

An Introductory Course on Stochastic Calculus

Xi Geng

Abstract

These are lecture notes for the stochastic calculus course at The University of Melbourne in Semester 1, 2021. They are designed at an introductory level of the subject. We aim at presenting the materials in a motivated way and developing the essential techniques at a minimal level of technicalities while maintaining mathematical precision as much as possible. Only very mild knowledge on probability measures and integration is needed, and all relevant tools are recalled in the appendix.

We follow the classical route of first introducing martingale notions and then using them to study Brownian motion and its stochastic calculus. As a result, the entire development has a very strong flavour of martingale methods. To keep the materials ideal for a one-semester course at an elementary level, we have omitted the discussion of several advanced but significant topics such as local times, stochastic calculus for semi-martingales, the Yamada-Watanabe theorem, martingale problems etc. One who is interested in some of these topics and wishes to dive deeper into the subject should consult the beautiful monographs [8, 9, 16, 20].

On the other hand, several approaches in these notes deviate from traditional texts. For instance, in the study of martingales, we adopt the elegant approach of D. Williams to prove most of the basic martingale theorems through gambling. For the construction of Brownian motion, we follow the original approach of N. Wiener based on Fourier series. We use the idea of Skorokhod's embedding to motivate a quick proof of Donsker's invariance principle which in turn recovers the classical central limit theorem as a byproduct. In the construction of stochastic integrals, our approach is largely inspired by H. McKean in which the integrals are constructed in one go via simple process approximations without the need of introducing any localisation argument (local martingales). To keep things elementary, in the construction of stochastic integrals we have to reluctantly give up the rather elegant Hilbert space approach of D. Revuz and M. Yor from the duality perspective. Any serious student should teach him/herself this

neat and deeper approach. For the Cameron-Martin-Girsanov's theorem, we reproduce the original calculation of R. Cameron and W. Martin which enables one to see how the particular shape of the transformation formula arises naturally. In the study of stochastic differential equations, we emphasise probabilistic properties of solutions as well as their applications rather than delving into the abstract theory of existence and uniqueness.

Last but not the least, the best part of these notes is in no doubt the list of exercises!

Contents

1	Introduction and preliminary notions	5
1.1	Motivation	5
1.1.1	A toy stock pricing model	5
1.1.2	Heat transfer and temperature distributions	8
1.2	Filtrations, stopping times and stochastic processes	10
1.3	The conditional expectation	21
1.4	Martingales	24
1.4.1	The martingale transform: discrete stochastic integration	26
1.4.2	The martingale convergence theorem	27
1.4.3	The optional sampling theorem	31
1.4.4	The maximal and L^p -inequalities	33
1.4.5	The continuous-time case	35
2	Brownian motion	37
2.1	The construction of Brownian motion	37
2.1.1	Some notions on Fourier series	38
2.1.2	Wiener's original idea	39
2.1.3	The mathematical construction	41
2.2	The strong Markov property and the reflection principle	45
2.3	Passage time distributions	51
2.4	Skorokhod's embedding theorem and Donsker's invariance principle	55
2.4.1	The heuristic proof of Donsker's invariance principle	56
2.4.2	Constructing solutions to Skorokhod's embedding problem	58
2.5	Erraticity of Brownian sample paths: an overview	65
2.5.1	Irregularity	65
2.5.2	Oscillations	66
2.5.3	The p -variation of Brownian motion	67
3	Stochastic integration	68
3.1	The essential structure: Itô's isometry	68
3.2	General construction of stochastic integrals	73
3.2.1	Step one: integration of simple processes	74
3.2.2	Step two: an exponential martingale property	75
3.2.3	Step three: a uniform estimate for integrals of simple processes	76
3.2.4	Step four: approximation by simple processes	78
3.2.5	Step five: completing the construction	80
3.3	Martingale properties	85

3.3.1	Square integrable martingales and the bracket process	85
3.3.2	Stochastic integrals as martingales	88
3.4	Itô's formula	90
3.5	Lévy's characterisation of Brownian motion	96
3.6	A martingale representation theorem	100
3.7	The Cameron-Martin-Girsanov transformation	107
3.7.1	Motivation: the original approach of Cameron and Martin	107
3.7.2	Girsanov's approach	111
4	Stochastic differential equations	118
4.1	Itô's theorem of existence and uniqueness	119
4.2	Strong Markov property, generator and the martingale formulation	127
4.3	Explicit solutions to linear SDEs	131
4.4	One-dimensional diffusions	135
4.4.1	Exit distribution and behaviour towards explosion	136
4.4.2	Bessel processes	141
4.5	Connection with partial differential equations	143
4.5.1	Dirichlet boundary value problems	144
4.5.2	Parabolic problems: the Feynman-Kac formula	146
5	Applications in risk-neutral pricing	150
5.1	Basic concepts	150
5.1.1	Stock price model and interest rate process	150
5.1.2	Portfolio process and value of portfolio	151
5.1.3	Risk-neutral measure and hedging	152
5.2	The Black-Scholes-Merton formula	155
5.2.1	Construction of the risk-neutral measure	155
5.2.2	Completeness of market model and construction of hedging strategies	156
5.2.3	Derivation of the Black-Scholes-Merton formula	158
5.3	The second fundamental theorem of asset pricing	160
6	List of exercises	163
A	Terminology from probability theory	198

1 Introduction and preliminary notions

In this chapter, we discuss some aspects of the motivation and introduce a few preliminary notions. The core concept is the mathematical way of describing the accumulation of information that evolves in time: *filtrations* and *stopping times*. This is an essential feature of stochastic calculus that is quite different from ordinary calculus.

1.1 Motivation

Stochastic calculus, in a restricted sense, is the theory of differential calculus for Brownian motion. The fundamental ideas were essentially due to K. Itô in the 1940s. Since then, the theory has been under vast development by many probabilists including several generations of Itô's students. Apart from its theoretical importance, the subject has also stimulated a wide range of applications in various areas. Among others, the most well-known application is the pricing of financial derivatives. We use two examples to motivate some of the fundamental ideas in stochastic calculus.

1.1.1 A toy stock pricing model

Consider a particular stock A. Let X_t ($t \geq 0$) denote its price at time t . The determination of X_t is subject to many external random factors. As a result, for each $t \geq 0$ the quantity X_t should be viewed as a random variable. What is a natural model describing the dynamics of X_t as a (random) function of t ?

We first look at a simplified situation where time is discrete, say $t = 0, 1, 2, \dots$. In this case, the price dynamics becomes a discrete sequence of random variables $\{X_n : n = 0, 1, 2, \dots\}$. As a toy model, we postulate that the price change $X_{n+1} - X_n$ in the next day relative to current price X_n is governed by two factors: a deterministic trend and a random perturbation. Mathematically, this is formulated as

$$\frac{X_{n+1} - X_n}{X_n} = \mu + \sigma \cdot \xi_{n+1}, \quad n = 0, 1, 2, \dots \quad (1.1)$$

Here μ is a deterministic real number representing the intrinsic trend of relative price change. If $\mu > 0$, the stock price tends to be increasing on average and otherwise if $\mu < 0$. The quantity $\sigma \cdot \xi_{n+1}$ is a random variable representing an external random force acting on the price change. Let us assume that $\{\xi_n : n \geq 1\}$ is an independent and identically distributed sequence with the two-point distribution

$$\mathbb{P}(\xi_n = 1) = \mathbb{P}(\xi_n = -1) = \frac{1}{2}.$$

The factor $\sigma > 0$ is a deterministic number so that $\sigma \cdot \xi_{n+1}$ quantifies the magnitude of the random kick on the relative price change. The larger σ is, the variance of the random perturbation gets larger, making the stock price more uncertain (risky).

To make the model more realistic, we shall treat time continuously. In this case, the discrete model (1.1) needs to be turned into an infinitesimal description:

$$\frac{X_{t+\delta t} - X_t}{X_t} = \mu \cdot \delta t + \sigma \cdot \delta B_t, \quad (1.2)$$

where $\{B_t : t \geq 0\}$ is a suitable stochastic process such that $\delta B_t = B_{t+\delta t} - B_t$ resembles an infinitesimal analogue of the random perturbation ξ_{n+1} . To motivate the shape of B_t , we first define the partial sum sequence

$$S_n \triangleq \xi_1 + \cdots + \xi_n, \quad n \geq 1.$$

Geometrically, $\{S_n : n \geq 1\}$ is a simple random walk on the set of integers. Note that $\xi_{n+1} = S_{n+1} - S_n$ is the “discrete differential” of S_n . It is now natural to think of the process $\{B_t\}$ as a “continuum limit” of the random walk $\{S_n\}$ normalised in a suitable way.

To understand this limiting procedure, let us restrict ourselves to the time horizon $[0, 1]$. Given $m \geq 1$, we divide $[0, 1]$ into m equal sub-intervals each with length $1/m$. One can naively construct an “approximating” continuous process $\{\tilde{B}_t^{(m)}\}$ on $[0, 1]$ by linearly interpolating the random walk over the partition, i.e. by defining

$$\tilde{B}_{k/m}^{(m)} \triangleq S_k, \quad k = 0, 1, 2, \dots, m$$

and requiring that $\tilde{B}^{(m)}$ is linear on each sub-interval $[(k-1)/m, k/m]$. However, this construction cannot converge in any sense (as $m \rightarrow \infty$), since

$$\text{Var}[\tilde{B}_1^{(m)}] = \text{Var}[S_m] = m \rightarrow \infty.$$

To expect convergence, one needs to rescale the process $\tilde{B}_t^{(m)}$ suitably. If we divide $\tilde{B}^{(m)}$ by \sqrt{m} , from the central limit theorem we know that

$$\frac{\tilde{B}_1^{(m)}}{\sqrt{m}} = \frac{S_m}{\sqrt{m}} \xrightarrow{\text{dist}} N(0, 1) \quad \text{as } m \rightarrow \infty.$$

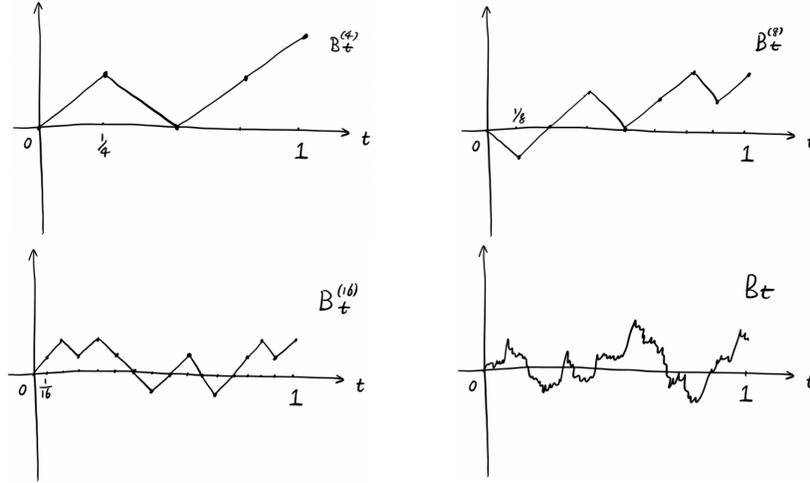


Figure 1.1: Random walks and Brownian motion.

Therefore, we should revise the definition of $\tilde{B}_t^{(m)}$ to be

$$B_t^{(m)} \triangleq \frac{\tilde{B}_t^{(m)}}{\sqrt{m}}, \quad t \in [0, 1].$$

It can then be shown that this sequence of processes $\{B_t^{(m)} : t \in [0, 1]\}$ converges to some limiting process $\{B_t : t \in [0, 1]\}$ in a suitable distributional sense as $m \rightarrow \infty$. Figure 1.1 provides the intuition.

The limiting process $\{B_t\}$ is known as the *Brownian motion*. It is not hard to convince ourselves that $B_t \sim N(0, t)$ for each t . Indeed, by the central limit theorem one has

$$B_t^{(m)} = \frac{\tilde{B}_t^{(m)}}{\sqrt{m}} = \sqrt{t} \cdot \frac{S_{tm}}{\sqrt{tm}} \xrightarrow{d} \sqrt{t} \cdot N(0, 1) = N(0, t)$$

as $m \rightarrow \infty$. In a similar way, one can heuristically see that $B_t - B_s \sim N(0, t - s)$ for $s < t$ and these increments are independent over disjoint time intervals. The Brownian motion plays a fundamental role in the theory of stochastic calculus and will be the central object of study in the next chapter.

In terms of the Brownian motion, the infinitesimal equation (1.2) for the stock price can be rewritten as

$$dX_t = \mu X_t dt + \sigma X_t dB_t.$$

This is an example of a *stochastic differential equation* (SDE). The renowned Black-Scholes option pricing formula is based on the use of such an SDE (cf. Section 5.2).

SDEs arise naturally in the description of the time evolution of systems that are subject to random perturbations. Understanding the meaning of these equations as well as properties of their solutions is a main objective of stochastic calculus. The theory is not a simple extension of ordinary calculus and requires new ideas, as the Brownian motion is a highly irregular (random) function and the differential dB_t makes no sense from the viewpoint of ordinary calculus.

1.1.2 Heat transfer and temperature distributions

Consider a rod of infinite length (modelled by the real line \mathbb{R}). The initial temperature distribution of the rod at time $t = 0$ is given by a function $f(x)$ (namely, $f(x)$ represents the initial temperature at position $x \in \mathbb{R}$). Suppose that heat transfers freely along the rod without any external heat source. How can we find the temperature distribution $u(t, x)$ of the rod at each time $t > 0$? From physical principles, $u(t, x)$ is the solution to the following partial differential equation (PDE)

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}, & t > 0, \\ u(0, x) = f(x). \end{cases}$$

The solution to the above PDE can be constructed in terms of the Brownian motion. Indeed, for each given $x \in \mathbb{R}$, let $\{B_t^x : t \geq 0\}$ be a Brownian motion starting at the position x . Then one has

$$u(t, x) = \mathbb{E}[f(B_t^x)], \quad t > 0, x \in \mathbb{R}. \quad (1.3)$$

Heuristically, the average value of the initial temperature distribution f over the Brownian motion at time t with starting position x gives the temperature $u(t, x)$.



This result extends to higher dimensions naturally.

Next, we consider a variant of the problem. Let $D \subseteq \mathbb{R}^2$ be a flat plate with boundary ∂D (e.g. a disk in \mathbb{R}^2). There is an external heat source acting on the boundary ∂D so that the temperature distribution over ∂D is fixed by

a given function $f : \partial D \rightarrow \mathbb{R}$ for all time. Suppose that heat transfers freely within the interior of the plate. After a sufficiently long period, what is the equilibrium temperature distribution inside the plate? From physical principles, the equilibrium temperature distribution $u : D \rightarrow \mathbb{R}$ satisfies the following PDE:

$$\begin{cases} \Delta u(x) = 0, & x \in D, \\ u(x) = f(x), & x \in \partial D, \end{cases} \quad (1.4)$$

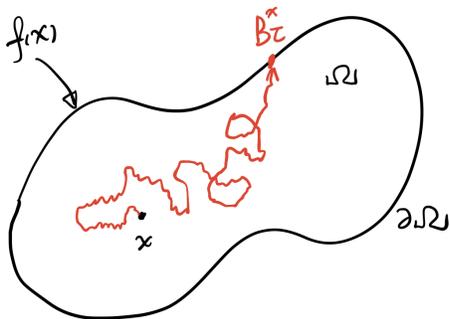
where $\Delta \triangleq \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}$ denotes the *Laplace operator*. The solution to this PDE can also be constructed by using the Brownian motion in \mathbb{R}^2 . More precisely, for each given $x = (x_1, x_2) \in D$, let $\{B_t^x : t \geq 0\}$ be a two-dimensional Brownian motion starting at the position x . Define

$$\tau \triangleq \inf\{t \geq 0 : B_t^x \in \partial D\}$$

to be the first time that the motion reaches the boundary of the plate. Then the solution to the PDE (1.4) is given by

$$u(x) = \mathbb{E}[f(B_\tau^x)], \quad x \in D.$$

Note that $B_\tau^x \in \partial D$ so that $f(B_\tau^x)$ is well-defined.



The above two problems do not involve the use of an SDE. However, it becomes relevant if the environment is not modelled by an Euclidean space (e.g. if the rod is curved or if the plate is a bended surface). In this case, in the PDE description of the temperature distribution, the Laplace operator needs to be replaced by a more general second order differential operator (say in the one-dimensional case):

$$\mathcal{A} = \frac{1}{2}a(x) \frac{d^2}{dx^2} + b(x) \frac{d}{dx}$$

with some suitable coefficient functions $a(x), b(x)$ that are related the geometry of the object. To obtain the PDE solution, the process B_t^x used in the above construction needs to be replaced by the solution to the SDE

$$\begin{cases} dX_t = b(X_t)dt + \sqrt{a(x)}dB_t, \\ X_0 = x, \end{cases}$$

where B_t is the Brownian motion.

We will revisit these motivating problems when we are acquainted with more tools from stochastic calculus.

1.2 Filtrations, stopping times and stochastic processes

Stochastic differential equations are used to describe the time evolution of random systems. Before studying them precisely, it is important to first have a mathematical way of describing the accumulation of information in the evolution of time. This leads us to the notion of filtrations.

From probability theory, we know that the proper mathematical way of describing information is through a σ -algebra. Let Ω be the underlying sample space. A class \mathcal{F} of subsets is called a σ -algebra over Ω , if it is stable under natural set operations, more precisely, if

- (i) $\Omega \in \mathcal{F}$;
- (ii) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$;
- (iii) $A_n \in \mathcal{F}$ for all $n \implies \cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Heuristically, a σ -algebra \mathcal{F} represents a collection of information. A subset A is \mathcal{F} -measurable (i.e. $A \in \mathcal{F}$) means that knowing the information provided by \mathcal{F} , one can determine whether the event A happens or not at each random experiment. More generally, a function $X : \Omega \rightarrow \mathbb{R}$ is \mathcal{F} -measurable means that knowing the information provided by \mathcal{F} , for each $a \in \mathbb{R}$ one can decide whether $X \leq a$ or not. In particular, one is then able to determine the value of X at each experiment. These interpretations (along with many others to be given in the sequel) are not mathematically precise. However, they are useful for developing the essential intuition behind the precise mathematical formulations.

Filtrations

The notion of a single σ -algebra does not take into account the evolution of time. To capture this situation, one can consider a family of σ -algebras which grows as

time increases.

Definition 1.1. Let Ω be a given sample space. A *filtration* over Ω is a family $\{\mathcal{F}_t : t \geq 0\}$ of σ -algebras such that

$$\mathcal{F}_s \subseteq \mathcal{F}_t \quad \forall s \leq t.$$

Heuristically, \mathcal{F}_t represents the accumulative information up to time t . In most situations, it is assumed that there is a given ultimate σ -algebra \mathcal{F} (the totality of information) and all the \mathcal{F}_t 's are contained in \mathcal{F} . There is also a *probability measure* \mathbb{P} defined on \mathcal{F} , i.e. a set function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ that satisfies:

- (i) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$;
- (ii) $\mathbb{P}(\Omega) = 1$;
- (iii) for any sequence $\{A_n : n \geq 1\} \subseteq \mathcal{F}$ of disjoint events (i.e. $A_m \cap A_n = \emptyset$ when $m \neq n$), one has

$$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

In probability theory, we call the triple $(\Omega, \mathcal{F}, \mathbb{P})$ a *probability space*. When it is equipped with a filtration $\{\mathcal{F}_t : t \geq 0\}$, we often refer to $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ as a *filtered probability space*.

Example 1.1. Although we are mostly interested in the continuous-time situation, we can also consider the discrete-time situation (i.e. the index set being $\mathbb{N} = \{0, 1, 2, \dots\}$). The definition of a filtration can be easily adapted to this case. As an example, consider the random experiment of tossing a coin repeatedly without stopping. The sample space Ω is defined by

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \omega_n = \text{H or T for each } n\}.$$

In other words, each generic outcome is an infinite sequence in which the n -th entry records the result of the n -th toss. A natural σ -algebra \mathcal{F} over Ω (the family of all legal events) should be the information generated by all the finite-step results. To be precise, for each $n \geq 1$ we define

$$A_n \triangleq \{\omega : \omega_n = H\}, \quad B_n \triangleq \{\omega : \omega_n = T\}.$$

A_n and B_n are the events corresponding to a specific result (“head” or “tail”) at the n -th toss. All these events A_n, B_n 's should be included as legal events in \mathcal{F} . As a result, we can define \mathcal{F} as the σ -algebra *generated* by the events

$$A_1, B_1, A_2, B_2, A_3, B_3, \dots,$$

i.e. the smallest σ -algebra containing all the events in the above list. This σ -algebra \mathcal{F} consists of the totality of information for the underlying random experiment. A natural choice of filtration is given by the results up to a specific step. More precisely, for each $n \geq 1$, we can define \mathcal{F}_n to be the σ -algebra generated by the events

$$A_1, B_1, \dots, A_n, B_n.$$

This is precisely the accumulative information provided by the first n tosses. We can also set $\mathcal{F}_0 \triangleq \{\emptyset, \Omega\}$ (the trivial information) as there is no meaningful information encoded at the initial time before the start of tossing. $\{\mathcal{F}_n : n \geq 0\}$ defines a filtration over Ω . The construction of a probability measure \mathbb{P} on \mathcal{F} requires the use of measure theory (Carathéodory's measure extension theorem). The main idea is that \mathbb{P} should satisfy (assuming that the coin is fair)

$$\mathbb{P}(\{\omega : \omega_1 = a_1, \dots, \omega_n = a_n\}) = \frac{1}{2^n} \tag{1.5}$$

for any arbitrary n and any choices of $a_1, \dots, a_n = \text{H or T}$. The measure extension theorem ensures the existence of a unique probability measure \mathbb{P} on \mathcal{F} that satisfies (1.5).

Stopping times

Sometimes we may consider information accumulated up to a random time rather than a deterministic time t . For instance, when predicting the behaviour of a volcano one relies on the dynamical data/information up to the next time of its eruption. However, the next eruption time is itself a random variable. In this case, we are talking about information up to a random time. The notion of a random time already appears in the discussion of the equilibrium temperature distribution in Section 1.1.2, where we have expressed the PDE solution in terms of the first time that the Brownian motion reaches the boundary. Since the Brownian motion is random, this hitting time is also a random variable.

Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a given filtered probability space. By a *random time* we shall mean a function $\tau : \Omega \rightarrow [0, +\infty]$. Allowing τ to achieve infinite value is convenient since it may not always be the case that τ is finite. In the volcano example, it is a theoretically possible outcome that the volcano never erupts in the future ($\tau(\omega) = +\infty$). Similar to the notion of random variables, to study its probabilistic properties one often needs to impose suitable measurability condition on a random time. Such a condition should respect the information flow given by the filtration $\{\mathcal{F}_t\}$.

Definition 1.2. A random time $\tau : \Omega \rightarrow [0, +\infty]$ is said to be an $\{\mathcal{F}_t\}$ -stopping time, if

$$\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \forall t \geq 0. \quad (1.6)$$

The idea behind the measurability condition (1.6) can be described as follows. For each given time t , suppose that we know the accumulative information up to time t . Then we are able to determine whether the event $\{\tau \leq t\}$ occurs or not. If it is the case that $\tau \leq t$, we can actually determine the exact value of τ . Indeed, since we can decide whether $\{\tau \leq s\}$ happens for every $s \in [0, t]$ (the information up to s is also known as part of \mathcal{F}_t), the value of τ can then be extracted from the turning point s at which $\{\tau \leq s\}$ happens while $\{\tau \leq s - \varepsilon\}$ fails for any $\varepsilon > 0$. On the other hand, if it is the case that $\tau > t$, no further implication on the value of τ can be made. In the volcano example, if we have observed its activity continuously for 100 days, we certainly know whether the volcano has erupted within this period of 100 days (i.e. $\tau \leq 100$) or not. If it does the given data should further tell us the exact eruption time, while if does not we cannot determine its eruption time by using the given information.

Apparently, every deterministic time is an $\{\mathcal{F}_t\}$ -stopping time. Moreover, one can construct new stopping times from the given ones. We use the notation $a \wedge b$ (respectively, $a \vee b$) to denote the minimum (respectively, the maximum) between two numbers a, b .

Proposition 1.1. *Suppose that σ, τ, τ_n are $\{\mathcal{F}_t\}$ -stopping times. Then*

$$\sigma + \tau, \sigma \wedge \tau, \sigma \vee \tau, \sup_n \tau_n$$

are all $\{\mathcal{F}_t\}$ -stopping times.

Proof. We only consider $\sigma + \tau$ and leave the other cases as an exercise. Consider the following decomposition:

$$\begin{aligned} \{\sigma + \tau > t\} &= \{\sigma = 0, \tau > t\} \cup \{0 < \sigma < t, \sigma + \tau > t\} \\ &\quad \cup \{\sigma \geq t, \tau > 0\} \cup \{\sigma > t, \tau = 0\}. \end{aligned} \quad (1.7)$$

The first and last events on the right hand side of (1.7) are clearly \mathcal{F}_t -measurable. The third event is in \mathcal{F}_t because

$$\{\sigma < t\} = \bigcup_{n \geq 1} \{\sigma \leq t - \frac{1}{n}\} \in \mathcal{F}_t.$$

For the second event, note that

$$\omega \in \{0 < \sigma < t, \sigma + \tau > t\} \implies \tau(\omega) > t - \sigma(\omega) > 0.$$

Since $\sigma(\omega) > 0$, one can choose $r \in (0, t) \cap \mathbb{Q}$, such that

$$\tau(\omega) > r > t - \sigma(\omega).$$

As a result, we see that

$$\{0 < \sigma < t, \sigma + \tau > t\} = \bigcup_{r \in (0, t) \cap \mathbb{Q}} \{\tau > r, t - r < \sigma < t\} \in \mathcal{F}_t.$$

□

It is helpful to re-examine the above property from the heuristic perspective. Let $t \geq 0$ be given and suppose that we know the accumulative information up to t . The criterion of being a stopping time is to see if we can decide whether $\{\sigma + \tau \leq t\}$ happens or not. Since σ, τ are both stopping times, the following scenarios are all decidable:

$$\sigma \leq t, \sigma > t, \tau \leq t, \tau > t.$$

If it is determined that either $\{\sigma > t\}$ or $\{\tau > t\}$ happens, then one decides that $\sigma + \tau > t$ since both σ, τ are non-negative. If it is determined that $\sigma \leq t$ and $\tau \leq t$, from the previous discussion on the definition of stopping times we know that the exact values of σ and τ are both decidable. As a result, the value of $\sigma + \tau$ can then be determined, which certainly allows us to decide if $\{\sigma + \tau \leq t\}$ happens or not.

One can use this kind of heuristic argument to discuss the other cases in Proposition 1.1. It also allows us to see e.g. why $\sigma - \tau$ may not necessarily be a stopping time (assuming $\sigma \geq \tau$). Indeed, we again suppose that the information up to t is presented. If one finds that $\sigma \leq t$, then $\{\sigma - \tau \leq t\}$ happens. However, if one finds that $\sigma > t$, no further implication on the value of σ can be made. In this case, $\sigma - \tau$ can either be smaller or larger than t and the occurrence of $\{\sigma - \tau \leq t\}$ is not decidable.

The most important class of stopping times is related to *hitting times* of a stochastic process (e.g. the one appearing in the plate heat transfer example). We will discuss this shortly after introducing the notion of stochastic processes.

σ -algebra at a stopping time

Let τ be a given $\{\mathcal{F}_t\}$ -stopping time. Recall that \mathcal{F}_t represents the information up to (the deterministic) time t . To generalise this idea, it is natural to talk about the accumulative information up to the stopping time τ . Mathematically, this should be defined by a suitable σ -algebra denoted as \mathcal{F}_τ .

The essential idea behind defining this σ -algebra is described as follows. First of all, an event $A \in \mathcal{F}_\tau$ means that knowing the information up to τ allows us to determine whether A happens or not. To rephrase this point properly in terms of the filtration $\{\mathcal{F}_t\}$, let t be a given fixed deterministic time. Suppose that we know the information up to time t . Since τ is a stopping time, we can determine whether $\{\tau \leq t\}$ has occurred or not. If it is the first case, the information up to τ is then known to us since we are given the information up to t and we have determined that $\tau \leq t$. In this scenario, we can decide whether A occurs or not. If it happens to be the second case ($\tau > t$), since we only have the information up to t , the information over the period $[t, \tau]$ is missing and we should not be able to decide whether A occurs in this scenario. To summarise, given the information up to time t , it is only in the scenario $\{\tau \leq t\}$ are we able to determine whether A occurs or not. This heuristic argument leads us to the following precise mathematical definition.

Definition 1.3. Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space and let τ be an $\{\mathcal{F}_t\}$ -stopping time. The σ -algebra at the stopping time τ is defined by

$$\mathcal{F}_\tau \triangleq \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \quad \forall t \geq 0\}.$$

The following fact justifies the definition of \mathcal{F}_τ .

Proposition 1.2. *The set class \mathcal{F}_τ is a σ -algebra.*

Proof. (i) Since τ is a stopping time, for each $t \geq 0$ we have

$$\Omega \cap \{\tau \leq t\} = \{\tau \leq t\} \in \mathcal{F}_t.$$

Therefore, $\Omega \in \mathcal{F}_\tau$.

(ii) Suppose $A \in \mathcal{F}_\tau$. Given an arbitrary $t \geq 0$, note that both of $\{\tau \leq t\}$ and $A \cap \{\tau \leq t\}$ belong to \mathcal{F}_t . As a result,

$$A^c \cap \{\tau \leq t\} = \{\tau \leq t\} \setminus (A \cap \{\tau \leq t\}) \in \mathcal{F}_t.$$

Therefore, $A^c \in \mathcal{F}_\tau$.

(iii) Let $A_n \in \mathcal{F}_\tau$ ($n \geq 1$). For each $t \geq 0$, we have

$$\left(\bigcup_{n=1}^{\infty} A_n\right) \cap \{\tau \leq t\} = \bigcup_{n=1}^{\infty} (A_n \cap \{\tau \leq t\}) \in \mathcal{F}_t$$

since $A_n \cap \{\tau \leq t\} \in \mathcal{F}_t$ for all n . Therefore, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_\tau$. \square

If $\tau \equiv t$ is a deterministic time, one can check by definition that $\mathcal{F}_\tau = \mathcal{F}_t$. In general, we have the following basic properties of \mathcal{F}_τ .

Proposition 1.3. *Suppose that σ, τ are two $\{\mathcal{F}_t\}$ -stopping times.*

(i) *Let $A \in \mathcal{F}_\sigma$. Then $A \cap \{\sigma \leq \tau\} \in \mathcal{F}_\tau$. In particular, if $\sigma \leq \tau$ then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.*

(ii) *We have $\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. In addition, the events*

$$\{\sigma < \tau\}, \{\sigma > \tau\}, \{\sigma \leq \tau\}, \{\sigma \geq \tau\}, \{\sigma = \tau\}$$

are all $\mathcal{F}_\sigma \cap \mathcal{F}_\tau$ -measurable.

Proof. (i) Let $t \geq 0$. Then we have

$$\begin{aligned} & A \cap \{\sigma \leq \tau\} \cap \{\tau \leq t\} \\ &= A \cap \{\sigma \leq \tau\} \cap \{\tau \leq t\} \cap \{\sigma \leq t\} \\ &= (A \cap \{\sigma \leq t\}) \cap \{\tau \leq t\} \cap \{\sigma \wedge t \leq \tau \wedge t\}. \end{aligned} \tag{1.8}$$

Since $A \in \mathcal{F}_\sigma$, by definition $A \cap \{\sigma \leq t\} \in \mathcal{F}_t$. In addition, $\{\tau \leq t\} \in \mathcal{F}_t$ as τ is a stopping time. For the last event in (1.8), the main observation is that both $\sigma \wedge t$ and $\tau \wedge t$ are \mathcal{F}_t -measurable. Indeed, for any $s \geq 0$ one has

$$\{\sigma \wedge t \leq s\} = \begin{cases} \{\sigma \leq s\} \in \mathcal{F}_s \subseteq \mathcal{F}_t, & \text{if } s < t; \\ \Omega \in \mathcal{F}_t, & \text{if } s \geq t. \end{cases}$$

This shows the \mathcal{F}_t -measurability of $\sigma \wedge t$ (and the same for $\tau \wedge t$). In particular, $\{\sigma \wedge t \leq \tau \wedge t\} \in \mathcal{F}_t$. We now see that the event defined by (1.8) is \mathcal{F}_t -measurable, and the claim follows from the definition of \mathcal{F}_τ .

(ii) Since $\sigma \wedge \tau$ is an $\{\mathcal{F}_t\}$ -stopping time, from Part (i) we know that $\mathcal{F}_{\sigma \wedge \tau} \subseteq \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. Conversely, let $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. Then we have

$$\begin{aligned} A \cap \{\sigma \wedge \tau \leq t\} &= A \cap (\{\sigma \leq t\} \cup \{\tau \leq t\}) \\ &= (A \cap \{\sigma \leq t\}) \cup (A \cap \{\tau \leq t\}) \in \mathcal{F}_t. \end{aligned}$$

Therefore, $A \in \mathcal{F}_{\sigma \wedge \tau}$. To prove the last claim, first note that (take $A = \Omega$ in Part (i))

$$\{\tau < \sigma\} = \{\sigma \leq \tau\}^c \in \mathcal{F}_\tau.$$

By replacing τ with $\sigma \wedge \tau$ in the above relation, we find

$$\{\tau < \sigma\} = \{\sigma \wedge \tau < \sigma\} \in \mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau.$$

All the other cases follow by symmetry and taking complement. \square

Remark 1.1. By taking $\sigma \equiv t$, we have

$$\{\tau \leq t\} \in \mathcal{F}_{\tau \wedge t} \subseteq \mathcal{F}_\tau \quad \forall t \geq 0.$$

In particular, τ is \mathcal{F}_τ -measurable.

One can re-examine the property $A \cap \{\sigma \leq \tau\} \in \mathcal{F}_\tau$ from the heuristic perspective. Suppose that the information up to τ is presented. If the event $\{\sigma \leq \tau\}$ occurs, the information up to σ is also known, and the occurrence of A is then decidable as $A \in \mathcal{F}_\sigma$. This is just recapturing the intuition behind the definition of \mathcal{F}_σ in a more general context.

Stochastic processes and their natural filtrations

Most of the basic objects in our study (Brownian motion, martingales, stochastic integrals, stochastic differential equations) are examples of a stochastic process.

Definition 1.4. A real-valued (continuous-time) *stochastic process* is a family $\{X_t : t \geq 0\}$ of random variables defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Remark 1.2. One can allow the index set to be of any shape. In our study, unless otherwise stated we always assume that the index set is $T = [0, \infty)$, so that $t \in T$ is interpreted as time and the stochastic process $\{X_t\}$ describes the time evolution of a random system.

Let $X = \{X_t : t \geq 0\}$ be a given stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. By definition, $X_t : \Omega \rightarrow \mathbb{R}$ is a random variable for each t . There is a more useful way of looking at a stochastic process: for each given $\omega \in \Omega$, one obtains a function

$$[0, \infty) \ni t \mapsto X_t(\omega).$$

This function (as a function of time) is called the *sample path* of the process X at the sample point ω . Note that the sample path depends on ω : one obtains different

sample paths for different ω 's. As a result, the function $t \mapsto X_t$ is considered as a *random function* of time. Equivalently, a stochastic process can be viewed as a “random variable” $\omega \mapsto X(\omega) = [t \mapsto X_t(\omega)]$ taking values in the *space of paths*.

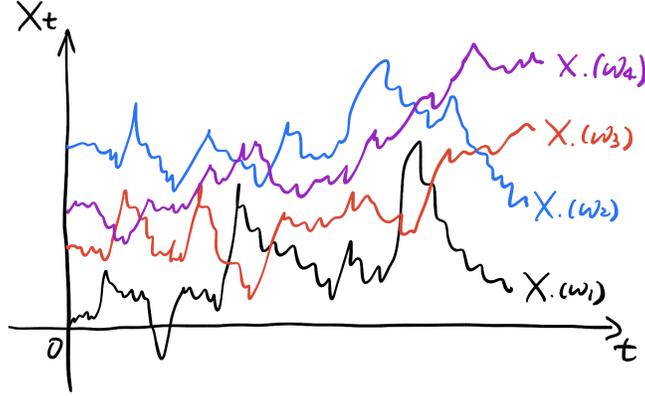


Figure 1.2: Sample paths of a stochastic process.

Under this viewpoint, for theoretical reasons one often needs to impose certain regularity assumptions on the sample paths. In most situations in our study, we shall assume that all sample paths of the underlying stochastic process are *continuous functions* of time. In this case, the process can be viewed as a “random variable” taking values in the space of continuous functions on $[0, \infty)$. Nonetheless, we point out that the theory of stochastic calculus for discontinuous processes (e.g. Lévy processes) is a rich subject of study. We do not discuss this situation in the current notes (cf. [2] for a general introduction).

It is often important to consider the relation between a stochastic process and a given filtration of information.

Definition 1.5. Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space and let $X = \{X_t\}$ be a given stochastic process. We say that X is $\{\mathcal{F}_t\}$ -*adapted*, if X_t is \mathcal{F}_t -measurable for each $t \geq 0$.

By the definition of adaptedness, given the information up to t we are able to determine the value of X_t . However, the more useful observation is the following. If we know the information up to t , then for every $s \leq t$ we also have the information up to s (since $\mathcal{F}_s \subseteq \mathcal{F}_t$). As a result, the value of X_s can be determined (for every $s \leq t$). In other words, the information up to t allows us to determine the *entire*

trajectory $s \mapsto X_s$ over the period $[0, t]$. Adaptedness is a basic condition that is assumed in most situations.

Given a stochastic process $X = \{X_t\}$, one can construct an associated natural filtration to encode the intrinsic accumulative information provided by X .

Definition 1.6. The *natural filtration* of the process $X = \{X_t\}$ is defined by

$$\mathcal{F}_t^X \triangleq \sigma(\{X_s : 0 \leq s \leq t\}), \quad t \geq 0,$$

where the right hand side denotes the smallest σ -algebra containing the following events:

$$\{X_s \leq a\} \quad \text{with } 0 \leq s \leq t, a \in \mathbb{R}.$$

Mathematically, \mathcal{F}_t^X is the smallest σ -algebra with respect to which all the X_s 's ($0 \leq s \leq t$) are measurable. Heuristically, \mathcal{F}_t^X is the intrinsic information carried by the trajectory of the process over the period $[0, t]$. It is trivial that X is always adapted to its natural filtration.

There is no reason to restrict ourselves to real-valued stochastic processes only. One can also consider stochastic processes taking values in \mathbb{R}^d (i.e. the position of a gas molecule at each time t has a stochastic process in \mathbb{R}^3). To conclude this section, we introduce an important example of stopping times. Let $X = \{X_t\}$ be an \mathbb{R}^d -valued stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$. We assume that X is $\{\mathcal{F}_t\}$ -adapted.

Proposition 1.4. *Suppose that every sample path of X is a continuous function. Let F be a given closed subset of \mathbb{R}^d . Define*

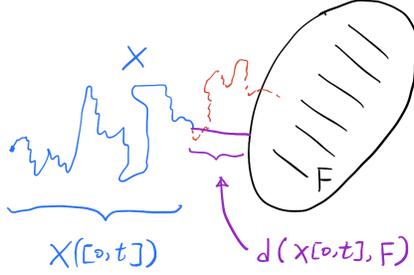
$$\tau \triangleq \inf\{t \geq 0 : X_t \in F\}$$

to the first time that the process hits the set F . Then τ is an $\{\mathcal{F}_t\}$ -stopping time.

Proof. Let $t \geq 0$ be given. Let us first observe that

$$\{\tau > t\} = \{d(X[0, t], F) > 0\}, \tag{1.9}$$

where $d(X[0, t], F)$ denotes the distance between the image of X on $[0, t]$ and the set F .



Indeed, if $\tau > t$, then $X([0, t]) \cap F = \emptyset$. Since both $X([0, t])$ and F are closed subsets of \mathbb{R}^d , they must be separated apart by a positive distance. Conversely, if $X([0, t])$ has a positive distance to F , by the continuity of sample paths the process will remain outside F at least within a sufficiently small amount of time after t . As a result, the hitting time of F must be strictly larger than t . Therefore, the relation (1.9) holds. The next observation is that

$$\{d(X[0, t], F) > 0\} = \bigcup_{n=1}^{\infty} \bigcap_{r \in [0, t] \cap \mathbb{Q}} \{d(X_r, F) > \frac{1}{n}\}, \quad (1.10)$$

which is again a simple consequence of the continuity of sample paths. Since X is $\{\mathcal{F}_t\}$ -adapted, we know that $X_r \in \mathcal{F}_r$. In particular,

$$\{d(X_r, F) > \frac{1}{n}\} \in \mathcal{F}_r \subseteq \mathcal{F}_t$$

for any $r \leq t$. Therefore, the right hand side of (1.10) and thus $\{\tau > t\}$ is \mathcal{F}_t -measurable. \square

Remark 1.3. The conclusion of Proposition 1.4 is in general not true if F is not assumed to be a closed subset (why?). Under what assumption can it be true for open subsets?

Remark 1.4. In more advanced texts, for technical reasons one often assumes that the underlying filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ satisfies the *usual conditions*, namely $\{\mathcal{F}_t\}$ is right continuous ($\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all t) and \mathcal{F}_0 contains all \mathbb{P} -null sets. These assumptions avoid many unpleasant technical issues related to hitting times, modification on null sets, space completeness etc. and it can be shown that they are not so restrictive at all. At the current level, we do not bother with these technical points and these assumptions will not be emphasised.

1.3 The conditional expectation

The idea of conditioning is essential in probability theory, as the a priori knowledge of partial information often changes the original distribution of the underlying random variable. Since different σ -algebras represent different amounts of information, it is natural to consider conditional probabilities/distributions *given a σ -algebra*. This leads us to the general notion of conditional expectation.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space. Let X be an integrable random variable on Ω (i.e. an \mathcal{F} -measurable function with finite expectation), and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub- σ -algebra of \mathcal{F} . Heuristically, \mathcal{G} contains a subset of information from \mathcal{F} . We want to define the conditional expectation of X given \mathcal{G} (denoted as $\mathbb{E}[X|\mathcal{G}]$).

Let us first make two extreme observations. If $\mathcal{G} = \{\emptyset, \Omega\}$ (the trivial σ -algebra), the information contained in \mathcal{G} is trivial. In this case, the most effective prediction of X given the information in \mathcal{G} is merely its mean value, i.e. $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}[X]$. Next, suppose that $\mathcal{G} = \mathcal{F}$ (the full information). Since X is \mathcal{F} -measurable, the information in \mathcal{F} allows us to determine the value of X at each random experiment. As a result, the prediction of X given \mathcal{F} should be the random variable X itself, i.e. $\mathbb{E}[X|\mathcal{F}] = X$. For those intermediate situations where \mathcal{G} is a non-trivial proper sub- σ -algebra of \mathcal{F} , it is reasonable to expect that $\mathbb{E}[X|\mathcal{G}]$ should be defined as a suitable random variable.

To motivate its definition, we first recapture an elementary situation. Suppose that A is a given event. The *conditional probability* of an arbitrary event B given A is defined as

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

When viewed as a set function, $\mathbb{P}(\cdot|A)$ is the *conditional probability measure* given A . The integral of X with respect to this conditional probability measure gives the *average value of X given the occurrence of A* :

$$\mathbb{E}[X|A] = \int_{\Omega} X d\mathbb{P}(\cdot|A) = \frac{\mathbb{E}[X\mathbf{1}_A]}{\mathbb{P}(A)}. \quad (1.11)$$

Now suppose that the given sub- σ -algebra \mathcal{G} is generated by a partition of Ω , say

$$\mathcal{G} = \sigma(A_1, A_2, \dots, A_n)$$

where $A_i \cap A_j = \emptyset$ and $\Omega = \cup_{i=1}^n A_i$. To define the random variable $\mathbb{E}[X|\mathcal{G}]$, the main idea is that on each event A_i the value of $\mathbb{E}[X|\mathcal{G}]$ should simply be the

average value of X given that A_i occurs. Mathematically, one has

$$\mathbb{E}[X|\mathcal{G}](\omega) = \sum_{i=1}^n c_i \mathbf{1}_{A_i}(\omega),$$

where

$$c_i \triangleq \mathbb{E}[X|A_i] = \frac{\mathbb{E}[X\mathbf{1}_{A_i}]}{\mathbb{P}(A_i)}, \quad i = 1, 2, \dots, n.$$

A key observation from this definition is that the integral of $\mathbb{E}[X|\mathcal{G}]$ on each event A_i coincides with the integral of X on the same event:

$$\int_{A_i} \mathbb{E}[X|\mathcal{G}] d\mathbb{P} = \int_{A_i} \sum_{j=1}^n c_j \mathbf{1}_{A_j}(\omega) d\mathbb{P} = c_i \mathbb{P}(A_i) = \mathbb{E}[X\mathbf{1}_{A_i}] = \int_{A_i} X d\mathbb{P}.$$

This observation motivates the following general definition of the conditional expectation.

Definition 1.7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space. Let X be an integrable random variable on Ω and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub- σ -algebra. The *conditional expectation of X given \mathcal{G}* is an integrable, \mathcal{G} -measurable random variable Y such that

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P} \quad \forall A \in \mathcal{G}. \quad (1.12)$$

This random variable is denoted as $\mathbb{E}[X|\mathcal{G}]$.

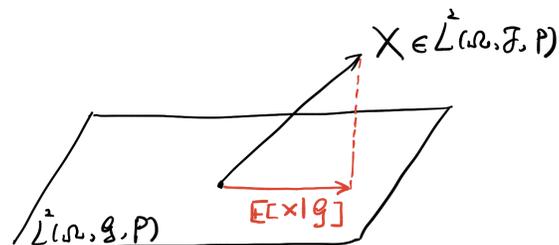
The existence and uniqueness of $\mathbb{E}[X|\mathcal{G}]$ is guaranteed by the so-called *Radon-Nikodym theorem* which we will not elaborate here. It is though useful to keep in mind that the uniqueness of $\mathbb{E}[X|\mathcal{G}]$ is understood in the following sense: if Y_1, Y_2 are two random variables satisfying the properties in Definition 1.7, then $Y_1 = Y_2$ *almost surely* (*a.s.*), i.e. $\mathbb{P}(Y_1 = Y_2) = 1$.

Although we do not discuss the general construction of $\mathbb{E}[X|\mathcal{G}]$, the following geometric intuition is enlightening. Let $H \triangleq L^2(\Omega, \mathcal{F}, \mathbb{P})$ denote the space of square integrable (i.e. having finite second moment), \mathcal{F} -measurable random variables. One can define natural notions of inner product, length and distance for elements in H :

$$\langle X, Y \rangle \triangleq \mathbb{E}[XY], \quad \|X\| \triangleq \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}[X^2]}, \quad d(X, Y) \triangleq \|X - Y\|.$$

Although this sounds abstract, the equipment of such a structure makes the space H analogous to the usual Euclidean space where elements are viewed as vectors

and lengths/angles can be measured. In particular, one can naturally talk about the orthogonal projection of a vector onto a given (closed) subspace. Now let $H_0 \triangleq L^2(\Omega, \mathcal{G}, \mathbb{P})$ denote the collection of random variables in H that are \mathcal{G} -measurable. Then H_0 is a closed subspace of H . It turns out that $\mathbb{E}[X|\mathcal{G}]$ is the orthogonal projection of X onto the subspace H_0 , i.e. the unique vector in H_0 that has minimal distance to X . Can you prove this fact?



By taking $A = \Omega$ in (1.12), it is clear that $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]]$. This is often known as the *law of total expectation*. We list several basic properties of the conditional expectation that are useful later on.

Theorem 1.1. *The conditional expectation satisfies the following properties. We always assume all the underlying random variables are integrable.*

- (i) *The map $X \mapsto \mathbb{E}[X|\mathcal{G}]$ is linear.*
- (ii) *If $X \leq Y$, then $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$. In particular,*

$$|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}].$$

- (iii) *If Z is \mathcal{G} -measurable, then*

$$\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}].$$

- (iv) *[The tower rule] If $\mathcal{G}_1 \subseteq \mathcal{G}_2$ are sub- σ -algebras of \mathcal{F} , then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbb{E}[X|\mathcal{G}_1].$$

- (v) *If X and \mathcal{G} are independent, then*

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X].$$

- (vi) *[Jensen's inequality] Let φ be a convex function on \mathbb{R} , i.e.*

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y)$$

for any $x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$. Then

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}]. \quad (1.13)$$

The proof of these properties are standard from the definition and we refer the reader to [21, Chapter 9] for the details. The intuition behind some of these properties are clear. For instance, for Property (iii), given the information in \mathcal{G} the value of Z is known. In other words, conditional on \mathcal{G} the random variable is frozen (treated as a constant) and can thus be moved outside the conditional expectation. In Property (v), by independence the knowledge of \mathcal{G} provides no meaningful information for the prediction of X . Therefore, the most effective prediction of X is its unconditional mean.

1.4 Martingales

A basic notion that describes the dynamics of a system under the effect of information growth is the concept of martingales. As we will see, the core techniques for studying stochastic integration and differential equations are based on martingale methods.

Heuristically, a martingale models the wealth process of a fair game. Mathematically, the fairness property can be described in terms of conditional expectation: given the information up to the present time, the conditional expectation of the future wealth is equal to the current wealth.

Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t : t \in T\})$ be a given filtered probability space, where T is a given subset of \mathbb{R} representing the index of time.

Definition 1.8. A real-valued stochastic process $X = \{X_t : t \in T\}$ is called an $\{\mathcal{F}_t\}$ -martingale (respectively, a *submartingale*/*supermartingale*) if the following properties hold true:

- (i) X is $\{\mathcal{F}_t\}$ -adapted;
- (ii) X_t is integrable for every $t \in T$;
- (iii) for every $s < t \in T$,

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s, \quad (\text{respectively } \geq / \leq). \quad (1.14)$$

Remark 1.5. This definition relies crucially on the underlying filtration. As a shorthand notation, we sometimes say that $\{X_t, \mathcal{F}_t : t \in T\}$ is a (sub/super)martingale. Note that a martingale with respect to one filtration may fail to be a martingale with respect to another.

Remark 1.6. In the discrete time context, the martingale property (1.14) is equivalent to $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ for all n (why?).

Example 1.2. Let $\{\xi_n : n = 1, 2, \dots\}$ denote an i.i.d. sequence of random variables with distribution

$$\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = \frac{1}{2}.$$

We consider the natural filtration associated with the process ξ :

$$\mathcal{F}_n \triangleq \sigma(\xi_1, \dots, \xi_n), \quad n \geq 1$$

with $\mathcal{F}_0 \triangleq \{\emptyset, \Omega\}$. Define $S_n \triangleq \xi_1 + \dots + \xi_n$ ($S_0 \triangleq 0$). Then $\{S_n, \mathcal{F}_n : n = 0, 1, 2, \dots\}$. It is clear that S_n is integrable and \mathcal{F}_n -measurable. To obtain the martingale property (1.14), for any $m > n$ we have

$$\begin{aligned} \mathbb{E}[S_m|\mathcal{F}_n] &= \mathbb{E}[S_n + \xi_{n+1} + \dots + \xi_m|\mathcal{F}_n] \\ &= \mathbb{E}[S_n|\mathcal{F}_n] + \mathbb{E}[\xi_{n+1} + \dots + \xi_m|\mathcal{F}_n] \\ &= S_n + \mathbb{E}[\xi_{n+1} + \dots + \xi_m] \\ &= S_n. \end{aligned}$$

A simple way of constructing a submartingale from a given martingale is to compose with convex functions. Recall that a function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is *convex* if

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y) \quad \forall \lambda \in [0, 1], x, y \in \mathbb{R}.$$

Proposition 1.5. *Let $\{X_t, \mathcal{F}_t : t \in T\}$ be a martingale (respectively, a submartingale). Suppose that $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function (respectively, a convex and increasing function). If $\varphi(X_t)$ is integrable for every $t \in T$, then $\{\varphi(X_t), \mathcal{F}_t : t \in T\}$ is a submartingale.*

Proof. The adaptedness and integrability conditions are clearly satisfied. To see the submartingale property, we apply Jensen's inequality (1.13) to find that

$$\mathbb{E}[\varphi(X_t)|\mathcal{F}_s] \geq \varphi(\mathbb{E}[X_t|\mathcal{F}_s]) \geq \varphi(X_s)$$

for any $s < t \in T$. □

Example 1.3. The functions

$$\varphi_1(x) = x^+ \triangleq \max\{x, 0\}, \quad \varphi_2(x) = |x|^p \quad (p \geq 1)$$

are convex on \mathbb{R} . As a result, if $\{X_t, \mathcal{F}_t\}$ is a martingale, then $\{X_t^+\}$ and $\{|X_t|^p\}$ ($p \geq 1$) are both $\{\mathcal{F}_t\}$ -submartingales, provided that $\mathbb{E}[|X_t|^p] < \infty$ for every t .

To develop the essential idea of martingales, in what follows we focus on the discussion of the discrete-time case. Some of the parallel results in the continuous-time case require only technical adaptations on which we will comment briefly in the sequel.

1.4.1 The martingale transform: discrete stochastic integration

We begin with a very useful construction of a class of martingales. This also provides the discrete version of stochastic integrals.

Let $T = \{0, 1, 2, \dots\}$ be the index set. We first introduce a few definitions.

Definition 1.9. Let $\{\mathcal{F}_n : n \geq 0\}$ be a filtration. A real-valued random sequence $\{C_n : n \geq 1\}$ is said to be $\{\mathcal{F}_n\}$ -predictable if C_n is \mathcal{F}_{n-1} -measurable for every $n \geq 1$.

Heuristically, predictability means that the future value C_{n+1} can be determined by the history up to the present time n .

Let $\{X_n : n \geq 0\}$ and $\{C_n : n \geq 1\}$ be two random sequences. We define another sequence $\{Y_n : n \geq 0\}$ by $Y_0 \triangleq 0$ and

$$Y_n \triangleq \sum_{k=1}^n C_k (X_k - X_{k-1}), \quad n \geq 1.$$

Definition 1.10. The sequence $\{Y_n : n \geq 0\}$ is called the *martingale transform* of $\{X_n\}$ by $\{C_n\}$. We often write Y_n as $(C \bullet X)_n$.

The martingale transform is a discrete-time version of stochastic integration as seen from the continuous/discrete comparison:

$$\int C_t dX_t \approx \sum_k C_k (X_{t_k} - X_{t_{k-1}}).$$

The following important result justifies its name.

Theorem 1.2. Let $\{X_n, \mathcal{F}_n : n \geq 0\}$ be a martingale (respectively, submartingale/supermartingale) and let $\{C_n : n \geq 1\}$ be an $\{\mathcal{F}_n\}$ -predictable random sequence which is uniformly bounded (respectively, bounded and non-negative). Then the martingale transform $\{(C \bullet X)_n, \mathcal{F}_n : n \geq 0\}$ is a martingale (respectively, submartingale, supermartingale).

Proof. We only consider the martingale case. Adaptedness and integrability are clear. To check the martingale property, we have

$$\begin{aligned}\mathbb{E}[(C \bullet X)_{n+1} | \mathcal{F}_n] &= \mathbb{E}[(C \bullet X)_n + C_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n] \\ &= (C \bullet X)_n + C_{n+1} \cdot (\mathbb{E}[X_{n+1} | \mathcal{F}_n] - X_n) \\ &= (C \bullet X)_n,\end{aligned}$$

where we have used the predictability of $\{C_n\}$ to reach the second last identity. \square

Remark 1.7. The boundedness of $\{C_n\}$ is not an essential assumption. It is imposed to guarantee the integrability of Y_n .

The following intuition of the martingale transform is particularly useful. Suppose that you are gambling over the time horizon $\{1, 2, \dots\}$. The quantity C_n represents your stake at game n . Predictability means that you are making your next decision on the stake amount C_{n+1} based on the information \mathcal{F}_n observed up to the present round n . The quantity $X_n - X_{n-1}$ represents your winning at game n per unit stake. As a result, Y_n is your total winning up to time n . Theorem 1.2 asserts that if the game is fair (i.e. $\{X_n, \mathcal{F}_n\}$ is a martingale) and if you are playing the game based on the intrinsic information carried by the game itself (predictability), then you cannot beat fairness (your wealth process $\{Y_n, \mathcal{F}_n\}$ is also a martingale).

As we will see below, Theorem 1.2 can be used as a unified approach to establish several fundamental results in martingale theory. These results were all due to J.L. Doob in the 1950s.

1.4.2 The martingale convergence theorem

The (sub/super)martingale property (1.14) exhibits certain kind of monotone behaviour. It is therefore reasonable to expect that a (sub/super)martingale converges in a suitable sense if its mean sequence does not explode in the long run.

Recall that a random sequence $\{X_n : n \geq 0\}$ is said to be convergent *almost surely* (a.s.) if it is convergent for every ω outside some event of zero probability. Equivalently, the event consisting of those ω 's at which $\{X_n(\omega)\}$ is not convergent has zero probability.

Before establishing the martingale convergence theorem, we first explain a general strategy of proving the almost sure convergence of a random sequence. Let $X = \{X_n : n \geq 0\}$ be a given random sequence. Then $\{X_n(\omega)\}$ is convergent if and only if

$$\underline{\lim}_{n \rightarrow \infty} X_n(\omega) = \overline{\lim}_{n \rightarrow \infty} X_n(\omega).$$

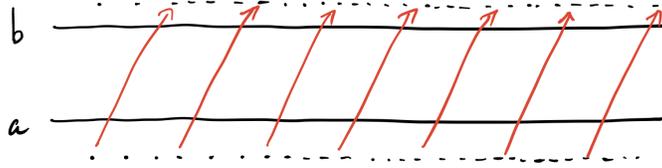
Therefore,

$$\begin{aligned} \{X_n \text{ is not convergent}\} &\subseteq \left\{ \liminf_{n \rightarrow \infty} X_n < \overline{\lim}_{n \rightarrow \infty} X_n \right\} \\ &\subseteq \bigcup_{\substack{a < b \\ a, b \in \mathbb{Q}}} \left\{ \liminf_{n \rightarrow \infty} X_n < a < b < \overline{\lim}_{n \rightarrow \infty} X_n \right\}. \end{aligned}$$

In order to prove that X_n converges almost surely, it suffices to show that

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} X_n < a < b < \overline{\lim}_{n \rightarrow \infty} X_n\right) = 0 \quad (1.15)$$

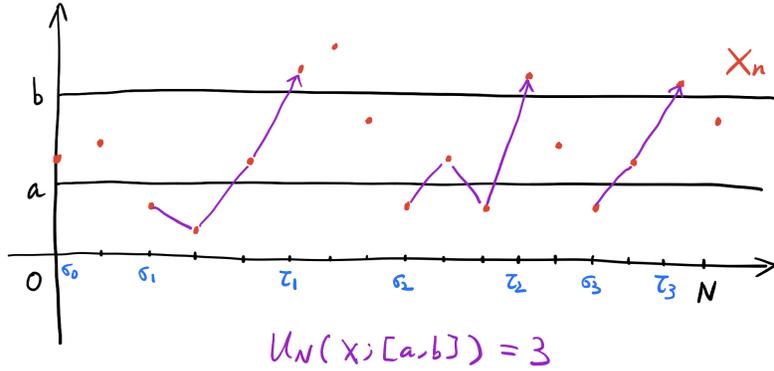
for every pair of given numbers $a < b$. Here comes the key observation: due to the definition of the liminf/limsup, the event in (1.15) implies that there is a subsequence of X_n lying below a while there is another subsequence of X_n lying above b . This further implies that, as n increases there must be infinitely many *upcrossings* by the sequence X_n from below the level a to above the level b .



From the above reasoning, the key step for proving the a.s. convergence of $\{X_n\}$ is to control its total upcrossing number with respect to the interval $[a, b]$, more specifically, to show that with probability one there are at most finitely many upcrossings with respect to $[a, b]$.

