerably more involved.

Theorem 3.3 (Novikov's Condition)**.** *Fix $T \in \mathbb{R}_+$. For $(Z^\phi_t)_{t\in[0,T]}$ to be a martingale, it suffices that*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \phi_u^2\,\mathrm{d}u\right)\right] \quad < \quad \infty. \tag{3.16}$$

In the proof of this theorem, we will use the following elementary lemma.

Lemma 3.3 (Non-negative local martingales)**.** *If $M = (M_t)_{t\in[0,T]}$ is a non-negative local martingale, then it is a super-martingale. Moreover, it is a martingale if and only if $\mathbb{E}[M_T] \geq \mathbb{E}[M_0]$.*

Proof. Let $(T_n)_{n\geq 1}$ be a localizing sequence, and let $0 \leq s \leq t \leq T$. For each $n \in \mathbb{N}$, we have

$$\mathbb{E}[M_{T_n\wedge t} \mid \mathcal{F}_s] \quad = \quad M_{T_n\wedge s}.$$

We now take $n \to \infty$. Since $T_n \to \infty$ a.-s., the conditional version of Fatou's Lemma yields

$$\mathbb{E}[M_t \mid \mathcal{F}_s] \quad \leq \quad M_s, \tag{3.17}$$

which shows that M is a super-martingale. Now, suppose that $\mathbb{E}[M_T] \geq \mathbb{E}[M_0]$. This forces the non-increasing map $t \mapsto \mathbb{E}[M_t]$ to be constant on $[0, T]$. In particular, for any $0 \leq s \leq t \leq T$, the non-negative variable $M_s - \mathbb{E}[M_t|\mathcal{F}_s]$ has zero mean, hence is null a.-s.. □

Proof of Theorem 3.3. Fix $0 < \varepsilon < 1$. It is straightforward to check that for all $0 \le t \le T$,

$$\left(Z_t^{(1-\varepsilon)\phi}\right)^{\frac{1}{1-\varepsilon^2}} \quad = \quad \left(Z_t^{\phi}\right)^{\frac{1}{1+\varepsilon}} \left(e^{\frac{1}{2}\int_0^t \phi_u^2 \, \mathrm{d}u}\right)^{\frac{\varepsilon}{1+\varepsilon}}.$$

In particular, we may choose $t = T \wedge T_n$, where $(T_n)_{n\ge 1}$ is a localizing sequence for $Z^{(1-\varepsilon)\phi}$. Taking expectations, and invoking Hölder's inequality, we arrive at

$$\mathbb{E}\left[\left(Z_{T\wedge T_n}^{(1-\varepsilon)\phi}\right)^{\frac{1}{1-\varepsilon^2}}\right] \quad \le \quad \mathbb{E}\left[Z_{T\wedge T_n}^{\phi}\right]^{\frac{1}{1+\varepsilon}} \mathbb{E}\left[e^{\frac{1}{2}\int_0^{T\wedge T_n} \phi_u^2 \, \mathrm{d}u}\right]^{\frac{\varepsilon}{1+\varepsilon}}. \tag{3.18}$$

Since $\mathbb{E}\left[Z_{T\wedge T_n}^{\phi}\right] = 1$, the right-hand side is bounded by $\mathbb{E}\left[e^{\frac{1}{2}\int_0^T \phi_u^2 \, \mathrm{d}u}\right]^{\frac{\varepsilon}{1+\varepsilon}}$ independently of n. This means that the sequence $(Z_{T\wedge T_n}^{(1-\varepsilon)\phi})_{n\ge 1}$ is bounded in L^p with $p = \frac{1}{1-\varepsilon^2} > 1$. Thus,

$$\mathbb{E}\left[Z_T^{(1-\varepsilon)\phi}\right] \quad = \quad \lim_{n\to\infty} \left[Z_{T\wedge T_n}^{(1-\varepsilon)\phi}\right] \quad = \quad 1. \tag{3.19}$$

In particular, $\mathbb{E}\left[(Z_T^{(1-\varepsilon)\phi})^p\right] \ge 1$, so using (3.18) with T instead of $T \wedge T_n$ yields

$$1 \quad \le \quad \mathbb{E}\left[Z_T^{\phi}\right]^{\frac{1}{1+\varepsilon}} \mathbb{E}\left[e^{\frac{1}{2}\int_0^T \phi_u^2 \, \mathrm{d}u}\right]^{\frac{\varepsilon}{1+\varepsilon}}.$$

Taking $\varepsilon \to 0$ yields $\mathbb{E}[Z_T^{\phi}] \ge 1$, which suffices to conclude thanks to the above lemma. □

3.5 Girsanov's Theorem

The interest of ensuring that the exponential local martingale Z^{ϕ} is a martingale – at least when restricted to a finite time horizon $[0, T]$ – is contained in the following fundamental result.

Theorem 3.4 (Girsanov's Theorem). *Fix $\phi \in \mathbb{M}^2_{\mathrm{LOC}}$ and $T \ge 0$, and suppose that the associated exponential local martingale $(Z_t^{\phi})_{t\in[0,T]}$ is a martingale. Then, the formula*

$$\forall A \in \mathcal{F}_T, \quad \mathbb{Q}(A) \quad := \quad \mathbb{E}\left[Z_T^{\phi} \mathbf{1}_A\right] \tag{3.20}$$

defines a probability measure on $(\Omega, \mathcal{F}_T)$, under which the process $X = (X_t)_{t\in[0,T]}$ defined by

$$X_t \quad := \quad B_t - \int_0^t \phi_u \, \mathrm{d}u, \tag{3.21}$$

is a $(\mathcal{F}_t)_{t\in[0,T]}-$Brownian motion (restricted to the time horizon $[0, T]$).

Let us make a number of important comments before proceeding to the proof of this result.

Remark 3.7 (Some useful comments).

1. *In practice, the most efficient way to verify the assumption is to check Novikov's Criterion:*

$$\mathbb{E}\left[e^{\frac{1}{2}\int_0^T \phi_u^2 \, \mathrm{d}u}\right] \quad < \quad +\infty. \tag{3.22}$$

2. *The statement that $\mathbb{Q}$ is a probability measure follows from the fact that $Z_T^{\phi} \ge 0$ and $\mathbb{E}[Z_T^{\phi}] = 1$.*

3. *By linearity and density, (3.20) implies that for any $\mathcal{F}_T-$measurable non-negative variable Y,*

$$\mathbb{E}^{\mathbb{Q}}[Y] = \mathbb{E}\left[YZ_T^{\phi}\right], \quad \textit{and} \quad \mathbb{E}[Y] = \mathbb{E}^{\mathbb{Q}}\left[\frac{Y}{Z_T^{\phi}}\right], \tag{3.23}$$

where $\mathbb{E}^{\mathbb{Q}}$ is expectation under $\mathbb{Q}$. This is useful for transferring computations between $\mathbb{Q}$ and $\mathbb{P}$.

