AN INTRODUCTION TO STOCHASTIC CALCULUS

Marta Sanz-Solé
Facultat de Matemàtiques
Universitat de Barcelona

September 4, 2017

Contents

1 A review of the basics on stochastic processes

This chapter is devoted to introduce the notion of stochastic processes and some general definitions related with this notion. For a more complete account on the topic, we refer the reader to [11]. Let us start with a definition.

Definition 1.1 *A stochastic process with state space S is a family $\{X_i, i \in I\}$ of random variables $X_i : \Omega \to S$ indexed by a set I.*

For a successful progress in the analysis of such an object, some further structure on the index set I and on the state space S is required. In this course, we shall mainly deal with the particular cases: $I = \mathbb{N}, \mathbb{Z}_+, \mathbb{R}_+$ and S either a countable set or a subset of $\mathbb{R}^d$, $d \geq 1$.
The basic problem statisticians are interested in, is the analysis of the probability law (mostly described by some parameters) of characters exhibited by populations. For a fixed character described by a random variable X, they use a finite number of independent copies of X -a sample of X. For many purposes, it is interesting to have samples of any size and therefore to consider sequences $X_n, n \geq 1$. It is important here to insist on the word *copies*, meaning that the circumstances around the different outcomes of X do not change. It is a static world. Hence, they deal with stochastic processes $\{X_n, n \geq 1\}$ consisting of independent and identically distributed random variables.
This is not the setting we are interested in here. Instead, we would like to give stochastic models for phenomena of the real world which evolve as time goes by. Stochasticity is a choice in front of a complete knowledge and extreme complexity. Evolution, in contrast with statics, is what we observe in most phenomena in Physics, Chemistry, Biology, Economics, Life Sciences, etc.
Stochastic processes are well suited for modeling stochastic evolution phenomena. The interesting cases correspond to families of random variables X_i which are not independent. In fact, the famous classes of stochastic processes are described by means of types of dependence between the variables of the process.

1.1 The law of a stochastic process

The probabilistic features of a stochastic process are gathered in the joint distributions of their variables, as given in the next definition.

Definition 1.2 *The finite-dimensional joint distributions of the process $\{X_i, i \in I\}$ consists of the multi-dimensional probability laws of any finite* *family of random vectors* $X_{i_1}, \ldots, X_{i_m}$*, where* $i_1, \ldots, i_m \in I$ *and* $m \geq 1$ *is arbitrary.*

Let us give an important example.

Example 1.1 *A stochastic process* $\{X_t, t \geq 0\}$ *is said to be Gaussian if its finite-dimensional joint distributions are Gaussian laws.*
Remember that in this case, the law of the random vector $(X_{t_1}, \ldots, X_{t_m})$ *is characterized by two parameters:*

$$\mu(t_1, \ldots, t_m) = E\left(X_{t_1}, \ldots, X_{t_m}\right) = \left(E(X_{t_1}), \ldots, E(X_{t_m})\right)$$
$$\Lambda(t_1, \ldots, t_m) = \left(\mathrm{Cov}(X_{t_i}, X_{t_j})\right)_{1 \leq i,j \leq m}.$$

In the sequel we shall assume that $I \subset \mathbb{R}_+$ and $S \subset \mathbb{R}$, either countable or uncountable, and denote by $\mathbb{R}^I$ the set of real-valued functions defined on I.
A stochastic process $\{X_t, t \geq 0\}$ can be viewed as a random vector

$$X : \Omega \to \mathbb{R}^I.$$

Putting the appropriate σ-field of events in $\mathbb{R}^I$, say $\mathcal{B}(\mathbb{R}^I)$, one can define, as for random variables, the law of the process as the mapping

$$P_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}^I).$$

Mathematical results from measure theory tell us that P_X is defined by means of a procedure of extension of measures on cylinder sets given by the family of all possible finite-dimensional joint distributions. This is a deep result.
In Example 1.1, we have defined a class of stochastic processes by means of the type of its finite-dimensional joint distributions. But, does such an object exist? In other words, could one define stochastic processes giving only its finite-dimensional joint distributions? Roughly speaking, the answer is yes, adding some extra condition. The precise statement is a famous result by Kolmogorov that we now quote.

Theorem 1.1 *Consider a family*

$$\{P_{t_1,\ldots,t_n}, t_1 < \ldots < t_n, n \geq 1, t_i \in I\} \tag{1.1}$$

where:

1. $P_{t_1,\ldots,t_n}$ *is a probability on* $\mathbb{R}^n$*,*

2. *if* $\{t_{i_1} < \ldots < t_{i_m}\} \subset \{t_1 < \ldots < t_n\}$*, the probability law* $P_{t_{i_1}\ldots t_{i_m}}$ *is the marginal distribution of* $P_{t_1\ldots t_n}$*.*

There exists a stochastic process $\{X_t, t \in I\}$ defined in some probability space, such that its finite-dimensional joint distributions are given by (1.1). That is, the law of the random vector $(X_{t_1}, \ldots, X_{t_n})$ is $P_{t_1,\ldots,t_n}$.

One can apply this theorem to Example 1.1 to show the existence of Gaussian processes, as follows.
Let $K : I \times I \to \mathbb{R}$ be a symmetric, nonnegative definite function. That means:

- for any $s, t \in I$, $K(t, s) = K(s, t)$;
- for any natural number n and arbitrary $t_1, \ldots, t_n \in I$, and $x_1, \ldots, x_n \in \mathbb{R}$,

$$\sum_{i,j=1}^{n} K(t_i, t_j) x_i x_j \geq 0.$$

Then there exists a Gaussian process $\{X_t, t \geq 0\}$ such that $E(X_t) = 0$ for any $t \in I$ and $\mathrm{Cov}\,(X_{t_i}, X_{t_j}) = K(t_i, t_j)$, for any $t_i, t_j \in I$.
To prove this result, fix $t_1, \ldots, t_n \in I$ and set $\mu = (0, \ldots, 0) \in \mathbb{R}^n$, $\Lambda = (K(t_i, t_j))_{1 \leq i,j \leq n}$ and

$$P_{t_1,\ldots,t_n} = N(0, \Lambda).$$

We denote by $(X_{t_1}, \ldots, X_{t_n})$ a random vector with law $P_{t_1,\ldots,t_n}$. For any subset $\{t_{i_1}, \ldots, t_{i_m}\}$ of $\{t_1, \ldots, t_n\}$, it holds that

$$A(X_{t_1}, \ldots, X_{t_n}) = (X_{t_{i_1}}, \ldots, X_{t_{i_m}}),$$

with

$$A = \begin{pmatrix} \delta_{t_1,t_{i_1}} & \cdots & \delta_{t_n,t_{i_1}} \\ \cdots & \cdots & \cdots \\ \delta_{t_1,t_{i_m}} & \cdots & \delta_{t_n,t_{i_m}} \end{pmatrix},$$

where $\delta_{s,t}$ denotes the Kronecker Delta function.
By the properties of Gaussian vectors, the random vector $(X_{t_{i_1}}, \ldots, X_{t_{i_m}})$ has an m-dimensional normal distribution, zero mean, and covariance matrix $A\Lambda A^t$. By the definition of A, it is trivial to check that

$$A\Lambda A^t = (K(t_{i_l}, t_{i_k}))_{1 \leq l,k \leq m}.$$

Hence, the assumptions of Theorem 1.1 hold true and the result follows.

1.2 Sample paths

In the previous discussion, stochastic processes are considered as random vectors. In the context of modeling, what matters are the observed values of the process. Observations correspond to fixed values of $\omega \in \Omega$. This new point of view leads to the next definition.

Definition 1.3 *The sample paths of a stochastic process* $\{X_t, t \in I\}$ *are the family of functions indexed by* $\omega \in \Omega$, $X(\omega) : I \to S$, *defined by* $X(\omega)(t) = X_t(\omega)$.

Sample paths are also called *trajectories*.

Example 1.2 *Consider random arrivals of customers at a store. We set our clock at zero and measure the time between two consecutive arrivals. They are random variables* $X_1, X_2, \dots$. *We assume* $X_i > 0$, *a.s. Set* $S_0 = 0$ *and* $S_n = \sum_{j=1}^n X_j$, $n \geq 1$. S_n *is the time of the n-th arrival. The process we would like to introduce is* N_t, *giving the number of customers who have visited the store during the time interval* $[0, t]$, $t \geq 0$.
Clearly, $N_0 = 0$ *and for* $t > 0$, $N_t = k$ *if and only if*

$$S_k \leq t < S_{k+1}.$$

The stochastic process $\{N_t, t \geq 0\}$ takes values on $\mathbb{Z}_+$. Its sample paths are increasing right continuous functions, with jumps at the random times S_n, $n \geq 1$, of size one. It is a particular case of a *counting process*. Sample paths of counting processes are always increasing right continuous functions, their jumps are natural numbers.

Example 1.3 *Evolution of prices of risky assets can be described by real-valued stochastic processes* $\{X_t, t \geq 0\}$ *with continuous, although very rough, sample paths. They are generalizations of the Brownian motion.*
The Brownian motion, also called Wiener process, is a Gaussian process $\{B_t, t \geq 0\}$ *with the following parameters:*

$$\begin{aligned} E(B_t) &= 0 \\ E(B_s B_t) &= s \wedge t, \end{aligned}$$

This defines the finite dimensional distributions and therefore the existence of the process via Kolmogorov's theorem (see Theorem 1.1).

Before giving a heuristic motivation for the preceding definition of Brownian motion, we introduce two further notions.

A stochastic process $\{X_t, t \in I\}$ has *independent increments* if for any $t_1 < t_2 < \ldots < t_k$ the random variables $X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}}$ are independent.

A stochastic process $\{X_t, t \in I\}$ has *stationary increments* if for any $t_1 < t_2$, the law of the random variable $X_{t_2} - X_{t_1}$ is the same as that of $X_{t_2 - t_1}$.

Brownian motion is termed after Robert Brown, a British botanist who observed and reported in 1827 the irregular movements of pollen particles suspended in a liquid. Assume that, when starting the observation, the pollen particle is at position $x = 0$. Denote by B_t the position of (one coordinate) of the particle at time $t > 0$. By physical reasons, the trajectories must be continuous functions and because of the erratic movement, it seems reasonable to say that $\{B_t, t \geq 0\}$ is a stochastic process. It also seems reasonable to assume that the change in position of the particle during the time interval $[t, t+s]$ is independent of its previous positions at times $\tau < t$ and therefore, to assume that the process has independent increments. The fact that such an increment must be stationary is explained by kinetic theory, assuming that the temperature during the experience remains constant.
The model for the law of B_t has been given by Einstein in 1905. More precisely, Einstein's definition of Brownian motion is that of a stochastic processes with independent and stationary increments such that the law of an increment $B_t - B_s$, $s < t$ is Gaussian, zero mean and $E(B_t - B_s)^2 = t - s$. This definition is equivalent to the one given before.

2 The Brownian motion

2.1 Equivalent definitions of Brownian motion

This chapter is devoted to the study of Brownian motion, the process introduced in Example 1.3 that we recall now.

Definition 2.1 *The stochastic process* $\{B_t, t \geq 0\}$ *is a one-dimensional Brownian motion if it is Gaussian, zero mean and with covariance function given by* $E(B_t B_s) = s \wedge t$.

The existence of such process is ensured by Kolmogorov's theorem. Indeed, it suffices to check that

$$(s,t) \to \Gamma(s,t) = s \wedge t$$

is nonnegative definite. That means, for any $t_i \geq 0$ and any real numbers a_i, $i, j = 1, \dots, m$,

$$\sum_{i,j=1}^{m} a_i a_j \Gamma(t_i, t_j) \geq 0.$$

But

$$s \wedge t = \int_0^\infty 1_{[0,s]}(r)\, 1_{[0,t]}(r)\, dr.$$

Hence,

$$\begin{aligned}
\sum_{i,j=1}^{m} a_i a_j (t_i \wedge t_j) &= \sum_{i,j=1}^{m} a_i a_j \int_0^\infty 1_{[0,t_i]}(r)\, 1_{[0,t_j]}(r)\, dr \\
&= \int_0^\infty \left(\sum_{i=1}^{m} a_i 1_{[0,t_i]}(r) \right)^2 dr \geq 0.
\end{aligned}$$

Notice also that, since $E(B_0^2) = 0$, the random variable B_0 is zero almost surely.

Each random variable B_t, $t > 0$, of the Brownian motion has a density, and it is

$$p_t(x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t}),$$

while for $t = 0$, its "density" is a Dirac mass at zero, $\delta_{\{0\}}$.
Differentiating $p_t(x)$ once with respect to t, and then twice with respect to x easily yields

$$\begin{aligned}
\frac{\partial}{\partial t} p_t(x) &= \frac{1}{2} \frac{\partial^2}{\partial x^2} p_t(x) \\
p_0(x) &= \delta_{\{0\}}.
\end{aligned}$$

This is the heat equation on $\mathbb{R}$ with initial condition $p_0(x) = \delta_{\{0\}}$. That means, as time evolves, the density of the random variables of the Brownian motion behaves like a diffusive physical phenomenon.
There are equivalent definitions of Brownian motion, as the one given in the nest result.

Proposition 2.1 *A stochastic process $\{X_t, t \geq 0\}$ is a Brownian motion if and only if*

(i) $X_0 = 0, a.s.,$

(ii) for any $0 \leq s < t$, the random variable $X_t - X_s$ is independent of the σ-field generated by $X_r, 0 \leq r \leq s$, $\sigma(X_r, 0 \leq r \leq s)$ and $X_t - X_s$ is a $N(0, t-s)$ random variable.

Proof: Let us assume first that $\{X_t, t \geq 0\}$ is a *Brownian motion*. Then $E(X_0^2) = 0$. Thus, $X_0 = 0$ a.s..
Let H_s and $\tilde{H}_s$ be the vector spaces included in $L^2(\Omega)$ spanned by $(X_r, 0 \leq r \leq s)$ and $(X_{s+u} - X_s, u \geq 0)$, respectively. Since for any $0 \leq r \leq s$

$$E\left(X_r\left(X_{s+u} - X_s\right)\right) = 0,$$

H_s and $\tilde{H}_s$ are orthogonal in $L^2(\Omega)$. Consequently, $X_t - X_s$ is independent of the σ-field $\sigma(X_r, 0 \leq r \leq s)$.
Since linear combinations of Gaussian random variables are also Gaussian, $X_t - X_s$ is normal, and $E(X_t - X_s) = 0$,

$$E(X_t - X_s)^2 = t + s - 2s = t - s.$$

This ends the proof of properties (i) and (ii).
Assume now that (i) and (ii) hold true. Then the finite dimensional distributions of $\{X_t, t \geq 0\}$ are multidimensional normal, since they are obtained by linear transformation of random vectors with Gaussian independent components. Moreover, for $0 \leq s \leq t$,

$$\begin{aligned} E(X_t X_s) &= E\left((X_t - X_s + X_s)X_s\right) = E\left((X_t - X_s)X_s\right) + E(X_s^2) \\ &= E(X_t - X_s)E(X_s) + E(X_s^2) = E(X_s^2) = s = s \wedge t. \end{aligned}$$

Remark 2.1 *We shall see later that Brownian motion has continuous sample paths. The description of the process given in the preceding proposition tell us that such a process is a model for a random evolution which starts from $x = 0$ at time $t = 0$, such that the qualitative change on time increments only depends on their length (stationary law), and that the future evolution of the process is independent of its past (Markov property).*

Remark 2.2 *The Brownian motion possesses several invariance properties. Let us mention some of them.*

- *If* $B = \{B_t, t \ge 0\}$ *is a Brownian motion, so is* $-B = \{-B_t, t \ge 0\}$.
- *For any* $\lambda > 0$, *the process* $B^\lambda = \left\{\frac{1}{\lambda} B_{\lambda^2 t}, t \ge 0\right\}$ *is also a Brownian motion. This means that zooming in or out, we will observe the same sort of behaviour. This is called the scaling property of Brownian motion.*
- *For any* $a > 0$, $B^{+a} = \{B_{t+a} - B_a,\ t \ge 0\}$ *is a Brownian motion.*

2.2 A construction of Brownian motion

There are several ways to obtain a Brownian motion. Here we shall give P. Lévy's construction, which also provides the continuity of the sample paths. Before going through the details of this construction, we mention an alternate.

Brownian motion as limit of a random walk

Let $\{\xi_j, j \in \mathbb{N}\}$ be a sequence of independent, identically distributed random variables, with mean zero and variance $\sigma^2 > 0$. Consider the sequence of partial sums defined by $S_0 = 0$, $S_n = \sum_{j=1}^n \xi_j$. The sequence $\{S_n, n \ge 0\}$ is a Markov chain, and also a martingale.

Let us consider the continuous time stochastic process defined by linear interpolation of $\{S_n, n \ge 0\}$, as follows. For any $t \ge 0$, let $[t]$ denote its integer value. Then set

$$Y_t = S_{[t]} + (t - [t])\xi_{[t]+1}, \tag{2.1}$$

for any $t \ge 0$.

The next step is to scale the sample paths of $\{Y_t, t \ge 0\}$. By analogy with the scaling in the statement of the central limit theorem, we set

$$B_t^{(n)} = \frac{1}{\sigma\sqrt{n}} Y_{nt}, \tag{2.2}$$

$t \ge 0$.

A famous result in probability theory -Donsker theorem- tell us that the sequence of processes $B_t^{(n)}, t \ge 0\}$, $n \ge 1$, converges in law to the Brownian motion. The reference sample space is the set of continuous functions vanishing at zero. Hence, proving the statement, we obtain continuity of the sample paths of the limit.

Donsker theorem is the infinite dimensional version of the above mentioned central limit theorem. Considering $s = \frac{k}{n}$, $t = \frac{k+1}{n}$, the increment $B_t^{(n)} - B_s^{(n)} = \frac{1}{\sigma\sqrt{n}}\xi_{k+1}$ is a random variable, with mean zero and variance $t -$

s. Hence $B_t^{(n)}$ is not that far from the Brownian motion, and this is what Donsker's theorem proves.

P. Lévy's construction of Brownian Motion

An important ingredient in the procedure is a sequence of functions defined on $[0,1]$, termed *Haar functions*, defined as follows:

$$\begin{aligned} h_0(t) &= 1, \\ h_n^k(t) &= 2^{\frac{n}{2}} 1_{[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}[} - 2^{\frac{n}{2}} 1_{[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}[}, \end{aligned}$$

where $n \geq 1$ and $k \in \{0, 1, \ldots, 2^n - 1\}$.
The set of functions (h_0, h_n^k) is a CONS of $L^2([0,1], \mathcal{B}([0,1]), \lambda)$, where λ stands for the Lebesgue measure. Consequently, for any $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$, we can write the expansion

$$f = \langle f, h_0 \rangle h_0 + \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle h_n^k, \tag{2.3}$$

where the notation $\langle \cdot, \cdot \rangle$ means the inner product in $L^2([0,1], \mathcal{B})([0,1]), \lambda)$.
Using (2.3), we define an isometry between $L^2([0,1], \mathcal{B}([0,1]), \lambda)$ and $L^2(\Omega, \mathcal{F}, P)$ as follows. Consider a family of independent random variables with law $N(0,1)$, (N_0, N_n^k). Then, for $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$, set

$$I(f) = \langle f, h_0 \rangle N_0 + \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle N_n^k.$$

Clearly,

$$E\left(I(f)^2\right) = \|f\|_2^2.$$

Hence I defines an isometry between the space of random variables $L^2(\Omega, \mathcal{F}, P)$ and $L^2([0,1], \mathcal{B}([0,1]), \lambda)$. Moreover, since

$$I(f) = \lim_{m \to \infty} \langle f, h_0 \rangle N_0 + \sum_{n=1}^{m} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle N_n^k,$$

the randon variable $I(f)$ is $N(0, \|f\|_2^2)$ and by Parseval's identity

$$E(I(f)I(g)) = \langle f, g \rangle, \tag{2.4}$$

for any $f, g \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$.

Theorem 2.1 *The process $B = \{B_t = I(1_{[0,t]}), t \in [0,1]\}$ defines a Brownian motion indexed by $[0,1]$. Moreover, the sample paths are continuous, almost surely.*

Proof: By construction $B_0 = 0$. Notice that for $0 \le s \le t \le 1$, $B_t - B_s = I(1_{]s,t]})$. Hence, by virtue or (2.4) the process $B_t - B_s$ is independent of any B_r, $0 < r < s$, and $B_t - B_s$ has a $N(0, t-s)$ law. By Proposition 2.1, we obtain the first statement.
Our next aim is to prove that the series appearing in

$$\begin{aligned} B_t = I\left(1_{[0,t]}\right) &= \langle 1_{[0,t]}, h_0\rangle N_0 + \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} \langle 1_{[0,t]}, h_n^k\rangle N_n^k \\ &= g_0(t) N_0 + \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k \end{aligned} \tag{2.5}$$

converges uniformly, a.s. In the last term we have introduced the *Schauder functions* defined as follows.

$$\begin{aligned} g_0(t) &= \langle 1_{[0,t]}, h_0\rangle = t, \\ g_n^k(t) &= \langle 1_{[0,t]}, h_n^k\rangle = \int_0^t h_n^k(s) ds, \end{aligned}$$

for any $t \in [0,1]$.
By construction, for any fixed $n \ge 1$, the functions $g_n^k(t)$, $k = 0, \ldots, 2^n - 1$, are positive, have disjoint supports and

$$g_n^k(t) \le 2^{-\frac{n}{2}}.$$

Thus,

$$\sup_{t \in [0,1]} \left| \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k \right| \le 2^{-\frac{n}{2}} \sup_{0 \le k \le 2^n - 1} |N_n^k|.$$

The next step consists in proving that $|N_n^k|$ is bounded by some constant depending on n such that when multiplied by $2^{-\frac{n}{2}}$ the series with these terms converges.
For this, we will use a result on *large deviations* for Gaussian measures along with the first Borel-Cantelli lemma.

Lemma 2.1 *For any random variable X with law $N(0,1)$ and for any $a \ge 1$,*

$$P(|X| \ge a) \le e^{-\frac{a^2}{2}}.$$

Proof: We clearly have

$$\begin{aligned} P(|X| \geq a) &= \frac{2}{\sqrt{2\pi}} \int_a^\infty dx e^{-\frac{x^2}{2}} \leq \frac{2}{\sqrt{2\pi}} \int_a^\infty dx \frac{x}{a} e^{-\frac{x^2}{2}} \\ &= \frac{2}{a\sqrt{2\pi}} e^{-\frac{a^2}{2}} \leq e^{-\frac{a^2}{2}}, \end{aligned}$$

where we have used that $1 \leq \frac{x}{a}$ and $\frac{2}{a\sqrt{2\pi}} \leq 1$.

□

We now move to the Borel-Cantelli's based argument. By the preceding lemma,

$$\begin{aligned} P\left(\sup_{0 \leq k \leq 2^n - 1} |N_n^k| > 2^{\frac{n}{4}} \right) &\leq \sum_{k=0}^{2^n-1} P\left(|N_n^k| > 2^{\frac{n}{4}} \right) \\ &\leq 2^n \exp\left(-2^{\frac{n}{2}-1} \right). \end{aligned}$$

It follows that

$$\sum_{n=1}^\infty P\left(\sup_{0 \leq k \leq 2^n - 1} |N_n^k| > 2^{\frac{n}{4}} \right) < +\infty,$$

and by the first Borel-Cantelli lemma

$$P\left(\liminf_{n \to \infty} \left\{ \sup_{0 \leq k \leq 2^n - 1} |N_n^k| \leq 2^{\frac{n}{4}} \right\} \right) = 1.$$

That is, a.s., there exists n_0, which may depend on ω, such that

$$\sup_{0 \leq k \leq 2^n - 1} |N_n^k| \leq 2^{\frac{n}{4}}$$

for any $n \geq n_0$. Hence, we have proved

$$\sup_{t \in [0,1]} \left| \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k \right| \leq 2^{-\frac{n}{2}} \sup_{0 \leq k \leq 2^n - 1} |N_n^k| \leq 2^{-\frac{n}{4}},$$

a.s., for n big enough, which proves the a.s. uniform convergence of the series (2.5).

□

Next we discuss how from Theorem 1.1 we can get a Brownian motion indexed by $\mathbb{R}_+$. To this end, let us consider a sequence $B^k, k \geq 1$ consisting of *independent* Brownian motions indexed by $[0,1]$. That means, for each $k \geq 1$, $B^k = \{B_t^k, t \in [0,1]\}$ is a Brownian motion and for different values of k, they are independent. Then we define a Brownian motion recursively as follows. Let $k \geq 1$; for $t \in [k, k+1]$ set

$$B_t = B_1^1 + B_1^2 + \cdots + B_1^k + B_{t-k}^{k+1}.$$

Such a process is Gaussian, zero mean and $E(B_t B_s) = s \wedge t$. Hence it is a Brownian motion.

we end this section by giving the notion of *d-dimensional Brownian motion*, for a natural number $d \geq 1$. For $d = 1$ it is the process we have seen so far. For $d > 1$, it is the process defined by

$$B_t = (B_t^1, B_t^2, \ldots, B_t^d),\ t \geq 0,$$

where the components are independent one-dimensional Brownian motions.

2.3 Path properties of Brownian motion

We already know that the trajectories of Brownian motion are a.s. continuous functions. However, since the process is a model for particles wandering erratically, one expects rough behaviour. This section is devoted to prove some results that will make more precise these facts.
Firstly, it is possible to prove that the sample paths of Brownian motion are γ-Hölder continuous. The main tool for this is Kolmogorov's continuity criterion (see e.g. [10][Theorem 2.1]):

Proposition 2.2 *Let $\{X_t, t \geq 0\}$ be a stochastic process satisfying the following property: for some positive real numbers α, β and C,*

$$E\left(|X_t - X_s|^\alpha\right) \leq C|t-s|^{1+\beta}.$$

Then almost surely, the sample paths of the process are γ-Hölder continuous with $\gamma < \frac{\beta}{\alpha}$.

The law of the random variable $B_t - B_s$ is $\mathrm{N}(0, t-s)$. Thus, it is possible to compute the moments, and we have

$$E\left((B_t - B_s)^{2k}\right) = \frac{(2k)!}{2^k k!}(t-s)^k,$$

for any $k \in \mathbb{N}$. Therefore, Proposition 2.2 yields that almost surely, the sample paths of the Brownian motion are γ–Hölder continuous with $\gamma \in (0, \frac{1}{2})$.

Nowhere differentiability

We shall prove that the exponent $\gamma = \frac{1}{2}$ above is sharp. As a consequence we will obtain a celebrated result by Dvoretzky, Erdös and Kakutani telling that a.s. the sample paths of Brownian motion are not differentiable. We gather these results in the next theorem.