We now define the upcrossing number mathematically. Consider the following two sequences of random times: $\sigma_0 \triangleq 0$,

$$\begin{aligned} \sigma_1 &\triangleq \inf\{n \geq 0 : X_n < a\}, & \tau_1 &\triangleq \inf\{n > \sigma_1 : X_n > b\}, \\ \sigma_2 &\triangleq \inf\{n > \tau_1 : X_n < a\}, & \tau_2 &\triangleq \inf\{n > \sigma_2 : X_n > b\}, \\ &\dots & & \\ \sigma_k &\triangleq \inf\{n > \tau_{k-1} : X_n < a\}, & \tau_k &\triangleq \inf\{n > \sigma_k : X_n > b\}, \\ &\dots & & \end{aligned}$$



Definition 1.11. Given $N \geq 0$, the *upcrossing number* $U_N(X; [a, b])$ with respect to the interval $[a, b]$ by the sequence $\{X_n\}$ up to time N is defined by the random number

$$U_N(X; [a, b]) \triangleq \sum_{k=1}^{\infty} \mathbf{1}_{\{\tau_k \leq N\}}.$$

Note that $U_N(X; [a, b]) \leq N/2$. Moreover, if $\{\mathcal{F}_n : n \geq 0\}$ is a filtration and X is $\{\mathcal{F}_n\}$ -adapted, then σ_k, τ_k are $\{\mathcal{F}_n\}$ -stopping times. In particular, $U_N(X; [a, b])$ is \mathcal{F}_N -measurable. The main result of controlling the quantity $U_N(X; [a, b])$ is stated as follows.

Proposition 1.6 (The Upcrossing Inequality). *Let $\{X_n, \mathcal{F}_n : n \geq 0\}$ be a supermartingale. Then the upcrossing number $U_N(X; [a, b])$ satisfies the following inequality:*

$$\mathbb{E}[U_N(X; [a, b])] \leq \frac{\mathbb{E}[(X_N - a)^-]}{b - a}, \quad (1.16)$$

where $x^- \triangleq \max\{-x, 0\}$.

Proof. The main idea is to construct a suitable martingale transform of $\{X_n\}$ (a suitable gambling strategy). Consider the gambling model where $X_n - X_{n-1}$ represents the winning at game n per unit stake. Let us construct a gambling strategy as follows: repeat the following two steps forever:

- (i) wait until X_n gets below a ;
- (ii) play unit stakes onwards until X_n gets above b and then stop playing.

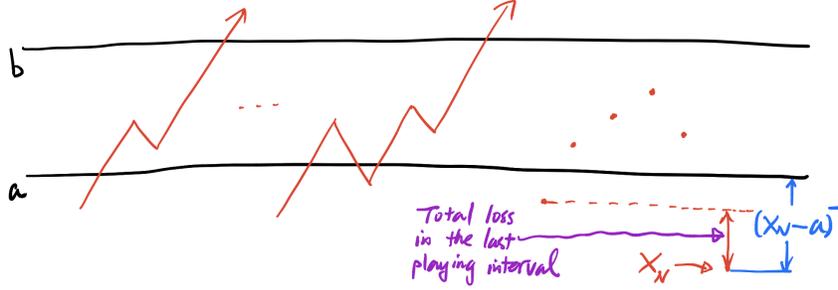
Mathematically, the strategy $\{C_n : n \geq 1\}$ is defined by the following equations:

$$C_1 \triangleq \mathbf{1}_{\{X_0 < a\}}$$

and

$$C_n \triangleq \mathbf{1}_{\{C_{n-1}=0\}} \mathbf{1}_{\{X_{n-1} < a\}} + \mathbf{1}_{\{C_{n-1}=1\}} \mathbf{1}_{\{X_{n-1} \leq b\}}, \quad n \geq 2.$$

Let $\{Y_n\}$ be the martingale transform of X_n by C_n . Then Y_N represents the total winning up to time N . Note that Y_N comes from two parts: the playing intervals corresponding to complete upcrossings, and the last playing interval corresponding to the last incomplete upcrossing (which might not exist).



The total winning Y_N from the first part is clearly bounded from below by $(b - a)U_N(X; [a, b])$. The total winning in the last playing interval (if it exists) is bounded from below by $-(X_N - a)^-$ (the worst scenario when a loss occurs). Consequently, we find that

$$Y_N \geq (b - a)U_N(X; [a, b]) - (X_N - a)^-.$$

On the other hand, from the construction it is clear that $\{C_n\}$ is a bounded, non-negative and $\{\mathcal{F}_n\}$ -predictable. According to Theorem 1.2, $\{Y_n, \mathcal{F}_n\}$ is a supermartingale. Therefore,

$$\mathbb{E}[Y_N] \leq \mathbb{E}[Y_0] = 0,$$

which then gives (1.16). □

Remark 1.8. There is also a version of the upcrossing inequality for the submartingale case. However, the proof of that case is quite different from what we give here. Since they both lead to the same convergence theorem, we only consider the supermartingale case.

Since $U_N(X; [a, b])$ is increasing in N , one can define the total upcrossing number for all time as

$$U_\infty(X; [a, b]) \triangleq \lim_{N \rightarrow \infty} U_N(X; [a, b]).$$

From the upcrossing inequality, if we assume that the supermartingale $\{X_n, \mathcal{F}_n\}$ is *bounded in L^1* , namely

$$\sup_{n \geq 0} \mathbb{E}[|X_n|] < \infty, \quad (1.17)$$

then

$$\mathbb{E}[U_\infty(X; [a, b])] = \lim_{N \rightarrow \infty} \mathbb{E}[U_N(X; [a, b])] \leq \frac{\sup_{n \geq 0} \mathbb{E}[|X_n|] + |a|}{b - a} < \infty.$$

In particular, $U_\infty(X; [a, b]) < \infty$ almost surely. It then follows from the relation

$$\left\{ \liminf_{n \rightarrow \infty} X_n < a < b < \overline{\lim}_{n \rightarrow \infty} X_n \right\} \subseteq \{U_\infty(X; [a, b]) = \infty\}$$

that (1.15) holds. As a result, we conclude that X_n is convergent almost surely. Let us denote the limiting random variable as X_∞ . From Fatou's lemma, under the L^1 -boundedness assumption (1.17) we also know that

$$\mathbb{E}[|X_\infty|] = \mathbb{E}\left[\lim_{n \rightarrow \infty} |X_n|\right] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[|X_n|] \leq \sup_{n \geq 0} \mathbb{E}[|X_n|] < \infty.$$

To summarise, we have established the following convergence result.

Theorem 1.3 (The Supermartingale Convergence Theorem). *Let $\{X_n, \mathcal{F}_n : n \geq 0\}$ be a supermartingale which is bounded in L^1 . Then X_n converges almost surely to an integrable random variable X_∞ .*

Remark 1.9. Since martingales are supermartingales and a submartingale is the negative of a supermartingale, it is immediate that the above convergence theorem is also valid for (sub)martingales.

1.4.3 The optional sampling theorem

It is reasonable to expect that the martingale property (1.14) remains valid even when we sample along stopping times. This is the content of the *optional sampling theorem*.

To elaborate this fact, let $\{X_n, \mathcal{F}_n : n \geq 0\}$ be a (sub/super)martingale and let τ be an $\{\mathcal{F}_n\}$ -stopping time. We introduce the *stopped process*

$$X_n^\tau \triangleq X_{\tau \wedge n} = \begin{cases} X_n, & n \leq \tau, \\ X_\tau, & n > \tau. \end{cases}$$

Theorem 1.4. *The stopped process X_n^τ is an $\{\mathcal{F}_n\}$ -(sub/super)martingale.*

Proof. As in the proof of the upcrossing inequality, we represent X_n^τ through a gambling model. The gambling strategy is constructed as follows: keep playing unit stake from the beginning and quit immediately after the time τ . Mathematically, the strategy is defined by

$$C_n \triangleq \mathbf{1}_{\{n \leq \tau\}}, \quad n \geq 1.$$

Then the total winning up to time n is given by $(C \bullet X)_n = X_{\tau \wedge n} - X_0$. The result follows immediately from Theorem 1.2. \square

Next, we consider the situation when we also stop our filtration at a stopping time. For simplicity, we only consider the situation where the underlying stopping times are both uniformly bounded.

Theorem 1.5 (The Optional Sampling Theorem). *Let $\{X_n, \mathcal{F}_n : n \geq 0\}$ be a martingale. Suppose that σ, τ are two bounded $\{\mathcal{F}_n\}$ -stopping times such that $\sigma \leq \tau$. Then X_σ (respectively, X_τ) is integrable, \mathcal{F}_σ -measurable (respectively, \mathcal{F}_τ -measurable) and*

$$\mathbb{E}[X_\tau | \mathcal{F}_\sigma] = X_\sigma. \quad (1.18)$$

Proof. Assume that $\sigma \leq \tau \leq N$ for some constant $N \geq 0$. The integrability and \mathcal{F}_σ -measurability of X_σ is left as an exercise. To obtain the martingale property, by the definition of conditional expectation one needs to show that

$$\int_F X_\tau d\mathbb{P} = \int_F X_\sigma d\mathbb{P} \quad \forall F \in \mathcal{F}_\sigma. \quad (1.19)$$

Let $F \in \mathcal{F}_\sigma$ be given fixed. Consider the gambling strategy of playing unit stake at each time step from $\sigma + 1$ until τ under the occurrence of F :

$$C_n \triangleq \mathbf{1}_F \mathbf{1}_{\{\sigma < n \leq \tau\}}, \quad n \geq 1.$$

The total winning by time N is $(C \bullet X)_N = (X_\tau - X_\sigma) \mathbf{1}_F$. On the other hand, $\{C_n\}$ is $\{\mathcal{F}_n\}$ -predictable since

$$F \cap \{\sigma < n \leq \tau\} = F \cap \{\sigma \leq n - 1\} \cap (\tau \leq n - 1)^c \in \mathcal{F}_{n-1}.$$

According to Theorem 1.2, $\{(C \bullet X)_n, \mathcal{F}_n\}$ is a martingale. In particular,

$$\mathbb{E}[(C \bullet X)_N] = \mathbb{E}[(X_\tau - X_\sigma) \mathbf{1}_F] = \mathbb{E}[(C \bullet X)_0] = 0.$$

This gives the desired property (1.19). \square

Remark 1.10. The above proof clearly applies to the sub/super martingale situation as well. Under suitable conditions, the result can be extended to the case of unbounded stopping times. We will not discuss this general situation (cf. [21, Sec. 10.10] and [5, Sec. 9.3]).

1.4.4 The maximal and L^p -inequalities

By using the optional sampling theorem for bounded stopping times, we derive two basic martingale inequalities that are important for the study of stochastic integrals and differential equations. In this part, we work with submartingales.

The core result is known as *the maximal inequality*. As a submartingale exhibits an increasing trend, it is not surprising that its running maximum can be controlled by the terminal value in some sense.

Theorem 1.6. *Let $\{X_n, \mathcal{F}_n : n \geq 0\}$ be a submartingale. For every $N \geq 0$ and $\lambda > 0$, the following inequality holds true:*

$$\mathbb{P}\left(\max_{0 \leq n \leq N} X_n \geq \lambda\right) \leq \frac{\mathbb{E}[X_N^+]}{\lambda}.$$

Proof. Let $\sigma \triangleq \inf\{n \leq N : X_n \geq \lambda\}$ denote the first time (up to N) that X_n exceeds the level λ . We set $\sigma = N$ if no such $n \leq N$ exists. Clearly σ is an $\{\mathcal{F}_n\}$ -stopping time bounded by N . By taking expectation on both sides of (1.18) (in the submartingale case), we have

$$\mathbb{E}[X_N] \geq \mathbb{E}[X_\sigma]. \quad (1.20)$$

On the other hand, we can write

$$X_\sigma = X_\sigma \mathbf{1}_{\{X_N^* \geq \lambda\}} + X_\sigma \mathbf{1}_{\{X_N^* < \lambda\}}$$

where

$$X_N^* \triangleq \max_{0 \leq n \leq N} X_n.$$

On the event $\{X_N^* \geq \lambda\}$, the process X_n does exceed λ at some $n \leq N$ and thus $X_\sigma \geq \lambda$. On the event $\{X_N^* < \lambda\}$ no such exceeding occurs and thus $X_\sigma = X_N$ ($\sigma = N$ in this case). As a result, we have

$$\begin{aligned} \mathbb{E}[X_N] &\geq \mathbb{E}[X_\sigma] = \mathbb{E}[X_\sigma \mathbf{1}_{\{X_N^* \geq \lambda\}}] + \mathbb{E}[X_\sigma \mathbf{1}_{\{X_N^* < \lambda\}}] \\ &\geq \lambda \mathbb{P}(X_N^* \geq \lambda) + \mathbb{E}[X_N \mathbf{1}_{\{X_N^* < \lambda\}}]. \end{aligned}$$

It follows that

$$\lambda \mathbb{P}(X_N^* \geq \lambda) \leq \mathbb{E}[X_N] - \mathbb{E}[X_N \mathbf{1}_{\{X_N^* < \lambda\}}] = \mathbb{E}[X_N \mathbf{1}_{\{X_N^* \geq \lambda\}}] \leq \mathbb{E}[X_N^+], \quad (1.21)$$

which yields the desired inequality. \square

An important corollary of the maximal inequality is the L^p -inequality for the running maximum. Before establishing the result, we first need the following lemma. Given a random variable X , we use $\|X\|_p \triangleq \mathbb{E}[|X|^p]^{1/p}$ to denote its L^p -norm ($p \geq 1$). We say that $X \in L^p$ if $\|X\|_p < \infty$.

Lemma 1.1. *Suppose that X, Y are two non-negative random variables such that*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[Y \mathbf{1}_{\{X \geq \lambda\}}]}{\lambda} \quad \forall \lambda > 0. \quad (1.22)$$

Then for any $p > 1$, we have

$$\|X\|_p \leq q \|Y\|_p, \quad (1.23)$$

where $q \triangleq p/(p-1)$ (so that $1/p + 1/q = 1$).

Proof. Suppose $\|Y\|_p < \infty$ for otherwise the result is trivial. We write

$$\mathbb{E}[X^p] = \mathbb{E}\left[\int_0^X p\lambda^{p-1} d\lambda\right] = \mathbb{E}\left[\int_0^\infty p\lambda^{p-1} \mathbf{1}_{\{X \geq \lambda\}} d\lambda\right].$$

By switching the order of the two integrals, we have

$$\begin{aligned} \mathbb{E}[X^p] &= \int_0^\infty p\lambda^{p-1} \mathbb{P}(X \geq \lambda) d\lambda \\ &\leq \int_0^\infty p\lambda^{p-2} \mathbb{E}[Y \mathbf{1}_{\{X \geq \lambda\}}] d\lambda \\ &= \mathbb{E}\left[Y \int_0^X p\lambda^{p-2} d\lambda\right] \\ &= \frac{p}{p-1} \mathbb{E}[Y X^{p-1}]. \end{aligned} \quad (1.24)$$

To proceed further, we assume for the moment that $X \in L^p$. According to Hölder's inequality (cf. Appendix (7)), we have

$$\mathbb{E}[Y X^{p-1}] \leq \|Y\|_p \|X^{p-1}\|_q = \|Y\|_p \|X\|_p^{p-1}.$$

The inequality (1.23) thus follows by dividing $\|X\|_p^{p-1}$ to the left hand side of (1.24). If $\|X\|_p = \infty$, we let $X^N \triangleq X \wedge N$ ($N \geq 1$). By considering the cases $\lambda > N$ and $\lambda \leq N$ separately, it is not hard to see that the condition (1.22) holds for the pair (X^N, Y) . The desired inequality (1.23) follows by first considering X^N and then applying the monotone convergence theorem. \square

The L^p -inequality for (sub)martingales is stated as follows.

Corollary 1.1. *Let $\{X_n, \mathcal{F}_n : n \geq 0\}$ be a non-negative submartingale. Let $p > 1$ and suppose that $X_n \in L^p$ for all n . Then for every $N \geq 0$, we have*

$$\|X_N^*\|_p \leq q \|X_N\|_p$$

where $X_N^* \triangleq \max_{0 \leq n \leq N} X_n$ and $q \triangleq p/(p-1)$. In particular, $X_N^* \in L^p$.

Proof. We have shown in (1.21) that

$$\mathbb{P}(X_N^* \geq \lambda) \leq \frac{\mathbb{E}[X_N \mathbf{1}_{\{X_N^* \geq \lambda\}}]}{\lambda}.$$

In particular, the condition (1.22) holds with $(X, Y) = (X_N^*, X_N)$. The result follows immediate from Lemma 1.1. \square

1.4.5 The continuous-time case

All the aforementioned results for discrete-time (sub/super)martingales hold in the continuous-time case, provided that suitable regularity conditions on the sample paths are imposed. In the current study, we assume that $X = \{X_t, \mathcal{F}_t : t \geq 0\}$ is a (sub/super)martingale such that *every sample path of X is continuous*. This assumption is almost sufficient for the purpose of diffusion theory and significantly simplifies several technical considerations (such as measurability properties). In what follows, we only point out the essential idea of extending the previous results to the continuous-time context. For the technical details, we refer the reader to [17, Sec. II.5].

The martingale convergence theorem. This can be established by using exactly the same idea based on upcrossing numbers. The only place which needs care is the definition of upcrossing numbers. Let $a < b$ be two real numbers. Given a finite subset $F \subseteq [0, \infty)$, we define $U_F(X; [a, b])$ to be the upcrossing number with respect to $[a, b]$ by the process $\{X_t : t \in F\}$, defined in the same way as in the discrete-time case. For a given time interval $I \subseteq [0, \infty)$, we set

$$U_I(X; [a, b]) \triangleq \sup\{U_F(X; [a, b]) : F \subseteq I, F \text{ is finite}\}.$$

This random variable $U_I(X; [a, b])$ records the upcrossing number for the process X over the time interval I . Since X has continuous sample paths, one can approximate $U_I(x; [a, b])$ by the upcrossing number over rational times in I . Now the

crucial observation is that for each fixed n , the discrete-time upcrossing inequality is uniform with respect to all finite subsets $F \subseteq [0, N]$. This allows us to take limit (by further approximating $[0, N] \cap \mathbb{Q}$ by finite subsets) to obtain the same upcrossing inequality for the random variable $U_{[0, N]}(X; [a, b])$.

The optional sampling theorem. The extension of this result to the continuous-time case is trickier. The main idea is to discretise the process as well as the given stopping times. More precisely, suppose that $\sigma, \tau < N$ with some deterministic number N . For each $n \geq 1$, we define

$$\sigma_n \triangleq \sum_{k=1}^n \frac{kN}{n} \mathbf{1}_{\{\frac{k-1}{n}N \leq \sigma < \frac{k}{n}N\}}$$

and similarly for τ_n . Then $\sigma_n \leq \tau_n$ are bounded $\{\mathcal{F}_{kN/n} : k = 0, 1, 2, \dots, n\}$ -stopping times taking values on the discrete-time grid $\{kN/n\}_{k=0}^n$. As a result, we can apply the discrete-time optional sampling theorem to get (say in the martingale case)

$$\mathbb{E}[X_{\tau_n} | \mathcal{F}_{\sigma_n}] = X_{\sigma_n}.$$

The next key observation is that $\sigma_n \downarrow \sigma$ and $\tau_n \downarrow \tau$ as $n \rightarrow \infty$. This allows us to take limit to obtain the desired property

$$\mathbb{E}[X_{\tau} | \mathcal{F}_{\sigma}] = X_{\sigma}.$$

We should however point out that the above limiting procedure is a non-trivial matter and relies on a tool known as the *backward martingale convergence theorem*.

The maximal and L^p -inequalities. The extension of this part is similar to the case of the upcrossing inequality. Firstly, the continuity of sample paths implies that

$$\sup_{t \in [0, N]} X_t = \sup_{t \in [0, N] \cap \mathbb{Q}} X_t.$$

In addition, the discrete-time maximal and L^p -inequalities are both uniform with respect to the restriction of the process X to any finite subset $F \subseteq [0, N]$. The continuous-time inequalities follow by approximating the index set $[0, N] \cap \mathbb{Q}$ by finite subsets.

Convention. Throughout the rest of the notes, a *continuous (sub/super)martingale* means a (sub/super)martingale indexed by continuous-time and *has continuous sample paths*.

2 Brownian motion

Based on principles of statistical physics, in 1905 A. Einstein discovered the mechanism governing the random movement of particles suspended in a fluid, a phenomenon first observed by the botanist R. Brown in 1827. Such a random motion is commonly known as the *Brownian motion*. In 1900, L. Bachelier first used the distribution of Brownian motion to model the Paris stock market and evaluate stock options. The precise mathematical construction of Brownian motion was due to N. Wiener in 1923.

Brownian motion is among the most important objects of study, as it lies at the intersection of almost all fundamental kinds of stochastic processes: it is a Gaussian process, a martingale, a Markov process, a diffusion process and a Lévy process. In addition, it creates a bridge connecting probabilistic methods with other branches of mathematics such as partial differential equations, harmonic analysis, differential geometry, group theory as well as applied areas such as physics and finance.

Before developing the theory of stochastic calculus which is essentially the differential calculus for Brownian motion, we must first spend some time investigating properties of the Brownian motion.

2.1 The construction of Brownian motion

In Section 1.1.1, we have motivated the Brownian motion as a suitable scaling limit of simple random walks and postulated its distributional properties. From the discussion over there, it is also reasonable to expect that the Brownian motion has continuous sample paths. The precise definition of Brownian motion is given as follows.

Definition 2.1. A stochastic process $B = \{B_t : t \geq 0\}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a (one-dimensional) *Brownian motion* if the following properties hold:

- (i) $\mathbb{P}(B_0 = 0) = 1$;
- (ii) $B_t - B_s \sim N(0, t - s)$ for any $s < t$;
- (iii) for any $n \geq 1$ and $t_1 < t_2 < \dots < t_n$, the increments

$$B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$$

are independent;

- (iv) almost every sample path of B is continuous, namely there exists a \mathbb{P} -null set N such that the function $t \mapsto B_t(\omega)$ is continuous for every $\omega \notin N$.

Remark 2.1. There is no harm to assume that $B_0(\omega) = 0$ and $t \mapsto B_t(\omega)$ is continuous for every $\omega \in \Omega$.

More generally, one can allow the Brownian motion to start from an arbitrarily given position x by requiring that $\mathbb{P}(B_0 = x) = 1$. We call this a *Brownian motion starting at x* . One can also define multidimensional Brownian motions: a Brownian motion in \mathbb{R}^d is a stochastic process $B = \{(B_t^1, \dots, B_t^d) : t \geq 0\}$ such that the components B^1, \dots, B^d are independent one-dimensional Brownian motions.

The Brownian motion has the following invariance properties. The proof is almost immediate from the definition and is left as an exercise.

Proposition 2.1. *Let $B = \{B_t : t \geq 0\}$ be a Brownian motion.*

- (i) *Translation invariance: for every $s \geq 0$, the process $\{B_{t+s} - B_s : t \geq 0\}$ is a Brownian motion.*
- (ii) *Reflection invariance: the process $-B$ is a Brownian motion.*
- (iii) *Scaling invariance: for each $\lambda > 0$, the process $\{\lambda^{-1}B_{\lambda^2 t} : t \geq 0\}$ is a Brownian motion.*

Before investigating deeper properties of Brownian motion, we first address the question of its existence. To this end, we follow the original idea of N. Wiener to construct the Brownian motion from the perspective of (random) Fourier series.

Theorem 2.1. *There exists a probability space on which a Brownian motion is defined.*

The rest of this section is devoted to the proof of Theorem 2.1. We only construct the Brownian motion on $[0, \pi]$ and let the reader think about how a Brownian motion on $[0, \infty)$ can be produced based on this construction.

2.1.1 Some notions on Fourier series

The classical *Fourier series* gives a formal expansion of a function $f : [-\pi, \pi] \rightarrow \mathbb{R}$ in terms of the elementary trigonometric functions:

$$f(t) \simeq \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt), \quad t \in [-\pi, \pi], \quad (2.1)$$

where the *Fourier coefficients* a_n, b_n are given by

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) dt, \quad a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos ntdt, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin ntdt.$$

The method of obtaining these expressions for the coefficients is easy: one simply observes that the trigonometric functions $\cos nt$ ($n \geq 0$), $\sin nt$ ($n \geq 1$) are orthogonal to each other:

$$\int_{-\pi}^{\pi} \cos mt \cdot \cos ntdt = \int_{-\pi}^{\pi} \sin mt \cdot \sin ntdt = 0 \quad \forall m \neq n$$

and

$$\int_{-\pi}^{\pi} \cos mtdt = \int_{-\pi}^{\pi} \sin ntdt = \int_{-\pi}^{\pi} \cos mt \cdot \sin ntdt = 0 \quad \forall m, n.$$

For instance, to compute b_n we multiply the equation (2.1) by $\sin nt$ and integrate over $[-\pi, \pi]$. The above orthogonality properties shows that

$$\int_{-\pi}^{\pi} f(t) \sin ntdt = b_n \cdot \int_{-\pi}^{\pi} \sin^2 ntdt = \pi b_n.$$

Note that if $f(t)$ is an even function on $[-\pi, \pi]$, one has $b_n = 0$ for all n and the expansion becomes

$$f(t) \simeq \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nt, \quad t \in [0, \pi] \quad (2.2)$$

with

$$a_0 = \frac{2}{\pi} \int_0^{\pi} f(t) dt, \quad a_n = \frac{2}{\pi} \int_0^{\pi} f(t) \cos ntdt. \quad (2.3)$$

2.1.2 Wiener's original idea

The key insight behind Wiener's construction of the Brownian motion $\{B_t : t \in [0, \pi]\}$ is to represent its "derivative" \dot{B}_t in terms of a (random) Fourier series. Before proceeding, we first make a note that the argument below is entirely formal (the derivative of Brownian motion makes no sense) and is intended for motivating the essential idea. The precise mathematical construction is given in the next part.

We begin by noting that the distribution of $B_{t+\delta t} - B_t$ is independent of t . As a result, the "derivative" \dot{B}_t behaves like a "constant function" and can thus be treated as an "even function". According to (2.2), one expects that

$$\dot{B}_t \simeq \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nt, \quad t \in [0, \pi].$$

Since B_t is a random, the Fourier coefficients a_0, a_n are themselves random variables. In view of (2.3), they are given by

$$a_0 = \frac{2}{\pi} \int_0^\pi \dot{B}_t dt, \quad a_n = \frac{2}{\pi} \int_0^\pi (\cos nt) \cdot \dot{B}_t dt. \quad (2.4)$$

To figure out the distribution of these coefficients, let us make the following key observation. Due to the special Gaussian nature of the Brownian motion, given any pair of square-integrable functions $f, g : [0, \pi] \rightarrow \mathbb{R}$, the random variables

$$X_f \triangleq \int_0^\pi f(t) \dot{B}_t dt, \quad X_g \triangleq \int_0^\pi g(t) \dot{B}_t dt$$

should be jointly Gaussian with mean zero. We claim that their covariance should be given by the L^2 -inner product of f and g :

$$\mathbb{E}[X_f X_g] = \int_0^\pi f(t) g(t) dt.$$

Indeed, let us simply assume that f, g are step functions:

$$f(t) = \sum_{i=1}^n c_i \mathbf{1}_{(u_{i-1}, u_i]}, \quad g(t) = \sum_{i=1}^n d_i \mathbf{1}_{(u_{i-1}, u_i]},$$

where $0 = u_0 < u_1 < \dots < u_{n-1} < u_n = \pi$ is a finite partition of $[0, \pi]$ and $c_i, d_i \in \mathbb{R}$. In this case, we have

$$X_f = \sum_{i=1}^n c_i \int_0^\pi \mathbf{1}_{(u_{i-1}, u_i]}(t) \dot{B}_t dt = \sum_{i=1}^n c_i \int_{u_{i-1}}^{u_i} \dot{B}_t dt = \sum_{i=1}^n c_i (B_{u_i} - B_{u_{i-1}})$$

and a similar expression holds for X_g . It follows from the definition of Brownian motion (Properties (ii) and (iii)) that

$$\begin{aligned} \mathbb{E}[X_f X_g] &= \mathbb{E}\left[\left(\sum_{i=1}^n c_i (B_{u_i} - B_{u_{i-1}})\right) \cdot \left(\sum_{j=1}^n d_j (B_{u_j} - B_{u_{j-1}})\right)\right] \\ &= \sum_{i=1}^n c_i d_i \mathbb{E}[(B_{u_i} - B_{u_{i-1}})^2] = \sum_{i=1}^n c_i d_i (u_i - u_{i-1}) \\ &= \int_0^\pi f(t) g(t) dt. \end{aligned}$$

As a consequence, we see that (by taking $f = g$)

$$X_f \sim N\left(0, \int_0^\pi f^2(t)dt\right) \quad (2.5)$$

and

$$X_f, X_g \text{ are independent if } \int_0^\pi f(t)g(t)dt = 0. \quad (2.6)$$

By applying (2.5) and (2.6) to (2.4) for $f = 1$ and $\cos nt$, one finds that

$$\frac{a_0}{2} \sim N\left(0, \frac{1}{\pi}\right), \quad a_n \sim N\left(0, \frac{2}{\pi}\right)$$

and these coefficients are all independent. To put it in an equivalent way, we can formally represent

$$\dot{B}_t \simeq \frac{1}{\sqrt{\pi}}\xi_0 + \sum_{n=1}^{\infty} \sqrt{\frac{2}{\pi}}\xi_n \cos nt \quad (2.7)$$

where $\{\xi_n : n \geq 1\}$ is an i.i.d. sequence of standard normal random variables. By integrating (2.7) from $[0, t]$, we obtain the following representation:

$$B_t \simeq \frac{t}{\sqrt{\pi}}\xi_0 + \sqrt{\frac{2}{\pi}} \sum_{n=1}^{\infty} \frac{\xi_n \sin nt}{n}, \quad t \in [0, \pi]. \quad (2.8)$$

This representation motivates the precise construction of the Brownian motion which we elaborate in what follows.

2.1.3 The mathematical construction

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space on which an i.i.d. sequence $\{\xi_n : n = 0, 1, 2, \dots\}$ of standard normal random variables are defined. The existence of such a probability space is a standard construction from measure theory which will not be discussed here (cf. Appendix (11)).

For each $n \geq 1$, we set

$$S_t^{(n)} \triangleq \frac{t}{\sqrt{\pi}}\xi_0 + \sqrt{\frac{2}{\pi}} \sum_{k=1}^n \frac{\xi_k \sin kt}{k}, \quad t \in [0, \pi].$$

Note that the stochastic process $S^{(n)}(t)$ is the partial sum (up to n) in the representation (2.8) and it has continuous sample paths since the function $\sin kt$ is

continuous. To construct the Brownian motion rigorously (with continuous sample paths), one needs to show that with probability one, $S^{(n)}$ converges *uniformly* on $[0, \pi]$. With no surprise the limiting continuous (random) function will then be a Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

However, directly proving the almost sure uniform convergence of $S^{(n)}$ is a rather challenging task. In what follows, we take a shortcut by proving the a.s. uniform convergence of $S^{(n)}$ along a subsequence. This will be enough to produce a Brownian motion in the limit.

Lemma 2.1. *With probability one, the sequence $\{S^{(2^m)} : m \geq 1\}$ is uniformly convergent on $[0, \pi]$.*

Proof. The following elegant complexification argument was due to K. Itô and H. McKean [8, Sec. 1.5]. For each $m < n$, we set

$$S_t^{(m,n)} \triangleq \sum_{k=m+1}^n \frac{\xi_k \sin kt}{k}, \quad C_{m,n} \triangleq \|S^{(m,n)}\|_\infty \triangleq \sup_{0 \leq t \leq \pi} |S_t^{(m,n)}|.$$

Note that

$$S_t^{(n)} - S_t^{(m)} = \sqrt{\frac{2}{\pi}} S_t^{(m,n)}.$$

By viewing $\sin kt$ as the imaginary part of the complex number e^{ikt} , we have

$$\begin{aligned} |S_t^{(m,n)}|^2 &= \left| \operatorname{Re} \left(\sum_{k=m+1}^n \frac{\xi_k e^{ikt}}{k} \right) \right|^2 \leq \left| \sum_{k=m+1}^n \frac{\xi_k e^{ikt}}{k} \right|^2 \\ &= \left(\sum_{k=m+1}^n \frac{\xi_k e^{ikt}}{k} \right) \cdot \overline{\left(\sum_{l=m+1}^n \frac{\xi_l e^{ilt}}{l} \right)} = \left(\sum_{k=m+1}^n \frac{\xi_k e^{ikt}}{k} \right) \left(\sum_{l=m+1}^n \frac{\xi_l e^{-ilt}}{l} \right) \\ &= \sum_{k=m+1}^n \frac{\xi_k^2}{k^2} + \sum_{m+1 \leq k \neq l \leq n} e^{i(k-l)t} \frac{\xi_k \xi_l}{kl} \\ &= \sum_{k=m+1}^n \frac{\xi_k^2}{k^2} + \sum_{j=1}^{n-m-1} (e^{ijt} + e^{-ijt}) \sum_{k=m+1}^{n-j} \frac{\xi_k \xi_{k+j}}{k(k+j)} \quad (j \triangleq |l-k|). \end{aligned}$$

It follows that

$$C_{m,n}^2 \leq \sum_{k=m+1}^n \frac{\xi_k^2}{k^2} + 2 \sum_{j=1}^{n-m-1} \left| \sum_{k=m+1}^{n-j} \frac{\xi_k \xi_{k+j}}{k(k+j)} \right|.$$

By using the Cauchy-Schwarz inequality (cf. Appendix (7)), we obtain

$$\begin{aligned}\mathbb{E}[C_{m,n}]^2 &\leq \mathbb{E}[C_{m,n}^2] \leq \sum_{k=m+1}^n \frac{1}{k^2} + 2 \sum_{j=1}^{n-m-1} \mathbb{E}\left[\left|\sum_{k=m+1}^{n-j} \frac{\xi_k \xi_{k+j}}{k(k+j)}\right|\right] \\ &\leq \sum_{k=m+1}^n \frac{1}{k^2} + 2 \sum_{j=1}^{n-m-1} \mathbb{E}\left[\left|\sum_{k=m+1}^{n-j} \frac{\xi_k \xi_{k+j}}{k(k+j)}\right|^2\right]^{1/2}.\end{aligned}$$

The expectation on the right hand side is computed as

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{k=m+1}^{n-j} \frac{\xi_k \xi_{k+j}}{k(k+j)}\right|^2\right] &= \mathbb{E}\left[\left(\sum_{k=m+1}^{n-j} \frac{\xi_k \xi_{k+j}}{k(k+j)}\right) \cdot \left(\sum_{l=m+1}^{n-j} \frac{\xi_l \xi_{l+j}}{l(l+j)}\right)\right] \\ &= \sum_{k=m+1}^{n-j} \frac{\mathbb{E}[\xi_k^2] \cdot \mathbb{E}[\xi_{k+j}^2]}{k^2(k+j)^2} = \sum_{k=m+1}^{n-j} \frac{1}{k^2(k+j)^2}.\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}[C_{m,n}]^2 &\leq \sum_{k=m+1}^n \frac{1}{k^2} + 2 \sum_{j=1}^{n-m-1} \left(\sum_{k=m+1}^{n-j} \frac{1}{k^2(k+j)^2}\right)^{1/2} \\ &\leq \frac{n-m}{m^2} + 2(n-m) \left(\frac{n-m}{m^4}\right)^{1/2} \leq \frac{3(n-m)^{3/2}}{m^2}.\end{aligned}$$

In particular, by taking $n = 2m$, we arrive at

$$\mathbb{E}[C_{m,2m}] \leq \sqrt{3}m^{-1/4}. \quad (2.9)$$

This estimate naturally leads us to the consideration of the subsequence $S^{(2^m)}$. Indeed, from (2.9) we know that

$$\mathbb{E}\left[\sum_{m=1}^{\infty} \|S^{(2^m)} - S^{(2^{m-1})}\|_{\infty}\right] = \sqrt{\frac{2}{\pi}} \sum_{m=1}^{\infty} \mathbb{E}[C_{2^{m-1}, 2^m}] \leq \sqrt{\frac{6}{\pi}} \sum_{m=1}^{\infty} 2^{-m/4} < \infty.$$

As a result,

$$\sum_{m=1}^{\infty} \|S^{(2^m)} - S^{(2^{m-1})}\|_{\infty} < \infty \quad \text{a.s.}$$

In particular, with probability one $\{S^{(2^m)} : m \geq 1\}$ is a Cauchy sequence under the uniform distance and is thus uniformly convergent. \square

According to Lemma 2.1, there is a \mathbb{P} -null set N , such that $S^{(2^m)}(\omega)$ is uniformly convergent at every $\omega \notin N$. We define a stochastic process $B = \{B_t : t \in [0, \pi]\}$ by

$$B_t(\omega) \triangleq \begin{cases} \lim_{m \rightarrow \infty} S_t^{(2^m)}(\omega), & \omega \notin N, \\ 0, & \omega \in N. \end{cases}$$

To complete the proof of Theorem 2.1, we need to check that B is indeed a Brownian motion on $[0, \pi]$. It is clear that $B_0 = 0$ since $S_0^{(n)} = 0$ for all n . In addition, since B is the uniform limit of $S^{(2^m)}$, we know that the sample paths of B are continuous. It remains to verify the desired distributional properties. For this purpose, we rely on the following observation whose proof is left as an exercise.

Proposition 2.2. *A stochastic process $\{X_t : t \geq 0\}$ is a Brownian motion if and only if it satisfies Properties (i), (iv) and additionally it is a Gaussian process (i.e. $(X_{t_1}, \dots, X_{t_n})$ is jointly Gaussian for any choices of n and $t_1 < \dots < t_n$) with covariance function*

$$\mathbb{E}[X_s X_t] = s \wedge t, \quad s, t \geq 0. \quad (2.10)$$

Since $S^{(2^m)}$ is a Gaussian process for every m , the limiting processes B is also Gaussian. In particular,

$$\mathbb{E}[B_s B_t] = \lim_{m \rightarrow \infty} \mathbb{E}[S_s^{(2^m)} S_t^{(2^m)}] = \frac{st}{\pi} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\sin ns \sin nt}{n^2}.$$

To verify the desired covariance property (2.10), the following analytical identity provides the last piece of the puzzle.

Lemma 2.2. *For any $s, t \in [0, \pi]$, we have*

$$s \wedge t = \frac{st}{\pi} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\sin ns \sin nt}{n^2}.$$

Proof. Let $f(u) \triangleq \mathbf{1}_{[0,s]}(u)$ and $g(u) \triangleq \mathbf{1}_{[0,t]}(u)$. By treating f, g as even functions on $[-\pi, \pi]$, their Fourier series expansions are given by (cf. (2.2), (2.3))

$$f(u) = \frac{s}{\pi} + \sum_{n=1}^{\infty} \frac{2 \sin ns}{n\pi} \cos nu, \quad g(u) = \frac{t}{\pi} + \sum_{n=1}^{\infty} \frac{2 \sin nt}{n\pi} \cos nu, \quad u \in [0, \pi].$$

It follows that

$$\begin{aligned} s \wedge t &= \int_0^\pi f(u)g(u)du = \frac{st}{\pi^2} \int_0^\pi du + \sum_{n=1}^\infty \frac{4 \sin ns \cdot \sin nt}{n^2\pi^2} \int_0^\pi \cos^2 nudu \\ &= \frac{st}{\pi} + \frac{2}{\pi} \sum_{n=1}^\infty \frac{\sin ns \sin nt}{n^2}. \end{aligned}$$

□

Remark 2.2. It is easy to show that $\{S_t^{(n)} : n \geq 1\}$ converges in L^2 for each fixed t . Indeed, from the following estimate

$$\mathbb{E}[(S_t^{(m)} - S_t^{(n)})^2] \leq \frac{2}{\pi} \sum_{k=m+1}^n \frac{\mathbb{E}[\xi_k^2] \sin^2 kt}{k^2} \leq \frac{2}{\pi} \sum_{k=m+1}^n \frac{1}{k^2}$$

one sees that $\{S_t^{(n)} : n \geq 1\}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. As a result, it converges to some $B(t) \in L^2$ for each given t . In a similar way as before, the process $\{B(t) : t \in [0, \pi]\}$ is Gaussian and satisfies the covariance property (2.10). However, it is not clear at all why it has continuous sample paths from this perspective. The main effort in the previous argument is to ensure this point by proving uniform convergence.

2.2 The strong Markov property and the reflection principle

The definition of Brownian motion given in the last section does not take into account the presence of a filtration. We shall extend the definition to include this situation.

Definition 2.2. Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t : t \geq 0\})$ be a given filtered probability space. A stochastic process $B = \{B(t) : t \geq 0\}$ is said to be an $\{\mathcal{F}_t\}$ -Brownian motion, if it satisfies the following properties:

- (i) $\mathbb{P}(B_0 = 0) = 1$;
- (ii) B is $\{\mathcal{F}_t\}$ -adapted;
- (iii) for any $s < t$, the random variable $B_t - B_s$ is independent of \mathcal{F}_s and is Gaussian distributed with mean zero and variance $t - s$.
- (iv) With probability one, B has continuous sample paths.

An $\{\mathcal{F}_t\}$ -Brownian motion is a Brownian motion in the sense of Definition 2.1. In addition, a Brownian motion in the sense of Definition 2.1 is a Brownian motion with respect to its natural filtration.

Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a given filtered probability space and let B be an $\{\mathcal{F}_t\}$ -Brownian motion.

The first fundamental property of Brownian motion is its Markov property. Heuristically, the Markov property means that *given the knowledge of the present state*, the history on the past does not provide any additional information on predicting the distribution of future states (the past and future are independent *given the knowledge of the present*). Mathematically, one can express the Markov property as

$$\mathbb{P}(B_{t+s} \in \Gamma | \mathcal{F}_t) = \mathbb{P}(B_{t+s} \in \Gamma | B_t), \quad \forall s, t \geq 0, \Gamma \in \mathcal{B}(\mathbb{R}). \quad (2.11)$$

In the Brownian context, this property is straight forward. Indeed, let us write

$$B_{t+s} = B_{t+s} - B_t + B_t.$$

Since $B_{t+s} - B_t$ is independent of the entire history \mathcal{F}_t , the distribution of the future state B_{t+s} is uniquely determined by the knowledge of current state B_t and the independent increment distribution $N(0, s)$.

The more useful property of Brownian motion is its *strong Markov property*, which asserts that the Markov property remains valid even if we take the present time as a stopping time. Instead of formulating the general strong Markov property, we directly state the following stronger result for the Brownian motion. Its proof relies on the optional sampling theorem for martingales.

Theorem 2.2. *Let $B = \{B_t\}$ be an $\{\mathcal{F}_t\}$ -Brownian motion and let τ be a finite $\{\mathcal{F}_t\}$ -stopping time. Then the process $B^{(\tau)} \triangleq \{B_{\tau+t} - B_\tau : t \geq 0\}$ is a Brownian motion which is independent of \mathcal{F}_τ .*

Remark 2.3. Theorem 2.2 implies the strong Markov property by expressing the future state $B_{\tau+t}$ as

$$B_{\tau+t} = B_{\tau+t} - B_\tau + B_\tau = B_t^\tau + B_\tau.$$

Since B_t^τ is independent of \mathcal{F}_τ , the distribution of $B_{\tau+t}$ is uniquely determined by the present state B_τ as well as the increment distribution $N(0, t)$.

Proof. There are two essential properties to check: the Brownian distribution of $B^{(\tau)}$ and its independence from \mathcal{F}_τ . To be more specific, let $0 = t_0 < t_1 < \dots < t_n$ be an arbitrary collection of indices. We want to check that:

(i) the random vector

$$X \triangleq (B_{t_1}^{(\tau)} - B_{t_0}^{(\tau)}, \dots, B_{t_n}^{(\tau)} - B_{t_{n-1}}^{(\tau)})$$

is independent of \mathcal{F}_τ ;

(ii) X has the required Gaussian distribution, i.e. $B_{t_k}^{(\tau)} - B_{t_{k-1}}^{(\tau)} \sim N(0, t_k - t_{k-1})$ for each k and the components of X are independent.

Since distribution and independence can both be characterised in terms of the characteristic function (cf. Appendix (9,10)), the required Properties (i) and (ii) are captured by the following claim in one go:

$$\mathbb{E}[\xi \cdot \exp(i \sum_{k=1}^n \theta_k (B_{t_k}^{(\tau)} - B_{t_{k-1}}^{(\tau)})] = \mathbb{E}[\xi] \cdot \exp(-\frac{1}{2} \sum_{k=1}^n \theta_k^2 (t_k - t_{k-1})) \quad (2.12)$$

for every bounded, \mathcal{F}_τ -measurable random variable ξ and every choice of $\theta_1, \dots, \theta_n \in \mathbb{R}$. Indeed, by taking $\xi = 1$ one obtains

$$\mathbb{E}[\exp(i \sum_{k=1}^n \theta_k (B_{t_k}^{(\tau)} - B_{t_{k-1}}^{(\tau)})] = \exp(-\frac{1}{2} \sum_{k=1}^n \theta_k^2 (t_k - t_{k-1}))$$

which yields the desired distribution of X . The equation (2.12) then becomes

$$\mathbb{E}[\xi \cdot \exp(i \sum_{k=1}^n \theta_k (B_{t_k}^{(\tau)} - B_{t_{k-1}}^{(\tau)})] = \mathbb{E}[\xi] \cdot \mathbb{E}[\exp(i \sum_{k=1}^n \theta_k (B_{t_k}^{(\tau)} - B_{t_{k-1}}^{(\tau)})]$$

which further implies that X and \mathcal{F}_τ are independent due to the arbitrariness of ξ and $\theta_1, \dots, \theta_n$.

We now proceed to establish (2.12). The essential idea is to make use of the optional sampling theorem for a suitable martingale. Given $\theta \in \mathbb{R}$, let us define the process

$$M_t^{(\theta)} = \exp(i\theta B_t + \frac{1}{2}\theta^2 t), \quad t \geq 0.$$

Then $\{M_t^{(\theta)}, \mathcal{F}_t\}$ is a martingale. Indeed,

$$\begin{aligned} \mathbb{E}[M_t^{(\theta)} | \mathcal{F}_s] &= \mathbb{E}[M_s^{(\theta)} \exp(i\theta(B_t - B_s) + \frac{1}{2}\theta^2(t-s)) | \mathcal{F}_s] \\ &= M_s^{(\theta)} \mathbb{E}[\exp(i\theta(B_t - B_s) + \frac{1}{2}\theta^2(t-s)) | \mathcal{F}_s] \\ &= M_s^{(\theta)}, \end{aligned}$$

where the last identity follows from the fact that $B_t - B_s \sim N(0, t - s)$ and it is independent of \mathcal{F}_s .

Next, we claim that

$$\mathbb{E}\left[e^{i\theta(B_{\sigma+t}-B_\sigma)}|\mathcal{F}_\sigma\right] = e^{-\theta^2 t/2} \quad (2.13)$$

for any finite $\{\mathcal{F}_t\}$ -stopping time σ and $t \geq 0$. This is a direct consequence of the optional sampling theorem in the case when σ is uniformly bounded. In fact, (2.13) is merely a rearrangement of the property

$$\mathbb{E}[M_{\sigma+t}^{(\theta)}|\mathcal{F}_\sigma] = M_\sigma^{(\theta)}$$

in this case. The extension to the case when σ is unbounded is technical and tedious, so we leave it to the end.

Assuming the correctness of (2.13), the desired claim (2.12) follows by taking conditional expectation and applying (2.13) recursively, starting from $\sigma \triangleq \tau + t_{n-1}$, $t \triangleq t_n - t_{n-1}$ and $\theta \triangleq \theta_n$. We unwind this process in the special case when $n = 2$:

$$\begin{aligned} & \mathbb{E}\left[\xi \exp\left(i\theta_1(B_{\tau+t_1} - B_\tau) + i\theta_2(B_{\tau+t_2} - B_{\tau+t_1})\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\xi \exp\left(i\theta_1(B_{\tau+t_1} - B_\tau) + i\theta_2(B_{\tau+t_2} - B_{\tau+t_1})\right)\right|\mathcal{F}_{\tau+t_1}\right]\right] \\ &= \mathbb{E}\left[\xi \exp\left(i\theta_1(B_{\tau+t_1} - B_\tau)\right) \cdot \mathbb{E}\left[\exp\left(i\theta_2(B_{\tau+t_1+(t_2-t_1)} - B_{\tau+t_1})\right)\right|\mathcal{F}_{\tau+t_1}\right]\right] \\ &= e^{-\theta_2^2(t_2-t_1)} \cdot \mathbb{E}\left[\xi \exp\left(i\theta_1(B_{\tau+t_1} - B_\tau)\right)\right] \\ &= e^{-\theta_2^2(t_2-t_1)} \cdot \mathbb{E}\left[\xi \mathbb{E}\left[\exp\left(i\theta_1(B_{\tau+t_1} - B_\tau)\right)\right|\mathcal{F}_\tau\right]\right] \\ &= e^{-\theta_2^2(t_2-t_1)} e^{-\theta_1^2 t_1} \cdot \mathbb{E}[\xi]. \end{aligned}$$

Now we return to prove (2.13) for the case when σ is unbounded. We first consider the bounded stopping time $\sigma \wedge N$ and observe that

$$\mathbb{E}\left[e^{i\theta(B_{\sigma \wedge N+t}-B_{\sigma \wedge N})}|\mathcal{F}_{\sigma \wedge N}\right] = e^{-\theta^2 t/2} \quad (2.14)$$

which is true for every N . Since σ is assumed to be finite, we know that $\sigma \wedge N \uparrow \sigma$ as $N \rightarrow \infty$. As a result, the property (2.13) should follow by sending $N \rightarrow \infty$ in (2.14). To make this limiting procedure rigorous, let $A \in \mathcal{F}_\sigma$ be given fixed. Then we have

$$A = A \cap \{\sigma < \infty\} = \bigcup_{N=1}^{\infty} (A \cap \{\sigma \leq N\}) \in \bigcup_{N=1}^{\infty} \mathcal{F}_\sigma \cap \mathcal{F}_N = \bigcup_{N=1}^{\infty} \mathcal{F}_{\sigma \wedge N}.$$

As a result, there exists $N_0 \geq 1$ such that

$$A \in \mathcal{F}_{\sigma \wedge N_0} \subseteq \mathcal{F}_{\sigma \wedge N} \quad \forall N \geq N_0.$$

For any such N , from the property (2.14) and the definition of the conditional expectation, we have

$$\int_A e^{i\theta(B_{\sigma \wedge N+t} - B_{\sigma \wedge N})} d\mathbb{P} = e^{-\theta^2 t/2} \mathbb{P}(A).$$

By using the dominated convergence theorem, as $N \rightarrow \infty$ we obtain

$$\int_A e^{i\theta(B_{\sigma+t} - B_{\sigma})} d\mathbb{P} = e^{-\theta^2 t/2} \mathbb{P}(A).$$

The claim (2.13) thus follows as $A \in \mathcal{F}_{\sigma}$ is arbitrary. \square

Remark 2.4. The strong Markov property remains valid if the stopping time is only assumed to be finite almost surely. The same theorem also holds for a multidimensional Brownian motion.

An interesting application of the strong Markov property is the *reflection principle*, which is a quite useful tool for deriving deeper distributional properties of Brownian functionals.

Let B be a Brownian motion of which all sample paths are continuous. Let $\{\mathcal{F}_t^B\}$ be its natural filtration. Given $x \neq 0$, we define

$$\tau_x \triangleq \inf\{t \geq 0 : B_t = x\}$$

to be the first hitting time of the level x . From Proposition 1.4, we know that τ_x is an $\{\mathcal{F}_t^B\}$ -stopping time. The following unboundedness property of Brownian motion implies that τ_x is finite a.s.

Lemma 2.3. *With probability one,*

$$\sup_{t \geq 0} B_t = +\infty, \quad \inf_{t \geq 0} B_t = -\infty.$$

Proof. We only need to consider the supremum as the other case follows from the fact that $-B$ is a Brownian motion. Let $M \triangleq \sup_{t \geq 0} B_t$. For each $\lambda > 0$, the process $B^{(\lambda)} \triangleq \{\lambda^{-1} B_{\lambda^2 t} : t \geq 0\}$ is also a Brownian motion. Since

$$\lambda^{-1} \sup_{t \geq 0} B_t = \lambda^{-1} \sup_{t \geq 0} B_{\lambda^2 t} = \sup_{t \geq 0} B_t^{(\lambda)},$$

we find $\lambda^{-1} M \stackrel{d}{=} M$. As a result,

$$\begin{aligned} \mathbb{P}(M > \lambda) &= \mathbb{P}(M > 1) \quad \forall \lambda > 0 \quad \xrightarrow{\lambda \rightarrow \infty} \quad \mathbb{P}(M = \infty) = \mathbb{P}(M > 1), \\ \mathbb{P}(M \leq \lambda) &= \mathbb{P}(M \leq 1) \quad \forall \lambda > 0 \quad \xrightarrow{\lambda \rightarrow 0} \quad \mathbb{P}(M \leq 0) = \mathbb{P}(M \leq 1). \end{aligned}$$

Since $M \geq 0$ almost surely, we conclude that

$$\mathbb{P}(M = 0 \text{ or } \infty) = 1. \quad (2.15)$$

To exclude the possibility of $M = 0$, first note that

$$\mathbb{P}(M = 0) \leq \mathbb{P}(B_1 \leq 0, B_u \leq 0 \forall u \geq 1). \quad (2.16)$$

On the other hand, since $t \mapsto B_{1+t} - B_1$ is a Brownian motion, from (2.15) we know that

$$\sup_{t \geq 0} (B_{1+t} - B_1) = 0 \text{ or } \infty \text{ a.s.} \quad (2.17)$$

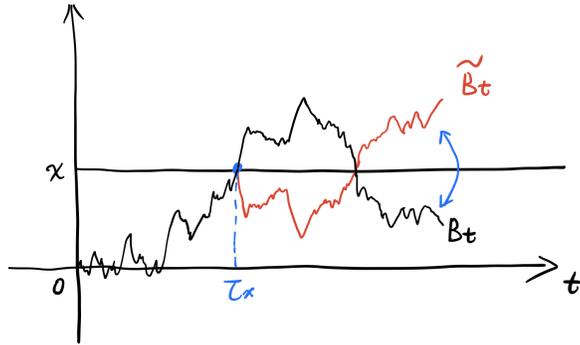
Under the occurrence of the event on the right hand side of (2.16), the second possibility in (2.17) is excluded and we thus obtain

$$\begin{aligned} \mathbb{P}(M = 0) &\leq \mathbb{P}(B_1 \leq 0, \sup_{t \geq 0} (B_{1+t} - B_1) = 0) \\ &= \mathbb{P}(B_1 \leq 0) \cdot \mathbb{P}(M = 0) \\ &= \frac{1}{2} \mathbb{P}(M = 0), \end{aligned}$$

where the first equality follows from the independence between B_1 and $\{B_{1+t} - B_1 : t \geq 0\}$. It follows that $\mathbb{P}(M = 0) = 0$ and hence $\mathbb{P}(M = \infty) = 1$. \square

Let us now define a ‘‘reflected’’ process \tilde{B} by

$$\tilde{B}_t \triangleq \begin{cases} B_t, & t < \tau_x; \\ 2x - B_t, & t \geq \tau_x. \end{cases} \quad (2.18)$$



The reflection principle of Brownian motion asserts that \tilde{B} is also a Brownian motion. The heuristic reason is simple to describe. Before the hitting time τ_x , one has $\tilde{B} = B$ being the original Brownian motion. After time τ_x , one can express B and \tilde{B} as

$$B_{\tau_x+t} = x + (B_{\tau_x+t} - x), \quad \tilde{B}_{\tau_x+t} = x - (B_{\tau_x+t} - x)$$

respectively. According to the strong Markov property, the process $\{B_{\tau_x+t} - x\}$ is a Brownian motion independent of the history up to time τ_x , and so is the process $\{-(B_{\tau_x+t} - x)\}$ by the reflection invariance of Brownian motion. As a result, the two processes B and \tilde{B} remain identically distributed after τ_x . Making this idea rigorous requires a little bit of extra effort.

Proposition 2.3. *The process \tilde{B} is a Brownian motion.*

Proof. Define two processes by

$$Y_t \triangleq B_t \mathbf{1}_{\{t \leq \tau_x\}}, \quad Z_t \triangleq (B_t - x) \mathbf{1}_{\{t > \tau_x\}}.$$

From the strong Markov property and the reflection invariance, we know that

$$(Y, \tau_x, Z) \stackrel{d}{=} (Y, \tau_x, -Z).$$

Next, let \mathcal{W} denote the space of all continuous paths and define a function

$$\Psi : \mathcal{W} \times [0, \infty) \times \mathcal{W} \rightarrow \mathcal{W}$$

by

$$\Psi(y, T, z)(t) \triangleq y_t \mathbf{1}_{\{t \leq T\}} + (x + z_t) \mathbf{1}_{\{t > T\}}, \quad t \geq 0.$$

One checks that

$$\Psi(Y, \tau_x, Z) = B, \quad \Psi(Y, \tau_x, -Z) = \tilde{B}.$$

Consequently, $B \stackrel{d}{=} \tilde{B}$. □

2.3 Passage time distributions

The next natural question is to understand the distribution of the hitting time τ_x . Hitting times are useful tools for studying the geometry of Markov processes and harmonic functions (potential theory), and they are often related to PDE problems. They also arise naturally in mathematical finance e.g. in the execution of an option when the asset price hits certain barrier.

Let B be a Brownian motion. Given $x > 0$, as before we define τ_x to be the first time that the process B hits the level x . By using a martingale method that is similar to the proof of the strong Markov property, one can derive the Laplace transform of τ_x .

Proposition 2.4. *The Laplace transform of τ_x is given by*

$$\mathbb{E}[e^{-\lambda\tau_x}] = e^{-x(\sqrt{2\lambda})}, \quad \lambda > 0. \quad (2.19)$$

Proof. Given fixed $\lambda > 0$, the process $t \mapsto M_t \triangleq \exp(\sqrt{2\lambda}B_t - \lambda t)$ is a martingale with respect to the natural filtration of B . By applying the optional sampling theorem to the martingale $\{M_t\}$ at the bounded stopping time $\tau_x \wedge n$, we have

$$\mathbb{E}[e^{\sqrt{2\lambda}B_{\tau_x \wedge n} - \lambda\tau_x \wedge n}] = 1 \quad \forall n \geq 1.$$

In addition, note that

$$e^{\sqrt{2\lambda}x - \lambda\tau_x} \mathbf{1}_{\{\tau_x < \infty\}} \xleftarrow{n \rightarrow \infty} e^{\sqrt{2\lambda}B_{\tau_x \wedge n} - \lambda\tau_x \wedge n} \leq e^{\sqrt{2\lambda}x}.$$

According to the dominated convergence theorem and the fact that τ_x is finite a.s., we obtain

$$\mathbb{E}[e^{\sqrt{2\lambda}x - \lambda\tau_x}] = 1,$$

which yields (2.19) after rearrangement. \square

A similar idea allows one to deal with the case of a double-barrier. Let $a < 0 < b$ be given fixed. Define τ_a, τ_b as the hitting time of the levels a, b respectively and set $\tau_{a,b} \triangleq \tau_a \wedge \tau_b$. Note that $\tau_{a,b}$ is the first time that the Brownian motion reaches the boundary of the interval $[a, b]$.

Proposition 2.5. *The Laplace transform of $\tau_{a,b}$ is given by:*

$$\mathbb{E}[e^{-\lambda\tau_{a,b}}] = \frac{\cosh((b+a)\sqrt{\lambda/2})}{\cosh((b-a)\sqrt{\lambda/2})}, \quad \lambda > 0, \quad (2.20)$$

where $\cosh x \triangleq \frac{e^x + e^{-x}}{2}$.