4. *It follows from the martingale property that we also have $\mathbb{E}^{\mathbb{Q}}[Y] = \mathbb{E}[YZ_t^{\phi}]$ for any $t \in [0,T]$ and any non-negative $\mathcal{F}_t-$measurable random variable Y.*

5. *The practical interest of Girsanov's Theorem is as follows: on our original probability space $(\Omega, \mathcal{F}, \mathbb{P})$, computing expectations about X is rather complicated. Moving to $(\Omega, \mathcal{F}, \mathbb{Q})$ turns X into a much simpler object, for which such computations become doable. One can then try to transfer the results back to $(\Omega, \mathcal{F}, \mathbb{P})$, using Formula (3.23). Practical examples will follow...*

6. *Under $\mathbb{Q}$, the process B is of course no longer a Brownian motion! Consequently, computing expectations under $\mathbb{Q}$ typically requires expressing all quantities of interest in terms of X only.*

7. *The result admits the following $T = \infty$ version: suppose that the whole process $Z^{\phi} = (Z_t^{\phi})_{t\geq 0}$ is a martingale (this is the case, for example, when (3.22) holds for each $t \geq 0$). Then, for each $t \geq 0$, the formula (3.20) can be used to define a probability measure $\mathbb{Q}_t$ on $(\Omega, \mathcal{F}_t)$. Moreover, for $0 \leq s \leq t$, the restriction of $\mathbb{Q}_t$ to $\mathcal{F}_s$ coincides with $\mathbb{Q}_s$, since*

$$\forall A \in \mathcal{F}_s, \quad \mathbb{Q}_t(A) = \mathbb{E}[Z_t^{\phi}\mathbf{1}_A] = \mathbb{E}[Z_s^{\phi}\mathbf{1}_A] = \mathbb{Q}_s(A).$$

where the second equality uses the martingale property. Thus, $(\mathbb{Q}_t)_{t\geq 0}$ is a consistent family of probability measures, and the Kolmogorov extension theorem guarantees that these measures are all restrictions of a common probability measure $\mathbb{Q}_\infty$ defined on $\mathcal{F}_\infty := \sigma\left(\bigcup_{t\geq 0}\mathcal{F}_t\right)$. On the probability space $(\Omega, \mathcal{F}_\infty, \mathbb{Q}_\infty)$, the whole process $X = (X_t)_{t\geq 0}$ is then a Brownian motion.

We now turn to the proof of Girsanov's Theorem.

Proof. Let us first settle the special case where $(\phi_t)_{t\in[0,T]}$ satisfies

$$\int_0^T \phi_u^2 \, \mathrm{d}u \leq C, \tag{3.24}$$

for some deterministic $C < \infty$. In particular, for any $\theta \in \mathbb{R}$, the shifted process $\phi + \theta$ satisfies Novikov's Criterion, so $(Z_t^{\phi+\theta})_{t\in[0,T]}$ is a martingale. Thus, for any $0 \leq s \leq t \leq T$, we have,

$$\mathbb{E}\left[Z_s^{\phi+\theta} | \mathcal{F}_t\right] = Z_s^{\phi+\theta}.$$

Since $Z_u^{\phi+\theta} = Z_u^{\phi} e^{\theta X_u - \frac{\theta^2 u}{2}}$ for all $u \geq 0$, we may rewrite this as follows:

$$\mathbb{E}\left[Z_s^{\phi} e^{\theta(X_t - X_s)} | \mathcal{F}_s\right] = e^{\frac{\theta^2}{2}(t-s)} Z_s^{\phi}.$$

In other words, for any $A \in \mathcal{F}_s$, we have

$$\mathbb{E}\left[Z_t^{\phi} e^{\theta(X_t - X_s)}\mathbf{1}_A\right] = e^{\frac{\theta^2}{2}(t-s)}\mathbb{E}\left[Z_s^{\phi}\mathbf{1}_A\right].$$

In view of Item 4 in the above remark, this may be further rewritten in terms of $\mathbb{Q}$ as follows:

$$\mathbb{E}^{\mathbb{Q}}\left[e^{\theta(X_t - X_s)}\mathbf{1}_A\right] = e^{\frac{\theta^2}{2}(t-s)}\mathbb{Q}(A).$$

Taking $A = \Omega$ shows that $X_t - X_s$ has distribution $\mathcal{N}(0, t-s)$, and the product form shows that $X_t - X_s$ is independent of $\mathcal{F}_s$. But this holds for any $0 \le s \le t \le T$, and X is continuous by construction, so $(X_t)_{t\in[0,T]}$ is indeed a $(\mathcal{F}_t)_{t\in[0,T]}-$Brownian motion under $\mathbb{Q}$. To address the general case, we of course introduce the truncated process $\phi^n_t := \phi_t \mathbf{1}_{T_n \ge t}$, where

$$T_n \quad := \quad \inf\left\{ t \ge 0 \colon \int_0^t \phi_u^2 \, \mathrm{d}u \ge n \right\}.$$

Since ϕ^n satisfies the condition (3.24) (with $C = n$), the first part of the proof implies that

$$\mathbb{E}\left[Z^{\phi}_{t\wedge T_n} e^{i\theta(X^n_t - X^n_s)} \mathbf{1}_A \right] \quad = \quad e^{\frac{-\theta^2}{2}(t-s)} \mathbb{E}\left[Z^{\phi}_{s\wedge T_n} \mathbf{1}_A \right],$$

for all $0 \le s \le t \le T$, $\theta \in \mathbb{R}$ and $A \in \mathcal{F}_s$, where $X^n_t := B_t - \int_0^{t\wedge T_n} \phi_u \, \mathrm{d}u$. Now, as $n \to \infty$, we have $T_n \uparrow +\infty$, hence $X^n_t \to X_t$ a.-s. Moreover, by Scheffe's Lemma, the a.-s. convergence $Z^{\phi}_{t\wedge T_n} \to Z^{\phi}_t$ also holds in L^1, for all $t \in [0,T]$. Thus, we may pass to the limit and obtain

$$\mathbb{E}\left[Z^{\phi}_t e^{i\theta(X_t - X_s)} \mathbf{1}_A \right] \quad = \quad e^{\frac{-\theta^2}{2}(t-s)} \mathbb{E}\left[Z^{\phi}_s \mathbf{1}_A \right].$$

Recalling the definition of $\mathbb{Q}$, this precisely means that

$$\mathbb{E}^{\mathbb{Q}}\left[e^{i\theta(X_t - X_s)} \mathbf{1}_A \right] \quad = \quad e^{\frac{-\theta^2}{2}(t-s)} \mathbb{Q}(A).$$

The fact that this is true for every $A \in \mathcal{F}_s$ and every $\theta \in \mathbb{R}$ shows that under $\mathbb{Q}$, the random variable $X_t - X_s$ has law $\mathcal{N}(0, t-s)$ and is independent of $\mathcal{F}_s$, as desired. $\square$

3.6 An application

Here is a good technical exercise to practice with Girsanov's Theorem.