Theorem 2.2 *Fix any $\gamma \in (\frac{1}{2}, 1]$; then a.s. the sample paths of $\{B_t, t \geq 0\}$ are nowhere Hölder continuous with exponent γ.*

Proof: Let $\gamma \in (\frac{1}{2}, 1]$ and assume that a sample path $t \to B_t(\omega)$ is γ-Hölder continuous at $s \in [0, 1)$. Then

$$|B_t(\omega) - B_s(\omega)| \leq C|t - s|^\gamma,$$

for any $t \in [0, 1]$ and some constant $C > 0$.
Let n big enough and let $i = [ns] + 1$; by the triangular inequality

$$\begin{aligned}\left|B_{\frac{j}{n}}(\omega) - B_{\frac{j+1}{n}}(\omega)\right| &\leq \left|B_s(\omega) - B_{\frac{j}{n}}(\omega)\right| \\ &\quad + \left|B_s(\omega) - B_{\frac{j+1}{n}}(\omega)\right| \\ &\leq C\left(\left|s - \frac{j}{n}\right|^\gamma + \left|s - \frac{j+1}{n}\right|^\gamma\right).\end{aligned}$$

Hence, by restricting $j = i, i+1, \ldots, i+N-1$, we obtain

$$\left|B_{\frac{j}{n}}(\omega) - B_{\frac{j+1}{n}}(\omega)\right| \leq C\left(\frac{N}{n}\right)^\gamma = \frac{M}{n^\gamma}.$$

Define

$$A^i_{M,n} = \left\{\left|B_{\frac{j}{n}}(\omega) - B_{\frac{j+1}{n}}(\omega)\right| \leq \frac{M}{n^\gamma}, j = i, i+1, \ldots, i+N-1\right\}.$$

We have seen that the set of trajectories where $t \to B_t(\omega)$ is γ-Hölder continuous at s is included in

$$\cup_{M=1}^\infty \cup_{k=1}^\infty \cap_{n=k}^\infty \cup_{i=1}^n A^i_{M,n}.$$

Next we prove that this set has null probability. Indeed,

$$\begin{aligned}P\left(\cap_{n=k}^\infty \cup_{i=1}^n A^i_{M,n}\right) &\leq \liminf_{n\to\infty} P\left(\cup_{i=1}^n A^i_{M,n}\right) \\ &\leq \liminf_{n\to\infty} \sum_{i=1}^n P\left(A^i_{M,n}\right) \\ &\leq \liminf_{n\to\infty} n\left(P\left(\left|B_{\frac{1}{n}}\right| \leq \frac{M}{n^\gamma}\right)\right)^N,\end{aligned}$$

where we have used that the random variables $B_{\frac{j}{n}} - B_{\frac{j+1}{n}}$ are $N\left(0, \frac{1}{n}\right)$ and independent. But

$$P\left(\left|B_{\frac{1}{n}}\right| \leq \frac{M}{n^\gamma}\right) = \sqrt{\frac{n}{2\pi}} \int_{-Mn^{-\gamma}}^{Mn^{-\gamma}} e^{-\frac{nx^2}{2}} dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-Mn^{\frac{1}{2}-\gamma}}^{Mn^{\frac{1}{2}-\gamma}} e^{-\frac{x^2}{2}} dx \leq Cn^{\frac{1}{2}-\gamma}.$$

Hence, by taking N such that $N\left(\gamma - \frac{1}{2}\right) > 1$,

$$P\left(\cap_{n=k}^{\infty} \cup_{i=1}^{n} A_{M,n}^{i}\right) \leq \liminf_{n\to\infty} nC\left[n^{\frac{1}{2}-\gamma}\right]^N = 0.$$

since this holds for any k, M, se get

$$P\left(\cup_{M=1}^{\infty} \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \cup_{i=1}^{n} A_{M,n}^{i}\right) = 0.$$

This ends the proof of the theorem.

□

Notice that, if a.s. the sample paths of Brownian motion were differentiable at some point, they would also be γ-Hölder continuous of degree $\gamma = 1$. This contradicts the preceding theorem.

What happens for $\gamma = \frac{1}{2}$? The answer to this question comes as a consequence of Paul Lévy's modulus of continuity result.
For a function $f : [0, \infty) \to \mathbb{R}$, the modulus of continuity is a way to describe its local smoothness. More precisely, let $\sigma : [0, \infty) \to [0, \infty)$ be such that $\sigma(0) = 0$ and strictly increasing. The function σ is a modulus of continuity for the function f if for any $0 \leq s \leq t$,

$$|f(s) - f(t)| \leq \sigma(|t - s|).$$

Theorem 2.3 *Let $\{B_t, t \geq 0\}$ be a Brownian motion. Then,*

$$P\left(\limsup_{h\to 0} \frac{\sup_{0\leq t\leq 1-h} |B(t+h) - B(t)|}{\sqrt{2h|\log h|}} = 1\right) = 1.$$

As a consequence, the sample paths of a Brownian motion are not Hölder continuous of degree $\gamma = \frac{1}{2}$. Indeed,

$$\frac{\sup_{0\leq t\leq 1-h} |B(t+h) - B(t)|}{\sqrt{h}}$$
$$= \frac{\sup_{0\leq t\leq 1-h} |B(t+h) - B(t)|}{\sqrt{2h|\log h|}} \sqrt{2}|\log h|,$$

which tends to ∞ as $h \to 0$.

Quadratic variation

The notion of quadratic variation provides a measure of the roughness of a function. Existence of variations of different orders are also important in procedures of approximation via a Taylor expansion and also in the development of infinitesimal calculus. We will study here the existence of quadratic variation, i.e. variation of order two, for the Brownian motion. As shall be discussed in more detail in the next chapter, this provides an explanation to the fact that rules of Itô's stochastic calculus are different from those of the classical differential deterministic calculus.

Fix a finite interval $[0,T]$ and consider the sequence of partitions given by the points $\Pi_n = (t_0^n = 0 \le t_1^n \le \ldots \le t_{r_n}^n = T)$, $n \ge 1$. We assume that

$$\lim_{n\to\infty} |\Pi_n| = 0,$$

where $|\Pi_n|$ denotes the *norm* of the partition Π_n:

$$|\Pi_n| = \sup_{j=0,\ldots,r_n-1} (t_{j+1} - t_j)$$

Set $\Delta_k B = B_{t_k^n} - B_{t_{k-1}^n}$. Under the preceding conditions on the sequence $(\Pi_n)_{n\ge 1}$ we have the following.

Proposition 2.3 *The sequence* $\{\sum_{k=1}^{r_n} (\Delta_k B)^2, n \ge 1\}$ *converges in* $L^2(\Omega)$ *to the deterministic random variable* T. *That is,*

$$\lim_{n\to\infty} E\left[\left(\sum_{k=1}^{r_n} (\Delta_k B)^2 - T\right)^2\right] = 0.$$

Proof: For the sake of simplicity, we shall omit the dependence on n. Set $\Delta_k t = t_k - t_{k-1}$. Notice that the random variables $(\Delta_k B)^2 - \Delta_k t$, $k = 1, \ldots, n$, are independent and centered. Thus,

$$\begin{aligned} E\left[\left(\sum_{k=1}^{r_n} (\Delta_k B)^2 - T\right)^2\right] &= E\left[\left(\sum_{k=1}^{r_n} \left((\Delta_k B)^2 - \Delta_k t\right)\right)^2\right] \\ &= \sum_{k=1}^{r_n} E\left[\left((\Delta_k B)^2 - \Delta_k t\right)^2\right] \\ &= \sum_{k=1}^{r_n} \left[3(\Delta_k t)^2 - 2(\Delta_k t)^2 + (\Delta_k t)^2\right] \\ &= 2\sum_{k=1}^{r_n} (\Delta_k t)^2 \le 2T|\Pi_n|, \end{aligned}$$

which clearly tends to zero as n tends to infinity. $\square$

This proposition, together with the continuity of the sample paths of Brownian motion yields

$$\sup_n \sum_{k=1}^{r_n} |\Delta_k B| = \infty, \text{a.}s..$$

Therefore, a.s. Brownian motion has infinite variation.
Indeed, assume that $V := \sup_n \sum_{k=1}^{r_n} |\Delta_k B| < \infty$. Then

$$\begin{aligned}\sum_{k=1}^{r_n} (\Delta_k B)^2 &\leq \sup_k |\Delta_k B| \left(\sum_{k=1}^{r_n} |\Delta_k B| \right) \\ &\leq V \sup_k |\Delta_k B|.\end{aligned}$$

We obtain $\lim_{n\to\infty} \sum_{k=1}^{r_n} (\Delta_k B)^2 = 0$. a.s., which contradicts the result proved in Proposition 2.3.

Remark 2.3 *In the particular case $\Pi_n \subset \Pi_{n+1}$, $n \geq 1$, the result on the quadratic variation of Brownian motion given in the preceding Proposition can be improved. In fact, the following stronger statement holds:*

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} (\Delta_k B)^2 = T, \ \ a.s. \tag{2.6}$$

For example, one can consider $\Pi_n = ((kT)2^{-n}, k = 0, \dots, 2^n)$.

Assume that the sequence of partitions $(\Pi_n)_{n\geq 1}$ satisfies the assumptions of Proposition 2.3. In addition, we assume that there exists $\gamma \in (0,1)$ such that $\sum_{n\geq 1} |\Pi_n|^\gamma < \infty$. Then (2.6) holds. Indeed, let $\lambda > 0$. Using Chebychev's inequality and the computations of the proof of Proposition 2.3, we have

$$\begin{aligned}P\left\{ \left| \sum_{k=1}^{r_n} (\Delta_k B)^2 - T \right| > \lambda \right\} &\leq \lambda^{-2} E \left(\left| \sum_{k=1}^{r_n} (\Delta_k B)^2 - T \right|^2 \right) \\ &\leq C \lambda^{-2} |\Pi_n|.\end{aligned}$$

Choose $\lambda = |\Pi_n|^{\frac{1-\gamma}{2}}$. Then

$$\sum_{n\geq 1} P\left\{ \left| \sum_{k=1}^{r_n} (\Delta_k B)^2 - T \right| > \lambda \right\} \leq C \sum_{n\geq 1} |\Pi_n|^\gamma < \infty.$$

Thus, (2.6) follows from Borel-Cantelli's Lemma.

The notion of quadratic variation presented before does not coincide with the usual notion of quadratic variation for real functions. In the latter case, there is no restriction on the partitions. Actually, for Brownian motion the following result holds:

$$\sup_{\Pi} \sum_{t_k \in \Pi} (\Delta_k B)^2 = +\infty, \ \ a.s.,$$

where the supremum is on the set of all partition of the interval $[0, T]$.

2.4 The martingale property of Brownian motion

We start this section by giving the definition of martingale for continuous time stochastic processes. First, we introduce the appropriate notion of filtration, as follows.
A family $\{\mathcal{F}_t, t \geq 0\}$ of sub σ–fields of $\mathcal{F}$ is termed a *filtration* if

1. $\mathcal{F}_0$ contains all the sets of $\mathcal{F}$ of null probability,

2. For any $0 \leq s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$.

If in addition

$$\cap_{s>t} \mathcal{F}_s = \mathcal{F}_t,$$

for any $t \geq 0$, the filtration is said to be *right-continuous.*

Definition 2.2 *A stochastic process* $\{X_t, t \geq 0\}$ *is a martingale with respect to the filtration* $\{\mathcal{F}_t, t \geq 0\}$ *if each variable belongs to* $L^1(\Omega)$ *and moreover*

1. X_t *is* $\mathcal{F}_t$*–measurable for any* $t \geq 0$,

2. for any $0 \leq s \leq t$, $E(X_t/\mathcal{F}_s) = X_s$.

If the equality in (2) is replaced by $\leq$ (respectively, $\geq$), we have a supermartingale (respectively, a submartingale).
Given a stochastic process $\{X_t, t \geq 0\}$, there is a natural way to define a filtration by considering

$$\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t), \ t \geq 0.$$

To ensure that the above property (1) for a filtration holds, one needs to complete the σ-field. In general, there is no reason to expect right-continuity. However, for the Brownian motion, the natural filtration possesses this property.

A stochastic process X with X_0 constant and constant mean, independent increments possesses the martingale property with respect to the natural filtration. Indeed, for $0 \le s \le t$,

$$E(X_t - X_s/\mathcal{F}_s) = E(X_t - X_s) = 0.$$

Hence, a Brownian motion possesses the martingale property with respect to the natural filtration.

Other examples of martingales with respect to the same filtration, related with the Brownian motion are

1. $\{B_t^2 - t, t \ge 0\}$,
2. $\{\exp\left(aB_t - \frac{a^2 t}{2}\right), t \ge 0\}$.

Indeed, for the first example, let us consider $0 \le s \le t$. Then,

$$\begin{aligned} E\left(B_t^2/\mathcal{F}_s\right) &= E\left((B_t - B_s + B_s)^2/\mathcal{F}_s\right) \\ &= E\left((B_t - B_s)^2/\mathcal{F}_s\right) + 2E\left((B_t - B_s)B_s/\mathcal{F}_s\right) \\ &\quad + E\left(B_s^2/\mathcal{F}_s\right). \end{aligned}$$

Since $B_t - B_s$ is independent of $\mathcal{F}_s$, owing to the properties of the conditional expectation, we have

$$\begin{aligned} E\left((B_t - B_s)^2/\mathcal{F}_s\right) &= E\left((B_t - B_s)^2\right) = t - s, \\ E\left((B_t - B_s)B_s/\mathcal{F}_s\right) &= B_s E\left(B_t - B_s/\mathcal{F}_s\right) = 0, \\ E\left(B_s^2/\mathcal{F}_s\right) &= B_s^2. \end{aligned}$$

Consequently,

$$E\left(B_t^2 - B_s^2/\mathcal{F}_s\right) = t - s.$$

For the second example, we also use the property of independent increments, as follows:

$$\begin{aligned} E\left(\exp\left(aB_t - \frac{a^2 t}{2}\right)/\mathcal{F}_s\right) &= \exp(aB_s) E\left(\exp\left(a(B_t - B_s) - \frac{a^2 t}{2}\right)/\mathcal{F}_s\right) \\ &= \exp(aB_s) E\left(\exp\left(a(B_t - B_s) - \frac{a^2 t}{2}\right)\right). \end{aligned}$$

Using the expression of the density of the random variable $B_t - B_s$, we write

$$\begin{aligned} E\left(\exp\left(a(B_t - B_s) - \frac{a^2 t}{2}\right)\right) &= \frac{1}{\sqrt{2\pi(t-s)}} \int_{\mathbb{R}} \exp\left(ax - \frac{at^2}{2} - \frac{x^2}{2(t-s)}\right) dx \\ &= \exp\left(\frac{a^2(t-s)}{2} - \frac{a^2 t}{2}\right) \\ &= \exp\left(-\frac{a^2 s}{2}\right), \end{aligned}$$

where the before last equality is obtained by using the identity

$$\frac{x^2}{2(t-s)} - ax + \frac{at^2}{2} = \frac{(x-a(t-s))^2}{2(t-s)} + \frac{a^2t}{2} - \frac{a^2(t-s)}{2}. \tag{2.7}$$

Therefore, we obtain

$$E\left(\exp\left(aB_t - \frac{a^2t}{2}\right) / \mathcal{F}_s\right) = \exp\left(aB_s - \frac{a^2s}{2}\right).$$

2.5 Markov property

For any $0 \le s \le t$, $x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$, we set

$$p(s,t,x,A) = \frac{1}{(2\pi(t-s))^{\frac{1}{2}}} \int_A \exp\left(-\frac{|x-y|^2}{2(t-s)}\right) dy. \tag{2.8}$$

Actually, $p(s,t,x,A)$ is the probability that a random variable, Normal, with mean x and variance $t-s$ take values on a fixed set A.
Let us prove the following identity:

$$P\{B_t \in A \,/\, \mathcal{F}_s\} = p(s,t,B_s,A), \tag{2.9}$$

which means that, conditionally to the past of the Brownian motion until time s, the law of B_t at a future time t only depends on B_s.
Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded measurable function. Then, since B_s is $\mathcal{F}_s$–measurable and $B_t - B_s$ independent of $\mathcal{F}_s$, we obtain

$$\begin{aligned} E\left(f(B_t)/\,\mathcal{F}_s\right) &= E\left(f\left(B_s + (B_t - B_s)\right) / \,\mathcal{F}_s\right) \\ &= E\left(f(x + B_t - B_s)\right)\Big|_{x=B_s}. \end{aligned}$$

The random variable $x + B_t - B_s$ is $\mathrm{N}(x, t-s)$. Thus,

$$E\left(f(x + B_t - B_s)\right) = \int_{\mathbb{R}} f(y) p(s,t,x,dy),$$

and consequently,

$$E\left(f(B_t)/\,\mathcal{F}_s\right) = \int_{\mathbb{R}} f(y) p(s,t,B_s,dy).$$

This yields (2.9) by taking $f = 1_A$.
With similar arguments, we can prove that

$$P\{B_t \in A \,/\, \sigma(B_s)\} = p(s,t,B_s,A),$$

which along with (2.9) yields

$$P\{B_t \in A \,/\, \mathcal{F}_s\} = P\{B_t \in A \,/\, \sigma(B_s)\} = p(s,t,B_s,A).$$

Going back to (2.8), we notice that the function $x \to p(s,t,x,A)$ is measurable, and the mapping $A \to p(s,t,x,A)$ is a probability.
Let us prove the additional property, called Chapman-Kolmogorov equation: For any $0 \le s \le u \le t$,

$$p(s,t,x,A) = \int_{\mathbb{R}} p(u,t,y,A)p(s,u,x,dy). \tag{2.10}$$

We recall that the sum of two independent Normal random variables, is again Normal, with mean the sum of the respective means, and variance the sum of the respective variances. This is expressed in mathematical terms by the fact that

$$\begin{aligned}[f_{\mathrm{N}(x,\sigma_1)} * f_{\mathrm{N}(y,\sigma_2)}](z) &= \int_{\mathbb{R}} f_{\mathrm{N}(x,\sigma_1)}(y) f_{\mathrm{N}(y,\sigma_2)}(z-y)dy \\ &= f_{\mathrm{N}(x+y,\sigma_1+\sigma_2)}(z).\end{aligned}$$

Using this fact, we obtain

$$\begin{aligned}\int_{\mathbb{R}} p(u,t,y,A)p(s,u,x,dy) &= \int_A dz\, \left(f_{\mathrm{N}(x,u-s)} * f_{\mathrm{N}(0,t-u)}\right)(z) \\ &= \int_A dz f_{\mathrm{N}(x,t-s)}(z) = p(s,t,x,A).\end{aligned}$$

proving (2.10).
This equation is the time continuous analogue of the property own by the transition probability matrices of a homogeneousMarkov chain. That is,

$$\Pi^{(m+n)} = \Pi^{(m)}\Pi^{(n)},$$

meaning that evolutions in $m+n$ steps are done by concatenating m-step and n-step evolutions. In (2.10) $m+n$ is replaced by the real time $t-s$, m by $t-u$, and n by $u-s$, respectively.
We are now ready to give the definition of a Markov process.
Consider a mapping

$$p : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to \mathbb{R}_+,$$

satisfying the properties

(i) for any fixed $s, t \in \mathbb{R}_+$, $A \in \mathcal{B}(\mathbb{R})$,

$$x \to p(s,t,x,A)$$

is $\mathcal{B}(\mathbb{R})$–measurable,

(ii) for any fixed $s, t \in \mathbb{R}_+$, $x \in \mathbb{R}$,

$$A \to p(s,t,x,A)$$

is a probability,

(iii) Equation (2.10) holds.

Such a function p is termed a Markovian transition function. Let us also fix a probability μ on $\mathcal{B}(\mathbb{R})$.

Definition 2.3 *A real valued stochastic process $\{X_t, t \in \mathbb{R}_+\}$ is a Markov process with initial law μ and transition probability function p if*

(a) the law of X_0 is μ,

(b) for any $0 \le s \le t$,

$$P\{X_t \in A/\mathcal{F}_s\} = p(s,t,X_s,A).$$

Therefore, we have proved that the Brownian motion is a Markov process with initial law a Dirac delta function at 0 and transition probability function p the one defined in (2.8).

Strong Markov property

Throughout this section, $(\mathcal{F}_t, t \ge 0)$ will denote the natural filtration associated with a Brownian motion $\{B_t, t \ge 0\}$ and stopping times will always refer to this filtration.

Theorem 2.4 *Let T be a stopping time. Then, conditionally to $\{T < \infty\}$, the process defined by*

$$B_t^T = B_{T+t} - B_T,\ t \ge 0,$$

is a Brownian motion independent of $\mathcal{F}_T$.

Proof: Assume that $T < \infty$ a.s. We shall prove that for any $A \in \mathcal{F}_T$, any choice of parameters $0 \le t_1 < \cdots < t_p$ and any continuous and bounded function f on $\mathbb{R}^p$, we have

$$E\left[1_A f\left(B^T_{t_1}, \cdots, B^T_{t_p}\right)\right] = P(A) E\left[f\left(B_{t_1}, \cdots, B_{t_p}\right)\right]. \tag{2.11}$$

This suffices to prove all the assertions of the theorem. Indeed, by taking $A = \Omega$, we see that the finite dimensional distributions of B and B^T coincide. On the other hand, (2.11) states the independence of 1_A and the random vector $(B^T_{t_1}, \cdots, B^T_{t_p})$. By a monotone class argument, we get the independence of 1_A and B^T.
The continuity of the sample paths of B implies, a.s.

$$\begin{aligned} & f\left(B^T_{t_1}, \cdots, B^T_{t_p}\right) = f\left(B_{T+t_1} - B_T, \ldots, B_{T+t_p} - B_T\right) \\ & = \lim_{n\to\infty} \sum_{k=1}^{\infty} 1_{\{(k-1)2^{-n} < T \le k2^{-n}\}} f\left(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \ldots B_{k2^{-n}+t_p} - B_{k2^{-n}}\right). \end{aligned}$$

Since f is bounded we can apply bounded convergence and write

$$\begin{aligned} & E\left[1_A f\left(B^T_{t_1}, \cdots, B_{t^T_p}\right)\right] \\ & \quad = \lim_{n\to\infty} \sum_{k=1}^{\infty} E\left[1_{\{(k-1)2^{-n} < T \le k2^{-n}\}} 1_A f\left(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \ldots B_{k2^{-n}+t_p} - B_{k2^{-n}}\right)\right]. \end{aligned}$$

Since $A \in \mathcal{F}_T$, the event $A \cap \{(k-1)2^{-n} < T \le k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$. Since Brownian motion has independent and stationary increments, we have

$$\begin{aligned} & E\left[1_{\{(k-1)2^{-n} < T \le k2^{-n}\}} 1_A f\left(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \ldots B_{k2^{-n}+t_p} - B_{k2^{-n}}\right)\right] \\ & = P\left[A \cap \{(k-1)2^{-n} < T \le k2^{-n}\}\right] E\left[f(B_{t_1}, \cdots B_{t_p})\right]. \end{aligned}$$

Summing up with respect to k both terms in the preceding identity yields (2.11), and this finishes the proof if $T < \infty$ a.s.
If $P(T = \infty) > 0$, we can argue as before and obtain

$$E\left[1_{A\cap\{T<\infty\}} f\left(B^T_{t_1}, \cdots, B^T_{t_p}\right)\right] = P(A \cap \{T < \infty\}) E\left[f\left(B_{t_1}, \cdots, B_{t_p}\right)\right].$$

$\square$

An interesting consequence of the preceding property is given in the next proposition.

Proposition 2.4 *For any $t > 0$, set $S_t = \sup_{s \le t} B_s$. Then, for any $a \ge 0$ and $b \le a$,*

$$P\left\{S_t \ge a, B_t \le b\right\} = P\left\{B_t \ge 2a - b\right\}. \tag{2.12}$$

As a consequence, the probability law of S_t and $|B_t|$ are the same.

Proof: Consider the stopping time

$$T_a = \inf\{t \ge 0, B_t = a\},$$

which is finite a.s. We have

$$P\{S_t \ge a, B_t \le b\} = P\{T_a \le t, B_t \le b\} = P\{T_a \le t, B^{T_a}_{t-T_a} \le b - a\}.$$

Indeed, $B^{T_a}_{t-T_a} = B_t - B_{T_a} = B_t - a$ and B and B^{T_a} have the same law. Moreover, we know that these processes are independent of $\mathcal{F}_{T_a}$.
This last property, along with the fact that B^{T_a} and $-B^{T_a}$ have the same law yields that (T_a, B^{T_a}) has the same distribution as $(T_a, -B^{T_a})$.
Define $H = \{(s, w) \in \mathbb{R}_+ \times \mathcal{C}(\mathbb{R}_+; \mathbb{R}); s \le t, w(t - s) \le b - a\}$. Then

$$\begin{aligned} P\{T_a \le t, B^{T_a}_{t-T_a} \le b - a\} &= P\{(T_a, B^{T_a}) \in H\} = P\{(T_a, -B^{T_a}) \in H\} \\ &= P\{T_a \le t, -B^{T_a}_{t-T_a} \le b - a\} = P\{T_a \le t, 2a - b \le B_t\} \\ &= P\{2a - b \le B_t\}. \end{aligned}$$

Indeed, by definition of the process $\{B^T_t, t \ge 0\}$, the condition $-B^{T_a}_{t-T_a} \le b - a$ is equivalent to $2a - b \le B_t$; moreover, the inclusion $\{2a - b \le B_t\} \subset \{T_a \le t\}$ holds true. In fact, if $T_a > t$, then $B_t \le a$; since $b \le a$, this yields $B_t \le 2a - b$. This ends the proof of (2.12).
For the second one, we notice that $\{B_t \ge a\} \subset \{S_t \ge a\}$. This fact along with (2.12) yield the validity of the identities

$$\begin{aligned} P\{S_t \ge a\} &= P\{S_t \ge a, B_t \le a\} + P\{S_t \ge a, B_t \ge a\} \\ &= 2P\{B_t \ge a\} = P\{|B_t| \ge a\}. \end{aligned}$$

The proof is now complete.

$\square$

3 Itô's calculus

Itô's calculus has been developed in the 50' by Kyoshi Itô in an attempt to give rigourous meaning to some differential equations driven by the Brownian motion appearing in the study of some problems related with continuous time Markov processes. Roughly speaking, one could say that Itô's calculus is an analogue of the classical Newton and Leibniz calculus for stochastic processes. In fact, in classical mathematical analysis, there are several extensions of the Riemann integral $\int f(x)dx$. For example, if g is an increasing bounded function (or the difference of two of these functions), Lebesgue-Stieltjes integral gives a precise meaning to the integral $\int f(x)g(dx)$, for some set of functions f. However, before Itô's development, no theory allowing nowhere differentiable integrators g was known. Brownian motion, introduced in the preceding chapter, is an example of stochastic process whose sample paths, although continuous, are nowhere differentiable. Therefore, Lebesgue-Stieltjes integral does not apply to the sample paths of Brownian motion.
There are many motivations coming from a variety of disciplines to consider stochastic differential equations driven by a Brownian motion. Such an object is defined as

$$\begin{aligned} dX_t &= \sigma(t, X_t)dB_t + b(t, X_t)dt, \\ X_0 &= x_0, \end{aligned}$$

or in integral form,

$$X_t = x_0 + \int_0^t \sigma(s, X_s)dB_s + \int_0^t b(s, X_s)ds. \tag{3.1}$$

The first notion to be introduced is that of *stochastic integral.* In fact, in (3.1) the integral $\int_0^t b(s, X_s)ds$ might be defined pathwise, but this is not the case for $\int_0^t \sigma(s, X_s)dB_s$, because of the roughness of the paths of the integrator. More explicitly, it is not possible to fix $\omega \in \Omega$, then to consider the path $\sigma(s, X_s(\omega))$, and finally to integrate with respect to $B_s(\omega)$.