Proof. Let $\lambda > 0$ be given fixed. The main observation is that both of the processes

$$M_t \triangleq e^{\sqrt{2\lambda}B_t - \lambda t}, \quad N_t \triangleq e^{-\sqrt{2\lambda}B_t - \lambda t}$$

are martingales. Similarly to the proof of Proposition 2.4, by applying the optional sampling theorem to the martingales $\{M_t\}$ and $\{N_t\}$ respectively, we end up with

$$\mathbb{E}[e^{\sqrt{2\lambda}B_{\tau_{a,b}} - \lambda\tau_{a,b}}] = 1, \quad \mathbb{E}[e^{-\sqrt{2\lambda}B_{\tau_{a,b}} - \lambda\tau_{a,b}}] = 1 \quad (2.21)$$

The first identity in (2.21) implies

$$\mathbb{E}[e^{\sqrt{2\lambda}a - \lambda\tau_{a,b}} \mathbf{1}_{\{\tau_a < \tau_b\}}] + \mathbb{E}[e^{\sqrt{2\lambda}b - \lambda\tau_{a,b}} \mathbf{1}_{\{\tau_b < \tau_a\}}] = 1.$$

The second identity in (2.21) implies

$$\mathbb{E}[e^{-\sqrt{2\lambda}a - \lambda\tau_{a,b}} \mathbf{1}_{\{\tau_a < \tau_b\}}] + \mathbb{E}[e^{-\sqrt{2\lambda}b - \lambda\tau_{a,b}} \mathbf{1}_{\{\tau_b < \tau_a\}}] = 1.$$

If we set

$$x \triangleq \mathbb{E}[e^{-\lambda\tau_{a,b}} \mathbf{1}_{\{\tau_a < \tau_b\}}], \quad y \triangleq \mathbb{E}[e^{-\lambda\tau_{a,b}} \mathbf{1}_{\{\tau_b < \tau_a\}}],$$

the above two equations become the linear system

$$\begin{cases} e^{a\sqrt{2\lambda}}x + e^{b\sqrt{2\lambda}}y = 1, \\ e^{-a\sqrt{2\lambda}}x + e^{-b\sqrt{2\lambda}}y = 1. \end{cases}$$

By solving the system for x, y and simplifying the result, we are led to

$$\mathbb{E}[e^{-\lambda\tau_{a,b}}] = x + y = \frac{\cosh((b+a)\sqrt{\lambda/2})}{\cosh((b-a)\sqrt{\lambda/2})}.$$

□

In some situations, we would like to know the entire distribution rather than just the Laplace transform. Although an inversion formula for the Laplace transform is theoretically available, implementing it in practice is often quite difficult. In what follows, we use the reflection principle to compute the probability density function of the hitting time τ_x . The double-barrier case is much more involved and will not be discussed here.

Let $S_t = \max_{0 \leq s \leq t} B_s$ denote the running maximum of Brownian motion up to time t . We start by establishing a general formula for the joint distribution of (S_t, B_t) . The distribution of τ_x then follows easily.

Proposition 2.6. *For any $x, y \geq 0$, we have*

$$\mathbb{P}(S_t \geq x, B_t \leq x - y) = \mathbb{P}(B_t \geq x + y) = \frac{1}{\sqrt{2\pi}} \int_{\frac{x+y}{\sqrt{t}}}^{\infty} e^{-u^2/2} du. \quad (2.22)$$

In particular, the joint density of (S_t, B_t) is given by

$$\mathbb{P}(S_t \in dx, B_t \in dy) = \frac{2(2x - y)}{\sqrt{2\pi t^3}} e^{-\frac{(2x-y)^2}{2t}} dx dy, \quad x > 0, x > y, \quad (2.23)$$

and the density of τ_x ($x > 0$) is given by

$$\mathbb{P}(\tau_x \in dt) = \frac{x}{\sqrt{2\pi t^3}} e^{-\frac{x^2}{2t}} dt, \quad t > 0. \quad (2.24)$$

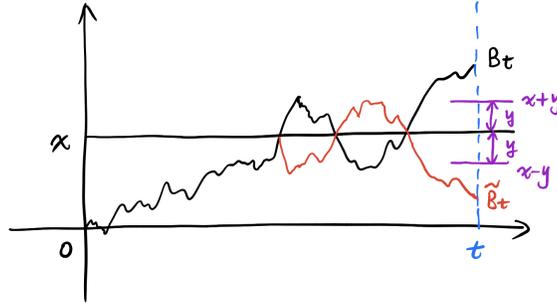
Proof. Let \tilde{B} be the reflected process with respect to the level x defined by (2.18), and we define the running maximum \tilde{S}_t of \tilde{B} accordingly. It is obvious that

$$\{S_t \geq x\} = \{\tilde{S}_t \geq x\} = \{\tau_x \leq t\}.$$

In addition, from the reflection principle (cf. Proposition 2.3), we know that \tilde{B} is also a Brownian motion. Therefore,

$$\begin{aligned} \mathbb{P}(S_t \geq x, B_t \leq x - y) &= \mathbb{P}(\tilde{S}_t \geq x, \tilde{B}_t \leq x - y) = \mathbb{P}(S_t \geq x, \tilde{B}_t \leq x - y) \\ &= \mathbb{P}(S_t \geq x, B_t \geq x + y) = \mathbb{P}(B_t \geq x + y), \end{aligned}$$

which yields the identity (2.22).



The claim (2.23) follows by differentiating the function $F(x, z) \triangleq \mathbb{P}(S_t \geq x, B_t \leq z)$, and (2.24) follows from the fact that

$$\begin{aligned} \mathbb{P}(\tau_x \leq t) &= \mathbb{P}(S_t \geq x) \\ &= \mathbb{P}(S_t \geq x, B_t \leq x) + \mathbb{P}(S_t \geq x, B_t > x) \\ &= \mathbb{P}(B_t \geq x + 0) + \mathbb{P}(B_t > x) \\ &= 2\mathbb{P}(B_t \geq x). \end{aligned}$$

□

Remark 2.5. From the relation

$$\mathbb{P}(S_t \geq x) = 2\mathbb{P}(B_t \geq x) = \mathbb{P}(B_t \geq x) + \mathbb{P}(B_t \leq -x) = \mathbb{P}(|B_t| \geq x),$$

one finds that $S_t \stackrel{\text{dist}}{=} |B_t|$. By explicit calculation based on the formula (2.23), one can also verify that

$$S_t - B_t \stackrel{\text{dist}}{=} |B_t|, \quad 2S_t - B_t \stackrel{\text{dist}}{=} R_t$$

for each fixed t , where $R_t \triangleq \sqrt{X_t^2 + Y_t^2 + Z_t^2}$ with (X, Y, Z) being a three-dimensional Brownian motion. A remarkable theorem of P. Lévy shows that

$$S - B \stackrel{\text{dist}}{=} |B|, \quad 2S - B \stackrel{\text{dist}}{=} R$$

as *stochastic processes* (cf. [16, Sec. VI.2]). These results are closely related to the study of local times and excursion theory.

2.4 Skorokhod's embedding theorem and Donsker's invariance principle

The Brownian motion can be constructed as the scaling limit of random walks. This fundamental result is known as *Donsker's invariance principle*. The main idea behind understanding this relation contains two parts: one first shows that a random walk can be embedded into a Brownian motion along a sequence of stopping times, and the convergence of the rescaled random walk towards the Brownian motion will then be a simple consequence of the continuity of Brownian sample paths. Using this perspective, one also recovers the classical central limit theorem as a byproduct!

We now make this idea more precise. Let F be a given fixed distribution on \mathbb{R} with mean zero and finite variance σ^2 .

Definition 2.3. A *random walk with step distribution* F is a random sequence given by $S_0 \triangleq 0$, and

$$S_n \triangleq X_1 + \cdots + X_n, \quad n \geq 1,$$

where $\{X_n : n \geq 1\}$ is an i.i.d. sequence of random variables with distribution F .

To think of the discrete random walk as an approximation of Brownian motion, one needs to turn it into a continuous process and rescale it according to the Brownian scaling property $\mathbb{E}[(B_t - B_s)^2] = t - s$. This motivates the following construction. Let $n \geq 1$ be given and we partition $[0, \infty)$ into the sub-intervals $[\frac{k-1}{n}, \frac{k}{n}]$ ($k \geq 1$). Define the *rescaled random walk* $S^{(n)} = \{S_t^{(n)} : t \geq 0\}$ to be the continuous process such that :

- (i) $S_{k/n}^{(n)} \triangleq S_k / (\sigma\sqrt{n})$ at each partition point k/n ;
- (ii) $S_t^{(n)}$ is linear within each sub-interval $[\frac{k-1}{n}, \frac{k}{n}]$.

Mathematically, we have

$$S_t^{(n)} = \frac{\sqrt{n}}{\sigma} \left(\left(\frac{k}{n} - t \right) S_{k-1} + \left(t - \frac{k-1}{n} \right) S_k \right), \quad t \in \left[\frac{k-1}{n}, \frac{k}{n} \right], k \geq 1. \quad (2.25)$$

This construction respects the Brownian scaling since

$$\mathbb{E}[(S_{k/n}^{(n)} - S_{(k-1)/n}^{(n)})^2] = \frac{1}{\sigma^2 n} \mathbb{E}[X_k^2] = \frac{1}{n}.$$

In vague terms, Donsker's invariance principle is stated as follows.

Theorem 2.3. *The sequence of continuous processes $S^{(n)} = \{S_t^{(n)} : t \geq 0\}$ “converges in distribution” to the Brownian motion $B = \{B_t : t \geq 0\}$ as $n \rightarrow \infty$.*

The precise mathematical meaning of the above distributional convergence requires the notion of weak convergence of probability measures and we will not elaborate it here.

2.4.1 The heuristic proof of Donsker's invariance principle

The core ingredient for proving Donsker's invariance principle is the following embedding theorem due to A. Skorokhod. Let $B = \{B_t : t \geq 0\}$ be a given Brownian motion and let $\{\mathcal{F}_t^B\}$ be its natural filtration.

Theorem 2.4 (Skorokhod's Embedding Theorem). *There exists an integrable $\{\mathcal{F}_t^B\}$ -stopping time τ , such that $B_\tau \stackrel{d}{=} F$ and $\mathbb{E}[\tau] = \sigma^2$.*

Let us now explain why Skorokhod's embedding theorem can be used to establish Donsker's invariance principle. We only outline the essential idea as the technical details require deeper tools from weak convergence theory.

The first point is the following result which allows one to embed the entire random walk into the Brownian motion. This is a consequence of the one-step embedding (Theorem 2.4) together with the strong Markov property of Brownian motion.

Corollary 2.1. *Let $\{S_n : n \geq 1\}$ be a random walk with step distribution F . Then there exists an increasing sequence $\{\tau_n : n \geq 1\}$ of integrable $\{\mathcal{F}_t^B\}$ -stopping times, such that $\{\tau_n - \tau_{n-1}\}$ are i.i.d. with mean σ^2 and $B_{\tau_n} \stackrel{\text{dist}}{=} S_n$ for every n .*

Proof. According to Skorokhod's embedding theorem, there exists an integrable $\{\mathcal{F}_t^B\}$ -stopping time τ_1 , such that $B_{\tau_1} \stackrel{d}{=} S_1$ and $\mathbb{E}[\tau_1] = \sigma^2$. On the other hand, from the strong Markov property we know that $t \mapsto B_t^{(\tau_1)} \triangleq B_{\tau_1+t} - B_{\tau_1}$ is a Brownian motion independent of $\mathcal{F}_{\tau_1}^B$. By applying Skorokhod's theorem again, we find an integrable $\{\mathcal{F}_t^{B^{(\tau_1)}}\}$ -stopping time τ'_2 ($\{\mathcal{F}_t^{B^{(\tau_1)}}\}$ denotes the natural filtration of $B^{(\tau_1)}$), such that $B_{\tau'_2}^{(\tau_1)} \stackrel{d}{=} F$ and $\mathbb{E}[\tau'_2] = \sigma^2$. Now we define $\tau_2 \triangleq \tau_1 + \tau'_2$. Since τ'_2 is constructed from the Brownian motion $B^{(\tau_1)}$, we see that $\tau_2 - \tau_1$ and τ_1 are i.i.d. In addition, we have

$$B_{\tau_2} = B_{\tau_1} + B_{\tau'_2}^{(\tau_1)} \stackrel{\text{dist}}{=} S_2.$$

The other τ_n 's are constructed in a similar inductive way. □

The other point is the continuity of Brownian sample paths. For simplicity let us assume that $\sigma^2 = 1$. Since we only concern with distributional properties, we may assume that the underlying random walk is defined by $S_n = B_{\tau_n}$ where $\{\tau_n\}$ is the sequence of stopping times given by Corollary 2.1. Let $S^{(n)} = \{S_t^{(n)} : t \geq 0\}$ be the corresponding rescaled version defined by (2.25). We also set $B^{(n)} \triangleq \{\frac{1}{\sqrt{n}}B_{nt} : t \geq 0\}$. Note that $B^{(n)}$ is again a Brownian motion. As a result, to establish Donsker's invariance principle it is enough to compare the distributions between $S^{(n)}$ and $B^{(n)}$.

Let us illustrate why $S_t^{(n)}$ and $B_t^{(n)}$ are closed to each other (given fixed t) as $n \rightarrow \infty$. Let k be the unique integer such that $t \in [\frac{k-1}{n}, \frac{k}{n})$. When n is large, we have $B_t^{(n)} \approx B_{\frac{k}{n}}^{(n)}$ and

$$S_t^{(n)} \approx S_{\frac{k}{n}}^{(n)} = \frac{1}{\sqrt{n}}S_k = \frac{1}{\sqrt{n}}B_{\tau_k} = B_{\frac{\tau_k}{n}}^{(n)}.$$

Now the crucial point is that

$$\frac{\tau_n}{n} = \frac{\tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_n - \tau_{n-1})}{n} \rightarrow \mathbb{E}[\tau_1] = 1$$

as a consequence of the strong law of large numbers since $\{\tau_n - \tau_{n-1}\}$ is an i.i.d. sequence. In particular, each summand $\frac{\tau_k - \tau_{k-1}}{n}$ is roughly equal to $\frac{1}{n}$ and we naturally expect that $\frac{\tau_k}{n} \approx \frac{k}{n}$ when n is large. Since Brownian sample paths are continuous, it follows that $B_{\tau_k/n}^{(n)} \approx B_{k/n}^{(n)}$ and thus $S_t^{(n)} \approx B_t^{(n)}$.

As a direct corollary of Donsker's invariance principle, one recovers the classical central limit theorem in the i.i.d. case.

Corollary 2.2. *Let $\{X_n : n \geq 1\}$ be an i.i.d. sequence with finite variance. Define $S_n \triangleq X_1 + \cdots + X_n$. Then*

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}[S_n]}} \xrightarrow{\text{dist}} N(0, 1) \quad \text{as } n \rightarrow \infty.$$

Proof. Define the rescaled random walk $S^{(n)}$ as before. It follows from Donsker's invariance principle that $S_1^{(n)} \xrightarrow{\text{dist}} B_1$ as $n \rightarrow \infty$, which is precisely the central limit theorem. \square

2.4.2 Constructing solutions to Skorokhod's embedding problem

The key missing piece in the previous discussion is the proof of Skorokhod's embedding theorem. We say that a stopping time τ is a *solution to Skorokhod's embedding problem for the distribution F* if it satisfies Theorem 2.4. As we will see, the approach we develop in what follows provides a tractable way of constructing τ explicitly.

The building block: two-point distributions

Solving Skorokhod's problem in full generality is quite challenging, but the starting observation is not hard. We begin by considering the simplest case when F is a *two-point distribution*, namely, F is the distribution of a random variable X that achieves only two values a and b ($a < 0 < b$). Since $\mathbb{E}[X] = 0$, the distribution of X is uniquely determined as

$$\mathbb{P}(X = a) = \frac{b}{b-a}, \quad \mathbb{P}(X = b) = \frac{-a}{b-a}. \quad (2.26)$$

In addition, we have $\mathbb{E}[X^2] = -ab$. In this case, a solution to Skorokhod's embedding problem can be constructed easily. Let us define

$$\tau_{a,b} \triangleq \inf\{t \geq 0 : B_t \notin (a, b)\}$$

to be the first time that B exists the interval (a, b) (cf. Section 2.3). Note that $\tau_{a,b}$ is an a.s. finite $\{\mathcal{F}_t^B\}$ -stopping time.

Proposition 2.7. *The stopping time $\tau_{a,b}$ gives a solution to Skorokhod's embedding problem for the two-point distribution (2.26):*

$$\mathbb{P}(B_{\tau_{a,b}} = a) = \frac{b}{b-a}, \quad \mathbb{P}(B_{\tau_{a,b}} = b) = \frac{-a}{b-a}, \quad \mathbb{E}[\tau_{a,b}] = -ab.$$

Proof. Note that both $\{B_t\}$ and $\{B_t^2 - t\}$ are $\{\mathcal{F}_t^B\}$ -martingales. By the optional sampling theorem, we have

$$\mathbb{E}[B_{\tau_{a,b} \wedge n}] = 0, \quad \mathbb{E}[B_{\tau_{a,b} \wedge n}^2] = \mathbb{E}[\tau_{a,b} \wedge n].$$

Since

$$|B_{\tau_{a,b} \wedge n}| \leq \max(|a|, |b|) \quad \forall n,$$

the dominated convergence theorem implies that

$$\mathbb{E}[B_{\tau_{a,b}}] = 0, \quad \mathbb{E}[B_{\tau_{a,b}}^2] = \mathbb{E}[\tau_{a,b}]. \quad (2.27)$$

On the other hand, by definition the random variable $B_{\tau_{a,b}}$ only achieves the two values a and b . The first identity in (2.27) uniquely identifies the distribution of $B_{\tau_{a,b}}$ as (2.26). The second identity in (2.27) shows that $\tau_{a,b}$ is integrable and $\mathbb{E}[\tau_{a,b}] = \mathbb{E}[X^2] = -ab$. \square

The binary splitting case

The main idea of solving Skorokhod's problem for a general distribution F is to approximate it by a binary splitting sequence so that one can essentially reduce the construction to the two-point case. To make it precise, we first introduce the following definition.

Definition 2.4. A sequence $\{X_n : n \geq 1\}$ of random variables is called *binary splitting* if for each $n \geq 1$, there exist a function $f_n : \mathbb{R} \times \{\pm 1\} \rightarrow \mathbb{R}$ and a $\{\pm 1\}$ -valued random variable D_n such that

$$X_n = f_n(X_{n-1}, D_n). \quad (2.28)$$

A binary splitting sequence $\{X_n\}$ is called a *binary splitting martingale* if it is a martingale with respect to its natural filtration.

When $n = 1$ the condition (2.28) reads $X_1 = f_1(D_1)$, and

$$X_2 = f_2(f_1(D_1), D_2), \quad X_3 = f_3(f_2(f_1(D_1), D_2), D_3) \text{ etc.}$$

In particular, X_1 takes (at most) two values, X_2 takes (at most) four values, X_3 takes (at most) eight values and so forth. If $\{X_n\}$ is a binary splitting martingale, the conditional distribution of X_n given (X_1, \dots, X_{n-1}) is supported on the two values $\{f_n(X_{n-1}, \pm 1)\}$ with mean

$$\mathbb{E}[X_n | X_1, \dots, X_{n-1}] = X_{n-1}.$$

This property uniquely determines the conditional distribution of $X_n | (X_1, \dots, X_{n-1})$. In this case, Skorokhod's embedding problem for the distribution of X_n can be solved explicitly: one uses the strong Markov property to recursively reduce the problem to the two-point case.

Lemma 2.4. *Let $\{X_n : n \geq 1\}$ be a binary splitting martingale such that $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] < \infty$ for every n . Then there exists an increasing sequence $\{\tau_n : n \geq 1\}$ of integrable, $\{\mathcal{F}_t^B\}$ -stopping times such that τ_n is a solution to Skorokhod's embedding problem for the distribution of X_n (i.e. $B_{\tau_n} \stackrel{d}{=} X_n$ and $\mathbb{E}[\tau_n] = \mathbb{E}[X_n^2]$) for every n .*

Proof. Since $X_1 = f_1(D_1)$ has a two-point distribution, the construction of τ_1 is contained in Proposition 2.7. Explicitly, we define

$$\tau_1 \triangleq \inf \{t \geq 0 : B_t \notin (f_1(-1), f_1(+1))\}.$$

Then $B_{\tau_1} \stackrel{\text{dist}}{=} X_1$ and $\mathbb{E}[\tau_1] = \mathbb{E}[X_1^2]$. Next, by the definition of binary splitting martingale, we know that the conditional distribution of X_2 given X_1 takes two values $\{f_2(X_1, \pm 1)\}$ with mean X_1 . In the meanwhile, we define

$$\tau_2 \triangleq \inf \{t \geq \tau_1 : B_t \notin (f_2(B_{\tau_1}, -1), f_2(B_{\tau_1}, +1))\}.$$

By the strong Markov property, the process $B^{(\tau_1)} \triangleq \{B_{\tau_1+t} - B_{\tau_1} : t \geq 0\}$ is a Brownian motion independent of $\mathcal{F}_{\tau_1}^B$. We now use Proposition 2.7 again to see that the conditional distribution of B_{τ_2} given B_{τ_1} takes two values $\{f_2(B_{\tau_1}, \pm 1)\}$ with mean B_{τ_1} . To put it in a simple form,

$$B_{\tau_2} |_{B_{\tau_1}} \stackrel{\text{dist}}{=} X_2 |_{X_1}.$$

Since $B_{\tau_1} \stackrel{\text{dist}}{=} X_1$, it follows that

$$(B_{\tau_1}, B_{\tau_2}) \stackrel{d}{=} (X_1, X_2). \quad (2.29)$$

To show $\mathbb{E}[\tau_2] = \mathbb{E}[X_2^2]$, we first note that

$$\begin{aligned} \mathbb{E}[\tau_2 - \tau_1] &= \mathbb{E}[(B_{\tau_2} - B_{\tau_1})^2] \quad (\text{by (2.27)}) \\ &= \mathbb{E}[(X_2 - X_1)^2] \quad (\text{by (2.29)}) \\ &= \mathbb{E}[X_2^2] + \mathbb{E}[X_1^2] - 2\mathbb{E}[X_1 X_2] \\ &= \mathbb{E}[X_2^2] + \mathbb{E}[X_1^2] - 2\mathbb{E}[X_1 \mathbb{E}[X_2 | X_1]] \\ &= \mathbb{E}[X_2^2] + \mathbb{E}[X_1^2] - 2\mathbb{E}[X_1^2] \quad (\text{martingale property}) \\ &= \mathbb{E}[X_2^2] - \mathbb{E}[X_1^2]. \end{aligned}$$

As a result,

$$\mathbb{E}[\tau_2] = \mathbb{E}[\tau_1] + \mathbb{E}[\tau_2 - \tau_1] = \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] - \mathbb{E}[X_1^2] = \mathbb{E}[X_2^2].$$

The general case can be treated by induction. We define

$$\tau_n \triangleq \inf \{t \geq \tau_{n-1} : B_t \notin (f_n(B_{\tau_{n-1}}, -1), f_n(B_{\tau_{n-1}}, +1))\}.$$

Then $B_{\tau_n} |_{(B_{\tau_1}, \dots, B_{\tau_{n-1}})}$ satisfies a two-point distribution supported on the values $\{f_n(B_{\tau_{n-1}}, \pm 1)\}$ with mean $B_{\tau_{n-1}}$. On the other hand, by definition the conditional distribution $X_n |_{(X_1, \dots, X_{n-1})}$ has exactly the same property. Therefore,

$$B_{\tau_n} |_{(B_{\tau_1}, \dots, B_{\tau_{n-1}})} \stackrel{d}{=} X_n |_{(X_1, \dots, X_{n-1})}.$$

Together with the induction hypothesis

$$(X_1, \dots, X_{n-1}) \stackrel{d}{=} (B_{\tau_1}, \dots, B_{\tau_{n-1}}),$$

we conclude that the same property holds for the n -tuple. By using the martingale property of $\{X_n\}$ in a similar way as before, we also have

$$\mathbb{E}[\tau_n] = \sum_{k=1}^n \mathbb{E}[\tau_k - \tau_{k-1}] = \sum_{k=1}^n (\mathbb{E}[X_k^2] - \mathbb{E}[X_{k-1}^2]) = \mathbb{E}[X_n^2].$$

□

Remark 2.6. For more general considerations, in the definition of a binary splitting martingale the function f_n is sometimes allowed to depend on the entire history (X_1, \dots, X_{n-1}) . We do not need this generality because in our case X_1, \dots, X_{n-2} are indeed uniquely determined by X_{n-1} .

Approximation by binary splitting martingales

Let us now return to the general situation where X is a given random variable with distribution F ($\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] < \infty$). The lemma below is the key step for solving Skorokhod's embedding problem for the distribution F .

Lemma 2.5. *There exists a binary splitting martingale $\{X_n : n \geq 1\}$ such that*

$$X_n \rightarrow X \text{ a.s. and in } L^2,$$

where convergence in L^2 means $\mathbb{E}[(X_n - X)^2] \rightarrow 0$ as $n \rightarrow \infty$.

We first explain how this approximation result leads to a proof of Theorem 2.4. Let $\{X_n\}$ be an approximating sequence given by the lemma and define the increasing sequence of stopping times $\{\tau_n\}$ according to Lemma 2.4. Set $\tau \triangleq \lim_{n \rightarrow \infty} \tau_n$. Since $\mathbb{E}[\tau_n] = \mathbb{E}[X_n^2]$ for all n and $X_n \rightarrow X$ in L^2 , by taking $n \rightarrow \infty$ we find $\mathbb{E}[\tau] = \mathbb{E}[X^2]$. This shows that $\tau < \infty$ a.s. and due to the continuity of Brownian sample paths we have

$$B_\tau = \lim_{n \rightarrow \infty} B_{\tau_n} \text{ a.s.}$$

Since $B_{\tau_n} \stackrel{d}{=} X_n$ for all n , we conclude that $B_\tau \stackrel{d}{=} X$. Consequently, τ is a solution to Skorokhod's embedding problem for X .

It now remains to prove Lemma 2.5. We only sketch the proof as the complete details require more technical considerations.

Sketch of the proof of Lemma 2.5. Suppose that the range of X is contained in an interval I and $0 = \mathbb{E}[X] \in I$. For any sub-interval $J \subseteq I$, we denote

$$c_J \triangleq \mathbb{E}[X | X \in J]$$

as the average of X given that it takes values in J (cf. 1.11). For instance, if X has a density f , then

$$c_J = \frac{\mathbb{E}[X \mathbf{1}_{\{X \in J\}}]}{\mathbb{P}(X \in J)} = \frac{\int_J x f(x) dx}{\int_J f(x) dx}.$$

We first construct X_1 as follows. Note that I is divided into two sub-intervals I_1, I_2 corresponding to the values of X that is below or above $0 = \mathbb{E}[X]$. We define X_1 to be the two-point random variable given by

$$X_1 \triangleq c_{I_1} \mathbf{1}_{\{X \in I_1\}} + c_{I_2} \mathbf{1}_{\{X \in I_2\}}.$$

Next we construct X_2 as follows. Note that the points c_{I_1}, c_{I_2} together with 0 further divide I into four sub-intervals J_1, J_2, J_3, J_4 . We define X_2 to be the four-point random variable given by

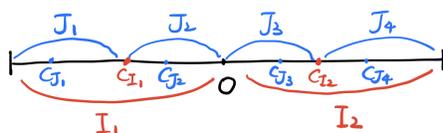
$$X_2 \triangleq c_{J_1} \mathbf{1}_{\{X \in J_1\}} + c_{J_2} \mathbf{1}_{\{X \in J_2\}} + c_{J_3} \mathbf{1}_{\{X \in J_3\}} + c_{J_4} \mathbf{1}_{\{X \in J_4\}}.$$

It can be checked that the conditional distribution of X_2 given X_1 is supported on two values with mean X_1 :

$$X_2|_{X_1=c_{I_1}} = X_2|_{X \in I_1} \in \{c_{J_1}, c_{J_2}\}, \quad \mathbb{E}[X_2|X_1 = c_{I_1}] = c_{I_1}$$

and

$$X_2|_{X_1=c_{I_2}} = X_2|_{X \in I_2} \in \{c_{J_3}, c_{J_4}\}, \quad \mathbb{E}[X_2|X_1 = c_{I_2}] = c_{I_2}.$$



The construction of X_3, X_4, \dots can be performed inductively in the same manner. In the n -th step, I is divided into 2^n sub-intervals K_1, \dots, K_{2^n} by all the possible values of X_1, X_2, \dots, X_{n-1} and the point 0. We define X_n by

$$X_n \triangleq \sum_{l=1}^{2^n} c_{K_l} \mathbf{1}_{\{X \in K_l\}}.$$

It is seen from the construction that X_1, \dots, X_{n-2} are determined by X_{n-1} and the conditional distribution of X_n given X_{n-1} is supported on two values.

This construction provides a natural approximation of X : when the range of X is partitioned into several sub-intervals say I_1, \dots, I_m , on each event $\{X \in I_l\}$ we use the corresponding conditional average value to approximate the actual value of X on this event. As a result, with no surprise one expects that X_n converges to X in a reasonable sense. What is less obvious is that $\{X_n\}$ is a binary splitting martingale. We let the curious reader explore the missing details. □

Example 2.1. Let X be a discrete uniform random variable on the four values $\{-2, -1, 1, 2\}$. Then X_1 is defined by

$$X_1 = c_1 \mathbf{1}_{\{X < 0\}} + c_2 \mathbf{1}_{\{X \geq 0\}} = c_1 \mathbf{1}_{\{X = -2 \text{ or } -1\}} + c_2 \mathbf{1}_{\{X = 1 \text{ or } 2\}},$$

where

$$c_1 = \mathbb{E}[X|X < 0] = -\frac{3}{2}, \quad c_2 = \mathbb{E}[X|X \geq 0] = \frac{3}{2}.$$

The next random variable X_2 is defined by

$$\begin{aligned} X_2 &= d_1 \mathbf{1}_{\{X < -3/2\}} + d_2 \mathbf{1}_{\{-3/2 \leq X < 0\}} + d_3 \mathbf{1}_{\{0 \leq X < 3/2\}} + d_4 \mathbf{1}_{\{X \geq 3/2\}} \\ &= d_1 \mathbf{1}_{\{X = -2\}} + d_2 \mathbf{1}_{\{X = -1\}} + d_3 \mathbf{1}_{\{X = 1\}} + d_4 \mathbf{1}_{\{X = 2\}} \\ &= (-2) \cdot \mathbf{1}_{\{X = -2\}} + (-1) \cdot \mathbf{1}_{\{X = -1\}} + 1 \cdot \mathbf{1}_{\{X = 1\}} + 2 \cdot \mathbf{1}_{\{X = 2\}} \\ &= X. \end{aligned}$$

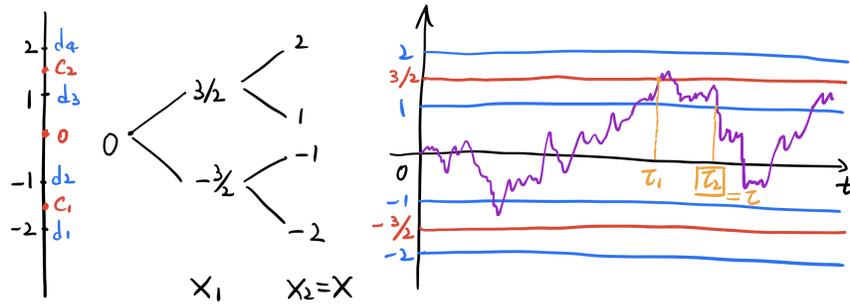
Note that

$$X_2|_{X_1 = -3/2} \in \{-2, -1\}, \quad X_2|_{X_1 = 3/2} \in \{1, 2\}.$$

Correspondingly,

$$\begin{aligned} \tau_1 &\triangleq \inf\{t \geq 0 : B_t \notin (-3/2, 3/2)\}, \\ \tau_2 &\triangleq \begin{cases} \inf\{t \geq \tau_1 : B_t \notin (-2, -1)\}, & \text{if } B_{\tau_1} = -3/2; \\ \inf\{t \geq \tau_1 : B_t \notin (1, 2)\}, & \text{if } B_{\tau_1} = 3/2. \end{cases} \end{aligned}$$

In this example, after two steps we have already recovered X . There is no need to go further as $X_n = X = X_2$ and $\tau_n = \tau_2$ for all $n \geq 3$. The stopping time τ_2 gives a solution to Skorokhod's problem for X .



Remark 2.7. If the distribution F is discrete and supported on finitely many points, the aforementioned construction of $\{X_n\}$ recovers X after finitely many steps (say n steps) and the corresponding τ_n gives a desired solution to Skorokhod's embedding problem for F .

Remark 2.8. If τ is an integrable $\{\mathcal{F}_t^B\}$ -stopping time τ , then B_τ is square integrable and

$$\mathbb{E}[B_\tau] = 0, \quad \mathbb{E}[B_\tau^2] = \mathbb{E}[\tau].$$

These are called *Wald's identities*. As a result, $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$ are necessary conditions for the existence of Skorokhod's embedding.

Remark 2.9. Skorokhod's embedding has important applications in mathematical finance (robust pricing and hedging). We refer the reader to [14] for an introduction as well as a discussion of other possible solutions to the embedding problem.

2.5 Erraticity of Brownian sample paths: an overview

The properties of Brownian motion we have dealt with so far are mostly distributional. On the other hand, sample path properties of Brownian motion is also a significant topic and has important implications on the development of stochastic calculus. In this section, we give an overview on what a generic Brownian sample path looks like. We do not provide proofs as they are quite technical and not so relevant to our main discussion. The essential point to have in mind is that:

Brownian sample paths are so irregular that ordinary differentiation/integration against Brownian motion fails in a fundamental way.

The goal of stochastic calculus is to provide a new approach for the differential calculus of Brownian motion.

In what follows, let $B = \{B_t : t \geq 0\}$ be a given fixed one-dimensional Brownian motion.

2.5.1 Irregularity

The following result provides some intuition towards how irregular Brownian sample paths can be (cf. [10, Sec. 2.9]).

Theorem 2.5. *With probability one, the following properties hold true.*

- (i) *the function $t \mapsto B_t(\omega)$ is nowhere differentiable.*
- (ii) *the set of local maximum/minimum points for the function $t \mapsto B_t(\omega)$ is countable and dense in $[0, \infty)$.*
- (iii) *the function $t \mapsto B_t(\omega)$ has no points of increase (t is a point of increase of a path $x : [0, \infty) \rightarrow \mathbb{R}$ if there exists $\delta > 0$ such that $x_s \leq x_t \leq x_u$ for all $s \in (t - \delta, t)$ and $u \in (t, t + \delta)$).*
- (iv) *for any given $x \in \mathbb{R}$, the level set $\{t \geq 0 : B_t(\omega) = x\}$ is closed, unbounded, has zero Lebesgue measure and does not contain isolated points.*

Remark 2.10. The first example of a continuous function that is nowhere differentiable was discovered by K. Weierstrass in the 1870s. Sample paths of stochastic processes provide rich examples of different kinds of pathological functions.

2.5.2 Oscillations

Hölder-continuity is a natural concept for describing the degree of oscillation for a given function. A function $f : I \rightarrow \mathbb{R}$ is said to be γ -Hölder continuous if

$$\sup_{s \neq t \in I} \frac{|f(t) - f(s)|}{|t - s|^\gamma} < \infty.$$

It is known that for any $0 < \gamma < 1/2$, with probability one every Brownian sample path is γ -Hölder continuous on any finite interval. Moreover, the Hölder-continuity fails when $\gamma = 1/2$ (cf. [16, Sec. I.2] for these results). To give a finer description on the precise rate of (local) oscillation for Brownian motion, one is led to the celebrated law of the iterated logarithm due to A. Khinchin (cf. [10, Sec. 2.9]).

Theorem 2.6 (Law of the Iterated Logarithm). *For every fixed $t \geq 0$, one has*

$$\mathbb{P}\left(\overline{\lim}_{h \downarrow 0} \frac{B_{t+h} - B_t}{\sqrt{2h \log \log 1/h}} = 1\right) = 1.$$

The law of iterated logarithm asserts that at a fixed time t , almost every sample path oscillates around t with order $\sqrt{2h \log \log 1/h}$. It is important to note that the underlying null set associated with this property *depends on t* . It is *not* true that Khinchin's law of the iterated logarithm holds uniformly in t with probability one. The uniform oscillation (the exact modulus of continuity) for Brownian sample paths was discovered by P. Lévy (cf. [10, Sec. 2.9]).

Theorem 2.7 (Lévy's Modulus of Continuity). *For each given $T > 0$, one has*

$$\overline{\lim}_{h \downarrow 0} \sup_{\substack{0 \leq s < t \leq T \\ t-s \leq h}} \frac{|B_t - B_s|}{\sqrt{2h \log 1/h}} = 1 \quad a.s.$$

The curious reader may wonder what the gap between Khinchin's law of the iterated logarithm and Lévy's modulus of continuity is. Indeed, to one's surprise the set of times at which Khinchin's law fails is not sparse at all: with probability one, the random set

$$\left\{t \in [0, 1] : \overline{\lim}_{h \downarrow 0} \frac{B_{t+h} - B_t}{\sqrt{2h \log 1/h}} = 1\right\}$$

is uncountable and dense in $[0, 1]$, and the random set

$$\left\{t \in [0, 1] : \overline{\lim}_{h \downarrow 0} \frac{B_{t+h} - B_t}{\sqrt{2h \log \log 1/h}} = \infty\right\}$$

has Hausdorff dimension one (cf. [15] for these deep results).

2.5.3 The p -variation of Brownian motion

Let $x : [0, \infty) \rightarrow \mathbb{R}$ be a continuous function. Given $p \geq 1$, the p -variation of x over $[s, t]$ is defined by the quantity

$$\|x\|_{p\text{-var};[s,t]} = \sup_{\mathcal{P}} \left(\sum_{t_k \in \mathcal{P}} |x_{t_k} - x_{t_{k-1}}|^p \right)^{1/p},$$

where the supremum is taken over all finite partitions \mathcal{P} of $[s, t]$. In particular, the 1-variation of x is its usual length. The classical Riemann-Stieltjes integration theory (cf. [1, Chap. 7]) and the semi-classical Young's integration theory (cf. [11, Chap. 1]) allows one to construct integrals of the form $\int_0^t y_s dx_s$ provided that the functions $x, y : [0, \infty) \rightarrow \mathbb{R}$ have finite p -variation with $1 \leq p < 2$. As a negative consequence, the following result (cf. [16, Sec. I.2]) destroys any hope for establishing differential calculus for Brownian motion from the classical viewpoint. We are thus led to the realm of stochastic calculus.

Theorem 2.8. *With probability one, Brownian sample paths have infinite p -variation on every finite interval for any $1 \leq p < 2$.*

Remark 2.11. By using the Hölder-continuity of Brownian sample paths, it is not hard to see that with probability one, Brownian sample paths have finite p -variation on every finite interval when $p > 2$. For the critical case of $p = 2$, one has $\|B\|_{2\text{-var};[s,t]} = \infty$ a.s. (cf. [6, Sec. 13.9]).

3 Stochastic integration

Let $B = \{B_t : t \geq 0\}$ be a one-dimensional Brownian motion defined on a given filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$. The primary goal of this chapter is to define the notion of stochastic integrals

$$\int_0^t \Phi_s dB_s \tag{3.1}$$

for a suitable class of stochastic processes Φ and to study their basic properties. Apart from its own interest, this is essential for the study of stochastic differential equations in the next chapter.

3.1 The essential structure: Itô's isometry

The underlying idea for constructing the integral (3.1) is natural: one considers a Riemann sum approximation and then pass to the limit in a reasonable way. To illustrate the main idea, let us assume that the integrand process Φ is $\{\mathcal{F}_t\}$ -adapted, has continuous sample paths and is uniformly bounded (i.e. $|\Phi_t(\omega)| \leq C$ with some $C > 0$ for all t, ω). We only work on the unit time interval $[0, 1]$ as there is no essential difference for a more general time horizon.

Suppose that

$$\mathcal{P} : 0 = t_0 < t_1 < \dots < t_{n-1} < t_n = 1$$

is a given finite partition of $[0, 1]$. It is natural to think of the Riemann sum

$$S_1^{\mathcal{P}} \triangleq \sum_{k=1}^n \Phi_{u_k} (B_{t_k} - B_{t_{k-1}})$$

as an approximation of $\int_0^1 \Phi_t dB_t$, where u_k is a given point in $[t_{k-1}, t_k]$ for each k . More generally, the Riemann sum

$$S_t^{\mathcal{P}} \triangleq \sum_{l=1}^{k-1} \Phi_{u_l} (B_{t_l} - B_{t_{l-1}}) + \Phi_{u_k} (B_t - B_{t_{k-1}}), \quad t \in [t_{k-1}, t_k]$$

is a natural candidate for an approximation of $\int_0^t \Phi_s dB_s$. Now the key point is that, unlike the definition of Riemann integrals u_k needs to be selected as the *left endpoint* of $[t_{k-1}, t_k]$, i.e. $u_k = t_{k-1}$. There are two basic reasons for doing so.

Reason 1. With the left endpoint selection, the process $\{S_t^{\mathcal{P}} : 0 \leq t \leq 1\}$ is an $\{\mathcal{F}_t\}$ -martingale. To see this, let us assume that $s < t$ and for simplicity that

$s, t \in [t_{k-1}, t_k]$ for some common k . By the definition of $S_t^{\mathcal{P}}$ (with $u_k \triangleq t_{k-1}$), one has

$$\begin{aligned}\mathbb{E}[S_t^{\mathcal{P}} | \mathcal{F}_s] &= \mathbb{E}\left[\sum_{l=1}^{k-1} \Phi_{t_{l-1}}(B_{t_l} - B_{t_{l-1}}) + \Phi_{t_{k-1}}(B_t - B_{t_{k-1}}) \middle| \mathcal{F}_s\right] \\ &= \sum_{l=1}^{k-1} \Phi_{t_{l-1}}(B_{t_l} - B_{t_{l-1}}) + \Phi_{t_{k-1}} \mathbb{E}[B_t - B_{t_{k-1}} | \mathcal{F}_s] \\ &= \sum_{l=1}^{k-1} \Phi_{t_{l-1}}(B_{t_l} - B_{t_{l-1}}) + \Phi_{t_{k-1}}(B_s - B_{t_{k-1}}) = S_s^{\mathcal{P}}.\end{aligned}$$

Therefore, $\{S_t^{\mathcal{P}}\}$ is a martingale. Combining with the next reason (Itô's isometry), the space of martingales provides a natural (Hilbert) structure on which one can investigate the convergence of $\{S_t^{\mathcal{P}}\}$ to obtain the definition of the integral process $t \mapsto \int_0^t \Phi_s dB_s$.

Before discussing the second reason, it is helpful to rewrite the definition of $S_t^{\mathcal{P}}$ in a slightly different way. With the left endpoint selection,

$$\Phi_t^{\mathcal{P}} \triangleq \sum_{k=1}^n \Phi_{t_{k-1}} \mathbf{1}_{(t_{k-1}, t_k]}(t)$$

is a natural approximation of the integrand $\{\Phi_t : 0 \leq t \leq 1\}$. Note that $\Phi^{\mathcal{P}}$ is a “simple” process in the sense that it is constant on each sub-interval $[t_{k-1}, t_k]$. In this case, the definition of the stochastic integral $\int_0^t \Phi_s^{\mathcal{P}} dB_s$ is obvious: one should define

$$\int_0^t \Phi_s^{\mathcal{P}} dB_s \triangleq \sum_{l=1}^{k-1} \Phi_{t_{l-1}}(B_{t_l} - B_{t_{l-1}}) + \Phi_{t_{k-1}}(B_t - B_{t_{k-1}}), \quad t \in [t_{k-1}, t_k].$$

Of course this is exactly $S_t^{\mathcal{P}}$ (with the left endpoint selection). In other words, the Riemann sum approximation can be viewed as the stochastic integral of a “simple” process.

Reason 2. The second reason for choosing the left endpoint, which is also a key feature of Itô's calculus, is the following so-called *Itô's isometry*:

$$\mathbb{E}\left[\left(\int_0^1 \Phi_t^{\mathcal{P}} dB_t\right)^2\right] = \mathbb{E}\left[\int_0^1 (\Phi_t^{\mathcal{P}})^2 dt\right]. \quad (3.2)$$

To see this, by the definition of $\Phi_t^{\mathcal{P}}$ one has

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^1 \Phi_t^{\mathcal{P}} dB_t\right)^2\right] &= \mathbb{E}\left[\left(\sum_{k=1}^n \Phi_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})\right)^2\right] \\ &= \sum_k \mathbb{E}\left[\Phi_{t_{k-1}}^2 (B_{t_k} - B_{t_{k-1}})^2\right] \\ &\quad + 2 \sum_{k < l} \mathbb{E}\left[\Phi_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})\Phi_{t_{l-1}}(B_{t_l} - B_{t_{l-1}})\right].\end{aligned}$$

By conditioning on $\mathcal{F}_{t_{k-1}}$, one finds that

$$\mathbb{E}\left[\Phi_{t_{k-1}}^2 (B_{t_k} - B_{t_{k-1}})^2\right] = \mathbb{E}\left[\Phi_{t_{k-1}}^2 \mathbb{E}[(B_{t_k} - B_{t_{k-1}})^2 | \mathcal{F}_{t_{k-1}}]\right] = \mathbb{E}[\Phi_{t_{k-1}}^2] \cdot (t_k - t_{k-1}).$$

In a similar way,

$$\mathbb{E}\left[\Phi_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})\Phi_{t_{l-1}}(B_{t_l} - B_{t_{l-1}})\right] = 0.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^1 \Phi_t^{\mathcal{P}} dB_t\right)^2\right] &= \sum_{k=1}^n \mathbb{E}[\Phi_{t_{k-1}}^2] (t_k - t_{k-1}) = \mathbb{E}\left[\sum_{k=1}^n \Phi_{t_{k-1}}^2 (t_k - t_{k-1})\right] \\ &= \mathbb{E}\left[\int_0^1 (\Phi_t^{\mathcal{P}})^2 dt\right].\end{aligned}$$

We now explain how Itô's isometry provides the essential structure for constructing the stochastic integral $\int_0^1 \Phi_t dB_t$ under the current assumption. For each $n \geq 1$, define $\Phi^{(n)} = \{\Phi_t^{(n)} : 0 \leq t \leq 1\}$ to be the left endpoint approximation of Φ with respect to the even partition of $[0, 1]$, i.e.

$$\Phi_t^{(n)} \triangleq \Phi_{\frac{k-1}{n}} \quad \text{for } t \in \left(\frac{k-1}{n}, \frac{k}{n}\right].$$

Recall that Φ has continuous sample paths. As a result, for each fixed (t, ω) one has

$$\Phi_t^{(n)}(\omega) \rightarrow \Phi_t(\omega) \quad \text{as } n \rightarrow \infty.$$

Since Φ is uniformly bounded, so is $\Phi^{(n)}$ as seen from its definition. According to the dominated convergence theorem, one concludes that

$$\mathbb{E}\left[\int_0^1 (\Phi_t^{(n)} - \Phi_t)^2 dt\right] \rightarrow 0 \tag{3.3}$$

as $n \rightarrow \infty$. In other words, the approximating sequence $\Phi^{(n)}$ converges to Φ in the sense of (3.3). Note that any convergent sequence must be *Cauchy*, in the sense that

$$\forall \varepsilon > 0, \exists N > 0 \text{ s.t. } \mathbb{E} \left[\int_0^1 (\Phi_t^{(n)} - \Phi_t^{(m)})^2 dt \right] < \varepsilon \quad \forall m, n > N.$$

According to Itô's isometry (3.2) for the “simple” process $\Phi^{(n)} - \Phi^{(m)}$, one has

$$\mathbb{E} \left[\left(\int_0^1 \Phi^{(n)} dB_t - \int_0^1 \Phi^{(m)} dB_t \right)^2 \right] = \mathbb{E} \left[\int_0^1 (\Phi_t^{(n)} - \Phi_t^{(m)})^2 dt \right].$$

As a consequence, the sequence of random variables $X_n \triangleq \int_0^1 \Phi_t^{(n)} dB_t$ is also a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ (the space of square integrable random variables). It is a standard fact from functional analysis that $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is *complete* under the metric

$$d(X, Y) \triangleq \sqrt{\mathbb{E}[(X - Y)^2]},$$

in the sense that every Cauchy sequence has a limit. Therefore, X_n converges to some $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ which will be taken as the definition of $\int_0^1 \Phi_t dB_t$. In the same way, for each fixed t , one can also define $\int_0^t \Phi_s dB_s$ as the L^2 -limit of $\int_0^t \Phi_s^{(n)} dB_s$.

The above argument outlines the essential idea behind the construction of the stochastic integral. However, there are several points that are not entirely satisfactory:

(i) Under the current assumption, it is an important fact that the stochastic integral $t \mapsto \int_0^t \Phi_s dB_s$ is a continuous martingale. However, this is not obvious at all from the above argument, although this point is clear when Φ is “simple” (e.g. when $\Phi = \Phi^P$). To establish this property carefully, one needs to prove convergence *at the process level* in a suitable space of martingales rather than just in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ for each fixed t .

(ii) The assumption that Φ is uniformly bounded is apparently too strong for many applications. This condition can largely be relaxed. As long as

$$\mathbb{E} \left[\int_0^1 \Phi_t^2 dt \right] < \infty, \tag{3.4}$$

one can still prove a lemma that there exists a sequence of “simple” processes $\Phi^{(n)}$ which converges to Φ in the sense of (3.3). After this, the argument for constructing the stochastic integral based on Itô's isometry is the same as before. However,

proving such a lemma is technically challenging (cf. Remark 3.3 below).

(iii) Even the integrability condition (3.4) itself may be too strong in some situations. In principle, we should be able to construct $\int_0^1 \Phi_t dB_t$ for any adapted process Φ with continuous sample paths. However, not all such processes satisfy (3.4) (e.g. $\Phi_t = e^{B_t^{100}}$). Is it possible to further relax the integrability condition (3.4) to allow a richer class of integrands?

There are several equivalent approaches to resolve the above points and give a complete construction of the stochastic integral in full generality. Among others, one elegant approach that contains the deeper insight into the martingale structure while at the same time avoids all the unpleasant technicalities of process approximation (Point (ii)) is given in [16, Sec. IV.2]. This approach relies on the method of Hilbert spaces and duality (the Riesz representation theorem). A less abstract approach that also respects Itô's isometry as the core structure can be found in [10, Sec. 3.2]. This approach develops the necessary technical points for resolving Points (i) and (ii). Both approaches require basic tools from Hilbert spaces. For relaxing the integrability condition (3.4), both of them make use of the notion of *local martingales*.

In what follows, we adopt an alternative but equivalent approach that was essentially due to McKean [13]. This approach is elementary in the sense that it does not involve any Hilbert space tools. In addition, one defines the integral $\int_0^t \Phi_s dB_s$ under a weaker integrability condition in one go without the need of localisation argument/local martingales. A price to pay is that the use of Itô's isometry becomes less transparent although this fact will be reproduced after the construction (cf. 3.5 below).

For those who are satisfied with the construction of $\int_0^t \Phi_s dB_s$ under the integrability condition (3.4) and do not need the more general construction in the next section, the following theorem summarises the essential properties of the stochastic integral (cf. Proposition 3.1 and Proposition 3.5 below).

Theorem 3.1. *Let $L^2(B)$ be the space of progressively measurable (cf. Definition 3.1 below) processes Φ that satisfy the integrability condition (3.4). Then the following statements hold true.*

(i) *The stochastic integral $\int_0^t \Phi_s dB_s$ is linear in Φ :*

$$\int_0^t (c\Phi_s + \Psi_s) dB_s = c \int_0^t \Phi_s dB_s + \int_0^t \Psi_s dB_s \quad \forall \Phi, \Psi \in L^2(B), c \in \mathbb{R}.$$

(ii) *With probability one, $t \mapsto \int_0^t \Phi_s dB_s$ is continuous.*

(iii) *The process $t \mapsto \int_0^t \Phi_s dB_s$ is a square integrable martingale and one has Itô's*

isometry:

$$\mathbb{E}\left[\left(\int_0^t \Phi_s dB_s\right)^2\right] = \mathbb{E}\left[\int_0^t \Phi_s^2 ds\right] \quad \forall t \in [0, 1].$$

3.2 General construction of stochastic integrals

Throughout this section, $B = \{B_t : t \geq 0\}$ is a one-dimensional $\{\mathcal{F}_t\}$ -Brownian motion defined on a given filtered probability space. We only work on the unit interval (there is no essential difference in the general case). More specifically, for a given stochastic process $\Phi = \{\Phi_t : t \in [0, 1]\}$, under suitable conditions we wish to define the integral

$$t \mapsto \int_0^t \Phi_s dB_s, \quad 0 \leq t \leq 1$$

as a stochastic process.

The conditions to be imposed on Φ contain two aspects: measurability and integrability.

Measurability. It is natural to expect that Φ should at least be adapted to the filtration $\{\mathcal{F}_t\}$. However, this simple condition turns out to be technically insufficient as one often needs extra measurability properties for the sample paths of Φ so that one can for instance consider the Lebesgue integral $\int_0^t \Phi_s(\omega) ds$ for each ω . The precise level of measurability is given by the following definition.

Definition 3.1. A stochastic process $\Phi = \{\Phi_t : t \geq 0\}$ is said to be $\{\mathcal{F}_t\}$ -progressively measurable if for every $t \geq 0$, the map

$$[0, t] \times \Omega \ni (s, \omega) \mapsto \Phi_s(\omega) \in \mathbb{R}$$

is measurable with respect to the product σ -algebra $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$.

To keep things simple, we will not mention and/or check this condition in any future context (all stochastic processes under consideration are either assumed or seen to be $\{\mathcal{F}_t\}$ -progressively measurable). It is a good exercise that if Φ is $\{\mathcal{F}_t\}$ -adapted and has continuous (or just right/left continuous) sample paths, then it is progressively measurable.

Integrability. In view of Itô's isometry (3.2), a natural integrability condition to be imposed on the integrand Φ is that

$$\mathbb{E}\left[\int_0^1 \Phi_t^2 dt\right] < \infty. \tag{3.5}$$

As pointed out in the last section (Point (iii)), such a condition may be too strong in several interesting situations. A reasonable integrability assumption for handling the most general case is given by

$$\int_0^1 \Phi_t^2 dt < \infty \quad \text{almost surely.} \quad (3.6)$$

For instance, (3.6) is satisfied for any process Φ with continuous sample paths. As we will see, under the stronger integrability condition (3.5) the process $\int_0^t \Phi_s dB_s$ is a martingale and satisfies Itô's isometry (cf. Proposition 3.5). However, under the weaker condition (3.6) these properties may fail even if the process $\int_0^t \Phi_s dB_s$ is integrable for all time! This is one of the deeper points in stochastic calculus and we will see a concrete counterexample in the context of stochastic differential equations (cf. Example 4.7 below).

In what follows, we develop the main steps for constructing the integral process $t \mapsto \int_0^t \Phi_s dB_s$ for any given $\Phi = \{\Phi_t : 0 \leq t \leq 1\}$ that is $\{\mathcal{F}_t\}$ -progressively measurable and satisfies the integrability condition (3.6). To emphasise the essential ideas, some technical details are omitted or presented with simplified assumptions. Providing all the fine details is a beneficial exercise for one who has basic training in real analysis and measure-theoretic probability.

3.2.1 Step one: integration of simple processes

Since we rely on an approximation idea, the first natural step is to construct the stochastic integral for simple integrands (i.e. step functions).

Definition 3.2. A process $\Phi = \{\Phi_t : 0 \leq t \leq 1\}$ is said to be *simple* if it can be expressed as

$$\Phi_t = \sum_{k=1}^n \xi_{k-1} \mathbf{1}_{(t_{k-1}, t_k]}(t), \quad (3.7)$$

where $0 = t_0 < t_1 < \dots < t_n = 1$ is a finite partition of $[0, 1]$, and ξ_{k-1} is a bounded, $\mathcal{F}_{t_{k-1}}$ -measurable random variable for each k . The class of simple processes is denoted as \mathcal{E} .

The definition of $\int_0^t \Phi_s dB_s$ for a simple process Φ is immediate.

Definition 3.3. Given a simple process Φ with representation (3.7), we define

$$\mathcal{I}(\Phi)_t \triangleq \int_0^t \Phi_s dB_s \triangleq \sum_{k=1}^n \xi_{k-1} (B_{t \wedge t_k} - B_{t \wedge t_{k-1}}), \quad 0 \leq t \leq 1.$$

From the definition, for $t \in (t_{k-1}, t_k]$ one has

$$\int_0^t \Phi_s dB_s = \sum_{l=1}^{k-1} \xi_{l-1} (B_{t_l} - B_{t_{l-1}}) + \xi_{k-1} (B_t - B_{t_{k-1}}).$$

Given $s < t$, we also define

$$\int_s^t \Phi_u dB_u \triangleq \int_0^t \Phi_u dB_u - \int_0^s \Phi_u dB_u.$$

The following properties are obvious from the definition.

Lemma 3.1. *The integral $\int_0^t \Phi_s dB_s$ does not depend on the specific representation of $\Phi \in \mathcal{E}$. In addition, the integral is linear in Φ (i.e. $\mathcal{I}(\Phi + \Psi)_t = \mathcal{I}(\Phi)_t + \mathcal{I}(\Psi)_t$) and has continuous sample paths.*

3.2.2 Step two: an exponential martingale property

Let $\Phi \in \mathcal{E}$. We introduce the exponential process

$$M_t \triangleq \exp \left(\int_0^t \Phi_s dB_s - \frac{1}{2} \int_0^t \Phi_s^2 ds \right).$$

The choice of its particular shape is due to the following result.

Lemma 3.2. *The process $\{M_t : 0 \leq t \leq 1\}$ is an $\{\mathcal{F}_t\}$ -martingale.*

Proof. Suppose that Φ is given by (3.7). Let $s < t$. We only discuss the case when $s, t \in (t_{k-1}, t_k]$ for some common k . In this case,

$$\int_s^t \Phi_u dB_u - \frac{1}{2} \int_s^t \Phi_u^2 du = \xi_{k-1} (B_t - B_s) - \frac{1}{2} \xi_{k-1}^2 (t - s),$$

and we have

$$\mathbb{E}[M_t | \mathcal{F}_s] = M_s \cdot \mathbb{E} \left[\exp \left(\xi_{k-1} (B_t - B_s) - \frac{1}{2} \xi_{k-1}^2 (t - s) \right) | \mathcal{F}_s \right].$$

Since $\xi_{k-1} \in \mathcal{F}_{t_{k-1}} \subseteq \mathcal{F}_s$ and $B_t - B_s$ is independent of \mathcal{F}_s , we find that

$$\mathbb{E} \left[\exp \left(\xi_{k-1} (B_t - B_s) \right) | \mathcal{F}_s \right] = \exp \left(\frac{1}{2} \xi_{k-1}^2 (t - s) \right).$$

Therefore,

$$\mathbb{E}[M_t | \mathcal{F}_s] = M_s.$$

If s, t belong to difference sub-intervals, say $s \in (t_{l-1}, t_l]$ and $t \in (t_{k-1}, t_k]$, one can argue in a similar way by conditioning on $\mathcal{F}_{t_{k-1}}, \mathcal{F}_{t_{k-2}}, \dots, \mathcal{F}_{t_l}, \mathcal{F}_s$ recursively. \square

A crucial fact about the exponential martingale $\{M_t\}$ we shall use is the following maximal inequality.

Lemma 3.3. *For any $\alpha, \beta > 0$, we have*

$$\mathbb{P}\left(\sup_{0 \leq t \leq 1} \left(\int_0^t \Phi_s dB_s - \frac{\alpha}{2} \int_0^t \Phi_s^2 ds \right) > \beta\right) \leq e^{-\alpha\beta}.$$

Proof. Define

$$M_t^\alpha \triangleq \exp\left(\alpha \int_0^t \Phi_s dB_s - \frac{\alpha^2}{2} \int_0^t \Phi_s^2 ds\right).$$

Lemma 3.2 shows that $\{M_t^\alpha\}$ is a martingale (with constant expectation one). According to Doob's maximal inequality (cf. Theorem 1.6),

$$\begin{aligned} & \mathbb{P}\left(\sup_{0 \leq t \leq 1} \left(\int_0^t \Phi_s dB_s - \frac{\alpha}{2} \int_0^t \Phi_s^2 ds \right) > \beta\right) \\ &= \mathbb{P}\left(\sup_{0 \leq t \leq 1} M_t^\alpha > e^{\alpha\beta}\right) \leq e^{-\alpha\beta} \mathbb{E}[M_1^\alpha] = e^{-\alpha\beta}. \end{aligned}$$

□

Remark 3.1. The consideration of such an exponential process is of fundamental importance in many problems (we have seen this in the discussion of Brownian passage time distributions). At the moment its martingale property comes from explicit calculation based on the distribution of Brownian motion. The deeper reason behind the calculation will be clearer from the viewpoint of Itô's formula as well as stochastic differential equations.