Exercise 3.3 (Joint distribution of $B_t^2, \int_0^t B_s^2 \, \mathrm{d}s$)**.** *In order to understand the joint distribution of B_t^2 and $\int_0^t B_s^2 \, \mathrm{d}s$ for fixed $t \ge 0$, one would naturally like to compute the following Laplace transform:*

$$L_t(a,b) \quad := \quad \mathbb{E}\left[\exp\left\{ -aB_t^2 - \frac{b^2}{2} \int_0^t B_u^2 \, du \right\} \right] \qquad (a, b, t \ge 0).$$

1. *Compute $L_t(a, 0)$ for all $a, t \ge 0$. We henceforth assume that $b > 0$.*

2. *Find $\psi \in \mathbb{M}^1_{loc}$ so that the process Z defined below is a local martingale:*

$$Z_t \quad := \quad \exp\left\{ -b \int_0^t B_u \, \mathrm{d}B_u - \int_0^t \psi_u \, \mathrm{d}u \right\}.$$

3. *Express Z_t in terms of the random variables B_t and $\int_0^t B_u^2 \, \mathrm{d}u$ only, and deduce that*

$$L_t(a,b) \quad = \quad \mathbb{E}\left[Z_t \exp\left\{ \left(\frac{b}{2} - a \right) B_t^2 \right\} \right] \exp\left(-\frac{bt}{2} \right).$$

4. *Fix $t \ge 0$ and construct a probability measure $\mathbb{Q}_t$ on $(\Omega, \mathcal{F}_t)$ under which the process $W = (W_s)_{s\in[0,t]}$ defined by $W_s := B_s + b \int_0^s B_u \, \mathrm{d}u$ is a Brownian motion. Show that for all $s \ge 0$,*

$$B_s \quad = \quad \int_0^s e^{b(u-s)} \, \mathrm{d}W_u.$$

5. *Determine the law of B_t under $\mathbb{Q}_t$ and deduce the formula*

$$L_t(a,b) \quad = \quad \frac{1}{\sqrt{\cosh(bt) + \frac{2a}{b}\sinh(bt)}}.$$

Chapter 4

Stochastic differential equations

4.1 Motivations

An ordinary differential equation (abbreviated as ODE) is an equation involving an unknown function $x = (x_t)_{t\geq 0}$ and its derivative. Such equations are massively used to model physical processes whose evolution in any infinitesimal time-interval $[t, t + \mathrm{d}t]$ only depends on the considered time t, and the current value x_t. In differential notation, they take the form

$$\mathrm{d}x_t \quad = \quad b(t, x_t)\,\mathrm{d}t \tag{4.1}$$

where the function $b\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ describes the underlying dynamics. The classical Picard-Lindelöf theorem gives a simple sufficient condition on b for such an equation to be well-posed, in the sense that it admits a unique solution that starts from each possible initial condition x_0.

Theorem 4.1 (Picard-Lindelöf). *Let $b\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be a measurable function satisfying*

(i) (uniform spatial Lipschitz continuity): there exists a constant $\kappa < \infty$ such that

$$\forall (t, x, y) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}, \quad |b(t, x) - b(t, y)| \quad \leq \quad \kappa |x - y|.$$

(ii) (local integrability in time): $\int_0^t |b(u, 0)|\,\mathrm{d}u < \infty$ for each $t \geq 0$.

Then, for each $z \in \mathbb{R}$, there exists a unique measurable function $x = (x_t)_{t\geq 0}$ satisfying

$$\forall t \geq 0, \quad x_t \quad = \quad z + \int_0^t b(u, x_u)\,\mathrm{d}u. \tag{4.2}$$

In many interesting situations however, the dynamics is intrinsically chaotic and unpredictable: it is then natural to add a random external influence to the above evolution equation, typically driven by a Brownian motion $B = (B_t)_{t\geq 0}$. This naturally leads to the following stochastic analogue of (4.1):

$$\mathrm{d}X_t \quad = \quad b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t,$$

where $X = (X_t)_{t\geq 0}$ is an (unknown) stochastic process, and $b\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $\sigma\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ are deterministic functions called the drift and diffusion coefficients, respectively. The mathematical study of such stochastic differential equations (SDE) is a rich and active topic, to which the present chapter only constitutes a modest introduction.

4.2 Existence and uniqueness

Our first task is to establish a stochastic analogue of the Picard-Lindelöf theorem, giving a simple sufficient condition for the well-posedness of a stochastic differential equation of the form

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \tag{4.3}$$

where $B = (B_t)_{t\geq 0}$ is a given $(\mathcal{F}_t)_{t\geq 0}-$Brownian motion on our filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and $b\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $\sigma\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ are two given measurable functions. By a solution to the stochastic equation (4.3), we will mean a progressive process $X = (X_t)_{t\geq 0}$ defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, satisfying $(b(t, X_t))_{t\geq 0} \in \mathbb{M}^1_{\mathrm{LOC}}$, $(\sigma(t, X_t))_{t\geq 0} \in \mathbb{M}^2_{\mathrm{LOC}}$, and

$$\forall t \geq 0, \quad X_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s = X_t. \tag{4.4}$$

Note that X is then necessarily an Itô process (in particular, it is continuous and adapted).

Theorem 4.2 (Existence and uniqueness). *Let $b, \sigma\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be measurable functions such that*

(i) (Uniform Lipschitz continuity in space): there exists $\kappa < \infty$ such that for all $(t, x, y) \in \mathbb{R}_+ \times \mathbb{R}^2$,

$$|b(t, x) - b(t, y)| \leq \kappa|x - y| \quad \text{and} \quad |\sigma(t, x) - \sigma(t, y)| \leq \kappa|x - y|.$$

(ii) (Local square-integrability in time): for all $t \geq 0$,

$$\int_0^t |b(u, 0)|^2\,\mathrm{d}u < \infty \quad \text{and} \quad \int_0^t |\sigma(u, 0)|^2\,\mathrm{d}u < \infty.$$

Then, for each initial condition $\zeta \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$, there exists a unique (up to indistinguishability) solution X to the SDE (4.3) satisfying $X_0 = \zeta$. Moreover, we have $X \in \mathbb{M}^2$.

As in the proof of the Picard-Lindelöf theorem, the uniqueness uses Gronwall's lemma:

Lemma 4.1 (Gronwall's Lemma). *Let $(x_t)_{t\in[0,T]}$ be a non-negative function in $L^1([0, T])$ satisfying*

$$\forall t \in [0, T], \quad x_t \leq \alpha + \beta \int_0^t x_u\,\mathrm{d}u,$$

for some constants $\alpha, \beta \geq 0$. Then, $x_t \leq \alpha e^{\beta t}$ for all $t \in [0, T]$.