3.1 Itô's integral

Throughout this section, we will consider a Brownian motion $B = \{B_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. We will also consider a filtration $(\mathcal{F}_t, t \geq 0)$ satisfying the following properties:

1. B is adapted to $(\mathcal{F}_t, t \geq 0)$,
2. the σ-field generated by $\{B_u - B_t, u \geq t\}$ is independent of $(\mathcal{F}_t, t \geq 0)$.

Notice that these two properties are satisfied if $(\mathcal{F}_t, t \geq 0)$ is the natural filtration associated to B.
We fix a finite time horizon T and define $L^2_{a,T}$ as the set of stochastic processes $u = \{u_t, t \in [0,T]\}$ satisfying the following conditions:

(i) u is adapted and jointly measurable in (t, ω), with respect to the product σ-field $\mathcal{B}([0,T]) \otimes \mathcal{F}$.

(ii) $\int_0^T E(u_t^2)dt < \infty$.

This is a Hilbert space with the norm $\|u\|_{L^2_{a,T}} = \left[\int_0^T E(u_t^2)dt\right]^{\frac{1}{2}}$, which coincides with the natural norm on the Hilbert space $L^2(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}([0,t]), dP \times d\lambda)$ (here λ stands for the Lebesgue measure on $\mathbb{R}$).
The notation $L^2_{a,T}$ evokes the two properties -adaptedness and square integrability- described before.
Consider first the subset of $L^2_{a,T}$ consisting of *step processes*. That is, stochastic processes which can be written as

$$u_t = \sum_{j=1}^n u_j 1_{[t_{j-1}, t_j[}(t), \tag{3.2}$$

with $0 = t_0 \leq t_1 \leq \cdots \leq t_n = T$ and where u_j, $j = 1, \ldots, n$, are $\mathcal{F}_{t_{j-1}}$–measurable square integrable random variables. We shall denote by $\mathcal{E}$ the set of these processes.
For step processes, the Itô stochastic integral is defined by the very natural formula

$$\int_0^T u_t dB_t = \sum_{j=1}^n u_j (B_{t_j} - B_{t_{j-1}}), \tag{3.3}$$

that we may compare with Lebesgue integral of simple functions. Notice that $\int_0^T u_t dB_t$ is a random variable. Of course, we would like to be able to consider more general integrands than step processes. Therefore, we must try to extend the definition (3.3). For this, we have to use tools from *Functional Analysis* based upon a very natural idea: If we are able to prove that (3.3) gives a continuous functional between two metric spaces, then the stochastic integral defined for the very particular class of step stochastic processes could be extended to a more general class given by the closure of this set with respect to a suitable norm. This is possible by one of the consequences of Hahn-Banach Theorem.
The idea of continuity is made precise by the

Isometry property:

$$E\left(\int_0^T u_t dB_t\right)^2 = E\left(\int_0^T u_t^2 dt\right). \tag{3.4}$$

Let us prove (3.4) for *step processes*. Clearly

$$\begin{aligned} E\left(\int_0^T u_t dB_t\right)^2 &= \sum_{j=1}^n E\left(u_j^2 (\Delta_j B)^2\right) \\ &\quad + 2\sum_{j<k} E(u_j u_k (\Delta_j B)(\Delta_k B)). \end{aligned}$$

The measurability property of the random variables u_j, $j = 1, \ldots, n$, implies that the random variables u_j^2 are independent of $(\Delta_j B)^2$. Hence, the contribution of the first term in the right hand-side of the preceding identity is equal to

$$\sum_{j=1}^n E(u_j^2)(t_j - t_{j-1}) = \int_0^T E(u_t^2) dt.$$

For the second term, we notice that for fixed j and k, $j < k$, the random variables $u_j u_k \Delta_j B$ are independent of $\Delta_k B$. Therefore,

$$E(u_j u_k (\Delta_j B)(\Delta_k B)) = E(u_j u_k (\Delta_j B)) E(\Delta_k B) = 0.$$

Thus, we have (3.4).
This property tell us that the stochastic integral is a continuous functional defined on $\mathcal{E}$, endowed with the norm of $L^2(\Omega \times [0,T])$, taking values on the set $L^2(\Omega)$ of square integrable random variables.

Other properties of the stochastic integral of step processes

1. The stochastic integral is a centered random variable. Indeed,

$$\begin{aligned} E\left(\int_0^T u_t dB_t\right) &= E\left(\sum_{j=1}^n u_j (B_{t_j} - B_{t_{j+1}}\right) \\ &= \sum_{j=1}^n E(u_j) E\left(B_{t_j} - B_{t_{j+1}}\right) = 0, \end{aligned}$$

   where we have used that the random variables u_j and $B_{t_j} - B_{t_{j+1}}$ are independent and moreover $E\left(B_{t_j} - B_{t_{j+1}}\right) = 0$.

2. *Linearity*: If u^1, u^2 are two step processes and $a, b \in \mathbb{R}$, then clearly $au^1 + bu^2$ is also a step process and

$$\int_0^T (au^1 + bu^2)(t)dB_t = a \int_0^T u^1(t)dB_t + b \int_0^T u^2(t)dB_t.$$

The next step consists of identifying a bigger set than $\mathcal{E}$ of random processes such that $\mathcal{E}$ is dense in the norm of the Hilbert space $L^2(\Omega \times [0,T])$. Since $L^2_{a,T}$ is a Hilbert space with respect to this norm, we have that $\bar{\mathcal{E}} \subset L^2_{a,T}$. The converse inclusion is also true. This is proved in the next Proposition, which is a crucial fact in Itô's theory.

Proposition 3.1 *For any $u \in L^2_{a,T}$ there exists a sequence $(u^n, n \ge 1) \subset \mathcal{E}$ such that*

$$\lim_{n\to\infty} \int_0^T E(u^n_t - u_t)^2 dt = 0.$$

Proof: Assume $u \in L^2_{a,T}$, bounded, and has continuous sample paths, a.s. An approximation sequence can be defined as follows:

$$u^n(t) = \sum_{k=0}^{[nT]} u\left(\frac{k}{n}\right) 1_{\left[\frac{k}{n}, \frac{k+1}{n}\right[}(t),$$

with the convention $\frac{[nT]+1}{n} := T$.
Clearly, $u^n \in L^2_{a,T}$ and by continuity,

$$\begin{aligned}\int_0^T |u^n(t) - u(t)|^2 dt &= \sum_{k=0}^{[nT]} \int_{\frac{k}{n}}^{\frac{k+1}{n}\wedge T} \left| u\left(\frac{k}{n}\right) - u(t)\right|^2 dt \\ &\le T \sup_k \sup_{t\in\Delta_k} |u^n(t) - u(t)|^2 \to 0,\end{aligned}$$

as $n \to \infty$, a.s. Then, the approximation result follows by bounded convergence.
In a second step, we assume that $u \in L^2_{a,T}$ is bounded. For any $n \ge 1$, let $\Psi_n(s) = n 1_{[0,\frac{1}{n}]}(s)$. The sequence $(\Psi_n)_{n\ge1}$ is an approximation of the identity. Consider

$$u^n(t) = \int_{-\infty}^{+\infty} \Psi_n(t-s)u(s)ds = (\Psi_n * u)(t) = \int_{t-\frac{1}{n}}^t u(s)ds,$$

where the symbol "$*$" denotes the convolution operator on $\mathbb{R}$. This defines a stochastic process u^n with continuous and bounded sample paths, a.s., and

$$\sup_{s\in[0,T]} |u^n(s) - u(s)|^2 \to 0, a.s.$$

By bounded convergence,

$$\lim_{n\to\infty} E\int_0^T |u^n(s)-u(s)|^2 ds \to 0.$$

Finally, consider $u \in L^2_{a,T}$ and define

$$u^n(t) = \begin{cases} 0, & u(t) < -n, \\ u(t), & -n \le u(t) \le n, \\ 0, & u(t) > n. \end{cases}$$

Clearly, $\sup_{\omega,t} |u^n(t)| \le n$ and $u^n \in L^2_{a,T}$. Moreover,

$$E\int_0^T |u^n(s)-u(s)|^2 ds = E\int_0^T |u(s)|^2 1_{\{|u(s)|>n\}} ds \to 0,$$

where we have used that for a function $f \in L^1(\Omega, \mathcal{F}, \mu)$,

$$\lim_{n\to\infty} \int_\Omega |f| 1_{|f|>n} d\mu = 0.$$

$\square$

By using the approximation result provided by the preceding Proposition, we can give the following definition.

Definition 3.1 *The Itô stochastic integral of a process $u \in L^2_{a,T}$ is*

$$\int_0^T u_t dB_t := L^2(\Omega) - \lim_{n\to\infty} \int_0^T u_t^n dB_t. \tag{3.5}$$

In order this definition to make sense, one needs to make sure that if the process u is approximated by two different sequences, say $u^{n,1}$ and $u^{n,2}$, the definition of the stochastic integral, using either $u^{n,1}$ or $u^{n,2}$ coincide. This is proved using the isometry property. Indeed

$$\begin{aligned} & E\left(\int_0^T u_t^{n,1} dB_t - \int_0^T u_t^{n,2} dB_t\right)^2 = \int_0^T E\left(u_t^{n,1} - u_t^{n,2}\right)^2 dt \\ & \le 2\int_0^T E\left(u_t^{n,1} - u_t\right)^2 dt + 2\int_0^T E\left(u_t^{n,2} - u_t\right)^2 dt \\ & \to 0. \end{aligned}$$

Consequently, denoting by $I^i(u) = L^2(\Omega) - \lim_{n\to\infty} \int_0^T u_t^{n,i} dB_t$, $i = 1, 2$ and using the triangular inequaltity, we have

$$\begin{aligned}\|I^1(u) - I^2(u)\|_2 \le \|I^1(u) - \int_0^T u_t^{n,1} dB_t\|_2 + \|\int_0^T u_t^{n,1} dB_t - \int_0^T u_t^{n,2} dB_t\|_2 \\ + \|I^2(u) - \int_0^T u_t^{n,2} dB_t\|_2,\end{aligned}$$

where $\|\cdot\|_2$ denotes the norm in $L^2(\Omega)$.
The right-hand side of this inequality tends to zero as $n \to \infty$. Hence the left-hand side is null.
By its very definition, the stochastic integral defined in Definition 3.1 satisfies the isometry property as well. Indeed, using again the notation $I(\cdot)$ for the stochastic integral, we have

$$\begin{aligned}\|I(u)\|_{L^2(\Omega)} &= \lim_{n\to\infty} \|I(u^n)\|_{L^2(\Omega)} \\ &= \lim_{n\to\infty} \|u^n\|_{L^2(\Omega\times[0,T])} = \|u\|_{L^2(\Omega\times[0,T])}.\end{aligned}$$

Moreover,

(a) stochastic integrals are centered random variables:

$$E\left(\int_0^T u_t dB_t\right) = 0,$$

(b) stochastic integration is a linear operator:

$$\int_0^T (au_t + bv_t)\, dB_t = a \int_0^T u_t dB_t + b \int_0^T v_t dB_t.$$

Remember that these facts are true for processes in $\mathcal{E}$, as has been mentioned before. The extension to processes in $L^2_{a,T}$ is done by applying Proposition 3.1. For the sake of illustration we prove (a).
Consider an approximating sequence u^n in the sense of Proposition 3.1. By the construction of the stochastic integral $\int_0^T u_t dB_t$, it holds that

$$\lim_{n\to\infty} E\left(\int_0^T u_t^n dB_t\right) = E\left(\int_0^T u_t dB_t\right),$$

Since $E\left(\int_0^T u_t^n dB_t\right) = 0$ for every $n \ge 1$, this concludes the proof.

We end this section with an interesting example.

Example 3.1 *For the Brownian motion B, the following formula holds:*

$$\int_0^T B_t dB_t = \frac{1}{2}\left(B_T^2 - T\right).$$

Let us remark that we would rather expect $\int_0^T B_t dB_t = \frac{1}{2}B_T^2$, by analogy with rules of deterministic calculus.
To prove this identity, we consider a particular sequence of approximating step processes based on the partition $\left\{\frac{jT}{n}, j = 0, \ldots, n\right\}$, as follows:

$$u_t^n = \sum_{j=1}^{n} B_{t_{j-1}} \; 1_{]t_{j-1},t_j]}(t),$$

with $t_j = \frac{jT}{n}$. Clearly, $u^n \in L^2_{a,T}$ and we have

$$\int_0^T E\left(u_t^n - B_t\right)^2 dt = \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} E\left(B_{t_{j-1}} - B_t\right)^2 dt$$
$$\leq \frac{T}{n}\sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} dt = \frac{T^2}{n}.$$

Therefore, $(u^n, n \geq 1)$ is an approximating sequence of B in the norm of $L^2(\Omega \times [0,T])$.
According to Definition 3.1,

$$\int_0^T B_t dB_t = \lim_{n\to\infty} \sum_{j=1}^{n} B_{t_{j-1}}\left(B_{t_j} - B_{t_{j-1}}\right),$$

in the $L^2(\Omega)$ norm.
Clearly,

$$\begin{aligned}\sum_{j=1}^{n} B_{t_{j-1}}\left(B_{t_j} - B_{t_{j-1}}\right) &= \frac{1}{2}\sum_{j=1}^{n}\left(B_{t_j}^2 - B_{t_{j-1}}^2\right) \\ &\quad - \frac{1}{2}\sum_{j=1}^{n}\left(B_{t_j} - B_{t_{j-1}}\right)^2 \\ &= \frac{1}{2}B_T^2 - \frac{1}{2}\sum_{j=1}^{n}\left(B_{t_j} - B_{t_{j-1}}\right)^2. \end{aligned} \qquad (3.6)$$

We conclude by using Proposition 2.3.

3.2 The Itô integral as a stochastic process

The indefinite Itô stochastic integral of a process $u \in L^2_{a,T}$ is defined as follows:

$$\int_0^t u_s dB_s := \int_0^T u_s 1_{[0,t]}(s) dB_s, \tag{3.7}$$

$t \in [0,T]$.
For this definition to make sense, we need that for any $t \in [0,T]$, the process $\{u_s 1_{[0,t]}(s), s \in [0,T]\}$ belongs to $L^2_{a,T}$. This is clearly true.
Obviously, properties of the integral mentioned in the previous section, like zero mean, isometry, linearity, also hold for the indefinite integral.
The rest of the section is devoted to the study of important properties of the stochastic process given by an indefinite Itô integral.

Proposition 3.2 *The process* $\{I_t = \int_0^t u_s dB_s, t \in [0,T]\}$ *is a martingale.*

Proof: We first establish the martingale property for any approximating sequence

$$I_t^n = \int_0^t u_s^n dB_s, t \in [0,T],$$

where u^n converges to u in $L^2(\Omega \times [0,T])$. This suffices to prove the Proposition, since $L^2(\Omega)$–limits of martingales are again martingales (this fact follows from Jensen's inequality).
Let u_t^n, $t \in [0,T]$, be defined by the right hand-side of (3.2). Fix $0 \le s \le t \le T$ and assume that $t_{k-1} < s \le t_k < t_l < t \le t_{l+1}$. Then

$$\begin{aligned} I_t^n - I_s^n &= u_k(B_{t_k} - B_s) + \sum_{j=k+1}^{l} u_j(B_{t_j} - B_{t_j-1}) \\ &\quad + u_{l+1}(B_t - B_{t_l}). \end{aligned}$$

Using properties (g) and (f), respectively, of the conditional expectation (see Appendix 1) yields

$$\begin{aligned} E\left(I_t^n - I_s^n / \mathcal{F}_s\right) &= E\left(u_k(B_{t_k} - B_s)/\mathcal{F}_s\right) + \sum_{j=k+1}^{l} E\left(E\left(u_j \Delta_j B/\mathcal{F}_{t_{j-1}}\right)/\mathcal{F}_s\right) \\ &\quad + E\left(u_{l+1} E\left(B_t - B_{t_l}/\mathcal{F}_{t_l}\right)/\mathcal{F}_s\right) \\ &= 0. \end{aligned}$$

This finishes the proof of the proposition. □

A proof not very different as that of Proposition 2.3 yields

Proposition 3.3 *For any process $u \in L^2_{a,T}$ and bounded,*

$$L^1(\Omega) - \lim_{n\to\infty} \sum_{j=1}^n \left(\int_{t_{j-1}}^{t_j} u_s dB_s\right)^2 = \int_0^t u_s^2 ds.$$

That means, the "quadratic variation" of the indefinite stochastic integral is given by the process $\{\int_0^t u_s^2 ds, t \in [0,T]\}$.

The isometry property of the stochastic integral can be extended in the following sense. Let $p \in [2,\infty[$. Then,

$$E\left(\int_0^t u_s dB_s\right)^p \le C(p) E\left(\int_0^t u_s^2 ds\right)^{\frac{p}{2}}. \tag{3.8}$$

Here $C(p)$ is a positive constant depending on p. This is Burkholder's inequality.
A combination of Burkholder's inequality and Kolmogorov's continuity criterion allows to deduce the continuity of the sample paths of the indefinite stochastic integral. Indeed, assume that $\int_0^T E\left(u_r\right)^p dr < \infty$, for any $p \in [2,\infty[$. Using first (3.8) and then Hölder's inequality (be smart!) implies

$$\begin{aligned} E\left(\int_s^t u_r dB_r\right)^p &\le C(p) E\left(\int_s^t u_r^2 dr\right)^{\frac{p}{2}} \\ &\le C(p)|t-s|^{\frac{p}{2}-1} \int_s^t E\left(u_r\right)^p dr \\ &\le C(p)|t-s|^{\frac{p}{2}-1}. \end{aligned}$$

Since $p \ge 2$ is arbitrary, with Theorem 1.1 we have that the sample paths of $\int_0^t u_s dB_s, t \in [0,T]$ are γ–Hölder continuous with $\gamma \in]0, \frac{1}{2}[$.

3.3 An extension of the Itô integral

In Section 3.1 we have introduced the set $L^2_{a,T}$ and we have defined the stochastic integral of processes of this class with respect to the Brownian motion. In this section we shall consider a large class of integrands. The notations and underlying filtration are the same as in Section 3.1.
Let $\Lambda^2_{a,T}$ be the set of real valued processes u adapted to the filtration $(\mathcal{F}_t, t \ge 0)$, jointly measurable in (t,ω) with respect to the product σ-field $\mathcal{B}([0,T]) \times \mathcal{F}$ and satisfying

$$P\left\{\int_0^T u_t^2 dt < \infty\right\} = 1. \tag{3.9}$$

Clearly $L^2_{a,T} \subset \Lambda^2_{a,T}$. Our aim is to define the stochastic integral for processes in $\Lambda^2_{a,T}$. For this we shall follow the same approach as in section 3.1. Firstly, we start with step processes $(u^n, n \geq 1)$ of the form (3.2) belonging to $\Lambda^2_{a,T}$ and define the integral as in (3.3). The extension to processes in $\Lambda^2_{a,T}$ needs two ingredients. The first one is an approximation result that we now state without giving a proof. Reader may consult for instance [1].

Proposition 3.4 *Let $u \in \Lambda^2_{a,T}$. There exists a sequence of step processes $(u^n, n \geq 1)$ of the form (3.2) belonging to $\Lambda^2_{a,T}$ such that*

$$\lim_{n\to\infty} \int_0^T |u_t^n - u_t|^2 dt = 0,$$

a.s.

The second ingredient gives a connection between stochastic integrals of step processes in $\Lambda^2_{a,T}$ and their quadratic variation, as follows.

Proposition 3.5 *Let u be a step processes in $\Lambda^2_{a,T}$. Then for any $\epsilon > 0$, $N > 0$,*

$$P\left\{\left|\int_0^T u_t dB_t\right| > \epsilon\right\} \leq P\left\{\int_0^T u_t^2 dt > N\right\} + \frac{N}{\epsilon^2}. \tag{3.10}$$

Proof: It is based on a truncation argument. Let u be given by the right-hand side of (3.2) (here it is not necessary to assume that the random variables u_j are in $L^2(\Omega)$). Fix $N > 0$ and define

$$v_t^N = \begin{cases} u_j, & \text{if } t \in [t_{j-1}, t_j[, \text{ and } \sum_{j=1}^n u_j^2(t_j - t_{j-1}) \leq N, \\ 0, & \text{if } t \in [t_{j-1}, t_j[, \text{ and } \sum_{j=1}^n u_j^2(t_j - t_{j-1}) > N, \end{cases}$$

The process $\{v_t^N, t \in [0,T]\}$ belongs to $L^2_{a,T}$. Indeed, by definition

$$\int_0^t |v_t^N|^2 dt \leq N.$$

Moreover, if $\int_0^T u_t^2 dt \leq N$, necessarily $u_t = v_t^N$ for any $t \in [0,T]$. Then by considering the decomposition

$$\begin{aligned} &\left\{\left|\int_0^T u_t dB_t\right| > \epsilon\right\} \\ &\quad = \left\{\left|\int_0^T u_t dB_t\right| > \epsilon, \int_0^T u_t^2 dt > N\right\} \cup \left\{\left|\int_0^T u_t dB_t\right| > \epsilon, \int_0^T u_t^2 dt \leq N\right\}, \end{aligned}$$

we obtain

$$P\left\{\left|\int_0^T u_t dB_t\right| > \epsilon\right\} \le P\left\{\left|\int_0^T v_t^N dB_t\right| > \epsilon\right\} + P\left\{\int_0^T u_t^2 dt > N\right\}.$$

We finally apply Chebychev's inequality along with the isometry property of the stochastic integral for processes in $L^2_{a,T}$ and get

$$P\left\{\left|\int_0^T v_t^N dB_t\right| > \epsilon\right\} \le \frac{1}{\epsilon^2} E\left(\int_0^T v_t^N dB_t\right)^2 \le \frac{N}{\epsilon^2}.$$

This ends the proof of the result.

□

The extension

Fix $u \in \Lambda^2_{a,T}$ and consider a sequence of step processes $(u^n, n \ge 1)$ of the form (3.2) belonging to $\Lambda^2_{a,T}$ such that

$$\lim_{n\to\infty} \int_0^T |u_t^n - u_t|^2 dt = 0, \tag{3.11}$$

in the convergence of probability.
By Proposition 3.5, for any $\epsilon > 0$, $N > 0$ we have

$$P\left\{\left|\int_0^T (u_t^n - u_t^m) dB_t\right| > \epsilon\right\} \le P\left\{\left|\int_0^T (u_t^n - u_t^m)^2 dt\right| > N\right\} + \frac{N}{\epsilon^2}.$$

Using (3.11), we can choose ϵ such that for any $N > 0$ and n, m big enough,

$$P\left\{\left|\int_0^T (u_t^n - u_t^m)^2 dt\right| > N\right\} \le \frac{\epsilon}{2}.$$

Then, we may take N small enough so that $\frac{N}{\epsilon^2} \le \frac{\epsilon}{2}$. Consequently, we have proved that the sequence of stochastic integrals of step processes

$$\left(\int_0^T u_t^n dB_t, n \ge 1\right) \tag{3.12}$$

is Cauchy in probability. The space $L^0(\Omega)$ of classes of finite random variables (a.s.) endowed with the convergence in probability is a complete metric space. For example, a possible distance is

$$d(X, Y) = E\left(\frac{|X - Y|}{1 + |X - Y|}\right).$$

Hence the sequence (3.12) does have a limit in probability. Then, we define

$$\int_0^T u_t dB_t = P - \lim_{n\to\infty} \int_0^T u_t^n dB_t. \tag{3.13}$$

It is easy to check that this definition is indeed independent of the particular approximation sequence used in the construction.

3.4 A change of variables formula: Itô's formula

Like in Example 3.1, we can prove the following formula, valid for any $t \geq 0$:

$$B_t^2 = 2\int_0^t B_s dB_s + t. \tag{3.14}$$

If the sample paths of $\{B_t, t \geq 0\}$ were sufficiently smooth -for example, of bounded variation- we would rather have

$$B_t^2 = 2\int_0^t B_s dB_s. \tag{3.15}$$

Why is it so? Consider a similar decomposition as the one given in (3.6) obtained by restricting the time interval to $[0,t]$. More concretely, consider the partition of $[0,t]$ defined by $0 = t_0 \leq t_1 \leq \cdots \leq t_n = t$,

$$\begin{aligned} B_t^2 &= \sum_{j=0}^{n-1} \left(B_{t_{j+1}}^2 - B_{t_j}^2\right) \\ &= 2\sum_{j=0}^{n-1} B_{t_j}\left(B_{t_{j+1}} - B_{t_j}\right) + \sum_{j=0}^{n-1}\left(B_{t_{j+1}} - B_{t_j}\right)^2, \end{aligned} \tag{3.16}$$

where we have used that $B_0 = 0$.

Consider a sequence of partitions of $[0,t]$ whose mesh tends to zero. We already know that

$$\sum_{j=0}^{n-1}\left(B_{t_{j+1}} - B_{t_j}\right)^2 \to t,$$

in the convergence of $L^2(\Omega)$. This gives the extra contribution in the development of B_t^2 in comparison with the classical calculus approach.

Notice that, if B were of bounded variation then, we could argue as follows:

$$\begin{aligned} \sum_{j=0}^{n-1}\left(B_{t_{j+1}} - B_{t_j}\right)^2 \leq & \sup_{0\leq j\leq n-1} |B_{t_{j+1}} - B_{t_j}| \\ & \times \sum_{j=0}^{n-1} |B_{t_{j+1}} - B_{t_j}|. \end{aligned}$$

By the continuity of the sample paths of the Brownian motion, the first factor in the right hand-side of the preceding inequality tends to zero as the mesh of the partition tends to zero, while the second factor remains finite, by the property of bounded variation.
Summarising. Differential calculus with respect to the Brownian motion should take into account second order differential terms. Roughly speaking

$$(dB_t)^2 = dt.$$

A precise meaning to this formal formula is given in Proposition 2.3.

3.4.1 One dimensional Itô's formula

In this section, we shall extend the formula (3.14) and write an expression for $f(t, X_t)$ for a class of functions f and a family of stochastic processes which include $f(x) = x^2$ and the Brownian motion, respectively. To illustrate the method of the proof, we start with the particular case given in the next assertion (see [6]).

Theorem 3.1 *Consider a function $f : \mathbb{R} \to \mathbb{R}$ belonging to $\mathcal{C}^2$, the class of continuous differentiable real functions up to order two. Then, for any $0 \le a \le t$,*

$$f(B_t) = f(B_a) + \int_a^t f'(B_s)dB_s + \frac{1}{2}\int_a^t f''(B_s)ds, \tag{3.17}$$

a.s.

The proof relies on two technical lemmas. In the sequel, we will consider a sequence of partitions $\Pi_n = \{0 = t_0 \le t_1 \le \cdots \le t_n = t\}$ such that $\lim_{n\to\infty} |\Pi_n| = 0$.

Lemma 3.1 *Let g be a real continuous function and $\lambda_i \in (0,1)$, $i = 1, \dots, n$. There exists a subsequence (denoted by (n), by simplicity) such that*

$$X_n := \sum_{i=1}^n \left[g(B_{t_{i-1}} + \lambda_i(B_{t_i} - B_{t_{i-1}})) - g(B_{t_{i-1}})\right](B_{t_i} - B_{t_{i-1}})^2,\ n \ge 1,$$

converges to zero a.s.