3.2.3 Step three: a uniform estimate for integrals of simple processes

Recall that a sequence of events $\{A_n : n \geq 1\}$ *eventually happens with probability one* if

$$\mathbb{P}(\exists N \text{ s.t. } A_n \text{ happens for all } n > N) = 1.$$

Note that the number N appearing in the above probability may depend on ω . If a property P_n (depending on n) eventually happens with probability one, we simply write

$$P_n \text{ eventually w.p.1.}$$

Lemma 3.4. Let $\{\Phi^{(n)} : n \geq 1\}$ be a sequence of simple processes such that

$$\int_0^1 (\Phi_t^{(n)})^2 dt \leq 2^{-n} \quad \text{eventually w.p.1.}$$

Then for any $\theta > 1$, we have

$$\sup_{0 \leq t \leq 1} \left| \int_0^t \Phi_s^{(n)} dB_s \right| \leq \left(\frac{1}{2} + \theta \right) 2^{-n/2} \sqrt{\log n} \quad \text{eventually w.p.1.} \quad (3.8)$$

Proof. For each given n , let us choose $\alpha_n \triangleq 2^{n/2} \sqrt{\log n}$ and $\beta_n \triangleq \theta 2^{-n/2} \sqrt{\log n}$. By Lemma 3.3, we have

$$\mathbb{P} \left(\sup_{0 \leq t \leq 1} \left(\int_0^t \Phi_s^{(n)} dB_s - \frac{\alpha_n}{2} \int_0^t (\Phi_s^{(n)})^2 ds \right) > \beta_n \right) \leq e^{-\alpha_n \beta_n} = n^{-\theta}.$$

Since $\theta > 1$, the above probability is summable over n and from the first Borel-Cantelli's lemma (cf. Theorem A.1 (i) in Appendix (5)) we have

$$\sup_{0 \leq t \leq 1} \left(\int_0^t \Phi_s^{(n)} dB_s - \frac{\alpha_n}{2} \int_0^t (\Phi_s^{(n)})^2 ds \right) \leq \beta_n \quad \text{eventually w.p.1.}$$

This implies that

$$\begin{aligned} \int_0^t \Phi_s^{(n)} dB_s &\leq \frac{\alpha_n}{2} \int_0^t (\Phi_s^{(n)})^2 ds + \beta_n \\ &\leq \frac{1}{2} 2^{n/2} \sqrt{\log n} \cdot \int_0^1 (\Phi_s^{(n)})^2 ds + \theta 2^{-n/2} \sqrt{\log n} \\ &\leq \left(\frac{1}{2} + \theta \right) 2^{-n/2} \sqrt{\log n} \quad \forall t \end{aligned}$$

eventually with probability one. By considering $-\Phi^{(n)}$ at the same time and using the simple fact that

$$\sup_x |f(x)| = \max \left\{ \sup_x f(x), \sup_x (-f(x)) \right\},$$

we conclude that

$$\left| \int_0^t \Phi_s^{(n)} dB_s \right| \leq \left(\frac{1}{2} + \theta \right) 2^{-n/2} \sqrt{\log n} \quad \forall t$$

eventually with probability one. □

Remark 3.2. The factors $(\frac{1}{2} + \theta)$ and $\sqrt{\log n}$ are of no importance. The crucial point is that a suitable choice of α_n, β_n allows one to use the first Borel-Cantelli's lemma and at the same time the right hand side of (3.8) should define a convergent series.

For conciseness, from now on we shall rephrase Lemma 3.4 as

$$\begin{aligned} \int_0^1 (\Phi_t^{(n)})^2 dt &\lesssim 2^{-n} \quad \text{eventually w.p.1.} \\ \implies \sup_{0 \leq t \leq 1} \left| \int_0^t \Phi_s^{(n)} dB_s \right| &\lesssim 2^{-n/2} \quad \text{eventually w.p.1.} \end{aligned}$$

The symbol \lesssim means the possibility of allowing some extra factor (e.g. $\sqrt{\log n}$ in the lemma) whose precise value is insignificant and does not affect the convergence of the relevant series.

3.2.4 Step four: approximation by simple processes

Let $\Phi = \{\Phi_t : 0 \leq t \leq 1\}$ be an $\{\mathcal{F}_t\}$ -progressively measurable process such that

$$\int_0^1 \Phi_t^2 dt < \infty \quad \text{a.s.}$$

The key for the construction of the stochastic integral $\int_0^t \Phi_s dB_s$ is the following lemma.

Lemma 3.5. *There exists a sequence $\{\Phi^{(n)} : n \geq 1\}$ of simple processes, such that*

$$\int_0^1 (\Phi_t^{(n)} - \Phi_t)^2 dt \leq 2^{-n} \quad \text{eventually w.p.1.} \quad (3.9)$$

Proof. For the sake of simplicity, we again assume that Φ is uniformly bounded (i.e. $|\Phi_t(\omega)| \leq C$ for all t, ω) and has continuous sample paths. For each $m \geq 1$, we partition $[0, 1]$ into m sub-intervals of equal length and consider the left endpoint approximation

$$\Psi_t^{(m)} \triangleq \Phi_{\frac{k-1}{m}} \quad \text{if } t \in \left(\frac{k-1}{m}, \frac{k}{m}\right]. \quad (3.10)$$

Then $\Psi^{(m)}$ is a simple process. As in the proof of (3.3), one sees that

$$\int_0^1 (\Psi_t^{(m)}(\omega) - \Phi_t(\omega))^2 dt \rightarrow 0 \quad \text{as } m \rightarrow \infty \quad (3.11)$$

for every ω . In particular,

$$\int_0^1 (\Psi_t^{(m)} - \Phi_t)^2 dt \rightarrow 0 \quad \text{in probability}$$

as $m \rightarrow \infty$. As a result, for each $n \geq 1$, there exists m_n such that

$$\mathbb{P}\left(\int_0^1 (\Psi_t^{(m_n)} - \Phi_t)^2 dt > 2^{-n}\right) \leq n^{-2}.$$

It follows from the first Borel-Cantelli lemma that

$$\int_0^1 (\Psi_t^{(m_n)} - \Phi_t)^2 dt \leq 2^{-n} \quad \text{eventually w.p.1.}$$

The sequence $\Phi^{(n)} \triangleq \Psi^{(m_n)}$ is desired. \square

Remark 3.3. The above case (Φ being continuous, uniformly bounded) has indeed been treated in Section 3.1 under the stronger integrability condition. The main challenging situation is when Φ is a general progressively measurable process under the weaker integrability condition. To construct $\Phi^{(n)}$ in this case, one first approximates Φ by processes with continuous sample paths and then uses the left endpoint approximation. The former part requires the following real analysis fact. Let $f : [0, 1] \rightarrow \mathbb{R}$ be a square integrable function (extend f to \mathbb{R} by setting $f(t) = 0$ when $t \notin [0, 1]$). For each $h > 0$, define the function

$$f^{(h)}(t) \triangleq \frac{1}{h} \int_{t-h}^t f(s) ds, \quad t \in [0, 1].$$

Then f_h is continuous and one can show that

$$\lim_{h \rightarrow 0} \int_0^1 (f^{(h)}(t) - f(t))^2 dt = 0.$$

Using this idea in our stochastic context, given Φ we first define $\Phi^{(h)}$ in the above way. Then $\Phi^{(h)}$ has continuous sample paths and approximates Φ in the above L^2 -sense. Next, since the definition of simple processes requires uniform boundedness, we can introduce the truncation

$$\Phi_t^{(h,N)} \triangleq \begin{cases} N, & \text{if } \Phi_t^{(h)} > N; \\ \Phi_t^{(h)}, & \text{if } -N \leq \Phi_t^{(h)} \leq N; \\ -N, & \text{if } \Phi_t^{(h)} < -N \end{cases}$$

before discretising it into a step function. Finally, given a finite partition \mathcal{P} of $[0, 1]$ we define the step process approximation $\Phi^{(h, N, \mathcal{P})}$ of $\Phi^{(h, N)}$ by (3.10). Then with probability one,

$$\lim_{h \rightarrow 0} \lim_{N \rightarrow \infty} \lim_{\text{mesh}(\mathcal{P}) \rightarrow 0} \int_0^1 (\Phi_t^{(h, N, \mathcal{P})} - \Phi_t)^2 dt = 0.$$

From this point on, the argument of extracting a subsequence $\Phi^{(n)}$ is not so different from the previous proof. We let the reader think about the fine details.

3.2.5 Step five: completing the construction

To see how one can define $\mathcal{I}(\Phi)_t = \int_0^t \Phi_s dB_s$ from Lemma 3.5, let $\{\Phi^{(n)}\}$ be a simple sequence satisfying (3.9). Then

$$\int_0^1 (\Phi_t^{(n+1)} - \Phi_t^{(n)})^2 dt \lesssim 2^{-n} \quad \text{eventually w.p.1.}$$

According to Lemma 3.4,

$$\sup_{0 \leq t \leq 1} |\mathcal{I}(\Phi^{(n+1)})_t - \mathcal{I}(\Phi^{(n)})_t| \lesssim 2^{-n} \quad \text{eventually w.p.1.}$$

Since the series of 2^{-n} is convergent, with probability one the sequence of continuous functions

$$[0, 1] \ni t \mapsto \mathcal{I}(\Phi^{(n)})_t \quad (n \geq 1)$$

form a Cauchy sequence with respect to the uniform distance. As a result, with probability one it has a well-defined uniform limit $\mathcal{I}(\Phi)$:

$$\mathbb{P}(\mathcal{I}(\Phi^{(n)})_t \text{ converges to } \mathcal{I}(\Phi)_t \text{ uniformly on } [0, 1]) = 1. \quad (3.12)$$

It remains to see that the limiting process $\{\mathcal{I}(\Phi)_t\}$ is independent of the choice of the approximating sequence $\{\Phi^{(n)}\}$. To simplify notation, we denote

$$\|f\| \triangleq \left(\int_0^1 f(t)^2 dt \right)^{1/2}, \quad \|f\|_\infty \triangleq \sup_{0 \leq t \leq 1} |f(t)|$$

for any generic function f .

Lemma 3.6. *Let $\{\Psi^{(n)} : n \geq 1\}$ be a sequence of simple processes such that*

$$\|\Psi^{(n)} - \Phi\|_2 \rightarrow 0 \quad \text{in probability.}$$

Then

$$\|\mathcal{I}(\Psi^{(n)}) - \mathcal{I}(\Phi)\|_\infty \rightarrow 0 \quad \text{in probability.} \quad (3.13)$$

Proof. It suffices to show that every subsequence of $\{\Psi^{(n)}\}$ contains a further subsequence along which (3.13) holds. Without loss of generality, it is enough to show that $\{\Psi^{(n)}\}$ itself contains such a subsequence. To this end, first note that

$$\|\Psi^{(n)} - \Phi^{(n)}\|_2 \rightarrow 0 \quad \text{in probability}$$

where $\{\Phi^{(n)}\}$ is defined in Lemma 3.5. Similar to the proof of that lemma, one finds a subsequence $\{k_n\}$ such that

$$\int_0^1 (\Psi_t^{(k_n)} - \Phi_t^{(k_n)})^2 dt \leq 2^{-n} \quad \text{eventually w.p.1.}$$

Lemma 3.4 then implies that

$$\|\mathcal{I}(\Psi^{(k_n)}) - \mathcal{I}(\Phi^{(k_n)})\|_\infty \lesssim 2^{-n} \quad \text{eventually w.p.1.}$$

It follows that

$$\|\mathcal{I}(\Psi^{(k_n)}) - \mathcal{I}(\Phi)\|_\infty \rightarrow 0 \quad \text{a.s.}$$

since $\mathcal{I}(\Phi^{(k_n)})$ satisfies this property (cf. (3.12)). □

Definition 3.4. A stochastic process $\Phi = \{\Phi_t : t \geq 0\}$ is said to be *Itô integrable* on $[0, 1]$ if it is an $\{\mathcal{F}_t\}$ -progressively measurable process and satisfies

$$\int_0^1 \Phi_t^2 dt < \infty \quad \text{a.s.}$$

Given an Itô integrable process Φ , the stochastic process $\mathcal{I}(\Phi)$ constructed in the above steps is called the *stochastic integral* (also known as the *Itô integral*) of Φ . It is often denoted as $\mathcal{I}(\Phi)_t = \int_0^t \Phi_s dB_s$.

Since continuous functions are bounded on finite intervals, any adapted process with continuous sample paths is Itô integrable. Lemma 3.6 shows that the stochastic integral is well-defined for any Itô integrable process. The same lemma also gives its linearity as it is true for simple processes (cf. Lemma 3.1).

Proposition 3.1. *The stochastic integral is linear:*

$$\mathcal{I}(\Phi + \Psi)_t = \mathcal{I}(\Phi)_t + \mathcal{I}(\Psi)_t, \quad \mathcal{I}(c\Phi)_t = c\mathcal{I}(\Phi)_t. \quad (3.14)$$

In addition, $\mathcal{I}(\Phi)$ has continuous sample paths.

The following result is a generalisation of Lemma 3.6. Unfortunately its proof is rather technical and non-inspiring (the idea is standard though: approximate Φ by simple processes and extract subsequences in a careful way). As a result we do not give the proof here.

Proposition 3.2. *Let $\Phi^{(n)}, \Phi$ ($n \geq 1$) be Itô integrable processes on $[0, 1]$. Suppose that*

$$\|\Phi^{(n)} - \Phi\|_2 \rightarrow 0 \quad \text{in prob.}$$

Then

$$\|\mathcal{I}(\Phi^{(n)}) - \mathcal{I}(\Phi)\|_\infty \rightarrow 0 \quad \text{in prob.}$$

As a useful corollary, one can justify the left endpoint Riemann sum approximation for the stochastic integral that is mentioned at the beginning of the chapter.

Corollary 3.1. *Let Φ be an Itô integrable process on $[0, 1]$ with left continuous and bounded sample paths (not necessarily uniform in ω). Given a partition \mathcal{P} of $[0, 1]$, let $\Phi^{(\mathcal{P})}$ denote the associated left endpoint approximation of Φ with respect to the partition \mathcal{P} . Then*

$$\|\mathcal{I}(\Phi^{(\mathcal{P})}) - \mathcal{I}(\Phi)\|_\infty \rightarrow 0 \quad \text{in prob.}$$

as the mesh size of the partition tends to zero.

Proof. Since Φ has left continuous sample paths, we have the pointwise convergence

$$\lim_{\text{mesh}(\mathcal{P}) \rightarrow 0} \Phi_t^{(\mathcal{P})}(\omega) = \Phi_t(\omega) \quad \forall (t, \omega) \in [0, 1] \times \Omega.$$

Since Φ also has bounded sample paths, the dominated convergence theorem implies that

$$\|\Phi^{(\mathcal{P})} - \Phi\|_2 \rightarrow 0 \quad \text{in prob.}$$

The claim thus follows from Proposition 3.2. □

Iterated integrals of Brownian motion: Part I

Before investigating deeper properties of stochastic integrals, we use one simple example to illustrate how the principle of stochastic integration differs from ordinary integration.

Consider the stochastic integral $\int_0^t B_s dB_s$. One may naively expect that this integral is equal to $\frac{1}{2}B_1^2$ from the perspective of ordinary calculus. This is indeed not true! To understand this, let

$$\mathcal{P} : 0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$$

be a given partition of $[0, t]$. Then one can write

$$\begin{aligned} B_t^2 &= \left(\sum_{k=1}^n (B_{t_k} - B_{t_{k-1}}) \right)^2 \\ &= \sum_{k=1}^n (B_{t_k} - B_{t_{k-1}})^2 + 2 \sum_{1 \leq i < j \leq n} (B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}}) \\ &= \sum_{k=1}^n (B_{t_k} - B_{t_{k-1}})^2 + 2 \sum_{j=1}^n B_{t_{j-1}} (B_{t_j} - B_{t_{j-1}}). \end{aligned}$$

The second term converges in probability to $2 \int_0^t B_t dB_t$ as a consequence of Corollary 3.1. However, the first term does not converge to zero! This is not too surprising if one keeps the relation

$$(B_{t_k} - B_{t_{k-1}})^2 \approx (t_k - t_{k-1})$$

in mind, which heuristically implies that

$$\sum_{k=1}^n (B_{t_k} - B_{t_{k-1}})^2 \approx \sum_{k=1}^n (t_k - t_{k-1}) = t_n - t_0 = t.$$

The following result justifies this property precisely. Recall that X_n converges to X in L^2 if

$$\mathbb{E}[|X_n - X|^2] \rightarrow 0$$

as $n \rightarrow \infty$.

Proposition 3.3. *Let $t > 0$ be fixed and let $\mathcal{P} = \{t_k\}_{k=0}^n$ be an arbitrary finite partition of $[0, t]$. Then*

$$\sum_k (B_{t_k} - B_{t_{k-1}})^2 \rightarrow t \quad \text{in } L^2$$

as the mesh size of \mathcal{P} tends to zero.

Proof. To simplify notation, we write $\Delta_k B \triangleq B_{t_k} - B_{t_{k-1}}$ and $\Delta_k t \triangleq t_k - t_{k-1}$. Then we have

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_k (\Delta_k B)^2 - t\right)^2\right] &= \mathbb{E}\left[\left(\sum_k ((\Delta_k B)^2 - \Delta_k t)\right)^2\right] \\
&= \sum_k \mathbb{E}\left[\left((\Delta_k B)^2 - \Delta_k t\right)^2\right] \\
&\quad + \sum_{i \neq j} \mathbb{E}\left[\left((\Delta_i B)^2 - \Delta_i t\right)\left((\Delta_j B)^2 - \Delta_j t\right)\right]. \\
&= \sum_k \left(\mathbb{E}[(\Delta_k B)^4] - 2\Delta_k t \cdot \mathbb{E}[(\Delta_k B)^2] + (\Delta_k t)^2\right) + 0 \\
&= 2 \sum_k (\Delta_k t)^2 \leq 2(\max_k \Delta_k t) \cdot \sum_k \Delta_k t \\
&= 2t \cdot \text{mesh}\mathcal{P}.
\end{aligned}$$

Therefore, the left hand side converges to zero as $\text{mesh}\mathcal{P} \rightarrow 0$. \square

Proposition 3.3 asserts the existence of non-zero *quadratic variation* for the Brownian motion. As a result of Proposition 3.3, one finds that

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

By writing it in a formal differential form, one has

$$dB_t^2 = 2B_t dB_t + dt. \tag{3.16}$$

From this simple example, one can already see that Itô's calculus differs from ordinary calculus (which normally reads $dx^2 = 2x dx$) by a second order term. The fundamental reason behind this new phenomenon is the existence of non-zero quadratic variation for the Brownian motion. This point will be clearer when we study Itô's formula in Section 3.4 below.

One can think of $\int_0^t B_s dB_s$ itself as an integrand and further integrate against the Brownian motion. By similar type of calculation, one finds that

$$\int_0^1 \left(\int_0^t B_s dB_s\right) dB_t = \frac{B_1^3 - 3B_1}{6}. \tag{3.17}$$

For higher order iterated integrals, it is not realistic to perform explicit calculation and one needs a more systematic method to evaluate them. As we will see in Section 4.3 below, these iterated Itô integrals are naturally related to the classical Hermite polynomials.

3.3 Martingale properties

Let us recapture what we have obtained so far. If $\Phi = \{\Phi_t : t \geq 0\}$ is an $\{\mathcal{F}_t\}$ -progressively measurable process such that with probability one

$$\int_0^t \Phi_s^2 ds < \infty \quad \forall t \geq 0,$$

then the stochastic integral $\mathcal{I}(\Phi)_t = \int_0^t \Phi_s dB_s$ ($t \geq 0$) is a well-defined $\{\mathcal{F}_t\}$ -adapted process with continuous sample paths, and the map $\Phi \mapsto \mathcal{I}(\Phi)$ is linear. To understand deeper properties of the stochastic integral, it is essential to look at it from the martingale perspective.

3.3.1 Square integrable martingales and the bracket process

We begin by introducing some general notions. The differential form (3.16) arises from the formal relation that $(dB_t)^2 = dt$. Another way to look at this property (as well as Proposition 3.3) is that the function $A_t \triangleq t$ makes the process.

One can ask the following more general question. Let $M = \{M_t, \mathcal{F}_t : t \geq 0\}$ be a continuous martingale. What is $(dM_t)^2$? This question is of fundamental importance for developing a general theory of differential calculus with respect to the martingale M . Inspired by the Brownian motion case, one should look for an increasing process A_t such that $M_t^2 - A_t$ is a martingale. The precise formulation of this property is the content of the *Doob-Meyer decomposition theorem*. We assume further that M is square integrable, i.e. $\mathbb{E}[M_t^2] < \infty$ for all t . The theorem asserts that one can extract a martingale part from M_t^2 and what is left is a pathwisely increasing process.

Theorem 3.2 (The Doob-Meyer Decomposition Theorem). *There exists a unique $\{\mathcal{F}_t\}$ -adapted stochastic process, denoted as $\langle M \rangle = \{\langle M \rangle_t : t \geq 0\}$, such that:*

- (i) $\langle M \rangle_0 = 0$ a.s.;
- (ii) $\mathbb{E}[\langle M \rangle_t] < \infty$ for all t ;
- (iii) with probability one, the sample paths of $\langle M \rangle$ are continuous and non-decreasing;
- (iv) the process $\{M_t^2 - \langle M \rangle_t, \mathcal{F}_t : t \geq 0\}$ is a martingale.

Remark 3.4. The proof of this theorem is rather involved and we refer the serious reader to [10, Sec. 1.4]. It is though easy to comprehend the discrete-time situation. Let $\{X_n, \mathcal{F}_n : n \geq 0\}$ be a submartingale ($X_n = M_n^2$ in the above context). We claim that there exists a unique $\{\mathcal{F}_n\}$ -predictable sequence $A = \{A_n : n \geq 0\}$

(i.e. $A_n \in \mathcal{F}_{n-1}$) satisfying Properties (i)–(iv) (stated in discrete-time in the obvious way). Indeed, if such a process A exists, by letting $Y_n \triangleq X_n - A_n$ one has

$$X_k - X_{k-1} = Y_k - Y_{k-1} + A_k - A_{k-1}.$$

Since $\{Y_n\}$ is a martingale and $\{A_n\}$ is predictable, by taking conditional expectation with respect to \mathcal{F}_{k-1} , one finds

$$A_k - A_{k-1} = \mathbb{E}[X_k - X_{k-1} | \mathcal{F}_{k-1}].$$

Summing over k from 1 to n , it follows that A_n has to be given by

$$A_n = \sum_{k=1}^n \mathbb{E}[X_k - X_{k-1} | \mathcal{F}_{k-1}]. \quad (3.18)$$

This not only shows the uniqueness of $\{A_n\}$ but also gives its existence defined explicitly by (3.18). Checking the desired properties of $\{A_n\}$ is routine.

Definition 3.5. The process $\langle M \rangle$ given by Theorem 3.2 is called the *quadratic variation process* of the martingale M .

Example 3.1. For the Brownian motion $\{B_t\}$, since $\{B_t^2 - t\}$ is a martingale we see that $\langle B \rangle_t = t$. The equation (3.15) suggests that this martingale can be expressed as a stochastic integral $2 \int_0^t B_s dB_s$.

Remark 3.5. As an extension of the relation $(dB_t)^2 = dt$, the differential calculus for M_t should respect the relation $(dM_t)^2 = d\langle M \rangle_t$. For instance, the analogue of the equation (3.15) should be given by

$$M_t^2 = 2 \int_0^t M_s dM_s + \langle M \rangle_t.$$

This formula also indicates what the martingale $M_t^2 - \langle M \rangle_t$ is: it is the stochastic integral $2 \int_0^t M_s dM_s$. A more general discussion is given in Section 3.5 below.

By using a standard idea of *polarisation*, one can generalise the notion of quadratic variation to the situation involving two different martingales. Let M, N be two continuous, square integrable, $\{\mathcal{F}_t\}$ -martingales. By the definition of the quadratic variation process, both of the following processes

$$(M + N)^2 - \langle M + N \rangle^2, \quad (M - N)^2 - \langle M - N \rangle^2$$

are martingales. In addition, note that

$$M_t N_t = \frac{(M_t + N_t)^2 - (M_t - N_t)^2}{4}.$$

As a result, if we define the process

$$\langle M, N \rangle_t \triangleq \frac{\langle M + N \rangle_t - \langle M - N \rangle_t}{4},$$

then $MN - \langle M, N \rangle$ is a martingale.

Definition 3.6. The process $\langle M, N \rangle$ is called the *bracket process* between the two martingales M, N .

Remark 3.6. Similar to the quadratic variation, it can be shown that the bracket process is the unique $\{\mathcal{F}_t\}$ -adapted process $A = \{A_t : t \geq 0\}$ satisfying the following properties:

- (i) $A_0 = 0$ a.s.;
- (ii) $\mathbb{E}[|A_t|] < \infty$ for all t ;
- (iii) with probability one, every sample path of A is continuous and has bounded total variation;
- (iv) $MN - A$ is an $\{\mathcal{F}_t\}$ -martingale.

The bracket process is a generalisation of the quadratic variation since $\langle M, M \rangle = \langle M \rangle$. Using this uniqueness property of $\langle M, N \rangle$, one can show that the bracket process behaves like a symmetric bilinear form.

Proposition 3.4. *The bracket process is symmetric and bilinear in (M, N) , i.e.*

$$\langle M, N \rangle = \langle N, M \rangle, \quad \langle aM_1 + M_2, N \rangle = a\langle M_1, N \rangle + \langle M_2, N \rangle.$$

Proof. We only verify $\langle M_1 + M_2, N \rangle = \langle M_1, N \rangle + \langle M_2, N \rangle$. This is a consequence of the fact that

$$(M_1 + M_2)N - \langle M_1, N \rangle - \langle M_2, N \rangle$$

is a martingale and the uniqueness of the bracket process stated in Remark 3.6. \square

Example 3.2. Suppose that X, Y are two independent $\{\mathcal{F}_t\}$ -Brownian motions. Then $\langle X, Y \rangle \equiv 0$. Indeed, given $s < t$ we have

$$\begin{aligned} \mathbb{E}[X_t Y_t | \mathcal{F}_s] &= \mathbb{E}[(X_t - X_s + X_s)(Y_t - Y_s + Y_s) | \mathcal{F}_s] \\ &= \mathbb{E}[(X_t - X_s)(Y_t - Y_s) | \mathcal{F}_s] + X_s \cdot \mathbb{E}[Y_t - Y_s | \mathcal{F}_s] \\ &\quad + Y_s \mathbb{E}[X_t - X_s | \mathcal{F}_s] + X_s Y_s \\ &= X_s Y_s. \end{aligned}$$

This shows that XY itself is a martingale and thus $\langle X, Y \rangle \equiv 0$.

Remark 3.7. The terminology of quadratic variation is justified by the fact that

$$\sum_{t_k \in \mathcal{P}} (M_{t_k} - M_{t_{k-1}})^2 \rightarrow \langle M \rangle_t \quad \text{in prob.} \quad (3.19)$$

for any fixed t and any finite partition of $[0, t]$ whose mesh size tends to zero (cf. Proposition 3.3 for the Brownian motion case). By using polarisation, (3.19) further implies that

$$\sum_{t_k \in \mathcal{P}} (M_{t_k} - M_{t_{k-1}})(N_{t_k} - N_{t_{k-1}}) \rightarrow \langle M, N \rangle_t \quad \text{in prob.}$$

For this reason the bracket process is also known as the *cross-variation process*. We do not prove these facts here and refer the reader to [16, Sec. IV.1].

3.3.2 Stochastic integrals as martingales

When the integrand $\Phi = \{\Phi_t : t \geq 0\}$ is a simple process, by definition it is uniformly bounded on each finite interval $[0, t]$ (as the ξ_{k-1} 's are assumed to be bounded). In particular, we know that

$$\mathbb{E}\left[\int_0^t \Phi_s^2 ds\right] < \infty \quad \forall t \geq 0. \quad (3.20)$$

On the other hand, one can prove the following fact.

Lemma 3.7. *Let $\Phi = \{\Phi_t : t \geq 0\}$ be a simple process. Then $\{\mathcal{I}(\Phi)_t\}$ is a square integrable martingale whose quadratic variation is given by*

$$\langle \mathcal{I}(\Phi) \rangle_t = \int_0^t \Phi_s^2 ds.$$

Proof. Let Φ be represented as (3.7). By performing the same type of explicit calculation as in Lemma 3.2, one shows that both $\{\mathcal{I}(\Phi)_t\}$ and $\{\mathcal{I}(\Phi)_t^2 - \int_0^t \Phi_s^2 ds\}$ are martingales. \square

It is interesting to ask whether a stochastic integral $\int_0^t \Phi_s dB_s$ is always a martingale in general? Of course it needs to be integrable to talk about the martingale property. However, even if we know/assume that $\int_0^t \Phi_s dB_s$ is integrable, this process may *fail* to be a martingale in general! This is counterintuitive in view of Lemma 3.7 and obtaining a counterexample is not an easy exercise (cf. Example 4.7 below).

Nonetheless, if we assume the stronger integrability condition (3.20), no surprise will be expected and we do have the nice martingale properties.

Proposition 3.5. *Suppose that $\Phi = \{\Phi_t : t \geq 0\}$ satisfies the condition (3.20). Then $\{\mathcal{I}(\Phi)_t\}$ is a square integrable martingale whose quadratic variation is given by*

$$\langle \mathcal{I}(\Phi) \rangle_t = \int_0^t \Phi_s^2 ds. \quad (3.21)$$

In particular, one has Itô's isometry

$$\mathbb{E}\left[\left(\int_0^t \Phi_s dB_s\right)^2\right] = \mathbb{E}\left[\int_0^t \Phi_s^2 ds\right]. \quad (3.22)$$

Proof. Let us just work on $[0, 1]$. The main technical point is that when choosing the approximating sequence $\Phi^{(n)}$ in the construction of the integral (cf. Lemma 3.5), if the condition (3.20) holds one can further require that

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[\int_0^1 (\Phi_t^{(n)} - \Phi_t)^2 dt\right] = 0. \quad (3.23)$$

This is already seen under the assumption that Φ has continuous sample paths and is uniformly bounded (cf. (3.3)). The general case is technically involved but is dealt with along the lines of Remark 3.3.

Under the extra property (3.23) for the approximating sequence, let us show that $\{\mathcal{I}(\Phi)_t\}$ is a martingale. Let $s < t$ and $A \in \mathcal{F}_s$. We need to check that

$$\mathbb{E}\left[\left(\int_0^t \Phi_u dB_u\right) \mathbf{1}_A\right] = \mathbb{E}\left[\left(\int_0^s \Phi_u dB_u\right) \mathbf{1}_A\right]. \quad (3.24)$$

Since $\{\mathcal{I}(\Phi^{(n)})_t\}$ is a martingale, the above property is true for $\Phi^{(n)}$. To pass to the limit, it suffices to show that

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[\left(\int_0^t (\Phi_u^{(n)} - \Phi_u) dB_u\right) \mathbf{1}_A\right] = 0. \quad (3.25)$$

For this purpose, we first use the following estimate as a consequence of the Cauchy-Schwarz inequality:

$$\left|\mathbb{E}\left[\left(\int_0^t (\Phi_u^{(n)} - \Phi_u) dB_u\right) \mathbf{1}_A\right]\right| \leq \sqrt{\mathbb{E}\left[\left(\int_0^t (\Phi_u^{(n)} - \Phi_u) dB_u\right)^2\right]}.$$

The right hand side converges to zero since

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^t (\Phi_u^{(n)} - \Phi_u)dB_u\right)^2\right] &\leq \varliminf_{m \rightarrow \infty} \mathbb{E}\left[\left(\int_0^t (\Phi_u^{(n)} - \Phi_u^{(m)})dB_u\right)^2\right] \quad (\text{Fatou's lemma}) \\ &= \varliminf_{m \rightarrow \infty} \mathbb{E}\left[\int_0^t (\Phi_u^{(n)} - \Phi_u^{(m)})^2 du\right] \\ &= \mathbb{E}\left[\int_0^t (\Phi_u^{(n)} - \Phi_u)^2 du\right], \end{aligned}$$

which tends to zero due to the choice of $\Phi^{(n)}$. Therefore, (3.25) holds and the martingale property (3.24) thus follows. A similar argument shows that $\mathcal{I}(\Phi)^2 - \int_0^t \Phi_s^2 ds$ is a martingale, hence yielding the last part of the proposition. \square

Due to the linearity of the stochastic integral and the usual Lebesgue integral, the property (3.21) implies that $\mathcal{I}(\Phi)\mathcal{I}(\Psi) - \int_0^t \Phi_t\Psi_t dt$ is a martingale for any Φ, Ψ satisfying the integrability condition (3.20). Therefore,

$$\langle \mathcal{I}(\Phi), \mathcal{I}(\Psi) \rangle_t = \int_0^t \Phi_s \Psi_s ds.$$

Remark 3.8. When Φ is Itô integrable without the stronger integrability condition (3.20), it is still possible to show that $t \mapsto \int_0^t \Phi_s^2 ds$ is the quadratic variation process of the stochastic integral $\int_0^t \Phi_s dB_s$, in the sense that

$$\sum_{t_k \in \mathcal{P}} \left(\int_{t_{k-1}}^{t_k} \Phi_s dB_s \right)^2 \rightarrow \int_0^t \Phi_s^2 ds$$

for any fixed t and any finite partition of $[0, t]$ whose mesh size tends to zero. Similarly, $\int_0^t \Phi_s \Psi_s ds$ is the cross-variation process of the two integrals $\mathcal{I}(\Phi), \mathcal{I}(\Psi)$ (cf. Remark 3.7). The best way to understand these properties (as well as Remark 3.7) is to put them in the general context of *local martingales*, which is beyond the scope of the current study (cf. [10, Sec. 1.5&3.2.D]).

3.4 Itô's formula

The special example of (3.15) can be rewritten compactly as

$$dB_t^2 = 2B_t dB_t + dt.$$

As we have pointed out before, this is different from the rule of ordinary calculus ($dx^2 = 2x dx$) due to the presence of the extra term dt . The occurrence of this

second order term is a consequence of the relation $(dB_t)^2 = dt$ which is considered as a non-negligible first order infinitesimal. This phenomenon does not exist in ordinary calculus as $(dx)^2$ is a negligible quantity comparing with the differential dx . It is a prominent feature of stochastic calculus, and as we have pointed out before the fundamental reason behind it is the existence of quadratic variation for the Brownian motion (cf. Proposition 3.3).

With this principle in mind, at a heuristic level it is not hard to derive what the general rule of stochastic calculus should look like. Let $f : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ be a given smooth function. From the formal Taylor expansion, we have

$$\begin{aligned} df(t, B_t) &= \frac{\partial f}{\partial t}(t, B_t)dt + \frac{\partial f}{\partial x}(t, B_t)dB_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, B_t)(dB_t)^2 \\ &\quad + \frac{\partial^2 f}{\partial t \partial x}(t, B_t)dB_t dt + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(t, B_t)(dt)^2 \\ &\quad + \frac{1}{3!}\frac{\partial^3 f}{\partial x^3}(dB_t)^3 + \frac{3}{3!}\frac{\partial^3 f}{\partial x^2 \partial t}(dB_t)^2 dt + \dots \end{aligned}$$

The key principle is the relation that $(dB_t)^2 = dt$ and all those terms of order higher than dt (e.g. $dB_t dt$, $(dB_t)^3 = dB_t dt$, $(dB_t)^2 dt = (dt)^2$ etc.) are considered negligible. As a result, we arrive at

$$df(t, B_t) = \partial_t f(t, B_t)dt + \partial_x f(t, B_t)dB_t + \frac{1}{2}\partial_x^2 f(t, B_t)dt.$$

This leads us to the renowned *Itô's formula*. We say that $f : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ is a $C^{1,2}$ -function if it is continuously differentiable in the time variable and twice continuously differentiable in the space variable.

Theorem 3.3 (Itô's formula for Brownian motion). *Let B be a one-dimensional Brownian motion. Let $f : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ be a $C^{1,2}$ -function. Then for any $s < t$, one has*

$$f(t, B_t) = f(s, B_s) + \int_s^t \partial_t f(u, B_u)du + \int_s^t \partial_x f(u, B_u)dB_u + \frac{1}{2} \int_s^t \partial_x^2 f(u, B_u)du.$$

Proof. For simplicity, we assume that $s = 0$, f does not depend on time and has bounded derivatives up to order three (the result is true under the original assumption but the proof is more technical). Let $\mathcal{P} = \{t_k\}_{k=0}^n$ be a given partition of $[0, t]$ and to simplify notation we define

$$\Delta_k B \triangleq B_{t_k} - B_{t_{k-1}}, \quad \Delta_k t \triangleq t_k - t_{k-1}$$

It follows from the third order Taylor expansion theorem that

$$\begin{aligned}
f(B_t) - f(B_0) &= \sum_{k=1}^n (f(B_{t_k}) - f(B_{t_{k-1}})) \\
&= \sum_{k=1}^n f'(B_{t_{k-1}}) \Delta_k B + \frac{1}{2} \sum_{k=1}^n f''(B_{t_{k-1}}) (\Delta_k B)^2 \\
&\quad + \frac{1}{6} \sum_{k=1}^n f^{(3)}(\xi_k) (\Delta_k B)^3
\end{aligned} \tag{3.26}$$

where $\xi_k \in [B_{t_{k-1}}, B_{t_k}]$.

In the first place, according to Corollary 3.1 the first summation in (3.26) converges to the stochastic integral $\int_0^t f'(B_s) dB_s$ as the partition mesh tends to zero. In addition, the last summation satisfies

$$\left| \sum_{k=1}^n f^{(3)}(\xi_k) (\Delta_k B)^3 \right| \leq C \sum_{k=1}^n |B_{t_k} - B_{t_{k-1}}|^3 \leq C \max_{1 \leq k \leq n} |B_{t_k} - B_{t_{k-1}}| \cdot \sum_{k=1}^n (B_{t_k} - B_{t_{k-1}})^2$$

where $C = \|f^{(3)}\|_\infty$. The term $\max_k |B_{t_k} - B_{t_{k-1}}|$ converges to zero a.s. as partition mesh tends to zero due to the continuity of Brownian sample paths. The last quadratic summation converges to t in probability by Proposition 3.3. Therefore, the error term

$$\sum_{k=1}^n f^{(3)}(\xi_k) (\Delta_k B)^3 \rightarrow 0 \quad \text{in prob.}$$

It remains to investigate the second summation in (3.26). We claim that

$$\sum_{k=1}^n f''(B_{t_{k-1}}) (B_{t_k} - B_{t_{k-1}})^2 \rightarrow \int_0^t f''(B_s) ds \quad \text{in prob.}$$

From Riemann integration theory, one has

$$\sum_{k=1}^n f''(B_{t_{k-1}}) (t_k - t_{k-1}) \rightarrow \int_0^t f''(B_s) ds.$$

As a result, it suffices to show that

$$M_{\mathcal{P}} \triangleq \sum_{k=1}^n f''(B_{t_{k-1}}) ((\Delta^k B)^2 - \Delta^k t) \rightarrow 0 \quad \text{in prob.} \tag{3.27}$$

as partition mesh tends to zero. To this end, note that

$$\begin{aligned}\mathbb{E}[M_{\mathcal{P}}^2] &= \sum_{k=1}^n \mathbb{E}[f''(B_{t_{k-1}})^2((\Delta^k B)^2 - \Delta^k t)^2] \\ &\quad + 2 \sum_{k < l} \mathbb{E}[f''(B_{t_{k-1}})f''(B_{t_{l-1}})((\Delta^k B)^2 - \Delta^k t)((\Delta^l B)^2 - \Delta^l t)].\end{aligned}$$

Every term in the second summation is equal to zero as seen by conditioning on $\mathcal{F}_{t_{l-1}}$. Since f'' is assumed to be uniformly bounded, it follows that

$$\mathbb{E}[M_{\mathcal{P}}^2] \leq C_1 \sum_{k=1}^n \mathbb{E}[(\Delta^k B)^2 - \Delta^k t]^2 = C_2 \sum_{k=1}^n (\Delta^k t)^2 \leq C_2 t \cdot \text{mesh}\mathcal{P} \rightarrow 0$$

where C_1, C_2 are suitable constants whose values are of no importance. Therefore, (3.27) holds.

The result thus follows by putting the above facts together and sending partition mesh to zero in (3.26). \square

Extension to Itô processes

Theorem 3.3 can be viewed as the chain rule for the composition of a function with a Brownian motion. For more interesting applications, we need to generalise the formula to cover the situation of composition with stochastic integrals.

Definition 3.7. A stochastic process $\{X_t : t \geq 0\}$ is called a (one-dimensional) *Itô process* if it can be written as

$$X_t = X_0 + \int_0^t \Phi_s dB_s + \int_0^t \Psi_s ds,$$

where B is a Brownian motion, Φ is an Itô integrable process and Ψ is a progressively measurable process such that with probability one $\int_0^t |\Psi_s| ds < \infty$ for all t .

To motivate Itô's formula in this context, let

$$X_t^i = X_0^i + \int_0^t \Phi_s^i dB_s + \int_0^t \Psi_s^i ds, \quad 1 \leq i \leq n \quad (3.28)$$

be n given Itô processes (with respect to the same Brownian motion). Let $f : [0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}$ be a given smooth function. We want to find the chain rule

for the composition $f(t, X_t^1, \dots, X_t^n)$ and compute it as another Itô process. The heuristic principle is the following. We compactly write

$$dX_t^i = \Phi_t^i dB_t + \Psi_t^i dt.$$

In the second order Taylor expansion of $f(t, X_t^1, \dots, X_t^n)$, we apply the following rule:

$$dX_t^i \cdot dt = 0, \quad dX_t^i \cdot dX_t^j = \Phi_t^i \Phi_t^j dt \quad \forall i, j.$$

This is reasonable in view of the relations

$$(dB_t)^2 = dt, \quad dB_t \cdot dt = 0, \quad (dt)^2 = 0.$$

Therefore, in its formal differential form Itô's formula reads

$$\begin{aligned} df(t, X_t^1, \dots, X_t^n) &= \partial_t f(t, X_t^1, \dots, X_t^n) dt + \sum_{i=1}^n \partial_{x_i} f(t, X_t^1, \dots, X_t^n) (\Phi_t^i dB_t + \Psi_t^i dt) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \partial_{x_i x_j}^2 f(t, X_t^1, \dots, X_t^n) \Phi_t^i \Phi_t^j dt. \end{aligned}$$

Theorem 3.4 (Itô's formula for Itô processes). *Let $\{X_t^i : t \geq 0\}$ ($1 \leq i \leq n$) be Itô processes given by (3.28) and let $f : [0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}$ be a $C^{1,2}$ -function. Then for any $s < t$, one has*

$$\begin{aligned} f(t, X_t^1, \dots, X_t^n) &= f(s, X_s^1, \dots, X_s^n) + \int_s^t \partial_u f(u, X_u^1, \dots, X_u^n) du \\ &\quad + \sum_{i=1}^n \int_s^t \partial_{x_i} f(u, X_u^1, \dots, X_u^n) \Phi_u^i dB_u + \sum_{i=1}^n \int_s^t \partial_{x_i} f(u, X_u^1, \dots, X_u^n) \Psi_u^i du \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \int_s^t \partial_{x_i x_j}^2 f(u, X_u^1, \dots, X_u^n) \Phi_u^i \Phi_u^j du \end{aligned} \tag{3.29}$$

Proof. Since one can approximate the integrands Φ^i, Ψ^i by simple processes (cf. Lemma 3.5), according to Proposition 3.2 it is enough to prove the formula when Φ^i, Ψ^i are simple. For the next simplification, the main observation is that the formula (3.29) is additive: if it is true for \int_s^t and \int_t^u then it is also true for \int_s^u .

Suppose for simplicity that the representations of Φ^i, Ψ^i (for all i) are over a common partition $\{t_k\}_{k=1}^m$, i.e.

$$\Phi_t^i = \xi_{k-1}^i, \quad \Psi_t^i = \eta_{k-1}^i \quad \text{when } t \in (t_{k-1}, t_k],$$

where $\xi_{k-1}^i, \eta_{k-1}^i$ are bounded and $\mathcal{F}_{t_{k-1}}$ -measurable. As a result of additivity, it is enough to prove (3.29) on each sub-interval $[t_{k-1}, t_k]$. But over such a sub-interval, one can view the processes X_t^i as

$$\begin{aligned} X_t^i &= X_{t_{k-1}}^i + \int_{t_{k-1}}^t \Phi_s^i dB_s + \int_{t_{k-1}}^t \Psi_s^i ds \\ &= X_{t_{k-1}}^i + \xi_{k-1}^i (B_t - B_{t_{k-1}}) + \eta_{k-1}^i (t - t_{k-1}). \end{aligned}$$

The crucial point here is that $X_{t_{k-1}}^i, \xi_{k-1}^i, \eta_{k-1}^i \in \mathcal{F}_{t_{k-1}}$ are regarded as frozen constants and $\tilde{B}_t \triangleq B_t - B_{t_{k-1}}$ ($t \in [t_{k-1}, t_k]$) is a Brownian motion independent of $\mathcal{F}_{t_{k-1}}$. Therefore, the process

$$f(t, X_t^1, \dots, X_t^n) - f(t_{k-1}, X_{t_{k-1}}^1, \dots, X_{t_{k-1}}^n) \quad (t_{k-1} \leq t \leq t_k)$$

is viewed as a function $g(t, \tilde{B}_t)$ (with the random variables from $\mathcal{F}_{t_{k-1}}$ being frozen). An application of Theorem 3.3 gives (3.29) over $[t_{k-1}, t_k]$. \square

Extension to higher dimensions

It is also useful to have a version of Itô's formula in higher dimensions. Let $B = \{(B_t^1, \dots, B_t^d) : t \geq 0\}$ be a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion. An Itô process with respect to the Brownian motion B takes the form

$$X_t = X_0 + \sum_{i=1}^d \int_0^t \Phi_s^i dB_s^i + \int_0^t \Psi_s ds.$$

Given n Itô processes of the above form and $f \in C^{1,2}$, with the following relations in mind one can easily write down the chain rule (Itô's formula) for the composition $f(t, X_t^1, \dots, X_t^n)$:

$$dB_t^i \cdot dB_t^j = \delta_{ij} \cdot dt \quad \text{where } \delta_{ij} \triangleq \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j. \end{cases} \quad (3.30)$$

The reason for the $i \neq j$ part in (3.30) is that $B^i B^j$ is a martingale when $i \neq j$ (cf. Example 3.2) and B^i, B^j have zero cross-variation in this case:

$$\lim_{\text{mesh } \mathcal{P} \rightarrow 0} \sum_{t_k \in \mathcal{P}} (B_{t_k}^i - B_{t_{k-1}}^i)(B_{t_k}^j - B_{t_{k-1}}^j) = 0 \quad \text{in prob.}$$

The explicit expression of the formula as well as its proof is left as an exercise.

As a simple application of Itô's formula, we now provide a precise explanation of the heat transfer problem that is discussed in the introduction (cf. (1.3)).

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a bounded continuous function with bounded derivative. Let $u(t, x)$ be the solution to the following PDE (heat equation):

$$\partial_t u(t, x) = \frac{1}{2} \partial_x^2 u(t, x), \quad (t, x) \in (0, \infty) \times \mathbb{R}$$

with initial condition $u(0, x) = f(x)$. Suppose that $\{B_t^x : t \geq 0\}$ is a one-dimensional Brownian motion starting at x (i.e. $B_0^x = x$) and $T > 0$ is fixed. By applying Itô's formula to the function $(t, y) \mapsto u(T - t, y)$ composed with B_t^x , one finds that

$$\begin{aligned} f(B_T^x) &= u(T, x) + \int_0^T \left(-\partial_t u(T - t, B_t^x) + \frac{1}{2} \partial_x^2 u(T - t, B_t^x) \right) dt \\ &\quad + \int_0^T u(T - t, B_t^x) dB_t^x. \end{aligned}$$

The Lebesgue integral term vanishes since u satisfies the heat equation. In addition, the stochastic integral is a martingale (why?) with zero expectation. Consequently, we arrive at

$$u(T, x) = \mathbb{E}[f(B_T^x)].$$

This provides a stochastic representation of the PDE solution $u(t, x)$.

3.5 Lévy's characterisation of Brownian motion

The notion of stochastic integrals can be further generalised to the situation where the integrator is a martingale. More specifically, let M be a continuous, square integrable, $\{\mathcal{F}_t\}$ -martingale. Given any $\{\mathcal{F}_t\}$ -progressively measurable process Φ that satisfies

$$\mathbb{P}\left(\int_0^t \Phi_s^2 d\langle M \rangle_s < \infty \quad \forall t \geq 0\right) = 1,$$

one can define the stochastic integral

$$\mathcal{I}(\Phi)_t = \int_0^t \Phi_s dM_s.$$

If Φ satisfies the stronger integrability condition

$$\mathbb{E}\left[\int_0^t \Phi_s^2 d\langle M \rangle_s\right] < \infty \quad \forall t \geq 0,$$

then $\mathcal{I}(\Phi)$ is a martingale and it satisfies the following generalised Itô's isometry:

$$\mathbb{E}\left[\left(\int_0^t \Phi_s dM_s\right)^2\right] = \mathbb{E}\left[\int_0^t \Phi_s^2 d\langle M \rangle_s\right].$$

Here the integral $\int_0^t \Phi_s^2 d\langle M \rangle_s$ is defined pathwisely for every ω . This makes sense in the classical way (more precisely, as a Riemann-Stieltjes integral) since the process $\langle M \rangle$ has non-decreasing sample paths. Note that the above facts are extensions of the Brownian case, as for the latter one has $\langle B \rangle_t = t$ (cf. Example 3.1).

In a similar way, one can also write down Itô's formula in this generalised context. Let $M^i = \{M_t^i\}$ ($1 \leq i \leq n$) be given continuous, square integrable, $\{\mathcal{F}_t\}$ -martingales. Let $f : [0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}$ be a $C^{1,2}$ -function. To obtain Itô's formula for the composition $f(t, M_t^1, \dots, M_t^n)$, the key principle is to apply the relations

$$dM_t^i \cdot dM_t^j = d\langle M^i, M^j \rangle_t \quad (3.31)$$

and

$$dM_t^i \cdot dt, d\langle M^i, M^j \rangle_t \cdot dM_t^k, (dt)^2 \text{ are all zero} \quad (3.32)$$

in the formal Taylor expansion of f (compare with (3.30) in the Brownian case). As a result, one has

$$\begin{aligned} f(t, M_t^1, \dots, M_t^n) &= f(0, M_0^1, \dots, M_0^n) + \int_0^t \partial_t f(s, M_s^1, \dots, M_s^n) du \\ &\quad + \sum_{i=1}^n \int_0^t \partial_{x_i} f(s, M_s^1, \dots, M_s^n) dM_s^i \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \int_0^t \partial_{x_i x_j}^2 f(s, M_s^1, \dots, M_s^n) d\langle M^i, M^j \rangle_s. \end{aligned} \quad (3.33)$$

The last integral is understood in the classical pathwise sense (as a Riemann-Stieltjes integral). A special and useful situation is the following *integration by parts* formula:

$$M_t N_t = M_0 N_0 + \int_0^t M_s dN_s + \int_0^t N_s dM_s + \langle M, N \rangle_t \quad (3.34)$$

for any square integrable martingales M, N . This is immediate from Itô's formula applied to the function $f(x, y) = xy$.

The most elegant way of establishing this more general theory is to use Hilbert space methods (Riesz representation theorem). We refer the reader to [16, Sec. IV.2] for the details. Let us use one example to illustrate the usefulness of this generalisation. Recall from (3.30) that a d -dimensional Brownian motion consists of d square integrable martingales whose bracket processes are given by $\langle B^i, B^j \rangle_t = \delta_{ij}t$. The following elegant result, which was due to P. Lévy, asserts that this property uniquely characterises the Brownian motion.

Theorem 3.5 (Lévy's characterisation of Brownian motion). *Let M^i ($1 \leq i \leq d$) be d continuous, square integrable, $\{\mathcal{F}_t\}$ -martingales and $M_0^i = 0$. Suppose that*

$$\langle M^i, M^j \rangle_t = \delta_{ij}t \quad \forall i, j.$$

Then (M^1, \dots, M^d) is a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion.

Proof. One needs to show the required distributional properties as well as the independence between $M_t - M_s$ and \mathcal{F}_s ($s < t$). Inspired by the proof of the strong Markov property (cf. (2.12)), in terms of characteristic functions all the desired properties are encoded in the following neat equation:

$$\mathbb{E}[e^{i\langle \theta, M_t - M_s \rangle_{\mathbb{R}^d}} | \mathcal{F}_s] = e^{-\frac{1}{2}|\theta|^2(t-s)} \quad \forall \theta \in \mathbb{R}^d. \quad (3.35)$$

It suffices to establish (3.35).

To this end, with given fixed $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$ we apply Itô's formula to the composition $Z_t \triangleq f(t, M_t^1, \dots, M_t^d)$, where the function f is defined by

$$f(t, x_1, \dots, x_d) \triangleq \exp\left(i \sum_{j=1}^d \theta_j x_j + \frac{1}{2}|\theta|^2 t\right).$$

Since $\langle M^i, M^j \rangle_t = \delta_{ij}t$ by assumption, the formula (3.33) yields

$$\begin{aligned} Z_t &= Z_0 + \frac{|\theta|^2}{2} \int_0^t Z_s ds + i \sum_{j=1}^d \theta_j \int_0^t Z_s dM_s^j \\ &\quad + \frac{1}{2} \sum_{j=1}^d (i\theta_j)^2 \int_0^t Z_s ds \\ &= 1 + i \sum_{j=1}^d \theta_j \int_0^t Z_s dM_s^j. \end{aligned}$$

The stochastic integral on the right hand side is a martingale. Therefore, $\{Z_t\}$ is a martingale and the desired equation (3.35) is merely a rearrangement of the martingale property

$$\mathbb{E}[Z_t | \mathcal{F}_s] = Z_s.$$

□

Example 3.3. The *sign function* is defined by

$$\operatorname{sgn}(x) \triangleq \begin{cases} 1, & x > 0; \\ -1, & x \leq 0. \end{cases}$$

The stochastic integral $W_t \triangleq \int_0^t \operatorname{sgn}(B_s) dB_s$ is a well-defined square integrable martingale. According to Proposition 3.5, one finds

$$\langle W \rangle_t = \int_0^t (\operatorname{sgn}(B_s))^2 ds = \int_0^t 1 ds = t.$$

By Theorem 3.5, the process W is also a Brownian motion. In addition, it can be shown that

$$|B_t| = W_t + L_t, \tag{3.36}$$

where L_t is the non-decreasing process defined by

$$L_t \triangleq \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbf{1}_{(-\varepsilon, \varepsilon)}(B_s) ds. \tag{3.37}$$

The formula (3.36) can be viewed as a generalisation of Itô's formula to the function $f(x) = |x|$ which is non-differentiable at the origin. The sign function can be viewed as the derivative of $f(x)$, leading to the stochastic integral term W_t in

the formula (3.36). L_t can be viewed as the second order term arising from the “second derivative” of $f(x)$ which should not be understood in any classical sense (in fact, $\frac{1}{2}f'' = \delta_0$ is a *generalised function* known as the *Dirac delta function* at the origin). The Lebesgue integral on the right hand side of (3.37) represents the amount of time (before t) that the Brownian motion stays in the region $(-\varepsilon, \varepsilon)$. As a result, L_t can be viewed as the occupation density (before t) of the Brownian motion at the level $x = 0$. This is known as the *local time* of B at $x = 0$. We refer the reader to [16, Chap. VI] for a discussion on these results as well as a beautiful introduction to the theory of local times. Local time theory is essential for the study of one-dimensional SDEs and diffusion processes on graphs.

3.6 A martingale representation theorem

There are several beautiful connections between continuous martingales and Brownian motion. In vague terms, we describe two types of fundamental results along this line. Let $M = \{M_t : t \geq 0\}$ be a continuous $\{\mathcal{F}_t\}$ -martingale.

- (i) *The Dambis-Dubins-Schwarz theorem*: M is the time-change of some Brownian motion, i.e. there exists a Brownian motion B such that $M_t = B_{\langle M \rangle_t}$.
- (ii) *The martingale representation theorem*: M can be written as an Itô integral $M_t = M_0 + \int_0^t \Phi_s dB_s$.

These results suggest that continuous martingales are not very general objects: they can be essentially reduced to Brownian motions and/or their stochastic integrals. As a consequence, the behaviour of continuous martingales is to some extent similar to the latter two objects.

In this section, we only elaborate one special type of martingale representation theorems. We also take this opportunity to introduce the elegant idea of Hilbert spaces.

Let $B = \{B_t : 0 \leq t \leq 1\}$ be a Brownian motion defined on some filtered probability space $(\Omega, \mathcal{F}_1^B, \mathbb{P}; \{\mathcal{F}_t^B : 0 \leq t \leq 1\})$. Here we particularly emphasise that the filtration $\{\mathcal{F}_t^B\}$ is the *natural filtration* associated with B . In other words, we work with the probability space which only carries the intrinsic information of the Brownian motion. The martingale representation theorem in this context is stated as follows.

Theorem 3.6. *Let $M = \{M_t : 0 \leq t \leq 1\}$ be a continuous, square integrable, $\{\mathcal{F}_t^B\}$ -martingale. Then there exists a unique $\{\mathcal{F}_t^B\}$ -progressively measurable pro-*

cess $\Phi = \{\Phi_t : 0 \leq t \leq 1\}$ satisfying

$$\mathbb{E}\left[\int_0^1 \Phi_t^2 dt\right] < \infty, \quad (3.38)$$

such that

$$M_t = M_0 + \int_0^t \Phi_s dB_s. \quad (3.39)$$

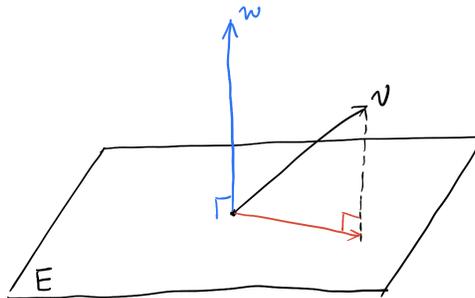
Hilbert spaces

We will use the notion of Hilbert spaces to prove Theorem 3.6. Let us first describe the intuition behind the abstract nonsense.

In Euclidean geometry, every element in \mathbb{R}^2 or \mathbb{R}^3 can be viewed as a vector. There is a notion of inner product $\langle v, w \rangle$ between two vectors v, w , which satisfies the following properties:

- (i) *symmetry*: $\langle v, w \rangle = \langle w, v \rangle$;
- (ii) *bilinearity*: $\langle cv_1 + v_2, w \rangle = c\langle v_1, w \rangle + \langle v_2, w \rangle$ where $c \in \mathbb{R}$ is a scalar;
- (iii) *positive definiteness*: $\langle v, v \rangle \geq 0$ where equality holds if and only if $v = 0$.

The inner product can be used to measure all sorts of geometric properties e.g. length ($|v| = \sqrt{\langle v, v \rangle}$), angle ($\angle_{v,w} = \frac{\langle v, w \rangle}{|v| \cdot |w|}$), orthogonality ($v \perp w \iff \langle v, w \rangle = 0$) etc. Given a vector v and a subspace $E \subseteq \mathbb{R}^3$, one can naturally talk about the orthogonal projection of v onto E . If E is a proper subspace (i.e. $E \neq \mathbb{R}^3$), one can find at least one non-zero vector w that is perpendicular to E .



The notion of Hilbert spaces generalises the above considerations.