Proof. Set $\kappa = \alpha + \beta \int_0^T x_u\,\mathrm{d}u$. An immediate induction shows that

$$\forall t \in [0, T], \quad x_t \leq \alpha \sum_{k=0}^{n-1} \frac{(\beta t)^k}{k!} + \kappa \frac{(\beta t)^n}{n!},$$

for every $n \in \mathbb{N}$. Sending $n \to \infty$ yields the result. □

Proof of uniqueness in Theorem 4.2. Suppose that X, Y are two solutions of (4.3) satisfying $X_0 = Y_0 = \zeta$. Fix $n \in \mathbb{N}$, and set $T_n := \inf\{t \geq 0\colon \int_0^t (X_u - Y_u)^2\,\mathrm{d}u \geq n\}$. For $t \geq 0$, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^{t\wedge T_n} \sigma(u, X_u)\,\mathrm{d}B_u - \int_0^{t\wedge T_n} \sigma(u, Y_u)\,\mathrm{d}B_u\right)^2\right] &\leq \mathbb{E}\left[\int_0^{t\wedge T_n} (\sigma(u, X_u) - \sigma(u, Y_u))^2\,\mathrm{d}u\right] \\
&\leq \kappa^2\,\mathbb{E}\left[\int_0^{t\wedge T_n} (X_u - Y_u)^2\,\mathrm{d}u\right].
\end{aligned}$$

On the other hand, by Cauchy-Schwartz,

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^{t\wedge T_n} b(u,X_u)\,\mathrm{d}u - \int_0^{t\wedge T_n} b(u,Y_u)\,\mathrm{d}u\right)^2\right] &\le t\mathbb{E}\left[\int_0^{t\wedge T_n}(b(u,X_u)-b(u,Y_u))^2\,\mathrm{d}u\right] \\ &\le \kappa^2 t\,\mathbb{E}\left[\int_0^{t\wedge T_n}(X_u-Y_u)^2\,\mathrm{d}u\right].\end{aligned}$$

Summing these two estimates and using $(u+v)^2 \le 2u^2+2v^2$, we obtain that

$$\mathbb{E}\left[(X_t-Y_t)^2\mathbf{1}_{(t\le T_n)}\right] \le 2\kappa^2(t+1)\int_0^t \mathbb{E}\left[(X_u-Y_u)^2\mathbf{1}_{(u\le T_n)}\right]\mathrm{d}u. \tag{4.5}$$

Note that the right-hand side is finite, by definition of T_n. Thus, we may invoke Gronwall's Lemma with $\alpha=0$, $\beta=2\kappa^2(T+1)$ and $x_t=\mathbb{E}\left[(X_t-Y_t)^2\mathbf{1}_{(t\le T_n)}\right]$ to conclude that

$$\forall t\ge 0,\quad \mathbb{E}\left[(X_t-Y_t)^2\mathbf{1}_{(t\le T_n)}\right] = 0.$$

Sending $n\to\infty$ yields $X\equiv Y$, as desired. □

Proof of existence in Theorem 4.2. Let us construct a sequence of approximate solutions $(X^n)_{n\ge 0}$ in $\mathbb{M}^2$ by setting $X^0\equiv 0$ and then inductively, for each $n\in\mathbb{N}$ and each $t\in\mathbb{R}_+$,

$$X_t^{n+1} := \zeta + \int_0^t \sigma\left(u,X_u^n\right)\mathrm{d}B_u + \int_0^t b\left(u,X_u^n\right)\mathrm{d}u. \tag{4.6}$$

Let us first check that this makes sense. Clearly, $X^0\equiv 0\in\mathbb{M}^2$. Now, fix $n\in\mathbb{N}$, and suppose we know that $X^n\in\mathbb{M}^2$. Then both integrands in (4.6) are in $\mathbb{M}^2$, because our assumptions imply

$$\begin{aligned}|b(u,X_u^n)|^2 &\le 4|b(u,0)|^2+4\kappa^2(X_u^n)^2\\ |\sigma(u,X_u^n)|^2 &\le 4|\sigma(u,0)|^2+4\kappa^2(X_u^n)^2.\end{aligned}$$

Since $t\mapsto\int_0^t\phi_u\,\mathrm{d}u$ and $t\mapsto\int_0^t\phi_u\,\mathrm{d}B_u$ are in $\mathbb{M}^2$ whenever $\phi\in\mathbb{M}^2$, we conclude that $X^{n+1}\in\mathbb{M}^2$, so our induction makes sense. Now, by a similar argument as for (4.5), we have for $n\ge 1$,

$$\forall t\in[0,T],\quad \mathbb{E}\left[\left(X_t^{n+1}-X_t^n\right)^2\right] \le C_T\int_0^t\mathbb{E}\left[\left(X_u^n-X_u^{n-1}\right)^2\right]\mathrm{d}u,$$

where $C_T=4\kappa^2(T+1)$. By an immediate induction, this implies

$$\mathbb{E}\left[\left(X_t^{n+1}-X_t^n\right)^2\right] \le \frac{M_T C_T^n t^{n-1}}{(n-1)!}$$

where $M_T:=\int_0^T\mathbb{E}\left[(X_u^1-X_u^0)^2\right]\mathrm{d}u$. This is more than enough to guarantee that

$$\sum_{n=0}^{\infty}\|X^{n+1}-X^n\|_{\mathbb{M}^2([0,T])} < \infty, \tag{4.7}$$

and hence that the sequence $(X^n)_{n\ge0}$ is convergent in the Hilbert space $\mathbb{M}^2([0,T])$. But this is true for each $T\ge 0$, so the limit is an element $X\in\mathbb{M}^2$, and passing to the limit in (4.6) yields

$$X_t = \zeta + \int_0^t\sigma\left(u,X_u\right)\mathrm{d}B_u + \int_0^t b\left(u,X_u\right)\mathrm{d}u, \tag{4.8}$$

for all $t\ge 0$, as desired. □

Remark 4.1 (Useful comments). *There are several things to note about the theorem.*

1. *Condition (ii) is only used in the proof of existence: it is not needed for the uniqueness part.*
2. *Thanks to Condition (i), the measurability of b, σ only needs to be checked w.r.t. the time variable.*
3. *If Conditions (i) and (ii) are only satisfied on some restricted time horizon $[0, T]$, then the above proof still yields existence and uniqueness of a restricted solution $X = (X_t)_{t\in[0,T]}$.*
4. *In particular, the conclusion of the theorem remains valid if the Lipschitz constant $\kappa = \kappa_t$ appearing in Condition (i) is allowed to depend on time t, as long as $\sup_{t\in[0,T]} \kappa_t < \infty$ for each $T \geq 0$.*
5. *Condition (ii) trivially holds in the homogeneous case where the coefficients $b(t,x), \sigma(t,x)$ do not depend on the time t and more generally, when they depend continuously on t.*
6. *Our construction shows that for each $t \geq 0$, X_t is $\sigma\left(\zeta, (B_s)_{0\leq s\leq t}\right)-$measurable. In other words,*

$$X_t \quad = \quad \Psi_t\left(\zeta, (B_s)_{s\in[0,t]}\right), \tag{4.9}$$

for some measurable $\Psi_t \colon \mathbb{R} \times \mathbb{R}^{[0,t]} \to \mathbb{R}$ which only depends on t and the coefficients b and σ.