Proof. Consider the random variable

$$Y_n = \max_{1\le i\le n, 0<\lambda<1} \left|g(B_{t_{i-1}} + \lambda(B_{t_i} - B_{t_{i-1}})) - g(B_{t_{i-1}})\right|,$$

Clearly

$$|X_n| \leq Y_n \sum_{i=1}^{n} (B_{t_i} - B_{t_{i-1}})^2.$$

By the continuity of the sample paths of Brownian motion, $(Y_n)_{n\geq 1}$ converges to zero a.s. Moreover, $\left\{\sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2, n \geq 1\right\}$ converges in $L^2(\Omega)$ to t. Hence, the lemma holds.

□

Lemma 3.2 *The hypotheses are the same as in Lemma 3.1. The sequence*

$$S_n := \sum_{i=1}^{n} g(B_{t_{i-1}}) \left[(B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i_1})\right],\ n \geq 1,$$

converges in probability to zero.

Proof. We will introduce a localization in Ω. With this, g may be assumed to be bounded and we will prove convergence to zero in $L^2(\Omega)$. Finally, we will remove the localization and obtain the result. This is a quite usual procedure in probability theory.
Fix $L > 0$ and set

$$A_{i-1}^{(L)} = \{|B(t_l)| \leq L, 0 \leq l \leq i-1\},\ 1 \leq i \leq n.$$

Notice that this is a decreasing family in the parameter i.
Define

$$S_{n,L} = \sum_{i=1}^{n} g(B_{t_{i-1}}) 1_{A_{i-1}^{(L)}} \left[(B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i_1})\right].$$

This is a localization of S_n.
To simplify the notation, we call

$$\begin{aligned} X_i &= (B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i_1}), \\ Y_i &= g(B_{t_{i-1}}) 1_{A_{i-1}^{(L)}} \left[(B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i_1})\right], \end{aligned}$$

$1 \leq 1 \leq n$. Our aim is to prove first that $E\left(|\sum_i Y_i|^2\right) \to 0$.
Let $\mathcal{F}_t = \sigma(B(s), 0 \leq s \leq t)$. Fix $1 \leq i < j \leq n$. Then

$$\begin{aligned} E(Y_i Y_j) &= E\left(E\left(Y_i Y_j / \mathcal{F}_{t_{j-1}}\right)\right) \\ &= E\left(Y_i g(B_{t_{j-1}}) 1_{A_{j-1}^{(L)}} E(X_j / \mathcal{F}_{t_{j-1}}\right) \\ &= 0. \end{aligned}$$

We clearly have $Y_i^2 \le \sup_{|x|\le L} |g(x)|^2 X_i^2$ and therefore,

$$E(Y_i^2) \le C(t_i - t_{i-1})^2 \sup_{|x|\le L} |g(x)|^2.$$

Consequently,

$$\begin{aligned} E(S_{n,L}^2) &= \sum_{i=1}^n E(Y_i^2) \\ &\le Ct \sup_{|x|\le L} |g(x)|^2 |\Pi_n|, \end{aligned}$$

which converges to zero as $n \to \infty$.
Next, we fix $\epsilon > 0$ and write

$$\begin{aligned} P\{|S_n| > \epsilon\} &= P\{|S_n| > \epsilon,\ S_n = S_{n,L}\} + P\{|S_n| > \epsilon,\ S_n \ne S_{n,L}\} \\ &\le P\{|S_{n,L}| > \epsilon\} + P\{S_n \ne S_{n,L}\}. \end{aligned}$$

Chebyshev's inequality yields

$$\lim_{n\to\infty} P\{|S_{n,L}| > \epsilon\} \le \epsilon^{-2} \lim_{n\to\infty} E(|S_{n,L}|^2) = 0.$$

Moreover, if ω is such that $S_n(\omega) \ne S_{n,L}(\omega)$ there exists $i = 1, \dots, n$, such that $\omega \in (A_{i-1}^{(L)})^c$. The family $(A_{i-1}^{(L)})_i$ decreases in i, consequently, $\omega \in (A_{n-1}^{(L)})^c$. Thus,

$$P\{S_n \ne S_{n,L}\} \le P(A_{n-1}^{(L)})^c \le P\left\{\sup_{0\le s\le t} |B(s)| > L\right\}.$$

We proved in Proposition 2.4 that the law of $\sup_{0\le s\le t} |B(s)|$ and $|B_t|$ are the same. Thus,

$$P\left\{\sup_{0\le s\le t} |B(s)| > L\right\} = P\{|B(t)| > L\} \le L^{-1} E(|B(t)|) = L^{-1}\sqrt{\frac{2t}{\pi}}.$$

This yields

$$\lim_{L\to\infty} P\{S_n \ne S_{n,L}\} = 0,$$

and ends the proof of the Lemma.

$\square$

Proof of Theorem 3.1

For simplicity, we take $a = 0$. We fix ω and consider a Taylor expansion up to the second order. This yields

$$
\begin{aligned}
f(B_t) - f(0) = & \sum_{i=1}^{n} f'(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) \\
& + \frac{1}{2}\sum_{i=1}^{n} f''(B_{t_{i-1}} + \lambda_i(B_{t_i} - B_{t_{i-1}}))(B_{t_i} - B_{t_{i-1}})^2.
\end{aligned}
$$

The stochastic process $\{u_t = f'(B_t), t \geq 0\}$ has continuous sample paths, a.s. Let

$$
u^n(t) = \sum_{i=1}^{n} f'(B_{t_{i-1}}) 1_{[t_{i-1}, t_i]}.
$$

By continuity,

$$
\lim_{n\to\infty} \int_0^t |u^n(s) - u(s)|^2 ds = 0, \ a.s.
$$

Therefore,

$$
\lim_{n\to\infty} \sum_{i=1}^{n} f'(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) = \lim_{n\to\infty} \int_0^t u^n(s) dB_s = \int_0^t u(s) dB_s,
$$

in probability, and also a.s. for some subsequence.

Next we consider the terms with the second derivative. By the triangular inequality, we have

$$
\begin{aligned}
& \left| \sum_{i=1}^{n} f''(B_{t_{i-1}} + \lambda_i(B_{t_i} - B_{t_{i-1}}))(B_{t_i} - B_{t_{i-1}})^2 - \int_0^t f''(B_s) ds \right| \\
\leq & \left| \sum_{i=1}^{n} f''(B_{t_{i-1}} + \lambda_i(B_{t_i} - B_{t_{i-1}}))(B_{t_i} - B_{t_{i-1}})^2 - \sum_{i=1}^{n} f''(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 \right| \\
+ & \left| \sum_{i=1}^{n} f''(B_{t_{i-1}}) \left[(B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1}) \right] \right| \\
+ & \left| \sum_{i=1}^{n} f''(B_{t_{i-1}})(t_i - t_{i-1}) - \int_0^t f''(B_s) ds \right|.
\end{aligned}
$$

The first term on the right-hand side of the preceding inequality converges to zero as $n \to \infty$, a.s. Indeed this follows from Lemma 3.1 applied to the function $g := f''$. With the same choice of g, Lemma 3.2 yields the a.s.

convergence to zero of a subsequence for the second term. Finally, the third term converges also to zero, a.s. by the classical result on approximation of Riemann integrals by Riemann sums.

□

We now introduce the class of *Itô Processes* for which we will prove a more general version of the Itô formula.

Definition 3.2 *Let $\{v_t, t \in [0,T]\}$ be a stochastic process, adapted, whose sample paths are almost surely Lebesgue integrable, that is $\int_0^T |v_t| dt < \infty$, a.s. Consider a stochastic process $\{u_t, t \in [0,T]\}$ belonging to $\Lambda^2_{a,T}$ and a random variable X_0. The stochastic process defined by*

$$X_t = X_0 + \int_0^t u_s dB_s + \int_0^t v_s ds, \tag{3.18}$$

$t \in [0,T]$ is termed an Itô process.

An alternative writing of (3.18) in *differential form* is

$$dX_t = u_t dB_t + v_t dt.$$

Let $\mathcal{C}^{1,2}$ denote the set of functions on $[0,T] \times \mathbb{R}$ which are jointly continuous in (t,x), continuous differentiable in t and twice continuous differentiable in x, with jointly continuous derivatives. Our next aim is to prove an Itô formula for the stochastic process $\{f(t,X_t), t \in [0,T\}]$, $f \in \mathcal{C}^{1,2}$ This will be an extension of (3.17).

Theorem 3.2 *Let $f : [0,T] \times \mathbb{R} \to \mathbb{R}$ be a function in $\mathcal{C}^{1,2}$ and X be an Itô process with decomposition given in (3.18). The following formula holds:*

$$\begin{aligned} f(t,X_t) = f(0,X_0) &+ \int_0^t \partial_s f(s,X_s) ds + \int_0^t \partial_x f(s,X_s) u_s dB_s \\ &+ \int_0^t \partial_x f(s,X_s) v_s ds + \frac{1}{2} \int_0^t \partial^2_{xx} f(s,X_s) u_s^2 ds. \end{aligned} \tag{3.19}$$

Formula (3.19) can also be written as

$$\begin{aligned} f(t,X_t) = f(0,X_0) &+ \int_0^t \partial_s f(s,X_s) ds + \int_0^t \partial_x f(s,X_s) dX_s \\ &+ \frac{1}{2} \int_0^t \partial^2_{xx} f(s,X_s) (dX_s)^2, \end{aligned} \tag{3.20}$$

or in *differential form* ,

$$df(t,X_t) = \partial_t f(t,X_t)dt + \partial_x f(t,X_t)dX_t + \frac{1}{2}\partial^2_{xx} f(t,X_t)(dX_t)^2, \tag{3.21}$$

where $(dX_t)^2$ is computed using the formal rule

$$\begin{aligned} dB_t \times dB_t &= dt, \\ dB_t \times dt &= dt \times dB_t = 0, \\ dt \times dt &= 0. \end{aligned}$$

Example 3.2 *Consider the function*

$$f(t,x) = e^{\mu t - \frac{\sigma^2}{2}t + \sigma x},$$

with $\mu, \sigma \in \mathbb{R}$.
Applying formula (3.19) to $X_t := B_t$ *-a Brownian motion- yields*

$$f(t,B_t) = 1 + \mu \int_0^t f(s,B_s)ds + \sigma \int_0^t f(s,B_s)dB_s.$$

Hence, the process $\{Y_t = f(t,B_t), t \geq 0\}$ *satisfies the equation*

$$Y_t = 1 + \mu \int_0^t Y_s ds + \sigma \int_0^t Y_s dB_s.$$

It is termed geometric Brownian motion. *The equivalent differential form of this identity is the linear stochastic differential equation*

$$\begin{aligned} dY_t &= \mu Y_t dt + \sigma Y_t dB_t, \\ Y_0 &= 1. \end{aligned} \tag{3.22}$$

Black and Scholes proposed as model of a market with a single risky asset with initial value $S_0 = 1$*, the process* $S_t = Y_t$*. We have seen that such a process is in fact the solution to a linear stochastic differential equation (see (3.22)).*

In the particular case where $f : \mathbb{R} \to \mathbb{R}$ is a function in $\mathcal{C}^2$ (twice continuously differentiable), Theorem 3.2 gives the following version of Itô's formula:

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_0^t f'(X_s)u_s dB_s + \int_0^t f'(X_s)v_s ds \\ &+ \frac{1}{2}\int_0^t f''(X_s)u_s^2 ds. \end{aligned} \tag{3.23}$$

Proof of Theorem 3.2

Let $\Pi_n = \{0 = t_0^n < \cdots < t_{p_n}^n = t\}$ be a sequence of increasing partitions such that $\lim_{n\to\infty} |\Pi_n| = 0$. First, we consider the decomposition

We can write

$$
\begin{aligned}
f(t, X_t) - f(0, X_0) &= \sum_{i=0}^{p_n-1} \left[f(t_{i+1}^n, X_{t_{i+1}}^n) - f(t_i, X_{t_i}^n) \right] \\
&= \sum_{i=0}^{p_n-1} \left[f(t_{i+1}^n, X_{t_i}^n) - f(t_i^n, X_{t_i}^n) \right] \\
&\quad + \left[f(t_{i+1}^n, X_{t_{i+1}}^n) - f(t_{i+1}^n, X_{t_i}^n) \right] \\
&= \sum_{i=0}^{p_n-1} \left[\partial_s f(\bar{t}_i^n, X_{t_i}^n)(t_{i+1}^n - t_i^n) \right] \\
&\quad + \left[\partial_x f(t_{i+1}^n, X_{t_i}^n)(X_{t_{i+1}}^n - X_{t_i}^n) \right] \\
&+ \frac{1}{2} \sum_{i=0}^{p_n-1} \partial_{xx}^2 f(t_{i+1}^n, \bar{X}_i^n)(X_{t_{i+1}}^n - X_{t_i}^n)^2.
\end{aligned}
\tag{3.24}
$$

with $\bar{t}_i^n \in]t_i^n, t_{i+1}^n[$ and $\bar{X}_i^n$ an intermediate (random) point on the segment determined by $X_{t_i}^n$ and $X_{t_{i+1}}^n$.

In fact, this follows from a Taylor expansion of the function f up to the first order in the variable s (or the *mean-value theorem*), and up to the second order in the variable x. The asymmetry in the orders is due to the existence of quadratic variation of the processes involved. The expresion (3.24) is the analogue of (3.16). The former is much simpler for two reasons. Firstly, there is no s-variable; secondly, f is a polynomial of second degree, and therefore it has an exact Taylor expansion. But both formulas have the same structure. When passing to the limit as $n \to \infty$, we expect

$$
\begin{aligned}
&\sum_{i=0}^{p_n-1} \partial_s f(\bar{t}_i^n, X_{t_i}^n)(t_{i+1}^n - t_i^n) \to \int_0^t \partial_s f(s, X_s) ds \\
&\sum_{i=0}^{p_n-1} \partial_x f(t_{i+1}^n, X_{t_i}^n)(X_{t_{i+1}}^n - X_{t_i}^n) \to \int_0^t \partial_x f(s, X_s) u_s dB_s \\
&\qquad + \int_0^t \partial_x f(s, X_s) v_s ds \\
&\sum_{i=0}^{p_n-1} \partial_{xx} f(t_{i+1}^n, \bar{X}_i^n)(X_{t_{i+1}}^n - X_{t_i}^n)^2 \to \int_0^t \partial_{xx}^2 f(s, X_s) u_s^2 ds,
\end{aligned}
$$

in some topology.
This actually holds in the a.s. convergence (by taking if necessary a subsequence). As in Theorem 3.1, the proof requires a localization in Ω. However, this can be avoided by assuming some additional assumptions as follows: the process v is bounded; $u \in L^2_{a,T}$; the partial derivatives $\partial_x f$, ∂^2_{xx} are bounded. We shall give a proof of the theorem under these additional hypotheses.

Checking the convergences

First term

$$\sum_{i=0}^{p_n-1} \partial_s f(\bar{t}^n_i, X_{t^n_i})(t^n_{i+1} - t^n_i) \to \int_0^t \partial_s f(s, X_s) ds, \tag{3.25}$$

a.s.
Indeed

$$\begin{aligned}
&\left| \sum_{i=1}^{p_n-1} \partial_s f(\bar{t}^n_i, X_{t^n_i})(t^n_{i+1} - t^n_i) - \int_0^t \partial_s f(s, X_s) ds \right| \\
&\quad = \left| \sum_{i=1}^{p_n-1} \int_{t^n_i}^{t^n_{i+1}} \left[\partial_s f(\bar{t}^n_i, X_{t^n_i}) - \partial_s f(s, X_s) \right] ds \right| \\
&\quad \le \sum_{i=1}^{p_n-1} \int_{t^n_i}^{t^n_{i+1}} \left| \partial_s f(\bar{t}^n_i, X_{t^n_i}) - \partial_s f(s, X_s) \right| ds \\
&\quad \le t \sup_{1 \le i \le p_n - 1} \sup_{s \in [t^n_i, t^n_{i+1}]} \left| \left(\partial_s f(\bar{t}^n_i, X_{t^n_i}) - \partial_s f(s, X_s) \right) 1_{[t^n_i, t^n_{i+1}]}(s) \right|.
\end{aligned}$$

The continuity of $\partial_s f$ along with that of the process X implies

$$\lim_{n \to \infty} \sup_{1 \le i \le p_n - 1} \sup_{s \in [t^n_i, t^n_{i+1}]} \left| \left(\partial_s f(\bar{t}^n_i, X_{t^n_i}) - \partial_s f(s, X_s) \right) 1_{[t^n_i, t^n_{i+1}]}(s) \right| = 0.$$

This gives (3.25).

Second term

We prove

$$\sum_{i=1}^{p_n-1} \partial_x f\left(t^n_i, X_{t^n_i}\right) \left(X_{t^n_{i+1}} - X_{t^n_i}\right) \to \int_0^t \partial_x f(s, X_s) dX_s \tag{3.26}$$

in probability, which amounts to check two convengences:

$$\sum_{i=1}^{p_n-1} \partial_x f\left(t^n_i, X_{t^n_i}\right) \int_{t^n_i}^{t^n_{i+1}} u_t dB_t \to \int_0^t \partial_x f(s, X_s) u_s dB_s, \tag{3.27}$$

$$\sum_{i=1}^{p_n-1} \partial_x f\left(t^n_i, X_{t^n_i}\right) \int_{t^n_i}^{t^n_{i+1}} v_t dt \to \int_0^t \partial_x f(s, X_s) v_s ds. \tag{3.28}$$

We start with (3.27). We have

$$
\begin{aligned}
&E\left|\sum_{i=1}^{p_n-1} \partial_x f\left(t_i^n, X_{t_i^n}\right) \int_{t_i^n}^{t_{i+1}^n} u_t dB_t - \int_0^t \partial_x f(s, X_s) u_s dB_s\right|^2 \\
&= E\left|\sum_{i=1}^{p_n-1} \int_{t_i^n}^{t_{i+1}^n} \left[\partial_x f\left(t_i^n, X_{t_i^n}\right) - \partial_x f(s, X_s)\right] u_s dB_s\right|^2 \\
&= \sum_{i=1}^{p_n-1} E\left|\int_{t_i^n}^{t_{i+1}^n} \left[\partial_x f\left(t_i^n, X_{t_i^n}\right) - \partial_x f(s, X_s)\right] u_s dB_s\right|^2 \\
&= \sum_{i=1}^{p_n-1} \int_{t_i^n}^{t_{i+1}^n} E\left\{\left[\partial_x f\left(t_i^n, X_{t_i^n}\right) - \partial_x f(s, X_s)\right] u_s\right\}^2 ds,
\end{aligned}
$$

where we have applied successively that the stochastic integrals on disjoint intervals are independent to each other, they are centered random variables, along with the isometry property.
By the continuity of $\partial_x f$ and the process X,

$$
\sup_{1\le i\le p_n-1} \sup_{s\in[t_i^n, t_{i+1}^n]} \left|\left(\partial_x f\left(t_i^n, X_{t_i^n}\right) - \partial_x f(s, X_s)\right) 1_{[t_i^n, t_{i+1}^n]}(s)\right| \to 0, \tag{3.29}
$$

a.s. Then, by bounded convergence,

$$
\sup_{1\le i\le p_n-1} \sup_{s\in[t_i^n, t_{i+1}^n]} E\left|\left[\partial_x f\left(t_i^n, X_{t_i^n}\right) - \partial_x f(s, X_s)\right] u_s 1_{[t_i^n, t_{i+1}^n]}(s)\right|^2 \to 0.
$$

Therefore we obtain (3.27) in $L^2(\Omega)$.
For the proof of (3.28) we write

$$
\begin{aligned}
&\left|\sum_{i=1}^{p_n-1} \partial_x f\left(t_i^n, X_{t_i^n}\right) \int_{t_i^n}^{t_{i+1}^n} v_t dt - \int_0^t \partial_x f(s, X_s) v_s ds\right| \\
&= \left|\sum_{i=1}^{p_n-1} \int_{t_i^n}^{t_{i+1}^n} \left[\partial_x f\left(t_i^n, X_{t_i^n}\right) - \partial_x f(s, X_s)\right] v_s ds\right| \\
&\le \sum_{i=1}^{p_n-1} \int_{t_i^n}^{t_{i+1}^n} \left|\partial_x f\left(t_i^n, X_{t_i^n}\right) - \partial_x f(s, X_s)\right| |v_s| ds.
\end{aligned}
$$

By virtue of (3.29) and bounded convergence, we obtain (3.28) in the a.s. convergence.
Third term

Set $f_{n,i} = \partial^2_{xx} f(t^n_{i+1}, \bar{X}^n_i)$. We have to prove

$$\sum_{i=0}^{p_n-1} f_{n,i}(X_{t^n_{i+1}} - X_{t^n_i})^2 \to \int_0^t \partial^2_{xx} f(s, X_s) u_s^2 ds. \tag{3.30}$$

This will be a consequence of the following convergences

$$\sum_{i=0}^{p_n-1} f_{n,i} \left(\int_{t^n_i}^{t^n_{i+1}} u_s dB_s \right)^2 \to \int_0^t \partial^2_{xx} f(s, X_s) u_s^2 ds, \tag{3.31}$$

$$\sum_{i=0}^{p_n-1} f_{n,i} \left(\int_{t^n_i}^{t^n_{i+1}} u_s dB_s \right) \left(\int_{t^n_i}^{t^n_{i+1}} v_s ds \right) \to 0, \tag{3.32}$$

$$\sum_{i=0}^{p_n-1} f_{n,i} \left(\int_{t^n_i}^{t^n_{i+1}} v_s ds \right)^2 \to 0, \tag{3.33}$$

in the a.s. convergence.
Let us start by arguing on (3.31). We have

$$E\left| \sum_{i=0}^{p_n-1} f_{n,i} \left(\int_{t^n_i}^{t^n_{i+1}} u_s dB_s \right)^2 - \int_0^t \partial^2_{xx} f(s, X_s) u_s^2 ds \right|$$
$$\leq (T_1 + T_2),$$

with

$$T_1 = E\left| \sum_{i=0}^{p_n-1} f_{n,i} \left[\left(\int_{t^n_i}^{t^n_{i+1}} u_s dB_s \right)^2 - \int_{t^n_i}^{t^n_{i+1}} u_s^2 ds \right] \right|,$$
$$T_2 = E\left| \sum_{i=0}^{p_n-1} f_{n,i} \int_{t^n_i}^{t^n_{i+1}} u_s^2 ds - \int_0^t \partial^2_{xx} f(s, X_s) u_s^2 ds \right|.$$

Since $\partial^2_{x_k, x_l} f$ is bounded,

$$T_1 \leq CE \sum_{i=0}^{p_n-1} \left| \left(\int_{t^n_i}^{t^n_{i+1}} u_s dB_s \right)^2 - \int_{t^n_i}^{t^n_{i+1}} u_s^2 ds \right|.$$

This tends to zero as $n \to \infty$ (see Proposition 3.3).

As for T_2, we have

$$T_2 = \left| \sum_{i=1}^{p_n-1} \int_{t_i^n}^{t_{i+1}^n} \left[f_{n,i} - \partial_{xx}^2 f(s, X_s) \right] u_s^2 ds \right|$$
$$\le \sum_{i=1}^{p_n-1} \int_{t_i^n}^{t_{i+1}^n} \left| f_{n,i} - \partial_{xx}^2 f(s, X_s) \right| u_s^2 ds.$$

Using the continuity property, we have

$$\sup_{0 \le 0 \le p_n-1} \sup_{s \in [t_i^n, t_{i+1}^n]} \left| f_{n,i} - \partial_{xx}^2 f(s, X_s) \right| \to 0,$$

a.s. Then, by bounded convergence we obtain that T_2 converges to zero as $n \to \infty$. Hence we have proved (3.31) in $L^1(\Omega)$ (and therefore also a.s. for some subsequence).
Next we prove (3.32) in $L^1(\Omega)$. Indeed

$$E \left| \sum_{i=0}^{p_n-1} f_{n,i} \left(\int_{t_i^n}^{t_{i+1}^n} u_s dB_s \right) \left(\int_{t_i^n}^{t_{i+1}^n} v_s ds \right) \right|$$
$$\le C \sum_{i=0}^{p_n-1} E \left(\left| \int_{t_i^n}^{t_{i+1}^n} u_s dB_s \right| \left| \int_{t_i^n}^{t_{i+1}^n} v_s ds \right| \right)$$
$$\le C \sum_{i=0}^{p_n-1} \left(E \int_{t_i^n}^{t_{i+1}^n} u_s^2 ds \right)^{\frac{1}{2}} \left(E \left(\int_{t_i^n}^{t_{i+1}^n} |v_s| ds \right)^2 \right)^{\frac{1}{2}}$$
$$\le C \sum_{i=0}^{p_n-1} |t_{i+1}^n - t_i^n|^{\frac{1}{2}} \left(\int_{t_i^n}^{t_{i+1}^n} E|u_s|^2 ds \right)^{\frac{1}{2}} \left(E \left(\int_{t_i^n}^{t_{i+1}^n} |v_s|^2 ds \right) \right)^{\frac{1}{2}}$$
$$\le C \sup_i |t_{i+1}^n - t_i^n|^{\frac{1}{2}} \left(\sum_{i=0}^{p_n-1} \int_{t_i^n}^{t_{i+1}^n} E|u_s|^2 ds \right)^{\frac{1}{2}}$$
$$\times \left(\sum_{i=0}^{p_n-1} \int_{t_i^n}^{t_{i+1}^n} E|v_s|^2 ds \right)^{\frac{1}{2}},$$

which tends to zero as $n \to \infty$ a.s.
The proof of (3.33) is very easy. Indeed

$$\left| \sum_{i=0}^{p_n-1} f_{n,i} \left(\int_{t_i^n}^{t_{i+1}^n} v_s ds \right)^2 \right| \le C \sup_i \left(\int_{t_i^n}^{t_{i+1}^n} |v_s| ds \right) \int_0^t |v_s| ds.$$

The first factor on the right-hand side of this inequality tends to zero as $n \to \infty$, while the second one is bounded a.s. Therefore (3.33) holds in the a.s. convergence.
This ends the proof of the Theorem.

$\square$

3.4.2 Multidimensional version of Itô's formula

Consider a m-dimensional Brownian motion $\{(B_t^1, \cdots, B_t^m), t \geq 0\}$ and a real-valued Itô processes, as follows:

$$dX_t^i = \sum_{l=1}^{m} u_t^{i,l} dB_t^l + v_t^i dt, \tag{3.34}$$

$i = 1, \ldots, p$. We assume that each one of the processes $u_t^{i,l}$ belong to $\Lambda^2_{a,T}$ and that $\int_0^T |v_t^i| dt < \infty$, a.s. Following a similar plan as for Theorem 3.2, we will prove the following:

Theorem 3.3 *Let $f : [0, \infty) \times \mathbb{R}^p \mapsto \mathbb{R}$ be a function of class $\mathcal{C}^{1,2}$ and $X = (X^1, \ldots, X^p)$ be given by (3.34). Then*

$$\begin{aligned} f(t, X_t) = f(0, X_0) &+ \int_0^t \partial_s f(s, X_s) ds + \sum_{k=1}^{p} \int_0^t \partial_{x_k} f(s, X_s) dX_s^k \\ &+ \frac{1}{2} \sum_{k,l=1}^{p} \int_0^t \partial_{x_k, x_l} f(s, X_s) dX_s^k dX_s^l, \end{aligned} \tag{3.35}$$

where in order to compute $dX_s^k dX_s^l$, we have to apply the following rules

$$\begin{aligned} dB_s^k dB_t^l &= \delta_{k,l} ds, \\ dB_s^k ds &= 0, \\ (ds)^2 &= 0, \end{aligned} \tag{3.36}$$

where $\delta_{k,l}$ denotes the Kronecker symbol.