Definition 3.8. Let H be a vector space over \mathbb{R} . An *inner product* over H is a function $\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{R}$ which satisfies the above Properties (i)–(iii). A vector space equipped with an inner product is called an *inner product space*.

Two elements v, w are said to be *orthogonal* (denoted as $v \perp w$) if $\langle v, w \rangle = 0$. By using the inner product, one can define the notion of length (more commonly known as a *norm*) by

$$\|v\| \triangleq \sqrt{\langle v, v \rangle}, \quad v \in H.$$

With this norm structure one can talk about convergence just like in Euclidean spaces: we say that v_n *converges to* v if $\|v_n - v\| \rightarrow 0$ as $n \rightarrow \infty$. A sequence $\{v_n : n \geq 1\}$ in H is said to be a *Cauchy sequence* in H if for any $\varepsilon > 0$, there exists $N \geq 1$ such that

$$m, n > N \implies \|v_m - v_n\| < \varepsilon.$$

Definition 3.9. A *Hilbert space* is a complete inner product space, i.e. an inner product space in which every Cauchy sequence converges.

Example 3.4. \mathbb{R}^d is a Hilbert space when equipped with the Euclidean inner product:

$$\langle x, y \rangle \triangleq \sqrt{(x_1 - y_1)^2 + \cdots + (x_d - y_d)^2}.$$

In finite dimensions there is no need to emphasise completeness as every finite dimensional inner product space is complete. The completeness property is essential when one considers infinite dimensional spaces such as a space of functions. In infinite dimensions it is also important to emphasise closedness when come to the use of subspaces: a subspace E is *closed* if $v_n \in E, v_n \rightarrow v \implies v \in E$. This notion is again not needed in finite dimensions as every subspace is closed in that case.

A favorite example of an infinite dimensional Hilbert space is the following.

Example 3.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$ be the space of square integrable random variables (we identify two elements X, Y if $X = Y$ a.s.). Define the inner product over H by

$$\langle X, Y \rangle_{L^2} \triangleq \mathbb{E}[XY], \quad X, Y \in H. \quad (3.40)$$

Then $(H, \langle \cdot, \cdot \rangle_{L^2})$ is a Hilbert space.

The most important thing to keep in mind is that all the properties we have mentioned in the Euclidean case remain valid in a general Hilbert space. For instance, recall that (cf. Section 1.3) the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ is the orthogonal projection of $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ onto the closed subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$. To prove the martingale representation theorem, we need one more definition and one specific fact.

Definition 3.10. A subset A of a Hilbert space H is said to be *total* if no non-zero elements in H can be perpendicular to every element in A , i.e.

$$v \perp w \quad \forall w \in A \implies v = 0.$$

Example 3.6. Any two non-colinear vectors in \mathbb{R}^2 form a total subset, since no non-zero vectors in the plane can be perpendicular to two linearly independent vectors at the same time.

Example 3.7. Let $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$ and let $A = \{\mathbf{1}_F : F \in \mathcal{F}\}$. Then A is a total subset. To see this, let $f \neq 0$ and $f \perp \mathbf{1}_F$ for all $F \in \mathcal{F}$. Then f is perpendicular to all *simple functions* of the form

$$\varphi = \sum_{k=1}^m a_k \mathbf{1}_{F_k}.$$

But any function (in particular, f) can be well approximated by simple functions. As a consequence, f is perpendicular to itself and thus $f = 0$.

Proposition 3.6. *Let H be a Hilbert space and let E be a closed, proper subspace of H . Then there exists $v \neq 0$ such that v is perpendicular to all elements in E .*

Proof of the martingale representation theorem

The proof of Theorem 3.6 again relies on the use of exponential martingales. Let \mathcal{T} denote the space of step functions on $[0, 1]$ of the form

$$f(t) = \sum_{k=1}^n c_{k-1} \mathbf{1}_{(t_{k-1}, t_k]}(t) \tag{3.41}$$

where $0 = t_0 < t_1 < \dots < t_n = 1$ is a partition of $[0, 1]$ and $c_{k-1} \in \mathbb{R}$. Given $f \in \mathcal{T}$, we define the exponential process

$$\mathcal{E}_t^f \triangleq \exp \left(\int_0^t f(s) dB_s - \frac{1}{2} \int_0^t |f(s)|^2 ds \right), \quad t \in [0, 1].$$

One checks that it is a martingale.

To prove Theorem 3.6, we make use of the Hilbert space $H \triangleq L^2(\Omega, \mathcal{F}_1^B, \mathbb{P})$ equipped with the inner product (3.40).

Lemma 3.8. *The set $\{\mathcal{E}_1^f : f \in \mathcal{T}\}$ is total in H .*

Proof. Let $Y \in H$ be such that $\mathbb{E}[Y\mathcal{E}_1^f] = 0$ for all $f \in \mathcal{T}$. We want to show that $Y = 0$. Since Y is \mathcal{F}_1^B -measurable, this is equivalent to showing that

$$\mathbb{E}[Y\mathbf{1}_A] = 0 \quad \forall A \in \mathcal{F}_1^B. \quad (3.42)$$

But \mathcal{F}_1^B is the σ -algebra generated by the Brownian motion. As a result, it is enough to consider those A 's of the form

$$A = \{\omega : (B_{t_1}, \dots, B_{t_n}) \in \Gamma\}$$

where $0 = t_0 < t_1 < \dots < t_n = 1$ is a partition of $[0, 1]$ and $\Gamma \in \mathcal{B}(\mathbb{R}^n)$. Given fixed t_i 's, we define the signed measure

$$\nu(\Gamma) \triangleq \mathbb{E}[Y\mathbf{1}_\Gamma(B_{t_1}, \dots, B_{t_n})], \quad \Gamma \in \mathcal{B}(\mathbb{R}^n).$$

To show that $\nu = 0$, we consider its *Fourier transform*

$$\begin{aligned} \hat{\nu}(\lambda_1, \dots, \lambda_n) &\triangleq \int_{\mathbb{R}^n} e^{i(\lambda_1 x_1 + \dots + \lambda_n x_n)} \nu(dx) \\ &= \mathbb{E}[Y e^{i(\lambda_1 B_{t_1} + \dots + \lambda_n B_{t_n})}], \quad (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n. \end{aligned}$$

The key observation is that the exponential random variable $e^{i(\lambda_1 B_{t_1} + \dots + \lambda_n B_{t_n})}$ can be related to \mathcal{E}_1^f for some $f \in \mathcal{T}$. Indeed, for any $f \in \mathcal{T}$ given by (3.41), one has

$$\begin{aligned} \mathcal{E}_1^f &= \exp\left(\sum_{k=1}^n c_{k-1}(B_{t_k} - B_{t_{k-1}}) - \frac{1}{2} \int_0^1 |f(t)|^2 dt\right) \\ &= \exp\left(\sum_{k=1}^n (c_{k-1} - c_k)B_{t_k} - \frac{1}{2} \int_0^1 |f(t)|^2 dt\right). \end{aligned}$$

As a result, by defining $c_{k-1} - c_k = i\lambda_k$ ($1 \leq k \leq n$) or equivalently

$$c_k \triangleq i\lambda_k + \dots + i\lambda_n,$$

the associated f satisfies

$$e^{i(\lambda_1 B_{t_1} + \dots + \lambda_n B_{t_n})} = \mathcal{E}_1^f \cdot \exp\left(\frac{1}{2} \int_0^1 |f(t)|^2 dt\right).$$

Therefore,

$$\hat{\nu}(\lambda_1, \dots, \lambda_n) = \exp\left(\frac{1}{2} \int_0^1 |f(t)|^2 dt\right) \cdot \mathbb{E}[Y\mathcal{E}_1^f] = 0$$

by the assumption on Y . Since the Fourier transform is one-to-one, we conclude that $\nu = 0$ and thus (3.42) follows. \square

Remark 3.9. If one is not familiar with the language of Fourier transform, here is a less enlightening argument. Functions of the form $e^{i(\lambda_1 x_1 + \dots + \lambda_n x_n)}$ (with arbitrary choices of $\lambda_1, \dots, \lambda_n$) are rich enough to generate the class of bounded continuous function. Hence

$$\begin{aligned} \mathbb{E}[Y e^{i(\lambda_1 B_{t_1} + \dots + \lambda_n B_{t_n})}] &= 0 \quad \forall \lambda_1, \dots, \lambda_n \\ \implies \mathbb{E}[Y f(B_{t_1}, \dots, B_{t_n})] &= 0 \quad \forall \text{bounded continuous } f. \end{aligned}$$

By using continuous functions to approximate indicator functions, one concludes that

$$\mathbb{E}[Y \mathbf{1}_\Gamma(B_{t_1}, \dots, B_{t_n})] = 0 \quad \forall \Gamma \in \mathcal{B}(\mathbb{R}^n).$$

Let us now introduce one more Hilbert space. We denote $L^2(B)$ as the space of progressively measurable processes $\Phi = \{\Phi_t : 0 \leq t \leq 1\}$ satisfying the integrability condition (3.38). $L^2(B)$ is a Hilbert space under the inner product

$$\langle \Phi, \Psi \rangle_{L^2(B)} \triangleq \mathbb{E} \left[\int_0^1 \Phi_t \Psi_t dt \right], \quad \Phi, \Psi \in L^2(B).$$

Proposition 3.7. *Let $Y \in H$. Then there exists a unique $\Phi \in L^2(B)$ such that*

$$Y = \mathbb{E}[Y] + \int_0^1 \Phi_t dB_t. \quad (3.43)$$

Proof. Existence. Let E denote the collection of all those $Y \in H$ which satisfies the desired property. It is clear that E is a subspace of H . We claim that E is closed. Let

$$Y_n = \mathbb{E}[Y_n] + \int_0^1 \Phi_t^{(n)} dB_t \in E \quad (3.44)$$

and $Y_n \rightarrow Y$ in H . According to Itô's isometry (3.22),

$$\begin{aligned} & \|\Phi^{(m)} - \Phi^{(n)}\|_{L^2(B)}^2 \\ &= \mathbb{E} \left[\int_0^1 (\Phi_t^{(m)} - \Phi_t^{(n)})^2 dt \right] = \mathbb{E} \left[\left(\int_0^1 (\Phi_t^{(m)} - \Phi_t^{(n)}) dB_t \right)^2 \right] \\ &= \text{Var}[Y_m - Y_n] \leq \mathbb{E}[(Y_m - Y_n)^2]. \end{aligned}$$

Since $\{Y_n\}$ is a Cauchy sequence in H (because it converges), we see that $\{\Phi^{(n)}\}$ is a Cauchy sequence in $L^2(B)$ and hence $\Phi^{(n)} \rightarrow \Phi$ for some $\Phi \in L^2(B)$. By taking $n \rightarrow \infty$ in (3.44), one finds

$$Y = \mathbb{E}[Y] + \int_0^1 \Phi_t dB_t$$

and thus $Y \in E$. This shows that E is closed. We also have

$$\{\mathcal{E}_1^f : f \in \mathcal{T}\} \subseteq E,$$

since

$$\mathcal{E}_1^f = 1 + \int_0^1 f(t)\mathcal{E}_t^f dB_t \quad (\text{so } \Phi_t = f(t)\mathcal{E}_t^f)$$

as a consequence of Itô's formula.

To prove the existence part of the proposition, one needs to show that $E = H$. Suppose on the contrary that $E \neq H$. From Proposition 3.6, there exists $Y \neq 0$ such that Y is perpendicular to all elements in E . In particular, $Y \perp \mathcal{E}_1^f$ for all $f \in \mathcal{T}$. Since $\{\mathcal{E}_1^f : f \in \mathcal{T}\}$ is total in H (cf. Lemma 3.8), we conclude that $Y = 0$ which is a contradiction. Therefore, $E = H$.

Uniqueness. Suppose that $Y \in H$ admits two representations (3.43) with some $\Phi, \Psi \in L^2(B)$. Then $\int_0^1 (\Phi_t - \Psi_t)dB_t = 0$. Itô's isometry implies that

$$\|\Phi - \Psi\|_{L^2(B)}^2 = \mathbb{E}\left[\left(\int_0^1 (\Phi_t - \Psi_t)dB_t\right)^2\right] = 0.$$

Therefore, $\Phi = \Psi$. □

We can now easily complete the proof of the martingale representation theorem.

Proof of Theorem 3.6. Let $M = \{M_t : 0 \leq t \leq 1\}$ be a square integrable continuous martingale. According to Proposition 3.7, there exists a unique $\Phi \in L^2(B)$ such that

$$M_1 = \mathbb{E}[M_1] + \int_0^1 \Phi_t dB_t.$$

In addition, the stochastic integral $\int_0^t \Phi_s dB_s$ is a martingale in this case. By conditioning on \mathcal{F}_t , we obtain

$$M_t = \mathbb{E}[M_1] + \int_0^t \Phi_s dB_s.$$

Note that since $B_0 = 0$, \mathcal{F}_0^B is the trivial σ -algebra $\{\emptyset, \Omega\}$ and thus $\mathbb{E}[M_1] = \mathbb{E}[M_0] = M_0$. The representation (3.39) thus follows. □

Example 3.8. According to (3.15) and (3.17), the stochastic integrable representations of B_1^2 and B_1^3 (in the sense of Proposition 3.7) are given by

$$B_1^2 = 1 + 2 \int_0^1 B_t dB_t, \quad B_1^3 = \int_0^1 (6 \int_0^t B_s dB_s + 3) dB_t.$$

The martingale representation theorem extends naturally to the case of multidimensional Brownian motion without essential difficulties. We only state the result and leave its proof to the reader.

Theorem 3.7. *Let $B = \{B_t : t \geq 0\}$ be a d -dimensional Brownian motion and let $\{\mathcal{F}_t^B\}$ be its natural filtration. Let $\{M_t : t \geq 0\}$ be a continuous, square integrable, $\{\mathcal{F}_t^B\}$ -martingale. Then there exist d $\{\mathcal{F}_t^B\}$ -progressively measurable processes $\Phi = (\Phi^1, \dots, \Phi^d)$ such that*

$$M_t = M_0 + \sum_{j=1}^d \int_0^t \Phi_s^j dB_s^j, \quad t \geq 0.$$

These Φ_j 's are unique in the sense that if $\Psi = (\Psi^1, \dots, \Psi^d)$ satisfies the same property, then with probability one

$$(\Phi_t^1(\omega), \dots, \Phi_t^d(\omega)) = (\Psi_t^1(\omega), \dots, \Psi_t^d(\omega)) \quad t\text{-a.e.}$$

Remark 3.10. It is natural to ask whether the integrand Φ can be constructed explicitly in the representation theorem. This is a challenging but important question whose solution, known as the *Clark–Ocone–Karatzas formula*, relies on techniques from stochastic calculus of variations (the Malliavin calculus). We refer the reader to [12, Sec. 1.6] for a discussion as well as its applications in mathematical finance.

3.7 The Cameron-Martin-Girsanov transformation

In this section, we discuss a rather useful technical in stochastic calculus: change of measure. Vaguely speaking, this technique allows one to eliminate the drift effect in an Itô process or a stochastic differential equation without changing the martingale part.

3.7.1 Motivation: the original approach of Cameron and Martin

It is well known that the Lebesgue measure on \mathbb{R}^d is translation invariant (translating a set along any direction does not change its volume). We begin by asking

the following question: if one considers the Gaussian measure (law of the standard d -dimensional Gaussian vector)

$$\nu(dx) = \frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2} dx,$$

how does ν behave under translation?

This can be figured out easily by explicit calculation. On the canonical probability space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \nu)$, the coordinate functions

$$\xi^i(x) \triangleq x_i, \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d$$

define a standard d -dimensional Gaussian vector

$$\xi = (\xi^1, \dots, \xi^d) \sim N(0, \text{Id}).$$

Given fixed $h \in \mathbb{R}^d$, we consider the translation map $T_h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined by $T_h(x) \triangleq x + h$. Let $\nu_h \triangleq \nu \circ (T_h)^{-1}$ denote the push-forward of ν by the map T_h , i.e.

$$\nu_h(A) = \nu(T_h^{-1}A) = \nu(\{x : x + h \in A\}) = \nu(A - h).$$

To compute ν_h explicitly, let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be an arbitrary test function. By the definition of ν_h ,

$$\begin{aligned} \int_{\mathbb{R}^d} f(y) \nu_h(dy) &= \int_{\mathbb{R}^d} f(x + h) \nu(dx) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x + h) e^{-|x|^2/2} dx \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(y) e^{|y-h|^2/2} dy \quad (y \triangleq x + h) \\ &= \int_{\mathbb{R}^d} f(y) e^{\langle h, y \rangle_{\mathbb{R}^d} - |h|^2/2} \nu(dy). \end{aligned}$$

As a result, ν_h is *absolutely continuous* with respect to ν with density function (cf. Appendix (8) for this terminology)

$$\frac{d\nu_h}{d\nu}(x) = e^{\langle h, x \rangle_{\mathbb{R}^d} - |h|^2/2}, \quad x \in \mathbb{R}^d. \quad (3.45)$$

This property is known as the *quasi-invariance* of Gaussian measures.

There is an equivalent way of looking that the above fact that is more relevant to us. If we define a new measure ν_h by the formula (3.45), then $\eta \triangleq \xi - h$ is a

standard Gaussian vector under ν_h . This is because the distribution of η under ν_h is the push-forward of ν_h by the map $T_{-h} : x \mapsto x - h$, which is nothing but just the original Gaussian measure ν .

We now generalise the previous calculation to the context of Brownian motion (the infinite dimensional situation). We first describe the canonical probability space as the infinite dimensional analogue of $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \nu)$. Let \mathcal{W} be the space of continuous paths $w : [0, 1] \rightarrow \mathbb{R}$ with $w_0 = 0$. By viewing \mathcal{W} as a sample space, one can define a canonical stochastic process (the *coordinate process*) $W = \{W_t : 0 \leq t \leq 1\}$ by

$$W_t(w) \triangleq w_t, \quad w \in \mathcal{W}.$$

There is a unique probability measure μ defined on the σ -algebra $\mathcal{B}(\mathcal{W})$ generated by the coordinate process W , under which the process W is a Brownian motion. Its existence can be seen as follows. Let $B = \{B_t : 0 \leq t \leq 1\}$ be a Brownian motion defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (cf. Theorem 2.1). Since B has continuous sample paths, it can be viewed as a “random variable” taking values in the path space \mathcal{W} . The measure μ is the *law* of B (the push-forward of \mathbb{P} by the map $B : \Omega \rightarrow \mathcal{W}$). Uniqueness is obvious since the distribution of Brownian motion is uniquely specified in its definition.

Definition 3.11. The probability space $(\mathcal{W}, \mathcal{B}(\mathcal{W}), \mu)$ is known as the *Wiener space*. The measure μ is known as the *Wiener measure*.

The Wiener measure can be viewed as the “standard Gaussian measure” on the path space $(\mathcal{W}, \mathcal{B}(\mathcal{W}))$, which is an extension of the finite dimensional situation.

We now fix a given path $h \in \mathcal{W}$ (a direction). Suppose that h has “nice” regularity properties and let us not bother with what they are at the moment. We again consider the translation map $T_h : \mathcal{W} \rightarrow \mathcal{W}$ defined by $T_h(w) = w + h$. Let μ_h denote the push-forward of μ by T_h .

To understand the relationship between μ_h and μ , we use finite dimensional approximation. For each $n \geq 1$, consider the partition

$$\mathcal{P}_n : 0 = t_0 < t_1 < \cdots < t_n = 1$$

of $[0, 1]$ into n sub-intervals with equal length $1/n$. Given $w \in \mathcal{W}$, let $w^{(n)} \in \mathcal{W}$ be the piecewise linear interpolation of w over the partition \mathcal{P}_n . More precisely, $w_{t_i}^{(n)} = w_{t_i}$ for $t_i \in \mathcal{P}_n$ and $w^{(n)}$ is linear on each sub-interval associated with \mathcal{P}_n . Given any test function $f : \mathcal{W} \rightarrow \mathbb{R}$, we define the approximation of f by $f^{(n)}(w) \triangleq f(w^{(n)})$. Note that $f^{(n)}$ depends only on the values $\{w_{t_1}, \dots, w_{t_n}\}$. Therefore, $f^{(n)}$ is essentially a function on \mathbb{R}^n : one can write $f^{(n)}(w) = H(w_{t_1}, \dots, w_{t_n})$ where $H : \mathbb{R}^n \rightarrow \mathbb{R}$.

Using the above notation, we perform the similar calculation as in the \mathbb{R}^d case. Note that under the Wiener measure, $w \mapsto (w_{t_1}, \dots, w_{t_n})$ is a Gaussian vector with density

$$\rho_{t_1, \dots, t_n}(x_1, \dots, x_n) = C \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right)$$

where

$$C \triangleq \frac{1}{(2\pi)^{n/2} \sqrt{t_1(t_2 - t_1) \cdots (t_n - t_{n-1})}}.$$

It follows that

$$\begin{aligned} & \int_{\mathcal{W}} f^{(n)}(w) \mu_h(dw) \\ &= \int_{\mathcal{W}} f^{(n)}(w + h) \mu(dw) \\ &= \int_{\mathcal{W}} H(w_{t_1} + h_{t_1}, \dots, w_{t_n} + h_{t_n}) \mu(dw) \\ &= C \int_{\mathbb{R}^n} H(x_1 + h_{t_1}, \dots, x_n + h_{t_n}) \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right) dx \\ &= C \int_{\mathbb{R}^n} H(y_1, \dots, y_n) \exp\left(\sum_{i=1}^n \frac{h_{t_i} - h_{t_{i-1}}}{t_i - t_{i-1}} \cdot (y_i - y_{i-1})\right. \\ &\quad \left. - \frac{1}{2} \sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{t_i - t_{i-1}} - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - y_{i-1})^2}{t_i - t_{i-1}}\right) dy \\ &= \int_{\mathcal{W}} f^{(n)}(w) \exp\left(\sum_{i=1}^n \frac{h_{t_i} - h_{t_{i-1}}}{t_i - t_{i-1}} \cdot (w_{t_i} - w_{t_{i-1}}) - \frac{1}{2} \sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{t_i - t_{i-1}}\right) \mu(dw), \end{aligned}$$

Here comes the crucial observation. As $n \rightarrow \infty$, one has $f^{(n)}(w) \rightarrow f(w)$, and it is natural to expect that

$$\begin{aligned} & \sum_{i=1}^n \frac{h_{t_i} - h_{t_{i-1}}}{t_i - t_{i-1}} \cdot (w_{t_i} - w_{t_{i-1}}) \rightarrow \int_0^1 h'_t dW_t, \\ & \sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{t_i - t_{i-1}} = \sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{(t_i - t_{i-1})^2} \cdot (t_i - t_{i-1}) \rightarrow \int_0^1 (h'_t)^2 dt, \end{aligned} \quad (3.46)$$

where the first limit is an Itô integral! Therefore, after taking limit we formally arrive at

$$\int_{\mathcal{W}} f(w) \mu_h(dw) = \int_{\mathcal{W}} f(w) \exp\left(\int_0^1 h'_t dW_t - \frac{1}{2} \int_0^1 (h'_t)^2 dt\right) \mu(dw).$$

This equation suggests that μ_h is absolutely continuous with respect to μ with density

$$\frac{d\mu_h}{d\mu} = \exp\left(\int_0^1 h'_t dW_t - \frac{1}{2} \int_0^1 (h'_t)^2 dt\right). \quad (3.47)$$

To rephrase this fact in an equivalent way, if we define μ_h by the formula (3.47), under the new measure μ_h the translated process

$$\tilde{W}_t \triangleq W_t - h_t = W_t - \int_0^t h'_s ds$$

becomes a Brownian motion. This is because the distribution of \tilde{W} under μ_h is the push-forward of μ_h by the map $T_{-h} : w \mapsto w - h$, which is exactly the Wiener measure μ .

The above discussion outlines the essential idea of R.H. Cameron and W.T. Martin's original work in 1944. The main technical difficulty lies in verifying the convergence in (3.46) for a suitable class of h . It turns out that the precise regularity assumption on h is given as follows: h needs to be absolutely continuous and $\int_0^1 (h'_t)^2 dt < \infty$. Cameron-Martin's theorem can now be stated below. We refer the reader to [19, Sec. 1.1] for a simplified modern proof.

Theorem 3.8. *Let \mathcal{H} be the space of absolutely continuous paths $h \in \mathcal{W}$ such that $\int_0^1 (h'_t)^2 dt < \infty$. Then for any given $h \in \mathcal{H}$, μ_h is absolutely continuous with respect to μ with density given by (3.47). In addition, $w_t - \int_0^t h'_s ds$ is a Brownian motion under μ^h .*

Remark 3.11. There is a deeper result which reveals that the infinite dimensional situation (the Brownian motion case) is drastically different from the finite dimensional case: the quasi-invariance property is only true along *directions in \mathcal{H}* and μ^h is *singular* to μ for all $h \notin \mathcal{H}$. The space \mathcal{H} , known as the *Cameron-Martin subspace*, plays a fundamental role in the stochastic analysis on the Wiener space.

3.7.2 Girsanov's approach

With the aid of martingale methods, I.V. Girsanov independently considered a similar problem but in a more general context which we now elaborate.

The following information is fixed throughout the whole discussion. Let $B = \{(B_t^1, \dots, B_t^d) : t \geq 0\}$ be a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $1 \leq i \leq d$ let $X^i = \{X_t^i : t \geq 0\}$ be an Itô integrable process with respect to B^i .

Inspired by the formula (3.47), we define the following exponential process (as appeared for several times before!):

$$\mathcal{E}_t^X \triangleq \exp\left(\sum_{i=1}^d \int_0^t X_s^i dB_s^i - \frac{1}{2} \int_0^t |X_s|^2 ds\right), \quad t \geq 0. \quad (3.48)$$

By applying Itô's formula to the exponential function, one finds that

$$\mathcal{E}_t^X = 1 + \sum_{i=1}^d \int_0^t \mathcal{E}_s^X X_s^i dB_s^i.$$

Although it is a stochastic integral, it may fail to be a martingale in general. The entire discussion in the sequel is based on the *assumption* that $\{\mathcal{E}_t^X : t \geq 0\}$ is a martingale. This will be the case for instance if

$$\mathbb{E}\left[\int_0^t (\mathcal{E}_s^X)^2 |X_s|^2 ds\right] < \infty$$

as suggested by Proposition 3.5. But this condition is hard to check as it involves the process \mathcal{E}_t^X itself. There is a neat sufficient condition due to A. Novikov, which only involves the process X_t . We sketch the proof and refer the reader to [10, Sec. 3.5.D] for the deeper details.

Theorem 3.9 (Novikov's condition). *Under the previous set-up, suppose that*

$$\mathbb{E}\left[\exp\left(\frac{1}{2} \int_0^t |X_s|^2 ds\right)\right] < \infty \quad \forall t \geq 0.$$

Then $\{\mathcal{E}_t^X : t \geq 0\}$ is an $\{\mathcal{F}_t\}$ -martingale.

Sketch of proof. Let us write $M_t \triangleq \sum_{i=1}^d \int_0^t X_s^i dB_s^i$. One finds that $\langle M \rangle_t = \int_0^t |X_s|^2 ds$. As a result, we can write

$$\mathcal{E}_t^X = e^{M_t - \langle M \rangle_t / 2}.$$

The strategy of proving the theorem contains the following steps.

(i) $\{\mathcal{E}_t^X\}$ is a supermartingale. To show that it is a martingale, it is enough to prove that $\mathbb{E}[\mathcal{E}_t^X] = 1$ for all t .

(ii) Define the process $B_s \triangleq M_{\tau_s}$, where τ_s is the inverse of the non-decreasing function $t \mapsto \langle M \rangle_t$. Then

$$\langle B \rangle_s = \langle M \rangle_{\tau_s} = s$$

by the definition of τ_s . It follows from Lévy's characterisation theorem that B is a Brownian motion. In other words, we have written M as the time-change of a Brownian motion: $M_t = B_{\langle M \rangle_t}$.

(iii) We know that $Z_s \triangleq e^{B_s - s/2}$ is a martingale. For fixed $t \geq 0$, by thinking of $\langle M \rangle_t$ as a stopping time, one expects from the optional sampling theorem that

$$\mathbb{E}[Z_{\langle M \rangle_t}] = \mathbb{E}[Z_0] = 1,$$

which is exactly the desired relation

$$\mathbb{E}[e^{M_t - \langle M \rangle_t/2}] = 1.$$

□

Remark 3.12. In the setting of Cameron-Martin, given $h \in \mathcal{H}$ (cf. Theorem 3.8), by definition we know that $\int_0^t (h'_s)^2 ds < \infty$ for every t . Therefore, Novikov's condition is satisfied and the process

$$\mathcal{E}_t^h \triangleq \exp \left(\int_0^t h'_s dB_s - \frac{1}{2} \int_0^t (h'_s)^2 ds \right)$$

is a martingale.

From now on, we make the following assumption exclusively.

Assumption 3.1. *The process $\mathcal{E}^X = \{\mathcal{E}_t^X\}$ is an $\{\mathcal{F}_t\}$ -martingale.*

Inspired by Cameron-Martin's formula (3.47), for each given $T > 0$ we define a new measure

$$\mathbb{Q}_T(A) \triangleq \mathbb{E}[\mathbf{1}_A \mathcal{E}_T^X], \quad A \in \mathcal{F}_T.$$

According to Assumption 3.1, \mathbb{Q}_T is a probability measure on (Ω, \mathcal{F}_T) . *Girsanov's transformation theorem* can now be stated as follows.

Theorem 3.10. *Define the translated process $\tilde{B} = (\tilde{B}^1, \dots, \tilde{B}^d)$ by*

$$\tilde{B}_t^i \triangleq B_t^i - \int_0^t X_s^i ds. \tag{3.49}$$

Under Assumption 3.1, for each $T > 0$ the process $\{\tilde{B}_t : 0 \leq t \leq T\}$ is a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion under the new measure \mathbb{Q}_T .

Example 3.9. Let $d = 1$ and $X_t \equiv c$ where $c \neq 0$ is a deterministic constant. Then Theorem 3.10 is satisfied and $\tilde{B}_t \triangleq B_t - ct$ ($0 \leq t \leq T$) is a Brownian motion under \mathbb{Q}_T . Note that under the old measure \mathbb{P} , the process \tilde{B} has the extra drift term $-ct$ (a Brownian motion with a drift). Girsanov's theorem thus enables one to eliminate the drift by working under the new measure \mathbb{Q}_T .

The rest of this section is devoted to the proof of Theorem 3.10. We begin with a useful lemma which enables us to compute conditional expectations under \mathbb{Q}_T .

Lemma 3.9. *Let $0 \leq s \leq t \leq T$. Suppose that Y is an \mathcal{F}_t -measurable random variable which is integrable with respect to \mathbb{Q}_T . Then we have:*

$$\tilde{\mathbb{E}}[Y|\mathcal{F}_s] = \frac{1}{\mathcal{E}_s^X} \mathbb{E}[Y\mathcal{E}_t^X|\mathcal{F}_s] \quad \mathbb{P} \text{ and } \mathbb{Q}_T \text{ a.s.}$$

where $\tilde{\mathbb{E}}$ denotes the expectation under \mathbb{Q}_T .

Proof. Given $A \in \mathcal{F}_s$, according to the martingale property of \mathcal{E}^X under \mathbb{P} ,

$$\begin{aligned} \tilde{\mathbb{E}}[Y\mathbf{1}_A] &= \mathbb{E}[Y\mathcal{E}_T^X\mathbf{1}_A] = \mathbb{E}[Y\mathcal{E}_t^X\mathbf{1}_A] \\ &= \mathbb{E}[\mathbb{E}[Y\mathcal{E}_t^X|\mathcal{F}_s]\mathbf{1}_A] = \mathbb{E}\left[\frac{\mathcal{E}_T^X}{\mathcal{E}_s^X} \mathbb{E}[Y\mathcal{E}_t^X|\mathcal{F}_s]\mathbf{1}_A\right] \\ &= \tilde{\mathbb{E}}\left[\frac{1}{\mathcal{E}_s^X} \mathbb{E}[Y\mathcal{E}_t^X|\mathcal{F}_s]\mathbf{1}_A\right]. \end{aligned}$$

The result thus follows. □

The following result is an important corollary of Lemma 3.9.

Corollary 3.2. *Let $M = \{M_t : 0 \leq t \leq T\}$ be an $\{\mathcal{F}_t\}$ -adapted process which is integrable with respect to \mathbb{Q}_T . Then M is a martingale under \mathbb{Q}_T if and only if $M \cdot \mathcal{E}^X$ is a martingale under \mathbb{P} .*

Proof. According to Lemma 3.9, one has

$$\mathcal{E}_s^X \cdot \tilde{\mathbb{E}}[M_t|\mathcal{F}_s] = \mathbb{E}[M_t\mathcal{E}_t^X|\mathcal{F}_s].$$

Therefore,

$$\tilde{\mathbb{E}}[M_t|\mathcal{F}_s] = M_s \iff \mathbb{E}[M_t\mathcal{E}_t^X|\mathcal{F}_s] = M_s\mathcal{E}_s^X.$$

□

The core of the proof of Girsanov's theorem is the following relation on martingale transformations.

Proposition 3.8. *Let $T > 0$ be fixed. Suppose that the processes X^i ($1 \leq i \leq d$) are uniformly bounded. Given any continuous \mathbb{P} -martingale $M = \{M_t : 0 \leq t \leq T\}$ with finite moments of all orders, we define the transformed process \tilde{M} by*

$$\tilde{M}_t \triangleq M_t - \sum_{i=1}^d \int_0^t X_s^i d\langle M, B^i \rangle_s, \quad 0 \leq t \leq T.$$

Then \tilde{M} is a square integrable martingale under \mathbb{Q}_T . In addition, the map $M \mapsto \tilde{M}$ preserves the bracket process:

$$\langle \tilde{M}, \tilde{N} \rangle^{\mathbb{Q}_T} = \langle M, N \rangle^{\mathbb{P}}.$$

Proof. To prove the first claim, by Corollary 3.2 it is equivalent to showing that $\tilde{M} \cdot \mathcal{E}^X$ is a martingale under \mathbb{P} . This is a consequence of the integration by parts formula (cf. (3.34)). To simplify notation we perform the calculation in differential form: under \mathbb{P} one has

$$\begin{aligned} d(\tilde{M}_t \mathcal{E}_t^X) &= \tilde{M}_t d\mathcal{E}_t^X + \mathcal{E}_t^X d\tilde{M}_t + d\mathcal{E}_t^X \cdot d\tilde{M}_t \\ &= \sum_{i=1}^d \tilde{M}_t \mathcal{E}_t^X X_t^i dB_t^i + \mathcal{E}_t^X dM_t - \sum_{i=1}^d \mathcal{E}_t^X X_t^i d\langle M, B^i \rangle_t \\ &\quad + \left(\sum_{i=1}^d \mathcal{E}_t^X X_t^i dB_t^i \right) \cdot \left(dM_t - \sum_{j=1}^d X_s^j d\langle M, B^j \rangle_s \right) \\ &= \sum_{i=1}^d \tilde{M}_t \mathcal{E}_t^X X_t^i dB_t^i + \mathcal{E}_t^X dM_t - \sum_{i=1}^d \mathcal{E}_t^X X_t^i d\langle M, B^i \rangle_t \\ &\quad + \sum_{i=1}^d \mathcal{E}_t^X X_t^i d\langle M, B^i \rangle_t \\ &= \sum_{i=1}^d \tilde{M}_t \mathcal{E}_t^X X_t^i dB_t^i + \mathcal{E}_t^X dM_t, \end{aligned} \tag{3.50}$$

where we have used the relations (3.31) and (3.32) to simplify the products of differentials. Note that the right hand side of (3.50) is a martingale as it consists of stochastic integrals. Therefore, $\tilde{M} \mathcal{E}^X$ is a martingale under \mathbb{P} .

The idea of proving the second claim is similar. Given two \mathbb{P} -martingales M, N , one needs to show that $\tilde{M}\tilde{N} - \langle M, N \rangle$ is a \mathbb{Q} -martingale, or equivalently, $(\tilde{M}\tilde{N} - \langle M, N \rangle)\mathcal{E}^X$ is a \mathbb{P} -martingale. This can be seen by similar but lengthier calculation as before: the integration by parts formula reveals that the product $(\tilde{M}\tilde{N} - \langle M, N \rangle)\mathcal{E}^X$ consists of martingale terms only and is thus a martingale. \square

Remark 3.13. The boundedness and integrability assumptions made in Proposition 3.8 is just for technical convenience to avoid the use of a localisation argument (one checks that all the relevant stochastic integrals satisfy the strong integrability condition and are thus martingales). These assumptions are not essential and the result remains valid in the general context of local martingales (cf. [10, Sec. 3.5.B]).

We are now ready to give the proof of Girsanov's theorem.

Proof of Theorem 3.10. From Theorem 3.8, we know that the process

$$\tilde{B}_t^i \triangleq B_t^i - \sum_{j=1}^d \int_0^t X_s^j d\langle B^i, B^j \rangle_s = B_t^i - \int_0^t X_s^i ds$$

is a martingale under \mathbb{Q} . In addition,

$$\langle \tilde{B}^i, \tilde{B}^j \rangle_t^{\mathbb{Q}} = \langle B^i, B^j \rangle_t^{\mathbb{P}} = \delta_{ij}t.$$

According to Lévy's characterization theorem, we conclude that $\tilde{B} = \{(\tilde{B}_t^1, \dots, \tilde{B}_t^d) : 0 \leq t \leq T\}$ is a d -dimensional, $\{\mathcal{F}_t\}$ -Brownian motion under \mathbb{Q}_T . \square

Remark 3.14. So far we have been working on the given fixed time horizon $[0, T]$. As a consequence of the martingale property, it is not hard to see that $\mathbb{Q}_{T_2}|_{\mathcal{F}_{T_1}} = \mathbb{Q}_{T_1}$ for any $T_1 < T_2$. One may wonder if there exists a single probability measure \mathbb{Q} on the "ultimate σ -algebra" $\mathcal{F}_\infty \triangleq \sigma(\cup_{t \geq 0} \mathcal{F}_t)$ such that $\mathbb{Q}|_{\mathcal{F}_T} = \mathbb{Q}_T$ for all $T > 0$ and the process \tilde{B} is a Brownian motion on $[0, \infty)$ under \mathbb{Q} . This is *not* true on general filtered probability spaces. It can be achieved though on the canonical path space equipped with the natural filtration of the coordinate process. Let Ω denote the space of continuous paths $w : [0, \infty) \rightarrow \mathbb{R}$ and let $B : B_t(w) \triangleq w_t$ be the coordinate process. We consider the filtered probability space $(\Omega, \mathcal{F}_\infty^B, \mathbb{P}; \{\mathcal{F}_t^B\})$ where $\{\mathcal{F}_t^B : t \geq 0\}$ is the natural filtration associated with B and \mathbb{P} is the law of Brownian motion. Let $c \neq 0$ be a fixed number (the drift). One can show that there exists a unique probability measure \mathbb{Q} on \mathcal{F}_∞^B such that

$$\tilde{B}_t \triangleq B_t - ct$$

is a Brownian motion on $[0, \infty)$ under \mathbb{Q} . This measure \mathbb{Q} extends the previous \mathbb{Q}_T 's to \mathcal{F}_∞^B . However, unlike each \mathbb{Q}_T which is absolutely continuous with respect to \mathbb{P} with density \mathcal{E}_T^X (on \mathcal{F}_T^B !), the measure \mathbb{Q} is *singular* to \mathbb{P} on \mathcal{F}_∞^B in the sense that there exists a measurable subset $\Lambda \in \mathcal{F}_\infty^B$ such that $\mathbb{Q}(\Lambda) = 1$ while $\mathbb{P}(\Lambda) = 0$! Can you find an example of such a Λ ?

4 Stochastic differential equations

A substantial part of stochastic calculus is related to the study of stochastic differential equations (SDEs). They provide natural ways for modelling the time evolution of systems that are subject to random effects. Apart from the applied side, there is also an important motivation from the mathematical perspective which we now describe.

Let \mathcal{A} be a second order differential operator on \mathbb{R}^n (acting on smooth functions on \mathbb{R}^n) defined by

$$\mathcal{A} = \frac{1}{2} \sum_{i,j=1}^n a^{ij}(x) \frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^n b^i(x) \frac{\partial}{\partial x^i},$$

where $a^{ij}(x), b^i(x)$ are given functions on \mathbb{R}^n . There are two basic questions one can raise naturally.

Question 1: How can one construct a Markov process X whose *generator* is \mathcal{A} , in the sense that

$$\lim_{t \rightarrow 0} \frac{1}{t} (\mathbb{E}[f(X_t) | X_0 = x] - f(x)) = (\mathcal{A}f)(x)$$

for all suitably regular functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$?

Question 2: How can one construct the *fundamental solution* to the parabolic PDE $\frac{\partial u}{\partial t} - \mathcal{A}^*u = 0$? Here \mathcal{A}^* denotes the *formal adjoint* of the operator \mathcal{A} , in the sense that

$$\int_{\mathbb{R}^n} (\mathcal{A}f)(x)g(x)dx = \int_{\mathbb{R}^n} f(x)(\mathcal{A}^*g)(x)dx$$

for all smooth functions f, g with compact support. The *fundamental solution* is the smallest positive solution to the equation

$$\begin{cases} \frac{\partial p}{\partial t}(t, x, y) - \mathcal{A}_y^*p(t, x, y) = 0, & t > 0; \\ p(0, x, y) = \delta_x(y), \end{cases} \quad (4.1)$$

where \mathcal{A}_y^* means that the differential operator acts on the y variable and δ_x is the Dirac delta function at x .

The first question, which is purely probabilistic, is of fundamental importance in the theory of Markov processes. The second question, which is purely analytical, is of fundamental importance in PDE theory. It is a remarkable fact that these two questions are essentially equivalent. At a formal level, if a Markov process $X =$

$\{X_t^x : t \geq 0, x \in \mathbb{R}^n\}$ (x records the starting position of X) solves Question 1 with a suitably regular transition density function $p(t, x, y) \triangleq \mathbb{P}(X_t^x \in dy)/dy$, then $p(t, x, y)$ solves Question 2. Conversely, if $p(t, x, y)$ is a solution to Question 2, then one can use standard methods in stochastic processes (Kolmogorov's extension theorem) to construct a Markov process with transition density function $p(t, x, y)$ and this Markov process solves Question 1.

It was originally suggested by P. Lévy that a probabilistic approach to these two questions could be possible. K. Itô and P. Malliavin carried out this program in a series of far-reaching works, in which the theory of SDEs was created and largely developed. The philosophy of the SDE approach can be summarised as follows. Let $a(x) = \sigma(x) \cdot \sigma^T(x)$ with some matrix-valued function $\sigma(x)$. Suppose that there exists a stochastic process X_t which solves the following SDE (written in matrix notation):

$$\begin{cases} dX_t^x = \sigma(X_t^x)dB_t + b(X_t^x)dt, & t \geq 0, \\ X_0 = x, \end{cases}$$

Then the Markov process $X = \{X_t^x\}$ solves Question 1, and its transition density function $p(t, x, y) \triangleq \mathbb{P}(X_t^x \in dy)/dy$ (if it exists and is suitably regular) solves Question 2. Even if the density function $p(t, x, y)$ fails to exist, the equivalence between the two questions remains valid as long as we interpret $p(t, x, y)$ in the distributional sense as a *generalised function*.

The above picture provides a natural mathematical motivation to develop the SDE theory in depth. This is the main theme of the present chapter.

4.1 Itô's theorem of existence and uniqueness

Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space and let $B = \{B_t : t \geq 0\}$ be a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion. The most classical and useful type of SDEs takes the form

$$dX_t = \sigma(t, X_t)dB_t + b(t, X_t)dt \tag{4.2}$$

with some initial condition given by a random variable $\xi \in \mathcal{F}_0$. Here X_t takes values in \mathbb{R}^n , the coefficients

$$\sigma : [0, \infty) \times \mathbb{R}^n \rightarrow \text{Mat}(n, d), \quad b : [0, \infty) \times \mathbb{R}^n \rightarrow \text{Mat}(n, 1)$$

are given functions where $\text{Mat}(n, d)$ denotes the space of $n \times d$ matrices. The equation (4.2) is thus written in matrix form. The notion of a solution should be understood in the following integral sense.

Definition 4.1. A stochastic process $X = \{X_t : t \geq 0\}$ is said to be a *solution* to the SDE (4.2) with initial condition ξ if it satisfies the following properties:

- (i) X is $\{\mathcal{F}_t\}$ -progressively measurable;
- (ii) on each finite interval, the process $t \mapsto \sigma(t, X_t)$ is Itô integrable and the process $t \mapsto b(t, X_t)$ is (pathwisely) Lebesgue integrable;
- (iii) X satisfies the following integral equation:

$$X_t^i = \xi^i + \sum_{j=1}^d \int_0^t \sigma_j^i(s, X_s) dB_s^j + \int_0^t b^i(s, X_s) ds, \quad t \geq 0, \quad i = 1, \dots, n. \quad (4.3)$$

When comes to the study of differential equations, before investigating solution properties one should first addresses its existence and uniqueness. In ODE theory, the standard assumption to ensure existence and uniqueness is the Lipschitz condition on the coefficient function. The same principle is true in the SDE case, and this is the content of Itô's theorem which we elaborate in this section. For simplicity, we assume that the Brownian motion and the equation are both one-dimensional. There are no essential difficulties to extend the result to higher dimensions.

Definition 4.2. A function $f : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ is said to

- (i) be *Lipschitz* in space if there exists a constant $K > 0$ such that

$$|f(t, x) - f(t, y)| \leq K|x - y| \quad \forall x, y, t; \quad (4.4)$$

- (ii) have *linear growth* in space if there exists a constant $K > 0$ such that

$$|f(t, x)| \leq K(1 + |x|) \quad \forall x, t. \quad (4.5)$$

Itô's existence and uniqueness theorem is stated as follows.

Theorem 4.1. *Suppose that the coefficient functions σ, b are Lipschitz and have linear growth in space. Given any square integrable initial condition $\xi \in \mathcal{F}_0$, there exists a unique solution to the SDE (4.2) in the sense of Definition 4.2.*

The simplest class of examples that satisfy the conditions in the theorem are linear SDEs. In this case, solutions can even be constructed explicitly (cf. Section 4.3 below). In a deeper way, the linear growth condition (4.5) ensures that the solution is finite for all time (non-explosion), while the Lipschitz condition (4.4) is to guarantee uniqueness. If one only assumes continuity of the coefficients, it is possible that a solution is only defined up to a finite time and then explodes to infinity. This phenomenon already appears in the ODE case as seen from the example below.

Example 4.1 (Explosion). Consider the equation

$$x_t = 1 + \int_0^t x_s^2 ds.$$

The unique solution is given by $x_t = \frac{1}{1-t}$. This solution is only defined on $[0, 1)$ and explodes to infinity at time $t = 1$.

The following result is a useful generalisation of Theorem 4.1.

Theorem 4.2. *Suppose that σ, b are continuous functions and locally Lipschitz in space, i.e. for any $x \in \mathbb{R}^n$ there exists a neighbourhood U_x of x in which σ, b are Lipschitz. Then for each given initial condition $\xi \in \mathcal{F}_0$, there exists a unique solution $X = \{X_t : t < e\}$ to the SDE (4.2) which is defined up to an $\{\mathcal{F}_t\}$ -stopping time e (known as the explosion time) and one has*

$$\lim_{t \uparrow e} |X_t| = \infty \quad \text{on } \{e < \infty\}.$$

In addition, if the linear growth condition (4.5) holds for σ, b , then the solution does not explode to infinity in finite time, i.e. $\mathbb{P}(e = \infty) = 1$.

Remark 4.1. A simple sufficient condition for Theorem 4.2 to hold (with possible explosion) is that σ, b are continuously differentiable functions.

Remark 4.2. The best way to understand Theorem 4.2 is to study the so-called *Yamada-Watanabe theorem* in depth. The theory asserts that

Strong "existence & uniqueness" \iff Weak existence + Pathwise uniqueness.

It is then seen that

Continuity of $\sigma, b \implies$ Weak existence

and

Local Lipschitz \implies Pathwise uniqueness.

We refer the reader to [9, Chap. IV] for the discussion of this deeper theory.

If one further relaxes the conditions on the coefficient functions, there is no guarantee on existence or uniqueness. These phenomena are again present in the ODE case, and with no surprise they will prevail in the stochastic context. We give two ODE examples, one for non-existence and one for non-uniqueness.

Example 4.2 (Non-existence). Consider the equation

$$x_t = \int_0^t f(x_s) ds \quad \text{where } f(x) \triangleq \begin{cases} 1, & x \leq 1 \\ -1, & x > 1. \end{cases}$$

Suppose that a solution x_t exists. Since $x_0 = 0$, before path x_t reaches the level $x = 1$ we have $x'_t = 1$. As a result, $x_t = t$ when $0 \leq t \leq 1$. We claim that $x_t = 1$ for all $t > 1$. Assume on the contrary that $x_{t_2} < 1$ for some $t_2 > 1$. Let $t_1 \triangleq \sup\{t < t_2 : x_t \geq 1\}$. Then $x_{t_1} = 1$ and $x_t < 1$ on $(t_1, t_2]$. It follows that $x'_t = 1$ on $[t_1, t_2]$ which clearly contradicts the fact that $x_{t_2} < x_{t_1}$. Therefore, $x_t \geq 1$ for all $t \geq 1$. Similarly, $x_t \leq 1$ on $[1, \infty)$ and thus $x_t = 1$ on this part. But this is impossible as it would imply $0 = x'_t = f(1) = 1$.

Example 4.3 (Non-uniqueness). Consider the equation

$$x_t = \int_0^t |x_s|^\alpha ds$$

where $\alpha \in (0, 1/2)$. Then both $x_t = 0$ and $x_t = ((1 - \alpha)t)^{\frac{1}{1-\alpha}}$ are solutions.

The rest of this section is devoted to the proof of Theorem 4.1.

The proof of Itô's theorem

The core ingredient of the proof is the following estimate, which is an immediate consequence of Doob's L^p -inequality for submartingales. Given a stochastic process $\{X_t : t \geq 0\}$, we introduce the notation

$$X_t^* \triangleq \sup_{0 \leq s \leq t} |X_s|, \quad t \geq 0.$$

Lemma 4.1. *Let $X = \{X_t : t \geq 0\}$ be an Itô process of the form*

$$X_t = \xi + \int_0^t \Phi_s dB_s + \int_0^t \Psi_s ds.$$

Then for each $T > 0$, there exists a constant $C_T > 0$ depending only on T , such that

$$\mathbb{E}[(X_t^*)^2] \leq C_T (\mathbb{E}[\xi^2] + \mathbb{E}[\int_0^t (\Phi_s^2 + \Psi_s^2) ds]) \quad \forall t \in [0, T].$$

Proof. We assume that the stochastic integral term is a square integrable martingale but the result is true in general. We first observe that

$$X_t^* \leq |\xi| + \sup_{0 \leq s \leq t} \left| \int_0^s \Phi_s dB_s \right| + \int_0^t |\Psi_s| ds,$$

and by applying Hölder's inequality to the last integral we find

$$\begin{aligned} \mathbb{E}[(X_t^*)^2] &\leq 3(|\xi|^2 + \sup_{0 \leq s \leq t} \left| \int_0^s \Phi_s dB_s \right|^2 + \left(\int_0^t |\Psi_s| ds \right)^2) \\ &\leq 3(|\xi|^2 + \sup_{0 \leq s \leq t} \left| \int_0^s \Phi_s dB_s \right|^2 + T \left(\int_0^t |\Psi_s|^2 ds \right)). \end{aligned}$$

The result thus follows from Corollary (1.1) (with $p = 2$) applied to the non-negative submartingale $\mathcal{I}(\Phi)^2$ and Itô's isometry. \square

With the aid of Lemma 4.1, the proof of Theorem 4.1 follows the same lines as in the ODE case. For the existence part, one uses Picard's iteration. For uniqueness, one relies on the Lipschitz condition and Gronwall's lemma. We use the same K to denote both constants appearing in the Lipschitz condition (4.4) and the linear growth condition (4.5).

We first prove uniqueness. Suppose that X, Y are both solutions to the SDE (4.2). Then

$$X_t - Y_t = \int_0^t (\sigma(s, X_s) - \sigma(s, Y_s)) dB_s + \int_0^t (b(s, X_s) - b(s, Y_s)) ds.$$

Given fixed $T > 0$, according to Lemma 4.1 and the Lipschitz condition (4.4), we have

$$\begin{aligned} \mathbb{E}[\left((X - Y)_t^* \right)^2] &\leq C_T \mathbb{E} \left[\int_0^t (\sigma(s, X_s) - \sigma(s, Y_s))^2 + (b(s, X_s) - b(s, Y_s))^2 \right] ds \\ &\leq 2C_T K^2 \int_0^t |X_s - Y_s|^2 ds \\ &\leq 2C_T K^2 \int_0^t \mathbb{E}[\left((X - Y)_s^* \right)^2] ds \end{aligned} \tag{4.6}$$

for any $t \in [0, T]$.

To proceed further, we first introduce a useful tool known as *Gronwall's lemma*.

Lemma 4.2. *Let $f : [0, \infty) \rightarrow \mathbb{R}$ be a non-negative continuous function. Suppose that there exists a constant $C > 0$, such that*

$$f(t) \leq C \int_0^t f(s) ds \quad \forall t \geq 0. \quad (4.7)$$

Then $f = 0$.

Proof. Since f is continuous, it is uniformly bounded, say $0 \leq f(t) \leq M$ for all t . By iterating the assumption (4.7), one finds

$$\begin{aligned} f(t) &\leq C \int_0^t (C \int_0^s f(v) dv) ds \leq C \int_0^t (C \int_0^s (C \int_0^v f(u) du) dv) ds \leq \dots \\ &\leq C^n \int_{0 < t_1 < \dots < t_n < t} f(t_1) dt_1 \dots dt_n \leq MC^n \int_{0 < t_1 < \dots < t_n < t} dt_1 \dots dt_n. \end{aligned}$$

We claim that

$$\int_{0 < t_1 < \dots < t_n < t} dt_1 \dots dt_n = \frac{t^n}{n!}. \quad (4.8)$$

Indeed, this integral is the volume of the n -dimensional simplex

$$\{(t_1, \dots, t_n) \in [0, t]^n : t_1 < t_2 < \dots < t_n\}.$$

But the n -dimensional cube $[0, t]$ can be divided into $n!$ congruent simplices (each one is obtained by permuting the condition $t_1 < \dots < t_n$). Therefore, as one particular simplex among the $n!$ ones, the formula (4.8) holds. It follows that

$$f(t) \leq \frac{MC^n t^n}{n!}.$$

Since this is true for all n , we conclude that $f = 0$. □

We now return to the uniqueness part (cf. (4.6)). By applying Gronwall's lemma to the function

$$f(t) = \mathbb{E}[\left((X - Y)_t^*\right)^2], \quad 0 \leq t \leq T,$$

we find that $f = 0$. In particular, $\mathbb{E}[\left((X - Y)_T^*\right)^2] = 0$, which implies that with probability one, $X = Y$ as a function of time over $[0, T]$. The uniqueness of solution follows since T is arbitrary.

Next, we prove the existence of solution. The essential idea is summarised conceptually as follows. We regard the right hand side of (4.3) as a transformation sending a given process X to another process

$$\mathcal{T}(X) \triangleq \xi + \int_0^\cdot \sigma(t, X_t)dB_t + \int_0^\cdot b(t, X_t)dt.$$

From this viewpoint, a solution to the SDE is merely a *fixed point* of the transformation \mathcal{T} , i.e. a process X satisfying $\mathcal{T}(X) = X$. The idea of locating the fixed point is very simple: one starts with a stupid initial guess $X^{(0)}$ and define inductively

$$X^{(n+1)} \triangleq \mathcal{T}(X^{(n)}) \tag{4.9}$$

to refine the guess. If we are able to show that the sequence $\{X^{(n)}\}$ converges to some process X , by taking limit in (4.9) we shall have $X = \mathcal{T}(X)$, and thus X is a desired solution. This procedure is known as *Picard's iteration*.

We now develop the mathematical details. It is enough to only work on a fixed interval $[0, T]$. Indeed, if we are able to construct solutions on $[0, T_1]$ and $[0, T_2]$ ($T_1 < T_2$), the uniqueness part implies that the two solutions coincide on the common interval $[0, T_1]$. This allows one to see that solutions defined on the intervals $[0, T]$ (with different T 's) are consistent, hence patching to a global solution on $[0, \infty)$.

Let T be fixed. Recall that $\xi \in \mathcal{F}_0$ is the initial condition. For $t \in [0, T]$ we set

$$X_t^{(0)} \triangleq \xi$$

and inductively define

$$X_t^{(n+1)} \triangleq \xi + \int_0^t \sigma(s, X_s^{(n)})dB_s + \int_0^t b(s, X_s^{(n)})ds, \quad n \geq 1. \tag{4.10}$$

In exactly the same way leading to (4.6), we find that

$$\mathbb{E}[\left((X^{(n+1)} - X^{(n)})_t^*\right)^2] \leq C_1 \int_0^t \mathbb{E}[\left((X^{(n)} - X^{(n-1)})_s^*\right)^2]ds$$

where C_1 is a suitable constant (depending on T and K). By iterating this in-

equality as in the proof of Lemma 4.2, we obtain

$$\begin{aligned} \mathbb{E}[\left((X^{(n+1)} - X^{(n)})_T^*\right)^2] &\leq C_1^n \int_{0 < t_1 < \dots < t_n < T} \mathbb{E}[\left((X^{(1)} - X^{(0)})_{t_1}^*\right)^2] dt_1 \cdots dt_n \\ &\leq C_1^n \mathbb{E}[\left((X^{(1)} - X^{(0)})_T^*\right)^2] \cdot \int_{0 < t_1 < \dots < t_n < T} dt_1 \cdots dt_n \\ &= \frac{C_1^n T^n}{n!} \cdot \mathbb{E}[\left((X^{(1)} - X^{(0)})_T^*\right)^2]. \end{aligned}$$

In addition, Lemma (4.1) together with the linear growth condition (4.5) imply that

$$\mathbb{E}[\left((X^{(1)} - X^{(0)})_T^*\right)^2] \leq C_2(1 + \mathbb{E}[\xi^2])$$

with a suitable constant C_2 . Therefore,

$$\mathbb{E}[\left((X^{(n+1)} - X^{(n)})_T^*\right)^2] \leq \frac{C_2 C_1^n T^n}{n!} (1 + \mathbb{E}[\xi^2]). \quad (4.11)$$

The right hand side of (4.5) defines a convergent series. As a result,

$$\sum_{n=0}^{\infty} \mathbb{E}[\left((X^{(n+1)} - X^{(n)})_T^*\right)^2] = \mathbb{E}\left[\sum_{n=0}^{\infty} \left((X^{(n+1)} - X^{(n)})_T^*\right)^2\right] < \infty.$$

In particular,

$$\sum_{n=0}^{\infty} \left((X^{(n+1)} - X^{(n)})_T^*\right)^2 < \infty \quad \text{a.s.}$$

This implies that with probability one, as continuous functions on $[0, T]$ the sequence $\{X^{(n)} : n \geq 0\}$ is a Cauchy sequence under the uniform distance. Therefore, with probability one $X^{(n)}$ converges uniformly to some continuous process X . By taking $n \rightarrow \infty$ in (4.10), we see that X satisfies the equation (4.3). This gives the existence part of Theorem 4.1.

Remark 4.3. It is not hard to see that the above argument yields the following estimate:

$$\mathbb{E}[(X_T^*)^2] \leq C_{T,K}(1 + \mathbb{E}[\xi^2]),$$

where $C_{T,K}$ is constant depending on T and K .

4.2 Strong Markov property, generator and the martingale formulation

Consider the following multidimensional SDE

$$\begin{cases} dX_t^x = \sigma(X_t^x)dB_t + b(X_t^x)dt, \\ X_0^x = x \in \mathbb{R}^n \end{cases} \quad (4.12)$$

written in matrix form and assume that the conditions of Theorem 4.1 are met. According to Theorem 4.1, the SDE (4.12) admits a unique solution for all time. Here we have assumed that the coefficients σ, b do not depend on time. The time-dependent situation can be reduced to the current case by introducing an extra (trivial) equation $dt = dt$ and regarding $t \mapsto (t, X_t^x)$ as the solution process.