Exercise 4.1 (Dependence in the initial condition). *Consider the homogeneous SDE*

$$\mathrm{d}X_t \quad = \quad b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t,$$

where b, σ are Lipschitz functions. Let X, Y be the solutions starting from $X_0, Y_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$, and

$$\psi_t \;:=\; \left(\frac{b(X_t) - b(Y_t)}{X_t - Y_t}\right)\mathbf{1}_{(X_t\neq Y_t)}, \quad \text{and} \quad \phi_t \;:=\; \left(\frac{\sigma(X_t) - \sigma(Y_t)}{X_t - Y_t}\right)\mathbf{1}_{(X_t\neq Y_t)} \tag{4.10}$$

1. *Establish the following identity: almost-surely, for all $t \geq 0$*

$$X_t - Y_t \quad = \quad (X_0 - Y_0)\exp\left\{\int_0^t \left(\psi_u - \frac{\phi_u^2}{2}\right)\mathrm{d}u + \int_0^t \phi_u\,\mathrm{d}B_u\right\}. \tag{4.11}$$

2. *Deduce that almost-surely, the overlap $\{t \geq 0 \colon X_t = Y_t\}$ is either equal to $\emptyset$, or to $\mathbb{R}_+$.*
3. *Prove the existence of a constant $c \in (0, \infty)$ such that for all $t \geq 0$ and all $p \geq 1$,*

$$\mathbb{E}\left[(X_t - Y_t)^p\right] \quad \leq \quad \mathbb{E}\left[(X_0 - Y_0)^p\right] e^{cp^2 t}.$$

4.3 Practical examples

Example 4.1 (Langevin equation). *The following SDE was proposed by Paul Langevin in 1908 to describe the random motion of a small particle in a fluid, due to collisions with the surrounding molecules:*

$$\mathrm{d}X_t \quad = \quad -bX_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t, \tag{4.12}$$

with $b, \sigma \in (0, \infty)$. This is an homogeneous SDE with $b(t,x) = -bx$ and $\sigma(t,x) = \sigma$. The above theorem ensures existence and uniqueness, for any initial condition $\zeta \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$. In fact,

$$\forall t \geq 0, \quad X_t \quad = \quad \zeta e^{-bt} + \sigma\int_0^t e^{-b(t-u)}\,\mathrm{d}B_u, \tag{4.13}$$

as can be checked by differentiating. Let us investigate the long-term behavior of X: the first term on the right-hand side tends to 0 a.-s. as $t \to \infty$, while the second term has law $\mathcal{N}\left(0, \frac{\sigma^2}{2b}(1-e^{-2bt})\right)$, so

$$X_t \xrightarrow[t\to\infty]{d} \mathcal{N}\left(0, \frac{\sigma^2}{2b}\right),$$

independently of the choice of the initial condition ζ. Thus, the process X mixes*: as time increases, the random variable X_t progressively forgets its initial distribution, and approaches a limit $\mathcal{N}\left(0, \frac{\sigma^2}{2b}\right)$.*

This observation suggests to start directly from $\zeta \sim \mathcal{N}\left(0, \frac{\sigma^2}{2b}\right)$. Recall that ζ is always assumed to be $\mathcal{F}_0-$measurable, hence independent of B. The right-hand side of (4.13) then belongs to the Gaussian space $\overline{vect(\zeta, B)}$, so X is a Gaussian process. Its mean is clearly 0, and its covariance is easily computed:

$$\forall s,t \geq 0, \qquad Cov(X_s, X_t) \;=\; \frac{\sigma^2}{2b}e^{-b|t-s|}. \tag{4.14}$$

A continuous centered Gaussian process with this covariance is called an Ornstein-Uhlenbeck process. Since its covariance only depends on $|t-s|$, its distribution is invariant under time-translation:

$$\forall a \geq 0, \quad (X_{t+a})_{t\geq 0} \;\stackrel{d}{=}\; (X_t)_{t\geq 0}.$$

This stationarity is a key property, which explains the importance of the Ornstein-Uhlenbeck process.

Example 4.2 (Geometric Brownian motion). *Fix $\zeta \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$, $\sigma, \mu \in \mathbb{R}$, and consider the SDE*

$$\mathrm{d}X_t \;=\; X_t\left(\sigma\,\mathrm{d}B_t + \mu\,\mathrm{d}t\right), \qquad X_0 = \zeta.$$

This homogeneous SDE with coefficients $b(t,x) = \mu x$ and $\sigma(t,x) = \sigma x$ has a unique solution $X = (X_t)_{t\geq 0}$. In light of what the answer would be in the deterministic case $\sigma = 0$, it is natural to expect a solution of the form $X_t = \zeta e^{Y_t}$, where Y is an Itô process. By Itô's formula,

$$\mathrm{d}\left(\zeta e^{Y_t}\right) \;=\; \zeta e^{Y_t}\left(\mathrm{d}Y_t + \frac{1}{2}\mathrm{d}\langle Y\rangle_t\right).$$

Writing $\mathrm{d}Y_t = \phi_t\,\mathrm{d}B_t + \psi_t\,\mathrm{d}t$ and identifying, we see that $\phi_t = \sigma$ and $\psi_t = \mu - \frac{\sigma^2}{2}$, yielding finally

$$X_t \;:=\; \zeta e^{\sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right)t}.$$

This important process is known as the geometric Brownian motion.

Example 4.3 (Black-Scholes process). *Fix $\zeta \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$ and two deterministic measurable bounded functions $\sigma = (\sigma_t)_{t\geq 0}$ and $\mu = (\mu_t)_{t\geq 0}$. Consider the inhomogeneous SDE*

$$\mathrm{d}X_t \;=\; X_t\left(\sigma_t\,\mathrm{d}B_t + \mu_t\,\mathrm{d}t\right), \qquad X_0 = \zeta.$$

The coefficients $b(t,x) = \mu_t x$ and $\sigma(t,x) = \sigma_t x$ satisfy the Lipshitz and square-integrability conditions, thanks to the boundedness of σ, μ. Thus, there is a unique solution X. As in the above example, it is natural to expect that $X_t = \zeta e^{Y_t}$, where Y is an Itô process. Writing $\mathrm{d}Y_t = \phi_t\,\mathrm{d}B_t + \psi_t\,\mathrm{d}t$, we have

$$\mathrm{d}\left(\zeta e^{Y_t}\right) \;=\; \zeta e^{Y_t}\left(\phi_t\,\mathrm{d}B_t + \frac{1}{2}\phi_t^2\,\mathrm{d}t + \psi_t\,\mathrm{d}t\right).$$

Thus, it suffices to choose $\phi_t = \sigma_t$ and $\psi_t = \mu_t - \frac{\sigma_t^2}{2}$, yielding finally

$$X_t \;=\; \zeta\exp\left\{\int_0^t \sigma_u\,\mathrm{d}B_u + \int_0^t\left(\mu_u - \frac{\sigma_u^2}{2}\right)\mathrm{d}u\right\}.$$

This natural generalization of the geometric Brownian motion is known as a Black-Scholes process.