We remark that the identity (3.36) is a consequence of the independence of the components of the Brownian motion.

Example 3.3 *Consider the particular case $m = 1$, $p = 2$ and $f(x, y) = xy$. That is, f does not depend on t and we have denoted a generic point of $\mathbb{R}$ by (x, y). Then the above formula (3.35) yields*

$$X_t^1 X_t^2 = X_0^1 X_0^2 + \int_0^t X_s^1 dX_s^2 + \int_0^t X_s^2 dX_s^1 + \int_0^t \left(u_s^1 u_s^2\right) ds. \tag{3.37}$$

4 Applications of the Itô formula

This chapter is devoted to give some important results that use the Itô formula in some parts of their proofs.

4.1 Burkholder-Davis-Gundy inequalities

Theorem 4.1 *Let $u \in L^2_{a,T}$ and set $M_t = \int_0^t u_s dB_s$. Define*

$$M_t^* = \sup_{s\in[0,t]} |M_s| .$$

Then, for any $p > 0$, there exist two positive constants c_p, C_p such that

$$c_p E\left(\int_0^T u_s^2 ds\right)^{\frac{p}{2}} \leq E\left(M_T^*\right)^p \leq C_p E\left(\int_0^T u_s^2 ds\right)^{\frac{p}{2}} . \tag{4.1}$$

Proof: We will only prove here the right-hand side of (4.1) for $p \geq 2$. For this, we assume that the process $\{M_t, t \in [0,T]\}$ is bounded. This assumption can be removed by a localization argument.
Consider the function

$$f(x) = |x|^p,$$

for which we have that

$$f'(x) = p|x|^{p-1}\text{sign}(x),$$
$$f''(x) = p(p-1)|x|^{p-2},$$

for $x \neq 0$. Then, according to (3.23) we obtain

$$|M_t|^p = \int_0^t p|M_s|^{p-1}\text{sign}(M_s)u_s dB_s + \frac{1}{2}\int_0^t p(p-1)|M_s|^{p-2}u_s^2 ds.$$

Applying the expectation operator to both terms of the above identity yields

$$E\left(|M_t|^p\right) = \frac{p(p-1)}{2} E\left(\int_0^t |M_s|^{p-2}u_s^2 ds\right). \tag{4.2}$$

We next apply Hölder's inequality to the expectation with exponents $\frac{p}{p-2}$ and $q = \frac{p}{2}$ and get

$$\begin{aligned} E\left(\int_0^t |M_s|^{p-2}u_s^2 ds\right) &\leq E\left((M_t^*)^{p-2}\int_0^t u_s^2 ds\right) \\ &\leq \left[E\left(M_t^*\right)^p\right]^{\frac{p-2}{p}} \left[E\left(\int_0^t u_s^2 ds\right)^{\frac{p}{2}}\right]^{\frac{2}{p}} . \end{aligned} \tag{4.3}$$

Doob's inequality (see Theorem 8.1) implies

$$E\left(M_t^*\right)^p \leq \left(\frac{p}{p-1}\right)^p E(|M_t|^p).$$

Hence, by applying (4.2), (4.3), we obtain

$$\begin{aligned} E\left(M_t^*\right)^p &\leq \left(\frac{p}{p-1}\right)^p E(|M_t|^p) \\ &\leq \left(\frac{p}{p-1}\right)^p \frac{p(p-1)}{2} E\left(\int_0^t |M_s|^{p-2} u_s^2 ds\right) \\ &\leq \left(\frac{p}{p-1}\right)^p \frac{p(p-1)}{2} \left[E\left(M_t^*\right)^p\right]^{\frac{p-2}{p}} \left[E\left(\int_0^t u_s^2 ds\right)^{\frac{p}{2}}\right]^{\frac{2}{p}}. \end{aligned}$$

Since $1 - \frac{p-2}{p} = \frac{2}{p}$, from this inequality we obtain

$$E\left(M_t^*\right)^p \leq \left(\left(\frac{p}{p-1}\right)^p \frac{p(p-1)}{2}\right)^{\frac{p}{2}} E\left(\int_0^t u_s^2 ds\right)^{\frac{p}{2}}.$$

This ends the proof of the upper bound.

□

4.2 Representation of L^2 Brownian functionals

We already know that for any process $u \in L^2_{a,T}$, the stochastic integral process $\{\int_0^t u_s dB_s, t \in [0,T]\}$ is a martingale. The next result is a kind of converse statement. In the proof we shall use a technical ingredient that we write without giving a proof.
In the sequel we denote by $\mathcal{F}_T$ the σ-field generated by $(B_t, 0 \leq t \leq T)$.

Lemma 4.1 *The vector space generated by the random variables*

$$\exp\left(\int_0^T f(t) dB_t - \frac{1}{2}\int_0^T f^2(t) dt\right),$$

$f \in L^2([0,T])$, *is dense in* $L^2(\Omega, \mathcal{F}_T, P)$.

Theorem 4.2 *Let* $Z \in L^2(\Omega, \mathcal{F}_T)$. *There exists a unique process* $h \in L^2_{a,T}$ *such that*

$$Z = E(Z) + \int_0^T h_s dB_s. \tag{4.4}$$

Hence, for any martingale $M = \{M_t, t \in [0,T]\}$ bounded in L^2, there exist a unique process $h \in L^2_{a,T}$ and a constant C such that

$$M_t = C + \int_0^t h_s dB_s. \tag{4.5}$$

Proof: We start with the proof of (4.4). Let $\mathcal{H}$ be the vector space consisting of random variables $Z \in L^2(\Omega, \mathcal{F}_T)$ such that (4.4) holds. Firstly, we argue the uniqueness of h. This is an easy consequence of the isometry of the stochastic integral. Indeed, if there were two processes h and h' satisfying (4.4), then

$$\begin{aligned} E\left(\int_0^T (h_s - h'_s)^2 ds\right) &= E\left(\int_0^T (h_s - h'_s) dB_s\right)^2 \\ &= 0. \end{aligned}$$

This yields $h = h'$ in $L^2([0,T] \times \Omega)$.
We now turn to the existence of h. Any $Z \in \mathcal{H}$ satisfies

$$E(Z^2) = (E(Z))^2 + E\left(\int_0^T h_s^2 ds\right).$$

From this it follows that if $(Z_n, n \geq 1)$ is a sequence of elements of $\mathcal{H}$ converging to Z in $L^2(\Omega, \mathcal{F}_T)$, then the sequence $(h_n, n \geq 1)$ corresponding to the representations is Cauchy in $L^2_{a,T}$. Denoting by h the limit, we have

$$Z = E(Z) + \int_0^T h_s dB_s.$$

Hence $\mathcal{H}$ is closed in $L^2(\Omega, \mathcal{F}_T)$.
For any $f \in L^2([0,T])$, set

$$\mathcal{E}_t^f = \exp\left(\int_0^t f_s dB_s - \frac{1}{2}\int_0^t f_s^2 ds\right).$$

The random variable $\int_0^t f_s dB_s$ is Gaussian, centered and with variance $\int_0^t f_s^2 ds$. Hence, $E\left(\mathcal{E}_t^f\right) = 1$. Then, by the Itô formula,

$$\mathcal{E}_t^f = 1 + \int_0^t \mathcal{E}_s^f f(s) dB_s.$$

Consequently, the representation holds for $Z := \mathcal{E}_T^f$ and also any linear combination of such random variables belong to $\mathcal{H}$. The conclusion follows from Lemma 4.1.

Let us now prove the representation (4.5). The random variable M_T belongs to $L^2(\Omega)$. Hence, by applying the first part of the Theorem we have

$$M_T = E(M_0) + \int_0^T h_s dB_s,$$

for some $h \in L^2_{a,T}$. By taking conditional expectations we obtain

$$M_t = E(M_T|\mathcal{F}_t) = E(M_0) + \int_0^t h_s dB_s, \; 0 \leq t \leq T.$$

The proof of the theorem is now complete.

$\square$

Example 4.1 *Consider $Z = B_T^3$. In order to find the corresponding process h in the integral representation, we apply first Itô's formula, yielding*

$$B_T^3 = \int_0^T 3B_t^2 dB_t + 3\int_0^T B_t dt.$$

An integration by parts gives

$$\int_0^T B_t dt = TB_T - \int_0^T t dB_t.$$

Thus,

$$B_T^3 = \int_0^T 3\left[B_t^2 + T - t\right] dB_t.$$

Notice that $E(B_T)^3 = 0$. Then $h_t = 3\left[B_t^2 + T - t\right]$.

4.3 Girsanov's theorem

It is well known that if X is a multidimensional Gaussian random variable, any affine transformation brings X into a multidimensional Gaussian random variable as well. The simplest version of Girsanov's theorem extends this result to a Brownian motion. Before giving the precise statement and the proof, let us introduce some preliminaries.

Lemma 4.2 *Let L be a nonnegative random variable such that $E(L) = 1$. Set*

$$Q(A) = E(\mathbf{1}_A L), \quad A \in \mathcal{F}. \tag{4.6}$$

Then, Q defines a probability on $\mathcal{F}$, equivalent to P, with density given by L. Reciprocally, if P and Q are two probabilities on $\mathcal{F}$ and $P \ll Q$, then there exists a nonnegative random variable L such that $E(L) = 1$, and (4.6) holds.

Proof: It is clear that Q defines a σ-additive function on $\mathcal{F}$. Moreover, since

$$Q(\Omega) = E(1_\Omega L) = E(L) = 1,$$

Q is indeed a probability.
Let $A \in \mathcal{F}$ be such that $Q(A) = 0$. Since $L > 0$, a.s., we should have $P(A) = 0$. Reciprocally, for any $A \in \mathcal{F}$ with $P(A) = 0$, we have $Q(A) = 0$ as well.
The second assertion of the lemma is Radon-Nikodym theorem.

□

If we denote by E_Q the expectation operator with respect to the probability Q defined before, one has

$$E_Q(X) = E(XL).$$

Indeed, this formula is easily checked for simple random variables and then extended to any random variable $X \in L^1(\Omega)$ by the usual approximation argument.

Consider now a Brownian motion $\{B_t, t \in [0,T]\}$. Fix $\lambda \in \mathbb{R}$ and let

$$L_t = \exp\left(-\lambda B_t - \frac{\lambda^2}{2} t\right). \tag{4.7}$$

Notice that $L_t = \mathcal{E}_t^f$ with $f = -\lambda$ (see section 4.2).
Itô's formula yields

$$L_t = 1 - \int_0^t \lambda L_s dB_s.$$

Hence, the process $\{L_t, t \in [0,T]\}$ is a positive martingale and $E(L_t) = 1$, for any $t \in [0,T]$. Set

$$Q(A) = E\left(1_A L_T\right), \quad A \in \mathcal{F}_T. \tag{4.8}$$

By Lemma 4.2, the probability Q is equivalent to P on the σ-field $\mathcal{F}_T$.
By the martingale property of $\{L_t, t \in [0,T]\}$, the same conclusion is true on $\mathcal{F}_t$, for any $t \in [0,T]$. Indeed, let $A \in \mathcal{F}_t$, then

$$\begin{aligned} Q(A) &= E\left(1_A L_T\right) = E\left(E\left(1_A L_T | \mathcal{F}_t\right)\right) \\ &= E\left(1_A E\left(L_T | \mathcal{F}_t\right)\right) \\ &= E\left(1_A L_t\right). \end{aligned}$$

Next, we give a technical result.

Lemma 4.3 *Let X be a random variable and let $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$ such that*

$$E\left(e^{iuX}|\mathcal{G}\right) = e^{-\frac{u^2\sigma^2}{2}}.$$

Then, the random variable X is independent of the σ-field $\mathcal{G}$ and its probability law is Gaussian, zero mean and variance σ^2.

Proof: By the definition of the conditional expectation, for any $A \in \mathcal{G}$,

$$E\left(1_A e^{iuX}\right) = P(A)e^{-\frac{u^2\sigma^2}{2}}.$$

In particular, for $A := \Omega$, we see that the characteristic function of X is that of a $\mathrm{N}(0, \sigma^2)$. This proves the last assertion.
Moreover, for any $A \in \mathcal{G}$,

$$E_A\left(e^{iuX}\right) = e^{-\frac{u^2\sigma^2}{2}},$$

saying that the law of X conditionally to A is also $\mathrm{N}(0, \sigma^2)$. Thus,

$$P\left((X \le x) \cap A\right) = P(A)P_A\left(X \le x\right) = P(A)P\left(X \le x\right),$$

yielding the independence of X and $\mathcal{G}$.

□

Theorem 4.3 *(Girsanov's theorem) Let $\lambda \in \mathbb{R}$ and set*

$$W_t = B_t + \lambda t.$$

In the probability space $(\Omega, \mathcal{F}_T, Q)$, with Q given in (4.8), the process $\{W_t, t \in [0,T]\}$ is a standard Brownian motion.

Proof: We will check that in the probability space $(\Omega, \mathcal{F}_T, Q)$, any increment $W_t - W_s$, $0 \le s < t \le T$ is independent of $\mathcal{F}_s$ and has $\mathrm{N}(0, t-s)$ distribution. That is, for any $A \in \mathcal{F}_s$,

$$E_Q\left(e^{iu(W_t - W_s)}1_A\right) = E_Q\left(1_A e^{-\frac{u^2}{2}(t-s)}\right) = Q(A)e^{-\frac{u^2}{2}(t-s)}.$$

The conclusion will follow from Lemma 4.3.
Indeed, writing

$$L_t = \exp\left(-\lambda(B_t - B_s) - \frac{\lambda^2}{2}(t-s)\right)\exp\left(-\lambda B_s - \frac{\lambda^2}{2}s\right),$$

we have

$$E_Q\left(e^{iu(W_t-W_s)}1_A\right) = E\left(1_A e^{iu(W_t-W_s)}L_t\right)$$
$$= E\left(1_A e^{iu(B_t-B_s)+iu\lambda(t-s)-\lambda(B_t-B_s)-\frac{\lambda^2}{2}(t-s)}L_s\right).$$

Since $B_t - B_s$ is independent of $\mathcal{F}_s$, the last expression is equal to

$$E\left(1_A L_s\right) E\left(e^{(iu-\lambda)(B_t-B_s)}\right) e^{iu\lambda(t-s)-\frac{\lambda^2}{2}(t-s)}$$
$$= Q(A)e^{\frac{(iu-\lambda)^2}{2}(t-s)+iu\lambda(t-s)-\frac{\lambda^2}{2}(t-s)}$$
$$= Q(A)e^{-\frac{u^2}{2}(t-s)}.$$

The proof is now complete.

$\square$

5 Local time of Brownian motion and Tanaka's formula

This chapter deals with a very particular extension of Itô's formula. More precisely, we would like to have a decomposition of the positive submartingale $|B_t - x|$, for some fixed $x \in \mathbb{R}$ as in the Itô formula. Notice that the function $f(y) = |y - x|$ does not belong to $\mathcal{C}^2(\mathbb{R})$. A natural way to proceed is to regularize the function f, for instance by convolution with an approximation of the identity, and then, pass to the limit. Assuming that this is feasible, the question of identifying the limit involving the second order derivative remains open. This leads us to introduce a process termed the *local time of B at x* introduced by Paul Lévy.

Definition 5.1 *Let $B = \{B_t, t \geq 0\}$ be a Brownian motion and let $x \in \mathbb{R}$. The local time of B at x is defined as the stochastic process*

$$\begin{aligned} L(t,x) &= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t 1_{(x-\epsilon, x+\epsilon)}(B_s) ds \\ &= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \lambda\{s \in [0,t] : B_s \in (x-\epsilon, x+\epsilon)\}, \end{aligned} \tag{5.1}$$

where λ denotes the Lebesgue measure on $\mathbb{R}$.

We see that $L(t,x)$ measures the time spent by the process B at x during a period of time of length t. Actually, it is the density of this occupation time.

We shall see later that the above limit exists in L^2 (it also exists a.s.), a fact that it is not obvious at all.
Local time enters naturally in the extension of the Itô formula we alluded before. In fact, we have the following result.

Theorem 5.1 *For any $t \geq 0$ and $x \in \mathbb{R}$, a.s.,*

$$(B_t - x)^+ = (B_0 - x)^+ + \int_0^t 1_{[x,\infty)}(B_s) dB_s + \frac{1}{2} L(t,x), \tag{5.2}$$

where $L(t,x)$ is given by (5.1) in the L^2 convergence.

Proof: The heuristics of formula (5.2) is the following. In the sense of distributions, $f(y) = (y-x)^+$ has as first and second order derivatives, $f'(y) = 1_{[x,\infty)}(y)$, $f''(y) = \delta_x(y)$, respectively, where δ_x denotes the Dirac delta measure. Hence we expect a formula like

$$(B_t - x)^+ = (B_0 - x)^+ + \int_0^t 1_{[x,\infty)}(B_s) dB_s + \frac{1}{2} \int_0^t \delta_x(B_s) ds.$$

However, we have to give a meaning to the last integral.

Approximation procedure
We are going to approximate the function $f(y) = (y-x)^+$. For this, we fix $\epsilon > 0$ and define

$$f_{x\epsilon}(y) = \begin{cases} 0, & \text{if } y \le x-\epsilon \\ \frac{(y-x+\epsilon)^2}{4\epsilon}, & \text{if } x-\epsilon \le y \le x+\epsilon \\ y-x & \text{if } y \ge x+\epsilon \end{cases}$$

which clearly has as derivatives

$$f'_{x\epsilon}(y) = \begin{cases} 0, & \text{if } y \le x-\epsilon \\ \frac{(y-x+\epsilon)}{2\epsilon}, & \text{if } x-\epsilon \le y \le x+\epsilon \\ 1 & \text{if } y \ge x+\epsilon \end{cases}$$

and

$$f''_{x\epsilon}(y) = \begin{cases} 0, & \text{if } y < x-\epsilon \\ \frac{1}{2\epsilon}, & \text{if } x-\epsilon < y < x+\epsilon \\ 0 & \text{if } y > x+\epsilon \end{cases}$$

Let $\phi_n, n \ge 1$ be a sequence of $\mathcal{C}^\infty$ functions with compact supports decreasing to $\{0\}$. For instance we may consider the function

$$\phi(y) = c \exp\left(-(1-y^2)^{-1}\right) 1_{\{|y|<1\}},$$

with a constant c such that $\int_{\mathbb{R}} \phi(z)dz = 1$, and then take

$$\phi_n(y) = n\phi(ny).$$

Set

$$g_n(y) = [\phi_n * f_{x\epsilon}](y) = \int_{\mathbb{R}} f_{x\epsilon}(y-z)\phi_n(z)dz.$$

It is well-known that $g_n \in \mathcal{C}^\infty$, g_n and g'_n converge uniformly in $\mathbb{R}$ to $f_{x\epsilon}$ and $f'_{x\epsilon}$, respectively, and g''_n converges pointwise to $f''_{x\epsilon}$ except at the points $x+\epsilon$ and $x-\epsilon$.

We then have an Itô's formula for g_n, as follows:

$$g_n(B_t) = g_n(B_0) + \int_0^t g'_n(B_s)dB_s + \frac{1}{2}\int_0^t g''_n(B_s)ds. \tag{5.3}$$

Convergence of the terms in (5.3) as $n \to \infty$

The function $f'_{x\epsilon}$ is bounded. The function g'_n is also bounded. Indeed,

$$\begin{aligned}|g'_n(y)| &= \int_{\mathbb{R}} f'_{x\epsilon}(y-z)\phi_n(z)dz \\ &= \int_{-\frac{1}{n}}^{\frac{1}{n}} f'_{x\epsilon}(y-z)\phi_n(z)dz \\ &\le 2\|f'_{x\epsilon}\|_\infty.\end{aligned}$$

Moreover,

$$\left|g'_n(B_s)1_{[0,t]} - f'_{x\epsilon}(B_s)1_{[0,t]}\right| \to 0,$$

uniformly in t and in ω. Hence, by bounded convergence,

$$E\int_0^t |g'_n(B_s) - f'_{x\epsilon}(B_s)|^2 \to 0.$$

Then, the isometry property of the stochastic integral implies

$$E\left|\int_0^t [g'_n(B_s) - f'_{x\epsilon}(B_s)]\, dB_s\right|^2 \to 0,$$

as $n \to \infty$.

We next deal with the second order term. Since the law of each B_s has a density, for each $s > 0$,

$$P\{B_s = x+\epsilon\} = P\{B_s = x-\epsilon\} = 0.$$

Thus, for any $s > 0$,

$$\lim_{n\to\infty} g''_n(B_s) = f''_{x\epsilon}(B_s),$$

a.s. Using Fubini's theorem, we see that this convergence also holds, for almost every s, a.s. In fact,

$$\begin{aligned}&\int_0^t ds\int_\Omega dP 1_{\{f''_{x,\epsilon}(B_s)\neq \lim_{n\to\infty} g''_n(B_s)\}} \\ &= \int_\Omega dP\int_0^t ds 1_{\{f''_{x,\epsilon}(B_s)\neq \lim_{n\to\infty} g''_n(B_s)\}} = 0.\end{aligned}$$

We have

$$\sup_{y\in\mathbb{R}} |g''_n(y)| \le \frac{1}{2\epsilon}.$$

Indeed,

$$
\begin{aligned}
|g_n''(y)| &= \frac{1}{2\epsilon}\left|\int_{\mathbb{R}} \phi_n(z)1_{(x-\epsilon,x+\epsilon)}(y-z)dz\right| \\
&\le \frac{1}{2\epsilon}\int_{y-x-\epsilon}^{y-x+\epsilon} |\phi_n(z)|dz \le \frac{2}{2\epsilon}.
\end{aligned}
$$

Then, by bounded convergence

$$
\int_0^t g_n''(B_s)ds \to \int_0^t f_{x\epsilon}''(B_s)ds,
$$

a.s. and in L^2.
Thus, passing to the limit the expression (5.3) yields

$$
f_{x\epsilon}(B_t) = f_{x\epsilon}(B_0) + \int_0^t f_{x\epsilon}'(B_s)dB_s + \frac{1}{2}\int_0^t \frac{1}{2\epsilon}1_{(x-\epsilon,x+\epsilon)}(B_s)ds. \tag{5.4}
$$

Convergence as $\epsilon \to 0$ of (5.4)
Since $f_{x\epsilon}(y) \to (y-x)^+$ as $\epsilon \to 0$ and

$$
|f_{x\epsilon}(B_t) - f_{x\epsilon}(B_0)| \le |B_t - B_0|,
$$

we have

$$
f_{x\epsilon}(B_t) - f_{x\epsilon}(B_0) \to (B_t - x)^+ - (B_0 - x)^+,
$$

in L^2.
Moreover,

$$
\begin{aligned}
E\left[\int_0^t \left(f_{x,\epsilon}'(B_s) - 1_{[x,\infty)}(B_s)\right)^2 ds\right] &\le E\left[\int_0^t 1_{(x-\epsilon,x+\epsilon)}(B_s)ds\right) \\
&\le \int_0^t \frac{2\epsilon}{\sqrt{2\pi s}}ds.
\end{aligned}
$$

that clearly tends to zero as $\epsilon \to 0$. Hence, by the isometry property of the stochastic integral

$$
\int_0^t f_{x,\epsilon}'(B_s)dB_s \to \int_0^t 1_{[x,\infty)}(B_s)dB_s,
$$

in L^2.
Consequently, we have proved that

$$
\int_0^t \frac{1}{2\epsilon}1_{(x-\epsilon,x+\epsilon)}(B_s)ds
$$

converges in L^2 as $\epsilon \to 0$ and that formula (5.2) holds.

$\square$

We give without proof two further properties of local time.

1. The property of local time as a density of occupation measure is made clear by the following identity, valid por any $t \geq 0$ and every $a \leq b$:

$$\int_a^b L(t,x)dx = \int_0^t 1_{(a,b)}(B_s)ds.$$

2. The stochastic integral $\int_0^t 1_{[x,\infty)}(B_s)dB_s$ has a jointly continuous version in $(t,x) \in (0,\infty) \times \mathbb{R}$. Hence, by (5.2) so does the local time $\{L(t,x),(t,x) \in (0,\infty) \times \mathbb{R}\}$.

The next result, which follows easily from Theorem 5.1 is known as *Tanaka's formula.*

Theorem 5.2 *For any $(t,x) \in [0,\infty) \times \mathbb{R}$, we have*

$$|B_t - x| = |B_0 - x| + \int_0^t sign(B_s - x)dB_s + L(t,x). \tag{5.5}$$

Proof: We will use the following relations: $|x| = x^+ + x^-$, $x^- = \max(-x,0) = (-x)^+$. Hence, by virtue of (5.2), we only need a formula for $(-B_t + x)^+$. Notice that we already have it, since the process $-B$ is also a Brownian motion. More precisely,

$$(-B_t + x)^+ = (-B_0 + x)^+ + \int_0^t 1_{[-x,\infty)}(-B_s)d(-B_s) + \frac{1}{2}L^-(t,-x),$$

where we have denoted by $L^-(t,-x)$ the local time of $-B$ at $-x$. We have the following facts:

$$\begin{aligned}
\int_0^t 1_{[-x,\infty)}(-B_s)d(-B_s) &= -\int_0^t 1_{(-\infty,x]}(B_s)dB_s, \\
L^-(t,-x) &= \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^t 1_{(-x-\epsilon,-x+\epsilon)}(-B_s)ds \\
&= \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^t 1_{(x-\epsilon,x+\epsilon)}(B_s)ds \\
&= L(t,x),
\end{aligned}$$

where the limit is in $L^2(\Omega)$.

Thus, we have proved

$$(B_t - x)^- = (B_0 - x)^- - \int_0^t 1_{(-\infty,x]}(B_s)dB_s + \frac{1}{2}L(t,x). \tag{5.6}$$

Adding up (5.2) and (5.6) yields (5.5). Indeed

$$1_{[x,\infty)}(B_s) - 1_{(-\infty,x]}(B_s) = \begin{cases} 1, & \text{if } B_s > x \\ -1 & \text{if } B_s < x \\ 0 & \text{if } B_s = x \end{cases}$$

which is identical to sign $(B_s - x)$.

□

6 Stochastic differential equations

In this section we shall introduce stochastic differential equations driven by a multi-dimensional Brownian motion. Under suitable properties on the coefficients, we shall prove a result on existence and uniqueness of solution. Then we shall establish properties of the solution, like existence of moments of any order and the Hölder property of the sample paths.

The setting

We consider a d-dimensional Brownian motion $B = \{B_t = (B_t^1, \dots, B_t^d), t \geq 0\}$, $B_0 = 0$, defined on a probability space $(\Omega, \mathcal{F}, P)$, along with a filtration $(\mathcal{F}_t, t \geq 0)$ satisfying the following properties:

1. B is adapted to $(\mathcal{F}_t, t \geq 0)$,
2. the σ-field generated by $\{B_u - B_t, u \geq t\}$ is independent of $(\mathcal{F}_t, t \geq 0)$.