Definition 4.3. The process $X = \{X_t^x : x \in \mathbb{R}^n, t \geq 0\}$ is called a (time-homogeneous) *Itô diffusion process* with diffusion coefficient σ and drift coefficient b .

Diffusion processes provide a rich and important class of strong Markov processes. Recall that a Markov process refreshes at any given deterministic time. We have defined the Markov property by (2.11) in Section 2.2. Similarly, a strong Markov process refreshes at any given stopping time. Mathematically, the strong Markov property is stated as

$$\mathbb{P}(X_{\tau+t} \in \Gamma | \mathcal{F}_\tau) = \mathbb{P}(X_{\tau+t} \in \Gamma | X_\tau) \quad \forall t \geq 0, \Gamma \in \mathcal{B}(\mathbb{R}^n)$$

where τ is any given finite $\{\mathcal{F}_t\}$ -stopping time. Heuristically, the Markov property of the solution $\{X_t^x\}$ is a simple consequence of uniqueness: the solution for the future is uniquely determined by the current location as a refreshed initial condition and forgets the entire past.

Theorem 4.3. *Itô diffusion processes are strong Markov processes.*

Proof. Let τ be a given finite $\{\mathcal{F}_t\}$ -stopping time. For any $t \geq 0$, we have

$$\begin{aligned} X_{\tau+t}^x &= x + \int_0^{\tau+t} \sigma(X_s^x)dB_s + \int_0^{\tau+t} b(X_s^x)ds \\ &= X_\tau^x + \int_\tau^{\tau+t} \sigma(X_s^x)dB_s + \int_\tau^{\tau+t} b(X_s^x)ds \\ &= X_\tau^x + \int_0^t \sigma(X_{\tau+u}^x)dB_u^{(\tau)} + \int_0^t b(X_{\tau+u}^x)du, \end{aligned} \quad (4.13)$$

where $B_u^{(\tau)} \triangleq B_{\tau+u} - B_\tau$. The key observation is that $t \mapsto X_{\tau+t}^x$ is the unique solution to the same SDE driven by the Brownian motion $B^{(\tau)}$ with initial condition X_τ^x . To elaborate this, we introduce the function $S(t, x, B)$ to denote the solution at time t for the SDE driven by B with initial condition x . Then (4.13) reads

$$X_{\tau+t}^x = S(t, X_\tau^x, B^{(\tau)}).$$

Since X_τ^x is \mathcal{F}_τ -measurable and $B^{(\tau)}$ is independent of \mathcal{F}_τ by the strong Markov property of Brownian motion, the conditional distribution of $X_{\tau+t}^x$ given \mathcal{F}_τ is uniquely determined by X_τ^x as well as the distribution of Brownian motion through the function S . This clearly implies the strong Markov property. \square

In the study of Markov processes, it is often important to consider the analytical perspective. Let $X = \{X_t^x\}$ be a given Itô diffusion process on \mathbb{R}^n . We use $C_b(\mathbb{R}^n)$ (respectively, $C_b^2(\mathbb{R}^n)$) to denote the space of functions on \mathbb{R}^n that are bounded and continuous (respectively, have bounded derivatives up to order two).

Definition 4.4. The *transition semigroup* of X is the family of linear operators $P_t : C_b(\mathbb{R}^n) \rightarrow C_b(\mathbb{R}^n)$ ($t \geq 0$) defined by

$$(P_t f)(x) \triangleq \mathbb{E}[f(X_t^x)], \quad f \in C_b(\mathbb{R}^n).$$

The *generator* of X is the linear operator

$$\mathcal{A}f \triangleq \lim_{t \rightarrow 0} \frac{P_t f - f}{t}$$

defined for those f 's for which the above limit exists in $C_b(\mathbb{R}^n)$.

Remark 4.4. Suppose that $\{X_n : n \in \mathbb{N}\}$ is a Markov chain with countable state space S and n -step transition probabilities

$$p_n(x, y) = \mathbb{P}(X_n = y | X_0 = x), \quad x, y \in S.$$

By arranging $(p_n(x, y))_{x, y \in S}$ as an $|S| \times |S|$ matrix, one obtains a linear transformation $P_n : \mathbb{R}^{|S|} \rightarrow \mathbb{R}^{|S|}$. Note that functions on S can be identified as (column) vectors in $\mathbb{R}^{|S|}$. From this perspective, the transition semigroup is a continuous-time extension of the notion of n -step transition probabilities. As a result, it encodes all information about the distribution of the Markov process $\{X_t^x\}$.

The precise study of semigroups, generators and their relationship relies on basic tools from Banach spaces. Here we only outline the essential ideas at a

semi-rigorous level. It is clear that $P_0 f = f$. A crucial property of the semigroup $\{P_t : t \geq 0\}$ is that

$$P_{s+t} = P_s \circ P_t \quad (4.14)$$

as seen from the Markov property:

$$\begin{aligned} P_{t+s}f(x) &= \mathbb{E}[f(X_{s+t}^x)] = \mathbb{E}[\mathbb{E}[f(X_{s+t}^x)|\mathcal{F}_s]] \\ &= \mathbb{E}[P_t f(X_s^x)] = (P_s \circ P_t f)(x). \end{aligned}$$

On the other hand, by the definition of the generator one has $\mathcal{A}f = \frac{dP_t}{dt}|_{t=0}$. It follows from the semigroup property (4.14) that

$$\frac{dP_t}{dt} = \lim_{s \rightarrow 0^+} \frac{P_{s+t} - P_t}{s} = \left(\lim_{s \rightarrow 0^+} \frac{P_s - \text{Id}}{s} \right) \circ P_t = \mathcal{A}P_t. \quad (4.15)$$

One can regard (4.15) as a linear ODE for P_t . Its solution is formally given by $P_t = e^{t\mathcal{A}}$. From this viewpoint, it is reasonable to expect that the semigroup (equivalently, the entire distribution of X) is uniquely determined by its generator \mathcal{A} . This explains why the generator plays an essential role in the study of Markov processes. The precise reconstruction of the semigroup $\{P_t\}$ from the generator \mathcal{A} is the content of the Hille-Yosida theorem in functional analysis (cf. [17, Sec. III.5]).

In the context of Itô diffusions, the generator \mathcal{A} can be computed explicitly by using Itô's formula.

Proposition 4.1. *Let $X = \{X_t^x\}$ be an Itô diffusion process define by the SDE (4.12). The generator of X is the second order differential operator given by*

$$(\mathcal{A}f)(x) \triangleq \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum_{i=1}^n b^i(x) \frac{\partial f(x)}{\partial x^i}, \quad f \in C_b^2(\mathbb{R}^n),$$

where $a(x)$ is the $n \times n$ matrix defined by $a(x) = \sigma(x) \cdot \sigma(x)^T$. In addition, for any given $f \in C_b^2(\mathbb{R}^n)$, the process

$$M_t^f \triangleq f(X_t^x) - f(x) - \int_0^t (\mathcal{A}f)(X_s^x) ds, \quad t \geq 0$$

is an $\{\mathcal{F}_t\}$ -martingale.

Proof. According to Itô's formula,

$$\begin{aligned} f(X_t^x) &= f(x) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s^x) dB_s^i + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s^x) b^i(X_s) ds \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s^x) \sigma_k^i(X_s^x) \sigma_k^j(X_s^x) ds, \end{aligned} \quad (4.16)$$

where we have used the relation

$$dX_t^i \cdot dX_t^j = \left(\sum_{k=1}^d \sigma_k^i dB_t^k + b^i dt \right) \times \left(\sum_{l=1}^d \sigma_l^j dB_t^l + b^j dt \right) = \sum_{k=1}^d \sigma_k^i \sigma_k^j dt.$$

Since the stochastic integral in (4.16) is a martingale, one finds that

$$\begin{aligned} \mathbb{E}[f(X_t^x)] &= f(x) + \sum_{i=1}^n \int_0^t \mathbb{E} \left[\frac{\partial f}{\partial x_i}(X_s^x) b^i(X_s^x) \right] ds \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \int_0^t \mathbb{E} \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(X_s^x) \sigma_k^i(X_s^x) \sigma_k^j(X_s^x) \right] ds. \end{aligned}$$

Now observe that

$$\begin{aligned} \frac{1}{t} \int_0^t \mathbb{E} \left[\frac{\partial f}{\partial x_i}(X_s^x) b^i(X_s^x) \right] ds &\rightarrow \frac{\partial f(x)}{\partial x_i} b^i(x), \\ \frac{1}{t} \int_0^t \mathbb{E} \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(X_s^x) \sigma_k^i(X_s^x) \sigma_k^j(X_s^x) \right] ds &\rightarrow \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \sigma_k^i(x) \sigma_k^j(x) \end{aligned}$$

as $t \rightarrow 0$ (why?). The first assertion of the proposition thus follows from the definition of the generator. The second assertion is an immediate consequence of the fact that M_t^f is the stochastic integral term in (4.16). \square

Example 4.4. A d -dimensional Brownian motion can be viewed as the solution to the trivial SDE $dB_t = dB_t$. In particular, $\sigma = \text{Id}$ and $b = 0$. According to Proposition 4.1, its generator is $\frac{1}{2}\Delta$ where $\Delta \triangleq \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_d^2}$ denotes the Laplace operator on \mathbb{R}^d .

Based on Proposition 4.1, the drift and diffusion coefficients have a simple interpretation. To simplify notation, we only consider the one-dimensional case. By the definition of \mathcal{A} , one has the following approximation:

$$(P_h f)(x) \approx f(x) + (\mathcal{A}f)(x)h$$

when h is small. By taking $f(x) = x$, one finds that $\mathcal{A}f(x) = b(x)$. Therefore,

$$\mathbb{E}[X_{t+h} - X_t | X_t = x] = (P_h f)(x) - f(x) \approx b(x)h$$

when h is small. In other words, $b(x)$ is the average “velocity vector” for the diffusion at position x . Similarly, one also finds that

$$\mathbb{E}[(X_{t+h} - X_t)^2 | X_t = x] \approx a(x)h.$$

As a result, starting at position x , in \sqrt{h} amount of time the diffusion will travel a distance of $\sqrt{a(x)h} = |\sigma(x)|\sqrt{h}$ on average.

Remark 4.5. In Proposition 4.1, the fact that M_t^f is a martingale for any $f \in C_b^2(\mathbb{R}^n)$, which is also known as the *martingale formulation* of the SDE (4.12), is of fundamental importance and has far-reaching applications in stochastic calculus as well as PDE theory. Indeed, this property uniquely characterises the distribution of the solution X . As a result, solving an SDE in a distributional sense is essentially equivalent to finding a distribution under which the aforementioned martingale property holds. This is a deep and rich topic in the theory of weak solutions to SDEs (*the martingale problem of Stroock-Varadhan*). We refer the reader to the monograph [20] for a thorough discussion.

4.3 Explicit solutions to linear SDEs

The simplest type of SDE examples are linear equations. In this case, solutions exist uniquely and can be constructed explicitly. For simplicity, we only consider the one-dimensional situation. A *one-dimensional linear SDE* takes the following general form:

$$dX_t = (A(t)X_t + a(t))dt + (C(t)X_t + c(t))dB_t \quad (4.17)$$

where $A, a, C, c : [0, \infty) \rightarrow \mathbb{R}$ are bounded, deterministic functions. It is plain to check that the conditions of Theorem 4.1 are satisfied. As a result, there exists a unique solution to the equation.

To solve the SDE (4.17) explicitly, just like the ODE case the key idea is to use an appropriate integrating factor to eliminate the linear dependence terms $A(t)X_t dt$ and $C(t)X_t dB_t$. To understand the method better, we first recapture the ODE situation:

$$\frac{dx_t}{dt} = A(t)x_t + a(t). \quad (4.18)$$

In order to solve (4.18), we introduce the *integrating factor* $z_t = e^{\int_0^t A(s)ds}$ so that when differentiating $z_t^{-1}x_t$ the linear term $A(t)x_t$ is eliminated. More explicitly, one finds

$$(z_t^{-1}x_t)' = z_t^{-1}a(t).$$

As a result, the solution to the ODE (4.18) is given by

$$x_t = z_t(x_0 + \int_0^t z_s^{-1}a(s)ds).$$

For the SDE (4.17), since there is an extra linear term $C(t)X_tdB_t$ apart from $A(t)X_tdt$, one should naturally add a “stochastic exponential” into the previous integrating factor z_t . Our multiple experience suggests that this stochastic exponential should be given by the exponential martingale

$$\exp\left(\int_0^t C(s)dB_s - \frac{1}{2}\int_0^t C(s)^2ds\right).$$

As a result, the *stochastic integrating factor* should be defined by

$$Z_t \triangleq \exp\left(\int_0^t A(s)ds + \int_0^t C(s)dB_s - \frac{1}{2}\int_0^t C(s)^2ds\right).$$

Now we proceed to check if both linear terms $A(t)X_tdt$ and $C(t)X_tdB_t$ are eliminated in the process $Z_t^{-1}X_t$. We first use Itô’s formula to obtain

$$dZ_t^{-1} = Z_t^{-1}\left(-A(t)dt - C(t)dB_t + \frac{1}{2}C(t)^2dt\right) + \frac{1}{2}Z_t^{-1}C(t)^2dt.$$

It then follows from the integration by parts formula that

$$\begin{aligned} d(Z_t^{-1}X_t) &= (dZ_t^{-1})X_t + Z_t^{-1}dX_t + (dZ_t^{-1})dX_t \\ &= Z_t^{-1}X_t\left(-A(t)dt - C(t)dB_t + \frac{1}{2}C(t)^2dt\right) + \frac{1}{2}Z_t^{-1}X_tC(t)^2dt \\ &\quad + Z_t^{-1}\left((A(t)X_t + a(t))dt + (C(t)X_t + c(t))dB_t\right) \\ &\quad + Z_t^{-1}\left(-C(t)dB_t\right)\left((C(t)X_t + c(t))dB_t\right) \\ &= Z_t^{-1}\left((a(t) - C(t)c(t))dt + c(t)dB_t\right). \end{aligned}$$

By integrating the above equation from 0 to t , we find that the solution to the SDE (4.17) is given by

$$X_t = Z_t\left(X_0 + \int_0^t Z_s^{-1}(a(s) - C(s)c(s))ds + \int_0^t Z_s^{-1}c(s)dB_s\right). \quad (4.19)$$

Example 4.5. Let $\gamma, \sigma > 0$ be given constants. The linear SDE

$$dX_t = -\gamma X_t dt + \sigma dB_t$$

is called the *Langevin equation* and the solution is known as the *Ornstein-Uhlenbeck process*. Its generator is given by

$$\mathcal{A} = \frac{\sigma^2}{2} \frac{d^2}{dx^2} - \gamma x \frac{d}{dx}.$$

According to (4.19),

$$X_t = X_0 e^{-\gamma t} + \sigma \int_0^t e^{-\gamma(t-s)} dB_s.$$

If X_0 is a Gaussian random variable with mean zero and variance η^2 , then X is a centered Gaussian process (why?) with covariance function

$$\begin{aligned} \rho(s, t) &\triangleq \mathbb{E}[X_s X_t] = e^{-\gamma(s+t)} \left(\eta^2 + \sigma^2 \int_0^s e^{2\gamma u} du \right) \\ &= \left(\eta^2 - \frac{\sigma^2}{2\gamma} \right) e^{-\gamma(s+t)} + \frac{\sigma^2}{2\gamma} e^{-\gamma(t-s)}, \quad s < t. \end{aligned}$$

In particular, if $\eta^2 = \sigma^2/(2\gamma)$, then X is also *stationary* in the sense that the distribution of $(X_{t_1+h}, \dots, X_{t_n+h})$ is independent of h for any given $n \geq 1$ and $t_1 < \dots < t_n$.

Example 4.6. Consider the SDE

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are given constants. The solution is given by

$$X_t = X_0 \exp \left(\mu t + \sigma B_t - \frac{1}{2} \sigma^2 t \right).$$

Its generator is given by

$$\mathcal{A} = \frac{1}{2} \sigma^2 x^2 \frac{d^2}{dx^2} + \mu x \frac{d}{dx}.$$

This is known as the *geometric Brownian motion*. It has important applications in modelling stock prices.

Iterated integrals of Brownian motion: Part II

In Section 3.2.5, we have computed $\int_0^t B_s dB_s$ and $\int_0^t (\int_0^s dB_v) dB_s$. Now we generalise the computation to the case of iterated Itô integrals of arbitrary orders. For each $n \geq 1$, we define

$$\begin{aligned} B_t^{(n)} &\triangleq \int_{0 < t_1 < \dots < t_n < t} dB_{t_1} \cdots dB_{t_n} \\ &\triangleq \int_0^t \left(\int_0^{t_n} \cdots \left(\int_0^{t_3} \left(\int_0^{t_2} dB_{t_1} \right) dB_{t_2} \right) \cdots dB_{t_{n-1}} \right) dB_{t_n}. \end{aligned}$$

To compute $B_t^{(n)}$, we again make use of the exponential martingale

$$Z_t^\lambda \triangleq e^{\lambda B_t - \lambda^2 t/2}, \quad \lambda \in \mathbb{R}^n, t \geq 0.$$

The main idea is to represent Z_t^λ in the following two different ways.

On the one hand, according to Itô's formula, Z_t^λ satisfies the following linear SDE

$$dZ_t^\lambda = \lambda Z_t^\lambda dB_t$$

with $Z_0^\lambda = 1$. The solution to this SDE can be represented in terms of the iterated integral series:

$$\begin{aligned} Z_t^\lambda &= 1 + \lambda \int_0^t Z_s^\lambda dB_s = 1 + \lambda \int_0^t \left(1 + \lambda \int_0^s Z_v^\lambda dB_v \right) dB_s \\ &= 1 + \lambda \int_0^t dB_s + \lambda^2 \int_0^t \left(\int_0^s Z_v^\lambda dB_v \right) dB_s \\ &= 1 + \lambda B_t^{(1)} + \lambda^2 \int_0^t \left(\int_0^s \left(1 + \lambda \int_0^u Z_u^\lambda dB_u \right) dB_v \right) dB_s \quad (4.20) \\ &= 1 + \lambda B_t^{(1)} + \lambda^2 B_t^{(2)} + \lambda^3 \int_0^t \left(\int_0^s \left(\int_0^v Z_u^\lambda dB_u \right) dB_v \right) dB_s \\ &\dots \end{aligned}$$

$$= \sum_{n=0}^{\infty} \lambda^n B_t^{(n)}. \quad (4.21)$$

On the other hand, let us introduce the function $F(x, t) \triangleq e^{xt - t^2/2}$. By viewing x as a parameter and t as the generic variable, one can write down the Taylor expansion of F as

$$F(x, t) = \sum_{n=0}^{\infty} H_n(x) t^n$$

where

$$H_n(x) \triangleq \frac{1}{n!} \frac{\partial F(x, t)}{\partial t} \Big|_{t=0} = \frac{(-1)^n}{n!} e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}.$$

It is not hard to see that $H_n(x)$ is a polynomial of degree n . By explicit calculation, the first few terms are given by

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = \frac{x^2 - 1}{2}, \quad H_3(x) = \frac{x^3 - 3x}{6} \text{ etc.}$$

Under this notation, we have

$$F\left(\frac{B_t}{\sqrt{t}}, \lambda\sqrt{t}\right) = Z_t^\lambda = \sum_{n=0}^{\infty} \lambda^n H_n\left(\frac{B_t}{\sqrt{t}}\right) t^{n/2}. \quad (4.22)$$

By comparing the coefficients of λ^n in (4.21) and (4.22), we conclude that

$$B_t^{(n)} = H_n\left(\frac{B_t}{\sqrt{t}}\right) t^{n/2}, \quad n = 1, 2, 3, \dots$$

Remark 4.6. In vague terms, the exponential martingale $e^{B_t - t^2/2}$ is the stochastic counterpart of the exponential function e^x , and $H_n\left(\frac{B_t}{\sqrt{t}}\right) t^{n/2}$ is the stochastic counterpart of the polynomial $x^n/n!$.

Remark 4.7. The polynomial $H_n(x)$ is known as the n -th *Hermite polynomial* over \mathbb{R} . They can also be obtained by orthogonalising the canonical polynomial system $\{1, x, x^2, \dots\}$ with respect to the standard Gaussian measure $\gamma(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ on \mathbb{R} . The family $\{H_n : n \geq 0\}$ provides the *spectral decomposition* for the generator $\mathcal{A} = \frac{d^2}{dx^2} - x \frac{d}{dx}$ of the standard Ornstein-Uhlenbeck process (cf. Example 4.5 with $\gamma = 1, \sigma = \sqrt{2}$), in the sense that H_n is an eigenfunction of \mathcal{A} with eigenvalue $-n$ (i.e. $\mathcal{A}H_n = -nH_n$) and $\{H_n : n \geq 0\}$ form an orthogonal basis of $L^2(\mathbb{R}, \gamma)$ (the space of square integrable functions with respect to γ). We refer the reader to [7, Sec. 2] for these details.

4.4 One-dimensional diffusions

In this section, we investigate further properties of one-dimensional SDEs. In this case, the behaviour of sample paths can be described more explicitly due to the availability of explicit solutions to suitable second order linear ODEs. The corresponding picture is less clear in multidimensions where the study is largely based on PDE methods.

In what follows, let $I = (l, r) \subseteq \mathbb{R}$ be a fixed open interval, where $-\infty \leq l < r \leq \infty$. We consider the following SDE

$$\begin{cases} dX_t = \sigma(X_t)dB_t + b(X_t)dt, \\ X_0 = x \in I, \end{cases}$$

where σ, b are continuously differentiable functions defined on I . In particular, σ, b are locally Lipschitz since as continuous functions σ', b' are bounded on compact sub-intervals of I . It follows from Theorem 4.2 that there exists a unique solution $X^x = \{X_t^x\}$ defined up to its intrinsic explosion time. Here explosion should be understood as *exiting the region I* , namely X_t^x is defined up to the stopping time

$$e_x \triangleq \inf\{t \geq 0 : X_t^x = l \text{ or } r\}.$$

When $I = (-\infty, \infty)$ this is the usual explosion time to infinity. On the event $\{e_x < \infty\}$, $\lim_{t \uparrow e_x} X_t^x$ exists and is equal to either l or r . According to Proposition 4.1, the generator of X is the second order differential operator given by

$$\mathcal{A}f = \frac{1}{2}\sigma^2 f'' + bf'.$$

From now on, we assume that the diffusion coefficient is *everywhere non-vanishing*, i.e. $\sigma^2 \neq 0$ on I . Heuristically, this ensures that X_t^x is “truly diffusive” and it locally behaves like a Brownian motion.

4.4.1 Exit distribution and behaviour towards explosion

Many basic questions in the study of diffusion processes are related to exit time distributions. Let $[a, b]$ be a fixed sub-interval of I ($a > l$ and $b < r$). Suppose that the starting point $x \in (a, b)$ and we define

$$\tau_x \triangleq \inf\{t \geq 0 : X_t^x \notin [a, b]\}.$$

Our first task is to compute $\mathbb{E}[\tau_x]$.

Proposition 4.2. *Let $M(x)$ ($x \in [a, b]$) be the unique solution to the following second order linear ODE*

$$\begin{cases} \frac{1}{2}\sigma(x)^2 M''(x) + b(x)M'(x) = -1, \\ M(a) = M(b) = 0. \end{cases} \quad (4.23)$$

Then $\mathbb{E}[\tau_x] = M(x)$.

Proof. Note that the ODE (4.23) is just a restatement of $\mathcal{A}M = -1$. By applying Itô's formula to $M(X_t^x)$ on $[0, \tau_x]$, we have

$$\begin{aligned} M(X_t^x) &= M(x) + \int_0^t \sigma(X_s^x) M'(X_s^x) dB_s + \int_0^t (\mathcal{A}M)(X_s^x) ds \\ &= M(x) + \int_0^t \sigma(X_s^x) M'(X_s^x) dB_s - t, \quad t < \tau_x. \end{aligned}$$

By taking expectation on both sides, we obtain

$$\mathbb{E}[M(X_{\tau_x \wedge t})] = M(x) - \mathbb{E}[\tau_x \wedge t] \quad \forall t \geq 0.$$

Since $M(X_{\tau_x}) = 0$ by the boundary conditions of M , the result follows by taking $t \rightarrow \infty$. \square

Proposition 4.2 implies that $\tau_x < \infty$ a.s. As a result, $X_{\tau_x}^x$ is well-defined and is equal to either a or b . Using a similar idea as before, one can compute the distribution of X_{τ_x} . For this purpose, we introduce the function

$$s(x) \triangleq \int_c^x \exp\left(-2 \int_c^y \frac{b(z)}{\sigma^2(z)} dz\right) dy, \quad x \in I,$$

where $c \in I$ is an arbitrary point whose particular choice does not affect any result in the sequel. The form of s comes from solving the homogeneous ODE $\mathcal{A}s = 0$ (one easily checks that $s(x)$ is a particular solution).

Definition 4.5. The function $s(x)$ is known as the *scale function* of the diffusion $\{X_t^x\}$.

Proposition 4.3. *The distribution of X_{τ_x} is given by*

$$\mathbb{P}(X_{\tau_x} = a) = \frac{s(b) - s(x)}{s(b) - s(a)}, \quad \mathbb{P}(X_{\tau_x} = b) = \frac{s(x) - s(a)}{s(b) - s(a)}. \quad (4.24)$$

Proof. By applying Itô's formula to $s(X_t^x)$ on $[0, \tau_x]$, we find

$$\mathbb{E}[s(X_{\tau_x \wedge t}^x)] = s(x) \quad \forall t \geq 0.$$

By taking $t \rightarrow \infty$, we obtain

$$s(x) = \mathbb{E}[s(X_{\tau_x}^x)] = s(a) \cdot \mathbb{P}(X_{\tau_x} = a) + s(b) \cdot \mathbb{P}(X_{\tau_x} = b).$$

Since

$$\mathbb{P}(X_{\tau_x} = a) + \mathbb{P}(X_{\tau_x} = b) = 1,$$

the result follows easily. \square

Remark 4.8. It is elementary to solve the boundary value problem (4.23) explicitly. Represented in a neat way, it can be seen that

$$M(x) = \int_a^b G_{a,b}(x, y)m(dy).$$

Here

$$G_{a,b}(x, y) \triangleq \frac{(s(x \wedge y) - s(a))(s(b) - s(x \vee y))}{s(b) - s(a)}, \quad x, y \in [a, b]$$

is the so-called *Green's function* and

$$m(dy) \triangleq \frac{2dy}{s'(y)\sigma^2(y)}, \quad y \in I$$

is the so-called *speed measure* of the diffusion.

Remark 4.9. The above analysis extends naturally to multidimensional diffusions. In this case, the equations $\mathcal{A}M = -1$ and $\mathcal{A}s = 0$ are PDEs whose explicit solutions are rarely available. Nonetheless, one can still use numerical methods to study their solutions and related exit distributions.

The scale function can also be used to study the behaviour of the diffusion towards explosion. Note that $s(x)$ is a strictly increasing function on I . We denote

$$s(l+) \triangleq \lim_{x \downarrow l} s(x), \quad s(r-) \triangleq \lim_{x \uparrow r} s(x).$$

Theorem 4.4. (i) Suppose that $s(l+) = -\infty$ and $s(r-) = \infty$. Then with probability one, we have

$$e_x = \infty, \quad \overline{\lim}_{t \rightarrow \infty} X_t^x = r, \quad \underline{\lim}_{t \rightarrow \infty} X_t^x = l.$$

(ii) Suppose that $s(l+) > -\infty$ and $s(r-) = \infty$. Then with probability one, we have

$$\lim_{t \uparrow e_x} X_t^x = l, \quad \sup_{t < e_x} X_t^x < r.$$

A parallel conclusion holds if the behaviours at l and r are switched.

(iii) Suppose that $s(l+)$ and $s(r-)$ are both finite. Then

$$\mathbb{P}\left(\lim_{t \uparrow e_x} X_t^x = l\right) = \frac{s(r-) - s(x)}{s(r-) - s(l+)}, \quad \mathbb{P}\left(\lim_{t \uparrow e_x} X_t^x = r\right) = \frac{s(x) - s(l+)}{s(r-) - s(l+)}.$$

Proof. (i) Let $[a, b]$ be an arbitrary sub-interval of I containing the starting point x . Define τ_x as before and we shall now write τ_x as $\tau_{a,b}$ to emphasise its dependence on a, b . Since

$$\{X_{\tau_{a,b}}^x = b\} \subseteq \left\{ \sup_{t < e} X_t^x \geq b \right\}$$

The second identity in (4.24) implies that

$$\frac{s(x) - s(a)}{s(b) - s(a)} \leq \mathbb{P}\left(\sup_{t < e} X_t^x \geq b\right) \quad \forall a. \quad (4.25)$$

Since $s(l+) = -\infty$, by taking $a \downarrow l$ we find that $\mathbb{P}\left(\sup_{t < e} X_t^x \geq b\right) = 1$. As this is true for all b , by further letting $b \uparrow r$ we obtain

$$\mathbb{P}\left(\sup_{t < e} X_t^x = r\right) = 1.$$

In a similar way,

$$s(r-) = \infty \implies \mathbb{P}\left(\inf_{t < e} X_t^x = l\right) = 1.$$

These two properties imply that with probability one, there are two subsequences of time, along one X_t^x approaches r while along the other X_t^x approaches l . This rules out the possibility of $e_x < \infty$ since on this event we know that X_t^x is convergent a.s. when $t \uparrow e_x$. The conclusion of Part (i) thus follows.

(ii) We have seen that

$$Y_t^{a,b} \triangleq s(X_{\tau_{a,b} \wedge t}^x) - s(l+)$$

is a (non-negative) martingale as a consequence of $\mathcal{A}s = 0$. In particular,

$$\mathbb{E}[s(X_{\tau_{a,b} \wedge t}^x) - s(l+) | \mathcal{F}_s] = s(X_{\tau_{a,b} \wedge s}^x) - s(l+)$$

for $s < t$. Since $\tau_{a,b} \uparrow e_x$ as $a \downarrow l$ and $b \uparrow r$, Fatou's lemma implies that

$$\mathbb{E}[s(X_{e_x \wedge t}^x) - s(l+) | \mathcal{F}_s] \leq \varliminf_{a \downarrow l, b \uparrow r} (s(X_{\tau_{a,b} \wedge s}^x) - s(l+)) = s(X_{e_x \wedge s}^x) - s(l+).$$

As a result, $\{s(X_{e_x \wedge t}^x) - s(l+) : t \geq 0\}$ is a non-negative supermartingale. Note that a non-negative supermartingale is always bounded in L^1 . By the martingale convergence theorem (cf. Theorem 1.3)

$$\lim_{t \rightarrow \infty} (s(X_{e_x \wedge t}^x) - s(l+)) \text{ exists a.s.}$$

Since $s(x)$ is strictly increasing, we conclude that $\lim_{t \uparrow e_x} X_t^x$ exists a.s. On the other hand, we have seen in Part (i) that $\mathbb{P}(\inf_{t < e_x} X_t^x = l) = 1$. As a result,

$$\mathbb{P}(\lim_{t \uparrow e_x} X_t^x = l) = 1.$$

This at the same time rules out the possibility of having a subsequence that converges to r , hence also yielding $\sup_{t < e_x} X_t^x < r$ a.s. The conclusion of Part (ii) thus follows.

(iii) By first letting $a \downarrow l$ and then $b \uparrow r$ in (4.25), we have

$$\mathbb{P}(\sup_{t < e_x} X_t^x = r) \geq \frac{s(x) - s(l+)}{s(r-) - s(l+)}.$$

Similarly,

$$\mathbb{P}(\inf_{t < e_x} X_t^x = l) \geq \frac{s(r-) - s(x)}{s(r-) - s(l+)}.$$

On the other hand, in Part (ii) we have shown that (under the assumption $s(l-) > -\infty$) $\lim_{t \uparrow e_x} X_t^x$ exists a.s. On this event, we have

$$\sup_{t < e} X_t^x = r \implies \lim_{t \uparrow e_x} X_t^x = r, \quad \inf_{t < e} X_t^x = l \implies \lim_{t \uparrow e_x} X_t^x = l.$$

Therefore,

$$\mathbb{P}(\lim_{t \uparrow e_x} X_t^x = r) \geq \frac{s(x) - s(l+)}{s(r-) - s(l+)}, \quad \mathbb{P}(\lim_{t \uparrow e_x} X_t^x = l) \geq \frac{s(r-) - s(x)}{s(r-) - s(l+)}. \quad (4.26)$$

Since

$$\frac{s(x) - s(l+)}{s(r-) - s(l+)} + \frac{s(r-) - s(x)}{s(r-) - s(l+)} = 1,$$

the two inequalities in (4.26) must both be equalities. The conclusion of Part (iii) thus follows. \square

Remark 4.10. Part (i) of Theorem 4.4 gives a simple non-explosion (i.e. $e = \infty$ a.s.) criterion. Although Part (ii) and (iii) describe the convergence properties at the boundary points l, r as $t \uparrow e_x$, it is not clear if $e_x < \infty$ with positive probability or not in these cases. More precise explosion tests were due to W. Feller (cf. [9, Sec. VI.3]).

4.4.2 Bessel processes

An important class of one-dimensional diffusions are Bessel processes. These are defined by taking the Euclidean norm of multidimensional Brownian motions. Let $B = \{(B_t^1, \dots, B_t^n) : t \geq 0\}$ be an n -dimensional Brownian motion (starting at some $\xi \in \mathbb{R}^n$). Define

$$\rho_t \triangleq (B_t^1)^2 + \dots + (B_t^n)^2. \quad (4.27)$$

Definition 4.6. The process $\{\rho_t : t \geq 0\}$ is called a *squared Bessel process* and $\{R_t \triangleq \sqrt{\rho_t} : t \geq 0\}$ is called a *Bessel process* (in dimension n).

There is a useful alternative characterisation of ρ_t as a one-dimensional diffusion. According to Itô's formula,

$$d\rho_t = 2 \sum_{i=1}^n B_t^i dB_t^i + ndt = 2\sqrt{\rho_t} \cdot \frac{\sum_{i=1}^n B_t^i dB_t^i}{\sqrt{\rho_t}} + ndt.$$

Let us introduce the process

$$W_t \triangleq \sum_{i=1}^n \int_0^t \frac{B_s^i}{\sqrt{\rho_s}} dB_s^i.$$

Lévy's characterisation theorem suggests that W is a one-dimensional Brownian motion, as seen from

$$dW_t \cdot dW_t = \left(\sum_{i=1}^n \frac{B_t^i dB_t^i}{\sqrt{\rho_t}} \right) \cdot \left(\sum_{j=1}^n \frac{B_t^j dB_t^j}{\sqrt{\rho_t}} \right) = \frac{\sum_{i=1}^n (B_t^i)^2}{\rho_t} dt = dt.$$

Therefore, ρ_t satisfies the SDE

$$d\rho_t = 2\sqrt{\rho_t} dW_t + ndt.$$

This is a one-dimensional diffusion on the interval $I = (0, \infty)$ which fits into the setting of the current section. With initial condition $\rho_0 = x \in I$, there is a uniquely well-defined solution up to the exit time e_x of I . We write ρ_t^x to keep track of the initial condition. To have a finer understanding about e_x , we define σ_x to be the actual explosion time to ∞ and let τ_0 be the first time of reaching 0. In particular, $e_x \triangleq \tau_0 \wedge \sigma_x$. Since $|\sqrt{x}| \leq \frac{1+|x|}{2}$, we know from Theorem 4.2 that $\sigma_x = \infty$ a.s. and thus $e_x = \tau_0$ a.s. Of course this point is trivial from the definition (4.27) in terms of the Brownian motion. In the one-dimensional case ($n = 1$), since the Brownian motion will a.s. visit the origin in finite time, we know that $\tau_0 < \infty$ a.s. The following result describes the situation when $n \geq 2$.

Theorem 4.5. For $n \geq 2$, we have $\mathbb{P}(\tau_0 = \infty) = 1$. As a result, with probability one an n -dimensional Brownian motion never return to the origin. In addition, for $n \geq 3$ we have

$$\lim_{t \rightarrow \infty} \rho_t^x = \infty \quad a.s. \quad (4.28)$$

Proof. Since $\sigma(x) = 2\sqrt{x}$ and $b(x) = n$, the scale function is found to be

$$s(x) = \int_1^x \exp\left(-\int_1^y \frac{2\alpha dz}{4z} dy\right) = \int_1^x y^{-n/2} dy.$$

In particular, $s(0+) = -\infty$ when $n \geq 2$. In addition,

$$s(\infty) \begin{cases} = \infty, & \text{if } n = 2, \\ < \infty, & \text{if } n \geq 3. \end{cases}$$

In the first case, Part (i) of Theorem 4.4 implies that $\tau_0 = e_x = \infty$ a.s. In the second case, Part (ii) of Theorem 4.4 implies that with probability one,

$$\inf_{t < e_x = \tau_0} \rho_t^x > 0, \quad \lim_{t \uparrow e_x = \tau_0} \rho_t^x = \infty.$$

As a result, τ_0 cannot be finite and at the same time (4.28) follows. \square

Remark 4.11. Since ρ_t^x is the sum of independent squared Gaussian random variables, it is straight forward to compute the Laplace transform of ρ_t^x as

$$\mathbb{E}[e^{-\lambda \rho_t^x}] = \frac{1}{(1 + 2\lambda t)^{n/2}} e^{-\frac{\lambda x}{(1+2\lambda t)}}, \quad \lambda \geq 0.$$

The reason for calling $R_t \triangleq \sqrt{\rho_t}$ a Bessel process is that its probability density function can be found explicitly by inverting the above Laplace transform of ρ_t^x and the resulting formula is given in terms of the classical Bessel functions. This can be seen from the viewpoint of Bessel-type ODEs and their Laplace transforms (cf. [4, Sec. A.5.2]).

Example 4.7. Consider the one-dimensional diffusion

$$dS_t = S_t^2 dB_t, \quad S_0 = x > 0.$$

Define $\rho_t \triangleq S_t^{-2}$. According to Itô's formula,

$$\rho_t = x^{-2} + 2 \int_0^t \sqrt{\rho_s} dW_s + 3t,$$

where $W_t \triangleq -B_t$. In particular, ρ_t is a three-dimensional squared Bessel process starting at x^{-2} . From Theorem 4.5, we know that ρ_t (respectively, S_t) neither explodes to infinity (respectively, hits zero) nor hits zero (respectively, explodes to infinity in finite time). We can write S_t as

$$S_t = \rho_t^{-1/2} = \frac{1}{\sqrt{X_t^2 + Y_t^2 + Z_t^2}} = x + \int_0^t S_u^2 dB_u,$$

where $\{(X_t, Y_t, Z_t)\}$ is a three-dimensional Brownian motion. From this expression, it is clear that $\mathbb{E}[S_t]$ cannot be constant in t (why?). Therefore, S_t is an Itô integral which fails to be a martingale. Another way of looking at the fact that S_t is an Itô integral is to observe that the function $f(x, y, z) = \frac{1}{\sqrt{x^2 + y^2 + z^2}}$ is harmonic on \mathbb{R}^3 (i.e. $\Delta f = 0$). In general, if f is a harmonic function on \mathbb{R}^n and B is an n -dimensional Brownian motion, $f(B_t)$ is always an Itô integral which is easily seen from Itô's formula.

4.5 Connection with partial differential equations

In this section, we explore the relationship between Itô diffusions and PDEs. As we will see, solutions to a class of elliptic and parabolic PDEs admit stochastic representations. As an example, we provide a mathematical explanation for the heat transfer problem discussed in Section 1.1.2.

Consider the following n -dimensional diffusion process:

$$\begin{cases} dX_t^x = \sigma(X_t^x)dB_t + b(X_t^x)dt, & t \geq 0, \\ X_0^x = x, \end{cases} \quad (4.29)$$

where B is a d -dimensional Brownian motion and the coefficient functions σ, b satisfy the Lipschitz condition (4.4). This implies existence and uniqueness by Theorem 4.1 (Lipschitz implies linear growth when σ, b does not depend on time). Recall from Proposition 4.1 that the generator of X is the second order differential operator

$$\mathcal{A}f(x) = \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum_{i=1}^n b^i(x) \frac{\partial f(x)}{\partial x_i},$$

where $a(x) \triangleq \sigma(x) \cdot \sigma^T(x)$.

4.5.1 Dirichlet boundary value problems

In PDE theory, one often considers elliptic boundary value problems associated with the operator \mathcal{A} . Let D be a bounded domain in \mathbb{R}^n . Let $g : \bar{D} \rightarrow \mathbb{R}^1$ and $f : \partial D \rightarrow \mathbb{R}^1$ be continuous functions, where \bar{D} denotes the closure of D and ∂D denotes its boundary. The so-called (*Dirichlet*) *boundary value problem* is to find a function $u \in C(\bar{D}) \cap C^2(D)$ (continuous on \bar{D} and twice continuously differentiable on D) that satisfies the following PDE:

$$\begin{cases} \mathcal{A}u = -g, & x \in D, \\ u = f, & x \in \partial D. \end{cases} \quad (4.30)$$

The existence of a solution u is well studied in PDE theory under suitable conditions on the coefficients as well as the boundary ∂D . We are interested in representing the solution u in terms of the diffusion process $\{X_t^x\}$, which also implies the uniqueness of (4.30).

Theorem 4.6. *Suppose that there exists $u \in C(\bar{D}) \cap C^2(D)$ which solves the boundary value problem (4.30). Let $\{X_t^x\}$ be the solution to the SDE (4.29). Suppose further that the exit time*

$$\tau_x \triangleq \inf\{t \geq 0 : X_t^x \notin D\}$$

is integrable for every given $x \in D$. Then the PDE solution u is given by

$$u(x) = \mathbb{E}\left[f(X_{\tau_x}^x) + \int_0^{\tau_x} g(X_s^x) ds\right]. \quad (4.31)$$

In particular, the solution to the boundary value problem (4.30) is unique in $C(\bar{D}) \cap C^2(D)$.

Proof. Fix $x \in D$. According to Itô's formula and the equation for u , we have

$$\begin{aligned} u(X_{\tau_x \wedge t}^x) &= u(x) + \sum_{i=1}^n \sum_{k=1}^d \int_0^{\tau_x \wedge t} \frac{\partial u}{\partial x_i}(X_s^x) \sigma_k^i(X_s^x) dB_s^k + \int_0^{\tau_x \wedge t} (\mathcal{A}u)(X_s^x) ds \\ &= u(x) + \sum_{i=1}^n \sum_{k=1}^d \int_0^{\tau_x \wedge t} \frac{\partial u}{\partial x_i}(X_s^x) \sigma_k^i(X_s^x) dB_s^k - \int_0^{\tau_x \wedge t} g(X_s^x) ds \end{aligned}$$

The integrability condition $\mathbb{E}[\tau_x] < \infty$ ensures that the stochastic integral term is indeed a martingale (we do not elaborate this technical point here). The result follows by taking expectation on both sides and sending $t \rightarrow \infty$. \square

It is useful to know when the exit time τ_x is integrable. Here is a simple sufficient condition.

Proposition 4.4. *Suppose that*

$$\inf_{x \in \overline{D}} a_{ii}(x) > 0$$

for some $i = 1, \dots, n$. Then $\mathbb{E}[\tau_x] < \infty$ for every $x \in D$.

Proof. Define

$$p \triangleq \inf_{x \in \overline{D}} a_{ii}(x), \quad q \triangleq \sup_{x \in \overline{D}} |b(x)|, \quad r \triangleq \inf_{x \in \overline{D}} x_i.$$

Let $\lambda > 2q/p$ be fixed and consider the function

$$h(x) \triangleq -e^{\lambda x_i}, \quad x \in \mathbb{R}^n.$$

Then one finds

$$-(\mathcal{A}h)(x) = e^{\lambda x_i} \cdot \left(\frac{1}{2} \lambda^2 a_{ii}(x) + \lambda b_i(x) \right) \geq \frac{1}{2} \lambda e^{\lambda r} (\lambda p - 2q) =: \gamma > 0$$

for every $x \in D$. On the other hand, given each $x \in D$, the process

$$t \mapsto h(X_{\tau_x \wedge t}^x) - h(x) - \int_0^{\tau_x \wedge t} (\mathcal{A}h)(X_s^x) ds$$

is a martingale (cf. Proposition 4.1). It follows that

$$\mathbb{E}[h(X_{\tau_x \wedge t}^x)] = h(x) + \mathbb{E} \left[\int_0^{\tau_x \wedge t} (\mathcal{A}h)(X_s^x) ds \right] \leq h(x) - \gamma \mathbb{E}[\tau_x \wedge t].$$

Therefore,

$$\mathbb{E}[\tau_x \wedge t] \leq \frac{h(x) - \mathbb{E}[h(X_{\tau_x \wedge t}^x)]}{\gamma} \leq \frac{2 \sup_{x \in \overline{D}} |h(x)|}{\gamma}.$$

Since this is true for all t , the result follows by taking $t \rightarrow \infty$. \square

Example 4.8. Suppose that $n = d$, $\sigma = \text{Id}$, $b = 0$. Then $\{X_t^x\}$ is a Brownian motion starting at x . In this case, $\mathcal{A} = \frac{1}{2} \Delta$. When $g = 0$, the solution u of the PDE (4.30) is represented as $u(x) = \mathbb{E}[f(X_{\tau_x}^x)]$. This completes the discussion of the heat transfer problem (1.4) introduced in Section 1.1.2.

4.5.2 Parabolic problems: the Feynman-Kac formula

Next we consider the parabolic situation. Let $T > 0$ be fixed, and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : [0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}$ be given continuous functions. We consider the following so-called *Cauchy problem*: find a function $u \in C([0, \infty) \times \mathbb{R}^n) \cap C^{1,2}([0, \infty) \times \mathbb{R}^n)$ ($C^{1,2}$ means continuously differentiable in t and twice continuously differentiable in x) that satisfies

$$\begin{cases} \frac{\partial u}{\partial t} = \mathcal{A}u + g, & (t, x) \in (0, \infty) \times \mathbb{R}^n, \\ u(0, x) = f(x), & x \in \mathbb{R}^n. \end{cases} \quad (4.32)$$

Here \mathcal{A} acts on u by differentiating with respect to the spatial variable. From PDE theory, under suitable conditions on the coefficients there exists a unique solution to (4.32). We are again interested in its stochastic representation. Suppose that the functions f, g satisfy the following polynomial growth condition:

$$|f(x)| \vee |g(t, x)| \leq C(1 + |x|^\mu) \quad \forall t, x \quad (4.33)$$

with some constants $C, \mu > 0$. Then we have the following renowned *Feynman-Kac formula*.

Theorem 4.7. *Let $u \in C([0, \infty) \times \mathbb{R}^n) \cap C^{1,2}([0, \infty) \times \mathbb{R}^n)$ be a solution to the Cauchy problem (4.32) which satisfies the polynomial growth condition:*

$$|u(t, x)| \leq K(1 + |x|^\lambda) \quad \forall t, x \quad (4.34)$$

with some constants $K, \lambda > 0$. Then u admits the following stochastic representation:

$$u(t, x) = \mathbb{E}\left[f(X_t^x) + \int_0^t g(t-s, X_s^x) ds\right]. \quad (4.35)$$

In particular, the solution to the Cauchy problem (4.32) is unique in the space of functions in $C([0, T] \times \mathbb{R}^n) \cap C^{1,2}([0, T] \times \mathbb{R}^n)$ that satisfy the polynomial growth condition (4.34).

Proof. The proof is essentially the same as in the elliptic case. Let $t > 0$ be fixed.

By applying Itô's formula to the process $s \mapsto u(t-s, X_s^x)$ ($0 \leq s \leq t$), one finds

$$\begin{aligned}
f(X_t^x) &= u(t, x) - \int_0^t \partial_t u(t-s, X_s^x) ds + \sum_{i,k} \int_0^t \partial_{x_i} u(t-s, X_s^x) \sigma_k^i(X_s^x) dB_s^k \\
&\quad + \int_0^t (\mathcal{A}u)(t-s, X_s^x) ds \\
&= u(t, x) + \sum_{i,k} \int_0^t \partial_{x_i} u(t-s, X_s^x) \sigma_k^i(X_s^x) dB_s^k - \int_0^t g(t-s, X_s^x) ds.
\end{aligned} \tag{4.36}$$

The result follows from taking expectation on both sides and rearrangement of terms. The polynomial growth condition for the functions f, g, u can be used to show that the stochastic integral in (4.36) is indeed a martingale. We omit the discussion on this technical point. \square

An immediate consequence of Theorem 4.6 (the elliptic problem) is that

$$f, g \geq 0 \implies u \geq 0.$$

Similar property holds for the Cauchy problem.

Remark 4.12. Although there are neat stochastic representations for PDE solutions, it is in general not efficient to prove existence of solutions in this way. The reason is that checking regularity properties for the function defined by the representation formula often involves a non-trivial amount of technicalities. A main benefit from these representation formulae is that one can investigate quantitative properties of the solution from the probabilistic viewpoint. On the applied side, it also enables one to simulate PDE solutions by using diffusion trajectories (the Monte-Carlo method) and study their numerical approximations.

To conclude this chapter, we briefly answer the two fundamental questions raised in the introduction of this chapter. Let $X = \{X_t^x\}$ be the solution to the SDE (4.29) where σ is such that $a = \sigma \cdot \sigma^T$.

Answer to Question 1. From Section 4.2, one knows that X is a Markov process whose generator is \mathcal{A} .

Answer to Question 2. Suppose that the transition density function

$$p(t, x, y) = \frac{\mathbb{P}(X_t^x \in dy)}{dy}$$

exists and is sufficiently regular. Since $X_0^x = x$, it is clear that $p(0, x, y) = \delta_x(y)$. According to the Feynman-Kac formula (cf. Theorem 4.7), for any initial condition $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the function

$$u(t, x) = \mathbb{E}[f(X_t^x)] = \int_{\mathbb{R}^n} p(t, x, y) f(y) dy$$

solves the Cauchy problem

$$\frac{\partial u}{\partial t} = \mathcal{A}u, \quad u(0, \cdot) = f.$$

Equivalently, one has

$$\int_{\mathbb{R}^n} \frac{\partial p}{\partial t} p(t, x, y) f(y) dy = \int_{\mathbb{R}^n} \mathcal{A}_x p(t, x, y) f(y) dy.$$

Since this is true for arbitrary f , it must hold true that

$$\frac{\partial p}{\partial t}(t, x, y) = \mathcal{A}_x p(t, x, y). \quad (4.37)$$

The above equation is known as *Kolmogorov's backward equation*. The term “backward” reflects the fact that the operator \mathcal{A}_x is acting on the backward variable x (the initial position). One needs to use a duality argument to justify the equation (4.1). Indeed, for fixed x let

$$\varphi(t, y) \triangleq p(t, x, y), \quad (t, x) \in [0, \infty) \times \mathbb{R}^n.$$

The Markov property implies that

$$\varphi(t + s, y) = \int_{\mathbb{R}^n} \varphi(t, z) p(s, z, y) dz.$$

According to the backward equation (4.37),

$$\begin{aligned} \frac{\partial \varphi}{\partial s}(t + s, y) &= \int_{\mathbb{R}^n} \varphi(t, z) \frac{\partial p}{\partial s}(s, z, y) dz = \int_{\mathbb{R}^n} \varphi(t, z) \mathcal{A}_z p(s, z, y) dz \\ &= \int_{\mathbb{R}^n} \mathcal{A}_z^* \varphi(t, z) p(s, z, y) dz \quad (\text{integration by parts}), \end{aligned}$$

where

$$\mathcal{A}^* : \varphi(\cdot) \mapsto \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x^i \partial x^j} (a^{ij}(\cdot) \varphi(\cdot)) - \sum_{i=1}^n \frac{\partial}{\partial x^i} (b^i(\cdot) \varphi(\cdot))$$

is the formal adjoint of operator \mathcal{A} . Since $p(0, z, y) = \delta_z(y)$, one obtains that

$$\frac{\partial p}{\partial t}(t, x, y) = \frac{\partial \varphi}{\partial s} \Big|_{s=0}(t + s, y) = \mathcal{A}_y^* \varphi(t, y) = \mathcal{A}_y^* p(t, x, y).$$

This equation is known as *Kolmogorov's forward equation* as the operator \mathcal{A}_y^* acts on the forward variable y (the future position). In the case when $p(t, x, y)$ does not exist, the probability measure $P(t, x, dy) \triangleq \mathbb{P}(X_t^x \in dy)$ is still the fundamental solution to the equation (4.1) in the distributional sense.

Remark 4.13. The existence and smoothness of probability density functions is a rich subject of study that was largely developed by P. Malliavin in the 1970s (the Malliavin calculus). It was shown by Malliavin that the transition density $p(t, x, y)$ exists and is smooth in the case when the generator \mathcal{A} is a *hypoelliptic operator*. This theorem provides a probabilistic approach to a renowned PDE result of L. Hörmander in the 1970s. We refer the reader to [19] for an elegant introduction to this subject.

5 Applications in risk-neutral pricing

In this chapter, we apply the methods of stochastic calculus to an important problem in mathematical finance: the pricing of derivative securities. To illustrate the essential ideas, we work with simplified assumptions/settings and do not pursue full generality. We refer the reader to [18, Chap. 5] for a thorough discussion, which is also an excellent source of learning mathematical finance.

5.1 Basic concepts

We first introduce the set-up and motivate the basic questions. We consider a financial market in which there are $n + 1$ fundamental assets: n stocks (risky assets) and 1 money market account (risk-free asset).

5.1.1 Stock price model and interest rate process

The price process for the n stocks are assumed to satisfy the following system of SDEs:

$$dS_t^i = \alpha^i(t)S_t^i dt + S_t^i \cdot \sum_{j=1}^d \sigma_j^i(t) dB_t^j, \quad i = 1, \dots, n. \quad (5.1)$$

Here $B = \{(B_t^1, \dots, B_t^d) : 0 \leq t \leq T\}$ is a d -dimensional Brownian motion defined on a given fixed filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t : 0 \leq t \leq T\})$. The coefficients $\{\alpha^i(t), \sigma_j^i(t)\}$ are given progressively measurable processes. The intuition behind this equation is that the relative change of stock price $\frac{dS_t^i}{S_t^i}$ is governed by two factors: a *mean rate of return* factor $\alpha^i(t)dt$ and a random force $\sum_j \sigma_j^i(t)dB_t^j$. Throughout the discussion, we assume that all the relevant integrands satisfy the stronger integrability condition (3.20) so that the stochastic integrals are martingales.

We make one more essential assumption: *the underlying filtration is generated by the Brownian motion B* . Heuristically, the role of B accounts for the intrinsic randomness of the market arising from the interactions among millions of individual actions. As a result, this filtration assumption means that there are no additional information/randomness other than the one intrinsically carried by the market itself.

For the money market account, we assume that the interest rate process is described by a given progressively measurable process $R = \{R_t : 0 \leq t \leq T\}$ and

we define the following *discount process*

$$D_t \triangleq e^{-\int_0^t R_s ds}, \quad 0 \leq t \leq T.$$

Mathematically, \$1 at time t is equivalent to $\$D_t$ at time 0. Note that D_t satisfies the ODE

$$dD_t = -R_t D_t dt, \quad D_0 = 1.$$

By using integration by parts, one easily finds that the *discounted stock price process* satisfies

$$\begin{aligned} d(D_t S_t^i) &= D_t dS_t^i + S_t^i dD_t + dD_t \cdot dS_t^i \\ &= D_t S_t^i ((\alpha^i(t) - R_t) dt + \sum_{j=1}^d \sigma_j^i(t) dB_t^j). \end{aligned} \quad (5.2)$$

5.1.2 Portfolio process and value of portfolio

Consider an agent with initial capital X_0 and she is managing a portfolio that consists of the n stocks and the money market account. Suppose that at time t the agent holds Δ_t^i shares of stock i ($i = 1, \dots, n$) and she borrows or invests money with interest rate R_t to finance this position. Let X_t denote the value of this portfolio at time t . The infinitesimal change of X_t must therefore satisfy

$$\begin{aligned} dX_t &= \sum_{i=1}^n \Delta_t^i dS_t^i + (X_t - \sum_{i=1}^n \Delta_t^i S_t^i) R_t dt \\ &= R_t X_t dt + \sum_{i=1}^n \Delta_t^i S_t^i ((\alpha^i(t) - R_t) dt + \sum_{j=1}^d \sigma_j^i(t) dB_t^j) \\ &= R_t X_t dt + \sum_{i=1}^n \frac{\Delta_t^i}{D_t} d(D_t S_t^i), \end{aligned} \quad (5.3)$$

where the last identity follows from the equation (5.2). As a result, the *discounted portfolio value* satisfies

$$d(D_t X_t) = D_t dX_t - R_t D_t X_t dt = \sum_{i=1}^n \Delta_t^i d(D_t S_t^i). \quad (5.4)$$

Definition 5.1. The process $\Delta = \{\Delta_t : 0 \leq t \leq T\}$ is called the *portfolio process*. Respectively, the process $X = \{X_t : 0 \leq t \leq T\}$ is called the *portfolio value process*.

Remark 5.1. The initial capital X_0 and portfolio process Δ can be decided at the agent's wish. However, the portfolio value process X is uniquely determined through the SDE (5.3). How to select a portfolio dynamically (i.e. deciding Δ) is an important question of study.

5.1.3 Risk-neutral measure and hedging

The goal of this chapter is to understand the pricing of derivative security. We must first give its definition.

Definition 5.2. A *derivative security with maturity* $T > 0$ is an asset whose payoff at time T is given by an \mathcal{F}_T -measurable random variable V_T . If an agent has a *long position* on a derivative security, the agent owns the security so that she will receive a payoff of amount V_T at the time T of maturity. Respectively, having a *short position* on the security means that the agent is obliged to pay someone an amount of V_T at maturity time T .

Example 5.1. A *European call option* on Stock X with strike price K and maturity T is a derivative security with payoff $V_T = \max\{S_T - K, 0\}$ at time T , where S_T is the price of the stock at maturity. This option gives its owner the right to purchase the stock with the fixed price K at time T , hence avoiding the risk of price increase above K . Indeed, if $S_T > K$, the buyer profits $S_T - K$ by exercising the option. If $S_T \leq K$, the buyer does not exercise the option as she can get the stock with a cheaper market price in this case. This explains the definition of the payoff V_T for a European call option. Similarly, a *European put option* on Stock X with strike price K and maturity T has payoff $V_T \triangleq \max\{K - S_T, 0\}$ at time T . It gives its owner the right to sell the stock at the fixed price K at time T .

Here comes a fundamental question in pricing theory.

Question: Suppose that a derivative security has payoff V_T at time T (maturity). What should its price be at the initial time and more generally at any given time $t < T$? E.g. how much do you need to pay today in order to have the right to buy a share of Google at \$2300 in a month from now (European call option)?

To answer this question, we need to introduce two essential concepts: risk-neutral measure and hedging strategy.

Definition 5.3. A probability measure $\tilde{\mathbb{P}}$ on \mathcal{F}_T^B is called a *risk-neutral measure* if it satisfies the following two properties:

- (i) $\tilde{\mathbb{P}}$ is equivalent to \mathbb{P} , i.e. $\mathbb{P}(A) = 0 \iff \tilde{\mathbb{P}}(A) = 0$ for any $A \in \mathcal{F}_T^B$;
- (ii) under $\tilde{\mathbb{P}}$, the discounted stock price process $\{D_t S_t^i\}$ is a martingale for every $i = 1, 2, \dots, n$.

Heuristically, the existence of a risk-neutral measure $\tilde{\mathbb{P}}$ suggests that the market is fair and there is no opportunity of earning free money by trading stocks (no arbitrage). This point can be made mathematically precise.