Exercise 4.2 (Change of variable). *Show that there is a unique Itô process $X = (X_t)_{t\geq 0}$ satisfying*

$$\mathrm{d}X_t \quad = \quad \left(\sqrt{1+X_t^2} + \frac{X_t}{2}\right) \mathrm{d}t + \sqrt{1+X_t^2}\, \mathrm{d}B_t, \qquad X_0 = x,$$

then determine it explicitly by means of the change of variable $Y_t = \mathrm{argsh}(X_t)$.

4.4 Markov property for diffusions

From now on, we focus on the homogeneous case (diffusions). Specifically, we fix two Lipschitz functions $b, \sigma \colon \mathbb{R} \to \mathbb{R}$, and we let $X = (X_t)_{t\geq 0}$ be the unique solution to the well-posed SDE

$$\begin{cases} \mathrm{d}X_t &= b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t \\ X_0 &= \zeta \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}). \end{cases} \tag{4.15}$$

The process X enjoys a fundamental *memoryless* property, which can be described as *independence between the past and future, given the present*. To formalize this, let us recall from (4.9) that we have

$$X_t \quad = \quad \Psi_t(\zeta, (B_u)_{u\in[0,t]}), \tag{4.16}$$

for some deterministic, measurable map $\Psi_t \colon \mathbb{R} \times \mathbb{R}^{[0,t]} \to \mathbb{R}$ which only depends on the coefficients b and σ. The following result shows that for any fixed time $s \geq 0$, the shifted process $\widetilde{X} = (X_{t+s})_{t\geq 0}$ solves a SDE with the same coefficients b and σ, but initialized with $\widetilde{X}_0 = X_s$ and driven by the shifted Brownian motion $\widetilde{B} = (B_{u+s} - B_s)_{u\geq 0}$ (which is independent of $\mathcal{F}_s$).

Theorem 4.3 (Invariance under time shift). *For any $s, t \geq 0$, we have*

$$X_{t+s} \quad = \quad \Psi_t(X_s, (B_{u+s} - B_s)_{u\in[0,t]}).$$

Proof. Fix $s \geq 0$ and define $\widetilde{B}_t := B_{t+s} - B_s$ for $t \geq 0$. Then, the change-of-variable formula

$$\int_s^{t+s} \phi_u\,\mathrm{d}B_u \quad = \quad \int_0^t \phi_{u+s}\,\mathrm{d}\widetilde{B}_u,$$

is clear in the elementary case where $\phi_t(\omega) = X(\omega)\mathbf{1}_{]u,v]}(t)$ with $X \in L^2(\Omega, \mathcal{F}_u, \mathbb{P})$. By linearity and density, it then extends to any $\phi \in \mathbb{M}^2_{\mathrm{LOC}}$. In particular, we can write

$$\begin{aligned} X_{t+s} &= X_s + \int_s^{t+s} b(X_u)\,\mathrm{d}u + \int_s^{t+s} \sigma(X_u)\,\mathrm{d}B_u \\ &= X_s + \int_0^t b(X_{u+s})\,\mathrm{d}u + \int_0^t \sigma(X_{u+s})\,\mathrm{d}\widetilde{B}_u. \end{aligned}$$

In other words, the process $\widetilde{X} := (X_{t+s})_{t\geq 0}$ solves the well-posed SDE

$$\begin{cases} \mathrm{d}\widetilde{X}_t &= b(\widetilde{X}_t)\,\mathrm{d}t + \sigma(\widetilde{X}_t)\,\mathrm{d}\widetilde{B}_t \\ \widetilde{X}_0 &= X_s \end{cases}$$

driven by the Brownian motion $\widetilde{B}$ on the filtered space $(\Omega, \mathcal{F}, (\widetilde{\mathcal{F}}_t)_{t\geq 0}, \mathbb{P})$, where $\widetilde{\mathcal{F}}_t := \mathcal{F}_{t+s}$. But this precisely means that $\widetilde{X}_t = \Psi_t(X_s, (\widetilde{B}_u)_{u\in[0,t]})$ for all $t \geq 0$. □

A direct consequence of Theorem 4.3 (along with the general Remark 1.6 about conditional expectation) is the following fundamental formula, which is known as the Markov property. Given $f \in L^\infty(\mathbb{R})$ and $t \geq 0$, we define a new function $P_t f \colon \mathbb{R} \to \mathbb{R}$ by

$$\forall x \in \mathbb{R}, \quad (P_t f)(x) \quad := \quad \mathbb{E}[f(X_t^x)], \tag{4.17}$$

where X^x denotes the unique solution to the SDE (4.15) with initial condition $\zeta = x$.

Corollary 4.1 (Markov property)**.** *For any $s, t \geq 0$ and any $f \in L^\infty$, we have*

$$\mathbb{E}[f(X_{t+s})|\mathcal{F}_s] \quad = \quad (P_t f)(X_s). \tag{4.18}$$

In particular, the map $(x, A) \mapsto (P_t \mathbf{1}_A)(x)$ is a transition kernel describing the conditional distribution of the future state X_{t+s} given that the current state is $X_s = x$. Note that by successive conditionings, the operators $(P_t)_{t\geq 0}$ actually allow one to recover the law of the entire process X from that of X_0. This motivates a deeper study of the family $(P_t)_{t\geq 0}$, called a semi-group because of the second property in the following lemma.

Lemma 4.2 (Properties of the semi-group)**.** *The family $(P_t)_{t\geq 0}$ enjoys the following properties:*

1. *P_t is a bounded linear operator from $L^\infty(\mathbb{R})$ to $L^\infty(\mathbb{R})$ for each $t \geq 0$.*
2. *We have $P_0 = Id$ and $P_{t+s} = P_t \circ P_s$ for all $s, t \geq 0$.*
3. *If f is continuous, then so is $t \mapsto P_t f(x)$ for each fixed $x \in \mathbb{R}$.*
4. *If f is monotone, then so is $P_t f$ for each $t \geq 0$.*
5. *If f is Lipschitz, then so is $P_t f$ for each $t \geq 0$.*
6. *If σ, b, f are in C_b^k for some $k \in \mathbb{N}$, then so is $P_t f$ for each $t \geq 0$.*

Proof. The linearity of P_t readily follows from that of $\mathbb{E}[\cdot]$. Moreover, for any $f \in L^\infty(\mathbb{R})$ and $t \geq 0$, the function $P_t f \colon x \mapsto \mathbb{E}[f(X_t^x)] = \mathbb{E}[(f \circ \Psi_t)(x, B)]$ is measurable by Fubini's theorem (because $f \circ \Psi_t$ is bounded and measurable). Since $\|P_t f\|_\infty \leq \|f\|_\infty$, the first assertion is proved. The fact that $P_0 = \mathrm{Id}$ is clear since $X_0^x = x$. To prove that $P_{t+s} = P_t \circ P_s$ for all $s, t \geq 0$, we write

$$(P_{t+s} f)(x) \quad = \quad \mathbb{E}[f(X_{t+s}^x)] \;=\; \mathbb{E}[(P_t f)(X_s^x)] \;=\; P_s(P_t f)(x),$$

where the second identity is obtained by taking expectations in (4.18). Finally, if f is continuous and bounded, then so is the random function $t \mapsto f(X_t^x)$, hence also its expectation. The last three assertions are consequences of the identity (4.11), and the details are left to the reader. □

4.5 Generator of a diffusion

It is an easy and pleasant exercise to show that any positive continuous function $t \mapsto p_t$ satisfying $p_{t+s} = p_t p_s$ must take the form $p_t = e^{\lambda t}$, where $\lambda := \lim_{t\to 0} \frac{p_t - p_0}{t}$. By analogy, it is natural to hope for a representation of our semi-group $(P_t)_{t\geq 0}$ under the form

$$P_t \quad = \quad e^{tL}, \tag{4.19}$$

where $L = \lim_{t\to 0} \frac{P_t - P_0}{t}$. Of course, at this level, those identities do not really make any sense and the analogy is purely formal. Nevertheless, this motivates the following fruitful definition.