We also consider functions

$$b : [0, \infty) \times \mathbb{R}^m \to \mathbb{R}^m, \quad \sigma : [0, \infty) \times \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^d; \mathbb{R}^m).$$

When necessary we will use the description

$$b(t, x) = \left(b^i(t, x)\right)_{1 \leq i \leq m}, \quad \sigma(t, x) = \left(\sigma_j^i(t, x)\right)_{1 \leq i \leq m, 1 \leq j \leq d}.$$

By a stochastic differential equation, we mean an expression of the form

$$\begin{aligned} dX_t &= \sigma(t, X_t) dB_t + b(t, X_t) dt, \ t \in (0, \infty), \\ X_0 &= x, \end{aligned} \tag{6.1}$$

where x is a m-dimensional random vector independent of the Brownian motion.

We can also consider any time value $u \geq 0$ as the initial one. In this case, we must write $t \in (u, \infty)$ and $X_u = x$ in (6.1). For the sake of simplicity we will assume here that x is deterministic.

The formal expression (6.1) has to be understood as follows:

$$X_t = x + \int_0^t \sigma(s, X_s) dB_s + \int_0^t b(s, X_s) ds, \tag{6.2}$$

or coordinate-wise,

$$X_t^i = x^i + \sum_{j=1}^d \int_0^t \sigma_j^i(s, X_s) dB_s^j + \int_0^t b^i(s, X_s) ds,$$

$i = 1, \ldots, m$.

Strong existence and path-wise uniqueness

We now give the notions of existence and uniqueness of solution that will be considered throughout this chapter.

Definition 6.1 *A m-dimensional stochastic process $(X_t, t \geq 0)$ measurable and $\mathcal{F}_t$-adapted is a strong solution to (6.2) if the following conditions are satisfied:*

1. *The processes $(\sigma_j^i(s, X_s), s \geq 0)$ belong to $L^2_{a,\infty}$, for any $1 \leq i \leq m$, $1 \leq j \leq d$.*

2. *The processes $(b^i(s, X_s), s \geq 0)$ belong to $L^1_{a,\infty}$, for any $1 \leq i \leq m$.*

3. *Equation (6.2) holds true for the fixed Brownian motion defined before, for any $t \geq 0$, a.s.*

Definition 6.2 *The equation (6.2) has a path-wise unique solution if any two strong solutions X_1 and X_2 in the sense of the previous definition are indistinguishable, that is,*

$$P\{X_1(t) = X_2(t), \textit{for any } t \geq 0\} = 1.$$

Hypotheses on the coefficients

We shall refer to (H) for the following set of hypotheses.

1. Linear growth:

$$\sup_t \left[|b(t,x)| + |\sigma(t,x)|\right] \leq L(1 + |x|). \tag{6.3}$$

2. Lipschitz in the x variable, uniformly in t:

$$\sup_t \left[|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)|\right] \leq L|x - y|. \tag{6.4}$$

In (6.3), (6.4), L stands for a positive constant.

6.1 Examples of stochastic differential equations

When the functions σ and b have a *linear structure*, the solution to (6.2) admits an explicit form. This is not surprising as it is indeed the case for ordinary differential equations. We deal with this question in this section. More precisely, suppose that

$$\sigma(t,x) = \Sigma(t) + F(t)x, \tag{6.5}$$

$$b(t,x) = c(t) + D(t)x. \tag{6.6}$$

Example 1 Assume for simplicity $d = m = 1$, $\sigma(t,x) = \Sigma(t)$, $b(t,x) = c(t) + Dx$, $t \geq 0$ and $D \in \mathbb{R}$. Now equation (6.2) reads

$$X_t = X_0 + \int_0^t \Sigma(s)dB_s + \int_0^t [c(s) + DX_s]ds,$$

and has a unique solution given by

$$X_t = X_0 e^{Dt} + \int_0^t e^{D(t-s)}(c(s)ds + \Sigma(s)dB_s). \tag{6.7}$$

To check (6.7) we proceed as in the deterministic case. First we consider the equation

$$dX_t = DX_t dt,$$

with initial condition X_0, which solution is

$$X_t = X_0 e^{Dt},\ t \geq 0.$$

The we use the *variation of constants procedure* and write

$$X_t = X_0(t)e^{Dt}.$$

A priori $X_0(t)$ may be random. However, since e^{Dt} is differentiable, the Itô differential of X_t is given by

$$dX_t = dX_0(t)e^{Dt} + X_0(t)e^{Dt}Ddt.$$

Equating the right-hand side of the preceding identity with

$$\Sigma(t)dB_t + (c(t) + X_t D)\, dt$$

yields

$$
\begin{aligned}
&dX_0(t)e^{Dt} + X_t D dt \\
&= \Sigma(t)dB_t + (c(t) + X_t D)\, dt,
\end{aligned}
$$

that is

$$
dX_0(t) = e^{-Dt} \left[\Sigma(t)dB_t + c(t)dt\right].
$$

In integral form

$$
X_0(t) = x + \int_0^t e^{-Ds} \left[\Sigma(s)dB_s + c(s)ds\right].
$$

Plugging the right-hand side of this equation in $X_t = X_0(t)e^{Dt}$ yields (6.7).

A particular example of the class of equations considered before is *Langevin Equation*:

$$
dX_t = \alpha dB_t - \beta X_t dt, \; t > 0,
$$

$X_0 = x_0 \in \mathbb{R}$, where $\alpha \in \mathbb{R}$ and $\beta > 0$. Here X_t stands for the velocity at time t of a free particle that performs a Brownian motion different from the B_t in the equation. The solution to this equation is given by

$$
X_t = e^{-\beta t}x_0 + \alpha \int_0^t e^{-\beta(t-s)} dB_s.
$$

Notice that $\{X_t, t \geq 0\}$ defines a Gaussian process.

6.2 A result on existence and uniqueness of solution

This section is devoted to prove the following result.

Theorem 6.1 *Assume that the functions σ, and b satisfy the assumptions (H). Then there exists a path-wise unique strong solution to (6.2).*

Before giving a proof of this theorem we recall a version of Gronwall's lemma that will be used repeatedly in the sequel.

Lemma 6.1 *Let $u, v : [\alpha, \beta] \to \mathbb{R}_+$ be functions such that u is Lebesgue integrable and v is measurable and bounded. Assume that*

$$
v(t) \leq c + \int_\alpha^t u(s)v(s)ds, \tag{6.8}
$$

for some constant $c \geq 0$ and for any $t \in [\alpha, \beta]$. Then

$$
v(t) \leq c \exp\left(\int_\alpha^t u(s)ds\right). \tag{6.9}
$$

Proof of Theorem 6.1
Let us introduce Picard's iteration scheme

$$X_t^0 = x,$$
$$X_t^n = x + \int_0^t \sigma(s, X_s^{n-1})dB_s + \int_0^t b(s, X_s^{n-1})ds, n \geq 1, \tag{6.10}$$

$t \geq 0$. Let us restrict the time interval to $[0,T]$, with $T > 0$. We shall prove that the sequence of stochastic processes defined recursively by (6.10) converges uniformly to a process X which is a strong solution of (6.2). Eventually, we shall prove path-wise uniqueness.
Step 1: We prove by induction on n that for any $t \in [0,T]$,

$$E\left\{\sup_{0\leq s\leq t} |X_s^n|^2\right\} < \infty. \tag{6.11}$$

Indeed, this property is clearly true if $n = 0$, since in this case X_t^0 is constant and equal to x. Suppose that (6.11) holds true for $n = 0, \ldots, m-1$. By applying Burkholder's and Hölder's inequality, we reach

$$\begin{aligned}
E\left\{\sup_{0\leq s\leq t} |X_s^n|^2\right\} &\leq C\Bigg[x + E\left(\sup_{0\leq s\leq t}\left|\int_0^s \sigma(u, X_u^{m-1})dB_u\right|^2\right) \\
&+ E\left(\sup_{0\leq s\leq t}\left|\int_0^s b(u, X_u^{m-1})du\right|^2\right)\Bigg] \\
&\leq C\Bigg[x + E\left(\int_0^t \left|\sigma(u, X_u^{m-1})\right|^2 du\right) \\
&+ E\left(\int_0^t \left|b(u, X_u^{m-1})\right|^2 du\right)\Bigg] \\
&\leq C\left[x + E\left(\int_0^t \left(1 + |X_u^{m-1}|^2\right) du\right)\right] \\
&\leq C\left[x + T + TE\left(\sup_{0\leq s\leq T} |X_s^{m-1}|^2\right)\right].
\end{aligned}$$

Hence (6.11) is proved.
The assumptions (H) along with (6.11) imply that $\sigma\left(s, X_s^{n-1}\right) \in L_{a,T}^2$ and $b\left(s, X_s^{n-1}\right) \in L^2(\Omega \times [0,T])$.
Step 2: As in Step 1, we prove by induction on n that

$$E\left\{\sup_{0\leq s\leq t} |X_s^{n+1} - X_s^n|^2\right\} \leq \frac{(Ct)^{n+1}}{(n+1)!}. \tag{6.12}$$

Indeed, consider first the case $n = 0$ for which we have

$$X_s^1 - x = \int_0^s \sigma(u,x)dB_u + \int_0^s b(s,x)ds.$$

Burkolder's inequality yields

$$\begin{aligned} E\left(\sup_{0\le s\le t}\left|\int_0^s \sigma(u,x)dB_u\right|^2\right) &\le C\int_0^t |\sigma(u,x)|^2 du \\ &\le Ct(1+|x|^2). \end{aligned}$$

Similarly,

$$\begin{aligned} E\left(\sup_{0\le s\le t}\left|\int_0^s b(u,x)du\right|^2\right) &\le Ct\int_0^t |b(u,x)|^2 du \\ &\le Ct^2(1+|x|^2). \end{aligned}$$

With this, (6.12) is established for $n = 0$.
Assume that (6.12) holds for natural numbers $m \le n-1$. Then, as we did for $n = 0$, we can consider the decomposition

$$E\left\{\sup_{0\le s\le t}|X_s^{n+1} - X_s^n|^2\right\} \le 2(A(t)+B(t)),$$

with

$$\begin{aligned} A(t) &= E\left\{\sup_{0\le s\le t}\left|\int_0^s \left(\sigma(u,X_u^n) - \sigma(u,X_u^{n-1})\right) dB_u\right|^2\right\}, \\ B(t) &= E\left\{\sup_{0\le s\le t}\left|\int_0^s \left(b(u,X_u^n) - b(u,X_u^{n-1})\right) du\right|^2\right\}. \end{aligned}$$

Using first Burkholder's inequality and then Hölder's inequality along with the Lipschitz property of the coefficient σ, we obtain

$$A(t) \le C(L)\int_0^t E(|X_s^n - X_s^{n-1}|^2)ds.$$

By the induction assumption we can upper bound the last expression by

$$C(L)\int_0^t \frac{(Cs)^n}{n!} \le C(T,L)\frac{(Ct)^{n+1}}{(n+1)!}.$$

Similarly, applying Hölder's inequality along with the Lipschitz property of the coefficient b and the induction assumption, yield

$$B(t) \leq C(T, L)\frac{(Ct)^{n+1}}{(n+1)!}.$$

Step 3: The sequence of processes $\{X_t^n, t \in [0,T]\}$, $n \geq 0$, converges uniformly in t to a stochastic process $\{X_t, t \in [0,T]\}$ which satisfies (6.2). Indeed, applying first Chebychev's inequality and then (6.12), we have

$$P\left\{\sup_{0\leq t\leq T}\left|X_t^{n+1} - X_t^n\right| > \frac{1}{2^n}\right\} \leq 2^{2n}\frac{(Ct)^{n+1}}{(n+1)!},$$

which clearly implies

$$\sum_{n=0}^{\infty} P\left\{\sup_{0\leq t\leq T}\left|X_t^{n+1} - X_t^n\right| > \frac{1}{2^n}\right\} < \infty.$$

Hence, by the first Borel-Cantelli's lemma

$$P\left\{\liminf_n \left\{\sup_{0\leq t\leq T}\left|X_t^{n+1} - X_t^n\right| \leq \frac{1}{2^n}\right\}\right\} = 1.$$

In other words, for each ω a.s., there exists a natural number $m_0(\omega)$ such that

$$\sup_{0\leq t\leq T}\left|X_t^{n+1} - X_t^n\right| \leq \frac{1}{2^n},$$

for any $n \geq m_0(\omega)$. The Weierstrass criterion for convergence of series of functions then implies that

$$X_t^m = x + \sum_{k=0}^{m-1}[X_t^{k+1} - X_t^k]$$

converges uniformly on $[0,T]$, a.s. Let us denote by $X = \{X_t, t \in [0,T]\}$ the limit. Obviously the process X has a.s. continuous paths.

To conclude the proof, we must check that X satisfies equation (6.2) on $[0,T]$. The continuity properties of σ and b imply the convergences

$$\sigma(t, X_t^n) \to \sigma(t, X_t),$$
$$b(t, X_t^n) \to b(t, X_t),$$

as $n \to \infty$, uniformly in $t \in [0,T]$, a.s.

Therefore,

$$\left|\int_0^t [b(s,X_s^n)-b(s,X_s)]\,ds\right| \le L\int_0^t |X_s^n - X_s|\,ds$$
$$\le L \sup_{0\le s\le t} |X_s^n - X_s| \to 0,$$

as $n\to\infty$, a.s. This proves the a.s. convergence of the sequence of the path-wise integrals.
As for the stochastic integrals, we will prove

$$\int_0^t \sigma(s,X_s^n)dB_s \to \int_0^t \sigma(s,X_s)dB_s \tag{6.13}$$

as $n\to\infty$ with the convergence in probability.
Indeed, applying the extension of Lemma 3.5 to processes of $\Lambda_{a,T}$, we have for each $\epsilon, N>0$,

$$P\left\{\left|\int_0^t (\sigma(s,X_s^n)-\sigma(s,X_s))\,dB_s\right| > \epsilon\right\}$$
$$\le P\left\{\int_0^t |\sigma(s,X_s^n)-\sigma(s,X_s)|^2\,ds > N\right\} + \frac{N}{\epsilon^2}.$$

The first term in the right-hand side of this inequality converges to zero as $n\to\infty$. Since $\epsilon, N>0$ are arbitrary, this yields the convergence stated in (6.13).
Summarising, by considering if necessary a subsequence $\{X_t^{n_k}, t\in[0,T]\}$, we have proved the a.s. convergence, uniformly in $t\in[0,T]$, to a stochastic process $\{X_t, t\in[0,T]\}$ which satisfies (6.2), and moreover

$$E\{\sup_{0\le t\le T}|X_t|^2\} < \infty.$$

In order to conclude that X is a strong solution to (6.1) we have to check that the required measurability and integrability conditions hold. This is left as an exercise to the reader.

Step 4: Path-wise uniqueness.

Let X_1 and X_2 be two strong solutions to (6.1). Proceeding in a similar way as in Step 2, we easily get

$$E\left(\sup_{0\le u\le t}|X_1(u)-X_2(u)|^2\right) \le C\int_0^t E\left(\sup_{0\le u\le s}|X_1(u)-X_2(u)|^2\right)ds.$$

Hence, from Lemma 6.1 we conclude

$$E\left(\sup_{0\le u\le T}|X_1(u)-X_2(u)|^2\right)=0,$$

proving that X_1 and X_2 are indistinguishable.

□

6.3 Some properties of the solution

We start this section by studying the L^p-moments of the solution to (6.2).

Theorem 6.2 *Assume the same assumptions as in Theorem 6.1 and suppose in addition that the initial condition is a random variable X_0, independent of the Brownian motion. Fix $p\in[2,\infty)$ and $t\in[0,T]$. There exists a positive constant $C=C(p,t,L)$ such that*

$$E\left(\sup_{0\le s\le t}|X_s|^p\right)\le C\left(1+E|X_0|^p\right). \tag{6.14}$$

Proof: From (6.2) it follows that

$$\begin{aligned}E\left(\sup_{0\le s\le t}|X_s|^p\right)\le C(p)\Big[E|X_0|^p+E\left(\sup_{0\le s\le t}\left|\int_0^s\sigma(u,X_u)dB_u\right|^p\right)\\ E\left(\sup_{0\le s\le t}\left|\int_0^s b(u,X_u)du\right|^p\right)\Big].\end{aligned}$$

Applying first Burkholder's inequality and then Hölder's inequality yield

$$\begin{aligned}E\left(\sup_{0\le s\le t}\left|\int_0^s\sigma(u,X_u)dB_u\right|^p\right)&\le C(p)E\left(\int_0^t|\sigma(s,X_s)|^2ds\right)^{\frac p2}\\ &\le C(p,t)E\left(\int_0^t|\sigma(s,X_s)|^pds\right)\le C(p,L,t)\int_0^t(1+E|X_s|^p)ds\\ &\le C(p,L,t)\left(1+\int_0^t E\left(\sup_{0\le u\le s}|X_u|^p\right)\right)ds.\end{aligned}$$

For the pathwise integral, we apply Hölder's inequality to obtain

$$\begin{aligned}E\left(\sup_{0\le s\le t}\left|\int_0^s b(u,X_u)du\right|^p\right)&\le C(p,t)E\left(\int_0^t|b(s,X_s)|^pds\right)\\ &\le C(p,L,t)\left(1+\int_0^t E\left(\sup_{0\le u\le s}|X_u|^p\right)ds\right).\end{aligned}$$

Define

$$\varphi(t) = E\left(\sup_{0\le s\le t} |X_s|^p\right).$$

We have established that

$$\varphi(t) \le C(p,L,t)\left(E|X_0|^p + 1 + \int_0^t \varphi(s)ds\right).$$

Then, with Lemma 6.1 we end the proof of (6.14).

□

It is clear that the solution to (6.2) depends on the initial value X_0. Consider two initial conditions X_0, Y_0 (remember that there should be m-dimensional random vectors independent of the Brownian motion). Denote by $X(X_0)$, $X(Y_0)$ the corresponding solutions to (6.2). With a very similar proof as that of Theorem 6.2 we can obtain the following.

Theorem 6.3 *The assumptions are the same as in Theorem 6.1. Then*

$$E\left(\sup_{0\le s\le t} |X_s(X_0) - X_s(Y_0)|^p\right) \le C(p,L,t)\left(E|X_0 - Y_0|^p\right), \tag{6.15}$$

for any $p \in [2,\infty)$, where $C(p,L,t)$ is some positive constant depending on p, L and t.

The sample paths of the solution of a stochastic differential equation possess the same regularity as those of the Brownian motion. We next discuss this fact.

Theorem 6.4 *The assumptions are the same as in Theorem 6.1. Let $p \in [2,\infty)$, $0 \le s \le t \le T$. There exists a positive constant $C = C(p,L,T)$ such that*

$$E\left(|X_t - X_s|^p\right) \le C(p,L,T)\left(1 + E|X_0|^p\right)|t-s|^{\frac{p}{2}}. \tag{6.16}$$

Proof: By virtue of (6.2) we can write

$$\begin{aligned} &E\left(|X_t - X_s|^p\right) \\ &\quad \le C(p)\left[E\left|\int_s^t \sigma(u,X_u)dB_u\right|^p + E\left|\int_s^t b(u,X_u)du\right|^p\right]. \end{aligned}$$

Burkholder's inequality and then Hölder's inequality with respect to Lebesgue measure on $[s,t]$ yield

$$
\begin{aligned}
E\left|\int_s^t \sigma(u,X_u)dB_u\right|^p &\le C(p)E\left(\int_s^t |\sigma(u,X_u)|^2 du\right)^{\frac{p}{2}}\\
&\le C(p)|t-s|^{\frac{p}{2}-1}E\left(\int_s^t |\sigma(u,X_u)|^p du\right)\\
&\le C(p,L,T)|t-s|^{\frac{p}{2}-1}\int_s^t \left(1+E(|X_u|^p)\right)du.
\end{aligned}
$$

By using the estimate (6.14) of Theorem 6.2, we have

$$
\begin{aligned}
\int_s^t \left(1+E(|X_u|^p)\right)du &\le C(p,T,L)|t-s|E\left(\sup_{0\le u\le t}|X_u|^p\right)\\
&\le C(p,T,L)|t-s|\left(1+E|X_0|^p\right). \qquad (6.17)
\end{aligned}
$$

This ends the estimate of the L^p moment of the stochastic integral.
The estimate of the path-wise integral follows from Hölder's inequality and (6.17). Indeed, we have

$$
\begin{aligned}
E\left|\int_s^t b(u,X_u)du\right|^p &\le |t-s|^{p-1}\int_s^t E\left(|b(u,X_u)|^p\right)du\\
&\le C(p,L)|t-s|^{p-1}\int_s^t \left(1+E(|X_u|^p)\right)du\\
&\le C(p,T,L)|t-s|^p\left(1+E|X_0|^p\right).
\end{aligned}
$$

Hence we have proved (6.16)

□

We can now apply Kolmogorov's continuity criterion (see Proposition 2.2) to prove the following.

Corollary 6.1 *With the same assumptions as in Theorem 6.1, we have that the sample paths of the solution to (6.2) are Hölder continuous of degree* $\alpha\in(0,\frac{1}{2})$.

Remark 6.1 *Assume that in Theorem 6.3 the initial conditions are deterministic and are denoted by* x *and* y*, respectively. An extension of Kolmogorov's continuity criterion to stochastic processes indexed by a multi-dimensional parameter yields that the sample paths of the stochastic process* $\{X_t(x), t\in[0,T], x\in\mathbb{R}^m\}$ *are jointly Hölder continuous in* (t,x) *of degree* $\alpha<\frac{1}{2}$ *in* t *and* $\beta<1$ *in* x*, respectively.*

6.4 Markov property of the solution

In Section 2.5 we discussed the Markov property of a real-valued Brownian motion. With the obvious changes $\mathbb{R}$ into $\mathbb{R}^n$, with arbitrary $n \geq 1$ we can see that the property extends to multi-dimensional Brownian motion. In this section we prove that the solution to the sde (6.2) inherits the Markov property from Brownian motion. To establish this fact we need some preliminary results.

Lemma 6.2 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(E, \mathcal{E})$ be a measurable space. Consider two independent sub-σ-fields of $\mathcal{F}$, denoted by $\mathcal{G}$, $\mathcal{H}$, respectively, along with mappings*

$$X : (\Omega, \mathcal{H}) \to (E, \mathcal{E}),$$

and

$$\Psi : (E \times \Omega, \mathcal{E} \otimes \mathcal{G}) \to \mathbb{R}^m,$$

with $\omega \mapsto \Psi(X(\omega), \omega)$ in $L^1(\Omega)$.
Then,

$$E\left(\Psi\left(X, \cdot\right) | \mathcal{H}\right) = \Phi(X)(\cdot), \tag{6.18}$$

with $\Phi(x)(\cdot) = E\left(\Psi(x, \cdot)\right)$.

Proof: Assume first that $\Psi(x, \omega) = f(x)Z(\omega)$ with a $\mathcal{G}$-measurable Z and a $\mathcal{E}$-measurable f. Then, by the properties of the mathematical expectation,

$$\begin{aligned} E\left(\Psi\left(X, \cdot\right) | \mathcal{H}\right) &= E\left(f(X(\cdot))Z(\cdot) | \mathcal{H}\right) \\ &= f(X(\cdot))E(Z). \end{aligned}$$

Indeed, we use that X is $\mathcal{H}$-measurable and that $\mathcal{G}$, $\mathcal{H}$ are independent.
Clearly,

$$f(X(\cdot))E(Z) = f(x)E(Z)|_{x=X(\cdot)} = E\left(\Psi(x, \cdot)\right)|_{x=X(\cdot)} = \Phi(X).$$

This yields (6.18).
The result extends to any $\mathcal{E} \otimes \mathcal{G}$-measurable function Ψ by a monotone class argument.

□

Lemma 6.3 *Fix $u \geq 0$ and let η be a $\mathcal{F}_u$-measurable random variable in $L^2(\Omega)$. Consider the SDE*

$$Y_t^\eta = \eta + \int_u^t \sigma\left(s, Y_s^\eta\right) dB_s \int_u^t b\left(s, Y_s^\eta\right) ds,$$

with coeffients σ and b satisfying the assumptions (H). Then for any $t \geq u$,

$$Y_t^{\eta(\omega)}(\omega) = X_t^{x,u}(\omega)|_{x=\eta(\omega)},$$

where $X_t^{x,u}$, $t \geq 0$, denotes the solution to

$$X_t^{x,u} = x + \int_u^t \sigma\left(s, X_s^{x,u}\right) dB_s \int_u^t b\left(s, X_s^{x,u}\right) ds. \tag{6.19}$$

Proof: Suppose first that η is a step function,

$$\eta = \sum_{i=1}^r c_i \mathbf{1}_{A_i}, \; A_i \in \mathcal{F}_u.$$

By virtue of the local property of the stochastic integral, on the set A_i, $X_t^{c_i,u}(\omega) = X_t^{x,u}(\omega)|_{x=\eta(\omega)} = Y_t^{\eta}(\omega)$.
Let now $(\eta_n, n \geq 1)$ be a sequence of simple $\mathcal{F}_u$-measurable random variables converging in $L^2(\Omega)$ to η. By Theorem 6.3 we have

$$L^2(\Omega) - \lim_{n\to\infty} Y_t^{\eta_n} = Y_t^{\eta}.$$

By taking if necessary a subsequence, we may assume that the limit is a.s.. Then, a.s.,

$$Y_t^{\eta(\omega)}(\omega) = \lim_{n\to\infty} Y_t^{\eta_n(\omega)}(\omega) = \lim_{n\to\infty} X_t^{x,u}(\omega)|_{x=\eta_n(\omega)} = X_t^{x,u}(\omega)|_{x=\eta(\omega)},$$

where in the last equality, we have applied the joint continuity in (t,x) of $X_t^{x,u}$.

$\square$

As a consequence of the preceding lemma, we have $X_t^{x,s} = X_t^{X_u^{x,s},u}$ for any $0 \leq s \leq u \leq t$, a.s.
For any $\Gamma \in \mathcal{B}(\mathbb{R}^m)$, set

$$p(s,t,x,\Gamma) = P\{X_t^{x,s} \in \Gamma\}, \tag{6.20}$$

so, for fixed $0 \leq s \leq t$ and $x \in \mathbb{R}^m$, $p(s,t,x,\cdot)$ is the law of the random variable $X_t^{x,s}$.