Definition 5.4. An *arbitrage* is a portfolio such that

$$X_0 = 0, \mathbb{P}(X_T \geq 0) = 1, \mathbb{P}(X_T > 0) > 0.$$

Theorem 5.1 (The first fundamental theorem of asset pricing). *Suppose that a risk-neutral measure $\tilde{\mathbb{P}}$ exists. Then there is no arbitrage opportunities.*

Proof. Let $X = \{X_t : 0 \leq t \leq T\}$ be an arbitrary portfolio value process (associated with a given portfolio Δ) such that $X_0 = 0$ and $X_T \geq 0$ a.s. Since $\{D_t S_t^i\}$ is a martingale under $\tilde{\mathbb{P}}$ for each i , so is the discounted value process $\{D_t X_t\}$ as a consequence of the equation (5.4). In particular

$$\tilde{\mathbb{E}}[D_T X_T] = X_0 = 0.$$

Since $D_T > 0$ and $X_T \geq 0$ a.s., one concludes that $\tilde{X}_T = 0$ a.s. under $\tilde{\mathbb{P}}$. By the equivalence between $\tilde{\mathbb{P}}$ and \mathbb{P} , one also has $\tilde{X}_T = 0$ a.s. under \mathbb{P} . Therefore, X cannot be an arbitrage. \square

Now consider an agent who holds a short position of the derivative security, i.e. she sells the security at the initial time and is subject to paying V_T to the buyer at time T . The agent may hedge her position by trading stocks continuously in time.

Definition 5.5. A *hedge* of a short position on the derivative security V_T is a choice of the initial capital X_0 and a suitable portfolio process Δ such that the terminal value of the portfolio is exactly V_T (i.e. $X_T = V_T$).

The pricing mechanism for a derivative security relies on the existence of risk-neutral measure and the availability of hedging strategy. Suppose that a hedge of the agent's short position on V_T exists and let us denote the associated portfolio value process as $V = \{V_t : 0 \leq t \leq T\}$. If a risk-neutral measure $\tilde{\mathbb{P}}$ exists, as in the proof of Theorem 5.1 the discounted value process $\{D_t V_t\}$ is a martingale under $\tilde{\mathbb{P}}$. Since V_T is the payoff of the security at time T , the martingale property

suggests that $D_t V_t$ is “equivalent” to $D_T V_T$ in a probabilistic sense and should thus be considered as the discounted time- t price of the derivative security. Since

$$D_t V_t = \tilde{\mathbb{E}}[D_T V_T | \mathcal{F}_t],$$

one finds that

$$V_t = \frac{1}{D_t} \tilde{\mathbb{E}}[D_T V_T | \mathcal{F}_t] = \tilde{\mathbb{E}}[e^{-\int_t^T R_s ds} V_T | \mathcal{F}_t]. \quad (5.5)$$

In particular,

$$V_0 = \tilde{\mathbb{E}}[e^{-\int_0^T R_s ds} V_T] \quad (5.6)$$

as \mathcal{F}_0^B is trivial.

It is not hard to see why V_0 given by (5.6) has to be the true price of the security at time 0. Suppose that the market price V'_0 does not coincide with V_0 , say $V'_0 > V_0$. A person who owns the security will then sell it to the market at time 0, and she is obliged to pay V_T to the buyer at time T . In the meanwhile, she can use an initial capital V_0 and hedging strategy Δ to create a portfolio. The terminal value of this portfolio is precisely V_T (by the definition of hedge), which covers her payment to the buyer. In this way, the person gains $V'_0 - V_0 > 0$ without any cost. The case when $V'_0 < V_0$ also leads to a free lunch. The liquidity and effectiveness of the market will eliminate all such arbitrage opportunities. As a result, the market price of the security will be adjusted to its theoretical value V_0 .

Equation (5.5) is the *pricing formula* for the derivative security with payoff V_T at maturity T . Our derivation of this formula is based on two assumptions:

- (i) A risk-neutral measure $\tilde{\mathbb{P}}$ exists;
- (ii) the derivative security can be hedged.

Remark 5.2. The pricing formula shows that if a hedge Δ exists its associated value process must be given by (5.5). At first glance, the existence of a hedge Δ seems to be irrelevant and does not come into the formula (5.5). However, this assumption is used when obtaining the martingale property of $\{D_t V_t\}$, as seen from equation (5.4) which involves Δ explicitly.

To yield a satisfactory theory, the following two questions need to be answered:

Question 1: When does a risk-neutral measure exist?

Question 2: Can a derivative security always be hedged?

As we will see, the first question can be studied by using Girsanov’s transformation theorem. For the second question, it turns out that the existence of hedge

is equivalent to the uniqueness of the risk-neutral measure. This is the content of the *second fundamental theorem of asset pricing*, which will be discussed in Section 5.3 below. Such a theorem is reasonable as the non-uniqueness of risk-neutral measures would make the pricing formula (5.5) ill-defined.

5.2 The Black-Scholes-Merton formula

To understand the two questions posted at the end of the last section, we first consider the case when $n = 1$ (there is only one stock in the market). More specifically, the stock price process is described by

$$dS_t = \alpha_t S_t dt + \sigma_t S_t dB_t \quad (5.7)$$

where $\{B_t : 0 \leq t \leq T\}$ is now a one-dimensional Brownian motion. We assume that the coefficient $\sigma_t \neq 0$ for every t so that S_t is “truly diffusive”. In this case, the aforementioned two questions can be answered affirmatively:

- (i) a risk-neutral measure always exists and can be found explicitly as a consequence of Girsanov’s theorem;
- (ii) a hedge of the short position on any derivative security V_T can be constructed from the martingale representation theorem.

In addition, in the situation when the coefficients α_t, σ_t are deterministic constants, one can explicitly write down the pricing formula for European options on the stock. This is the content of the renowned *Black-Scholes-Merton formula*.

5.2.1 Construction of the risk-neutral measure

We want to find an equivalent probability measure $\tilde{\mathbb{P}}$ under which the discounted stock price process $\{D_t S_t\}$ is a martingale. To this end, in view of (5.2) one can write

$$d(D_t S_t) = \sigma_t (D_t S_t) \cdot (\theta_t dt + dB_t) \quad (5.8)$$

where the process $\theta_t \triangleq \frac{\alpha_t - R_t}{\sigma_t}$ is called the *market price of risk*. The idea of constructing $\tilde{\mathbb{P}}$ is very simple: we want the process

$$\tilde{B}_t \triangleq B_t + \int_0^t \theta_s ds$$

to be a Brownian motion under $\tilde{\mathbb{P}}$ so that $\{D_t S_t\}$ is a martingale as a stochastic integral against $\{\tilde{B}_t\}$. This is exactly the content of Girsanov’s theorem (3.10).

Namely, let us define the exponential martingale

$$Z_t \triangleq \exp \left(- \int_0^t \theta_s dB_s - \frac{1}{2} \int_0^t \theta_s^2 ds \right),$$

and introduce a new probability measure

$$\tilde{\mathbb{P}}(A) \triangleq \mathbb{E}[\mathbf{1}_A Z_T], \quad A \in \mathcal{F}_T^B. \quad (5.9)$$

According to Girsanov's theorem, $\{\tilde{B}_t\}$ is a Brownian motion under $\tilde{\mathbb{P}}$. As a result, $\{D_t S_t\}$ is a $\tilde{\mathbb{P}}$ -martingale. This proves the existence of a risk-neutral measure in the current context.

Simple algebra shows that under $\tilde{\mathbb{P}}$,

$$dS_t = R_t S_t dt + \sigma_t S_t d\tilde{B}_t. \quad (5.10)$$

In other words, the mean rate of return for the stock is changed to the interest rate R_t under the new measure $\tilde{\mathbb{P}}$. As a result, under the risk-neutral measure trading stocks is essentially equivalent to trading in the money bank account. This is also consistent with the first fundamental theorem of asset pricing which asserts that there is no arbitrary opportunities.

5.2.2 Completeness of market model and construction of hedging strategies

To study hedging strategies, we first give the following definition.

Definition 5.6. The market model (5.1) is said to be *complete* if any derivative security with an \mathcal{F}_T^B -measurable payoff V_T at maturity T can be hedged.

Consider a derivative security with payoff V_T at maturity. The following result is crucial for constructing a hedge of V_T . Recall that $\tilde{\mathbb{P}}$ is the risk-neutral measure defined by (5.9).

Lemma 5.1. Let $\tilde{M} = \{\tilde{M}_t : 0 \leq t \leq T\}$ be a martingale under $\tilde{\mathbb{P}}$. Then there exists a progressively measurable process $\Gamma = \{\Gamma_t : 0 \leq t \leq T\}$ such that

$$\tilde{M}_t = \tilde{M}_0 + \int_0^t \Gamma_s d\tilde{B}_s.$$

Proof. This is not an immediate application of the martingale representation theorem, as the filtration is generated by B rather than \tilde{B} even though \tilde{B} is a $\tilde{\mathbb{P}}$ -Brownian motion. But one can deal with this issue easily. Indeed, from Proposition 3.8 we know that

$$M_t \triangleq \tilde{M}_t - \int_0^t \theta_s d\langle \tilde{M}, \tilde{B} \rangle_s$$

is a martingale under \mathbb{P} . Here one regards $\tilde{\mathbb{P}}$ as the old measure, \tilde{B} as the old Brownian motion and the original measure

$$d\mathbb{P} = \exp\left(\int_0^T \theta_t d\tilde{B}_t - \frac{1}{2} \int_0^T \theta_t^2 dt\right) d\tilde{\mathbb{P}}$$

as the new measure. Note that the transformed Brownian motion under \mathbb{P} is precisely the original B . According to the martingale representation theorem (cf. Theorem 3.6) under \mathbb{P} , one has

$$M_t = M_0 + \int_0^t \Gamma_s dB_s$$

for some $\Gamma = \{\Gamma_t : 0 \leq t \leq T\}$. It follows that

$$d\langle M, B \rangle_t = dM_t \cdot dB_t = \Gamma_t dt.$$

Since

$$\langle \tilde{M}, \tilde{B} \rangle^{\tilde{\mathbb{P}}} = \langle M, B \rangle^{\mathbb{P}},$$

one obtains that

$$\begin{aligned} \tilde{M}_t &= M_t + \int_0^t \theta_s d\langle \tilde{M}, \tilde{B} \rangle_s = M_t + \int_0^t \theta_s d\langle M, B \rangle_s \\ &= M_0 + \int_0^t \Gamma_s dB_s + \int_0^t \Gamma_s \theta_s ds \\ &= \tilde{M}_0 + \int_0^t \Gamma_s d\tilde{B}_s. \end{aligned}$$

□

Now consider a derivative security with payoff V_T at maturity T . We define V_t by the pricing formula (5.5). Then the process $\{D_t V_t\}$ is a martingale under $\tilde{\mathbb{P}}$.

As a consequence of Lemma 5.1, there exists a progressively measurable process $\{\Gamma_t\}$ such that

$$D_t V_t = V_0 + \int_0^t \Gamma_s d\tilde{B}_s. \quad (5.11)$$

On the other hand, equations (5.4) and (5.8) imply that

$$d(D_t V_t) = \Delta_t \sigma_t D_t S_t d\tilde{B}_t. \quad (5.12)$$

By comparing (5.11) and (5.12), we find that

$$\Gamma_t = \Delta_t \sigma_t D_t S_t. \quad (5.13)$$

Therefore, the hedge strategy is given by

$$\Delta_t = \frac{\Gamma_t}{\sigma_t D_t S_t}.$$

This proves the existence of a hedge for the security derivative V_T . By definition, we conclude that the one-dimensional market model (5.7) is complete.

5.2.3 Derivation of the Black-Scholes-Merton formula

We now consider the simplest situation where the coefficients of the stock model (5.7) and the interest rate process are all deterministic constants, say

$$\alpha_t \equiv \alpha \in \mathbb{R}, \quad \sigma_t \equiv \sigma \neq 0, \quad R_t \equiv r > 0.$$

Consider an European call option with strike price K at maturity T (cf. Example 5.1). By definition, its payoff at time T is given by $V_T = (S_T - K)^+$ ($x^+ \triangleq \max\{x, 0\}$). We are going to derive an explicit formula for its price V_t at each $t < T$.

First of all, by solving the equation (5.10) under $\tilde{\mathbb{P}}$ in this case, one finds that (cf. Example 4.6)

$$S_t = \exp\left(\sigma \tilde{B}_t + \left(r - \frac{1}{2}\sigma^2\right)t\right).$$

In particular,

$$S_T = S_t \cdot \exp\left(\sigma \sqrt{\tau} Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right) \quad (5.14)$$

where we have set $\tau \triangleq T - t$ (the time to maturity) and

$$Y \triangleq \frac{\tilde{B}_T - \tilde{B}_t}{\sqrt{T - t}} \sim N(0, 1).$$

According to the pricing formula (5.5),

$$V_t = \tilde{\mathbb{E}}[e^{-r(T-t)}(S_T - K)^+ | \mathcal{F}_t]. \quad (5.15)$$

In view of (5.14), S_T consists of two parts: the current price $S_t \in \mathcal{F}_t$ and the standard normal random variable Y that is independent of \mathcal{F}_t . As a result, the conditional expectation (5.15) can be evaluated explicitly as

$$V_t = c(t, S_t),$$

where the function $c(t, x)$ is defined by the formula

$$\begin{aligned} c(t, x) &\triangleq \tilde{\mathbb{E}}[e^{-r\tau}(x \cdot \exp(\sigma\sqrt{\tau}Y + (r - \frac{1}{2}\sigma^2)\tau) - K)^+] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-r\tau}(x \cdot \exp(-\sigma\sqrt{\tau}y + (r - \frac{1}{2}\sigma^2)\tau) - K)^+ e^{-y^2/2} dy. \end{aligned}$$

Note that $x \cdot \exp(-\sigma\sqrt{\tau}y + (r - \frac{1}{2}\sigma^2)\tau) > K$ if and only if

$$y < d_-(\tau, x) \triangleq \frac{1}{\sigma\sqrt{\tau}} \left(\log \frac{x}{K} + (r - \frac{1}{2}\sigma^2)\tau \right).$$

Therefore,

$$\begin{aligned} c(t, x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_-(\tau, x)} x e^{-y^2/2 - \sigma\sqrt{\tau}y - \sigma^2\tau/2} dy - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_-(\tau, x)} e^{-r\tau} K e^{-y^2/2} dy \\ &= x\Phi(d_+(\tau, x)) - K e^{-r\tau} \Phi(d_-(\tau, x)), \end{aligned}$$

where

$$d_+(\tau, x) \triangleq d_-(\tau, x) + \sigma\sqrt{\tau} = \frac{1}{\sigma\sqrt{\tau}} \left(\log \frac{x}{K} + (r + \frac{1}{2}\sigma^2)\tau \right)$$

and Φ is the cumulative distribution function of $N(0, 1)$. To summarise, we have obtained the following *Black-Scholes-Merton option pricing formula* for European call options.

Theorem 5.2. *The price of a European call option with strike price K and time to maturity $\tau = T - t$ is given by*

$$V_t = S_t \cdot \Phi(d_+(\tau, S_t)) - e^{-r\tau} K \Phi(d_-(\tau, S_t)),$$

where S_t is the price of the stock at the current time t .

Remark 5.3. European put options are priced in a similar way and a parallel explicit formula is also available. We omit the discussion on this case.

5.3 The second fundamental theorem of asset pricing

The generalisation of the previous calculations to the case with n stocks is almost straight forward with two major exceptions:

- (i) a risk-neutral measure needs not always exist;
- (ii) the market model needs not always be complete (hedge may not always be possible).

To see this, we first recall that the discounted stock price process $\{D_t S_t^i : i = 1, \dots, n\}$ satisfies the SDE (5.2). Let us introduce the following linear algebraic system

$$\alpha^i(t) - R_t = \sum_{j=1}^d \sigma_j^i(t) \theta_t^j, \quad i = 1, \dots, n,$$

with unknowns $\theta = \{\theta_t^j : j = 1, \dots, d\}$. In matrix notation,

$$\alpha_t - \mathbf{R}_t = \sigma_t \cdot \theta_t, \quad (5.16)$$

where α_t, θ_t are written as column vectors, $\mathbf{R}_t \triangleq (R_t, \dots, R_t)^T$ and σ_t is the matrix $\sigma_j^i(t)$ being its (i, j) -entry. The system (5.16) is called the *market price of risk equation*. If the system is solvable (not always true though!), one can rewrite the SDE (5.2) as

$$d(D_t S_t^i) = D_t S_t^i \cdot \left(\sum_{j=1}^d \sigma_j^i(t) \cdot (\theta_t^j dt + dB_t^j) \right).$$

Starting from this, one can apply the change of measure associated with the exponential martingale

$$Z_t \triangleq \exp \left(- \sum_{j=1}^d \int_0^t \theta_s^j dB_s^j - \frac{1}{2} \sum_{j=1}^d \int_0^t (\theta_s^j)^2 ds \right).$$

Along the same lines as in the one-dimensional case, the following result can be established easily.

Theorem 5.3. *Suppose that the market price of risk equation (5.16) is solvable. Then there exists a risk-neutral measure \mathbb{P} and thus no arbitrage opportunities are available. In addition, $\{D_t X_t\}$ is a \mathbb{P} -martingale for any portfolio value process $\{X_t\}$.*

In general, it can be proved that if the market price of risk equation fails to be solvable, there will be arbitrage opportunities in the market. As a consequence, the following three statements are equivalent:

- (i) the market price of risk equation is solvable;
- (ii) there exists a risk-neutral measure;
- (iii) there is no arbitrage.

We only use an example to illustrate this point. The general proof of “(iii) \implies (i)” is contained in [3, Sec. 6.2].

Example 5.2. Suppose that $n = 2, d = 1$ and the processes $\alpha^i(t), \sigma_j^i(t), R_t$ are deterministic constants. In this case, the market price of risk equation becomes

$$\begin{cases} \alpha^1 - r = \sigma^1 \cdot \theta, \\ \alpha^2 - r = \sigma^2 \cdot \theta. \end{cases}$$

This system is solvable if and only if

$$\frac{\alpha^1 - r}{\sigma^1} = \frac{\alpha^2 - r}{\sigma^2}.$$

Suppose that

$$\mu \triangleq \frac{\alpha^1 - r}{\sigma^1} - \frac{\alpha^2 - r}{\sigma^2} > 0.$$

We consider the portfolio process $\Delta = \{(\Delta_t^1, \Delta_t^2)\}$ defined by

$$\Delta_t^1 = \frac{1}{S_t^1 \sigma^1}, \quad \Delta_t^2 \triangleq -\frac{1}{S_t^2 \sigma^2}.$$

Direct calculation shows that

$$d(D_t X_t) = D_t(dX_t - rX_t dt) = \mu D_t dt$$

in this case. Since $\mu, D_t > 0$, this implies that the discounted process $\{D_t X_t\}$ is strictly increasing. In other words, one can earn faster than the interest rate for sure and this leads to an arbitrage opportunity.

Next, we consider a derivative security with payoff V_T at maturity T . In a similar way, the parallel equation of (5.13) for the hedging strategy $\Delta = \{(\Delta_t^1, \dots, \Delta_t^n)\}$ is given by

$$\Gamma_t^j = \sum_{i=1}^n D_t \Delta_t^i S_t^i \sigma_j^i(t), \quad (5.17)$$

where Γ is now a d -dimensional process such that

$$D_t V_t = V_0 + \sum_{j=1}^d \int_0^t \Gamma_s^j d\tilde{B}_s^j. \quad (\tilde{B}_t^j \triangleq B_t^j + \int_0^t \theta_s^j ds)$$

In matrix notation, the algebraic system (5.17) is expressed as

$$\Gamma_t = \sigma_t^T \cdot (DS\Delta)_t,$$

where Γ_t , $(DS\Delta)_t$ are written as column vectors and $(DS\Delta)_t^i \triangleq D_t S_t^i \Delta_t^i$. The system (5.17) is called the *hedging equation*. This system need not always be solvable. If it is solvable for every given V_T , a hedging strategy always exists and by definition the market model is complete. As seen by the following result, this is essentially related to the uniqueness of risk-neutral measures. We only sketch the proof and refer the reader to [18, Sec. 5.4.4] for the complete details.

Theorem 5.4 (The second fundamental theorem of asset pricing). *Suppose that the market price of risk equation (5.16) is solvable so that a risk-neutral measure exists. Then the market model is complete if and only if there is a unique risk-neutral measure.*

Sketch of proof. Sufficiency. The uniqueness of risk-neutral measures implies that the coefficient matrix σ_t for the system (5.16) is injective (i.e. defining an injective linear transform from \mathbb{R}^d into \mathbb{R}^n). For if it were not the case, there will be more than one solutions to the system (5.16), leading to different risk-neutral measures which is a contradiction. But from linear algebra, we know that the injectivity of σ_t is equivalent to the surjectivity of σ_t^T , and the latter is merely a restatement of the solvability for the hedging equation (5.17) for every given Γ .

Necessity. Suppose that the market model is complete. Let $\tilde{\mathbb{P}}_1$ and $\tilde{\mathbb{P}}_2$ be two risk-neutral measures. Consider a derivative security whose payoff is $V_T = \frac{1}{D_T} \mathbf{1}_A$ where $A \in \mathcal{F}_T^B$ is given fixed. By the assumption, there exists a hedge of V_T whose associated portfolio value process is denote as $\{V_t\}$. Since $\{D_t V_t\}$ is a martingale under both $\tilde{\mathbb{P}}_1$ and $\tilde{\mathbb{P}}_2$, one finds that

$$\tilde{\mathbb{P}}_1(A) = \tilde{\mathbb{E}}_1[D_T V_T] = V_0 = \tilde{\mathbb{E}}_2[D_T V_T] = \tilde{\mathbb{P}}_2(A).$$

Therefore, $\tilde{\mathbb{P}}_1 = \tilde{\mathbb{P}}_2$. □

6 List of exercises

This chapter is a vital part of the course. It contains a total of 50 carefully chosen exercises that are intimately related to materials in the main text. With only a few exceptions, most of the problems can be solved without the need of knowing deeper tools from stochastic calculus that are not covered at the current introductory level (e.g. measure-theoretic techniques, localisation techniques, topological/functional analytic considerations). To keep abstract contents at a minimum level, many problems are put on concrete settings and involve explicit analysis. However, almost none of the problems are routine and most of them require deeper thinking but are certainly reachable for one who understands the main text well.

I must admit that most of these problems are by no mean my original invention. Some of them (or their variants) are so fundamental and inspiring that they have been included in almost every classical textbook. I am deeply indebted to many great probabilists (K. Chung, W. Feller, N. Ikeda, K. Itô, F. Le Gall, H. McKean, D. Revuz, L. Rogers, D. Stroock, S. Varadhan, S. Watanabe, D. Williams, M. Yor among others), from whom I learned the elegant theories and gained the tremendous joy of solving puzzles back to my own student time. It is now the time for you to have fun!

Exercise 6.1. Let X, Y be random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let \mathcal{G} be a sub- σ -algebra of \mathcal{F} . All relevant expectations are assumed to exist.

(i) Show that $\mathbb{E}[X\mathbb{E}[Y|\mathcal{G}]] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]]$.

(ii) Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a Borel measurable function. Suppose that X is \mathcal{G} -measurable, and Y, \mathcal{G} are independent. Show that

$$\mathbb{E}[f(X, Y)|\mathcal{G}] = \varphi(X),$$

where $\varphi(x) \triangleq \mathbb{E}[f(x, Y)]$ for $x \in \mathbb{R}$.

(iii) Suppose that $\sigma(\sigma(X), \mathcal{G})$ and \mathcal{H} are independent. Show that

$$\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}].$$

(iv) Let X, Y be random variables that satisfy

$$\mathbb{E}[X|Y] = Y, \quad \mathbb{E}[Y|X] = X \quad \text{a.s.}$$

Show that $\mathbb{P}(X = Y) = 1$.

Exercise 6.2. Let \mathcal{W} be the space of continuous paths $w : [0, \infty) \rightarrow \mathbb{R}$. The coordinate process $X = \{X_t : t \geq 0\}$ on \mathcal{W} is defined by $X_t(w) \triangleq w_t$. Let $\mathcal{B}(\mathcal{W}) \triangleq \sigma(X_t : t \geq 0)$ and let $\{\mathcal{F}_t^X : t \geq 0\}$ be the natural filtration of X .

(i) Given $t \geq 0$, define \mathcal{G} to be the class of subsets $A \in \mathcal{B}(\mathcal{W})$ that satisfy the following property: for any $w, w' \in \mathcal{W}$, if $w \in A$ and $w_s = w'_s$ for all $s \in [0, t]$, then $w' \in A$. Show that $\mathcal{G} = \mathcal{F}_t^X$.

(ii) Let $\tau : \mathcal{W} \rightarrow [0, \infty]$ be a $\mathcal{B}(\mathcal{W})$ -measurable map. Show that τ is an $\{\mathcal{F}_t^X\}$ -stopping time if and only if the following property holds: for any $w, w' \in \mathcal{W}$ with $w = w'$ on $[0, \tau(w)] \cap [0, \infty)$, one has $\tau(w') = \tau(w)$.

(iii) Let τ be an $\{\mathcal{F}_t^X\}$ -stopping time and let $A \in \mathcal{B}(\mathcal{W})$. Show that $A \in \mathcal{F}_\tau^X$ if and only if for every $w \in \mathcal{W}$, $w \in A \iff w^{\tau(w)} \in A$, where $w^{\tau(w)}$ is the stopped path defined by $w_t^{\tau(w)} \triangleq w_{\tau(w) \wedge t}$ ($t \geq 0$).

(iv) Let τ be an $\{\mathcal{F}_t^X\}$ -stopping time. By using the previous parts, show that $\mathcal{F}_\tau^X = \sigma(X_{\tau \wedge t} : t \geq 0)$.

Exercise 6.3. A sequence of integrable random variables $\{X_n : n \geq 0\}$ is said to be *uniformly integrable*, if

$$\lim_{\lambda \rightarrow \infty} \sup_{n \geq 0} \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > \lambda\}}] = 0.$$

(i) Suppose that there exists a constant $M > 0$ such that $\mathbb{E}[X_n^2] \leq M$ for all n . Show that $\{X_n\}$ is uniformly integrable.

(ii) Let $\{X_n\}$ be a uniformly integrable sequence such that $X_n \rightarrow X$ a.s. for some random variable X . Show that X is integrable and X_n converges to X in L^1 .

(iii) Let X be an integrable random variable and let $\{\mathcal{F}_n\}$ be a given filtration. Define $X_n \triangleq \mathbb{E}[X | \mathcal{F}_n]$. Show that $\{X_n\}$ is a uniformly integrable, $\{\mathcal{F}_n\}$ -martingale. Deduce that X_n converges to some random variable Y both a.s. and in L^1 . How do you describe Y ?

(iv) Let X_n and X be random variables such that $X_n \rightarrow X$ a.s. Suppose that there exists a non-negative, integrable random variable Y with $|X_n| \leq Y$ for all n . Let $\{\mathcal{F}_n\}$ be an arbitrarily given filtration and set $\mathcal{F}_\infty \triangleq \sigma(\cup_n \mathcal{F}_n)$. Show that $\mathbb{E}[X_n | \mathcal{F}_n]$ converges to $\mathbb{E}[X | \mathcal{F}_\infty]$ a.s. and in L^1 .

Exercise 6.4. At the initial time $n = 0$, an urn contains one black ball and one white ball. At each time $n \geq 1$, a ball is chosen from the urn uniformly at random and it is replaced back to the urn along with a new ball of the same colour. Let B_n

(respectively, M_n) denote the number of black balls (respectively, the proportion of black balls) in the urn right after time n .

- (i) Find the distribution of B_n .
- (ii) Show that $\{M_n : n \geq 0\}$ is a martingale with respect to its natural filtration.
- (iii) Show that M_n is convergent both a.s. and in L^1 . What is the limiting distribution of M_n ?

Exercise 6.5. Tom and Jerry are gambling with each other. Their initial capitals at time $n = 0$ are a and b respectively, where a, b are given positive integers. At each round $n \geq 1$, either Tom wins one dollar from Jerry or the otherwise. Suppose that Tom's winning probability at each round is $p \in (0, 1)$ and assume that $p \neq 1/2$. All rounds are assumed to be independent. The game is finished if either one of them goes bankrupt. Let τ be the time that the game is finished and let γ be the probability that Tom first goes bankrupt.

- (i) Let $\{X_n : n \geq 1\}$ be an i.i.d. sequence such that

$$\mathbb{P}(X_n = 1) = p, \quad \mathbb{P}(X_n = -1) = 1 - p.$$

Define $S_n \triangleq X_1 + \cdots + X_n$ where $S_0 \triangleq a$. Find two real numbers α, β , such that

$$M_n \triangleq \alpha^{S_n}, \quad N_n \triangleq S_n - \beta n$$

are martingales with respect to their natural filtrations.

- (ii) Use the result of Part (i) to compute γ and $\mathbb{E}[\tau]$.
- (iii) Solve Part (ii) again under the assumption that $p = 1/2$.

Exercise 6.6. This problem provides an enlightening method of simulating the number e . Let $\{X_n : n \geq 1\}$ be an i.i.d. sequence of uniform random variables over $[0, 1]$. Define $S_0 \triangleq 0$ and $S_n \triangleq X_1 + \cdots + X_n$ for $n \geq 1$. Let $\tau \triangleq \inf\{n : S_n > 1\}$.

- (i) Let $f : [0, 2] \rightarrow \mathbb{R}$ be a function such that

$$f(x) = \int_0^1 f(x+t)dt + 1 \quad \forall x \in [0, 1]. \tag{6.1}$$

Show that $\{f(S_{\tau \wedge n}) + \tau \wedge n : n \geq 0\}$ is a martingale with respect to the natural filtration of $\{S_n\}$.

- (ii) Use the result of Part (i) to show that $e = \mathbb{E}[\tau]$.

Exercise 6.7. Consider the probability space $([0, 1], \mathcal{B}([0, 1]), dt)$ where dt is the Lebesgue measure.

(i) Let $\mathcal{P} : 0 = t_0 < t_1 < \dots < t_n = 1$ be a finite partition of $[0, 1]$ and define

$$\mathcal{F}_{\mathcal{P}} \triangleq \sigma(\{[t_{k-1}, t_k] : k = 1, \dots, n\}).$$

Compute $\mathbb{E}[\varphi | \mathcal{F}_{\mathcal{P}}]$ for any given integrable function $\varphi : [0, 1] \rightarrow \mathbb{R}$.

(ii) Let $\gamma : [0, 1] \rightarrow \mathbb{R}$ be an absolutely continuous function and define

$$\|\gamma\|_{1\text{-var}} \triangleq \sup_{\mathcal{P}} \sum_{t_i \in \mathcal{P}} |\gamma(t_i) - \gamma(t_{i-1})| < \infty,$$

where the supremum is taken over all finite partitions of $[0, 1]$. For each $n \geq 1$, let $\mathcal{P}_n = \{k/2^n\}_{k=0}^{2^n}$ be the n -th dyadic partition of $[0, 1]$. Define $\gamma_n : [0, 1] \rightarrow \mathbb{R}$ to be the linear interpolation of γ over \mathcal{P}_n , i.e. $\gamma_n = \gamma$ on the partition points and γ_n is linear on each sub-interval $[(k-1)/2^n, k/2^n]$. Show that $\{\gamma_n' : n \geq 1\}$ is a martingale on $([0, 1], \mathcal{B}([0, 1]), dt)$ with respect to a suitable filtration.

(iii) By using the martingale convergence theorem, show that

$$\lim_{n \rightarrow \infty} \|\gamma_n - \gamma\|_{1\text{-var}} = 0.$$

Exercise 6.8. The aim of this problem is to study recurrence/transience of Markov chains by using martingale methods. Let $X = \{X_n : n \geq 0\}$ be a Markov chain with a countable state space S and one-step transition matrix $P = (P_{ij})_{i,j \in S}$. A function $f : S \rightarrow \mathbb{R}$ is said to be P -superharmonic if

$$(Pf)(i) \triangleq \sum_{j \in S} P_{ij} f(j) \leq f(i)$$

for all $i \in S$, provided that Pf is well defined.

(i) Let $f : S \rightarrow [0, \infty)$ be a given function. Define

$$Y_n^f \triangleq f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (Pf - f)(X_k), \quad Y_0^f \triangleq 0.$$

Show that $\{Y_n^f\}$ is a martingale with respect to the natural filtration of X . Conclude that

$$f \text{ is } P\text{-superharmonic} \implies \{f(X_n)\} \text{ is a supermartingale.}$$

(ii) Let $F \subseteq S$. Define $\tau \triangleq \inf\{n \geq 1 : X_n \in F\}$. Let $f : S \rightarrow [0, \infty)$ be P -superharmonic outside F . Show that

$$\mathbb{E}[f(X_{\tau \wedge n}) | X_0 = i] \leq f(i) \quad \forall i \in F^c, n \geq 0.$$

(iii) Suppose that X is irreducible. If X is recurrent, show that any non-negative P -superharmonic function must be constant. Conversely, if X is transient, for each fixed $j \in S$ show that the function

$$S \ni i \mapsto G(i, j) \triangleq \sum_{n=0}^{\infty} P_{ij}^n \quad (P_{ij}^n \triangleq \mathbb{P}(X_n = j | X_0 = i))$$

is P -superharmonic and non-constant. Conclude that X is recurrent if and only if all non-negative P -superharmonic functions are constant.

(iv) Let $j \in S$ be a fixed state. Let $\{B_m : m \geq 1\}$ be an increasing family of non-empty subsets of S such that $j \in B_0$ and for each m , with probability one X (starting at j) exits B_m in finite time. Suppose that there exists $f : S \rightarrow [0, \infty)$ which is P -superharmonic on $\{j\}^c$ and

$$a_m \triangleq \inf_{i \notin B_m} f(i) \rightarrow \infty \quad \text{as } m \rightarrow \infty.$$

By using the result of Part (ii), show that

$$f(j) \geq a_m \mathbb{P}(\tau_m \leq \rho_j | X_0 = j) \quad \forall m \geq 1,$$

where $\tau_m \triangleq \inf\{n \geq 1 : X_n \notin B_m\}$ and $\rho_j \triangleq \inf\{n \geq 1 : X_n = j\}$. Conclude that j is recurrent in this case.

(v) Let X be the simple random walk on \mathbb{Z} (i.e. with probability 1/2 jumping left/right). By constructing a suitable function f in Part (iv), show that X is recurrent.

(vi) Let X be the simple random walk on \mathbb{Z}^2 (i.e. with probability 1/4 jumping up/down/left/right in each step). By considering the function

$$f(k) \triangleq \begin{cases} \log(k_1^2 + k_2^2 - 1/2), & k = (k_1, k_2) \neq (0, 0); \\ \kappa, & k = (0, 0) \end{cases}$$

with a suitably chosen κ , use the result of Part (iv) to show that X is recurrent.

(vii-a) Let X be the simple random walk on \mathbb{Z}^3 . Define the function

$$f_\alpha(k) \triangleq (\alpha^2 + |\mathbf{k}|^2)^{-1/2}, \quad \mathbf{k} = (k_1, k_2, k_3) \in \mathbb{Z}^3$$

where $\alpha > 0$ is a parameter to be specified later on. By algebraic manipulation, show that f_α is superharmonic if and only if

$$\left(1 - \frac{1}{M}\right)^{-1/2} \geq \frac{1}{6} \sum_{i=1}^3 \frac{(1 + 2x_i)^{1/2} + (1 - 2x_i)^{1/2}}{(1 - 4x_i^2)^{1/2}} \quad \forall \mathbf{k} \in \mathbb{Z}^3,$$

where $M \triangleq 1 + \alpha^2 + |\mathbf{k}|^2$ and $x_i \triangleq k_i/M$ ($i = 1, 2, 3$).

(vii-b) Show that

$$\frac{1}{2}((1 + \xi)^{1/2} + (1 - \xi)^{1/2}) \leq 1 - \frac{1}{8}\xi^2$$

for all $\xi \in [-1, 1]$. Conclude that

$$\frac{1}{6} \sum_{i=1}^3 \frac{(1 + 2x_i)^{1/2} + (1 - 2x_i)^{1/2}}{(1 - 4x_i^2)^{1/2}} \leq \frac{1}{3} \sum_{i=1}^3 \frac{1}{(1 - 4x_i^2)^{1/2}} - \frac{1}{6}|\mathbf{x}|^2.$$

(vii-c) Show that there exists a universal constant $\beta < 1$, such that $|2x_i| \leq \beta$ for all $i = 1, 2, 3$ and all $\alpha \geq 1$. In addition, by considering the function $(1 - y)^{-1/2}$ ($y \in [0, \beta^2]$), show that

$$\frac{1}{3} \sum_{i=1}^3 \frac{1}{(1 - 4x_i^2)^{1/2}} \leq 1 + \frac{2}{3}|\mathbf{x}|^2 + C|\mathbf{x}|^4$$

where C is some universal constant.

(vii-d) By combining all the previous steps with the extra observation that

$$\left(1 - \frac{1}{M}\right)^{-1/2} \geq 1 + \frac{1}{2M},$$

show that f_α is superharmonic provided that α is large enough. Use the result of Part (iii) to conclude that X is transient.

(vii-e) By reducing to the three-dimensional case or otherwise, show that the simple random walk on \mathbb{Z}^d ($d \geq 4$) is also transient.

Exercise 6.9. (i) Show that $\log t \leq t/e$ for every $t > 0$ and conclude that

$$a \log^+ b \leq a \log^+ a + \frac{b}{e} \quad \forall a, b > 0,$$

where $\log^+ t = \max\{\log t, 0\}$ ($t > 0$).

(ii) Let $\{X_t, \mathcal{F}_t : t \geq 0\}$ be a continuous, non-negative submartingale. Given $T > 0$, set

$$X_T^* \triangleq \sup_{t \in [0, T]} X_t.$$

Let $\rho : [0, \infty) \rightarrow \mathbb{R}$ be a continuous, increasing function with $\rho(0) = 0$. Show that

$$\mathbb{E}[\rho(X_T^*)] \leq \mathbb{E}\left[X_T \int_0^{X_T^*} \lambda^{-1} d\rho(\lambda)\right].$$

(iii) By choosing a function $\rho(t)$ suitably, show that

$$\mathbb{E}[X_T^*] \leq \frac{e}{e-1} (1 + \mathbb{E}[X_T \log^+ X_T]).$$

Exercise 6.10. A casino is proposing a new game called ALPHABETALPHA. The dealer rolls a die with 26 faces (one letter per face) repeatedly. At every round, precisely one gambler enters the game and she bets in the following way. She bets \$1 on the first letter A in the string ALPHABETALPHA. She quits if she loses, while if she wins the dealer pays her \$26 dollars and she bets all this amount on the second letter L at the next round. If she loses she quits while if she wins again the dealer pays her \$26² and she further bets all the money on the third letter P at the next round. The strategy continues and she quits the game when she loses at some point or she wins the entire string. Let \mathcal{F}_n denote the σ -algebra generated by the outcomes of the first n tosses.

(i) Let $M = \{M_n : n \geq 1\}$ be the net gain of the casino up to the n -th round. Show that M is a martingale with uniformly bounded increments, i.e. there exists $C > 0$ such that $|M_n(\omega) - M_{n-1}(\omega)| \leq C$ for all ω, n .

(ii) Let τ be the first time that the string ALPHABETALPHA appears. Show that there exist $N \geq 1$ and $\varepsilon > 0$, such that

$$\mathbb{P}(\tau \leq N + n | \mathcal{F}_n) \geq \varepsilon \quad \text{a.s.}$$

for all $n \geq 1$.

(iii) Show that the property in Part (ii) implies that

$$\mathbb{P}(\tau > kN + r) \leq (1 - \varepsilon)^{k-1} \mathbb{P}(T > N + r) \quad \forall k, r \geq 1.$$

Use this inequality to deduce that $\mathbb{E}[\tau] < \infty$.

(iv) Prove the following extension of the optional sampling theorem: if X is a

discrete-time martingale with uniformly bounded increments and σ is an integrable stopping time, then $\mathbb{E}[X_\sigma] = \mathbb{E}[X_0]$.

(v) Use the above steps to compute $\mathbb{E}[\tau]$.

Exercise 6.11. Let B be a one-dimensional Brownian motion.

(i) Define

$$X_t \triangleq \begin{cases} tB_{1/t}, & t > 0; \\ 0, & t = 0, \end{cases}$$

Show that $\{X_t : t \geq 0\}$ is also a Brownian motion.

(ii) Show that with probability one, there exist two sequences of positive times $s_n \downarrow 0$, $t_n \downarrow 0$, such that $B_{s_n} < 0$ and $B_{t_n} > 0$ for every n .

(iii) Show that with probability one, $t \mapsto B_t$ is not differentiable at $t = 0$. Conclude that with probability one, $t \mapsto B_t$ is non-differentiable everywhere outside a set of zero Lebesgue measure.

(iv) Let $s < u < t$. Compute $\mathbb{E}[B_u | \sigma(B_s, B_t)]$.

Exercise 6.12. Let B be a one-dimensional $\{\mathcal{F}_t\}$ -Brownian motion.

(i) Suppose that τ is an integrable, $\{\mathcal{F}_t\}$ -stopping time. Show that $\mathbb{E}[B_\tau] = 0$ and $\mathbb{E}[B_\tau^2] = \mathbb{E}[\tau]$.

(ii) Find an example of a stopping time τ for which $\mathbb{E}[B_\tau] \neq 0$.

(iii) Find two stopping times $\sigma \leq \tau$ with $\mathbb{E}[\sigma] < \infty$, such that

$$\mathbb{E}[B_\sigma^2] > \mathbb{E}[B_\tau^2].$$

Exercise 6.13. Let $B = \{(X_t, Y_t) : t \geq 0\}$ be a two-dimensional Brownian motion starting at the point $(0, 1)$. Let τ be the first time that B hits the x -axis.

(i) Show that X_τ is a standard Cauchy random variable, i.e. with probability density function $\frac{1}{\pi(1+x^2)}$ ($x \in \mathbb{R}$).

(ii) Use the result of Part (i) to compute the characteristic function of the standard Cauchy distribution.

(iii) For each $r > 1$, define $\tau_r \triangleq \inf\{t : Y_t = r\}$. Show that the process $\{X_{\tau_r} : r > 1\}$ has independent increments.

(iv) Denote $\mathbb{H} \triangleq \{(x, y) : y > 0\}$ as the upper-half plane. Let f be a bounded, uniformly continuous function on \mathbb{R} . Use the result of Part (i) to show that the

unique harmonic function $u(x, y)$ on \mathbb{H} (i.e. $\Delta u = 0$ on \mathbb{H}) that satisfies $u|_{\partial\mathbb{H}} = f$ is given by

$$u(x, y) = \int_{\mathbb{R}} P_y(x - t) f(t) dt,$$

where

$$P_y(x) \triangleq \frac{y}{\pi(x^2 + y^2)}, \quad (x, y) \in \mathbb{H}$$

is the so-called *Poisson kernel* on the upper-half plane.

Exercise 6.14. Let $\mathcal{B}_b(\mathbb{R})$ denote the space of bounded, Borel measurable functions on \mathbb{R} . Define a family of linear operators $P_t : \mathcal{B}_b(\mathbb{R}) \rightarrow \mathcal{B}_b(\mathbb{R})$ ($t \geq 0$) by

$$(P_t f)(x) \triangleq \mathbb{E}[f(x + B_t)], \quad f \in \mathcal{B}_b(\mathbb{R}),$$

where B is a one-dimensional $\{\mathcal{F}_t\}$ -Brownian motion starting at the origin.

(i) Show that $P_{t+s} = P_t \circ P_s$ for any $s, t \geq 0$.

(ii) Let τ be a finite $\{\mathcal{F}_t\}$ -stopping time. Show that

$$\mathbb{E}[f(B_{\tau+t}) | \mathcal{F}_\tau] = P_t f(B_\tau) \quad \forall f \in \mathcal{B}_b(\mathbb{R}).$$

(iii) Let σ, τ be two finite $\{\mathcal{F}_t\}$ -stopping times such that $\tau \in \mathcal{F}_\sigma$ and $\sigma \leq \tau$. Show that

$$\mathbb{E}[f(B_\tau) | \mathcal{F}_\sigma] = P_{\tau-\sigma} f(x) |_{x=B_\sigma} \quad \forall f \in \mathcal{B}_b(\mathbb{R}).$$

Is it always true that $B_\tau - B_\sigma$ and \mathcal{F}_σ are independent?

Exercise 6.15. The aim of this problem is to prove Khinchin's law of the iterated logarithm for Brownian motion:

$$\mathbb{P}\left(\overline{\lim}_{t \rightarrow 0} \frac{B_t}{\sqrt{2t \log \log 1/t}} = 1\right) = 1.$$

(i) By using the exponential martingale $e^{\alpha B_t - \alpha^2 t/2}$, show that

$$\mathbb{P}\left(\sup_{0 \leq s \leq t} \left(B_s - \frac{\alpha s}{2}\right) > \beta\right) \leq e^{-\alpha\beta} \quad \forall t, \alpha, \beta > 0.$$

(ii) Define $h(t) \triangleq \sqrt{2t \log \log 1/t}$. Let $0 < \theta, \delta < 1$ be given. By taking

$$\alpha \triangleq (1 + \delta)\theta^{-n} h(\theta^n), \quad \beta \triangleq h(\theta^n)/2$$

in Part (i), show that with probability one,

$$\sup_{0 \leq s \leq \theta^{n-1}} B_s \leq \left(\frac{1+\delta}{2\theta} + \frac{1}{2} \right) h(\theta^n) \quad \text{for all sufficiently large } n.$$

Conclude that with probability one,

$$B_t \leq \left(\frac{1+\delta}{2\theta} + \frac{1}{2} \right) h(t) \quad \text{for all sufficiently small } t.$$

Use this to deduce that $\overline{\lim}_{t \rightarrow 0} \frac{B_t}{\sqrt{2t \log \log 1/t}} \leq 1$ a.s.

(iii) By using the second Borel-Cantelli lemma, show that with probability one,

$$B_{\theta^n} \geq (1 - \sqrt{\theta})h(\theta^n) + B_{\theta^{n+1}} \quad \text{for infinitely many } n.$$

Use this fact and the result of Part (i) to deduce that with probability one,

$$B_{\theta^n} \geq (1 - \sqrt{\theta})h(\theta^n) - 2h(\theta^{n+1}) \quad \text{for infinitely many } n.$$

Conclude that $\overline{\lim}_{t \rightarrow 0} \frac{B_t}{\sqrt{2t \log \log 1/t}} \geq 1$ a.s.

Exercise 6.16. Let $\{B_t : t \geq 0\}$ be a one-dimensional Brownian motion. Define the running maximum process

$$S_t \triangleq \sup_{0 \leq s \leq t} B_s, \quad t \geq 0.$$

(i) Establish the following analytical fact. Let $b : [0, \infty) \rightarrow \mathbb{R}$ be a continuous function with $b(0) = 0$ and define

$$s(t) \triangleq \sup_{0 \leq s \leq t} b(s), \quad t \geq 0.$$

Let I be an open interval on $[0, \infty)$ that does not intersect the set $\{t : s(t) = b(t)\}$. Show that

$$\int_0^\infty (s(u) - b(u)) \mathbf{1}_I(u) ds(u) = 0.$$

Deduce that

$$\int_0^\infty (s(u) - b(u)) f(u) ds(u) = 0$$

for any bounded, Borel measurable function $f : [0, \infty) \rightarrow \mathbb{R}$.

(ii) Let f be a bounded, continuous function on $[0, \infty)$ and define $F(x) \triangleq \int_0^x f(y)dy$. Use the result of Part (i) to show that

$$(S_t - B_t)f(S_t) = F(S_t) - \int_0^t f(S_u)dB_u.$$

(iii) For each given $\lambda > 0$, show that the process

$$e^{-\lambda S_t} + \lambda(S_t - B_t)e^{-\lambda S_t}$$

is a martingale with respect to the Brownian filtration.

(iv) Let $c > 0$ and define $\tau \triangleq \inf\{t : S_t - B_t = c\}$. Show that $\tau < \infty$ a.s. and S_τ has the exponential distribution with parameter $1/c$.

Exercise 6.17. Let B be a one-dimensional Brownian motion and let $S_1 \triangleq \sup_{0 \leq t \leq 1} B_t$. Define

$$\sigma \triangleq \sup\{t < 1 : B_t = 0\}, \quad \tau \triangleq \inf\{t \geq 0 : B_t = S_1\}$$

and

$$\zeta \triangleq \inf\{t > 1 : B_t = 0\}.$$

(i) Let $0 < s < t$. Show that

$$\mathbb{P}(B_u \neq 0 \forall u \in [s, t]) = \frac{2}{\pi} \arccos \sqrt{\frac{s}{t}}.$$

(ii) Show that both of σ and τ are distributed according to the arcsine law:

$$\mathbb{P}(\sigma \leq t) = \mathbb{P}(\tau \leq t) = \frac{2}{\pi} \arcsin \sqrt{t}, \quad t \in [0, 1].$$

(iii) Show that the joint density function of (σ, ζ) is given by

$$f_{\sigma, \zeta}(s, t) = \frac{1}{2\pi} s^{-1/2} (t-s)^{-3/2}, \quad 0 < s < 1 < t.$$

Find the probability density function of $\zeta - \sigma$ (the length of the Brownian excursion across the time $t = 1$).

(iv) Show that with probability one, there are two sequences $s_n < t_n < \sigma$ such that

$$s_n \uparrow \sigma, \quad B(s_n) < 0, \quad B(t_n) > 0.$$

Is $t \mapsto B_{\sigma+t} - B_\sigma$ a Brownian motion?

Exercise 6.18. Let B^x be a one-dimensional Brownian motion starting at $x \in \mathbb{R}$. Define $p_t(y) \triangleq \frac{1}{\sqrt{2\pi t}} e^{-y^2/2t}$.

(i) Show that the transition density function of the Brownian motion starting at $x > 0$ before hitting zero is given by

$$\mathbb{P}(B_t^x \in dy, t < \tau_0^x) = (p_t(y-x) - p_t(y+x))dy, \quad y > 0,$$

where $\tau_0^x \triangleq \inf\{t : B_t^x = 0\}$. In addition, given $y > 0$ show that

$$\mathbb{P}(\tau_0^x > t | X_t = y) = 1 - e^{-2xy/t}.$$

(ii) Let $\alpha, \beta > 0$. Show that

$$\mathbb{P}(B_t \leq \alpha t + \beta \quad \forall t \in [0, 1] | B_0 = B_1 = 0) = 1 - e^{-2\beta(\alpha+\beta)}.$$

(iii) Define

$$S_t \triangleq \sup_{0 \leq s \leq t} B_s^0, \quad I_t \triangleq \inf_{0 \leq s \leq t} B_s^0.$$

It is known that the probability density function of B_t^0 before exiting the interval (w, v) ($w < 0, v > 0$) is given by

$$\begin{aligned} \mathbb{P}(B_t^0 \in dy, t < \sigma_{w,v}) &= \sum_{n=-\infty}^{\infty} (p_t(2nv - 2nw - y) \\ &\quad - p_t(2(n-1)v - 2nw + y))dy, \quad y \in (w, v), \end{aligned} \quad (6.2)$$

where $\sigma_{w,v} \triangleq \inf\{t : B_t^0 \notin (w, v)\}$. Use this formula to show that the probability density function of $R_t \triangleq S_t - I_t$ (the range of Brownian motion) is given by

$$f_{R_t}(r) = 8 \sum_{n=1}^{\infty} (-1)^{n-1} n^2 p_t(nr), \quad r > 0.$$

Exercise 6.19. Let B be a one-dimensional Brownian motion.

(i) Show that $X \triangleq \{|B_t| : t \geq 0\}$ is a strong Markov process with respect to its natural filtration. In addition, its transition density function is given by

$$\mathbb{P}(X_t \in dy | X_0 = x) = \sqrt{\frac{2}{\pi t}} \exp\left(-\frac{x^2 + y^2}{2t}\right) \cosh\left(\frac{xy}{t}\right) dy, \quad y > 0.$$

(ii) Let $S_t \triangleq \sup_{0 \leq s \leq t} B_s$. By using the strong Markov property of Brownian motion, show that the two-dimensional process $t \mapsto (S_t, S_t - B_t)$ is a strong Markov process with respect to its natural filtration. In addition, given $f : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ and $x_0, y_0 > 0$, find an integral expression of

$$\mathbb{E}[f(S_{s+t}, S_{s+t} - B_{s+t}) | (S_s, S_s - B_s) = (x_0, y_0)].$$

(iii) Use the above results to show that the two processes X and $S - B$ have the same distribution in the sense that

$$\mathbb{E}[f(X_{t_1}, \dots, X_{t_n})] = \mathbb{E}[f(S_{t_1} - B_{t_1}, \dots, S_{t_n} - B_{t_n})]$$

for any choices of $n, t_1 < \dots < t_n$ and bounded, measurable $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Exercise 6.20. Let $c \neq 0$ be a given fixed real number. Consider the process $X_t^x \triangleq B_t^x + ct$ where B^x is a one-dimensional Brownian motion starting at $x \in \mathbb{R}$.

(i) Let \mathcal{W} be the space of continuous paths on $[0, \infty)$ equipped with the natural filtration $\{\mathcal{F}_t\}$ of the coordinate process. Let \mathbb{Q}^x (respectively, \mathbb{P}^x) denote the law of X^x (respectively, B^x) over \mathcal{W} . Show that for each $t \geq 0$, when restricted to \mathcal{F}_t the probability measure \mathbb{Q}^x is absolutely continuous with respect to \mathbb{P}^x with density function

$$\frac{d\mathbb{Q}^x}{d\mathbb{P}^x} \Big|_{\mathcal{F}_t}(w) = e^{c(w_t - x) - c^2 t/2}, \quad w \in \mathcal{W}.$$

Is it true that \mathbb{Q}^x is absolutely continuous to \mathbb{P}^x on $\mathcal{B}(\mathcal{W})$ (the σ -algebra generated by the coordinate process on $[0, \infty)$)?

(ii) Define $S_t = \sup_{0 \leq s \leq t} X_s^0$. Compute the joint density function of (S_t, X_t^0) .

(iii) Given $\theta \in \mathbb{R}$ and $\lambda > 0$, define the process $M_t \triangleq \exp(\theta X_t^0 - \lambda t)$. Find the suitable relation between θ and λ , under which the process $\{M_t\}$ is a martingale.

(iv) Let $a \neq 0$ and define $\tau_a^0 \triangleq \inf\{t : X_t^0 = a\}$. By using the exponential martingale in Part (iii), compute the Laplace transform of τ_a^0 and $\mathbb{P}(\tau_a^0 < \infty)$.

(v) Suppose that $c > 0$. Let $\sigma \triangleq \sup\{t : X_t^0 = 0\}$. Show that

$$\mathbb{P}(\sigma > t | \mathcal{F}_t^B) = e^{-2c \max\{X_t^0, 0\}},$$

where $\{\mathcal{F}_t^B : t \geq 0\}$ is the natural filtration of $\{B_t^0\}$. Use this relation to derive the probability density function of σ .

Exercise 6.21. (i) Find a solution to Skorokhod's embedding problem for the discrete uniform distribution on $\{-2, -1, 0, 1, 2\}$.

(ii) Describe the way of constructing a solution to Skorokhod's embedding problem for the uniform distribution over $(-1, 1)$.

(iii) Let $f(x)$ be a given probability density function on \mathbb{R} with mean zero and finite variance σ^2 . Define

$$\mu(dx, dy) \triangleq c(y-x)f(x)f(y)dxdy, \quad x < 0 < y$$

where $c > 0$ is the normalising constant so that μ is a joint probability density function. Suppose that there exists a random vector (X, Y) such that (X, Y) is μ -distributed and it is independent of the Brownian motion B . Use (X, Y) and B to construct a random time τ such that B_τ has probability density function f and $\mathbb{E}[\tau] = \sigma^2$.

Exercise 6.22. Let M be a square integrable continuous martingale with respect to its natural filtration.

(i) Define $\langle M \rangle_\infty \triangleq \lim_{t \rightarrow \infty} \langle M \rangle_t$. Show that $\mathbb{E}[\langle M \rangle_\infty] < \infty$ if and only if there exists a square integrable random variable M_∞ such that

$$\lim_{t \rightarrow \infty} \mathbb{E}[|M_t - M_\infty|^2] = 0.$$

(ii) Show that M is a Gaussian process if and only if the quadratic variation of M is deterministic (i.e. there exists a continuous, non-decreasing function $f : [0, \infty) \rightarrow \mathbb{R}$ vanishing at the origin, such that with probability one $\langle M \rangle_t = f(t)$ for all t). In this case, M has independent increments.

Exercise 6.23. Let B^x be a d -dimensional Brownian motion starting at x .

(i) Let f be an arbitrary smooth function on \mathbb{R}^d with compact support. Show that the process

$$M_t \triangleq f(B_t^x) - f(x) - \frac{1}{2} \int_0^t (\Delta f)(B_s^x) ds$$

is a martingale with respect to the natural filtration of B^x .

(ii) Show that $f(x) \triangleq \log|x|$ (in dimension $d = 2$) and $f(x) = |x|^{2-d}$ (in dimension $d \geq 3$) are harmonic functions on $\mathbb{R}^d \setminus \{0\}$ (i.e. $\Delta f(x) = 0$ for every $x \neq 0$).

(iii) Let $0 < a < |x| < b$. Define τ_a^x (respectively, τ_b^x) to be the hitting time of

the sphere $S_a \triangleq \{y \in \mathbb{R}^d : |y| = a\}$ (respectively, the sphere S_b) by the Brownian motion B^x . By choosing a suitable f based on Parts (i) and (ii), compute the probabilities $\mathbb{P}(\tau_a^x < \tau_b^x)$ and $\mathbb{P}(\tau_a^x < \infty)$ in all dimensions $d \geq 2$.

(iv) Let U be a non-empty, bounded open subset of \mathbb{R}^d . Define $\sigma_U^x = \sup\{t : B_t^x \in U\}$ to be the last time that B^x visits U . Show that $\mathbb{P}(\sigma_U^x = \infty) = 1$ in dimension $d = 2$, while $\mathbb{P}(\sigma_U^x < \infty) = 1$ in dimension $d \geq 3$.

(v) Let $y \in \mathbb{R}^d$ and define $\zeta_y^x \triangleq \inf\{t > 0 : B_t^x = y\}$. Show that $\mathbb{P}(\zeta_y^x < \infty) = 0$ in all dimensions $d \geq 2$.

Exercise 6.24. Let $\{\mu_t : t \geq 0\}$ and $\{\sigma_t : t \geq 0\}$ be given uniformly bounded, progressively measurable processes defined on some given filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$.

(i) Construct an Itô process X that satisfies the following equation

$$X_t = 1 + \int_0^t X_s \mu_s ds + \int_0^t X_s \sigma_s dB_s, \quad t \geq 0.$$

Show that such a process X is unique.

(ii) Suppose that there exists a constant $C > 0$ such that $\sigma_t(\omega) \geq C$ for all t and ω . Given fixed $T > 0$, construct a probability measure \mathbb{Q}_T on \mathcal{F}_T that is equivalent to \mathbb{P} , under which X becomes an $\{\mathcal{F}_t\}$ -martingale.

Exercise 6.25. Let $B = \{B_t : 0 \leq t \leq 1\}$ be a one-dimensional Brownian motion.

(i) Let $Y \triangleq \int_0^1 B_t dt$. Find the unique progressively measurable process Φ such that

$$Y = \mathbb{E}[Y] + \int_0^1 \Phi_t dB_t.$$

(ii) Define $S_1 \triangleq \sup_{0 \leq t \leq 1} B_t$. By writing $\mathbb{E}[S_1 | \mathcal{F}_t^B]$ as a function of (t, S_t, B_t) , find the unique progressively measurable process Φ such that

$$S_1 = \mathbb{E}[S_1] + \int_0^1 \Phi_t dB_t.$$

Exercise 6.26. Let Φ be an Itô integrable process with respect to a one-dimensional Brownian motion B . Suppose that $\int_0^\infty \Phi_t^2 dt < \infty$ a.s.