Definition 4.1 (Generator)**.** *The generator of the semi-group $(P_t)_{t\geq 0}$ is the linear operator L defined by*

$$\forall x \in \mathbb{R}, \qquad (Lf)(x) \quad := \quad \lim_{t\to 0} \frac{(P_t f)(x) - f(x)}{t}, \tag{4.20}$$

for all $f \in L^\infty(\mathbb{R})$ such that the limit exists. Those functions form a vector space denoted, by $Dom(L)$.

The interest of this definition is summed up in the following important result. We recall that $\mathcal{C}_c^2(\mathbb{R})$ denotes the vector space of twice continuously differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$ with compact support, and that this space is dense in $L^\infty(\mathbb{R})$.

Theorem 4.4 (Properties of the generator). *Let $f \in \mathcal{C}_c^2(\mathbb{R})$. Then,*

1. *Lf is well-defined and given by the formula:*

$$\forall x \in \mathbb{R}, \quad Lf(x) \quad = \quad b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x). \tag{4.21}$$

2. *The function $P_t f$ is in $\mathrm{Dom}(L)$ for all $t \geq 0$, and it satisfies Kolmogorov's equation:*

$$\forall x \in \mathbb{R}, \qquad \frac{d}{dt}(P_t f)(x) \quad = \quad (P_t L f)(x) \quad = \quad (L P_t f)(x). \tag{4.22}$$

3. *The process $M = (M_t)_{t\geq 0}$ defined as follows is a continuous square-integrable martingale:*

$$\forall t \geq 0, \qquad M_t \quad := \quad f(X_t) - f(X_0) - \int_0^t (Lf)(X_u)\,\mathrm{d}u. \tag{4.23}$$

Proof. Let us define $Rf := bf' + \frac{1}{2}\sigma^2 f''$. Applying Itô's formula to f and X, we find

$$\begin{aligned}
\mathrm{d}f(X_t) \quad &= \quad f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)\,\mathrm{d}\langle X\rangle_t \\
&= \quad \left(b(X_t)f'(X_t) + \frac{1}{2}\sigma^2(X_t)f''(X_t)\right)\,\mathrm{d}t + f'(X_t)\sigma(X_t)\,\mathrm{d}B_t \\
&= \quad (Rf)(X_t)\,\mathrm{d}t + f'(X_t)\sigma(X_t)\,\mathrm{d}B_t.
\end{aligned}$$

In other words, for all $t \geq 0$,

$$f(X_t) - f(X_0) - \int_0^t (Rf)(X_u)\,\mathrm{d}u \quad = \quad \int_0^t \sigma(X_u)f'(X_u)\,\mathrm{d}B_u. \tag{4.24}$$

Now, the fact that $f \in \mathcal{C}_c(\mathbb{R})$ easily ensures that the function $u \mapsto (Rf)(X_u)$ and $u \mapsto \sigma(X_u)f'(X_u)$ are in $\mathbb{M}^1$ and $\mathbb{M}^2$, respectively. In particular, the right-hand side of (4.24) is a square-integrable martingale. Taking expectations and using Fubini's theorem, we deduce that

$$\mathbb{E}\left[f(X_t)\right] \quad = \quad f(X_0) + \int_0^t \mathbb{E}\left[(Rf)(X_u)\right]\,\mathrm{d}u.$$

Recalling the definition of P_t, we deduce that

$$\forall x \in \mathbb{R}, \qquad (P_t f)(x) \quad = \quad f(x) + \int_0^t (P_u R f)(x)\,\mathrm{d}u. \tag{4.25}$$

But for each fixed $x \in \mathbb{R}$, the function $u \mapsto P_u Rf(x)$ is continuous (Lemma 4.2), so we conclude that $t \mapsto (P_t f)(x)$ is continuously differentiable on $\mathbb{R}_+$, with derivative

$$\frac{\partial}{\partial t}(P_t f)(x) \quad = \quad P_t R f(x). \tag{4.26}$$

Since the left-hand side equals $\lim_{h\to 0} \frac{1}{h}\left((P_h P_t f)(x) - P_t(x)\right)$, we see that $P_t f \in \mathrm{Dom}(L)$ and that $LP_t f = P_t R f$. Finally, taking $t = 0$ shows that $Lf = Rf$, and the proof is complete. $\square$

Remark 4.2 (Extension). *Since X is square-integrable (Theorem 4.2), the definition of $P_t f$ given at (4.17) actually extends to all measurable functions f with quadratic growth (i.e. $|f(x)| \le K(1+x^2)$ for some constant K), and the general properties of $(P_t)_{t\ge 0}$ established in Lemma 4.2 remain valid for this extended definition. In particular, our proof of Theorem 4.4 carries over to the case where $f \in \mathcal{C}_b^2(\mathbb{R})$, meaning that f is twice continuously differentiable with f, f', f'' being bounded.*

Remark 4.3 (Martingales). *The family of martingales given by (4.23) is of course extremely useful for studying the process X. For example, in the case of Brownian motion, we obtain that*

$$t \quad \mapsto \quad f(B_t) - \frac{1}{2}\int_0^t f''(B_u)\,\mathrm{d}u,$$

is a martingale for any $f \in \mathcal{C}_b^2(\mathbb{R})$, generalizing the two simple cases $(B_t)_{t\ge0}$ and $(B_t^2 - t)_{t\ge 0}$.

Remark 4.4 (Fokker-Planck equation). *Writing h_t for the distribution of X_t, the equation (4.22) gives*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} f(z)h_t(\,\mathrm{d}z) \quad = \quad \int_{\mathbb{R}} \left(b(z)f'(z) + \frac{\sigma^2(z)}{2}f''(z)\right) h_t(\,\mathrm{d}z), \tag{4.27}$$

for any $f \in \mathcal{C}_b^2$. Integrating by parts in the sense of distributions, we obtain Fokker-Planck's equation:

$$\frac{\partial h_t}{\partial t} \quad = \quad L^\star h_t, \qquad \text{where} \quad L^\star h = \frac{1}{2}\left(h\sigma^2\right)'' - (bh)'. \tag{4.28}$$

In particular, the equation $L^\star h = 0$ characterizes those distributions h which are stationary.