Theorem 6.5 *The stochastic process $\{X_t^{x,s}, t \geq s\}$ is a Markov process with initial distribution $\mu = \delta_{\{x\}}$ and transition probability function given by (6.20).*

Proof: According to Definition 2.3 we have to check that (6.20) defines a Markovian transition function and that

$$P\{X_t^{x,s} \in \Gamma | \mathcal{F}_u\} = p(u,t,X_u^{x,s},\Gamma). \tag{6.21}$$

We start by proving this identity. For this, we shall apply Lemma 6.3 in the following setting:

$$\begin{aligned}
&(E,\mathcal{E}) = (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)),\\
&\mathcal{G} = \sigma\left(B_{r+u} - B_u, r \geq 0\right), \ \mathcal{H} = \mathcal{F}_u,\\
&\Psi(x,\omega) = 1_\Gamma\left(X_t^{x,u}(\omega)\right), \ u \leq t,\\
&X := X_u^{x,s}, \ s \leq u.
\end{aligned}$$

The property of independent increments of the Brownian motion clearly yields that the σ-fields $\mathcal{G}$ and $\mathcal{H}$ defined above are independent and $X_u^{x,s}$ is $\mathcal{F}_u$-measurable. Moreover,

$$\begin{aligned}
\Phi(x) &= E\left(\Psi(x,\cdot)\right) = E\left(1_\Gamma\left(X_t^{x,u}\right)\right)\\
&= P\{X_t^{x,u} \in \Gamma\} = p(u,t,x,\Gamma).
\end{aligned}$$

Thus, Lemma 6.3 and then Lemma 6.2 yield

$$\begin{aligned}
P\left\{X_t^{x,s} \in \Gamma | \mathcal{F}_u\right\} &= P\left\{X_t^{X_u^{x,s},u} \in \Gamma | \mathcal{F}_u\right\}\\
&= E\left(1_\Gamma\left(X_t^{X_u^{x,s},u}\right) | \mathcal{F}_u\right) = E\left(\Psi\left(X_u^{x,s},\omega\right) | \mathcal{F}_u\right)\\
&= \Phi\left(X_u^{x,s}\right) = p(u,t,X_u^{x,s},\Gamma).
\end{aligned}$$

Since $x \mapsto X_t^{x,s}$ is continuous a.s., the mapping $x \mapsto p(s,t,\cdot,\Gamma)$ is also continuous and thus measurable. Moreover, by its very definition $\Gamma \mapsto p(s,t,x,\Gamma)$ is a probability. We now prove that Chapman-Kolmogorov's equation is satisfied (see (2.10)).
Indeed, fix $0 \leq s \leq u \leq t$; by property (c) of the conditional expectation we have

$$\begin{aligned}
p(s,t,x,\Gamma) &= E\left(1_\Gamma(X_t^{x,s})\right)\\
&E\left(E\left(1_\Gamma(X_t^{x,s}) | \mathcal{F}_u\right)\right) = E\left(P\left\{X_t^{x,s} \in \Gamma\right\} | \mathcal{F}_u\right).
\end{aligned}$$

By (6.21) this last expression is $E\left(p(u,t,X_u^{x,s},\Gamma)\right)$. But

$$E\left(p(u,t,X_u^{x,s},\Gamma)\right) = \int_{\mathbb{R}^m} p(u,t,y,\Gamma)\mathcal{L}_{X_u^{x,s}}(dy),$$

where $\mathcal{L}_{X_u^{x,s}}$ denotes the probability law of $X_u^{x,s}$. By definition

$$\mathcal{L}_{X_u^{x,s}}(dy) = p(s,u,x,dy).$$

Therefore,

$$p(s,t,x,\Gamma) = \int_{\mathbb{R}^m} p(u,t,y,\Gamma)p(s,u,x,dy).$$

The proof of the theorem is now complete.

□

7 Numerical approximations of stochastic differential equations

In this section we consider a fixed time interval $[0,T]$. Let $\pi = \{0 = \tau_0 < \tau_1 < \ldots < \tau_N = T\}$ be a partition of $[0,T]$. *The Euler-Maruyama scheme* for the SDE (6.2) based on the partition π is the stochastic process $X^\pi = \{X_t^\pi, t \in [0,T]\}$ defined iteratively as follows:

$$\begin{aligned} X_{\tau_{n+1}}^\pi &= X_{\tau_n}^\pi + \sigma(\tau_n, X_{\tau_n}^\pi)(B_{\tau_{n+1}} - B_{\tau_n}) + b(\tau_n, X_{\tau_n}^\pi)(\tau_{n+1} - \tau_n), \\ &\quad n = 0, \cdots, N-1 \\ X_0^\pi &= x, \end{aligned} \tag{7.1}$$

Notice that the values $X_{\tau_n}^\pi$, $n = 0, \cdots, N-1$ are determined by the values of B_{τ_n}, $n = 1, \ldots, N$.
We can extend the definition of X^π to any value of $t \in [0,T]$ by setting

$$X_t^\pi = X_{\tau_j}^\pi + \sigma(\tau_j, X_{\tau_j}^\pi)(B_t - B_{\tau_j}) + b(\tau_j, X_{\tau_j}^\pi)(t - \tau_j), \tag{7.2}$$

for $t \in [\tau_j, \tau_{j+1})$.
The stochastic process $\{X_t^\pi, t \in [0,T]\}$ defined by (7.2) can be written as a stochastic differential equation. This notation will be suitable for comparing with the solution of (6.2). Indeed, for any $t \in [0,T]$ set $\pi(t) = \sup\{\tau_l \in \pi; \tau_l \leq t\}$; then

$$X_t^\pi = x + \int_0^t \left[\sigma\left(\pi(s), X_{\pi(s)}^\pi\right) dB_s + b\left(\pi(s), X_{\pi(s)}^\pi\right) ds\right]. \tag{7.3}$$

The next theorem gives the rate of convergence of the Eurler-Maruyama scheme to the solution of (6.2) in the L^p norm.

Theorem 7.1 *We assume that the hypotheses (H) are satisfied. Moreover, we suppose that there exists $\alpha \in (0,1)$ such that*

$$|\sigma(t,x) - \sigma(s,x)| + |b(t,x) - b(s,x)| \leq C(1+|x|)|t-s|^{\alpha}, \tag{7.4}$$

where C is some positive contant.
Then, for any $p \in [1,\infty)$

$$E\left(\sup_{0\leq t\leq T} |X_t^{\pi} - X_t|^p\right) \leq C(T,p,x)|\pi|^{\beta p}, \tag{7.5}$$

where $|\pi|$ denotes the norm of the partition π and $\beta = \frac{1}{2} \wedge \alpha$.

Proof: We shall apply the following result, that can be argued in a similar way as in Theorem 6.2

$$\sup_{\pi} E\left(\sup_{0\leq s\leq t} |X_s^{\pi}|^p\right) \leq C(p,T). \tag{7.6}$$

Set

$$Z_t = \sup_{0\leq s\leq t} |X_s^{\pi} - X_s|.$$

Applying Burkholder's and Hölder's inequality we obtain

$$\begin{aligned} E(Z_t^p) \leq 2^{p-1}\Big\{ & t^{\frac{p}{2}-1} \int_0^t E\left(\left|\sigma(\pi(s), X_{\pi(s)}^{\pi}) - \sigma(s, X_s)\right|^p\right) ds \\ & + t^{p-1} \int_0^t E\left(\left|b(\pi(s), X_{\pi(s)}^{\pi}) - b(s, X_s)\right|^p\right) ds \Big\}. \end{aligned}$$

The assumptions on the coefficients yield

$$\begin{aligned} \left|\sigma(\pi(s), X_{\pi(s)}^{\pi}) - \sigma(s, X_s)\right| &\leq \left|\sigma(\pi(s), X_{\pi(s)}^{\pi}) - \sigma(\pi(s), X_{\pi(s)})\right| \\ &\quad + \left|\sigma(\pi(s), X_{\pi(s)}) - \sigma(s, X_{\pi(s)})\right| + \left|\sigma(s, X_{\pi(s)}) - \sigma(s, X_s)\right| \\ &\leq C_T \left[\left|X_{\pi(s)}^{\pi} - X_{\pi(s)}\right| + \left(1 + \left|X_{\pi(s)}\right|\right)|s - \pi(s)|^{\alpha} + \left|X_{\pi(s)} - X_s\right|\right] \\ &\leq C_T \left[Z_s + \left(1 + \left|X_{\pi(s)}\right|\right)\left((s-\pi(s))^{\alpha} + \left|X_{\pi(s)} - X_s\right|\right)\right], \end{aligned}$$

and similarly for the coefficient b.
Hence, we have

$$\begin{aligned} & E\left(\left|\sigma(\pi(s), X_{\pi(s)}^{\pi}) - \sigma(s, X_s)\right|^p\right) \\ & \quad \leq C(p,T)\left[E(|Z_s|^p) + (1+|x|^p)\left(|\pi|^{\alpha p} + |\pi|^{\frac{p}{2}}\right)\right] \\ & \quad \leq C(p,T)\left[E(|Z_s|^p) + (1+|x|^p)\left(|\pi|^{\alpha} + |\pi|^{\frac{1}{2}}\right)^p\right], \end{aligned}$$

and a similar estimate for b. Consequently,

$$E(Z_t^p) \leq C(p,T,x)\left[\int_0^t E(Z_s^p)ds + \left(|\pi|^\alpha + |\pi|^{\frac{1}{2}}\right)^p T\right].$$

With Gronwall's lemma we conclude

$$E(Z_t^p) \leq C(p,T,x)|\pi|^{\beta p},$$

with $\beta = \frac{1}{2} \wedge \alpha$.

$\square$

Remark 7.1 *If the coefficients σ and b do not depend on t, with a similar proof, we obtain $\beta = \frac{1}{2}$ in (7.4).*

Assume that the sequence of partitions of $[0,T]$, $(\pi_n, n \geq 1)$ satisfies the following property: there exists $\gamma \in (0,\beta)$ and $p \geq 1$ such that

$$\sum_{n\geq 1} |\pi_n|^{(\beta-\gamma)p} < \infty. \tag{7.7}$$

Then, Chebyshev's inequality and (7.4) imply

$$\sum_{n=0}^{\infty} P\left\{|\pi_n|^{-\gamma} \sup_{0\leq s\leq T} |X_s^{\pi_n} - X_s| > \epsilon\right\} \leq C(p,T,x)\epsilon^{-p} \sum_{n=0}^{\infty} |\pi_n|^{(\beta-\gamma)p}$$
$$C(p,T,x)\epsilon^{-p}.$$

Then, Borel-Cantelli's lemma then yields

$$|\pi_n|^{-\gamma} \sup_{0\leq s\leq T} |X_s^{\pi_n} - X_s| \to 0, \ a.s., \tag{7.8}$$

as $n \to \infty$, uniformly in $t \in [0,T]$.
For example, for the sequence of dyadic partitions, $|\pi_n| = 2^{-n}$ and for any $\gamma \in (0,\beta)$ and $\gamma \in (0,\beta)$, $p \geq 1$, (7.7) holds.

8 Continuous time martingales

In this chapter we shall study some properties of martingales (respectively, supermartingales and submartingales) whose sample paths are continuous. We consider a filtration $\{\mathcal{F}_t, t \geq 0\}$ as has been introduced in section 2.4 and refer to definition 2.2 for the notion of martingale (respectively, supermartingale, submartingale). We notice that in fact this definition can be extended to families of random variables $\{X_t, t \in \mathbb{T}\}$ where $\mathbb{T}$ is an ordered set. In particular, we can consider discrete time parameter processes.
We start by listing some elementary but useful properties.

1. For a martingale (respectively, supermartingale, submartingale) the function $t \mapsto E(X_t)$ is a constant (respectively, decreasing, increasing) function.

2. Let $\{X_t, t \geq 0\}$ be a martingale and let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Assume further that $f(X_t) \in L^1(\Omega)$, for any $t \geq 0$. Then the stochastic process $\{f(X_t), t \geq 0\}$ is a submartingale. The same conclusion holds true for a submartingale if, additionally the convex function f is increasing.

The first assertion follows easily from property (c) of the conditional expectation. The second assertion can be proved using Jensen's inequality, as follows. Assume first that $\{X_t, t \geq 0\}$ is a martingale, and fix $0 \leq s \leq t$. Then by applying the functionn f to the identity $E(X_t|\mathcal{F}_s) = X_s$ along with the convexity of f, we obtain

$$E\left(f(X_t)|\mathcal{F}_s\right) \geq f\left(E(X_t|\mathcal{F}_s)\right) = f(X_s).$$

If $\{X_t, t \geq 0\}$ is a submartingale, we consider the inequality $E(X_t|\mathcal{F}_s) \geq X_s$. Since f is increasing and convex, we have

$$E\left(f(X_t)|\mathcal{F}_s\right) \geq f\left(E(X_t|\mathcal{F}_s)\right) \geq f(X_s).$$

8.1 Doob's inequalities for martingales

In the first part of this section, we will deal with discrete parameter martingales indexed by $\{0, 1, \ldots, N\}$.

Proposition 8.1 *Let $\{X_n, 0 \leq n \leq N\}$ be a submartingale. For any $\lambda > 0$, the following inequalities hold:*

$$\begin{aligned}\lambda P\left(\sup_n X_n \geq \lambda\right) &\leq E\left(X_N 1_{(\sup_n X_n \geq \lambda)}\right) \\ &\leq E\left(|X_N| 1_{(\sup_n X_n \geq \lambda)}\right). \qquad (8.1)\end{aligned}$$

Proof: Consider the stopping time

$$T = \inf\{n : X_n \geq \lambda\} \wedge N.$$

Then

$$\begin{aligned}
E(X_N) \geq E(X_T) &= E\left(X_T 1_{(\sup_n X_n \geq \lambda)}\right) \\
&\quad + E\left(X_T 1_{(\sup_n X_n < \lambda)}\right) \\
&\geq \lambda P\left(\sup_n X_n \geq \lambda\right) + E\left(X_N 1_{(\sup_n X_n < \lambda)}\right).
\end{aligned}$$

By substracting $E\left(X_N 1_{(\sup_n X_n < \lambda)}\right)$ from the first and last term before we obtain the first inequality of (8.1). The second one is obvious.

□

As a consequence of this proposition we have the following.

Proposition 8.2 *Let $\{X_n, 0 \leq n \leq N\}$ be either a martingale or a positive submartingale. Fix $p \in [1, \infty)$ and $\lambda \in (0, \infty)$. Then,*

$$\lambda^p P\left(\sup_n |X_n| \geq \lambda\right) \leq E\left(|X_N|^p\right). \tag{8.2}$$

Moreover, for any $p \in]1, \infty)$,

$$E\left(|X_N|^p\right) \leq E\left(\sup_n |X_n|^p\right) \leq \left(\frac{p}{p-1}\right)^p E\left(|X_N|^p\right). \tag{8.3}$$

Proof: Without loss of generality, we may assume that $E|X_N|^p) < \infty$, since otherwise (8.2) holds trivially.
According to property 2 above, the process $\{|X_n|^p, 0 \leq n \leq N\}$ is a submartingale and then, by Proposition 8.1 applied to the process $X_n := |X_n|^p$,

$$\begin{aligned}
\mu P\left(\sup_n |X_n|^p \geq \mu\right) &= \mu P\left(\sup_n |X_n| \geq \mu^{\frac{1}{p}}\right) \\
&= \lambda^p P\left(\sup_n |X_n| \geq \lambda\right) \leq E\left(|X_N|^p\right),
\end{aligned}$$

where for any $\mu > 0$ we have written $\lambda = \mu^{\frac{1}{p}}$.
We now prove the second inequality of (8.3); the first is obvious.
Set $X^* = \sup_n |X_n|$, for which we have

$$\lambda P\left(X^* \geq \lambda\right) \leq E\left(|X_N| 1_{(X^* \geq \lambda)}\right).$$

Fix $k > 0$. Fubini's theorem yields

$$\begin{aligned}
E\left((X^* \wedge k)^p\right) &= E\left(\int_0^{X^*\wedge k} p\lambda^{p-1} d\lambda\right) \\
&= \int_\Omega dP \int_0^\infty 1_{\{\lambda \le X^* \wedge k\}} p\lambda^{p-1} d\lambda \\
&= p\int_0^k d\lambda \lambda^{p-1} \int_{\{\lambda \le X^*\}} dP \\
&= p\int_0^k d\lambda \lambda^{p-2} \lambda P\left(X^* \ge \lambda\right) \\
&\le p\int_0^k d\lambda \lambda^{p-2} E\left(|X_N| 1_{(X^* \ge \lambda)}\right) \\
&= pE\left(|X_N| \int_0^{k\wedge X^*} \lambda^{p-2} d\lambda\right) \\
&= \frac{p}{p-1} E\left(|X_N| \left(X^* \wedge k\right)^{p-1}\right).
\end{aligned}$$

Applying Hölder's inequality with exponents $\frac{p}{p-1}$ and p yields

$$E\left((X^* \wedge k)^p\right) \le \frac{p}{p-1} \left[E\left((X^* \wedge k)^p\right)\right]^{\frac{p-1}{p}} \left[E\left(|X_N|^p\right)\right]^{\frac{1}{p}}.$$

Consequently,

$$\left[E\left((X^* \wedge k)^p\right)\right]^{\frac{1}{p}} \le \frac{p}{p-1} \left[E\left(|X_N|^p\right)\right]^{\frac{1}{p}}.$$

Letting $k \to \infty$ and using monotone convergence, we end the proof.

□

It is not difficult to extend the above results to martingales (submartingales) with continuous sample paths. In fact, for a given $T > 0$ we define

$$\begin{aligned}
D &= \mathbb{Q} \cap [0, T], \\
D_n &= D \cap \left\{\frac{k}{2^n}, k \in \mathbb{Z}_+\right\},
\end{aligned}$$

where $\mathbb{Q}$ denotes the set of rational numbers.
We can now apply (8.2), (8.3) to the corresponding processes indexed by D_n.
By letting n to ∞ we obtain

$$\lambda^p P\left(\sup_{t\in D} |X_t| \ge \lambda\right) \le \sup_{t\in D} E\left(|X_t|^p\right), \ p \in [1, \infty)$$

and

$$E\left(\sup_{t\in D}|X_t|^p\right) \le \left(\frac{p}{p-1}\right)^p \sup_{t\in D} E\left(|X_t|^p\right),\ p \in]1,\infty).$$

By the continuity of the sample paths we can finally state the following result.

Theorem 8.1 *Let $\{X_t, t \in [0,T]\}$ be either a continuous martingale or a continuous positive submartingale. Then*

$$\lambda^p P\left(\sup_{t\in[0,T]}|X_t| \ge \lambda\right) \le \sup_{t\in[0,T]} E\left(|X_t|^p\right), p \in [1,\infty), \tag{8.4}$$

$$E\left(\sup_{t\in[0,T]}|X_t|^p\right) \le \left(\frac{p}{p-1}\right)^p \sup_{t\in[0,T]} E\left(|X_t|^p\right) \tag{8.5}$$

$$= \left(\frac{p}{p-1}\right)^p E(|X_T|^p), p \in]1,\infty). \tag{8.6}$$

Inequality (8.4) is termed Doob's maximal inequality, while (8.6) is called Doob's L^p inequality.

8.2 Quadratic variation of a continuous martingale

In this section we construct the quadratic variation of a continuous martingale.
We start by a very simple consequence of the martingale property.

Lemma 8.1 *Let $\{N_t, t \ge 0\}$ be a martingale with respect to a filtration $(\mathcal{F}_t)_{t\ge0}$. The following identity holds: for any $0 \le s \le t$*

$$E\left(N_t^2 - N_s^2\right) = E\left((N_t - N_s)^2\right), \tag{8.7}$$

$$E\left(N_t^2 - N_s^2|\mathcal{F}_s\right) = E\left((N_t - N_s)^2\mathcal{F}_s\right), \tag{8.8}$$

Proof: By developing the square,

$$E\left((N_t - N_s)^2\right) = E\left(N_t^2 + N_s^2 - 2N_sN_t\right).$$

From the martingale property,

$$E(N_sN_t) = E\left(E(N_sN_t|\mathcal{F}_s)\right) = E(N_s^2).$$

Hence, the (8.7) follows. The proof of (8.8) follows by similar arguments. □

The next statement gives an idea of the roughness of the sample paths of a continuous local martingale.

Proposition 8.3 *Let N be a continuous bounded martingale, null at zero and with sample paths of bounded variation, a.s. Then N is indistinguishable from the constant process* 0.

Proof: Fix $t > 0$ and consider a partition $0 = t_0 < t_1 < \cdots < t_p = t$ of $[0, t]$. Then

$$\begin{aligned}
E(N_t^2) &= \sum_{i=1}^{p} E| \left[N_{t_i}^2 - N_{t_{i-1}}^2\right] \\
&= \sum_{i=1}^{p} E\left[\left(N_{t_i} - N_{t_{i-1}}\right)^2\right] \\
&\le E\left[\left(\sup_i |N_{t_i} - N_{t_{i-1}}|\right) \sum_{i=1}^{p} |N_{t_i} - N_{t_{i-1}}|\right] \\
&\le CE\left[\left(\sup_i |N_{t_i} - N_{t_{i-1}}|\right)\right],
\end{aligned}$$

where the second identity above is a consequence of Lemma 8.1 before.
By considering a sequence of partitions whose mesh tends to zero, the preceding estimate yields $E(N_t^2) = 0$, by the continuity of the sample paths of N. This finishes the proof of the proposition.

□

Throughout this section we will consider a fixed $t > 0$ and an increasing sequence of partitions of $[0, t]$ whose mesh tends to zero. Points of the n-th partition will be generically denoted by t_k^n, $k = 0, 1, \ldots, p_n$. We will also consider a continuous martingale M and define

$$\begin{aligned}
\langle M\rangle_t^n &= \sum_{k=1}^{p_n} \left(M_{t_k^n} - M_{t_{k-1}^n}\right)^2, \\
(\Delta_k^n M)_t &= M_{t_k^n} - M_{t_{k-1}^n}.
\end{aligned}$$

Theorem 8.2 *Let M be a continuous bounded martingale. Then the sequence $(\langle M\rangle_t^n, n \ge 1)$, $t \in [0, T]$, converges uniformly in t, in probability, to a continuous, increasing process $\langle M\rangle = (\langle M\rangle_t, t \in [0, T])$, such that $\langle M\rangle_0 = 0$. That is, for any $\epsilon > 0$,*

$$P\left\{\sup_{t\in[0,T]} |\langle M\rangle_t^n - \langle M\rangle_t| > \epsilon\right\} \to 0, \tag{8.9}$$

as $n \to \infty$. The process $\langle M\rangle$ is unique satisfying the above conditions and that $M^2 - \langle M\rangle$ is a continuous martingale.

Proof: Uniqueness follows from Proposition 8.3. Indeed, assume there were two increasing processes $\langle M_i\rangle$, $i = 1, 2$ satisfying that $M^2 - \langle M_i\rangle$ is a continuous martingale. By taking the difference of these two processes we get that $\langle M_1\rangle - \langle M_2\rangle$ is of bounded variation and, at the same time a continuous martingale. Hence $\langle M_1\rangle$ and $\langle M_2\rangle$ are indistinguishable.

The next objective is to prove that $(\langle M\rangle_t^n, n \geq 1)$, $t \in [0, T]$ is, uniformly in t, a Cauchy sequence in probability.

Let $m > n$. We have the following:

$$\begin{aligned}
E\left(|\langle M\rangle_t^n - \langle M\rangle_t^m|^2\right) &= E\left(\left|\sum_{k=1}^{p_n}\left[(\Delta_k^n M)_t^2 - \sum_{j:t_j^m\in[t_{k-1}^n,t_k^n[}(\Delta_j^m M)_t^2\right]\right|^2\right)\\
&= 4E\left(\left|\sum_{k=1}^{p_n}\left[\sum_{j:t_j^m\in[t_{k-1}^n,t_k^n[}(\Delta_j^m M)_t\left(M_{t_j^m} - M_{t_{k-1}^n}\right)\right]\right|^2\right)\\
&= 4E\left(\sum_{k=1}^{p_n}\sum_{j:t_j^m\in[t_{k-1}^n,t_k^n[}(\Delta_j^m M)_t^2\left(M_{t_j^m} - M_{t_{k-1}^n}\right)^2\right)\\
&\leq 4E\left(\sup_k \sup_{j:t_j^m\in[t_{k-1}^n,t_k^n[}\left(M_{t_j^m} - M_{t_{k-1}^n}\right)^2\sum_{j=1}^{p_m}(\Delta_j^m M)_t^2\right)\\
&\leq 4\left[E\left(\sup_k \sup_{j:t_j^m\in[t_{k-1}^n,t_k^n[}\left(M_{t_j^m} - M_{t_{k-1}^n}\right)^4\right)E\left(\sum_{j=1}^{p_m}(\Delta_j^m M)_t^2\right)^2\right]^{\frac{1}{2}}.
\end{aligned}$$

Let us now consider the last expression. The first factor tends to zero as n and m tends to infinity, because M is continuous and bounded. The second factor is easily seen to be bounded uniformly in m. Thus we have proved

$$\lim_{n,m\to\infty} E\left(|\langle M\rangle_t^n - \langle M\rangle_t^m|^2\right) = 0, \tag{8.10}$$

for any $t \in [0, T]$.

One can easily check that for any $n \geq 1$, $\{M_t^2 - \langle M\rangle_t^n, t \in [0, T]\}$ is a continuous martingale. Indeed, let $0 \leq s \leq t$, then

$$E\left(M_t^2 - M_s^2|\mathcal{F}_s\right) = E\left((M_t - M_s)^2|\mathcal{F}_s\right) = E\left(\langle M\rangle_t^n - \langle M\rangle_s^n|\mathcal{F}_s\right).$$

Therefore $\{\langle M\rangle_t^n - \langle M\rangle_t^m, t \in [0, T]\}$ is a martingale for any $n, m \geq 1$. Hence, Doob's inequality yields

$$E\left(\sup_{t\in[0,T]}|\langle M\rangle_t^n - \langle M\rangle_t^m|^2\right) \leq 4E\left(|\langle M\rangle_T^n - \langle M\rangle_T^m|^2\right).$$

Consequently,

$$P\{\sup_{0\le t\le T}|\langle M\rangle_t^n-\langle M\rangle_t^m|>\epsilon\}\le\epsilon^{-2}E\left(\sup_{0\le t\le T}|\langle M\rangle_t^n-\langle M\rangle_t^m|^2\right)$$
$$\le 4\epsilon^{-2}E\left(|\langle M\rangle_T^n-\langle M\rangle_T^m|^2\right).$$

This last expression tends to zero as n,m tend to infinity. Since the convergence in probability is metrizable, we see that there exists a process $\langle M\rangle$ satisfying the required conditions.

□

Remark Assume that the martingale M in Theorem 8.2 is bounded in L^2. Then, by a stopping argument the boundedness assumption on M can be removed.

Example Consider the stochastic integral process $M=\left\{\int_0^t\varphi(s)dB_s, s\in[0,T]\right\}$, with $\varphi\in L^2_{a,T}$. It was proved in Section 3 that the process M is a continuous martingale with respect to the filtration generated by the Brownian motion B. By applying the extension of Theorem 8.2 to continuous martingales bounded in L^2 we can see that

$$\langle M\rangle_t=\int_0^t\varphi(s)^2ds.$$

Indeed, the right-hand side of this equality defines an increasing process vanishing at $t=0$. Moreover, by Itô's formula,

$$M_t^2-\int_0^t\varphi(s)^2ds=2\int_0^tM_s\varphi(s)dB_s,$$

and the right-hand side of this identity defines a continuous martingale.

Given two continuous martingales M and N we define the cross variation by

$$\langle M,N\rangle_t=\frac{1}{2}\left[\langle M+N\rangle_t-\langle M\rangle_t-\langle N\rangle_t\right], \tag{8.11}$$

$t\in[0,T]$.

From the properties of the quadratic variation (see Theorem 8.2) we see that the process $\{\langle M,N\rangle_t, t\in[0,T]\}$ is a process of bounded variation and it is the unique process (up to indistinguishability) such that $\{M_tN_t-\langle M,N\rangle_t, t\in[0,T]\}$ is a continuous martingale. It is also clear that

$$\lim_{n\to\infty}\sum_{k=1}^{p_n}\left(M_{t_k^n}-M_{t_{k-1}^n}\right)\left(N_{t_k^n}-N_{t_{k-1}^n}\right)=\langle M,N\rangle_t, \tag{8.12}$$

uniformly in $t \in [0,T]$ in probability.
This result together with Schwarz's inequality imply

$$|\langle M,N\rangle_t| \le \sqrt{\langle M\rangle_t \langle N\rangle_t}.$$

and more generally, by setting $\langle M,N\rangle_s^t = \langle M,N\rangle_t - \langle M,N\rangle_s$, for $0 \le s \le t \le T$, then

$$|\langle M,N\rangle_s^t| \le \sqrt{\langle M\rangle_s^t \langle N\rangle_s^t}.$$

This inequality (a sort of Cauchy-Schwarz's inequality) is a particular case of the result stated in the next proposition.