(i) Show that

$$\int_0^\infty \Phi_t dB_t \triangleq \lim_{t \rightarrow \infty} \int_0^t \Phi_s dB_s$$

exists a.s. In addition, if $\mathbb{E}[\int_0^\infty \Phi_t^2 dt] < \infty$, then $\mathbb{E}[\int_0^\infty \Phi_t dB_t] = 0$ and

$$\mathbb{E}[(\int_0^\infty \Phi_t dB_t)^2] = \mathbb{E}[\int_0^\infty \Phi_t^2 dt].$$

(ii) Suppose that $\mathbb{E}[\exp(\frac{1}{2} \int_0^\infty \Phi_t^2 dt)] < \infty$. Show that

$$\mathbb{E}[\exp(i \int_0^\infty \Phi_t dB_t + \frac{1}{2} \int_0^\infty \Phi_t^2 dt)] = 1.$$

(iii) Construct an example of Φ which is not identically zero, such that $\int_0^\infty \Phi_t^2 dt < \infty$ while $\int_0^\infty \Phi_t dB_t = 0$.

Exercise 6.27. Consider a stochastic integral $M_t = \int_0^t \Phi_s dB_s$ defined over a right continuous filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ (i.e. $\mathcal{F}_{t+} = \mathcal{F}_t$ for all t). Let $A_t \triangleq \int_0^t \Phi_s^2 ds$ and suppose that $A_\infty = \infty$ a.s. For each $t \geq 0$, define

$$C_t \triangleq \inf\{s \geq 0 : A_s > t\}$$

to be the *generalised inverse* of A and set $W_t \triangleq M_{C_t}$.

(i) Show that C_t is an $\{\mathcal{F}_s\}$ -stopping time.

(ii) Let $t_1 < \dots < t_n$ and $\theta_1, \dots, \theta_n \in \mathbb{R}$. Show that

$$\mathbb{E}[\exp(i \sum_{j=1}^n \theta_j W_{t_j})] = \exp(-\frac{1}{2} \sum_{j,k=1}^n \theta_j \theta_k t_j \wedge t_k).$$

Conclude that $\{W_t : t \geq 0\}$ is a Brownian motion and thus $M_t = W_{A_t}$ is the time-change of a Brownian motion.

(iii) Let τ be an $\{\mathcal{F}_t\}$ -stopping time. Show that

$$\mathbb{P}(\sup_{0 \leq t \leq \tau} |M_t| \geq x, A_\tau \leq y) \leq 2e^{-\frac{x^2}{2y}}, \quad \forall x, y > 0.$$

How do you interpret this property heuristically?

Exercise 6.28. Consider the n -order iterated Itô integral

$$I_n \triangleq \int_0^1 \left(\int_0^{t_n} \cdots \left(\int_0^{t_2} dB_{t_1} \right) \cdots dB_{t_{n-1}} \right) dB_{t_n}.$$

- (i) Compute the variance of I_n .
- (ii) Show that $\mathbb{E}[I_m I_n] = 0$ if $m \neq n$.
- (iii) Let $\tau \triangleq \inf\{t \geq 0 : B_t \notin (-1, 1)\}$. By considering suitable Hermite polynomials, compute $\mathbb{E}[\tau^2]$.

Exercise 6.29. Let $B = \{(X_t, Y_t) : 0 \leq t \leq T\}$ be a two-dimensional Brownian motion. Define

$$L \triangleq \frac{1}{2} \left(\int_0^T X_t dY_t - \int_0^T Y_t dX_t \right).$$

The aim of this problem is to compute the characteristic function of L .

- (i) If $x, y : [0, T] \rightarrow \mathbb{R}$ are smooth paths, what is the geometric interpretation of the integral $\frac{1}{2} \int_0^T (x_t y'_t - y_t x'_t) dt$?
- (ii) Write $L = \int_0^T \langle AB_t, dB_t \rangle$ with a suitable deterministic 2×2 matrix A , where B_t is written as a column vector and $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on \mathbb{R}^2 .
- (iii) Define

$$h(\gamma, \mu) \triangleq \mathbb{E}[e^{i\langle \gamma, B_T \rangle + \mu L}], \quad \gamma \in \mathbb{R}^2, \mu \in \mathbb{R}.$$

Show that there exists $c > 0$, such that $h(\gamma, \mu)$ is finite for all $\gamma \in \mathbb{R}^2$ and $\mu \in (-c, c)$.

- (iv) Fix $\mu \in (-c, c)$ as in Part (iii). Let $k(t)$ be the solution to the following ODE

$$\begin{cases} k'(t) = -\mu^2 - k(t)^2, & 0 \leq t \leq T; \\ k(T) = 0, \end{cases}$$

and set $K(t) \triangleq \begin{pmatrix} k(t) & 0 \\ 0 & k(t) \end{pmatrix}$. Construct a probability measure $\tilde{\mathbb{P}}$ on \mathcal{F}_T , under which

$$\tilde{B}_t \triangleq B_t - \int_0^t (K(s) + \mu A) B_s ds, \quad 0 \leq t \leq T$$

is a Brownian motion and

$$h(\gamma, \lambda) = \tilde{\mathbb{E}} \left[\exp \left(i \langle \gamma, B_T \rangle - \frac{1}{4} \int_0^T k(t) \langle B_t, dB_t \rangle - \frac{1}{8} \int_0^T k'(t) |B_t|^2 dt \right) \right],$$

(v) By applying Itô's formula to the process $Y_t \triangleq k(t)|B_t|^2$ and finding $k(t)$ explicitly, show that $h(0, \mu) = \frac{1}{\cos \mu T/2}$. Conclude that the characteristic function of L is given by

$$\mathbb{E}[e^{i\lambda L}] = \frac{1}{\cosh \lambda T/2}, \quad \lambda \in \mathbb{R}.$$

(vi) Let $H(t) \in \text{Mat}(2, 2)$ be the solution to the following matrix equation:

$$\begin{cases} H'(t) = (K(t) + \mu A)H(t), \\ H(0) = \text{Id}. \end{cases}$$

Show that $B_t = H(t) \int_0^t H(s)^{-1} d\tilde{B}_s$. Use this fact to obtain a closed-form expression of $h(\gamma, \mu)$.

(vii) Use the result of Part (vi) to show that the conditional characteristic function of L given $B_T = z$ ($z \in \mathbb{R}^2$) is expressed as

$$\mathbb{E}[e^{i\lambda L} | B_T = z] = \frac{\lambda T}{2 \sinh(\lambda T/2)} \exp\left(\frac{|z|^2}{2T} \left(1 - \frac{\lambda T}{2} \coth \frac{\lambda T}{2}\right)\right), \quad \lambda \in \mathbb{R}.$$

Exercise 6.30. Let B be a d -dimensional Brownian motion starting at the origin. Let $S \triangleq \{x \in \mathbb{R}^d : |x| = 1\}$ denote the unit sphere in \mathbb{R}^d .

(i) Define $\tau \triangleq \inf\{t : B_t \in S\}$. Describe the distribution of B_τ . Show that B_τ and τ are independent.

(ii) Consider the process $X_t \triangleq B_t + ct$ where $c \in \mathbb{R}^d$ is a given fixed vector. Define $\sigma \triangleq \inf\{t : X_t \in S\}$. By using Girsanov's transformation theorem or otherwise, show that X_σ and σ are independent. How do you convince yourself about this property heuristically?

(iii) Show that the probability density function of X_σ with respect to the normalised uniform measure on S is given by

$$f_{X_\sigma}(x) = \mathbb{E}[e^{-|c|^2 \tau/2}] \cdot \exp(\langle c, x \rangle_{\mathbb{R}^d}), \quad x \in S.$$

Exercise 6.31. The aim of this problem is to explore how the ideas of stochastic calculus can be applied to prove theorems in complex analysis and algebra. A function $f : U \rightarrow \mathbb{C}$ defined on an open subset $U \subseteq \mathbb{C}$ is said to be *differentiable* at $z_0 \in U$ if there exists a complex number w_0 such that

$$\lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0} = w_0.$$

In this case, we denote w_0 as $f'(z_0)$. The function f is said to be *holomorphic* in U if it is differentiable at every point in U . Usual algebraic rules for real differentiation hold in the same way for complex variables. A basic property of holomorphic functions is the following so-called *Cauchy-Riemann* equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \quad \forall (x, y) \in U,$$

where we have expressed f in terms of its real and imaginary parts: $f(z) = u(x, y) + iv(x, y)$ ($z = x + iy$). It is also true that $f'(z) = \partial_x u + i\partial_x v$.

(i) [*Liouville's Theorem*] Suppose that f is a uniformly bounded, holomorphic function on \mathbb{C} . Let B be a two-dimensional Brownian motion. Explain why $\{\operatorname{Re}f(B_t)\}$ and $\{\operatorname{Im}f(B_t)\}$ are both bounded martingales. By using the martingale convergence theorem for $\{\operatorname{Re}f(B_t)\}$ or otherwise, conclude that f must be a constant function.

(ii) [*Fundamental Theorem of Algebra*] Use the result of Part (i) to show that every non-constant polynomial $f : \mathbb{C} \rightarrow \mathbb{C}$ has a root, i.e. $f(z) = 0$ for some $z \in \mathbb{C}$.

(iii) [*Maximum Principle for Harmonic Functions*] Let $u : U \rightarrow \mathbb{R}$ be a harmonic function on a bounded, connected open subset $U \subseteq \mathbb{C}$. Suppose that

$$\bar{B}(z_0, r) \triangleq \{z : |z - z_0| \leq r\} \subseteq U.$$

By considering the process $\{u(B_t^{z_0}) : t \geq 0\}$ (B^{z_0} is a Brownian motion starting at z_0), show that

$$u(z_0) = \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta}) d\theta.$$

Use this result to show that u can only attain maximum/minimum at the boundary of U unless u is a constant function.

Exercise 6.32. This problem is a continuation of Exercise 6.31. Let $B_t = X_t + iY_t$ be a two-dimensional Brownian motion. Let f be a non-constant holomorphic function on \mathbb{C} .

(i) Show that

$$f(B_t) = f(B_0) + \int_0^t f'(B_s) dB_s,$$

where f' is the complex derivative of f , the integral is the Itô integral but the product $f'(B_s)dB_s$ is understood as the complex multiplication.

(ii) It is known from complex analysis that there are at most countably many points at which $f' = 0$. Define

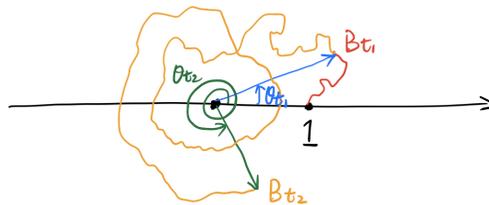
$$C_t \triangleq \int_0^t |f'(B_s)|^2 ds.$$

Show that with probability one, $t \mapsto C_t$ is strictly increasing and $C_\infty = \infty$.

(iii) By extending the method for Exercise 6.27 (ii), show that there exists a two-dimensional Brownian motion W such that

$$f(B_t) = f(B_0) + W_{C_t}.$$

Exercise 6.33. Let $B_t = X_t + iY_t$ be a two-dimensional Brownian motion starting at $z = 1$. Define $t \mapsto \theta_t \in \mathbb{R}$ to be the unique continuous determination of the argument of $B_t \in \mathbb{C}$ with $\theta_0 = 0$ (the total *winding number* of B around the origin up to time t)



In other words, $\{\theta_t\}$ is a process with continuous sample paths such that

$$B_t = R_t e^{i\theta_t}, \quad \theta_0 = 0,$$

where $R_t \triangleq |B_t|$. The aim of this problem is to establish the renowned *Spitzer's law*:

$$\frac{2\theta_t}{\log t} \xrightarrow{\text{dist.}} C \quad \text{as } t \rightarrow \infty \quad (6.3)$$

where C denotes the standard Cauchy distribution.

(i) Under complex multiplication, define

$$L_t \triangleq \int_0^t B_s^{-1} dB_s, \quad t \geq 0.$$

Explain why L_t is well-defined for all time. Show that there exists a two-dimensional Brownian motion (β, γ) starting at the origin, such that

$$L_t = \beta_{C_t} + i\gamma_{C_t},$$

where $C_t \triangleq \int_0^t R_s^{-2} ds$.

(ii) By using integration by parts for $B_t \cdot e^{-L_t}$, conclude that $B_t = e^{L_t}$.

(iii) Show that $\operatorname{Re} L_t = \log R_t$. By using the SDE of the Bessel process R or otherwise, show that $t \mapsto \int_0^t e^{2\beta_s} ds$ is the inverse function of $t \mapsto C_t$. Use this fact to conclude that the processes β and R generate the same σ -algebra. As a result, γ and R are independent.

(iv) For each $r > 1$, define $\sigma_r \triangleq \inf\{t : |B_t| = r\}$. By using the result of Exercise 6.13 (i), show that $\theta_{\sigma_r}/\log r$ is a standard Cauchy random variable.

(v) Let W be a two-dimensional Brownian motion starting at the origin, and set $\zeta \triangleq \inf\{t : |W_t| = 1\}$. Show that

$$\theta_t - \theta_{\sigma_{\sqrt{t}}} \xrightarrow{\text{dist.}} \operatorname{Im}\left(\int_{\zeta}^1 \frac{dW_s}{W_s}\right) \quad \text{as } t \rightarrow \infty,$$

where $\operatorname{Im}(\cdot)$ denotes the imaginary part of a complex number. As a result,

$$(\theta_t - \theta_{\sigma_{\sqrt{t}}})/\log t \rightarrow 0 \quad \text{in prob.}$$

as $t \rightarrow \infty$. Use this property and Part (iv) to conclude the result of (6.3). Compare this asymptotic phenomenon with Exercise 6.41 (ii).

Exercise 6.34. The aim of this problem is to obtain Itô's formula for $f(B_t)$ where B is a one-dimensional Brownian motion and $f(x) \triangleq |x|$. Note that f is not differentiable at $x = 0$.

(i) For each $\varepsilon > 0$, define

$$f_\varepsilon(x) \triangleq \begin{cases} |x|, & |x| \geq \varepsilon; \\ \frac{1}{2}(\varepsilon + x^2/\varepsilon), & |x| < \varepsilon. \end{cases}$$

Sketch the graph of f_ε and show that f_ε converges to f uniformly on \mathbb{R} as $\varepsilon \rightarrow 0$.

(ii) Show that

$$f_\varepsilon(B_t) = f_\varepsilon(B_0) + \int_0^t f'_\varepsilon(B_s) dB_s + \frac{1}{2\varepsilon} |\{s \in [0, t] : B_s \in (-\varepsilon, \varepsilon)\}|,$$

where $|\cdot|$ denotes the Lebesgue measure of a set.

(iii) For each $t \geq 0$, show that

$$\int_0^t f'_\varepsilon(B_s) \mathbf{1}_{(-\varepsilon, \varepsilon)}(B_s) dB_s \rightarrow 0$$

in the sense of L^2 .

(iv) Show that the limit

$$L_t \triangleq \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} |\{s \in [0, t] : B_s \in (-\varepsilon, \varepsilon)\}|$$

exists in L^2 and

$$|B_t| = |B_0| + \int_0^t \operatorname{sgn}(B_s) dB_s + L_t,$$

where

$$\operatorname{sgn}(x) \triangleq \begin{cases} -1, & x \leq 0; \\ 1, & x > 0. \end{cases}$$

(v) Show that the process $\{L_t : t \geq 0\}$ has non-decreasing sample paths, and its induced Lebesgue-Stieltjes measure is *carried by the zero level set* $\{t : B_t = 0\}$, i.e.

$$\int_0^\infty \mathbf{1}_{\{B_t \neq 0\}}(t) dL_t = 0.$$

Exercise 6.35. (i) Derive a general formula for the solution to the following linear SDE:

$$dX_t = (AX_t + a(t))dt + \sigma(t)dB_t.$$

Here A is a deterministic $n \times n$ matrix, the coefficients $a(t), \sigma(t)$ are bounded, deterministic functions taking values in $\operatorname{Mat}(n, 1)$ and $\operatorname{Mat}(n, d)$ respectively, and B is a d -dimensional Brownian motion.

(ii) Use the formula obtained in Part (i) to solve the equation of stochastic harmonic oscillator:

$$\begin{cases} dX_t = Y_t dt, \\ m dY_t = -kX_t dt - cY_t dt + \sigma dB_t. \end{cases}$$

where m, k, c, σ are given positive constants and B is a one-dimensional Brownian motion.

Exercise 6.36. Consider the following stochastic population growth model:

$$dX_t = X_t(K - X_t)dt + X_t dB_t, \quad X_0 = x > 0,$$

where $K > 0$ is a given constant.

- (i) Let $Y_t = \log X_t$. Establish the equation for Y_t .
- (ii) Solve the SDE for X_t explicitly. Discuss the possibility of reaching zero and/or exploding to infinity.

Exercise 6.37. (i) Consider the differential operator

$$\mathcal{A} = \frac{1}{2} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + b^i(x) \frac{\partial}{\partial x_i}$$

where $a : \mathbb{R}^n \rightarrow \text{Mat}(n, n)$ and $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$ are given bounded functions. Let X be an \mathbb{R}^n -valued process which satisfies the martingale formulation with generator \mathcal{A} , i.e. the process

$$M_t^f \triangleq f(X_t) - f(X_0) - \int_0^t (\mathcal{A}f)(X_s) ds$$

is a martingale for every given $f \in C_b^2(\mathbb{R}^n)$. What is the quadratic variation process of M^f ?

- (ii) Consider the case when $n = 1$:

$$\mathcal{A} = \frac{1}{2} a(x) \frac{d^2}{dx^2} + b(x) \frac{d}{dx},$$

where $a \in C_b^1$ is assumed to be a strictly positive function. Identify a suitable function $H : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$W_t \triangleq H(X_t) - H(X_0) - \int_0^t (\mathcal{A}H)(X_s) ds$$

is a Brownian motion and

$$X_t = X_0 + \int_0^t \sqrt{a(X_s)} dW_s + \int_0^t b(X_s) ds.$$

Conclude that a process X satisfies the martingale formulation with generator \mathcal{A} if and only if it is an Itô diffusion process with diffusion coefficient \sqrt{a} and drift coefficient b with respect to some Brownian motion.

Exercise 6.38. Let $B = \{B_t : 0 \leq t \leq 1\}$ be a one-dimensional Brownian motion.

(i) Define $X_t \triangleq B_t - tB_1$ ($0 \leq t \leq 1$). Show that X_t is a Gaussian process and compute its covariance function $\rho(s, t) \triangleq \mathbb{E}[X_s X_t]$.

(ii) Find the solution $\{Y_t : 0 \leq t < 1\}$ to the SDE

$$\begin{cases} dY_t = dB_t - \frac{Y_t}{1-t} dt, & 0 \leq t < 1, \\ Y_0 = 0. \end{cases}$$

Show that Y_t has the same law as X_t ($0 \leq t < 1$). Conclude that $\lim_{t \uparrow 1} Y_t = 0$ almost surely.

(iii) Given $x, y \in \mathbb{R}$, define the process

$$X^{x,y} = \{X_t^{x,y} \triangleq B_t^x + t(y - B_1^x) : 0 \leq t \leq 1\}$$

where $B^x = \{B_t^x : 0 \leq t \leq 1\}$ is a one-dimensional Brownian motion starting at x . Let μ_x (respectively, $\mu_{x,y}$) be the law of the process B^x (respectively, $X^{x,y}$) over the space \mathcal{W} of continuous paths on $[0, 1]$. Show that $\mu_{x,y}$ coincides with the conditional distribution of B^x given $B_1^x = y$, in the sense that

$$\mathbb{E}^{\mu_x}[\Phi | B_1^x] = \mathbb{E}^{\mu_{x,y}}[\Phi] \quad \mu_x\text{-a.s.}$$

for all bounded measurable functions $\Phi : \mathcal{W} \rightarrow \mathbb{R}$.

(iv) Use the result of Part (iii) to show that

$$\mathbb{P}\left(\sup_{0 \leq t \leq 1} X_t^{0,0} \geq x\right) = e^{-2x^2}, \quad x \geq 0.$$

Exercise 6.39. Consider the one-dimensional diffusion

$$\begin{cases} dX_t = 3X_t^2 dt - 2X_t^{3/2} dB_t, & t \geq 0 \\ X_0 = x > 0. \end{cases}$$

(i) Find the solution X_t explicitly. Conclude that X_t never reaches zero.

(ii) Let $e \triangleq \inf\{t : X_t = \infty\}$ be its explosion time. Show that $\mathbb{P}(e < \infty) = 1$ and $\mathbb{E}[e] = \infty$.

Exercise 6.40. Let X, Y be two Itô processes (possibly with respect to multidimensional Brownian motions). Define a new type of stochastic integration by

$$Z_t \triangleq \int_0^t X_s \circ dY_s \triangleq \lim_{\text{mesh}(\mathcal{P}) \rightarrow 0} \sum_{t_i \in \mathcal{P}} \frac{X_{t_{i-1}} + X_{t_i}}{2} \cdot (Y_{t_i} - Y_{t_{i-1}}),$$

where \mathcal{P} denotes an arbitrary partition of $[0, t]$.

(i) Show that Z is also an Itô process and identify its decomposition into the sum of an Itô integral and a Lebesgue integral.

(ii) For any smooth function f on \mathbb{R}^n , show that

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) \circ dX_s.$$

In particular, the integral $X \circ dY$ is consistent with the rules of ordinary calculus.

(iii) A *vector field* on \mathbb{R}^n is a function $V : \mathbb{R}^n \rightarrow \mathbb{R}^n$ that specifies a vector $V(x) = (V^1(x), \dots, V^n(x))$ at every point $x \in \mathbb{R}^n$. A vector field V can also be viewed as a differential operator on smooth functions by defining

$$(Vf)(x) \triangleq \sum_{i=1}^n V^i(x) \frac{\partial f(x)}{\partial x_i}, \quad f \in C^1(\mathbb{R}^n).$$

Let $B = \{(B_t^1, \dots, B_t^d) : t \geq 0\}$ be a d -dimensional Brownian motion and let V_1, \dots, V_d be d vector fields on \mathbb{R}^n with bounded derivatives of all orders. Show that there exists a unique \mathbb{R}^n -valued progressively measurable process $X = \{X_t\}$ that satisfies the following equation

$$X_t^i = x_i + \sum_{\alpha=1}^d V_\alpha^i(X_t) \circ dB_t^\alpha, \quad i = 1, \dots, n$$

with given initial condition $X_0 = x \in \mathbb{R}^n$. In addition, show that the generator of X is given by

$$\mathcal{A}f = \frac{1}{2} \sum_{\alpha=1}^d V_\alpha(V_\alpha f)$$

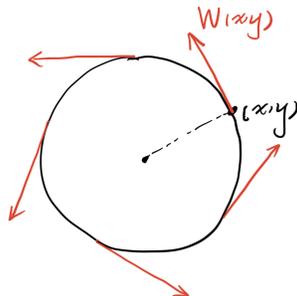
and

$$f(X_t) = f(x) + \sum_{\alpha=1}^d \int_0^t (V_\alpha f)(X_s) \circ dB_s^\alpha$$

for any suitably smooth function f .

Exercise 6.41. Let B be a one-dimensional Brownian motion.

(i) Consider the vector field $W(x, y) = (-y, x)$ on \mathbb{R}^2 .



Let $X_t \in \mathbb{R}^2$ be the solution to the SDE

$$dX_t = W(X_t) \circ dB_t$$

with initial condition $X_0 = (1, 0)$ defined in the sense of Exercise 6.40 (iii) ($n = 2$, $d = 1$). Show that X_t stays on the unit circle, i.e. $|X_t| = 1$ for all time. What if the integral “ \circ ” is replaced by the Itô integral?

(ii) Let $N_t \in \mathbb{Z}$ ($t \geq 0$) denote the total number of loops that X_t has winded around the origin (an anti-clockwise loop is counted as $+1$ and a clockwise loop is counted as -1). Show that N_t/\sqrt{t} converges to $N(0, \frac{1}{4\pi^2})$ in distribution as $t \rightarrow \infty$.

(iii) Given $a, b > 0$, let $\{(X_t, Y_t) : t \geq 0\}$ be the unique solution to the SDE

$$\begin{cases} dX_t = -\frac{1}{2}X_t dt - \frac{a}{b}Y_t dB_t, \\ dY_t = -\frac{1}{2}Y_t dt + \frac{b}{a}X_t dB_t \end{cases}$$

with initial condition $(X_0, Y_0) = (a, 0)$. Show that (X_t, Y_t) stays on the standard ellipse $\{(x, y) : x^2/a^2 + y^2/b^2 = 1\}$ for all time. Find the expected amount of time that the process takes to complete one exact loop.

Exercise 6.42. Define for $r > 0$ the sphere $S_r \triangleq \{(x, y, z) : x^2 + y^2 + z^2 = r^2\}$. Let $\{e_1, e_2, e_3\}$ be the standard orthonormal basis of \mathbb{R}^3 . For $\alpha = 1, 2, 3$, define the vector field W_α on \mathbb{R}^3 by setting $W_\alpha(\xi)$ ($\xi \in S_r$) to be the orthogonal projection

of $r \cdot e_\alpha$ onto the tangent plane of S_r at ξ . Let $X = \{X_t^\xi : \xi \in S_r, t \geq 0\}$ be the solution to the SDE

$$\begin{cases} dX_t^\xi = \sum_{\alpha=1}^3 W_\alpha(X_t^\xi) \circ dB_t^\alpha, & t \geq 0; \\ X_0^\xi = \xi \in S_r \end{cases} \quad (6.4)$$

defined in the sense of Exercise 6.40 (iii) ($n = d = 3$).

(i) Show that X^ξ lives on S_r for all time.

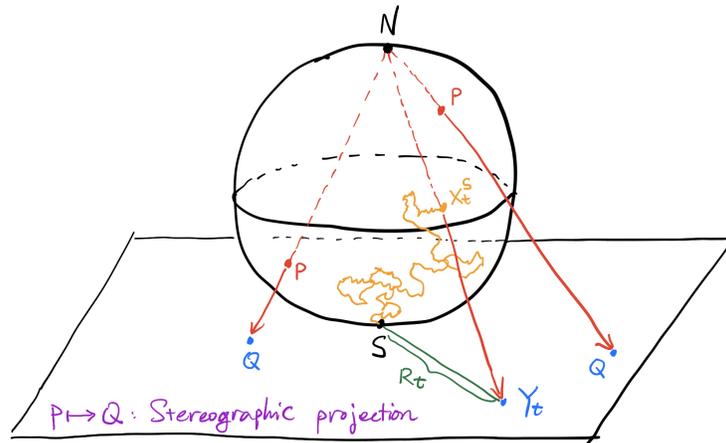
(ii) The *spherical Laplacian* Δ_{S_1} on S_1 is defined by

$$(\Delta_{S_1} f)(\xi) \triangleq (\Delta \hat{f})(\xi), \quad f \in C^2(S_1),$$

where $\hat{f} : \mathbb{R}^3 \setminus \{0\} \rightarrow \mathbb{R}$ is the extension of f defined by $\hat{f}(\eta) \triangleq f(\eta/|\eta|)$ and Δ is the usual Laplace operator on \mathbb{R}^3 . Show that the generator of $\{X^\xi : \xi \in S_1\}$ is $\frac{1}{2}\Delta_{S_1}$ in the sense that

$$\lim_{t \rightarrow 0} \frac{\mathbb{E}[f(X_t^\xi)] - f(\xi)}{t} = \frac{1}{2}(\Delta_{S_1} f)(\xi) \quad \forall f \in C^2(S_1).$$

(iii) Let X_t^S be the unique solution to the SDE (6.4) starting at the south pole of S_1 . Consider the *stereographic projection* defined in the figure below which maps $S_1 \setminus \{\text{north pole}\}$ bijectively onto the plane \mathbb{R}^2 .



Let Y_t denote the projected point of X_t^S on the plane and let R_t be the distance between Y_t and the south pole. Establish a one-dimensional SDE for the process R .

(iv) What is the probability that the process X_t^S hits the north pole in finite time?

Exercise 6.43. Consider the following one-dimensional SDE

$$\begin{cases} dX_t^x = \sigma(X_t^x)dB_t + b(X_t^x)dt, & t \geq 0; \\ X_0^x = x \in \mathbb{R}, \end{cases} \quad (6.5)$$

where the coefficients $\sigma, b : \mathbb{R} \rightarrow \mathbb{R}$ satisfy the Lipschitz condition. Define the function $\varphi(x, t)$ by the differential equation

$$\frac{\partial \varphi(x, t)}{\partial t} = \sigma(\varphi(x, t)), \quad \varphi(x, 0) = x.$$

(i) Let $w : [0, \infty) \rightarrow \mathbb{R}$ be a smooth path with $w_0 = 0$. Define the path $t \mapsto \xi_t^x$ by the differential equation

$$\frac{d\xi_t^x}{dt} = b(\varphi(\xi_t^x, w_t)) \exp\left(-\int_0^{w_t} \sigma'(\varphi(\xi_t^x, s))ds\right), \quad \xi_0^x = x. \quad (6.6)$$

Show that the function $t \mapsto \varphi(\xi_t^x, w_t)$ is the solution to the equation

$$dx_t = (\sigma(x_t)w_t' + b(x_t))dt, \quad x_0 = x.$$

(ii) How do you adapt the method of Part (i) to construct the solution to the SDE (6.5)?

(iii) Suppose that $b - \frac{1}{2}\sigma' \cdot \sigma = 0$ and $\sigma > 0$ on \mathbb{R} . Show that for each fixed $t > 0$, the random variable X_t^x has a probability density function. Obtain a formula for this density.

Exercise 6.44. Consider the following SDE:

$$\begin{cases} dX_t = dB_t + b(X_t)dt, & t \geq 0; \\ X_0 = 0, \end{cases}$$

where B is a one-dimensional Brownian motion and $b \in C_b^1(\mathbb{R})$.

(i) Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a given function. Derive a formula for computing $\mathbb{E}[f(X_t)]$ in terms of a Brownian motion.

(ii) Suppose that $\int_{\mathbb{R}} |b(x)|dx < \infty$. Show that with probability one,

$$\overline{\lim}_{t \rightarrow \infty} X_t = \infty, \quad \underline{\lim}_{t \rightarrow \infty} X_t = -\infty.$$

(iii) Let $\alpha < 0 < \beta$. Define $\tau \triangleq \inf\{t : X_t \notin (\alpha, \beta)\}$. Compute the expected value of τ and the probability that X_t exits the interval (α, β) from the left end point α .

(iv) Let $c > 0$ be a given parameter. In each of the following scenarios, discuss the limiting behaviour of X_t as $t \rightarrow \infty$:

(iv-a) $b(x) = 0$ when $x \leq 0$ and $b(x) = c/x$ when $x \geq 1$;

(iv-b) $b(x) = c/x$ when $|x| \geq 1$.

Exercise 6.45. Consider the following interest rate model:

$$\begin{cases} dR_t = (1 - R_t)dt + \sqrt{R_t}dB_t, & t \geq 0, \\ R_0 = r > 0. \end{cases}$$

(i) Define $X_t \triangleq R_t e^t$. Find the SDE for X_t .

(ii) Show that there exists a deterministic time change $t \mapsto c(t)$ (i.e. a continuous, increasing function with $c(0) = 0$) such that $\rho = \{\rho_t \triangleq X_{c(t)} : t \geq 0\}$ is a squared Bessel process. Conclude that $R_t = e^{-t} \rho_{(e^t - 1)/4}$.

(iii) Let S_t be the solution to the following SDE:

$$\begin{cases} dS_t = S_t dt + S_t^{3/2} dB_t, & t \geq 0, \\ S_0 = x > 0. \end{cases}$$

By investigating the relationship between S_t and R_t , represent S_t in terms of a Bessel process. Conclude that $\{S_t\}$ never reaches zero nor explodes to infinity in finite time.

Exercise 6.46. The aim of this problem is to study the evolution of random ellipses in the plane (the behaviour of Brownian motion taking values in the space of ellipses with unit area). In this problem, we use matrix notation exclusively.

In planar Euclidean geometry, it is known that there is a one-to-one correspondence between the space \mathcal{E} of ellipses centered at the origin with unit area and the space \mathcal{S} of 2×2 positive definite, symmetric matrices with determinant one, which is given by

$$\text{matrix } Y \in \mathcal{S} \longleftrightarrow \text{ellipse } E^Y = \{z \in \mathbb{R}^2 : z^T Y z = 1\} \in \mathcal{E}.$$

Under this correspondence, the major (respectively, minor) semi-axis of the ellipse E^Y is equal to the larger (respectively, smaller) eigenvalue of Y . If one orthogonally diagonalises Y by writing

$$Y = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix} \begin{pmatrix} \cos \gamma & \sin \gamma \\ -\sin \gamma & \cos \gamma \end{pmatrix} \quad (6.7)$$

with $\lambda > 1$ being the larger eigenvalue of Y , then the vector $(\cos \gamma, \sin \gamma)$ represents the direction of the minor semi-axis.

(i) Let $B_t \triangleq \begin{pmatrix} W_t^1/\sqrt{2} & W_t^2 \\ W_t^3 & -W_t^1/\sqrt{2} \end{pmatrix}$ where (W_t^1, W_t^2, W_t^3) is a three-dimensional Brownian motion. Consider the $\text{Mat}(2, 2)$ -valued SDE

$$dG_t = dB_t \cdot G_t + \frac{1}{4}G_t dt$$

with initial condition given by a fixed non-identity matrix. Show that with probability one, G_t has determinant one for all t .

(ii) Define $Y_t \triangleq G_t^T G_t$. Show that

$$dY_t = G_t^T (dB_t + dB_t^T) G_t + 2Y_t dt.$$

(iii) Write the eigenvalues of Y_t as $e^{\pm U_t}$ ($U_t \geq 0$). By using the relation $\text{Tr}(Y_t) = e^{U_t} + e^{-U_t}$, show that U_t satisfies a one-dimensional SDE

$$dU_t = \sqrt{2}d\beta_t + \coth U_t dt,$$

where $\{\beta_t\}$ is a suitable one-dimensional Brownian motion. Write down the generator \mathcal{A} of $\{U_t\}$.

(iv) Show that with probability one, U_t never reaches zero and U_t/t converges to 1 as $t \rightarrow \infty$. Interpret this property geometrically in the context of ellipses.

(v) Let γ_t be the continuous determination of the angle appearing in the orthogonal diagonalisation (6.7) of the matrix Y_t . By extending the method for Exercise 6.27 (ii), show that there exists a one-dimensional Brownian motion W that is independent of $\{U_t\}$, such that

$$\gamma_t = W\left(\int_0^t \frac{1}{2} \text{csch}^2 U_s ds\right).$$

Use this fact and Part (iv) to deduce that $\gamma_\infty \triangleq \lim_{t \rightarrow \infty} \gamma_t$ exists a.s. What is the distribution of γ_∞ ? Interpret γ_t and γ_∞ geometrically in the context of ellipses.

(vi) Let $(x_0, y_0) \in \mathbb{R}^2$ be a fixed vector. Define $(x_t, y_t)^T \triangleq G_t \cdot (x_0, y_0)^T$ and set $R_t \triangleq \sqrt{x_t^2 + y_t^2}$. Show that R_t has a log-normal distribution and use this to derive the p -th moment Lyapunov exponent of R_t as

$$\lim_{t \rightarrow \infty} t^{-1} \log \mathbb{E}[R_t^p] = \frac{p(p+2)}{4}, \quad \forall p > 0.$$

Exercise 6.47. Let $a : [0, \infty) \rightarrow S_n^+$ and $b : [0, \infty) \rightarrow \mathbb{R}^n$ be given bounded, continuous functions, where S_n^+ denotes the space of $n \times n$ symmetric, positive definite matrices. For each $t \geq 0$, define the second order differential operator \mathcal{A}_t by

$$\mathcal{A}_t f \triangleq \frac{1}{2} \sum_{i,j=1}^n a_{ij}(t) \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^n b^i(t) \frac{\partial f}{\partial x_i}.$$

(i) Construct a stochastic process $X = \{X_t : t \geq 0\}$, such that

$$M_t^f \triangleq f(X_t) - f(X_0) - \int_0^t (\mathcal{A}_s f)(X_s) du$$

is a martingale for any function $f \in C_b^2(\mathbb{R}^n)$. What is the distribution of X_t ?

(ii) Let $\varphi : \mathbb{R}^n \rightarrow \infty$ be a given smooth function with compact support. Find a stochastic representation of the solution to the following Cauchy problem:

$$\begin{cases} \frac{\partial u}{\partial t} - \mathcal{A}_t u = 0, & (t, x) \in (0, \infty) \times \mathbb{R}^n; \\ u(0, x) = \varphi(x), & x \in \mathbb{R}^n. \end{cases}$$

Exercise 6.48. Let (a, b) be a finite open interval on \mathbb{R} . Let $K : [a, b] \rightarrow [0, \infty)$ be a given continuous function.

(i) Find a stochastic representation to the following initial-boundary value problem (heat equation with cooling coefficient K):

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} - K(x)u, & (t, x) \in (0, \infty) \times (a, b), \\ u(t, a) = 0, \quad u(t, b) = 0, & t \geq 0, \\ u(0, x) = f(x), & a \leq x \leq b, \end{cases}$$

where f is a given continuous function with compact support inside (a, b) .

(ii) Let $\{B_t^x : t \geq 0\}$ be a one-dimensional Brownian motion starting at $x \in (a, b)$.

Define $\tau^x \triangleq \inf\{t : B_t^x \notin (a, b)\}$. By using the result of Part (i) for $K \equiv 0$ and solving the corresponding initial-boundary value problem analytically, show that

$$\mathbb{P}(B_t^x \in dy, t < \tau^x) = \sum_{n=0}^{\infty} \frac{2}{b-a} e^{-\frac{n^2\pi^2 t}{2(b-a)^2}} \sin\left(\frac{n\pi(x-a)}{b-a}\right) \sin\left(\frac{n\pi(y-a)}{b-a}\right) dy$$

for $t > 0$, $y \in (a, b)$.

Exercise 6.49. Let $D \triangleq \{x \in \mathbb{R}^d : |x| < 1\}$ be the unit open ball on \mathbb{R}^d . Denote $L^2(D)$ as the Hilbert space of square integrable functions on D with respect to the Lebesgue measure. It is a classical result in PDE theory (*spectral decomposition of the Dirichlet Laplacian*) that there exist a sequence of real numbers

$$0 < \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq \cdots \uparrow \infty$$

and an orthonormal basis $\{\phi_n : n \geq 0\} \subseteq C_b^\infty(D)$ of $L^2(D)$, such that

$$\begin{cases} -\Delta\phi_n = \lambda_n\phi_n & \text{in } D, \\ \phi_n = 0 & \text{on } \partial D \end{cases}$$

for each n . In addition, $\phi_0 > 0$ everywhere in D , and there exists $N \geq 1$ such that the series

$$\sum_{n=0}^{\infty} \frac{1}{\lambda_n^N} \phi_n(x)\phi_n(y)$$

converges absolutely and uniformly on $\bar{D} \times \bar{D}$.

(i) For each $f \in C(\bar{D})$ vanishing at the boundary ∂D , verify that the function

$$u(t, x) \triangleq \sum_{n=0}^{\infty} e^{-\frac{1}{2}\lambda_n t} \langle f, \phi_n \rangle_{L^2(D)} \phi_n(x)$$

is the solution to the initial-boundary value problem

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{2}\Delta u = 0, & (t, x) \in (0, \infty) \times D, \\ u(0, x) = f(x), & x \in D, \\ u(t, x) = 0, & (t, x) \in [0, \infty) \times \partial D. \end{cases}$$

(ii) Define

$$\tau_x \triangleq \inf\{t \geq 0 : |B_t^x| \geq 1\}$$

where B^x denotes a Brownian motion starting at $x \in \mathbb{R}^d$. Under the same assumption as in Part (i), show that

$$u(t, x) = \mathbb{E}[f(B_t^x); \tau_x > t], \quad (t, x) \in [0, \infty) \times \bar{D}.$$

(iii) Show that there exists $c > 0$ such that $\mathbb{E}[e^{c\tau_x}] < \infty$ for all $x \in D$.

(iv) Show that

$$\mathbb{P}\left(\sup_{0 \leq t \leq 1} |B_t^0| < \varepsilon\right) \sim C e^{-\lambda_0/2\varepsilon^2} \quad \text{as } \varepsilon \rightarrow 0,$$

where $C \triangleq \phi_0(0) \int_D \phi_0(x) dx$ and the notation $\varphi(\varepsilon) \sim \psi(\varepsilon)$ means $\lim_{\varepsilon \rightarrow 0} \frac{\varphi(\varepsilon)}{\psi(\varepsilon)} = 1$.

Exercise 6.50. Consider the following one-dimensional SDE

$$\begin{cases} dX_t = b(X_t)dt + dB_t, & 0 \leq t \leq 1; \\ X_0 = 0, \end{cases}$$

where $b : \mathbb{R} \rightarrow \mathbb{R}$ satisfies the Lipschitz condition. Let $\phi : [0, 1] \rightarrow \mathbb{R}$ be a twice continuously differentiable function with $\phi_0 = 0$. The aim of this problem is to study the asymptotic behaviour of the probability

$$C_\phi(\varepsilon) \triangleq \mathbb{P}(\|X - \phi\|_\infty < \varepsilon)$$

and identify those paths ϕ for which $C_\phi(\varepsilon)$ is maximised in the asymptotics $\varepsilon \rightarrow 0$.

(i) Define $\psi_t \triangleq \phi_t - \int_0^t b(\phi_s) ds$. Construct a probability measure \mathbb{Q} under which the process $\tilde{B}_t \triangleq B_t - \psi_t$ is a Brownian motion. Show that

$$\begin{aligned} C_\phi(\varepsilon) &= \exp\left(-\frac{1}{2} \int_0^1 \dot{\psi}_t^2 dt\right) \\ &\cdot \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^1 \dot{\psi}_t dB_t + \int_0^1 \dot{\psi}_t^2 dt\right) \mathbf{1}_{\{\|X - \phi\|_\infty < \varepsilon\}}\right]. \end{aligned}$$

(ii) By using integration by parts, show that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^1 \dot{\psi}_t dB_t + \int_0^1 \dot{\psi}_t^2 dt\right) \mid \|X - \phi\|_\infty < \varepsilon\right] = 1.$$

Hence conclude that

$$C_\phi(\varepsilon) \sim \exp\left(-\frac{1}{2} \int_0^1 \dot{\psi}_t^2 dt\right) \cdot \mathbb{Q}(\|X - \phi\|_\infty < \varepsilon) \quad \text{as } \varepsilon \rightarrow 0.$$

(iii) Define $\tilde{b}(t, x) \triangleq b(x + \phi_t) - b(\phi_t)$ and $Y_t \triangleq X_t - \phi_t$. Show that Y_t satisfies the following SDE:

$$dY_t = d\tilde{B}_t + \tilde{b}(t, Y_t)dt.$$

(iv) Construct another probability measure $\tilde{\mathbb{Q}}$ under which Y_t is a Brownian motion. Conclude that

$$\mathbb{Q}(\|X - \phi\|_\infty < \varepsilon) \sim \mathbb{E}^\mu \left[\exp \left(\int_0^1 \tilde{b}(t, \beta_t) d\beta_t \right) \mathbf{1}_{\{\|\beta\|_\infty < \varepsilon\}} \right] \quad \text{as } \varepsilon \rightarrow 0,$$

where β is the canonical Brownian motion and μ is the Wiener measure.

(v) By using integration by parts, show that

$$\begin{aligned} & \mathbb{E}^\mu \left[\exp \left(\int_0^1 \tilde{b}(t, \beta_t) d\beta_t \right) \mid \|\beta\|_\infty < \varepsilon \right] \\ & \sim \exp \left(- \int_0^1 b'(\phi_t) dt \right) \cdot \mathbb{E}^\mu \left[\exp \left(- \int_0^1 \beta_t b'(\beta_t + \phi_t) d\beta_t \right) \mid \|\beta\|_\infty < \varepsilon \right] \quad \text{as } \varepsilon \rightarrow 0. \end{aligned}$$

(vi) Show that

$$\mathbb{E}^\mu \left[\exp \left(- \int_0^1 \beta_t b'(\beta_t + \phi_t) d\beta_t \right) \mid \|\beta\|_\infty < \varepsilon \right] \sim \exp \left(\frac{1}{2} \int_0^1 b'(\phi_t) dt \right).$$

Hence conclude that

$$\mathbb{Q}(\|X - \phi\|_\infty < \varepsilon) \sim \exp \left(- \frac{1}{2} \int_0^1 b'(\phi_t) dt \right) \cdot \mathbb{P}^\mu(\|\beta\|_\infty < \varepsilon) \quad \text{as } \varepsilon \rightarrow 0.$$

(vii) Combine the previous steps as well as the result of Exercise 6.49 (iv) to conclude that

$$C_\phi(\varepsilon) \sim \frac{4}{\pi} \exp \left(- \frac{1}{2} \int_0^1 ((\dot{\phi}_t - b(\phi_t))^2 + b'(\phi_t)) dt \right) \cdot e^{-\frac{\pi^2}{8\varepsilon^2}} \quad \text{as } \varepsilon \rightarrow 0.$$

(viii) Fix $y \in \mathbb{R}$. Denote $\mathcal{I}_{0,y}$ as the space of twice continuously differentiable functions ϕ such that $\phi_0 = 0$ and $\phi_1 = y$. Suppose that $\phi^y \in \mathcal{I}_{0,y}$ minimises the function

$$J(\phi) \triangleq \int_0^1 ((\dot{\phi}_t - b(\phi_t))^2 + b'(\phi_t)) dt, \quad \phi \in \mathcal{I}_{0,y}.$$

Show that ϕ^y satisfies the following ODE

$$\ddot{\phi}_t = b(\phi_t)b'(\phi_t) + \frac{1}{2}b''(\phi_t).$$

(ix) Identify the minimiser $\phi^y \in \mathcal{I}_{0,y}$ of $J(\phi)$ in the cases when $b = 0$ and $b(x) = \alpha x$ ($\alpha \neq 0$). Heuristically, ϕ^y maximises the “probability”

$$\mathcal{I}_{0,y} \ni \phi \mapsto \lim_{\varepsilon \rightarrow 0} C_\phi(\varepsilon) e^{\frac{\pi^2}{8\varepsilon}}$$

among all paths in $\mathcal{I}_{0,y}$. In other words, the paths ϕ^y ($y \in \mathbb{R}$) are the “mostly preferred paths” for the diffusion $\{X_t\}$.

Appendix A Terminology from probability theory

We collect a few notions and tools from probability theory that are quoted in these notes.

(1) A *sample space* is a non-empty subset Ω . A σ -*algebra* is a class \mathcal{F} of subsets such that

$$\Omega \in \mathcal{F}; A \in \mathcal{F} \implies A^c \in \mathcal{F}; A_n \in \mathcal{F} \forall n \implies \cup_n A_n \in \mathcal{F}.$$

A *probability measure* on \mathcal{F} is a set function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ such that:

- (i) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$;
- (ii) $\mathbb{P}(\Omega) = 1$;
- (iii) for any sequence $\{A_n : n \geq 1\} \subseteq \mathcal{F}$ of disjoint events (i.e. $A_m \cap A_n = \emptyset$ when $m \neq n$), one has

$$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Such a triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

(2) The *Borel σ -algebra* over \mathbb{R}^n , denoted as $\mathcal{B}(\mathbb{R}^n)$, is the smallest σ -algebra (equivalently, the intersection of all those σ -algebras) containing the following class of subsets:

$$(a, b] \triangleq \{x = (x_1, \dots, x_n) : a_i < x_i \leq b_i \forall i\}, \quad a, b \in \mathbb{R}^n.$$

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be *Borel measurable* if

$$f^{-1}A \triangleq \{x \in \mathbb{R}^n : f(x) \in A\} \in \mathcal{B}(\mathbb{R}^n) \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

(3) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given fixed probability space. An \mathbb{R}^n -*valued random variable* is a function $X : \Omega \rightarrow \mathbb{R}^n$ such that $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$ for all $x \in \mathbb{R}^n$. Let $\{X_t : t \in \mathcal{T}\}$ be a given family of random variables. The σ -*algebra generated by this family*, denoted as $\sigma(X_t : t \in \mathcal{T})$, is the smallest σ -algebra containing the following class of subsets:

$$\{\omega : X_t(\omega) \in A\}, \quad t \in \mathcal{T}, A \in \mathcal{B}(\mathbb{R}).$$

More generally, the notation $\sigma(\dots)$ denotes the σ -algebra generated by (i.e. the smallest σ -algebra containing) whatever is listed inside the bracket.

(4) A subset $N \subseteq \Omega$ is called a \mathbb{P} -null set if there exists $E \in \mathcal{F}$ such that $\mathbb{P}(E) = 0$ and $N \subseteq E$. A property \mathbf{P} is said to hold *almost surely (a.s.)* or *with probability one* if the set of ω for which \mathbf{P} does not hold is a \mathbb{P} -null set. For instance, a random series $\sum_{n=1}^{\infty} X_n$ is convergent a.s. if the set

$$\left\{ \omega \in \Omega : \sum_{n=1}^{\infty} X(\omega) \text{ is not convergent} \right\}$$

is a \mathbb{P} -null set.

(5) Let $\{A_n : n \geq 1\} \subseteq \mathcal{F}$ be a sequence of events. We define

$$\overline{\lim}_{n \rightarrow \infty} A_n \triangleq \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{ \omega : \text{there are infinitely many } n \text{ s.t. } \omega \in A_n \}$$

to be the event that “ A_n occurs infinitely often”. Respectively, we define

$$\underline{\lim}_{n \rightarrow \infty} A_n \triangleq \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \{ \omega : \omega \in A_n \text{ for all sufficiently large } n \}$$

to be the event that “ A_n happens eventually”. By definition,

$$\left(\overline{\lim}_{n \rightarrow \infty} A_n \right)^c = \underline{\lim}_{n \rightarrow \infty} A_n^c$$

is the event that “there are at most finitely many A_n ’s happening” or equivalently “ A_n does not happen for all sufficiently large n ”. It is often the case that these events are either \mathbb{P} -null sets or have probability one. One particular situation, which is the content of the *(first) Borel-Catelli lemma*, is used in the construction of stochastic integrals.

Theorem A.1. (i) [The first Borel-Catelli lemma] Suppose that

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty.$$

Then

$$\mathbb{P}\left(\overline{\lim}_{n \rightarrow \infty} A_n \right) = 0.$$

In other words, with probability one there are at most finitely many A_n ’s happening.

(ii) [The second Borel-Cantelli lemma] Suppose that $\{A_n : n \geq 1\}$ are independent and

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty.$$

Then

$$\mathbb{P}\left(\overline{\lim}_{n \rightarrow \infty} A_n\right) = 1.$$

(6) The *mathematical expectation* is an integration map $X \mapsto \mathbb{E}[X]$ defined as follows. If X is a simple random variable, say $X = \sum_{i=1}^m c_i \mathbf{1}_{A_i}$ then

$$\mathbb{E}[X] \triangleq \sum_{i=1}^m c_i \mathbb{P}(A_i).$$

For a general non-negative random variable X , one approximates it by an increasing sequence $\{X_n : n \geq 1\}$ of simple functions and define

$$\mathbb{E}[X] \triangleq \lim_{n \rightarrow \infty} \mathbb{E}[X_n].$$

If X has arbitrary sign, its expectation is defined as

$$\mathbb{E}[X] \triangleq \mathbb{E}[X^+] - \mathbb{E}[X^-],$$

where $X^+ \triangleq \max\{X, 0\}$ and $X^- \triangleq \max\{-X, 0\}$ are both non-negative and $X = X^+ - X^-$. Sometimes $\mathbb{E}[X]$ is also denoted as $\int_{\Omega} X d\mathbb{P}$. We say that X is *integrable* if $\mathbb{E}[X]$ is finite.

(7) Let $p, q > 1$ be such that $1/p + 1/q = 1$. *Hölder's inequality* asserts that

$$|\mathbb{E}[XY]| \leq \|X\|_p \|Y\|_q \tag{A.1}$$

for any random variables $X \in L^p, Y \in L^q$, where $\|X\|_p \triangleq (\mathbb{E}[|X|^p])^{1/p}$ and we say $X \in L^p$ if $\|X\|_p < \infty$. By taking $p = q = 2$ and $Y = 1$, (A.1) becomes

$$|\mathbb{E}[X]|^2 \leq \mathbb{E}[X^2]. \tag{A.2}$$

We refer to (A.2) as the *Cauchy-Schwarz inequality*. It is a simple consequence of Hölder's inequality that $[1, \infty) \ni p \mapsto \|X\|_p$ is non-decreasing.

(8) Let \mathbb{P}, \mathbb{Q} be two probability measures on \mathcal{F} . We say that \mathbb{Q} is absolutely continuous with respect to \mathbb{P} if

$$A \in \mathcal{F}, \mathbb{P}(A) = 0 \implies \mathbb{Q}(A) = 0.$$

In this case, the *Radon-Nikodym theorem* asserts that there exists a non-negative \mathbb{P} -integrable function $X : \Omega \rightarrow \mathbb{R}$, denoted as $\frac{d\mathbb{Q}}{d\mathbb{P}}$ (the *density function* of \mathbb{Q} with respect to \mathbb{P}), such that

$$\mathbb{Q}(A) = \int_A X d\mathbb{P} \quad \forall A \in \mathcal{F}.$$

(9) A random vector (X_1, \dots, X_n) is *jointly Gaussian* with mean vector μ and covariance matrix Σ if its joint density function is given by

$$f_{X_1, \dots, X_n}(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^T \cdot \Sigma^{-1} \cdot (x - \mu)\right), \quad x \in \mathbb{R}^n.$$

This is equivalent to saying that

$$\mathbb{E}[e^{i(\theta_1 X_1 + \dots + \theta_n X_n)}] = \exp\left(\theta^T \cdot \mu - \frac{1}{2} \theta^T \cdot \Sigma \cdot \theta\right) \quad \forall \theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^n.$$

(10) Let X, Y be random variables. We say that X, Y are *independent* if any of the following three equivalent statements hold true:

- (i) $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$ for any $x, y \in \mathbb{R}$;
- (ii) $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ for any bounded, Borel measurable functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$;
- (iii) $\mathbb{E}[e^{i(sX+tY)}] = \mathbb{E}[e^{isX}]\mathbb{E}[e^{itY}]$ for any $s, t \in \mathbb{R}$.

Two sub- σ -algebras $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ are said to be independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \quad \forall A \in \mathcal{G}, B \in \mathcal{H}.$$

One can also talk about the independence between X and a sub- σ -algebra \mathcal{G} :

$$\mathbb{P}(\{X \in A\} \cap B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(B) \quad \forall A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{G}.$$

More generally, a family of random variables $\{X_t : t \in \mathcal{T}\}$ and \mathcal{G} are independent if

$$\mathbb{P}(\{(X_{t_1}, \dots, X_{t_n}) \in \Gamma\} \cap B) = \mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in \Gamma) \cdot \mathbb{P}(B)$$

for any choices of $n \geq 1$, $t_1, \dots, t_n \in \mathcal{T}$, $\Gamma \in \mathcal{B}(\mathbb{R}^n)$ and $B \in \mathcal{G}$. An equivalent characterisation, which is used in the proofs of the strong Markov property for Brownian motion and Lévy's characterisation theorem, is the following statement in terms of joint characteristic functions:

$$\mathbb{E}[\xi \cdot e^{i(\theta_1 X_{t_1} + \dots + \theta_n X_{t_n})}] = \mathbb{E}[\xi] \mathbb{E}[e^{i(\theta_1 X_{t_1} + \dots + \theta_n X_{t_n})}]$$

for any choices of n , $t_i \in \mathcal{T}$, $\theta_i \in \mathbb{R}$ and bounded, \mathcal{G} -measurable ξ .

(11) The notion of product spaces is useful for constructing independent random variables. Let $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ ($n \geq 1$) be given probability spaces. Define

$$\Omega^\infty \triangleq \prod_{n=1}^{\infty} \Omega_n = \{\omega = (\omega_1, \omega_2, \dots) : \omega_n \in \Omega_n \forall n\}.$$

Let \mathcal{F}^∞ be the smallest σ -algebra over Ω containing the following subsets:

$$A_1 \times A_2 \times \dots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \dots \quad (\text{A.3})$$

for all choices of $n \geq 1$, $A_i \in \Omega_i$ ($1 \leq i \leq n$). Then there exists a unique probability measure \mathbb{P}^∞ on \mathcal{F}^∞ , such that the probability of the event (A.3) is equal to

$$\mathbb{P}_1(A_1) \mathbb{P}_2(A_2) \dots \mathbb{P}_n(A_n).$$

The triple $(\Omega^\infty, \mathcal{F}^\infty, \mathbb{P}^\infty)$ is known as the *product probability space* of the given sequence.

One can easily construct independent sequences by using the idea of product spaces. For instance, take $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ for every n where μ is the standard Gaussian measure on \mathbb{R} . On the resulting product space $(\Omega^\infty, \mathcal{F}^\infty, \mathbb{P}^\infty)$, define

$$X_n : \Omega^\infty \rightarrow \mathbb{R}, \quad X_n(\omega) \triangleq \omega_n.$$

Then $\{X_n : n \geq 1\}$ is an i.i.d. sequence of standard normal random variables.

(12) An important question one often encounters is the interchange of the limit sign and the expectation. This is the content of the following result which contains three basic principles of such kind.

Theorem A.2. *Let X_n ($n \geq 1$) and X, Y be random variables.*

(i) [Fatou's Lemma] *Suppose that $X_n \geq 0$ a.s. Then*

$$\mathbb{E}\left[\liminf_{n \rightarrow \infty} X_n\right] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n].$$

(ii) [Monotone Convergence Theorem] Suppose that

$$X_n \geq 0, X_n \uparrow X \text{ a.s.}$$

Then

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

(iii) [Dominated Convergence Theorem] Suppose that

$$X_n \rightarrow X, |X_n| \leq Y \text{ a.s.}$$

and Y is integrable. Then

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

References

- [1] T.M. Apostol, *Mathematical analysis*, Second Edition, Addison-Wesley Publishing Company, 1974.
- [2] D. Applebaum, *Lévy processes and stochastic calculus*, Second Edition, Cambridge University Press, 2009.
- [3] N.H. Bingham and R. Kiesel, *Risk-Neutral Valuation*, Springer Finance, 1998.
- [4] M. Chesney, M. Jeanblanc and M. Yor, *Mathematical methods for financial markets*, Springer Finance, 2009.
- [5] K.L. Chung, *A course in probability theory*, Third Edition, Academic Press, 2001.
- [6] P.K. Friz and N.B. Victoir. *Multidimensional stochastic processes as rough paths: theory and applications*, Cambridge University Press, 2010.
- [7] X. Geng, The Wiener-Itô chaos decomposition and multiple Wiener integrals, [Unpublished online notes](#), 2020.
- [8] K. Itô and H.P. McKean, *Diffusion processes and their sample paths*, Springer, 1965.
- [9] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*, Second Edition, North-Holland Mathematical Library, 1989.

- [10] I.A. Karatzas and S.E. Shreve, *Brownian motion and stochastic calculus*, Springer-Verlag, 1988.
- [11] T.J. Lyons, M.J. Caruana and Thierry Lévy, *Differential equations driven by rough paths*, *École d'Été de Probabilités de Saint-Flour XXXIV* Springer, 2007.
- [12] P. Malliavin and A. Thalmaier, *Stochastic calculus of variations in mathematical finance*, Springer Finance, 2006.
- [13] H.P. McKean, *Stochastic integrals*, American Mathematical Society, 1969.
- [14] J. Obłój, On some aspects of Skorokhod embedding problem and its applications in mathematical finance, Notes for the 5th European Summer School in Financial Mathematics, 2012.
- [15] S. Orey and S. J. Taylor, How often on a Brownian path does the law of iterated logarithm fail?, *Proc. London Math. Soc.* **28** (3) (1974): 174–192.
- [16] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, Springer-Verlag, 1991.
- [17] L.C.G. Rogers and D. Williams, *Diffusions, Markov processes and martingales*, Volumes 1 and 2, Second Edition, Cambridge University Press, 2000.
- [18] S.E. Shreve, *Stochastic calculus for finance II*, Springer, 2004.
- [19] I. Shigekawa, *Stochastic analysis*, American Mathematical Society, 2004.
- [20] D.W. Stroock and S.R.S. Varadhan, *Multidimensional diffusion processes*, Springer-Verlag, 1979.
- [21] D. Williams, *Probability with martingales*, Cambridge University Press, 1991.