Exercise 4.3 (Stationary distribution). *Check that the Gaussian distribution $\mathcal{N}(0, \frac{\sigma^2}{2b})$ is stationary for the Langevin equation $\mathrm{d}X_t = -bX_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$ with $b, \sigma > 0$.*

4.6 Connection with partial differential equations

The Kolmogorov equation (4.22) is the starting point of a far-reaching connection between SDEs and partial differential equations (PDEs), which we now uncover. Consider again our diffusion

$$\begin{cases} \mathrm{d}X_t^x &= b(X_t^x)\,\mathrm{d}t + \sigma(X_t^x)\,\mathrm{d}B_t \\ X_0^x &= x, \end{cases} \tag{4.29}$$

where $b, \sigma \colon \mathbb{R} \to \mathbb{R}$ are Lipschitz functions. Now, fix $f \in L^\infty(\mathbb{R})$, and consider the PDE

$$\begin{cases} \dfrac{\partial v}{\partial t}(t,x) &= b(x)\,\dfrac{\partial v}{\partial x}(t,x) + \dfrac{\sigma^2(x)}{2}\,\dfrac{\partial^2 v}{\partial x^2}(t,x) \\ v(0,x) &= f(x), \end{cases} \tag{4.30}$$

where $v \in \mathcal{C}^{1,2}(\mathbb{R}_+ \times \mathbb{R})$ is unknown.

Theorem 4.5 (Connection). *The evolutions (4.29) and (4.30) are linked as follows:*

1. *If v is a bounded solution to the PDE (4.30), then we must have*

$$\forall (t,x) \in \mathbb{R}_+ \times \mathbb{R}, \qquad v(t,x) \quad = \quad \mathbb{E}[f(X_t^x)]. \tag{4.31}$$

2. *If $b, \sigma, f \in \mathcal{C}_b^2$, then conversely, the function (4.31) is a bounded solution to (4.30).*

Proof. Fix $v \in \mathcal{C}^2(\mathbb{R}_+ \times \mathbb{R})$ and $(T,x) \in \mathbb{R}_+ \times \mathbb{R}$, and consider the stochastic process $M = (M_t)_{t\in[0,T]}$ defined by $M_t := v(T-t, X_t^x)$. By Ito's formula, we have

$$\begin{aligned} \mathrm{d}M_t &= -\frac{\partial v}{\partial t}(T-t,X_t^x)\,\mathrm{d}t + \frac{\partial v}{\partial x}(T-t,X_t^x)\,\mathrm{d}X_t^x + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}(T-t,X_t^x)\,\mathrm{d}\langle X\rangle_t^x \\ &= \left(\frac{\partial v}{\partial x}(T-t,X_t^x)b(X_t^x) + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}(T-t,X_t^x)\sigma^2(X_t^x) - \frac{\partial v}{\partial t}(T-t,X_t^x)\right)\mathrm{d}t \\ &\quad + \frac{\partial v}{\partial x}(T-t,X_t^x)\sigma(X_t^x)\,\mathrm{d}B_t. \end{aligned}$$

In particular, if v solves (4.30), then the drift term is zero so M is a local martingale. If moreover v is bounded, then M is bounded, so it is a martingale. Thus, $\mathbb{E}[M_T] = \mathbb{E}[M_0]$, i.e.

$$\mathbb{E}[f(X_T^x)] \;=\; v(T,x).$$

This proves the first claim. Now, assume that $b, \sigma, f \in \mathcal{C}_b^2$. Then the last item in Lemma 4.2 ensures that $P_t f$ is in $\mathcal{C}_b^2(\mathbb{R})$ for all $t \geq 0$, so we may replace f with $P_t f$ in (4.21) to obtain

$$LP_t f(x) \;=\; b(x)(P_t f)'(x) + \frac{1}{2}\sigma^2(x)(P_t f)''(x).$$

Thus, the Kolmogorov equation $\frac{\partial}{\partial t}P_t f(x) = LP_t f(x)$ (along with the trivial identity $P_0 f(x) = f(x)$) precisely assert that the bounded function $v(t,x) := (P_t f)(x)$ solves the PDE (4.30). □

The interest of this connection between SDEs and PDEs is two-fold: on the one hand, one can use tools from PDE theory to understand the distribution of X_t^x. Conversely, the probabilistic representation (4.31) offers a practical way to numerically solve the PDE (4.30), by simulation. Here is an important extension, which incorporates a zero-order term into our PDE.

Theorem 4.6 (Feynman-Kac's formula). *Let $v \in \mathcal{C}^{1,2}(\mathbb{R}_+ \times \mathbb{R})$ be a bounded solution to the PDE*

$$\begin{cases} \dfrac{\partial v}{\partial t}(t,x) &= -h(x)v(t,x) + b(x)\dfrac{\partial v}{\partial x}(t,x) + \dfrac{\sigma^2(x)}{2}\dfrac{\partial^2 v}{\partial x^2}(t,x) \\ v(0,x) &= f(x), \end{cases} \tag{4.32}$$

where $f, h \colon \mathbb{R} \to \mathbb{R}$ are measurable, with h non-negative. Then, we have the representation

$$\forall (t,x) \in \mathbb{R}_+ \times \mathbb{R}, \qquad v(t,x) \;=\; \mathbb{E}\left[f(X_t^x)e^{-\int_0^t h(X_u^x)\,\mathrm{d}u}\right]. \tag{4.33}$$

Proof. Fix $T \geq 0$ and $x \in \mathbb{R}$, and consider the stochastic process $(M_t)_{t\in[0,T]}$ defined by

$$M_t \;:=\; V_t e^{-\int_0^t h(X_u^x)\,\mathrm{d}u}, \qquad \text{with} \qquad V_t := v(T-t, X_t^x).$$

As above, using Ito's formula and the fact that v solves the PDE (4.32), we find for $t \in [0,T]$,

$$\mathrm{d}V_t \;=\; h(X_t^x)V_t\,\mathrm{d}t + \sigma(X_t^x)\frac{\partial v}{\partial x}(T-t,X_t^x)\,\mathrm{d}B_t,$$

Consequently,

$$\begin{aligned} \mathrm{d}M_t &= e^{-\int_0^t h(X_u^x)\,\mathrm{d}u}\left(\mathrm{d}V_t - h(X_t^x)V_t\,\mathrm{d}t\right) \\ &= e^{-\int_0^t h(X_u^x)\,\mathrm{d}u}\sigma(X_t^x)\frac{\partial v}{\partial x}(T-t,X_t^x)\,\mathrm{d}B_t. \end{aligned}$$

Thus, $(M_t)_{t\in[0,T]}$ is a local martingale. Since it is bounded (v is bounded and $h \geq 0$), it is in fact a true martingale. In particular, $\mathbb{E}[M_T] = \mathbb{E}[M_0]$, i.e.

$$\mathbb{E}\left[f(X_T^x)e^{-\int_0^T h(X_u^x)\,\mathrm{d}u}\right] \;=\; v(T,x).$$

Since $T \geq 0$ and $x \in \mathbb{R}$ are arbitrary, the claim is proved. □