Proposition 8.4 *Let M, N be two continuous martingales and H, K be two measurable processes. Then for any $t \ge 0$,*

$$\left| \int_0^t H_s K_s \ d\langle M,N\rangle_s \right| \le \left(\int_0^t H_s^2 d\langle M\rangle_s \right)^{\frac{1}{2}} \left(\int_0^t K_s^2 d\langle N\rangle_s \right)^{\frac{1}{2}}. \tag{8.13}$$

Proof: Consider a finite partition of $[0,t]$, given by $\{t_0 = 0 < t_1 < \ldots < t_r = t\}$, and assume first that the stochastic processes H, K are step processes and described by this partition, as follows:

$$\begin{aligned} H &= H_0 1_{\{0\}} + H_1 1_{]0,t_1]} + \ldots H_r 1_{]t_{r-1},t_r]}, \\ K &= K_0 1_{\{0\}} + K_1 1_{]0,t_1]} + \ldots K_r 1_{]t_{r-1},t_r]}, \end{aligned}$$

with H_i, K_i, $i = 0, \ldots, r$ bounded measurable random variables. Then

$$\int_0^t H_s K_s \ d\langle M,N\rangle_s = \sum_{i=1}^r H_i K_i \langle M,N\rangle_{t_i}^{t_{i+1}}.$$

Thus,

$$\begin{aligned} \left| \int_0^t H_s K_s \ d\langle M,N\rangle_s \right| &\le \le \sum_{i=1}^r |H_i K_i| \left| \langle M,N\rangle_{t_i}^{t_{i+1}} \right| \\ &\le \sum_{i=1}^r |H_i K_i| \left(\langle M\rangle_{t_i}^{t_{i+1}} \right)^{\frac{1}{2}} \left(\langle N\rangle_{t_i}^{t_{i+1}} \right)^{\frac{1}{2}}. \end{aligned}$$

By applying Schwarz's inequality, the last expression is bounded by

$$\left(\int_0^t H_s^2 d\langle M\rangle_s \right)^{\frac{1}{2}} \left(\int_0^t H_s^2 d\langle N\rangle_s \right)^{\frac{1}{2}}.$$

The general case follows from an approximation argument.

□

The bounded variation process $\langle M, N\rangle$ gives rise to the total variation measure $d\|\langle M, N\rangle\|_s$ defined by

$$d\|\langle M, N\rangle\|_s = d(\langle M, N\rangle^+)(s) + d(\langle M, N\rangle^-)(s),$$

where $\langle M, N\rangle_s^+$, $\langle M, N\rangle_s^-$, denote the increasing functions such that

$$\langle M, N\rangle_s = \langle M, N\rangle_s^+ - \langle M, N\rangle_s^-.$$

It is worth noticing that (8.13) can be extended to the following inequality.

Kunita-Watanabe inequality

$$\int_0^t |H_s|\,|K_s|\ \ \|d\langle M, N\rangle_s\| \leq \left(\int_0^t H_s^2 d\langle M\rangle_s\right)^{\frac{1}{2}} \left(\int_0^t K_s^2 d\langle N\rangle_s\right)^{\frac{1}{2}}. \qquad (8.14)$$

More details can be found in [10].

9 Stochastic integrals with respect to continuous martingales

This chapter aims to give an outline of the main ideas of the extension of the Itô stochastic integral to integrators which are continuous martingales. We start by describing precisely the spaces involved in the construction of such a notion. Throughout the chapter, we consider a fixed probability space $(\Omega, \mathcal{F}, P)$ endowed with a filtration $(\mathcal{F}_t, t \geq 0)$.
We denote by $\mathbb{H}^2$ the space of continuous martingales M, indexed by $[0, T]$, with $M_0 = 0$ a.s. and bounded in $L^2(\Omega)$. That is,

$$\sup_{t\in[0,T]} E(|M_t|^2) < \infty.$$

This is a Hilbert space endowed with the inner product

$$(M, N)_{\mathbb{H}^2} = E[\langle M, N\rangle_T].$$

A stochastic process $(X_t, t \geq 0)$ is said to be progressively measurable if for any $t \geq 0$, the mapping $(s, \omega) \mapsto X_s(\omega)$ defined on $[0, t] \times \Omega$ is measurable with respect to the σ-field $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$.
For any $M \in \mathbb{H}^2$ we define $L^2(M)$ as the set of progressively measurable processes H such that

$$E\left[\int_0^\infty H_s^2 d\langle M\rangle_s\right] < \infty.$$

Notice that this in an L^2 space of measurable mappings defined on $\mathbb{R}_+ \times \Omega$ with respect to the measure $dPd\langle M\rangle$. Hence it is also a Hilbert space, the natural inner product being

$$(H, K)_{L^2(M)} = E\left[\int_0^\infty H_s K_s d\langle M\rangle_s\right].$$

The standard Brownian motion belongs to the space $\mathbb{H}^2$ and the space $L^2(M)$ will play the same role as $L^2_{a,T}$ in the Itô theory of stochastic integration with respect to Brownian motion.
Let $\mathcal{E}$ be the linear subspace of $L^2(M)$ consisting of processes of the form

$$H_s(\omega) = \sum_{i=0}^{p} H_i(\omega) 1_{]t_i,t_{i+1}]}(s), \tag{9.1}$$

where $0 = t_0 < t_1 < \ldots < t_{p+1}$, and for each i, H_i is a $\mathcal{F}_{t_i}$-measurable, bounded random variable.
Stochastic processes belonging to $\mathcal{E}$ are termed *elementary*. There are related with $L^2(M)$ as follows.

Proposition 9.1 *Fix $M \in \mathbb{H}^2$. The set $\mathcal{E}$ is dense in $L^2(M)$.*

Proof: We will prove that if $K \in L^2(M)$ and is orthogonal to $\mathcal{E}$ then $K = 0$. For this, we fix $0 \le s < t \le T$ and consider the process

$$H = F 1_{]s,t]},$$

with F a $\mathcal{F}_s$-measurable and bounded random variable.
Saying that K is orthogonal to H in $L^2(M)$ can be written as

$$E\left[\int_0^T H_u K_u d\langle M\rangle_u\right] = E\left[F\int_s^t K_u d\langle M\rangle_u\right] = 0.$$

Consider the stochastic process

$$X_t = \int_0^t K_u d\langle M\rangle_u.$$

Notice that $X_t \in L^1(\Omega)$. In fact,

$$\begin{aligned} E|X_t| &\le E\int_0^t |K_u| d\langle M\rangle_u \\ &\le \left(E\int_0^t |K_u|^2 d\langle M\rangle_u\right)^{\frac{1}{2}} \left(E\langle M\rangle_t^2\right)^{\frac{1}{2}}. \end{aligned}$$

We have thus proved that $E\left(F(X_t - X_s)\right) = 0$, for any $0 \le s < t$ and any $\mathcal{F}_s$-measurable and bounded random variable F. This shows that the process $(X_t, t \ge 0)$ is a martingale. At the same time, $(X_t, t \ge 0)$ is also a process of bounded variation. Hence

$$\int_0^t K_u d\langle M\rangle_u = 0, \ \forall t \ge 0,$$

which implies that $K = 0$ in $L^2(M)$.

$\square$

Stochastic integral of processes in $\mathcal{E}$

Proposition 9.2 *Let $M \in \mathbb{H}^2$, $H \in \mathcal{E}$ as in (9.1). Define*

$$(H.M)_t = \sum_{i=0}^p H_i \left(M_{t_{i+1}\wedge t} - M_{t_i \wedge t}\right).$$

Then

(i) $H.M \in \mathbb{H}^2$.

(ii) The mapping $H \mapsto H.M$ extends to an isometry from $L^2(M)$ to $\mathbb{H}^2$.

The stochastic process $\{(H.M)_t, t \geq 0\}$ is called the stochastic integral of the process H with respect to M and is also denoted by $\int_0^t H_s dM_s$.

Proof of (i): The martingale property follows from the measurability properties of H and the martingale property of M. Moreover, since H is bounded, $(H.M)$ is bounded in $L^2(\Omega)$.

Proof of (ii): We prove first that the mapping

$$H \in \mathcal{E} \mapsto H.M$$

is an isometry from $\mathcal{E}$ to $\mathbb{H}^2$.
Clearly $H \mapsto H.M$ is linear. Moreover, $H.M$ is a finite sum of terms like

$$M_t^i = H_i \left(M_{t_{i+1} \wedge t} - M_{t_i \wedge t} \right)$$

each one being a martingale and orthogonal to each other. It is easy to check that

$$\langle M^i \rangle_t = H_i^2 (\langle M \rangle_{t_{i+1} \wedge t} - \langle M \rangle_{t_i \wedge t}).$$

Hence,

$$\langle H.M \rangle_t = \sum_{i=0}^{p} H_i^2 (\langle M \rangle_{t_{i+1}} - \langle M \rangle_{t_i}).$$

Consequently,

$$\begin{aligned} E\langle H.M \rangle_T = \|H.M\|_{\mathbb{H}^2}^2 &= E\left[\sum_{i=0}^{p} H_i^2 (\langle M \rangle_{t_{i+1} \wedge T} - \langle M \rangle_{t_i \wedge T}) \right] \\ &= E\left[\int_0^T H_s^2 d\langle M \rangle_s \right] \\ &= \|H\|_{L^2(M)}. \end{aligned}$$

Since $\mathcal{E}$ is dense in $L^2(M)$ this isometry extends to a unique isometry from $L^2(M)$ into $\mathbb{H}^2$. The extension is termed the stochastic integral of the process H with respect to M.

□

10 Appendix 1: Conditional expectation

Roughly speaking, a conditional expectation of a random variable is the mean value with respect to a modified probability after having incorporated some *a priori* information. The simplest case corresponds to conditioning with respect to an event $B \in \mathcal{F}$. In this case, the conditional expectation is the mathematical expectation computed on the modified probability space $(\Omega, \mathcal{F}, P(\cdot/B))$.
However, in general, additional information cannot be described so easily. Assuming that we know about some events $B_1, \ldots, B_n$ we also know about those that can be derived from them, like unions, intersections, complementaries. This explains the election of a σ-field to keep known information and to deal with it.
In the sequel, we denote by $\mathcal{G}$ an arbitrary σ-field included in $\mathcal{F}$ and by X a random variable with finite expectation ($X \in L^1(\Omega)$). Our final aim is to give a definition of the conditional expectation of X given $\mathcal{G}$. However, in order to motivate this notion, we shall start with more simple situations.

Conditional expectation given an event

Let $B \in \mathcal{F}$ be such that $P(B) \neq 0$. The conditional expectation of X given B is the real number defined by the formula

$$E(X/B) = \frac{1}{P(B)} E(1_B X). \tag{10.1}$$

It immediately follows that

- $E(X/\Omega) = E(X)$,
- $E(1_A/B) = P(A/B)$.

With the definition (10.1), the conditional expectation coincides with the expectation with respect to the conditional probability $P(\cdot/B)$. We check this fact with a discrete random variable $X = \sum_{i=1}^{\infty} a_i 1_{A_i}$. Indeed,

$$\begin{aligned} E(X/B) &= \frac{1}{P(B)} E\left(\sum_{i=1}^{\infty} a_i 1_{A_i \cap B}\right) = \sum_{i=1}^{\infty} a_i \frac{P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} a_i P(A_i/B). \end{aligned}$$

Conditional expectation given a discrete random variable
Let $Y = \sum_{i=1}^{\infty} y_i 1_{A_i}$, $A_i = \{Y = y_i\}$. The conditional expectation of X given Y is the random variable defined by

$$E(X/Y) = \sum_{i=1}^{\infty} E(X/Y = y_i) 1_{A_i}. \tag{10.2}$$

Notice that, knowing Y means knowing all the events that can be described in terms of Y. Since Y is discrete, they can be described in terms of the basic events $\{Y = y_i\}$. This may explain the formula (10.2).
The following properties hold:

(a) $E\left(E(X/Y)\right) = E(X)$;

(b) if the random variables X and Y are independent, then $E(X/Y) = E(X)$.

For the proof of (a) we notice that, since $E(X/Y)$ is a discrete random variable

$$\begin{aligned} E\left(E(X/Y)\right) &= \sum_{i=1}^{\infty} E(X/Y = y_i) P(Y = y_i) \\ &= E\left(X \sum_{i=1}^{\infty} 1_{\{Y=y_i\}}\right) = E(X). \end{aligned}$$

Let us now prove (b). The independence of X and Y yields

$$\begin{aligned} E(X/Y) &= \sum_{i=1}^{\infty} \frac{E(X 1_{\{Y=y_i\}})}{P(Y = y_i)} 1_{A_i} \\ &= \sum_{i=1}^{\infty} E(X) 1_{A_i} = E(X). \end{aligned}$$

In the sequel, we shall denote by $\sigma(Y)$ the σ-field generated by a random variable Y. It consist of sets of the form $\{\omega : Y(\omega) \in B\}$, where B is a Borel set of $\mathbb{R}$.
The next proposition states two properties of the conditional expectation that motivates the Definition 10.1.

Proposition 10.1 *1. The random variable $Z := E(X/Y)$ is $\sigma(Y)$-measurable; that is, for any Borel set $B \in \mathcal{B}$, $Z^{-1}(B) \in \sigma(Y)$,*

2. for any $A \in \sigma(Y)$, $E\left(1_A E(X/Y)\right) = E(1_A X)$.

Proof: Set $c_i = E(X/\{Y = y_i\})$ and let $B \in \mathcal{B}$. Then

$$Z^{-1}(B) = \cup_{i:c_i \in B}\{\omega : Z(\omega) = c_i\} = \cup_{i:c_i \in B}\{\omega : Y(\omega) = y_i\} \in \sigma(Y),$$

proving the first property.
To prove the second one, it suffices to take $A = \{Y = y_k\}$. In this case

$$\begin{aligned} E\left(1_{\{Y=y_k\}}E(X/Y)\right) &= E\left(1_{\{Y=y_k\}}E(X/Y = y_k)\right) \\ &= E\left(1_{\{Y=y_k\}}\frac{E(X1_{\{Y=y_k\}})}{P(Y = y_k)}\right) = E(X1_{\{Y=y_k\}}). \end{aligned}$$

□

Conditional expectation given a σ-field

Definition 10.1 *The conditional expectation of X given $\mathcal{G}$ is a random variable Z satisfying the properties*

1. *Z is $\mathcal{G}$-measurable; that is, for any Borel set $B \in \mathcal{B}$, $Z^{-1}(B) \in \mathcal{G}$,*
2. *for any $G \in \mathcal{G}$,*
$$E(Z1_G) = E(X1_G).$$

We will denote the conditional expectation Z by $E(X/\mathcal{G})$.
Notice that the conditional expectation is not a number but a random variable. There is nothing strange in this, since conditioning depends on the observations.
Condition (1) tell us that events that can be described by means of $E(X/\mathcal{G})$ are in $\mathcal{G}$. Whereas condition (2) tell us that on events in $\mathcal{G}$ the random variables X and $E(X/\mathcal{G})$ have the same mean value.

The existence of $E(X/\mathcal{G})$ is not a trivial issue. You should trust mathematicians and believe that there is a theorem in measure theory -the Radon-Nikodym Theorem- which ensures the existence and uniqueness of such a random variable (out of a set of probability zero).
Before stating properties of the conditional expectation, we are going to explain how to compute it in two particular situations.

Example 10.1 *Let $\mathcal{G}$ be the σ-field (actually, the field) generated by a finite partition $G_1, \ldots, G_m$. Then*

$$E(X/\mathcal{G}) = \sum_{j=1}^{m} \frac{E(X1_{G_j})}{P(G_j)} 1_{G_j}. \tag{10.3}$$

Formula (10.3) can be checked using Definition 10.1. It tell us that, on each generator of $\mathcal{G}$, the conditional expectation is constant; this constant is weighted by the *mass* of the generator ($P(G_j)$).

Example 10.2 *Let* $\mathcal{G}$ *be the* σ*-field generated by random variables* $Y_1, \ldots, Y_m$*, that is, the* σ*-field generated by events of the form* $Y_1^{-1}(B_1), \ldots, Y_1^{-1}(B_m)$*, with* $B_1, \ldots, B_m$ *arbitrary Borel sets. Assume in addition that the joint distribution of the random vector* $(X, Y_1, \ldots, Y_m)$ *has a density* f*. Then*

$$E(X/Y_1, \ldots, Y_m) = \int_{-\infty}^{\infty} x f(x/Y_1, \ldots, Y_m) dx, \tag{10.4}$$

with

$$f(x/y_1, \ldots, y_m) = \frac{f(x, y_1, \ldots, y_m)}{\int_{-\infty}^{\infty} f(x, y_1, \ldots, y_m) dx}. \tag{10.5}$$

In (10.5), we recognize the conditional density of X given $Y_1 = y_1, \ldots, Y_m = y_m$. Hence, in (10.4) we first compute the conditional expectation $E(X/Y_1 = y_1, \ldots, Y_m = y_m)$ and finally, replace the real values $y_1, \ldots, y_m$ by the random variables $Y_1, \ldots, Y_m$.

We now list some important properties of the conditional expectation.

(a) Linearity: for any random variables X, Y and real numbers a, b

$$E(aX + bY/\mathcal{G}) = aE(X/\mathcal{G}) + bE(Y/\mathcal{G}).$$

(b) Monotony: If $X \leq Y$ then $E(X/\mathcal{G}) \leq E(Y/\mathcal{G})$.

(c) The mean value of a random variable is the same as that of its conditional expectation: $E(E(X/\mathcal{G})) = E(X)$.

(d) If X is a $\mathcal{G}$-measurable random variable, then $E(X/\mathcal{G}) = X$

(e) Let X be independent of $\mathcal{G}$, meaning that any set of the form $X^{-1}(B)$, $B \in \mathcal{B}$ is independent of $\mathcal{G}$. Then $E(X/\mathcal{G}) = E(X)$.

(f) Factorization: If Y is a bounded, $\mathcal{G}$-measurable random variable,

$$E(YX/\mathcal{G}) = YE(X/\mathcal{G}).$$

(g) If $\mathcal{G}_i$, $i = 1, 2$ are σ-fields with $\mathcal{G}_1 \subset \mathcal{G}_2$,

$$E(E(X/\mathcal{G}_1)/\mathcal{G}_2) = E(E(X/\mathcal{G}_2)/\mathcal{G}_1) = E(X/\mathcal{G}_1).$$

(h) Assume that X is a random variable independent of $\mathcal{G}$ and Z another $\mathcal{G}$-measurable random variable. For any measurable function $h(x,z)$ such that the random variable $h(X,Z)$ is in $L^1(\Omega)$,

$$E(h(X,Z)/\mathcal{G}) = E(h(X,z))|_{Z=z}.$$

We give some proofs.
Property (a) follows from the definition of the conditional expectation and the linearity of the operator E. Indeed, the candidate $aE(X/\mathcal{G})+bE(Y/\mathcal{G})$ is $\mathcal{G}$-measurable. By property 2 of the conditional expectation and the linearity of E,

$$\begin{aligned} E(1_G[aE(X/\mathcal{G})+bE(Y/\mathcal{G})]) &= aE(1_G X) + bE(1_G Y) \\ &= E(1_G[aX+bY]). \end{aligned}$$

Property (b) is a consequence of the monotonicity property of the operator E and a result in measure theory telling that, for $\mathcal{G}$-measurable random variables Z_1 and Z_2, satisfying

$$E(Z_1 1_G) \le E(Z_2 1_G),$$

for any $G \in \mathcal{G}$, we have $Z_1 \le Z_2$. Indeed, under the standing assumptions, for any $G \in \mathcal{G}$,

$$E(1_G X) \le E(1_G Y).$$

Then, property 2 of the conditional expectation yields

$$E(1_G E(X/\mathcal{G})) = E(1_G X) \le E(1_G Y) = E(1_G E(Y/\mathcal{G})).$$

By applying the above mentioned property to $Z_1 = E(X/\mathcal{G})$, $Z_2 = E(Y/\mathcal{G})$, we obtain the result.
Taking $G = \Omega$ in condition (2) above, we prove (c). Property (d) is obvious. Constant random variables are measurable with respect to any σ-field. Therfore $E(X)$ is $\mathcal{G}$-measurable. Assuming that X is independent of $\mathcal{G}$, yields

$$E(X1_G) = E(X)E(1_G) = E(E(X)1_G).$$

This proves (e).
For the proof of (f), we first consider the case $Y = 1_{\tilde{G}}$, $\tilde{G} \in \mathcal{G}$. Claiming (f) means that we propose as candidate $E(YX/\mathcal{G}) = 1_{\tilde{G}} E(X/\mathcal{G})$. Clearly $1_{\tilde{G}} E(X/\mathcal{G})$ is $\mathcal{G}$-measurable. Moreover

$$E(1_G 1_{\tilde{G}} E(X/\mathcal{G})) = E(1_{G\cap\tilde{G}} E(X/\mathcal{G})) = E(1_{G\cap\tilde{G}} X).$$

The validity of the property extends by linearity to simple random variables. Then, by monotone convergence, to positive random variables and, finally, to random variables in $L^1(\Omega)$, by the usual decomposition $X = X^+ - X^-$.
For the proof of (g), we notice that since $E(X/\mathcal{G}_1)$ is $\mathcal{G}_1$-measurable, it is $\mathcal{G}_2$-measurable as well. Then, by the very definition of the conditional expectation,

$$E(E(X/\mathcal{G}_1)/\mathcal{G}_2) = E(X/\mathcal{G}_1).$$

Next, we prove that

$$E(X/\mathcal{G}_1) = E(E(X/\mathcal{G}_2)/\mathcal{G}_1).$$

For this, we fix $G \in \mathcal{G}_1$ and we apply the definition of the conditional expectation. This yields

$$E(1_G E(E(X/\mathcal{G}_2)/\mathcal{G}_1)) = E(1_G E(X/\mathcal{G}_2)) = E(1_G X).$$

Property (h) is very intuitive: Since X is independent of $\mathcal{G}$ in does not enter the game of conditioning. Moreover, the measurability of Z means that by conditioning one can suppose it is a constant.
Let us give a proof of this property in the particular case $h(x, z) = 1_A(x)1_B(z)$, where A, B are Borel sets. In this case we have

$$\begin{aligned} E(h(X,Z)/\mathcal{G}) &= E(1_A(X)1_B(Z)/\mathcal{G}) = 1_B(Z)E(1_A(X)/\mathcal{G}) \\ &= 1_B(Z)E(1_A(X)). \end{aligned}$$

Moreover,

$$E(h(X,z)) = E(1_A(X)1_B(z)) = 1_B(z)E(1_A(X)).$$

Therefore

$$E(h(X,z))\Big|_{z=Z} = 1_B(Z)E(1_A(X)).$$

.

11 Appendix 2: Stopping times

Throughout this section we consider a fixed filtration (see Section 2.4) $(\mathcal{F}_t, t \geq 0)$.

Definition 11.1 *A mapping $T : \Omega \to [0, \infty]$ is termed a stopping time with respect to the filtration $(\mathcal{F}_t, t \geq 0)$ is for any $t \geq 0$*

$$\{T \leq t\} \in \mathcal{F}_t.$$

It is easy to see that if S and T are stopping times with respect to the same filtration then $T \wedge S$, $T \vee S$ and $T + S$ are also stopping times.

Definition 11.2 *For a given stopping time T, the σ-field of events prior to T is the following*

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \le t\} \in \mathcal{F}_t, \textit{for all } t \ge 0\}.$$

Let us prove that $\mathcal{F}_T$ is actually a σ-field. By the definition of stopping time $\Omega \in \mathcal{F}_T$. Asuume that $A \in \mathcal{F}_T$. Then

$$A^c \cap \{T \le t\} = \{T \le t\} \cap (A \cap \{T \le t\})^c \in \mathcal{F}_t.$$

Hence with any A, $\mathcal{F}_T$ also contains A^c.
Let now $(A_n, n \ge 1) \subset \mathcal{F}_T$. We clearly have

$$(\cup_{n=1}^{\infty} A_n) \cap \{T \le t\} = \cup_{n=1}^{\infty} (A_n \cap \{T \le t\}) \in \mathcal{F}_t.$$

This completes the proof.

Some properties related with stopping times

1. Any stopping time T is $\mathcal{F}_T$-measurable. Indeed, let $s \ge 0$, then

$$\{T \le s\} \cap \{T \le t\} = \{T \le s \wedge t\} \in \mathcal{F}_{s \wedge t} \subset \mathcal{F}_t.$$

2. If $\{X_t, t \ge 0\}$ is a process with continuous sample paths, a.s. and $(\mathcal{F}_t)$-adapted, then X_T is $\mathcal{F}_T$-measurable. Indeed, the continuity implies

$$X_T = \lim_{n \to \infty} \sum_{i=0}^{\infty} X_{i2^{-n}} 1_{\{i2^{-n} < T \le (i+1)2^{-n}\}}.$$

   Let us now check that for any $s \ge 0$, the random variable $X_s 1_{\{s<T\}}$ is $\mathcal{F}_T$-measurable. This fact along with the property $\{T \le (i+1)2^{-n}\} \in \mathcal{F}_T$ shows the result.

   Let $A \in \mathcal{B}(\mathbb{R})$ and $t \ge 0$. The set

$$\{X_s \in A\} \cap \{s < T\} \cap \{T \le t\}$$

   is empty if $s \ge t$. Otherwise it is equal to $\{X_s \in A\} \cap \{s < T \le t\}$, which belongs to $\mathcal{F}_t$.

References

[1] P. Baldi: Equazioni differenziali stocastiche e applicazioni. Quaderni dell'Unione Matematica Italiana 28. Pitagora Editrici. Bologna 2000.

[2] K.L. Chung, R.J. Williams: Introduction to Stochastic Integration, 2nd Edition. Probability and Its Applications. Birkhäuser, 1990.

[3] R. Durrett: Stochastic Calculus, a practical introduction. CRC Press, 1996.

[4] L.C. Evans: An Introduction to Stochastic Differential Equations. http://math.berkeley.edu/ evans/SDE.course.pdf

[5] I. Karatzas, S. Shreve: Brownian Motion and Stochastic Calculus. Springer Verlag, 1991.

[6] H.H. Kuo: Introduction to Stochastic Integration. Universitext. Springer 2006.

[7] G. Lawler: Introduction to Stochastic Processes. Chapman and Hall/CRC, 2nd Edition, 2006.

[8] J.-F. Le Gall: Mouvement Brownien et calcul stochastiques. Notes de Cours de DEA 1996-1997. http://www.dma.ens.fr/ legall/

[9] B. Øksendal: Stochastic Differential Equations; an Introduction with Applications. Springer Verlag, 1998.

[10] D. Revuz, M. Yor: Continuous martingales and Brownian motion. Springer Verlag, 1999.

[11] P. Todorovic: An Introduction to Stochastic Processes and Their Applications. Springer Series in Statistics. Springer Verlag, 1992,

[12] R. L. Schilling, L. Partzsch: Bronian Motion: an introduction to stochastic processes. 2nd Edition. De Gruyter, 2014.